LCG Service Challenges Planning for Tier2 Sites

Jamie Shiers

IT-GD, CERN

Introduction

- Roles of Tier2 sites

- Services they require

- Timescale for involving T2s in LCG SCs

- Possible procedure

The Problem (or at least part of it...)

- SC1 – December 2004
- SC2 – March 2005

Neither of these involve T2s or even the experiments – just basic infrastructure

- SC3 – from July 2005 involves 2 Tier2s
  - + experiments' software + catalogs + other additional stuff
- SCn – completes at least 6 months prior to LHC data taking should involve all Tier1s and ~all Tier2s

- Not clear how many T2s there will be

- Maybe 50; maybe 100 – still a huge number to add!

Tier2 Roles

- Tier2 roles vary by experiment, but include:

  - **Production of simulated data;**
  - **Production of calibration constants;**
  - Active role in end-user analysis

- Must also consider services offered to T2s by T1s

  - e.g. safe-guarding of simulation output;
  - Delivery of analysis input.

- No fixed dependency between a given T2 and T1

  - But 'infinite flexibility' has a cost…

T2 Functionality

(At least) two distinct cases:

- ## Simulation output
  - This is relatively straightforward to handle
  - Most simplistic case: associate a T2 with a given T1
    - Can be reconfigured
    - Logical unavailability of a T1 could eventually mean that T2 MC production might stall
  - More complex scenarios possible
    - But why? Make it as simple as possible, but no simpler…

- ## Analysis
  - Much less well understood and likely much harder…

T1/T2 Roles

Tier1	Tier2

• Keep certain portions of RAW, ESD, sim ESD
 	
• Full copies of AOD + TAG, calibration data
 	
• Official physics group large scale data analysis
 	
• ALICE + LHCb:
   • also contribute to simulation
	• Keep certain portions of AOD and full copies of TAG for real + simulated data
	   • LHCb: sim only at T2s
	
	• Selected ESD samples
	
	• Produce simulated data
	
	• General end-user analysis

Based on "T1 Services for T2 Centres" document

(Just type this into Google)

Analysis

- Certain subsets of the data will be distributed across T0 and T1s
- **Must allow equal access to all data regardless of users' and its location**
- But this does not imply same physical network connectivity between every T2 and every T1…

- A model whereby data is handed between T1s rather than directly from 'remote' T1 to T2 may be much more affordable and manageable

- Needs further discussion and 'analysis'

Basic T2 Services

- T2s will need to provide services for data up- and down-load
- Assume that this uses the same components as between T0 and T1s
- Assume that this also includes an SRM interface to local disk pool manager
  - This should / could also be provided through LCG
- Networking requirements need to be further studied but current estimates suggest 1Gbit connections will satisfy needs of a given experiment
  - Can be dramatically affected by analysis model, heavy ion data processing etc

T2 Timeline

- Assume a small number of T2s fairly early in SC3

- Assume at least one of these will be from the UK

- SC3 preparations will be complex

- Need to start identifying candidate sites and ramp up procedure now

- Given the number of T2s, hope that this experience can be (extensively) re-used

T2 Model

- As much as possible, work through existing structures
  - GridPP in the UK, INFN in Italy, HEPiX, etc.

- Probably need also some 'regional' events
  - e.g. fall HEPiX for North American sites

- Complemented with workshops at CERN and elsewhere

Which Candidate T2 Sites?

- Would be useful to have:
  - Good local support from relevant experiment(s)
  - Some experience with disk pool mgr and file transfer s/w
  - 'Sufficient' local CPU and storage resources
  - 1Gbit/s network connection to T1 desirable
  - Manpower available to participate in SC3+
    - And also define relevant objectives?
- First T2 site(s) will no doubt be a learning process
- Hope to (semi-)automate this so that adding new sites can be achieved with low overhead

Summary

- The first T2 sites need to be actively involved in Service Challenges from Summer 2005

- ~All T2 sites need to be successfully integrated just over one year later

- Some candidate T2s already exist