LHCb DC06 Activity

o   DC06 aims

o   DC06 activities

LHCb DC'06

Challenge (using the LCG production services):

a) Distribution of RAW data from CERN to Tier-1's

b) Reconstruction/stripping at Tier-1's including CERN

c) DST distribution to CERN & other Tier-1's

Ongoing over the summer

LHCb Tier-1's:

CNAF	~~NIKHEF~~
GridKa	PIC
IN2P3	RAL

LHCb DC'06

Distribution of RAW data from CERN

~ 40 TB of storage replicated over Tier-1 sites (each file on only one Tier-1)

CERN MSS SRM → local MSS SRM endpoint

Reconstruction/stripping at Tier-1's including CERN

~300 kSI2k.months needed to reconstruct & strip events

Output:

- rDST (from recons): ~10 TB (accumulate) at Tier-1 sites where produced, on MSS

- DST (from stripping): ~1.2TB in total on disk SE

DST distribution to CERN & all other Tier-1's

- 1.2 TB of DST will be replicated to a disk based SE at each Tier-1 & CERN

Reconstruction (CERN & Tier-1's)

500 events per file

Assume a job 10k events per job (20-40 i/p files)

- ~11 hours/job

- 2.0 GB/job input; 0.5 GB/job output

- Output stored at Tier-1 where job runs

- 10k jobs

Comparison with computing model:

2-3 GB (single) input files - ~100k events

2-3 GB output data

Jobs duration ~36 hours

~30k recons jobs per month during data taking

<u>Stripping (CERN & Tier-1's)</u>

Assume a job 40k events per job (4 i/p files)

     ~2 hours/job

     10 GB input/job; ~1.5 GB data output per job

     output distributed to ALL Tier-1's & CERN

     2.5k jobs

Comparison with computing model:

Jobs duration ~36 hours

10 input rDST files + 10 input RAW files - ~1M events

50 GB data input per job; ~7GB data output per job

~3k stripping jobs per month during data taking

Recons & Stripping (CERN & Tier-1′s)

Need to be able to prioritise recons/stripping jobs for LHCb

LHCb prioritisation dealt with by DIRAC workload management system

Done for production vs recons – now need to address analysis

SRM -based storage

Implementation across back-ends technologies e.g. CASTOR, dCache, …

- Would′ve benefited from SRM v2.2

Tape and disk storage endpoints at all T1 sites

File Access & GFAL

Aim was to use GFAL for a consistent access layer across all sites
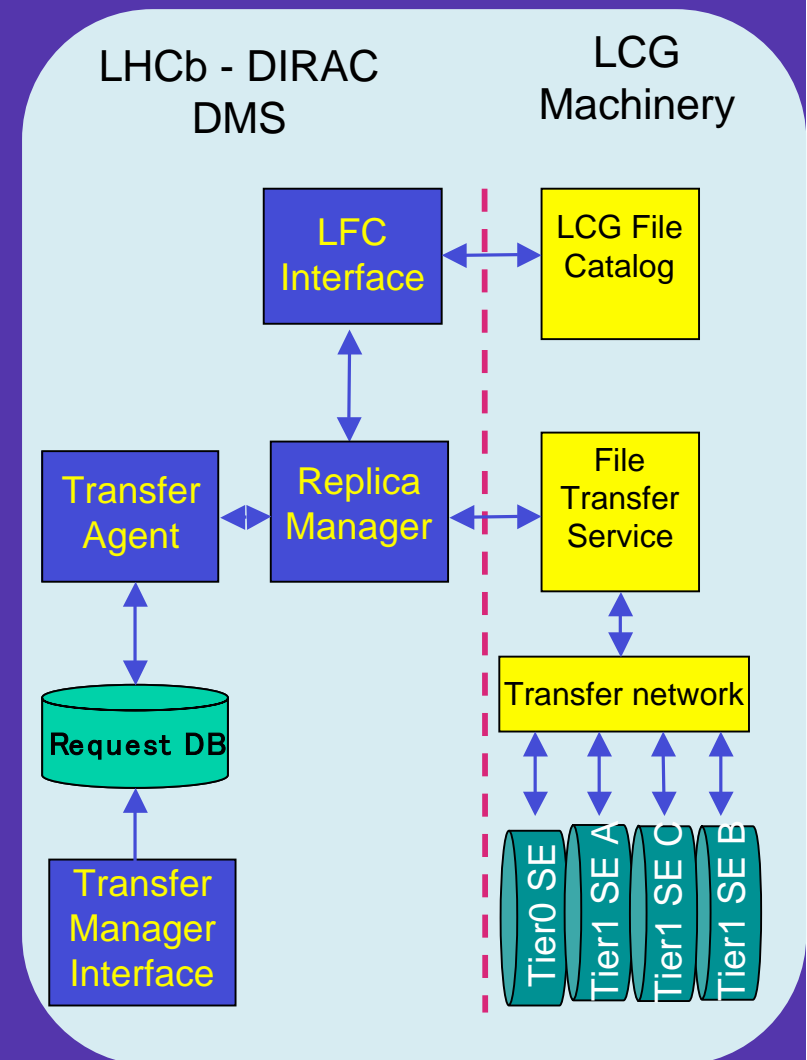
Still issues associated with ROOT plug-in

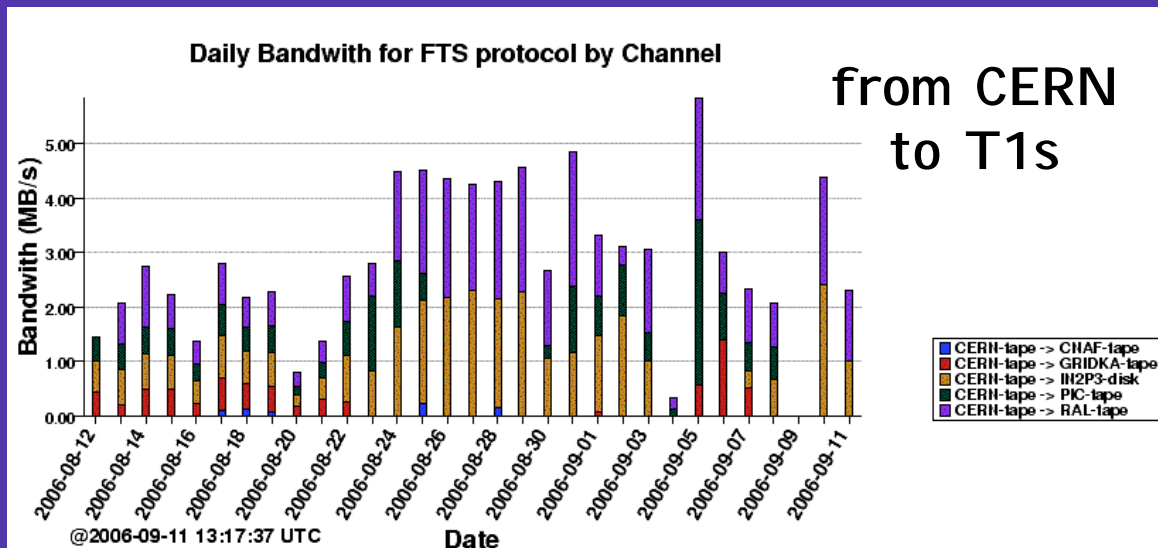- Still being debugged

Gsidcap protocol issues

- Issues to do with callback & firewalls - NIKHEF/SARA problem - beta-version of dCache currently undergoing tests

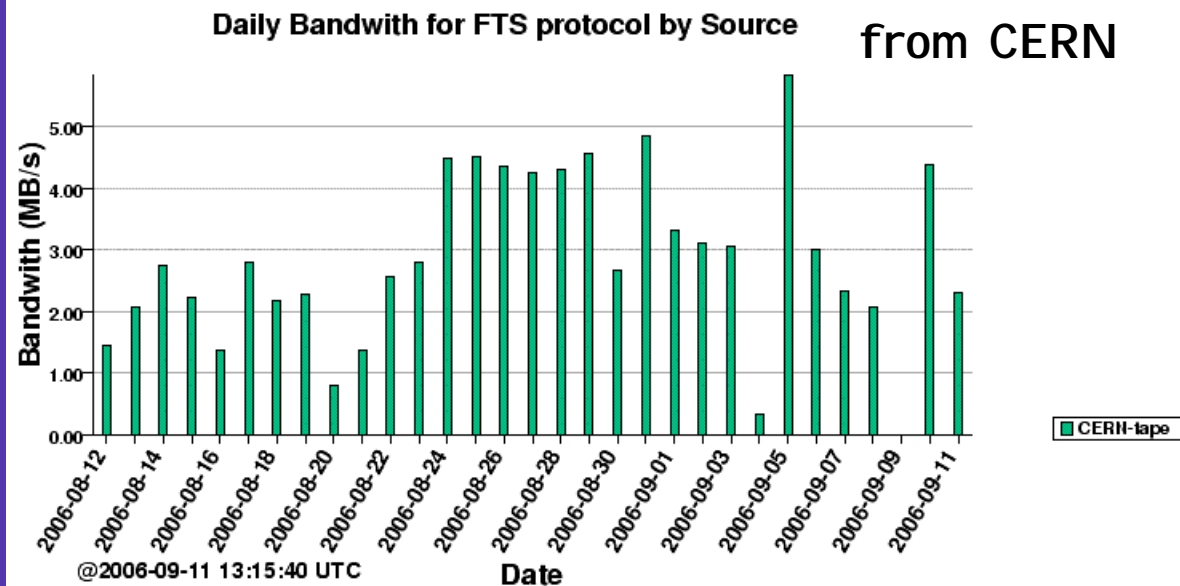- Backdoor way of using protocol in ROOT - undergoing tests

FTS

- FTS (gLite File Transfer Service) is used by the Dirac Data Management
  - Reconstruction and Stripping jobs are created by a transformation agent automatically
    - The DIGI files are sent to a T1 site, automatically a corresponding reconstruction job is sent to that site
  - Successfully tested (many times!) for reconstruction
  - Will also be used for stripping jobs

LHCb - DIRAC DMS

LCG Machinery

LFC Interface

LCG File Catalog

Replica Manager

File Transfer Service

Transfer Agent

Request DB

Transfer network

Transfer Manager Interface

Tier0 SE

Tier1 SE A

Tier1 SE C

Tier1 SE B

DIGI data transfer to T1s

from CERN to T1s

Daily Bandwith for FTS protocol by Channel

- CERN-tape -> CNAF-tape
- CERN-tape -> GRIDKA-tape
- CERN-tape -> IN2P3-disk
- CERN-tape -> PIC-tape
- CERN-tape -> RAL-tape

@2006-09-11 13:17:37 UTC

Over the last month

Daily Bandwith for FTS protocol by Source

from CERN

- CERN-tape

@2006-09-11 13:15:40 UTC

FTS status

- Many problems on the sites
  - Storage, bandwidth issues also affect file transfers
  - Sites have fewer problems with the FTS (compared to reconstruction)
- Recent problems
  - RAL gridFTP doors crashed
    - Twice in 3 days (2, 4 September 2006)
    - Okay now
  - GridKa : gridFTP server overload
  - Lyon : Power supply failure
  - CNAF : SE problem
- After initial period we planning to increase the data transfer rate to/from the stable sites

CERN

- Mostly stable at present
  - 19% of simulated events
  - 25% of reconstructed events
- Recent issues
  - Ability of production manager to write to SE
    - ACL maintained and updated "by hand"
    - Currently handled on a case-by-case basis
  - Uploads to CERN-CASTOR via gridFTP
    - Problem transferring from non-Tier1 to T0
    - Investigated by CERN / Cisco.
    - Understood & solved since 24 August (HTAR issue)
    - Occasional timeouts still seen
      - Now affects very few jobs
      - Job output is never uploaded – lost data
  - High loads on CERN RB
    - No longer a problem

GridKa

- Problems running reconstruction job until early September, no problem on simulation
  - 4% of simulated events
  - 1% of reconstructed events
- Problems accessing storage since 1 August 2006 (now fixed)
  - Problems with lcg-gt used within the jobs
    - Diagnosed as overload of gridFTP server when the transfer requests pile up for more than a few minutes
  - Affects both access of data and upload of data at GridKa
- Fixed
  - Data accessed locally
  - Output also uploaded locally

RAL

- Site currently fine / stable
  - 10% of simulated events
  - 24% of reconstructed events

- Problems in last month
  - Slow access of data in late July / early August
    - Solved by configuring additional LHCb dedicated disk server

- Upcoming issue
  - Move to CASTOR from dCache
    - No timescale yet – CASTOR not yet stable
    - dCache will end in early 2007 (~ April ?)
    - Need to discuss with RAL about moving our data to new SE

- Data transfer

CNAF

- Problems running reconstruction jobs, but no problem on simulation
    - 11% of simulated events
    - few reconstructed events
- Problems running reconstruction jobs with Castor-2
    - Castor-2 installation at CNAF is different from CERN
        - CERN: Single instance of DB and LSF queue for different experiments
        - CNAF: All the experiments share the same instance of DB and LSF
    - Problems
        - Bug on the Stager Data Base → now fixed
        - Usage of "pure" disk with Castor-2
            - No migration to tape, no Garbage Collector if the space becomes full all the new requests become pending jobs (LSF queue overloaded); other VO interference!
        - The Castor2 DB is overloaded even if there is no external activities
            - The problem seems to be fixed, reconstruction jobs are running

PIC

- Site currently fine / stable
  - 2 % of simulated events
  - 19 % of reconstructed events
- Problems during last month
  - PIC was unable to process reconstruction job due to software installation problem (now fixed)

IN2P3

- ## Site currently fine / stable
  - 3% of simulated events
  - 30% of reconstructed events
- ## Site mainly used as DIRAC site
  - Working with IN2P3 support to get the LCG setup working
    - Grid accessible queues need to become longer (done beginning of Sept)
    - Now is working as DIRAC and LCG site
- ## Problems during last month
  - Batch system at Lyon went down for few days

NIKHEF/SARA

- Problems running reconstruction jobs, but no problem with simulations
  - 3% of simulated events
  - No reconstructed events
- Problems accessing data
  - accessing files stored in the WAN connected Storage from WN via dCache is not available
    - The NIKHEF firewall doesn't allow that
  - Waiting to test patched version of the dCache client
    - This version will not require inbound connectivity on the WN anymore

Summary of Reconstruction

Site	Low lumi jobs	Low lumi events	High lumi jobs	High lumi events	High+low
CERN	305	3.05M	288	2.88M	25%
CNAF	-	-	14	0.14M	<1%
GRIDKA	-	-	28	0.30M	1%
IN2P3(DIRAC)	464	4.64M	251	2.51M	30%
NIKHEF/SARA	-	-	-	-	-
PIC	178	1.78M	273	2.73M	19%
RAL	282	2.82M	281	2.81M	24%
TOTAL		12.3M		11.3M	

Data access problematic at many Tier-1 centres

Conclusions

- **Simulation**
  - running smoothly for several months
  - Continuing to run production for LHCb physics book

- **Reconstruction**
  - Difficult to run reconstruction jobs at all T1's
  - 24M of events reconstructed at CERN(25%), RAL(24%), PIC(19%), IN2P3(30%)

- **Stripping**
  - The rDST files are ready to be processed - waiting to be able to get the ancestor for a given rDST

- **Data Transfer**
  - No particular problem related to the applications used to transfer data
  - The failures are mainly related to problems at the sites
    - SE, power supply failure, gridFTP servers overloaded