Cloud project

Lab 3: Kafka connects for Redis and MySQL

Objective:

- Deploy Tabular and key-Value data storage to GKE.
- Get familiar with Key-Value data storage
- Get familiar with Kafka Connectors and their configuration.
- Configure and use Kafka source connector to Redis.
- Configure and user MySQL sink and source Kafka connectors.

Lab3 repository: https://github.com/goergedaoud/SOFE4630U-tut3.git

(Links to an external site.)

Procedure:

1. **Watch the first three videos for Kafka connectors (focus on the concepts, not the details) from https://www.confluent.io/blog/kafka-connect-tutorial/**

Describe the following:

- **Sink and Source connectors**
  - A Source Connector is responsible for bringing data into Kafka (with the aid of Source Tasks), whereas a Sink Connector is responsible for getting data out of Kafka (with the help of Sink Tasks).
- **The applications/advantages of using Kafka Connectors with data storage.**
  - Data-Centric Pipeline - Connect pulls or pushes data to Kafka using meaningful data abstractions.
  - Connect can be used with both streaming and batch-oriented systems on a single node (standalone) or scaled to a company-wide service (distributed).
  - Reusability and Extensibility - Connect makes use of existing connectors or extends them to meet your specific requirements, resulting in a faster time to market.
- **How do Kafka connectors maintain availability?**
  - Ingests whole databases and feeds table updates to Kafka topics using the source connector. A source connector can also collect metrics from all of your application servers and store them in Kafka topics, allowing for low-latency stream processing.

- **List the popular Kafka converters for values and the properties/advantages of each.**
    - On the route into or out of Kafka, converters serialize or deserialize the data.

3. Search the internet to answer the following question:

- **What's a Key-Value (KV) database?**
    - A key-value database is a nonrelational database that stores data using a simple key-value mechanism. Data is stored in a key-value database as a collection of key-value pairs, with a key serving as a unique identifier. Both keys and values can be any type of object, from simple to sophisticated compound objects.
- **What are KV databases' advantages and disadvantages?**
    - A key-value database is a nonrelational database that stores data using a simple key-value mechanism. Data is stored in a key-value database as a collection of key-value pairs, with a key serving as a unique identifier. Both keys and values can be any type of object, from simple to sophisticated compound objects.
- **List some popular KV databases**
    - Amazon DynamoDB
    - Amazon ElastiCache
    - Redis
    - Couchbase
    - ScyllaDB
    - Aerospike
    - Hbase
    - InterSystems IRIS

4. Follow the following videos to deploy and use Redis and MySQL databases using GKE.

- https://youtu.be/qVD1uVKMZYc
- (Links to an external site.)

-
- https://youtu.be/GBvGjVLbYIs
- (Links to an external site.)

- https://youtu.be/9R3eEAIpOtk
- (Links to an external site.)

5. Follow the following video to set up sink and source Kafka connectors to the deployed MySQL database.

- https://youtu.be/B9sgQOQoGHA
- (Links to an external site.)

- https://youtu.be/yLxfDw9Yh_A
- (Links to an external site.)

6. Follow the following video to set up a Kafka connector to the deployed Redis database.

https://youtu.be/dJWUsAVigR4

(Links to an external site.)

7. Now, you will store a dataset into cloud storage. The dataset has to be sent into Kafka topics and connectors have to be configured to automatically store the dataset into the data storage. The producer that will send the dataset to Kafka topics should run on your local machine as it will simulate real sensors while Kafka, connectors, and data storage should be on the cloud. Use MySQL for the CSV files and Redis for images. Feel free to update the Yaml files from the given repository to fit your dataset.
8. Use the sensors, images, ground truth Pose in the lastest session in http://robots.engin.umich.edu/nclt/
9. (Links to an external site.)
10. as your dataset.

Record a video showing the configuration of Kafka connectors, producers' python script, a proof of successfully stored data into data storage.

9. List some possible applications that can be implemented by using the uploaded dataset.

IT is applicable to a variety of AI applications, machine learning. Practical uses could include object detection or medical databases. By storing the data on the cloud, it could privately be authorized that anyone from anywhere can access it and have unlimited services.