UNLEASHING DATA POTENTIAL WITH DATA DIVINITY: FRAMEWORK FOR EFFICIENT FINTECH-BNPL DATA LAKE

by

Durga Rathinasamy, B.Tech

Under the Guidance of
Vijayakumar Varadarajan, PhD

DISSERTATION
Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

JULY, 2023

UNLEASHING DATA POTENTIAL WITH DATA DIVINITY: FRAMEWORK FOR EFFICIENT FINTECH-BNPL DATA LAKE

by

Durga Rathinasamy

APPROVED BY

Luka Lesko, PhD

<Chair's Name, Degree>, Chair

Iva Bulјubašić

<Member's Name, Degree>, Committee Member

<Vijayakumar Varadarajan, PhD>,
Mentor/ Committee Member

RECEIVED/APPROVED BY:

<Associate Dean's Name, Degree>, Associate Dean

Dedication

To my fellow researchers and industry colleagues, whose collaboration, discussions, and shared insights have enriched our understanding and propelled our research forward. Your intellectual camaraderie and collective pursuit of knowledge have been inspiring. This research paper reflects the collaborative efforts that have shaped my academic endeavors.

Acknowledgments

The completion of this research paper would not have been possible without the support and contributions of various individuals and organizations. I want to express our sincere gratitude to all those who have played a significant role in the successful completion of this study.

I am also indebted to the participants of this study, whose willingness to share their time, experiences, and perspectives has been instrumental in generating meaningful data. Their cooperation and involvement are deeply appreciated.

Each contribution has been invaluable, and I am genuinely grateful for the collaborative efforts that have made this study possible.

ABSTRACT

UNLEASHING DATA POTENTIAL WITH DATA DIVINITY: FRAMEWORK FOR EFFICIENT FINTECH-BNPL DATA LAKE

Durga Rathinasamy
2023

Dissertation Chair: <Chair's Name>
Co-Chair: <If applicable. Co-Chair's Name>

According to a recent Statista report by (Norrestad, 2022), the Fintech industry has experienced significant growth, with a 68% increase in Fintech startups from 2018 to 2021. Additionally, the global Fintech market is projected to grow at a compound annual growth rate (CAGR) of 23.58% from 2021 to 2025, as stated in the (*81 Key Fintech Statistics 2021/2022: Market Share & Data Analysis*, n.d.). This rapid growth is expected to lead to a surge in digital transactions within a short period. In order to effectively compete and make strategic business decisions with a wide range of insights, it is essential for Fintech companies to establish and maintain foolproof Cloud Data Lakes. However, building and managing a data lake has become problematic due to inappropriate management decisions, non-strategic approaches, technical solutions, and inefficient governance. To address these challenges and their impact on the Fintech business, this research paper aims to propose a conceptual framework and strategies for

identifying pain points in building a cloud data lake within the Fintech industry, specifically to BNPL - Buy-Now-Pay-Later. The proposed framework will offer a comprehensive 360º view for managing constraints in Fintech-BNPL dthe ata lake.

TABLE OF CONTENTS

List of Tables xii

List of Figures xiv

CHAPTER I: INTRODUCTION 1

1.1. Introduction 1
1.2. Overview 1
1.3. Research Problem 4
1.4. Research question and objectives 5
1.5. Significance of the Study 6
1.6. Summary 8

CHAPTER II: REVIEW OF LITERATURE 9

2.1. Literature Review on Data lake 9
2.1.1. Data Lake 9
2.1.2. Architecture 11
2.1.2.1. Overview of Data Lake Architecture 11
2.1.2.2. Framework of Data Lake Architecture 12
2.1.3. Cloud Computing and Cloud Data Lake 16
2.1.3.1. Cloud Computing Overview 16
2.1.3.2. Cloud Data Lake 19
2.1.4. Key Studies on the data lake 20
2.1.4.1. Metadata management framework for successful data lake 20
2.1.4.2. Criteria for evaluating a good data lake 26
2.1.4.3. Criticism of data lake 26
2.1.4.4. Multi-cloud and challenges 27
2.1.4.5. Data Security Implications with Cloud 28
2.2. Literature Review on Fintech 30
2.2.1. Fintech and its evolution 30
2.2.2. Fintech Verticals 32
2.2.3. Introduction to Payments and Lending 33
2.2.3.1. Overview of Payment and Lending in Fintech 34
2.2.3.2. Advancements in Payment and Lending Technologies 36
2.2.4. Key studies on Payments & lending 38
2.2.4.1. Challenges in the Payment and Lending Industry 39
2.2.4.2. Emerging Trends in Payment and Lending Technologies 40
2.2.5. Buy-Now-Pay-Later 42
2.2.5.1. Overview of Buy-Now-Pay-Later 42
2.2.5.2. Growth of Buy-Now-Pay-Later 43
2.2.6. Key studies on Buy-now-pay-later 47

2.2.6.1. Consumer Behavior in Buy-Now-Pay-Later 47
2.2.6.2. Implications of Buy-Now-Pay-Later on Debt Management 50
2.3. Chapter Summary 52
2.3.1. Summary of Literature Review on Data Lake and cloud computing 52
2.3.2. Summary of Literature Review on Fintech, Payments & Lending and Buy-Now-Pay-Later 54

CHAPTER III: METHODOLOGY 60

3.1. Overview of the Research Problem 60
3.1.1. Buy-Now-Pay-Later & Data lake 60
3.1.2. Integration of Buy-Now-Pay-Later and Data Lake 66
3.2. Research Design 68
3.2.1. Data Collection 68
3.2.2. Fintech-BNPL data lake canvas 70
3.2.3. BNPL data lake canvas validation 71
3.2.3.1. Cost optimization with BNPL data lake canvas 71
3.2.3.2. 'Data Maturity Model' with BNPL data lake canvas 73
3.2.3.3. Data as an Asset (DaaA) with BNPL data lake canvas 78
3.2.4. Interpretation of results 79
3.2.5. Research Design Limitations 80
3.2.6. Chapter Summary 81

CHAPTER IV: RESULTS 83

4.1. Research case 83
4.1.1. Research study 83
4.1.2. Research Participants 84
4.2. Factors Impacting the successful strategy for Fintech data lake 85
4.2.1. Descriptive Analysis of the Fintech – Buy Now Pay Later - Business 85
4.2.1.1. Fintech and BNPL Business 85
4.2.1.2. Fintech and BNPL Business – Survey Results 87
4.2.1.2.1. Fintech challenges 88
4.2.1.2.2. BNPL challenges 90
4.2.1.2.3. BNPL Risk Management 91
4.2.1.2.4. BNPL Customer Journey 92
4.2.1.2.5. BNPL Subject areas 93
4.2.1.2.6. BNPL cost management 94
4.2.1.3. Conclusion 96

4.2.2. Descriptive Analysis of the Engineering Aspects for BNPL Data lake 98
4.2.2.1. Data lake engineering 98
4.2.2.2. Data Engineering / Architecture Factors– Survey Results 102
4.2.2.2.1. BNPL data lake preference 102
4.2.2.2.2. BNPL data lake architecture preference 104
4.2.2.2.3. BNPL data lake data structure preference 104
4.2.2.2.4. BNPL data lake data variety and data model 106
4.2.2.3. Data lake ML factors – Survey Results 107
4.2.2.4. Conclusion 109
4.2.3. Influencing Factors for Business and Data lake Framework 110
4.2.3.1. BNPL cost management 116
4.2.3.2. BNPL Time to value 118
4.2.3.3. BNPL Time to Market 121
4.2.3.4. BNPL data quality 124
4.2.3.5. BNPL data security 127
4.2.3.6. BNPL data governance 131
4.3. Results from Fintech – BNPL & Engineering Survey 135
4.3.1. Fintech – BNPL Data lake canvas 135
4.3.1.1. Data enemies 142
4.3.1.1.1. Fintech – BNPL drivers 143
4.3.1.1.2. Data lake drivers 147
4.3.1.2. Data Divinity 151
4.3.1.2.1. Objective 1 - Cost Optimization using Fintech – BNPL Data lake canvas 151
4.3.1.2.2. Objective 2 - Data Maturity Model using Fintech – BNPL Data lake canvas 154
4.3.1.2.3. Objective 3 - Data as an Asset (DaaA) using Fintech – BNPL Data lake canvas 164
4.4. Chapter Summary 173

CHAPTER V: DISCUSSION 175

5.1. Discussions of findings from the questionnaire 175
5.1.1. Fintech - BNPL 175
5.1.2. Fintech and BNPL Business – Survey insights 176
5.1.2.1. Fintech challenges 176
5.1.2.2. BNPL challenges 178
5.1.2.3. BNPL Risk management 181
5.1.2.4. BNPL Customer Journey 183
5.1.2.5. BNPL Subject areas 185
5.2. Discussions of findings from the questionnaire – Data engineering 188
5.2.1. Data engineering 188

5.2.2. Data Engineering / Architecture factors– Survey Insights 188
5.2.2.1. BNPL data lake preference 188
5.2.2.2. BNPL data lake architecture preference 190
5.2.2.3. BNPL data lake data structure preference 196
5.2.2.4. BNPL data lake data variety and data model 197
5.2.3. Data lake ML factors – Survey Insights 201
5.3. Comments on responses 206
5.4. Discussions of findings on cost optimization 206
5.4.1. Discussion of results 206
5.5. Discussions of findings on the Data Maturity Model (DMM) 207
5.5.1. Discussion of results 208
5.6. Discussions of findings on Data as an Asset (DaaA) 211
5.7. Conclusion & Insights 211

CHAPTER VI: SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS 213

6.1 Summary 213
6.2 Implications and Recommendations for Future Research 213
6.3 Conclusion 214

APPENDIX A LIST OF TABLES 216

APPENDIX B LIST OF FIGURES 237

APPENDIX C INFORMED CONSENT FORM 241

APPENDIX D QUESTIONAIRE WITH CONSENT FORM 242

APPENDIX E INTERVIEW QUESTIONS - EVALUATION ON DATA LAKE CANVAS 259

APPENDIX F RESPONDENT RESULTS FOR FINTECH – BNPL INDUSTRY CHALLENGES 261

APPENDIX G RESPONDENT RESULTS FROM ENGINEERING EXPERTS ON SUCCESSFUL STRATEGY FOR FINTECH - BNPL DATA LAKE FRAMEWORK 271

APPENDIX H RESPONDENT RESULTS FOR FINTECH – BNPL DATA LAKE – INFLUENCING FACTORS 281

APPENDIX I DATA LAKE CANVAS 295

APPENDIX J EVALUATION OF RESULTS FOR DATA LAKE CANVAS 306

REFERENCES 315

LIST OF TABLES

Table 1: Fintech Verticals – Market scope and growth 1
Table 2: (ZHAO, 2021) - Metadata management systems for data lake 20
Table 3: (Chelliah and Surianarayanan, 2021) - Challenges in multi-cloud and respective solution approaches 27
Table 4: Phased growth of a BNPL 45
Table 5: BNPL Consumer behavior 47
Table 6: BNPL implications on debt management 50
Table 7: Summary of Objectives & Gaps in BNPL Data lake studies 63
Table 8: Research methodology 68
Table 9: Objective 1 - Hypothesis testing strategy 72
Table 10: Objective 2 - Hypothesis testing strategy 74
Table 11: Objective 3 - Hypothesis testing strategy 78
Table 12: Survey participants – Industry and Role 84
Table 13: Survey questionnaire – Fintech and BNPL Industry view on Data lake 85
Table 14: Survey questionnaire – Data engineering view on Data lake 98
Table 15: Survey questionnaire – Influencing factors on Data lake 110
Table 16: Fintech challenges 144
Table 17: BNPL Challenges 144
Table 18: Risk Management 144
Table 19: Customer journey 145
Table 20: Key subject areas 145
Table 21: BNPL cost increase factors 146
Table 22: BNPL cost-effective factors 146
Table 23: BNPL cost measurement factors 147
Table 24: BNPL Data lake preference 148
Table 25: Preferred scalable BNPL architecture 148
Table 26: Preferred data model for BNPL data lake 149
Table 27: Preferred data structure for BNPL data lake 149
Table 28: Critical elements for effective AI/ML 150
Table 29: Features stores expected in BNPL data lake 150

Table 30: Current Observation from the traditional data lake methods 152
Table 31: Expected state with data lake canvas – new responses 153
Table 32: Current DQI from the traditional data lake methods 154
Table 33: Expected DQI with data lake canvas – new responses 155
Table 34: Current DSI from the traditional data lake methods 156
Table 35: Expected DSI with data lake canvas – new responses 156
Table 36: Current DGI from the traditional data lake methods 158
Table 37: Expected DGI with data lake canvas – new responses 158
Table 38: Current TTM from the traditional data lake methods – TTM_1I 160
Table 39: Expected TTM with data lake canvas – new responses – TTM_1I 160
Table 40: Current TTM from the traditional data lake methods – TTM_2I 162
Table 41: Expected TTM with data lake canvas – new responses – TTM_2I 162
Table 42: Unit of Economies – Non-Fintech company 166
Table 43: Operational cost – New response with data lake canvas 207
Table 44: Data Quality – New response with data lake canvas 208
Table 45: Data Security – New response with data lake canvas 208
Table 46: Data Governance – New response with data lake canvas 209
Table 47: TTM_1I – New response with data lake canvas 210
Table 48: TTM_2I – New response with data lake canvas 210

LIST OF FIGURES

Figure 1: Realigned Fintech Model with data lake 4
Figure 2: Realigned Data Lake Organization 11
Figure 3: (Giebler, Corinna et al., 2021a) Data Lake Architecture Framework (DLAF) 13
Figure 4: (John and Misra, 2017) Layers in a Data Lake 13
Figure 5: Buy-now-pay-later business flow 42
Figure 6: Potential growth of BNPL in Online, In-store, and Mobile 45
Figure 7: Fintech data lake – Connect the Dots 62
Figure 8: Fintech business challenges 89
Figure 9: Fintech challenges grouped by role 89
Figure 10: Fintech challenges grouped by role and organization type 90
Figure 11: BNPL challenges 90
Figure 12: BNPL challenges grouped by role and organization type 91
Figure 13: Risk management factors grouped by Role and Organization type 92
Figure 14: Risk management factors grouped by Role and Organization type 92
Figure 15: Customer journey factors 93
Figure 16: Customer journey factors grouped by Role and Organization type 93
Figure 17: BNPL subject areas 94
Figure 18: BNPL subject areas grouped by Role and Organization type 94
Figure 19: BNPL cost increase 95
Figure 20: BNPL cost-effectiveness 96
Figure 21: BNPL data lake measuring factors 96
Figure 22: BNPL data lake preference 103
Figure 23: BNPL data lake preference grouped by Role and Organization type 103
Figure 24: BNPL data lake architecture preference 104
Figure 25: BNPL data lake data structure algorithm preference 105
Figure 26: BNPL data lake data structure importance 105
Figure 27: BNPL data lake structured/semi-structured data management 106
Figure 28: BNPL data lake unstructured data management 107
Figure 29: BNPL data lake preferred data model 107

Figure 30: BNPL data lake critical elements for AI/ML 108
Figure 31: BNPL data lake – Feature stores 109
Figure 32: BNPL data lake – Operational cost 118
Figure 33: BNPL data lake – Time to value 120
Figure 34: BNPL data lake – Lack of SME 121
Figure 35: BNPL data lake – TTM – Build from scratch 123
Figure 36: BNPL data lake – TTM – Migrate without modernization 124
Figure 37: BNPL data lake – TTM – Migrate with modernization 124
Figure 38: BNPL data lake – Data Quality areas 127
Figure 39: BNPL data lake – Data Quality issues 127
Figure 40: BNPL data lake – Data Security view by Fintech experts 129
Figure 41: BNPL data lake – Data Security view by Data engineering experts 130
Figure 42: BNPL data lake – Data Security drivers 130
Figure 43: BNPL data lake – Data Security issues 131
Figure 44: BNPL data lake – Garbage dump 134
Figure 45: BNPL data lake – Data Governance 134
Figure 46: BNPL data lake – Garbage dump reasons 135
Figure 47: BNPL data lake – data asset 135
Figure 48: Fintech-BNPL data lake canvas pillars 137
Figure 49a: Fintech – BNPL data lake architecture 138
Figure 49b: Fintech - BNPL data lake canvas 138
Figure 50: BNPL data lake canvas – Unit of Economies 141
Figure 49c: Fintech - BNPL data enemies identification framework 143
Figure 51: Operational Expense distribution 153
Figure 52: Data quality distribution 155
Figure 53: Data security distribution 157
Figure 54: Data governance distribution 159
Figure 55: TTM – Migration without modernization distribution 161
Figure 56: TTM – Build from scratch distribution 163

CHAPTER I:

INTRODUCTION

1.1. Introduction

Chapter 1 provides an overview of the Fintech and Data lake challenges. It also highlights the scope of the growing Fintech market, especially with Payments and Lending, and the significance of making data-driven business decisions more effectively and efficiently with data lake.

1.2. Overview

FinTech combines Finance and Technology, where the market share has increased massively. It has enabled growth in the various streams of FinTech, including payments, digital banking, Digital wealth management, Capital markets, Equity crowdfunding, InsureTech, and PropTech.

Fintech verticals in relevance to Big data and cloud computing are as below.

Table 1: Fintech Verticals – Market scope and growth

Fintech Vertical	Market Scope	Expected growth / Market share
Payments Technology	Disruption of the payments is due to the evolving need for contactless, real-time payments with enhanced features like BNPL, biometrics, and other latest technologies.	Digital payments will have a total transaction value of US$8,502.00bn in 2022, and users will be 4,929.55m users by 2025
Digital Banking	Cutting-edge technologies have made digital banking easier for Neobanks,	Market to Reach $30.1 Billion by 2026

	Challenger-banks, New-banks, and non-banks.	
Digital Wealth Management	Robo advisors using big data and analytics are radically changing wealth management in serving the HNI and UHNI investors.	The market projection will expand at a CAGR of 13.9% between 2022 and 2027 to reach a value of USD 10,268.9 million by 2027.
Capital Markets	Investors continue to embrace technology with the growth in crypto trading, algorithmic trading, HTF, AI/ML, and RPA for post-trade functions.	The global domestic equity market capitalization value is 116.78 trillion USD.
Fintech Lending & Equity crowdfunding	Fintech Lending is very popular with individuals and SMEs. On the other hand, equity crowdfunding is popular with start-ups and innovative products to raise funds. It has revolutionized the P2P lending/crowdfunding platform with ML, Big data & Analytics for a seamless lending/crowdfunding process and credit scoring mechanism.	The global peer-to-peer (P2P) lending market generated $67.9 billion in 2019 and will reach $558.9 billion by 2027, registering a CAGR of 29.7% from 2020 to 2027.
InsurTech	InsurTech has enabled seamless experience with digitization and automation with AI / ML & RPA for data collection, loss assessment, cost	The global insurtech market size will value USD 2.72 billion in 2020. It will expand at a compound annual growth

	estimation, damage analysis through image recognition, automated self-service guidance, and more.	rate (CAGR) of 48.8% from 2021 to 2028
PropTech	PropTech allows individuals and companies to make acquisition and disposal decisions and manage a real estate portfolio. Apart from AR / VR & IoT, drones modernizing real estate, Proptech converges many functions with Fintech. In the Fintech arena, Blockchain aids in data tracking and reaching immutable data in pricing and ownership rights and using Big data & Analytics for financial process management.	The value of the real estate tech deals worldwide is 8.4 billion USD from 2014 to 2020

On the technology end, data and analytics have taken a paradigm shift with AI & ML, and Data lake is the critical enabler for achieving this shift. Data lake, amalgamated with Cloud solutions, brings a new dimension to solving business solutions.

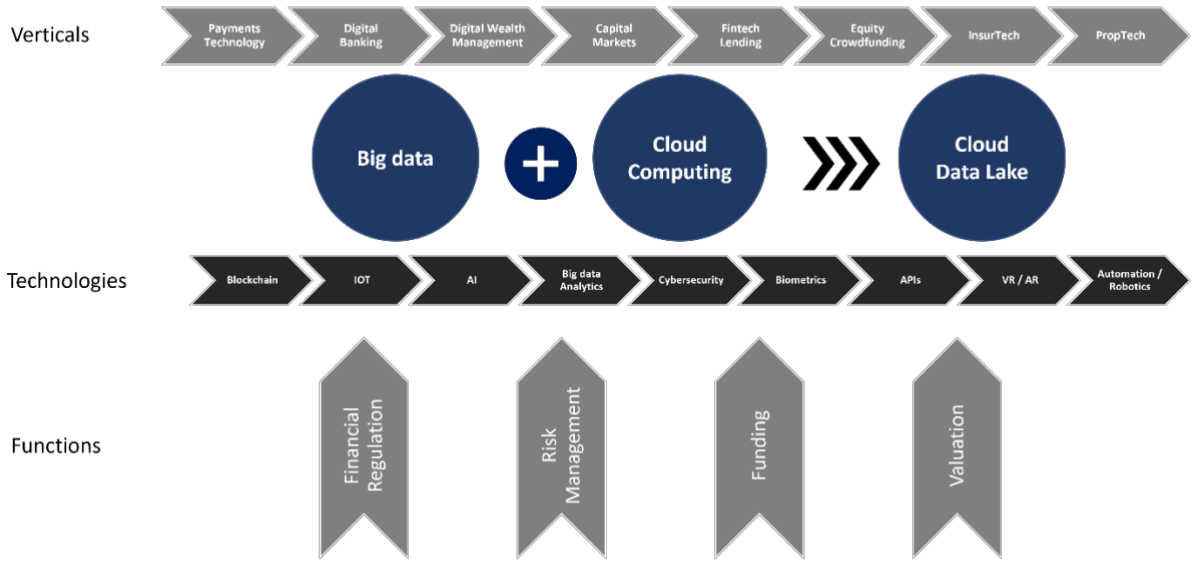


Figure 1: Realigned Fintech Model with data lake

Fintech's ascent has led to a significant market share, with widespread distribution of Fintech verticals. Most Fintech start-ups fall into the broad categories of Payments and Lending. "House of Debt" from Credit Suisse has predicted that lending will grow 5x in the next 10 years, going to $3 trillion from $600 billion. According to a report by Allied Market Research, the global BNPL market size values at $7.3 billion in 2019 and projects to reach $33.6 billion by 2027, growing at a CAGR of 21.2% from 2020 to 2027.

Considering the opportunity in the Lending market, Buy-Now-Pay-Later (BNPL) is disrupting traditional lending systems. Disruption with BNPL comes with its challenges & opportunities, which makes data a critical asset in business decisions. Hence the need for a successful strategy for building a BNPL data lake is inevitable.

1.3. Research Problem

The research study is to bridge the gap between the FinTech BNPL and data lake by building a strategy to build a successful Fintech data lake.

Managing data lakes in any organization, especially the Fintech BNPL, is cost intense. Traditional approaches considering only technology factors like on-prem vs. cloud solutions and real-time vs. batch leads to wastage of resources. The proposed framework will lead to cost-effective data lake management, allowing businesses to leverage 'Data as an Asset' (DaaA) more productively.

With slight modifications to the subject data requirements, it can utilize the framework to construct various Fintech data lake use cases.

1.4. Research question and objectives

The research is to understand - How to build a successful data lake in Fintech BNPL?

Hence the goal is to Build a BNPL – Fintech Data lake framework based on the industry's best practices and challenges. The purpose is to comprehensively analyze existing studies on challenges faced by BNPL and data lake, collect survey responses, and develop a framework to address those challenges.

The three main Objectives to achieve the goal includes,

- ***Objective 1*** – Testing the framework and generating comparative statistics with traditional approaches vs. the BNPL data lake framework for cost optimization. Develop conclusions and insights for Data leaders and Executive management for cost optimization, both approaches.
- ***Objective 2*** - Testing the framework and generating comparative statistics with traditional approaches vs. BNPL data lake framework for Data Maturity Model (DMM). Develop conclusions and insights for Data

leaders and Executive Management for DMM with traditional and new framework.

- ***Objective 3*** - Develop conclusions and insights for Data leaders and Executive management for defining 'Data as an Asset' via 'Data as a Service (DaaS) and 'Data as a Product' (DaaP).

1.5. Significance of the Study

The struggles the BNPL Fintech organizations face are multi-fold, including BNPL and data lake Challenges.

This research bridges the gap between the BNPL and data lake Challenges with the proposed framework for the BNPL Fintech data lake, favorable for cost optimization and building 'Data as an Asset' - with an efficient Data Maturity Model.

BNPL Challenges

1. Risk Management – It is the nucleus of the BNPL business. The risk management life cycle, from risk assessment, credit checks, transaction monitoring, collection policies, compliance, economic conditions, etc., is crucial for the business's success. Incorrect risk assessment & credit scoring, fraudulent transactions by impersonating, and not adhering to compliance and regulations brings a colossal impact P&L of the organization, including the closure of business.
2. BNPL Customers – Understanding the BNPL customers for acquisition, retention, and providing a seamless customer experience is essential. Given the competition in the lending market of BNPL, sales and marketing business functions will choke if Net Promoter Score (NPS) scores fail with customers.

3. BNPL subject data – Comprehension of the core data subjects of BNPL is essential to bring only the data required for making business decisions.

Data lake Challenges

1. Cost Management

   The cost to build the data lake with On-prem or cloud solutions is a challenge to control and manage. With the on-prem, the licensing cost on the CAPEX (Capital Expense), and the cloud solution, the cost on the OpEx (Operational Expense) is a constant challenge for the technology team to align with the financial budget of the company goals.

2. Data Swamp

   Often technology team tends to bring every other data assuming it to be an asset without understanding the business use case, or bring the data without being used for any business use case.

3. Time to Value

   The time between the data inception to generate data-driven business insights consumes more than the value of the business use case.

4. Data security

   Data vulnerability comes with a vast digital footprint, a risk that incurs reputational costs and penalties when PII / PI information gets leaked.

5. Data Governance

   With data security at stake comes data governance, crucial factors considered are data privacy, data security, data regulations as per the local bodies, data management, data lineage, and data quality. With heterogeneous data, the complexity of data governance is multifold.

6. Data engineering/architecture

   The multitude of technologies and tools for data storage and pipelines has made decision-making with data engineering and architecture much more complex.

The study offers to consider the 360 views of BNPL Business and BNPL data coupled with technology to build the data lake rather than blindsided by technology alone.

1.6. Summary

The BNPL market has huge opportunities with the increased adoption of online shopping, changing consumer expectations on flexible payment options, and expanding BNPL services into new markets. It will increase the BNPL scope with the e-commerce business and unravels into travel, health care, ed-tech, and other industries.

Using the BNPL data for risk management, customer experience, fraud protection, and economic impact prediction is inevitable. It showcases the need for data lake to align with BNPL business needs, the latest technology, and architecture requirements. BNPL data lake business needs are driven by Time to Market and cost associated with deriving 'Data as an Asset.'

The study aims to build a framework by evaluating the BNPL business drivers, associated data elements, and Data Engineering aspects of data lake.

CHAPTER II:

REVIEW OF LITERATURE

Chapter 2 provides an inclusive literature review on Fintech, Buy-now-pay-later (BNPL) in finance, and data lake in technology. It starts with data lake technology for making data-driven business decisions. It covers the data lake architectures and existing frameworks on the various aspects required for consideration for building the data lake. It also covers cloud computing and the factors considered for the data lake. The chapter then focuses on the general Fintech concept and its evolution, focusing on payments and lending, which is the trending Fintech arena. The chapter then focuses on BNPL, the businesses leveraging BNPL, and its challenges. Finally, the chapter concludes with a summary of how a data lake can help address BNPL challenges.

2.1. Literature Review on Data lake

2.1.1. Data Lake

Definition

According to (Hai, Quix, and Jarke, 2021), A data lake is a flexible, scalable data storage and management system which ingests and stores raw data from heterogeneous sources in their original format and provides query processing and data analytics in an on-the-fly manner.

Also (Hai, Quix, and Jarke, 2021) pose the differences between the data warehouse and data lake. Former serves structured ETL process for OLAP and OLTP use cases with schema on write approach. On the other hand, the latter supports the load-as-is of heterogeneous data supporting Bigdata 3Vs for various use cases, not limiting to OLAP and OLTP with a schema-on-read approach.

Various factors, including cost, SMB or Enterprise, time to market, and many more, determine whether to implement it as an on-premise or cloud data lake.

Characteristics of Data Lake

(Inmon, 2016) provides the details on how the data lake should be structured to avoid the garbage dump of the data. It has to be structured ponds to make it a usable data lake.

- Raw data pond
    - Most organizations manage the data lake with only the raw data, which ideally should be different. Also, the forever persistence of raw data should never happen. They are transient once they converge to any ponds based on the nature of the data.
- Analog data pond
    - These attribute to the data generated by the machine or any other device with ample metrics with repetitive structure.
- Application data pond
    - Application data pond attribute to data generated from an application with a repetitive structure from the transactional process. It happens whenever any business event is triggered and measured by the application.
- Textual data pond
    - Textual data pond attribute to unstructured data associated with an application with a non-repetitive structure.
- Archival data pond
    - Warm and cold data from the analog, application, and textual data ponds gets into the archival data ponds.

(Gorelik, 2019) has provided details on how the Enterprise will evolve with the data lake.

Figure 2 shows the realigned data lake organization below.

- Data Puddle
    - A Data puddle is a single-purpose data mart. It is the first step in developing and adapting a data lake.
- Data Pond
    - A Data Pond is a collection of data puddles with various classifications.
- Data lake
    - A Data lake enables the self-service data and aims to pull together all the information from the Enterprise.
- Data Ocean
    - Data Ocean enables self-service data and data-driven decision-making for all enterprise data.

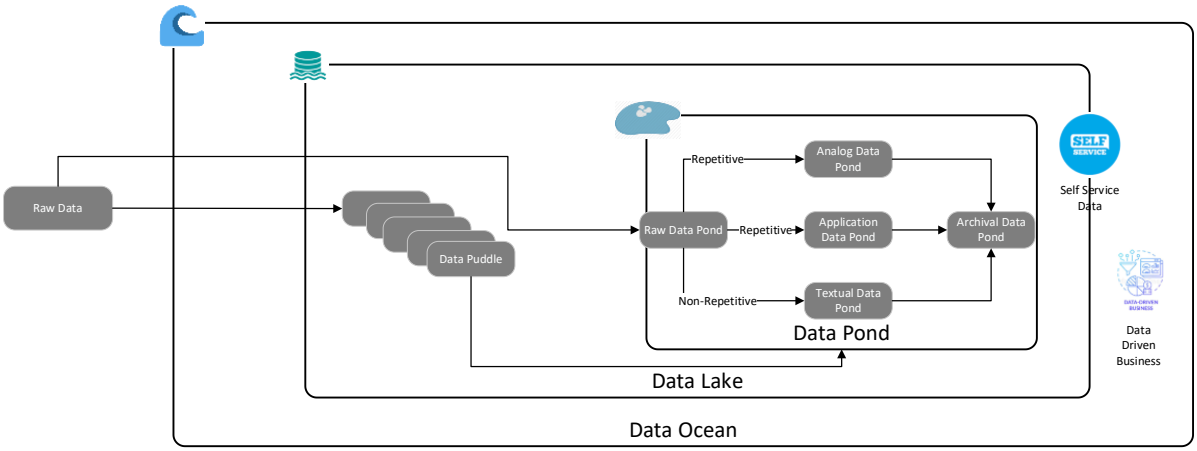


Figure 2: Realigned Data Lake Organization

2.1.2. Architecture

2.1.2.1. Overview of Data Lake Architecture

The topology of the data lake has been ambiguous since the data inception of data usage. Therefore, laying out an appropriate topology with various terminology while building the data lake takes much effort.

- Data Discovery

- Data Ingestion
- Data Integration
- Data Extraction
- Data Transformation
- Data Versioning
- Data Storage
- Data Model
- Data Access
- Metadata Management
- Data Quality
- Data Security
- Data Governance

Nevertheless, (Giebler, Corinna et al., 2021a) and (John and Misra, 2017) provide the framework and Lambda architecture to define the topology for the Data lake, as shown in Figure 3 and Figure 4.

2.1.2.2. Framework of Data Lake Architecture

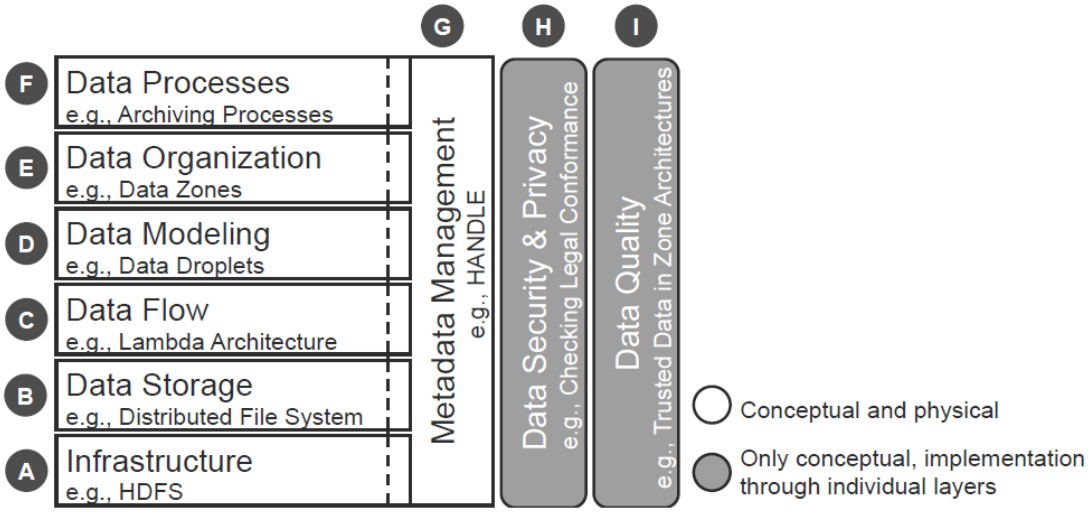


Figure 3: (Giebler, Corinna et al., 2021a) Data Lake Architecture Framework (DLAF)

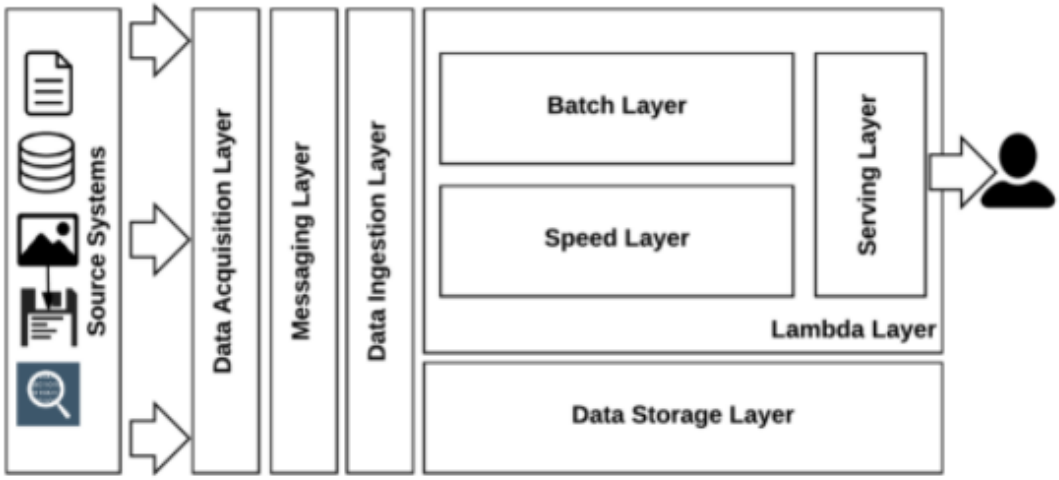


Figure 4: (John and Misra, 2017) Layers in a Data Lake

Let us review each of the blocks of Lambda architecture with the framework blocks embedded. Lambda architecture enables the scale-out architecture with high performance and low latency expectations, providing consistent data for both batches and real-time processing.

Data Acquisition Layer

Data discovery is a prerequisite for establishing this layer where the heterogeneous data and data sources should be supported. It should support structured, semi-structured, and unstructured data with faster connectivity with any source technology stack on-prem or cloud.

Messaging Layer

Messaging layer is responsible for the guaranteed delivery of messages with the Queues (point-point communication) and Topics (publish/subscribe) type of communication.

Data Ingestion Layer

The Data Ingestion layer is responsible for consuming the data with low latency to deliver the required transformed data to Lambda Layer. Data This layer handles integration and transformation before providing the data to the lambda layer.

Lambda Layer

Based on the business use case, the Lambda layer supports batch and real-time processing via the speed layer. Furthermore, it is available for the users via the servicing layer in either pull or push mode.

Data Storage Layer

The storage layer can be a relational datastore or distributed data store (NoSQL) based on the requirement. A stand-alone data model adapting a data lake with a data puddle is a good start for the structured data store. However, a unified data model is necessary to reach the maturity level.

Horizontal Integration

Below blocks cuts across the layers of the Lambda data lake architecture.

Data Infrastructure

With the dawn of big-data tools and technologies, the organization should perform concrete product capability analysis. Tools & technologies cut across the layers and depend on the requirements' nature, including storage, processing, database, compute, DevOps, and other factors.

Metadata Management

In a data lake, metadata is the structure information to describe all the data within the data lake.

It is essential to catalog the data for easy search and query on the data & schema with the vast datasets. According to (Cherradi et al., 2022), there are five significant functionalities for data management: Data Enrichment, Data Indexing, Data versioning, Number of descriptive variables, and Data accessibility. It also helps with data governance. (Weintraub et al., 2021) Provides a solution for indexing in the data lake for optimized query execution for the users to retrieve a specific record from the data lake.

Data Security & Privacy

Data security and privacy are essential for online businesses, legal communities, and policymakers. As per (Zouari et al., 2021), various techniques are available for data privacy in a data lake, like access control, anonymization, Blockchain, encryption, machine learning, and the definition of policies. Also, according to (Zouari et al., 2021) study, Blockchain is the most used technique for privacy preservation, and its use still needs to be improved in data lake space.

Data Quality & Governance

As per (Zouari et al., 2021), there are three data curation tasks: 1) Data curation, including data enrichment and data cleaning (dedupe, repair, and other cleansing processes). 2) Curation tasks related to metadata extraction and modeling 3) Curation tasks related to schemas like extraction, matching, mapping, and evolution.

2.1.3. Cloud Computing and Cloud Data Lake

2.1.3.1. Cloud Computing Overview

According to Gartner David Mitchell Smith's "The Gartner Hype Cycle for Cloud Computing." the cloud's future is distributed and ubiquitous, forming the foundation of the business. As described by (Akhtar et al., 2021), characteristics of the cloud include various computing models. Service models and storage models offered are as below.

Computing Models

Public cloud: Cloud providers offer computing infrastructure on their premises, and the customers can leverage the services via the Internet without maintaining them. Multiple tenants share the service however offers high scalability with a pay-as-you-go feature.

Private Cloud: It is the use by a company to host the infrastructure on its premises with a maximum level of protection and control.

Hybrid Cloud: It is a combination of both public cloud and private cloud. It typically hosts the business-sensitive and critical applications on its servers and the secondary applications in the cloud provider's location.

Multi-Cloud: Leveraging the multi-cloud service providers to de-risk and use multiple technology capabilities.

Edge Cloud: Decentralizing the computing power to the clients and devices at the network edge to lower the processing cost and have a low latency experience for customers.

Distributed Cloud: It is a public computing service that allows execution of public cloud in multiple locations (on-premises, other cloud service providers, third-party data centers) managed from a single site.

Community Cloud: It is for numerous customers to collaborate on community-owned projects and apps with centralized cloud infrastructure.

Service Models

IaaS (Infrastructure as a Service): Provides on-demand resources like compute, storage, and networking.

PaaS (Platform as a Service): The service provider offers access to hardware and software tools.

SaaS (Software as a Service): Provides fully functional software managed by the cloud provider.

XaaS (Everything as a Service): Refers to the wide range of products, tools, and technology.

UCaaS (Unified Communication as a Service): Provides communication needs like emails, video conferencing, and instant messaging with a mobile suite for organizations.

STaaS (STorage as a Service): Allows businesses to consume storage as needed for their multimedia, data repositories, data backup, and data recovery.

FaaS (Functions as a Service): Allows developers to create apps and 'Functions' with business logic.

SECaaS (SECurity as a Service): Provides full responsibility for the security to provide secured access to the apps and services.

TEaS (Test Environment as a Service): Provides an on-demand test environment for business to test their software and apps using a web browser.

CaaS (Communication as a Service): Provides a single vendor to handle all communication needs.

AIaaS (Artificial Intelligence as a Service): Offers the AI platform to allow businesses to implement and grow AI business approaches.

NaaS (Network as a Service): Enables secure access to network infrastructure.

DaaS (Desktop as a Service): Offers virtual desktops to customers.

DRaaS (Disaster Recovery as a Service): Provides the replication of physical and virtual servers for failover in any disaster.

MaaS (Monitoring as a Service): Provides the service to protect the assets 24 hours a day.

MDaaS (Model as a Service): Provides the service to run the simulation models

Decisioning on the service is based on the below factors (Wulf et al., 2021)

- Cost savings
- Strategic importance
- Reduced time to market
- Flexibility
- Access to specialized resources
- Focus on core competencies
- Security risks

Storage Models

The cloud offers both managed and unmanaged storage. In addition, it offers block storage, object storage, and file storage, which can be utilized based on business application needs.

Availability Zones

As mentioned by (Maurya et al., 2021), availability zones provided by the service providers offer flexibility with the data center setup, which is independent.

Pricing

Cloud service providers operate with Opex with a 'Pay-as-you-go' flexible pricing model.

2.1.3.2. Cloud Data Lake

A cloud data lake is a large, centralized repository of raw data stored on a cloud-based platform such as Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform. The data in a cloud data lake is typically stored in its original format and is not preprocessed or structured. Data scientists and analysts can easily access, process, and analyze it.

Cloud data lake design can handle large volumes of structured, semi-structured, and unstructured data, such as log files, sensor data, social media feeds, and other data generated in real-time. The essential advantage of a cloud data lake is that it enables organizations to store and manage vast amounts of data at a low cost while providing easy access for analysis and processing.

One of the primary benefits of using a cloud data lake is that it provides a highly scalable and cost-effective way to store and manage data. Cloud data lake design can be highly scalable, so they can easily accommodate large volumes of data as they grow over time. Additionally, the cloud's pay-as-you-go pricing model lets organizations pay only for

their storage and computing resources when they build cloud data lake on cloud-based infrastructure.

Overall, cloud data lakes provide a powerful and flexible platform for organizations to store and manage large volumes of data while providing easy access for analysis and processing.

2.1.4. Key Studies on the data lake

2.1.4.1. Metadata management framework for successful data lake

Metadata management becomes crucial with the data lake characteristic of schema-on-read, data from various modes, and heterogeneous data sources. Hence this is tedious and should have proper governance to avoid the data swamp. (ZHAO, 2021) provides various metadata management systems for data lake from various research papers, as shown in Table 2.

Table 2: (ZHAO, 2021) - Metadata management systems for data lake

S.No	Researcher	Metadata Management Framework	Description	Metadata Management System	Implementation Strategy
1	Walker and Alrehamy (2015)	Personal data lake	Framework has unified data storage and metadata management to help users analyze and query personal metadata.	Integrated with personal data lake	JSON Object Graph databased with four nodes for metadata, raw data, Semantics, and identified.

2	Hai et al. (2016)	Constance Data Lake	The framework helps users to discover, extract and summarize metadata from structured and semi-structured data (relational databases, JSON, spreadsheets, and XML).	Integrated with constance data lake	Extract explicit and implicit metadata and semantic annotations Cluster the schemas based on the distance between them
3	(Quix et al., 2016)	GEMMS (Generic and Extensible Metadata Management System)	GEMMS extracts metadata automatically from the datasets the data lake repository loads in their original format.	Built with the persistence component	MongoDB handles three main functions 1) metadata extraction, 2) Metadata transformation, and 3) Metadata Storage.
4	(Halevy et al., 2016a,b)	GOODS (Google Dataset Search)	The post-hoc system organizes the datasets that Google generates and uses. The post-hoc manner allows the system to collect and aggregate	integrated with the GOODS system	It has four principal services that include 1) Search engine, 2) per-dataset profile, 3) monitoring service, 4) Annotation service

			metadata about datasets after the creation, access, or updating of different pipelines without interfering with dataset owners' or users'		
5	Suriarachchi and Plale (2016b)	Data Lake of Suriarachchi	They focus on data provenance metadata when they transform processes for data lakes. A data product is stored to get the data lineage from the metadata.	Not a complete metadata management system	Ingest API and graph database.
6	Maccioni and Torlone (2018)	KAYAK Framework	KAYAK helps data scientists to define, execute and optimize data preparation pipelines in a data lake.	Not available	

7	Diamantini et al. (2018)	Metadata model of Diamantini	Metadata model that can structure unstructured data to extract thematic views from heterogeneous and generally unstructured data sources	Not available	
8	Sawadogo et al. (2019)	MEDAL (MEtadata model for DAta Lakes)	Provides evaluation criteria through a list of features for data lake metadata systems and a metadata typology	Not available	
9	Eichler et al. (2020)	HANDLE	HANDLE is a generic model for metadata which enables comprehensive metadata management.	Not available	Neo4j graph DB

10	Scholly et al. (2021)	goldMEDAL	Provide a generic metadata model and enable the data lineage tracing with the concept of process	Applied to 3 data lake systems	1. HOUDAL (Public Housing Data Lake) - Metadata with Neo4j DB 2. AUDAL, which is a textual and tabular data lake – stores documents in MongoDB and Metadata with Neo4j DB 3. Archaeological Data Lake - Implemented with Apache Atlas framework
11	(Zhao, Megdiche, et al., 2021)	Metadata model of Zhao et al., 2021	Provides a metadata model for ingestion with the IoT data lake using the Neo4j framework.	Integrated with IoT data lake	1. Metadata management with Neo4j to support both batch and streaming raw data 2. Front-end application built with HTML/CSS and JavaScript to perform data analytics 3. The back-end uses API Neoviz for visualizing the metadata with data lineage enabled.

12	(Zhao, Ravat, et al., 2021)	Analysis-oriented metadata model	Provides a metadata model related to machine learning analysis on the description information of datasets and their attributes	Applied to the data lake	1. Manage the metadata to be applied across phases of 1) Data exploration, 2) Data preparation, 3) Modeling, and algorithms 2. Metadata storage with Neo4j DB

2.1.4.2. Criteria for evaluating a good data lake

(Sawadogo and Darmont, 2021a) and (Nargesian et al., 2019a) highlight the factors required to build a successful data lake.

Benefits of Data Lake:

- Cheap storage
- Data Durability
- Preservation of Data fidelity as original data
- Heterogenous data
- Schema-on-read
- Flexibility and agility with a broader range of analysis
- Real-time data ingestion
- High scalability
- Fault tolerance
- On-the-fly analysis
- Discover links, correlations, and cross-analysis between heterogeneous data.

2.1.4.3. Criticism of data lake

(Sawadogo and Darmont, 2021a) and (Nargesian et al., 2019a) highlight the data lake's factors with the data evolution as the new oil.

- Data inconsistency due to data integration from multiple sources without any transformation
- Lack of methodological and technical Standards to manage heterogeneous data
- Requires an efficient metadata system for ensuring data access for schema-on-read
- The incompleteness of a comprehensive schema or data catalog

- Versioning is a cross-cutting concern over all stages of a data lake
- Expertise in performing on-the-fly analysis
- Lack of SMEs (Data scientists and others.)

2.1.4.4. Multi-cloud and challenges

With Enterprise applications relying on cloud platforms, the platform should accustom to the latest cloud trends and any innovation related to technology or industry. Hence, multi-cloud is an essential shift that organizations should consider, hence the data security.

Multi-Cloud Paradigm

As per Gartner, 76% of Enterprises use multiple cloud providers; hence, it is an essential aspect of the cloud journey of an organization.

(Chelliah and Surianarayanan, 2021) provides the critical drivers for the multi-cloud paradigm below.

- Fast-paced cloud journey
- Fast-paced cloud tool ecosystem
- Varied cloud service offerings that are value-adding
- Powerful platforms for cloud environments
- Smart networks
- Affordability
- Distributed and decentralized computing models
- Breakthrough digital technologies

Furthermore, the implementation comes with the challenges by (Chelliah and Surianarayanan, 2021), and the possible solutions as shown in Table 3.

Table 3: (Chelliah and Surianarayanan, 2021) - Challenges in multi-cloud and respective solution approaches

#	Challenge	Solution approaches
1	Interoperability and portability(Nogueira, E et al., 2016)	(i) Open APIs and Standards (ii) The Open Cloud Computing Interface (iii) Automation (iv) DevOps through CI/CD pipeline (v) Infrastructure as Code (IaC) (vi) Microservices Architecture (vii) Spinnaker for Multi-Cloud Software Delivery (viii) Containerization (ix) Serverless computing and management across Multiple Clouds (x) Service Resiliency Frameworks and Libraries for Multi-Clouds (xi) Service mesh orchestration
2	Application & data integration(Senda Romdhani, 2019)	Application and Data Integration Platforms
3	Multi-cloud orchestration(Ming Lu et al., 2018)	(i) intelligent brokers (ii)Container Clustering and Orchestration Platforms
4	Multi-Cloud Monitoring, Measurement and Management(E. Rios et al., 2016)	(i) Multi-Cloud Management and Governance Platforms (ii) Multi-Cloud Monitoring and Measurement Tools (iii) AI-Inspired Log and Operational Analytics Platforms for Multi-Clouds
5	Identity and Access Management(I.Indu et al., 2018)	(i) Next-Generation Identity and Access Management (IAM) Solutions (ii) Edge Cloud Integration with Traditional Clouds (iii) Multi-cloud security

2.1.4.5. Data Security Implications with Cloud

The cloud offers numerous tools and technology with which data governance and data breaches lead to cybercrimes. Access control is highly critical when outsourcing data to the cloud happens. As mentioned by (Akhtar et al., 2021), data breaches occur in many ways, including

- Service hijacking (phishing, fraud),
- data leaks due to sharing and no proper data destruction,
- DDoS - Distributed Denial of Service attacks,
- DoS - Denial of Service attacks,
- Cryptojacking,
- Cross Virtual Machine (VM) Side-Channel Attacks,
- Snooping,

- Trojan horse,
- Shared Technology Vulnerabilities, and many more.

(Akhtar et al., 2021) also, provide various cloud offers on security via

- consistent security updates,
- encryption mechanisms,
- formal change control process,
- access control,
- attribute-based encryption,
- Homomorphic encryption,
- Secure Data Destruction,
- Multi-Authority Attribute-Based Encryption (MA-ABE),
- encryption of backups,
- encrypted search and database,
- Built-in firewalls,
- Hierarchical Attribute Set Based Encryption (HASBE),
- AI Tools and auto-patching,
- Written security policies plan,
- redundancy (ultra-backed-up data),
- ranked keyword search,
- Proper usage of administrative privileges,
- key management strategy,
- data concealment,
- strategies for a secure transition to the cloud,
- Third-party security testing,
- Reliability of hard drive,

- enable two-factor authentication,
- Deploy Multi-Factor Authentication (MFA)

2.2. Literature Review on Fintech

2.2.1. Fintech and its evolution

Definition of Fintech

Fintech combines Finance and Technology with cutting-edge digital technology to improve financial services. Fintech has been in use since the 19th century, and the context of the term 'Fintech' has been used relative to the time period. It represented Fedwire transactions, SWIFT transactions, and the online revolution.

Fintech companies leverage technology to offer innovative and more efficient solutions to traditional financial services such as banking, insurance, investments, and payments.

Fintech encompasses various financial products and services, including mobile banking, digital payments, peer-to-peer lending, robo-advisory services, crowdfunding, cryptocurrency, and other evolving innovations. Often Fintech solutions are designed to be user-friendly, accessible, and cost-effective, making financial services more accessible to a wider range of people.

The rise of Fintech has disrupted the traditional financial industry, and many Fintech companies are competing with established financial institutions to offer more innovative and customer-centric solutions. The growing popularity of Fintech has also led to increased investment in the sector, as well as partnerships and collaborations between Fintech companies and traditional financial institutions.

Brief history of Fintech

The introduction of credit cards and the development of automated teller machines (ATMs) in the 1950s and 1960s marked the beginning of Fintech.

However, the term "Fintech" did not become widespread until the early 21st century, when new technologies such as mobile devices, cloud computing, and Blockchain became financial services tools.

In the early days of Fintech, the focus was primarily on automating manual processes and improving the efficiency of existing financial services. In the 1990s, online banking and brokerage services emerged, allowing customers to manage their finances online.

The 2008 financial crisis marked a turning point for the Fintech industry, as consumers became more wary of traditional financial institutions and sought more innovative and accessible financial services. It led to the rise of peer-to-peer lending platforms, mobile banking apps, and digital payment services, which offered consumers a more convenient and cost-effective alternative to traditional banking services.

The Fintech industry has expanded recently, with new technologies such as artificial intelligence and machine learning applied to financial services. Fintech has also become a primary focus of investment, with venture capital firms and other investors pouring billions of dollars into the sector. Today, Fintech is a major disruptor of the traditional financial industry, and many established financial institutions are partnering with Fintech companies to offer more innovative and customer-centric services. (Scardovi, 2016) shows how Fintech changes consumer journey across all aspects of life digitally.

‘Fintech’ term gained its traction in the 21st century with the advanced technologies coming into existence. The current revolution is with varied technologies, including

Cryptocurrency, Blockchain, machine learning, Artificial Intelligence, Big-data, and cloud computing across various verticals, including Open banking, Neo-banking, Payments as a service, Banking as a service, and lending.

2.2.2. Fintech Verticals

Fintech evolution is mainly due to cost, efficiency, low latency, flexibility, simplicity, and innovation. These factors have brought either a new business model or a business transformation with the existing model.

Fintech is a broad and diverse industry encompassing various financial products and services. Some of the critical Fintech verticals include:

1. *Payments and money transfer:* This includes digital payment solutions, such as mobile wallets, peer-to-peer payments, and online payment platforms.
2. *Lending and financing:* It includes alternative lending platforms, such as peer-to-peer lending, crowdfunding, and invoice financing, as well as digital banking and lending services.
3. *Personal finance and wealth management:* It includes robo-advisory services, digital wealth management platforms, and financial planning tools.
4. *Insurance*: This includes digital insurance platforms, such as insurtech startups, that offer automated underwriting and claims processing, as well as new insurance products and services.
5. *Blockchain and cryptocurrency:* This includes blockchain-based financial solutions, such as cryptocurrency exchanges, digital wallets, and decentralized finance (DeFi) platforms.
6. *Regtech and compliance*: This includes solutions that use technology to automate

and streamline regulatory compliance and risk management, such as Know Your Customer (KYC) and Anti-Money Laundering (AML) solutions.

7. *Banking and financial infrastructure:* This includes technology platforms that provide infrastructure and support for financial institutions, such as banking-as-a-service (BaaS) platforms, core banking systems, and payment gateways.

Fintech verticals are emerging and evolving as the industry grows and matures.

2.2.3. Introduction to Payments and Lending

With the various verticals of Fintech, the majority of the startups fall under the scope of Payments and Lending due to their significant market share. Hence the literature review focuses on payments and lending, specifically with Buy-now-Pay-Later.

Payments and lending are two critical verticals within the Fintech industry that have experienced significant growth and innovation in recent years.

Payments:

The payments vertical encompasses various digital payment solutions that enable individuals and businesses to send and receive money electronically. It includes mobile wallets, peer-to-peer payment platforms, online payment gateways, and digital currencies like cryptocurrency.

Fintech payment solutions aim to provide faster, more efficient, and more cost-effective than traditional payment methods such as cash or checks. Many Fintech payment platforms offer additional features such as rewards programs, budgeting tools, and fraud protection.

(Scardovi, 2017) highlights payments industry disruption in developing paperless societies and the progressive dematerialization of cash.

Lending:

The lending vertical encompasses a range of alternative lending platforms that use technology to connect borrowers with investors or lenders. It includes peer-to-peer lending platforms, crowdfunding platforms, and online lending platforms.

Fintech lending platforms offer borrowers an alternative to traditional banks and financial institutions, providing faster approval times, more flexible terms, and often lower interest rates. For investors and lenders, Fintech lending platforms offer new investment opportunities and the potential for higher returns.

Payments and lending are two critical areas of innovation within the Fintech industry, with new technologies and business models continuing to emerge and disrupting traditional financial services.

2.2.3.1. Overview of Payment and Lending in Fintech

Payments

The various payment players have evolved with the payment disruption, which includes,

Payment Service Provider (PSP)

Service providers help merchants to accept debit/credit payments, e-wallets, ACH transfers, and payment modes. PSPs are intermediate actors between customers, merchants, card networks, and banks. PSPs have enabled several payment methods, cross-border transactions with numerous currencies, process secured transactions, fraud protection, merchant reports, and account opening for merchants with acquiring banks - E.g., PayPal, and Stripe.

Payment Gateway

It is the technology merchants use to accept any debit card or credit card purchases from customers by collecting and verifying the customer's card information. Payment gateways deal with only online transactions, i.e., ec-commerce and card-not-present transactions. E.g., Stripe, Adyen, HDFC, ICICI, Razorpay.

Payment Aggregator

Merchant aggregator is a third-party payment solution offering merchant onboarding service using a Merchant Identification Number (MID) without an individual merchant account with a bank or financial services provider—E.g., Billdesk, PayU, Innoviti, PayPal, Gpay, Amazon Pay.

Payment Processor

Service that is responsible for facilitating the transactions between the merchant, issuing bank, and acquiring bank. POS transactions leverage the payment processor - E.g., Stripe, PayPal, Square.

Payment Facilitator

PayFac is a service provider for merchant accounts via two options: a Payment Service Provider (PSP) or an Independent Sales Organization (ISO).

Lending

Lending is the pillar of any financial institution, including Banks, NBFC, or Fintechs across SMBs and Enterprise, forming the core of the transaction life cycle. Lending practices and policies significantly impact the country's economy and globe, which accelerates the growth or is the peril of a recession.

Fintech has created new lending models and streamlined traditional lending processes in the lending segment, making loans more accessible and affordable to borrowers. Fintech

lending platforms include peer-to-peer lending, crowdfunding, and online lending platforms.

Peer-to-peer lending platforms connect borrowers with individual investors, bypassing traditional banks and financial institutions. Crowdfunding platforms allow individuals and businesses to raise funds from many investors through online platforms. Online lending platforms provide borrowers access to funds through a fast and convenient online application process.

Overall, Fintech has transformed how people make payments and access credit, making financial services more accessible, affordable, and efficient. As Fintech continues to evolve, new payment and lending solutions will emerge, providing consumers and businesses with even greater convenience and flexibility.

With the increased digital payments footprint, Fintech's Lending industry has evolved to a greater extent. The disruption involves investor loans, business loans, mortgages, and P2P lending.

When the credit card penetration is low, Buy-Now-Pay-Later offerings in lending Fintech reach out to a more extensive customer base. Also, BNPL now offers the smaller ticket transactions that used to be the target before.

2.2.3.2. Advancements in Payment and Lending Technologies

The Fintech industry has witnessed significant advancements in payment and lending technologies in recent years. Some of the key developments in these areas include:

1. *Mobile payments*: The rise of mobile payments has been one of the most significant advancements in payment technologies. Mobile wallets, such as Apple Pay, Google Wallet, and Samsung Pay, allow users to store their payment

information on their smartphones and make purchases by simply tapping their device at a point-of-sale terminal.

2. *Peer-to-peer payments*: Peer-to-peer payment platforms, such as Venmo, PayPal, and Cash App, have become increasingly popular, allowing users to send and receive money from one another through a mobile app.

3. *Blockchain and cryptocurrencies*: The emergence of Blockchain and cryptocurrencies, such as Bitcoin, Ethereum, and Ripple, has disrupted traditional payment systems by enabling secure, decentralized transactions.

4. *Contactless payments*: The COVID-19 pandemic has accelerated the adoption of contactless payments, such as NFC-enabled cards and mobile wallets, as consumers seek to minimize physical contact with surfaces.

5. *Digital lending platforms*: Digital lending platforms, such as LendingClub, Prosper, and SoFi, use technology to match borrowers with investors or lenders, providing faster approval times and more flexible loan terms than traditional banks.

6. *Crowdfunding platforms:* Crowdfunding platforms like Kickstarter and Indiegogo enable individuals and businesses to raise funds from many investors through online platforms.

7. *Open banking*: Open banking allows customers to share their banking data securely with third-party providers. It has created opportunities for Fintech companies to develop innovative payment and lending solutions.

These advancements have transformed the payments and lending landscape, providing consumers and businesses with faster, more convenient, and more accessible financial

services. As Fintech continues to evolve, new payment and lending technologies will likely emerge, creating even more excellent opportunities for innovation and disruption in the industry.

2.2.4. Key studies on Payments & lending

Digital is changing the consumer journey across the payment's life cycle (Scardovi, 2017). It has changed drastically, especially post-pandemic. Fintech has created 'new economy' even with the burst of the DotCom financial bubble.

Disruption of the Payments industry is with the four main innovation trends: new intermediaries entering the payments market, new payments rail, technology-enabled customer discovery in the front-end, and potential obsolescence of traditional back-end technology. (Scardovi, 2016)

Front-end innovations improve the process for clients and merchants via open-loop, streamlined, and closed-loop payment solutions. (Scardovi, 2016)
Payment gateway is the core in the payments (Gulati and Srivastava, 2007) discuss payment industry disruption in e-commerce, e-governance, and e-procurement and the gateway can help to boost the economy. (Gulati and Srivastava, 2007) proposes a National internet e-commerce payment gateway that can support all banks and transactions.

P2P Peer-to-Peer Lending has rapidly gained market share in consumer and SMB lending since the financial crisis, becoming a significant segment for credit supply. The disruption with Fintech lending industry is due to Mortgage lending and shadow banks, P2P lending, digital instruments. P2P platform substitutes as well as complement banks. (Agarwal and Zhang, 2020)

Post the subprime lending crisis 2008, tightened regulations increased the burden on the banking business with the immediate effect of reducing the accessibility to loans. It has enabled alternative lending models using cutting-edge technology for credit scoring and predictive analytics. Funding circle business model where the clients benefit from lower funding costs by direct matching with small businesses. SoFi business model targets students offering job placement and career advice services. Kabbage business model targets small e-commerce businesses for quicker and easier access to credit. (Scardovi, 2017)

2.2.4.1. Challenges in the Payment and Lending Industry

Like any other industry, the payment and lending industry faces several challenges. Some of the key challenges include:

1. *Regulatory compliance*: The payment and lending industry is highly regulated, and Fintech companies must comply with a range of regulations at the national and international levels. Compliance with these regulations can be complex and time-consuming, and compliance can result in significant penalties.

2. *Cybersecurity*: The payment and lending industry is a prime target for cybercriminals due to the sensitive financial information transmitted and stored by payment and lending platforms. Fintech companies must invest heavily in cybersecurity to protect their systems and customers.

3. *Fraud prevention*: Fraud is a significant challenge in the payment and lending industry, and Fintech companies must employ advanced fraud prevention technologies to identify and prevent fraudulent transactions.

4. *Customer acquisition and retention*: The payment and lending industry is highly competitive, and Fintech companies must invest in customer acquisition and

retention strategies to attract and retain customers.

5. *Payment infrastructure*: The payment infrastructure in many countries must be updated and more cohesive, making it difficult for Fintech companies to provide seamless payment services. Fintech companies must therefore invest in building robust payment infrastructure or partner with traditional financial institutions to offer their services.

6. *Credit risk*: Lending is inherently risky, and Fintech companies must carefully assess credit risk to ensure that they lend to borrowers who are likely to repay their loans. Fintech companies must also develop effective collection strategies to recover any unpaid loans.

These challenges require Fintech companies to invest heavily in technology, talent, and infrastructure to provide safe, reliable, and accessible payment and lending services to their customers.

The existing mobile payment system functions within an intricate and diverse network, utilizing a complex infrastructure. It engages in competition to generate and provide value to its customers. Moreover, the utilization of multiple virtual currencies in online transactions raises concerns about potential risks to the financial system. It is primarily due to regulators' challenges in effectively supervising and monitoring these currencies. (Suryono et al., 2020)

2.2.4.2. Emerging Trends in Payment and Lending Technologies

The Fintech industry is constantly evolving, and several emerging trends in payment and lending technologies are shaping the future of financial services. Some of the key trends include:

1. *Embedded finance*: Embedded finance refers to integrating financial services into non-financial platforms such as e-commerce marketplaces, social media, and ride-hailing apps. This trend will likely increase the adoption of digital payments and lending solutions by making them more convenient and accessible to consumers.

2. *Buy now, pay later*: Buy now, pay later (BNPL) solutions are becoming increasingly popular, especially among younger consumers. BNPL providers, such as Affirm, Afterpay, and Klarna, allow customers to pay for their purchases in installments, often without interest or fees. This trend is likely to continue as consumers seek more flexible payment options.

3. *Digital wallets*: Digital wallets are becoming more sophisticated, offering users a range of features such as rewards programs, budgeting tools, and personalized offers. Fintech companies are also exploring digital wallets for cross-border payments, further expanding their potential use cases.

4. *Decentralized finance*: Decentralized finance (DeFi) refers to a range of financial applications built on blockchain technology. DeFi platforms offer users a range of financial services, including lending, borrowing, and trading, without the need for intermediaries such as banks. This trend is likely to increase as blockchain technology becomes more widely adopted.

5. *Contactless payments*: Contactless payments have become more popular during the COVID-19 pandemic, and this trend is likely to continue as consumers seek more hygienic payment options. Fintech companies are also exploring biometric authentication, such as facial recognition and fingerprint scanning, further to enhance the security and convenience of contactless payments.

These emerging trends will likely drive continued growth and innovation in the Fintech industry, providing consumers and businesses with even greater convenience, flexibility, and accessibility in their financial services.

2.2.5. Buy-Now-Pay-Later

2.2.5.1. Overview of Buy-Now-Pay-Later

With the evidence that BNPL is an emerging trend with payments and lending technologies, the research focuses on BNPL and the related literature review.

Buy-Now-Pay-Later is a lending service offered to customers to purchase goods and services with interest and fees in installments, as shown in Figure 5. BNPL is gaining more traction as it enables one to purchase with no upfront cost and then pay in EMI. It is an attractive option for those who need help to afford to pay a substantial upfront amount.

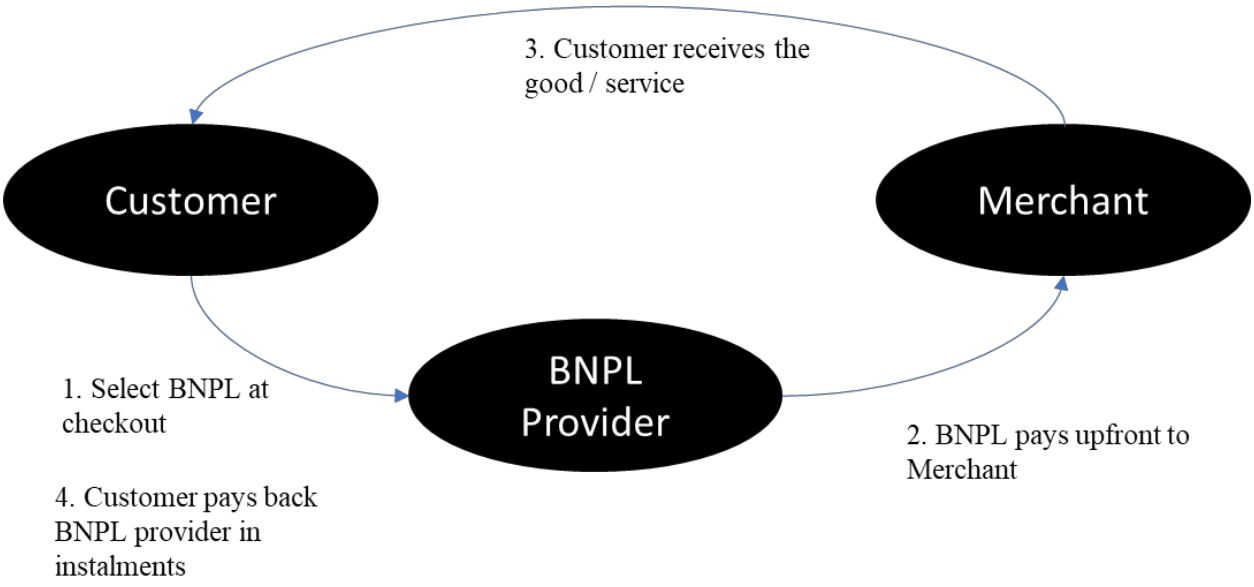


Figure 5: Buy-now-pay-later business flow

Installment options vary from provider to provider. However, it falls under either of the four models, namely Card Linked Financing (CLF), Off-Card Financing (OCF), Virtual Rent to Own Models (VRTO), and Integrated Online Shopping Apps (IOSA).

BNPL can be either for B2B or B2C customers. In either of the business models, customer journey, customer experience, credit scoring, and risk management are the core BNPL business drivers.

BNPL providers typically make their money by charging merchants a percentage of the purchase price and paying interest and fees to consumers who do not pay on time. Some providers also generate revenue by offering additional financial services like credit reporting and identity verification.

While BNPL has many benefits, such as convenience and flexibility, it can also pose risks to consumers who may overspend or need more clarification on the terms and conditions of their loans. Many BNPL providers have implemented responsible lending practices, such as credit checks and affordability assessments, to mitigate the risks. They have partnered with financial education providers to help consumers understand their financial obligations better.

BNPL is a rapidly growing payment method changing how consumers shop and pay for goods and services. As the industry continues to evolve, it will be necessary for BNPL providers to maintain responsible lending practices and to work closely with regulators to ensure that they are meeting their obligations to consumers.

2.2.5.2. Growth of Buy-Now-Pay-Later

The buy-now-pay-later (BNPL) industry has experienced tremendous growth in recent years, particularly in the e-commerce sector. The growth can attribute to several factors:

1. *Convenience*: BNPL offers consumers a more convenient payment option than traditional credit cards or loans. It allows them to spread payments over time and avoid significant upfront costs.

2. *Accessibility*: BNPL providers have made it easier for consumers to access credit, particularly those who may not have access to traditional credit cards or loans due to their credit history or income.

3. *User-friendly*: Many BNPL providers offer a seamless and user-friendly checkout experience, making it easy for consumers to use their online services.

4. *Marketing strategies*: BNPL providers have invested heavily in marketing their services, particularly on social media platforms, which has helped to raise awareness and drive adoption.

5. *Pandemic-driven online shopping:* The COVID-19 pandemic has accelerated the shift to online shopping, and BNPL has become an attractive payment option for consumers looking for more flexibility when purchasing online.

According to a report by Allied Market Research, the global BNPL market, value at $7.3 billion in 2019. The expected growth is to reach $33.6 billion by 2027, growing at a CAGR of 21.2% during the forecast period. The report attributes the growth to the increasing adoption of e-commerce and the rise in smartphone usage, particularly in emerging markets.

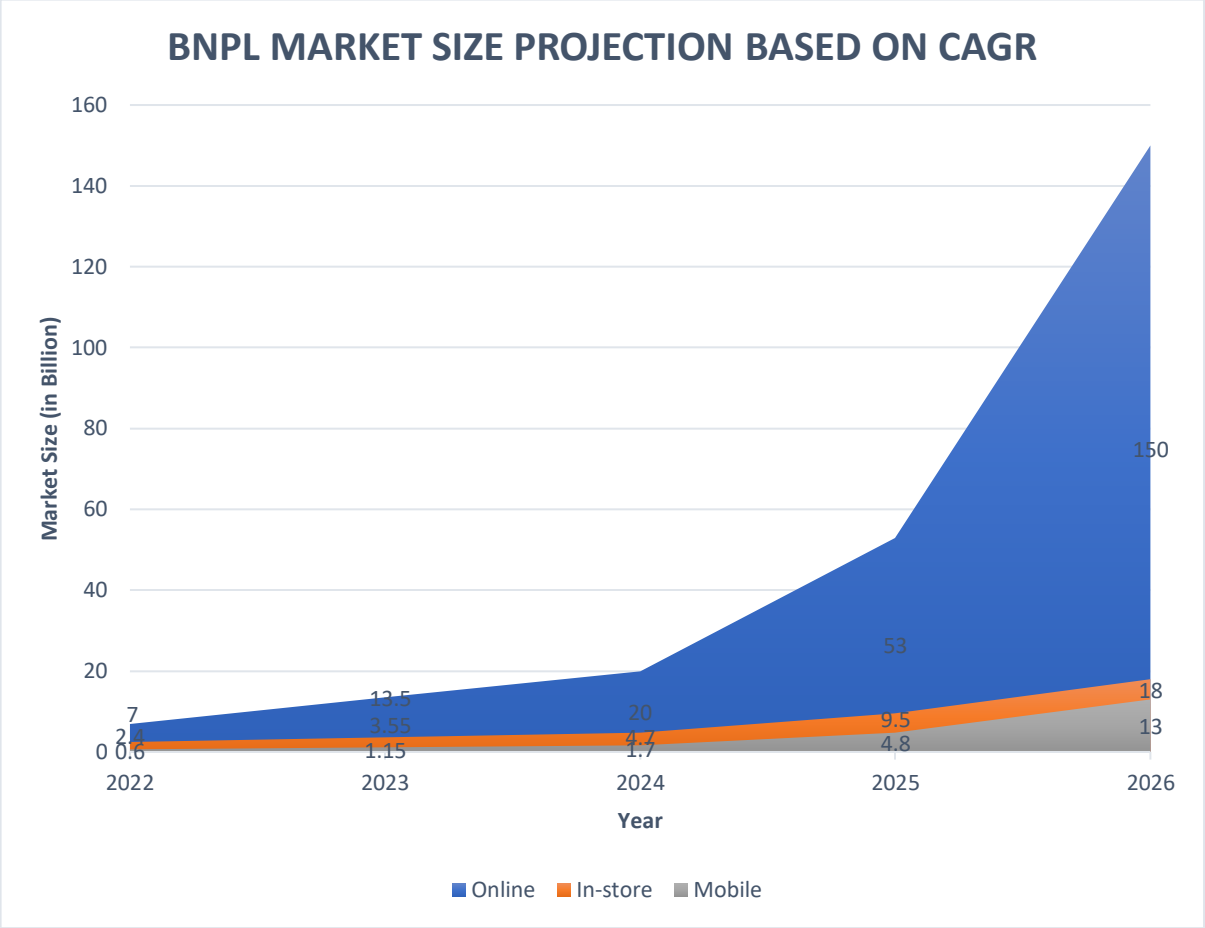


Figure 6: Potential growth of BNPL in Online, In-store, and Mobile

Figure 6 assumes a CAGR of 20% for Online, a CAGR of 15% for in-store, and a CAGR of 25% for mobile. The expected market size is to grow from $7 billion in 2022 to $150 billion in 2026 for online, $2.4 billion in 2022 to $18 billion in 2026 for In-store, $600 million in 2022 to $13 billion in 2026 for mobile. As it infers, the growth in the mobile space is even more significant than in the online or in-store spaces, representing a considerable opportunity for BNPL providers, merchants, and consumers.

Table 4: Phased growth of a BNPL

BNPL Growth	Online	In-store

Year 1	BNPL added to e-commerce checkout	
Year 2	Expands to more retailers, payment processors, more customer base	Offered in-store
Year 3	Gains popularity with a younger demographic and increases the adoption rate.	Expansion in in-store and gains traction thru' in app
Year 4	Consolidates market share with top players acquiring customers	Adoption reaches critical mass and becomes a mainstream option

Table 4 shows the evolution of the BNPL industry over four years, highlighting key milestones each year. In the first year, BNPL options add to e-commerce checkout processes, and in the second year, BNPL expands to more retailers and payment processors, attracting more customers. In the third year, BNPL gains popularity with younger demographics and increases adoption rates, primarily through mobile apps. By the fourth year, BNPL has consolidated its market share with top players acquiring competitors, and the payment option has become mainstream.

However, the rapid growth of the BNPL industry has also raised concerns about its impact on consumer debt and financial stability. Regulators in some countries, such as the UK and Australia, have already introduced measures to ensure BNPL providers operate responsibly and protect consumers from debt risks. More regulations will likely be in as the industry continues to evolve.

2.2.6. Key studies on Buy-now-pay-later

2.2.6.1. Consumer Behavior in Buy-Now-Pay-Later

Table 5: BNPL Consumer behavior

BNPL Consumer Behavior	Description
Browsing products	Consumers browse online stores for products they are interested in purchasing.
Adding to cart	Once they find something they want to buy, consumers add it to their online shopping cart.
Selecting BNPL option	At checkout, consumers choose the BNPL option, which allows them to pay for their purchases in installments over time.
Completing purchase	Consumers complete their purchase using the BNPL option, often with a few clicks or taps.
Repaying installments	Consumers must repay the BNPL provider per the agreed-upon repayment terms and schedule.
Monitoring account	Consumers may check their accounts regularly to monitor repayment status, upcoming payments, and available credit.
Using multiple providers	Some consumers may use multiple BNPL providers to spread their purchases and repayments across different services.
Managing budget	Consumers may use BNPL services as part of their overall budget management strategy, carefully keeping track of their expenses and repayments.

The table highlights some key behaviors consumers exhibit when using BNPL services, from browsing products to managing their budgets. It illustrates how BNPL has become a popular payment option for many consumers, offering flexibility and convenience when purchasing. However, consumers should be aware of the potential risks associated with BNPL services, such as high-interest rates and fees, and carefully manage their finances to avoid overextending themselves.

Influence of Consumer behavior in the buy-now-pay-later (BNPL) industry is by several factors, including:

1. *Convenience*: Consumers are attracted to BNPL because it offers a convenient and flexible payment option, allowing them to spread out the cost of their purchases over time.
2. *Affordability*: Many BNPL providers offer interest-free or low-interest payment plans, making it more affordable for consumers to make purchases they may not be able to afford upfront.
3. *Online shopping*: BNPL has become increasingly popular in e-commerce, where consumers seek convenient and secure online payment options.
4. *Budgeting*: BNPL can also help consumers to budget and manage their finances more effectively by allowing them to plan their payments well ahead of time.
5. *Trust*: Consumer trust is critical in the BNPL industry, and providers that offer transparent pricing and clear terms and conditions are more likely to attract and retain customers.

However, there are also concerns about the impact of BNPL on consumer debt and financial stability. Consumers who rely too heavily on BNPL may find themselves in debt and struggling to make payments, which can have long-term consequences for their financial health.

To mitigate these risks, many BNPL providers have implemented responsible lending practices, such as credit checks and affordability assessments, to ensure that consumers only borrow what they can afford to repay. Providers are also partnering with financial education providers to help consumers understand their financial obligations and make informed decisions about their borrowing.

The desire for convenience, affordability, and flexibility drives Consumer behavior in the BNPL industry, but concerns about debt and financial stability also influence it. As the industry continues to grow, it will be necessary for providers to maintain responsible lending practices and work closely with regulators to ensure that they are meeting their obligations to consumers.

Post-COVID-19, consumers' online shopping has increased, especially with e-commerce platforms. Several leading players adopted BNPL business strategies more effectively to serve the consumers. (Market Research Report, 2022).

BNPL has enabled expanded avenues of customers in addition to millennials and Gen Z consumers, extending to baby boomers and affluent consumers. BNPL also targets financially underserved consumers due to bad credit or no credit. (Alcazar and Bradford, 2021)

Unlike credit cards, BNPL does not require considering consumers' repayment capability. In addition, BNPL encourages impulsive buying due to the available credit. BNPLs have risks associated with identity fraud when there are unreported BNPL loans to credit

bureaus. With the opportunity to serve a wide range of consumers, credit risk increases with BNPL products. (Alcazar and Bradford, 2021)

Due to the risk associated with unregulated financial innovation, real-time monitoring of transaction data to monitor markets and evaluate risks is inevitable. (Guttman-Kenney et al., 2022)

2.2.6.2. Implications of Buy-Now-Pay-Later on Debt Management

Buy-Now-Pay-Later (BNPL) services allow consumers to make purchases without paying for them upfront instead of in installments over time. While this service can benefit consumers who need to purchase but do not have the funds available, it can also affect debt management.

Table 6: BNPL implications on debt management

Strategy	Description
Budgeting	Create a budget that includes repayment of BNPL debts. Ensure that enough income to meet the expenses and BNPL payments is available.
Prioritizing payments	Prioritize BNPL debt repayment over other expenses to ensure that payments are received and incur additional fees.
Negotiating terms	Contact the BNPL provider to negotiate repayment terms if there are struggles to make payments.
Avoiding new purchases	Repay the existing debts to make any new purchases with BNPL services.

Tracking payments	Keep track of BNPL payments, payment dates, and the total amount owed to stay on top of the debt.
Seeking help	Seek help from a financial advisor or credit counseling service if there are struggles with BNPL debt or managing finances.

Table 6 highlights some common strategies for managing BNPL debt, which can help consumers avoid falling into a debt trap. It is essential to carefully manage BNPL debts to avoid incurring high-interest rates and fees that can quickly add up. Consumers should also be aware of the potential impact on their credit score and financial health and take steps to manage their debt responsibly.

Here are some potential implications of BNPL on debt management:

1. Increased debt: Since BNPL services allow consumers to make purchases without paying for them upfront, they can lead to increased debt if consumers need to be more careful with their spending. Consumers may make purchases they would not usually make if they had to pay for them upfront, which could lead to more debt.
2. Interest charges: BNPL services often come with interest charges, which can add up quickly if payments are made late. Consumers who fail to make timely payments may also be subject to late fees, which can further increase their debt.
3. Missed payments: Since BNPL payments are in installments over time, consumers can easily forget when payments are due. Missed payments can lead to late fees, interest charges and damage a consumer's credit score.
4. Limited credit options: Using BNPL services can limit a consumer's ability to access other forms of credit, such as credit cards or loans. Lenders may view BNPL services as a form of debt, making it harder for consumers to qualify for

other forms of credit in the future.

BNPL services can have both positive and negative implications on debt management. Consumers who use BNPL services should be mindful of spending and ensure they plan to pay off their debt promptly to avoid interest charges and late fees. It is also essential to keep track of payments and due dates to ensure timely payments, which can further damage credit scores and increase debt.

2.3. Chapter Summary

2.3.1. Summary of Literature Review on Data Lake and cloud computing

Data lake

According to Inmon (2010), a data lake is a centralized repository that allows for storing structured and unstructured data at any scale. The data can be stored in its original format, allowing flexibility and agility in processing and analyzing it.

In their data lake architecture study, Lakshmanan et al. (2015) state that data lakes' are built on a distributed file system, which allows for storing data in multiple formats such as CSV, JSON, Parquet, and Avro. The architecture of data lakes also allows various tools and technologies to process and analyze the data.

Ghiassi and Lee (2018) highlight the primary advantages of using a data lake, including cost-effectiveness, scalability, flexibility, and agility. With a data lake, organizations can store large amounts of data without worrying about data silos or data schema changes, enabling faster and more efficient data analysis.

According to a study by Zhu et al. (2018), some of the challenges associated with data lakes include data quality, governance, and security. Without proper governance and security measures, data lakes can become a "data swamp" where data becomes inaccessible or unusable due to poor data quality or data security concerns.

In their study, Mishra et al. (2018) provides examples of use cases for data lakes, including customer analytics, fraud detection, and predictive maintenance. Data lakes are increasing in various industries, including healthcare, finance, and retail.

Overall, the literature suggests that data lakes are an effective solution for storing and processing large amounts of data. However, implementing proper governance, security, and data quality measures ensures that the data in the data lake remains usable and accessible.

Cloud computing

According to Armbrust et al. (2010), cloud computing uses remote servers and the Internet to store, manage, and process data, applications, and services.

In their study, Zhang et al. (2010) describe the architecture of cloud computing, which includes front-end devices, back-end servers, and the cloud itself. Cloud services classify into three services, namely, infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS).

In their research, Mell and Grance (2011) highlight the primary advantages of cloud computing, including cost-effectiveness, scalability, flexibility, and agility. With cloud computing, organizations can reduce their IT infrastructure costs while also having the ability to scale up or down as needed.

According to a study by Subashini and Kavitha (2011), some of the challenges associated with cloud computing include data security, privacy, and compliance. There are also concerns about vendor lock-in and the potential for service disruptions.

Marston et al. (2011) provide examples of use cases for cloud computing, including data storage, data processing, and application development. Cloud computing is increasing in various industries, including healthcare, finance, and education.

Overall, the literature suggests that cloud computing is an effective solution for organizations to reduce IT infrastructure costs while also having the ability to scale up or down as needed. However, implementing proper security and compliance measures protects data and applications.

In conclusion, **Cloud computing** will be a trendsetter; hence, appreciating the cloud's characteristics is crucial. These characteristics will be the foundation for data innovation on products and services. Also, they enable the Fintech businesses, from small to Enterprise, to set up their data platform. They are decisive in enabling the operational and analytics Fintech use cases.

With the advent of cloud computing, a multi-cloud paradigm is crucial, hence the associated data security. There is also more scope to explore Cloud-native architecture, edge computing, distributed cloud, hybrid cloud computing, cloud migration, and implementation challenges. The cloud journey is a massive paradigm shift for established on-prem systems, which needs a proper cloud migration strategy.

2.3.2. Summary of Literature Review on Fintech, Payments & Lending and Buy-Now-Pay-Later

Fintech

Fintech refers using of technology to provide financial services, according to (Lin et al., 2022). Fintech includes various verticals - payments, lending, insurance, wealth management, and personal finance.

Payments: In their study, (El Haddad et al., 2018) examine the payments vertical of Fintech, which includes mobile payments, digital wallets, and peer-to-peer payments. The study highlights the impact of Fintech on traditional payment systems and the benefits of increased financial inclusion.

Lending: In their research, (Gai et al., 2018) examine the lending vertical of Fintech, which includes peer-to-peer lending, crowdfunding, and online lending platforms. The study discusses the benefits of increased access to credit for individuals and small businesses.

Insurance: According to a study (PwC, 2019), the insurance vertical of Fintech includes insurtech, which uses technology to improve the efficiency of insurance processes, and digital insurance, which allows customers to purchase insurance online. The study highlights the potential for insurtech to reduce costs and improve customer experiences.

Wealth Management: In their study, (Anshari et al., 2022) examine the wealth management vertical of Fintech, which includes robo-advisors, automated investment platforms, and digital wealth management services. The study discusses the benefits of increased access to wealth management services and the potential for robo-advisors to improve investment outcomes.

Personal Finance: According to a study (McKinsey, 2022), the personal finance vertical of Fintech includes digital banking, personal financial management, and financial education. The study highlights the potential for Fintech to improve financial literacy and increase financial inclusion.

The Fintech literature suggests that Fintech is transforming the financial services industry by increasing access to financial services, improving efficiency, and reducing costs. The various verticals of Fintech, including payments, lending, insurance, wealth management,

and personal finance, offer numerous benefits for individuals and businesses. However, addressing privacy, security, and regulatory compliance concerns is essential.

Payments & Lending

(Akhtar et al., 2020) examine the impact of digital payments on financial inclusion in developing countries. The study highlights the potential for digital payments to increase access to financial services, reduce transaction costs, and improve the efficiency of payment systems.

(Berger et al., 2019) examine the impact of Fintech on small business lending. The study discusses the benefits of increased access to credit for small businesses but also highlights the potential for increased risks and regulatory challenges.

(Hua and Huang, 2021) examine the interplay between digital payments and lending. The study discusses the benefits of digital payments in reducing transaction costs and improving financial inclusion, as well as the potential for digital lending to increase access to credit for underserved populations.

(Delabarre, Maxime, 2021) examine the role of Fintech in payments and lending in emerging markets. The study discusses the potential for Fintech to improve financial inclusion in these markets but also highlights the challenges of regulatory compliance and consumer protection.

The payments & lending literature suggests that digital payments and lending are transforming the financial services industry by increasing access to financial services, reducing transaction costs, and improving efficiency. However, addressing the concerns about regulatory challenges, risks, and consumer protection ensures these technologies' responsible and sustainable use.

BNPL

According to a study by (Gerrans et al., 2022), BNPL (Buy-Now-Pay-Later) is a payment method that allows consumers to make purchases and pay for them in installments over time.

In their research, (Wang, 2022) examine the factors contributing to the popularity of BNPL among consumers. The study highlights the convenience and flexibility of BNPL, as well as the appeal of interest-free payments.

According to a study by (Alcazar and Bradford, 2021) , BNPL can influence consumer behavior by encouraging impulse buying and increasing spending. The study discusses the importance of educating consumers about the risks and responsibilities associated with BNPL.

(Aalders, 2023) examine the regulatory challenges associated with BNPL. The study discusses the need for consumer protection and transparency in BNPL practices and the potential for regulatory fragmentation across different jurisdictions.

(Alcazar and Bradford, 2021) examine the financial implications of BNPL for consumers and merchants. The study (Khan and Vilary Mbanyi, 2022) discusses the potential benefits of increased sales and customer loyalty for merchants and the potential risks of increased debt and financial instability for consumers, especially millenials.

The BNPL literature suggests that BNPL is a popular payment method that offers convenience and flexibility for consumers. However, there are also concerns about the potential for BNPL to encourage impulse buying and increase consumer debt. Proper regulation and consumer education are necessary to ensure the responsible and sustainable use of BNPL.

Fintech, BNPL & Data lake - Gaps

For any Fintech, from startup to Enterprise, to evolve and adapt to data lake should go thru' the evolution from data puddles, transform to data pond, adapt to the data lake, and revolutionize with data ocean to be future-proof.

There are various orchestrations of the data lake and multiple layers. Based on the business functionality and whether they are small, medium, or Enterprise, the data lake ponds, puddle, data lake, or data-ocean need to be federated. There are well-defined architectures and frameworks available. However, which fits the best in which state of business affairs needs in-depth analysis. For example, various ponds are applicable to avoid the garbage dump, but how Fintech should align the data lake based on the nature of business still needs more clarity.

Furthermore, there are various factors to consider making a practical decision, as cleaning up or re-orchestrating will cost time and resources. Correspondingly, on the multiple layers, some layers, like data ingestion or data storage, must be established with high efficiency and quick time-to-market. In contrast, some layers, like data quality or data governance, require more effort to set up the long-term solution to make it infallible.

Literature reviews demonstrate that the studies are on individual areas of Fintech, data lake, or cloud computing to a great extent. However, there are a few to stitch Fintech with cloud computing or data lake with the cloud. There are minimal touchpoints to connect Fintech data platform requirements for the cloud data lake. Studies must cover the business drivers in Fintech verticals and cloud data strategy. Furthermore, there is a need for the cloud data strategy cookbook for Fintech to establish the cloud data lake.

Engineering practices on the data lake layers are studied. However, it does not get to the cloud data lake, how to avoid building a monolithic data lake, and how to enable domain-

driven design with data mesh. The various innovations with Data mesh, data fabric, and others are the key to unlocking the future of data.

Data security and governance are always a concern with a data lake, which needs a strategic approach with Fintech-BNPL. Moreover, the maturity model of the data lake has yet to be defined and assessed, which requires fast-changing technology stacks for a data lake.

Finally, there should be a process for measuring the maturity model of the cloud data lake built regularly. It is required to evaluate how accustomed the architecture and data orchestration for users, operations, innovations, technology adaptations, and data & Fintech-BNPL trends are.

We need the data lake drivers for defining innovation strategy to bring the Fintech – BNPL disruption.

CHAPTER III:

METHODOLOGY

Chapter 3 provides a comprehensive overview of the research design and methodology applied to address the research problem and the research objectives outlined in Chapter 1. This chapter starts with summarizing the research problem, purpose, and questions. It then delves into the research design. The chapter then highlightsthe limitations of the research design and concludes with a summary of the chapter.

3.1. Overview of the Research Problem

The research study is to bridge the gap between the FinTech BNPL and Data Lake by building a strategy to build a successful Fintech data lake.

Managing data lakes in any organization, especially the Fintech BNPL, is cost intense. Traditional approaches considering only technology factors like on-prem vs. cloud solution, real-time vs. batch, etc., leads to wastage of resources. The proposed framework will lead to cost-effective Data lake management, allowing businesses to leverage 'Data as an Asset' (DaaA) more productively with a Data Maturity Model (DMM) as a Data Divinity factor.

The framework can eventually be used for building any Fintech data lake use cases with minor modifications to the subject data requirements.

3.1.1. Buy-Now-Pay-Later & Data lake

With the growing market share and digital footprint with BNPL, the BNPL data lake is crucial for any Fintech BNPL startup. Fintech BNPL should build the data lake quickly, with a fast Time-to-market, to enable critical data-driven business decisions, unlike in the past when building a data warehouse took a long time.

Buy-Now-Pay-Later (BNPL) services collect essential customer data, including transaction history, payment information, and personal details. This data can improve customer experience, prevent fraud, and improve business decisions. It is where a data lake comes into play.

A data lake is a centralized repository that allows organizations to store and analyze large volumes of structured and unstructured data. The design accommodates diverse data types, such as customer data collected by BNPL services. Here are some ways that a data lake can be beneficial for BNPL services:

1. *Improved customer experience:* By analyzing customer data stored in a data lake, BNPL services can gain insights into customer behavior, preferences, and needs. This information can be used to personalize customer experiences, offer targeted promotions, and improve customer satisfaction.

2. *Fraud prevention*: Data lakes can help BNPL services detect and prevent fraud by identifying suspicious transactions and patterns. Using machine learning algorithms, data lakes can analyze large volumes of data in real-time to identify fraudulent activities and take proactive measures to prevent them.

3. *Business intelligence:* By analyzing customer data stored in a data lake, BNPL services can gain insights into market trends, customer behavior, and competition. This information makes better business decisions, develop new products and services, and improve overall business performance.

4. *Regulatory compliance*: BNPL services are subject to strict regulatory requirements, including data protection and privacy regulations. A data lake can help ensure regulatory compliance by providing a secure and scalable platform for storing and managing customer data.

The research study is to connect the dots between the Business drivers and Technology drivers of data lake as shown in Figure 7.

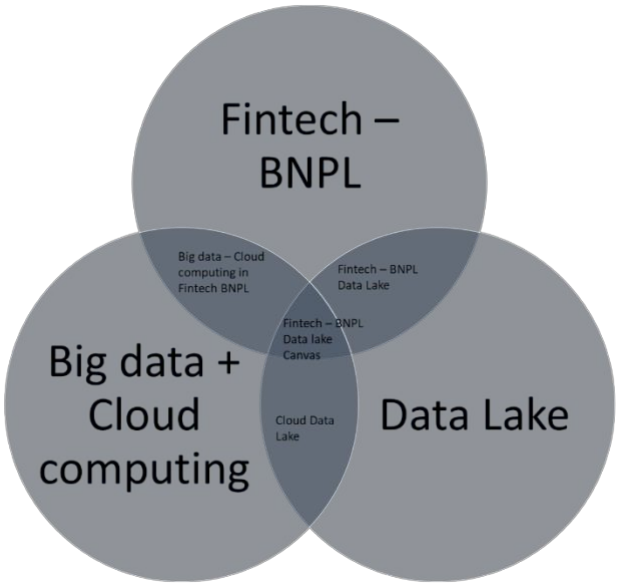


Figure 7: Fintech data lake – Connect the Dots

A literature review on Data Lake reveals several key insights and trends:

Definition and Characteristics: Data lake is a centralized repository for storing and analyzing large volumes of raw data in its native format. It is a scalable, cost-effective, and flexible solution that enables businesses to store and analyze large volumes of structured and unstructured data.

Benefits: Data Lake provides several benefits to businesses, including improved decision-making, personalized customer experience, scalability, better data governance, and cost-effectiveness. It also enables businesses to perform advanced analytics and gain insights into customer behavior, preferences, and market trends.

Implementation: The implementation of data lake involves several steps, including data

ingestion, data processing, and data analysis. Businesses must ensure proper data governance and security to protect customer data and comply with regulatory requirements.

Challenges: Data Lake implementation also comes with several challenges, including data quality, data integration, data security, and data governance. It is essential to address these challenges to ensure the successful implementation and adoption of data lake solutions.

Use Cases: Data lake has been successfully implemented in various industries, including finance, healthcare, retail, and telecommunications. Use cases range from fraud detection and prevention to personalized marketing and customer experience.

The literature review highlights the importance of data lake in current business operations. By providing a centralized repository for storing and analyzing large volumes of data, a data lake enables businesses to gain insights into customer behavior, preferences, and market trends. However, businesses must address several challenges to successfully implement and adopt data lake solutions.

Table 7: Summary of Objectives & Gaps in BNPL Data lake studies

Study	Objective	Gaps
(Parne, 2021)	Provides cloud computing strategy, impact in banking and financial institutions and discusses the significant reliance of cloud computing	Lack of addressing the high cost for micro enterprises to adopt it

(Oberoi et al., 2021)	Provides insights into how cloud computing can be used in the banking industry, the various business models associated with it, and the challenges the banking industry faces in adopting this technology.	Lack of standardization to address the challenges
(Imerman and Fabozzi, 2020)	Showcases FinTech Ecosystem and a conceptual framework for FinTech innovation.	Lack of addressing the business challenges with data technology
(Khan and Vilary Mbanyi, 2022)	Studies buy now, pay later (BNPL) and its influence on millennials buying behavior and consumption when mobile shopping.	Lack of connecting consumer behavior with data technology
(Guttman-Kenney et al., 2022)	Analysis of an example of how consumer financial protection regulators can use realtime transactions data to monitor markets and evaluate potential risks - especially (largely) unregulated, financial innovations such as BNPL	Lack of framework for addressing BNPL regulation and security issues connecting with data technologies
(Alshahri, 2022)	Offers "buy now,p pay later" (BNPL) payment methods for B2B and provide solutions (e.g. liquidity, automation, digital payments) for online and offline SMEs	Lack of data to capture the business drivers on the payment methods and solutions with cost optimization

(Aalders, 2023)	Uses mechanisms and conditions framework of affordances and walkthrough method to analyze how popular BNPL products define responsible lending and spending	Lack of brining the consumer behavior and regulations to handle with data lake
(Vinoth et al., 2022)	Examines several cloud computing applications in banking and e-commerce, as well as the security issues associated with them	Lack of data to protect business from security threats
(Hai et al., 2021b)	Provides a comprehensive overview of research questions for designing and building data lakes.	Lack of other architecture aspects on stream data lakes, integrate data lakes with machine learning and data science
(Sawadogo and Darmont, 2021b)	Provides a comprehensive state of the art of the different approaches to data lake design focusing on data lake architectures and metadata management, which are key issues in successful data lakes.	1. Data integration and transformation aspects having recurring issues. 2. Data governance principles are indeed currently seldom turned into actual solutions.
(Nargesian et al., 2019b)	Discusses how data lakes are introducing problems including dataset discovery and how they are changing the requirements for classic problems including data extraction, data cleaning, data integration, data versioning, and metadata	Lack of framework to handle the data issues

	management.	
(Giebler, Corinna et al., 2021b)	Introduces the data lake architecture framework.	Lack of its connection to the business drivers and challenges
(Kumar et al., 2021)	Discusses various available data storage options, suitability, and limitations with cloud.	Lack of the underlying need of the business driving the selection of database

As presented by the data in Table 7, despite the availability of information on the topic, there are gaps in the research performed regarding understanding the performance and usefulness of the Fintech - BNPL data lake.

3.1.2. Integration of Buy-Now-Pay-Later and Data Lake

The integration of Buy-Now-Pay-Later (BNPL) services and data lakes can provide several benefits to businesses. By storing and analyzing customer data in a data lake, businesses can gain insights into customer behavior and preferences, develop personalized marketing strategies, and prevent fraud. Here are some ways that BNPL and data lake integration can work together:

1. *Data collection*: BNPL services collect a significant amount of customer data, including transaction history, payment information, and personal details. This data fed into a data lake can be analyzed and used to develop insights and business strategies.
2. *Analytics*: Data lakes can provide businesses with advanced analytics capabilities, allowing them to analyze large volumes of data in real-time. It can help

businesses identify trends, develop personalized marketing strategies, and optimize business operations.

3. *Fraud prevention*: By integrating BNPL services with a data lake, businesses can monitor transactions and identify suspicious activities. Using machine learning algorithms, data lakes can analyze large volumes of data and detect fraudulent transactions, allowing businesses to take corrective actions and prevent future fraud.

4. *Personalization*: By analyzing customer data stored in a data lake, businesses can develop personalized marketing strategies and tailor their offerings to customer needs and preferences. It can help businesses attract new customers and increase customer loyalty.

5. *Regulatory compliance*: BNPL services are subject to strict regulatory requirements, including data protection and privacy regulations. Businesses can use a data lake to comply with these regulations and protect customer data.

Integrating BNPL services and data lakes can give businesses valuable insights into customer behavior and preferences, helping them develop personalized marketing strategies, prevent fraud, and optimize business operations. It can also help businesses comply with regulatory requirements and protect customer data, ensuring a secure and seamless customer experience.

3.1 Research Question

The research is to understand - How to build a successful data lake in Fintech BNPL, bridging the business and technology drivers? Hence the goal is to Build a BNPL – Fintech Data lake framework based on the industry's best practices and challenges by analyzing existing studies and collecting survey responses.

The three main Objectives to achieve the goal includes,

- ***Objective 1*** – Validate the canvas for Data leaders and Executive management to optimize the Operation Expense (OpEx) cost more productively by comparing the existing model and the proposed canvas.
- ***Objective 2*** – Validate the canvas for Data leaders and Executive management to estimate the current Data Maturity Model (DMM) that the organization owns by comparing the existing model and the proposed canvas.
- ***Objective 3*** - Validate the canvas by generating comparative statistics with an existing model to measure 'Data as an Asset.'

3.2. Research Design

The research design uses a mixed methodology approach, combining historical data analysis, simulation, and statistical analysis to arrive at meaningful conclusions and recommendations for the Fintech BNPL data lake. The study is to conduct in five stages, including a research survey based on the current data lake challenges, data analysis on the independent variables, building the Fintech-BNPL data lake framework, validation of framework, and interpretation of results.

3.2.1. Data Collection

The research survey and analysis focus on the challenges of BNPL & data lake, which are independent variables. The responses will be used for building the framework, which will be used for validating Hypothesis 1 and 2.

Below are the data collection techniques followed for various research factors for building the framework.

Table 8: Research methodology

S.No	Research Factors	Research Methodology	Primary data collection
1	Build a framework model to collect BNPL data with key business subject areas.	Exploratory & Qualitative research	Survey
2	Possible Data Lake adaptability for BNPL Fintech (Mapping SMB Vs. Large enterprises with Data puddle, data pond, data lake, data ocean, and cloud solution - Cloud service, storage, zone)	Exploratory & Qualitative research	Survey
4	Framework to avoid Garbage dump in BNPL - Fintech data lake	Exploratory & Qualitative research	Survey
5	Factors to consider for building sustaining BNPL Fintech data lake with long-term efficiency	Exploratory & Qualitative research	Survey
6	Factors to consider for building BNPL Fintech Data Lake with quick Time to Market	Exploratory & Qualitative research	Survey

7	Aspects for Setting up New Cloud data lake Vs. Cloud Data Lake migration	Exploratory & Qualitative research	Survey
8	Evaluate the framework for better cost management	Quantitative research	Survey and Interviews
9	Evaluate the framework for Data Maturity Model	Quantitative research	Survey and Interviews
10	Evaluate the framework for Data as an Asset (DaaA)	Quantitative research	Survey and Interviews

3.2.2. Fintech-BNPL data lake canvas

In order to achieve the first objective of the study: To build the Fintech-BNPL data lake canvas, current challenges of Fintech – BNPL and Data lake were identified. The identified factors are the independent variables, and their possible outcomes are the dependent variables. These data have been fed into the survey questionnaire and circulated to get the responses.

Survey responses are from, Fintech leaders, Engineering leaders, Product leaders, Product owners, Technology leaders, CXOs, Data architects, Data Engineers, Data Scientists, Data leaders, and other data practitioners,

Insights are derived from the survey responses to derive the BNPL Data Lake canvas with various analytics using Python scripts.

1. Frequency analysis – Identification of the most popular response for the

identified question to determine the most common response and the overall distribution of the responses. It applies to the responses, either multiple choice or with the Likert scale. Also, to comprehend the pattern & trends in the data.

2. Cross-Tabulation –Analysis of the responses from various data practitioners with a data matrix showing the number of respondents who fall into each combination of responses. This analysis is to be at the role and industry levels to analyze the data in the matrix to identify patterns or relationships between the variables and present the results in a table or chart format, highlighting any significant findings.

3. Cluster analysis - Identify patterns and relationships in Fintech and data engineering survey data and classify data into meaningful groups or clusters based on the independent variables identified. Bar charts and comparison analysis charts are to be used for analysis.

3.2.3. BNPL data lake canvas validation

Survey responses and analysis are to build the BNPL data lake canvas. The perusal of the canvas is to perform three hypotheses testing by achieving Objectives 2, 3, and 4.

3.2.3.1. Cost optimization with BNPL data lake canvas

In order to achieve the second objective of the study: Validate the canvas for Data leaders and Executive management to optimize the Operation Expense (OpEx) cost in a more productive way by comparing the existing model and the proposed canvas.

Various strategies to perform the non-parametric tests, plot, and simulate using Python

scripts to compare the data for the existing and data lake canvas models. Simulate the responses on cost management based on the survey responses. Build two groups of data – existing and data lake canvas models- to perform Post-hoc analysis using the Kolmogrov-Smirnov test.

Table 9: Objective 1 - Hypothesis testing strategy

Strategy	Rules
Opex_cost $f(x1) = \uparrow D(a)$ where $5 \leq a \leq 10$	% Sampling to evaluate based % of the responses for Opex cost increase by 5-10%
Opex_cost $f(x2) = \uparrow D(a)$ where $10 \leq a \leq 20$	% Sampling to evaluate based % of the responses for Opex cost increase by 10-20%
Opex_cost $f(x3) = \uparrow D(a)$ where $20 \leq a \leq 40$	% Sampling to evaluate based % of the responses for Opex cost increase by 20-40%
Opex_cost $f(x4) = \downarrow D(a)$ where $5 \leq a \leq 10$	% Sampling to evaluate based % of the responses for Opex cost decrease by 5-10%
Opex_cost $f(x5) = \downarrow D(a)$ where $10 \leq a \leq 20$	% Sampling to evaluate based % of the responses for Opex cost decrease by 10-20%
Opex_cost $f(x6) = \downarrow D(a)$ where $20 \leq a \leq 40$	% Sampling to evaluate based % of the responses for Opex cost decrease by 20-40%

Opex_cost $f(x7) = D(a)$ where a is undefined	% Sampling to evaluate based % of the responses for Opex cost decrease by 20-40%

Hypothesis

H0 - Existing approach(es) leads to the same Operational Expense (OpEx) as compared to the proposed framework

Statically,

$H0 – \mu T\ f(x) = \mu F\ f(x)$

Where μT is the mean opex of the existing approach(es), and μF is the mean opex of the proposed framework being tested.

Alternative Hypothesis

$H1 - \mu T\ f(x) > \mu F\ f(x)$

3.2.3.2. 'Data Maturity Model' with BNPL data lake canvas

In order to achieve the third objective of the study: Validate the canvas for Data leaders and Executive management to estimate the current Data Maturity Model (DMM) that the organization owns by comparing the existing model and the proposed canvas.

Various strategies to perform the non-parametric tests, plot, and simulate using Python scripts to compare the data for the existing and data lake canvas models. Simulate the

responses on cost management based on the survey responses. Build two groups of data – existing and data lake canvas models- to perform Post-hoc analysis using the Kolmogorov-Smirnov test.

Table 10: Objective 2 - Hypothesis testing strategy

Strategy	Rules
f(DQI1) = ↑ D(q) where $0 \leq q \leq 20$	% Sampling to evaluate based % of the responses for data quality issues in the data lake between 0-20%
f(DQI2) = ↑ D(q) where $20 \leq q \leq 40$	% Sampling to evaluate based % of the responses for data quality issues in the data lake between 20-40%
f(DQI3) = ↑ D(q) where $q \geq 40$	% Sampling to evaluate based % of the responses for data quality issues in the data lake >40%
f(DQI4) = D(q) where q is undefined	% Sampling to evaluate based % of the responses for data quality issues in the data lake due to other reasons and not measured
f(DSI1) = ↑ D(s) where $0 \leq s \leq 1$	% Sampling to evaluate based % of the responses for data security issues in the data lake between 0-1%
f(DSI2) = ↑ D(s) where $2 \leq s \leq 5$	% Sampling to evaluate based % of the responses for data security issues in the data lake between 2-5%

f(DSI3) = ↑ D(s) where 5 ≤ s ≤ 10	% Sampling to evaluate based % of the responses for data security issues in the data lake between 5-10%
f(DSI4) = D(s) where s is undefined	% Sampling to evaluate based % of the responses for data security issues in the data lake due to other reasons and not measured
f(DGI1) = ↑ D(g) where 5 ≤ g ≤ 10	% Sampling to evaluate based % of the responses for garbage dump in the data lake increase by 5-10%
f(DGI2) = ↑ D(g) where 10 ≤ g ≤ 20	% Sampling to evaluate based % of the responses for garbage dump in the data lake increase by 10-20%
f(DGI3) = ↑ D(g) where 20 ≤ g ≤ 40	% Sampling to evaluate based % of the responses for garbage dump in data lake increase by 20-40%
f(DGI4) = ↓ D(g) where 5 ≤ g ≤ 10	% Sampling to evaluate based % of the responses for garbage dump in the data lake decrease by 5-10%
f(DGI5) = ↓ D(g) where 10 ≤ g ≤ 20	% Sampling to evaluate based % of the responses for garbage dump in the data lake decrease by 10-20%
f(DGI6) = ↓ D(g) where 20 ≤ g ≤ 40	% Sampling to evaluate based % of the responses for garbage dump in the data lake decrease by 20-40%

f(DGI7) = D(g) where g is undefined	% Sampling to evaluate based % of the responses for garbage in the data lake due to other reasons
$f(TTM_1 1)$ = ↑ D(t) where $0 \leq t \leq 1$	% Sampling to evaluate based % of the responses for the effort required to migrate data without modernization for startup < 1 month
$f(TTM_1 2)$ = ↑ D(t) where $0 \leq t \leq 2$	% Sampling to evaluate based % of the responses for the effort required to migrate data without modernization for SMB < 2 months
$f(TTM_1 3)$ = ↑ D(t) where $0 \leq t \leq 3$	% Sampling to evaluate based % of the responses for the effort required to migrate data without modernization for Enterprise < 3 months
$f(TTM_1 4)$ = ↑ D(t) where t is undefined	% Sampling to evaluate based % of the responses for the effort required to migrate data without modernization which can be other
$f(TTM_2 1)$ = ↑ D(t) where $0 \leq t \leq 2$	% Sampling to evaluate based % of the responses for the effort required to build data lake from scratch for startup < 2 months
$f(TTM_2 2)$ = ↑ D(t) where $t \geq 2$	% Sampling to evaluate based % of the responses for the effort required to build data lake from scratch for

	startup > 2 months
$f(TTM_2 3) = \uparrow D(t)$ where $3 \leq t \leq 4$	% Sampling to evaluate based % of the responses for effort required to build data lake from scratch for SMB < 3 - 4 months
$f(TTM_2 4) = \uparrow D(t)$ where $0 \leq t \leq 6$	% Sampling to evaluate based % of the responses for the effort required to build data lake from scratch for SMB is 6 months
$f(TTM_2 5) = \uparrow D(t)$ where $t \geq 6$	% Sampling to evaluate based % of the responses for the effort required to build data lake from scratch for Enterprise > 6 months
$f(TTM_2 6) = \uparrow D(t)$ where $t \geq 12$	% Sampling to evaluate based % of the responses for the effort required to build data lake from scratch for Enterprise > 12 months
$f(TTM_2 7) = \uparrow D(t)$ where t is undefined	% Sampling to evaluate based % of the responses for the effort required to build data lake from scratch, which can be other

Hypothesis 2

H0 - Existing approach(es) leads to measuring the 'Data Maturity Model' (DMM) as compared to the proposed framework

H0 – DMM[T] > DMM[F]

$DMM[T] = DQI[T] + DSI[T] + DGI[T] + TTM_1[T] + TTM_2[T]$

Where DQI is the Data Quality Index score, DSI is the Data Security Index score, DGI is the Data Governance Index Score, TTM_1I, TTM_2I is the Time To Market Index score from the traditional framework [T], and the proposed framework [F]

Alternative Hypothesis

H1 - DMM[T] < DMM[F]

Assumption: Scales and weights are equal

3.2.3.3. Data as an Asset (DaaA) with BNPL data lake canvas

In order to achieve the fourth objective of the study: Validate the canvas by generating comparative statistics with an existing model to measure 'Data as an Asset.'

Various strategies to perform the non-parametric tests, plot, and simulate using Python scripts to compare the data for the existing and data lake canvas models. Simulate the responses on cost management based on the survey responses. Build two groups of data – existing and data lake canvas models- to perform Post-hoc analysis using the Kolmogorov-Smirnov test.

Table 11: Objective 3 - Hypothesis testing strategy

Strategy	Rules

f(DaaS)	% Sampling to evaluate based on the number of data services derived from the data lake
f(DaaP)	% Sampling to evaluate based on the number of data products derived from the data lake

Hypothesis 3

H0 - Traditional approach(es) leads to measuring 'Data as an Asset' (DaaA) as compared to the proposed framework

H0 – DaaA[T] > DaaA[F]

DaaA[T] = DaaP[T] + DaaS[T]

Where DaaP is Data as a Product, and DaaS is Data as a Service from the traditional framework [T] and proposed framework [F]

Alternative Hypothesis

H1 - DaaA[T] < DaaA[F]

3.2.4. Interpretation of results

The results from the first objective are to derive the Fintech-BNPL data lake canvas with the critical factors based on strategic groups. The purpose of the BNPL data lake canvas is to help address the challenges of data lake effectively. It includes cost optimization, a

data maturity model to address data quality, data security, data governance, time to market, lifetime value challenges, and building Data as an Asset (DaaA) based on BNPL subject areas. It also helps in supporting the data lake with architecture considerations.

To prove the effectiveness of the BNPL data lake canvas is by first, second, and third objectives.

3.2.5. Research Design Limitations

It is important to note that the research design for this study has certain limitations to consider when interpreting the findings.

The data used for the study is limited to the survey responses on BNPL and data lake from the relevant industry data practitioners and simulation based on the response groups. Though it is across various data practitioners and industry experts, the results of this study may need to be more consistent when tested on other Fintech areas or other groups.

The research design attempts to eliminate the sample and look-forward bias by training the model and using the data based on the survey response.
However, the results may vary when applied in certain situations due to issues such as
The research design attempts to understand how the small, medium, and enterprise Fintech build the data lake and the architecture they use, the executive leaders' data-driven business decisions, etc.

The research design uses a survey for Fintech industry experts and data engineering

experts to ensure consistency between Fintech - BNPL and data lake. The research design can be leveraged for any Fintech data lake; however, it is important to consider that the subject areas for each Fintech may vary.

3.2.6. Chapter Summary

The chapter provided an overview of the research problem with Fintech - BNPL data lake, where there is no current framework for building an effective BNPL data lake. This study aims to build a Fintech data lake canvas and evaluate its effectiveness by addressing the associated challenges on cost optimization, data maturity model, and Data as an Asset.

The chapter reiterated the study's objectives: analyzing the existing survey data on cost-effectiveness, data quality, data security, data governance, time to value, time to market, data as a product, and data as a service.

The chapter introduces the research design of this study, which consists of four objectives. The survey data uses Python scripts with data exploration techniques of frequency analysis, and cross-tabulation.

The first, second, and third objective is to evaluate data lake canvas effectiveness with hypothesis testing using Python scripts. Achieving these three objectives is by simulating the data based on the survey data responses by a strategic grouping of parameters with each objective.

The chapter also discussed the research design limitations. The data is limited to the Fintech BNPL data lake and may not be consistent with other Fintech subject areas. The study eliminates sample and look-forward bias by strategic sampling, but the results may vary due to different Fintech subjects. The data lake primary data used in

the research is consistent with other studies, but different Fintech verticals may lead to different results.

CHAPTER IV:

RESULTS

Chapter 4 provides the significant findings of this research. It starts by accessing the results from the survey responses on the Fintech / Buy-Now-Pay Later factors for the data lake and the Data engineering aspects for the data lake. Both the survey responses are compared to derive the factors that impact the successful strategy for the Fintech - Buy Now Pay Later data lake canvas. Finally, the chapter will provide a descriptive analysis of the results to provide the Fintech – Buy Now Pay Later data lake canvas based on the results described in the previous chapters.

4.1. Research case

This section will take a closer look at the parameters impacting the Fintech data lake and evaluate the stability of the canvas based on the responses. These analyses provide valuable insights into the factors considered for the Fintech data lake canvas.

Factors that affect the Fintech – data lake falls into two streams of study impacting the data lake aspects 1) Fintech business aspects and 2) data engineering aspects. Hence the survey was conducted with Fintech industry experts and data technology experts. The survey responses will be analyzed and combined to define the data lake canvas.

4.1.1. Research study

The research identifies the challenges of data lake from the business and engineering perspectives. Identified challenges are the independent variables influencing the data lake success strategy. The design of the questionnaire for each category depends on the independent variables.

4.1.2. Research Participants

Survey participants vary for both surveys, with few overlapping roles in Fintech companies, including the BFSI sector. Table 12 highlights the targeted experts for the two surveys on Fintech and Data engineering.

Table 12: Survey participants – Industry and Role

Survey	**Role**
Fintech Survey	Data Leader
	CTO / CXO
	Engineering Leader
	Product Leader
	IT / Technology Leader
Data Engineering Survey	Data Engineer
	Data Analyst
	Data Architect
	Engineering Manager
	Data scientist
	Product Manager
	Data Leader

Survey participants include Fintech and data-leading technology companies where the organization can be any of the below.

1. Enterprise
2. SMB

3. Startup

4.2. Factors Impacting the successful strategy for Fintech data lake

As described in Chapter 2, the challenges of the Fintech data lake are thoroughly analyzed to identify the influencing factors, which are as below.

1. Fintech business challenges
2. Buy-Now-Pay-Later business challenges

4.2.1. Descriptive Analysis of the Fintech – Buy Now Pay Later - Business

4.2.1.1. Fintech and BNPL Business

To understand the data lake aspects of Fintech – BNPL, it is essential to comprehend Fintech challenges, BNPL challenges, business aspects associated with the critical function areas, and business data required for making data-driven decisions. Hence survey has the questions in Table 13.

Table 13: Survey questionnaire – Fintech and BNPL Industry view on Data lake

Q.No	Factor Type	Question	Expected Responses (Can be more than one)
1	Fintech Challenge	What are the critical Fintech Business Challenges?	1. Data Security 2. Competition 3. Compliance and Regulation 4. Customer Retention and Customer

			Experience 5. Service Personalization 6. Data and AI integration 7. Blockchain Integration 8. Others
2	BNPL challenge	What are the critical **BNPL** challenges?	1. Risk Management 2. Fraud Protection 3. Current economic/geopolitical / Inflation factors 4. Customer Acquisition 5. Customer Retention 6. Other
3	BNPL challenge	What are the focus areas for Risk Management in BNPL?	1. Fraud Protection 2. Risk Scoring 3. Interest Rates 4. Missed payments 5. Increased debt for FinTech 6. Risk of debt spirals across multiple BNPL providers

			7. Other
4	BNPL challenge	What are Key factors in the BNPL customer retention journey?	1. Customer loyalty 2. Customer satisfaction 3. Customer referrals 4. Merchant brand exposure 5. Improved efficiency in Managing returns 6. Other
5	BNPL challenge	Key Business subject data of BNPL are	1. Payments & Schedule 2. Customer 3. Scoring 4. Risk Levels & category 5. Product Catalog of BNPL 6. Financial Calculation 7. Limits 8. Master data & Metadata 9. Other

4.2.1.2. Fintech and BNPL Business – Survey Results

In the first phase of the data analysis, survey responses were reviewed carefully, and then the frequency checks and the cross-tabulation with role and organization. Key insights from the survey responses are as below.

4.2.1.2.1. Fintech challenges

1. The data presented in Figure 8 shows that Data security and Compliance & regulations are the critical Fintech challenges for the enterprise organization BNPL data lake.

2. The data presented in Figure 9 and Figure 10 shows that Engineering leaders, data leaders, product leaders, and Fintech leaders consider data security and Compliance & security as critical challenges.

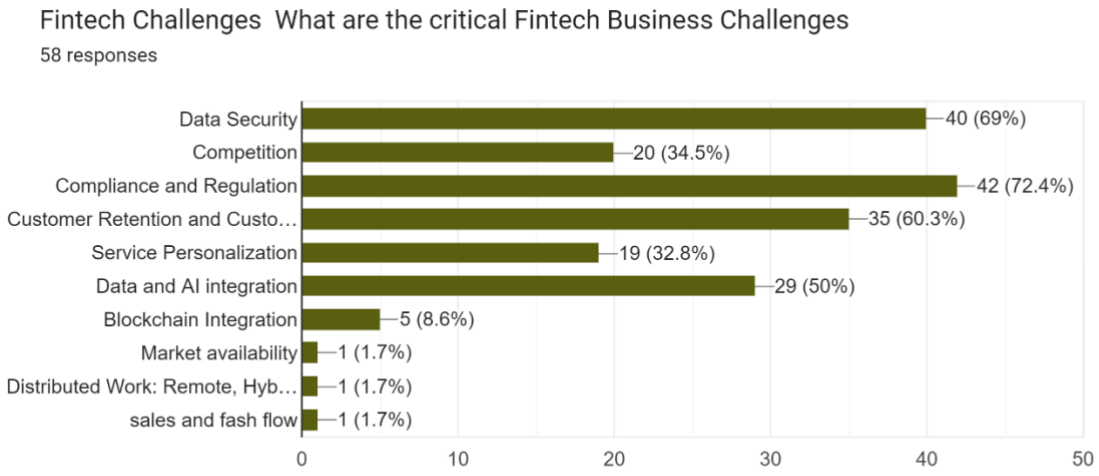


Figure 8: Fintech business challenges

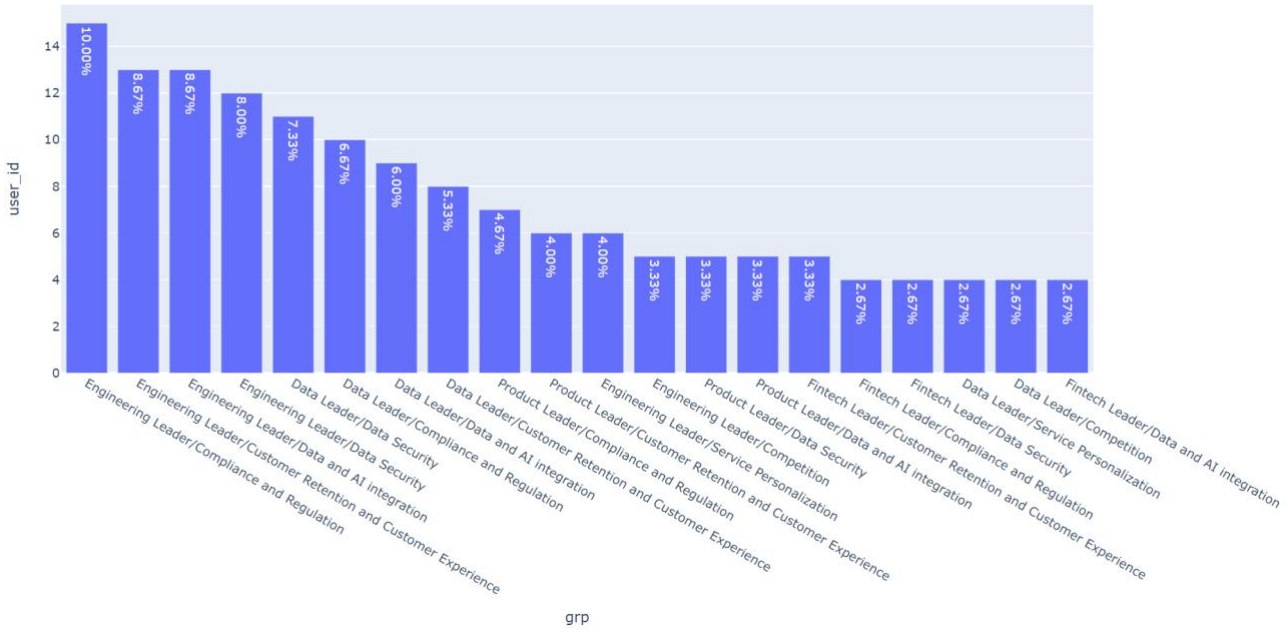


Figure 9: Fintech challenges grouped by role

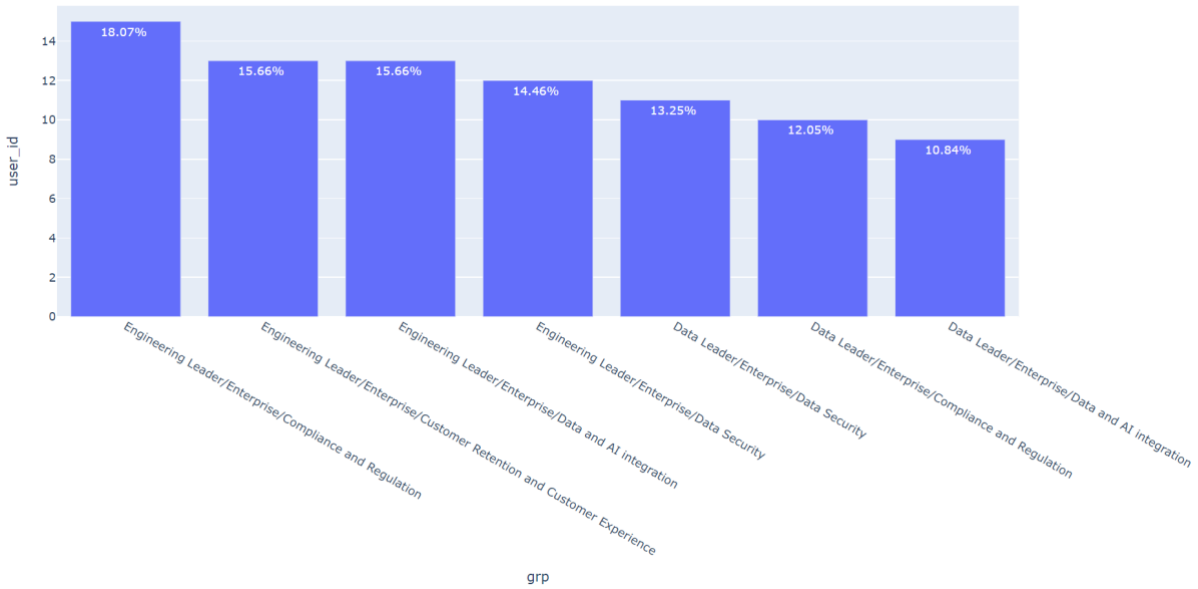


Figure 10: Fintech challenges grouped by role and organization type

4.2.1.2.2. BNPL challenges

1. The Figure 11 and Figure 12 shows that Risk management and fraud protection are the key BNPL challenge for the enterprise.

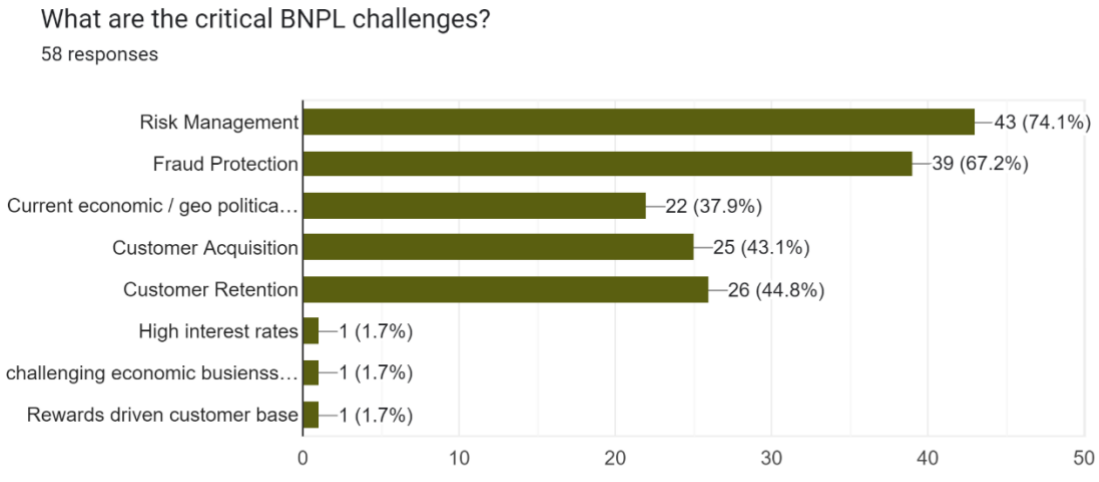


Figure 11: BNPL challenges

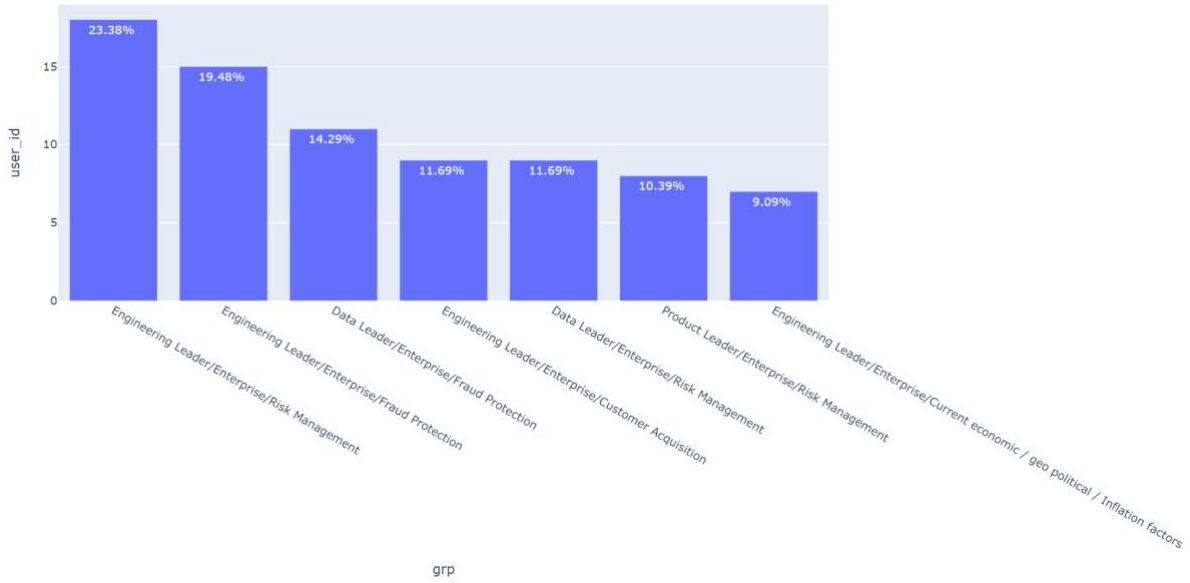


Figure 12: BNPL challenges grouped by role and organization type

4.2.1.2.3. BNPL Risk Management

1. The Figure 13 and Figure 14 shows that Fraud protection and risk scoring are the critical factors for BNPL risk management for the enterprise.

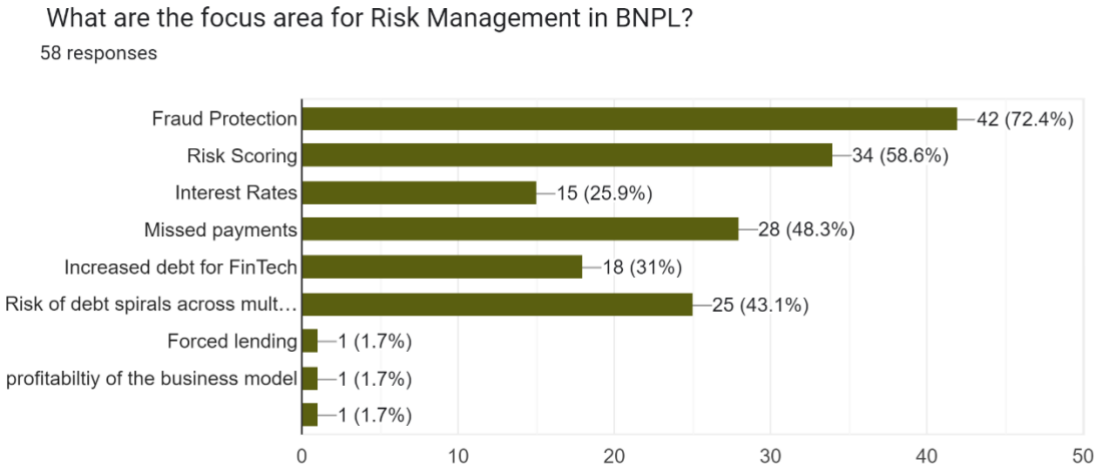


Figure 13: Risk management factors grouped by Role and Organization type

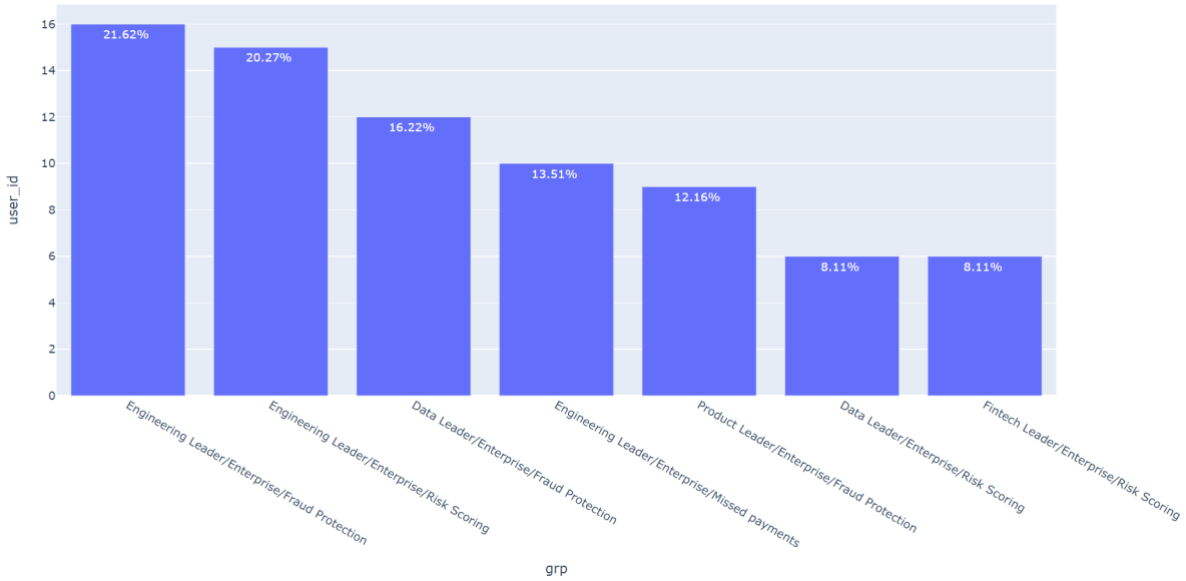


Figure 14: Risk management factors grouped by Role and Organization type

4.2.1.2.4. BNPL Customer Journey

1. The Figure 15 and Figure 16 shows that Customer satisfaction and improving efficiency in managing returns are the critical factors for the BNPL customer journey for the enterprise.

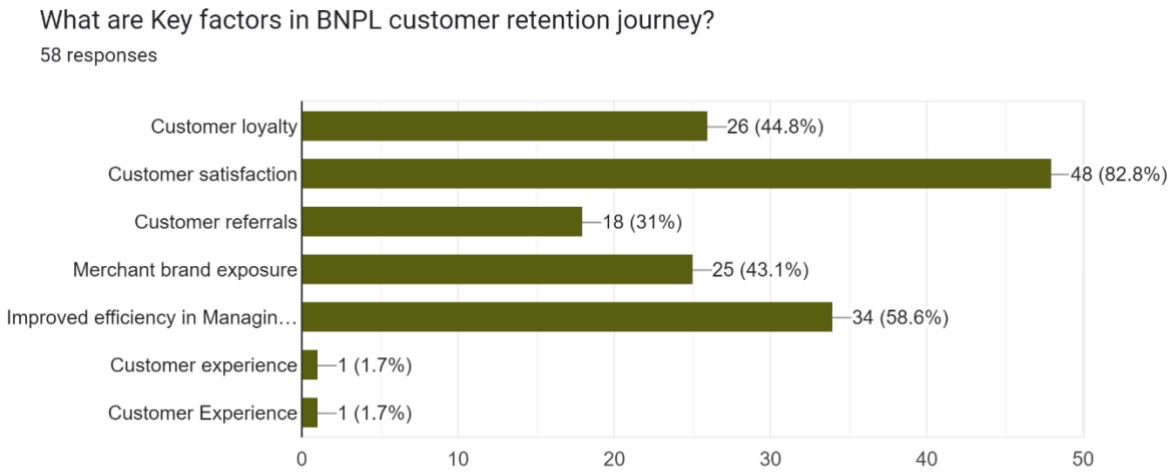


Figure 15: Customer journey factors

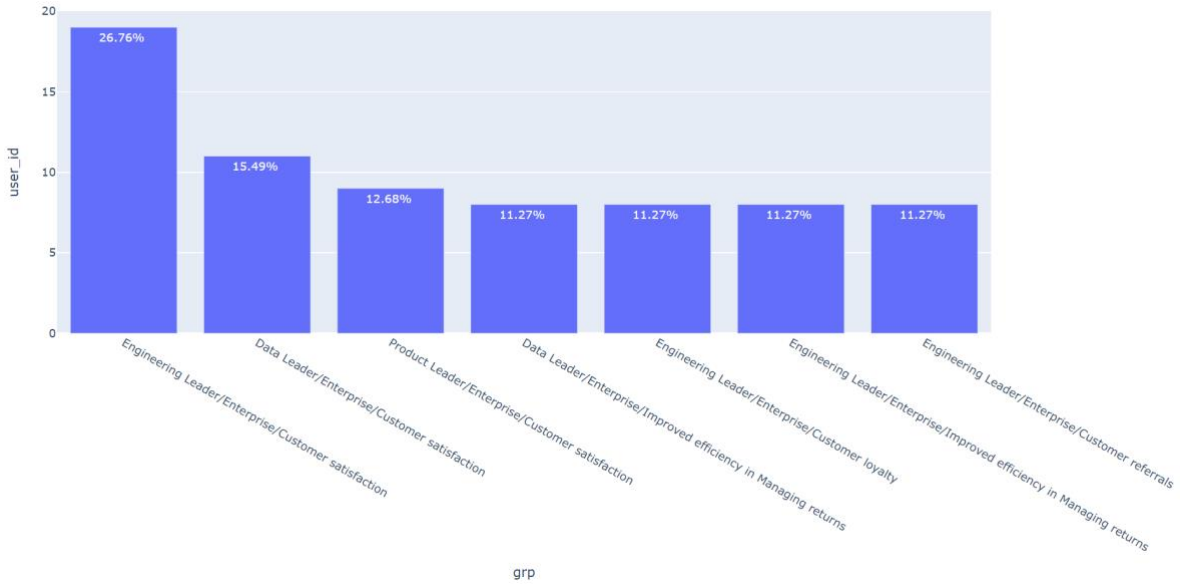


Figure 16: Customer journey factors grouped by Role and Organization type

4.2.1.2.5. BNPL Subject areas

1. The Figure 17 and Figure 18 shows that Payments & schedule and, Risk levels & category are the critical factors for BNPL subject areas for the enterprise.

2. It also shows that the customer and scoring subject areas are essential BNPL subject areas for the enterprise.

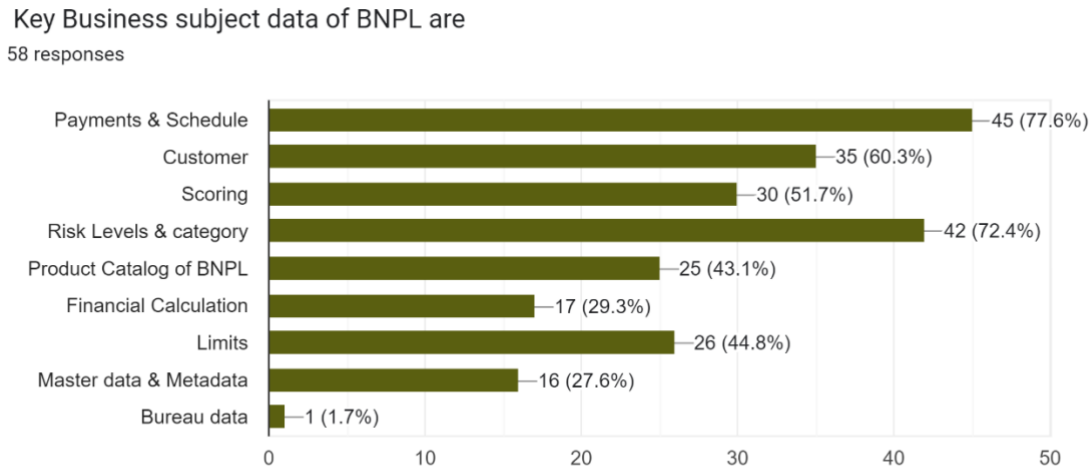


Figure 17: BNPL subject areas

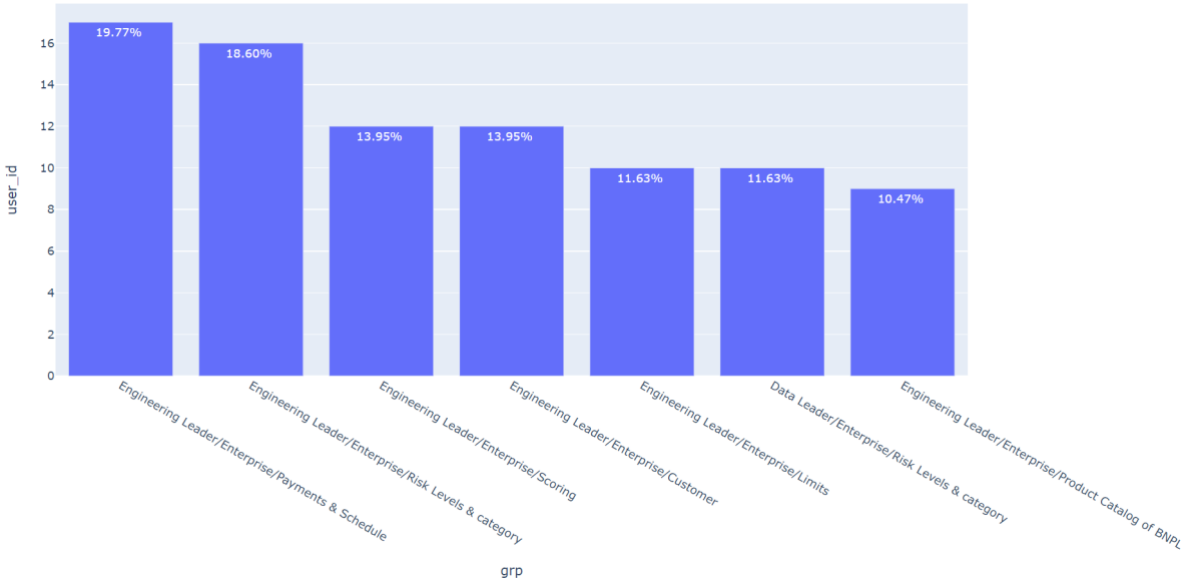


Figure 18: BNPL subject areas grouped by Role and Organization type

4.2.1.2.6. BNPL cost management

1. Based on the data presented in Figure 19, it is clear that BNPL cost is due to risk management and fraud protection strategies that the BNPL providers need to perform.

2. The data results in Figure 19 show that customer acquisition & retention, the global economic situation, and operation expenses contribute to the BNPL cost.

3. The data results in Figure 20 show that BNPL cost effectiveness is driven by good customer acquisition and retention strategies defined, availability of technology & tools and proper risk management.

4. The results in the Figure 21 shows that BNPL data lake cost is measured with the licensing cost and from the cloud-offered tools.

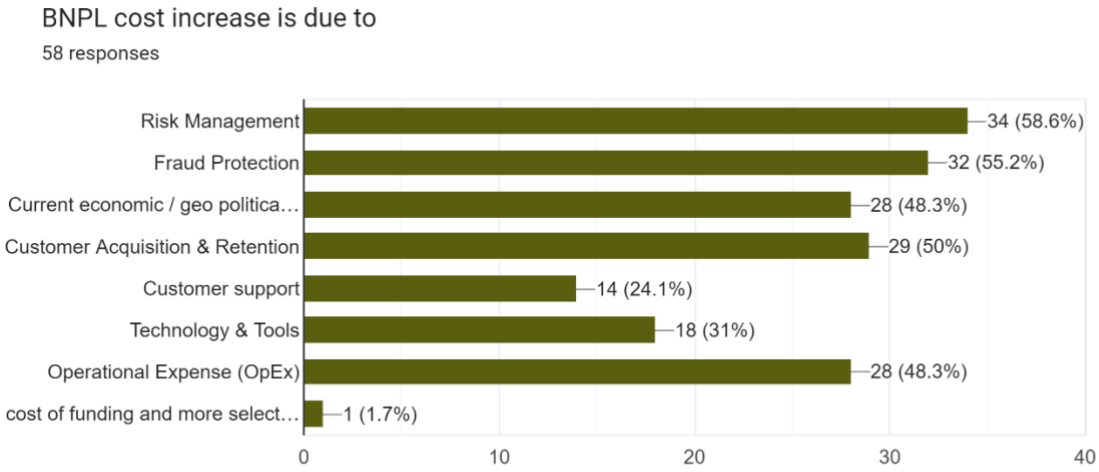


Figure 19: BNPL cost increase

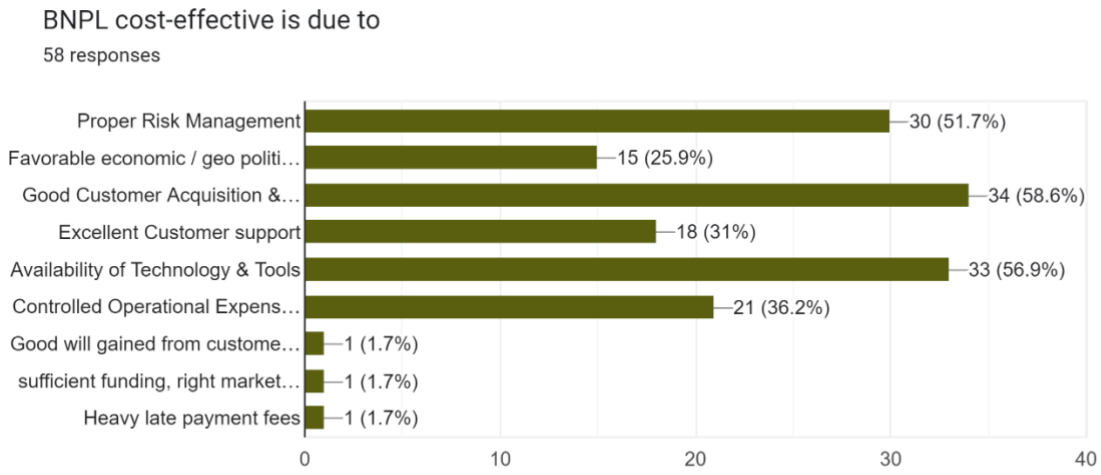


Figure 20: BNPL cost-effectiveness

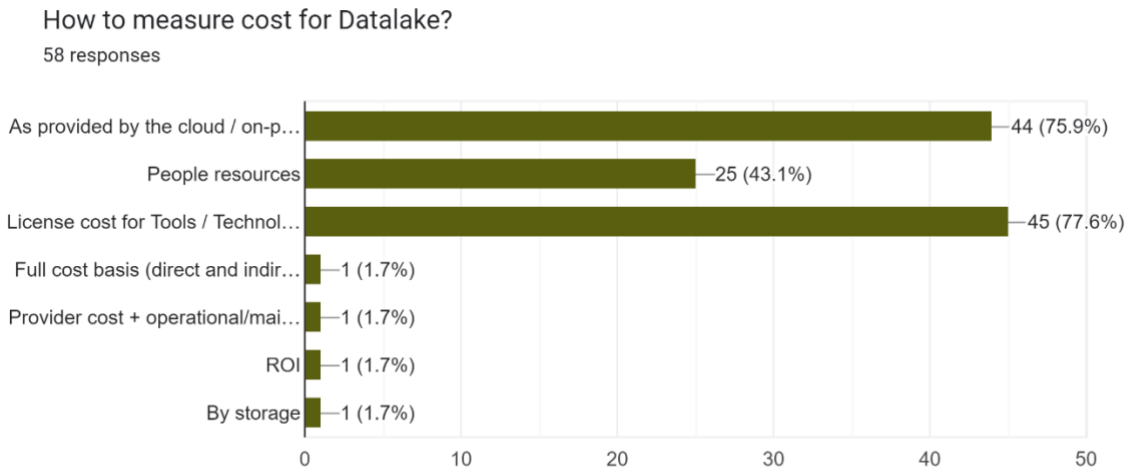


Figure 21: BNPL data lake measuring factors

4.2.1.3. Conclusion

The results show that the below are the critical drivers for a successful Fintech – BNPL data lake strategy.

a. Data security is crucial for Fintech data lake to build customer trust and protect sensitive information. By implementing these best practices, Fintech data lake can help ensure their systems are secure and their customers' data is protected.

b. Compliance and regulation are critical considerations for Fintech companies. By complying with regulatory requirements and implementing robust compliance programs, Fintech data lake can build trust within the organization and ensure they operate legally and securely.

c. BNPL data lake needs to implement robust risk management measures to ensure they can operate safely and sustainably in a highly competitive and rapidly evolving market. By managing credit, fraud, operational, Compliance, and reputation risks effectively, the BNPL data lake can build data divinity within the data.

d. BNPL data lake must take a proactive and holistic approach to fraud protection to ensure it can operate safely and sustainably in a highly competitive and rapidly evolving market. By implementing multi-factor authentication, transaction monitoring, and data privacy and security measures, the BNPL data lake can be built robustly.

e. Risk scoring is a critical component of BNPL data lake - risk management. Using data analytics and statistical models to assess creditworthiness and determine risk, BNPL providers can make more informed decisions about extending credit and managing credit risk using the data lake. By collecting data from various sources, using machine learning models, considering multiple factors, scoring risk in real-time and continuously monitoring creditworthiness, the BNPL data lake can improve accuracy, reduce risk, and build trustable data.

f. Customer satisfaction in a BNPL data lake involves collecting, analyzing, and leveraging customer data to optimize customer experience and drive business growth.

g. BNPL data lake should ensure to capture of the payment obligations and any potential fees or penalties.

h. BNPL data lake may also have to build different risk models that assess the likelihood of default on a case-by-case basis. These models may consider factors such as the consumer's credit score, income, employment status, and other financial obligations.

4.2.2. Descriptive Analysis of the Engineering Aspects for BNPL Data lake

4.2.2.1. Data lake engineering

To understand the data lake aspects for Fintech – BNPL, it is essential to understand engineering challenges with architecture, ML, data governance, and others that are required for enabling data-driven decisions. Hence the questions in Table 14 are part of the survey.

Table 14: Survey questionnaire – Data engineering view on Data lake

Q.No	Factor Type	Question	Expected Responses (Can be more than one)
1	Data lake architectural / engineering factors	Your preference for BNPL data lake	1. On-prem Data lake 2. On-prem Datawarehouse 3. Private cloud data lake 4. Public cloud data lake

			5. Hybrid cloud data lake 6. Multi-cloud data lake 7. Cloud data lake based on service offering (IaaS, PaaS, SaaS, etc.)
2	Data lake architectural / engineering factors	Preferred scalable architectures for data lake	1. Kappa Architecture (For speed layer) 2. Data mesh architecture (For decentralized data lake) 3. Dynamo Architecture - Distributed Hash Table (DHT) (For columnar, Key-value store) 4. GFS / HDFS Architecture (For distributed file system) 5. Event-driven

			Architecture (For asynchronous messaging) 6. Microservices Architecture (for Data API) 7. Chubby Architecture (For locking service) 8. Data warehousing architecture
3	Data lake architectural / engineering factors	Which DSA works best for handling data in distributed systems	1. SST - Sorted String Table 2. LSM - Log-Structured Merge 3. B+ 4. B Trees 5. Memtable 6. I prefer going with what is offered by the distributed file system 7. Other
4	Data lake architectural / engineering	Structured / semi-structured data lake is easy to source, maintain	Rating 1 to 10

	factors	and manage.	
5	Data lake architectural / engineering factors	Unstructured data lake is easy to source, maintain and manage	Rating 1 to 10
6	Data lake architectural / engineering factors	Underlying data structure and the algorithm are important for data lake design.	Rating 1 to 10
7	Data lake architectural / engineering factors	Preferred data model for BNPL data lake	1. Relational model 2. Document model 3. Graph model 4. Polyglot / Multi-model 5. Other
8	Data lake ML factors	What are the critical elements from Data lake for effective Al/ML?	1. Tools and Technology 2. Self-service Business Intelligence 3. ML pipelines 4. Feature Store 5. Expected sample and population of data 6. Data Quality 7. Data Discovery

			8. Data Integration 9. Other
9	Data lake ML factors	What are BNPL Feature Stores expected in Data lake?	1. Customer 360 feature store 2. Merchant 360 feature store 3. Finance feature store (Book-keeping, financial metrics, etc.) 4. Payments / Transactions feature store 5. Product & Pricing feature store 6. Other

4.2.2.2. Data Engineering / Architecture Factors– Survey Results

4.2.2.2.1. BNPL data lake preference

1. The data presented in Figure 22 and Figure 23 shows that the hybrid and private cloud data lake are the most preferred for the enterprise organization BNPL.
2. Based on the data presented, Cloud data lake is the most preferred in the current trend over the On-prem solution.

3. The data results also show that multi-cloud data lake and Cloud data lake offerings are preferred modes within the cloud options based on the organization's needs.

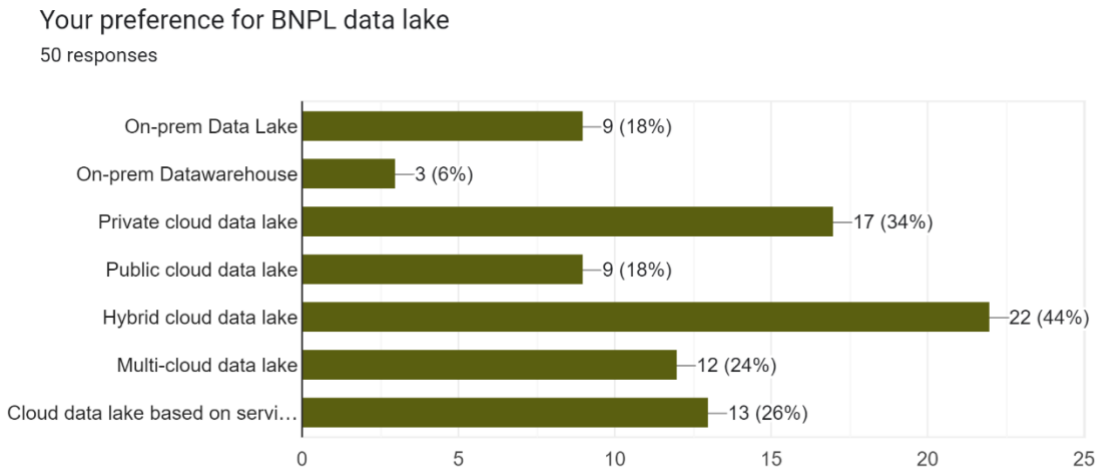


Figure 22: BNPL data lake preference

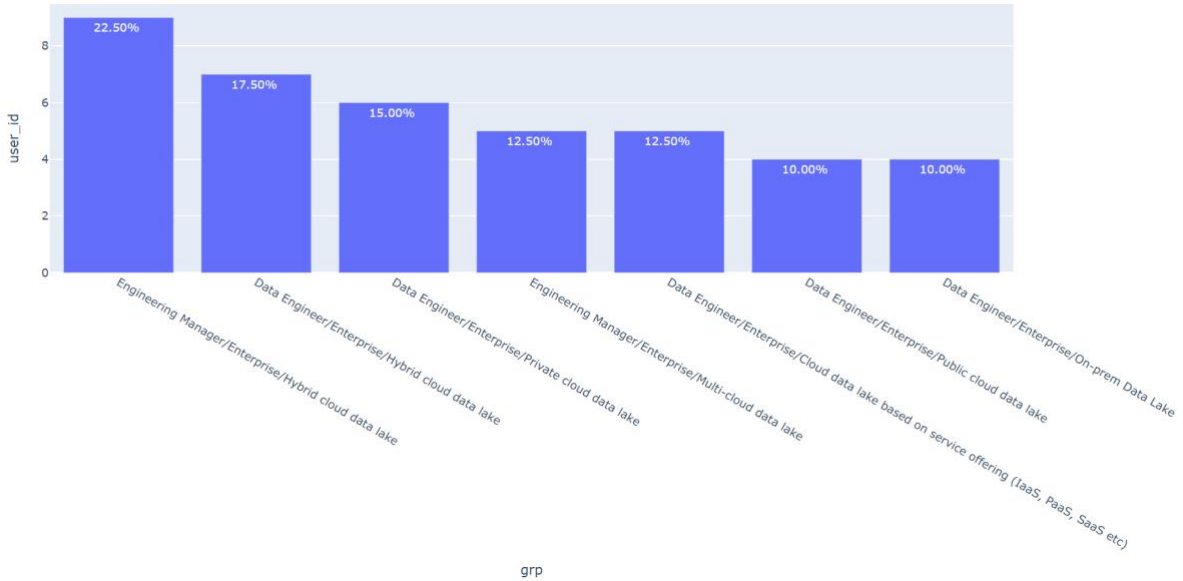


Figure 23: BNPL data lake preference grouped by Role and Organization type

4.2.2.2.2. BNPL data lake architecture preference

1. The data presented in Figure 24 shows that the microservices and data mesh architecture is the most preferred for the enterprise organization BNPL data lake.

2. Based on the data presented, GFS / HDFS architecture is preferred for the file system.

3. The data results also show that event-driven architecture is preferred for BNPL architecture.

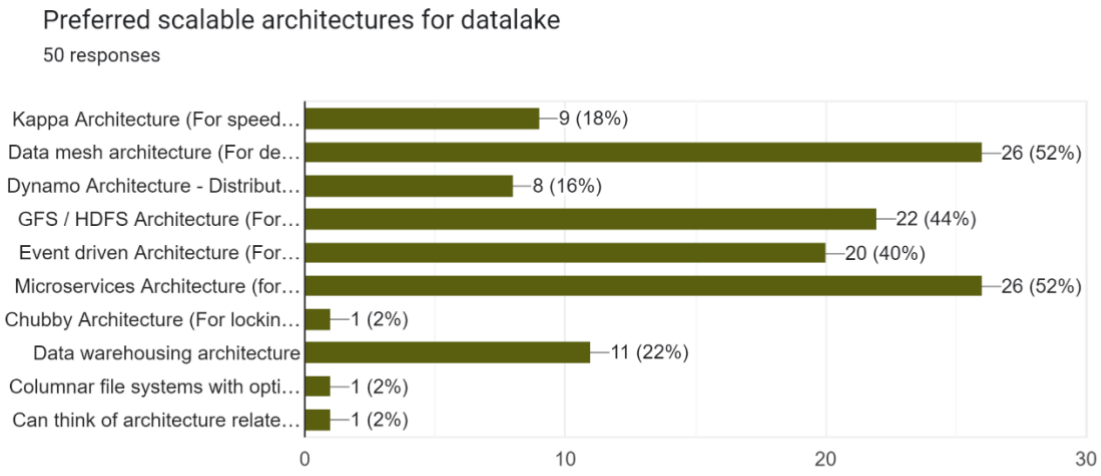


Figure 24: BNPL data lake architecture preference

4.2.2.2.3. BNPL data lake data structure preference

1. Based on the data presented in Figure 25, it shows that the data structure offered by the distributed file system is the most preferred for the BNPL data lake. Hence organization wants to spend less effort in building distributed file system. However, the data presented in the Figure 26 shows that the underlying data structure is significant for data lake. Hence, understanding the underlying data

structure involved with each distributed file system for the BNPL data lake use case is crucial.

2. Based on the data presented, if any specific data structure is to be chosen, then B+, B trees, and Memtable are the preferred data structure.

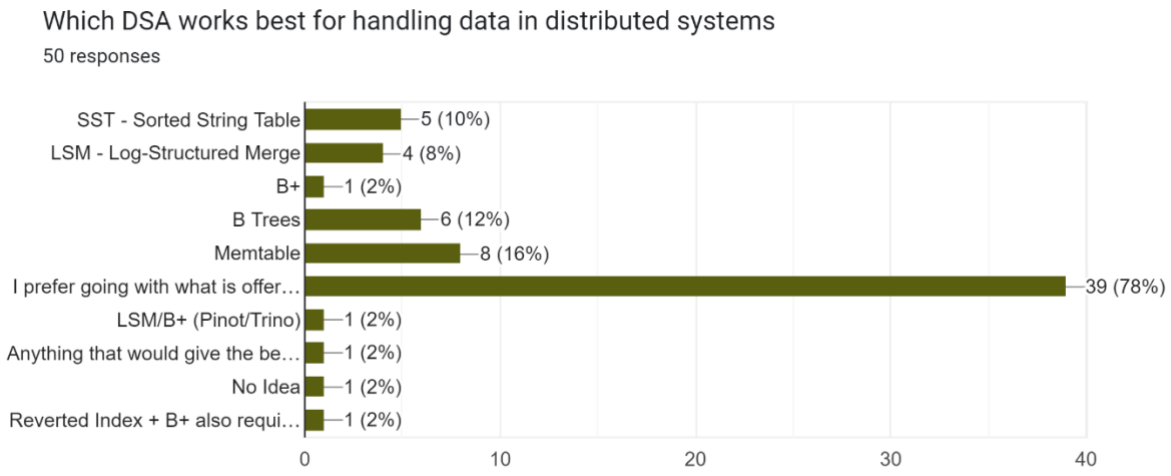


Figure 25: BNPL data lake data structure algorithm preference

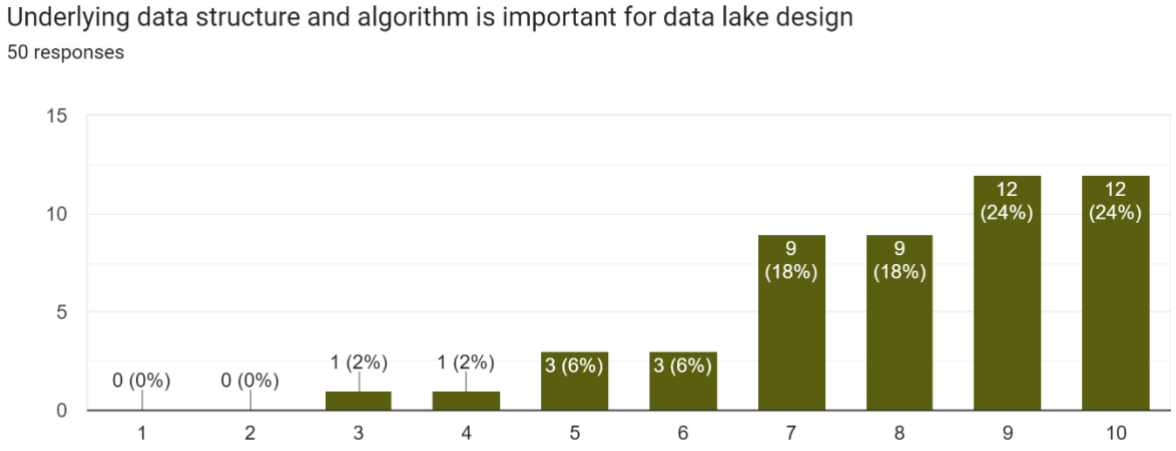


Figure 26: BNPL data lake data structure importance

4.2.2.2.4. BNPL data lake data variety and data model

1. Based on the data presented in Figure 27 and Figure 28, managing structured / semi-structured data is much easier and more manageable than unstructured data.

2. Based on the data presented in Figure 29, a relational data model is most preferred over the No-SQL data model.

3. The data results in Figure 29 also show that the NoSQL used for unstructured data has the Document and graph model and the Polyglot model preferred.

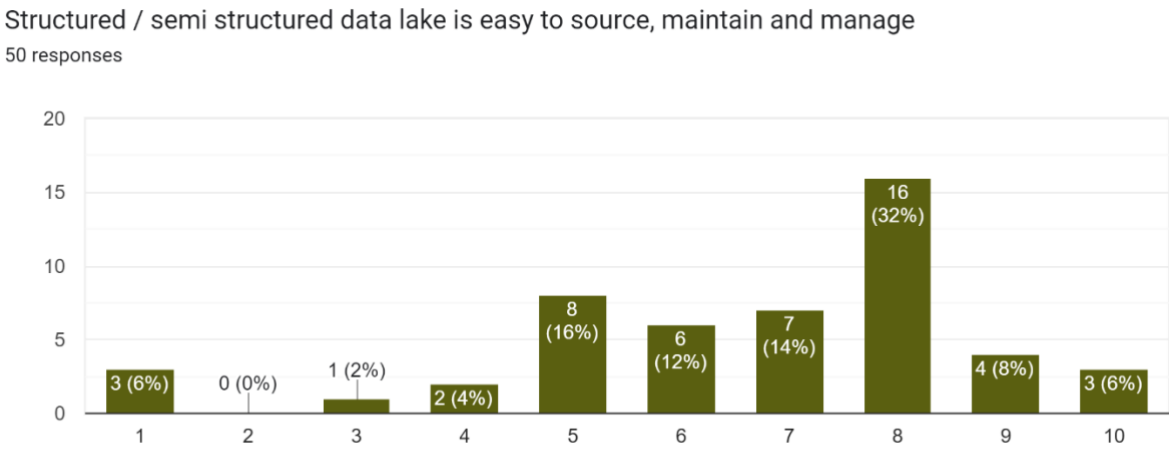


Figure 27: BNPL data lake structured/semi-structured data management

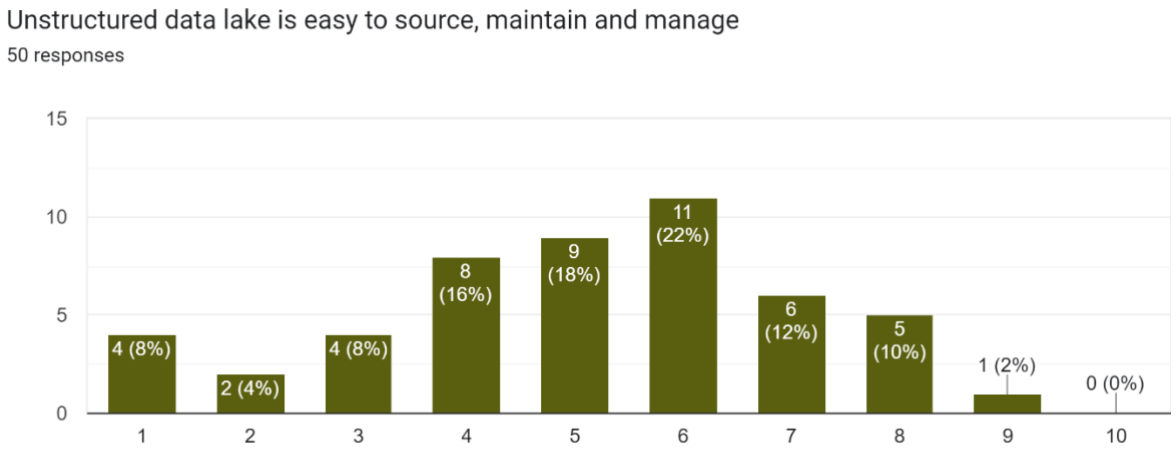


Figure 28: BNPL data lake unstructured data management

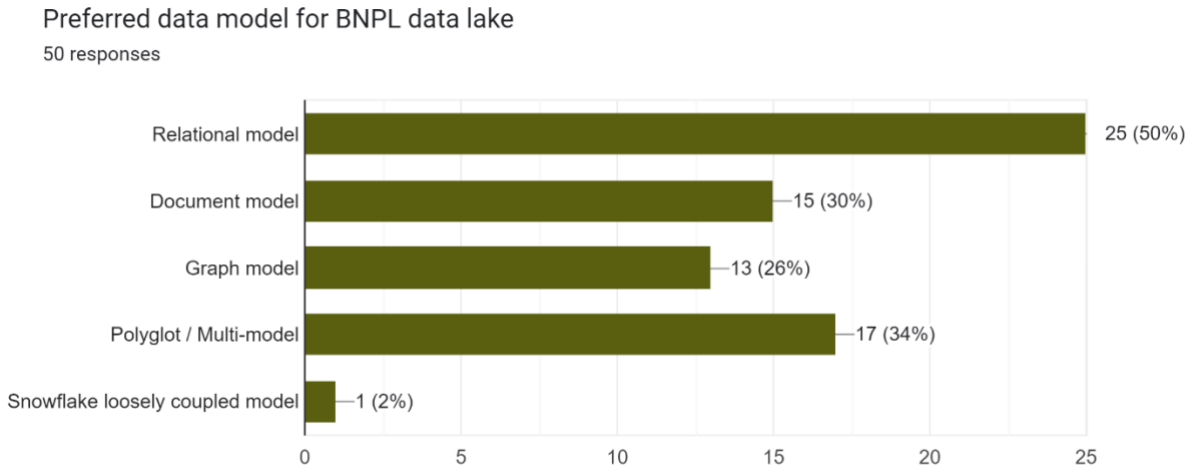


Figure 29: BNPL data lake preferred data model

4.2.2.3. Data lake ML factors – Survey Results

1. Based on the data presented in Figure 30, data quality is the primary concern for BNPL data lake – ML stakeholders.

2. Also, the data results show that significant drivers for effective AI/ML are ML pipelines, related tools & technologies, and data integrations.

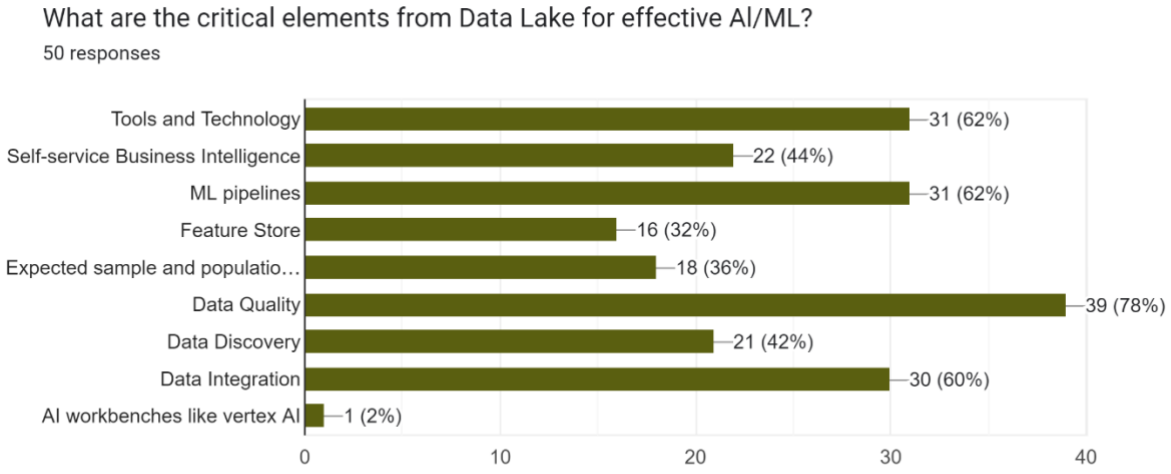


Figure 30: BNPL data lake critical elements for AI/ML

3. The data presented in Figure 31 shows that the feature store is very required for

payments & transactions, Customer 360, Product & Pricing, and Merchant 360. Based on the data presented in Figure, a feature store is very much required.

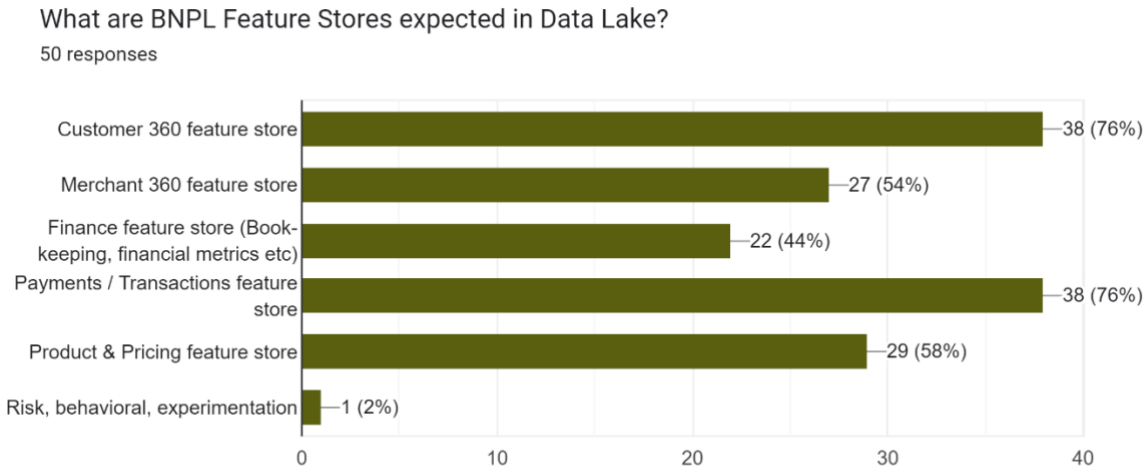


Figure 31: BNPL data lake – Feature stores

4.2.2.4. Conclusion

The results show that the below are the critical drivers for a successful Fintech – BNPL data lake strategy from the engineering standpoint.

Data lake preference – on whether an On-prem and cloud data lake should be preferred for BNPL. The results show that the cloud data lake is the most preferred within which the private, hybrid, and multi-cloud are the most desired.

Data lake architecture preference – Time to market and scalability are the core drivers for choosing the data lake architecture. Data mesh architecture, microservice architecture, and event-driven architecture are preferred for meeting these criteria.

Underlying Data structure is very important and preferred based on the existing distributed file system offerings rather than building one.

Structured and unstructured data are preferred for the BNPL data lake though managing structured data is relatively easier than unstructured data. Also, the data model to support them is crucial for data management; hence the right decision to be taken between the SQL and NoSQL data model based on the nature of the data.

For effective AI/ML on the BNPL data lake, data quality, data integration, data pipelines, and the tools & technologies supporting them are the key drivers. Also, the features store is preferred for Payments & transactions, Customer 360, Product & pricing, and Merchant 360.

4.2.3. Influencing Factors for Business and Data lake Framework

Common factors influence both Business and Technology, especially with Fintech – BNPL, including the following.

1. BNPL cost management
2. BNPL time to value
3. BNPL time to market
4. BNPL data quality
5. BNPL data security
6. BNPL data governance

Hence it is essential to understand those parameters addressed with the questions in Table 15 as part of the survey.

Table 15: Survey questionnaire – Influencing factors on Data lake

Q.No	Factor Type	Question	Expected Responses (Can be more than one)
1	Cost Management	BNPL cost increase is due to	1. Risk Management 2. Fraud Protection 3. Current economic/geopolitical / Inflation factors 4. Customer Acquisition & Retention 5. Customer support 6. Technology & Tools 7. Operational Expense (OpEx) 8. Other
2	Cost Management	BNPL's cost-effective is due to	1. Proper Risk Management 2. Favorable economic/geopolitical / Inflation factors 3. Good Customer Acquisition & Retention strategy 4. Excellent Customer support 5. Availability of Technology & Tools 6. Controlled Operational

			Expense (OpEx) 7. Other
3	Cost Management	How to measure the cost for Data lake?	1. As provided by the cloud / on-prem provider 2. People resources 3. License cost for Tools / Technologies 4. Other
4	Cost Management	With effective data management and governance, will the OpEx cost increase or decrease?	1. Increase by 5 - 10% 2. Increase by 10 - 20% 3. Increase by 20 - 40% 4. Decrease by 5 - 10% 5. Decrease by 10 - 20% 6. Decrease by 20 - 40% 7. Other
5	Time to Value	BNPL Data - Reasons for the delay in Time to Value	1. Delay / missing Data pipeline setup 2. Delay/missing Data Analytics & Insights 3. Improper AI/ML approaches 4. Other
6	Time to Value	The lack of SME is due to	1. Data Engineers for building data-intensive applications 2. Data Analysts in BNPL 3. Data Architects for breaking

			monolithic and centralized architectures 4. Data Scientists for niche BNPL skill 5. Other
7	Data Security	BNPL data security is at stake due to	1. Data leaks 2. Fraud transactions by impersonating lenders 3. Hackers 4. Poor/missing data infrastructure
8	Data Quality	Significant areas causing BNPL Data quality issues	1. Payments & Schedule 2. Customer 3. Scoring 4. Risk Levels & category 5. Product Catalog of BNPL 6. Financial Calculation 7. Limits 8. Master data & Metadata 9. Other
9	Data Quality	With effective data quality and data engineering practices, will garbage dump increase or decrease?	1. Increase by 5 - 10% 2. Increase by 10 - 20% 3. Increase by 20 - 40% 4. Decrease by 5 - 10% 5. Decrease by 10 - 20%

			6. Decrease by 20 - 40%
10	Time to Market	Effort required to build data lake from scratch	1. Startup - < 2 months 2. Startup - > 2 months 3. SMB - 3 - 4 months 4. SMB - 6 months 5. Enterprise > 6 months 6. Enterprise > 12 months
11	Time to Market	Effort required to migrate data lake to new tech stack without modernization	1. Startup - < 1 month 2. SMB < 2 months 3. Enterprise < 3 months 4. Other
12	Time to Market	Effort required to migrate data lake to new tech stack with modernization	1. Startup - < 4 months 2. SMB - 6 - 8 months 3. Enterprise > 12 months
13	Data Governance	BNPL Data is valuable and usable with	1. Source-aligned data (raw data without any transformation) 2. Consumer-aligned data (fit-for-purpose data) 3. Aggregate-domain data (OLAP - analytical mart) 4. Data discovery & Data Catalog Management

			5. Data Privacy 6. Other
14	Data Governance	Federated governance is the key to a distributed architecture, and the key governance policy is	1. Standards as Code 2. Policies as code 3. Automated Tests 4. Automated Monitoring 5. Data Privacy 6. Other
15	Data Governance	Reasons for Garbage dump in BNPL data lake	1. Building data lake just for the sake of it with all sources of data 2. Never transforming the raw data into meaningful insights 3. Redundant data 4. Duplicate data 5. Missing data governance 6. Not interpreting business value from data
16	Data Quality	% of data quality issues in data lake	1. 0-20% 2. 20-40% 3. >40% 4. Data quality is not measured though implemented 5. Other

17	Data Security	BNPL data security is at stake due to	1. Data leaks 2. Fraud transactions by impersonating lenders 3. Hackers 4. Poor/missing data infrastructure 5. Other
18	Data Security	BNPL security drivers are	1. Best encryption/decryption protocols 2. On-prem data lake solution 3. Private Cloud data lake 4. Public / Multi-Cloud data lake with security on clusters, nodes, files, and data attributes 5. Other
19	Data Security	% of data security issues in data lake	1. 0-1% 2. 2-5% 3. 5-10% 4. Data Security is not measured though implemented 5. Other

4.2.3.1. BNPL cost management

1. Based on the results in Figure 32, effective data governance with a BNPL data lake reduces the operational excellence between 5% -20% based on the effectiveness of data governance.

2. For effective cost management in BNPL, it is essential to understand the parameters contributing to the cost. Cost management in BNPL refers to the strategies and techniques used to manage the costs associated with offering BNPL services to customers.

    a. One of the critical aspects of cost management in BNPL is risk management. Since BNPL providers assume the risk of default by customers, it is crucial to assess and manage this risk to minimize losses. This can include implementing credit checks, setting credit limits, and monitoring customer behavior and payment patterns closely.

    b. Another critical aspect of cost management in BNPL is optimizing the business's cost structure. This can include reducing overhead costs such as rent and salaries, streamlining operational processes, and negotiating favorable terms with suppliers and partners.

    c. Additionally, BNPL providers can use technology and data analytics to improve cost management. For example, they can use machine learning algorithms to analyze customer behavior and predict default rates & fraudulent transactions or implement automated systems for collections and customer service to reduce labor costs.

    d. Overall, effective cost management in BNPL is essential for ensuring the long-term viability and profitability of the business while providing a valuable

service to customers.

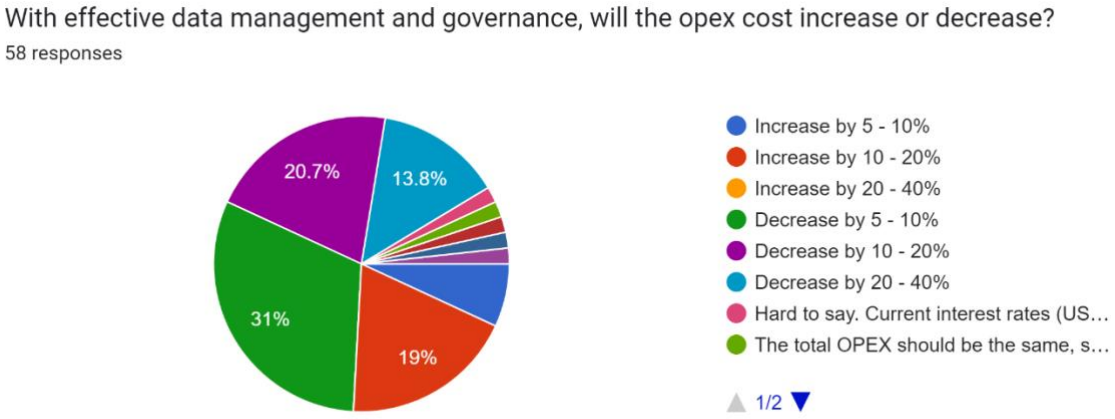
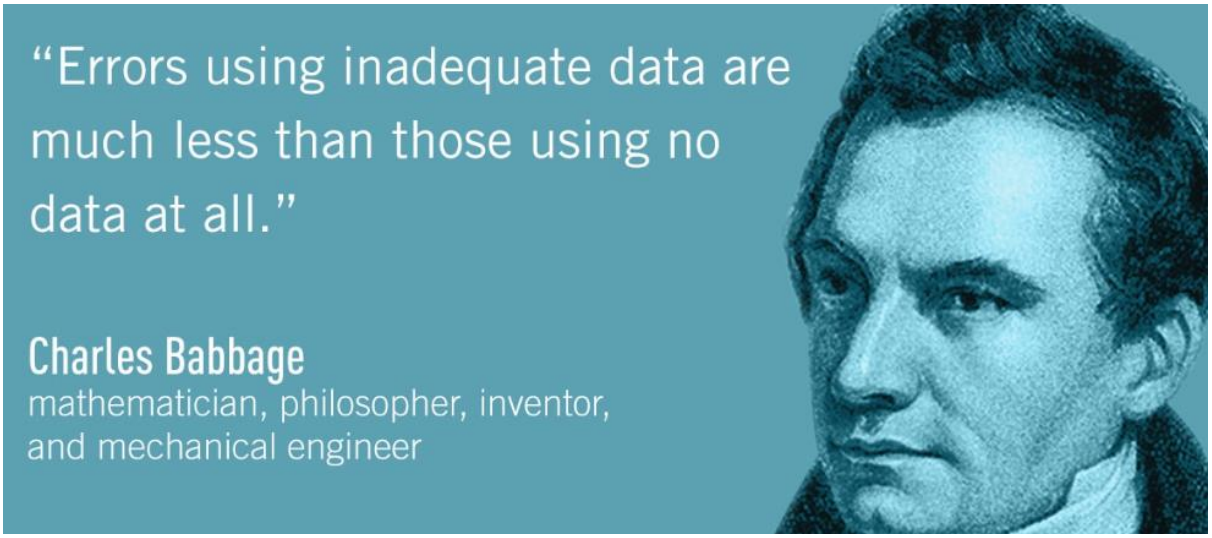


Figure 32: BNPL data lake – Operational cost

4.2.3.2. BNPL Time to value



'Lack of SMEs' is one of the attributes contributing to BNPL's time to value. The data in the Figure 34 shows that the lack of SMEs resides with data architects, data scientists, and data engineers.

Time to Value (TTV) in BNPL data lake refers to the time it takes for a BNPL provider

to realize the benefits and value of their data lake investments. This includes the time it takes to extract insights from the data, make data-driven decisions, and derive tangible business benefits from the insights.

1. Based on the data presented in Figure 33, the reason for BNPL data lake time to value is either delay / missing analytics or delay / missing data pipelines, or improper AI/ML pipelines.
2. To reduce TTV (Time-To-Value) in the BNPL data lake, providers can adopt the following strategies:
    a. Set clear objectives: BNPL providers should clearly define their business objectives for the data lake and set key performance indicators (KPIs) to measure the success of the data lake initiatives. This will help focus the data analytics efforts and ensure they align with business goals.
    b. Prioritize use cases: BNPL providers should prioritize use cases that provide the most significant business value and impact, such as fraud detection, customer segmentation, and risk management, and focus their data analytics efforts on these use cases first.
    c. Ensure data quality: To ensure that the insights derived from the data lake are accurate and reliable, BNPL providers should ensure that the data stored in the data lake is high quality, complete, and up-to-date.
    d. Automate data processing: BNPL providers can use automation tools and techniques, such as machine learning algorithms, to automate data processing and analysis, enabling faster and more accurate insights.
    e. Collaborate across functions: BNPL providers should foster collaboration between data analysts, business stakeholders, and IT teams to ensure that insights are shared and acted upon quickly and that the data lake is used to drive business

value.

f. By reducing TTV in the BNPL data lake, providers can accelerate the time it takes to realize business benefits and derive value from their data investments. This will enable them to make data-driven decisions more quickly and effectively, improving their competitiveness and bottom line.

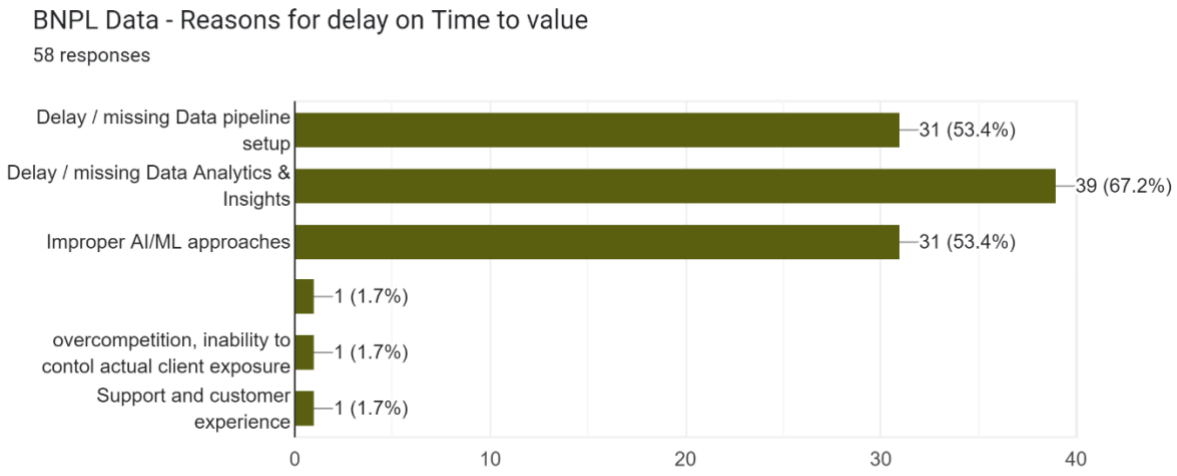


Figure 33: BNPL data lake – Time to value

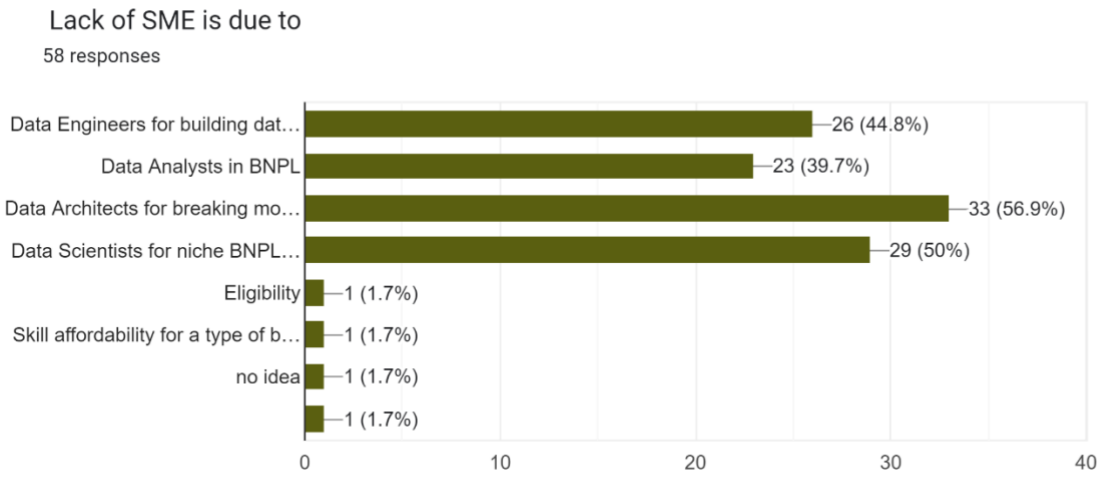


Figure 34: BNPL data lake – Lack of SME

4.2.3.3. BNPL Time to Market

Time to Market (TTM) in the BNPL data lake refers to the time it takes to develop and launch new data-driven products or services that can provide value to customers. Reducing TTM in the BNPL data lake is crucial for staying ahead of the competition and

meeting customers' changing needs and expectations.

Time to market varies based on the nature of building the data lake for a Startup, SMB, or Enterprise.

a. When the data lake requires to be built from scratch

b. When the data lake requires to be migrated without modernization. (Lift and shift)

c. When the data lake requires to be migrated with modernization

1. Based on the results in Figure 35, it appears that for building the data lake from scratch it takes 6-12 months for Enterprise, Startup in < 2 months

2. Based on the results in Figure 36, it appears that for migrating the data lake without modernization as Lift & Shift it takes < 3 months for Enterprise.

3. Based on the results in Figure 37, it appears that for migrating the data lake with modernization is treated as complex compared with the other options and takes more than 12 months for Enterprise.

4. Some of the strategies that can help reduce TTM (Time-To-Market) in the BNPL data lake:
   a. Agile development process: BNPL providers can adopt an agile development process involving close collaboration between developers, data analysts, and business stakeholders to quickly develop and test new data-driven products and services.
   b. Data quality: BNPL providers should ensure that the data stored in the data lake is accurate, complete, and up-to-date to enable faster and more reliable data analysis

and decision-making.

c. Automation: BNPL providers can use automation tools and platforms to accelerate the development and deployment of data-driven products and services, such as machine learning models for fraud detection or customer segmentation.
d. Scalability: BNPL providers should design the data lake architecture to be scalable and flexible to easily accommodate new data sources and analytics tools as the business grows and evolves.
e. Data governance: BNPL providers should establish clear data governance policies and procedures to ensure that data is used ethically and in Compliance with regulatory requirements while also enabling faster data analysis and decision-making.
f. By reducing TTM in the BNPL data lake, providers can accelerate innovation and improve their ability to deliver data-driven products and services that provide value to customers while maintaining data quality, governance, and security.

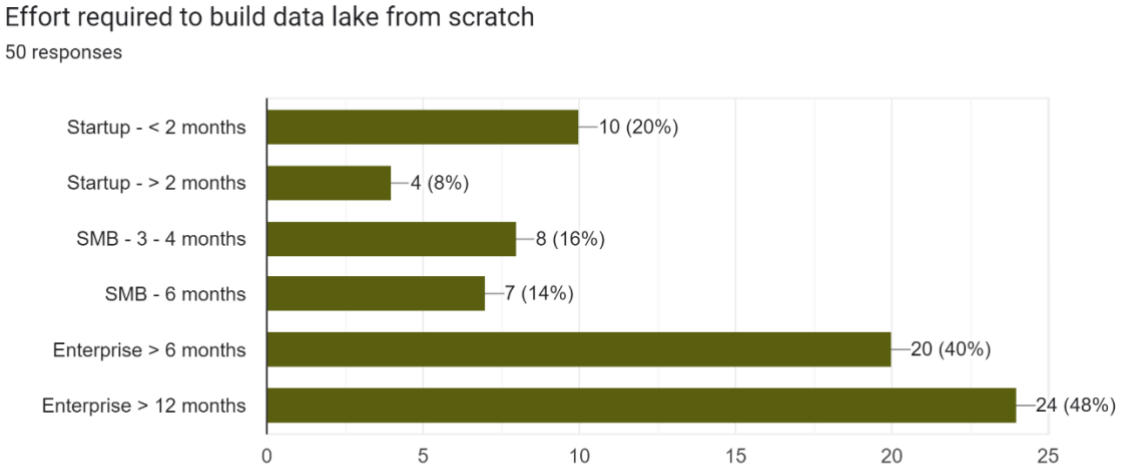


Figure 35: BNPL data lake – TTM – Build from scratch

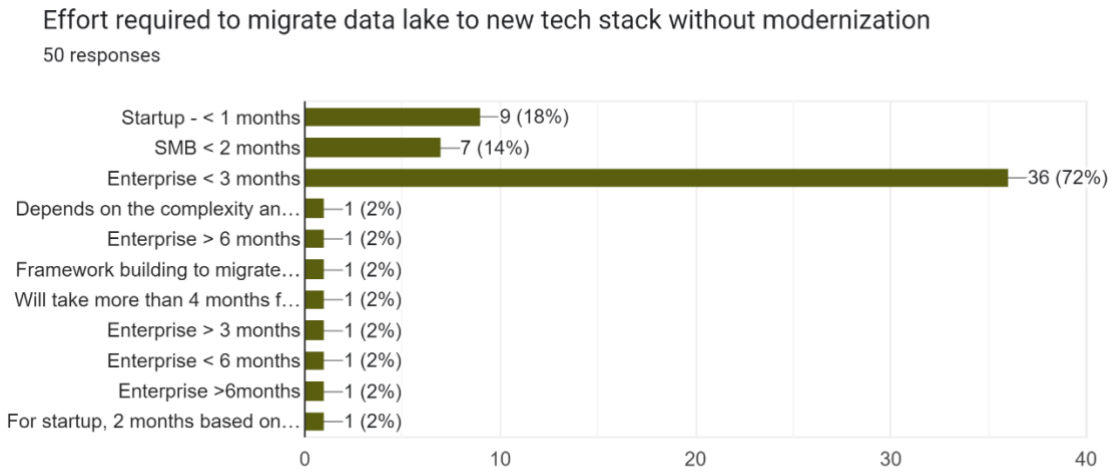


Figure 36: BNPL data lake – TTM – Migrate without modernization

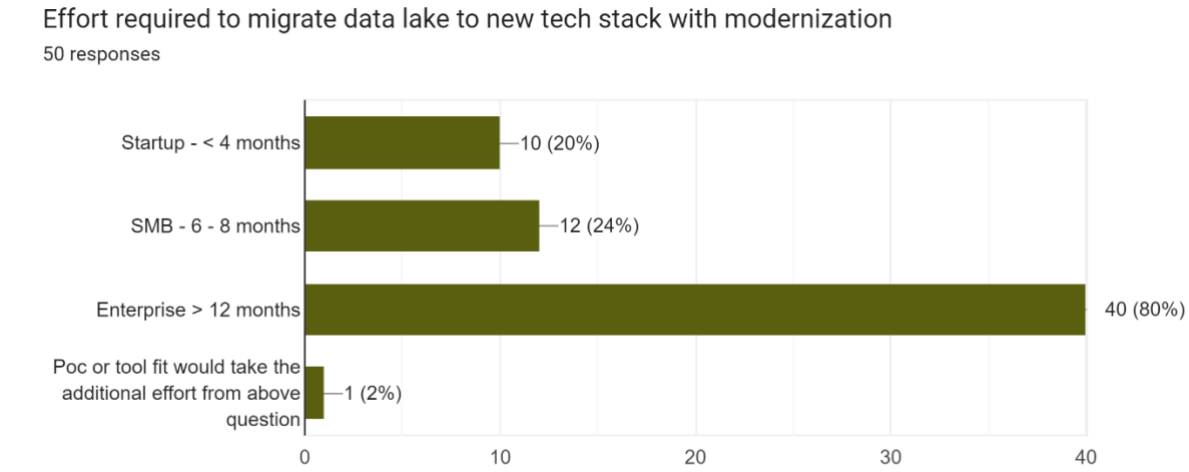
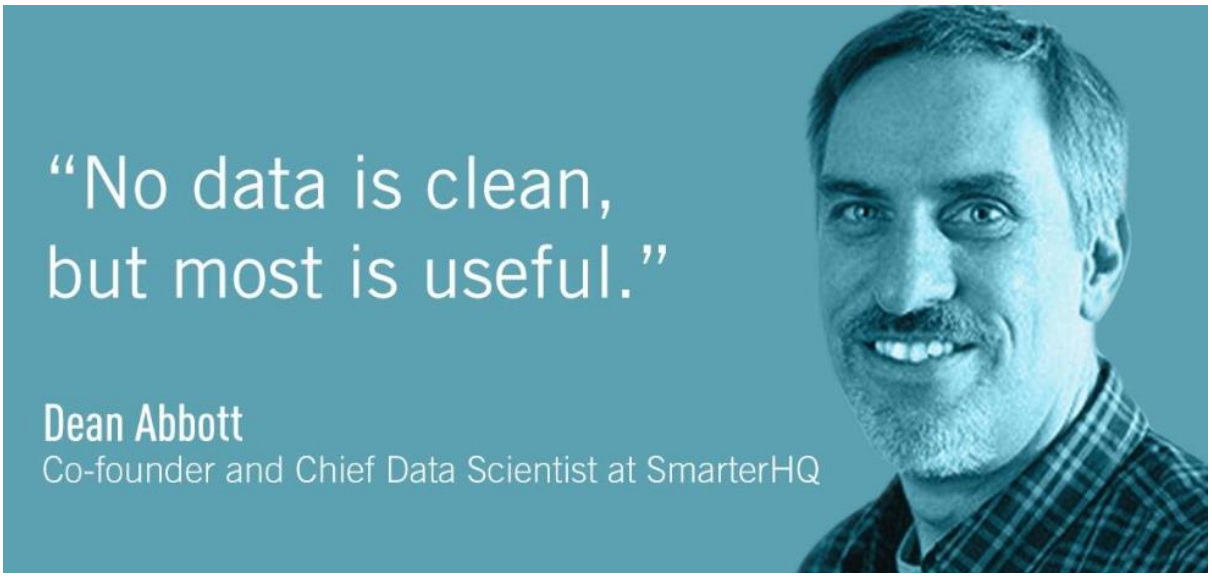


Figure 37: BNPL data lake – TTM – Migrate with modernization

4.2.3.4. BNPL data quality



Data quality is an essential factor in the success of a BNPL data lake. Poor data quality can negatively impact the accuracy and reliability of the data lake's insights, leading to incorrect decisions, missed opportunities, and financial losses.

1. Based on the results from the Figure 38, data quality issues concern the subject area risk levels & category and payments & schedule considerably more than the other area.

2. Based on the results from the Figure 39, the % of data quality issues in the BNPL data lake are between 0%-40%. The results show that data quality is not being given importance within the data lake to have zero-defect tolerance.

3. Some of the strategies that BNPL providers can use to ensure data quality in their data lake:

    a. Data profiling and cleansing: BNPL providers can use data profiling techniques to identify data quality issues, such as missing values, duplicates, and inconsistencies, and then clean the data to ensure that it is accurate,

complete, and consistent.

b. Data validation: BNPL providers should establish validation rules to ensure the data stored in the data lake meets predefined quality standards. For example, data validation rules can check that the data is within valid ranges or conform to specific formats.

c. Data governance: BNPL providers should establish clear data governance policies and procedures to ensure that data is managed and used consistently and compliant and that data quality is maintained over time.

d. Data lineage: BNPL providers should establish data lineage, which tracks the data's journey from its origin to its final destination. It will help to ensure data accuracy and enable data traceability for audit and compliance purposes.

e. Data security: BNPL providers should establish data security measures to ensure that the data stored in the data lake is secure and protected from unauthorized access, misuse, or theft.

f. By ensuring data quality in their data lake, BNPL providers can derive accurate and reliable insights from the data, leading to better decision-making, improved business performance, and increased customer satisfaction.

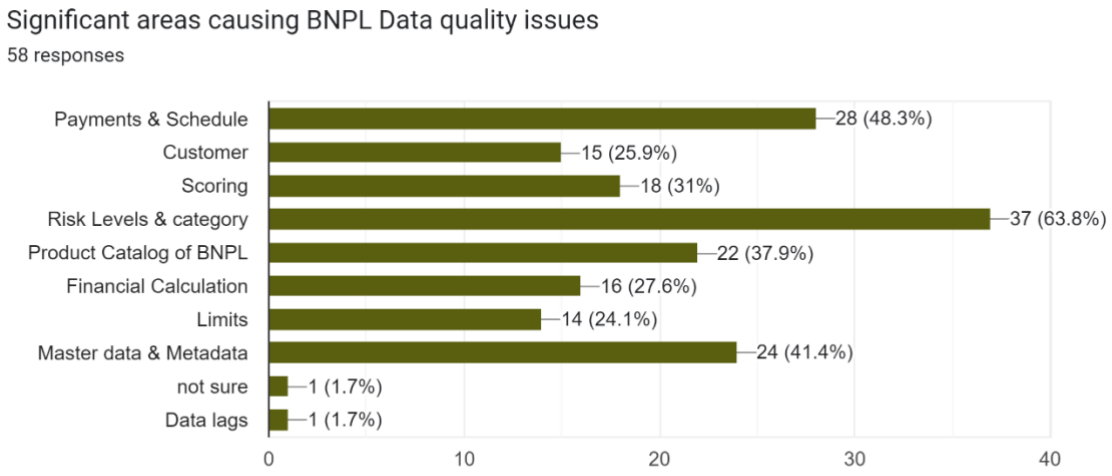


Figure 38: BNPL data lake – Data Quality areas

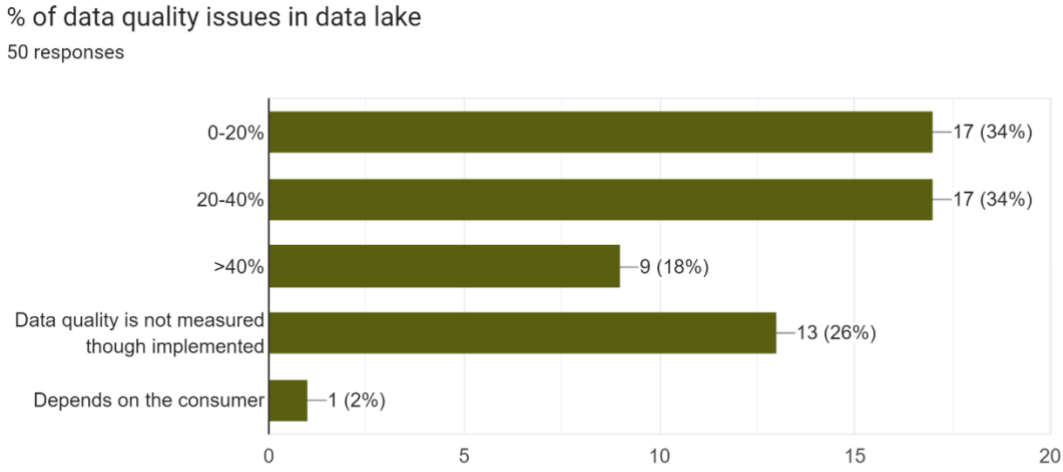


Figure 39: BNPL data lake – Data Quality issues

4.2.3.5. BNPL data security

Data security is crucial in the BNPL data lake to protect sensitive and confidential data from unauthorized access, misuse, or theft. Given the humongous amount of data stored

in the data lake, data security is a top priority for BNPL providers to maintain customer trust and comply with regulatory requirements.

1. Based on the results from the Figure 40 and Figure 41, data security is at risk is due to the fraud transactions impersonating lenders which is also linked to the hackers, a major risk and the infrastructure in use for protecting the security which gets vulnerable for data leaks.

2. Based on the results from the Figure 42, data security drivers for protection are with the implementation of best encryption & decryption protocols with the best cloud infrastructure.

3. Based on the results from the Figure 43, data security issues are mostly not measured though implemented and face a threat level with issues ranging majorly 5 – 10%.
4. Some of the strategies that BNPL providers can use to ensure data security in their data lake:
    a. Access control: BNPL providers can implement access control mechanisms to restrict access to the data lake based on user roles and permissions. It ensures that only authorized personnel can access sensitive data and perform actions on the data lake.
    b. Encryption: BNPL providers can use encryption to protect the data stored in the data lake from unauthorized access. It includes data encryption at rest and in transit, which helps to protect data from being intercepted or stolen during transmission.
    c. Network security: BNPL providers should ensure their network infrastructure is secure, with firewalls, intrusion detection and prevention systems, and other security measures to protect against external threats.

d. Data masking: BNPL providers can use data masking techniques to obscure sensitive data in the data lake from unauthorized users. It includes techniques such as data redaction, substitution, and scrambling, which help to protect sensitive data from unauthorized access.
e. Data backup and recovery: BNPL providers should establish data backup and recovery processes to protect data from data loss or corruption. It includes regular backups, disaster recovery plans, and testing of recovery procedures to ensure data availability in case of a disaster.
f. By ensuring data security in their data lake, BNPL providers can protect sensitive and confidential data, maintain customer trust, and comply with regulatory requirements. It will help to mitigate the risks of data breaches, financial losses, and reputational damage.

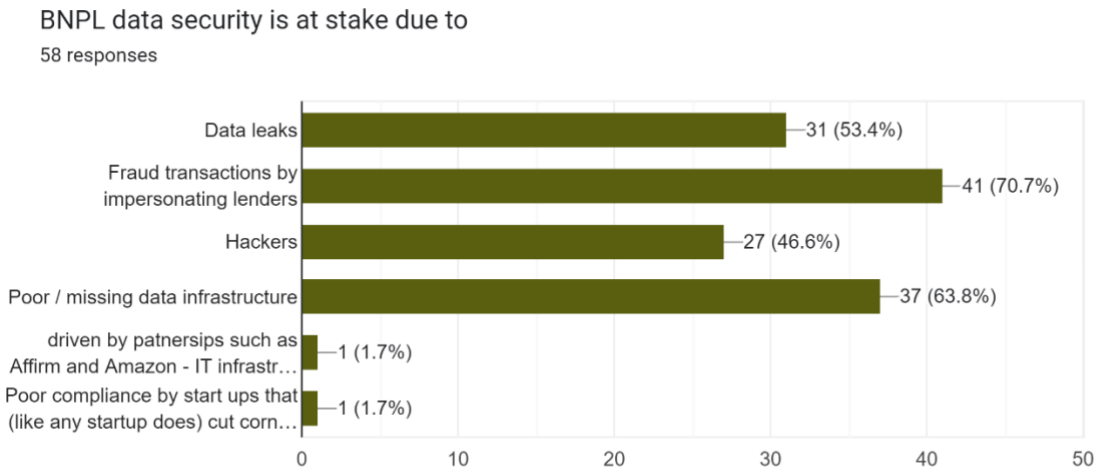


Figure 40: BNPL data lake – Data Security view by Fintech experts

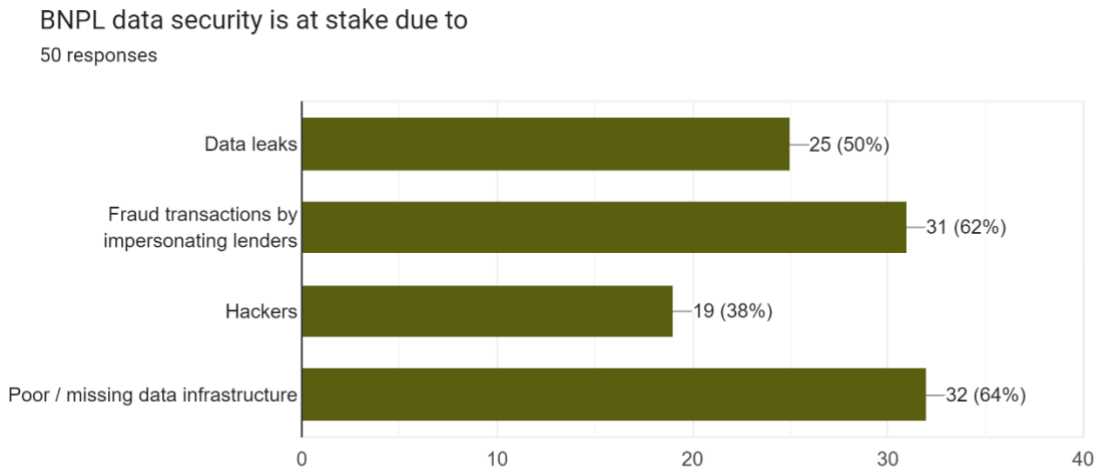


Figure 41: BNPL data lake – Data Security view by Data engineering experts

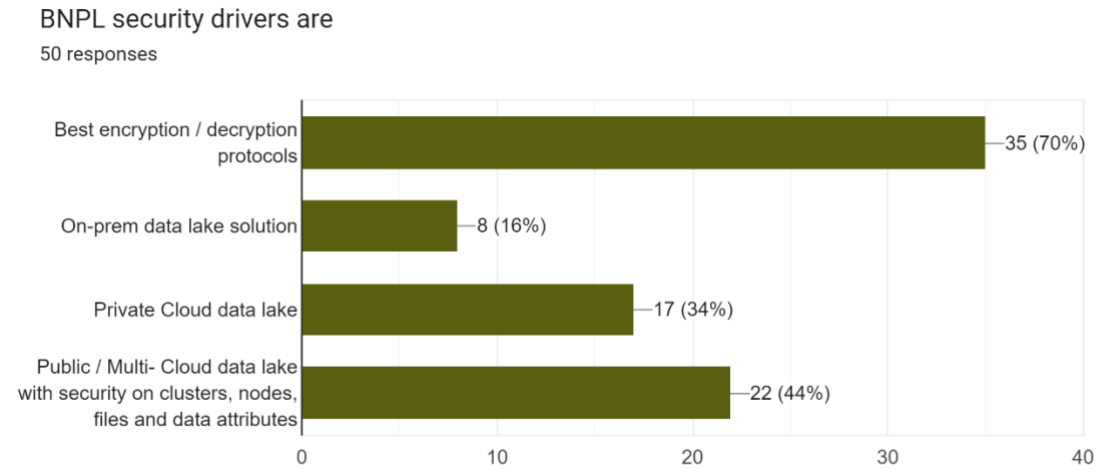


Figure 42: BNPL data lake – Data Security drivers

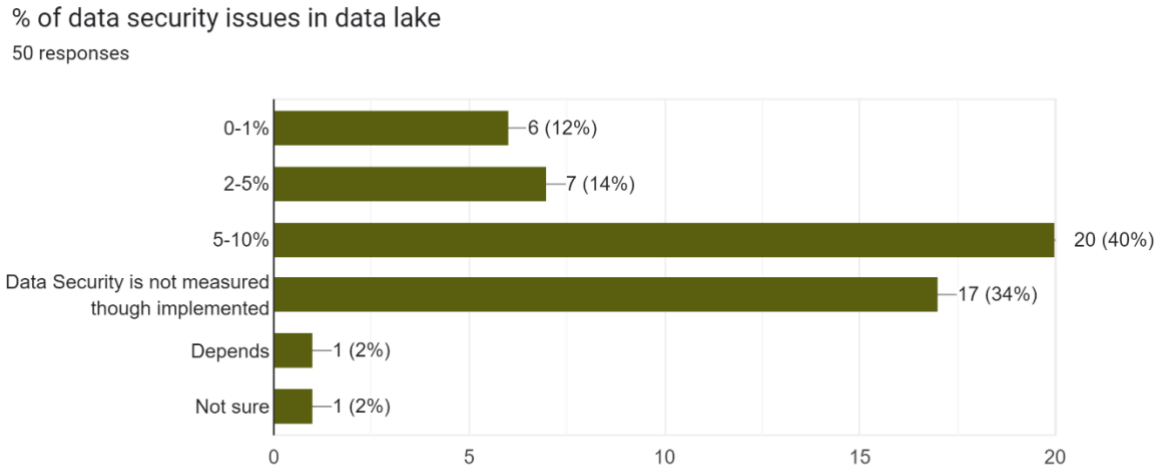


Figure 43: BNPL data lake – Data Security issues

4.2.3.6. BNPL data governance

Data governance is managing the availability, usability, integrity, and security of the data used in an organization. In a BNPL data lake, data governance is critical to ensure that the data used for various analytics and business intelligence purposes is accurate, consistent, and compliant with regulatory requirements.

1. Based on the results in Figure 44, with the effective data governance policies in place garbage dump decrease majorly between the range of 10 – 40%.

2. Based on the results in Figure 45, federated data governance should enable automated monitoring with policies and standards as a code, and support data privacy.

3. Based on the results in Figure 46, garbage dump within the BNPL data lake is due to missing data governance causing huge volume of data sourced without any need leading to missing insights or duplicate the data.

4. Based on the results in Figure 47, BNPL data lake to be as a data asset is driven by having the fit-for-purpose data and analytical data mart. Results show that it is also driven by data discovery & data catalog management and data privacy.

5. Some of the strategies that BNPL providers can use to establish data governance in their data lake:

    a. Data quality management: BNPL data lake should establish data quality management processes to ensure that data is accurate, complete, consistent, and relevant for the intended purpose. It includes data profiling, cleansing, and validation techniques to improve data quality.

    b. Data classification and metadata management: BNPL data lake should classify data based on sensitivity, confidentiality, and regulatory requirements. They should also establish metadata management processes to properly document, label, and track data.

    c. Data lifecycle management: BNPL data lake should establish data lifecycle

management processes to manage the data flow from creation to deletion or archival. It includes data retention policies, data archiving, and data disposal procedures.

d. Data access and security management: BNPL data lake should establish access and security management processes to ensure that only authorized personnel can access the data lake. It includes user authentication and authorization, data encryption, and data masking techniques to protect sensitive data.

e. Data governance policies and procedures: BNPL data lake should establish clear data governance policies and procedures that define the roles, responsibilities, and processes for managing the data lake. It includes data governance committees, data stewardship roles, and data governance frameworks.

By establishing data governance processes in their data lake, BNPL providers can ensure that the data used for various analytics and business intelligence purposes is accurate, consistent, and compliant with regulatory requirements. It helps to improve the quality of insights derived from the data, leading to better decision-making and improved business performance.

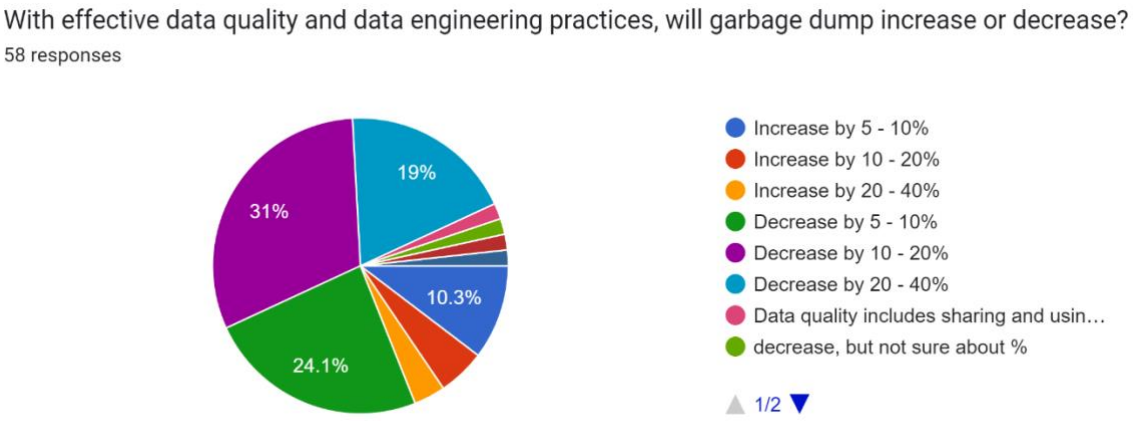


Figure 44: BNPL data lake – Garbage dump

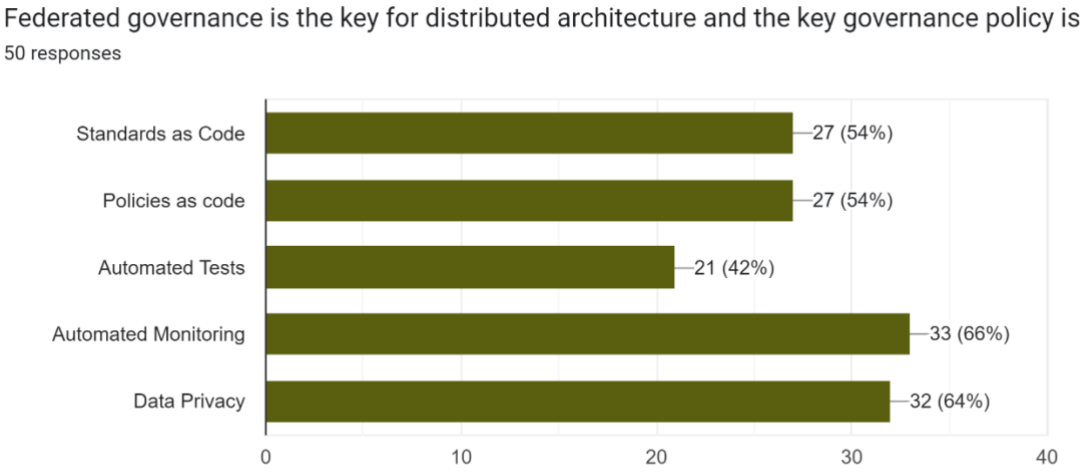


Figure 45: BNPL data lake – Data Governance

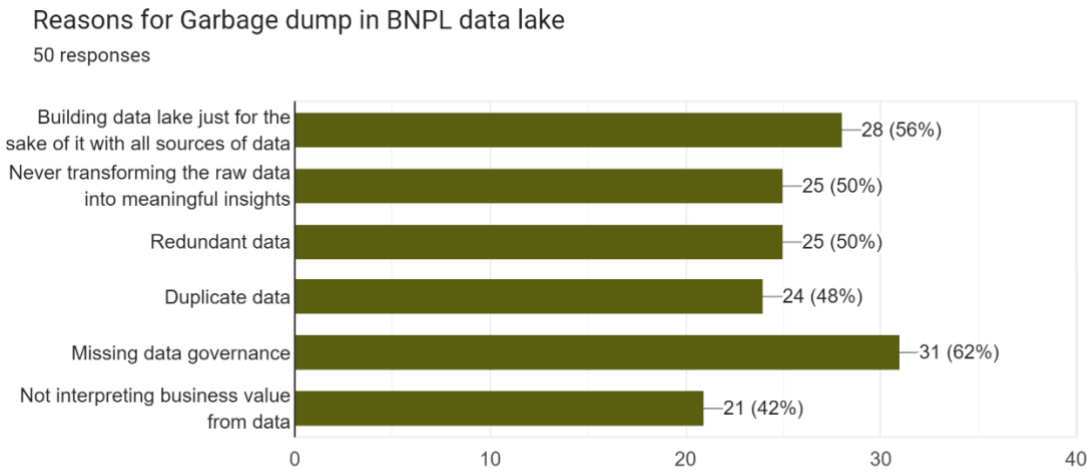


Figure 46: BNPL data lake – Garbage dump reasons

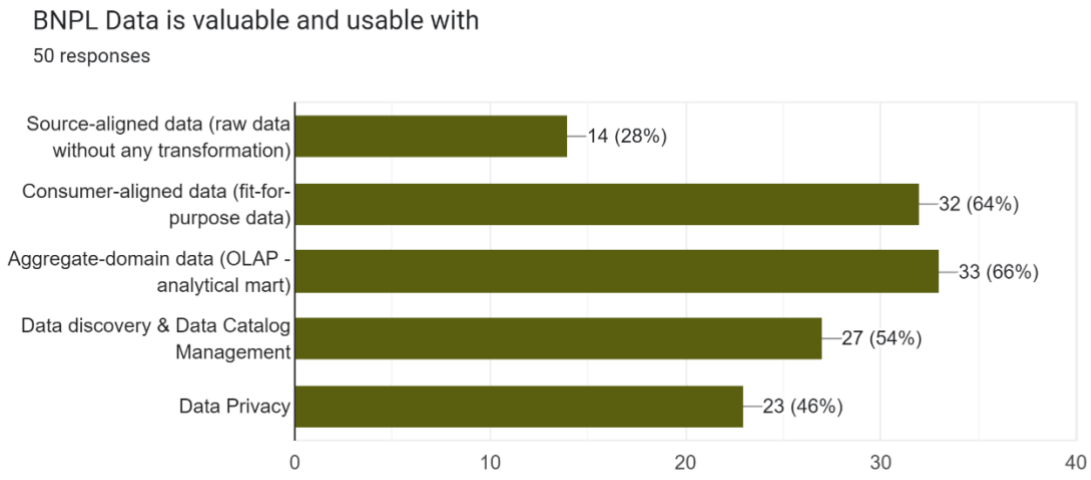


Figure 47: BNPL data lake – data asset

4.3. Results from Fintech – BNPL & Engineering Survey

4.3.1. Fintech – BNPL Data lake canvas

Based on the results, the Fintech – BNPL data lake canvas combines the factors based on the above evaluation. The canvas combines two zones, namely data enemies and data

divinity.

Unique Value Proposition

The USP of the Fintech – BNPL data lake canvas is to reap the ROI by identifying the data enemies to transform them into divine data to increase the Fintech – Data lake efficiency by 5-10X times with cost optimization by 30-40%, increase the Data Maturity Model (DMM) efficiency by 50%, generate Data as an Asset (DaaA) by 5X times.

Data lake canvas

Data lake canvas USP is driven by the 3 main pillars as in Figure 48 which bridges the gap between the Fintech-BNPL and data lake elements to identify the data enemies to transform to the data divine state with a set of drivers and enablers which are quantifiable.

Figure 49a is the realignment of the Fintech Data Lake architecture from the key studies discussed in Chapter II and the research responses.

The foundation of data lake canvas is driven by Data enemies and Data Divinity as in Figure 49b which helps to drive the data lake architecture.

Data enemies – This zone captures Fintech-BNPL drivers that includes business challenges, motivators and cost management factors. It also captures the data lake drivers that includes data engineering, architectural and AI / ML standards & practices.

Data Divinity –This zone ensures to attain the data divinity to guarantee the success of Fintech – BNPL data lake. When we achieve quantifiable business, technology, and techno-functional benefits through foundational drivers and enablers, we define it as the success of Fintech - BNPL data lake.

Unit of economies (for drivers & enablers):

The unit of economies in Figure 50 brings the drivers and enablers from the technology and Fintech business together with the unit of measurement as count and cost metric. This

helps to derive the expense, revenue, and net profit by establishing data lake. This becomes the foundation to enable the Data Divinity state of data lake. It can be measured at any frequency as required by the organization as monthly, quarterly, yearly or on ad-hoc basis. It does not adhere to the financial accounting standards however can be used to define the Key Performance Indicators (KPI) for Fintech – BNPL data lake.

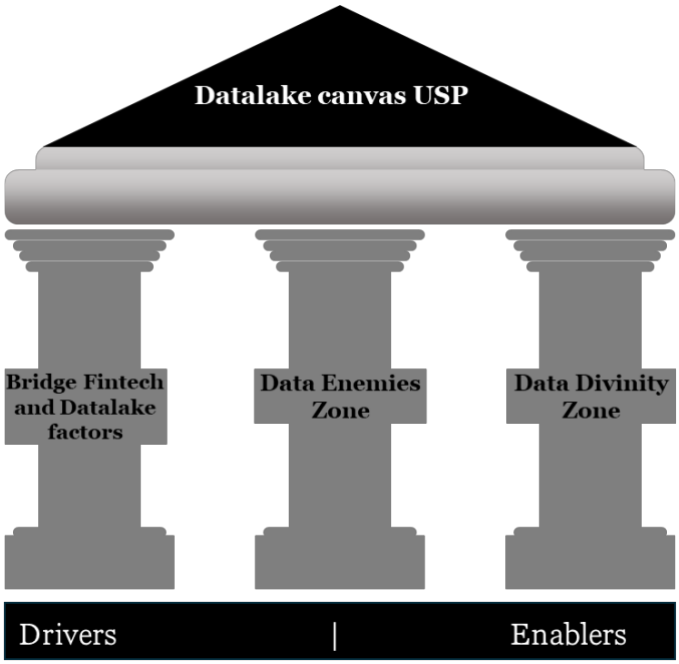


Figure 48: Fintech-BNPL data lake canvas pillars

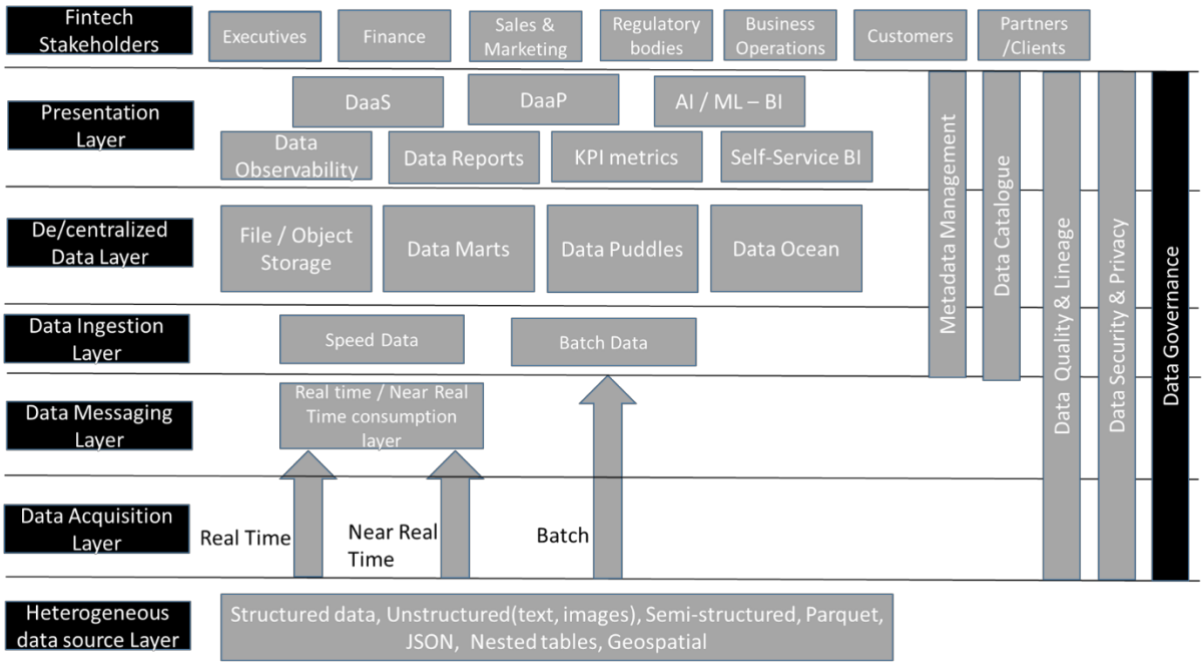


Figure 49a: Fintech – BNPL data lake architecture

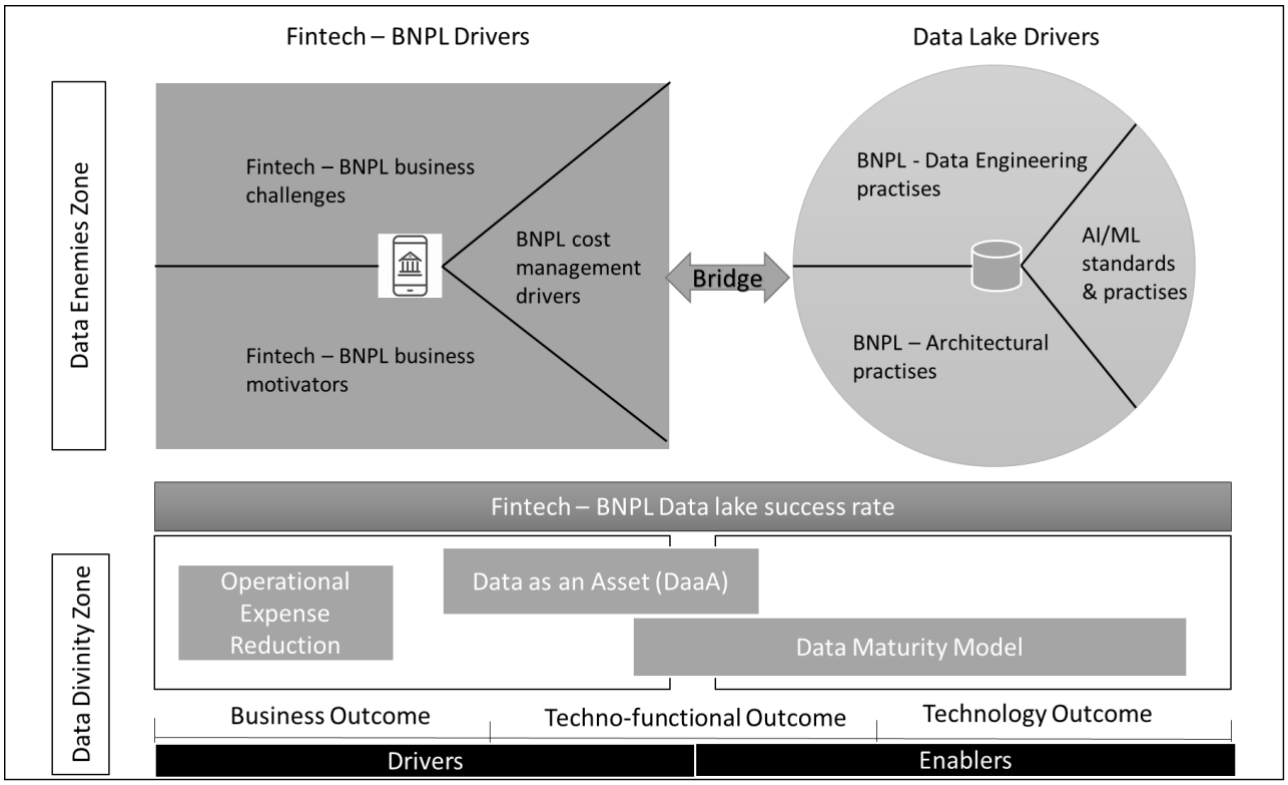


Figure 49b: Fintech - BNPL data lake canvas

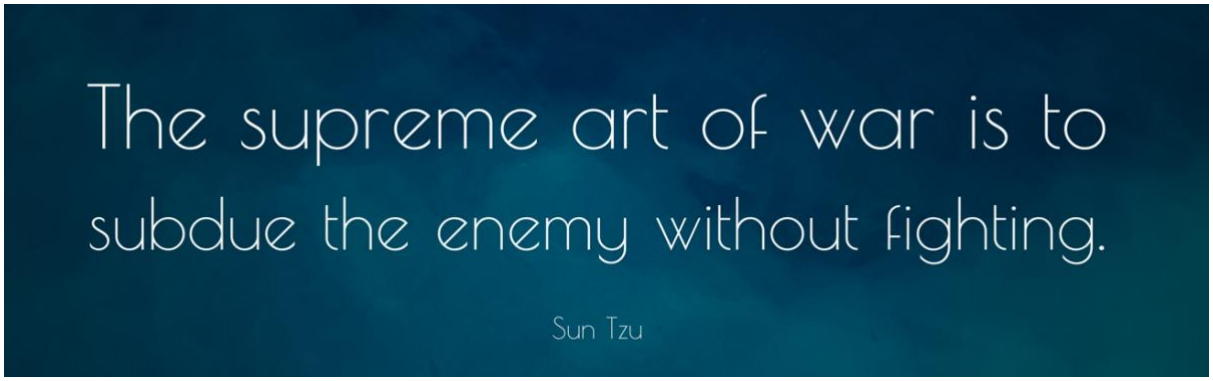
Unit of Economies					
Cost Drivers	**Unit**	**Count**	**Amount**		**Description**
Server	No.of servers and server cost	0	$	-	Defines the clusters, nodes, servers based on on-prem or cloud setup
Computing power	CPU and RAM power				Defines the RAM, CPU, cores used with on-prem or cloud setup
Storage	Storage cost	0	$	-	Defines the space utilization in DBMS - SQL / NoSQL, distributed file system
Database License Cost	Licensing cost	0			Defines the license cost either perpetual or subscription cost for database
BI / Data observability - License / Open-source cost	For e.g., **Tableau license cost**				Defines the license cost either perpetual or subscription cost for BI tool used
	Total tableau users & cost	0	$	-	
	Tableau Desktop	0	$	-	
	Tableau Explorer (Admin, Creator)	0	$	-	
	Tableau Viewer	0	$	-	
	For e.g., for open-source - Dremio / Superset				
	DCU used	0	$	-	
Data Quality	# of data rules, data quality issues / day, cost per defect	0	$	-	Defines the data quality rules set, onboarded, executed and its associated cost

Data Security	# of data security policies, data security issues / year, cost per defect	0	$	-	Defines the data security rules set as policies, Security Group, IAM. Defines the security rules onboarded, executed and its associated cost.
Infra setup cost			$	-	Defines the cost required to setup the application based on on-prem or cloud
Migration Cost			$	-	Defines the cost required to migrate the application from on-prem to cloud based on lift & shift or with modernization
Data Pipeline maintenance cost			$	-	Defines the cost required to maintain the application based on on-prem or cloud
Enablers					
Resources			$	-	Defines the people resources or others apart from above required
L&D			$	-	Defines the L&D resources either as online, offline and on-job trainings required
Total Cost					
Revenue Drivers					
Data as a Product	# of data products and revenue generated	0	$	-	Defines no. of data products and revenue generated
Data as a Service	# of data services and revenue generated	0	$	-	Defines no. of data services and revenue generated
In built tools	# of home-grown tools built and used, and revenue generated	0	$	-	Defines the home-grown tools built and used, and revenue generated

Risk Management	# of data services and revenue generated	$	-	Defines no. of data services and revenue generated for managing the risk management of BNPL
Fraud Protection	# of data services and revenue generated	$	-	Defines no. of data services and revenue generated for managing the Fraud Protection of BNPL
Compliance and Regulation	# of data services and revenue generated	$	-	Defines no. of data services and revenue generated for managing the Compliance and Regulation of BNPL
Customer Retention	# of data services and revenue generated	$	-	Defines no. of data services and revenue generated for managing the Customer Retention of BNPL
Customer Acquisition	# of data services and revenue generated	$	-	Defines no. of data services and revenue generated for managing the Customer Acquisition of BNPL
Strategic Drivers				
Time to Market	# of Strategic initiatives for deriving quick TTM for a business	$	-	Defines no. Strategic initiatives for deriving quick TTM for BNPL business
Time to Value	# of Strategic initiatives for deriving quick TTV for a business	$	-	Defines no. Strategic initiatives for deriving quick TTV for BNPL business
Total Revenue				
Net Profit				

Figure 50: BNPL data lake canvas – Unit of Economies

4.3.1.1. Data enemies



Data enemies zone comprises Fintech – BNPL drivers and data lake drivers. Most of the time in the organization, these drivers operate individually. For defining and deriving the successful strategy for the Fintech-BNPL data lake, there should be a mechanism for bridging them where business and data lake drivers should converge. Each of the blocks within Business and data lake drivers has many factors to consider for building a successful data lake.

Data enemies fall under two categories: the business drivers, which are challenges and problem areas, and data lake drivers, which must be set and follow the principles, standards, procedures, and processes.

Each of these factors is to consider an enemy to identify which one needs to be winning amongst multiple factors to be measured.

The strategy to identify the winner and stack rank the factors by grouping them in the right bucket. This strategy applies the Bing Fa technique to identify enemies per the Art of War by Sun Tzu. We can bucketize the factor into four categories per the Bing Fa technique as defined in Figure 49c.

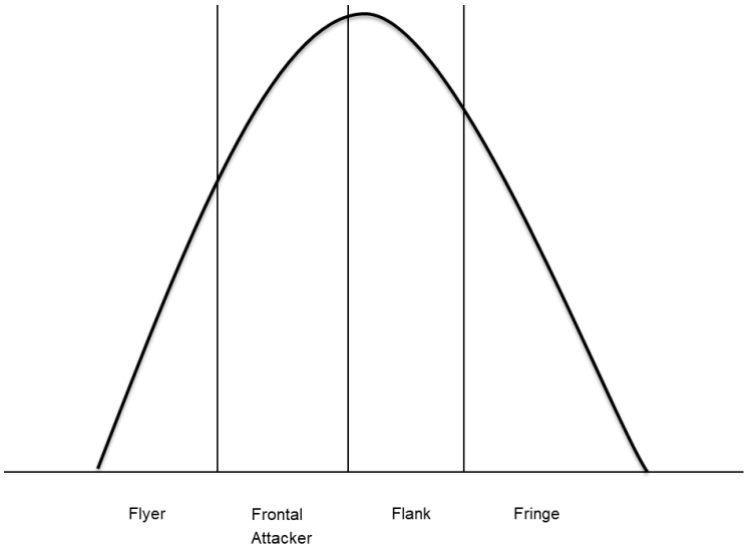


Figure 49c: Fintech - BNPL data enemies identification framework

1. Fringe → The factor that is not directly involved in a business's or data lake's primary objective but has a certain level of influence and can potentially impact the outcome.
2. Flank → The factor where potential threats or opportunities can arise
3. Frontal Attacker → The factor that is a competitor or an opposing force directly challenging the data lake goals or objectives.
4. Flyer → The factor which performs sudden or unexpected change or event significantly impacts the data lake's objectives.

4.3.1.1.1. Fintech – BNPL drivers

Fintech – BNPL business challenges

Business challenges are the key pain areas that need attention for run-the-business. Table 16 and Table 17 represents the data enemies of factors with the categorization based on the business challenges.

Table 16: Fintech challenges

Fintech challenges	**Enemy Bucket**
Compliance & Regulation	Flyer
Data Security	Frontal attacker
Customer Retention and acquisition Data and AI integration Competition Service personalization	Flank
Blockchain integration Market availability Sales and cash flow	Fringe

Table 17: BNPL Challenges

BNPL challenges	**Enemy Bucket**
Risk Management	Flyer
Fraud Protection	Frontal attacker
Customer Retention and acquisition Economic and geo-political situation	Flank
High-interest rate	Fringe

Fintech – BNPL business motivators

Table 18, Table 19 and Table 20 represents the data enemies' zone of business drivers that are the key pain areas that influence run-the-business.

Table 18: Risk Management

Risk Management	Enemy Bucket
Fraud Protection	Flyer
Risk Scoring	Frontal attacker
Missed payments Risk of debt spirals across multiple BNPL provides Interest rates Increased debt for Fintech companies	Flank
Forced lending Profitability of the business model	Fringe

Table 19: Customer journey

Customer journey	Enemy Bucket
Customer satisfaction	Flyer
Improved efficiency in managing returns	Frontal attacker
Customer loyalty Merchant brand exposure Customer referrals	Flank
Customer experience	Fringe

Table 20: Key subject areas

Key subject areas	Enemy Bucket
Payments & schedule	Flyer
Risk levels & category	Frontal attacker

Customer Scoring Product catalog of BNPL Limits	Flank
Financial calculation Master data and metadata Bureau data	Fringe

BNPL cost management drivers

Table 21, Table 22 and Table 23 represents the data enemies' zone of cost management drivers.

Table 21: BNPL cost increase factors

BNPL cost increase factors	**Enemy Bucket**
Risk management	Flyer
Fraud protection	Frontal attacker
Operational expense Customer retention & acquisition Current economical and geo-political situation	Flank
Technology & Tools Customer support Cost of funding	Fringe

Table 22: BNPL cost-effective factors

BNPL cost-effective factors	Enemy Bucket
Good Customer Acquisition & Retention strategy Availability of Technology & Tools	Flyer
Proper Risk Management Controlled Operational Expense (OpEx)	Frontal attacker
Favorable economic/geopolitical / Inflation factors Excellent Customer support	Flank
Goodwill gained from customers Sufficient funding and the right market Heavy late payment fees	Fringe

Table 23: BNPL cost measurement factors

BNPL cost measurement factors	Enemy Bucket
License cost for Tools & technology	Flyer
As provided by the cloud / on-prem	Frontal attacker
People resources	Flank
Full cost basis (direct + indirect costs) Provider cost + Operational / maintenance cost ROI Storage	Fringe

4.3.1.1.2. Data lake drivers

Data architectural practices

Table 24, Table 25 and Table 26 represents the data enemies' zone of data architectural practices.

Table 24: BNPL Data lake preference

Data lake preference	**Enemy Bucket**
Hybrid cloud data lake	Flyer
Private cloud data lake	Frontal attacker
Multi-cloud data lake Cloud data lake based on service offering (IaaS, PaaS, SaaS, etc.)	Flank
On-prem data lake Public data lake On-prem data warehouse	Fringe

Table 25: Preferred scalable BNPL architecture

Preferred scalable BNPL architecture	**Enemy Bucket**
Microservices architecture Data mesh architecture	Flyer
GFS / HDFS architecture Event-driven architecture	Frontal attacker
Data warehousing architecture Kappa architecture Dynamo architecture	Flank

Chubby architecture Columnar file systems	Fringe

Table 26: Preferred data model for BNPL data lake

Preferred data model for BNPL data lake	**Enemy Bucket**
Relational model	Flyer
Polyglot/multi-model	Frontal attacker
Document model Graph model	Flank
Snowflake model	Fringe

Data engineering practices

Table 27 represents the data enemies' zone of data engineering practices.

Table 27: Preferred data structure for BNPL data lake

Preferred data structure for BNPL data lake	**Enemy Bucket**
Offered by the distributed file system	Flyer
B trees Memtable	Frontal attacker
SST LSM	Flank
Reverted index B+ B+	Fringe

The other factors for data engineering practices include

a. Ease of sourcing, maintaining, and managing the structured & semi-structured

data

b. Ease of sourcing, maintaining, and managing the unstructured data

c. Importance of the underlying data structure

AI / ML standards & practices

Table 28 and Table 29 represents the data enemies' zone of AI / ML practices.

Table 28: Critical elements for effective AI/ML

Critical elements for effective AI/ML	**Enemy Bucket**
Data quality	Flyer
Tools & technology ML pipelines Data integration	Frontal attacker
Self-service business intelligence Data discovery	Flank
Feature store Expected sample and population of data	Fringe

Table 29: Features stores expected in BNPL data lake

Features stores expected in BNPL data lake	**Enemy Bucket**
Payments & transactions Customer 360	Flyer
Product & pricing Merchant 360	Frontal attacker
Finance	Flank

Risk, behavioral, and experimentation	Fringe

4.3.1.2. Data Divinity

When the data enemies are well managed, the data divinity zone is materialized. The data divinity zone is where the business, technology, and techno-functional benefits reap Return on Investment (ROI).

Data Divinity is achieved when the business problems are turned into solutions, and data lake drivers are set with the principles, standards, procedures, and processes. This state is achieved with cost optimization, Data Maturity Model, and 'Data as an Asset' factors at the expected levels aligned with the organization's goals.

4.3.1.2.1. Objective 1 - Cost Optimization using Fintech – BNPL Data lake canvas

Hypothesis test

Build two groups of data – traditional and canvas models- to perform Post-hoc analysis using the Kolmogorov-Smirnov test.

As mentioned in Chapter 3, below is the hypothesis we need to test for Operational expense cost.

H0 - Existing approach(es) leads to the same Operational Expense (OpEx) as compared to the proposed framework

Statically,

H0 – $\mu T\ f(x) = \mu F\ f(x)$

Where μT is the mean OpEx of the existing approach(es), and μF is the mean OpEx of the proposed framework being tested.

Alternative Hypothesis

H1 - $\mu T\ f(x) > \mu F\ f(x)$

The Kolmogorov-Smirnov (KS) test is a statistical test widely used to determine if a sample comes from a specific probability distribution or if two samples come from the same distribution. It is a non-parametric test, meaning it makes no assumptions about the underlying distribution of the data. The two-sample KS test is one of the most valuable and general non-parametric methods for comparing two samples, as it is sensitive to differences in the empirical cumulative distribution functions of the two samples.

Table 30 represents the current observation of operational cost from the traditional data lake methods derived from the survey responses. Table 31 represents the desired state of operational cost with the data lake canvas derived from the interviews with experts.

Table 30: Current Observation from the traditional data lake methods

bnpl_opex_cost	Current state
Increase by 5 - 10%	11%
Increase by 10 - 20%	5%
Increase by 20 - 40%	3%
Decrease by 5 - 10%	24%
Decrease by 10 - 20%	31%

Decrease by 20 - 40%	19%
Others	7%

Table 31: Expected state with data lake canvas – new responses

bnpl_opex_cost	Desire State
Increase by 5 - 10%	10%
Increase by 10 - 20%	2%
Increase by 20 - 40%	0%
Decrease by 5 - 10%	15%
Decrease by 10 - 20%	40%
Decrease by 20 - 40%	28%
Others	5%

Statistical Result

In our case, p-value = 0.00373, test statistic D = 0.32758, and Significant level α = 0.05.

p-value < α, proves our alternative hypothesis **H1 - μT *f*(x) > μF *f*(x)**

D > α, proves that two samples come from different distributions, which is also evident from the distribution of 2 samples in Figure 51.

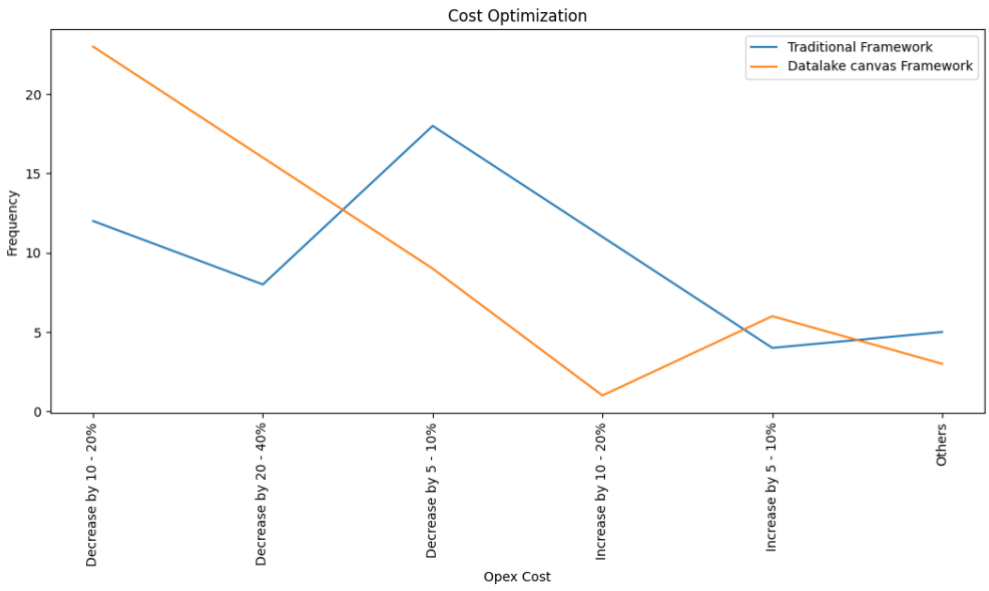


Figure 51: Operational Expense distribution

4.3.1.2.2. Objective 2 - Data Maturity Model using Fintech – BNPL Data lake canvas

Data Quality Index

Hypothesis test

H0 - Existing approach(es) leads to the same Data quality Index (DQI[T]) as compared to the proposed framework DQI[F]

Statically,

H0 – μDQI [T]$f(x)$ = μDQI[F] $f(x)$

Where μDQI [T] is the mean DQI of the existing approach(es), and μDQI [F]is the mean DQI of the proposed framework being tested.

Alternative Hypothesis

H1 - μDQI [T] $f(x)$ > μDQI [F] $f(x)$

Table 32 represents the current observation of data quality issues from the traditional data lake methods derived from the survey responses. Table 33 represents the desired state of data quality issues with the data lake canvas derived from the interviews with experts.

Table 32: Current DQI from the traditional data lake methods

Data Quality Issues	Current state
0-20%	28%
20-40%	26%
>40%	16%
Data quality is not measured though implemented	16%
20-40%; Data quality is not measured though implemented	6%
0-20%; Depends on the consumer	4%
0-20%; Data quality is not measured though implemented	2%
20-40%;>40%	2%

Table 33: Expected DQI with data lake canvas – new responses

Data Quality Issues	Desire State
0-20%	75%
20-40%	15%
>40%	10%

Statistical Result

In our case, p-value = 0.011, test statistic D = 0.32 and Significant level α = 0.05.

- p-value < α, proves our alternative hypothesis **H1 - µDQI [T] *f*(x) > µDQI [F] *f*(x)**
- D > α, proves that two samples come from different distributions which is also evident from the distribution of 2 samples in Figure 52.

Figure 52: Data quality distribution

- **Value of DQI[F] = 1** as we are rejecting the null hypothesis

Data Security Index

Hypothesis test

H0 - Existing approach(es) leads to the same Data security Index (DSI[T]) as compared

to the proposed framework DSI[F]

Statically,

H0 – μDSI [T]$f(x)$ = μDSI[F] $f(x)$

Where μDSI [T] is the mean DSI of the existing approach(es), and μDSI [F]is the mean DQI of the proposed framework being tested.

Alternative Hypothesis

H1 - μDSI [T] $f(x)$ > μDSI [F] $f(x)$

Table 34 represents the current observation of data security issues from the traditional data lake methods derived from the survey responses. Table 35 represents the desired state of data security issues with the data lake canvas derived from the interviews with experts.

Table 34: Current DSI from the traditional data lake methods

Data Security Issues	Current state
0-1%	11%
2-5%	13%
5-10%	39%
Data security is not measured though implemented	33%
Depends	2%
Not sure	2%

Table 35: Expected DSI with data lake canvas – new responses

Data Security Issues	Desire State
0-1%	30%
2-5%	40%
5-10%	20%
Data security is not measured though implemented	4%
Depends	2%
Not sure	4%

Statistical Result

In our case, p-value = 0.0001, test statistic D = 0.423 and Significant level α = 0.05.

- p-value < α, proves our alternative hypothesis **H1 - µDSI [T] *f*(x) > µDSI [F] *f*(x)**
- D > α, proves that two samples come from different distributions which is also evident from the distribution of 2 samples in Figure 53.

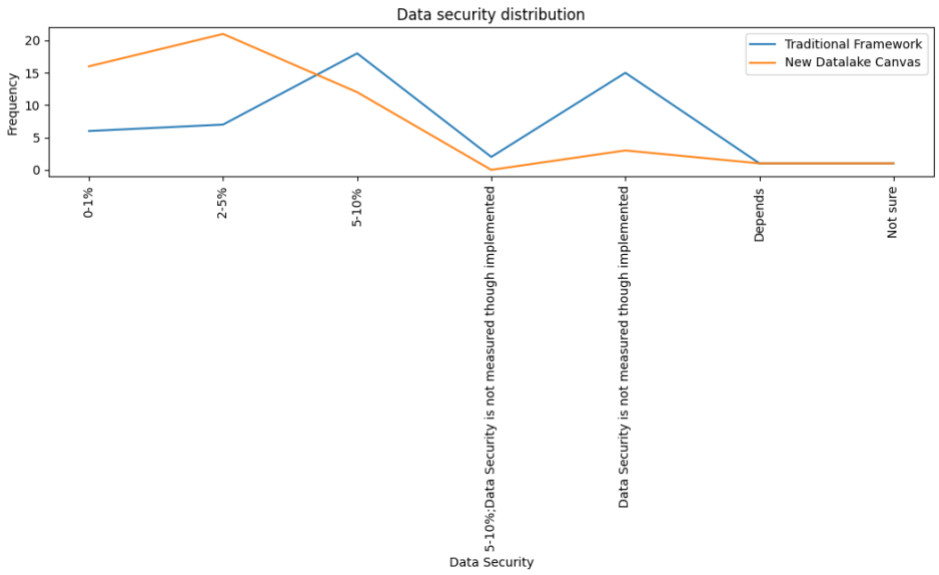


Figure 53: Data security distribution

- **Value of DSI[F] = 1** as we are rejecting the null hypothesis

Data Governance Index

Hypothesis test

H0 - Existing approach(es) leads to the same Data governance Index (DGI[T]) as compared to the proposed framework DGI[F]

Statically,

H0 – µDGI [T]*f*(x) = µDGI[F] *f*(x)

Where µDGI [T] is the mean DGI of the existing approach(es), and µDGI [F]is the mean

DQI of the proposed framework being tested.

Alternative Hypothesis

H1 - μDGI [T] $f(x)$ > μDGI [F] $f(x)$

Table 36 represents the current observation of data governance issues from the traditional data lake methods derived from the survey responses. Table 37 represents the desired state of data governance issues with the data lake canvas derived from the interviews with experts.

Table 36: Current DGI from the traditional data lake methods

Data governance Issues	Current state
Decrease by 10 - 20%	31%
Decrease by 20 - 40%	19%
Decrease by 5 - 10%	24%
Increase by 10 - 20%	5%
Increase by 20 - 40%	4%
Increase by 5 - 10%	10%
Others	7%

Table 37: Expected DGI with data lake canvas – new responses

Data governance Issues	Desire State
Decrease by 10 - 20%	10%
Decrease by 20 - 40%	10%
Decrease by 5 - 10%	40%
Increase by 10 - 20%	10%
Increase by 20 - 40%	10%
Increase by 5 - 10%	10%
Others	10%

Statistical Result

In our case, p-value = 0.009, test statistic D = 0.2966 and Significant level α = 0.05.

- p-value < α, proves our alternative hypothesis **H1 - μDGI [T] $f(x)$ > μDGI [F] $f(x)$**

- D > α, proves that two samples come from different distributions which is also evident from the distribution of 2 samples in Figure 54.

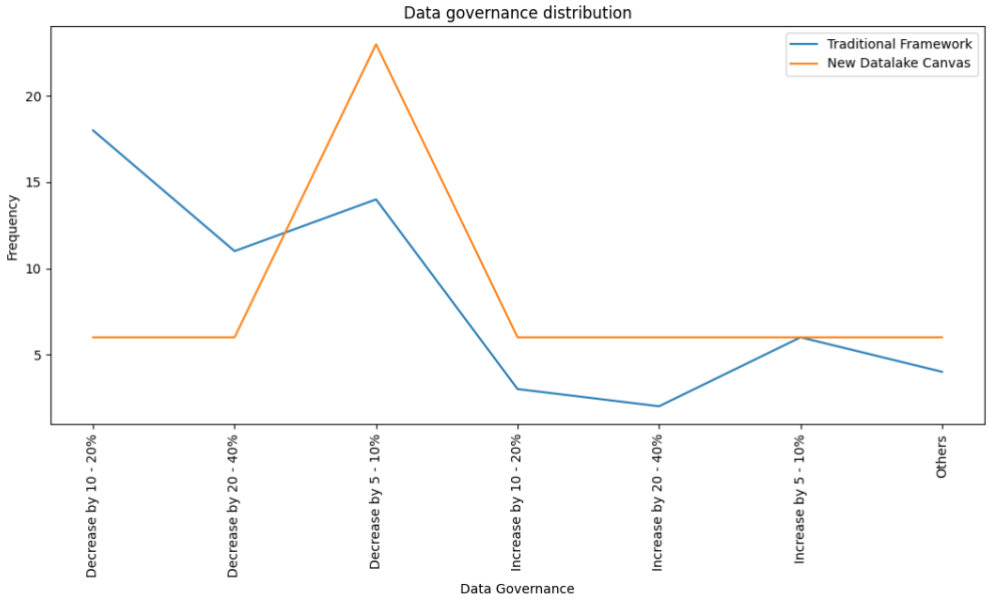


Figure 54: Data governance distribution

- **Value of DGI[F] = 1** as we are rejecting the null hypothesis

Time to Market Index

Scoring test for Lift & shift to cloud – TTM_1I

H0 - Existing approach(es) leads to the same Time to Market Index for Lift & Shift to cloud ($TTM_1I[T]$) as compared to the proposed framework TTM_1I [F]

Statically,

H0 – $TTM_1I[T] f(x) = TTM_1I[F]\ f(x)$

Where $TTM_1[T]$ is the index score of TTM_1I of the existing approach(es), and $TTM_1I[F]$is the index score of TTM of the proposed framework being tested.

The index score is based on the individual characteristic and their relative weightage where it is expected to be > 5

Alternative Hypothesis

H1 – $TTM_1I[T]\ f(x) < TTM_1I[F]\ f(x)$

Table 38 represents the current observation of Time to Market for migrating data lake to cloud without modernization as Lift & Shift (TTM_1I) from the traditional data lake methods derived from the survey responses. Table 39 represents the desired state of Time to Market for migrating data lake to cloud without modernization as Lift & Shift (TTM_1I) with the data lake canvas derived from the interviews with experts.

Table 38: Current TTM from the traditional data lake methods – TTM_1I

Time To Market (without modernization)	Scoring	Current state
Others	0	14%
Enterprise > 6 months	1	2%
Enterprise < 6 months	3	5%
Startup - < 1 months	6	19%
SMB < 2 months	8	21%
Enterprise < 3 months	10	39%

Table 39: Expected TTM with data lake canvas – new responses – TTM_1I

Time To Market (without modernization)	Scoring	Desire State
Startup - < 1 months	6	20%
SMB < 2 months	8	25%
Enterprise < 3 months	10	55%

Distribution comparison

- New distribution in Figure 55 shows that the scoring is > 5 hence the **value of $TTM_1I[F] = 1$** as we are rejecting the null hypothesis

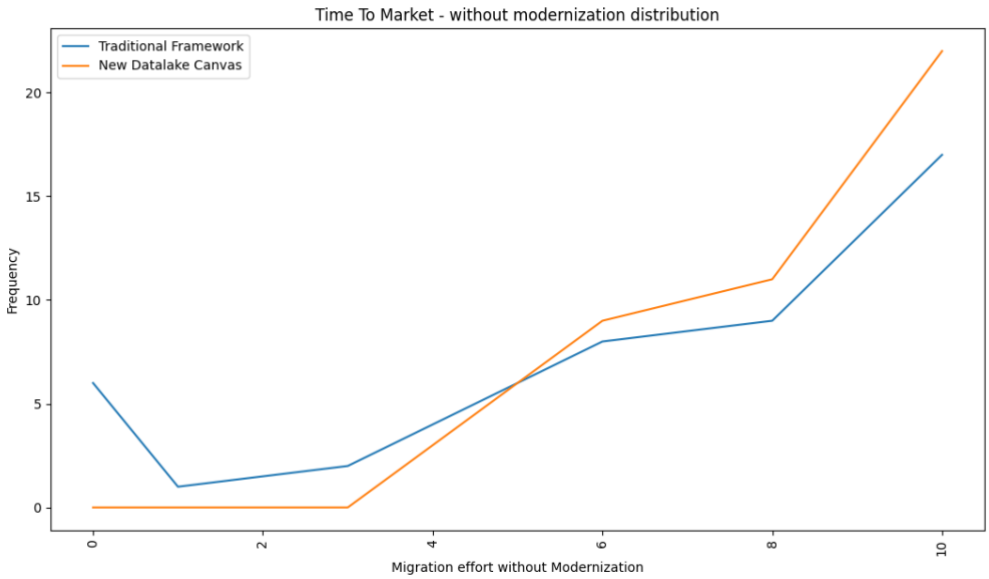


Figure 55: TTM – Migration without modernization distribution

Scoring test for Building data lake from scratch – TTM_2I

H0 - Existing approach(es) leads to the same Time to Market Index for building data lake from scratch (TTM_2I [T]) as compared to the proposed framework TTM_2I [F]

Statically,

H0 – $TTM_2I[T] f(x) = TTM_2I$ [F] $f(x)$

Where $TTM_2I[T]$ is the index score of TTM with the existing approach(es), and $TTM_2I[F]$is the index score of TTM with the proposed framework being tested.

The index score is based on their individual characteristic and their relative weightage where it is expected to be > 5

Alternative Hypothesis

H1 – $TTM_2I[T] f(x) < TTM_2I[F] f(x)$

Where the allowable index score is > 5

Table 40 represents the current observation of Time to Market for building data lake from scratch (TTM_2I) from the traditional data lake methods derived from the survey

responses. Table 41 represents the desired state of Time to Market for building data lake from scratch (TTM_2I) with the data lake canvas derived from the interviews with experts.

Table 40: Current TTM from the traditional data lake methods – TTM_2I

Time To Market (Build from Scratch)	Scoring	Current state
Startup - > 2 months	3	4%
SMB - 6 months	4	12%
Enterprise > 12 months	6	34%
Startup - < 2 months	7	18%
SMB - 3 - 4 months	8	14%
Enterprise > 6 months	10	18%

Table 41: Expected TTM with data lake canvas – new responses – TTM_2I

Time To Market (Build from Scratch)	Scoring	Desired state
Enterprise > 12 months	6	16%
Startup - < 2 months	7	24%
SMB - 3 - 4 months	8	28%
Enterprise > 6 months	10	30%

Distribution comparison

- New distribution in Figure 56 shows that the scoring is > 5 hence the **value of $TTM_2I[F] = 1$** as we are rejecting the null hypothesis

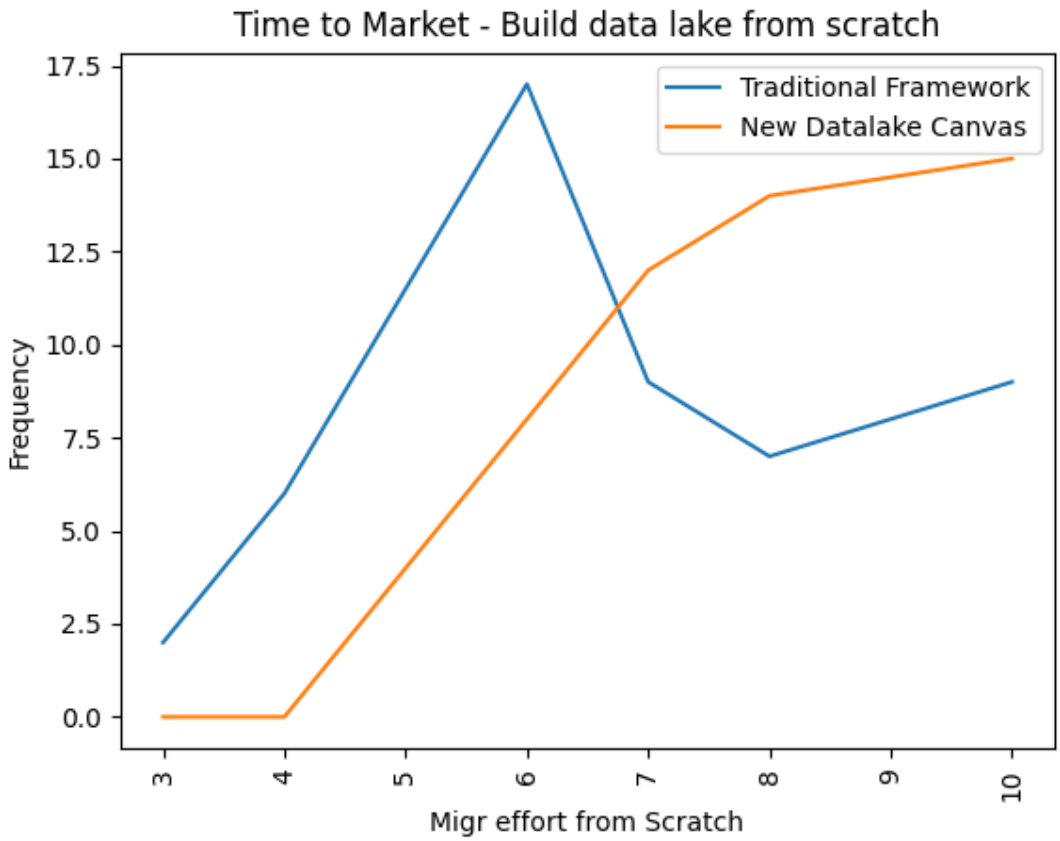


Figure 56: TTM – Build from scratch distribution

Data Maturity Model Index

Hypothesis test

H0 - Existing approach(es) leads to measuring the 'Data Maturity Model' (DMM) as compared to the proposed framework

H0 – DMM[T] > DMM[F]

$DMM[T] = DQI[T] + DSI[T] + DGI[T] + TTM_1I[T] + TTM_2I[T]$

Where DQI is the Data Quality Index score, DSI is the Data Security Index score, DGI is the Data Governance Index Score, and TTMI is the Time To Market Index score from the

traditional framework [T] and proposed framework [F]

DQI[T] = 1 if the null hypothesis is to accept and 0 if the null hypothesis is to reject, i.e., DQI[F] = 1

DSI[T] = 1 if the null hypothesis is to accept and 0 if the null hypothesis is to reject, i.e., DSI[F] = 1

DGI[T] = 1 if the null hypothesis is to accept and 0 if the null hypothesis is to reject, i.e., DGI[F] = 1

$TTM_1I[T]$ = 1 if the null hypothesis is to accept and 0 if the null hypothesis is to reject, i.e., $TTM_1I[F]$ = 1

$TTM_2I[T]$ = 1 if the null hypothesis is to accept and 0 if the null hypothesis is to reject, i.e., $TTM_2I[F]$ = 1

Alternative Hypothesis

H1 - DMM[T] < DMM[F]

Assumption: Scales and weights are equal

Statistical Result of Data Maturity Model

We know that with the above hypothesis testing in Sec 7.2.3.2.1, 7.2.3.2.2, 7.2.3.2.3, 7.2.3.2.4

DMM[T] < DMM[F] as

DMM[T] = 0
DMM[F] = 1 + 1 + 1 + 1 + 1 = 5

4.3.1.2.3. Objective 3 - Data as an Asset (DaaA) using Fintech – BNPL Data lake canvas

Hypothesis test

H0 - Traditional approach(es) leads to measuring ‘Data as an Asset’ (DaaA) as compared to the proposed framework

H0 – DaaA[T] > DaaA[F]

DaaA[T] = DaaP[T] + DaaS[T]

Where DaaP is Data as a Product, and DaaS is Data as a Service from the traditional framework [T] and proposed framework [F]

Alternative Hypothesis

H1 - DaaA[T] < DaaA[F]

Result of DaaA

DaaA cannot be proved with the survey responses as they are categorical and cannot be quantified. However, they can be used for making inferences that cannot be essentially used for the hypothesis.

To prove the hypothesis, it has to be measured against the unit and cost factors; hence, it can be evaluated with the ‘Unit of economies’ framework built as part of the data lake canvas.

Comprehending the sensitivity of the information with the ‘Unit of economies’, this couldn’t be collected for any Fintech. However, to evaluate the framework, this is validated with non-Fintech company data lake.

Results in the Table for ‘Unit of economies’ are collated from a non-Fintech company's ‘Head of Analytics.’

Table 42: Unit of Economies – Non-Fintech company

Unit of Economies						
Cost Drivers	**Unit**	**Count**		**Amount**	**Comments**	**Description**
Server	No.of servers and server cost	0	$	70,00,000.00	This is Azure cloud not sure count will be applicable	Defines the clusters, nodes, servers based on on-prem or cloud setup
Computing power	CPU and RAM power	0	$	-		Defines the RAM, CPU, cores used with on-prem or cloud setup
Storage	Storage cost	0	$	3,00,000.00		Defines the space utilization in DBMS - SQL / NoSQL, distributed file system
Database License Cost	Licensing cost	70	$	90,00,000.00		Defines the license cost either perpetual or subscription cost for database
BI / Data obervability - License / Open source cost	Total Power BI users & cost	30	$	12,000.00	30 users	Defines the license cost either perpetual or subscription cost for BI tool used
	Desktop BI	50	$	5,000.00	50 users	
	Power BI Explorer (Admin, Creator)	0	$	1,200.00		

	Power BI Viewer	20000	$	-	This is part of office 365 for entire organization hence cost not captured under here	
	Tableau Viewer	0	$	-		
	For eg, for open-source - Dremio / Superset				We don't use any of the open-source solution for security reasons	
	DCU used	0	$	-		
Data Quality	# of data rules, data quality issues / day, cost per defect	1200	$	3,66,240.00	280 euro per defect approximately each product suite approximately 100 defects / month would need fixing	Defines the data quality rules set, onboarded, executed and its associated cost
Data Security	# of data security policies, data security issues / year, cost per defect	0	$	-	Unknown Not measured	Defines the data security rules set as policies, Security Group, IAM. Defines the security rules onboarded, executed and its associated cost.

Infra setup cost		0	$ -	Unknown Not measured	Defines the cost required to setup the application based on on-prem or cloud
Migration Cost		0	$ -	Unknown Not measured	Defines the cost required to migrate the application from on-prem to cloud based on lift & shift or with modernization
Data Pipeline maintenance cost		0	$ -	Unknown Not measured	Defines the cost required to maintain the application based on on-prem or cloud
Enablers					
Resources		0	$15,000		Defines the people resources or others apart from above required
L&D		0	$10,000	This includes HR, IT support and other staff	Defines the L&D resources either as online, offline and on-job trainings required
Total Cost			$ 1,66,72,440.00		
Revenue Drivers					
Data as a Product	# of data products and revenue generated	7	$ 5,00,00,000.00	7 Data product suite	Defines no. of data products and revenue generated
Data as a Service	# of data services and	0	$ -	Not used by this team though used	Defines no. of data services and revenue generated

	revenue generated				in org by other teams	
In built tools	# of home-grown tools built and used, and revenue generated	0	$	-		Defines the home grown tools built and used, and revenue generated
Risk Management	# of data services and revenue generated	0	$	-		Defines no. of data services and revenue generated for managing the risk management of BNPL
Fraud Protection	# of data services and revenue generated	0	$	-		Defines no. of data services and revenue generated for managing the Fraud Protection of BNPL
Compliance and Regulation	# of data services and revenue generated	0	$	-		Defines no. of data services and revenue generated for managing the Compliance and Regulation of BNPL
Customer Retention	# of data services and revenue generated	7	$	17,70,00,000.00	7 Data product suite	Defines no. of data services and revenue generated for managing the Customer Retention of BNPL
Customer Acquisition	# of data services and	7	$	7,70,00,000.00	7 Data product suite	Defines no. of data services and revenue generated for managing

	revenue generated		the Customer Acquisition of BNPL
Strategic Drivers			
Time to Market	# of Strategic initatives for deriving quick TTM for a business	$ -	Defines no. Strategic initatives for deriving quick TTM for BNPL business
Time to Value	# of Strategic initatives for deriving quick TTV for a business	$ -	Defines no. Strategic initatives for deriving quick TTV for BNPL business
Total Revenue		$ 30,40,00,000.00	
Net Profit		$ 28,73,27,560.00	

Results show that the unit of economies helps to measure all the drivers and enablers of data lake, which are measurable with unit and cost metrics. For the non-Fintech company measures, it shows that the unit and cost per year for DaaP and DaaS shows that with the traditional framework, it wasn't measured though the data was available. However, with data lake canvas, the factor on unit and cost metric could be put together to measure across; hence the below statistical result.

DaaA[T] = DaaP[T] + DaaS[T] = 0 + 0 as they were not measured in the traditional framework

DaaA[F] = DaaP[F] + DaaS[F] = 7 + 0 (there are no DaaS being used)

DaaA[T] < DaaA[F] hence rejecting the null hypothesis.

Unit of Economies

Additionally, the unit of economies from the table 31 proves that the profit is due to the data lake setup. It shows that with the traditional framework, the drivers and enablers were not measured though they are available in different forms, and some aren't. 'Unit of economies' infers the below

1. Cost drivers are being measured
    a. Licensing cost and Data quality cost metrics are available
    b. Data security metrics are not measured in the organization
    c. Infra setup, migration and maintenance cost are not being measured
2. Enablers are being measured
3. Revenue metrics are being measured
    a. DaaP is available and measured with revenue
    b. DaaS is not available though it is being used by the other department

c. Total revenue is measured

4. Strategic driver metrics are not available
5. There is a net profit with the cost drivers (with data lake) and revenue drivers (with Business motivators)

The framework has helped them identify, categorize and measure them with the suitable unit of measure to derive the cost metrics as expense, revenue, and profit though it is not expected to be as per the financial accounting standards. It also helps to comprehend what they are missing to setup, missing to measure, and focus areas to improve with data lake and business.

4.3.1.2.4. Conclusion with Data Divinity

Data divinity factor = Business outcome + Techno-functional outcome + Technology outcome = Operation Expense reduction + DaaA + DMM

In this case, hypothesis testing with the data lake canvas has been proved and below is the data divinity factor

Data divinity factor = 1 + 7 (No. of DaaA) + 5 (DMM from the above hypothesis)

= 13

Where 1 is the factor from the hypothesis testing proven for Opex with data lake canvas, 5 is the factor from DMM from DQI, DSI, DGI, TTM_1I and TTM_2I, 7 is the factor from the 'Unit of economies' for the calculation of DaaA with DaaP and DaaS.

More the data divinity factor better the data lake built. Upper limit for the data divinity factor is as defined by the business, techno-functional and technology outcome. There can be more factors to consider for each of the outcome based on the criticality factors defined by organization. Business outcome can include other cost drivers apart from

OpEx, techno-functional outcome can include more factors into DaaA and others, and technology factors can include more factors into DMM.

4.4. Chapter Summary

This chapter presented the results of the 3 key strategies used to assess the performance of the data lake canvas using the operational cost (OpEx), Data Quality Index (DQI), Data Security Index (DSI), Data Governance Index (DGI), Time To Market (TTM), Data as a Product (DaaP), Data as a Service (DaaS).

The result indicates that the OpEx can be majorly decreased by 10 - 20% or decreased by 20 - 40% based on the focussed discussions with industry Fintech and engineering experts and can be used as a benchmark return for further comparisons.

The result indicates that the DMM can be measured with DQI, DSI, DGI, TTM_1I, and TTM_2I, which can be used as a benchmark for further comparisons. Results also indicate that

- Data quality issues can be majorly brought within 0-20%,
- Data security issues can be brought majorly within 0-1% & 2-5%,
- Data governance issues can be decreased majorly by 5 - 10%, T
- Time to market for lift-shift can be done for Startup - < 1 month, SMB < 2 months, Enterprise < 3 months
- Time to market for building data lake from scratch can be done for Startup - < 2 months, SMB - 3 - 4 months, Enterprise > 6 months

The result indicates that the DaaA can be measured with DaaP and DaaS with the unit of economies framework using the drivers and enablers.

The various factors analyzed in this chapter provided varying cost and index scoring results, which we shall discuss in the next chapter.

CHAPTER V:

DISCUSSION

Chapter 5 presents a comprehensive discussion and conclusion of the research findingspresented in the previous chapters. It begins by summarizing the results and dives straight into their in-depth interpretation. Furthermore, the chapter presents the implications of the research findings for industry experts and practitioners in the Fintech – Buy Now Pay Later data lake. Finally, the chapter provides recommendations for future research in the area.

5.1. Discussions of findings from the questionnaire

In this study, the questionnaires are from two sets of professionals: the Fintech – BNPL experts and the Data engineering experts. The aim is to bridge the gap between the business and data lake technology, which has not been previously achieved. The results of the responses are for building the data lake canvas and validating the hypothesis and the unit of economies for drivers and enablers. Also, for validating the hypothesis the selected questions from the questionnaire are re-evaluated with the respondents to measure against the data lake canvas for cost optimization and DMM.

As the skillset required for the questionnaire is a niche, target responses are 100 and 50 from each group of experts, namely the Fintech-BNPL and data engineering experts.

5.1.1. Fintech - BNPL

The responses for the Fintech – BNPL questionnaire are the base to build the Fintech-BNPL drivers for the data lake canvas and are the responses based on the traditional framework. In addition, 2 of the quantifiable responses on operational cost and garbage dump are used for validating the hypothesis on cost optimization and data governance, respectively. The collection of responses for these two questions is against the data lake

canvas for the hypothesis testing.

5.1.2. Fintech and BNPL Business – Survey insights

Identification and prioritization of the Fintech and BNPL business challenges and motivators in the data enemies' zone is identified as per Sun Tzu's art of war. (Sridhar, 2022)

5.1.2.1. Fintech challenges

1. The results show that Data security is a critical concern for Fintech companies as they deal with sensitive financial and personal information. Hence the data lake needs to address the following.

    a. User encryption: Encryption converts data into a code to prevent unauthorized access. Fintech companies should use encryption to protect sensitive data in transit and at rest in the data lake.

    b. Implement multi-factor authentication: multi-factor authentication requires users to provide multiple forms of identification to access their accounts. It can help prevent unauthorized access to sensitive data.

    c. Regularly update software: Fintech companies should update their software to address vulnerabilities and fix security issues.

    d. Conduct regular security audits: Regular security audits can help Fintech companies identify potential security risks and vulnerabilities in their systems.

    e. Implement access controls: Access controls can limit who has access to sensitive data within the data lake. Fintech companies should implement access controls to ensure that only authorized personnel can access

sensitive data.

f. Have a response plan for security incidents: In the event of a security incident, the data lake team should have a response plan to minimize the impact and prevent further damage.

g. Overall, data security is crucial for Fintech data lake to build trust with their customers and protect their sensitive information. By implementing these best practices, Fintech data lake can help ensure their systems are secure and their customers' data is protected.

Implementation of security measures in Fintech companies is via encryption, access controls, data governance policies, compliance with data protection laws, and cloud computing technologies. (Rehman et al., 2023)

2. The results show that compliance and regulation are essential components of Fintech. Fintech data lake must manage financial transactions and sensitive personal information and comply with various regulations to ensure they operate legally and securely along with data security. Some of the key Compliance and regulatory considerations in the Fintech data lake:

    a. Know your regulatory environment: Fintech data lake must understand the regulatory environment in which they are set-up.

    b. Maintain data privacy and security: Fintech data lake must comply with data privacy regulations to protect the sensitive personal information of their customers. It includes implementing data security measures and giving customers transparency and control over their data.

c. Stay up-to-date on regulatory changes: Fintech data lake should stay informed of regulatory changes and adapt its compliance programs accordingly. It includes monitoring changes in laws and regulations and industry best practices.

d. Overall, Compliance and regulation are critical considerations for Fintech companies. Fintech data lake can build trust within the organization and ensure they operate legally and securely by complying with regulatory requirements and implementing robust compliance programs.

Financial compliance and regulation are crucial for the innovation and success of Fintech, with unprecedented opportunities with regulation. It includes regulations using big data and analytics for reporting, stimulating a new generation of "RegTech" companies. (Bu et al., 2022)

5.1.2.2. BNPL challenges

1. The results show that BNPL carries risks that need to be managed. Some of the key risk management considerations for the BNPL data lake:

    a. Credit risk: BNPL data lake to help evaluate the credit risk, which is the risk of customers defaulting on their payments. To manage this risk, the BNPL data lake should have credit assessment data and appropriate credit scoring models to assess customers' creditworthiness.

    b. Fraud risk: BNPL data lake to help evaluate fraud risk, which is the risk of customers engaging in fraudulent activities such as identity theft and chargebacks. To manage this risk, the BNPL data lake to have robust fraud detection and prevention measures, such as multi-factor authentication and

transaction monitoring.

c. Operational risk: BNPL data lake help to evaluate operational risk - the risk of disruptions to business operations due to internal or external factors. To manage this risk, the BNPL data lake must have a business continuity plan to ensure it can continue operating during an interruption, such as a cyberattack or natural disaster.

d. Compliance risk: BNPL data lake to help evaluate compliance risk, which is the risk of non-compliance with laws, regulations, and industry standards. To manage this risk, the BNPL data lake to evaluate applicable laws and regulations and adhere to its data, such as data privacy, consumer protection, and anti-money laundering (AML).

e. Reputation risk: BNPL data lake to help evaluate reputation risk, which is the risk of damage to their reputation due to negative publicity or customer complaints. To manage this risk, the BNPL data lake enables customer service and transparency; to ensure the data is available with low latency to resolve customer complaints and inquiries.

f. Overall, BNPL data lake needs to implement robust risk management measures to ensure they can operate safely and sustainably in a highly competitive and rapidly evolving market. By managing credit, fraud, operational, Compliance, and reputation risks effectively, the BNPL data lake can build data divinity within the data.

Studies show that BNPL as point-of-sale loans post the threat of point-of-fail causing substantial dangers for consumers. SBPC investigation on student loans identified a

range of unaccredited and dubious schools that market point-of-sale financing, such as BNPL credit, as a variety of student loans, particularly institutions in the tech-focused “bootcamp” space. (Center, 2022)

2. The results show that BNPL carries the risk of fraudulent activities such as identity theft and chargebacks. Key fraud protection measures that BNPL data lake should consider implementing

    a. Multi-factor authentication: BNPL data lake should implement multi-factor authentication to verify the identity and prevent unauthorized account access.

    b. Transaction monitoring: BNPL data lake should monitor transactions for suspicious activity, such as high-value purchases, unusual transaction patterns, and purchases from high-risk locations. It can help detect and prevent fraudulent transactions before they occur.

    c. Data privacy and security: BNPL data lake should implement robust data privacy and security measures to protect the sensitive personal information of their customers. It includes encrypting data in transit and at rest, restricting sensitive data access, and monitoring security breaches.

    d. Overall, the BNPL data lake must take a proactive and holistic approach to fraud protection to ensure it can operate safely and sustainably in a highly competitive and rapidly evolving market. By implementing multi-factor authentication, transaction monitoring, and data privacy and security measures, the BNPL data lake can be robust.

BNPL consumers choose credit cards as the primary instrument, and there are loopholes

in the loan application system that the fraud imposters piggyback on. Hence there should be a better online BNPL solution to avoid fraudulent transactions. (Xing et al., 2019)

5.1.2.3. BNPL Risk management

1. The results show that Fraud protection is critical to Buy Now, Pay Later (BNPL) risk management. Fraud protection as part of the risk management in the BNPL data lake is explained under the BNPL challenges section.

2. The results show that Risk scoring is critical to Buy Now, Pay Later (BNPL) risk management. Risk scoring involves using data analytics and statistical models to assess customers' creditworthiness and determine the risk level associated with providing them with credit. Key considerations for risk scoring in the BNPL data lake:

    a. Data sources: BNPL data lake should collect data from various sources, including credit bureaus, social media, and other publicly available data. It can help create a more comprehensive picture of a customer's creditworthiness and financial stability.

    b. Machine learning models: BNPL data lake should build feature stores to use machine learning models to analyze data and identify patterns that can predict creditworthiness and assess risk. These models can be trained on historical data to improve accuracy over time.

    c. Multiple factors: BNPL data lake should bring multiple factors into the feature store when scoring risk, including credit history, income, employment status, and payment history. The BNPL data lake can create a more accurate and comprehensive customer risk profile by considering

multiple factors.

d. Real-time scoring: BNPL data lake should enable to score risk in real-time as customers apply for credit. It can help BNPL providers make more informed decisions about whether to extend credit to customers and on what terms.

e. Continuous monitoring: BNPL data lake should monitor customers' creditworthiness and adjust risk scores as necessary. It can help identify customers at increased risk of default and allow BNPL providers to take appropriate action, such as adjusting credit limits or offering alternative payment plans.

f. In summary, Risk scoring is a critical component of BNPL data lake - risk management. Using data analytics and statistical models to assess creditworthiness and determine risk, BNPL providers can make more informed decisions about extending credit and managing credit risk using the data lake. By collecting data from various sources, using machine learning models, considering multiple factors, scoring risk in real-time, and continuously monitoring creditworthiness, the BNPL data lake can improve accuracy, reduce risk, and build trustable data.

According to survey data from the United States, approximately one-third of BNPL users have experienced missed payments, resulting in a subsequent decline in their credit scores. However, it is important to note that this information does not establish a direct cause-and-effect relationship between BNPL usage and credit score changes. (Guttman-Kenney et al., 2022). Hence scoring model in the data lake needs to be robust.

5.1.2.4. BNPL Customer Journey

1. The results show that a customer journey is critical in BNPL (Buy Now Pay Later) data lake to capture the customer's various touchpoints or interactions with the BNPL service provider throughout their buying and payment process.

    a. In a BNPL data lake, customer journey data can be collected from various sources, such as e-commerce platforms, mobile apps, and payment gateways. This data can include the customer's browsing behavior, product selection, payment preferences, payment history, and customer service interactions.

    b. By analyzing the customer journey data, the BNPL data lake can gain insights into the customer's preferences, pain points, and behavior patterns. This information can improve the overall customer experience and better tailor marketing campaigns to target specific customer segments.

    c. For example, analyzing customer journey data might reveal that a particular segment of customers tends to abandon their purchase at the payment stage. By identifying this issue, the BNPL data lake can focus on improving its payment process or offering more flexible payment options to reduce the likelihood of cart abandonment.

Big data and machine learning (ML) algorithms significantly drive various fintech innovations that revolve around the consumer journey. Employing machine prediction to determine loan eligibility enables investors to achieve higher return rates and provides borrowers with limited alternative funding options with increased opportunities for

securing funds. (Fu et al., 2021)

BNPL companies assert that they are more accountable than credit cards due to their inclusive nature and fairer loan terms. They attribute these claims to the platform-based structure of BNPL products. However, it is worth noting that these companies define responsible consumers as those who consistently meet their repayment obligations. This redefinition of responsible consumption promotes higher spending and establishes the use of BNPL credit as a normalized practice for such consumption. (Aalders, 2023)

2. The results show that customer satisfaction is critical for the BNPL data lake; where the key considerations for ensuring customer satisfaction in the BNPL data lake are:

    a. Transparency: BNPL data lake should have data about BNPL fees, payment terms, and any penalties for late payments. It can help to build models on customers' data, understand their obligations and avoid unexpected fees or charges, ensuring data transparency for its customers.

    b. Security: BNPL data lake should implement robust security measures to protect customers' personal and financial information. It includes using encryption, restricting access to sensitive data, and monitoring for security breaches.

    c. Overall Customer satisfaction in a BNPL data lake involves collecting, analyzing, and leveraging customer data to optimize the customer experience and drive business growth.

Customer satisfaction is driven by the psychological determinants of four dimensions, temporal, spatial, probabilistic, and social, and indicates how close consumers feel.

These dimensions drive transparency and security for customer satisfaction. (Relja et al., 2023)

3. By focusing on these key considerations, BNPL data lake can improve customer satisfaction, build loyalty, and differentiate itself from competitors. Ultimately, the success of the BNPL data lake depends on its ability to meet the needs and expectations of its customers.

4. Ways to improve efficiency in managing returns in the BNPL data lake:

    a. Returns data: BNPL data lake should have the data on the returns of the products, its timelines, conditions & restrictions on returns, and returns status.

    b. Regular Evaluation and Improvement: BNPL data lake should enable to build models to track return rates and identify any areas of the return process that may be causing delays or inefficiencies.

5. By implementing these strategies, merchants can improve the efficiency of managing returns in BNPL, leading to a better customer experience and reduced costs associated with returns.

5.1.2.5. BNPL Subject areas

1. The payment and schedule in the BNPL data lake may vary depending on the provider and the specific terms of the agreement. Some of the general aspects to consider:

    a. Payment Amounts: BNPL data lake should have various payment options, including equal payments, variable payments, or balloon payments (one large

final payment). Typically, the payments are spread out over a few weeks or months.

b. Payment Schedule: BNPL data lake should have a payment schedule that will depend on the agreement between the consumer and the BNPL provider, which can be bi-weekly or monthly.

c. Late Payment Fees: Late payment fees are a common feature of BNPL agreements which should be in the BNPL data lake. It should cover the aspects that if a consumer misses a payment, they may charge a fee, and their credit score may impact.

d. Interest Rates: BNPL data lake should have interest charges on the outstanding balance that may be fixed or variable, which can vary depending on the provider and the consumer's credit score.

e. Duration of Payments: BNPL data lake should capture the length of the BNPL agreement can vary, but typically, it ranges from a few weeks to several months. Consumers should be aware of the length of the agreement, as it will impact the total amount paid and the interest charged.

f. In summary, the BNPL data lake should ensure to capture the payment obligations and any potential fees or penalties.

BNPL programs offer various repayment options, including different cycles, amounts, and approaches to handling overdue payments. However, this diversity can lead to consumer uncertainty. BNPL providers often present themselves as interest- and fee-free platforms, which exempts them from conducting comprehensive affordability assessments (Relja et al., 2023). Hence, they are incredibly critical to have them in the

data lake.

2. BNPL (Buy Now Pay Later) services can categorize into various risk levels based on the consumer's likelihood of default or non-payment. Some considerations on categories based on risk level in the BNPL data lake:

    a. Low Risk: For low-risk levels, the BNPL data lake should have data on retailers and good credit scores, offering lower interest rates and more favorable payment terms.

    b. Moderate Risk: For moderate risk levels, the BNPL data lake should have data on a broader range of retailers and more lenient credit assessments offering higher interest rates and penalties for missed payments.

    c. High Risk: For high risk, the BNPL data lake should have data on riskier retailers and consumers with lower credit scores offering very high-interest rates and more aggressive debt collection practices.

    d. In summary, the BNPL data lake may also have to build different risk models that assess the likelihood of default on a case-by-case basis. These models may consider factors such as the consumer's credit score, income, employment status, and other financial obligations.

The large size of small and medium-sized enterprises (SMEs) presents challenges to obtaining credit risk assessments having a direct impact on cash flow. The greater scale and size of business customers then individual consumers pose increased risks, which can significantly impact on consumer reputation. Therefore, BNPL aims to mitigate these businesses risks by leveraging AI-enabled technology for instant credit and fraud checks, ensuring seamless payment processes in the data lake. (Alshahri, 2022)

5.2. Discussions of findings from the questionnaire – Data engineering

5.2.1. Data engineering

The responses to the data engineering questionnaire are used for building the data engineering drivers for the data lake canvas. They are considered to be the responses based on the traditional framework. In addition, 4 of the quantifiable responses on data quality, data security, effort required for building the data lake from scratch and without modernization are used for validating the hypothesis on DMM. For the hypothesis testing, responses are collated against the data lake canvas for these 4 questions.

5.2.2. Data Engineering / Architecture factors– Survey Insights

5.2.2.1. BNPL data lake preference

1. A hybrid cloud data lake in BNPL (Buy-Now-Pay-Later) refers to a data storage and management system combining private and public cloud infrastructure to store and process large amounts of data related to BNPL services.

    a. A hybrid cloud data lake in BNPL can streamline this process by providing a scalable and flexible data storage and analysis platform. By combining private and public cloud infrastructure, BNPL providers can take advantage of the cost-effectiveness and agility of public cloud services while maintaining the security and control of private cloud infrastructure.

    b. Overall, a hybrid cloud data lake in BNPL can help providers improve their data management capabilities, reduce costs, and provide customers with more personalized and efficient services.

2. A private cloud data lake in BNPL (Buy-Now-Pay-Later) refers to a data storage and management system hosted and managed within a private cloud

infrastructure. The data is stored within the BNPL provider's infrastructure and not on a third-party server.

a. In the context of BNPL, a private cloud data lake can provide several benefits. Firstly, it can provide greater control and security over the data, as it is not on a public cloud server storage that may be vulnerable to breaches or unauthorized access. It is crucial, given the sensitive nature of customer financial data.

b. Secondly, a private cloud data lake can also provide greater flexibility and customization, as the BNPL provider has more control over the infrastructure and can tailor it to their specific needs. It can improve the speed and efficiency of data processing and analysis.

c. However, one potential disadvantage of a private cloud data lake is that it may be more expensive to set up and maintain than a public cloud service. Additionally, it may offer different scalability and agility than a public cloud service, limiting the ability to quickly adapt to changing business needs.

d. Overall, a private cloud data lake can be a viable option for BNPL providers looking to store and manage large amounts of customer data, mainly if security and control are top priorities.

3. A multi-cloud data lake in BNPL (Buy-Now-Pay-Later) refers to a data storage and management system that uses multiple cloud providers to store and manage data. In this scenario, BNPL providers may use different cloud services for different aspects of their data management, such as storage, processing, and

analysis.

a. One of the main advantages of a multi-cloud data lake is that it can provide greater flexibility and redundancy. Using multiple cloud providers, BNPL providers can avoid relying on one provider for all their data needs. It can help mitigate data loss or downtime due to outages or other issues with a single cloud provider.

b. Additionally, a multi-cloud data lake can also provide greater cost-effectiveness, as BNPL providers can select the most cost-effective cloud service for each aspect of their data management. For example, they may use one cloud provider for storage, another for processing, and yet another for analysis.

c. However, a multi-cloud data lake requires greater management and coordination between cloud providers. BNPL providers must ensure that their data is properly integrated and managed across different cloud platforms, which can require additional resources and expertise.

d. Overall, a multi-cloud data lake can be a viable option for BNPL providers looking to store and manage large amounts of customer data, mainly if they prioritize flexibility, redundancy, and cost-effectiveness. However, it does require additional management and coordination to ensure optimal performance and data integrity.

5.2.2.2. BNPL data lake architecture preference

1. Microservices architecture approach emphasizes creating small, independent, and loosely-coupled services that work together to form a BNPL daya lake application

or system. In the context of a data lake for BNPL (Buy-Now-Pay-Later), microservices architecture can create a modular and scalable data management system that can handle large volumes of data while maintaining flexibility and agility. Some of the key benefits of using microservices architecture for a data lake in BNPL include the following:

a. Scalability: Microservices architecture allows BNPL providers to scale individual components of the data lake independently, allowing them to handle large volumes of data more efficiently.

b. Flexibility: Microservices architecture allows BNPL providers to add, remove, or modify components of the data lake without impacting the entire system, making it easier to adapt to changing business needs.

c. Resilience: Microservices architecture allows BNPL providers to design the data lake to minimize the impact of failures, with each service designed to be resilient and fault-tolerant.

d. Faster time-to-market: Microservices architecture allows BNPL providers to develop and deploy new features or services more quickly, as each service can be developed and tested independently.

e. Improved resource utilization: Microservices architecture allows BNPL providers to optimize resource utilization by allocating resources only to the services that need them rather than the entire system.

f. However, implementing a microservices architecture for a data lake in BNPL can be complex and requires high expertise. BNPL providers must consider factors such as service discovery, inter-service communication,

and data governance when designing and implementing a microservices-based data lake.

g. Overall, a microservices architecture can be a viable option for BNPL providers looking to create a scalable and flexible data lake that can handle large volumes of customer data while maintaining agility and resilience. However, it requires careful planning and execution to ensure optimal performance and data integrity.

A global cloud microservices platform enables businesses to scale applications horizontally. By deploying microservices independently, organizations have the flexibility to add or remove services as needed, effectively utilizing resources and optimizing costs. (Elkholy and A. Marzok, 2022)

2. Data mesh architecture is a relatively new data management approach that emphasizes creating data products by cross-functional teams of domain experts. In the context of a data lake for BNPL (Buy Now Pay Later), data mesh architecture can create a more decentralized and collaborative approach to data management that empowers domain experts to take ownership of their data products and improve the overall quality of the data. Some of the key benefits of using data mesh architecture for a data lake in BNPL include the following:

    a. Improved data quality: Data mesh architecture allows domain experts to take ownership of their data products, leading to better data quality and accuracy.

    b. Increased agility: Data mesh architecture allows domain experts to create and update data products more quickly and efficiently, leading to faster

time-to-market and greater agility.

c. Decentralized governance: Data mesh architecture allows for a more decentralized approach to data governance, with domain experts taking ownership of their data products and collaborating with others to ensure data consistency and accuracy.

d. Better alignment with business needs: Data mesh architecture allows domain experts to focus on the specific business needs of their domain, resulting in more relevant and valuable data products.

e. However, implementing a data mesh architecture for a data lake in BNPL can be complex and requires a significant shift in organizational culture and mindset. When designing and implementing a data mesh-based data lake, BNPL providers must consider data ownership, data discovery, and data product management factors.

f. Overall, a data mesh architecture can be a promising approach for BNPL providers looking to create a more collaborative and agile data lake that aligns more closely with business needs and improves data quality. However, it requires careful planning and execution to ensure optimal performance and data integrity.

Data mesh is a socio-technical methodology that focuses on decentralized analytics data management. To effectively handle this decentralized approach, data mesh relies on the automation capabilities offered by a self-service data infrastructure platform.(Panigrahy et al., 2023)

3. Event-driven architecture (EDA) is an approach to software architecture that

emphasizes using events to trigger and communicate between services or components. In the context of a data lake for BNPL (Buy Now Pay Later), event-driven architecture can create a system that responds to events in real-time, allowing for faster data processing and analysis. Some of the key benefits of using event-driven architecture for a data lake in BNPL include the following:

a. Real-time processing: Event-driven architecture allows for real-time processing and data analysis, which can be critical in a fast-paced industry like BNPL.

b. Scalability: Event-driven architecture can be highly scalable, as services can be added or removed to meet changing demands.

c. Flexibility: Event-driven architecture allows for a more flexible and agile system, as services can be easily modified or replaced to meet changing business needs.

d. Better resource utilization: Event-driven architecture can optimize resource utilization by only processing events that require attention rather than processing all data all the time.

e. However, implementing an event-driven architecture for a data lake in BNPL can be complex and requires careful planning and execution. BNPL providers must consider factors such as event sourcing, data consistency, and event-driven workflows when designing and implementing an event-driven data lake.

f. Overall, an event-driven architecture can be a promising approach for BNPL providers looking to create a more real-time, scalable, and flexible

data lake that can handle the demands of a fast-paced industry. However, it requires careful planning and execution to ensure optimal performance and data integrity.

While event-driven architectures have long been recognized as the go-to choice for decoupled, adaptable, and progressive architectures, it is only in recent times that enterprises have started embracing them specifically for implementing distributed microservices.(Oliveira Rocha, 2022)

4. GFS (Google File System) and HDFS (Hadoop Distributed File System) are file systems designed for large-scale data processing and storage. In the context of a data lake for BNPL (Buy Now Pay Later), GFS/HDFS architecture can create a distributed file system to store and process vast amounts of data. Some of the key benefits of using GFS/HDFS architecture for a data lake in BNPL include the following:

    a. Scalability: GFS/HDFS architecture can scale horizontally to support the storage and processing needs of large volumes of data.

    b. Fault tolerance: GFS/HDFS architecture includes data replication and fault tolerance features, ensuring data protection against failures and errors.

    c. Data locality: GFS/HDFS architecture allows for data locality, where data is stored closer to the required computation, reducing network traffic and improving performance.

    d. Flexibility: GFS/HDFS architecture is flexible and can be used with a wide range of processing frameworks, allowing for various use cases and applications.

e. However, implementing GFS/HDFS architecture for a data lake in BNPL can be complex and requires careful planning and execution. BNPL providers must consider factors such as data partitioning, data replication, and data consistency when designing and implementing a GFS/HDFS-based data lake.

f. Overall, GFS/HDFS architecture can be a powerful approach for BNPL providers looking to create a scalable, fault-tolerant, and flexible data lake that can handle the demands of large-scale data processing and storage. However, it requires careful planning and execution to ensure optimal performance and data integrity.

5.2.2.3. BNPL data lake data structure preference

1. B-trees, B+ trees, Reverted Index B+ trees, and Memtable are all data structures used for organizing and storing large amounts of data in the BNPL data lake. Each structure has unique advantages and disadvantages, and the optimal choice for a BNPL data lake will depend on the specific requirements and use cases. A brief comparison of these data structures:

    a. B-trees: B-trees are a balanced tree structure that stores data in sorted order. Databases and file systems use them for indexing data. B-trees can be efficient for searching and inserting data but require a large amount of memory.

    b. B+ trees: B+ trees are similar to B-trees, but they store data only in the leaf nodes, with pointers to the next leaf node. It allows for more efficient disk access and data retrieval, especially for range queries. Databases and file systems use them for indexing data.

c. Reverted Index B+ trees: Reverted Index B+ trees are similar to B+ trees but optimized for write-heavy workloads. They use "reverted indexes" to reduce the amount of disk access needed for writes, improving performance.

d. Memtables: Memtables are an in-memory data structure for storing data in a key-value format. Distributed databases such as Apache Cassandra use them. Memtables can be very efficient for write-heavy workloads but can be limited by the amount of available memory.

e. In summary, B-trees and B+ trees are well-suited for indexing and searching data, while Reverted Index B+ trees are optimized for write-heavy workloads. Memtables are ideal for write-heavy workloads that require fast in-memory storage but can be limited by available memory. The choice of data structure will depend on the specific requirements and use cases of the BNPL data lake.

5.2.2.4. BNPL data lake data variety and data model

1. In a BNPL data lake, structured, semi-structured, and unstructured data may all be present, and each type of data presents its unique challenges and opportunities. A brief comparison of these data types:

    a. Structured data: Structured data is organized in a specific format, with a fixed schema that defines the relationships between data elements. Structured data is typically stored in relational databases and can be queried using SQL. In a BNPL data lake, structured data may include customer profiles, transaction details, and payment history. Structured data is generally easier to manage and analyze than semi-structured or

unstructured data. However, scaling can be more challenging and may only capture some information required for advanced analytics.

b. Semi-structured data: Semi-structured data has some organizational structure but does not conform to a fixed schema like structured data. Semi-structured data is often stored in JSON, XML, or other document-based formats and can be queried using NoSQL databases. In a BNPL data lake, semi-structured data may include data from web logs, customer feedback forms, and social media interactions. Semi-structured data can be more flexible than structured data, allowing for more complex analysis and modeling, but it can also be more challenging to manage and query.

c. Unstructured data: Unstructured data has no organizational structure and is typically stored as raw text or binary files. In a BNPL data lake, unstructured data may include images, videos, audio recordings, and other multimedia content. Unstructured data is difficult to manage and analyze, but it can provide valuable insights when combined with other data sources using machine learning and natural language processing techniques.

d. Overall, each type of data presents unique challenges and opportunities in a BNPL data lake. BNPL providers must carefully consider their data architecture and management strategies to effectively handle structured, semi-structured, and unstructured data in their data lake.

2. SQL and NoSQL work as database technologies in BNPL data lakes. A brief comparison of these two types of databases:

a. SQL databases: SQL databases are based on the relational data model and use SQL (Structured Query Language) to access and manipulate data. SQL databases are typically used for structured data and provide a predefined schema for the data. In a BNPL data lake, SQL databases may be used for storing and analyzing transaction data, customer profiles, and other structured data. SQL databases are well-suited for data consistency and transactions, and they can provide a high level of security and data integrity.

b. NoSQL databases: NoSQL databases are not based on the relational data model and use other methods to store and access data. NoSQL databases are typically used for semi-structured and unstructured data and can handle large volumes of data with high velocity and variety. In a BNPL data lake, NoSQL databases may store web logs, social media data, and other unstructured data. NoSQL databases are well-suited for scalability, high availability, and performance, but they may provide a different level of consistency and data integrity than SQL databases.

c. In summary, SQL databases best suit for structured data that requires consistency and transactions. In contrast, NoSQL databases best suit for semi-structured and unstructured data that require scalability and performance. BNPL providers must carefully consider their data needs and use cases when selecting the appropriate database technology for their data lake. In some cases, a hybrid approach that combines SQL and NoSQL databases may be the best solution for a BNPL data lake.

3. In a BNPL data lake, different data models can be used to organize and analyze

data. A brief comparison of the document data model, graph data model, and polyglot data model:

a. Document data model: The document data model is a collection of self-contained documents, typically in JSON or XML format. Each document contains all the data for a single entity, organized into fields and nested structures. In a BNPL data lake, the document data model can store semi-structured and unstructured data, such as customer feedback and web logs. The document data model best suits data that can represent a single object with complex attributes and provides flexibility for evolving schemas and dynamic data.

b. Graph data model: The graph data model represents data as nodes and edges and captures complex relationships between entities. In a BNPL data lake, the graph data model can analyze transaction networks, customer behavior patterns, and social media interactions. The graph data model is well-suited for data involving many-to-many relationships and provides powerful tools for graph-based queries and analysis.

c. Polyglot data model: The polyglot data model involves using multiple data models within the same data lake. This approach allows BNPL providers to use the best data model for each type of data and leverage the strengths of multiple models. In a BNPL data lake, the polyglot data model can store structured, semi-structured, and unstructured data in separate data stores, each optimized for their respective data models. The polyglot data model is well-suited for data that has diverse data types and varying access patterns.

d. In summary, the document data model best suits semi-structured and unstructured data that can represent a single object, the graph data model best suits data with complex relationships, and the polyglot data model best suits data with diverse data types and varying access patterns. BNPL providers must carefully consider their data needs and use cases when selecting the appropriate data model for their data lake.

5.2.3. Data lake ML factors – Survey Insights

1. Data quality which is a major concern for AI / ML in the BNPL data lake is discussed in detail in section above.

2. Some tools and technologies that can be used in a BNPL data lake to enable effective AI/ML and ML pipelines:

    a. Data integration tools: Data integration tools can help to combine data from various sources, such as transactional data, customer data, and third-party data. Examples of data integration tools include Apache NiFi, Talend, and Informatica.

    b. Data ingestion tools: Data ingestion tools can help to efficiently and reliably ingest data into a data lake, including streaming data. Examples of data ingestion tools include Apache Kafka, AWS Kinesis, and Google Cloud Pub/Sub.

    c. Data processing frameworks: Data processing frameworks can help to transform and process data at scale and provide a distributed computing infrastructure for AI/ML workloads. Examples of data processing frameworks include Apache Spark, Apache Flink, and Apache Beam.

d. Machine learning frameworks: Machine learning frameworks can help to build and deploy AI/ML models at scale and provide tools for model training, validation, and deployment. Examples of machine learning frameworks include TensorFlow, PyTorch, and Scikit-learn.

e. Data visualization tools: Data visualization tools can help to explore and understand data and provide interactive dashboards and visualizations for AI/ML insights. Examples of data visualization tools include Tableau, Power BI, and D3.js.

f. Cloud platforms: Cloud platforms can provide a scalable and secure infrastructure for a BNPL data lake and enable seamless integration with AI/ML services. Examples of cloud platforms include AWS, Google Cloud, and Microsoft Azure.

g. Data governance and security tools: Data governance and security tools can help ensure data quality, privacy, and Compliance and provide data access control and auditing tools. Examples of data governance and security tools include Apache Ranger, AWS Lake Formation, and Google Cloud Data Loss Prevention.

h. In summary, many tools and technologies can be used in a BNPL data lake to enable effective AI/ML. BNPL providers must carefully consider their data needs and use cases when selecting the appropriate tools and technologies for their data lake.

3. A feature store is a centralized repository for storing and managing features (i.e., input variables) used in machine learning models. In the context of BNPL payments and transactions, a feature store can store and manage features related

to customer behavior, payment history, transactional data, and other relevant data points. Here are some potential use cases for a feature store in a BNPL data lake:

a. Fraud detection: A feature store can store and manage features related to transaction history, such as transaction amount, location, and frequency, as well as customer behavior, such as browsing history and device information. These features can build models for fraud detection and prevention.
b. Risk assessment: A feature store can store and manage features related to creditworthiness, such as credit score, payment history, and debt-to-income ratio, as well as other relevant data points, such as employment history and income. These features can build models for risk assessment and credit scoring.
c. Customer segmentation: A feature store can store and manage features related to customer behavior, such as browsing and purchasing history, as well as demographic and geographic data. These features can build models for customer segmentation and targeting.
d. Personalization: A feature store can store and manage features related to customer preferences and behavior, such as purchase history and product reviews, as well as contextual data, such as location and time of day. These features can build models for personalized product recommendations and marketing.
e. To build and manage a feature store for BNPL payments and transactions, BNPL providers must consider data quality, feature engineering, versioning, and integration with machine learning workflows. Some popular feature store solutions include Feast, Hopsworks, and Tecton.

3. In addition to the use cases for fraud detection, risk assessment, customer

segmentation, and personalization discussed earlier, a feature store in a BNPL data lake can also build customer 360 and merchant 360 views. Here is how a feature store can solve these business use cases:

a. Customer 360: A feature store can store and manage features related to customer behavior, such as purchase history, browsing history, and product reviews, as well as demographic and geographic data, such as age, gender, and location. These features can build models for customer 360 views, which provide a comprehensive view of a customer's interactions with the BNPL provider across multiple touchpoints. It can help BNPL providers to understand customer preferences, behavior, and needs and to deliver a more personalized customer experience.
b. Merchant 360: A feature store can also store and manage features related to merchant behavior, such as transaction history, customer feedback, and product inventory, as well as demographic and geographic data, such as location and industry. These features can then build models for merchant 360 views, which provide a comprehensive view of a merchant's interactions with the BNPL provider across multiple touchpoints. It can help BNPL providers to understand merchant needs and preferences and to deliver a more personalized merchant experience.
c. To build a feature store for customer 360 and merchant 360 use cases, BNPL providers will need to ensure that data from various sources (e.g., transactional data, browsing data, feedback data) is integrated and stored in a centralized repository and that features are engineered

to capture relevant customer and merchant behaviors. BNPL providers can leverage existing feature store solutions, such as the ones mentioned earlier, or build their feature store using open-source technologies such as Apache Hadoop or Apache Spark.

a. A feature store can also be used in a BNPL data lake to support product and pricing use cases. Here is how:

    d. Data integration: A feature store can integrate and store data from various sources, such as transactional data, product data, customer data, and external data sources (e.g., weather data, economic data). This data can create features that capture relevant product and pricing information.

    e. Feature engineering: A feature store can engineer features that capture relevant product and pricing information, such as product attributes, pricing history, customer demand, market trends, and competitor pricing. These features can build machine learning models to help BNPL providers make data-driven product and pricing decisions.

    f. Real-time updates: A feature store design can support real-time updates, ensuring that the data used for product and pricing decisions are up-to-date and accurate.

    g. Model Management: A feature store can be integrated with model management workflows to support model training, validation, and deployment. It can help BNPL providers build models that provide real-time product and pricing recommendations.

    h. Using a feature store in a BNPL data lake for product and pricing use cases, BNPL providers can make data-driven decisions to improve

customer engagement and loyalty. They can use machine learning models to optimize pricing strategies, create personalized product recommendations, and identify product gaps in the market. BNPL providers can leverage existing feature store solutions, such as Feast, Hopsworks, and Tecton, or build their feature store using open-source technologies such as Apache Hadoop or Apache Spark.

5.3. Comments on responses

Some of the responses from the survey do not fit into the analysis for making meaningful insights. They are considered outliers and either grouped under minority or removed from the data analysis.

Some of the responses from the survey are with multiple answers as the multiple choice was enabled. Those responses are handled in the Python code to handle them accordingly for data analysis.

Some responses might be skewed compared to the new ones on the data lake canvas.

5.4. Discussions of findings on cost optimization

Cost optimization can be considered with various factors, and operational cost (OpEx) is considered for the research study. Responses from the survey are considered for the traditional approach, and for the data lake canvas, new responses are considered based on the inputs from the experts.

5.4.1. Discussion of results

New responses are collated by having focused discussions with experts, including fintech leaders, data leaders, engineering leads, analytical leaders, and data architects. Table 43 shows the new responses from the experts as the desired state for operational cost with

the data lake canvas.

Table 43: Operational cost – New response with data lake canvas

Cost Optimization response	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
Increase by 5 - 10%	0	0	0	1	10%
Increase by 10 - 20%	0	0	0	0	2%
Increase by 20 - 40%	0	0	0	0	0%
Decrease by 5 - 10%	1	0	0	0	15%
Decrease by 10 - 20%	1	1	1	1	40%
Decrease by 20 - 40%	0	1	2	0	28%
Others	0	1	0	0	5%
Total Response	2	3	3	2	10

Most of the experts were clear that it would decrease the cost, but only some considered the increase % as well, and others also marked where the experts do not believe in the data lake canvas. Hence the 2% for 'Increase by 10 - 20%' and 5% for 'Others' is based on the expert inputs on the overall % though it does not account for their individual response. This new response % is simulated with the overall response against the traditional framework used for hypothesis testing and data comparison.

5.5. Discussions of findings on the Data Maturity Model (DMM)

Data Maturity Model (DMM) is considered for four factors: data quality, data security, data governance, and time to market. The four factors can be calculated in different ways, and for the research study, the responses from the survey are considered for calculating the index score. Responses from the survey are considered for the traditional approach, and for the data lake canvas, new responses are considered based on the inputs from the experts.

5.5.1. Discussion of results

New responses are collated by having focused discussions with experts, which include fintech leaders, data leaders, engineering leads, analytical leaders, and data architects.

Data Quality

Table 44 shows the new responses from the experts as the desired state for data quality issues with the data lake canvas.

Table 44: Data Quality – New response with data lake canvas

Data Quality Issues	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
0-20%	0	2	3	2	75%
20-40%	1	1	0	0	15%
>40%	1	0	0	0	10%
Others	0	0	0	0	0%
Total Responses	2	3	3	2	100%

Most experts were clear that it would decrease the data quality with the data lake canvas.

Data Security

Table 45 shows the new responses from the experts as the desired state for data security issues with the data lake canvas.

Table 45: Data Security – New response with data lake canvas

Data Security Issues	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
0-1%	0	1	1	1	30%
2-5%	1	1	1	1	40%
5-10%	1	1	0	0	20%

Data security is not measured though implemented	0	0	1	0	4%
Depends	0	0	0	0	2%
Not sure	0	0	0	0	4%
Total Response	2	3	3	2	100%

Most experts were clear that it would decrease data security, but few do not believe in the data lake canvas. Hence the 4% for 'Data security is not measured though implemented', 2% for 'Depends', and 4% for 'Not sure' based on the expert inputs on the overall % though it does not account for their individual response.

This new response % is simulated with the overall response against the traditional framework, which is used for hypothesis testing and data comparison.

Data Governance

Table 46 shows the new responses from the experts as the desired state for data governance issues with the data lake canvas.

Table 46: Data Governance – New response with data lake canvas

Data governance Issues	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
Decrease by 10 - 20%	0	0	1	0	10%
Decrease by 20 - 40%	0	1	0	0	10%
Decrease by 5 - 10%	1	1	1	1	40%
Increase by 10 - 20%	0	1	0	0	10%
Increase by 20 - 40%	1	0	0	0	10%
Increase by 5 - 10%	0	0	1	0	10%
Others	0	0	0	1	10%
Total Response	2	3	3	2	100%

Most experts expected the data governance issues to reduce by 5-10% with the data lake

canvas, but other options were also expected.

Time to Market

Table 47 shows the new responses from the experts as the desired state for the Effort to migrate to the data lake without modernization – TTM_1I with the data lake canvas.

Table 47: TTM_1I – New response with data lake canvas

Time To Market (without modernization)	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
Startup - < 1 months	1	1	0	0	20%
SMB < 2 months	1	1	0	1	25%
Enterprise < 3 months	0	1	3	1	55%
Total Response	2	3	3	2	100%

Table 48 shows the new responses from the experts as the desired state for the Effort to build a data lake from scratch – TTM_2I with the data lake canvas.

Table 48: TTM_2I – New response with data lake canvas

Time To Market (without modernization)	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
Enterprise > 12 months	1	1	0	0	15%
Startup - < 2 months	1	0	1	0	25%
SMB - 3 - 4 months	0	1	1	1	30%
Enterprise > 6 months	0	1	1	1	30%
Total Response	2	3	3	2	100%

This new response % is simulated with the overall response against the traditional framework, which is used for hypothesis testing and data comparison.

5.6. Discussions of findings on Data as an Asset (DaaA)

Data as an Asset (DaaA) is considered a factor of DaaP and DaaP for which the inferences can be made from the survey responses. However, they are not quantifiable or can be evaluated to prove DaaA in the data lake. Hence the unit of economies framework is utilized to measure the DaaP and DaaS.

Due to the sensitivity of the Fintech industry, the results presented are from the non-Fintech industry. However, the inputs from non-Fintech can be studied to compare their usefulness. The metrics on the cost are in their simplest form to derive the useful and do not reflect the financial standards.

5.7. Conclusion & Insights

In this study, we aimed to answer the research question: How to bridge the gap between Fintech-BNPL and Data lake with a successful strategy to build a Fintech data lake?

To accomplish this, we have the data enemies and data divinity defined, and we have three objectives: to study the cost optimization, to study the DMM with 4 factors, and to measure DaaA.

In order to accomplish the first objective, an analysis of survey responses on the Operational expense is studied with the industry experts. The comparison is between the traditional approach and the data lake canvas dataset through statistical tests. Statistical tests showed that the operational expense of traditional framework samples is significantly different from the data lake canvas. Also, the comparison suggests that the OpEx can be majorly decreased by 10 - 20% or decreased by 20 - 40%.

In order to accomplish the second objective, an analysis of survey responses on the

DMM is studied with industry experts based on 4 factors. The comparison is between the traditional approach and the data lake canvas dataset with statistical tests for the 3 factors of data quality, data security, and data governance. Statistical tests showed that the traditional framework samples are significantly different from the data lake canvas. The scoring test is done for TTM_1I and TTM_2I for the traditional approach and data lake canvas dataset. Also, scoring with weightage suggests that the DMM with the data lake canvas provide DQ issues majorly to be within 0 – 20%, DS issues majorly to be within 0 – 1% & 2-5%, DG issues majorly to be within 5 – 10%, TTM_1 (lift-shifT) expected to be Startup - < 1 months, SMB < 2 months, Enterprise < 3 months, TTM_2 (build data lake from scratch) expected to be Startup - < 2 months, SMB - 3 - 4 months, Enterprise > 6 months.

In order to accomplish the third objective, 'unit of economies' is measured for DaaP and DaaS to quantify the DaaA as a unit of measure with count and cost metric.

It can be considered a reliable benchmark of successful strategy for building Fintech-BNPL data lake – Data Divinity during this period and as a reference for future comparisons.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

Chapter 5 presents a summary, implications, limitations, and recommendations for future research findings presented in the previous chapters. It summarizes the research paper and dives straight into their in-depth interpretation. Furthermore, the chapter presents the implications of the research findings for fintech-BNPL, data leaders, and data engineering experts while building a Fintech-data lake. Finally, the chapter provides recommendations for future research in the area.

6.2 Implications and Recommendations for Future Research

The results presented in this study have certain limitations that should be acknowledged in order to fully understand and interpret the findings.

The results presented in this study on the BNPL challenges are at a high level. Further research can be done in specific subject areas (e.g., risk management) to add business drivers with sub-categories that are more specific so the data lake canvas can be made even more robust.

The research was limited to operational expenses (OpEx) with the survey responses and the new responses from the focused discussions. It may limit the generalizability of the findings on cost optimization as different with more extensive focused discussions may produce different results. Therefore, it is essential to back-test the data on a larger set of samples before considering the results of this research to be true.

The research tried to avoid any look-forward bias and skewness based on the inputs from the responses and focused discussions. However, due to the handling of data to

remove the outliers and other factors, results might vary when used in the application.

The research did test the objective results statistically using non-parametric tests. However, the new responses are simulated based on the industry experts' inputs. However, future researchers can consider running tests with new responses without simulation.

The research used non-Fintech evaluation for the unit for economies to prove the usefulness of measuring the DaaP and DaaS as DaaA. However, future researchers can consider evaluating the framework with a Fintech company.

The research has coined the new term ‘Data Divinity’, a factor to measure the success of data lake with 3 strategies of OpEx, DaaA and DMM. However, future researchers can consider expanding the factors under business, techno-functional and technology outcomes.

Despite the limitations outlined in this section, the research presented in this study provides valuable insights and potential avenues for further exploration in the field of Fintech data lake. Further research, building on the findings and methods presented here, has the potential to strengthen and deepen our understanding of the strategies and techniques for a successful Fintech data lake, which can be leveraged for any data lake.

6.3 Conclusion

The chapter presented a comprehensive discussion and conclusion of the research findings. It summarized the results of the research and provided in-depth interpretation. The chapter also discusses the implications of the research findings for Fintech leaders, data leaders, data engineering experts, and practitioners in data

& analytics. It provides recommendations for future research in the area.

The chapter evaluates the data lake canvas with Data enemies and data divinity to achieve via 3 strategies. The chapter also used bar charts, non-parametric testing, and data comparison charts to study the data.

The research aimed to answer the question: How to bridge the gap between Fintech-BNPL and Data lake with a successful strategy to build a Fintech data lake? The study found that comprehending and identifying data enemies is essential to get to the data divinity state. Data divinity is achieved with 3 strategies on cost optimization, DMM, and DaaA with enablers and drivers. However, the results showed that the performance of the data divinity depends on the specific parameters used.

The chapter also acknowledges the limitations of the study, including the use of non-Fintech data and biases from focused discussions with simulation, which may limit the generalizability of the findings.

APPENDIX A

LIST OF TABLES

Table 1: Fintech Verticals – Market scope and growth

Fintech Vertical	Market Scope	Expected growth / Market share
Payments Technology	Disruption of the payments is due to the evolving need for contactless, real-time payments with enhanced features like BNPL, biometrics, and other latest technologies.	Digital payments will have a total transaction value of US$8,502.00bn in 2022, and users will be 4,929.55m users by 2025
Digital Banking	Cutting-edge technologies have made digital banking easier for Neobanks, Challenger-banks, New-banks, and non-banks.	Market to Reach $30.1 Billion by 2026
Digital Wealth Management	Robo advisors using big data and analytics are radically changing wealth management in serving the HNI and UHNI investors.	The market projection will grow at a CAGR of 13.9% between 2022 and 2027 to reach a value of USD 10,268.9 million by 2027.
Capital Markets	Investors continue to embrace technology with the growth in crypto trading, algorithmic trading, HTF, AI/ML, and RPA for post-trade functions.	The global domestic equity market capitalization value is 116.78 trillion USD.

Fintech Lending & Equity crowdfunding	Fintech Lending is very popular with individuals and SMEs. On the other hand, equity crowdfunding is popular with start-ups and innovative products to raise funds. It has revolutionized the P2P lending/crowdfunding platform with ML, Big data & Analytics for a seamless lending/crowdfunding process and credit scoring mechanism.	The global peer-to-peer (P2P) lending market generated $67.9 billion in 2019 and will reach $558.9 billion by 2027, registering a CAGR of 29.7% from 2020 to 2027.
InsurTech	InsurTech has enabled seamless experience with digitization and automation with AI / ML & RPA for data collection, loss assessment, cost estimation, damage analysis through image recognition, automated self-service guidance, and more.	The global insurtech market size will value USD 2.72 billion in 2020. It will expand at a compound annual growth rate (CAGR) of 48.8% from 2021 to 2028
PropTech	PropTech allows individuals and companies to make acquisition and disposal decisions and manage a real estate portfolio. Apart from AR / VR & IoT, drones modernizing real estate, Proptech converges many functions with Fintech. In the Fintech arena, Blockchain aids in data tracking and reaching immutable data in pricing and	The value of the real estate tech deals worldwide is 8.4 billion USD from 2014 to 2020

ownership rights and using Big data & Analytics for financial process management.

Table 2: (ZHAO, 2021) - Metadata management systems for data lake

S.No	Researcher	Metadata Management Framework	Description	Metadata Management System	Implementation Strategy
1	Walker and Alrehamy (2015)	Personal data lake	Framework has unified data storage and metadata management to help users analyze and query personal data metadata.	Integrated with personal data lake	JSON Object Graph databased with four nodes for metadata, raw data, Semantics, and identified.
2	Hai et al. (2016)	Constance Data Lake	The framework helps users to discover, extract and summarize metadata from structured and semi-structured data (relational databases,	Integrated with constance data lake	Extract explicit and implicit metadata and semantic annotations Cluster the schemas based on the distance between them

			JSON, spreadsheets, and XML).		
3	(Quix et al., 2016)	GEMMS (Generic and Extensible Metadata Management System)	GEMMS extracts metadata automatically from the datasets the data lake repository loads in their original format.	Built with the persistence component	MongoDB handles three main functions 1) metadata extraction, 2) Metadata transformation, and 3) Metadata Storage.
4	(Halevy et al., 2016a,b)	GOODS (Google Dataset Search)	The post-hoc system organizes the datasets that Google generates and uses. The post-hoc manner allows the system to collect and aggregate metadata about datasets after the creation, access, or updating of different pipelines without interfering with	integrated with the GOODS system	It has four principal services that include 1) Search engine, 2) per-dataset profile, 3) monitoring service, 4) Annotation service

			dataset owners' or users'		
5	Suriarachchi and Plale (2016b)	Data Lake of Suriarachchi	They focus on data provenance metadata when they transform processes for data lakes. A data product is stored to get the data lineage from the metadata.	Not a complete metadata management system	Ingest API and graph database.
6	Maccioni and Torlone (2018)	KAYAK Framework	KAYAK helps data scientists to define, execute and optimize data preparation pipelines in a data lake.	Not available	
7	Diamantini et al. (2018)	Metadata model of Diamantini	Metadata model that can structure unstructured data to extract thematic views from heterogeneous and	Not available	

			generally unstructured data sources		
8	Sawadogo et al. (2019)	MEDAL (MEtadata model for DAta Lakes)	Provides evaluation criteria through a list of features for data lake metadata systems and a metadata typology	Not available	
9	Eichler et al. (2020)	HANDLE	HANDLE is a generic model for metadata which enables comprehensive metadata management.	Not available	Neo4j graph DB
10	Scholly et al. (2021)	goldMEDAL	Provide a generic metadata model and enable the data lineage tracing with the concept of process	Applied to 3 data lake systems	1. HOUDAL (Public Housing Data Lake) - Metadata with Neo4j DB 2. AUDAL, which is a textual and tabular data lake - documents are stored in

					MongoDB and Metadata with Neo4j DB 3. Archaeological Data Lake - Implemented with Apache Atlas framework
11	(Zhao, Megdiche, et al., 2021)	Metadata model of Zhao et al., 2021	Provides a metadata model for ingestion with the IoT data lake using the Neo4j framework.	Integrated with IoT data lake	4. Metadata management with Neo4j to support both batch and streaming raw data 5. Front-end application built with HTML/CSS and JavaScript to perform data analytics 6. The back-end uses API Neoviz for visualizing the metadata with data lineage enabled.
12	(Zhao, Ravat, et al., 2021)	Analysis-oriented metadata model	Provides a metadata model related to machine learning analysis on the description	Applied to data lake	3. Manage the metadata to be applied across phases of 1) Data exploration, 2) Data preparation, 3)

	information of datasets and their attributes	Modeling, and algorithms 4. Metadata storage with Neo4j DB

Table 3: (Chelliah and Surianarayanan, 2021) - Challenges in multi-cloud and respective solution approaches

#	Challenge	Solution approaches
1	Interoperability and portability(Nogueira, E et al., 2016)	(i) Open APIs and Standards (ii) The Open Cloud Computing Interface (iii) Automation (iv) DevOps through CI/CD pipeline (v) Infrastructure as Code (IaC) (vi) Microservices Architecture (vii) Spinnaker for Multi-Cloud Software Delivery (viii) Containerization (ix) Serverless computing and management across Multiple Clouds (x) Service Resiliency Frameworks and Libraries for Multi-Clouds (xi) Service mesh orchestration
2	Application & data integration(Senda Romdhani, 2019)	Application and Data Integration Platforms
3	Multi-cloud orchestration(Ming Lu et al., 2018)	(i) intelligent brokers (ii)Container Clustering and Orchestration Platforms
4	Multi-Cloud Monitoring, Measurement and Management(E. Rios et al., 2016)	(i) Multi-Cloud Management and Governance Platforms (ii) Multi-Cloud Monitoring and Measurement Tools (iii) AI-Inspired Log and Operational Analytics Platforms for Multi-Clouds
5	Identity and Access Management(I.Indu et al., 2018)	(i) Next-Generation Identity and Access Management (IAM) Solutions (ii) Edge Cloud Integration with Traditional Clouds (iii) Multi-cloud security

Table 4: Phased growth of a BNPL

BNPL Growth	Online	In-store
Year 1	BNPL added to e-commerce checkout	
Year 2	Expands to more retailers, payment processors, more customer base	Offered in-store

Year 3	Gains popularity with a younger demographic and increases the adoption rate.	Expansion in in-store and gains traction thru' in app
Year 4	Consolidates market share with top players acquiring customers	Adoption reaches critical mass and becomes a mainstream option

Table 5: BNPL Consumer behavior

BNPL Consumer Behavior	Description
Browsing products	Consumers browse online stores for products they are interested in purchasing.
Adding to cart	Once they find something they want to buy, consumers add it to their online shopping cart.
Selecting BNPL option	At checkout, consumers choose the BNPL option, which allows them to pay for their purchases in installments over time.
Completing purchase	Consumers complete their purchase using the BNPL option, often with a few clicks or taps.
Repaying installments	Consumers must repay the BNPL provider per the agreed-upon repayment terms and schedule.
Monitoring account	Consumers may check their accounts regularly to monitor repayment status, upcoming payments, and available credit.

Using multiple providers	Some consumers may use multiple BNPL providers to spread their purchases and repayments across different services.
Managing budget	Consumers may use BNPL services as part of their overall budget management strategy, carefully keeping track of their expenses and repayments.

Table 6: BNPL implications on debt management

Strategy	**Description**
Budgeting	Create a budget that includes repayment of BNPL debts. Ensure that enough income to meet the expenses and BNPL payments is available.
Prioritizing payments	Prioritize BNPL debt repayment over other expenses to ensure that payments are received and incur additional fees.
Negotiating terms	Contact the BNPL provider to negotiate repayment terms if there are struggles to make payments.
Avoiding new purchases	Repay the existing debts to make any new purchases with BNPL services.
Tracking payments	Keep track of BNPL payments, payment dates, and the total amount owed to stay on top of the debt.

Seeking help	Seek help from a financial advisor or credit counseling service if there are struggles with BNPL debt or managing finances.

Table 7: Summary of Objectives & Gaps in BNPL Data lake studies

Study	Objective	Gaps
(Parne, 2021)	Provides cloud computing strategy, impact on banking and financial institutions, and discusses the significant reliance on cloud computing	Lack of addressing the high cost for micro enterprises to adopt it
(Oberoi et al., 2021)	Provides insights into cloud computing usage in the banking industry, the various business models associated with it, and the challenges the banking industry faces in adopting this technology.	Lack on standardization to address the challenges
(Imerman and Fabozzi, 2020)	Showcases FinTech Ecosystem and a conceptual framework for FinTech innovation.	Lack of addressing the business challenges with data technology
(Khan and Vilary Mbanyi, 2022)	Studies on buy now, pay later (BNPL) and its influence on millennials buying behavior and consumption when mobile shopping	Lack of connecting the consumer behavior with data technology
(Guttman-Kenney et al., 2022)	Analysis of an example of how consumer financial protection regulators can use	Lack of framework for addressing BNPL regulation and

	realtime transactions data to monitor markets and evaluate potential risks - especially (largely) unregulated, financial innovations such as BNPL	security issues connecting with data technologies
(Alshahri, 2022)	Offers "buy now, pay later" (BNPL) payment methods for B2B and provide solutions (e.g., liquidity, automation, digital payments) for online and offline SMEs	Lack of data to capture the business drivers on the payment methods and solutions with cost optimization
(Aalders, 2023)	Uses mechanisms and conditions framework of affordances and walkthrough method to analyze how popular BNPL products define responsible lending and spending	Lack of bringing consumer behavior and regulations to handle with data lake
(Vinoth et al., 2022)	Examines several cloud computing applications in banking and e-commerce, as well as the security issues associated with them	Lack of data to protect businesses from security threats
(Hai et al., 2021b)	Provides a comprehensive overview of research questions for designing and building data lakes.	Lack of other architecture aspects on streaming data lakes, and integrating data lakes with machine learning and data science.
(Sawadogo and Darmont, 2021b)	Provides a comprehensive state of the art of the different approaches to data lake	1. Data integration and transformation aspects have

	design focusing on data lake architectures and metadata management, which are critical issues in successful data lakes.	recurring issues. 2. Data governance principles are indeed currently seldom turned into actual solutions.
(Nargesian et al., 2019b)	Discusses how data lakes are introducing problems, including dataset discovery, and how they are changing the requirements for classic problems, including data extraction, cleaning, integration, versioning, and metadata management.	Lack of framework to handle the data issues
(Giebler, Corinna et al., 2021b)	Introduces the data lake architecture framework.	Lack of connection to the business drivers and challenges
(Kumar et al., 2021)	Discusses various available data storage options, suitability and limitations with cloud.	Lack of the underlying need of the business driving the selection of a database

Table 8: Research methodology

S.No	Research Factors	Research Methodology	Primary data collection
1	Build a framework model to collect BNPL data with key business subject	Exploratory & Qualitative research	Survey

	areas.		
2	Possible Data Lake adaptability for BNPL Fintech (Mapping SMB Vs. Large enterprises with Data puddle, data pond, data lake, data ocean, and cloud solution - Cloud service, storage, zone)	Exploratory & Qualitative research	Survey
4	Framework to avoid Garbage dump in BNPL - Fintech data lake	Exploratory & Qualitative research	Survey
5	Factors to consider for building sustaining BNPL Fintech data lake with long-term efficiency	Exploratory & Qualitative research	Survey
6	Factors to consider for building BNPL Fintech Data Lake with quick Time to Market	Exploratory & Qualitative research	Survey
7	Aspects for Setting up New Cloud data lake Vs. Cloud Data Lake migration	Exploratory & Qualitative research	Survey
8	Evaluate the framework for better cost management	Quantitative research	Survey and Interviews

9	Evaluate the framework for Data Maturity Model	Quantitative research	Survey and Interviews
10	Evaluate the framework for Data as an Asset (DaaA)	Quantitative research	Survey and Interviews

Table 9: Objective 1 - Hypothesis testing strategy

Strategy	**Rules**
Opex_cost f(x1) = ↑D(a) where $5 \leq a \leq 10$	% Sampling to evaluate based % of the responses for Opex cost increase by 5-10%
Opex_cost f(x2) = ↑D(a) where $10 \leq a \leq 20$	% Sampling to evaluate based % of the responses for Opex cost increase by 10-20%
Opex_cost f(x3) = ↑D(a) where $20 \leq a \leq 40$	% Sampling to evaluate based % of the responses for Opex cost increase by 20-40%
Opex_cost f(x4) = ↓D(a) where $5 \leq a \leq 10$	% Sampling to evaluate based % of the responses for Opex cost decrease by 5-10%
Opex_cost f(x5)= ↓D(a) where $10 \leq a \leq 20$	% Sampling to evaluate based % of the responses for Opex cost decrease by 10-20%

Opex_cost f(x6) = ↓D(a) where 20 ≤ a ≤ 40	% Sampling to evaluate based % of the responses for Opex cost decrease by 20-40%
Opex_cost f(x7) = D(a) where a is undefined	% Sampling to evaluate based % of the responses for Opex cost decrease by 20-40%

Table 10: Objective 2 - Hypothesis testing strategy

Strategy	**Rules**
f(DQI1) = ↑ D(q) where 0 ≤ q ≤ 20	% Sampling to evaluate based % of the responses for data quality issues in data lake between 0-20%
f(DQI2) = ↑ D(q) where 20 ≤ q ≤ 40	% Sampling to evaluate based % of the responses for data quality issues in data lake between 20-40%
f(DQI3) = ↑ D(q) where q ≥40	% Sampling to evaluate based % of the responses for data quality issues in data lake >40%
f(DQI4) = D(q) where q is undefined	% Sampling to evaluate based % of the responses for data quality issues in data lake due to other reasons and not measured
f(DSI1) = ↑ D(s) where 0 ≤ s	% Sampling to evaluate based % of the responses for

≤ 1	data security issues in data lake between 0-1%
f(DSI2) = ↑ D(s) where $2 \leq s \leq 5$	% Sampling to evaluate based % of the responses for data security issues in data lake between 2-5%
f(DSI3) = ↑ D(s) where $5 \leq s \leq 10$	% Sampling to evaluate based % of the responses for data security issues in data lake between 5-10%
f(DSI4) = D(s) where s is undefined	% Sampling to evaluate based % of the responses for data security issues in data lake due to other reasons and not measured
f(DGI1) = ↑ D(g) where $5 \leq g \leq 10$	% Sampling to evaluate based % of the responses for garbage dump in data lake increase by 5-10%
f(DGI2) = ↑ D(g) where $10 \leq g \leq 20$	% Sampling to evaluate based % of the responses for garbage dump in data lake increase by 10-20%
f(DGI3) = ↑ D(g) where $20 \leq g \leq 40$	% Sampling to evaluate based % of the responses for garbage dump in data lake increase by 20-40%
f(DGI4) = ↓ D(g) where $5 \leq g \leq 10$	% Sampling to evaluate based % of the responses for garbage dump in data lake decrease by 5-10%
f(DGI5) = ↓ D(g) where $10 \leq g \leq 20$	% Sampling to evaluate based % of the responses for garbage dump in data lake decrease by 10-20%

$f(DGI6) = \downarrow D(g)$ where $20 \leq g \leq 40$	% Sampling to evaluate based % of the responses for garbage dump in data lake decrease by 20-40%
$f(DGI7) = D(g)$ where g is undefined	% Sampling to evaluate based % of the responses for garbage in data lake due to other reasons
$f(TTM_1 1) = \uparrow D(t)$ where $0 \leq t \leq 1$	% Sampling to evaluate based % of the responses for effort required to migrate data without modernization for startup < 1 month
$f(TTM_1 2) = \uparrow D(t)$ where $0 \leq t \leq 2$	% Sampling to evaluate based % of the responses for effort required to migrate data without modernization for SMB < 2 months
$f(TTM_1 3) = \uparrow D(t)$ where $0 \leq t \leq 3$	% Sampling to evaluate based % of the responses for effort required to migrate data without modernization for Enterprise < 3 months
$f(TTM_1 4) = \uparrow D(t)$ where t is undefined	% Sampling to evaluate based % of the responses for effort required to migrate data without modernization which can be other
$f(TTM_2 1) = \uparrow D(t)$ where $0 \leq t \leq 2$	% Sampling to evaluate based % of the responses for effort required to build data lake from scratch for startup < 2 months

$f(TTM_2 2) = \uparrow D(t)$ where $t \geq 2$	% Sampling to evaluate based % of the responses for effort required to build data lake from scratch for startup > 2 months
$f(TTM_2 3) = \uparrow D(t)$ where $3 \leq t \leq 4$	% Sampling to evaluate based % of the responses for effort required to build data lake from scratch for SMB < 3 - 4 months
$f(TTM_2 4) = \uparrow D(t)$ where $0 \leq t \leq 6$	% Sampling to evaluate based % of the responses for effort required to build data lake from scratch for SMB is 6 months
$f(TTM_2 5) = \uparrow D(t)$ where $t \geq 6$	% Sampling to evaluate based % of the responses for effort required to build data lake from scratch for Enterprise > 6 months
$f(TTM_2 6) = \uparrow D(t)$ where $t \geq 12$	% Sampling to evaluate based % of the responses for effort required to build data lake from scratch for Enterprise > 12 months
$f(TTM_2 7) = \uparrow D(t)$ where t is undefined	% Sampling to evaluate based % of the responses for effort required to build data lake from scratch which can be other

Table 11: Objective 3 - Hypothesis testing strategy

Strategy	Rules
f(DaaS)	% Sampling to evaluate based on the number of data services derived from the data lake
f(DaaP)	% Sampling to evaluate based on the number of data products derived from the data lake

APPENDIX B

LIST OF FIGURES

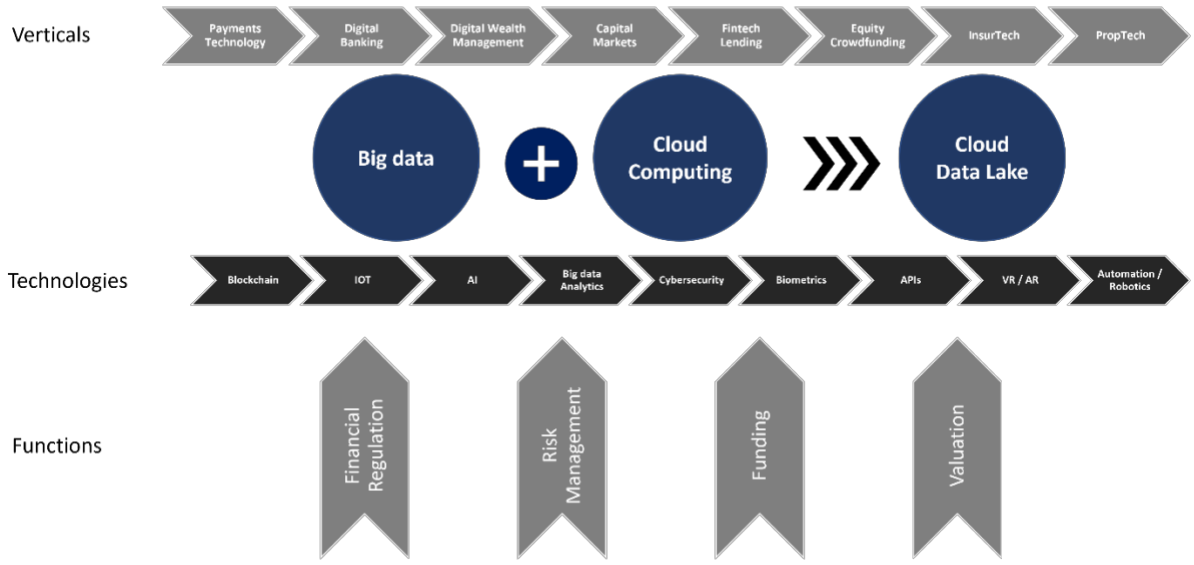


Figure 1: Realigned Fintech Model with data lake

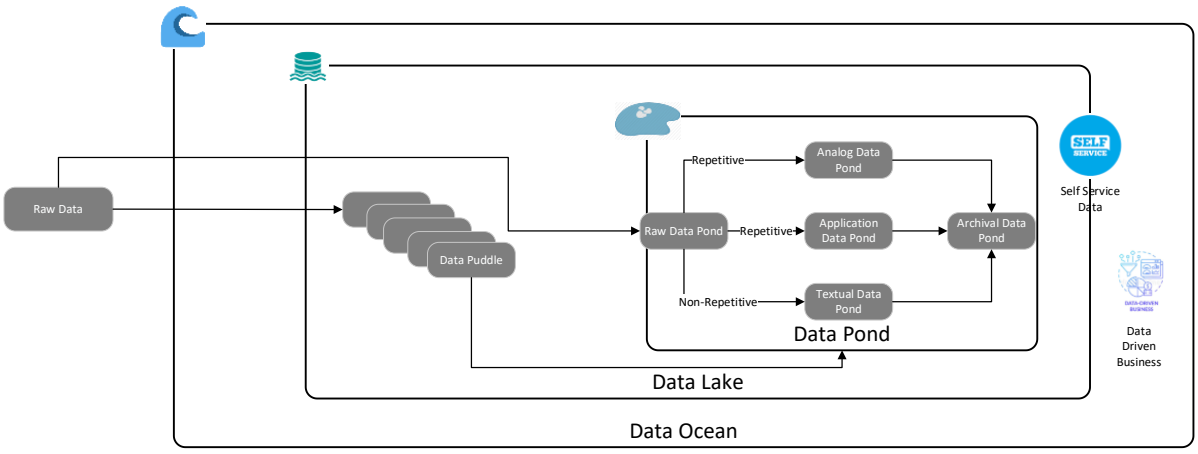


Figure 2: Realigned Data Lake Organization

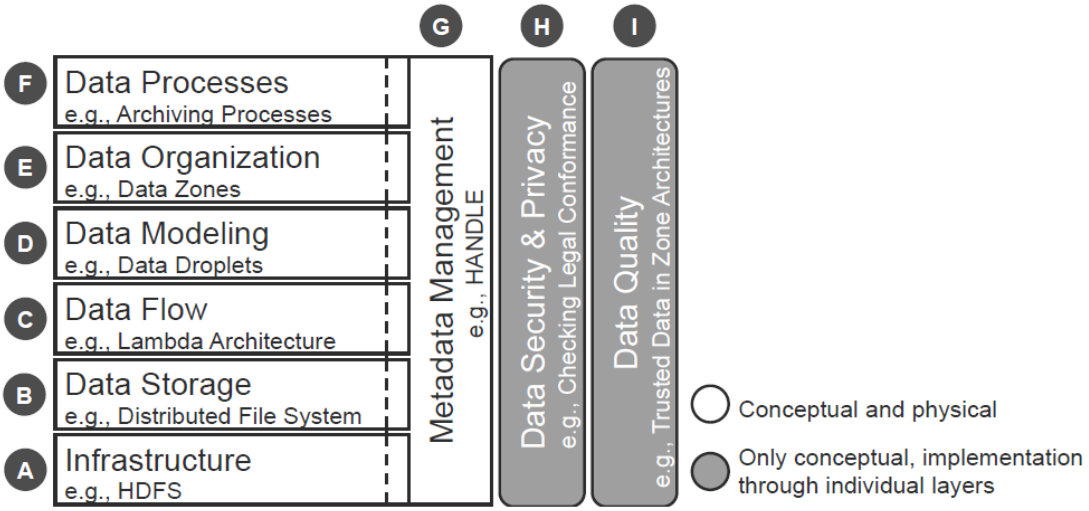


Figure 3: (Giebler, Corinna et al., 2021a) Data Lake Architecture Framework (DLAF)

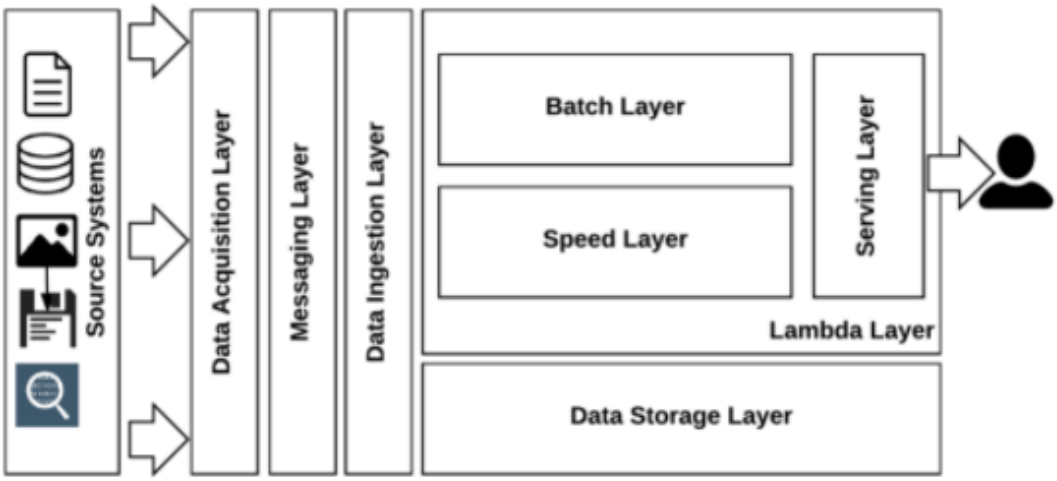


Figure 4: (John and Misra, 2017) Layers in a Data Lake

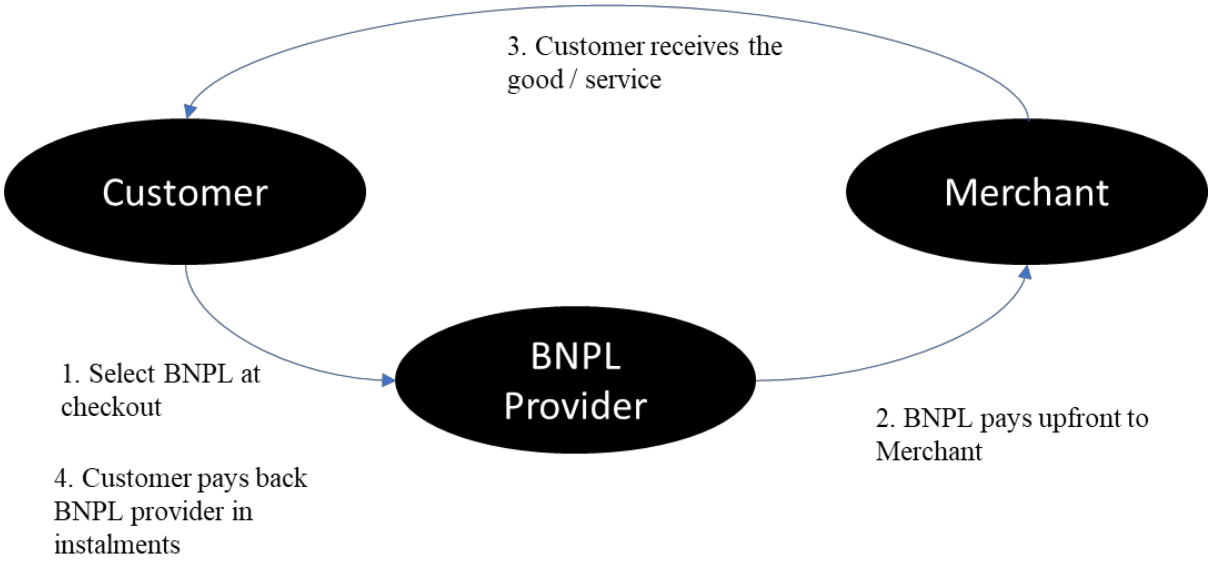


Figure 5: Buy-now-pay-later business flow

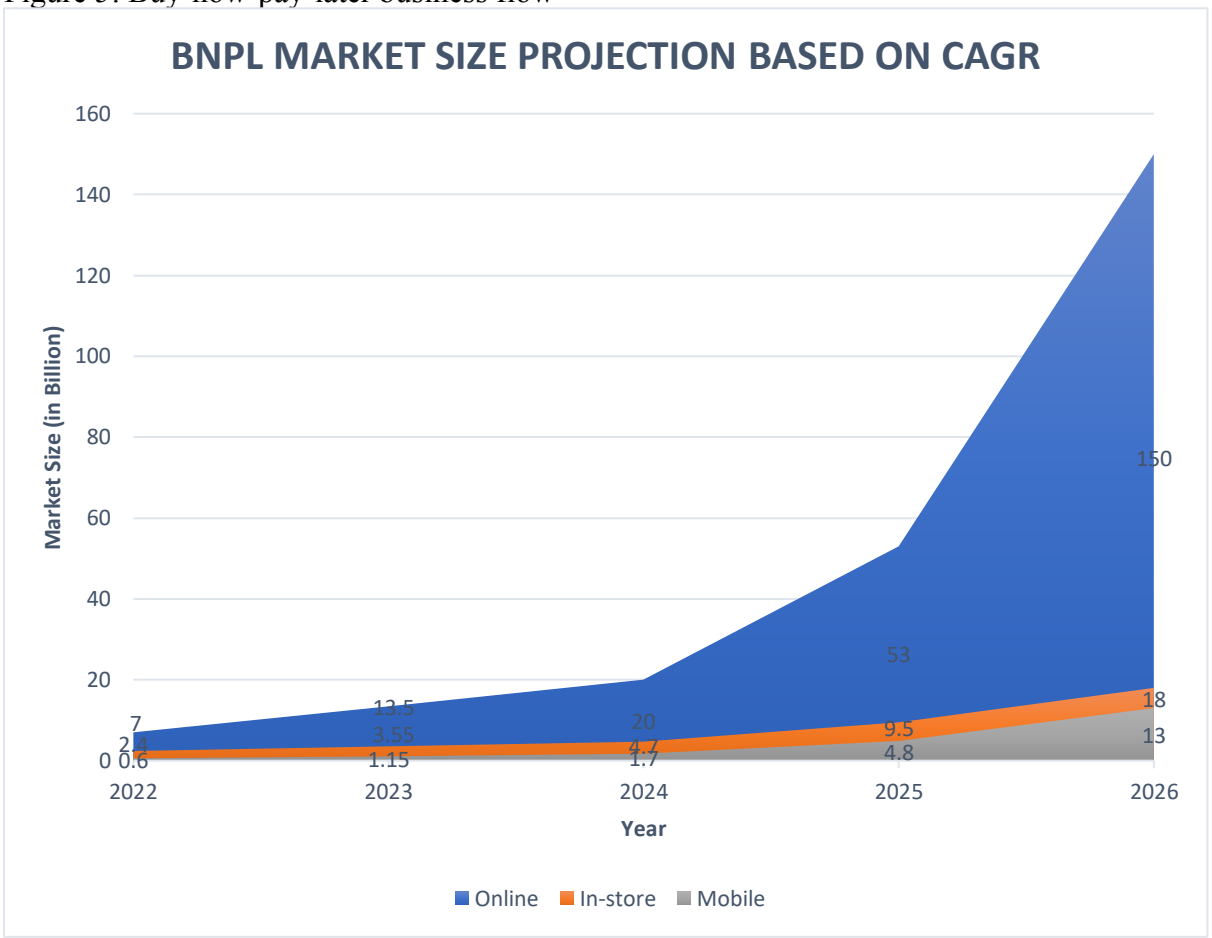


Figure 6: Potential growth of BNPL in Online, In-store, and Mobile

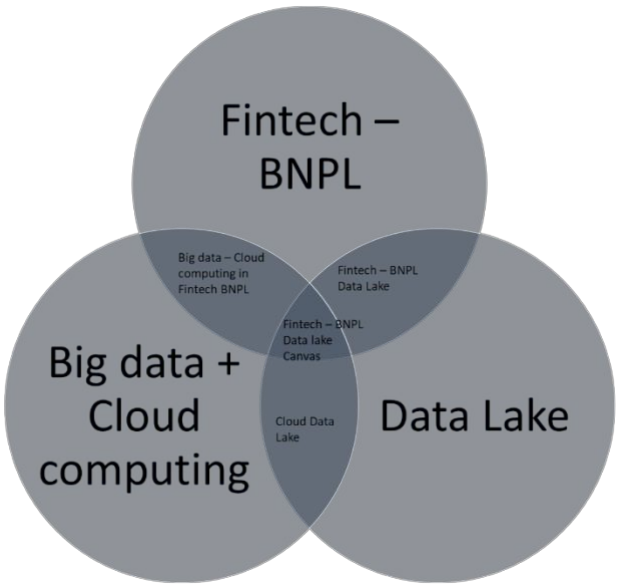


Figure 7: Fintech data lake – Connect the Dots

APPENDIX C

INFORMED CONSENT FORM

I, ……. agree to be interviewed for the research which will be conducted by

……...a doctorate students at the Swiss School of Business and Management, Geneva, Switzerland.

I certify that I have been told of the confidentiality of information collected for this research and the anonymity of my participation; that I have been given satisfactory answers to my inquiries concerning research procedures and other matters; and that I have been advised that I am free to withdraw my consent and to discontinue participation in the research or activity at any time without prejudice.

I agree to participate in one or more electronically recorded interviews for this research. I understand that such interviews and related materials will be kept completely anonymous and that the results of this study may be published in any form that may serve its best.

I agree that any information obtained from this research may be used in any way thought best for this study.

…… ……

Signature of Interviewee Date

APPENDIX D

QUESTIONAIRE WITH CONSENT FORM

I am Durga Rathinasamy, pursuing Doctorate in Business Administration on Fintech Industry challenges and strategies to explore for developing an effective strategical framework with Data Lake.

You're one of the industry (or Fintech) / data engineering leader identified for the survey. Kindly request your support in completing this research survey aimed at collecting data towards building the strategic framework.

Thank you in advance for your valuable time and effort. Your details are completely confidential and will be destroyed once the analysis is complete. For any clarification, please drop your query to durga1@ssbm.ch.

Thanks,

Durga R

- Survey with Industry (or Fintech) Leaders - Study on Successful Strategy for Fintech Data lake Framework

Q.No	Factor Type	Question	Expected Responses (Can be more than one)
1	Fintech Challenge	What are the critical Fintech Business Challenges?	1. Data Security 2. Competition 3. Compliance and Regulation 4. Customer Retention and Customer

			Experience 5. Service Personalization 6. Data and AI integration 7. Blockchain Integration 8. Others
2	BNPL challenge	What are the critical **BNPL** challenges?	1. Risk Management 2. Fraud Protection 3. Current economic/geopolitical / Inflation factors 4. Customer Acquisition 5. Customer Retention 6. Other
3	BNPL challenge	What are the focus areas for Risk Management in BNPL?	1. Fraud Protection 2. Risk Scoring 3. Interest Rates

			4. Missed payments 5. Increased debt for FinTech 6. Risk of debt spirals across multiple BNPL providers 7. Other
4	BNPL challenge	What are Key factors in the BNPL customer retention journey?	1. Customer loyalty 2. Customer satisfaction 3. Customer referrals 4. Merchant brand exposure 5. Improved efficiency in Managing returns 6. Other
5	BNPL challenge	Key Business subject data of BNPL are	1. Payments & Schedule 2. Customer 3. Scoring 4. Risk Levels &

			category 5. Product Catalog of BNPL 6. Financial Calculation 7. Limits 8. Master data & Metadata 9. Other

- Survey with Data Engineering experts - Study on Successful Strategy for Fintech Data lake Framework

Q.No	Factor Type	Question	Expected Responses (Can be more than one)
1	Data lake architectural / engineering factors	Your preference for BNPL data lake	1. On-prem Data lake 2. On-prem Datawarehouse 3. Private cloud data lake 4. Public cloud data

			lake 5. Hybrid cloud data lake 6. Multi-cloud data lake 7. Cloud data lake based on service offering (IaaS, PaaS, SaaS, etc.)
2	Data lake architectural / engineering factors	Preferred scalable architectures for data lake	1. Kappa Architecture (For speed layer) 2. Data mesh architecture (For decentralized data lake) 3. Dynamo Architecture - Distributed Hash Table (DHT) (For columnar, Key-value store) 4. GFS / HDFS

			Architecture (For distributed file system) 5. Event-driven Architecture (For asynchronous messaging) 6. Microservices Architecture (for Data API) 7. Chubby Architecture (For locking service) 8. Data warehousing architecture
3	Data lake architectural / engineering factors	Which DSA works best for handling data in distributed systems	1. SST - Sorted String Table 2. LSM - Log-Structured Merge 3. B+ 4. B Trees 5. Memtable

			6. I prefer going with what is offered by the distributed file system 7. Other
4	Data lake architectural / engineering factors	Structured / semi-structured data lake is easy to source, maintain and manage.	Rating 1 to 10
5	Data lake architectural / engineering factors	Unstructured data lake is easy to source, maintain and manage	Rating 1 to 10
6	Data lake architectural / engineering factors	Underlying data structure and the algorithm are important for data lake design.	Rating 1 to 10
7	Data lake architectural / engineering factors	Preferred data model for BNPL data lake	1. Relational model 2. Document model 3. Graph model 4. Polyglot / Multi-

			model 5. Other
8	Data lake ML factors	What are the critical elements from Data lake for effective Al/ML?	1. Tools and Technology 2. Self-service Business Intelligence 3. ML pipelines 4. Feature Store 5. Expected sample and population of data 6. Data Quality 7. Data Discovery 8. Data Integration 9. Other
9	Data lake ML factors	What are BNPL Feature Stores expected in Data lake?	1. Customer 360 feature store 2. Merchant 360 feature store 3. Finance feature store

			(Book-keeping, financial metrics, etc.) 4. Payments / Transactions feature store 5. Product & Pricing feature store 6. Other

- Influencing factors for data lake canvas from questionnaire

Q.No	Factor Type	Question	Expected Responses (Can be more than one)
1	Cost Management	BNPL cost increase is due to	1. Risk Management 2. Fraud Protection 3. Current economic/geopolitical / Inflation factors 4. Customer Acquisition & Retention 5. Customer support

			6. Technology & Tools 7. Operational Expense (OpEx) 8. Other
2	Cost Management	BNPL's cost-effective is due to	1. Proper Risk Management 2. Favorable economic/geopolitical / Inflation factors 3. Good Customer Acquisition & Retention strategy 4. Excellent Customer support 5. Availability of Technology & Tools 6. Controlled Operational Expense (OpEx) 7. Other
3	Cost Management	How to measure the cost for Data lake?	1. As provided by the cloud / on-prem provider 2. People resources

			3. License cost for Tools / Technologies 4. Other
4	Cost Management	With effective data management and governance, will the OpEx cost increase or decrease?	1. Increase by 5 - 10% 2. Increase by 10 - 20% 3. Increase by 20 - 40% 4. Decrease by 5 - 10% 5. Decrease by 10 - 20% 6. Decrease by 20 - 40% 7. Other
5	Time to Value	BNPL Data - Reasons for the delay in Time to Value	1. Delay / missing Data pipeline setup 2. Delay/missing Data Analytics & Insights 3. Improper AI/ML approaches 4. Other
6	Time to Value	The lack of SME is due to	1. Data Engineers for building data-intensive applications

			2. Data Analysts in BNPL 3. Data Architects for breaking monolithic and centralized architectures 4. Data Scientists for niche BNPL skill 5. Other
7	Data Security	BNPL data security is at stake due to	1. Data leaks 2. Fraud transactions by impersonating lenders 3. Hackers 4. Poor/missing data infrastructure
8	Data Quality	Significant areas causing BNPL Data quality issues	1. Payments & Schedule 2. Customer 3. Scoring 4. Risk Levels & category 5. Product Catalog of BNPL 6. Financial Calculation

			7. Limits 8. Master data & Metadata 9. Other
9	Data Quality	With effective data quality and data engineering practices, will garbage dump increase or decrease?	1. Increase by 5 - 10% 2. Increase by 10 - 20% 3. Increase by 20 - 40% 4. Decrease by 5 - 10% 5. Decrease by 10 - 20% 6. Decrease by 20 - 40%
10	Time to Market	Effort required to build data lake from scratch	1. Startup - < 2 months 2. Startup - > 2 months 3. SMB - 3 - 4 months 4. SMB - 6 months 5. Enterprise > 6 months 6. Enterprise > 12 months
11	Time to Market	Effort required to migrate data lake to new tech stack	1. Startup - < 1 month 2. SMB < 2 months

		without modernization	3. Enterprise < 3 months 4. Other
12	Time to Market	Effort required to migrate data lake to new tech stack with modernization	1. Startup - < 4 months 2. SMB - 6 - 8 months 3. Enterprise > 12 months
13	Data Governance	BNPL Data is valuable and usable with	1. Source-aligned data (raw data without any transformation) 2. Consumer-aligned data (fit-for-purpose data) 3. Aggregate-domain data (OLAP - analytical mart) 4. Data discovery & Data Catalog Management 5. Data Privacy 6. Other
14	Data Governance	Federated governance is the key to a distributed architecture, and the	1. Standards as Code 2. Policies as code 3. Automated Tests

		key governance policy is	4. Automated Monitoring 5. Data Privacy 6. Other
15	Data Governance	Reasons for Garbage dump in BNPL data lake	1. Building data lake just for the sake of it with all sources of data 2. Never transforming the raw data into meaningful insights 3. Redundant data 4. Duplicate data 5. Missing data governance 6. Not interpreting business value from data
16	Data Quality	% of data quality issues in data lake	1. 0-20% 2. 20-40% 3. >40% 4. Data quality is not measured though implemented

			5. Other
17	Data Security	BNPL data security is at stake due to	1. Data leaks 2. Fraud transactions by impersonating lenders 3. Hackers 4. Poor/missing data infrastructure 5. Other
18	Data Security	BNPL security drivers are	1. Best encryption/decryption protocols 2. On-prem data lake solution 3. Private Cloud data lake 4. Public / Multi-Cloud data lake with security on clusters, nodes, files, and data attributes 5. Other
19	Data Security	% of data security issues in data lake	1. 0-1% 2. 2-5%

			3. 5-10% 4. Data Security is not measured though implemented 5. Other

APPENDIX E

INTERVIEW QUESTIONS - EVALUATION ON DATA LAKE CANVAS

Name		**Job Title**	
Industry		**Role**	

What is the new response for the below survey question considering the data lake canvas

[1] With effective data management and governance, will the opex cost increase or decrease?
[2] With effective data quality and data engineering practices, will garbage dump increase or decrease?
[3] Effort required to build data lake from scratch
[4] Effort required to migrate data lake to new tech stack without modernization
[5] % of data quality issues in data lake
[6] % of data security issues in data lake

APPENDIX F

RESPONDENT RESULTS FOR FINTECH – BNPL INDUSTRY CHALLENGES

Table 12: Survey participants – Industry and Role

Survey	**Role**
Fintech Survey	Data Leader
	CTO / CXO
	Engineering Leader
	Product Leader
	IT / Technology Leader
Data Engineering Survey	Data Engineer
	Data Analyst
	Data Architect
	Engineering Manager
	Data scientist
	Product Manager
	Data Leader

Table 13: Survey questionnaire – Fintech and BNPL Industry view on Data lake

Q.No	**Factor Type**	**Question**	**Expected Responses (Can be more than one)**
1	Fintech Challenge	What are the critical Fintech Business Challenges?	1. Data Security 2. Competition

			3. Compliance and Regulation 4. Customer Retention and Customer Experience 5. Service Personalization 6. Data and AI integration 7. Blockchain Integration 8. Others
2	BNPL challenge	What are the critical **BNPL** challenges?	1. Risk Management 2. Fraud Protection 3. Current economic/geopolitical / Inflation factors 4. Customer Acquisition 5. Customer Retention 6. Other
3	BNPL challenge	What are the focus areas for Risk Management in BNPL?	1. Fraud Protection 2. Risk Scoring 3. Interest Rates 4. Missed payments 5. Increased debt for

			FinTech 6. Risk of debt spirals across multiple BNPL providers 7. Other
4	BNPL challenge	What are Key factors in the BNPL customer retention journey?	1. Customer loyalty 2. Customer satisfaction 3. Customer referrals 4. Merchant brand exposure 5. Improved efficiency in Managing returns 6. Other
5	BNPL challenge	Key Business subject data of BNPL are	1. Payments & Schedule 2. Customer 3. Scoring 4. Risk Levels & category 5. Product Catalog of BNPL 6. Financial Calculation 7. Limits 8. Master data & Metadata 9. Other

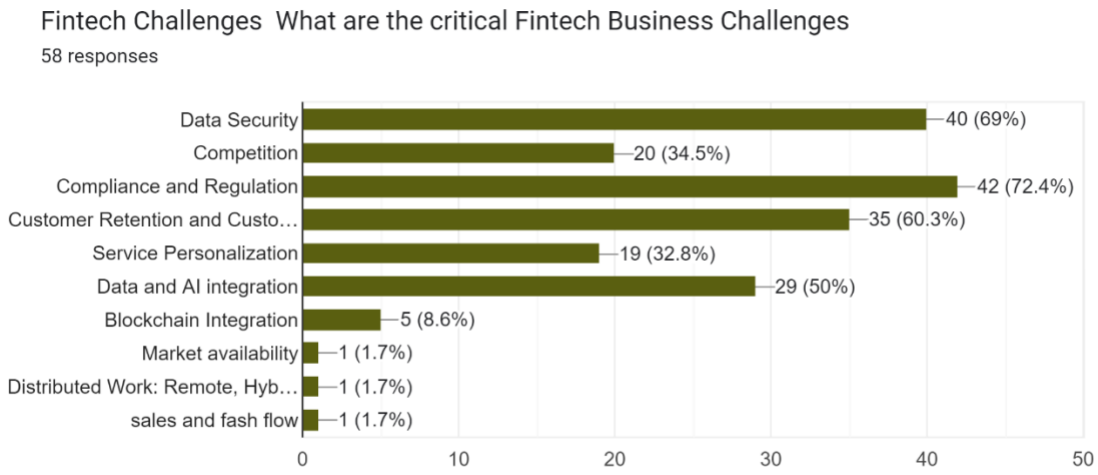


Figure 8: Fintech business challenges

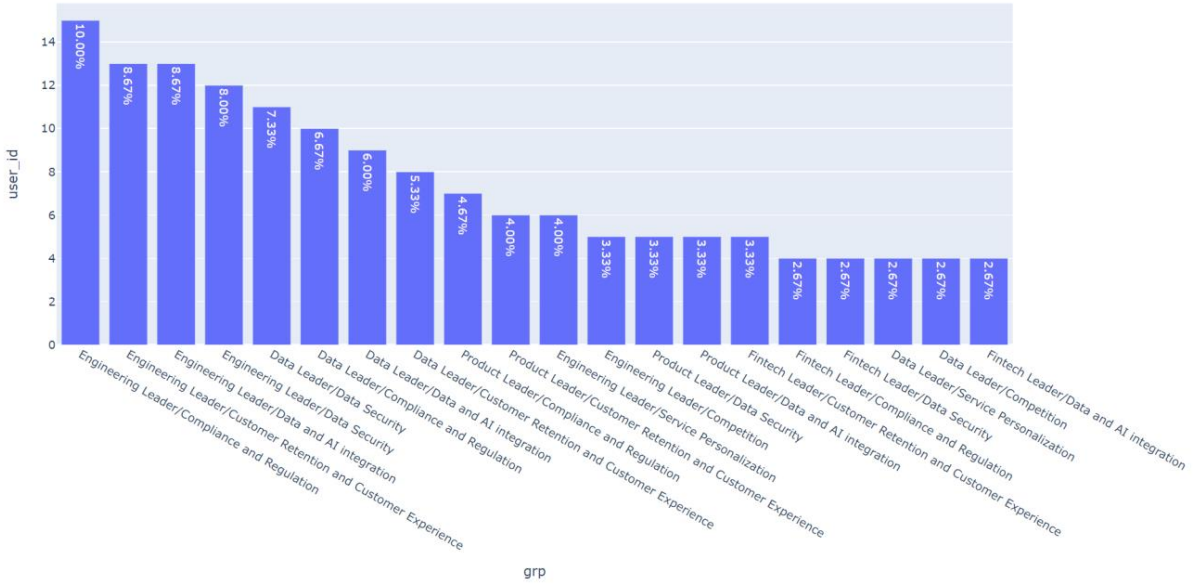


Figure 9: Fintech challenges grouped by role

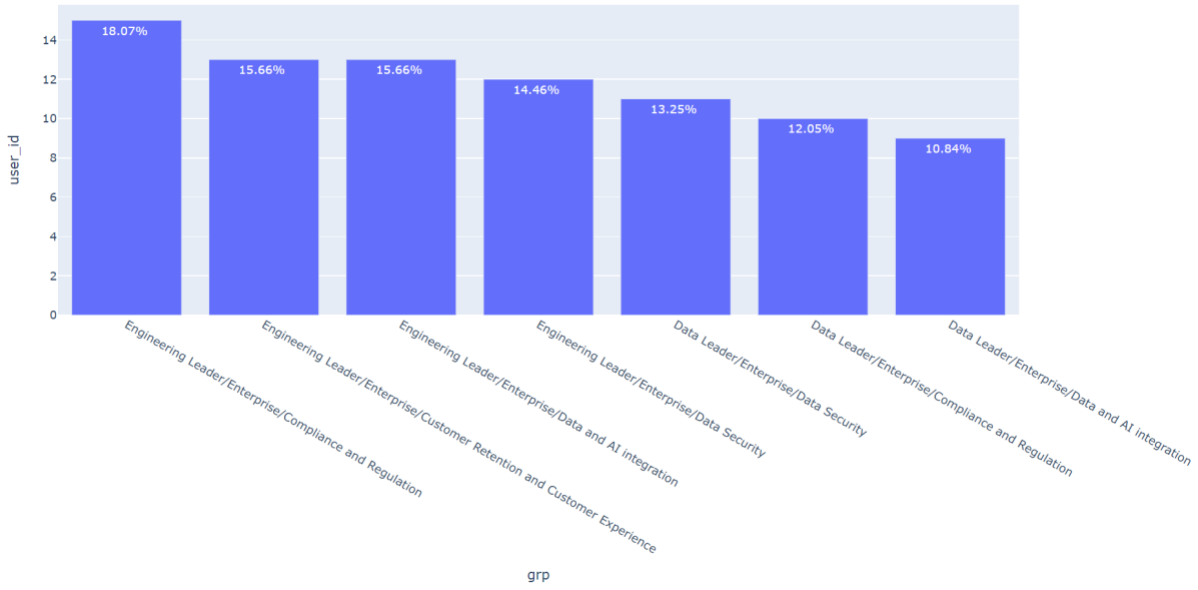


Figure 10: Fintech challenges grouped by role and organization type

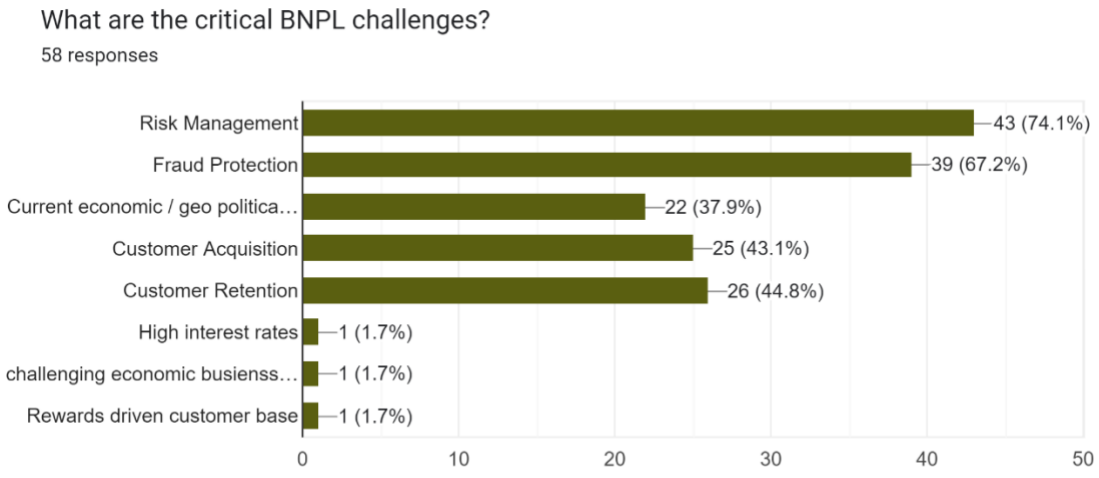


Figure 11: BNPL challenges

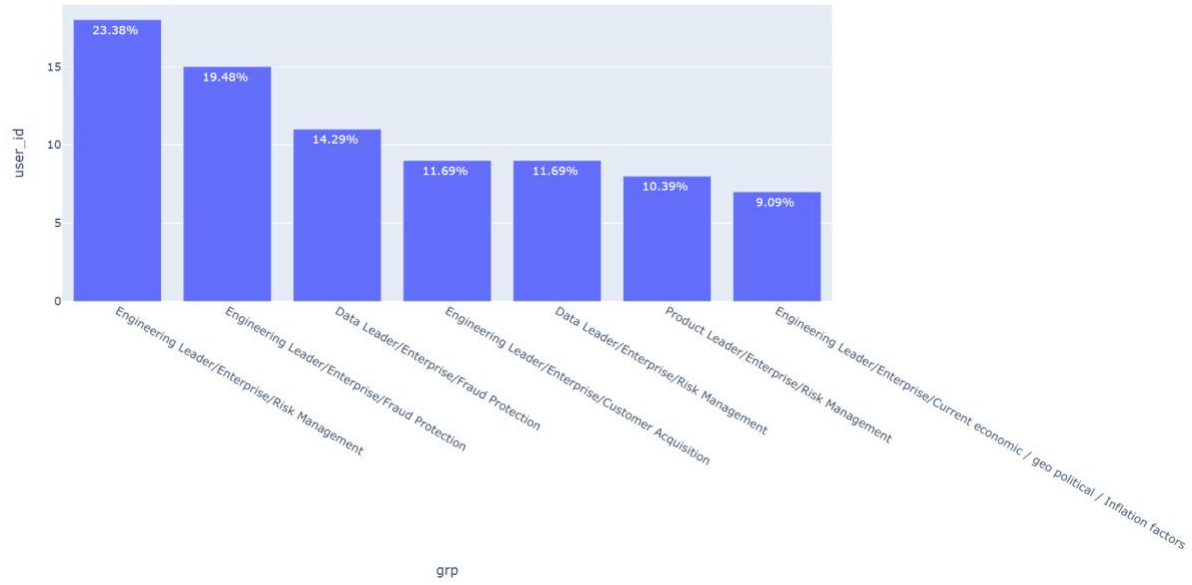


Figure 12: BNPL challenges grouped by role and organization type

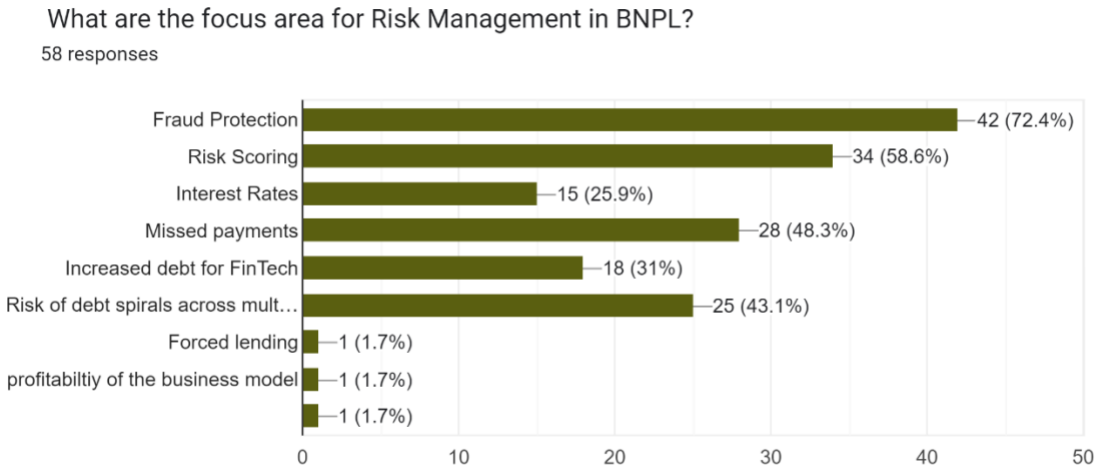


Figure 13: Risk management factors grouped by Role and Organization type

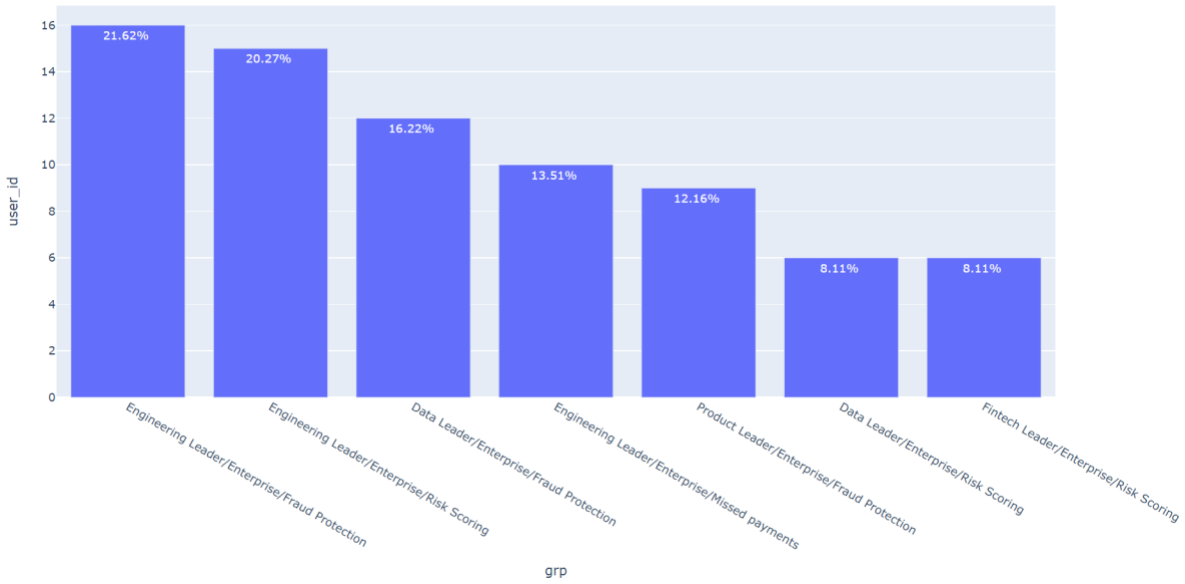


Figure 14: Risk management factors grouped by Role and Organization type

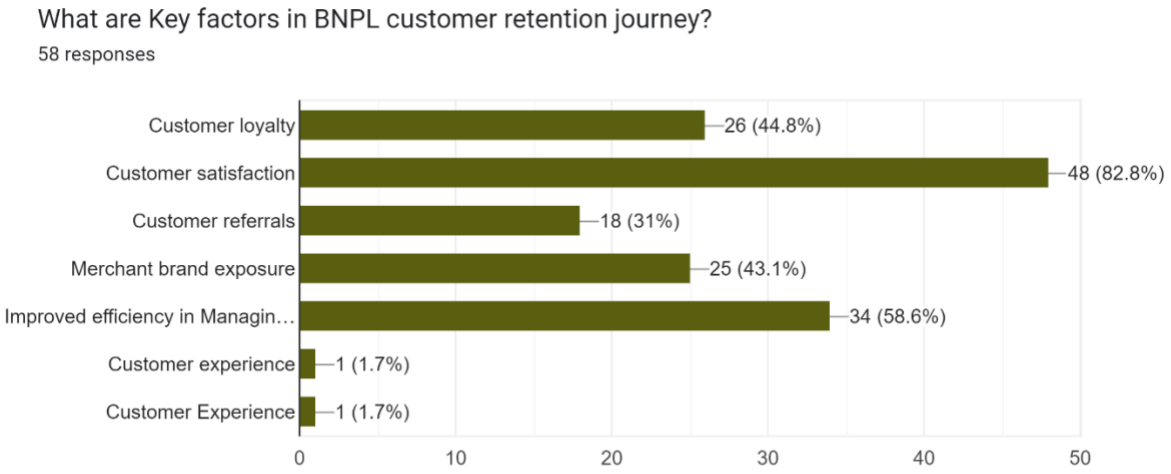


Figure 15: Customer journey factors

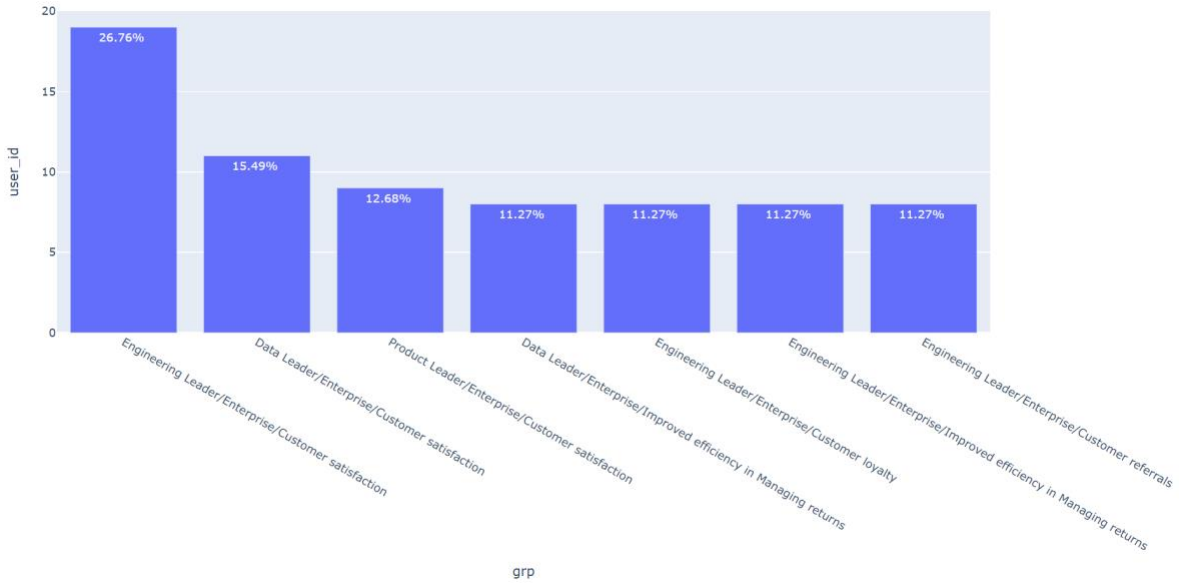


Figure 16: Customer journey factors grouped by Role and Organization type

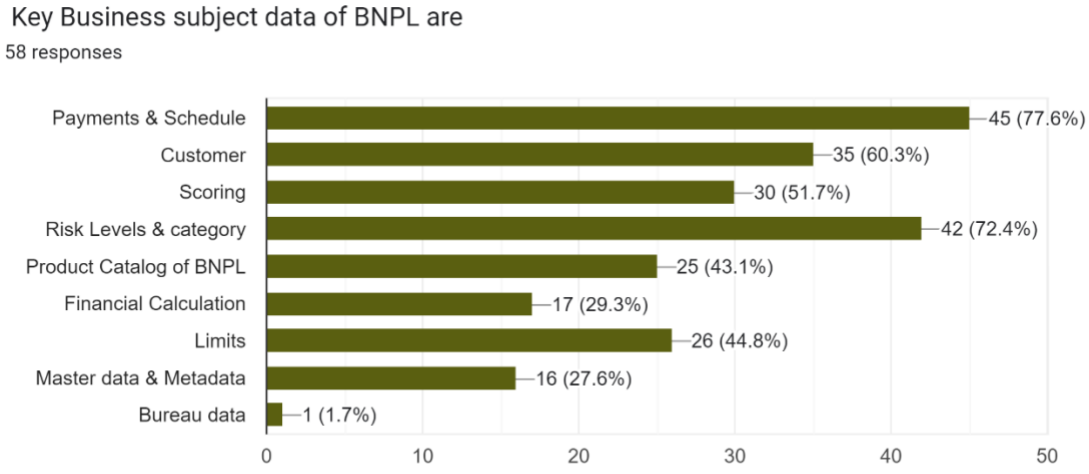


Figure 17: BNPL subject areas

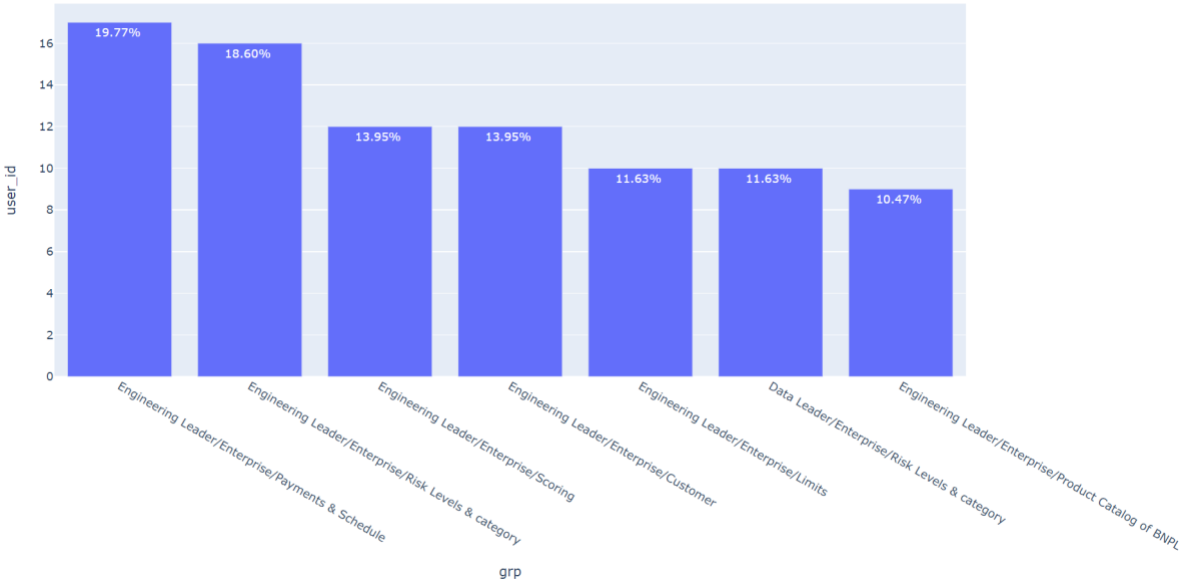


Figure 18: BNPL subject areas grouped by Role and Organization type

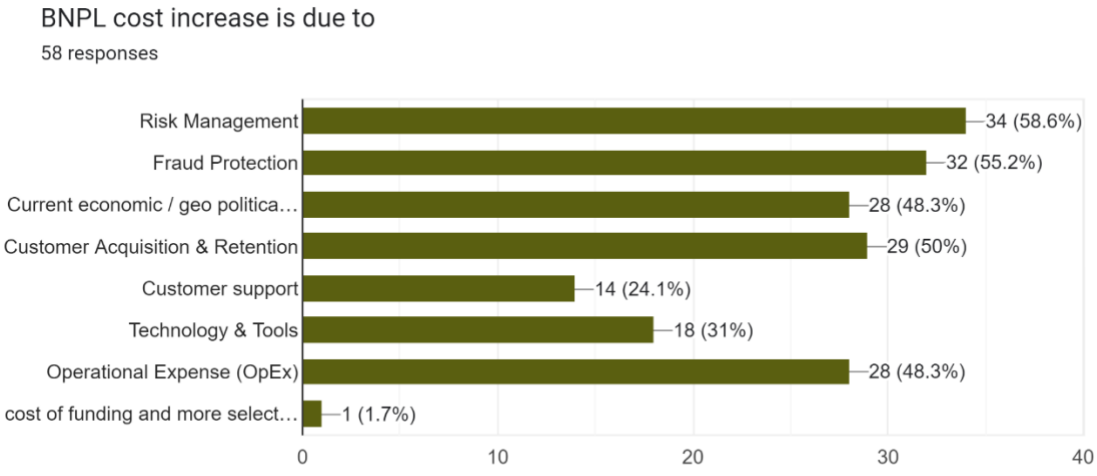


Figure 19: BNPL cost increase

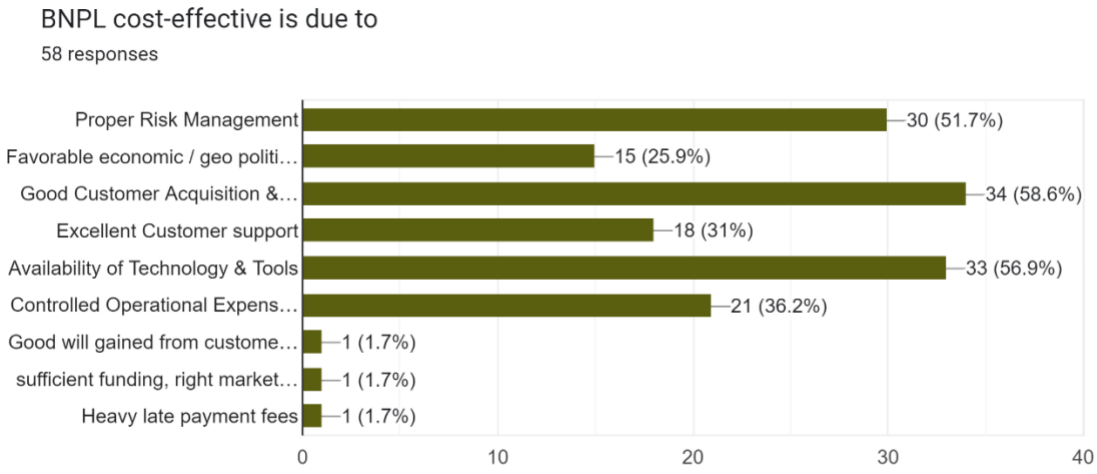


Figure 20: BNPL cost-effectiveness

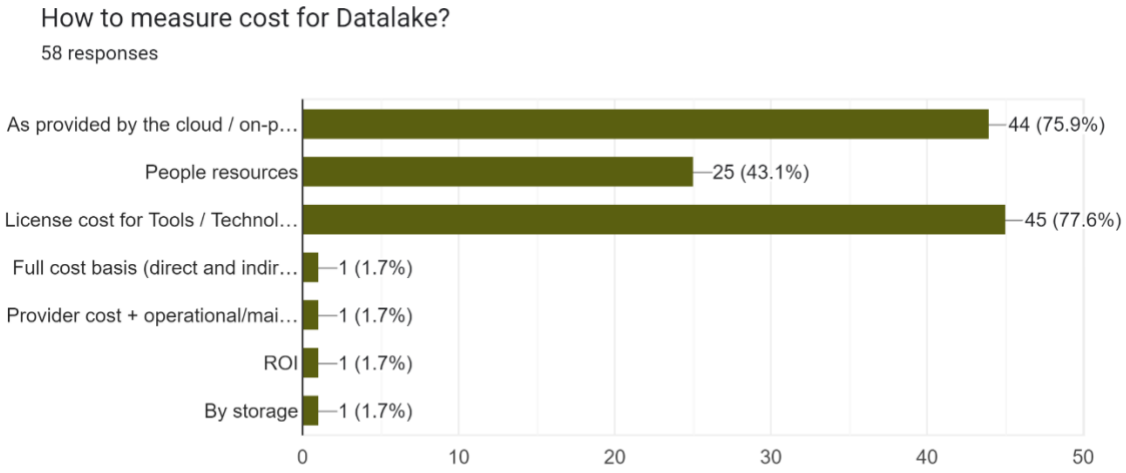


Figure 21: BNPL data lake measuring factors

APPENDIX G

RESPONDENT RESULTS FROM ENGINEERING EXPERTS ON SUCCESSFUL STRATEGY FOR FINTECH - BNPL DATA LAKE FRAMEWORK

Table 14: Survey questionnaire – Data engineering view on Data lake

Q.No	Factor Type	Question	Expected Responses (Can be more than one)
1	Data lake architectural / engineering factors	Your preference for BNPL data lake	1. On-prem Data lake 2. On-prem Datawarehouse 3. Private cloud data lake 4. Public cloud data lake 5. Hybrid cloud data lake 6. Multi-cloud data lake 7. Cloud data lake based on service offering (IaaS, PaaS, SaaS, etc.)
2	Data lake architectural /	Preferred scalable architectures for data	1. Kappa Architecture (For speed layer)

engineering factors	lake	2. Data mesh architecture (For decentralized data lake) 3. Dynamo Architecture - Distributed Hash Table (DHT) (For columnar, Key-value store) 4. GFS / HDFS Architecture (For distributed file system) 5. Event-driven Architecture (For asynchronous messaging) 6. Microservices Architecture (for Data API) 7. Chubby Architecture (For locking service) 8. Data warehousing architecture

3	Data lake architectural / engineering factors	Which DSA works best for handling data in distributed systems	1. SST - Sorted String Table 2. LSM - Log-Structured Merge 3. B+ 4. B Trees 5. Memtable 6. I prefer going with what is offered by the distributed file system 7. Other
4	Data lake architectural / engineering factors	Structured / semi-structured data lake is easy to source, maintain and manage.	Rating 1 to 10
5	Data lake architectural / engineering factors	Unstructured data lake is easy to source, maintain and manage	Rating 1 to 10
6	Data lake architectural / engineering factors	Underlying data structure and the algorithm are important for data lake design.	Rating 1 to 10
7	Data lake	Preferred data model for	1. Relational model

	architectural / engineering factors	BNPL data lake	2. Document model 3. Graph model 4. Polyglot / Multi-model 5. Other
8	Data lake ML factors	What are the critical elements from Data lake for effective Al/ML?	1. Tools and Technology 2. Self-service Business Intelligence 3. ML pipelines 4. Feature Store 5. Expected sample and population of data 6. Data Quality 7. Data Discovery 8. Data Integration 9. Other
9	Data lake ML factors	What are BNPL Feature Stores expected in Data lake?	1. Customer 360 feature store 2. Merchant 360 feature store 3. Finance feature store (Book-keeping, financial metrics, etc.)

4. Payments / Transactions feature store
5. Product & Pricing feature store
6. Other

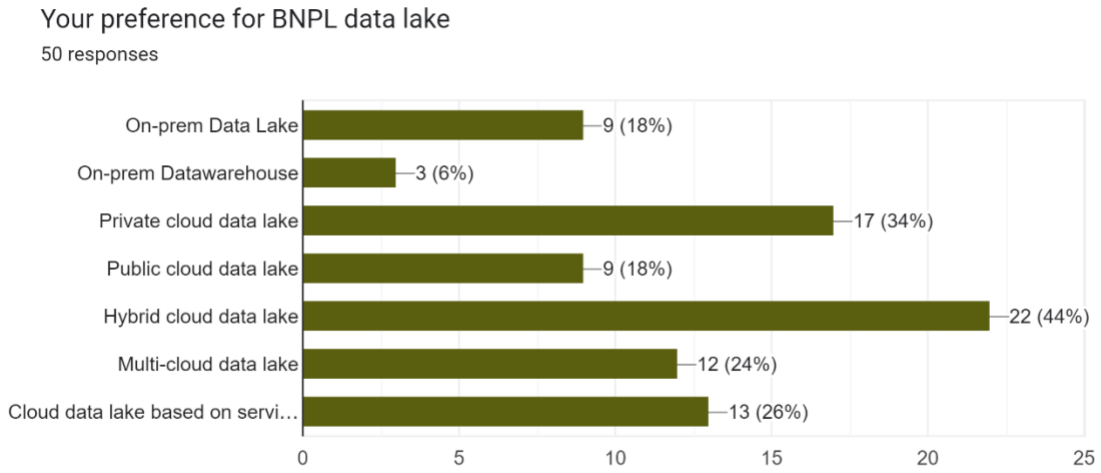


Figure 22: BNPL data lake preference

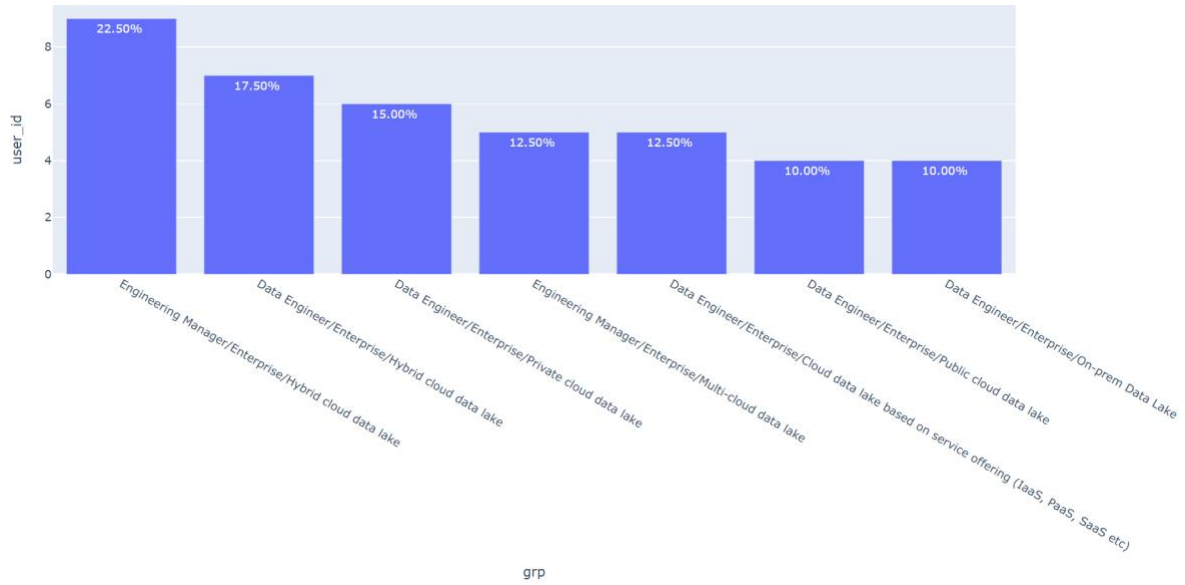


Figure 23: BNPL data lake preference grouped by Role and Organization type

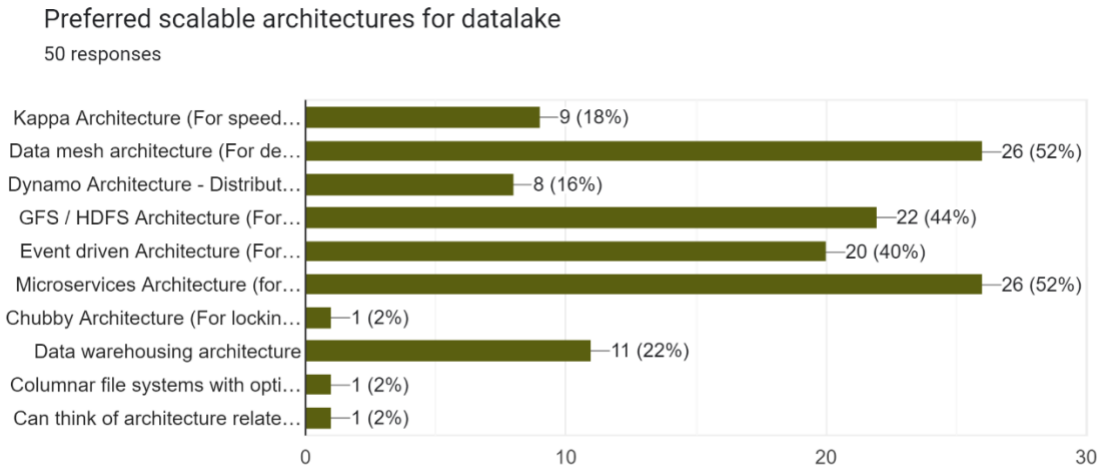


Figure 24: BNPL data lake architecture preference

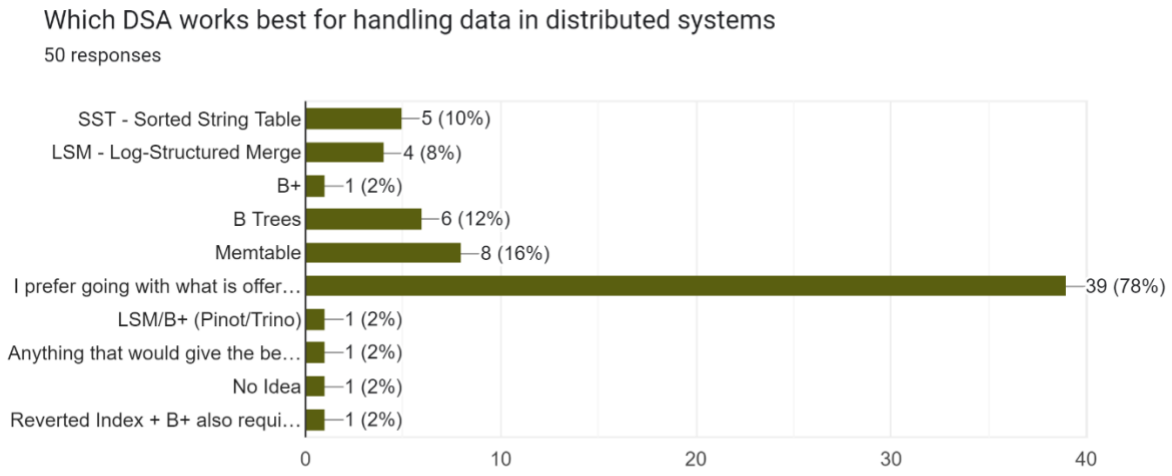


Figure 25: BNPL data lake data structure algorithm preference

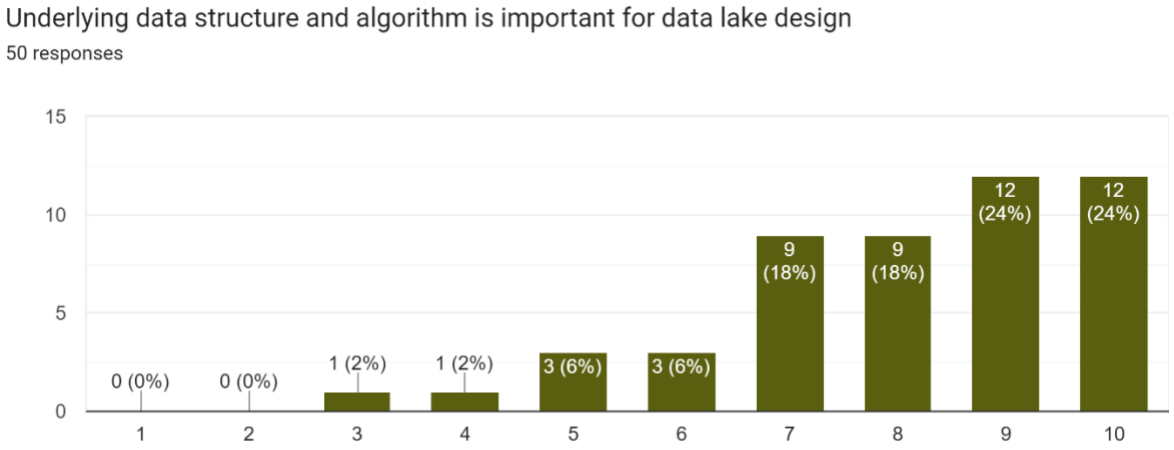


Figure 26: BNPL data lake data structure importance

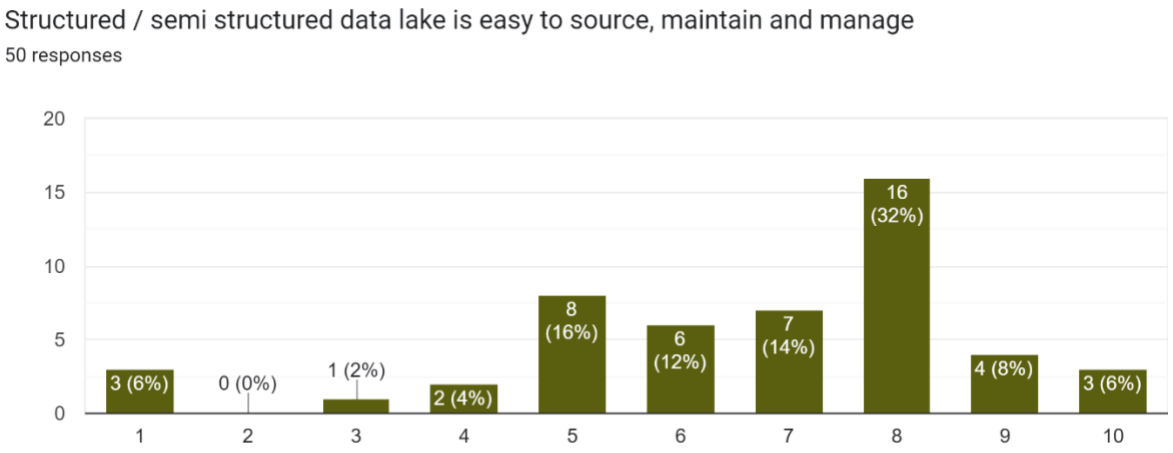


Figure 27: BNPL data lake structured/semi-structured data management

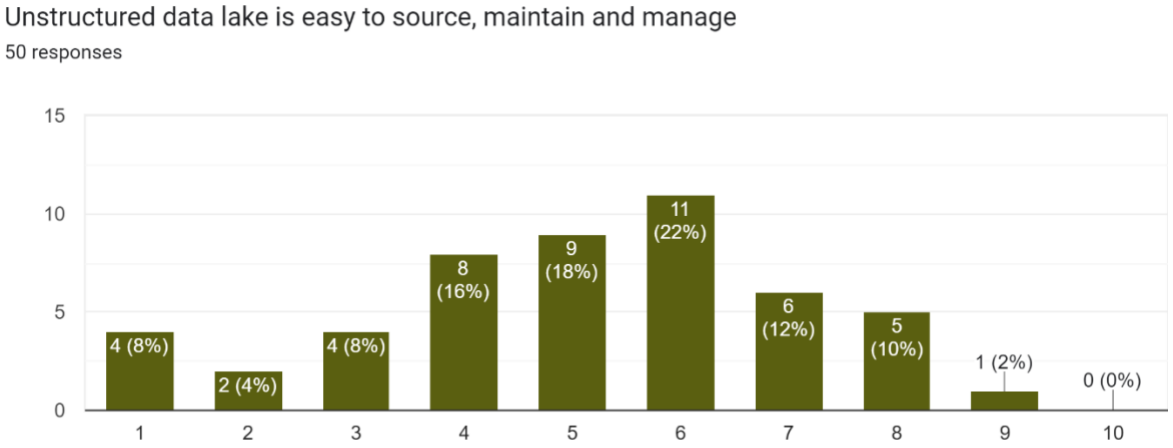


Figure 28: BNPL data lake unstructured data management

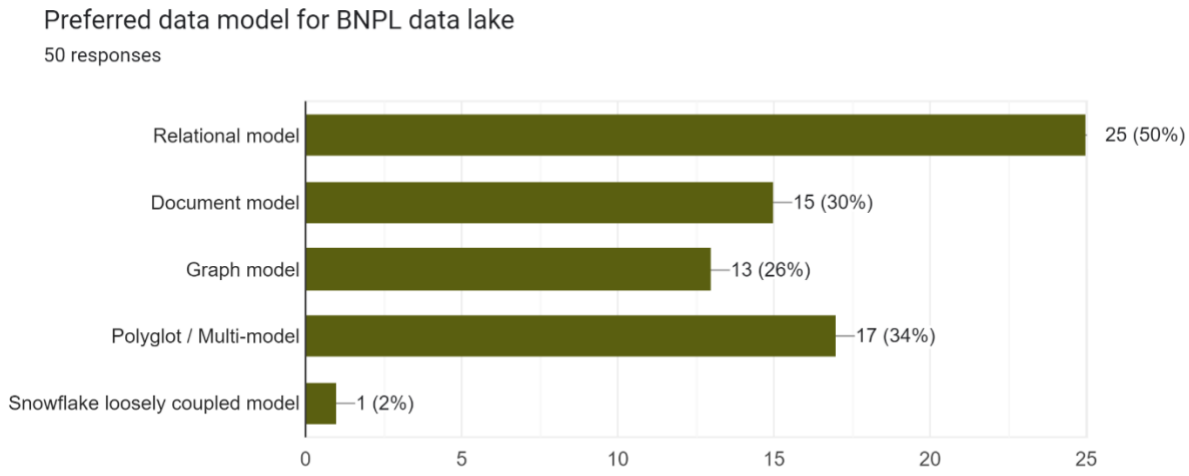


Figure 29: BNPL data lake preferred data model

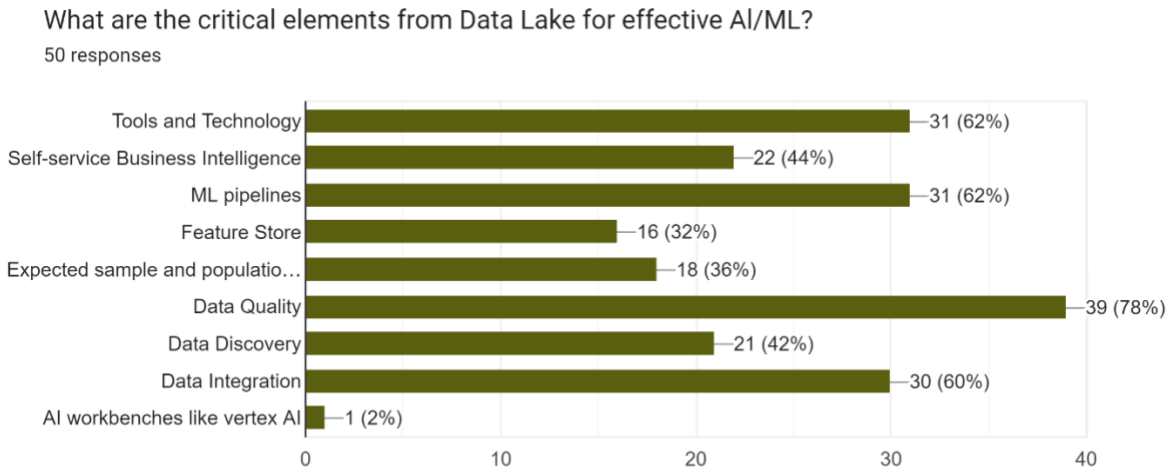


Figure 30: BNPL data lake critical elements for AI/ML

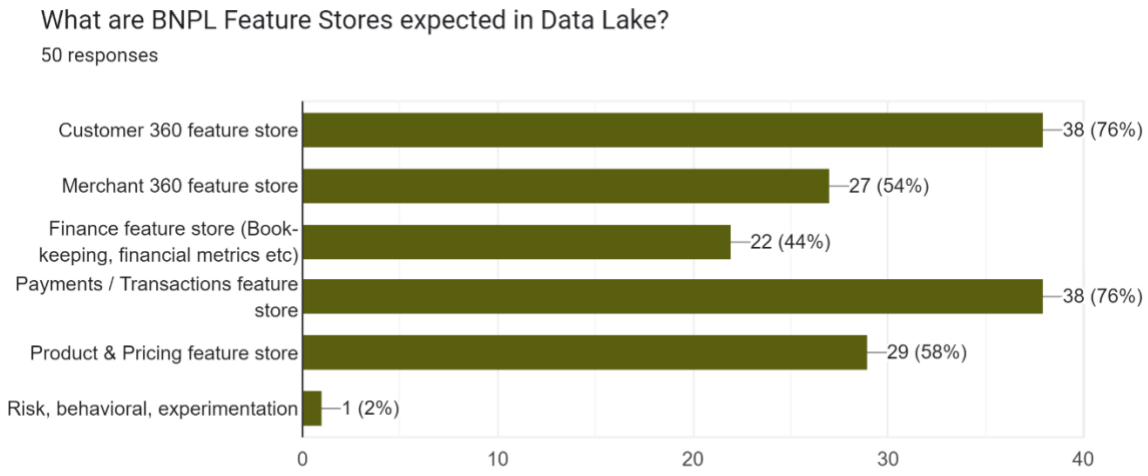


Figure 31: BNPL data lake – Feature stores

APPENDIX H

RESPONDENT RESULTS FOR FINTECH – BNPL DATA LAKE – INFLUENCING FACTORS

Table 15: Survey questionnaire – Influencing factors on Data lake

Q.No	Factor Type	Question	Expected Responses (Can be more than one)
1	Cost Management	BNPL cost increase is due to	1. Risk Management 2. Fraud Protection 3. Current economic/geopolitical / Inflation factors 4. Customer Acquisition & Retention 5. Customer support 6. Technology & Tools 7. Operational Expense (OpEx) 8. Other
2	Cost Management	BNPL's cost-effective is due to	1. Proper Risk Management 2. Favorable economic/geopolitical / Inflation factors 3. Good Customer Acquisition

			& Retention strategy 4. Excellent Customer support 5. Availability of Technology & Tools 6. Controlled Operational Expense (OpEx) 7. Other
3	Cost Management	How to measure the cost for Data lake?	1. As provided by the cloud / on-prem provider 2. People resources 3. License cost for Tools / Technologies 4. Other
4	Cost Management	With effective data management and governance, will the OpEx cost increase or decrease?	1. Increase by 5 - 10% 2. Increase by 10 - 20% 3. Increase by 20 - 40% 4. Decrease by 5 - 10% 5. Decrease by 10 - 20% 6. Decrease by 20 - 40% 7. Other
5	Time to Value	BNPL Data - Reasons for the delay in Time to Value	1. Delay / missing Data pipeline setup 2. Delay/missing Data Analytics & Insights 3. Improper AI/ML approaches

			4. Other
6	Time to Value	The lack of SME is due to	1. Data Engineers for building data-intensive applications 2. Data Analysts in BNPL 3. Data Architects for breaking monolithic and centralized architectures 4. Data Scientists for niche BNPL skill 5. Other
7	Data Security	BNPL data security is at stake due to	1. Data leaks 2. Fraud transactions by impersonating lenders 3. Hackers 4. Poor/missing data infrastructure
8	Data Quality	Significant areas causing BNPL Data quality issues	1. Payments & Schedule 2. Customer 3. Scoring 4. Risk Levels & category 5. Product Catalog of BNPL 6. Financial Calculation 7. Limits 8. Master data & Metadata 9. Other

9	Data Quality	With effective data quality and data engineering practices, will garbage dump increase or decrease?	1. Increase by 5 - 10% 2. Increase by 10 - 20% 3. Increase by 20 - 40% 4. Decrease by 5 - 10% 5. Decrease by 10 - 20% 6. Decrease by 20 - 40%
10	Time to Market	Effort required to build data lake from scratch	1. Startup - < 2 months 2. Startup - > 2 months 3. SMB - 3 - 4 months 4. SMB - 6 months 5. Enterprise > 6 months 6. Enterprise > 12 months
11	Time to Market	Effort required to migrate data lake to new tech stack without modernization	1. Startup - < 1 month 2. SMB < 2 months 3. Enterprise < 3 months 4. Other
12	Time to Market	Effort required to migrate data lake to new tech stack with modernization	1. Startup - < 4 months 2. SMB - 6 - 8 months 3. Enterprise > 12 months
13	Data Governance	BNPL Data is valuable and usable with	1. Source-aligned data (raw data without any transformation) 2. Consumer-aligned data (fit-

			for-purpose data) 3. Aggregate-domain data (OLAP - analytical mart) 4. Data discovery & Data Catalog Management 5. Data Privacy 6. Other
14	Data Governance	Federated governance is the key to a distributed architecture, and the key governance policy is	1. Standards as Code 2. Policies as code 3. Automated Tests 4. Automated Monitoring 5. Data Privacy 6. Other
15	Data Governance	Reasons for Garbage dump in BNPL data lake	1. Building data lake just for the sake of it with all sources of data 2. Never transforming the raw data into meaningful insights 3. Redundant data 4. Duplicate data 5. Missing data governance 6. Not interpreting business value from data
16	Data Quality	% of data quality	1. 0-20%

		issues in data lake	2. 20-40% 3. >40% 4. Data quality is not measured though implemented 5. Other
17	Data Security	BNPL data security is at stake due to	1. Data leaks 2. Fraud transactions by impersonating lenders 3. Hackers 4. Poor/missing data infrastructure 5. Other
18	Data Security	BNPL security drivers are	1. Best encryption/decryption protocols 2. On-prem data lake solution 3. Private Cloud data lake 4. Public / Multi-Cloud data lake with security on clusters, nodes, files, and data attributes 5. Other
19	Data Security	% of data security issues in data lake	1. 0-1% 2. 2-5% 3. 5-10% 4. Data Security is not

measured though implemented 5. Other

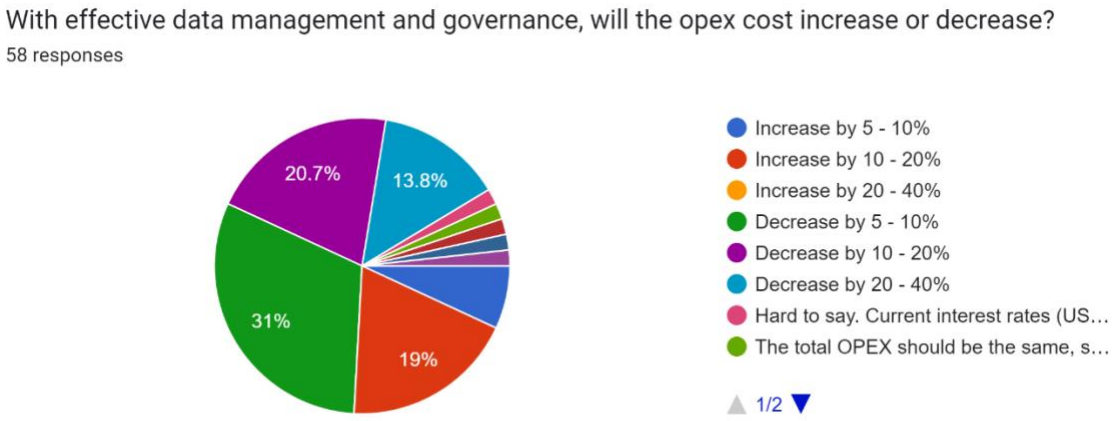


Figure 32: BNPL data lake – Operational cost

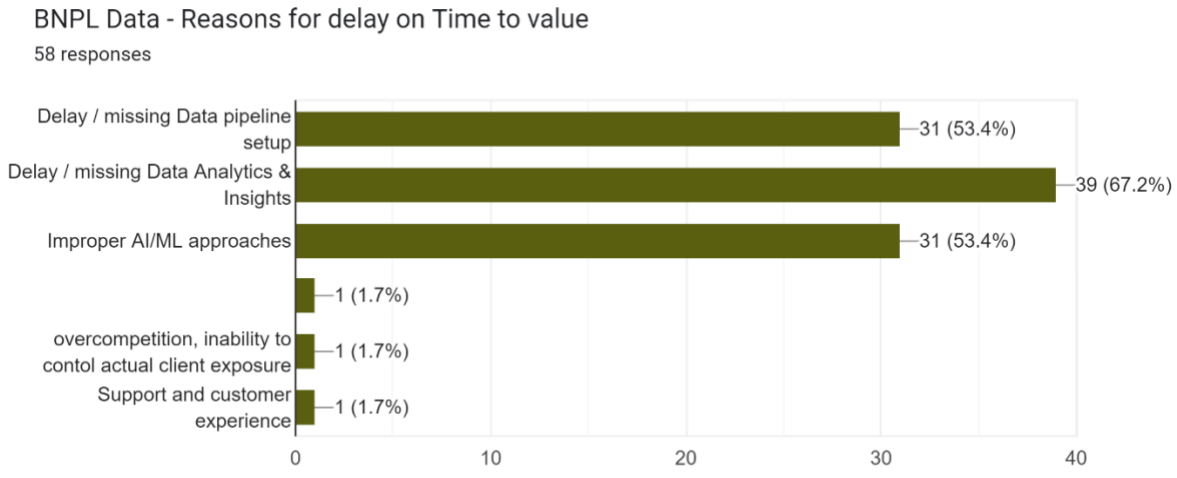


Figure 33: BNPL data lake – Time to value

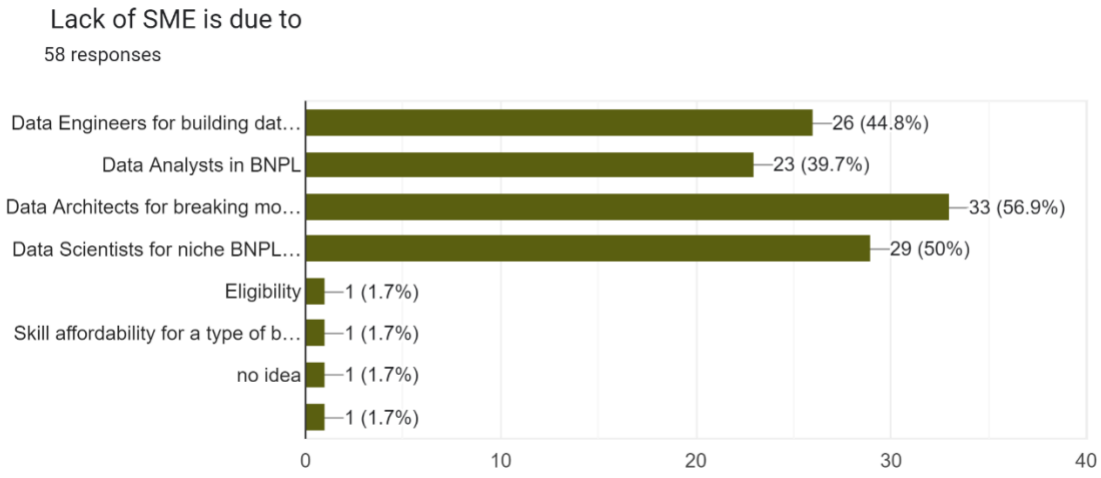


Figure 34: BNPL data lake – Lack of SME

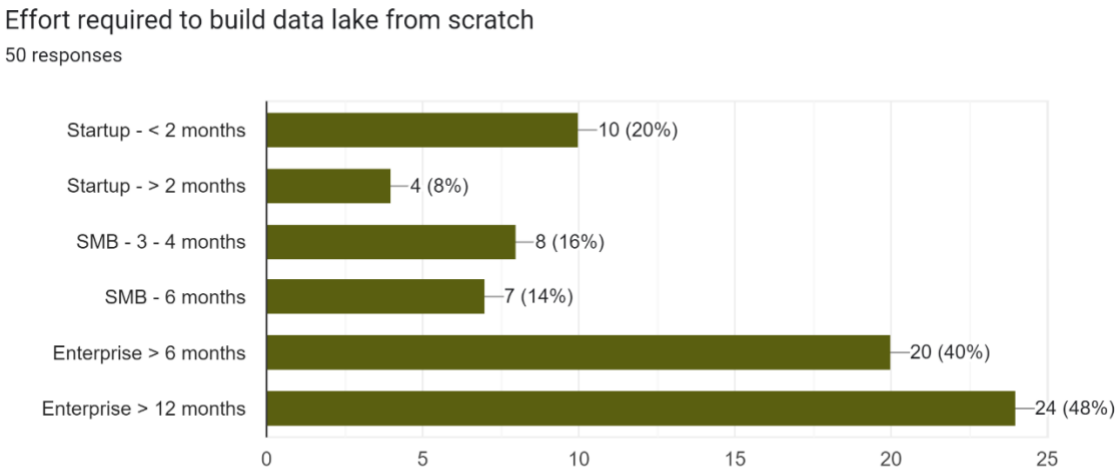


Figure 35: BNPL data lake – TTM – Build from scratch

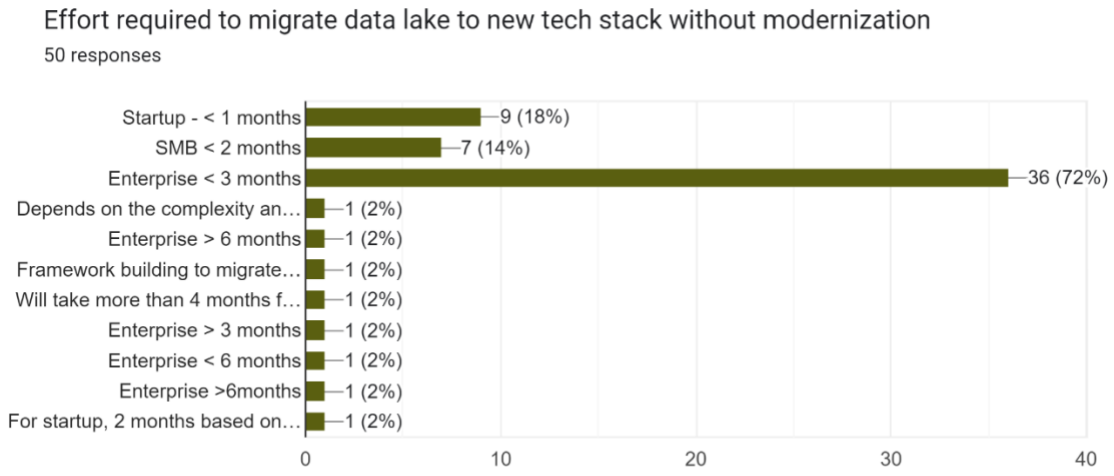


Figure 36: BNPL data lake – TTM – Migrate without modernization

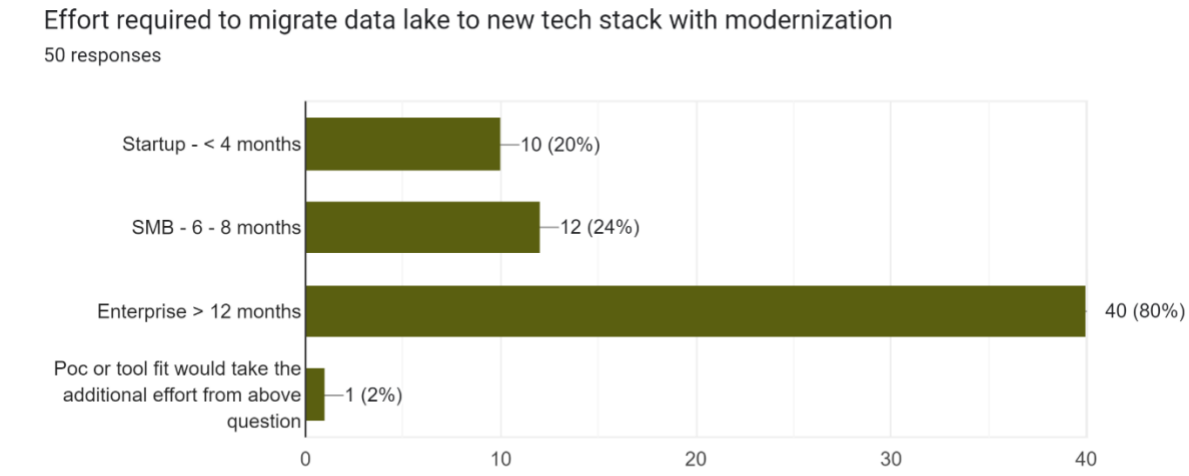


Figure 37: BNPL data lake – TTM – Migrate with modernization

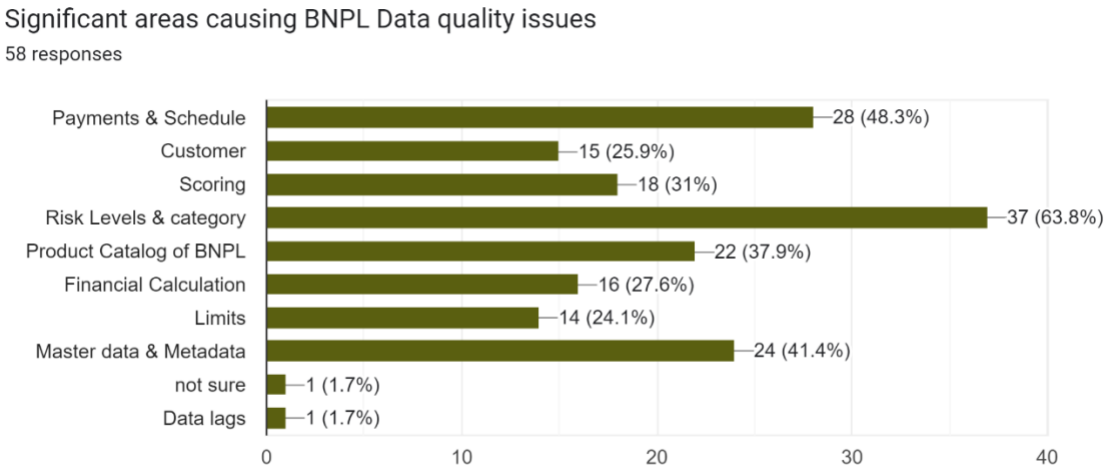


Figure 38: BNPL data lake – Data Quality areas

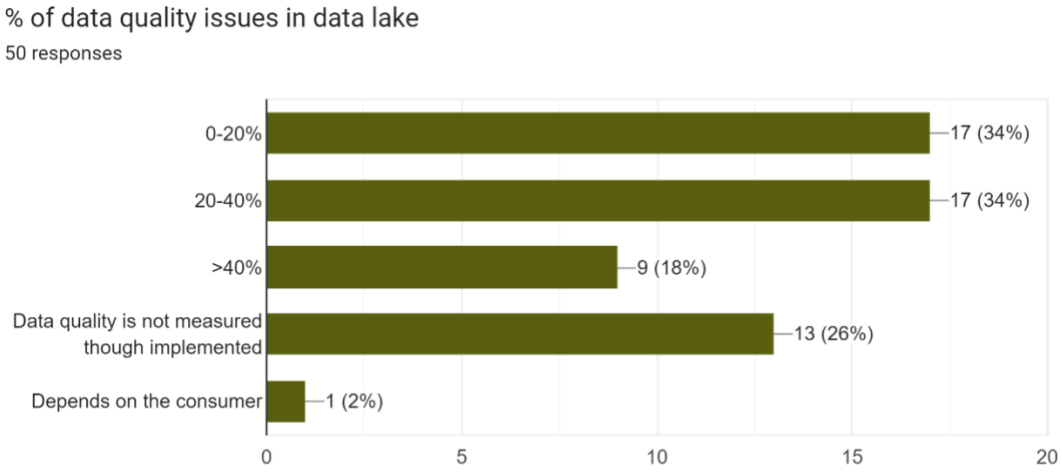


Figure 39: BNPL data lake – Data Quality issues

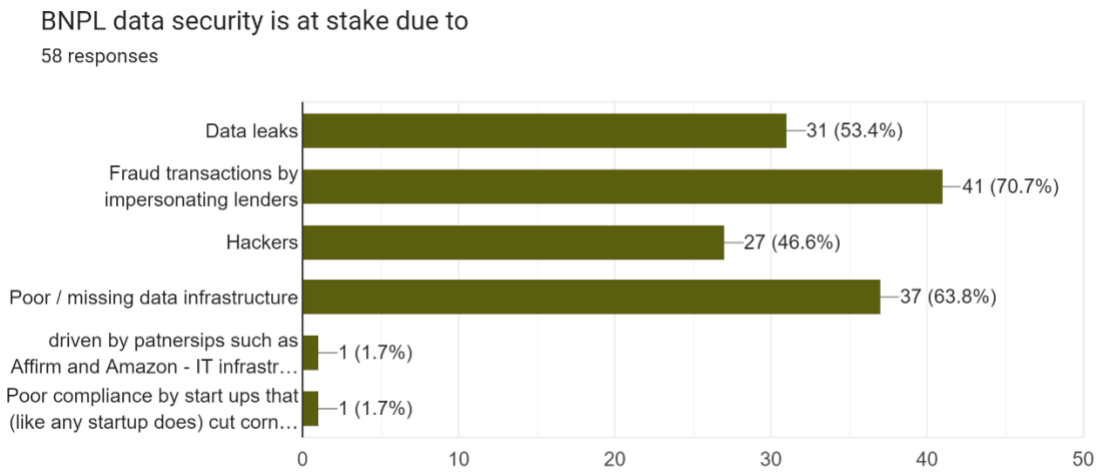


Figure 40: BNPL data lake – Data Security view by Fintech experts

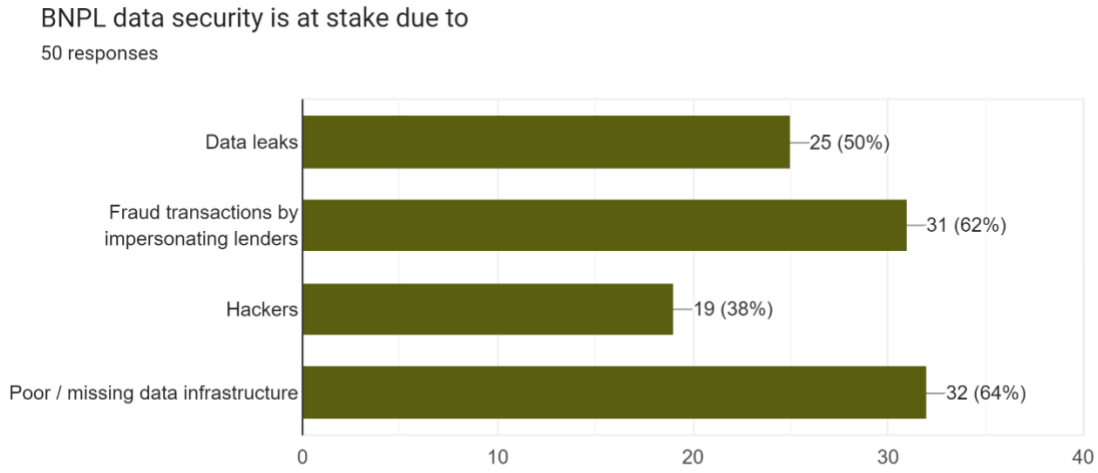


Figure 41: BNPL data lake – Data Security view by Data engineering experts

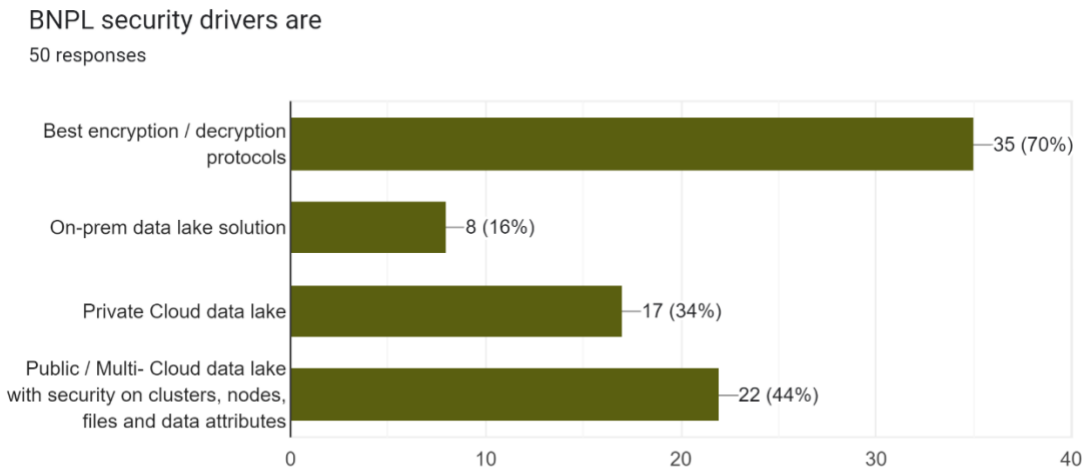


Figure 42: BNPL data lake – Data Security drivers

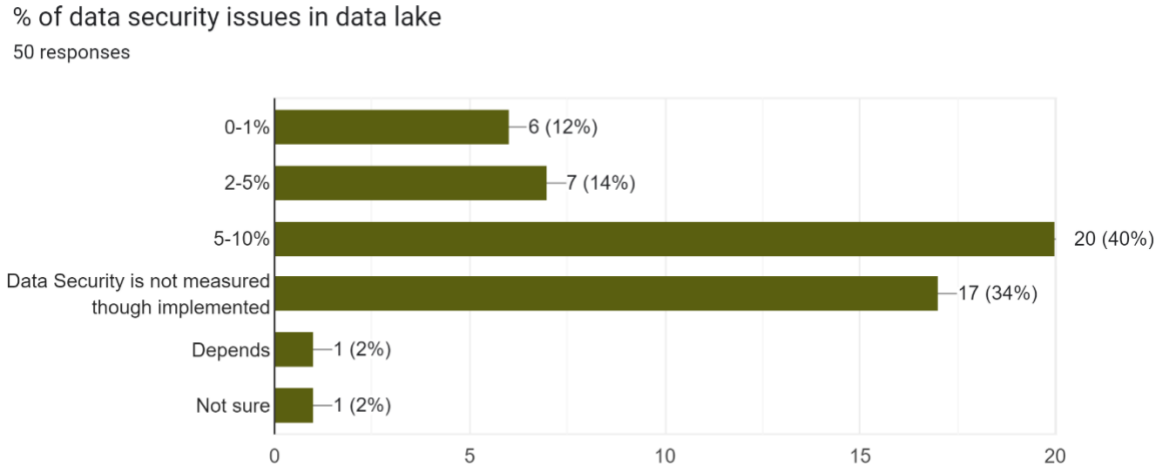


Figure 43: BNPL data lake – Data Security issues

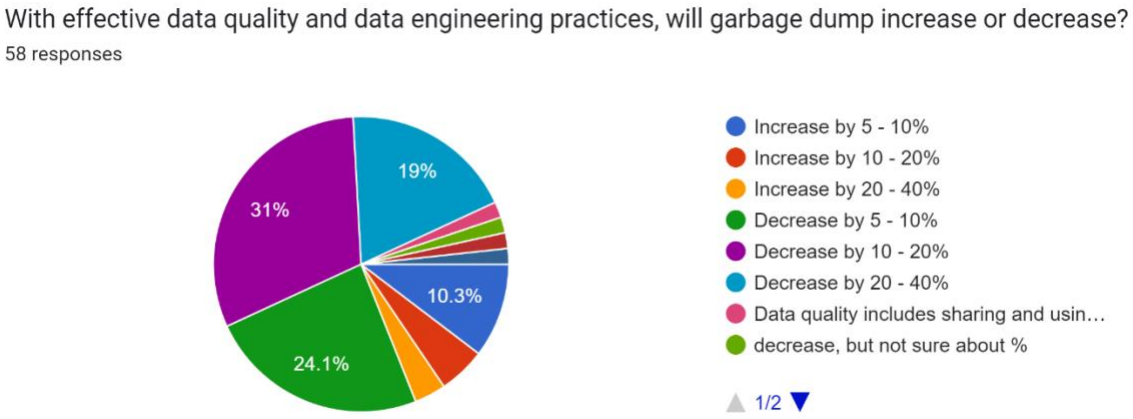


Figure 44: BNPL data lake – Garbage dump

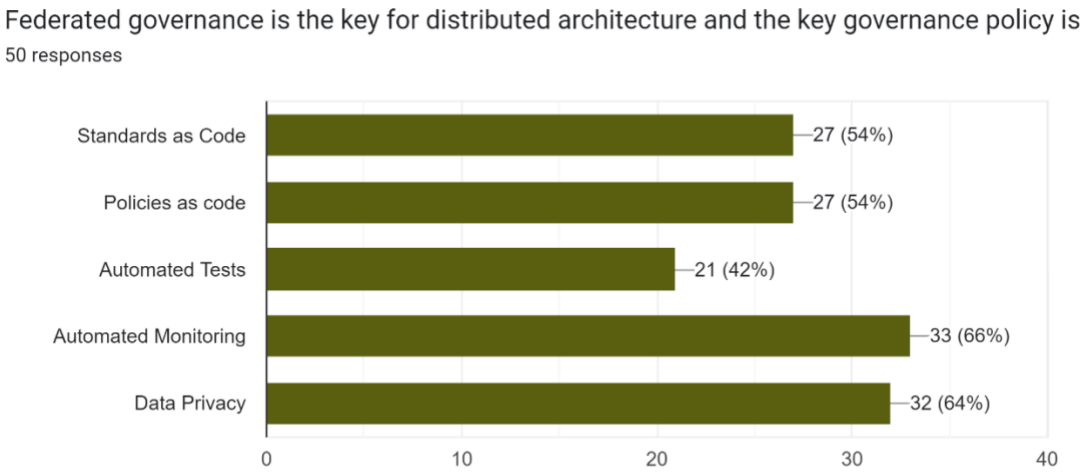


Figure 45: BNPL data lake – Data Governance

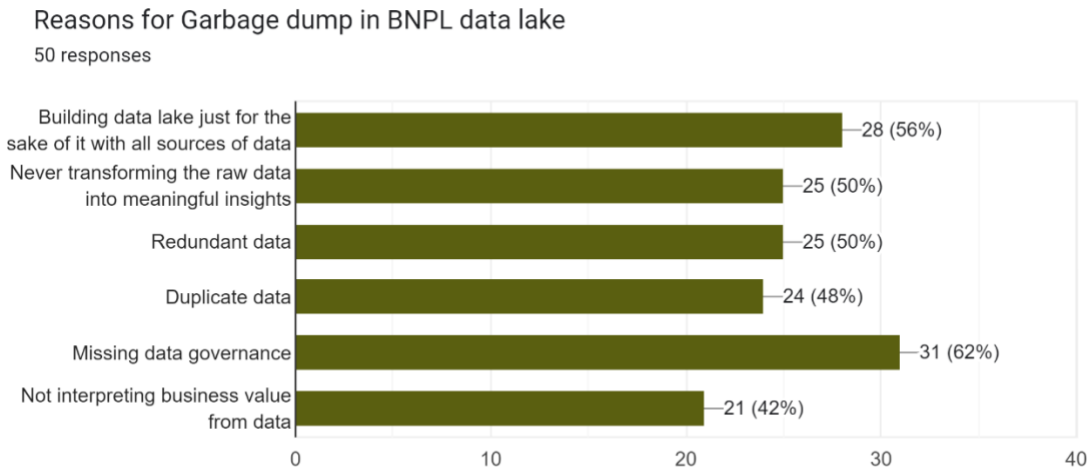


Figure 46: BNPL data lake – Garbage dump reasons

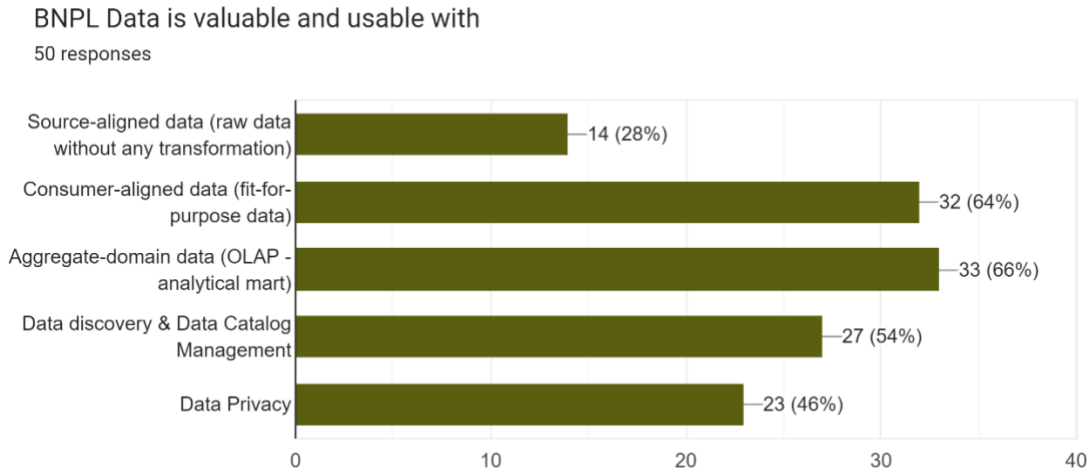


Figure 47: BNPL data lake – data asset

APPENDIX I

DATA LAKE CANVAS

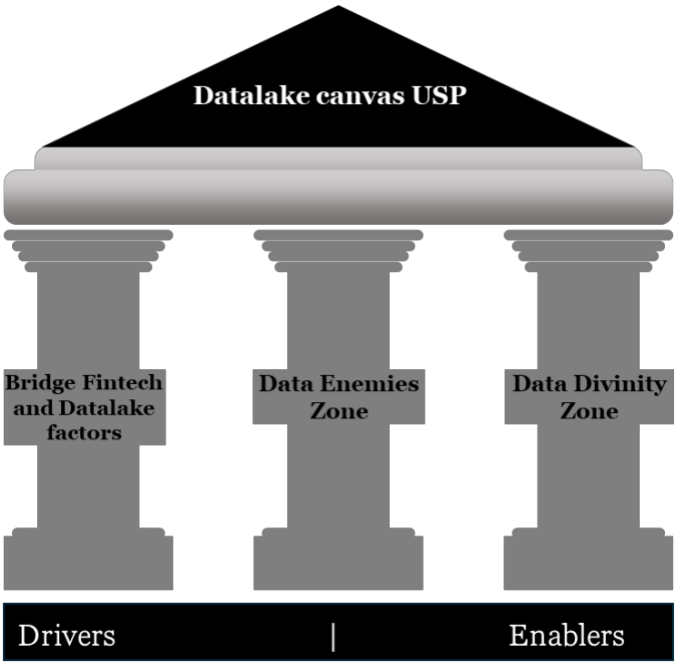


Figure 48: Fintech-BNPL data lake canvas pillars

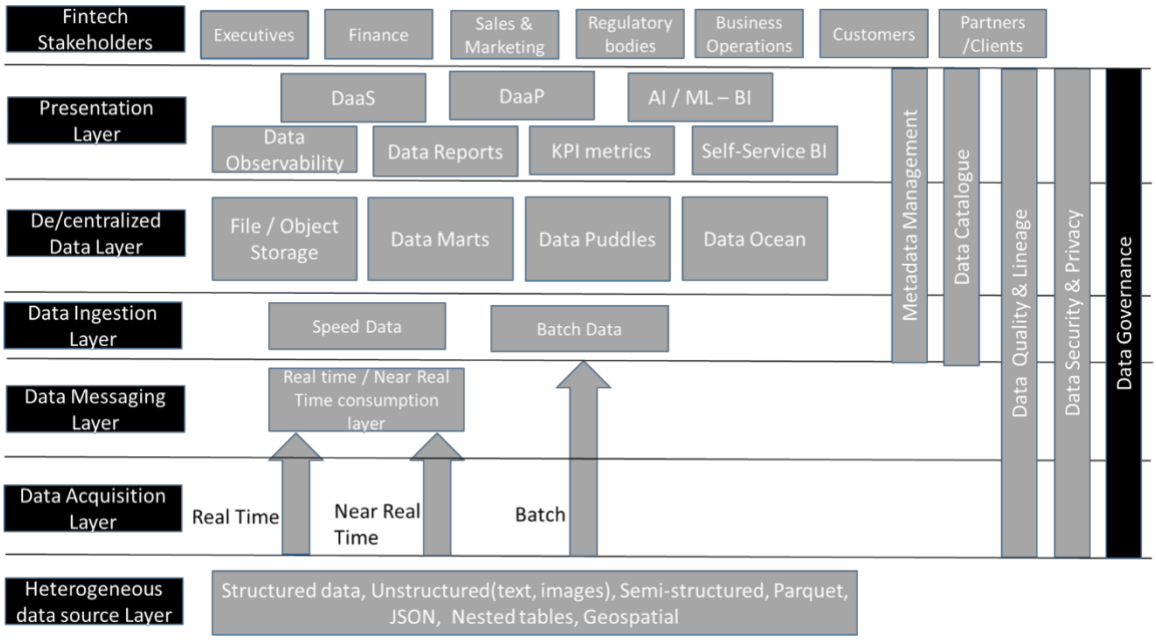


Figure 49a: Fintech-BNPL data lake architecture

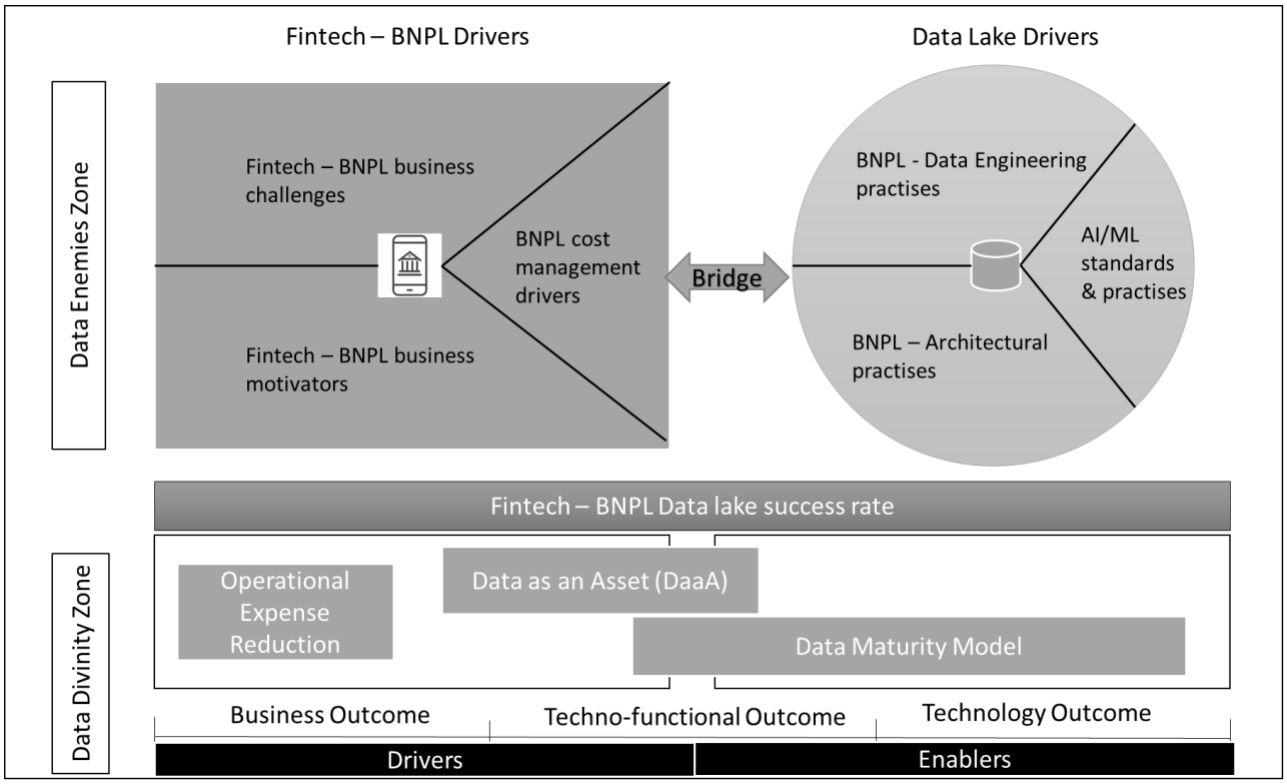


Figure 49b: Fintech - BNPL data lake canvas

Unit of Economies				
Cost Drivers	**Unit**	**Count**	**Amount**	**Description**
Server	No.of servers and server cost	0	$ -	Defines the clusters, nodes, servers based on on-prem or cloud setup
Computing power	CPU and RAM power			Defines the RAM, CPU, cores used with on-prem or cloud setup
Storage	Storage cost	0	$ -	Defines the space utilization in DBMS - SQL / NoSQL, distributed file system
Database License Cost	Licensing cost	0		Defines the license cost either perpetual or subscription cost for database
BI / Data obervability - License / Open source cost	For eg, **Tableau license cost**			Defines the license cost either perpetual or subscription cost for BI tool used
	Total tableau users & cost	0	$ -	
	Tableau Desktop	0	$ -	
	Tableau Explorer (Admin, Creator)	0	$ -	
	Tableau Viewer	0	$ -	
	For eg, for open-source - Dremio / Superset			
	DCU used	0	$ -	
Data Quality	# of data rules, data quality issues / day, cost per defect	0	$ -	Defines the data quality rules set, onboarded, executed and its associated cost
Data Security	# of data security policies, data security issues / year, cost per defect	0	$ -	Defines the data security rules set as policies, Security Group, IAM. Defines the security rules onboarded, executed and its associated cost.
Infra setup cost			$ -	Defines the cost required to setup the application based on on-prem or cloud
Migration Cost			$ -	Defines the cost required to migrate the application from on-prem to cloud based on lift & shift or with modernization
Data Pipeline maintenance cost			$ -	Defines the cost required to maintain the application based on on-prem or cloud
Enablers				
Resources			$ -	Defines the people resources or others apart from above required
L&D			$ -	Defines the L&D resources either as online, offline and on-job trainings required
Total Cost				
Revenue Drivers				
Data as a Product	# of data products and revenue generated	0	$ -	Defines no. of data products and revenue generated
Data as a Service	# of data services and revenue generated	0	$ -	Defines no. of data services and revenue generated
In built tools	# of home grown tools built and used, and revenue generated	0	$ -	Defines the home grown tools built and used, and revenue generated
Risk Management	# of data services and revenue generated		$ -	Defines no. of data services and revenue generated for managing the risk management of BNPL
Fraud Protection	# of data services and revenue generated		$ -	Defines no. of data services and revenue generated for managing the Fraud Protection of BNPL
Compliance and Regulation	# of data services and revenue generated		$ -	Defines no. of data services and revenue generated for managing the Compliance and Regulation of BNPL
Customer Retention	# of data services and revenue generated		$ -	Defines no. of data services and revenue generated for managing the Customer Retention of BNPL
Customer Acquisition	# of data services and revenue generated		$ -	Defines no. of data services and revenue generated for managing the Customer Acquisition of BNPL
Strategic Drivers				
Time to Market	# of Strategic initatives for deriving quick TTM for a business		$ -	Defines no. Strategic initatives for deriving quick TTM for BNPL business
Time to Value	# of Strategic initatives for deriving quick TTV for a business		$ -	Defines no. Strategic initatives for deriving quick TTV for BNPL business
Total Revenue				
Net Profit				

Figure 50: BNPL data lake – Unit of Economies

Table 16: Fintech challenges

Fintech challenges	**Enemy Bucket**

Compliance & Regulation	Flyer
Data Security	Frontal attacker
Customer Retention and acquisition Data and AI integration Competition Service personalization	Flank
Blockchain integration Market availability Sales and cash flow	Fringe

Table 17: BNPL Challenges

BNPL challenges	**Enemy Bucket**
Risk Management	Flyer
Fraud Protection	Frontal attacker
Customer Retention and acquisition Economic and geo-political situation	Flank
High-interest rate	Fringe

Table 18: Risk Management

Risk Management	**Enemy Bucket**
Fraud Protection	Flyer
Risk Scoring	Frontal attacker
Missed payments	Flank

Risk of debt spirals across multiple BNPL provides Interest rates Increased debt for Fintech companies	
Forced lending Profitability of the business model	Fringe

Table 19: Customer journey

Customer journey	**Enemy Bucket**
Customer satisfaction	Flyer
Improved efficiency in managing returns	Frontal attacker
Customer loyalty Merchant brand exposure Customer referrals	Flank
Customer experience	Fringe

Table 20: Key subject areas

Key subject areas	**Enemy Bucket**
Payments & schedule	Flyer
Risk levels & category	Frontal attacker
Customer Scoring Product catalog of BNPL Limits	Flank

Financial calculation Master data and metadata Bureau data	Fringe

Table 21: BNPL cost increase factors

BNPL cost increase factors	**Enemy Bucket**
Risk management	Flyer
Fraud protection	Frontal attacker
Operational expense Customer retention & acquisition Current economical and geo-political situation	Flank
Technology & Tools Customer support Cost of funding	Fringe

Table 22: BNPL cost-effective factors

BNPL cost-effective factors	**Enemy Bucket**
Good Customer Acquisition & Retention strategy Availability of Technology & Tools	Flyer
Proper Risk Management Controlled Operational Expense (OpEx)	Frontal attacker
Favorable economic/geopolitical / Inflation factors	Flank

Excellent Customer support	
Goodwill gained from customers Sufficient funding and the right market Heavy late payment fees	Fringe

Table 23: BNPL cost measurement factors

BNPL cost measurement factors	**Enemy Bucket**
License cost for Tools & technology	Flyer
As provided by the cloud / on-prem	Frontal attacker
People resources	Flank
Full cost basis (direct + indirect costs) Provider cost + Operational / maintenance cost ROI Storage	Fringe

Table 24: BNPL Data lake preference

Data lake preference	**Enemy Bucket**
Hybrid cloud data lake	Flyer
Private cloud data lake	Frontal attacker
Multi-cloud data lake Cloud data lake based on service offering (IaaS, PaaS, SaaS, etc.)	Flank
On-prem data lake Public data lake	Fringe

On-prem data warehouse

Table 25: Preferred scalable BNPL architecture

Preferred scalable BNPL architecture	**Enemy Bucket**
Microservices architecture Data mesh architecture	Flyer
GFS / HDFS architecture Event-driven architecture	Frontal attacker
Data warehousing architecture Kappa architecture Dynamo architecture	Flank
Chubby architecture Columnar file systems	Fringe

Table 26: Preferred data model for BNPL data lake

Preferred data model for BNPL data lake	**Enemy Bucket**
Relational model	Flyer
Polyglot/multi-model	Frontal attacker
Document model Graph model	Flank
Snowflake model	Fringe

Table 27: Preferred data structure for BNPL data lake

Preferred data structure for BNPL data lake	Enemy Bucket
Offered by the distributed file system	Flyer
B trees Memtable	Frontal attacker
SST LSM	Flank
Reverted index B+ B+	Fringe

Table 28: Critical elements for effective AI/ML

Critical elements for effective AI/ML	Enemy Bucket
Data quality	Flyer
Tools & technology ML pipelines Data integration	Frontal attacker
Self-service business intelligence Data discovery	Flank
Feature store Expected sample and population of data	Fringe

Table 29: Features stores expected in BNPL data lake

Features stores expected in BNPL data lake	Enemy Bucket

Payments & transactions Customer 360	Flyer
Product & pricing Merchant 360	Frontal attacker
Finance	Flank
Risk, behavioral, and experimentation	Fringe

APPENDIX J

EVALUATION OF RESULTS FOR DATA LAKE CANVAS

Table 30: Current Observation from the traditional data lake methods

bnpl_opex_cost	Current state
Increase by 5 - 10%	11%
Increase by 10 - 20%	5%
Increase by 20 - 40%	3%
Decrease by 5 - 10%	24%
Decrease by 10 - 20%	31%
Decrease by 20 - 40%	19%
Others	7%

Table 31: Expected state with data lake canvas – new responses

bnpl_opex_cost	Desire State
Increase by 5 - 10%	10%
Increase by 10 - 20%	2%
Increase by 20 - 40%	0%
Decrease by 5 - 10%	15%
Decrease by 10 - 20%	40%
Decrease by 20 - 40%	28%
Others	5%

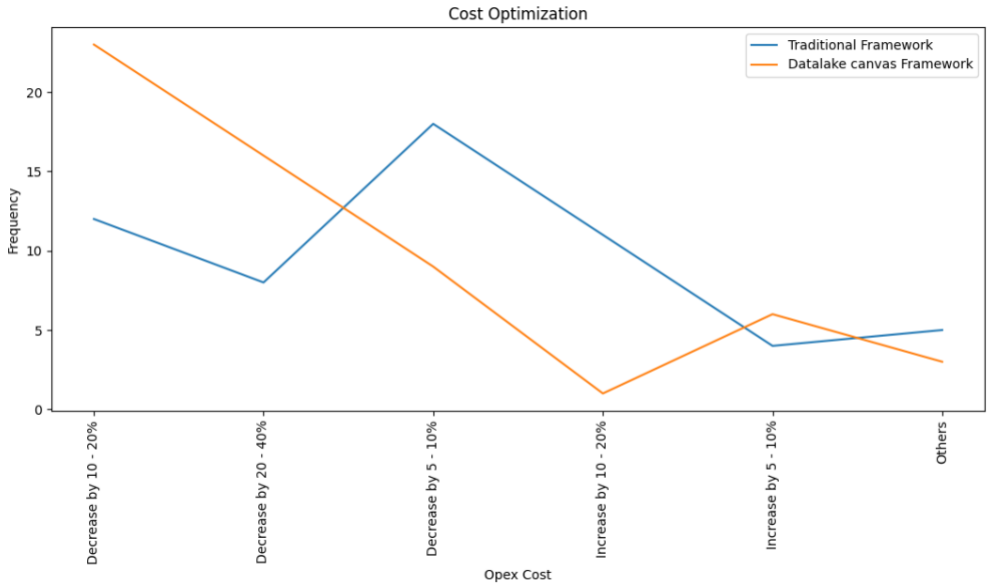


Figure 51: Operational Expense distribution

Table 32: Current DQI from the traditional data lake methods

Data Quality Issues	Current state
0-20%	28%
20-40%	26%
>40%	16%
Data quality is not measured though implemented	16%
20-40%; Data quality is not measured though implemented	6%
0-20%; Depends on the consumer	4%
0-20%; Data quality is not measured though implemented	2%
20-40%;>40%	2%

Table 33: Expected DQI with data lake canvas – new responses

Data Quality Issues	Desire State
0-20%	75%
20-40%	15%
>40%	10%

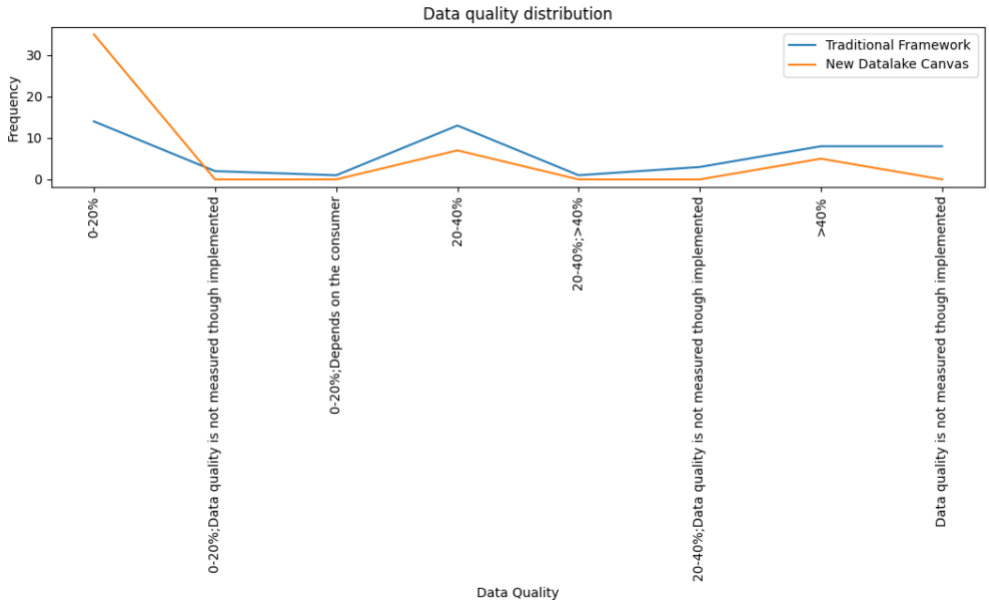


Figure 52: Data quality distribution

Table 34: Current DSI from the traditional data lake methods

Data Security Issues	Current state
0-1%	11%
2-5%	13%
5-10%	39%
Data security is not measured though implemented	33%
Depends	2%
Not sure	2%

Table 35: Expected DSI with data lake canvas – new responses

Data Security Issues	Desire State
0-1%	30%
2-5%	40%
5-10%	20%
Data security is not measured though implemented	4%
Depends	2%
Not sure	4%

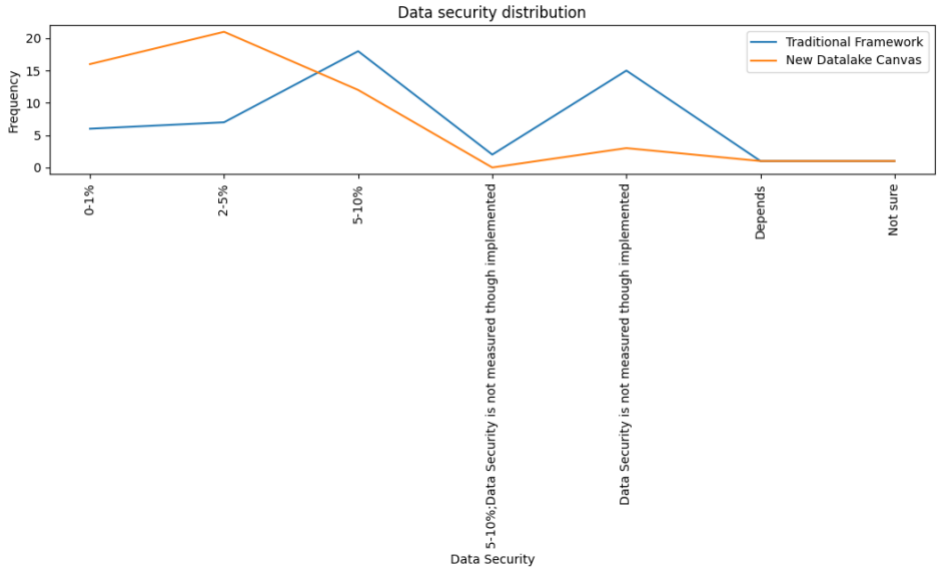


Figure 53: Data security distribution

Table 36: Current DGI from the traditional data lake methods

Data governance Issues	Current state
Decrease by 10 - 20%	31%
Decrease by 20 - 40%	19%
Decrease by 5 - 10%	24%
Increase by 10 - 20%	5%
Increase by 20 - 40%	4%
Increase by 5 - 10%	10%
Others	7%

Table 37: Expected DGI with data lake canvas – new responses

Data governance Issues	Desire State
Decrease by 10 - 20%	10%
Decrease by 20 - 40%	10%
Decrease by 5 - 10%	40%
Increase by 10 - 20%	10%
Increase by 20 - 40%	10%
Increase by 5 - 10%	10%
Others	10%

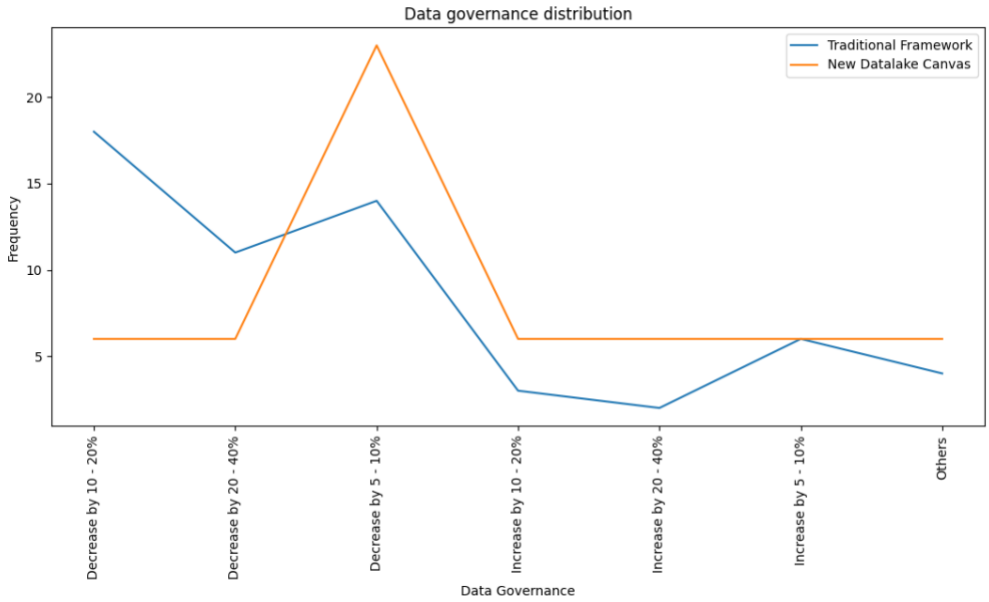


Figure 54: Data governance distribution

Table 38: Current TTM from the traditional data lake methods – TTM_1I

Time To Market (without modernization)	Scoring	Current state
Others	0	14%
Enterprise > 6 months	1	2%
Enterprise < 6 months	3	5%
Startup - < 1 months	6	19%
SMB < 2 months	8	21%
Enterprise < 3 months	10	39%

Table 39: Expected TTM with data lake canvas – new responses – TTM_1I

Time To Market (without modernization)	Scoring	Desire State
Startup - < 1 months	6	20%
SMB < 2 months	8	25%
Enterprise < 3 months	10	55%

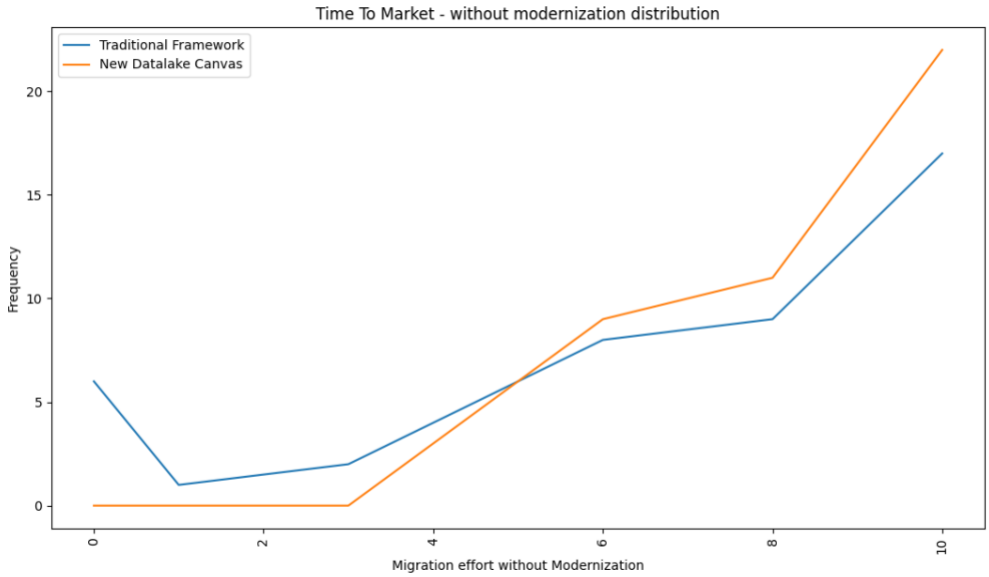


Figure 55: TTM – Migration without modernization distribution

Table 40: Current TTM from the traditional data lake methods – TTM_2I

Time To Market (Build from Scratch)	Scoring	Current state
Startup - > 2 months	3	4%
SMB - 6 months	4	12%
Enterprise > 12 months	6	34%
Startup - < 2 months	7	18%
SMB - 3 - 4 months	8	14%
Enterprise > 6 months	10	18%

Table 41: Expected TTM with data lake canvas – new responses – TTM_2I

Time To Market (Build from Scratch)	Scoring	Desired state
Enterprise > 12 months	6	16%
Startup - < 2 months	7	24%
SMB - 3 - 4 months	8	28%
Enterprise > 6 months	10	30%

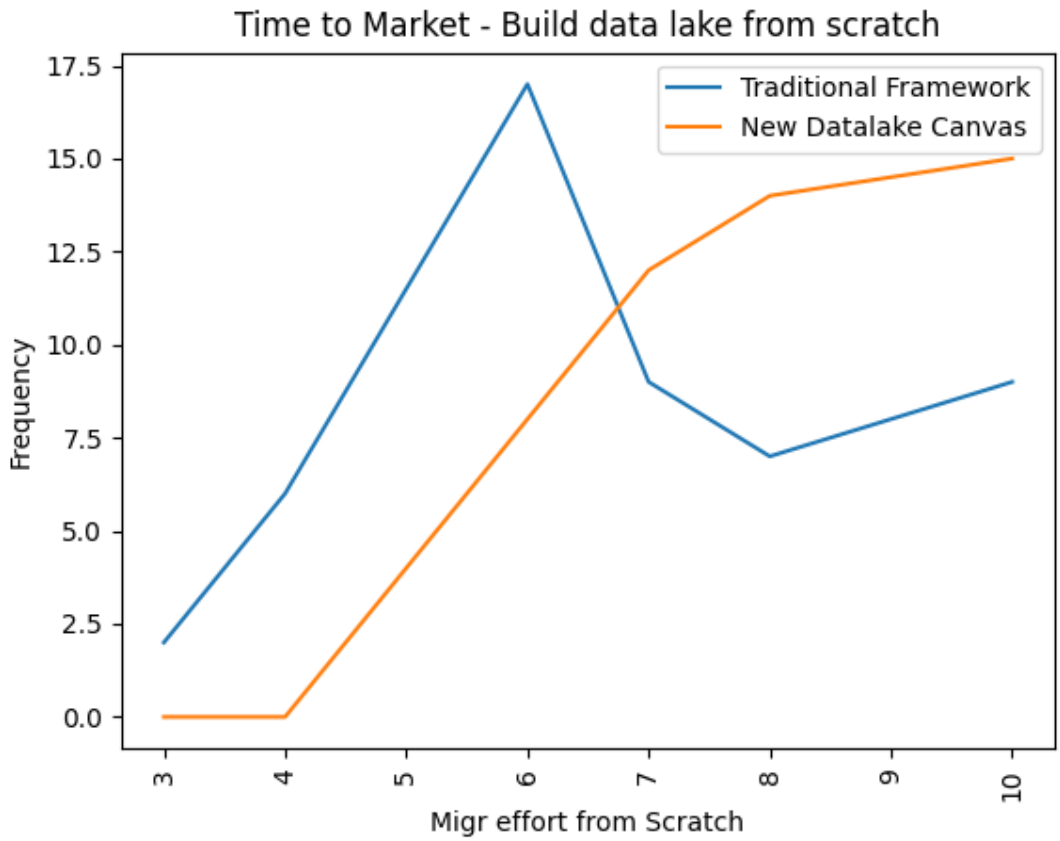


Figure 56: TTM – Build from scratch distribution

Table 43: Operational cost – New response with data lake canvas

Cost Optimization response	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
Increase by 5 - 10%	0	0	0	1	10%
Increase by 10 - 20%	0	0	0	0	2%
Increase by 20 - 40%	0	0	0	0	0%
Decrease by 5 - 10%	1	0	0	0	15%
Decrease by 10 - 20%	1	1	1	1	40%
Decrease by 20 - 40%	0	1	2	0	28%
Others	0	1	0	0	5%
Total Response	2	3	3	2	10

Table 44: Data Quality – New response with data lake canvas

Data Quality Issues	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
0-20%	0	2	3	2	75%
20-40%	1	1	0	0	15%
>40%	1	0	0	0	10%
Others	0	0	0	0	0%
Total Responses	2	3	3	2	100%

Table 45: Data Security – New response with data lake canvas

Data Security Issues	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
0-1%	0	1	1	1	30%
2-5%	1	1	1	1	40%
5-10%	1	1	0	0	20%
Data security is not measured though implemented	0	0	1	0	4%
Depends	0	0	0	0	2%
Not sure	0	0	0	0	4%
Total Response	2	3	3	2	100%

Table 46: Data Governance – New response with data lake canvas

Data governance Issues	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
Decrease by 10 - 20%	0	0	1	0	10%
Decrease by 20 - 40%	0	1	0	0	10%
Decrease by 5 - 10%	1	1	1	1	40%

Increase by 10 - 20%	0	1	0	0	10%
Increase by 20 - 40%	1	0	0	0	10%
Increase by 5 - 10%	0	0	1	0	10%
Others	0	0	0	1	10%
Total Response	2	3	3	2	100%

Table 47: TTM_1I – New response with data lake canvas

Time To Market (without modernization)	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
Startup - < 1 months	1	1	0	0	20%
SMB < 2 months	1	1	0	1	25%
Enterprise < 3 months	0	1	3	1	55%
Total Response	2	3	3	2	100%

Table 48: TTM_2I – New response with data lake canvas

Time To Market (without modernization)	Fintech leader	Data Leader	Engineering Leader	Data Architect / Engineer / Analyst	New response %
Enterprise > 12 months	1	1	0	0	15%
Startup - < 2 months	1	0	1	0	25%
SMB - 3 - 4 months	0	1	1	1	30%
Enterprise > 6 months	0	1	1	1	30%
Total Response	2	3	3	2	100%

REFERENCES

1. 81 Key Fintech Statistics 2021/2022: Market Share & Data Analysis, n.d.
2. Aalders, R., 2023. Buy now, pay later: redefining indebted users as responsible consumers. Information, Communication & Society 26, 941–956. https://doi.org/10.1080/1369118X.2022.2161830
3. Agarwal, S., Zhang, J., 2020. FinTech, Lending and Payment Innovation: A Review.
4. Akhtar, Dr.N., Kerim, Dr.B., Perwej, Dr.Y., Tiwari, Dr.A., Praveen, Dr.S., 2021. A Comprehensive Overview of Privacy and Data Security for Cloud Storage. IJSRSET 113–152. https://doi.org/10.32628/IJSRSET21852
5. Akhtar, M.H., Chaudhry, I.S., Sheikh, M.R., 2020. Does Islamic Banking Augment Financial Inclusion in Pakistan? A Reinforcement Analysis. READS 6, 739–758. https://doi.org/10.47067/reads.v6i4.275
6. Alcazar, J., Bradford, T., 2021. The Rise of Buy Now, Pay Later: Bank and Payment Network Perspectives and Regulatory Considerations. Payments System Research Briefing 1–6.
7. Alshahri, A., 2022. B2B BNPL Payment Solution in MENA: A Startup business case. https://doi.org/10.13140/RG.2.2.18450.73927
8. Anshari, M., Almunawar, M.N., Masri, M., 2022. Digital Twin: Financial Technology's Next Frontier of Robo-Advisor. JRFM 15, 163. https://doi.org/10.3390/jrfm15040163
9. Berger, A.N., Makaew, T., Roman, R.A., 2019. Do Business Borrowers Benefit from Bank Bailouts?: The Effects of TARP on Loan Contract Terms. Financial Management 48, 575–639. https://doi.org/10.1111/fima.12222

10. Bu, Y., Li, H., Wu, X., 2022. Effective regulations of FinTech innovations: the case of China. Economics of Innovation and New Technology 31, 751–769. https://doi.org/10.1080/10438599.2020.1868069
11. Center, S.B.P., 2022. Point of Fail: How a Flood of "Buy Now, Pay Later" Student Debt is Putting Millions at Risk. Student Borrower Protection Center Research Paper.
12. Chelliah, P.R., Surianarayanan, C., 2021. Multi-Cloud Adoption Challenges for the Cloud-Native Era: Best Practices and Solution Approaches. International Journal of Cloud Applications and Computing 11, 67–96. https://doi.org/10.4018/IJCAC.2021040105
13. Cherradi, M., EL Haddadi, A., Routaib, H., 2022. Data Lake Management Based on DLDS Approach, in: Ben Ahmed, M., Teodorescu, H.-N.L., Mazri, T., Subashini, P., Boudhir, A.A. (Eds.), Networking, Intelligent Systems and Security, Smart Innovation, Systems and Technologies. Springer Singapore, Singapore, pp. 679–690. https://doi.org/10.1007/978-981-16-3637-0_48
14. David Mitchell Smith, 2022. The Gartner Hype Cycle for Cloud Computing.
15. Delabarre, Maxime, 2021. FinTech in the Financial Market.
16. El Haddad, G., Aimeur, E., Hage, H., 2018. Understanding Trust, Privacy and Financial Fears in Online Payment, in: 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). Presented at the 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering

(TrustCom/BigDataSE), IEEE, New York, NY, USA, pp. 28–36. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00015

17. Elkholy, M., A. Marzok, M., 2022. Trusted Microservices: A Security Framework for Users' Interaction with Microservices Applications. JISCR 5, 135–143. https://doi.org/10.26735/QOPM9166
18. Fu, R., Huang, Y., Singh, P.V., 2021. Crowds, Lending, Machine, and Bias. Information Systems Research 32, 72–92. https://doi.org/10.1287/isre.2020.0990
19. Gai, K., Qiu, M., Sun, X., 2018. A survey on FinTech. Journal of Network and Computer Applications 103, 262–273. https://doi.org/10.1016/j.jnca.2017.10.011
20. Gerrans, P., Baur, D.G., Lavagna-Slater, S., 2022. Fintech and responsibility: Buy-now-pay-later arrangements. Australian Journal of Management 47, 474–502. https://doi.org/10.1177/03128962211032448
21. Giebler, Corinna, Gröger, Christoph, Hoos, Eva, Eichler, Rebecca, Schwarz, Holger, Mitschang, Bernhard, 2021a. The Data Lake Architecture Framework. https://doi.org/10.18420/BTW2021-19
22. Giebler, Corinna, Gröger, Christoph, Hoos, Eva, Eichler, Rebecca, Schwarz, Holger, Mitschang, Bernhard, 2021b. The Data Lake Architecture Framework. https://doi.org/10.18420/BTW2021-19
23. Gorelik, A., 2019. The enterprise big data lake: delivering the promise of big data and data science, First edition. ed. iO'Reilly Media, Inc, Sebastopol, California.
24. Gulati, V.P., Srivastava, S., 2007. The empowered internet payment gateway, in: International Conference on E-Governance. Citeseer, pp. 98–107.
25. Guttman-Kenney, B., Firth, C., Gathergood, J., 2022. Buy Now, Pay Later (BNPL)...On Your Credit Card. SSRN Journal. https://doi.org/10.2139/ssrn.4001909

26. Hai, R., Quix, C., Jarke, M., 2021a. Data lake concept and systems: a survey. arXiv:2106.09592 [cs].
27. Hai, R., Quix, C., Jarke, M., 2021b. Data lake concept and systems: a survey. arXiv:2106.09592 [cs].
28. Hua, X., Huang, Y., 2021. Understanding China's Fintech sector: development, impacts and risks. The European Journal of Finance 27, 321–333. https://doi.org/10.1080/1351847X.2020.1811131
29. Imerman, M.B., Fabozzi, F.J., 2020. A Conceptual Framework for Fintech Innovation. SSRN Journal. https://doi.org/10.2139/ssrn.3543810
30. Inmon, W.H., 2016. Data lake architecture: designing the data lake and avoiding the garbage dump, First edition. ed. Technics Publications, Basking Ridge, NJ.
31. John, T., Misra, P., 2017. Data lake for enterprises: leveraging Lambda architecture for building enterprise data lake. Packt, Birmingham, UK.
32. Khan, A., Vilary Mbanyi, A., 2022. Millennial's fashion buying behavior from Buy Now, Pay Later perspective: A study of Buy Now, Pay Later (BNPL) and its influence on millennials buying behavior andconsumption when mobile shopping.
33. Kumar, E.S., Kesavan, S., Naidu, R.Ch.A., Kumar R, S., Latha, 2021. Comprehensive Analysis of Cloud based Databases. IOP Conf. Ser.: Mater. Sci. Eng. 1131, 012021. https://doi.org/10.1088/1757-899X/1131/1/012021
34. Lin, H., Chen, C., Chiu, Y., Lin, T., 2022. How financial technology (Fintech) can improve the business performance of securities firms by using the dynamic data envelopment analysis modified model. Manage Decis Econ 43, 1113–1132. https://doi.org/10.1002/mde.3443
35. Maurya, S., Mufti, T., Kumar, D., Mittal, P., Gupta, R., 2021. A Study on Cloud Computing: A Review, in: Proceedings of the 2nd International Conference on

ICT for Digital, Smart, and Sustainable Development, ICIDSSD 2020, 27-28 February 2020, Jamia Hamdard, New Delhi, India. EAI, New Delhi, India. https://doi.org/10.4108/eai.27-2-2020.2303253

36. McKinsey, 2022. The 2022 McKinsey Global Payments Report.
37. Nargesian, F., Zhu, E., Miller, R.J., Pu, K.Q., Arocena, P.C., 2019a. Data lake management: challenges and opportunities. Proc. VLDB Endow. 12, 1986–1989. https://doi.org/10.14778/3352063.3352116
38. Nargesian, F., Zhu, E., Miller, R.J., Pu, K.Q., Arocena, P.C., 2019b. Data lake management: challenges and opportunities. Proc. VLDB Endow. 12, 1986–1989. https://doi.org/10.14778/3352063.3352116
39. Norrestad, F., 2022. Number of Fintech startups worldwide from 2018 to November 2021, by region.
40. Oberoi, A., Dave, Y., Patel, B., Anas, M., 2021. Cloud Computing in Banking Sector – A Case Study.
41. Oliveira Rocha, H.F., 2022. Practical Event-Driven Microservices Architecture: Building Sustainable and Highly Scalable Event-Driven Microservices. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-7468-2
42. Panigrahy, S., Dash, B., Thatikonda, R., 2023. From Data Mess to Data Mesh: Solution for Futuristic Self-Serve Platforms. INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING 12. https://doi.org/10.17148/IJARCCE.2023.124121
43. Parne, P., 2021. Cloud Computing Strategy and Impact in Banking/Financial Services, in: Computer Science and Information Technology Trends. Presented at the 5th International Conference on Computer Science and Information

Technology (COMIT 2021), Academy and Industry Research Collaboration Center (AIRCC), pp. 37–45. https://doi.org/10.5121/csit.2021.111704

44. PwC, 2019. Global Fintech Report 2019.
45. Rehman, F.U., Attaullah, H.M., Ahmed, F., Ali, S., 2023. Data Defense: Examining Fintech's Security and Privacy Strategies, in: INTERACT 2023. Presented at the INTERACT 2023, MDPI, p. 3. https://doi.org/10.3390/engproc2023032003
46. Relja, R., Ward, P., Zhao, A.L., 2023. Understanding the psychological determinants of buy-now-pay-later (BNPL) in the UK: a user perspective. IJBM. https://doi.org/10.1108/IJBM-07-2022-0324
47. Sawadogo, P., Darmont, J., 2021a. On data lake architectures and metadata management. J Intell Inf Syst 56, 97–120. https://doi.org/10.1007/s10844-020-00608-7
48. Sawadogo, P., Darmont, J., 2021b. On data lake architectures and metadata management. J Intell Inf Syst 56, 97–120. https://doi.org/10.1007/s10844-020-00608-7
49. Scardovi, C., 2017. Digital Transformation in Financial Services, 1st ed. 2017. ed. Springer International Publishing : Imprint: Springer, Cham. https://doi.org/10.1007/978-3-319-66945-8
50. Scardovi, C., 2016. Restructuring and Innovation in Banking - Fin Tech Innovation and the Disruption of the Global Financial System - Chapter 2.
51. Sridhar, S., 2022. Winning without Waging War.
52. Suryono, R.R., Budi, I., Purwandari, B., 2020. Challenges and Trends of Financial Technology (Fintech): A Systematic Literature Review. Information 11, 590. https://doi.org/10.3390/info11120590

53. Vinoth, S., Vemula, H.L., Haralayya, B., Mamgain, P., Hasan, M.F., Naved, M., 2022. Application of cloud computing in banking and e-commerce and related security threats. Materials Today: Proceedings 51, 2172–2175. https://doi.org/10.1016/j.matpr.2021.11.121
54. Wang, R., 2022. Business Model Innovation in Swedish FinTech Industry: A case Study of Klarna.
55. Weintraub, G., Gudes, E., Dolev, S., 2021. Indexing cloud data lakes within the lakes, in: Proceedings of the 14th ACM International Conference on Systems and Storage. ACM, Haifa Israel, pp. 1–1. https://doi.org/10.1145/3456727.3463828
56. Wulf, F., Lindner, T., Strahringer, S., Westner, M., 2021. IaaS, PaaS, or SaaS? The Why of Cloud Computing Delivery Model Selection: Vignettes on the Post-Adoption of Cloud Computing, in: Proceedings of the 54th Hawaii International Conference on System Sciences, 2021. pp. 6285–6294.
57. Xing, Y., Chen, H., Zhuang, X., 2019. Australian Online BNPL Services Research: Building Gain Value Model of Individual Credit Background, in: Proceedings of the 2019 International Conference on Information Technology and Computer Communications. Presented at the ITCC 2019: 2019 International Conference on Information Technology and Computer Communications, ACM, Singapore Singapore, pp. 45–51. https://doi.org/10.1145/3355402.3355410
58. ZHAO, Y., 2021. Metadata Management for Data Lake Governance (Theses). Université Toulouse 1 Capitole (UT1 Capitole).
59. Zouari, F., Kabachi, N., Boukadi, K., Ghedira Guegan, C., 2021. Data Management in the Data Lake: A Systematic Mapping, in: 25th International Database Engineering & Applications Symposium. ACM, Montreal QC Canada, pp. 280–284. https://doi.org/10.1145/3472163.3472173