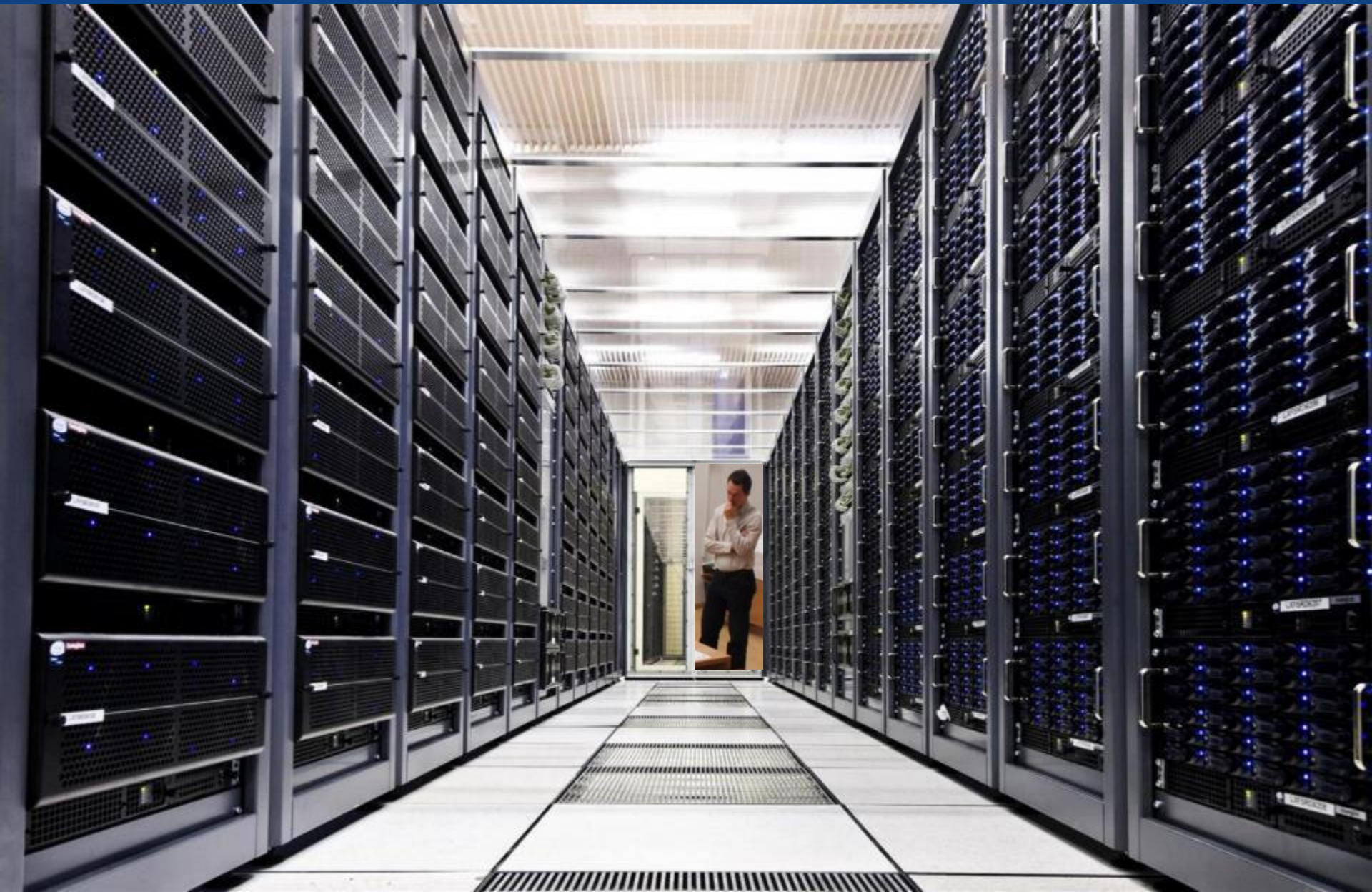







Data archiving



(The result of two weeks' clickin' and googlin')



Big data on tape

Company		
<p>Tape libraries</p> 	<p>2 x Oracle StorageTek SL8500</p>	<p>2 x TS4500 1 x TS3500</p>
<p>Tape drives</p> 	<p>20 x Oracle StorageTek T10000D</p>	<p>46 x IBM TS1155 20 x IBM LTO-8</p>
<p>Tape cartridges</p> 	<p>10000 x T2 - 8 TB</p>	<p>13000 x IBM 3592JD - 15 TB 6000 x IBM 3592JC - 7 TB 500 x LTO-7M - 9 TB 160 x LTO-8 - 12 TB</p>
<p>Current amount of data in the archive: 250 PB</p>		



Tape library - Oracle StorageTek SL8500





Comparison of the tools

splunk®

- primarily commercial software
- not widely used in the IT department
- knowledge lost with departure of a colleague
- used version obsolete
- data loading limitations

GET ACTIONABLE INTELLIGENCE

Splunk® Enterprise

STARTS AT

\$173

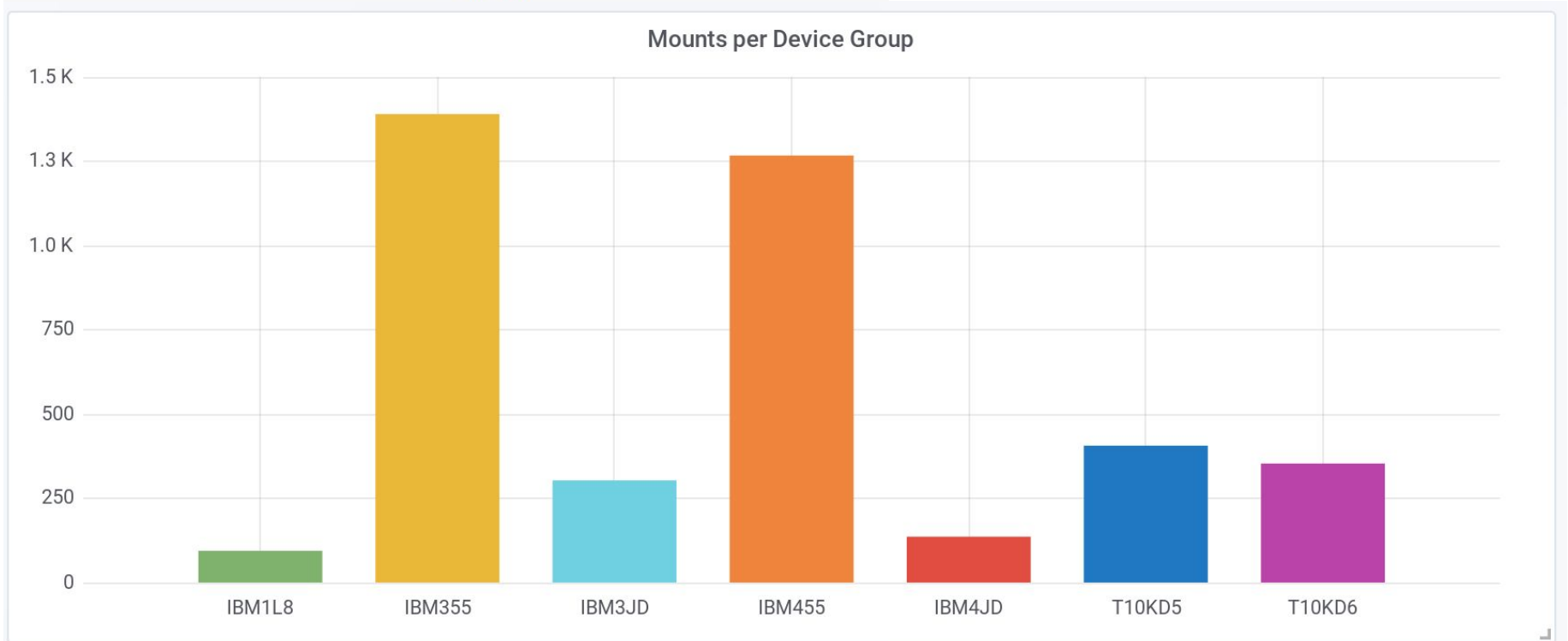
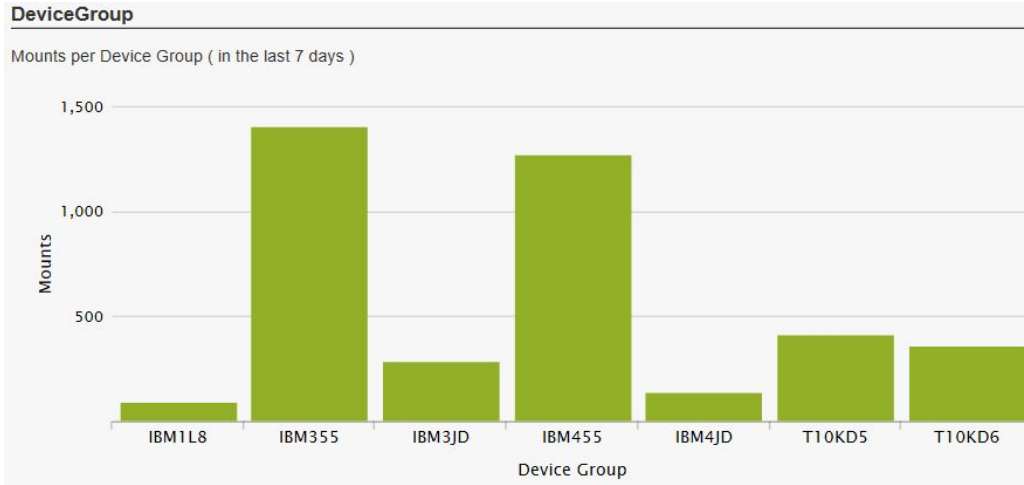
Per [Ingested GB](#), Per Month, Billed Annually*

Grafana

- open source with wide community
- underlying infrastructure used by majority of IT department services
- different design / capabilities = different constraints



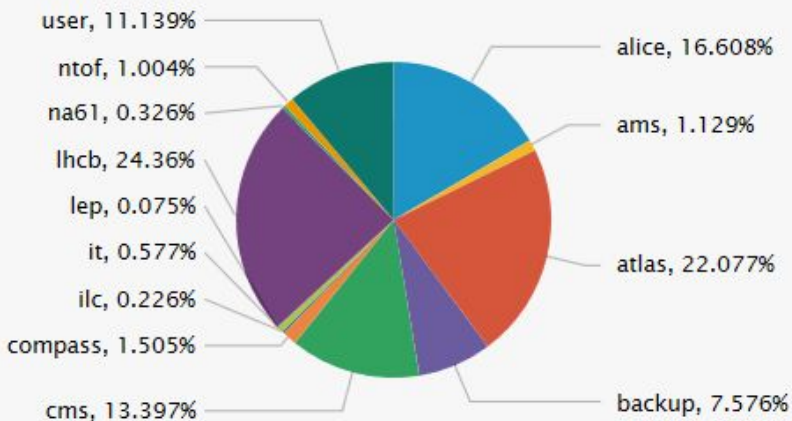
Graph comparison: bars



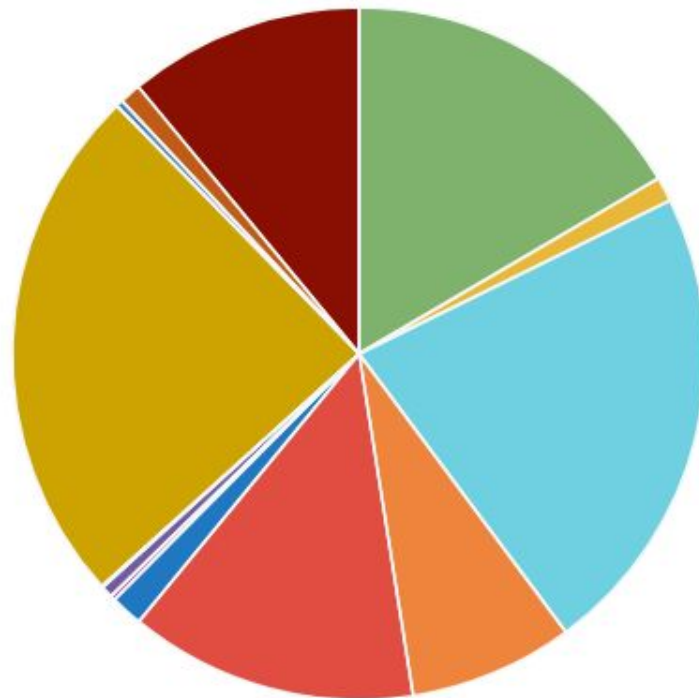


Graph comparison: pie

Mounts per VO (in the last 7 days)



Mounts per VO



current percentage

VO	current	percentage
ALICE	654	17%
AMS	45	1%
ATLAS	874	22%
BACKUP	303	8%
CMS	531	13%



Nested SELECT



Data Source

monit_idb_tape ▾

▸ Options

▸ Help

▸ Query Inspector

▾ B

```
SELECT mean(average_transfer_rate_per_mount) FROM (SELECT
"payload.dataVolume.num" / "payload.transferTime.num" AS
average_transfer_rate_per_mount FROM "tape_sessions_finished" WHERE
("payload.status.str" = 'failure' OR "payload.status.str" = 'success') AND
"payload.clientHost.str" =~ /^c2/ AND $timeFilter) GROUP BY time(1d), "payload.vo.str"
fill(0)
```



FORMAT AS

Time series ▾

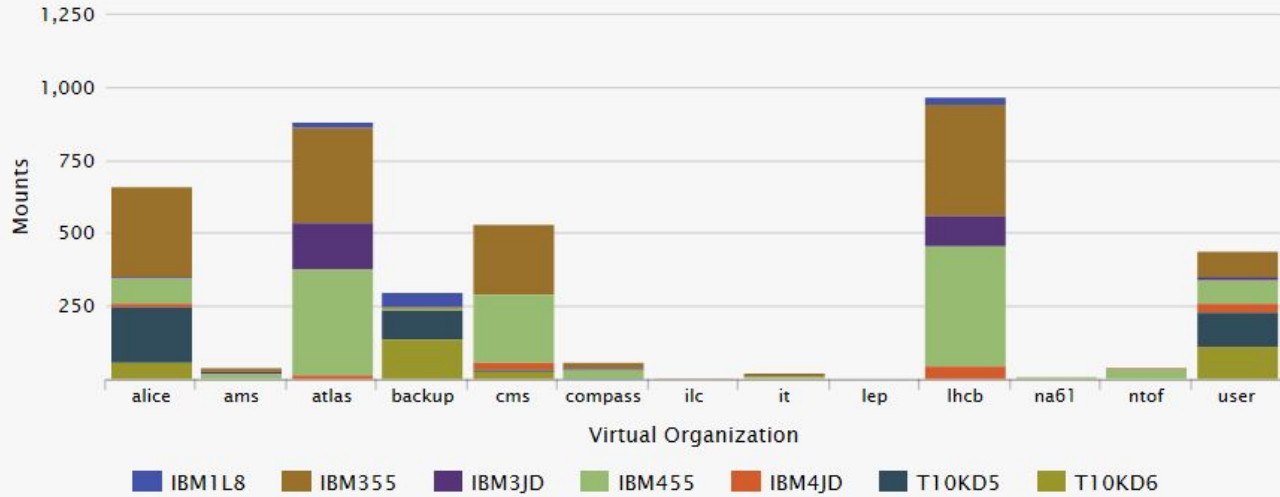
ALIAS BY

[[tag_payload.vo.str]]



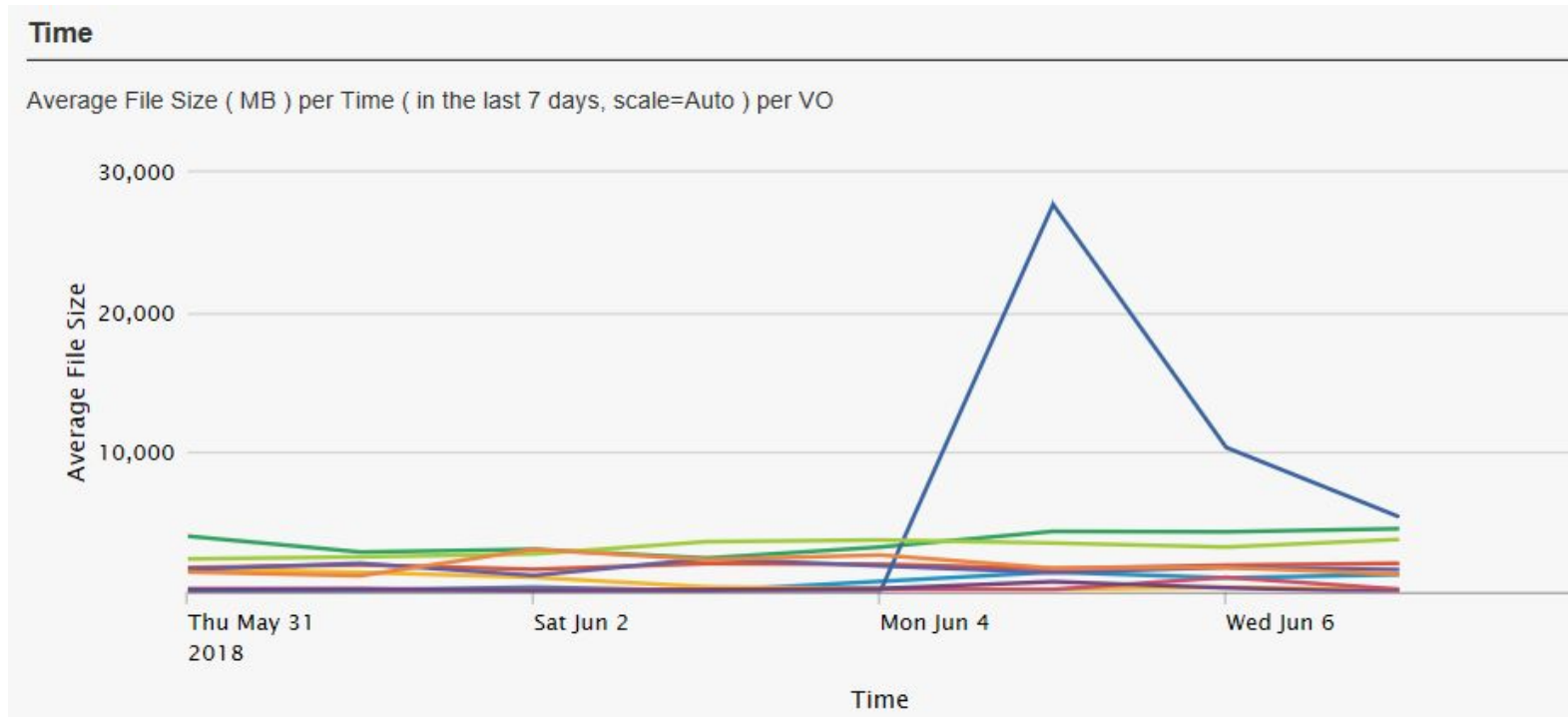
Graph comparison: stacked

MOUNTS per Device Group per VO (in the last 7 days)





Graph comparison: lines



Service Unavailable

The server is temporarily unable to service your request due to maintenance downtime or capacity problems. Please try again later.



Our challenge

1. Choose any programming language you are familiar with.
2. Create and run an external program to extract data from InfluxDB and for each virtual organisation compute:
 - a) the number of all READ mounts,
 - b) the number of unique tape READ mounts.
3. Return output in JSON file format.
4. Learn to handle all error conditions!





Required output

```
{
  "storageservice": {
    "storageshares": [
      {
        "name": "ALICE",
        "timestamp": 1528358958,
        "totalmounts24h": 90,
        "uniquemounts24h": 42,
        "vos": [
          "alice"
        ]
      },
      {
        "name": "ATLAS",
        "timestamp": 1528358958,
        "totalmounts24h": 160,
        "uniquemounts24h": 154,
        "vos": [
          "atlas"
        ]
      },
      {
        "name": "CMS",
        "timestamp": 1528358958,
        "totalmounts24h": 86,
        "uniquemounts24h": 68,
        "vos": [
          "cms"
        ]
      },
      {
        "name": "LHCB",
        "timestamp": 1528358958,
        "totalmounts24h": 153,
        "uniquemounts24h": 136,
        "vos": [
          "lhcb"
        ]
      }
    ]
  }
}
```

```
{
  "name": "ALICE",
  "timestamp": 1528358958,
  "totalmounts24h": 90,
  "uniquemounts24h": 42,
  "vos": [
    "alice"
  ]
},
```




Pros and cons of our project

Pros

- learning UNIX shell and scripting
- improving Python/Java skills
- working with real live data
- creating useful dashboards and code

Cons

- monotony of Splunk-Grafana conversion
(but somebody had to do it ...)

A hand is shown holding a thin black line that slopes upwards from left to right. The word "PROS" is written in green on the lower part of the line, and the word "CONS" is written in red on the upper part of the line.



Thank you for your attention.