

Complete the following tasks associated with data-pipeline.

8 points/task

(Note: don't run queries directly on the database, instead run them through/using Python).

Download the '*Employees.csv*' file available on blackboard.

1. Write a python code to read the *employees.csv* file into a data-frame. Loop through each row in the data-frame, and insert/store the records into a table of sqlite3 database.
2. In the table (from step-1), drop at-least 5 employees, update (some) information of at-least 5 employees, and then add at-least 5 new employees. Perform at-least one analysis task using group-by clause.
3. Read the Employees-data of sqlite3 table (from step-2 above) into a data-frame. Loop through each row in the data-frame, and convert the rows into documents. Store each of these documents into a mongodb-collection.
4. In the collection (from step-3), drop at-least 5 employees, update (some) information of at-least 5 employees, and add at-least 5 new employees. Perform at-least one analysis task using group-by clause.
5. Read the Employees-data of sqlite3 table (from step-2 above), and write it to a csv file (comma-delimited). Similarly, read each document of the mongodb-collection (from step-4 above), and write it to a text file in JSON (key-value) format.

Submission:

1. Make sure to include/print all the **output-data** in the notebook.
2. Either submit the notebook-file (***.ipynb**) or copy your python-code and output to word/pdf document, and upload the document file through the submission link provided on blackboard.