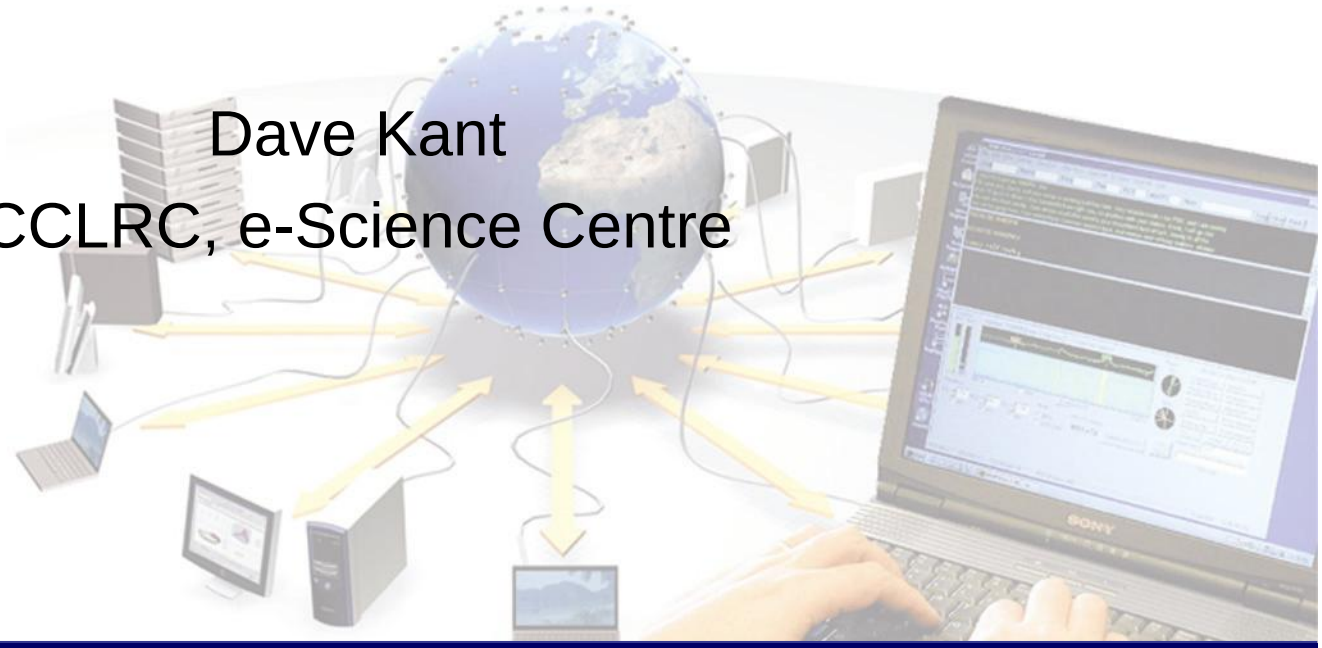
Enabling Grids for E-science in Europe

Accounting in LCG/EGEE
Can We Gauge Grid Usage via RBs?

Dave Kant
CCLRC, e-Science Centre



APEL in LCG/EGEE

Enabling Grids for E-science in Europe

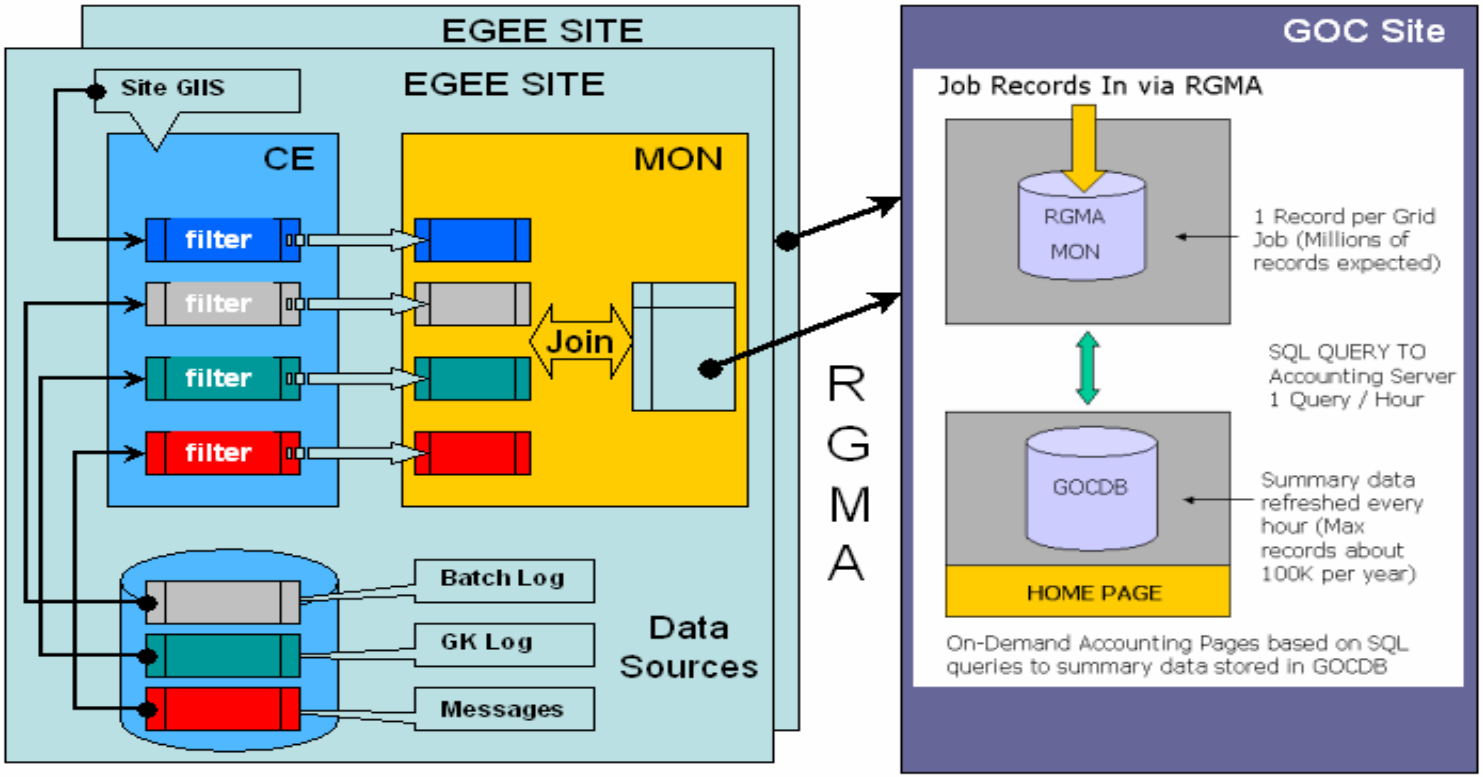
- Overview
- Accounting in LCG/EGEE
- Data Consolidation
- Determining Grid Usage at the RB Level.

APEL, Job Accounting Flow Diagram

[1] Build Job Accounting Records at site.

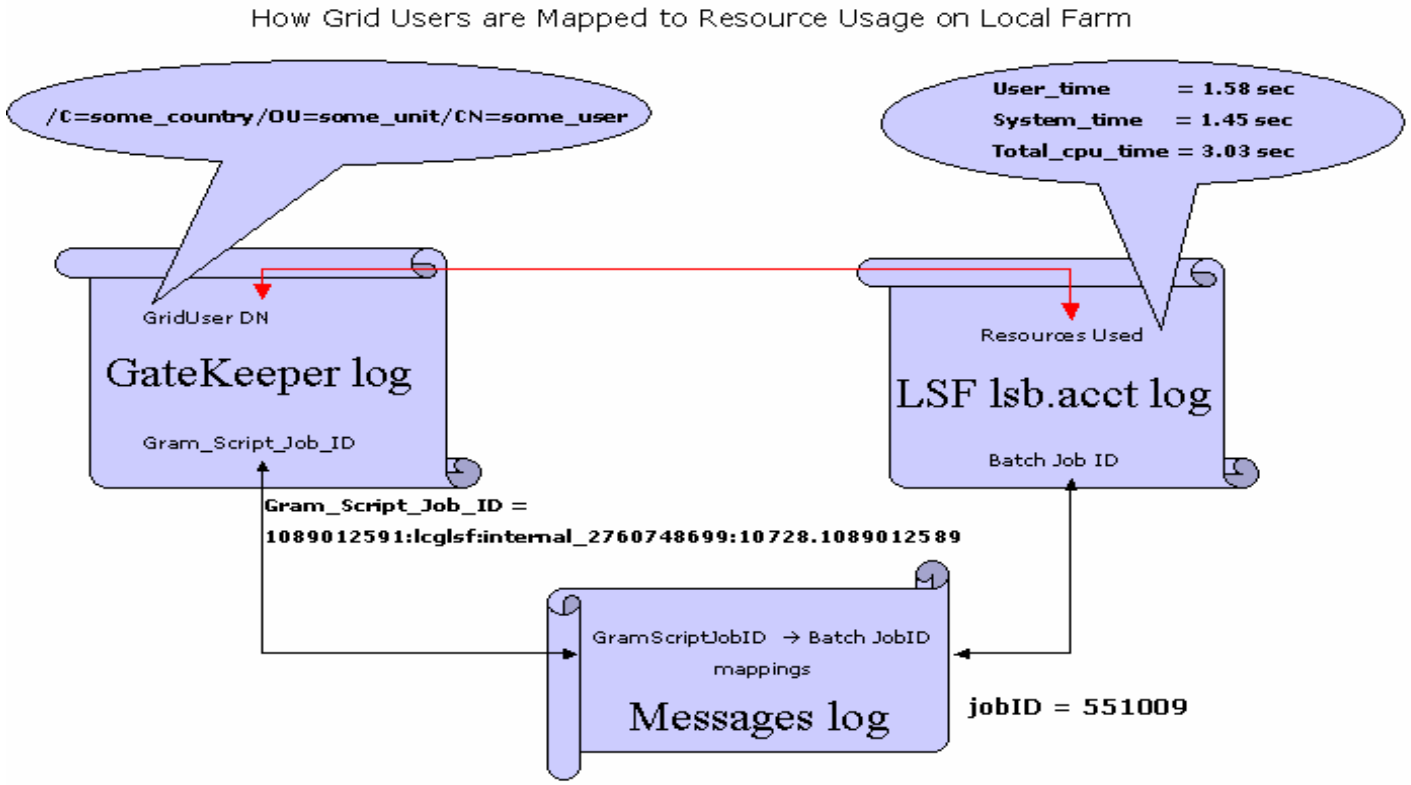
[2] Send Job Records to a central repository

[3] Data Aggregation



Accounting for Grid Jobs

- Build Job Records at Site
- APEL mapping grid users to the resource usage on local farms



Consolidation of Data

GOC

Job Records In via RGMA

RGMA

MON

1 Record per Grid Job (Millions of records expected)

SQL QUERY TO
Accounting Server
1 Query / Hour

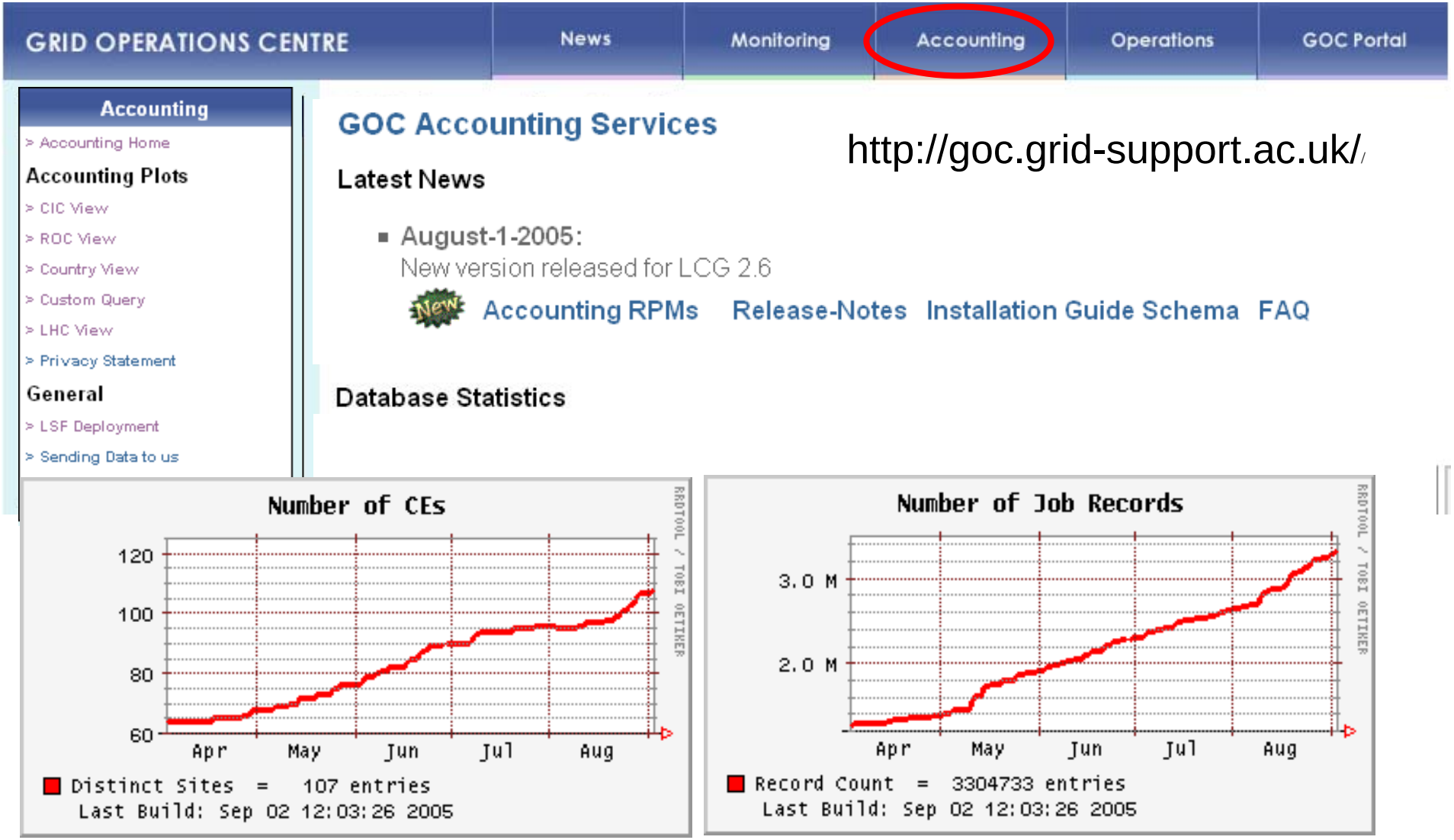
User queries

Graphs

Home Page

Summary data refreshed every hour (Max records about 100K per year)

On-Demand Accounting Pages based on SQL queries to summary data



http://goc.grid-support.ac.uk/

127 Sites publishing data (Oct 10 2005)

3.9 Million Job records

~ 100K records per week (period June – Oct 2005)

Batch Support in APEL

Enabling Grids for E-science in Europe

Currently Available in LCG 2.6

- OpenPBS, Torque, PBSPro and Vanilla PBS
  - ~90% Sites in LCG/EGEE
- Load Share Facility (Versions 5 and 6)
  - CERN, Italy

Work in Progress (LCG 2.7)

- Condor
  - Canada, Cambridge
- Sun Grid Engine
  - Imperial College

Demos of Accounting Aggregation

Enabling Grids for E-science in Europe

Global views of resource consumption.

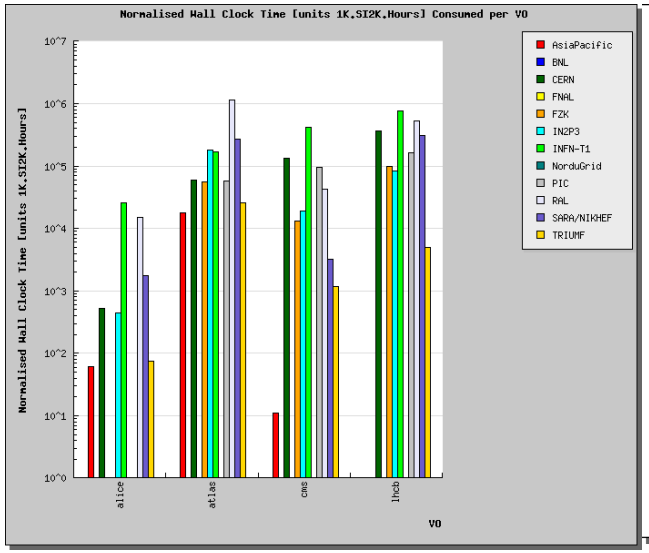
- LCG View
- **http://goc.grid-support.ac.uk/gridsite/accounting/tree/treeview.php**
  - Shows Aggregation for each LHC VO
    - *Requirements driven by CRRB / Kors Bos*
    - *Tier-1 and Country entry points*
    - *All data normalised in units of 1000 . SI2000 . Hour*
    - *Tabular Summaries per Tier1/ Country*
- GridPP View
- **http://goc.grid-support.ac.uk/gridsite/accounting/tree/gridppview.php**
  - Shows Aggregation for Rocs and Organisations
- CESGA View
- http://www.egee.cesga.es/EGEE-SA1-SWE/accounting/reports/
  - Prototype for EGEE View of resource consumption?

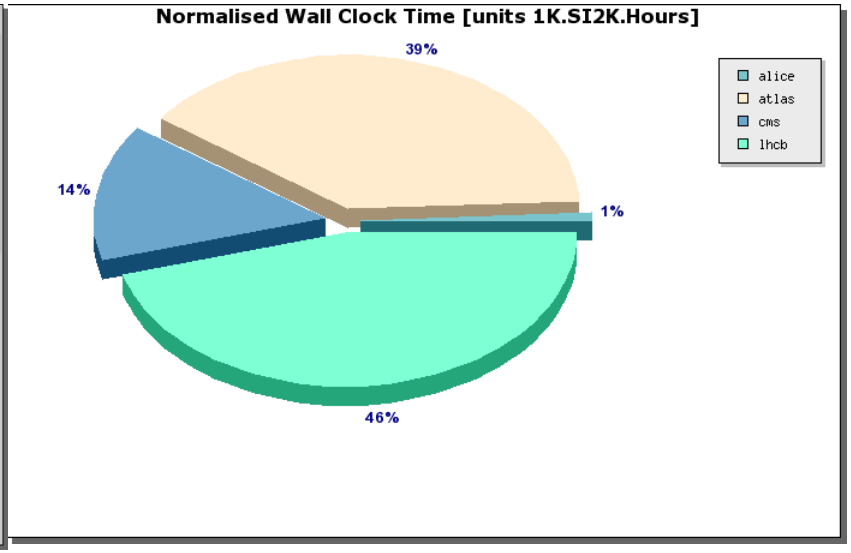
LHC View: Data Aggregation For VOs per Tier1, per Country

LHC Hierarchical Tree

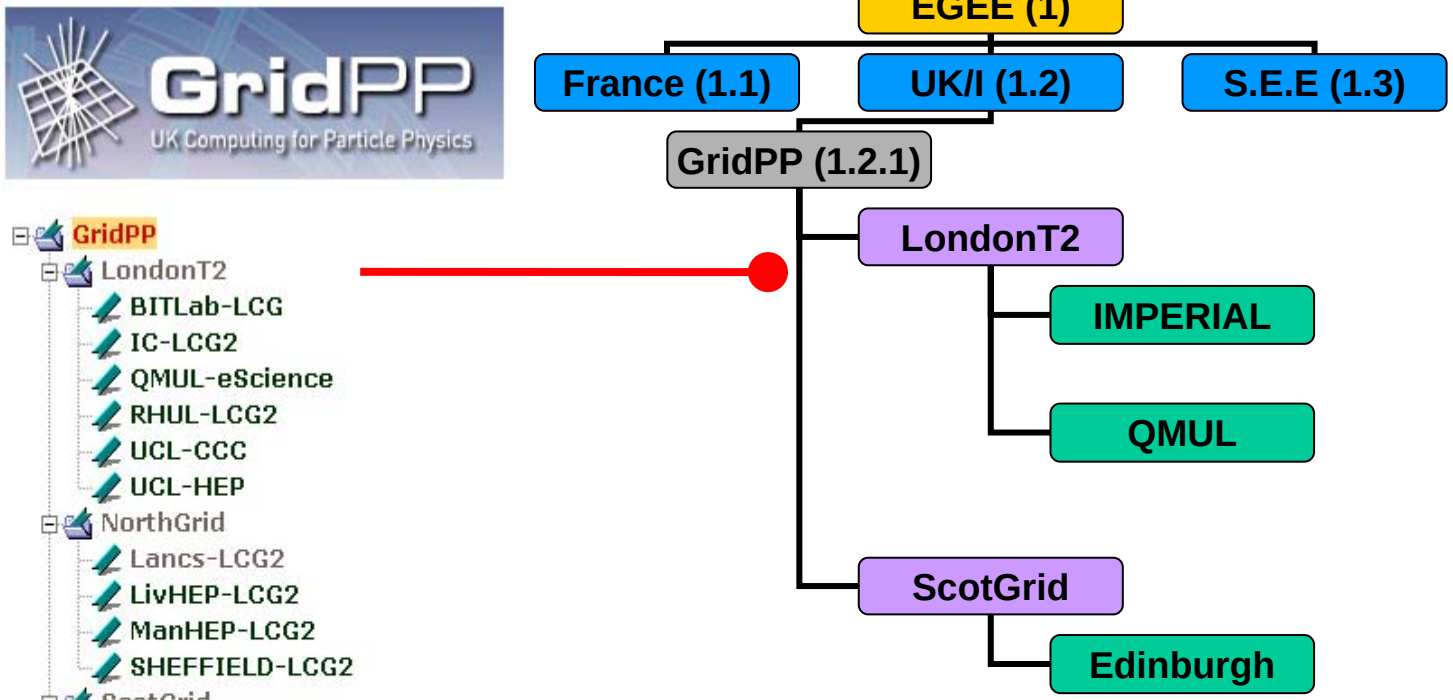
- Tier1
  - AsiaPacific
  - BNL
  - CERN
  - CNAF
  - FNAL
  - FZK
  - IN2P3
  - NorduGrid
  - PIC
  - RAL
  - SARA/NIKHEF
  - TRIUMF
- Regions
- Countries

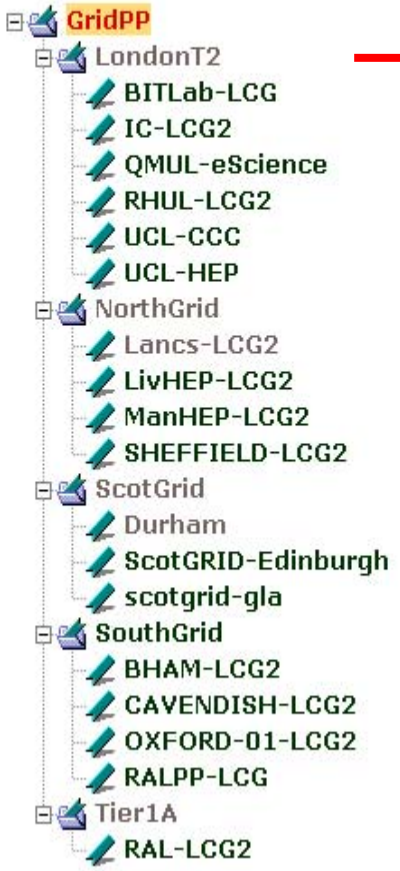
Normalised Wall Clock Time [units 1K.SI2K.Hours]					
Tier-1	Site	alice	atlas	cms	lhcb
AsiaPacific		62	17643	11	0
BNL		0	0	0	0
CERN		528	59438	134900	365990
FNAL		0	0	0	0
FZK		0	55167	13360	100432
IN2P3		439	183735	18787	83212
INFN-T1		25396	170752	424697	765625
NorduGrid		0	0	0	0
PIC		0	56900	94447	161707
RAL		15174	1165126	42936	535577
SARA/NIKHEF		1741	266563	3206	313196
TRIUMF		76	25401	1175	5035
Click here for a csv dump of this table					





Organisation Structures





Consolidated views of resource usage

- Tier1 and Tier2 within an EGEE ROC

Aggregation of Data for GridPP Community

CPU Quantity - NormalisedCPUTime | Start year - 2005 | Start month - 1 | End year - 2005 | End month - 9

Refresh

- GridPP
  - LondonT2
    - BITLab-LCG
    - IC-LCG2
    - QMUL-eScience
    - RHUL-LCG2
    - UCL-CCC
    - UCL-HEP
  - NorthGrid
    - Lancs-LCG2
    - LivHEP-LCG2
    - ManHEP-LCG2
    - SHEFFIELD-LCG2
  - ScotGrid
    - Durham
    - ScotGRID-Edinburgh
    - scotgrid-gla
  - SouthGrid
    - BHAM-LCG2
    - CAVENDISH-LCG2
    - OXFORD-01-LCG2
    - RALPP-LCG
  - Tier1A
    - RAL-LCG2

Normalised Wall Clock Time [units 1K.SI2K.Hours]

Region	Site	alice	atlas	babar	biomed	cdf	cms	dteam	dzero	hone	ilc	lhcb	pheno	sixt	zeus
	BHAM-LCG2	16	79	767	9809	0	1241	377	0	2224	0	10691	0	0	0
	BITLab-LCG	0	50	0	0	0	61	642	0	0	0	277	0	0	0
	BRISTOL-PP-LCG	0	12	0	0	0	2	8	0	0	0	83	0	0	0
	CAVENDISH-LCG2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Durham	0	42	0	0	0	14	40	0	0	0	2458	0	0	0
	hp-bristol	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	IC-LCG2	20	4071	0	7	0	2131	1246	0	0	0	17406	0	2	72
	Lancs-LCG2	0	29051	282	24010	0	9047	379	0	0	0	199361	0	0	0
	LivHEP-LCG2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ManHEP-LCG2	2	248	5442	9046	0	18	124	823	0	0	876	0	0	0
	OXFORD-01-LCG2	0	18126	851	22171	0	4813	1257	0	0	0	35166	0	0	9088
	QMUL-eScience	2	71411	251	0	0	1095	229	0	0	0	320123	0	0	0
	RAL-LCG2	15174	1047016	105366	50553	0	26403	1990	72793	30934	0	510105	0	0	138532
	RALPP-LCG	6646	26146	17637	5873	0	10751	460	168	9302	0	35677	0	0	11824
	RHUL-LCG2	1137	8219	0	0	0	11594	281	0	0	68	105909	0	0	0
	ScotGRID-Edinburgh	0	662	158	620	0	196	81	0	0	0	1107	0	0	0
	scotgrid-gla	0	1436	0	11382	0	9	18	0	0	0	23	0	0	1154
	SHEFFIELD-LCG2	4306	109446	0	0	0	11093	182	0	0	0	1068	0	0	0
	UCL-CCC	0	32057	0	0	0	24141	161	0	0	0	36319	0	0	9427
	UCL-HEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Summary	alice	atlas	babar	biomed	cdf	cms	dteam	dzero	hone	ilc	lhcb	pheno	sixt	zeus
Summed Usage	27303	1348072	130754	133471	0	102609	7475	73784	42460	68	1276649	0	2	170097
As A Percentage	0.8%	40.7%	3.9%	4%	0%	3.1%	0.2%	2.2%	1.3%	0%	38.5%	0%	0%	5.1%

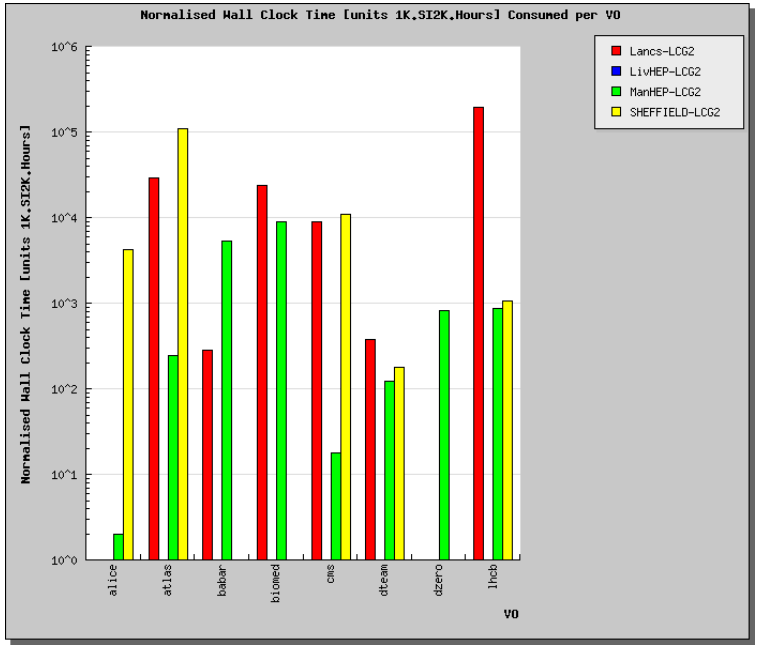
Click here for a csv dump of this table

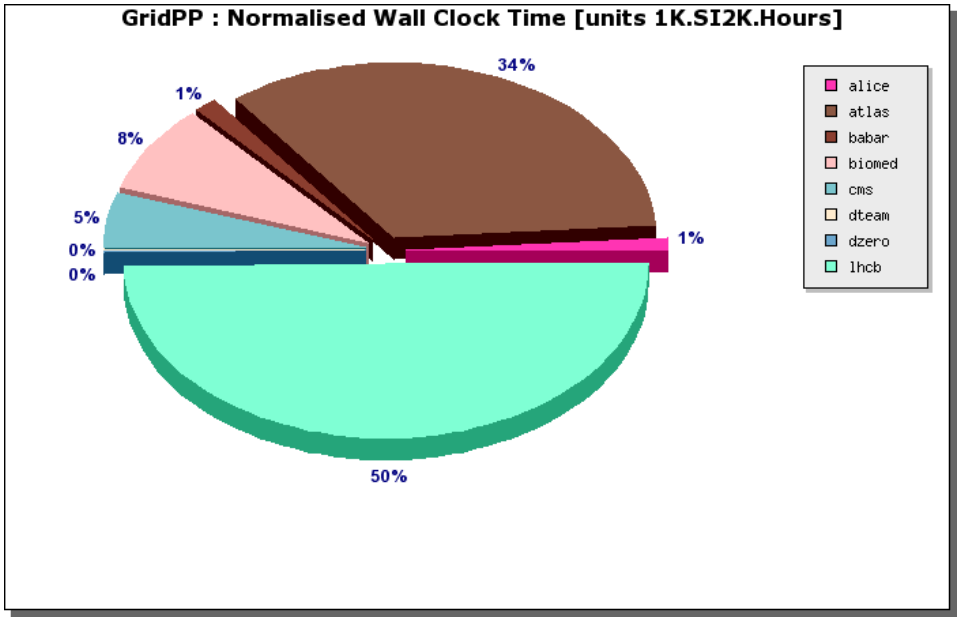
Aggregation of Data for Tier2

Normalised Wall Clock Time [units 1K.SI2K.Hours]									
Region	Site	alice	atlas	babar	biomed	cms	dteam	dzero	lhcb
	Lancs-LCG2	0	29051	282	24010	9047	379	0	199361
	LivHEP-LCG2	0	0	0	0	0	0	0	0
	ManHEP-LCG2	2	248	5442	9046	18	124	823	876
	SHEFFIELD-LCG2	4306	109446	0	0	11093	182	0	1068

Summary	alice	atlas	babar	biomed	cms	dteam	dzero	lhcb
Summed Usage	4308	138745	5724	33056	20158	685	823	201305
As A Percentage	1.1%	34.3%	1.4%	8.2%	5%	0.2%	0.2%	49.7%

Click here for a csv dump of this table

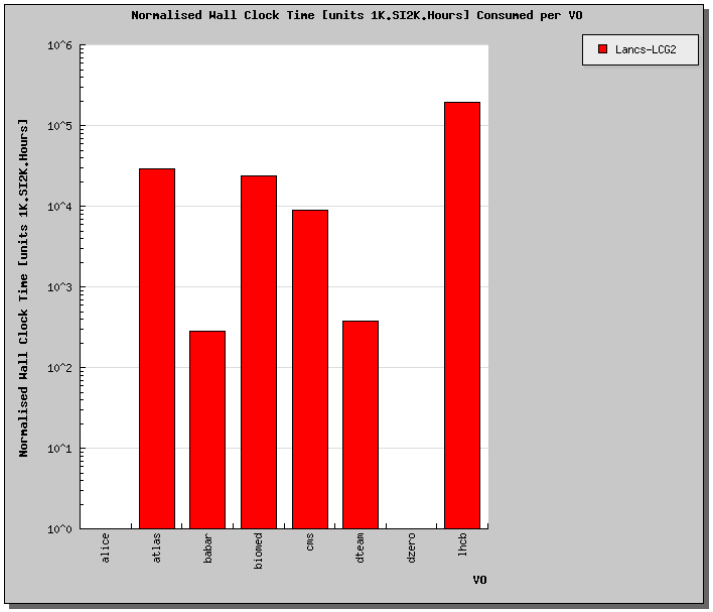




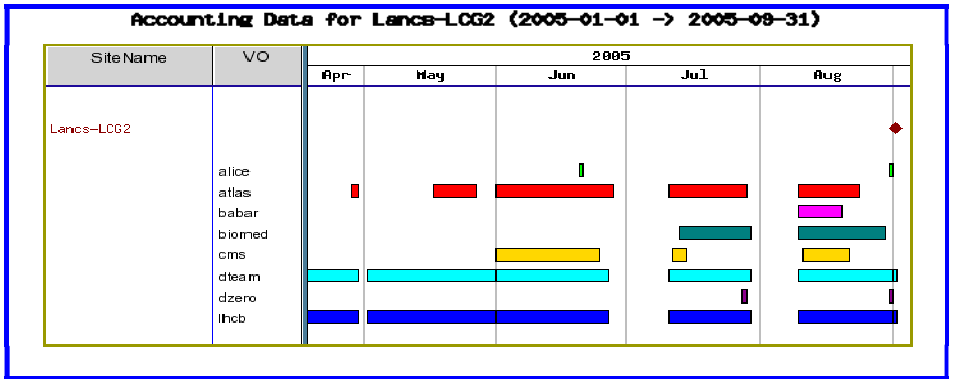
Data Aggregation at Site Level

VO	Njobs	CPU(Hours)	CPU (SI2K.Hours)	WCT(Hours)	WCT (SI2K.Hours)	RecordStart	RecordEnd
alice	1	0	0	0	0	2005-06-20	2005-06-20
alice	1	0	0	0	0	2005-08-31	2005-08-31
atlas	2	0	0	0	0	2005-04-28	2005-04-29
atlas	12	0	0	0	0	2005-05-17	2005-05-26
atlas	805	1	1	25258	29046	2005-06-01	2005-06-27
atlas	129	0	0	3	4	2005-07-11	2005-07-28
atlas	34	0	0	1	1	2005-08-10	2005-08-23
babar	51	243	280	245	282	2005-08-10	2005-08-19
biomed	906	9148	10521	9333	10733	2005-07-13	2005-07-29
biomed	782	11482	13204	11545	13277	2005-08-10	2005-08-29
cms	4	0	0	7	8	2005-06-01	2005-06-24
cms	210	4265	4905	4449	5116	2005-07-12	2005-07-14
cms	103	3331	3831	3411	3923	2005-08-11	2005-08-21
dteam	525	0	0	15	17	2005-04-18	2005-04-29
dteam	1250	1	2	108	125	2005-05-02	2005-05-31
dteam	798	1	1	167	192	2005-06-01	2005-06-26
dteam	173	0	0	21	24	2005-07-11	2005-07-29
dteam	472	1	1	18	21	2005-08-10	2005-08-31
dteam	17	0	0	0	0	2005-09-01	2005-09-01
dzero	1	0	0	0	0	2005-07-28	2005-07-28
dzero	2	0	0	0	0	2005-08-31	2005-08-31
lhcb	1174	3	4	283	325	2005-04-18	2005-04-29
lhcb	7663	77798	89468	85123	97891	2005-05-02	2005-05-31
lhcb	3317	36253	41691	36959	42503	2005-06-01	2005-06-26
lhcb	147	4	5	68	78	2005-07-11	2005-07-29
lhcb	3236	49698	57152	50847	58474	2005-08-10	2005-08-31
lhcb	6	77	88	78	90	2005-09-01	2005-09-01

Breakdown of data per Vo per month showing Njobs, CPUt, WCT, record history



Total CPU Usage per VO



Gantt Chart NB:Gaps across all VOs consistent with scheduled downdowns in GocDB

http://www.egee.cesga.es/EGEE-SA1-SWE/accounting/reports/

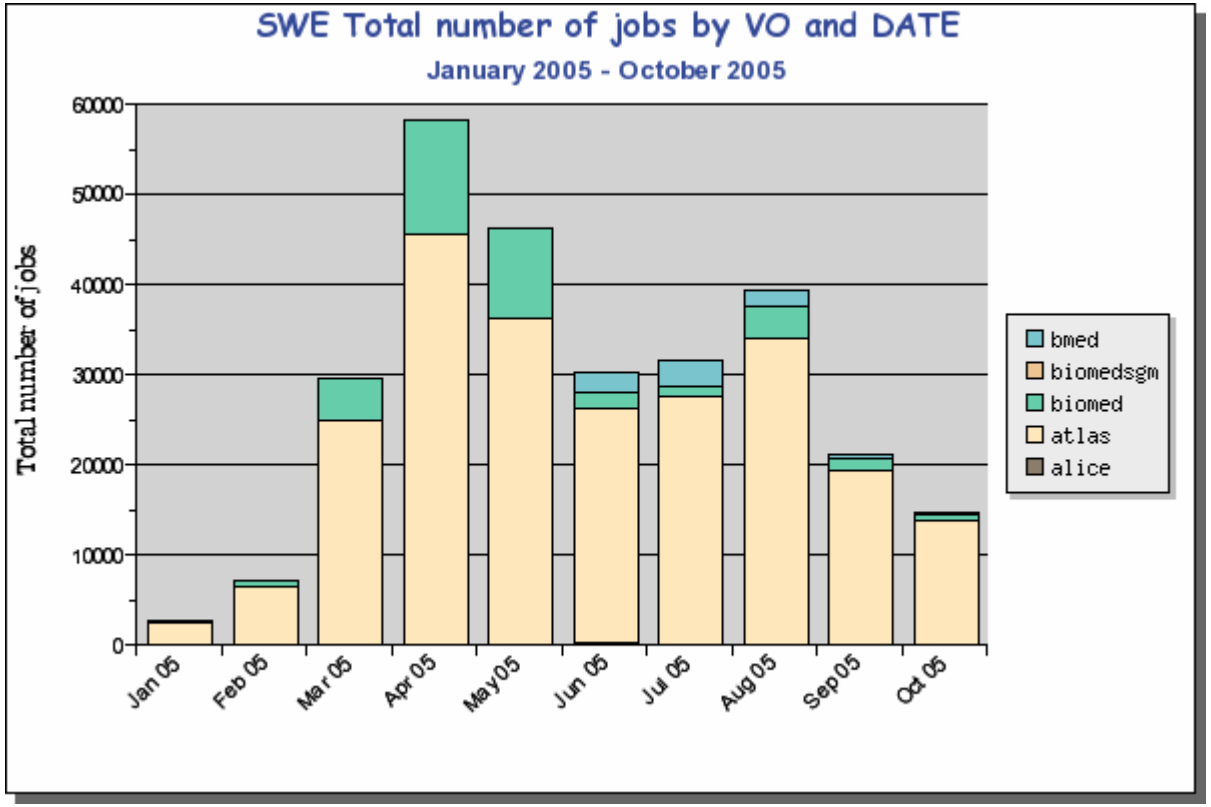
Enabling Grids for E-sciencE

- JOBS
  - Number of jobs
- CPU TIME
- ELAPSED TIME
- MEMORY
- CUSTOM
- REPORTS
- USERS

SWE Total number of jobs run by VO and DATE

January 2005 - October 2005

	Jan 05	Feb 05	Mar 05	Apr 05	May 05	Jun 05	Jul 05	Aug 05	Sep 05	Oct 05	Total
alice	0	0	0	0	0	124	0	1	0	0	125
atlas	2,535	6,500	24,790	45,567	36,280	26,109	27,462	34,064	19,355	13,719	236,381
biomed	177	713	4,794	12,675	9,870	1,715	1,269	3,579	1,223	679	36,694
biomedsgm	1	0	0	0	0	0	0	0	0	0	1
bmed	0	0	0	0	11	2,308	2,932	1,711	571	296	7,829
Total	2,713	7,213	29,584	58,242	46,161	30,256	31,663	39,355	21,149	14,694	281,030



Provides reporting for the SWE federation.

Protected area for users to get usage at dn level.

Determining Usage of the Grid

Enabling Grids for E-science in Europe

- How many different ways are there to submit a job on the grid?
- Broadly speaking, two categories: Indirect and Direct
- INDIRECT: using a resource broker and a JDL
  - Example: via a known “core” RB for SFT Monitoring Jobs
  - Or via a private RB to target national resources
- DIRECT: globus-job-run by specifying the target resource
  - Example: Atlas production run earlier in 2005
    - which required 1 hr cpu time, but took 7 seconds to submit via a RB. Thus, the RBs were a bottleneck.
    - CPU Usage per site for this run available from the Atlas production database.
    - Significant usage via this path: 50% Rome LCG Production

Determining Usage of the Grid

Indirect via RB

RB

L&B

L&B captures usage via state change

DIRECT JOB SUBMISSION

globus_job_run

COMPUTING ELEMENT/ BATCH FARM MANAGER

wn wn wn

Worker Node Pool

APEL / DGAS capture job usage at the end point.

Comparing Numbers: APEL vs JRA2

- Can RB L&B be used to gauge use of the grid?
  - Not all RBs have RGMA publishers installed
  - RBs are not the only gateway to grid resoucres
- Sum over RBs to find total usage for each site and compare with the numbers at the end point.
- Example: SWE (PIC) Tier-1 that has deployed APEL (complete dataset).

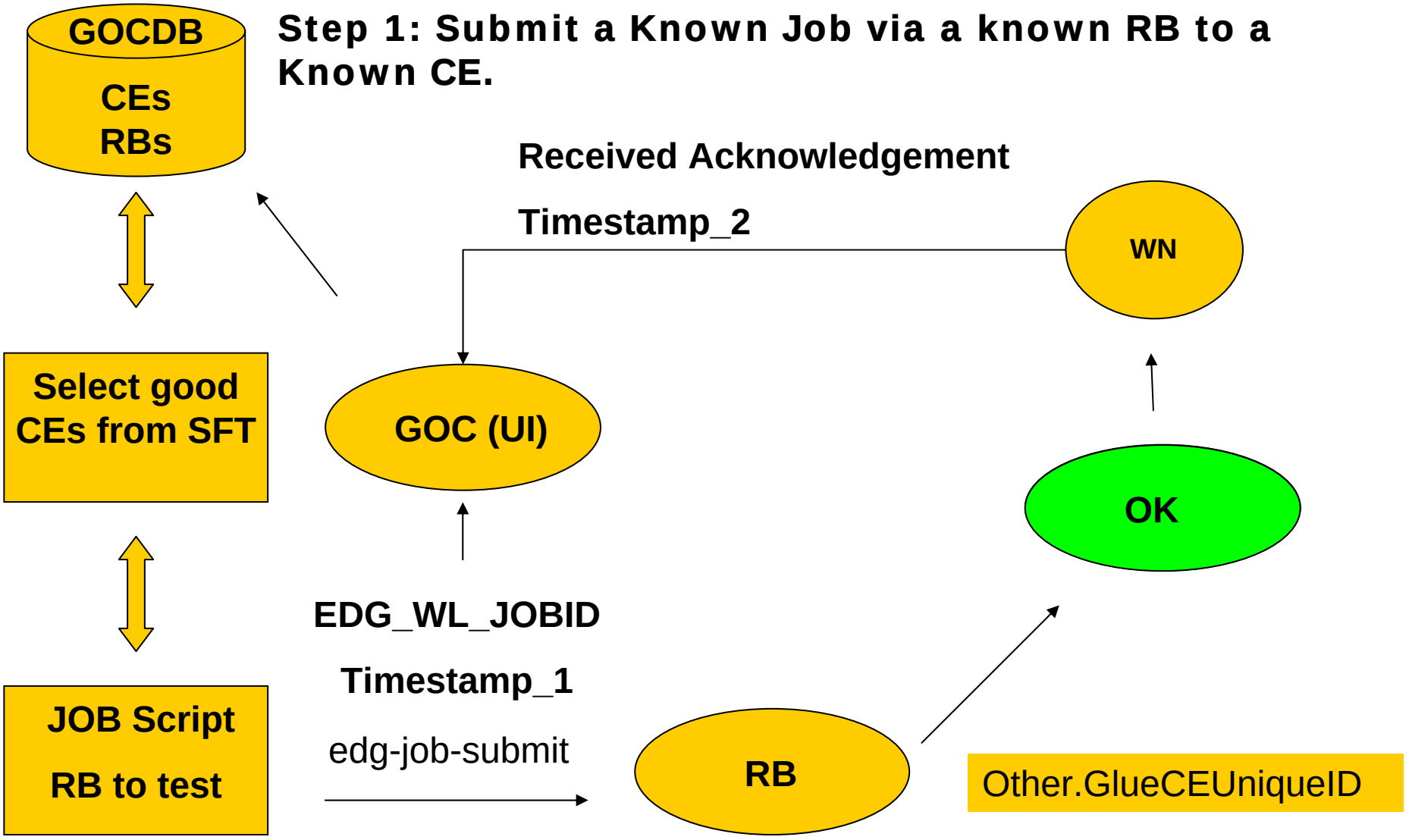
PERIOD	Total (Njobs)	ATLAS	CMS	DTEAM	LHCB	BIOMED
FEB 2005	LB = 20 APEL = 8922	**886**	**2455**	**4076**	**1505**	**0**
AUG 2005	LB = 4327 APEL = 14233	**998**	**4294**	**3352**	**3878**	**1711**

Service Metrics for RB

- Rather than gauge grid usage on a site or VO basis, why not try it on a job by job basis?
- Metric WG wants to monitor availability of RBs using test jobs that have a know finite WCT on the target resoucre.
- Compare APEL information with L&B information for these test jobs.

RB Service Monitoring : Metrics WG

Enabling Grids for E-science in Europe



DK2

Data Flow Diagram

A graphical means of presenting, describing or analyzing a process.

Dave Kant, 3/28/2004

RB L&B Data

- RGMA query of JobStatusRaw table which holds L&B data
- Record States of RB for the specified Job
  - Submitted
  - Waiting
  - Ready
  - Scheduled
  - Running
  - Done
  - Cleared
- Correlate time in running state with the APEL WCT

Alternative Approaches

Enabling Grids for E-science in Europe

- Compare on a job-by-job basis via the WMS job ID
  - DGAS – APEL integration
  - APEL support for Condor
- Any other ways?