

Wholesale Client 360 Foundation for leading US Financial Services Major

Business Value

- 40% faster data acquisition through data ingestion framework
- 30% more accurate matching through customizable rule based customer matching framework that allows probabilistic matching functions
- Reduction in development and support cost for adding new sources due to ingestion framework
- License cost due to open source technology



Challenge

- No single view of Wholesale customer across business.
- Fragmented data landscape that does not provide a comprehensive view of the customer.
- Need for data integrity maintained for customers across systems.
- No single view of customer product holdings.
- Inability in determining profitability resulting in sub-optimal pricing, segmentation and targeting.

Solution

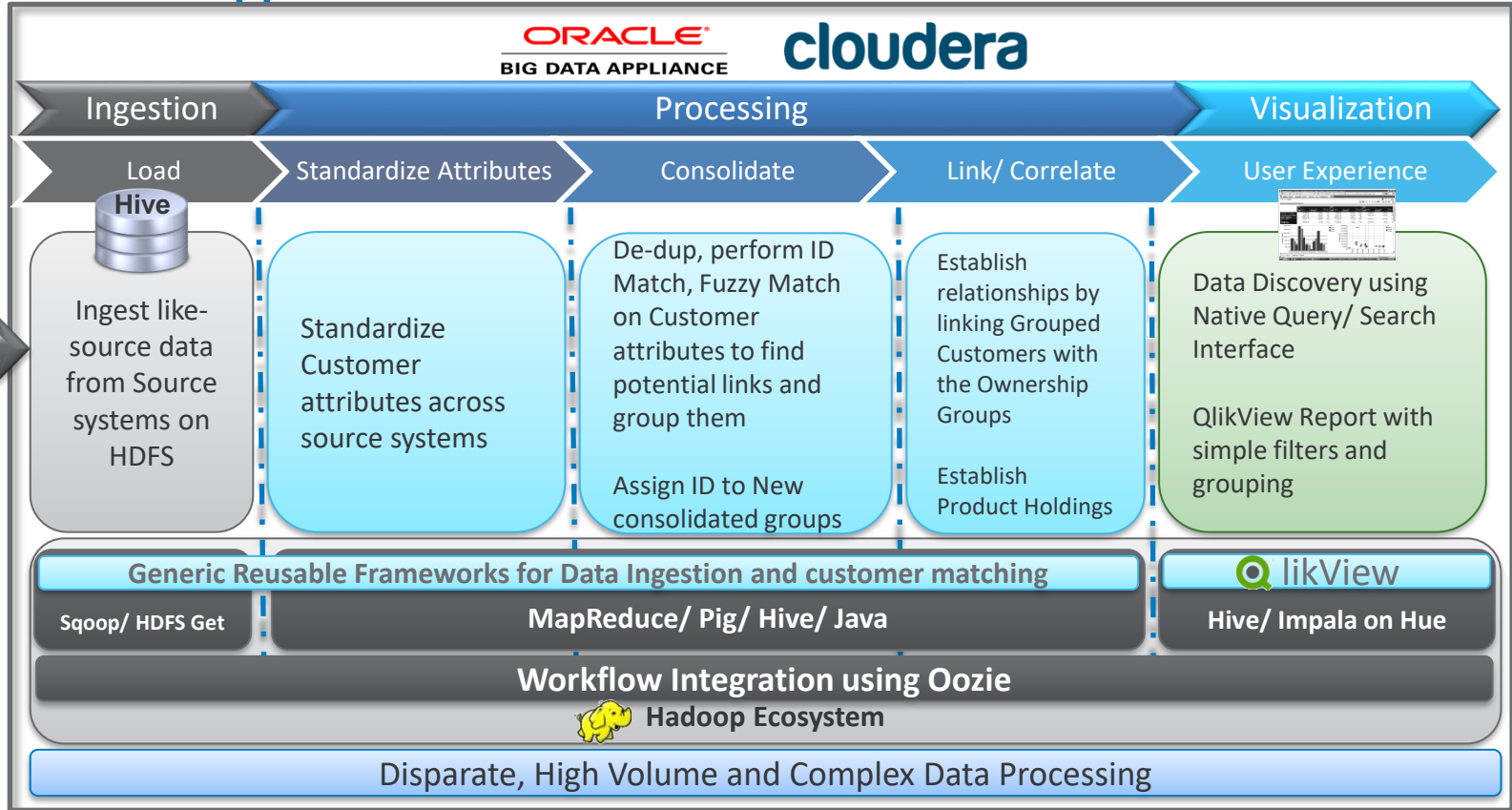
- Reusable and extensible data ingestion framework, Attribute Value Matching framework to prescribe golden attributes and customizable and rule-based customer matching framework for identity resolution and correlation.
- Ingested data stored in Avro format to allow for schema evolution to address any changes in the source data formats.
- The customer matching framework leveraged deterministic and probabilistic matching algorithms such as Jaro-winkler, Levenstein, Soundex and Double metaphone.

High Level Solution Approach

14 Systems, 4000+ tables, 4TB daily volume



Mainframe
DB2
Oracle
SQL Server
LUW Servers



Future Use cases across Wholesale Client Lifecycle - Different perspectives



Data Ingestion Framework for large Card Company in US

Business Value

- Promote reuse and access to authentic data, accelerate development of analytics, develop an architecture for importing data from various sources and make it available for analytics in a consistent manner.
- Provide an industry standard, compliance oriented, secure framework to support the Big Data platform's offerings to business users
- Platform with capabilities for efficient storage and retrieval of huge volume of data.
- Reduce storage cost.
- Platform with capabilities to perform real-time analytics using CEP (Complex Event Processing) on huge volume of transactional data



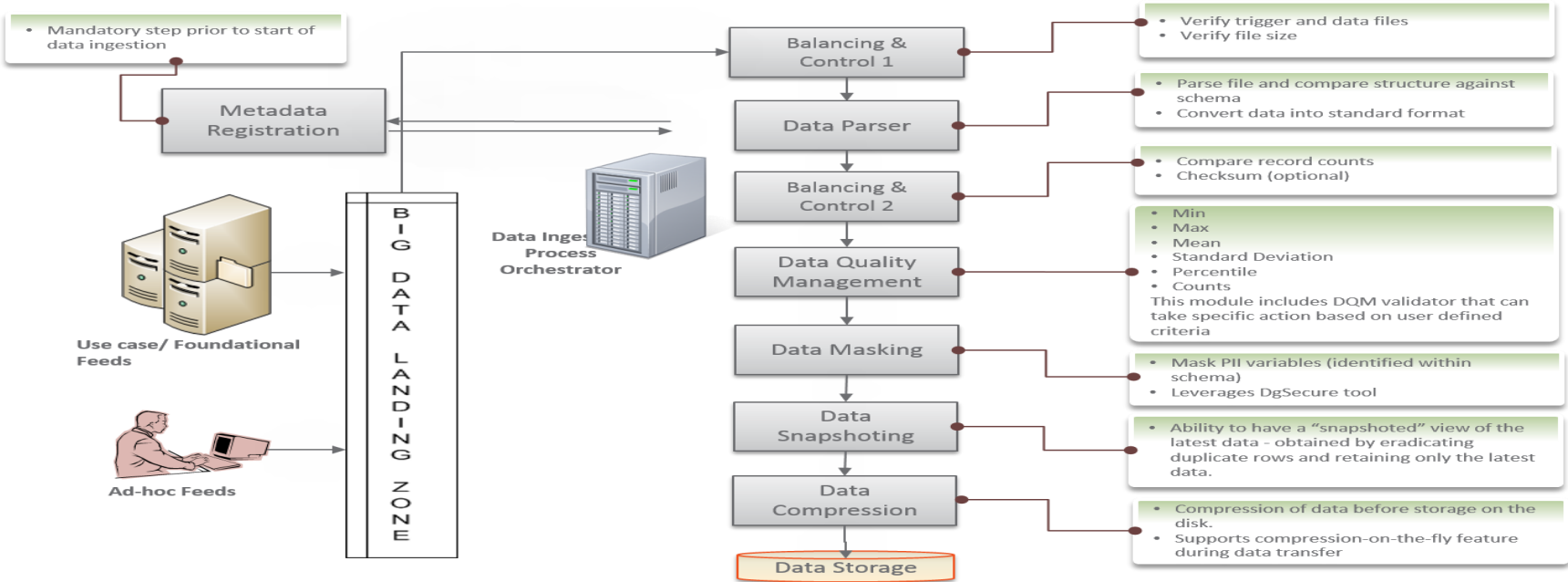
Challenge

- Resolve technical pain points with data coming in from various sources to enhance interoperability.
- Handle large volumes of data and support distributed processing.
- Support innovative solutions across the enterprise and fully utilize Big Data capabilities
- Scalable framework for complex, large scale analytics

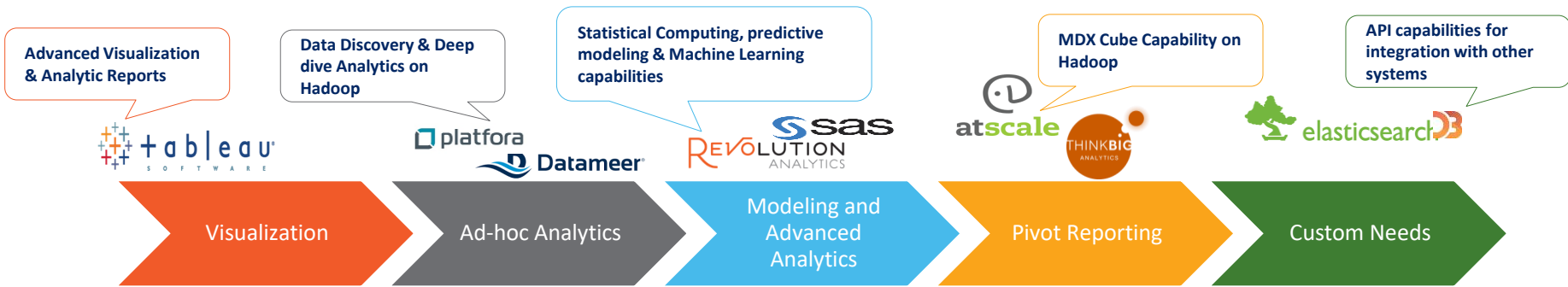
Solution

- Set up process to periodically ingest incremental and historical data into Hadoop.
- MapReudce program to Implement processes to manage data reconciliation, balance and control and error/exception handling and finally load into Hive tables.
- Designed an application data model for Metadata management to store the layout of Hive tables.
- Provide a GUI interface to allow users to browse and search existing data in Hadoop

Data Lake – Ingestion Framework

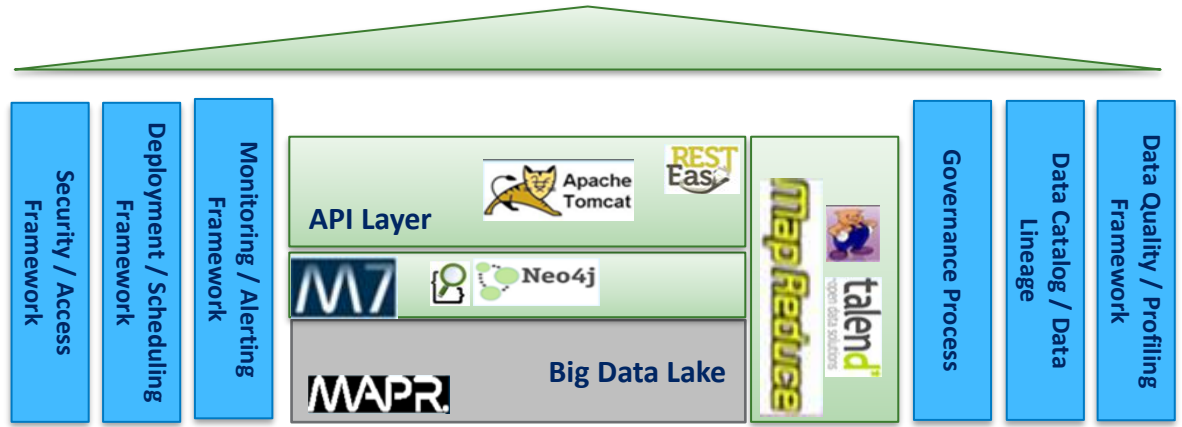


Reporting Data - Self-service on Big Data Landscape



kognitio

- In-memory Analytical Platform for Big Data - MPP
- High-Performance Analytics
- SQL on Hadoop
- Commodity Hardware



Data & Analytics Lab @ a World's Leading Retailer

Business Value

- Reduced the Time to insights from 3 days to 1 Hour leveraging the power of Hadoop
- Statistical models were having a lot of sampling and randomness bias. Using Full data to execute the models provided accurate results and an uplift in campaign effectiveness

Challenge

Client wanted to setup an agile analytics environment that can help them perform data experiments and respond to business needs faster

Key challenges faced by client

- Huge volumes of enterprise and external data
- Requirements first approach to BI leading to longer cycle times for analytical insights

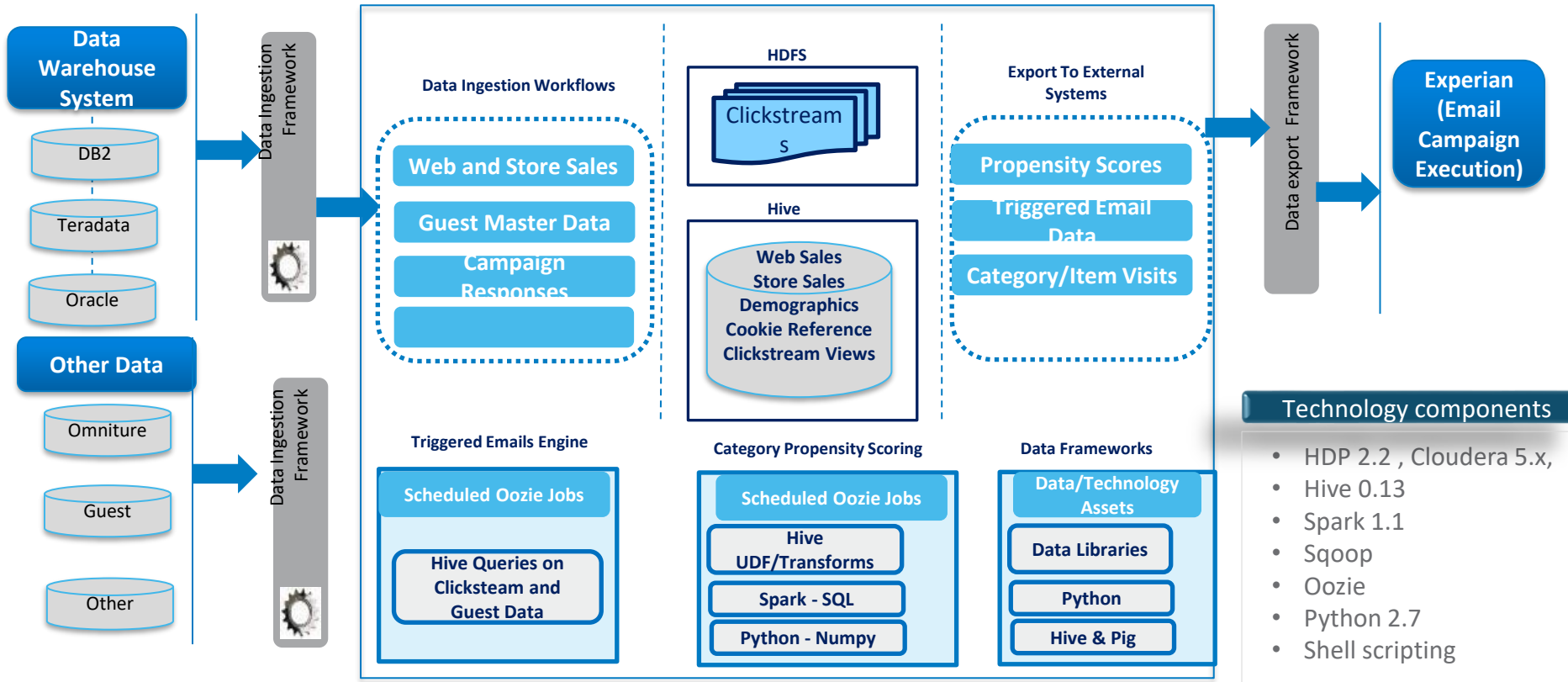


Solution

- Set up the Marketing Data labs for the customer by Leveraging a 300 Node Hadoop Cluster to improve model build performance, hence the client was able to perform multiple simulations before arriving at the final customer segment
- Fast Failure approach to perform Data Experiments quickly
- Helped the customer derive value out of the Site catalyst and Clickstream raw data available

End to End Solution Blueprint Marketing Analytics and Campaigns

HDP 2.2 Cluster (300 Nodes)



- Technology components**
- HDP 2.2 , Cloudera 5.x,
 - Hive 0.13
 - Spark 1.1
 - Sqoop
 - Oozie
 - Python 2.7
 - Shell scripting

Marketing Insights Platform for Global Beverages Company

Business Value

- Created Interactive Dashboards to Provide Insights
 - Brand Equity Dashboards
 - Volume and Market Share
 - Consumption Growth Dashboards
 - Brand & Corporate P&L

Challenge

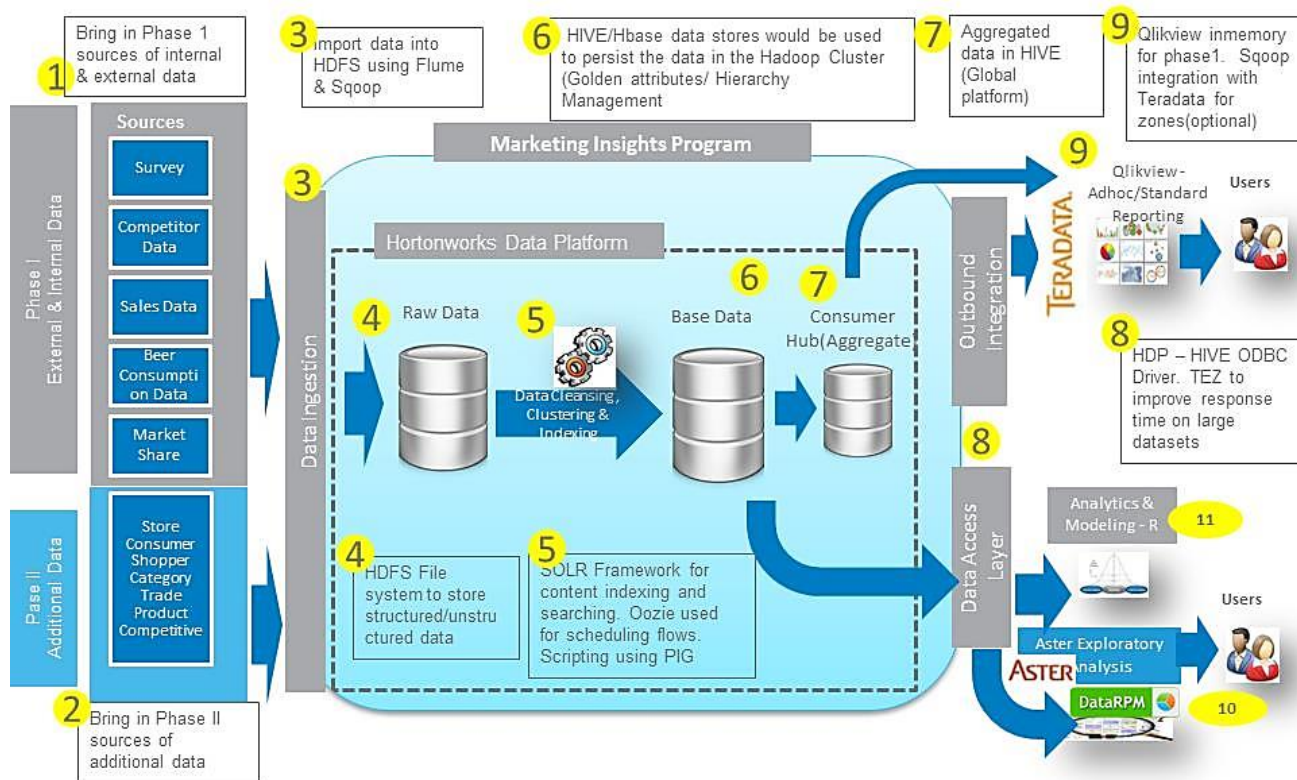
- Client was facing challenges in business growth due to
- Significant competition and pricing pressures
 - Fragmentation of market with new entrants in new categories
 - Category growth flat to declining in some segments
- Strategic objective that drove the analytics initiative
- Increase penetration of Brands
 - Increase # of occasions for consumption
 - Increase value from each occasion



Solution

- Designed Scalable, Flexible Platform Based on Hortonworks
- Delivered Program Using Agile and Iterative Approach
- Created a common data integration layer as part of Phase I
- Built in exploratory analysis capabilities as part of Phase II
- Enabled business to predict future state by predictive analytics based on market conditions

Designed Scalable, Flexible Platform Based on Hortonworks



Phase I: Create a common data integration layer to integrate both internal and external data. Build a common data layer which act as single source of truth for consumer insights. Provide visibility across data by adhoc and standard reporting.

Phase II: Analyze business actions by exploratory analysis. Better insights and root cause analysis and their impact on KPI's

Phase III: Drive Business and predict the future state of business by predictive analysis. Identify and respond to market conditions quickly, minimize risk and improved resource allocation.

Next Generation Analytics Platform at the largest footwear retailer in US

Business Value

- Auto scaled platform / infrastructure to meet seasonality
- Next Generation platform for an on demand exploration
- Efficient analytics cycle time and self service analytics
- Capability to identify consumer across channels and understand activity, interest, products for personalized campaigns resulting in **increase in product visit ratio by ~ 40% & product visit conversion ratio by ~20%**
- Move from “Blast Communication “ to Personalized Omni Channel Experience effective Cost Per Action(eCPA)

Challenge

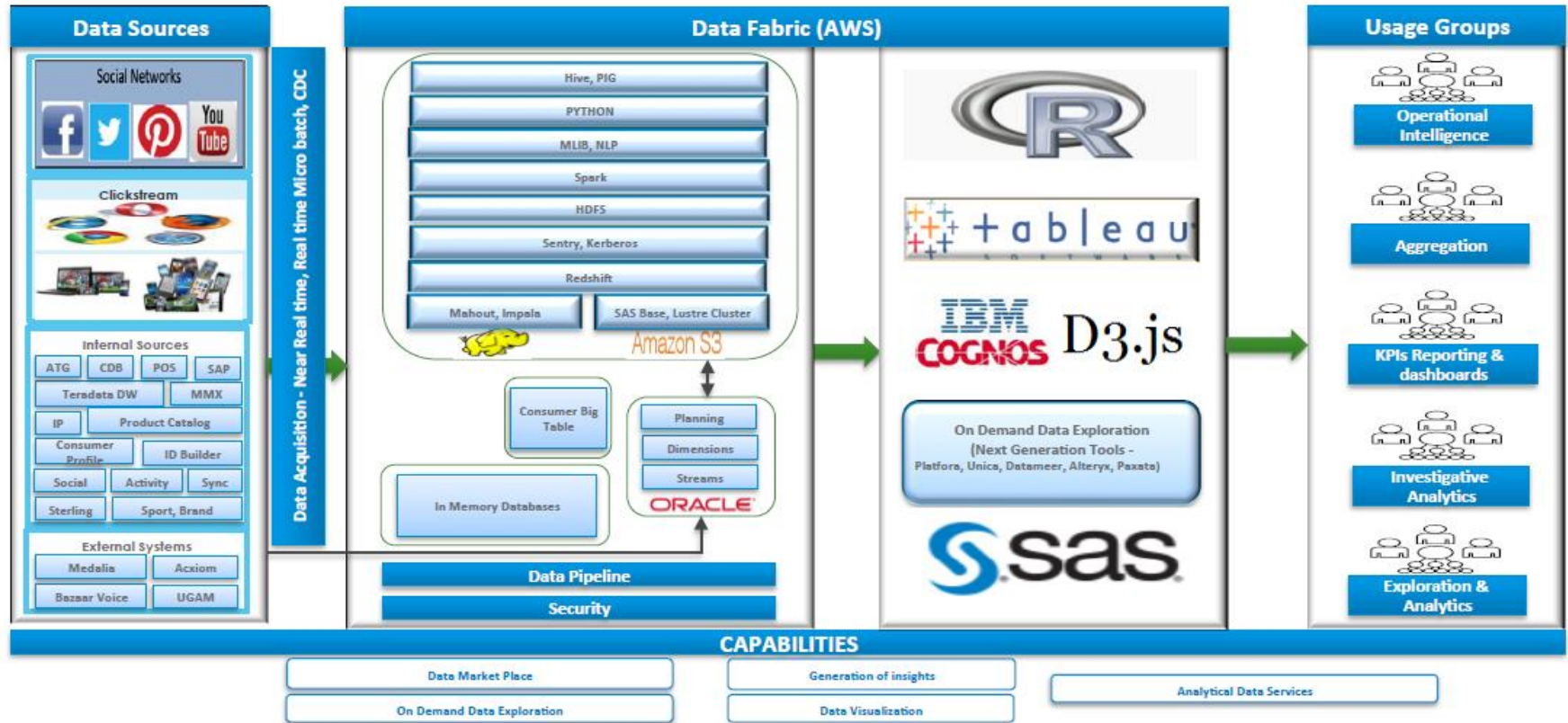
- Build next generation platform to for
 - Build a unique digital profile of a consumer.
 - Consumer Analytics - Build digital consumer profile tapping in to social, click stream, mobile apps and traditional data sources and hence improving consumer experience , brand recognition and campaign effectiveness via personalization.
 - SKU Forecasting, Store Curves Size Profile Optimization, Labor optimization etc.
- Scalable and fault tolerant system that can handle any high volume data



Solution

- Conducted platform evaluation for exercise for best fit tool
- Implemented a Cloud based solution leveraging Amazon Simple Storage Service (S3), Machine learning by AWS, Scala, Spark, Redshift, DynamoDB to minimize costs and leverage the elastic scalability features available
- Implemented various ingestion patterns to ingest data from a variety of on premise and external source onto Cloud
- Enabled Next Gen analytics, Social analytics and advance visualization using D3JS & Tableau

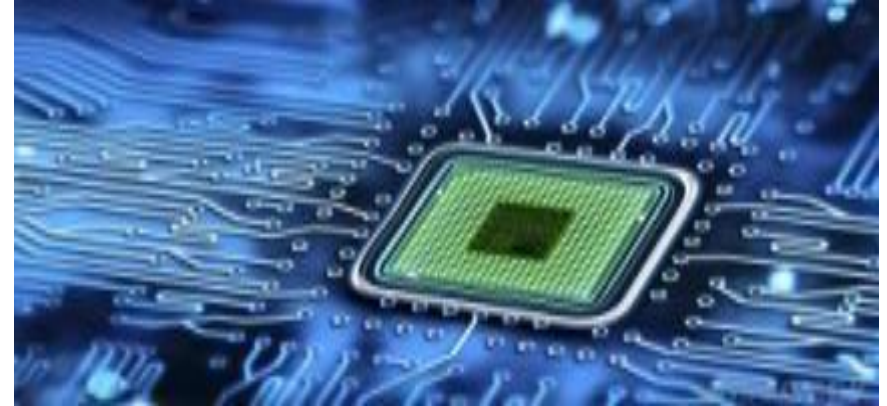
Solution Architecture



Augmented DW leveraging Big Data for a Major US High Tech Manufacturer

Business Value

- Overall annual operational cost for processing reduced from 300M to 80M (from 50-200K per TB to 1.5-10K per TB)
- Loading of 100GB data reduced from 2 hours to 10minutes
- Data processing of 100GB to 30GB reduced from 4 to 2 hours



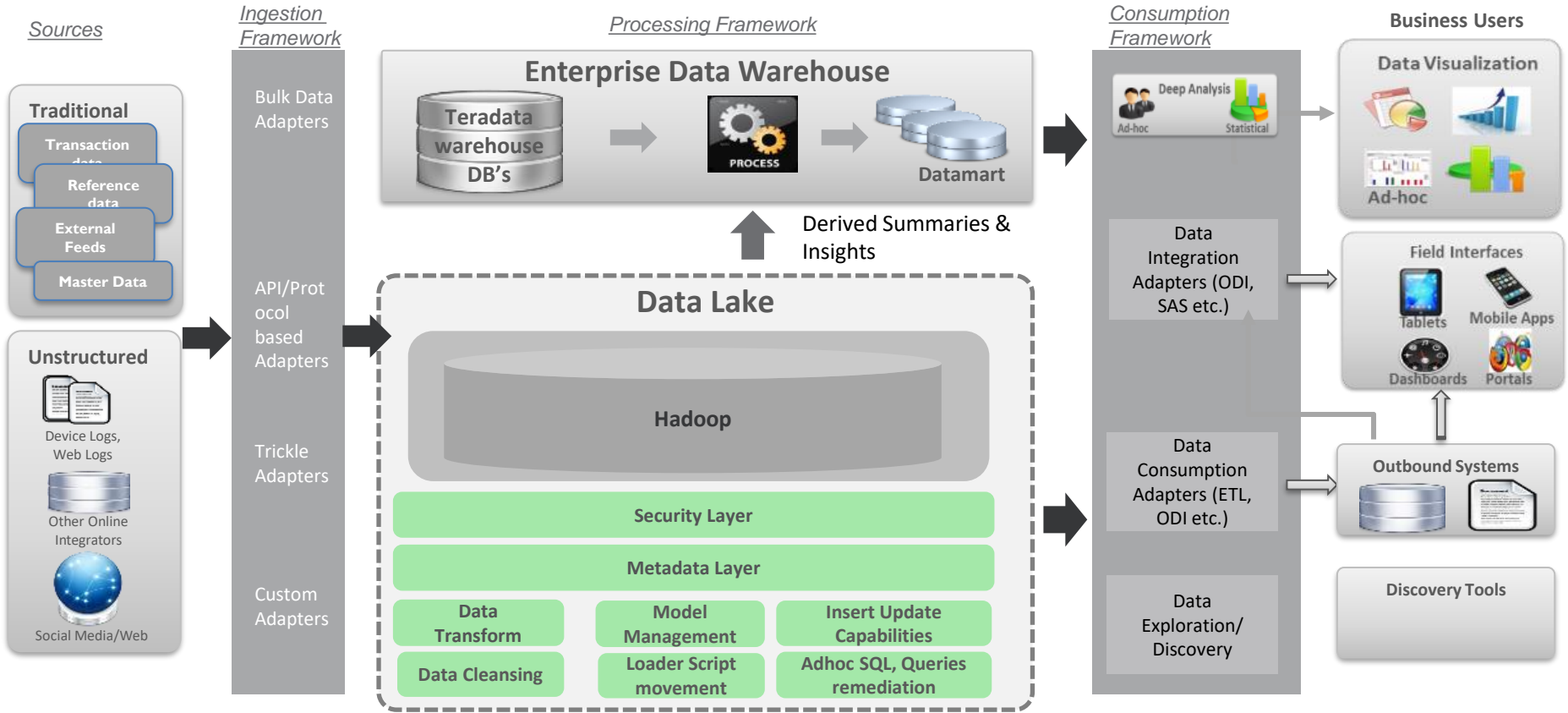
Challenge

- Escalating data processing costs of its large EDW
- Exponential increase in the Infrastructure cost due to data demand growth
 - Analytic and reporting queries need to scan large data (75% of total data)
 - Increased demand to process unstructured data
- Longer load times than desired

Solution

- Implemented solution based on Big Data technologies to augment current EDW
- Leveraged Informatica for extracts from source systems and loads
- ETL and analytics done on Hadoop cluster
- Reporting needs supported by Teradata
 - Processed data from Hadoop cluster was consumed by reporting systems

Solution Architecture



Big data Lake Implementation for US Insurance Firm

Business Value

- Movement of non-business critical data from Teradata helped in reduced storage, processing costs. Free-up server resources to new onboarding applications
- Showcasing big data foundation blocks viz., accelerators and framework helped business groups to perform iterative analysis of data elements
- Phased implementation with minimal disruption to business systems helped in better adoption of big data



Challenge

Current state architecture constraints' on processing claims, premium, policies does not help in timely, high quality data

- Increased cost of RDBMS based systems that needed quick solution
- Tech-stack limitations in current state that inhibit gathering business value of data in quick manner
- Increased IT Costs for supporting existing Ab-initio and Teradata, Oracle systems.

Solution

- Created big data COE blueprint to establish process, standards for big data adoption across enterprise
- Created XML Ingestion tool to help business users analyze loss data in Hadoop and prescribe additional elements
- Deployed "Active Archive tool" to help bi-directional data move between Teradata/Oracle/SQL Server to Hadoop
- Created a Big Data platform reference architecture design