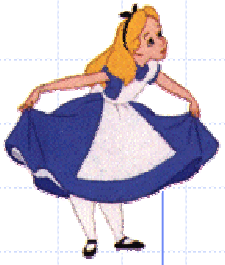


AliEn

@GRID

Predrag Buncic

P.Saiz, J-E. Revsbech
R.Piskac, V.Sego, L. Aphecetche

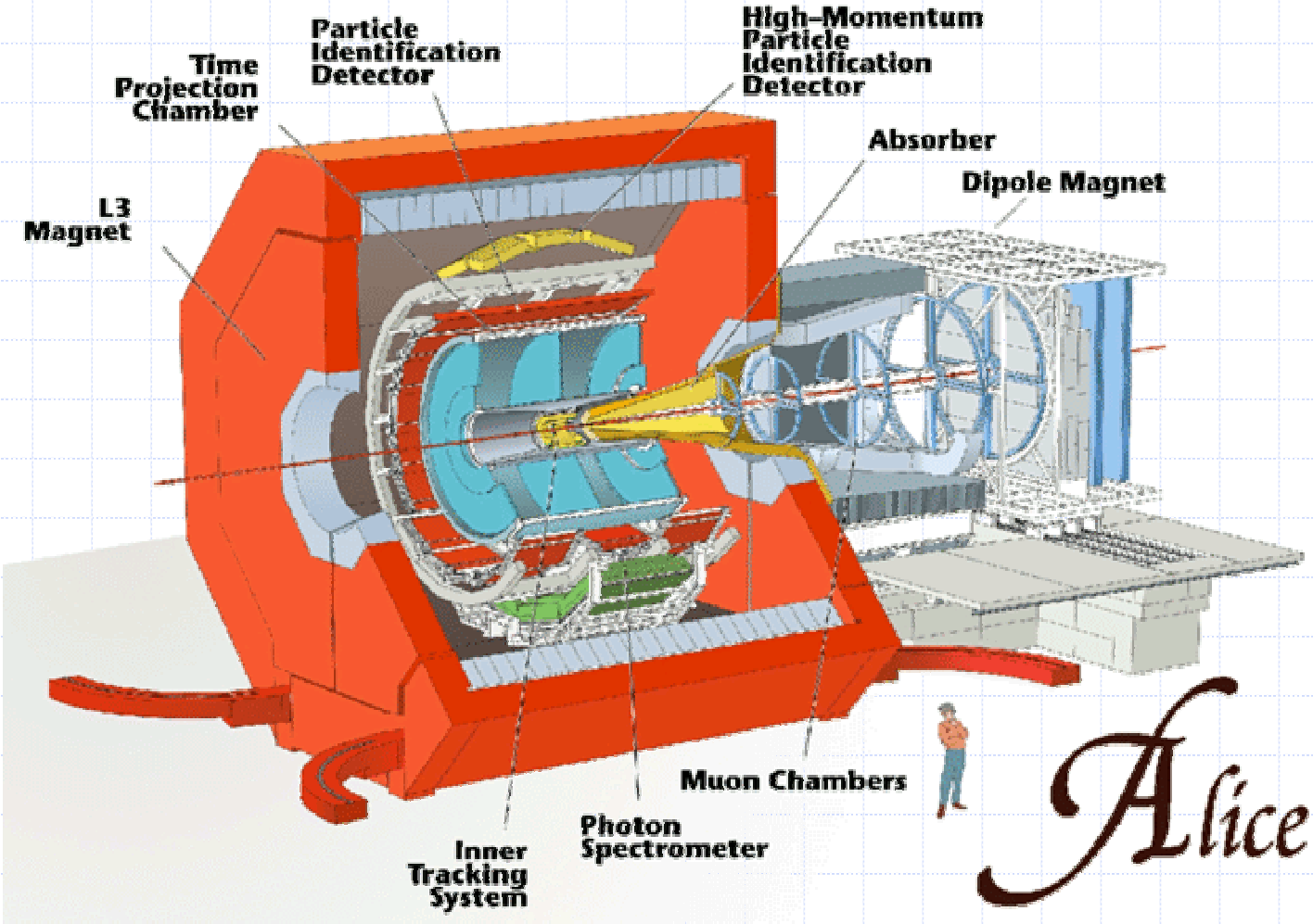
@GRID

CERN - LHC

Construction

@GRID

ALICE Experiment



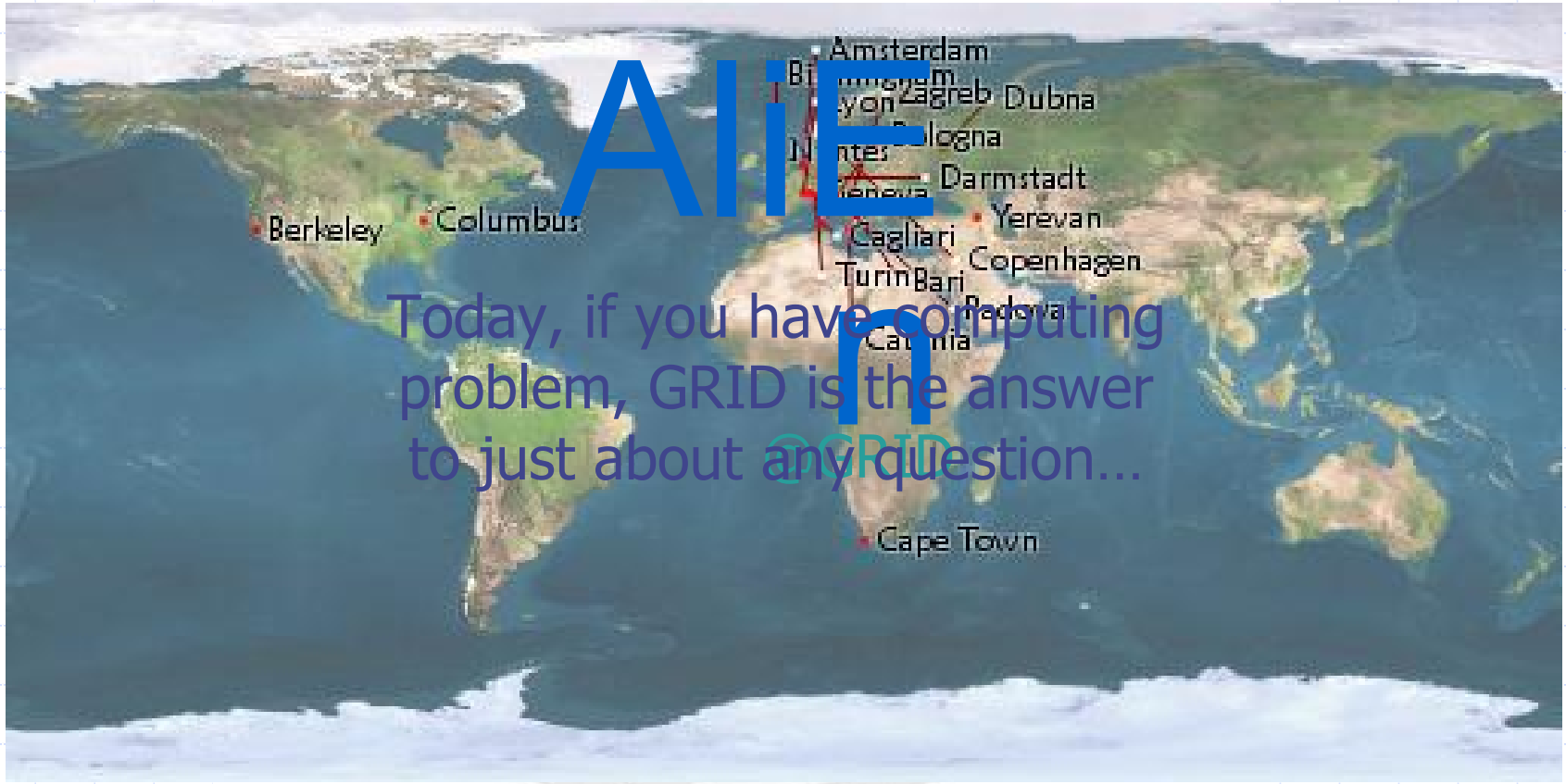
Alice Computing

- Simulation, Data Challenges & Reconstruction
  - Centrally managed production of background eventS
  - Distributed processing and event storage
- Event mixing
  - Not necessarily centrally managed
  - Once background events exist, the subsequent requests for event mixing must be routed to the location which holds required input
- Analysis
  - Using AliEn API, PROOF will locate optimal site(s) for macro execution, try to execute it in parallel, collect the output and return it to the user (or register it in the catalogue)

ALICE specific requirements

- Event size 2GB (simulated events), possibly split into several physical files, 20000 background events required
- Event size 40MB (Pb+Pb), 1MB (p+p), 10^9 files/year (x n, n>2)

Solution?



AliEn @GRID

Today, if you have computing problem, GRID is the answer to just about any question…

AliEn@GRID

- First implementation of Alice World Wide Computing Model
- A lightweight, simplified but functionally equivalent alternative to full blown GRID
- Partial solution which is applicable to our boundary conditions and current requirements for simulation, reconstruction
- Gaining recognition within the Collaboration and in wider community

@GRID

Architecture

- In brief:
  - File catalogue built on top of RDBMS with user interface that mimics the file system
  - Authentication module which supports various authentication methods
  - Task queue which holds commands to be executed in the system (commands, inputs and outputs are all registered in catalogue)
  - Metadata catalogue
  - Services that support above components
  - C/C++/perl API
  - 100% perl5 (95% reusable opens source modules)

Components

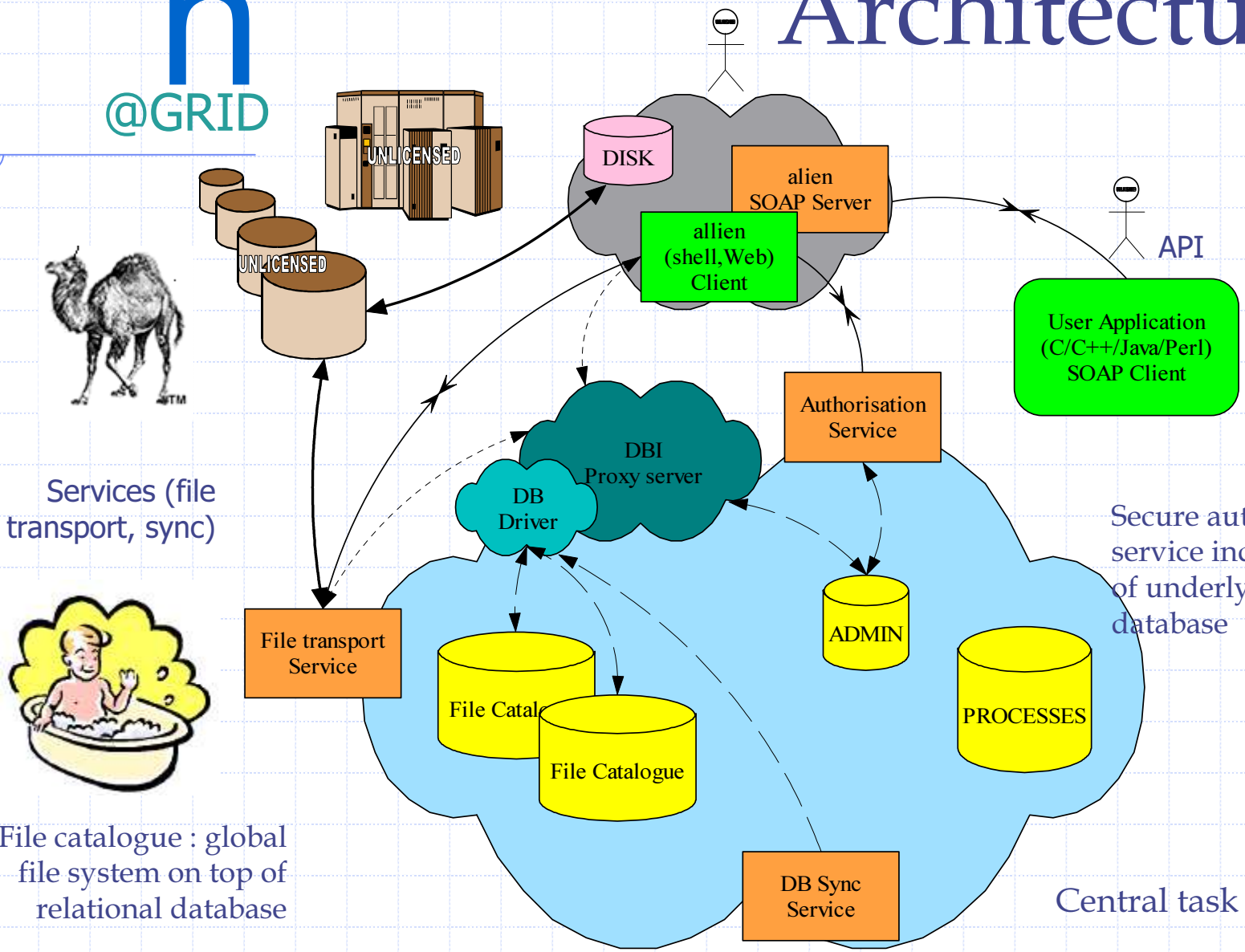
- AliEn Components (RPMs):
  - AliEn-Base
    - Contains external modules (more than 100, including Globus)
  - AliEn-Client
    - Basic client functionality, needed to access LFNs
  - AliEn-Server
    - AliEn Server, one per Virtual Organization
  - AliEn-SE
    - Storage Element, must be installed on sites which provide MSS functionality
  - AliEn-CE
    - Computing Element, must be installed if site wants to participate in production

@GRID

Components (…)

- Optional components:
  - AliEn-GUI
    - Graphical User interface, optional component
  - AliEn-Monitor
    - Monitor is required by Server to enable advanced RB features
  - AliEn-Portal
    - This is AliEn Web site
  - AliEn-Alice
    - Description of Alice Virtual Organization
      - Packages
      - Commands
- The RPMs can be found at http://alicedb.cern.ch/GRID/current

Architecture



Services (file transport, sync)

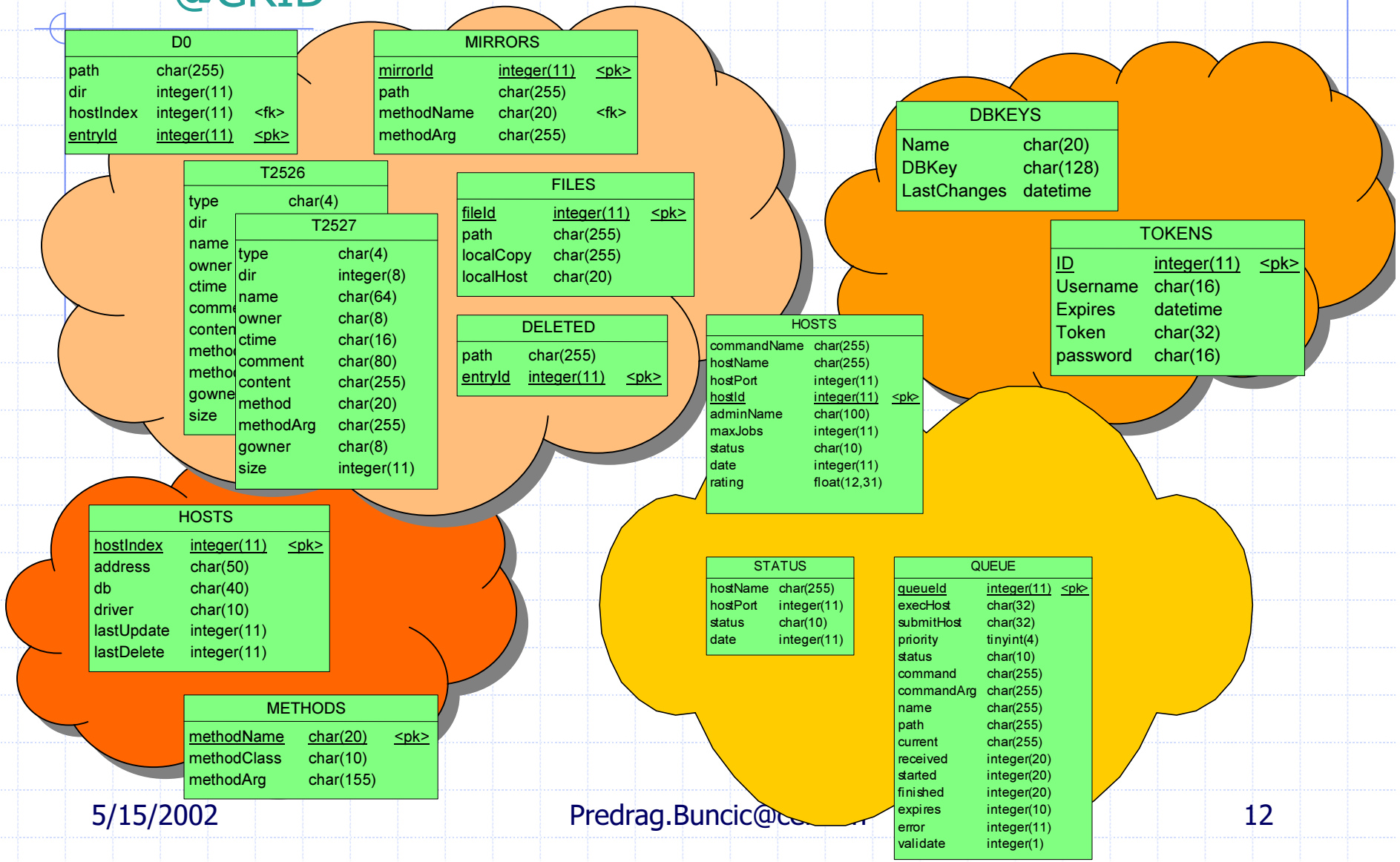
Secure authentication service independent of underlying database

File catalogue : global file system on top of relational database

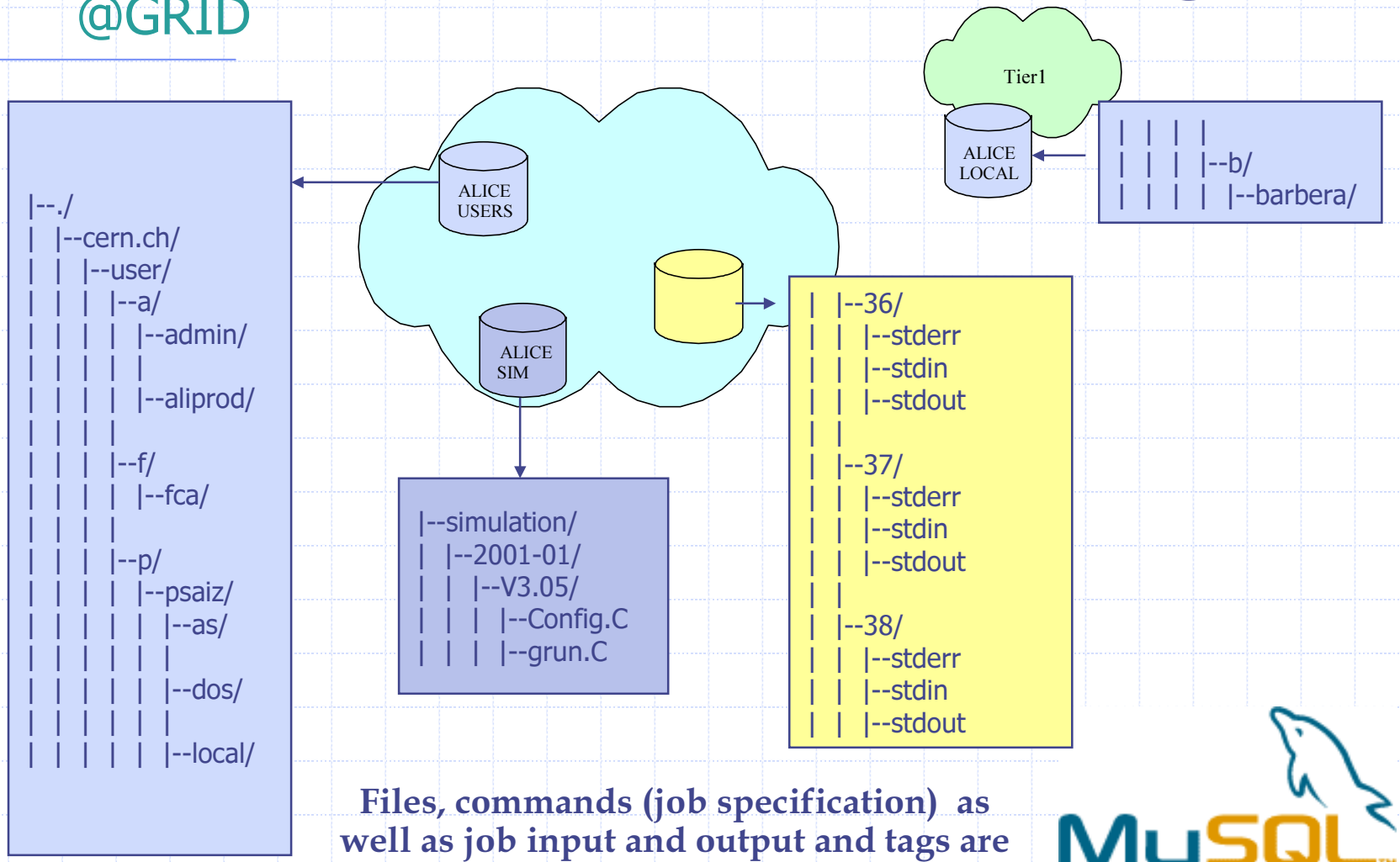
Central task queue

SQL Backend

@GRID



File catalogue



Files, commands (job specification) as well as job input and output and tags are stored in the catalogue

File organization

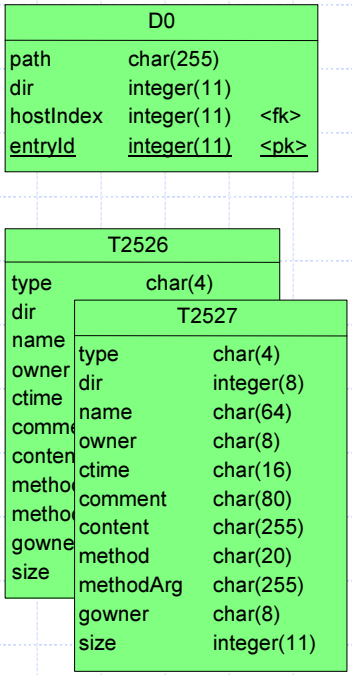
```
[tbed0007d.cern.ch] /alice/simulation/2001-02/V3.06/00001/ > tree
|--./
|  |--00001/
|  |  |--galice.root
|  |
|  |--00002/
|  |  |--galice.root
|  |  .....
|  |--Config.C
|  |--grun.C
```

```
[tbed0007d.cern.ch] /proc/33608/ > tree
|--./
|  |--stderr
|  |--stdin
|  |--stdout
```

Forgotten wisdom: by organizing files into directory structure one can already tell a lot about file content, define cleanup and access policy and optimize access performance

Tags

```
--./
|  |--r3418_01-01.ds
|  |--r3418_02-02.ds
|  |--r3418_03-03.ds
|  |--r3418_04-04.ds
|  |--r3418_05-05.ds
|  |--r3418_06-06.ds
|  |--r3418_07-07.ds
|  |--r3418_08-08.ds
|  |--r3418_09-09.ds
|  |--r3418_10-10.ds
|  |--r3418_11-11.ds
|  |--r3418_12-12.ds
|  |--r3418_13-13.ds
|  |--r3418_14-14.ds
|  |--r3418_15-15.ds
```

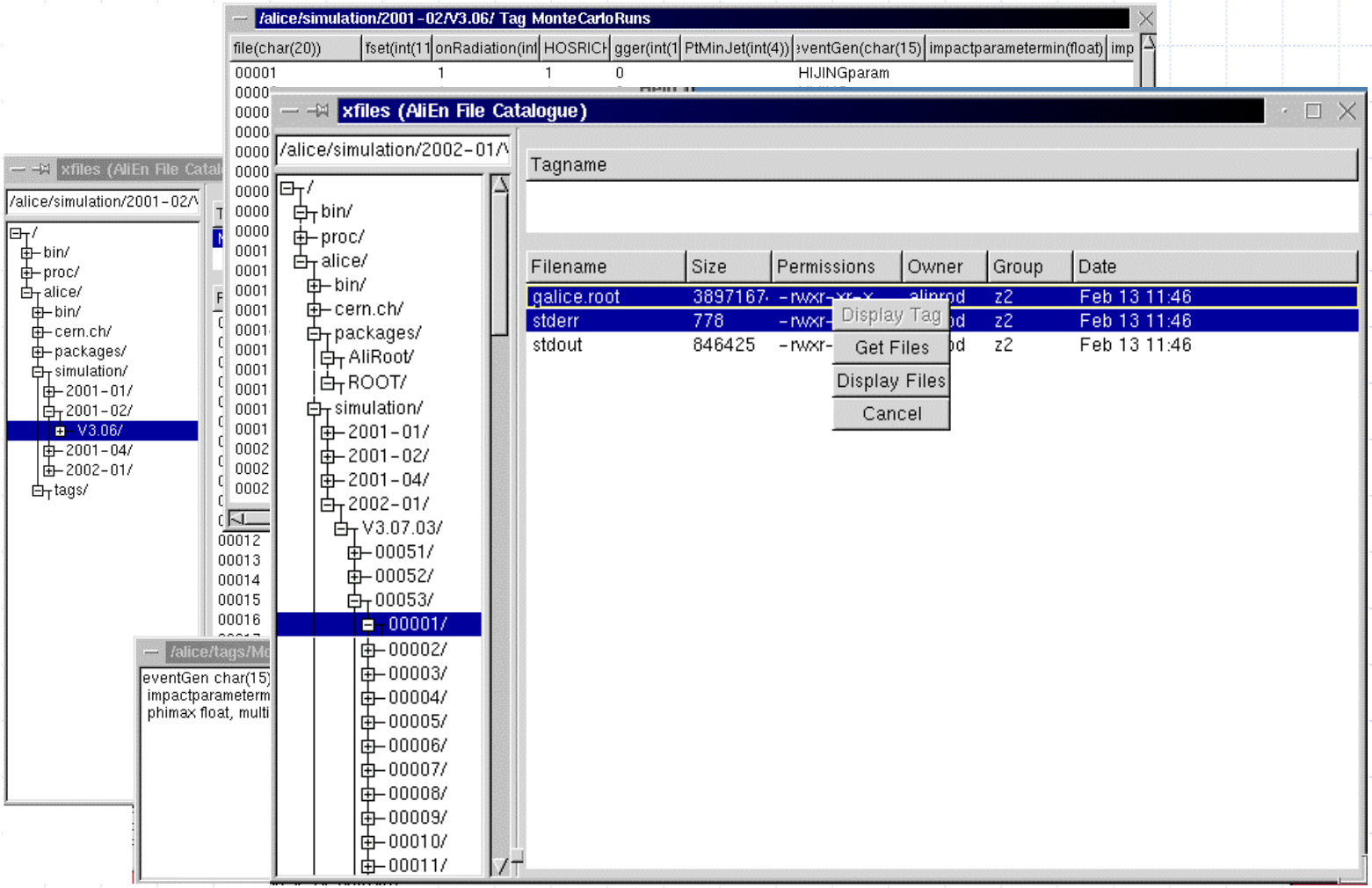


- The file catalogue on its own does not know anything about file content
- It is possible to add an additional information to describe file properties (metadata)
- In AliEn environment this can be achieved by attaching an arbitrary number of TAG table(s) to the corresponding directory table

lfn://alien.cern.ch/alice/simulation/2001%/V3.05/%/galice.root?npart>1000#mytag

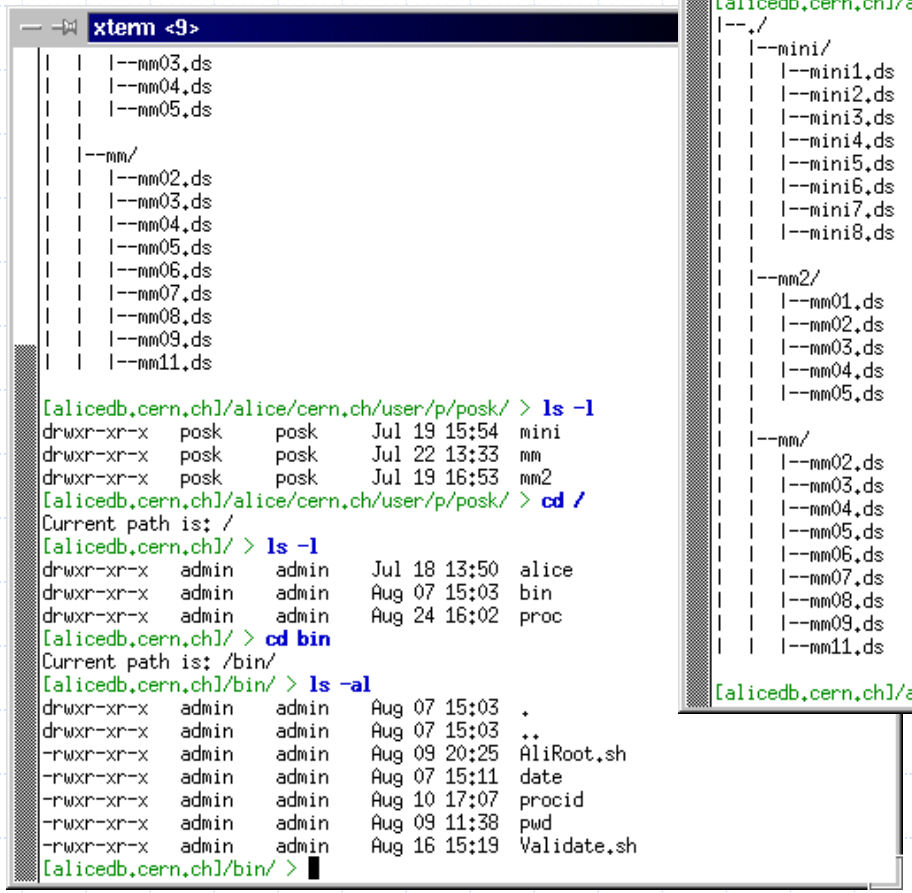
The search will first select all tables on the basis of the file name selection and then locates all tables that correspond to "mytag" definition, apply selection and finally return only the list of files for which the attribute search has been successful

GUI: AliEn Xfiles



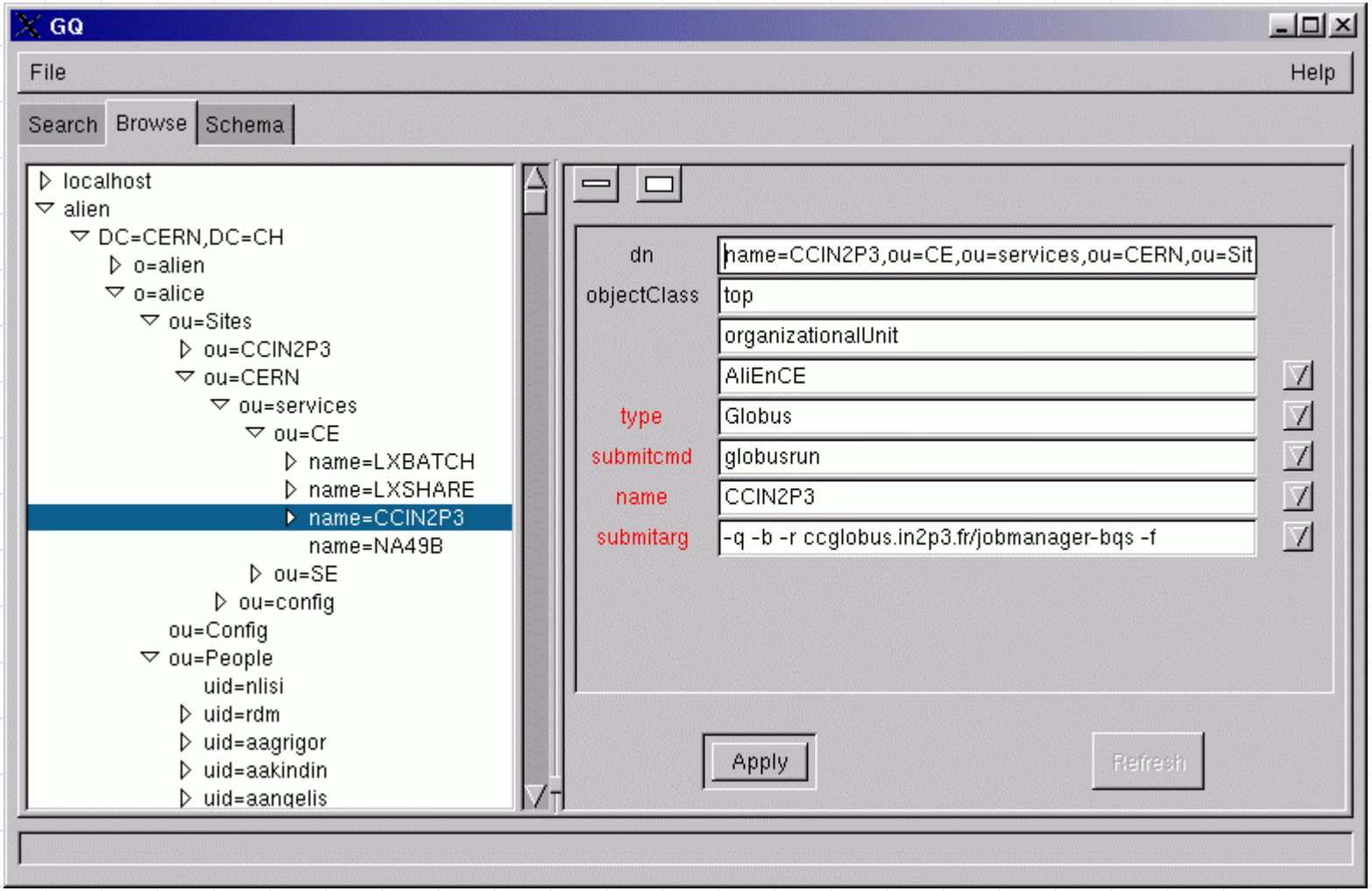
Command Interface

@GRID



@GRID

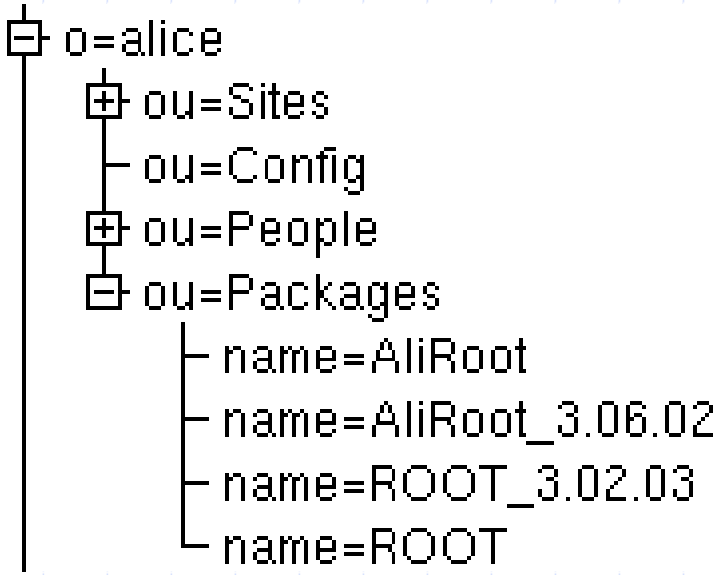
OpenLDAP®
http://www.OpenLDAP.org



Organisation Cfg.

An Organisation defines:

- Sites
- People
- Service evironment
- Packages



Site Cfg.

Site defines:

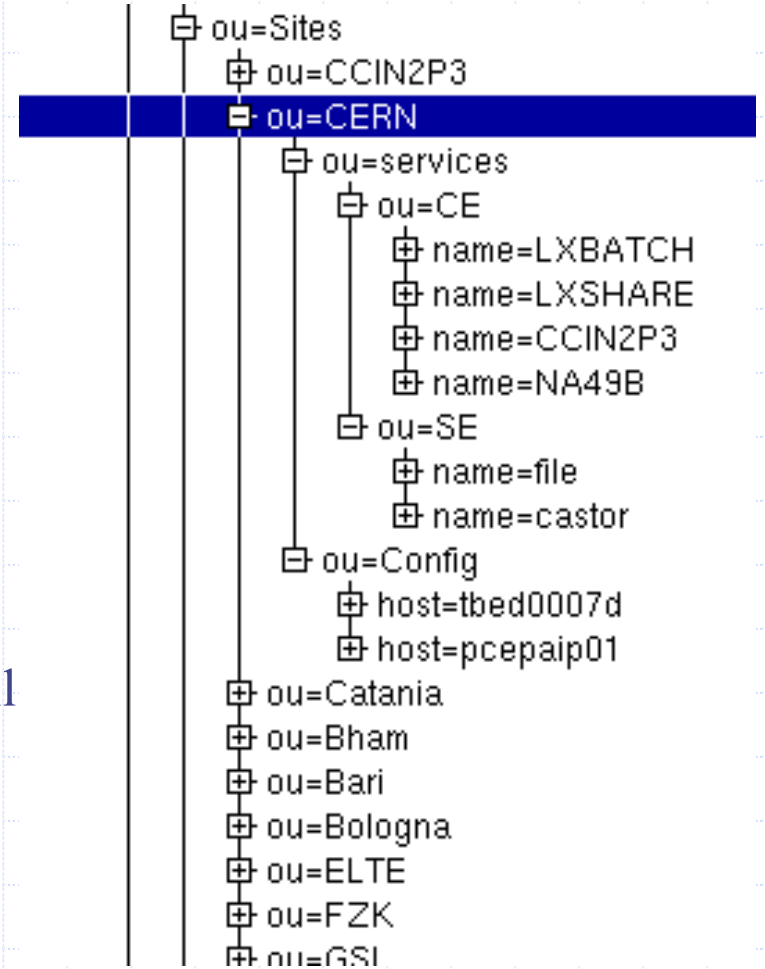
- Log directories
- Packages to install
- CE
- SE
- Host configuration

CE defines:

- Type of a local queue
- Command and arguments to submit, check the status and kill jobs (if different than default)

SE defines:

- Type of MSS
- Base directory



Authentication - SASL

- SASL is the Simple Authentication and Security Layer, a method for adding authentication support to connection-based protocols.
- To use SASL, a protocol includes a command for identifying and authenticating a user to a server and for optionally negotiating protection of subsequent protocol interactions.
- If its use is negotiated, a security layer is inserted between the protocol and the connection.
- It can be used on the client or server side to provide authentication (See RFC 2222 for more information)
- OpenLDAP v2.x uses SASL

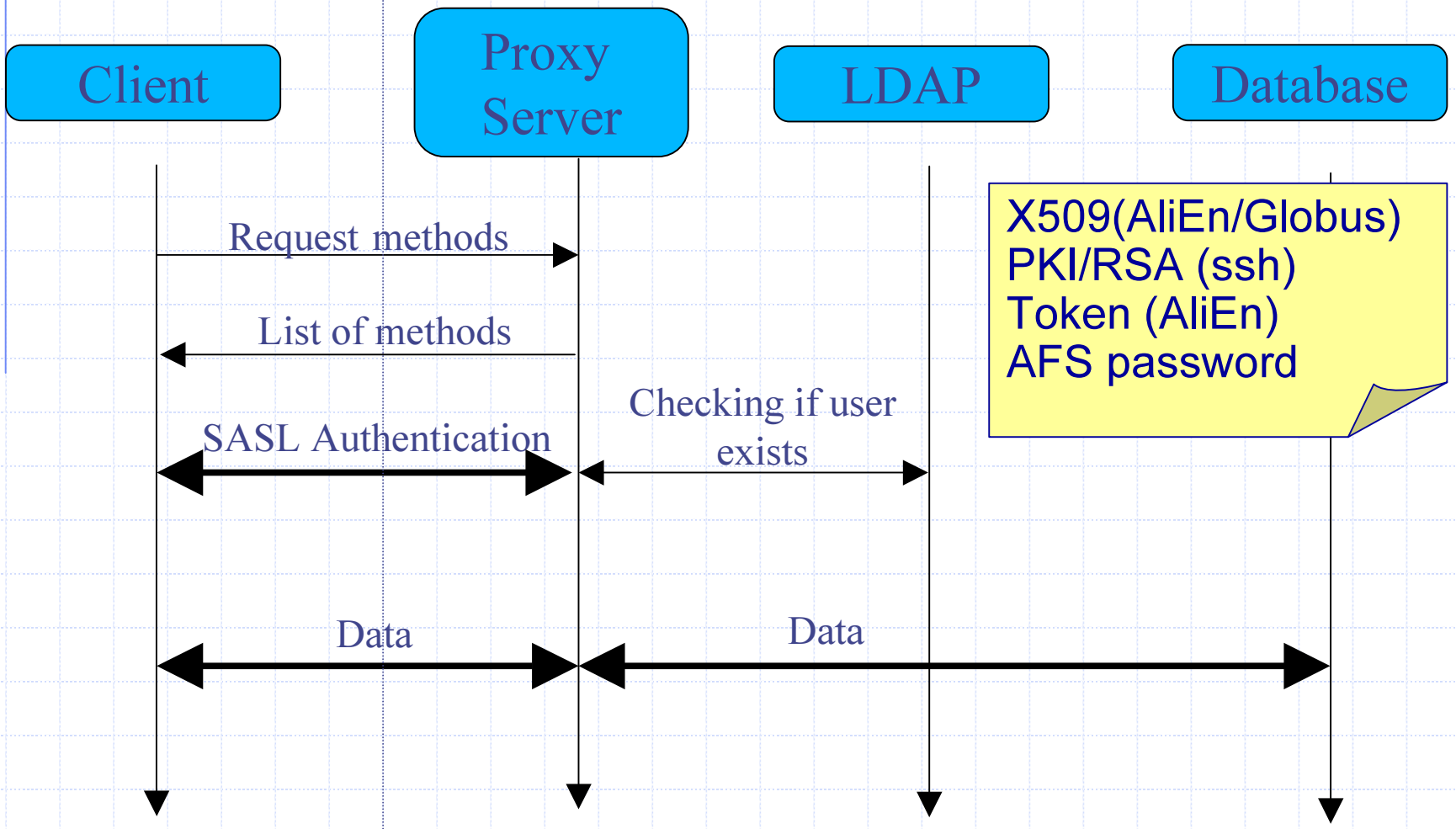
SASL mechanisms

- The following mechanisms are included in Cyrus SASL distribution:
  - ANONYMOUS
  - CRAM-MD5
  - KERBEROS_V4
  - PLAIN
  - SCRAM-MD5 (deprecated)
  - GSSAPI (MIT Kerberos 5 or Heimdal Kerberos 5)
  - DIGEST-MD5
  - LOGIN (unsupported)
  - SRP (unsupported, may not work)
- Globus MDS 2.1 uses Globus/GSI implementation of GSSAPI

AliEn SASL implementation

- AliEn now has perl module with implementation GSSAPI
- This allows us to use
  - all SASL authentication schemes
  - old AliEn authentication (token, AFS password, SSH)
  - X509 certificates
  - Globus/GSI (credential delegation)
- AliEn distribution now includes necessary Globus/MDS/GSI software
- This allows us to develop secure Peer-To-Peer File Transfers based on machine/protocol/user certificates and LDAP based configuration management

Authentication



Token or PKI Key

To obtain either a private key or a token run

alien CreateKeys
alien UpdateToken

This will prompt you for a password. The password will be send to the AliEn authentication server using SOAP over SSL.

The Authentication server will check that you exists in our LDAP-server and that your password is correct. Checked with PAM.

For VO Alice the password is the CERN AFS password

Certificates

Currently AliEn will trust certificates signed by Cern CA or AliEn CA. This means that normal certificates issued by CERN will be valid as credential.

A certificate is is requested by typing (on DataGrid machine):

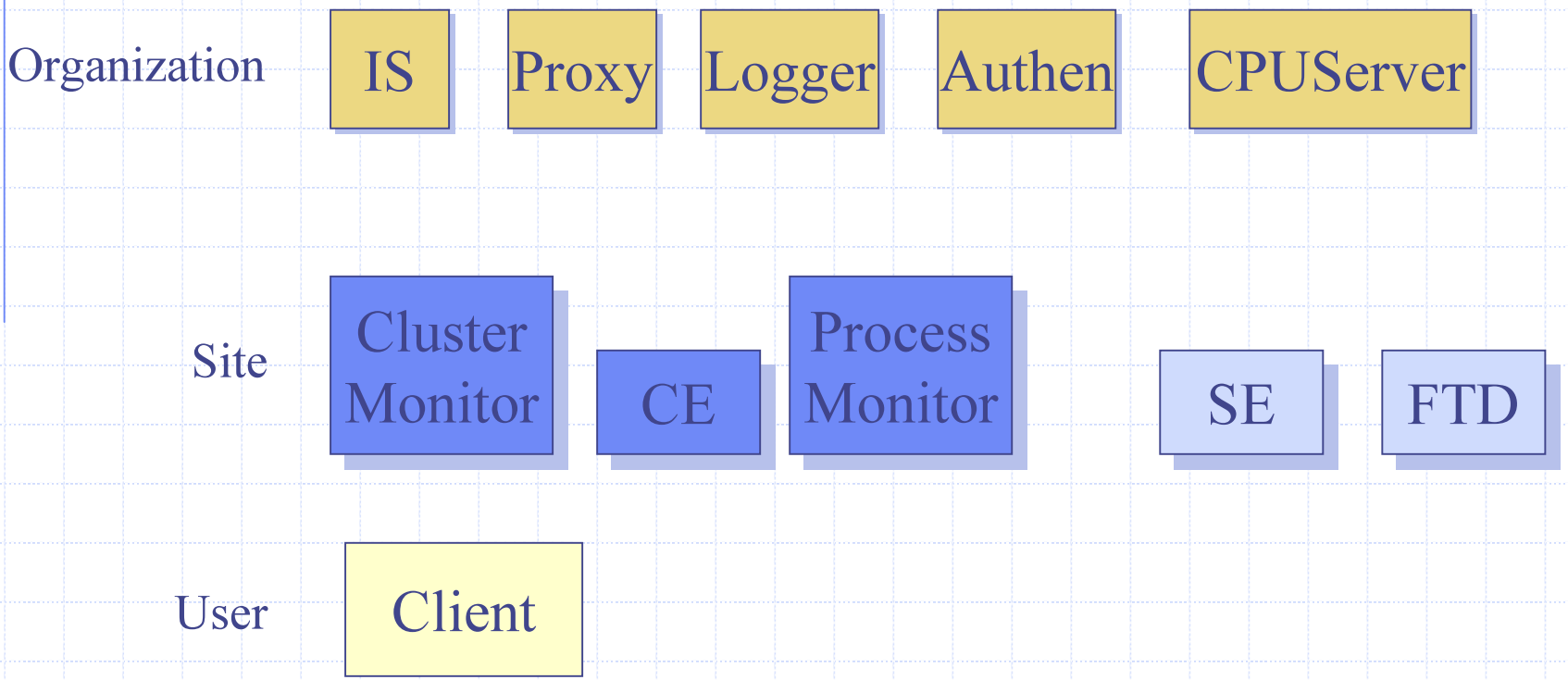
grid-cert-request

To register your certificate for use with AliEn, do:

alien RegisterCert

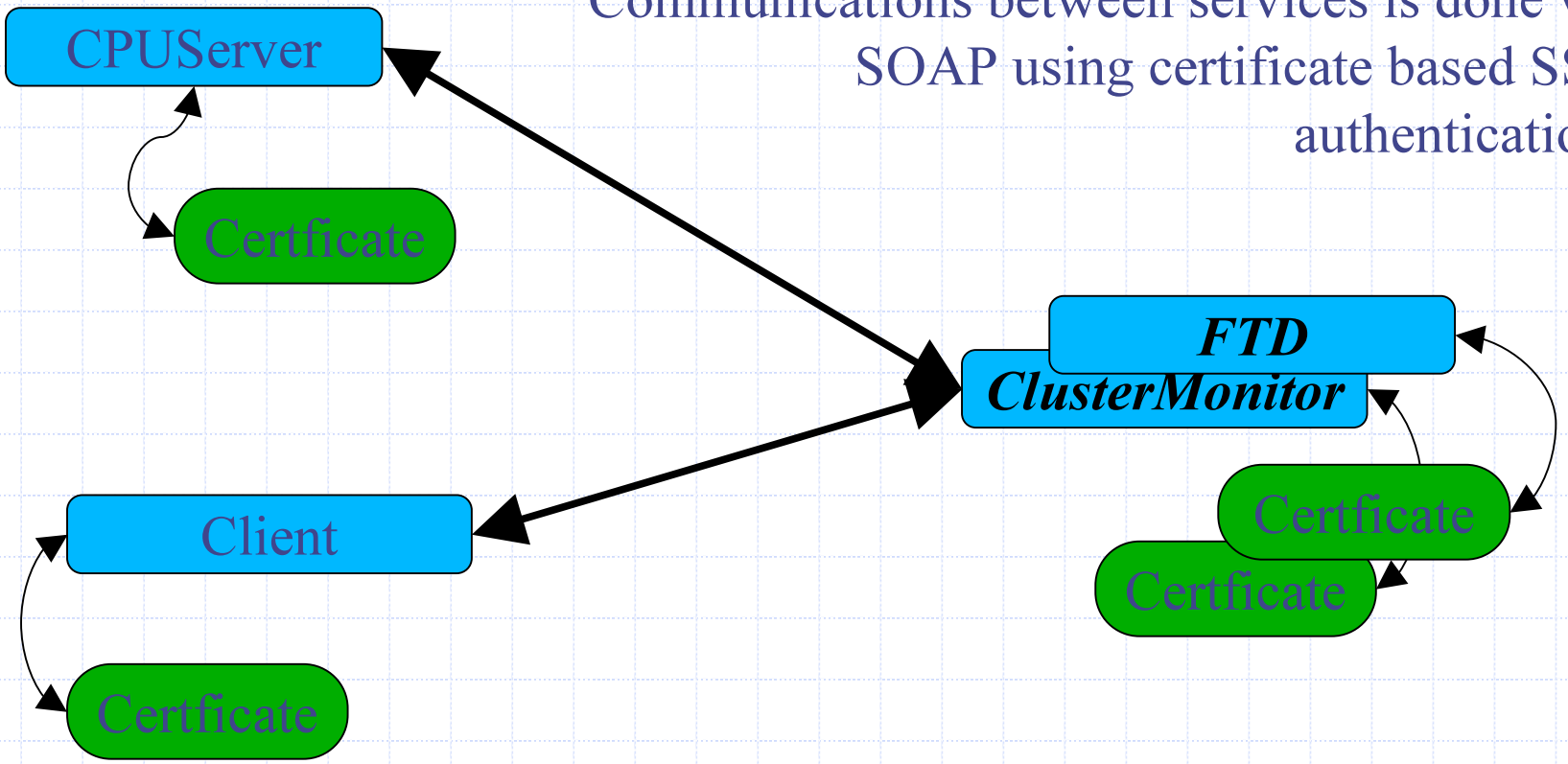
@GRID

Services

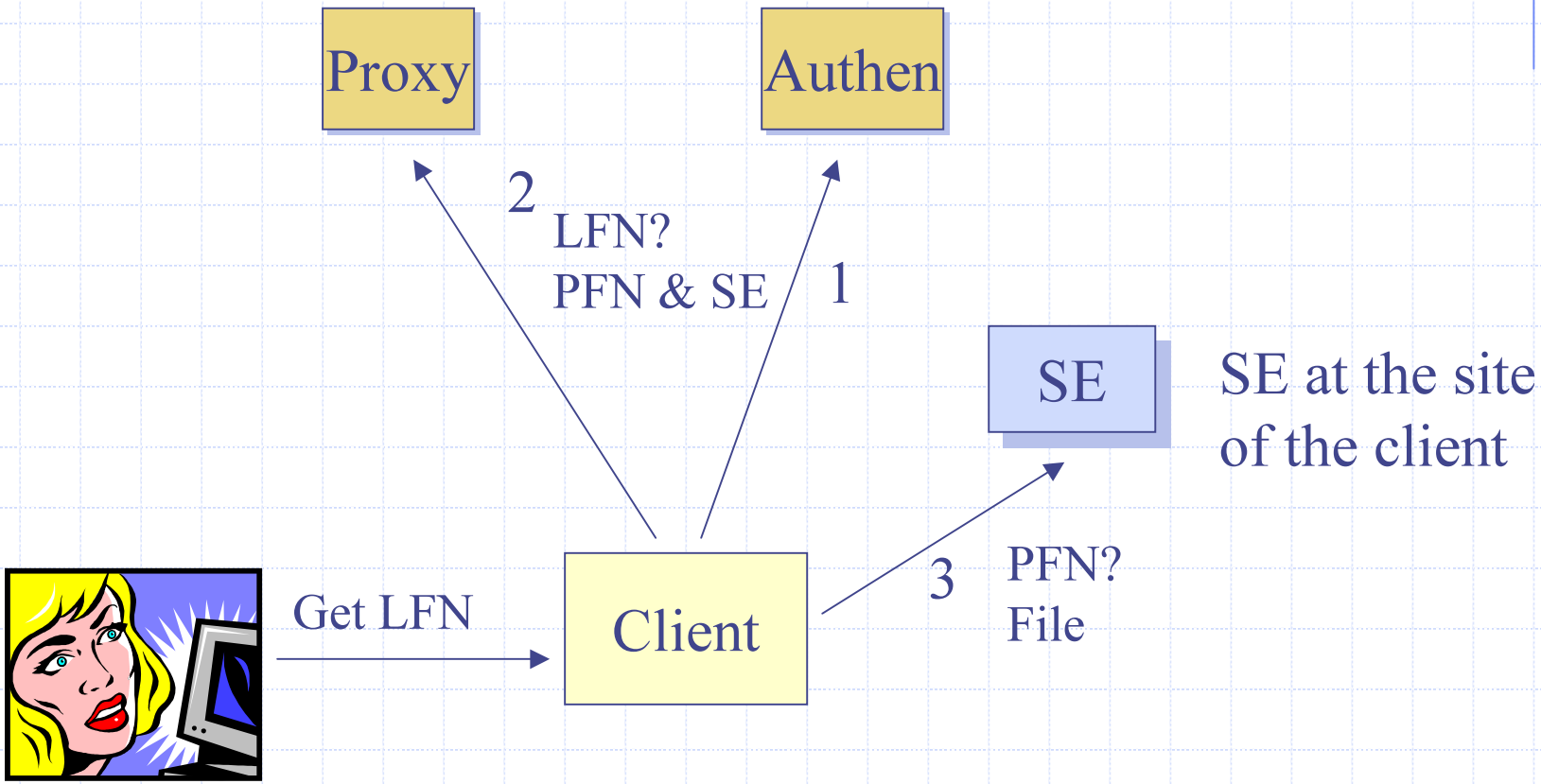


Services

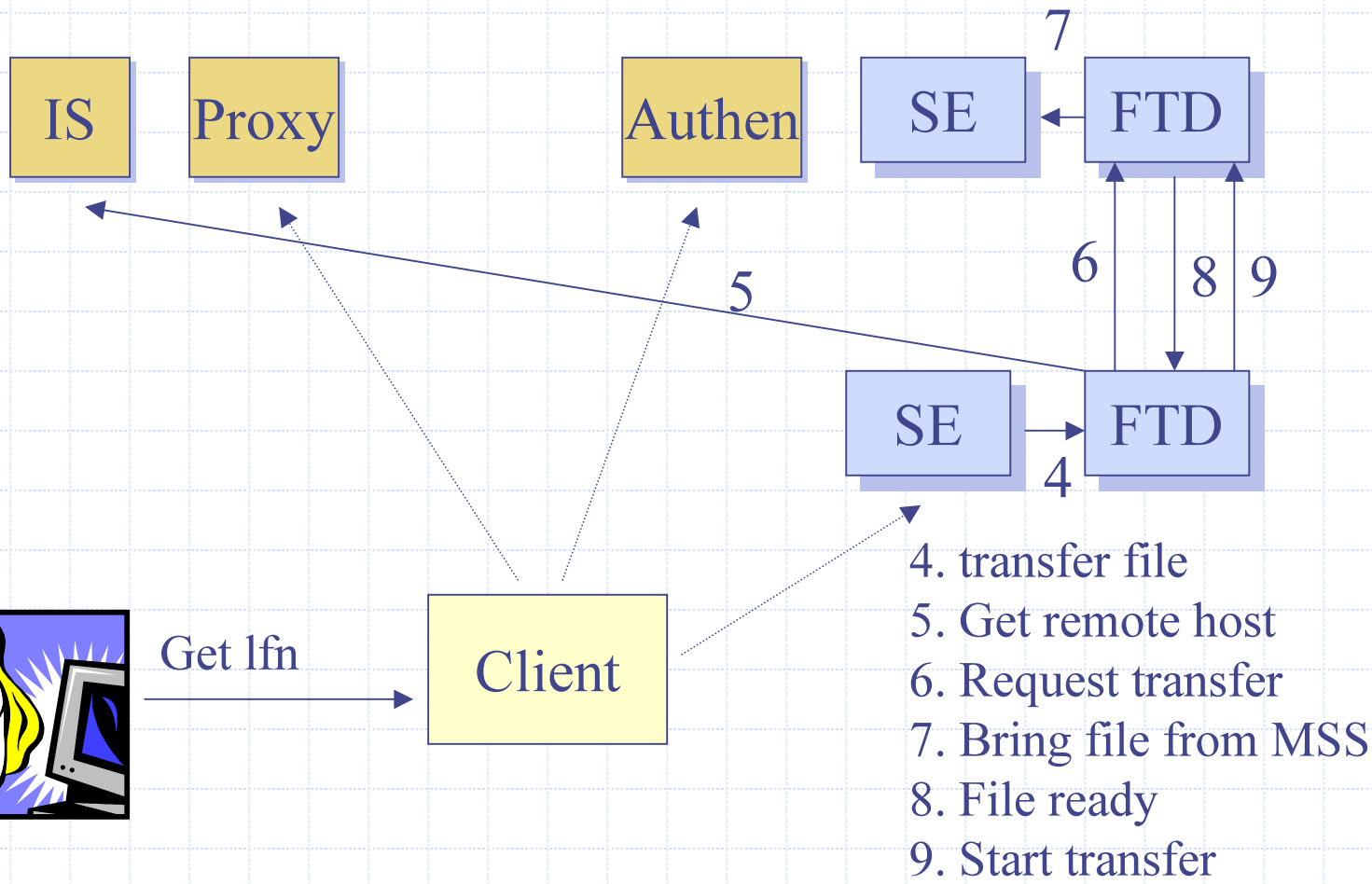
Communications between services is done via SOAP using certificate based SSL authentication.



Getting a file (from local SE)



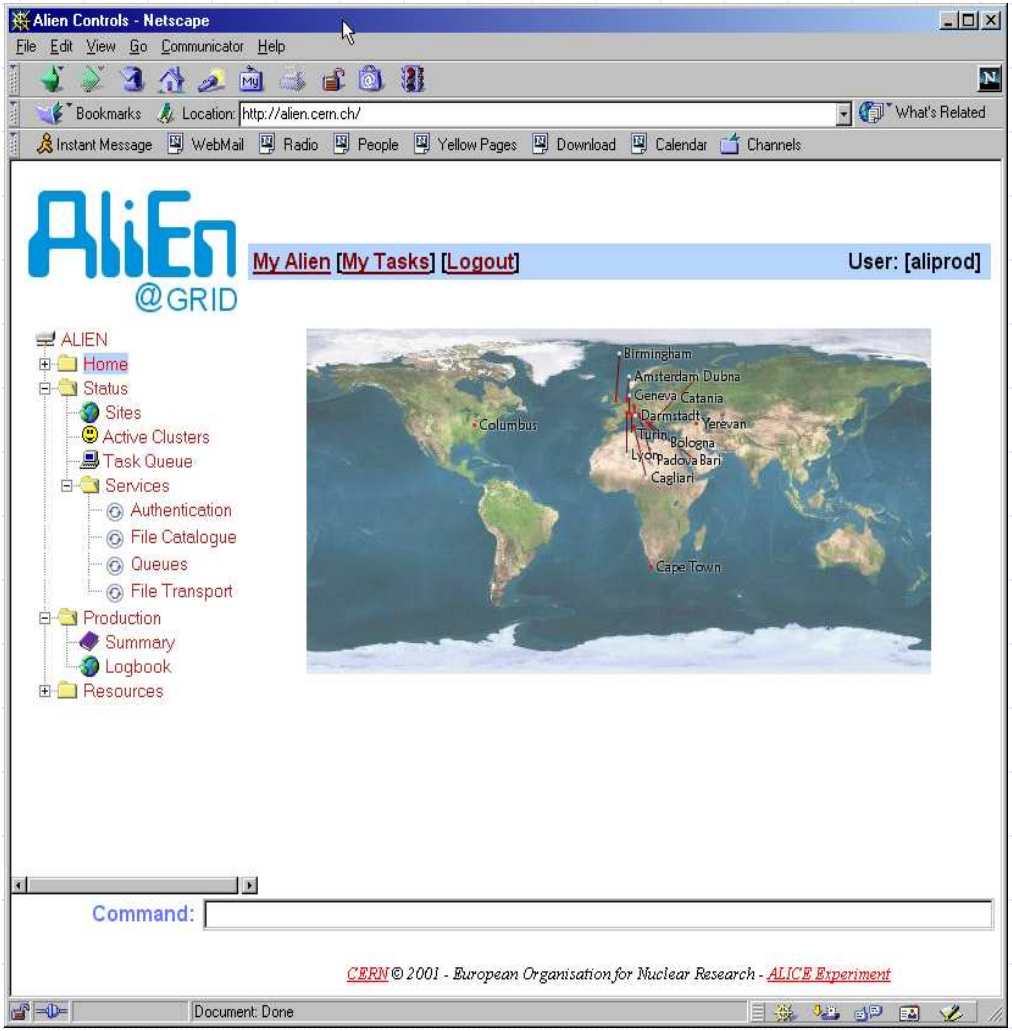
Getting a file (from remote SE)



Portal

http://alien.cern.ch

- Generic Web portal
- User can
  - interact with alien
  - submit jobs
  - check jobs status
- Administrator can
  - configure system
  - monitor status
  - check syslog
  - update distribution



Sending a job…

Production Edit+

ROUND	TAG	COMMENT	STATUS	PATH
2001-01	V3.05	PPR simulation	TESTING	/alice/simulation/2001-01/V3.05

Runs Edit+

Action	Run	Todo	Done	Failed	STARTED	Config	Macro
Resume	2	10	0	0	Sat Aug 25 10:39	Config.C	grun.C

None
Submit
Pause
Resume
Stop

Submit Reset

Task Queue

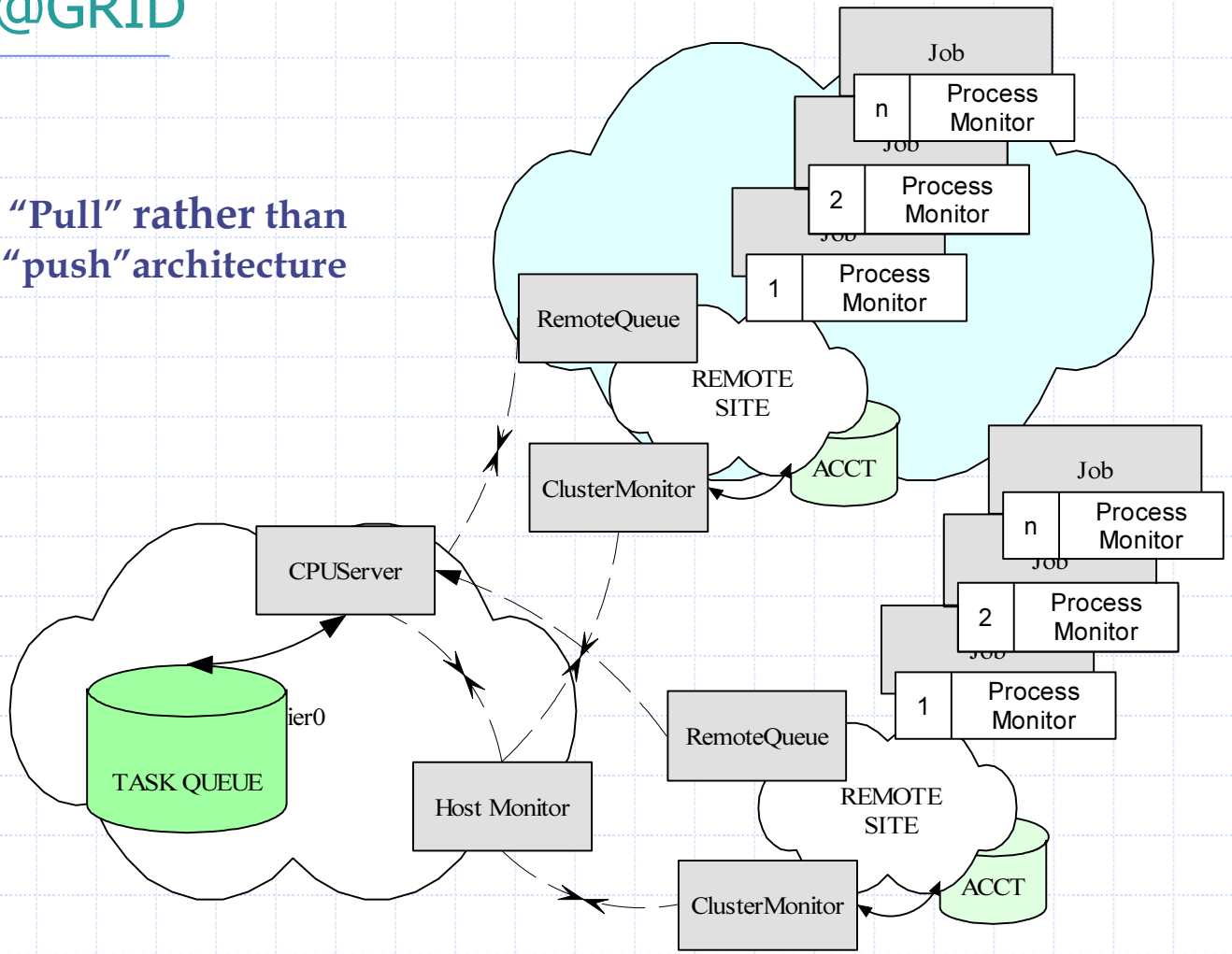
[ALL][EXPIRED][WAITING][DONE][KILLED][ERROR]

Task Queue

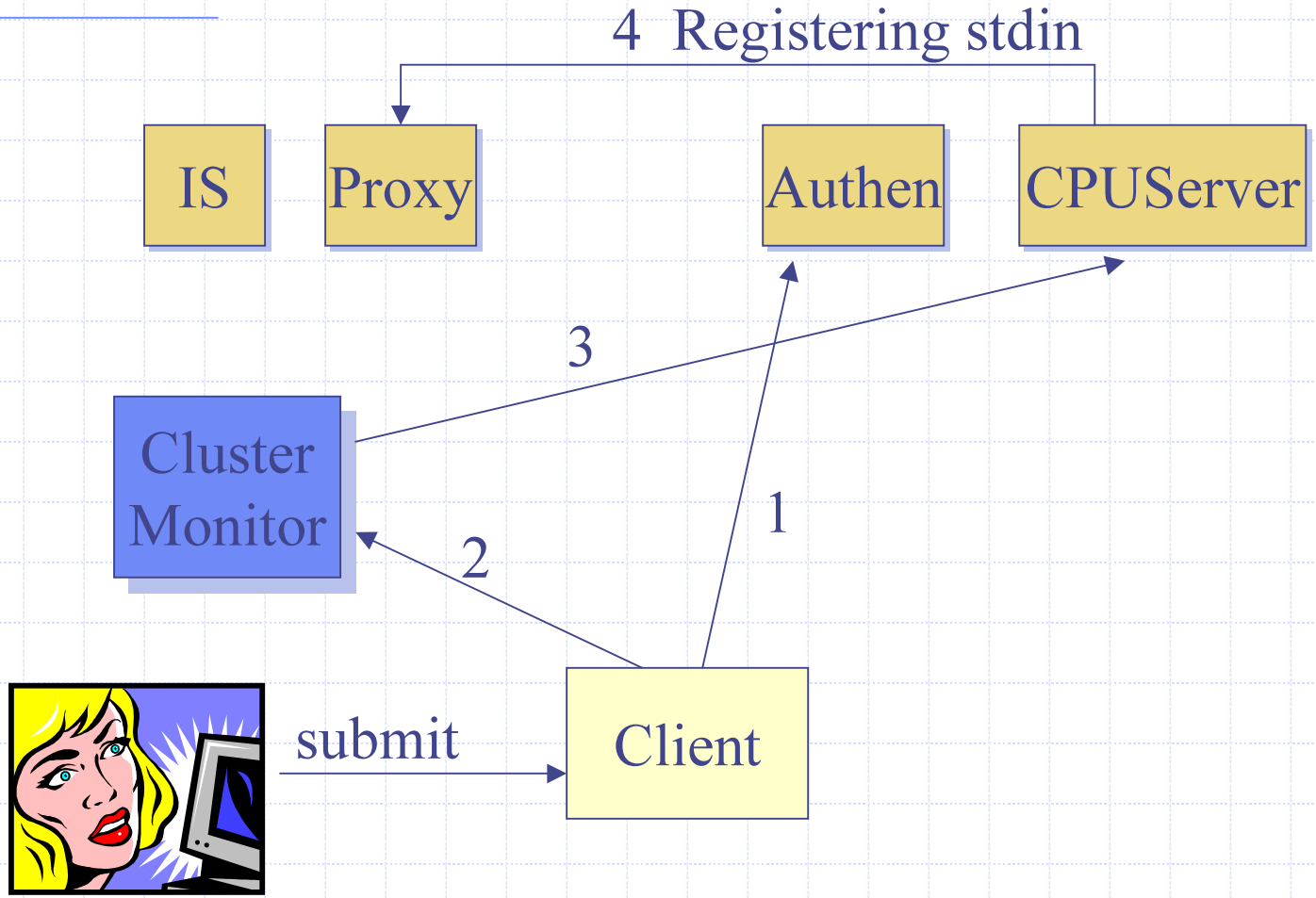
ID	USER	EXEC	STATUS	COMMAND	START	TIME
30	psaiz	pcepaip15.cern.ch	ERROR	AliRoot.sh	Fri Sep 14 10:21	447:38:42
31	psaiz		KILLED	AliRoot.sh		00:00:00
32	psaiz		KILLED	AliRoot.sh		00:00:00
33	psaiz	pcepaip15.cern.ch	ERROR	AliRoot.sh	Fri Sep 14 09:57	448:02:59
34	psaiz	pcepaip15.cern.ch	ERROR	AliRoot.sh	Fri Sep 14 10:00	447:59:12
35	psaiz	pcepaip15.cern.ch	KILLED	AliRoot.sh	Fri Sep 14 10:05	447:54:06
36	psaiz	pcepaip15.cern.ch	DONE	AliRoot.sh	Fri Sep 14 10:23	00:00:19
37	psaiz	pcepaip15.cern.ch	DONE	AliRoot.sh	Fri Sep 14 10:27	00:00:49
38	psaiz	pcepaip15.cern.ch	DONE	AliRoot.sh	Fri Sep 14 11:22	00:00:15
39	psaiz	pcepaip15.cern.ch	KILLED	AliRoot.sh	Fri Sep 14 10:18	447:41:16
40	psaiz	pcepaip15.cern.ch	DONE	AliRoot.sh	Fri Sep 14 11:25	00:00:18

Task Queue

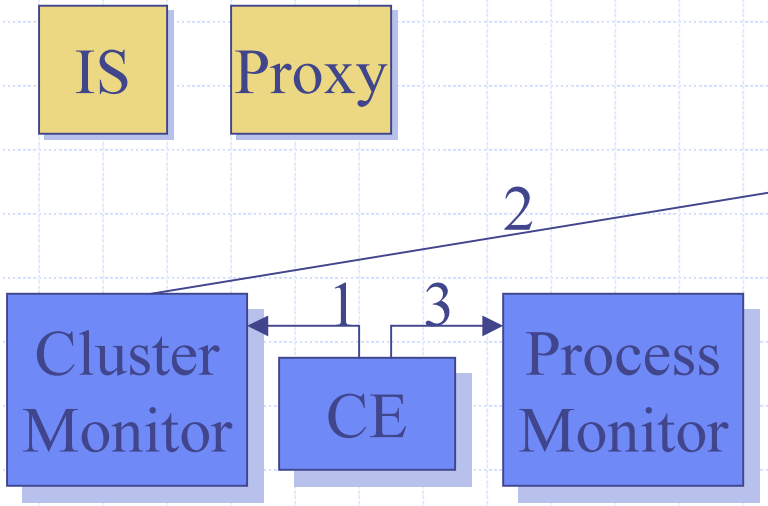
"Pull" rather than "push"architecture



Submiting a job



Executing a job



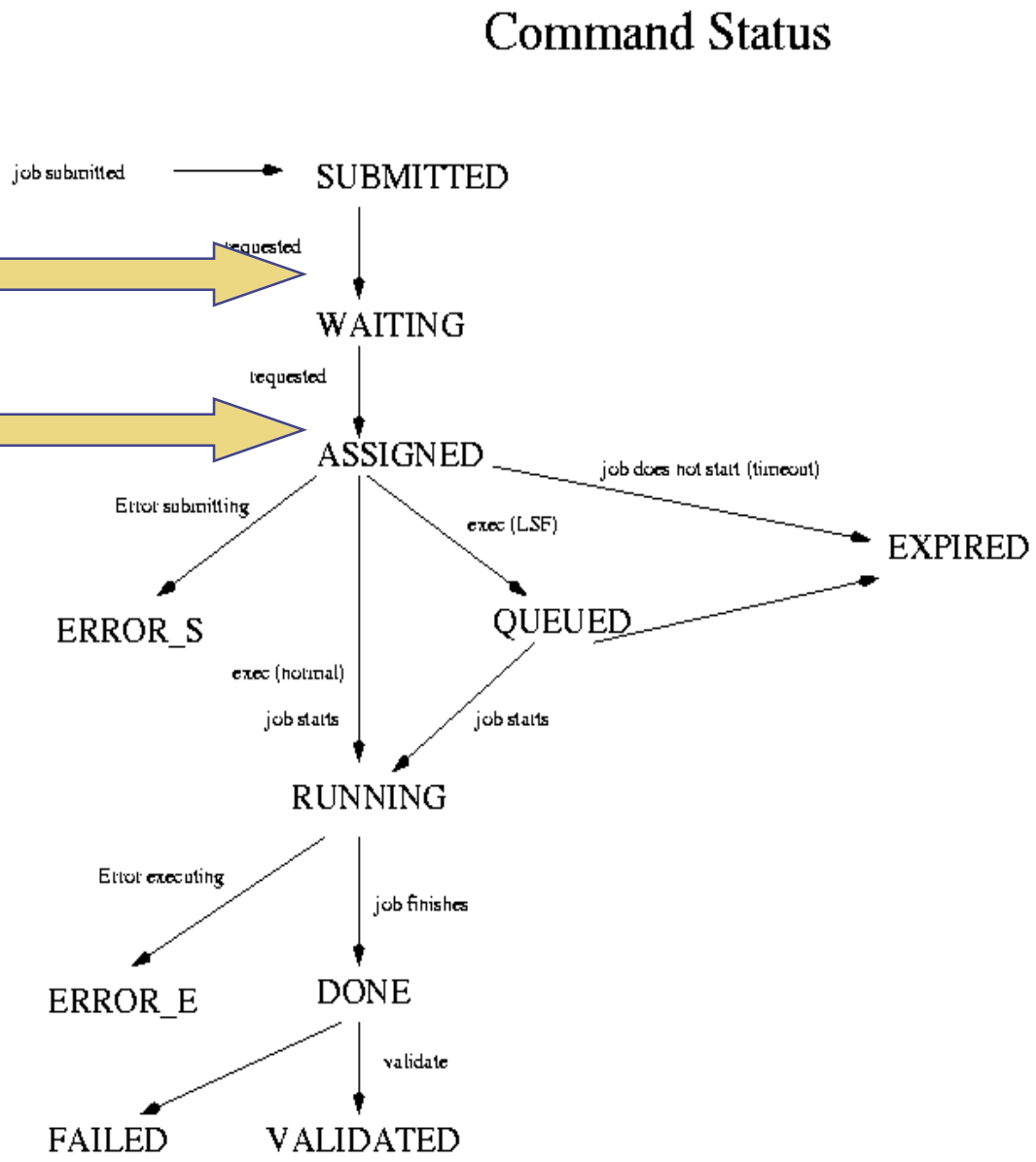
Possible Local Queues:

- LSF
- PBS
- BQS
- Globus
- CONDOR
- DQS

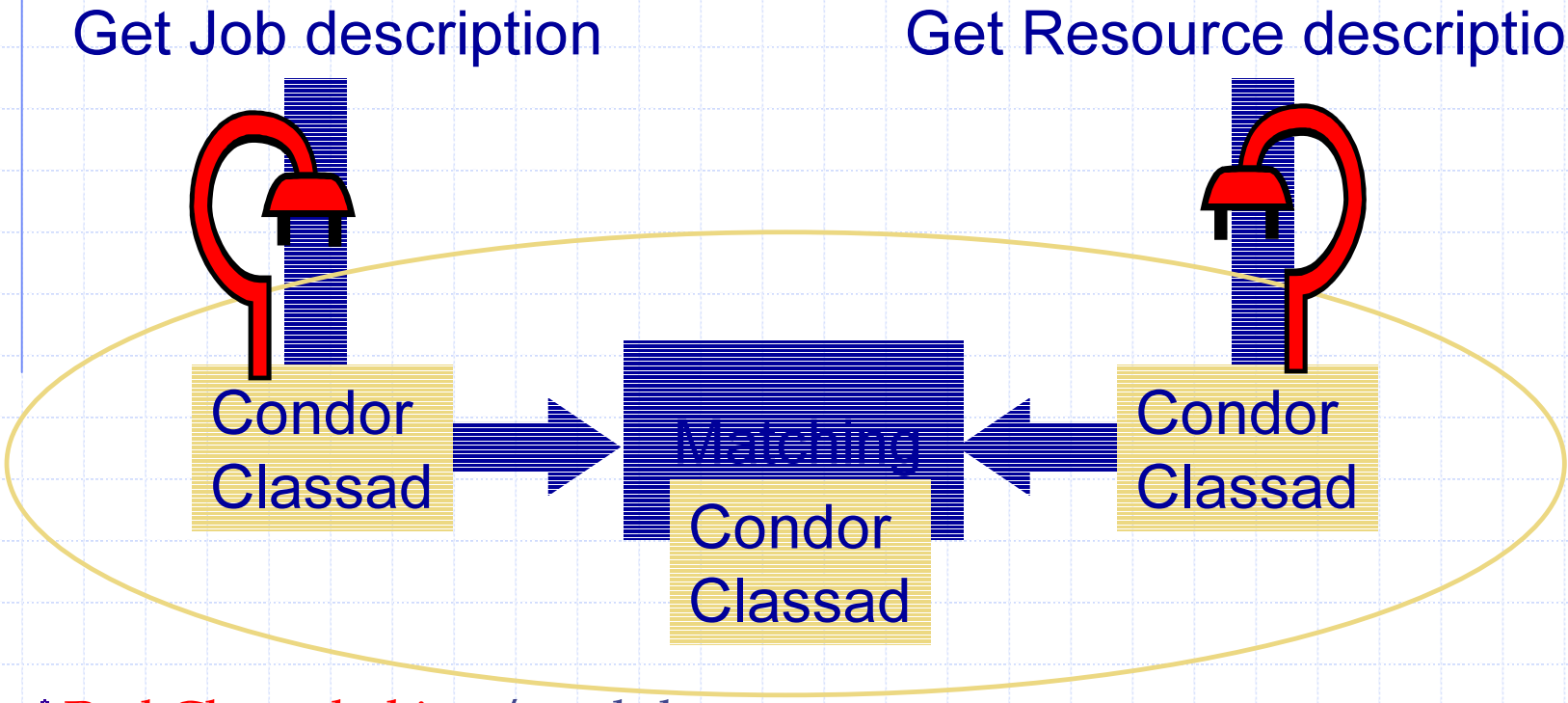
@GRID

Resource Broker

Optimizer

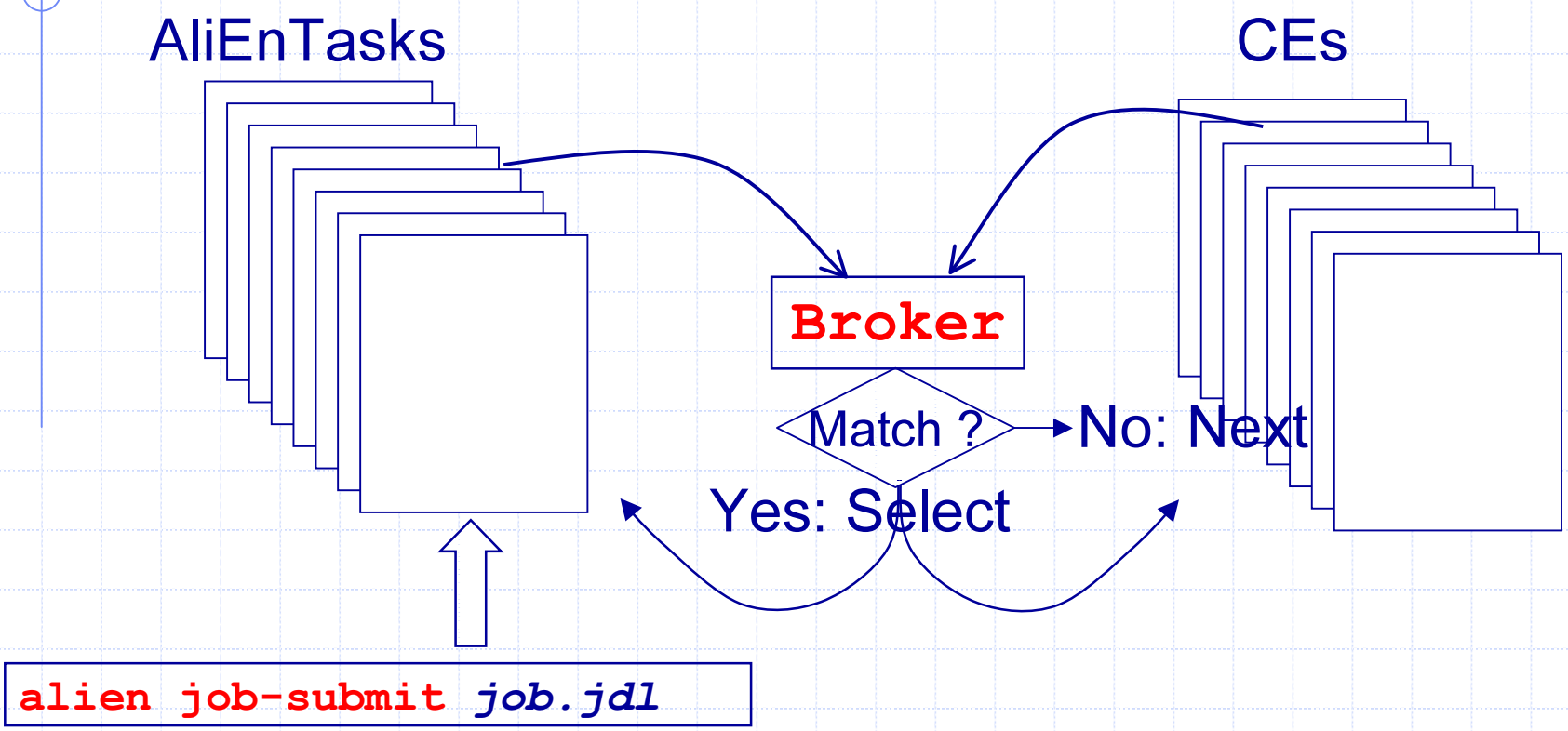


Resource broking



- Perl Classad object/module
  - wrapped C++ Condor Classad library using an « automated » tool : SWIG (Simplified Wrapper and Interface Generator, www.swig.org)

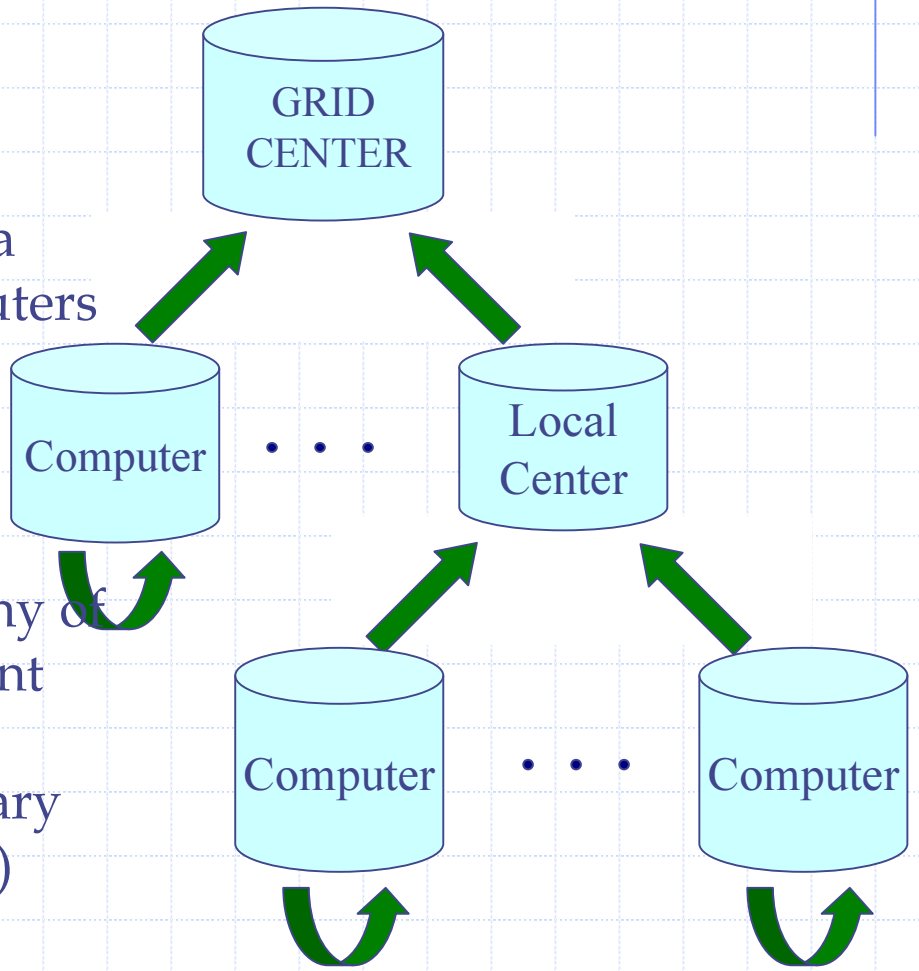
Resource Broker



CE contacts CPUServer and presents its own ClassAd,
Resource Broker will match them against job ClassAds and select
the most appropriate job to run on that CE

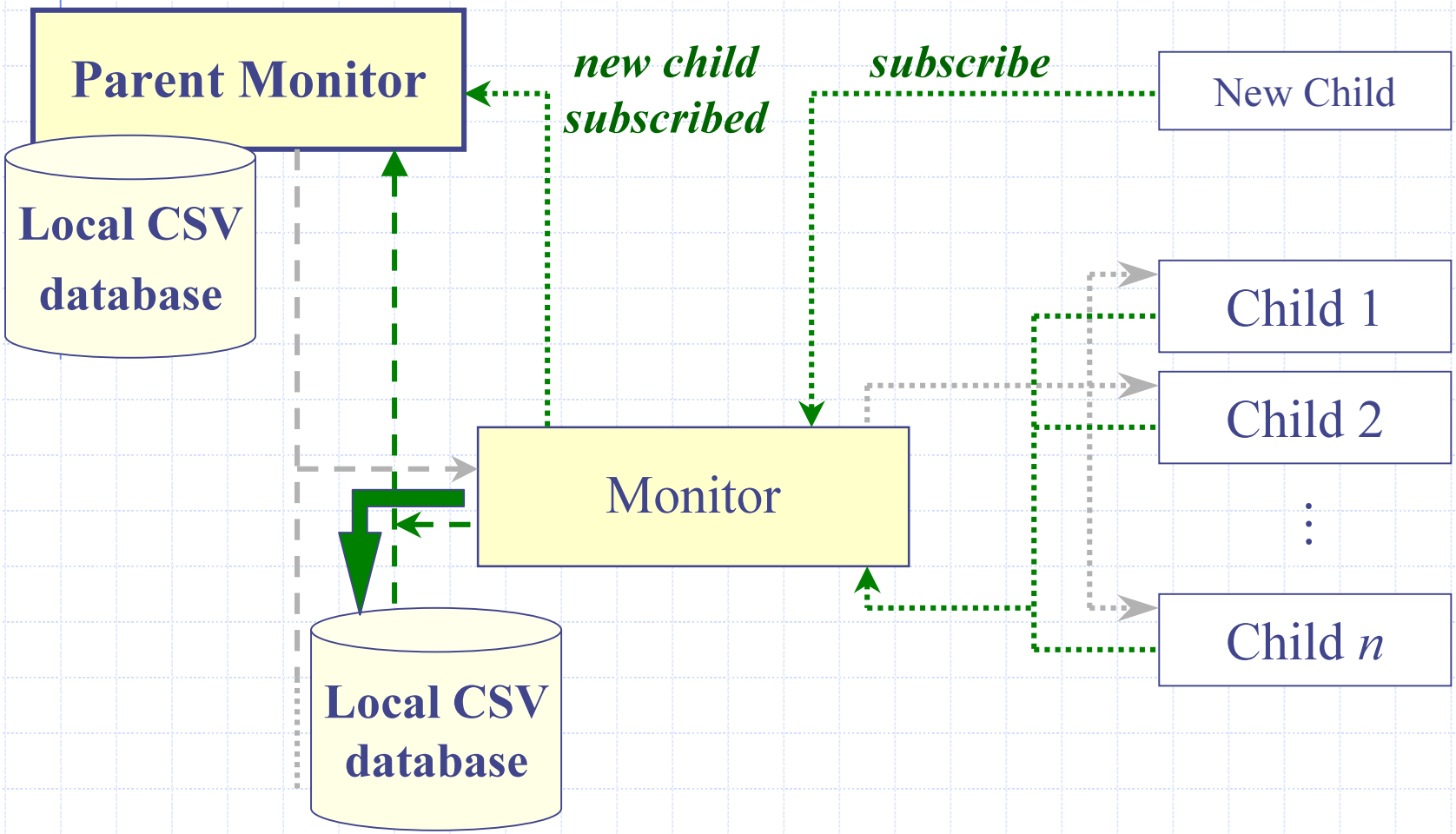
Monitoring

- In order to develop and deploy sensible Resource Broker we need monitoring framework
- Frequent data updates, large data volume for large number of computers

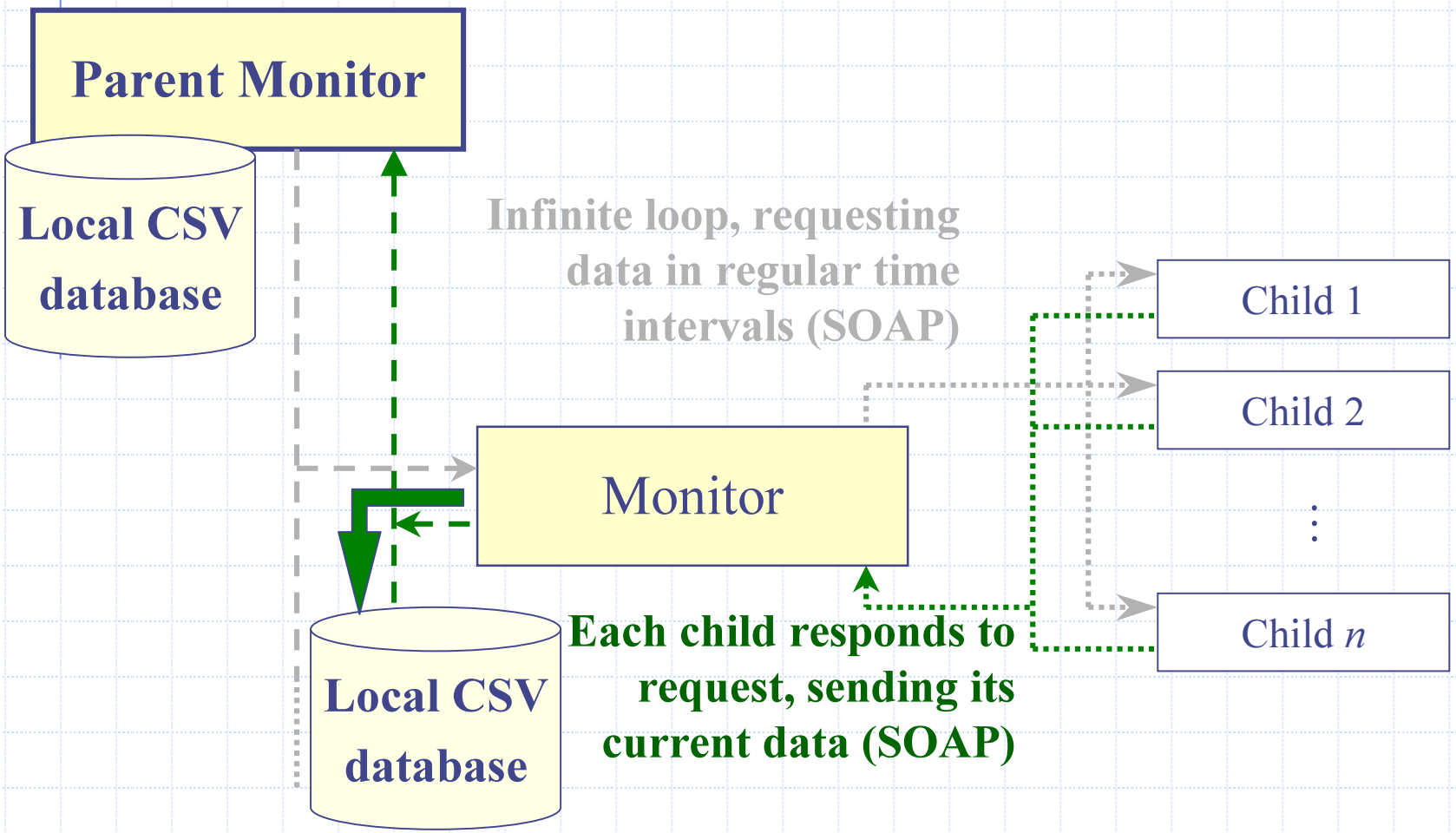


- The idea is to implement hierarchy of clients and servers where each client (child) maintains the history of measurements reports the summary information to upper layer (parent) using SOAP protocol

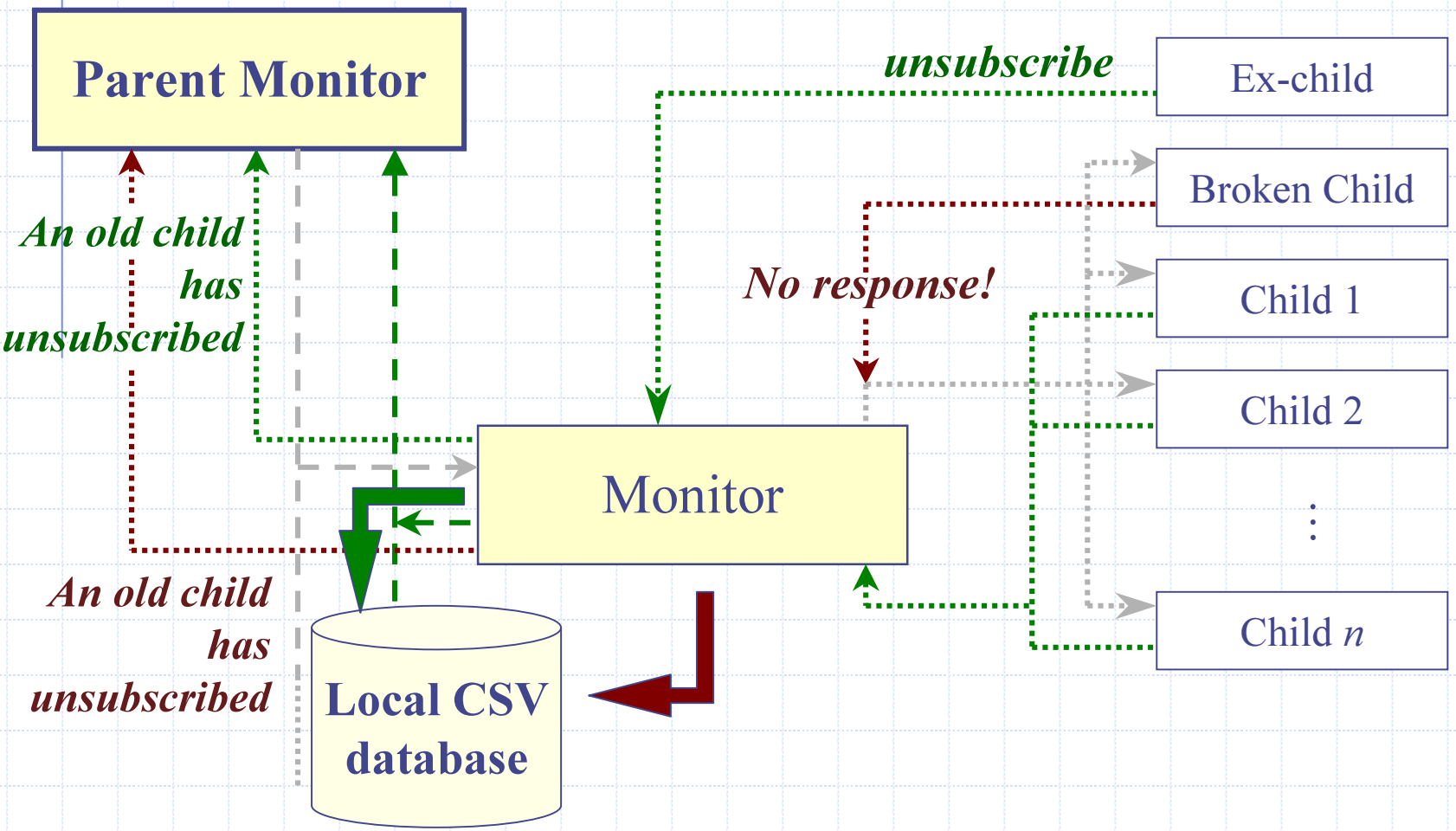
Subscribe


Parent Monitor
Local CSV database
new child subscribed
subscribe
New Child
Child 1
Child 2
Child n
Monitor
Local CSV database

Data collection

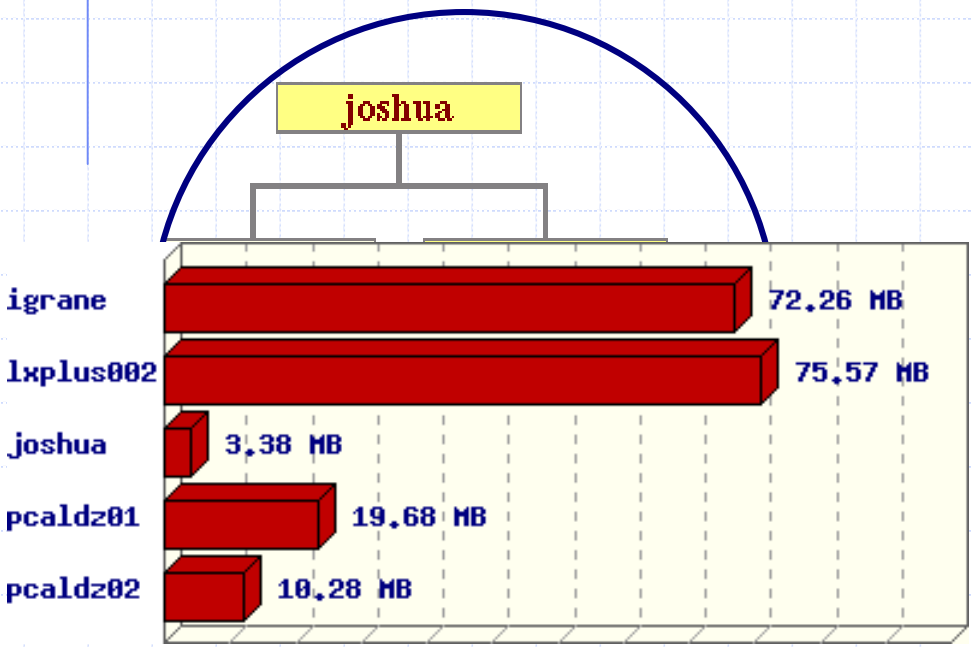


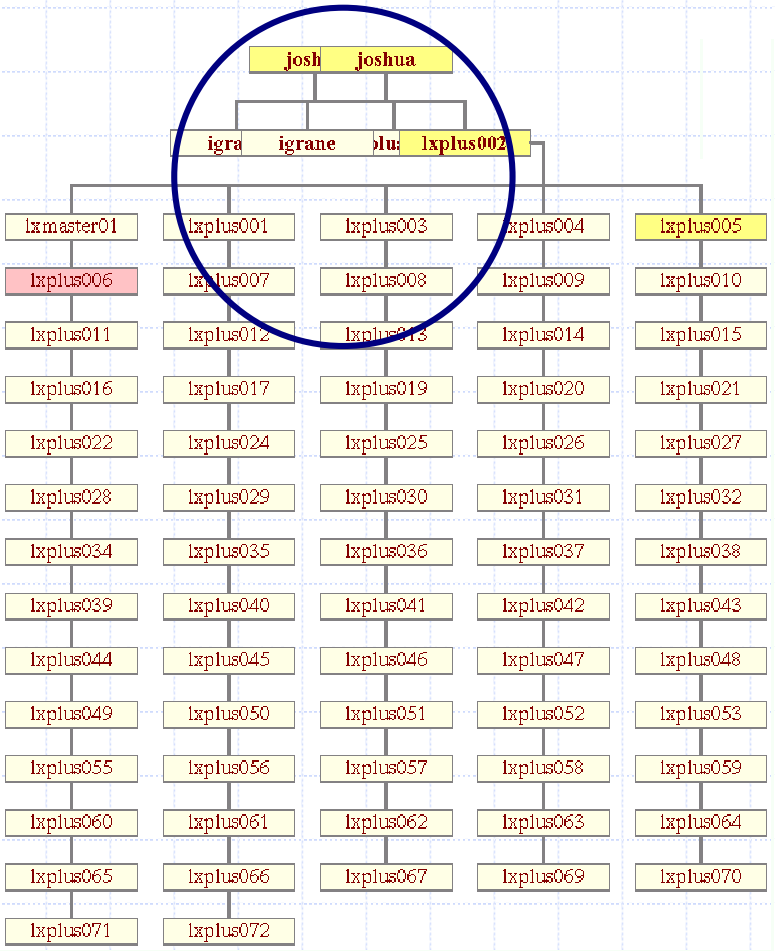
Disconnect and failure



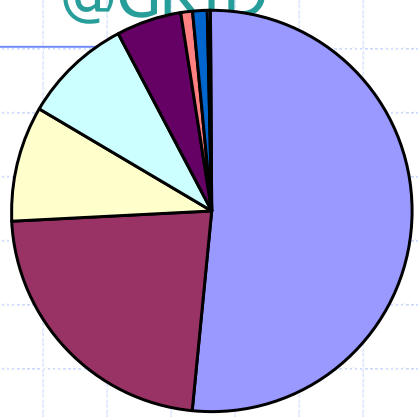
Data representation

- Preview of the whole GRID

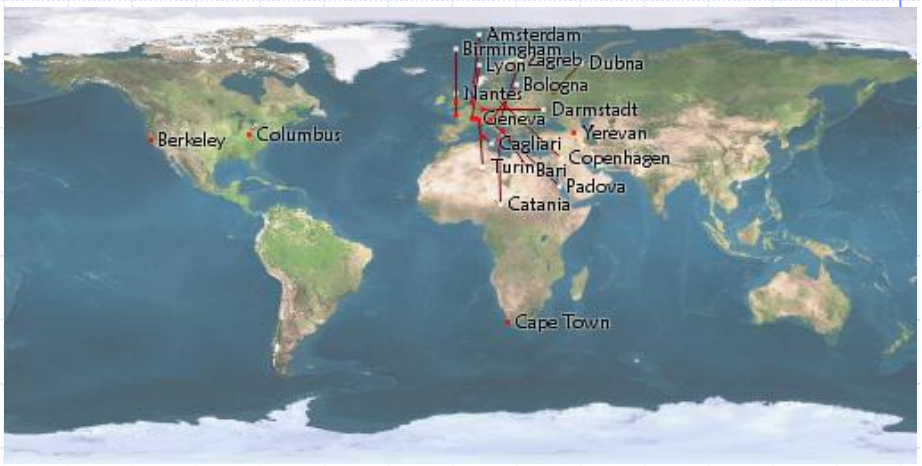




Production Summary



10^5 CPU hours



13 clusters, 9 sites

- 5682 events validated, 118 failed (2%)
- Up to 300 concurrently running jobs worldwide (5 weeks)
- 5 TB of data generated and stored at the sites with mass storage capability (CERN 73%,CCIN2P3 14%, LBL, 14%, OSC 1%)
- GSI, Karlsruhe, Dubna, Nantes, Budapest, Bari, Zagreb, Birmingham(?), Calcutta in addition ready by now

The Spiral Model

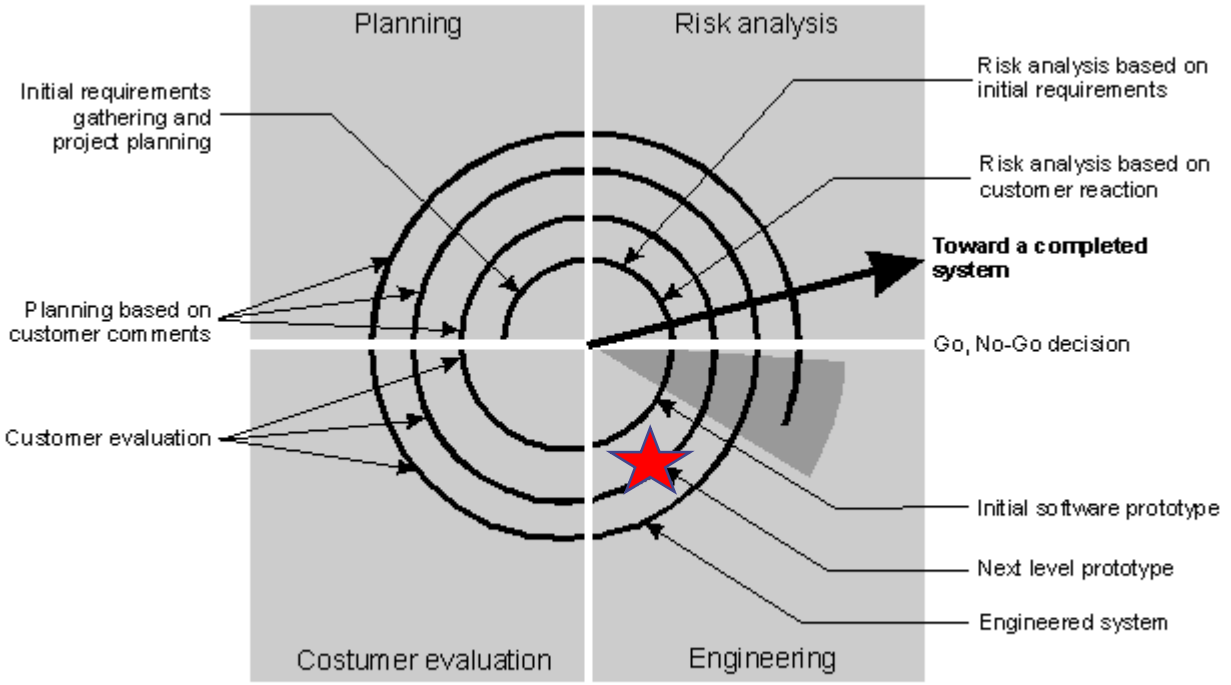


Figure 3 Program Structure Diagrams depicted on the spiral model ([PRES92] p 29

1. Planning: determination of objectives, alternatives and constraints
2. Risk analysis: analysis of alternatives and identification/resolution of risks
3. Engineering: development of the "next level" product;
4. Customer evaluation: assessment of the results of engineering.

Plans

- ✓ C/C++ API (January)
- ✓ GSI/Globus certificate authentication (March)
- Adding new services
  - ✓ Monitoring and exception handling (February)
  - ✓ Use Condor ClassAds for resource description and matching(February)
  - Queue optimization (March)
  - Support for interactive jobs (April)
- Virtual datasets (May)
- Disk pool/cache manager (June)
- Modular “kernel” using POE (July)
- Implementation of Web services (August)
  - ✓ SOAP (Simple Object Access Protocol)
  - WSDL (Web Services Description Language)
  - UDDI (Universal Description Discovery & Integration)

How to proceed?

Follow up ongoing in-house developments (DataGRID)
Maintain compatibility and use standard solutions

Keep Alice users happy

Look towards the future...

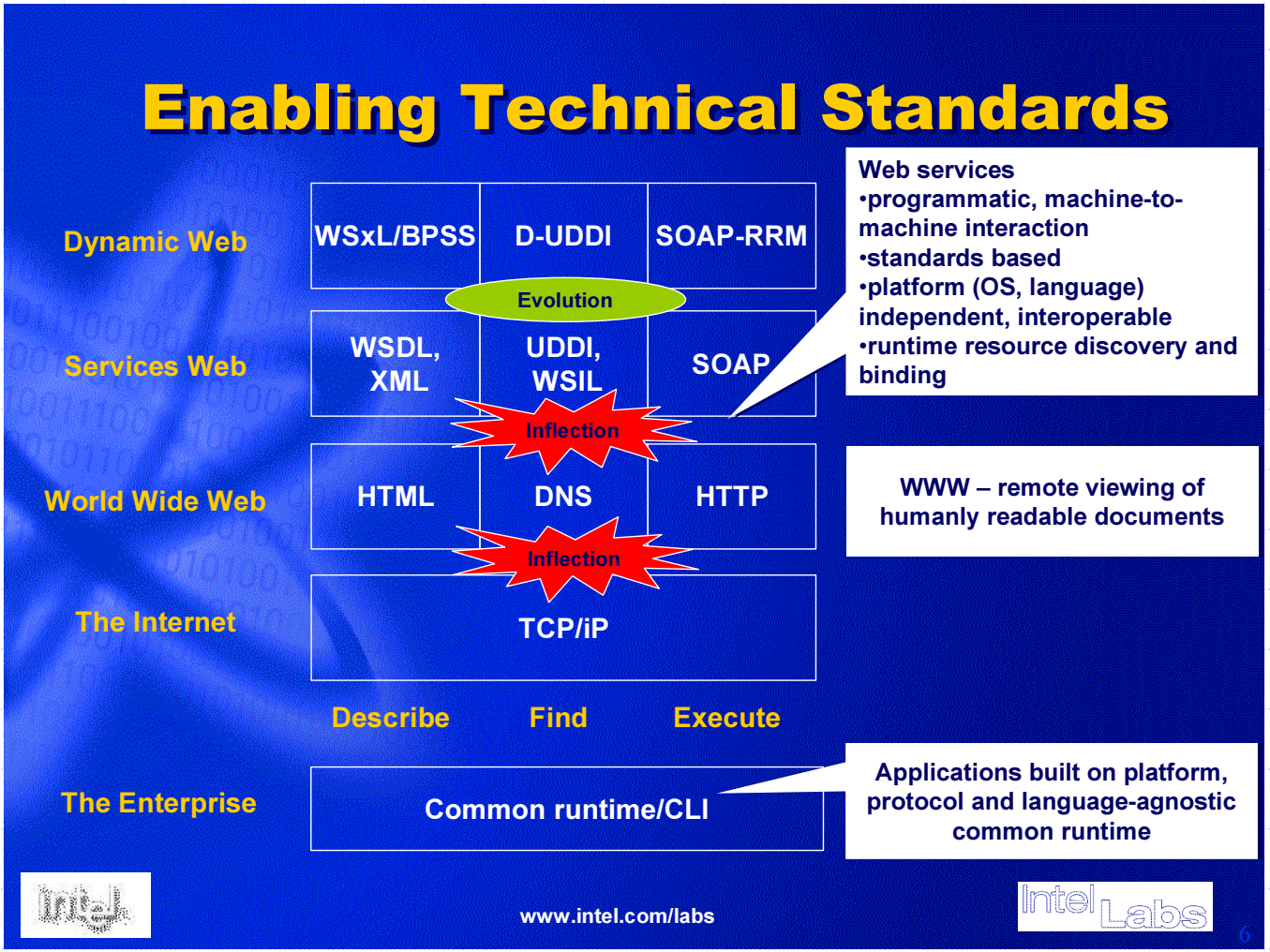
History

Summary: Evolution of Grid Technologies

- Initial exploration (1996-1999; Globus 1.0)
  - Extensive appln experiments; core protocols
- Data Grids (1999-??; Globus 2.0+)
  - Large-scale data management and analysis
- Open Grid Services Architecture (2001-??, Globus 3.0)
  - Integration w/ Web services, hosting environments, resource virtualization
  - Databases, higher-level services
- Radically scalable systems (2003-??)
  - Sensors, wireless, ubiquitous computing

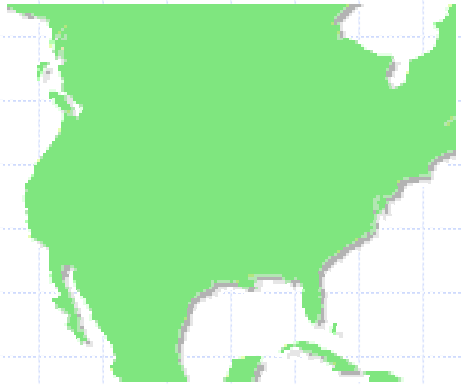
@GRID

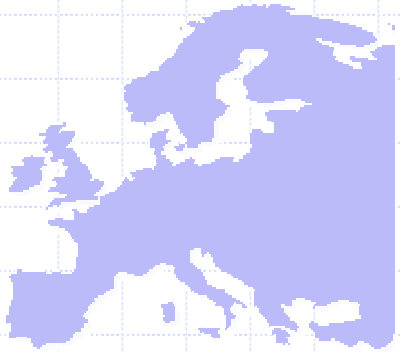
Future?


Enabling Technical Standards
Dynamic Web
WSxL/BPSS
D-UDDI
SOAP-RRM
Evolution
Services Web
WSDL, XML
UDDI, WSIL
SOAP
Inflection
World Wide Web
HTML
DNS
HTTP
Inflection
The Internet
TCP/iP
Describe
Find
Execute
The Enterprise
Common runtime/CLI
Web services
•programmatic, machine-to-machine interaction
•standards based
•platform (OS, language) independent, interoperable
•runtime resource discovery and binding
WWW – remote viewing of humanly readable documents
Applications built on platform, protocol and language-agnostic common runtime
www.intel.com/labs
Intel Labs

AliEn as a meta-GRID

AliEn User Interface		
iVDGL stack	AliEn stack	EDG stack





n
@GRID

Conclusions

- AliEn framework is a lightweight, simplified but functionally equivalent alternative to full blown GRID
- It is a partial solution which solves boundary conditions and current requirements for simulation, reconstruction for next generation HEP experiment(s)
- AliEn has passed first field tests and gearing up for next production
- We have picked up right direction (SOAP, Web services, standard components) one year ago and that gives us the competitive edge
- Following mainstream developments in computing pays off but one always must preserve critical judgment and not rush to implement latest and greatest buzzwords
- On mid to long term, ALICE remains committed to integrate AliEn with DataGRID solutions once they become available

The goal is not to deliver the most exciting and advanced computing exercise but to deliver data to ALICE users