Big Data Technologies
Assignment - 11

EMR CONNECTION

```
Last login: Sat Nov 18 00:10:51 on ttys011
[(base) sailavanyanarthu@Sailavanyas-Air ~ % cd downloads
[(base) sailavanyanarthu@Sailavanyas-Air downloads % chmod 400 assign11.pem
(base) sailavanyanarthu@Sailavanyas-Air downloads % ssh -i assign11.pem ec2-user@ec2-3-145-9-40.us-east-2.compute.amazonaws.com

Last login: Sat Nov 18 06:15:23 2023 from 207.237.235.14
      ,        #_
    ~\_   ####_        Amazon Linux 2
   ~~  \_#####\
   ~~     \###|        AL2 End of Life is 2025-06-30.
   ~~      \#/ ___
    ~~     V~' '->
     ~~~         /     A newer version of Amazon Linux is available!
       ~~._.   _/
          _/ _/        Amazon Linux 2023, GA and supported until 2028-03-15.
        _/m/'            https://aws.amazon.com/linux/amazon-linux-2023/

16 package(s) needed for security, out of 31 available
Run "sudo yum update" to apply all updates.

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:::::EEEEEEEEE::::E M:::::M             M:::::M   R:::R      R::::R
E::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR
```

hbase shell command

```
[[ec2-user@ip-172-31-9-176 ~]$ hbase shell
log4j:ERROR setFile(null,true) call failed.
java.io.FileNotFoundException: /var/log/hbase/hbase.log (Permission denied)
        at java.io.FileOutputStream.open0(Native Method)
        at java.io.FileOutputStream.open(FileOutputStream.java:270)
        at java.io.FileOutputStream.<init>(FileOutputStream.java:213)
        at java.io.FileOutputStream.<init>(FileOutputStream.java:133)
        at org.apache.log4j.FileAppender.setFile(FileAppender.java:282)
        at org.apache.log4j.FileAppender.activateOptions(FileAppender.java:166)
        at org.apache.log4j.DailyRollingFileAppender.activateOptions(DailyRollingFileAppender.java:216)
        at org.apache.log4j.config.PropertySetter.activate(PropertySetter.java:284)
        at org.apache.log4j.config.PropertySetter.setProperties(PropertySetter.java:160)
        at org.apache.log4j.config.PropertySetter.setProperties(PropertySetter.java:100)
        at org.apache.log4j.PropertyConfigurator.parseAppender(PropertyConfigurator.java:802)
        at org.apache.log4j.PropertyConfigurator.parseCategory(PropertyConfigurator.java:738)
        at org.apache.log4j.PropertyConfigurator.configureRootCategory(PropertyConfigurator.java:630)
        at org.apache.log4j.PropertyConfigurator.doConfigure(PropertyConfigurator.java:516)
        at org.apache.log4j.PropertyConfigurator.doConfigure(PropertyConfigurator.java:577)
```

Exercice - 1

```
[hbase:001:0> create 'csp554Tb1', 'cf1', 'cf2'
Created table csp554Tb1
Took 1.2880 seconds
=> Hbase::Table - csp554Tb1
[hbase:002:0> describe 'csp554Tb1'
Table csp554Tb1 is ENABLED
csp554Tb1
COLUMN FAMILIES DESCRIPTION
{NAME => 'cf1', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCA
CHE => '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', MIN_VERSIONS => '0', BLOCKCA
CHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}

2 row(s)
Quota is disabled
Took 0.1733 seconds
hbase:003:0>
```

Exercise - 2

```
[hbase:003:0> put 'csp554Tb1', 'Row1', 'cf1:name', 'sam'
Took 0.1240 seconds
[hbase:004:0> put 'csp554Tb1', 'Row2', 'cf1:name', 'Ahmed'
Took 0.0113 seconds
[hbase:005:0> put 'csp554Tb1', 'Row1', 'cf2:job', 'Pilot'
Took 0.0070 seconds
[hbase:006:0> put 'csp554Tb1', 'Row2', 'cf2:job', 'Doctor'
Took 0.0079 seconds
[hbase:007:0> put 'csp554Tb1', 'Row1', 'cf2:level', 'LZ3'
Took 0.0061 seconds
[hbase:008:0> put 'csp554Tb1', 'Row2', 'cf2:level', 'AR7'
Took 0.0072 seconds
[hbase:009:0> scan 'csp554Tb1'
ROW                               COLUMN+CELL
 Row1                             column=cf1:name, timestamp=2023-11-18T06:24:54.821, value=sam
 Row1                             column=cf2:job, timestamp=2023-11-18T06:26:30.002, value=Pilot
 Row1                             column=cf2:level, timestamp=2023-11-18T06:28:20.348, value=LZ3
 Row2                             column=cf1:name, timestamp=2023-11-18T06:25:23.026, value=Ahmed
 Row2                             column=cf2:job, timestamp=2023-11-18T06:27:30.561, value=Doctor
 Row2                             column=cf2:level, timestamp=2023-11-18T06:28:50.270, value=AR7
2 row(s)
Took 0.0498 seconds
hbase:010:0>
```

Exercise - 3

```
[hbase:010:0>
[hbase:011:0> get 'csp554Tb1', 'Row1', {COLUMN=>'cf2:level'}
COLUMN                               CELL
 cf2:level                           timestamp=2023-11-18T06:28:20.348, value=LZ3
1 row(s)
Took 0.0227 seconds
[hbase:012:0>
```

Exercise - 4

```
hbase:012:0>
hbase:013:0> get 'csp554Tb1', 'Row2', {COLUMN=>'cf1:name'}
COLUMN                               CELL
 cf1:name                            timestamp=2023-11-18T06:25:23.026, value=Ahmed
1 row(s)
Took 0.0152 seconds
[hbase:014:0>
```

Exercise - 5

```
[hbase:014:0>
[hbase:015:0> scan 'csp554Tb1', {LIMIT=>2}
ROW                               COLUMN+CELL
 Row1                             column=cf1:name, timestamp=2023-11-18T06:24:54.821, value=sam
 Row1                             column=cf2:job, timestamp=2023-11-18T06:26:30.002, value=Pilot
 Row1                             column=cf2:level, timestamp=2023-11-18T06:28:20.348, value=LZ3
 Row2                             column=cf1:name, timestamp=2023-11-18T06:25:23.026, value=Ahmed
 Row2                             column=cf2:job, timestamp=2023-11-18T06:27:30.561, value=Doctor
 Row2                             column=cf2:level, timestamp=2023-11-18T06:28:50.270, value=AR7
2 row(s)
Took 0.0198 seconds
hbase:016:0>
```

Submitted by:
Sailavanya Narthu
A20516764
snarthu@hawk.iit.edu