Execution Plan

1. Data Preparation

* Datasets: Upload your datasets (e.g., e-commerce website, mobile app, and POS data) into Azure Blob Storage as raw CSV files.
  + Create separate folders for each data source (e.g., /web\_data/, /app\_data/, /pos\_data/).
* Additional Data: Include supporting files like product details and customer feedback.

2. Azure Services Setup

Step 1: Azure Blob Storage

* Store raw datasets here.
* Use a hierarchical structure for organization:

perl

Copy code

raw\_data/

web/

app/

pos/

Step 2: Azure Data Lake Storage Gen2

* Set up this as your staging area for structured, semi-structured, and unstructured data after initial transformations.

Step 3: Azure SQL Database

* Store intermediate and cleaned data for operational analytics.
* Create tables for transactional and metadata storage.

Step 4: Azure Synapse Analytics

* Configure a Synapse SQL Pool for your data warehouse.
* Create fact and dimension tables for the reporting layer.

Step 5: Azure Data Factory

* Build pipelines for:
  + Extracting data from Azure Blob Storage.
  + Transforming data using Data Flows or Azure Databricks.
  + Loading processed data into Azure SQL Database and Synapse.

Step 6: Power BI

* Connect Power BI to Azure Synapse to create dashboards for insights like:
  + Sales trends.
  + Customer segmentation.
  + Feedback sentiment analysis.

3. Data Transformation Steps

1. Data Cleaning (via Azure Databricks or Data Flows):
   * Remove duplicates and irrelevant rows.
   * Standardize date formats and categorical fields.
   * Handle missing data via imputation or omission.
2. Data Normalization:
   * Normalize customer details (merge from different sources based on CustomerID).
   * Standardize product details across sources.
3. Data Integration:
   * Merge datasets from different sources (web, app, POS) into a unified schema.
4. Data Validation:
   * Check for anomalies and apply validation rules (e.g., order totals matching the sum of line items).

4. Architecture Design

Below is the high-level architecture for your project:

plaintext

Copy code

+---------------------------+

| E-commerce Website |

+---------------------------+

|

V

+----------------+ +---------------------------+

| Point-of-Sale |---------> | Mobile App |

+----------------+ +---------------------------+

|

V

+--------------------------+

| Azure Blob Storage |

+--------------------------+

|

V

+-------------------------------+

| Azure Data Factory (ADF) |

| Pipelines for Orchestration |

+-------------------------------+

|

+-----------------------------+-----------------------------+

| |

V V

+----------------+ +--------------------+

| Azure Databricks| | Data Flows (ADF) |

+----------------+ +--------------------+

| |

V V

+----------------+ +---------------------+

| Azure Data Lake| | Azure SQL Database |

| Storage Gen2 | +---------------------+

+----------------+

|

V

+---------------------------+

| Azure Synapse Analytics |

| (Centralized Warehouse) |

+---------------------------+

|

V

+----------------+

| Power BI |

+----------------+

5. Implementation Steps

1. Ingest Data:
   * Upload datasets to Azure Blob Storage.
   * Configure ADF pipelines for ingestion.
2. Transform Data:
   * Use Azure Databricks for complex transformations, like:
     + Deduplication.
     + Joining datasets.
     + Calculating derived fields.
   * Use ADF Data Flows for simpler transformations, like column renaming or data type conversion.
3. Store Transformed Data:
   * Load intermediate cleaned data into Azure SQL Database.
   * Load final transformed data into Synapse for reporting.
4. Build Data Model in Synapse:
   * Create star or snowflake schemas with fact and dimension tables.
   * Example:
     + Fact table: FactSales (fields: TransactionID, Date, Amount, CustomerID).
     + Dimension tables: DimCustomer, DimProduct, DimRegion.
5. Visualize in Power BI:
   * Connect Power BI to Synapse using a direct query.
   * Build interactive dashboards for KPIs:
     + Sales trends by region/product.
     + Customer segmentation (e.g., based on purchase frequency).
     + Sentiment analysis of customer feedback (if NLP is applied).

6. Testing and Deployment

* Test pipelines for each data source.
* Validate data accuracy post-transformation.
* Monitor performance and optimize Synapse queries if needed