**Data Terms:**

1. Replayability: the capacity to replay or reprocess previously created or recorded events or data. Replicability in the context of event-driven architecture is the ability to replay events for different uses, such testing, debugging, or reprocessing.

2. 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 might mean keeping event schemas backwards or forwards compatible to provide interoperability across various iterations of event producers and consumers.

3. Throughput: The speed at which information may be sent or processed inside a system; this speed is commonly expressed in terms of data units per second, or messages per second. 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.

4. Latency: The duration required for data to move from one location to another or for a system function to be finished. 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 almost real-time systems want low latency.

5. Volume: The volume or amount of data produced, handled, or kept in a system or during a length of time. Volume in event-driven architecture describes the size or quantity of events that the system generates and consumes. Effectively managing large numbers of events is essential for performance and scalability.

6. Velocity: The rate of change or frequency of data updates is a common way to quantify the pace 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.

7. Event-Driven Architecture (EDA): A software architectural paradigm where user actions, system events, or communications from other components control the information flow and behaviour. EDA's asynchronous communication between components happens through events, enabling responsiveness, scalability, and loose coupling.

8. Veracity: The precision, dependability, and credibility of information included 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 might produce erroneous judgments or results.

9. Variety: the variety or heterogeneity of sources, formats, data kinds, or other information inside a dataset or system. Variety in event-driven architecture refers to the existence of several event kinds, each having its own payload, metadata, and schema. In order to maintain flexibility and interoperability, handling diversity entails integrating and adapting various event types and structures.

3. The 5 V's of data - Volume, Velocity, Variety, Veracity, and Value

1. Volume: The sheer quantity or magnitude of data created, saved, processed, and examined inside a certain system or environment is referred to as volume.   
   Example: Every day, a huge amount of data is processed by e-commerce platforms like Amazon. This covers everything from product views to user interactions to purchases to reviews and more. Large volumes of data are gathered and analysed by the platform to better understand consumer behaviour, provide product recommendations, enhance inventory control, and customize the user experience.

.

1. Velocity: The pace at which data is created, processed, and transferred within a system is referred to as its velocity. It frequently highlights how data processing occurs in real-time or very close to it.   
   Example: Because tweets are created and shared so quickly, social media sites like Twitter have high data velocity. Twitter has to analyse and distribute data fast in order to provide timely updates on users' timelines, trending topics, and alerts, given the millions of users that make updates, share information, and participate in conversations.
2. Variety: Variety in a dataset or system refers to the heterogeneity or diversity of data kinds, formats, sources, and structures. It includes data that is unstructured, semi-structured, and structured.   
   For instance, a healthcare facility handles many kinds of data, such as semi-structured lab results, unstructured medical pictures, and structured patient records. Healthcare workers may get thorough insights into patient health, treatment results, and disease trends by integrating and evaluating these many data sources.
3. Veracity: The terms "accuracy," "reliability," "consistency," and "trustworthiness" describe the qualities of data. It highlights the accuracy, dependability, and usefulness of the data while considering elements like completeness and correctness.   
   Example: To control risk and comply with regulations, a financial institution needs accurate and trustworthy data. Maintaining openness and confidence with stakeholders, identifying fraudulent activity, and making well-informed choices all depend on the accuracy of financial data.
4. Value: The usefulness, relevance, and insights gained by studying and understanding data are referred to as value. It focuses on the useful advantages and results that individuals and businesses may obtain from data-driven insights.   
   An illustration of this would be a retail analytics platform that examines internet activity, demographic information, and past purchases to spot patterns and trends. Retailers may improve consumer happiness, personalize marketing efforts, and optimize product offers by utilizing this information, which will ultimately lead to revenue growth and a competitive edge.

4.

|  |  |  |
| --- | --- | --- |
| Aspect | Data Ingestion | Data Integration |
| Definition | gathering, importing, and putting away unprocessed data from several sources in a system for processing or storing data. | integrating and harmonizing data from many systems, formats, and sources to produce a cohesive, consistent, and unified picture of the data. |
| Purpose | Acquiring and loading raw data into a central repository. | integrating and harmonizing many data sources to provide a cohesive and useful dataset for analysis and judgment. |
| Activities | obtaining raw data, processing it, and adding it to a destination data repository. | collecting, transforming, and adding unprocessed data to a target data repository |
| Use-case Example | For analysis, a retail business gathers data from social media, CRM systems, POS systems, and online transactions, among other sources. | After ingesting data, the business combines customer and sales information to analyse buying trends and tailor advertising campaigns. |