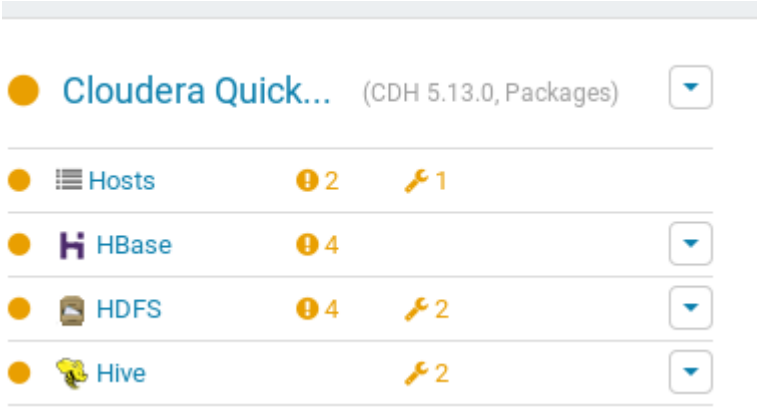
Practical 5 - Install Hive and use Hive Create and store structureddatabases.

Step 1: Start Cloudera

Ensure that your Cloudera services HIVE , HDFS is up and running.



Step 2: Create Employee File

Create a file named employee.txt with the following content:

```
[cloudera@quickstart ~]$ cat > /home/cloudera/employee.txt
1~Sachine~Pune~Product Engineering~100000~Big Data
2~Gaurav~Bengalore~Sales~90000~CRM
3~Manish~Chennai~Recruiter~125000~HR
4~Bhushan~Hyderabad~Developer~50000~BFSI^Z
```

Step 3: View the Create file

```
[cloudera@quickstart ~]$ cat /home/cloudera/employee.txt
1~Sachine~Pune~Product Engineering~100000~Big Data
2~Gaurav~Bengalore~Sales~90000~CRM
3~Manish~Chennai~Recruiter~125000~HR
```

Step 4: List All Files in HDFS

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 9 items
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 10:29 /benchmarks
drwxrwxrwx   - hbase    supergroup          0 2024-02-21 22:53 /hbase
drwxrwxrwx   - hdfs     supergroup          0 2024-02-22 00:50 /inputdirectory
drwxrwxrwx   - hdfs     supergroup          0 2024-02-18 03:18 /mapreduce
drwxrwxrwx   - solr     solr                0 2017-10-23 10:32 /solr
drwxr-xr-x   - cloudera supergroup          0 2024-02-21 22:41 /sqoop_import_dat
a
drwxrwxrwt   - hdfs     supergroup          0 2024-01-06 06:56 /tmp
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 10:31 /user
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 10:31 /var
```

Step 5: Create HDFS Directory(in my case already created)

```
[cloudera@quickstart ~]$ sudo -u hdfs hadoop fs -mkdir /inputdirectory
mkdir: `/inputdirectory': File exists
```

Step 6: List HDFS Directory(input directory should be present)

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
```

Step 7: Set Permissions for the Directory

```
[cloudera@quickstart ~]$ sudo -u hdfs hadoop fs -chmod -R 777 /inputdirectory

[cloudera@quickstart ~]$
```

Step 8: Verify Permissions(ignore)

HDFS DFS -LS /

Step 9: Move the File to HDFS

```
[cloudera@quickstart ~]$ sudo -u hdfs hadoop fs -put /home/cloudera/employee.txt
 /inputdirectory
put: `/inputdirectory/employee.txt': File exists
```

Step 10: Check File in HDFS

```
[cloudera@quickstart ~]$ hdfs dfs -ls /inputdirectory
Found 1 items
-rwxrwxrwx   1 hdfs supergroup        123 2024-02-22 00:50 /inputdirectory/emplo
yee.txt
```

Step 11: Read File in HDFS

```
[cloudera@quickstart ~]$ hadoop fs -cat /inputdirectory/employee.txt
1~Sachine~Pune~Product Engineering~100000~Big Data
2~Gaurav~Bengalore~Sales~90000~CRM
3~Manish~Chennai~Recruiter~125000~HR
[cloudera@quickstart ~]$ hdfs dfs -ls /
```

Step 12: Enter Hive Shell

```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.13.0.jar!/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive>
```

Step 13: Show Databases in Hive

```
hive> show databases;
OK
default
organization
Time taken: 1.541 seconds, Fetched: 2 row(s)
```

Step 14: Create Database

```
hive>  create database organisation;
OK
```

Step 15: Show Databases Again

```
hive> show databases;
OK
default
organisation
organization
Time taken: 0.021 seconds, Fetched: 3 row(s)
```

Step 16: Change Database

```
hive> use organisation;
OK
```

Step 17: Create Table in Hive

```
hive> CREATE TABLE employee (
    >     name STRING,
    >     city STRING,
    >     department STRING,
    >     salary INT,
    >     domain STRING
    > ) ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY '~';
OK
Time taken: 0.402 seconds
```

Step 18: Show Tables in Hive

```
hive> show tables;
OK
employee
Time taken: 0.078 seconds, Fetched: 1 row(s)
```

Step 19: Read Content from Employee Table (Empty)

```
hive> select * from employee;
OK
Time taken: 0.455 seconds
```

Step 20: Load Data into Employee Table

```
hive> load data inpath '/inputdirectory/employee.txt' overwrite into table employee;
Loading data to table organisation.employee
Table organisation.employee stats: [numFiles=1, numRows=0, totalSize=123, rawDataSize=0]
OK
Time taken: 0.497 seconds
```

Step 21: Read Content from Employee Table (With Data)

```
hive>  select * from employee;
OK
1       Sachine Pune    NULL    100000
2       Gaurav  Bengalore       NULL    90000
3       Manish  Chennai NULL    125000
Time taken: 0.078 seconds, Fetched: 3 row(s)
hive>
```