

Case Study

Data Lake for NBFC

A solution approach by Dataeaze Systems



About the client

NBFC having offerings in a wide variety of financial investment and insurance services ranging from : Equity, Mutual Fund, Insurance, Wealth Management. Total AUM managed across all businesses is more than \$2 billion



Client need of Data Lake



Make consolidated analysis possible

Client business contains multiple business units, systems supporting these business units were developed over period of a decade using variety of database systems and data models.

This variety of data across systems made it very difficult to achieve consolidated analysis of client across all businesses.

Extracting reports across businesses involved manual process of data extraction and consolidation which lasted for few days.

In order to solve the above problem, it was decided to build a consolidated data lake and data mart.



Handling scale of data

Client has large scale of data, with approximate 200k+ customers, total 50+ million transactions across all businesses with daily few thousand transactions being added.

Investment data reporting contains KPIs which require revisit of all transactions everytime, serving such reports at large scale of data was becoming difficult with existing database systems, which needed massive scale up resulting into increase in costs.

In order to solve this, it was decided to build open source big data lake which is made to serve large scale data processing.



Building new age digital solutions

Client has entered into a new age digital wealth management offering which is targeted for million+ customers. This offering allows client for self serve wealth management.

Considering the scale of data, it was decided to make use of modern big data platform for this.

How has dataeaze helped the client?

Built open source big data lake

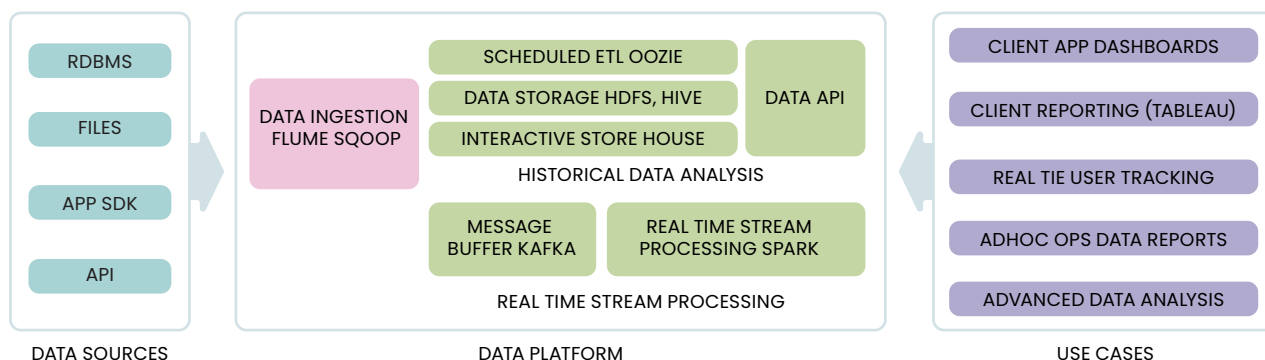
We at dataeaze have built an open source big data lake for the customer.

Have built automated data ingestion process to simplify the complexity of data pull from a variety of data sources.

Data marts are built on top of this data lake which simplify serving reporting, analysis as per business needs. Automated ETL keep on refreshing data mart at fixed interval without manual intervention.

Have enabled real time stream processing capability which serves use cases which need analytics done as soon as action happens at source.

Solution overview



Technology stack

Solution stack based on hadoop based big data ecosystem

Data ingestion

Apache sqoop, flume, custom data access components

Data storage

Hadoop distributed file system, Hive, Hbase, RDBMS for data mart and real time access

Real time stream processing

Apache Kafka, Spark streaming

Data processing

Apache mapreduce, hive, spark

Data access and BI tool

RESTful APIs, Tableau BI Reporting

Benefits to the client

- Enabled consolidated data reporting : With big data lake, consolidated data reporting across all businesses became possible, earlier it was very time taking process.
- Reduced time of making reports available : Earlier reporting processes of manual data extraction and excel based reporting used to take a few days for a new report. With the current system all data is always automatically pre-populated till yesterday. Reporting is done through BI tools so effective report time is in a few minutes.
- Possible to serve growing scale of data : Earlier systems were serving business as well as reporting, it made it impossible to serve growing data with advanced reporting needs. Current data lake is made for handling big data, reports which used to take hours in the earlier system finish within a few minutes with the new system. With properly designed data marts, common needs are pre-calculated and extracting data from those is possible within a few seconds.
- New business units based on a modern data platform : Client has started a new digital wealth management offering which is self-serve for customers with an app based interface. This new offering is enabled by this data mart.
- Self serve BI Reporting : New data lake and data mart has enabled use of a popular BI tool (tableau is being used), which has made reporting very easy, quick and with a new age interface.
- Enable advanced data offering : This data lake is capable of serving predictive analytics and machine learning. Advanced use cases of personalisation, recommendation and prediction are being served on this now.
- Cost saving : This is based on open source stack ie. without licensing costs. Only cost the client bears is of hardware and of maintenance.

dataeaze

www.dataeaze.io

About dataeaze

Dataeaze helps its customers build an analytics data platform around modern big data ecosystem. Dataeaze systems is focused on making it easy for organisations to work with data. Organisations assisted by Dataeaze get benefit of quick bring up of robust data platform with analytics capabilities brought up as per need.

contactus@dataeaze.io