

HBASE

```
ssh -i /c/Users/Ankan\ Mazumdar/Downloads/emr_key_pair.pem
hadoop@ec2-18-190-0.us-east-2.compute.amazonaws.com
```

hbase shell

```

ankan Mazumdar@DESKTOP-CMULEBA MINGW~
$ ssh -i /C:/Users/Ankan/Mazumdar/Downloads/emr_key_pair.pem hadoop@ec2-18-221-191-0.us-east-2.compute.amazonaws.com
The authenticity of host 'ec2-18-221-191-0.us-east-2.compute.amazonaws.com (18.221.191.0)' can't be established.
ED25519 key fingerprint is SHA256:etzx0AwLk03rHcy2Un9Q3yn5m4aJowTMqNHL5FnN1A4.
This key is not known by any other names.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-18-221-191-0.us-east-2.compute.amazonaws.com' (ED25519) to the list of known hosts.

      _-|  _-|  )
      _|  _-|  /
      _|  _-|  |
      _|  _-|  |

Amazon Linux 2 AMI

https://aws.amazon.com/amazon-linux-2/
36 package(s) needed for security, out of 60 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEE MMMMMMMM          MMMMMMMM RRRRRRRRRRRRRRRR
E:::EEEEEEEEEEEEEE M:::MM          M:::MM R:::R R:::R R:::R
EE:::EEEEEEEEEEEEEE M:::MM          M:::MM R:::RRRRRR:::R
E:::E EEEEE M:::MM          M:::MM RR:::R R:::R
E:::E M:::MM M:::MM M:::MM M:::MM R:::R R:::R
E:::EEEEEEEEEEEEEE M:::MM M:::MM M:::MM M:::MM R:::RRRRRR:::R
E:::EEEEEEEEEEEEEE M:::MM M:::MM M:::MM R:::RRRRRR:::R
E:::EEEEEEEEEEEEEE M:::MM M:::MM M:::MM R:::RRRRRR:::R
E:::E M:::MM M:::MM M:::MM R:::R R:::R
E:::E EEEEE M:::MM MMM M:::MM R:::R R:::R
EE:::EEEEEEEEEEEEEE M:::MM M:::MM R:::R R:::R
E:::EEEEEEEEEEEEEE M:::MM M:::MM RR:::R R:::R
EEEEEEEEEEEEEEEEEE MMMMMMMM          MMMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-40-61 ~]$ hbase shell
-bash: hbase shell: command not found
[hadoop@ip-172-31-40-61 ~]$ hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-reload4j-1.7.33.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.4.17-amzn-1, rUnknown, Tue Aug 22 19:10:50 UTC 2023
Took 0.0521 seconds

```

Exercise 1

```
create 'csp554Tbl', {NAME => 'cf1'}, {NAME => 'cf2'}
describe 'csp554Tbl'
```

```
hbase:001:0> create 'csp554Tbl', {NAME => 'cf1'}, {NAME => 'cf2'}
Created table csp554Tbl
Took 6.5518 seconds
=> Hbase:Table - csp554Tbl
hbase:002:0> describe 'csp554Tbl'
Table csp554Tbl is ENABLED
csp554Tbl
COLUMN FAMILIES DESCRIPTION
{NAME => 'cf1', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MI
_Versions => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
{NAME => 'cf2', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MI
_Versions => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
2 row(s)
Quota is disabled
Took 0.7283 seconds
```

Exercise 2

```
put 'csp554Tbl', 'Row1', 'cf1:Name', 'Sam'
put 'csp554Tbl', 'Row2', 'cf1:Name', 'Ahmed'
put 'csp554Tbl', 'Row1', 'cf2:Job', 'Pilot'
put 'csp554Tbl', 'Row2', 'cf2:Job', 'Doctor'
```

```
put 'csp554Tbl', 'Row1', 'cf2:Level', 'LZ3'
put 'csp554Tbl', 'Row2', 'cf2:Level', 'AR7'
scan 'csp554Tbl'
```

```
hbase:003:0> put 'csp554Tbl', 'Row1', 'cf1:Name', 'Sam'
Took 0.7883 seconds
hbase:004:0> put 'csp554Tbl', 'Row2', 'cf1:Name', 'Ahmed'
Took 0.0317 seconds
hbase:005:0> put 'csp554Tbl', 'Row1', 'cf2:Job', 'Pilot'
Exercise 4Took 0.0167 seconds

hbase:006:0> put 'csp554Tbl', 'Row2', 'cf2:Job', 'Doctor'
Took 0.0138 seconds
hbase:007:0> put 'csp554Tbl', 'Row1', 'cf2:Level', 'LZ3'
Took 0.0146 seconds
hbase:008:0> put 'csp554Tbl', 'Row2', 'cf2:Level', 'AR7'
Took 0.0125 seconds
hbase:009:0> scan 'csp554Tbl'
ROW
Row1
Row1
Row1
Row2
Row2
Row2
2 row(s)
Took 0.4816 seconds
COLUMN+CELL
column=cf1:Name, timestamp=2023-11-29T01:13:04.826, value=Sam
column=cf2:Job, timestamp=2023-11-29T01:13:14.271, value=Pilot
column=cf2:Level, timestamp=2023-11-29T01:13:14.568, value=LZ3
column=cf1:Name, timestamp=2023-11-29T01:13:11.931, value=Ahmed
column=cf2:Job, timestamp=2023-11-29T01:13:14.427, value=Doctor
column=cf2:Level, timestamp=2023-11-29T01:13:14.733, value=AR7
```

Exercise 3

```
get 'csp554Tbl', 'Row1', {COLUMN => ['cf1', 'cf2:Level']}
```

```
hbase:018:0> get 'csp554Tbl', 'Row1', {COLUMN => ['cf2:Level']}
COLUMN
cf2:Level
CELL
timestamp=2023-11-30T20:52:16.324, value=LZ3
1 row(s)
Took 0.0136 seconds
```

Exercise 4

```
get 'csp554Tbl', 'Row2', {COLUMN => ['cf1:Name']}
```

```
hbase:012:0> get 'csp554Tbl', 'Row2', {COLUMN => ['cf1:Name']}
COLUMN
cf1:Name
CELL
timestamp=2023-11-29T01:13:11.931, value=Ahmed
1 row(s)
Took 0.0195 seconds
```

Exercise 5

```
scan 'csp554Tbl', {LIMIT => 2}
```

```
hbase:013:0> scan 'csp554Tbl', {LIMIT=>2}
ROW
Row1
Row1
Row1
Row2
Row2
Row2
2 row(s)
Took 0.0860 seconds
COLUMN+CELL
column=cf1:Name, timestamp=2023-11-29T01:13:04.826, value=Sam
column=cf2:Job, timestamp=2023-11-29T01:13:14.271, value=Pilot
column=cf2:Level, timestamp=2023-11-29T01:13:14.568, value=LZ3
column=cf1:Name, timestamp=2023-11-29T01:13:11.931, value=Ahmed
column=cf2:Job, timestamp=2023-11-29T01:13:14.427, value=Doctor
column=cf2:Level, timestamp=2023-11-29T01:13:14.733, value=AR7
```

CASSANDRA

Exercise 1

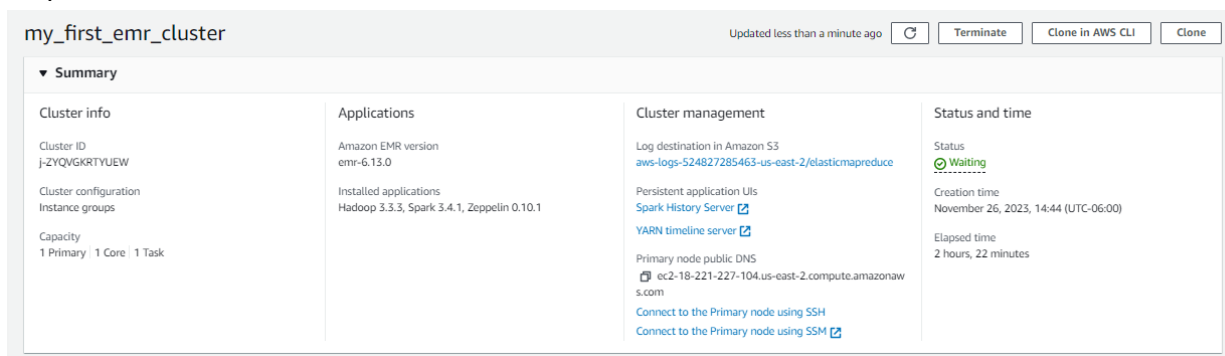
Apache Cassandra stands out as a prominent distributed database, known for its transactional capabilities, scalability, and robust availability. It effectively manages some of the largest datasets globally, distributed across numerous nodes spanning multiple datacenters. The widespread adoption of Cassandra in big data applications can be attributed to various factors, including its fault-tolerant peer-to-peer architecture, adaptable data model stemming from the BigTable data model, and the user-friendly Cassandra Query Language (CQL) that simplifies declarative queries.

This paper introduces an innovative methodology for data modeling in Apache Cassandra driven by queries. It then proceeds to compare this approach with traditional data modeling techniques. The proposed methodology significantly deviates from traditional relational data modeling by prioritizing application workflow and access patterns in the modeling process. Notable distinctions include emphasis on data nesting and duplication. The paper provides a detailed exploration of the Cassandra data model, covering conceptual data modeling and application workflows. Subsequent sections delve into query-driven mapping, physical modeling, and Chebotko diagrams.

Additionally, the paper introduces a robust data modeling tool named KDM, designed to automate complex, error-prone, and time-consuming data modeling tasks. KDM also facilitates CQL generation. The paper concludes by outlining future directions and potential avenues for further enhancing their proposed solution.

Exercise 2

Step A



The screenshot displays the Amazon EMR console for a cluster named 'my_first_emr_cluster'. The cluster is in a 'Waiting' status. The console shows various details including the cluster ID, configuration, capacity, applications, and management links.

Cluster info	Applications	Cluster management	Status and time
Cluster ID j-ZYQVGKRTYUEW	Amazon EMR version emr-6.13.0	Log destination in Amazon S3 aws-logs-524827285463-us-east-2/elasticmapreduce	Status Waiting
Cluster configuration Instance groups	Installed applications Hadoop 3.3.3, Spark 3.4.1, Zeppelin 0.10.1	Persistent application UIs Spark History Server YARN timeline server	Creation time November 26, 2023, 14:44 (UTC-06:00)
Capacity 1 Primary 1 Core 1 Task		Primary node public DNS ec2-18-221-227-104.us-east-2.compute.amazonaws.com Connect to the Primary node using SSH Connect to the Primary node using SSM	Elapsed time 2 hours, 22 minutes

Step B

Cass-Term

```
wget https://archive.apache.org/dist/cassandra/3.11.2/apache-cassandra-3.11.2-bin.tar.gz
tar -xzf apache-cassandra-3.11.2-bin.tar.gz
```

```

Ankan Mazumdar@DESKTOP-CHUL8BA WING64 ~
$ ssh -i /c/Users/Ankan/Mazumdar/Downloads/emr_key_pair.pem hadoop@ec2-18-224-213-198.us-east-2.compute.amazonaws.com

  _   _  _   _
 | | | | | | | |
 | |_| | | |_| |
 |  __| |  __|
 | |__| | |__| |
 |_____|_|_____|

Amazon Linux 2 AMI

https://aws.amazon.com/amazon-linux-2/
53 package(s) needed for security, out of 85 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::::: M::::::::: M::::::::: R:::::::::::::::::R
EE::::::::EEEEEEEEEEE M::::::::: M::::::::: R::::::::RRRRRR::::R
E::::E EEEEE M::::::::: M::::::::: RR::::R R::::R
E::::E M::::::::: M M::M M::M M::M R::R R::::R
E::::::::EEEEEEEEEE M::M M::M M::M M::M R::RRRRRR::::R
E::::::::::::::::: M::M M::M M::M M::M R::::::::RRR
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 M M M::M R::R R::::R
EE::::::::EEEEEEEEEE M::M M::M R::R R::::R
E::::::::::::::::: M::M M::M RR::::R R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-46-79 ~]$ wget https://archive.apache.org/dist/cassandra/3.11.2/apache-cassandra-3.11.2-bin.tar.gz
--2023-11-29 03:29:06-- https://archive.apache.org/dist/cassandra/3.11.2/apache-cassandra-3.11.2-bin.tar.gz
Resolving archive.apache.org (archive.apache.org)... 65.108.204.189, 2a01:4f9:1a:a084::2
Connecting to archive.apache.org (archive.apache.org)|65.108.204.189|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 38436262 (37M) [application/x-gzip]
Saving to: 'apache-cassandra-3.11.2-bin.tar.gz'

100%[=====] 38,436,262 16.3MB/s in 2.2s

2023-11-29 03:29:09 (16.3 MB/s) - 'apache-cassandra-3.11.2-bin.tar.gz' saved [38436262/38436262]

[hadoop@ip-172-31-46-79 ~]$ tar -xzf apache-cassandra-3.11.2-bin.tar.gz
apache-cassandra-3.11.2/bin/
apache-cassandra-3.11.2/conf/
apache-cassandra-3.11.2/conf/triggers/
apache-cassandra-3.11.2/doc/
apache-cassandra-3.11.2/doc/cqlsh/

```

apache-cassandra-3.11.2/bin/cassandra &

```

[hadoop@ip-172-31-46-79 ~]$ apache-cassandra-3.11.2/bin/cassandra &
[1] 15321
[hadoop@ip-172-31-46-79 ~]$ OpenJDK 64-Bit Server VM warning: Cannot open file apache-cassandra-3.11.2/bin/../logs/gc.log due to No such file or directory

CompilerOracle: dontinline org/apache/cassandra/db/Columns$Serializer.serializeLargeSubset (Lorg/apache/cassandra/io/util/DataInputPlus;Lorg/apache/cassandra/db/Columns;I)Lorg/apac
e/cassandra/db/Columns;
CompilerOracle: dontinline org/apache/cassandra/db/Columns$Serializer.serializeLargeSubset (Ljava/util/Collection;ILorg/apache/cassandra/db/Columns;ILorg/apache/cassandra/io/util/Data
OutputPlus;)V
CompilerOracle: dontinline org/apache/cassandra/db/Columns$Serializer.serializeLargeSubsetSize (Ljava/util/Collection;ILorg/apache/cassandra/db/Columns;I)I
CompilerOracle: dontinline org/apache/cassandra/db/commitlog/AbstractCommitLogSegmentManager.advanceAllLocatingFrom (Lorg/apache/cassandra/db/commitlog/CommitLogSegment;J)
CompilerOracle: dontinline org/apache/cassandra/db/transform/BaseIterator.tryGetMoreContents ()Z
CompilerOracle: dontinline org/apache/cassandra/db/transform/StoppingTransformation.stop ()V
CompilerOracle: dontinline org/apache/cassandra/db/transform/StoppingTransformation.stopInPartition ()V
CompilerOracle: dontinline org/apache/cassandra/io/util/BufferedDataOutputStreamPlus.doFlush ()V
CompilerOracle: dontinline org/apache/cassandra/io/util/BufferedDataOutputStreamPlus.writeExcessSlow ()V
CompilerOracle: dontinline org/apache/cassandra/io/util/BufferedDataOutputStreamPlus.writeSlow ()J)V
CompilerOracle: dontinline org/apache/cassandra/io/util/RebufferingInputStream.readPrimitiveSlowly (I)I
CompilerOracle: inline org/apache/cassandra/db/rows/UnfilteredSerializer.serializeRowBody (Lorg/apache/cassandra/db/rows/Row;ILorg/apache/cassandra/db/SerializationHeader;Lorg/apache/
cassandra/io/util/DataOutputPlus;)V
CompilerOracle: inline org/apache/cassandra/io/util/Memory.checkBounds ()JJ)V
CompilerOracle: inline org/apache/cassandra/io/util/SafeMemory.checkBounds ()JJ)V
CompilerOracle: inline org/apache/cassandra/utils/AsymmetricOrdering.selectBoundary (Lorg/apache/cassandra/utils/AsymmetricOrdering;Lorg/apache/cassandra/utils/AsymmetricOrdering;O;I)I
CompilerOracle: inline org/apache/cassandra/utils/AsymmetricOrdering.strictnessOfLessThan (Lorg/apache/cassandra/utils/AsymmetricOrdering;O;I)I
CompilerOracle: inline org/apache/cassandra/utils/BloomFilter.indexes (Lorg/apache/cassandra/utils/IFilter/FilterKey;)J
CompilerOracle: inline org/apache/cassandra/utils/BloomFilter.setIndexes ()JJJ)V
CompilerOracle: inline org/apache/cassandra/utils/ByteBufferUtil.compare (Ljava/nio/ByteBuffer;B)I
CompilerOracle: inline org/apache/cassandra/utils/ByteBufferUtil.compare ([Ljava/nio/ByteBuffer;)I
CompilerOracle: inline org/apache/cassandra/utils/ByteBufferUtil.compareUnsigned (Ljava/nio/ByteBuffer;Ljava/nio/ByteBuffer;)I
CompilerOracle: inline org/apache/cassandra/utils/ByteBufferUtil.compareUnsigned (Ljava/nio/ByteBuffer;Ljava/nio/ByteBuffer;)I
CompilerOracle: inline org/apache/cassandra/utils/FastByteOperations$UnsafeOperations.compareTo (Ljava/lang/Object;JILjava/nio/ByteBuffer;)I
CompilerOracle: inline org/apache/cassandra/utils/FastByteOperations$UnsafeOperations.compareTo (Ljava/lang/Object;JILjava/nio/ByteBuffer;)I
CompilerOracle: inline org/apache/cassandra/utils/vint/VintCoding.encodeVInt ()I)V
INFO [main] 2023-11-29 03:29:37,546 YamlConfigurationLoader.java:89 ~ Configuration location: file:/home/hadoop/apache-cassandra-3.11.2/conf/cassandra.yaml
INFO [main] 2023-11-29 03:29:38,248 Config.java:495 ~ Node configuration: [allocate_tokens_for_keyspace=null; authenticator=AllowAllAuthenticator; authorizer=AllowAllAuthorizer; auto_
bootstrap=true; auto_snapshot=true; back_pressure_enabled=false; back_pressure_strategy=org.apache.cassandra.net.RateBasedBackPressure(high_ratio=0.9, factor=5, flow=FAST); batch_size
_fail_threshold_in_kb=50; batch_size_warn_threshold_in_kb=5; batchlog_replay_throttle_in_kb=1024; broadcast_address=null; broadcast_rpc_address=null; buffer_pool_use_heap_if_exhausted
=true; cas_contention_timeout_in_ms=1000; cdc.enabled=false; cdc_free_space_check_interval_ms=250; cdc_raw_directory=null; cdc_total_space_in_mb=0; client_encryption_options=<REDACTED
>; cluster_name=Test Cluster; column_index_cache_size_in_kb=2; column_index_size_in_kb=64; commit_failure_policy=stop; commitlog_compression=null; commitlog_directory=null; commitlog_
max_compression_buffers_in_pool=3; commitlog_periodic_queue_size=5; commitlog_segment_size_in_mb=32; commitlog_sync=periodic; commitlog_sync_batch_window_in_ms=900; commitlog_sync_ne

```

Step C

Cqlsh-Term

apache-cassandra-3.11.2/bin/cqlsh

```
Ankan Mazumdar@DESKTOP-CMULEBA MINGW64 ~
$ ssh -i /c/Users/Ankan\ Mazumdar/Downloads/emr_key_pair.pem hadoop@ec2-18-224-213-198.us-east-2.compute.amazonaws.com
```

```

  _| _|_ )
  _| C /   Amazon Linux 2 AMI
  _|\_|_|_
```

```
https://aws.amazon.com/amazon-linux-2/
```

```

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M M::::::::M R::::::::::::R
EE::::::::EEEEEEEE::::E M::::::::M M::::::::M R::::::::RRRRR::::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::::M R::RRRRRR::::R
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 MMMMMMMM MMMMMMMM RRRRRRR RRRRRR
```

```
[hadoop@ip-172-31-46-79 ~]$ 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.
```

```
CREATE TABLE A20541357.Music (
  artistName text,
  albumName text,
  numberSold int,
  Cost int,
  PRIMARY KEY (artistName, albumName)
) WITH CLUSTERING ORDER BY (albumName DESC);
```

```
Ankan Mazumdar@DESKTOP-CMULEBA MINGW64 ~/Downloads/Big data CSP 554 Assignment 6-9/Homework9
$ scp -i /c/Users/Ankan\ Mazumdar/Downloads/emr_key_pair.pem init.cql ex2.cql ex3.cql ex4.cql ex5.cql hadoop@ec2-3-20-239-198.us-east-2.compute.amazonaws.com:/home/hadoop/
init.cql 100% 102 3.2KB/s 00:00
ex2.cql 100% 176 3.5KB/s 00:00
ex3.cql 100% 546 15.8KB/s 00:00
ex4.cql 100% 55 1.1KB/s 00:00
ex5.cql 100% 63 2.0KB/s 00:00
```

Exer 1

```
source './init.cql';
describe keyspaces;
USE A20541357;
```

```

cqlsh> source './init.cql';
cqlsh> describe keyspaces;

a20541357 system_schema system_auth system system_distributed system_traces

cqlsh> USE 20541357
...
cqlsh> USE 20541357;
Improper USE command.
cqlsh> USE a20541357
...
cqlsh> USE a20541357;
cqlsh:a20541357> source './ex2.cql';
```

Exer 2

```
source './ex2.cql';
DESCRIBE TABLE Music;
```

```
cqlsh:a20541357> source './ex2.cql';
cqlsh:a20541357> DESCRIBE TABLE Music;

CREATE TABLE a20541357.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';
```

Exercise 3

```
source './ex3.cql';
SELECT * FROM Music;
```

```
cqlsh:a20541357> source './ex3.cql';
cqlsh:a20541357> 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)
```

Exercise 4

```
source './ex4.cql';
```

```
cqlsh:a20541357> source './ex4.cql';

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

(2 rows)
```

Exercise 5

```
source './ex5.cql';
```

```
cqlsh:a20541357> 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)
cqlsh:a20541357> |
```

MONGODB

Step A

my_first_emr_cluster Updated less than a minute ago [Refresh](#) [Terminate](#) [Clone in AWS CLI](#) [Clone](#)

▼ Summary

Cluster info	Applications	Cluster management	Status and time
Cluster ID j-ZYQVGKRTYUEW Cluster configuration Instance groups Capacity 1 Primary 1 Core 1 Task	Amazon EMR version emr-6.13.0 Installed applications Hadoop 3.3.3, Spark 3.4.1, Zeppelin 0.10.1	Log destination in Amazon S3 aws-logs-524827285463-us-east-2/elasticmapreduce Persistent application UIs Spark History Server YARN timeline server Primary node public DNS ec2-18-221-227-104.us-east-2.compute.amazonaws.com Connect to the Primary node using SSH Connect to the Primary node using SSM	Status Waiting Creation time November 26, 2023, 14:44 (UTC-06:00) Elapsed time 2 hours, 22 minutes

STEP B

```
scp -i /c/Users/Ankan\ Mazumdar/Downloads/emr_key_pair.pem mongoex.tar mongodb-org-4.2.repo
hadoop@ec2-18-218-171-77.us-east-2.compute.amazonaws.com:/home/hadoop/
mongoex.tar
```

```
Ankan Mazumdar@DESKTOP-CMULEBA MINGW64 ~/Downloads/Big data CSP 554 Assignment 6-9/Homework9
$ scp -i /c/Users/Ankan\ Mazumdar/Downloads/emr_key_pair.pem mongoex.tar mongodb-org-4.2.repo hadoop@ec2-18-218-171-77.us-east-2.compute.amazonaws.com:/home/hadoop/
mongoex.tar 100% 14KB 357.0KB/s 00:00
mongodb-org-4.2.repo 100% 197 6.0KB/s 00:00
Ankan Mazumdar@DESKTOP-CMULEBA MINGW64 ~/Downloads/Big data CSP 554 Assignment 6-9/Homework9
$ |
```

Step C

Init-Term

```
sudo cp mongodb-org-4.2.repo /etc/yum.repos.d
tar -xvf mongoex.tar
```

```
[hadoop@ip-172-31-46-235 ~]$ sudo cp mongodb-org-4.2.repo /etc/yum.repos.d
[hadoop@ip-172-31-46-235 ~]$ tar -xvf mongoex.tar
./._demo1.js
demo1.js
demo2.js
demo3.js
demo4.js
demo5.js
demo6.js
demo7.js
demo8.js
demo9.js
load.js
```

Step D – Install and start MongoDB

`sudo yum install -y mongodb-org-4.2.15 mongodb-org-server-4.2.15 mongodb-org-shell-4.2.15 mongodb-org-mongos-4.2.15 mongodb-org-tools-4.2.15`

```
[hadoop@ip-172-31-46-235 ~]$ sudo yum install -y mongodb-org-4.2.15 mongodb-org-server-4.2.15 mongodb-org-shell-4.2.15 mongodb-org-mongos-4.2.15 mongodb-org-tools-4.2.15
Loaded plugins: extras_suggestions, langpacks, priorities, update-motd
14 packages excluded due to repository priority protections
Resolving Dependencies
--> Running transaction check
--> Package mongodb-org.x86_64 0:4.2.15-1.amzn1 will be installed
--> Package mongodb-org-mongos.x86_64 0:4.2.15-1.amzn1 will be installed
--> Package mongodb-org-server.x86_64 0:4.2.15-1.amzn1 will be installed
--> Package mongodb-org-shell.x86_64 0:4.2.15-1.amzn1 will be installed
--> Package mongodb-org-tools.x86_64 0:4.2.15-1.amzn1 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

=====================================================================================================================================
 Package Arch Version Repository Size
=====================================================================================================================================
Installing:
mongodb-org x86_64 4.2.15-1.amzn1 mongodb-org-4.2 6.0 k
mongodb-org-mongos x86_64 4.2.15-1.amzn1 mongodb-org-4.2 15 M
mongodb-org-server x86_64 4.2.15-1.amzn1 mongodb-org-4.2 26 M
mongodb-org-shell x86_64 4.2.15-1.amzn1 mongodb-org-4.2 17 M
mongodb-org-tools x86_64 4.2.15-1.amzn1 mongodb-org-4.2 32 M
Transaction Summary
-----
Install 5 Packages

Total download size: 89 M
Installed size: 252 M
Downloading packages:
warning: /mnt/var/cache/yum/x86_64/2/mongodb-org-4.2/packages/mongodb-org-4.2.15-1.amzn1.x86_64.rpm: Header V3 RSA/SHA1 Signature, key ID 058f8b6b: NOKEY
Public key for mongodb-org-4.2.15-1.amzn1.x86_64.rpm is not installed
(1/5): mongodb-org-4.2.15-1.amzn1.x86_64.rpm | 6.0 kB 00:00:00
(2/5): mongodb-org-mongos-4.2.15-1.amzn1.x86_64.rpm | 15 MB 00:00:00
(3/5): mongodb-org-shell-4.2.15-1.amzn1.x86_64.rpm | 17 MB 00:00:00
(4/5): mongodb-org-server-4.2.15-1.amzn1.x86_64.rpm | 26 MB 00:00:00
(5/5): mongodb-org-tools-4.2.15-1.amzn1.x86_64.rpm | 32 MB 00:00:00
Total | 89 MB 00:00:01
Retrieving key from https://www.mongodb.org/static/pgp/server-4.2.asc
Importing GPG key 0x058f8b6b:
```

`sudo systemctl start mongod`

```
[hadoop@ip-172-31-46-235 ~]$ sudo systemctl start mongod
[hadoop@ip-172-31-46-235 ~]$ |
```

Step E – Start the MongoDB Shell (Command Line Interpreter) Mongo


```
[hadoop@ip-172-31-46-235 ~]$ mongo
MongoDB shell version v4.2.15
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("5072503e-18a7-498a-8ba7-80973958d313") }
MongoDB server version: 4.2.15
Welcome to the MongoDB shell.
For interactive help, type "help".
For more comprehensive documentation, see
  https://docs.mongodb.com/
Questions? Try the MongoDB Developer Community Forums
  https://community.mongodb.com
Server has startup warnings:
2023-11-29T05:35:00.790+0000 I CONTROL [initandlisten]
2023-11-29T05:35:00.790+0000 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2023-11-29T05:35:00.790+0000 I CONTROL [initandlisten] **           Read and write access to data and configuration is unrestricted.
2023-11-29T05:35:00.790+0000 I CONTROL [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).
```

Step G-
use assignment;
load('./load.js');
db.unicorns.find();
oad('./demo1.js');

```
> use assignment;
switched to db assignment
> load('./load.js');
true
> db.unicorns.find():
{ "_id" : ObjectId("6566cdf60f8ddc27d15b928b"), "name" : "Horny", "dob" : ISODate("1992-03-13T07:47:00Z"), "loves" : [ "carrot", "papaya" ], "weight" : 600, "gender" : "m", "vampires" : 63 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b928c"), "name" : "Aurora", "dob" : ISODate("1991-01-24T13:00:00Z"), "loves" : [ "carrot", "grape" ], "weight" : 450, "gender" : "f", "vampires" : 43 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b928d"), "name" : "Unicrom", "dob" : ISODate("1973-02-09T22:10:00Z"), "loves" : [ "energon", "redbull" ], "weight" : 984, "gender" : "m", "vampires" : 182 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b928e"), "name" : "Roooooondles", "dob" : ISODate("1979-08-18T18:44:00Z"), "loves" : [ "apple" ], "weight" : 575, "gender" : "m", "vampires" : 99 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b928f"), "name" : "Solnara", "dob" : ISODate("1985-07-04T02:01:00Z"), "loves" : [ "apple", "carrot", "chocolate" ], "weight" : 550, "gender" : "f", "vampires" : 80 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9290"), "name" : "Ayna", "dob" : ISODate("1998-03-07T08:30:00Z"), "loves" : [ "strawberry", "lemon" ], "weight" : 733, "gender" : "f", "vampires" : 40 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9291"), "name" : "Kenny", "dob" : ISODate("1997-07-01T10:42:00Z"), "loves" : [ "grape", "lemon" ], "weight" : 690, "gender" : "m", "vampires" : 39 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9292"), "name" : "Raleigh", "dob" : ISODate("2005-05-03T00:57:00Z"), "loves" : [ "apple", "sugar" ], "weight" : 421, "gender" : "m", "vampires" : 2 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9293"), "name" : "Leia", "dob" : ISODate("2001-10-08T14:53:00Z"), "loves" : [ "apple", "watermelon" ], "weight" : 601, "gender" : "f", "vampires" : 33 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9294"), "name" : "Pilot", "dob" : ISODate("1997-03-01T05:03:00Z"), "loves" : [ "apple", "watermelon" ], "weight" : 650, "gender" : "m", "vampires" : 54 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9295"), "name" : "Nimue", "dob" : ISODate("1999-12-20T16:15:00Z"), "loves" : [ "grape", "carrot" ], "weight" : 540, "gender" : "f" }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9296"), "name" : "Dunx", "dob" : ISODate("1976-07-18T18:18:00Z"), "loves" : [ "grape", "watermelon" ], "weight" : 704, "gender" : "m", "vampires" : 165 }
>
> load('./demo1.js');
true
```

Exercise 1
db.unicorns.find({weight:{<500}})

```
true
> db.unicorns.find({weight:{<500}})
{ "_id" : ObjectId("6566cdf60f8ddc27d15b928c"), "name" : "Aurora", "dob" : ISODate("1991-01-24T13:00:00Z"), "loves" : [ "carrot", "grape" ], "weight" : 450, "gender" : "f", "vampires" : 43 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9292"), "name" : "Raleigh", "dob" : ISODate("2005-05-03T00:57:00Z"), "loves" : [ "apple", "sugar" ], "weight" : 421, "gender" : "m", "vampires" : 2 }
```

Exercise 2
db.unicorns.find({loves:"apple"})

```
> db.unicorns.find({loves:"apple"})
{ "_id" : ObjectId("6566cdf60f8ddc27d15b928e"), "name" : "Rooooooodles", "dob" : ISODate("1979-08-18T18:44:00Z"), "loves" : [ "apple" ], "weight" : 575, "gender" : "m", "vampires" : 99 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b928f"), "name" : "Solnara", "dob" : ISODate("1985-07-04T02:01:00Z"), "loves" : [ "apple", "carrot", "chocolate" ], "weight" : 550, "gender" : "f", "vampires" : 80 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9292"), "name" : "Raleigh", "dob" : ISODate("2005-05-03T00:57:00Z"), "loves" : [ "apple", "sugar" ], "weight" : 421, "gender" : "m", "vampires" : 2 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9293"), "name" : "Leia", "dob" : ISODate("2001-10-08T14:53:00Z"), "loves" : [ "apple", "watermelon" ], "weight" : 601, "gender" : "f", "vampires" : 33 }
{ "_id" : ObjectId("6566cdf60f8ddc27d15b9294"), "name" : "Pilot", "dob" : ISODate("1997-03-01T05:03:00Z"), "loves" : [ "apple", "watermelon" ], "weight" : 650, "gender" : "m", "vampires" : 54 }
```

Exercise 3

```
db.unicorns.insertOne({name:"Malini", dob:"2008-03-11", loves:["pears", "grapes"], weight:450,
gender:"F", vampires:23, horns:1})
db.unicorns.find({name:"Malini"})
```

```
> db.unicorns.insertOne({name:"Malini", dob:"2008-03-11", loves:["pears", "grapes"],weight:450, gender:"F", vampires:23, horns:1})
{
  "acknowledged" : true,
  "insertedId" : ObjectId("65690333e2993dfcde5dae0e")
}
> db.unicorns.find({name:"Malini"})
{ "_id" : ObjectId("65690333e2993dfcde5dae0e"), "name" : "Malini", "dob" : "2008-03-11", "loves" : [ "pears", "grapes" ], "weight" : 450, "gender" : "F", "vampires" : 23, "horns" : 1 }
```

Exercise 4

```
db.unicorns.updateOne({name:"Malini"}, {$set:{loves:["pears", "grapes", "apricots"]}})
db.unicorns.find({name:"Malini"})
```

```
> db.unicorns.updateOne({name:"Malini"}, {$set:{loves:["pears", "grapes", "apricots"]}})
{ "acknowledged" : true, "matchedCount" : 1, "modifiedCount" : 1 }
> db.unicorns.find({name:"Malini"})
{ "_id" : ObjectId("6566ce2a0f8ddc27d15b9298"), "name" : "Malini", "dob" : "2008-03-11", "loves" : [ "pears", "grapes", "apricots" ], "weight" : 450, "gender" : "F", "vampires" : 23, "horns" : 1 }
```

Exercise 5

```
db.unicorns.deleteMany({weight:{$gt:600}})
db.unicorns.find({weight:{$gt:600}})
```

```
>
> db.unicorns.deleteMany({weight:{$gt:600}})
{ "acknowledged" : true, "deletedCount" : 6 }
> db.unicorns.find({weight:{$gt:600}})
> |
```