

Name: Namratha Renjal

Student ID: 23268158

Email: [Namratha.renjal2@mail.dcu.ie](mailto:Namratha.renjal2@mail.dcu.ie)

Gitlab repository: [https://gitlab.com/renjaln2/ca675\\_assignment1](https://gitlab.com/renjaln2/ca675_assignment1)

# CA675 – Cloud Technologies – Assignment 1

## SPAM DETECTION SYSTEM

### Task 1: Cloud Infrastructure Setup (AWS)

- Created an EMR cluster of three EC2 instances, with one core and two task nodes.
- All the EC2 instances are of type m4.large having 2vCPUs and 8GiB of memory.
- Installed Hadoop 3.3.6, Hive 3.1.3, Pig 0.17.0 and Spark 3.4.1 on all instances.

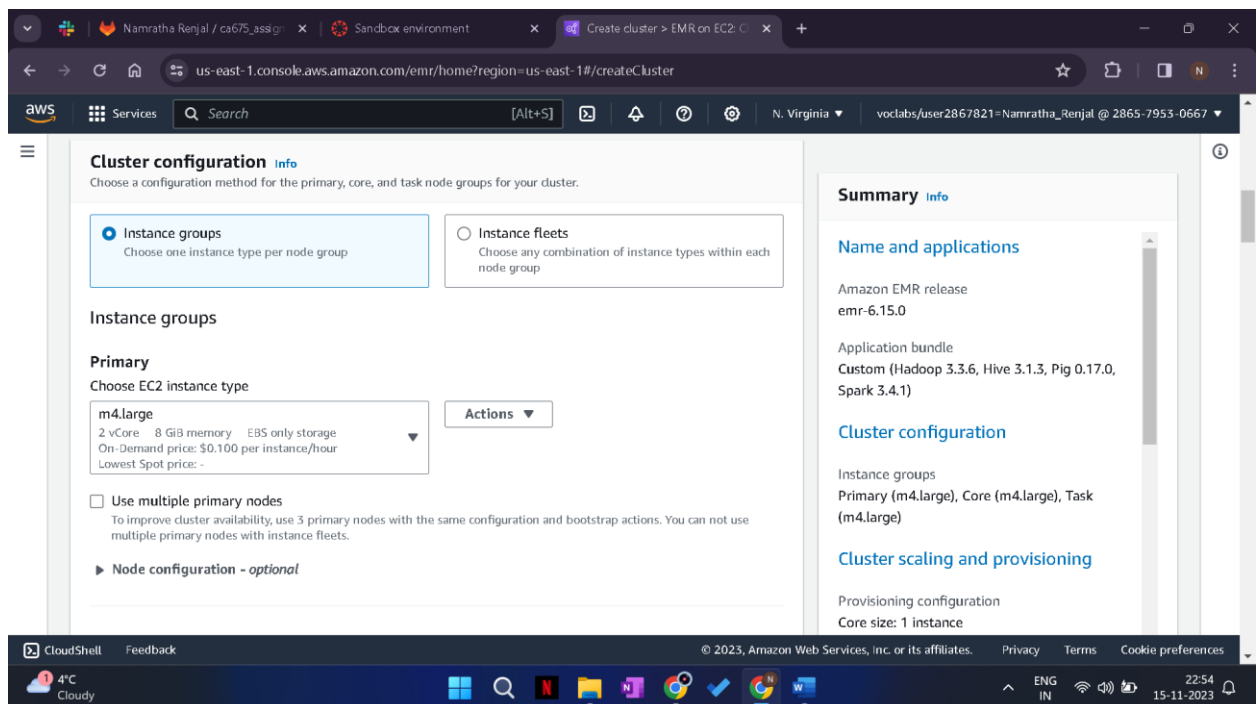


Figure 1: Instance type for primary node

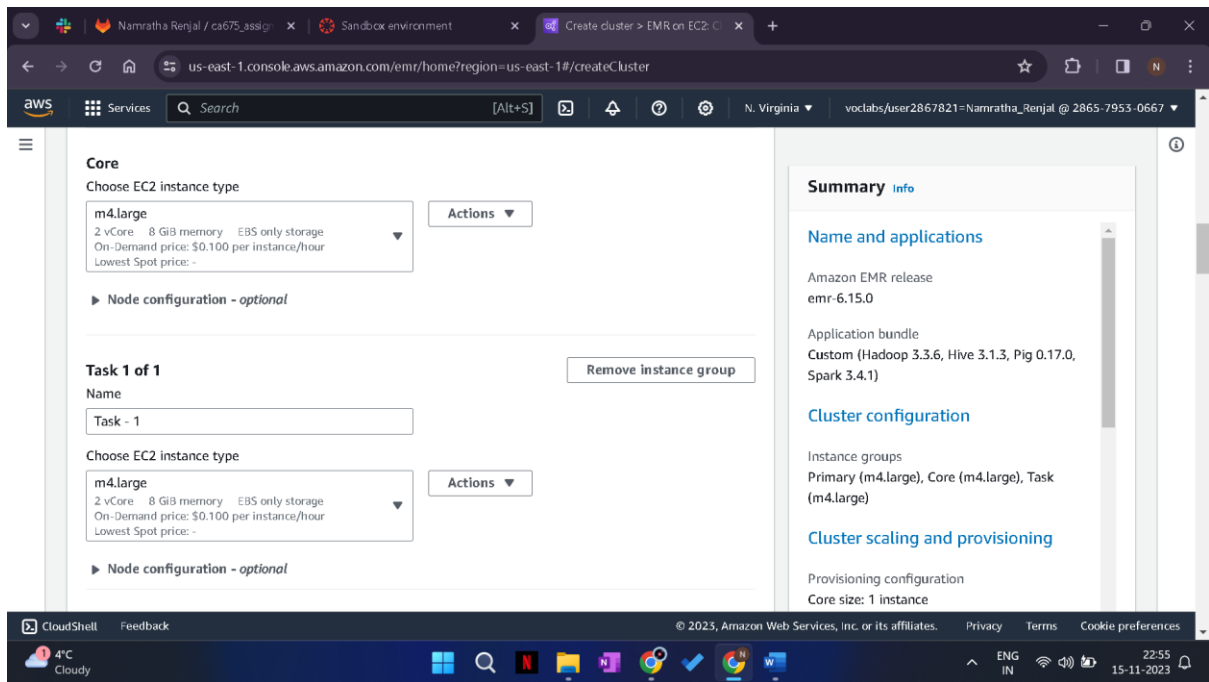


Figure 2: Instance types for core and task node

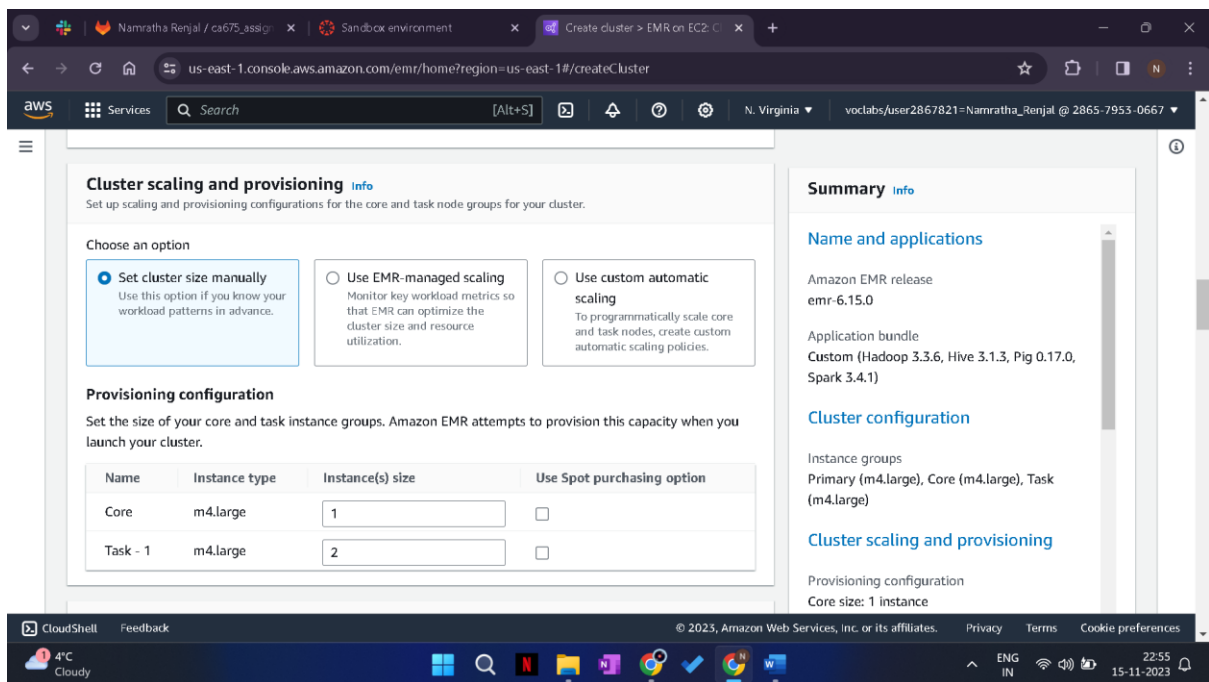


Figure 3: Size of cluster. Number of core and task nodes.

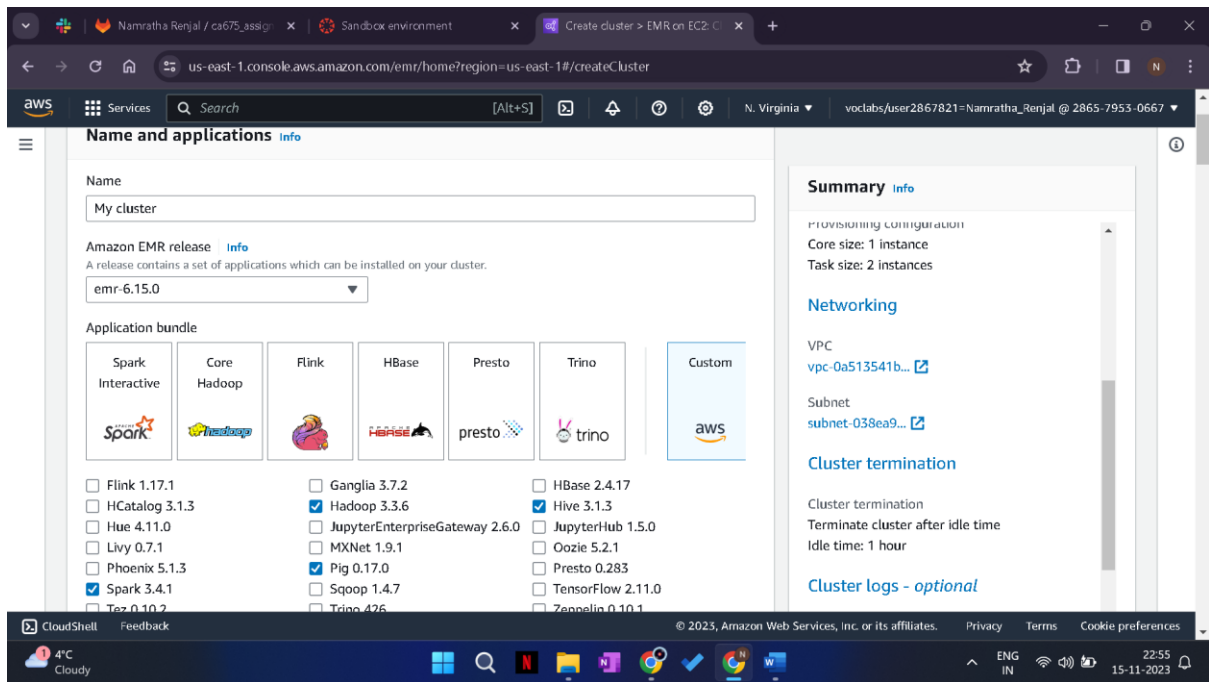


Figure 4: Applications installed in each instance

## Task 2: Dataset

- Picked youtube comments dataset from Kaggle since it had comment data along with account username.
- Dataset link: <https://www.kaggle.com/datasets/goneeee/youtube-spam-classifiedcomments>
- The data from the five files were merged into a single CSV file, the merged file is available in the gitlab repository.
- The dataset has five fields
  1. COMMENT\_ID: unique string values for each comment
  2. AUTHOR: string value identifying the account associated with the comment
  3. DATE: date of comment
  4. CONTENT: the comment string
  5. CLASS: 0 for ham and 1 for spam comments
- The dataset was then uploaded to an S3 bucket.
- Hive table was created for further processing and the dataset from the S3 bucket was loaded into the table
- Hive QL Query to create table:
 

```
CREATE TABLE youtube_comments(
  comment_id STRING,
  author STRING,
  comment_date date,
  comment_content STRING,
  label INT
) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://mycloudassignmentbucket/assignmentdata'
```

TBLPROPERTIES('skip.header.line.count'='1');

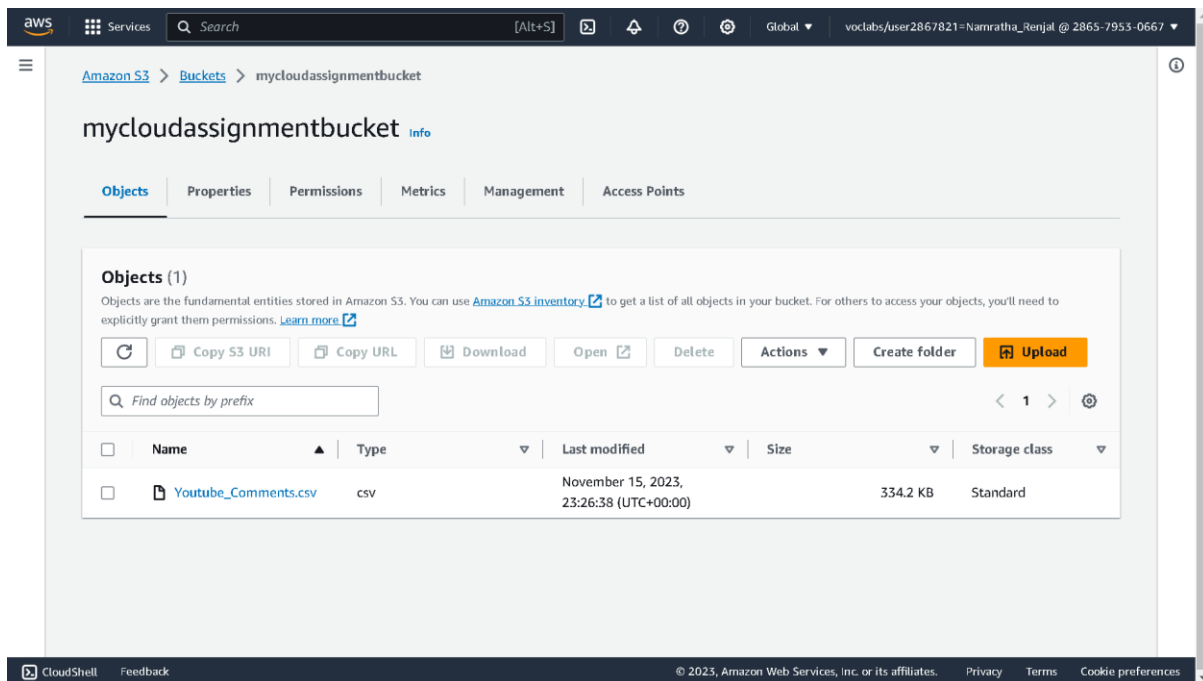


Figure 5: Dataset uploaded into AWS S3 bucket

- Hive QL query to load data from CSV to table  
LOAD DATA INPATH 'S3://mycloudassignmentbucket/Youtube\_Comments.csv'  
INTO TABLE youtube\_comments;

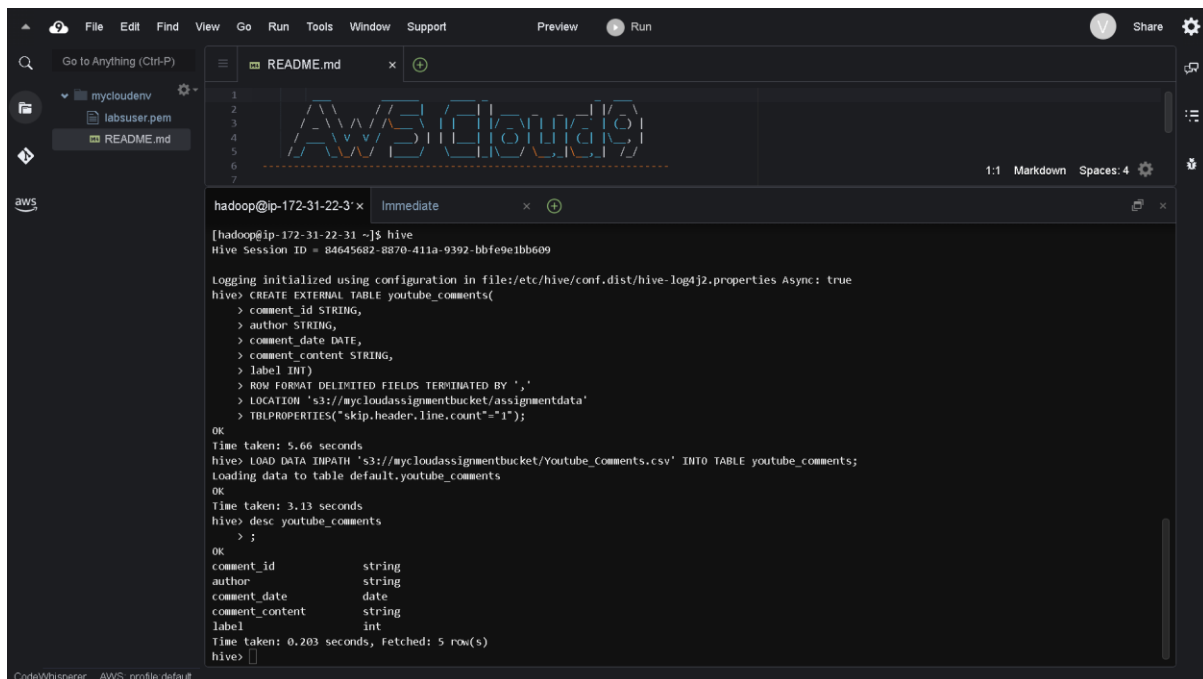


Figure 6: Table created in Hive to store data

```

-- README.md
-- CREATE TABLE youtube_comments (
--   author STRING,
--   comment_date DATE,
--   comment_content STRING,
--   label INT)
-- ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
-- LOCATION 's3://mycloudassignmentbucket/assignmentdata'
-- TBLPROPERTIES("skip.header.line.count"="1");
OK
Time taken: 5.66 seconds
hive> LOAD DATA INPATH 's3://mycloudassignmentbucket/Youtube_Comments.csv' INTO TABLE youtube_comments;
Loading data to table default.youtube_comments
OK
Time taken: 3.13 seconds
hive> desc youtube_comments
;
OK
comment_id      string
author          string
comment_date    date
comment_content string
label           int
Time taken: 0.203 seconds, Fetched: 5 row(s)
hive> select * from youtube_comments limit 10;
OK
LZ0PqHly8h80UyxuaD6hIGQYnQ96IUCG-AYMqPjpu  Julius RM      2013-11-07      Huh anyway check out this you[tube] channel: kobyoshi02 1
LZ0PqHly8h80UyxuaD6hIGQYnQ96IUCG-AYMqPjpu  adam riyati    2013-11-07      Hey guys check out my new channel and our first vid THIS IS US THE MO
HkYsY!!! I'm the sockey in the white shirt please leave a like comment and please subscribe!!!! 1
LZ0PqHly8h80UyxuaD6hIGQYnQ96IUCG-AYMqPjpu  Evgeny Murashkin  2013-11-08      Just for test I have to say murdev.com 1
z13jhp0b0xncus12p220zKasxwvzjaz04        Eliino Melender  2013-11-09      me shaking my sexy ass on my channel enjoy ^.^ 1
z13fubaploujthgqj04chlngvzatt3r3da        GsWega 2013-11-10  watch?v=vta8Gv6tWQ check this out . 1
LZ0PqHly8h80UyxuaD6hIGQYnQ96IUCG-AYMqPjpu  Jason Haddad    2013-11-26      Hey check out my new website!! This site is about kids stuff. kidsmedi
ausa . com 1
z131fd05vmd11c123te5uz2mq1g1brz04        ferleack ferles  2013-11-27      Subscribe to my channel 1
z122wfzgt30fhubn04cdn3xfx2axzngs140k      Bob Kanowski    2013-11-28      i turned it on mute as soon as i came on i just wanted to check the views...0
z13tt1jcrqesxk2o234ghbgzxywz1z7i04        Cony 2013-11-28  You should check my channel for Funny VIDEOS!! 1
z12avveb4xqilrsix04chxvilljryduwag0       BeBe Burkey     2013-11-28      and u should.d check my channel and tell me what I should do next! 1
Time taken: 2.917 seconds, Fetched: 10 row(s)
hive>

```

Figure 7: Dataset loaded into Hive table

## Task 3: Clean and process the data using Hive

- The comment text is converted to all lower case to make the dataset uniform and records with empty content field or empty author field or empty label were dropped.

```

hive> CREATE TABLE cleaned_comments AS
> SELECT author, lower(comment_content) AS content, label
> FROM youtube_comments WHERE author IS not null AND comment_content IS not null;
Query ID = hadoop_20231116001142_e53453dc-14e8-4508-9cdd-adac3f369ef8
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1700089803344_0003)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1          1          0          0          0          0
-----
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 22.61 s
-----
Moving data to directory hdfs://ip-172-31-22-31.ec2.internal:8020/user/hive/warehouse/cleaned_comments
OK
Time taken: 36.607 seconds

```

Figure 8: Hive QL query to convert content to lowercase and remove records with null values

```
hive> select * from cleaned_comments limit 10;
OK
Julius MM      huh anyway check out this you[tube] channel: kobyoshi02 1
adam riyati    hey guys check out my new channel and our first vid this is us the  monkeys!!! i'm the monkey in the white shirtplease leave a like co
mmment and please subscribe!!!! 1
Evgeny Murashkin      just for test i have to say murdev.com 1
ElNino Melendez me shaking my sexy ass on my channel enjoy ^_^ 1
GSMega watch?v=vtarggygtwq check this out . 1
Jason Haddad hey check out my new website!! this site is about kids stuff. kidsmediausa . com 1
ferleck ferles subscribe to my channel 1
Bob Kanowski i turned it on mute as soon as i came on i just wanted to check the views... 0
cony you should check my channel for funny videos!! 1
BeBe Burkey and u should.d check my channel and tell me what i should do next! 1
Time taken: 0.151 seconds, Fetched: 10 row(s)
```

Figure 9: Output after running query

- Some comments were observed to have HTML tags which are not relevant to ham or spam classification. Therefore, HTML tags are removed from the comment content fields.

```
hive> CREATE TABLE comments_without_html AS
> SELECT author, regexp_replace(content, '<[^>+>+', '') AS content, label
> FROM cleaned_comments;
Query ID = hadoop_20231116002651_2a68c5f6-1c84-45b0-905d-23d4d7cbb00
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1700089803344_0004)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01 [=====] 100% ELAPSED TIME: 8.79 s
-----
Moving data to directory hdfs://ip-172-31-22-31.ec2.internal:8020/user/hive/warehouse/comments_without_html
OK
Time taken: 22.262 seconds
```

Figure 10: HiveQL query to remove HTML tags

- Some comments were also observed to contain URLs which were removed from the contents field.

```
hive> CREATE TABLE comments_without_urls AS
> SELECT author, regexp_replace(content, 'http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|!*"\\(\\)|)(?:%[0-9a-fA-F][0-9a-fA-F]))+', 'url') AS content
> FROM comments_without_html;
Query ID = hadoop_20231116200653_d501d544-f400-4dcd-8818-d27e5a5aa3d0
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1700163558598_0002)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01 [=====] 100% ELAPSED TIME: 7.17 s
-----
Moving data to directory hdfs://ip-172-31-7-113.ec2.internal:8020/user/hive/warehouse/comments_without_urls
OK
Time taken: 7.894 seconds
```

Figure 11: Hive QL Query to remove URLs

- Special characters are removed from comment content since weights cannot be assigned to them.

```
hive> CREATE TABLE comments_cleaned AS
> SELECT author, regexp_replace(content, '[^a-zA-Z0-9\\s]', '') as content, label
> FROM comments_without_url;
Query ID = hadoop_20231116003714_86a18ba2-2904-4372-9a6e-c1c8dc669ae2
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1700089803344_0005)

-----
VERTICES    MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container    SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 01/01 [=====>>] 100% ELAPSED TIME: 6.83 s
-----
Moving data to directory hdfs://ip-172-31-22-31.ec2.internal:8020/user/hive/warehouse/comments_cleaned
OK
Time taken: 7.594 seconds
```

Figure 12: Hive QL query to remove special characters

```
hive> select * from comments_cleaned limit 10;
OK
Julius NM      huh anyway check out this youtube channel kobyoshi02    1
adam riyati    hey guys check out my new channel and our first vid this is the monkeys in the monkey in the white shirtplease leave a like commen
t and please subscribe 1
Evgeny Murashkin      just for test i have to say murdevcom    1
ElNino Melendez me shaking my sexy ass on my channel enjoy    1
GSMega watchvvtargvgvtwq check this out    1
Jason Haddad hey check out my new website this site is about kids stuff kidsmediausa com    1
ferleck ferles subscribe to my channel    1
Bob Kanowski i turned it on mute as soon is i came on i just wanted to check the views    0
Cony you should check my channel for funny videos    1
BeBe Burkey and u should check my channel and tell me what i should do next    1
Time taken: 0.127 seconds, Fetched: 10 row(s)
```

Figure 13: Data after cleaning

- The cleaned dataset is then exported to a CSV file in the S3 bucket for further processing

```
hive> INSERT OVERWRITE LOCAL DIRECTORY 's3://mycloudassignmentbucket/assignmentdata/Youtube_Comments_Cleaned.csv'
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
> SELECT * FROM comments_cleaned;
Query ID = hadoop_20231116004051_fef8587a-867a-4eb7-ad6f-262fe0b309db
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1700089803344_0005)

-----
VERTICES    MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container    SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 01/01 [=====>>] 100% ELAPSED TIME: 9.89 s
-----
OK
Time taken: 10.869 seconds
```

Figure 14: Hive QL query to export data to CSV

## Task 4: Ham and Spam using Hive

- Comments were labelled ham or spam using a bag of words

```
hive> CREATE TABLE spam_candidates AS
> SELECT
>   author,content,
>   CASE
>     WHEN content LIKE '%my channel%' OR
>     content LIKE '%click here%' OR
>     content LIKE '%check%' OR
>     content LIKE '%limited time offer%' OR
>     content LIKE '%promo%' OR
>     content LIKE '%url%' OR
>     content LIKE '%youtube%' OR
>     content LIKE '%free%' OR
>     content LIKE '%amazing%' OR
>     content LIKE '%win%' OR
>     content LIKE '%warning%'
>   THEN 1
>   ELSE 0
>   END AS is_spam
> FROM comments_cleaned;
Query ID = hadoop_20231116213624_7d538236-d5f3-4caf-ba63-2bf3971300b4
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1700163558598_0005)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01 [=====>>] 100%  ELAPSED TIME: 6.84 s
-----
Moving data to directory hdfs://ip-172-31-7-113.ec2.internal:8020/user/hive/warehouse/spam_candidates
OK
Time taken: 15.358 seconds
```

Figure 15: query to use bag of words and classify data as ham or spam

```
hive> select * from spam_candidates limit 10;
OK
Julius NM      huh anyway check out this youtube channel kobyoshi02      1
adam riyati    hey guys check out my new channel and our first vid this is us the  monkeys im the monkey in the white shirtplease leave a like co
mment and please subscribe      1
Evgeny Murashkin      just for test i have to say murdevcom      0
ElMino Melendez me shaking my sexy ass on my channel enjoy      1
GSMega watchvvtargvgvbwq check this out      1
Jason Haddad    hey check out my new website this site is about kids stuff kidsmediausa com      1
ferleck ferles subscribe to my channel      1
Bob Kanowski    i turned it on mute as soon is i came on i just wanted to check the  views      1
Cony           you should check my channel for funny videos      1
BeBe Burkey     and u shouldd check my channel and tell me what i should do next      1
Time taken: 0.24 seconds, Fetched: 10 row(s)
hive> select count(*) from spam_candidates;
OK
1956
Time taken: 0.213 seconds, Fetched: 1 row(s)
```

Figure 16: spam or ham classified data using bag of words



- Dataset was then split into two tables, one containing all the spam comments and the other containing all the ham comments.

```
hive> CREATE TABLE spam_comments AS
> SELECT * FROM spam_candidates
> WHERE is_spam = 1;
Query ID = hadoop_20231116212251_9010363b-d684-4f10-b9b9-ee23191d2f8e
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1700163558598_0004)

-----
VERTICES    MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container    SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 8.17 s
-----
Moving data to directory hdfs://ip-172-31-7-113.ec2.internal:8020/user/hive/warehouse/spam_comments
OK
Time taken: 17.0 seconds
```

Figure 17: Hive QL query to select all spam classified comments and store it in spam\_comments table

```
hive> select * from spam_comments limit 10;
OK
Julius NM      huh anyway check out this youtube channel kobyoshi02      1
adam riyati    hey guys check out my new channel and our first vid this is us the monkeys im the monkey in the white shirtplease leave a like commen
t and please subscribe 1
Evgeny Murashkin      just for test i have to say murdevcom      1
ElNino Melendez me shaking my sexy ass on my channel enjoy      1
G5Mega watchvvtarggygtwq check this out      1
Jason Haddad hey check out my new website this site is about kids stuff kidsmediausa com      1
ferleck ferles subscribe to my channel      1
Cony you should check my channel for funny videos      1
BeBe Burkey and u shouldd check my channel and tell me what i should do next      1
Huckyduck hey subscribe to me      1
Time taken: 0.107 seconds, Fetched: 10 row(s)
```

Figure 18: Data from spam\_comments table

```
hive> CREATE TABLE ham_comments AS
> SELECT * FROM spam_candidates
> WHERE is_spam = 0;
Query ID = hadoop_20231116212418_864a5461-4247-4667-b25e-d6a9404c68a4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1700163558598_0004)

-----
VERTICES    MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container    SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 6.47 s
-----
Moving data to directory hdfs://ip-172-31-7-113.ec2.internal:8020/user/hive/warehouse/ham_comments
OK
Time taken: 7.09 seconds
```

Figure 19: HiveQL query to select all ham comments and store it in ham\_comments table

```
hive> select * from ham_comments limit 10;
OK
Bob Kanowski      i turned it on mute as soon is i came on i just wanted to check the  views      0
Zielimeek21       im only checking the views      0
zhichao wang      i think about 100 millions of the views come from people who only wanted to  check the views      0
Owen Lai          just checking the views 0
Brandon Pryor      i dont even watch it anymore i just come here to check on 2 billion or not      0
DropShotsk8r      came here to check the views goodbye      0
Tasha Lucius      2 billioncoming soon      0
Jason Provencal    why dafuq is a korean song so big in the usa does that mean we support  koreans last time i checked they wanted to bomb us      0
crestpee          now its 1884034783 views pls comment the view count the next hour p      0
Phuc Ly go here to check the views 3      0
Time taken: 0.178 seconds, Fetched: 10 row(s)
```

Figure 20: Data from ham\_comments table

- The spam\_comments table was queried to find the accounts that generated the highest number of spam comments.

```
hive> SELECT author, count(*)
> FROM spam_comments
> GROUP BY author
> ORDER BY count(*) DESC
> LIMIT 10;
Query ID = hadoop_20231116005347_291beead-33e1-41be-834f-341dd20de9e1
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1700089803344_0006)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1 .....	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2 .....	container	SUCCEEDED	2	2	0	0	0	0
Reducer 3 .....	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 03/03 [=====]>>] 100% ELAPSED TIME: 7.26 s
-----
OK
M.E.S      8
Shadrach Grentz 7
Louis Bryant 7
DanteBTV 6
Hidden Love 5
LuckyMusiqLive 5
Derek Moya 5
Alldailyvines 4
Laura Brown 4
Scott Johnson 4
Time taken: 8.224 seconds, Fetched: 10 row(s)
```

Figure 21: HiveQL query to find top 10 spam accounts and the output generated

- The ham\_comments table was queried to find the accounts that generated the highest number of ham comments.

```
hive> SELECT author, count(*)
> FROM ham_comments
> GROUP BY author
> ORDER BY count(*) DESC
> LIMIT 10;
Query ID = hadoop_20231116005520_8777893c-54c9-4569-a4cf-1a5babf6de90
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1700089803344_0006)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1 .....	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2 .....	container	SUCCEEDED	2	2	0	0	0	0
Reducer 3 .....	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 03/03 [=====>>] 100% ELAPSED TIME: 7.51 s
OK
5000palo 7
Marshmallow Kingdom 3
Seth Ryan 3
Sonny Carter 2
Alain Bruno 2
Brian Brai 2
Paul Crowder 2
Pepe The Meme King 2
janet rangel 2
Eric Gonzalez 2
Time taken: 7.985 seconds, Fetched: 10 row(s)
```

Figure 22: HiveQL query to find top 10 ham accounts and the output

## References

[1] Youtube SPAM CLASSIFIED-COMMENTS, Goneee

<https://www.kaggle.com/datasets/goneee/youtube-spam-classifiedcomments>

[2] How to Write a Hive SQL Query to Create Tables & Query Data?

<https://www.projectpro.io/recipes/explain-write-sql-code-create-hive-table-query-data>

[3] Hive Tutorial: Working with Data in Hadoop

<https://www.simplilearn.com/tutorials/hadoop-tutorial/hive>

# Appendix

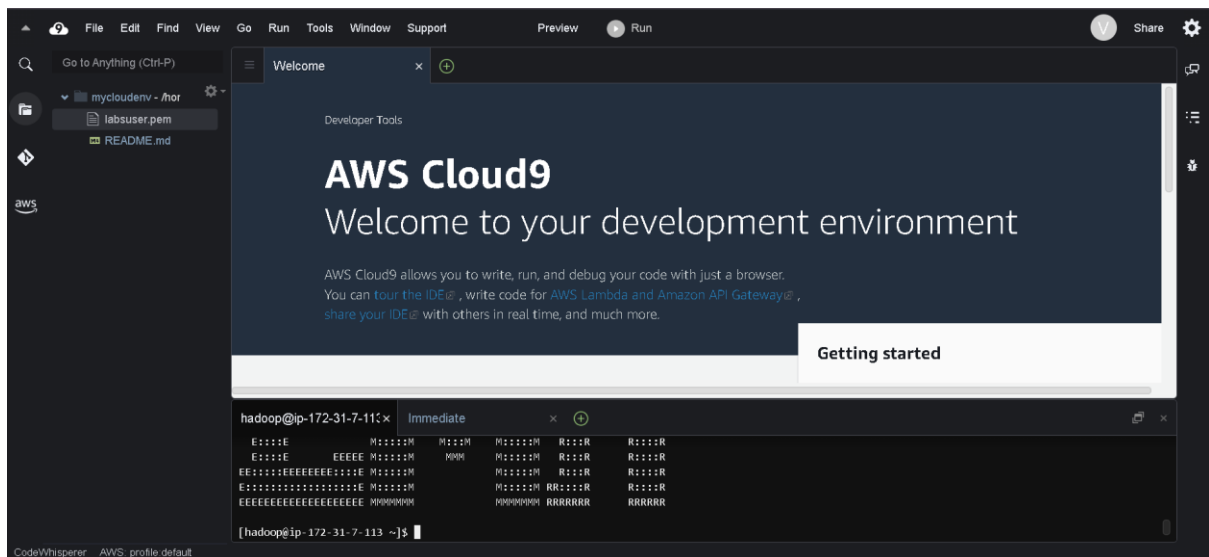


Figure 23: Cloud9 instance

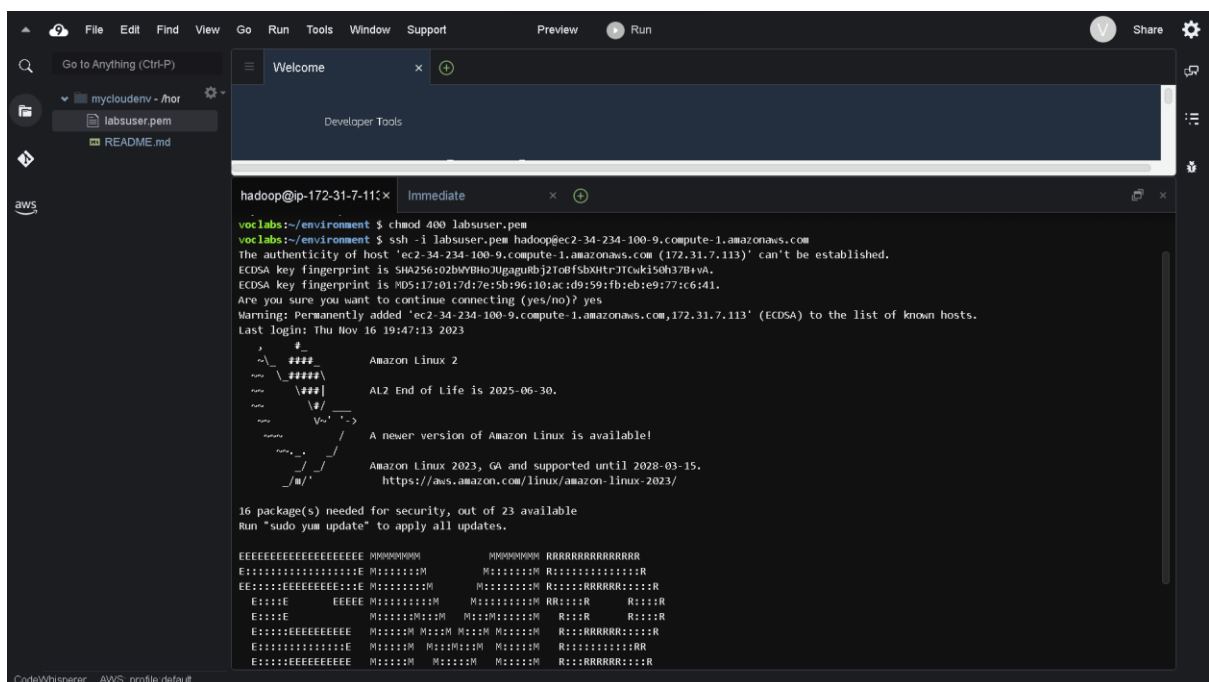
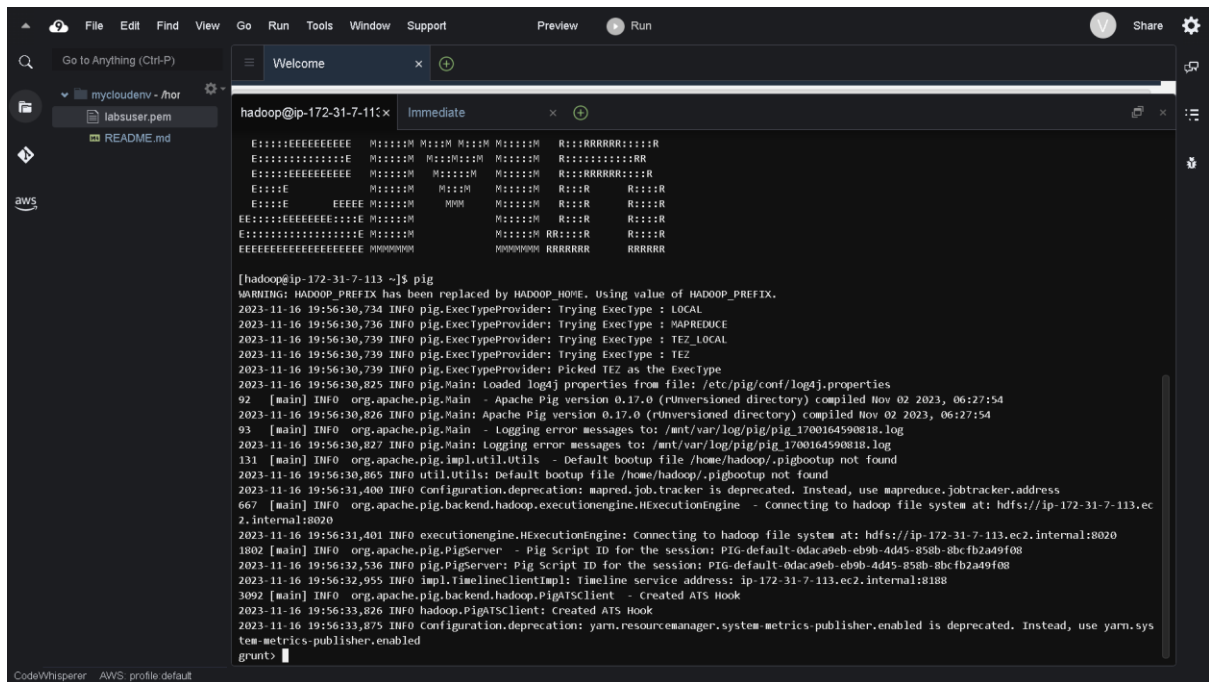


Figure 24: EMR startup



```
hadoop@ip-172-31-7-113:~$ pig
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2023-11-16 19:56:30,734 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2023-11-16 19:56:30,736 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2023-11-16 19:56:30,739 INFO pig.ExecTypeProvider: Trying ExecType : TEZ_LOCAL
2023-11-16 19:56:30,739 INFO pig.ExecTypeProvider: Trying ExecType : TEZ
2023-11-16 19:56:30,739 INFO pig.ExecTypeProvider: Picked TEZ as the ExecType
2023-11-16 19:56:30,825 INFO pig.Main: Loaded log4j properties from file: /etc/pig/conf/log4j.properties
92 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r/unversioned directory) compiled Nov 02 2023, 06:27:54
93 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r/unversioned directory) compiled Nov 02 2023, 06:27:54
93 [main] INFO org.apache.pig.Main - Logging error messages to: /mnt/var/log/pig/pig_1700164590818.log
2023-11-16 19:56:30,827 INFO pig.Main: Logging error messages to: /mnt/var/log/pig/pig_1700164590818.log
131 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/.pigbootstrap not found
2023-11-16 19:56:30,865 INFO util.Utils: Default bootstrap file /home/hadoop/.pigbootstrap not found
2023-11-16 19:56:31,480 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
667 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - connecting to hadoop file system at: hdfs://ip-172-31-7-113.ec2.internal:8020
2023-11-16 19:56:31,401 INFO executionengine.HExecutionEngine: Connecting to hadoop file system at: hdfs://ip-172-31-7-113.ec2.internal:8020
1802 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-0daca9eb-eb9b-4d45-858b-8bcfb2a49f08
2023-11-16 19:56:32,536 INFO pig.PigServer: Pig Script ID for the session: PIG-default-0daca9eb-eb9b-4d45-858b-8bcfb2a49f08
2023-11-16 19:56:32,955 INFO impl.TimelineClientImpl: Timeline service address: ip-172-31-7-113.ec2.internal:8188
3092 [main] INFO org.apache.pig.backend.hadoop.PigATSClient - Created ATS Hook
2023-11-16 19:56:33,826 INFO hadoop.PigATSClient: Created ATS Hook
2023-11-16 19:56:33,875 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt>
```

Figure 25: Verification of pig installation