**Data Terms:**

1. **Replay ability**: Replay or reprocessing of previously recorded or created events or data is possible. The capacity to replay events for different purposes, such debugging, testing, or reprocessing, is referred to as replay ability in the context of event-driven architecture**.**

2. **Versioning**: the process of tracking changes in software, data, or schemas across time by giving them distinct versions or identifiers. To ensure interoperability between various versions of event producers and consumers, versioning in event-driven architecture may refer to preserving backward or forward compatibility of event schemas.

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. 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.

4. **Latency:** How long does it take for data to move from one place to another or for a system function to be finished? The time lag between an event occurring and being processed by event consumers is known as latency in event-driven architecture. Real-time or near-real-time systems typically want low latency.

5. **Volume**: the volume of data that is produced, processed, and kept in a system or over time. Within the context of event-driven architecture, volume describes the amount or size of events that the system generates and consumes. Scalability and performance are critically dependent on the effective handling of large numbers of events.

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

7. **Event-Driven Architecture (EDA):** a software design paradigm where signals from other components, system events, or user actions control behaviour and information flow. EDA allows for loose coupling, scalability, and responsiveness by facilitating asynchronous communication between components via events.

8. **Veracity**: the quality, dependability, and credibility of the data in a dataset or system. The guarantee that event data is accurate, consistent, and free from mistakes or inconsistencies that could result in wrong decisions or consequences is known as veracity in event-driven architecture.

9**. Variety**: a system's or dataset's heterogeneity, or the variety of data sources, formats, types, or architectures. The existence of various event kinds, each having a unique schema, payload, or metadata, is referred to as variety in event-driven architecture. To maintain interoperability and flexibility, handling diversity entails accommodating and integrating various event types and structures.

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

1. **Volume:** The sheer amount or size of data created, saved, processed, and evaluated 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 such as Amazon. User interactions, product views, transactions, reviews, and other data are included in this. To better understand consumer behavior, make product recommendations, enhance inventory management, and customize the user experience, the platform gathers and analyzes enormous volumes of data.

1. **Velocity:** The term "velocity" describes the rate at which data is created, processed, and transferred within a system. It frequently highlights how data processing is done in real-time or almost in real-time.

**Example:** The quick creation and distribution of tweets on social media sites like Twitter results in high data velocity. To guarantee fast updates on users' timelines, trending topics, and notifications, Twitter needs to analyze and distribute data swiftly considering the millions of users that post updates, share information, and participate in conversations.

1. **Variety:** In a dataset or system, variety is the range or heterogeneity of data sources, formats, types, and structures. It includes unstructured as well as semi-structured data.

**Example:** Numerous data types are handled by a healthcare organization, such as semi-structured lab reports, unstructured medical imaging, and structured patient records. Gaining complete insights into patient health, treatment outcomes, and disease patterns is possible for healthcare practitioners through the integration and analysis of these varied data sources.

1. **Veracity:** Veracity refers to the accuracy, reliability, consistency, and trustworthiness of data. It emphasizes the quality and reliability of data, considering factors such as completeness, correctness, and relevance.

**Example:** For the purposes of risk management and regulatory compliance, a financial institution needs accurate and trustworthy data. Making educated judgments, identifying fraudulent activity, and upholding transparency and stakeholder trust all depend on the accuracy of financial data.

1. **Value:** Value refers to the usefulness, relevance, and insights derived from analyzing and interpreting data. It focuses on the practical benefits and outcomes that data-driven insights can provide to organizations and individuals.

**Example:** A retail analytics platform analyzes customer purchase history, demographic data, and online behavior to identify patterns and trends. By leveraging these insights, retailers can personalize marketing campaigns, optimize product offerings, and enhance customer satisfaction, ultimately driving revenue growth and competitive advantage.

4.

|  |  |  |
| --- | --- | --- |
| Aspect | Data Ingestion | Data Integration |
| Definition | Collecting, importing, and storing raw data from various sources into a data storage or processing system. | Combining and harmonizing data from different sources, formats, or systems to create a unified, consistent, and coherent view of the data. |
| Purpose | Acquiring and loading raw data into a central repository. | Harmonizing and combining disparate data sources to create a unified and actionable dataset for analysis and decision-making. |
| Activities | Extracting, transforming, and loading raw data into a target data repository. | Cleansing, transforming, and enriching data to ensure its quality, accuracy, and relevance. |
| Use-case Example | A retail company ingests data from POS systems, online transactions, CRM systems, social media, etc., for analysis. | After ingesting data, the company integrates sales data with customer data to analyse purchasing patterns and personalize marketing campaigns. |