Azure cloud data warehouse implementation

Private equity firm

Built robust and scalable Azure cloud data warehouse by integrating source systems across all investment strategies (CITCO, IHS Markit, ExaVault, and SharePoint) into a central data repository. Also, deployed an automated reporting suite in Power BI for fund performance and portfolio monitoring.

Private equity firm needs a scalable cloud data warehouse

Picture this…

You're looking for a robust and scalable cloud data warehouse by integrating source systems across all investment strategies into a central data repository. Currently, you rely on front-end reports provided by individual source systems which are aggregated on offline Excel sheets to track fund performance.

You turn to Accordion.

We partner with your team to build a robust and scalable Azure cloud data warehouse by integrating source systems across all investment strategies into a central data repository, including:

1) Collaborating with the PE, Real Estate, and Credit portfolio operations and deal teams to understand and define the Fund and portfolio companies' performance metrics
2) Building pipelines to ingest data from source systems (Citco, Exavault, IHS Markit (S&P Global), and SharePoint Excel files) to create a single central repository of the data on Azure Synapse Data Warehouse. Implementing delta refresh via CDC to process only the changes, multiple times in a day.
3) Creating user friendly data models for custom report building and creating a single source of truth in the form of fact and dimension tables across functions
4) Building automated self-serve dashboards on Power BI for Fund Performance and portfolio monitoring to effectively manage the investments in real-time

Your value is enhanced.

You have a central automated repository with clean, sanitized and transformed data that helped in 80% reduction in person-hours needed for report preparation. You have reduced the ETL sync time by 66%, thereby providing nearly real time financials by multiple refreshes in a day. You also have better visibility into the fund and portfolio performance by creating holistic dashboards with a flexibility to stakeholders to extract essential data as per requirements from a central data repository.

AZURE CLOUD DATA WAREHOUSE IMPLEMENTATION

KEY RESULT

- ~80% reduction in manual effort
- ~66% of ETL sync time reduced

VALUE LEVERS PULLED

- Azure Cloud Data Warehouse
- Power BI Dashboards
- Fund performance
- Portfolio monitoring

Azure Data Warehouse implementation for a PE Firm with more than $40B AUM across PE, Real estate & Credit

AZURE CLOUD DATA WAREHOUSE IMPLEMENTATION

Situation

- Client relied on front-end reports provided by individual source systems and then aggregated them on offline Excel sheets to track fund performance. There was an opportunity in automating the reporting to provide the Partner's group better transparency and insights to business performance
- Partnered with the client to build robust and scalable cloud data warehouse by integrating source systems across all investment strategies into a central data repository

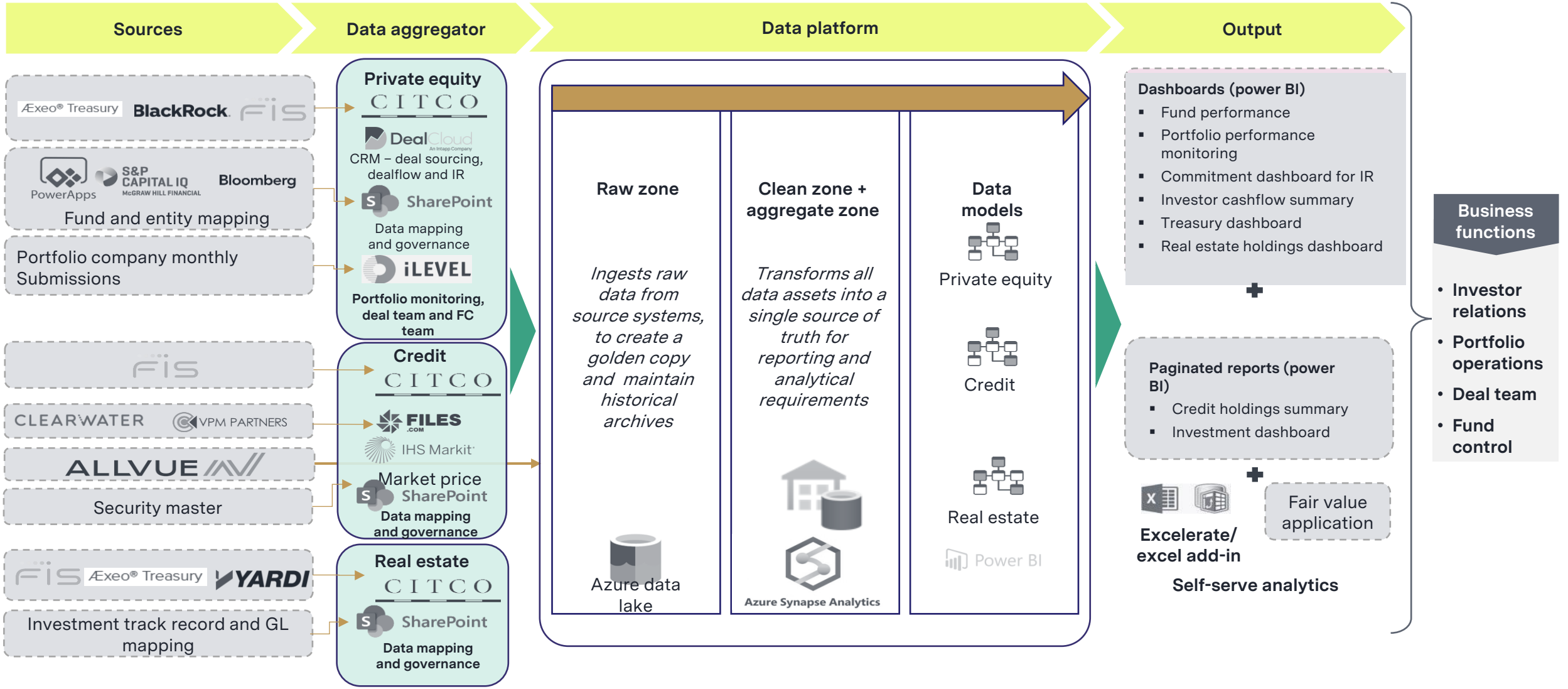
Accordion Value Add

- Collaborated with the PE, Real Estate, and Credit portfolio operations and deal teams to understand and define the Fund and Portfolio Companies' performance metrics
- Built pipelines to ingest data from source systems (Citco, Exavault, IHS Markit (S&P Global), and SharePoint Excel files) to create a single central repository of the data on Azure Synapse Data Warehouse. Implemented multiple times a day delta refresh via CDC to process only the changes
- Created user friendly data models for custom report building and created single source of truth in the form of fact and dimension tables across functions
- Built automated self-serve dashboards on Power BI for Fund Performance and Portfolio Monitoring to effectively manage their investments in real-time

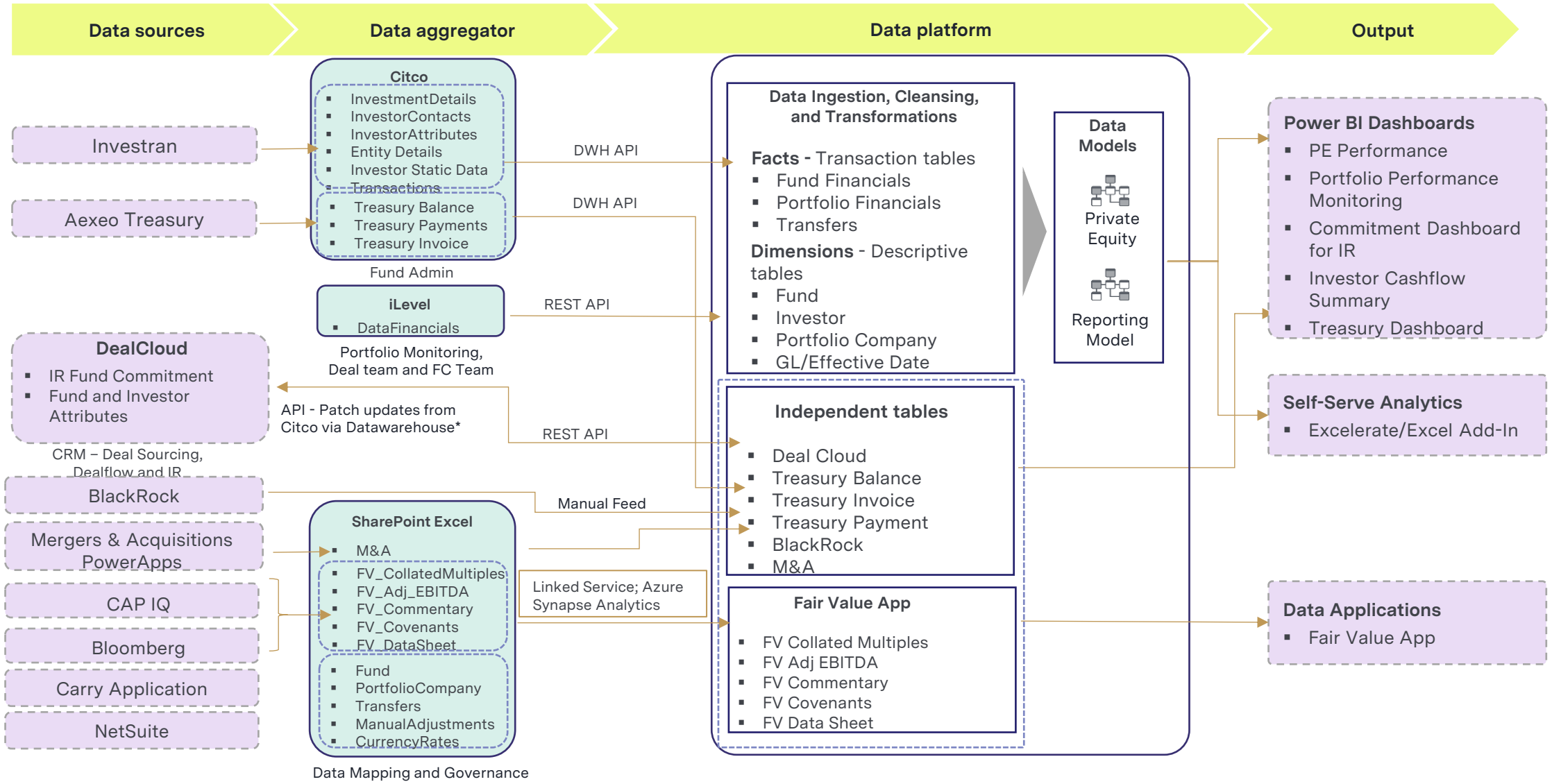
Impact

- The central automated repository with clean, sanitized and transformed data helped in 80% reduction in person-hours needed for report preparation
- Intra-day delta loads reduced the ETL sync time by 66%, thereby providing near real time financials refreshed multiple times a day
- Provided better visibility into the fund and portfolio performance by creating holistic dashboards and provided flexibility to stakeholders to extract essential data as per requirements from a central data repository

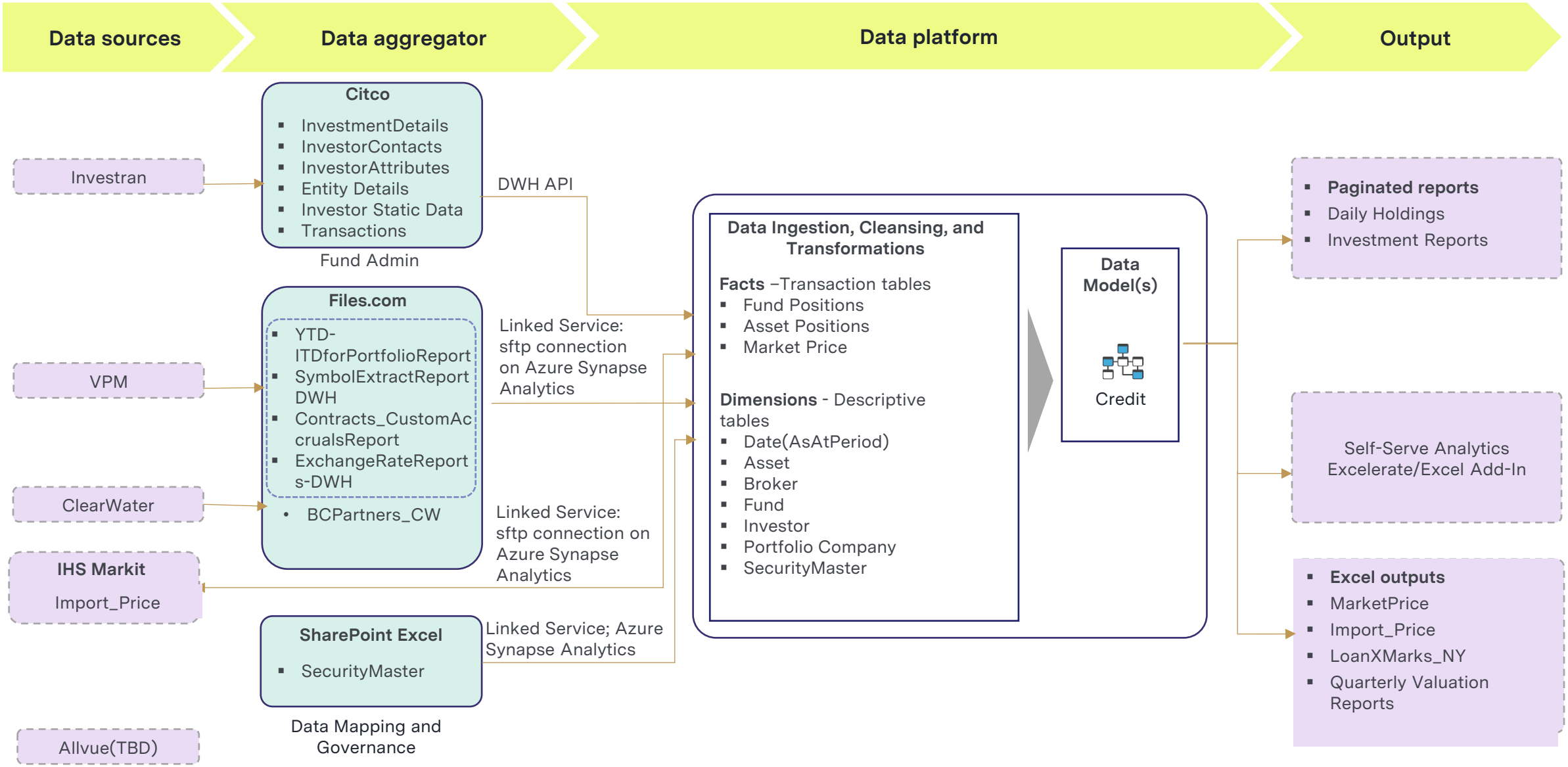
Data architecture diagram of the implemented infrastructure



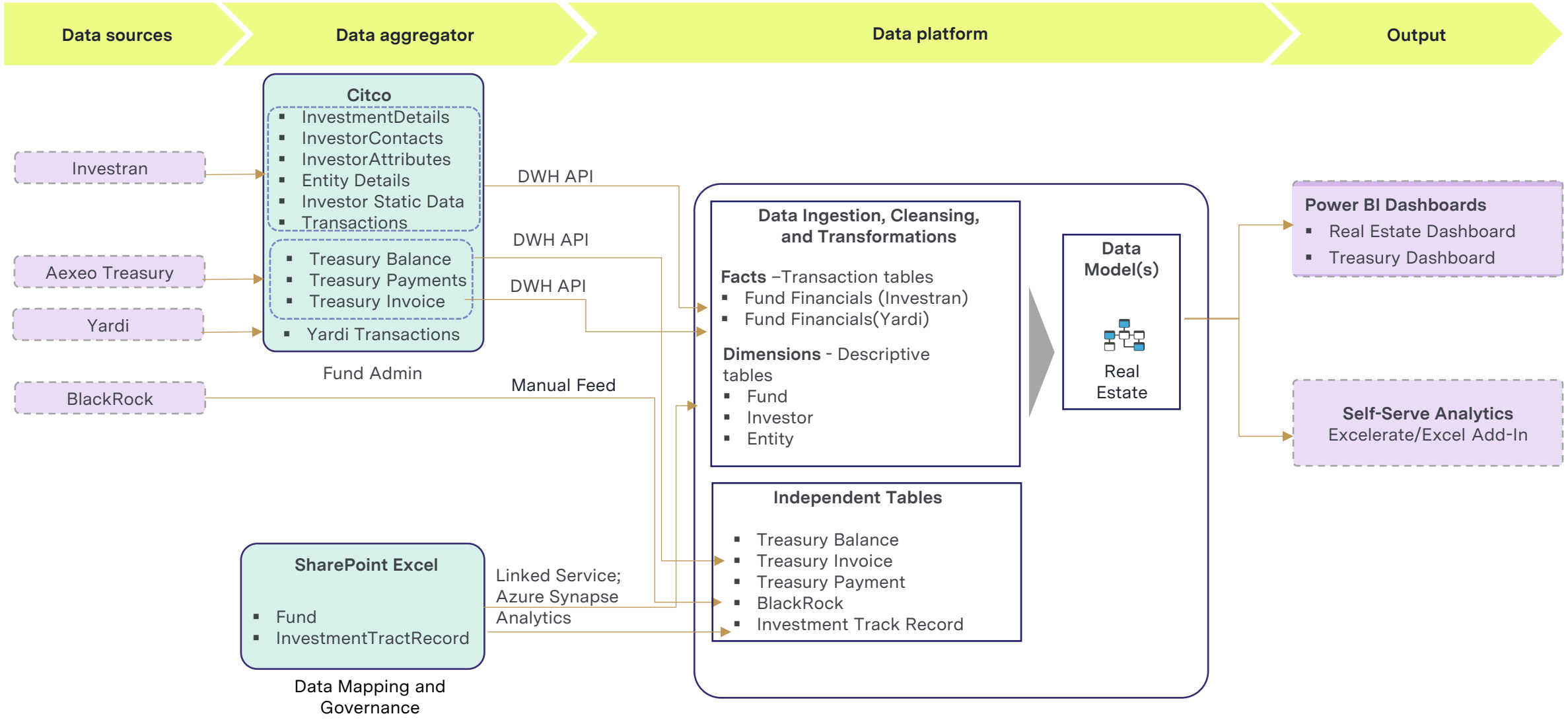
System architecture – Private equity



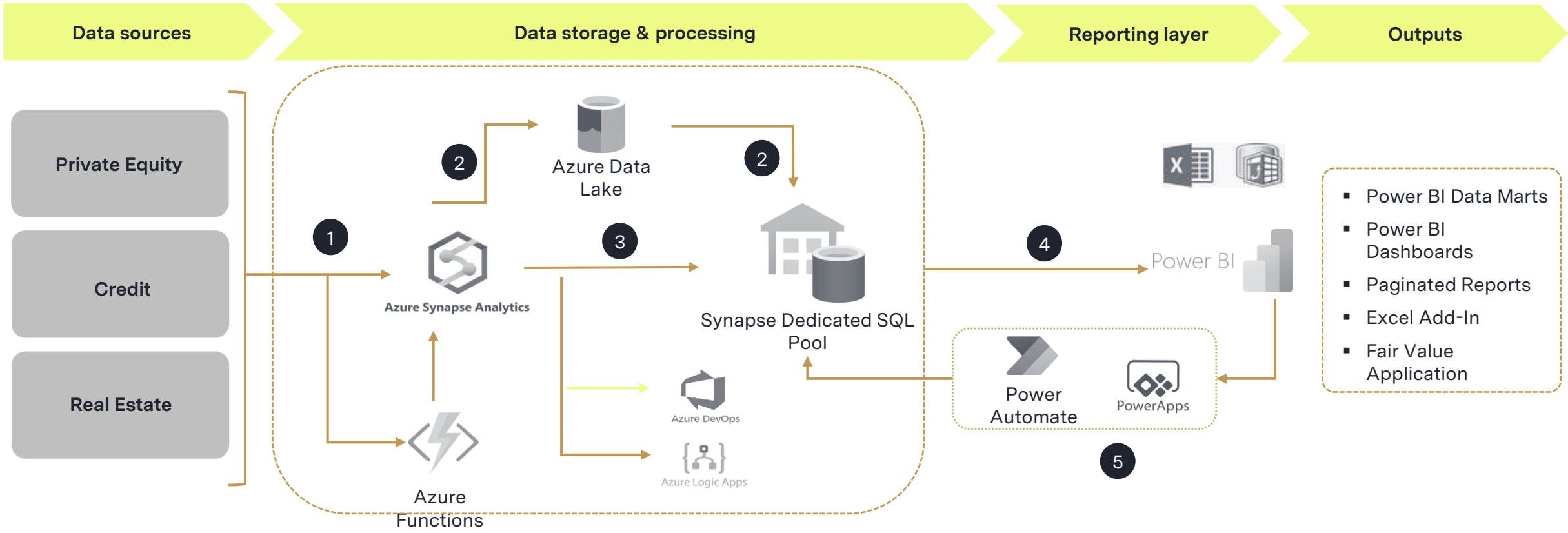
System architecture – Credit



System architecture – Real estate

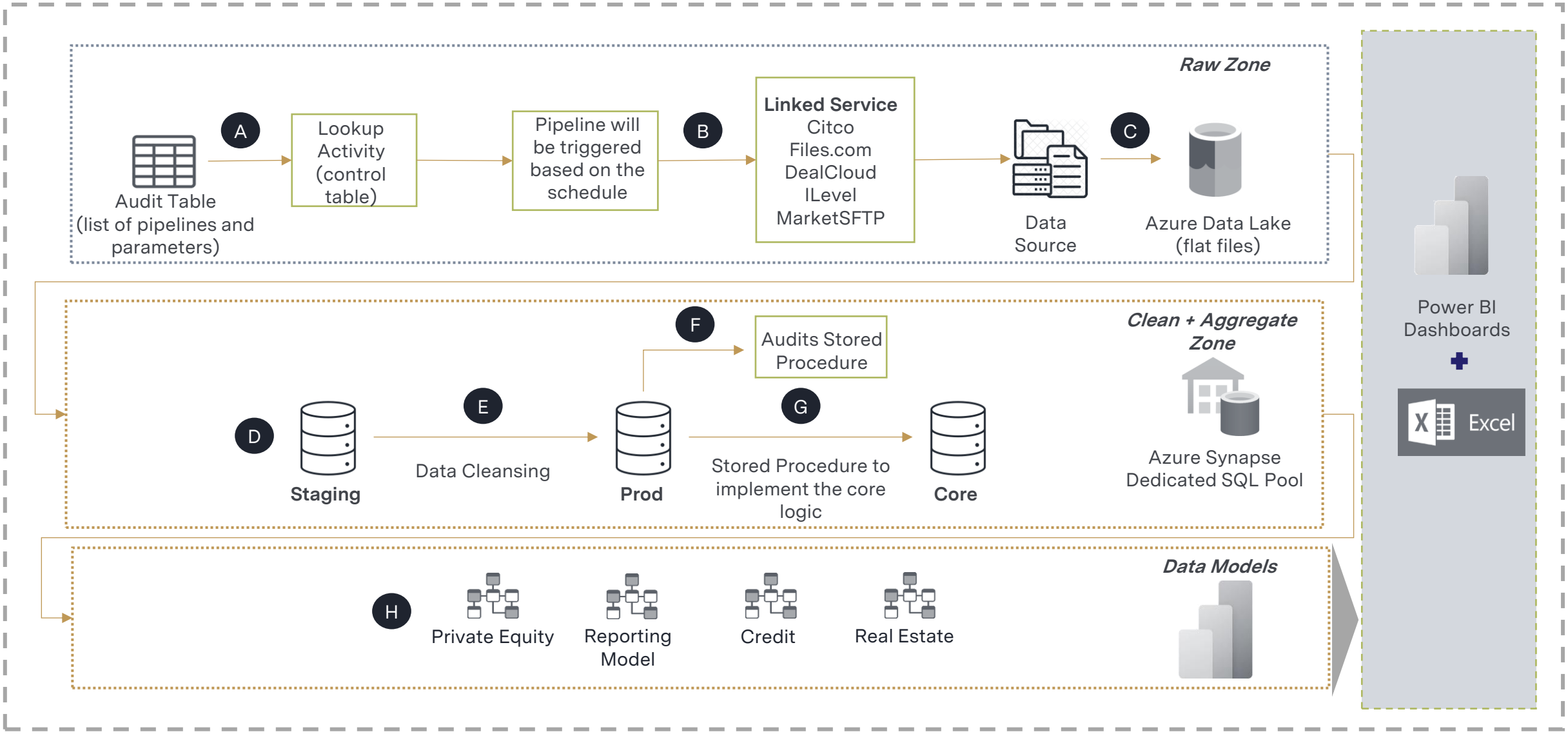


High level design



1. Data is extracted from source systems using Azure Synapse Analytics and ingested into Azure Data Lake and Synapse Dedicated SQL Pool
2. Raw data is first saved in Azure Data Lake as flat files which will then be ingested into staging tables on Synapse Dedicated SQL Pool
3. Staging tables in Azure Synapse for transformations & subsequent data modelling
4. Final reporting views from Azure Synapse Dedicated SQL Pool are uploaded into Power BI for visualization
5. User has flexibility to edit operational data. Subsequently, data is reprocessed back to the data warehouse to enable real time visualizations while ensuring data governance

Low level design (1/2)



Low level design (2/2)

A. Control table is a compilation of data pertaining to the dynamic content utilized in the dynamic synapse pipeline, output of lookup from this control table to be utilized further in the pipeline process.

B. Linked services for source are set up which is utilized by datasets in pipelines to run.

C. Data is ingested into the data lake in the form of flat files. Additionally, archives are maintained to safeguard historical data.

D. Copy data activity is used to absorb data into the staging tables in Synapse Dedicated SQL pool from flat files in the data lake.

E. Data from staging layer is moved to prod layer which can be an incremental load or full load based on the size and type of data we ingest from different source systems.

F. Lookup activities for source, staging, and prod audits, which ensures integrity of data after the cleaning process.

G. The logic to transform source data into dimensions and facts is executed, and an enterprise data model is created in the core layer.

H. Customized data models are developed for each business stream to enable end-users to connect and analyze data using Excel pivots. Additionally, Power BI dashboards are developed utilizing these data models to provide real-time insights of KPIs.