Data Warehouse and BI Infrastructure Design and Implementation

Chain of healthcare facilities for mental and behavioral health

Built data lake and data warehouse capturing Slowly Changing Dimension type-4 on Azure platform to accommodate data from multiple data sources and Built a Power BI reporting suite by migrating existing dashboards from Excel to Power BI

Case study - Sage data integration

Data Warehouse and BI Infrastructure Design and Implementation

Situation

- Client had financial data stored on SAGE Intacct. Custom reports were built on the platform manually by the client and were used across the organization
- Partnered with client to set up automatic delivery of the data from SAGE INTACCT to Azure file share using Data Delivery Service (DDS), to get daily load of the files automatically

Accordion Value Add

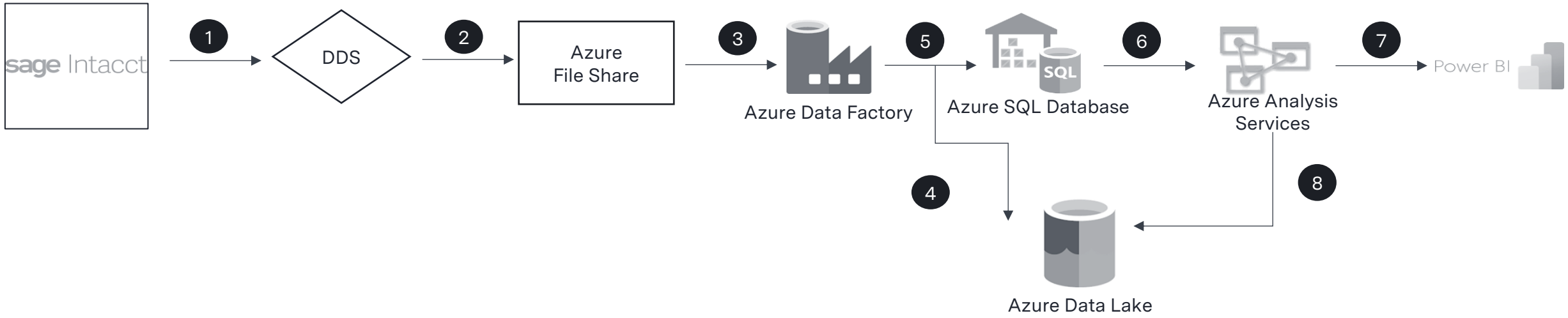
- Set up the automatic Data Delivery Service (DDS) in SAGE to get full load and incremental load of company and general ledger objects on daily basis
- Built 6 automated pipelines from SAGE to capture daily load (full load & incremental) of data with different file structures
- Took backup of daily files in Data lake for archival purpose
- Captured SCD type-4 while storing the data in data warehouse tables to track historic changes and recorded them in separate history tables

Impact

- The automated data warehouse architecture helped the client save time and efforts in creating, maintaining and running custom reports on SAGE
- Data from data warehouse can be used to create multiple financial reports with desired cuts for filtering than being limited by the source

Datawarehouse architecture design

Data Source | **Data Storage & Processing** | **Data Visualization**



1. Data is extracted from SAGE Intacct using automatic DDS (Data Delivery Service) in the form of CSV files
2. CSV files are saved in the specific folders for Full Load and Incremental files in Azure File Share
3. Data is extracted from Azure File Share using Azure data Factory pipelines and ingested into Azure Data lake and Azure SQL Database
4. Raw data is saved in Azure Data as flat files which serves as “Source” of raw data across the company
5. Data is ingested in Azure SQL Database for transformations & subsequent data modelling
6. Data is prepared for creating Fact & Dimension tables and relevant KPIs & metrics
7. Using the Fact & Dimension tables, final data models will be generated for visualization
8. Periodic data backups and data archiving will be done for the dimensional models from Azure Analysis Services in the Azure Data Lake.

Learnings

- There is a limitation for file upload size in Azure file share. If the file to be uploaded is **>4MB** in size, then the file will come as blank sheet and will not contain any data
- In order to resolve this issue, the files needed to be split after 10,000 records while uploading data from SAGE to Azure File share
- The Incremental Load from SAGE General Ledger object will have 3 different files that will contain newly created, deleted and updated entries in 3 different files having different file structure
- Schedule time for SAGE Data Delivery Service (DDS) trigger cannot be added automatically. The trigger will consider the timestamp of Data Delivery Service creation to be the automatic trigger time for daily runs