1.
2. select count(id) from Log;
   9082213

```
bda_assignment1=# select count(id) from Log;
  count
---------
 9082213
(1 row)
```

3. select count(id),logging_level from Log where logging_level like 'WARN' group by (logging_level);
   123571

```
bda_assignment1=# select count(id),logging_level from Log where logging_level li
ke 'WARN' group by (logging_level);
 count  | logging_level
--------+---------------
 123571 | WARN
(1 row)
```

4. select count(*) from (select distinct (regexp_matches(description,'api.github.com/[a-zA-Z0-9-]+/[a-zA-Z0-9-]+/')) as url from Log) as temp ;

   55310

```
bda_assignment1=# select count(*) from (select distinct (regexp_matches(descript
ion,'api.github.com/[a-zA-Z0-9-]+/[a-zA-Z0-9-]+/')) as  url from Log) as temp ;
 count
-------
 55310
(1 row)
```

5. select downloader_id,count(*) from Log where description like '%http:%' or description like '%https:%' group by downloader_id ORDER BY COUNT(*) DESC limit 10;

```
 downloader_id | count
---------------+-------
            13 | 80391
             4 | 17905
            18 | 17847
            10 | 17787
            40 | 17737
            38 | 17479
            47 | 17473
            39 | 17467
            25 | 17353
```

1 | 17324

```
bda_assignment1=# select downloader_id,count(*) from Log where description like
'%http:%' or description like '%https:%' group by downloader_id ORDER BY COUNT(*
) DESC limit 10;
 downloader_id | count
---------------+-------
            13 | 80391
             4 | 17905
            18 | 17847
            10 | 17787
            40 | 17737
            38 | 17479
            47 | 17473
            39 | 17467
            25 | 17353
             1 | 17324
(10 rows)
```

6. select downloader_id,count(*) from Log where description like '%http:%' or description like '%https:%' and description like 'Failed%' group by downloader_id ORDER BY COUNT(*) DESC limit 10;

```
bda_assignment1=# select downloader_id,count(*) from Log where description like
'%http:%' or description like '%https:%' and description like 'Failed%' group by
 downloader_id ORDER BY COUNT(*) DESC limit 10;
 downloader_id | count
---------------+-------
            13 | 74838
            21 |  1301
            40 |  1045
             9 |   344
            42 |   343
            18 |   340
             4 |   329
            25 |   321
             6 |   316
            10 |   314
(10 rows)
```

7. select count(*),temp.time from (select extract(hour from time) as time from Log) as temp group by temp.time order by count(*) DESC limit 1;

   2500754

```
bda_assignment1=# select count(*),temp.time from (select extract(hour from time
) as time from Log) as temp group by temp.time order by count(*) DESC limit 1;
  count   | time
---------+------
 2500754 |   10
(1 row)
```

8. select count(*),substring(cast(temp.url as text),22) from (select (regexp_matches(description,'api.github.com/repos/[a-zA-Z0-9-]+/[a-zA-Z0-9-]+/')) as url from Log) as temp where array_length(temp.url,1) > 0 group by temp.url order by count(*) DESC limit 1;

74744 | /greatfakeman/Tabchi/}

```
bda_assignment1=# select count(*),substring(cast(temp.url as text),22) from (sel
ect (regexp_matches(description,'api.github.com/repos/[a-zA-Z0-9-]+/[a-zA-Z0-9-]
+/')) as  url from Log) as temp where array_length(temp.url,1) > 0 group by  tem
p.url order by count(*) DESC limit 1;
 count |       substring
-------+------------------------
 74744 | /greatfakeman/Tabchi/}
(1 row)
```

9. select count(*),substring(cast(temp.access as text),11,11) from (select (regexp_matches(description,'Access: [0-9a-zA-Z]+,')) as access from Log where description like 'Failed%') as temp group by temp.access order by count(*) DESC limit 10;

```
 count |  substring
-------+-------------
 74838 | ac6168f8776
  1263 | 46f11b5791b
  1045 | 9115020fb01
   350 | c1240f63b5b
   343 | 8993d227f49
   340 | 2776f3ba0a5
   340 | bd72b2479f9
   329 | 5c7cf6cbe46
   319 | f0580432e57
   315 | 78e09e3bff3
(10 rows)
```

```
bda_assignment1=# select count(*),substring(cast(temp.access as text),11,11) fro
m (select (regexp_matches(description,'Access: [0-9a-zA-Z]+,')) as  access from
Log where description like 'Failed%') as temp group by temp.access order by coun
t(*) DESC limit 10;
 count |  substring
-------+-------------
 74838 | ac6168f8776
  1263 | 46f11b5791b
  1045 | 9115020fb01
   350 | c1240f63b5b
   343 | 8993d227f49
   340 | 2776f3ba0a5
   340 | bd72b2479f9
   329 | 5c7cf6cbe46
   319 | f0580432e57
   315 | 78e09e3bff3
(10 rows)
```

10. CREATE INDEX downloader_id_index ON Log (downloader_id);
    select count(temp.url) from (select downloader_id,(regexp_matches(description,'api.github.com/[a-zA-Z0-9-]+/[a-zA-Z0-9-]+/')) as url from Log) as temp where temp.downloader_id = 22 ;

    16500

```
bda_assignment1=# select count(temp.url) from (select downloader_id,(regexp_matc
hes(description,'api.github.com/[a-zA-Z0-9-]+/[a-zA-Z0-9-]+/')) as  url from Log
) as temp where temp.downloader_id = 22 ;
 count
-------
 16500
(1 row)
```

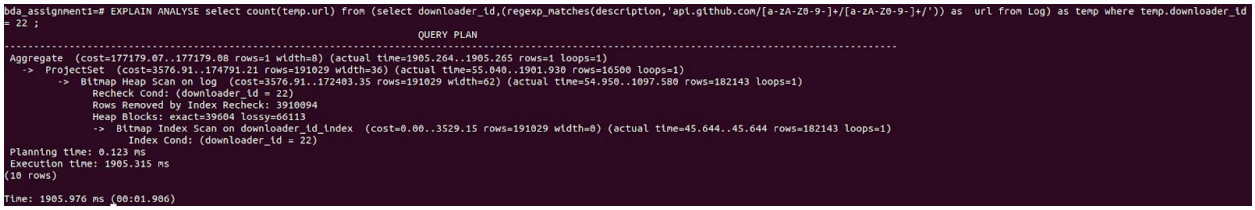
Time taken was measured using EXPLAIN ANALYSE option of POSTGRESQL

```
                                                              QUERY PLAN
----------------------------------------------------------------------------------------------------------------------------------------------------
 Aggregate  (cost=177179.07..177179.08 rows=1 width=8) (actual time=1839.348..1839.348 rows=1 loops=1)
   ->  ProjectSet  (cost=3576.91..174791.21 rows=191029 width=36) (actual time=47.712..1836.054 rows=16500 loops=1)
         ->  Bitmap Heap Scan on log  (cost=3576.91..172403.35 rows=191029 width=62) (actual time=47.584..1064.116 rows=182143 loops=1)
               Recheck Cond: (downloader_id = 22)
               Rows Removed by Index Recheck: 3910094
               Heap Blocks: exact=39604 lossy=66113
               ->  Bitmap Index Scan on downloader_id_index  (cost=0.00..3529.15 rows=191029 width=0) (actual time=30.453..30.453 rows=182143 loops=1)
                     Index Cond: (downloader_id = 22)
 Planning time: 0.106 ms
 Execution time: 1839.388 ms
(10 rows)

Time: 1839.889 ms (00:01.840)
bda_assignment1=#
```

11. DROP INDEX downloader_id_index
    select count(temp.url) from (select downloader_id,(regexp_matches(description,'api.github.com/[a-zA-Z0-9-]+/[a-zA-Z0-9-]+/')) as url from Log) as temp where temp.downloader_id = 22 ;

    16500

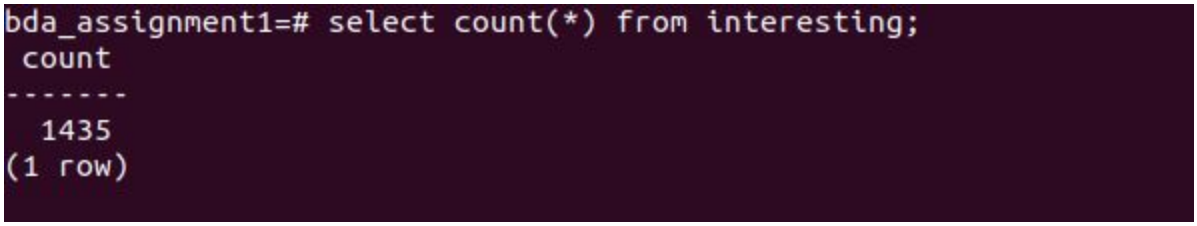


12. COPY interesting FROM
    '/home/gyaneshanand/Desktop/BDA/Assignments/Assignment1/important-repos.csv'
    DELIMITERS ',' csv;

    COPY interesting FROM '/home/suyash-singh/Documents/BDA/important-repos.csv'
    DELIMITERS ',' csv;

    CREATE TABLE interesting(id INT ,URL TEXT, OwnerId INT , Name TEXT , Language TEXT, created_at TIMESTAMP,forked_from INT, deleted INT ,updated_at TIMESTAMP);

    1435



13. select count(newfile.id) from (select temp.id,cast(split_part(substring(cast(temp.url as text),22),'/}',1) as text) as account from (select id,(regexp_matches(description,'api.github.com/repos/[a-zA-Z0-9\-.]+/[a-zA-Z0-9\-.]+/')) as url from Log) as temp where array_length(temp.url,1) > 0) as newfile where EXISTS (select id from interesting where interesting.url like '%' || newfile.account || '%');

18278

```
bda_assignment1=# select count(newfile.id) from (select temp.id,cast(split_part(
substring(cast(temp.url as text),22),'/}',1) as text ) as account from (select i
d,(regexp_matches(description,'api.github.com/repos/[a-zA-Z0-9\-.]+/[a-zA-Z0-9\-
.]+/')) as  url from Log) as temp where array_length(temp.url,1) > 0) as newfile
 where EXISTS (select  id from interesting where interesting.url like '%' || new
file.account || '%' );
 count
-------
 18278
(1 row)

Time: 1699360.222 ms (28:19.360)
```

14. select newfile.account,count(*) from (select temp.id,temp.description,cast(split_part(substring(cast(temp.url as text),22),'/}',1) as text) as account from (select id,description,(regexp_matches(description,'api.github.com/repos/[a-zA-Z0-9\-.]+/[a-zA-Z0-9\-.]+/')) as url from Log) as temp where array_length(temp.url,1) > 0) as newfile where newfile.description like 'Failed%'and EXISTS (select id from interesting where interesting.url like '%' || newfile.account || '%') group by newfile.account order by count(*) DESC limit 10;

```
bda_assignment1=# select newfile.account,count(*) from (select temp.id,temp.desc
ription,cast(split_part(substring(cast(temp.url as text),22),'/}',1) as text ) a
s account from (select id,description,(regexp_matches(description,'api.github.co
m/repos/[a-zA-Z0-9\-.]+/[a-zA-Z0-9\-.]+/')) as  url from Log) as temp where arra
y_length(temp.url,1) > 0) as newfile where newfile.description like 'Failed%'and
 EXISTS (select  id from interesting where interesting.url like '%' || newfile.a
ccount || '%' ) group by newfile.account order by count(*) DESC limit 10;
                        account                         | count
--------------------------------------------------------+-------
 /asmagin/sitecore-foundation-codegeneration-composition |     5
 /wireapp/wire-ios                                      |     3
 /Jmercier13/jobgenerator                               |     2
 /couchbaselabs/mobile-training-todo                    |     2
 /biggerlion/LockTest                                   |     2
 /georkom/AndoPrime                                     |     2
 /GoogleCloudPlatform/google-cloud-intellij             |     2
 /baharboutique/baharboutique.github.io                 |     2
 /InsidePointSistemas/MeuProjeto                        |     2
 /ipfs/js-ipfs                                          |     2
(10 rows)
```

With newfile as (select temp.id,cast(split_part(substring(cast(temp.url as text),22),'/}',1) as text) as account from (select id,(regexp_matches(descreption,'api.github.com/repos/[a-zA-Z0-9\-.]+/[a-zA-Z0-9\-.]+/')) as url

```
from Log) as temp where array_length(temp.url,1) > 0) select count(newfile.id) from newfile
where EXISTS (select  id from interesting where interesting.url like '%' || newfile.account || '%' );
```

```
with newfile as (select temp.id,temp.descreption,cast(split_part(substring(cast(temp.url as
text),22),'/}',1) as text ) as account from (select
id,descreption,(regexp_matches(descreption,'api.github.com/repos/[a-zA-Z0-9\-.]+/[a-zA-Z0-9\-.]
+/')) as  url from Log) as temp where array_length(temp.url,1) > 0) select
newfile.account,count(*) from newfile where newfile.descreption like 'Failed%'and EXISTS
(select  id from interesting where interesting.url like '%' || newfile.account || '%' ) group by
newfile.account order by count(*) DESC limit 5;
```