Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Title - Multi node Cassandra cluster

Aim - To setup a multi node Cassandra cluster

Introduction & Theory -

A cluster in Cassandra is one of the shells in the whole Cassandra database. Many Cassandra Clusters combine together to form the database in Cassandra. A Cluster is basically the outermost shell or storage unit in a database. The Cassandra Cluster contains many different layers of storage units.

The following are different layers in a Cassandra Cluster:

a. Node Cluster

Node is the second layer in a cluster. This layer basically comprises of systems or computers or storage units. Each cluster may contain many nodes or systems. These systems or nodes are connected together.

They collectively share data through the replication in Cassandra and independently as well. The replication factor in the nodes allows the user to have a redundancy for the data stored.

b. Keyspace

The keyspace is the next layer of the storage. In a node, there are many keyspaces. These keyspaces are basically the outermost storage unit in a system. They contain the main data. The data distributed according to their properties or areas.

c. Column Families

The next layer is the column families. The keyspace is further divided into column families. These column families have different areas or headings under which the data is distributed. In a keyspace, these column families are categorized into different headings or titles.

These titles further contain different layers of storage units. These column families can also be characterized by tables. The column families differ from the tables through their APIs.

d. Rows

The next layer in the database is of the Rows as according to column families. The Rows are basically the classification under which the column family is divided. These classifications, in turn, create specific distribution criteria for the entries.

e. Column

This is the innermost layer in a database. The column basically is divided into different titles or headings. These headings contain the main data regarding the specific entry.

Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Procedure -

Setup a multi-node Cassandra Cluster (Take 3 machines in lab or

on a single machine). Give your PRN as cluster name. Follow the

steps given in below link

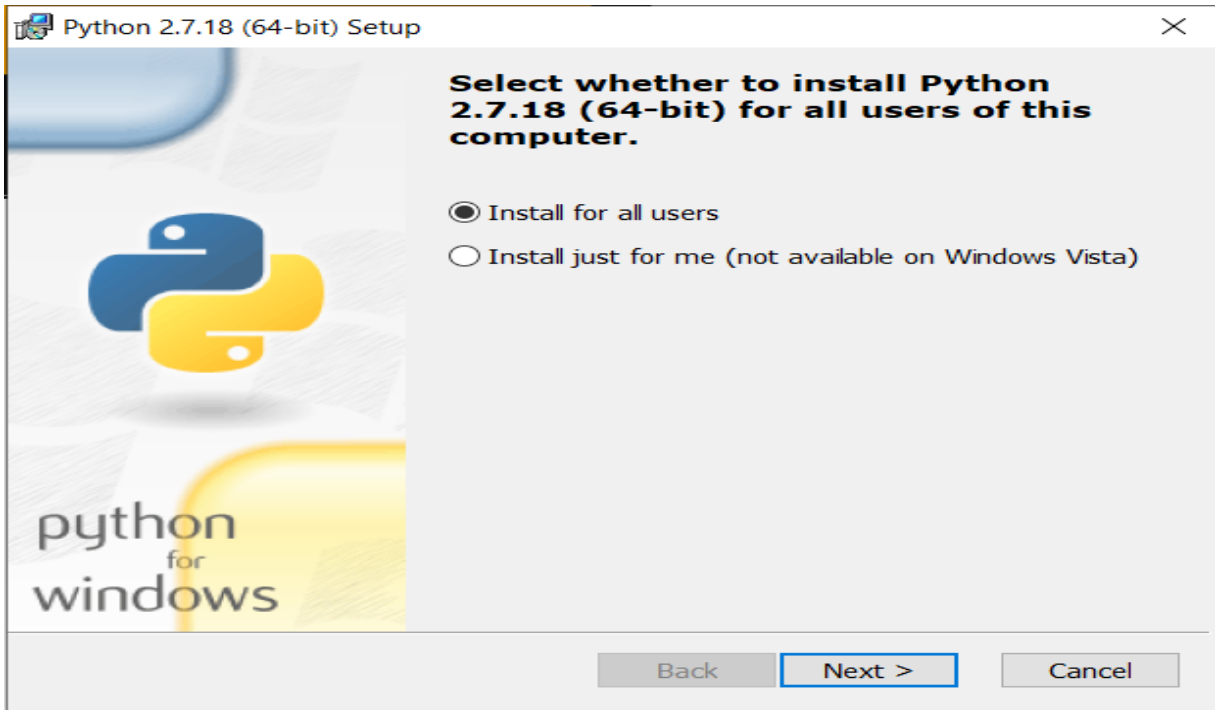
https://extendit.us/articles/steps-configure-multiple-nodes-

cassandra-single-windows-machine

1. Setup a multi-node Cassandra Cluster (Take 3 machines in lab or on single machine). Give your PRN as cluster name. Follow the steps given in below link https://extendit.us/articles/steps-configure-multiple-nodes-cassandra-single-windows-machine

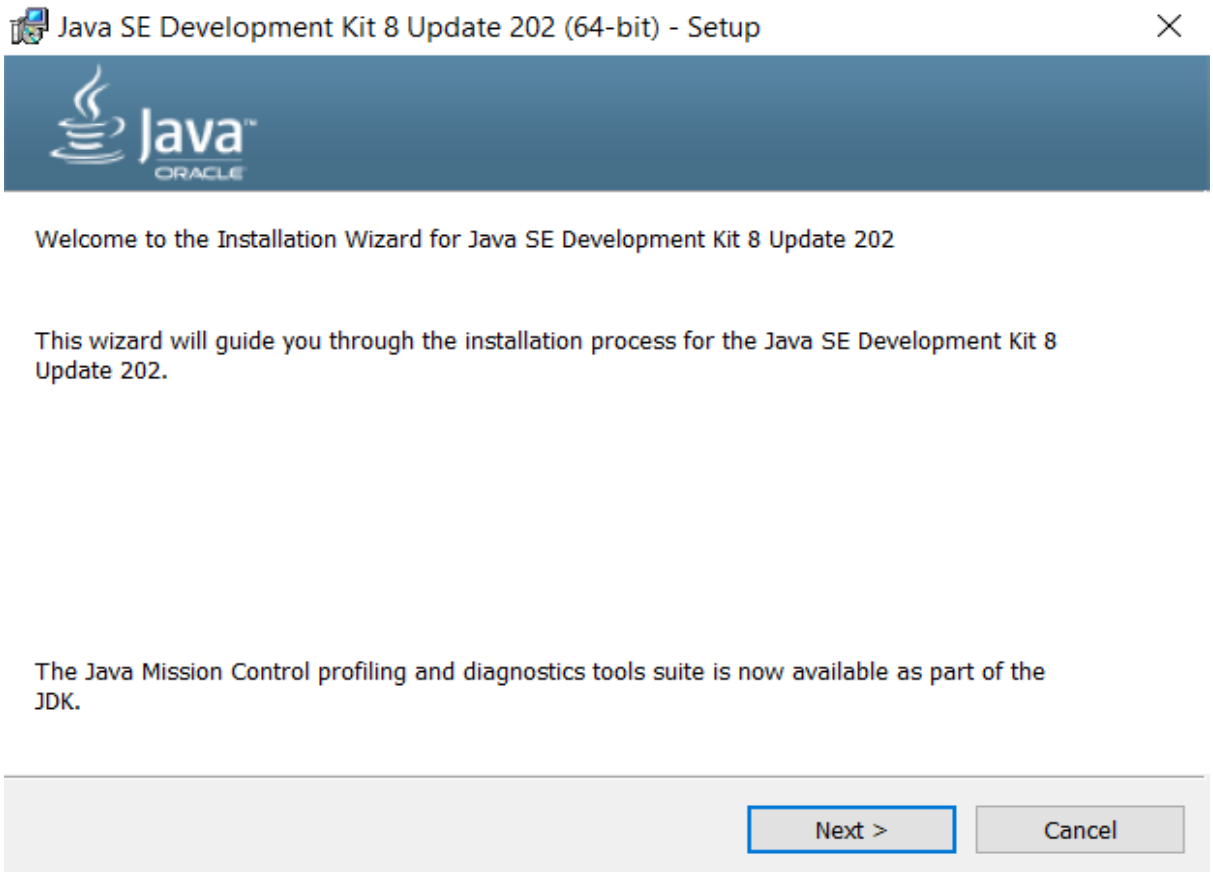
Ans :

Installation of python 2.7.18 on node1

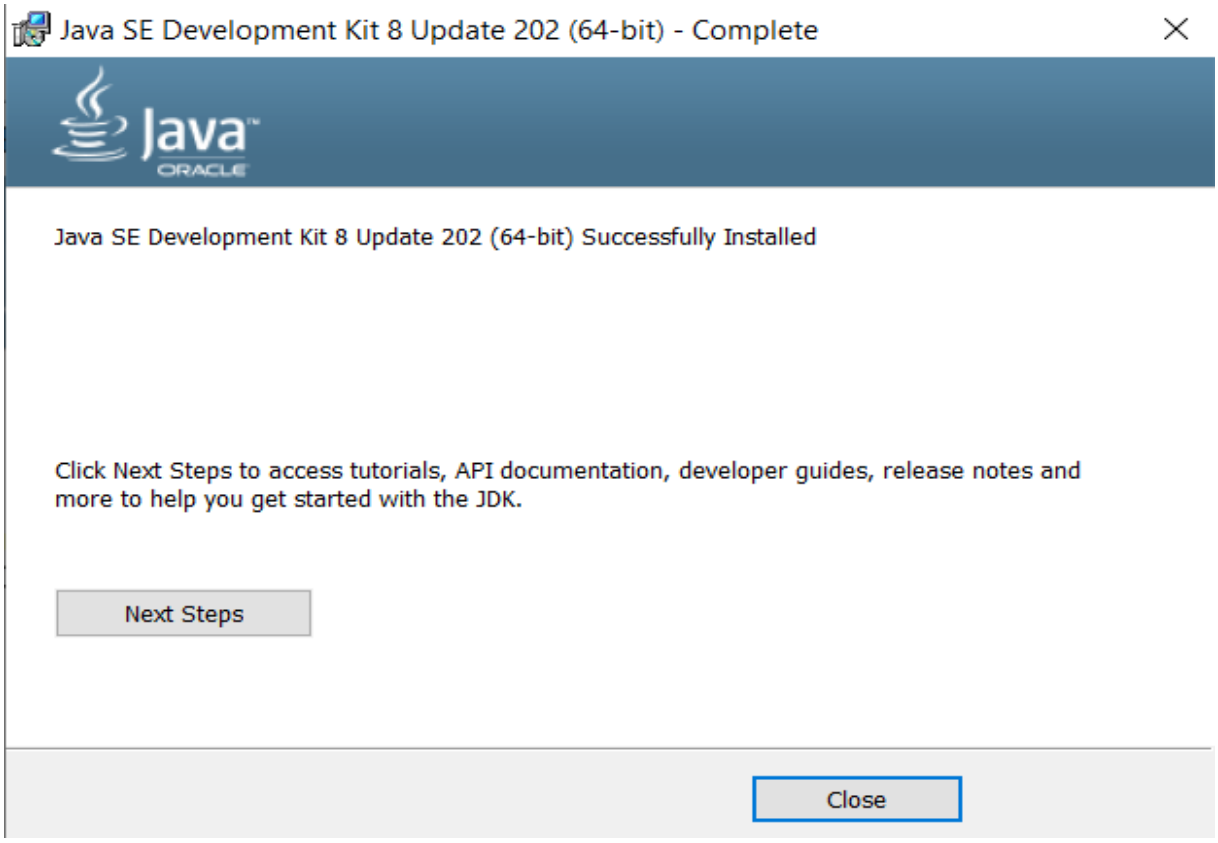


Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Installation of JDK 8 on node1



Name:Tushar Patil
PRN:2020BTECS00075
ADS10



Set environment variable for python, JDK on node1

Same procedure is repeated for node2 and node3

Hostname: DBE2743 node1

Changed – seeds, listen_address, rpc_address in cassandra.yaml file on node1.

```
seed_provider:
    # Addresses of hosts that are deemed contact points.
    # Cassandra nodes use this list of hosts to find each other and learn
    # the topology of the ring.  You must change this if you are running
    # multiple nodes!
    - class_name: org.apache.cassandra.locator.SimpleSeedProvider
      parameters:
          # seeds is actually a comma-delimited list of addresses.
          # Ex: "<ip1>,<ip2>,<ip3>"
          - seeds: "10.4.2.109, 10.4.2.108, 10.4.2.111"
```

```
listen_address: 10.4.2.109
```

```
# For security reasons, you should not expose this port to the internet.  Firewall it if needed.
rpc_address: 10.4.2.109
```

Changed hostname to IP address in cqlshrc.sample file on node1.

Name:Tushar Patil
PRN:2020BTECS00075
ADS10

```
[connection]

;; The host to connect to
hostname = 10.4.2.109

;; The port to connect to (9042 is the native protocol default)
port = 9042

;; Always connect using SSL - false by default
; ssl = true

;; A timeout in seconds for opening new connections
; timeout = 10

;; A timeout in seconds for executing queries
; request_timeout = 10
```

Changed DEFAULT_HOST to IP address in cqlsh.py file on node1.

```
DEFAULT_HOST = '10.4.2.109'
DEFAULT_PORT = 9042
DEFAULT_SSL = False
DEFAULT_CONNECT_TIMEOUT_SECONDS = 5
DEFAULT_REQUEST_TIMEOUT_SECONDS = 10
```

Hostname: DBE2744 node2

Changed – seeds, listen_address, rpc_address in cassandra.yaml file on node2.

```
# any class that implements the SeedProvider interface and has a
# constructor that takes a Map<String, String> of parameters will do.
seed_provider:
    # Addresses of hosts that are deemed contact points.
    # Cassandra nodes use this list of hosts to find each other and learn
    # the topology of the ring.  You must change this if you are running
    # multiple nodes!
    - class_name: org.apache.cassandra.locator.SimpleSeedProvider
      parameters:
          # seeds is actually a comma-delimited list of addresses.
          # Ex: "<ip1>,<ip2>,<ip3>"
          - seeds: "10.4.2.108,10.4.2.109,10.4.2.111"
```

```
listen_address: 10.4.2.108
```

Name:Tushar Patil
PRN:2020BTECS00075
ADS10

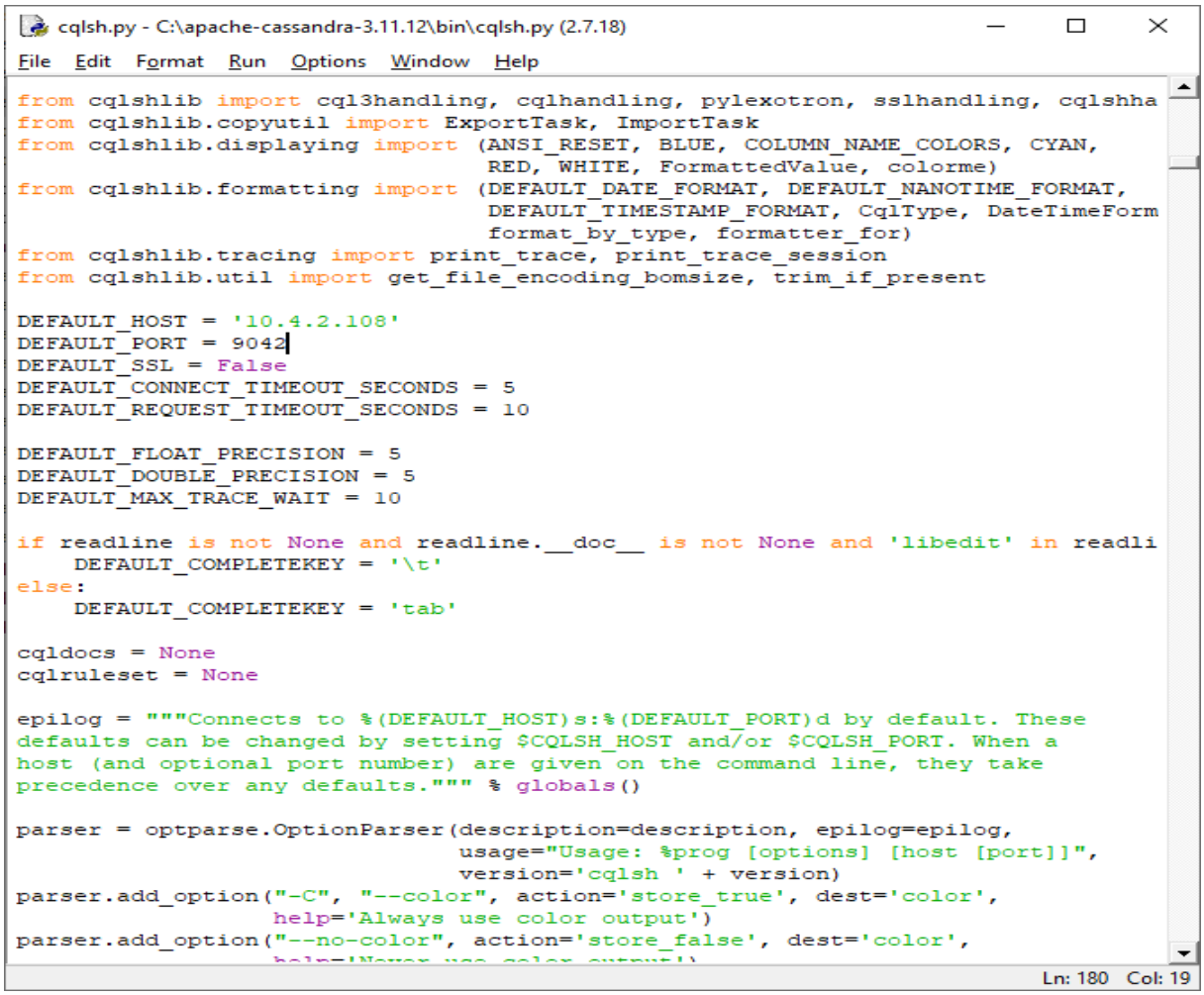
```
# Whether to start the thrift rpc server.
start_rpc: false

# The address or interface to bind the Thrift RPC service and native transport
# server to.
#
# Set rpc_address OR rpc_interface, not both.
#
# Leaving rpc_address blank has the same effect as on listen_address
# (i.e. it will be based on the configured hostname of the node).
#
# Note that unlike listen_address, you can specify 0.0.0.0, but you must also
# set broadcast_rpc_address to a value other than 0.0.0.0.
#
# For security reasons, you should not expose this port to the internet.  Firewall it if needed.
rpc_address: 10.4.2.108

# Set rpc_address OR rpc_interface, not both. Interfaces must correspond
# to a single address, IP aliasing is not supported.
# rpc_interface: eth1

# If you choose to specify the interface by name and the interface has an ipv4 and an ipv6 address
# you can specify which should be chosen using rpc_interface_prefer_ipv6. If false the first ipv4
# address will be used. If true the first ipv6 address will be used. Defaults to false preferring
# ipv4. If there is only one address it will be selected regardless of ipv4/ipv6.
# rpc_interface_prefer_ipv6: false

# port for Thrift to listen for clients on
```

Changed hostname to IP address in cqlshrc.sample file on node2.

```
[connection]

;; The host to connect to
hostname = 10.4.2.108

;; The port to connect to (9042 is the native protocol default)
port = 9042

;; Always connect using SSL - false by default
; ssl = true

;; A timeout in seconds for opening new connections
; timeout = 10

;; A timeout in seconds for executing queries
; request_timeout = 10
```

Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Changed DEFAULT_HOST to IP address in cqlsh.py file on node2.



Hostname: DBE2745 node3

Changed – seeds, listen_address, rpc_address in cassandra.yaml file on node3.

```
seed_provider:
    # Addresses of hosts that are deemed contact points.
    # Cassandra nodes use this list of hosts to find each other and learn
    # the topology of the ring.  You must change this if you are running
    # multiple nodes!
    - class_name: org.apache.cassandra.locator.SimpleSeedProvider
      parameters:
          # seeds is actually a comma-delimited list of addresses.
          # Ex: "<ip1>,<ip2>,<ip3>"
          - seeds: "10.4.2.111, 10.4.2.109, 10.4.2.108"
```

```
# Setting listen_address to 0.0.0.0 is always wrong.
#
listen_address: 10.4.2.111
```

```
# For security reasons, you should not expose this port to the internet.  Firewall it if needed.
rpc_address: 10.4.2.111
```

Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Changed hostname to IP address in cqlshrc.sample file on node3.

```
[connection]

;; The host to connect to
hostname = 10.4.2.111

;; The port to connect to (9042 is the native protocol default)
port = 9042

;; Always connect using SSL - false by default
; ssl = true

;; A timeout in seconds for opening new connections
; timeout = 10

;; A timeout in seconds for executing queries
; request_timeout = 10
```

Changed DEFAULT_HOST to IP address in cqlsh.py file on node3.

Name:Tushar Patil
PRN:2020BTECS00075
ADS10

cqlsh.py - C:\apache-cassandra-3.11.12-bin\apache-cassandra-3.11.12\bin\cqlsh.py (2.7.18)

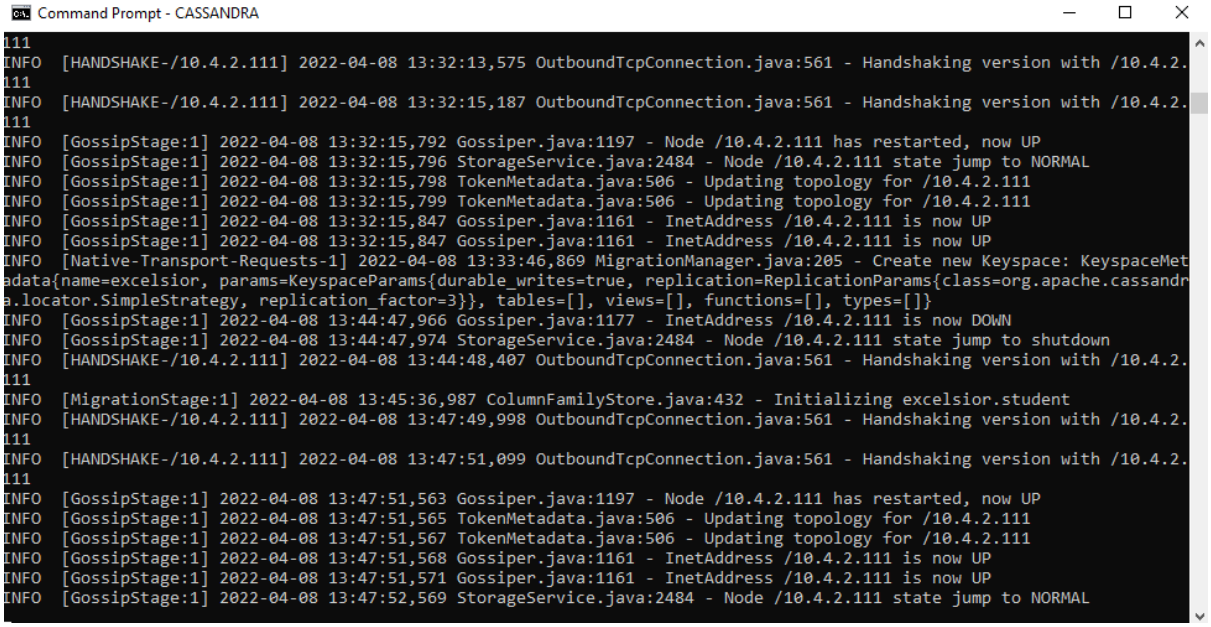
File Edit Format Run Options Window Help

```
from cassandra.util import datetime_from_timestamp

# cqlsh should run correctly when run out of a Cassandra source tree,
# out of an unpacked Cassandra tarball, and after a proper package install.
cqlshlibdir = os.path.join(CASSANDRA_PATH, 'pylib')
if os.path.isdir(cqlshlibdir):
    sys.path.insert(0, cqlshlibdir)

from cqlshlib import cql3handling, cqlhandling, pylexotron, sslhandling, cqlshha
from cqlshlib.copyutil import ExportTask, ImportTask
from cqlshlib.displaying import (ANSI_RESET, BLUE, COLUMN_NAME_COLORS, CYAN,
                                 RED, WHITE, FormattedValue, colorme)
from cqlshlib.formatting import (DEFAULT_DATE_FORMAT, DEFAULT_NANOTIME_FORMAT,
                                 DEFAULT_TIMESTAMP_FORMAT, CqlType, DateTimeForm
                                 format_by_type, formatter_for)
from cqlshlib.tracing import print_trace, print_trace_session
from cqlshlib.util import get_file_encoding_bomsize, trim_if_present

DEFAULT_HOST = '10.4.2.111'
DEFAULT_PORT = 9042
DEFAULT_SSL = False
DEFAULT_CONNECT_TIMEOUT_SECONDS = 5
DEFAULT_REQUEST_TIMEOUT_SECONDS = 10

DEFAULT_FLOAT_PRECISION = 5
DEFAULT_DOUBLE_PRECISION = 5
DEFAULT_MAX_TRACE_WAIT = 10

if readline is not None and readline.__doc__ is not None and 'libedit' in readli
    DEFAULT_COMPLETEKEY = '\t'
else:
    DEFAULT_COMPLETEKEY = 'tab'

cqldocs = None
cqlruleset = None

epilog = """Connects to %(DEFAULT_HOST)s:%(DEFAULT_PORT)d by default. These
defaults can be changed by setting $CQLSH_HOST and/or $CQLSH_PORT. When a
host (and optional port number) are given on the command line, they take
precedence over any defaults.""" % globals()
```

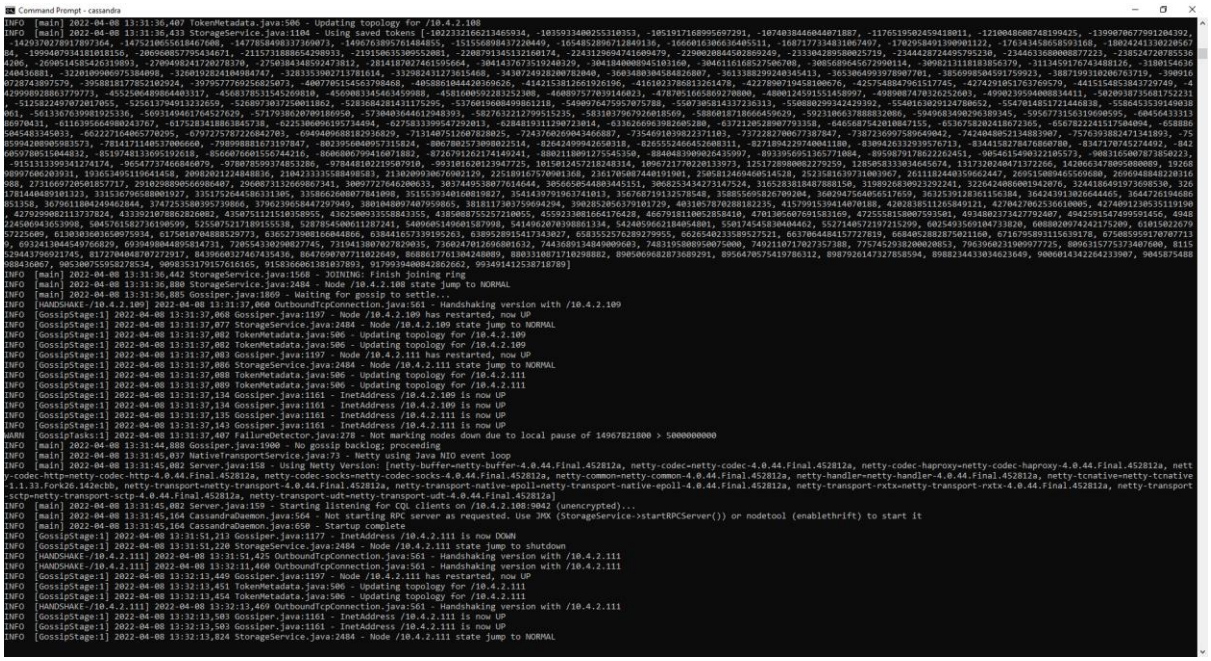
Ln: 195 Col: 0

Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Started cassandra server on node1

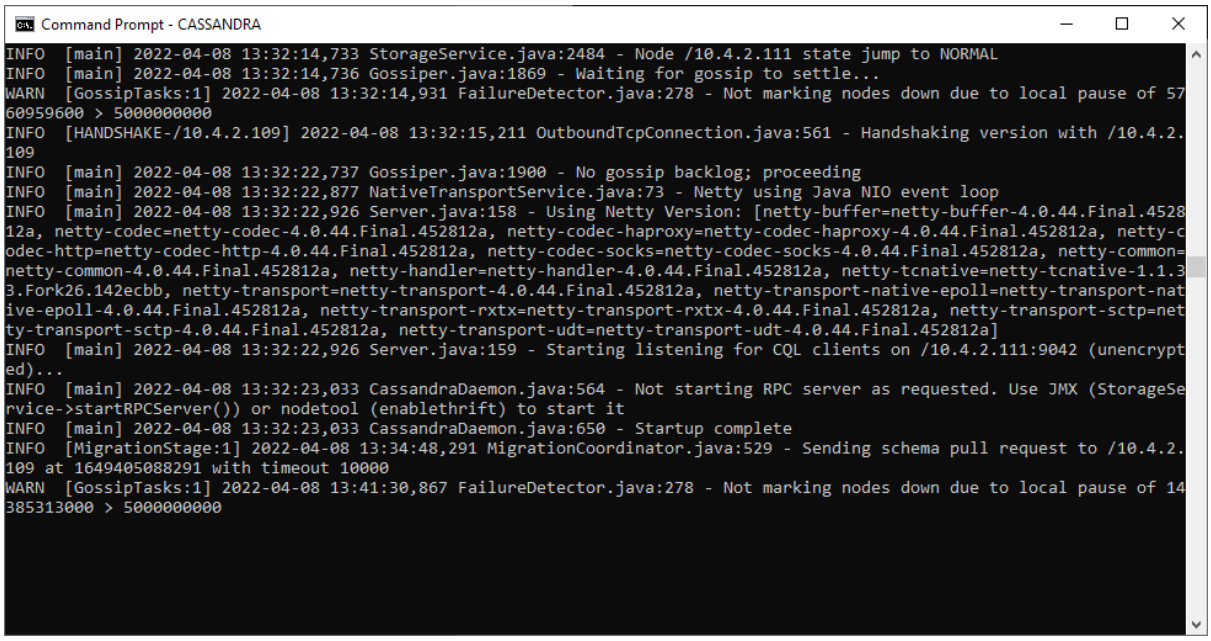


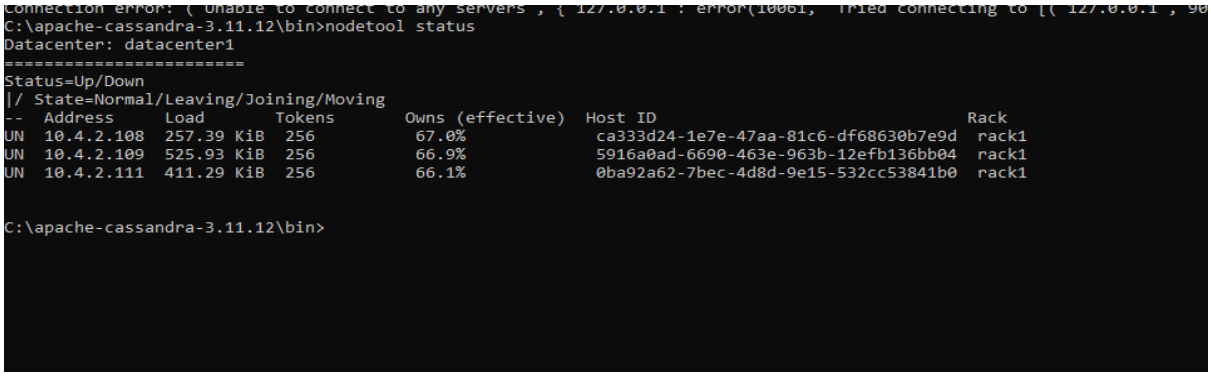
Started cassandra server on node2



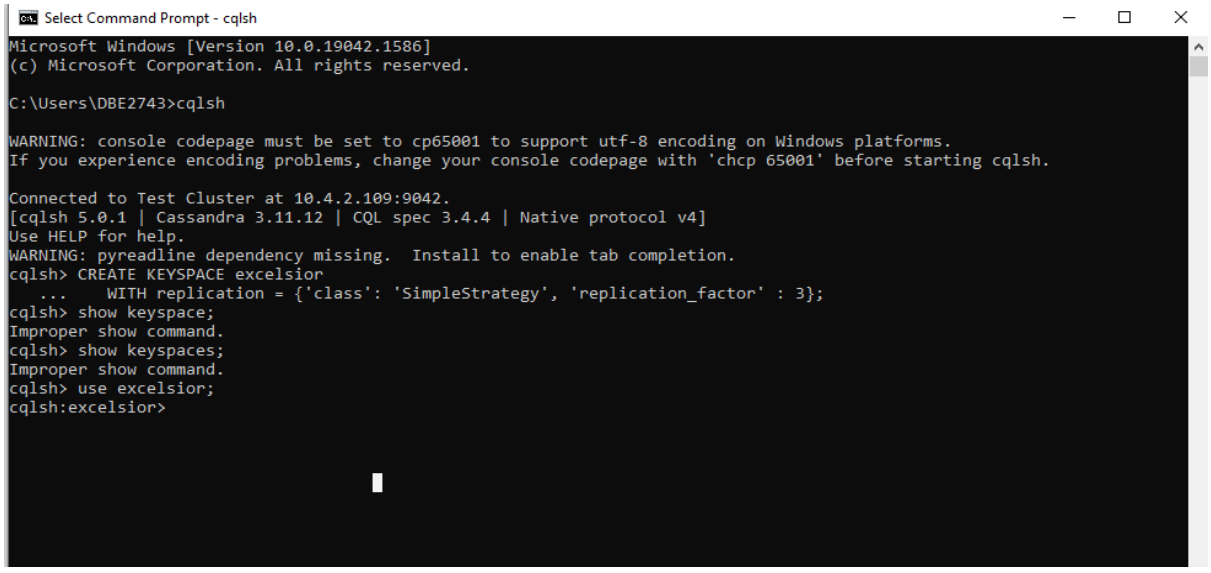
Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Started cassandra server on node3



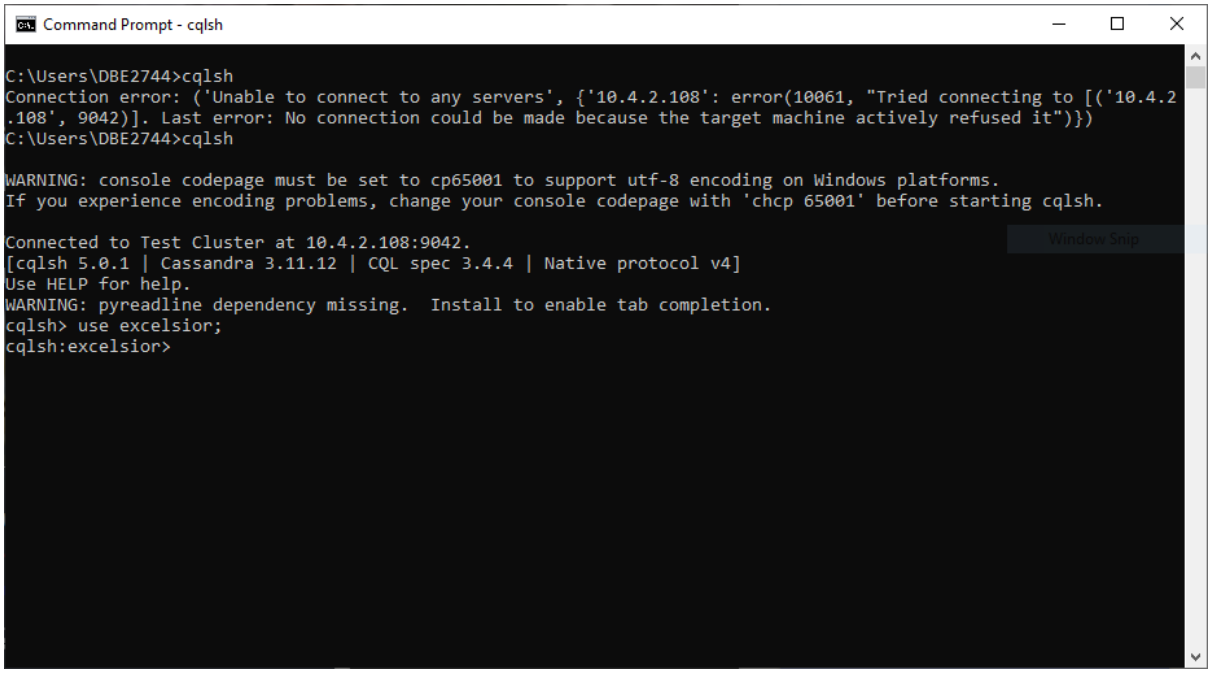


Created keyspace excelsior on node1:

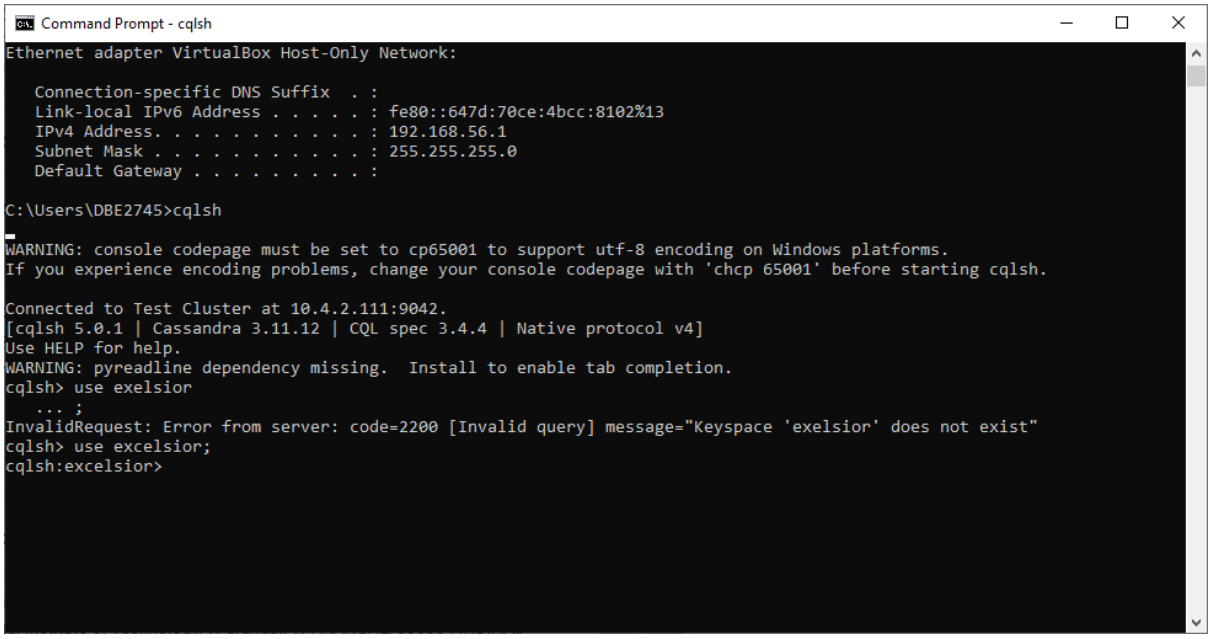


Accessing keyspace excelsior from node2

Name:Tushar Patil
PRN:2020BTECS00075
ADS10

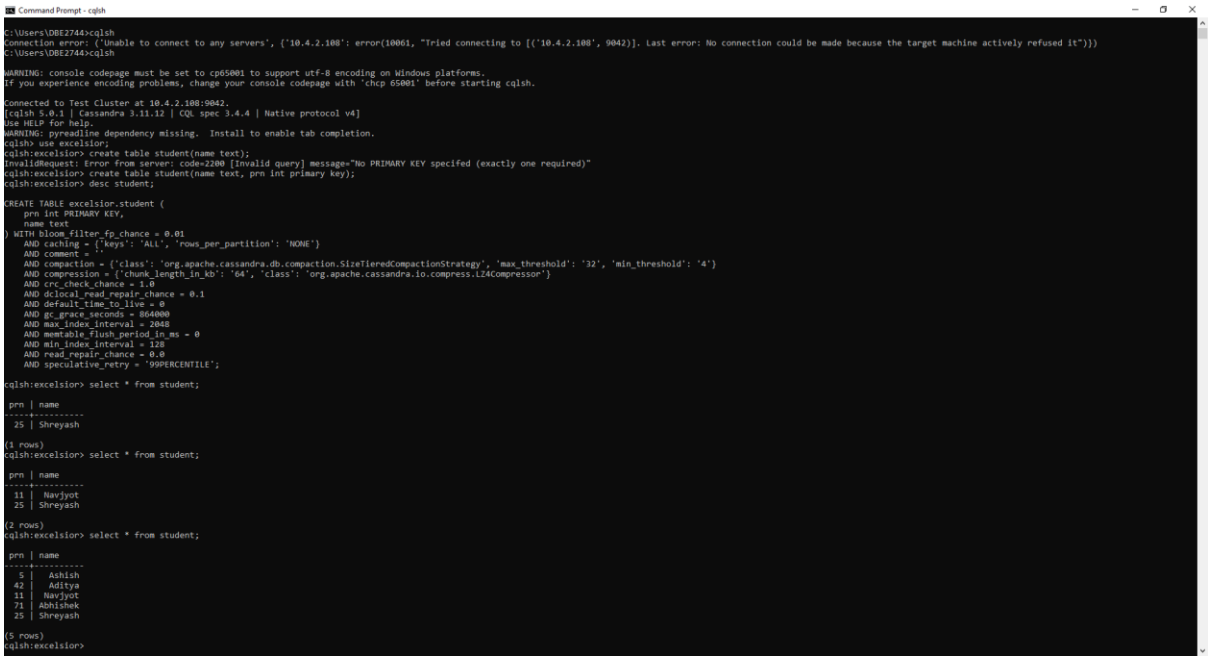


Accessing keyspace excelsior from node3

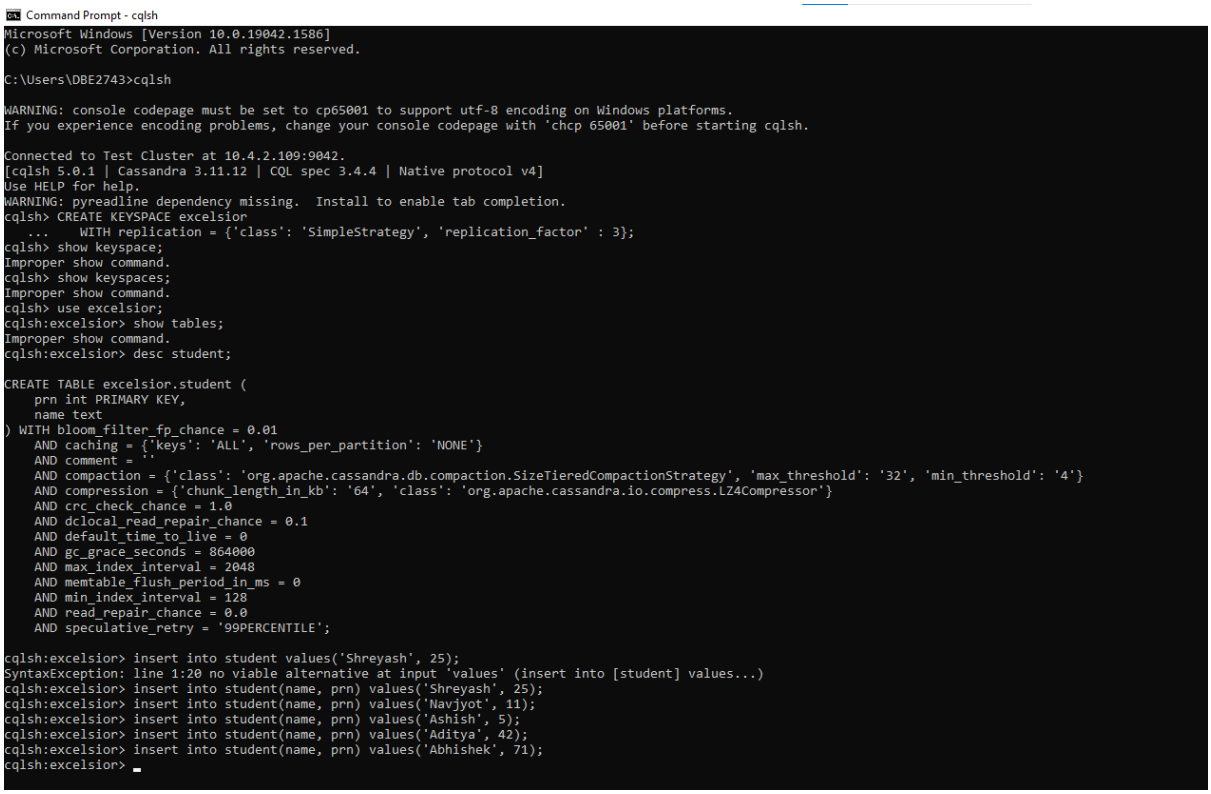


Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Created table from node1 and inserted records from node2:

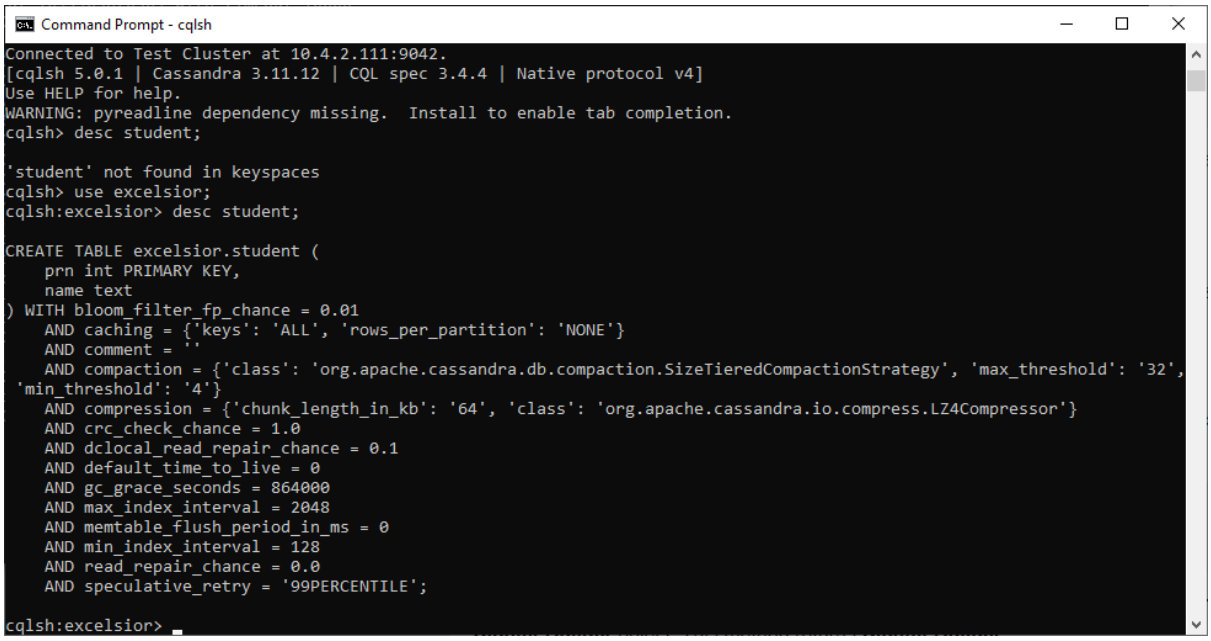


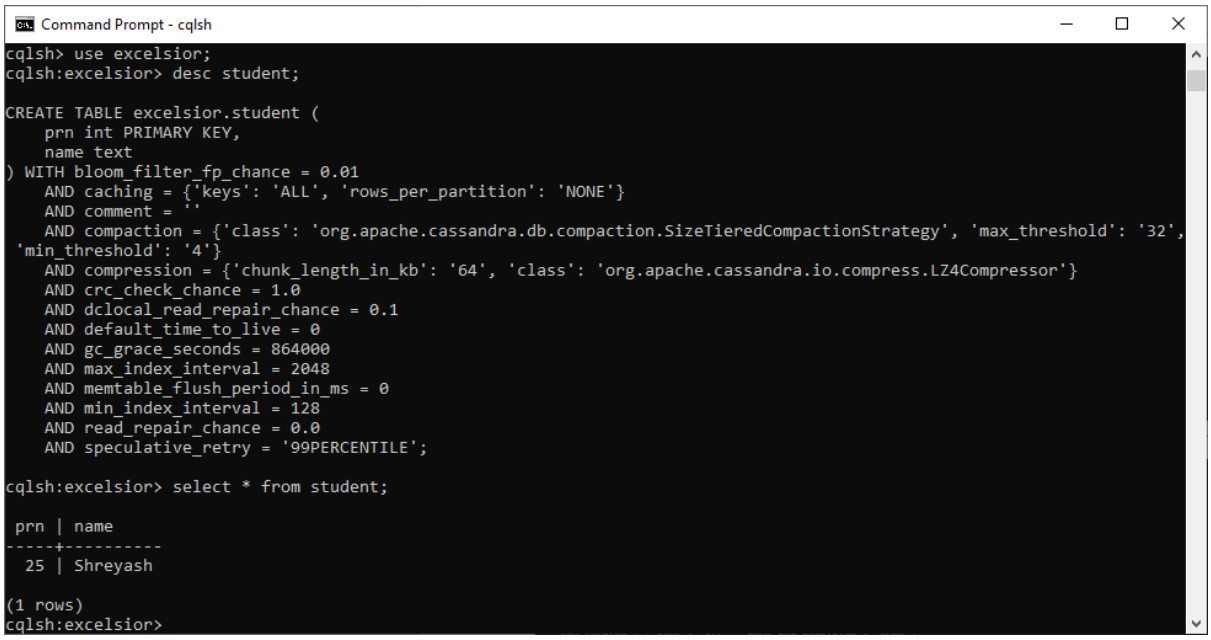
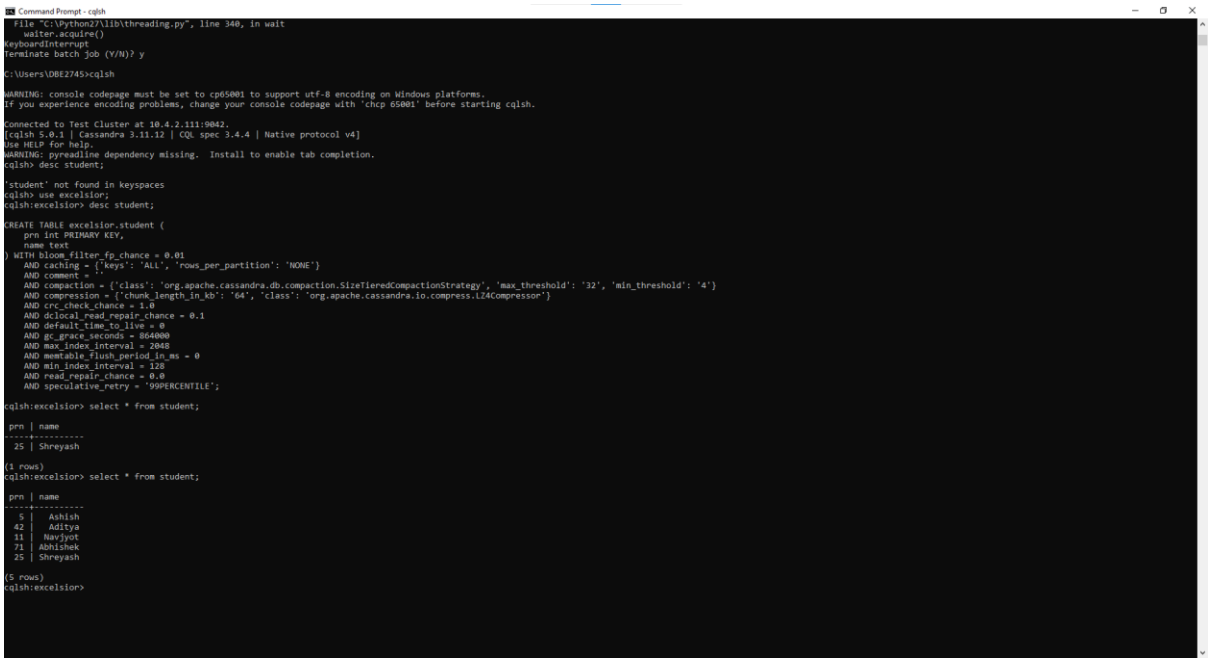
Record insertion from node2:



Name:Tushar Patil
PRN:2020BTECS00075
ADS10

Checking table and inserted records from node3





Name:Tushar Patil
PRN:2020BTECS00075
ADS10