Videos:
https://drive.google.com/drive/folders/1y5C2Rx6zwPaUfBhe7g3_4B6g6X9mX8iX?usp=sharing

Github: Yusuf-shaik/Cloud-Lab-3 (github.com)

1. Describe the following:

- Sink and Source connectors.
    - Source: any external system where data is being imported from. Eg: a MySQL database where the data is being fed into Kafka. Azure IoT hub is also an example of this.
    - Sink: an external system that Kafka will export data to eg a data lake storage that will contain all data dumps.
    - These connectors are used as a pipeline between Kafka and producers & consumers.
- The applications/advantages of using Kafka Connectors with data storage.
    - Data are available for stream processing with low latency.
    - REST interface.
    - Distributed and scalable by default.
    - Streaming and batch integration.
    - Kafka connectors are premade components that allow developers to connect external applications as sources and sync to their Kafka event stream. This allows us to decouple the system and keep difficult components of our system separate. Developers will connect their Kafka cluster to an IoT hub and define it as the source, then define an SQL database as the sink. All data streamed from the IoT hub source will then be saved in the SQL table sink. This enables developers to integrate Kafka seamlessly into their applications and connect popular components with ease due to the ready-made connectors.
- How do Kafka connectors maintain availability?
    - Kafka connectors are able to remain available because they can be defined as distributed. This allows them to be created in a cluster with workers and tasks separated. All config data is stored in a synchronized config file, so if a worker dies, a new worker can be spawned with the same tasks and offset data as the previous worker. Multiple clusters can also be created to distribute workload (eg: one cluster for streaming to elastic search, another to dump the data in hdfs).
- List the popular Kafka converters for values and the properties/advantages of each.

Converter	Details
Avro	Schema-based and allows for fast binary serialization.
Protobuf	Allows both binary serialization and JSON serialization. ~6 times faster than Avro.

String	Simple data transmission with low effort, but not schema-based with organized data.
JSON	Commonly used in most applications, any microservice communicating will offer JSON communication and serialization which will ensure fewer conflictions with dependent applications.
ByteArray	Is commonly used in event streaming applications.

2. Search the internet to answer the following question:

- What's a Key-Value (KV) database?
    - The key-value database is a database similar to a hashmap or a dictionary that allows you to save values, strings, objects, bitmaps, etc.. with an associated key.
- What are KV databases' advantages and disadvantages?
    - Advantages: Access is O(1) since there is only one key for each value in the database. Write and read operations are faster. The values can be anything
    - Disadvantage: Data can only be accessed with that single key, data is not structured like in an SQL table where data can be manipulated and queried based upon its values. If the user does not have the key, they cannot access the data based on the values associated with all keys. In SQL, a user can query based on any field by simply adding it to the select statement. It is also not optimized for lookup
- List some popular KV databases.
    - Redis, Couchbase, Amazon DynamoDB, Azure table storage, Azure Cosmos DB, Cassandra

9. The datasets from the link provided could be used to map a full-scale 3d image of the values collected. The image dataset contains images of all places traversed, and it can be used to create a full-scale 3d map of the location. The ground truth pose is the information at a certain location, and that info can be used combined with the 3d map to create an interactive display for the user. This interactive display will be a 3d map of the area, allowing the user to click on any area and get the information gathered by the ground truth pose.