



<http://cern.ch/arda>

CHIPP Meeting, September 15, 2004

“ARDA Status (ALICE Subgroup)”

Derek Feichtinger



eGEE

Enabling Grids for
E-science in Europe

www.eu-egee.org



cern.ch/lcg



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

Contents



1. ALICE PDC04 short overview

1. Strategy
2. Phase II purpose and tasks
3. Status
4. Statistics, LCG contribution

2. ARDA and EGEE

3. ARDA-ALICE activities

1. Goals and status
2. New C++ API to gLite
3. Interactive analysis

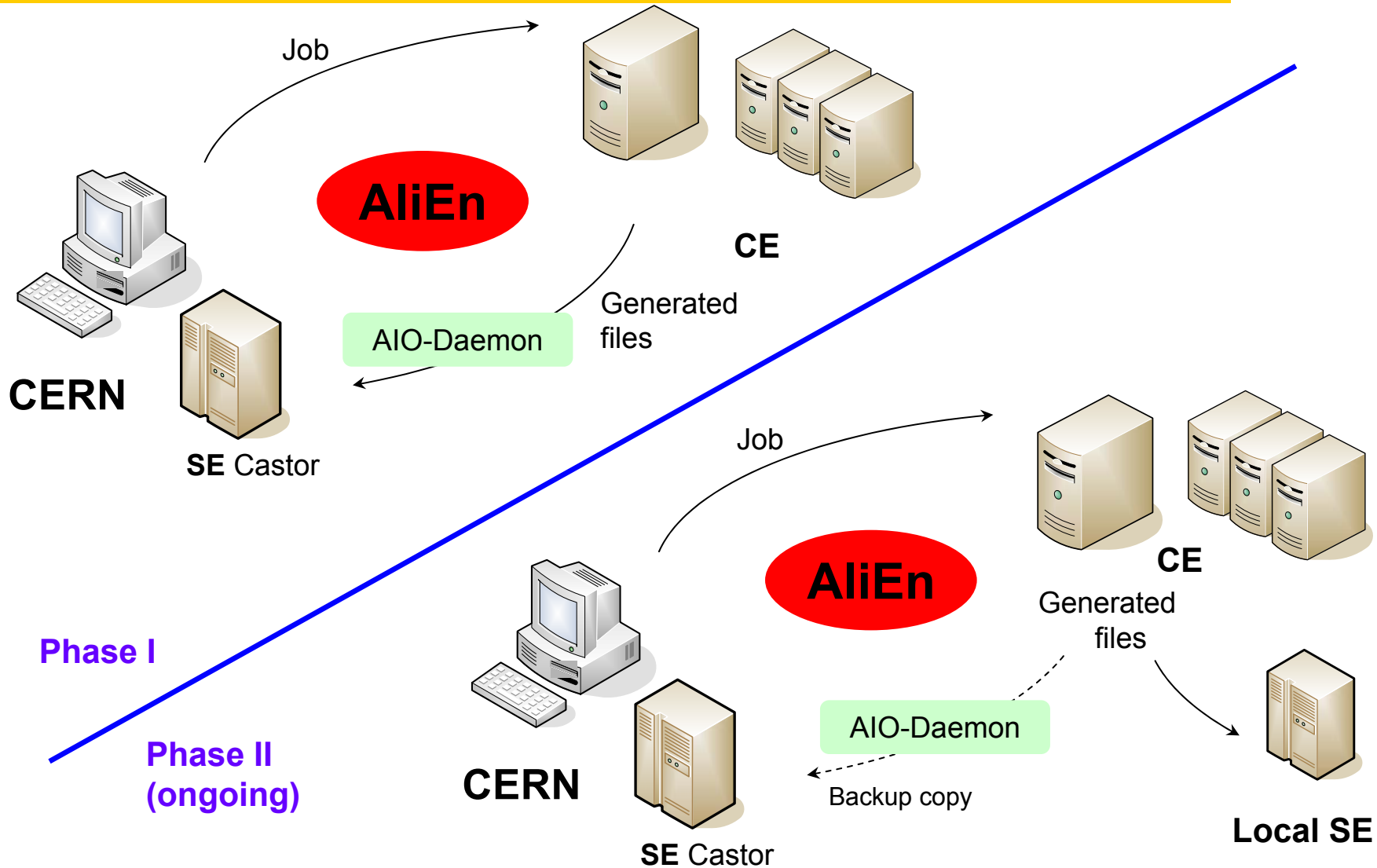




PDC04 Phase II purpose and tasks

- Mixing of signal events with different physics content into the underlying Pb+Pb events (underlying events are reused several times)
- Test of:
 - Standard production of signal events
 - Stress test of network and file transfer tools
 - Storage at remote SEs, stability (crucial for phase 3)
- Conditions, jobs:
 - 62 different conditions
 - 340K jobs, 15.2M events
 - 10 TB produced data
 - 200 TB data transfer from CERN
 - 500 MSI2K hours CPU

ALICE PDC04 Strategy



PDC04 Phase II Status

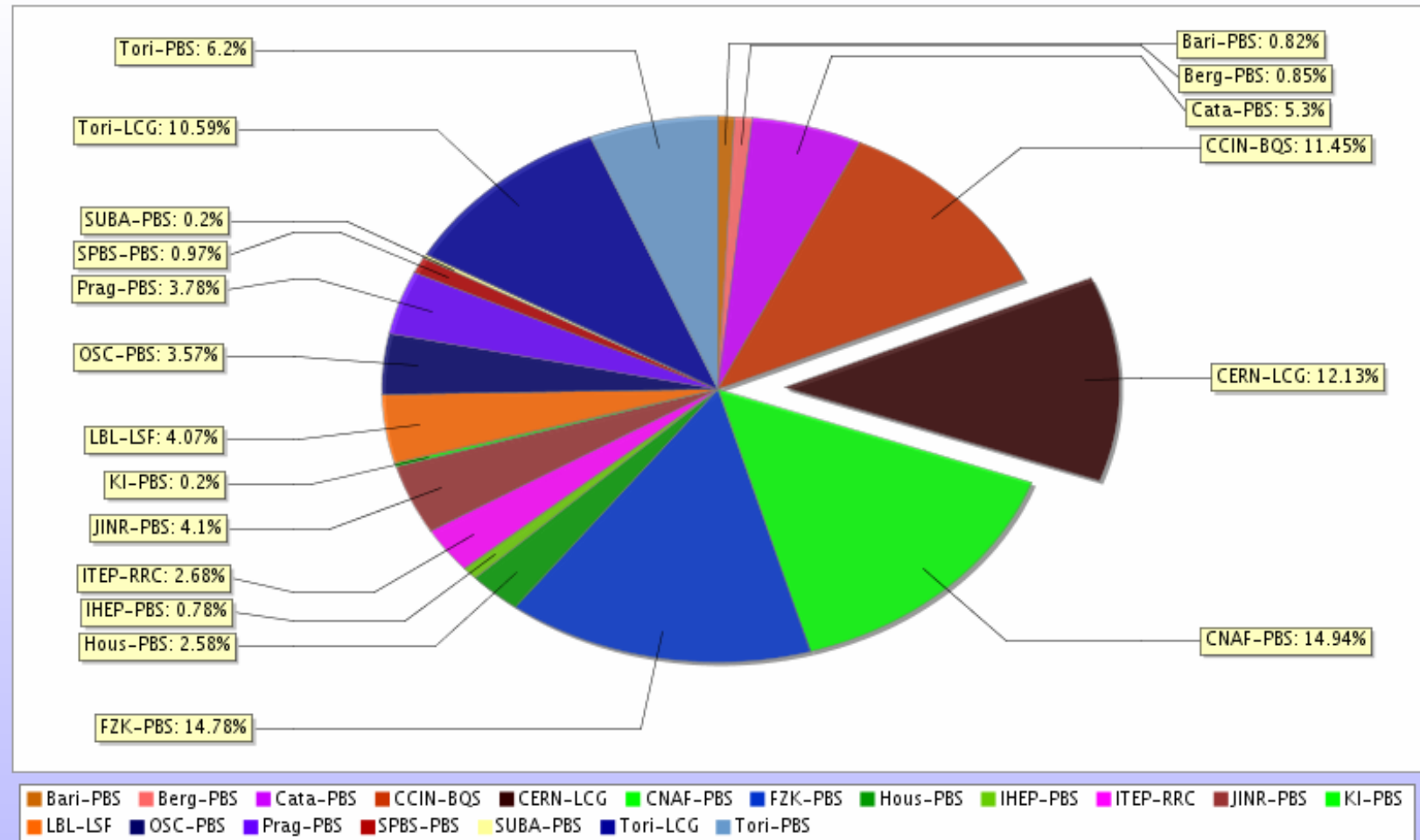


- Phase 2 statistics (start July 2004 – end September 2004):
 - Number of jobs: 75K (number of done jobs/day is accelerating)
 - Number of output files: 375K data, 300K log
 - Data volume: 3.2 TB at local SEs, 3.2 TB at CERN (backup)
 - Job duration: 2h 30min cent1, 1h 20min per1:
 - Careful profiling of AliRoot and cleaning up of the code has reduced the processing time by a factor of 2!

PDC04 Phase I CPU contributions



CPU contribution



LCG contribution



- ***LCG is seen as a single big CE by AliEN***
- ***Phase I (completed) – LCG contribution***
 - Successfully completed jobs 11k (out of 56k total)
 - Average job duration 7.5 hours
 - CPU used 57 MSI-2k hours
 - Number of jobs sent to LCG 14.5k (about 76% efficiency)
 - Storage: 26TB at CERN in CASTOR, seen as *AliEn* SE
 - Registration: *AliEn* Data Catalogue
- ***Phase II (ongoing) – LCG contribution***
 - Ca. 62% efficiency

ARDA Project



ARDA in a nutshell



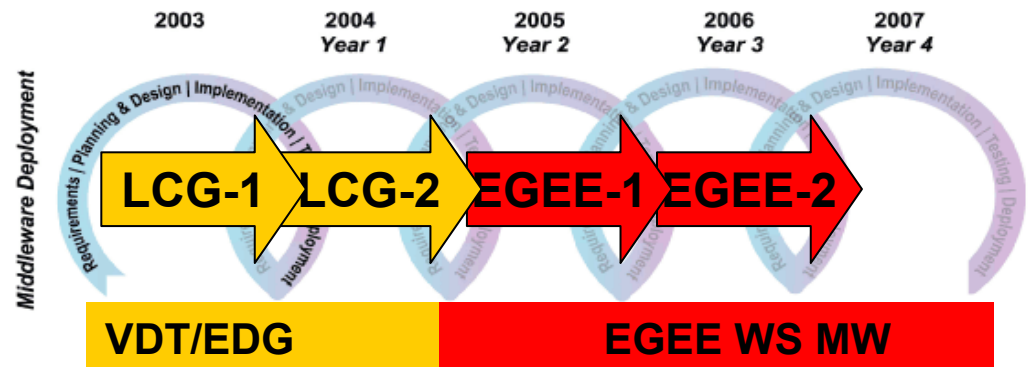
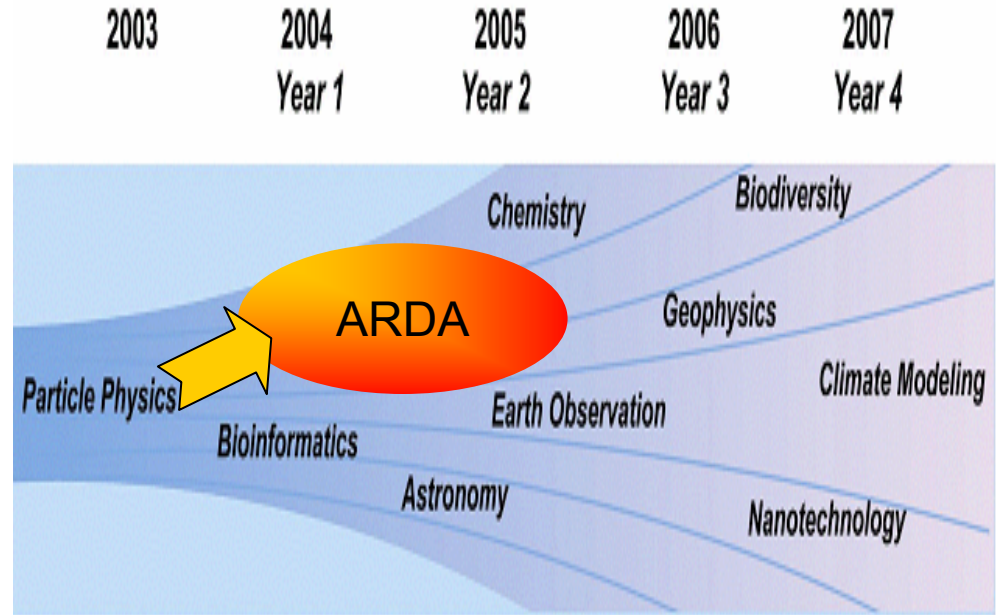
- ARDA is an LCG project whose main activity is to enable LHC analysis on the grid
- ARDA is coherently contributing to EGEE NA4 (using the entire CERN NA4-HEP resource)
- Use the grid software as it matures (EGEE project)
 - ARDA should be the key player in the evolution from LCG2 to the EGEE infrastructure
 - Provide early and continuous feedback (guarantee the software is what experiments expect/need)
- Use the last years experience/components both from Grid projects (LCG, VDT, EDG) and experiments middleware/tools (Alien, Dirac, GAE, Octopus, Ganga, Dial,...)
 - Help in adapting/interfacing (direct help within the experiments)
 - Every experiment has different implementations of the standard services, but:
 - Used mainly in production environments
 - Few expert users
 - Coordinated update and read actions
 - ARDA
 - Interface with the EGEE middleware
 - Verify (help to evolve to) such components to analysis environments
 - Many users (Robustness might be an issue)
 - Concurrent “read” actions (Performance will be more and more an issue)
- One prototype per experiment
 - A Common Application Layer might emerge in future
 - ARDA emphasis is to enable each of the experiment to do its job
- Provide a forum for discussion
 - Comparison on results/experience/ideas
 - Interaction with other projects
 - ...

The experiment interfaces agree with the ARDA project leader the work plan and coordinate the activity on the experiment side (users)

EGEE and ARDA



- Strong links already established between EDG and LCG. It will continue in the scope of EGEE
- The core infrastructure of the LCG and EGEE grids will be operated as a single service, and will grow out of LCG service
 - LCG includes many US and Asia partners
 - EGEE includes other sciences
 - Substantial part of infrastructure common to both
- Parallel production lines as well
 - LCG-2
 - 2004 data challenges
 - Pre production prototype
 - EGEE MW
 - ARDA playground for the LHC experiments



Preliminary activities



- Existing system as starting point
 - Every experiment has different implementations of the standard services
 - Used mainly in production environments
 - Few expert users
 - Coordinated update and read actions
 - ARDA
 - Interface with the EGEE middleware
 - Verify (help to evolve to) such components to analysis environments
 - Many users
 - » Robustness
 - Concurrent “read” actions
 - » Performance
- One prototype per experiment
 - A Common Application Layer might be used in the future
 - ARDA emphasis is to enable the experiment to do its job

Glite “disclosed” May 18th 😊

Since the beginning

Time consuming (see next section)

| Milestone | Date | Description |
|-----------|----------------|--|
| 1.6.2-5 | May 2004 | E2E x prototype definition agreed with the experiment |
| 1.6.6-9 | September 2004 | E2E x prototype using basic EGEE middleware |
| 1.6.14-17 | November 2004 | E2E x prototype improved functionality |
| 1.6.18 | December 2004 | E2E prototype for experiment x, capable of analysis |
| 1.6.19 | December 2005 | E2E prototype for experiment x, capable of analysis and production |

ARDA @ Regional Centres



- “Deployability” is a key factor of MW success
- A few Regional Centres will have the responsibility to provide early installation for ARDA to supplement the LCG preproduction service
- Stress and performance tests could be ideally located outside CERN
 - This is for experiment-specific components (e.g. a Meta Data catalogue)
 - Leverage on Regional Centre local know how
 - Data base technologies
 - Web services
 - ...
- Ease the interaction with “the rest of” HEP?
 - DESY
 - Non LHC experiments?
- Running ARDA pilot installations
 - Experiment data available where the experiment prototype is deployed
 - CERN, RAL, all Tier1s... The strategy is not clear yet
- As for the “Forum activities”, no milestones proposed for these activities

Subgroup Goals & status: ALICE



- Goals

- Developing a new C++ API, which interacts efficiently with the present system (service based)
 - gLite plugin for the ROOT data analysis framework (<http://root.cern.ch/>).
- Interactive Analysis: Building on the proof of concept done via ROOT/PROOF and AliEn in SC'03.

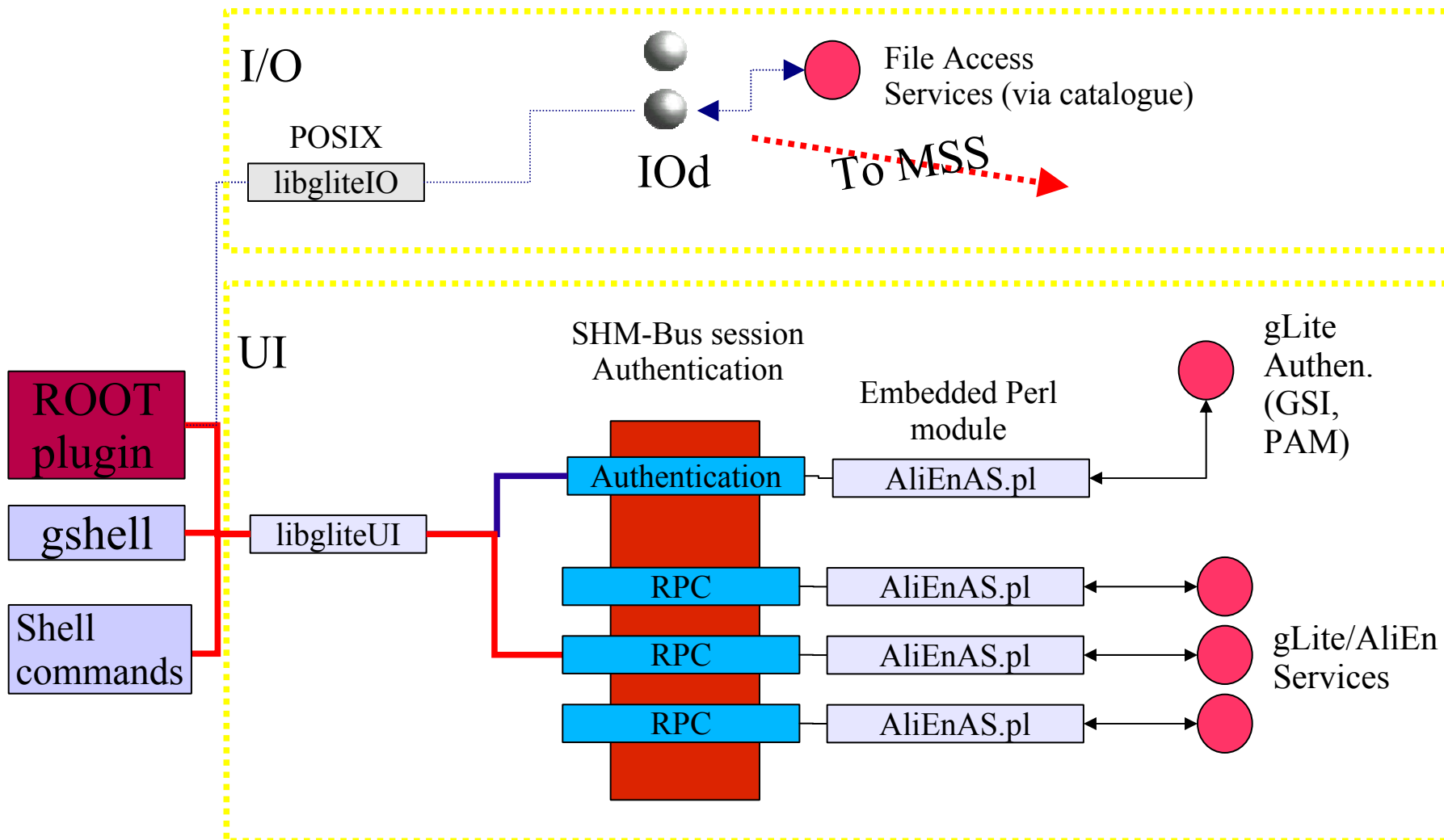


More automatization. Demonstration at SC'04 in November (Pittsburgh)

- Status

- API via grid access daemon deployed close to the core services. Currently integrating with gLite build system. Nearly finished.
- Subset of POSIX I/O via glite-IO-daemon (EGEE ported version of AIO-daemon).

Grid Access Library



ROOT/PROOF

<http://www.slac.stanford.edu/econf/C0303241/proc/papers/TULT003.PDF>



- Loop over data is coordinated by PROOF
 - Allows efficient parallelization
 - Data is a collection of ROOT Trees.
 - User just specifies a 'Selector' describing which events to select and how to process them.
- A Selector script can be run
 - In batch
 - Interactive ROOT _____ 1 machine
 - Interactive ROOT + PROOF _____ cluster
 - Interactive or batch ROOT + PROOF + gLite _____ Grid

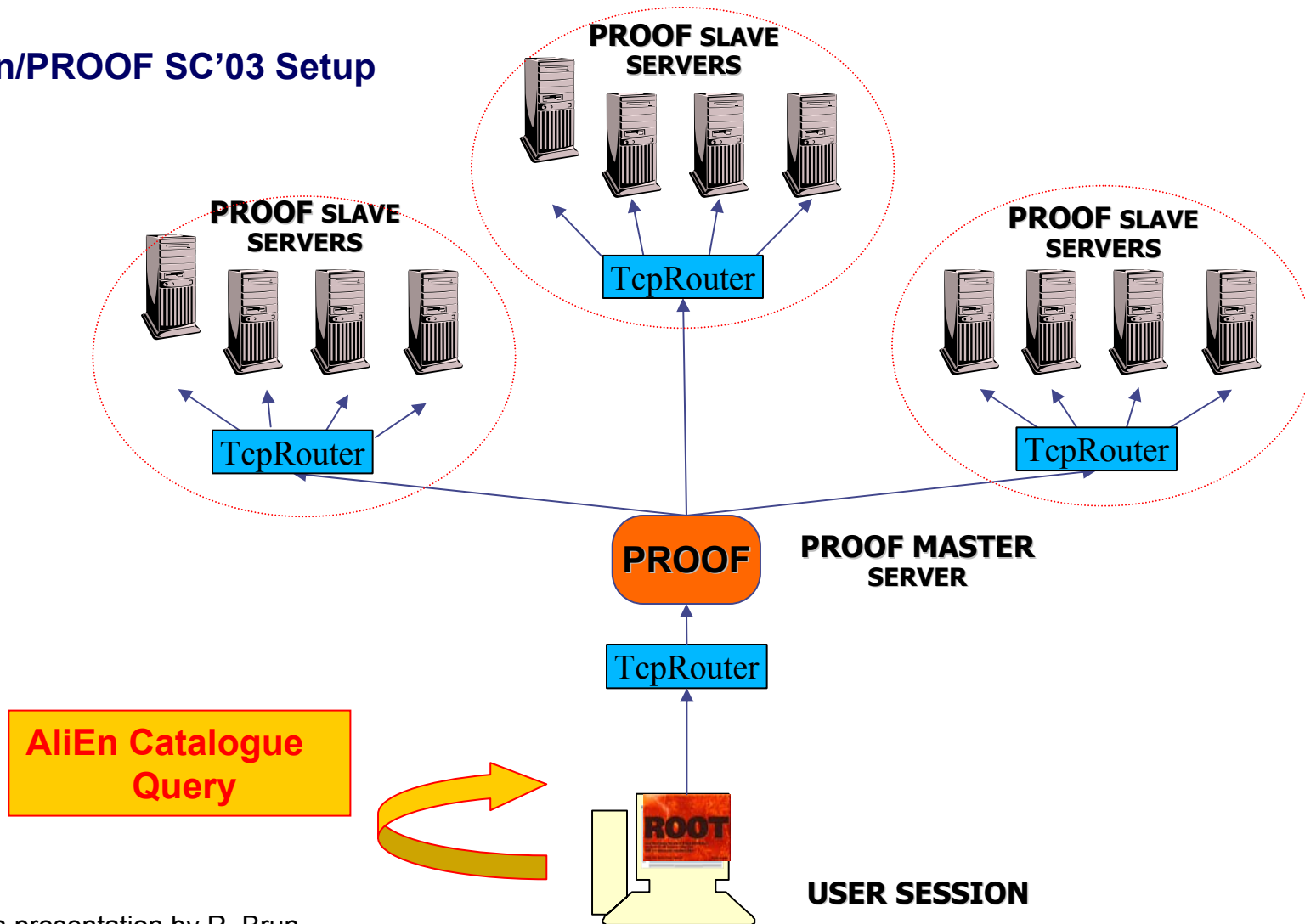
➡ Test selector locally on small data set, then use via PROOF
- A Selector script can be
 - Interpreted `tree.Process("myselector.C")`
 - Or compiled `tree.Process("myselector.C++")`

Slide modified from a presentation by R. Brun

ALICE: AliEn/ROOT/PROOF



AliEn/PROOF SC'03 Setup



Slide from a presentation by R. Brun

ARDA team



- Massimo Lamanna
- Birger Koblitz

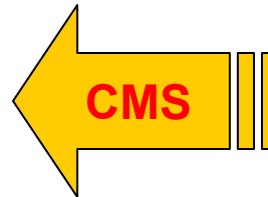
- Derek Feichtinger
- Andreas Peters



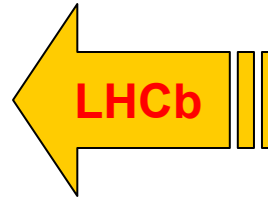
- Dietrich Liko
- Frederik Orellana



- Julia Andreeva
- Juha Herrala



- Andrew Maier
- Kuba Moscicki



- Andrey Demichev
- Viktor Pose



- Wei-Long Ueng
- Tao-Sheng Chen



Experiment interfaces

Piergiorgio Cerello (ALICE)
David Adams (ATLAS)
Lucia Silvestris (CMS)
Ulrik Egede (LHCb)