



BIG DATA ANALYTICS

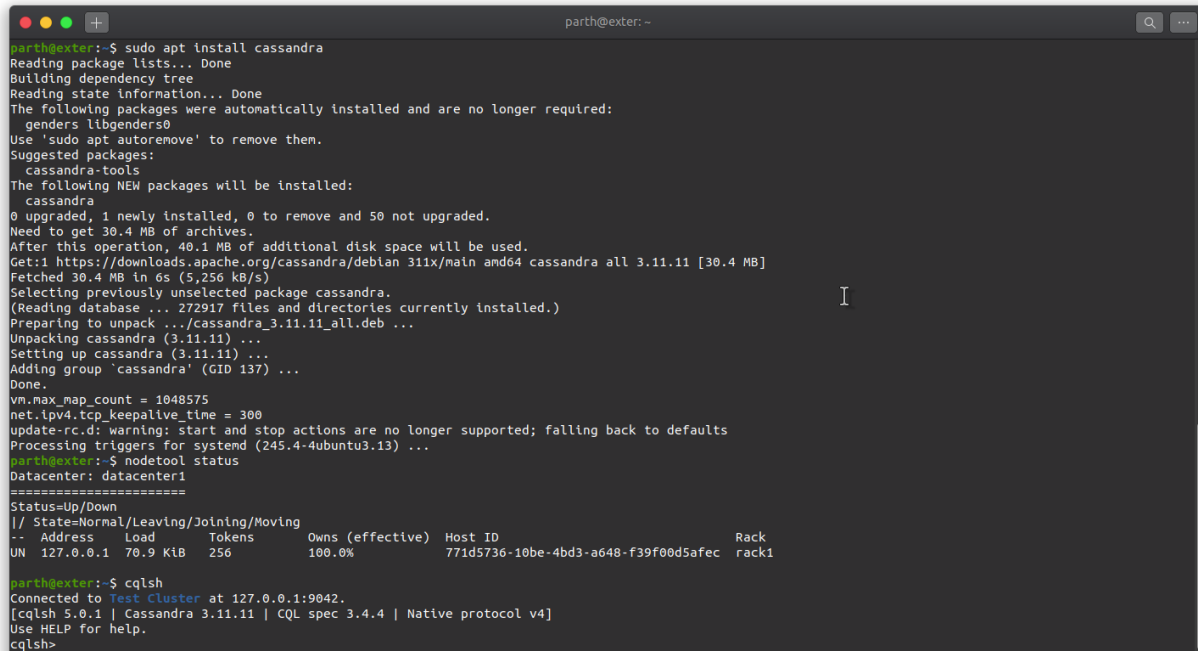
Practical 9

Setup Cassandra environment in your system and perform CRUD operation.

Prepared By

18BCE151 - Parth Kapadia

Install Apache Cassandra



```
parth@exter:~$ sudo apt install cassandra
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  genders libgenders0
Use 'sudo apt autoremove' to remove them.
Suggested packages:
  cassandra-tools
The following NEW packages will be installed:
  cassandra
0 upgraded, 1 newly installed, 0 to remove and 50 not upgraded.
Need to get 30.4 MB of archives.
After this operation, 40.1 MB of additional disk space will be used.
Get:1 https://downloads.apache.org/cassandra/debian 311x/main amd64 cassandra all 3.11.11 [30.4 MB]
Fetched 30.4 MB in 6s (5,256 kB/s)
Selecting previously unselected package cassandra.
(Reading database ... 272917 files and directories currently installed.)
Preparing to unpack .../cassandra_3.11.11_all.deb ...
Unpacking cassandra (3.11.11) ...
Setting up cassandra (3.11.11) ...
Adding group 'cassandra' (GID 137) ...
Done.
vm.max_map_count = 1048575
net.ipv4.tcp_keepalive_time = 300
update-rc.d: warning: start and stop actions are no longer supported; falling back to defaults
Processing triggers for systemd (245.4-4ubuntu3.13) ...
parth@exter:~$ nodetool status
Datacenter: datacenter1
=====
Status=Up/Down
// State=Normal/Leaving/Joining/Moving
-- Address      Load      Tokens     Owns (effective)  Host ID                               Rack
UN 127.0.0.1    70.9 KiB   256        100.0%            771d5736-10be-4bd3-a648-f39f00d5afec  rack1

parth@exter:~$ cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.11 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh>
```

Commands to Run

- **help**
- **CAPTURE**
- **USE**
- **Create keyspace**
- **Create Table**
- **INSERT**
- **UPDATE**
- **DELETE**

```

prac9 > ≡ cassandra.txt
1  ## help
2
3  help
4
5  -----
6
7  ## CAPTURE
8
9  CAPTURE '/home/parth/parth/practicals/BDA/prac9/cassandra_output.txt'
10
11 capture off;
12
13 -----
14
15 ## Describe
16
17 describe cluster;
18 describe keyspaces;
19 describe tables;
20
21 -----
22
23 ## Create Keyspace
24
25 CREATE KEYSPACE bda9
26 WITH replication = {'class':'SimpleStrategy', 'replication_factor' : 3};
27
28 USE bda9;
29
30 -----
31
32 ## Create Table
33
34
35 CREATE TABLE emp(
36     emp_id int PRIMARY KEY,
37     emp_name text,
38     emp_city text,
39     emp_sal varint,
40     emp_phone varint
41 );
42
43 -----
44
45 ## Insert
46
47 INSERT INTO emp (emp_id, emp_name, emp_city,
48     emp_phone, emp_sal) VALUES(1,'ram', 'Hyderabad', 9848022338, 50000);

```

```

50 -----
51
52 ## Update
53
54 UPDATE emp SET emp_city='Delhi',emp_sal=50000
55 | ... WHERE emp_id=2;
56
57 -----
58
59 ## Find with where
60
61 Why is there a need to set secondary index?
62 CREATE INDEX ON emp(emp_sal);
63 SELECT * FROM emp WHERE emp_sal=50000;
64
65 -----

```

```
66
67  ## Find all
68
69  SELECT * from emp;
70
71  -----
72
73  ## Delete
74
75  DELETE emp_sal FROM emp WHERE emp_id=1;
76
77  DELETE FROM emp WHERE emp_id=1;
78  -----
79
```

Output

```
parth@exter: ~
cqlsh> help

Documented shell commands:
=====
CAPTURE CLS          COPY DESCRIBE EXPAND LOGIN SERIAL SOURCE UNICODE
CLEAR   CONSISTENCY DESC EXIT      HELP   PAGING SHOW   TRACING

CQL help topics:
=====
AGGREGATES          CREATE_KEYSPACE      DROP_TRIGGER         TEXT
ALTER_KEYSPACE      CREATE_MATERIALIZED_VIEW DROP_TYPE            TIME
ALTER_MATERIALIZED_VIEW CREATE_ROLE           DROP_USER            TIMESTAMP
ALTER_TABLE         CREATE_TABLE         FUNCTIONS            TRUNCATE
ALTER_TYPE          CREATE_TRIGGER       GRANT               TYPES
ALTER_USER          CREATE_TYPE          INSERT              UPDATE
APPLY              CREATE_USER          INSERT_JSON         USE
ASCII              DATE                INT                 UUID
BATCH              DELETE              JSON
BEGIN             DROP_AGGREGATE      KEYWORDS
BLOB              DROP_COLUMNFAMILY  LIST_PERMISSIONS
BOOLEAN          DROP_FUNCTION      LIST_ROLES
COUNTER         DROP_INDEX         LIST_USERS
CREATE_AGGREGATE  DROP_KEYSPACE      PERMISSIONS
CREATE_COLUMNFAMILY DROP_MATERIALIZED_VIEW REVOKE
CREATE_FUNCTION   DROP_ROLE          SELECT
CREATE_INDEX      DROP_TABLE         SELECT_JSON

cqlsh>
```

```
parth@exter: ~
Functions aggregates indexes types columns

Keyspace system_auth
-----
resource_role_permissions_index role_permissions role_members roles

Keyspace system
-----
available_ranges peers batchlog transferred_ranges
batches compaction_history size_estimates hints
prepared_statements sstable_activity built_views
"IndexInfo" peer_events range_xfers
views_builds_in_progress paxos local

Keyspace system_distributed
-----
repair_history view_build_status parent_repair_history

cqlsh> CREATE TABLE emp(
...     emp_id int PRIMARY KEY,
...     emp_name text,
...     emp_city text,
...     emp_sal varint,
...     emp_phone varint
... );
InvalidRequest: Error from server: code=2200 [Invalid query] message="No keyspace has been specified. USE a keyspace, or explicitly specify keyspace.table name"
cqlsh>
cqlsh> CREATE KEYSPACE bda9
... WITH replication = {'class': 'SimpleStrategy', 'replication_factor' : 3};
cqlsh> USE bda9
... ;
cqlsh:bda9> CREATE TABLE emp(
...     emp_id int PRIMARY KEY,
...     emp_name text,
...     emp_city text,
...     emp_sal varint,
...     emp_phone varint
... );
cqlsh:bda9>
```

```
... WITH replication = {'class':'SimpleStrategy', 'replication_factor' : 3};
cqlsh> USE bda9
... ;
cqlsh:bda9> CREATE TABLE emp(
...     emp_id int PRIMARY KEY,
...     emp_name text,
...     emp_city text,
...     emp_sal varint,
...     emp_phone varint
... );
cqlsh:bda9> INSERT INTO emp (emp_id, emp_name, emp_city,
...     emp_phone, emp_sal) VALUES(1,'ran', 'Hyderabad', 9848022338, 50000);
cqlsh:bda9> INSERT INTO emp (emp_id, emp_name, emp_city,
...     emp_phone, emp_sal) VALUES(2,'robin', 'Hyderabad', 9848022339, 40000);
cqlsh:bda9> SELECT * FROM emp;
cqlsh:bda9> SELECT * FROM emp;
cqlsh:bda9> describe tables;

emp

cqlsh:bda9> describe table emp;
SyntaxException: line 1:0 no viable alternative at input 'desribe' ([desribe]...)
cqlsh:bda9> describe table emp;

CREATE TABLE bda9.emp (
  emp_id int PRIMARY KEY,
  emp_city text,
  emp_name text,
  emp_phone varint,
  emp_sal varint
) 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:bda9> select * from emp;
cqlsh:bda9>
```

```
CREATE TABLE bda9.emp (
  emp_id int PRIMARY KEY,
  emp_city text,
  emp_name text,
  emp_phone varint,
  emp_sal varint
) 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:bda9> select * from emp;
cqlsh:bda9> UPDATE emp SET emp_city='Delhi',emp_sal=50000
... WHERE emp_id=2;
cqlsh:bda9> select * from emp;
cqlsh:bda9> SELECT emp_name, emp_sal from emp;
cqlsh:bda9> SELECT * FROM emp WHERE emp_sal=50000;
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"
cqlsh:bda9> CREATE INDEX ON emp(emp_sal);
cqlsh:bda9> SELECT * FROM emp WHERE emp_sal=50000;
cqlsh:bda9> DELETE emp_sal FROM emp WHERE emp_id=3;
cqlsh:bda9> SELECT * FROM emp;
cqlsh:bda9> DELETE emp_sal FROM emp WHERE emp_id=1;
cqlsh:bda9> SELECT * FROM emp;
cqlsh:bda9> DELETE FROM emp WHERE emp_id=1;
cqlsh:bda9> SELECT * FROM emp;
cqlsh:bda9> capture off;
cqlsh:bda9> SELECT * FROM emp;

emp_id | emp_city | emp_name | emp_phone | emp_sal
-----+-----+-----+-----+-----
2 | Delhi | robin | 9848022339 | 50000

(1 rows)
cqlsh:bda9>
```

prac9 > ≡ cassandra_output.txt

```
1
2  emp_id | emp_city | emp_name | emp_phone | emp_sal
3  -----+-----+-----+-----+-----
4      1 | Hyderabad |      ram | 9848022338 | 50000
5      2 | Hyderabad |     robin | 9848022339 | 40000
6
7 (2 rows)
8
9  emp_id | emp_city | emp_name | emp_phone | emp_sal
10 -----+-----+-----+-----+-----
11      1 | Hyderabad |      ram | 9848022338 | 50000
12      2 | Hyderabad |     robin | 9848022339 | 40000
13
14 (2 rows)
15
16  emp_id | emp_city | emp_name | emp_phone | emp_sal
17 -----+-----+-----+-----+-----
18      1 | Hyderabad |      ram | 9848022338 | 50000
19      2 | Hyderabad |     robin | 9848022339 | 40000
20
21 (2 rows)
22
23  emp_id | emp_city | emp_name | emp_phone | emp_sal
24 -----+-----+-----+-----+-----
25      1 | Hyderabad |      ram | 9848022338 | 50000
26      2 |      Delhi |     robin | 9848022339 | 50000
27
28 (2 rows)
29
30  emp_name | emp_sal
31  -----+-----
32      ram |   50000
33     robin |   50000
34
35 (2 rows)
36
37  emp_id | emp_city | emp_name | emp_phone | emp_sal
38  -----+-----+-----+-----+-----
39      1 | Hyderabad |      ram | 9848022338 | 50000
40      2 |      Delhi |     robin | 9848022339 | 50000
41
42 (2 rows)
43
44  emp_id | emp_city | emp_name | emp_phone | emp_sal
45  -----+-----+-----+-----+-----
46      1 | Hyderabad |      ram | 9848022338 | 50000
47      2 |      Delhi |     robin | 9848022339 | 50000
48
```