1] Understanding Hive tables

1) Download the attached file in **STAGING\_AREA**

2) Extract it

3) Create a folder **LABS\_AREA**/lab7.1

4) Put the extracted contents in above location

5) Wait for the instructions.

* Code/Dataset

data/Lab7.1.rar

2] Understanding partitions

1) Download the RAR in **STAGING\_AREA.**

2) Extract the contents

3) Pull the contents in **LABS\_HOME**/demos

4) Refer the lab manual

5) Wait for the instructions

* Code/Dataset

data/partitionandskewtables.rar

3] Bucketed Table

1) Download the dataset in **STAGING\_AREA.**

**2)** Pull the contents in **LABS\_HOME**/demos

**3)**  Wait for the instructions

* Code/Dataset

data/dept\_loc.txt

4] Analyzing BigData with Hive

1) Refer the lab manual

2) Follow the instructions in lab manual

3) Translate the paths given in the lab manual to paths applicable for the cloudera sandbox.

**Note -** No data for this lab, the data needed is already there with you through previous labs.

**Timeline =** 60 Minutes.

Ref - <https://cwiki.apache.org/confluence/display/hive/languagemanual+udf>

5] Writing a Hive UDF

1) Extract the rar in **STAGING\_AREA**

2) Create a folder **LABS\_AREA**/labs/lab11.1

3) Put the extracted contents in above folder

4) Refer the lab manual

5) Wait for the instructions from an Instructor

* Code/Dataset

data/HiveUDF.rar

6] Computing ngrams

1) Refer the lab manual

2) Wait for the instructions

<https://cwiki.apache.org/confluence/display/hive/languagemanual+udf>

7] Joining data in Hive

1) Refer lab manual

2) Follow instructions in the lab manual and implement it.

Time Line = 30 Minutes

* Code/Dataset

data/Lab7.4.rar

8] Computing ngrams of Emails in Avro Format

1) Download the rar in **STAGING\_AREA**

2) Move the extracted contents to **LABS\_AREA**/lab7.5

3) Refer the lab manual

4) Wait for the instructions

* Code/Dataset

data/Lab7.5.rar

9] Advanced Hive Programming

1) Download the rar on STAGING\_AREA

2) Move the extracted contents to LABS\_AREA/lab9.1

3) Refer the lab manual

* Code/Dataset

data/Lab9.1.rar

* TimeLine = 60 Minutes

10] Skipping Headers

1. Download the attached file in LABS\_HOME/demos
2. Wait for the instructions

* Code/Dataset

data/log2.txt

Some more links related to Hive

11) Hive Complex Data Types with Examples

<http://bigdatariding.blogspot.com/2014/02/hive-complex-data-types-with-examples.html>

12) Custom Map/Reduce Scripts

<https://cwiki.apache.org/confluence/display/Hive/Tutorial#Tutorial-CustomMap/ReduceScripts>

13) LanguageManual DDL

<https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-CreateTableCreate/Drop/TruncateTable>

LanguageManual DML

<https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML>

LanguageManual UDF

<https://cwiki.apache.org/confluence/display/hive/languagemanual+udf>

Hadoop different file formats

<https://www.linkedin.com/pulse/hadoop-file-formats-when-what-use-rajesh-dangi-pmp/>