



LCG Service Challenges: Planning for Tier2 Sites

Jamie Shiers
IT-GD, CERN





Introduction

- Roles of Tier2 sites
- Services they require (and offer)
- Timescale for involving T2s in LCG SCs
- Simplest (useful) T2 Model
- 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. Must involve all Tier1s and ~all Tier2s
 - Not clear how many T2s there will be
 - 💣 Maybe **50**; maybe **100** – still a *huge* number to add!
 - ALICE: 15? ,ATLAS: 30, CMS: 25, LHCb: 15; overlap?



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

- 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

Tier2

- 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
 - May even be a star configuration
- Needs further discussion and 'analysis' 😊

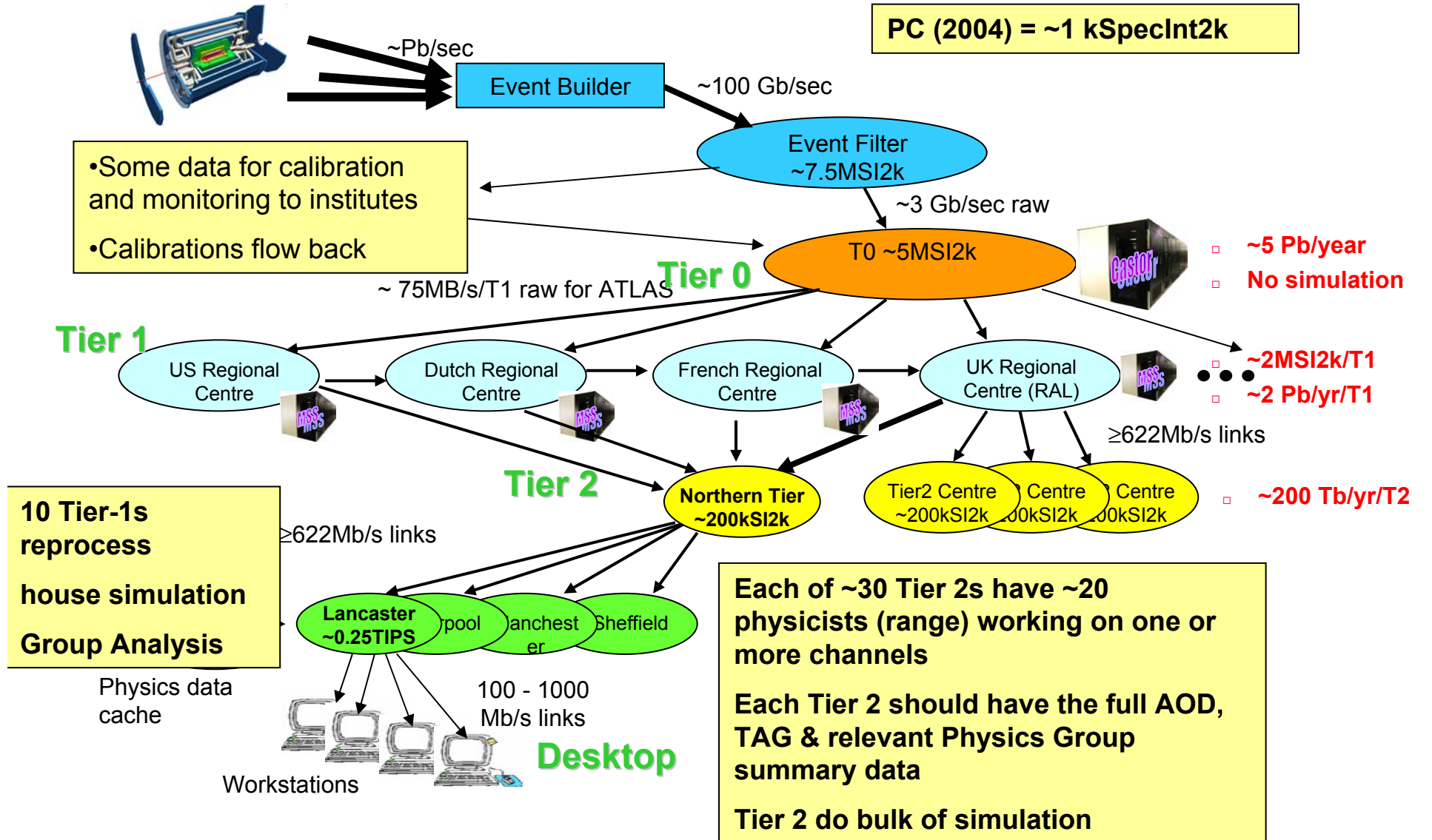


Analysis Example (ATLAS)

- A typical analysis scenario might begin with the physicist issuing a query against a **very large tag dataset**, e.g. the latest **reconstruction** of all data **taken to date**. For example, the query might be for events with **three leptons** and **missing transverse energy** above some threshold. The result of this query is used to define a dataset with the AOD information for these events. The analyst could then provide an **Athena algorithm** to make further event selection by refining the **electron** quality or missing **transverse** energy calculations. The new output dataset might be used to create an n-tuple for further analysis or the AOD data for the selected events could be copied into new files. A subset of particularly striking events identified in one of these samples could be used to construct a dataset that includes the ESD and perhaps even RAW data for these events. The physicist might then redo the electron reconstruction for these events and then use it to create a new AOD collection or n-tuple.



The (ATLAS) System





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 / these 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
 - 💣 T3 network requirements so far ignored(?)



Network Requirements (ATLAS)

Source	Inbound from CERN (MB/s)	Outbound to CERN (MB/s)
RAW	30.4	
ESD Versions	20	1.41
AOD versions	18	0.28
TAG Versions	0.18	0
Group DPD		0.81
Total CERN/AverageTier-1	68.58	2.51



T1-T1 Requirements (ATLAS)

Source	Inbound from other Tier-1s (MB/s)	Outbound to other Tier-1s (MB/s)
ESD Versions	12	14
AOD versions	20.8	2.08
TAG Versions	0.21	0.02
Group DPD		4
Total Tier-1 to Tier-1	33	19



T2 Specifications (CMS)

- T2 centres should have WAN connectivity **in the range** of 1GB/s or **more** to satisfy CMS analysis requirements
- T2 centres will require relatively sophisticated **disk cache management** systems, or explicit and enforceable local policy, to ensure sample latency on disk is adequate and to avoid disk/WAN thrashing



CMS Network Requirements

- We are pushing available networks to their limits in the Tier-1/Tier-2 connections
 - Tier-0 needs $\sim 2 \times 10 \text{ Gb/s}$ links **for CMS**
 - Each Tier-1 needs $\sim 10 \text{ Gb/s}$ links
 - Each Tier-2 needs 1 Gb/s for its incoming traffic
- There will be extreme upward pressure on these numbers as the distributed computing becomes more and more useable and effective

(Presentation to LHCC review of Computing Models)



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

Possible T2 Procedure(?)

➤ Work through existing structures where possible

- GridPP in the UK, INFN in Italy, HEPiX, etc.
- USATLAS, USCMS etc.
- Probably need also some 'regional' events
 - e.g. fall HEPiX for North American sites
- Complemented with workshops at CERN and elsewhere

➤ A UK T2 is clearly a strong option

- © Well established GridPP infrastructure + EasyJet connections
 - Need to look beyond the basic 'file transfer' service and consider also good local experiment support / involvement



A Simple T2 Model

N.B. this may vary from region to region

- Each T2 is configured to upload MC data **to** and download data **via** a given T1
- In case the T1 is logical unavailable, **wait and retry**
 - MC production might eventually stall
- For data download, **retrieve** via **alternate** route / T1
 - Which may well be at lower speed, but hopefully rare
- Data residing at a T1 other than 'preferred' T1 is transparently delivered through appropriate network route
 - T1s are expected to have at least as good interconnectivity as to T0

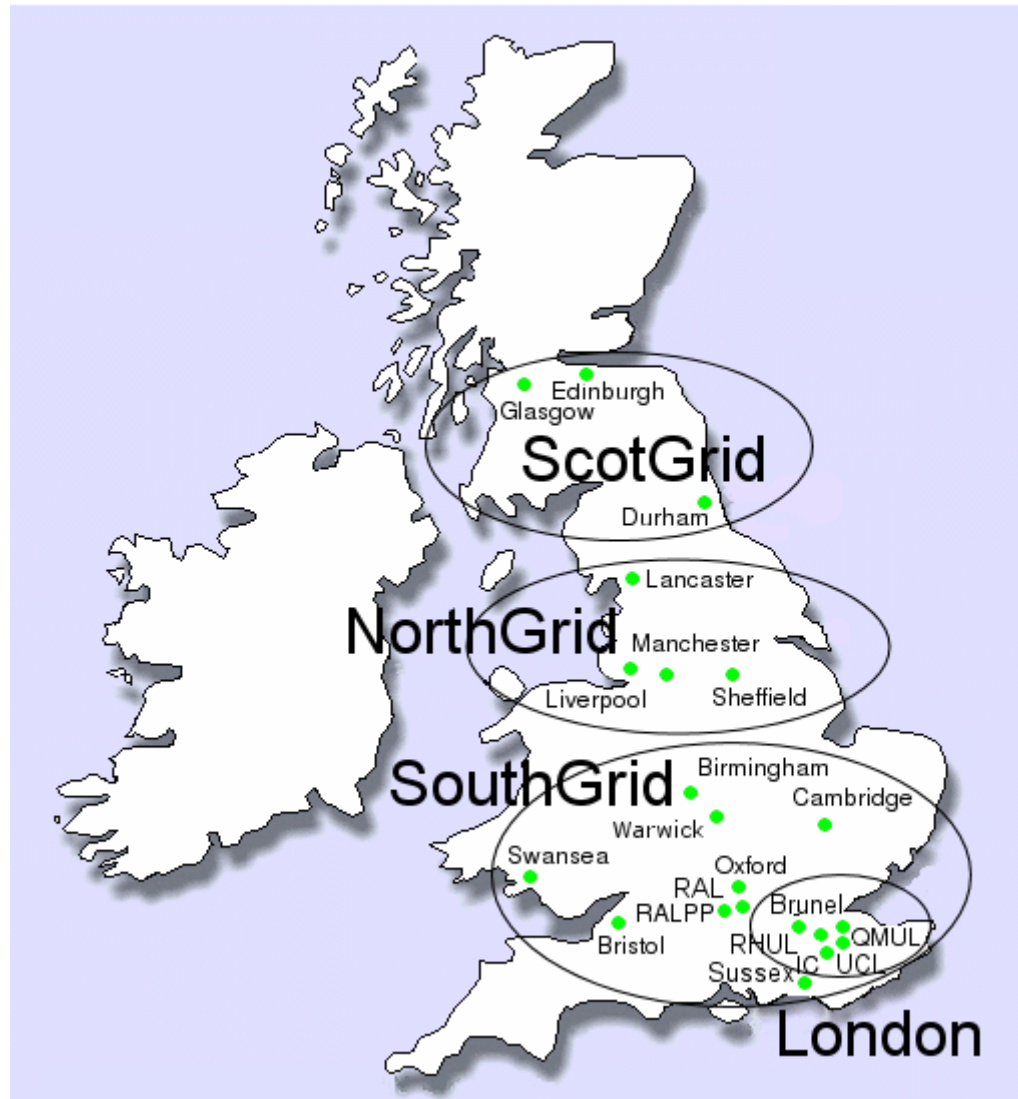


Which Candidate T2 Sites?

- 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



Which T2(s)?





Which T2s cont.





Which Experiments?

- Meetings with LHCb (4th) and ALICE (7th) March
 - ALICE have indicated that they would like to be involved in SC3 from the start...
 - LHCb have a 'free slot' for the production team in summer
- Meetings with ATLAS and CMS will take place asap
 - Proposal that experiments' plans are presented at both March (Lyon) and April (Taipei) SC meetings
- LCG Storage Management workshop 5-7 April
 - Will focus on concrete needs for Service Challenges
 - File transfer s/w, storage management s/w, experiment-specific solutions (hopefully) built on top of standards
 - ☺ Stephen Burke has been invited as "loose cannon"



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
- Adding the T2s and integrating the experiments' software in the SCs will be a massive effort!
- **Due to the well-established GridPP infrastructure, the UK would be an excellent place to pioneer this work**