



BigData tools for the monitoring of the ATLAS EventIndex

**Evgeny Alexandrov¹, Andrei Kazymov¹, Fedor
Prokoshin², on behalf of the ATLAS
collaboration**

¹Joint Institute for Nuclear Research, Dubna, Russia.

²Centro Científico Tecnológico de Valparaíso-CCTVal,
Universidad Técnica Federico Santa María.

GRID Conference at JINR
11 September 2018

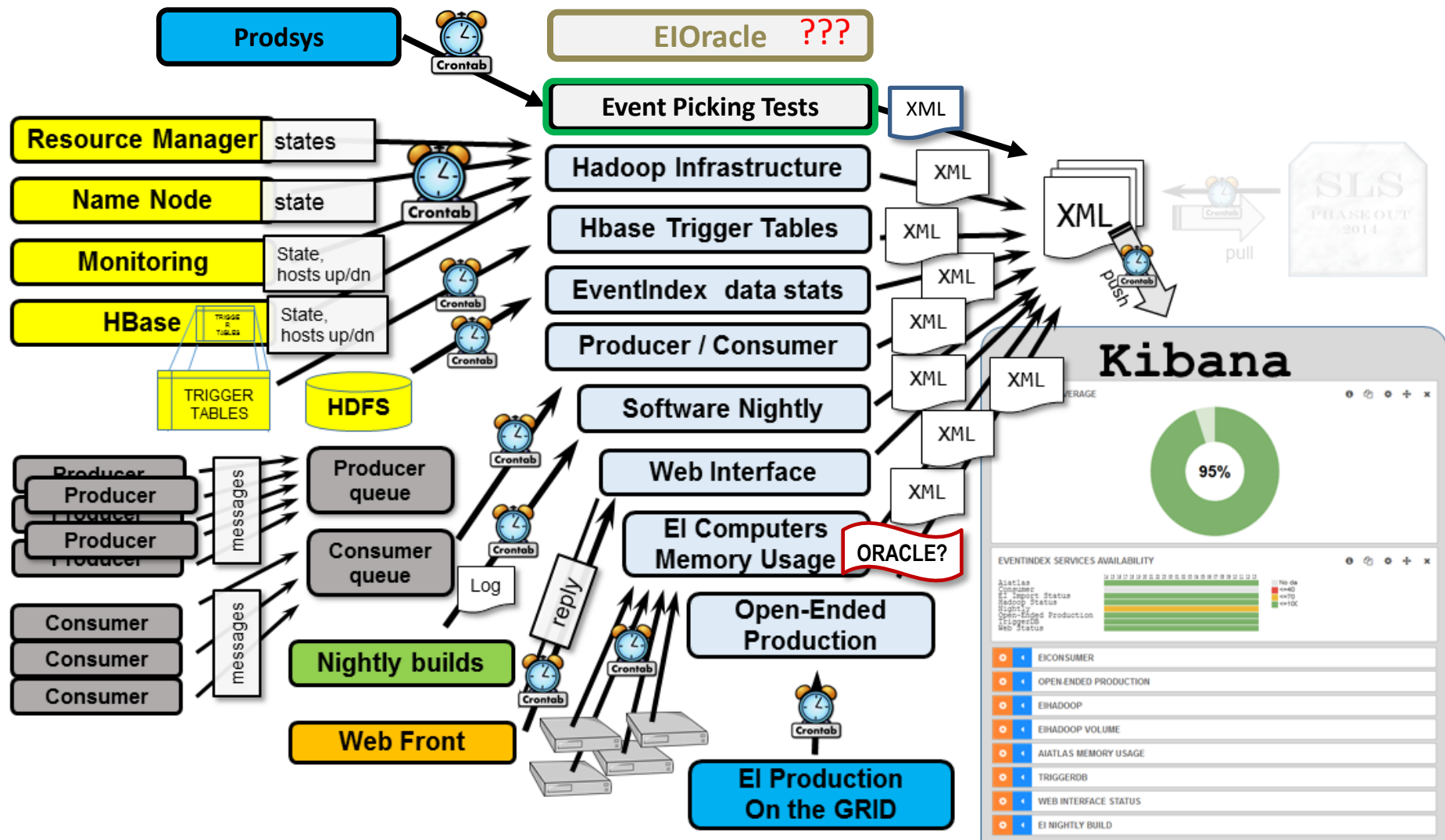


Introduction

- The EventIndex is the complete catalogue of all ATLAS events, keeping the references to all files that contain a given event in any processing stage.
- The ATLAS EventIndex collects event information from data both at CERN and Grid sites.
- It uses the Hadoop system to store the results, and web services to access them.
- Its successful operation depends on a number of different components.
- Each component has completely different sets of parameters and states and requires a special approach.

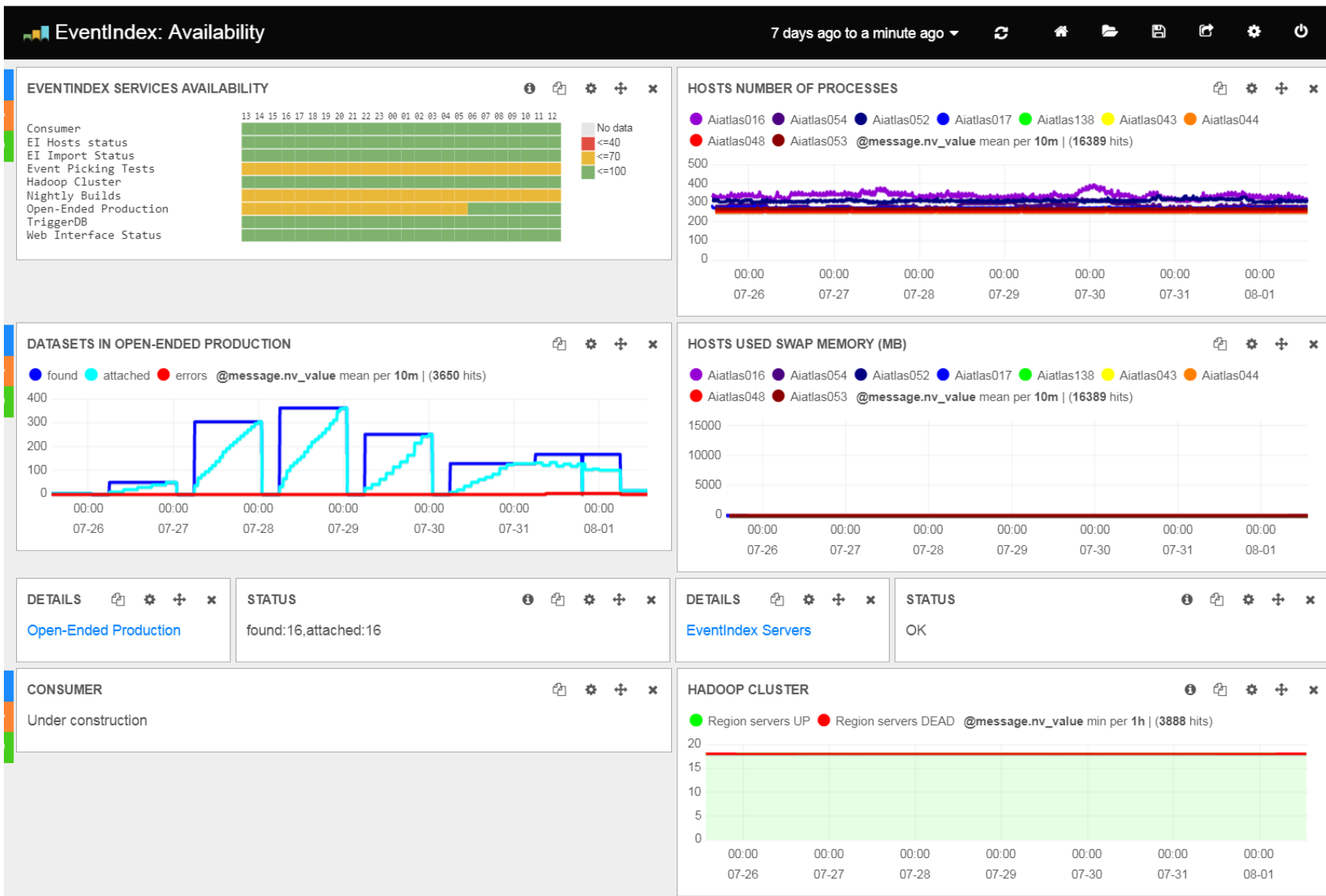


Monitoring Components





Old Monitoring System based on Kibana





Disadvantages of Kibana

Slow dashboard retrieving time:

- for two days period: 15 seconds;
- for 7 days period: 1 minute 30 seconds;
- for a longer periods: it may take tens of minutes and eventually get stuck.

Not very comfortable way of editing the dashboard's page



Grafana

Grafana is one of the most popular packages for visualizing monitoring data.

Uses modern technologies:

- back-end is written using Go programming language;
- front-end is written on typescript and uses angular approach.

The following datasources are officially supported:

- Graphite
- Elasticsearch
- CloudWatch
- InfluxDB
- OpenTSDB
- Prometheus
- MySQL
- Postgres
- Microsoft SQL Server (MSSQL)



InfluxDB

InfluxDB is InfluxData's open source time series database designed to handle high write and query loads.

Uses modern technologies:

- it is written on GO;
- It has the possibility of working in cluster mode;
- availability of libraries for a large number of programming languages (Python, JavaScript, PHP, Haskell and others);
- SQL-like query language, with which you can perform various operations with time series (merging, splitting).

CERN IT Monitoring team supports Grafana + InfluxDB.



Common structure of monitoring

Script for storing data

```
Python Shell
File Edit Shell Debug Options Windows Help
Python 3.2.5 (default, May 15 2013, 21:04:03) [AMD64 v1500 32 bit (Intel)] on win...
>>>
Type "copyright", "credits" or "license()" for more information.
>>> import random
>>> random.randint(1,20)
5
>>> random.randint(1,20)
18
>>> random.randint(1,20)
18
>>> |
```

HTTP



All data presented as JSON and structured for Monit service

Monit service for collecting data



HTTP ?



HTTP



Grafana

HTTP



User





Monitoring Components



	Component	Producer	Status	Description
1	Open-Ended Production	atlevindtc.py	Done	
2	Consumer	atlevindcon.py	None*	No info for new version
3	EI Import Status	atlevindimp.py	Done	
4	Hadoop Cluster	atlevindhad.py	Done	
5	TriggerDB	atlevindtdb.py	Done	
6	Web Interface Status	atlevindweb.py	Done	
7	Health of EventIndex computers	InfluxDB	Done	
8	Nightly Builds	atlevindsoft.py	Done	
9	Event Picking Tests	atlevindpic.py	Done	
10	EI Data Volumes	atlevindvol.py	Done	
	<i>Main Page</i>	<i>All</i>	<i>Done</i>	

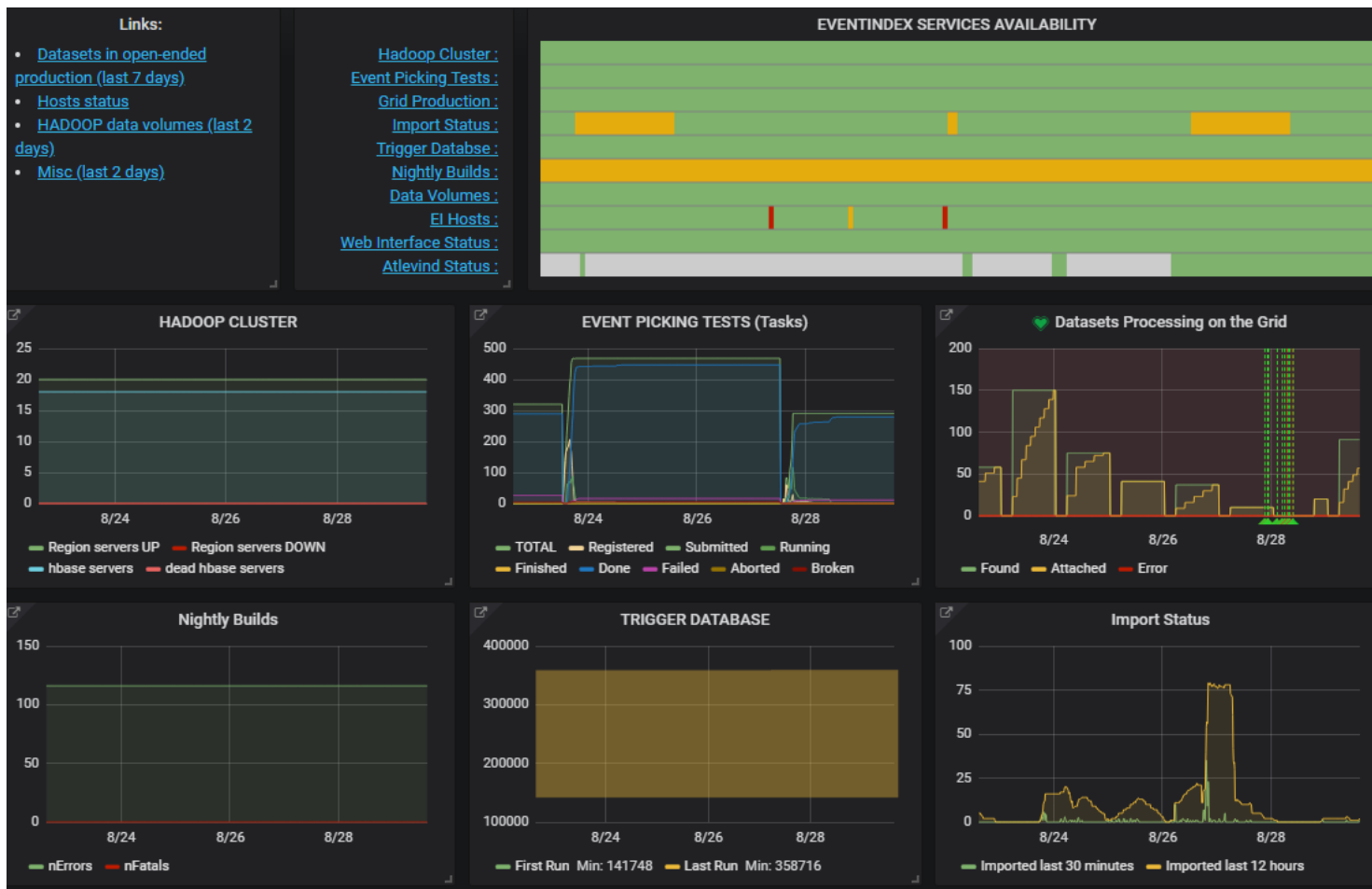
*Component was completely re-created on a new platform, no monitoring available yet.



Main Page(1)



“Home”->select “Atlas EventIndex”

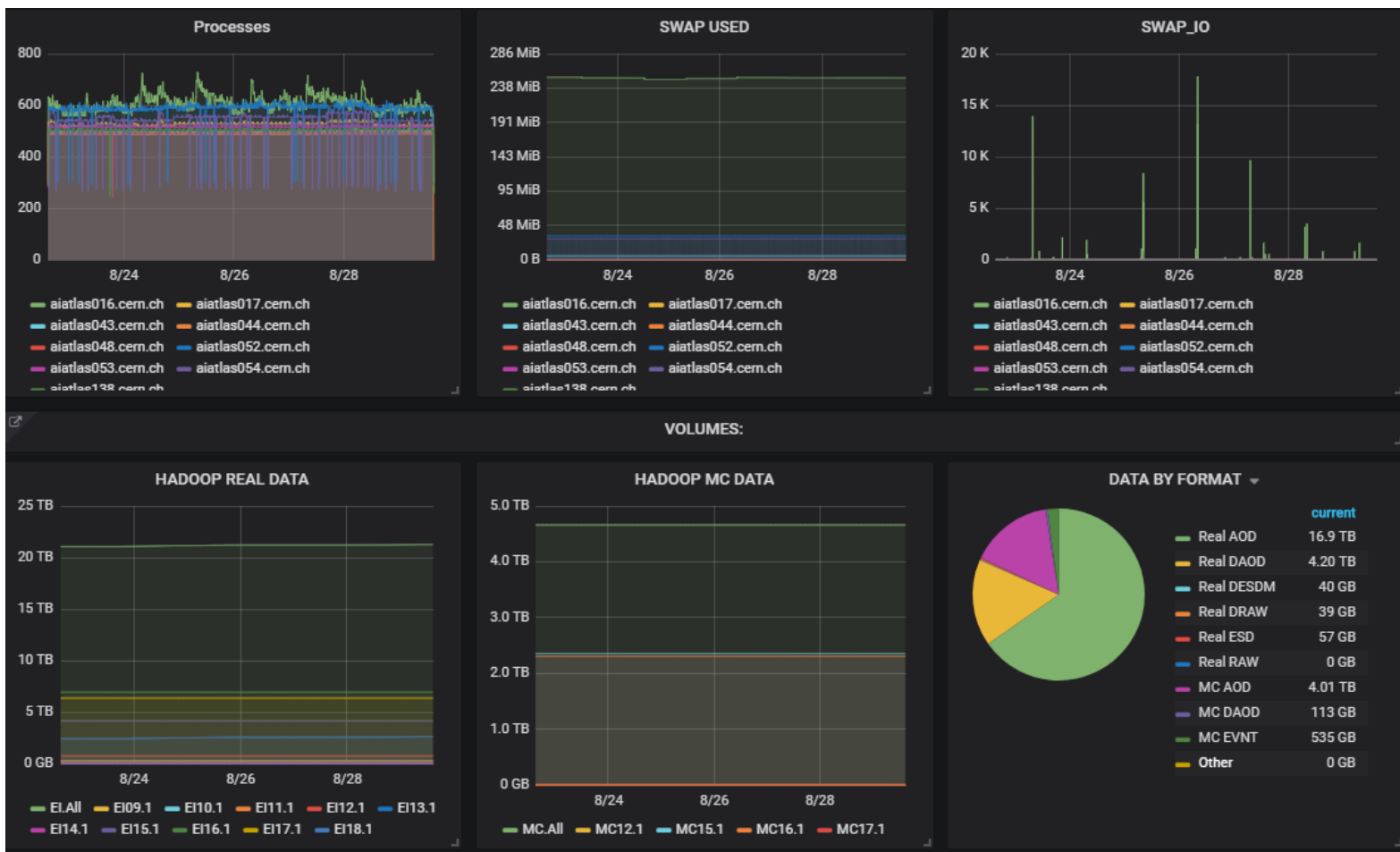


URL: <https://monit-grafana.cern.ch/d/RBMMmOnmz/atlas-eventindex?orgId=6>

Requires CERN SSO Sign-in



Main Page(2)



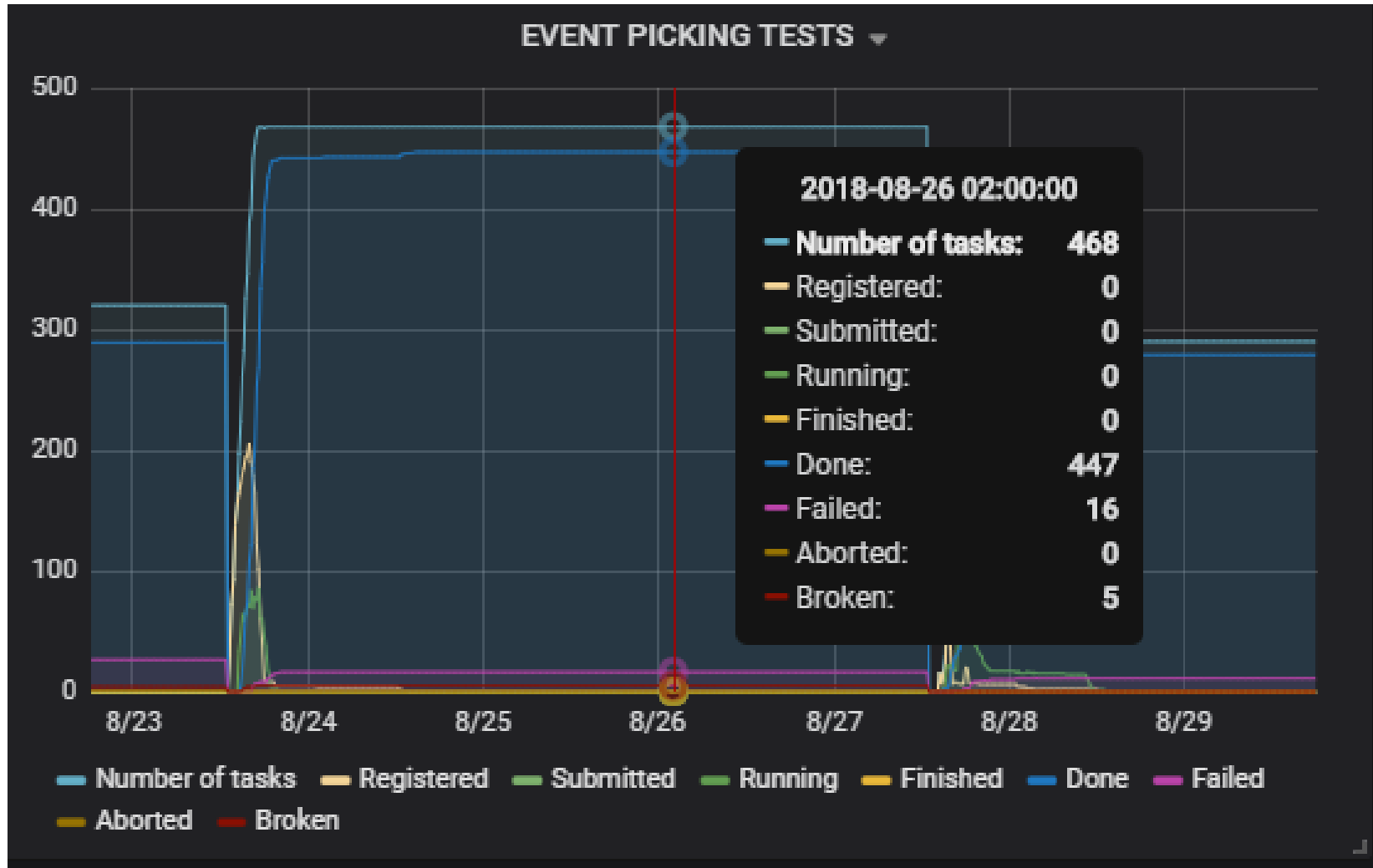


Hadoop Cluster



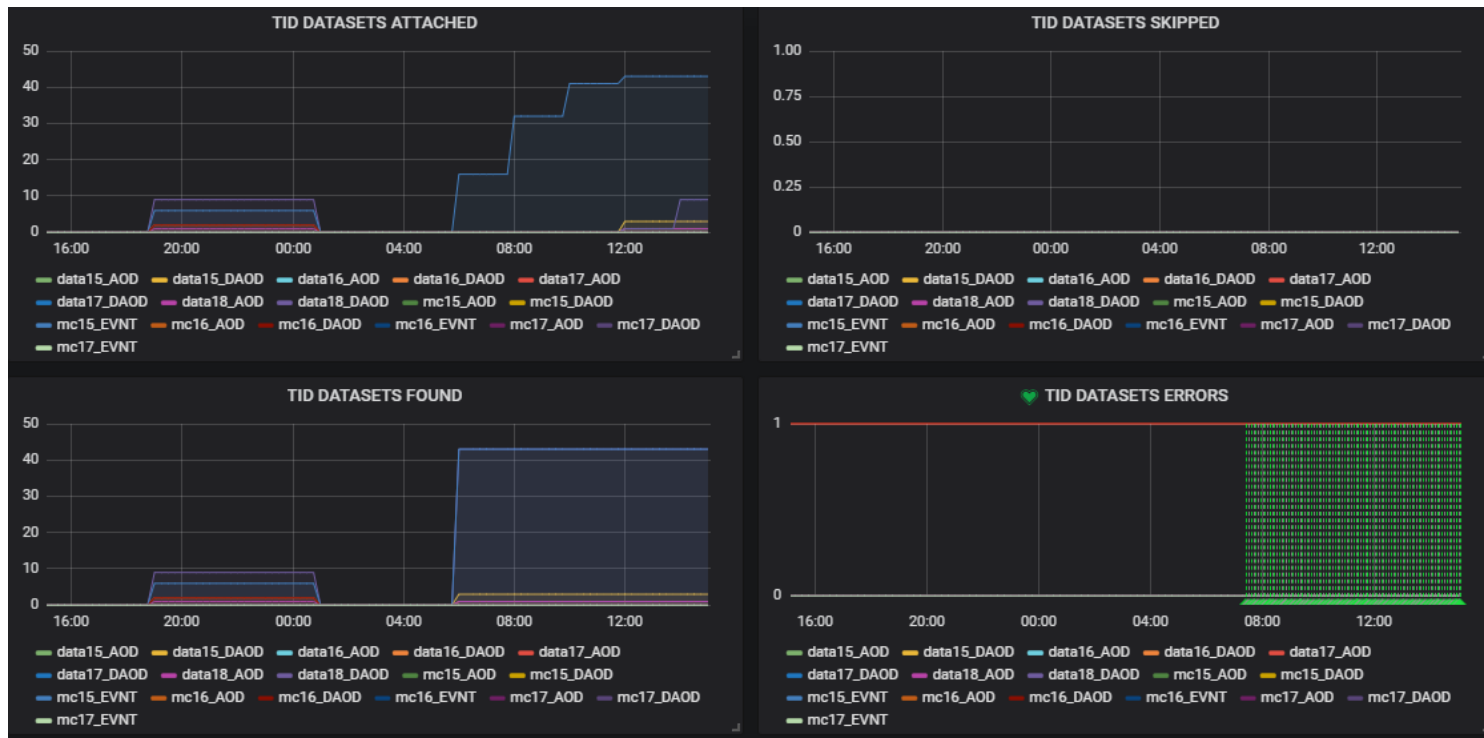


Event Picking Tests





EI production on the GRID and import

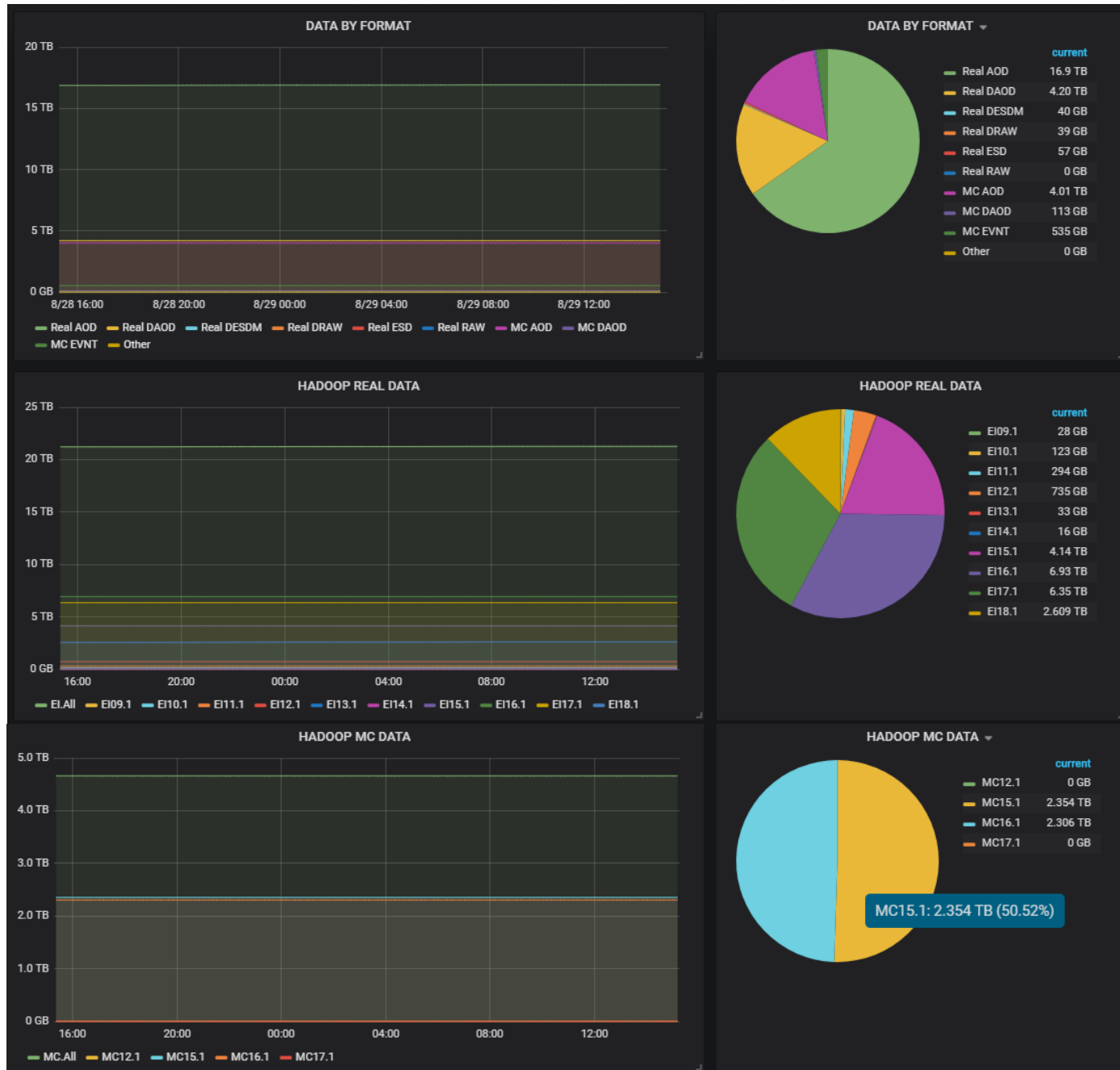


EI Import Status





Data Volumes





Health of EventIndex computers structure of monitoring

Script for storing data

```
Python Shell
File Edit Shell Debug Options Window Help
Python 3.2.5 (default, May 15 2013, 21:04:03) [AMD64 v1500 32 bit (Intel)] on win...
>>>
Type "copyright", "credits" or "license()" for more information.
>>> import random
>>> random.randint(1,20)
5
>>> random.randint(1,20)
18
>>> random.randint(1,20)
18
>>> |
```

HTTP



All data presented as JSON and structured for Monit service

Monit service for collecting data



Implemented by Monit group and is invisible for us

HTTP ?



User

HTTP



Grafana

HTTP

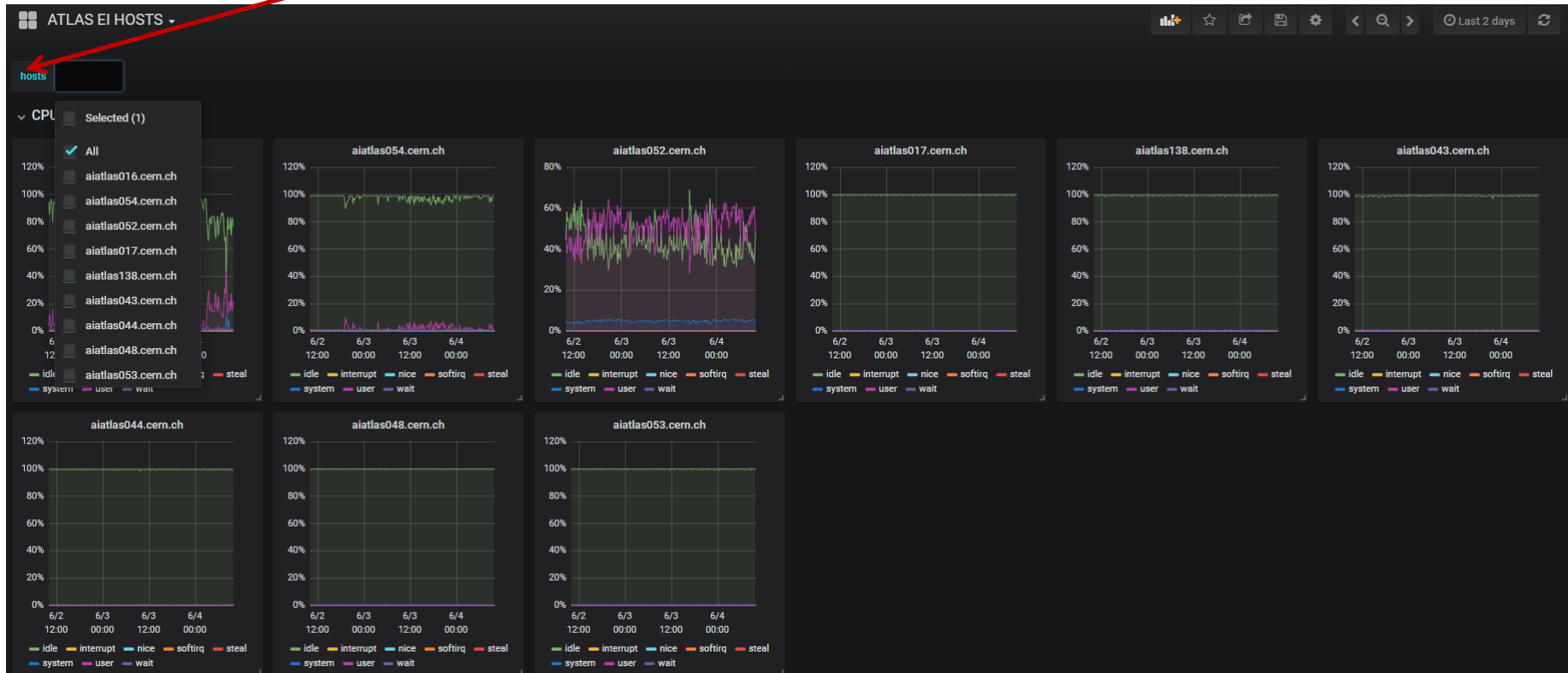


InfluxDB



Health of EventIndex computers

Select host





Conclusion

- After half year of development, renovated and improved monitoring services of the ATLAS EventIndex were successfully implemented, tested and put in production.
- Development efforts will continue, to adapt for possible configuration changes and new components support.

Thanks for your attention!