Scott Berry

Hive DDL Operations

```
hive> CREATE TABLE access.access (id INT, first STRING, last STRING);
OK
Time taken: 0.506 seconds
```

```
hive> show tables from access;
OK
access
Time taken: 0.07 seconds, Fetched: 1 row(s)
```

```
hive> show tables from access '.*s';
OK
access
Time taken: 0.093 seconds, Fetched: 1 row(s)
```

```
hive> describe access.access;
OK
id                      int
first                   string
last                    string
Time taken: 0.099 seconds, Fetched: 3 row(s)
```

```
hive> alter table access.access rename to access.users;
OK
Time taken: 0.065 seconds
```

```
hive> show tables from access;
OK
users
Time taken: 0.069 seconds, Fetched: 1 row(s)
```

```
hive> alter table access.users add columns (email string comment 'email column added');
OK
Time taken: 0.1 seconds
```

```
hive> describe access.users;
OK
id                      int
first                   string
last                    string
email                   string                  email column added
Time taken: 0.074 seconds, Fetched: 4 row(s)
```

```
hive> alter table access.users replace columns (identifier int, first_name string, last_name string);
OK
Time taken: 0.084 seconds
```

```
hive> describe access.users;
OK
identifier              int
first_name              string
last_name               string
Time taken: 0.072 seconds, Fetched: 3 row(s)
```

```
hive> alter table access.users replace columns (identifier int, first_name string);
OK
Time taken: 0.064 seconds
```

```
hive> describe access.users;
OK
identifier              int
first_name              string
Time taken: 0.07 seconds, Fetched: 2 row(s)
```

```
hive> drop table access.users;
OK
Time taken: 0.27 seconds
```

```
hive> show tables from access;
OK
Time taken: 0.071 seconds
```

Hive DML

```
hive> CREATE TABLE access.users (id INT, name STRING, address STRING, city STRING);
```

(Using users CSV from Module 12)

```
hive> LOAD DATA LOCAL INPATH 'Downloads/users.csv' OVERWRITE INTO TABLE access.users;
```

```
hive> CREATE EXTERNAL TABLE ex_users (id INT, name STRING, address STRING, city STRING) LOCATION 'user/hive/users';
```

Querying Data

```
hive> SELECT * FROM access.users;
10331   John    1021 College Ave        Bloomington
74358   Scott   Bird St Washington
62529   Carly   543 Walnut Ave  Bloomington
19338   Jennifer        394 10th St     Bedford
87511   Jessica 967 Woodlawn Ave        Bedford
```

```
hive> SELECT t.id FROM access.users t WHERE t.city='Bloomington';
10331
62529
```

```
hive> INSERT OVERWRITE DIRECTORY 'tmp/hdfs_out' SELECT t.id FROM access.users t WHERE t.city='Bloomington';
```

```
hive> INSERT OVERWRITE LOCAL DIRECTORY '/tmp/hdfs_out' SELECT t.id FROM access.users t WHERE t.city='Bloomington';
```

(Couldn't resolve GROUP BY error)

```
hive> SELECT t.city, count(*) FROM access.users t GROUP BY t.city;
Ended Job = job_local606155403_0010 with errors
Error during job, obtaining debugging information...
FAILED: Execution Error, return code 2 from org.apache.hadoop.hive.ql.exec.mr.MapRedTask
```

(Created a graduation status table)

```
hive> CREATE TABLE access.users2 (id INT, grad STRING);
hive> LOAD DATA LOCAL INPATH 'Downloads/users2.csv' OVERWRITE INTO TABLE access.users2;
```

```
hive> FROM access.users t1 JOIN access.users2 t2 ON (t1.id = t2.id) SELECT t1.id, t1.name, t2.grad;
SLF4J: Found binding in [jar:file:/usr/local/Cellar/hive/3.1.3/libexec/lib/log4j-slf4j-impl-2.17.1.j
SLF4J: Found binding in [jar:file:/usr/local/Cellar/hadoop/3.3.2/libexec/share/hadoop/common/lib/sl
nder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
10331   John    Graduated
74358   Scott   Not Graduated
62529   Carly   Graduated
19338   Jennifer        Not Graduated
87511   Jessica Graduated
```

HiveQL Scripts

```
  ▯   ▯ ~
  hive -f Downloads/script.q
```

```
MapReduce Jobs Launched:
Stage-Stage-3:  HDFS Read: 0 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
10331    John     Graduated
74358    Scott    Not Graduated
62529    Carly    Graduated
19338    Jennifer         Not Graduated
87511    Jessica Graduated
Time taken: 21.964 seconds, Fetched: 5 row(s)
```

```
hive -e "FROM access.users t1 JOIN access.users2 t2 ON (t1.id = t2.id) SELECT t1.id, t1.name, t2.grad"
```

```
MapReduce Jobs Launched:
Stage-Stage-3:  HDFS Read: 0 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
10331    John     Graduated
74358    Scott    Not Graduated
62529    Carly    Graduated
19338    Jennifer         Not Graduated
87511    Jessica Graduated
Time taken: 21.454 seconds, Fetched: 5 row(s)
```

```
hive> CREATE TABLE access.bloom_users (id INT, name STRING, address STRING);
```

```
hive> INSERT OVERWRITE TABLE access.bloom_users SELECT id, name, address FROM access.users where city='Bloomington';
```

```
hive> SELECT * from access.bloom_users;
10331    John     1021 College Ave
62529    Carly    543 Walnut Ave
```

(Unable to resolve error for INSERTs)

```
hive> INSERT INTO TABLE access.bloom_users VALUES (96391, 'Ryan', '6860 N Dunn St' );
Ended Job = job_local869346472_0003 with errors
Error during job, obtaining debugging information...
FAILED: Execution Error, return code 2 from org.apache.hadoop.hive.ql.exec.mr.MapRedTask
```

```
hive> INSERT INTO TABLE access.bloom_users VALUES (76301, 'Julia', '10009 N Washington St' ), (96391, 'Ryan', '6860 N Dunn St' );
Ended Job = job_local642997622_0004 with errors
Error during job, obtaining debugging information...
FAILED: Execution Error, return code 2 from org.apache.hadoop.hive.ql.exec.mr.MapRedTask
```