

D&A TRANSFORMATION TALES

# Streamlining Data Migration with **Data Lakes**



## BACKGROUND

Our client is a global leader in audit and advisory for top leading fortune 500 companies. They approached Aspire to streamline their data migration process from various sources such as files, DD, Stream to build a data warehouse in Azure Synapse within a specific time frame. The challenges they faced included:

- Redundant data from different divisions and company codes
- Lack of documentation support for various sources
- High data volume than expected for incremental load
- Inefficient current and history data management
- Inability to identify and classify data received for merge/overwrite/append
- Multitude of data ingestion tools utilized based on different input sources
- Inability to manage, identify and secure access from external resources



## BACKGROUND

To overcome these challenges our data and analytics experts suggested streamlining their data challenges with Data Lakes that would enable the client to:

- Manage a single version of truth for their reporting requirements
- Use Azure synapse to produce BI reporting and dashboards
- Appropriate handling of historical and current data
- Scheduled and monitored job loads
- Implement data governance, data catalog and track lineage

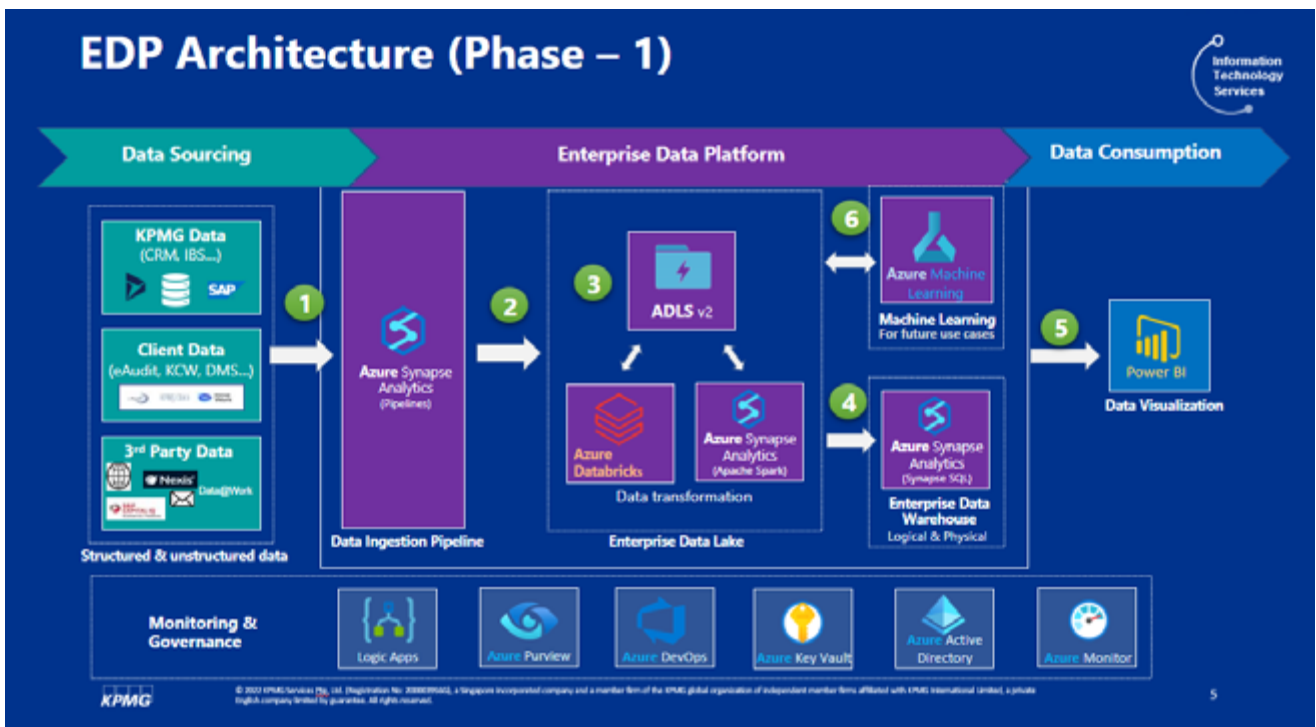




## THE SOLUTION

After taking the client's needs and challenges into consideration, Aspire's Data and Analytics team curated a well-crafted plan to build and deploy Data Lake that would help the client process, store, clean, share, analyze, model and monetize their datasets using BI and ML solutions. The solution involved:

- Synapse pipeline to load data from flat file and database to Azure Data Lake
- Azure Data Lake to store raw as well as delta data
- Synapse spark to process raw and delta data in data lake
- Generic/re-usable synapse spark notebooks
- Synapse spark Container workspace for ETL processing & writing data into data warehouse
- Azure synapse data to visualize BI reports
- ADF and Azure monitor to schedule and monitor daily /weekly/Monthly job load
- Microsoft Purview for data governance, data catalog & track lineage





## TECHNOLOGY SNAPSHOT

- Enterprise Cloud Platform – Azure Cloud Platform
- **Azure Durable Functions** for Orchestration and data ingestion
- **Enterprise storage** – Data Lake storage for raw data and processed data sets
- **Databricks** for in-memory data transformation and processing and all CRUD Operations
- **Enterprise DWH**– Azure Synapse Database
- Cloud Identity and secure - AAD, Azure key vault
- Cloud monitor - Azure Monitor
- **Microsoft Purview** - Data governance, data discovery, sensitive data classification, and end-to-end data lineage
- Azure Devops for CI/CD integration
- Logic apps for email triggering



## BUSINESS BENEFITS

- Involved heterogeneous sources to produce powerful reports in milli seconds.
- High return on investment with quality reports.
- Higher query performance and improved historical insight.
- Enhanced data quality and consistency.
- Increased the power and speed of data analytics
- Microsoft Purview reduced the need for manual and customised data discovery and classification





Aspire Systems is a global technology services firm serving as a trusted technology partner for our customers. We work with some of the world's most innovative enterprises and independent software vendors, helping them leverage technology and outsourcing in our specific areas of expertise. Our core philosophy of "Attention. Always." communicates our belief in lavishing care and attention on our customer and employees.

**For more info contact:** [info@aspire.com](mailto:info@aspire.com) or visit [www.aspiresys.com](http://www.aspiresys.com)

**USA**  
+ 1 630 368 0970

**SINGAPORE**  
+65 3163 3050

**INDIA**  
+91 44 6740 4000

**BELGIUM**  
+ 32 3 204 1942

**NETHERLANDS**  
+ 31 (0)30 800 92 16

**POLAND**  
+48 58 732 77 71

**MEXICO**  
+52 222 980 0115