

CSC 343, Exam II Review

Exam II Notes

- You may bring one page of notes (front and back) to the exam. This page may be handwritten or typed.
- Computer access will not be permitted during the exam.
- Cell phones must be put away at all times.
- Don't hesitate to contact me if you have any questions!

Exam II Outline

- MySQL – queries (SELECT statements) that may contain any of the following:
 - WHERE clause
 - ORDER BY (desc)
 - GROUP BY
 - INNER JOIN
 - Aliases
 - Aggregate functions (COUNT, SUM, MIN, MAX, AVG)
 - DISTINCT
 - LIMIT
- Sqoop
 - Importing all tables
 - Importing a single table
 - Importing selected columns and rows from a single table
 - Importing the results of a query
 - Changing the field terminator
 - Changing the number of mappers
 - Selecting a column to determine how the mappers split up the work
- Hive/Impala
 - Conceptual differences between the two
 - Performing queries using SQL (all MySQL concepts above)
 - Syntax for queries involving maps and structures
 - Table creation using DDL
 - CREATE statement (with options to change table location and default field delimiter)
 - External vs. Managed (Internal) tables
 - Changing column names
 - Inserting data into a table
 - Updating Impala Metadata

Exam II Practice

Use the mySQLtutorial (<http://www.mysqltutorial.org/tryit/>) sample database to answer the following questions:

1. Find the total payment amount for each customer (use column names of *customer_number* and *total*, and display the corresponding values)
2. Find the average payment amount for each customer (use column names of *customer_name* and *total*, and display the corresponding values).
3. Output the customer name and phone number for all customers who have spent more than \$100,000.

Use Hadoop (Sqoop/Hive/Impala) to answer the remaining questions:

4. Import the *auth_user* table from the *hue* database into the HDFS directory `/user/cloudera/hue/auth_user`.
5. Import the *id*, *username*, and *last_login* columns from the *auth_user* table into the HDFS directory `/user/cloudera/hue/user_info`
6. Import the *id*, *username*, and *last_login* columns from the *auth_user* table into the HDFS directory `/user/cloudera/hue/active/auth_user`, for active users only.
7. Using Impala or Hive, and the sample customers table, display the number of customers with an e-mail preference of HTML.
8. Using Impala, output a list of customer names and shipping address (city, state, zip) for customers whose shipping address is in Indiana (IN).