XRootD and Singularity setups for Grid integration of pledged and opportunistic resources on shared HPC systems

11th Annual Meeting Physics at the Terascale
DESY
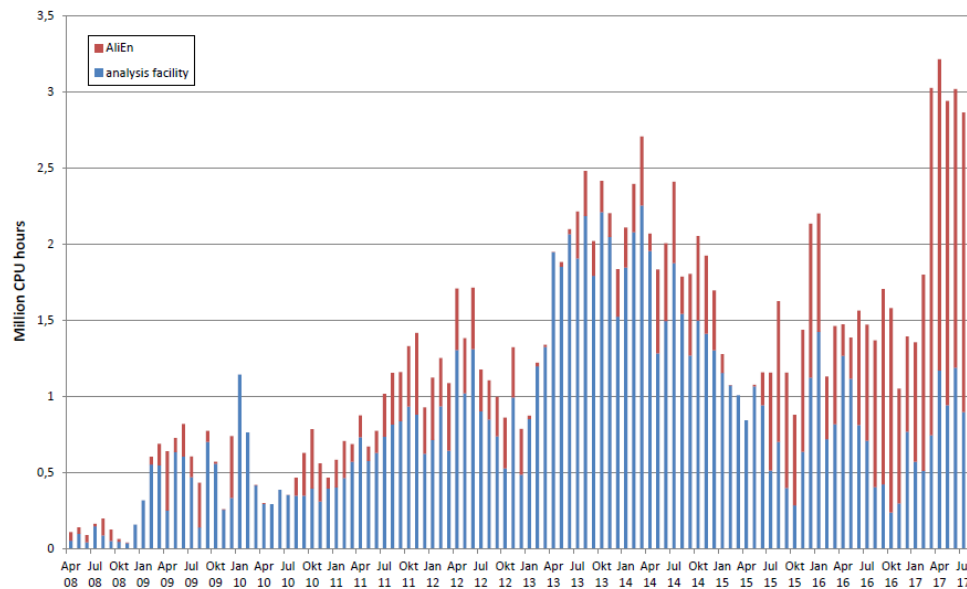### November 2017

Kilian Schwarz
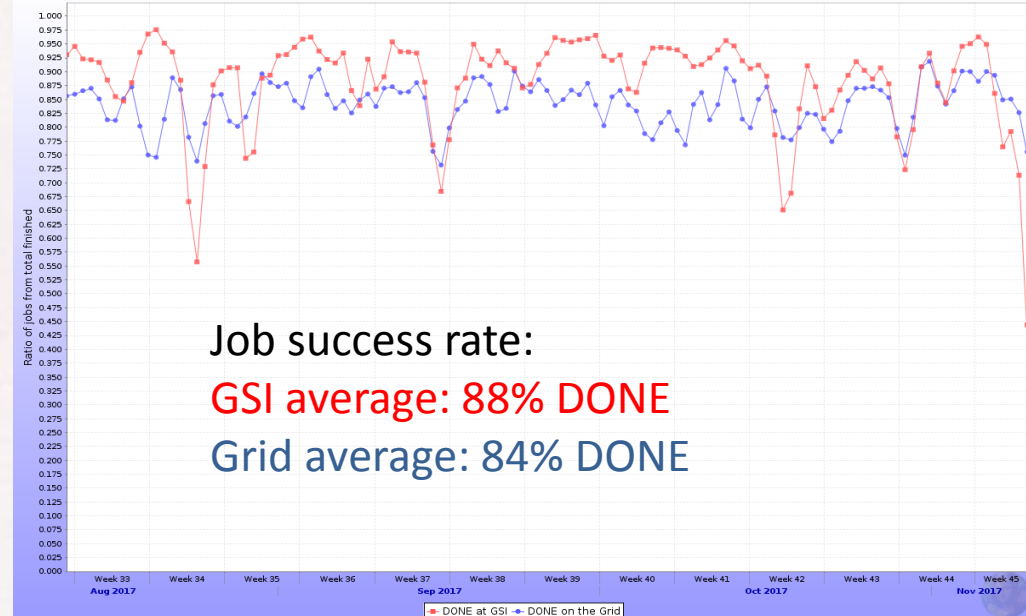
GSI Helmholtzzentrum für Schwerionenforschung GmbH

ALICE T2@GSI Darmstadt: 1/3 Grid, 2/3 NAF

ALICE @ GSI: CPU time

GSI vs Grid

Job success rate:
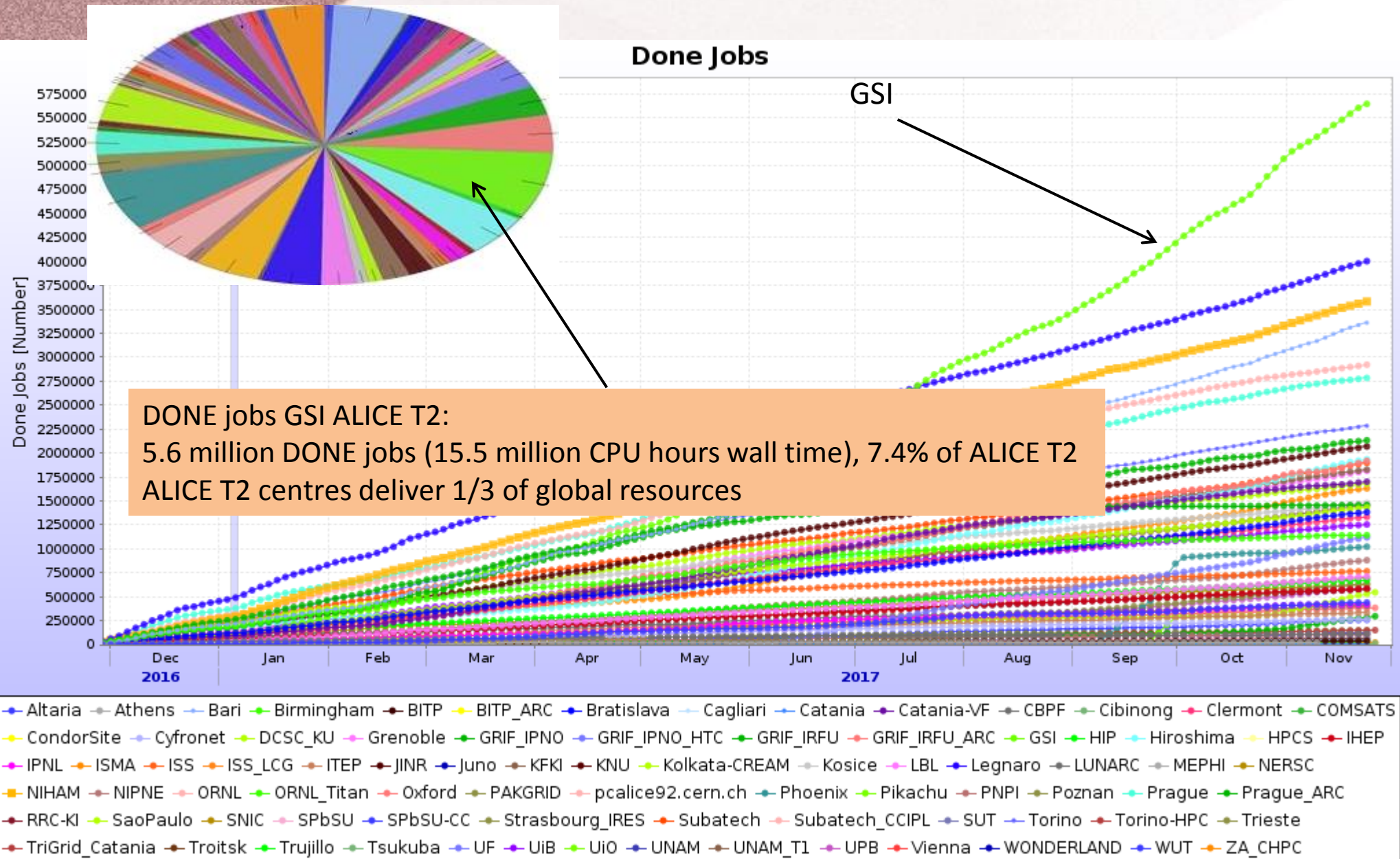GSI average: 88% DONE
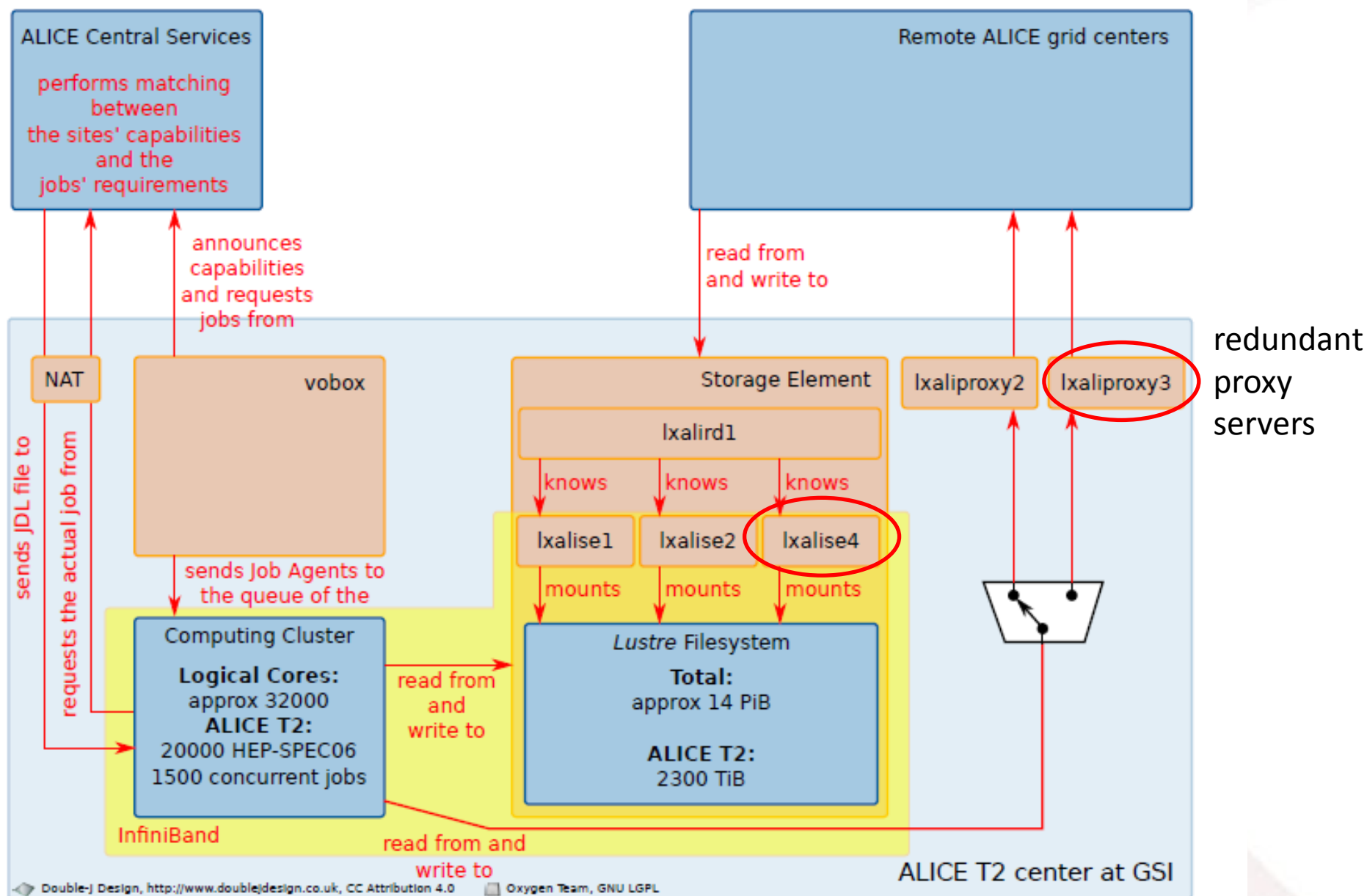Grid average: 84% DONE

Green IT Cube

- T2: 7% of ALICE T2 (2017:
  CPU: 20kHS06 (5%, pledged),
  26kHS06 (7%, delivered), disk: 2.3 PB)
  - about 1/3 Grid (red), 2/3 NAF (blue)
- on average: 2045 jobs
- GSI T2 Efficiency (incl. CPUtime /Wall time ratio) > 10% higher than Grid

Job contribution (last year)
GSI: largest Tier2 centre in ALICE

Done Jobs

GSI

DONE jobs GSI ALICE T2:
5.6 million DONE jobs (15.5 million CPU hours wall time), 7.4% of ALICE T2
ALICE T2 centres deliver 1/3 of global resources

ALICE T2 Centre at GSI

GSI Storage Contribution

AliEn name	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage
ALICE::GSI::SE2	2.3 PB	1.926 PB	382.5 TB	83.76%	46,501,158	FILE	2.3 PB	1.963 PB	344.5 TB	85.37%

- during the last year about 2.4 PB have been written, 8.5 PB read
  - significant increase compared to 2016
  - saturation of dedicated 10 Gbs LHCOne network link
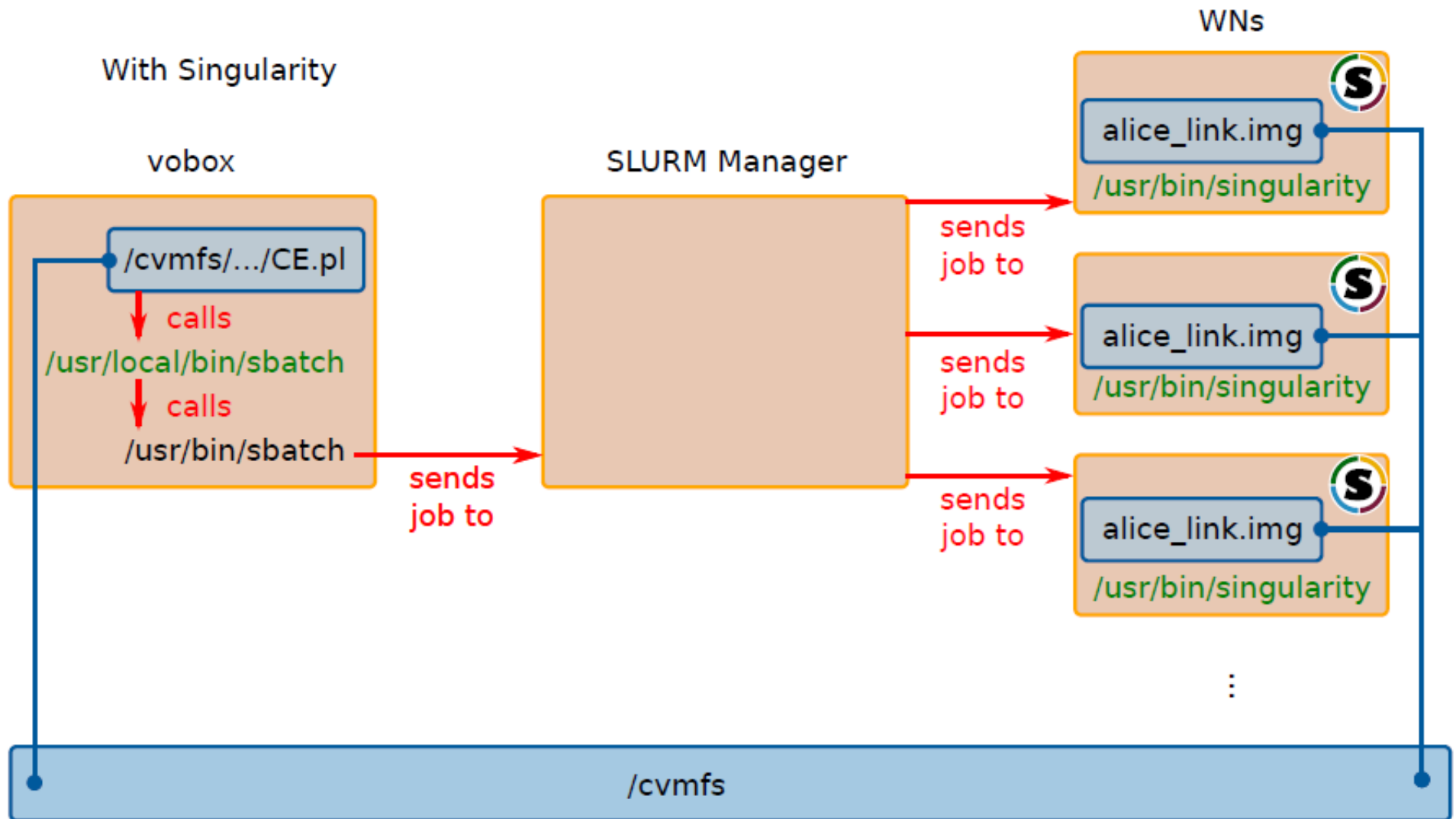

Network traffic on ALICE::GSI::SE2
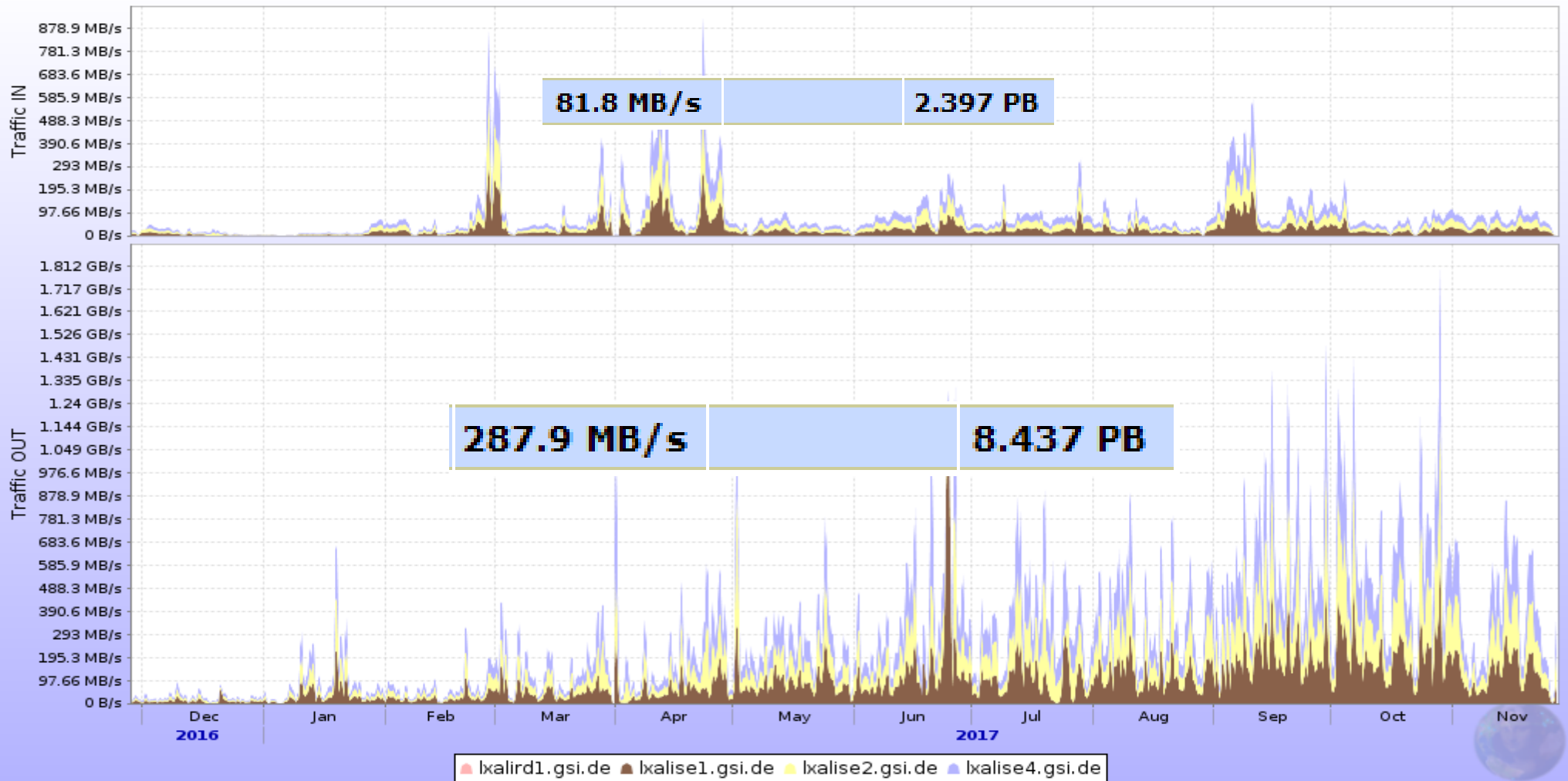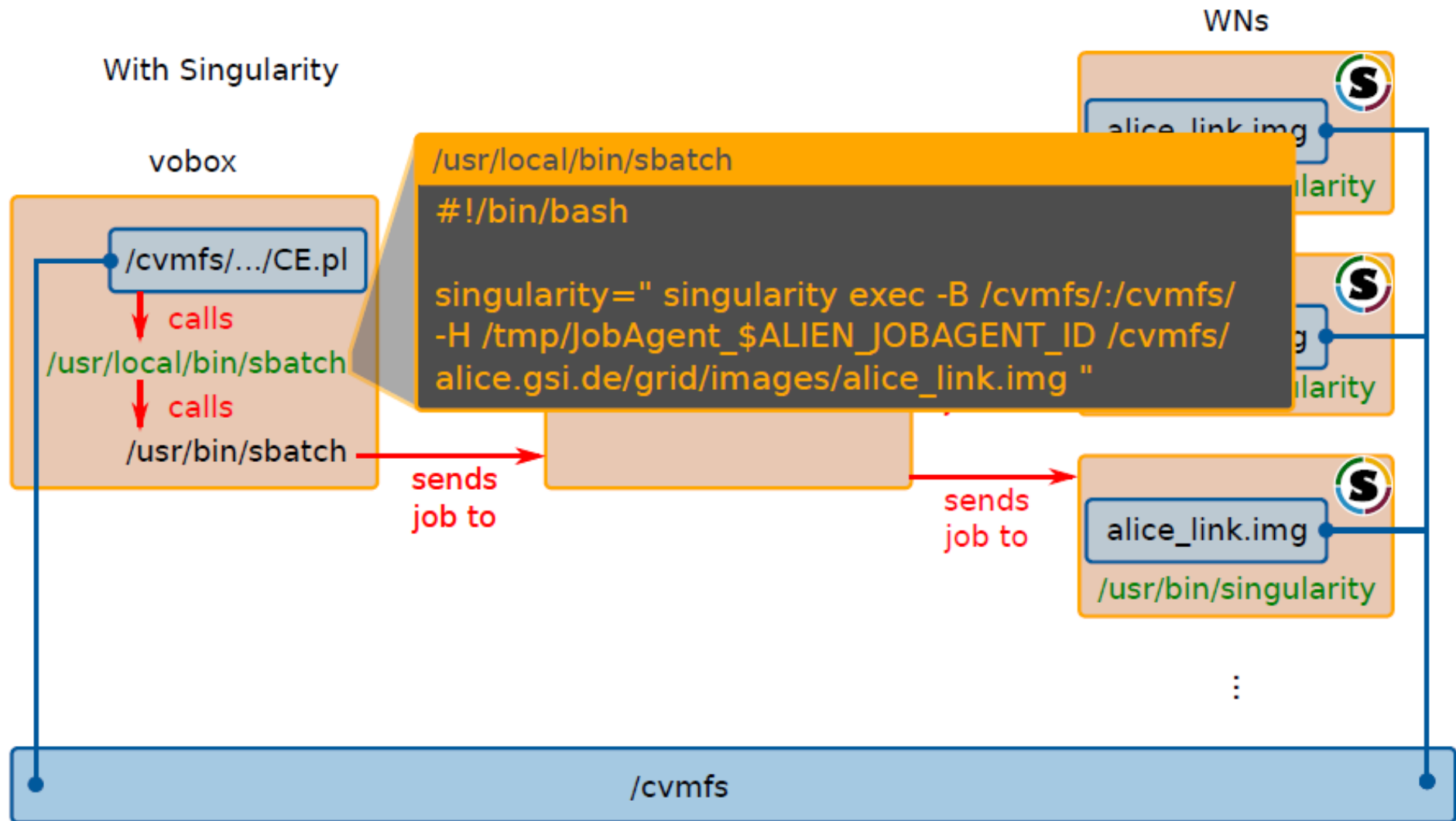
GSI ALICE T2 – *Singularity*

- Linux containers are self-contained execution environments
    - they use their own, isolated CPU, memory, block I/O, and network resources
    - they share the kernel of the host operating system.
    - result: a lightweight virtual machine
    - 6 to 8 times as many containers as VMs on same hardware possible
- Virtual Machines (VMs):
    - duplicate instances of the same OS and redundant boot volumes.

Source: https://www.infoworld.com

GSI ALICE T2 – Singularity

With Singularity

vobox

/cvmfs/.../CE.pl

calls

/usr/local/bin/sbatch

calls

/usr/bin/sbatch → sends job to

sends job to

WNs

alice_link.img

/usr/bin/singularity

alice_link.img

/usr/bin/singularity

/cvmfs

```
/usr/local/bin/sbatch

#!/bin/bash

singularity=" singularity exec -B /cvmfs/:/cvmfs/
-H /tmp/JobAgent_$ALIEN_JOBAGENT_ID /cvmfs/
alice.gsi.de/grid/images/alice_link.img "
```
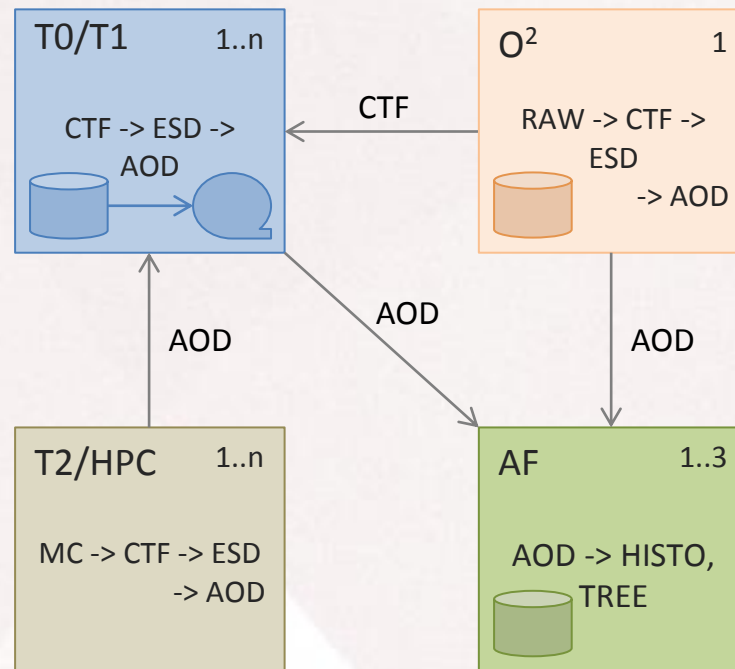
ALICE Analysis Facility Prototype

Something new!

ALICE Analysis Facility
project from Helmholtz
Data Federation (HDF)

Now: prototype based on
existing ALICE setup

Run 3 Computing Model
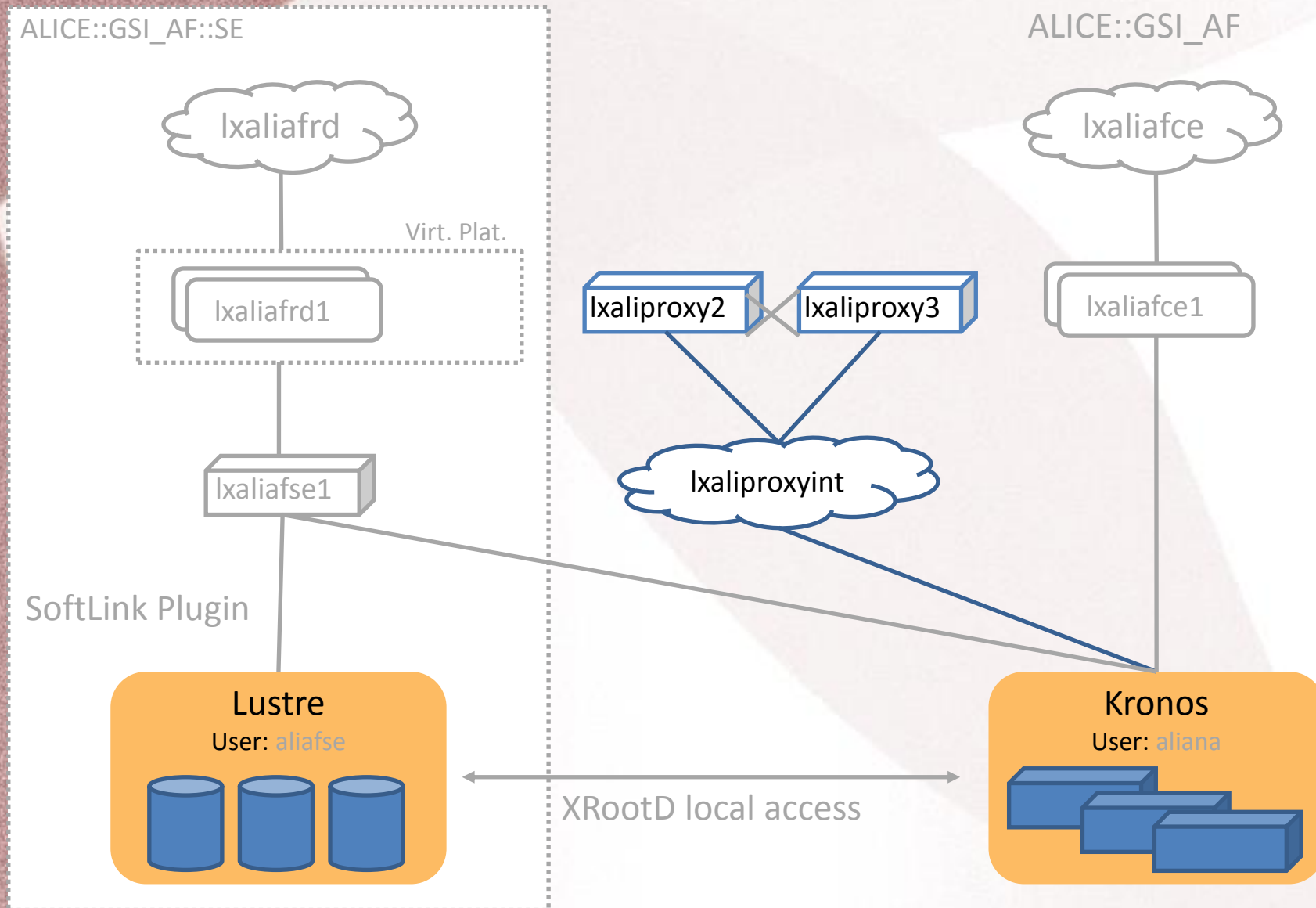

Grid Tiers mostly specialized for given role

- O2 facility (2/3 of reconstruction and calibration), T1s (1/3 of reconstruction and calibration, archiving to tape), T2s (simulation)

- All AODs will be collected on the specialized Analysis Facilities (AF) capable of processing ~1 PB of data within ½ day timescale

The goal is to minimize data movement and optimize processing efficiency

(P. Buncic, ALICE T1/T2 Workshop, Strasbourg, May 2017)

ALICE Analysis Facility
Prototype

Initial Resources
- 0.7 PB disk space
- up to 1000 job slots
  (taken from Tier 2 allocation)

Initial dataset
(TBC with local ALICE group)
- Full AOD set 2015 Pb-Pb

ALICE::GSI_AF::SE

ALICE::GSI_AF

lxaliafrd

lxaliafce

Virt. Plat.

lxaliafrd1

lxaliproxy2   lxaliproxy3

lxaliafce1

lxaliproxyint

lxaliafse1

SoftLink Plugin

Lustre
User: aliafse

Kronos
User: aliana

XRootD local access

GSI ALICE T2 – XRootD Plugins

XRootD Client Plug-in – XrdOpenLocal:
Clients should open a file directly from Lustre
if at GSI
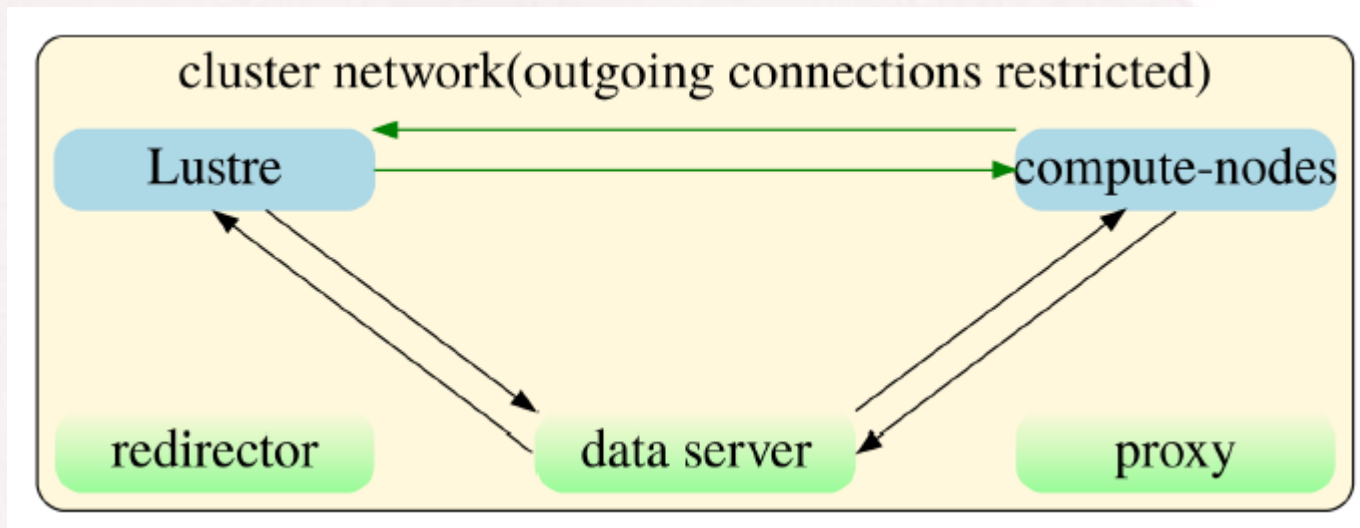
Available as Client and **Server (Redirector) Plugin.**
Clients will still need a new XRootD Client, though
Needed Client code in XRootD base starting with version 4.8
(see Bachelor thesis Paul-Niklas Kramp)

XRootD Disk Cache on the Fly

Based on existing ingredients: Disk Cache on the Fly for Opportunistic Resources integrated into production systems

local Disk Cache

Grid Jobs on the local HPC Cluster
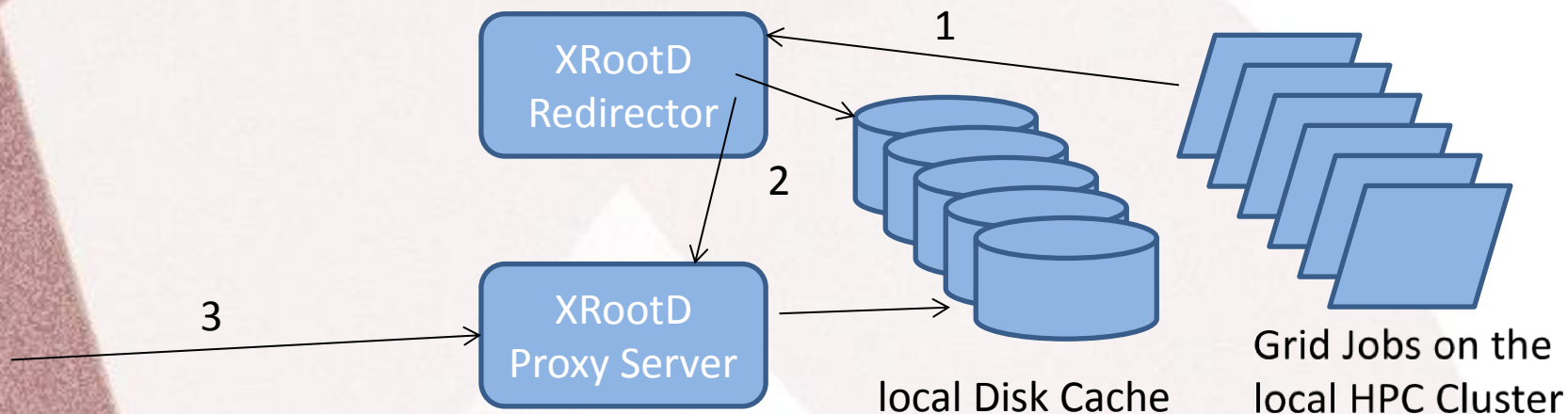
1. Xrd Client Plug-In: Redirect to local Redirector
2. Xrd Server (Redirector) Plug-In: Redirect to local or Redirect to Proxy
3. Xrd Forward Proxy: read from remote and write on local disk cache

New Scientific Journal:
Computing and Software for Big Science

Target Group:
Scientists working in the border region
between Physics and Computer Science

URL: http://www.springer.com/physics/particle+and+nuclear+physics/journal/41781