

step 1:

create database named custom using hive command prompt:

```
Disha — java -Xmx256m -Djava.net.preferIPv4Stack=true -Djava.security.krb5.realm= -Djava.security.krb5.kdc= -Dhadoop.log.dir=/usr/local/Cell...
java -Xmx10...2.jar -x local    java -Xmx10...2.jar -x local    ~ -- -bash    ~ -- java -x spark-shell    java -Xmx25...e.cli.CliDriver    +
Time taken: 0.431 seconds
hive> show databases
> ;
OK
acadmild_db
custom
default
Time taken: 0.056 seconds, Fetched: 3 row(s)
hive> use custom;
OK
Time taken: 0.144 seconds
hive> show tables;
OK
temperature_data
Time taken: 0.419 seconds, Fetched: 1 row(s)
hive> CREATE DATABASE custom;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.q1.exec.DDLTask. Database custom already exists
hive> DELETE DATABASE custom;
Usage: delete [FILE|JAR|ARCHIVE] <value> [<value>]*
Query returned non-zero code: 1, cause: null
hive> DROP DATABASE IF EXISTS custom CASCADE;
Fri Sep 15 13:23:42 IST 2017 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Fri Sep 15 13:23:42 IST 2017 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Fri Sep 15 13:23:42 IST 2017 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
OK
Time taken: 7.871 seconds
hive> show databases;
OK
acadmild_db
default
Time taken: 0.135 seconds, Fetched: 2 row(s)
hive> CREATE DATABASE custom;
OK
Time taken: 1.134 seconds
hive> █
```

2. copy dataset.txt to hadoop using put command and then check the data in the hadoop path using cat command:

```
java -Xmx10...2.jar -x local    java -Xmx10...2.jar -x local    ~ -- -bash    ~ -- java - spark-shell    java -Xmx25...e.cli.CliDriver    +

[-rmdir [--ignore-fail-on-non-empty] <dir> ...]
[-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>]|[--set <acl_spec> <path>]]
[-setfattr {-n name [-v value] | -x name} <path>]
[-setrep [-R] [-w] <rep> <path> ...]
[-stat [format] <path> ...]
[-tail [-f] <file>]
[-test [-d] <path>]
[-text [-ignoreCrc] <src> ...]
[-touchz <path> ...]
[-truncate [-w] <length> <path> ...]
[-usage [cmd ...]]

Generic options supported are
-conf <configuration file>      specify an application configuration file
-D <property=value>             use value for given property
-fs <file:///hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
-jt <local|resourceManager:port> specify a ResourceManager
-files <comma separated list of files> specify comma separated files to be copied to the map reduce cluster
-libjars <comma separated list of jars> specify comma separated jar files to include in the classpath.
-archives <comma separated list of archives> specify comma separated archives to be unarchived on the compute machines.

The general command line syntax is
command [genericOptions] [commandOptions]

Chaitanyas-MacBook-Pro:~ Disha$ hadoop fs -cat /hive_data/temperature_data.txt
17/09/15 13:31:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
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
Chaitanyas-MacBook-Pro:~ Disha$
```

3. 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.

commad for table creation:

create external table temperature\_data

```
(
dt string,
zipcode BIGINT,
temperature int
)
row format delimited
fields terminated by ','
location '/hive_data/';
```

```

Time taken: 7.871 seconds
[hive> show databases;
OK
acadgild_db
default
Time taken: 0.135 seconds, Fetched: 2 row(s)
[hive> CREATE DATABASE custom;
OK
Time taken: 1.134 seconds
[hive> create external table temperature_data
> (
> dt string,
> zipcode BIGINT,
> temperature int
> )
> row format delimited
> fields terminated by ','
[> location '/hive_data/';
OK
Time taken: 1.13 seconds
[hive> show databases;
OK
acadgild_db
custom
default
Time taken: 0.122 seconds, Fetched: 3 row(s)
[hive> show create table temperature_data;
OK
CREATE EXTERNAL TABLE `temperature_data` (
  `dt` string,
  `zipcode` bigint,
  `temperature` int)
ROW FORMAT SERDE
  'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'
WITH SERDEPROPERTIES (
  'field.delim'=',',
  'serialization.format'=',')
STORED AS INPUTFORMAT
  'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION
  'hdfs://localhost:9000/hive_data'
TBLPROPERTIES (
  'numFiles'='1',
  'totalSize'='419',
  'transient_lastDdlTime'='1505462696')
Time taken: 0.25 seconds, Fetched: 19 row(s)

```

---