CSP554–Big Data Technologies

Assignment #11

Worth: 60 points

Due by midnight of next class (August 1ˢᵗ)

Assignments should be uploaded via the Canvas portal.

Set-Up:

Step A – Start an Dataproc cluster (Chooose)

Start up an Dataproc cluster as previously, but instead of choosing the "Core Hadoop" configuration chose the "Spark Interactive" configuration (see below), otherwise proceed as before.

Step B – Download the assignment software (mongoex.tar, mongodb-org-4.2.repo) to master node

Download "mongodb-example.zip" (included as a file with the assignment) to your PC or MAC. Now, using "scp" copy this file to the DATAPROC master node using something like the following (just an example):

scp -i ./Dataproc-key-pair-2.cer mongodb-example.zip hadoop@<master-node-ip>:/home/hadoop/

Then enter this into Init-Term to unzip mongodb-examples.zip:

unzip mongodb-examples.zip

Now download "mongodb-org-7.0.repo" (included as a file with the assignment) to your PC or MAC. Now, using "scp" copy this file to the DATAPROC master node using something like the following (just an example):

```
scp -i ./Dataproc-key-pair-2.cer mongodb-org-7.0.repo hadoop@<master-
node-ip>:/home/hadoop/
```

Step C – Install assignment software (mongoex.zip, mongodb-org-7.0.repo)
Ensure you are running Linux 2023, which is crucial for installing MongoDB 7.0

Enter the following into a terminal window which you have connected to the DATAPROC master node. Going forward we will call this terminal connection Init-Term:

sudo cp mongodb-org-7.0.repo /etc/yum.repos.d

Verify if MongoDB 7.0 was added to the Yum repository list:
sudo yum repolist

Step D – Install and start MongoDB

Enter the following into Init-Term to install MongoDB:

sudo yum install -y mongodb-org mongodb-mongosh-shared-openssl3 openssl mongodb-org-database-tools-extra mongodb-database-tools mongodb-org-tools mongodb-org-server mongodb-org-mongos mongodb-org-database

Verify that mongodb was installed correctly:
rpm -qa | grep mongodb-*

mongodb-org-database-tools-extra-7.0.8-1.xxxx86_64
mongodb-org-server-7.0.8-1.xxxx86_64
mongodb-org-mongos-7.0.8-1.xxxx86_64
mongodb-org-database-7.0.8-1.xxxx86_64
mongodb-mongosh-shared-openssl3-2.2.5-1.el8.x86_64
mongodb-database-tools-100.9.4-1.x86_64
mongodb-org-tools-7.0.8-1.xxxx86_64
mongodb-org-7.0.8-1.xxxx86_64

Now enter this into Init-Term to start mongodb:

sudo systemctl start mongod

Step E – Start the MongoDB Shell (Command Line Interpreter)
Open a second terminal connection to the DATAPROC master node. Going forward we will call this terminal connection: CLI-Term.

You will use this terminal window to start and run the mongodb shell as follows:

mongosh

Step F – Edit mongo query language files
Open a third terminal connection to the DATAPROC master node. Going forward we will call this terminal connection: CLI-Term. You will use this terminal window to run the 'vi' editor to create your Mongo code files.

As an alternative you could edit your MongoDB code files on your PC/MAC and then 'scp' them to the DATAPROC mater node.

Step G – Setting up the assignment database

Now, in the MongoDB shell, using the CLI-Term, create a database called "assignment" by entering the following into the MongoDB shell:

    use assignment;

This will set the shell variable 'db' to this new database.

Load a collection called 'unicorns' with sample data by executing the script load.js in the MongoDB shell as follows (don't cut and paste this, type it in manually):

    load('./load.js');

Note, look at the content of the script file (via the other terminal window you have opened to the COMPUTE instance) to see how each unicorn is described.

Confirm this has all worked by executing the following command in the MongoDB shell:

    db.unicorns.find();

NOTE: The files named "demo*.js" (also included in the mongoex.tar file) provide examples of operating in the unicorn collection. These are a VERY good idea to review and understand, and they will present you with information that will help you complete the assignment. Also, try them out by typing something like

    load(./demo1.js');

Exercises:
Exercise 1) (5 points)
Write a command that finds all unicorns having weight less than 500 pounds. Include the code you executed and some sample output as the result of this exercise. Recall

you can place the command, if you choose, into a file, say 'ex1.js' and execute it with the load command as above and similarly for the following exercises.

Exercise 2) (5 points)
Write a command that finds all unicorns who love apples. Hint, search for "apple". Include the code you executed and some sample output as the result of this exercise.

Exercise 3) (5 points)
Write a command that adds a unicorn with the following attributes to the collection. Note dob means "Date of Birth."

Attribute	Value(s)
name	Malini
dob	11/03/2008
loves	pears, grapes
weight	450
gender	F
vampires	23
horns	1

Include the code you executed to insert this unicorn into the collection and the output of a find command showing it is in the collection.

Exercise 4) (5 points)
Write a command that updates the above record to add apricots to the list of things Malini loves. Include the code you executed and some sample output showing the addition.

Exercise 5) (5 points)
Write a command that deletes all unicorns with a weight of more than 600 pounds. Include the code you executed and some sample output resulting from this exercise.

Exercise 6) (35 points)
Read an article on "Canvas" in the "Articles" section titled **Modeling temporal aspects of sensor data for MongoDB NoSQL database** (35 pages!) Summarize the article in 1-page and not more. (Please refer to "**How to Summarize a Research Article**")