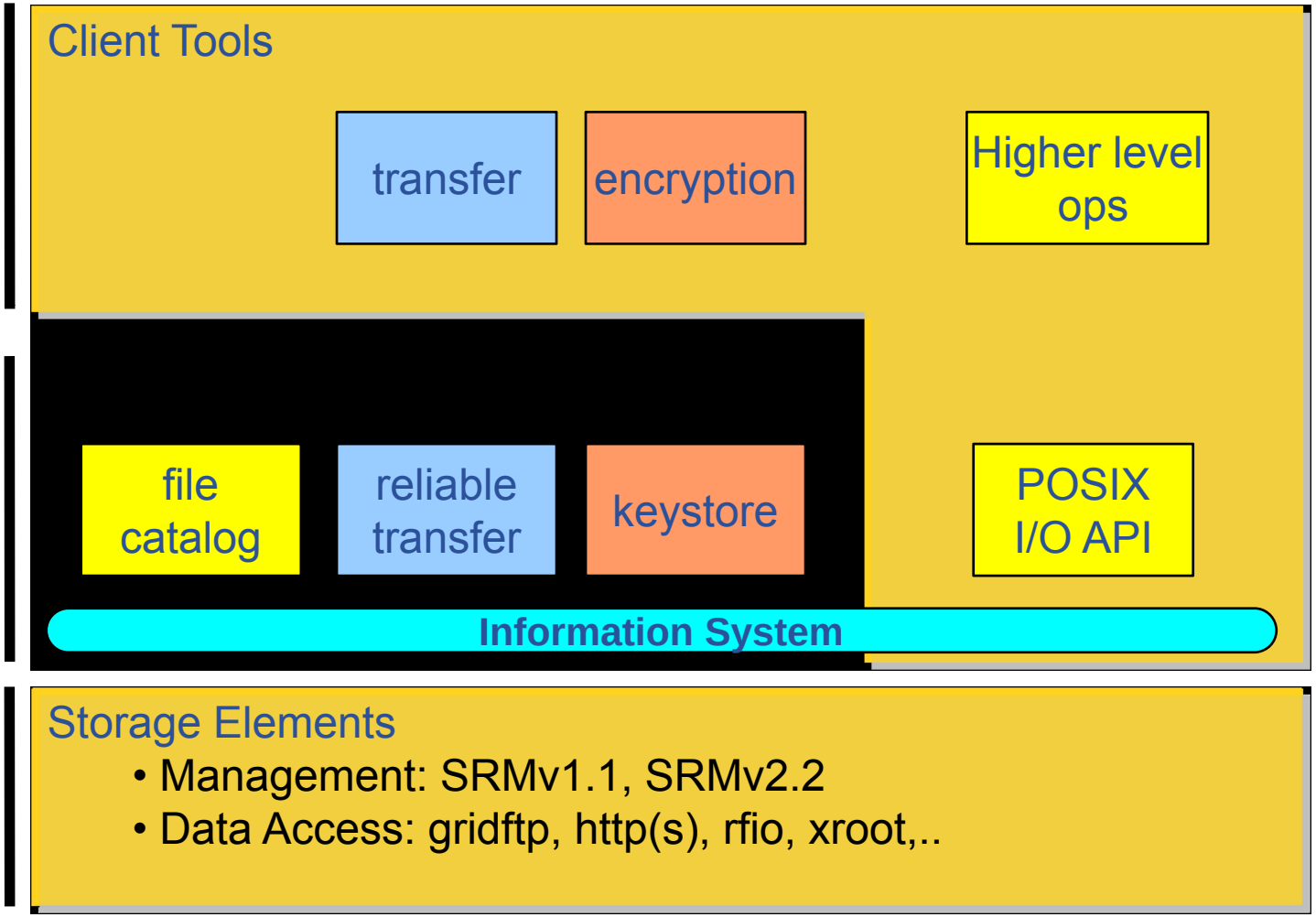
Data Management

Ákos Frohner
Presented by German Cancio
CERN, IT-DM

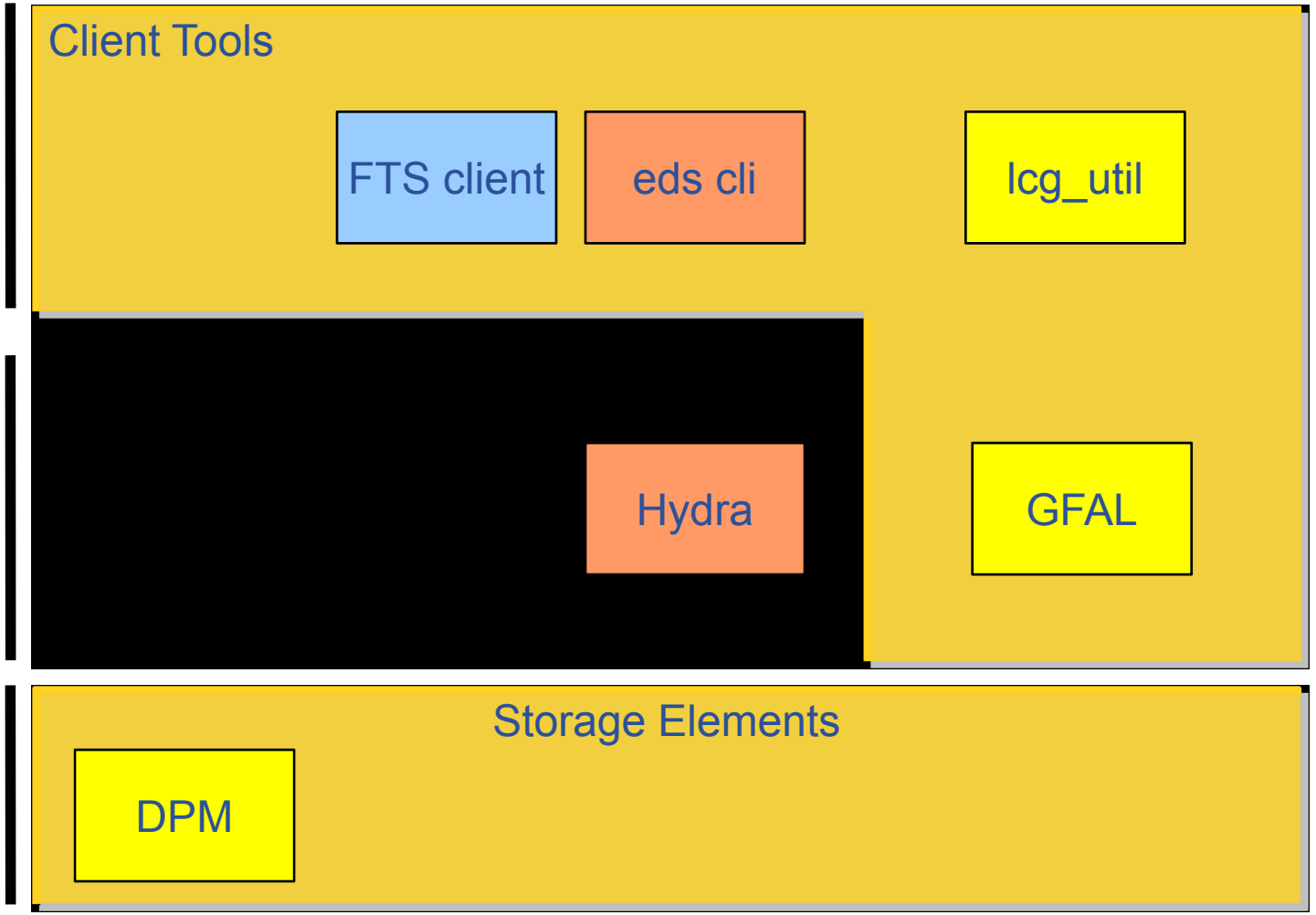
Outline

- EGEE project
- Data Management Software Stack
- Storage Element: DPM
- File Catalog: LFC
- Transfer: FTS
- Clients: GFAL, lcg_util
- Future directions

DM Software Stack



gLite DM Software Stack



Deployment Status

- GFAL, lcg_util clients on all worker nodes ~72,000
- 46 instances of LFC is used by two major VOs
- 15 FTS instances manage several 100s of MB/sec (up to 1.5GB/sec) traffic each, via 636 channels
- 190 DPM instances manage up to 360TB data on a single SE

Common Features

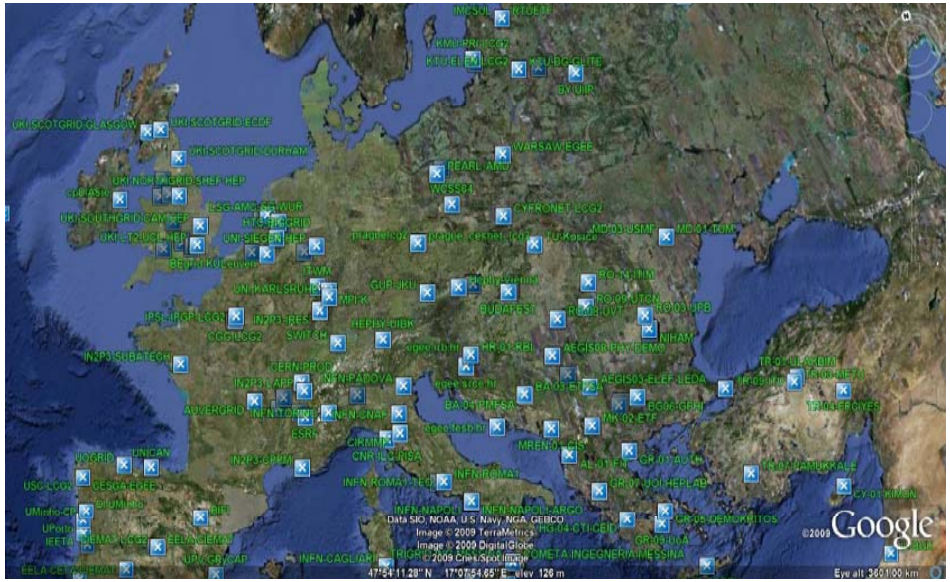
- **Security**
  - GSI authentication
  - VOMS authorization, including FQANs
- **Information system**
  - BDII info providers and queries
- **Storage protocols**
  - SRM v1.1 and SRMv2.2 support client or server respectively
  - gridftp

Outline

- EGEE project
- Data Management Software Stack
- **Storage Element: DPM**
- File Catalog: LFC
- Transfer: FTS
- Clients: GFAL, lcg_util
- Future directions

SE: DPM

- **Manages disk-only storage**
  - Deployments up to few 100TB data
  - Easy to install and manage
  - Interoperability with other storage systems
- **SRM v1.1 and v2.2 support**
- **Data access**
  - Gridftp
  - Secure rfio
  - https with re-direction
  - Xroot (special)
- **IPv6 ready**



DPM: features

- **POSIX authorization semantics**
  - user = certificate DN
  - group = VOMS FQAN
  - Independent of the underlying OS
  - Full ACL and secondary group support
- **Disk pool management**
  - Pool/space protection
  - Garbage collector
  - Replication of hot files

DPM v1.7.x

- **DPM v1.7.2 includes**
  - Compliant with the WLCG SRM MoU
  - Checksum support: Adler32, MD5, Crc32
  - srmCopy support
  - Multi-platform ready (packaging for RedHat, Debian, Solaris)
- **Upcoming in v1.7.3:**
  - Update of the xrootd plug-in for Alice
  - gLite release on Debian

DPM future releases

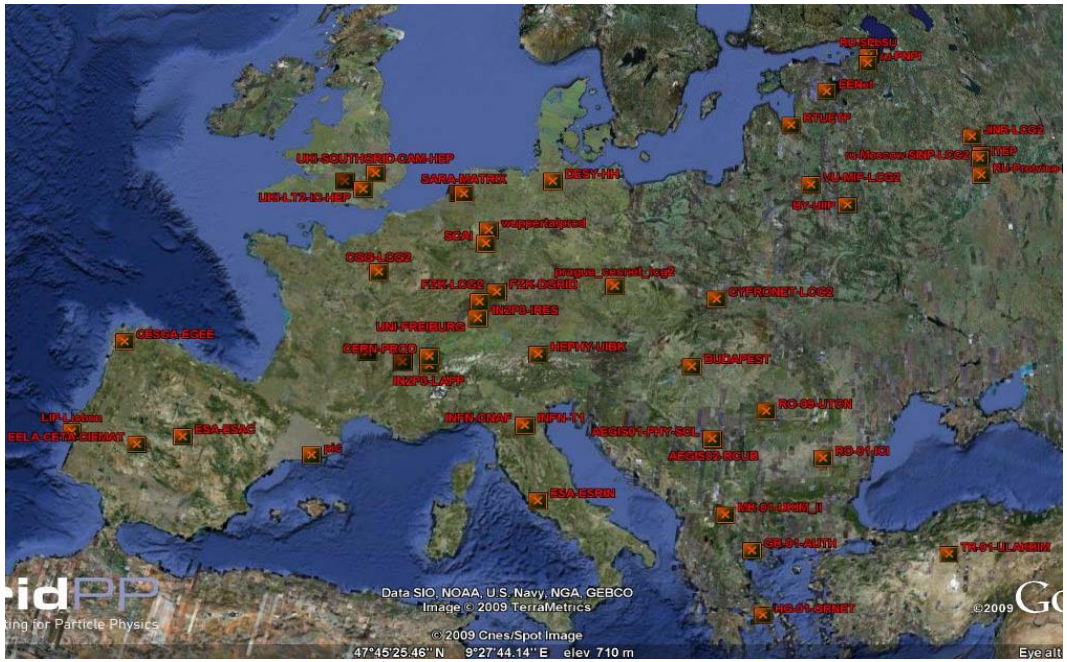
- **v1.8.0:**
  - On-the-fly checksum calculation (transfers)
  - Faster draining of disk servers
  - DB maintenance tools
  - User/VO banning
  - multi-VO xroot support
- **v1.9.x**
  - Better filesystem selection
  - Per-VO “Admin” role
  - Quota support

Outline

- EGEE project
- Data Management Software Stack
- Storage Element: DPM
- **File Catalog: LFC**
- Transfer: FTS
- Clients: GFAL, lcg_util
- Future directions

File Catalog: LFC

- **Logical to many physical file (SURL) mapping**
- **POSIX authorization semantics (see DPM)**
- **C and Python API (improved)**
- **Deployment models**
  - Central with DB replication
  - Local catalogs
- **OS & DB:**
  - As with DPM
- **IPv6 ready**



LFC v1.7.2 and after

Current release of LFC

- Multi-platform support
- Focus is on performance by bulk and compound methods

v1.5.x from gLite 3.0:

lfc_delreplica(replica)

lfc_unlink(name)

lfc_readdir(directory)

v1.6.x: up to 10 times speedup

lfc_delfilesbyguid(array of guids)

lfc_delfilesbyname(array of names)

lfc_delfilesbypattern(pattern)

lfc_delreplicas(array of replicas)

lfc_getreplicas(array of guids)

lfc_readdirxp(directory, pattern, SE)
reading subset of a directory

lfc_getreplicasl(array of lfns)

v1.7.x: (for LHCb and Atlas)

lfc_getreplicasx(array of lfns)
extended lfc_getreplicasl()

lfc_delreplicasbysfn(array of sfns)

lfc_registerfiles(array of entries)

Outline

- EGEE project
- Data Management Software Stack
- Storage Element: DPM
- File Catalog: LFC
- **Transfer: FTS**
- Clients: GFAL, lcg_util
- Future directions

Transfer: FTS

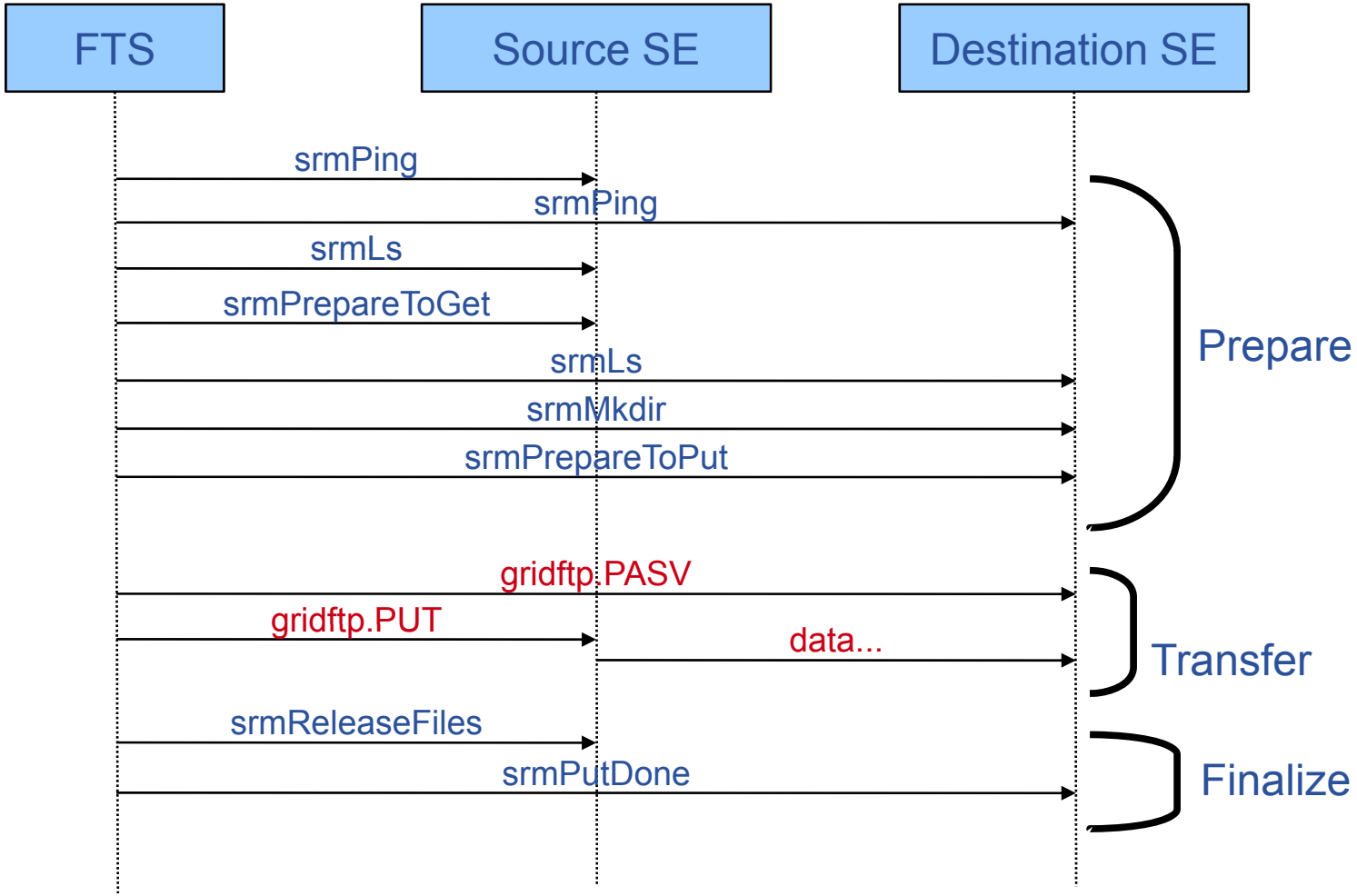
- **Reliable File Transfer Service**
  - Bulk data transfers between SRM compliant storage elements
  - Multi-VO service to balance network/SE utilization
  - Prevent overloading network/SE resources
  - Service monitoring and statistics

FTS: Channel

- **Single direction queue for transfer jobs**
- **Between**
  - CERN-RAL — single sites
  - siteA-[T2region] — new management tools
  - CERN-* — catch-all channel
- **Parameters**
  - Detailed timeouts — proportional to filesize
  - VO shares
  - Priorities
  - Limit of parallel transfers — srm/gridftp split!

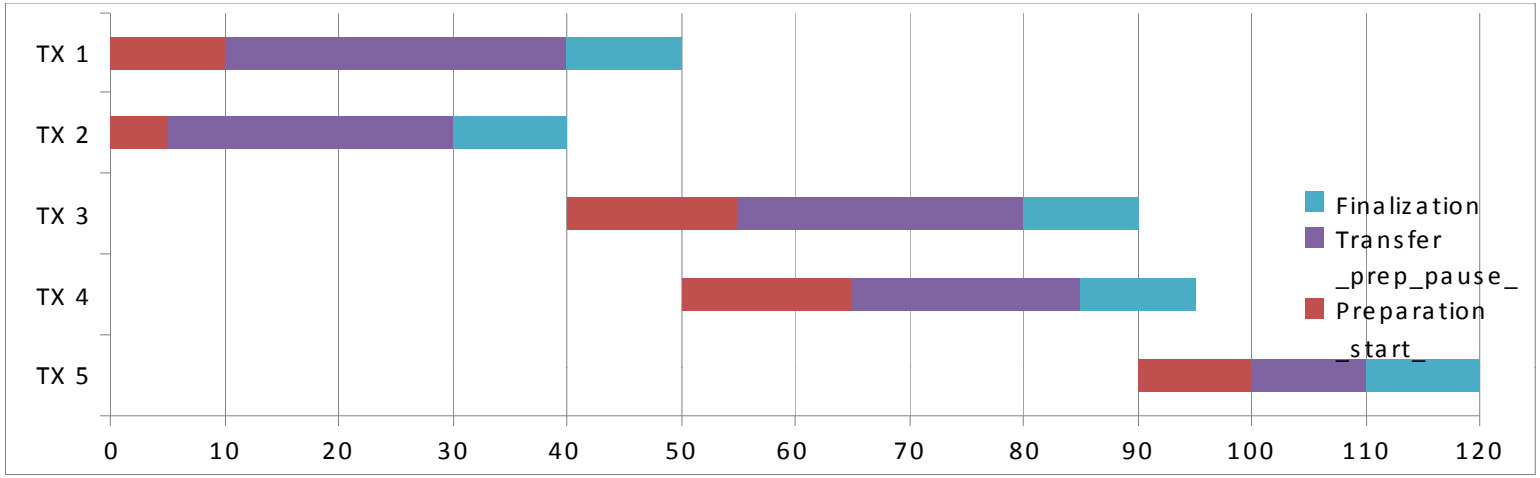
FTS: srm/gridftp split

FTS <= 2.1: channel = sum(Prepare + Transfer + Finalize)
FTS >= 2.2: channel = sum(Transfer)

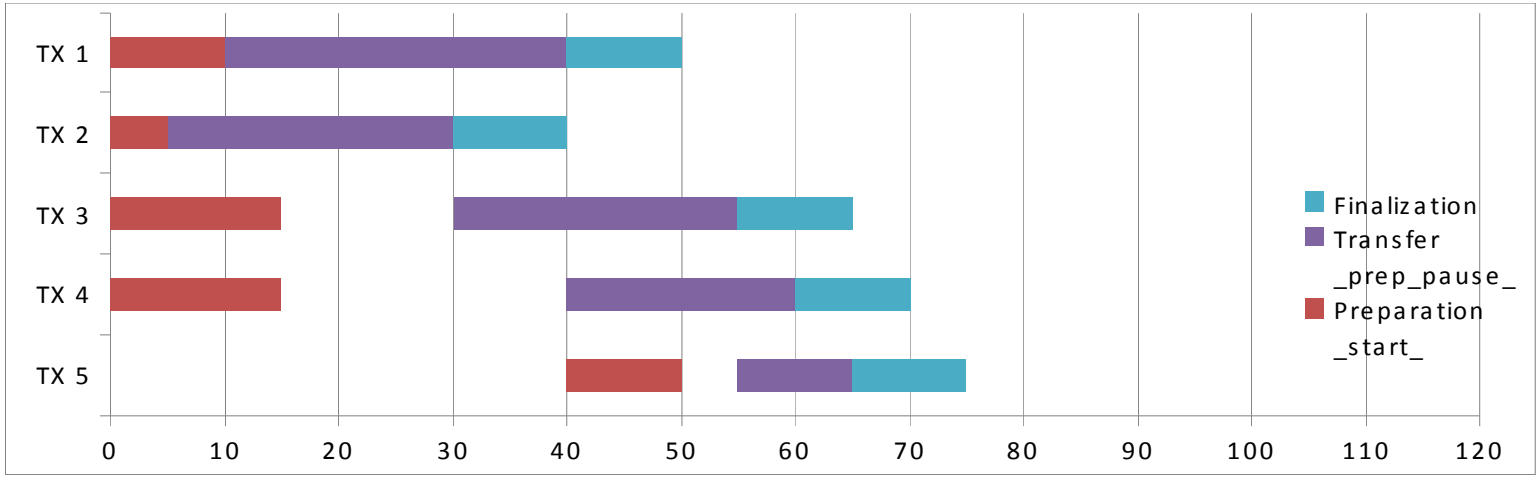


FTS 2.2 speedup

FTS 2.1:



FTS 2.2 with split srm/gridftp interactions:



FTS 2.2

- **Current (2.2) release of FTS (48 fixes/features)**
  - srm/gridftp split in 'urlcopy' channels
  - CLI tools for channel configuration
  - Implementing the WLCG SRM MoU
  - Smarter timeouts for large files
  - Multi-platform release
- **Upcoming releases**
  - Checksum support (FTS 2.2+)
  - Avoiding SRM overload (aka 'srm busy')
  - New WSDL and Python client library (FTS 2.3)
  - Live transfer information (FTS 2.3)

Outline

- EGEE project
- Data Management Software Stack
- Storage Element: DPM
- File Catalog: LFC
- Transfer: FTS
- **Clients: GFAL, lcg_util**
- Future directions

Client: GFAL, lcg_util

- **Hides the complexity for users**
- **GFAL**
  - POSIX-like I/O functions (open(), read(), close())
  - SRM abstraction layer
  - C, Python APIs and CLI
- **lcg_util**
  - Covering most common use cases:
    - File creation, registration, replication, deletion,...
  - C, Python APIs and CLI

GFAL v1.11.x and lcg_util v1.7.x

- **Current release:**
  - Detailed timeout support
  - Fall-back on secondary replica
    - If first is not available
    - Classified replicas (local, remote)
  - Improved Python interface
  - Implementing the WLCG SRM MoU
- **Upcoming releases**
  - Checksum support
  - Asynchronous srm-ls
  - Avoiding SRM overload (aka 'srm busy')
- **In the long term:**
  fully synchronize the behavior with FTS

Outline

- EGEE project
- Data Management Software Stack
- Storage Element: DPM
- File Catalog: LFC
- Transfer: FTS
- Clients: GFAL, lcg_util
- **Future directions**

Moving to “DM Product Team”

Increased responsibility: testing for release

- **DPM current practice**
  - Internal certification on SLC4, SLC5
  - Internal test-suite, lcg_util, FTS
- **FTS current practice**
  - Unit tests, integration tests, deployment tests
  - Pilot service on SLC4 (and SLC5) with LHC experiments
- **What are the resources consumed for certification?**
  - These will need to be substracted from dvp effort
- **Node types split up?**
- **What is the “product” expected by SA3?**
  - Versioned tarball, Source RPM, Binary RPM
  - ETICS configuration

Software Process

Thinking beyond EGEE-III: what services will be available for the developers?

- **CVS/SVN: provided by CERN**
  - DVCS could be used in parallel
- **Savannah: provided by CERN**
  - alternatives would require migration of current items
- **ETICS: who would run it after EGEE-III?**
  - LFC is built without ETICS in NorduGrid and VDT
  - FTS is only built with ETICS currently
- **Metapackages, configuration**
  - Will come from the development cluster, part of the release
- **Documentation, web pages, support lists**
  - Provided by the development cluster, private resources

Future directions

- **Priorities: focus on**
  - Stability
  - Reliability
  - Maintainability

  } rather than on new tools/functionality
- **Improved administrative tools**
  - Real time monitoring of services
  - Automating regular procedures (e.g. cleanup)
  - Friendly (web) interface for configuration
  - Resource protection (quota, limits)
- **Better integration**
  - Status feedback from services (i.e. “srm busy”)
  - Framework consolidation / integration (libraries)