



GRID planning in LHCb

***Andrei Tsaregorodtsev,
CPPM, Marseille***

EU DataGrid Internal Project Conference,
Barcelona, 12-16 May 2003



Outline

- ◆ **The LHCb Data Challenges**
- ◆ **The LHCb MC production system - DIRAC**
- ◆ **Using of the EDG and LCG-1 platforms**
- ◆ **Future plans**
- ◆ **Conclusion**



DC 2003 (1)

-
- ◆ **Data to support Physics and Trigger TDRs for September 2003**
 - ◆ **“GEANT3 simulation” + “C++ reconstruction”**
 - ◆ **50 days running from 13 Feb 2003-04-15**
 - ◆ **17 centres in production configuration (CERN, UK, FRANCE, Italy, Germany, Holland, Russia, Brazil)**
 - ◆ **36,600 jobs ran**
 - ◆ **34,000 ran successfully 92% success**
 - **2% 50 days running from 13 Feb 2003-04-15 LHCb s/w problems**
 - **6% mixture data transfer + site + farm problems**



DC 2003 (2)

-
- ◆ **45M events fully simulated and reconstructed**
 - ~5 sec of LHC collisions
 - ◆ **~250,000 files produced**
 - 100 MB /file
 - ~20 TB data (Castor, HPSS)
 - ◆ **Total CPU used 1.4 M hours 1GHz PC (equivalent to ~1000 CPUs for 50 days)**

 - ◆ **DC2003 production continues at low level**
 - ~30K events per day;
 - Additional physics channels;
 - Special detector running conditions.



DC 2004 (1)

- ◆ **When:** Q1/Q2 2004
- ◆ **What for:** DC 2004 will be used to justify the LHCb Computing Model to be presented in the Computing TDR in Q1/2005
- ◆ **DC scale:**
 - “3 x DC3”
 - 50M B-events + 100M minimum bias events;
 - ~500K files produced;
 - ~200TB of data (~100TB to be stored in MSS);
 - >2000 concurrent jobs in >20 centers continuously running during 3-4 months



DC 2004 (1)

Goals:

- ◆ **New GEANT4 based simulation**
- ◆ **“Realistic” dataflow of recorded raw/triggered events;**
 - New raw event data model;
- ◆ **Different dataflows:**
 - Various streams for physics subgroups;
 - Data stripping, clustering;
 - Event tag collections.
- ◆ **Tests of performance of the simulation and reconstruction software;**
- ◆ **Tests of the production tools.**

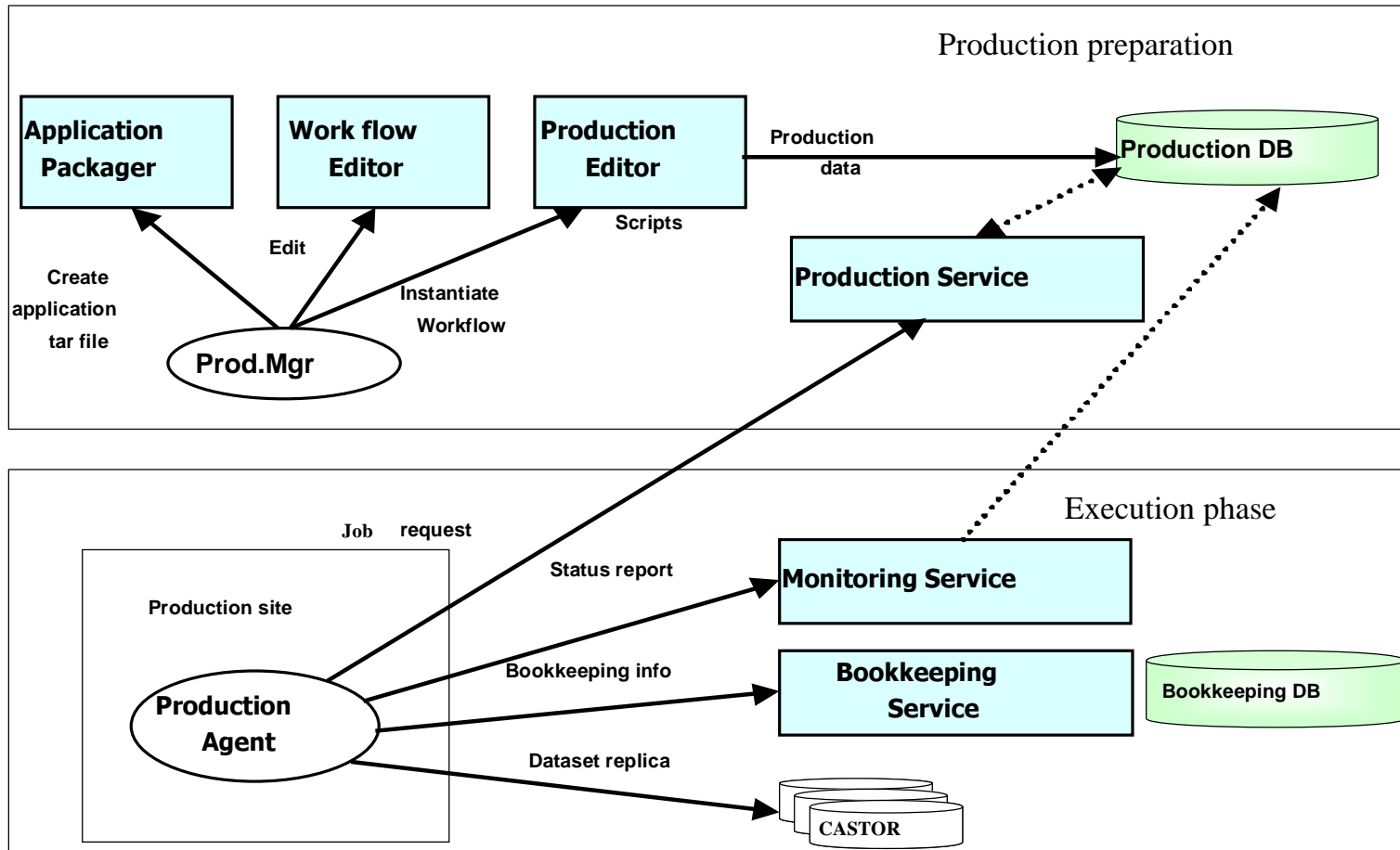


DIRAC

- ◆ **DC2003 and DC2004 will be done with DIRAC**
- ◆ **Distributed MC production system for LHCb**
 - Production tasks definition and steering;
 - Software installation on production sites;
 - Job scheduling and monitoring;
 - Data transfers and bookkeeping.
- ◆ **Automates most of the production tasks**
 - Minimum participation of local production managers
- ◆ ***PULL* rather than *PUSH* concept for jobs scheduling**



DIRAC architecture





DIRAC interface to the Grid (1)

- ◆ **DIRAC has EDG testbed as one of its production sites:**
 - Via Agent running on a Grid portal;
 - We use our own job monitoring and logging facilities;
 - Data produced on the Grid are copied to Castor and their metadata is added to the LHCb bookkeeping database to be available outside the Grid.

- ◆ **We will continue to keep EDG and later LCG-1 as a “back-end” computing resource to DIRAC**
 - Use as one whole “black box”;
 - Synchronization of the Grid and DIRAC Replica Catalog.



DIRAC interface to the Grid (2)

- ◆ **We would like to explore the possibility to use the Grid services with not necessarily using the central RB:**
 - **Using DIRAC to schedule jobs directly to CE's**
 - **Interested in using the Grid services:**
 - **DataSet Metadata Catalog (if any);**
 - **Data management (Replica Manager, RLS);**
 - **R-GMA based information system;**
 - **Security infrastructure.**



Conclusions

- ◆ **DIRAC system behaved successfully in the DC just finished and is being prepared for the future DCs;**
- ◆ **LCG-1 platform is seen as one of the “back-end” production site to DIRAC;**
- ◆ **We are also studying the possibility to use some of the Grid middleware services as components of our production system.**