

Shayan sepasdar

100722542

EDA or Exploratory data analysis is a basic technique method used for data analysis, this technique is mainly used in the data preparation phase, also it provides graphical visualization of data which is very helpful for identifying data patterns and comparison.

Advantages: High scalability , decouple services based on user need .

Disadvantages: poor or improper use of EDA could misguide a problem.(high complexity)

EDA won't be effective if we are dealing with high-dimensional data.

KAFKA

Kafka **topic** is the feed to which a producer pushes messages/records, and a consumer subscribes to and consumes messages from it.

Partition Kafka breaks topic logs up into partitions for speed, scalability, and size.

Leader is a broker that is chosen to own the partition in the cluster

Cluster consists of at least three brokers in order to allow broker to provide enough redundancy, commit messages to disk and be responsible for fetching requests.

Kafka **Brokers** are the servers that together form a Kafka **cluster** by sharing information with each other.

In a Kafka cluster, one of the brokers serves as the **Controller** which is responsible for managing the states of partitions.

A **Producer** is an application which a source of data being pushed to one or more Kafka topics

A **Consumer** is a client application that fetches data from Kafka topics.

Zookeeper is installed on Kafka brokers for maintaining configuration, processing, and managing data flow.

Kafka **consumer group** is basically several consumers who can read data in parallel from a Kafka topic.

Prepare a video showing the codes that generated topics, produce messages, and consume them in both NodeJS and python. Your video should display the possible producer and consumer scenarios.

<https://drive.google.com/file/d/1pKUol3EX1F5XlirYMTDnCh1r7ROWPrXK/view?usp=sharing>

https://drive.google.com/file/d/17x1d1edGr_daPeIRWNYOtRvvyIJJS0QS/view?usp=sharing

Docker persistent volume

I fixed this problem by using the docker volume method , using this method avoid losing any of the messages/data when the created images are down. This simply happens because the data is now kept outside of the container which makes them independent of the actions inside the container.