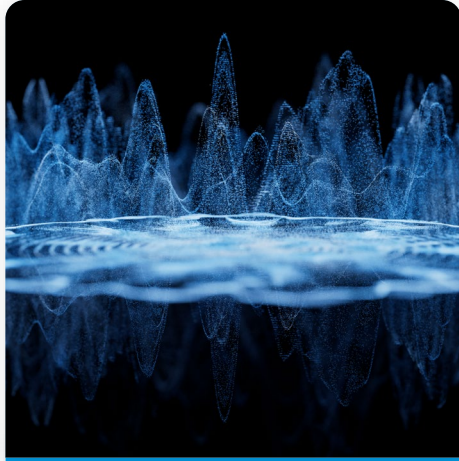


Creation of GxP Compliant Data Lake

Client Success Story



Industry

Pharmaceutical
(Human Pharma)

Client

One of the leading
international
pharmaceutical company

Challenges

Implementing modern cloud platform to build deploy lakes and create 2 data lakes on top of the platform.

- Limited access to data due to on prem solution
- Consolidate data from various commercial systems from manufacturing and supply chain areas
- Multiple solutions connecting to multiple data sources and therefore duplicating the data steams
- Very long data onboarding and consumption cycle
- Mix of different technologies required bigger number of operational teams, high SLAs

Solution

- Create highly configurable, metadata driven, scalable, high uptime data, distributed globally lake to ingest from over 250 different data sources
- Actively maintained and extended by C&F:
 - Integrating new data source
 - Cleansing, curating and refining ingested data
 - Monitoring availability and data quality
 - Integration with Colibra for data governance
 - Supporting business in daily work Technologies used:
 - AWS (Kubernetes, Spark, S3, Airflow)

Benefits

- Decrease the TCO spending on infrastructure by creating centralized lake
- Democratization of refined data stored in Data Lake in parquet standard
- Faster delivery and cheaper cycle thanks to semi-automated deployments
- Reduce Operational Costs, by having single support team across all data in Data Lake
- Improved decision-making accuracy by improving timeliness
- Enabler for autonomous supply chain planning