ATLAS Data Challenges

LCG-GDB

September 8th 2004

ATLAS DC; Grid and Operations teams

ATLAS Data challenges

- DC1 (2002-2003)
  - Put in place full software chain
    - Simulation of the data
    - Reconstruction
  - Production system
    - Tools (bookkeeping; monitoring; …)
    - Intensive use of Grid
- DC2 (Summer 2004)
  - New software
  - New "automated" production system
  - Full use of Grids
  - Test of Computing Model
- DC3 (Spring 2006)
  - Final test before data taking

ATLAS-DC2 operation

- Consider DC2 as a three-part operation:
  - part I: production of simulated data (July-September 2004)
    - running **on "Grid"**
    - **Worldwide**
  - part II: test of Tier-0 operation (October 2004)
    - Do in 10 days what "should" be done in 1 day when real data-taking start
    - Input is "Raw Data" like
    - output (ESD+AOD) will be **distributed to Tier-1s in real time** for analysis
  - part III: test of distributed analysis on the Grid (Oct.-Dec. 2004)
    - access to event and non-event data from anywhere in the world both in organized and chaotic ways
- Requests
  - ~30 Physics channels (10 Millions of events)
  - Several millions of events for calibration (single particles and physics samples)

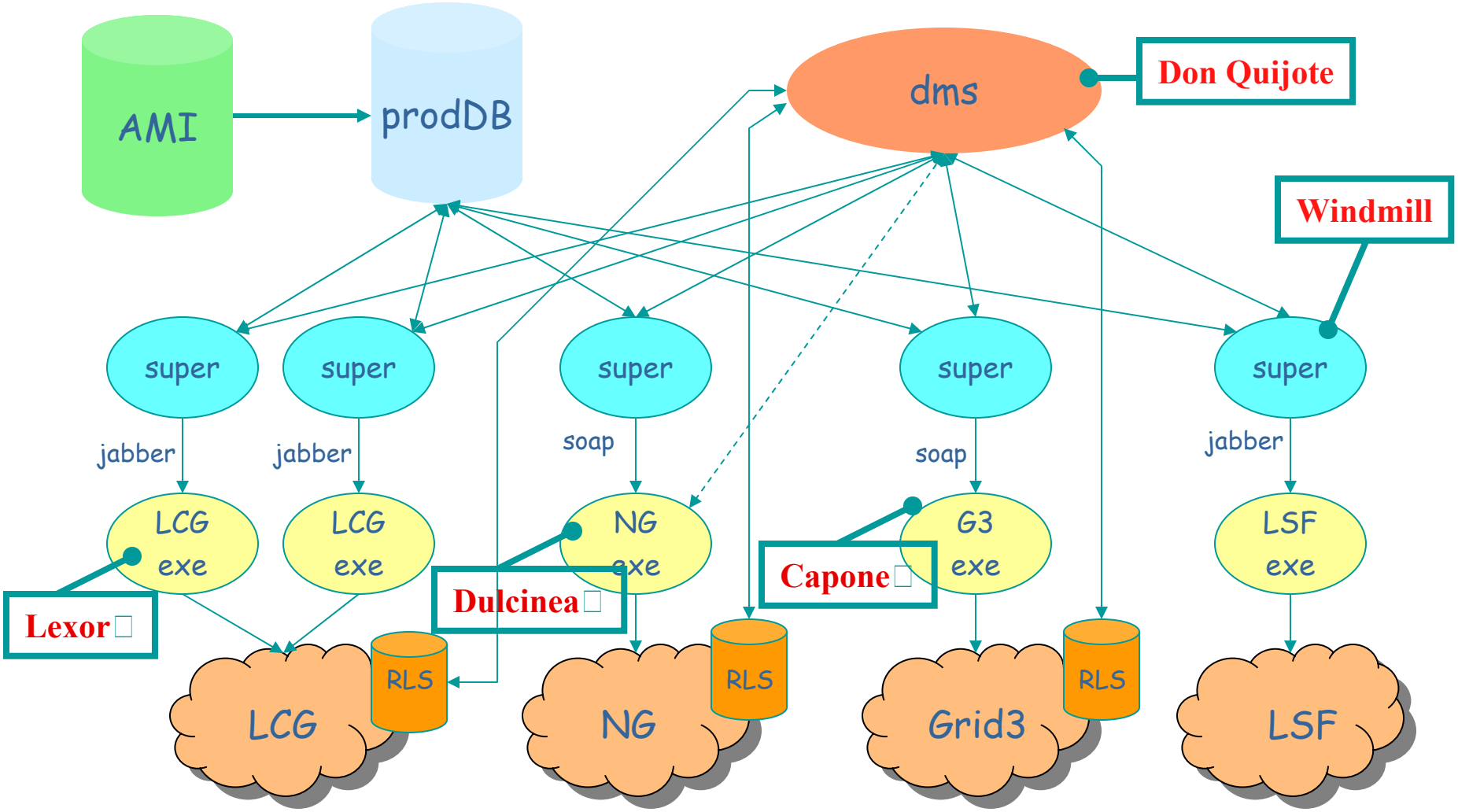
More on Phase I: Data preparation

- DC2 Phase I
  - Part 1: Event generation
    - Physics processes --> 4-momentum of particles
    - Several Event generators (Pythia; Herwig; …)
  - Part 2: Detector simulation (Geant4)
    - Tracking of particles through the detector
    - Records interaction of particle with sensitive elements of the detector
  - Part 3: Pile-up and digitization
    - Pile-up: superposition of "background" events with the "signal" event
    - Digitization: response of the sensitive elements of the detector
    - Output, called byte-stream data, "looks-like" "Raw Data"
  - Part 4: Data transfer (to CERN Tier-0)
    - ~35 TB in 4 weeks
  - Part 5: Event mixing
    - Physics events are "mixed" in "ad-hoc" proportion

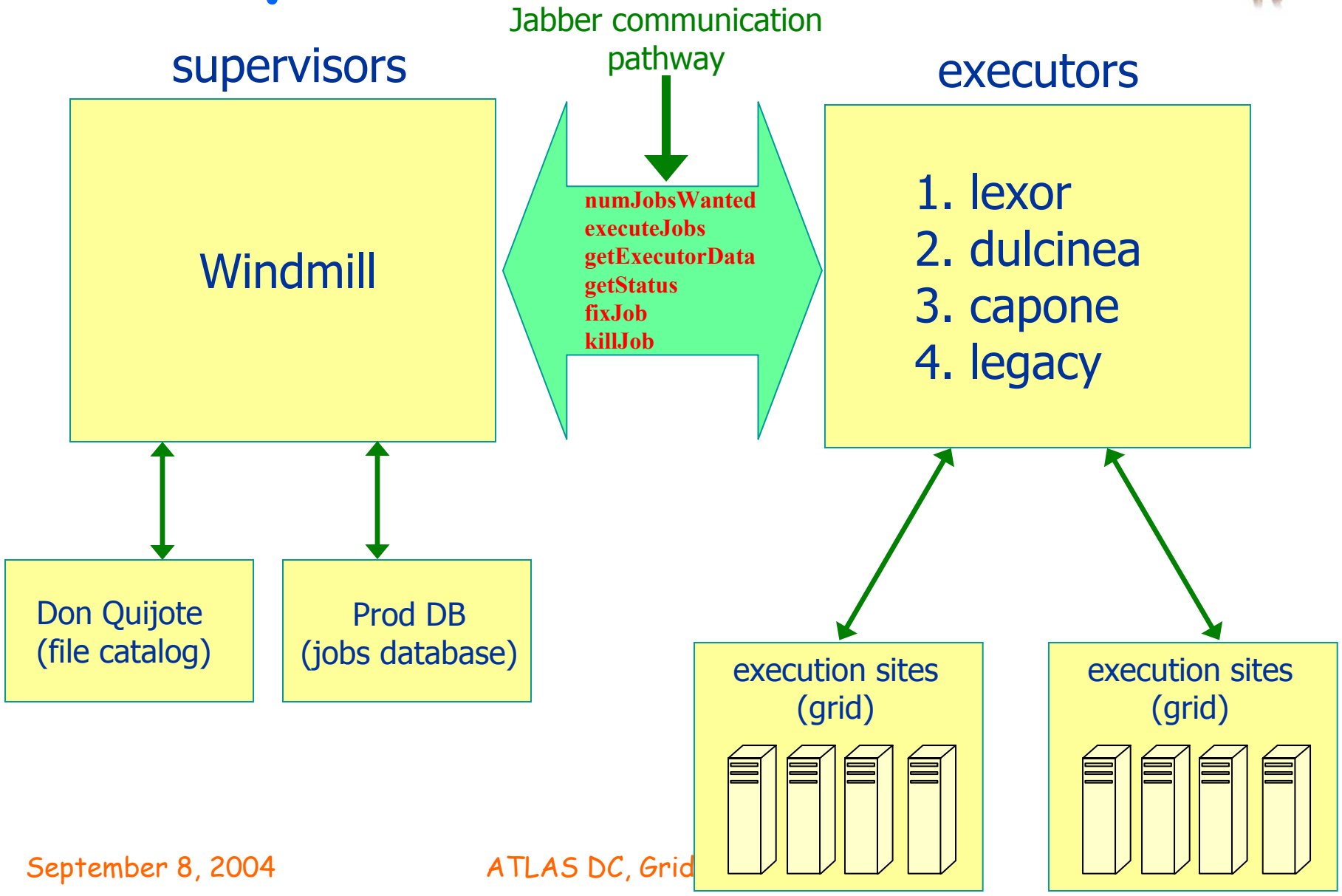
ATLAS Production System

- Automated version of previous ATLAS DC1 production system
- Components
  - Supervisor: Windmill (US)
  - Executors (one per Grid or "legacy batch") :
    - Capone (Grid3) (US)
    - Dulcinea (NorduGrid) (Scandinavia)
    - Lexor (LCG) (Italy)
    - "Legacy systems" (Germany-FZK; France-Lyon)
  - Data Management System (DMS): Don Quijote (CERN)
  - Bookkeeping: AMI (LPSC-Grenoble)
  - Production Data base (Oracle)
    - Definition and status of the jobs

ATLAS Production system



Supervisor -Executors

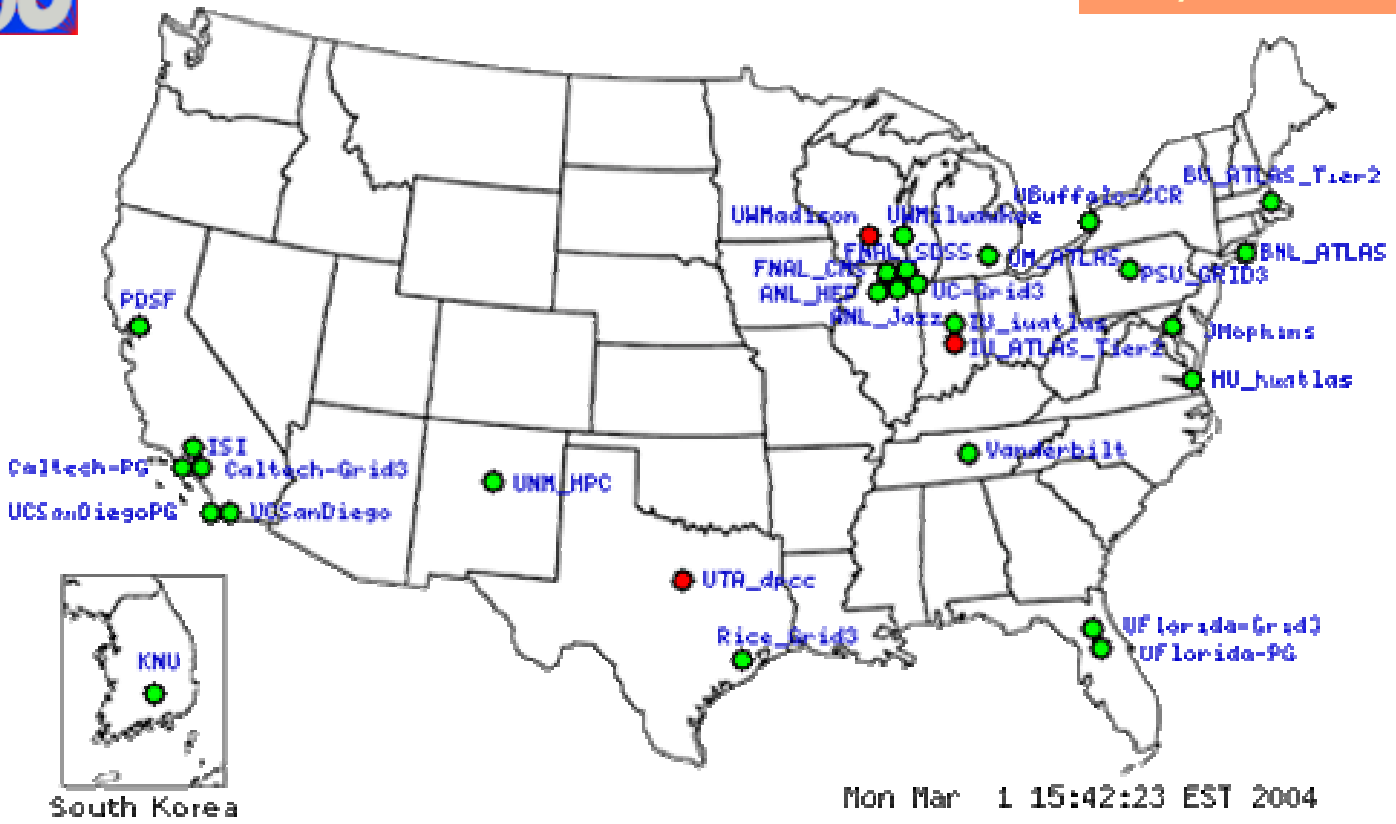


ATLAS DC2 Phase I

- Started beginning of July and still running
- On 3 Grids
  - LCG
    - Including some non-ATLAS sites (Legnaro, Torino)
    - Using in production mode the LCG-Grid-Canada interface
      - 3 sites are accessible through this interface(TRIUMF)
        - Uni. Victoria, Uni. Alberta and WestGrid(SFU/TRIUMF)
  - NorduGrid
    - Several Scandinavian super-computer resources
  - Grid3
    - Harnessing opportunistic computing resources that are not dedicated to ATLAS (e.g. US CMS sites)

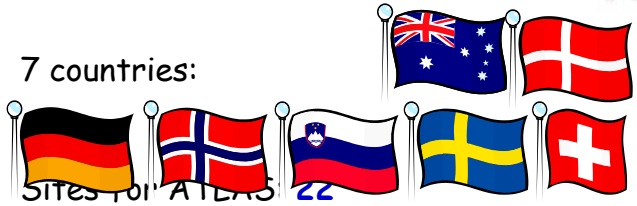
Current Grid3 Status (3/1/04)
(http://www.ivdgl.org/grid2003)

- 28 sites, multi-VO
- shared resources
- ~2000 CPUs
- dynamic – roll in/out



NorduGrid & Co. Resources:

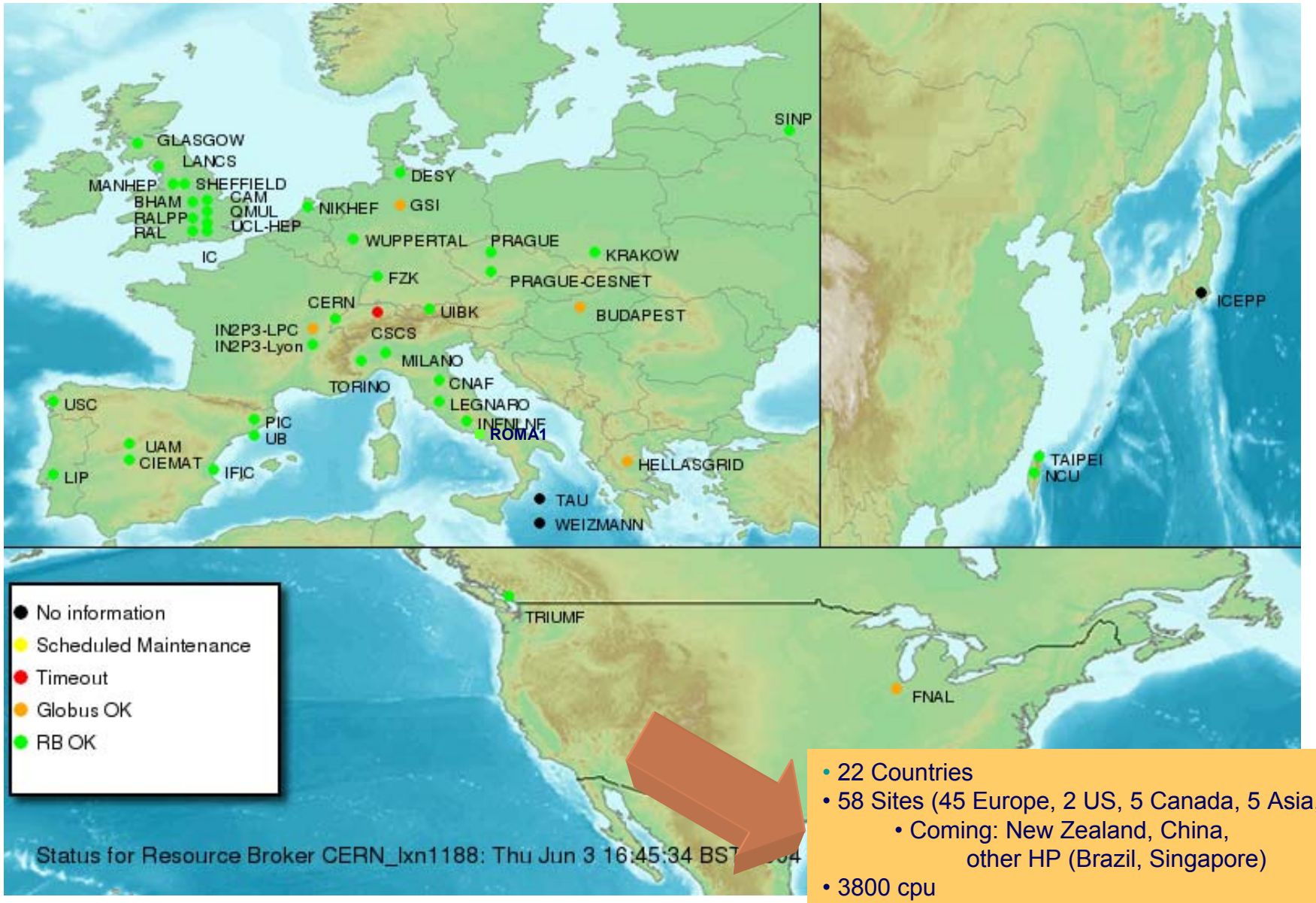
- 7 countries:



- Sites for ATLAS: 22
  - Dedicated: 3, the rest is shared
- CPUs for ATLAS: ~3280
  - Effectively available: ~800
- Storage Elements for ATLAS: 10
  - Capacity: ~14 TB, all shared

Cluster	running	waiting	finished	failed	(%)	total
atlas.hpc.unimelb.edu.au	28	86	641	20	(3%)	828
brenta.ijs.si	50	30	3200	217	(7%)	3562
bluesmoke.nsc.liu.se	48	70	1949	145	(7%)	2354
lxsrv9.lrz-muenchen.de	6	56	695	70	(10%)	1051
hypatia.uio.no	56	18	835	106	(13%)	1011
hagrid.it.uu.se			3550	508	(14%)	5325
benedict.aau.dk	46	41	2050	326	(16%)	2292
grid.uio.no	13	22	580	90	(16%)	726
sigrid.lunarc.lu.se	16	84	2542	441	(17%)	3510
sg-access.pdc.kth.se		58	2736	491	(18%)	2876
lheppc10.unibe.ch	12	14	455	82	(18%)	576
fire.ii.uib.no	10	12	838	163	(19%)	1073
farm.hep.lu.se	45	70	911	214	(23%)	1120
ingrid.hpc2n.umu.se	7		3507	886	(25%)	3774
fe10.dcsc.sdu.dk			1052	342	(33%)	1058
genghis.hpc.unimelb.edu.au		8	608	336	(55%)	653
morpheus.dcgc.dk	17	17	456	289	(63%)	490
charm.hpc.unimelb.edu.au			718	456	(64%)	916
atlas.fzk.de	15	23	77	52	(68%)	115
hive.unicc.chalmers.se			34	26	(76%)	34
lscf.nbi.dk	16	17	188	147	(78%)	221
grid.fi.uib.no			1	1	(100%)	1
TOTAL	385	626	27623	5408	(20%)	33566

Sites in LCG-2: 4 June 2004



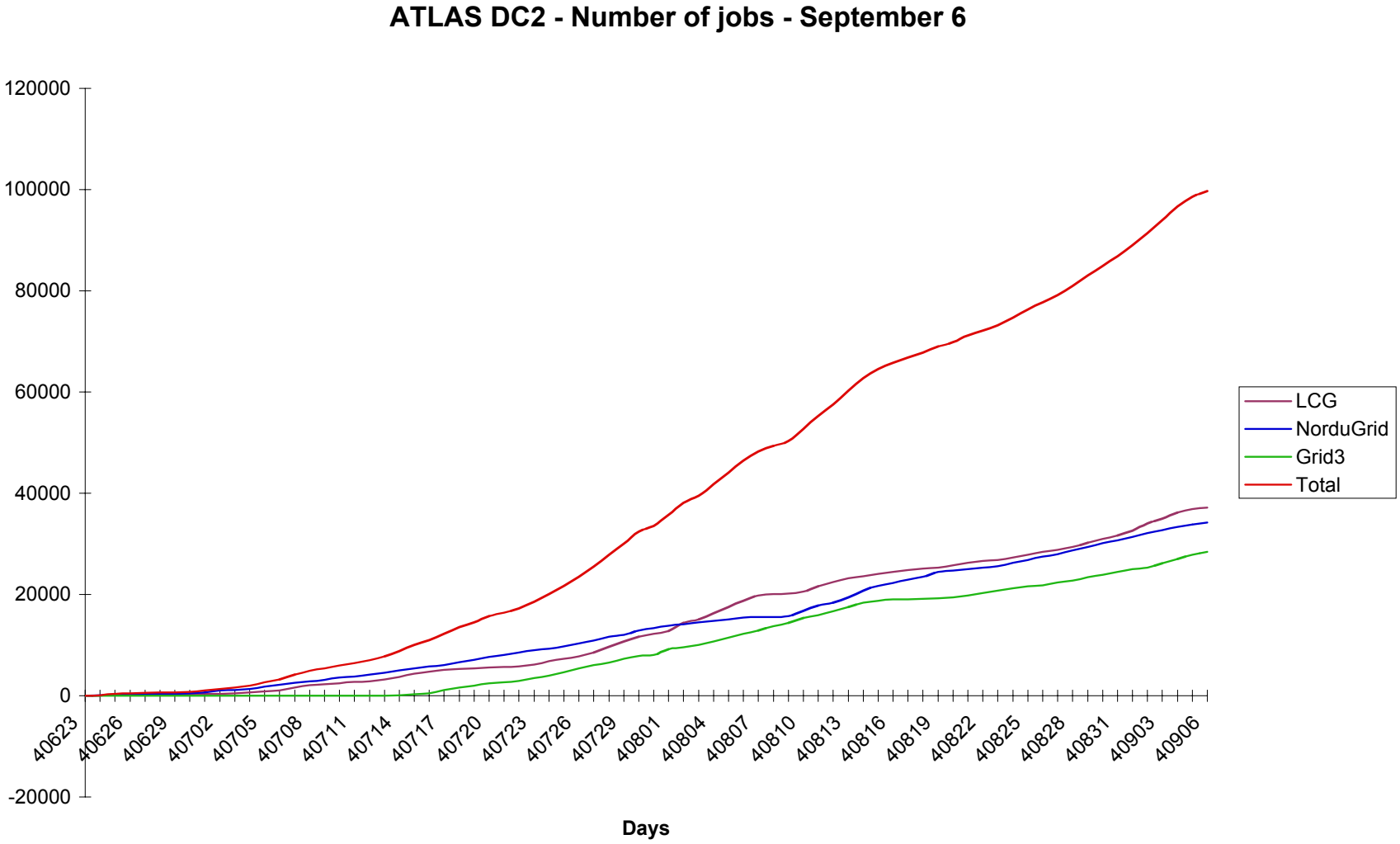
- 22 Countries
- 58 Sites (45 Europe, 2 US, 5 Canada, 5 Asia, 1 HP)
  - Coming: New Zealand, China, other HP (Brazil, Singapore)
- 3800 cpu

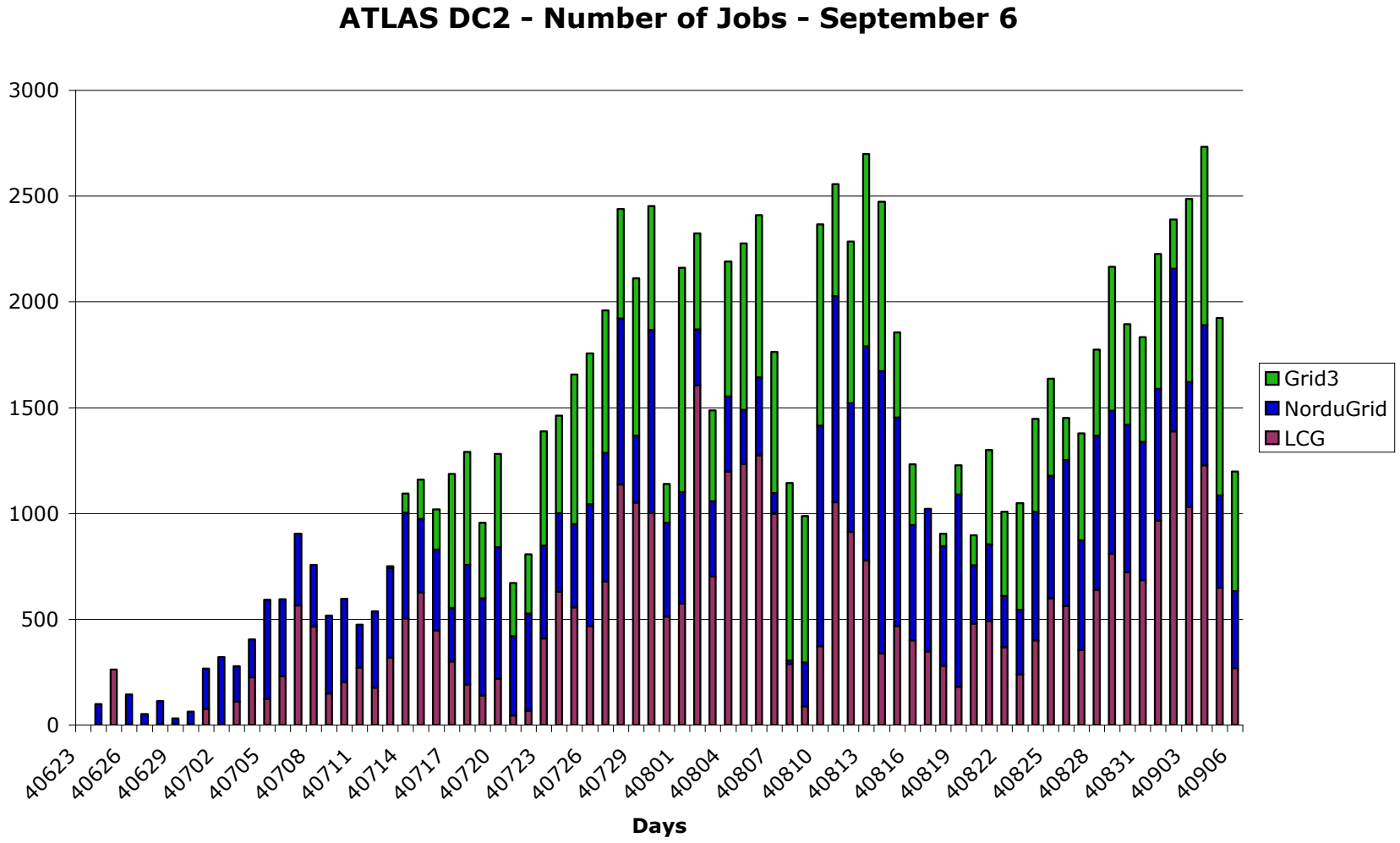
ATLAS DC2 Phase I

- Main difficulties at the initial phase
  - For all Grids
    - Debugging the Production System
    - On LCG and Grid3 several instances of the Supervisor have to be run for better coping with the instability of the system. As a consequence the Production System was more difficult to handle.
  - LCG
    - Mis-configuration of sites; Information system (wrong or missing information); Job submission and Resource Broker; Jobs ranking.
    - Data management(copy & register); Stage in/out problems
  - NorduGrid
    - Replica Location Service (Globus) hanging several times per day
    - Mis-configuration of sites
    - Access to the conditions database
  - Grid3
    - Data Management - RLS interactions
    - Software distribution problems
    - Load on gatekeepers
    - Some problems with certificates (causing jobs to abort)
  - Good collaboration with Grid teams to solve the problems

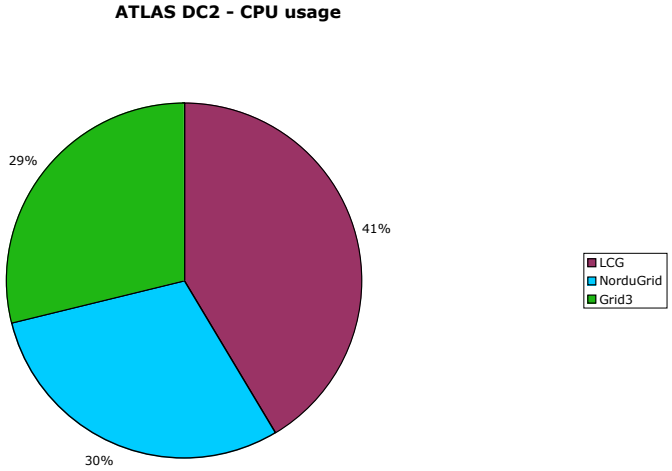
ATLAS DC2 Phase I

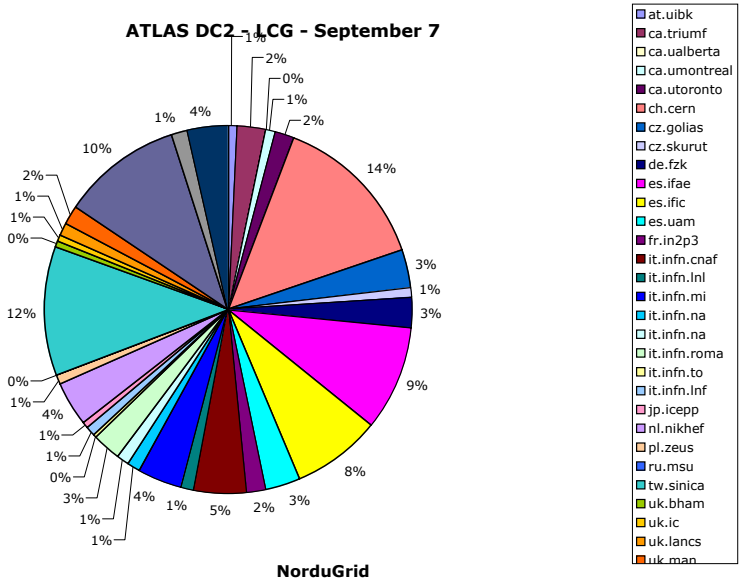
- Not all problems solved
  - NorduGrid
    - RLS; Access to the conditions database; Storage elements died …
  - Grid3
    - Try to avoid single points of failure (adding new servers)
    - Lack of storage management in some sites
  - LCG
    - Still some problems with resource broker and information system
    - And data management (copy and register) and stage in/out problems
  - For all
    - Slowness of the response of the Production Database
      - Problem that appears after ~6 weeks of running and which is still not fully understood (mix software and hardware problems? being worked with IT-DB).
      - Has been solved!
- Consequences: we did not succeed (yet) to run as many jobs as expected per day
- Nevertheless should be completed by end-September and is "Grid" only

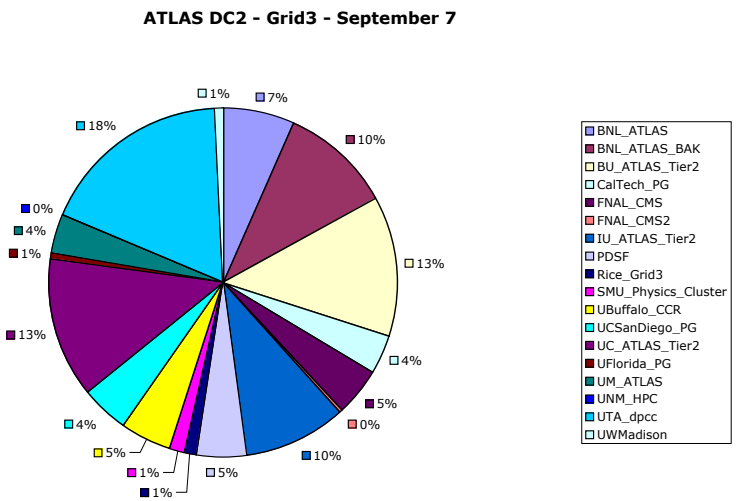

ATLAS DC2 - Number of jobs - September 6
120000
100000
80000
60000
40000
20000
0
-20000
40623
40626
40629
40702
40705
40708
40711
40714
40717
40720
40723
40726
40729
40801
40804
40807
40810
40813
40816
40819
40822
40825
40828
40831
40903
40906
Days
LCG
NorduGrid
Grid3
Total

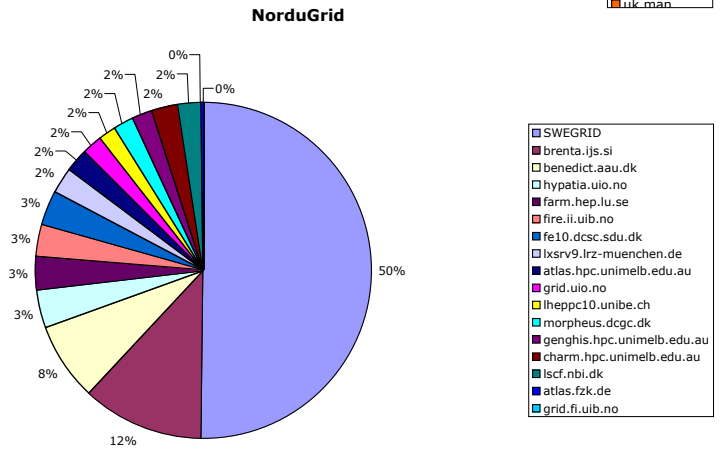

ATLAS DC2 - Number of Jobs - September 6
3000
2500
2000
1500
1000
500
0
40623
40626
40629
40702
40705
40708
40711
40714
40717
40720
40723
40726
40729
40801
40804
40807
40810
40813
40816
40819
40822
40825
40828
40831
40903
40906
Days
Grid3
NorduGrid
LCG

CPU usage & Jobs

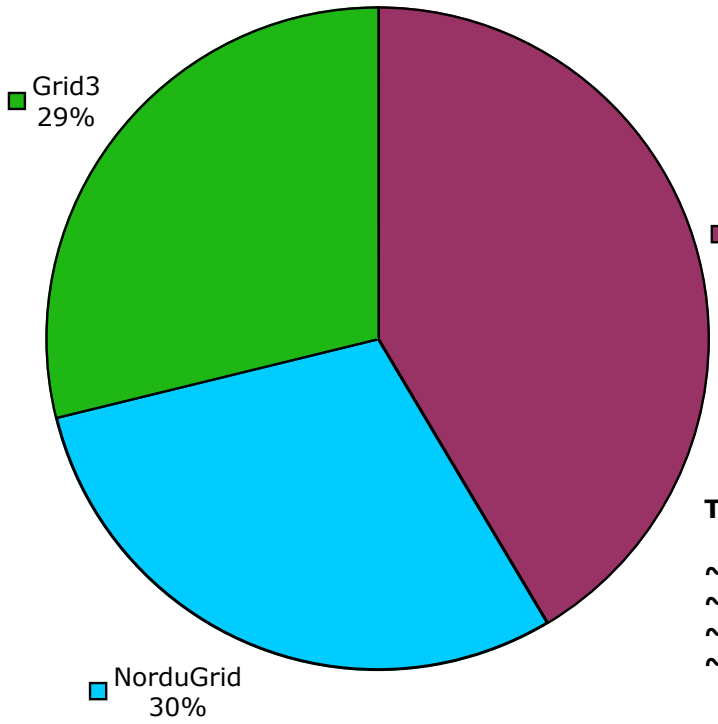








ATLAS DC2 - CPU usage

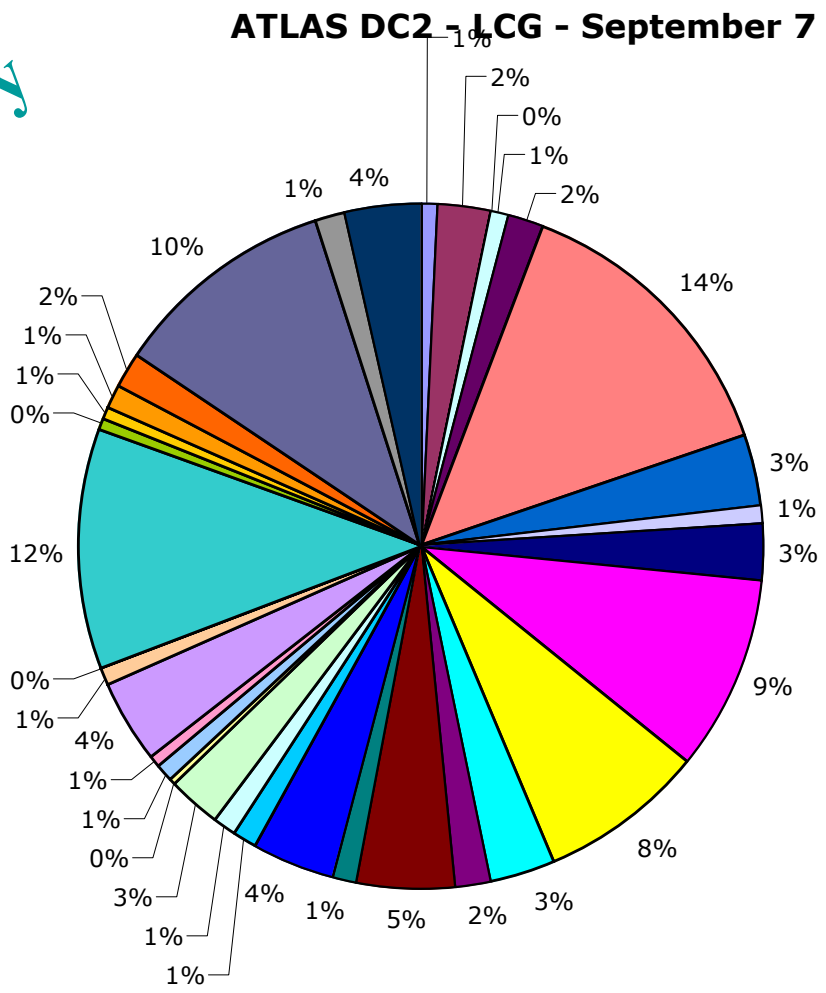


Total:

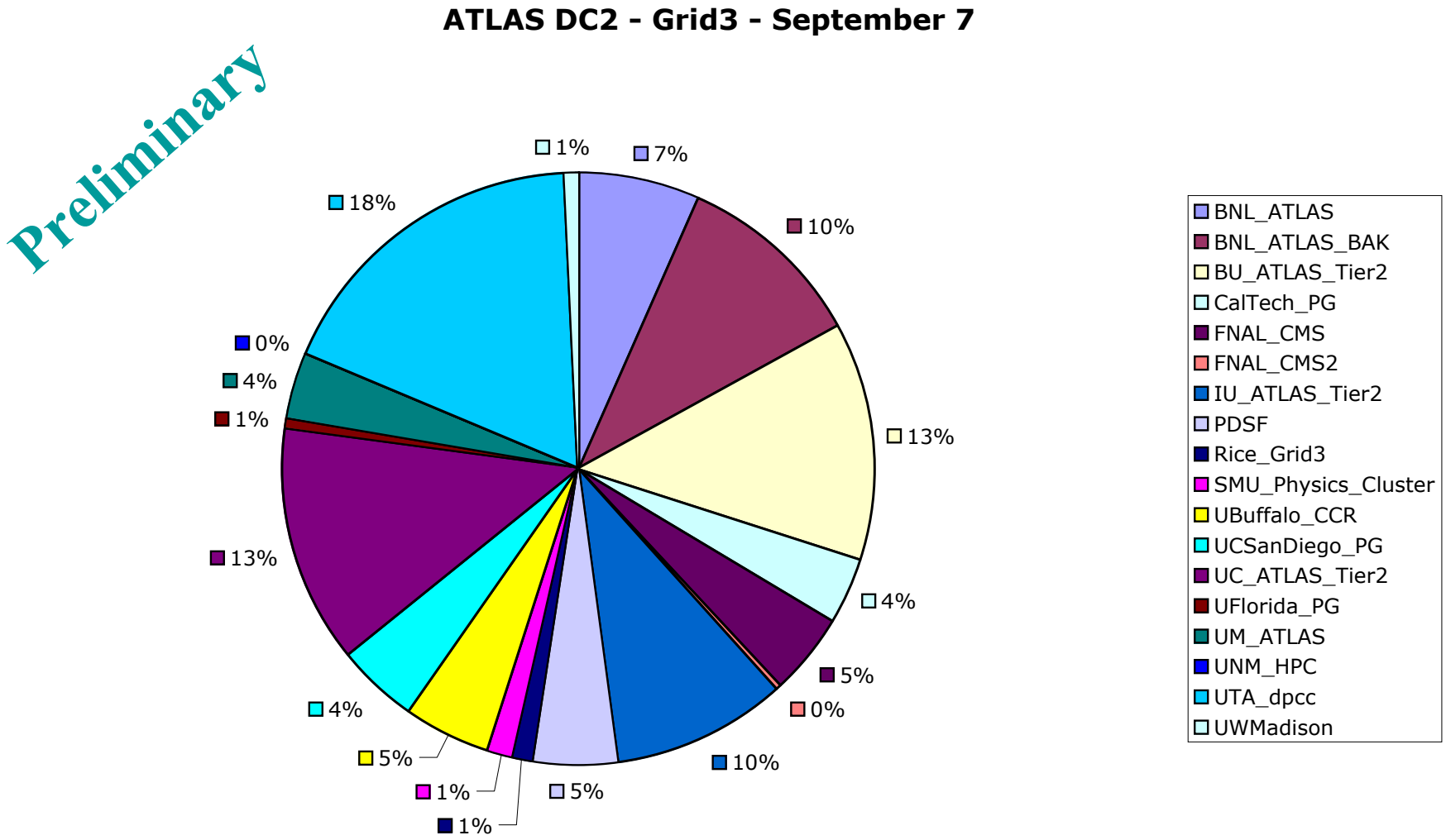
~ 1350 kSI2k.months
~ 95000 jobs
~ 7.7 Million events fully simulated (Geant4)
~ 22 TB

Jobs distribution on LCG

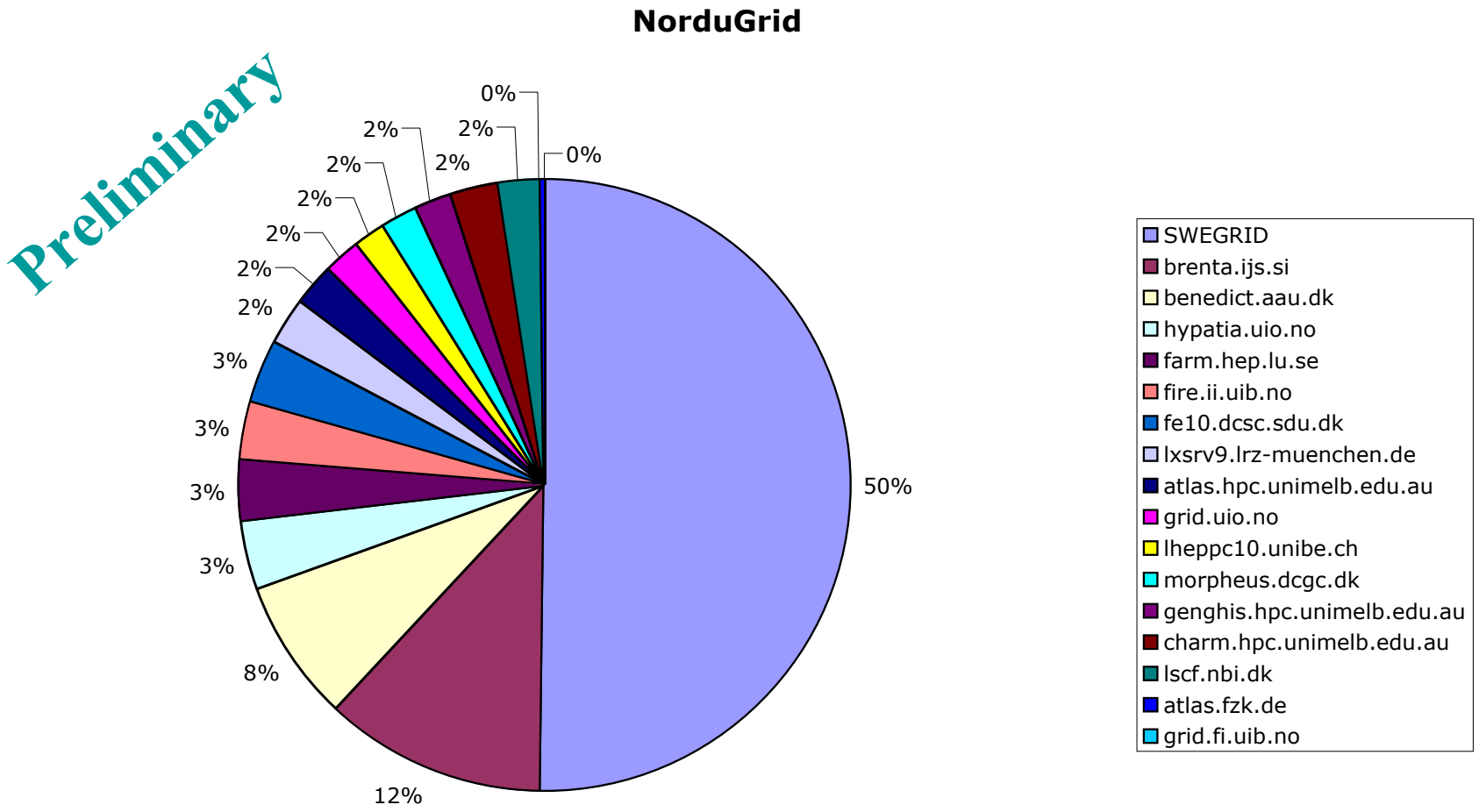
Preliminary

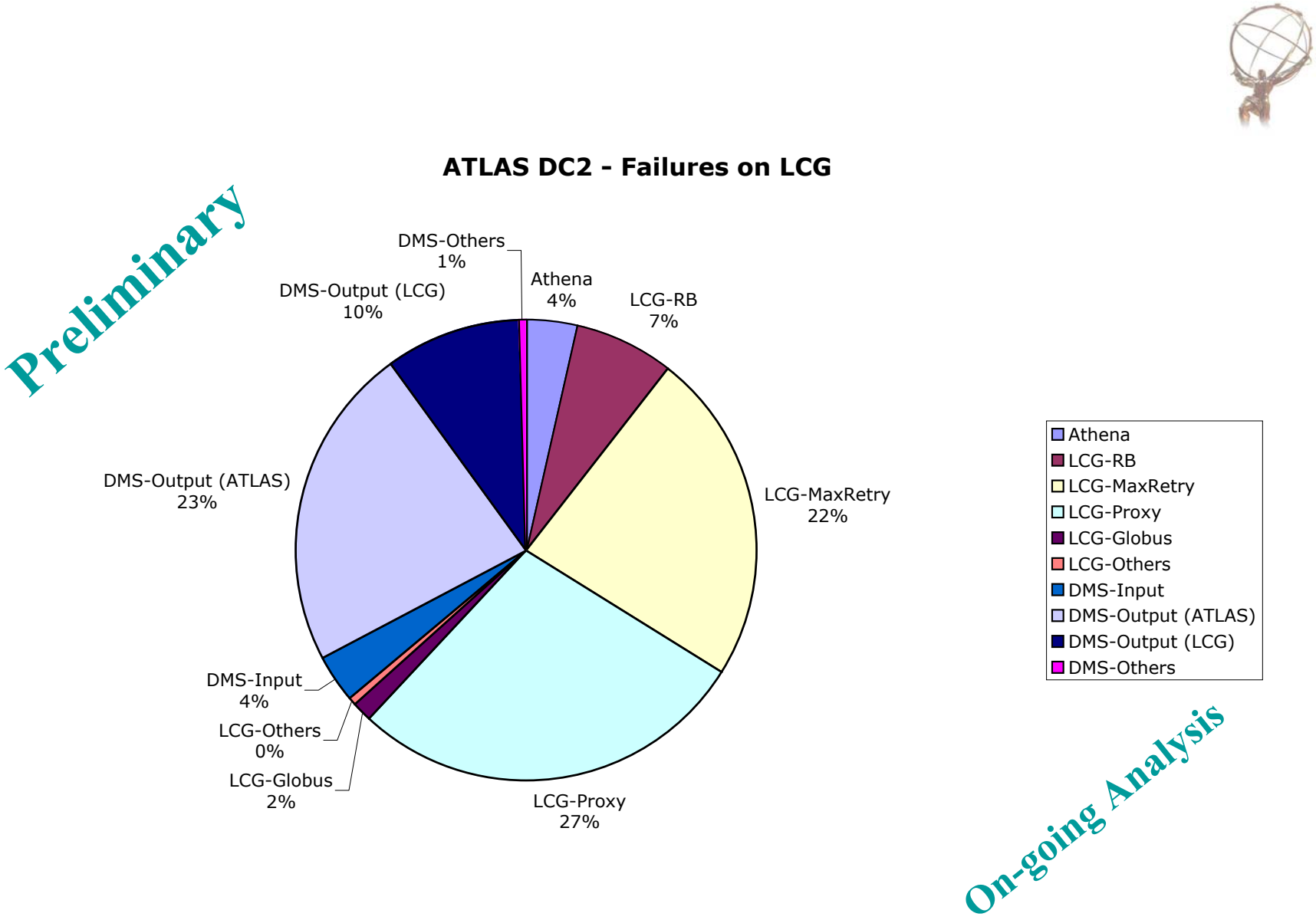


Jobs distribution on Grid3



Jobs distribution on NorduGrid




ATLAS DC2 - Failures on LCG
Preliminary
DMS-Others 1%
Athena 4%
LCG-RB 7%
DMS-Output (LCG) 10%
DMS-Output (ATLAS) 23%
LCG-MaxRetry 22%
DMS-Input 4%
LCG-Others 0%
LCG-Globus 2%
LCG-Proxy 27%
Athena
LCG-RB
LCG-MaxRetry
LCG-Proxy
LCG-Globus
LCG-Others
DMS-Input
DMS-Output (ATLAS)
DMS-Output (LCG)
DMS-Others
On-going Analysis

Summary (1)

- Major efforts on the past few months
  - Redesign of the ATLAS Event Data Model and Detector Description
  - Integration of the LCG components (G4; POOL; …)
  - Introduction of the **Production System**
    - Interfaced with 3 Grid flavors (and "legacy" systems)
- Delays in all activities have affected the schedule of DC2
  - Note that Combined Test Beam is ATLAS 1st priority
  - And DC2 schedule was revisited
    - To wait for the readiness of the software and of the Production system

Summary (2)

- DC2
  - About 80% of the Geant4 simulation foreseen for PhaseI has been completed using only Grid and using the 3 flavors coherently; Pile-up just starting
  - The 3 Grids have been proven to be usable for a real production and this is a major achievement
- BUT
  - Phase I progressing slower than expected and it's clear that all the involved elements (Grid middleware; Production System; deployment and monitoring tools over the sites) need improvements
  - It's one of the goals of the Data Challenges to identify these problems as early as possible.
- Phase II (Tier0 exercise) is scheduled for October