**Design Question**

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.

**Requirements**

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.

Graphical user interface, application

Description automatically generated

**Architecture Description:**

**Two Data Flow mechanisms have been defined:**

* POS- data processing (batch processing) , we can directly connect to source system using ADF and loading the data into SQL database on periodic time basis.
* For handling Millions/Billions of online orders, we should use kafka/EventHubs Publish/Subscribe messaging queue system in order to handle streaming data. we need a proper queuing mechanism otherwise there would be bottleneck in the system. We can use spark structure streaming for processing realtime data on batches on time basis and finally storing them into blob storage/ADLS/Cosmos.
* Finally both the data has been combined and consumed using Azure synapse and reports and dashboards are being created in Power BI to give the near real time view to the end users.

**Why Such a System has been designed:**

* The above system can handle billions of write events per day as we are using Kafka streaming and processing using Spark streaming is works on distributed, Scalable, Fault tolerant and in memory data processing engine.
* This system would give near to real time metrics to the customer as the data processing is being done continuously with minimal time lag.
* As this is based on cloud, so no maintenance window required as such and hence no downtime as such during the business hours.
* No issues of data reprocessing or no processing of data as the same is being monitored in the Azure Data Factory.

**Cost:**

* When we think about cost, Databricks wouldn’t be a good approach for processing and now a day they costlier. So, we must maintain Azure Kubernates or Confluent Kafka.
* Azure Synapse analytics is also costlier but we can get as package right(ADF, Spark and Power BI), we can maintain all resources cost at one go and also cheaper when compare induvial.