



Migration of Data Lake to Azure for one of the leading Fortune 500 Organization

Objectives

- Modernizing the Data Lake
- Bringing Data Lake close to source System and Reporting platforms
- Separation and isolation of workloads – Data Ingestion, Data Curation, Data Consumption, Analytical Sandboxes, Advanced Analytics, Analytical Data Extracts
- Increasing the Scalability -Scale up and Scale out compute resources independently based on the workload
- Provide additional capabilities - Advanced Analytics, Visual data Exploration



Solution

- Migration of all **major 21 domains**
- L0 integration
- Developed Robust Autoloader Framework for different source file types such as Parquet, CSV, TXT and Excel
- External Vendor files ingestion
- QDRM, Claims Workflow, MCG, Edge Server, ODS ingestion
- Data quality dashboard creation
- Creation of artifacts with detailed workflow name, frequency, AutoSys job id and source.
- Curation pipeline creation for Key domains such as Member, Claims, Provider and Auth
- Streamlining the current outbound data extraction that was created for various downstream from the Cloudera
- Implementation of BDR copy automation framework
- Creation of Secured view and data automation
- Reporting migration aligned to Azure Cloud Infrastructure



Benefits

- Successfully migrated **2000+ PowerBI reports**
- Converted **18,000** parquet files into delta tables
- Improved scalability by allowing compute resources to be scaled up or down independently based on the workload.
- **12,000 tables** were migrated from an on-premises system to Azure data bricks by using the BDR automation framework.
- Additional capabilities like Advanced Analytics and Visual Data Exploration are provided
- The BDR automation framework **reduces manual effort by 8 hours** to create and load 1500 tables whereas at the same time per person only can create and load 30 to 40 tables manually
- Automation of the data pipeline development process improves reusability, productivity and reduces costs.
- Ingestion Framework that provides, workflow orchestration, Time driven automation, workflow parallelize, resource providing, Data flow management, Scheduling, monitoring, visibility and Alerting



For more information, please visit www.infinite.com

