

Data Warehouse Implementation for an Innovative Consumer Goods business

Digital
Services

Success Story

Client

The client is one of the innovative business in the Consumer Goods industry with various product categories under different brand names ranging from Personal Care, Professional Care, Food & Beverages and Dairy Products.

Overview

Owing to the nature of the business, data arising from hundreds of thousands of small-scale points of sale transactions, SKUs etc. is created in high volumes. However, the data was residing in silos needing a technology solution to create a centralized data warehouse with high speed querying to support a decision making and planning system for Sales and Marketing operations. Indium software delivered an infrastructure that empowered the client with a dynamic support system for on-demand analytics.

Status Quo

Owing to the diverse product range, large volumes of data associated with the products and brands contributed to the stack of their data sources. However, the data was lying in silos and needed a technology handle to create a centralized data warehouse that would be a potential business decision and planning system for Sales and Marketing operations.

Business Requirements

- Design and develop a robust Data Lake solution that supports high-speed Data Querying for Data Discovery and Analytics.
- Consume data from various data sources including SAP, Enterprise Data etc. to model data that acts as a precursor for required data analytics use cases.
- Empower the client with a dynamic decision support system for on-demand Analysis.

Indium Software's Approach and Implementation

- Indium Software structured a robust and highly scalable Big Data ecosystem to address the Data Storage and Processing needs of the clients business.
Key objectives:

- Non-volatile and real-time querying Data Warehouse support.
- Provision high availability needs of Data Warehouse for On-demand reporting and analytics.

Business

Data Warehouse

Domain

Consumer Goods

Tools

SQOOP – ETL
HIVE – Spark (Data Warehouse)
Spark – Machine Learning
Hbase / HDFS – Data Lake
NoSQL

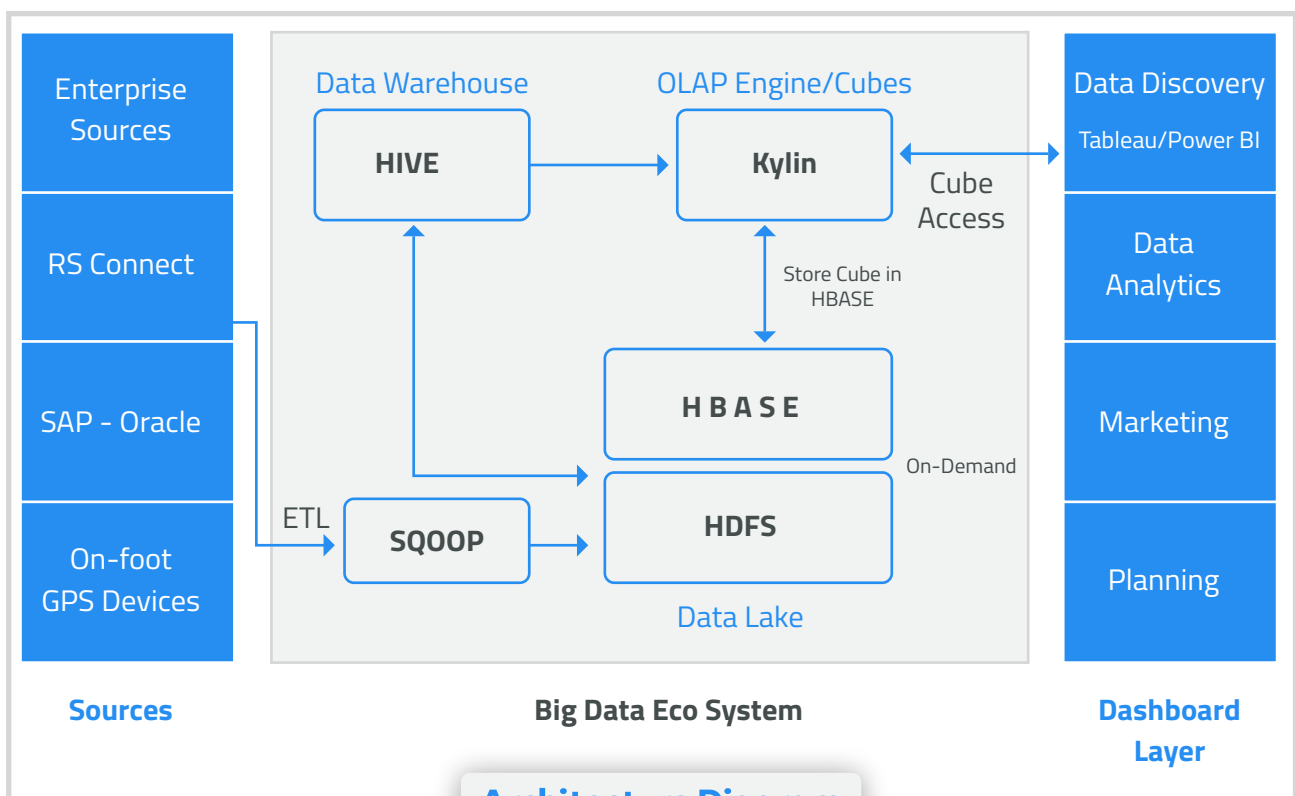
Key Highlights

- Query latency for 10+ billion rows reduced by 70%, with response within seconds
- A future-proof Big Data ecosystem with high scalability features at a low cost

- Performed Data cleansing and Integration prior to storage in the Big Data - HDFS.
- Implemented HDFS using SQOOP (Big Data ecosystem) for Data Acquisition, ETL Processes.
- Used HIVE to create data tables and pre-join Star Schema for cube building.
- The solution uses Apache Kylin which acts as a high-speed OLAP Engine that is designed to reduce query latency on Hadoop.
- Kylin supports integration with BI tools such as Tableau/ Power BI.
- Utilize HBase to store and query Data Cubes.
- The solution supports the transfer of large data volumes from the client's operational systems to HDFS in specified intervals (Data Refresh).

Business Impact

- Architected a future-proof Big Data ecosystem with high scalability features.
- Enabling Cubes improved faster data slice and dice for business users. Analysis parameters can be prepared on demand with high responsiveness.
- Query latency for 10+ billion rows reduced by 70%, with a response within seconds.
- On the fly, querying enabled with improved visibility into the data inventory.
- Low-cost infrastructure with the ability to run on commodity-grade hardware.





INDIA

Chennai | Bengaluru | Mumbai
Toll-free: 1800-123-1191

USA

Cupertino | Princeton
Toll-free: 1 888 207 5969

UK

London

SINGAPORE

+65 9630 7959



General Inquiries
info@indiumsoftware.com

Sales Inquiries
sales@indiumsoftware.com