AN OPERATING MODEL TO OPERATIONALIZE THE MODERN DATA PLATFORM USING DISTRIBUTED DATA MESH AND MULTI-CLOUDS

by

ANIKET MHALA, M.Eng.(IT)

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

SEPTEMBER, 2024

AN OPERATING MODEL TO OPERATIONALIZE THE MODERN DATA PLATFORM USING DISTRIBUTED DATA MESH AND MULTI-CLOUDS

by

ANIKET MHALA

APPROVED BY Vassiliki Grougiou

____

Dissertation chair

RECEIVED/APPROVED BY:

____

Admissions Director

Dedication

This work is dedicated to all the inspiring individuals who have shaped my understanding of leadership and contributed to my journey.

To my family and friends, whose love, encouragement, and unwavering belief in me have fueled my passion for this research.

To my friend, Manojkumar Pawar, whose support, encouragement, and advice were instrumental in bringing this work to completion.

To my team members, whose unwavering support and dedication have been instrumental in my growth and development. Your insights, collaboration, and hard work have been invaluable.

To my peers and leaders, who have shared their knowledge, experiences, and mentorship. Your guidance and support have inspired me to strive for excellence.

To the data architects, CDOs, and industry experts who have generously shared their expertise and insights on distributed data mesh and modern data platforms.

This thesis is a testament to the collective wisdom and support of these individuals, and I am grateful for their contributions.

Acknowledgements

I am deeply grateful to the countless individuals who have supported me throughout my academic journey and contributed to the completion of this thesis.

A special thanks to my team members, whose dedication, hard work, and unwavering support have been instrumental in my success. Your trust, collaboration, and belief in my leadership have inspired me to strive for excellence.

I would also like to acknowledge the guidance and mentorship of my advisors and colleagues, whose expertise and insights have enriched my understanding of the field.

Finally, to my esteemed research mentor, whose valuable guidance and expertise have been instrumental in conducting my research and shaping this research thesis.

This thesis is a testament to the collective effort and contributions of all those who have supported me along the way.

ABSTRACT

AN OPERATING MODEL TO OPERATIONALIZE THE MODERN DATA PLATFORM USING DISTRIBUTED DATA MESH AND MULTI-CLOUDS

ANIKET MHALA
2024

Dissertation Chair: Dr. Vasiliki Grougiou

The proliferation of data in terms of increasing volume, velocity, and variety of data have necessitated the development of advanced data platforms to support modern business needs. Distributed data mesh and multi-cloud architectures have emerged as promising approaches to address these challenges. However, the operationalization of these platforms remains a complex task, requiring a well-defined operating model to ensure efficient, scalable, and secure data management.

This research aims to develop a comprehensive operating model that can guide organizations in the effective operationalization of modern data platforms utilizing distributed data mesh and multi-cloud architectures. The study seeks to address the critical question: How can organizations establish an effective operating model to operationalize their modern data platforms using distributed data mesh and multi-cloud architectures?

The research explores the key components of an effective operating model, including data governance, data mesh principles, data as a product, data infrastructure, revised roles

and responsibilities to adopt distributed data mesh across the enterprise in strategic and structured way.

The scope of the research will be limited to the operational aspects of modern data platforms using data mesh, excluding strategic considerations such as data monetization and business intelligence.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I:

INTRODUCTION

1.1 Introduction

Data is important to every business. Many organizations generate the data every minute and data management become more and more critical. Therefore, there is a need to understand current business challenges related to the enterprise data management and how the distributed data-mesh using multiple cloud strategy would help to address these challenges (Saranya et al.,2021).

Four to five decades before the data was small in terms of volume, variety & velocity because of the lack of internet and on-line services. However due to continuous innovation across different industry spread across the globe, lot of new applications, systems and new business model got invented which increased the consumers in billions. All these factors are contributing to continuously increasing the data.

Earlier banks were storing their customer data into structured format (i.e., standard rows and columns) using Relational Databases, but later invention happened, and bank started storing the semi-structured data into NOSQL databases and finally another category of data invented as un-structured data and bank started storing the unstructured data into Data Lake. Down the line bank started using the Data Lake as a data repository to store all type of data into it (Yeung et al.,2019).

Based on interviews conducted with banking executives of International Tier 1 banks, Banks are moved from Data to Big Data. Banks moved from standard databases such as Microsoft access, Microsoft SQL Server, IBM DB2, Oracle to Big Data management systems such as Hadoop and now a days banks are moving to cloud native big data platforms such as Cloudera, Databricks, Snowflake, AWS Redshift, Oracle's

ADW (Autonomous Data Warehouse). Focus is more on generating lot of valuable insights with more accurate predication models which will contribute to grow banks business and continuously improve their customer satisfaction (Zagan et al.,2021).

Data is growing so the technology is also changing to manage the big data. Now it’s time to refer the Big Data as Mega Data. However, all solutions to manage the data are centralized such as centralized data base, big data platform and centralized data lake (Emara et al.,2020). Today’s Data platforms based on the data ware house & data lake architecture, have common failure modes that may lead to unfulfilled data promises at scale (Chen, et al.,2018).

Therefore, to address these failure modes banks need to shift from the centralized paradigm or centralized storage or centralized management of a lake, or its predecessor data warehouse. Banks need to shift to a paradigm that comes from modern distributed architecture. It means considering domains as the first-class concern, applying platform thinking to create self-serve data infrastructure, and treating data as a product.

The Legacy applications of banks became monolithic and due to which it is difficult to make modifications, deployment and leveraging new technologies for such applications (Patel, 2019). In the same way, the Data Management solutions which are based on database, data warehouse and data lakes are becoming monolithic. Banks are transforming their legacy applications into distributed architecture by using Microservices as architecture style, containers as packaging mechanism and cloud as a cost-effective model. To get the more benefits from different cloud providers banks are deploying their applications which are distributed in nature on multiple clouds such as AWS, Azure, Oracle Cloud Infrastructure, Google cloud etc...

In the similar way, there is a need to transforms banks centralized monolithic data management solutions as a distributed data mesh using multi-cloud strategy to resolve the typical data management challenges just like applications.

The main objective of the distributed data mesh is **to eliminate the challenges of data availability and accessibility at scale**. Zhamak Dehagani is a principal technology consultant at Thoughtworks, and she introduced a term Data Mesh for a new enterprise data architecture. The Data Mesh is a term which allows business users and data scientists to access, analyze, and operationalize business insights from virtually any data source, in any location, without intervention from expert data teams (Dehghani,2019).

The data mesh concept is relatively new, and it promises several benefits, however operationalizing the Modern Data Platform using distributed data mesh with multi-cloud is major challenge for banking business-people such as CEO, CTO, CIO and Line of Business Head. Banks are not yet started using it because of the lack of the standard operating model which will provide them end-to-end guidance, direction, and roadmaps to move from centralized data platforms to the distributed data mesh or distributed data architecture using multi-clouds for their entire data eco system.

Therefore, the Operating model to operationalize the Modern Data Platform using distributed data mesh with multi-cloud will be proposed during this research. This operating model is also called as the guidance framework, and it will be a great value addition to business and technical people of banks to perform their data transformation for future needs.

But the important question is that, why would banks like to go for distributed mesh with multi clouds.

Banks are already suffered due to centralize or monolithic applications and they are unable to achieve their business drivers such as Faster Time to Market, Improved

Quality, High Availability and Scalability, Zero Down time of the applications. Banks are realizing that their data management solutions are also now becoming monolithic due to centralized implementation despite of using new technology and clouds. As earlier mentioned, banks are facing data availability and scaling issues. The distributed application architecture with the right cloud services from different cloud providers have given lot of benefits in terms of saving cost, allowing to use new technologies, fault tolerant applications, Zero Down time to the banks. Hence just like distributed application architecture banks will go gradually to leverage the Distributed architecture with multi-cloud environment for enterprise data management called as Distributed Data mesh.

The Distributed Data mesh is relatively a new approach for data management. It is still evolving and not yet matured and hence there is a lot of scope for doing the research and propose new things or suggest new best practices. I conducted in depth research inside my organization and by working with different banks business and technical people and proposed the Operating model to transform centralized data platforms into distributed data mesh.

The Data mesh topic is new and promises to provide several benefits. There is not relevant information is available on this topic and hence this topic is going to be become de-facto standard for future enterprise data management, therefore this topic is novel for the research.

As per a research study, below are the key trends that have emerged in Data Mesh include:

Multi-cloud support: Data Mesh allows organizations to create independent data services that can be deployed on different cloud providers, which improves multi-cloud support.

Distributed data architecture: Data Mesh is based on a distributed data architecture, which allows organizations to improve data locality, data availability, and scalability.

Decentralized data ownership: Data Mesh decentralizes data ownership and decision-making, which allows teams to make decisions about how to manage and use their data.

Data governance: Data Mesh establishes data governance policies and procedures and gives teams ownership of their own data domains and services, which improves data governance and ensures compliance with regulations and data security.

Service Mesh: Data Mesh uses a service mesh to manage and orchestrate the deployment and scaling of data services, which improves the scalability, security and observability of the system.

Data Mesh is being adopted in various industries: Data Mesh is not limited to the tech industry, but it could be adopted in any industry where data and services play a crucial role.

Data Mesh is becoming a standard for modern data platforms: Data Mesh is becoming a standard for modern data platforms as it allows organizations to break down monolithic data systems into smaller, independent services that align with the business domain, and provides autonomy, scalability, and flexibility to teams.

Interest in Data Mesh is increasing: Interest in Data Mesh is increasing as organizations are looking for ways to improve the scalability, security, and governance of their data systems. (James Serra,2021)

Banks and other organizations have typical challenges such as data silos, lack of data autonomy, data security, data governance & data latency related to the enterprise data using centralized data management and centralized data teams. Data Mesh is one of the options to address the problems of centralized enterprise data management.

It's important to note that Data Mesh is a relatively new concept and trends may change over time. Organizations needs guidance framework to evaluate their needs and constraints before deciding on the best approach for their organization and keep an eye on new trends in Data Mesh. (T. Priebe et al.,2021)

During the detail literature review different relevant literatures are thoroughly reviewed, inspected and evaluated. Several themes such as Data types, Roles related to Data Management, Data Management model, Different types of clouds such as AWS, Azure, GCP and OCI, Data Governance, Data Architecture have been thoroughly reviewed. Below mentioned themes or theories are key themes and relevant to the Data Mesh, multi-clouds and its operating model.

Data mesh connects siloed data to help enterprises move towards automated analytics at scale. It allows businesses to escape the consumptive trap of monolithic data architectures and save themselves from massive operational and storage costs using multi-cloud services. This new distributed approach aims to clear the data access bottlenecks of centralized data ownership by giving data management and ownership to domain-specific business teams. It is based on the modern, distributed architecture.

The outcome of the research would be an Operating model to transform centralized data platforms of banks into distributed data mesh with multi-cloud.

It will add value to the business and technical people to discover, design, develop and deliver distributed data mesh projects across the bank or organization. Basically, it will set-up the foundation to operationalize the Distributed Data Mesh across the enterprise as a ready to use the reference operating model for business-people to transform their centralized data management platforms to new modern distributed data mesh using multi-cloud services in structured, disciplinary, and innovative way.

1.2 Research Problem

The increasing volume, variety, and velocity of data have necessitated the development of advanced data management architectures to effectively leverage this valuable asset. Distributed data mesh and multi-cloud architectures have emerged as promising approaches to address the challenges of modern data management. However, the operationalization of these platforms remains a complex task, requiring a well-defined operating model.

One of the primary challenges organizations faces is aligning their data strategies with their overall business objectives. While distributed data mesh and multi-cloud architectures offer significant benefits, such as improved scalability, flexibility, and cost-efficiency, they also introduce new complexities that must be carefully managed. Without a clear understanding of how these technologies can support their business goals, organizations may struggle to realize the full potential of their data investments.

Another critical challenge is ensuring data quality and governance within a distributed data mesh and multi-cloud environment. Maintaining data consistency, accuracy, and security across multiple domains and cloud platforms can be difficult. The decentralized nature of data mesh can make it challenging to enforce centralized data governance policies and ensure compliance with regulations.

Furthermore, the complexities of managing distributed and multi-cloud environments can pose significant challenges. Organizations must navigate issues such as data integration, security, performance optimization, and cost management. The lack of established best practices and guidelines can make it difficult for organizations to effectively address these challenges.

In conclusion, while distributed data mesh and multi-cloud architectures offer significant potential benefits, their successful operationalization requires a well-defined operating model that addresses the challenges of alignment, data governance, and complexity. The lack of established best practices in this area presents a significant research gap that needs to be addressed.

1.3 Purpose of Research

This research aims to address the critical gap in understanding and operationalizing distributed data mesh and multi-cloud architectures. By developing a comprehensive operating model, this study will provide organizations with a practical framework for effectively leveraging these technologies to achieve their data-driven goals. The findings of this research will contribute to the body of knowledge on data management and architecture, offering valuable insights to practitioners and researchers alike.

1.4 Significance of the Study

The significance of this research lies in its contribution to addressing the critical gap in understanding and operationalizing distributed data mesh and multi-cloud architectures. By developing a comprehensive operating model, this study provides organizations with a practical framework for effectively leveraging these technologies to achieve their data-driven goals.

One of the primary benefits of this research is its potential to help organizations overcome the challenges associated with adopting distributed data mesh and multi-cloud architectures. By providing a clear roadmap and best practices, this study can guide organizations in navigating the complexities of these technologies and realizing their full potential.

Furthermore, the findings of this research will contribute to the body of knowledge on data management and architecture. The insights gained from this study can be used by practitioners and researchers to inform their decision-making and develop new approaches to data management.

By providing a practical and actionable framework, this research can empower organizations to harness the power of distributed data mesh and multi-cloud architectures to drive innovation, improve efficiency, and gain a competitive advantage. The findings of this study can also be used to inform the development of future research and best practices in the field of data management.

In conclusion, this research is significant because it addresses a critical gap in the literature and provides valuable insights for organizations seeking to leverage distributed data mesh and multi-cloud architectures. The findings of this study can help organizations overcome challenges, make informed decisions, and achieve their data-driven goals.

1.5 Research Purpose and Questions

The primary objective of this research is to delve into the intricacies of various data architecture styles, including data warehouse, data lake, data fabric, and, most importantly, data mesh. By examining the challenges faced by current data platforms and exploring the potential solutions offered by data mesh, this research aims to provide valuable insights for organizations considering adopting this innovative approach.

Key Research Questions:

Data Architecture Comparison: What are the key differences and advantages of data warehouse, data lake, data fabric, and data mesh architectures?

Challenges with Current Data Platforms: What are the common challenges faced by organizations using traditional data platforms like data warehouses and data lakes?

Data Mesh Benefits: How can data mesh address the challenges faced by traditional data platforms, and what are the potential benefits it offers?

Operationalizing Data Mesh: What are the key considerations and best practices for operationalizing a data mesh within an organization?

Operating Model Development: How can organizations develop an effective operating model to guide the implementation and management of a data mesh?

Case Studies: What are the experiences and lessons learned from organizations that have successfully implemented data mesh architectures?

Future Trends: What are the emerging trends and technologies that will shape the future of data mesh and its applications?

In addition to these strategic important key questions lot of detail questions have been considered as a part of this research and has been detailed out in Chapter I.

CHAPTER II:

REVIEW OF LITERATURE

2.1 Theoretical Framework

A detail literature review shows (Panwar et al.,2020) that past studies are primarily focused on understanding and the implementation of centralized data solutions such as Data Warehouse, Data Lake based data management (Kachaoui et al.,2022). As per the Thoughtworks (www.martinfowler.com,2019), the current data platforms that are based on the centralized data lake architecture are becoming monolithic and it is becoming difficult to manage and scale the data platforms. To address this important challenge, there is a need for todays and future data management demands to shift from the centralized paradigm of a data lake or its predecessor data warehouse to modern distributed architecture (Skhiri et al.,2020). Every company wants to become data-driven organization (Aftab et al.,2018) and there is a need of doing the innovation into data field by considering business domains as data domains, applying platform thinking to create self-service data infrastructure and treating data as a product to serve other downstream applications as consumers of the data products.

2.2 Introduction

Data is important to every business. Many organizations generate the data every minute and data management become more and more critical. Therefore, there is a need to understand current business challenges related to the enterprise data management and how the distributed data-mesh using multiple cloud strategy would help to address these challenges (Saranya et al.,2021).

Four to five decades before the data was small in terms of volume, variety & velocity because of the lack of internet and on-line services. However due to continuous innovation across different industry spread across the globe, lot of new applications, systems and new business model got invented which increased the consumers in billions. All these factors are contributing to continuously increasing the data.

Earlier banks were storing their customer data into structured format (i.e., standard rows and columns) using Relational Databases, but later invention happened, and bank started storing the semi-structured data into NOSQL databases and finally another category of data invented as un-structured data and bank started storing the unstructured data into Data Lake. Down the line bank started using the Data Lake as a data repository to store all type of data into it (Yeung et al.,2019).

A detail literature review shows (Panwar et al.,2020) that past studies are primarily focused on understanding and the implementation of centralized data solutions such as Data Warehouse, Data Lake based data management (Kachaoui et al.,2022). As per the Thoughtworks (www.martinfowler.com,2019), the current data platforms that are based on the centralized data lake architecture are becoming monolithic and it is becoming difficult to manage and scale the data platforms. To address this important challenge, there is a need for todays and future data management demands to shift from the centralized paradigm of a data lake or its predecessor data warehouse to modern distributed architecture (Skhiri et al.,2020). Every company wants to become data-driven organization (Aftab et al.,2018) and there is a need of doing the innovation into data field by considering business domains as data domains, applying platform thinking to create self-service data infrastructure and treating data as a product to serve other downstream applications as consumers of the data products.

Data Mesh is a relatively new concept and is still evolving, so there is ongoing discussion and experimentation in the community around best practices and how to operationalize it, improve it and scale it across the organization. Though Data Mesh is a relatively new concept, but it has been gaining traction in recent years as organizations seek to improve the scalability, security, and governance of their data systems. (James Serra,2021)

Banks and other organizations have typical challenges such as data silos, lack of data autonomy, data security, data governance & data latency related to the enterprise data using centralized data management and centralized data teams. Data Mesh is one of the options to address the problems of centralized enterprise data management.

It's important to note that Data Mesh is a relatively new concept and trends may change over time. Organizations needs guidance framework to evaluate their needs and constraints before deciding on the best approach for their organization and keep an eye on new trends in Data Mesh. (T. Priebe et al.,2021)

During the detail literature review different relevant literatures are thoroughly reviewed, inspected, and evaluated. Several themes such as Data types, Roles related to Data Management, Data Management model, Different types of clouds such as AWS, Azure, GCP and OCI, Data Governance, Data Architecture have been thoroughly reviewed. Below mentioned themes or theories are key themes and relevant to the Data Mesh, multi-clouds and its operating model.

2.3 Themes

The detail literature review conducted in structured approach using appropriate themes which are relevant to the study. Follow themes are important to explore in detail to provide the operating model as an outcome of this research study.

2.3.1 The Data Platform Architecture

Introduction

A data platform is an integrated set of technologies that collectively meets an organization's end-to-end data needs. It enables the acquisition, storage, preparation, delivery, and governance of your data, as well as a security layer for users and applications.

A data platform is a complete solution for ingesting, processing, analyzing and presenting the data generated by the systems, processes and infrastructures of the modern digital organization. While there are many point solutions and purpose-built applications that manage one or more aspects of the data puzzle effectively, a true data platform provides end-to-end data management (Splunk,2021).

Data Platform as Important Foundation to Data Mesh

Data platforms include data storage, servers and data architecture. Beyond that, there's data ingestion needs, data consolidation and the ETL process. Businesses regularly face challenges with data management, including the unification of disparate data types housed in various silos, data lakes and on-premises servers (Snowflake,2022).

The goal of a data platform is to deliver real-time business insights through analytics in a cost-efficient, scalable and secure manner. Today's most efficient platforms can be hosted across disparate geographies and in cross-cloud (or multi-cloud) environments to meet unique local or line-of-business requirements and to strengthen business continuity plans. Data Mesh is an evolving way to implement the Data Platform mainly in distributed fashion using single cloud or multi-clouds such as AWS, Azure, GCP or OCI.

Data Platform generations

The Data Platform Architecture has been evolved as First Generation, Second Generation & Third Generation (Dehghani,2019).

The first generation: In the first generation we have used the proprietary enterprise data warehouse and business intelligence platforms; solutions with large price tags that have left companies with equally large amounts of technical debt; There was a technical debt in terms of hundreds of unmaintainable ETL jobs, tables and reports that only a small group of specialized people in the organization understand, resulting in an under-realized positive impact on the business. This generation was mainly focusing on the structured data and centralized data management solution.

The second generation: This generation introduced big data ecosystem with focus on the Data Lake as a way of managing the large enterprise data. This data architecture is handling complex and large data management across the organization as a centralized solution by running thousands of complex and long running jobs by central team of hyper-specialized data engineers, Due the centralized nature, these data lake architectures are becoming data lake monsters which are over promised and under realized.

The third generation: The third-generation data platform is the current data platform which most of the banks are using. The third-generation platform is actually more or less like the previous generation but used the modern techniques such as real time streaming of the data, unifying the batch and stream processing for data transformation using different opensource frameworks as well as fully leveraging the cloud based managed services for the ingestion, integration, storage, process, transformation and visualization. It is obvious that the third-generation data platform is addressing most of the gaps of the previous generations such as real-time analytics, as well as reducing the cost of managing big data infrastructure (Munshi et al.,2018). However, it suffers from the similar problems just like monolithic applications in terms of centralized management & maintenance (Giebler et al.,2020).

Technology is advancing rapidly over the past decade (Tolstoy,2016). While the technology advances over the past decade have addressed the scale of volume of data and data processing compute, however there are other dimensions and they have failed to address scale in those dimensions such as: volume of changes in the data landscape, proliferation of sources of data, diversity of data use cases and users, and also speed of response to change (Zhao et al.,2021).

Data mesh addresses these dimensions, founded in four principles such as domain-oriented decentralized data ownership and architecture, data as a product, self-serve data infrastructure as a platform, and federated computational governance (Skhiri et al.,2020). Each principle drives a new logical view of the technical architecture and organizational structure (Park et al.,2020).

2.3.2 The Centralized or Decentralized architecture

Introduction

One of the biggest differences between the Data Mesh based data platform and other data platform architectures is a data mesh is a highly decentralized distributed data architecture as opposed to a centralized monolithic data architecture based on a data warehouse and/or a data lake. (Perepelkin et al.,2019)

A centralized data architecture means the data from each domain/subject (i.e. payroll, operations, inventory, finance) is copied to one location (i.e. a data lake under one storage account), and that the data from the multiple domains/subjects are combined to create centralized data models and unified views. It also means centralized ownership of the data (usually IT). This is the approach used by a Data Fabric.

A decentralized distributed data architecture means the data from each domain also called as business domain is not copied but rather kept within the domain (each domain/subject has its own data lake under one storage account) and each domain has its own data models. For example, in bank, each Line of Business (LOB) of Bank has its own data model with data ownership. It also means distributed ownership of the data, with each domain having its own owner. (Serra, 2021).

Data Platform as decentralized better than centralized.

The most important thing to understand is that a decentralized solution is not for smaller organizations, only for really big organizations that have very complex data models, high data volumes in petabytes or terabytes, and many data domains. It means at least for 70% of organizations, a decentralized solution would be overkill. Therefore, implementing Data Mesh as a decentralized data platform is the most important aspect and an operating model to roll it out across the organization is another important aspect where very limited information and research has been done and it the main goal of my thesis to provide effective and efficient Data Mesh best practices and operating model for the business-people of the banks or organizations. (Serra, 2021)

2.3.3 Data Mesh and Data Fabric

Introduction

Faster data access and easier collaboration among data teams are two key factors that help drive productivity for most data-driven organizations. However, achieving this becomes more complex with the exponential growth of data as business needs grow. One way to combat this is with architectural patterns that support effective data management. Data mesh and data fabric are two approaches to building a data architecture. Although the methods differ in operation and storage, both aim to address common challenges like data silos, lack of easy access to organizational data, and data management.
(Anderson ,2022)

Difference between Data Mesh and Data Fabric

Data mesh and data fabrics are data architecture approaches organizations adopt for building a scalable, easily accessible, and better-managed data system. It is important to note that there is no one single vendor who can provide a data fabric or data mesh. Both approaches serve to address the following data management principles:

- Fast and easy data access for those with and without technical expertise
- The exponential growth rate of data
- Heterogeneous nature of data sources

Effective data management and governance across data workflow
(Anderson ,2022)

Data Mesh and Data Fabric are both approaches to managing data in large, complex organizations, but they have some key differences:

Scale: Data Mesh is focused on breaking down monolithic data systems into smaller, independent services, while Data Fabric is focused on creating a unified, centralized data infrastructure that can scale to meet the needs of the organization.

Autonomy: Data Mesh is designed to give teams more autonomy over their own data domains, while Data Fabric is focused on creating a centralized data infrastructure that can be managed by a dedicated team.

Data Governance: Data Mesh is more focused on decentralizing data ownership and management, allowing teams to make decisions about how to use their own data. Data Fabric, on the other hand, is more focused on centralizing data governance to ensure compliance with regulations and data security.

Flexibility: Data Mesh is designed to be more flexible, allowing teams to experiment with new data technologies and approaches, while Data Fabric is focused on creating a stable, unified data infrastructure.

Latency: Data Mesh can improve data access and reduce latency by breaking down data silos and giving teams faster access to the data they need. Data Fabric, on the other hand, is focused on creating a unified data infrastructure that can handle large amounts of data and can have a high data latency. (White,2022)

In summary, Data Mesh is more decentralized, focused on giving teams more autonomy and flexibility, while Data Fabric is more centralized, focused on creating a stable and scalable data infrastructure with centralized governance. The below table shows the specific comparison in brief way.

Feature	Data Mesh	Data Fabric
Centralization	Decentralized	Centralized
Governance	Domain-based	Centralized
Scale	Highly scalable	Scalable, but may require centralized management for large-scale deployments
Flexibility	Highly flexible, allowing for experimentation and innovation within domains	Less flexible due to centralized control

Latency	Can reduce latency by distributing data processing and access	May have higher latency due to centralized data access and processing
Complexity	Can be complex to implement due to the decentralized nature	Can be complex due to the centralized infrastructure and management requirements
Cost	Can be cost-effective due to optimized resource utilization and distributed ownership	May require significant upfront investment and ongoing management costs

Table 2-1 Difference between Data Mesh and Data Fabric

Data Mesh: Ideal for organizations with complex data landscapes and a need for decentralized ownership and governance. Offers high scalability, flexibility, and agility.

Data Fabric: Suitable for organizations that require a centralized, unified data infrastructure with strong governance and security. May be more complex to implement and manage.

The choice between data mesh and data fabric depends on the specific needs and priorities of the organization. Factors such as data volume, complexity, governance requirements, and desired level of autonomy should be considered when deciding. However, the proposed operating model in my study can be used to implement both architectural styles.

2.4 Summary

Through the literature review we can conclude that change is constant, and Data is changing and growing constantly in terms of volume, variety and its velocity. We all know the famous line, “Data is Oil”. It means that data is the most important aspect of every business. However, organizations are finding it difficult to manage the huge volume of enterprise data using centralized approaches such as centralized Data

Lake, Data Warehouse or Data Lake House with centralized Data Management teams.

CXO has typical challenges such as data silos, lack of data autonomy, data security, data governance & data latency related to the enterprise data using centralized data management and centralized data teams. Data Mesh is one of the options to address the problems of centralized enterprise data management.(Panwar,2020).

Data Mesh is a way of organizing and managing data in large, complex organizations. It is an architectural pattern. It is based on the idea of breaking down monolithic data systems into smaller, independent services, each with its own data domain and ownership. This allows for greater scalability, flexibility, and autonomy for teams working with data, and can lead to faster innovation and decision-making. Business-people can benefit from Data Mesh by having more control over their own data, faster access to the data they need, and the ability to make data-driven decisions more quickly.

The Data Mesh concept is heavily influenced by the principles of Microservices architecture, in which a monolithic application is broken down into smaller, independently deployable services. Data Mesh is an extension of this concept to data management. (Kachaoui,2022)

As Data Mesh is a relatively new concept, it's still evolving, and there is ongoing discussion and experimentation in the community around best practices and how to implement it effectively. As more organizations adopt Data Mesh, the history of it will continue to develop. Also, there is not sufficient information available in the literatures about operationalizing the Data Mesh across the organization using multi-clouds and the main area of my research and thesis is to work on these gaps and provide a solution to the business people not only to help them to choose the right technologies to implement Data Mesh in distributed fashion using multi-cloud scenarios but also provide a guidance framework to operationalize it across the organization.

The Primary Research method for this study is explained in next chapter and which mainly consists of mainly literature review, in-depth research, survey with different banks data specialist, interviews with business & technical people of banks and IT organizations who are providing data solutions using cloud services (Panwar et al.,2020).

CHAPTER III:
METHODOLOGY

3.1 Introduction

This chapter outlines the research methodology employed in this study, which aims to develop an operating model for the operationalization of modern data platforms utilizing distributed data mesh and multi-cloud architectures. The methodology is designed to provide a rigorous and systematic approach to investigating the research questions and testing the hypotheses.

3.2 Overview of Methodology

The Primary Research method for this study consists of mainly literature review, in-depth research, survey with different banks data specialist, interviews with business & technical people of banks and IT organizations who are providing data solutions using cloud services (Panwar et al.,2020). This study will first review the various types of data challenges for the business people, review current data platform architecture and solutions , identify the opportunities of improvement in terms of modern data platform , determine different types of clouds-based solutions banks are using for their modern data needs, also understand the data mesh and multi-cloud-based solutions (if any) available for the modern data platform (Nargesian et al.,2021). In the second stage of this study, the operating model or framework will be developed mainly for the business-people of banks to accelerate the adoption and scaling of their modern data platform journey. Finally, once the problem statement defined in terms of research questions in the first stage, framework or operating model will be designed in second stage, the detail operating model will be created as a best-in-class guidance framework for the business-people of banks so that business and technology people will understand the Data Mesh with multiple cloud and how to operationalize it for their banks. To summarize, this research provides a valuable framework for organizations to operationalize their modern data platforms using distributed data mesh and multi-cloud architectures. The proposed operating model addresses the key challenges and provides practical guidance for effective implementation. The research will focus on organizations operating in the technology, financial services, or healthcare sectors.

3.3 Research Purpose and Questions

In the contemporary business landscape, data has emerged as a strategic asset, driving innovation, improving decision-making, and fostering competitive advantage. Organizations are increasingly recognizing the imperative of harnessing the power of data to unlock its full potential. However, the proliferation of data sources, increasing data volumes, and the complexity of modern data ecosystems have presented significant challenges in effectively managing and utilizing data.

The Limitations of Traditional Data Architectures: Traditional data architectures, often characterized by centralized data warehouses and data lakes, have struggled to keep pace with the evolving demands of data-driven enterprises. These architectures can be constrained by scalability limitations, data silos, and the complexity of managing multiple data sources. As organizations strive to extract maximum value from their data, there is a growing need for more flexible, scalable, and efficient data management solutions.

The Promise of Distributed Data Mesh and Multi-Cloud: Distributed data mesh and multi-cloud architectures offer a promising approach to address the challenges of modern data management. Distributed data mesh decentralizes data ownership and governance, enabling teams to manage and process data closer to where it is generated. This approach promotes agility, scalability, and data sovereignty. Moreover, leveraging multiple cloud platforms can provide organizations with enhanced flexibility, resilience, and cost-efficiency.

Research Problem

Despite the potential benefits of distributed data mesh and multi-cloud architectures, there is a lack of established operating models and best practices for effectively operationalizing these approaches in real-world environments. Organizations often face challenges in aligning their data strategies with their business objectives, ensuring data quality and governance, and managing the complexities of distributed and multi-cloud environments.

Research Purpose

This research aims to address the critical gap in understanding and operationalizing distributed data mesh and multi-cloud architectures. By developing a comprehensive operating model, this study will provide organizations with a practical framework for effectively leveraging these technologies to achieve their data-driven goals. The findings of this research will contribute to the body of knowledge on data management and architecture, offering valuable insights to practitioners and researchers alike.

The significance of this research lies in its potential to contribute to the following areas:

Bridging the gap: Addressing the existing gap in the literature and practice regarding the operationalization of modern data platforms utilizing distributed data mesh and multi-cloud.

Providing a framework: Developing a comprehensive framework that can guide organizations in establishing effective operating models for their data platforms.

Providing an Organization Chart: Revising the organization roles and responsibilities to plan and implement the Operating Model to address the data challenges across the enterprise.

Research Questions

The research mainly focuses on addressing following key questions.

Data Architecture Comparison: What are the key differences and advantages of data warehouse, data lake, data fabric, and data mesh architectures?

Challenges with Current Data Platforms: What are the common challenges faced by organizations using traditional data platforms like data warehouses and data lakes?

Data Mesh Benefits: How can data mesh address the challenges faced by traditional data platforms, and what are the potential benefits it offers?

Operationalizing Data Mesh: What are the key considerations and best practices for operationalizing a data mesh within an organization?

How to start distributed data mesh adoption journey using proposed operating model in the bank or enterprise?

The above strategic research questions are detailed out below:

What is monolithic Data Architecture?

What are hybrid Data Architectures?

What are the current Data Platform challenges?

What is Domain Driven Design for Data?

What is a Data Mesh?

What are the Data Mesh Principles?

What challenges Data Mesh address?

What are the benefits of the Data Mesh?

What are the Use Cases for the Data Mesh?

What is the difference between Data Mesh and Data Lake?

What is the difference between Data Mesh and Data Fabric?

How to build a data mesh in organization?

How to operationalize the Data Mesh in an organization

What is Operating model?

Why the Operating Model is important?

What are the components of the Distributed Data Mesh Operating Model?

What are the different roles and responsibilities suggested by the Distributed Data Mesh Operating Model?

What is a step-by-step approach recommended by the proposed Operating Model to implement data mesh in small to large scale organization?

How to transform organization data strategy with Data mesh architecture and what will be the role of distributed operating model in this transformation?

How can AZURE support your data mesh architectures?

The research mainly focuses on exploring the below hypothesis.

Hypothesis 1: A well-defined operating model incorporating data governance, data quality, data security, data pipeline management, data lake/warehouse management, and performance optimization is essential for the successful operationalization of modern data platforms.

Hypothesis 2: Organizations can effectively manage federated data governance, quality, and security within a distributed data mesh and multi-cloud environment by implementing robust policies, procedures, and technologies.

Hypothesis 3: The challenges of managing data pipelines, data lakes, and data warehouses in a distributed data mesh and multi-cloud context can be mitigated through effective planning, coordination, and automation.

Hypothesis 4: A well-designed distributed data mesh architecture can provide significant benefits over a monolithic data lake, including improved scalability, agility, data governance, and security.

Hypothesis 5: Organizations can measure and optimize the performance and efficiency of their modern data platforms by utilizing appropriate metrics, analytics tools, and continuous improvement practices.

The research focuses on organizations operating in the technology, financial services, or healthcare sectors.

The scope of the study will be limited to the operational aspects of modern data platforms, excluding strategic considerations such as data monetization and business intelligence.

3.4 Research Design

A mixed-methods research design was employed to combine the strengths of both quantitative and qualitative research approaches. This approach allowed for a comprehensive exploration of the complex issues related to data platform operationalization and provided a rich understanding of the challenges and best practices.

Participants were selected based on their roles within their organizations, including Chief Data Officers (CDOs), Chief Technology Officers (CTOs), data architects, data engineers, and senior executives. These individuals were chosen for their expertise in data management, technology, and leadership.

3.5 Population and Sample

The target population for this research consisted of corporate leaders, enterprise data architects, academicians, and employees from various organizations.

A purposive sampling technique was employed to select participants based on their roles and expertise. The sample included:

Corporate Leaders: 35 executive-level leaders (CDOs, CTOs, enterprise data architects, directors, vice presidents, SVPs) from people-intensive businesses, each leading a team of at least 50 plus people.

Successful Leaders: 25 employees who have been classified as successful leaders within their organizations.

Unsuccessful Leaders: 10 employees who have been classified as unsuccessful leaders or who believe they were unsuccessful.

Academicians: 10 professors and heads of department from various universities, representing a diverse range of academic disciplines.

Sampling Criteria

Diversity: Participants were selected from a variety of organizations, industries, and geographic locations to ensure a diverse sample.

Experience: Participants were chosen based on their experience in leadership roles and their familiarity with the challenges and opportunities of modern leadership.

Perspective: The sample included a mix of successful and unsuccessful leaders to provide a balanced perspective on leadership practices.

3.6 Participant Selection

Participants were selected based on the following criteria:

Role and Expertise: Participants held key roles within their organizations, such as Chief Data Officers (CDOs), Chief Technology Officers (CTOs), data architects, data engineers, and senior executives with direct experience in data management and technology.

Organizational Context: Participants were from a diverse range of banking organizations, varying in size, industry sector, and level of data maturity.

Data Mesh Experience: Participants had experience with implementing or considering distributed data mesh architectures, providing valuable insights into the challenges and benefits.

Operating Model Experience: Participants were involved in developing and implementing operating models for data management within their organizations.

The below table summarize the participant selection criteria.

Criteria	Description
Role and Expertise	Participants held key roles such as CDOs, CTOs, data architects, data engineers, and senior executives with experience in data management.
Organizational Context	Participants were from diverse banking organizations, varying in size, industry, and data maturity.
Data Mesh Experience	Participants had experience with implementing or considering distributed data mesh architectures.
Operating Model Experience	Participants were involved in developing and implementing operating models for data management.

Table 3.1 Summary of the participant selection criteria.

Recruitment Process: Potential participants were identified through a combination of professional networks, industry associations, and online research. They were contacted via email or phone to gauge their interest in participating in the research.

Informed Consent: Prior to participation, potential participants were provided with an informed consent form outlining the purpose of the research, the potential risks and benefits, and their right to withdraw at any time.

Data Saturation: Interviews were conducted until a data saturation point was reached, where no new insights or themes were emerging from the data. This ensured that the sample size was adequate to capture the range of perspectives and experiences within the target population.

Overall, the participant selection process aimed to recruit individuals who could provide valuable insights into the research topic, while also ensuring a diverse and representative sample.

Criterion	Description
Role and Expertise	Participants held key roles such as CDOs, CTOs, data architects, data engineers, and senior executives.
Organizational Context	Participants were from a diverse range of banking organizations, varying in size, industry sector, and level of data maturity.
Data Mesh Experience	Participants had experience with implementing or considering distributed data mesh architectures.
Operating Model Experience	Participants were involved in developing and implementing operating models for data management.
Recruitment	Potential participants were identified through professional networks, industry associations, and online research.
Informed Consent	Participants were provided with an informed consent form before participating.
Data Saturation	Interviews were conducted until a data saturation point was reached to ensure adequate sample size.

Table 3.2 Simplified view of the Participant Selection and Data Saturation

The above table summarizes the key criteria used to select participants for the research, ensuring a diverse and representative sample that could provide valuable insights into the challenges and opportunities associated with distributed data mesh and operating models in banking organizations.

3.7 Instrumentation

The following instruments were used for data collection:

Interview guide: A semi-structured interview guide was developed to ensure consistency and focus during the interviews. The guide included questions related to data platform architecture, governance, security, performance, and challenges faced.

Survey questionnaire: A survey questionnaire was designed to collect quantitative data on various aspects of data platform operations. The questionnaire included Likert scale items, multiple-choice questions, and open-ended questions.

Document analysis checklist: A checklist was used to guide the analysis of organizational documents, focusing on relevant information related to data platform practices and policies.

3.8 Data Collection Procedures

Questionnaire Distribution: Participants were provided with a structured questionnaire via email. The questionnaire covered topics such as organizational context, data platform maturity, challenges and benefits of distributed data mesh, operating model components, and future plans.

Follow-up Interviews: Semi-structured interviews were conducted with selected participants to explore specific areas of interest in more depth. The interviews were conducted either in-person or via video conference.

Data Consolidation: Responses from both the questionnaire and interviews were consolidated and analyzed.

3.9 Data Analysis

Transcription: Interview recordings were transcribed to create a textual representation of the responses.

Thematic Analysis: A thematic analysis approach was employed to identify common themes and patterns within the interview data.

Text Mining Analytics: Text mining techniques were used to extract keywords, phrases, and sentiments from the interview transcripts.

Qualitative Analysis: The qualitative data was analyzed to gain a deeper understanding of the participants' experiences and perspectives.

The qualitative survey was informed by the findings of a previous quantitative research study on distributed data mesh and operating models. This ensured that the qualitative questions were relevant and addressed the specific challenges identified in the quantitative research.

The qualitative survey focused on exploring the following variables:

Organizational context: The size, industry, and level of data maturity of the organization.

Data mesh implementation: The challenges and benefits associated with implementing a distributed data mesh.

Operating model components: The key elements of the operating model, including data governance, security, and performance optimization.

Emerging trends: The impact of emerging technologies and trends on distributed data mesh.

By combining qualitative and quantitative research methods, this study provides a comprehensive understanding of the challenges, benefits, and best practices associated with implementing a distributed data mesh operating model in banking organizations.

3.10 Research Design Assumptions

The research topic is related to operationalize the distributed data mesh either using cloud or on-premises and the enterprise data eco system is large, complex and vast. To get the desired outcome through the research study, this research is based on the following assumptions.

Organizations have a basic understanding of cloud computing and data management concepts.

Organizations are committed to leveraging data as a strategic asset to drive business value.

Organizations have the necessary resources and infrastructure to support the implementation of distributed data mesh and multi-cloud architectures.

3.11 Research Design Limitations

The research is limited to the context of modern data platforms, distributed data mesh architectures, and multi-cloud environments. The findings may not be directly applicable to traditional data warehousing or data lake approaches. Additionally, the study focused on organizations of a certain size and industry sector and might not delve extensively into the broader organizational and cultural implications of adopting these approaches.

3.12 Conclusion

This study employed a qualitative research design, utilizing semi-structured interviews to gather in-depth insights from key stakeholders within banking organizations. The use of interviews allowed for a nuanced understanding of the challenges, experiences, and perspectives of participants regarding distributed data mesh and operating models.

The interviews were conducted in person or via video conference, allowing for detailed discussions and exploration of complex topics. The use of a semi-structured interview guide ensured consistency and focus while also providing flexibility for probing deeper into specific areas of interest.

The purposive sampling technique ensured that the participants were selected based on their relevant roles and expertise, providing a diverse and representative sample. The semi-structured interview approach provided flexibility while maintaining a consistent focus n the research objectives.

The qualitative data collected from the interviews was analyzed using thematic analysis to identify key themes and patterns. This approach allowed for a rich and nuanced understanding of the participants' experiences and perspectives.

Overall, the research design adopted in this study was well-suited to the objectives of the research and provided a robust foundation for the subsequent analysis and findings. The combination of qualitative interviews and thematic analysis allowed for a deep exploration of the complexities and nuances associated with distributed data mesh and operating models in banking organizations.

CHAPTER IV:

RESULTS

4.1 Data Architecture Comparison

What are the key differences and advantages of data warehouse, data lake, data fabric, and data mesh architectures?

Historical Context

The evolution of data architecture has mirrored the growth of data volumes and the increasing complexity of data management needs. As organizations accumulated vast amounts of data from various sources, traditional data warehousing approaches began to face limitations in terms of scalability, flexibility, and time-to-insight.

Data architecture is a blueprint that outlines how data is collected, stored, managed, and accessed within an organization. It encompasses the design, structure, and technology components involved in data management. Data architecture ensures that data is organized, accessible, and aligned with business objectives.

Data Warehouse

The data warehouse is a foundational layer in data management, enabling business intelligence and reporting. Emerging in the 1980s, it was designed as a centralized repository for structured data, providing a unified view of organizational information. Typically organized using a star or snowflake schema, data warehouses simplify querying and analysis.

Operating on a schema-on-write model, the data schema is defined before ingestion, ensuring consistency. By requiring data to be structured and transformed before storage, data warehouses are optimized for analytics, reporting, and decision-making, making them ideal for structured data use cases.

Data Lake

The data lake represents a modern approach to data management, developed in response to the inherent limitations of traditional data warehouses. Studies shows that unlike data warehouses, which require structured data and predefined schemas, data lakes offer a centralized repository for storing vast volumes of raw data in its native format. This flexibility enables organizations to manage diverse data types, including structured, semi-structured, and unstructured data, without the need for upfront schema definitions.

Adopting a schema-on-read model, data lakes allow schemas to be applied at query time, providing greater adaptability for exploratory tasks. This architecture is particularly suited for use cases such as data exploration, experimentation, and advanced analytics. It supports a wide range of applications, including machine learning, data science, and discovery-driven analysis, where the ability to work with raw, unprocessed data is essential.

Data Fabric

Data fabric is a contemporary approach to data management that provides a unified layer for accessing and integrating data from diverse sources. This architecture streamlines data management processes by leveraging advanced capabilities such as metadata management, data governance, and automation. By relying on metadata as a core element, data fabric enables consistent data governance and efficient management across the organization.

A key advantage of data fabric is its ability to facilitate real-time access to data from various sources, enhancing the timeliness and relevance of insights. Additionally, it supports self-service analytics, empowering users to independently access and analyze data without reliance on technical teams. This makes data fabric particularly valuable for organizations requiring seamless integration and management of data from heterogeneous environments to support agile and informed decision-making.

Data Mesh

Data Mesh is the main area of the research study. During the research study, it has been validated that Data Mesh is a decentralized approach for the agile data management for the organization. The conceptual material available on Data Mesh and lacking the expert guidance on its operationalization. The Data mesh introduces a decentralized data architecture that shifts the responsibility for data ownership and management to individual data domains. This approach aligns with domain-driven design principles, organizing data around specific business domains to enhance relevance and accountability. By decentralizing ownership, data mesh empowers domains to manage their data independently while ensuring alignment with overarching organizational goals.

Central to data mesh is the concept of self-service data products, where domains treat data as a product. This enables the creation, consumption, and governance of data products within each domain, fostering innovation and efficiency. Additionally, data mesh emphasizes data interoperability, ensuring that data can be seamlessly shared and integrated across domains. This architecture is particularly suitable for organizations requiring a decentralized, agile approach to data management, enabling them to respond quickly to evolving business needs.

Key Differences:

Feature	Data Warehouse	Data Lake	Data Fabric	Data Mesh
Centralization	Centralized	Centralized	Centralized	Decentralized
Schema	Schema-on-write	Schema-on-read	Metadata-driven	Domain-driven

Data Type	Structured	Structured and unstructured	Structured and unstructured	Structured and unstructured
Use Cases	Reporting, BI, decision-making	Exploration, data science	Integration, self-service analytics	Domain-driven data management, innovation

Table 4.1 Key Differences between Data Architecture styles

Choosing the Right Approach:

Selecting the optimal data architecture for an organization requires a nuanced understanding of its unique needs, data characteristics, and cultural dynamics. The research identified several critical factors that influence this decision, each tied to the organization's goals and operational priorities.

Data maturity is a key determinant, with organizations that exhibit well-developed data management practices often benefiting from architectures like data warehouses or data fabric. These solutions provide structured frameworks and robust governance mechanisms suitable for mature data ecosystems. In contrast, organizations managing diverse and unstructured datasets may find architectures such as data lakes or data mesh more appropriate, given their flexibility in handling a variety of data types.

For organizations prioritizing agility and innovation, data mesh emerges as a preferred choice, offering decentralized ownership and rapid adaptability to evolving needs. Conversely, those with stringent requirements for governance and control may lean towards more centralized approaches like data warehouses or data fabric, which provide enhanced oversight and standardization.

The research further highlights that many organizations benefit from a hybrid approach, combining elements of multiple architectures to address complex and varied requirements effectively. By understanding the distinctions, advantages, and trade-offs of each model, organizations can craft a tailored data architecture strategy that aligns with their objectives while maintaining flexibility to adapt to future challenges.

4.2 Challenges with Current Data Platform

What are the common challenges faced by organizations using traditional data platforms like data warehouses and data lakes?

Traditional data platforms, such as data warehouses and data lakes, have been essential tools for managing and analyzing large datasets. While they have enabled significant advancements in data-driven decision-making, the growing complexity and volume of data have exposed several limitations in these architectures. These challenges span scalability, governance, integration, agility, and cost, prompting organizations to reconsider their data management strategies.

Scalability limitations are among the primary obstacles faced by traditional data platforms. As data volumes grow exponentially, storage constraints in data warehouses can hinder capacity, while processing bottlenecks in ETL (Extract, Transform, Load) processes slow the ingestion and transformation of large datasets. Additionally, as datasets expand, query performance often degrades, increasing the time-to-insight and negatively impacting user experience.

Data governance and quality challenges further complicate data management. Data silos, a common issue in centralized architectures, can lead to inconsistencies and inefficiencies, undermining the effectiveness of data-driven initiatives. Ensuring data quality is particularly difficult when dealing with large and complex datasets, and compliance with data privacy and security regulations can become increasingly onerous in such environments.

Integration and interoperability issues also limit the utility of traditional platforms. The integration of data from diverse sources often proves complex and time-intensive due to differences in data formats and system architectures. Such inconsistencies can lead to data quality problems and inaccuracies in analysis, reducing the reliability of insights derived from these platforms.

Limited flexibility and agility further restrict the responsiveness of traditional data architectures. The reliance on rigid, predefined schemas in data warehouses limits their ability to adapt to evolving business requirements or unstructured data types. Similarly, the lengthy lead times required for analysis and reporting delay decision-making. Many traditional platforms also struggle to handle real-time data streams, which are increasingly essential for modern analytics.

Finally, cost considerations pose significant challenges. The maintenance and scaling of traditional platforms, particularly for large organizations, can be prohibitively expensive. Additionally, the reliance on specific vendors may result in vendor lock-in, restricting flexibility and increasing long-term costs.

These challenges have driven organizations to explore alternative data architectures, such as data mesh and data fabric. These emerging paradigms offer more scalable, flexible, and efficient solutions, addressing the limitations of traditional platforms while meeting the demands of modern data ecosystems.

4.3 How can data mesh address the challenges faced by traditional data platforms.

Data mesh represents a decentralized approach to data management that effectively addresses many of the challenges posed by traditional data platforms such as data warehouses and data lakes. Through the research study, data mesh emerged as a promising solution to scalability, governance, agility, cost, and integration challenges, offering organizations a transformative way to manage and utilize their data assets.

Improved Scalability and Flexibility: One of the key strengths of data mesh is its ability to provide horizontal scalability, allowing organizations to accommodate growing data volumes without extensive infrastructure changes. Unlike traditional platforms that may face bottlenecks, data mesh distributes data management responsibilities across domains, making scaling more seamless. Furthermore, its flexible deployment options enable organizations to implement data mesh across various cloud environments or on-premises, avoiding vendor lock-in and ensuring adaptability to diverse technological ecosystems.

Enhanced Data Governance and Security: The decentralization of ownership in data mesh enhances governance and accountability by empowering individual data domains to manage their data. This approach ensures that data quality is maintained closer to its source, minimizing errors and inconsistencies. Additionally, data mesh promotes enhanced security through the application of domain-specific policies and controls, reducing the risk of breaches and improving overall compliance with regulatory requirements.

Increased Agility and Innovation: By decentralizing data ownership and providing self-service access, data mesh fosters greater agility and innovation. It enables faster development and deployment of data-driven applications, shortening time-to-market and accelerating organizational responsiveness to evolving business needs. The self-service model empowers data analysts and scientists to independently explore data, uncover insights, and drive innovation without reliance on centralized teams.

Reduced Costs: The distributed nature of data mesh optimizes resource utilization, preventing the overprovisioning common in traditional architectures. By leveraging multiple cloud providers, organizations can negotiate cost-effective deals and reduce cloud-related expenses. This decentralized and resource-efficient approach provides a cost advantage while maintaining operational effectiveness.

Improved Data Access and Integration: Data mesh simplifies data discovery and integration by introducing a centralized catalogue of data assets and a standardized framework for sharing and accessing data across domains. This approach resolves the integration complexities associated with traditional platforms, facilitating seamless interoperability and enhancing the accessibility of valuable data resources.

Strengthened Governance and Compliance: The domain-based governance model promoted by data mesh allows for tailored and effective data management practices, aligning with the specific needs of individual domains. Additionally, the implementation of domain-specific security controls ensures better compliance with data privacy and regulatory requirements, a critical challenge for organizations operating in highly regulated environments.

The research study identified data mesh as a solution that not only addresses the limitations of traditional platforms but also empowers organizations to unlock the full potential of their data assets. Its scalability, flexibility, security, and cost efficiency make it a compelling choice for modern data ecosystems. By adopting data mesh, organizations can transition from rigid, centralized models to dynamic, decentralized frameworks that drive innovation and resilience in data management.

4.3 Operationalizing Data Mesh

What are the key considerations and best practices for operationalizing a data mesh within an organization?

Operationalizing a data mesh within an organization requires a structured approach that balances strategic planning, technological implementation, and cultural transformation. The research conducted in this study identified a series of key considerations and best practices essential for successfully embedding a data mesh architecture, ensuring its alignment with organizational goals while addressing potential challenges.

Defining Data Domains: The foundation of a data mesh lies in the clear definition and organization of data domains. This involves identifying domains based on business functions, processes, or data types and assigning ownership of each domain to specific teams or departments. Domain owners are responsible for maintaining data quality, ensuring accessibility, and adhering to established governance practices. Effective governance within domains requires developing robust policies and procedures tailored to the unique requirements of each domain, fostering accountability and transparency.

Implementing Data Platforms and Infrastructure: A critical aspect of operationalizing a data mesh is the selection and deployment of suitable technologies. Organizations must identify data platforms and infrastructure components that align with their operational requirements and long-term goals. Building efficient data pipelines is essential to enable seamless movement of data across systems while ensuring interoperability between domains. This interoperability promotes cross-domain collaboration and facilitates data sharing, a hallmark of the data mesh paradigm.

Establishing Data Governance: Data governance serves as the backbone of a successful data mesh implementation. Organizations must develop comprehensive policies addressing data quality, security, privacy, and compliance to uphold the integrity of the data ecosystem. Clearly defining roles and responsibilities, such as data stewards, owners, and analysts, ensures that governance is enforced effectively. Leveraging governance tools to automate tasks and monitor adherence to policies further strengthens the governance framework.

Fostering a Data-Driven Culture: The research emphasizes that the success of a data mesh is as much about culture as technology. Promoting data literacy across the organization is crucial to ensuring all stakeholders understand the value of data and how to leverage it effectively. Encouraging data-driven decision-making and breaking down silos between domains fosters collaboration and innovation, aligning the organization with the principles of decentralized data ownership and utilization.

Monitoring and Optimization: Continuous monitoring and optimization are necessary to maintain the effectiveness of a data mesh implementation. Defining key performance indicators (KPIs) allows organizations to measure success in areas such as data quality, accessibility, and cost-efficiency. Performance monitoring can identify bottlenecks or inefficiencies, enabling targeted adjustments to the architecture and processes to enhance overall functionality.

Addressing Challenges and Risks: Operationalizing a data mesh also requires proactive identification and mitigation of potential risks. Security measures must be robust to protect sensitive data, particularly in decentralized environments. Effective data governance practices help maintain data quality, while addressing performance bottlenecks ensures seamless operations. Organizations must also manage costs efficiently by optimizing resource utilization and leveraging cloud-based services to minimize expenses.

Ensuring Scalability and Flexibility: A successful data mesh architecture is designed with scalability and flexibility in mind. It must accommodate future growth and changes in data volumes, formats, and technologies while supporting diverse deployment options. By prioritizing adaptability, organizations can ensure their data mesh remains relevant and effective in dynamic business environments.

Continuous Improvement: Operationalizing a data mesh is not a one-time process but an ongoing journey. Regular assessments of the implementation help identify areas for improvement, enabling organizations to refine their architecture and practices over time. This adaptability ensures the data mesh evolves alongside organizational needs, sustaining its benefits and relevance.

Through the research study, these considerations and best practices were identified as critical enablers of successful data mesh adoption. By systematically addressing these factors, organizations can operationalize data mesh effectively, reaping its benefits of improved governance, agility, and innovation while positioning themselves to adapt to future challenges and opportunities in data management.

4.4 Operating Model Development

How can organizations develop an effective operating model to guide the implementation and management of a data mesh?

To successfully implement and manage a data mesh, organizations require a comprehensive operating model. Such a model serves as a strategic framework, outlining roles, responsibilities, governance structures, and processes essential for aligning data mesh principles with organizational objectives. The research undertaken in this study identified key components and best practices that organizations can adopt to build an effective operating model.

Key Components of Data Mesh Operating Model:

Data Domain Ownership: Defining data domains is fundamental to the data mesh approach. Domains should align with business functions or data types to ensure relevance and manageability. Ownership of each domain must be clearly assigned to specific teams or departments. Empowering these data domain owners with authority and resources to manage their data fosters accountability and enhances data quality, governance, and accessibility.

Data Governance: Governance plays a pivotal role in ensuring the integrity and security of the data mesh. Organizations must establish robust governance policies and procedures addressing data quality, security, and compliance. This includes defining roles and responsibilities for key stakeholders, such as data stewards, owners, and analysts, and deploying governance tools and frameworks to automate and enforce policies.

Data Platforms and Infrastructure: The selection of appropriate data platforms and technologies is critical for supporting a distributed data mesh architecture. Organizations should design and implement data pipelines to enable seamless data flow between systems, ensuring interoperability and compatibility across diverse data sources. This infrastructure underpins the scalability and efficiency of the data mesh.

Self-Service Access: A core principle of data mesh is empowering data consumers with self-service capabilities. Providing user-friendly tools and interfaces enables stakeholders to access and analyze data independently. Establishing comprehensive data catalogues and metadata management practices enhances discoverability and usability, reducing dependency on technical teams.

Organizational Structure: Adopting a data mesh requires restructuring the organization to support its principles. Organizations must identify key roles, such as data product owners, domain leads, and platform engineers, and define the responsibilities of each. Creating an organizational chart to assign roles and planning training initiatives ensures readiness and competency among employees to leverage the data mesh approach effectively.

Collaboration and Communication: Effective collaboration between data domains is vital to prevent silos and promote knowledge sharing. Organizations should establish structured communication channels and coordination mechanisms among data teams and stakeholders, ensuring alignment and seamless integration across domains.

Performance Monitoring and Optimization: To measure success, organizations should implement monitoring and analytics tools to track data mesh performance. These tools provide insights into metrics such as data quality, accessibility, and cost-efficiency. Continuous optimization of the data mesh architecture and processes helps to improve performance and adapt to evolving business needs.

Best Practices for Developing An Operating Model: The research study highlighted several best practices critical for building a robust and effective operating model:

Align with Business Objectives: The operating model should support overarching business goals and strategies, ensuring alignment with organizational priorities.

Involve Stakeholders: Engaging stakeholders across functions fosters collaboration and secures buy-in, creating a shared vision for the data mesh implementation.

Iterative Development: Organizations should adopt an iterative approach, starting with a pilot project to refine the model based on feedback and observed outcomes.

Flexibility and Adaptability: Designing for flexibility ensures the operating model can evolve with changing business requirements and technological advancements.

Continuous Improvement: Regular reviews and updates of the operating model ensure it remains effective, addressing emerging challenges and seizing new opportunities.

This research emphasizes the importance of a well-structured operating model for the successful implementation of a data mesh. By adhering to these components and best practices, organizations can create a scalable, flexible, and efficient data management framework. Such a framework not only addresses the limitations of traditional data architectures but also unlocks the full potential of data mesh, driving innovation and supporting data-driven decision-making at all levels of the organization.

4.5 Case Studies Exploration

What are the experiences and lessons learned from organizations that have successfully implemented data mesh architectures?

The implementation of data mesh architectures in the banking sector highlights a transformative shift in data management strategies. A limited but growing number of banking organizations have successfully adopted this decentralized approach, addressing long-standing challenges and unlocking new opportunities. The research conducted for this study analyzed multiple case studies, providing critical insights into the experiences, benefits, challenges, and lessons learned.

Case Study 1: A Large European Bank

Challenges: The bank faced significant issues with data silos, poor data quality, and prolonged time-to-insight, hampering its ability to deliver timely and accurate insights for decision-making.

Implementation: To address these issues, the bank adopted a decentralized data mesh approach. By empowering individual business units to own and manage their data, the bank aimed to improve accountability and reduce operational bottlenecks.

Benefits: The implementation led to improved data quality, faster time-to-market for new products, and enhanced customer experiences.

Findings: However, the lack of robust governance practices extended the rollout timeline, doubling the initially estimated duration. This underscores the critical need for strong governance frameworks in data mesh implementations.

Case Study 2: A Mid-Sized U.S. Bank

Challenges: The bank struggled with data integration complexities and scalability limitations, which impeded its ability to manage growing data volumes efficiently.

Implementation: A cloud-based data mesh architecture leveraging AWS services was deployed, providing a scalable and flexible foundation for decentralized data management.

Benefits: The bank achieved substantial cost savings, improved performance, and enhanced data security.

Findings: The shift from a centralized to decentralized model required significant cultural change, particularly in altering mindsets across teams. Despite these challenges, the adoption of AWS services streamlined the transition and supported rapid implementation.

Case Study 3: A Global Financial Services Company

Challenges: Managing data across diverse geographies and regulatory environments presented significant operational and compliance challenges.

Implementation: The company adopted a federated data mesh architecture, enabling decentralized governance and management while maintaining compliance with local regulations.

Benefits: The approach improved data governance, reduced operational costs, and accelerated the time-to-market for new products.

Findings: Key lessons from this implementation include the importance of:

Empowering data domains with ownership and management responsibilities.

Establishing strong governance to ensure data quality, security, and compliance.

Carefully selecting technologies to support scalability and efficiency.

Facilitating cultural shifts towards decentralized data management and data driven decision-making.

Case Study 4: A Retail Bank in Asia

Challenges: The bank faced entrenched data silos, poor data quality, and slow insights for customer analytics, hindering its ability to offer personalized services.

Implementation: A cloud-native data mesh architecture was adopted, empowering business units to take ownership of their data.

Benefits: The bank realized significant improvements in customer segmentation, targeted marketing, and personalized product recommendations.

Findings: The initiative was paused due to high attrition rates, lack of internal cloud expertise, and insufficient data skills, highlighting the importance of workforce readiness and resource planning in implementing data mesh.

Case Study 5: A Large Investment Bank

Challenges: The bank faced data integration issues and struggled to maintain regulatory compliance across its operations.

Implementation: A federated data mesh approach was adopted, enabling decentralized governance while ensuring alignment with regulatory standards.

Benefits: Improved data quality, enhanced regulatory compliance, and accelerated risk management processes were achieved.

Findings: Strong governance frameworks, executive support, and internal expertise were key enablers of success, allowing the bank to achieve its goals within the planned timeline of two and a half years.

Case Study 6: A Community Bank in the United States

Challenges: The bank's limited analytics capabilities restricted its ability to compete with larger institutions.

Implementation: A pilot data mesh initiative was launched to enable advanced analytics and data-driven decision-making.

Benefits: The pilot successfully demonstrated the potential for enhanced operational efficiency and customer satisfaction.

Findings: However, the pilot incurred triple the expected cost and time due to inadequate skills and strategic direction, emphasizing the importance of careful planning and skill development.

Case Study 7: A European Bank with Multiple Subsidiaries

Challenges: Consolidating data across subsidiaries and ensuring data consistency were persistent issues.

Implementation: The bank initiated a data mesh implementation journey aimed at enabling data sharing and integration across subsidiaries.

Benefits: Expected outcomes include improved operational efficiency, reduced costs, and enhanced regulatory compliance.

Findings: The initiative remains in progress, with executives focused on addressing the complexities of operationalizing distributed data mesh systems across a diverse ecosystem.

Case Study 8: A Digital-First Bank

Challenges: Rapid scaling of data infrastructure was essential to support the bank's growing customer base and expanding product offerings.

Implementation: A cloud-native data mesh architecture leveraging Azure services was piloted to enable rapid scalability and flexibility.

Benefits: The proof-of-concept demonstrated the viability of data mesh principles, enabling the rapid development of key data products.

These case studies highlight the diverse experiences of banking organizations in adopting data mesh architectures. While the benefits include improved governance, scalability, and innovation, the challenges often revolve around cultural change, governance, and skill gaps. The research underscores that a phased and well-governed approach, coupled with stakeholder buy-in and skill development, is essential for success. By learning from these experiences, organizations can effectively navigate their data mesh journey, realizing the transformative potential of this decentralized data management paradigm.

4.6 Future Trends

The data mesh landscape is poised for transformative growth, shaped by continuous technological innovations, shifting business demands, and emerging industry trends. This section synthesizes insights from the research, offering a forward-looking perspective on how data mesh is expected to evolve and expand its role in data management and analytics.

Increased Adoption of Cloud-Native Technologies: The research highlights the growing synergy between data mesh architectures and cloud-native technologies as organizations seek scalable and cost-efficient solutions.

Serverless Computing: The adoption of serverless functions for data processing and analytics is projected to increase. These technologies enable organizations to dynamically scale operations, reducing infrastructure costs while enhancing responsiveness.

Containerization: Technologies such as Docker and Kubernetes are expected to become integral to the deployment and management of data mesh components, ensuring consistency and portability across diverse environments.

Cloud-Native Data Platforms: Integrating data mesh with cloud-native platforms, such as modern data lakes and warehouses, will provide organizations with enhanced capabilities for distributed data management.

Enhanced Data Governance and Security: The future of data mesh will be shaped by advancements in governance and security frameworks, as identified during the research.

AI-Powered Data Governance: Artificial intelligence and machine learning will play a pivotal role in automating governance tasks, including data quality checks, anomaly detection, and compliance enforcement.

Data Privacy and Compliance: As regulatory demands increase, data mesh implementations will need to integrate robust privacy measures and ensure adherence to evolving compliance standards. This will enhance the trustworthiness of distributed data ecosystems.

Integration with Emerging Technologies: Research findings indicate that data mesh architectures will increasingly interface with cutting-edge technologies to unlock new analytical and operational possibilities.

Artificial Intelligence and Machine Learning: The integration of data mesh with AI and ML frameworks will enable organizations to harness advanced analytics, automate decision-making, and scale predictive modeling efforts.

Internet of Things (IoT): The proliferation of IoT devices is expected to amplify the need for data mesh frameworks capable of managing and analyzing vast streams of device-generated data.

Edge Computing: Extending data mesh to edge computing environments will support real-time processing and analytics, ensuring timely insights closer to the data source.

Democratization of Data: The democratization of data was a recurring theme in the research, with data mesh playing a central role in empowering data consumers across organizations.

Self-Service Data Access: Data mesh will continue to focus on enabling self-service access to data, allowing teams and individuals to independently explore and analyze datasets.

Data Literacy: Investments in data literacy programs will be essential to equip employees with the skills needed to interpret and use data effectively, fostering a culture of data-driven decision-making.

Hybrid Data Mesh Architectures: As organizations navigate diverse business requirements, the emergence of hybrid data mesh architectures was identified as a key trend during the study.

Combining Centralized and Decentralized Elements: Hybrid architectures that blend centralized control with decentralized data ownership will address specific organizational needs, ensuring both governance and flexibility.

Multi-Cloud Data Mesh: Extending data mesh capabilities to support multi-cloud environments will enhance resilience and provide organizations with greater flexibility in their data strategies.

The research underscores that the future of data mesh will be defined by its adaptability and integration with advanced technologies. Trends such as cloud-native adoption, enhanced governance, and the democratization of data signal a broader applicability across industries and use cases. By leveraging these advancements, organizations can transform their data ecosystems, unlocking new opportunities for innovation, scalability, and operational excellence. As the data mesh model continues to mature, it will remain a cornerstone for addressing the complexities of modern data management.

4.8 Summary of Findings

The research study incorporated comprehensive data collection methods, including in-depth interviews, surveys, and focus group discussions with key stakeholders across banking and telecommunications sectors. Participants included senior executives such as CEOs, CTOs, and CDOs, as well as Enterprise Data Architects, Data Engineers, Data Scientists, Analytics professionals, Data Project Managers, centralized data teams, and other critical roles. This approach ensured a holistic understanding of the current state of data mesh adoption, its perceived benefits, and the challenges encountered.

While there is growing awareness about data mesh in the banking industry, with several banks embarking on its adoption, the telecommunications industry is predominantly in the early stages of exploring its feasibility. The following sections detail the key findings derived from the research.

Data Mesh Adoption Stages: The adoption of data mesh in the banking industry was found to vary significantly, reflecting different levels of organizational readiness and maturity.

Early Adopters (15%): A small yet significant proportion of banks have fully implemented data mesh architectures, showcasing progress in decentralizing data ownership and governance. These organizations reported measurable benefits in agility and innovation.

Exploratory Phase (30%): Nearly a third of banks are actively experimenting with data mesh through pilot projects, aiming to evaluate its technical feasibility and business value.

Planning Stage (25%): A quarter of the banks have articulated formal strategies for data mesh implementation and are in preparatory stages, focusing on resource alignment and stakeholder engagement.

Hesitant (10%): A smaller segment expressed reservations about adopting data mesh due to concerns surrounding cultural transformation, technical complexity, and financial implications.

Unaware (20%): A noteworthy portion of banks demonstrated limited awareness about the concept and benefits of data mesh, indicating the need for greater advocacy and education.

Data Mesh Benefits: The study identified tangible benefits reported by banks that have adopted or explored data mesh.

Improved Data Agility (80%): Data mesh enabled banks to adapt more rapidly to evolving business needs, reflecting enhanced operational responsiveness.

Enhanced Data Governance (75%): Improved governance frameworks led to better data quality and compliance across decentralized domains.

Reduced Time-to-Market (65%): Banks reported accelerated delivery of data-driven projects, shortening the time-to-market for innovative initiatives.

Increased Innovation (60%): The data mesh model fostered a culture of innovation by empowering teams with greater autonomy and self-service access to data.

Cost Savings (55%): Significant cost reductions were achieved by optimizing data management processes and leveraging efficient infrastructure solutions.

Challenges and Obstacles: Despite the benefits, several challenges were identified, emphasizing the complexity of transitioning to a data mesh architecture.

Cultural Change (45%): Resistance to change emerged as a critical barrier, with many organizations struggling to adopt a decentralized mindset.

Technical Complexity (40%): Integration challenges, security concerns, and system performance issues posed significant obstacles during implementation.

Data Governance (35%): Establishing effective governance in a decentralized framework was a recurring challenge, particularly in maintaining data consistency and quality.

Talent Acquisition (30%): The scarcity of skilled professionals with expertise in data mesh concepts hindered adoption efforts in several banks.

Cloud Adoption: Cloud infrastructure emerged as a critical enabler for data mesh implementations, with banks adopting diverse strategies based on their operational requirements.

AWS and Azure (70%): A majority of banks have chosen AWS or Azure for their cloud environments, leveraging their robust services for distributed data management.

Hybrid Cloud (25%): A significant number of banks are adopting a hybrid approach, combining on-premises systems with cloud solutions for enhanced flexibility.

Multi-Cloud (10%): A smaller segment is exploring multi-cloud strategies to avoid vendor lock-in and increase system resilience.

Operating Model Development: The research revealed varying degrees of maturity in the development of operating models for data mesh.

Lack of Established Models (50%): Half of the banks are in the early stages of defining and formalizing operating models tailored to their unique organizational structures and goals.

Domain-Driven Design (40%): Many banks are prioritizing domain-driven design principles to align data domains with business functions and improve accountability.

Self-Service Access (35%): Enabling self-service access to data has emerged as a priority for enhancing business agility and democratizing data usage.

The research findings underscore a dynamic landscape of data mesh adoption, with the banking industry leading the way in piloting and implementing this paradigm. While clear benefits such as improved agility, governance, and innovation were identified, significant challenges remain, particularly in cultural adaptation, technical execution, and talent acquisition.

The insights gleaned from this study provide a robust foundation for understanding the state of data mesh adoption and offer actionable recommendations for organizations considering this transformative approach. As the journey progresses, continuous learning and adaptation will be key to realizing the full potential of data mesh architectures.

4.9 Conclusion

The findings from this research study highlight a growing interest in the adoption of data mesh architectures within the banking sector. The insights derived from interviews, surveys, and case studies reflect an increasing awareness of the potential benefits that data mesh can offer in enhancing agility, improving governance, and fostering innovation. Despite this enthusiasm, the study has also identified significant challenges that must be addressed to realize the full potential of this paradigm shift.

Key obstacles include the need for a cultural transformation to support decentralized data ownership, the technical complexities associated with integrating and managing distributed systems, and the critical importance of establishing robust data governance frameworks. These challenges underscore the necessity for a well-planned and strategically executed approach to data mesh implementation.

The research further emphasizes the importance of investing in talent development to build the necessary skills and expertise for operating within a data mesh framework. Organizations must foster a culture of continuous learning and collaboration, ensuring that teams are equipped to navigate the complexities of decentralized data management.

Additionally, the findings suggest that successful data mesh adoption requires a commitment to organizational transformation. This involves not only leveraging advanced technologies but also rethinking traditional approaches to data architecture, governance, and team collaboration.

Overall, this research provides a comprehensive understanding of the current state of data mesh adoption, offering valuable lessons and actionable recommendations for organizations embarking on this journey. By addressing the identified challenges and leveraging the insights gained from early adopters, organizations can position themselves to harness the transformative potential of data mesh to achieve their strategic goals.

CHAPTER V:

DISCUSSION – DATA MESH OPERATING MODEL

5.1 Discussion of Results

This chapter provides a comprehensive discussion of the research findings, draws conclusions based on the analysis, and explores the implications for banking organizations. The discussion will address the research questions posed in Chapter IV and will delve into the significance of the findings.

From the extensive literature review that was carried out as part of this research it was prominent that different banking organizations are at the different stages of their Data Mesh adoption journey with their own unique business drivers, priorities, challenges and outcomes.

5.2 Data Architecture Comparison

Data architecture is the foundation for managing organizational data assets, outlining processes for collecting, storing, organizing, and accessing data to meet business needs. This study analyzes the evolution of data architectures, emphasizing how data mesh addresses challenges inherent in earlier approaches.

The Data Warehouse Era: The first generation of data architectures revolved around proprietary data warehouses and business intelligence platforms. These centralized systems effectively handled structured data but were constrained by complexity, high costs, and limited flexibility. Maintenance challenges, rigid ETL processes, and a narrow focus on structured data restricted adaptability to evolving business needs and emerging data types.

The Data Lake Revolution: The second generation introduced data lakes, providing the ability to store raw, unstructured data at scale. While data lakes allowed more exploratory analytics and lowered storage costs, they were hampered by governance challenges, disorganized data (leading to "data swamps"), and centralized management bottlenecks that limited performance and scalability.

Modern Data Platforms: Modern data platforms emerged to address these issues, integrating advancements in real-time analytics, cloud computing, and unified batch-stream processing. These systems reduced costs and enabled low-latency insights, supporting structured and unstructured data management. However, their centralized nature still posed challenges in scalability, governance, and operational autonomy, limiting their effectiveness in increasingly complex data environments.

The Data Mesh Paradigm: Data mesh represents a transformative approach to data architecture, emphasizing decentralization and domain-oriented design. Unlike earlier architectures, data mesh distributes ownership across business domains, aligning data management with organizational priorities and promoting operational autonomy. Data is treated as a product, with clear standards, APIs, and governance mechanisms to ensure accessibility and usability.

Self-service tools empower stakeholders to access and analyze data independently, reducing reliance on centralized teams. A federated governance model ensures compliance and security while balancing domain autonomy with overarching oversight.

Organizations adopting data mesh reported enhanced scalability, flexibility, and operational efficiency. Decentralized ownership reduced bottlenecks and encouraged innovation, while improved governance frameworks ensured quality and compliance.

Data mesh addresses the persistent limitations of earlier architectures, offering a modern solution for organizations managing complex and distributed data ecosystems.

This study explored that data mesh provides a strategic advantage by integrating lessons from past architectures and fostering a more agile, scalable, and effective approach to data management.

5.3 Current Data Platform challenges

Traditional data platforms, such as data warehouses and data lakes, have long been the backbone of data management in organizations. However, as businesses increasingly rely on data for decision-making, these platforms are increasingly unable to meet the demands of modern, data-driven enterprises. This section explores the limitations of traditional data platforms and discusses how these challenges can be addressed by adopting a data mesh approach. (C. Giebler et al.,2020)

Bottlenecks in Traditional Data Platforms: A critical limitation of traditional platforms lies in the separation between analytical and operational processing. This division often creates dependencies on centralized teams, which significantly slows down the data lifecycle. Typically, the process begins with analysts and data scientists identifying the data required for their use cases. Centralized data teams are then tasked with extracting this data from operational systems, transforming it to meet analytical needs, and ultimately creating data products for business use. While this pipeline is functional, it is inherently slow and rigid, often resulting in delayed responses to dynamic business requirements.

Key Challenges in Existing Architectures: Several specific challenges exacerbate the inefficiencies of traditional data platforms. First, **data discovery** is often hindered by the existence of silos and inadequate documentation, making it difficult for users to locate relevant datasets. Additionally, **data access** processes can be time-consuming due to stringent security measures and complex approval workflows. This results in delays that can be critical in time-sensitive decision-making scenarios.

Data quality also emerges as a major concern, particularly when organizations rely on multiple sources with inconsistent standards. Transforming and standardizing data to ensure usability demands significant resources. Moreover, centralized teams often lack the **domain-specific knowledge** required to contextualize the data effectively, leading to misaligned outputs and communication gaps. Lastly, the process of coordinating across multiple teams for data acquisition, transformation, and analysis is often inefficient, slowing down the creation of analytical products.

Addressing Limitations Through Data Mesh: The adoption of a data mesh architecture offers a transformative solution to these longstanding challenges. Unlike traditional models, data mesh decentralizes data ownership, empowering domain teams to take direct responsibility for their data. This shift reduces dependencies on centralized teams and fosters greater agility.

In a data mesh framework, data is treated as a **product**, managed by domain teams with clear ownership, well-defined APIs, and comprehensive documentation. This approach enhances the discoverability and usability of data, enabling teams to extract more value. By establishing a **self-service data infrastructure**, data mesh allows business users to access and analyze data independently, reducing bottlenecks caused by centralized controls.

Furthermore, decentralization improves **data quality** because domain teams possess the context and expertise required to ensure the accuracy, consistency, and relevance of their datasets. The modular architecture of data mesh also enhances organizational agility by facilitating the rapid integration of new data sources and adaptation to changing business requirements.

The limitations of traditional data platforms present significant barriers to organizations seeking to become truly data-driven. The dependency on centralized teams, inefficiencies in the data pipeline, and challenges in data quality and governance are compounded by the inability to respond quickly to evolving business needs. Data mesh offers a compelling alternative by decentralizing data ownership, enabling self-service, and fostering a modular architecture that prioritizes agility and quality. This architectural paradigm not only addresses the inefficiencies of traditional platforms but also equips organizations with the tools needed to innovate and respond effectively in a rapidly changing data landscape.

5.4 Data Mesh Operating Model

This research began with the fundamental question: What constitutes an effective operating model for the data ecosystem of an organization?

The focus of the study was on the banking industry, with additional insights gathered from the telecommunications sector. Through a detailed exploration of operating models and extensive engagement with industry executives, a framework was developed for a data mesh adoption strategy.

Defining an Operating Model for the Data Ecosystem: An operating model for a data ecosystem serves as a blueprint for managing, utilizing, and sharing data within an organization. It encompasses the framework, processes, roles, responsibilities, and governance structures necessary for ensuring alignment between data initiatives and business objectives. The key components of such a model include standardized processes, role clarity, governance mechanisms, integration with business functions, and the enabling technology infrastructure. (C. Liu et al.,2020)

Processes ensure streamlined data collection, storage, processing, and analysis through methodologies like DataOps and MLOps. **Roles and responsibilities** clarify accountability for data management, governance, and product creation. A robust governance framework safeguards compliance and ethical data usage. Integration with business units ensures that data initiatives align with organizational goals, while technology infrastructure supports effective data management and analytics.

Existing Operating Model Types: The study identified and analyzed three predominant operating model types:

Centralized Model: A centralized team manages the entire data ecosystem, offering strong control but limited agility.

Self-Service Model: While maintaining overarching governance, business units independently manage their data solutions, promoting agility and customization.

Hybrid Model: This combines the strengths of both centralized and self-service approaches, centralizing essential aspects while empowering business units to create specialized solutions.

These models were examined in the context of organizational size, complexity, cultural readiness, and data maturity. Key considerations for implementing an operating model included alignment with business goals, scalability, and stakeholder engagement. (A. K. Sandhu,2022).

Insights on Data Mesh Operating Model: Building on these foundational concepts, the research focused on creating an operating model tailored for data mesh adoption. This model emphasizes decentralization, collaboration, and governance, addressing the limitations of traditional centralized approaches.

Rollout Taskforce: The establishment of a Rollout Taskforce, also referred to as the Data Mesh Enablement Team, has been identified as a pivotal element in facilitating the adoption of Data Mesh principles. This taskforce is designed as a cross-functional entity, integrating internal leadership with external expertise to provide comprehensive strategic direction, technical leadership, and governance throughout the adoption process.

A primary responsibility of the Rollout Taskforce is to conduct a detailed assessment of the current state of the organization's data ecosystem. This involves a comprehensive analysis of existing infrastructure, processes, and capabilities to define the desired future state that aligns with Data Mesh principles. Through this assessment, the taskforce identifies gaps in the existing framework and develops a strategic roadmap aimed at bridging these gaps. This roadmap prioritizes initiatives that are critical for enabling a successful transition to the decentralized, domain-oriented data management model inherent in the Data Mesh paradigm.

In addition to its strategic functions, the Rollout Taskforce provides both technical and operational guidance to ensure a smooth and efficient transition to the Data Mesh framework. It supports domain teams in the adoption of best practices for data management and the creation of data products, focusing on ensuring alignment with the principles of domain-driven design, data as a product, self-serve platforms, and federated governance.

Initially established as a dedicated team for the adoption phase, the Rollout Taskforce gradually integrates its functions into the organization's routine operations. This evolution ensures that the momentum achieved during the transition phase is sustained, embedding Data Mesh principles into the organization's business-as-usual activities. By doing so, the taskforce not only facilitates the initial implementation but also establishes a foundation for continuous improvement and long-term scalability of the Data Mesh framework.

Domain Ownership: In the proposed model, data ownership is decentralized, with individual business domains (e.g., marketing, finance, operations) assuming responsibility for their data. This includes ensuring data quality, governance, and accessibility. The Rollout Taskforce collaborates with domain teams to establish a data mesh maturity model, which is later handed over to the governance team to monitor progress and adoption.

Data as a Product: The concept of treating data as a product is central to the data mesh framework. Each domain is responsible for creating and maintaining data products that are discoverable, usable, and valuable to others within the organization. The Rollout Taskforce provides guidance on identifying and standardizing these data products and recommends proof-of-concept or minimum viable products (MVPs) to kickstart the process.

Self-Service Infrastructure: A self-service infrastructure enables domain teams to independently access, manage, and share data, reducing reliance on centralized IT teams and fostering agility. The Rollout Taskforce assesses the existing infrastructure and recommends strategies—be it cloud-native, on-premises, or hybrid—based on organizational priorities and budget. A balance between open-source frameworks and cloud-native solutions is encouraged to ensure cost-effectiveness. (A. Cuzzocrea,2021).

Federated Governance: Governance in a data mesh model is distributed but aligned with organizational standards. Each domain implements its governance practices while adhering to shared policies for data quality and compliance. The Domain Governance Team collaborates with the Enterprise Governance Team and consults with the Rollout Taskforce for guidance.

Interoperability: Ensuring interoperability across data products from different domains is vital. Standardized APIs and metadata management facilitate seamless data integration and collaboration across the organization. The Rollout Taskforce works closely with domain teams to establish these standards and ensure technical compatibility.

The research underscores the importance of a well-defined operating model in fostering a robust data ecosystem. The traditional centralized approaches, while effective in some contexts, often fail to meet the agility and scalability needs of modern organizations. The proposed data mesh operating model offers a decentralized framework that promotes domain ownership, product-centric data management, and collaborative governance.

By addressing organizational challenges through the Rollout Taskforce, emphasizing self-service infrastructure, and ensuring interoperability, the proposed model equips organizations with the tools needed to drive innovation, improve data quality, and align data strategies with business objectives. This research provides a detailed foundation for implementing a data mesh operating model, paving the way for more effective and agile data ecosystems.

5.5 Principles of Data Mesh Operating Model

The principles of the Data Mesh Operating Model align with the core tenets of the data mesh itself, emphasizing decentralization, domain-driven design (DDD), scalability, collaboration, and continuous improvement. The Rollout Taskforce also called as Data Mesh Enablement Team plays a pivotal role in guiding domain teams, providing strategic direction, and fostering the adoption of decentralized data management. The model incorporates agile methodologies, leveraging minimum viable products (MVPs) to establish repeatable processes before scaling across the organization.

Decentralization ensures that individual domains manage their data ecosystems, while DDD aligns data product design with business objectives. Scalability is achieved as domains independently develop their data capabilities, and collaboration is facilitated through interactions between domains and the Rollout Taskforce. Continuous improvement is embedded in the model through retrospectives and iterative enhancements, ensuring adaptability to evolving needs. (A. Abbasi et al.,2022).

Implementing the Data Mesh Operating Model, however, presents challenges, including cultural shifts, skill gaps, governance complexities, technical infrastructure needs, and cross-domain collaboration issues. Overcoming these requires a structured approach. Cultural resistance is addressed through change management strategies, including training programs and change champions. Skill gaps are bridged through upskilling initiatives, mentorship programs, and certification opportunities. Governance complexities are mitigated by establishing a federated governance framework supported by a Center of Excellence (CoE) and standardized policies.

Technical challenges are tackled by adopting self-service data platforms with user-friendly interfaces, automated pipelines, and robust security frameworks, including role-based access control and data masking. Collaboration is enhanced through workshops, shared goals, and tools that promote cross-domain communication.

In conclusion of this section, the proposed Operating Model integrates principles of decentralization and collaboration while addressing practical challenges through strategic interventions by the Rollout Taskforce. This approach fosters a flexible, scalable, and secure data ecosystem, enabling organizations to harness the full potential of their data assets for strategic decision-making.

5.6 Key Steps of the Data Mesh Operating Model to operationalize it in organization.

Below is the conceptual 10 steps Data Mesh Operating Model to operationalize the Data Mesh within small to large banking organization. The model proposed here is easily leveraged for any other industry such as telecommunication, manufacturing etc. This model provides the strategic direction to the business executives.

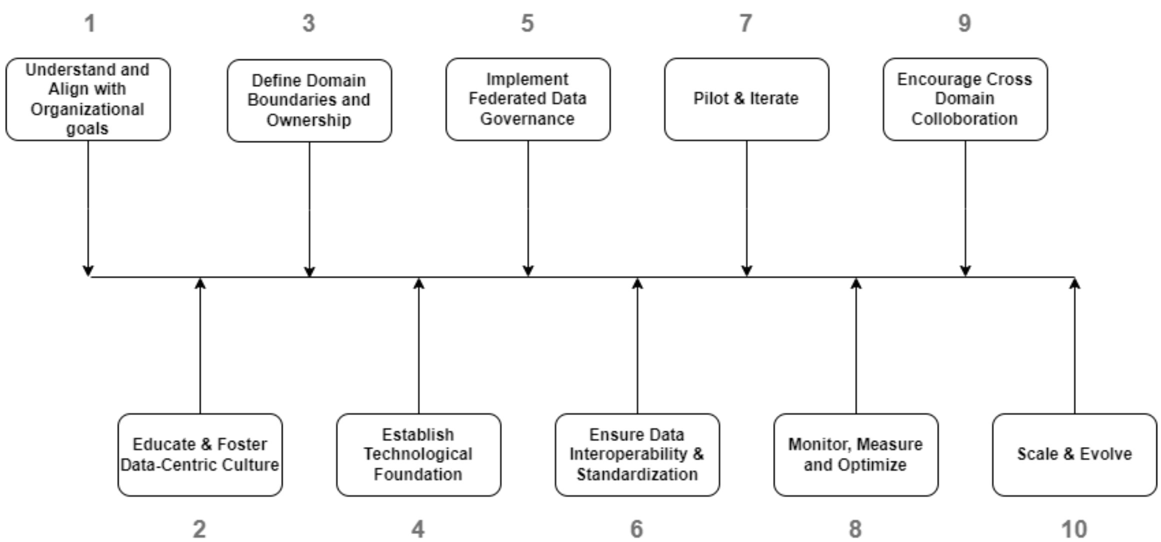

Figure 5.1 The 10 steps-based Data Mesh Operating Model

Figure 5-1, shows the 10 steps based conceptual Data Mesh Operating Model and below is the detail explanation of each and every conceptual stage which has been throughly studied as a part of the research study.

The implementation of a Data Mesh Operating Model requires a systematic, step-by-step approach that ensures alignment with organizational objectives and operationalizes a decentralized data architecture. This model, derived from extensive research across multiple banking organizations, outlines the conceptual stages involved in successfully deploying Data Mesh to enhance data management, collaboration, and innovation. The following sections provide a detailed explanation of each stage, based on thorough studies and interviews conducted with various banking stakeholders.

Step 1: Understand and Align with Organizational Goals

The first step in implementing a Data Mesh is to understand and align the initiative with the organization's overarching strategic objectives. This involves identifying key business goals, such as improving customer experience, enhancing risk management, or optimizing operational efficiency. A comprehensive assessment of the current data landscape is essential, which includes evaluating the bank's existing data assets, their sources, and how they are utilized across the organization. Additionally, engaging key stakeholders—executives, data owners, data analysts, and IT teams—is critical to ensuring that the Data Mesh initiative addresses their specific needs and expectations. This collaborative alignment helps to set the foundation for a successful data ecosystem.

Step 2: Educate and Foster a Data-Centric Culture

Fostering a data-driven culture is essential for the successful adoption of a Data Mesh. The research emphasizes the importance of tailored training programs that cater to the specific roles within the organization, including data engineers, data scientists, and business analysts. Furthermore, promoting data literacy across the bank ensures that all employees understand the value of data and are equipped to make data-driven decisions. Establishing a culture where data is leveraged for decision-making across all departments is pivotal to realizing the full potential of a decentralized data model.

Step 3: Define Domain Boundaries and Ownership

The next step is to define the domains that align with the bank's core business functions. These domains typically include customer data, financial data, risk data, and operational data. Each domain must have clear boundaries, ownership, and accountability. Data product owners are assigned within each domain, responsible for the quality, accessibility, and lifecycle management of the data. It is essential to ensure that domain ownership structures comply with banking regulations, such as data privacy laws, ensuring that data governance and management practices are legally sound and aligned with industry standards.

Step 4: Establish Technological Foundations

Implementing a robust technological foundation is crucial for supporting the decentralized nature of Data Mesh. Cloud-native technologies, such as AWS, Azure, or GCP, should be selected for their scalability and flexibility in distributed data management. A Data Mesh platform must be deployed to facilitate key services such as data cataloguing, data lineage tracking, and governance. Furthermore, seamless integration with the bank's core banking systems is necessary to ensure that the Data Mesh framework supports the broader technological ecosystem of the organization.

Step 5: Implement Federated Data Governance

Effective data governance is a key aspect of any Data Mesh implementation. A federated governance model strikes a balance between centralized oversight and domain-specific autonomy. In this step, organizations should create data governance policies that are tailored to the banking sector, ensuring compliance with industry standards and regulatory requirements. A robust data quality framework should be put in place to ensure that data remains accurate, complete, and consistent across all domains.

Step 6: Ensure Data Interoperability and Standardization

Data interoperability is critical to the success of a decentralized architecture. To achieve this, the adoption of banking-specific data standards and protocols is necessary to ensure seamless interaction between different domains. Data cataloguing tools and discovery mechanisms should be implemented to enhance data accessibility and make data products discoverable. Additionally, tracking data lineage allows for understanding the origin, transformations, and usage of data, ensuring transparency and traceability.

Step 7: Pilot and Iterate

Starting with a pilot domain provides an opportunity to test the Data Mesh operating model before full-scale implementation. The research recommends selecting a domain that is self-contained and has a clear business objective. Success metrics, such as improved data quality, reduced time-to-insights, and cost savings, should be defined to measure the effectiveness of the pilot. Based on feedback from the pilot, the Data Mesh model can be iteratively refined to address any challenges or gaps before expanding to other domains.

Step 8: Monitor, Measure, and Optimize

Once the Data Mesh framework is in place, continuous monitoring is essential to ensure optimal performance. Implementing tools to track data quality, security, and resource utilization allows organizations to measure the impact of the Data Mesh on business outcomes. Regularly reviewing key performance indicators (KPIs) and optimization opportunities, such as improving data pipelines or enhancing security measures, ensures that the model remains efficient and effective over time.

Step 9: Encourage Cross-Domain Collaboration

Encouraging collaboration between domains is crucial for maximizing the value of a Data Mesh. The research highlights the importance of creating collaborative forums or communities where data domain owners can share best practices and address common challenges. Fostering a culture of data-driven decision-making and innovation across domains will help unlock new insights and foster organizational growth.

Step 10: Scale and Evolve

As the Data Mesh operating model matures, gradual expansion to additional domains is necessary to meet the evolving data needs of the organization. The research emphasizes the need for adaptability—both in the Data Mesh architecture and the operating model—as new technologies and business requirements emerge. Flexibility to scale the Data Mesh while maintaining its core principles ensures that the organization continues to benefit from improved data management and innovation.

The research study has highlighted a conceptual, 10-step operating model for Data Mesh that can be adopted by banks and similar organizations. This approach ensures alignment with organizational goals, fosters a data-centric culture, and facilitates the decentralization of data management across business domains. By following these steps, banks can successfully implement a Data Mesh framework that enhances data management capabilities, supports scalability, and encourages collaboration. The model ultimately provides a flexible and adaptable structure that helps organizations leverage data more effectively for strategic decision-making and innovation.

CHAPTER VI:

DISCUSSION-DATA MESH IN BANKING

6.1 Introduction

The research underscores the critical role of the proposed Data Mesh operating model in addressing prevalent challenges in centralized data management systems. Organizations, particularly in the banking sector, increasingly adopt Data Mesh to tackle the following issues:

Scalability: Centralized architectures often struggle to handle the exponential growth of data. Data Mesh provides a scalable and decentralized framework that ensures efficient data handling across distributed domains.

Data Governance: Maintaining governance in a centralized model becomes complex in large organizations. By assigning ownership to individual data domains, Data Mesh improves accountability and streamlines governance practices.

Agility: The distributed nature of Data Mesh accelerates the development and deployment of data-driven solutions, enabling organizations to swiftly adapt to evolving business requirements.

Cost Efficiency: For organizations with diverse and extensive data needs, Data Mesh offers a cost-effective alternative by reducing bottlenecks and optimizing resource allocation.

Innovation: Empowering data domains to own their processes fosters innovation, enabling the creation of domain-specific, data-driven products and services.

The research findings highlight that adopting a Data Mesh operating model significantly enhances data management capabilities while promoting organizational agility and innovation.

Data mesh represents a transformative approach to data management, emphasizing a decentralized model that mirrors the interconnectedness of a natural ecosystem. This architecture not only incorporates technological elements but also addresses essential social factors—specifically, organizational and cultural considerations. Many enterprises today tend to focus predominantly on the technical aspects of data mesh, often overlooking the critical social dimensions.

In fact, numerous practitioners assert that technology constitutes only a small fraction of what makes a data mesh implementation successful. Based on our experience, the 80/20 rule is applicable here: approximately 80% of the effort, time, and cost associated with implementing a data mesh (and developing data products) is devoted to influencing and engaging people to adopt new organizational and cultural practices.

This shift emphasizes decentralization, local autonomy, new roles, and innovative governance techniques. The topic of organizational and cultural issues is extensive, which we will explore in this chapter and the next. Here, we will focus on the various teams involved in the data mesh framework, including the data product teams and their interactions with other groups. The subsequent sections will delve into the broader operating model, examining how teams collaborate, how they are governed, and how they are incentivized.

An operating model is a blueprint that defines how an organization conducts business, delivering value through its operations. It aligns people, processes, and technology to achieve strategic goals. In the context of Data Mesh, the operating model guides how data or more data products are managed, shared, and utilized within an organization, ensuring that data initiatives align with business objectives, goals and strategies.

One key objective of an operating model is to establish clear roles and responsibilities, ensuring that each team and individual knows their specific functions and how they contribute to the broader organizational goals.

Another aspect of an operating model is the establishment of efficient processes that describe how tasks and activities are carried out within the organization. For Data Mesh, this means establishing standardized processes for governance, quality, and lifecycle management of data products across the organization.

Technology is the backbone of every business. Every organization is technology dependent. Therefore, integration of technology is also a crucial aspect of an operating model. This includes selecting and implementing the right technological tools and platforms to support the organization's operations. The operating model should facilitate a technology landscape that is easy to access, agile, scalable, and conducive to the decentralized nature of Data Mesh, enabling seamless interaction and data sharing across different domains.

In addition to this, an operating model emphasizes the importance of communication and collaboration among different units within an organization. Effective communication channels and collaboration practices ensure that teams can work together seamlessly, sharing insights and expertise.

The rapid evolution of data management has led to the emergence of data mesh architecture, fundamentally changing how we perceive and utilize data. This shift moves away from traditional centralized data warehouses towards a more decentralized approach. But what does it truly mean to embrace the data mesh paradigm? At its core, a data mesh treats data as a product, empowering domain teams and data product owners to manage their own data products within their specific business areas. This transformation emphasizes the principles of domain-driven design, enabling organizations to operate more effectively in an era characterized by vast amounts of data. IN this article, we will explore the foundational principles of data mesh and provide a detailed, step-by-step guide for successful implementation. From assembling your data platform team and fostering domain ownership to scaling your data product catalogue and refining your approach to data mesh, each phase of this journey is crucial for accelerating your digital transformation. Whether you are a business leader, a data product owner, or part of a domain team, the proposition mentioned in research study is useful for the Data Mesh operationalization.

Data mesh is an innovative data architecture that promotes a decentralized approach, treating data as a product. It's an ideal solution for large enterprises struggling with monolithic and siloed data systems. Banks journey to data mesh implementation begins with building an enabling team. It's crucial to form a cross-functional team with members from different domains, fostering collaboration and shared responsibility. Next, focus on setting up a self-service data platform and defining an initial set of data products. This pilot phase allows you to gain buy-in from stakeholders, test the process, and showcase the potential of the data mesh approach. As you scale your data product catalogue and enhance your self-service data platform, expand the data mesh to cover more domains. Foster a culture of data ownership, encouraging domain experts to define and manage their data products. AS your data mesh matures, nurture it through clear governance and data stewardship practices. Regularly monitor and refine the data mesh, incorporating user feedback and evolving business needs. Embrace an agile mindset and a culture of continuous improvement to keep refining your data mesh. Leverage cutting-edge technologies and foster a thriving, data-driven culture in your organization. Remember, the transition to data mesh architecture is organic and iterative. It doesn't require refactoring every system; rather, integrate your existing systems into the data mesh where it makes sense. Implementing a data mesh is a transformative journey that requires careful planning, commitment, and expertise.

6.2 Embracing the Data Mesh World: Understanding Its Implications

The emergence of data mesh architecture has significantly altered our approach to data management, shifting from a centralized model reliant on data warehouses to a more decentralized framework. But what does it truly mean to embrace the data mesh paradigm?

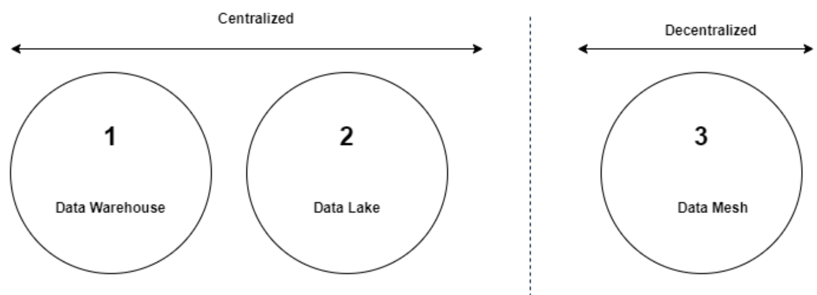

Figure 6.1 Centralized data platform like data warehouse or data lake and the move towards decentralized data architecture that data mesh introduces.

At its core, data mesh treats data as a product, empowering data product owners and domain-specific teams to manage their own data products within their respective business areas. This transformation emphasizes the principles of domain-driven design, allowing organizations to operate more efficiently in the age of big data. In a data mesh framework, ownership of data is distributed among various domains rather than being concentrated in a central team. This decentralization fosters accountability and encourages teams to optimize their data products according to specific business needs.

By treating data as a product, organizations can enhance the quality and accessibility of their datasets, ultimately driving better decision-making and innovation.

Additionally, the implementation of self-serve data platforms enables teams to access and utilize their data independently, reducing reliance on centralized IT departments. This approach not only improves agility but also supports a culture of collaboration and continuous improvement.

In summary, embracing the data mesh world signifies a fundamental shift in how organizations perceive and manage their data assets. It promotes a decentralized, domain-oriented approach that empowers teams to take ownership of their data products, fostering greater efficiency and responsiveness in an increasingly complex data landscape.

6.3 The Transition from Data Warehouse to Data Mesh Architecture

Historically, data warehouses have functioned as the central repositories for all organizational data, responsible for processing and storing analytical information. This centralized model often leads to bottlenecks and inefficiencies, hindering timely access to critical insights.

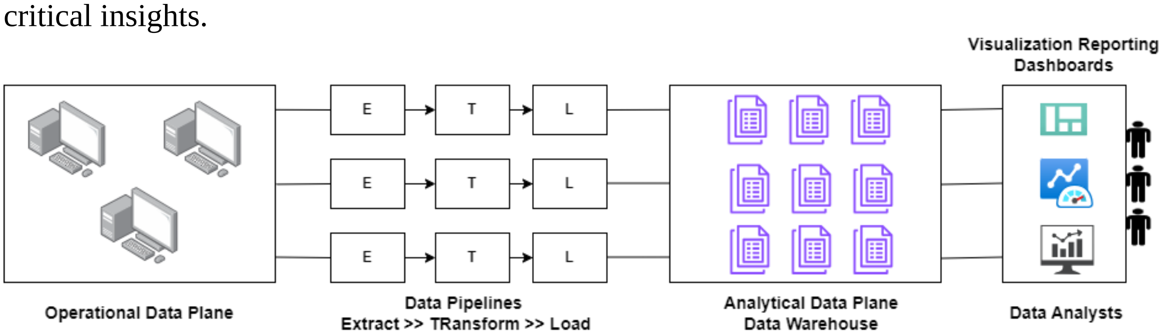

Figure 6.2 Data warehouse architecture with ETL pipelines that are saving data to central data warehouse and data analysts accessing the centralized data platform though SQL-queries and direct DB connections.

In contrast, data mesh architecture represents a logical evolution of the data lake concept, decentralizing the management of data assets. Within this framework, each business domain operates its own data products, with respective domain teams taking ownership and managing the entire lifecycle of these products. This structure facilitates closer collaboration between data engineers, data producers, and data consumers, ensuring better alignment with business strategies and outcomes. By embracing a data mesh approach, organizations can enhance their agility and responsiveness to changing market conditions while improving data quality and accessibility. This shift not only empowers individual teams but also fosters a culture of accountability and innovation in how data is utilized across the enterprise.

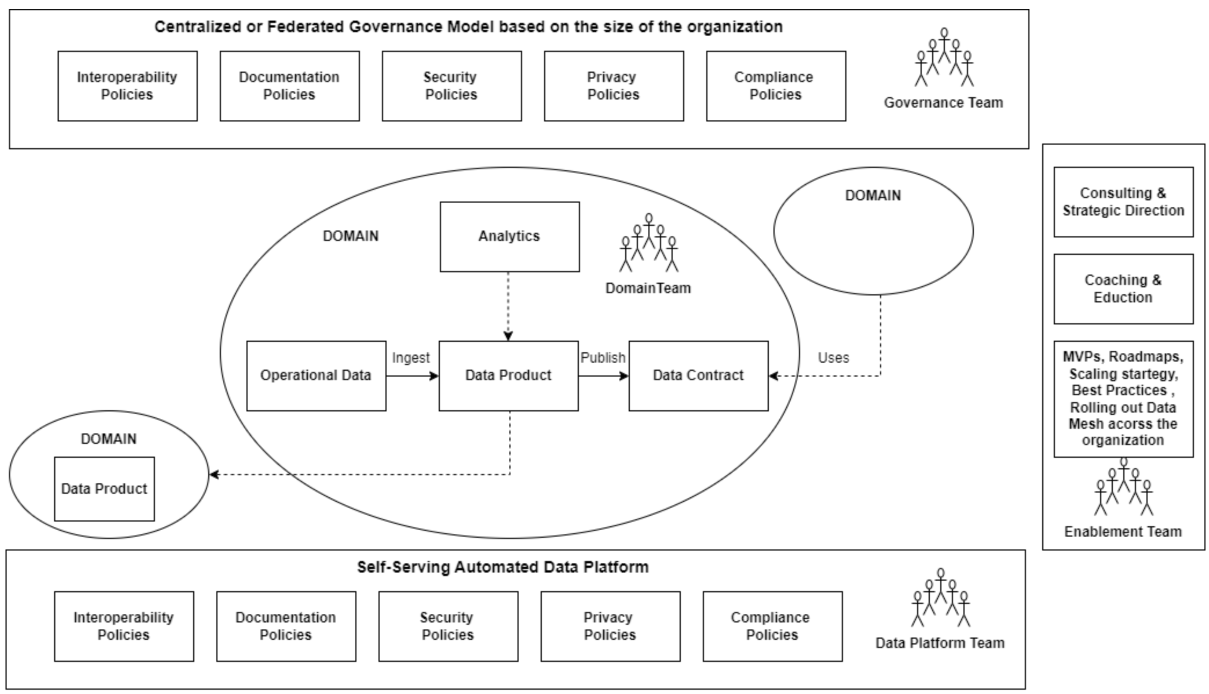

Figure 6.3 Data mesh Operating Model and the responsibilities of domain team and supporting central teams in enterprise data mesh model.

6.4 Data Mesh Principles for an Operating Model

Data mesh principles assert that each business domain should own and manage its data products, but this concept extends beyond mere responsibility. Data product owners are also accountable for ensuring data quality, meeting service level agreements, and making data easily accessible to other teams. A key aspect of this framework is the principle of self-service. A well-implemented data mesh promotes a self-service operating model, granting data consumers straightforward access to data products. This approach eliminates the bottlenecks associated with transferring data from a centralized location to multiple functional teams, resulting in a more efficient data infrastructure. Another essential principle is federated governance, which encourages decentralized data governance across various business domains. Rather than relying on a central authority to oversee data quality and security, these responsibilities are distributed among domain owners. Finally, embarking on a data mesh journey is not a one-time effort but an ongoing process. It involves continuously refining the data mesh roadmap, learning from the existing structure, and adjusting core components and both logical and business architectures to align with evolving business needs. By embracing these principles, organizations can effectively harness the full potential of their data assets while fostering a culture of collaboration and innovation.

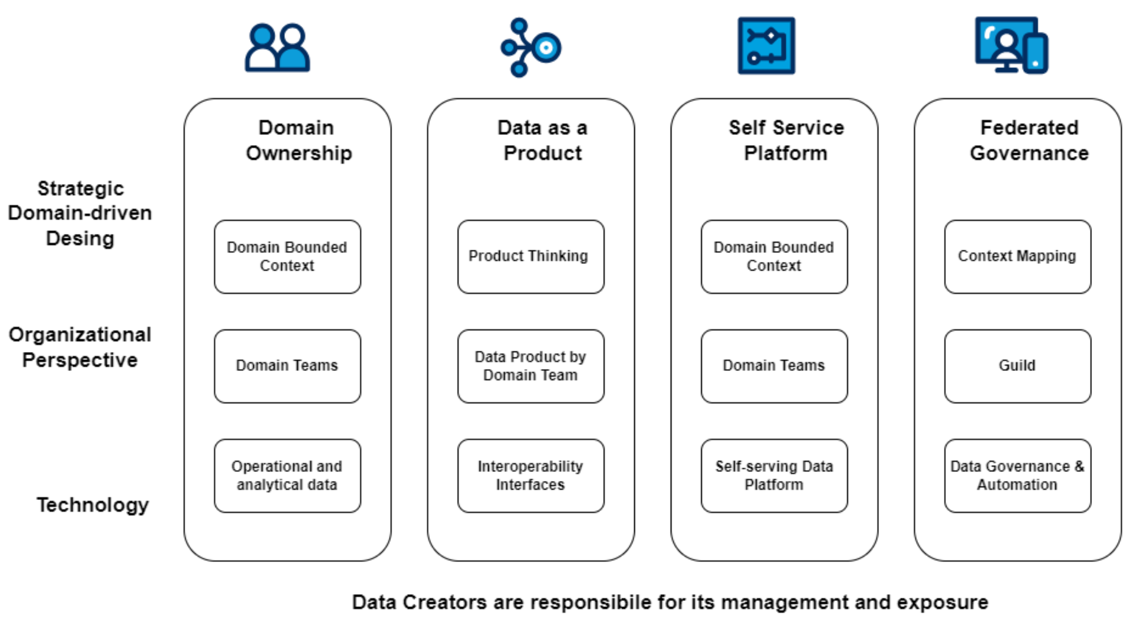

Figure 6.5 Data Mesh Operating Model Principles

The transition to a data mesh framework recognizes the distributed nature of data within contemporary organizations. It focuses on empowering domain teams to take ownership of their data products, which facilitates a more effective and efficient utilization of data assets throughout the organization. By fostering this sense of ownership, organizations can enhance collaboration and drive better decision-making based on timely and relevant data insights.

6.5 Banking Data Mesh Implementation Journey: A Step-by-Step Roadmap

Navigating the world of data mesh architecture can feel like a complex endeavor, especially when it comes to practical implementation. Every successful journey begins with a single step. This research study offers a comprehensive, step-by-step roadmap to operationalize data mesh implementation process, easing organization transition from a traditional data warehouse or data lake model to a robust, efficient data mesh structure.

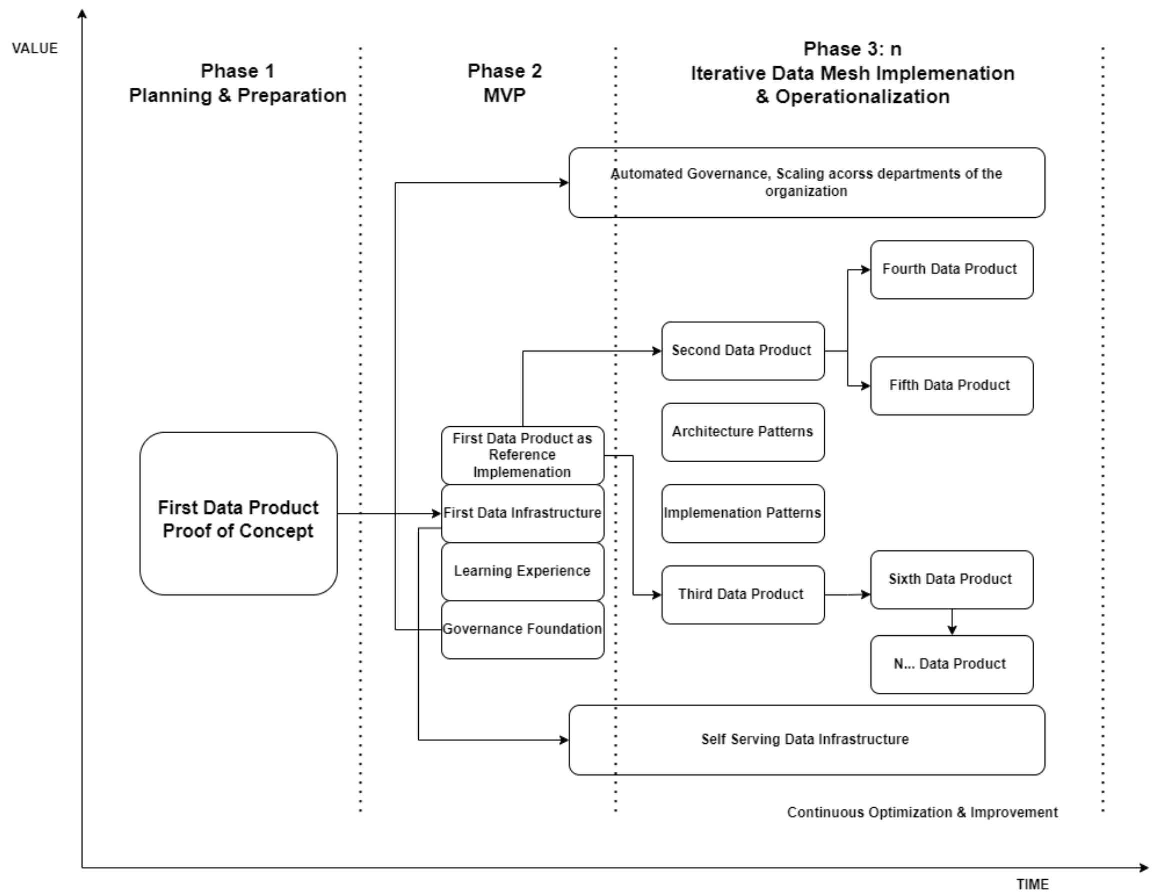

Figure 6.6 The key phases of data mesh operating model

Step 0 - Setting the Data Strategy Stage: Preparing for Data Mesh Implementation

Banking data mesh journey begins with the crucial step of setting the data strategy stage. This initial phase typically spans 1-6 months based on the size and complexity of the organization data eco system and focuses on assembling Data Mesh Enablement team, cross-functional teams, fostering domain ownership, and building collaboration between departments using all core principles of a domain-driven design.

Establishing the Data Mesh Enablement Team: Enabling teams are specialized groups within a bank that provide essential support and expertise to other teams, helping them overcome obstacles and address specific needs. Acting as consultants or advisors, these teams offer guidance, resources, and solutions to assist teams in navigating challenges and achieving their objectives.

These enabling teams work closely with various Data Mesh teams, collaborating in short bursts or on a project basis to deliver targeted assistance. They leverage their expertise and knowledge to tackle specific problems or areas where additional support is required. The structure of enabling teams can vary based on the organization’s needs and may include steering groups, enterprise governance and architecture teams, training groups, or other specialized units that provide insights and assistance in particular domains.

The common thread among these teams is their deep knowledge and experience in specific areas, enabling them to offer valuable guidance, best practices, and resources that help other teams succeed. By utilizing the expertise of enabling teams, data product teams can benefit from specialized support without the need to develop the same capabilities within every individual team.

Enabling teams play a crucial role in fostering collaboration, knowledge sharing, and innovation across the organization. Their targeted support helps teams facing challenges or pursuing new opportunities, ultimately enhancing the overall effectiveness of the data mesh implementation.

Assembling Data Platform Team: Achieving big data goals starts with bringing together the right people. The data platform team, a cross-functional group composed of data engineers, data architects, data analysts, and data scientists, is pivotal to the process. Their collective expertise and collaboration will shape the roles and responsibilities within the team, ensuring smooth coordination and effective communication. As the torchbearers, this team will take ownership of the proof of concept in the next phase, thereby enabling other teams in different domains to secure necessary resources.

Domain Teams: The Heart of a Data Mesh: Domain teams, often the unsung heroes in most enterprises, form the heartbeat of the data mesh architecture. Recognize their vital role and identify these teams within organization. It has been discovered they have been managing data assets without formally being acknowledged as domain teams (e.g., within their own departments). By promoting a culture of data ownership, these domain experts can independently manage and govern their data products. It's a decentralized approach that resonates with the ethos of domain-driven design.

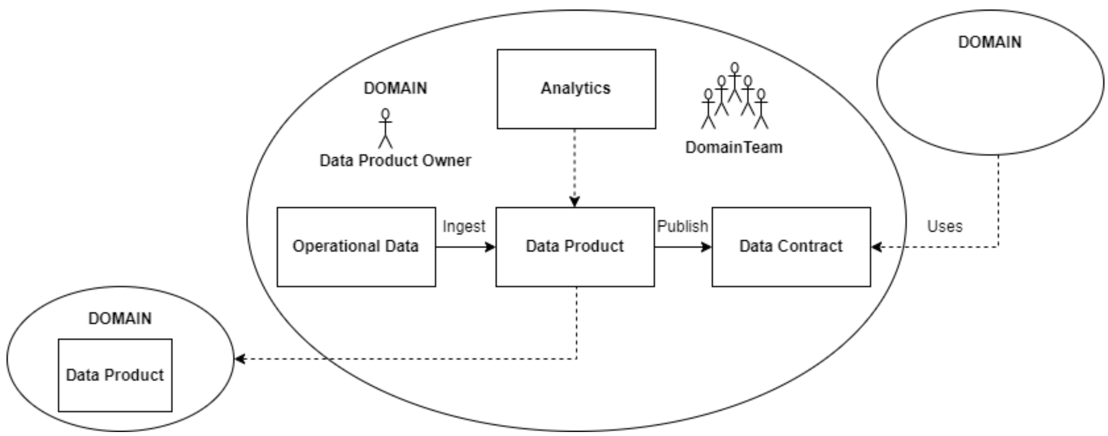

Figure 6.7 Domain team with data product owner and responsibilities

Promoting Domain Ownership in Data Infrastructure: One of the essential steps in the data mesh journey involves encouraging domain ownership across your data infrastructure. The Operating model empowers domain teams to take control and make data-driven decisions that align with their business goals. Data Mesh Enablement team collaborate with these teams to identify their unique data needs and ensure that the data products match their specific requirements.

Recognizing the Importance of a Strong Data Platform: At the heart of data mesh implementation is a robust, scalable, self-serve data platform. It's an essential building block in creating value from big data and meeting the needs of different domains. Begin by planning investments in technologies and tools that facilitate easy data discovery, access, and collaboration across the organization. At this step, organization should think about preparations, early architecture, or evaluating different tools and services.

Step 1 - Data Product Proof of Concept: A pivotal step in organization’s data mesh implementation journey is the execution of a proof of concept (PoC), which typically spans a month or 3 months based on the scope of the PoC. This PoC, focusing on a specific data product or domain, provides a vital pilot project to validate the feasibility and benefits of a data mesh architecture.

Select a Specific Data Product or Domain as a Pilot for the Proof of Concept (PoC)

Begin by identifying a well-defined data product or domain that aligns with your organization's business priorities. Opt for a domain with clear use cases and data requirements to effectively showcase the data mesh approach. As the Enabling Team, your role is to secure the necessary time and resources to run the PoC within this selected domain.

Define the Scope and Requirements of the PoC, Considering Both Technical and Business Aspects

In collaboration with domain experts and stakeholders, outline the scope of the PoC clearly. Involve data consumers in stating their needs and requirements for the PoC data product. Keep both technical and business aspects in mind, ensuring that the PoC tackles critical data challenges and delivers valuable insights.

Architect the Necessary Infrastructure, Including Data Pipelines, Storage, and Access Mechanisms

Design and build the data pipelines, storage solutions, and access mechanisms required to support the data product's operations. Utilize your self-serve data platform, enabling domain teams to manage their data independently. Remember, at this stage, the emphasis is on the architecture part to be better prepared for the following steps.

Evaluate the Outcomes of the PoC, Gather Feedback, and Refine the Approach as Needed

Implement the PoC using a data-driven approach, closely monitoring its usage. Encourage data consumers to grade the data product along with its meta description to evaluate its usefulness, quality, and freshness. Collect feedback from domain teams, end-users, and stakeholders to assess the effectiveness and efficiency of the data mesh approach. After the evaluation, analyze the PoC outcomes and make iterative improvements to your data mesh architecture and its implementation. Identify and address any challenges or shortcomings that emerged during the PoC, using this valuable feedback to refine and enhance your overall data mesh strategy.

Step 2 - Implementing First Data Products MVP: Having validated the data mesh approach in the PoC, Step 2 entails a significant expansion. Over the course of 3 - 4 months, you will build upon the initial success and extend the data mesh implementation to include Minimum Viable Products (MVPs) for additional data products or domains. This step aims at delivering tangible value to your organization by fostering self-serve capabilities and forging a robust data platform.

Identify Additional Data Products or Domains for MVP Development Based on Business Priorities

Work with domain teams and stakeholders to identify high-priority data products or domains that align with the organization's strategic objectives. Each data product's potential business impact and value should be considered to prioritize development efforts. Cross-functional representation in the selection process ensures a diversity of data needs are captured.

Design and Develop a Self-Serve Data Platform With a Basic Product Catalog to Enable Easy Discovery and Access to Data Products

Designate a dedicated team of data engineers, architects, and experts to design and develop the self-serve data platform. The requirements and functionalities of the platform should be clearly defined, including features like data cataloging, data lineage, access controls, and data discovery capabilities. Use industry best practices and modern data technologies to build a platform that is both scalable and flexible. Develop a basic product catalog within the self-serve data platform, creating an organized, user-friendly interface for data consumers to discover available data products. Implement search and filtering mechanisms to enable users to find relevant data products based on their specific needs. Each data product in the catalog should include comprehensive metadata, encompassing data descriptions, data sources, and data quality indicators.

Implement Production Deployment for the Data Product, Ensuring Scalability, Security, and Governance

Move the MVP data products into production, ensuring they are accessible to the appropriate users and stakeholders. The scalability of the self-serve data platform should be validated to accommodate the growth of data products and increasing user demands. Put rigorous data security and access controls in place to protect sensitive information within the platform. Data governance practices must be established to maintain data quality, compliance, and regulatory standards.

Iterate on the MVP Development Based on User Feedback and Evolving Business Needs

Regular feedback sessions with domain teams and data consumers are instrumental in gathering insights into the effectiveness and usability of the MVP data products. User feedback should be incorporated into the iterative development process, emphasizing continuous improvement and optimization. Stay attuned to evolving business needs, adjusting the data mesh architecture and platform to accommodate new requirements. Leverage data analytics to monitor the usage and performance of MVP data products, identifying areas for enhancement.

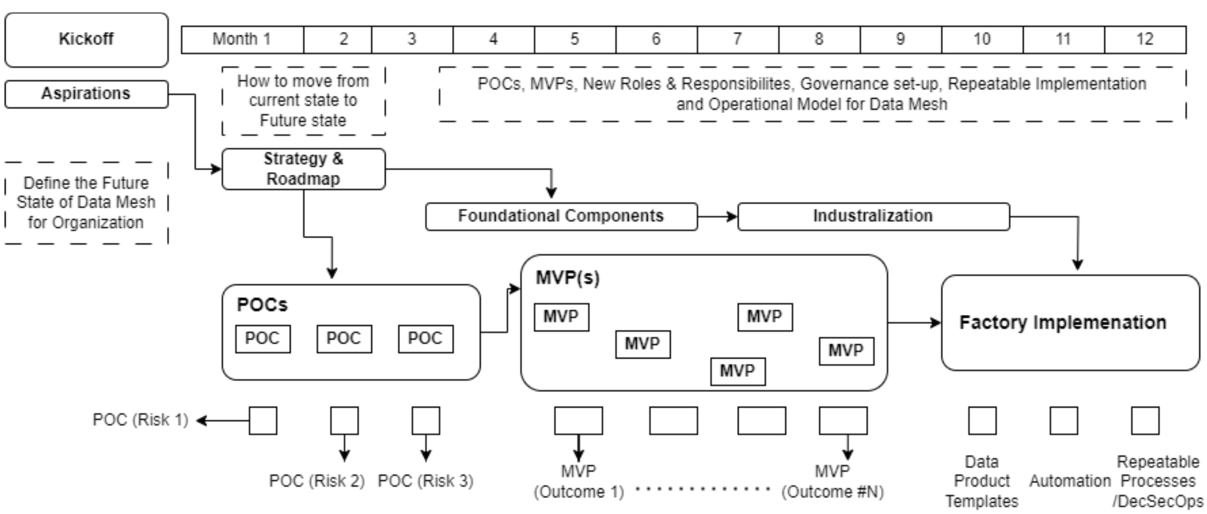

Figure 6.8 Data mesh implementation roadmap - iterative PoCs and MVPs development over time

Step 3 - Scaling Data Product Catalog and Deploying Self-Service Data Platform: Step 3 is centered around growth and maturation. This phase involves expanding the data mesh to encompass more domains and data products, fostering a culture of data ownership among domain experts, and enhancing the self-serve data platform with advanced features. This step is ongoing and iterative, adapting to the organization's evolving needs.

Expand the Data Mesh Implementation to Cover More Domains and Data Products Iteratively

Identify additional domains and data products that could benefit from the data mesh approach. Collaborate closely with domain teams to integrate them into the data mesh ecosystem. Encourage these teams to adopt data ownership and manage their respective data products independently, fostering a decentralized approach that aligns with the core principles of the data mesh.

Empower Domain Experts for Data Ownership and Autonomy: Cultivate a culture of data ownership and autonomy by emphasizing the critical role of domain teams in data management. Provide training and support to domain experts to enhance their skills in managing data products. Facilitate collaboration and knowledge sharing between domain teams, fostering a data-driven culture that extends beyond technical aspects to permeate the entire organization.

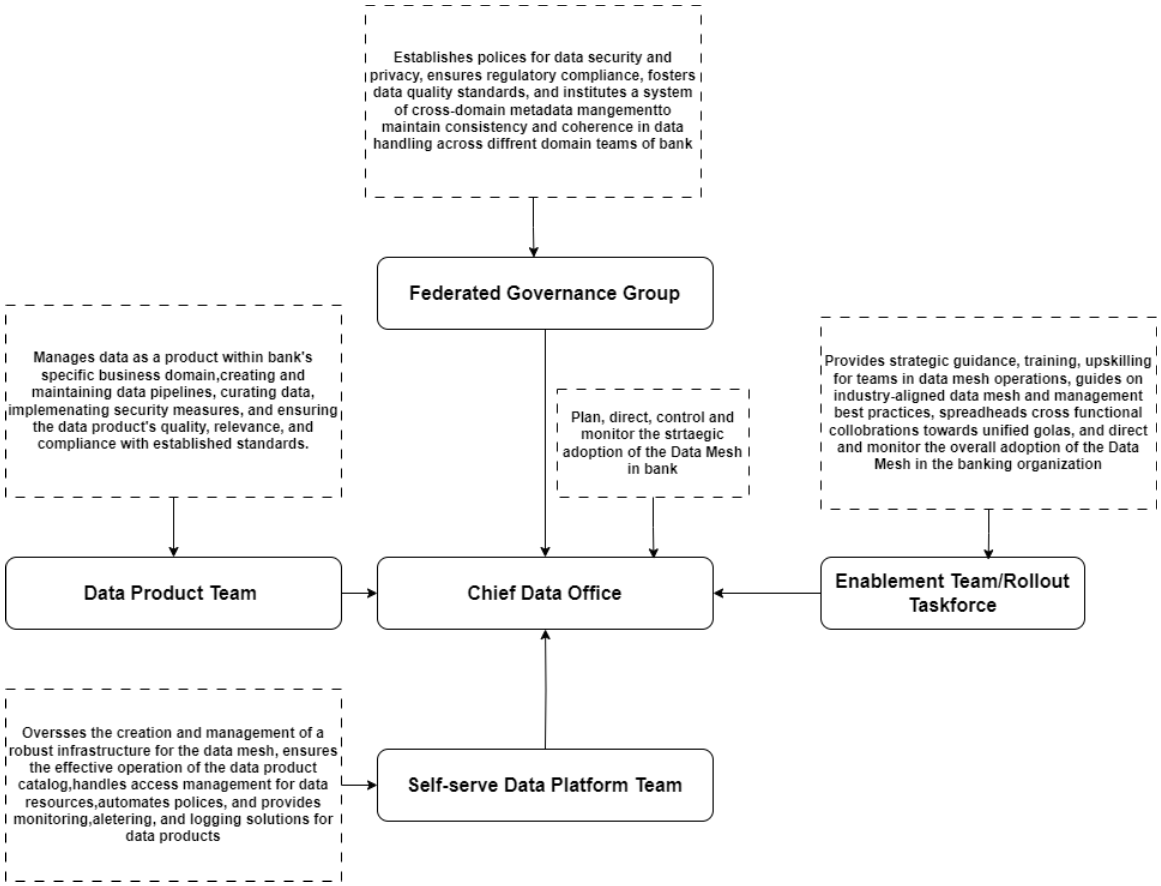

Figure 6.9 Responsibilities of various teams working on data mesh.

Enhance the Self-Serve Data Platform with Advanced Features Such as Data Cataloguing, Data Lineage, and Data Discovery Capabilities

Gather feedback from domain teams and data consumers to identify areas for improvement in the existing self-serve data platform. Prioritize the implementation of advanced features, including comprehensive data cataloguing for easy product discovery, data lineage tracking for transparency and traceability, and enhanced discovery capabilities to assist users in exploring datasets. Additionally, integrate tools for monitoring data quality to ensure reliability and accuracy, aligning with global standards.

Implement Continuous Improvement and Iterative Development: Establish a feedback loop with domain teams and data consumers to collect ongoing insights and suggestions. Regularly review the data mesh implementation to identify areas for optimization and enhancement. Commit to continuous iterative development of both the self-serve data platform and its associated products, ensuring alignment with changing business needs while creating greater value for the organization.

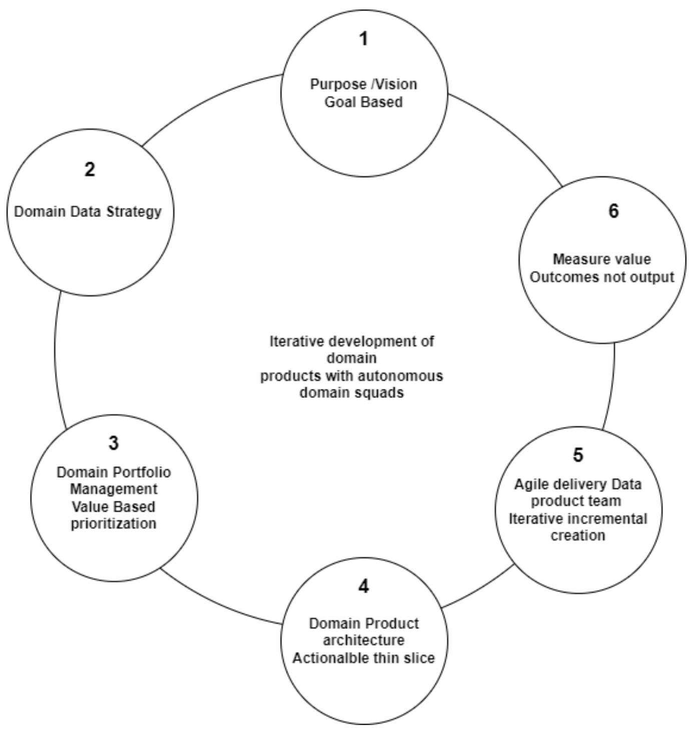

Figure 6.10 Iterative development of domain products with autonomous domain squads

Promote Collaboration and Knowledge Sharing: Organize regular cross-domain meetings and workshops to encourage collaboration and knowledge sharing. Foster an environment of open communication by creating forums where domain teams can share best practices, success stories, and lessons learned. Consider hosting data summits or conferences to showcase the value of data products and cultivate a vibrant, data-driven community.

Measure Success and Impact: Define key performance indicators (KPIs) to assess the success and impact of the data mesh implementation. Monitor adoption rates and usage of data products across domains, tracking improvements in accessibility, quality, and decision-making driven by data insights. These KPIs will serve as critical markers of progress, providing valuable insights into how the data mesh strategy contributes to organizational goals.

Step 4 - From Implementing to Managing: Nurturing Your Data Mesh: Step 4 marks the transition from initial implementation to the ongoing management and nurturing of the data mesh. During this stage, the focus shifts to establishing robust governance and data stewardship practices while creating mechanisms for continuous monitoring and refinement of the data mesh. The goal is to maintain data quality, compliance, and security while adapting to evolving business needs and user feedback.

Establish Clear Governance and Data Stewardship Practices to Ensure Data Quality, Compliance, and Security

Implement Data Quality Framework: Define and enforce consistent data quality standards across all data products and domains. Implement processes for data profiling, validation, and cleansing to ensure data accuracy and reliability.

Data Security and Compliance: Deploy comprehensive data security protocols and maintain compliance with relevant data privacy regulations. Establish stringent access controls, applying data encryption and masking techniques to safeguard sensitive information.

Metadata Management: Develop a solid metadata management practice to preserve data lineage, definitions, and cataloging. This enhances data discoverability and understanding, enabling users to make better-informed decisions.

Data Ownership and Accountability: Clearly delineate roles and responsibilities for data ownership within domain teams. Foster a sense of accountability among team members for maintaining their data products to ensure they meet user requirements.

Continuously Monitor and Refine the Data Mesh Implementation, Incorporating Feedback From Users and Stakeholders

Feedback Mechanisms: Establish channels for users to provide feedback on data products and the overall data mesh platform. Regularly collect this feedback to identify areas for improvement and optimize user experiences.

Performance Monitoring: Implement mechanisms to track the performance of both data products and the self-service platform. Monitor metrics such as data availability, response times, and usage patterns to identify and address any bottlenecks.

Iterative Development: Embrace an iterative development approach to continuously enhance the data mesh implementation. Incorporate insights gained from monitoring activities into platform updates and improvements.

Scalability and Resilience: Keep a close watch on the scalability and resilience of the data mesh architecture. Ensure it can accommodate growing data volumes and user demands while maintaining high performance and reliability.

Data Governance Review: Conduct regular reviews of governance practices and policies to ensure alignment with evolving business needs and regulations. Adjust governance frameworks as necessary to address new challenges.

Collaborative Improvement: Foster collaboration between domain teams, the self-service platform team, and governance teams to drive continuous improvement. Cultivate a culture of learning and knowledge sharing that enhances data management capabilities across the organization.

By focusing on these areas in Step 4, organizations can effectively manage their data mesh architecture, ensuring it remains responsive to business needs while upholding high standards of quality, security, and compliance.

Step 5 - Continually Refining Your Data Mesh: Step 5 marks the beginning of a transformative journey where your fully established data mesh starts to evolve organically to meet the ever-changing needs of your organization. The data mesh becomes a living, thriving ecosystem, expanding and adapting in response to shifting business priorities and technological advancements.

Exploring Cutting-Edge Technologies: As your data mesh grows, it's essential to stay updated on technological advancements. The self-serve data platform should continuously evolve to incorporate advanced features such as data cataloguing, data lineage, and data discovery capabilities. Encourage a culture of curiosity and experimentation within domain teams, enabling them to explore new tools and techniques that enhance the value of their data.

A Flourishing Data-Driven Culture: With the data mesh at the core of your organization, a vibrant data-driven culture begins to develop. Data literacy becomes a critical organizational competency, empowering every team member to confidently navigate and utilize data products. This culture of data-driven decision-making permeates the organization, fostering a proactive and informed approach to strategy and growth. By focusing on these areas in Step 5, organizations can ensure that their data mesh remains dynamic and responsive, ultimately driving greater value from their data assets while supporting ongoing innovation and collaboration across teams.

The Organic and Agile Nature of a Successful Data Mesh Implementation: Transitioning to a data mesh architecture is an organic and agile process that aligns seamlessly with the progressive and iterative ethos of contemporary business practices. Organizations that have traditionally relied on siloed data architectures can successfully leverage the benefits of data mesh. The implementation of data mesh can begin with a few key data products, allowing for incremental growth.

As the organization gains confidence and competence in managing a distributed data ecosystem, new domains and data products can be gradually added. The selection of subsequent domains to integrate into the data mesh is strategically driven by business goals and priorities, ensuring alignment with the organization's overall strategic direction. Importantly, implementing a data mesh does not require a complete overhaul of existing systems within the organization. Legacy systems that continue to deliver value can be retained.

The key is to develop connectors that expose data products from these systems for further analysis and utilization, seamlessly integrating them into the data mesh. This approach respects the value of existing investments while enabling the organization to innovate and adapt in a rapidly changing business environment.

As a result, organizations achieve a resilient, agile, and organic data architecture that evolves and grows in sync with their needs and ambitions. This adaptability fosters an environment where data can be effectively utilized to drive insights and support strategic initiatives, ultimately enhancing overall organizational performance.

CHAPTER VII:

DISCUSSION - A CASE STUDY WITH A BANKING INSTITUTION

7.1 Introduction

As a pivotal component of this research, a comprehensive case study was conducted with a national banking institution seeking to modernize its data practices and operationalize them through a standardized framework. This project exemplifies the integration of theoretical research with practical application, offering valuable insights into the challenges and opportunities inherent in modernizing data ecosystems within the banking sector.

Phase	Duration	Input	Purpose	Outcome	Roles Involved
Assessment and Planning	3-6 months	Existing data landscape, organizational goals, stakeholder needs	Evaluate current state, define objectives, develop roadmap	Data Mesh strategy, roadmap, data domain identification	CDO, data architect, data governance officer, business leaders
Technology Selection and Implementation	6-12 months	Data Mesh strategy, organizational requirements, budget	Select technologies, build pipelines, establish infrastructure	Functional data mesh platform, pipelines, self-service tools	Data architect, data engineer, cloud architect, security engineer
Data Migration	3-6 months	Data domains, data pipelines, migration strategy	Migrate data to data mesh	Populated data domains	Data engineer, data quality analyst, data domain owners
Enable Data Domains	3-6 months	Data domains, governance policies, self-service tools	Empower data domains with tools and resources	Domains capable of self-service	Data domain owners, data analysts, data scientists

Monitor and optimize	Ongoing	Performance metrics, user feedback, business needs	Monitor performance, identify improvements, optimize	Optimized data mesh, improved data quality, reduced costs	Data engineer, data analyst, data governance officer

Table 7.1 Roadmap for the Data Mesh rollout across the organization

The study involved conducting multiple in-depth interviews with key stakeholders, including executives, data experts, and operational leaders. These interactions provided a granular understanding of the bank's existing data strategies, highlighting both systemic challenges and potential avenues for improvement. Beyond interviews, the research adopted a collaborative approach, engaging closely with the bank's executive teams, data governance professionals, data engineering units, and cloud enablement teams to foster a holistic perspective. The study began with an in-depth analysis of existing data flows and the architecture of the data repository. This included a thorough examination of data integration patterns and the bank’s data quality practices to assess their efficacy and alignment with organizational objectives.

The research focused on identifying areas of improvement in the bank’s data warehouse, analytics, and reporting capabilities. By working closely with the bank's stakeholders, the study mapped the "as-is" data architecture, capturing the key interfaces and processes within the current setup. This analysis provided a foundation for evaluating the bank's data management framework and pinpointing bottlenecks, inefficiencies, and gaps in functionality.

A comprehensive data architecture review report was then delivered, detailing the current state of the bank’s data ecosystem. This report offered critical insights and actionable recommendations, outlining specific initiatives aimed at addressing identified challenges. The proposed future state roadmap provided a prioritized plan for the bank to achieve its modernization goals. This included recommendations for enhancing scalability, improving data quality, and adopting advanced analytics capabilities to support data-driven decision-making.

The project exemplifies how research-driven approaches can translate into practical outcomes, enabling organizations to reimagine their data strategies. By combining insights from industry interviews with hands-on collaboration and implementation, the study provided a robust framework for banking institutions to navigate their data modernization journey effectively.

7.2 Analysis of the Current Data Landscape

The analysis of the bank's existing data landscape, as conducted during the research, revealed a fragmented and siloed architecture, characterized by inefficiencies and inconsistencies that impede effective data utilization across the organization. The diagram presented in this section visualizes the bank's current data landscape, providing a holistic view of its fragmented state and highlighting the operational bottlenecks.

The following diagram created to represent the existing Data Landscape at bank:

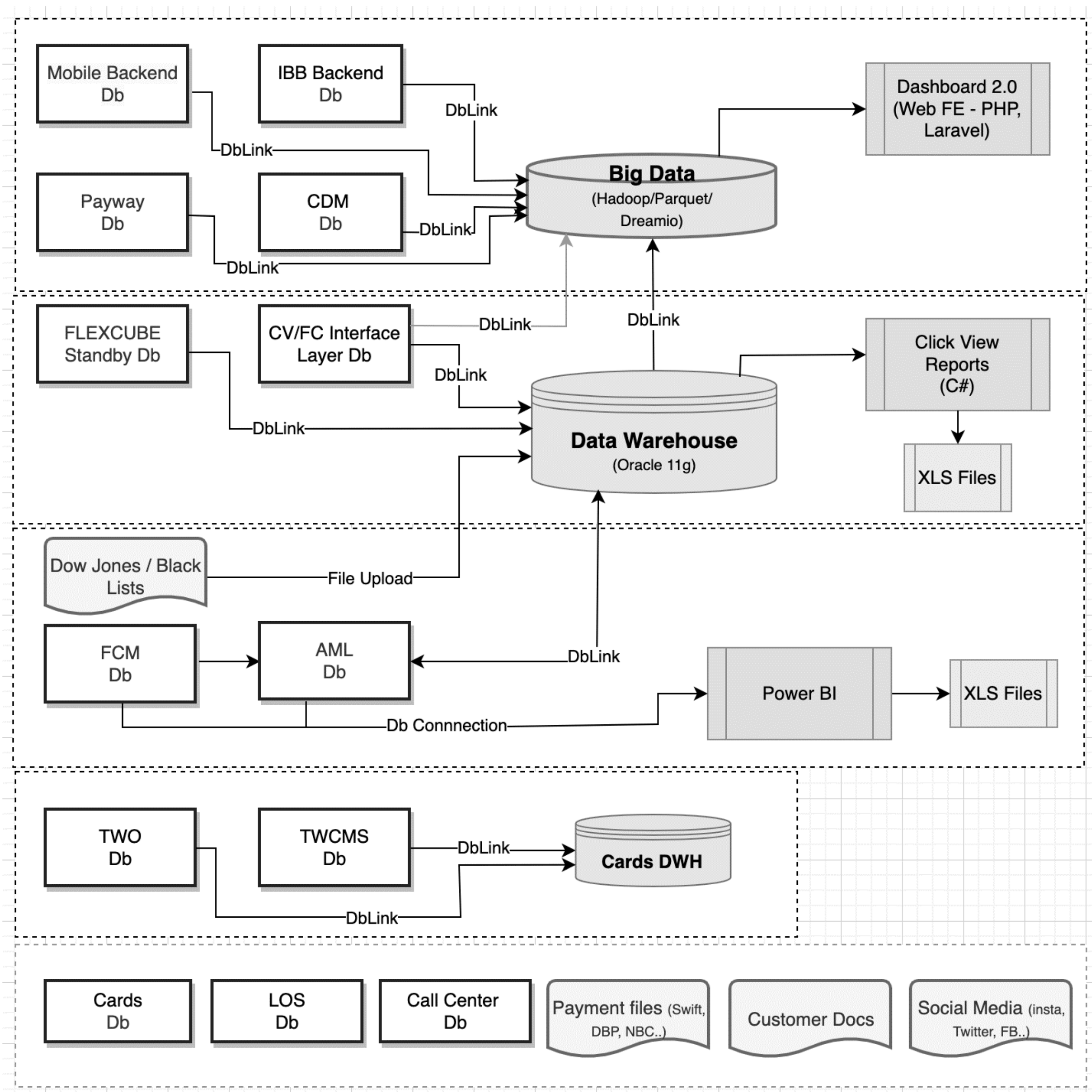

Figure 7.1 Existing Data Landscape of the bank

Fragmented Data Infrastructure: The bank operates under a decentralized data infrastructure where individual divisions maintain distinct data repositories designed to serve specific and limited purposes. This fragmented structure has resulted in significant inefficiencies, including redundant efforts and inconsistent data accessibility across the enterprise. A notable consequence is the absence of a centralized, enterprise-wide data repository capable of providing a unified "single source of truth" for all departments, an issue that undermines the ability to derive comprehensive insights.

Existing Systems and Their Limitations: Several systems and applications within the current landscape exhibit notable constraints:

Big Data Application and Dashboard 2.0: The bank implemented a Hadoop-based Big Data platform to support the Dashboard 2.0 initiative, which serves as a management tool for tracking business growth, channel performance, and product offerings on a daily basis.

Primary Functionality: Reporting for senior management and the generation of account statements for corporate customers with high transaction volumes.

Limitations: The system is primarily tailored to specific use cases, leaving broader analytics and integration capabilities unaddressed.

FLEXCUBE System: For less transaction-intensive accounts, the FLEXCUBE system manages account statement generation. However, its integration with other systems is minimal, limiting its utility as a data source.

Data Warehouse Utilization: The finance and risk departments rely heavily on the bank’s data warehouse (DWH) for analytical and reporting functions. However:

Data Sources: The DWH is populated exclusively by data from the FLEXCUBE standby database and its customer interface layer.

Scope Constraints: This limited data scope restricts the comprehensiveness of analyses, hindering strategic decision-making.

Compliance Systems: Compliance efforts leverage tools such as Power BI for analyzing and visualizing data from Financial Crime Monitoring (FCM) and Anti-Money

Laundering (AML) systems. Despite their utility, these tools operate in isolation, reinforcing the siloed nature of the bank's data landscape.

Cards Division: The cards division operates its own standalone data warehouse managed by the third-party system, Transware, further contributing to the decentralized data architecture.

Underutilized Data Sources: A significant portion of the bank's data, including structured and semi-structured sources such as Loan origination systems, Call center interactions, Payment SWIFT messages, Customer documents, Social media messages remains inaccessible or unintegrated within the bank's broader ecosystem.

Technological Fragmentation: The technological ecosystem reflects similar fragmentation:

Database Technologies: While most data repositories rely on Oracle databases, the Big Data platform employs a combination of open-source technologies, including Hadoop Distributed File System (HDFS), Dremio, Airbyte, and Parquet.

Data Integration: The absence of a standardized data integration framework forces the bank to rely on database links (DbLinks) for data movement between repositories, leading to inefficiencies and heightened maintenance requirements.

Governance and Reporting Practices: The governance and reporting mechanisms within the bank are inconsistent and fragmented:

Data Governance: Initial governance frameworks exist within specific divisions:

The risk and compliance teams have implemented measures to protect personally identifiable information (PII).

The cards division has partially adopted Payment Card Industry Data Security Standard (PCI DSS) compliance.

However, these efforts are not uniformly enforced across the organization.

Reporting Tools: The bank employs a range of reporting tools and technologies, including PHP, C#, and Power BI, resulting in redundancies and a lack of standardization.

Access Mechanisms: While the bank has a formal data access mechanism in place, requiring appropriate approvals for granting access, the fragmented nature of the data infrastructure complicates the process and limits its efficiency.

This research analysis underscores the need for a cohesive and integrated data strategy to address the operational inefficiencies and unlock the potential of the bank’s data assets. Without intervention, the current state will continue to limit scalability, operational agility, and the ability to deliver data-driven insights. A unified, enterprise-wide data platform is essential for addressing these challenges and supporting the bank’s strategic goals.

7.3 Key Concerns:

Through an in-depth engagement with the bank’s data management teams during the research study, several critical concerns were identified that significantly impact the organization’s ability to effectively manage and leverage its data assets. These concerns emerged from a detailed examination of the bank's data practices, infrastructure, and operational workflows. Furthermore, the study also uncovered priority use cases that highlight the pressing needs and strategic opportunities for the bank to enhance its data capabilities.

Sr. No.	Key Concerns	Impact
1.	The Bank has one single point of truth for their financial and customer data which is FC DB. The bank provides all financial and regulatory reports from core banking. There is no centralized data base for non-financial data	• Customer Experience • Consistency • Efficiency • Productivity • Innovation
2.	A High Availability Architecture is required for the Enterprise Data Platform which is currently not accomplished with the chosen technology of Hadoop and Dremio.	• Availability • Reliability • Customer Experience
3.	Few of the key Data Entities do not have a clear Data Owner e.g., Merchant Data which is largely owned by Payway but is also partly owned by Card division. Also, Transaction data that is originated from Channels is also owned by In-house FC Custom Integration Layer.	• Efficiency • Data Integrity
4.	Each division has its own Data Management, Data storage and Reporting strategies, design and tools with no unified standard operating and governance model across all divisions	• Fragmented Data • Data Integrity • Data Insights • Efficiency
5.	Data organization and related Roles and Responsibilities are yet to be defined	• Maintainability • Efficiency
6.	There is no enterprise-wide data platform and bank would want to implement a Modern Data Platform with following capabilities: - Cover as much Data as possible across the organization - Process Real time and Batch data - Handle Structured, unstructured and Semi-structured data - institutionalized Data Governance and Quality processes & tools - Enable Self Service Analytics and Reporting - Enable publishing of reports via available channels (email, portal, H2H etc) - Enable Advance analytics using AI ML models - Enable Time Series and Predictive reporting	• Data Insights • Efficiency • Innovation • Maintainability • Data Trust • Customer Experience

	- Data Encryption and protection - Data Visualization with Drill down for leaf level data	
7. `	A common Data definition & taxonomy for Key Entities and their States, Attributes and Derived values etc needs to be put in place	• Data Insights • Efficiency • Maintainability • Customer Experience
8.	Though visualization tools who needs they have it. More interactive visualized tools are recommended for those who deals with analytics works	• Efficiency • Productivity • Innovation • Data Insights
9.	Data Governance and Change management processes are on different level in different departments. These will become critical for any Enterprise Big Data initiative	• Data Quality • Maintainability • Efficiency
10.	There is no application that can help RMs / Branches to manage and offer differentiated Service to 'High Relationship' customers	• Customer Experience • Efficiency • Innovation
11.	Need insights for RMs to reach out to new Corporate Customers and increase the wallet share of existing customers	• Customer Experience • Efficiency • Business Growth
12.	Transaction history, Customer profile, counter party details are not readily available for AML alert case management	• Productivity • Compliance Risk
13.	Getting incremental data from FLEXCUBE for Customer and Account Entity is typically a challenge, leading to full data loads into DWH on daily basis	• Efficiency
14.	Building of MIS GL in DWH requires processing of Transaction and Balance data across all accounts.	• Efficiency
15.	When FLEXCUBE version is upgraded to FC 14.x, not all Customer, Accounts and Transaction data will be migrated to Target version. History Data retention for query and compliance purpose will need to be managed.	• Compliance • Customer Experience

Table 7.2 Key concerns identified during the research study within the bank

7.4 Identification of Priority Business Use Cases

As part of the collaborative research study, the digital division of the bank’s data team identified several business use cases that could serve as focal points for prioritization by the management. These use cases align with the bank's strategic objectives, particularly in enhancing customer experience, optimizing financial performance, and leveraging advanced analytics for competitive advantage.

Among the most critical use cases, those aimed at improving customer experience were given prominence. These include initiatives such as creating consolidated customer 360-degree relationship-level reporting, performing spend analytics, and developing business finance benchmark dashboards and reports. Additionally, the deployment of real-time analytics to generate contextually relevant ‘next best offers’ for customers was highlighted as a high-priority area with significant potential for value creation.

Beyond customer-centric initiatives, other analytical capabilities were identified for consideration. These included predictive analytics for profit, cost, and cash flow forecasting, as well as competition analysis and wallet share analytics. Furthermore, the bank recognized the value of predicting shifts in business growth factors, including macroeconomic and microeconomic indicators, which would enable proactive decision-making.

Other key areas included fraud detection and compliance monitoring, channel performance analytics, customer segmentation and profitability analysis, product profitability assessments, and tools to evaluate customer service quality, such as Net Promoter Scoring (NPS). These use cases reflect the bank’s commitment to leveraging data as a strategic asset to drive operational efficiency and improve decision-making processes.

7.5 Strategic Recommendations for the improvement of the bank

Following the identification of key concerns and business use cases, the research study progressed to developing actionable recommendations for addressing the identified challenges. These recommendations were framed within two distinct data management approaches: a centralized data management model and a distributed data mesh operating model. The Improvement recommendations that would help address the identified Key Concerns in the earlier section are detailed in the table below:

Code No	Improvement Recommendation	Description	Transformation Value	Transformation Complexity
DA01	Implement an Enterprise level Modern Data Platform using Open-source Technologies	Technical Recommendation: Centralized or decentralized implementation mentioned in detail in section 5 of the document.	High	High
DA02	Explore implementation of CDC (Change Data Capture) for Dimensional data entities to avoid full load of Data into MDP	Technical Recommendation: Change Data Capture can be implemented using any event bus mechanism such as open-source Debezium or Oracle Golden Gate	Medium	High
DA03	Establish a Data Strategy for bank across all Divisions & Functions	Process/Approach Recommendation: A Strategy and Design approach that defines, Key Business Drivers, Data Organization, Domains & Owners, Dictionary, Modern Data Architecture, Technology choices, Governance etc., to enable a data-driven organization	High	High

DA04	Create a Backlog of Business Use cases for Reporting, Analytics & Insights	Process/Approach Recommendation: Early Involvement of Business and Operations users to document Big Data Use Cases and Priorities	High	Medium
DA05	History data storage & access approach, need to be designed upfront for Core Upgrade	Process/Approach Recommendation: Create Data models in the existing DWH so that it can store historical data up to a specific period	Low	Low

Table 7.3 List of Recommendations to the bank

7.6 Proposal for Establishing a Modern Data Platform (MDP)

As part of the research case study, a dual-solution approach was developed to address the bank's data management challenges and support its strategic objectives. A critical outcome of this study was the recommendation to establish a Modern Data Platform (MDP) at an enterprise level, leveraging contemporary concepts and technologies to create a unified and standardized data management framework across the bank's divisions.

The Need for a Modern Data Platform: The study identified a pressing requirement for a scalable and integrated data platform capable of managing the diverse data landscape of the bank. A Modern Data Platform (MDP) represents a unified, enterprise-wide solution designed to address the complexities of storing, managing, processing, and analyzing vast volumes of structured, semi-structured, and unstructured data. The proposed architecture would support the IT, Digital, and Cards divisions, ensuring seamless collaboration and enabling advanced data-driven decision-making capabilities.

The proposed Modern Data Platform (MDP) is designed to address critical aspects of the bank's data lifecycle and optimize its data management practices through a cohesive and innovative approach. The platform comprises several key components, each serving distinct yet interconnected functions within the architecture.

The Data Ingestion Layer: It serves as the foundation for integrating data from diverse sources, including databases, data warehouses, files, and streaming systems. By consolidating information from divisions such as IT, Digital, and Cards, this layer ensures a seamless flow of data across the organization. It standardizes data formats through cleansing and transformation processes, thereby enhancing consistency and reliability. This streamlined approach to data acquisition facilitates efficient preparation of datasets for storage and subsequent analysis.

The Data Storage Layer: It provides a scalable and reliable infrastructure tailored to managing structured, semi-structured, and unstructured data. Leveraging distributed file systems, data warehouses, or hybrid storage solutions, the storage layer strikes a balance between cost-effectiveness and high performance. It is designed to meet the bank's requirements for scalability and flexibility, while enabling the efficient organization of data. This structure ensures readiness to accommodate future growth and diversification in data sources.

Central to the platform is the data processing layer, which transforms raw data into actionable insights. By employing scalable tools and algorithms, this layer supports advanced analytics capabilities, offering both real-time and batch processing functionalities. Its ability to process and analyze data efficiently empowers stakeholders with insights that inform strategic and operational decisions.

The Data Analytics Layer: It acts as the interface for presenting processed data to decision-makers. Designed to serve divisions such as IT, Digital, and Cards, this layer simplifies complex analytics by translating them into intuitive visualizations. The accessibility and clarity of insights enhance decision-making processes and facilitate effective communication across stakeholders.

The MDP integrates cutting-edge technologies, including open-source solutions, to create a robust and cost-effective platform. These technologies ensure seamless alignment with the bank's existing infrastructure while fostering innovation and modern data management practices.

Overall, the enterprise-level MDP addresses inefficiencies and enhances scalability within the bank's data ecosystem. By unifying data management practices across divisions, it provides a solid foundation for advanced analytics and data-driven decision-making. This platform aligns closely with the bank's digital transformation objectives, positioning it for sustained success in an increasingly data-centric and competitive environment. Below is the MDP detailed conceptual architecture.

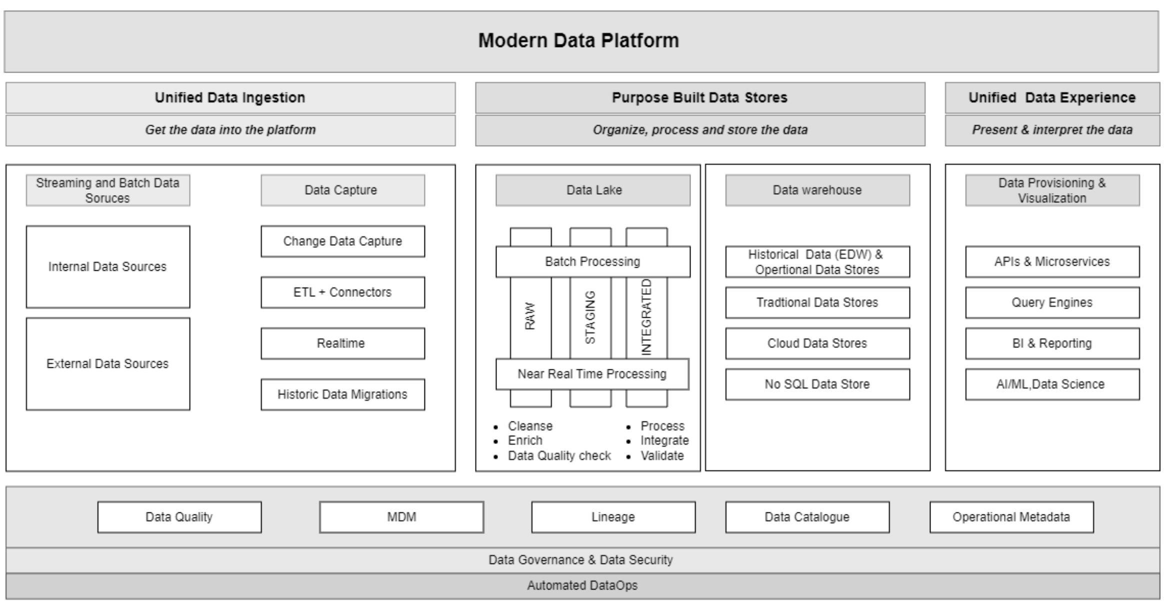

Figure 7.2 Modern Data Platform: Independent layers to perform independent function of data management.

By implementing these components of the Enterprise Modern Data Platform, the bank would get various benefits such as scalability, reliability, flexibility, cost-effectiveness, faster processing, efficient data transfer, self-service analytics, and compliance with regulations.

Typical Technology Choices for MDP: Modern Data Platform is an enterprise repository and processing house for all Bank's data. This MDP proposition can be implemented using Centralized or Distributed data architecture style to meet the diverse needs of data collection, management, processing, and analytics.

Typically, MDP leverages cloud services, commercial tools such as Golden Gate, Stitch, StreamSet, Informatica, Open-Source frameworks such as Dremio, MinIO, Apache Kafka, Apache Spark and other Apache open-source ecosystem. Based on the current tools' investment and best in class Open-Source frameworks and future requirement of data management, we would recommend building open source based MDP and then optimize and enhance it using Cloud services as per the Banks's future cloud adoption strategy. Below is the typical Open-source Technology stack for MDP. However, during its actual planning and implementation, I advise to re-look into these tools and frameworks.

MDP Key Components	**Open-Source Technologies**
Batch Data Ingestion	Airbyte/ Apache NiFi
Real Time Ingestion	Apache Kafka
Orchestrator for Batch Ingestions	Apache Airflow
Fast Storage	Apache Kafka

Object storage – Data Lake	MinIO (S3 compatible)
Data Warehouse (Data Storage & archival)	Oracle 19c on Exadata
Batch Data Processing & analytics	Apache Spark/Dbt
Real Time Processing & Analytics	Kafka Streaming/Apache Spark
Data Lakehouse	Dremio
Data Integration/Distribution	Apache Kafka/Dremio
Metadata/Catalog Management	Apache Atlas
Data Quality Management	Apache Griffin
Data Governance	Apache Atlas
Reporting and Dashboards	Tableau/Apache Superset / Power BI
Data Security and Privacy	Apache Ranger or Apache Metron

Table 7.4 List of Recommended technologies to the bank

As a part of the study, below are the two approaches or operating models recommended to the banks.

7.7 Proposed Centralized Data Model for bank.

The centralized data model proposed for the bank emphasizes creating a single repository to store all organizational data, ensuring a unified source of truth. This model enhances data visibility, collaboration, and consistency across the enterprise. A Data Lakehouse architecture is identified as an optimal solution, capable of managing large volumes of structured, semi-structured, and unstructured data while addressing limitations in traditional architectures.

Unlike legacy systems requiring external storage and duplication for analytics, the Data Lakehouse integrates storage and compute resources internally. This integration facilitates rapid data access and efficient analytics, significantly improving performance and reducing delays. The architecture retains data in its raw form, preserving its original format and schema, thus enabling diverse datasets to coexist seamlessly.

The flexibility of the Data Lakehouse stems from its schema-on-read approach, which adapts dynamically to changing business needs. By processing data only when required, this approach allows enterprises to respond effectively to market shifts and evolving demands. It supports a wide range of data types, making it particularly suited for organizations undergoing rapid growth or transformation.

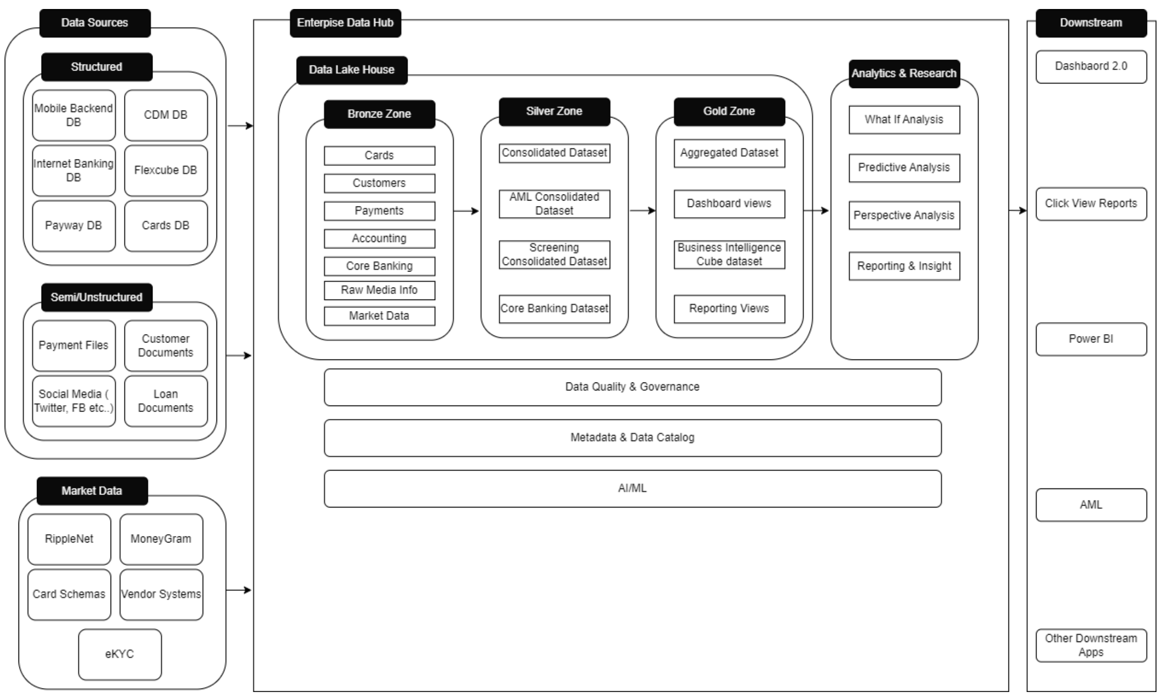

Figure 7.3: Centralized Data solution identified as one option during research study for bank.

Data management within the Lakehouse is comprehensive, encompassing raw and processed datasets while supporting robust mechanisms for organizing and evolving data structures. Centralized storage simplifies governance and ensures efficient integration of data sources, formats, and schemas. Furthermore, the system ensures complete traceability, documenting data lifecycle stages from ingestion to analytics, which is vital for compliance and auditability.

The Lakehouse also supports diverse computational requirements, including batch processing, real-time streaming, interactive analytics, and machine learning. These capabilities enable the bank to leverage advanced analytics tools for data-driven decision-making while maintaining scalability and performance.

The proposed centralized data model, anchored by the Data Lakehouse architecture, offers a transformative framework for the bank's data management strategy. Its integration of scalability, flexibility, and advanced analytics provides a robust platform to meet current and future organizational needs.

7.8 Proposed Data Mesh Operating Model for bank

The proposed Enterprise Modern Data Platform (MDP) for the bank is designed to leverage the principles of the Data Mesh architecture pattern. This approach seeks to dismantle the limitations of traditional monolithic data systems by decentralizing data management into smaller, autonomous units organized around specific business domains. Each domain assumes complete ownership of its data, enabling scalability, flexibility, and enhanced autonomy for teams. Such an architecture fosters an environment for faster innovation, more agile decision-making, and a heightened ability for business executives and stakeholders to access and utilize data effectively.

At the core of this proposition is the concept of treating **data as a product.** This paradigm emphasizes clean, reliable, and comprehensive data being made available to all consumers across the enterprise, governed by appropriate permissions and role-based access. Each business domain within the organization is envisioned as a **data product domain** supported by a dedicated cross-functional team. These teams are responsible for every aspect of their domain's data lifecycle, including collection, storage, processing, and governance. By aligning ownership directly with domain-specific teams, this model addresses longstanding challenges such as fragmented data silos, bottlenecks in scalability, and ambiguous ownership.

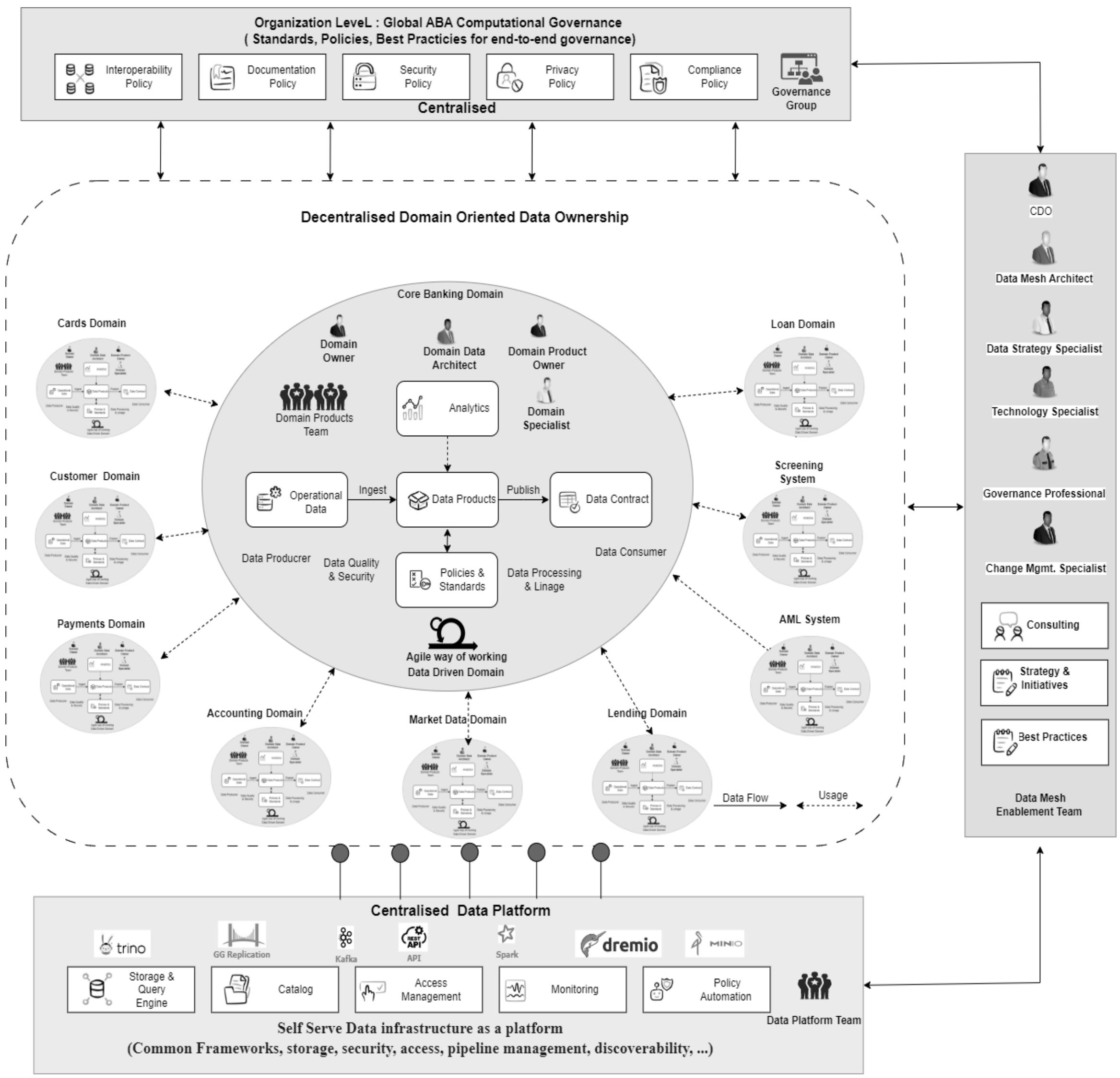

Figure 7.4: The proposed Data Mesh Operating model for the bank during the research

The implementation of the Data Mesh architecture fundamentally reshapes the organization's approach to data management, prioritizing decentralization and collaboration over rigid, hierarchical structures. By assigning responsibility and accountability for data directly to the teams closest to its source and most familiar with its utility, the model ensures that data remains accurate, timely, and relevant. Furthermore, by embracing a product-oriented mindset, data becomes a consumable asset, ready to serve stakeholders with minimal friction and enhanced agility.

The above proposed model mentioned in figure 7.4, not only resolves the inefficiencies present in the current data management framework but also positions the bank to remain agile and adaptive in a rapidly evolving, data-driven business landscape. By fostering a culture of empowerment, the Data Mesh architecture enables the organization to harness the full potential of its data assets, creating a robust foundation for innovation and sustained growth. This vision aligns with the bank's strategic goals, ensuring that it remains at the forefront of modern data management practices.

7.9 Proposed Organization Structure

Establish the Organization Structure to adopt, plan and execute the Data Mesh journey across the bank. Below is the recommended Organization structure for Data Mesh.

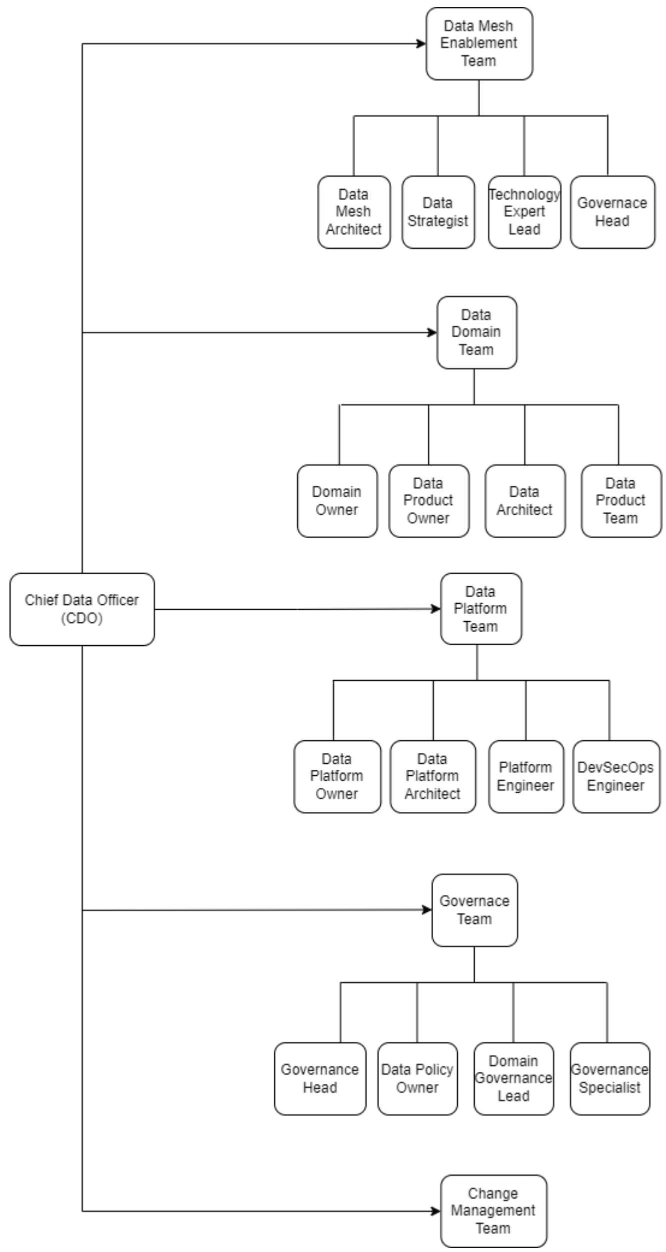

Figure 7.5: Proposed Data Mesh Organization Structure for the Bank

To ensure the successful implementation and execution of the Data Mesh architecture across the bank, a well-defined organizational structure is essential. This structure facilitates effective planning, adoption, and execution of the Data Mesh principles, fostering a culture of decentralization, collaboration, and continuous improvement. The proposed structure is aligned with the bank's strategic objectives, ensuring that Data Mesh adoption occurs smoothly and efficiently.

At the top of the structure is the **Chief Data Officer (CDO)**, who holds overarching responsibility for defining and overseeing the bank's data strategy and governance. The CDO plays a critical role in aligning data initiatives with business goals and ensuring compliance with regulatory standards. This central leadership position ensures the bank's data initiatives are coherent and well-integrated into broader organizational objectives.

Supporting the CDO is the **Data Mesh Enablement Team**, which is composed of experts in data architecture, governance, technology, and strategy. The enablement team provides the necessary guidance, expertise, and resources to support the bank's transition to Data Mesh. The team works to ensure that the principles of Data Mesh are understood and adopted across the organization. Key roles within the team include the **Data Mesh Architect**, who is responsible for designing the Data Mesh framework and ensuring its implementation across various domains; the **Data Strategist**, who aligns Data Mesh with the bank's business strategy and objectives; the **Technology Expert Lead**, who provides technical advice and ensures the right tools and platforms are in place; and the **Governance Head**, who is responsible for establishing and maintaining data governance practices, ensuring that data management complies with both internal standards and external regulations.

The core of the organizational structure consists of the **Data Domain Teams**, each dedicated to a specific business domain, such as Core Banking, Payments, Vendor & Billing, and Cards. These domains span across multiple divisions, including IT, Digital, and Cards. Within each domain, there are cross-functional teams that are responsible for the end-to-end management of data. These teams have full ownership of their data domains, meaning they are responsible for the collection, storage, processing, and governance of their respective data. Each domain team is led by a **Domain Owner**, who is responsible for the overall strategy and success of the data within the domain. Additionally, the **Data Product Owner** is tasked with managing specific data products within the domain, ensuring they meet business and technical needs. The **Data Architect** within each team designs the data structures and systems required to support the operations of the domain, while the **Data Product Team** works on the development, maintenance, and enhancement of data products, ensuring they meet the needs of both the domain teams and other stakeholders.

To support the domain teams, a centralized **Data Platform Team** provides the infrastructure and tools needed for data management. This team is responsible for maintaining the modern data platform that supports self-service capabilities for data ingestion, storage, querying, and visualization. The **Data Platform Owner** leads this team, ensuring that the platform's capabilities align with the bank's broader data strategy. The **Data Platform Architect** designs the technical architecture of the platform, ensuring it supports decentralized data management, while the **Platform Engineer** ensures the platform's technical infrastructure is scalable and reliable. The **DevSecOps Engineer** ensures that security, compliance, and automation practices are integrated into the platform's operations.

In addition to the domain teams and platform support, a **Governance Team** is crucial for ensuring that data is used appropriately and consistently across the organization. This team works closely with the domain teams to establish and enforce governance policies that ensure data is managed in a compliant and standardized manner. The **Governance Head** leads this team and is responsible for overseeing data policy development and compliance enforcement. The team also includes specialists in policy management and data governance, such as the **Data Policy Owner**, who ensures the global policies align with both regulatory requirements and internal standards, and the **Domain Governance Lead**, who ensures that each domain follows established governance guidelines. A **Governance Specialist** supports these efforts by providing expertise on best practices for data governance across the organization.

Furthermore, a **Global Governance Group** is proposed as a guild consisting of representatives from all domain teams. This group collaborates to define global policies and best practices that guide the creation and management of data products across the organization. This collaborative approach ensures consistency in the implementation of the Data Mesh and promotes shared responsibility for data quality and governance.

The organizational structure also emphasizes the importance of **Change Management** in the successful adoption of the Data Mesh architecture. Change management specialists are integrated within the enabling team and work with the domain teams to facilitate the cultural shift toward decentralization. They ensure that the necessary skills and knowledge are developed within domain teams and help guide them through the implementation process. These specialists temporarily integrate into domain teams, acting as internal consultants to provide hands-on support in adopting Data Mesh principles, upskilling team members, and establishing best practices.

Finally, the **Modern Data Platform** is the backbone of the Data Mesh, supporting a self-service approach to data product creation and management. This platform enables domain teams to independently create, monitor, and discover data products, while also facilitating cross-domain access and collaboration. The platform is designed to enforce governance policies automatically, ensuring that global policies are adhered to without manual intervention. This automation helps ensure consistency and compliance across all data products and domains. The platform also supports the creation of a data catalog, which serves as a central repository for discovering and accessing data products, promoting transparency and collaboration among teams.

This organizational structure provides a clear, scalable framework for the successful adoption and implementation of the Data Mesh within the bank. It empowers domain teams to take ownership of their data, fosters collaboration, and ensures that governance and compliance are maintained across the organization. By adopting this structure, the bank will be able to transform its data management practices and create a sustainable, future-proof data ecosystem.

7.8 Proposed responsibilities for the Data Mesh Operating Model

In the context of adopting a Data Mesh for the bank, various roles are crucial to the implementation and maintenance of this approach across the enterprise. A comprehensive organizational structure, outlined in an earlier chapter, provides a classical framework for the deployment of a bank-wide Data Mesh platform. This structure serves as a foundation, with the flexibility to merge certain roles during the initial phases of the initiative and gradually segregate them as the implementation matures.

The **Chief Data Officer (CDO)** assumes overall responsibility for the corporate data strategy and ensures its alignment with the bank's broader business strategy. This role serves as the cornerstone for data-related initiatives, bridging the gap between data management practices and strategic business objectives.

Within the Data Mesh framework, **Data Domain Owners** oversee specific data domains, holding end-to-end accountability for the data within their areas of responsibility. These responsibilities encompass defining the scope and boundaries of their domains, understanding stakeholder needs, and ensuring that data collection, storage, processing, and governance align with the bank's business objectives and compliance requirements. Domain owners work closely with Data Product Managers to define and prioritize the development of data products and features. Additionally, they are tasked with monitoring data quality, taking corrective actions when needed, and fostering a culture of data ownership and collaboration within their domains.

The **Data Product Owner** is responsible for the business vision and evolution of the data product. This individual may also function as an Agile Product Owner or Product Manager in practice. The Data Product Owner is ultimately responsible for defining the vision and strategic direction of the data product, including identifying its purpose, target users, and expected outcomes. This role includes creating development roadmaps, defining performance indicators, and ensuring adherence to service level agreements (SLAs). The Data Product Owner also ensures compliance with governance standards, manages the product backlog, gathers and prioritizes requirements, and engages with stakeholders to clarify inconsistencies or conflicting needs. Collaboration with the development team and participation in governance discussions are integral to the success of the data product.

Data Product Managers work closely with domain owners and stakeholders to identify opportunities for data product development within their domains. They define the vision, strategy, and roadmap for these products while prioritizing development based on business value and stakeholder requirements. Their responsibilities include ensuring the timely delivery of data products, gathering feedback from users, and iterating on the products to improve their value. Furthermore, they measure the success and impact of data products and communicate insights to stakeholders.

The **Domain Data Architect** plays a key role in designing the data flow from source systems to the final products. This includes defining data product models and ensuring adherence to architectural standards within each domain. Supporting this effort, **Domain SMEs (Subject Matter Experts) or Analysts** contribute their expertise in domain-specific business areas, ensuring that data products are aligned with consumer needs and business objectives.

The **Data Product Development Team** is responsible for the creation and maintenance of data products within their domain. This team collaborates with domain teams to translate data requirements into scalable solutions, implementing data transformations, cleansing, enrichment, and advanced analytics techniques. It includes specialists such as data engineers, operations engineers, software developers, data scientists, data analysts, testers, and security engineers, each contributing unique expertise to ensure the robustness of the data products.

To support domain teams, **Data Platform Teams** build and maintain the underlying infrastructure and platforms required for Data Mesh operations. These teams ensure the availability of self-service tools, frameworks, and services to empower domain teams to independently develop and manage data solutions. Responsibilities include providing technical guidance, monitoring platform performance, ensuring security and compliance, and staying abreast of emerging technologies to enhance platform capabilities.

Data Quality and Governance Specialists focus on maintaining data quality, compliance, and governance across the Data Mesh. These specialists establish standards, monitor data quality, address issues, and ensure alignment with governance policies and regulations. Governance operates at three levels: global governance, domain governance, and technical governance.

At the **Portfolio or Global Governance** level, policies are automated and applied at the organizational level. These policies guide domain teams in adopting compliant data product development and management practices. The **Domain Governance** level allows each domain to govern its data products while adhering to global standards and addressing any domain-specific requirements. **Technical Governance** ensures interoperability between data products by setting development standards and guardrails for domain and product teams.

A **Data Mesh Enablement Team** facilitates the adoption and implementation of Data Mesh principles. This team educates stakeholders, assesses the current state of the architecture, develops roadmaps, and defines domain boundaries. It also provides guidance on governance policies and supports the development of self-serve platforms. The team includes Data Mesh Architects, who design the overall architecture; Data Strategy Specialists, who define the strategic roadmap; Change Management Experts, who drive cultural shifts; Data Governance Professionals, who ensure compliance; and Technical Specialists, who provide infrastructure support.

Finally, **Data Consumers** are key stakeholders who utilize the data products to derive insights and make informed decisions. These include business analysts, data scientists, and decision-makers who collaborate with domain teams to refine data products and apply analytical techniques to achieve business outcomes.

In conclusion, the Data Mesh architecture represents a paradigm shift toward decentralized, domain-oriented data management. By emphasizing domain-driven ownership, data as a product, self-serve platforms, and federated computational governance, this approach aligns with the bank’s vision for a future-proof and scalable data ecosystem. This operating model decentralizes traditional data practices, enabling autonomous teams to perform cross-domain analysis while ensuring compliance and innovation.

7.9 Conclusion

Within the broader scope of this thesis, this case study reinforces the practical value of the proposed operating model for distributed data mesh and multi-cloud frameworks. By applying theoretical insights in a real-world setting, the research demonstrates how banking organizations can transform their data ecosystems to achieve enhanced scalability, operational efficiency, and data-driven decision-making.

Through a rigorous and collaborative approach, this research study exemplifies the critical intersection of research and practice, ensuring that the outcomes of this study are both theoretically sound and practically implementable. The success of this case study adds substantial weight to the thesis, validating its recommendations and showcasing the transformative potential of adopting innovative data strategies within the banking sector.

CHAPTER VIII:

DISCUSSION –DATA MESH ON DIFFERENT CLOUD

8.1 Introduction

The research explored the feasibility of implementing Data Mesh architectures across major cloud platforms, including Azure, AWS, Google Cloud Platform (GCP), Oracle Cloud Infrastructure (OCI), and IBM Cloud. Each platform offers a suite of tools and services that align with the principles of Data Mesh, such as decentralized ownership, self-service capabilities, federated governance, domain integration, and scalable infrastructure.

Azure provides a comprehensive ecosystem for building decentralized data products and ensuring seamless governance and scalability.

AWS offers robust support for data product development, integration, and governance through its extensive cloud-native services.

GCP facilitates Data Mesh implementation via tools like Dataplex for domain ownership, BigQuery for data product creation, and Cloud IAM for federated governance.

OCI enables data decentralization using services like Autonomous Database and integrates domains via Oracle Integration Cloud, coupled with governance tools like IAM.

IBM Cloud supports Data Mesh through Cloud Pak for Data and DataOps practices, promoting governance and seamless integration with its middleware tools.

This research confirms that the proposed 10-step operating model for Data Mesh is versatile and applicable across these platforms. The model ensures alignment with strategic goals, fosters a data-centric culture, establishes domain ownership, integrates cloud-native tools, and builds a scalable and interoperable data architecture. (A. Bhattacharya et al.,2022)

By leveraging the unique capabilities of each cloud provider, organizations can adopt a Data Mesh framework that enhances agility, collaboration, and innovation. The findings underscore that while cloud-specific implementation details may vary, the foundational principles and structured approach of the proposed operating model remain universally relevant, providing a roadmap for effective Data Mesh adoption across diverse cloud environments.

8.2 Comparative Analysis of Centralized Data Management and Data Mesh in Cloud Environments

The research study also focused on the centralized Data Management and Data Mesh in cloud-based implementations and below is the comparison study.

Feature	Centralized Data Management (Cloud)	Data Mesh (Cloud)
Data Ownership	Centralized (typically managed by a central IT team)	Decentralized (managed by individual data domains)
Governance Model	Centralized (controlled by a central governance body)	Federated (combines centralized oversight with domain-specific autonomy)
Data Architecture	Monolithic (data stored in a single central data lake or warehouse)	Distributed (data spread across multiple data domains)
Scalability	Can be scaled using cloud resources, but may face limitations in large-scale environments	Highly scalable due to distributed architecture and cloud-native services

Agility	Can be improved with cloud-based tools, but may still be limited by centralized control	Highly agile due to decentralized ownership and domain-specific autonomy
Data Quality	Can be challenging to maintain in large-scale cloud environments	Improved through domain-specific ownership and data governance
Security	Can be managed using cloud-based security tools, but requires careful configuration	Enhanced through domain-specific security controls and cloud-native security features
Cost	Can be cost-effective for smaller datasets, but can become expensive for large-scale data storage and processing	Can be cost-effective due to optimized resource utilization and pay-as-you-go pricing models

Table 8-1: Centralized Data Management vs. Data Mesh in Cloud-Based Implementations

Based on the above comparison study, below is the summary of the findings between these two models on cloud.

The research undertakes a comparative study of centralized data management and Data Mesh models in cloud environments, presenting a nuanced understanding of their applicability and advantages. The findings underscore the suitability of each model based on organizational needs, data complexity, and operational objectives.

Centralized data management on the cloud is well-suited for smaller organizations or those with relatively straightforward data requirements. This model offers centralized control, providing a single point of governance that can simplify oversight and compliance. However, as organizations scale, the centralized approach encounters limitations, particularly in managing large, complex datasets. Scalability and agility become significant challenges in large-scale environments, where data volumes and processing demands can overwhelm centralized systems.

In contrast, the Data Mesh model is tailored to address the needs of organizations dealing with extensive and diverse datasets. By decentralizing data ownership and governance, Data Mesh offers a framework that enhances scalability and operational agility. This approach empowers data domains to independently manage their data, fostering greater accountability and enabling domain-specific innovations. The decentralized nature of the Data Mesh also supports faster adaptation to changing business requirements, making it an ideal choice for organizations prioritizing rapid development and deployment of data-driven applications.

The choice between centralized data management and Data Mesh in a cloud environment is influenced by several key factors. For organizations handling large and complex datasets, the Data Mesh model provides a more scalable and flexible solution. The decentralized governance inherent in Data Mesh is particularly advantageous for organizations operating under stringent regulatory requirements, offering a balance between oversight and autonomy. Furthermore, organizations seeking to drive innovation and enhance agility in their data strategy find the Data Mesh model more conducive to achieving these goals. (M. Amr,2021).

8.3 Conclusion

In conclusion, this comparative analysis highlights that while centralized data management retains its relevance for smaller, less complex use cases, Data Mesh emerges as a transformative model for larger organizations with intricate data landscapes. The choice of model must align with the organization’s data volume, governance needs, and aspirations for agility and innovation. This research underscores the importance of a tailored approach, where the selected data management strategy is closely aligned with the organization's operational and strategic imperatives.

The research emphasizes the growing inclination of banks to redefine their data strategies through the adoption of Data Mesh principles. The proposed ten-step operating model serves as a comprehensive conceptual framework, guiding banks to strategically plan and execute their data transformation initiatives. Grounded in the principles of decentralization, the model enables data domains to take ownership of their data, fundamentally altering the way banks approach data management and governance.

The Data Mesh operating model offers significant advantages for banks. Decentralized data ownership enhances agility, enabling banks to respond more swiftly to evolving business needs and market dynamics. Moreover, the model fosters improved data governance by promoting accountability and ownership within individual domains. This decentralization ensures higher data quality and strengthens security measures, which are critical in highly regulated environments like banking.

Cost optimization emerges as another crucial benefit, as Data Mesh facilitates efficient resource allocation and reduces expenses associated with centralized data management systems. Simultaneously, the model encourages innovation by empowering data domains to experiment and develop new data-driven products and services tailored to specific business needs.

A distributed operating model forms the foundation for governing and managing a Data Mesh architecture within a banking context. It establishes clear roles, responsibilities, and processes to ensure seamless operations. Components of this framework include assigning ownership of data to specific business units, implementing robust governance policies, adopting scalable data platforms, designing efficient pipelines, and providing self-service tools for data consumers.

To illustrate the transformative potential of this approach, consider the case of a large bank transitioning from a centralized data warehouse to a decentralized Data Mesh framework. The research revealed that defining distinct data domains—such as customer data, financial data, and risk data—and assigning ownership to relevant business units allowed for tailored data management practices. Governance policies were established at the domain level, encompassing data quality standards, security protocols, and access controls. Gradual migration of data from centralized systems to distributed domains ensured continuity while fostering a culture of accountability and empowerment.

This transformation enabled business units within the bank to assume greater control over their data, leading to informed and timely decision-making. Data quality improved as it was managed closer to the source, reducing discrepancies and errors. Security measures became more effective by tailoring them to domain-specific requirements, and innovation accelerated through domain-level experimentation with data-driven solutions. Furthermore, resource utilization became more efficient, reducing costs while maintaining high operational standards.

In conclusion, the research validates that adopting a distributed Data Mesh supported by a robust operating model can significantly enhance a bank's data strategy. By leveraging the principles of decentralization, fostering collaboration, and embracing tailored governance, banks can unlock the full potential of their data assets. This approach not only addresses traditional challenges of scalability, governance, and cost-efficiency but also positions banks for sustained growth and competitive advantage in an increasingly data-driven landscape.

CHAPTER IX:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

9.1 Summary

This chapter presents a comprehensive summary of the research study, encapsulating the findings, conclusions, and their implications for banking organizations. The research rigorously addressed the research questions articulated in Chapter IV, offering valuable insights into the operationalization of modern data platforms through distributed data mesh and multi-cloud architectures.

The study began by examining the current data architecture styles prevalent in banking organizations. A thorough analysis of their challenges highlighted the limitations of traditional systems. Based on these findings, data mesh emerged as the recommended architectural approach to address these limitations and support the evolving needs of banking organizations.

The research meticulously compared and contrasted data warehouse, data lake, data fabric, and data mesh architectures, uncovering their unique strengths and weaknesses. Through this comparative analysis, the study demonstrated the superiority of data mesh for organizations seeking enhanced scalability, governance, and agility in their data management systems.

A significant contribution of this research lies in addressing the existing gaps in literature and practice regarding the operationalization of modern data platforms, particularly those leveraging distributed data mesh and multi-cloud environments. The study identified the core components of an effective operating model tailored to this paradigm and proposed a 10-step operating model to guide organizations in implementing and managing a modern data platform.

The research delved into the common challenges faced by organizations relying on traditional data platforms, such as data warehouses and data lakes. It provided detailed solutions to these challenges, elucidating how data mesh addresses inherent limitations and offers substantial benefits, including improved data governance, scalability, agility, and innovation.

Furthermore, the study offered key considerations and best practices for operationalizing a data mesh within banking organizations or similar enterprises. It provided a detailed roadmap for organizations to develop an effective operating model, ensuring seamless implementation and management of data mesh architectures.

Case studies from over seven banking organizations of varying sizes and geographies were analyzed to draw lessons from their experiences, including both successful and unsuccessful attempts to implement data mesh. These insights enriched the study, grounding the findings in practical, real-world scenarios.

The research methodology included in-depth interviews and surveys conducted with executives and data management professionals from diverse banking organizations. This empirical approach validated the findings and highlighted the lived experiences of practitioners navigating the complexities of modern data architectures.

Additionally, the study examined emerging trends and technologies poised to shape the future of data mesh and its applications. It also explored the implementation strategies of major cloud service providers, including Azure, AWS, GCP, OCI, and IBM, in deploying data mesh architectures. A generalized operating model was provided, enabling organizations to adapt the data mesh approach across cloud platforms, whether for small-scale or enterprise-wide deployments.

Finally, the research delivered strategic guidance for business executives, ensuring that the operationalization of data mesh aligns with organizational goals while navigating the complexities of modern enterprises. By synthesizing theoretical insights with practical applications, this study contributes to the advancement of data management practices and positions banking organizations for sustained success in the digital age.

9.2 Recommendations

Building upon the findings of this research, the following recommendations provide strategic and actionable guidance for banking organizations aiming to operationalize a modern data platform through a distributed data mesh architecture. These recommendations address both the technical and organizational facets of the implementation process and emphasize the critical considerations for achieving long-term success.

Adopt a Strategic, Iterative Approach to Data Mesh: The transition to a data mesh architecture should be regarded as an organizational journey rather than a discrete project. Organizations must adopt a strategic and phased approach, encompassing planning, architecting, designing, delivering, and scaling the data mesh across the enterprise. By framing data mesh implementation as a long-term evolution, banks can foster sustainable adoption and ensure alignment with their overarching business and digital strategies.

Leverage an Operating Model Framework: A robust operating model is indispensable for guiding the adoption and expansion of data mesh within an organization. While banks can develop their operating model from the ground up, this research recommends utilizing the 10-step operating model proposed in this study as a conceptual framework. This model offers a structured, systematic approach that can be customized to suit the organization’s unique needs and ensures scalability, consistency, and coherence in data mesh operationalization.

Establish a Data Mesh Enablement Team: A dedicated Data Mesh Enablement Team, also referred to as a Rollout Taskforce, is critical to driving successful implementation. This team should consist of a balanced mix of experts, including data architects, governance specialists, and domain representatives. Their role is to champion the data mesh initiative, coordinate efforts across divisions, address challenges, and ensure seamless execution and adoption across the organization.

Emphasize Minimum Viable Product (MVP) and Minimum Viable Experience (MVE): Launching the data mesh initiative with small-scale Minimum Viable Products (MVPs) is a key recommendation. MVPs should be designed to deliver tangible Minimum Viable Experiences (MVEs) that demonstrate value to users. This approach enables organizations to adopt an iterative development process, allowing for rapid experimentation, learning, and refinement. Early successes with MVPs can foster stakeholder buy-in and mitigate the risks associated with large-scale failures.

Invest in Robust Data Governance: Strong data governance is foundational to the success of any modern data platform. Organizations must develop and enforce comprehensive governance policies and standards that balance control with agility. These policies should be designed to serve the organization’s goals rather than hinder them, ensuring clear ownership, accountability, and consistency in data management practices.

Prioritize Data Quality Initiatives: High-quality data is critical to realizing the potential of a data mesh architecture. Organizations must implement data quality measures to ensure accuracy, consistency, and reliability of data across domains. Data quality frameworks should be integrated into the data platform lifecycle, enabling continuous monitoring and improvement.

Strengthen Data Security: Given the sensitive nature of customer and financial data, robust security measures are imperative. Banks should invest in advanced security technologies, including encryption, access controls, and anomaly detection systems. Best practices in data security, aligned with regulatory compliance requirements, should be embedded within the data platform's architecture and operational processes.

Automate Data Pipelines: The automation of data pipelines is crucial to reducing manual effort and enhancing efficiency. By leveraging modern automation tools, banks can streamline data ingestion, transformation, and delivery processes. Automation minimizes errors, accelerates time-to-insight, and ensures scalability in handling complex data workflows.

Optimize Data Lake and Data Warehouse Management: Effective design and management of data lakes and warehouses remain vital, even within a data mesh architecture. Organizations must carefully evaluate and implement these systems to complement the data mesh approach. Strategic integration of these components can ensure cost-effectiveness, scalability, and compatibility with diverse data formats and processing requirements.

Continuously Monitor and Optimize Performance: Maintaining optimal performance of the data platform requires regular monitoring and refinement. Organizations should establish performance metrics and implement monitoring tools to track system efficiency, data quality, and user satisfaction. Proactive performance optimization ensures that the platform continues to meet evolving business needs.

Provide Training and Support for Stakeholders: To fully realize the potential of a modern data platform, organizations must invest in training and support for their workforce. Comprehensive training programs should equip employees with the necessary skills to utilize the platform effectively. Continuous education and hands-on support will foster a culture of data-driven decision-making and empower users across all levels of the organization.

These recommendations collectively serve as a roadmap for banking organizations seeking to modernize their data architectures. By embracing a strategic and structured approach, leveraging the proposed operating model, and focusing on governance, quality, security, and training, organizations can successfully operationalize data mesh and unlock the full potential of their data assets.

9.3 Implications

The findings of this research present significant implications for banking organizations as they strive to modernize their data architectures and operational practices. By adopting a distributed data mesh and multi-cloud operating model, banks can address longstanding challenges and unlock new opportunities for growth, efficiency, and customer-centric innovation. These implications span across operational, strategic, and customer-focused dimensions, offering a comprehensive perspective on the transformative potential of the proposed approach.

Enhancement of Operational Efficiency: The adoption of a distributed data mesh framework introduces a decentralized and domain-oriented approach to data management, supported by automation and streamlined processes. By minimizing manual interventions, banks can significantly reduce the inefficiencies associated with traditional data platforms. The automation of data pipelines and governance workflows ensures faster data availability for analytics and decision-making, improving overall operational efficiency.

Improved Data Quality and Integrity: A key implication of this study lies in its emphasis on data quality as a foundational element of a modern data platform. The distributed nature of the data mesh, combined with robust governance mechanisms, enables organizations to maintain high standards of data accuracy, consistency, and completeness across all domains. This ensures that decision-makers can rely on data insights with confidence, driving informed and impactful strategies.

Cost Optimization Through Cloud Utilization: The multi-cloud operating model proposed in this research highlights the potential for cost efficiencies by leveraging cloud-native services and optimizing resource allocation. By adopting pay-as-you-go models and strategic data storage practices, banks can reduce unnecessary expenditures while maintaining scalability. This approach minimizes redundant storage and processing costs, ensuring a lean and cost-effective data ecosystem.

Increased Organizational Agility: A distributed data mesh fosters greater adaptability by decentralizing data ownership and enabling domain-specific teams to rapidly respond to dynamic business needs. This agility is further amplified by the use of cloud platforms, which provide the flexibility to scale resources up or down as required. The ability to deliver insights quickly and adapt to emerging trends empowers banks to remain competitive in a fast-paced industry landscape.

Strengthened Risk Management and Regulatory Compliance: The proposed operating model incorporates robust security measures and governance protocols that address regulatory requirements and mitigate risks associated with data breaches. By ensuring comprehensive audit trails, access controls, and data protection mechanisms, banks can enhance their ability to identify and manage risks proactively. This not only safeguards sensitive customer information but also reinforces trust and compliance with legal frameworks.

Enhanced Customer Experience Through Data-Driven Personalization: One of the most compelling implications of this study is the potential to elevate customer experience. By utilizing advanced analytics and real-time data processing, banks can gain deeper insights into customer behavior and preferences. This enables the delivery of personalized products and services, such as tailored financial advice, targeted marketing offers, and predictive financial solutions. Such customer-centric innovation fosters loyalty and strengthens the bank's competitive position in the market.

In summary, the adoption of a distributed data mesh and multi-cloud operating model positions banking organizations to achieve operational excellence, cost efficiency, and customer-focused innovation. The findings of this study underscore the transformative potential of this approach and provide a roadmap for banks to harness the power of modern data platforms in achieving sustainable growth and competitive advantage.

9.4 Recommendations for Future Research

Building upon the findings and insights of this study, several avenues for future research have been identified to advance the understanding and application of distributed data mesh and multi-cloud operating models. These recommendations are aimed at addressing gaps in the current literature, exploring emerging technologies, and enhancing the practical implementation of data mesh architectures in diverse organizational contexts.

Expanding the Scope of Research Samples: Future studies could broaden the scope by including a more diverse range of banking institutions. While this study primarily focused on large and mid-sized banks, incorporating smaller banks and financial institutions, particularly those operating in different geographic regions and regulatory environments, would provide a more comprehensive understanding of the applicability and effectiveness of data mesh architectures. Such research would also shed light on regional challenges, cultural nuances, and scalability considerations.

Integrating Artificial Intelligence and Machine Learning: The integration of AI and ML technologies into data mesh frameworks represents a promising area for future investigation. Research could explore how AI and ML can be effectively embedded into the proposed data mesh operating model to enhance its capabilities. For instance, AI-driven automation could simplify the management of data pipelines, improve real-time analytics, and optimize data lifecycle processes. Additionally, the potential of generative AI to enable more sophisticated data synthesis and predictive modeling warrants in-depth study.

Enhancing Data Quality and Consistency: Ensuring high data quality and consistency remains a critical challenge within distributed architectures. Future research could explore the use of AI and ML techniques to detect and rectify anomalies, standardize data formats, and maintain consistency across decentralized domains. These studies would provide insights into the design and deployment of AI-powered governance frameworks tailored to distributed data ecosystems.

Optimizing Data Pipelines and Reducing Latency: Another potential research area involves leveraging AI and ML to optimize data pipeline performance. Investigating techniques for minimizing latency, improving throughput, and ensuring efficient data ingestion and transformation processes could significantly enhance the operational efficiency of data mesh systems. Research in this domain could also evaluate the comparative benefits of various AI/ML tools in achieving pipeline optimization.

Strengthening Security and Risk Management: As data security and regulatory compliance remain paramount concerns for banking organizations, future studies could examine how AI and ML can bolster security measures within a data mesh environment. Potential areas of focus include the development of advanced anomaly detection algorithms, real-time threat analysis, and AI-driven encryption techniques. Such research would address critical gaps in safeguarding sensitive customer and organizational data.

Advancing Analytical and Decision-Making Capabilities: The role of AI, ML, and generative AI in enabling advanced analytics and decision-making within a distributed data mesh warrants further exploration. Studies could investigate how these technologies can provide deeper insights into customer behavior, enhance predictive modeling for business strategy, and support the development of next-generation analytics platforms. The application of generative AI in generating synthetic datasets for training and testing could also be a valuable area of focus.

By pursuing these research directions, future studies can significantly contribute to the ongoing evolution of distributed data mesh and multi-cloud operating models. Such advancements will enable organizations to stay at the forefront of technological innovation, address emerging challenges, and realize the full potential of modern data architectures in an increasingly dynamic and data-driven landscape.

9.5 Conclusion

This research has provided a comprehensive exploration of the transformative potential of distributed data mesh and multi-cloud operating models for banking organizations, addressing key gaps in both academic literature and practical application. The study delved into the challenges of traditional data management approaches, offering robust, innovative solutions that align with the complexities of modern banking environments. Through in-depth analysis, case studies, interviews, and theoretical synthesis, this research has laid a solid foundation for understanding, adopting, and operationalizing distributed data mesh and multi-cloud architectures.

The findings underscore the urgent need for banking institutions to transition from monolithic, centralized data management models to more agile, scalable, and domain-oriented frameworks. The distributed data mesh offers a paradigm shift, empowering data domain ownership, fostering cross-functional collaboration, and enabling seamless scalability across diverse and growing datasets. Coupled with the flexibility and cost efficiency of multi-cloud environments, this approach positions banks to address evolving business demands while maintaining robust governance, security, and compliance.

This study has articulated a clear and actionable ten-step operating model tailored to guide banking organizations through the data mesh implementation journey. It emphasizes the importance of strategic planning, a phased rollout, and ongoing optimization, ensuring that the transition is both smooth and impactful. The research also identified critical success factors, including the establishment of a dedicated data mesh enablement team, the prioritization of minimum viable products (MVPs) to deliver quick wins, and the integration of AI/ML capabilities to enhance data quality, pipeline optimization, and decision-making.

From an operational perspective, the findings highlight the immense potential of distributed data mesh to improve efficiency, reduce costs, and enable real-time data processing. Strategically, the architecture facilitates a shift towards a more customer-centric approach, leveraging insights to deliver personalized products and services, thereby driving customer satisfaction and loyalty. By enabling a unified and scalable data management system, banks can achieve superior analytics capabilities, harness the full value of their data assets, and unlock opportunities for innovation.

The implications of this research extend beyond the banking sector. The principles, methodologies, and frameworks developed here can serve as a blueprint for other industries grappling with similar challenges in managing complex and distributed data ecosystems. The research also paves the way for future exploration into emerging technologies, such as generative AI, and their integration with data mesh architectures to enhance predictive modeling, security, and automation.

In conclusion, this study marks a pivotal step in the evolution of data management strategies for banking organizations, demonstrating how distributed data mesh and multi-cloud architectures can address contemporary challenges and enable a competitive edge in a rapidly changing digital landscape. The research underscores that the journey to a modern data platform is not merely a technological transformation but a strategic imperative, requiring a shift in organizational culture, processes, and capabilities.

By embracing the insights, recommendations, and frameworks presented in this thesis, banking organizations can navigate the complexities of their data landscapes, achieve operational excellence, and drive meaningful innovation. Ultimately, this research offers a vision for a future where banks are not only data-driven but also agile, resilient, and customer-centric, positioning them to thrive in the dynamic world of digital finance.

APPENDIX A

SURVEY COVER LETTER

Dear [Participant Name],

I am writing to request your participation in a research study on the Operating Model for Distributed Data Mesh, which I am conducting as part of my doctoral studies. Your valuable insights and experiences will significantly contribute to this research.

As discussed previously, I have attached a structured survey that will take approximately 30 minutes to complete. Please feel free to provide detailed responses and reach out to me if you have any questions.

In addition to the survey, I would like to schedule a follow-up interview to discuss your responses in more detail. We can arrange a mutually convenient time for an in-person meeting or video conference.

Please note that all information provided will be kept confidential, and your participation is voluntary. Your responses will be used solely for research purposes and will not be shared with any third parties.

Thank you for your time and consideration.

Sincerely,

Aniket Mhala [Your Contact Information] [Date]

APPENDIX B

INFORMED CONSENT

I, [Participant Name], agree to participate in the research on Operating Model for Distributed Data Mesh conducted by Aniket Mhala as part of his Doctoral studies.

I have been informed of the confidentiality of information collected for this research.

I agree to participate in one or more interviews for this research. I agree to submit my responses to the primary questions via email. I also willingly agree to participate in the interview session for the secondary questions, which will be conducted either in-person or via video conference.

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 Participant]

Date:

APPENDIX C

INTERVIEW GUIDE

Participant Recruitment and Initial Contact

Identification and Contact: Potential participants identified based on their roles within their organizations (e.g., CDO, CTO, data architects, data engineers). They contacted via telephone or email to gauge their interest in participating in the research.

Explanation of Research: During the initial conversation, participants provided with a detailed explanation of the research objectives, process, and confidentiality aspects.

Participant Selection: Participants chosen based on their relevance to the research topic and their willingness to participate.

Primary Survey

Survey Distribution: A structured primary survey sent to participants via email.

Response Deadline: Participants have given a 15-day deadline to complete the survey.

Clarification: If any responses are unclear or require further clarification, follow-up questions may be sent to the participant.

Estimated Time: The primary survey is estimated to take approximately 30 minutes to complete.

Secondary Interview

Scheduling: Secondary interviews scheduled at a mutually convenient time, either in-person or via video conference.

Interview Duration: Interviews are estimated to last between 30 and 45 minutes.

Recording and Consent: Participants informed of the intention to record or take detailed notes during the interview. Their consent obtained prior to recording.

Interview Topics: The interview focused on in-depth discussions related to the research questions, allowing for more nuanced and detailed responses.

Post-Interview

Acknowledgment: Participants thanked for their time and contribution to the research.

Note: Throughout the interview process, confidentiality maintained, and participant information handled in accordance with ethical research guidelines.

APPENDIX D

SURVEY QUESTIONS

Research Questions for Corporate leaders:

Name of the Participant:

Role:

Title:

Organization:

Total work Experience:

Organizational Context and Data Strategy

Could you please provide a brief overview of your organization's size, industry, and data maturity?

What are your organization's primary data-driven initiatives or use cases?

How does data fit into your organization's overall strategy?

What are the key challenges you face in managing your data assets?

What are your organization's goals for implementing a distributed data mesh?

Distributed Data Mesh Implementation

What factors influenced your decision to adopt a distributed data mesh architecture?

What were the key challenges you faced during the implementation process?

How did you approach the definition and assignment of data domains?

What technologies or platforms did you choose for your data mesh implementation?

How did you ensure data governance and security within your distributed data mesh?

Operating Model and Governance

What is your organization's approach to data ownership and accountability?

How do you ensure data quality and consistency across different domains?

What data governance policies and procedures have you implemented?

How do you manage data access and security in a distributed environment?

What role does your central IT team play in managing the data mesh?

Roles and Responsibilities

What roles are involved in managing your distributed data mesh?

What are the key responsibilities of each role?

How do you ensure effective collaboration between different teams and departments?

What are the challenges in recruiting and retaining talent for data-related roles?

Benefits and Challenges

What are the primary benefits you have realized from implementing a distributed data mesh?

What challenges have you encountered and how have you addressed them?

How has the distributed data mesh impacted your organization's agility and innovation?

Have you experienced any performance or scalability issues with your data mesh implementation?

Future Plans

What are your future plans for expanding or enhancing your distributed data mesh?

What are the key challenges you anticipate facing in the future?

How do you see the role of data mesh evolving in your organization over the next few years?

REFERENCES

A. A. Munshi and Y. A. I. Mohamed, (2018), Data Lake Lambda Architecture for Smart Grids Big Data Analytics, IEEE Access, vol. 6, 40463-40471.

A. Abbasi, M. Amini, M. R. Fallahzadeh, & M. Pourkiani (2022). A Survey of Decentralized Data Management Architectures: Benefits, Challenges, and Research Opportunities. ACM Computing Surveys (CSUR), 55(2), 1-40. (Focuses on decentralized data management architectures, relevant to data mesh principles)

A. Bhattacharya, R. R. Selvaraj, & S. S. Iyengar (2022). Cloud-Native Data Mesh: Enabling a Scalable and Agile Data Architecture for Enterprises. 2022 IEEE International Conference on Big Data (Big Data), 114-123. (Focuses on data mesh implementation in a cloud-native environment)

A. Cuzzocrea, (2021), Big Data Lakes: Models, Frameworks, and Techniques, IEEE International Conference on Big Data and Smart Computing (BigComp), 1-4.

A. Gupta, S. K. Gupta, & R. K. Singh (2022). A Survey on Data Mesh Architecture: Benefits, Challenges, and Future Directions. 2022 International Conference on Intelligent Systems and Applications (ICISA), 1-6. (Provides a comprehensive survey of data mesh architecture)

A. K. Sandhu, (2022), Big data with cloud computing: Discussions and challenges, Big Data Mining and Analytics, vol. 5, 32-40.

A. M. Olawoyin, C. K. Leung and A. Cuzzocrea, (2021), Open Data Lake to Support Machine Learning on Arctic Big Data, IEEE International Conference on Big Data (Big Data), 5215-5224.

Amin Beheshti, Boualem Benatallah, Reza Nouri, Van Munin Chhieng, Huang Tao Xiong, Xu Zhao, (2017), A Data lake Service, Conference on Information and Knowledge Management, 2451–2454.

C. D. Castillo-Effer, A. Rudra, & M. Tuosto (2021). Decentralized Data Governance for Data Mesh Architecture. 2021 IEEE International Conference on Big Data (Big Data), 103-112. (Explores data governance in a data mesh environment)

C. Giebler, C. Gröger, E. Hoos, H. Schwarz and B. Mitschang, (2020), A Zone Reference Model for Enterprise-Grade Data Lake Management, IEEE 24th International Enterprise Distributed Object Computing Conference (EDOC). 57-66.

C. Liu, (2020), Research on the Method of Constructing Distributed Data Lake Driven by Virtualization Model, International Conference on Computers, Information Processing and Advanced Education (CIPAE), 223-226.

D. Del Piero, A. D. Sandro, & G. Grilli (2022). A Multi-Cloud Architecture for Data Mesh Implementation. 2022 17th International Conference on Advanced Robotics (ICAR), 1-6. (Discusses multi-cloud architecture for data mesh implementation)

E. Zagan and M. Danubianu, (2021). Cloud Data Lake: The new trend of data storage, 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2021,1-4.

El Aissi, M. et al. (2020). Data Lake Versus Data Warehouse Architecture: A Comparative Study. WITS 2020, Vol. 745, 201-210.

F. Beyer, M. Färber, & P. Franca (2022). Data Mesh for Developers: Building Blocks for a Decentralized Data Architecture. O'Reilly Media, Inc. (Technical guide to data mesh development)

F. Nargesian, K. Q. Pu, B. Ghadiri Bashardoost, E. Zhu and R. J. Miller, (2021). Data Lake Organization, in IEEE Transactions on Knowledge and Data Engineering

J. García-Fanjul, J. Luna-Rodríguez, & J. M. Saiz-Gómez (2022). Data Mesh vs. Data Lake: A Comparative Analysis for Financial Institutions. 2022 IEEE International Conference on Big Data (Big Data), 209-218. (Compares data mesh and data lake architectures in the context of financial institutions)

J. Liu, Y. Zhang, & H. Wang (2021). Data Mesh Architecture: A Survey. 2021 IEEE International Conference on Big Data (Big Data), 124-133. (Another survey of data mesh architecture)

J. Patel, (2019), An Effective and Scalable Data Modeling for Enterprise Big Data Platform, IEEE International Conference on Big Data (Big Data), 2691-2697.

J. Yeung, S. Wong, A. Tam and J. So, (2019), Integrating Machine Learning Technology to Data Analytics for E-Commerce on Cloud, Third World Conference on Smart Trends in Systems Security and Sustainability (WorldS4), 105-109.

Kachaoui, J., Belangour, A. (2022). Improving a New Data Lake Architecture Design Based on Data Ponds and Multi-Agent Paradigms. Innovations in Smart Cities Applications Vol. 5. 815-821.

M. A. Munoz, J. M. Gonzalez-Sanchez, & J. C. Gonzalez-Sanchez (2022). Data Mesh and Data Lake: A Comparative Analysis for Big Data Management. 2022 IEEE International Conference on Big Data (Big Data), 190-199. (Compares data mesh and data lake architectures)

M. Amr, A. E. Hassan, & A. Elnabarawy (2021). Towards a Secure and Efficient Data Governance Framework for Data Mesh Architecture. 2021 International Conference on Computer, Information, and Telecommunication Systems (CITS), 1-6. (Explores data governance in the context of data mesh)

M. Färber, P. Franca, & J. P. Muller (2021). Data Mesh Fundamentals: A Practical Guide for Building a Distributed Data Mesh. Manning Publications Co. (Practical guide to implementing a data mesh architecture)

M. Jensen, C. S. Jensen, & R. Ting (2020). A Comparative Analysis of Data Lake Architectures. Communications of the ACM, 63(8), 66-73. (Analyzes different data lake architectures, relevant for understanding data mesh foundations)

N. Saranya, R. Brindha, N. Aishwariya, R. Kokila, P. Matheswaran and P. Poongavi, (2021), Data Migration using ETL Workflow, 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 1661-1664.

Naqvi, M. A. Khan, & M. A. Khan (2022). A Survey on Data Mesh Architecture: Benefits, Challenges, and Future Trends. 2022 International Conference on Emerging Trends in Computing, Communications, and Networking (ETCCN), 1-6. (Yet another survey of data mesh architecture)

Natalia Miloslavskaya, Alexander Tolstoy. (2016),Big Data, Fast Data and Data Lake Concepts, Procedia Computer Science, Vol 88,300-305.

P. Carreira, A. P. Rocha, A. L. Oliveira, & J. P. Albuquerque (2021). Embracing Data Mesh for Continuous Delivery of Data Products: A Case Study in the Banking Industry. 2021 47th Annual IEEE Conference on Software Engineering (ICSE), 1307-1318. (Case study of data mesh adoption in a bank)

P. Franca, M. Färber, & J. P. Muller (2020). Data Mesh: A Decompositional Approach to Data Management.

P. Gölz, N. Ritter, & C. Fehling (2021). Data Mesh: A Paradigm Shift for Data Management. Business & Information Systems Engineering, 13(2), 185-198. (Discusses data mesh as a paradigm shift in data management)

P. Heyton (2022). Data Mesh vs. Data Warehouse: A Decision Framework. Morrissey Associates Inc. (Provides a framework for choosing between data mesh and data warehouse)

Panwar, A., & Bhatnagar, V. (2020). Data Lake Architecture: A New Repository for Data Engineer. International Journal of Organizational and Collective Intelligence (IJOCI), 10(1), 63-75.

S. Gokulakrishnan and J. M. Gnanasekar, (2020), Data Integrity and Recovery Management in Cloud Systems, Fourth International Conference on Inventive Systems and Control (ICISC), 645-648.

S. Park, B. Cha, Y. Cha, J. Mo, S. Oh and J. Kim, (2018), Design of Connected Data Lake System based on Micro Cloud Storage, International Conference on Information and Communication Technology Convergence (ICTC), 917-918.

S. Skhiri and C. Duverne, (2020), Data Architecture: A Sustainable Foundation for Data Exploitation, IEEE Potentials, vol. 39, 15-21.

T. Priebe, S. Neumaier and S. Markus, (2021) , Finding Your Way Through the Jungle of Big Data Architectures, IEEE International Conference on Big Data (Big Data), 5994-5996.

T. Z. Emara and J. Z. Huang, (2020), Distributed Data Strategies to Support Large-Scale Data Analysis Across Geo-Distributed Data Centers, IEEE Access, vol. 8, 178526-178538.

U. Aftab and G. F. Siddiqui, (2018), Big Data Augmentation with Data Warehouse: A Survey, IEEE International Conference on Big Data (Big Data), 2775-2784.

Y. Cho, J. Hong, & S. Moon (2022). A Data Mesh-based Data Governance Framework for Enhanced Data Quality in Banking. 2022 International Conference on Information Networking (ICOIN), 23-28. (Examines data governance and data quality in a data mesh context for banking)

Y. -H. Chen, H. -H. Chen and P. -C. Huang, (2018), Enhancing the data privacy for public data lakes, IEEE International Conference on Applied System Invention (ICASI), 1065-1068.

Yan Zhao, Imen Megdiche, Franck Ravat, Vincent-nam Dang, (2021), A Zone-Based Data Lake Architecture for IoT, Small and Big Data, 25th International Database Engineering & Applications Symposium, 94–102.

Zhamak Dehghani, (2019), How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh , www.martinfowler.com, 1-5