SOFE4630 Cloud Computing (Winter 2022 - Dr. M. El-darieby)

Lab 3: Data Ingestion Software - Kafka Clusters

Mar 18, 2022

Group 6

Cassidy Linhares	100615025
Iliya Karac	100703933
Kaushik Ramani	100651855
Tanzir Hossain	100709113

Links

Title	**Link**
Github	lab 3 repo

Step 7/8	video folder

Describe the Following:

- **Sink and Source connectors.**
    - Sink Connectors: they are responsible for getting data from something through kafka and into the database by following a schema of how the database should be.
    - Source Connector: responsible for pulling data from a database and converting it to a schema which kafka does and then kafka pipes it to a consumer
- **The applications/advantages of using Kafka Connectors with data storage.**
    - The advantage of using a service like confluent is that you can bypass most programming and set up the connections through a user interface most likely. But this only applies to a selection of mainstream services. If you want to use a private system as your source or system you will need kafka connectors. Fortunately they are also fairly simple to use. Instead of rewriting the code for every connection you can use a connector and declare all the connections you need. They also record data automatically which is useful in data storage.
- **How do Kafka connectors maintain availability?**
    - A kafka connector is a worker object that runs in a java virtual machine. Many of these worker objects can be made to balance the load of the tasks. This also means that there are more connectors working which allows for availability. If a worker were to go down its specifications are stored in kafka and a new worker with the same specifications as the previous one exactly where the old one went down.
- **List the popular Kafka converters for values and the properties/advantages of each.**

  String - converts to string. Good for immediate processing. Not legible as it is not structured

  JSON - default format. Very legible and supported by many systems. Lower character count compared to String

  Protobuf - Protocol Buffers, or Protobuf, are a smaller/faster/simpler replacement for XML files. Supported by many systems.

Answer the Following:

- **What's a Key-Value (KV) database?**

  A key-value database is a no-sql database that instead of columns and rows it uses a key to value pairing to store data. All the data entries can follow a specific format but are allowed to have vastly different attributes that may be unrelated. The database is also non relational meaning that the key values pairs cannot be related to each other by a specified attribute.

- **What are KV databases' advantages and disadvantages?**
  - **Advantages:** can store any value type or any shape of data, simple, fast, reliable
  - **Disadvantages:** not optimized for index querying or querying by value, collects junk if not set up correctly, non-relational
- **List some popular KV databases.**
  - MongoDB
  - DynamoDB
  - Redis
  - Aerospike
  - Hbase

Task 4-6 screenshots provided in individual reports

List some applications that can use the data set from Step 8:

- Path finding through VR
- Autonomous vehicles
- Computer Vision
- EDGE device processing
- Object detection