CSCE 5300 Introduction to Big Data and DataScience
ICE-6

Lesson Title: Cassandra

Lesson Description: Independent Column based No SQL Tool - Cassandra

Lesson Overview:

- Cassandra Introduction
- Strengths
- Architecture
- DataTypes
- Commands

In Class Exercise

Cassandra (CQL Command Line)

Write a document and include your CQL screen shots for the commands below. Use the datasets provided to you in class (employees data) and create a KEYSPACE with the name "testICE6" or use appropriate KEYSPACE according to your choice. You have the flexibility to change data according to your requirements.

To create KEYSPACE :

CREATE KEYSPACE KEYSPACENAME WITH REPLICATION = {'class' : 'SimpleStrategy'
,'replication_factor' : 3 }

Two key parameters to take note in the class and the replication factor. The class defines the replication strategy i.e. algorithm used to determine the placement of the replicas. The replication factor determines the number of replicas.

DESC KEYSPACES

Execute the above command to check if keyspace is created or not.

USE KEYSPACENAME

Next, let's connect to the created keyspace with the help of the USE command. The USE command enables switching context between keyspaces.

ICE Questions:

Write queries to answer the following questions. Your answer should include the Query the output and an explanation of how/why you drafted the query.

1. What is the total number of employees in the dataset?
2. What is the average salary of all employees in the Sales department?
3. Which employee has the highest salary in the Marketing department?
4. What is the average number of years that employees in the Engineering department have been with the company?
5. How many employees have a job title of "teamlead"?
6. Which manager has the most employees reporting to them?
7. How many employees have a manager with a salary greater than $50,000?
8. Update the department of employee with employee_id 3 to "Sales".
9. Delete all employees who have a salary less than $30,000.
10. Select the total number of employees who work in the Marketing department.

ICE Submission Guidelines

1. ICE Submission is individual.
2. The documentation should include screenshots of your code/queries and results.
3. Provide an explanation of the exercise for each question as per your understanding.
4. The similarity score for your document should be less than 15%.
5. Submit the source code (if any) properly commented and documentation (.pdf/.doc) with explanation and screenshot of source code/queries having input logic and output results.
6. Submission after the deadline is considered as late submission.