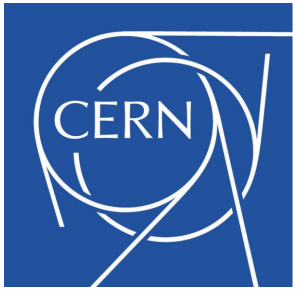
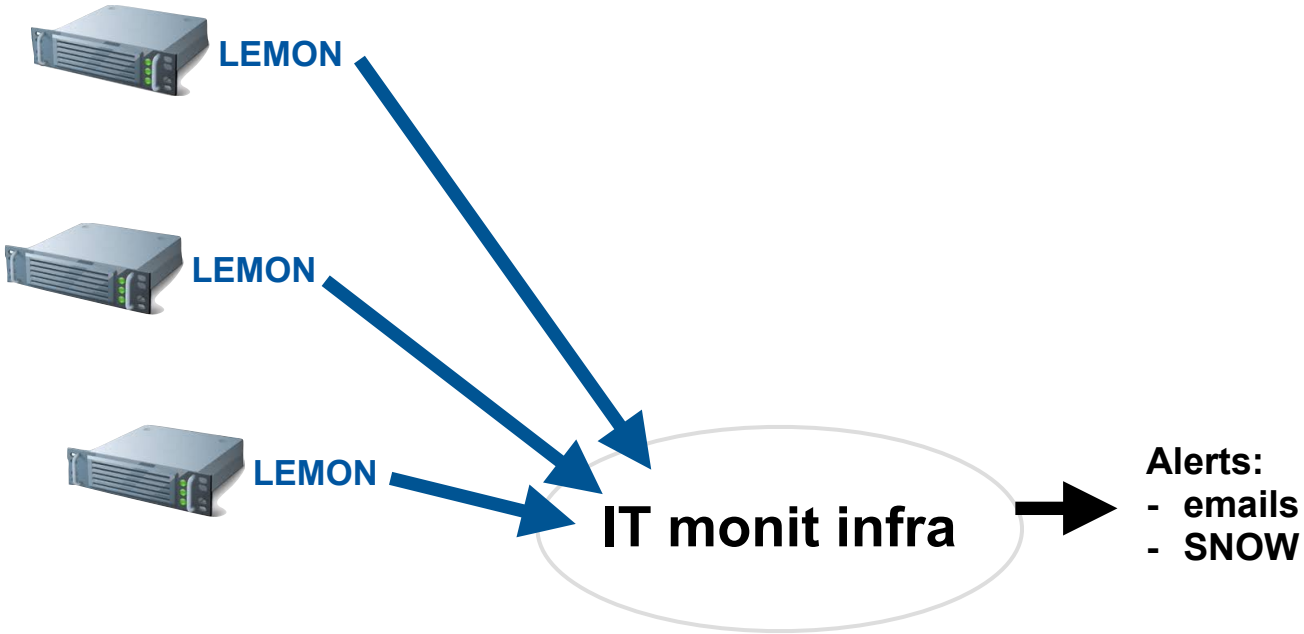
Upgrading monitoring and alerting for the tape infrastructure

Daniel Lanza

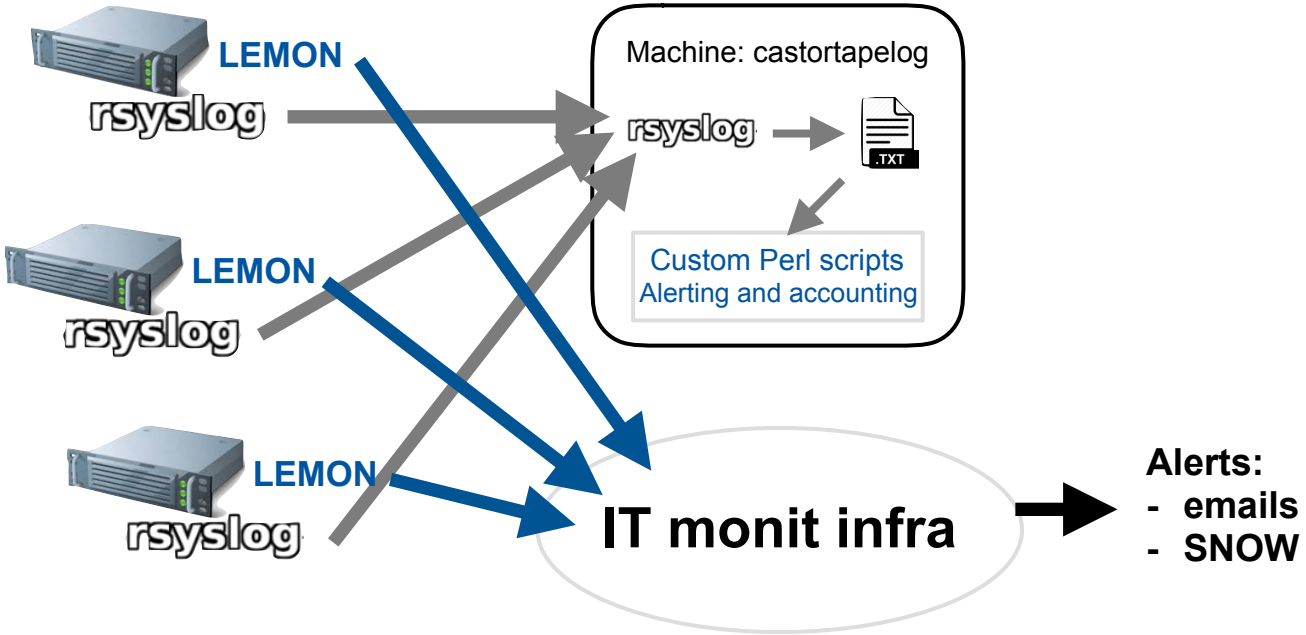
IT-ST group meeting, 12nd June, 2018



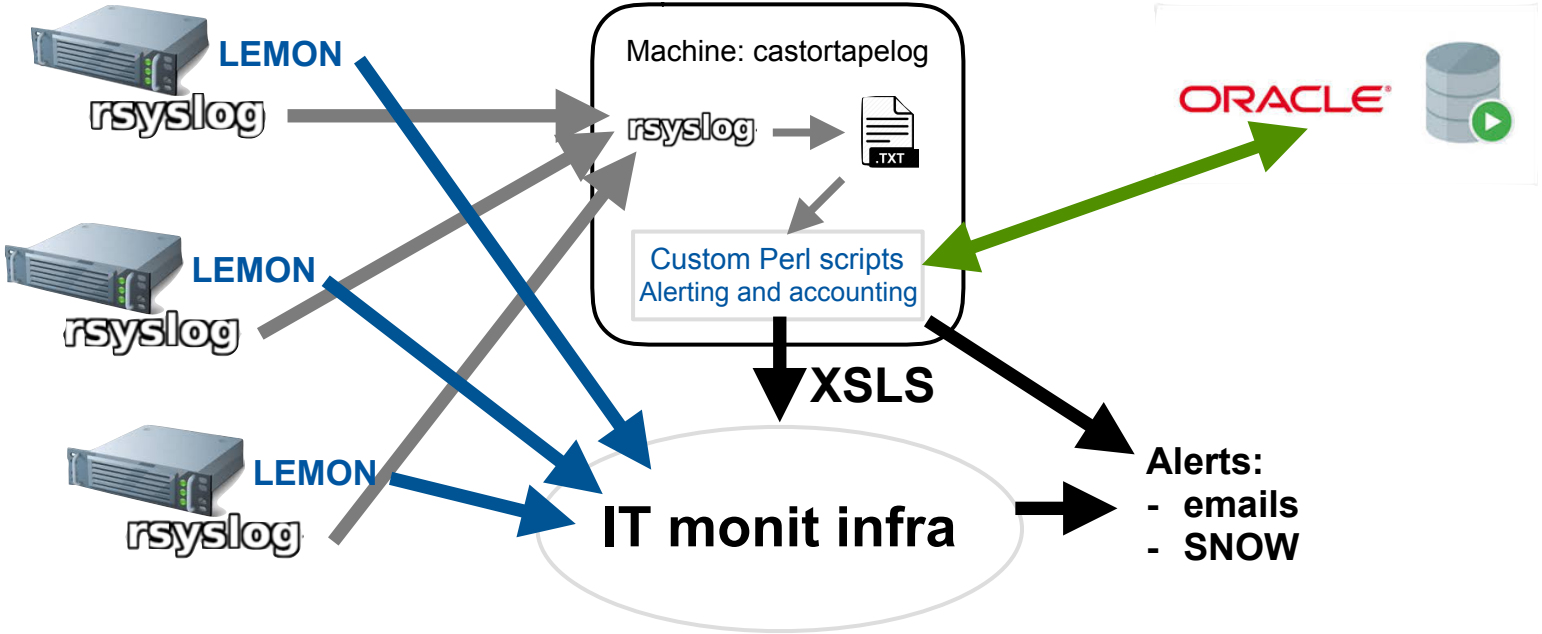
Tape monitoring architecture



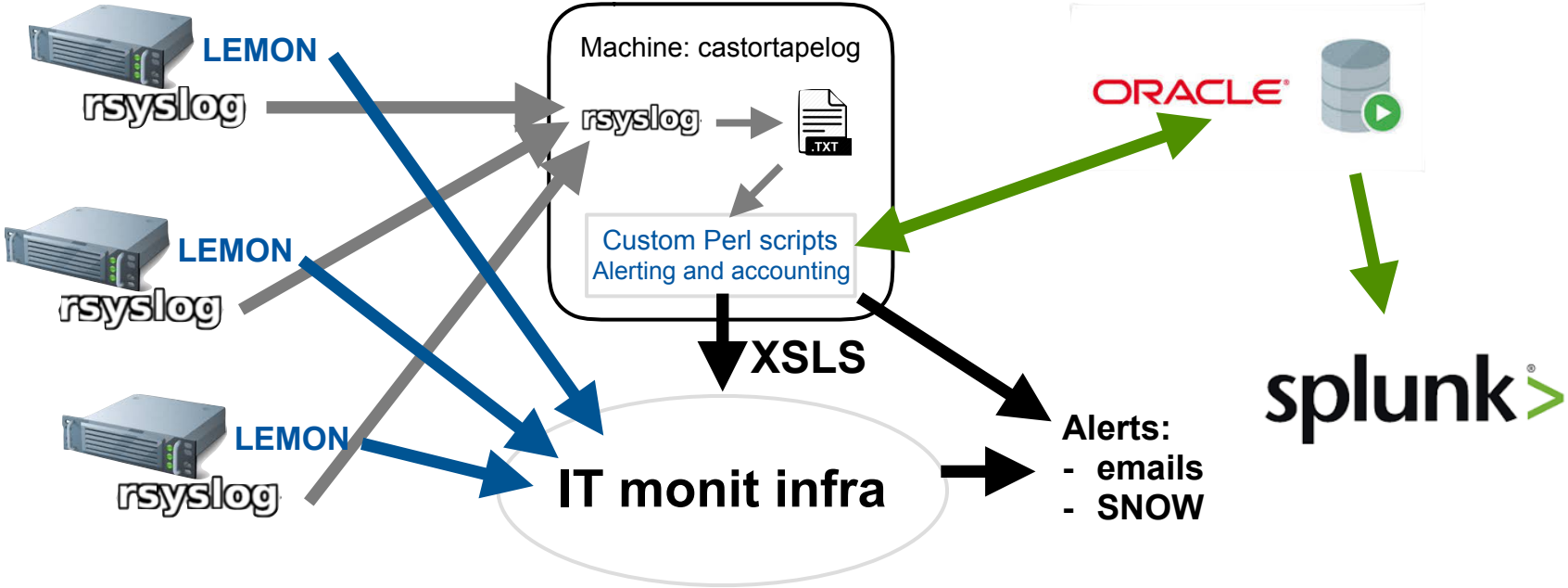
Tape monitoring architecture



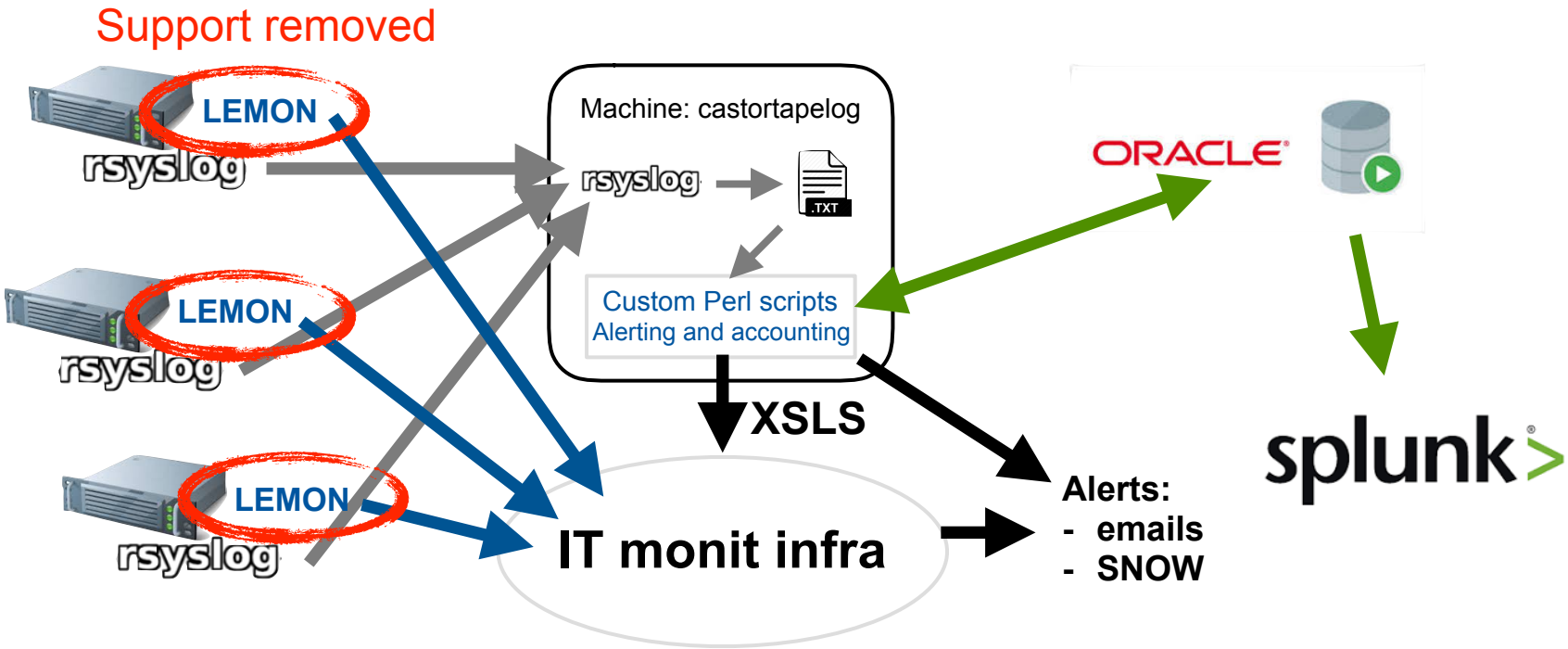
Tape monitoring architecture



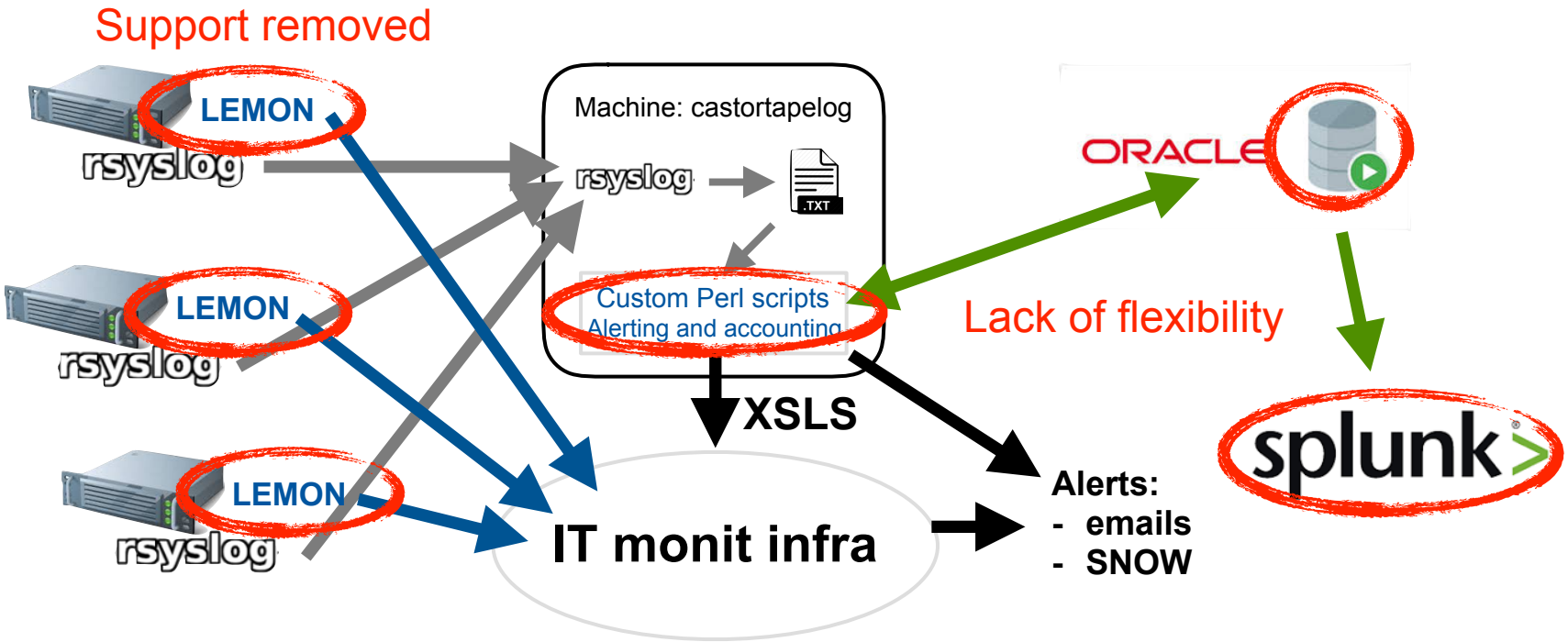
Tape monitoring architecture



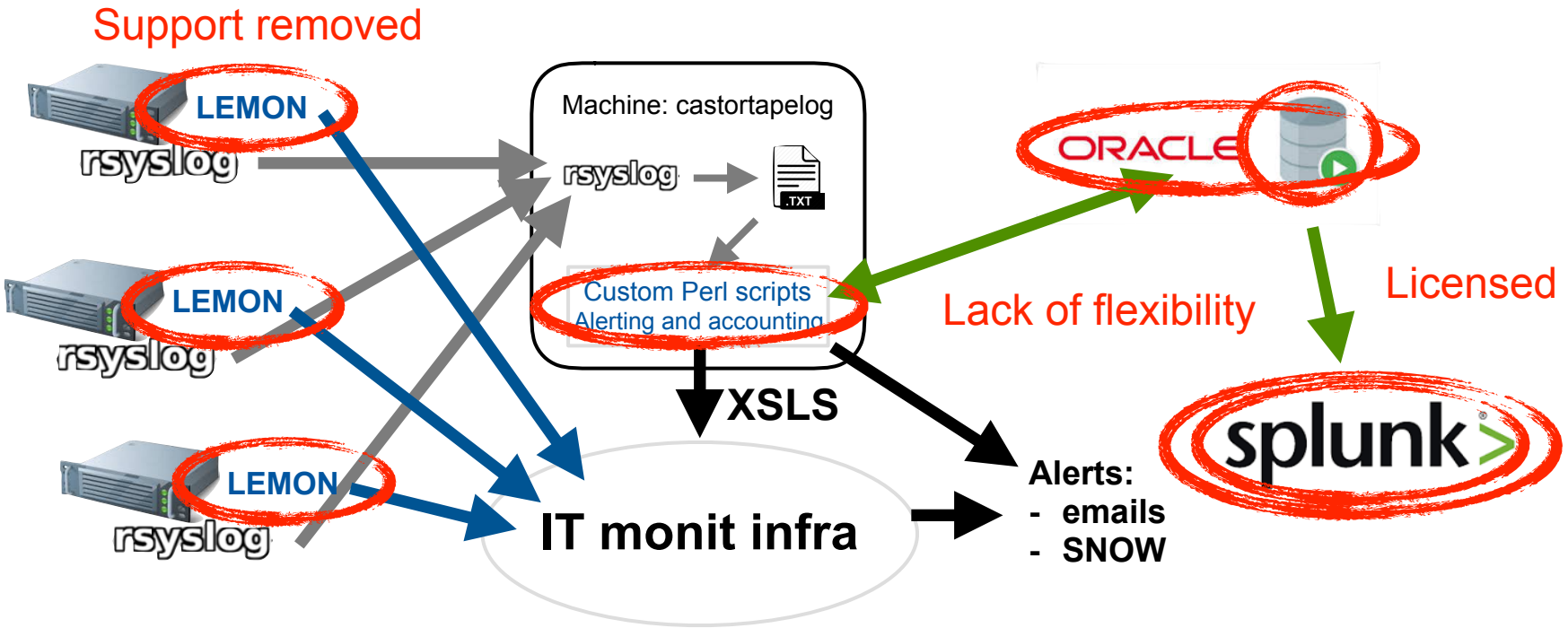
Tape monitoring architecture



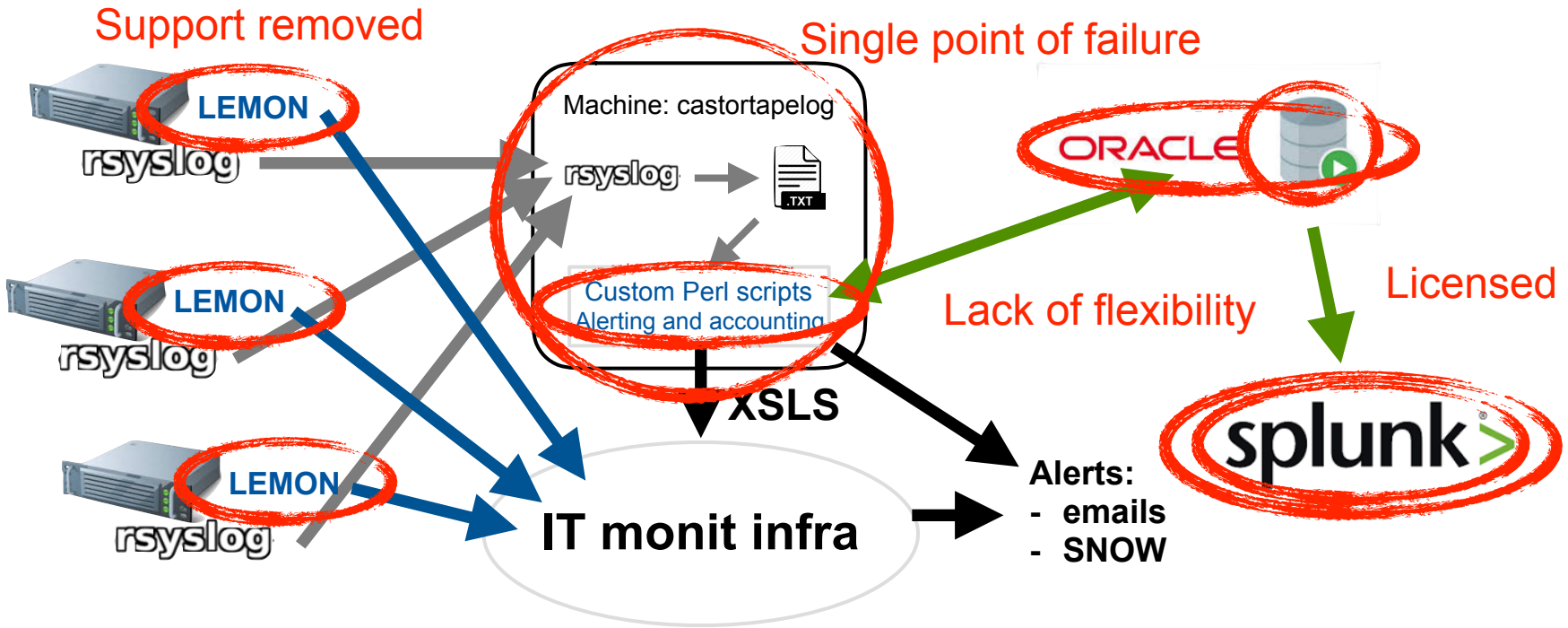
Tape monitoring architecture



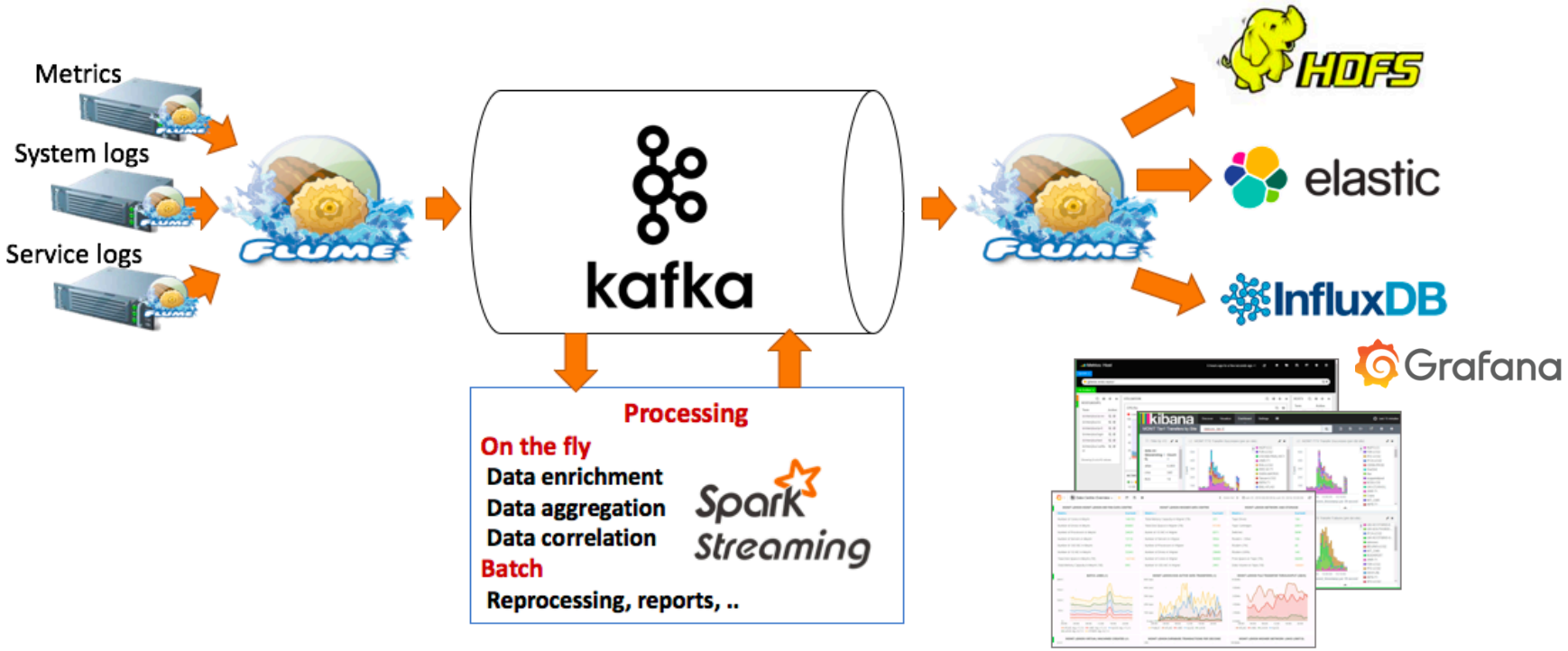
Tape monitoring architecture



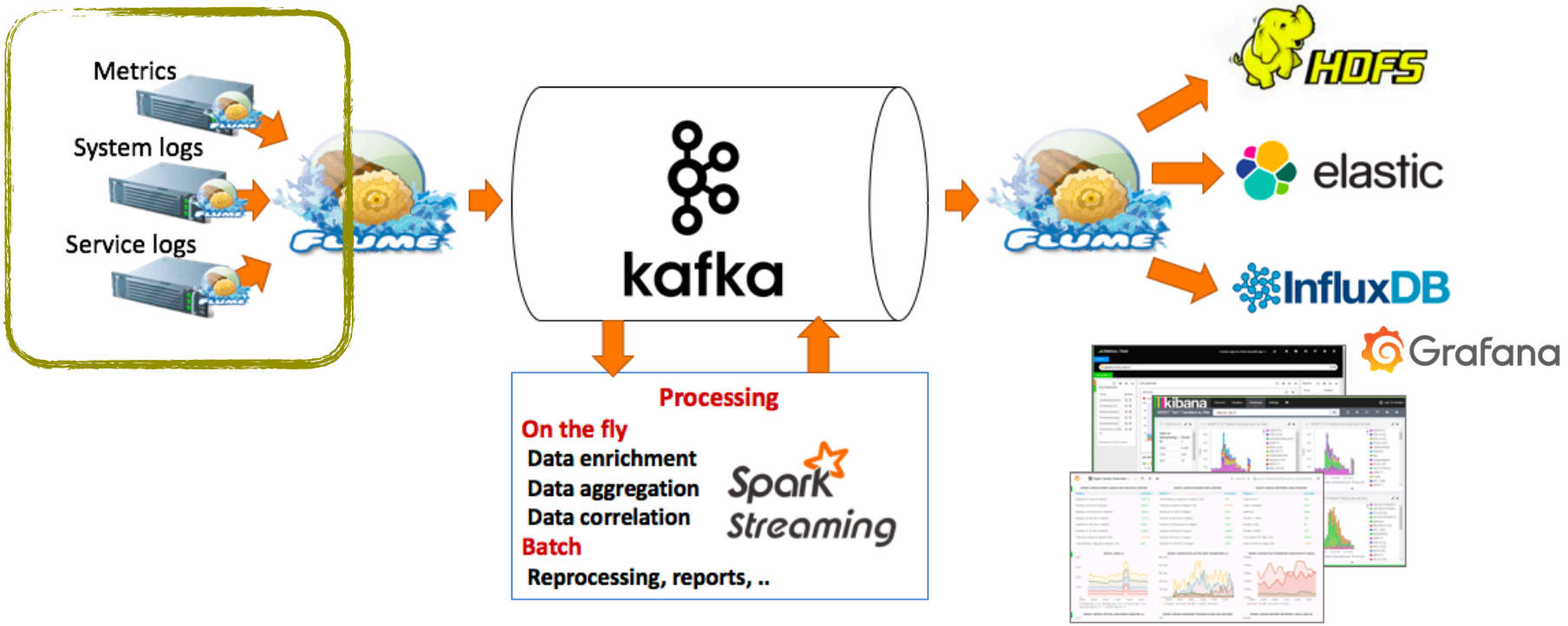
Tape monitoring architecture



IT monitoring architecture



IT monitoring architecture



Storage monitoring module

```
include ::storage_monitoring::<your_hostgroup>
```

- Make monitoring more independent
  - Provide flexibility
  - Changes are supervised
- Components
  - Tape inventory dump (every day)
  - EOS metrics
  - Disk performance metrics
  - SMART (SSDs) disk metrics
  - SNMP traps from tape libraries
  - General logs and metrics collector
- Anyway, you can use components directly

```
class storage_monitoring::tapeserver (
  $exdemon = {},
  $logs    = {},
  $metrics = {},
) {

  include ::storage_monitoring::component::performance_metrics

  class { '::storage_monitoring::component::logs_collector' :
      logs     => $logs,
      metrics  => $metrics,
  }

  include ::storage_monitoring::component::smart_info_report

  exdemon::conf{ 'tape':
    conf => $exdemon,
  }

}
```

```
class{ '::storage_monitoring::component::logs_collector' :
  logs     => $logs,
  metrics  => $metrics,
}
```

logs and metrics collection: parsing


```
2017/08/11 02:43:09.000000 tpsrv101 info tapeserverd[2931]: LVL="Info" TID="2931"
MSG="Tape session finished" volReqId="39154496" dgn="IBM4JD" driveUnit="I4JD0402"
clientHost="c2alice-1.cern.ch" clientPort="62801" clientType="TAPE_GATEWAY"
TPVID="I56144" volumeMode="READ" density="10TC" mountTime="26.697614"
positionTime="28.307989" status="success" tapePoolName="r_alice" vo="ALICE"

2017/08/11 02:43:15.000000 tpsrv202 info tapeserverd[1686]: LVL="Info" TID="1686"
MSG="Tape session finished" volReqId="39154444" dgn="IBM3JD" driveUnit="I3JD0543"
clientHost="c2cms-2.cern.ch" clientPort="62801" clientEuid="14029" TPVID="I52935"
driveTransferSpeedMBps="27.072294" status="success" tapePoolName="r_cms_fam" vo="CMS"

2017/08/11 02:43:16.000000 tpsrv238 info tapeserverd[1295]: LVL="Info" TID="1295"
MSG="Tape session finished" volReqId="39154490" dgn="IBM4JD" driveUnit="I4JD0409"
clientHost="c2cms-2.cern.ch" clientPort="62801" clientEuid="14029" clientEgid="1474"
clientType="TAPE_GATEWAY" TPVID="I45639" volumeMode="READ" density="7000GC"
```

logs and metrics collection: parsing


```
2017/08/11 02:43:09.0000
MSG="Tape session finish
clientHost="c2alice-1.ce
TPVID="I56144" volumeMod
positionTime="28.307989"

2017/08/11 02:43:15.0000
MSG="Tape session finish
clientHost="c2cms-2.cern
driveTransferSpeedMBps="

2017/08/11 02:43:16.0000
MSG="Tape session finish
clientHost="c2cms-2.cern
clientType="TAPE_GATEWAY
```

```
"data": {
  "schemaVersion": "20171123",
  "timestamp_fraction": 0,
  "timestamp_ns": 1528707230000000000,
  "logLevel": "unknown",
  "processName": "tapeserverd",
  "processPid": 1874,
  "payload": {
    "LVL": {
      "str": "Info"
    },
    "density": {
      "str": "15TC"
    },
    "waitInstructionsTime": {
      "num": 0.204764
    },
    "waitDataTime": {
      "num": 0
    },
```

```
VL="Info" TID="2931"
driveUnit="I4JD0402"
"TAPE_GATEWAY"
.697614"
ce" vo="ALICE"

VL="Info" TID="1686"
driveUnit="I3JD0543"
4029" TPVID="I52935"
ame="r_cms_fam" vo="CMS"

VL="Info" TID="1295"
driveUnit="I4JD0409"
4029" clientEgid="1474"
nsity="7000GC"
```

logs and metrics collection: filter and extraction

```
storage_monitoring::tapeserver::logs:
  /var/log/castor/diskmanagerd.log:
    logs:
      type: diskmanagerd
  /var/log/castor/gcd.log:
    logs:
      type: gcd
  /var/log/castor/tapeserverd.log:
    logs:
      type: tapeserverd
    tape_session_finished:
      filter:
        payload.MSG.str: Tape session finished
      tags:
        - hostname
        - payload.driveUnit.str
        - payload.dgn.str
        - payload.clientHost.str
        - payload.volumeMode.str
        - payload.density.str
        - payload.status.str
        - payload.tapePoolName.str
        - payload.vo.str
```

logs and metrics collection: filter and extraction


```
201 ####
202 storage_monitoring::cta::frontend::logs:
203   /var/log/cta/cta-frontend.log:
204     source:
205       socket: "127.0.0.1:6139"
206     logs:
207       type: cta-frontend
208     ctafrontend_cta-frontend_started:
209       filter:
210         payload.MSG.str: cta-frontend started
211       tags:
212         - hostname
213         - instance
214     ctafrontend_queuedArchiveRequest:
215       filter:
216         payload.MSG.str: Queued archive request
217       tags:
218         - hostname
219         - instance
220         - payload.instanceName.str
221         - payload.storageClass.str
222         - payload.policyName.str
```

Monitoring: alerting options

- Local threshold plugin

- On plot alerts

- More complex use cases

Monitoring team: alerting options

- Local threshold plugin

- On plot alerts

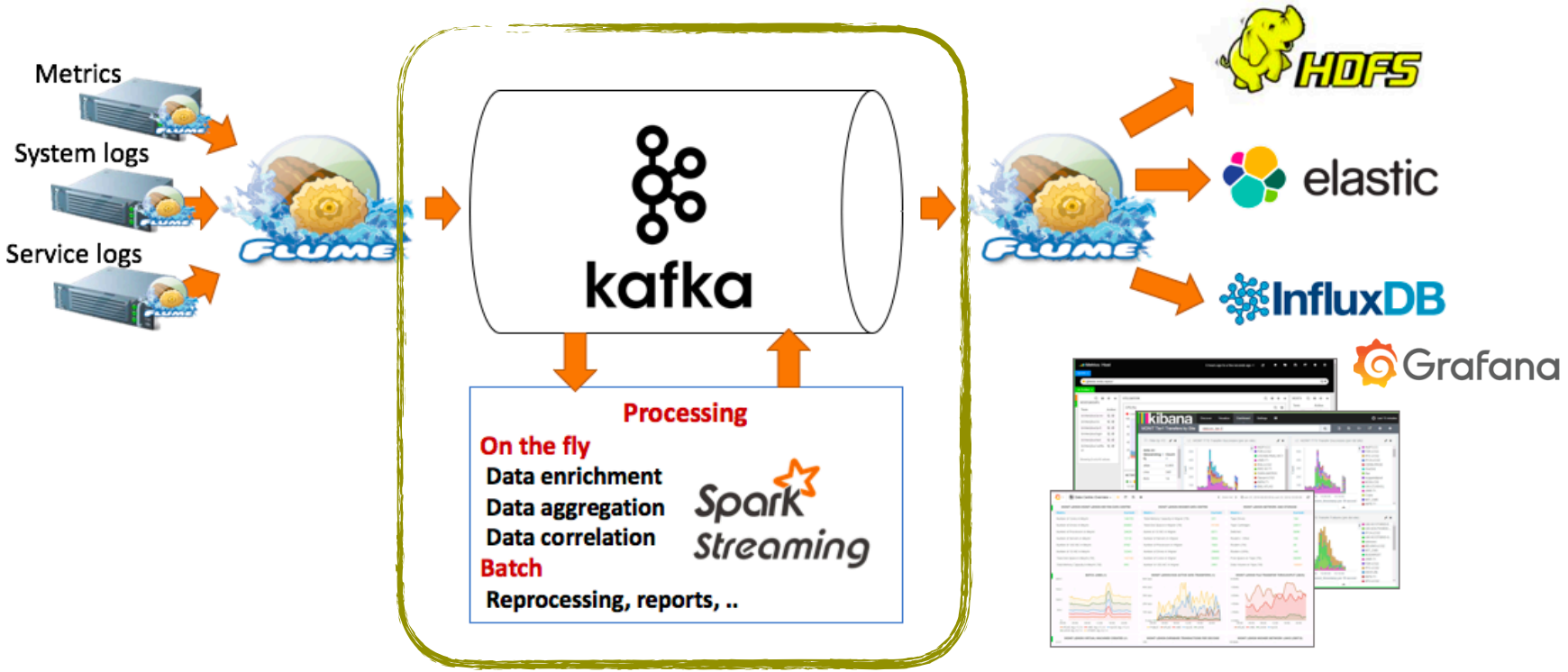
- More complex use cases

Do you have simple use cases?

Why?

- Current use cases
  - Aggregations from different systems
  - For example:
    - A tape failed in 3 different drives
    - Tape failed 15 times in a row
- Ambitions
  - Answer more complex scenarios
    - Is the system back after failure?
  - General purpose solution for alerting
- Configure from Puppet

IT monitoring architecture



A metrics monitor: ExDeMon

- Make use of current provided services.
  - Consume metrics from kafka
  - Notify to elastic, GNI, email, Rundeck, Mattermost,…
- Real-time Spark Streaming
- Aggregation of metrics coming from different systems/machines.
- Flexible metric analytic engine

Mattermost

ExDeMon BOT 5:11 AM

tapeserverd logs not coming from tpsrv233

Rundeck BOT 5:11 AM

Collection system on tpsrv233 was restarted

ExDeMon BOT 6:11 AM

tapeserverd logs are coming again from tpsrv233

RUNDECK

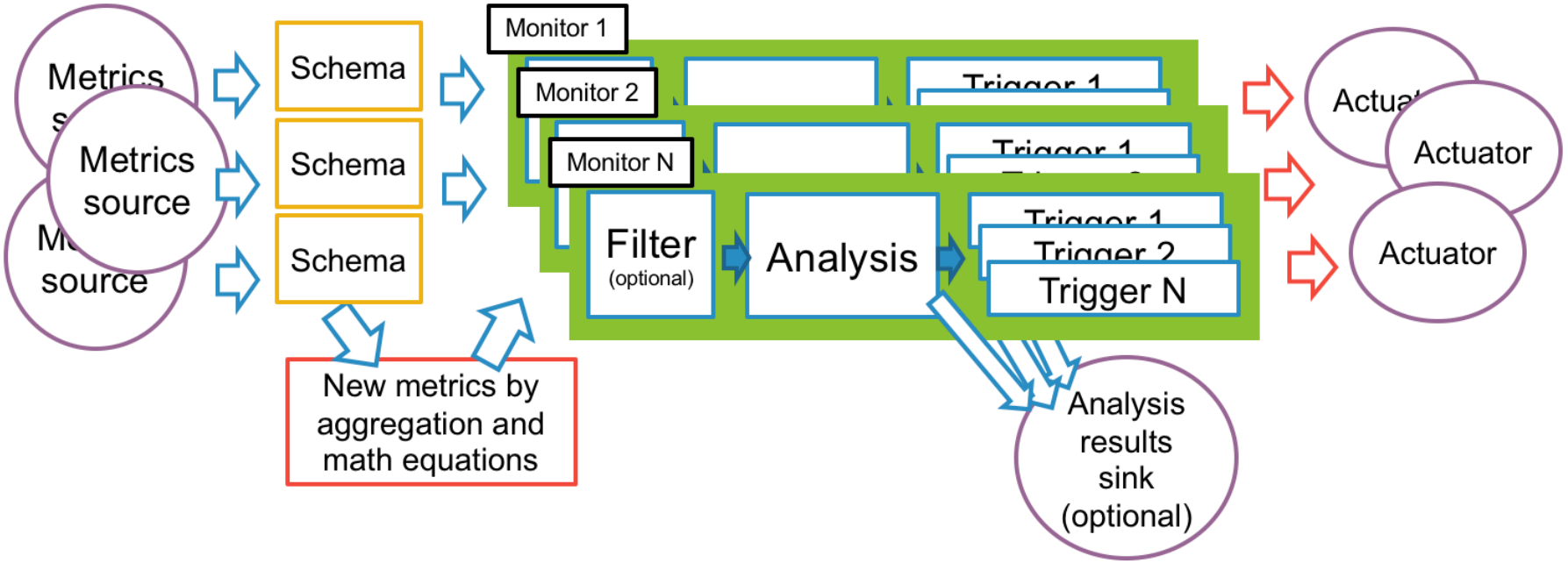
ExDeMon BOT 11:03 AM

A tape (L70027) reported media alerts with at least 3 drives (I1L80531@tpsrv119,I1L80242@tpsrv116,I1L80533@tpsrv114,)

Rundeck BOT 11:03 AM

Tape L70027 was successfully disabled. (edited)

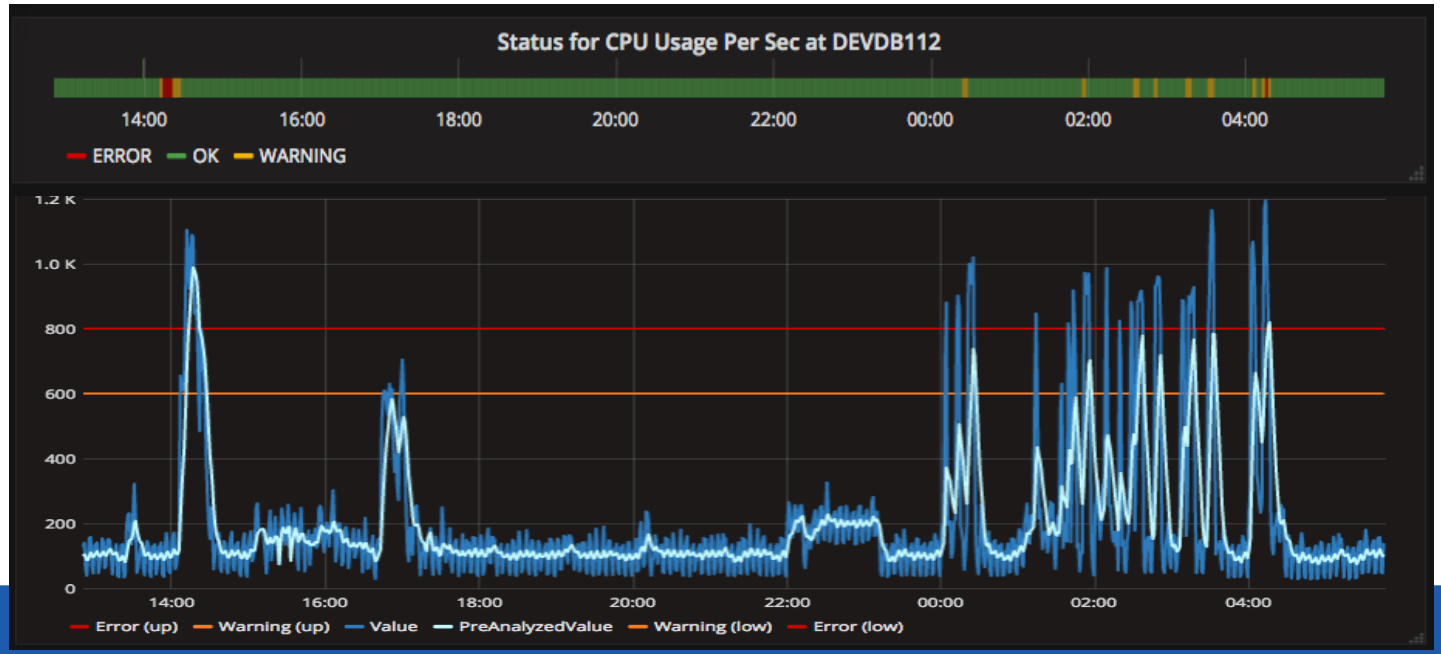
ExDeMon: extract, define and monitor



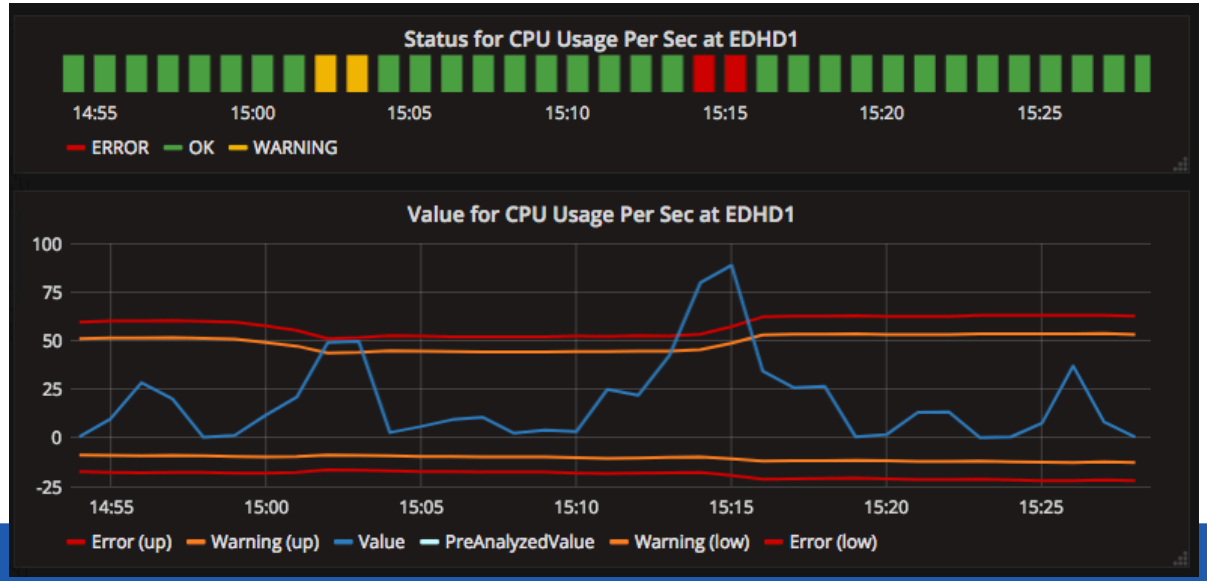
- Which metrics should process this monitor?
  - CPU Usage of all databases, Cache Hit % of MySQL databases, all machines in PROD, …
- Filter metrics by attributes
- Accepts exact value or regular expressions
- If not applied, all metrics are processed

- Is the metric behaving as it should?
- Determine the status of each incoming metric
  - OK, WARNING, ERROR, EXCEPTION
- Results can be sent to external storage
- Internal store for stateful analysis

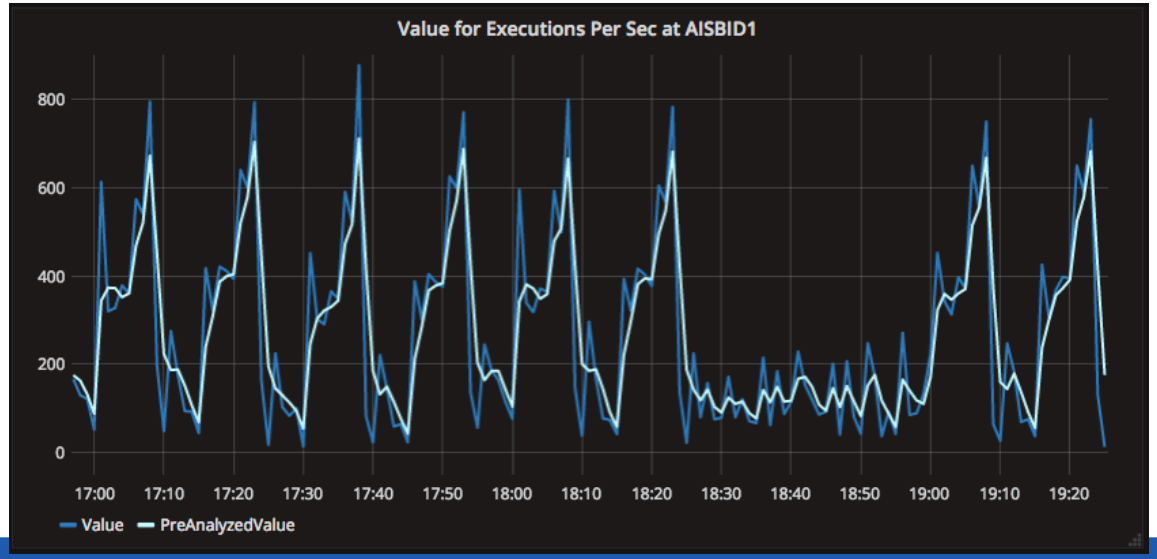
- Type: thresholds fixed manually



- Type: learning from recent behaviour (average and variance)
  - Errors when metric does not behave as it used to

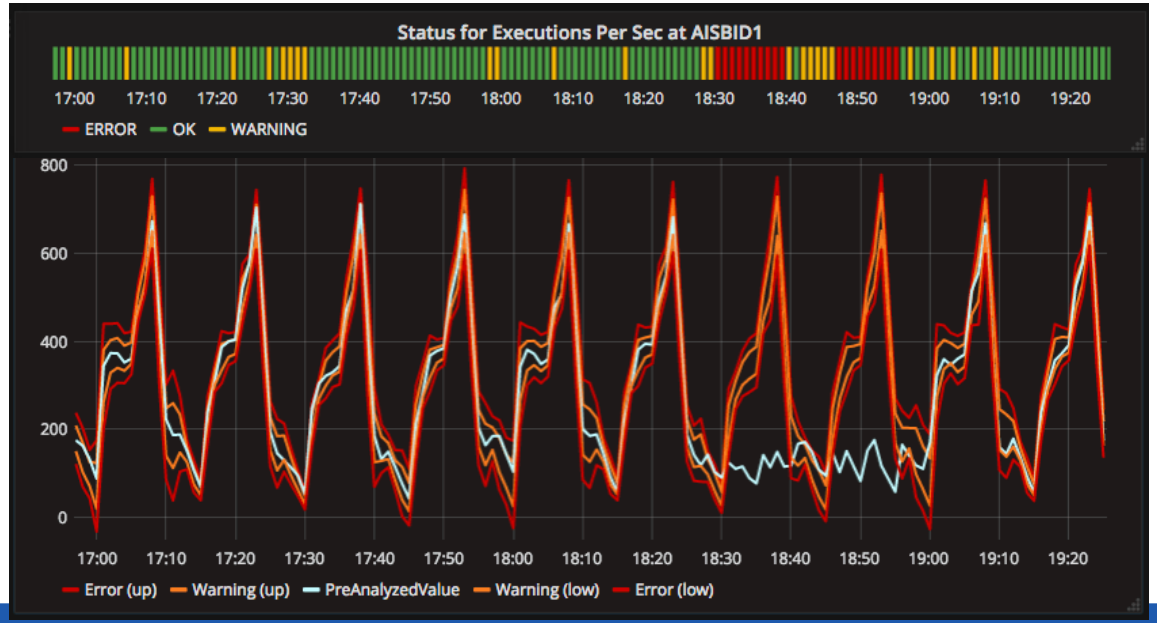


- Type: learning a season (hour, day, week)



- Type: learning a season (hour, day, week)

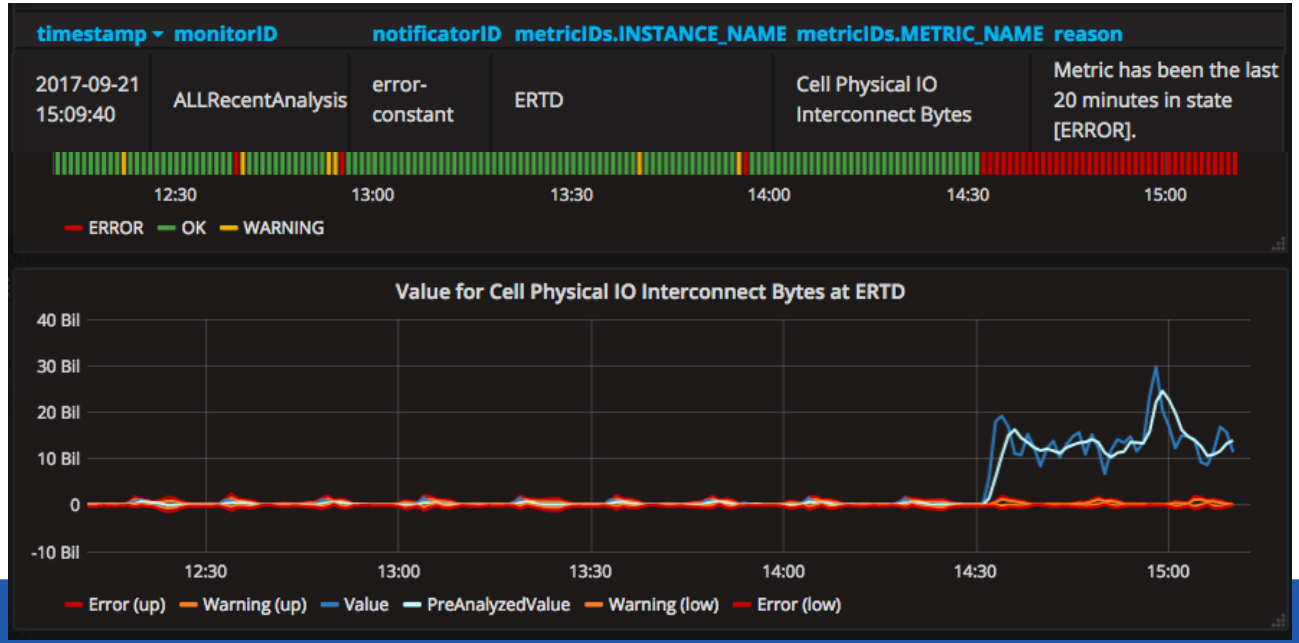
- Type: learning a season (hour, day, week)



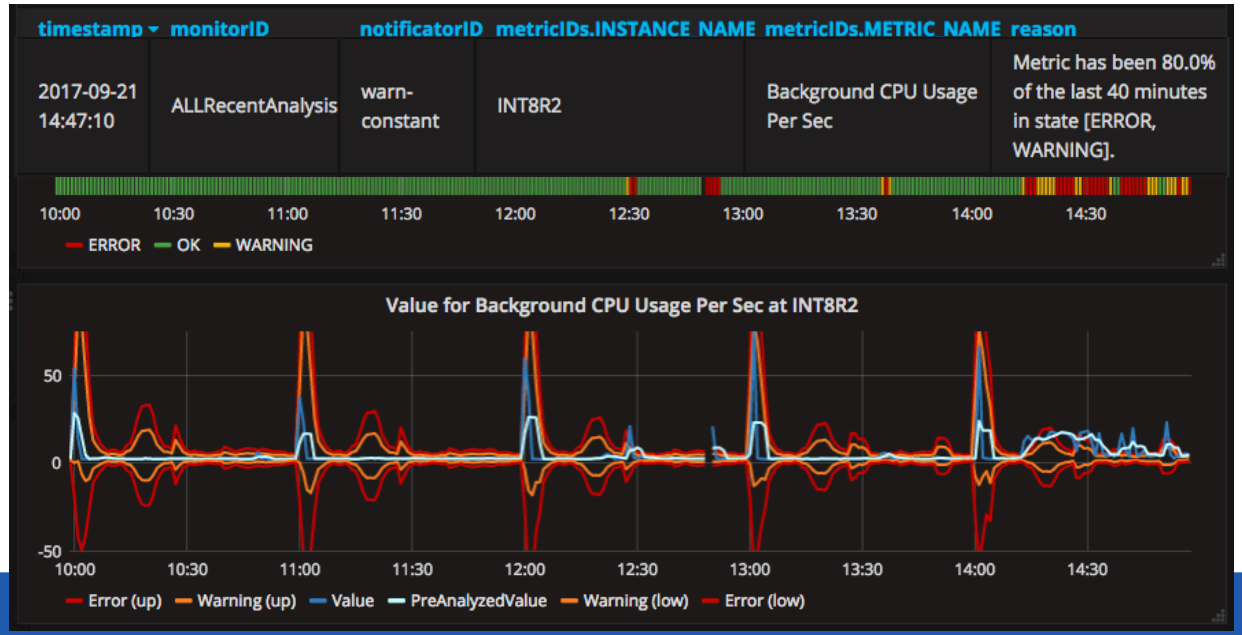
- Should we inform anyone?
  - Metric has been in ERROR for 20 minutes…

- Determines when an action is triggered
- Based on metric statuses
- Actions are processed by actuators

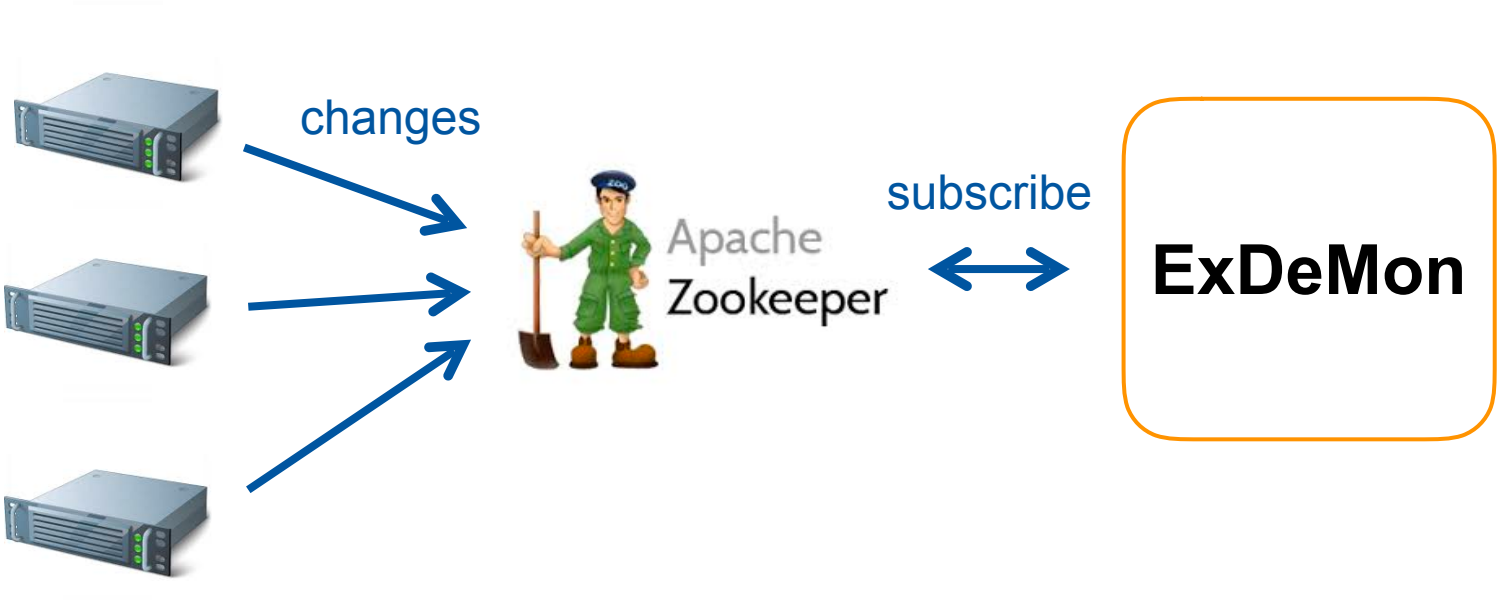
- Type: constant status for certain period



- Type: percentage of the period in certain statuses



ExDeMon: configuration

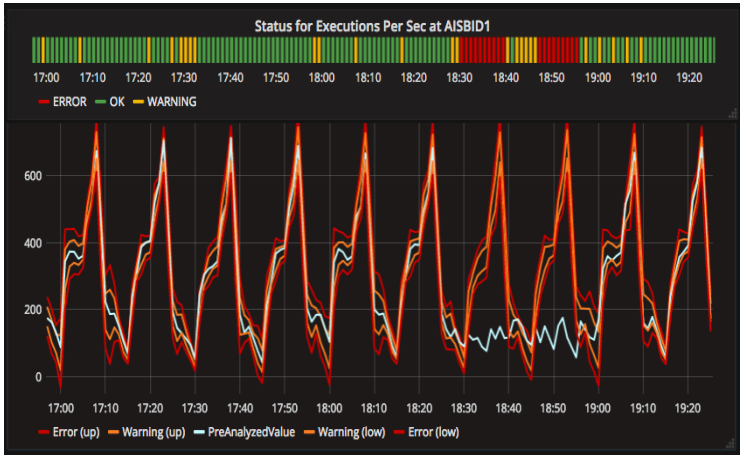


Configuration

puppet

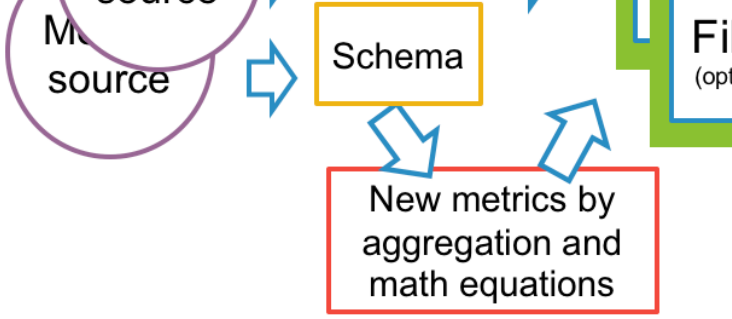

```
exdemon::monitor{ 'executions': }
```

```
monitor_executions:
  filter:
    attribute:
      HOSTNAME: .*
      METRIC_NAME: Read Bytes Per Sec
  analysis:
    type: seasonal
    season: hour
  triggers:
    ew-perc:
      type: percentage
      statuses: ERROR WARNING
      percentage: 80
      period: 10m
      actuators: email rundeck
```



CERN

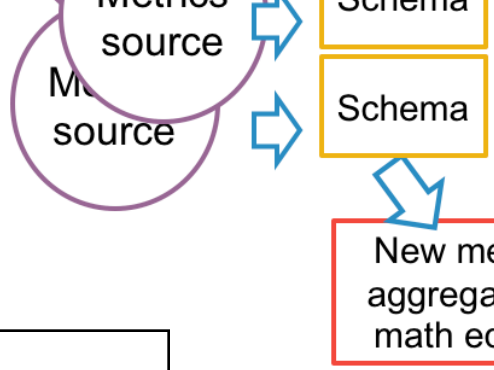
Defined metrics



- Configuration can be updated while running
  - Add/remove defined metrics or change any parameter

- Aggregations
  - Different metrics
  - Value along time

Defined metrics configuration



- Ratio read/write per instance

Metric's attributes
- HOSTNAME
- DISK
- METRIC_NAME

```
metric_disk_ratio:
  metrics:
    groupby: HOSTNAME DISK
  value: readbytes / writebytes
  variables:
    readbytes:
      filter.attribute.METRIC_NAME: Read Bytes Per Sec
    writebytes:
      filter.attribute.METRIC_NAME: Write Bytes Per Sec
```

Results will be a metric per HOSTNAME and DISK with:
- HOSTNAME
- DISK
- Value = "`Read Bytes Per Sec`" / "`Write Bytes Per Sec`"
  - Values from the same instance
  - Each variable updated when new value arrives

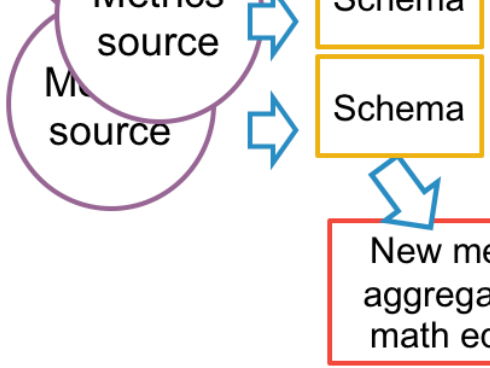
Defined metrics configuration

- Count machines running per cluster

Metric's attributes
- HOSTNAME
- CLUSTER_NAME
- METRIC_NAME

```
metric_cluster_machines:
  metrics:
    groupby: CLUSTER_NAME
  when: BATCH
  variables:
    value:
      filter.attribute.METRIC_NAME: Status
      aggregate: count
      attributes: HOSTNAME
      expire: 5m
```

Results will be a metric per CLUSTER_NAME with:
- CLUSTER_NAME
- Value = count(diff(HOSTNAME))
  - New metrics with same HOSTNAME are updated.
  - If not updated after 5 minutes, they are removed.

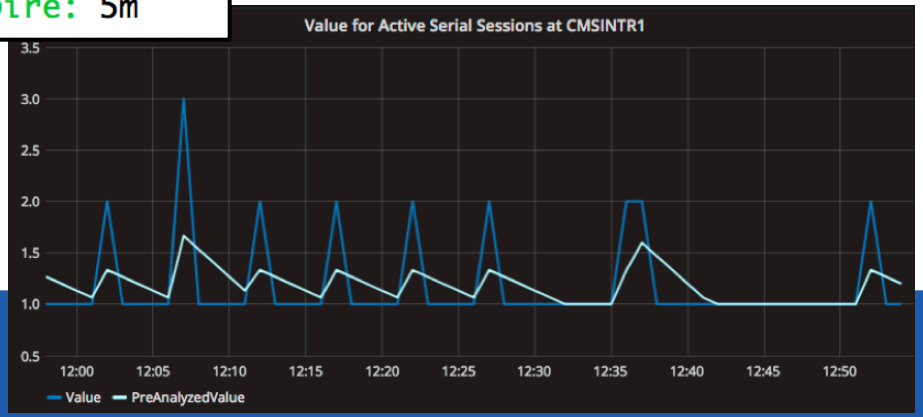


Defined metrics configuration

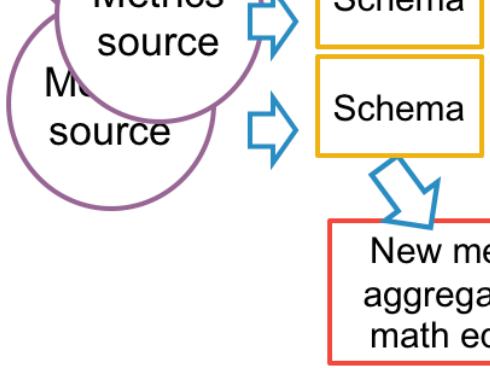
- Average value along time for each metric

Metric's attributes
- HOSTNAME
- DISK
- METRIC_NAME

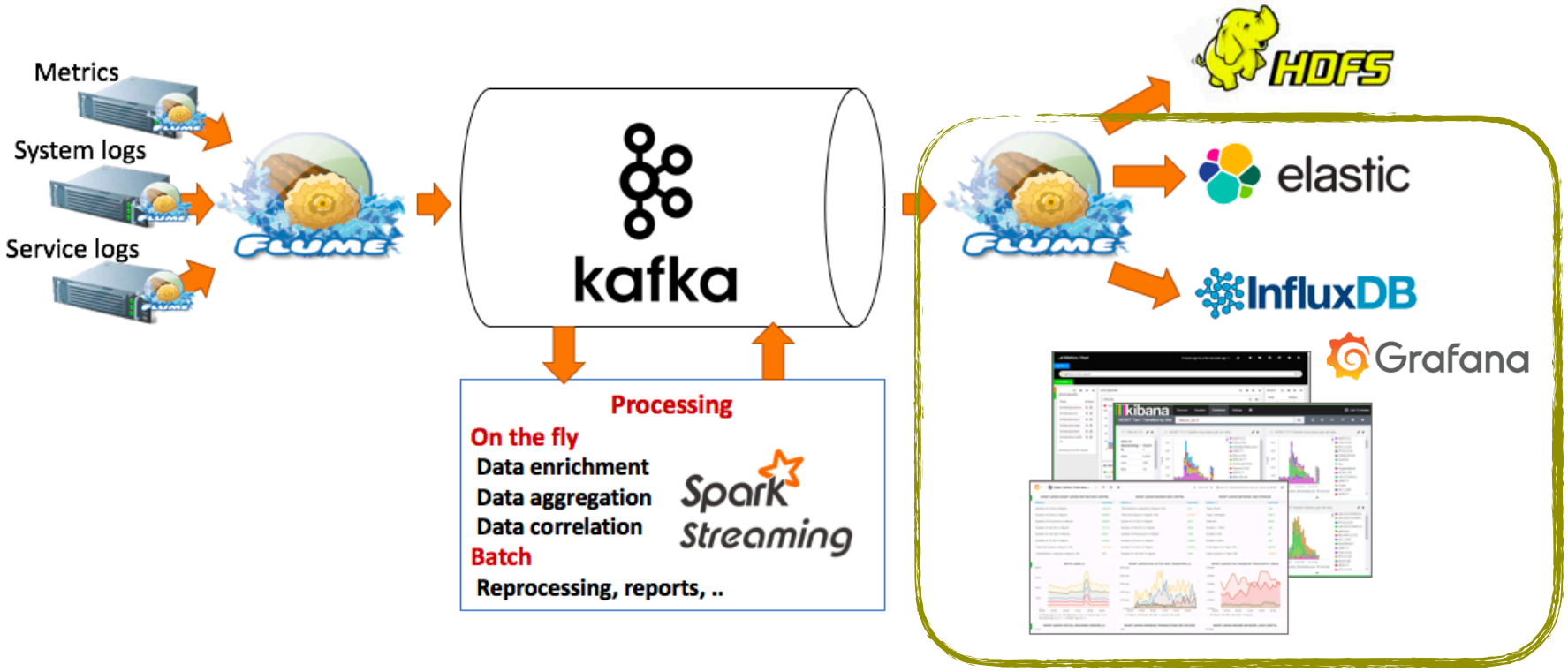
```
1 metric_avg_5m:
2   metrics:
3     groupby: ALL
4   variables:
5     value:
6       aggregate: avg
7       expire: 5m
```

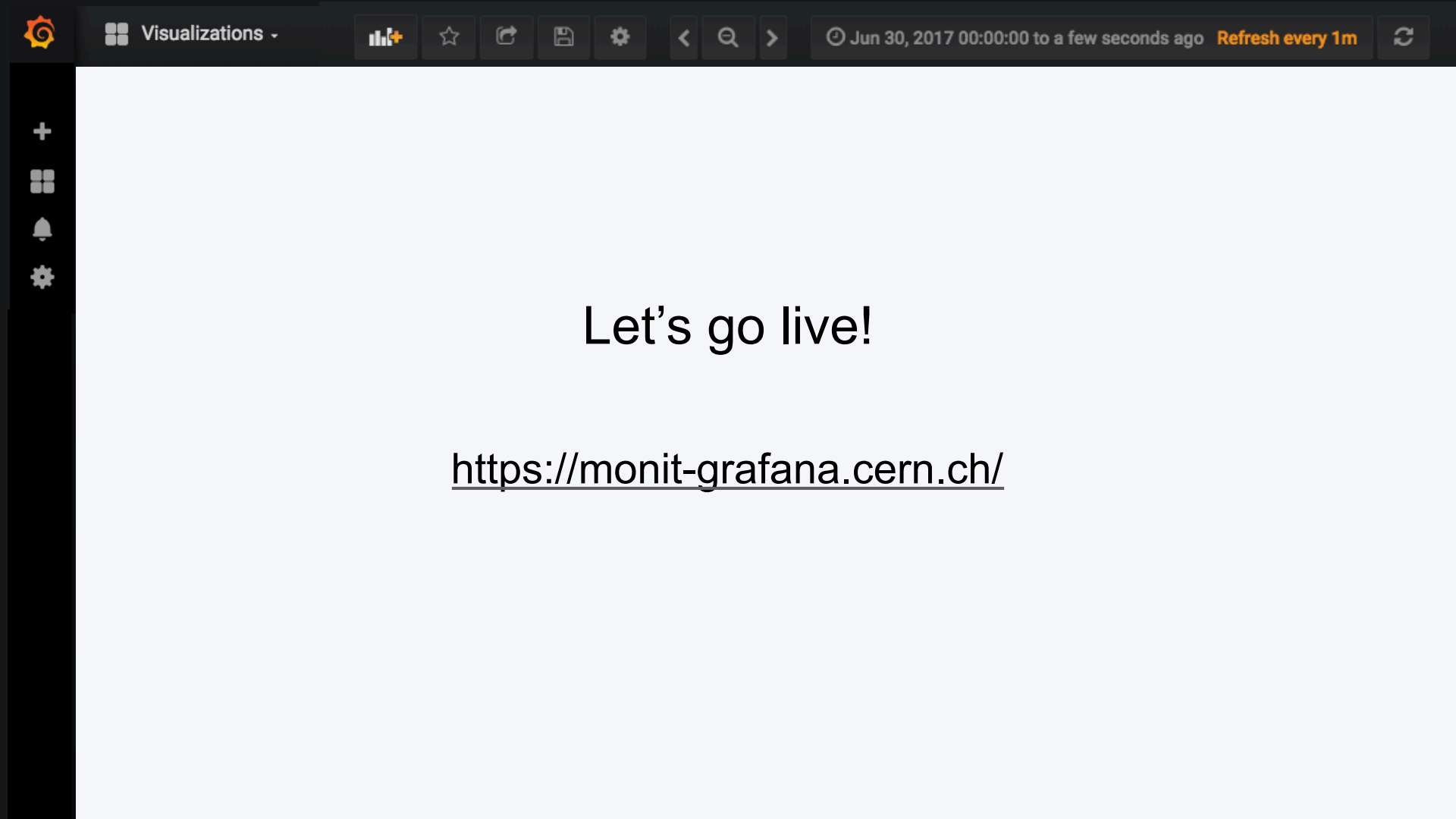


Results will be a metric per incoming metric with:
- HOSTNAME
- DISK
- METRIC_NAME
- Value = avg(values from the last 5 minutes)



IT monitoring architecture



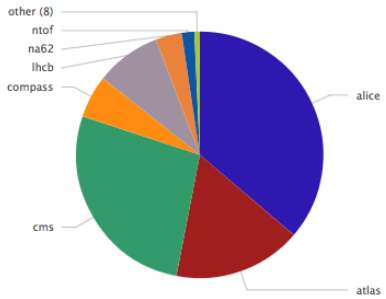


Let’s go live!

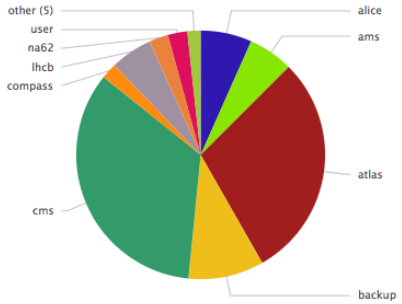
https://monit-grafana.cern.ch/

Splunk

Transfered Data Amount per Virtual Organization for READ Requests

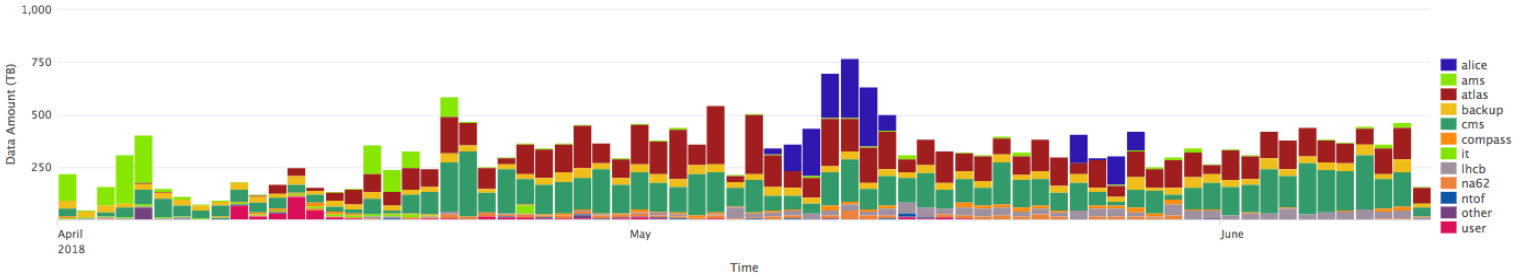


Transfered Data Amount per Virtual Organization for WRITE Requests

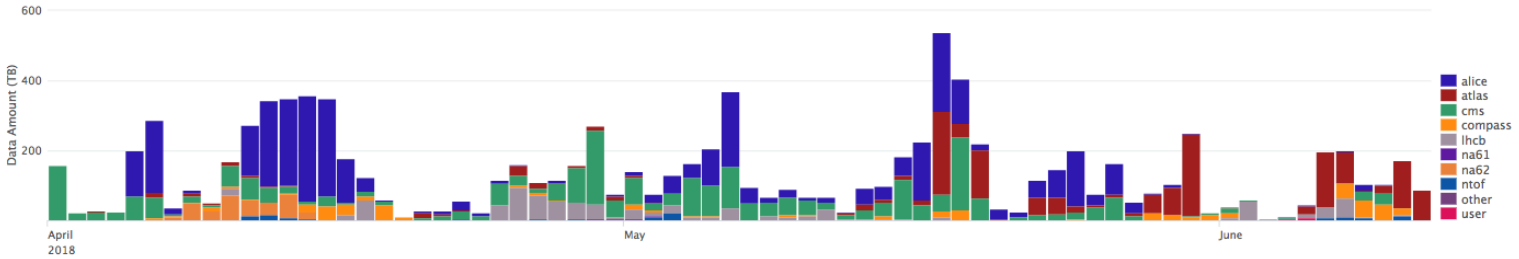


2 minutes ago

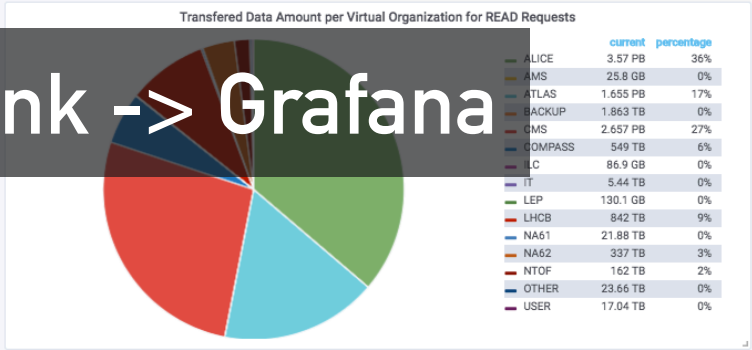
Transfered Data Amount per Virtual Organization for WRITE Requests

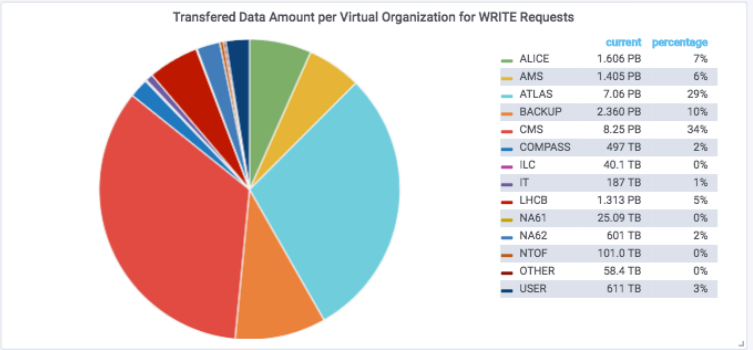


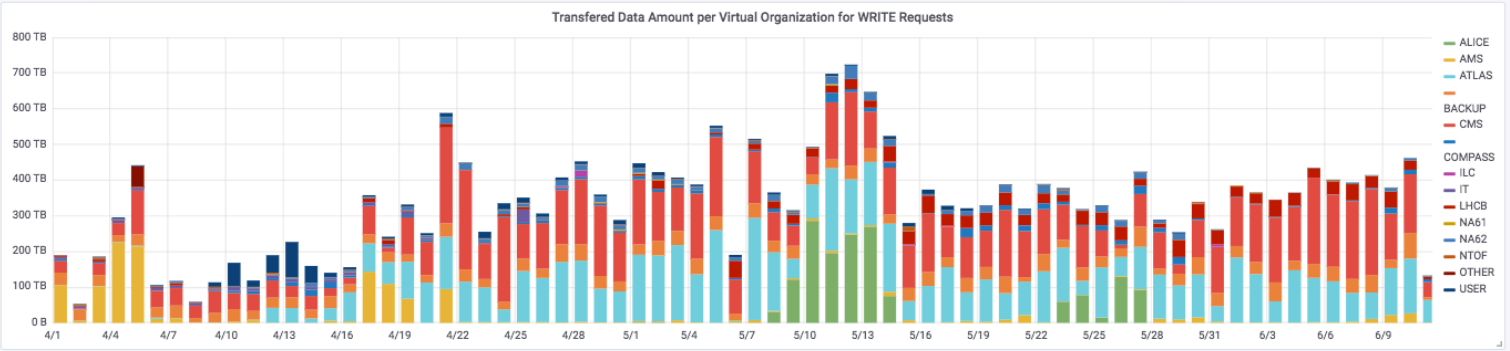
Transfered Data Amount per Virtual Organization for READ Requests

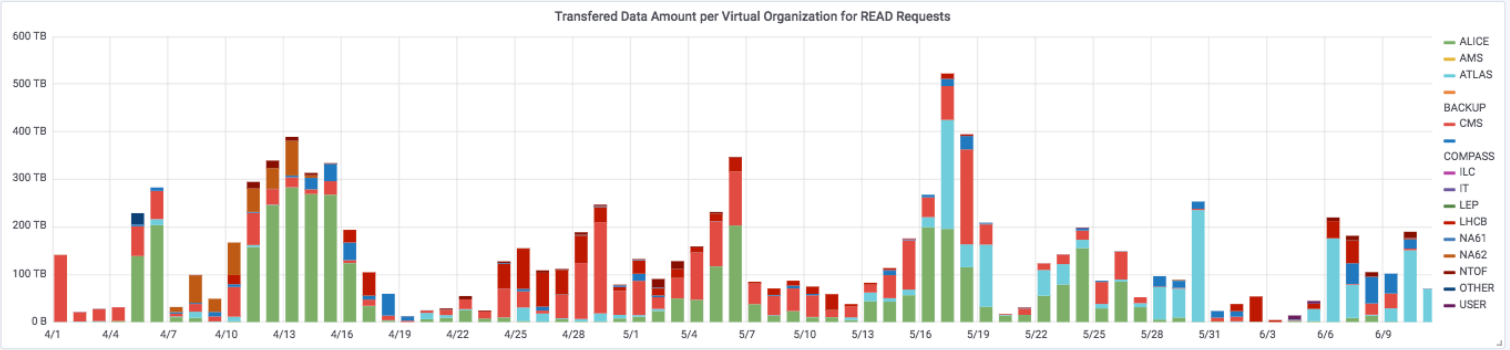


Splunk -> Grafana

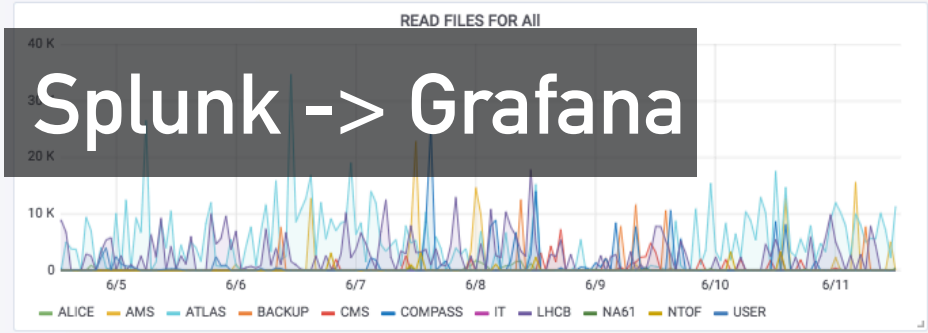


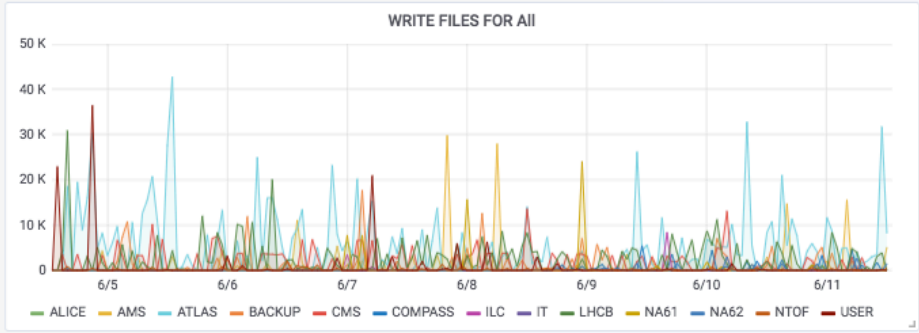


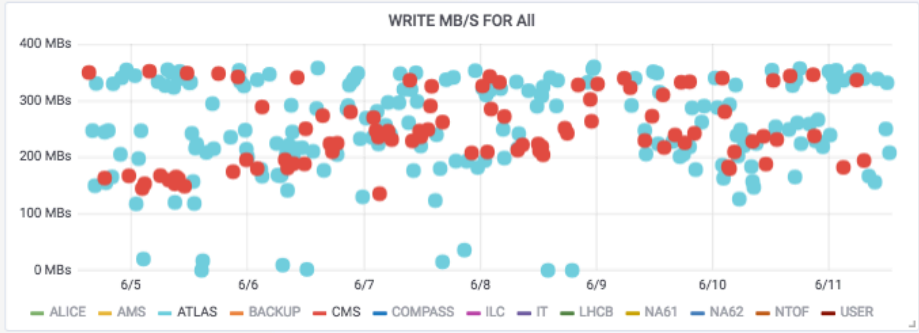


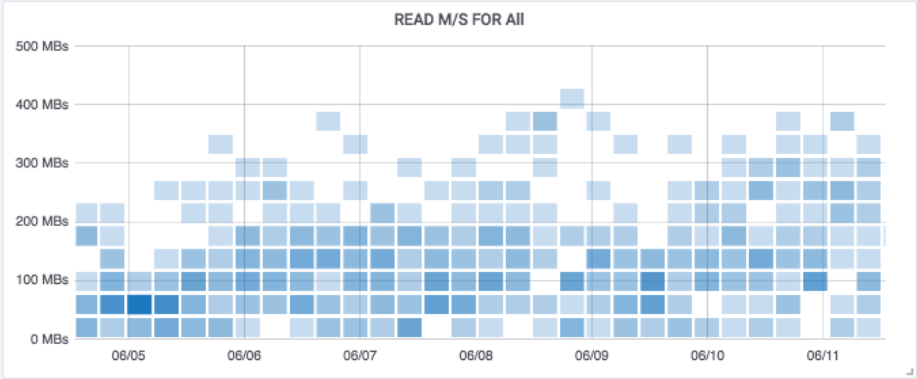


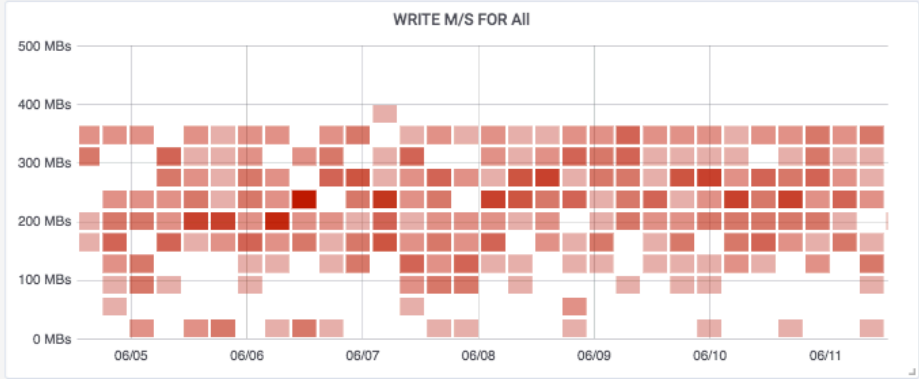
Splunk -> Grafana



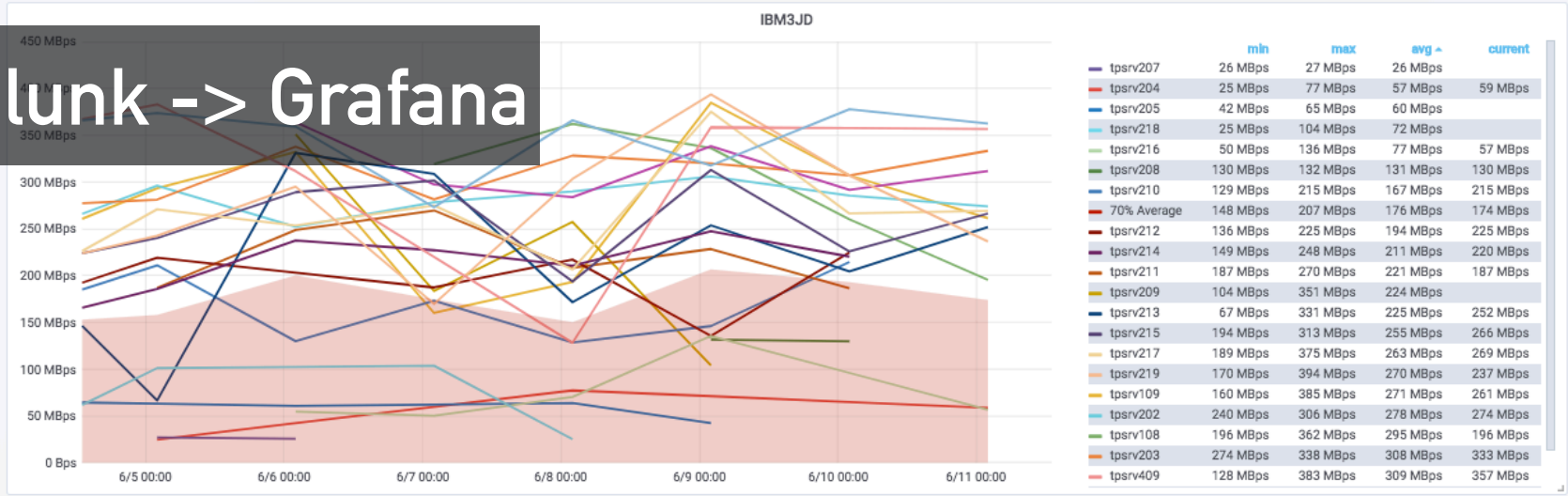




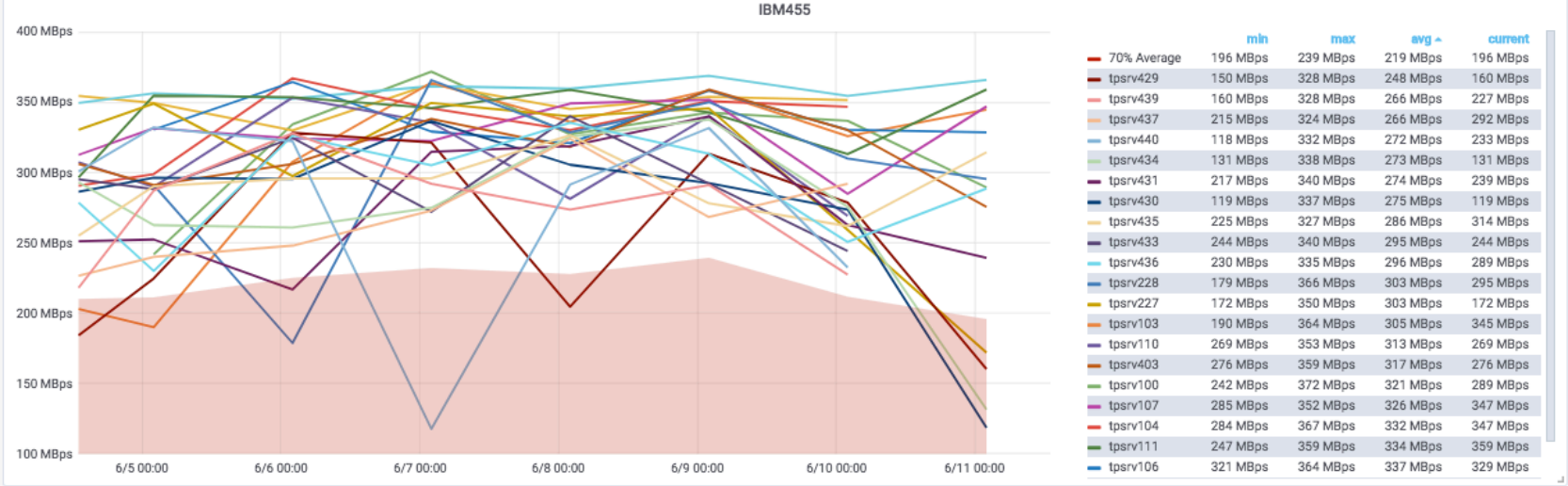




Splunk -> Grafana

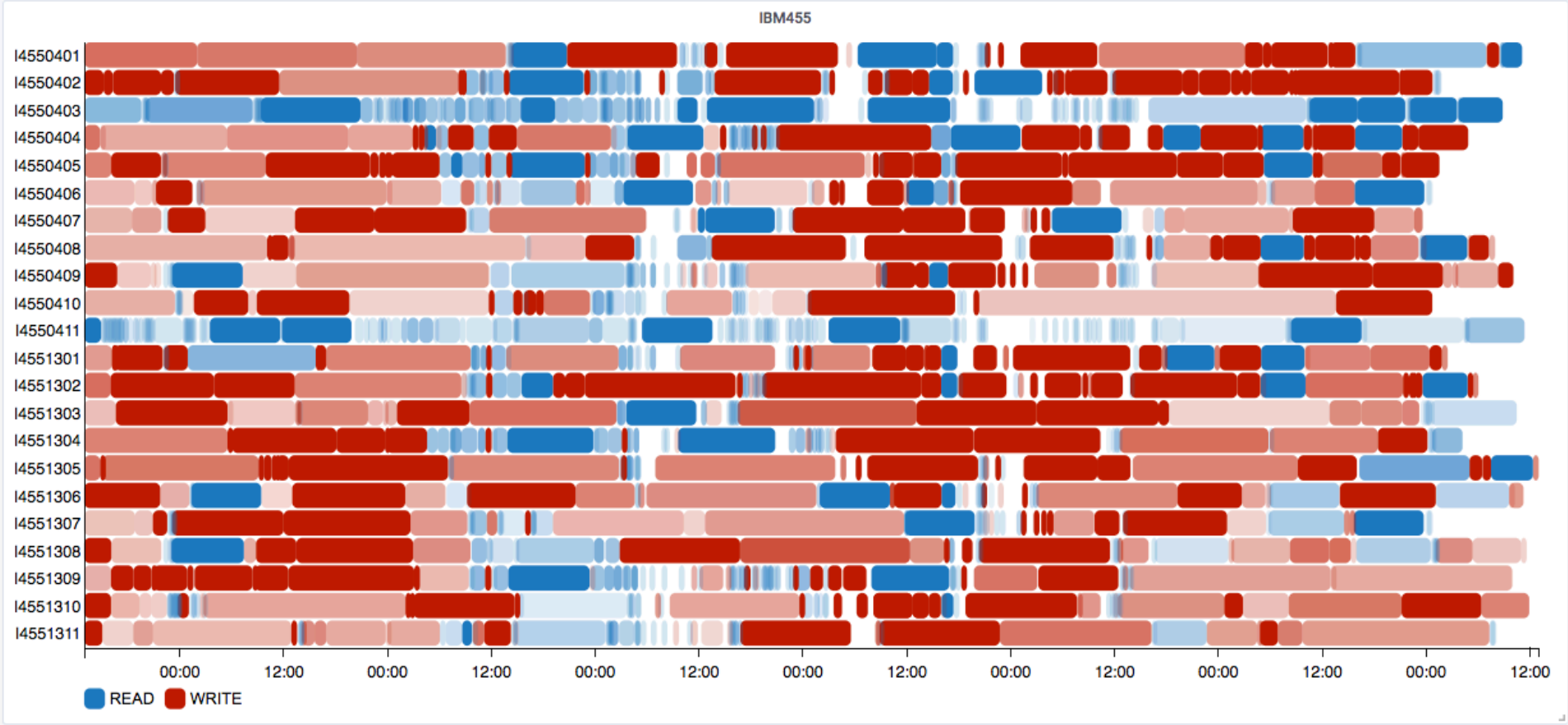


	min	max	avg	current
tpsrv207	26 MBps	27 MBps	26 MBps	
tpsrv204	25 MBps	77 MBps	57 MBps	59 MBps
tpsrv205	42 MBps	65 MBps	60 MBps	
tpsrv218	25 MBps	104 MBps	72 MBps	
tpsrv216	50 MBps	136 MBps	77 MBps	57 MBps
tpsrv208	130 MBps	132 MBps	131 MBps	130 MBps
tpsrv210	129 MBps	215 MBps	167 MBps	215 MBps
70% Average	148 MBps	207 MBps	176 MBps	174 MBps
tpsrv212	136 MBps	225 MBps	194 MBps	225 MBps
tpsrv214	149 MBps	248 MBps	211 MBps	220 MBps
tpsrv211	187 MBps	270 MBps	221 MBps	187 MBps
tpsrv209	104 MBps	351 MBps	224 MBps	
tpsrv213	67 MBps	331 MBps	225 MBps	252 MBps
tpsrv215	194 MBps	313 MBps	255 MBps	266 MBps
tpsrv217	189 MBps	375 MBps	263 MBps	269 MBps
tpsrv219	170 MBps	394 MBps	270 MBps	237 MBps
tpsrv109	160 MBps	385 MBps	271 MBps	261 MBps
tpsrv202	240 MBps	306 MBps	278 MBps	274 MBps
tpsrv108	196 MBps	362 MBps	295 MBps	196 MBps
tpsrv203	274 MBps	338 MBps	308 MBps	333 MBps
tpsrv409	128 MBps	383 MBps	309 MBps	357 MBps

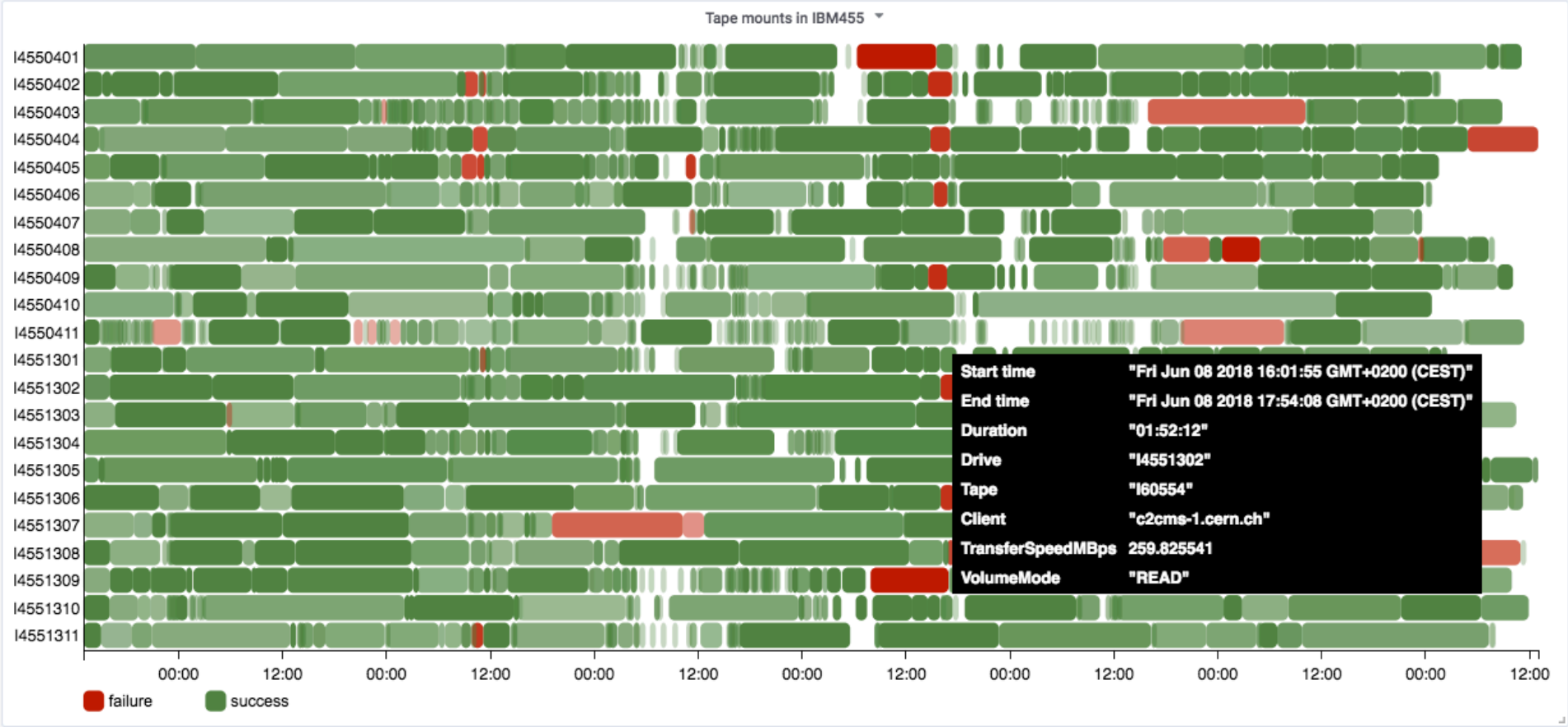


	min	max	avg	current
70% Average	196 MBps	239 MBps	219 MBps	196 MBps
tpsrv429	150 MBps	328 MBps	248 MBps	160 MBps
tpsrv439	160 MBps	328 MBps	266 MBps	227 MBps
tpsrv437	215 MBps	324 MBps	266 MBps	292 MBps
tpsrv440	118 MBps	332 MBps	272 MBps	233 MBps
tpsrv434	131 MBps	338 MBps	273 MBps	131 MBps
tpsrv431	217 MBps	340 MBps	274 MBps	239 MBps
tpsrv430	119 MBps	337 MBps	275 MBps	119 MBps
tpsrv435	225 MBps	327 MBps	286 MBps	314 MBps
tpsrv433	244 MBps	340 MBps	295 MBps	244 MBps
tpsrv436	230 MBps	335 MBps	296 MBps	289 MBps
tpsrv228	179 MBps	366 MBps	303 MBps	295 MBps
tpsrv227	172 MBps	350 MBps	303 MBps	172 MBps
tpsrv103	190 MBps	364 MBps	305 MBps	345 MBps
tpsrv110	269 MBps	353 MBps	313 MBps	269 MBps
tpsrv403	276 MBps	359 MBps	317 MBps	276 MBps
tpsrv100	242 MBps	372 MBps	321 MBps	289 MBps
tpsrv107	285 MBps	352 MBps	326 MBps	347 MBps
tpsrv104	284 MBps	367 MBps	332 MBps	347 MBps
tpsrv111	247 MBps	359 MBps	334 MBps	359 MBps
tpsrv106	321 MBps	364 MBps	337 MBps	329 MBps

Grafana: new Gantt panel



Grafana: new Gantt panel



Wrap-up

- Done
  - Logs and metrics collection system: simplified, homogeneous and performance
  - Alerting with ExDeMon has been working for months (gathering confidence)
  - Plots on Splunk replicated in Grafana
    - New Gantt panel
    - More than a year of old data imported, expect to finish importing this month
- Future steps
  - After validate the dashboards, shut down Splunk
  - Integrate alerting with email and actions on tape/drives
  - Remove Oracle databases from the equation
  - Spread ExDeMon to other services along the department?
- Interest in any component?
  - Easy! pass by 31/2-02 or drop an email at daniel.lanza@cern.ch