CREATING A SCALABLE PLATFORM TO ENABLE DATA SCIENCE, ML AND AI AND ENSURING COMPLIANCE

IMPACT

- Over 50 TB data ingested; analysed every week
- ML models which used to take up to 5 days now run in a matter of minutes or seconds
- Petabyte scale storage capacity
- Enabled predictive maintenance and continuous asset monitoring
- Optimized operations and supply chain
- Centralized data governance and compliance

A Fortune 500 company and the largest private sector corporation in India, Reliance Industries Limited (RIL) has businesses across the entire energy and materials value chain. RIL operates under different business segments namely exploration and production, petroleum refining and marketing, petrochemicals, textiles, retail, and telecommunications. They are now focused on building technology platforms that will create opportunities and avenues for India and all its citizens to realize their true potential.

The search for a single source of truth

Business teams across Reliance Industries processed and analysed information in silos, using disparate applications. "We lacked a single, common view of the data and there was no enterprise-wide data modelling in place. Each business function worked with data in pockets," said Deveshri Patel, General Manager – Digital transformation, Reliance Industries Limited. "As a result, every team had their own version of the data collected, which in many cases was out of date with limited traceability to its source. Disparate reports in traditional formats like excel existed, and we lacked a single source of truth across the organization. For instance, several business units and functions relied on older monolithic RDBMS based BI solutions and exported excel spreadsheets to analyse and make decisions on everything from crude sourcing, refining and supply chain operations."

Recognizing the need for an organization-wide data strategy and repository, RIL created a centralized Cloudera data lake on HDP to bring together and integrate data across various Oil to Chemicals (O2C) businesses. With the hydrocarbons data lake, provisioning access and availability to curated real-time business data improved significantly, from cycle times of several days to real-time data access with millisecond delays and zero data handling and processing errors. This enabled the roll out of several real-time business operation centres across RIL, significantly bettering business performance and lowering operational risk through increased process transparency and timely decision making.

A data backbone for Data Science, ML and AI

Considering the number of users, magnitude of use cases, volume, velocity and variety of data, RIL built an open source based integrated automation solution to cater to data acquisition, streaming, data engineering, data distribution, performance optimization, process mining, machine learning, data governance, data cataloguing, metadata management, data archival and data disposal. Leveraging process mining, RIL built a common data model which facilitated real-time, end-to-end data views of its processes and operations.

The Hydrocarbons data lake was a scalable, microservices-based, and containerized Kubernetes-based cloud architecture built on top of a Hadoop-based enterprise data lake. The shift to open source-based technologies formed the core of the Hydrocarbons cloud platforms and the data lake enabled RIL to integrate end to end business processes that cut across business and functional silos with real-time data provisioning and a common end-to-end data model which was independent of the source systems. It facilitated cross business and cross system workflows and operational insights and allows for the embedding of unique API based open source AI/ML functionalities which were not available in legacy IT solutions. RIL has now embedded an

30%

Savings in overall storage costs

"Cloudera has been one of the pillars that has helped us harness the power of our data to the extent where business managers and end users of the hydrocarbon data lake to make informed decisions basis near real-time data on business-critical information related to markets, crude prices and qualities, refining processes and decisions. We now have the ability to optimize consumer pricing of fuel and other petrochemical products that can be optimally produced within available capacity constraints."

Deveshri Patel, General Manager – Digital Transformation, Reliance Industries Limited

About Cloudera
At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.
Learn more at cloudera.com

enterprise data governance model which allows for organization-wide self-service information sharing, whilst still meeting data security and privacy requirements.

They can provide the common architectural building blocks using which future business use cases can be developed in an accelerated, agile approach with limited new developments.

Several new digital skills and capabilities, ranging from design thinking, domain-driven design, agile and secure DevOps to data engineering, visualization and ML Ops were developed and rolled out across the hydrocarbons business. These proved instrumental in transitioning the business and IT functions from a 'buy & configure' to a 'design & code' operating model with a shared cloud native architecture.

The hydrocarbons data lake has become RIL's data storage backbone for data visualization and data science capabilities for both structured and unstructured data. These capabilities have enabled the rollout of hundreds of business use cases with newer use cases deployed every day. These use cases range from procurement (e.g. demand forecasting prediction), to supply chain (e.g. freight & route optimization), trading (e.g. scenario modelling & sourcing optimizations), refining & marketing (e.g. IoT integrations, predictive maintenance, plant optimizations), project management (e.g. demand planning), petrochemicals (e.g. margin optimizations), finance, treasury, HR, corporate services and internal audit.

Unlocking the value of data
RIL utilizes Cloudera's HDP to provide end-to-end data solutions, cater to inbound and outbound data requirements across diverse data sources and destinations ranging from structured, unstructured, streaming and web data and support building data science solutions on enterprise data. From a business perspective the impact has been tremendous in unlocking the value of RIL's data and to support the transition of the hydrocarbons business to a truly integrated O2C business. The hydrocarbons data lake:

- Acts as a single source of curated, near real-time data for outbound data services for a variety of micro-service-based application databases, data analytics and data science solutions, data visualization engines for both structured and unstructured data for the end to end O2C business.

- Enables RIL to build reusable enterprise data models to support end-to-end data pipelines irrespective of source system, process, project or use-case. It has enabled the company with end-to-end data mining and data science capabilities to monetise enterprise data and optimise business processes across O2C.

- Different hydrocarbons businesses, Predictive maintenance, IoT, supply chain optimization, and functions have built their Operations Centres on top of the Hydrocarbons data lake as a common data foundation. In addition, end user self-service analytics capabilities have been rolled out to enable the end users to build unlimited number of metrics to explore data with limited IT intervention.

- Enables common O2C enterprise data governance, cataloguing, lineage management, archival and retention processes centralized with the Hydrocarbons data lake implementation which supports the transition to a cloud enabled microservices application architecture which enables: improved customer insight, reduced process cycle times and drive operational efficiencies, optimized assets and working capital.

- Machine learning models today leverage the data lake to provide business insights and predictions which used to take up to 5 days to generate and consolidate outputs in excel sheets for business managers to analyse and make decisions. Today, the same models run in a matter of minutes or seconds.

- Drives data storage cost efficiencies, as the cost per terabyte for extracting, storing, processing, managing and governing data has gone down considerably – with around 30% savings in overall storage costs by moving data load from existing tools used for data analytics.

CLOUDERA

Cloudera, Inc. 5470 Great America Pkwy, Santa Clara, CA 95054 USA. cloudera.com