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

Lab 3: Data Storage

Mar 18, 2022

Group 6

Cassidy Linhares	100615025

Describe the following:

Name	Description
Sink Connector	The sink connector is how data gets put into Kafka to go to a database. Kafka just needs a rough idea of the schema for the database and then takes the data from the producer, that should roughly match the schema, and kafka pipes it into a database that you set up the connector to.
Source Connector	The source connecter pulls data from the database into kafka and kafka pipelines it to whoever is calling for it. It uses a schema to fit the data into its designated shape and gives that to a consumer to use
Advantages of Kafka Connectors	Connectors are really good for transferring/pipelining data from one database to another (in the event you are migrating databases) It is also good for decoupling the databases from any producers and consumers. It can also help slow the rate of data that may be coming in too quickly. Also good for scaling to use multiple databases and create event driven architecture
How does Kafka connectors maintain availability	If one of your sources or sinks goes down, Kafka can replay events in the order they came in since Kafka stores messages as well
List the popular Kafka converters for values and the properties/advantages of each	Popular Kafka converters: mysql, redis, amazonS3, hadoop. Postgresql, elastic search Advantages: elastic search, and redis are no-sql so good for storing documents, elastic search and hadoop are good for fast querying or processing of a lot of data MySQL, Postgresql are sql databases and have shape to them which is good for defining objects

Search the Internet for the answers to the following questions:

What is a Key-Value Database?

A Key-Value Database is a type of non relational database and uses a key:value method to store data. Like a dictionary or map or json object except the value can be anything. The values are retrieved through the keys

What are the advantages and disadvantages of a Key-Value Database?

Advantages: Key value databases are easy to partition and allow horizontal scaling very easily. It is very fast for read and write operations since this data structure is said to have a read time of O(1). They are reliable because of how redundant they are.
Disadvantages: Querying can be a challenge since the only way to query is by key (thus the creation of documents). Also, there is no standard language (like SQL) that can be used to query data, meaning it can differ from database to database. The values are sometimes blob types so you may not be able to query by value

List some popular Key-Value databases?

- Redis
- Mongo DB
- Amazon DynamosDB
- Aerospike
- Google firebase
- Elastic search

Video Links:

Title	Link
Step 4.1	mysql k8 prt1
Step 4.2	mysql k8 prt2
Step 4.3	redis k8
Step 5.1	mysql source
Step 5.2	mysql sink
Step 6	redis
Step 7	Didnt do because the file was too huge to download and every step kept breaking and I had to sped 3h+ on step 4 &5 because of that so I gave up by this point