Enabling Grids for E-science in Europe

www.eu-egee.org

Cern School of Computing, Aug 30 – Sep 10 2004

EGEE Applications
The Future EGEE Middleware

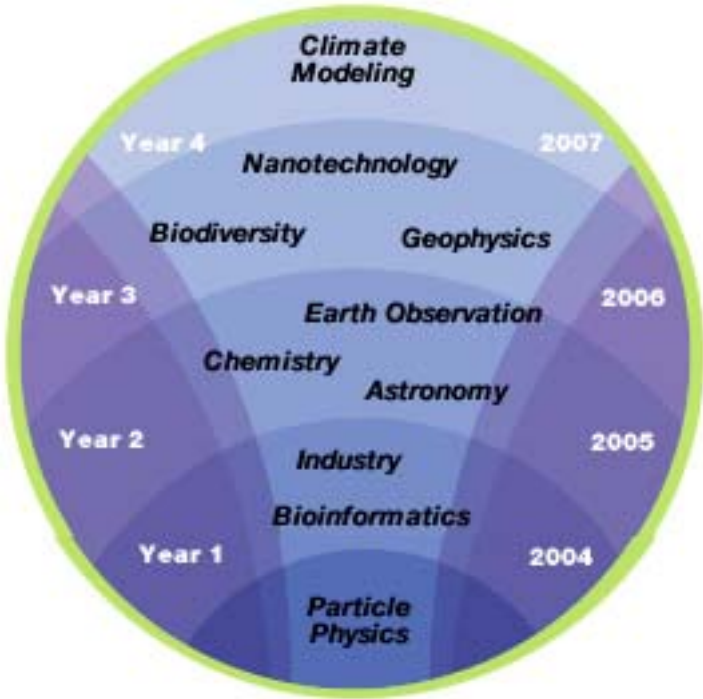
Erwin Laure
EGEE Deputy Middleware Manager

gLite

EGEE is a project funded by the European Union under contract IST-2003-508833

EGEE Applications

- EGEE Scope : ALL-Inclusive for academic applications (open to industrial and socio-economic world as well)
- The major success criterion of EGEE: how many satisfied users from how many different domains ?
- 5000 users (3000 after year 2) from at least 5 disciplines
- Two pilot applications selected to guide the implementation and certify the performance and functionality of the evolving infrastructure: **Physics & Bioinformatics**

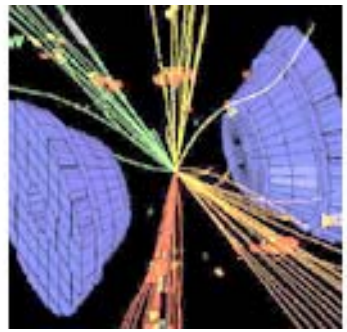


Application domains and timelines are for illustration only

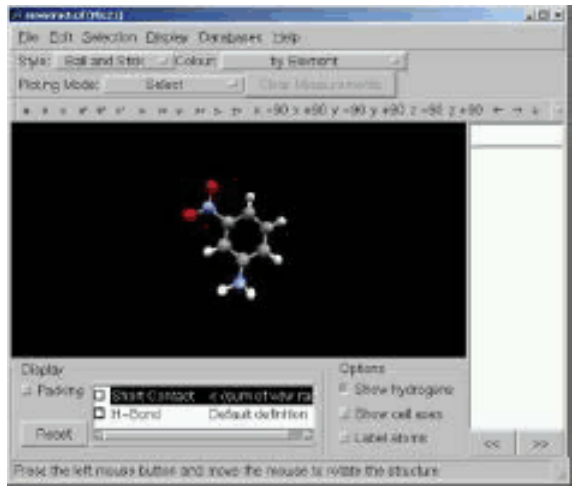
The pilot applications

Enabling Grids for E-science in Europe

- **High Energy Physics (HEP)** with LHC Computing Grid (www.cern.ch/lcg) relies on a Grid infrastructure to store and analyse petabytes (10^{15} bytes) of real and simulated data. LCG is a major source of resources, requirements and a hard deadlines with no conventional solution available



- In **Biomedics** several communities are facing equally daunting challenges to cope with the flood of bioinformatics and healthcare data. Need to access large and distributed non-homogeneous data and important on-demand computing and security requirements



Biomedical Applications

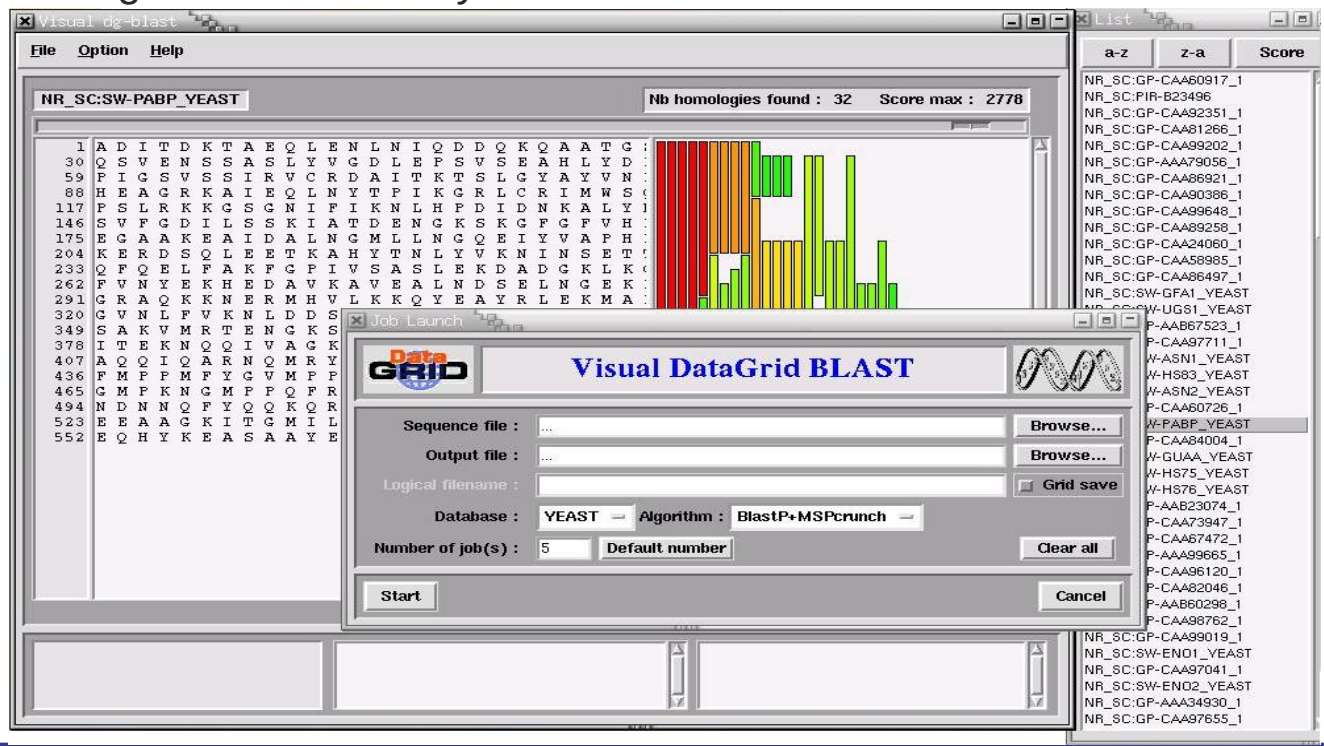
Enabling Grids for E-science in Europe

- Some key applications and their characteristics
  - Bioinformatics: gene/proteome databases distributions
  - Medical applications (screening, epidemiology...): image databases distribution
  - Parallel algorithms for medical image processing, simulation, etc
  - Interactive application (human supervision or simulation)

  - Security/privacy constraints
  - Heterogeneous data formats (genomics, proteomics, image formats)
  - Frequent data updates
  - Complex data sets (medical records)
  - Long term archiving requirements

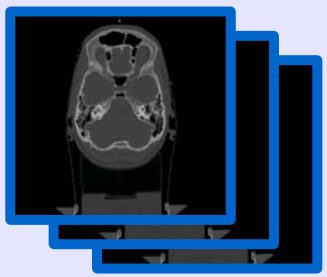
BLAST – comparing DNA or protein sequences

- BLAST is the first step for analysing new sequences: to compare DNA or protein sequences to other ones stored in personal or public databases. Ideal as a grid application.
  - Requires resources to store databases and run algorithms
  - Can compare one or several sequence against a database in parallel
  - Large user community



Monte carlo simulation for radiotherapy planning

Enabling Grids for E-science in Europe



Scanner slices: DICOM format

Concatenation

Image: text file

Binary file: Image.raw Size 19M

Retrieving of root output files from CEs

the CE

User interface

Anonymisation

Database

GATE

Computing Element

Storage Element

CCIN2P3

GATE

Computing Element

Storage Element

RAL

GATE

Computing Element

Storage Element

NIKHEF

GATE

Computing Element

Storage Element

MARSEILLE

LHC Experiments

Enabling Grids for E-science in Europe

ATLAS

CMS

ALICE

LHCb

Storage –

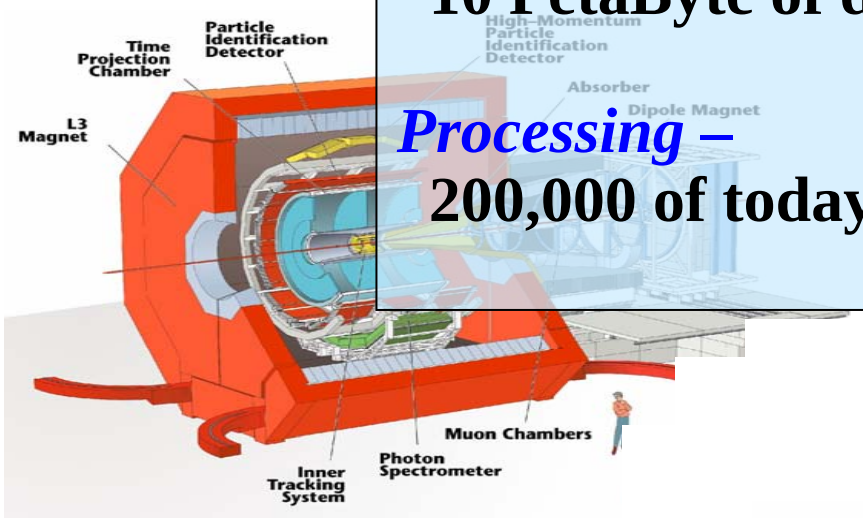
Raw recording rate 0.1 – 1 GByte/s

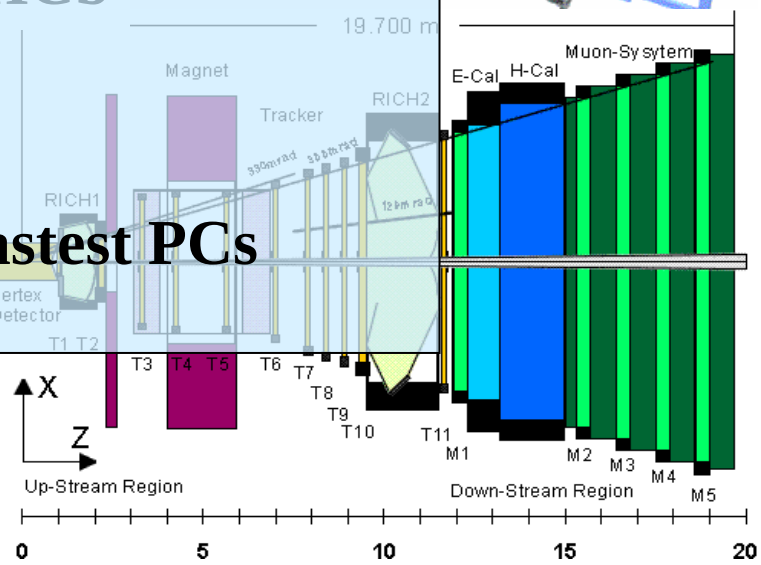
Accumulating at 5-8 PetaByte/year

10 PetaByte of disk

Processing –

200,000 of today's fastest PCs





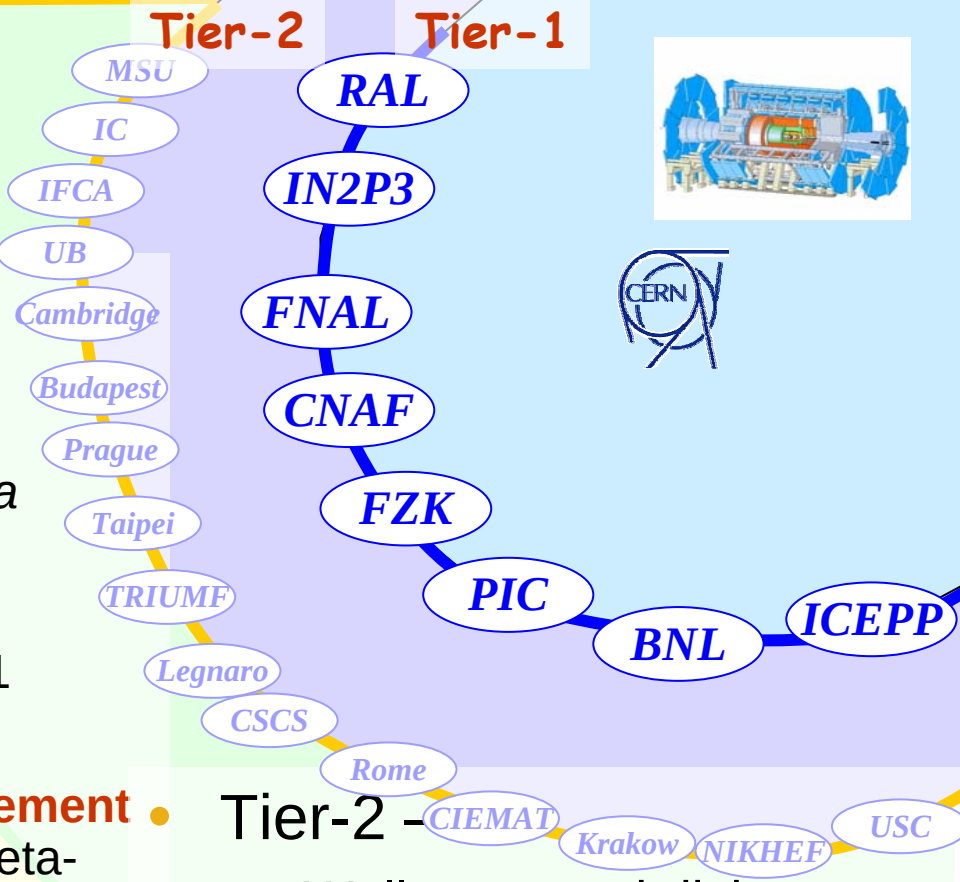
HEP Applications and the Grid

Enabling Grids for E-science in Europe

- Have been running large distributed computing systems for many years
- Now the focus for the future is on computing for LHC and hence we have the LCG (LHC computing grid project)
- In addition to the 4 LHC experiments (ATLAS,ALICE,CMS,LHCb) other current HEP experiments use grid technology e.g. Babar,CDF,D0.., and don't forget Theory and other new HEP experiments..
- LHC experiments are currently executing large scale data challenges(DCs) involving thousands of processors world-wide and generating many Terabytes of data
- Moving to so-called 'chaotic' use of grid with individual user analysis (thousands of users interactively operating within experiment VOs)

desktops portables

small centres



LHC Computing Model (simplified!!)

- Tier-0 – the accelerator centre
  - Filter → *raw data*
  - Reconstruction → *summary data* (ESD)
  - Record *raw data* and *ESD*
  - Distribute *raw* and *ESD* to Tier-1
- Tier-1 –
  - Permanent storage and **management** of *raw*, *ESD*, calibration data, meta-data, analysis data and databases → **grid-enabled data service**
  - Data-heavy analysis
  - Re-processing raw → ESD
  - National, regional support

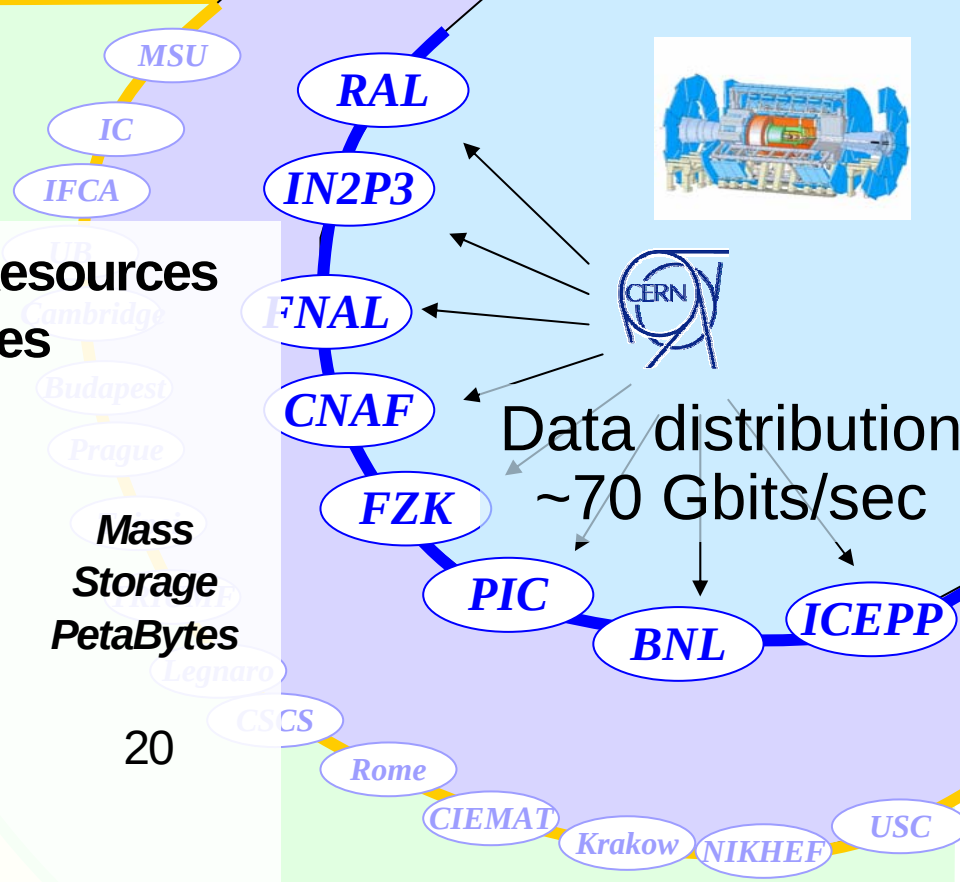
  "online" to the data acquisition process high availability, long-term commitment managed mass storage
- Tier-2 –
  - Well-managed disk storage – grid-enabled
  - Simulation
  - End-user analysis – batch and interactive
  - High performance parallel analysis (PROOF)

Current estimates of Computing Resources needed at Major LHC Centres

First full year of data - 2008

	*Processing M SI2000***	*Disk PetaBytes*	*Mass Storage PetaBytes*
CERN	20	5	20
Major data handling centres (Tier 1)	45	20	18
Other large centres (Tier 2)	40	12	5
Totals	**105**	**37**	**43**

*** Current fast processor ~1K SI2000*



Characteristics of CMS Data Challenge DC04 (just completed)……run with LCG-2 and CMS resources world-wide (US Grid3 was a major component)

Enabling Grids for E-science in Europe

- **Pre-Challenge Production (Phase 1) – simulation generation and digitisation**
  - After 8 months of continuous running:
    - 750,000 jobs
    - 3,500 KSI2000 months
    - 700,000 files
    - 80 TB of data

- **Data Challenge (Phase 2)**
  - **Ran the full data reconstruction and distribution chain at 25 Hz**
  - **Achieved**
    - 2,200 jobs/day (about 500 CPU's) running at Tier-0
    - Total 45,000 jobs Tier-0 and 1
    - 0.4 files/s registered to RLS (with POOL metadata)
    - Total 570,000 files registered to RLS
    - 4 MB/s produced and distributed to each Tier-1

LHC Computing Grid Project
A Collaboration

Enabling Grids for E-science in Europe

Building and operating the LHC Grid – a collaboration between

- The physicists and computing specialists from the LHC experiment
- The projects in Europe and the US that have been developing Grid middleware
- The regional and national computing centres that provide resources for LHC
- The research networks

Researchers

Software Engineers

Service Providers

More Information

- NA4 – Application Identification and Support
  - http://egee-na4.ct.infn.it/index.php

Future EGEE middleware

Enabling Grids for E-science in Europe

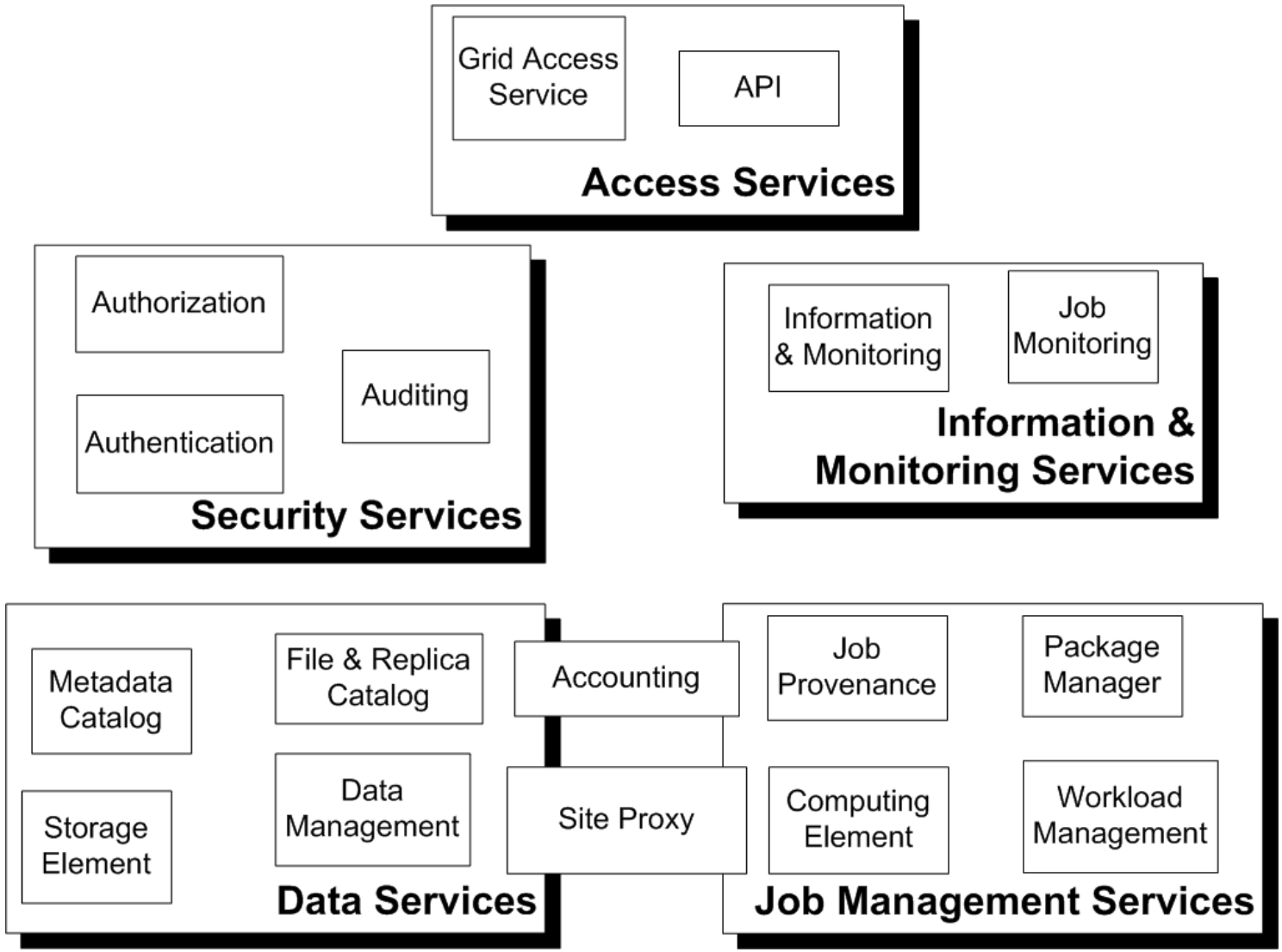
Architecture Guiding Principles

Enabling Grids for E-science in Europe

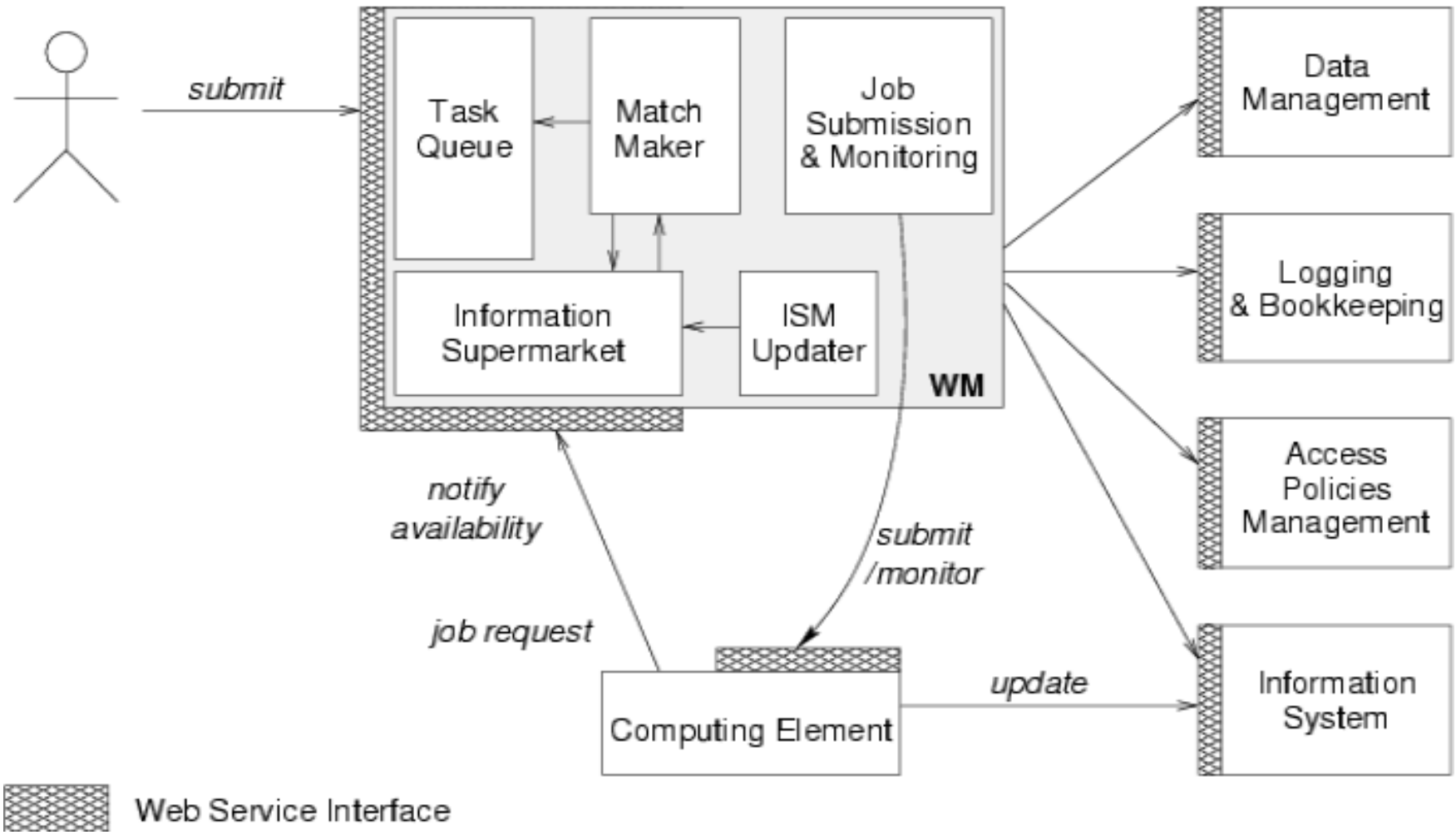
- Lightweight (existing) services
  - Easily and quickly deployable
- Interoperability
  - Allow for multiple implementations
- Resilience and Fault Tolerance
- Co-existence with deployed infrastructure
  - Run as an application (e.g. on LCG-2; Grid3)
  - Reduce requirements on site components
    - Basically globus and SRM
  - Co-existence (and convergence) with LCG-2 and Grid3 are essential for the EGEE Grid service
- Service oriented approach
  - WSRF still being standardized
  - No mature WSRF implementations exist to date, no clear picture about the impact of WSRF hence: start with plain WS
    - WSRF compliance is not an immediate goal, but we follow the WSRF evolution
    - WS-I compliance is important

gLite Services

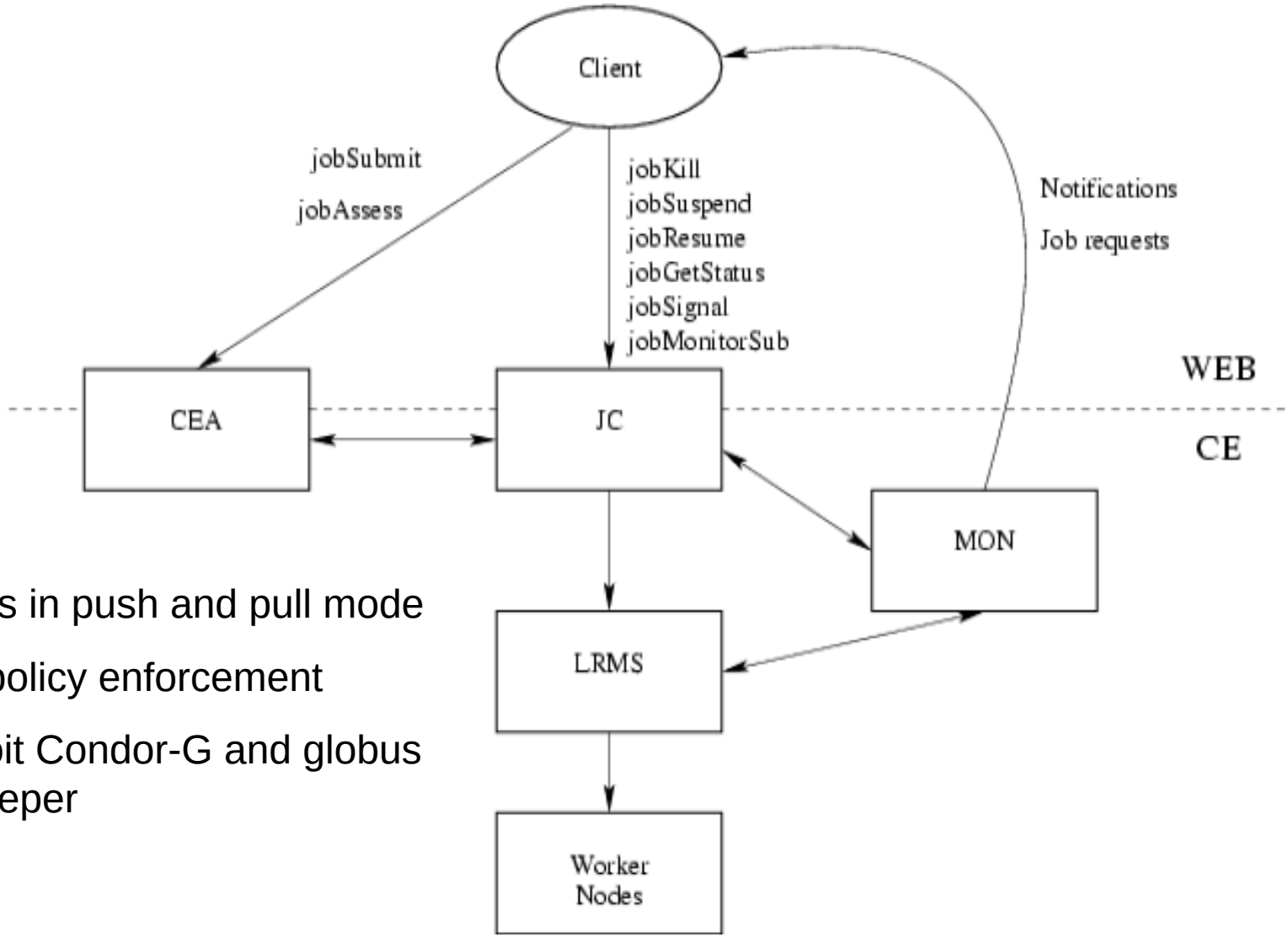
Enabling Grids for E-science in Europe



WMS


submit
Task Queue
Match Maker
Job Submission & Monitoring
Information Supermarket
ISM Updater
WM
Data Management
Logging & Bookkeeping
Access Policies Management
Information System
notify availability
submit /monitor
job request
update
Computing Element
Web Service Interface

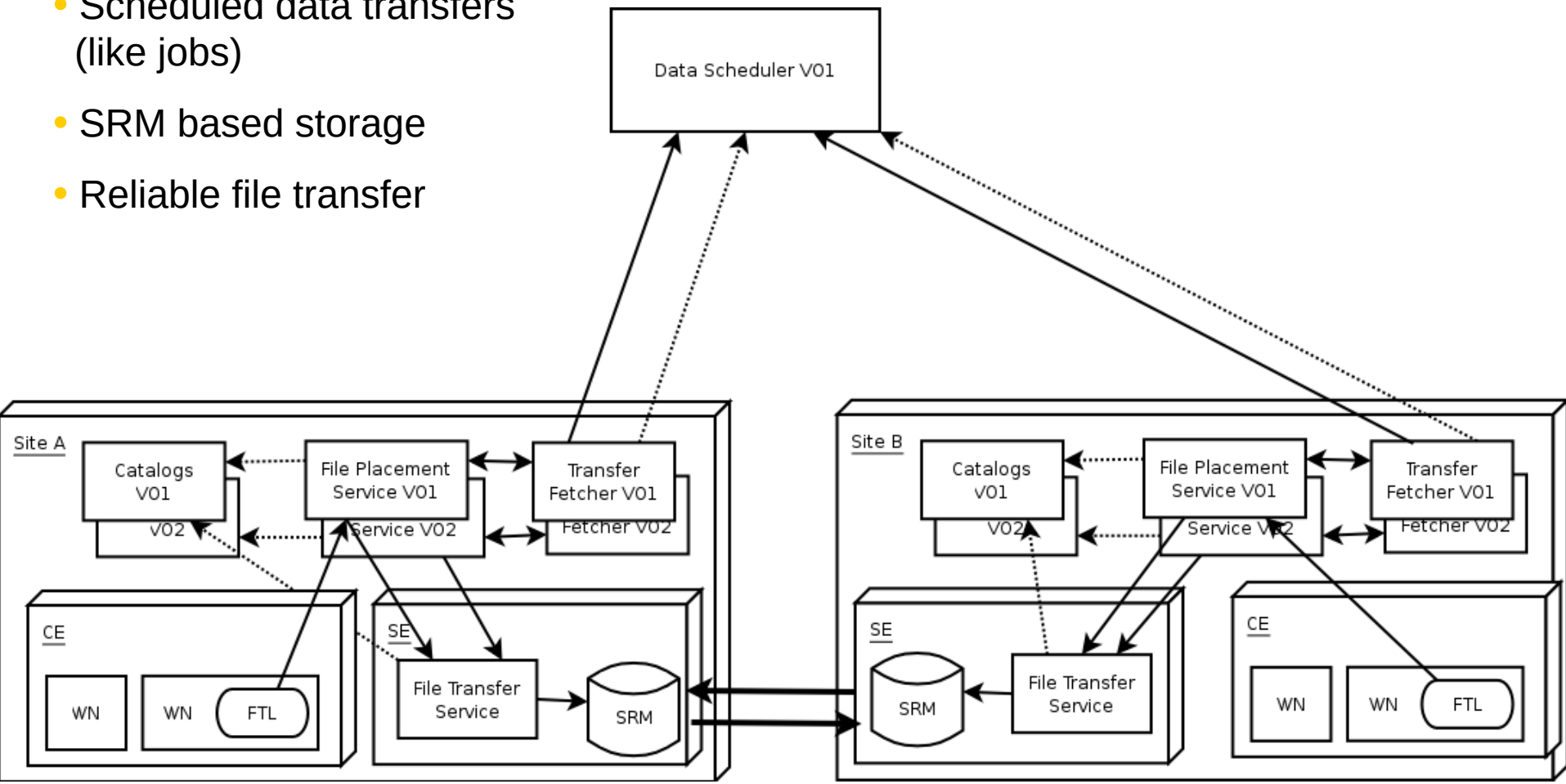
CE



- Works in push and pull mode
- Site policy enforcement
- Exploit Condor-G and globus gatekeeper

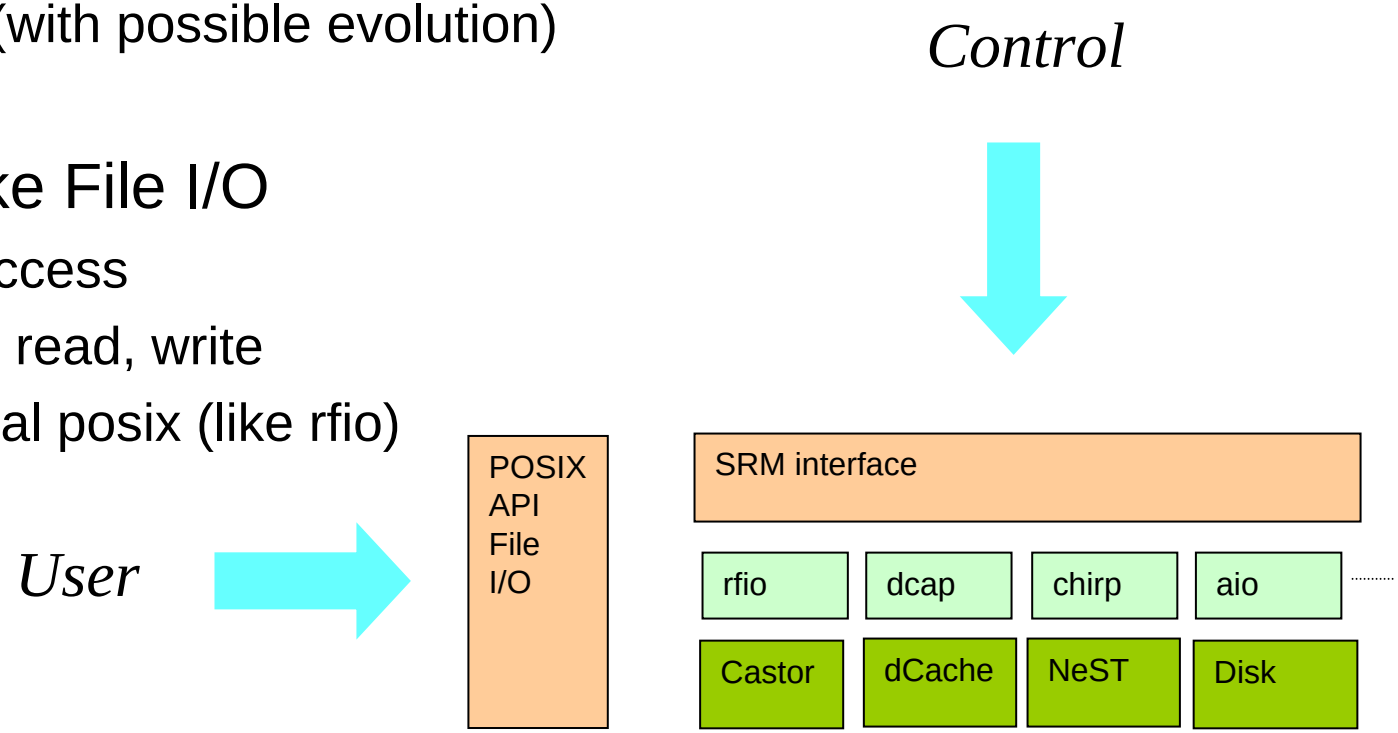
Data Management

- Scheduled data transfers (like jobs)
- SRM based storage
- Reliable file transfer



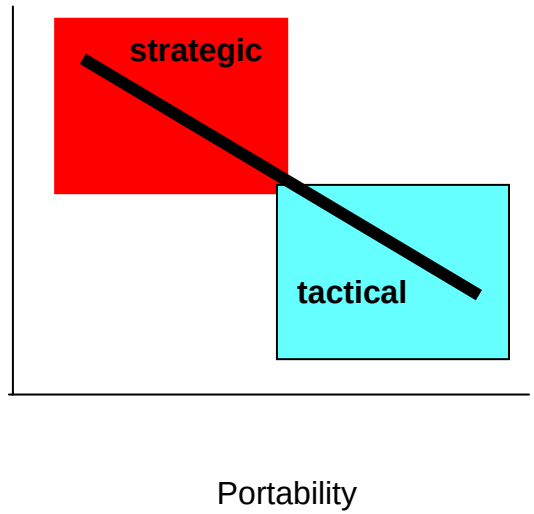
Storage Element Interfaces

- SRM interface
  - Management and control
  - SRM (with possible evolution)
- Posix-like File I/O
  - File Access
  - Open, read, write
  - Not real posix (like rfio)



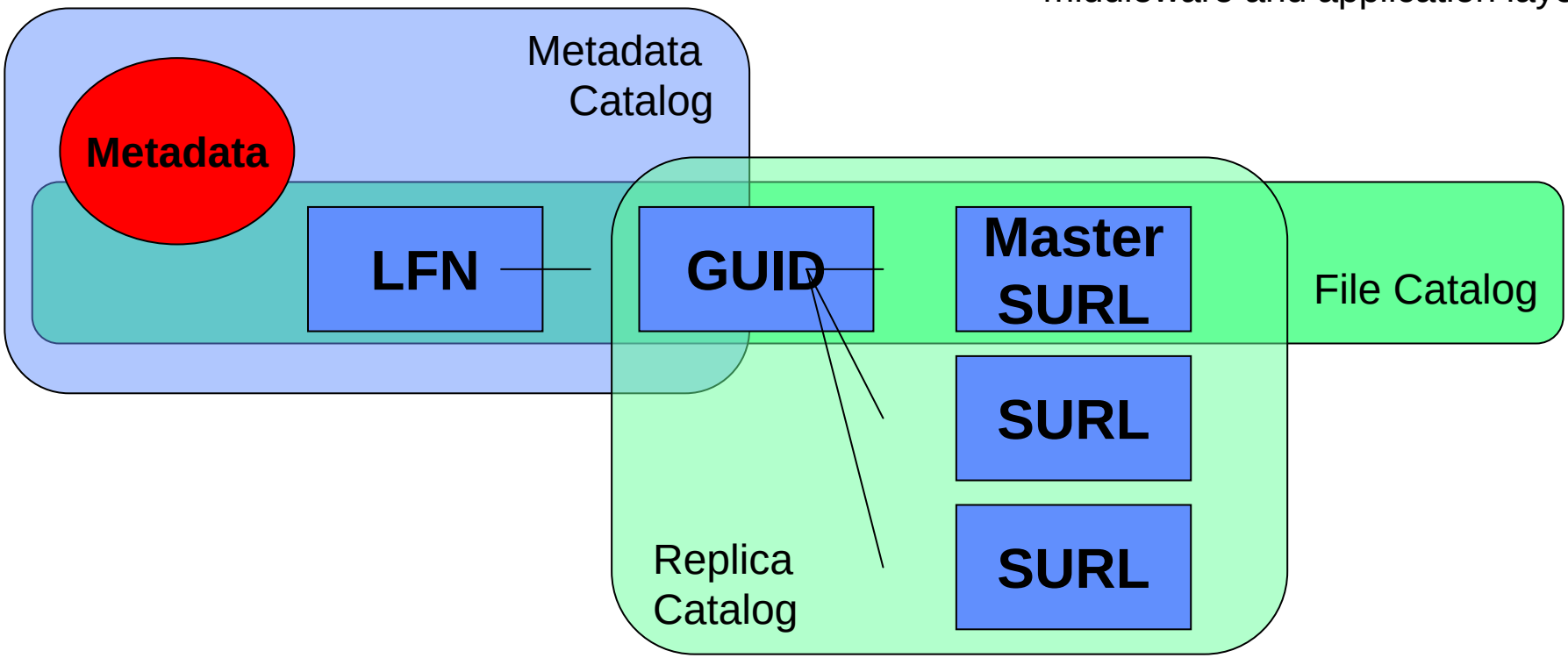
Storage Element

- ‘Strategic’ SE
  - High QoS: reliable, safe..
  - Has usually an MSS
  - Place to keep important data
  - Needs people to keep running
  - Heavyweight
- ‘Tactical’ SE
  - Volatile, ‘lightweight’ space
  - Enables sites to participate in an opportunistic manner
  - Best effort
  - Collaboration with LCG Deployment

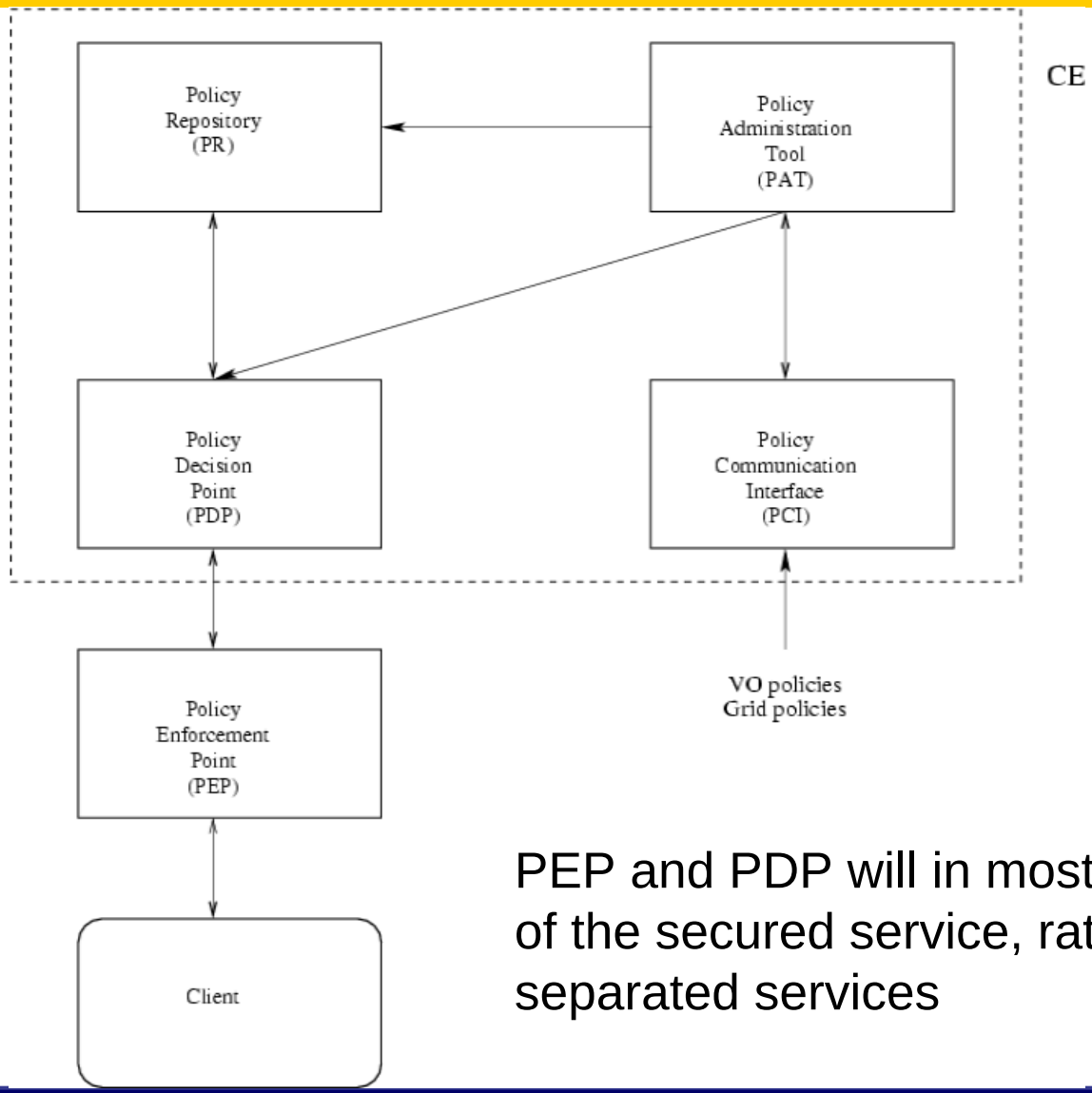


Catalogs

- File Catalog
  - Filesystem-like view on logical file names
- Replica Catalog
  - Keep track of replicas of the same file
- (Meta Data Catalog)
  - Attributes of files on the logical level
  - Boundary between generic middleware and application layer



Security

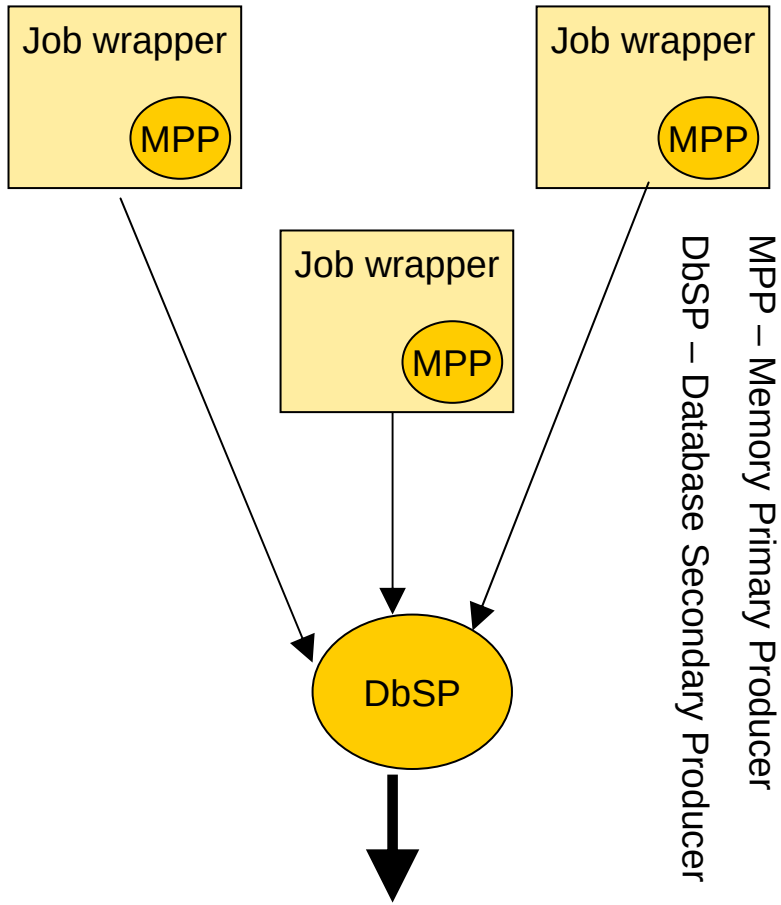


PEP and PDP will in most cases be part of the secured service, rather than separated services

Information and Monitoring

- R-GMA for
  - Information system and system monitoring
  - Application Monitoring
- No major changes in architecture
  - But re-engineer and harden the system
- Co-existence and interoperability with other systems is a goal
  - E.g. MonaLisa

e.g: D0 application monitoring:



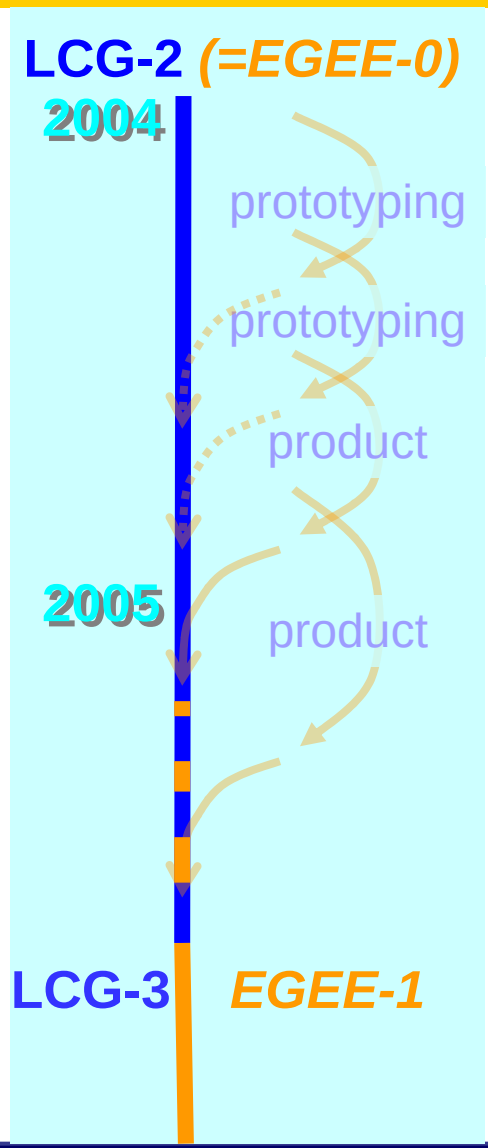
gLite and LCG-2

LCG-2

focus on production, large-scale data handling

- The service for the 2004 data challenges
- Provides experience on operating and managing a global grid service
- Development programme driven by data challenge experience
  - Data handling
  - Strengthening the infrastructure
  - Operation, VO management
- Evolves to LCG-3 as components progressively replaced with new middleware

  -- target is to minimise the discontinuities of migration to the new generation
- Aim for migration plan by end of year



gLite

focus on analysis

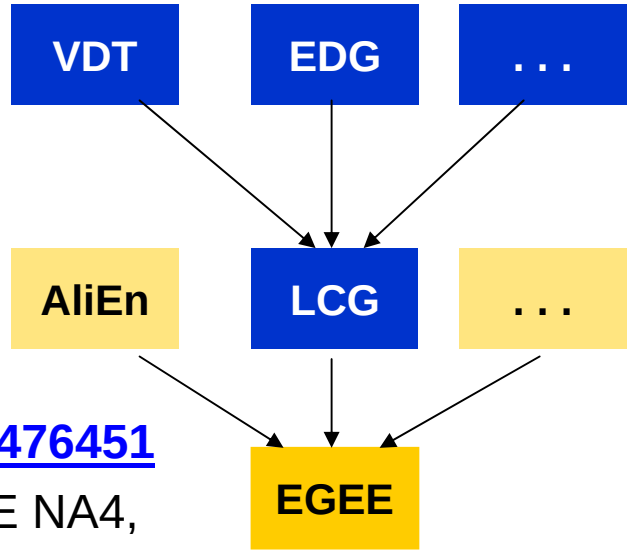
- Developed by EGEE project in collaboration with VDT (US)
- LHC applications and users closely involved in prototyping & development (ARDA project)
- Short development cycles
- Co-existence with LCG-2
- Profit as far as possible from LCG-2 infrastructure, experience

→ Ease deployment – avoid separate hardware

- As far as possible - completed components integrated in LCG-2

→ improved testing, easier displacement of LCG-2

Approach



- Exploit experience and components from existing projects
  - AliEn, VDT, EDG, LCG, and others
- Design team works out architecture and design
  - Architecture: **https://edms.cern.ch/document/476451**
  - Feedback and guidance from EGEE PTF, EGEE NA4, LCG GAG, LCG Operations, LCG ARDA
- Components are initially deployed on a prototype infrastructure
  - Small scale (CERN & Univ. Wisconsin)
  - Get user feedback on service semantics and interfaces
- After internal integration and testing components are delivered to SA1 and deployed on the pre-production service

Deployment considerations

Enabling Grids for E-science in Europe

- Interoperability and co-existence
  - Exploit different service implementations
    - E.g. Castor and dCache SRM implementations
  - Require minimal support from development environment
    - Sites required to run globus and SRM (might not be required for tactical storage)
  - Flexible service deployment
    - Multiple services running on the same physical machine (if possible)
- Platform support
  - Goal is to have portable middleware
  - Building & Integration on RHEL 3 and windows
  - Initial testing (at least 3 sites) using different Linux flavors (including free distributions)
- Service autonomy
  - User may talk to services directly or through other services (like access service)

Summary

- Next generation middleware being designed and assembled
  - Prototype first tangible outcome
    - **BUT this is a PROTOTYPE!**
  - Architectural and design work well advanced
  - Incremental changes to prototype
    - Feedback from applications and operations essential!
      - Work with HEP via ARDA started May 18
      - Work with Biomed started June 15
      - Operations feedback via pre-production service
- Detailed release plan worked out
  - **https://edms.cern.ch/document/468699**
- First components for pre-production service during summer
- Continuous integration and testing scheme defined and being adopted
- Technology Risk
  - Will WS allow for all upcoming requirements?
  - Divergence to standards

Links

- JRA1 homepage
  - http://egee-jra1.web.cern.ch/egee-jra1/
- Architecture document
  - https://edms.cern.ch/document/476451/
- Release plan
  - https://edms.cern.ch/document/468699
- Prototype installation
  - http://egee-jra1.web.cern.ch/egee-jra1/Prototype/testbed.htm