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

* These are simplified system design diagram for event driven architecture which triggers an email survey as soon as a purchase transaction at Walmart.
  + Walmart Point of Sale (POS) System
  + Event Producer
  + Broker
  + Event Handler Service
  + Email Service
  + Survey Response Collector
* The Walmart Point of Sale (POS) system creates an event and forwards it to the Event Producer upon the completion of a purchase transaction.
* The transaction event is published by the Event Producer to the Message Broker.
* The events from the Message Broker are subscribed to by the Event Handler Service, which then handles them.
* The Event Handler Service notifies the Email Service of a transaction that satisfies the requirements for starting a survey (such as a high-value purchase or a first-time customer).
* An email survey is created by the Email Service and sent to the client.
* The Survey Response Collector handles the incoming survey replies, who also stores them for later analysis.

1. Capture all the new data terms from this class and create a data terms catalog.
2. Repeatability: The capacity to replay or reprocess previously created or recorded events or data. Replay ability in the context of event-driven architecture is the ability to replay events for different uses, such testing, debugging, or reprocessing.
3. Versioning: The process of tracking changes over time in data, schemas, or software components by giving them distinct identities or versions. Versioning in the context of event-driven architecture can refer to keeping event schemas backwards or forwards compatible to guarantee interoperability between various iterations of event producers and consumers.
4. Throughput: Usually expressed in terms of data units per unit of time (such as messages per second), throughput is the speed at which data can be handled or moved inside a system. The term "throughput" in event-driven architecture describes the system's ability to manage and process a specific number of events in a predetermined amount of time.
5. Latency: The time it takes for a system activity to be finished or for data to move from its source to its destination. The term "latency" in event-driven architecture describes the interval of time between an event's occurrence and event consumers' processing of it. Many times, real-time or nearly real-time systems want low latency.
6. Volume: The total amount of data produced, processed, or kept in a system or over an extended length of time. Volume in event-driven architecture describes the size or quantity of events that the system generates and consumes. Scalability and performance depend heavily on the capacity to handle large numbers of events in an effective manner.
7. Velocity: Usually expressed in terms of the rate of change or the frequency of data updates, velocity is the speed at which data is created, processed, or transferred within a system. Velocity in event-driven architecture describes how quickly events are created, sent, and consumed in reaction to modifications or interactions in the system or in its surroundings.
8. Event-Driven design (EDA): This paradigm for software design bases information flow and behavior on events, such as system events, user activities, or communications from other components. EDA's asynchronous communication between components is facilitated by events, which permits scalability, responsiveness, and loose coupling.
9. Veracity: The precision, dependability, and credibility of the data in a dataset or system. Veracity in event-driven architecture is the guarantee that event data is accurate, consistent, and devoid of mistakes or inconsistencies that could produce erroneous judgments or results.
10. Variety: Within a system or dataset, the heterogeneity or diversity of data kinds, formats, sources, or architectures. Variety in event-driven architecture refers to the existence of various event kinds, each having its own payload, metadata, and schema. To maintain flexibility and interoperability, handling diversity entails integrating and accommodating various event types and structures.

3.Read, understand and write about 5 V’s of Data with example scenario for each.

The five V's of data volume, velocity, variety, veracity, and value represent essential traits that characterize the opportunities and difficulties related to data management and analysis.

Volume: Volume is the total amount of data that is produced, processed, and stored. Both organized and unstructured data are included in this. An example scenario would be: Every second, a social media site such as Facebook gathers a massive amount of data about user interactions, posts, comments, and media uploads. Effectively handling this enormous volume of data is essential to delivering a flawless user experience and obtaining insightful information.

Velocity: The rate at which data must be processed and analyzed in addition to the rate at which it is generated is referred to as velocity. Typical Situation: Real-time data feeds are essential for financial trading firms to make quick choices about purchasing and disposing of securities. To take advantage of short-lived opportunities and efficiently manage risks, they must handle enormous quantities of market data with very little delay.

Veriety: The term "variety" describes the various forms and origins of data, encompassing unstructured, semi-structured, and structured data. Typical Situation: An online retailer such as Amazon collects information from a variety of sources, including clickstream data, customer reviews, website interactions, and transactional data from customers. They are able to enhance consumer satisfaction, optimize their supply chain, and personalize recommendations by analyzing this wide variety of data types.

Veracity: Veracity refers to the accuracy, dependability, and trustworthiness of data. Making sure the data being evaluated is of the highest caliber and devoid of biases or inaccuracies is crucial. Example Situation: To make well-informed decisions about diagnosis and treatments, a healthcare provider consults patient data. Ensuring the accuracy of this data is essential to upholding patient safety and providing efficient medical care.

Value: Value is the term used to describe the conclusions and useful information that can be drawn from data to inform business choices and provide observable results. Typical Situation: To gather information on equipment performance, energy consumption, and product quality, a manufacturing corporation uses Internet of Things (IoT) sensors throughout its production facilities. They can find chances to streamline procedures, cut down on downtime, and enhance product quality by evaluating this data, which will eventually boost productivity and profitability.

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

Although they have different functions, data integration and ingestion are both essential procedures in the field of data management.

Data Integration: The process of gathering and importing data into a data lake or storage system from multiple sources is known as data ingestion. The process entails taking raw data out of the source systems, converting it into an analysis-ready format, and then loading it into a location where it can be further processed and stored.

Real-world Use case for Data Ingestion: Let us consider a retail corporation that wishes to enhance its marketing strategy by analyzing the purchase behavior of its customers. The business may receive data from a variety of sources, including social media platforms, internet transactions, point-of-sale systems in physical locations, and consumer feedback forms. This raw data must be fed into a consolidated data repository or data lake. It contains information such as consumer demographics, purchase history, product reviews, and social media activities. After being ingested, the data might undergo additional processing and analysis to reveal patterns, behaviors, and preferences of the client base.

Data Integration: The practice of merging data from several sources and giving users a cohesive view of this data is known as data integration, in contrast. It entails combining data from several sources, resolving conflicts and inconsistencies, and presenting the information in a logical and cohesive manner.

Real – World Use case for Data Integration: Using the retail organization as an example again, integrating data obtained from several sources is the next stage towards developing a holistic understanding of customer behavior. Data from many systems, including marketing automation platforms, inventory management systems, and CRM databases, may be dispersed throughout the organization. Through data integration, the organization can combine information from several unrelated sources, settle any disputes, and provide a single customer profile that offers a comprehensive picture of every client's interaction with the company. The customer experience can be enhanced overall, inventory management can be optimized, and marketing campaigns can be made more personalized with the use of this connected data.