Cloud Architecture Structure:

A Step-by-Step Journey through Modern Design

By Ranjeet Singh

LinkedIn GitHub

Table of Contents

Introduction

Bombay spices, central to Mumbai's culinary tradition, are known for their vibrant flavours and rich aromas. Key spices like turmeric, cumin, coriander, and garam masala are essential in creating the unique and memorable tastes of Indian cuisine. These spices add depth and health benefits, making them indispensable in both traditional and modern cooking.

Objectives

- Real-Time Data Processing and Monitoring: - Implement real-time data ingestion and processing to track inventory levels, sales transactions, and customer interactions.

- Advanced Analytics and Machine Learning: - Utilize advanced analytics and machine learning to extract actionable insights from the data, such as predicting customer behaviour, optimizing inventory, and enhancing marketing strategies.

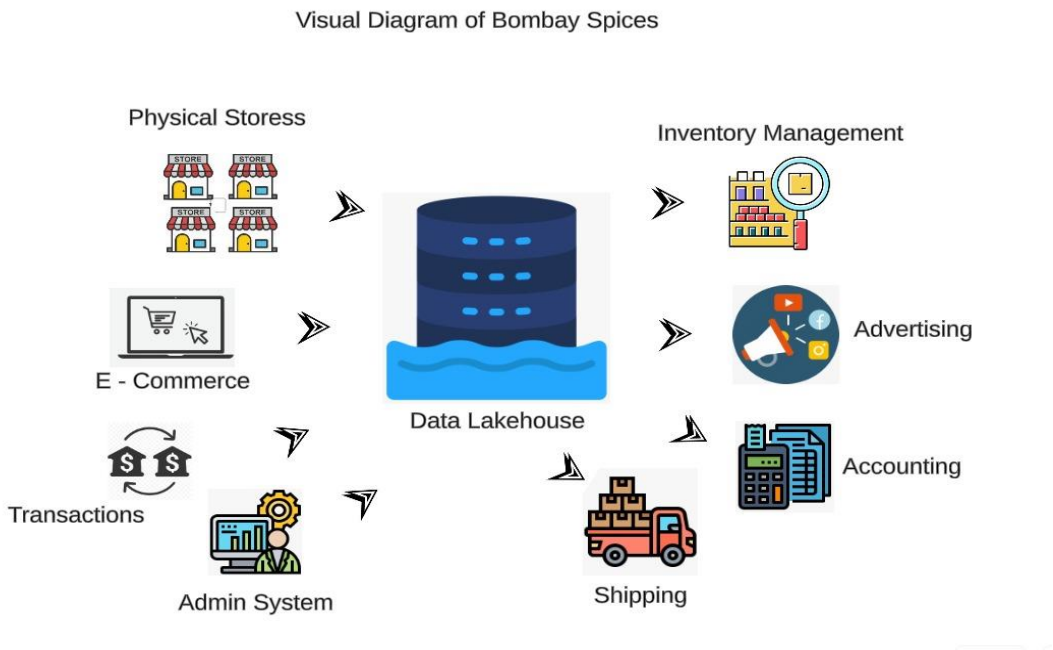
- Operational Efficiency and Automation: - Collecting, processing, analysing, and reporting data from online sales, in-store purchases, and inventory changes manually is timeconsuming and prone to errors. Operational efficiency and automation streamline these tasks, saving time and reducing errors.

Goals

- Build a scalable and efficient data processing and analytics platform.
- Integrate data from multiple sources for comprehensive analysis.
- Provide real-time insights to drive informed decision-making across the organization.

The Vision of the Cloud Architecture:

- From the outlook of the vision, it’s noticed that there are three main layers in the vision, which tell different information about the architecture.
- Data sources (First layer): In the first layer, data are collected from physical stores, E-commerce, Transaction (Every transaction related to store either online or offline) and Admin System. They are good sources of information to know about the business’ growth.



- Centralized Data Lake (Second layer): The centralized data lake is not just a storage solution but a strategic asset that empowers Bombay Spices to harness the full potential of its data. By providing a unified, scalable, and efficient platform for data management, the data lake is essential in driving the organization's vision of real-time insights, advanced analytics, and operational excellence.
- Advanced Analytics and End user consumption (Third and fourth layer): In Advanced analytics, everything has been done to clear and clean the data for insights by using different methods and techniques, Furthermore, in consumption layer, all good data is used for marketing level and other purposes.

Building the Architecture phase by phase:

Phase 1: Data Ingestion and ETL (Extract, Transform, Load)

The architecture begins with some data sources such as E-commerce, and structured data (like transactions and inventory).

E-commerce data stream is ingested using Event Hubs, which allows for high-throughput data ingestion, suitable for real time analytics.

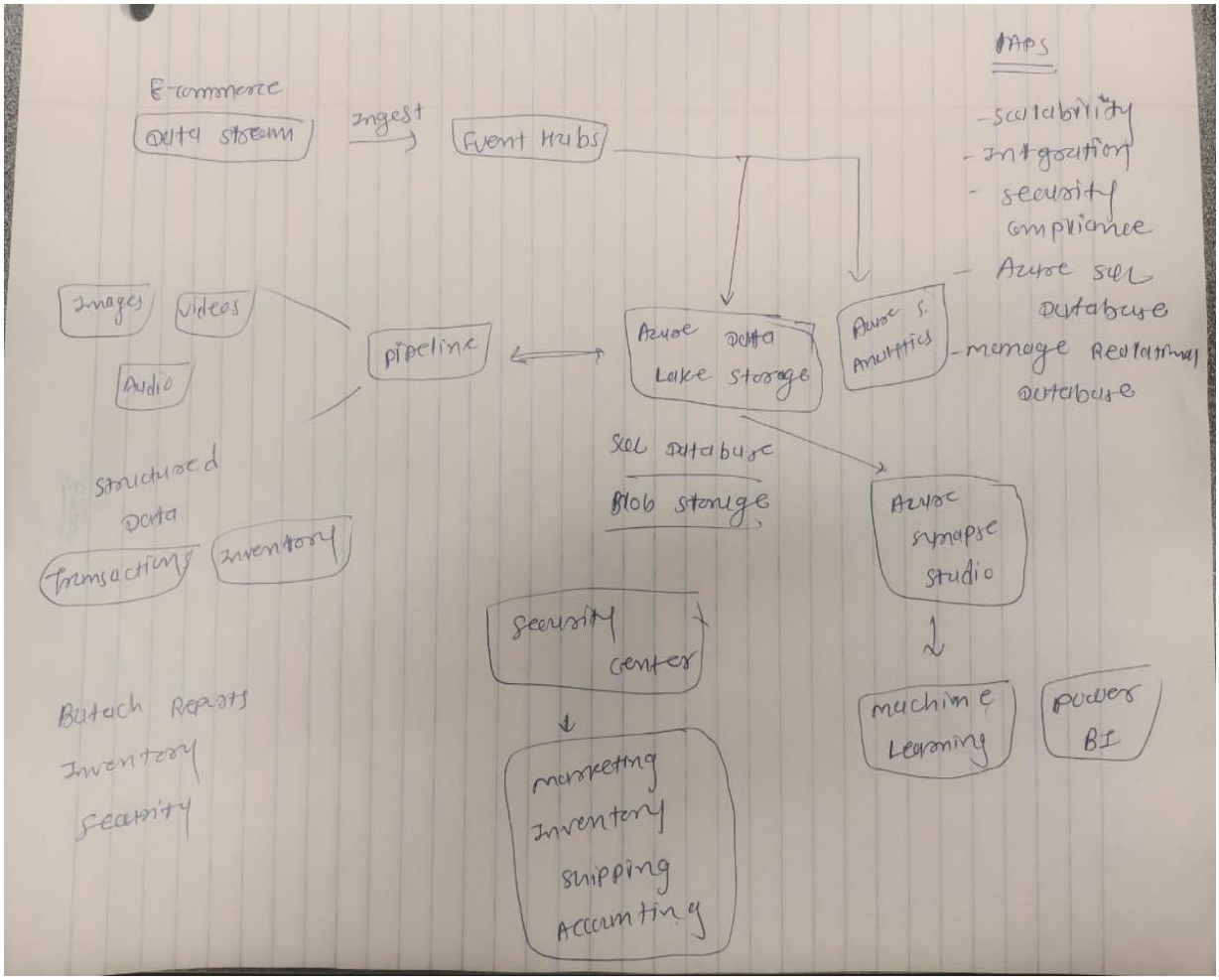
Azure Data Lakehouse: - I'm using Azure Data Lake Storage, which is designed to large volume of data. It's ideal for storing raw data, making it available for future processing and analysis. In Lakehouse, I used SQL database and blob storage to store the data.

Reporting and Predictive Analytics: - Azure Synapse Studio is utilized for orchestrating analytics workflows and integrating data with advanced processing techniques. Power Bi is used for business intelligence, enabling the visualization and reporting of insights gained from the data. This helps in making data-driven decisions across the organization.

The Transformation Begins: As I'm transitioned from phase 1 to final phase, the primary setup to reveal its limitations. During first phase, I didn't consider storing the traffic data which is coming from through social media and website. Second thing is that I considered the structured and unstructured data at the beginning, but it is not really needed, the reason is data lake can store everything without knowing the type of data. In the data Lakehouse, I didn't use bronze, silver and gold layer to clean the data for further processing. And I want to add one thing more is that giving security to our data is important but don't worry on this stage because when you work on cloud, they provide basic security to protect the data.

Let's go to final phase and check the changes I made after the first phase.

First phase of Architecture:



Final phase: - Optimization and Finalization

The final phase was dedicated to refining, optimizing, and finalizing the architecture to ensure it was both efficient and scalable for future growth.

Data Sources:

E-commerce (Streaming Data Sources):

- Data from the e-commerce platform, which could include customer interactions, transactions, and inventory data, is streamed in real-time.

Physical Stores:

- Data from physical stores, including transactions and inventory, is collected. This data might come from point-of-sale systems, admin systems, or other in-store data collection methods.

Data ingestion:

- Azure Event Hubs**:** This service allows for the real-time capture of event data and is designed to handle large volumes of data streams.
- Azure Data Factory: Data from physical stores is ingested using Azure Data Factory. This service is responsible for orchestrating and automating data movement and transformation, making it ideal for integrating various on-premises and cloud data sources.

Data Storage:

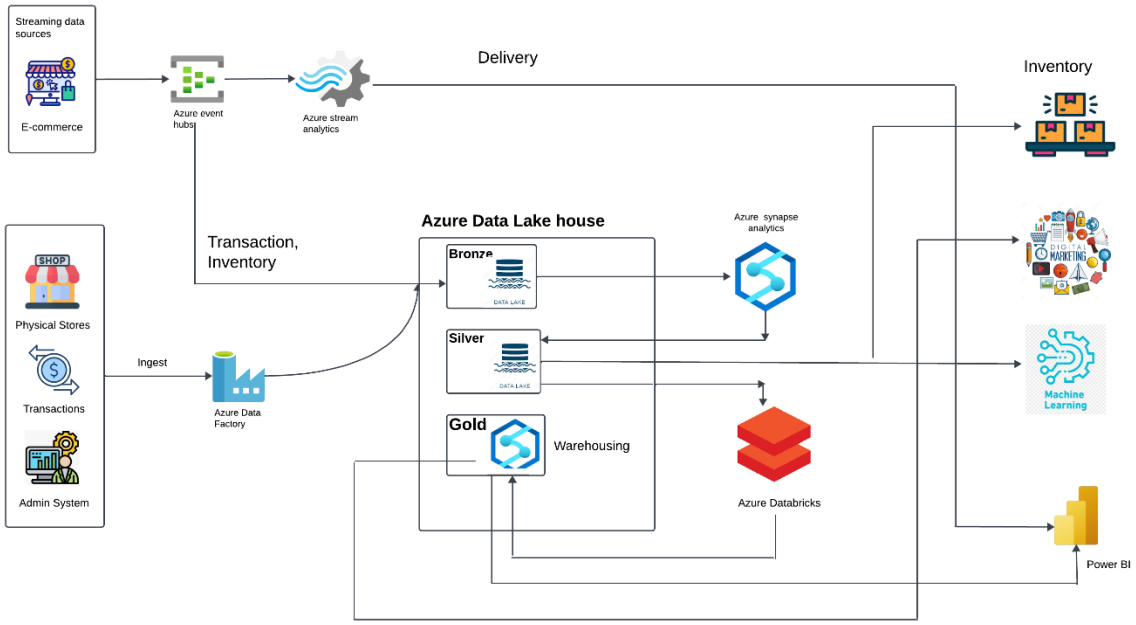
- **Azure Data Lakehouse (Bronze, Silver, Gold Layers):**
    - **Bronze Layer:** This is the raw data storage where data is ingested in its original format. It serves as the foundational layer for all ingested data.
    - **Silver Layer:** Here, the data is cleaned, transformed, and organized into a more structured format, making it ready for analysis.
    - **Gold Layer:** This layer is used for warehousing and storing highly refined, aggregated, and ready-to-use data for business applications and reporting.

Data Processing and Analytics:

- Azure Synapse Analytics:
    - This service allows for big data analytics and data warehousing, providing powerful tools for querying and analysing large datasets.
- Azure Databricks: This service is used for advanced data processing, including machine learning and deep analytics. It supports collaborative data science and is tightly integrated with Azure's data storage and processing services.

Consumer layer: After processing and all clear data send to different teams to utilize the data. From Silver layer data is directly share to machine learning team. With that data they

will make some predictive models and all data share to analysis team to find some useful insights which will help to grow the business. Advertising team will target the audience to grow the sales of business.



This architecture is a robust and scalable solution for managing and analysing large volumes of data from diverse sources. By leveraging Azure's suite of data processing, storage, and analytics tools, the architecture supports real-time decision-making, advanced analytics, and business intelligence, enabling the organization to stay competitive and responsive to market demands.

THANK YOU