Group 2

1. Describe the following:
   - **Sink and Source connectors**: **Sinks** basically take data from Kafka topics and deliver them to other systems such as Hadoop, Elasticsearch, or batch systems. **Sources** are essentially stream tables, databases, or message brokers that take data from a system.
   - **The applications/advantages of using Kafka Connectors with data storage**: The great thing about Kafka connectors is that it is very easy for external systems to import and export data. Since they are ready-to-use they can be applied in situations where we require data streaming between other systems or transfer data quickly between systems.
   - **How do Kafka connectors maintain availability**: Instances for each connector copies data, and through having the connector break a job into multiple tasks this coordination allows for scalability and parallelism with minimal configuration. Also there are no states stored within these tasks thus meaning that they can be re-coordinated or assigned at any time providing it with high availability.
   - **List the popular Kafka converters for values and the properties/advantages of each**: JSON schema converter is used when reading or consumption of JSON data from Kafka topics into a sink connection (Connect Sink).

2. **Search the internet to answer the following question:**
   - **What's a Key-Value (KV) database**: They are essentially a form of non relational databases that utilize what is called a simple "key-value method" when storing data.
   - **What are KV databases' advantages and disadvantages**:
     - **Advantage**: is that the read and write operations are quick due to the data format being simple.
     - **Advantage**: is that the data has no requirement of type. It could be anything like flexible schema's or JSON type data.
     - **Disadvantage**: is that to store multiple values a parser is required and the optimization of the database is only for a single key and value data.
     - **Disadvantage:** is that looking up specific key's for a single piece of data is not optimized and would require the entire collection to be scanned or separate index values to be created for referencing.
   - **List some popular KV databases**:
     - Amazon DynamoDB
     - Oracle NoSQL Database
     - InfinityDB, Redis, Aerospike
     - Oracle Berkeley DB

- Riak KV
- Voldemort

3. **List some possible applications that can be implemented by using the uploaded dataset:** Given that this dataset is able to simulate real sensors and save states in a database there are a few real-world applications that you can utilize with this.
    - Website activity tracking
    - Logistics for businesses and economics
    - Real-time data streaming for large presentations, google meets, or screen sharing
    - Real-time shipments of products
    - Monitoring star clusters and planet coordinates for astronomy
    - Monitoring cars during intersections
    - Land mapping for agriculture