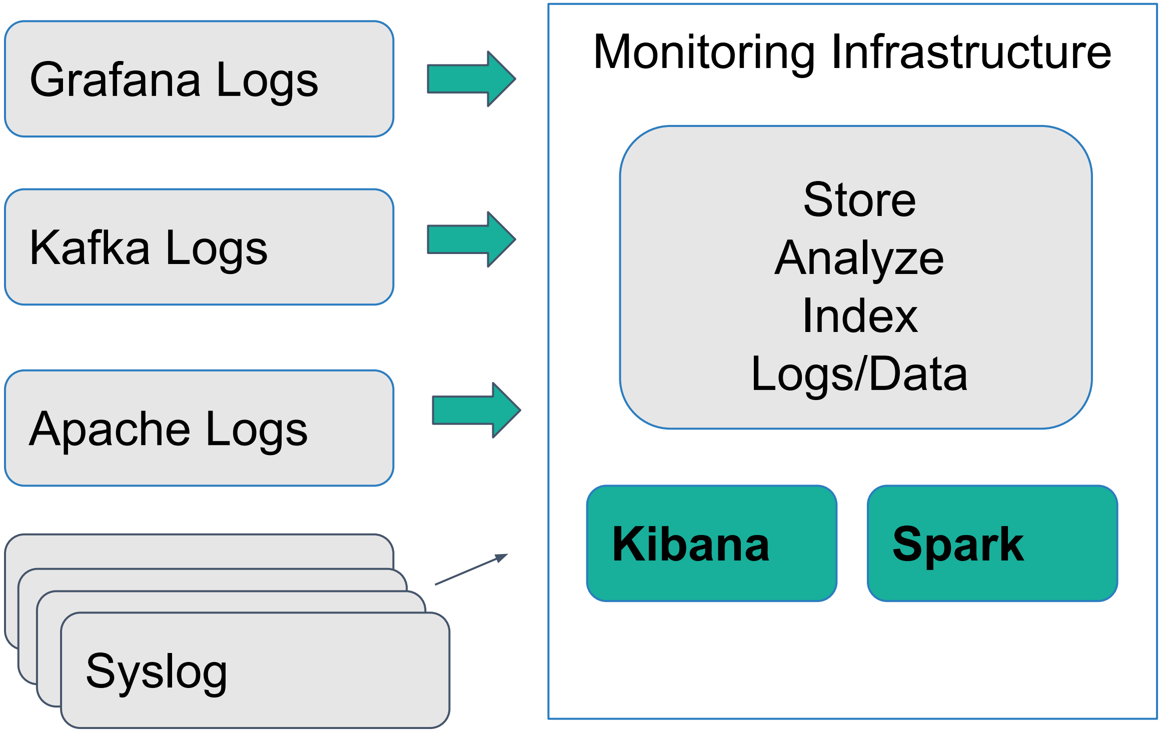
ANALYSIS AND MACHINE LEARNING ON LOGS OF THE MONITORING INFRASTRUCTURE

AUTHOR:
Mert Ozer

SUPERVISOR:
Borja Garrido Bear

GOAL OF THE PROJECT

IMPROVE UTILISATION OF MONITORING INFRASTRUCTURE



1. **Import** logs in Monitoring Infrastructure
2. **Visualize** with Kibana to understand and optimize
3. **Analyze** (ML) log data with Spark for anomalies detection

1. IMPORT LOGS

DATA INGESTION

Unstructured raw log

```
[03/Aug/2017:16:53:33
+0200] "GET
/api/search?limit=10&quer
y=&tag=wlcg HTTP/1.1"
200 211
"https://monit-grafana-dev.
cern.ch/dashboard/db/def
ault?orgId=1" "Mozilla/5.0
(X11; Linux x86_64)
AppleWebKit/537.36
(KHTML, like Gecko)
Chrome/60.0.3112.78
Safari/537.36
```

Structured JSON data

metadata.type:apache
data.request:/api/search ? limit = 10 & query
= & tag = wlcg
data.verb:GET
data.response:200
data.referrer:"https://monit-grafana-dev.cern
ch/dashboard/db/default?orgId=1"
data.agent:"Mozilla/5.0 (X11; Linux x86_64)
AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/60.0.3112.78 Safari/537.36"
data.bytes:211
data.httpversion: 1.1
metadata.timestamp: [1501772013000]

Logstash

RAW LOGS

MANY TYPES AND FORMATS

Grafana

```
t=2017-08-01T23:59:07+0200 lvl=info msg="Request Completed" logger=context
userId=43 orgId=3 method=GET path=/query status=502 remote_addr=", ::1"
time_ms=10034 size=0
```

```
t=2017-08-01T23:57:10+0200 lvl=info msg="Path Plugins" logger=settings
path=/var/lib/grafana/plugins
```

APACHE kafka

```
[2017-08-10 09:43:34,814] INFO Rolled new log segment for 'condor_raw_metric-8' in 10
ms. (kafka.log.Log)
```

```
[2017-08-01 23:58:43,060] INFO Deleting index
/var/spool/kafka/rucio_raw_tracer-2/00000000000258723415.index.deleted
(kafka.log.OffsetIndex)
```

Apache

```
[Tue Aug 01 03:23:01.180514 2017] [ssl:warn] [pid 2210] AH01909: RSA certificate
configured for monit-grafana-dev.cern.ch:443 does NOT include an ID which matches the
server name
```

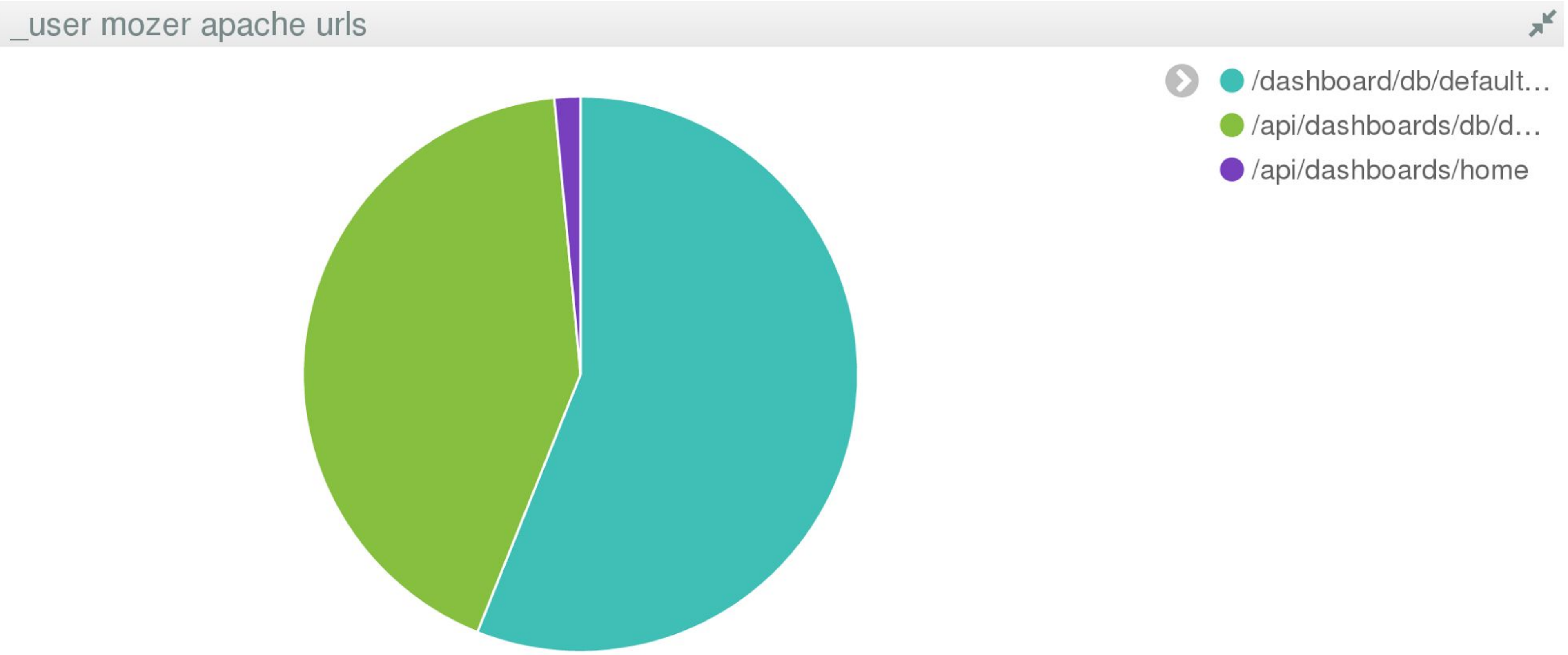
```
[03/Aug/2017:16:53:33 +0200] "GET /api/search?limit=10&query=&tag=wlcg HTTP/1.1"
200 211 "https://monit-grafana-dev.cern.ch/dashboard/db/default?orgId=1" "Mozilla/5.0
(X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/60.0.3112.78
Safari/537.36
```

2. ANALYSIS OBJECTIVES

- Most used dashboards
- Hourly visualization of grafana users
- Abnormal behaviour in services
- Anomaly detection in kafka clusters

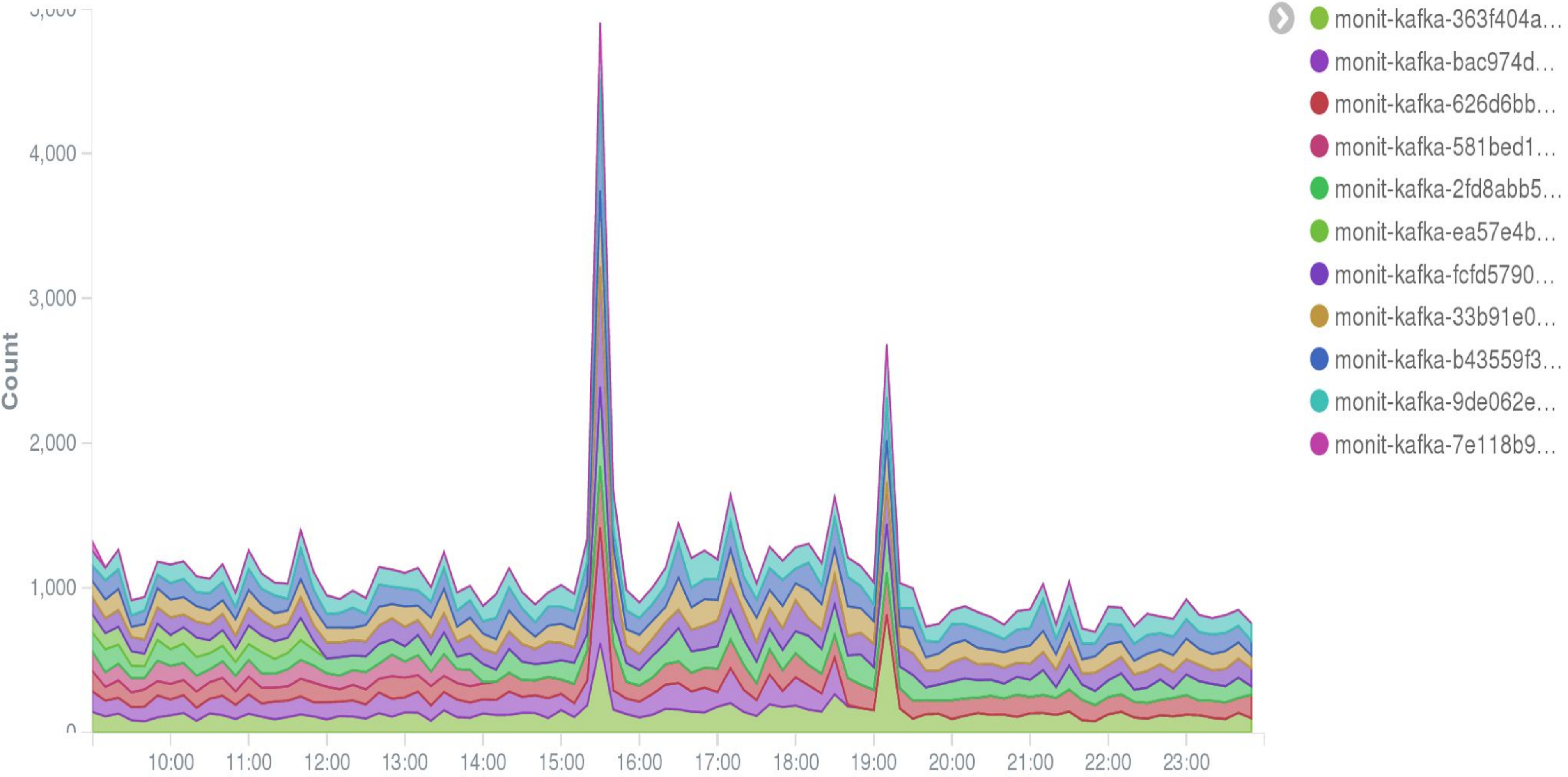
MOST USED DASHBOARDS

IMPROVE PERFORMANCE AND USER EXPERIENCE



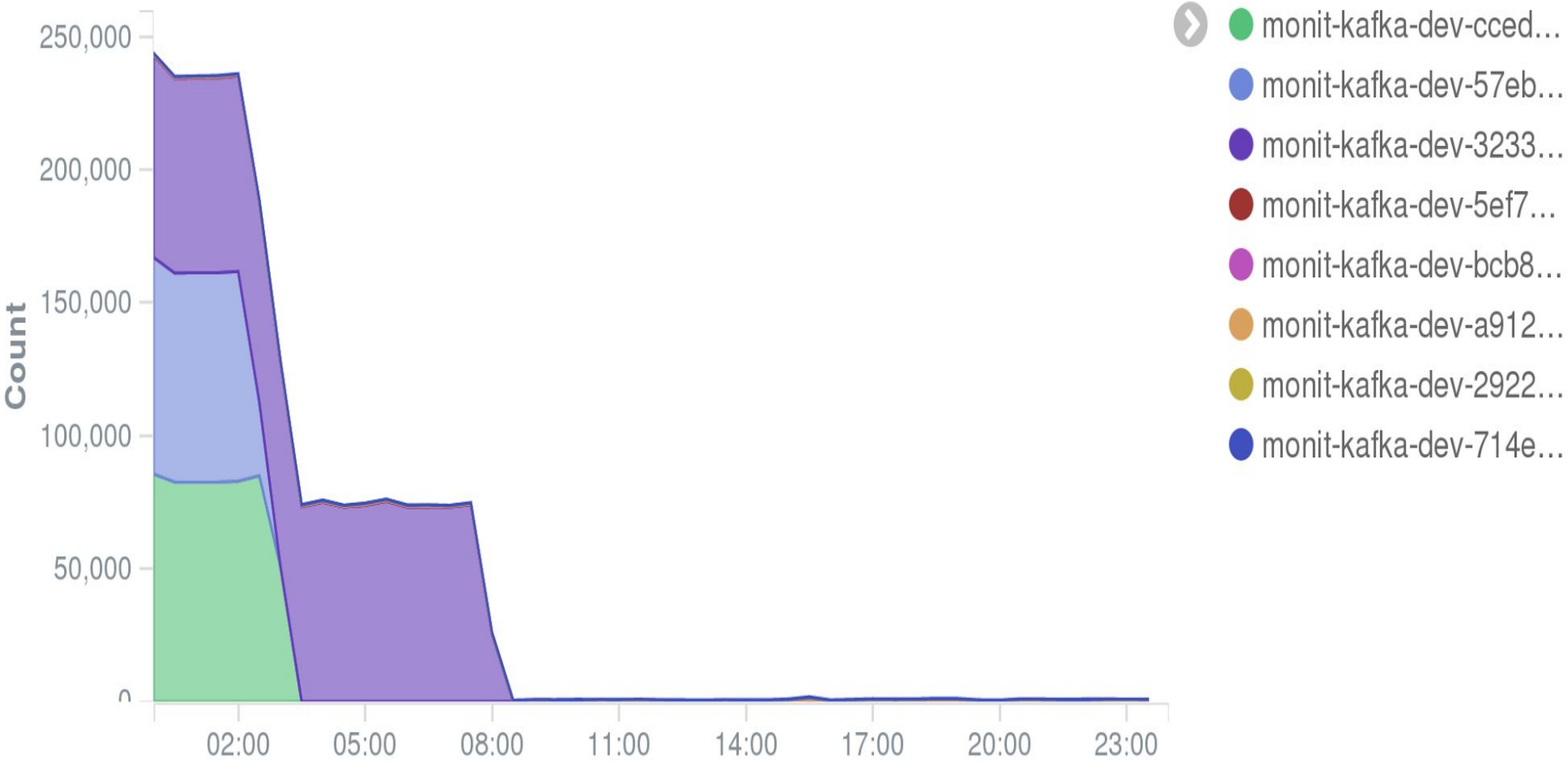
SPOT ISSUES WITH KAFKA

NORMAL BEHAVIOUR IN KAFKA CLUSTER



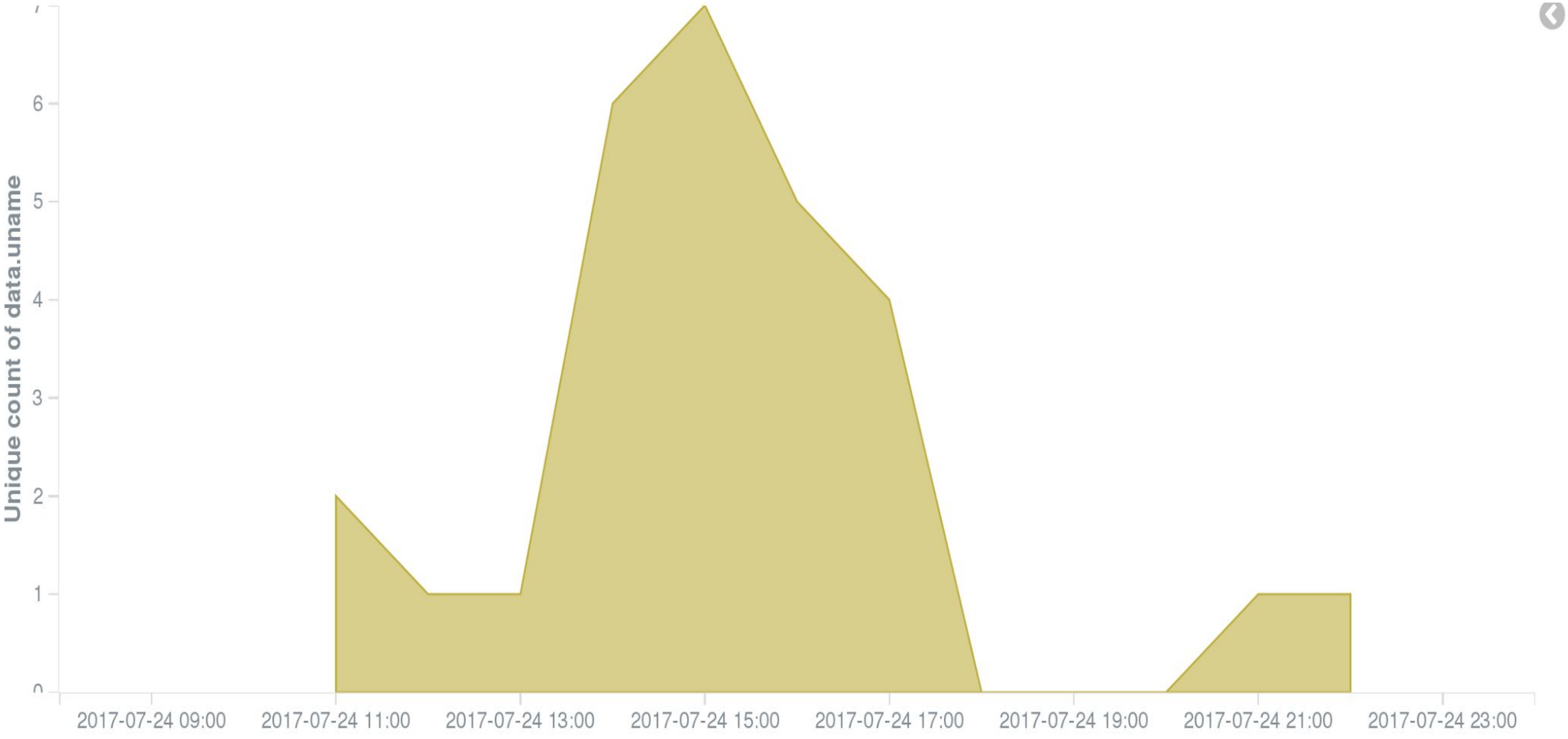
SPOT ISSUES WITH KAFKA

ABNORMAL BEHAVIOUR IN KAFKA CLUSTER



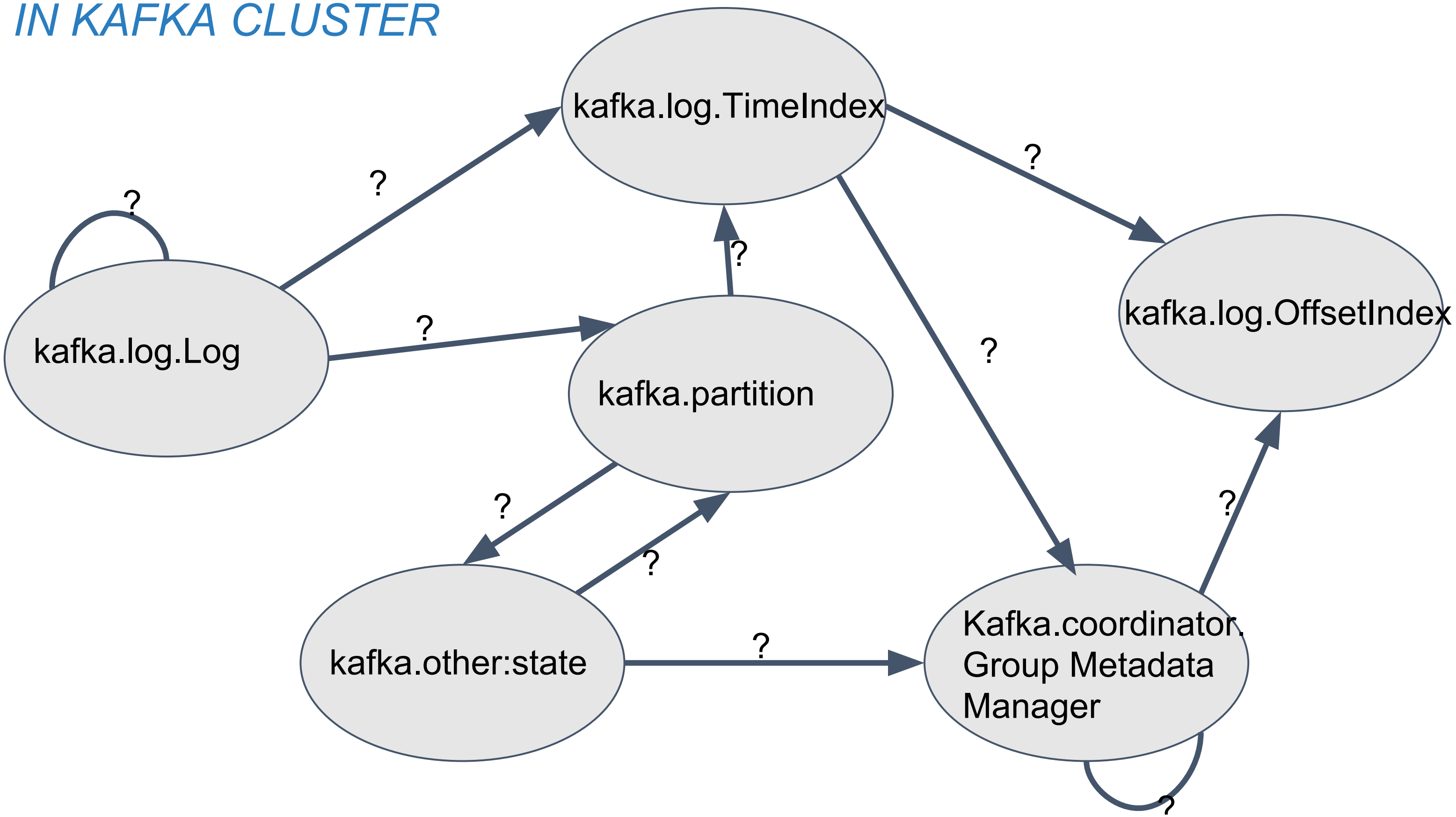
HOURLY GRAFANA USAGE

HOURLY VISUALIZATION OF UNIQUE GRAFANA USERS



3. ANOMALY DETECTION

IN KAFKA CLUSTER



Nodes:

Indicates current log state name

Edges:

Probability of transition to another state.

ONGOING WORK

- Finding anomalies in Kafka clusters
- More visualisations
- Machine learning to complement visualisations

SUMMARY

1. **Import** logs in Monitoring Infrastructure
2. **Visualize** with Kibana to understand and optimize monitored services
3. **Analyze** (ML) log data with Spark for anomalies detection

Big Thanks to:

CERN Openlab for this amazing experience
My supervisor **Borja Garrido Bear**
And to my team members.

mertozer94@gmail.com