

dCache – protocol developments and plans

Paul Millar

(on behalf of the dCache team)

2015-01-13 pre-GDB: on Data Management & Data Preservation



Why am I here?

- We, WLCG, are re-evaluating the “protocol zoo”
 - From a dCache point-of-view:
 - dCache provides significant storage **for WLCG**.
 - dCache provides significant storage capacity **for many other communities**.
 - The dCache team needs to be efficient – not waste resources.
-

as a consequence, the dCache mantra

Use standards.

What should we do with SRM?

- There has been **considerable** investment in SRM,
 - SRM has some **unique features**:
 - Transfer protocol negotiation,
 - Asynchronous operations,
 - Bulk operations,
 - Transactions / 2-stage-commit upload.
 - (Space reservations, Access-Latency, ...)
 - Battle tested, with over 10 years of production use
 - it works.
-

SRM unique features

If WLCG continues to makes use of SRM unique features then it must either:

- Continue using SRM for those unique features
- Invent a new protocol to support unique features

(NB. a protocol extension == a new protocol)

Inventing a new protocol is **bad**: takes effort away from all software teams

WLCG has a finite effort → implementing a new protocol means all WLCG storage will suffer.

Recommendation: if WLCG uses SRM-unique features then stay with SRM for those features

We (dCache) are continuing to improve SRM, both server and client.

Non-unique SRM features

- **Storage accounting** (WebDAV+RFC-4331)

Not equivalent, but if space-reservations tied to namespace, it may be used instead. Along with DPM, we plan to implement support for this.

- **Direct data transfers** (not actually an SRM feature)

- Suggest using WebDAV for WAN and NFS for LAN (like GPFS+StoRM).
- We plan to drop dcap support once NFS is proven to work in production.

- **3rd party transfers** (FTP, WebDAV+extension, xrootd+extension, NFSv4.2)

Solution space is quite complex. Suggest using FTP for now and watch how WebDAV develops. N.B. SRM two-stage commit can prevent dark data.

- **Namespace operations** (FTP, WebDAV, xrootd, NFS)

Recommend bulk operations use SRM; for a “small” number of operations, any protocol is fine – suggest WebDAV or NFS if available.

WebDAV

- We have added **3rd-party transfer** support
 - Supports current specification.
 - Working with FTS/DPM developers to verify this.
 - Some limitations of the current approach, some development may be needed.
 - Will be adding support for **RFC 4331**.
 - Provides information like `du -ks <directory>`.
 - For many VOs, may be a substitute for SRM space accounting.
-

NFS

- Deployed at DESY for **CMS** (WNs and NAF) and **other users**:
 - “Interactive” NAF has CMS permanently mounted.
 - Grid WN was for a limited period, with some fraction of WNs using NFS.
 - Operations are basically OK, but performance is under investigation.
 - When network is working well, performance is comparable with dcap
 - Statistics show a slight decrease in performance, but a good starting point.
 - NFS protocol also allows the clients (WNs) reading data through the door, which is used as a fall-back if there's a problem. This works, but there is a large impact on performance. We're working on this by:
 - tuning the client to be more resilient to minor, transitory problems,
 - reducing this impact when the client falls back and reads through the door.
-

xrootd

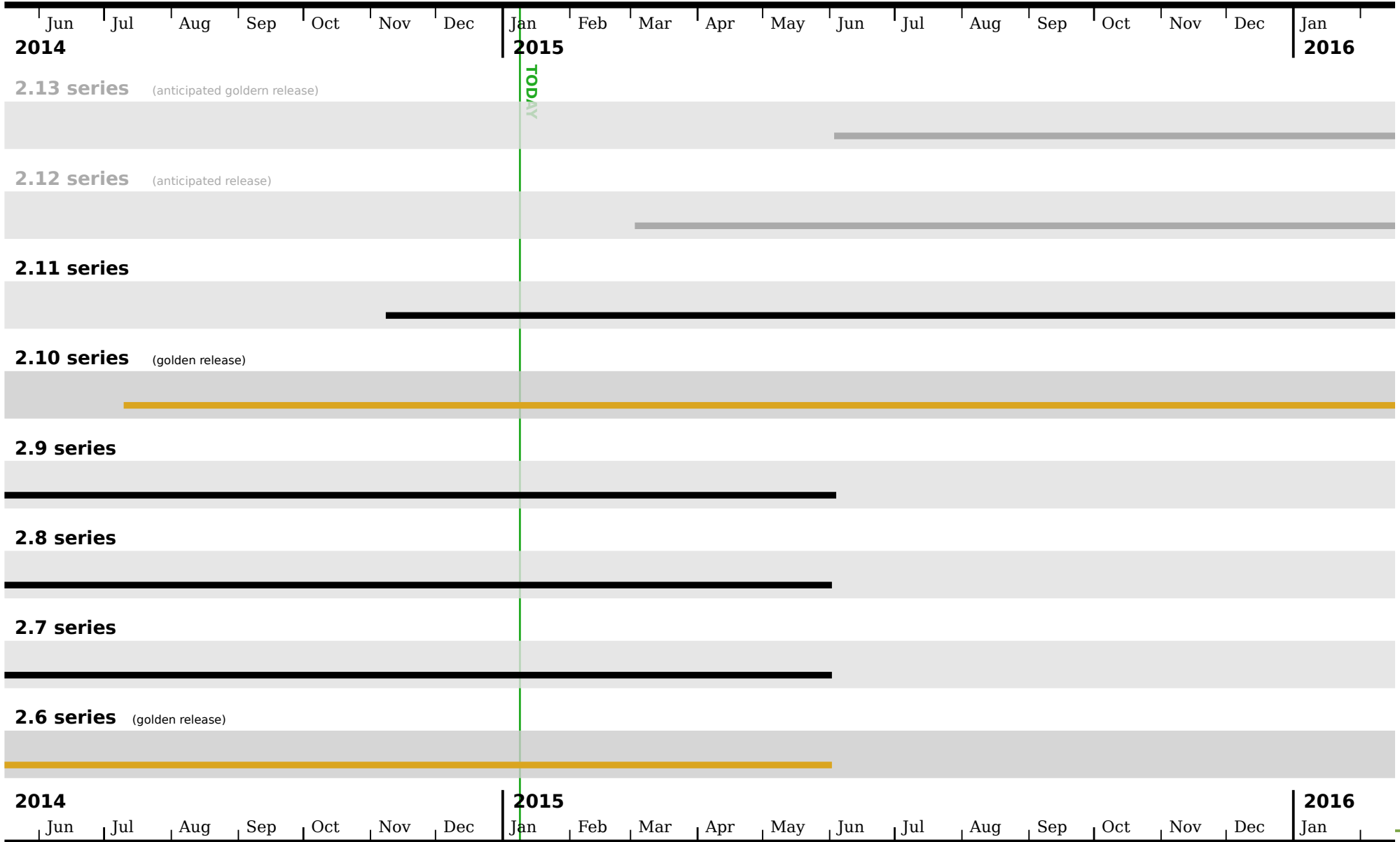
- Plugins allow communities to **extend** basic functionality:
 - CMS, ATLAS and ALICE make use of this
 - Used for monitoring, name-to-name translation (federation), access-control, redirection, ...
 - Plugin framework will change in the near future:
dCache v2.12 (~1st March) or v2.13 (~1st July)
(don't panic)
 - **3rd-party copy**: currently no road-map.
-

Authentication

- Many people looking at **SAML** (or OpenID-Connect) to replace X.509
 - dCache team are investigating this in collaboration with LSDMA, EGI FedCloud and OGF.
 - Medium term solutions will likely involve **gateway services**:
 - Infrastructure continues to use X.509; online CAs generate X.509 certificates.
 - Does not exclude any of the listed protocols.
-

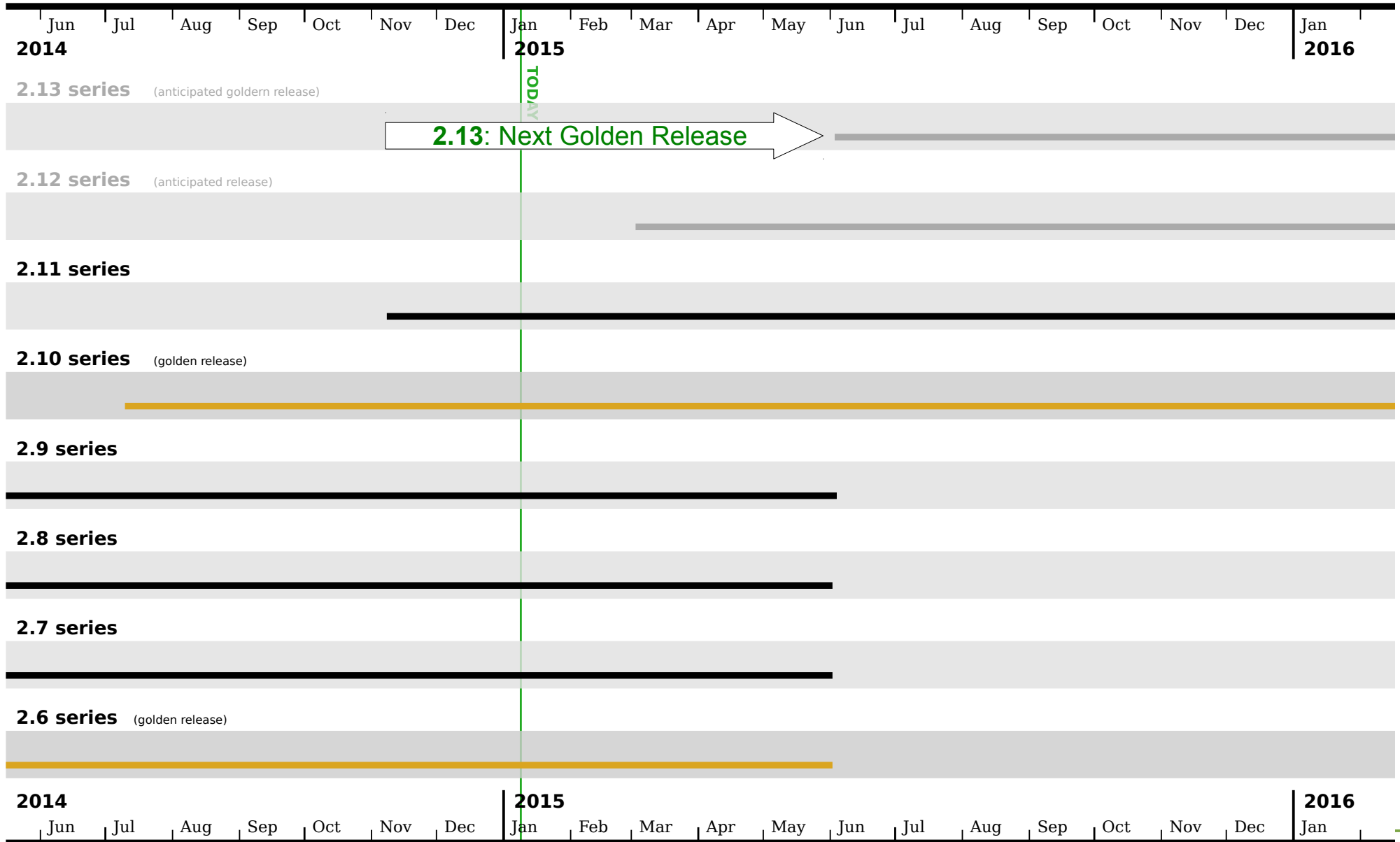
dCache server releases

... along with the series support durations.



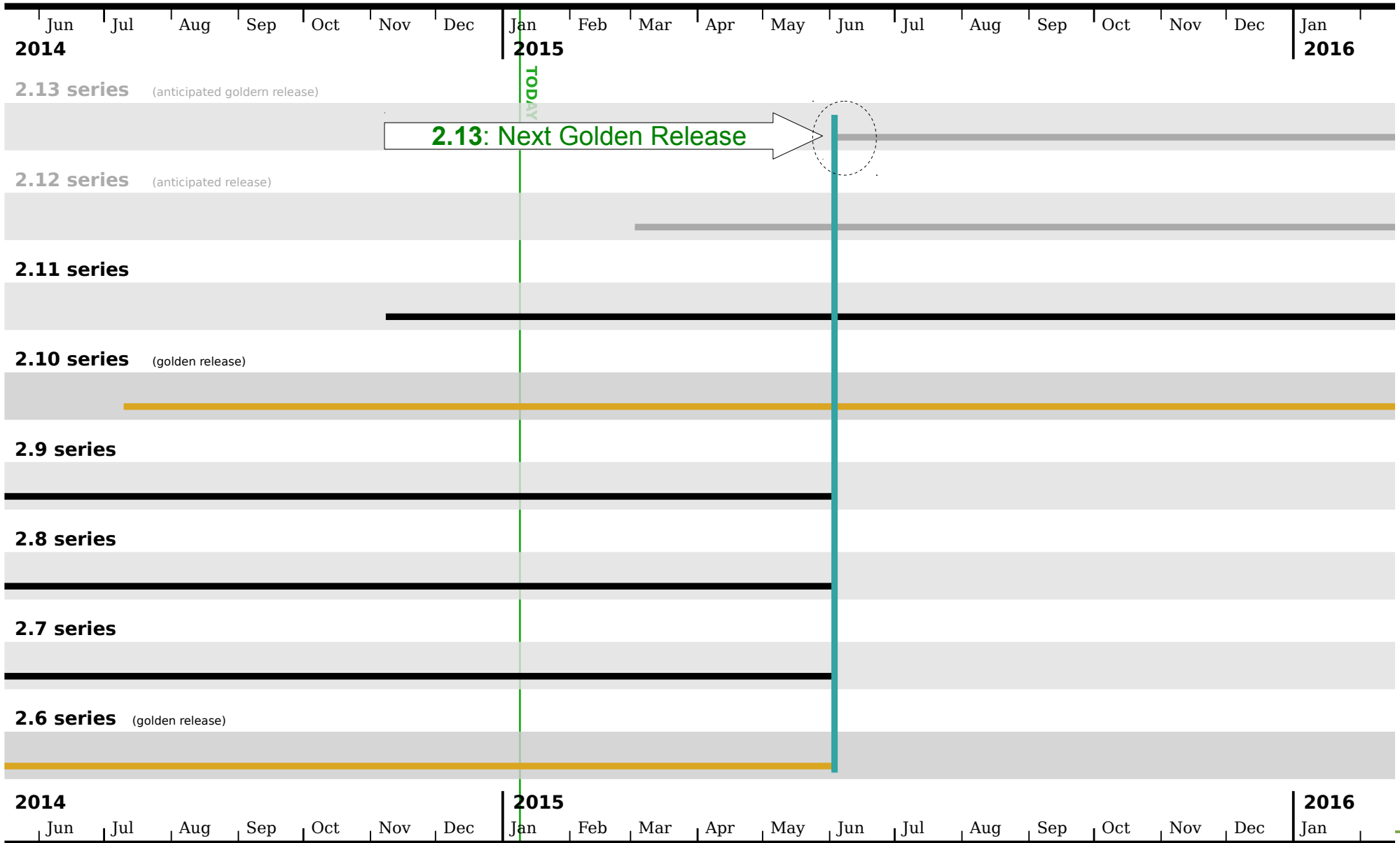
dCache server releases

... along with the series support durations.



dCache server releases

... along with the series support durations.



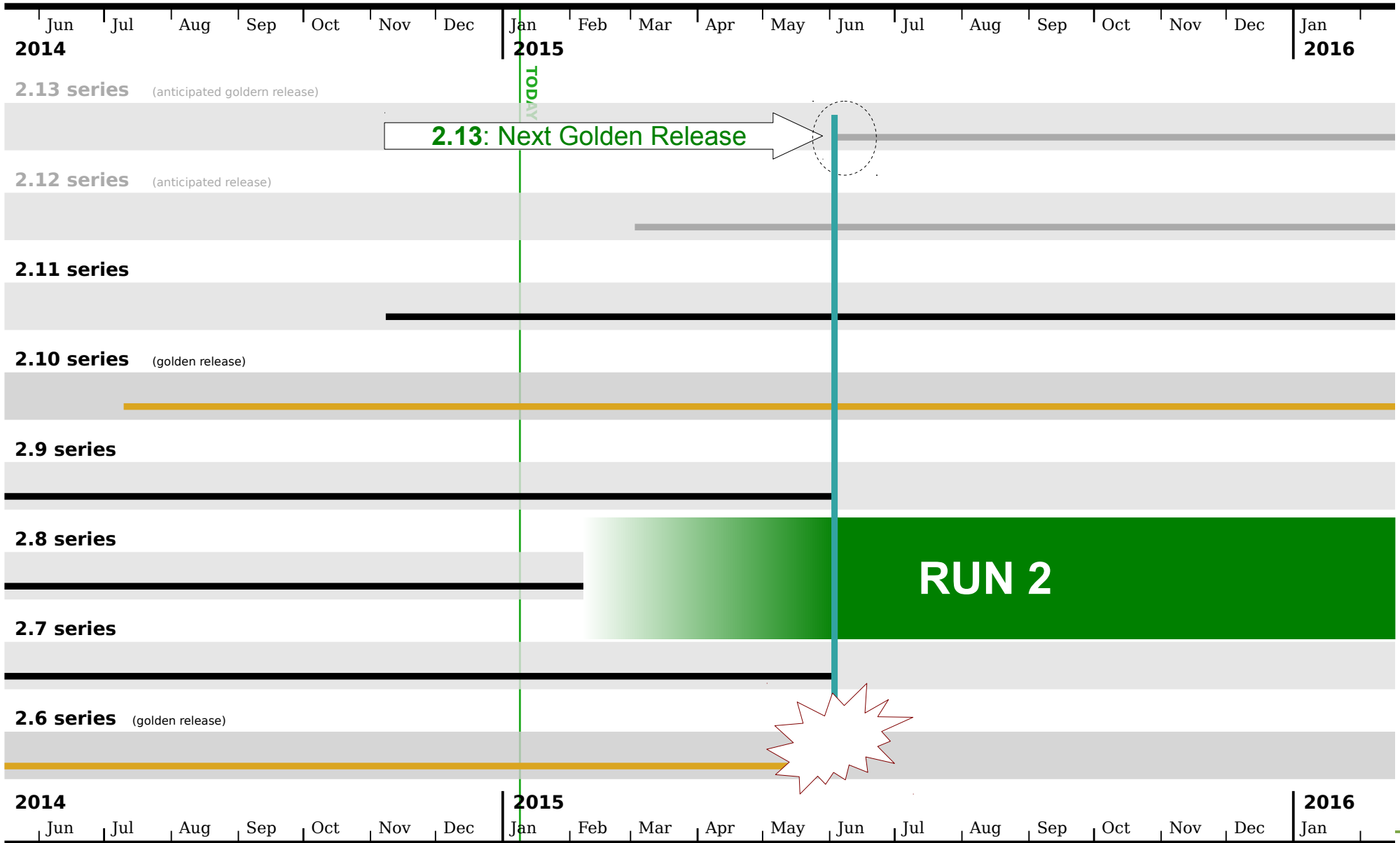
dCache server releases

... along with the series support durations.



dCache server releases

... along with the series support durations.



One more thing...

Prometheus

- The goal:
 - try to discover bugs before software hits production services
- The problem:
 - We can't test all (ATLAS, CMS, LHCb, Alice, ...) use-cases.
- The offer: **prometheus.desy.de**
 - A test instance, rebuilt daily (data is lost overnight),
 - Always the tip of development branch (currently 2.12.0-SNAPSHOT),
 - Anyone** from **atlas**, **cms**, **lhcb**, **alice** and **dteam** VOs can use this service **right now**.
 - Verify their software-stack works with the next major-version dCache.
 - Bug-fixes (which will be rolled out on production services) are available first in prometheus.
- If you're interested, start testing – contact me if there's any problems.



Thanks for listening, any questions?

Backup slides



DCAP vs NFS performance

