LHC Experiments Grid Integration Plans

C.Grandi

INFN - Bologna

Introduction

What do LHC experiments have to do:

- Build a distributed computing system for the experiment
- Test the prototypes:
  - tools evaluation
  - data challenges: tests with big data flows

What do LHC experiments have today:

Software

- EDG release 1.1 (1.2 coming)
- VDT release 1.0 (includes Globus, Condor-G, GDMP, …)
- Other tools for distributed computing (e.g. Web services, …)

Test facilities

- EDG test bed
- EDT supported experiment test beds
- US GriPhyN/PPDG/iVDGl experiment test beds

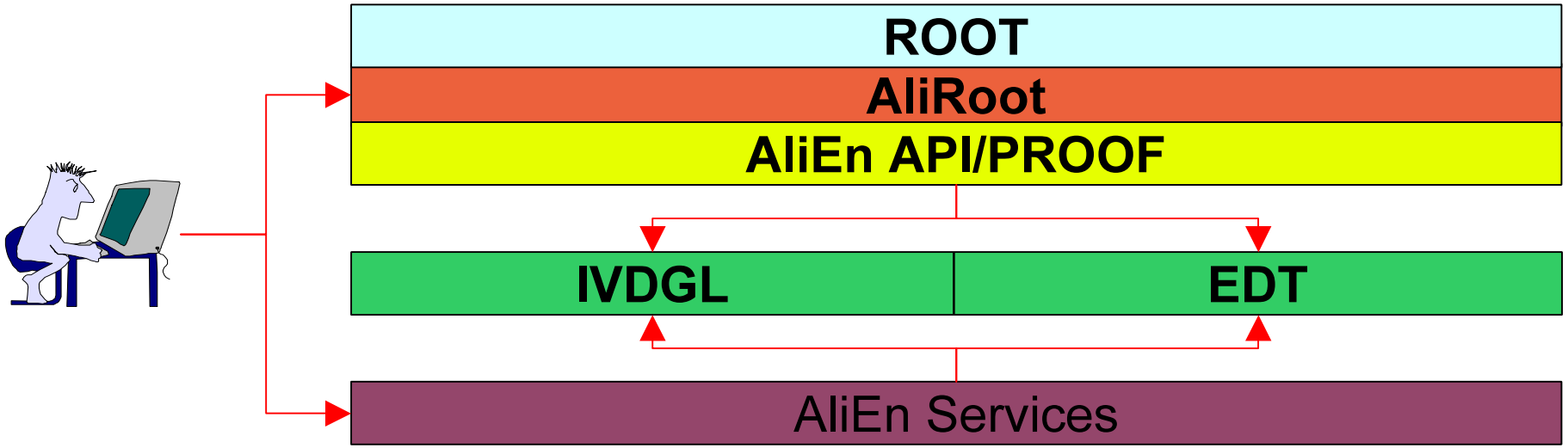
ALICE: Overview

Job submission/control/monitoring: AliEn @GRID

Application: AliRoot ⟷ Data Catalogue: AliEn

Data Persistency: Root

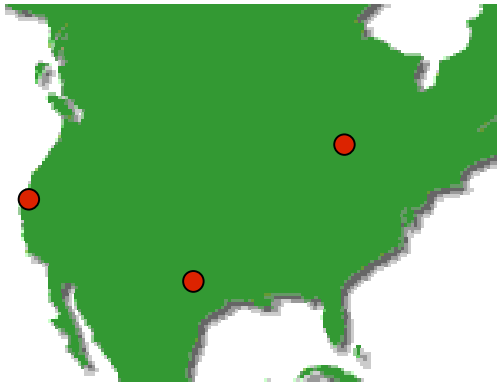
Job parallelisation: Root/PROOF



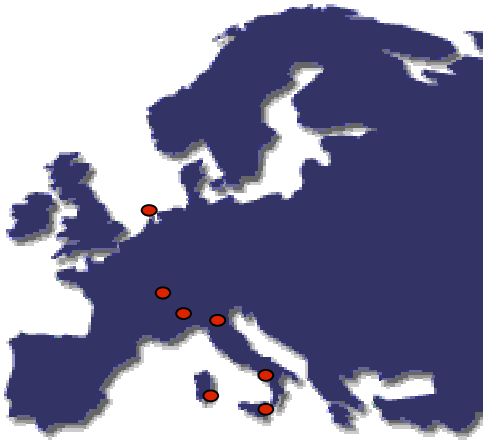
By P.Cerello

ALICE: Sites

Houston Univ. - Texas
OSU/OSC - Ohio
LBL - California

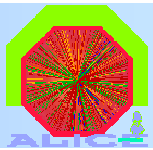


CERN - CH
INFN Cagliari/Catania/
CNAF/Salerno/Torino - I
NIKHEF - NL (?)



By P.Cerello

ALICE: Integration Items



Job submission:

- EDG/Resource Broker as an AliEn client
- Condor-G as an AliEn client

Data Management:

- register/access the AliEn Data Catalogue from an EDG/iVDGL/... Job

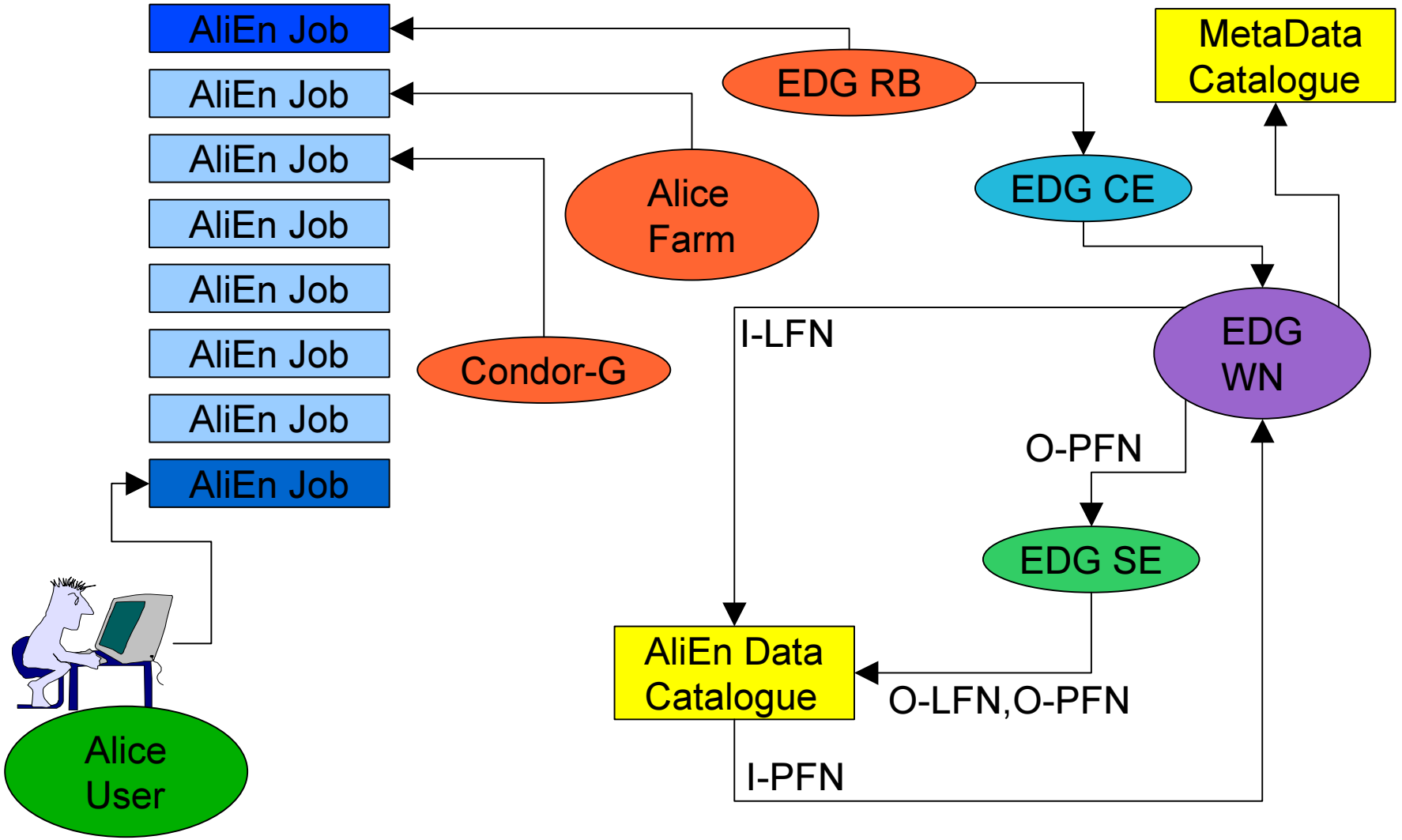
MetaData Catalogue: implement a Job MetaData Catalogue

- Test multiple accesses and concurrent updates
- Evaluate Spitfire to manage it

Spawn PROOF sub-jobs on any GRID

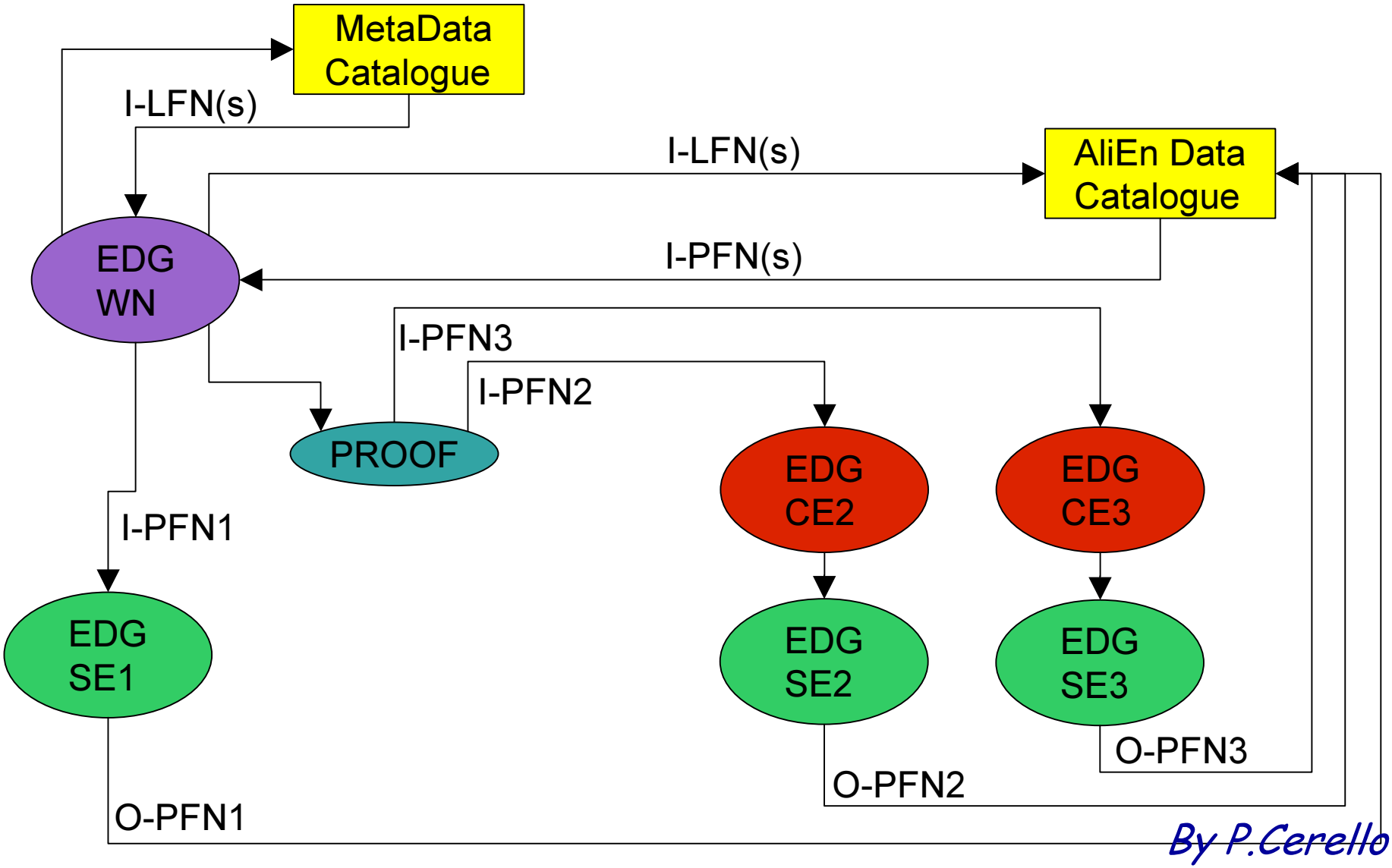
By P.Cerello

ALICE: Integration Plans



By P.Cerello

ALICE: Integration Plans



ATLAS: DC1 & 2

DC1/1: simulation of 10**7 evts for HLT TDR started 15 April. Bulk of production will start after 15 June

DC1/2: stress is on new software (20-9 till Xmas)

Use of currently avalable Grid tools favoured but not mandatory: start with EDG 1.2 in EU and VDT 1.0 in US:

- Some US sites deploy EDG software and vice-versa.
- Integration of Magda (US data catalog+some replica management) with EDG RC
- Critical issue the EDG stability and reliability

DC2: start spring 03

- Use EDG release 2
- Grid API to be inserted in ATHENA framework
- Magda and WP2 convergency plan to be detailed and executed
- Integrate VDT components

ATLAS: Overview

Main tools for data management on the grid:

Magda - MAnager for Grid Data

GANGA - Gaudi and Grid Alliance (with LHCb)

Other tools in development:

- GridView - simple script tool to monitor status of testbed (Java version being developed)
- Gripe - unified user accounts
- Pacman - package management and distribution tool
- Grappa - web portal based on active notebook tech.
- GRAT (GRid Application Toolkit): distribution of ATLAS software on the Grid

Integration with EDG is being discussed

ATLAS: Magda

Magda: MAnager for Grid-based Data

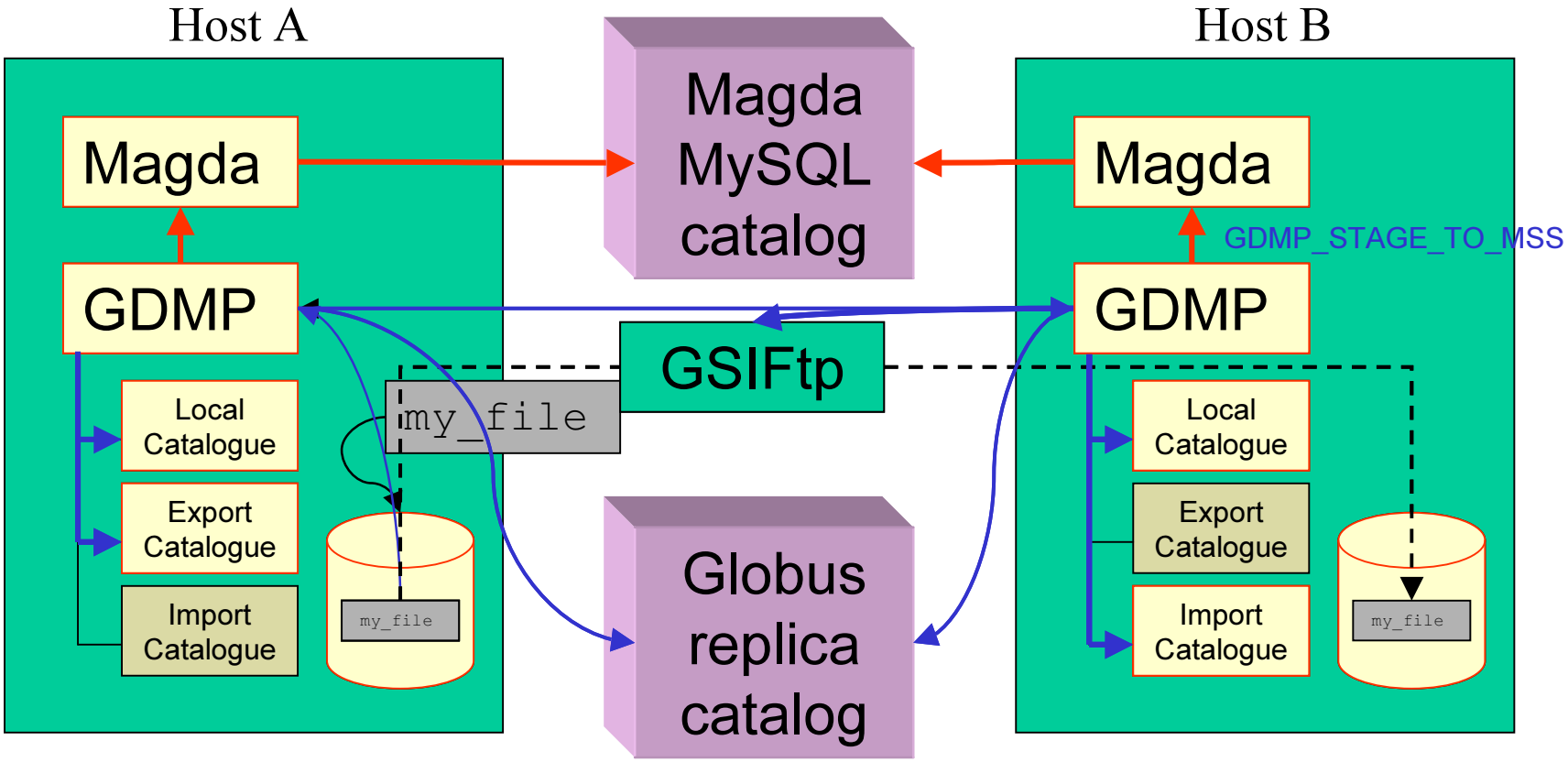
- Deliverable for the Particle Physics Data Grid
- MySQL database at the core of the system
- User interaction via command line and *web interface*
- File replication via GSIFtp
- Currently in use for ATLAS distributed data management prototyping and design
- 260k files, ~11TB cataloged at present

Synchronization between Magda & EDG catalogues

- ATLAS applications query Magda's MySQL catalogue;
- EDG Resource Broker queries the Globus Replica Catalog.

By D.Rebatto

ATLAS: Magda



By D.Rebatto

CMS: Overview

Build a unique CMS-GRID framework (EU+US)

EU and US grids not interoperable today

- **Help from DataTAG-iVDGL-GLUE**
- **Work in parallel in EU and US**

Main US activities:

- **MOP**
- **Virtual Data System**
- **Interactive Analysis**

Main EU activities:

- **Integration of IMPALA with EDG WP1+WP2 sw.**
- **Batch Analysis: user job submission & analysis farm**

CMS Production Framework

CMS: Schedule

2000-01: A few sites in production. Some grid tools used "in production" (GDMP).

2002: World-Wide productions. Prototypes of "distributed" production sites deployed in US and EU. DAQ TDR.

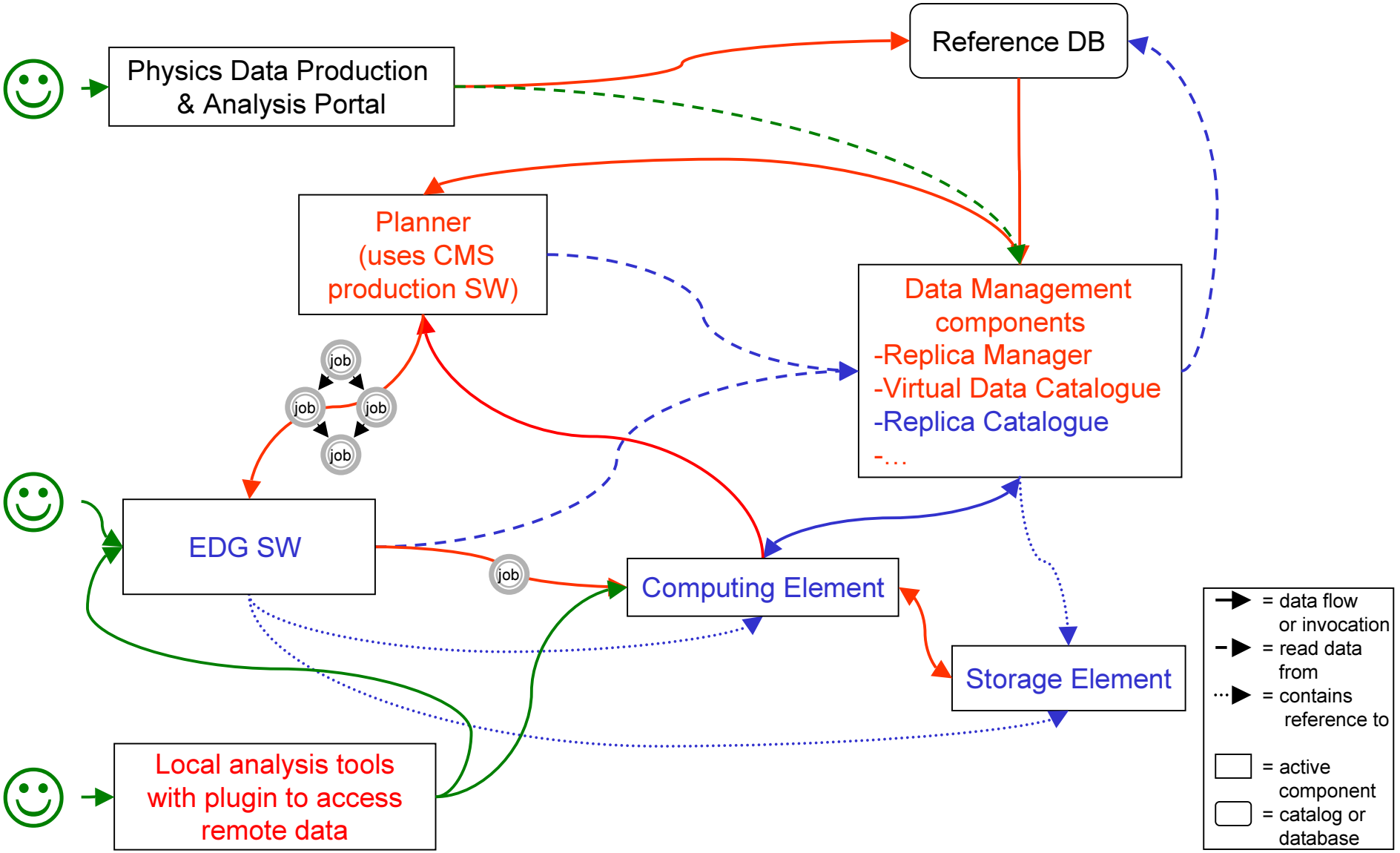
2003: Integration activities: US/EU grid interoperability. First grid-enabled sites used "in production".

2004: 5% data challenge. Deployment of delivered LCG prototypes. CCS TDR.

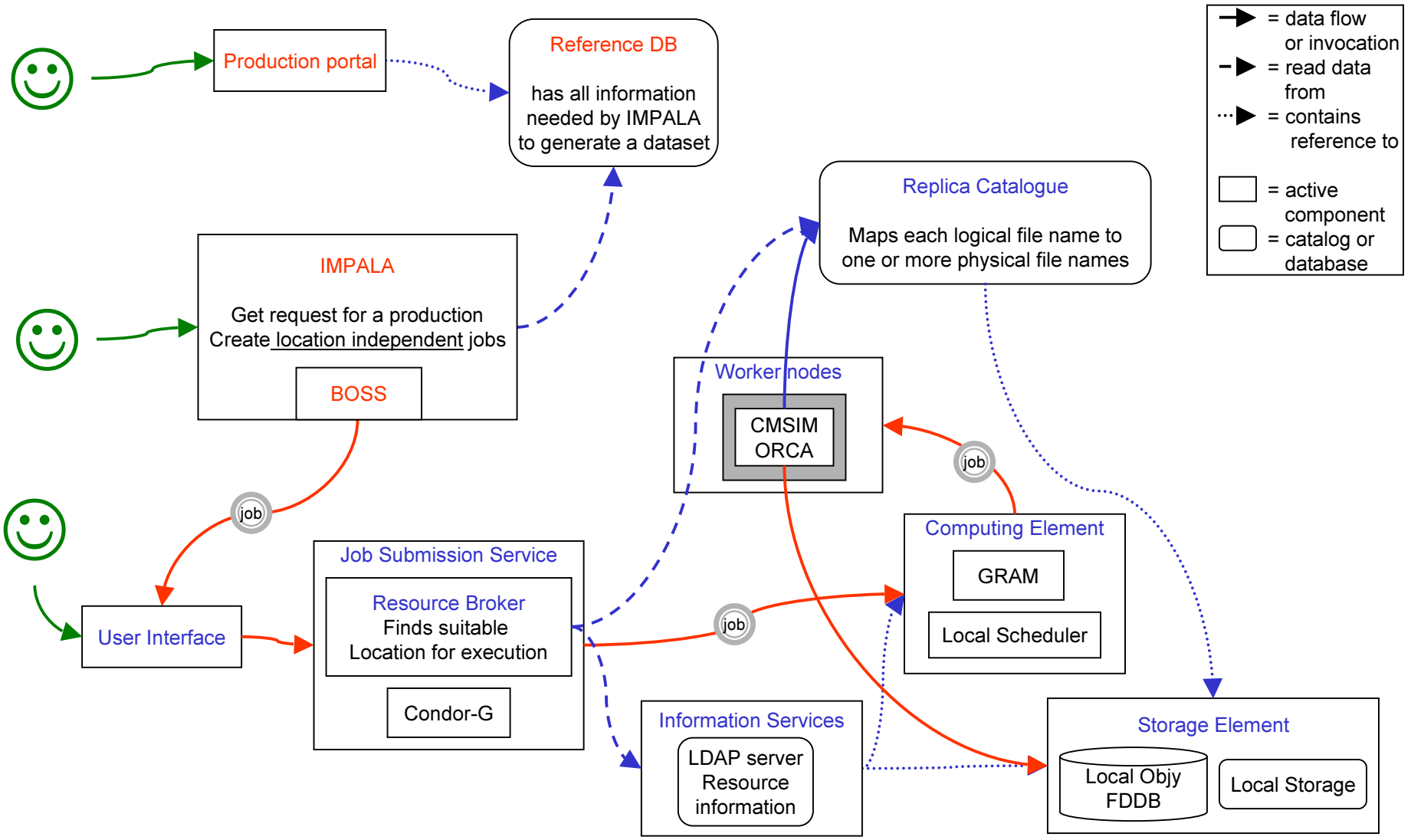
2005: preparation for 20% data challenge. Build "final" computing environment. Physics TDR.

Long term schedule being revised now!

CMS: EU-US integration 02



CMS: EDG integration 02



LHCb: Overview

GANGA (Gaudi and Grid Alliance) :

- In collaboration with Atlas developing user interface for all levels of user (physicist,production manager,developer)

Control and Monitoring system for distributed data production:

- Using PVSS today

Data challenges:

- DC1: 3-30 june + 26 august -22 september 2002
- DC2: 2003
- DC3: 2004
- DC4: 2005

LHCb: DC1

Computing tests

- prototype of new data management databases
- configuration of LHCb environment at remote sites and installation kits
- monitoring and control with PVSS
- integrability of EDG testbed into our production system
- data quality checking, evaluate DaVinci based tools
- use "push" MC job submission initially, perhaps test first version of pull mechanism during second half
- executables to be tested in production: sicbmc, Brunel, DaVinci

Data production

Produce useful data, for example for Physics studies with this facility, if the software is ready…

Resources:

100 – 150 CPU at CERN & ~300-350 CPUs outside (Bologna, IN2P3, RAL, Oxford, Bristol, Edinburgh, Nikhef, Cambridge, Barcelona, Moscow, Amsterdam VU?)

By J.Closier

LHCb: DC2 (2003)

Computing tests

- production version of new data management databases
- implementation of "pull" MC job submission system, exploiting EDG resource broker
- stress test Grid "philosophy" of submitting jobs without worrying where the job will run and where the output will be stored
- monitoring and control, evaluate EDG middleware (how does it work with PVSS? can it replace it?)
- make job submission tools independent of remote sites
- test of first Ganga prototype
- test of some Gaudi-Grid interfaces e.g. event selector
- test of automatic data quality histogram creation and checking in production
- executables to be tested: (sicbmc), Gauss, Brunel, DaVinci, Gaudi

Resources

Same sites as DC1 + Switzerland, Germany, Poland

By J.Closier

LHCb: DC3&4 (2004/5)

Computing tests for DC3

- test of first prototype of the analysis model (display events that were analysed on the grid)
- test of Ganga production version
- full use of Grid in production
- further tests of Gauss in production
- tests of data quality checking in production
- executables to be tested: Gauss, Brunel, DaVinci, Gaudi

Computing tests for DC4

test of production analysis model

the other points also mentioned previously

By J.Closier

Summary

Experiments are integrating Grid tools in their computing environment

- **Most of the experiments are using Grid tools delivered by more than one Grid Project**
- **Most of the experiments are not planning to use the tools *out of the box* but to adapt them to their needs**

- **Need help by middleware experts for *interoperability***
- **Need *non-standard* facilities to develop and test the experiment computing environment. Help by middleware experts is needed for that too**