LHC Computing Model Summary: Update +Tier2 Plans

Jamie Shiers, CERN-IT-GD

28 February 2005

Summary of Key Questions / Issues

- Zero suppression, HLT efficiency factors in raw event sizes
- Variation in TAG sizes (1-10KB)

[Stick with Computing Model Numbers on these for now]

[Revisit TAG issues when we get to end-user analysis]

- **Heavy ion model**
- **Better understanding of T2 (and T1-T2) issues**

Overview of Heavy Ion running

Experiment	SIM	SIMESD	RAW	Trigger	RECO	AOD	TAG
ALICE	300MB	2.1MB	12.5MB	100Hz	2.5MB	250KB	10KB
ATLAS			5MB	50Hz			
CMS			7MB	50Hz	1MB	200KB	TBD
LHCb	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Heavy Ion Questions / Uncertainties

- Heavy Ion computing models less well established than for pp running
- I am concerned about model for 1st/2nd/3rd pass reconstruction and data distribution
- ***"We therefore require that these data (Pb-Pb) are reconstructed at the CERN T0 and exported over a four-month period after data taking. This should leave enough time for a second and third reconstruction pass at the Tier 1's"*** **(ALICE)**
- Heavy Ion model has major impact on those Tier1's supporting these experiments
  - All bar LHCb!
- Critical to clarify these issues as soon as possible…

Heavy Ion Model Revisited

- [pp data taking and inline reconstruction as before (7 months)]
- No 'first pass' reconstruction during Heavy Ion data taking (1 month)
- **Full first pass reconstruction completed at least 6 months prior to next year's Heavy Ion data taking**
  - 4 months shutdown + 2 month overlap with pp run
- **Data distribution during 4 months**
  - + overlap for reconstructed data generated during pp run
- **2nd and 3rd pass reconstruction overlaps with pp run**
  - Should not overlap with next year's Heavy Ion run
- Network requirement:
  - Raw Data distributed over 4 (+1) month period
- CPU requirement:
  - overlap of Heavy Ion reconstruction passes with pp needs

Tier 2 issues

Roles of T0 / T1 / T2

Plans for adding Tier 2 sites to LCG Service Challenges

Services offered by / required by T2s

Network issues

Summary of Tier0/1/2 Roles

- Tier0 (CERN): safe keeping of RAW data (first copy); first pass reconstruction, distribution of RAW data and reconstruction output to Tier1; reprocessing of data during LHC down-times;

- Tier1: safe keeping of a proportional share of RAW and reconstructed data; large scale reprocessing and safe keeping of corresponding output; distribution of data products to Tier2s and safe keeping of a share of simulated data produced at these Tier2s;

- Tier2: Handling analysis requirements and proportional share of simulated event production and reconstruction.

N.B. there are differences in roles by experiment
Essential to test using complete production chain of each!

Tier2 Plans

- SC3 should include a couple of T2s
- SC4 should complete with essentially all T2s on board
- How many? 50 – 100? [Draft compilation – Kors]
- Cannot use 'T1 model' for adding these
- Suggestion:
  - Work through bodies such as GridPP and INFN
  - Use this experience to provide guidance for adding others
  - Use HEPiX, regional / national events and workshops

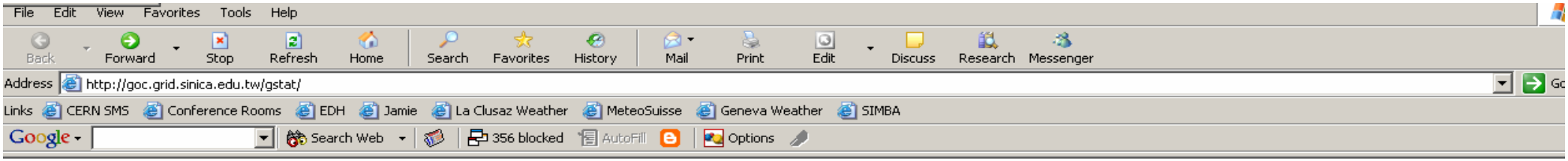
➢ **Do not leave until last minute!**

	ALICE	ATLAS	CMS	LHCb
Parameters:				
Number of Tier-1s	4	6	6	5
Number of Tier-2s	20	24	25	15

T2 Plans

- Discussions at GridPP T2 board meeting March 2nd
- Session(s) at / around HEPiX May 9 – 13 in FZK
- INFN T2 workshop foreseen end-May in Bari
- T2-fest at Fall HEPiX (Sep 19+ at SLAC)
- For other T2s, organise workshop(s) at CERN
  - October 2005?

➢ **Clarify roles of T2s, requirements in terms of services, services required from T1, network topology issues etc.**



GIIS Monitor 15:51:12 01/25/05 GMT

home alert table service regional rgma access service metrics misc pages help

ALBERTA-LCG2 il	BEIJING-LCG2 ok	BG01-IPP ok	BHAM-LCG2 ok	BITLab-LCG ct	BNL-LCG ct
BUDAPEST ok	CARLETONU-LCG2 il	CAVENDISH-LCG2 il	CCIN2P3-LCG2 ct	CEA-DAPNIA-SACLAY na	CERN-LCG2 ok
CESGA-EGEE is	CGG-LCG2 ok	CIEMAT-LCG2 ok	CNAF-LCG2 wl	CNB-LCG2 il	CSCS-LCG2 na
csTCDie is	CY01-LCG2 is	CYFRONET-LCG2 ok	DESYPRO ct	Durham is	ekplcg2 ok
FNAL-LCG2 ct	FZK-LCG2 ok	GR-01-AUTH ok	GSI-LCG2 is	HEPHY-UIBK ok	HG-01-GRNET ok
HPC2N ok	HPTC-LCG2 il	HU-BERLIN na	IC-LCG2 ok	IFCA-LCG2 il	IFIC-LCG2 ok
IISAS-Bratislava il	IN2P3-CPPM na	IN2P3-LAL ok	IN2P3-LAPP na	IN2P3-LPC il	INFN-BARI na
INFN-BOLOGNA na	INFN-BOLOGNA-CMS na	INFN-CAGLIARI na	INFN-CATANIA il	INFN-CNAF il	INFN-FERRARA na
INFN-FRASCATI ok	INFN-LNL-LCG il	INFN-MILANO-LCG2 ok	INFN-NAPOLI na	INFN-NAPOLI-ATLAS il	INFN-PADOVA il
INFN-PERUGIA na	INFN-PISA na	INFN-ROMA1 ok	INFN-ROMA1-VIRGO na	INFN-ROMA2 na	INFN-TORINO-LCG2 ct
INTA-CAB ok	IPSL-IPGP-LCG2 ok	ITEP ok	JINR-LCG2 il	Lancs-LCG2 ct	LIP-LCG2 ok
LivHEP-LCG2 il	ManHEP-LCG2 ok	NCP-LCG2 il	NIKHEF.NL ok	NSC ct	OXFORD-01-LCG2 ok
PIC-LCG2 ok	POZNAN-LCG2 ct	Prague-CESNET ok	Prague-LCG2 il	QMUL-eScience ok	RAL-LCG2 ok
RALPP-LCG ok	RHUL-LCG2 il	ROGRID-ICI il	RU-Moscow-KIAM-LCG2 ct	ru-Moscow-SINP-LCG2 il	RU-Protvino-IHEP ok
ru-PSN-LCG2 il	RWTH-Aachen-Test il	SARA-LCG2 ok	SCAI ok	ScotGRID-Edinburgh ok	scotgrid-gla il
SHEFFIELD-LCG2 na	Taiwan-IPAS-LCG2 ok	Taiwan-LCG2 ok	TAU-LCG2 ok	TIFR-LCG2 na	TOKYO-LCG2 il
TORONTO-LCG2 ok	TRIUMF-GC-LCG2 is	TRIUMF-LCG2 ok	TW-NCUHEP na	UAM-LCG2 ct	UB-LCG2 ok
UCL-CCC ct	UCL-HEP na	UIOWA-LCG2 ok	Umontreal-LCG2 ok	Uni-Wuppertal ok	UPV-GRyCAP il
USC-LCG2 ok	WARSAW-LCG2 is	WEIZMANN-LCG2 il			

	sites	countries	totalCPU	freeCPU	runJob	waitJob	seAvail TB	seUsed TB	maxCPU	avgCPU
Total	111	31	9884	5074	2095	789	4784.84	1930.36	19600	9120

Code	Color	Alert Severity
0	#FFFFFF	_
10	#FAFFFA	OK
20	#EEFFFF	INFO
30	#EDEFFF	NOTE
40	#FFFFCC	WARN

http://goc.grid.sinica.edu.tw/gstat/

T2 Requests

- So far, both Prague and "Russian Tier2 cluster" have made requests to join Service Challenges
- Foresee T2s from other regions
  - Interest from Italy, UK, US, …
- It is clear that we need a plan…

Backup slides

- Copy of some of the main slides from the January Computing Model summary (RAL Service Challenge meeting)

http://agenda.cern.ch/fullAgenda.php?ida=a045745

Goals

- Present key features of LHC experiments' Computing Models in a consistent manner
- High-light the commonality
- Emphasize the key differences
- Define these 'parameters' in a central place (LCG web)
  - Update with change-log as required
- Use these parameters as input to requirements for Service Challenges
- To enable partners (T0/T1 sites, experiments, network providers) to have a clear understanding of what is required of them

➢ **Define precise terms and 'factors'**

Where do these numbers come from?

- Based on Computing Model presentations given to GDB in December 2004 and to T0/T1 networking meeting in January 2005
- Documents are those publicly available for January LHCC review
  - Official website is protected

➢ **Some details may change but the overall conclusions do not!**

- Part of plan is to understand how sensitive overall model is to variations in key parameters
- Iteration with experiments is on-going
  - i.e. I have tried to clarify any questions that I have had

➢ **Any mis-representation or mis-interpretation is entirely my responsibility**

- Sanity check: compare with numbers from MoU Task Force

Nominal	These are the raw figures produced by multiplying e.g. event size x trigger rate.
Headroom	A factor of 1.5 that is applied to cater for peak rates.
Efficiency	A factor of 2 to ensure networks run at less than 50% load.
Recovery	A factor of 2 to ensure that backlogs can be cleared within 24 - 48 hours and to allow the load from a failed Tier1 to be switched over to others.
Total Requirement	**A factor of 6 must be applied to the nominal values to obtain the bandwidth that must be provisioned.** **Arguably this is an over-estimate, as "Recovery" and "Peak load" conditions are presumably relatively infrequent, and can also be smoothed out using appropriately sized transfer buffers.** **But as there may be under-estimates elsewhere…**

All numbers presented will be
nominal unless explicitly specified

High Level Overview

- All experiments assume a Grid-based solution – i.e. LCG
- Computing Models can be viewed as that proposed by MONARC with Grid extensions
- Largely similar functions for Tier0 / Tier1 / Tier2
- ...but there are important differences…
- First focus on commonality
- ➢ **Differences stress absolute necessity for including all main experiment Use Cases into (later, but not much) Service Challenges**
- 'We' cannot run experiments' offline frameworks…
- Requires significant commitment from them… now…
- Have started discussions with experiments on this basis.

Contacts with Experiments

- Using names from CM documents:
- ALICE: F. Carminati, Y. Schutz
- ATLAS: R. Jones (+ others)
- CMS: C. Grandi, D. Stickland, L. Taylor
- LHCb: Nick Brook
- Also contacting production teams (see later)

Summary of Tier0/1/2 Roles

- Tier0 (CERN): safe keeping of RAW data (first copy); first pass reconstruction, distribution of RAW data and reconstruction output to Tier1; reprocessing of data during LHC down-times;

- Tier1: safe keeping of a proportional share of RAW and reconstructed data; large scale reprocessing and safe keeping of corresponding output; distribution of data products to Tier2s and safe keeping of a share of simulated data produced at these Tier2s;

- Tier2: Handling analysis requirements and proportional share of simulated event production and reconstruction.

N.B. there are differences in roles by experiment
Essential to test using complete production chain of each!

Tier-1 Centres (January 2004)

				ALICE	ATLAS	CMS	LHCb	
1	GridKa	Karlsruhe	Germany	X	X	X	X	4
2	CCIN2P3	Lyon	France	X	X	X	X	4
3	CNAF	Bologna	Italy	X	X	X	X	4
4	NIKHEF/SARA	Amsterdam	Netherlands	X	X		X	3
5	Nordic	Distributed	Dk, No, Fi, Se	X	X			1
6	PIC	Barcelona	Spain		X	X	X	3
7	RAL	Didcot	UK	X	X	X	X	4
8	Triumf	Vancouver	Canada		X			1
9	BNL	Brookhaven	US		X			1
10	FNAL	Batavia, Ill.	US			X		1
11	ASCC	Taipei	Taiwan		X	X		2
				6	10	7	6	

x – announced at January GDB

Overview of pp running

Experiment	SIM	SIMESD	RAW	Trigger	RECO	AOD	TAG
ALICE	400KB	40KB	1MB	100Hz	200KB	50KB	10KB
ATLAS	2MB	500KB	1.6MB	200Hz	500KB	100KB	1KB
CMS	2MB	400KB	1.5MB	150Hz	250KB	50KB	10KB
LHCb		400KB	25KB	2KHz	75KB	25KB	1KB

pp questions / uncertainties

- Trigger rates essentially independent of luminosity
  - Explicitly stated in both ATLAS and CMS CM docs
- Uncertainty (at least in my mind) on issues such as zero suppression, compaction etc of raw data sizes
  - Discussion of these factors in CMS CM doc p22:
- RAW data size ~300kB (Estimated from MC)
  - Multiplicative factors drawn from CDF experience
    - MC Underestimation factor 1.6
    - HLT Inflation of RAW Data, factor 1.25
    - Startup, thresholds, zero suppression,…. Factor 2.5
  - Real initial event size more like **1.5MB**
    - Could be anywhere between 1 and 2 MB
  - Hard to deduce when the even size will fall and how that will be compensated by increasing Luminosity

➢ i.e. total factor = 5 for CMS raw data

- N.B. must consider not only Data Type (e.g. result of Reconstruction) but also how it is used
  - e.g. compare how Data Types are used in LHCb compared to CMS
- All this must be plugged into the meta-model!

Overview of Heavy Ion running

Experiment	SIM	SIMESD	RAW	Trigger	RECO	AOD	TAG
ALICE	300MB	2.1MB	12.5MB	100Hz	2.5MB	250KB	10KB
ATLAS			5MB	50Hz			
CMS			7MB	50Hz	1MB	200KB	TBD
LHCb	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Heavy Ion Questions / Uncertainties

- Heavy Ion computing models less well established than for pp running
- I am concerned about model for 1st/2nd/3rd pass reconstruction and data distribution
- ***"We therefore require that these data (Pb-Pb) are reconstructed at the CERN T0 and exported over a four-month period after data taking. This should leave enough time for a second and third reconstruction pass at the Tier 1's"*** **(ALICE)**
- Heavy Ion model has major impact on those Tier1's supporting these experiments
  - All bar LHCb!
- Critical to clarify these issues as soon as possible…

Data Rates from MoU Task Force

MB/Sec	RAL	FNAL	BNL	FZK	IN2P3	CNAF	PIC	T0 Total
ATLAS	106.87	0.00	173.53	106.87	106.87	106.87	106.87	707.87
CMS	69.29	69.29	0.00	69.29	69.29	69.29	69.29	415.71
ALICE	0.00	0.00	0.00	135.21	135.21	135.21	0.00	405.63
LHCb	6.33	0.00	0.00	6.33	6.33	6.33	6.33	31.67
T1 Totals MB/sec	182.49	69.29	173.53	317.69	317.69	317.69	182.49	1560.87
T1 Totals Gb/sec	1.46	0.55	1.39	2.54	2.54	2.54	1.46	12.49
Estimated T1 Bandwidth Needed								
(Totals * 1.5*(headroom))*2(capacity)*	4.38	1.66	4.16	7.62	7.62	7.62	4.38	37.46
Assumed Bandwidth Provisioned	**10.00**	**10.00**	**10.00**	**10.00**	**10.00**	**10.00**	**10.00**	**70.00**

Spreadsheet used to do this calculation will be on Web.

Table is in

http://cern.ch/LCG/MoU%20meeting%20March%2010/Report_to_the_MoU_Task_Force.doc

Data Rates using CM Numbers

Steps:

- Take Excel file used to calculate MoU numbers
- Change one by one the Data Sizes as per latest CM docs
- See how overall network requirements change
- ➢ Need also to confirm that model correctly reflects latest thinking
- ➢ And understand how sensitive the calculations are to e.g. changes in RAW event size, # of Tier1s, roles of specific Tier1s etc.
- This will take several iterations but will need to converge relatively rapidly to satisfy request from 'Networkers' (see below)
- [Did want to do this 'live' now, but think it makes sense for LHCC review to be made public – the models are still changing!]

Base Requirements for T1s

- **Provisioned bandwidth comes in units of 10Gbits/sec although this is an evolving parameter**
  - *From* Reply to Questions from Computing MoU Task Force…
- Since then, some parameters of the Computing Models have changed
- Given the above quantisation, relatively insensitive to small-ish changes
- Important to understand implications of multiple-10Gbit links, particularly for sites with Heavy Ion programme

- **For now, need plan for 10Gbit links to all Tier1s**