**HLD Document**

**Distributed Pizza Restaurant Management System**

**Scenario:**

A Pizza Restaurant chain “Pizza House” has more than 2000 stores across the country. Each store manages its own inventory of raw materials. Each store prepares pizzas, side dishes, etc. and sells them along with ready to eat products such as cookies, drinks, etc. The sale can happen by Point of Sale (POS) or Online. The online transactions would be flowing in real time whereas the transactions made by POS can be synced every 15 minutes in batches. They offer pick-up and deliveries by 3rd party providers.

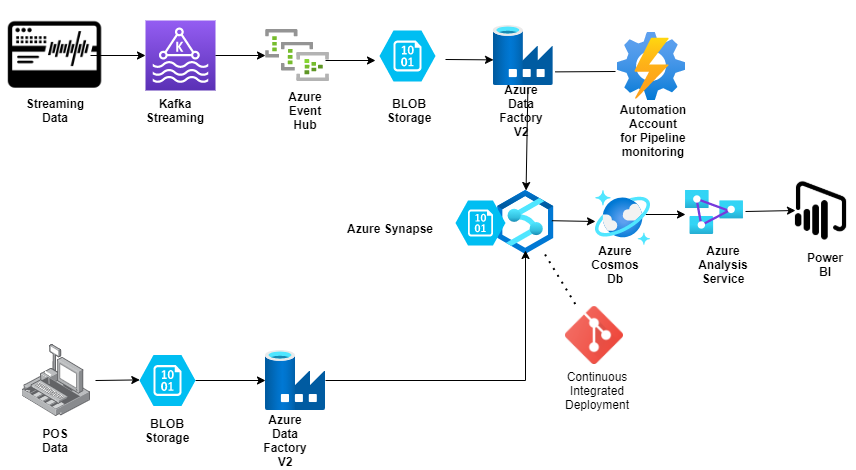
At the head office of the restaurant chain, management is concerned with the logistics of ordering, stocking and selling products while maximizing profits as well as understanding their marketing & communications. Several promotional schemes such as temporary price reductions, ads in newspapers, displays etc., also keep rising. Considering the huge data volumes (hundreds of GB per month) and the variety of the data they have; management wants the architecture to be robust enough to handle the varying data loads.

Design a cloud data platform to process and deliver insights based on the above. Please provide a high level solution design for the architecture. Feel free to choose any cloud provider you want.

**Major Considerations**

1. Handle large write volume: Billions of write events per day.
2. Handle large read/query volume: Millions of merchants wish to gain insight into their business. Read/Query patterns are time-series related metrics.
3. Provide metrics to customers with at most one hour delay.
4. Run with minimum downtime.
5. Have the ability to reprocess historical data in case of bugs in the processing logic.

**Process Flow Diagram**



**Explanation:**

POS Data:

Data is being loaded into Blob Storage and then moved to Datawarehouse using ADF. In the datawarehouse , it is being combined with streaming data to provide meaningful insights in form of Power BI reports.

Streaming Data:

Data is being consumed by Kafka topics as it can handle unlimited no. of topics simultaneously. Then, it is being processed one by one using Event Hubs. Finally, this data is being moved to the datawarehouse as well to be consumed by Power BI Reports

**Design Considerations:**

* Handling billions of writes per day through Kafka and ADF
* Simultaneous read and write using an scalable storage solution – Azure Blob Storage and finally using Azure Cosmos Db & Azure Analysis Service.
* Provide Metrics through Power BI Dashboards with minimal delay.
* No downtime as such.
* Event hubs and ADF have logics to make sure reprocessing in case of errors.