



BIG DATA  
DEVELOPMENT

**ACADGILD**

# Session 8

## HIVE BASICS

### Assignment 1

---

## [Big Data Hadoop and Spark Development](#)

*Assignment 1 – You must perform the given tasks.*

### **Table of Contents**

1.	Introduction	3
2.	Objective	3
3.	Prerequisites:	3
4.	Associated Data Files	3
5.	Problem Statement	3
6.	Expected Output	4
7.	Approximate Time to Complete Task	4

## 1. Introduction

In this assignment, you need to perform the given tasks.

## 2. Objective

This assignment will help you to consolidate the concepts learnt in the session 4.

## 3. Prerequisites:

None

## 4. Associated Data Files

<https://drive.google.com/file/d/0Bxr27gVaXO5sa0JBamZXdkpYUFk/view?usp=sharing>

## 5. Problem Statement

### Task 1

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.

### Task 2

- Fetch date and temperature from temperature\_data where zip code is greater than 300000 and less than 399999.
- Calculate maximum temperature corresponding to every year from temperature\_data table.
- Calculate maximum temperature from temperature\_data table corresponding to those years which have at least 2 entries in the table.
- Create a view on the top of last query, name it temperature\_data\_vw.
- Export contents from temperature\_data\_vw to a file in local file system, such that each file is '|' delimited.

## **6. Expected Output**

Solution report with commands, explanation to commands and screenshot for output. Report shall be in PDF format. Submitted in GitHub

## **7. Approximate Time to Complete Task**

6 hours.