Grid3
and a few thoughts on federating grids

Rob Gardner
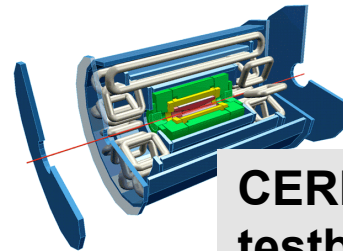
University of Chicago

rwg@hep.uchicago.edu

GGF PGM

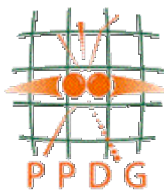May 7, 2004

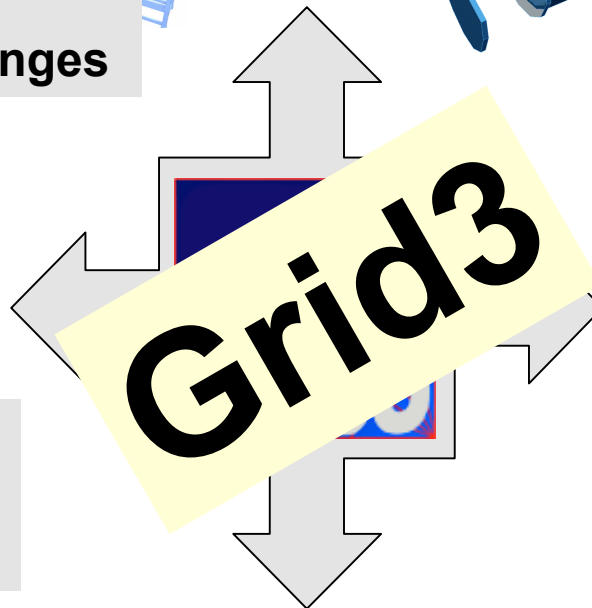# Grid2003: an application grid laboratory

CERN LHC: USCMS testbeds & data challenges

CERN LHC: US ATLAS testbeds & data challenges

Grid3

end-to-end HENP applications

virtual data research

virtual data grid laboratory

Grid3 at a Glance

- Grid environment built from core Globus and Condor middleware, as delivered through the Virtual Data Toolkit (VDT)
    - GRAM, GridFTP, MDS, RLS, VDS, VOMS, …

- …equipped with VO and multi-VO security, monitoring, and operations services

- …allowing federation with other Grids where possible, eg. CERN LHC Computing Grid (LCG)
    - USATLAS: GriPhyN VDS execution on LCG sites
   - USCMS: storage element interoperability (SRM/dCache)

Grid3 – a snapshot of sites

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

4

Grid Design
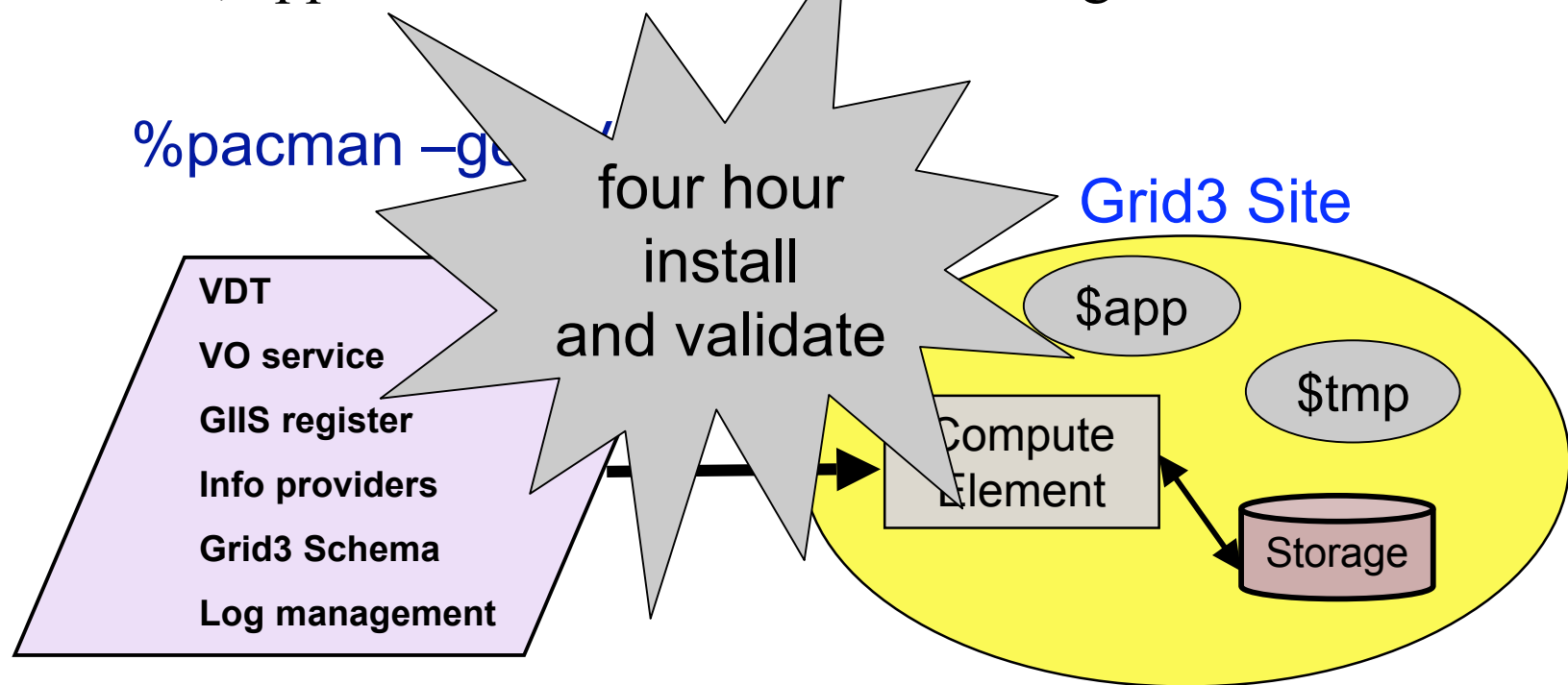
- **Simple approach:**
  - Sites consisting of
    - Computing element (CE)
    - Storage element (SE)
    - Information and monitoring services
  - VO level, and multi-VO
    - VO information services
    - Operations (iGOC)
- **Minimal use of grid-wide systems**
  - No centralized resource broker, replica/metadata catalogs, or command line interface
    - to be provided by individual VO's
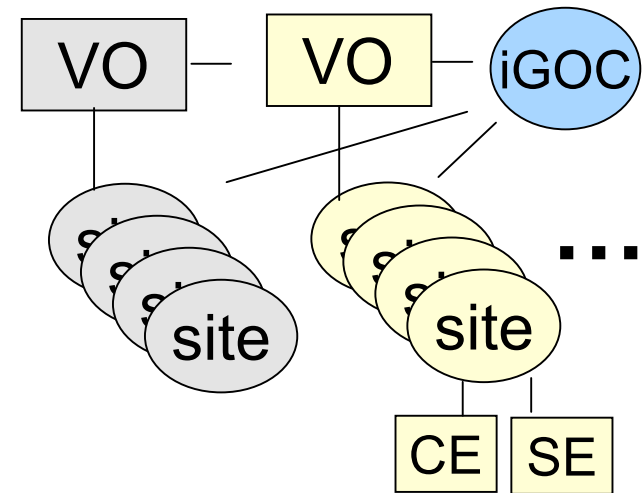
Site Services and Installation

- Goal is to install and configure with minimal human intervention
- Use Pacman tool and distributed software "caches"
- Registers site with VO and Grid3 level services
- Accounts, application install areas & working directories

%pacman –g

four hour install and validate

Grid3 Site

VDT

VO service

GIIS register

Info providers

Grid3 Schema

Log management

$app

$tmp

Compute Element

Storage

VDT 1.1.4

- Virtual Data System 1.2.3

  ClassAds 0.9.5

  Condor 6.6.1

  EDG CRL Update 1.2.5

  EDG Make Gridmap 2.1.0

  Fault Tolerant Shell (ftsh) 2.0.0

  Globus 2.4.3 plus patches

  GLUE Information providers, (CVS version 1.79, 25-July-2003)

  GLUE Schema 1.1, extended version 1

GPT 3.1

GSI-Enabled OpenSSH 3.0

Java SDK 1.4.1

KX509 2031111

Monalisa 0.95

MyProxy 1.11

Netlogger 2.2

PyGlobus 1.0

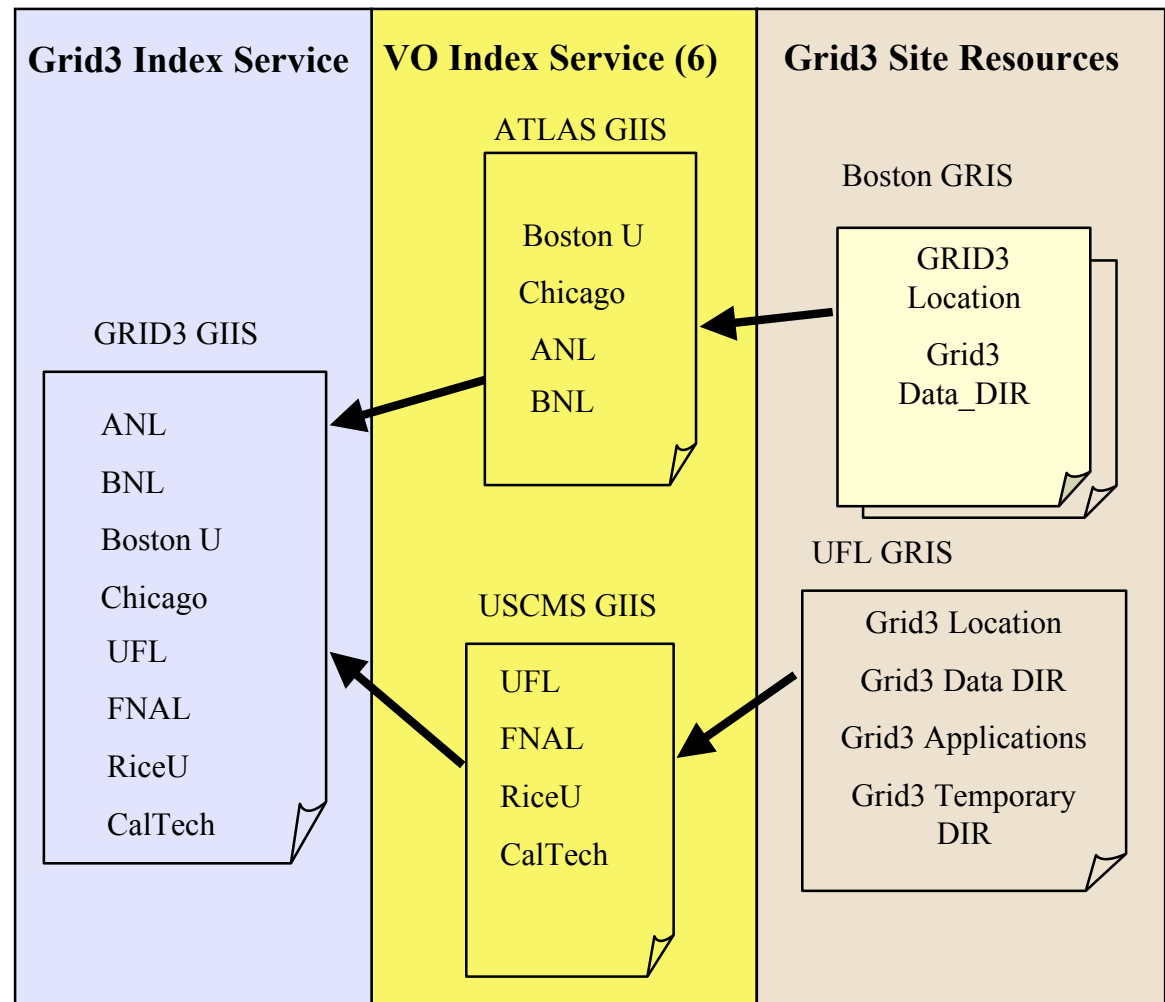PyGlobus URL Copy 1.1.2.11

RLS 2.1.4

UberFTP 1.3

Multi-VO Security Model

- DOEGrids Certificate Authority
- PPDG and iVDGL Registration Authorities, with VO or site sponsorship
- Automated multi-VO authorization, using VOMS
- Each VO manages a service and it's members
- Each Grid3 site is able to generate a Globus gridmap file with an authenticated SOAP query to each VO service
- Site-specific adjustments or mappings

VOMS servers

SDSS

US CMS

Grid3 gridmap

Grid3 Sites

US ATLAS

BTeV

LSC

iVDGL

8
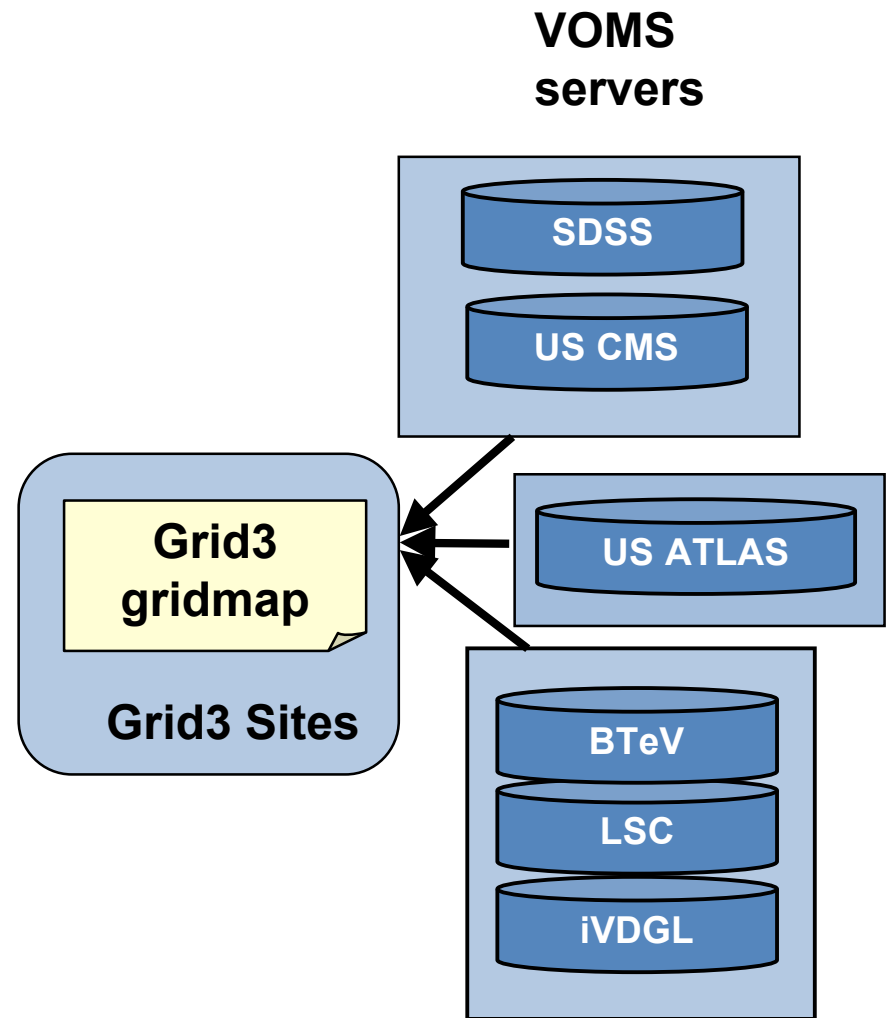
Information Services

- Information Index Service (GIIS)
- Automated Resource registration to Index Service
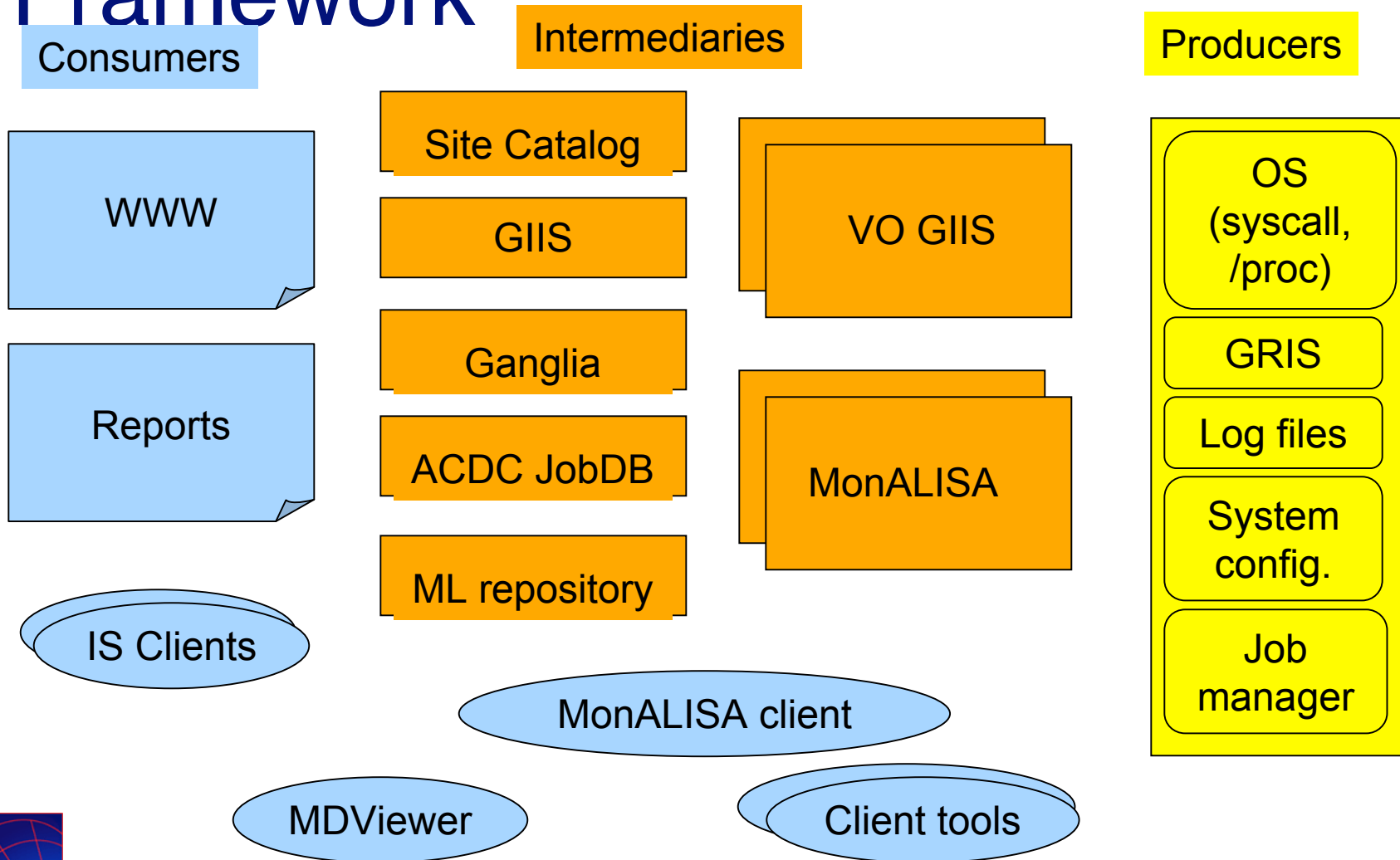- MDS Grid3 Schema development & Information Provider verification
- MDS tuning

Grid3 Index Service	VO Index Service (6)	Grid3 Site Resources
	ATLAS GIIS	Boston GRIS
GRID3 GIIS	Boston U	GRID3 Location
ANL	Chicago	Grid3 Data_DIR
BNL	ANL	
Boston U	BNL	
Chicago		UFL GRIS
UFL	**USCMS GIIS**	Grid3 Location
FNAL	UFL	Grid3 Data DIR
RiceU	FNAL	Grid3 Applications
CalTech	RiceU	Grid3 Temporary DIR
	CalTech	

Monitoring Services and Tools

- ## Globus MDS

  - GLUE and Grid3 schema based information services

- ## Ganglia

  - cluster monitoring information such as CPU and network load, memory and disk usage

- ## MonALISA

  - Agent based monitoring gathering system

- ## ACDC Job Monitoring System

  - Grid submitted jobs to query the job managers and collect information

- ## Metrics Data Viewer (MDViewer)

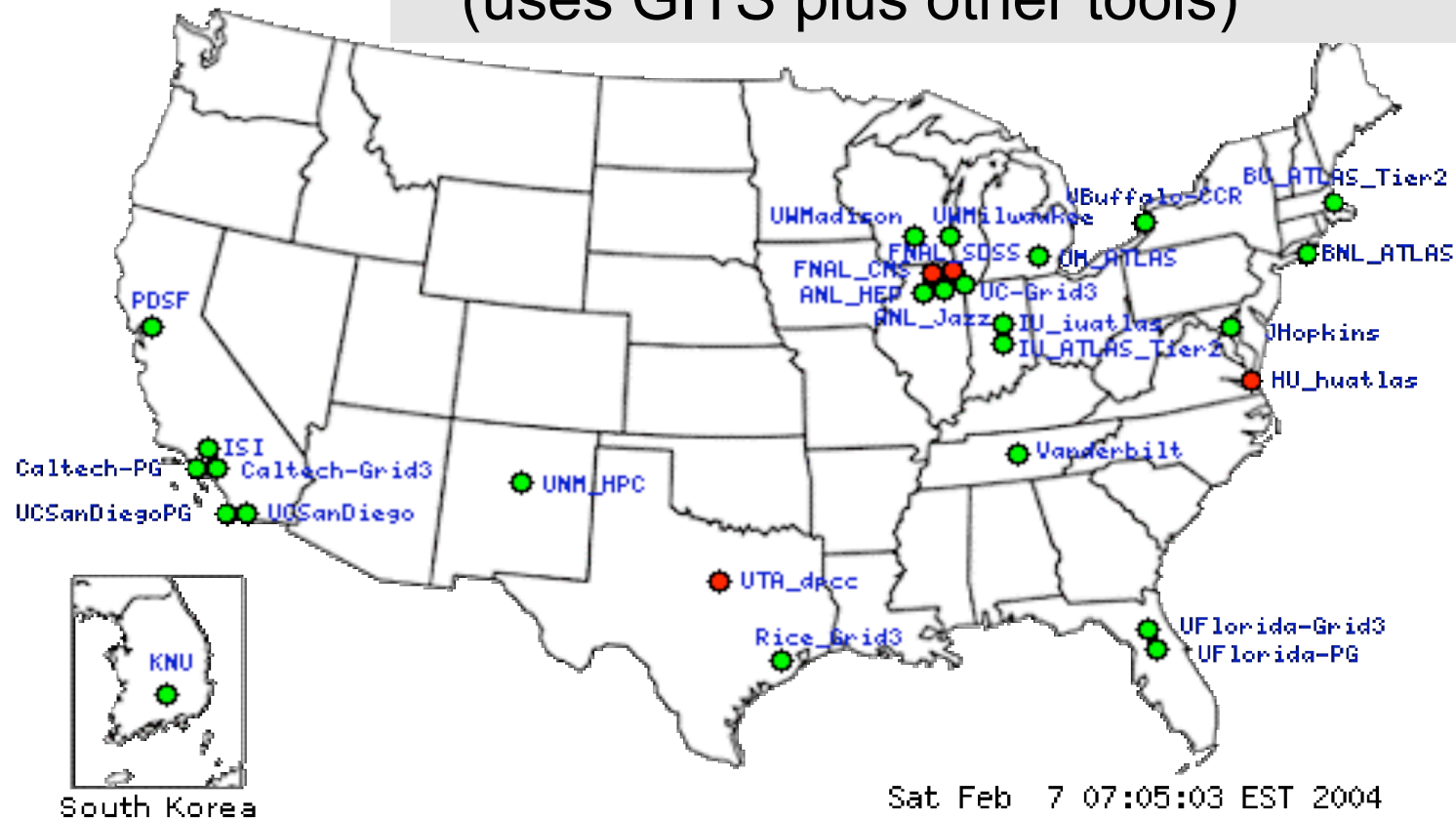 - Plots information collected by the different monitoring systems

Collective Monitoring Framework

Consumers

Intermediaries

Producers

WWW

Reports

IS Clients

Site Catalog

GIIS

Ganglia

ACDC JobDB

ML repository

VO GIIS

MonALISA

MonALISA client

MDViewer

Client tools

OS (syscall, /proc)

GRIS

Log files

System config.

Job manager

Site Catalog Tool

periodic validation checks & status display
(uses GITS plus other tools)

iVDGL Operations Center (iGOC)

- Co-located with Abilene NOC, hosted by Indiana University
- Hosts/manages multi-VO services:
  - top level Ganglia collectors
  - MonALISA web server and archival service
  - top level GIIS
  - VOMS servers for iVDGL, BTeV, SDSS
  - Site Catalog service
  - iVDGL Pacman caches
- Trouble ticket systems
  - phone (24 hr), web and email based collection and reporting system
  - Investigation and resolution of grid middleware problems at the level of 16-20 contacts per week
- Weekly operations meetings for troubleshooting

Scientific Applications

- Rule: install through grid jobs, encourage self-contained environments

- 11 applications
  - LHC (ATLAS, CMS)
  - Astrophysics (LIGO/SDSS)
  - Biochemical
    - Molecular X-ray diffraction
    - GADU/Gnare: compares protein sequences
  - Computer Science
    - Adaptive data placement and scheduling algorithms
    - Grid Exerciser
  - Over 100 users authorized to run on Grid3
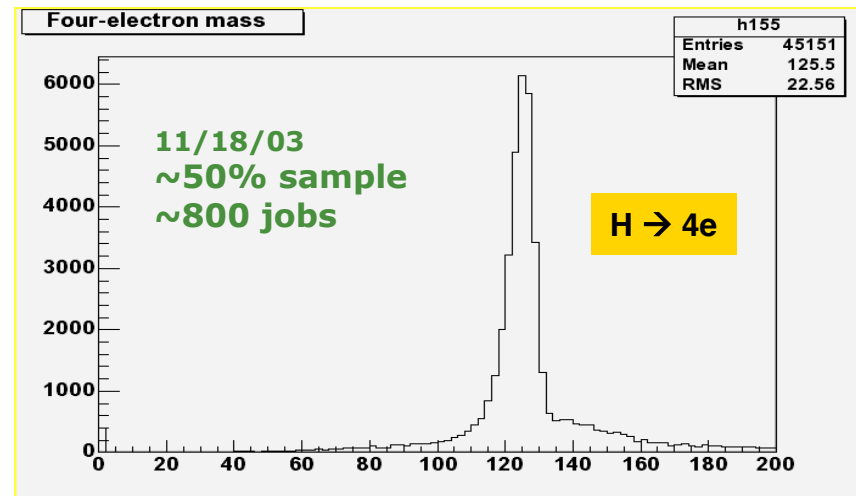
Physics Workflow on Grid3

- **Higgs → 4 lepton sample**
  - Simulate and reconstruct
  - 2000 jobs (X 6 subjobs); 100~200 events per job

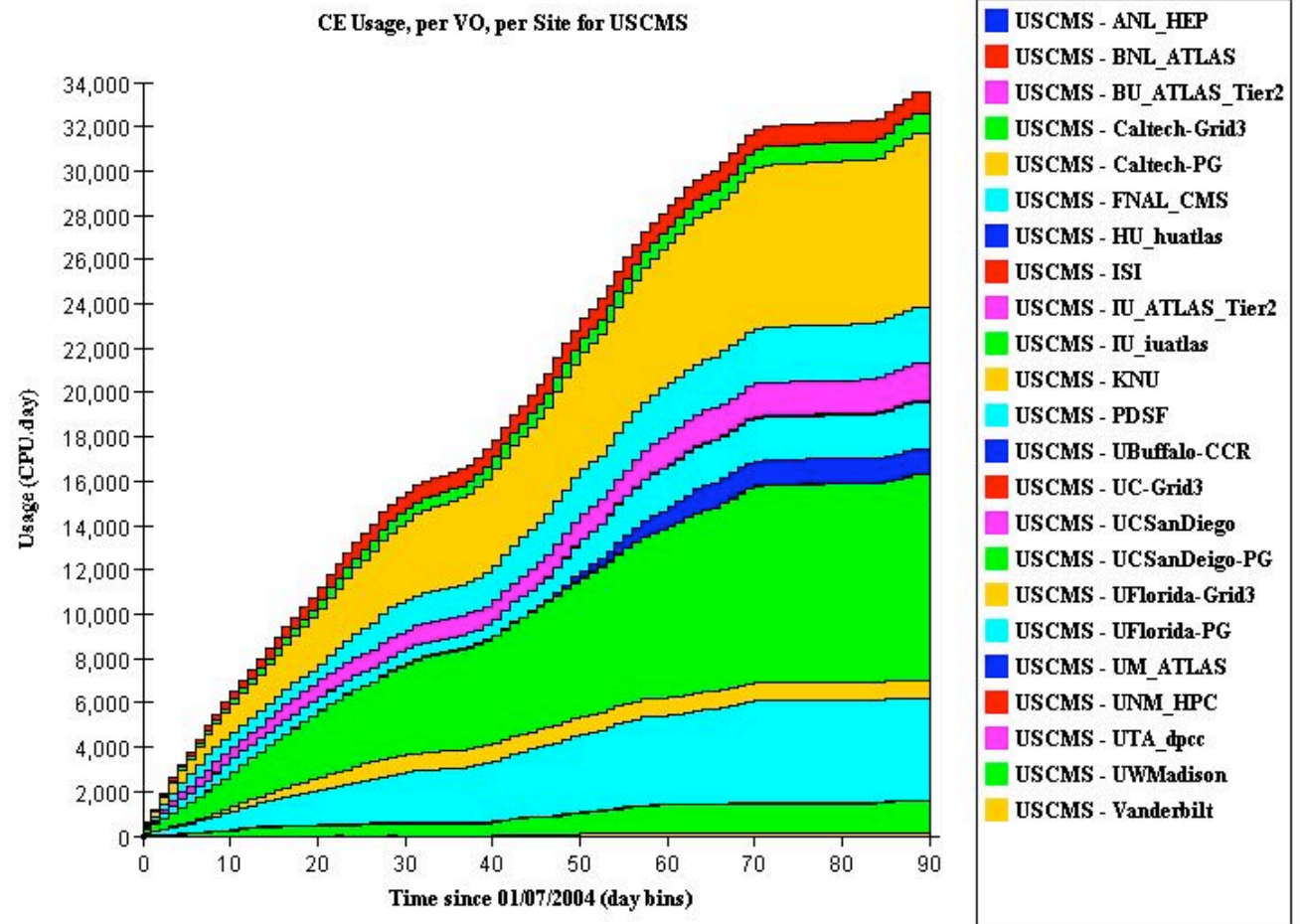- **Workflow**
  - Jobs defined, stored and registered in RLS
  - Chimera, Pegasus, and Condor-G to submit to Grid3
  - Data pushed to SE at Brookhaven, registered RLS
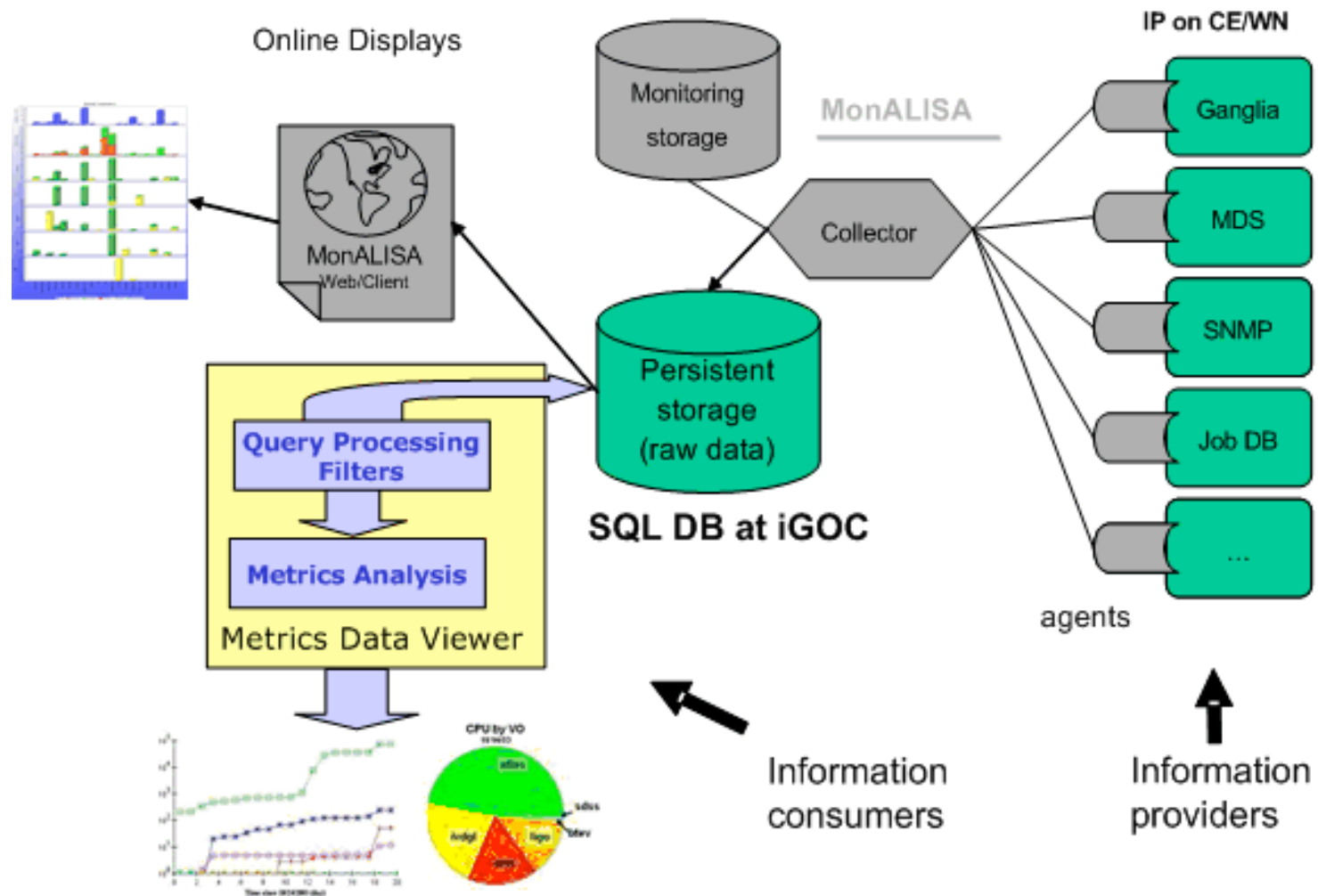 - Datasets assembled, passed to physicist for analysis

- **Grid3 resources used**
  - 16 sites, ~1500 CPUs exercised; peak ~800 jobs over three week period

Four-electron mass

h155
Entries   45151
Mean   125.5
RMS   22.56

11/18/03
~50% sample
~800 jobs

H → 4e

Large Scale Grid3 Operations

- **USCMS DC04**

- **First three months of 2004**

- **~40% of non-CMS resources**

CE Usage, per VO, per Site for USCMS

Legend:
- USCMS - ANL_HEP
- USCMS - BNL_ATLAS
- USCMS - BU_ATLAS_Tier2
- USCMS - Caltech-Grid3
- USCMS - Caltech-PG
- USCMS - FNAL_CMS
- USCMS - HU_huatlas
- USCMS - ISI
- USCMS - IU_ATLAS_Tier2
- USCMS - IU_iuatlas
- USCMS - KNU
- USCMS - PDSF
- USCMS - UBuffalo-CCR
- USCMS - UC-Grid3
- USCMS - UCSanDiego
- USCMS - UCSanDeigo-PG
- USCMS - UFlorida-Grid3
- USCMS - UFlorida-PG
- USCMS - UM_ATLAS
- USCMS - UNM_HPC
- USCMS - UTA_dpcc
- USCMS - UWMadison
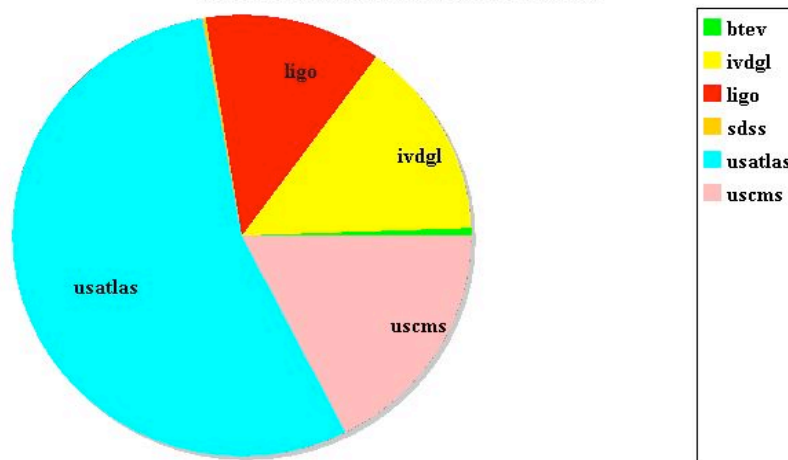- USCMS - Vanderbilt

Usage (CPU.day) vs. Time since 01/07/2004 (day bins)

Grid3 Metrics (1)

■ Goal: associate activity with VOs and sites

Grid3 Metrics (2)

Average Number of CPUs per VO, 3month

Legend:
- btev
- ivdgl
- ligo
- sdss
- usatlas
- uscms

Usage: Running Jobs

- USCMS
- iVDgL
- SDSS
- ATLAS
- BTeV
- LIGO

Y-axis: Usage (CPU)
X-axis: Time since 10/26/2003 (day bins)

- **Hardware resources**
  - Total of 2762 CPUs
    - Maximum CPU count
    - Off project > 60%
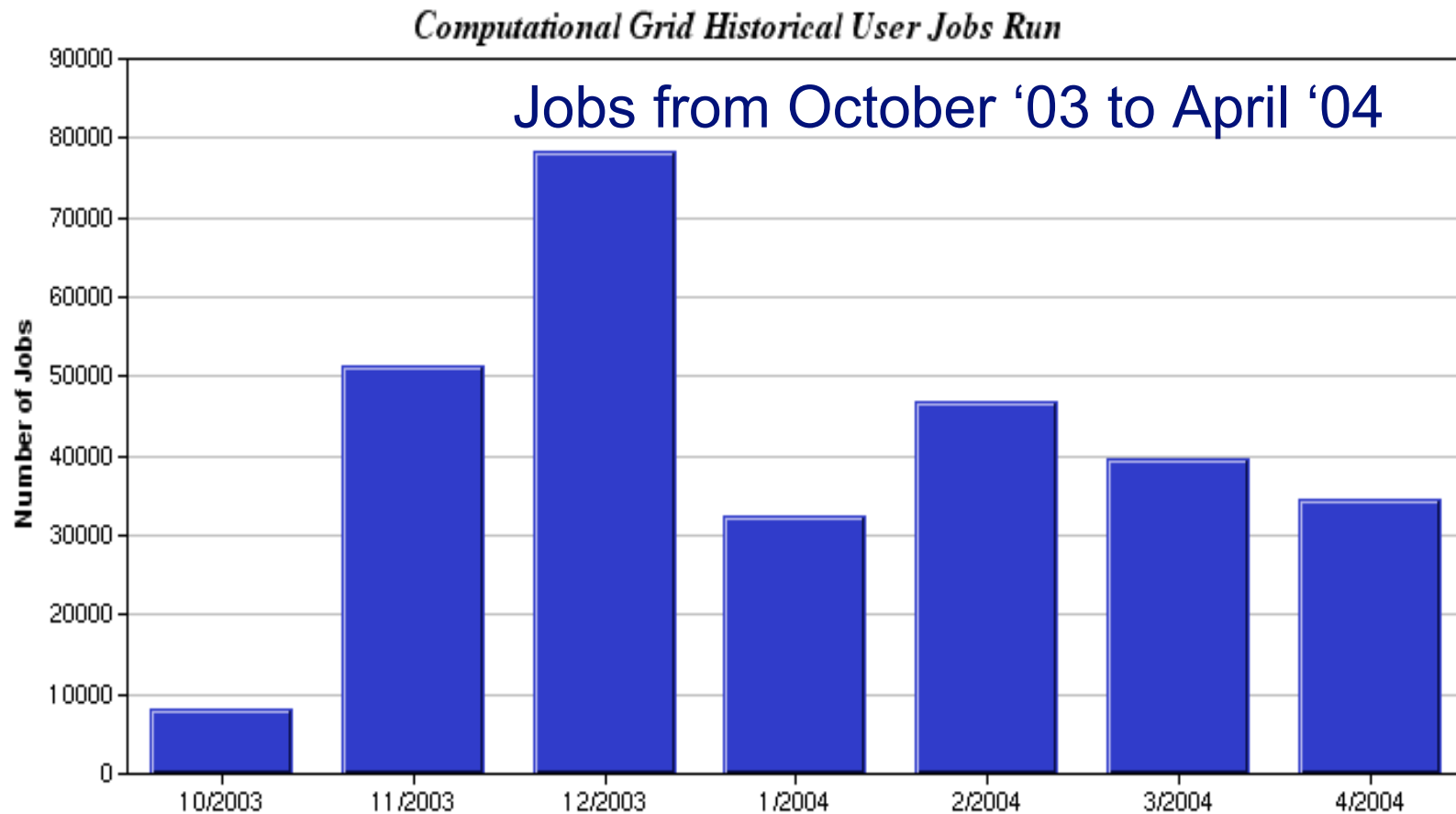 - Total of 27 sites
    - 27 administrative domains
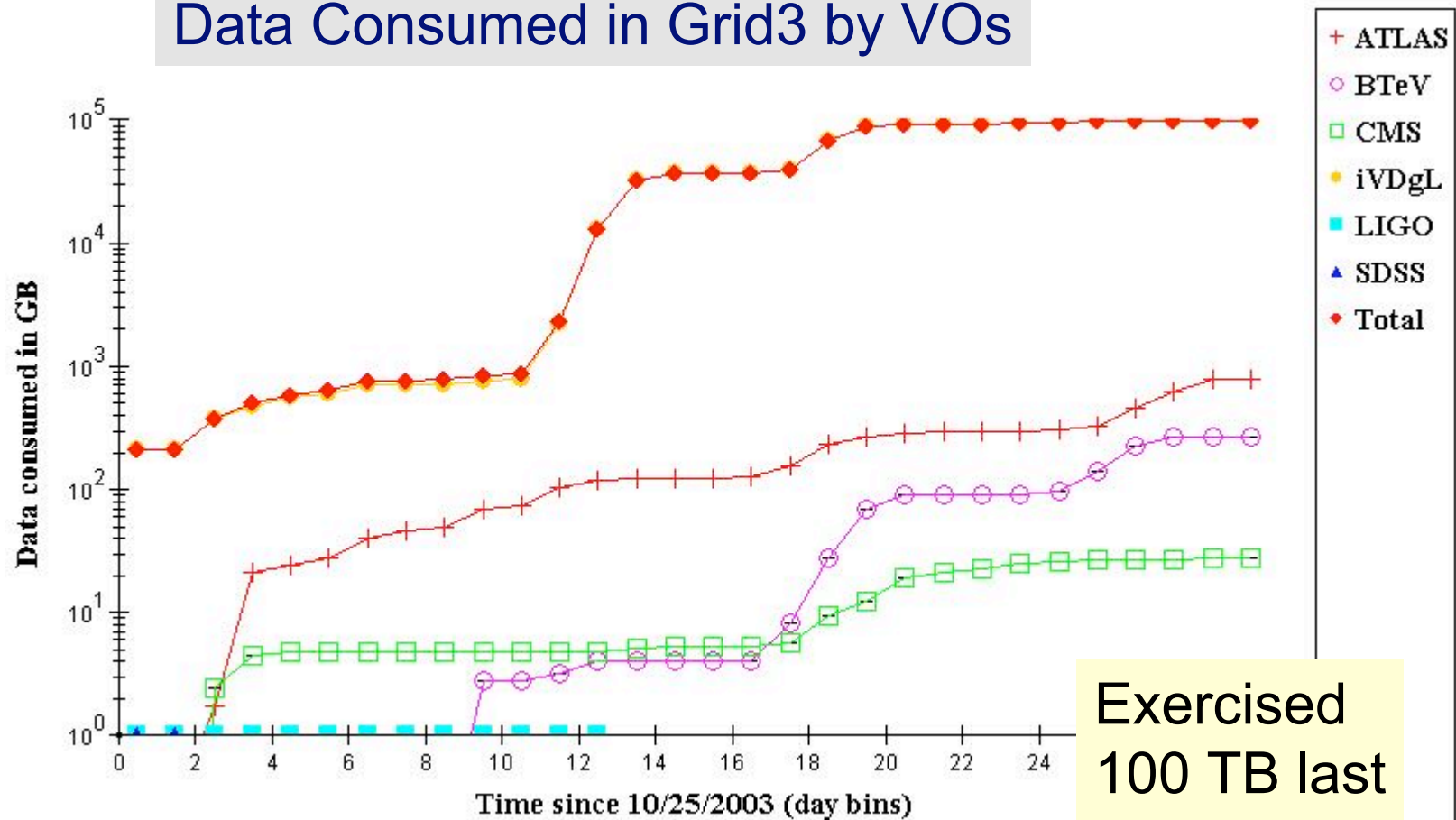   - US and Korea
- **Running jobs**
  - Peak number of jobs 1100
  - During SC2003 various applications were running simultaneously across various Grid3 sites
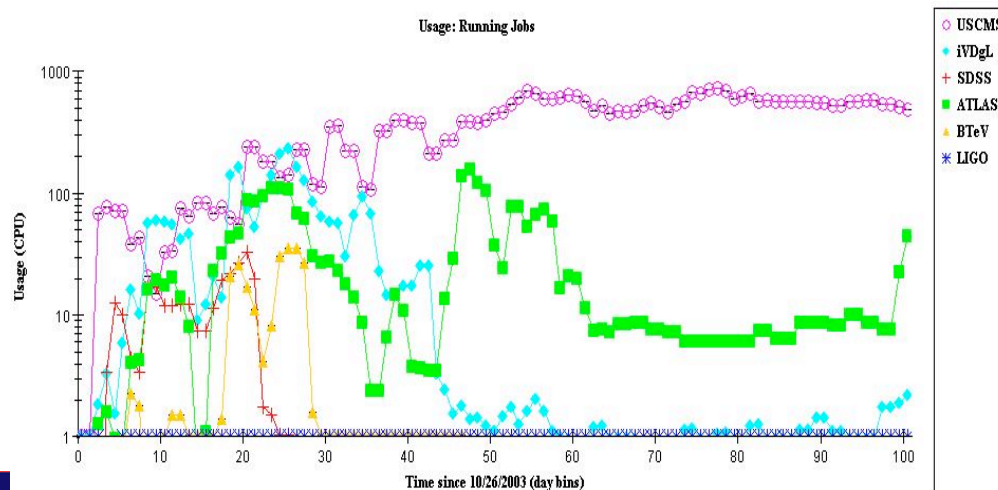
Grid3 Metrics (3)

Computational Grid Historical User Jobs Run
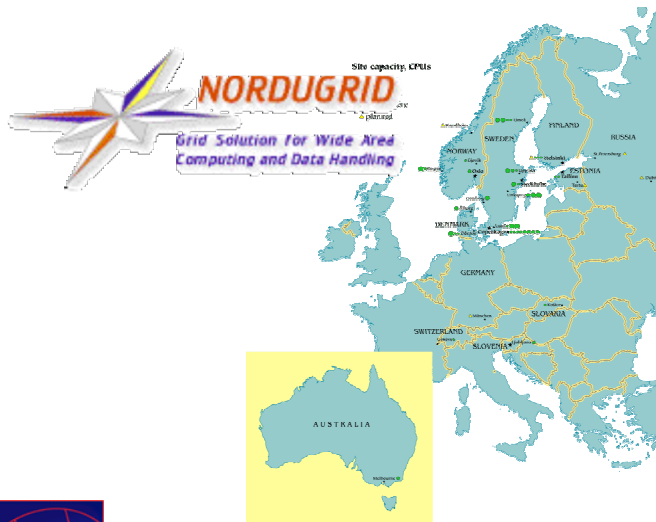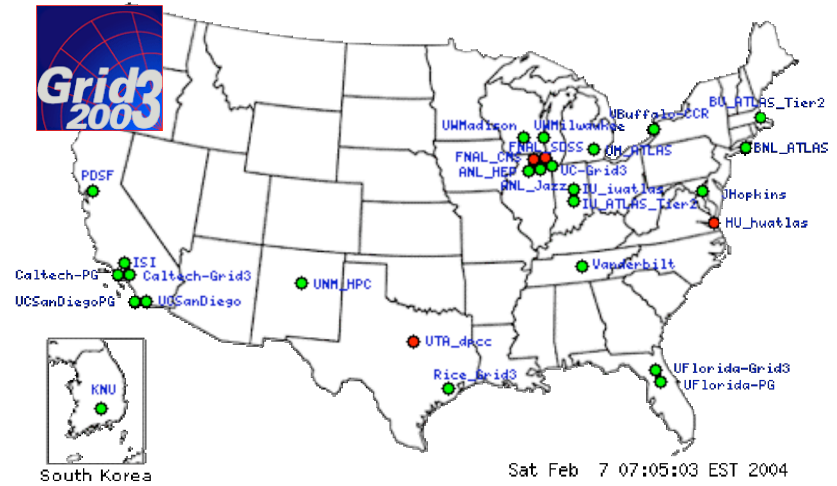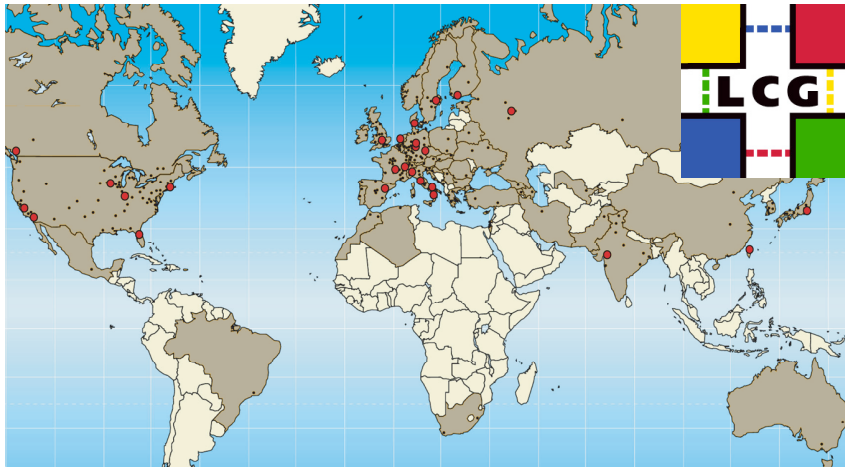
Jobs from October '03 to April '04

Grid3 Metrics (4)

Data Consumed in Grid3 by VOs

Exercised
100 TB last
November

Federation of Grids

Few thoughts on Federation

- First and foremost, scientists belong to VOs which span Grids…
  - …an LHC scientist may wish (and be authorized to) access resources from the LHC Computing Grid (LCG), UK Grid, NorduGrid, APGrid, participating US Grids like Grid3 and its successors (i.e. Open Science Grid)
  - US LHC physicists may wish to gain opportunistic access to other resources like TeraGrid, campus-based Grids, etc
- What is the best way to enable this? Many issues.
 - authorization, accounting, catalogs, (workflow)

Workflow between Grids

- Workflow exercise Grid3 and LCG
    - Grid3 produced output staged on disk at Brookhaven lab in NY
    - Use VDT client to submit to an LCG server (bypass RB)
    - Jobs executing on LCG:
        - Input files registered at BNL-RLI
        - Stage data from BNL to local scratch area on LCG grid
        - Run Athena-reconstruction on LCG
        - Write data to local storage element on LCG
        - Copy back to disk cache at NY, RLS register
        - 3rd-party transfer between LCG SE and BNL gridftp servers
    - Post job: Dataset catalog and analysis

- Many lessons learned about site architectural differences, service configurations; *Concrete steps towards combined use of LCG and US Grid3*

ATLAS Production System for DC2
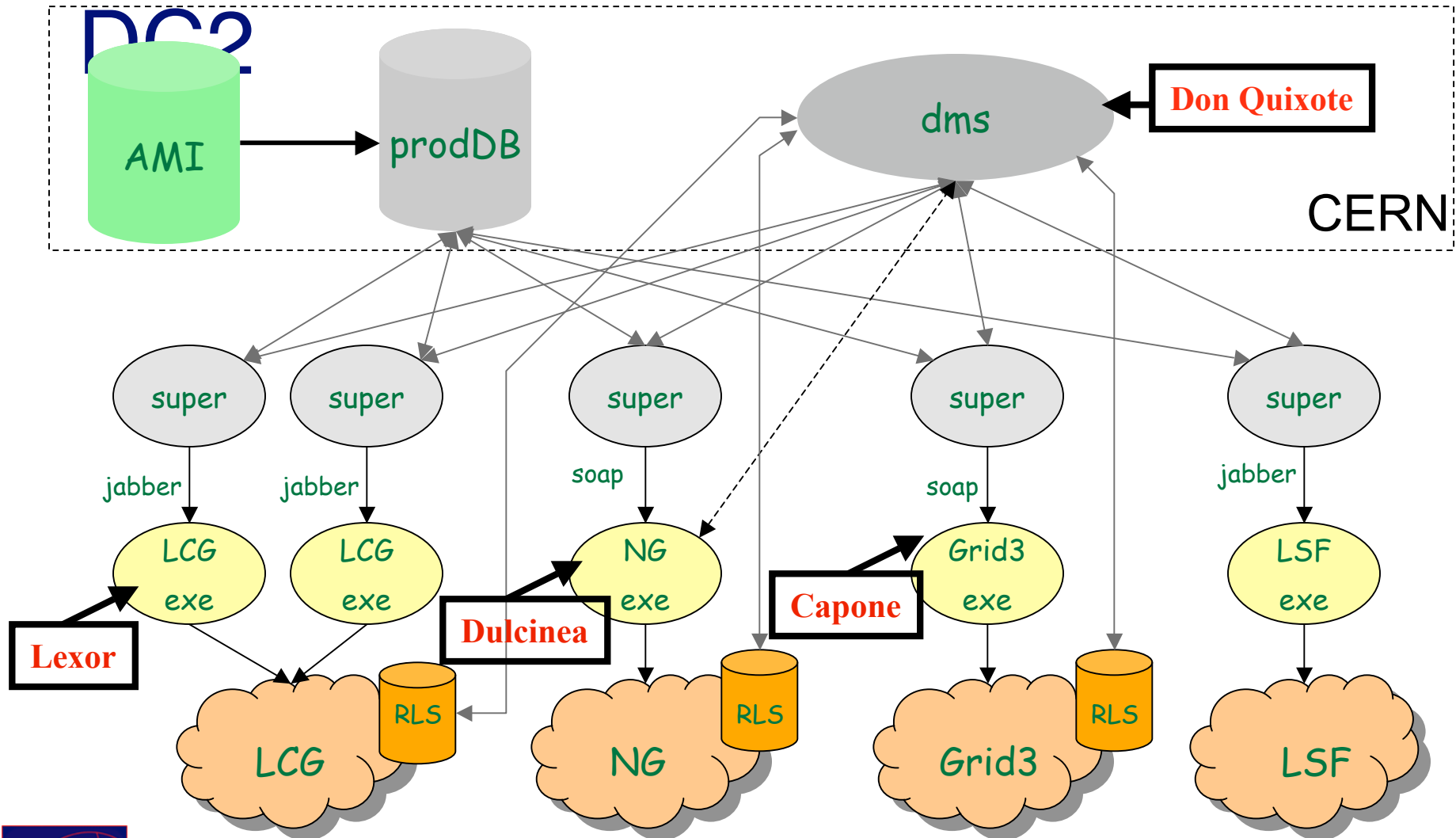
system implemented, production starting this week

Conclusions

- Grid2003 project created a persistent, production scale grid infrastructure, Grid3

- Continuous use since 10/03

- Grid3 will undergo adiabatic upgrades, and continue operations for 2004 (Grid3+)

- Outlook:
    - Follow-on project will involve more sites and VOs
    - Larger "consortia" forming, Open Science Grid (OSG)
    - New projects will be defined in this larger framework

Acknowledgements

- the entire Grid2003 team which did all the work!
  - site administrators
  - application developers
  - iGOC team
  - coordinators and project teams
- Slides I've taken from presentations by:
  - Leigh Grundhoefer (Indiana, iVDGL)
  - Ian Fisk (Fermilab, USCMS)
  - Ruth Pordes (Fermilab, PPDG)
  - Jorge Rodriguez (Florida, USCMS/iVDGL)