Assignment objective:

The document provides a concise approach to analyzing and understanding the current state of data in a financial institution. Its objective is to identify the target data maturity stage and formulate a strategy and execution plan.

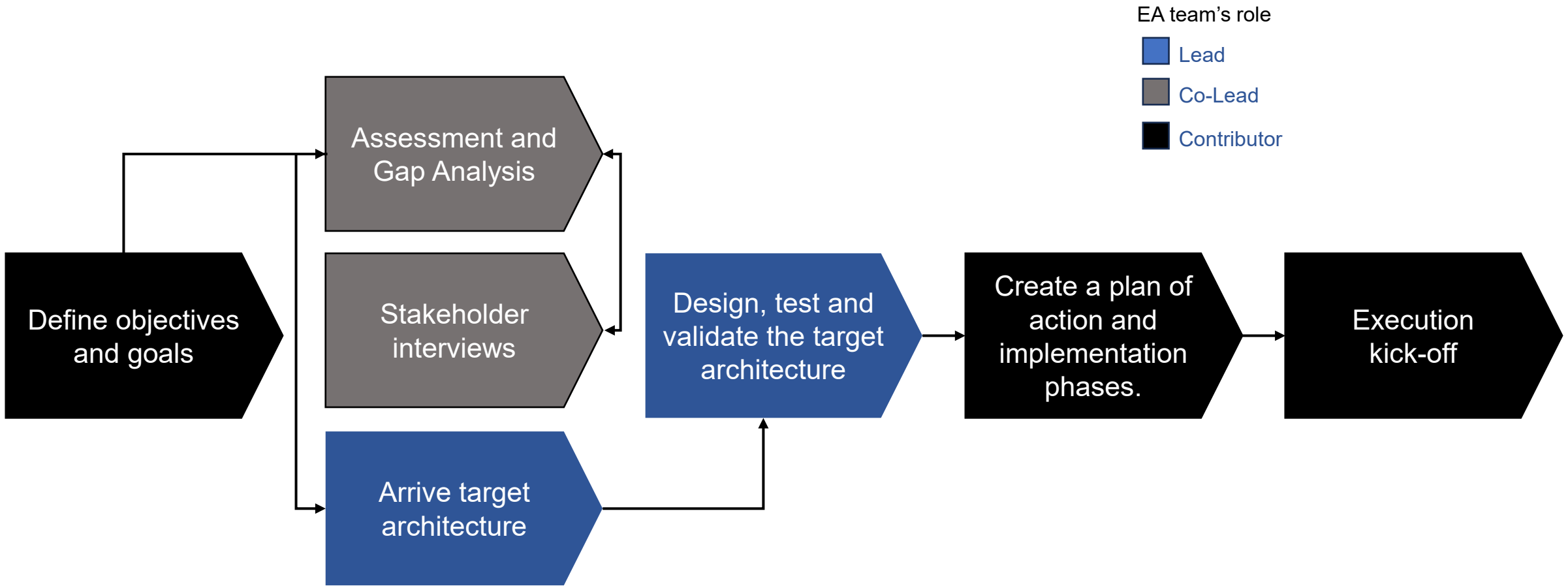
Key discussion points

1. What are the key goals from a business and data engineering perspective?

2. What is the plan for initiating the first step of the transformation? What are the key use cases?

3. How can we effectively govern the program to ensure enterprise-wide impact and seamless implementation?

Approach: To initiate the transformation process, we will follow a five-step approach to tackle the main challenges and establish the desired outcome.



NOTE: To ensure the successful execution of this approach, it is important to clarify and agree upon the roles and responsibilities. Ideally, a program manager, reporting to a data board, who will work closely with business, data, and tech teams, will streamline the process.

Objective: Our goal is to create a future-ready data infrastructure that empowers the bank to leverage data as a strategic asset for innovation and competitive advantage.

The growing importance of making strategic decisions based on data is crucial for various reasons.

Powerful customer insights for a personalized experience	• Customer segmentation • Personalization • Digital-first journeys
Product development and faster market entry and capitalization	• Product feature scoping • Pricing optimization • Product performance analysis
Risk, security and compliance mitigation and remediation	• Regulatory reporting • Data Compliance • Risk & Compliance analytics
Operational efficiency, and increased productivity.	• Operational reporting • Integrate to RPA and Process streamlining • Marketing & Campaigns
Competitive advantage and strategic market expansion	• Competitor analysis • Market entry strategies

.. which requires a robust technology infrastructure that enables key engineering goals listed below.

1. Achieve **faster data integration** and data warehousing by integrating them into a sustainable data lake house.
2. Robust database architecture is needed to **scale and manage** increasing data and analytic needs of business
3. **Real-time data processing** and streaming capabilities to enable instant insights and actions.
4. **Data security and compliance** is organically plugged into every engineering process.
5. Balanced approach **to leverage cloud** and hybrid technologies to maximize technology ROI.
6. Embed machine learning & AI capabilities within the architecture to enable **self-serve analytics** for business.
7. **Faster deployment** and technology operating model to achieve quick idea-to-market
8. Friendlier **data governance** mechanisms to maintain high data quality, access controls and metadata mgt..

Plan of action: Within the first 180 days, our goal is to conduct a thorough discovery, identify key gaps, and begin executing the initial steps to kickstart the transformation process.

Discovery

10 weeks

Interview key stakeholders to understand the business needs, data gaps, and priorities for the upcoming years.

Collaborate and co-create a target data architecture, conduct gap analysis based on interviews, and arrive key action items

Establish governance forums, guiding principles for engineering teams, and define key tech objectives

Document the artifacts

Design

12 weeks

Design technology landscape (on-prem v/s cloud v/s hybrid) and define key solution components

Conduct impact analysis on transitioning from AS-IS to TARGET

Validate key technology design assumptions through a rapid R&D exercise to test and validate target view

Arrive a transition plan, considering the new and existing reusable components

Arrive an investment plan – Budget, plan of action, and execution approach

Institutionalize

2+ weeks

Share the final design with the Data Board, seek their approval, and identify potential investments.

In the second round of socialization, meet with key business heads to ensure that they are aligned with the approach.

Arrive a subsequent 180 days plan of execution for transitioning to the new data architecture with minimal business impact

NOTE

1. To ensure that we are working towards a collective objective, we need to establish an operating model with the data team.
2. Seek an initial check-in with the potential data sponsor (CXO) to obtain feedback for this plan of action.

Use cases: While following the 180-day plan, we must identify quick wins and demonstrate immediate value. There are three different domains for use cases.

This is a list of ideas for consideration. The forum may choose 2-3 ideas for immediate action.

Business

One-obligor customer to understand the corporate customer relationship cashflow and optimization methods

Analytics to understand **customer behaviors**, channel interactions, and personalize offers.

Conduct near-real-time **fraud analysis** on a chosen product category and its transactions

Reporting and analytical uses case cases pertain to operation efficiency such as TAT monitoring, resource forecasting etc.

Data-Engineering

Conduct a rapid exercise to define key **data quality gaps** and potential action to remediate

Choose a high-impact business use case that has the potential to utilize both structured and non-structured data. Create a **data pipeline to demonstrate the value of modern data architecture** principles.

Create a minor exercise to **auto-evolve schemas** to accommodate changing data structures seamlessly.

Tech Infrastructure

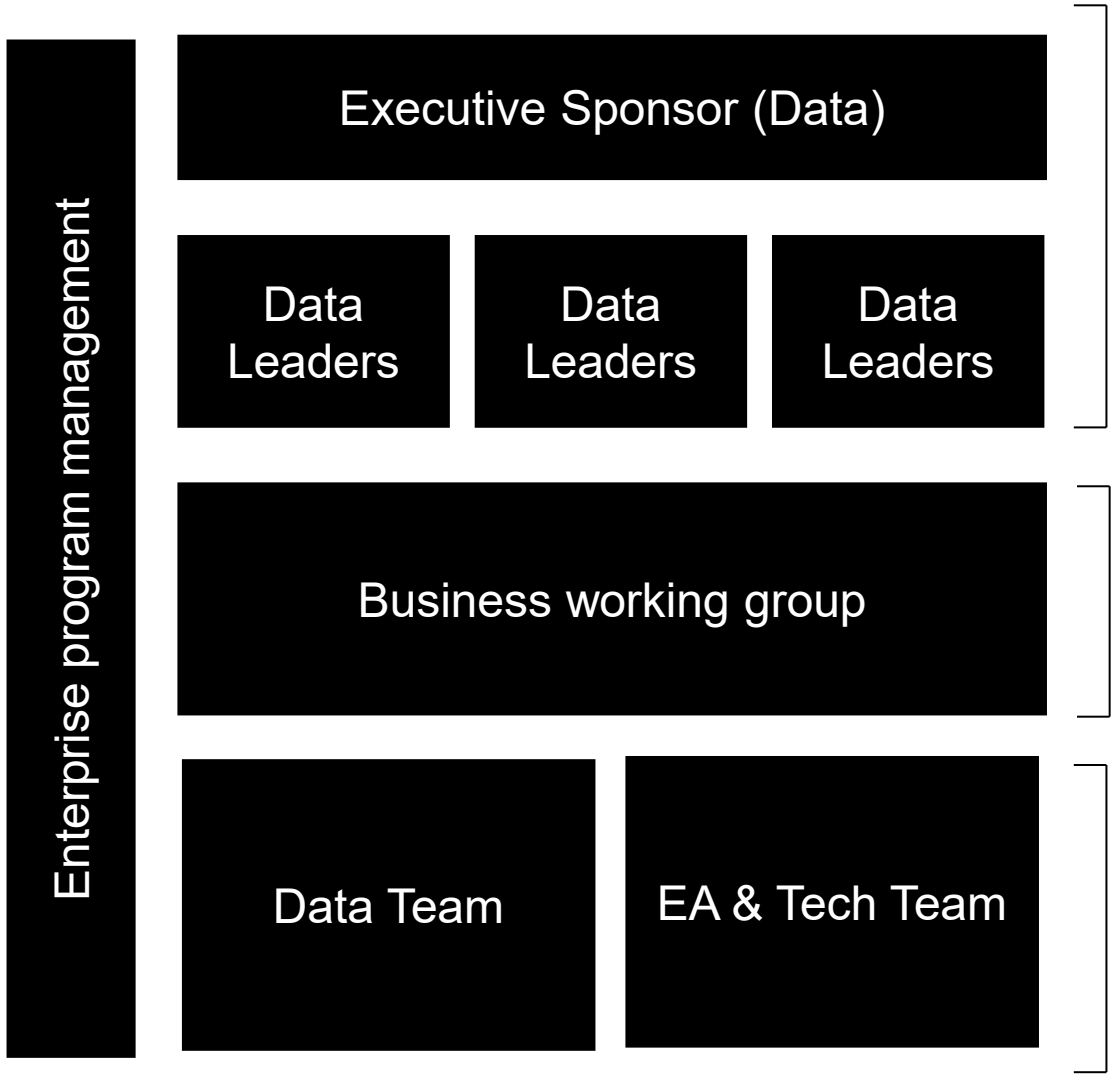
Identify a manual data ingestion area and automate the **data ingestion and transformation** pipeline to speed up data delivery.

Evaluate the current **data formats** and test new formats with the intention of standardizing data formats.

Analyze the storage and implement data **compression techniques** to reduce storage costs and enhance data retrieval speeds.

NOTE: The points listed above are a few ideas that I understood during our discussion. During the earlier phase, this list can be refined with the help of business, data, and technology teams

Governance: To ensure strategic and continuous delivery, it is essential to establish organized and continuous organization governance.



Data Board for strategic advisory and governance

- An executive sponsor, along with data leaders, defines key direction and governance of the data program
- Data Leaders are business heads of each segments drives the business priorities of their respective business units in conjunction with their counterparts in business working group

Subject matter experts to drive use cases and prioritization

- Senior subject matter experts from respective business units will drive high-impact use cases
- Definition of requirements, testing, and validation methods

Design & Execution Experts

- Head of Data leads the engineering execution and building use cases based on business demand
- Enterprise Architecture team leads the technology, security, and infrastructure engineering roadmaps that will empower the ecosystem.

Gap Analysis: A comprehensive framework for assessing the maturity of technology in the data and analytics space.

Integration and Storage

- Scan all data sources
- Identify and categorize different data types (Struct vs Non-Struct)
- Analyze the volumes of data, data growth rate etc.
- Assess the storage infrastructure and costing policies
- Assess big data storage infrastructure and retention policies

Data Modelling

- Conduct an analysis of different data models in place and identify the compatibility of each with current and future business needs
- Data availability and time-to-transition SLAs
- Conduct performance analysis – Normalization, consistency, completeness etc.

Data Access & Querying

- Identify data access processes by tech and business users
- Analyze the tools and tech used by engineering team for data preparation, transformation and analysis
- Identify real-time and data streaming capabilities

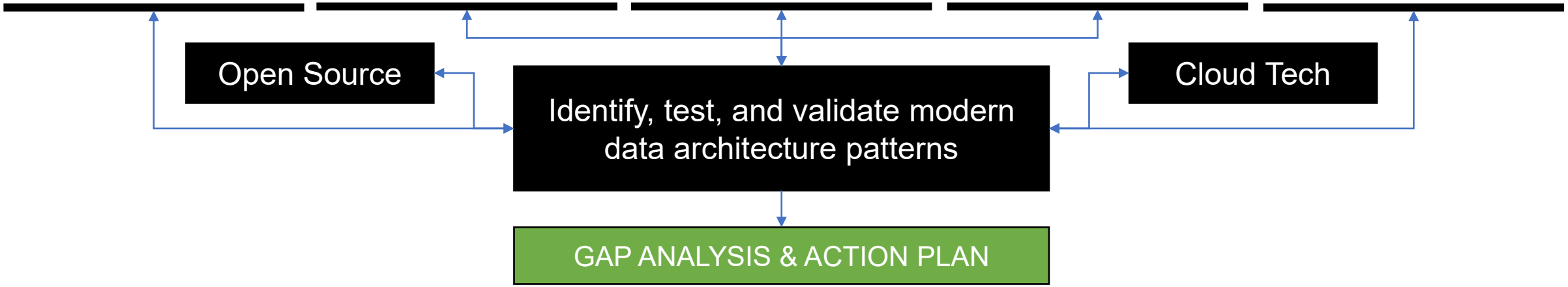
Scalability & Performance

Conduct a thorough analysis on key data performance KPIs such as

- Latency
- Throughput
- Query response rate
- Resource utilization
- Indexing & Concurrency
- Backup & recovery
- Caching and retention
- Data availability

Security, Privacy & Compliance

- GDPR or similar compliance compatibility
- Data quality
- Identify the access controls and operating model for business and engineering teams
- Assess the data auditing process.



Thank you