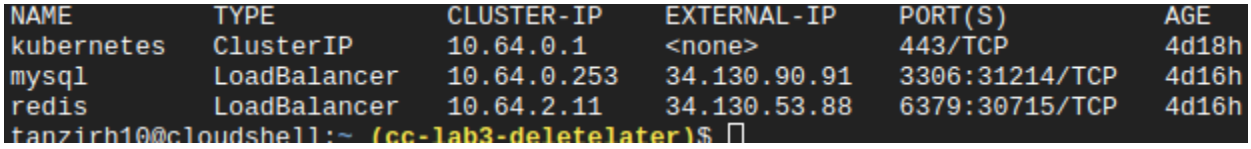
1. Watch the first three videos for Kafka connectors (focus on the concepts, not the details) from https://www.confluent.io/blog/kafka-connect-tutorial/ (Links to an external site.).
2. Describe the following:
    - Sink and Source connectors.
        - Source connectors allow you to import data from any relational database into a kafka topic
        - Sink connectors allow you to export data from kafka topics to any database
    - The applications/disadvantages of using Kafka Connectors with data storage.
        - Pros
            - No use of code
            - Easy way of integration
        - Cons
            - Available connectors can have some limitations that are not very apparent
            - Documentation can be poor and be lacking
    - How do Kafka connectors maintain availability?
        - In the event of failures, processing will be saved from the last committed offsets and tasks will recover
        - In most of the failure cases, progress will retrigger when back up
    - List the popular Kafka converters for values and the properties/advantages of each.
        - JSON
            - Simple text making it suitable and safe for many platforms and systems
            - Self describing
            - Compact and easy to work with
        - Avro
            - Faster to load
            - Doesnt require typing or serialization
            - Easier to parse
        - Protobuf
            - Simplest and faster and smaller
            - Support for rpc
            - Having a predefined and larger structure, when compared to JSON, set of data types, messages serialized on Protobuf can be automatically validated by the code that is responsible to exchange them.

3. Search the internet to answer the following question:
    - What's a Key-Value (KV) database?
        - A key-value database stores data as a collection of key-value pairs in which a key serves as a unique identifier. Both keys and values can be anything, ranging from simple objects to complex compound objects.
    - What are KV databases' advantages and disadvantages?
        - Advantages
            - Simple data format
            - Values can be anything
        - Disadvantages
            - Not optimized for lookup
            - Optimized for only single key and value
    - List some popular KV databases.
        - DynamoDB
        - InfinityDB
        - Redis
        - Oracle NoSQL DB
4. Follow the following videos to deploy and use Redis and MySQL databases using GKE.



Mysql

```
mysql: [Warning] Using a password on the command line interface can be insecure.
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 646
Server version: 8.0.28 MySQL Community Server - GPL

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> select * from myDB.test
    -> ;
+----+-------+---------------+------------+---------------------+
| id | name  | email         | department | modified            |
+----+-------+---------------+------------+---------------------+
|  1 | alice | alice@abc.com | eng.       | 2022-03-17 21:42:21 |
|  2 | bob1  | bob1@abc.com  | sales      | 2022-03-17 21:42:21 |
|  3 | bob2  | bob2@abc.com  | sales      | 2022-03-17 21:42:21 |
|  4 | bob3  | bob3@abc.com  | sales      | 2022-03-17 21:42:21 |
|  5 | bob4  | bob4@abc.com  | sales      | 2022-03-17 21:42:21 |
|  6 | bob5  | bob5@abc.com  | sales      | 2022-03-17 21:42:21 |
|  7 | bob6  | bob6@abc.com  | sales      | 2022-03-17 21:42:21 |
|  8 | bob7  | bob7@abc.com  | sales      | 2022-03-17 21:42:21 |
|  9 | bob8  | bob8@abc.com  | sales      | 2022-03-17 21:42:21 |
| 10 | bob9  | bob9@abc.com  | sales      | 2022-03-17 21:42:21 |
+----+-------+---------------+------------+---------------------+
10 rows in set (0.03 sec)

mysql>
```

Redis

```
tanzirh10@cloudshell:~/SOFE4630U-tut3/GKE (cc-lab3-deletelater)$ redis-cli -h 34.130.53.88
34.130.53.88:6379> auth SOFE4630U
OK
34.130.53.88:6379> keys *
1) "__kafka.offset.ToRedis.4"
2) "key1"
3) "k1"
4) "Course"
5) "OntarioTech"
34.130.53.88:6379>
```

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

MySqlSourceConnector_0	• Running	Source	lcc-0xxk86	MySqlSource	1	0	0B/s	0	0

MySqlSourceConnector_0

See in Stream lineage

Overview | Settings

Pause | Delete

Running

This connector is running.

Messages processed

30

Total in last 7 days

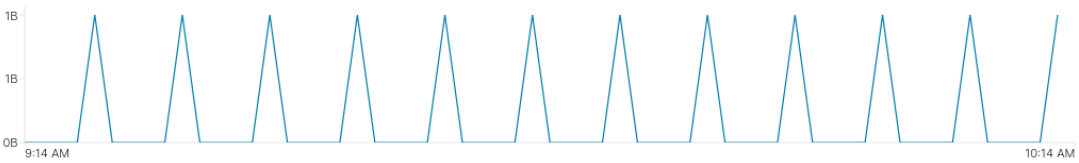
Last hour

Explore metrics

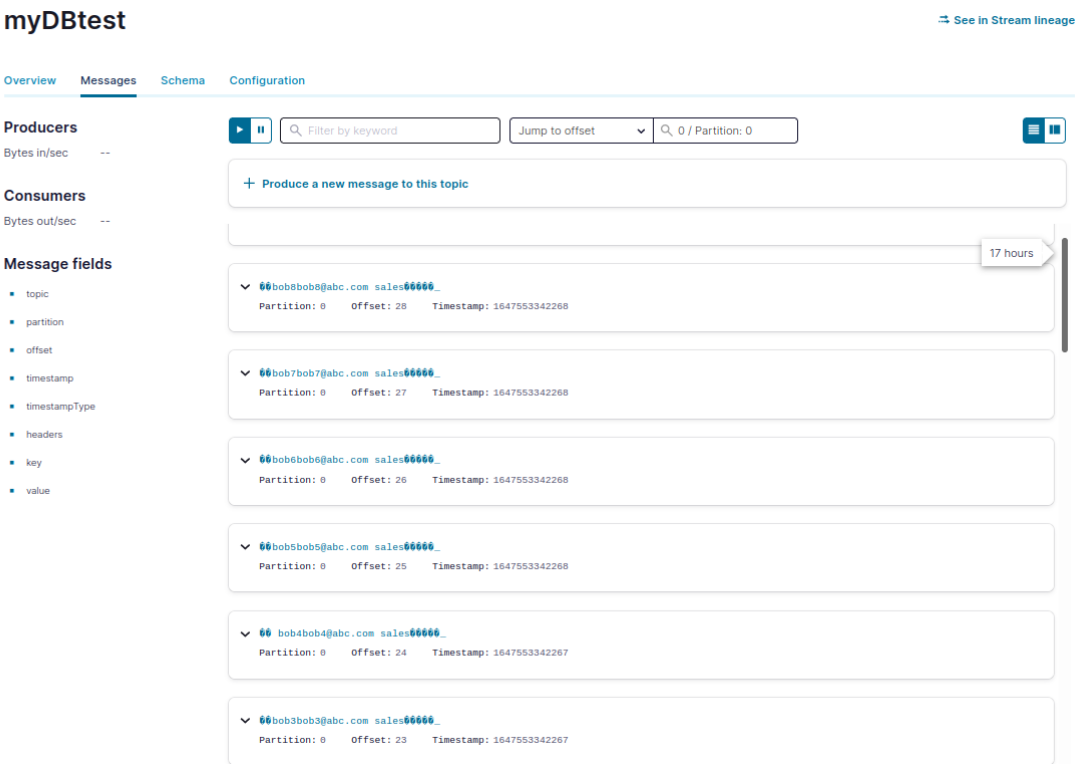
Throughput

Messages (messages/sec)

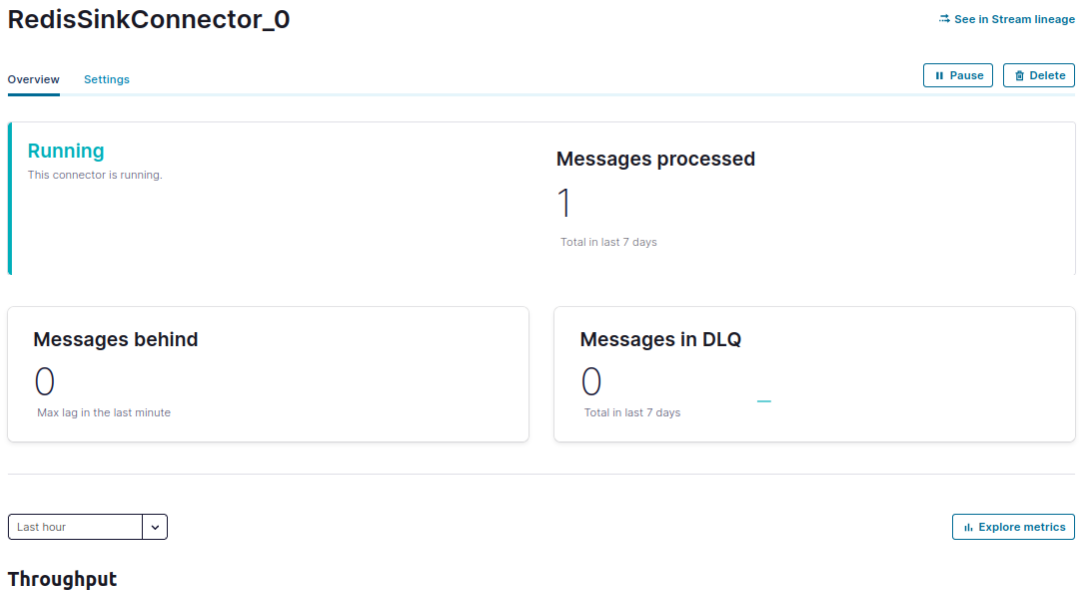
Production (bytes/sec)

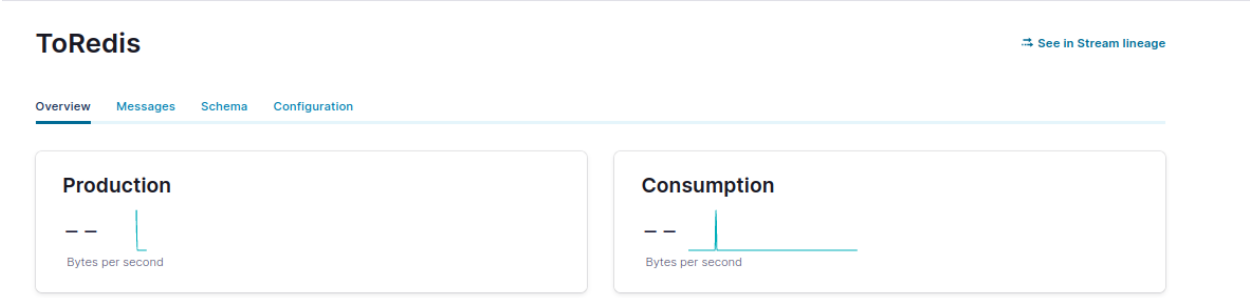


myDBtest	1	--	--	✔



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





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.

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

I was not able to successfully upload a dataset into cloud storage as working with the csv proved to be a hassle. The csv conversion to avro proved to be a hassle with very little documentation online as working with a csv made it so I couldn't work with single json dictionaries and had to work with json arrays . I also had problems with working with the schema provided in the MYSQL folder in the connectors folder

8. List some possible applications that can be implemented by using the uploaded dataset.
    - The dataset used is primarily to help in the research of robotics. This dataset helps in evaluating different robotic algorithms, help with tasks involving mapping, navigation and various other tasks that the field of robotics research in. It can also be used in various other ways including but not limited to: computer vision, obstacle detection and tracking, and prediction of changes in environments/appearances