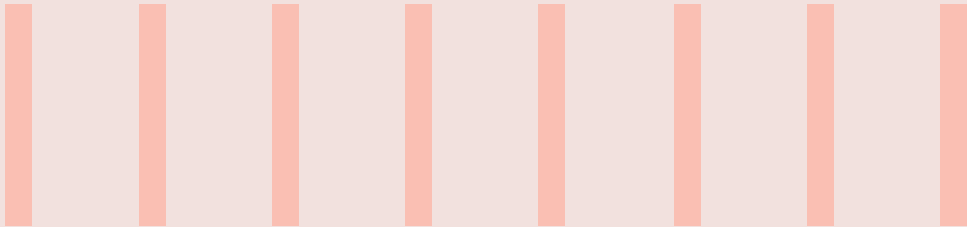




NEXT GENERATION DATA PLATFORM FOR AMERICAS LEADING INSURANCE CLIENT



The situation: Challenges with On-premise Data Lake

The insurance client has built an on-premise data lake (Hortonworks Hadoop platform) with data from multiple data sources, transformation processes and few models and was facing challenges with





growing needs and supporting the current loads.

The data science community and other analysts were not happy with the limitations of tools and performance constraints.

The management was having challenges with cost of the infrastructure increasing continuously and was keen to check the feasibility to migrate to any cloud environment in secured way without having disrupt much.

The need: Support the next gen Analytic requirements

The client has to design the solution in next gen environment for supporting the below personas requirement with efficient usage of resources and less maintenance effort

Persona	Main Tools	Main Activities	Sample Teams
 Data Scientist	<ul style="list-style-type: none"> Machine Learning / Deep Learning Frameworks Spark, R, Python, SAS, Notebooks, GPU 	<ul style="list-style-type: none"> Utilize cutting edge AI tooling to discover competitive data insights Create, train and evaluate machine learning models 	<ul style="list-style-type: none"> SDA DSAL
 Advanced Analyst	<ul style="list-style-type: none"> Spark, R, Python, Notebooks Dbeaver, WinSQL, SQL Developer SAS 	<ul style="list-style-type: none"> Use data prep and data analysis tools to develop prepared data for other business users. Often combine different data sources to create blended datasets for analysis Create machine learning models 	<ul style="list-style-type: none"> Customer Analytics Pricing Actuary
 SQL Analyst	<ul style="list-style-type: none"> Dbeaver, WinSQL, SQL Developer, SAS Excel 	<ul style="list-style-type: none"> Query raw and prepared data Share results with others 	<ul style="list-style-type: none"> Personal Lines Claims
 I/S User	<ul style="list-style-type: none"> Spark, Python, Java Dbeaver, WinSQL, SQL Developer 	<ul style="list-style-type: none"> Productionize data pipelines, reports and models Perform production deployments Provide production support 	<ul style="list-style-type: none"> I/S Support Development

The Objective: Digital transformation of Enterprise

The client main objective was to transform the enterprise digitally and first step was to consolidate the data lake by enabling governance and providing access to different personas efficiently with appropriate security measures.



The Solution: Next gen data lake in AWS

Infosys teams collaborated with client in designing the solution to migrate the existing data lake layer.

Key project requirements were:

Data Scope

- Raw Layer (~100 Sources, ~10,000+ Data Sets)
 - Structured Data – RDBMS & Flat File Sources
 - Semi-Structured Data – XML Files, JSON Files
 - Unstructured Data – Audio WAV Files, Chat Transcripts, Images from P8
- Transform layer for 6 subject Areas(Personal lines, Claims, Customer etc.)
- Migration of deployed models and Apps
- Historical data migration of 60Tb

Processes

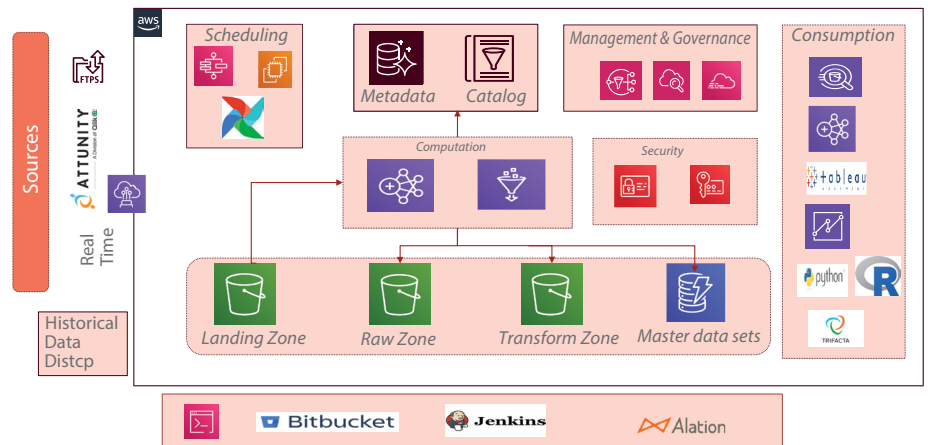
- Ingestion of data into Cloud (Land, Cleanse and Create Active / History data copies)
- Historical as well as Incremental data loads
- Metadata Delivery (Operational, Technical and Business Metadata)
- Checkout and Data Quality Processes
- Role Based Security and User Access

- High Availability and Disaster Recovery Processes

User Migration

- Data Scientists First; Business Users Later
- Migration of users with relevant data access to different groups

Identified as the AWS cloud partner for secure data migration and building pipelines for transforming the journey



Details of Solution:

- Delivered end to end cloud migration solution, starting from architecture, source analysis, ingestion and history data loads
- History data was migrated using combination of Hadoop dist. copy and ingestion framework
- Cost effective storage layers using Amazon S3 for Landing, Raw and Curated data
- Built metadata driven ingestion framework on EMR with all checks and balances features
- Data processing framework on GLUE and EMR
- Glue Data Catalog was used as centralized metadata repository and integrated with existing Data Governance tools
- Automated deployment of AWS Infrastructure using cloud formation
- Framework built for automated deployment of workflows and schedules



The outcome: Enhanced Analytical capabilities

Business Benefits:

- Enable self-service for Data Exploration by Data Scientists and Data Analyst communities
- Accelerate Predictive Modeling project lifecycle by reducing data acquisition and exploration time
- Reduce project cycle time by access to flexible compute infrastructure. Train / Retrain predictive models in less time by having access to on demand higher compute
- Single point of data access for both structured and unstructured data to uncover deeper and broader insights
- Timely access to data – access to on demand compute infrastructure assures on time data ingestion that meets SLA's

IT Benefits

- Agility and flexibility offered by cloud and resulting improvement in time to value for delivery of data solutions
- Enablement of emerging technologies such as Machine Learning and Artificial Intelligence available in the cloud and acceleration of their application to solve business problems
- Efficient time to market for infrastructure provisioning; scaling compute resources up and down on demand
- Separation of compute and storage and associated flexibility and lowering of infrastructure costs by 40%
- Ability to show (and potentially charge) business divisions the costs associated to their use of the infrastructure resources and data – such as data egress cost, compute and storage cost etc.

Quotes:

Appreciation from CEO

"As we embark on a Digital Transformation of our enterprise, data is going to be the fuel that will drive it. I was very impressed to hear the team was able to deliver a first release in just 6 months for the implementation of the Data Lake, that's fantastic. I know a lot of people have contributed to this project but your leadership and commitment made a very significant impact on the success of this project. I want to congratulate you and also thank you for your efforts.

Please convey my congratulations and appreciation to the rest of the team.

I am looking forward to seeing our Data Lake grow and drive our Digital Transformation."

Infosys Cobalt is a set of services, solutions and platforms for enterprises to accelerate their cloud journey. It offers over 14,000 cloud assets, over 200 industry cloud solution blueprints and a thriving community of cloud business and technology practitioners to drive increased business value. With Infosys Cobalt, regulatory and security compliance, along with technical and financial governance comes baked into every solution delivered.

For more information, contact askus@infosys.com



© 2021 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.