SOFE 4630U
Cloud Computing

Project Milestone #2
2022/02/15

Data Ingestion Software and Kafka Clusters

Jerusha Macwan: 100723319

EDA and its advantages and disadvantages:

An event driven architecture is popular in current applications created using microservices.. It utilizes events to initiate and interact across decoupled services. An event is a state change or update, such as when an item is added to a shopping cart on an online shopping website. The three main components of event driven architecture include an event producer, event consumers and brokers. Producer and consumer services are decoupled, allowing for autonomous scaling, updating and deployment.

Advantages:
- It is easily scalable providing quick responses to events.
- Due to decoupling, the services are interoperable, that is if failure of one component occurs, the other will continue to function.
- It speeds up the development process.
- It helps cut costs for companies.

Disadvantages:
- It is complex and takes time to understand.
- Errors might be difficult to detect.
- It's common to have duplicate names or naming problems when dealing with a large number of publishers and subscribers.

Cluster: A group of one or more Kafka brokers or servers running kafka make up a kafka cluster.
Broker: Supports and processes subscriber requests and ensures that data is replicated throughout the cluster. A cluster can contain multiple brokers.

Topic: It is a category or data structure in which messages are stored and published.

Replica: A replica is the redundant element of a topic partition. A partition often contains multiple replicas. Replication occurs at partition level.

Partition: Topics are separated into partitions each of which contain messages in a fixed order. A distinct offset is given and attributed to each message in a partition.

Zookeeper: A zookeeper manages and maintains the servers in the kafka cluster. There can be several zookeepers in a cluster.

Controller: One of the brokers in the cluster, acts as the controller, handling the partition and replica states as well as administrative activities such as reallocating partitions. Only one controller server is active in the cluster at any given moment.

Leader: One server serves as the leader of each partition, while others function as followers. Leader is in charge of all read-write queries for the partition, while the followers duplicate the leader.

Consumer: A consumer subscribes to a topic and retrieves messages from a kafka topic.
Producer: These produce events and push messages into kafka topics.

Consumer Group: A group of consumers which is assigned with a partition of a topic.