An overview of production with grid technology accomplished by the LHC experiments

F Harris (OXFORD/CERN)

Presented by I Augustin (CERN)

Structure of talk

- Physics data production with grid tools 2002-2003, lessons learned and thinking for the future
  - ALICE
  - ATLAS
  - CMS
  - LHCb
- Summary

ALICE production in 2002

http://alien.cern.ch

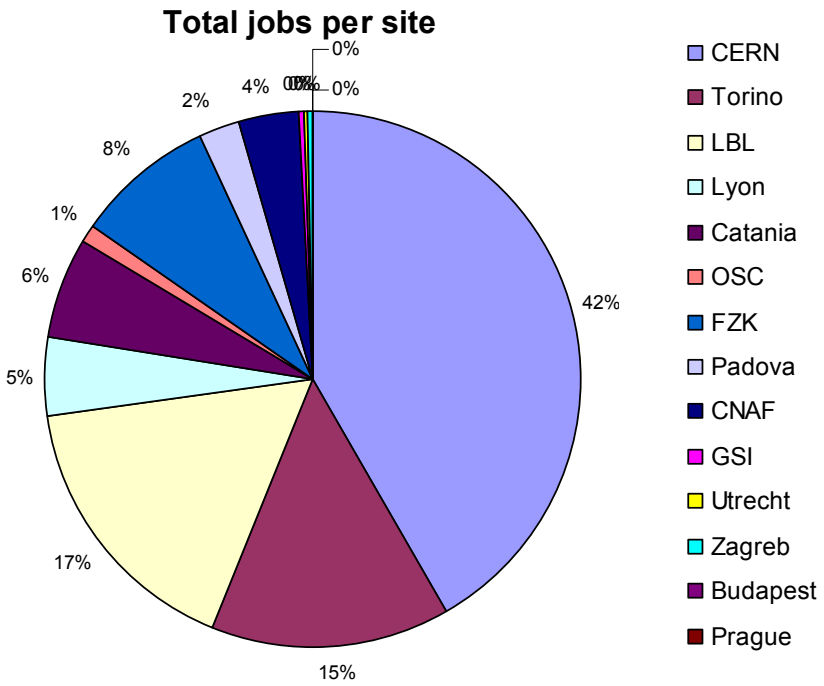
- 35 sites configured, with~14 contributing with CPU cycles
- 4 sites provided mass storage capability
- Fully distributed production controlled from one point
- 12 production rounds in 12 months

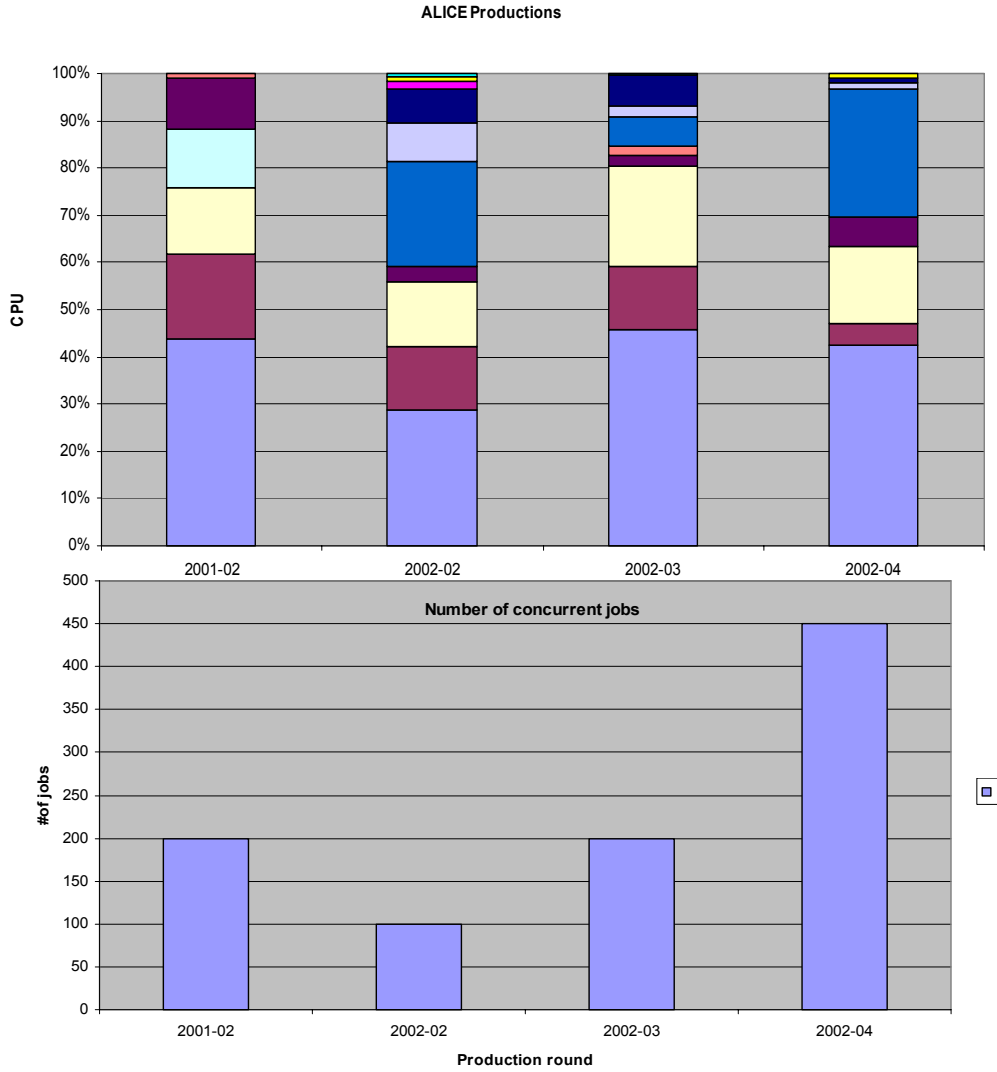
Edit +

ROUND	TAG	COMMENT	STATUS	COMMAND	Statistics
2001-01	V3.05	Test-Round	TESTING	AliRoot	view Chart
2001-02	V3.06	PPR-Production	DONE	AliRoot	view Chart
2002-01	V3.07.03	EMCAL-Production	DONE	AliRoot	view Chart
2002-02	V3.08.03	Proton-proton minimum bias for charm	DONE	AliRoot	view Chart
2002-03	V3.08.Rev.01	PPR production	STARTED	AliRoot	view Chart
2002-04	V3.08.Rev.01	p-p minimum bias	DONE	AliRoot	view Chart

http://alien.cern.ch/Alien/main?task=production

ALICE Production Summary 2001-2





22713 jobs, ~12CPUh/job,
~1GB output/job
up to 450 concurrently running jobs
Achieved well over 90% total efficiency

ALIEN and interfacing to EDG and LCG

- The interface between AliEn and EDG (1.4) has been successfully tested in 2003
- ALICE has also carried out a test production on EDG 1.4 (at a time when support for application TB was diminishing - Mar toMay 2003)
  - **Aim was 5000 jobs 12-24 hours per job**
  - **The average efficiency was only ~35%, so did not complete the task**
  - **Jobs failed generally for services failures (*RB overloaded, WN disk full, LDAP failure etc*)**
  - **Jobs generally lasted longer than the time-scale stability of the application testbed**
- The interface with LCG-1 is being developed now and tests are proceeding
- Will do more evaluations on EDG App TB when it has stabilised

ALIEN- a taste of future developments and some ARDA thinking

- modeling and simulation of AliEn job scheduling and execution (aiming to verify Alice computing model and perhaps come with cost estimator for AliEn queries). Work in progress.
- verify scalability of file and metadata catalogue
- use alternative transport protocols (jabber) to route SOAP messages in order to reduce need for inbound connectivity on gatekeeper nodes
- interface to LCG (at CE level and SE). Work in progress.
- use monitoring information to build high level optimizers and controls
- Portal service for AliEn (should make it easier to build Portals for multiple V.O.). Work in progress.
- **With ARDA in mind**
  - making AliEn services OGSI compatible
  - expose several more components as Grid services

ATLAS DC1 (July 2002-April 2003)

- **DC1 was divided in 3 phases**
  - **Phase 1 (July-August 2002)**
    - **Event generation and detector simulation**
  - **Phase 2 (December 2002 – April 2003)**
    - **Pile-up production**
  - **Phase 3 (April-Sept 2003)**
    - **Reconstruction**

- **Worldwide exercise with many participating institutes and developing use of grid technology in US Grids, Nordugrid and EDG Application Testbed**

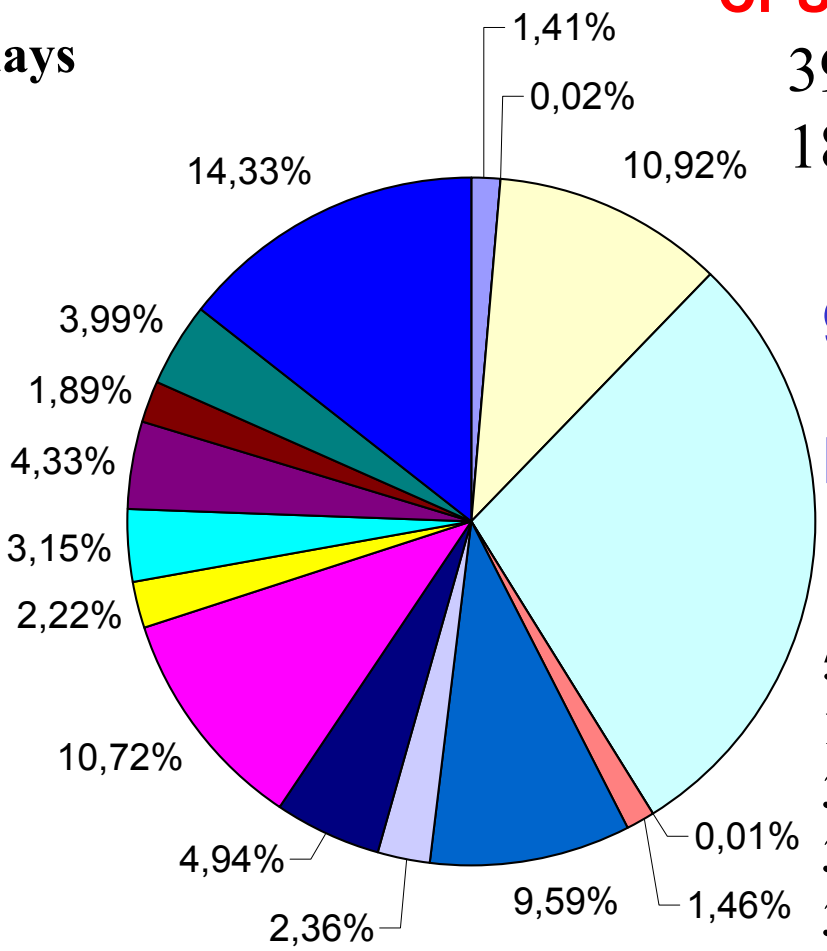
ATLAS DC1 Phase 1 : July-August 2002

3200 CPU's
110 kSI95
71000 CPU days

1. Australia
2. Austria
3. Canada
4. CERN
5. Czech Republic
6. France
7. Germany
8. Israel
9. Italy
10. Japan
11. Nordic
12. Russia
13. Spain
14. Taiwan
15. UK
16. USA

Contribution to the overall CPU-time (%) per country

39 Sites in
18 Countries



grid tools used at 11 sites (5% of production)

5*10*^{7} events generated
1*10*^{7} events simulated
3*10*^{7} single particles
30 Tbytes
35 000 files

ATLAS DC1 production on the Grid (shared between Nordugrid, US Atlas Grid and later Atlas-EDG)

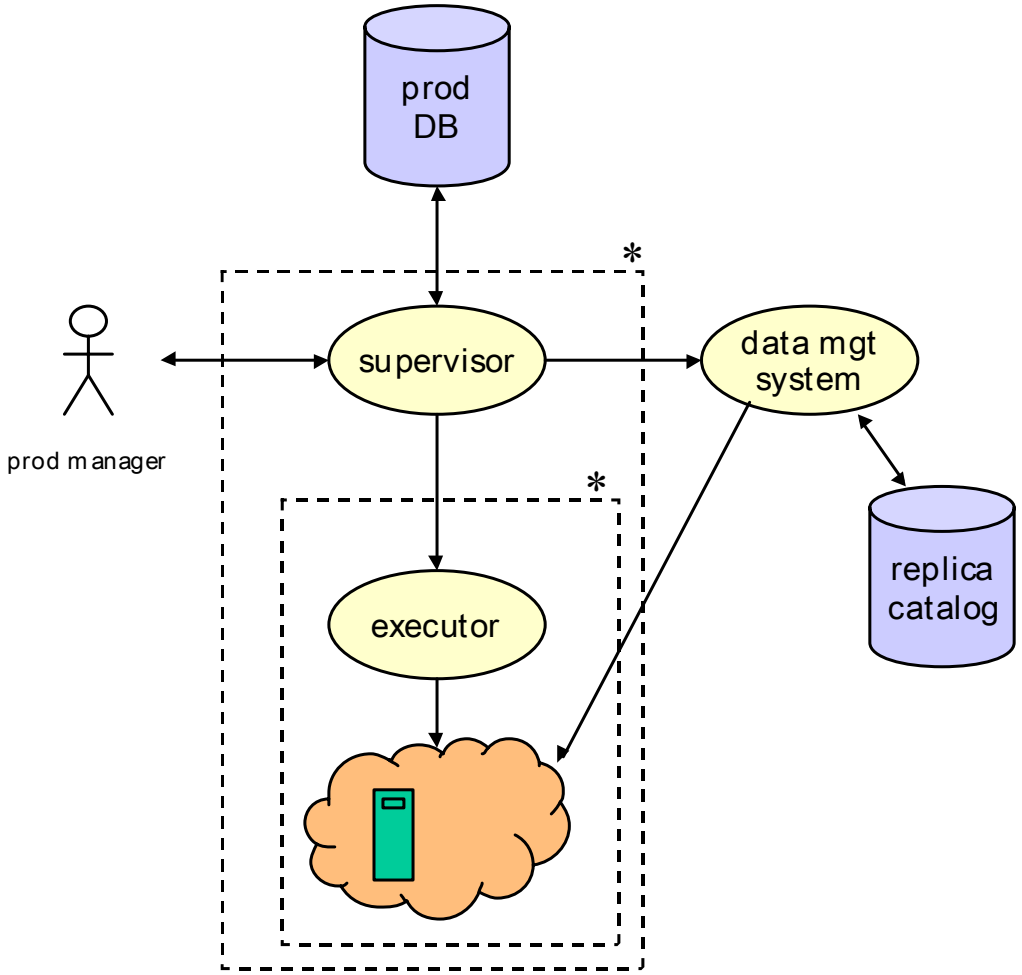
- **Phase 1 (summer 2002)**
  - **NorduGrid (Bergen, Grendel, Ingvar, ISV, NBI,Oslo,Lund,LSCF)**
    - **all Scandinavian production done on the Grid - ~1300 jobs**
  - **US-ATLAS-Grid (LBL, UTA, OU)**
    - **~10% of US production**

- **Phase 2-3 (late 2002 – summer 2003)**
  - **NorduGrid (full pile-up production & reconstruction) - ~ 4000 jobs**
  - **US ATLAS-Grid (BNL, LBNL, Boston U., UTA, Indiana U., Oklahoma U, Michigan U., ANL, SMU)**
    - **Pile-up: (?25% US total) - ~ 6000 jobs**
    - **Reconstruction (25% of US total) ~ 800 jobs**
  - **ATLAS-EDG: in June 2003 successful reconstruction (250 jobs) on well controlled set of sites (CNAF,Lyon,Milan,Rome,Cambridge) – 90% efficiency**

Comments on performance of Grid systems for Atlas DC1

- Nordugrid
  - **Has run very reliably (90+ % efficiency) with customised MDS but developed own job submission , data management and user interfaces**
  - **Supported by small team**
  - **Were major 'grid' producer for Atlas**
- EDG
  - **First efforts helped debug EDG 1.N s/w**
  - **Reconstruction in June 2003 ran with good efficiency (90+ % efficiency) on 5 well controlled sites using EDG 1.4 s/w**
- US Atlas (VDT-all sites running same version)
  - **Simple approach at start- ->reached 80% efficiency in non labour intensive operation**
    - **Did not use resource discovery to optimize job submission. Only for checking s/w and configurations.**
    - **In the beginning, the biggest problems were in site configuration + s/w install.**
    - **Later the biggest problems were hardware and network failures.**
    - **Most of code in GRAT(Grid Application Toolkit) devoted to error recovery, cleanup, data verification etc. – allows to run 10 site production with 1 person.**

Schematic of New ATLAS DC2 System - integrating use of LCG.Nordugrid and US production



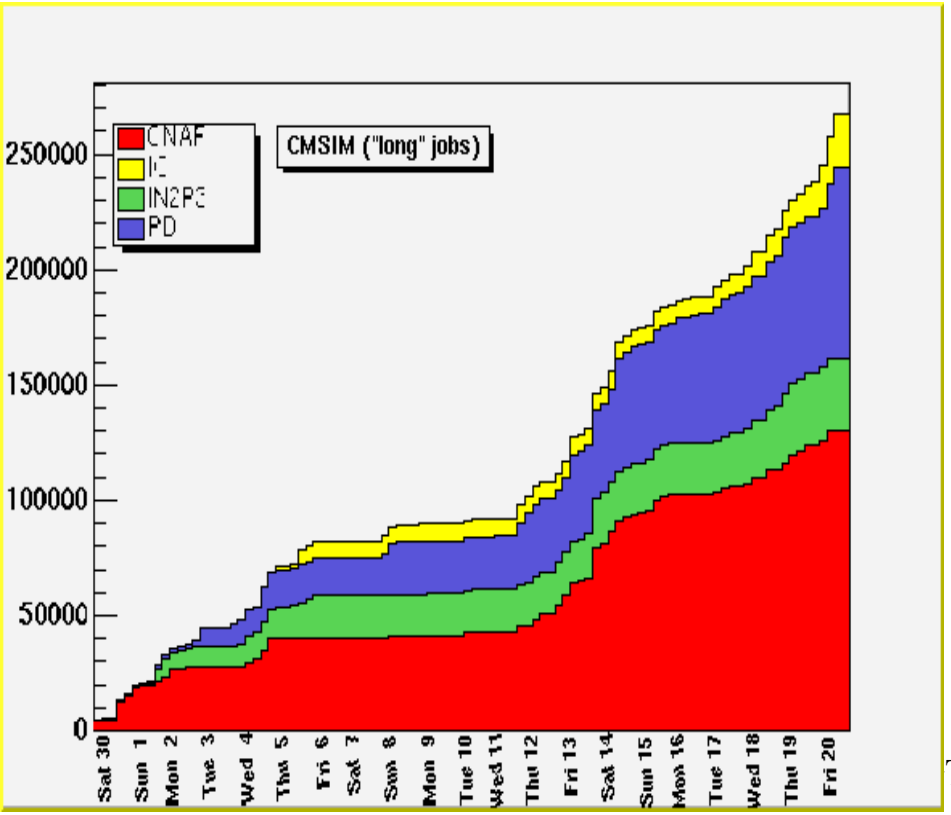
- Main features
  - Common production database for all of ATLAS
  - Common ATLAS supervisor run by all facilities/managers
  - Common data management system a la Magda
  - Executors developed by middleware experts (LCG, NorduGrid, Chimera teams)
  - Final verification of data done by supervisor

Pioneeering CMS Grid productions end 2002

CMS/EDG Stress Test on EDG testbed & CMS sites

Top-down approach: more functionality but less robust, large manpower needed

260,000 events in 3 weeks

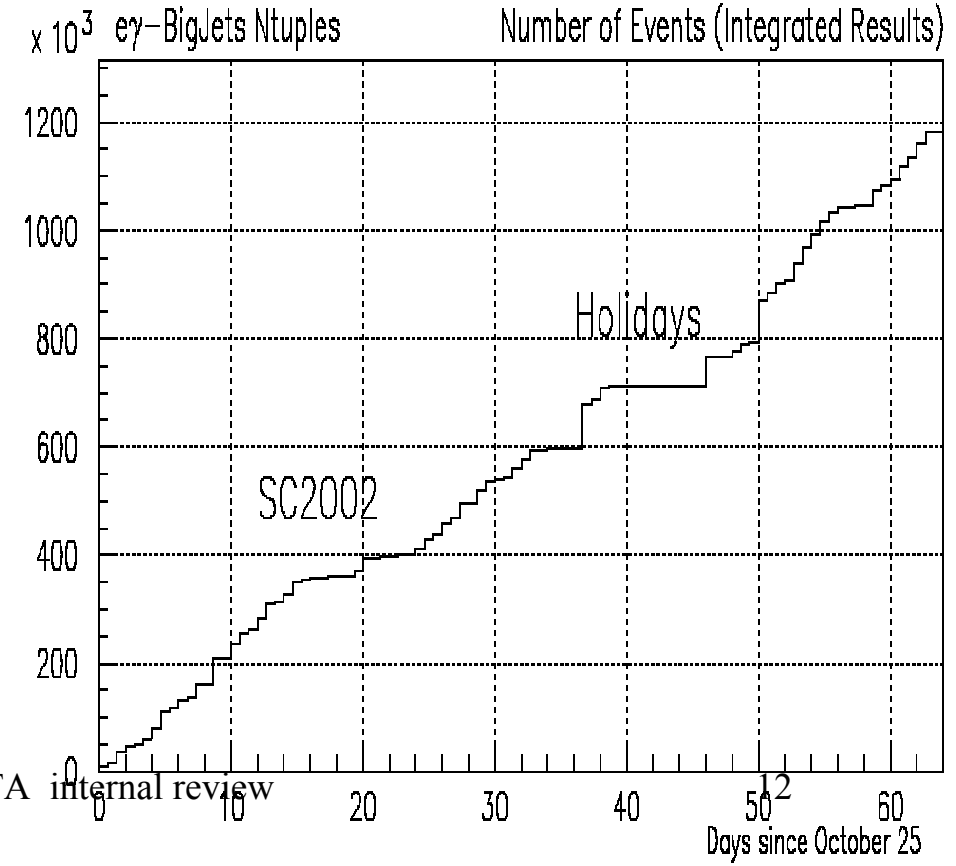


USCMS IGT Production in the US

(VDT based)

Bottom-up approach: less functionality but more stable, little manpower needed

1.2 million events in 2 months →



Grids used by CMS for production in 2003

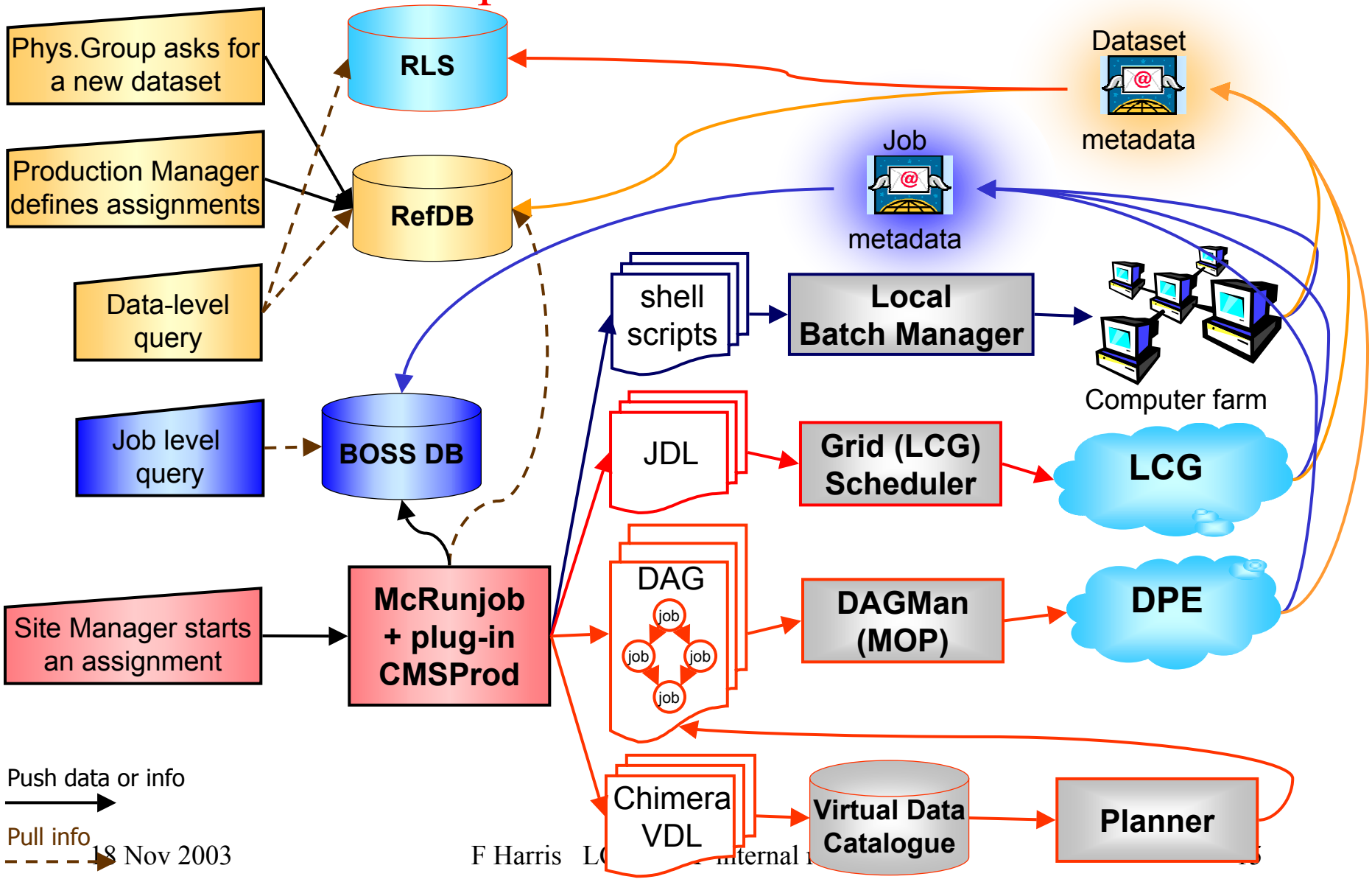
- CMS/LCG-0 is a CMS-wide testbed based on the LCG pilot distribution (LCG-0), owned by CMS (joint CMS/LCG/Datatag effort)
  - Red Hat 7.3
  - Components from VDT 1.1.6 and EDG 1.4.X
  - GLUE schemas and info providers (DataTAG)
  - VOMS
  - RLS
  - Monitoring: GridICE by DataTAG
  - R-GMA (as BOSS transport layer for specific tests)
- Currently configured as a CMS RC and producing data for PCP
- 14 sites configured

- **USCMS MOP(Monte Carlo Production)**
  - **VDT 1.1.8**
  - **MOP permits DAG description of jobs**
  - **Production grid: FNAL, Caltech, U.Florida, UCSD**
  - **Development grid: also Iowa, Rice, MIT**
- **Currently configured as a CMS RC and producing data for PCP**

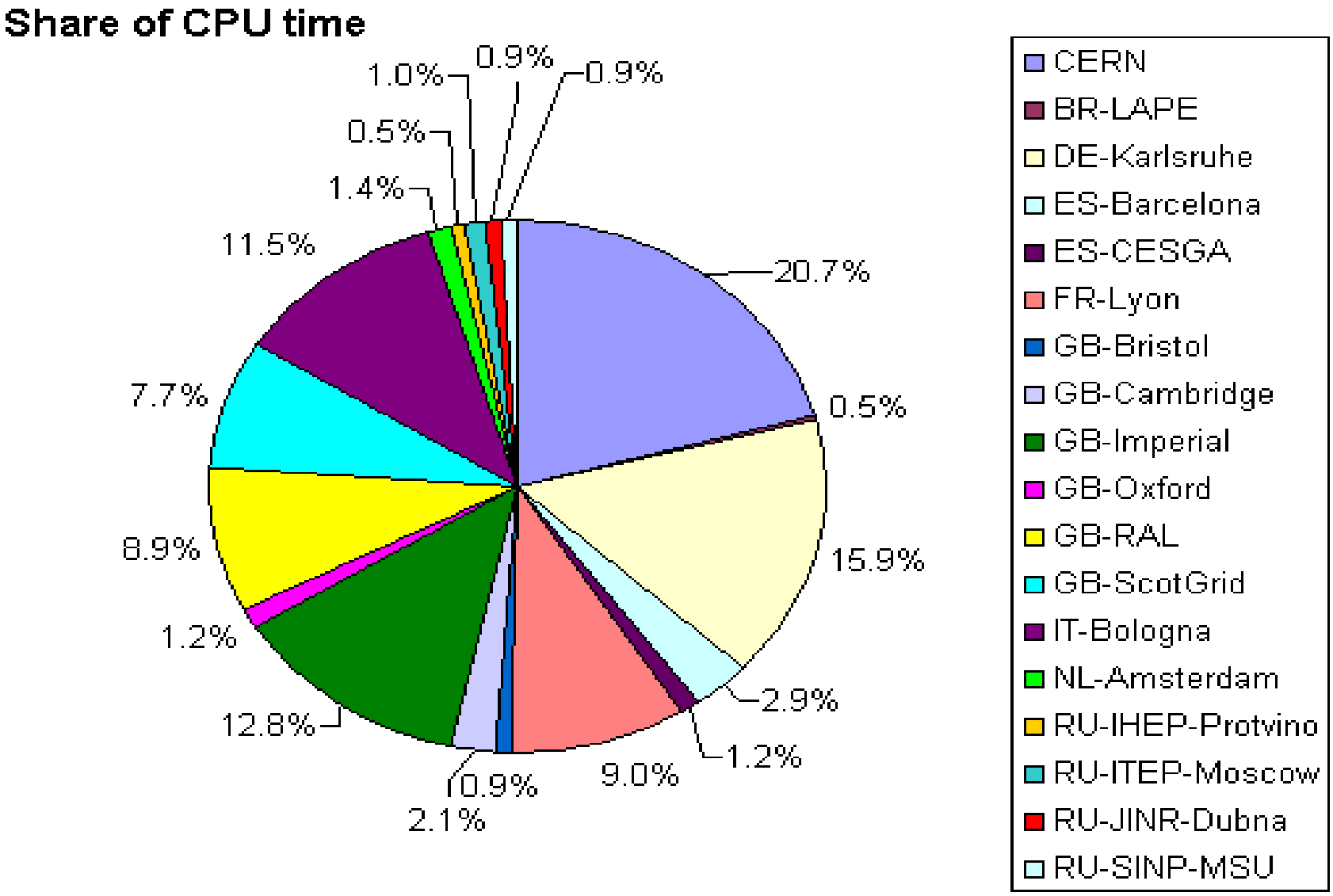
Update of CMS grid production (as of Nov 2003) and some comments on performance

- LCG-0 work
  - **500K Pythia 2000 jobs 8 hr**
  - **1.5M CMSIM 6000 jobs 10 hr.**
  - **This has been a substantial piece of work in preparation for LCG-1**
  - **Had substantial improvements in efficiency compared to first EDG stress test**
  - **Networking and site configuration were problems**
  - **Problems from time to time with unavailability of RLS**
- USCMS-MOP
  - **7.7M Pythia 30000jobs 1.5 mins**
  - **2.3M CMSIM 9000jobs 10 hr.**
  - **Recent version of MOP has Condor-G matchmaking DAG description of job to site attributes (number of CPUs,software environment). Previously job submission was manual**
- Grid production 8% of total CMS production

CMS OCTOPUS Production System – integrating all production modes



LHCb DIRAC based production in 2003



LHCb DIRAC DC'03 – Results and comments

- Distributed MC production system for LHCb
  - Production tasks definition and steering
  - Software installation on production sites
  - Job scheduling and monitoring
  - Data transfers
- Automates most of the production tasks
  - minimum participation of local production managers
- ***PULL*** rather than ***PUSH*** concept for tasks scheduling (as for ALIEN)

- A total of 47 Million events have been produced in two months
- 18 centres participate and 80% of CPU outside CERN
- 36 600 jobs have been run and each job:
  - producing between 250 and 500 events
  - using from 32 to 56 hours on 1GHz PC
- Data Challenge co-ordinated from CERN
- None of the 18 centres are LHCb dedicated
- External manpower needed to perform DC'03 – integrates to about ~1FTE

LHCb use of EDG Application service

- Standard LHCb production jobs were used for the tests (at a time when support for testbed diminishing)
  - Jobs of different statistics with 8 steps workflow.
- Jobs submitted to 4 EDG testbed Resource Brokers
  - keeping ~50 jobs per broker
- Software installed for each job

Job type (hours)	Total	Success	Success rate
Mini (0.2)	190	113	59%
Short (6)	171	102	59%
Medium (24)	1195	346	29%
Total	1556	561	36%

Total of ~300K events produced - EDG testbed already a "competitive" LHCb production site

Comments on DIRAC performance and needed developments

- Has run very reliably at 92+% efficiency over several months on 20 sites
- Lacks data management tools
- Requires outbound IP connectivity from WNs
- Lack security checking – e.g. of software in production database which is leaded into WNs

Summary 1

- All LHC experiments have well developed distributed ***batch production*** systems which have already produced substantial data (with grid technology beginning to make a major contribution in area of 'batch production', ***user analysis*** coming as a big challenge
- In past 2 years considerable experience has been gained with middleware coming from US and European grid projects, and from 'home-grown' developments in experiments. Consequently the community has a much better grasp of the problems.
- Major issues are
  - To have some real optimisation everyone needs a robust, scalable information system. We keep our fingers crossed that RGMA and/or MDS3 deliver…currently being evaluated
  - Site management, configuration and certification tools are essential. This area remains a major source of errors
  - Error detection, reporting and recovery are still very basic or non-existent (though applications have done good work e.g. GRAT,BOSS,CHIMERA…)
  - Scalability of middleware as configurations and N users grow
  - Application Software installation at sites
  - Support of a variety of mass storage devices

Summary 2

- **The PULL systems used by ALICE and LHCb have been very efficient since they do not try to do the optimisation associated with PUSH models which in turn require stable and scalable information systems, and which are very sensitive to problems of scalability**
- **The systems developed in the US (VDT based) and in Nordugrid have benefited from starting with simple prototypes which could get on the air quickly without requiring a lot of manpower – and then developed incrementally. Important tools such as GRAT, Chimera have been developed**
- **The EDG project has had goals bearing fruit on a longer term – the RGMA information system currently being commissioned, the distributed RLS system, and advanced resource brokering (e.g. DAGMAN support) – the first RB and DM products are in LCG-1 and LCG-2**
- **We must run in a multi-grid world!**
- **We should pool all expertise gained – we look to the ARDA project to harmonise future developments in time for the Data Challenges -**

- **Thanks to experiments for full cooperation in providing information**
  - **ALICE** P Buncic(CERN),P Cerello(Turin)
  - **ATLAS** K De(Univ of Texas, Arlington),O Smirnova(Lund), G Negri(Milan),GPoulard(CERN)
  - **CMS** C.Grandi(Bol),G.Graham(FNAL),D.Bradley(Wisc),A.Fanfani(Bol)
  - **LHCb** N Brook(Bristol),A Tsaregorodtsev(Marseille)