ASSIGNMENT - 12

Exercise 1:

The paper presents Apache Cassandra's first query-driven big data modeling methodology, as well as important data modeling principles, mapping rules, and mapping patterns to guide logical data modeling, visual diagrams for Cassandra logical and physical data models, and a data modeling tool that automates the entire data modeling process. Because of its linear scalability, ease of multi-datacenter deployment, and low downtime, Cassandra is employed by many web-scale organizations for online transaction processing.

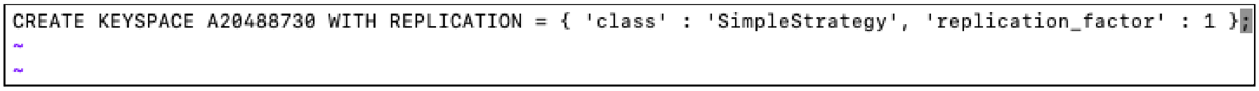
Apache Cassandra is a leading transactional, scalable, and highly-available distributed database. It is known to manage some of the world's largest datasets on clusters with many thousands of nodes deployed across multiple data centers. Cassandra data management use cases include product catalogs and playlists, sensor data and Internet of Things, messaging and social networking, recommendation, personalization, fraud detection, and numerous other applications that deal with time series data.

It covers query-driven methodology for Apache Cassandra, as well as important Data Modelling rules, patterns, and guidelines, as well as visual representations for Cassandra logical and physical data models and a data modelling automation tool.

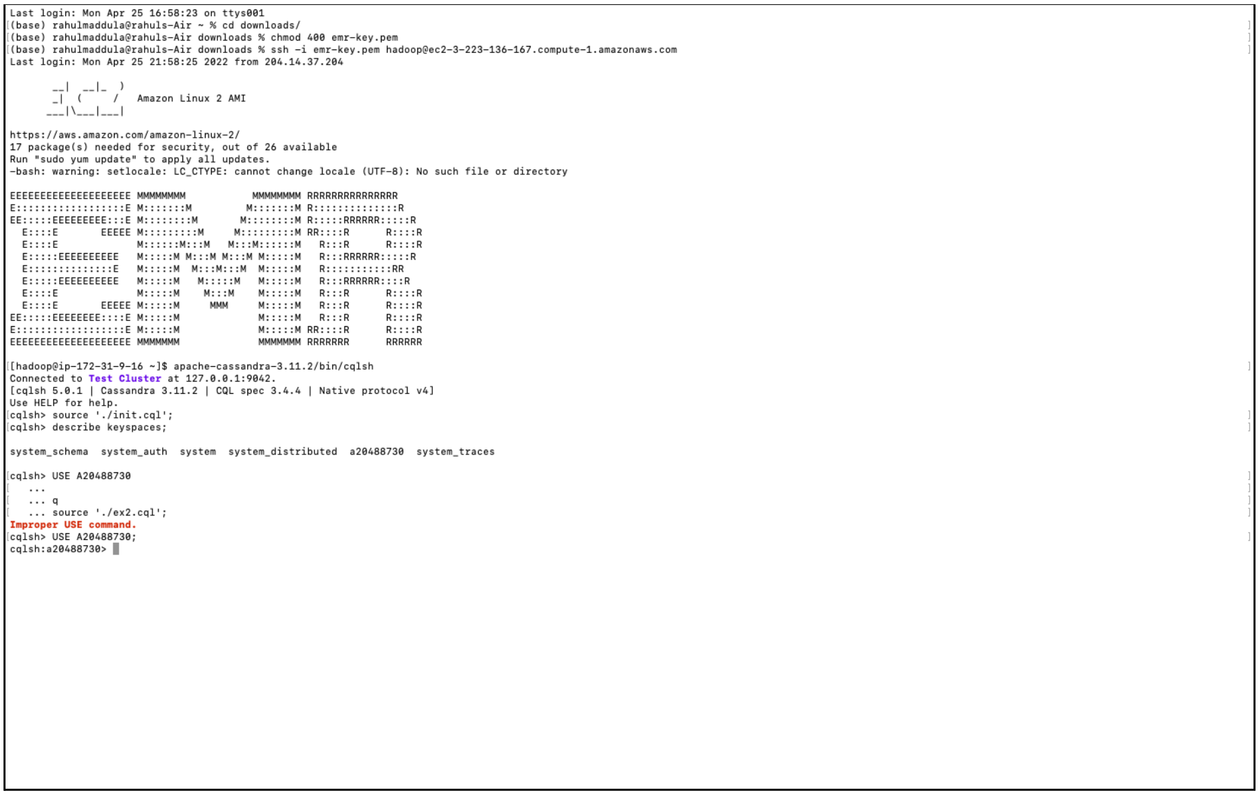
The data modeling approaches proposed in this study were vastly different from traditional relational data modeling strategies. It introduced Chebotko Diagrams, a new visualization tool for capturing complex logical and physical data models, and addressed the necessity of physical data modeling. This document for Cassandra defines and establishes data nesting, data duplication, mapping rules, and mapping patterns. Chebotko diagrams depict a database schema design as a mix of individual table schemas and query-driven application workflow transitions. Chebotko Diagrams provide improved overall readability, superior intelligibility for intricate data models, and better expressivity for both table schemas and their associated application queries when compared to standard CQL schema design scripts. It is described how to go from technology-independent conceptual models to Cassandra-specific logical data models.

Finally, it showed about the one of the good data modeling tool called KDM, a sophisticated data modeling tool that automates some of the most difficult, error-prone, and time-consuming data modeling tasks, including conceptual-to-logical mapping, logical-to-physical mapping, and CQL creation. By utilizing mapping patterns and unique algorithms, the application automates the most sophisticated, error-prone, and time-consuming data modeling processes. Advanced users can take advantage of KDM's advanced features, such as automatic schema construction in the presence of type hierarchies, n-array relationship types, explicit roles, and alternate keys.

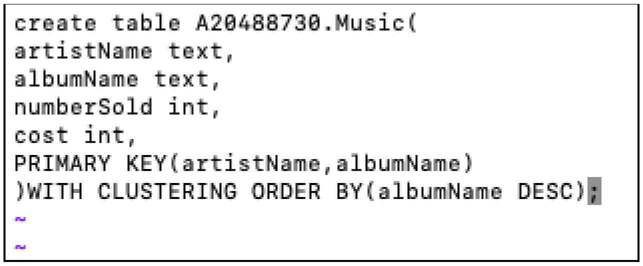
Exercise 2:



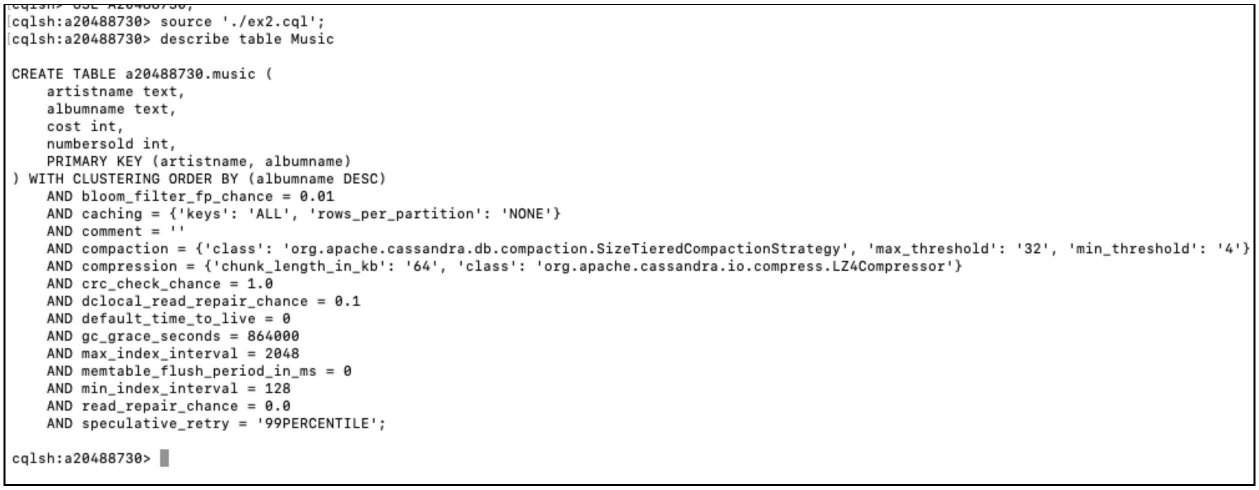
```
CREATE KEYSPACE A20488730 WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
~
~
```



```
Last login: Mon Apr 25 16:58:23 on ttys001
(base) rahulmaddula@rahuls-Air ~ % cd downloads/
(base) rahulmaddula@rahuls-Air downloads % chmod 400 emr-key.pem
(base) rahulmaddula@rahuls-Air downloads % ssh -i emr-key.pem hadoop@ec2-3-223-136-167.compute-1.amazonaws.com
Last login: Mon Apr 25 21:58:25 2022 from 204.14.37.204

       __|  __|_  )
       _|  (     /   Amazon Linux 2 AMI
      ___|\___|___|

https://aws.amazon.com/amazon-linux-2/
17 package(s) needed for security, out of 26 available
Run "sudo yum update" to apply all updates.
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE:::::EEEEEEEEE:::E M::::::::M       M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M     M:::::::::M RR::::R      R::::R
  E::::E             M::::::M:::M   M:::M::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE   M:::::M M:::M M:::M M:::::M   R:::RRRRRR:::::R
  E::::::::::::::E   M:::::M  M:::M:::M  M:::::M   R:::::::::::RR
  E:::::EEEEEEEEEE   M:::::M   M:::::M   M:::::M   R:::RRRRRR::::R
  E::::E             M:::::M    M:::M    M:::::M   R:::R      R::::R
  E::::E       EEEEE M:::::M     MMM     M:::::M   R:::R      R::::R
EE:::::EEEEEEEE::::E M:::::M             M:::::M   R:::R      R::::R
E::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-9-16 ~]$ apache-cassandra-3.11.2/bin/cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.2 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> source './init.cql';
cqlsh> describe keyspaces;

system_schema  system_auth  system  system_distributed  a20488730  system_traces

cqlsh> USE A20488730
   ...
   ... q
   ... source './ex2.cql';
Improper USE command.
cqlsh> USE A20488730;
cqlsh:a20488730> 
```



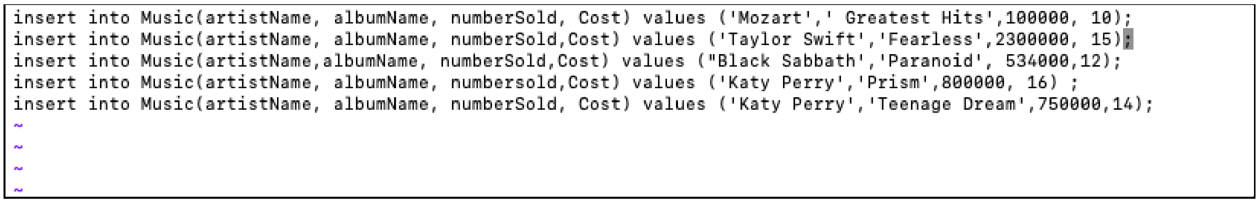
```
create table A20488730.Music(
artistName text,
albumName text,
numberSold int,
cost int,
PRIMARY KEY(artistName,albumName)
)WITH CLUSTERING ORDER BY(albumName DESC);
~
~
```



```
cqlsh> USE A20488730;
[cqlsh:a20488730> source './ex2.cql';
[cqlsh:a20488730> describe table Music

CREATE TABLE a20488730.music (
    artistname text,
    albumname text,
    cost int,
    numbersold int,
    PRIMARY KEY (artistname, albumname)
) WITH CLUSTERING ORDER BY (albumname DESC)
    AND 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:a20488730> 
```

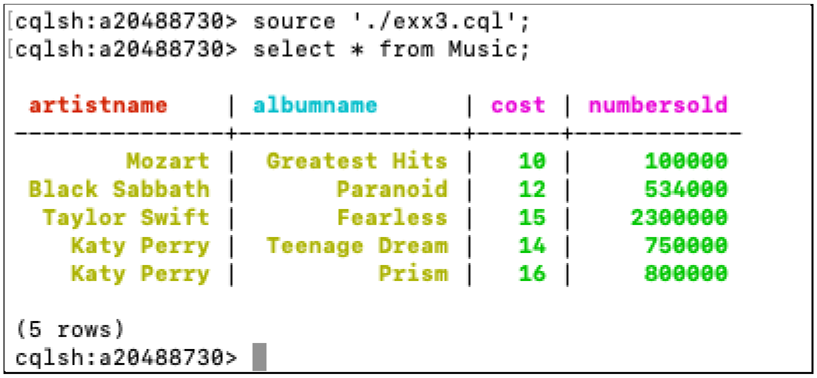
Exercise 3:

a)



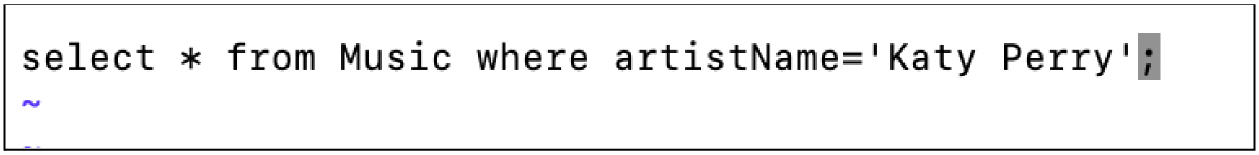
```
insert into Music(artistName, albumName, numberSold, Cost) values ('Mozart',' Greatest Hits',100000, 10);
insert into Music(artistName, albumName, numberSold,Cost) values ('Taylor Swift','Fearless',2300000, 15);
insert into Music(artistName,albumName, numberSold,Cost) values ("Black Sabbath','Paranoid', 534000,12);
insert into Music(artistName, albumName, numbersold,Cost) values ('Katy Perry','Prism',800000, 16) ;
insert into Music(artistName, albumName, numberSold, Cost) values ('Katy Perry','Teenage Dream',750000,14);
~
~
~
~
```

b)

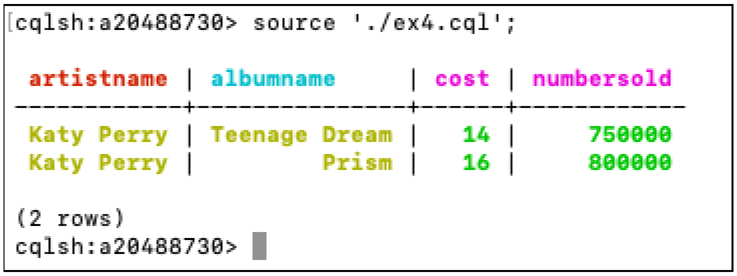


```
[cqlsh:a20488730> source './exx3.cql';
[cqlsh:a20488730> select * from Music;

 artistname    | albumname       | cost | numbersold
---------------+-----------------+------+------------
        Mozart |  Greatest Hits  |   10 |     100000
 Black Sabbath |        Paranoid |   12 |     534000
  Taylor Swift |        Fearless |   15 |    2300000
    Katy Perry |   Teenage Dream |   14 |     750000
    Katy Perry |           Prism |   16 |     800000

(5 rows)
cqlsh:a20488730> 
```

Exercise 4:

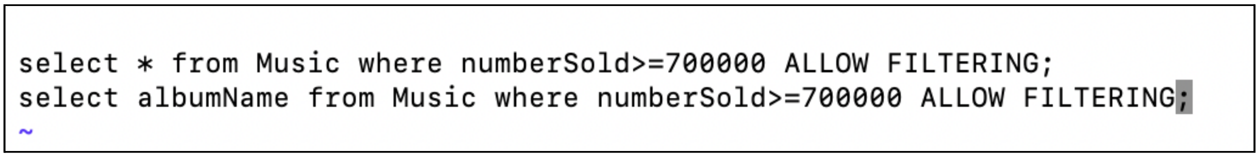


```
select * from Music where artistName='Katy Perry';
~
```

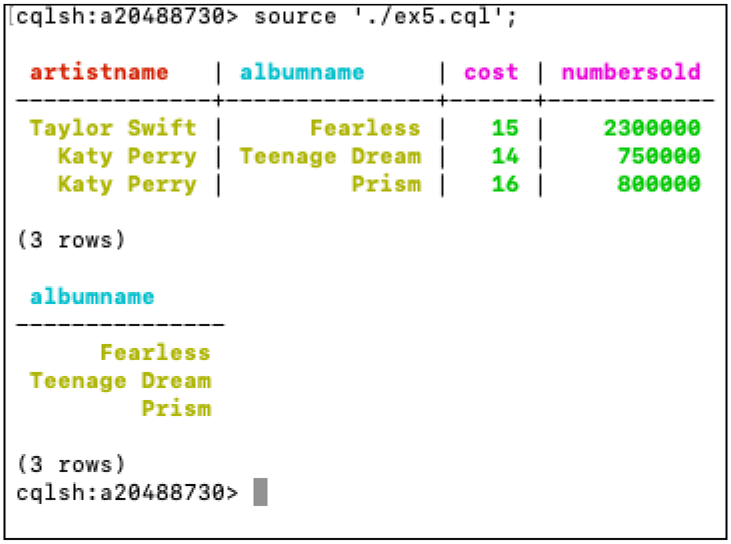


```
[cqlsh:a20488730> source './ex4.cql';

 artistname | albumname     | cost | numbersold
------------+---------------+------+------------
 Katy Perry | Teenage Dream |   14 |     750000
 Katy Perry |         Prism |   16 |     800000

(2 rows)
cqlsh:a20488730> 
```

Exercise 5:



```
select * from Music where numberSold>=700000 ALLOW FILTERING;
select albumName from Music where numberSold>=700000 ALLOW FILTERING;
~
```



```
[cqlsh:a20488730> source './ex5.cql';

 artistname   | albumname     | cost | numbersold
--------------+---------------+------+------------
 Taylor Swift |      Fearless |   15 |    2300000
   Katy Perry | Teenage Dream |   14 |     750000
   Katy Perry |         Prism |   16 |     800000

(3 rows)

 albumname
---------------
      Fearless
 Teenage Dream
         Prism

(3 rows)
cqlsh:a20488730> 
```