



Design Architecture &
Technology Acceleration

ENVIRONMENT, SOCIAL & GOVERNANCE

Synechron
Digital / Business Consulting / Technology

FinLabs

Table of Content

No.	CONTENT
1	EXECUTIVE SUMMARY
2	HIGHLIGHTS - ESG SOLUTION ARCHITECTURE
3	HIGH LEVEL ARCHITECTURE
4	ACTIVITY DIAGRAM
5	DOMAIN DRIVEN APPROACH
6	COMMANDS/QUERIES DESIGN PATTERN
7	CONCEPTUAL VIEW
8	MICROSERVICE ARCHITECTURE
9	CHANGE DATA CAPTURE (CDC)
10	DATA FABRIC ARCHITECTURE
11	DATA INGESTION FRAMEWORK
12	DEPLOYMENT DESIGN
13	TECHNOLOGY CHOICES

Accelerator briefing

Why we are building an Environment, Social & Governance accelerator

- Institutional firms have started to embrace ESG, but are struggling with moving targets and lack of standards
- No uniformity and aligned definitions in terms of ESG data from the legislator
- Lots of data sources available but how to make sense of this data?
- No mainstream platform which provides the functionality of showing the ESG score at portfolio level and/or benchmarking of ESG data



Executive Summary

We can help you to define, design and execute technology implementation that drives business growth

Objective	Bank intends to choose best portfolio based on ESG scores of instruments. Need to take off manual efforts and suggest best portfolio for investment to portfolio manager.
-----------	---

Key Challenges	Our Solution Approach
<ul style="list-style-type: none">Lack of Standardization in ESG ratings as too many agencies available with Different ESG rating of their own.Lack of compliance or regulation in ESG space .Difficulty in choosing between different instrument options for portfolio.Test Data models easily.High turnaround time for change management and fixes	<ul style="list-style-type: none">Develop & Perform in-house Normalization Method by fetching ESG score from multiple agencies. Also develop an in-house solution in data science for recommendation with Companies having better ESG ratings.Build a platform to identify correlation between different ESG standard of ratings by developing data science and data engineering component.Solution will give an alternative monitoring data which will help the asset manager to decide which are the best instruments/investments to add into his portfolio.DevOps will help data scientist to change model, deploy this model easily so, as to test it in multiple environments.Greater flexibility by leveraging best fit technology on Microservices based architecture.

Key Drivers	Reduce time to market by implementing best DevOps Practices.	Corelate between different ESG agencies with Data science components.	Leverage different data engineering methodologies.	Increase reusability and flexibility by following Microservices architecture.

What features this accelerator offer?

Portfolio Analysis



- Feature:**
Portfolio Analysis against different ESG metrics.
Benchmarking using different ESG scores and indexes

Recommendation & Simulation



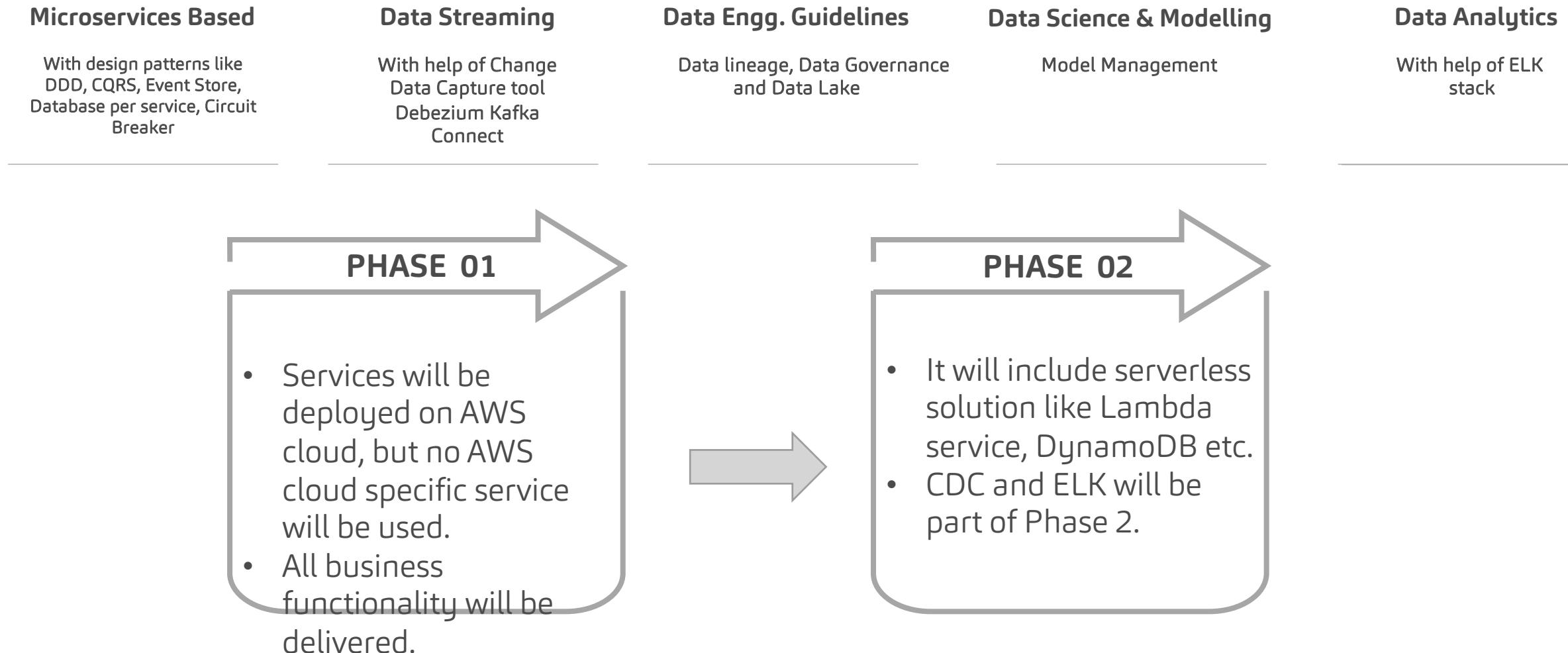
- Feature:**
Supports the endeavour to improve the portfolio ESG score by proposing an “alternative portfolio” with minimal changes and also taking into account client preferences

Investment Action and Reporting

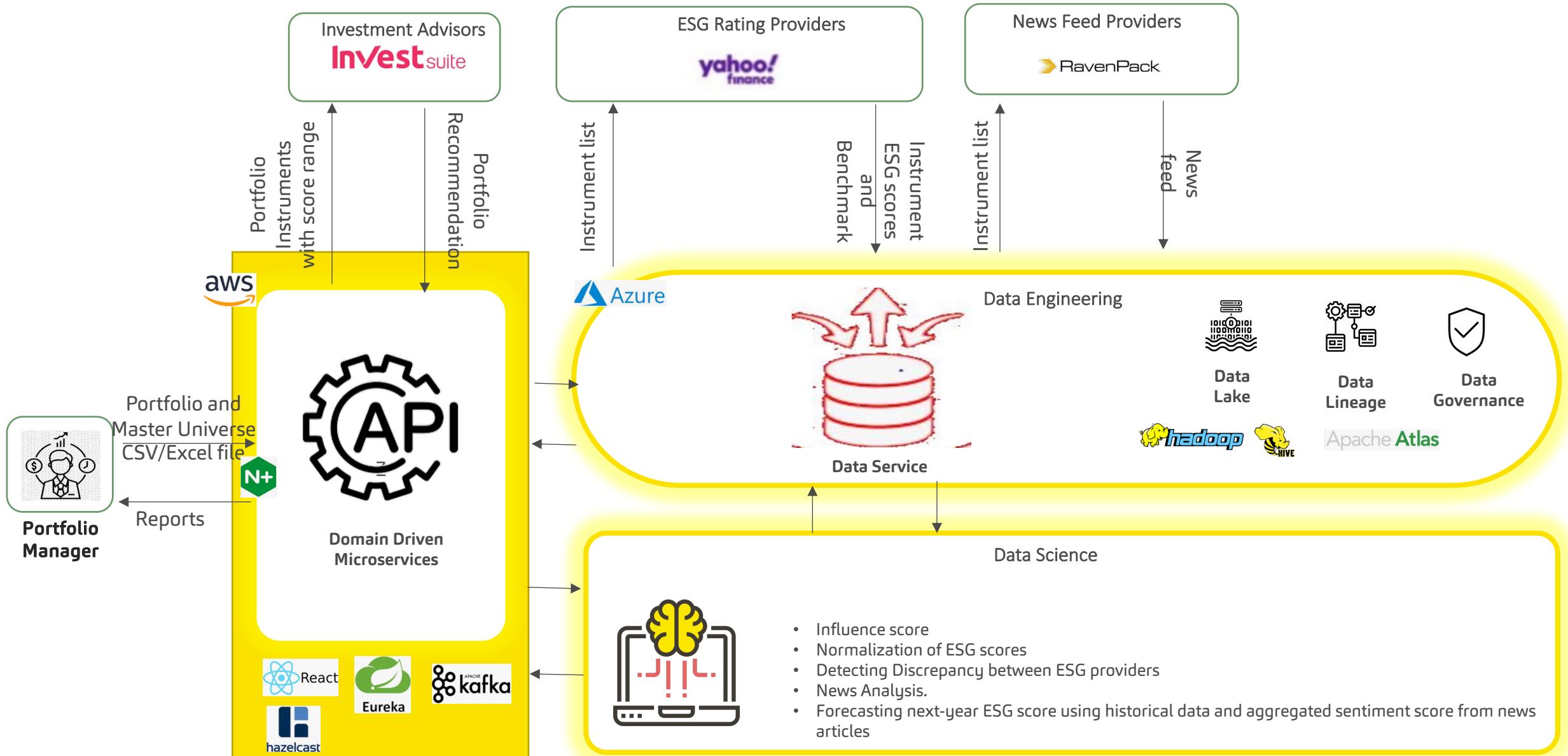


- Feature:**
Easily extract reporting overviews of the recommendation/investment actions for further discussion or approval

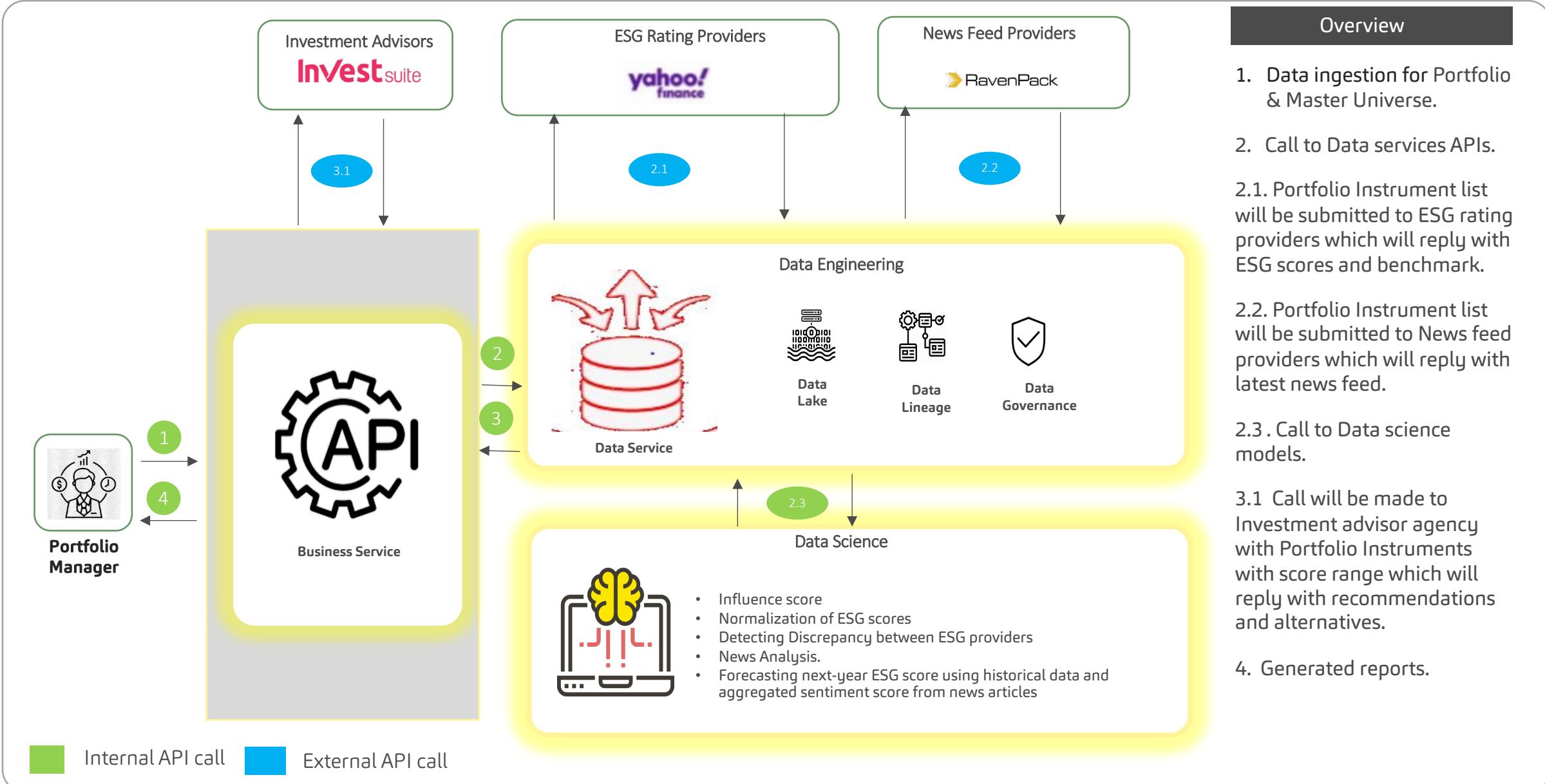
Highlights - ESG Solution Architecture



High level Architecture

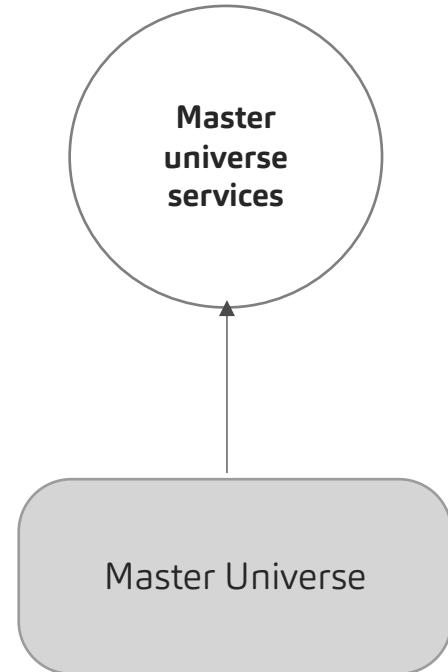
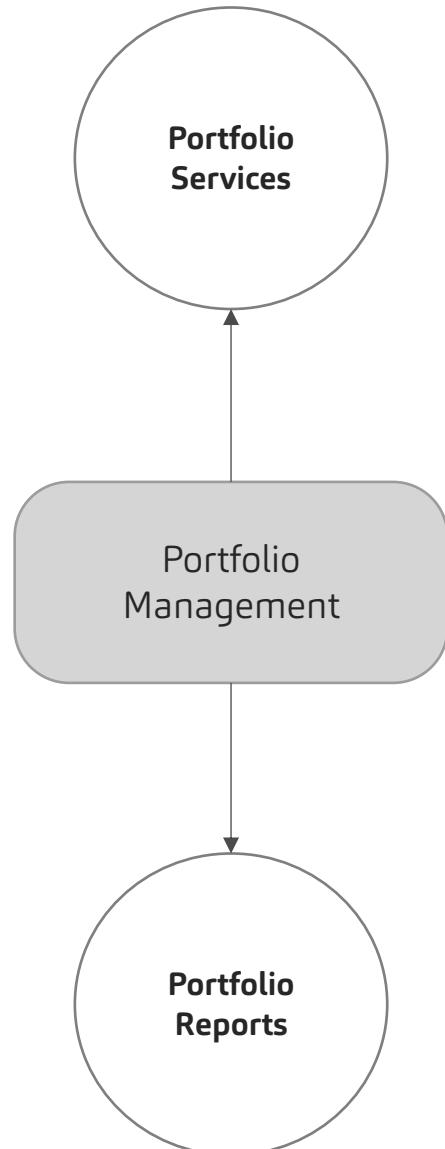


Activity Diagram



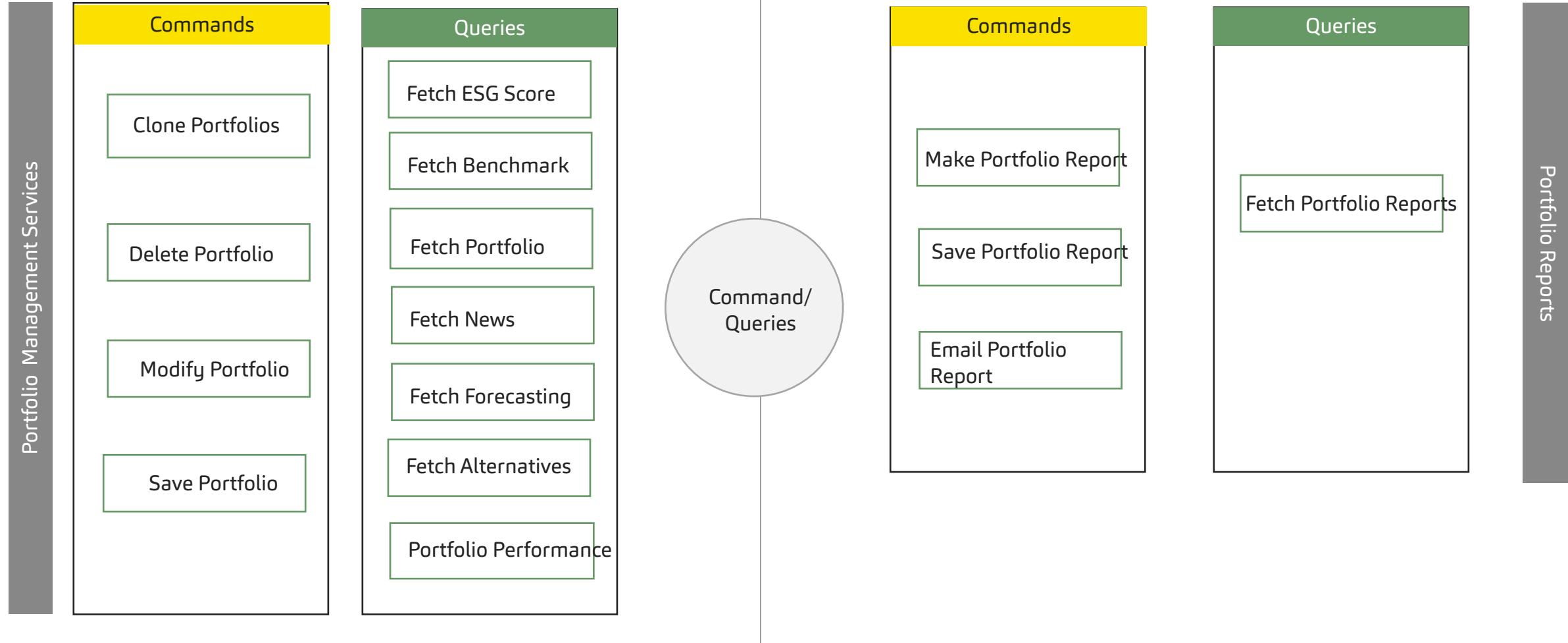
Domain Driven Approach

Sub Domains & Bounded Context



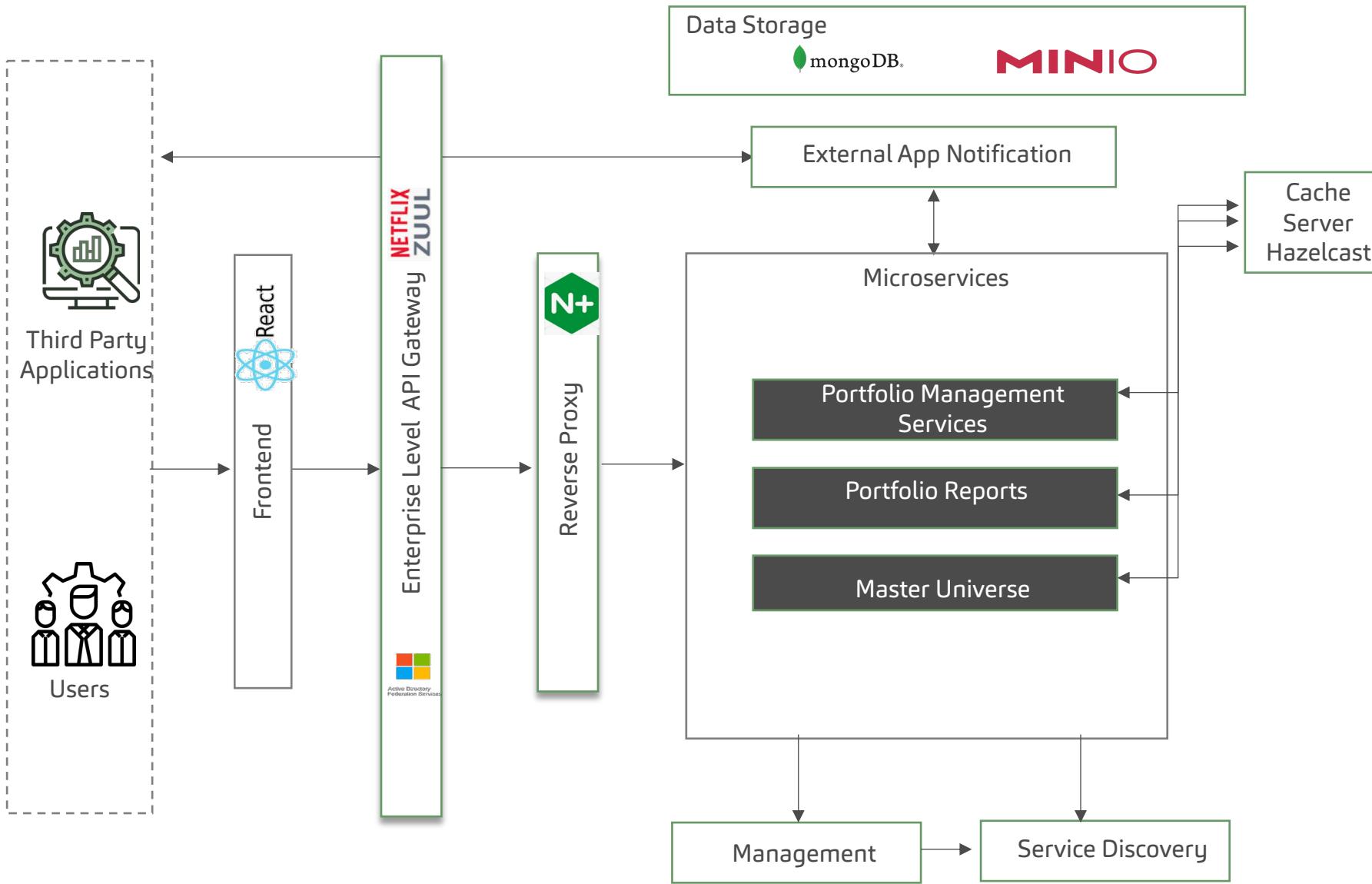
Portfolio Hub – Commands/Queries

Commands and Queries represent any kind of operations within the bounded context.



Conceptual View - Current

All bounded context becomes Service Family & Inside that there are microservices built in.

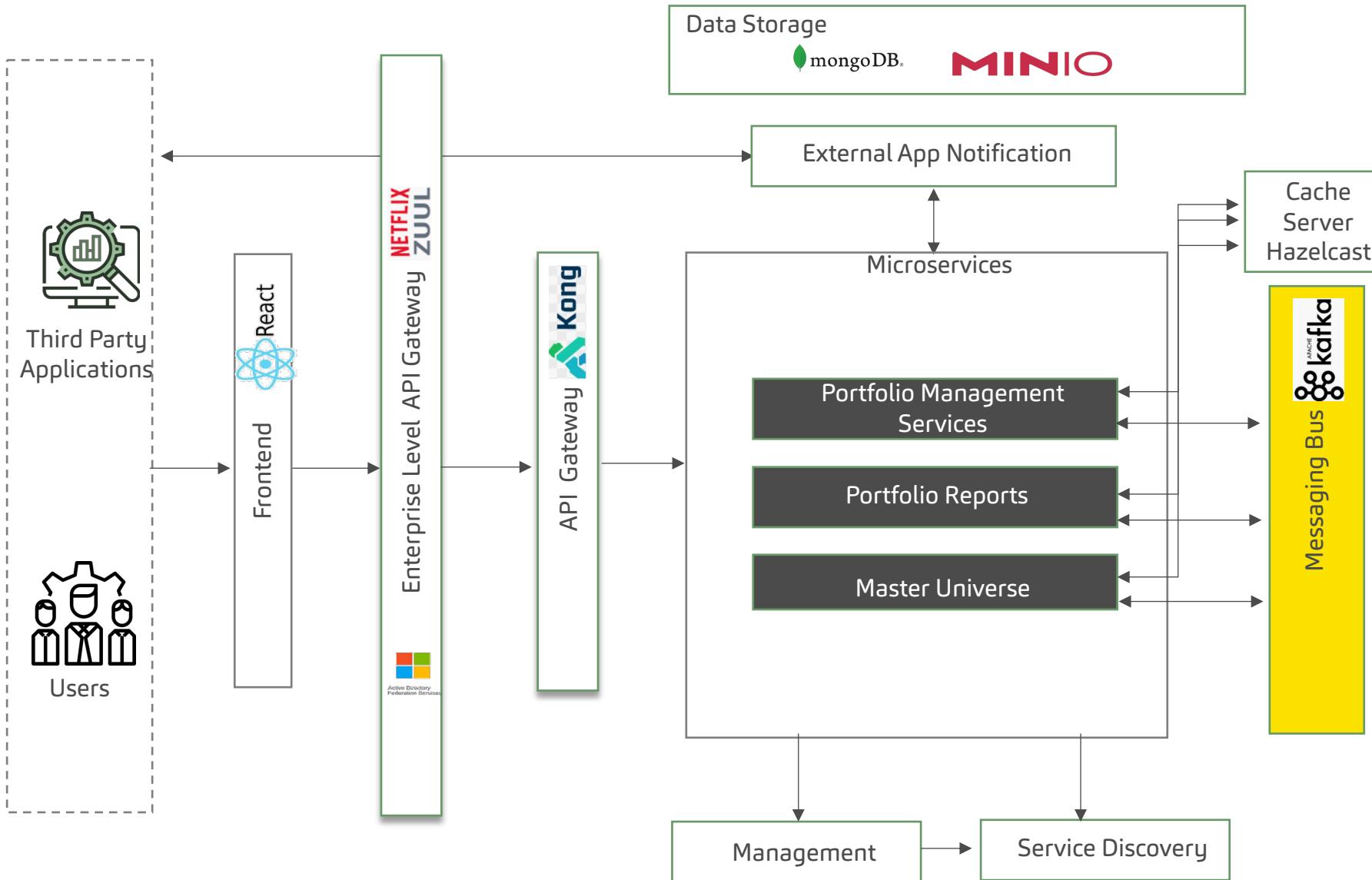


Overview

- **Enterprise Level API Gateway** – A single-entry point at enterprise level which is responsible for SSO with the help of ADFS.
- **Reverse Proxy**– It will act as Application-level Gateway which acts as an entry point for all backend services.
- **Management** – Balances services on nodes and identify failures.
- **Service Discovery** – A guide to find the route of communication between microservices.
- **Microservice** – Identified by domain and submain.
- **External App Notification** – This will be event-based system that will notify the data engineering component of any changes that have been made.
- **Cache Server Hazelcast** – To manage distributed cache.

Conceptual View - Future

All bounded context becomes Service Family & Inside that there are microservices built in.

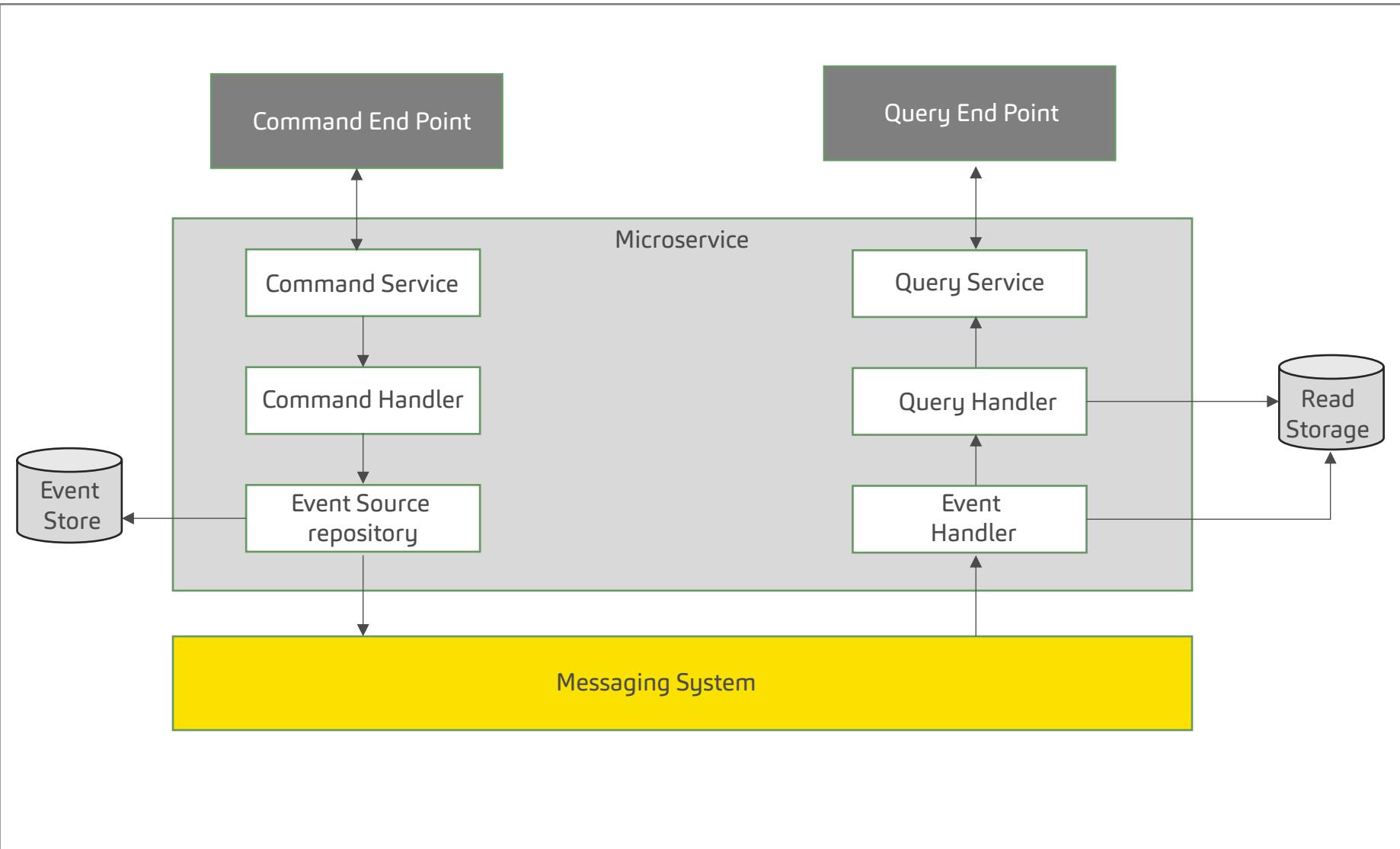


Overview

- **Enterprise Level API Gateway** – A single-entry point at enterprise level which is responsible for SSO with the help of ADFS.
- **API Management** – Application-level Gateway which acts as management tool for all backend services.
- **Management** – Balances services on nodes and identify failures.
- **Service Discovery** – A guide to find the route of communication between microservices.
- **Microservice** – Identified by domain and submain.
- **Messaging Bus** – A publisher subscriber implementation for propagating events between the microservices.
- **External App Notification** – This will be event-based system that will notify the data engineering component of any changes that have been made.

Microservice Architecture

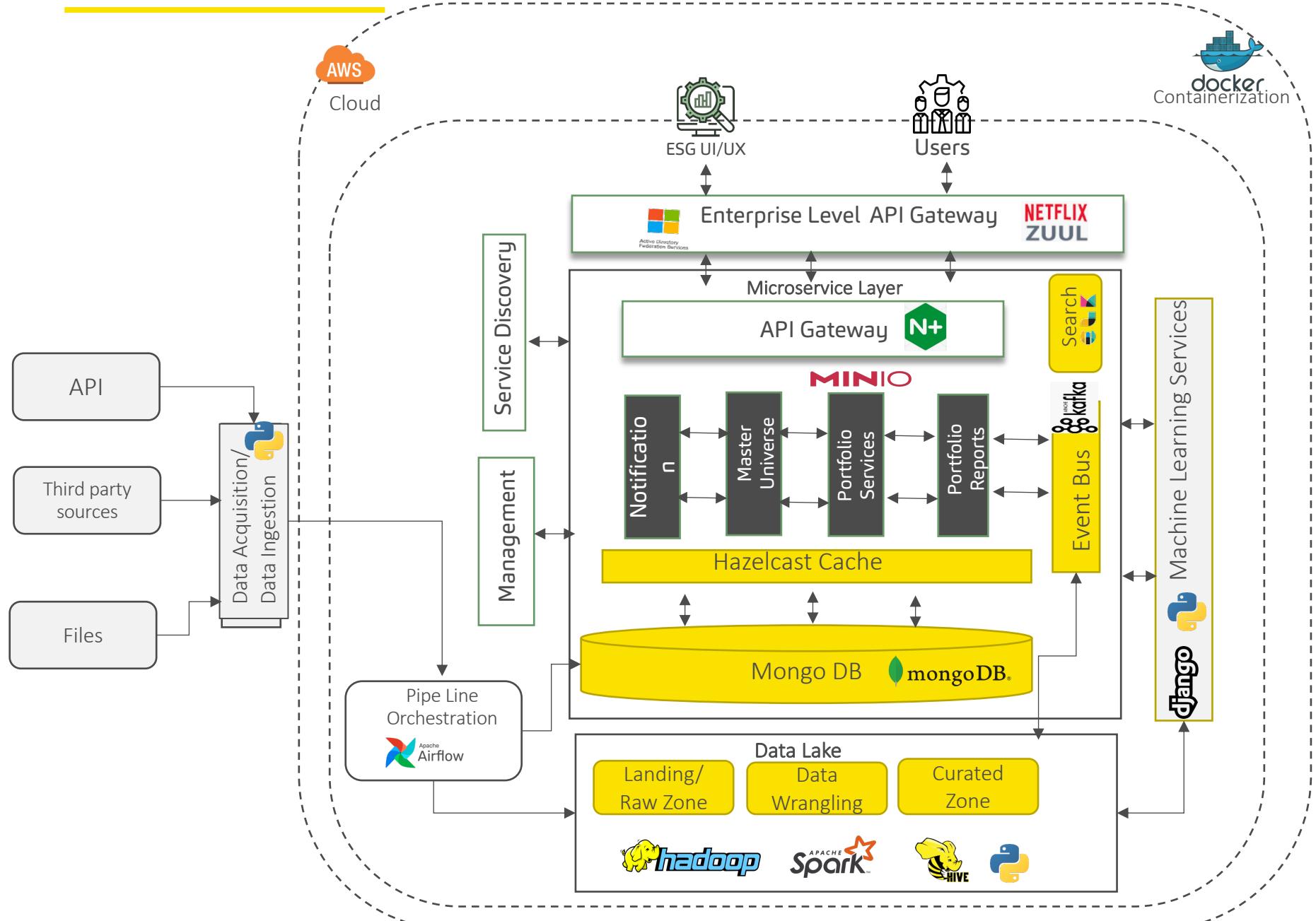
DDD with event sourcing and CQRS



Overview

- We will use two patterns:
 - Event Sourcing
 - CQRS
- Event Sourcing – Everything should be traceable (Ex. Changes in configuration) If you are modifying portfolio for better ESG percentage.
- What we captured is final state, but what if you want to know detail steps?
- At any given point of time. Whenever there is change, that change will be recorded as an event and not as data.
- **Axon framework** will be used for implementing CQRS and Event store.

Data Fabric Architecture



Overview

- Automated Data Ingestion:** The data ingestion service will perform ingestion either scouting for files or will be triggered through an orchestration pipeline using Airflow.
- Data Lake:** All the raw data will be stored in Hadoop Data Lake; any data intensive computations will be performed on Hadoop.
- Data Exploration :** Hive Tables will be created for all the datasets created on Hadoop so that application team, Developers, Business users and Data scientists can write SQL queries to explore the data using Hue.
- Data Wrangling:** All data intensive pre-processing required for ML models will be performed on Hadoop before being consumed.
- Microservice Layer:** Micro services will be leveraged when services will be deployed on cloud.
- Data Archiving:** Historical and System logs & Unstructured data can be moved to Hadoop.

Data Ingestion framework

Overview

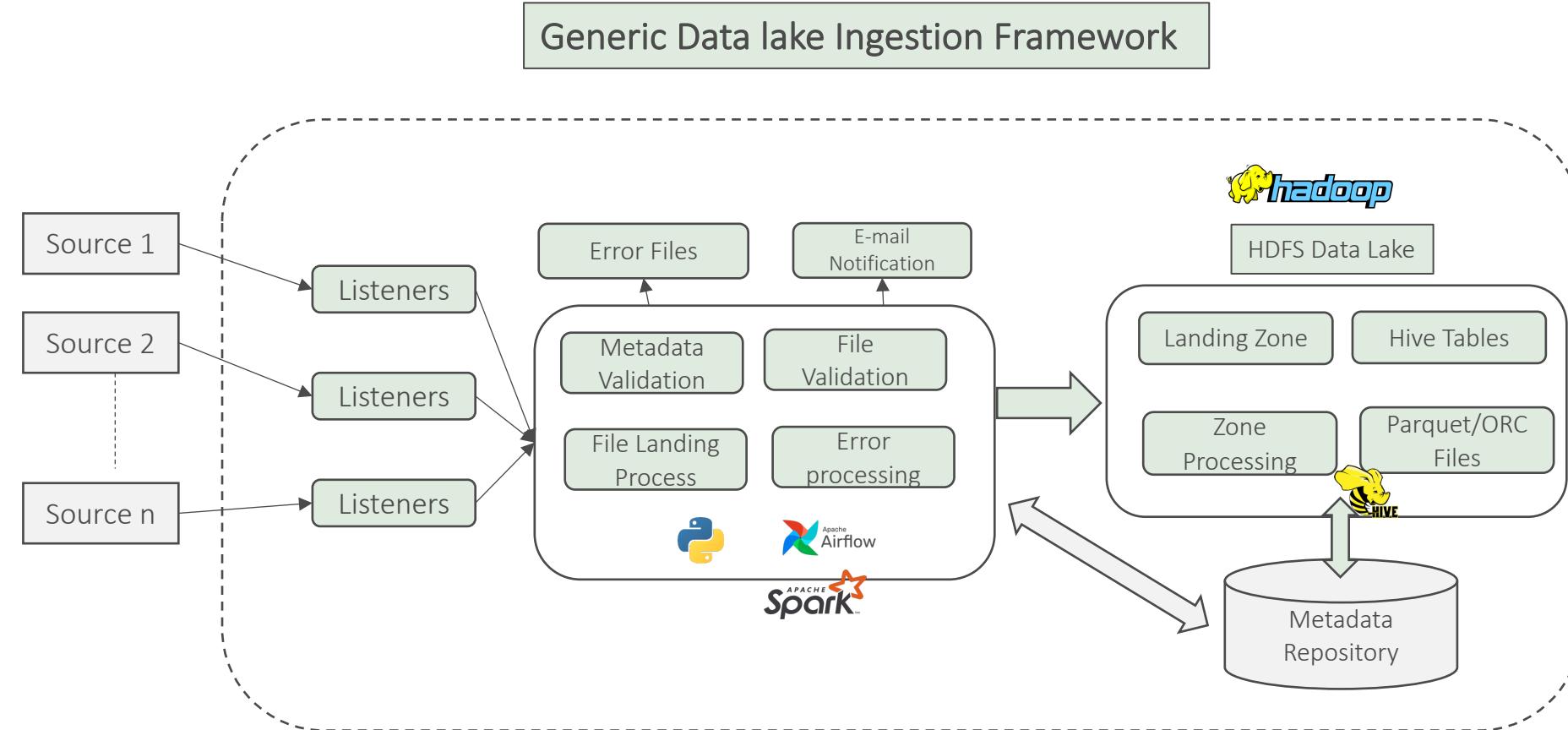
Automated Data Ingestion: The data ingestion service will perform ingestion either scouting for files or will be triggered through a orchestration pipeline using Airflow.

Multiple Data Formats: Multiple file formats of files can be extracted based on the needs and ingested on to Hadoop platform.

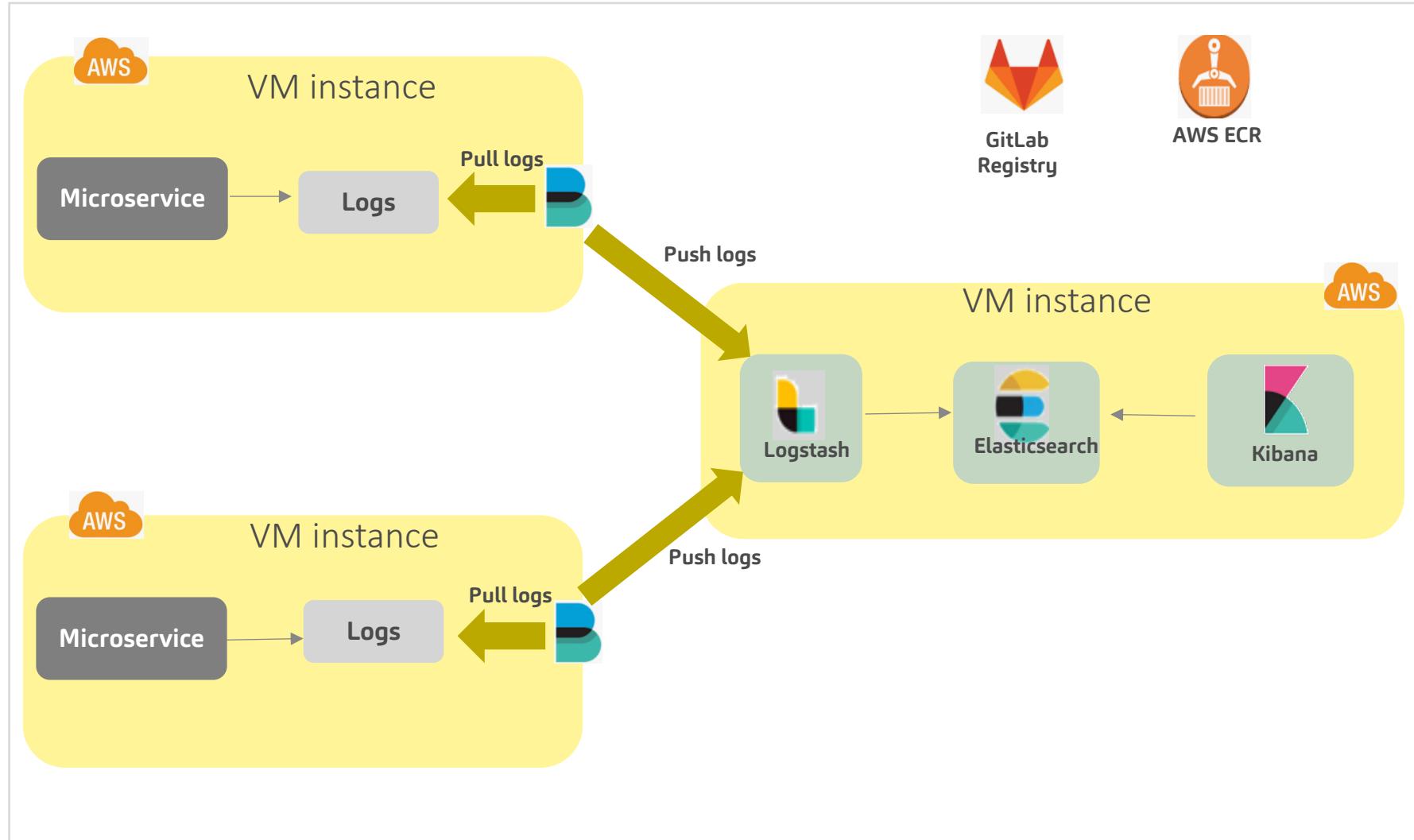
Hive Tables: Hive tables are created automatically once files are landed on HDFS.

Hue Interface: Hue interface will be provisioned to write SQL like queries to explore Hive tables.

Metadata Validation: Each time a file is ingested, it is validated for Metadata before its gets into the correct partition. If the validation fails then E-Mail notifications will be sent to designated group or person for manual intervention and trigger error processing.



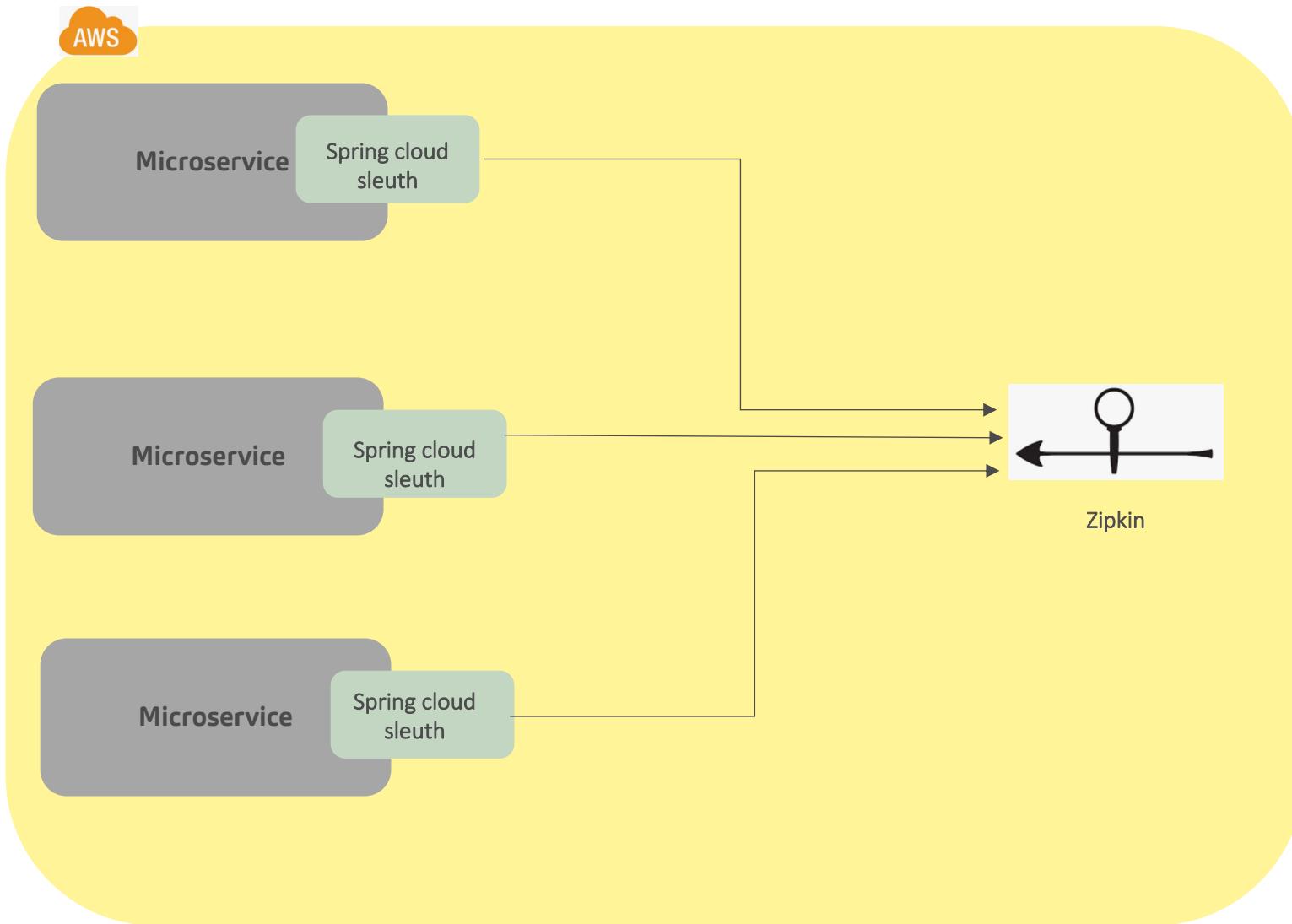
Centralized Logging Architecture



Architectural Components

- AWS ECR container file beat image.
- Gitlab registry containers docker file for microservices. These docker file will have reference to Filebeat image from AWS ECR.
- Microservices are running on multiple servers.
- ELK is running on separate AWS server.
- Filebeat – Instead of running Logstash on multiple server. As it's heavy process and consumes lot of resources, Filebeat are installed which are lightweight data shippers.
- Logstash – It will collect the logs from different Filebeat agents, transform to JSON messages and fed into Elasticsearch.
- Elasticsearch - Elasticsearch is a search engine based on the Lucene library.
- Kibana – Data Visualization dashboard for ElasticSearch

Distributed Tracing Architecture

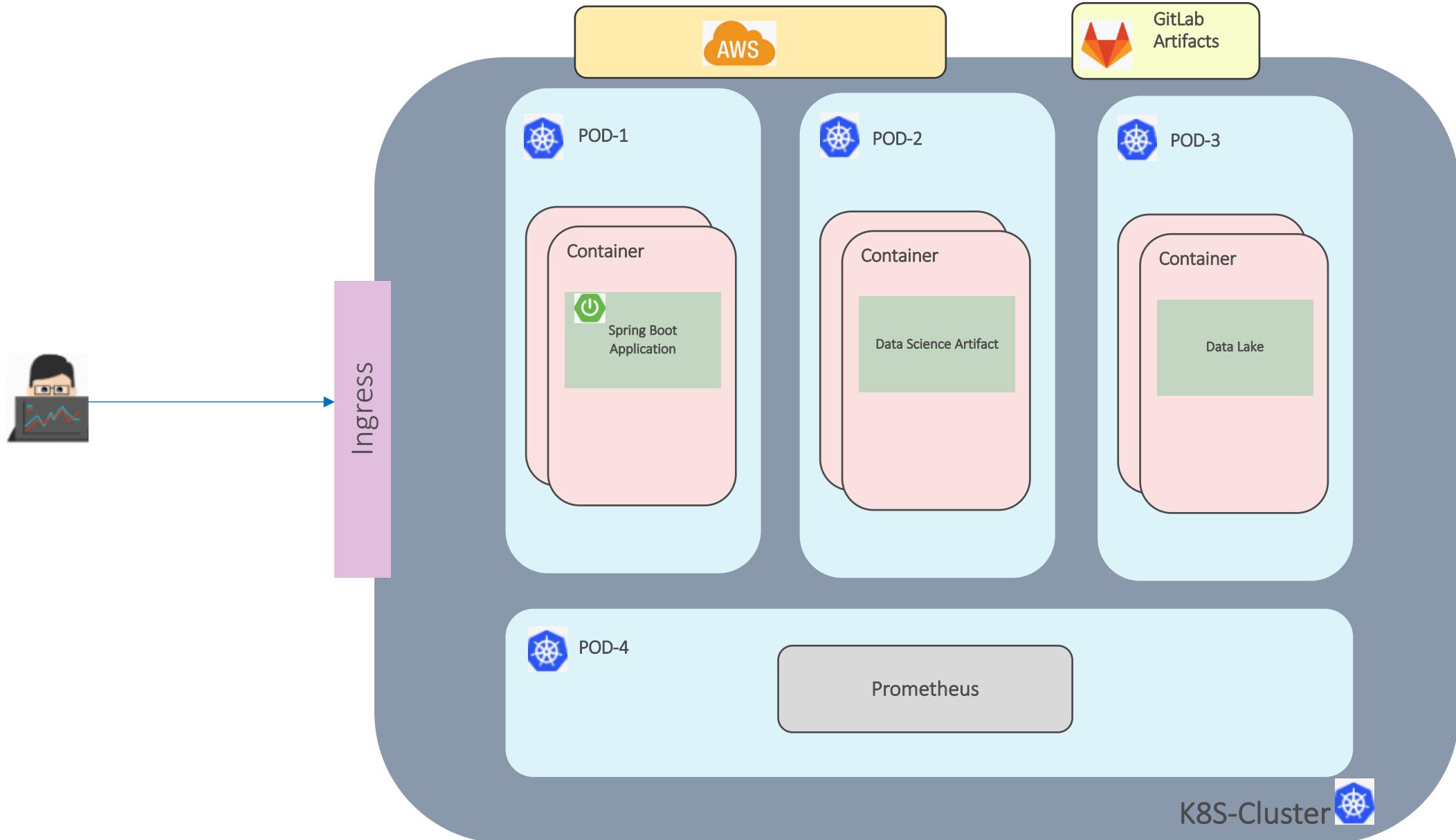


Architectural Components

- Spring Cloud Sleuth: A Spring Cloud library that lets you track the progress of subsequent microservices by adding trace and span id's on the appropriate HTTP request headers. Sleuth adds two types of IDs to the log file, one called a **trace ID** and the other called a **span ID**.
- Zipkin: A Java-based distributed tracing application that helps gather timing data for every request propagated between independent services. It has a simple management console where we can find a visualization of the time statistics generated by subsequent services.

Deployment Diagram

Structure of run-time system



Technology Choices

Frontend



Discovery



API Gateway



Messaging



CDC Tool



Data Science



Framework



Data Lake



Log Tracing



Analytics



Programming Language



Cache



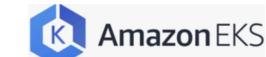
Storage



Cloud



DevOps



ESG:

Influence score – measure the influence of companies and their underlying factors to overall portfolio scores

decomposition of weighted average formula

influence score has been calculated based on holding % and ESG score

Normalization and comparison ESG score from multiple providers

Min-max normalization

Mean normalization

Z-score normalization

Detecting discrepancy between ESG providers

using different outliers detection methods

News analysis

Applying Named Entity Recognition to identify company name for news articles

Classifying articles in order to filter out only ESG related news (Labor issues, Pollution, natural disasters, etc)

Measure sentiment score

Forecasting next-year ESG score using historical data and aggregated sentiment score from news articles

Using econometric models to capture the impact across different industries

Showing the difference between models

Explain how the sentiment score is driving predictions

Current Module

Future Modules

Core Data Science Components include –

- Normalizing ESG scores across providers; and
- Predict future values– **Miroslav to add here.**

Data Science methods employed –

- ESG scores are provided by various providers using their proprietary methods which often do match across
- Sometimes, the direction of scores in opposite direction provides similar risk categories (Example – Provider A – higher score indicates higher risk, Provider B – Lower score indicates higher risk)
- There is a huge variation in scores across companies considered in our analytics; within one provider and across providers
- Used ‘Z Score’ to normalize ESG scores, which considers both central tendency (Average) and Dispersion (Standard deviation) of data and provides an intuitive meaning of how far an individual company ESG score from its mean in terms of standard deviation
- An outlier is computed as a combination of Normalized Score and Directionality between providers and flagged for user consideration
- Those companies which are outliers in a given portfolio is highlighted for User attention.

Thank You

 Contact Person: xxxx

 Email: xxxx

 Phone: +xxxx

 Location: Pune, India



New York

Synechron, Inc.
11 Times Square, Suite 3301
New York, NY 10036
PHONE +1.212.619.5200

India

Synechron Technologies Pvt. Ltd.
CEDAR Building, Ascendas IT Park,
Rajiv Gandhi Infotech Park,
Hinjewadi Phase III, Pune 411057, India
Phone +91 20 4290 1000