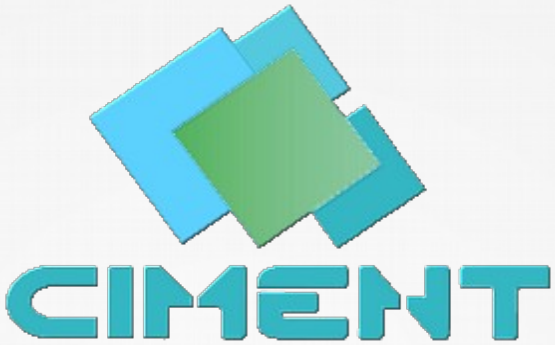


IRODS use case : Ciment High Performance Computing center

B.Bzeznik / X.Briand

Irods users group meeting

11/06/2015

Plan

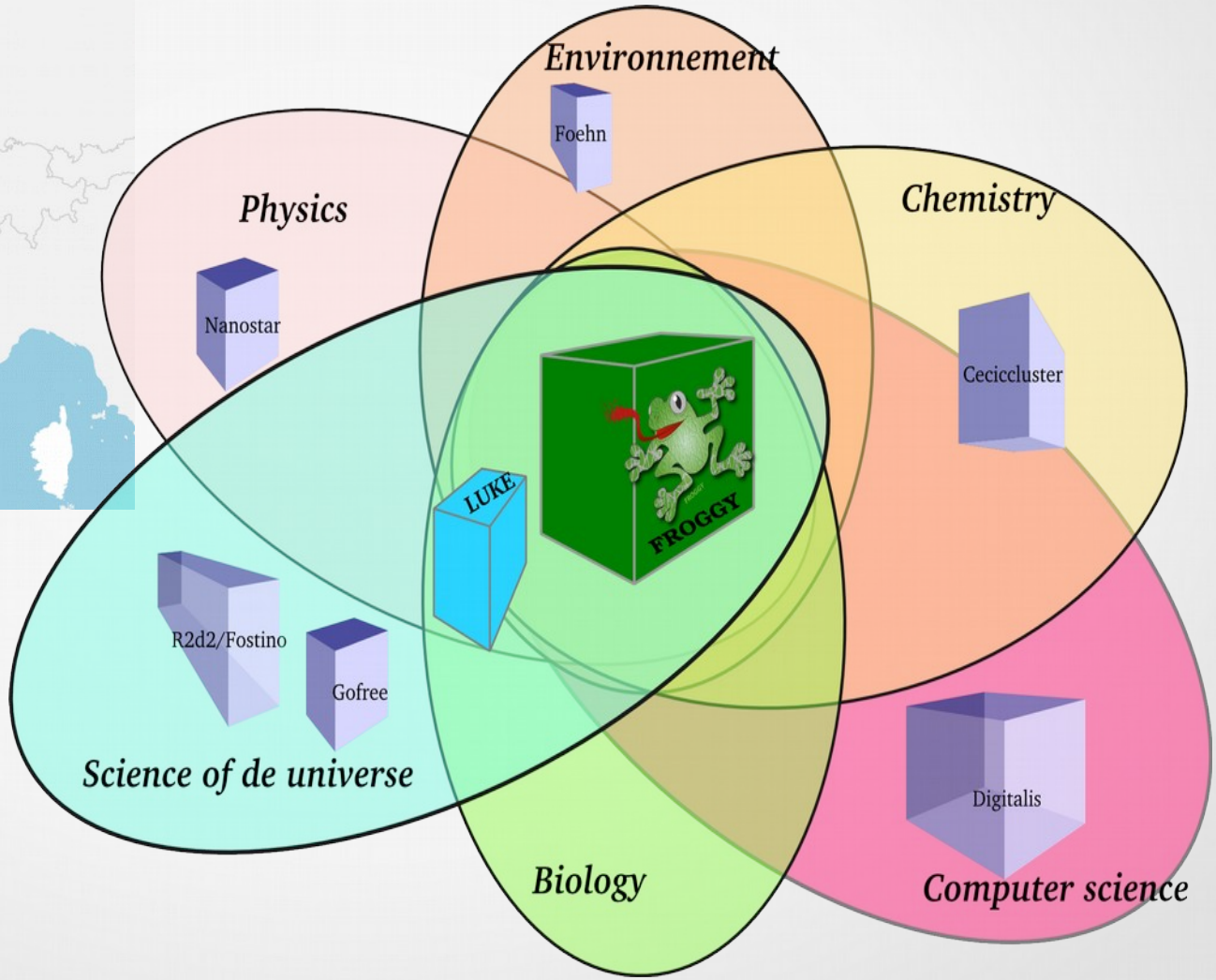
- **What is CIMENT?**
  - What is provided by CIMENT?
  - The batch scheduler: OAR
- **How does iRODS works on CIMENT platforms?**
  - Infrastructure
  - Cigri middleware: accessing ressources
  - Configuration of the IRODS grid environment
  - Load questions
  - Cirods and ciget: a usage of Pyrods API
- **Somes scientific applications**
  - Seismology
  - Rosetta mission
  - Ecology
  - Particle physics

CIMENT

What is CIMENT?

CIMENT : high performance computing center of the university of Grenoble

UNIVERSITE JOSEPH FOURIER
SCIENCES TECHNOLOGIE MEDECINE



CIMENT

Computing platforms

HPC platform  	**Data processing platform**  	**Other thematic platforms**
Froggy	*Luke*	*r2d2,foehn,ceciccluster,digitalis,...*
3200 Xeon SB cores @2.6Ghz +18 GPUS K20m	~400 cores - hetherogeneous	~3000 cores hetherogeneous into 10 small clusters
High performance distributed storage (Lustre): 90TB	Local scratches on nodes : 450 TB	NFS filesystems: a few TB per cluster
Infiniband FDR network	10 GBE network	Infiniband QDR networks

Common storage (IRODS) : 1PB

CIMENT

Jobs scheduling : OAR

- OAR is a resources and jobs management system
  - Made in Grenoble (LIG)
  - On all CIMENT platforms

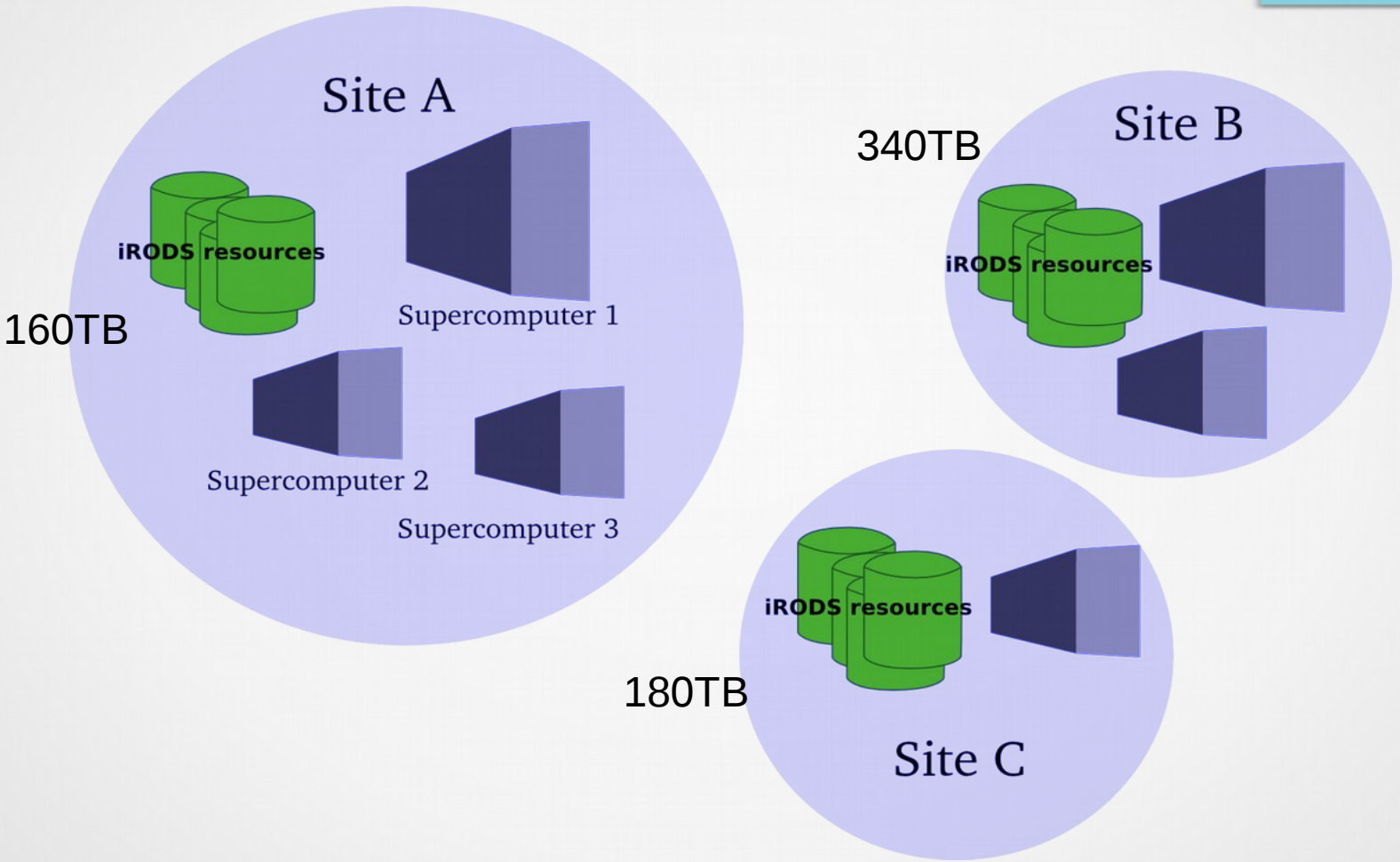
- Best-effort JOBS
  - Definition: opportunistic jobs having a 0 priority, that may be killed to let other jobs keep the resources
  - Optimization of the load of the platforms (best-effort jobs is a way to make use of the free resources)

CIMENT

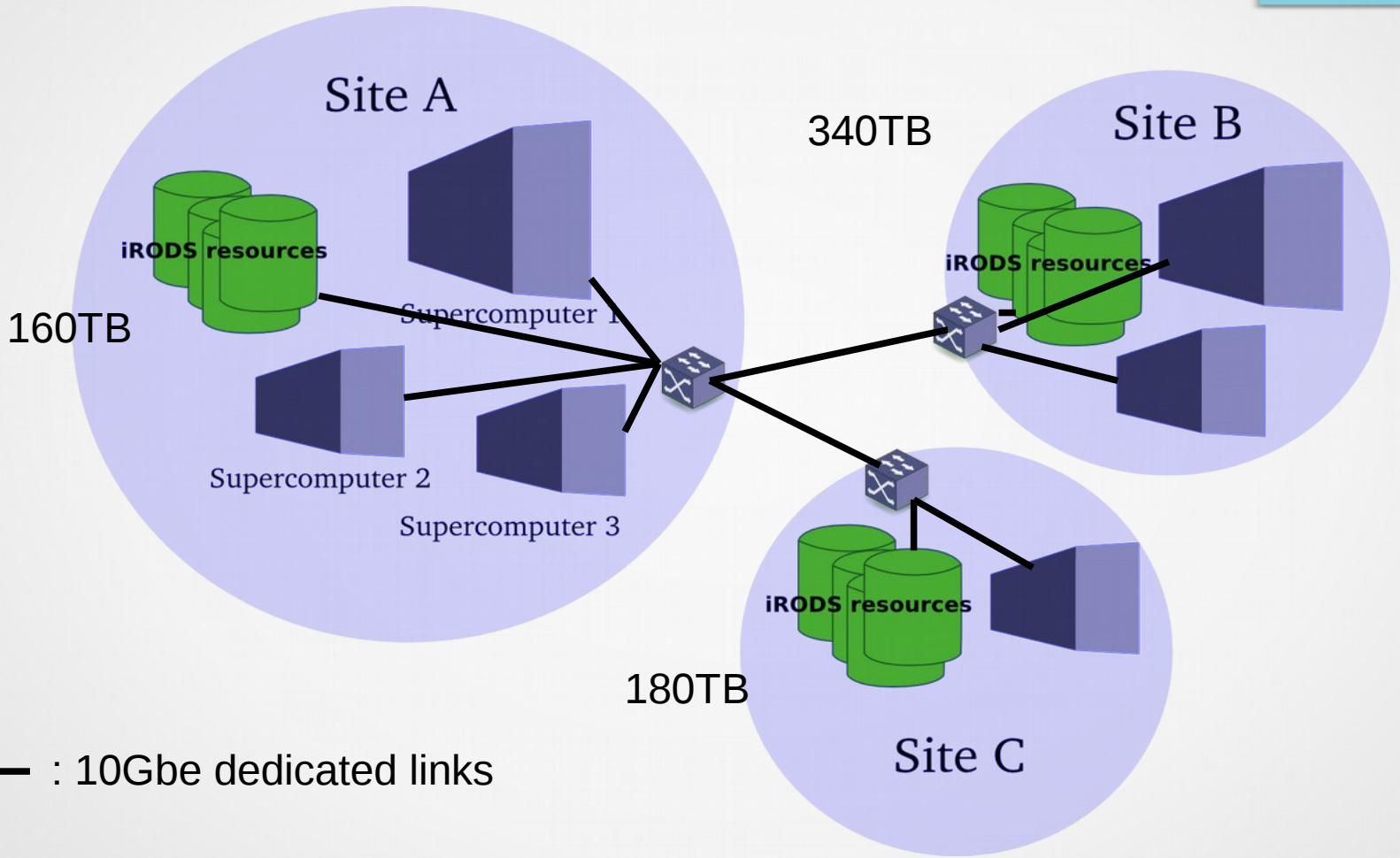
How does IRODS works on CIMENT platforms?

CIMENT

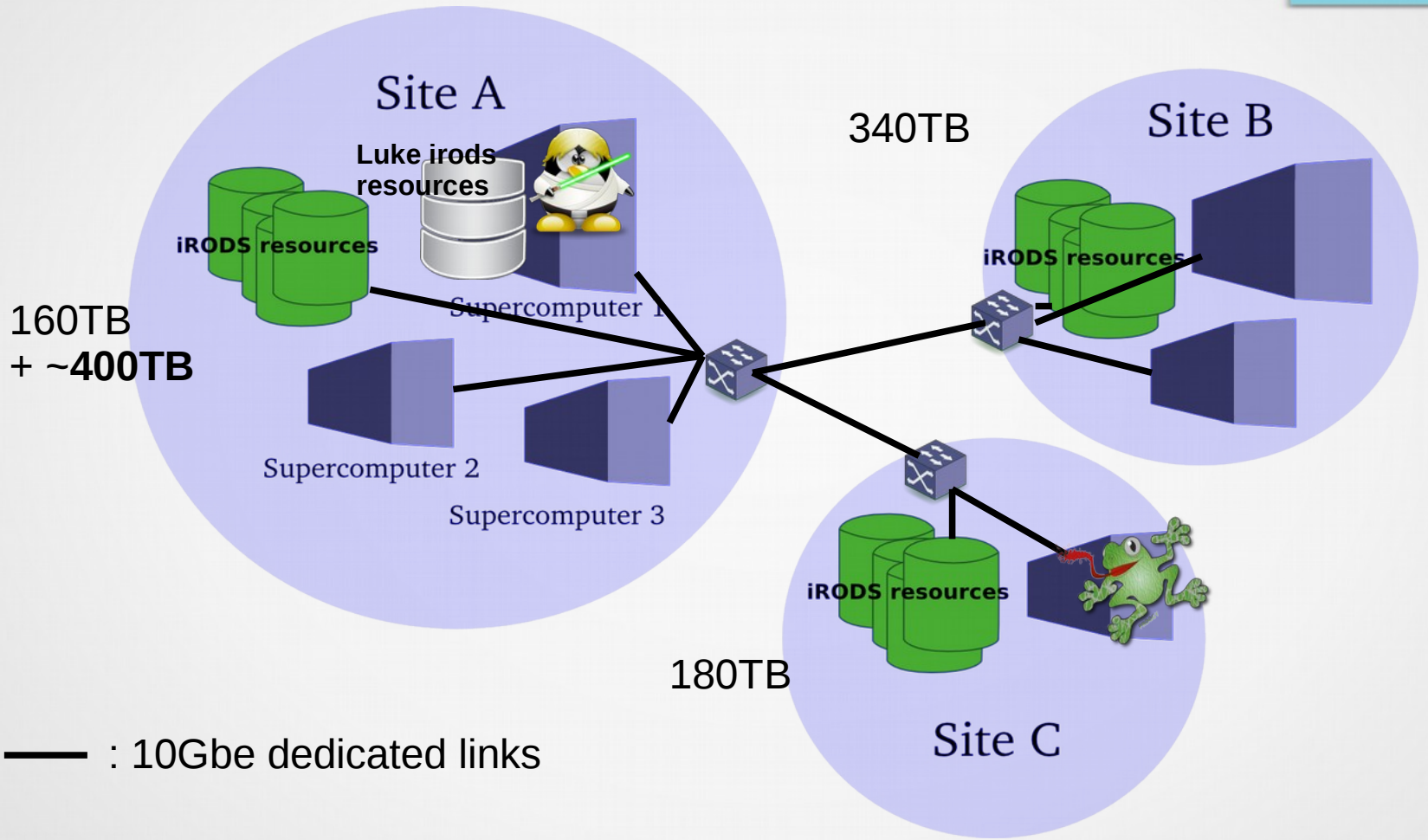
CIMENT IRODS infrastructure



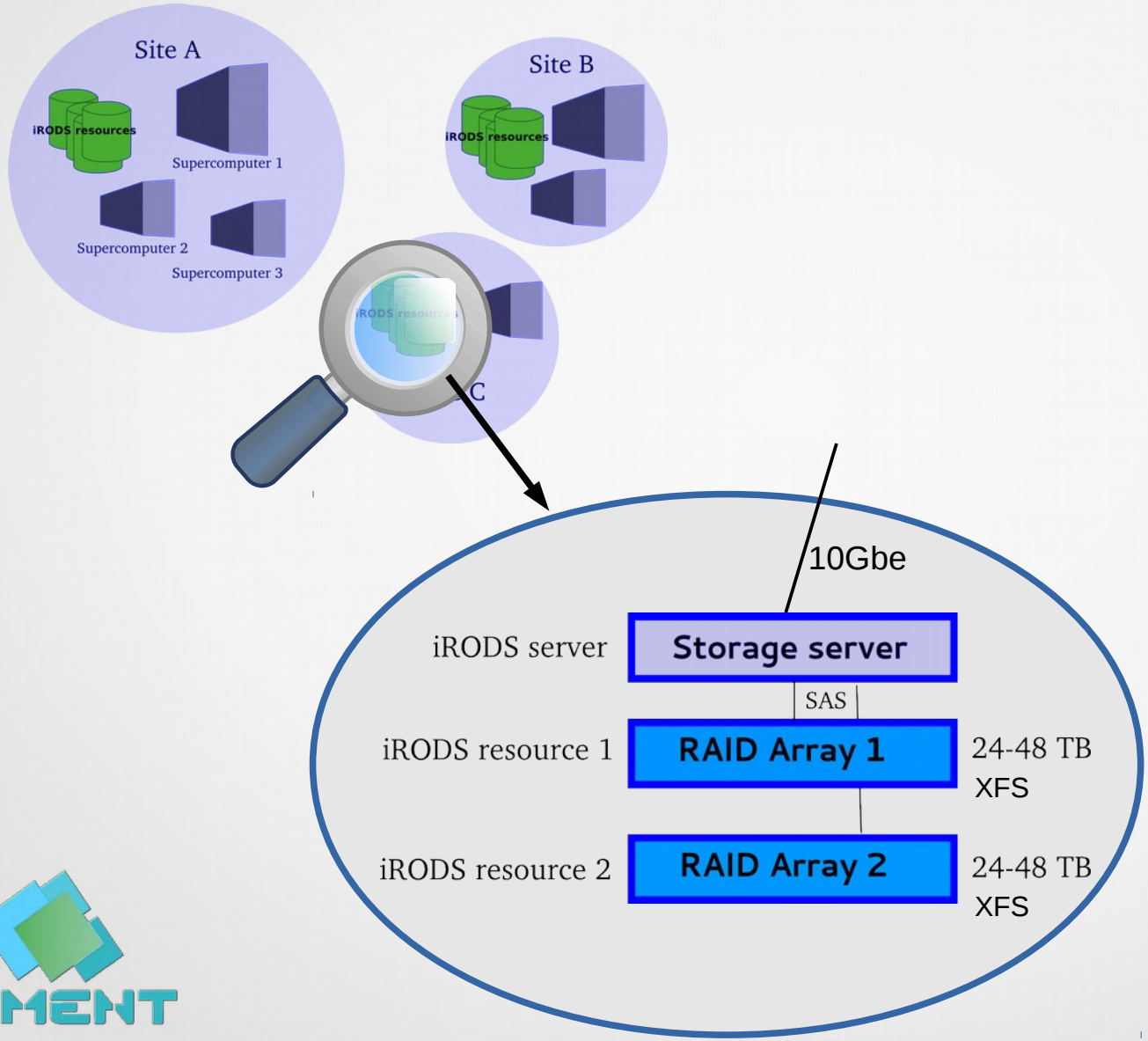
CIMENT IRODS infrastructure



CIMENT IRODS infrastructure



CIMENT IRODS infrastructure



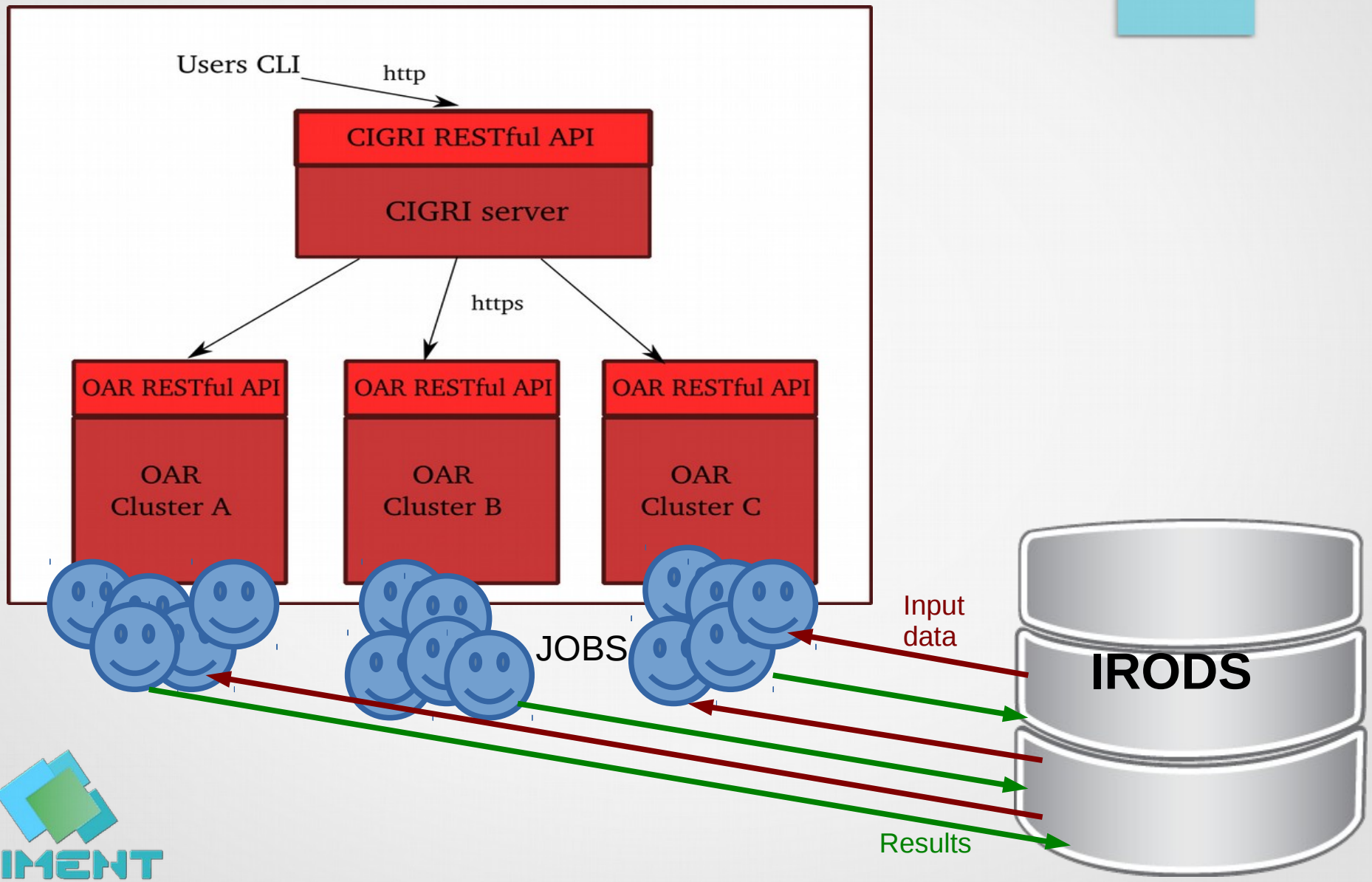
X 3 to 5
(depending on the site)

The CIGRI middleware

- A lightweight grid middleware
- Developed at GRENOBLE
- Focuses on “bag-of-tasks” jobs type
- Optimized for high throughput: millions of small computations, weakly parallel (or embarrassingly parallel)
- Manages well the “best-effort” mode of OAR (automatic resubmits)

CIMENT

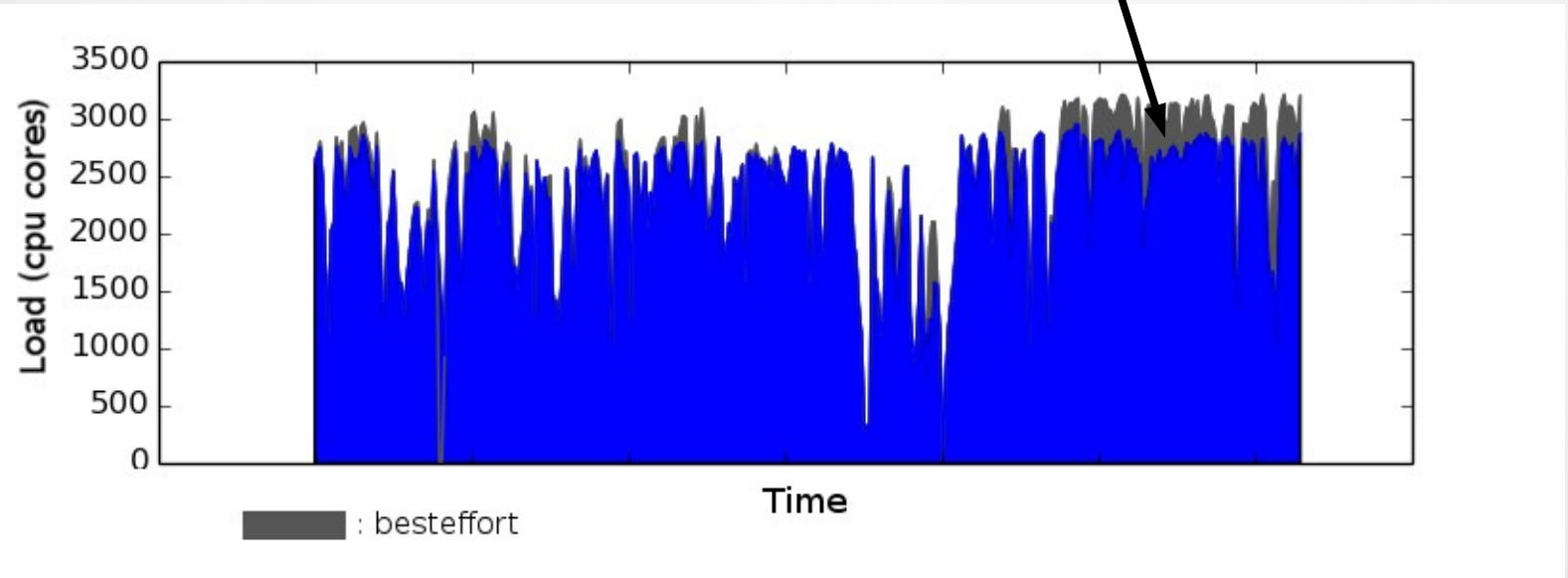
The CIGRI middleware



The CIGRI middleware

Cigri Jobs

Load sample of one HPC platform



CIMENT IRODS configuration

-
-
- Files distributed on sites by default

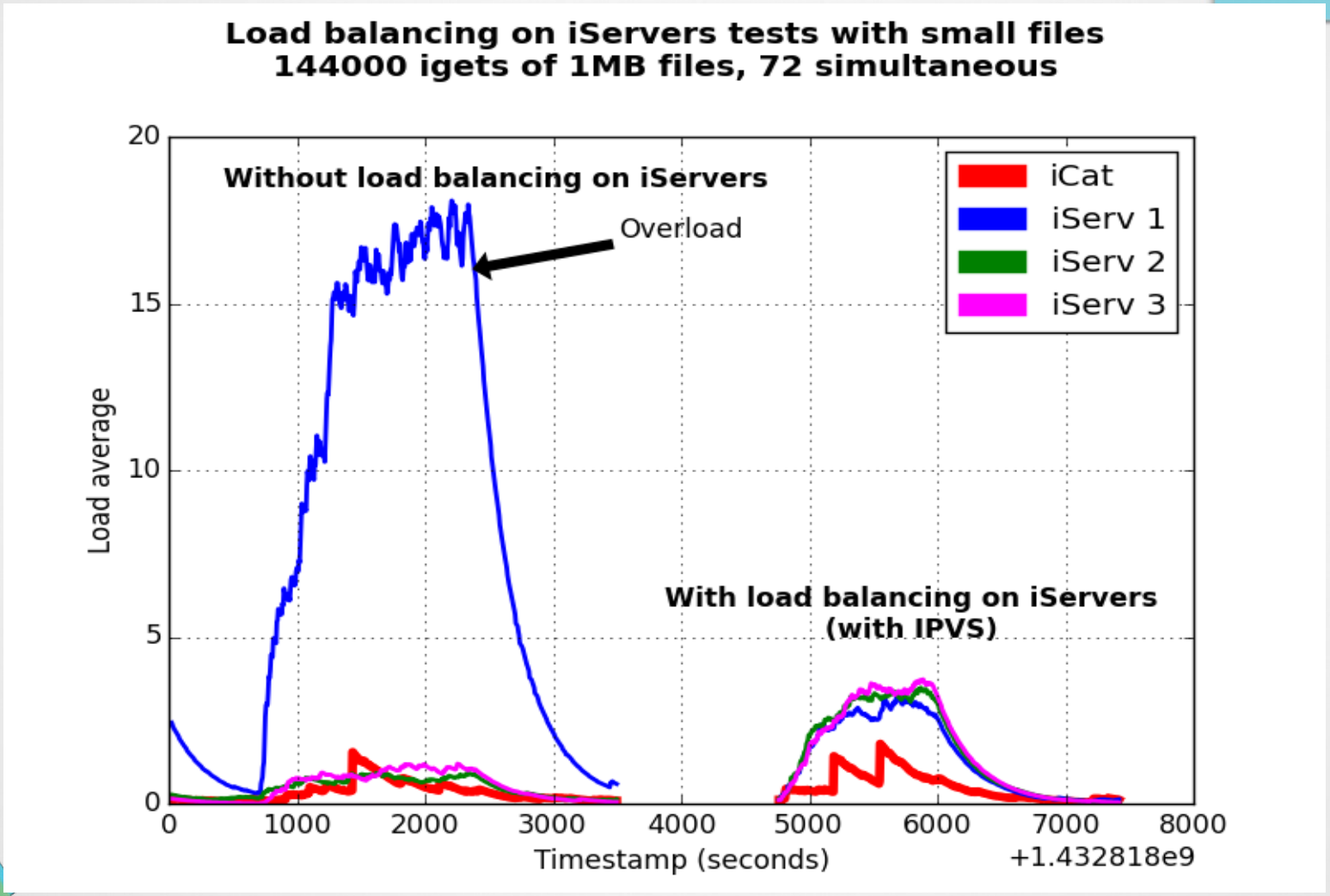
IRODS 3.3.1 under heavy load at CIMENT

- **Small files** (<32MB) problems: overload of the “proxy” iServ and bad performances on “iget -r” with very small files
- **Big files** problems: overload of resources
- **Current (not always satisfaying) solutions:**
  - connectControl.config → maxConnections
  - secure_i(get|put) wrapper → retry with incremental delay
  - Limiting the number of jobs
  - Ciget
- **Possible enhanced solutions:**
  - Load balancing on iServers
  - Irods resources included in the batch management (OAR)
  - Other irods side solutions? (queue management? rate limiting?...)

CIMENT

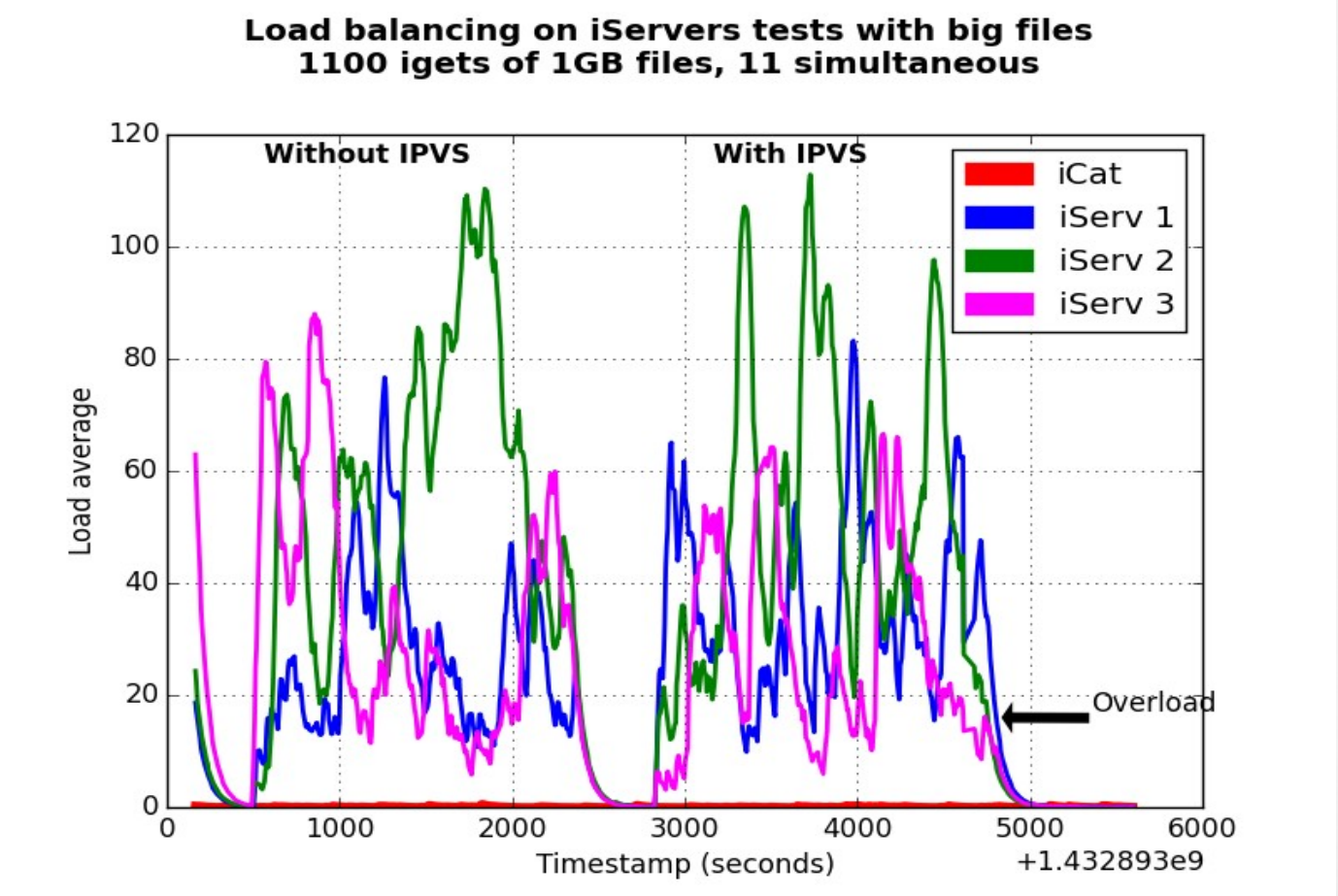
Load balancing : considered solutions

IPVS efficiency (1/2)



IPVS efficiency (2/2)

Or uselessness in this case actually...



… we simply illustrate the “big files overload of resources” problem

IPVS : conclusion

Cirods : python library to optimize the use of large set of small files and custom meta-data

CIMENT

Cirods : tests results

CIMENT

Problems or missing features

- Staging abandonned: overload of network link
- Errors on heavy load
- Retry feature does not solve overload problems
- Global rate/connexion limits missing

CIMENT

Future works

- IRODS 4
- Cirods → Python API with irods 4?

CIMENT

Example applications

Whisper

Rosetta

Ecology

CIMENT

Particle physics : « Exotics »

Conclusion