

marininvp

2023-04-09

Загрузка данных

```
new_data <- arrow::read_csv_arrow("D:\\ProjectsRstudio\\2_sem_R\\LAB1\\gowiththeflow_20190826.csv", schema=schema
(timestamp=int64(),src=utf8(),dst=utf8(),port=int32(),bytes=int32()))
```

```
head(new_data, 10)
```

```
## # A tibble: 10 × 5
##       timestamp src          dst          port bytes
##       <int64> <chr>       <chr>       <int> <int>
## 1 1578326400001 13.43.52.51 18.70.112.62    40 57354
## 2 1578326400005 16.79.101.100 12.48.65.39     92 11895
## 3 1578326400007 18.43.118.103 14.51.30.86     27   898
## 4 1578326400011 15.71.108.118 14.50.119.33     57  7496
## 5 1578326400012 14.33.30.103 15.24.31.23    115 20979
## 6 1578326400012 18.121.115.31 13.56.39.74     92  8620
## 7 1578326400014 16.108.75.29 14.34.34.69     65 46033
## 8 1578326400018 12.46.104.126 16.25.76.33    123  1500
## 9 1578326400021 12.43.98.93 18.85.31.68     79   979
## 10 1578326400021 14.32.60.107 12.30.62.113     72 1036
```

```
filter(new_data, str_detect(src, "^(12|13|14)\\.\\.\\."), str_detect(dst, "^(12|13|14)\\.\\.\\.") %>% select(src, bytes) %>% group_by(src) %>% summarise(bytes = sum(bytes)) %>% slice_max(bytes) %>% select(src)
```

```
## # A tibble: 1 × 1
##   src
##   <chr>
## 1 13.37.84.125
```

```
new_data %>% select(src,dst,bytes) %>% filter(str_detect(src,"^((12|13|14)\\.\\.)",str_detect(dst,"^((12|13|14)\\.\\.)",negate=TRUE)) %>%  
group_by(src) %>% summarise(loads = sum(bytes)) %>% arrange(desc(loads)) %>% head(10)
```

```
## # A tibble: 10 × 2
##   src                      loads
##   <chr>                   <dbl>
## 1 13.37.84.125 10625497574
## 2 13.48.72.30 1727525696
## 3 14.51.75.107 1656207638
## 4 14.51.30.86 1647330708
## 5 12.59.25.34 1617316039
## 6 12.56.32.111 1600604429
## 7 14.57.50.29 1589314261
## 8 13.39.46.94 1576854685
## 9 12.58.68.102 1545945975
## 10 12.45.94.34 1501437235
```