ATLAS Data Challenge 2 (April-July 2004)

10 November 2003

Dario Barberis

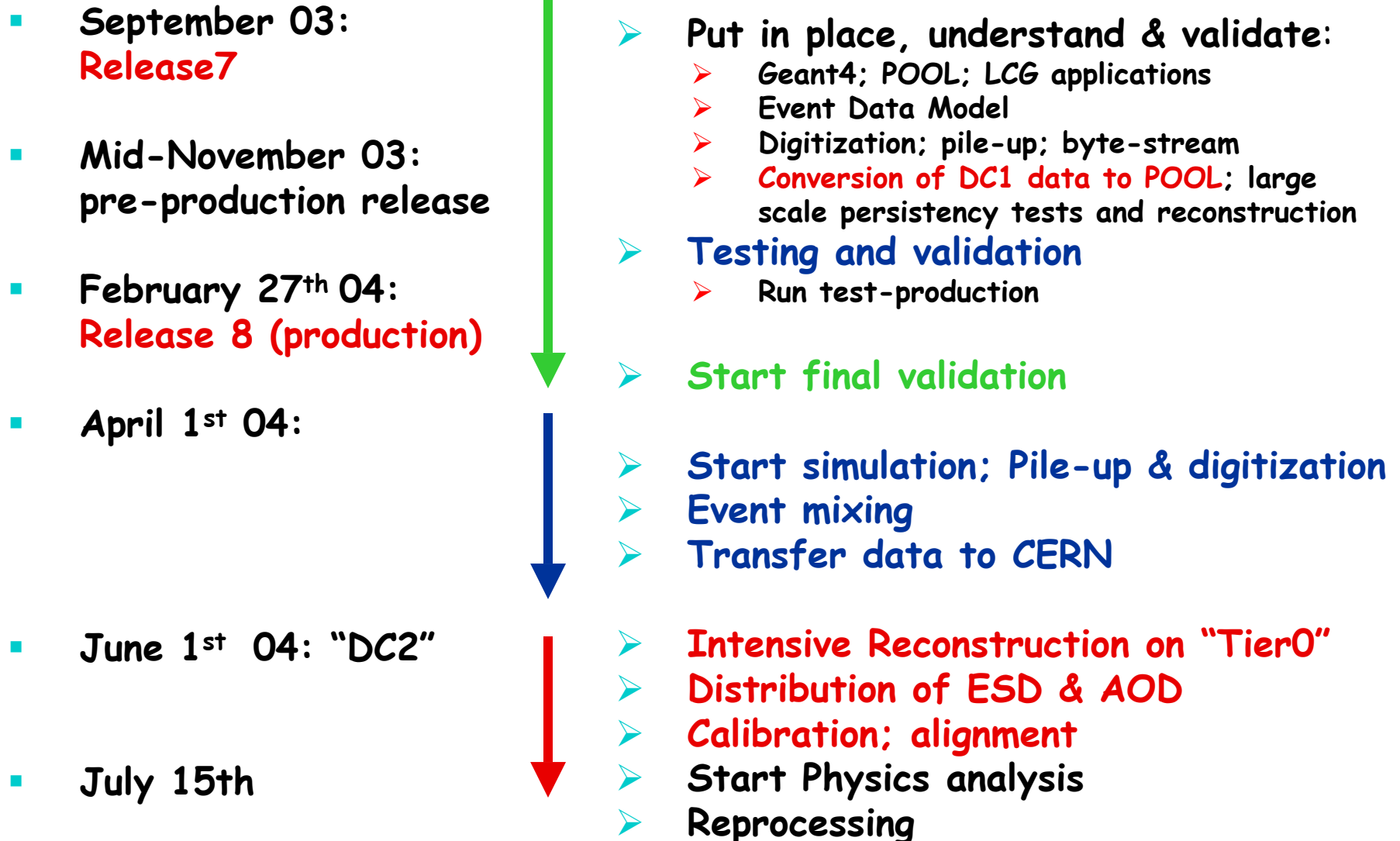### CERN & Genoa University/INFN

Data Challenges

- DC1 completed earlier this year
  - some productions still going on using the DC1 software and organization
    - they are now called "continuous productions"
    - we foresee a semi-automatic system for these medium-sized productions

- 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 (~10 days, "10% test")
  - reconstruction results distributed to Tier-1s and analysed on Grid
  - needs delivery of well-tested software by March for simulation and May for reconstruction
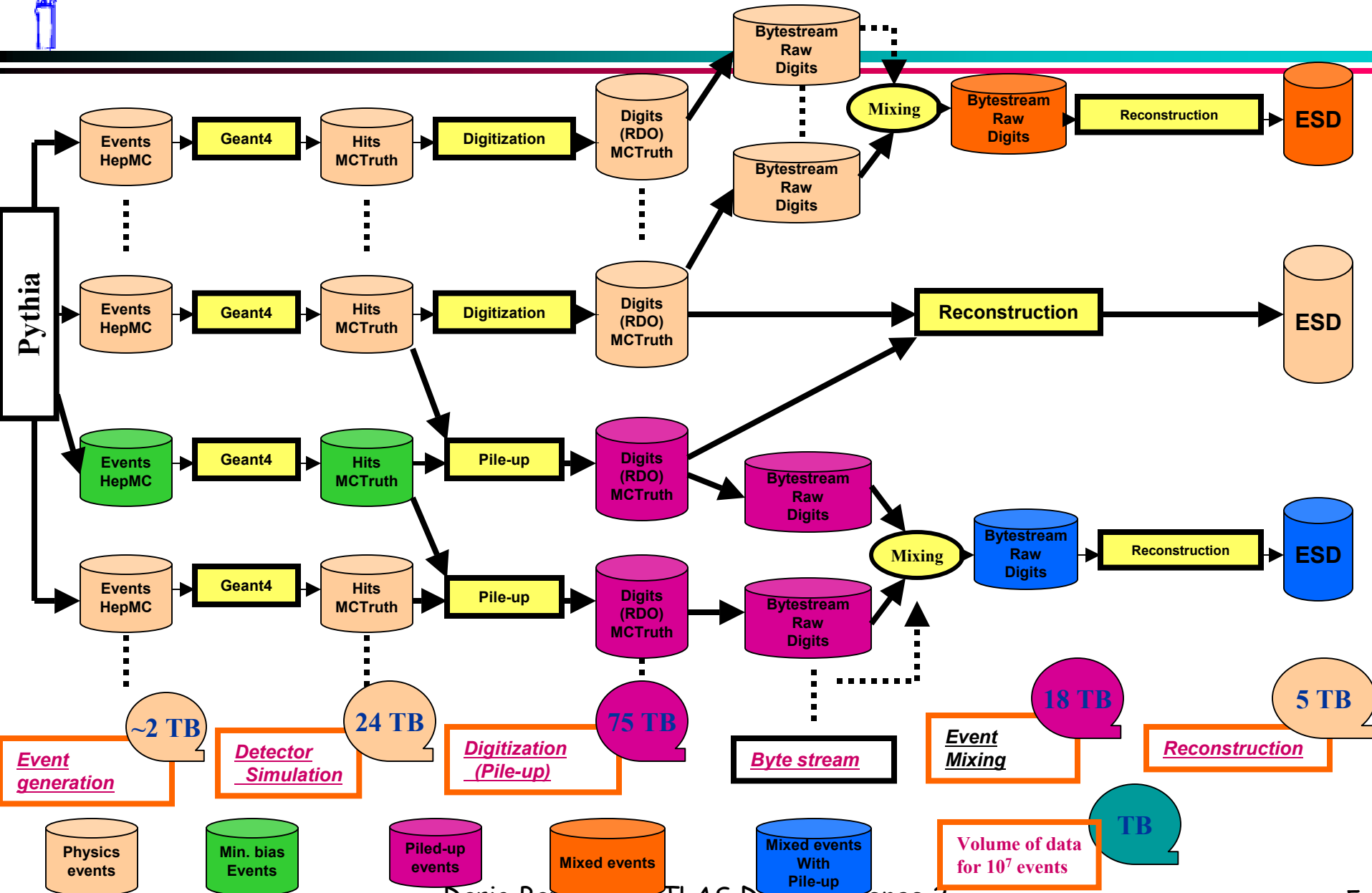
DC2: April – July 2004

- At this stage the goal includes:
  - ➤ Full use of Geant4; POOL; LCG applications
  - ➤ Pile-up and digitization in Athena
  - ➤ Deployment of the complete Event Data Model and the Detector Description
  - ➤ Simulation of full ATLAS and 2004 combined Testbeam
  - ➤ Test the calibration and alignment procedures
  - ➤ **Use widely the GRID middleware and tools**
  - ➤ **Large scale physics analysis**
  - ➤ **Computing model studies (document end 2004)**
  - ➤ **Run as much as possible of the production on LCG-1/2**

DC2: Scenario & Time scale

- **September 03:**
  Release7

- **Mid-November 03:**
  pre-production release

- **February 27th 04:**
  Release 8 (production)

- **April 1st 04:**

- **June 1st 04: "DC2"**

- **July 15th**

➢ **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

➢ **Start final validation**

➢ **Start simulation; Pile-up & digitization**
➢ **Event mixing**
➢ **Transfer data to CERN**

➢ **Intensive Reconstruction on "Tier0"**
➢ **Distribution of ESD & AOD**
➢ **Calibration; alignment**
➢ **Start Physics analysis**
➢ **Reprocessing**

Task Flow for DC2 data

DC2 and Grid tools

- **Much work already done:**
  - o Magda (BNL); AMI (Grenoble) used already in DC1
  - o Other tools in different stages of development and test:
    - ➢ GANGA (ATLAS-LHCb UK main effort)
    - ➢ Chimera (US) [exploit Virtual Data ideas (DC1)]; DIAL (BNL)
    - ➢ AtCom used to generate the jobs (batch and grid flavors)
    - ➢ GRAT (US); Grappa (US)
  - o A coherent view of tools use and integration is emerging but need more work and thinking

- **LCG-1/2**
  - o We intend to use and contribute to validate LCG-1/2 components when they become available
  - o ATLAS-EDG task force has become ATLAS-LCG testing task force

DC2 resources (based on Geant3 numbers)

Process	No. of events	Time	CPU power	No. of machines 500SI2k	Data volume	At CERN	Off site
		months	kSI2k		TB	TB	TB
Simulation	10^7	2	260		24	8	16
Pile-up (*) & Digitization	10^7	2	175		(75)	(25)	(50)
Byte-stream	10^7	2			18	18	12
Total	10^7	2	435	870	42 (+57)	26 (+57)	28 (+38)
Reconstruction	10^7	0.5	600	1200	5	5	5

* To be kept if no "0" suppression in calorimeters

ATLAS DC2 and POOL/RLS

- ATLAS will use POOL in Data Challenge 2

- The detailed interface of POOL in DC2 to grid services has yet to be worked out precisely, depending on technical issues and the timescale of the development of data catalogs.

- Ultimately there must be a commonly accessible data catalog for all of DC2 data for all of ATLAS.

- Flavours of RLS may be one choice, MAGDA may be a fall-back, or some other work-around can be developed, but there must be ultimately one populated catalog.

- If the Grid configurations vary from region to region, we need to put in a lot more work to define the common interfaces and services carefully and understand the definition of interoperability: this is NOT the experiments' task!

- We will work to minimize grid middleware divergence: we should have different middleware at different sites ONLY if it is REALLY necessary!

DC2: Summary

- Schedule still as it was before summer
  - o Will be re-evaluated in December (S/W week)

- Still uncertainties on resources
  - o CPU
    - ➢ timing for Geant4 not yet known (likely factor 1.5-2)
    - ➢ Impact of new EDM not evaluated
  - o Storage
    - ➢ Impact from POOL and EDM not known
    - ➢ "zero suppression" still a concern

- Resources needed for "combined test-beam" activities "Monte Carlo" and "real data" not taken in to account

- Trigger/DAQ "tests" not taken into account