CASSANDRA

Code to create streaming keyspace:

CREATE KEYSPACE streaming
WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };

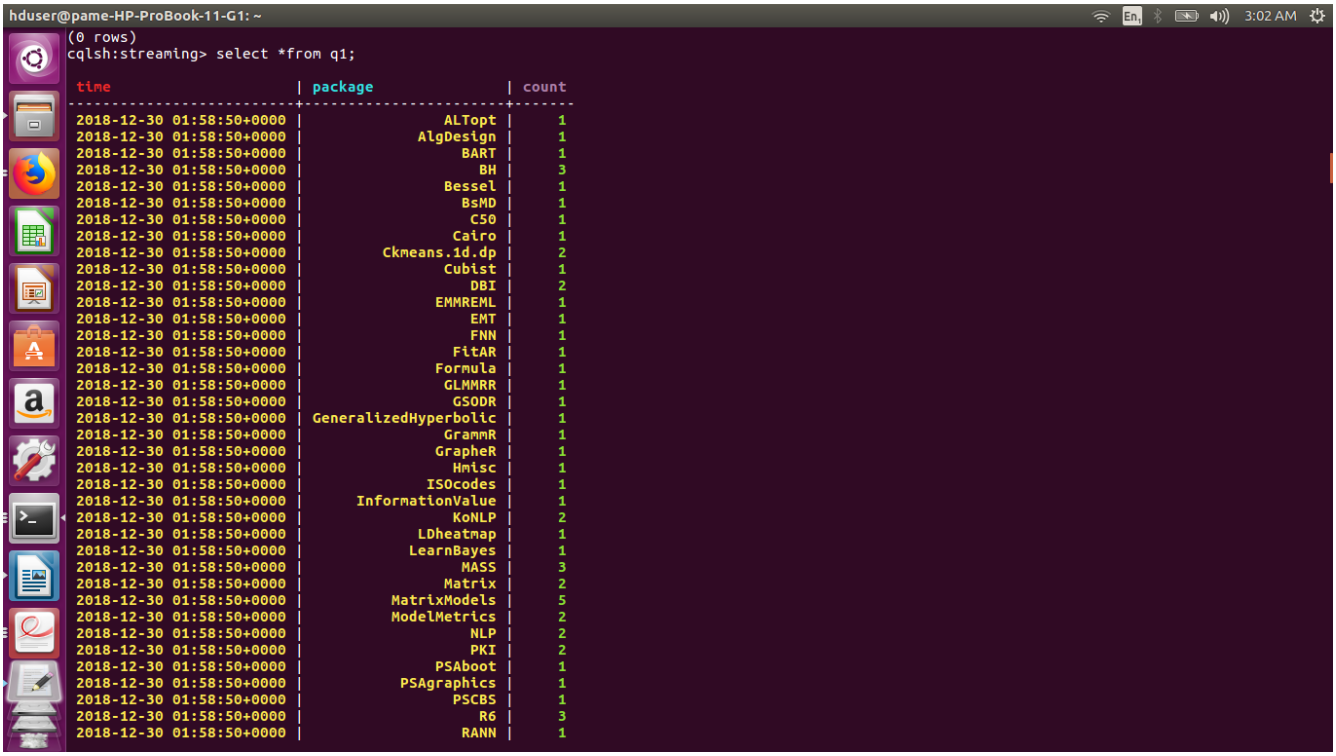
Streaming Computations:

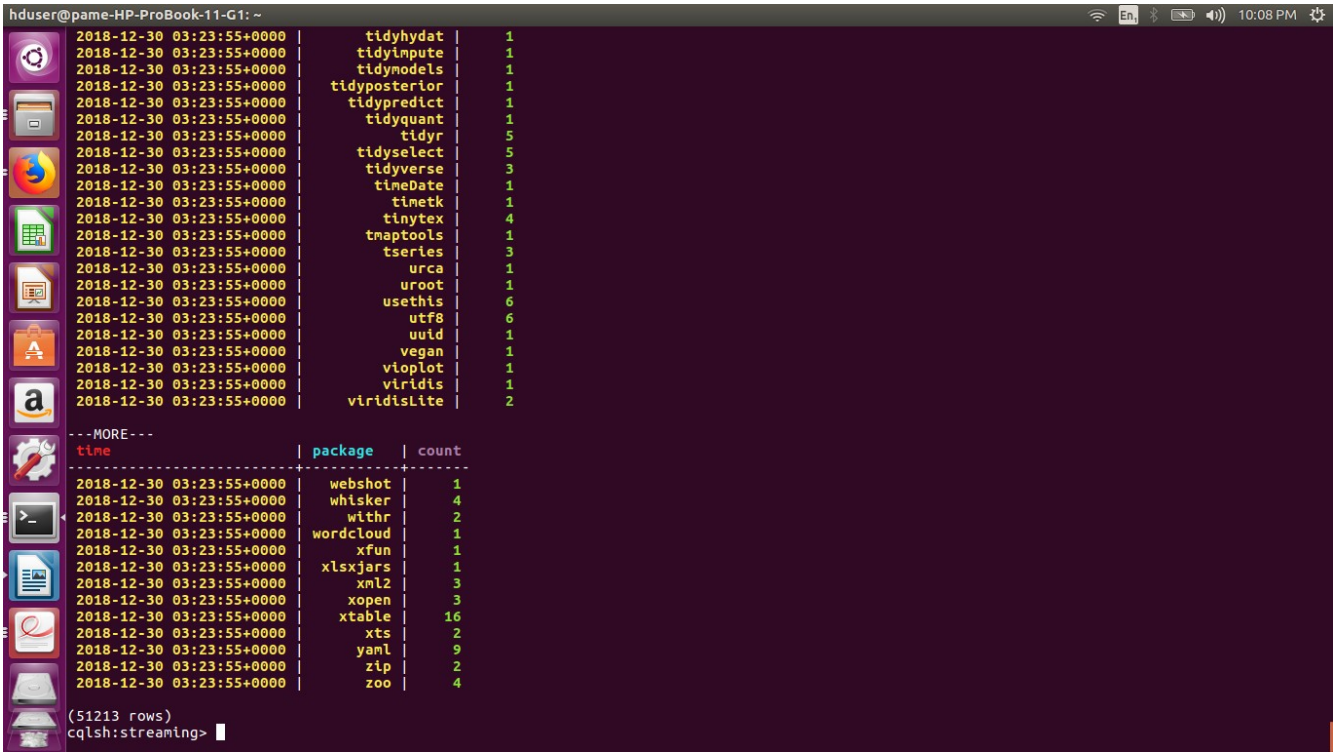
Q1. To calculate the number of downloads of each package.

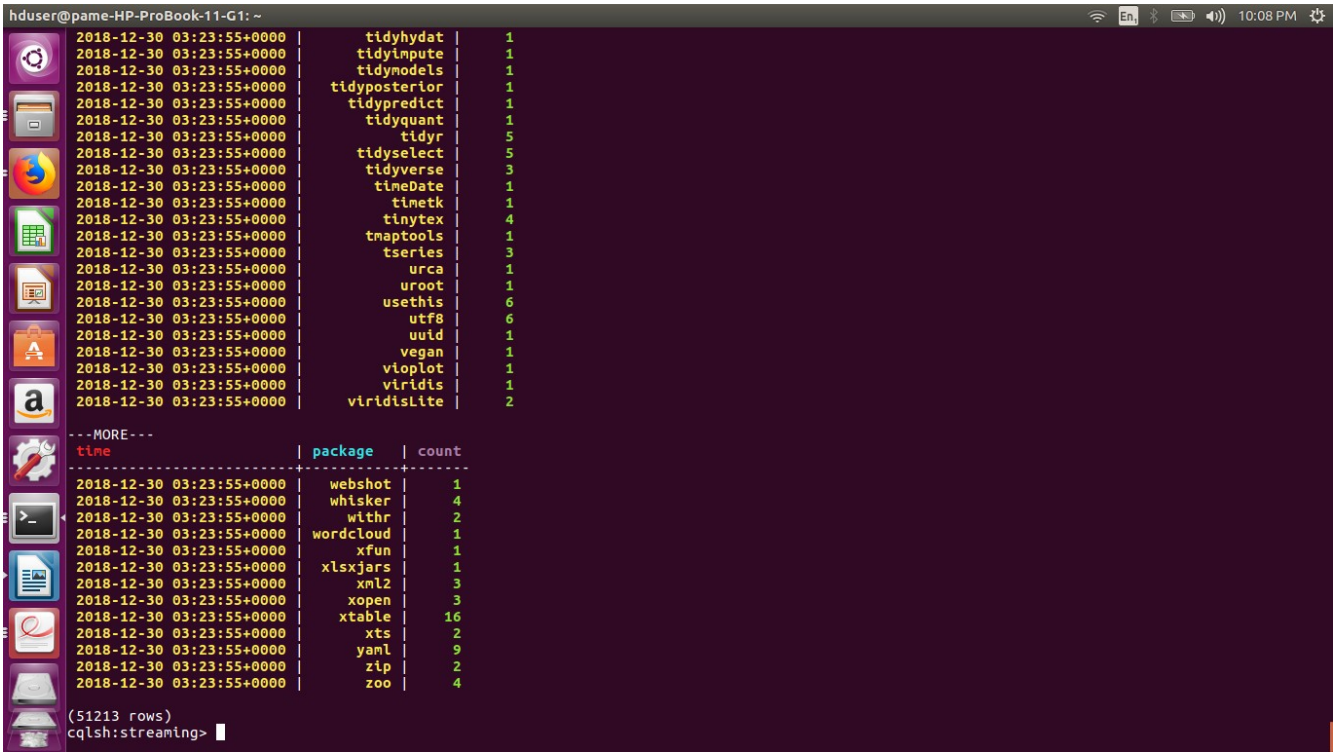
CQL Code Table:

:cqlsh:streaming> CREATE TABLE q1 (time text, package text, count int, PRIMARY KEY(time, package));

ScreenShots:




hduser@pame-HP-ProBook-11-G1: ~
10:08 PM
2018-12-30 03:23:55+0000 | tidyhydat | 1
2018-12-30 03:23:55+0000 | tidyimpute | 1
2018-12-30 03:23:55+0000 | tidymodels | 1
2018-12-30 03:23:55+0000 | tidyposterior | 1
2018-12-30 03:23:55+0000 | tidypredict | 1
2018-12-30 03:23:55+0000 | tidyquant | 1
2018-12-30 03:23:55+0000 | tidyr | 5
2018-12-30 03:23:55+0000 | tidyselect | 5
2018-12-30 03:23:55+0000 | tidyverse | 3
2018-12-30 03:23:55+0000 | timeDate | 1
2018-12-30 03:23:55+0000 | timetk | 1
2018-12-30 03:23:55+0000 | tinytex | 4
2018-12-30 03:23:55+0000 | tmaptools | 1
2018-12-30 03:23:55+0000 | tseries | 3
2018-12-30 03:23:55+0000 | urca | 1
2018-12-30 03:23:55+0000 | uroot | 1
2018-12-30 03:23:55+0000 | usethis | 6
2018-12-30 03:23:55+0000 | utf8 | 6
2018-12-30 03:23:55+0000 | uuid | 1
2018-12-30 03:23:55+0000 | vegan | 1
2018-12-30 03:23:55+0000 | vioplot | 1
2018-12-30 03:23:55+0000 | viridis | 1
2018-12-30 03:23:55+0000 | viridisLite | 2
---MORE---
time | package | count
2018-12-30 03:23:55+0000 | webshot | 1
2018-12-30 03:23:55+0000 | whisker | 4
2018-12-30 03:23:55+0000 | withr | 2
2018-12-30 03:23:55+0000 | wordcloud | 1
2018-12-30 03:23:55+0000 | xfun | 1
2018-12-30 03:23:55+0000 | xlsxjars | 1
2018-12-30 03:23:55+0000 | xml2 | 3
2018-12-30 03:23:55+0000 | xopen | 3
2018-12-30 03:23:55+0000 | xtable | 16
2018-12-30 03:23:55+0000 | xts | 2
2018-12-30 03:23:55+0000 | yaml | 9
2018-12-30 03:23:55+0000 | zip | 2
2018-12-30 03:23:55+0000 | zoo | 4
(51213 rows)
cqlsh:streaming>

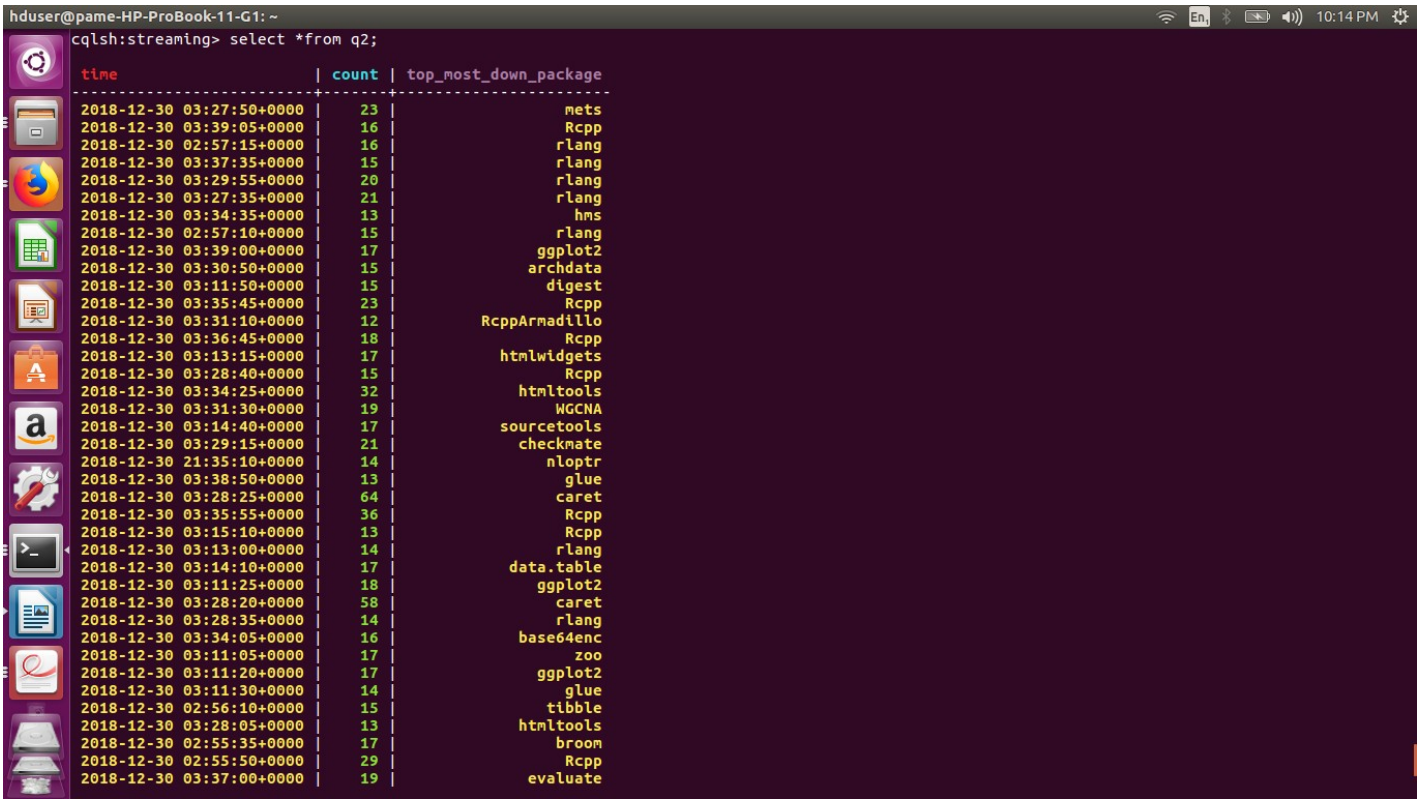

hduser@pame-HP-ProBook-11-G1: ~
10:08 PM
2018-12-30 03:23:55+0000 | tidyhydat | 1
2018-12-30 03:23:55+0000 | tidyimpute | 1
2018-12-30 03:23:55+0000 | tidymodels | 1
2018-12-30 03:23:55+0000 | tidyposterior | 1
2018-12-30 03:23:55+0000 | tidypredict | 1
2018-12-30 03:23:55+0000 | tidyquant | 1
2018-12-30 03:23:55+0000 | tidyr | 5
2018-12-30 03:23:55+0000 | tidyselect | 5
2018-12-30 03:23:55+0000 | tidyverse | 3
2018-12-30 03:23:55+0000 | timeDate | 1
2018-12-30 03:23:55+0000 | timetk | 1
2018-12-30 03:23:55+0000 | tinytex | 4
2018-12-30 03:23:55+0000 | tmaptools | 1
2018-12-30 03:23:55+0000 | tseries | 3
2018-12-30 03:23:55+0000 | urca | 1
2018-12-30 03:23:55+0000 | uroot | 1
2018-12-30 03:23:55+0000 | usethis | 6
2018-12-30 03:23:55+0000 | utf8 | 6
2018-12-30 03:23:55+0000 | uuid | 1
2018-12-30 03:23:55+0000 | vegan | 1
2018-12-30 03:23:55+0000 | vioplot | 1
2018-12-30 03:23:55+0000 | viridis | 1
2018-12-30 03:23:55+0000 | viridisLite | 2
---MORE---
time | package | count
2018-12-30 03:23:55+0000 | webshot | 1
2018-12-30 03:23:55+0000 | whisker | 4
2018-12-30 03:23:55+0000 | withr | 2
2018-12-30 03:23:55+0000 | wordcloud | 1
2018-12-30 03:23:55+0000 | xfun | 1
2018-12-30 03:23:55+0000 | xlsxjars | 1
2018-12-30 03:23:55+0000 | xml2 | 3
2018-12-30 03:23:55+0000 | xopen | 3
2018-12-30 03:23:55+0000 | xtable | 16
2018-12-30 03:23:55+0000 | xts | 2
2018-12-30 03:23:55+0000 | yaml | 9
2018-12-30 03:23:55+0000 | zip | 2
2018-12-30 03:23:55+0000 | zoo | 4
(51213 rows)
cqlsh:streaming>

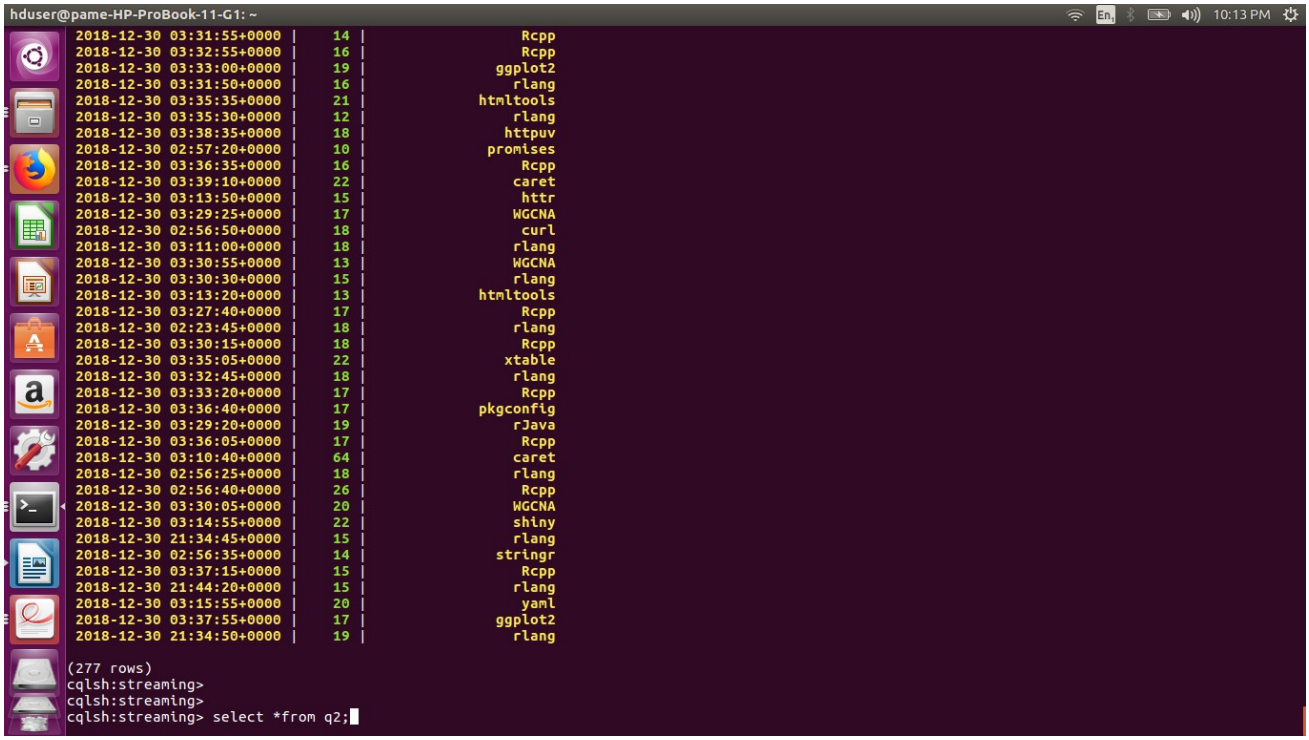
Q2.To find the top most downloaded package.

CQL Code Table:

:cqlsh:streaming> CREATE TABLE q2 (time text, top_most_down_package text, count int, PRIMARY KEY(time, count));

ScreenShots:



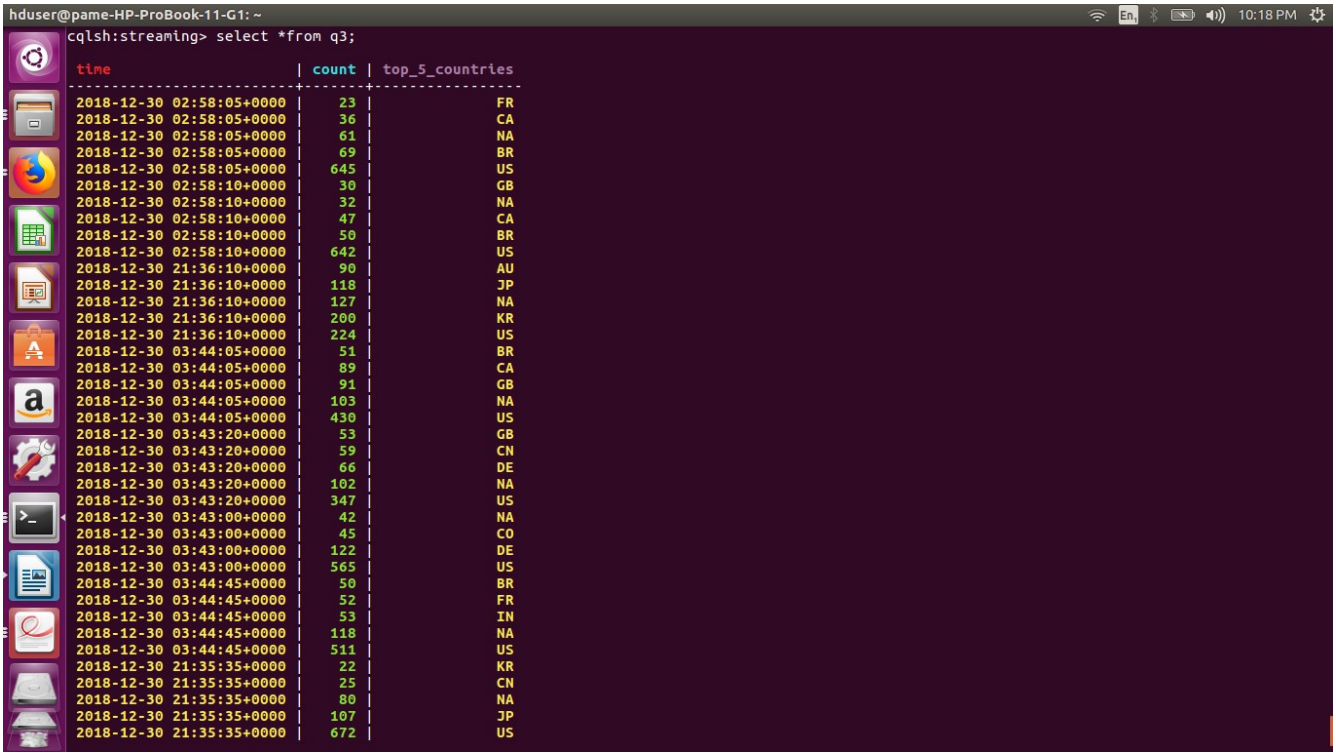


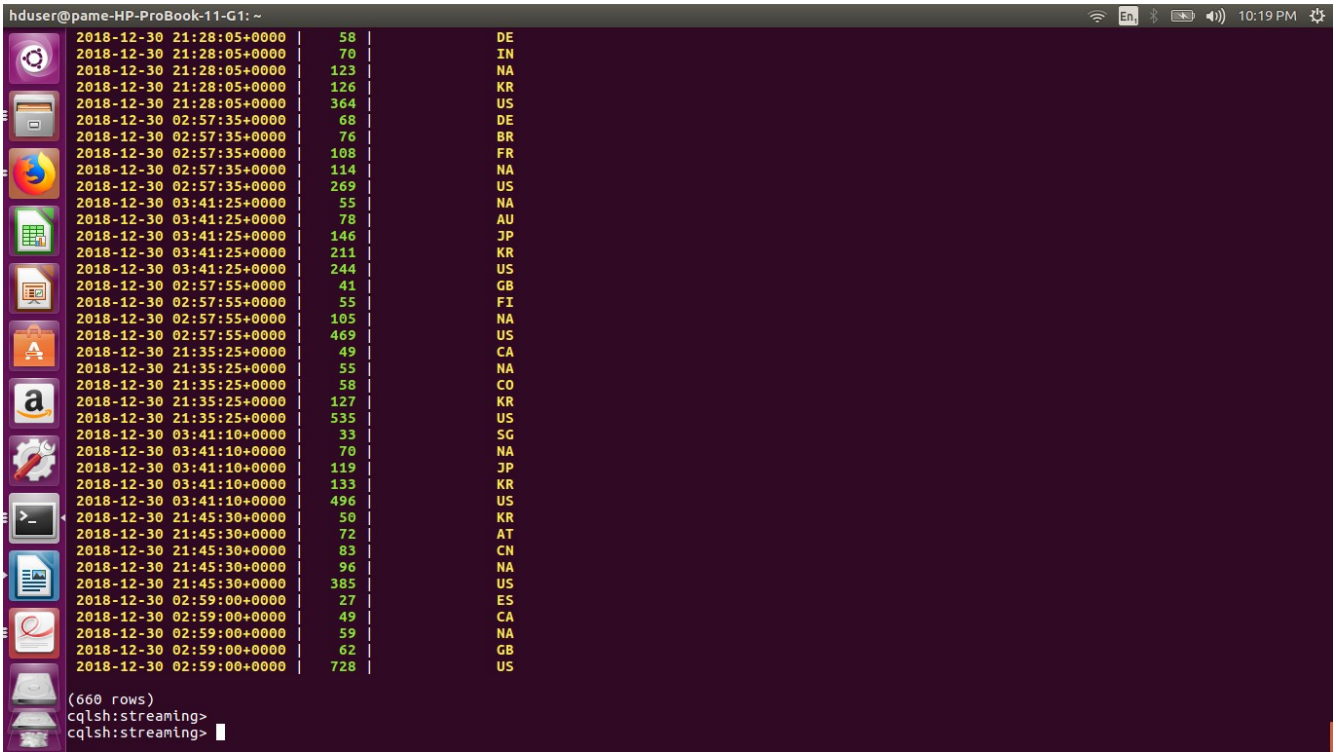
Q3. To find the top 5 countries along with number of downloads.

CQL Code Table:

:cqlsh:streaming> CREATE TABLE q3 (time text, top_5_countries text, count int, PRIMARY KEY(time, count));

ScreenShots:


hduser@pame-HP-ProBook-11-G1: ~
10:18 PM
cqlsh:streaming> select *from q3;
time | count | top_5_countries
2018-12-30 02:58:05+0000 | 23 | FR
2018-12-30 02:58:05+0000 | 36 | CA
2018-12-30 02:58:05+0000 | 61 | NA
2018-12-30 02:58:05+0000 | 69 | BR
2018-12-30 02:58:05+0000 | 645 | US
2018-12-30 02:58:10+0000 | 30 | GB
2018-12-30 02:58:10+0000 | 32 | NA
2018-12-30 02:58:10+0000 | 47 | CA
2018-12-30 02:58:10+0000 | 50 | BR
2018-12-30 02:58:10+0000 | 642 | US
2018-12-30 21:36:10+0000 | 90 | AU
2018-12-30 21:36:10+0000 | 118 | JP
2018-12-30 21:36:10+0000 | 127 | NA
2018-12-30 21:36:10+0000 | 200 | KR
2018-12-30 21:36:10+0000 | 224 | US
2018-12-30 03:44:05+0000 | 51 | BR
2018-12-30 03:44:05+0000 | 89 | CA
2018-12-30 03:44:05+0000 | 91 | GB
2018-12-30 03:44:05+0000 | 103 | NA
2018-12-30 03:44:05+0000 | 430 | US
2018-12-30 03:43:20+0000 | 53 | GB
2018-12-30 03:43:20+0000 | 59 | CN
2018-12-30 03:43:20+0000 | 66 | DE
2018-12-30 03:43:20+0000 | 102 | NA
2018-12-30 03:43:20+0000 | 347 | US
2018-12-30 03:43:00+0000 | 42 | NA
2018-12-30 03:43:00+0000 | 45 | CO
2018-12-30 03:43:00+0000 | 122 | DE
2018-12-30 03:43:00+0000 | 565 | US
2018-12-30 03:44:45+0000 | 50 | BR
2018-12-30 03:44:45+0000 | 52 | FR
2018-12-30 03:44:45+0000 | 53 | IN
2018-12-30 03:44:45+0000 | 118 | NA
2018-12-30 03:44:45+0000 | 511 | US
2018-12-30 21:35:35+0000 | 22 | KR
2018-12-30 21:35:35+0000 | 25 | CN
2018-12-30 21:35:35+0000 | 80 | NA
2018-12-30 21:35:35+0000 | 107 | JP
2018-12-30 21:35:35+0000 | 672 | US

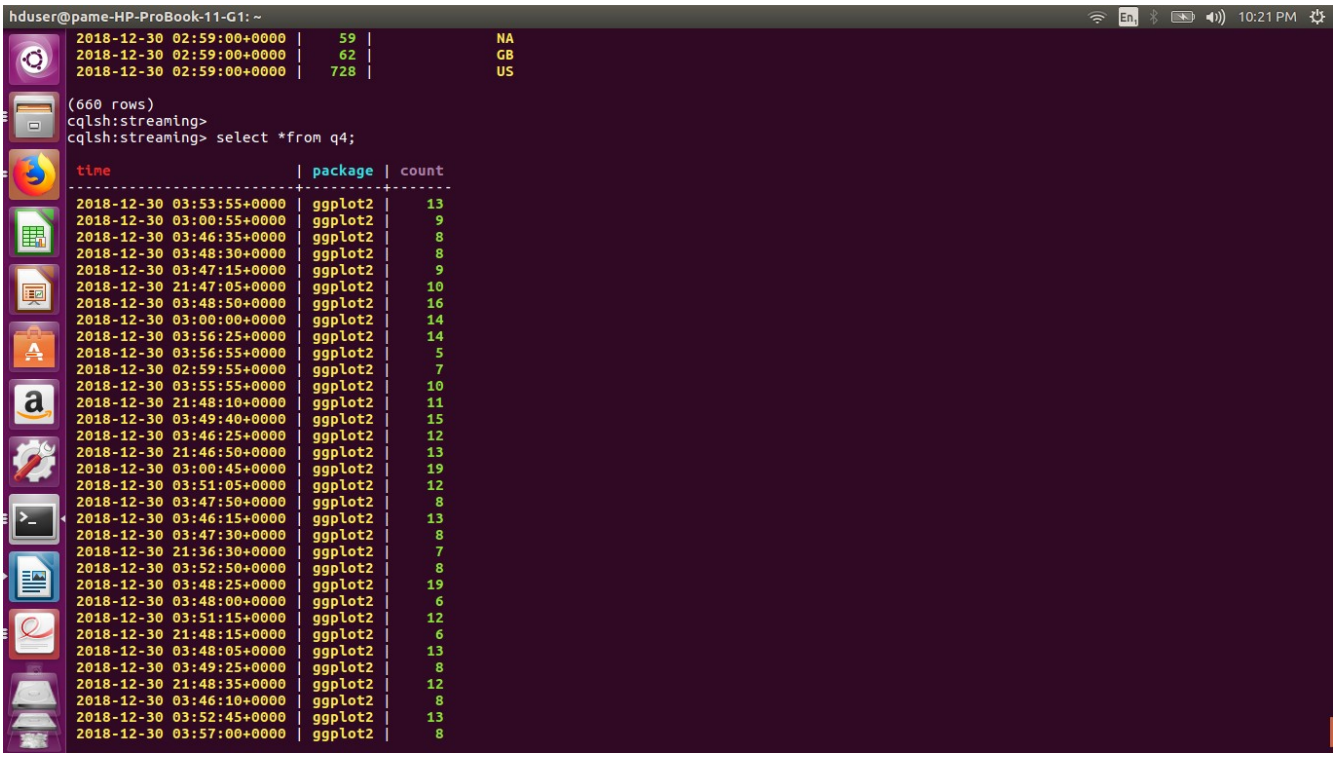

hduser@pame-HP-ProBook-11-G1: ~
10:19 PM
2018-12-30 21:28:05+0000 | 58 | DE
2018-12-30 21:28:05+0000 | 70 | IN
2018-12-30 21:28:05+0000 | 123 | NA
2018-12-30 21:28:05+0000 | 126 | KR
2018-12-30 21:28:05+0000 | 364 | US
2018-12-30 02:57:35+0000 | 68 | DE
2018-12-30 02:57:35+0000 | 76 | BR
2018-12-30 02:57:35+0000 | 108 | FR
2018-12-30 02:57:35+0000 | 114 | NA
2018-12-30 02:57:35+0000 | 269 | US
2018-12-30 03:41:25+0000 | 55 | NA
2018-12-30 03:41:25+0000 | 78 | AU
2018-12-30 03:41:25+0000 | 146 | JP
2018-12-30 03:41:25+0000 | 211 | KR
2018-12-30 03:41:25+0000 | 244 | US
2018-12-30 02:57:55+0000 | 41 | GB
2018-12-30 02:57:55+0000 | 55 | FI
2018-12-30 02:57:55+0000 | 105 | NA
2018-12-30 02:57:55+0000 | 469 | US
2018-12-30 21:35:25+0000 | 49 | CA
2018-12-30 21:35:25+0000 | 55 | NA
2018-12-30 21:35:25+0000 | 58 | CO
2018-12-30 21:35:25+0000 | 127 | KR
2018-12-30 21:35:25+0000 | 535 | US
2018-12-30 03:41:10+0000 | 33 | SG
2018-12-30 03:41:10+0000 | 70 | NA
2018-12-30 03:41:10+0000 | 119 | JP
2018-12-30 03:41:10+0000 | 133 | KR
2018-12-30 03:41:10+0000 | 496 | US
2018-12-30 21:45:30+0000 | 50 | KR
2018-12-30 21:45:30+0000 | 72 | AT
2018-12-30 21:45:30+0000 | 83 | CN
2018-12-30 21:45:30+0000 | 96 | NA
2018-12-30 21:45:30+0000 | 385 | US
2018-12-30 02:59:00+0000 | 27 | ES
2018-12-30 02:59:00+0000 | 49 | CA
2018-12-30 02:59:00+0000 | 59 | NA
2018-12-30 02:59:00+0000 | 62 | GB
2018-12-30 02:59:00+0000 | 728 | US
(660 rows)
cqlsh:streaming>
cqlsh:streaming>

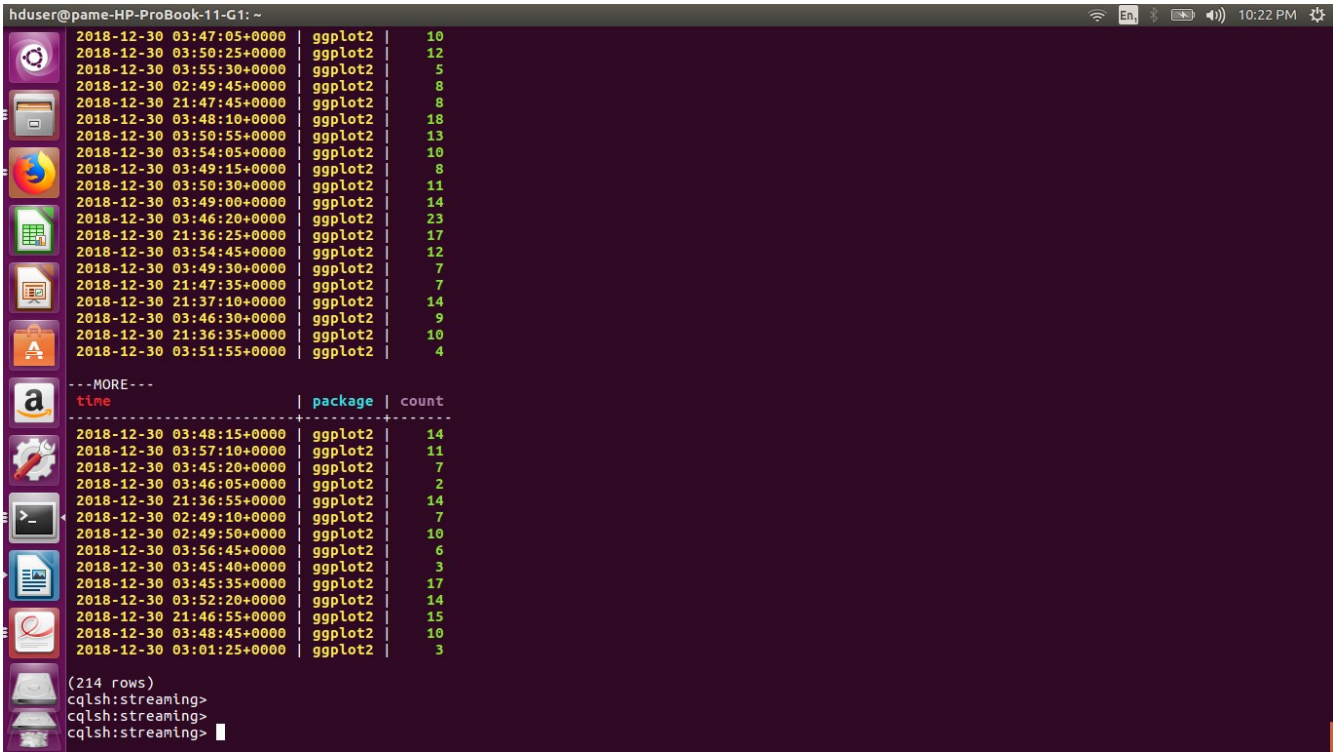
Q4. To find number of downloads for ggplot2 package.

CQL Code Table:

:cqlsh:streaming> CREATE TABLE q4 (time text, package text, count int, PRIMARY KEY(time, package));

ScreenShots:




hduser@pame-HP-ProBook-11-G1: ~
10:22 PM
2018-12-30 03:47:05+0000 | ggplot2 | 10
2018-12-30 03:50:25+0000 | ggplot2 | 12
2018-12-30 03:55:30+0000 | ggplot2 | 5
2018-12-30 02:49:45+0000 | ggplot2 | 8
2018-12-30 21:47:45+0000 | ggplot2 | 8
2018-12-30 03:48:10+0000 | ggplot2 | 18
2018-12-30 03:50:55+0000 | ggplot2 | 13
2018-12-30 03:54:05+0000 | ggplot2 | 10
2018-12-30 03:49:15+0000 | ggplot2 | 8
2018-12-30 03:50:30+0000 | ggplot2 | 11
2018-12-30 03:49:00+0000 | ggplot2 | 14
2018-12-30 03:46:20+0000 | ggplot2 | 23
2018-12-30 21:36:25+0000 | ggplot2 | 17
2018-12-30 03:54:45+0000 | ggplot2 | 12
2018-12-30 03:49:30+0000 | ggplot2 | 7
2018-12-30 21:47:35+0000 | ggplot2 | 7
2018-12-30 21:37:10+0000 | ggplot2 | 14
2018-12-30 03:46:30+0000 | ggplot2 | 9
2018-12-30 21:36:35+0000 | ggplot2 | 10
2018-12-30 03:51:55+0000 | ggplot2 | 4
---MORE---
time | package | count
2018-12-30 03:48:15+0000 | ggplot2 | 14
2018-12-30 03:57:10+0000 | ggplot2 | 11
2018-12-30 03:45:20+0000 | ggplot2 | 7
2018-12-30 03:46:05+0000 | ggplot2 | 2
2018-12-30 21:36:55+0000 | ggplot2 | 14
2018-12-30 02:49:10+0000 | ggplot2 | 7
2018-12-30 02:49:50+0000 | ggplot2 | 10
2018-12-30 03:56:45+0000 | ggplot2 | 6
2018-12-30 03:45:40+0000 | ggplot2 | 3
2018-12-30 03:45:35+0000 | ggplot2 | 17
2018-12-30 03:52:20+0000 | ggplot2 | 14
2018-12-30 21:46:55+0000 | ggplot2 | 15
2018-12-30 03:48:45+0000 | ggplot2 | 10
2018-12-30 03:01:25+0000 | ggplot2 | 3
(214 rows)
cqlsh:streaming>
cqlsh:streaming>
cqlsh:streaming>