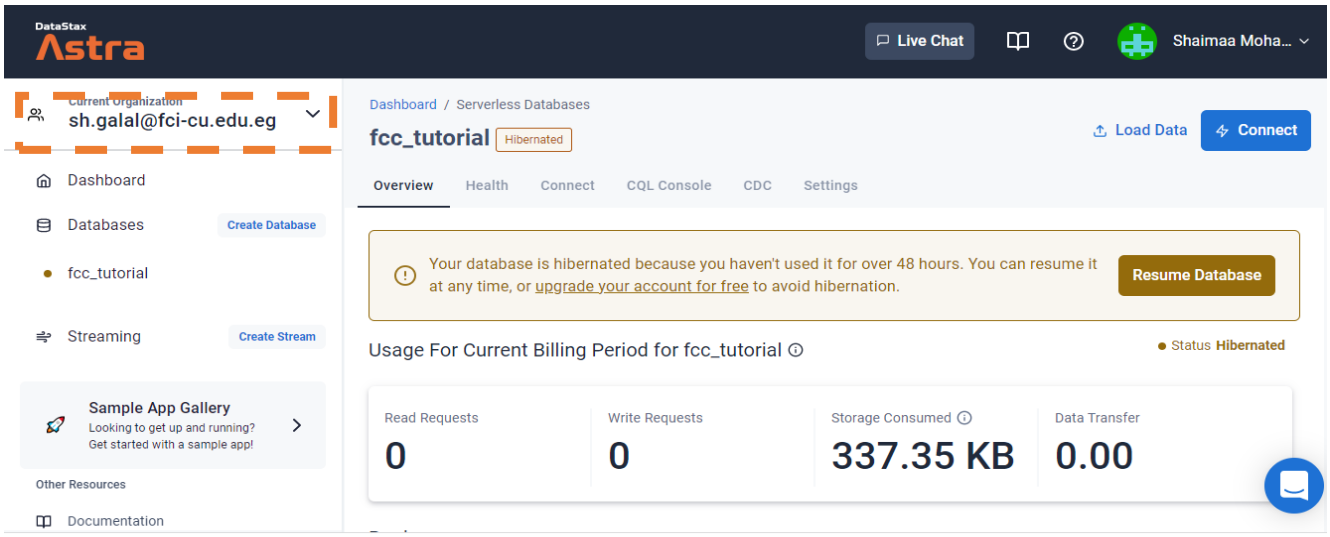
Cairo University
Faculty of Computers and Artificial Intelligence
SIS314 Advanced Database – Fall (2024) – Dr. Shaimaa Galal

Assignment 2: Cassandra DB

Students number: Max 5 (groups must be in the same lab)	**Deadline**: 28 December

Consider implementing the following using Cassandra DB (DataStax Astra DB).

1. Create a KeySpace named “Movies” with a replication factor 1 and simple strategy.
   Document a screenshot for this step showing your used email in that screenshot as in the figure below



For the next points use the CQL shell (Except points 5 and 6) and deliver the CQL Statements that performs the following + screenshot of CQLSH : (CQL 5 marks – python code 5 marks)

2. Create a column-family “Movie” with columns (Id int, name text, movie-cast map, movie-poster blob). The cast column is a map with (director(s) – actors(s) – music-cast-person(s)). Set the first row TTL to 7 days.
3. Check the schema of the “Movie” column-family.
4. Populate the Movie table with 3 real movies of your choice to populate all the columns except the movie poster (you might use https://www.imdb.com/ to collect the data).
5. Write a python function that connects to your KeySpace, transform the movie poster (extracted from IMDB) to blob datatype and then update the movie-poster column for the three inserted rows. The Python function should read the images from a folder on your machine, transform it to blob and update your DB rows.
6. Write another python function to query the movies given certain director or actor (as input form the program) and display the row results including the image (**suggestion:** you can output the image to output folder). Discuss if you can use “Like” or “Contains” CQL operators to implement this query (proof by CQL command).
   Helping-tip: Cassandra supports few programming languages. You can use any other supported language if you **do not prefer python**.
7. Update a certain movie actors list to add another actor to the list.
8. Update the first row TTL to 3 seconds. After the 3 second re-query the table then illustrate your findings.