AAA: CMS federated data access, operations and development

Engin Eren, DESY
on behalf of the CMS experiment

Physics at the Terascale

Introduction to AAA

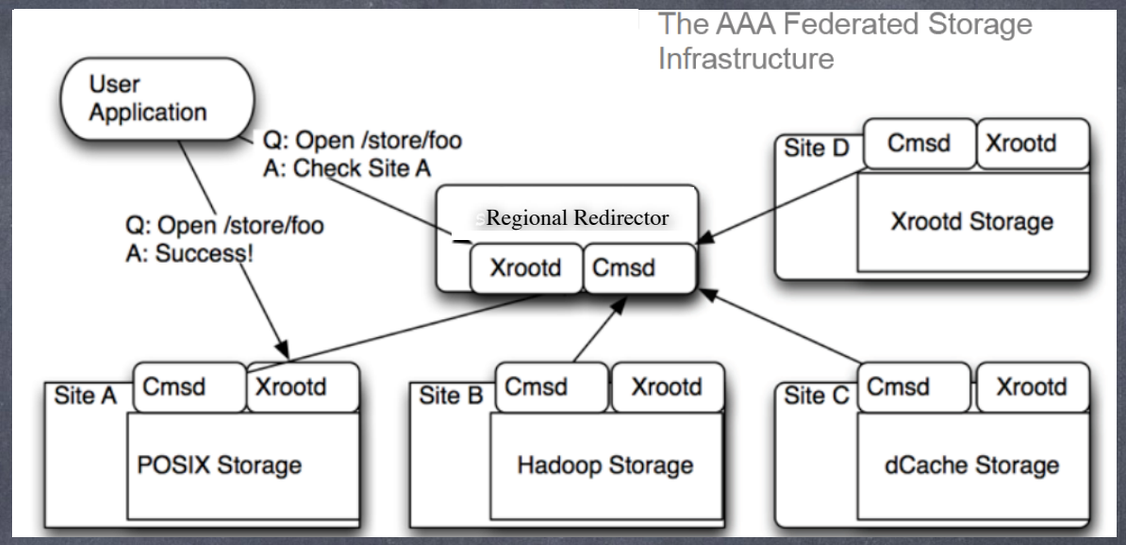
- AAA project focuses on development and deployment of storage sources, which can be accessible for <u>any data, anywhere at anytime.</u>

- Why AAA ?
  - ✳ CMS sends “jobs to where the data is”. (Design decision, 10 years ago.)
  - ✳ Some drawbacks :
    - Jobs at a site may fail for data access (accidents etc.)
    - If a small number of sites host the “needed (popular)” data, queue waiting time might be long.
    - Data transfer may be necessary before submitting jobs.

- The ultimate goals and motivations of AAA project :
  - ✳ A user does not need to know where the data actually is.
  - ✳ No access failures regardless of data location.
  - ✳ No burden for physicists who want to access **remote** data.

AAA Structure and Federation

- XrootD infrastructure spans all of the T1 and T2 sites (+ some of T3s currently).
- Each XrootD server of a site is interfaced with local storage system to export CMS namespace. Current federation supports the sites with dCache, DPM, Hadoop, StoRm, Lustre and Castor.
- Site storage servers should subscribe to a **site redirector.**



✓ Site redirectors from each site should subscribe to **regional** redirector:

- US : hosted by FNAL and Nebraska
- EU : hosted by Bari, Pisa and Paris

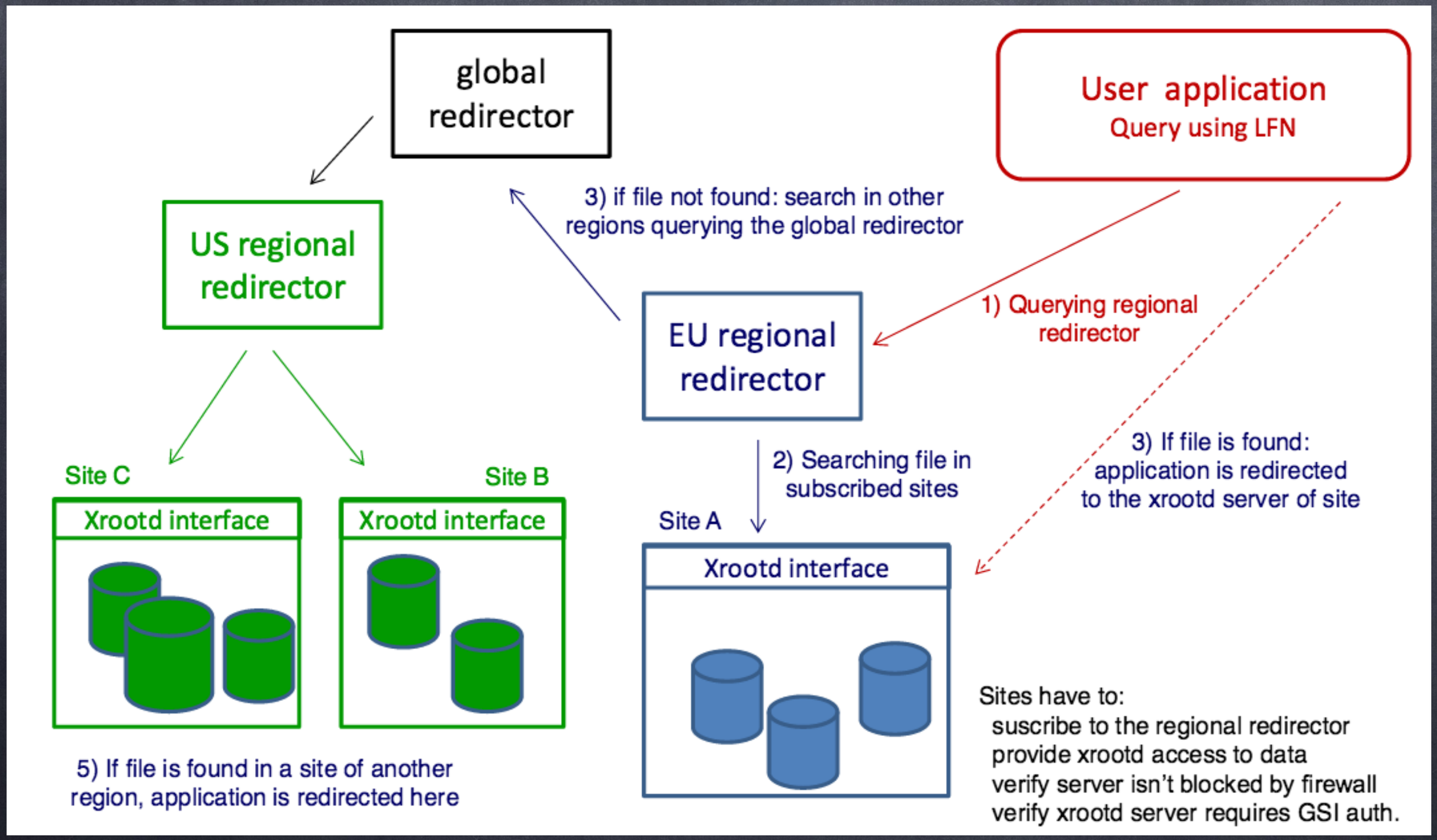
- Tree-structured federated storage system :

✓ A user can request a file from regional redirector.

✓ Regional redirector queries the site redirectors in the tree and redirects the user to the server where the data is.

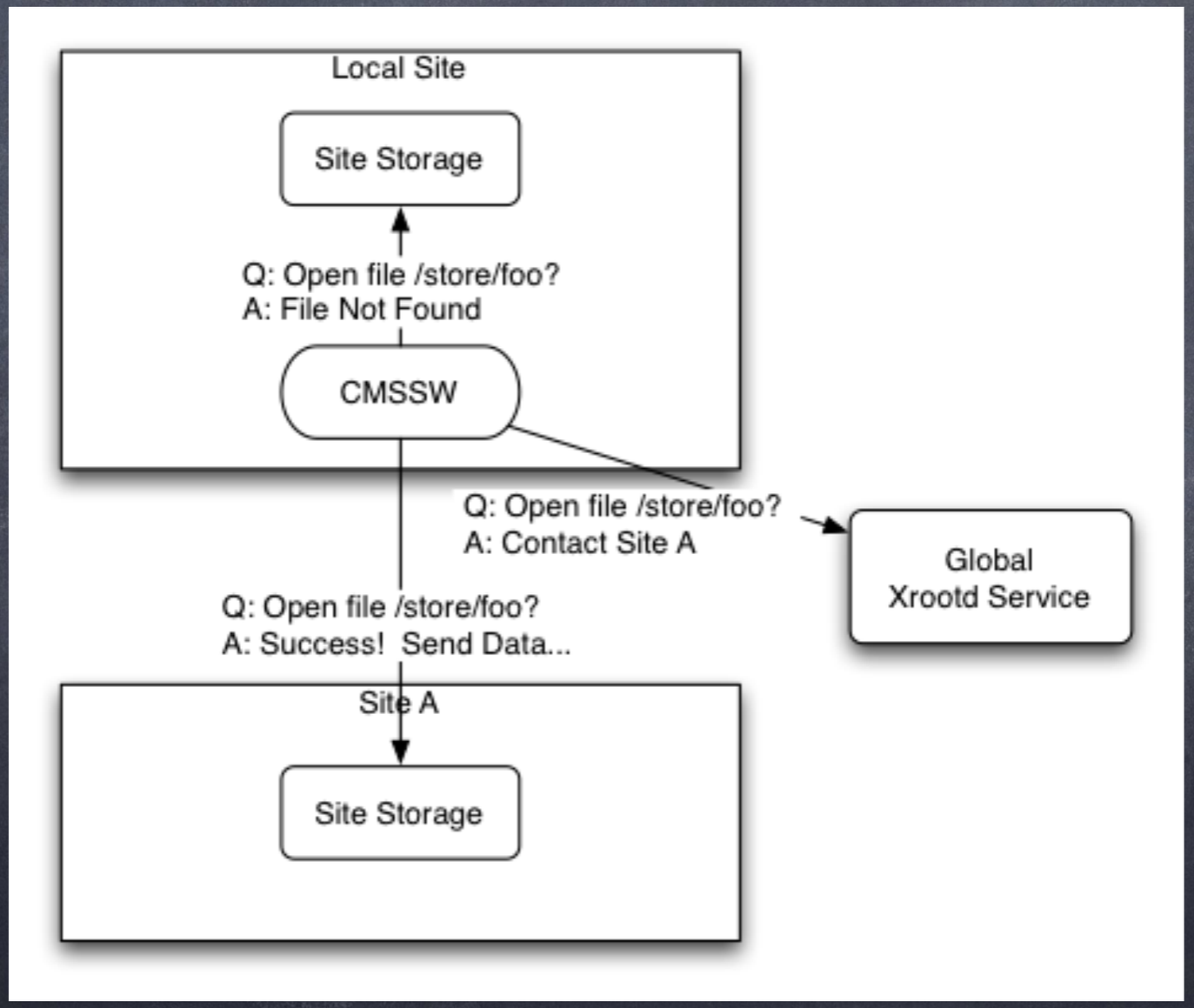
✓ This allows users to access any on-disk data <u>without knowing its location.</u>

AAA Redirectors Schema



Fallback Access

- Fallback access allows CMSSW to retry opening a file from different location if it cannot be opened by the primary source.
- This usually happens due to :
  - ✓ Accident : the local site's copy is corrupted or deleted.
  - ✓ On purpose : the jobs scheduler is overflowing jobs to another site.



Monitoring

- SAM Access Test
- HammerCloud Tests

SAM Access Tests For XrootD Federation

- Site Availability Monitor (SAM) is an automatic system which allows to execute tests on services located at sites.

Legends for Metric Result Status

Status:	OK	WARNING	CRITICAL	UNKNOWN	MAINTENANCE
Legend:	OK	W	C	U	MT

Note: darkest colors: test is 0 - 12 hours old, ... lightest colors: test is more that 12 hours old

Legend	Metric Name
1	org.cms.glexec.WN-gLExec (/cms/Role_pilot)
2	org.cms.WN-squid (/cms/Role_lcgadmin)
3	org.cms.SRM-GetPFNFromTFC (/cms/Role_production)
4	org.cms.WN-mc (/cms/Role_production)
5	org.cms.WN-basic (/cms/Role_lcgadmin)
6	org.cms.WN-xrootd-access (/cms/Role_lcgadmin)
7	org.cms.WN-swinst (/cms/Role_lcgadmin)
8	org.cms.WN-frontier (/cms/Role_lcgadmin)
9	org.cms.WN-env (/cms/Role_lcgadmin)
10	org.cms.WN-xrootd-fallback (/cms/Role_lcgadmin)
11	org.cms.WN-analysis (/cms/Role_lcgadmin)
12	org.cms.SRM-VOPut (/cms/Role_production)
13	org.cms.SRM-VOGet (/cms/Role_production)
14	org.sam.CONDOR-JobSubmit (/cms/Role_lcgadmin)
15	org.sam.CONDOR-JobSubmit (/cms/Role_production)

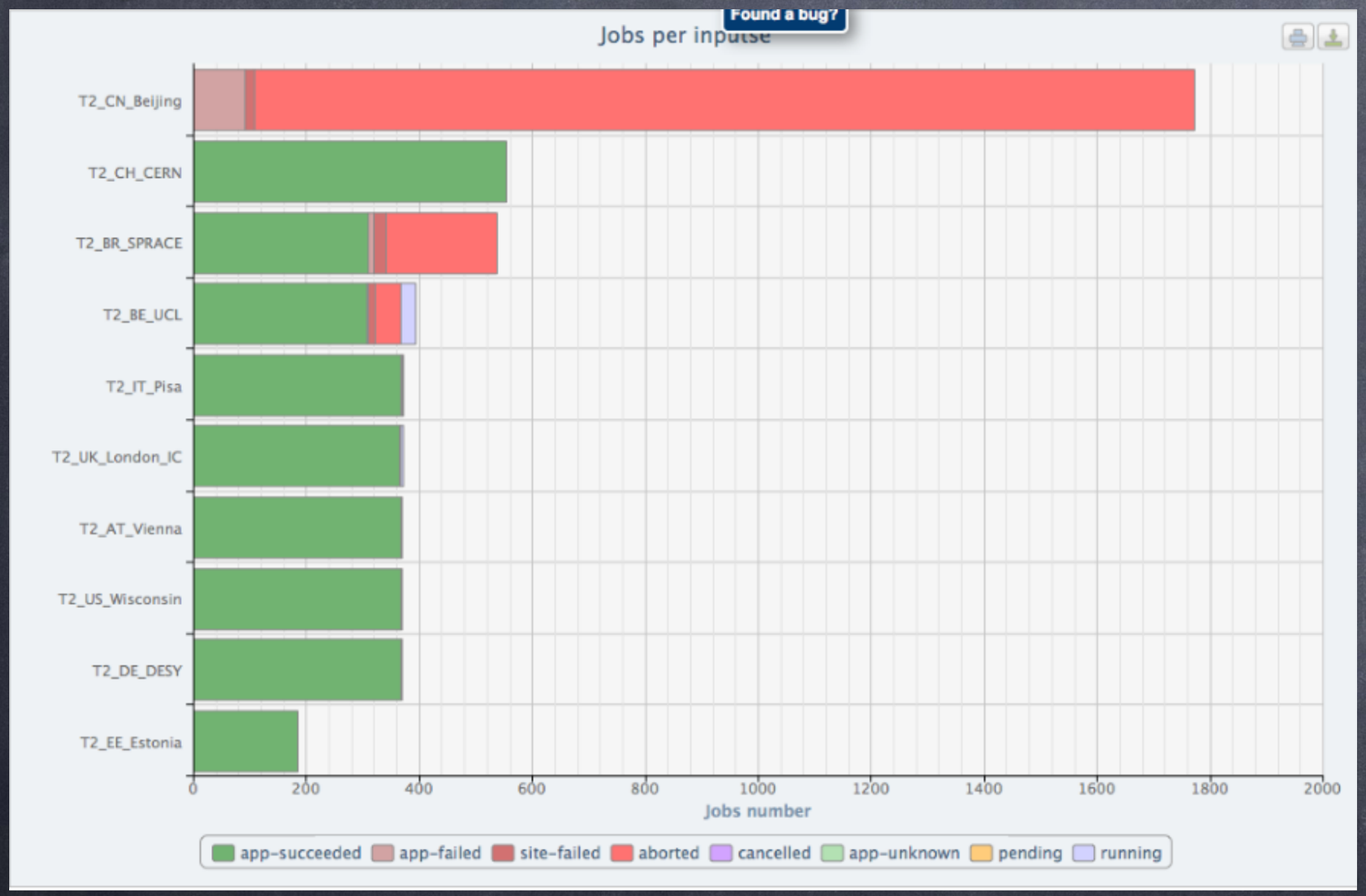
Link to data

Sitename	Flavour	Host status in profile	Hosts	3	12	13									
T2_DE_DESY	SRMv2	OK	dcache-se-cms.desy.de	OK	OK	OK									
				1	2	4	5	6	7	8	9	10	11	14	15
	CREAM-CE	WARNING	grid-cr0.desy.de	W	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
		WARNING	grid-cr1.desy.de	W	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
		WARNING	grid-cr2.desy.de	W	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
		WARNING	grid-cr3.desy.de	W	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
				1	2	4	5	6	7	8	9	10	11	14	15
	ARC-CE	WARNING	grid-arc0.desy.de	W	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

- ✓ SAM XrootD Fallback Test : checks whether fallback configuration of sites are working properly.
- ✓ SAM XrootD Access Test : determines whether the site is exporting files via the redirector. A short cmsRun job tries to access a standard file.

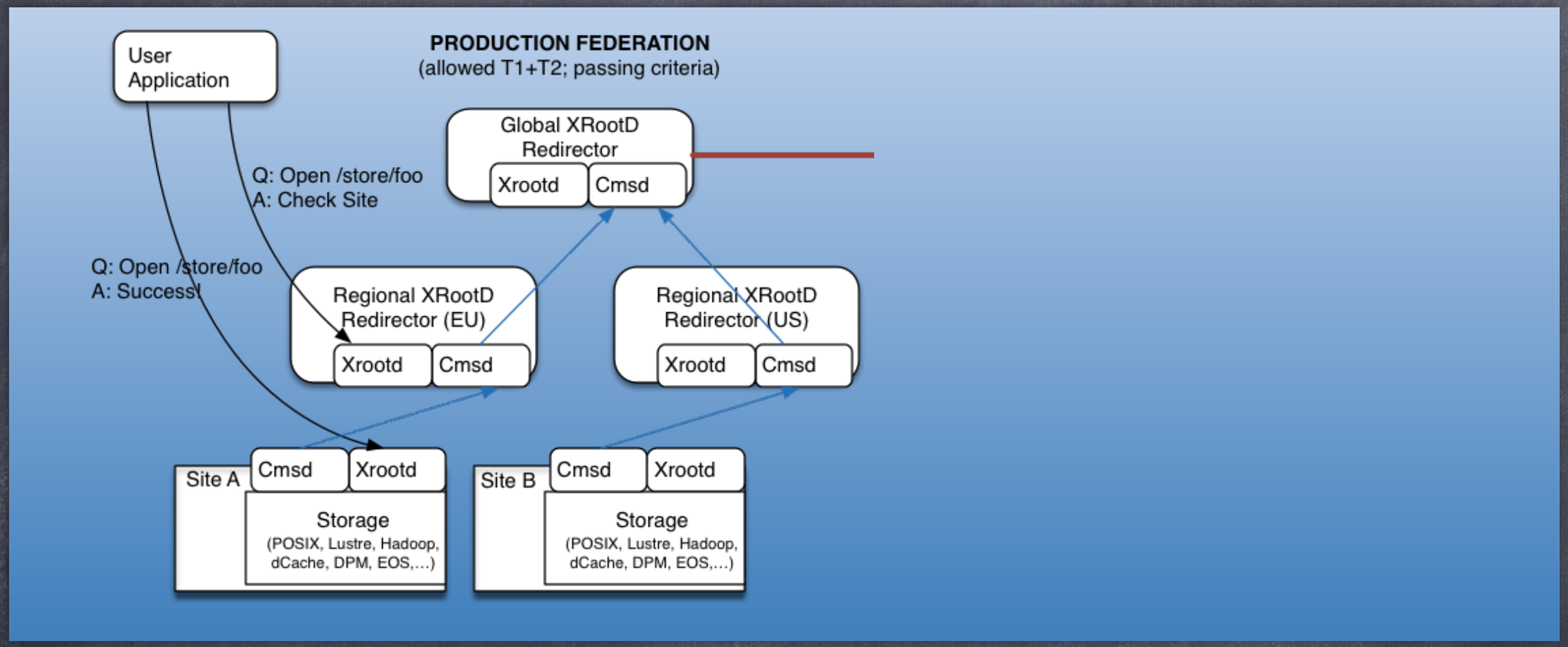
Hammer Cloud Tests

- HammerCloud (HC) is a distributed-analysis testing system for WLCG sites. HC tests your system and reports the results.

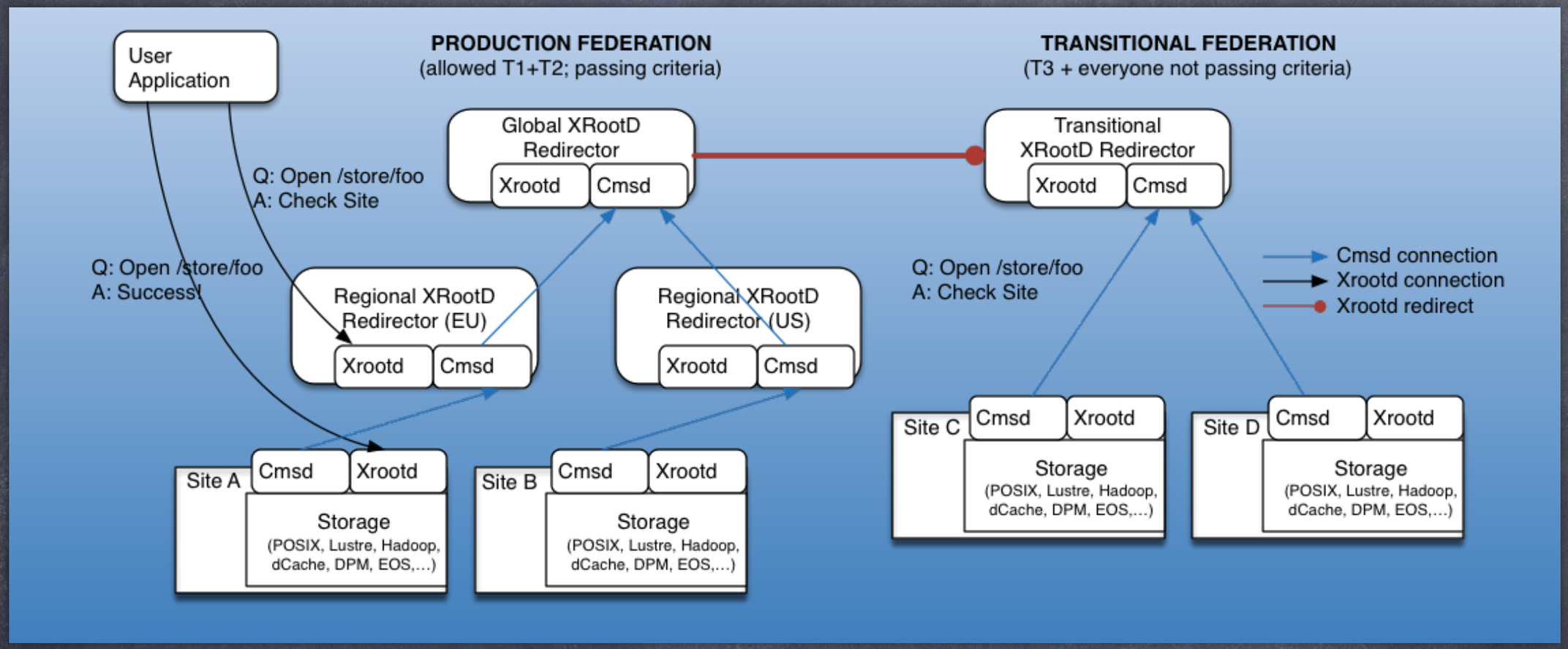


- Succes Rate Evaluation :
  - ✓ %90 is OK for T1s
  - ✓ %80 is OK for T2s
  - ✓ Below these, it is BAD.

AAA-Transitional Federation

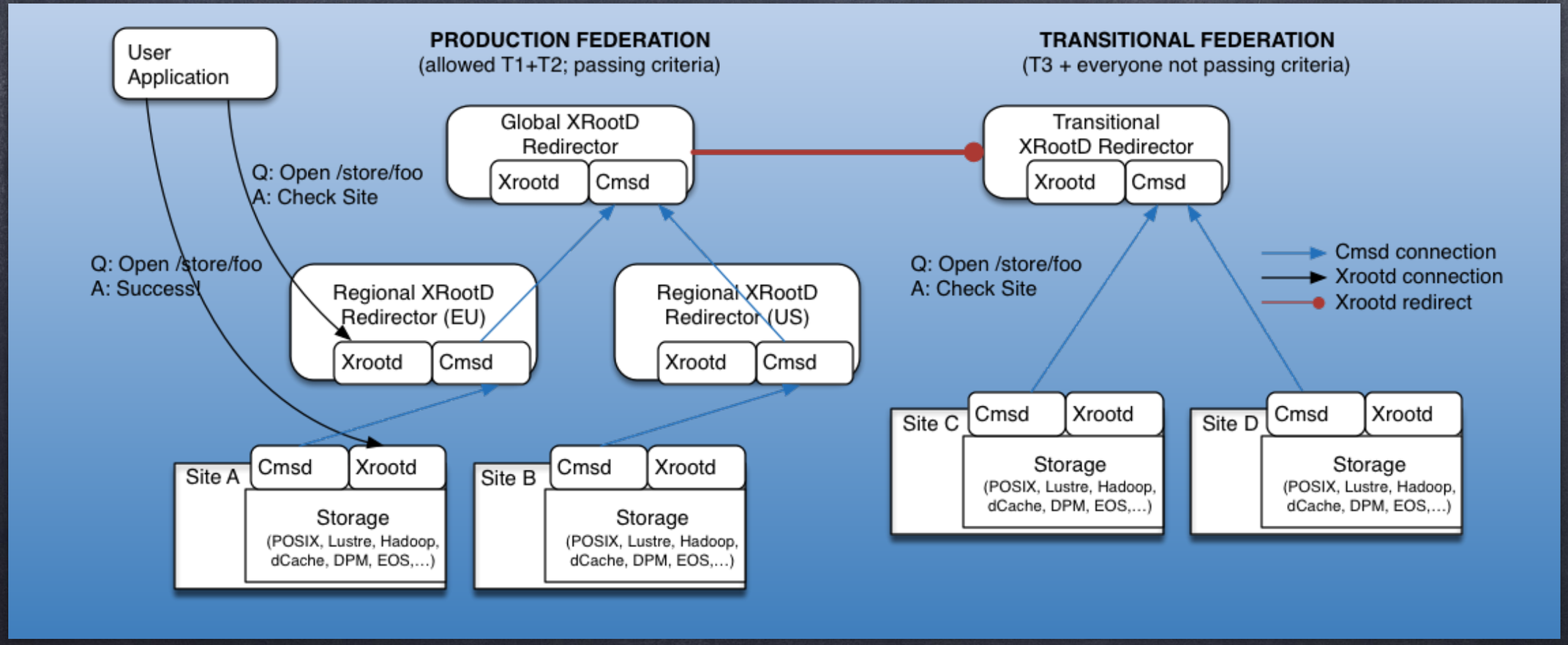


AAA-Transitional Federation



AAA-Transitional Federation

- Need for Transitional Federation : Separation between “well performing” and “weakly performing” sites.
- Transitional Federation is an isolated AAA data federation from the main (production) one. It consists of sites which are failing the criteria for to be in production federation :
  - ✓ SAM access test results > 50% for past two weeks.
  - ✓ HC Success rate > 80% for past two weeks.
  - ✓ Do not have AAA-WAN Access GGUS ticket open for longer than two weeks.
- When production federation site ends up in transitional federation, one-way access to the data (which are not anywhere else in production fed.) are allowed.

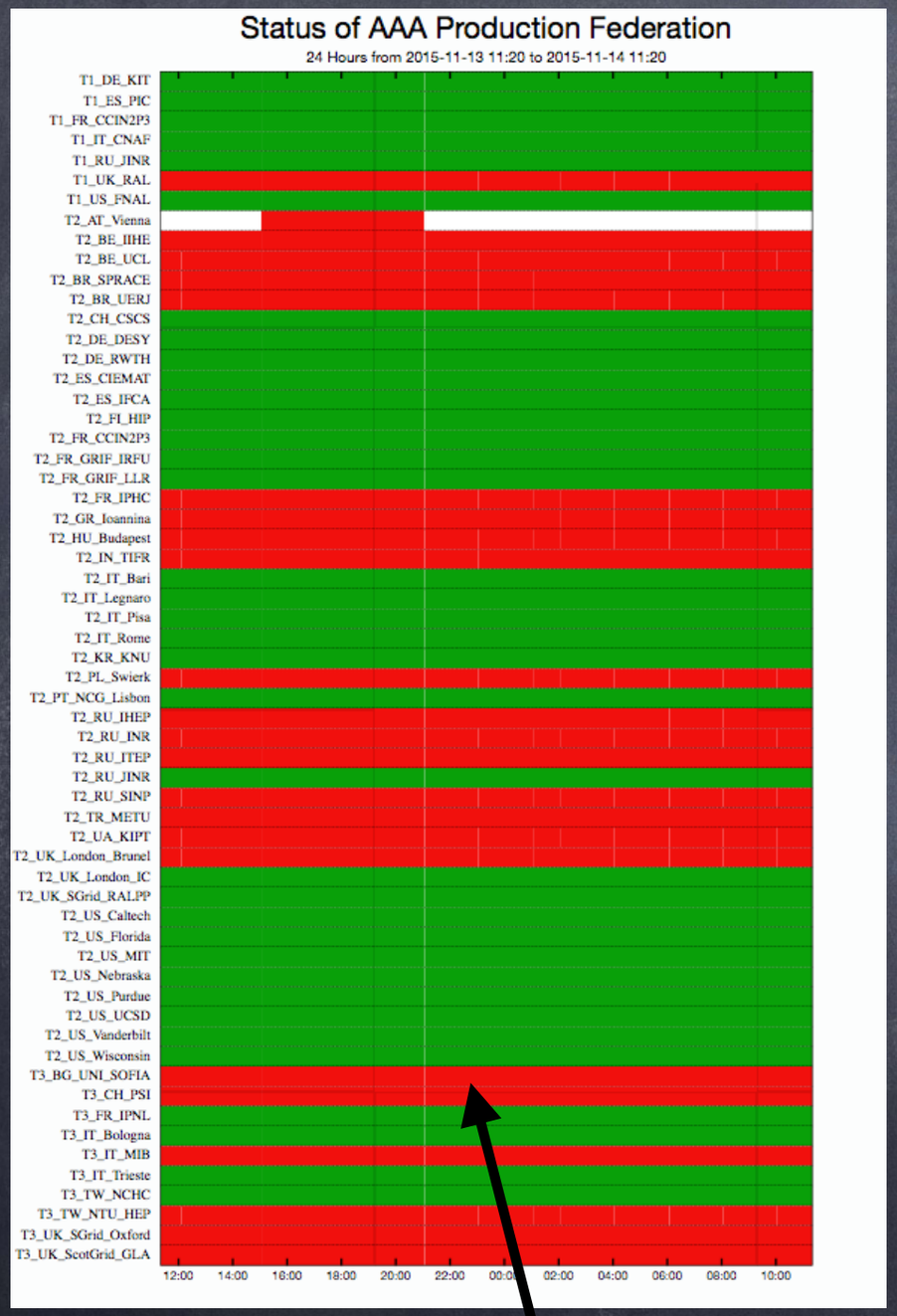


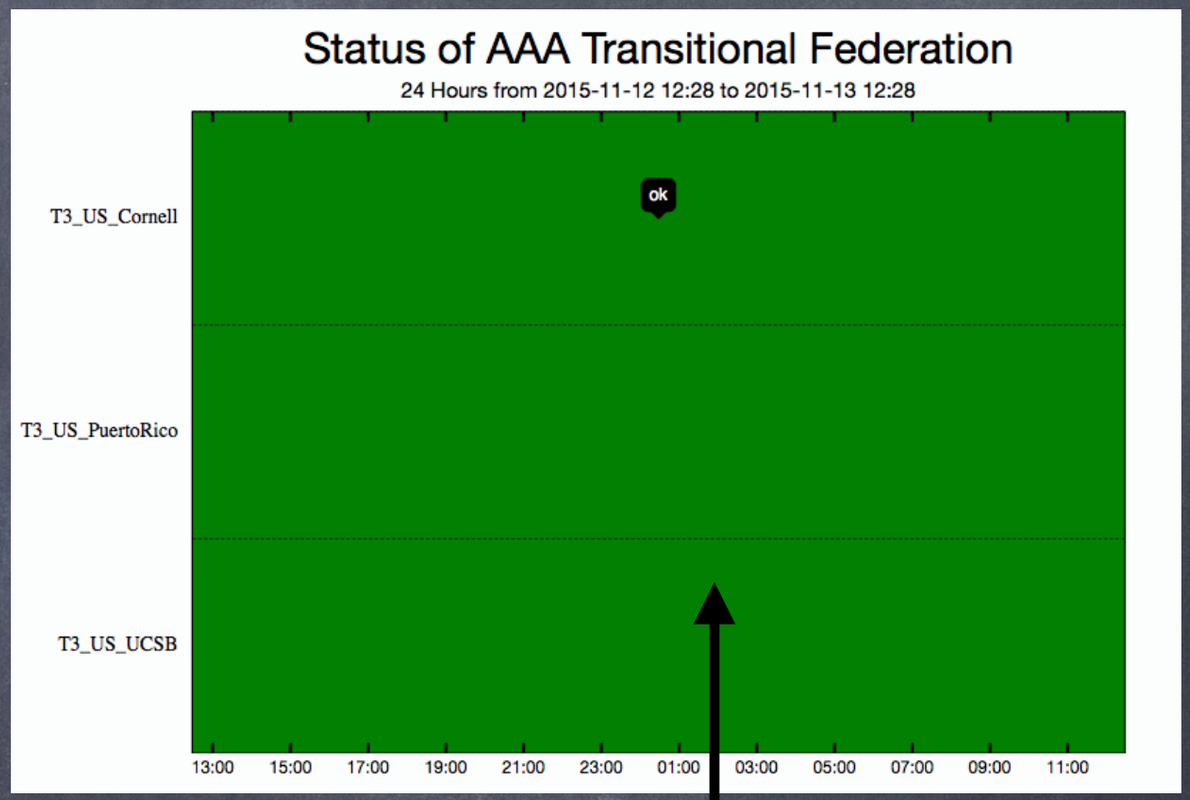
How to manage sites in federation ?

- **Production Federation** :
  - ✓ Make list of allowed sites join the federation (access list) and distribute across regional level of redirectors.
- **Transitional Federation :**
  - ✓ No access list needed, let anyone join (by default help T3 sites)
  - ✓ Use as temporary room for the new sites joining (later) production until they pass the criteria.
  - ✓ Use as temporary (waiting) room for the sites which were disallowed in production federation based on bad results of the tests and other metric defined from SAM and HC tests.

Metric for Transitional & Production Federation in Site Status Board (SSB)

- Based on same criteria, there is an implementation on SSB : production and transitional federation.





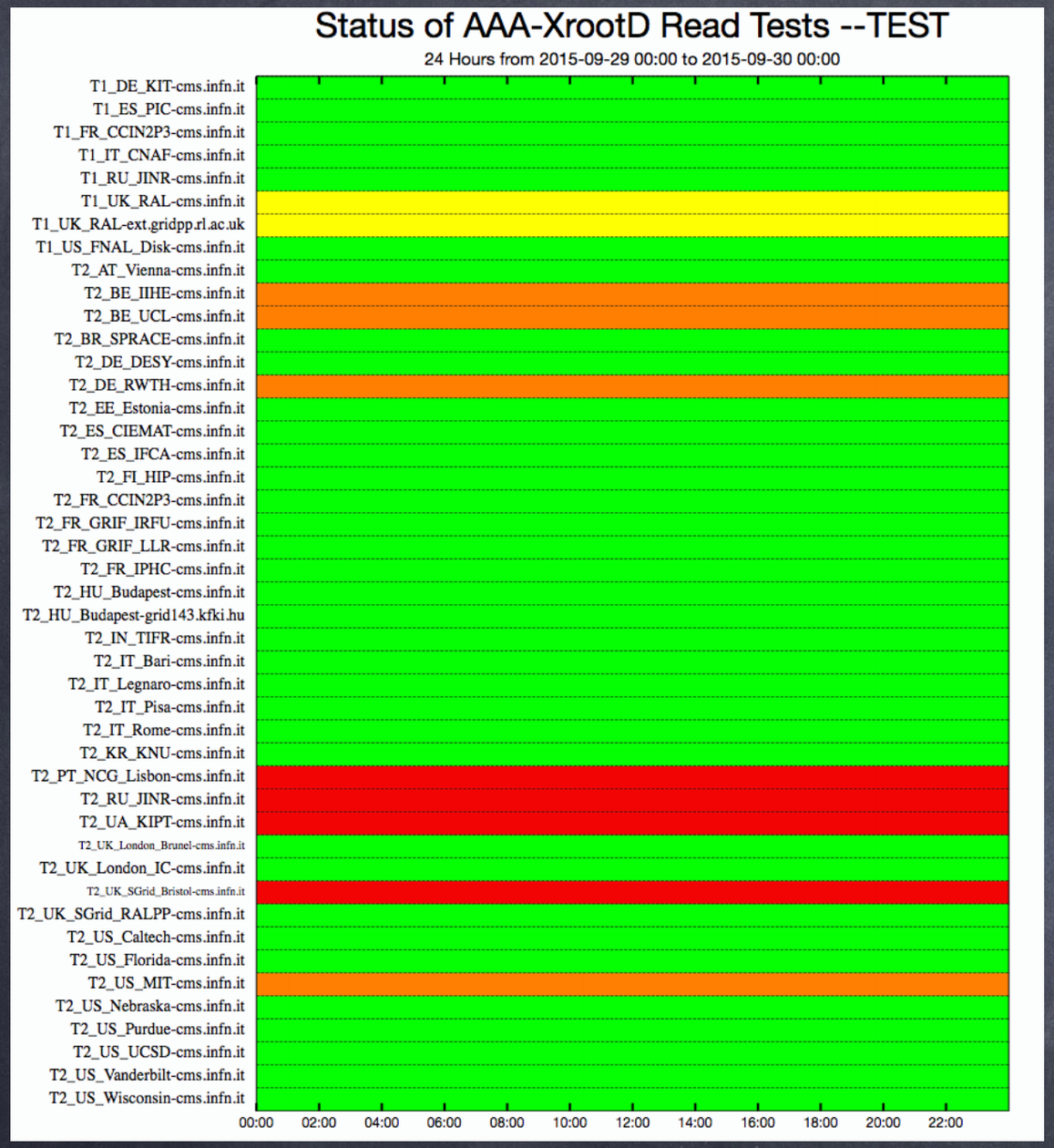
- “Red” sites are possible candidate for transitional federation.

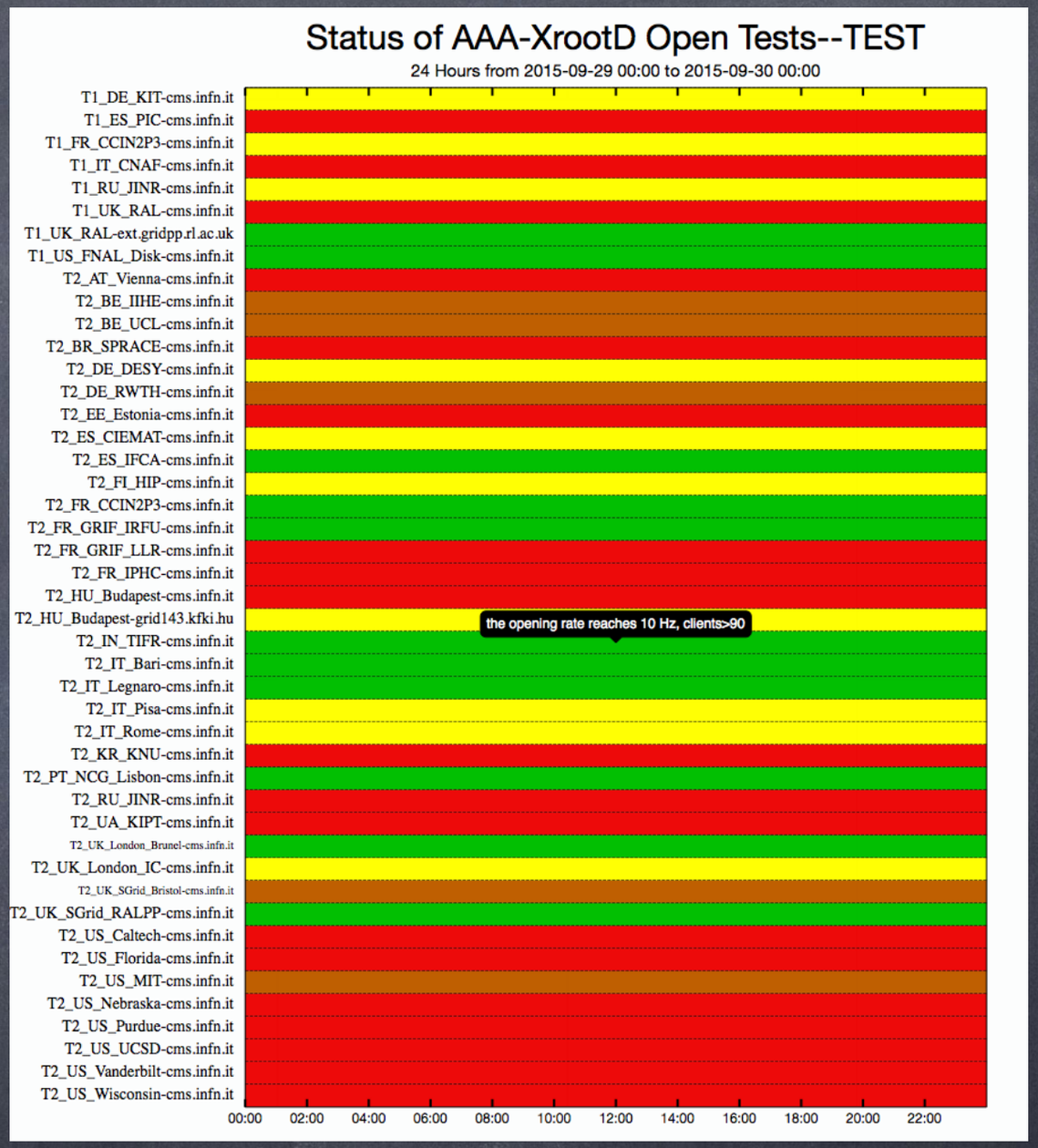
- “Green” sites are possible candidate to move back to production federation. (Except T3s!).

Scale Tests

- AAA infrastructure generates a network traffic and additional load on storage systems.
- It is needed to validate that the AAA component of the infrastructure is capable of handling this traffic.
- Scale tests measure **the rates** that files at a CMS site can be **opened** and **read**.
- Scale tests run every week.
- Minimum performance expectations:
  - ✓ An average job opens a new file once per 1000 seconds and reads from a file at an average rate 0.25MB/s.
  - ✓ Assuming a worst case of 100,000 jobs opening files at a site at once gives a benchmark of 100Hz for file-opening rates.
  - ✓ For file reading, an assumption of 600 simultaneous jobs actively reading files from a site given a total rate of 150MB/s.

Scale Tests : Metrics @ SSB





- the reading rate is lower than 150 MB/s, clients < 600
- the reading rate is lower than 150 MB/s, clients = 600
- the reading rate reaches 150 MB/s
- completely failed test, no plots produced

- the opening rate is lower than 10 Hz
- the number of simultaneous clients is lower than 90
- the opening rate reaches 10 Hz, clients>90
- completely failed test, no plots produced

Summary

- AAA is in production.
  - ★ All T1s and (almost all) T2s are in the federation.
- Monitoring is in place.
  - ★ Operations are being trained to use it.
- Transitional federation is in its integration phase.
  - ★ Moving “weakly performing” sites will start soon.