



# ATLAS DC2

26 January 2004

Dario Barberis  
CERN & Genoa University



# Data Challenge 2

- DC2 operation in 2004:
  - distributed production of ( $>10^7$ ) simulated events in April-June
  - events sent to CERN in ByteStream (raw data) format to Tier-0
  - (possibly) "prompt" alignment/calibration and (certainly) reconstruction processes run on prototype Tier-0 in a short period of time ( $\sim 10$  days, "10% data flow test")
  - reconstruction results distributed to Tier-1s and analysed on Grid
- Main "new" software to be used (wrt DC1 in 2002/2003):
  - Geant4-based simulation, pile-up and digitization in Athena
  - complete "new" EDM and Detector Description interfaced to simulation and reconstruction
  - POOL persistency
  - LCG-1/2 Grid infrastructure
  - Distributed Production and Analysis environment



# DC2: Scenario & Time scale

September 03: **Release7**

**Put in place, understand & validate:**

Geant4; POOL; LCG applications

Event Data Model

Digitization; pile-up; byte-stream

**Conversion of DC1 data to POOL**; large scale persistency tests and reconstruction

**Testing and validation**

Run test-production

February 27<sup>th</sup> 04: **Release 8 (production)**

**Start final validation**

April 1<sup>st</sup> 04:

**Start simulation; Pile-up & digitization**

**Event mixing**

**Transfer data to CERN**

June 1<sup>st</sup> 04: “DC2”

**Intensive Reconstruction on “Tier0”**

**Distribution of ESD & AOD**

**Calibration; alignment**

July 15th

**Start Physics analysis**

**Reprocessing**



# Phases of DC2 operation

- Consider DC2 as a three-part operation:
  - part I: production of simulated data (April-June 2004)
    - needs Geant4, digitization and pile-up in Athena, POOL persistency
    - "minimal" reconstruction just to validate simulation suite
    - will run on any computing facilities we can get access to around the world
  - part II: test of Tier-0 operation (June-July 2004)
    - needs full reconstruction software following RTF report design, definition of AODs and TAGs
    - (calibration/alignment and) reconstruction will run on Tier-0 prototype as if data were coming from the online system (at 10% of the rate)
    - output (ESD+AOD) will be distributed to Tier-1s in real time for analysis
  - in parallel: run distributed reconstruction on simulated data
    - this is useful for the Physics community as MC truth info is kept
  - part III: test of distributed analysis on the Grid (July-Sept. 2004)
    - access to event and non-event data from anywhere in the world both in organized and chaotic ways



# DC2 resources

Process	No. of events	Time	CPU power	Data volume	At CERN	Off site	
		months	kSI2k	TB	TB	TB	
Simulation	$10^7$	2	600	25	5	20	Phase I (Apr-Jun)
Pile-up & Digitization	$10^7$	2	400	75	15	60	
ByteStream Production	$10^7$	2	(small)	20	20	16	
Total Phase I	$10^7$	2	1000	120	40	96	
Reconstr. Tier-0	$10^7$	0.5	600	5	5	10	Phase II (Jun-Jul)
Reconstr. Tier-1	$10^7$	2	150	5	0	5	
Total	$10^7$			130	45	111	

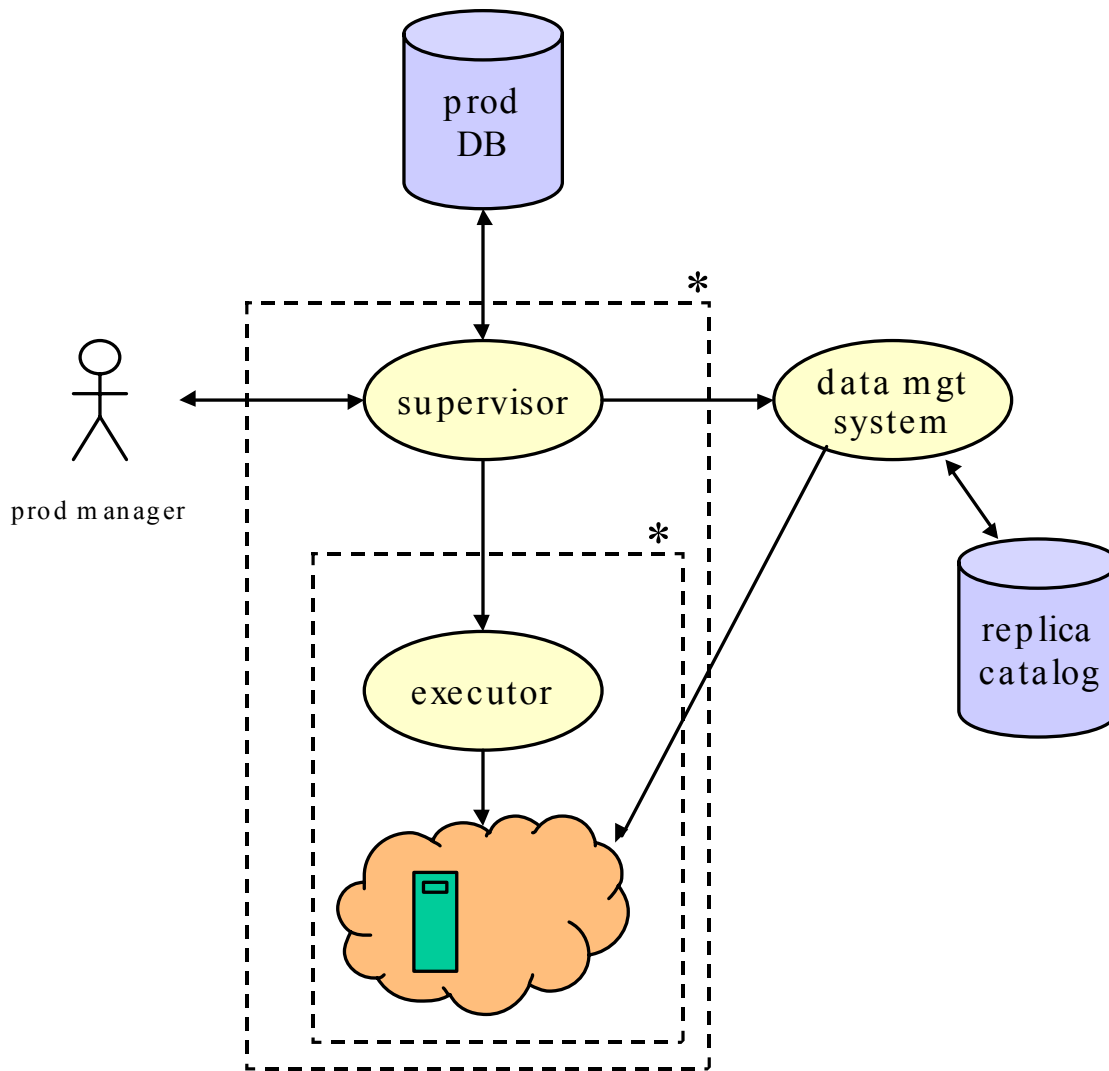


## DC2 resources - notes

- CPU needs are now based on Geant4 processing times
- We assume 20% of simulation is done at CERN
- All data in ByteStream format are copied to CERN for the Tier-0 test (Phase II)
- Event sizes (except ByteStream format) are based on DC1/Zebra format: events in POOL may be larger (will know by end Jan.)
- Reconstruction output is exported in 2 copies from CERN Tier-0
- Output of parallel reconstruction on Tier-1s, including links to MC truth, remains local and is accessed for analysis through the Grid(s)



# New Production System (1)

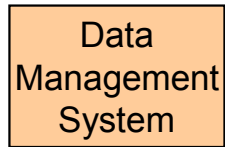


- DC1 production in 2002/2003 was done mostly with traditional tools (scripts)
  - Manpower intensive!
- Main features of new system:
  - Common production database for all of ATLAS
  - Common ATLAS supervisor run by all facilities/managers
  - Common data management system
  - Executors developed by middleware experts (LCG, NorduGrid, Chimera teams)
  - Final verification of data done by supervisor

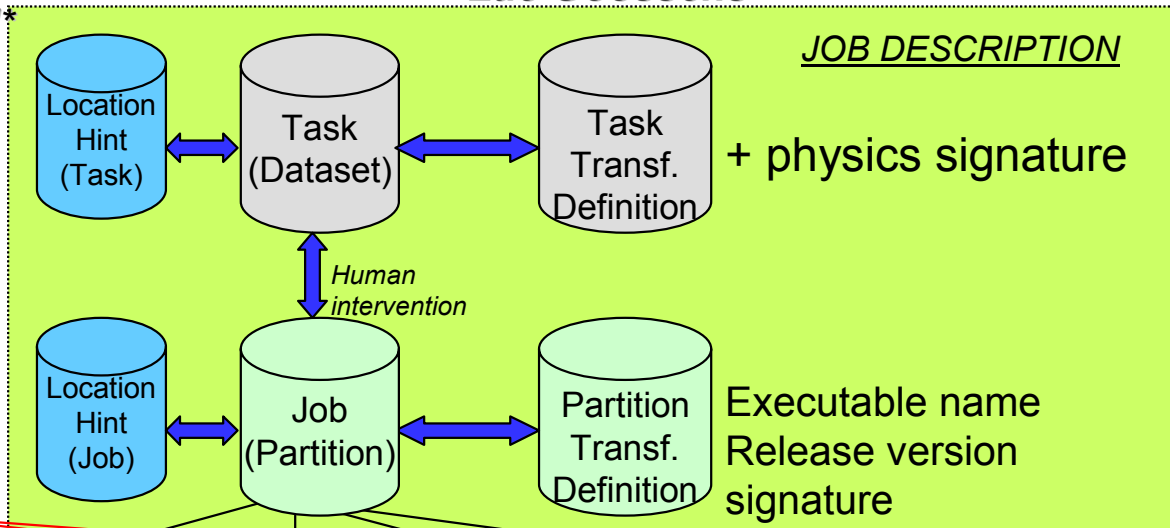
# New Production System (2)

Task = [job]\*

Dataset = [partition]\*



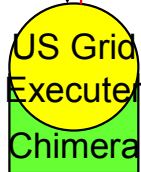
Luc Goossens



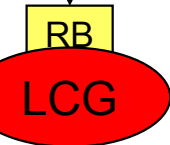
Kaushik De



Rob Gardner



Alessandro De Salvo



Oxana Smirnova



Luc Goossens

