Session 6

Assignment 1

|  |  |
| --- | --- |
| **Prepared For:** | AcadGild |
|  |  |
| **Document Approval:** | **AcadGild** |
|  |  |
|  |  |
|  |  |
|  |  |
| **Project Title:** | Session 6 – Assignment 1 |
|  |  |
| **Prepared By:** | Duncan Burgess |
|  |  |
|  | dburgess@duncb.com |
|  |  |
| **Primary Engineer:** | Duncan Burgess |
|  |  |
| **Document Reference:** | **Session 6 – Assignment 1** |
|  |  |
| **Start Date:** | 20/09/2017 |
|  |  |
|  |  |

# 

# Contents

[Contents 2](#_Toc493666736)

[Change History 3](#_Toc493666737)

[1. Problem Statement 4](#_Toc493666738)

[2. Creating Database and tables in Hive 5](#_Toc493666739)

[2.1. Create the database 5](#_Toc493666740)

[2.2. Create the tables required 5](#_Toc493666741)

[3. Import Data 5](#_Toc493666742)

# Change History

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Document Revision** | **Date** | **Authored By** | **Authorised By** | **Sections Affected** | **Reason for Change** |
| Rev 01 | 20/09/2017 | Duncan Burgess |  | All | Initial release. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# Problem Statement

Create a database named 'custom'.

Create a table named temperature\_data inside custom having below fields:

1. date (mm-dd-yyyy) format

2. zip code

3. temperature

The table will be loaded from comma-delimited file.

Load the dataset.txt (which is ',' delimited) in the table.

**dataset\_Session 14.txt'**

10-01-1990,123112,10

14-02-1991,283901,11

10-03-1990,381920,15

10-01-1991,302918,22

12-02-1990,384902,9

10-01-1991,123112,11

14-02-1990,283901,12

10-03-1991,381920,16

10-01-1990,302918,23

12-02-1991,384902,10

10-01-1993,123112,11

14-02-1994,283901,12

10-03-1993,381920,16

10-01-1994,302918,23

12-02-1991,384902,10

10-01-1991,123112,11

14-02-1990,283901,12

10-03-1991,381920,16

10-01-1990,302918,23

12-02-1991,384902,10

# Creating Database and tables in Hive

To display the database working in I set

***Set hive.cli.print.current.db = true***

## Create the database

*create database custom;*

**Use the database created**

*use custom;*

## Create the tables required

*hive (custom)> create table if not exists temperature\_data*

*> (*

*> date string,*

*> postcode int,*

*> temperature int*

*> )*

*> row format delimited*

*> fields terminated by ',';*

*OK*

*Time taken: 0.056 seconds*

**Check tables**

*hive (custom)> describe temperature\_data;*

*OK*

*date                    string*

*postcode                int*

*temperature              int*

*Time taken: 0.054 seconds, Fetched: 3 row(s)*

# Import Data

*hive (custom)> LOAD DATA*

*> LOCAL INPATH '/home/acadgild/Datasets/dataset\_Session 14.txt'*

*> INTO TABLE temperature\_data;*

*Loading data to table custom.temperature\_data*

*Table custom.temperature\_data stats: [numFiles=1, totalSize=437]*

*OK*

*Time taken: 0.181 seconds*

**Check results**

*hive (custom)> select \* from temperature\_data;*

*OK*

*10-01-1990 123112 10*

*14-02-1991 283901 11*

*10-03-1990 381920 15*

*10-01-1991 302918 22*

*12-02-1990 384902 9*

*10-01-1991 123112 11*

*14-02-1990 283901 12*

*10-03-1991 381920 16*

*10-01-1990 302918 23*

*12-02-1991 384902 10*

*10-01-1993 123112 11*

*14-02-1994 283901 12*

*10-03-1993 381920 16*

*10-01-1994 302918 23*

*12-02-1991 384902 10*

*10-01-1991 123112 11*

*14-02-1990 283901 12*

*10-03-1991 381920 16*

*10-01-1990 302918 23*

*12-02-1991 384902 10*

*Time taken: 0.042 seconds, Fetched: 20 row(s)*