

```
[training@localhost ~]$ sudo su
[root@localhost training]# hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
Hive history file=/tmp/root/hive_job_log_root_202401080322_1285639828.txt
hive> create database crime rate
> create database crime rate;
FAILED: ParseException line 1:22 mismatched input 'rate' expecting EOF near 'cr
ime'

hive> create database crimerate_data
> create database crimerate_data;
FAILED: ParseException line 2:0 mismatched input 'create' expecting EOF near 'cr
imerate_data'

hive> create database crimerate_data;
[1]: Stopped hive
[root@localhost training]# hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
Hive history file=/tmp/root/hive_job_log_root_202401080322_1622643534.txt
hive> show databases;
OK
default
Time taken: 3.065 seconds
hive> create database crime;
OK
Time taken: 0.698 seconds
hive> use crime
> ;
OK
Time taken: 0.023 seconds
hive> create table nor_table(in int, off int, grp string, des string, dist string, rep int, sho string, occ string, year stri
ng);
FAILED: ParseException line 1:23 mismatched input 'in' expecting Identifier near '(' in column specification

hive> create table nor_table(in int, off int, grp string, des string, dist string, rep int, sho string, occ string, yea string, month int,
Display all 409 possibilities? (y or n)
!
(
<
ALTER
BY
DATETIME
EXTENDED
INPATH
LIKE
OF
PARTITIONS
SERDE
TABLES
TRUE
UNION
acos(
asc
binary
ceil(
context_ngrams(
date
=
)
<=>
AND
CAST
DELIMITED
EXTERNAL
INPUTFORMAT
LIMIT
ON
PARTITIONS
SERDE
TABLESAMPLE
TRUE
UNION
acos(
asc
binary
ceil(
context_ngrams(
date
SELEMS
);
<=>
ARRAY
CLUSTER
DESC
FALSE
INSERT
LINES
OR
REDUCE
SET
TBLPROPERTIES
USING
all
asin(
boolean
clustered
conv(
date_add(
SELEMS
);
<=>
ARRAY
CLUSTER
DESC
FALSE
INSERT
LINES
OR
REDUCE
SET
TBLPROPERTIES
USING
all
asin(
boolean
clustered
conv(
date_add(
KEYS
AS
CLUSTERED
DIRECTORIES
FIELDS
INT
LOAD
ORDER
REGEXP
SHOW
TEMPORARY
WHERE
alter
assert true(
boolean(
coalesce(
cos(
datediff(
VALUES
+
=
ASC
COLLECTION
DIRECTORY
FLOAT
INTO
LOCAL
OUTER
REPLACE
SMALLINT
SORT
TERMINATED
WITH
and
atan(
avg(
buckets
collection
count(
day(
Selems$
-
=
BINARY
COLUMNS
DISTINCT
FORMAT
IS
LOCATION
OUTER
RIGHT
SORTED
TEXTFILE
TIMESTAMP
\
array
between(
by
columns
covar_samp(
dayofmonth(
$values$
>=
BOOLEAN
CREATE
DISTRIBUTE
FULL
JOIN
MCKK
OVERWRITE
ROW
STORED
TINYINT
TO
^
array contains(
bigint(
case(
concat(
create union(
degrees(
%
/
ADD
BUCKET
DATA
DROP
FUNCTION
KEYS
NOT
PARTITION
SELECT
STRING
TRANSFORM
abs(
as
bin(
ceil(
concat_ws(
data
desc
&
:
ALL
BUCKETS
DATE
EXPLAIN
GROUP
LEFT
NULL
PARTITIONED
SEQUENCEFILE
TABLE
TRANSFORM
```

```
Applications | File Edit View VM Jobs Help | CentOS 7 64-bit | Clone of Cloudera-Utility-T... | Mon Jan 8, 4:26 AM
root@localhost: /home/training

hive> show tables;
OK
nor_data
Time taken: 0.069 seconds
hive> load data local inpath '/crime.csv' into nor_data;
FAILED: ParseException line 1:41 mismatched input 'nor_data' expecting TABLE near 'into' in load statement

hive> load data local inpath '/crime.csv' into table nor_data;
FAILED: SemanticException Line 1:23 Invalid path ''/crime.csv'': No files matching path file:/crime.csv
hive> load data local inpath '/crime.csv' into table nor_data;
FAILED: SemanticException Line 1:23 Invalid path ''/crime.csv'': No files matching path file:/crime.csv
hive> load data local inpath '/crime.csv' into table nor_data;
Copying data from file:/crime.csv
Copying file: file:/crime.csv
Loading data to table crime.nor_data
OK
Time taken: 1.949 seconds
hive> create external table ext_data(INC_NUMBER int,OFF_CODE int,OFF_CODE_GROUP string,OFFENSE_DES string,DIST string,REP_AREA int,OCU_DATE TIMESTAMP,Moun int,DOW string,UCR string,Lat int,Lon int,loc string)row format delimited fields
terminated by ','location '/';
OK
Time taken: 0.04 seconds
hive> show tables;
OK
ext_data
nor_data
Time taken: 0.057 seconds
hive> create external table ext_data(INC_NUMBER int,OFF_CODE int,OFF_CODE_GROUP string,OFFENSE_DES string,DIST string,REP_AREA int,OCU_DATE TIMESTAMP,Moun int,DOW string,UCR string,Lat int,Lon int,loc string)row format delimited fields
load data local inpath '/crime.csv' into table ext_data;
Copying data from file:/crime.csv
Copying file: file:/crime.csv
Loading data to table crime.ext_data
OK
Time taken: 1.391 seconds
hive> create table static_data(INC_NUMBER int,OFF_CODE int,OFF_CODE_GROUP string,OFFENSE_DES string,DIST string,REP_AREA int,OCU_DATE TIMESTAMP,Moun int,DOW string,Lat int,Lon int,loc string)partitioned by(UCR string) row format delimit
ed fields terminated by ',';
OK
Time taken: 0.028 seconds
hive> describe formatted static_data;
OK
# col_name          data_type          comment
inc_number          int                None
off_code            int                None
off_code_group      string             None
offense_des         string             None
dist                string             None
rep_area            int                None
occu_date           timestamp          None
moun                int                None
dow                 string             None
lat                 int                None
lon                 int                None
loc                 string             None
```

```
Applications | File Edit View VM Jobs Help | CentOS 7 64-bit | Clone of Cloudera-Utility-T... | Mon Jan 8, 4:26 AM
root@localhost: /home/training

lon int None
loc string None

# Partition Information
# col_name data_type comment
ucr string None

# Detailed Table Information
Database: crime
Owner: root
CreateTime: Mon Jan 08 04:23:53 EST 2024
LastAccessTime: UNKNOWN
Protect Mode: None
Retention: 0
Location: hdfs://localhost:8020/user/hive/warehouse/crime.db/static_data
Table Type: MANAGED_TABLE
Table Parameters:
  transient_lastDdlTime 1704705833

# Storage Information
SerDe Library: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat: org.apache.hadoop.mapred.TextInputFormat
OutputFormat: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed: No
Num Buckets: -1
Bucket Columns: []
Sort Columns: []
Storage Desc Params:
  field.delim
  serialization.format
hive> create table dai_data(INC_NUMBER int,OFF_CODE int,OFF_CODE_GROUP string,OFFENSE_DES string,DIST string,REP_AREA int,OCCE_DATE TIMESTAMP,Moun int,DOW string,Lat int,Lon int,loc string)partitioned by(UCR string) row format delimited
fields terminated by ',';
OK
Time taken: 0.051 seconds
hive> show tables;
OK
dai_data
ext_data
nor_data
static_data
Time taken: 0.044 seconds
hive> describe formatted dai_data;
OK
# col_name data_type comment
inc_number int None
off_code int None
off_code_group string None
offense_des string None
dist string None
rep_area int None
occe_date timestamp None
moun int None
dow string None
lat int None
lon int None
loc string None
```

```
Applications | File Edit View VM Jobs Help | CentOS 7 64-bit | Clone of Cloudera-Utility-T... | Mon Jan 08, 4:26 AM
root@localhost: /home/training
File Edit View Search Terminal Help
fields terminated by ',';
OK
Time taken: 0.051 seconds
hive> show tables;
OK
dai_data
ext_data
nor_data
static_data
Time taken: 0.044 seconds
hive> describe formatted dai_data;
OK
# col_name      data_type      comment
inc number      int            None
off code        int            None
off code group  string         None
offense_des     string         None
dist            string         None
rep area        int            None
occu date       timestamp     None
noun            int            None
dow             string         None
lat             int            None
lon             int            None
loc             string         None

# Partition Information
# col_name      data_type      comment
ucr             string         None

# Detailed Table Information
Database:        crime
Owner:           root
CreateTime:      Mon Jan 08 04:25:00 EST 2024
LastAccessTime:  UNKNOWN
Protect Mode:    None
Retention:       0
Location:        hdfs://localhost:8020/user/hive/warehouse/crime.db/dai_data
Table Type:      MANAGED_TABLE
Table Parameters:
    transient_lastDdlTime    1704705900

# Storage Information
SerDe Library:   org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:     org.apache.hadoop.mapred.TextInputFormat
OutputFormat:    org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed:      No
Num Buckets:     -1
Bucket Columns:  []
Sort Columns:    []
Storage Desc Params:
    field.delim      ,
    serialization.format
Time taken: 0.054 seconds
hive>
```



