# 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

Handle large write volume: Billions of write events per day.

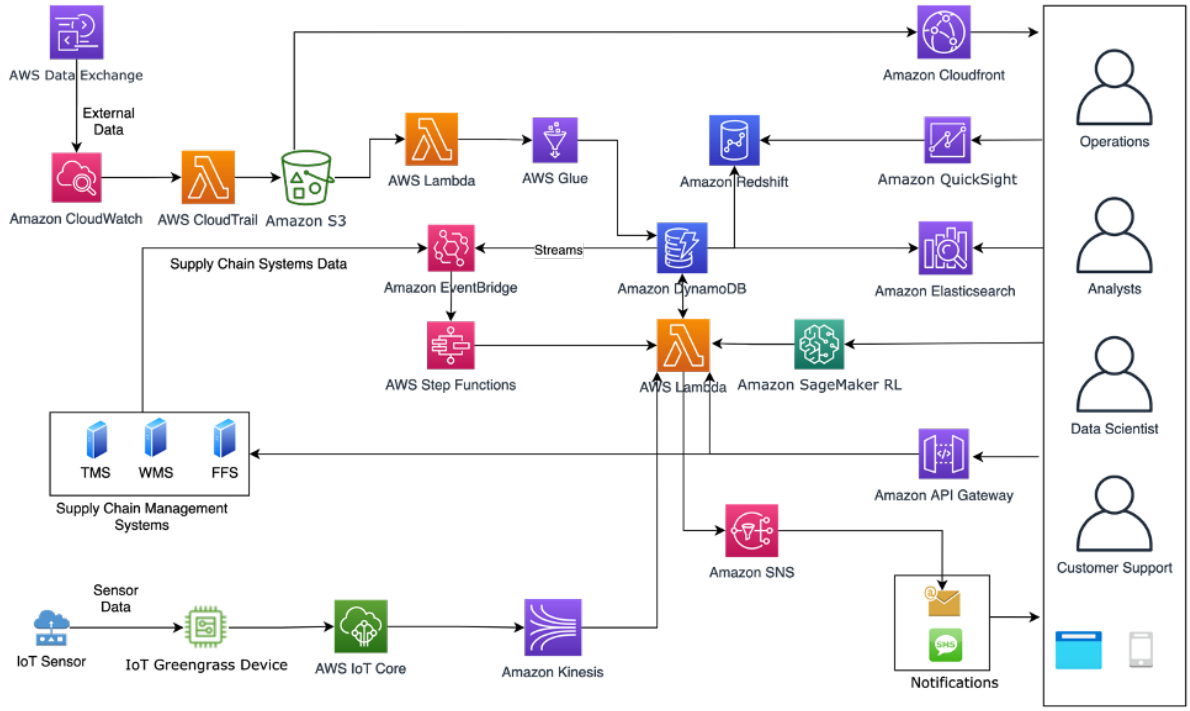
Handle large read/query volume: Millions of merchants wish to gain insight into their business. Read/Query patterns are time-series related metrics.

Provide metrics to customers with at most one hour delay.

Run with minimum downtime.

Have the ability to reprocess historical data in case of bugs in the processing logic.

# Architecture




Raw Data Collection -> streaming service / Data transfer -> perform Etl -> warehouse -> Reporting

**Raw Data Collection:**

It receives data from client applications. This should get high volumes of data and perform writes in fixed intervals and streams.

Here we have two types of data

1. Dynamic data:
    - Inventory data
    - Product based data
    - Media related data (advertisements, promos)
2. Static data:
    - Transactional data
    - Calendar data

For handling dynamic data, we need streaming service data pipeline where it includes some operations like upsert, insert, append, delete, vaccum.

Amazon kinesis can handle large streams of records in real time.

With Amazon Kinesis Data Streams, there are no servers to manage. The on-demand mode eliminates the need to provision or manage capacity required for running applications.

With Amazon Kinesis Data Streams, there are no servers to manage. The on-demand mode eliminates the need to provision or manage capacity required for running applications.

Use built-in integrations with other AWS services to create analytics, server less, and application integration solutions on AWS quickly.

**Implementation of data processing system:**

1. Data cleaning, validation, reformatting, etc.
2. Implement data analytics, ml model.

Here we can use last processed time, CDC, delta.

Raw data from on premises can be written based on events using data exchange and went through data processing system.

Monitor performance metrics and information on API calls, data latency, and error rates from the API Gateway dashboard, which allows you to visually monitor calls to your services using AWS Cloud watch.

Since we are using AWS as it is easy to scale. Based on the amount of data we can increase the cluster size add number of servers assign Iam roles.

Dynamic content delivery with a CDN. Amazon CloudFront is a fast content delivery network (CDN) service that securely delivers data, videos, applications.

Machine Downtime Monitor on AWS is a new solution that provides a dashboard to industrial customers to help them monitor the health of factory floor equipment, which enables easier identification of machine breakdowns and quicker response, improving asset availability.

Since we are using multiple data formats we can write data to different storages like s3, redshift, dynamodb

Amazon QuickSight allows everyone in your organization to understand your data by asking questions in natural language, exploring through interactive dashboards, or automatically looking for patterns and outliers powered by machine learning.