**1.Create a system design diagram for event driven architecture which trigger an email survey as soon as a purchase transaction at Walmart.**

Flow:

* A customer completes a purchase transaction at a Walmart store or online platform.
* The point-of-sale (POS) system or e-commerce platform captures details of the transaction, including the items purchased, transaction ID, customer ID (if available), and timestamp.
* The transaction details are formatted into an event message, which includes relevant metadata and is published to a message broker or event streaming platform. This event serves as a trigger for downstream processes.
* Event processing services subscribe to the message broker and receive the transaction event in real-time. These services are responsible for analysing and interpreting the event data.
* A filtering mechanism may be applied to the incoming events to identify transactions that meet specific criteria for triggering the email survey. For example, only transactions above a certain value or involving specific products may qualify.
* If the transaction event meets the predefined criteria for triggering the survey (e.g., high-value purchase), the event processing service generates a command or event to initiate the survey process.
* Upon receiving the trigger event, a survey generation service dynamically creates an email survey tailored to the customer and their recent purchase. The survey may include questions about their shopping experience, satisfaction with products, and likelihood of recommending Walmart.
* The survey email is sent to the customer's email address associated with their Walmart account. The email delivery service handles the transmission of the survey email, ensuring it reaches the customer's inbox promptly.
* Customers who receive the survey email can respond by providing feedback through the survey form. Their responses are collected and recorded in a database or analytics platform.
* The survey response data is aggregated and analysed to gain insights into customer satisfaction, product preferences, and shopping trends. Reports and dashboards may be generated to visualize survey results and track key metrics over time.
* The event-driven architecture continuously monitors transaction events and survey responses to adapt and optimize the survey process. This may involve refining trigger conditions, adjusting survey content, or incorporating feedback to enhance the customer experience.

**2.Capture all the new data terms from this class and create a data terms catalog.**

1. **Event-Driven Architecture (EDA):** Event-Driven Architecture is a design pattern where the production, detection, consumption, and reaction to events are central to the architecture. In EDA, components communicate through events, which represent significant occurrences or state changes within the system. This approach enables loosely coupled and scalable systems, as components can react to events asynchronously, decoupling producers and consumers.
2. **Latency:** Latency refers to the time delay between the initiation of a process and its completion. In the context of systems and networks, latency often refers to the time taken for data to travel from its source to its destination. High latency can impact system performance and user experience, especially in real-time or interactive applications where responsiveness is critical.
3. **Versioning:** Versioning refers to the practice of assigning unique identifiers or labels to different versions of software, data, or other artifacts. It allows for the management and tracking of changes over time, facilitating rollback to previous versions, comparison between versions, and ensuring compatibility between different versions. Versioning is crucial in maintaining consistency and reliability in distributed systems and APIs.
4. **Veracity:** Veracity refers to the accuracy, reliability, and trustworthiness of data. In the context of big data and analytics, veracity emphasizes the need to ensure that data used for decision-making is accurate and free from errors, biases, or inconsistencies. Veracity is essential for maintaining the integrity of analytical insights and making informed decisions based on reliable data.
5. **Orchestrating:** Orchestrating refers to the process of coordinating and managing the execution of tasks or workflows within a system or application. In event-driven architectures, orchestrating involves coordinating the flow of events and actions across different components or microservices to achieve a specific outcome or business process. This may involve defining event-driven workflows, specifying the order of execution, handling errors, and managing dependencies between tasks.

**3.write about 5 V’s of Data with example scenario for each.**

**Volume:** Volume refers to the sheer amount of data generated and collected. With the proliferation of digital devices, sensors, and online activities, organizations are accumulating vast volumes of data at an unprecedented rate.

Example Scenario: A social media platform like Facebook collects massive volumes of data every day, including user posts, comments, likes, shares, and messages. The platform's data storage infrastructure must be capable of handling petabytes of data to store and manage this vast volume of user-generated content.

**Velocity:** Velocity refers to the speed at which data is generated, collected, and processed. Many modern data sources produce data in real-time or at high frequencies, requiring organizations to ingest, analyze, and respond to data quickly.

Example Scenario: A financial trading firm relies on real-time market data to make split-second trading decisions. Streaming data feeds from stock exchanges, news sources, and social media platforms are continuously analyzed to identify trading opportunities and execute transactions within milliseconds.

**Variety:** Variety refers to the diversity of data types and sources. Data can be structured (e.g., databases), semi-structured (e.g., JSON, XML), or unstructured (e.g., text, images, videos). Dealing with diverse data formats and sources presents challenges for data integration and analysis.

Example Scenario: A retail company collects data from various sources, including sales transactions (structured data), customer reviews (unstructured text data), social media mentions (semi-structured data), and sensor data from IoT devices (structured and semi-structured data). Integrating and analyzing this diverse dataset allows the company to gain insights into customer behavior and preferences.

**Veracity:** Veracity refers to the quality, accuracy, and reliability of data. Big data often comes from disparate sources with varying levels of accuracy and trustworthiness. Ensuring data veracity is crucial for making informed decisions and avoiding erroneous conclusions.

Example Scenario: A healthcare organization aggregates patient data from electronic health records (EHRs), medical devices, and wearable sensors. However, the accuracy and completeness of this data can vary due to factors such as data entry errors, device malfunctions, and patient compliance. Data cleansing and quality assurance processes are essential to ensure the veracity of the data used for clinical decision-making.

**Value:** Value refers to the ultimate goal of big data analytics: extracting actionable insights and deriving value from data. Organizations invest in big data technologies and analytics to uncover patterns, trends, and correlations that can drive business growth, innovation, and competitive advantage.

Example Scenario: An e-commerce platform analyzes customer browsing behavior, purchase history, and demographic data to personalize product recommendations and targeted marketing campaigns. By leveraging big data analytics, the platform improves customer engagement, increases conversion rates, and maximizes revenue.

**4.What is Data Ingestion vs Data Integration, explain them with a real-world use-case**.

Data ingestion and data integration are both important processes in the field of data management, but they serve different purposes.

**Data Ingestion:**

Data ingestion refers to the process of collecting and importing raw data from various sources into a storage system or data lake. The goal of data ingestion is to gather data quickly and efficiently without necessarily modifying its format or structure. This raw data can come from a wide range of sources, including databases, files, streaming data sources, APIs, sensors, and more.

**Use Case:**

Consider a retail company that operates both physical stores and an online e-commerce platform. The company collects data from multiple sources, including point-of-sale (POS) systems in stores, website interactions, mobile app usage, social media channels, customer feedback forms, and third-party vendors.

To effectively analyze this diverse dataset and gain insights into customer behavior, inventory management, sales trends, and marketing effectiveness, the company needs to ingest data from all these sources into a centralized data lake. Data ingestion tools and processes are used to efficiently collect and import data from each source into the data lake in its raw form, without any transformation.

For example, data from POS systems may be ingested as structured transaction records, while data from social media channels may be ingested as unstructured text data. By ingesting all this data into a centralized data lake, the company can create a unified view of its operations and customers, which can then be used for further analysis and decision-making.

**Data Integration:**

Data integration, on the other hand, involves combining data from different sources and formats, transforming it as necessary, and then loading it into a target system such as a data warehouse or a business intelligence platform. The goal of data integration is to create a unified and consistent view of data across the organization, enabling better decision-making and analysis.

**Use Case :**

Building upon the previous example of the retail company, once data has been ingested into the data lake from various sources, data integration processes come into play to combine and consolidate this data. For instance:

* Data from POS systems, website interactions, and mobile app usage may be integrated to create a comprehensive view of customer purchase behavior across different channels.
* Customer feedback data from different sources may be integrated and analyzed together to identify common themes and sentiment trends.
* Inventory data from stores and warehouses may be integrated with sales data to optimize stock levels and prevent stockouts or overstocking.

By integrating data from diverse sources and formats, the retail company can create a cohesive and actionable dataset that can drive business decisions, improve operational efficiency, and enhance customer experience.