Graphical user interface

Description automatically generated

**Assignment 3- Hive & Sqoop**

**Task 2: Week 5 – Sqoop**

**Objective:** Use Sqoop to transfer data to MySQL from Hadoop or vice versa

**Task:**

1. Import a table from MySQL to HDFS (Choose the parameter as you wish)
2. Demonstrate export command

* Troubleshooting is part of the evaluation. Please research and troubleshoot your issue.

**Note:** All screen shots must have time stamp otherwise you get zero

**Task 2: Week 6: Hive**

1. Upload Book.xml in HDFS then create a table in Hive and parse it.
2. RUN Three queries (on your choice) i.e run a query to display all authors.

**Note:** Processing XML data in Hive using XML SerDe XML has been one of the most important data structures and has been used for quite a long time for data transfers and storage. Parsing XML data and then processing it is always a tricky task as parsing XML is one of the costliest operations. Hive does not have any built-in support for XML data processing, but many organizations and individuals have made open-source contributions to XML SerDe .

**Hint:** If you are using EMR and its same version as mine which I posted the recording then you can use the attached jar file otherwise you need to research and find the right version. There are various XML SerDe that have been made available by open-source developers. (<https://search.maven.org/>)

**Note**

1. You can use any platform
2. Troubleshooting is part of the evaluation. Please research and troubleshoot your issue.

**Part 2 : Deliverable**

1. Its group based. One report per group
2. Must be PDF format.
3. Format of the report as following
   1. Table of contents
   2. Body of report (which has screen shots **with explanation and timestamp. These screen shots proof that you have done the activity successfully)**
   3. Achievement (Write a paragraph and explain what you have learned in this activity)
   4. All you AWS screen shots must have your login name (top right corner)
4. Follow the rubrics.
5. All screen shots must have time stamp otherwise you do not get mark
6. There will be 10% deduction for each day delay. After three days I do not accept the assignment and you get zero.
7. Whoever is uploading must complete the following table. If this table is not completed, you do not get mark for the assignment. This is to make sure everyone participates in the activity. Keep it at the end of PDF

|  |
| --- |
| **Name of students who has participated in the assignment.** |
| **Student name:** |
| **Student name:** |
| **Student name:** |
| **Student name:** |
| **Student name:** |

|  |
| --- |
| **Name of students who has not participated in the assignment.** |
| **Student name:** |
| **Student name:** |
| **Student name:** |
| **Student name:** |
| **Student name:** |