

```
bmace@bmace-Precision-T1700:~$ cd cassandra
bmace@bmace-Precision-T1700:~/cassandra$ cd apache-cassandra-3.11.0/
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0$ cd bin
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ ./cqlsh
bash: ./cqlsh: No such file or directory
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ ./cqlsh.
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ ./cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.4 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> CREATE KEYSPACE students WITH REPLICATION={
... 'class':'SimpleStrategy',
... 'replication_factor':1};
cqlsh> DESCRIBE KEYSPACES;

students system_schema system_auth system system_distributed system_traces

cqlsh> SELECT * FROM system_schema.keyspaces;
InvalidRequest: Error from server: code=2200 [Invalid query] message='unconfigured table schema_keyspaces'
cqlsh> SELECT * FROM system_schema.keyspaces;

keyspace_name | durable_writes | replication
-----
system_auth | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_distributed | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_traces | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}
students | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}

(6 rows)
cqlsh> USE students;
cqlsh:students> create table Student_info(
... rollno int PRIMARY KEY,
... stdname text,
... DateofJoining timestamp,
... LastExamPercent double);
...
[1]+ Stopped ./cqlsh
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ ./cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.4 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> USE students
...
[2]+ Stopped ./cqlsh
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ ./cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.4 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> DESCRIBE KEYSPACES

students system_schema system_auth system system_distributed system_traces

cqlsh> SELECT * FROM system_schema.keyspaces
... ;

keyspace_name | durable_writes | replication
-----
system_auth | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_distributed | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_traces | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}
students | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}

(6 rows)
cqlsh> USE students
cqlsh:students> create table Student_info(
... rollno int PRIMARY KEY,
... stdname text,
... DateofJoining timestamp,
... LastExamPercent double);
SyntaxException: line 2:12 no viable alternative at input '.' (create table Student_info(
cqlsh:students> create table Student_info(rollno int PRIMARY KEY, stdname text, DateofJoining timestamp, LastExamPercent double);
cqlsh:students> DESCRIBE TABLES;

student_info

cqlsh:students> DESCRIBE TABLE student_info;

CREATE TABLE students.student_info (
rollno int PRIMARY KEY,
dateofjoining timestamp,
lastexampercent double,
stdname text
) WITH bloom_filter_fp_chance = 0.01
AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
AND comment = ''
AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
AND crc_check_chance = 1.0
AND dclocal_read_repair_chance = 0.1
AND default_time_to_live = 0
AND gc_grace_seconds = 864000
AND max_index_interval = 2048
AND memtable_flush_period_in_ms = 0
AND min_index_interval = 128
AND read_repair_chance = 0.0
AND speculative_retry = '99PERCENTILE';

cqlsh:students> BEGIN BATCH
... INSERT INTO
... student_info(RollNo,StdName,DateofJoining,LastExamPercent)
... VALUES (1,'Michael Storm','2012-03-29', 69.6)
... INSERT INTO student_info
... (RollNo,StdName,DateofJoining,LastExamPercent) VALUES
... (2,'Stephen Fox','2013-02-27', 72.5)
... APPLY BATCH;
InvalidRequest: Error from server: code=2200 [Invalid query] message='Undefined column name stdname'
cqlsh:students> SELECT * from student_info;

rollno | dateofjoining | lastexampercent | stdname
-----
1 | 2012-03-28 18:30:00.000000+0000 | 69.6 | Michael Storm
2 | 2013-02-26 18:30:00.000000+0000 | 72.5 | Stephen Fox

(2 rows)
cqlsh:students> script
[3]+ Stopped ./cqlsh
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ script
Script started, file is typescript
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ stopecript
stopecript: command not found
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ script
Script started, file is typescript
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ exit
exit
Script done, file is typescript
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ script output.txt
Script started, file is output.txt
bmace@bmace-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ ./cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.4 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> DESCRIBE KEYSPACES;

students system_schema system_auth system system_distributed system_traces

cqlsh> SELECT * FROM system_schema.keyspaces
... ;

keyspace_name | durable_writes | replication
-----
system_auth | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_distributed | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_traces | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}
students | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}

(6 rows)
cqlsh> USE students;
cqlsh:students> create table Student_Info(
... RollNo int PRIMARY KEY,
... StdName text,
... DateofJoining timestamp,
... LastExamPercent double);
AlreadyExists: Table 'students.student_info' already exists
cqlsh:students> DESCRIBE TABLES;

student_info

cqlsh:students> DESCRIBE TABLE student_info;

CREATE TABLE students.student_info (
rollno int PRIMARY KEY,
dateofjoining timestamp,
lastexampercent double,
stdname text
) WITH bloom_filter_fp_chance = 0.01
AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
AND comment = ''
AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
AND crc_check_chance = 1.0
AND dclocal_read_repair_chance = 0.1
AND default_time_to_live = 0
AND gc_grace_seconds = 864000
AND max_index_interval = 2048
AND memtable_flush_period_in_ms = 0
AND min_index_interval = 128
AND read_repair_chance = 0.0
AND speculative_retry = '99PERCENTILE';

cqlsh:students> SELECT * from student_info;

rollno | dateofjoining | lastexampercent | stdname
```

```
(2 rows)
cqlsh:students> UPDATE student_info SET stdname='David Sheen' WHERE rollno=2 ;
cqlsh:students> SELECT * FROM student_info WHERE rollno in (1,2,3);

rollno | dateofjoining          | lastexampercent | stdname
-----|-----|-----|-----
1 | 2012-03-28 18:30:00.000000+0000 | 69.6 | Michael Storm
2 | 2013-02-26 18:30:00.000000+0000 | 72.5 | David Sheen

(2 rows)
cqlsh:students> INSERT INTO student_info
... (RollNo,StdName,DateofJoining,LastExamPercent) VALUES
... (2,'Stephen Fox','2013-02-27', 72.5);
InvalidRequest: Error from server: code=2200 [Invalid query] message="Undefined column name stdname"
cqlsh:students> INSERT INTO student_info (rollno,stdname,dateofjoining,lastexampercent) VALUES (2,'Stephen Fox','2013-02-27', 72.5);
cqlsh:students> SELECT * from student_info;

rollno | dateofjoining          | lastexampercent | stdname
-----|-----|-----|-----
1 | 2012-03-28 18:30:00.000000+0000 | 69.6 | Michael Storm
2 | 2013-02-26 18:30:00.000000+0000 | 72.5 | Stephen Fox

(2 rows)
cqlsh:students> INSERT INTO student_info (rollno,stdname,dateofjoining,lastexampercent) VALUES (3,'Stephen Fox','2013-02-27', 72.5);
cqlsh:students> SELECT * from student_info;

rollno | dateofjoining          | lastexampercent | stdname
-----|-----|-----|-----
1 | 2012-03-28 18:30:00.000000+0000 | 69.6 | Michael Storm
2 | 2013-02-26 18:30:00.000000+0000 | 72.5 | Stephen Fox
3 | 2013-02-26 18:30:00.000000+0000 | 72.5 | Stephen Fox

(3 rows)
cqlsh:students> INSERT INTO student_info (rollno,stdname,dateofjoining,lastexampercent) VALUES (4,'Stephen Fox','2013-02-27', 72.5);
cqlsh:students> UPDATE student_info SET stdname='David Sheen' WHERE rollno=2 ;
cqlsh:students> SELECT * from student_info;

rollno | dateofjoining          | lastexampercent | stdname
-----|-----|-----|-----
1 | 2012-03-28 18:30:00.000000+0000 | 69.6 | Michael Storm
2 | 2013-02-26 18:30:00.000000+0000 | 72.5 | David Sheen
3 | 2013-02-26 18:30:00.000000+0000 | 72.5 | Stephen Fox
4 | 2013-02-26 18:30:00.000000+0000 | 72.5 | Stephen Fox

(4 rows)
cqlsh:students> CREATE INDEX ON student_info(stdname);
cqlsh:students> SELECT * FROM student_info WHERE stdname='Stephen Fox'
... ;

rollno | dateofjoining          | lastexampercent | stdname
-----|-----|-----|-----
4 | 2013-02-26 18:30:00.000000+0000 | 72.5 | Stephen Fox
3 | 2013-02-26 18:30:00.000000+0000 | 72.5 | Stephen Fox

(2 rows)
cqlsh:students> CREATE INDEX ON student_info(lastexampercent);
cqlsh:students> SELECT rollno,lastexampercent FROM student_info LIMIT 2;

rollno | lastexampercent
-----|-----
1 | 69.6
2 | 72.5

(2 rows)
cqlsh:students> SELECT rollno,stdname AS "Name" FROM student_info
... ;

rollno | Name
-----|-----
1 | Michael Storm
2 | David Sheen
4 | Stephen Fox
3 | Stephen Fox

(4 rows)
cqlsh:students> UPDATE student_info SET stdname="Samaira", lastexampercent=85.7 WHERE rollno=3;
SyntaxException: line 1:141 no viable alternative at input ',' (UPDATE student_info SET stdname=["Samair"]a,...)
cqlsh:students> UPDATE student_info SET stdname="Samaira",lastexampercent=85.7 WHERE rollno=3;
SyntaxException: line 1:141 no viable alternative at input ',' (UPDATE student_info SET stdname=["Samair"]a,...)
cqlsh:students> UPDATE student_info SET stdname="Samaira" lastexampercent=85.7 WHERE rollno=3;
SyntaxException: line 1:142 no viable alternative at input 'lastexampercent' (UPDATE student_info SET stdname=["Samair"]a lastexampercent...)
cqlsh:students> UPDATE student_info
... SET stdname="Samaira" lastexampercent=85.7 WHERE rollno=3;
SyntaxException: line 2:22 no viable alternative at input 'lastexampercent' (UPDATE student_infoSET stdname=["Samair"]a lastexampercent...)
cqlsh:students> UPDATE student_info
SET stdname="Samaira",lastexampercent=85.7
WHERE rollno=3;
cqlsh:students> SELECT rollno,stdname,lastexampercent
FROM student_info
WHERE rollno=3;

rollno | stdname | lastexampercent
-----|-----|-----
3 | Samaira | 85.7

(1 rows)
cqlsh:students> DELETE lastexampercent FROM student_info WHERE rollno=4;
cqlsh:students> SELECT * FROM student_info where rollno=4;

rollno | dateofjoining          | lastexampercent | stdname
-----|-----|-----|-----
4 | 2013-02-26 18:30:00.000000+0000 | null | Stephen Fox

(1 rows)
cqlsh:students> DELETE FROM student_info WHERE rollno=4;
cqlsh:students> SELECT * FROM student_info where rollno=4;

rollno | dateofjoining | lastexampercent | stdname
-----|-----|-----|-----

(0 rows)
cqlsh:students> SELECT * FROM student_info
... ;

rollno | dateofjoining          | lastexampercent | stdname
-----|-----|-----|-----
1 | 2012-03-28 18:30:00.000000+0000 | 69.6 | Michael Storm
2 | 2013-02-26 18:30:00.000000+0000 | 72.5 | David Sheen
3 | 2013-02-26 18:30:00.000000+0000 | 85.7 | Samaira

(3 rows)
cqlsh:students> CREATE TABLE project_details(
... project_id int,
... project_name text,
... stdname text,
... rating double,
... duration int,
... PRIMARY KEY(project_id,project_name)
... );
cqlsh:students> BEGIN BATCH
... INSERT INTO project_details VALUES (1,'MS data migration', 'David Sheen',3.5, 720)
... INSERT INTO project_details VALUES (1,'MS data WarehouseSET stdname="Samaira" lastexampercent=85.7 WHERE rollno=3;', 'David Sheen',3.5, 720)
... APPLY BATCH;
SyntaxException: line 2:28 no viable alternative at input 'VALUES' (BEGIN BATCHINSERT INTO [project_details] VALUES...)
cqlsh:students> BEGIN BATCH
... INSERT INTO project_details VALUES (1,'MS data migration', 'David Sheen',3.5, 720)
... INSERT INTO project_details VALUES (1,'MS data Warehouse', 'David Sheen',3.5, 1440)
... INSERT INTO project_details VALUES (2,'SAP Reporting', 'Stephen Fox',4.2, 3000)
... INSERT INTO project_details VALUES (2,'SAP BI DW', 'Stephen Fox',4, 9000)
... APPLY BATCH;
SyntaxException: line 2:29 no viable alternative at input 'VALUES' (BEGIN BATCH INSERT INTO [project_details] VALUES...)
cqlsh:students> BEGIN BATCH INSERT INTO project_details(project_id, project_name, stdname, rating, duration) VALUES (1,'MS data migration', 'David Sheen',3.5, 720) INSERT INTO project_details(project_id, project_name, stdname, rating, duration) VALUES (1,'MS data Warehouse', 'David Sheen',3.5, 1440)
cqlsh:students> SELECT * FROM project_details;

project_id | project_name | duration | rating | stdname
-----|-----|-----|-----|-----
1 | MS data Warehouse | 1440 | 3.9 | David Sheen
1 | MS data migration | 720 | 3.5 | David Sheen
2 | SAP BI DW | 9000 | 4 | Stephen Fox
2 | SAP Reporting | 3000 | 4.2 | Stephen Fox

(4 rows)
cqlsh:students> SELECT * FROM project_details WHERE project_id=1;

project_id | project_name | duration | rating | stdname
-----|-----|-----|-----|-----
1 | MS data Warehouse | 1440 | 3.9 | David Sheen
1 | MS data migration | 720 | 3.5 | David Sheen

(2 rows)
cqlsh:students> SELECT * FROM project_details WHERE project_id IN (1,2)
ORDER BY project_name DESC;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot page queries with both ORDER BY and a IN restriction on the partition key; you must either remove the ORDER BY or the IN and sort client side, or disable paging for this query"
cqlsh:students> PAGING OFF;
Disabled Query paging.
cqlsh:students> SELECT * FROM project_details WHERE project_id IN (1,2) ORDER BY project_name DESC;

project_id | project_name | duration | rating | stdname
-----|-----|-----|-----|-----
2 | SAP Reporting | 3000 | 4.2 | Stephen Fox
2 | SAP BI DW | 9000 | 4 | Stephen Fox
1 | MS data migration | 720 | 3.5 | David Sheen
1 | MS data Warehouse | 1440 | 3.9 | David Sheen

(4 rows)
cqlsh:students> PAGING ON;
Now Query paging is enabled
Page size: 100
cqlsh:students> exit
bmce@bmce-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ exit
Script done, file is output.txt
bmce@bmce-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$
```

```
mysql-precision-t1700:~$ typescript lab2.txt
typescript: command not found
mysql-precision-t1700:~$ script lab2.txt
Script started, file is lab2.txt
mysql-precision-t1700:~/cassandra$ cd apache-cassandra-3.11.0/
mysql-precision-t1700:~/cassandra/apache-cassandra-3.11.0$ ./qlsh
bash: ./qlsh: No such file or directory
mysql-precision-t1700:~/cassandra/apache-cassandra-3.11.0$ ls
bin      conf  interface  lib  NEWS.txt  pylib
CHANGES.txt  doc  javadoc  LICENSE.txt  NOTICE.txt  tools
mysql-precision-t1700:~/cassandra/apache-cassandra-3.11.0$ cd bin
mysql-precision-t1700:~/cassandra/apache-cassandra-3.11.0/bin$ ./qlsh
Connected to Test Cluster at 127.0.0.1:9042.
[qlsh 5.0.1 | Cassandra 3.11.5 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
qlsh> SELECT * FROM system_schema.keyspaces
... /

keyspace_name | durable_writes | replication
-----
system_auth | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_distributed | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system_traces | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}
students | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}

(7 rows)
qlsh> CREATE KEYSPACE employee WITH REPLICATION={('class': 'SimpleStrategy', 'replication_factor': 1)};
SyntaxException: line 1:48 no viable alternative at input '{' (... KEYSPACE employee WITH REPLICATION={('class':...})
qlsh> CREATE KEYSPACE employee WITH REPLICATION={('class': 'SimpleStrategy', 'replication_factor': 1)};
qlsh> USE employee;
qlsh:employee> CREATE TABLE employee_info (emp_id int PRIMARY KEY, emp_name text, designation text, DateofJoining timestamp, salary double, Dept_name text );
SyntaxException: line 1:21 no viable alternative at input '<' (CREATE TABLE [employee]...)
qlsh:employee> CREATE TABLE employee_info (emp_id int PRIMARY KEY, emp_name text, designation text, DateofJoining timestamp, salary double, Dept_name text );
qlsh:employee> DESCRIBE TABLES;

employee_info

qlsh:employee> DESCRIBE TABLE employee_info;

CREATE TABLE employee.employee_info (
  emp_id int PRIMARY KEY,
  dateofjoining timestamp,
  dept_name text,
  designation text,
  emp_name text,
  salary double
) WITH bloom filter fp_chance = 0.01
AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
AND comment = ''
AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
AND crc_check_chance = 1.0
AND dclocal_read_repair_chance = 0.1
AND default_time_to_live = 0
AND gc_grace_seconds = 864000
AND max_index_interval = 2048
AND memtable_flush_period_in_ms = 0
AND min_index_interval = 128
AND read_repair_chance = 0.0
AND speculative_retry = '99PERCENTILE';

qlsh:employee> BEGIN BATCH
... /
... /
... /
SyntaxException: line 2:0 mismatched input ';' expecting K APPLY (BEGIN BATCH){...}
qlsh:employee> BEGIN BATCH INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (100, 'Eva Smart', 'HOD', '2020-02-28', 50000.00, 'CSE' )
... INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (120, 'Kelsey Stephens', 'Professor', '2020-03-08', 49000.00, 'CSE' )
... INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (120, 'Vince Poole', 'Professor', '2020-03-08', 49000.00, 'CSE' )
... INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (120, 'Phyllis Douglas', 'Assistant Professor', '2020-03-08', 47000.00, 'CSE' )
... INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (121, 'Laurel Dittman', 'Associate Professor', '2020-03-08', 48000.00, 'CSE' )
... INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (121, 'Horace Rice', 'Associate Professor', '2020-03-08', 48000.00, 'CSE' )
... INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (123, 'Phyllis Douglas', 'Assistant Professor', '2020-03-10', 47000.00, 'CSE' )
... APPLY BATCH ;
qlsh:employee> SELECT * from employee_info
... /

emp_id | dateofjoining | dept_name | designation | emp_name | salary
-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | 49000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | 47000
121 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Horace Rice | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | 50000

(4 rows)
qlsh:employee> BEGIN BATCH INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (109, 'Kelsey Stephens', 'Professor', '2020-03-08', 49000.00, 'CSE' ) INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (110, 'Kelsey Stephens', 'Professor', '2020-03-08', 49000.00, 'CSE' )
qlsh:employee> SELECT * from employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | salary
-----
110 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | 49000
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | 49000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | 47000
122 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Laurel Dittman | 48000
121 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Horace Rice | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | 49000

(7 rows)
qlsh:employee> DELETE from employee_info WHERE emp_id=110;
qlsh:employee> SELECT * from employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | salary
-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | 49000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | 47000
122 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Laurel Dittman | 48000
121 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Horace Rice | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | 49000

(6 rows)
qlsh:employee> UPDATE employee_info SET emp_name='Edwin Mendoza',Dept_name='ISE' WHERE emp_id=121;
qlsh:employee> SELECT * from employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | salary
-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | 49000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | 47000
122 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Laurel Dittman | 48000
121 | 2020-03-07 18:30:00.000000+0000 | ISE | Associate Professor | Edwin Mendoza | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | 49000

(6 rows)
qlsh:employee> SELECT * FROM employee_info ORDER BY salary ASC;
InvalidRequest: Error from server: code=2200 [Invalid query] message="ORDER BY is only supported when the partition key is restricted by an EQ or an IN."
qlsh:employee> SELECT * FROM employee_info WHERE emp_id=109 ORDER BY salary ASC;
SyntaxException: line 1:41 no viable alternative at input ';' (... FROM employee_info WHERE emp_id=109)
qlsh:employee> SELECT * FROM employee_info WHERE salary>0 ORDER BY salary ASC;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this query despite the performance unpredictability, use ALLOW FILTERING"
qlsh:employee> SELECT * FROM employee_info WHERE salary > 0, emp_id IN 8 ORDER BY salary ASC;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this query despite the performance unpredictability, use ALLOW FILTERING"
qlsh:employee> SELECT * FROM employee_info WHERE salary > 0, emp_id IN 8 ORDER BY salary ASC;
SyntaxException: line 1:44 mismatched input ',', expecting EOF (... employee_info WHERE salary > 0],[...])
qlsh:employee> SELECT * FROM employee_info WHERE salary > 0, emp_id IN 8 ORDER BY salary ASC;
SyntaxException: line 1:44 mismatched input ',', expecting EOF (... employee_info WHERE salary > 0],[...])
qlsh:employee> SELECT * FROM employee_info WHERE emp_id IN 100, 1000 ORDER BY salary ASC;
SyntaxException: line 1:44 no viable alternative at input ';' (... (SELECT * FROM employee_info WHERE [emp_id] IN...))
qlsh:employee> SELECT * FROM employee_info WHERE emp_id IN (100, 1000) ORDER BY salary ASC;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Order by is currently only supported on the clustered columns of the PRIMARY KEY, got salary"
qlsh:employee> SELECT * FROM employee_info WHERE salary>0 ORDER BY salary ASC ALLOW FILTERING;
InvalidRequest: Error from server: code=2200 [Invalid query] message="ORDER BY is only supported when the partition key is restricted by an EQ or an IN."
qlsh:employee> ALTER TABLE employee_info ADD projects set<text>
... /
qlsh:employee> SELECT * from employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | projects | salary
-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | {'Project1', 'Project2'} | 48000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | {'Project2'} | 47000
122 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Laurel Dittman | null | 48000
121 | 2020-03-07 18:30:00.000000+0000 | ISE | Associate Professor | Edwin Mendoza | null | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | null | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | null | 49000

(6 rows)
qlsh:employee> CREATE INDEX ON employee_info (salary);
qlsh:employee> SELECT * FROM employee_info ORDER BY salary
... /
InvalidRequest: Error from server: code=2200 [Invalid query] message="ORDER BY is only supported when the partition key is restricted by an EQ or an IN."
qlsh:employee> DESC INDEX employee.employee_info_salary_idx
CREATE INDEX employee_info_salary_idx ON employee.employee_info (salary);
qlsh:employee> DESC INDEX employee.employee_info_salary_idx;
CREATE INDEX employee_info_salary_idx ON employee.employee_info (salary);
qlsh:employee> DROP INDEX IF EXISTS employee.employee_info_salary_idx
... /
qlsh:employee> UPDATE employee_info SET projects=projects+(Project1) WHERE emp_id=100;
SyntaxException: line 1:52 no viable alternative at input ')' (UPDATE employee_info SET projects=projects+{Project1}...)
qlsh:employee> UPDATE employee_info SET projects=projects+(Project1) WHERE emp_id=100;
SyntaxException: line 1:52 no viable alternative at input ')' (UPDATE employee_info SET projects=projects+{Project1}...)
qlsh:employee> UPDATE employee_info SET projects=projects+(Project1) WHERE emp_id=100;
qlsh:employee> UPDATE employee_info SET projects=projects+(Project1) WHERE emp_id=109;
qlsh:employee> UPDATE employee_info SET projects=projects+(Project1,Project2) WHERE emp_id=120;
qlsh:employee> UPDATE employee_info SET projects=projects+(Project2) WHERE emp_id=123;
qlsh:employee> UPDATE employee_info SET projects=projects+(Project2,Project4) WHERE emp_id=122;
qlsh:employee> UPDATE employee_info SET projects=projects+(Project4) WHERE emp_id=121;
qlsh:employee> INSERT INTO employee_info (emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (122, 'Conrad Menzie', 'Intern', '2020-03-28', 10000.00, 'CSE') USING TTL 30 ;
qlsh:employee> SELECT * FROM employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | projects | salary
-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | {'Project1', 'Project2'} | 48000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | {'Project2'} | 47000
122 | 2020-03-27 18:30:00.000000+0000 | CSE | Intern | Conrad Menzie | {'Project2', 'Project4'} | 10000

(3 rows)
```

```
ISE | Associate Professor | Edwin Mendoza | ('Project4') | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | ('Project1') | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | ('Project1') | 49000

(6 rows)
cqlsh:employee> INSERT INTO employee_info(emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (101, 'Conrad Menzie', 'Intern', '2020-03-28',10000.00,'CSE') USING TTL 30 ;
cqlsh:employee> SELECT * FROM employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | projects | salary
-----|-----|-----|-----|-----|-----|-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | ('Project1', 'Project2') | 48000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | ('Project2') | 47000
122 | 2020-03-07 18:30:00.000000+0000 | null | null | null | ('Project2', 'Project4') | null
121 | 2020-03-07 18:30:00.000000+0000 | ISE | Associate Professor | Edwin Mendoza | ('Project4') | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | ('Project1') | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | ('Project1') | 49000
101 | 2020-03-27 18:30:00.000000+0000 | CSE | Intern | Conrad Menzie | null | 10000

(7 rows)
cqlsh:employee> SELECT * FROM employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | projects | salary
-----|-----|-----|-----|-----|-----|-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | ('Project1', 'Project2') | 48000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | ('Project2') | 47000
122 | 2020-03-07 18:30:00.000000+0000 | null | null | null | ('Project2', 'Project4') | null
121 | 2020-03-07 18:30:00.000000+0000 | ISE | Associate Professor | Edwin Mendoza | ('Project4') | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | ('Project1') | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | ('Project1') | 49000

(6 rows)
cqlsh:employee> INSERT INTO employee_info(emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (122, 'Laurel Dittman', 'Associate Professor', '2020-03-08',48000.00,'CSE') ;
cqlsh:employee> SELECT * FROM employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | projects | salary
-----|-----|-----|-----|-----|-----|-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | ('Project1', 'Project2') | 48000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | ('Project2') | 47000
122 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Laurel Dittman | ('Project2', 'Project4') | 48000
121 | 2020-03-07 18:30:00.000000+0000 | ISE | Associate Professor | Edwin Mendoza | ('Project4') | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | ('Project1') | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | ('Project1') | 49000

(6 rows)
cqlsh:employee> INSERT INTO employee_info(emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (101, 'Conrad Menzie', 'Intern', '2020-03-28',10000.00,'CSE') USING TTL 30 ;
cqlsh:employee> INSERT INTO employee_info(emp_id, emp_name, designation, DateofJoining, salary, Dept_name) VALUES (101, 'Conrad Menzie', 'Intern', '2020-03-28',10000.00,'CSE') USING TTL 30 ;
cqlsh:employee> SELECT * FROM employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | projects | salary
-----|-----|-----|-----|-----|-----|-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | ('Project1', 'Project2') | 48000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | ('Project2') | 47000
122 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Laurel Dittman | ('Project2', 'Project4') | 48000
121 | 2020-03-07 18:30:00.000000+0000 | ISE | Associate Professor | Edwin Mendoza | ('Project4') | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | ('Project1') | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | ('Project1') | 49000
101 | 2020-03-27 18:30:00.000000+0000 | CSE | Intern | Conrad Menzie | null | 10000

(7 rows)
cqlsh:employee> SELECT * FROM employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | projects | salary
-----|-----|-----|-----|-----|-----|-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | ('Project1', 'Project2') | 48000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | ('Project2') | 47000
122 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Laurel Dittman | ('Project2', 'Project4') | 48000
121 | 2020-03-07 18:30:00.000000+0000 | ISE | Associate Professor | Edwin Mendoza | ('Project4') | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | ('Project1') | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | ('Project1') | 49000
101 | 2020-03-27 18:30:00.000000+0000 | CSE | Intern | Conrad Menzie | null | 10000

(7 rows)
cqlsh:employee> SELECT * FROM employee_info ;

emp_id | dateofjoining | dept_name | designation | emp_name | projects | salary
-----|-----|-----|-----|-----|-----|-----
120 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Vince Poole | ('Project1', 'Project2') | 48000
123 | 2020-03-09 18:30:00.000000+0000 | CSE | Assistant Professor | Phyllis Douglas | ('Project2') | 47000
122 | 2020-03-07 18:30:00.000000+0000 | CSE | Associate Professor | Laurel Dittman | ('Project2', 'Project4') | 48000
121 | 2020-03-07 18:30:00.000000+0000 | ISE | Associate Professor | Edwin Mendoza | ('Project4') | 48000
100 | 2020-02-27 18:30:00.000000+0000 | CSE | HOD | Eva Smart | ('Project1') | 50000
109 | 2020-03-07 18:30:00.000000+0000 | CSE | Professor | Kelsey Stephens | ('Project1') | 49000

(6 rows)
cqlsh:employee> exit
bmuce@bmuce-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ ./cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.5 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> CREATE KEYSPACE library WITH REPLICATION={'class':'SimpleStrategy','replication_factor':1};
SyntaxException: line 1:41 no viable alternative at input '{' (CREATE KEYSPACE library WITH [REPLICATION]=...)
cqlsh> CREATE KEYSPACE library WITH REPLICATION={'class':'SimpleStrategy','replication_factor':1};
cqlsh> USE library;
cqlsh:library> CREATE TABLE library_info(
...
SyntaxException: line 2:0 no viable alternative at input ';' (CREATE TABLE library_info{;})
cqlsh:library> CREATE TABLE library_info(stud_id int PRIMARY KEY, stud_name text, book_name text, bookname text, book_id int, DateofIssue timestamp);
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot mix counter and non counter columns in the same table"
cqlsh:library>
cqlsh:library> ALTER TABLE library_info ADD counter_value counter;
ConfigurationException: Cannot add a counter column (counter_value) in a non counter column family
cqlsh:library> DROP TABLE library_info;
cqlsh:library> DESCRIBE TABLES;
<empty>

cqlsh:library> CREATE TABLE library_info(stud_id int, stud_name text, book_name text, bookname text, book_id int, DateofIssue timestamp, counter_value counter , PRIMARY KEY (stud_id));
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot mix counter and non counter columns in the same table"
cqlsh:library> CREATE TABLE library_info(stud_id int PRIMARY KEY, stud_name text, bookname text, book_id int, DateofIssue timestamp, counter_value counter);
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot mix counter and non counter columns in the same table"
cqlsh:library>
cqlsh:library> CREATE TABLE library_info(stud_id int PRIMARY KEY, counter_value counter);
cqlsh:library> ALTER TABLE library_info ADD stud_name text, book_name text, bookname text, book_id int, DateofIssue timestamp ;
SyntaxException: line 1:25 no viable alternative at input 'stud_name' (ALTER TABLE [library_info] stud_name...)
cqlsh:library> ALTER TABLE library_info ADD stud_name text, book_name text, bookname text, book_id int, DateofIssue timestamp ;
SyntaxException: line 1:43 mismatched input ',', expecting EOF (... TABLE library_info ADD stud_name text[, book_name...)
cqlsh:library> ALTER TABLE library_info stud_name text
...
SyntaxException: line 1:25 no viable alternative at input 'stud_name' (ALTER TABLE [library_info] stud_name...)
cqlsh:library> DROP TABLE library_info ;
cqlsh:library> CREATE TABLE library_info(stud_id int, stud_name text, book_name text, bookname text, book_id int, DateofIssue timestamp, counter_value counter , PRIMARY KEY (stud_id,stud_name,timestamp), book_idbook_name,);
SyntaxException: line 1:206 no viable alternative at input ',' (...(stud_id,stud_name,timestamp), book_idbook_name[,...])
cqlsh:library> CREATE TABLE library_info(stud_id int, stud_name text, book_name text, bookname text, book_id int, DateofIssue timestamp, counter_value counter , PRIMARY KEY (stud_id,stud_name), book_id,book_name,DateofIssue);
SyntaxException: line 1:186 no viable alternative at input ',' (... KEY (stud_id,stud_name), book_id[,...])
cqlsh:library> CREATE TABLE library_info(stud_id int, stud_name text, book_name text, bookname text, book_id int, DateofIssue timestamp, counter_value counter , PRIMARY KEY (stud_id,stud_name,book_id,book_name,DateofIssue);
SyntaxException: line 1:185 no viable alternative at input ',' (... KEY (stud_id,stud_name),book_id[,...])
cqlsh:library> CREATE TABLE library_info(stud_id int, stud_name text, book_name text, book_id int, DateofIssue timestamp, counter_value counter , PRIMARY KEY (stud_id,stud_name),book_id,book_name,DateofIssue);
SyntaxException: line 1:170 no viable alternative at input ',' (... KEY (stud_id,stud_name),book_id[,...])
cqlsh:library> CREATE TABLE library_info(stud_id int, stud_name text, book_name text, book_id int, DateofIssue timestamp, counter_value counter , PRIMARY KEY ((stud_id,stud_name),book_id,book_name,DateofIssue);
SyntaxException: line 1:194 mismatched input ';' expecting ')' (...stud_name),book_id,book_name,DateofIssue);)
cqlsh:library> CREATE TABLE library_info(stud_id int, stud_name text, book_name text, book_id int, DateofIssue timestamp, counter_value counter , PRIMARY KEY ((stud_id,stud_name),book_id,book_name,DateofIssue));
cqlsh:library> exit
bmuce@bmuce-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ stop
Command 'stop' not found, but there are 19 similar ones.

bmuce@bmuce-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ script
Script started, file is typescript
bmuce@bmuce-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ script stop
Script started, file is stop
bmuce@bmuce-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ exit
exit
Script done, file is stop
bmuce@bmuce-Precision-T1700:~/cassandra/apache-cassandra-3.11.0/bin$ exit
exit
Script done, file is lab2.txt
bmuce@bmuce-Precision-T1700:~$
```

```
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.10 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> CREATE KEYSPACE Library WITH REPLICATION={
... 'class':'SimpleStrategy',
... 'replication_factor':1};
cqlsh> DESCRIBE KEYSPACES

system_schema system_auth system_library system_distributed system_traces

cqlsh> SELECT * FROM sytem.schema_keyspaces ;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Keyspace sytem does not exist"
cqlsh> SELECT * FROM system.schema_keyspaces ;
InvalidRequest: Error from server: code=2200 [Invalid query] message="unconfigured table schema_keyspaces"
cqlsh> SELECT * FROM system_schema.keyspaces ;

keyspace_name | durable_writes | replication
-----
system_auth | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_library | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_distributed | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system_system | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_traces | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}

(6 rows)
cqlsh> USE Library;
cqlsh:library> create table library_info(
... studid int PRIMARY KEY,
... countval counter,
;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot mix counter and non counter columns in the same table"
cqlsh:library> create table library_info(
... studid int PRIMARY KEY,
... countval counter);
cqlsh:library> DEscribe table library_info

CREATE TABLE library.library_info (
studid int PRIMARY KEY,
countval counter
) WITH bloom_filter_fp_chance = 0.01
AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
AND comment = ''
AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
AND crc_check_chance = 1.0
AND dclocal_read_repair_chance = 0.1
AND default_time_to_live = 0
AND gc_grace_seconds = 864000
AND max_index_interval = 2048
AND memtable_flush_period_in_ms = 0
AND min_index_interval = 128
AND read_repair_chance = 0.0
AND speculative_retry = '99PERCENTILE';

cqlsh:library> SELECT * FROM Library_info
... ;

studid | countval
-----+-----
(0 rows)
1';
SyntaxException: Only expressions of the form X = X +<value> are supported.
;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Invalid STRING constant (001) for "studid" of type int"
;
cqlsh:library> SELECT * FROM Library_info;

studid | countval
-----+-----
111 | 1

(1 rows)
;
cqlsh:library> SELECT * FROM Library_info;

studid | countval
-----+-----
111 | 2

(1 rows)
cqlsh:library> ALTER library_info
... DROP countval
... ;
SyntaxException: line 1:6 no viable alternative at input 'library_info' ([ALTER] library_info...)
cqlsh:library> ALTER table library_info
... DROP countval;
cqlsh:library> SELECT * FROM Library_info;

studid
-----
(0 rows)
cqlsh:library> ALTER TABLE library_info ADD studname text;
ConfigurationException: Cannot add a non counter column (studname) in a counter column family
cqlsh:library> DROP tablelibrary_info;
SyntaxException: line 1:5 no viable alternative at input 'tablelibrary_info' ([DROP] tablelibrary_info...)
cqlsh:library> DROP table library_info;
cqlsh:library> DESCRIBE COLUMNFAMILIES;

<empty>

cqlsh:library> create table library_info(
... studid int PRIMARY KEY,
... studname text,
... bookname text,
... bookid int,
... dateofissue timestamp) ;
cqlsh:library> SELECT * FROM Library_info;

studid | bookid | bookname | dateofissue | studname
-----+-----+-----+-----+-----
(0 rows)
cqlsh:library> BEGIN BATCH
... INSERT INTO
... library_info(
... studid,studname,bookname ,bookid,dateofissue)
... VALUES
... (1,'Michael Storm','BDA Beginner',10000,'2012-03-29')
issue) VALUES
... (2,'Stephen Fox','Cassandra 3.1.11',10001,'2012-03-30')
... APPLY BATCH;
cqlsh:library> SELECT * FROM Library_info;

studid | bookid | bookname | dateofissue | studname
```

```

1 | 10000 | BDA Beginner | 2012-03-29 00:00:00.000000+0000 | Michael Storm
2 | 10001 | Cassandra 3.1.11 | 2012-03-30 00:00:00.000000+0000 | Stephen Fox

(2 rows)
cqlsh:library> ALTER TABLE library_info add counter int
... ;
cqlsh:library> ;
SyntaxException: line 1:0 no viable alternative at input ';' ([;])
cqlsh:library> SELECT * FROM Library_info;

studid | bookid | bookname | counter | dateofissue | studname
-----+-----+-----+-----+-----+-----
1 | 10000 | BDA Beginner | null | 2012-03-29 00:00:00.000000+0000 | Michael Storm
2 | 10001 | Cassandra 3.1.11 | null | 2012-03-30 00:00:00.000000+0000 | Stephen Fox

(2 rows)
cqlsh:library> ;
SyntaxException: line 1:0 no viable alternative at input ';' ([;])
cqlsh:library> UPDATE library_info SET
... counter=2 WHERE studid=1
... ;
cqlsh:library> UPDATE library_info SET
... counter=2 WHERE studid=1
... ;
cqlsh:library> UPDATE library_info SET
... counter=1 WHERE studid=2;
cqlsh:library> SELECT * FROM Library_info;

studid | bookid | bookname | counter | dateofissue | studname
-----+-----+-----+-----+-----+-----
1 | 10000 | BDA Beginner | 2 | 2012-03-29 00:00:00.000000+0000 | Michael Storm
2 | 10001 | Cassandra 3.1.11 | 1 | 2012-03-30 00:00:00.000000+0000 | Stephen Fox

(2 rows)
cqlsh:library> SELECT bookname, counter FROM Library_info where studid=1;

bookname | counter
-----+-----
BDA Beginner | 2

(1 rows)
cqlsh:library> SELECT studid,bookname, counter FROM Library_info where studid=1
... ;

studid | bookname | counter
-----+-----+-----
1 | BDA Beginner | 2

(1 rows)
cqlsh:library> COPY library_info(studid,studname,bookname ,bookid,dateofissue)
... TO '/content/libinfo.csv' WITH HEADER = TRUE ;
Using 1 child processes

Starting copy of library.library_info with columns [studid, studname, bookname, bookid, dateofissue].
Processed: 2 rows; Rate: 3 rows/s; Avg. rate: 3 rows/s
2 rows exported to 1 files in 0.689 seconds.
cqlsh:library> TRUNCATE library_info ;
cqlsh:library> SELECT * FROM library_info;

studid | bookid | bookname | counter | dateofissue | studname
-----+-----+-----+-----+-----+-----

(0 rows)
cqlsh:library> COPY library_info(studid,studname,bookname ,bookid,dateofissue)
... FROM '/content/libinfo.csv' WITH HEADER = TRUE ;
Invalid syntax at line 2, char 5
FROM '/content/libinfo.csv' WITH HEADER = TRUE ;
^
cqlsh:library> FROM '/content/libinfo.csv' WITH HEADER = TRUE ;
SyntaxException: line 1:0 no viable alternative at input 'FROM' ([FROM]...)
cqlsh:library> COPY library_info(studid,studname,bookname ,bookid,dateofissue)
... FROM '/content/libinfo.csv' WITH HEADER = TRUE ;
Using 1 child processes

Starting copy of library.library_info with columns [studid, studname, bookname, bookid, dateofissue].
Processed: 2 rows; Rate: 4 rows/s; Avg. rate: 6 rows/s
2 rows imported from 1 files in 0.356 seconds (0 skipped).
cqlsh:library> SELECT * FROM library_info;

studid | bookid | bookname | counter | dateofissue | studname
-----+-----+-----+-----+-----+-----
1 | 10000 | BDA Beginner | null | 2012-03-29 00:00:00.000000+0000 | Michael Storm
2 | 10001 | Cassandra 3.1.11 | null | 2012-03-30 00:00:00.000000+0000 | Stephen Fox

(2 rows)
cqlsh:library> exit

```

```
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.10 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
ation_factor':1});
cqlsh> describe keyspaces;

system_schema system_auth system system_distributed wmetro system_traces

cqlsh> select * from system_schema.keyspaces;

keyspace_name | durable_writes | replication
-----
system_auth | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_distributed | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system | True | {'class': 'org.apache.cassandra.locator.LocalStrategy'}
wmetro | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_traces | True | {'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}

(6 rows)
cqlsh> use wmetro
... ;
ile bigint,current_position text,skill set<text>);

cqlsh:wmetro> select * from drivers;
InvalidRequest: Error from server: code=2200 [Invalid query] message="unconfigured table drivers"
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile | password | skill
-----
(0 rows)
cqlsh:wmetro> begin batch
ill) values('Swen','123abc',9000000001,'Wellington',{'Passenger'})
ill) values('Fred','234bcd',9000000002,'Upper Hutt',{'Passenger'})
ill) values('Jane','345cde',9000000003,'Lower Hutt',{'Goods'})
... apply batch;
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile | password | skill
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'}
Swen | Wellington | 9000000001 | 123abc | {'Passenger'}
Jane | Lower Hutt | 9000000003 | 345cde | {'Goods'}

(3 rows)
text);
cqlsh:wmetro> begin batch
tenance','Gulliver');
Hutt','Gulliver');
r Hutt','Kiwi Rail');
ington','Kiwi Rail');
... apply batch;
cqlsh:wmetro> select * from vehicles;

vehicle_id | status | type
-----
WR1236 | Upper Hutt | Kiwi Rail
WR1234 | Maintenance | Gulliver
WR1235 | Lowe Hutt | Gulliver
WR1237 | Wellington | Kiwi Rail

(4 rows)
ine_name, station_name));
cqlsh:wmetro> begin batch
,longitude,time,distance) values('LINE1',2,'Wellington',-40.28,174.77,0605,0.0)

,longitude,time,distance) values('LINE2',2,'Upper Hutt',-41.12,175.07,1550,0.0)
... apply batch;
cqlsh:wmetro> select * from time_table;

line_name | station_name | distance | latitude | longitude | service_num | time
-----
LINE2 | Upper Hutt | 0 | -41.12 | 175.07 | 2 | 1550
LINE2 | Wellington | 34.3 | -40.28 | 174.77 | 2 | 1130
LINE1 | Upper Hutt | 34.3 | -41.12 | 175.07 | 2 | 1025
LINE1 | Wellington | 0 | -40.28 | 174.77 | 2 | 605

(4 rows)
nce counter);
ere vehicle_id=1234 ;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Invalid INTEGER constant (1234) for "vehicle_id" of type text"
ere vehicle_id="WR1234" ;
SyntaxException: line 1:90 no viable alternative at input ';' (...total_distance + 3000 where vehicle_id=["WR123]4" );
ere vehicle_id='WR1234';
ere vehicle_id='WR1235';
ere vehicle_id='WR1236';
ere vehicle_id='WR1237';
cqlsh:wmetro> select * from VEHICLE_USAGE;

vehicle_id | total_distance
-----
WR1236 | 3500
WR1234 | 3000
WR1235 | 5000
WR1237 | 5300

(4 rows)
itude double,longitude double,speed double);
cqlsh:wmetro> begin batch
es(14,'2020-03-10',-42.30,171.50,76)
es(14,'2020-03-10 10:07:40+1300',-42.50,171.90,75)
es(14,'2020-03-10 10:30:25+1300',-44.50,160.60,75)
... apply batch ;
cqlsh:wmetro> select * from data_point;

sequence | day | latitude | longitude | speed
-----
2020-03-09 21:30:25.000000+0000 | 14 | -44.5 | 160.6 | 75
2020-03-09 21:07:40.000000+0000 | 14 | -42.5 | 171.9 | 75
2020-03-10 00:00:00.000000+0000 | 14 | -42.3 | 171.5 | 76

(3 rows)
ill) values('William','678bcf',9000000004,'Wellington',{'Goods','Passenger'})
... ;
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile | password | skill
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'}
Swen | Wellington | 9000000001 | 123abc | {'Passenger'}
William | Wellington | 9000000004 | 678bcf | {'Goods', 'Passenger'}
Jane | Lower Hutt | 9000000003 | 345cde | {'Goods'}

(4 rows)
```

```

cqlsh:wmetro> create index on driver(password);
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile      | password | skill
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'}
Sven | Wellington | 9000000001 | 123abc | {'Passenger'}
William | Wellington | 9000000004 | 678bcf | {'Goods', 'Passenger'}
Jane | Lower Hutt | 9000000003 | 345cde | {'Goods'}

(4 rows)
cqlsh:wmetro> update driver set password='789gfh' where driver_name='William';
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile      | password | skill
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'}
Sven | Wellington | 9000000001 | 123abc | {'Passenger'}
William | Wellington | 9000000004 | 789gfh | {'Goods', 'Passenger'}
Jane | Lower Hutt | 9000000003 | 345cde | {'Goods'}

(4 rows)
='William';
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile      | password | skill
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'}
Sven | Wellington | 9000000001 | 123abc | {'Passenger'}
William | Maintenance | 9000000004 | 789gfh | {'Goods', 'Passenger'}
Jane | Lower Hutt | 9000000003 | 345cde | {'Goods'}

(4 rows)
cqlsh:wmetro> update driver set skill={'Passenger'} where driver_name='Jane';
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile      | password | skill
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'}
Sven | Wellington | 9000000001 | 123abc | {'Passenger'}
William | Maintenance | 9000000004 | 789gfh | {'Goods', 'Passenger'}
Jane | Lower Hutt | 9000000003 | 345cde | {'Passenger'}

(4 rows)
Jane';
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile      | password | skill
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'}
Sven | Wellington | 9000000001 | 123abc | {'Passenger'}
William | Maintenance | 9000000004 | 789gfh | {'Goods', 'Passenger'}
Jane | Lower Hutt | 9000000003 | 345cde | {'Goods', 'Passenger'}

(4 rows)
password='789gfh';

[applied]
-----
True

cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile      | password | skill
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'}
Sven | Wellington | 9000000001 | 123abc | {'Passenger'}
William | Maintenance | 9000000004 | 678bcf | {'Goods', 'Passenger'}
Jane | Lower Hutt | 9000000003 | 345cde | {'Goods', 'Passenger'}

(4 rows)
cqlsh:wmetro> alter table driver add workdays int ;
cqlsh:wmetro> update driver set workdays=22 where driver_name='William'
... update driver set workdays=22 where driver_name='Fred';
SyntaxException: line 2:0 mismatched input 'update' expecting EOF (...workdays=22 where driver_name='William'[update]...)
cqlsh:wmetro> update driver set workdays=22 where driver_name='Fred';
cqlsh:wmetro> update driver set workdays=22 where driver_name='William';
cqlsh:wmetro> update driver set workdays=21 where driver_name='Jane';
cqlsh:wmetro> update driver set workdays=22 where driver_name='Sven';
cqlsh:wmetro> select * from driver;

driver_name | current_position | mobile      | password | skill | workdays
-----
Fred | Upper Hutt | 9000000002 | 234bcd | {'Passenger'} | 22
Sven | Wellington | 9000000001 | 123abc | {'Passenger'} | 22
William | Maintenance | 9000000004 | 678bcf | {'Goods', 'Passenger'} | 22
Jane | Lower Hutt | 9000000003 | 345cde | {'Goods', 'Passenger'} | 21

(4 rows)
ington','Diesel');
cqlsh:wmetro> select * from vehicles;

vehicle_id | status | type
-----
WR1236 | Upper Hutt | Kiwi Rail
WR1234 | Maintenance | Gulliver
WR1238 | Wellington | Diesel
WR1235 | Lowe Hutt | Gulliver
WR1237 | Wellington | Kiwi Rail

(5 rows)
;
cqlsh:wmetro> select * from vehicles;

vehicle_id | status | type
-----
WR1236 | Upper Hutt | Kiwi Rail
WR1234 | Upper Hutt | Gulliver
WR1238 | Wellington | Diesel
WR1235 | Lowe Hutt | Gulliver
WR1237 | Wellington | Kiwi Rail

(5 rows)
cqlsh:wmetro> alter table vehicle add daily_distance ;
SyntaxException: line 1:39 no viable alternative at input ';' (alter table vehicle add daily_distance [;])
cqlsh:wmetro> alter table vehicles add daily_distance;
SyntaxException: line 1:39 no viable alternative at input ';' (alter table vehicles add daily_distance[;])
cqlsh:wmetro> alter table vehicles add daily_distance int;
cqlsh:wmetro> update vehicles set daily_distance=180 where vehicle_id="WR1236";
SyntaxException: line 1:64 no viable alternative at input ';' (...set daily_distance=180 where vehicle_id=["WR1236"]);
cqlsh:wmetro> select* from vehicles;
InvalidRequest: Error from server: code=2200 [Invalid query] message="unconfigured table veicles"
cqlsh:wmetro> select * from vehicles;

vehicle_id | daily_distance | status | type
-----
WR1236 | null | Upper Hutt | Kiwi Rail
WR1234 | null | Upper Hutt | Gulliver

```



```

WR1238 | null | Wellington | Diesel
WR1235 | null | Lowe Hutt | Gulliver
WR1237 | null | Wellington | Kiwi Rail

(5 rows)
;
;
;
;
;
cqish:wmetro> select * from vehicles;

vehicle_id | daily_distance | status | type
-----
WR1236 | 180 | Upper Hutt | Kiwi Rail
WR1234 | 200 | Upper Hutt | Gulliver
WR1238 | 300 | Wellington | Diesel
WR1235 | 50 | Lowe Hutt | Gulliver
WR1237 | 300 | Wellington | Kiwi Rail

(5 rows)
HICLE_ID='WR1235';
cqish:wmetro> SELECT * FROM VEHICLE_USAGE;

vehicle_id | total_distance
-----
WR1236 | 3500
WR1234 | 3000
WR1235 | 5050
WR1237 | 5300

(4 rows)

cqish:wmetro> begin batch
,longitude,time,distance) values('LINE3',2,'Melling',-43.20,171.60,1540,330)
,longitude,time,distance) values('LINE3',2,'Waterloo',-53.50,165.70,0940,40,0)
,longitude,time,distance) values('LINE4',2,'Melling',-43.20,171.60,1540,0)
)
... apply batch ;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Unmatched column names/values"
,longitude,time,distance) values('LINE3',2,'Melling',-43.20,171.60,1540,330.0)
... ;
cqish:wmetro> begin batch
)
,longitude,time,distance) values('LINE4',2,'Melling',-43.20,171.60,1540,0.0)
... apply batch ;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Unmatched column names/values"
;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Unmatched column names/values"
,longitude,time,distance) values('LINE3',2,'Waterloo',-53.50,165.70,0940,0.0);
,longitude,time,distance) values('LINE4',2,'Melling',-43.20,171.60,1540,0.0)
... ;
0);
InvalidRequest: Error from server: code=2200 [Invalid query] message="Unmatched column names/values"
,longitude,time,distance) values('LINE4',2,'Waterloo',-53.50,165.70,0940,330.0)
... ;
cqish:wmetro> SELECT * FROM TIME_TABLE;

line_name | station_name | distance | latitude | longitude | service_num | time
-----
LINE2 | Upper Hutt | 0 | -41.12 | 175.07 | 2 | 1550
LINE2 | Wellington | 34.3 | -40.28 | 174.77 | 2 | 1130
LINE3 | Melling | 330 | -43.2 | 171.6 | 2 | 1540
LINE3 | Waterloo | 0 | -53.5 | 165.7 | 2 | 940
LINE4 | Melling | 0 | -43.2 | 171.6 | 2 | 1540
LINE4 | Waterloo | 330 | -53.5 | 165.7 | 2 | 940
LINE1 | Upper Hutt | 34.3 | -41.12 | 175.07 | 2 | 1025
LINE1 | Wellington | 0 | -40.28 | 174.77 | 2 | 605

(8 rows)
' and station_name='Waterloo';
SyntaxException: line 1:7 no viable alternative at input 'table' (update [table]...)
station_name='Waterloo';
station_name='Melling';
cqish:wmetro> begin batch
station_name='Waterloo'
station_name='Melling'
... apply batch ;
es(15,'2020-03-04',-52.10,168.90,67);
cqish:wmetro> select * from time_table;

line_name | station_name | distance | latitude | longitude | service_num | time
-----
LINE2 | Upper Hutt | 0 | -41.12 | 175.07 | 2 | 1550
LINE2 | Wellington | 34.3 | -40.28 | 174.77 | 2 | 1130
LINE3 | Melling | 330 | -43.2 | 171.6 | 1 | 1540
LINE3 | Waterloo | 0 | -53.5 | 165.7 | 1 | 940
LINE4 | Melling | 0 | -43.2 | 171.6 | 1 | 1540
LINE4 | Waterloo | 330 | -53.5 | 165.7 | 1 | 940
LINE1 | Upper Hutt | 34.3 | -41.12 | 175.07 | 2 | 1025
LINE1 | Wellington | 0 | -40.28 | 174.77 | 2 | 605

(8 rows)
cqish:wmetro> SELECT * FROM DATA_POINT;

sequence | day | latitude | longitude | speed
-----
2020-03-04 00:00:00.000000+0000 | 15 | -52.1 | 168.9 | 67
2020-03-09 21:30:25.000000+0000 | 14 | -44.5 | 160.6 | 75
2020-03-09 21:07:40.000000+0000 | 14 | -42.5 | 171.9 | 75
2020-03-10 00:00:00.000000+0000 | 14 | -42.3 | 171.5 | 76

(4 rows)
cqish:wmetro> exit

```

COMMANDS EXECUTED

```

create keyspace wmetro with replication={'class':'SimpleStrategy','replication_factor':1};
describe system_schema.keyspaces;
use wmetro;

create table driver(driver_name text primary key,password text,mobile bigint,current_position text,skill set<text>);
begin batch
insert into driver(driver_name,password,mobile,current_position,skill) values('Sven','123abc',9000000001,'Wellington',{'Passenger'})
insert into driver(driver_name,password,mobile,current_position,skill) values('Fred','234bcd',9000000002,'Upper Hutt',{'Passenger'})
insert into driver(driver_name,password,mobile,current_position,skill) values('Jane','345cde',9000000003,'Lower Hutt',{'Goods'})
apply batch;
select * from driver;

create table vehicles(vehicle_id text primary key,status text,type text);
begin batch
insert into vehicles(vehicle_id,status,type) values('WR1234','Maintenance','Gulliver');
insert into vehicles(vehicle_id,status,type) values('WR1235','Lowe Hutt','Gulliver');

```

```

insert into vehicles(vehicle_id,status,type) values('WR1236','Upper Hutt','Kiwi Rail');
insert into vehicles(vehicle_id,status,type) values('WR1237','Wellington','Kiwi Rail');
apply batch;
select * from vehicles;

create table time_table(line_name text, service_num int,station_name text,latitude double,longitude double,time int,distance double, primary key(line_name, station_name));
begin batch
insert into
time_table(line_name,service_num,station_name,latitude,longitude,time,distance) values('LINE1',2,'Wellington',-40.28,174.77,0605,0.0)
insert into time_table(line_name,service_num,station_name,latitude,longitude,time,distance) values('LINE1',2,'Upper Hutt',-41.12,175.07,1025,34.3)
insert into time_table(line_name,service_num,station_name,latitude,longitude,time,distance) values('LINE2',2,'Wellington',-40.28,174.77,1130,34.3)
insert into time_table(line_name,service_num,station_name,latitude,longitude,time,distance) values('LINE2',2,'Upper Hutt',-41.12,175.07,1550,0.0)
apply batch;
select * from time_table;

create table vehicle_usage(vehicle_id text primary key,total_distance counter);
update vehicle_usage set total_distance = total_distance + 3000 where vehicle_id='WR1234';
update vehicle_usage set total_distance = total_distance + 5000 where vehicle_id='WR1235';
update vehicle_usage set total_distance = total_distance + 3500 where vehicle_id='WR1236';
update vehicle_usage set total_distance = total_distance + 5300 where vehicle_id='WR1237';
select * from VEHICLE_USAGE;

CREATE table data_point(day int,sequence timestamp primary key,latitude double,longitude double,speed double);
begin batch
insert into data_point(day,sequence,latitude,longitude,speed) values(14,'2020-03-10',-42.30,171.50,76)
insert into data_point(day,sequence,latitude,longitude,speed) values(14,'2020-03-10 10:07:40+1300',-42.50,171.90,75)
insert into data_point(day,sequence,latitude,longitude,speed) values(14,'2020-03-10 10:30:25+1300',-44.50,160.60,75)
apply batch;
select * from data_point;

insert into driver(driver_name,password,mobile,current_position,skill) values('William','678bcf',9000000004,'Wellington',{'Goods','Passenger'})
select * from driver;

update driver set password='678bcf' where driver_name='William' if password='789ghf';
select * from driver;

update driver set current_position='Maintenance' where driver_name='William';
select * from driver;

update driver set skill={'Passenger','Goods'} where driver_name='Jane';
select * from driver;

alter table driver add workdays int ;
update driver set workdays=22 where driver_name='William';
update driver set workdays=22 where driver_name='Fred';
update driver set workdays=21 where driver_name='Jane';
update driver set workdays=22 where driver_name='Sven';

insert into vehicles(vehicle_id,status,type) values('WR1238','Wellington','Diesel');
select * from vehicles;

update vehicles set status='Upper Hutt' where vehicle_id='WR1234';
select * from vehicles;

alter table vehicles add daily_distance int;
update vehicles set daily_distance = 200 where vehicle_id='WR1234';
update vehicles set daily_distance = 180 where vehicle_id='WR1236';
update vehicles set daily_distance = 300 where vehicle_id='WR1237';
update vehicles set daily_distance = 50 where vehicle_id='WR1235';
update vehicles set daily_distance = 300 where vehicle_id='WR1238';
select * from vehicles;

UPDATE VEHICLE_USAGE SET TOTAL_DISTANCE=TOTAL_DISTANCE+50 WHERE VEHICLE_ID='WR1235';
SELECT * FROM VEHICLE_USAGE;

begin batch
insert into time_table(line_name,service_num,station_name,latitude,longitude,time,distance) values('LINE3',2,'Melling',-43.20,171.60,1540,330.0)
insert into time_table(line_name,service_num,station_name,latitude,longitude,time,distance) values('LINE3',2,'Waterloo',-53.50,165.70,0940,0.0)
insert into time_table(line_name,service_num,station_name,latitude,longitude,time,distance) values('LINE4',2,'Melling',-43.20,171.60,1540,0.0)
insert into time_table(line_name,service_num,station_name,latitude,longitude,time,distance) values('LINE4',2,'Waterloo',-53.50,165.70,0940,330.0)
apply batch ;
SELECT * FROM TIME_TABLE;

update time_table set service_num = 1 where line_name='LINE4' and station_name='Waterloo';
update time_table set service_num = 1 where line_name='LINE4' and station_name='Melling';
update time_table set service_num = 1 where line_name='LINE3' and station_name='Waterloo';
update time_table set service_num = 1 where line_name='LINE3' and station_name='Melling';
select * from time_table;

insert into data_point(day,sequence,latitude,longitude,speed) values(15,'2020-03-04',-52.10,168.90,67);
SELECT * FROM DATA_POINT;

EXIT

```

1. Create a new collection

```
use student
```

2. Insert a value

```
db.student.insert({
  "Name" : "abc",
  "RollNo:" : 1,
  "Age" : 21,
  "ContactNo" : "9000000001",
  "EmailId": "abc@example.com"
})
```

3. Insert multiple values at once

```
var Mystudents = [
  {
    "Name" : "def",
    "RollNo:" : 2,
    "Age" : 21,
    "ContactNo" : "9000000002",
    "EmailId": "def@example.com"
  },
  {
    "Name" : "ghi",
    "RollNo:" : 3,
    "Age" : 22,
    "ContactNo" : "9000000003",
    "EmailId": "ghi@example.com"
  },
  {
    "Name" : "jkl",
    "RollNo:" : 4,
    "Age" : 21,
    "ContactNo" : "9000000004",
    "EmailId": "jkl@example.com"
  },
  {
    "Name" : "mno",
    "RollNo:" : 5,
    "Age" : 22,
    "ContactNo" : "9000000005",
    "EmailId": "mno@example.com"
  },
]
```

```
db.student.insert(Mystudents);
```

4. Print all current values

```
db.getCollection('student').find({}).forEach(printjson)
```

```
{
  "_id" : ObjectId("60aaca0014c26c4c599fc88e"),
  "Name" : "abc",
  "RollNo:" : 1,
  "Age" : 21,
  "ContactNo" : "90000000001",
  "EmailId" : "abc@example.com"
}
{
  "_id" : ObjectId("60aacb7214c26c4c599fc893"),
  "Name" : "def",
  "RollNo:" : 3,
  "Age" : 21,
  "ContactNo" : "9000000002",
  "EmailId" : "def@example.com"
}
{
  "_id" : ObjectId("60aacb7214c26c4c599fc894"),
  "Name" : "ghi",
  "RollNo:" : 3,
  "Age" : 22,
  "ContactNo" : "9000000003",
  "EmailId" : "ghi@example.com"
}
{
  "_id" : ObjectId("60aacb7214c26c4c599fc895"),
  "Name" : "jkl",
  "RollNo:" : 4,
  "Age" : 21,
  "ContactNo" : "9000000004",
  "EmailId" : "jkl@example.com"
}
{
  "_id" : ObjectId("60aacb7214c26c4c599fc896"),
  "Name" : "mno",
  "RollNo:" : 5,
  "Age" : 22,
  "ContactNo" : "9000000005",
  "EmailId" : "mno@example.com"
}
```

5. Update RollNo of a student

```
db.student.update({"Name" : "abc"},{$set: { "RollNo" : 10}})
```

```
db.getCollection('student').find({"RollNo":10}).forEach(printjson)
```

```
{
  "_id" : ObjectId("60aaca0014c26c4c599fc88e"),
  "Name" : "abc",
  "RollNo:" : 10,
  "Age" : 21,
  "ContactNo" : "90000000001",
  "EmailId" : "abc@example.com"
}
```

6. Update Name of a student

```
db.student.update({"Name" : "abc"},{$set: { "Name" : "FEM"}})
```

```
db.getCollection('student').find({"Name" : "FEM"}).forEach(printjson)
```

```
{
  "_id" : ObjectId("60aaca0014c26c4c599fc88e"),
  "Name" : "FEM",
  "RollNo:" : 1,
  "Age" : 21,
  "ContactNo" : "90000000001",
  "EmailId" : "abc@example.com"
}
```

7. Export to json

```
mongoexport --db testdb --collection student --out C:\Users\Desktop\student.json
{"_id":{"$_oid":"60aaca0014c26c4c599fc88e"},"Name":"FEM","age":21.0,"Contact":"9000000001","email":"abc@example.com","RollNo":10.0}
{"_id":{"$_oid":"60aacb7214c26c4c599fc893"},"Name":"def","RollNo":2.0,"Age":21.0,"ContactNo":"9000000002","EmailId":"def@lab.com"}
{"_id":{"$_oid":"60aacb7214c26c4c599fc894"},"Name":"ghi","RollNo":3.0,"Age":22.0,"ContactNo":"9000000003","EmailId":"ghi@lab.com"}
{"_id":{"$_oid":"60aacb7214c26c4c599fc895"},"Name":"jkl","RollNo":4.0,"Age":21.0,"ContactNo":"9000000004","EmailId":"jkl@lab.com"}
{"_id":{"$_oid":"60aacb7214c26c4c599fc896"},"Name":"mno","RollNo":5.0,"Age":22.0,"ContactNo":"9000000005","EmailId":"mno@lab.com"}
```

8. Drop student

```
db.getCollection('student').drop()
```

9. Import from exported file

```
mongoimport --db stud --collection stud C:\Users\Desktop\student.json
```

10. Print values

```
use stud
db.getCollection('stud').find({}).forEach(printjson)
```

```
{
  "_id" : ObjectId("60aaca0014c26c4c599fc88e"),
  "Name" : "abc",
  "RollNo" : 1,
  "Age" : 21,
  "ContactNo" : "90000000001",
  "EmailId" : "abc@example.com"
}
{
  "_id" : ObjectId("60aacb7214c26c4c599fc893"),
  "Name" : "def",
  "RollNo" : 3,
  "Age" : 21,
  "ContactNo" : "90000000002",
  "EmailId" : "def@example.com"
}
{
  "_id" : ObjectId("60aacb7214c26c4c599fc894"),
  "Name" : "ghi",
  "RollNo" : 3,
  "Age" : 22,
  "ContactNo" : "90000000003",
  "EmailId" : "ghi@example.com"
}
{
  "_id" : ObjectId("60aacb7214c26c4c599fc895"),
  "Name" : "jkl",
  "RollNo" : 4,
  "Age" : 21,
  "ContactNo" : "90000000004",
  "EmailId" : "jkl@example.com"
}
{
  "_id" : ObjectId("60aacb7214c26c4c599fc896"),
  "Name" : "mno",
  "RollNo" : 5,
  "Age" : 22,
  "ContactNo" : "90000000005",
  "EmailId" : "mno@example.com"
}
```

Overview 'localhost:9000' (active)

Started:	Mon Jun 14 03:46:43 +0530 2021
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled:	Tue Jul 07 00:14:00 +0530 2020 by brahma from branch-3.3.0
Cluster ID:	CID-be544c5e-c20e-4a5a-b2e1-8773b2447853
Block Pool ID:	BP-248846901-172.28.0.2-1623622203021

Summary

Security is off.

Safemode is off.

2 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 2 total filesystem object(s).

Heap Memory used 85.6 MB of 394 MB Heap Memory. Max Heap Memory is 3.17 GB.

Non Heap Memory used 56.95 MB of 60.49 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	107.72 GB
Configured Remote Capacity:	0 B
DFS Used:	28 KB (0%)
Non DFS Used:	39.74 GB
DFS Remaining:	67.96 GB (63.09%)

DataNode on 93c45e96a80e:9866

Cluster ID:	CID-be544c5e-c20e-4a5a-b2e1-8773b2447853
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af

Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	BP-248846901-172.28.0.2-1623622203021	RUNNING	0s	an hour	0 B (128 MB)

Volume Information

Directory	Storage Type	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/content/datanode	DISK	28 KB	67.96 GB	0 B	0 B	0

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -cat /dir1/file1.txt
Sample text
hello line 1
hello line 2
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hadoop fs -appendToFile ~/Desktop/file2.txt /dir1/file1.txt
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -cat /dir1/file1.txt
Sample text
hello line 1
hello line 2
This is a sample text
line 1
line 2
This is a new file
I plan to append this
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -cat /dir1/file1.txt
This Sample text
hello line 1
hello line 2
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```



```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -cat /dir1/file1.txt
Sample text
hello line 1
hello line 2
This is a sample text
line 1
line 2
This is a new file
I plan to append this
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -cp /dir1/file1.txt /sample/file2.txt
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -cat /sample/file2.txt
Sample text
hello line 1
hello line 2
This is a sample text
line 1
line 2
This is a new file
I plan to append this
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -copyFromLocal ~/Desktop/file1.txt /dir1
copyFromLocal: `/dir1/file1.txt': File exists
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -ls /dir1
Found 1 items
-rw-r--r--  1 hduser supergroup      38 2021-04-19 03:40 /dir1/file1.txt
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hadoop fs -count -e /dir1/file1.txt
      0          1          38 Replicated /dir1/file1.txt
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hadoop fs -count -e /dir1
      1          1          38 EC: /dir1
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```

```
is hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -mkdir /dir1
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x  - hduser supergroup          0 2021-04-19 03:27 /dir1
drwxr-xr-x  - hduser supergroup          0 2021-04-19 02:49 /user
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hadoop fs -du -h -v /sample/file1.txt
SIZE    DISK_SPACE_CONSUMED_WITH_ALL_REPLICAS    FULL_PATH_NAME
11Show Applications    /sample/file1.txt
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```

Terminal
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin\$ hadoop fs -mkdir /dir2
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin\$ hadoop fs -mv /dir1/file1.txt /dir2
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin\$ hadoop fs -ls /dir2
Found 1 items
-rw-r--r-- 1 hduser supergroup 118 2021-04-19 03:50 /dir2/file1.txt
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin\$

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hadoop fs -rm /dir2/file1.txt
```

```
Deleted /dir2/file1.txt
```

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hadoop fs -rm -R /dir1
```

```
Deleted /dir1
```

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hadoop fs -ls /
```

```
Found 3 items
```

drwxr-xr-x	-	hduser	supergroup	0	2021-04-19 04:36	/dir2
drwxr-xr-x	-	hduser	supergroup	0	2021-04-19 04:26	/sample
drwxr-xr-x	-	hduser	supergroup	0	2021-04-19 02:49	/user

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```

truncate /sample/file1.txt to 3. No such file or directory

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -cat /sample/file1.txt
```

Sample text

hello line 1

hello line 2

This is a sample text

line 1

line 2

This is a new file

I plan to append this

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hadoop fs -truncate 3 /sample/file1.txt
```

Truncating /sample/file1.txt to length: 3. Wait for block recovery to complete before further updating this file.

```
hduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$ hdfs dfs -cat /sample/file1.txt
```

```
Samhduser@ubuntu:~/Downloads/hadoop-3.2.1/bin$
```



```
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -mkdir /input_dir
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-05-04 08:31 /input_dir
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ cd ..
hadoop@ubuntu:~/hadoop$ cd hadoop-3.3.0/
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -mkdir /input_dir
mkdir: `/input_dir': File exists
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-05-04 08:31 /input_dir
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ cd bin
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/bin$ cd ..
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ cd sbin
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -mkdir /input_dir
mkdir: `/input_dir': File exists
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-05-04 08:31 /input_dir
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ cd ..
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ cd share/hadoop/mapreduce/
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/share/hadoop/mapreduce$ hdfs dfs -put ~/input.txt /input_dir/input.txt
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/share/hadoop/mapreduce$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-05-04 08:32 /input_dir
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-3.3.0.jar wordcount /input_dir /output_dir
2021-05-04 08:33:36,338 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2021-05-04 08:33:36,789 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1620142240400_0001
2021-05-04 08:33:37,050 INFO input.FileInputFormat: Total input files to process : 1
2021-05-04 08:33:37,146 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-04 08:33:37,347 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1620142240400_0001
2021-05-04 08:33:37,347 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-04 08:33:37,649 INFO conf.Configuration: resource-types.xml not found
2021-05-04 08:33:37,649 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-04 08:33:38,131 INFO impl.YarnClientImpl: Submitted application application_1620142240400_0001
2021-05-04 08:33:38,193 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1620142240400_0001/
2021-05-04 08:33:38,193 INFO mapreduce.Job: Running job: job_1620142240400_0001
2021-05-04 08:33:50,677 INFO mapreduce.Job: Job job_1620142240400_0001 running in uber mode : false
2021-05-04 08:33:50,681 INFO mapreduce.Job: map 0% reduce 0%
2021-05-04 08:34:01,923 INFO mapreduce.Job: map 100% reduce 0%
2021-05-04 08:34:11,111 INFO mapreduce.Job: map 100% reduce 100%
2021-05-04 08:34:12,142 INFO mapreduce.Job: Job job_1620142240400_0001 completed successfully
2021-05-04 08:34:12,239 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=38
    FILE: Number of bytes written=527741
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=152
    HDFS: Number of bytes written=20
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=8813
    Total time spent by all reduces in occupied slots (ms)=6239
    Total time spent by all map tasks (ms)=8813
    Total time spent by all reduce tasks (ms)=6239
    Total vcore-milliseconds taken by all map tasks=8813
    Total vcore-milliseconds taken by all reduce tasks=6239
    Total megabyte-milliseconds taken by all map tasks=9024512
    Total megabyte-milliseconds taken by all reduce tasks=6388736
  Map-Reduce Framework
    Map input records=3
    Map output records=9
    Map output bytes=79
    Map output materialized bytes=38
    Input split bytes=106
    Combine input records=9
    Combine output records=3
    Reduce input groups=3
    Reduce shuffle bytes=38
    Reduce input records=3
    Reduce output records=3
    Spilled Records=6
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=182
    CPU time spent (ms)=2120
    Physical memory (bytes) snapshot=458768384
    Virtual memory (bytes) snapshot=5063098368
    Total committed heap usage (bytes)=351272960
    Peak Map Physical memory (bytes)=267317248
    Peak Map Virtual memory (bytes)=2528882688
    Peak Reduce Physical memory (bytes)=191451136
    Peak Reduce Virtual memory (bytes)=2534215680
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=46
  File Output Format Counters
    Bytes Written=20
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/share/hadoop/mapreduce$ hdfs dfs -ls /output_dir
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-05-04 08:34 /output_dir/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 20 2021-05-04 08:34 /output_dir/part-r-00000
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/share/hadoop/mapreduce$ hdfs dfs -cat /output_dir/*
bear 3
car 2
deer 4
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/share/hadoop/mapreduce$ hdfs dfs -cat /output_dir/part-r-00000
bear 3
car 2
deer 4
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/share/hadoop/mapreduce$
```

```
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ sbin/start-sh
bash: sbin/start-sh: No such file or directory
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ jps
3187 NodeManager
2807 SecondaryNameNode
3515 Jps
2491 NameNode
2620 DataNode
3052 ResourceManager
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ cd sbin
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -mkdir /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-05-05 02:51 /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -put ~/input.txt /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hadoop jar ~/Desktop/wordcount.jar wordcount /input /output
2021-05-05 02:53:07,566 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2021-05-05 02:53:08,539 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
2021-05-05 02:53:08,559 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1620207953365_0001
2021-05-05 02:53:09,047 INFO input.FileInputFormat: Total input files to process : 1
2021-05-05 02:53:09,293 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-05 02:53:09,984 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1620207953365_0001
2021-05-05 02:53:09,984 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-05 02:53:10,248 INFO conf.Configuration: resource-types.xml not found
2021-05-05 02:53:10,248 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-05 02:53:10,613 INFO impl.YarnClientImpl: Submitted application application_1620207953365_0001
2021-05-05 02:53:10,712 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1620207953365_0001/
2021-05-05 02:53:10,713 INFO mapreduce.Job: Running job: job_1620207953365_0001
2021-05-05 02:53:25,000 INFO mapreduce.Job: Job job_1620207953365_0001 running in uber mode : false
2021-05-05 02:53:25,002 INFO mapreduce.Job: map 0% reduce 0%
2021-05-05 02:53:39,008 INFO mapreduce.Job: map 100% reduce 0%
2021-05-05 02:53:47,098 INFO mapreduce.Job: map 100% reduce 100%
2021-05-05 02:53:48,123 INFO mapreduce.Job: Job job_1620207953365_0001 completed successfully
2021-05-05 02:53:48,257 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=38
        FILE: Number of bytes written=527269
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=148
        HDFS: Number of bytes written=20
        HDFS: Number of read operations=8
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=10842
        Total time spent by all reduces in occupied slots (ms)=4966
        Total time spent by all map tasks (ms)=10842
        Total time spent by all reduce tasks (ms)=4966
        Total vcore-milliseconds taken by all map tasks=10842
        Total vcore-milliseconds taken by all reduce tasks=4966
        Total megabyte-milliseconds taken by all map tasks=11102208
        Total megabyte-milliseconds taken by all reduce tasks=5085184
    Map-Reduce Framework
        Map input records=3
        Map output records=9
        Map output bytes=79
        Map output materialized bytes=38
        Input split bytes=102
        Combine input records=9
        Combine output records=3
        Reduce input groups=3
        Reduce shuffle bytes=38
        Reduce input records=3
        Reduce output records=3
        Spilled Records=6
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=254
        CPU time spent (ms)=1140
        Physical memory (bytes) snapshot=458858496
        Virtual memory (bytes) snapshot=5065277440
        Total committed heap usage (bytes)=383254528
        Peak Map Physical memory (bytes)=269565952
        Peak Map Virtual memory (bytes)=2530467840
        Peak Reduce Physical memory (bytes)=189292544
        Peak Reduce Virtual memory (bytes)=2534809600
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=46
    File Output Format Counters
        Bytes Written=20
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -ls /output
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-05-05 02:53 /output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 20 2021-05-05 02:53 /output/part-r-00000
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -cat /output/part-r-00000
bear 3
car 2
deer 4
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$
```

```
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ jps
2544 NameNode
2865 SecondaryNameNode
3284 NodeManager
2677 DataNode
3447 Jps
3132 ResourceManager
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -mkdir /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-05-05 10:03 /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -put ~/annual_usage.txt /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hadoop jar ~/Desktop/annualusage.jar abc.abc /input /output
2021-05-05 10:06:15,493 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2021-05-05 10:06:17,926 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
2021-05-05 10:06:18,017 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1620234033694_0001
2021-05-05 10:06:19,440 INFO input.FileInputFormat: Total input files to process : 1
2021-05-05 10:06:19,801 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-05 10:06:22,294 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1620234033694_0001
2021-05-05 10:06:22,294 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-05 10:06:23,084 INFO conf.Configuration: resource-types.xml not found
2021-05-05 10:06:23,085 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-05 10:06:24,339 INFO impl.YarnClientImpl: Submitted application application_1620234033694_0001
2021-05-05 10:06:24,536 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1620234033694_0001/
2021-05-05 10:06:24,538 INFO mapreduce.Job: Running job: job_1620234033694_0001
2021-05-05 10:07:01,013 INFO mapreduce.Job: Job job_1620234033694_0001 running in uber mode : false
2021-05-05 10:07:01,017 INFO mapreduce.Job: map 0% reduce 0%
2021-05-05 10:07:22,768 INFO mapreduce.Job: map 100% reduce 0%
2021-05-05 10:07:41,087 INFO mapreduce.Job: map 100% reduce 100%
2021-05-05 10:07:45,195 INFO mapreduce.Job: Job job_1620234033694_0001 completed successfully
2021-05-05 10:07:45,563 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=527181
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=457
    HDFS: Number of bytes written=24
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17272
    Total time spent by all reduces in occupied slots (ms)=14905
    Total time spent by all map tasks (ms)=17272
    Total time spent by all reduce tasks (ms)=14905
    Total vcore-milliseconds taken by all map tasks=17272
    Total vcore-milliseconds taken by all reduce tasks=14905
    Total megabyte-milliseconds taken by all map tasks=17686528
    Total megabyte-milliseconds taken by all reduce tasks=15262720
  Map-Reduce Framework
    Map input records=5
    Map output records=5
    Map output bytes=45
    Map output materialized bytes=39
    Input split bytes=109
    Combine input records=5
    Combine output records=3
    Reduce input groups=3
    Reduce shuffle bytes=39
    Reduce input records=3
    Reduce output records=3
    Spilled Records=6
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=351
    CPU time spent (ms)=3330
    Physical memory (bytes) snapshot=442978304
    Virtual memory (bytes) snapshot=5061464064
    Total committed heap usage (bytes)=317718528
    Peak Map Physical memory (bytes)=257789952
    Peak Map Virtual memory (bytes)=2529673216
    Peak Reduce Physical memory (bytes)=185188352
    Peak Reduce Virtual memory (bytes)=2531790848
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=348
  File Output Format Counters
    Bytes Written=24
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ cd sbin
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -ls /output
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-05-05 10:07 /output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 24 2021-05-05 10:07 /output/part-r-00000
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -cat /output/part-r-00000
1979 25
1980 29
1981 34
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$
```

```
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ jps
2544 NameNode
2865 SecondaryNameNode
3284 NodeManager
2677 DataNode
3447 Jps
3132 ResourceManager
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -mkdir /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-05-05 10:03 /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hdfs dfs -put ~/annual_usage.txt /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ hadoop jar ~/Desktop/annualusage.jar abc.abc /input /output
2021-05-05 10:06:15,493 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2021-05-05 10:06:17,926 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
2021-05-05 10:06:18,017 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1620234033694_0001
2021-05-05 10:06:19,440 INFO input.FileInputFormat: Total input files to process : 1
2021-05-05 10:06:19,801 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-05 10:06:22,294 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1620234033694_0001
2021-05-05 10:06:22,294 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-05 10:06:23,084 INFO conf.Configuration: resource-types.xml not found
2021-05-05 10:06:23,085 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-05 10:06:24,339 INFO impl.YarnClientImpl: Submitted application application_1620234033694_0001
2021-05-05 10:06:24,536 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1620234033694_0001/
2021-05-05 10:06:24,538 INFO mapreduce.Job: Running job: job_1620234033694_0001
2021-05-05 10:07:01,013 INFO mapreduce.Job: Job job_1620234033694_0001 running in uber mode : false
2021-05-05 10:07:01,017 INFO mapreduce.Job: map 0% reduce 0%
2021-05-05 10:07:22,768 INFO mapreduce.Job: map 100% reduce 0%
2021-05-05 10:07:41,087 INFO mapreduce.Job: map 100% reduce 100%
2021-05-05 10:07:45,195 INFO mapreduce.Job: Job job_1620234033694_0001 completed successfully
2021-05-05 10:07:45,563 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=527181
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=457
    HDFS: Number of bytes written=24
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=17272
    Total time spent by all reduces in occupied slots (ms)=14905
    Total time spent by all map tasks (ms)=17272
    Total time spent by all reduce tasks (ms)=14905
    Total vcore-milliseconds taken by all map tasks=17272
    Total vcore-milliseconds taken by all reduce tasks=14905
    Total megabyte-milliseconds taken by all map tasks=17686528
    Total megabyte-milliseconds taken by all reduce tasks=15262720
  Map-Reduce Framework
    Map input records=5
    Map output records=5
    Map output bytes=45
    Map output materialized bytes=39
    Input split bytes=109
    Combine input records=5
    Combine output records=3
    Reduce input groups=3
    Reduce shuffle bytes=39
    Reduce input records=3
    Reduce output records=3
    Spilled Records=6
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=351
    CPU time spent (ms)=3330
    Physical memory (bytes) snapshot=442978304
    Virtual memory (bytes) snapshot=5061464064
    Total committed heap usage (bytes)=317718528
    Peak Map Physical memory (bytes)=257789952
    Peak Map Virtual memory (bytes)=2529673216
    Peak Reduce Physical memory (bytes)=185188352
    Peak Reduce Virtual memory (bytes)=2531790848
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=348
  File Output Format Counters
    Bytes Written=24
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ cd sbin
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -ls /output
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-05-05 10:07 /output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 24 2021-05-05 10:07 /output/part-r-00000
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -cat /output/part-r-00000
1979 25
1980 29
1981 34
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$
```

```
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ jps
2789 SecondaryNameNode
3510 Jps
3176 NodeManager
3048 ResourceManager
2569 DataNode
2411 NameNode
hadoop@ubuntu:~/hadoop/hadoop-3.3.0$ cd sbin
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -mkdir /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-05-05 11:28 /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -put ~/sortwords_input.txt /input
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hadoop jar ~/Desktop/sortwords.jar sortwords.TopN /input /output
2021-05-05 11:33:18,034 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2021-05-05 11:33:21,923 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1620239174642_0001
2021-05-05 11:33:23,404 INFO input.FileInputFormat: Total input files to process : 1
2021-05-05 11:33:23,731 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-05 11:33:27,154 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1620239174642_0001
2021-05-05 11:33:27,155 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-05 11:33:28,437 INFO conf.Configuration: resource-types.xml not found
2021-05-05 11:33:28,438 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-05 11:33:29,765 INFO impl.YarnClientImpl: Submitted application application_1620239174642_0001
2021-05-05 11:33:29,936 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1620239174642_0001/
2021-05-05 11:33:29,940 INFO mapreduce.Job: Running job: job_1620239174642_0001
2021-05-05 11:34:43,278 INFO mapreduce.Job: Job job_1620239174642_0001 running in uber mode : false
2021-05-05 11:34:43,929 INFO mapreduce.Job: map 0% reduce 0%
2021-05-05 11:35:23,067 INFO mapreduce.Job: map 100% reduce 0%
2021-05-05 11:35:40,352 INFO mapreduce.Job: map 100% reduce 100%
2021-05-05 11:35:42,395 INFO mapreduce.Job: Job job_1620239174642_0001 completed successfully
2021-05-05 11:35:42,693 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=167
FILE: Number of bytes written=527455
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=183
HDFS: Number of bytes written=20
HDFS: Number of read operations=8
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=32866
Total time spent by all reduces in occupied slots (ms)=13362
Total time spent by all map tasks (ms)=32866
Total time spent by all reduce tasks (ms)=13362
Total vcore-milliseconds taken by all map tasks=32866
Total vcore-milliseconds taken by all reduce tasks=13362
Total megabyte-milliseconds taken by all map tasks=33654784
Total megabyte-milliseconds taken by all reduce tasks=13682688
Map-Reduce Framework
Map input records=5
Map output records=15
Map output bytes=131
Map output materialized bytes=167
Input split bytes=112
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=167
Reduce input records=15
Reduce output records=3
Spilled Records=30
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=267
CPU time spent (ms)=2400
Physical memory (bytes) snapshot=424538112
Virtual memory (bytes) snapshot=5063954432
Total committed heap usage (bytes)=325582848
Peak Map Physical memory (bytes)=236724224
Peak Map Virtual memory (bytes)=2529677312
Peak Reduce Physical memory (bytes)=187813888
Peak Reduce Virtual memory (bytes)=2534277120
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=71
File Output Format Counters
Bytes Written=20
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -ls /output
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-05-05 11:35 /output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 20 2021-05-05 11:35 /output/part-r-00000
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$ hdfs dfs -cat /output/part-r-00000
car 5
bear 4
dear 3
hadoop@ubuntu:~/hadoop/hadoop-3.3.0/sbin$
```

1. map()

```
prathibha@ubuntu:~$ spark-shell
2021-05-21 22:06:01,172 WARN util.Utils: Your hostname, ubuntu resolves to a loopback address: 127.0.1.1; using 192.168.113.128 instead (on interface ens33)
2021-05-21 22:06:01,175 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
2021-05-21 22:06:05,167 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://192.168.113.128:4040
Spark context available as 'sc' (master = local[*], app id = local-1621660004423).
Spark session available as 'spark'.
Welcome to

  ____      _
 / ___|    / \
| |  | |  / _ \
| |  | | / ___ \
| |  | |/_/   \_\
|_|  |_____/___\_\

version 3.1.1

Using Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 1.8.0_292)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val input = sc.parallelize(List(1,2,3,4))
input: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at <console>:24

scala> val result = input.map(x => x*x)
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[1] at map at <console>:25

scala> println(result.collect().mkString(","))
[Stage 0:>                                     (0 + 0) / 2[Stage 0:>                                (0 + 2) / 2
1,4,9,16

scala>
```

```
scala> val input = sc.parallelize(List(1,2,3,4))
```

```
input: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at <console>:24
```

```
scala> val result = input.map(x => x*x)
```

```
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[1] at map at <console>:25
```

```
scala> println(result.collect().mkString(","))
```

```
1,4,9,16
```

2. flatmap()

```
scala> val lines = sc.parallelize(List("hello world","hi"))
lines: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD[2] at parallelize at <console>:24

scala> val words = lines.flatMap(line => line.split(" "))
words: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[3] at flatMap at <console>:25

Terminal
scala> words.first()
res1: String = h
```

```
scala> val lines = sc.parallelize(List("hello world","hi"))
```

```
lines: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD[2] at parallelize at
<console>:24
scala> val words = lines.flatMap(line => line.split(""))
words: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[3] at flatMap at <console>:25

scala> words.first()
res1: String = h
```

3. filter()

```
scala> val input = sc.parallelize(List(1,2,3,4))
input: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize
at <console>:24

scala> val result = input.filter(x => x != 1)
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[1] at filter at <conso
le>:25

scala> println(result.collect().mkString(","))
[Stage 0:>                                     (0 + 0) / 2
[Stage 0:>                                     (0 + 2) / 2
2,3,4
```

```
scala> val input = sc.parallelize(List(1,2,3,4))
input: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at <console>:24

scala> val result = input.filter(x => x != 1)
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[1] at filter at <console>:25

scala> println(result.collect().mkString(","))
2,3,4
```

4. distinct()

```
scala> val input4 = sc.parallelize(List(1,2,2,3,3,4))
input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[2] at parallelize
at <console>:24

scala> val result = input4.distinct()
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[5] at distinct at <con
sole>:25

scala> println(result.collect().mkString(","))
4,2,1,3
```

```
scala> val input4 = sc.parallelize(List(1,2,2,3,3,4))
```

input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[2] at parallelize at <console>:24

scala> val result = input4.distinct()

result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[5] at distinct at <console>:25

scala> println(result.collect().mkString(","))

4,2,1,3

5. union()

```
scala> val input4 = sc.parallelize(List(1,2,3))
input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[6] at parallelize
at <console>:24
```

```
scala> val input5 = sc.parallelize(List(3,4,5))
input5: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[7] at parallelize
at <console>:24
```

```
scala> val result= input4.union(input5)
result: org.apache.spark.rdd.RDD[Int] = UnionRDD[8] at union at <console>:27
```

```
scala> println(result.collect().mkString(","))
```

1 2 3 3 4 5
Show Applications

scala> val input4 = sc.parallelize(List(1,2,3))

input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[6] at parallelize at <console>:24

scala> val input5 = sc.parallelize(List(3,4,5))

input5: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[7] at parallelize at <console>:24

scala> val result= input4.union(input5)

result: org.apache.spark.rdd.RDD[Int] = UnionRDD[8] at union at <console>:27

scala> println(result.collect().mkString(","))

1,2,3,3,4,5

6. intersection()


```
scala> val input4 = sc.parallelize(List(1,2,3))
input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at <console>:24

scala> val input5 = sc.parallelize(List(3,4,5))
input5: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[1] at parallelize at <console>:24

scala> val result= input4.intersection(input5)
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[7] at intersection at <console>:27

scala> println(result.collect().mkString(", "))
[Stage 0:>
/ 2[Stage 0:>      (0 + 2) / 2[Stage 0:>      (0 + 1) / 2[Stage 0:>      (0 + 2) / 2][Stage 1:>      (0 + 0)
                  (0 + 2) / 2][Stage 1:>      (0 + 2) / 2[Stage 2:>
                  (0 + 2) / 2
3
```

```
scala> val input4 = sc.parallelize(List(1,2,3))
input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[8] at parallelize at :24
```

```
scala> val input5 = sc.parallelize(List(3,4,5))
input5: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[9] at parallelize at :24
```

```
scala> val result = input4.intersection(input5)
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[16] at intersection at :27
```

```
scala> println(result.collect().mkString(", "))
3
```

7. subtract()

```
scala> val input4 = sc.parallelize(List(1,2,3))
input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[8] at parallelize at <console>:24

scala> val input5 = sc.parallelize(List(3,4,5))
input5: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[9] at parallelize at <console>:24

scala> val result= input4.subtract(input5)
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[13] at subtract at <console>:27

scala> println(result.collect().mkString(", "))
2,1
```

```
scala> val input4 = sc.parallelize(List(1,2,3))
input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[8] at parallelize at :24
```

```
scala> val input5 = sc.parallelize(List(3,4,5))
input5: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[9] at parallelize at :24
```

```
scala> val result = input4.subtract(input5)
result: org.apache.spark.rdd.RDD[Int] = MapPartitionsRDD[20] at subtract at :27
```

```
scala> println(result.collect().mkString(", "))
2,1
```

8. cartesian()

```
scala> val input4 = sc.parallelize(List(1,2,3))
input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[14] at parallelize at <console>:24

scala> val input5 = sc.parallelize(List(3,4,5))
input5: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[15] at parallelize at <console>:24

scala> val result= input4.cartesian(input5)
result: org.apache.spark.rdd.RDD[(Int, Int)] = CartesianRDD[16] at cartesian at <console>:27









scala> println(result.collect().mkString(", "))
(1,3),(1,4),(1,5),(2,3),(3,3),(2,4),(2,5),(3,4),(3,5)
```

```
scala> val input4 = sc.parallelize(List(1,2,3))
input4: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[8] at parallelize at :24
```

```
scala> val input5 = sc.parallelize(List(3,4,5))
input5: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[9] at parallelize at :24
```

```
scala> val result = input4.cartesian(input5)
result: org.apache.spark.rdd.RDD[(Int, Int)] = CartesianRDD[21] at cartesian at :27
```

```
scala> println(result.collect().mkString(", "))
(1,3),(1,4),(1,5),(2,3),(2,4),(2,5),(3,3),(3,4),(3,5)
```

MTanwirS Lab 8			ec3407d 14 days ago	 History
..				
	JoinDriver.java	Lab 8		14 days ago
	JoinReducer.java	Lab 8		14 days ago
	Posts.java	Lab 8		14 days ago
	TextPair.java	Lab 8		14 days ago
	Users.java	Lab 8		14 days ago
	sampleposts.tsv	Lab 8		14 days ago
	sampleusers.tsv	Lab 8		14 days ago

```

package mapreduce_join;

import java.io.*;

import org.apache.hadoop.io.*;

public class TextPair implements WritableComparable<TextPair> {

    private Text first;
    private Text second;

    public TextPair() {
        set(new Text(), new Text());
    }

    public TextPair(String first, String second) {
        set(new Text(first), new Text(second));
    }

    public TextPair(Text first, Text second) {
        set(first, second);
    }

    public void set(Text first, Text second) {
        this.first = first;
        this.second = second;
    }

    public Text getFirst() {
        return first;
    }

    public Text getSecond() {
        return second;
    }

    @Override
    public void write(DataOutput out) throws IOException {
        first.write(out);
        second.write(out);
    }

    @Override
    public void readFields(DataInput in) throws IOException {
        first.readFields(in);
        second.readFields(in);
    }

    @Override
    public int hashCode() {
        return first.hashCode() * 163 + second.hashCode();
    }

    @Override
    public boolean equals(Object o) {
        if (o instanceof TextPair) {
            TextPair tp = (TextPair) o;
            return first.equals(tp.first) && second.equals(tp.second);
        }
        return false;
    }

    @Override
    public String toString() {
        return first + "\t" + second;
    }

    @Override
    public int compareTo(TextPair tp) {
        int cmp = first.compareTo(tp.first);
        if (cmp != 0) {
            return cmp;
        }
        return second.compareTo(tp.second);
    }
} // ^^ TextPair

// vv TextPairComparator
public static class Comparator extends WritableComparator {

    private static final Text.Comparator TEXT_COMPARATOR = new Text.Comparator();

    public Comparator() {
        super(TextPair.class);
    }

    @Override
    public int compare(byte[] b1, int s1, int l1,
                      byte[] b2, int s2, int l2) {

        try {
            int firstL1 = WritableUtils.decodeVIntSize(b1[s1]) + readVInt(b1, s1);
            int firstL2 = WritableUtils.decodeVIntSize(b2[s2]) + readVInt(b2, s2);
            int cmp = TEXT_COMPARATOR.compare(b1, s1, firstL1, b2, s2, firstL2);
            if (cmp != 0) {
                return cmp;
            }
            return TEXT_COMPARATOR.compare(b1, s1 + firstL1, l1 - firstL1,
                                           b2, s2 + firstL2, l2 - firstL2);
        } catch (IOException e) {
            throw new IllegalArgumentException(e);
        }
    }
}

static {
    WritableComparator.define(TextPair.class, new Comparator());
}
// ^^ TextPairComparator

// vv TextPairFirstComparator
public static class FirstComparator extends WritableComparator {

    private static final Text.Comparator TEXT_COMPARATOR = new Text.Comparator();

    public FirstComparator() {
        super(TextPair.class);
    }

    @Override
    public int compare(byte[] b1, int s1, int l1,
                      byte[] b2, int s2, int l2) {


        try {
            int firstL1 = WritableUtils.decodeVIntSize(b1[s1]) + readVInt(b1, s1);
            int firstL2 = WritableUtils.decodeVIntSize(b2[s2]) + readVInt(b2, s2);
            return TEXT_COMPARATOR.compare(b1, s1, firstL1, b2, s2, firstL2);
        }
    }
}

```

```
    } catch (IOException e) {
        throw new IllegalArgumentException(e);
    }
}

@Override
public int compare(WritableComparable a, WritableComparable b) {
    if (a instanceof TextPair && b instanceof TextPair) {
        return ((TextPair) a).first.compareTo(((TextPair) b).first);
    }
    return super.compare(a, b);
}
}
// ^^ TextPairFirstComparator

// vv TextPair
}
```

 main ▾

BDAIab / LAB 8 / Users.java / <> Jump to ▾

Go to file

...

 MTanwirS Lab 8

Latest commit ec3487d 14 days ago  History

👤 1 contributor

29 lines (22 sloc) | 879 Bytes

RawBlame

```
1 package mapreduce_join;
2
3 import java.io.IOException;
4 import java.util.Iterator;
5 import org.apache.hadoop.conf.Configuration;
6 import org.apache.hadoop.fs.FSDataInputStream;
7 import org.apache.hadoop.fs.FSDataOutputStream;
8 import org.apache.hadoop.fs.FileSystem;
9 import org.apache.hadoop.fs.Path;
10 import org.apache.hadoop.io.LongWritable;
11 import org.apache.hadoop.io.Text;
12 import org.apache.hadoop.mapred.*;
13
14 import org.apache.hadoop.io.IntWritable;
15
16 public class Users extends MapReduceBase implements Mapper<LongWritable, Text, TextPair, Text> {
17
18     @Override
19     public void map(LongWritable key, Text value, OutputCollector<TextPair, Text> output, Reporter reporter)
20         throws IOException
21     {
22
23         String valueString = value.toString();
24         String[] SingleNodeData = valueString.split("\t");
25         output.collect(new TextPair(SingleNodeData[0], "1"), new Text(SingleNodeData[1]));
26     }
27 }
28
29
```

The screenshot shows the Databricks IDE interface. On the left is a file explorer showing a directory structure with files like spark-class.cmd, spark-shell, and spark-submit. The main area displays the Spark shell output, which includes the Spark version (3.1.2) and Scala version (2.12.10). The output text is as follows:
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at <http://9a4efbc34f91:4040>
Spark context available as 'sc' (master = local[*], app id = local-1623674006872).
Spark session available as 'spark'.
Welcome to

Spark version 3.1.2

Using Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 11.0.11)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
At the bottom, a status bar indicates 68.78 GB available disk space.

```

scala>
scala> val textFile = sc.textFile("/home/naman/textfile.txt")
textFile: org.apache.spark.rdd.RDD[String] = /home/naman/textfile.txt MapPartitionsRDD[1] at textFile at <console>:24

scala> val counts = textFile.flatMap(line => line.split(" ")).map(word => (word, 1)).reduceByKey(_ + _)
counts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at <console>:25

scala> import scala.collection.immutable.ListMap
import scala.collection.immutable.ListMap

scala> val sorted=ListMap(counts.collect.sortWith(_. _2 > _. _2):_*)
sorted: scala.collection.immutable.ListMap[String,Int] = ListMap(agarwal -> 3, name -> 3, naman -> 2, is -> 2, my -> 1)

scala> println(sorted)
ListMap(agarwal -> 3, name -> 3, naman -> 2, is -> 2, my -> 1)

scala> for((k,v)<-sorted) { if(v>2) { print(k+",") print(v) println() } }
<console>:27: error: value print is not a member of Unit
for((k,v)<-sorted) { if(v>2) { print(k+",") print(v) println() } }
                             ^

scala> for((k,v)<-sorted) { if(v>2) { print(k+",") print(v) println() } }
<console>:27: error: value print is not a member of Unit
for((k,v)<-sorted) { if(v>2) { print(k+",") print(v) println() } }
                             ^

scala> for((k,v)<-sorted)
| {
|   if(v>2)
|   {
|       print(k+",")
|       print(v)
|       println()
|   }
| }
agarwal,3
name,3
scala>

```