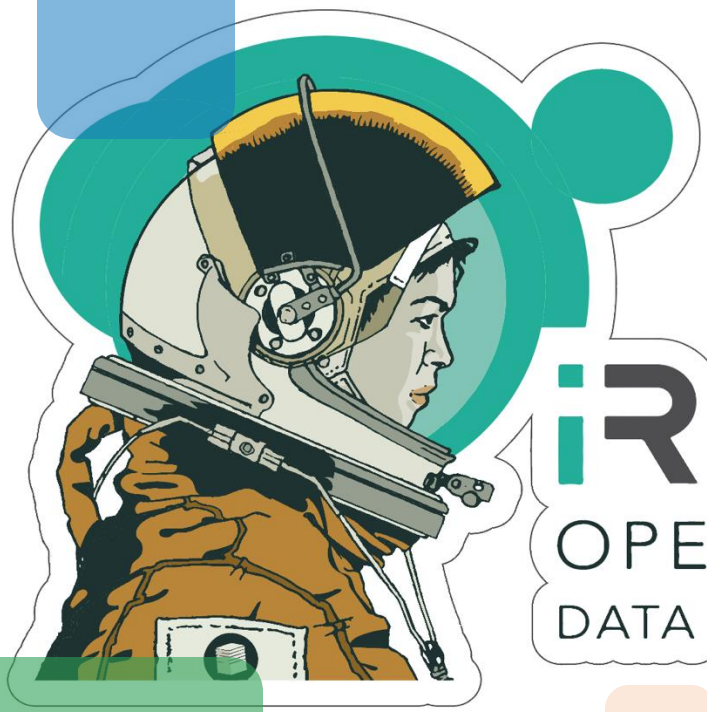
Enhancing iRODS Monitoring

Francisco Morales &
Alice Stuart-Lee

June 2025



iRODS
OPEN SOURCE
DATA MANAGEMENT

SURF

Who are we?

Collaborative organisation for ICT in Dutch education and research

DMS Team

Data management services

Francisco Morales

DevOps Engineer

francisco.morales@surf.nl

Alice Stuart-Lee

Technical Advisor

alice.stuartlee@surf.nl

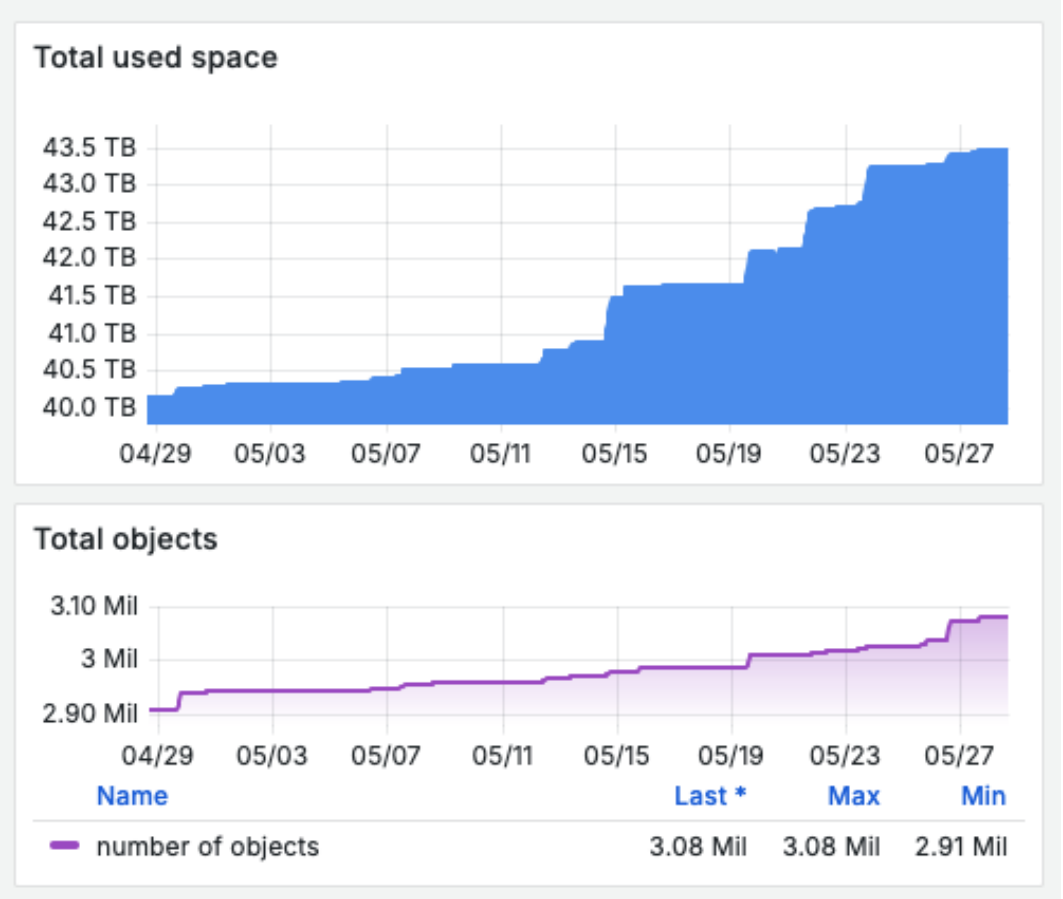
Claudio Cacciari

Sr. Technical Advisor

claudio.cacciari@surf.nl

Motivation

- 11 iRODS instances in production
  - Each with own zone
  - Importance of robust monitoring
- Visualised data
  - Service insights for us
  - Accessibility for customers



Our dashboards

- Choice of Grafana
- Standard set of dashboards
- Customers have read-access
- Demo...

SURF

To the dashboards...

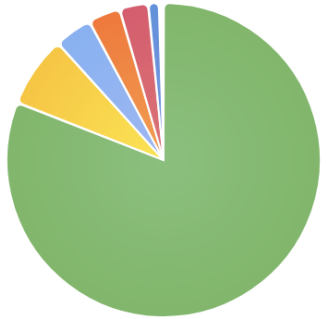
Service rdm_yoda_1 | Overall status | Configuration | Users | Storage resources | Data integrity | Last 2 days | Refresh

iRODS service

up

surf-yoda.irods.surfsara.nl

up 20 days, 6:2



irodsResc

UP

irodsResc: % used space

374.53%

irodsResc: used sp...

3.75 TB

irodsResc: data inte...

8.33%

irodsResc: % dupli...

5.69%

surfObjStore1

UP

surfObjStore1: % used s...

4.15%

surfObjStore1: used...

41.48 GB

surfObjStore1: data ...

6.80%

surfObjStore1: % d...

64.91%

surfArchive

UP

surfArchive: % used sp...

0.12%

surfArchive: used s...

11.55 GB

surfArchive: data in...

91.1%

surfArchive: % dup...

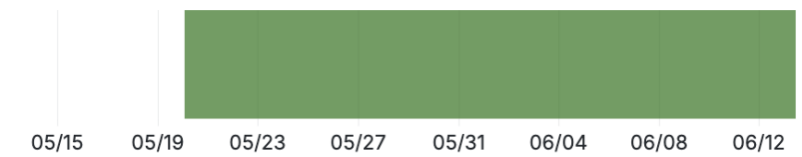
0.00%

Service rdm_yoda_1

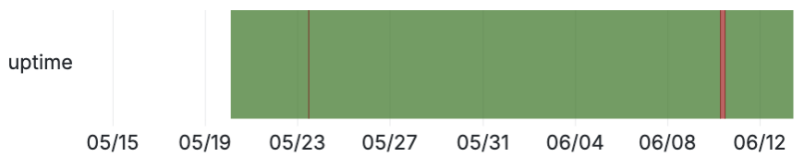
Last 30 days

Refresh

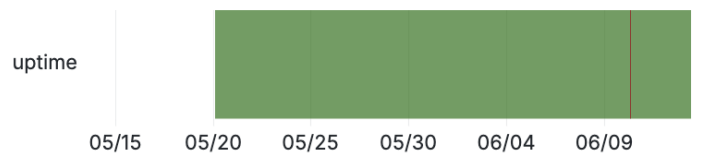
Disk resource uptime



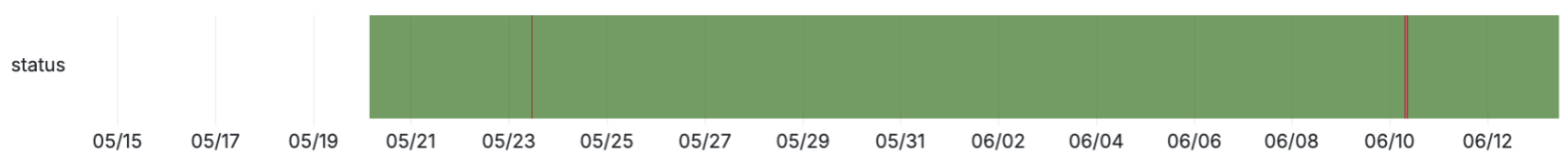
Object store resource uptime



Archive resource uptime



iRODS service uptime



Total used space

3.80 TB

Trash

2.41 TB

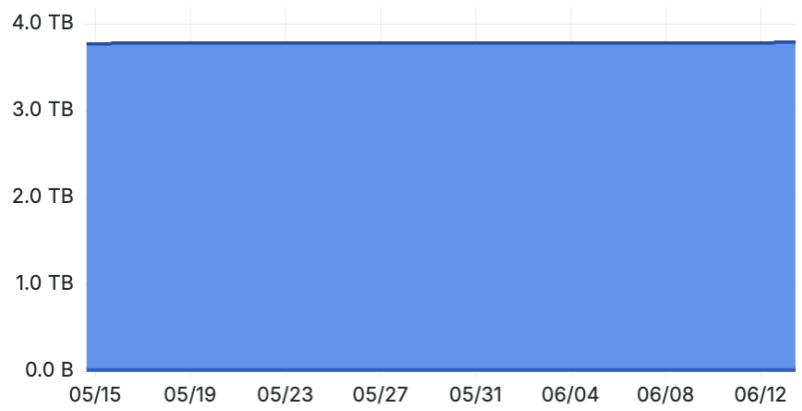
Total objects

32362

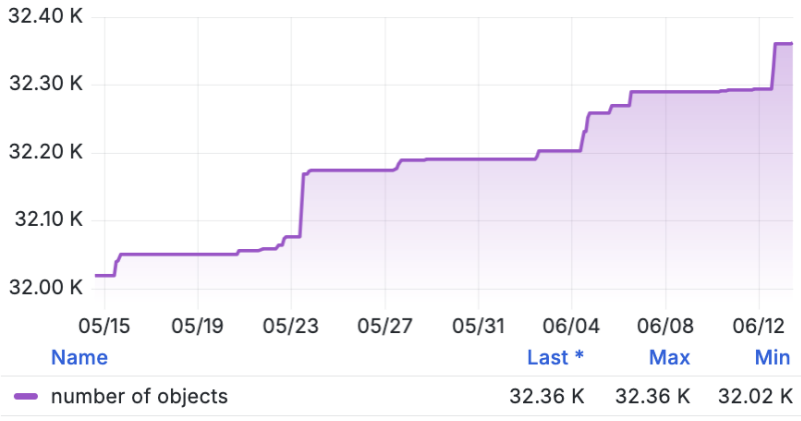
Total collections

8413

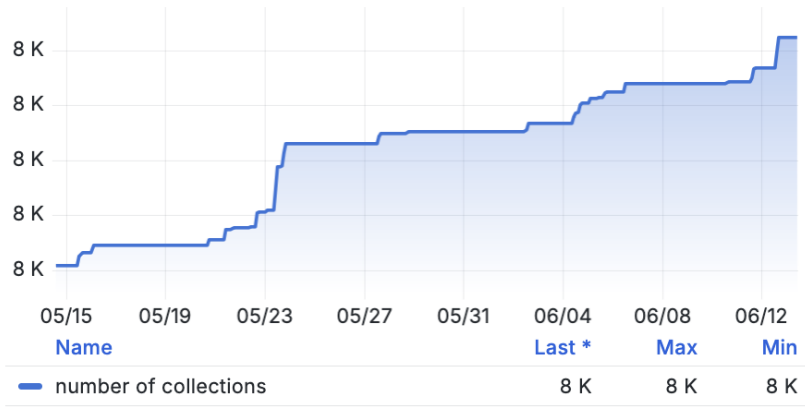
Total used space



Total objects



Total collections



Total duplicated space

254.49 GB

Total duplicated objects

10.3 K

Total duplicates

21.6 K

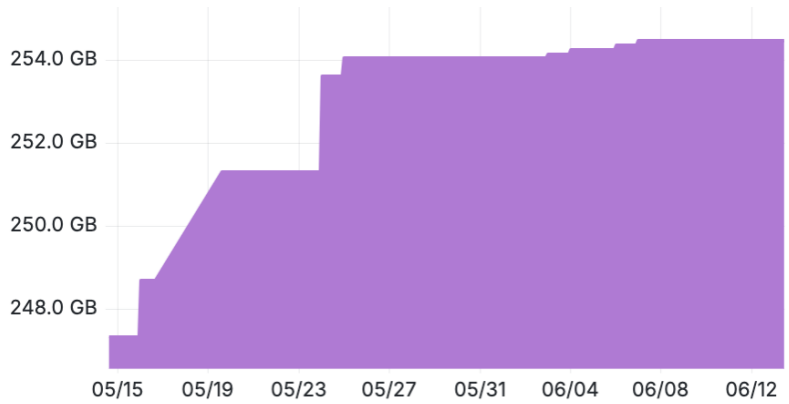
Total revisions space

71.11 GB

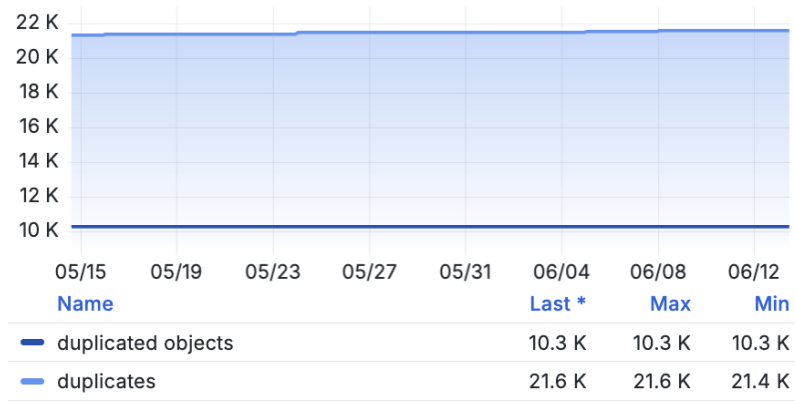
Total revisions objects

10.3 K

Total duplicated space

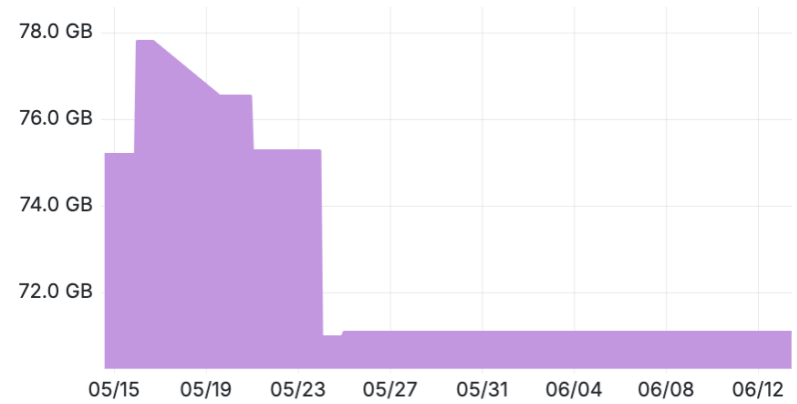


Total duplicates and duplicated

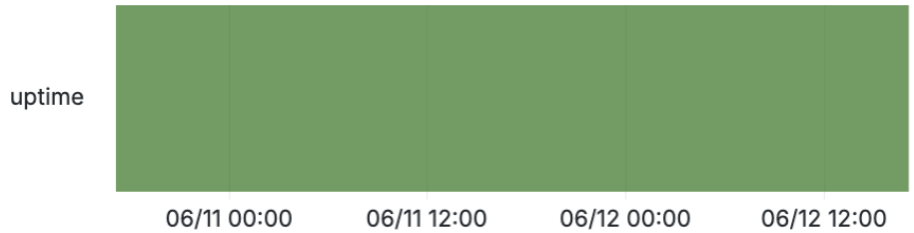


Name	Last *	Max	Min
duplicated objects	10.3 K	10.3 K	10.3 K
duplicates	21.6 K	21.6 K	21.4 K

Total revisions duplicated space



Resource irodsResc



quota

1 TB

used

3.75 TB

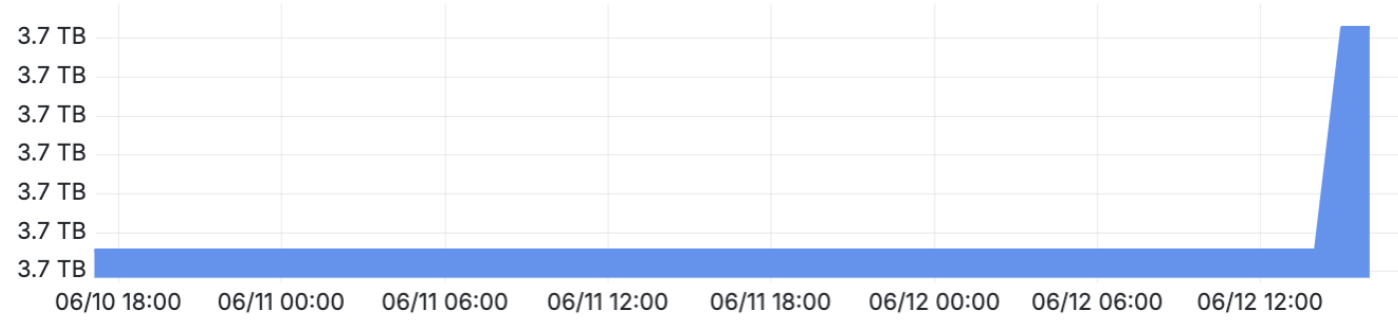
free

-2.75 TB

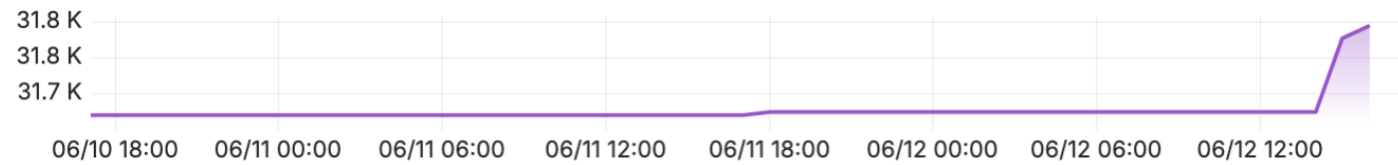
duplicated

213.01 GB

Total used space

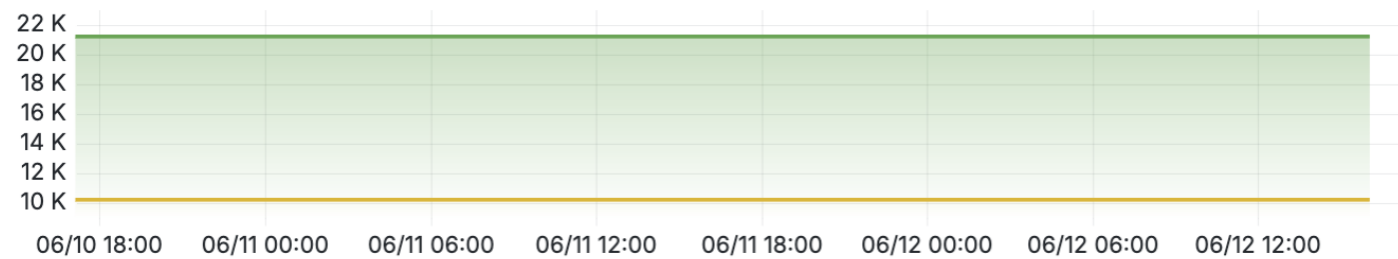


Total objects



Name	Last *	Max	Min
number of objects	31.8 K	31.8 K	31.7 K

Total duplicates



Name	Last *	Max	Min
number of duplicates	21.2 K	21.2 K	21.2 K
number of duplicated objects	10.3 K	10.3 K	10.3 K

Service

hostname	port range	zone	port	version	hash matching policy
surf-yoda.irods.surfsara.nl	[20000,20199]	yoda	1247	rods4.2.12	compatible

Disk quota	Object Store quota	Archive quota
1 TB	1 TB	10 TB

available rule engines

legacy rule language, default policy

available rule engines

Service rdm_yoda_1

Last 2 days

Refresh

Total groups

75

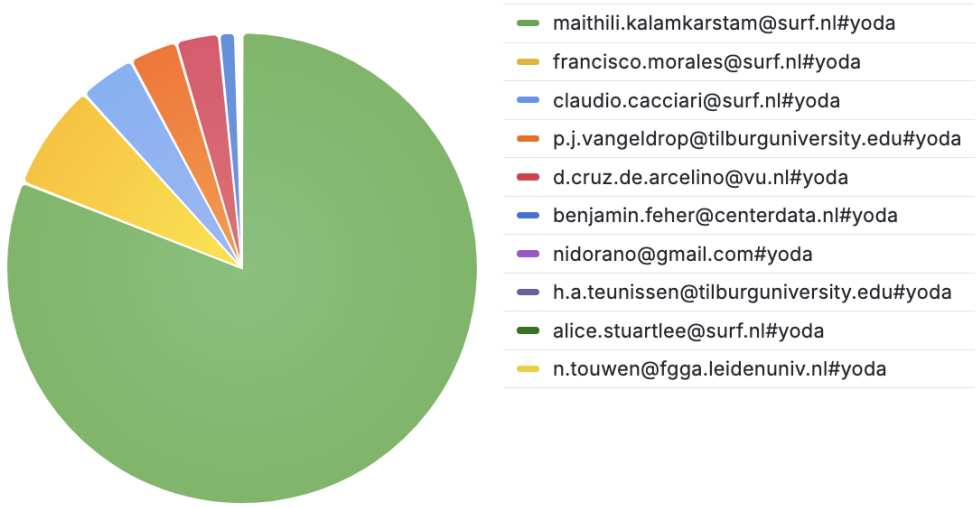
Total users

65

Total admins

3

TOP 10 users per used space

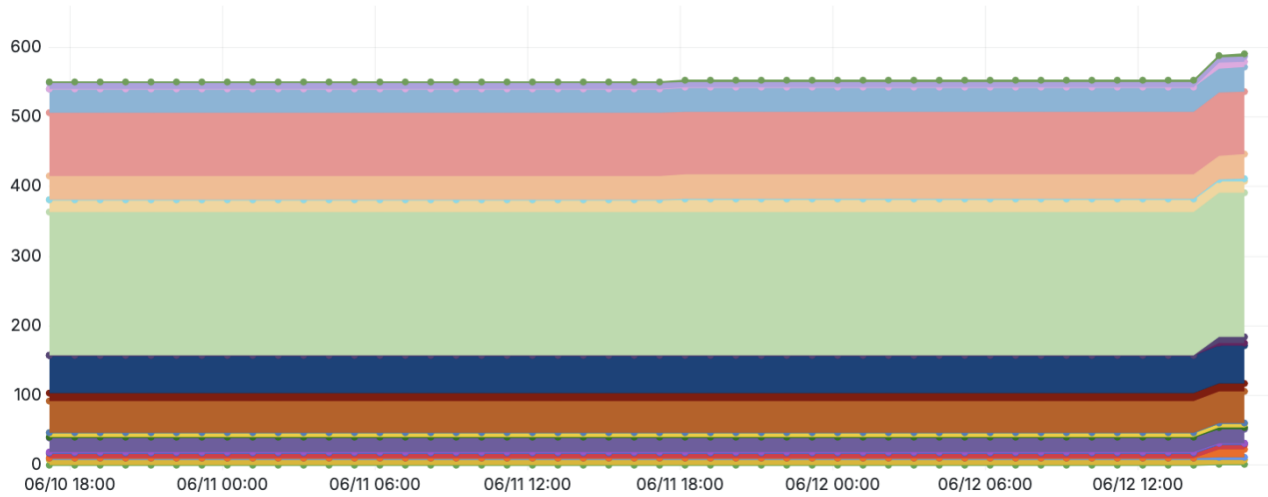


Data per user

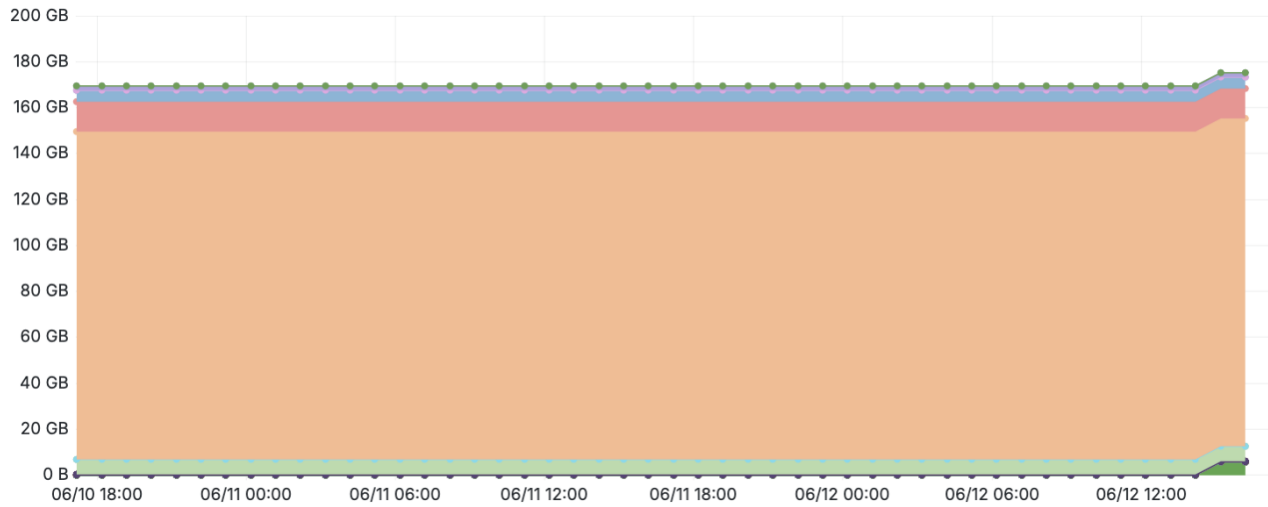
user	# objects	size ↓
maithili.kalamkarstam@surf.nl#yoda	35	143 GB
francisco.morales@surf.nl#yoda	90	12.8 GB
claudio.cacciari@surf.nl#yoda	206	6.71 GB
p.j.vangeldrop@tilburguniversity.edu#yoc	1	5.71 GB
d.cruz.de.arcelino@vu.nl#yoda	35	5.08 GB
benjamin.feher@centerdata.nl#yoda	8	1.82 GB
nidorano@gmail.com#yoda	3	105 MB
h.a.teunissen@tilburguniversity.edu#yoda	45	92.0 MB
alice.stuartlee@surf.nl#yoda	20	44.1 MB
n.touwen@fgga.leidenuniv.nl#yoda	6	9.54 MB
n.c.degroot@tilburguniversity.edu#yoda	4	9.42 MB

Data distribution per size

Object number trend



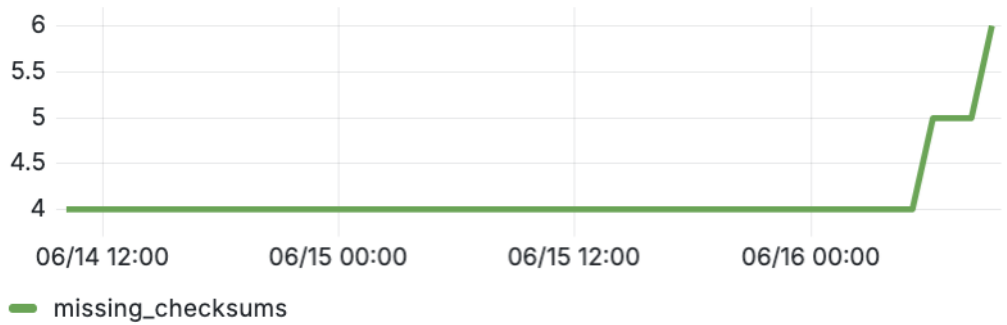
Data size trend



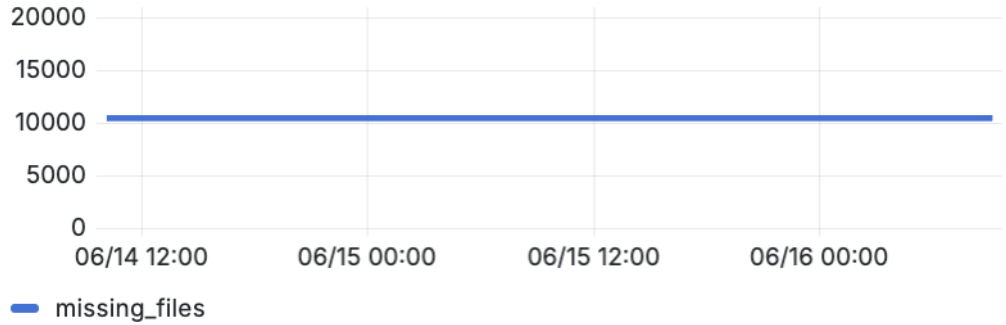
Resource irodsResc

Total objects	Missing checksums	Intermediate replicas	Missing files	Missing objects
31783	6	2	10478	43
	Percentage 0.019%	Percentage 0.006%	Percentage 32.967%	Percentage 0.201%

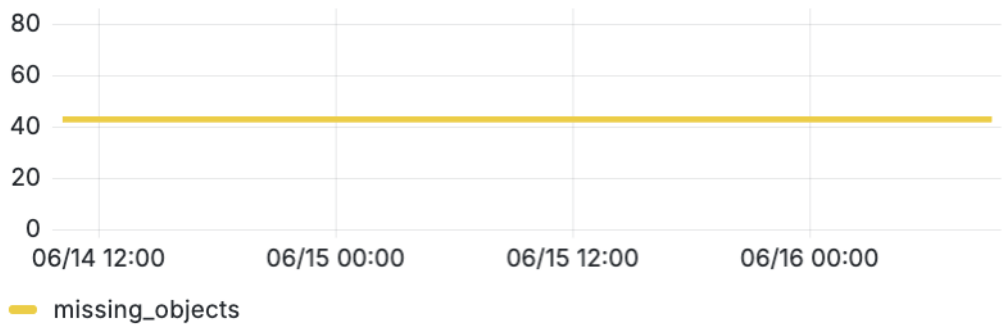
Missing checksums



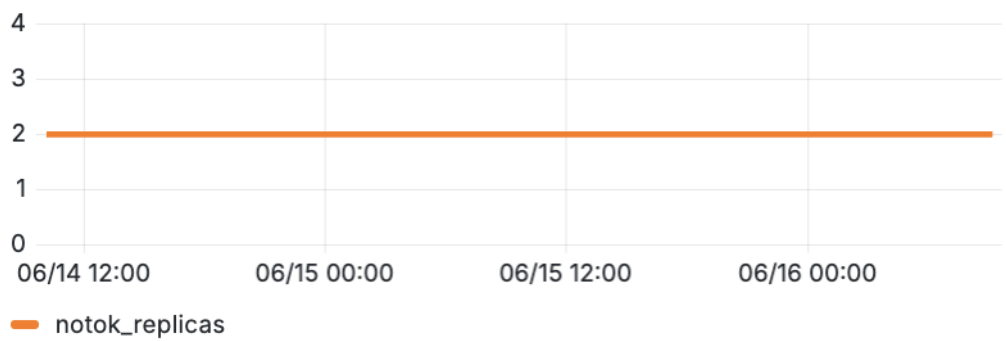
Missing files



Missing objects

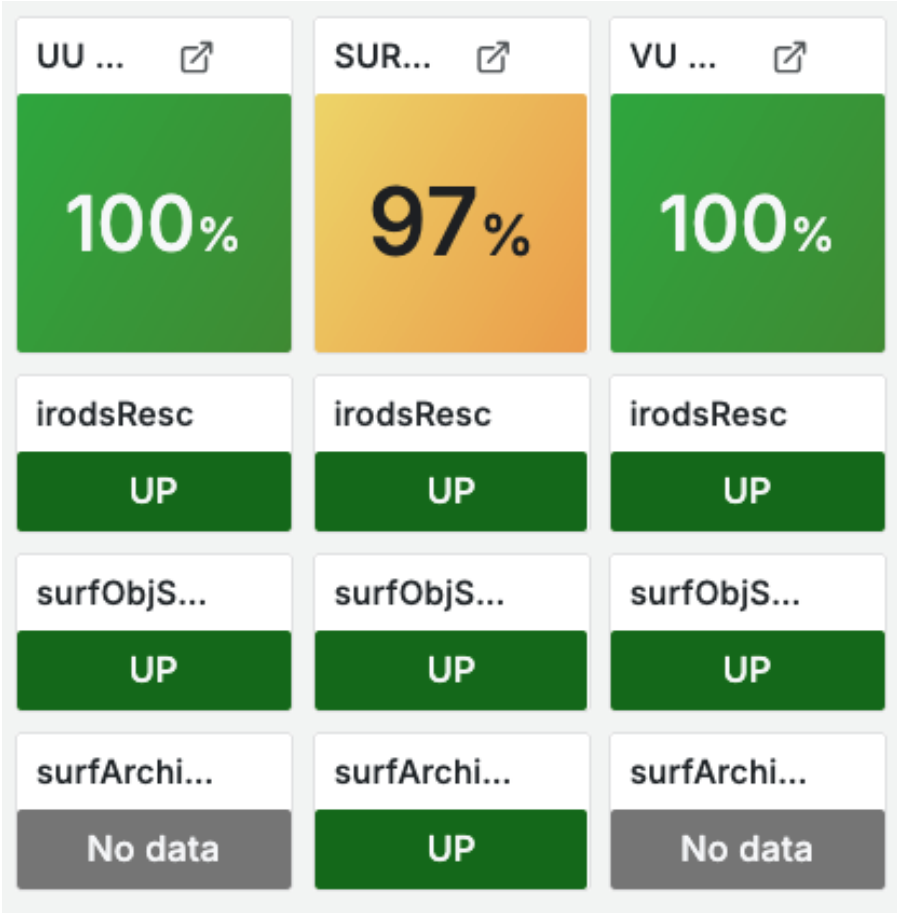


Intermediate replicas



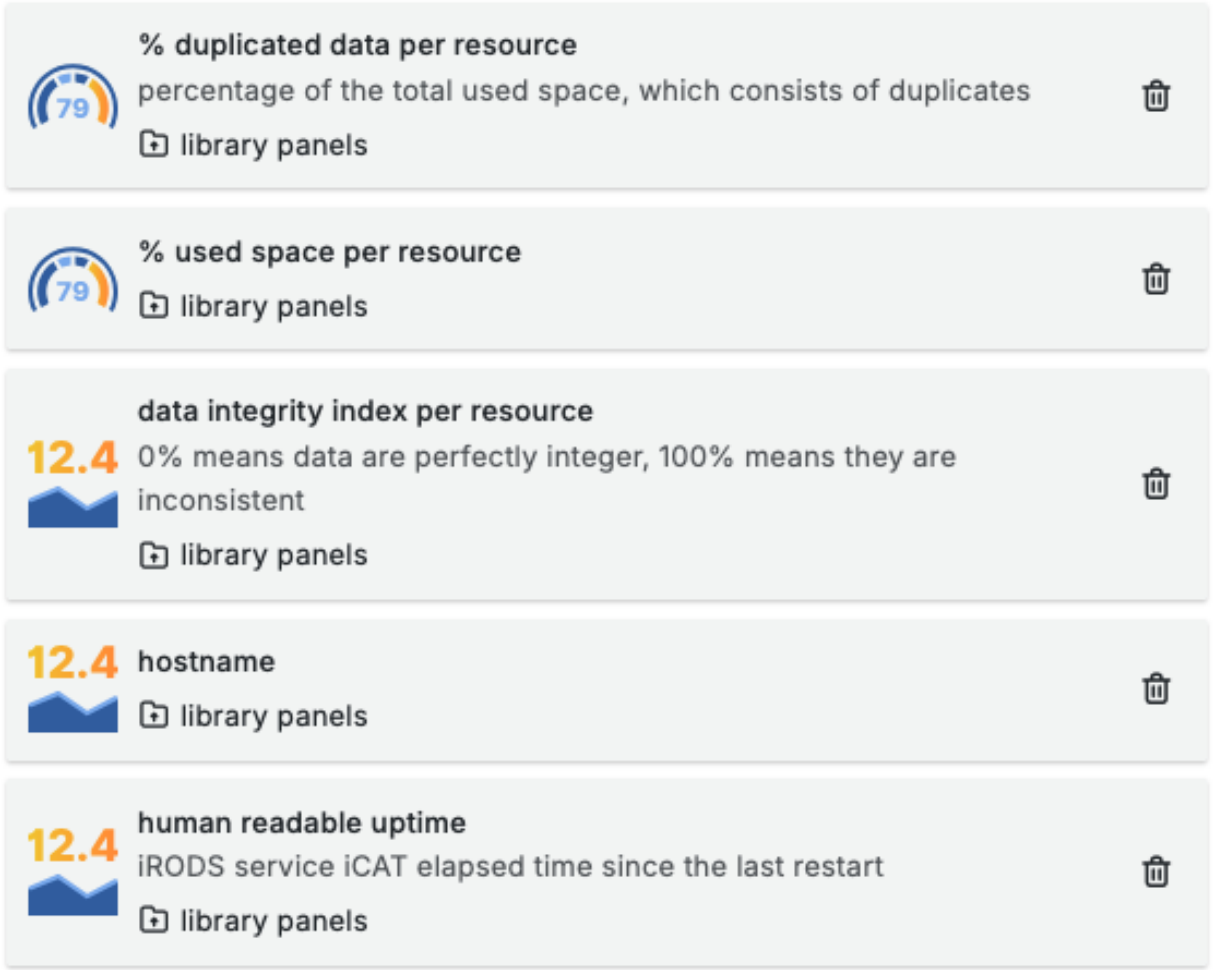
Alerts

- Passive signalling: overview dashboards
- Active signalling: alerts
  - Servers and resources uptime
    - Every 5 minutes
    - Email alerts to team
  - Next: storage usage



Automation

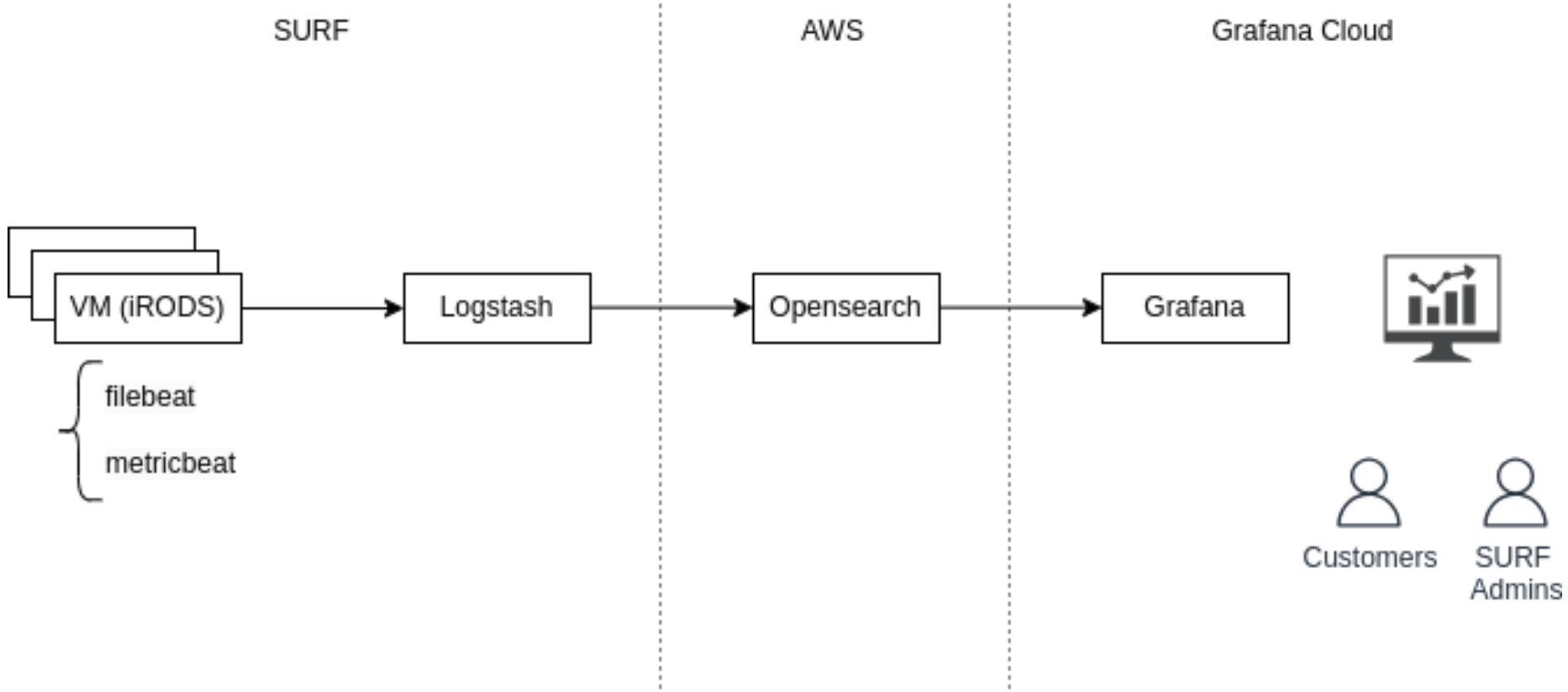
- Grafana library panels
  - Dashboard variables to make customer-specific
- Tailored automation
  - Python script/Grafana API
  - (Re)create dashboards
    - Dashboard templates
    - User config



Extracting the data

SURF

SURF Monitoring: Architecture



SURF Monitoring: metrics collection

- iRODS metrics:
  - Python iRODS client scripts
  - iRODS rule files
  - Bash scripts
  - Cronjob daily tasks for script execution

- System metrics
  - Via metricbeat

SURF Monitoring: pros & cons

- Pros:
  - Each VM collects and reports its own metrics. Customer isolation.
  - Opensearch and Grafana are outside SURF infrastructure. Grafana dashboards are still visible (with historic data) if SURF infra is unreachable.
  - Grafana dashboard access via SSO

- Cons
  - Logstash: single point of failure.
  - Opensearch shards are costly

SURF Monitoring: Operations

- Monitoring configuration management via internal git repo. Parameters are isolated per customer.
- Configuration deployment via salt
- User access: credentials via SSO and secure access via SSL.

SURF Monitoring: Next steps

- Better error handling/reporting (WIP) along the monitoring stack
- Add new metrics
  - Uptime of different components
- Dedicated analytics compute resources
  - Computing checksums, comparing large lists are CPU/memory intensive tasks
- Collect iRODS metrics via audit plugin

Questions?

SURF