***Text

Description automatically generated with low confidence***

***Replication***

Student Name: Laiba Asif

Student ID: R00201303

Module: NoSQL Data

Architectures

Date: 10/02/23

To set up the distributed environment using MongoDB scripts, I followed the steps below

1. From the resources folder, download the script files create\_nodes.bat, setup\_cluster.bat, insert\_collection.bat, shard\_collection.bat, and remove\_cluster.bat.
2. I opened the command prompt and went to the directory where the script files were saved.
3. I reboot the computer before beginning the process to guarantee a clean environment.
4. The machine name in the script files should be changed to reflect the name of my machine. I use a text editor to open each script file, look for the machine name placeholder, and then change it to the name of my computer.
5. Run the command C: to open a new command prompt and go to the "C:" drive.
6. Go to "This PC" -> right-click and choose "Properties" -> copy the computer name to find the name of the logical lab machine.
7. I adjust the directory paths in the’ mkdir’ commands to point to a directory where I wrote permissions on the C: drive after opening the script file create\_nodes.bat. The script's paths should be updated to reflect the intended directory.
8. Save the create\_nodes.bat script file after any changes have been made.
9. To ensure a clean environment, run the remove\_cluster.bat script first. Any current MongoDB processes or data will be deleted by this script.
10. From the command prompt, execute the create\_nodes.bat script. For each replica set node, this script will start MongoDB instances on various ports and construct the required directories.

* Based on the supplied directory paths in the script, the script will generate directories for each replica set node, such as dublin0, dublin1, dublin2, and so forth.
* The allotted port numbers for each node, such as 27000 for dublin0, 27001 for dublin1, and so on, will be used to launch each MongoDB instance.
* The output of the script will show that each MongoDB instance has successfully started up.

1. Check the output from the command prompt to ensure that the processes are active and functioning. Messages confirming the successful startup of MongoDB instances on the specified ports should be present.
2. Use a new command prompt to execute the remaining scripts (setup\_cluster.bat, insert\_collection.bat, and shard\_collection.bat). The scripts in question will set up the cluster, add information to the MongoDB collection, and shard the collection using the designated shard key.

setup\_cluster.bat

* The given replica set names (ending with your student ID) and port numbers will be used by the script to establish connections to each MongoDB instance and set them up as replica set nodes.
* The supplied configuration servers (cfg0, cfg1, cfg2) and their corresponding port numbers will be used to configure the replica set nodes.
* For the supplied database and collection, the script will start the replica sets and activate sharding.

insert\_collection.bat

* By establishing a connection to the MongoDB cluster, the script will insert the given dataset into the chosen database and collection.
* The information from the JSON file is imported into the collection using the mongoimport command with the --jsonArray parameter.
* The script's output will show how many documents were inserted.

shard\_collection.bat

* The script will establish a connection to the MongoDB cluster and set up sharding for the chosen collection and database.
* It specifies the shard key and enables sharding for the collection using the mongosh command and JavaScript code.
* The script's output will show that sharding has been enabled successfully.

1. I review the output from the scripts to ensure that the cluster setup, data insertion, and sharding processes are completed successfully.
2. After all the scripts have run, open a new command prompt and type ‘mongo’ command this will launch the Mongo client. The MongoDB shell will open as a result.
3. To check that the cluster is operating properly, use the MongoDB shell to query the data in the collection.
4. I ran the relevant queries in the MongoDB shell utilising aggregation pipelines, map-reduce, or other query operators to respond to the specified inquiries about New Yorkers' culinary preferences, market prospects, and rivals.
5. After ensuring that the cluster is operating properly and responding to queries, I add block comments to the script files to record the actions taken and to describe what occurs when each script is being executed.

I set up a simulated distributed MongoDB cluster on my machine by running the scripts; I then import data, partition the collection, establish replica sets, and query the data to analyze information about New York restaurants.

Prior to making any edits or executing the scripts, its important to routinely save the progress and backup any vital data.