# Rucio & Kubernetes
## Lessons Learned from the ESCAPE Full Dress Rehearsal

Riccardo Di Maria

CERN

November 26th, 2020 - Rucio Development Meeting, CERN

# Science Projects

# ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

**EUROPEAN OPEN SCIENCE CLOUD**
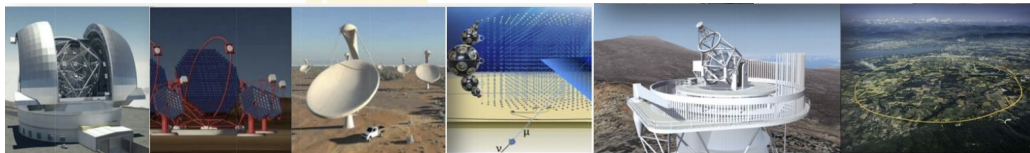
Horizon2020
European Union Funding
for Research & Innovation

# Data Centres

# Project Goals

- Prototype an infrastructure adapted to exabyte-scale needs of large science projects.
- Ensure sciences **drive** the development of EOSC.
- Address FAIR data management principles.

November 26th, 2020

Riccardo.Di.Maria@cern.ch

# The ESCAPE Project Work Packages

- **Management, Innovation, Networking and Dissemination (MIND)**: coordination and management.

- **Data Infrastructure for Open Science (DIOS)**: a scalable federated data infrastructure (DataLake) as the basis of an open science for the ESFRI projects within ESCAPE.

- **Open-source scientific Software and Service Repository (OSSR)**: the repository of scientific software services of the research infrastructures concerned by the ESCAPE project.

- **Virtual Observatory - connecting ESFRI projects to EOSC through VO framework (VO)**: astronomical high-level products archive and related services.

  @rucio @swan

- **ESFRI Science Analysis Platform (ESAP)**: a flexible science platform for the analysis of open access data.

- **Citizen Science - engagement and communication (CS)**: an open gateway dedicated to the public through Citizen Science and communication actions.

# WP2: Data Infrastructure for Open Science (DIOS)

- Data Lake Infrastructure and Federation Services - <u>Riccardo Di Maria (CERN)</u>

- Data Lake Orchestration Service - Patrick Fuhrmann/Paul Millar (DESY)

- Integration with Compute Services - Yan Grange (ASTRON-NWO-I)

- Networking - Rosie Bolton (SKAO)

- Authentication and Authorization - Andrea Ceccanti (INFN)

<u>Xavier Espinal (CERN)</u> as WP leader and Rosie Bolton (SKAO) as deputy

Riccardo.Di.Maria@cern.ch

# ESCAPE DataLake Components

**From CERN Team**

Rucio

@martin @mario @thomas

XCache

@david    @andy

**From CERN-IT**

FTS

@rizart

perfSONAR

@rizart

Grafana

@rizart

OpenStack

CRIC

@panos

OracleDB

k8s

From ESCAPE Partners

IAM

@ceccanti

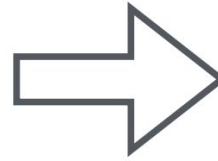# AuthN/Z in the ESCAPE DataLake

**Current, X.509 based AAI**
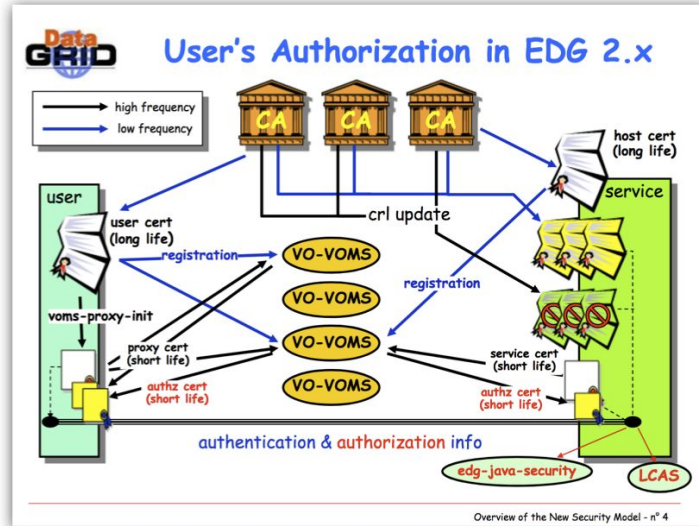
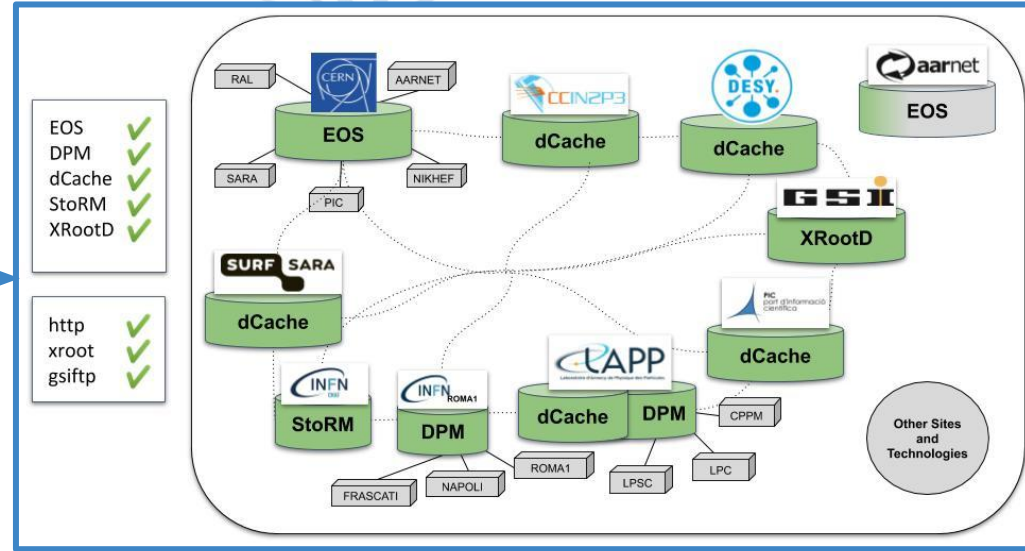**Future, token-based AAI**

**Move beyond X.509**
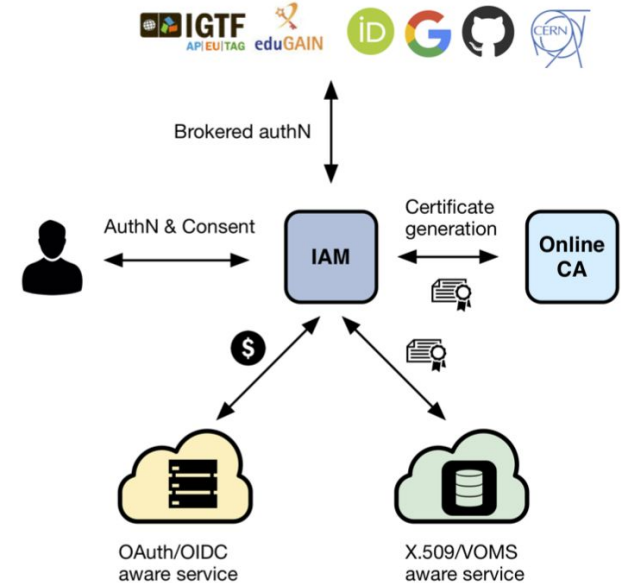
# Approach: leverage and build upon the WLCG experience

# ESCAPE DataLake
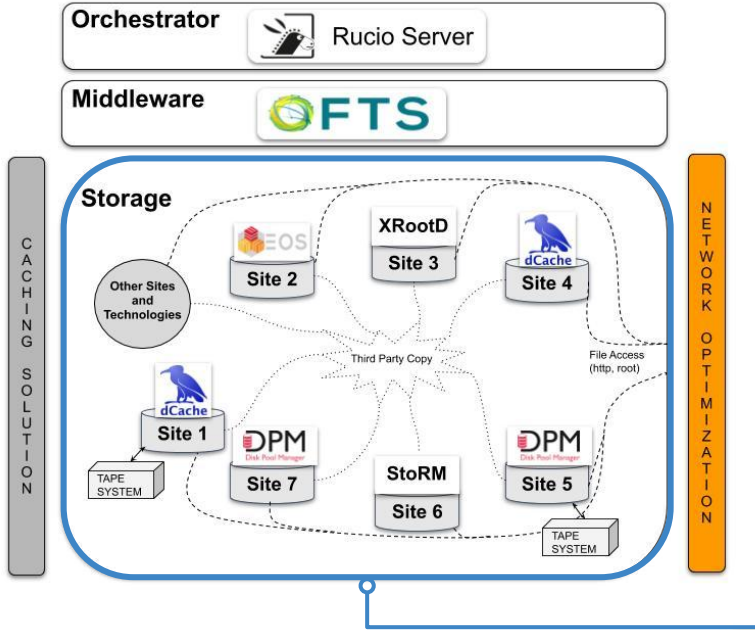


- Hiding complexity and providing transparent access to data.
- Heterogeneous federated storage and operations model.
- Some centers joining even if not funded by ESCAPE.

Further info: https://wiki.escape2020.de/index.php/WP2_-_DIOS#Datalake_Status

November 26th, 2020

Riccardo.Di.Maria@cern.ch

# ESCAPE DataLake

| RSE | Quota | WM |
|-----|-------|-----|
| ALPAMED-DPM | 100 TB | 10 TB |
| CNAF-STORM | 10 TB | 1 TB |
| DESY-DCACHE | 40 TB | 4 TB |
| EULAKE-1 | 300 TB | 30 TB |
| GSI-ROOT | 1 TB | 10 GB |
| IN2P3-CC-DCACHE | 60 TB | 1 TB |
| INFN-NA-DPM | 68 TB | 5 TB |
| INFN-NA-DPM-FED | 46 TB | 5 TB |
| INFN-ROMA1 | 2 TB | 200 GB |
| LAPP-DCACHE | 10 TB | 1 TB |
| LAPP-WEBDAV | 100 GB | 90 GB |
| PIC-DCACHE | 28 TB | 27.99 TB |
| PIC-INJECT | 28 TB | 27.99 TB |
| SARA-DCACHE | 98 TB | 140 GB |

**ESCAPE DataLake**

- Total Quota:

  **891 TB**

- Watermark:

  **113.44 TB**

- 10+ RSEs
- 9 sciences
- 50+ accounts

November 26th, 2020          Riccardo.Di.Maria@cern.ch

# DataLake **24-hour Full Dress Rehearsal** Objectives

**Goal:** exercise covering **experiment data workflow** needs on a single day (data injection, replication, and access).

**Three fold goal:** perspective from **scientists**, perspective from **sites**, and the assessment of the **ESCAPE DataLake tools and services** under **pseudo-production conditions**: RUCIO, FTS, CRIC, IAM, perfSONAR, monitoring, QoS, clients, etc.

## DataLake Objectives

- Stable infrastructure:
  10 sites, 5 storages technologies, 3 protocols ✓
- Monitoring: automated tests ✓
- 1M files - to demonstrate stable and sizeable data movement ✓
- 3 QoS: CRIC as reference point ✓

## Compute Integration Objectives

- Interactive access to files in DataLake
  e.g. using JupyterLab Notebook ✓
- Batch access to files in DataLake ✓
- Data access through caching layer ⊖
- Process data locally downloaded ✓
- Use cases covering small-to-large files range; simple from compute perspective (focus on mimicking data flow) ✓

## QoS Objectives

- Demonstrate compute-driven QoS staging ⊖
- Demonstrate cost-performance QoS trade-off ✓
- Demonstrate VO-specific workflows (data lifecycles) ✓
- Demonstrate data injection with targeted QoS ✓
- Demonstrate computational match-making ⊖
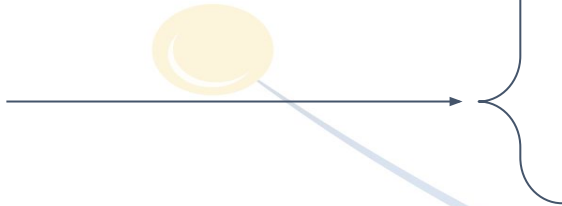
## Network and Monitoring Objectives

- Test suite development ✓
- Rucio events into a dashboard ✓
- Dashboard adjustments ✓
- Develop and run rucio-level tests ✓
- Develop example data lifecycle definitions and scripts ✓

## AAI Testbed Objectives

- User enrollment flow in place ✓
- X.509/VOMS AuthN/Z in place ✓
- Token-based AuthN/AuthZ in place ⊖
- Continuous monitoring tools to assess that AuthN/Z work as expected at sites ✓

Riccardo.Di.Maria@cern.ch

ESCAPE
European Science Cluster of Astronomy &
Particle physics ESFRI research infrastructures

- OpenStack VMs:

  ○ 1 master: 4 CPU, 8 GB RAM

  ○ 10 nodes: 8 CPU, 16 GB RAM

- K8s cluster:

  ○ filebeat (per node) and logstash for cluster monitoring

  ○ rucio-client with root account and admin privileges for DataLake managing

  ○ escape-crons pod

- IAM-Rucio sync

- IAM-Gridmap (EOS) sync

- CRIC-Rucio sync

- noise production (100MB file upload per RSE + add rule per RSE)

- Gfal SAM and FTS tests

Riccardo.Di.Maria@cern.ch

# Kubernetes Cluster @ CERN

- Rucio (HELM-charts-based):

  - UI (escape-rucio.cern.ch)

  - Auth Server

  - Main Server (2)

  - Daemons:

    - **Abacus Account** [updating account (counter) usages]

    - **Abacus Collection Replica** [updating collection replica]

    - **Abacus RSE** [updating RSE (counter) usages]

    - **Conveyor Submitter** (3 x 4 threads) [managing non-tape file transfers - preparing and submitting jobs]

    - **Conveyor Poller** (3 x 4 threads) [checking status of submitted transfers]

- **Conveyor Finisher** (2 threads) [updating Rucio internal state for finished transfers]

- **Hermes** [delivering messages via STOMP to a message broker]

- **Judge Injector** (2) [asynchronously injecting replication rules]

- **Judge Evaluator** (3 x 3 threads) [executing and reevaluating replication rules]

- **Judge Repairer** (2 x 5 threads) [repairing stuck replication rules]

- **Judge Cleaner** (2 x 5 threads) [cleaning expired replication rules]

- **Reaper2** (2 x 4 threads) [deleting replicas]

- **Transmogrifier** [creating replication rules for DIDs matching a subscription]

- **Undertaker** [managing (deleting) expired DIDs]

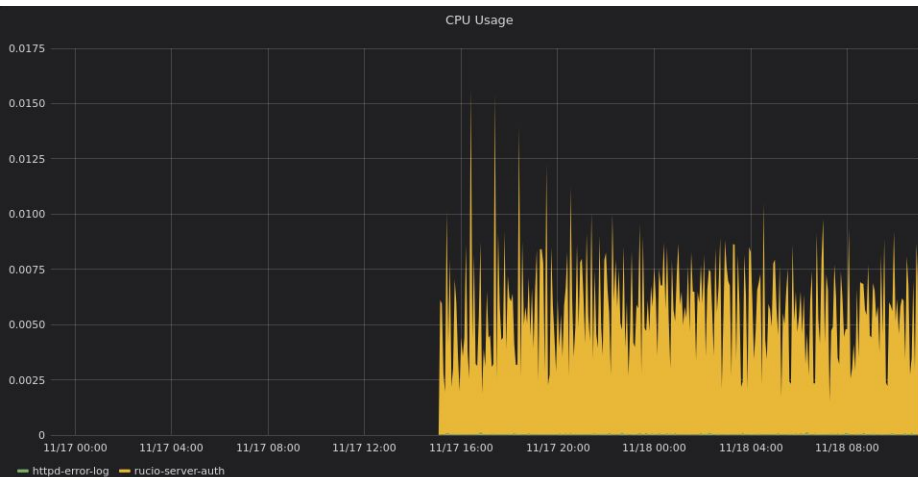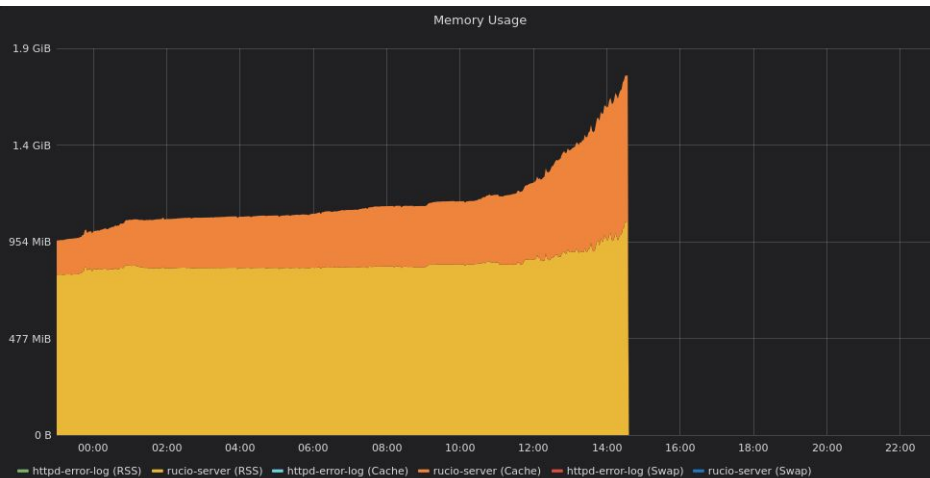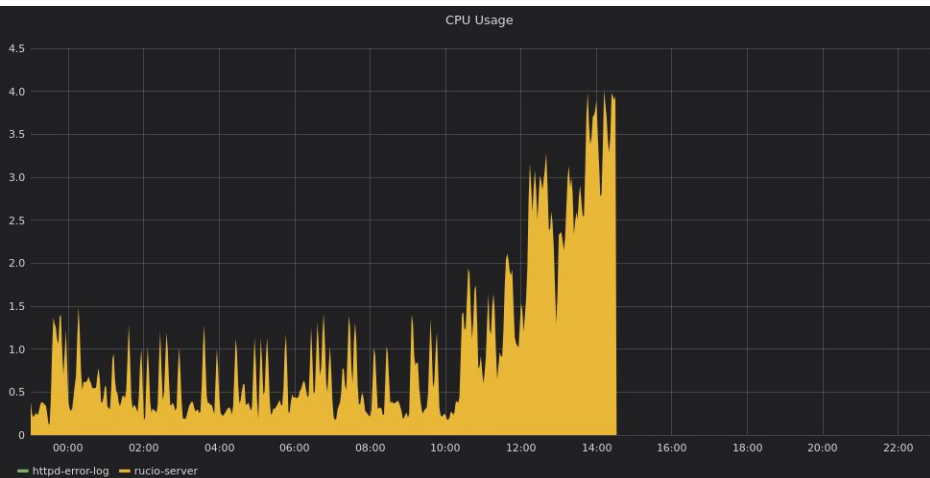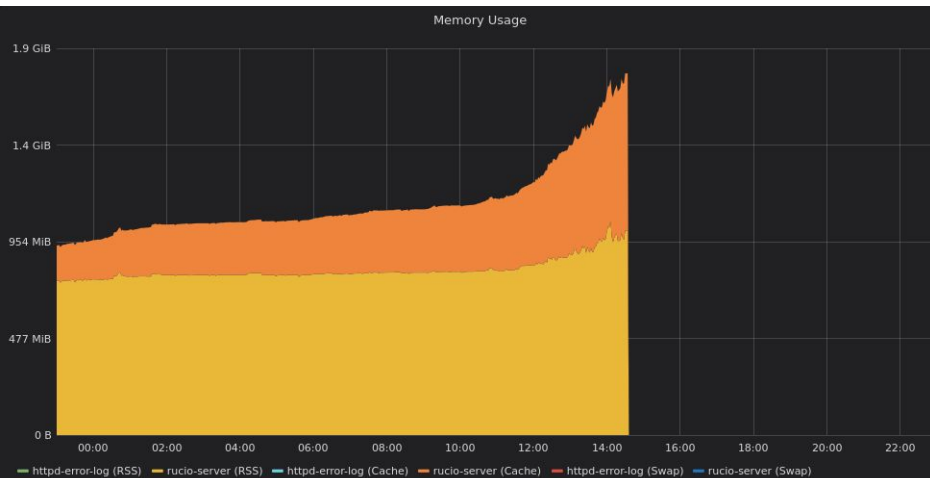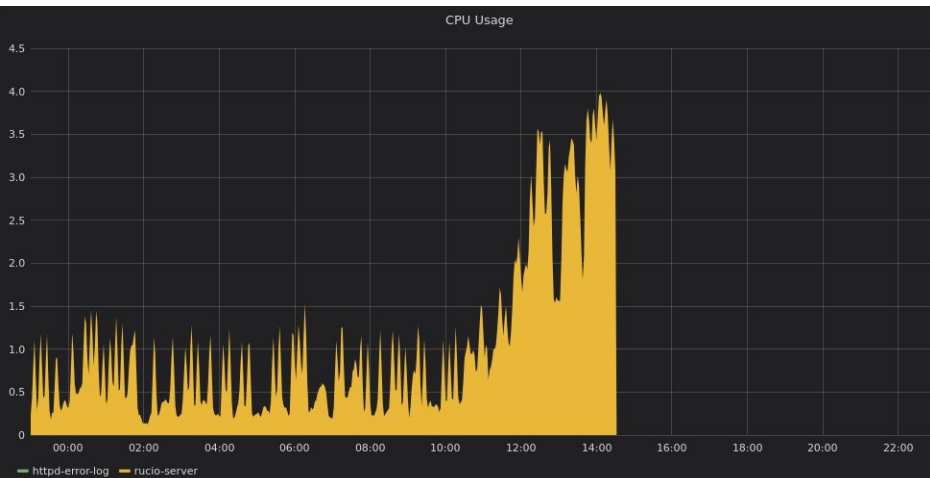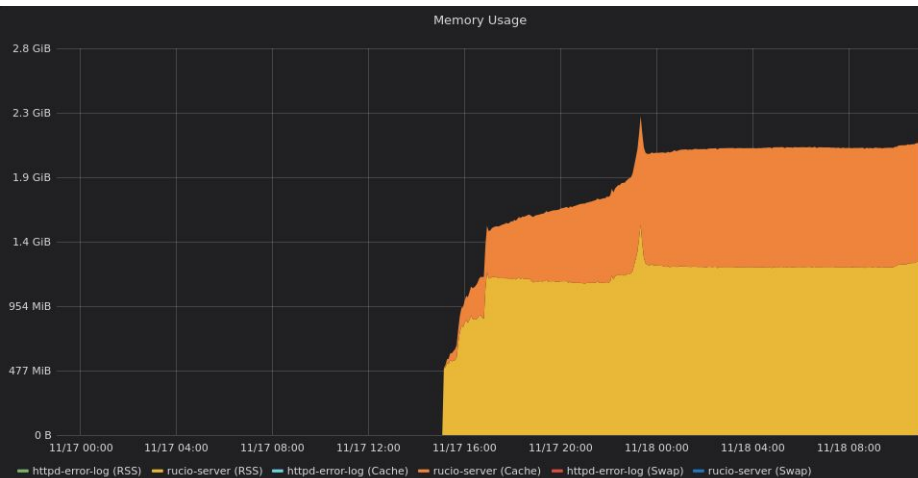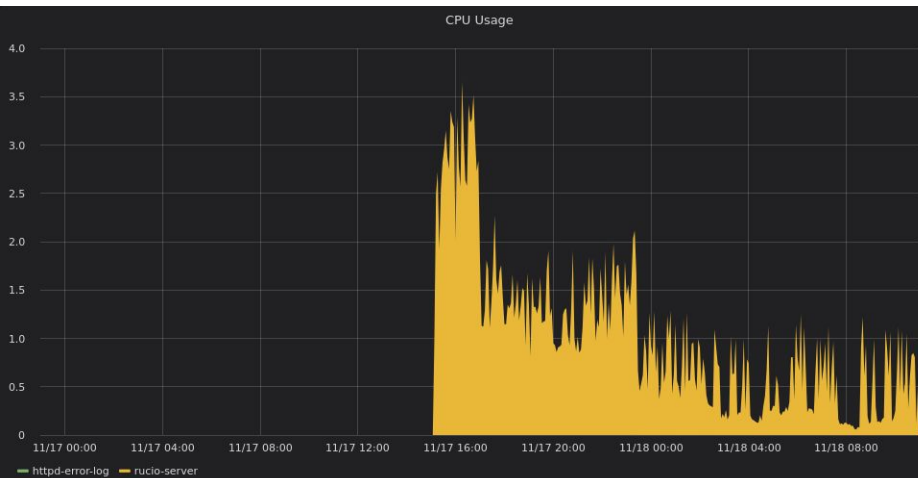| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| ui | - | - | 0.004 (0.06) \| 550 MiB | OK |
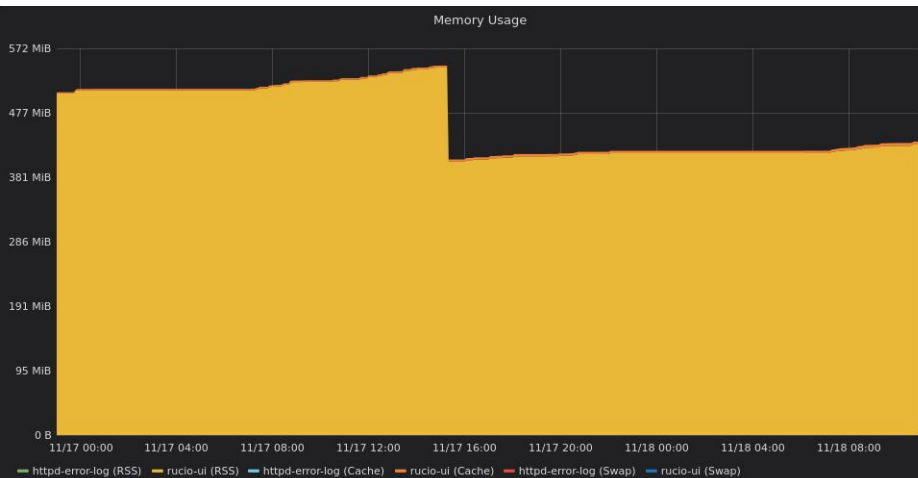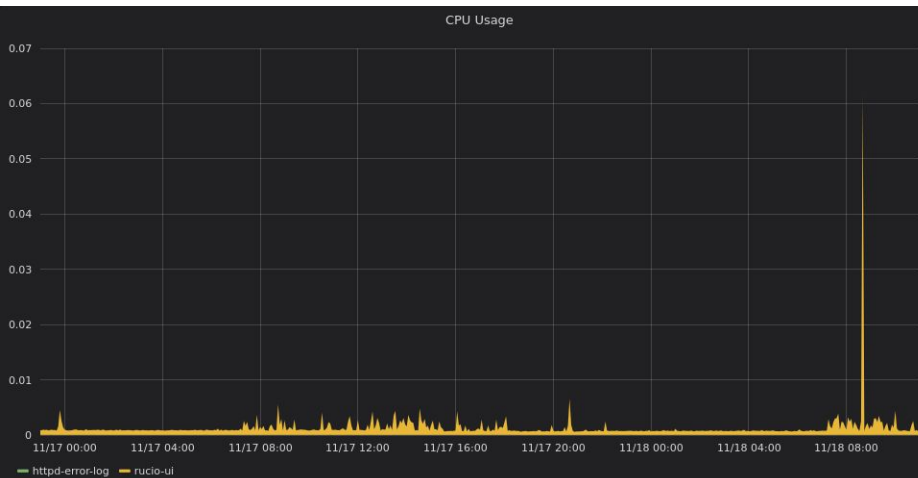
# Auth - before setting resources requests/limits

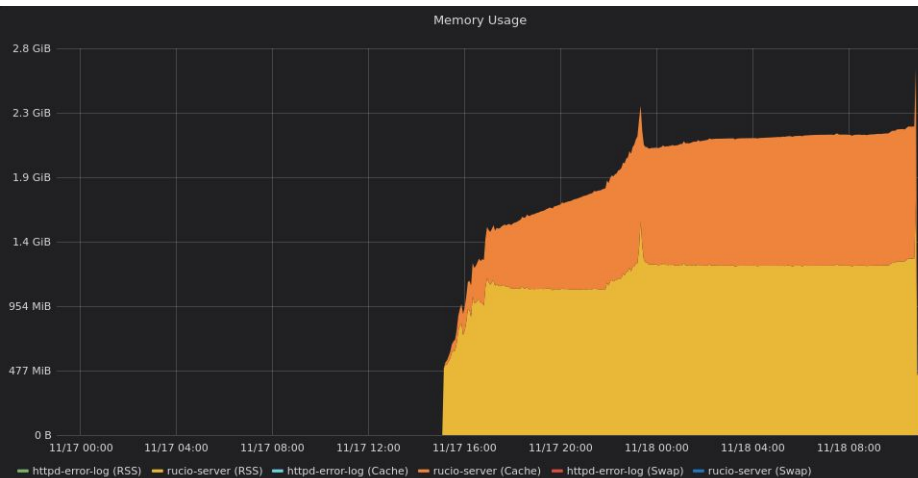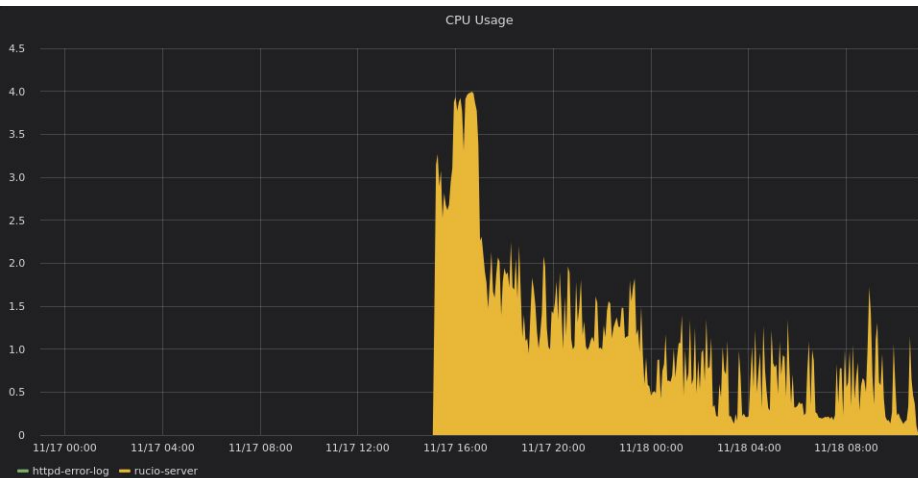| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| server-auth | - | - | 0.02 (0.12) \| 550 MiB (1.25 GiB) | Errors and restart due to no limits set. |
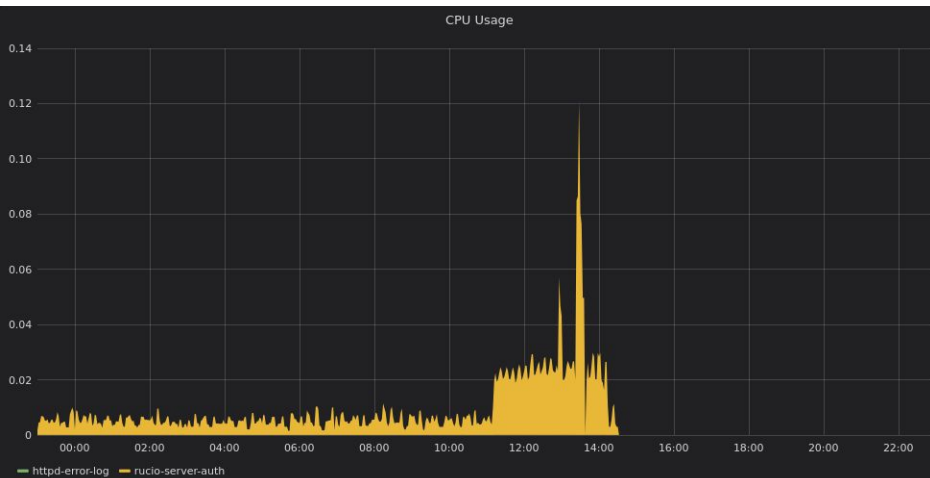
# Auth - after setting resources requests/limits

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| server-auth | 4 \| 2500 MiB | 4 \| 2500 MiB | 0.02 \| 500 MiB | OK |
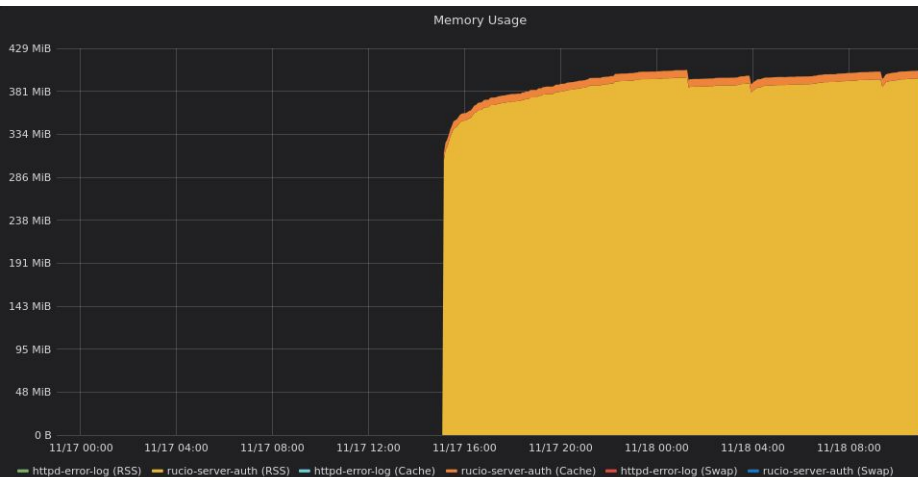
November 26th, 2020          Riccardo.Di.Maria@cern.ch

Funded by the European Union's
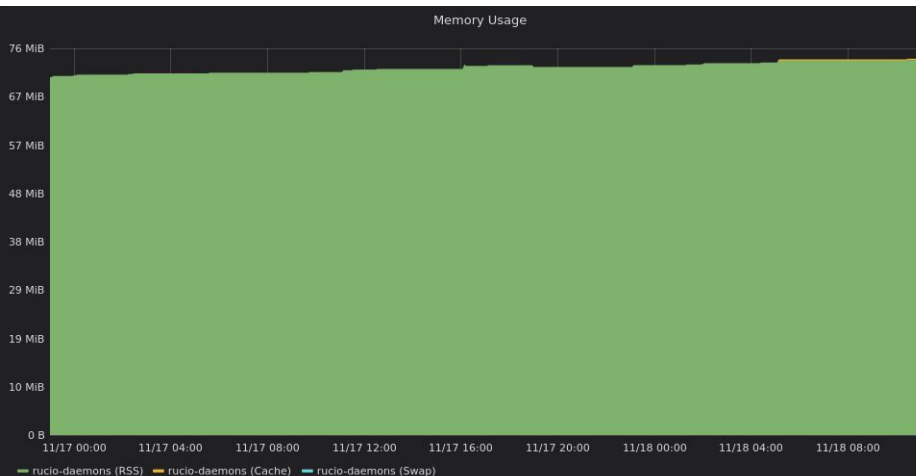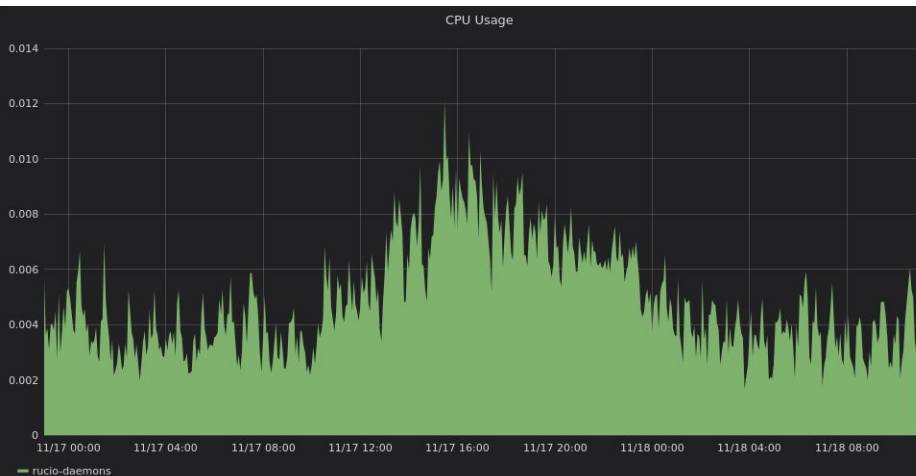Horizon 2020 - Grant N° 824064

# Server - before setting resources requests/limits
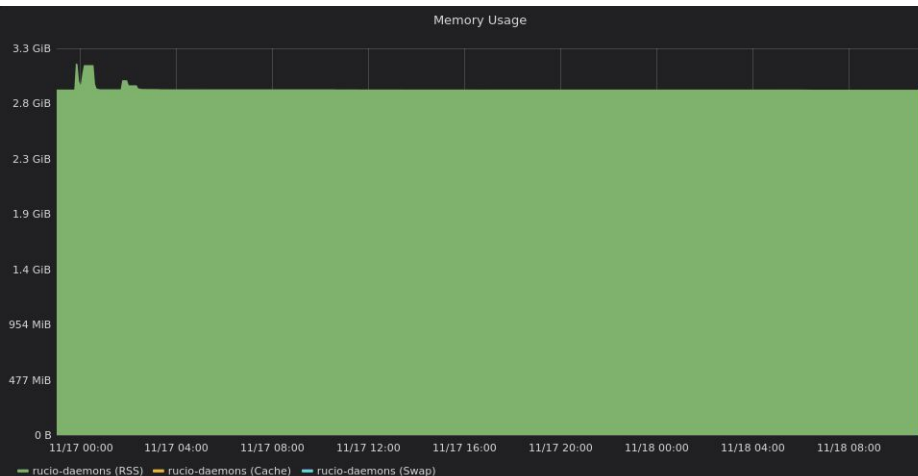
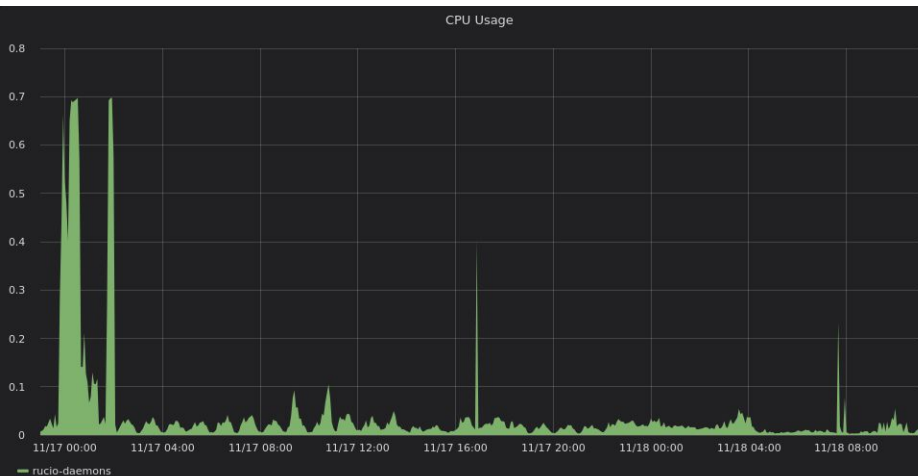| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---------|-----------------------------------|----------------------------------|------------------------------------------------------|-------------------|
| server (2) | - | - | 2-4 \| 1.25 GiB (2 GiB) | Manual restart due to no limits set. |

# Server - before setting resources requests/limits

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| server (2) | - | - | 2-4 \| 1.25 GiB (2 GiB) | Manual restart due to no limits set. |

November 26th, 2020

Riccardo.Di.Maria@cern.ch

# Server - after setting resources requests/limits

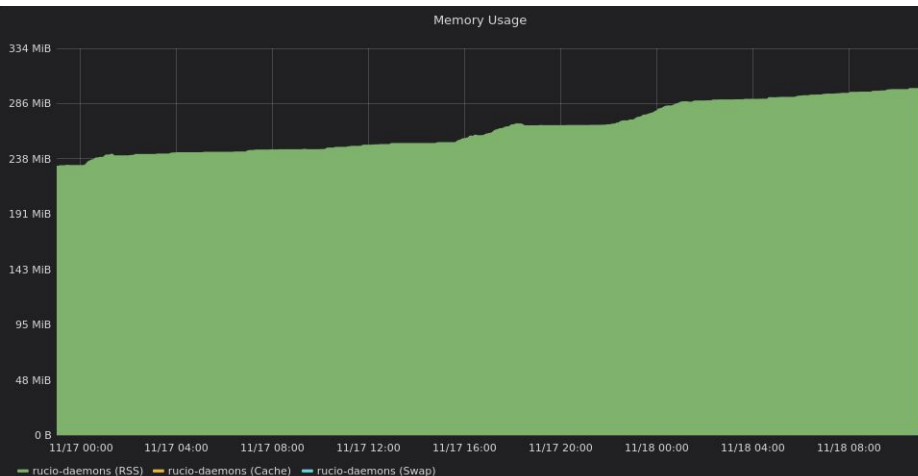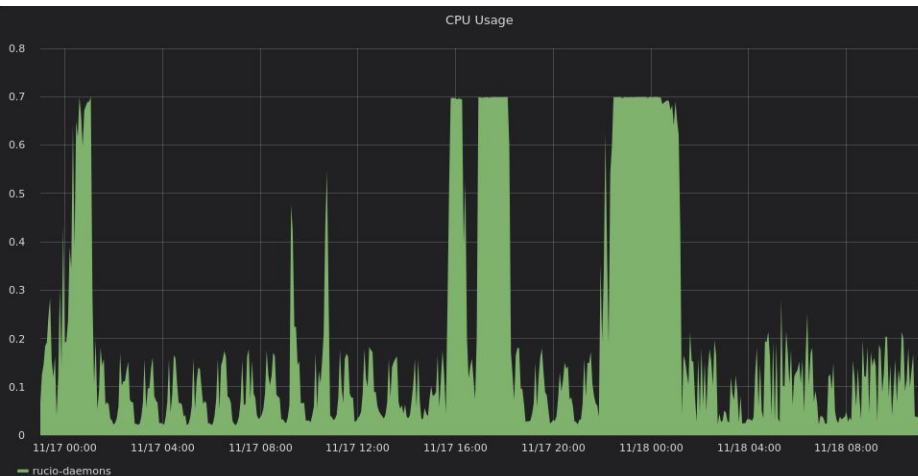| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| server (2) | 4 \| 2500 MiB | 4 \| 2500 MiB | 2-4 \| 1-2 GiB (>2.5 GiB cache) | 2 restarts due to memory limit. |

# Server - after setting resources requests/limits
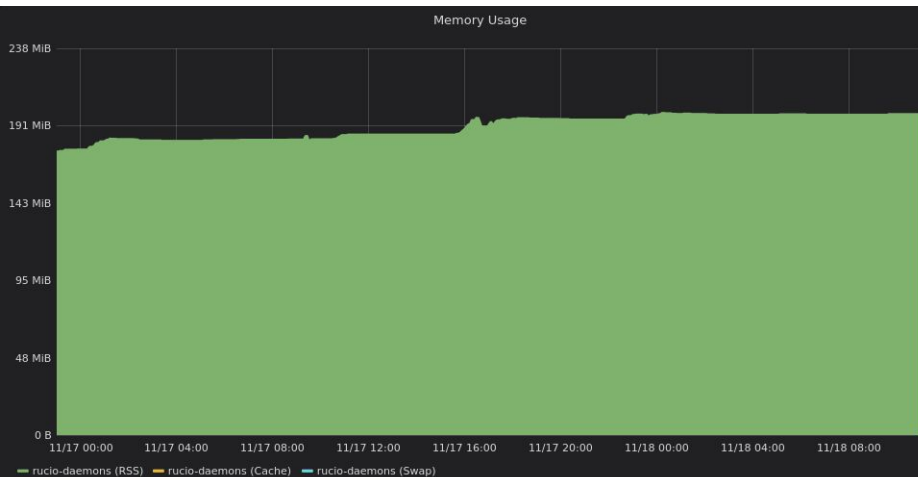
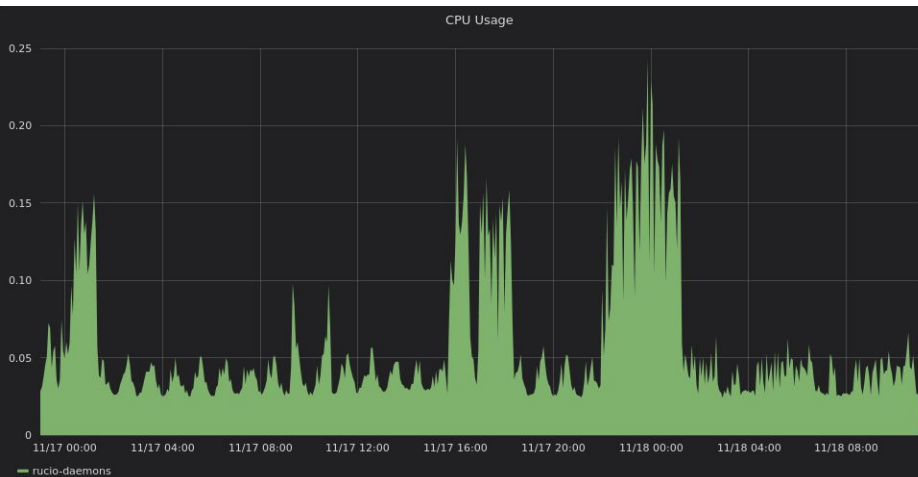| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---------|-----------------------------------|----------------------------------|------------------------------------------------------|-------------------|
| server (2) | 4 \| 2500 MiB | 4 \| 2500 MiB | 2-4 \| 1-2 GiB (>2.5 GiB cache) | 2 restarts due to memory limit. |

# Daemon: Abacus Account

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| abacus-account | 0.70 \| 200 MiB | 0.70 \| 200 MiB | 0.01 \| 80 MiB | OK |

# Daemon: Abacus Collection Replica
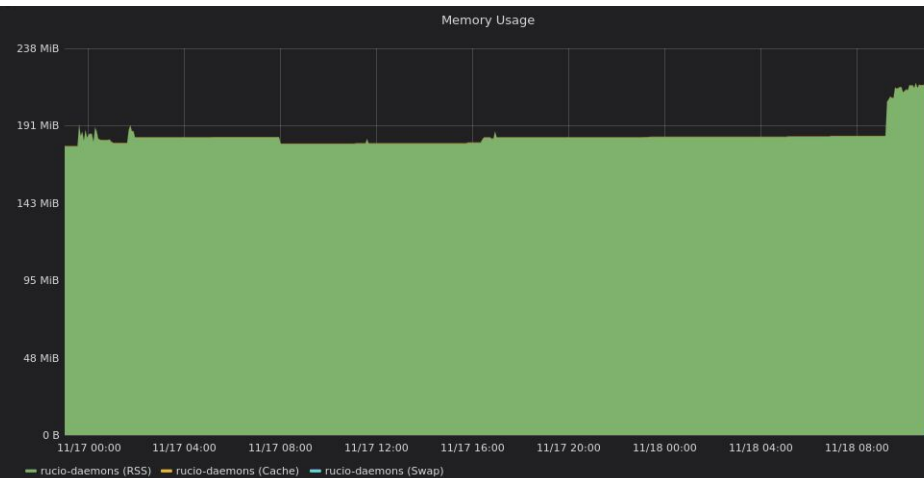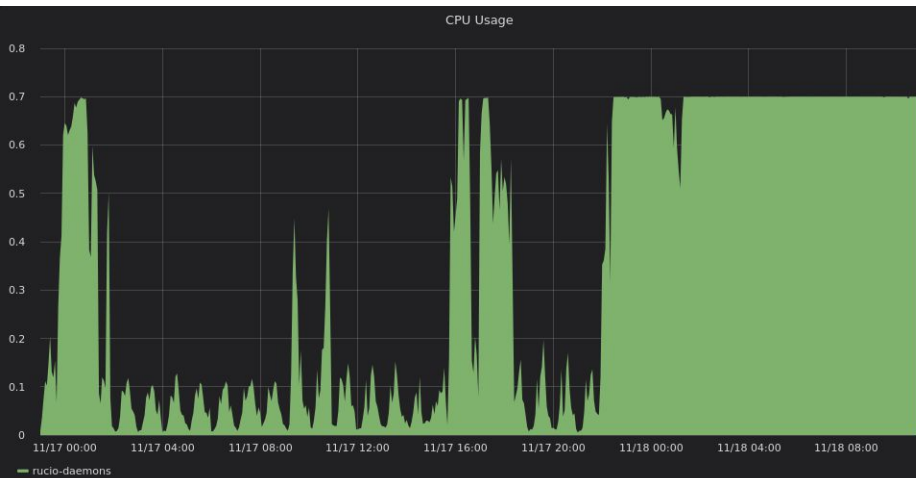
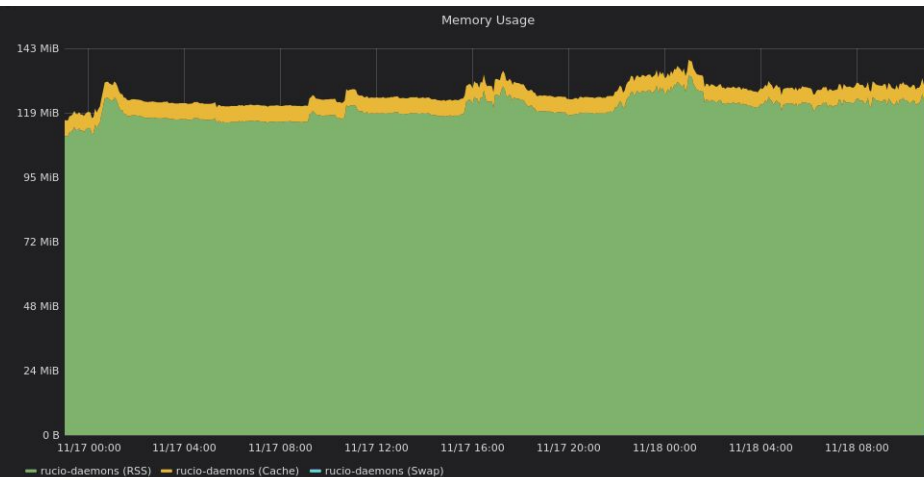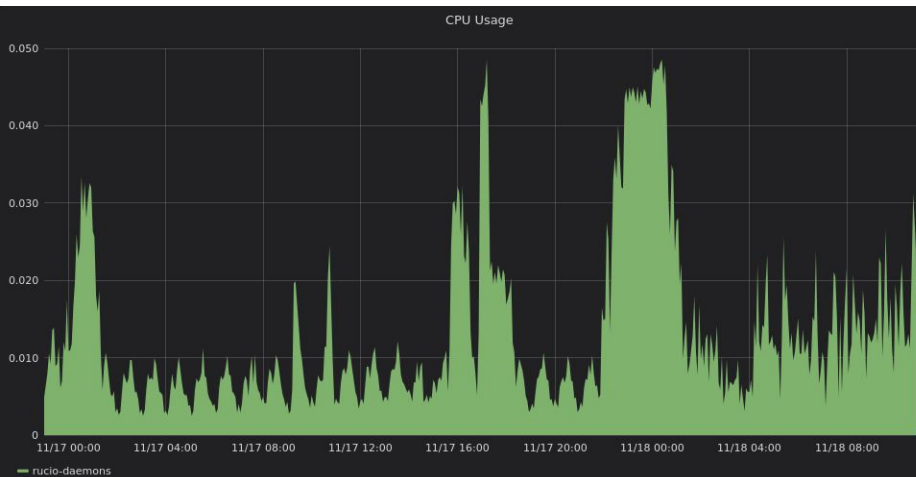| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| abacus-collection-replica | 0.70 \| 7000 MiB | 0.70 \| 7000 MiB | 0.01-0.2 (0.7) \| 3 GiB | OK<br>1 restart due to OOMKilled but not shown in plot. Not understood. |

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---------|-----------------------------------|----------------------------------|-----------------------------------------------------|-------------------|
| abacus-rse | 0.70 \| 200 MiB | 0.70 \| 200 MiB | 0.01 \| 80 MiB | OK |

November 26th, 2020

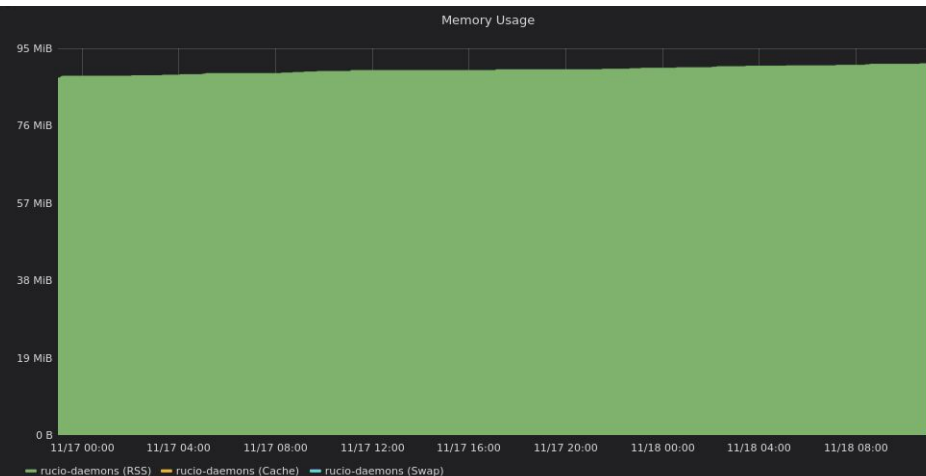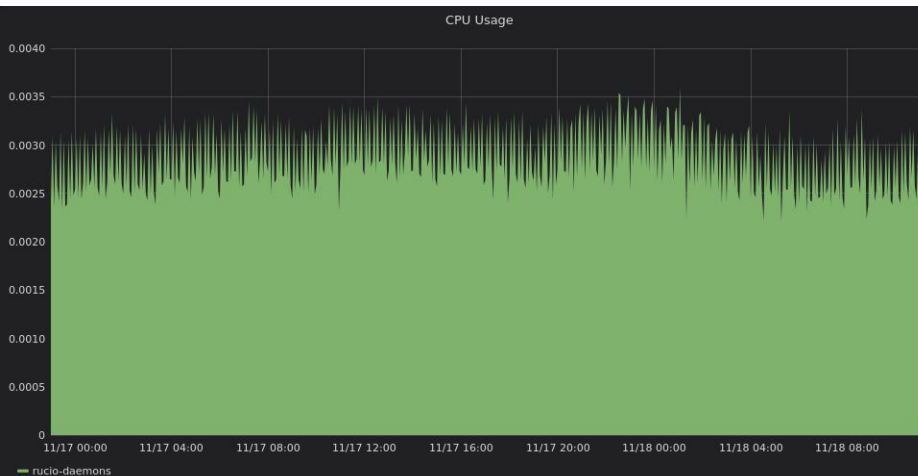Riccardo.Di.Maria@cern.ch
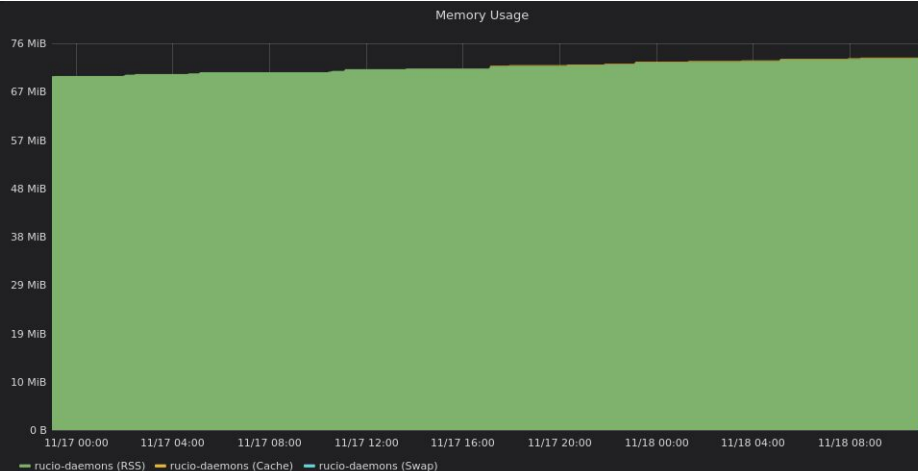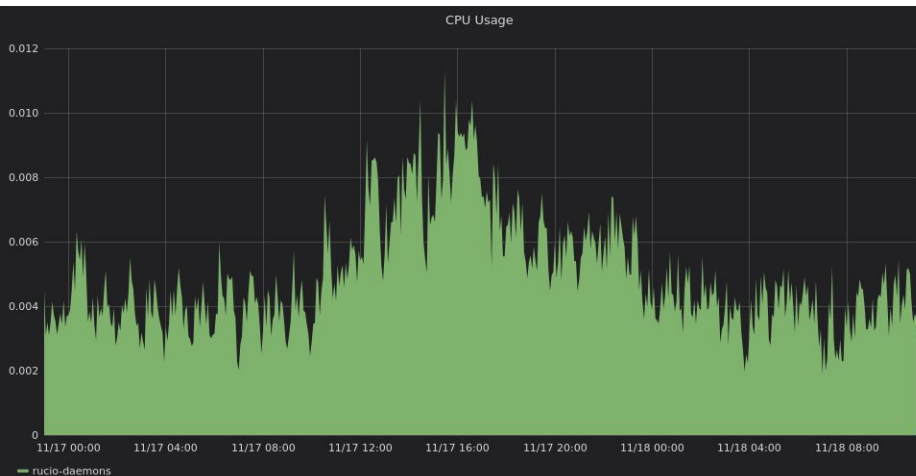
# Daemon: Conveyor Submitter

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| conveyor-submitter (3 x 4 threads) | 0.70 \| 800 MiB | 0.70 \| 800 MiB | 0.2-0.7 \| 300 MiB | OK |

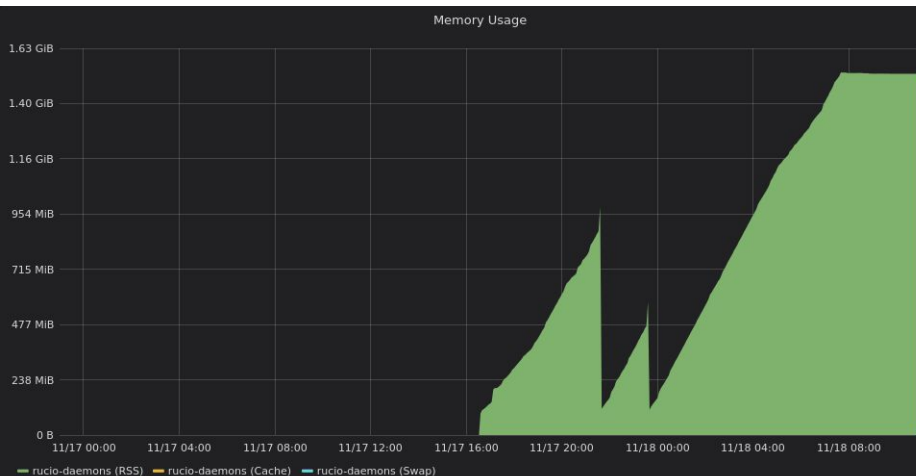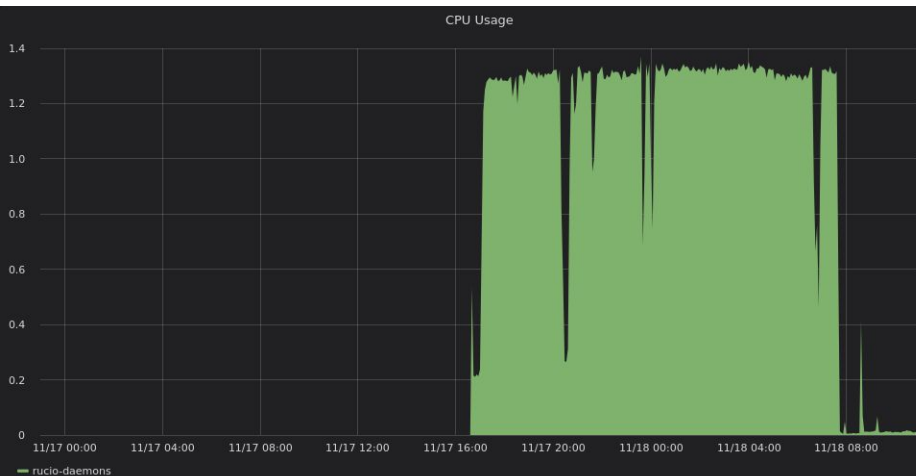November 26th, 2020          Riccardo.Di.Maria@cern.ch

# Daemon: Conveyor Poller

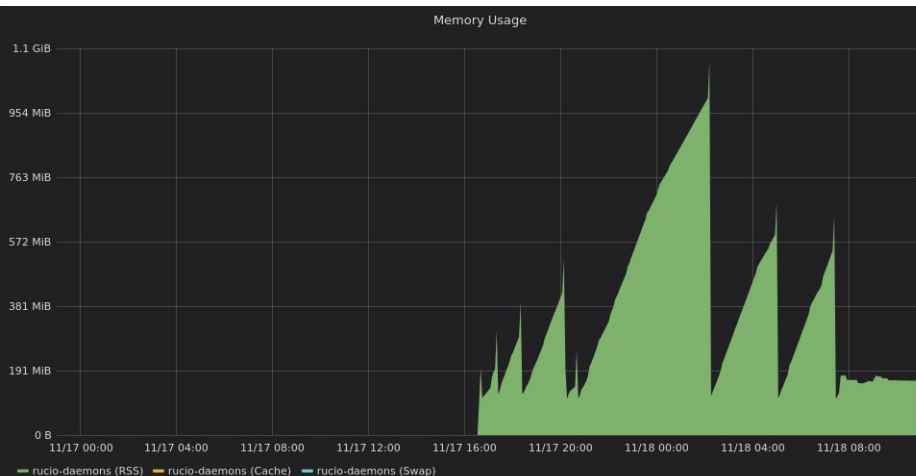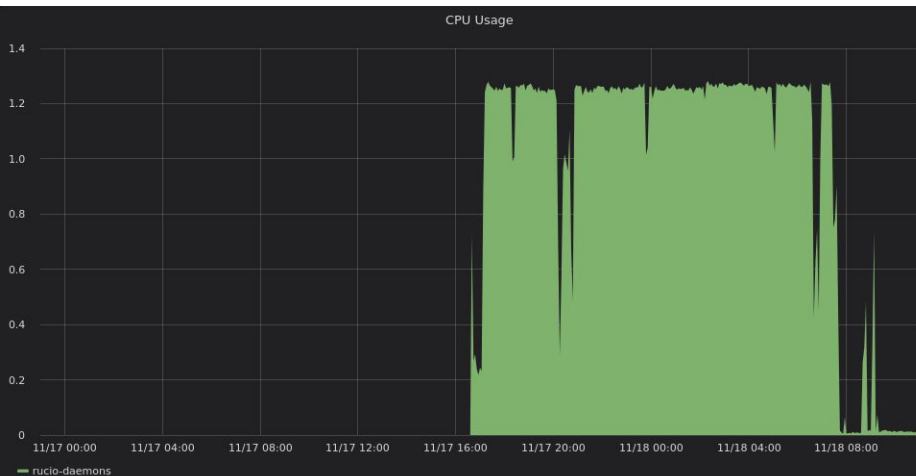| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| conveyor-poller (3 x 4 threads) | 0.70 \| 800 MiB | 0.70 \| 800 MiB | 0.05-0.25 (0.4) \| 200 MiB | OK |

# Daemon: Conveyor Finisher

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---------|------------------------------------|----------------------------------|------------------------------------------------------|-------------------|
| conveyor-finisher (2 threads) | 0.70 \| 400 MiB | 0.70 \| 400 MiB | 0.1-0.7 (0.7 maintained) \| 200 MiB | OK Probably worth to increase CPU. |

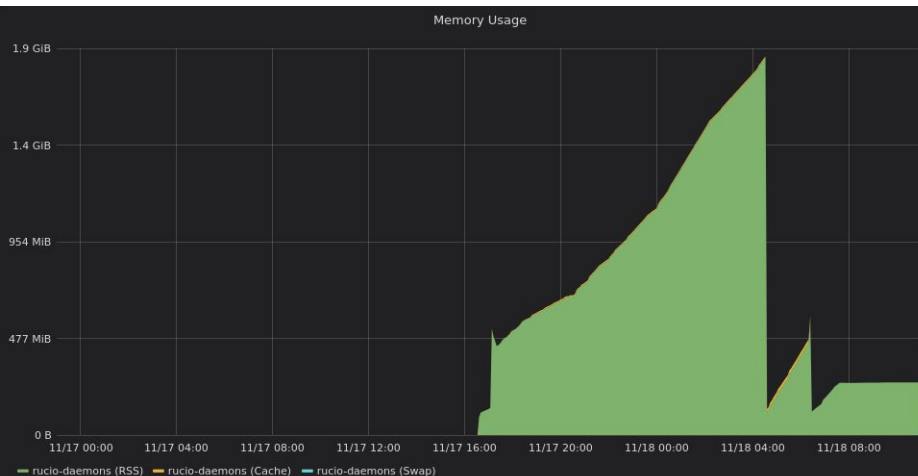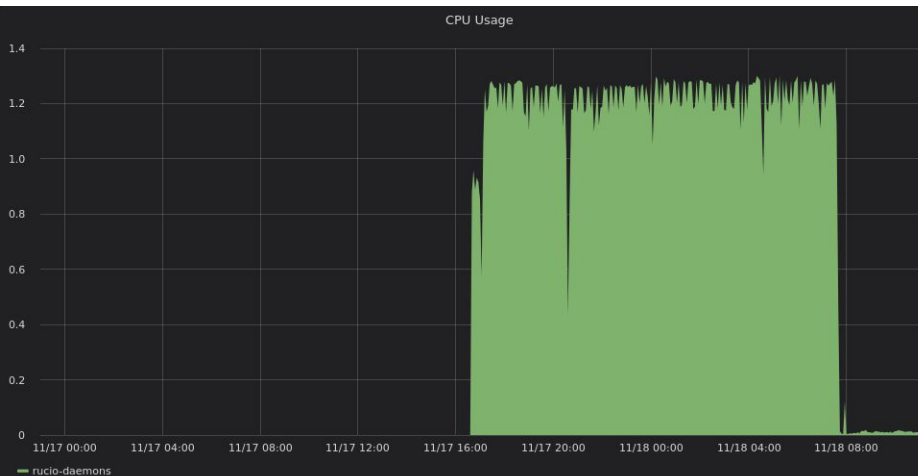November 26th, 2020          Riccardo.Di.Maria@cern.ch

# Daemon: Hermes

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| hermes | 0.70 \| 700 MiB | 0.70 \| 700 MiB | 0.05 \| 140 MiB | OK |

# Daemon: Judge Injector

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| judge-injector (2) | 0.70 \| 8000 MiB | 0.70 \| 8000 MiB | 0.003 \| 100 MiB | OK<br>Limits are so high because was problematic in the past. |

Riccardo.Di.Maria@cern.ch

# Daemon: Judge Evaluator

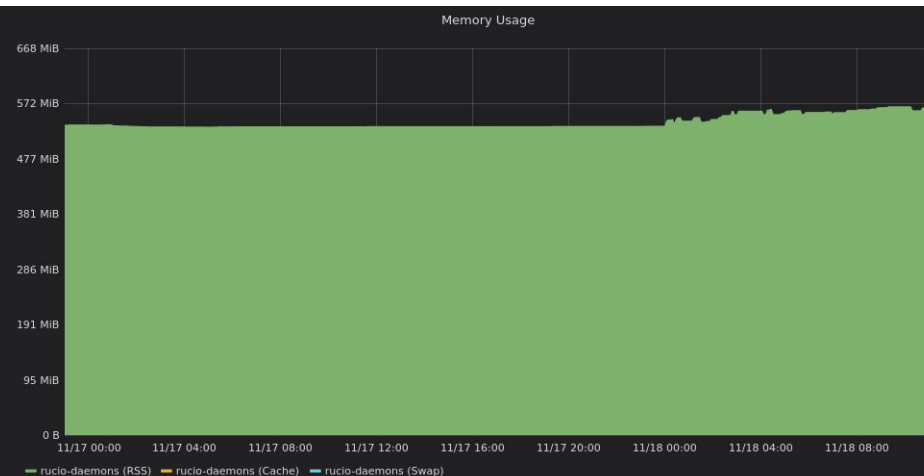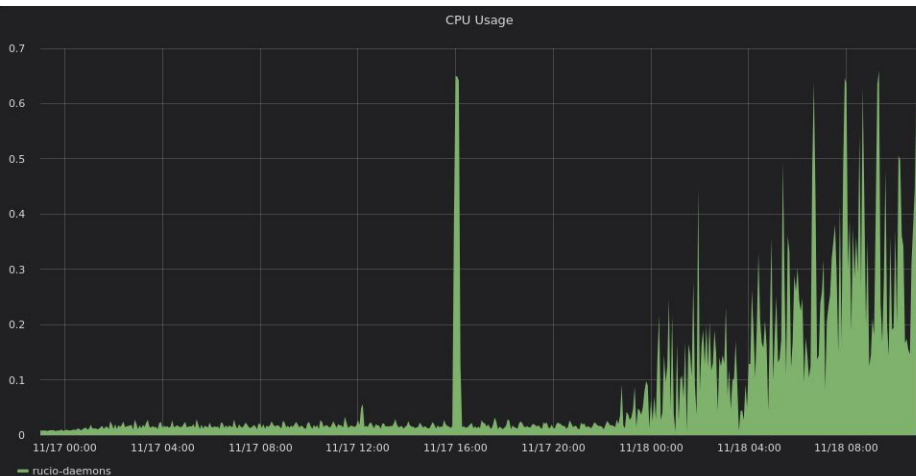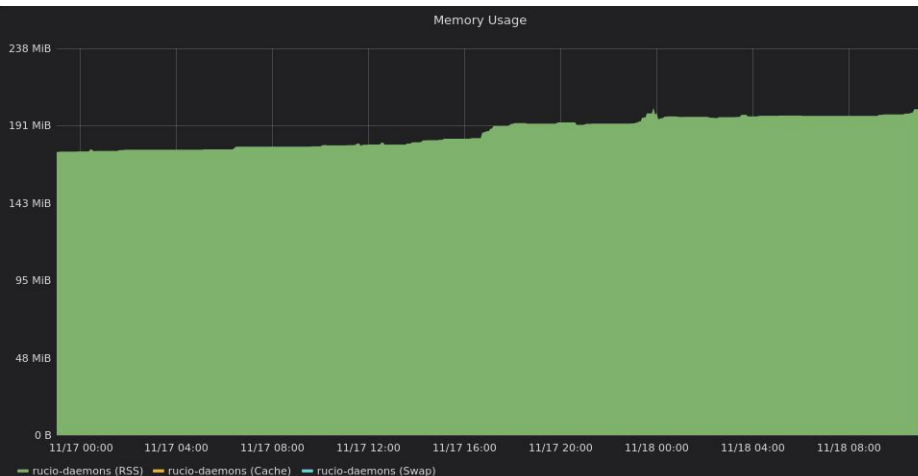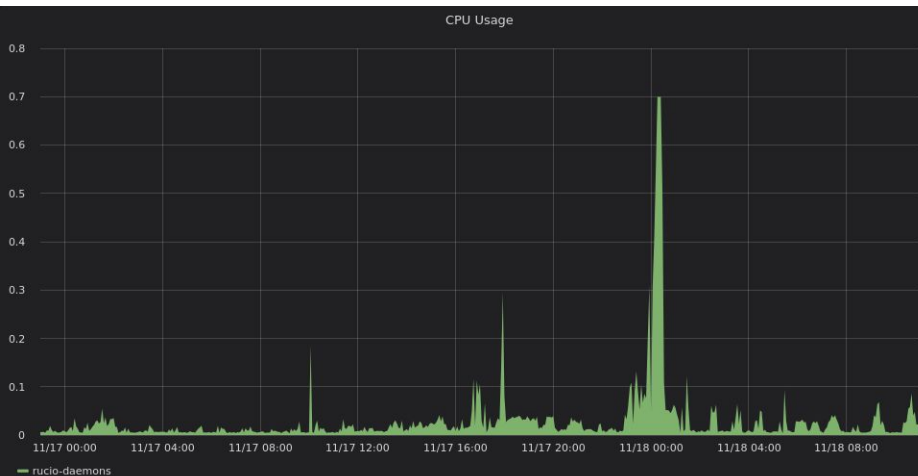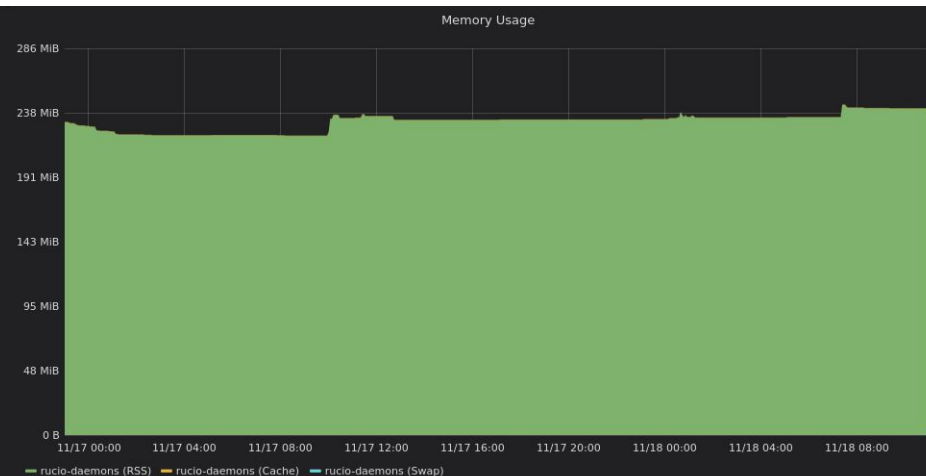| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| judge-evaluator (3 x 3 threads) | 2 \| 2000 MiB | 2 \| 2000 MiB | 1.3 \| 1-2 GiB | Problematic 2, 9, 2 restarts. Tried different configurations (2 x 5 threads). Not understood. |

# Daemon: Judge Evaluator

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| judge-evaluator (3 x 3 threads) | 2 \| 2000 MiB | 2 \| 2000 MiB | 1.3 \| 1-2 GiB | Problematic 2, 9, 2 restarts. Tried different configurations (2 x 5 threads). Not understood. |



November 26th, 2020          Riccardo.Di.Maria@cern.ch
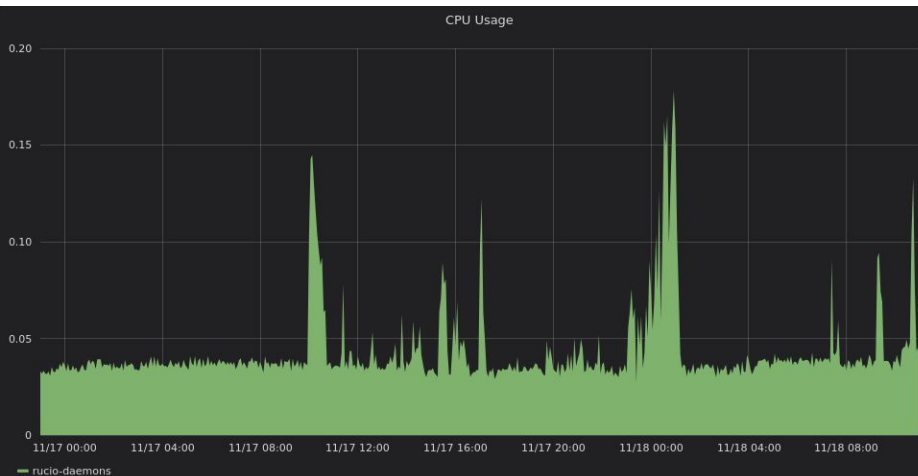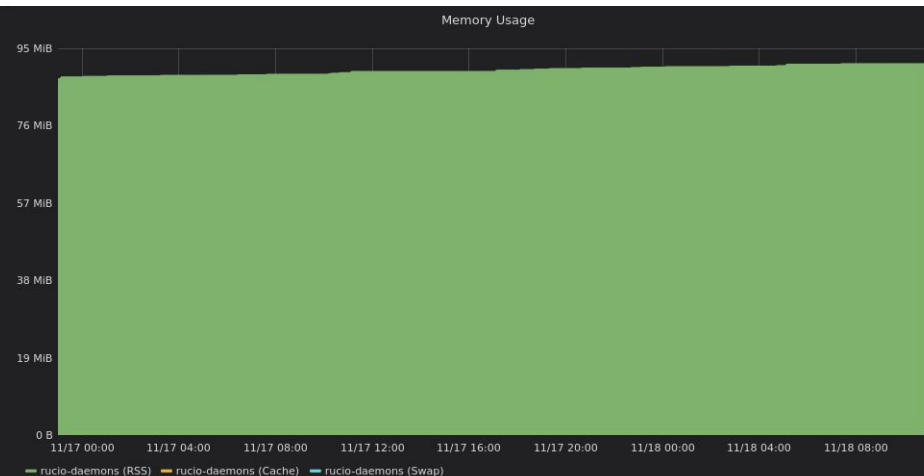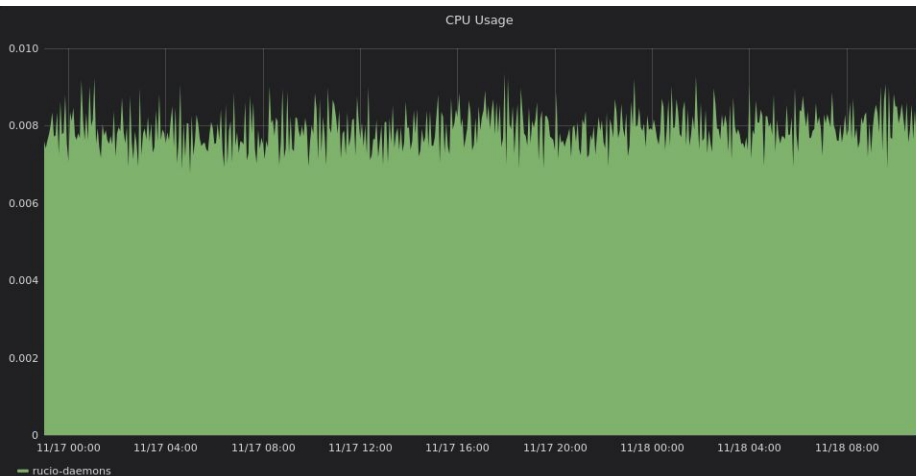
# Daemon: Judge Evaluator

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| judge-evaluator (3 x 3 threads) | 2 \| 2000 MiB | 2 \| 2000 MiB | 1.3 \| 1-2 GiB | Problematic 2, 9, 2 restarts. Tried different configurations (2 x 5 threads). Not understood. |

# Daemon: Judge Repairer

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| judge-repairer (2 x 5 threads) | 0.70 \| 1000 MiB | 0.70 \| 1000 MiB | 0.2 (0.7) \| 600 MiB | OK |

# Daemon: Judge Cleaner

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| judge-cleaner (2 x 5 threads) | 0.70 \| 500 MiB | 0.70 \| 500 MiB | 0.1 (0.7) \| 200 MiB | OK |

# Daemon: Reaper2

| Service | Resources Requests [CPU | Memory] | Resources Limits [CPU | Memory] | FDR Usage of Resources [CPU (peak) | Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| reaper2 (2 x 4 threads) | 0.70 | 400 MiB | 0.70 | 400 MiB | 0.05 (0.2) | 250 MiB | OK |

Riccardo.Di.Maria@cern.ch

# Daemon: Transmogrifier

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| transmogrifier | 0.70 \| 200 MiB | 0.70 \| 200 MiB | 0.01 \| 100 MiB | OK |

# Daemon: Undertaker

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| undertaker | 0.70 \| 400 MiB | 0.70 \| 400 MiB | 0.03 (0.7) \| 100 MiB (200 MiB) | OK Memory limit raised at 400 MiB due to OOMKilled at 200 MiB. |
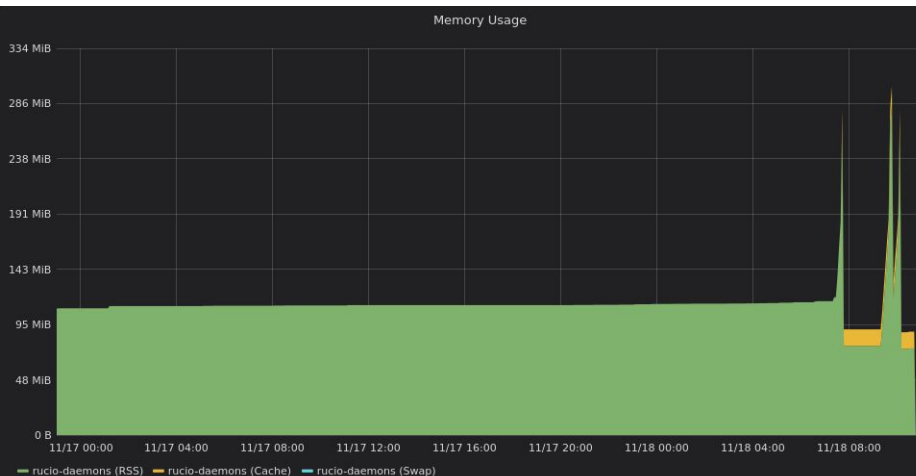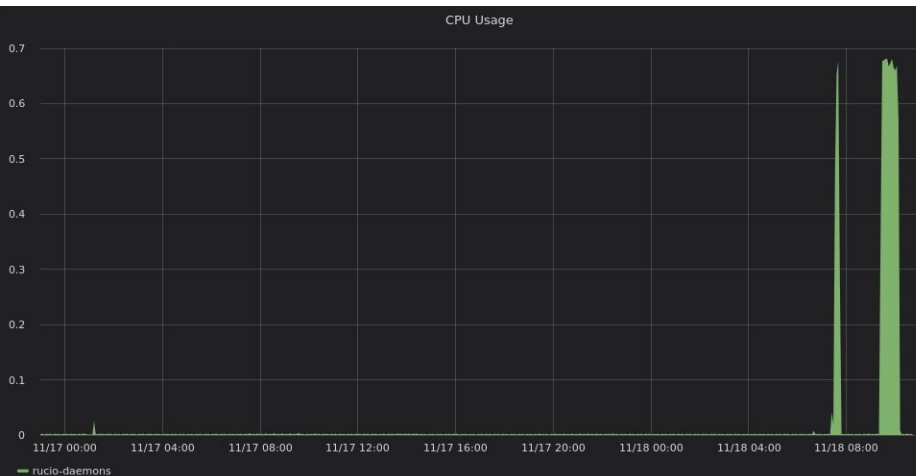
- Rucio → (#replicas) [CPU (limits)|Memory]:

  - UI → (1) [0.1|500 (800) MiB]

  - Auth Server → (2) [0.2 (*1)|0.5 (1) GiB]

  - Main Server → (2) [2 (4)|2 (4) GiB]

  - Daemons:

    - **Abacus Account** → (1) [0.1|150 MiB]

    - **Abacus Collection Replica** → (1) [1|4 (*8) GiB]

    - **Abacus RSE** → (1) [0.1|150 MiB]

    - **Conveyor Submitter** → (3 x 4 threads) [0.8|400 MiB]

    - **Conveyor Poller** → (3 x 4 threads) [0.5|250 MiB]

    - **Conveyor Finisher** → (1 x 2 threads) [1(*1.5)|250 (*500) MiB]

- **Hermes** → (1) [0.1|200 MiB]

- **Judge Injector** → (2) [0.4 (*1)|1 (*10) GiB]

- **Judge Evaluator** → (3 x 3 threads) [2|3 GiB]

- **Judge Repairer** → (2 x 5 threads) [1|800 MiB]

- **Judge Cleaner** → (2 x 5 threads) [1|400 MiB]

- **Reaper2** → (2 x 4 threads) [0.4|400 MiB]

- **Transmogrifier** → (1) [0.1|200 MiB]

- **Undertaker** → (1) [1|400 MiB]

- Total → (29) [23.4|27 GiB]
  → *26.7|54.55 GiB

- OpenStack VMs → (6 nodes) [8|16 GiB]

Riccardo.Di.Maria@cern.ch

# Daemon: Conveyor Finisher - FDR Takeaway

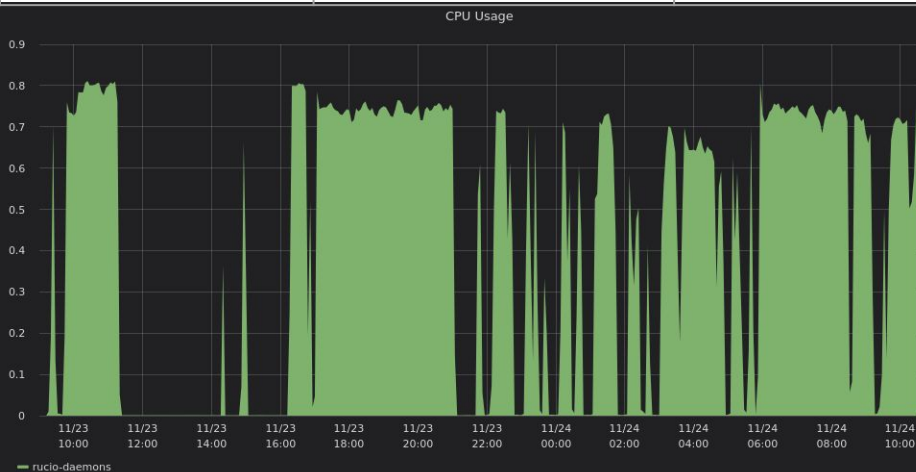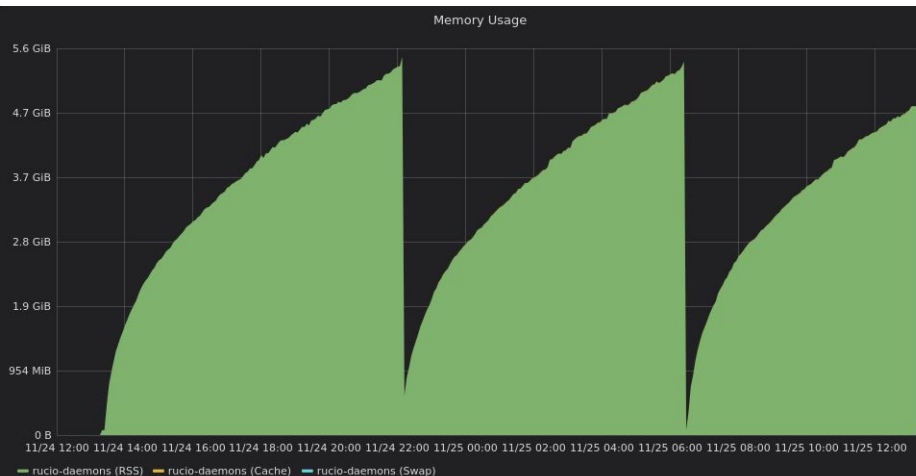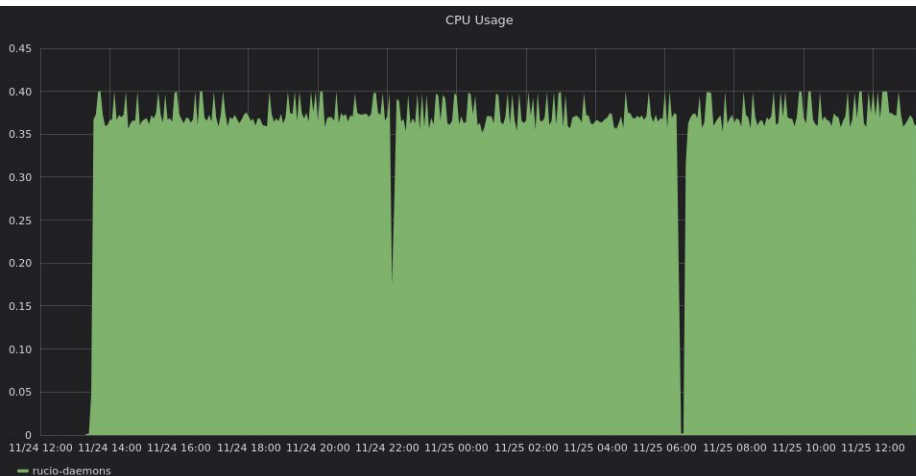| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| conveyor-finisher (2 threads) | 1 \| 250 MiB | 1 \| 250 MiB | 0.7 (1) \| 200 (400) MiB | OK-ish<br>2 restarts.<br>Limits changed again to:<br>(*1.5) \| (*500) MiB |

# Daemon: Abacus Collection Replica - FDR Takeaway

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | Post-FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---------|-----------------------------------|----------------------------------|----------------------------------------------------------|-------------------|
| abacus-collection-replica | 1 \| 4 GiB | 1 \| 6 GiB | 0.75 \| 4.5 (>6) GiB | OK-ish Restarts due to OOMKilled. Cannot keep up. Millions of rows in UPDATED_COL_REP. |



November 26th, 2020     Riccardo.Di.Maria@cern.ch

# Daemon: Abacus Collection Replica - FDR Takeaway

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | Post-FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| abacus-collection-replica | 1 \| 4 GiB | 1 \| 6 GiB | 0.8 \| 4.5 (>6) GiB | Restarts due to OOMKilled. Cannot keep up. Millions of rows in UPDATED_COL_REP. Limits changed again to: (*8) GiB |

Riccardo.Di.Maria@cern.ch

# Daemon: Judge Injector - FDR Takeaway

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | Post-FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| judge-injector (2) | 0.2 \| 400 MiB | 0.4 \| 5.5 GiB | 0.03 \| 80 MiB | OK |
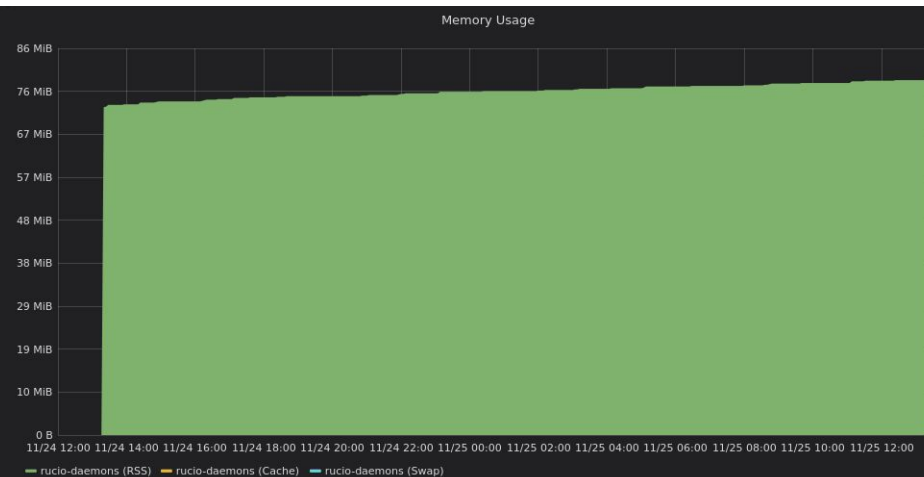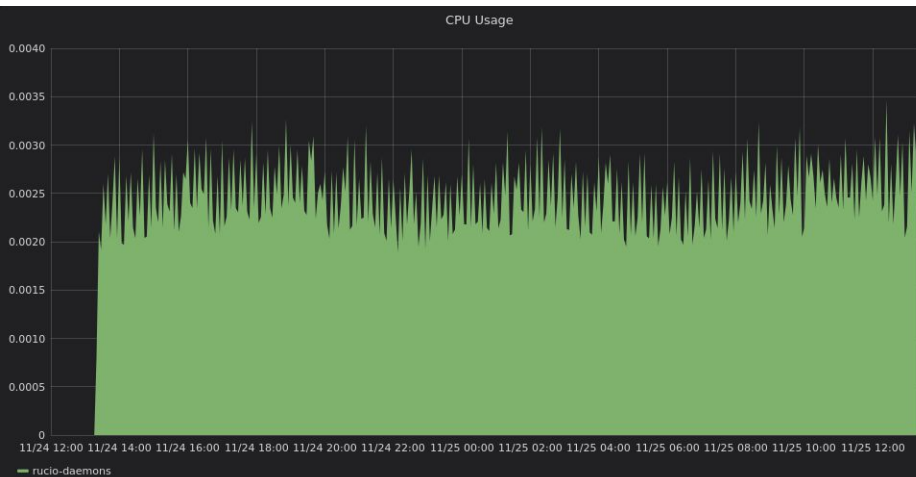
# Daemon: Judge Injector - FDR Takeaway

| Service | Resources Requests [CPU \| Memory] | Resources Limits [CPU \| Memory] | Post-FDR Usage of Resources [CPU (peak) \| Memory (peak)] | Restarts/Comments |
|---|---|---|---|---|
| judge-injector (2) | 0.2 \| 400 MiB | 0.4 \| 5.5 GiB | 0.4 \| >6 GiB | Problematic<br><br>Limits changed again to:<br>0.4(*1) \| 1(*10) GiB |

# Rucio Events

November 26th, 2020

Riccardo.Di.Maria@cern.ch

# Rucio Stats

# Rucio Stats

November 26th, 2020

Riccardo.Di.Maria@cern.ch

# Rucio Stats

| Experiment | Number of Files/DIDs Injected | Experiment | Number of Files/DIDs Injected |
|---|---|---|---|
| CTA | 508.9 K | MAGIC | 516+773 |
| LSST | 164.8 K | LOFAR | 426 |
| SKA | 127.7 K | CMS | 385 |
| EGO-VIRGO | 14.9 K | FAIR | 18+124 |
| ATLAS | 3308 | | |

November 26th, 2020

Riccardo.Di.Maria@cern.ch

# On-Going and To-Do-ASAP

- **Upgrade infrastructure as a primary goal → Rucio(daemons) + k8s + CERN-GitOps + DB**

- Improve tests (incl. HammerCloud) and monitoring

- **Collaboration for a second/test rucio instance (DB at CERN for the time being)**

- **Enable or develop more features, e.g. Rucio multi-VO, tokens, etc.**

- Investigate data corruption

- HammerCloud ready to run realistic research infrastructure workloads

- Real data distribution and analysis for non-HEP RI (LOFAR, CTA, LSST, MAGIC)

- Ability to plug heterogeneous clouds (commercial)

- Changing QoS within a site and across sites

Riccardo.Di.Maria@cern.ch

Start 02/2019

End 08/2022

Pilot DataLake design and implementation plan ✓

Workshop pilot DataLake performance analysis

Expanded Prototype. Verify RI data access from compute platforms (inc. clouds)

**September 2019**

**August 2020**

**November 2020**

**January 2021**

**February 2021**

**June 2021**

Initial DataLake implementation with at least 3 centers ✓

Analysis and assessment of the pilot DataLake

Full analysis and assessment of the full DataLake prototype

Riccardo.Di.Maria@cern.ch

# Backup

November 26th, 2020      Riccardo.Di.Maria@cern.ch

# Next Steps

- 2020: Pilot DataLake Functional Data Transfer Tests Machinery.
  - Data Transfer Tests Machinery to demonstrate stable and sizeable data movement across sites in the DataLake.
    - Moving bulk data from A to B (to C), changing QoS within a site and across sites (on demand, by policy).
    - DataLake data can be accessed by clients (e.g. simple workloads).
  - *Performance* monitoring in place (e.g. transfer matrix).
    - Show transfer metrics: bandwidth, number of files, success/errors, perfSONAR, etc.
      - Allowing us to debug the infrastructure: network, storage, data management tools, etc.
    - Basis for an operations and deployment model.

- 2021: Extending the Pilot to a Full Prototype DataLake.
  - Automated infrastructure testing (based on HC) ready to run realistic research infrastructure workloads.
  - Real data distribution and analysis for several non-HEP RI and HL-LHC reference workloads.
  - Ability to plug comercial cloud resources into the DataLake infrastructure.
  - Caching mechanism deployed, vanilla software suite ready and deployed on several sites.
  - Bonus: PoC integration of SA and AUS sites into the DataLake infrastructure (SKA/LOFAR data).

- Continue the **synergy and feedback** with the WLCG DOMA R&D projects.