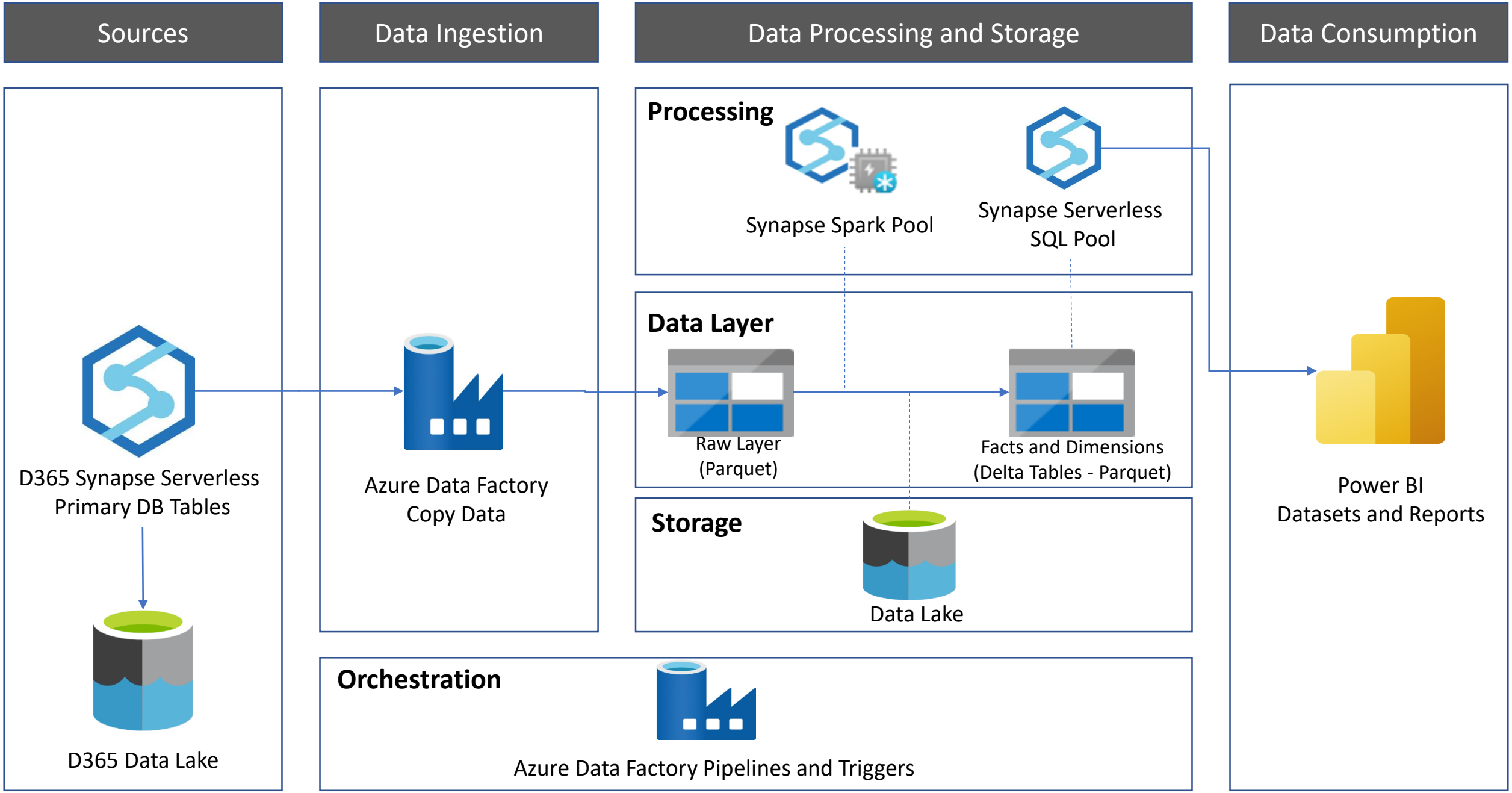
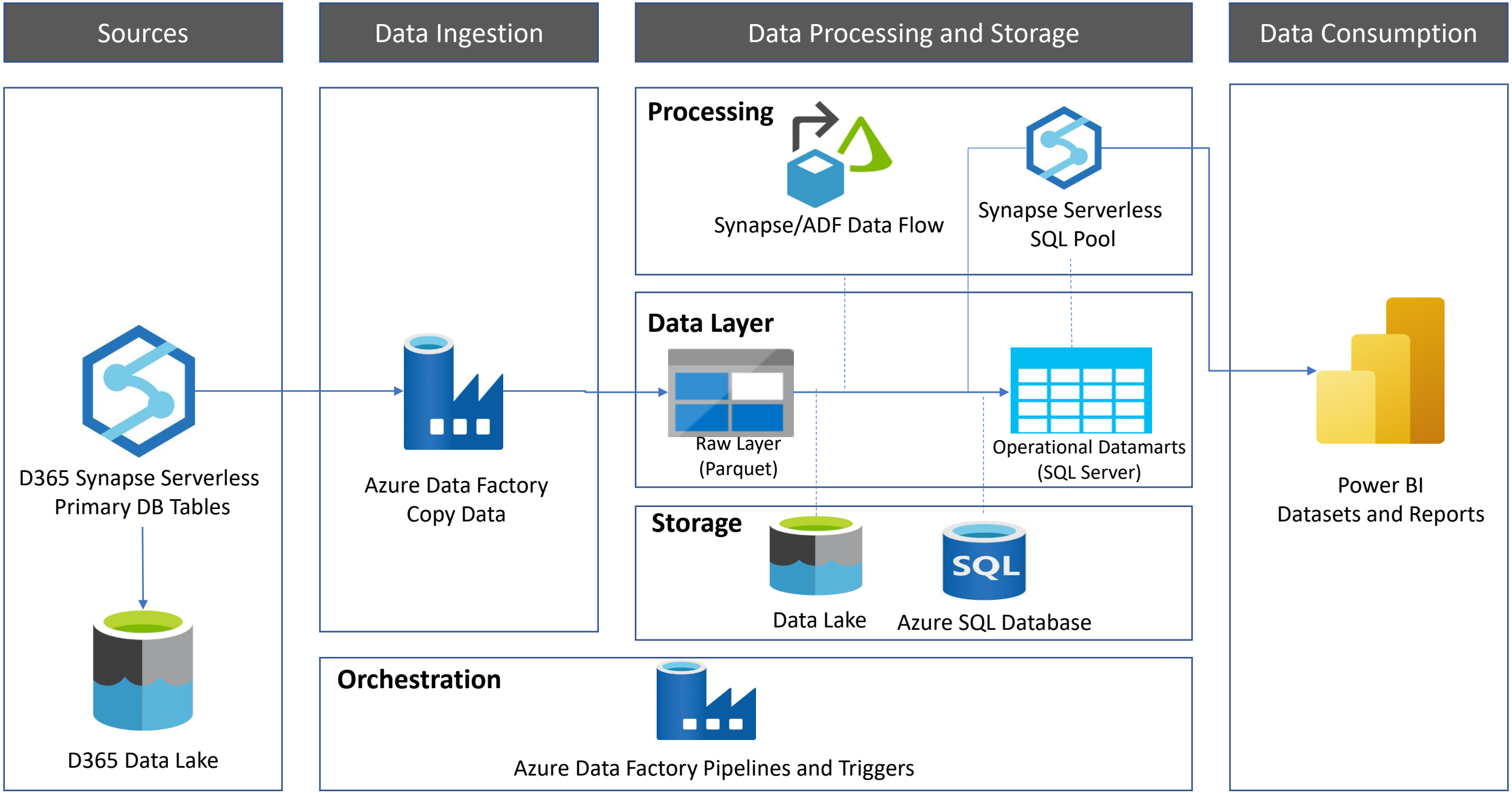
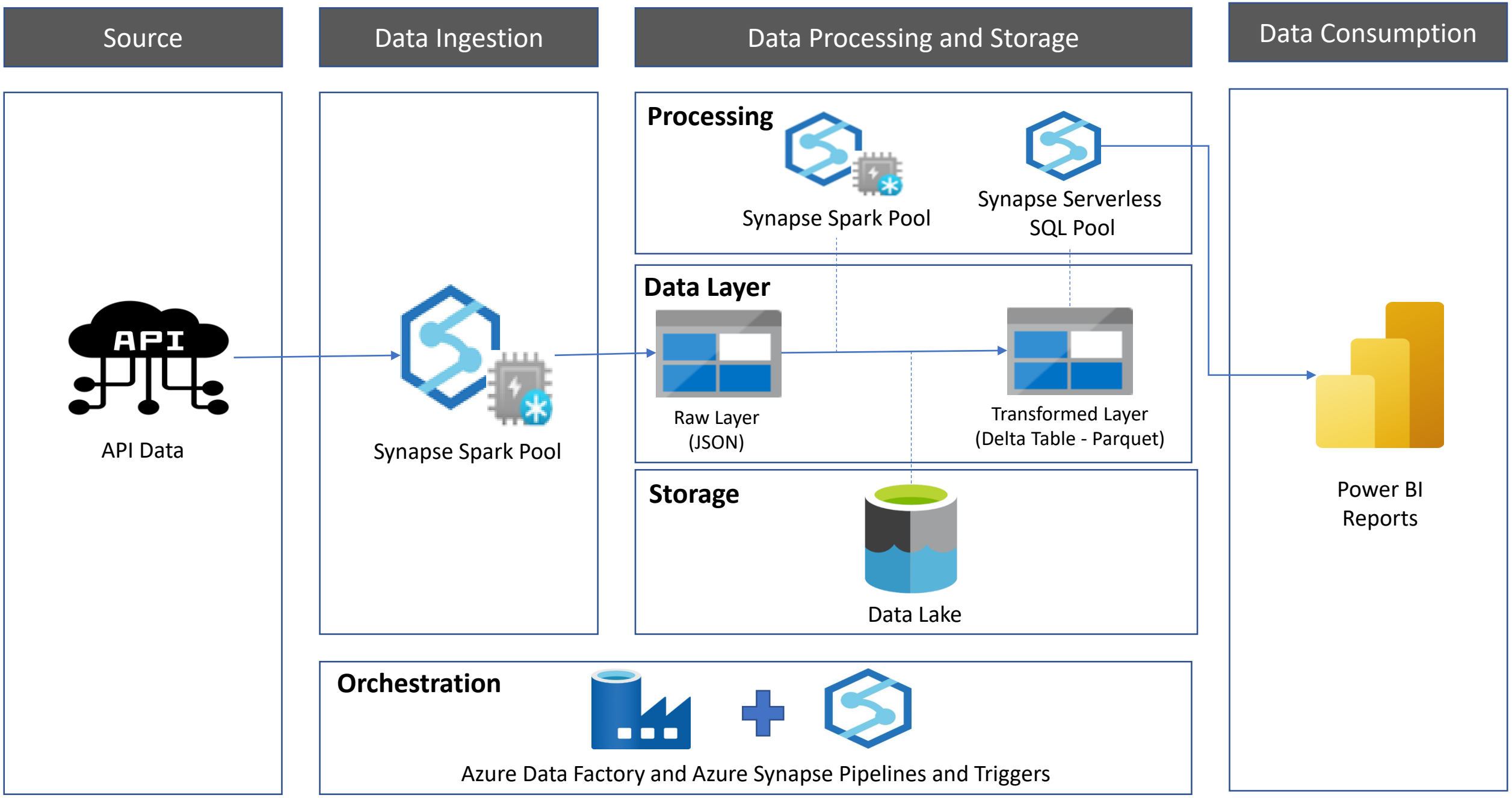
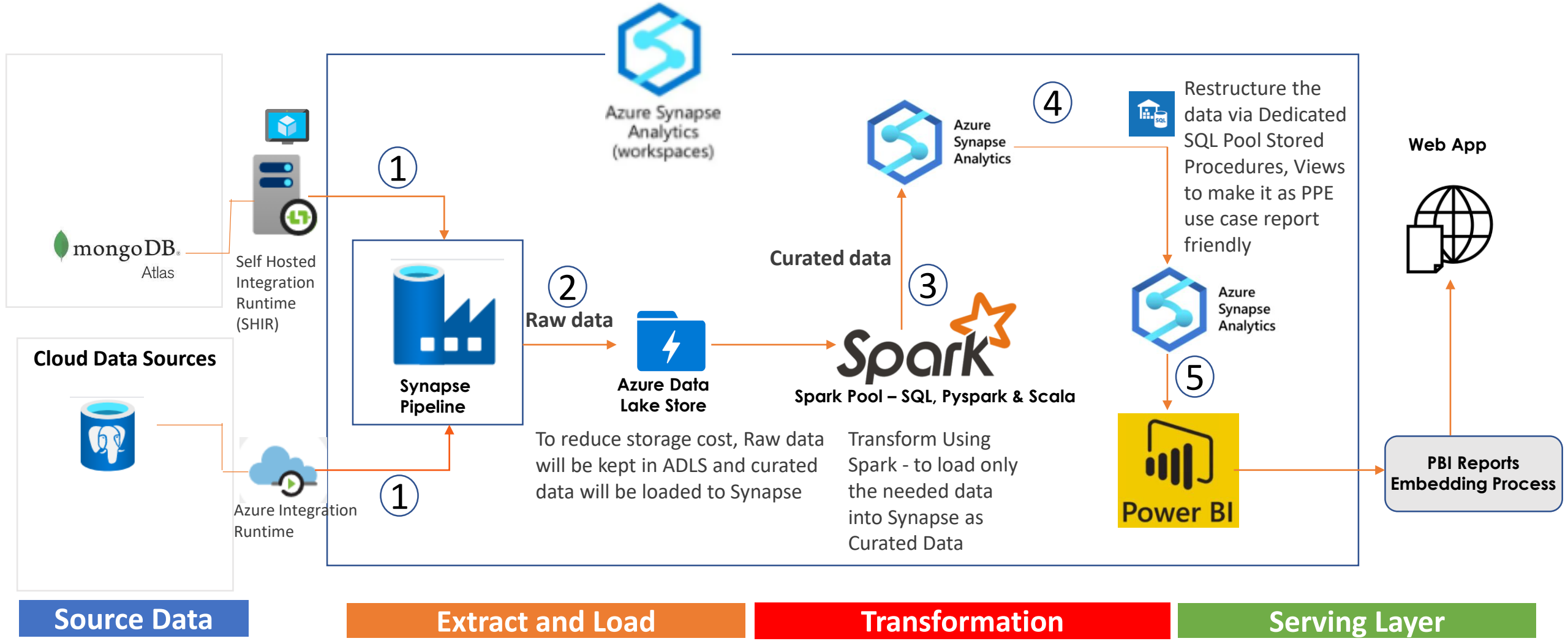

Sources
Data Ingestion
Data Processing and Storage
Data Consumption
Processing
Synapse Spark Pool
Synapse Serverless SQL Pool
Data Layer
Raw Layer (Parquet)
Facts and Dimensions (Delta Tables - Parquet)
Storage
Data Lake
D365 Synapse Serverless Primary DB Tables
D365 Data Lake
Azure Data Factory Copy Data
Power BI Datasets and Reports
Orchestration
Azure Data Factory Pipelines and Triggers

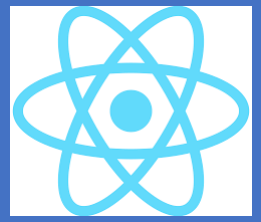

Sources
Data Ingestion
Data Processing and Storage
Data Consumption
Processing
Synapse/ADF Data Flow
Synapse Serverless SQL Pool
Data Layer
Raw Layer (Parquet)
Operational Datamarts (SQL Server)
Storage
Data Lake
Azure SQL Database
SQL
D365 Synapse Serverless Primary DB Tables
D365 Data Lake
Azure Data Factory Copy Data
Orchestration
Azure Data Factory Pipelines and Triggers
Power BI Datasets and Reports


Source
Data Ingestion
Data Processing and Storage
Data Consumption
API
API Data
Synapse Spark Pool
Processing
Synapse Spark Pool
Synapse Serverless SQL Pool
Data Layer
Raw Layer (JSON)
Transformed Layer (Delta Table - Parquet)
Storage
Data Lake
Power BI Reports
Orchestration
Azure Data Factory and Azure Synapse Pipelines and Triggers


Azure Synapse Analytics (workspaces)
mongoDB. Atlas
Self Hosted Integration Runtime (SHIR)
Cloud Data Sources
Azure Integration Runtime
1
1
Synapse Pipeline
2
Raw data
Azure Data Lake Store
To reduce storage cost, Raw data will be kept in ADLS and curated data will be loaded to Synapse
Curated data
3
Spark
Spark Pool – SQL, Pyspark & Scala
Transform Using Spark - to load only the needed data into Synapse as Curated Data
4
Azure Synapse Analytics
Restructure the data via Dedicated SQL Pool Stored Procedures, Views to make it as PPE use case report friendly
Azure Synapse Analytics
5
Power BI
Web App
PBI Reports Embedding Process
Source Data
Extract and Load
Transformation
Serving Layer

AWS EC2 Linux Server

Presentation Layer



React JS

Artifacts / Pages :

- Connections
- Projects
- Datasets
  - Profiling
  - Validation

Application Layer

Python

Flask - API

Capital One Profiler

Great Expectations

Database Layer

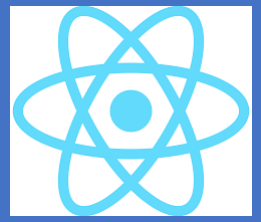
PostgreSQL

Metadata Tables:

- Connections
- Projects
- Datasets
- Dataset Details
- Testcase Master
- Testcase Details
- Testcase Result
- Profile Result

AWS EC2 Linux Server

Presentation Layer



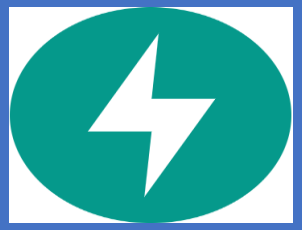
React JS

Artifacts / Pages :

- Connections
- Pipelines
  - Source
  - Sink
  - Transformation
  - Masking
  - Validation
- Migration
- Scheduling

Application Layer

Python



Fast - API

Snowpark

Great Expectations

Database Layer

Snowflake

Metadata Tables:

- Connections
- Applications
- Entity Master
- Attributes Master
- Testcase Master
- Testcase Details
- Testcase Result
- Query Transformation
- Masking Master
- Masking Details
- Schedule Master
- Schedule Logs