**Design Solution:**

Diagram

Description automatically generated

We have 2 Data Sources:

1. Online Transactions
2. Point of Sale from 2000 stores across country

**Design**:

Components and data flow steps for ONLINE transactions:

1. I will create **Event Hub** to capture online transactions,
2. The output from **Azure Event Hub** will be sent to **Azure Stream Analytics** which will create a job that will run 24 hours
3. Output from **Azure Stream Analytics** will then be pushed to **Azure Function App**. This function app will be responsible for writing data to the **Redis Cache**.
4. From **Redis Cache** data will be synced to the **ADLS** periodically (every 1 hour (or sooner depending on cost and resource utilization).
5. Push data from **ADLS** to **Synapse** for further insight

Components and data flow steps for POS transactions:

1. I will create Event Hub with CAPTURE option. Events will be written to ADLs once this options is enabled.
2. From ADLS, I will configure Azure Data Factory (ADF) to read these events every 15 mins
3. Push data from **ADLS** to **Synapse** for further insight

Requirements and Solution:

1. Handle large write volume: Billions of write events per day.
   * From Event Hub and Azure Stream Analytics, I will be able to write data to ADLS periodically (every 15 mins to 1 hour).
2. Handle large read/query volume: Millions of merchants wish to gain insight into their business. Read/Query patterns are time-series related metrics.
   * To handle large read and query optimization, I will take help of ADF to copy data to my Staging Area, then use Databricks and Spark to Transform data. With the help of Delta Tables and Delta Lakes, I will be able to place data to a “Single Source of Truth” location i.e., in Synapse Analytics.
3. Provide metrics to customers with at most one-hour delay.
   * From Synapse Analytics, based on PowerBI reports or directly from Synapse I can publish metrics that can be accessed by the customers and merchants with fresh data information with the maximum delay of 1 hour (time from Online streaming data from Redis to ADLS)
4. Run with minimum downtime.
   * No downtime as Azure is always up and running. DR can be handled by making use of Azure storage replication feature
5. Have the ability to reprocess historical data in case of bugs in the processing logic.
   * All historical data will be present in the “single source of truth” location as well as ADLS. This data can be used at any given time to debug and resync. We should also define a periodic weekly or monthly archive process of Synapse and ADLS data.

Note: In this diagram, security concerns and other smaller details are excluded.