SDJ INTERNATIONAL COLLEGE

Lab Assignment of Big Data and Hadoop

Sr.No.	Problems
1	**Perform the following operations using MongoDB:** 1. **Create**: Insert multiple documents into a collection called employees. Each document should have the following fields: name, position, and salary. Insert the following employees: a. name: "Alice", position: "Manager", salary: 70000 b. name: "Bob", position: "Developer", salary: 55000 c. name: "Charlie", position: "HR", salary: 45000 2. **Read**: Query the employees collection to find all employees with a salary greater than 50000. 3. **Update**: Increase the salary of all employees whose position is "Developer" by 10%. 4. **Delete**: Remove the employee named "Charlie" from the employees collection.
2	**Write a MongoDB script to perform the following CRUD operations:** 1. **Create**: Insert a new document into a collection called students. The document should include the following fields: name, age, and course. Insert a document with values: o name: "John Doe" o age: 22 o course: "Computer Science" 2. **Read**: Retrieve all documents from the students collection where the age is greater than 20. 3. **Update**: Update the course field of the student named "John Doe" to "Data Science". 4. **Delete**: Delete all documents from the students collection where the age is less than 21.
3	**A. Write MongoDB commands to perform the following operations on a collection named books:** 1. **Create**: Insert a document representing a book with the fields title, author, publishedYear, and copiesSold. Use the following data: o title: "The Alchemist" o author: "Paulo Coelho" o publishedYear: 1988 o copiesSold: 65000000 2. **Read**: Find all books in the books collection that were published before the year 2000. 3. **Update**: Add a new field rating with a value of 4.8 to the document where

	the title is "The Alchemist". 4. **Delete**: Delete all books from the collection where copiesSold is less than 1 million.
4	Program to calculate average weather temperature of city from csv file
5	Write a MapReduce program that reads a CSV file containing sales transactions with columns: transaction_id, product_id, quantity, and price. Calculate the total revenue generated for each product and output it as product_id, total_revenue.
6	Given a CSV file with columns: movie_id, user_id, rating, write a MapReduce program to calculate the average rating for each movie. Output the results as movie_id, average_rating.
7	**Calculating Maximum and Minimum Temperature** • Using a weather dataset with columns: city, date, temperature, write a MapReduce program to find the maximum and minimum temperature recorded for each city. • Output the results as city, max_temperature, min_temperature.
8	Create a database named company_db and within it, create a table called employees with the following columns: employee_id (INT), name (STRING), department (STRING), and salary (FLOAT). Ensure the table stores data in text format. Query: 1. Write a Hive query to retrieve all employee details from the employees table where the department is IT. 2. Write a Hive query to delete all records from the employees table where the salary is below 30000. 3. Update the location column for all employees whose department is HR to Head Office. 4. Delete the department record where department_id = 3.
9	Create a database named university_db and within it, create a table called students with the following columns: student_id (INT), name (STRING), course (STRING), and marks (FLOAT). Ensure the table stores data in text format. **Query:** 1. Write a Hive query to retrieve all student details from the students table where the course is Computer Science. 2. Write a Hive query to delete all records from the students table where the marks are below 50. 3. Update the marks column for all students whose course is Mathematics by adding 10 bonus points. 4. Delete the record from the students table where the student_id = 102.
10	Create a database named retail_db and within it, create a table called products with the following columns: product_id (INT), product_name (STRING), category (STRING), and price (FLOAT). Ensure the table stores data in text format.

	Query: 1. Write a Hive query to retrieve all product details from the products table where the category is Electronics. 2. Write a Hive query to delete all records from the products table where the price is below 100. 3. Update the price column for all products in the Furniture category by increasing it by 5%. 4. Delete the product record from the products table where the product_id = 5.
11	Create a database named hospital_db and within it, create a table called patients with the following columns: patient_id (INT), patient_name (STRING), age (INT), and disease (STRING). Ensure the table stores data in text format. **Query:** 1. Write a Hive query to retrieve all patient details from the patients table where the disease is Diabetes. 2. Write a Hive query to delete all records from the patients table where the age is below 18. 3. Update the disease column for all patients whose disease is Hypertension to Chronic Hypertension. 4. Delete the record from the patients table where the patient_id = 8.