Back to summary page

Data virtualization strategy design

Client is a U.S.-based company that provides variety of experiences related to boating, dining, and transportation, particularly in coastal areas and major cities in US.

- Diagnosed the existing data infrastructure, uncovering integration gaps and inefficiencies due to data silos across various acquired entities.
- Recommended a Data Mesh architecture leveraging Data Virtualization to provide accessible, reliable, and real-time data integration without the need for a centralized data warehouse.
- Developed a decentralized data strategy that empowers business units to own and manage their data, enhancing data consistency and decision-making.

Tourism and Marine transport company needs to hedge its data architecture

DATA MANAGEMENT

Picture this…

You're looking for an optimal architecture for data and reporting to be centrally available on a real-time basis. But with data scattered across various acquired companies and systems—like Microsoft Dynamics GP, NPS, UKG, and Anchor—you find yourself facing the challenge of building a centralized data warehouse. The complexity and inconsistency due to differing business definitions across your business units lead to fragmented and delayed reporting.

You turn to Accordion.

We partner with your team to address the complexities of integrating diverse data sources across your organization. Rather than pushing towards a traditional centralized data warehouse with intricate ELT processes, we propose a Data Virtualization solution aligned with a Data Mesh architecture. This approach avoids the pitfalls of centralized data warehousing, offering a more agile, scalable, and cost-effective data strategy tailored to your unique needs.

1) We assess your current data sources and infrastructure, identifying gaps and challenges in integrating data across your organization's diverse systems.
2) Working closely with your business teams, we design a Data Mesh architecture, emphasizing decentralized data ownership and governance empowering each business unit to take responsibility for its data ensures that the data is accurate, reliable, and meaningful within its specific context.
3) The design allows data from different systems to remain where it is while still being accessible and integrated in real-time through virtualization, ensuring data consistency and eliminating the need for heavy data movement and transformations.

Your value is enhanced.

With our solution, your organization now benefits from a modern, flexible data architecture. The Data Mesh and virtualization approach provide a single source of truth, real-time insights, and improved decision-making capabilities, aligning your data strategy with your business goals. This setup has drastically cut down on the overhead of managing centralized data and allowed your teams to leverage data as an asset without the constraints of traditional data management approaches.

KEY RESULT

- 100% automation of data integration
- Reduced infrastructure costs by 60% through the adoption of Data Virtualization and Data Mesh architecture

VALUE LEVERS PULLED

- Data Virtualization
- Data Mesh
- Data Governance

Approach - Diagnostic phase

EXPLORATION

- Identified **data sources, servers** and other **third-party systems**
- Accessed client's data systems to **deep-dive and explore** the data directly
- Developed an understanding of **existing infrastructure and data flows**
- **Conducted technology stakeholder interviews** to understand current-state data sources
- Connected with functional heads to **understand the desired future state of reporting**

ASSESSMENT

- Defined **existing process flows and workflows** for every data system
- **Identified current process and data gaps** by assessing existing metrics and data quality across functional areas including finance, sales & marketing, operations, etc.
- **Assessed data governance processes** to document data capture methodology
- Reviewed existing reports to identify **automation and enhancement opportunities**

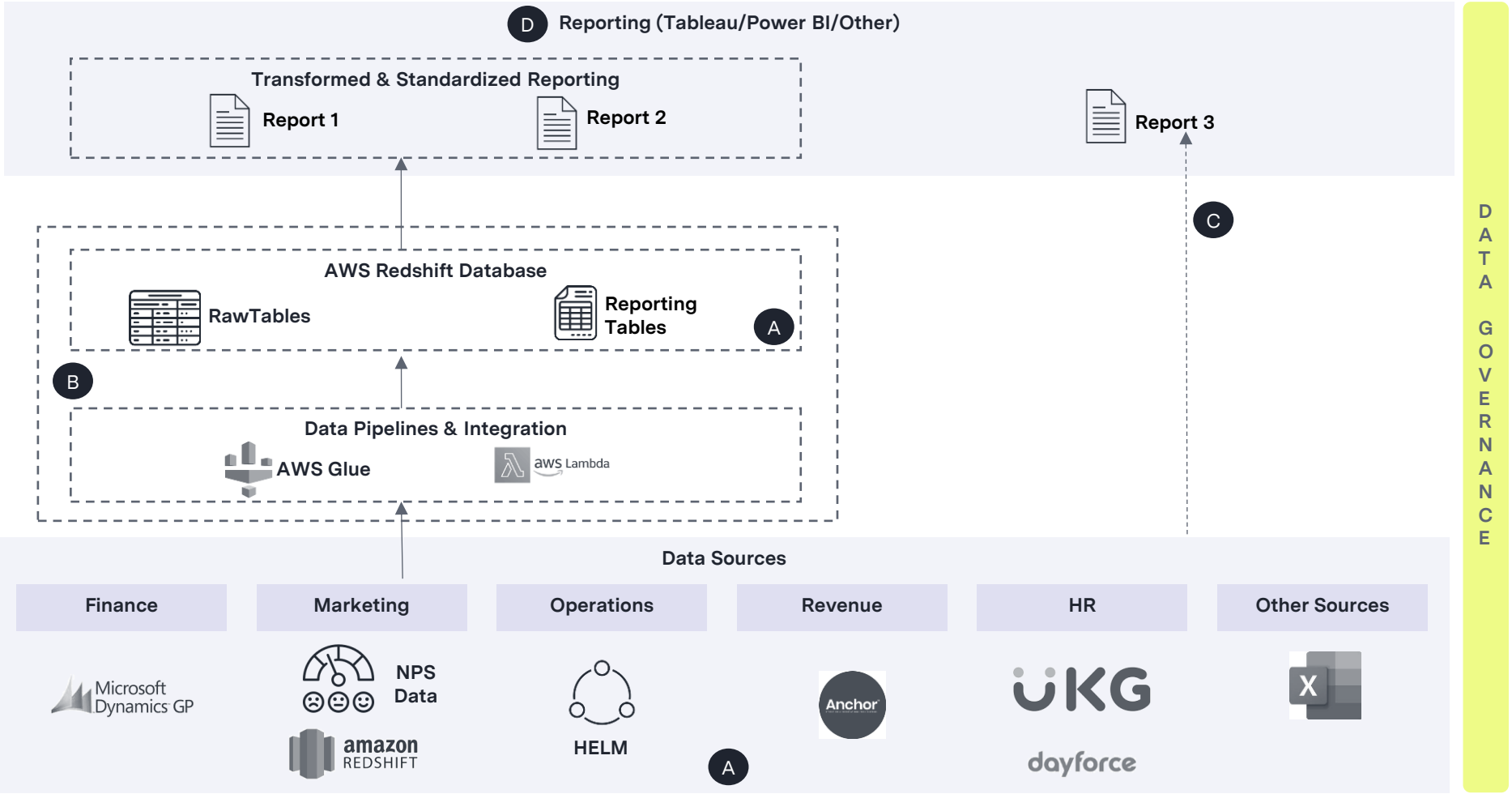
DESIGN

- Designed a **Data Virtualization** strategy using **Data Mesh to replace centralized ELT and traditional data warehousing.**
- Developed a data access framework that empowers **business units to manage their data as a product, enhancing governance.**
- Created an adaptable environment with **Data Virtualization tools for seamless, real-time data integration without centralization.**
- Established a **decentralized, domain-oriented reporting suite** that aligns with the specific needs of each functional area, including finance, operations, sales & marketing, and more.

DOCUMENTATION & RECOMMENDATION

- Created high-level and low-level architecture diagrams showcasing the **Data Mesh and Virtualization framework.**
- Created **data flow and sample workflows** for all source systems for ease of reference
- **Provided recommendations on data framework** to establish a single source of truth and enable actionable reporting
- Advised on **enhancing data governance models to ensure data quality, security, and compliance** while enabling agile decision-making.

Current state data architecture – Data Mesh



Key observations

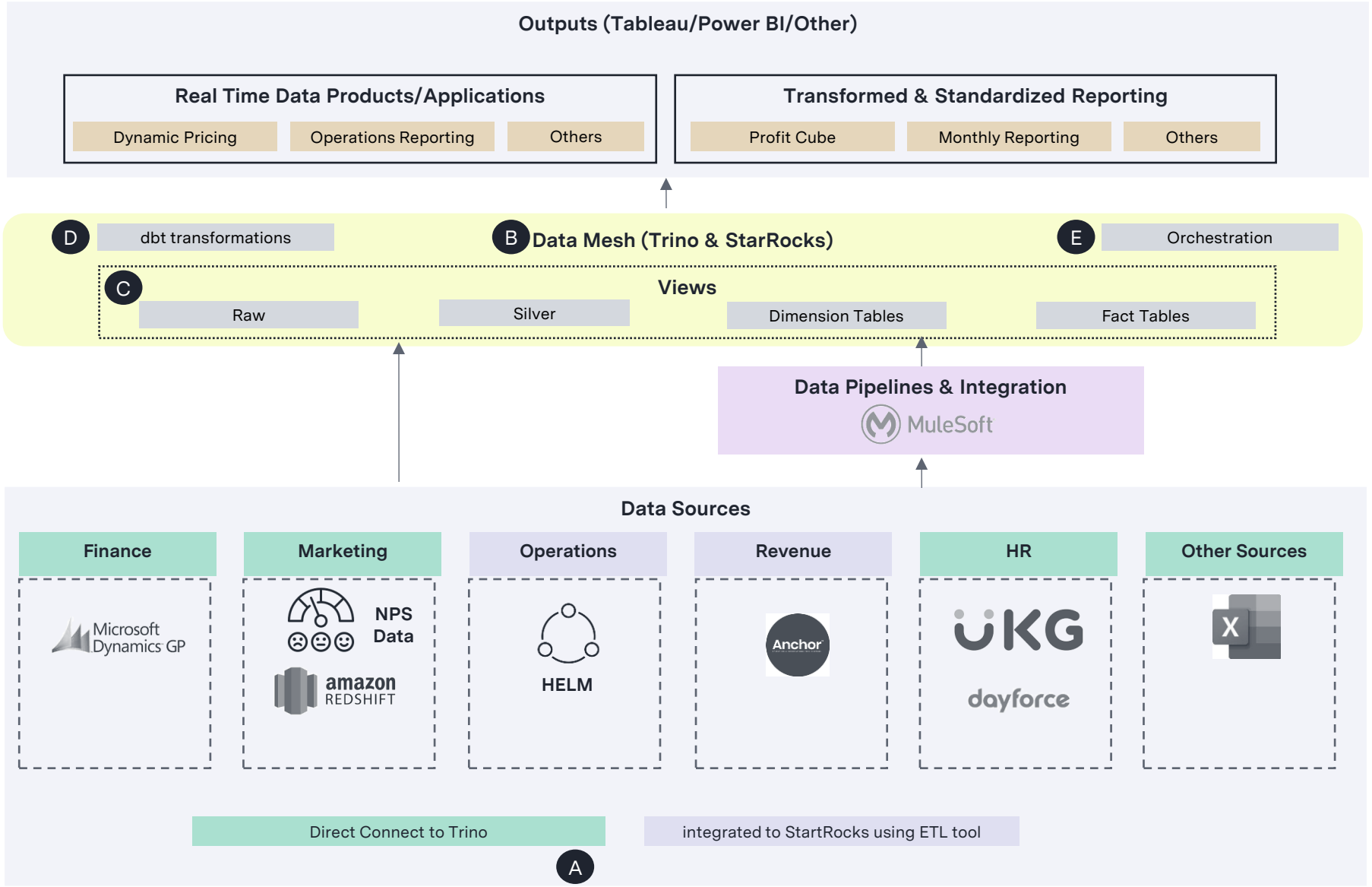
A. Data is spread across multiple sources (Redshift and other data sources), leading to challenges in data governance, integration, and consistent reporting.

B. The Marketing team has developed an isolated data infrastructure on AWS, creating silos and limiting data sharing and unified analysis across the organization.

C. The absence of a single source of truth has led different business units to generate their own reports directly from raw data, resulting in inconsistent KPIs, varied business rules, and fragmented reporting standards.

D. There is a need to establish clear ownership and responsibilities for managing the reporting suite and ensuring consistent data governance across all business units.

Future state data architecture – Data Mesh



F DATA GOVERNANCE

Recommendations

- **A** Use Trino to directly connect with databases and sources via JDBC connectors, replacing traditional ELT processes and reducing the need for intermediate storage
- **B** Leverage StartRocks as a Lakehouse to support the Data Mesh architecture, providing a scalable and centralized platform for managing data across various domains
- **C** Organize data into virtual views within the Data Mesh, including layers such as Raw, Silver, and Reporting, to enhance data quality and consistency without physically transforming or moving the data.
- **D** Utilize dbt for data transformation and Git for version control, to establish a robust framework for developing, managing and scaling data transformation workflows in data mesh
- **E** Leverage an orchestration tool to manage end to end data workflows within data mesh.
- **F** Establish a data governance committee consisting of a Governance Lead, Domain Data Stewards, and implementation teams to oversee and guide the Data Mesh strategy, ensuring adherence to governance standards and optimizing data value across the organization