Getting the most out of EGEE, OSG and ....

Sanjay Padhi

University of Wisconsin-Madison

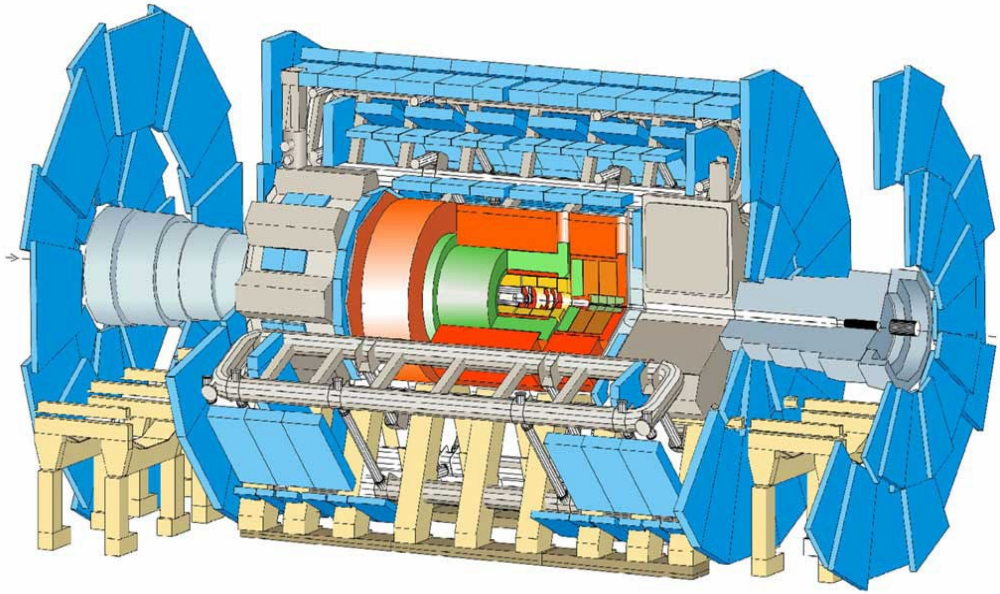
OSG Consortium All Hands Meeting

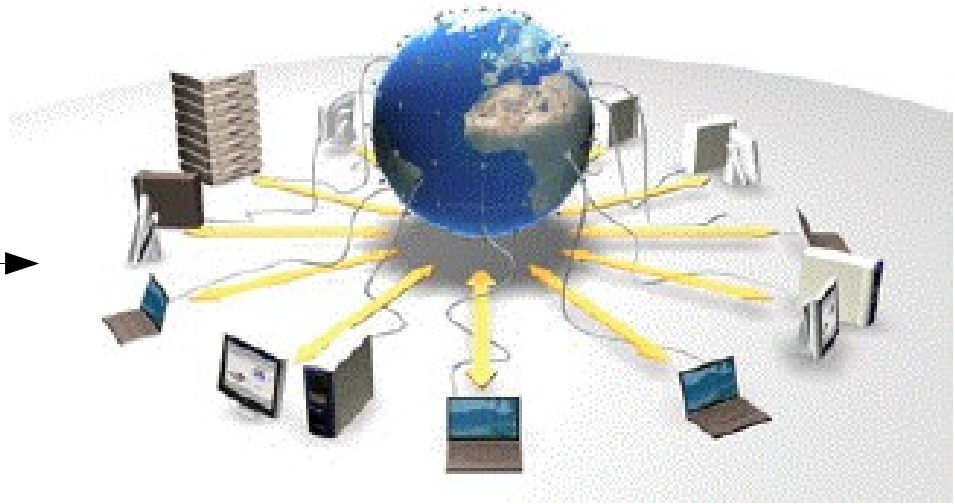
University of California, San Diego 5 – 8 March 2007

Many thanks to R. Walker, Sau Lan Wu, Miron Livny,

the Condor Team & several members of the ATLAS Collaboration

Getting the most out of EGEE, OSG and ....

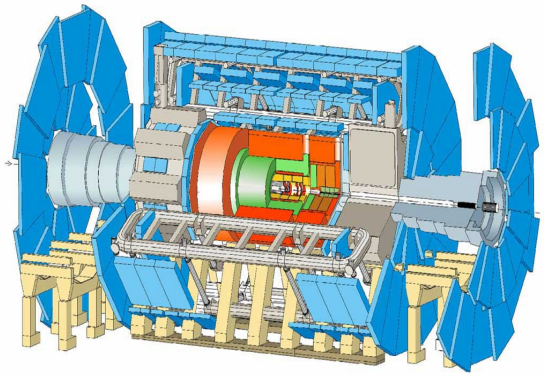


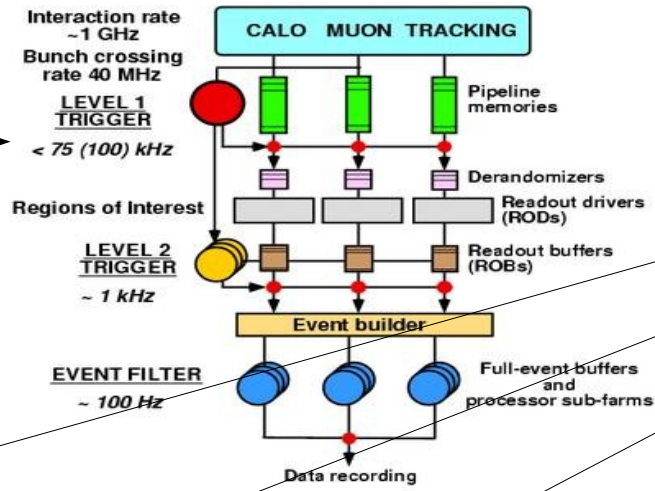


- HEP is evolving :
  - Physics analysis is no longer at the Experimental site.
- Physicists are evolving:
  - Illiteracy in Computing is no longer possible.
- Collaborations are evolving:
  - Computing using local batch systems are just not enough.

<u>Inter-operable Grid Computing is a necessity</u>

Classical LHC Tier Model





Archival and Distribution of RAW DATA

CERN - TIER 0

TIER 1 hosts and provide long term access & archival - subset of RAW DATA

TW

FR

ES

IT

NL

. . .

US

TIER 2s – provide calibration constants, simulation & analysis

Tier2 Center

2 Center

Tier2 Center

Tier 3 Institutes

Tier 3 Insti

Tier 3 Institutes

RAW Size ~ 1.6MB (1600 TB/year)
ESD Size ~ 1.0MB (1000 TB/year)
AOD Size ~ 0.1MB (100 TB/year)
Number of Events (2008) - 1000M

For all this to work:

Need a strong relationship between scientists and "the Grid Community" ....

A Physicist's Need ....

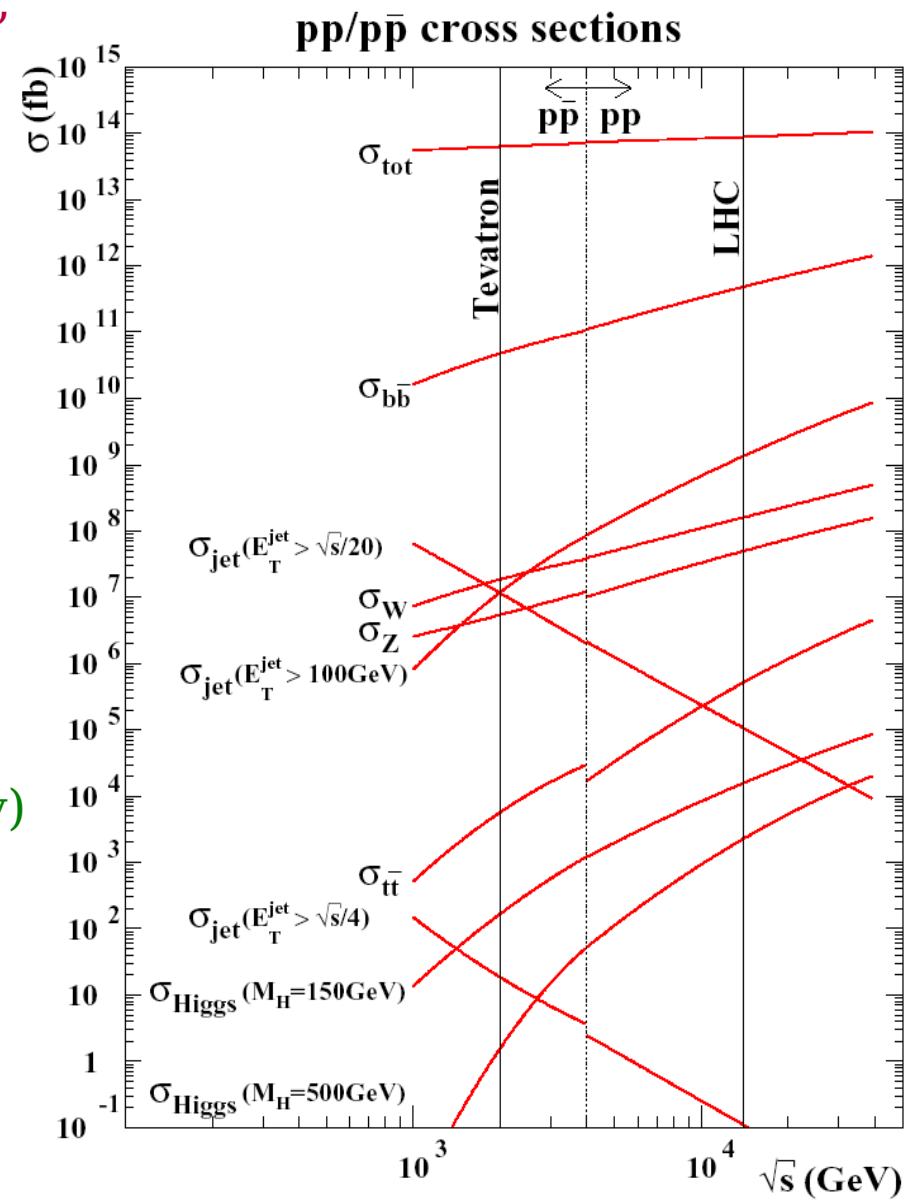
- Search for SUSY (SuperSymmetry)
  - Symmetry between fermions (matter) and bosons (forces)

New Physics (SUSY) is expected to have "small" cross sections

Standard Physics processes have relatively <u>large uncertainties</u> but with <u>huge cross section</u>

[10 pb^{-1} ~ 1 day at 1/100 of the design lumi.]

- Total inelastic: ~ 0.1 barn
- Inclusive QCD multijets ~ 10^5 nb (10^9/day)
- Inclusive bbar: ~ 10^3 nb (10^7/day)
- Inclusive γ/Z/W: ~ 0.1/1.5/10 nb (1- 10k /day)
- Inclusive top ~ 0.89 nb (8.9 k /day)
- Z(νν) + Jets ~ 0.3 nb (3 k /day)



<u>Compare this to 1 TeV SUSY ~ 200 pb</u>

Need to have an accurate understanding of the background – to claim a discovery

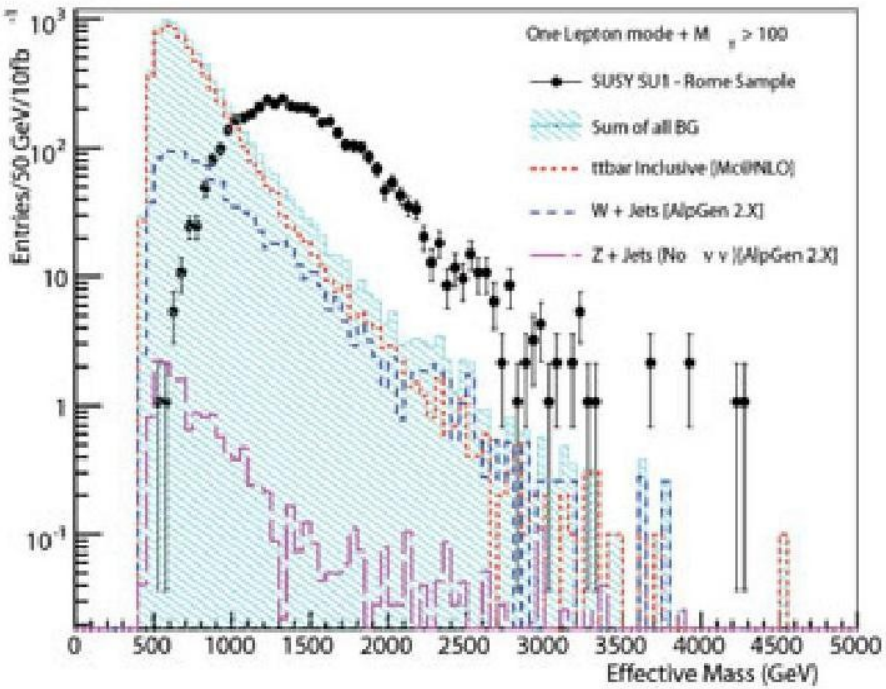
It all started here in OSG ...

To discover evidence for SUSY, Padhi will compare data from the ATLAS experiment against simulations of what would be expected using the Standard Model—the conventional, well-understood model describing the physics of the universe. Simulating these expected Standard Model results requires innumerable numbers of particles to be generated and their interactions with all of ATLAS' many sub-detectors to be recorded and analyzed.

"Simulation of these multijet events involves CPU-intensive exact matrix element calculations," said Padhi. "There were no dedicated CPU farms to complete these calculations, so to perform these studies I had to use the grid."

Accurately simulating the search for supersymmetry required Padhi to create a gateway to three different grid environments from his desk at CERN. He used the Virtual Data Toolkit to create an access point to resources from the Open Science Grid, LHC Computing Grid and the University of Wisconsin-Madison's Condor pool.

"The most difficult part was to make a grid which is interoperable, such that the requirements of all existing grid flavors could be included," explained Padhi. "This was done by modifying the current VDT, and I consumed more than 215 CPU years in less than two months using resources from OSG and Madison's Condor Pool."



A simulation of an inclusive search for supersymmetry completed using grid resources. This plot shows a clear enhancement of the one-lepton

-- SCIENCE GRID THIS WEEK 2006 – Katie Yurkewicz

In order to fulfill ATLAS 2007 Timeline ...

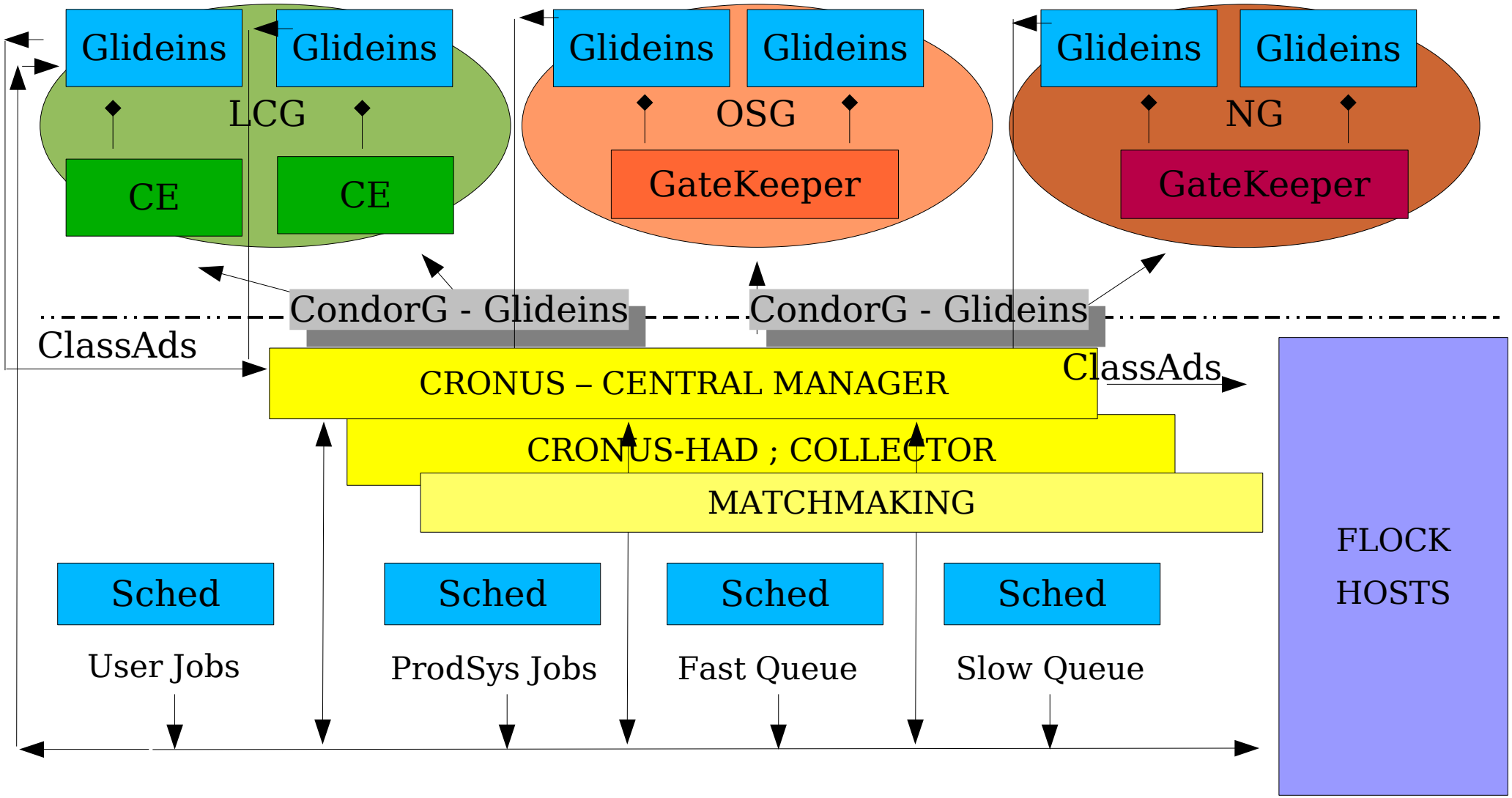
- Running continuously throughout the year (increasing rates):
  - Simulation production
  - Cosmic ray data-taking (detector commissioning)
- January to June:
  - Data streaming tests
- February and May:
  - Intensive Tier-0 tests
- From February onwards:
  - Data Distribution tests
- From March onwards:
  - Distributed Analysis (intensive tests)
- May to July:
  - Calibration Data Challenge
- June to October:
  - Full Dress Rehearsal
- November:
  - GO!

Dario Barberis: ATLAS Computing Plans for 2007

CRONUS was introduced !!! [Motivated by Igor, Frank (CDF-CAF)]

– Based on Condor Glide-in Technology

The concept of late binding is intrinsic to Condor via the ClassAds



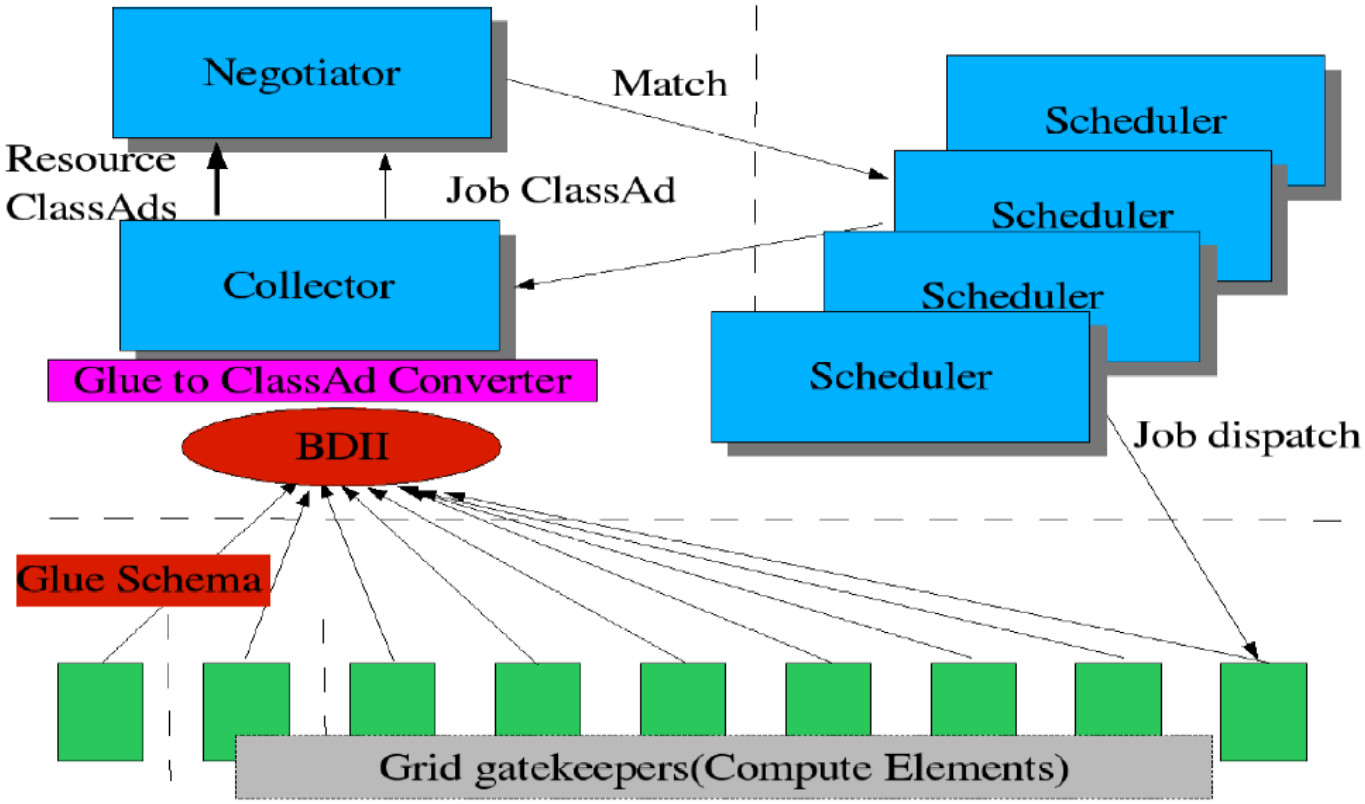
Private Resources

Glide-Ins using Condor-G

- Condor Startup Glide-Ins* are submitted using Condor-G

  - No need to re-invent the wheels !!!

- Glide-in Startup jobs ensures the needed resources and the ATLAS Environment

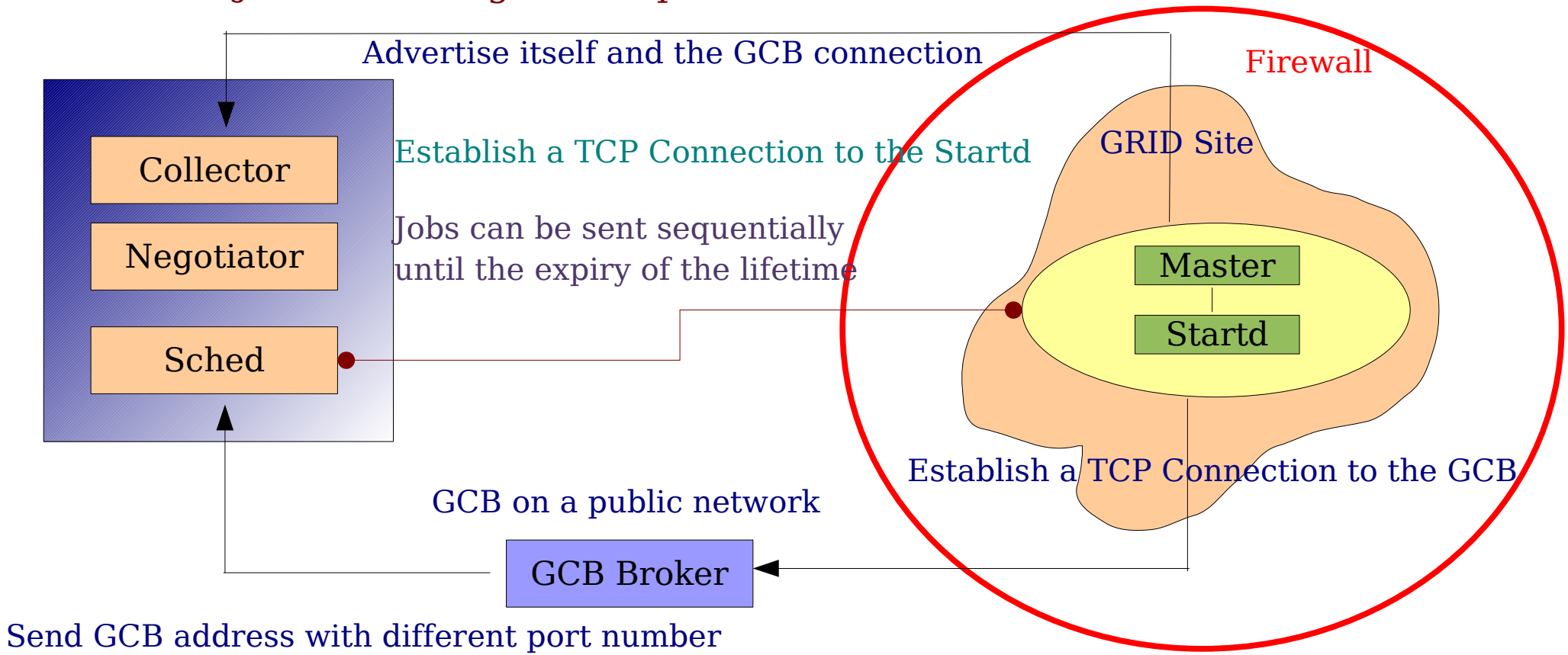


LCG-BDII: atlas-bdii.cern.ch; OSG-BDII: data.grid.iu.edu; NG-BDII: arc-bdii.cern.ch

* Glide-Ins are properly configured Startds, blindly follow Master-Worker relationship

GCB is a Condor proxy service to the Startds :

- Part of Condor distribution, fully integrated in Condor
- Just few configuration parameters.

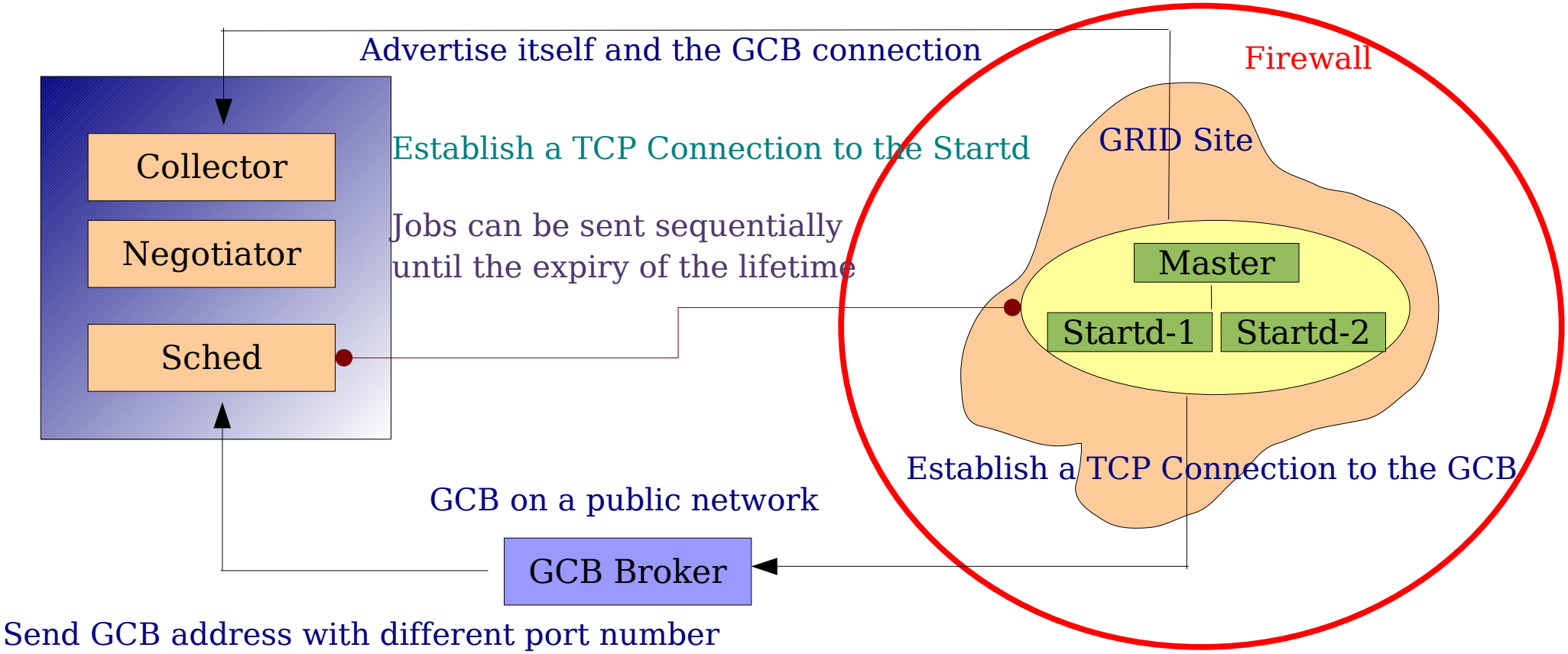


For scalability reasons – use multiple GCB instances

Introducing Multiple Startds (2)

STARTD-1 : Responsible for “standard” production job

STARTD-2 : Responsible for instant DATA transfer/subscription or error recovery jobs

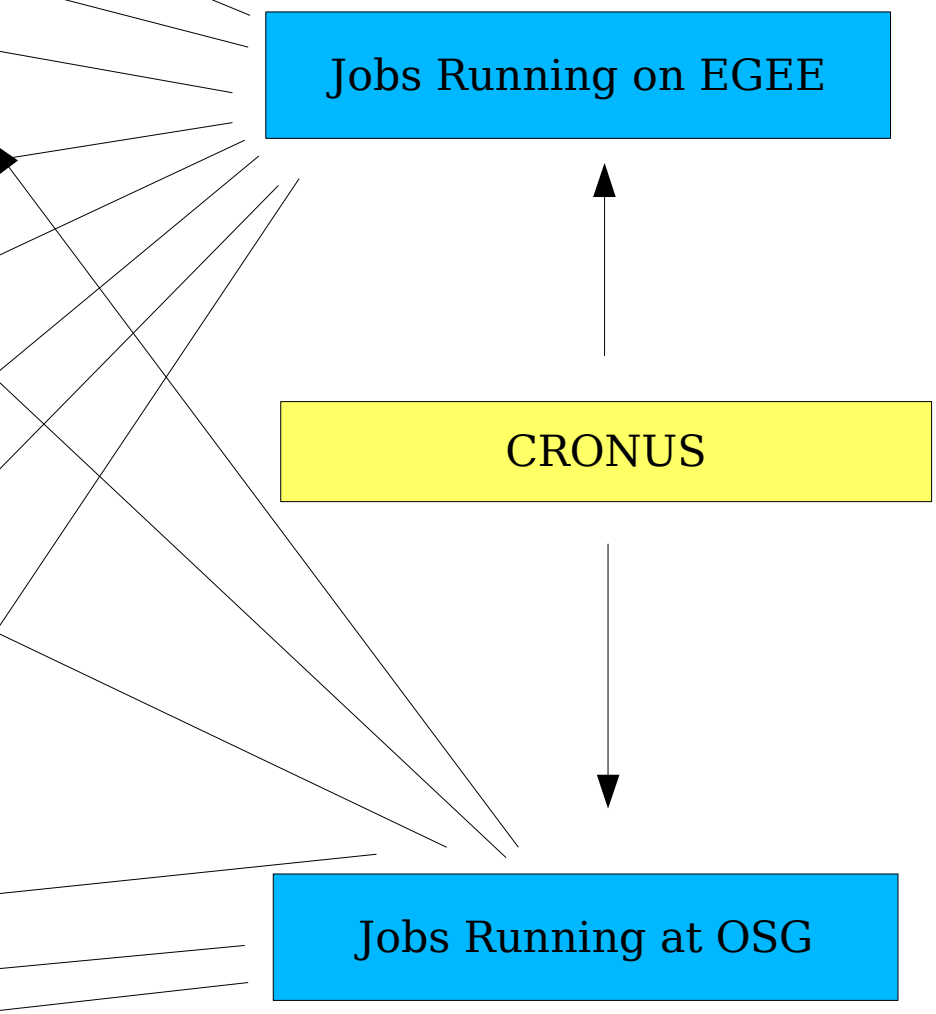


<u>STARTD -2 is mostly Idle:</u>

- Work under progress to exploit BitTorrent
- or any other low level protocoles like rfcp, dccp .. etc

Example Virtual Cluster: Uniformity across various flavours of Grid – OSG, EGEE ..

FileSystemDomain = "bigmac-lcg-ce.physics.utoronto.ca"
FileSystemDomain = "cclcgceli02.in2p3.fr"
FileSystemDomain = "ce101.cern.ch"
FileSystemDomain = "ce102.cern.ch"
FileSystemDomain = "cern.ch"
FileSystemDomain = "cit-gatekeeper.ultralight.org"
FileSystemDomain = "epgce1.ph.bham.ac.uk"
FileSystemDomain = "ft.uam.es"
FileSystemDomain = "gate02.grid.umich.edu"
FileSystemDomain = "grid003.ft.uam.es"
FileSystemDomain = "grid-ce0.desy.de"
FileSystemDomain = "grid-ce1.desy.de"
FileSystemDomain = "grid-ce2.desy.de"
FileSystemDomain = "hamptonu.edu"
FileSystemDomain = "ifaece01.pic.es"
FileSystemDomain = "lancs.ac.uk"
FileSystemDomain = "lcg002.ihep.ac.cn"
FileSystemDomain = "lcg2ce.ific.uv.es"
FileSystemDomain = "lcgce01.gridpp.rl.ac.uk"
FileSystemDomain = "lcgce02.nic.ualberta.ca"
FileSystemDomain = "lcg-ce0.ifh.de"
FileSystemDomain = "qmul.ac.uk"
FileSystemDomain = "teraport.edu"
FileSystemDomain = "tier2-osg.uchicago.edu"
FileSystemDomain = "tp-osg.uchicago.edu"
FileSystemDomain = "vampire.accre.vanderbilt.edu"
FileSystemDomain = "vanderbilt.edu"
FileSystemDomain = "zeus02.cyf-kr.edu.pl"



– Fully inter-operable Virtual batch System

CRONUS- Dynamics of Information System - ClassAds

Based on TiersOfATLASCache into ClassAds @ the WNs

MACHINE ClassAd

```
MyType = "Machine"
TargetType = "Job"
Name = "8505@ccwali04.in2p3.fr"
Machine = "ccwali04.in2p3.fr"
Rank = 0.000000
CpuBusy = FALSE
IS_GLIDEIN = TRUE
DaemonStopTime = DaemonStartTime + 172740
ATLAS_SOFT = "/afs/in2p3.fr/group/atlas/sw/software"
ATHENA_VERSION = "10.0.1,10.0.4,11.0.3,11.0.41,11.0.42,11.0.5,11.2.0,12.0.1,12.0.2,12.0.3,12.0.31,12.2.0"
GLOBAL_SE = "LYON"
GLOBAL_SE_DISK = "srm://ccsrm.in2p3.fr:8443/pnfs/in2p3.fr/data/atlas/disk/dq2/"
GLOBAL_SE_TAPE = "srm://ccsrm.in2p3.fr:8443/pnfs/in2p3.fr/data/atlas/dq2/"
LOCAL_SE_DISK = "srm://ccsrm.in2p3.fr:8443/pnfs/in2p3.fr/data/atlas/disk/dq2/"
LOCAL_SE_TAPE = "srm://ccsrm.in2p3.fr:8443/pnfs/in2p3.fr/data/atlas/dq2/"
LOCAL_CENTER = "TIER1"
```

CRONUS

Job ClassAd

```
universe = vanilla
executable = EvGenproduction.sh
transfer_files = always
output = job/out.mc11test.005030.Jimmy_jetsJ1.evgen.EVNT.v12000301._00002
error = job/err.mc11test.005030.Jimmy_jetsJ1.evgen.EVNT.v12000301._00002
log   = job/SubmissionLog2
stream_output = False
stream_error = False
notification = never
Requirements  = (StringlistMember("12.0.31",ATHENA_VERSION) && ( GLOBAL_SE == "LYON" )) && (Arch == "INTEL")
WhenToTransferOutput = ON_EXIT
OnExitRemove = TRUE
arguments = "mc11test.005030.Jimmy_jetsJ1.evgen.v12000301_tid3326 \
 5030 5000 5000 2 DC3.005030.Jimmy_jetsJ1.py \
mc11test.005030.Jimmy_jetsJ1.evgen.EVNT.v12000301._00002.pool.root NONE NONE NONE"
```

CRONUS- Dynamics of Information System - ClassAds

One can even match to the Tier2s – Info from the WNs

MACHINE ClassAd

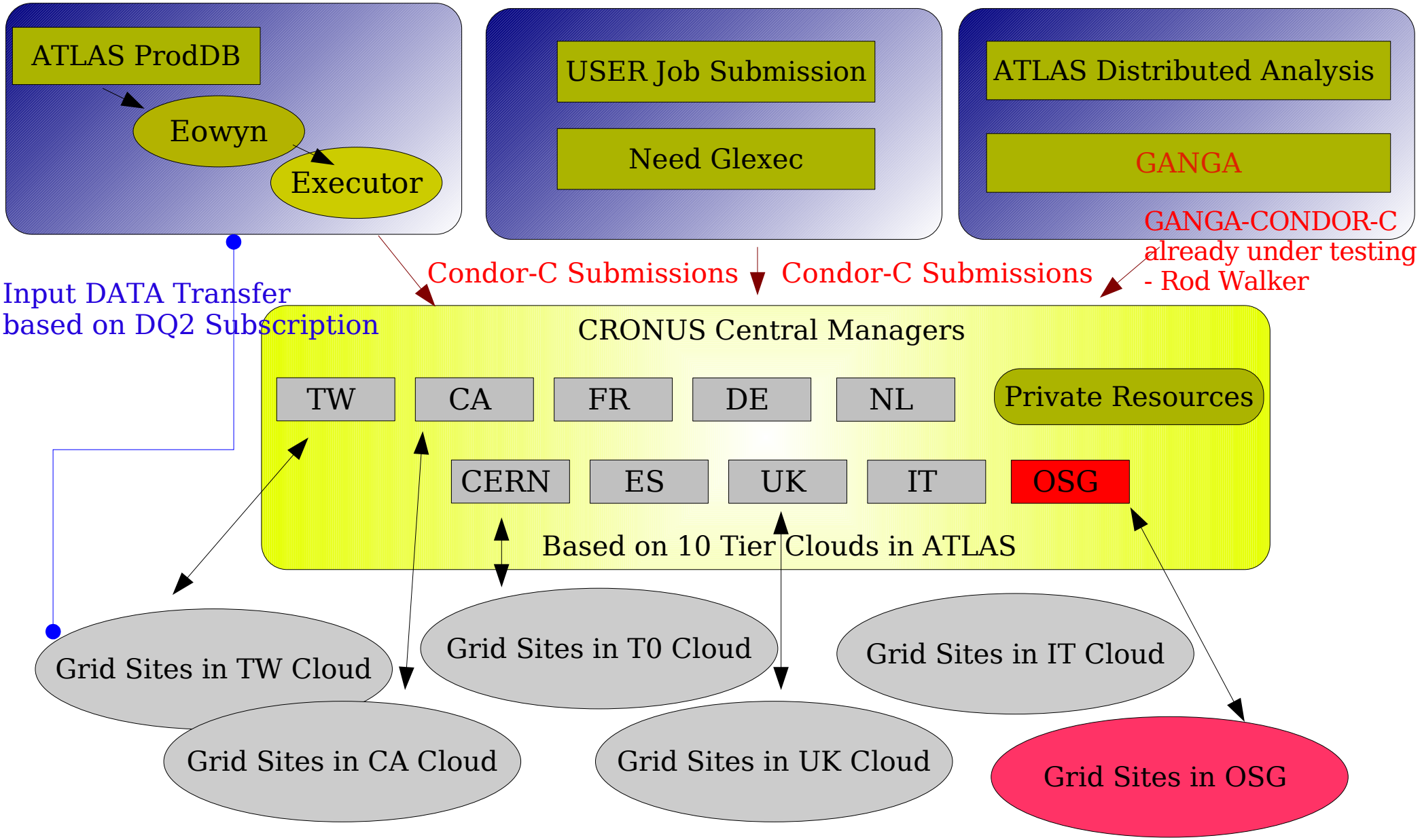
```
Name = "2414@lcg021.ihep.ac.cn"
Machine = "lcg021.ihep.ac.cn"
Rank = 0.000000
CpuBusy = FALSE
IS_GLIDEIN = TRUE
DaemonStopTime = DaemonStartTime + (2879 * 60)
ATLAS_SOFT = "/opt/exp_soft/atlas/software"
ATHENA_VERSION = "11.0.42,11.0.5,12.0.2,12.0.3,12.0.31"
GLOBAL_SE = "LYON"
GLOBAL_SE_DISK = "srm://ccsrm.in2p3.fr:8443/pnfs/in2p3.fr/data/atlas/disk/dq2/"
GLOBAL_SE_TAPE = "srm://ccsrm.in2p3.fr:8443/pnfs/in2p3.fr/data/atlas/dq2/"
LOCAL_SE_DISK = "srm://atlasse01.ihep.ac.cn:8443/pnfs/ihep.ac.cn/data/atlas/"
LOCAL_SE_TAPE = "srm://atlasse01.ihep.ac.cn:8443/pnfs/ihep.ac.cn/data/atlas/"
LOCAL_CENTER = "TIER2"
MyType = "Machine"
```

CRONUS

Job ClassAd

```
universe = vanilla
executable = EvGenproduction.sh
transfer_files = always
output = job/out.mc11test.005030.Jimmy_jetsJ1.evgen.EVNT.v12000301._00002
error = job/err.mc11test.005030.Jimmy_jetsJ1.evgen.EVNT.v12000301._00002
log   = job/SubmissionLog2
stream_output = False
stream_error = False
notification = never
Requirements  = (StringlistMember("12.0.31",ATHENA_VERSION) && ( GLOBAL_SE == "LYON" )) && (Arch == "INTEL")
WhenToTransferOutput = ON_EXIT
OnExitRemove = TRUE
arguments = "mc11test.005030.Jimmy_jetsJ1.evgen.v12000301_tid3326 \
 5030 5000 5000 2 DC3.005030.Jimmy_jetsJ1.py \
mc11test.005030.Jimmy_jetsJ1.evgen.EVNT.v12000301._00002.pool.root NONE NONE NONE"
```

CRONUS - ATLAS Production Executor



Cronus inherits - Condor built-in features

- Global fairshare at user and group level
- Post-submit priority manipulations
- Guaranteed ATLAS - ATHENA Environment
- Live CPU, Walltime and RAM requirement and display
- Accommodates the usage of both Grid and Private resources
- Reduces grid related errors by more than 50%
- Homogeneous Environment across all grid flavours – EGEE, OSG ... etc.
- Relies in a Condor agent in two ways
  - Uses agent as service for reliably executing user/production jobs
  - With Master-Worker inheritance solves the problem reliably.
- Credential Management – glexec – Still to come

Cronus inherits - Condor built-in features

- Live log updates:
  - User can know at any instant of time at what stage the actual job is ...

```
[spadhi@glow-s007 ~]$
[spadhi@glow-s007 ~]$
[spadhi@glow-s007 ~]$ tail -10  /atlas/JOBS/4542/2036330/2/stdout
<?xml version="1.0" ?>
<POOLFILECATALOG><File ID="D6D3288F-BF92-DB11-B4DC-003048209597"><physical><pfn filetype="ROOT_All" name="mc12.005802.JF17_pythia_jet_filter.
evgen.EVNT.v12003105_tid004541._01066.pool.root"/></physical><logical/></File></POOLFILECATALOG>
<?xml version="1.0" ?>
<!DOCTYPE POOLFILECATALOG  SYSTEM "InMemory">
<POOLFILECATALOG>
<File ID="D6D3288F-BF92-DB11-B4DC-003048209597"><physical><pfn filetype="ROOT_All" name="mc12.005802.JF17_pythia_jet_filter.evgen.EVNT.v12003
105_tid004541._01066.pool.root"/></physical><logical/></File>
</POOLFILECATALOG>

2007-02-21 12:10:17,843 runPyJT     : INFO
2007-02-21 12:10:17,844 runPyJT     : INFO     Running TRF, log is: misal1_mc12.005802.JF17_pythia_jet_filter.digit.log.v12003105_tid004542._
06474.job.log.2
[spadhi@glow-s007 ~]$
```

Apart from efficiency improvements,

- Cronus also helps much better control of the job priorities

The standard grid-queue's are very unpredictable,

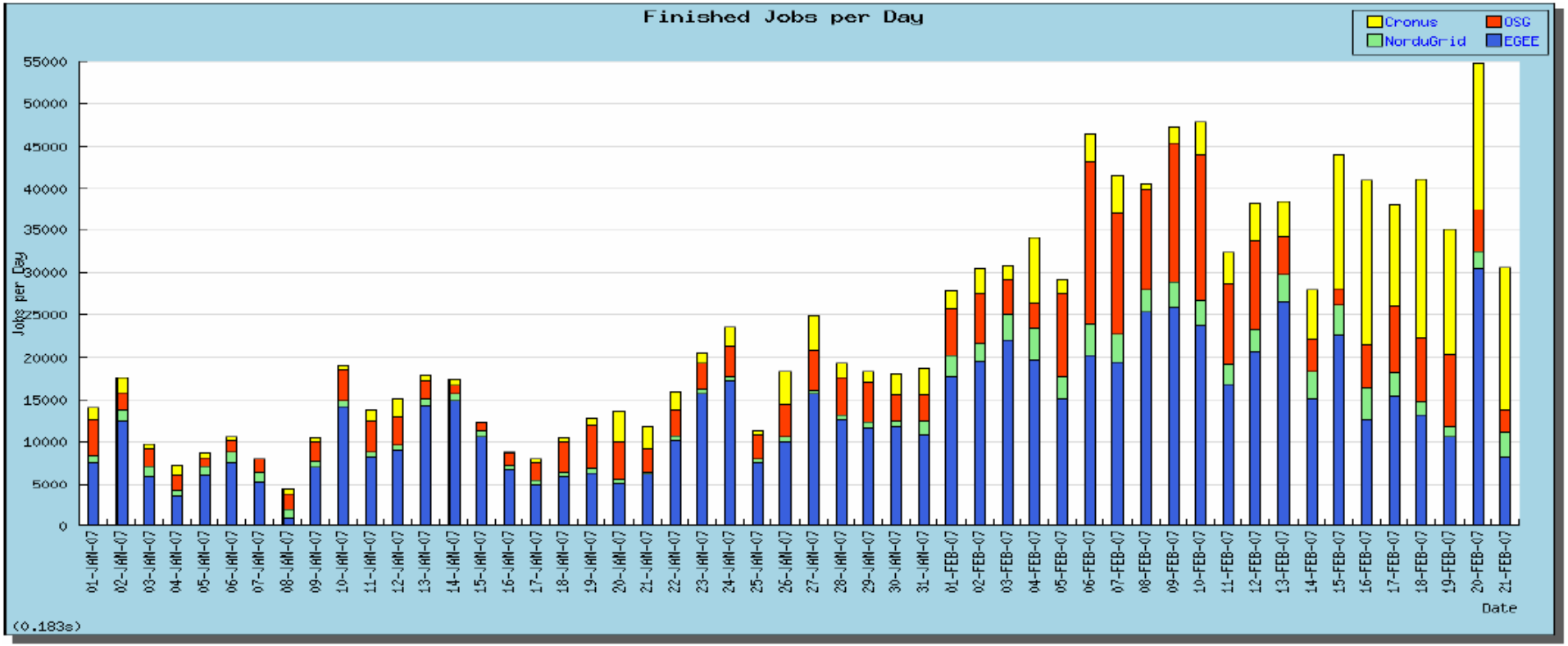
- Cronus glide-ins completely removes them from the game

Cronus in ATLAS Production – Jobs finished Per day

Fully integrated into ATLAS Monitoring System

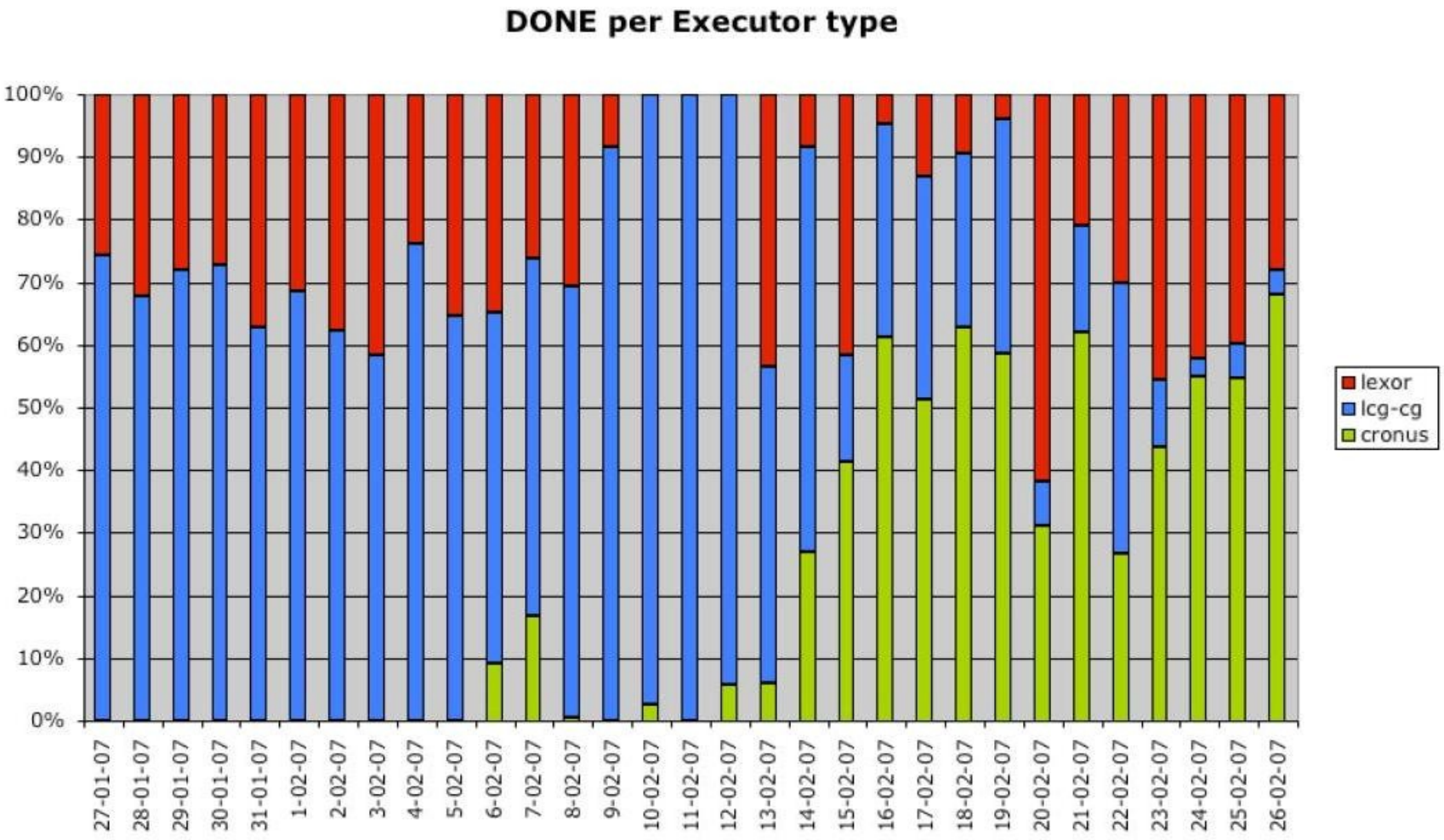
Peter Jenni (ATLAS Overview Week)

The computing and software developments have made huge steps to be ready to receive the data, from the combined cosmics running and the first run at the end of the year



CRONUS OSG NorduGrid EGEE

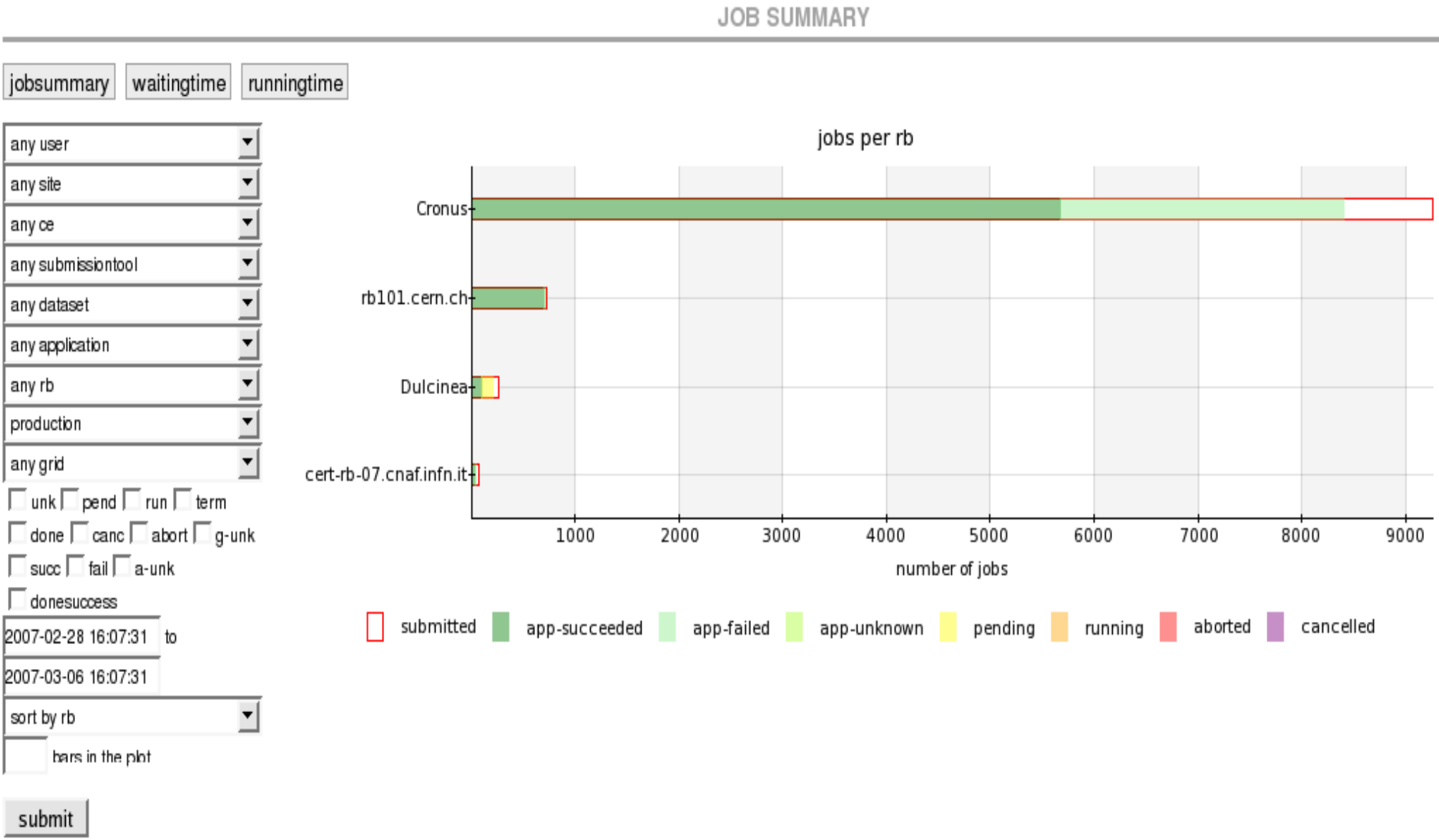
Report by Guido Negri – ATLAS-LCG-EGEE Task Force Meeting 27th Feb 2007



Instances of previous slide have been grouped into three main categories

ARDA Dashboard Monitoring

Thanks to Massimo Lamanna and his team.



Summary and Outlook

The pre-production mode using CRONUS was found to be very successful !!!!

- Currently more than 50% of the ATLAS production jobs are DONE using Cronus

Use Condor Hawkeye to actually run the executor – Error checking .. etc

Run two VMs (Virtual Machines on each CPU) in Cronus:

- One for User/Production Job
- Second to be used completely for the DATA transfer

Plans to provide User support via GANGA:

- Condor-C (already built-in feature in Cronus)

Glide-ins are planned to be submitted using Glite-WMS

- Phase out the Condor-G executor soon ....

Expand Cronus to use effectively in inter-operable mode across OSG ...

- Cronus actually originated using OSG based on the Physics Need

Summary and Outlook

CRONUS – means time .....

According to Greek mythology

- CRONUS was the father of ZEUS who sent ATHENA to be in charge of ATLAS.