

Submitted by T D Prasad

RStudioLogFiles\_Analysis On hive.....

Q)Count the number of distinct users each day(with one week data) ?

Query:

```
select dates, count(distinct(ip_id)) from rstudios group by dates ;
```

out\_put:

"2019-02-04" 72675

"2019-02-05" 76464

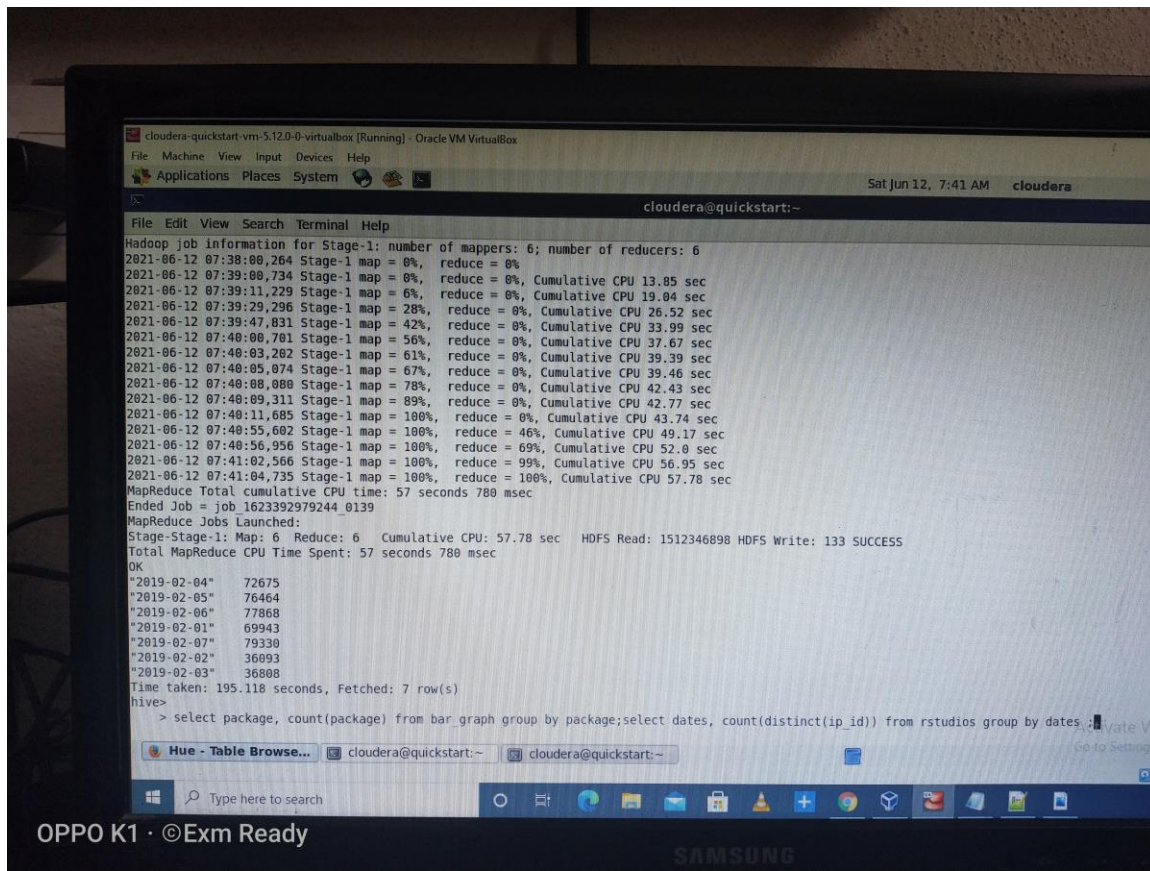
"2019-02-06" 77868

"2019-02-01" 69943

"2019-02-07" 79330

"2019-02-02" 36093

"2019-02-03" 36808



Q)Number of individual users each day with one week data?

Query:

```
select dates, count(dates) from rstudios group by dates;
```

out\_put:

```
"2019-02-04" 2864816
```

"2019-02-05" 3026029

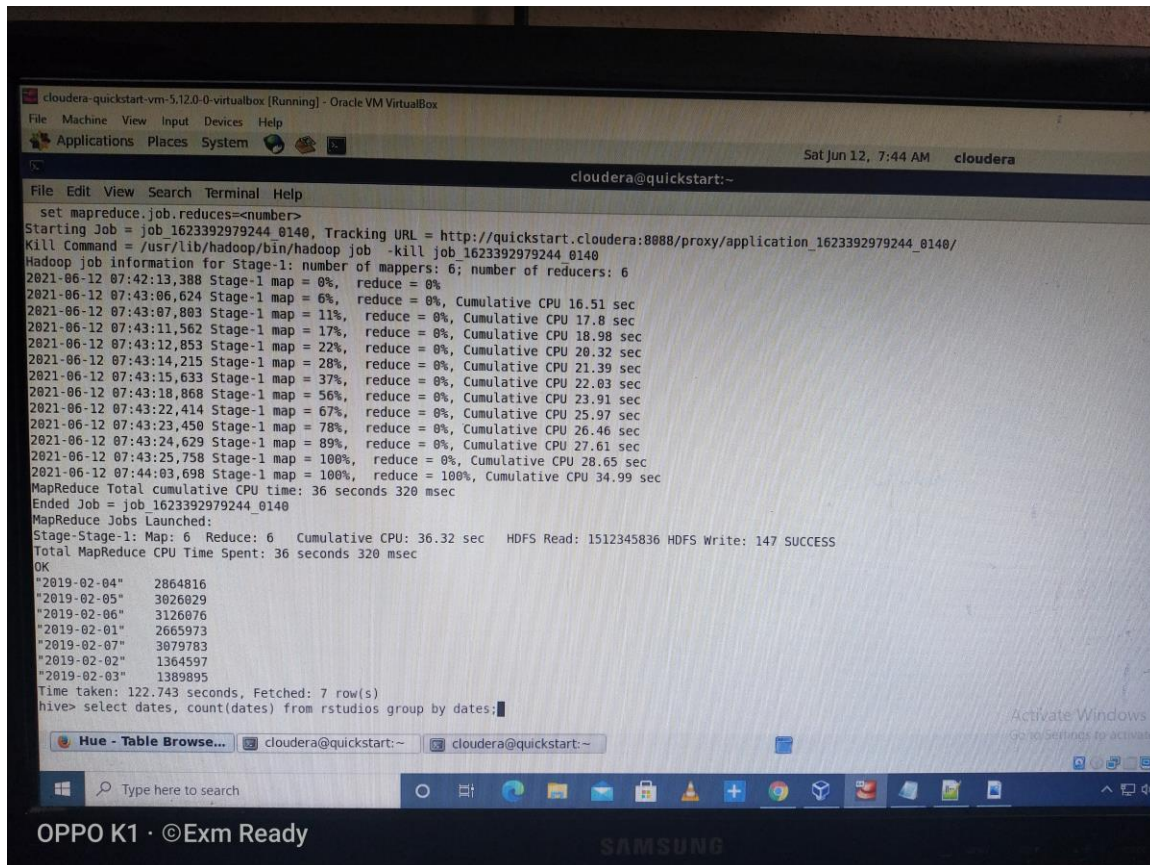
"2019-02-06" 3126076

"2019-02-01" 2665973

"2019-02-07" 3079783

"2019-02-02" 1364597

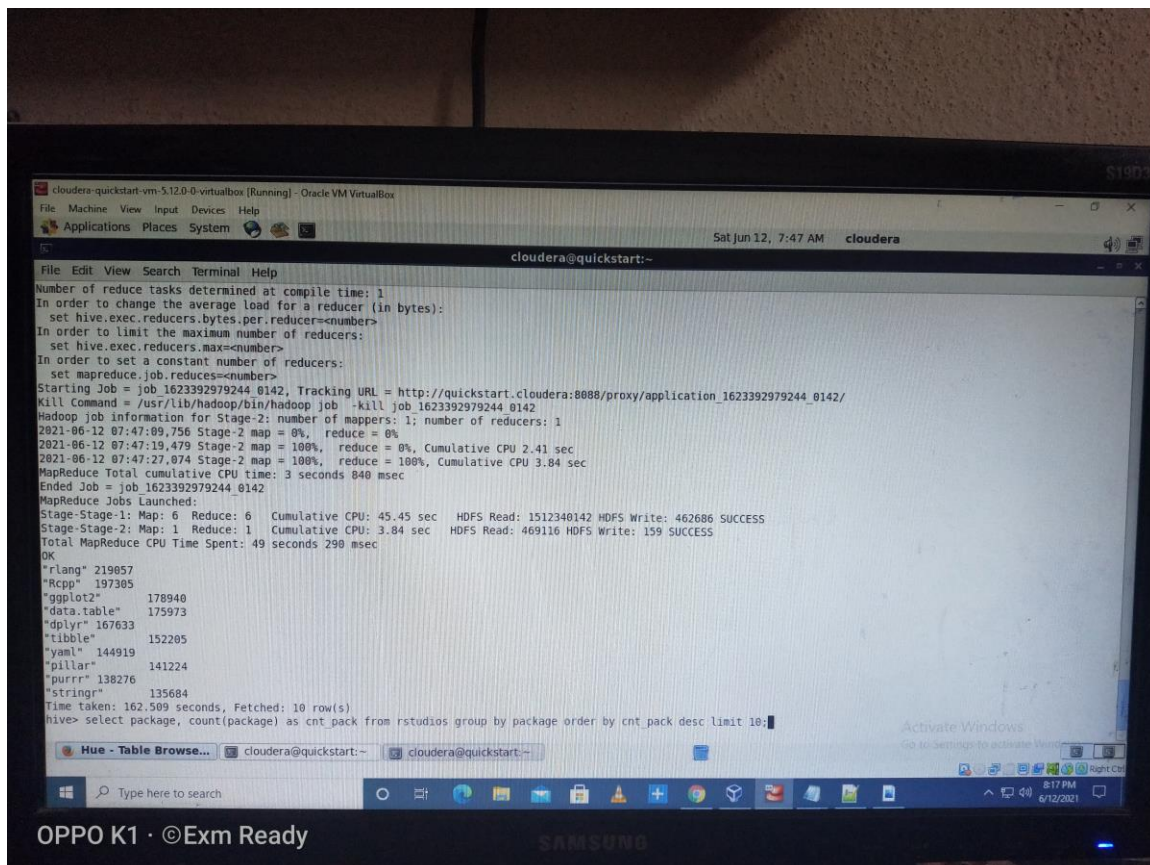
"2019-02-03" 1389895



Q) Top 10 packages downloaded on oneweek data?

select package, count(package) as cnt\_pack from rstudios group by package order by cnt\_pack desc limit 10;

"rlang" 219057  
"Rcpp" 197305  
"ggplot2" 178940  
"data.table" 175973  
"dplyr" 167633  
"tibble" 152205  
"yaml" 144919  
"pillar" 141224  
"purrr" 138276  
"stringr" 135684



Q) Top 10 packages downloaded (or) Count the number of occurrences of different packages?

"rlang" 38986 1

"dplyr" 34366 2

"Rcpp" 34349 3

"data.table" 33805 4

"ggplot2" 31533 5

"stringr" 29301 6

"purrr" 29149 7

"tibble" 27270 8

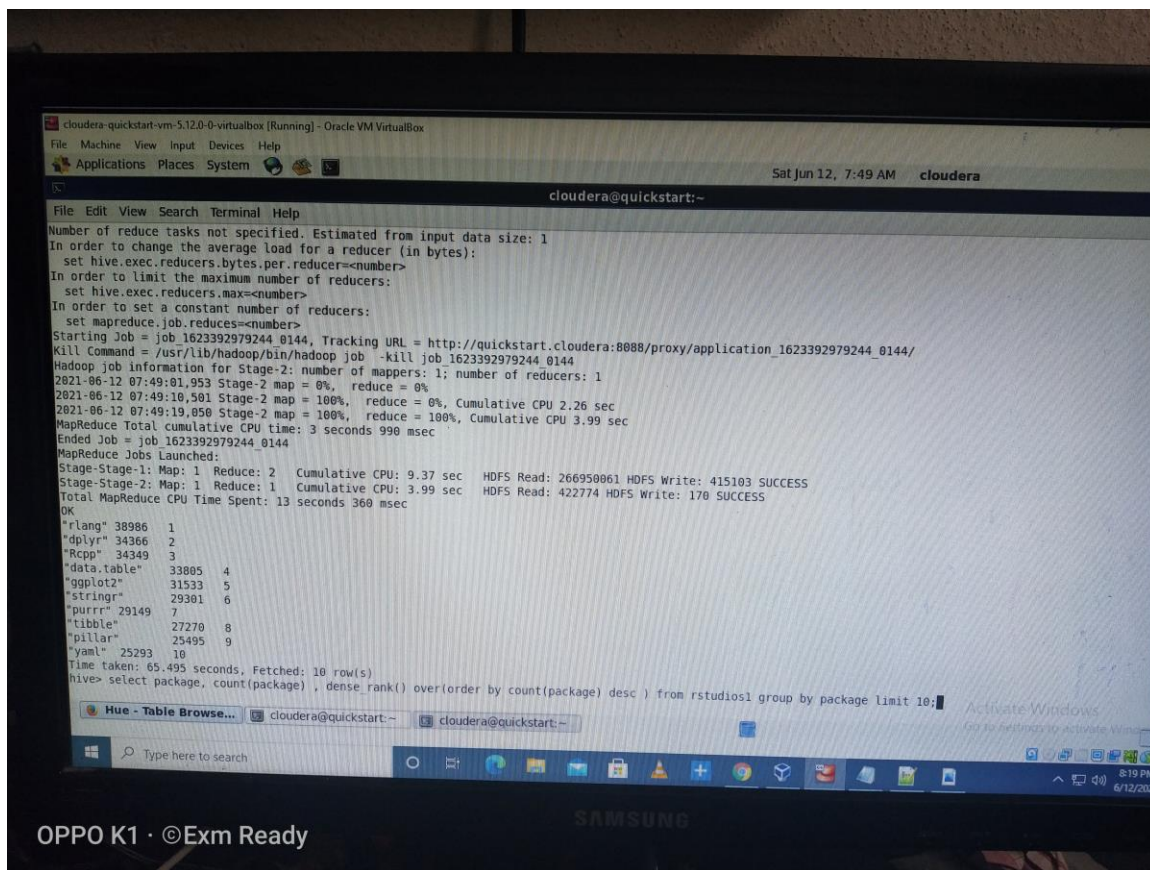
"pillar" 25495 9

"yaml" 25293 10

Time taken: 66.742 seconds, Fetched: 10 row(s)

```
hive> select package, count(package) , dense_rank() over(order by count(package) desc ) from rstudios1
group by package limit 10;
```





Q)Count the number of occurrences of different packages by operating system?

Q)Top 100 packages os wise as well?

Query:

```

select * from (select package, r_os, count(package) as cnt, rank() over(partition by r_os order
by count(package) desc ) as rank from rstudios1 group by r_os, package) as p where p.rank < 2;

```

"promises"	"cygwin"	7	1
"cli"	"darwin10.8.0"	6	1
"ggplot2"	"darwin10.8.0"	6	1
"data.table"	"darwin11.4.2"	5	1
"htmltools"	"darwin11.4.2"	5	1
"cellranger"	"darwin13"	2	1
"fs"	"darwin13"	2	1
"memoise"	"darwin13"	2	1
"StanHeaders"	"darwin13"	2	1
"hms"	"darwin13.1.0"	1	1
"Rcpp"	"darwin13.1.0"	1	1
"ggplot2"	"darwin13.1.0"	1	1
"kpeaks"	"darwin13.1.0"	1	1
"readxl"	"darwin13.1.0"	1	1
"rlang"	"darwin13.4.0"	1360	1
"Rcpp"	"darwin14.5.0"	8	1
"shiny"	"darwin14.5.0"	8	1
"rlang"	"darwin15.6.0"	4699	1
"tibble"	"darwin16.1.0"	2	1
"remotes"	"darwin16.1.0"	2	1
"sp"	"darwin16.1.0"	2	1
"readr"	"darwin16.1.0"	2	1
"magrittr"	"darwin16.4.0"	2	1
"rstudioapi"	"darwin16.6.0"	1	1

"unittest"	"darwin16.6.0"	1	1
"wordcloud"	"darwin16.6.0"	1	1
"stringr"	"darwin16.6.0"	1	1
"mnormt"	"darwin16.6.0"	1	1
"dplyr"	"darwin16.6.0"	1	1
"bitops"	"darwin16.6.0"	1	1
"bindr"	"darwin16.6.0"	1	1
"RYandexTranslate"	"darwin16.6.0"	1	1
"NORMA"	"darwin16.6.0"	1	1
"GGally"	"darwin16.6.0"	1	1
"RcppParallel"	"darwin16.6.0"	1	1
"abc"	"darwin16.6.0"	1	1
"clipr"	"darwin16.6.0"	1	1
"generics"	"darwin16.6.0"	1	1
"greybox"	"darwin16.6.0"	1	1
"gridExtra"	"darwin16.6.0"	1	1
"htmltools"	"darwin16.6.0"	1	1
"htmlwidgets"	"darwin16.6.0"	1	1
"mclust"	"darwin16.6.0"	1	1
"pkgconfig"	"darwin16.6.0"	1	1
"xopen"	"darwin16.6.0"	1	1
"viridisLite"	"darwin16.6.0"	1	1
"data.table"	"darwin16.7.0"	18	1
"stringr"	"darwin16.7.0"	18	1
"rlang"	"darwin16.7.0"	18	1



"Rcpp"	"darwin16.7.0"	18	1
"xfun"	"darwin16.7.0"	18	1
"clipr"	"darwin17.0.0"	1	1
"usethis"	"darwin17.0.0"	1	1
"covr"	"darwin17.0.0"	1	1
"devtools"	"darwin17.0.0"	1	1
"gbm"	"darwin17.0.0"	1	1
"htmlwidgets"	"darwin17.0.0"	1	1
"lme4"	"darwin17.0.0"	1	1
"rio"	"darwin17.0.0"	1	1
"rvest"	"darwin17.0.0"	1	1
"TH.data"	"darwin17.0.0"	1	1
"ape"	"darwin17.0.0"	1	1
"fracdiff"	"darwin17.0.0"	1	1
"Rcpp"	"darwin17.0.0"	1	1
"caret"	"darwin17.0.0"	1	1
"bit"	"darwin17.2.0"	1	1
"modelr"	"darwin17.2.0"	1	1
"markdown"	"darwin17.2.0"	1	1
"EasyABC"	"darwin17.2.0"	1	1
"tidyr"	"darwin17.2.0"	1	1
"data.table"	"darwin17.3.0"	4	1
"stringr"	"darwin17.4.0"	12	1
"dplyr"	"darwin17.5.0"	5	1
"stringr"	"darwin17.6.0"	14	1

"ggplot2"	"darwin17.7.0"	25	1
"cli"	"darwin18.0.0"	11	1
"rlang"	"darwin18.2.0"	75	1
"polynom"	"freebsd11.2"	5	1
"ps"	"freebsd12.0"	1	1
"rmarkdown"	"freebsd12.0"	1	1
"rlang"	"linux-gnu"	13006	1
"pillar"	"linux-gnueabihf"	4	1
"lubridate"	"linux-gnueabihf"	4	1
"shiny"	"linux-gnueabihf"	4	1
"jsonlite"	"linux-gnueabihf"	4	1
"purrr"	"linux-gnueabihf"	4	1
"readr"	"linux-gnueabihf"	4	1
"promises"	"linux-gnueabihf"	4	1
"stringi"	"linux-musl"	2	1
"rlang"	"mingw32"	19336	1
"stringr"	"openbsd6.4"	1	1
"modelr"	"openbsd6.4"	1	1
"stringi"	"solaris2.10"	1	1
"metamisc"	NA	1149	1

Time taken: 67.026 seconds, Fetched: 94 row(s)

