

Centralised Data Lake



Impact



**Estimated
Spend
Reduction by
\$15,000 per
month**



**70% reduction
in time to
market**



**Meeting
SLA efficiency
85%**

Methods



Hortonworks Data platform



Hadoop HDFS



Apache Atlas

Apache Atlas

Apache Atlas



Apache Knox



Apache Ranger



Apache HBase

Developed a centralised data lake platform for a large insurance company

- The client is a large insurance company and had their data on multiple (20+) systems and in multiple formats which resulted in data duplicacy.
- Quantiphi helped them in migration of data from multiple on-prem data sources to cloud and then built a cost effective and scalable centralized data lake platform on cloud by setting up an ingestion pipeline with security protocols and enabled them to visualize & draw insights from in real time.
- In the platform, billing alerts were enabled for any threshold violations, lambda functions were used to start/stop services as per business hours & Glue was used for fully automated & highly scalable heavy ETL jobs.
- KMS and AWS Certificate Manager was to encrypt data at rest and during transit, respectively and encrypted AMIs was used to achieve root volume encryption.

Challenges

- Different teams used multiple tools which resulted in increased cost of ownership support resources.
- Integrating data from systems of different sizes and configuration.
- Large scale data ingestion took required higher turnaround time.
- Insurance claim data is highly confidential and needs to be secured in transit.

Result

- By building a cloud Data Lake solution, most of the existing data is structured, traditional infrastructure is available and querying the data is also easier.
- It enables us to draw actionable insights from the data.
- Single click deployment was enabled which saved time to setup the entire platform.