Report
On
Steel Strips Cloud Architecture

-Mann Trivedi

Table Of Content

1. **Introduction:**

Manufacturing Cloud Architecture refers to the design and structure of a cloud-based infrastructure that supports and enhances the entire manufacturing process. It encompasses the integration of various digital tools, applications, and systems that work together to optimize operations, improve decision-making, and drive innovation within manufacturing organizations.

2. Mission and Objective :

Mission: To design and implement a scalable, robust, and integrated data architecture that seamlessly connects diverse data sources, processes real-time and batch data efficiently, and delivers actionable insights, empowering informed decision-making and operational excellence.

Objectives:

1. **Unified Data Integration:** Establish a centralized system to connect IoT devices, ERP systems, on-premises databases, streaming platforms, and web-scraped inputs, ensuring real-time and batch data ingestion.

2. **Scalability and Flexibility:** Design an architecture capable of handling growing data volumes and accommodating diverse data types (structured, semi-structured, and unstructured).

3. **Real-Time Data Processing:** Implement real-time data pipelines to ingest and analyze streaming data from sensors, energy consumption metrics, and workforce tracking.

4. **Data Accuracy and Consistency:** Use robust ETL/ELT processes and data validation techniques to ensure high data quality for inventory, production, and customer-related datasets.

5. **Automated Insights Delivery:** Enable timely decision-making through scheduled reports, dynamic dashboards, and automated alerts for critical business processes like production, inventory, and supplier evaluation.

6. **Interoperability:** Ensure seamless data sharing with external systems through standardized formats like CSV, Excel, or APIs, fostering collaboration across teams and systems.

7. **Performance Monitoring:** Implement tools for monitoring KPIs such as production efficiency, energy usage, and manpower utilization, driving continuous improvement.

3.Data Model:

Sources :

1. **IoT Integration:** Leverage IoT technologies to connect Customer Relationship Management (CRM) system. Automate the flow of data related to customer inquiries, sales orders, and feedback directly into the system. Enhance customer experience through real-time updates and actionable insights derived from IoT-enabled devices.

2. **On-Premises Data Loading :** Utilize Extract, Transform, Load (ETL) tools to import data from static sources, such as CSV file.Focus on loading critical datasets like inventory levels and product specifications. Ensure the accuracy and reliability of on-premises data by implementing validation checks during the ETL process.

3. **Streaming Data Integration:** Employ streaming platforms like Apache Kafka to ingest real-time data stream.Examples include monitoring energy consumption metrics, collecting sensor data from machinery, and tracking workforce activity.Real-time insights enable proactive decision-making and optimization of operations.

4. **ERP System Connectivity:** Establish direct integration with Enterprise Resource Planning (ERP) system.Access essential data such as Bills of Materials (BOM), production orders, operations sequences, and purchase entries.Streamline workflows by ensuring consistent and synchronized data between ERP systems and other platforms.

5. **Web Scrapping :** Use web forms and web scraping tools to gather external data, including market surveys and competitive intelligence. Facilitate manual data entry processes, especially for non-structured data like images or forms that require human input and validation.

Sink :

1. **Reports:** Generate scheduled and automated reports covering critical business areas such as:

▪ Dispatch details for logistics and supply chain management.

▪ Supplier evaluation metrics to assess vendor performance.

▪ Plant performance reports to analyze operational efficiency and productivity.

2. **Automated Notifications:** Set up automated alerts and notifications for key triggers, such as:

▪ Requests for quotations (RFQs) to suppliers or clients.

▪ Production alerts for stakeholders to act on significant updates or deviations.Ensure stakeholders remain informed without manual intervention, enhancing responsiveness.

3. **Dashboards:** Create interactive and visually appealing dashboards for real-time monitoring of key performance indicators (KPIs).Common KPIs include:

▪ Production Efficiency: Compare actual production with planned targets.

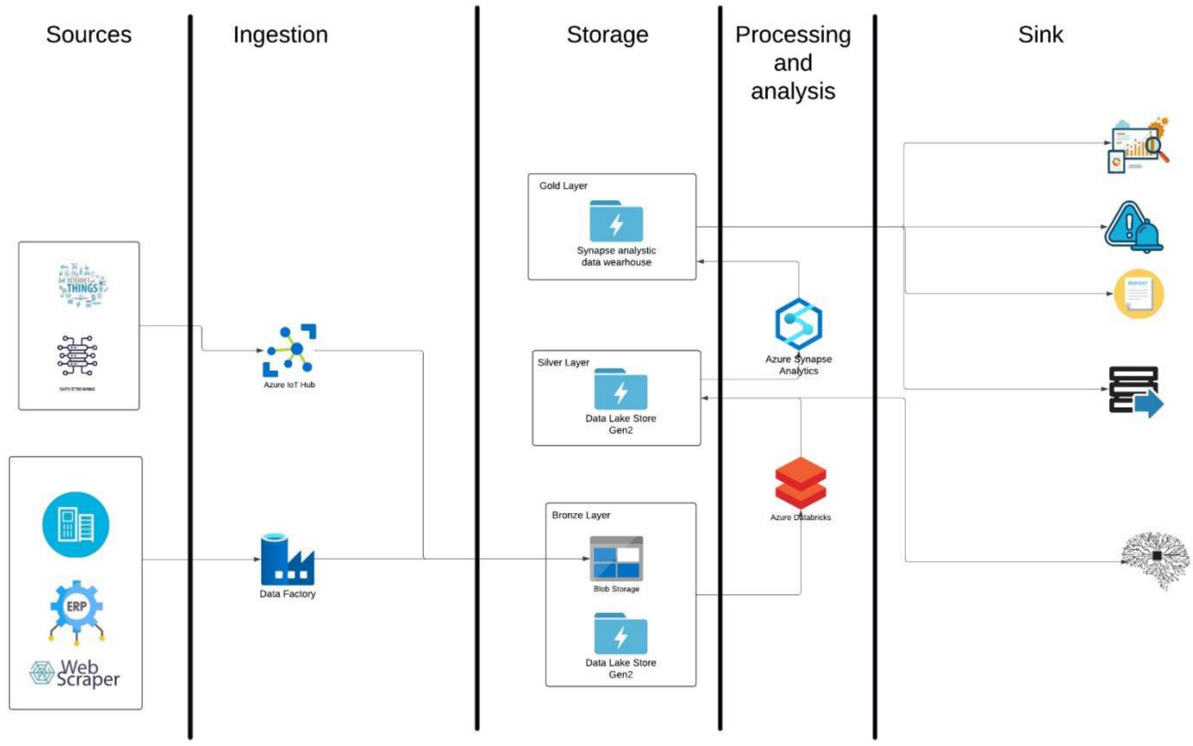
▪ Inventory Levels: Monitor stock availability to avoid overstocking or shortages.

▪ Energy Usage: Track energy consumption trends for cost and sustainability management.

▪ Manpower Utilization: Analyze workforce productivity to optimize labor allocation.

4. **Data Sharing:** Facilitate data export to external systems or platforms as needed.Supported formats include CSV files, Excel sheets, or via APIs for seamless integration.Ensure the processed data is readily available for analysis, reporting, or further processing in other applications.

4.Cloud Architecture:



Sources:

Data that is organized into defined fields like rows and columns, with a predefined schema.

On-Premises Data Loading: Often involves structured data from relational databases (e.g., SQL, Oracle).

ERP System Connectivity: ERP systems typically store highly structured data in predefined schemas.

Semi-Structured Data :

Data that has some organizational properties but does not fit neatly into tables (e.g., JSON, XML, sensor logs).

IoT Integration: IoT devices often generate semi-structured data like JSON messages or log files.

Streaming Data Integration: Streaming data may include a mix of structured and semi-structured formats like Kafka messages or telemetry data.

Ingestion :

Data that lacks a predefined structure, such as text, images, or videos.

Web Scraping and Manual Data Capture: Web scraping can yield unstructured data like free-form text or images from websites, while manual capture often involves unstructured notes or observations.

In the provided diagram, **Azure Data Factory** and **Event Hub** are used for **data ingestion** due to their unique roles in handling different types of data and workloads:

1. **Azure Data Factory**:

Purpose: Used for **batch data ingestion** from structured or semi-structured sources like databases, ERP systems, or web scrapers.

Reasons for Use:

Extract, Transform, Load (ETL/ELT): Automates the movement and transformation of large amounts of data from various sources to the data lake (Bronze Layer) or other storage layers.

Integration with Multiple Sources: Can pull data from databases, REST API s, SaaS applications, and file systems.

Data Orchestration: Allows you to define workflows and pipelines that control when, where, and how the data moves into the data lake.

It is handling data ingestion for **ERP systems**, **web scrapers**, or other similar structured data sources.

It moves this data into **Azure Blob Storage** or **Data Lake Store Gen2** in the **Bronze Layer**.

2. **Azure Event Hub**:

Purpose: Used for **real-time data ingestion** from high-velocity, unstructured sources like IoT devices or streaming applications.

Reasons for Use:

Event Streaming: Captures and ingests millions of events per second with low latency.

Scalability: Can handle large-scale, time-sensitive data from Internet of Things (IoT) devices, sensors, or other applications.

Integration with Azure Ecosystem: Works well with services like Azure Stream Analytic and Azure Functions to process events in real time.

It ingests **real-time streaming data** from IoT devices or data sources and places it in the **Bronze Layer** (Data Lake Gen2).

Event Hub focuses on real-time ingestion of high-velocity data (e.g., IoT and sensor data).

Bronze Layer:

1. Blob Storage :

Purpose: The Bronze Layer is often used as the **raw data ingestion layer** in a data lake house architecture. Blob Storage is cost-effective and designed for storing massive amounts of unstructured or semi-structured data.

Advantages:

Cost-effective storage: Blob Storage offers a cheap way to store raw, unprocessed data.

Scalability: It can handle large amounts of data from various sources, such as IoT devices, web scraping, and ERP systems.

Native integration: It integrates well with Azure Data Factory, Azure Data bricks,and Azure Synapse for further processing.

Silver Layer:

2. Data Lake Store Gen2: It is used for Bronze and Silver Layers

Why Data Lake Store Gen2?

Hierarchical namespace: Allows for improved performance and fine-grained access control over folders and files.

Unified storage: Supports both structured (CSV, JSON, Parquet) and unstructured data.

Big data processing: Optimized for use with big data frameworks such as Spark, Hive, and Databricks.

3. Synapse Analytic: It is used for the Gold Layer

Purpose: The Gold Layer is designed for **curated, aggregated, and ready- to-consume data** for analytic, reporting, and business intelligence.

Why Azure Synapse Analytic?

Data warehouse capabilities: Synapse is an enterprise-grade data warehouse optimized for structured, relational data.

High performance: It supports complex analytical queries with high speed using massive parallel processing (MPP).

Integration: Works seamlessly with Azure Data Lake and Data bricks, making it easy to query the Gold Layer data.

Built-in analytic: Synapse can run both on-demand and provisioned queries using SQL or Spark, reducing latency for downstream applications Like dashboards, predictive analytic, and reporting.

How They Work Together

Ingestion (Blob Storage): Raw data is collected and stored in the most cost effective manner.

Processing (Data Lake Gen2): Data is refined and structured through ETL/ELT pipelines. The hierarchical name space of Gen2 allows for better organization.

Consumption (Synapse Analytic): Transformed data is loaded into Synapse for business intelligence, machine learning, and advanced analytic.

1. Databricks :

Purpose: Databricks is a unified analytic platform built for big data processing and machine learning. It is particularly strong in handling unstructured and semi - structured data and providing tools for data scientists and engineers.

Uses in the Architecture:

Data Transformation: Used to clean, pre process, and transform raw data (bronze layer) into more structured and refined datasets (silver layer).

Data Science and Machine Learning: Provides a collaborative environment for data scientists to build machine learning models.

Scalability: Handles large-scale distributed data processing using Spark.

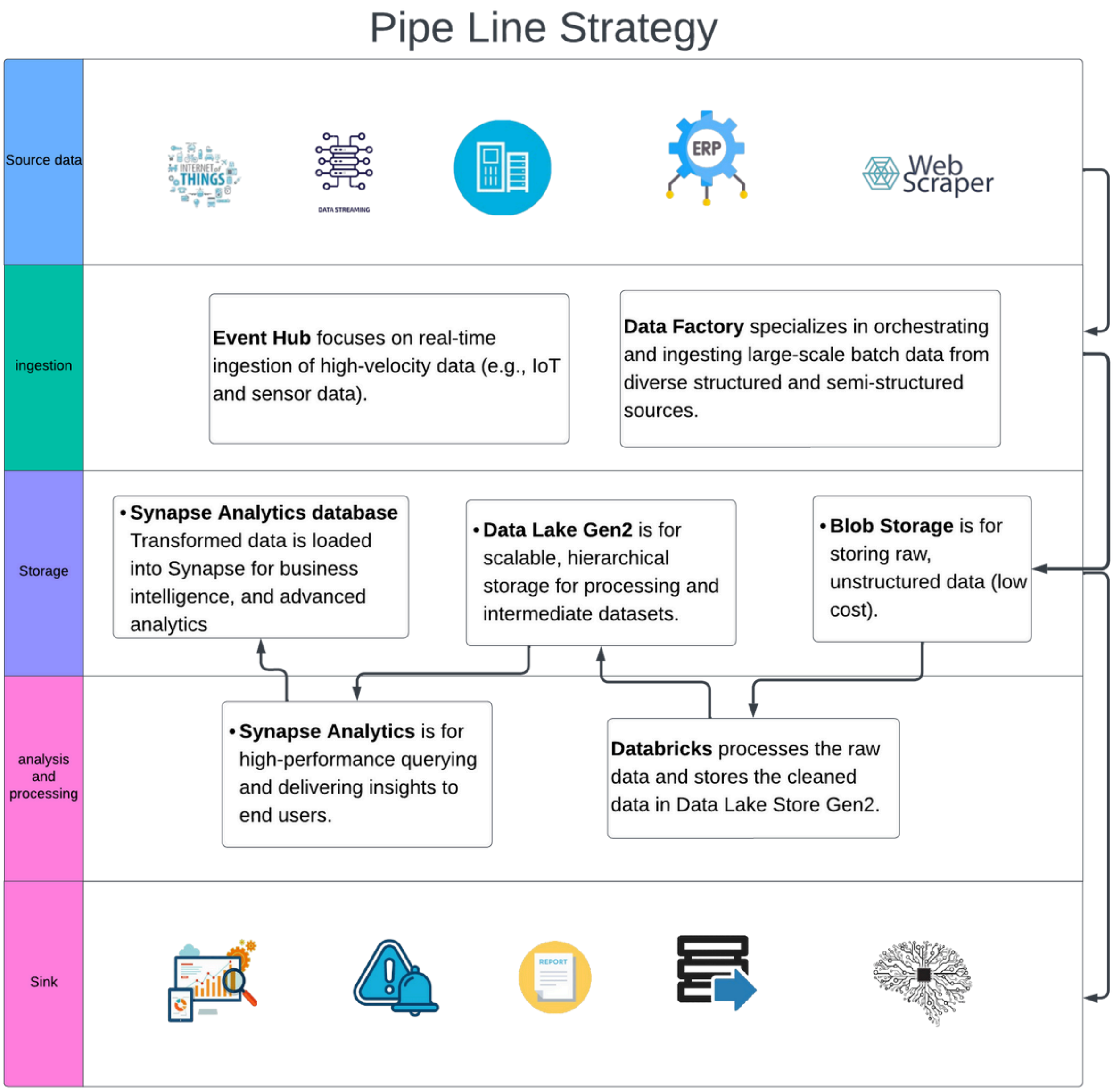
Notebook Integration: Offers a developer-friendly interface with notebooks for exploratory analysis.

Machine learning comes into play primarily during the Silver and Gold Layers, where:

In the **Silver Layer**, you can use ML models for pre processing steps like outlier detection, normalization, or feature extraction.

In the **Gold Layer**, ML is used to make predictions, detect patterns, or generate insights from the refined data.

5. Pipeline Strategy:



6. Conclusion:

This initiative establishes a scalable, robust data architecture, integrating diverse sources and enabling real-time processing. It ensures data accuracy, automated insights delivery, and seamless interoperability. By empowering decision-making through dynamic dashboards, reports, and alerts, the system enhances operational excellence, fosters collaboration, and supports continuous improvement across critical business functions.