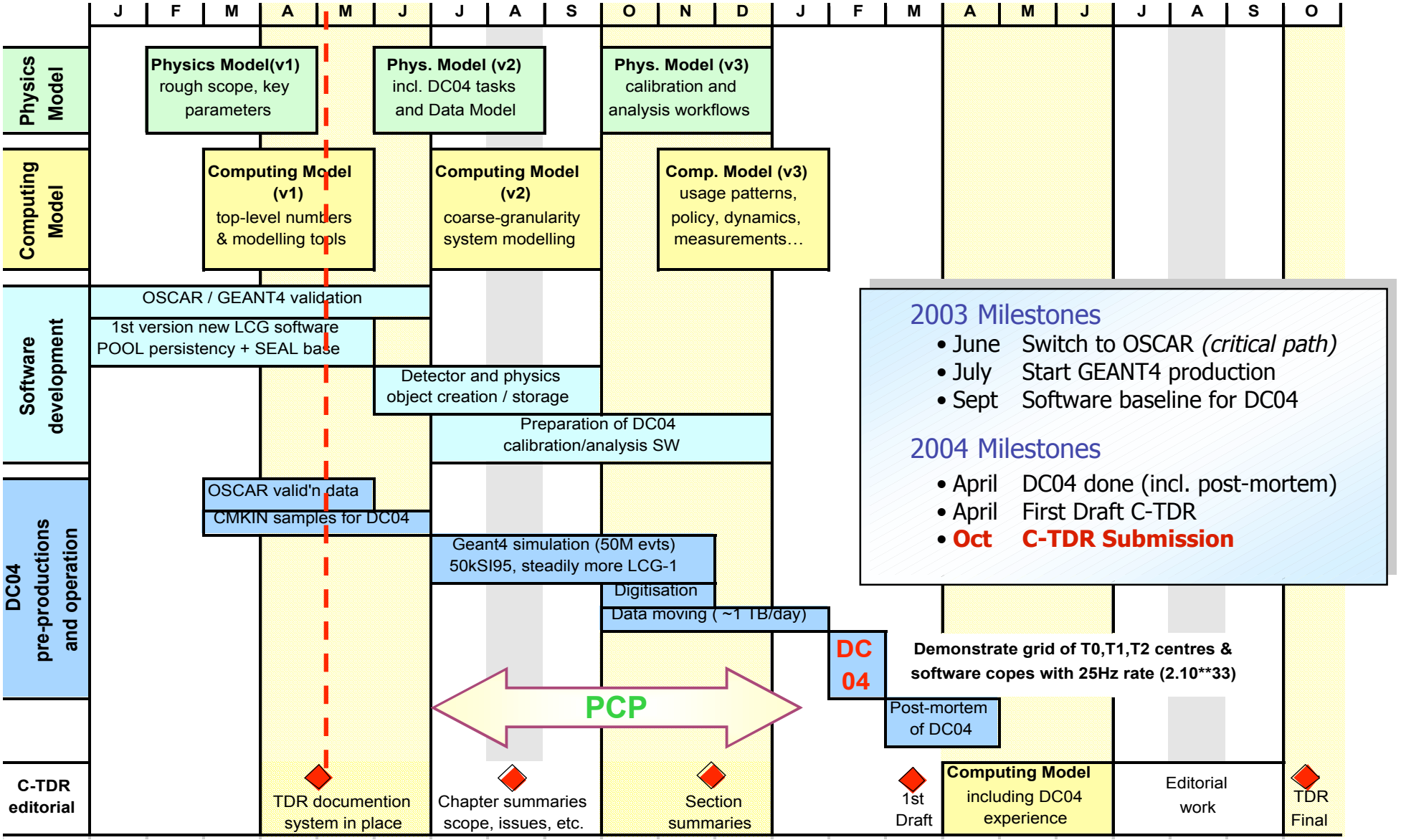
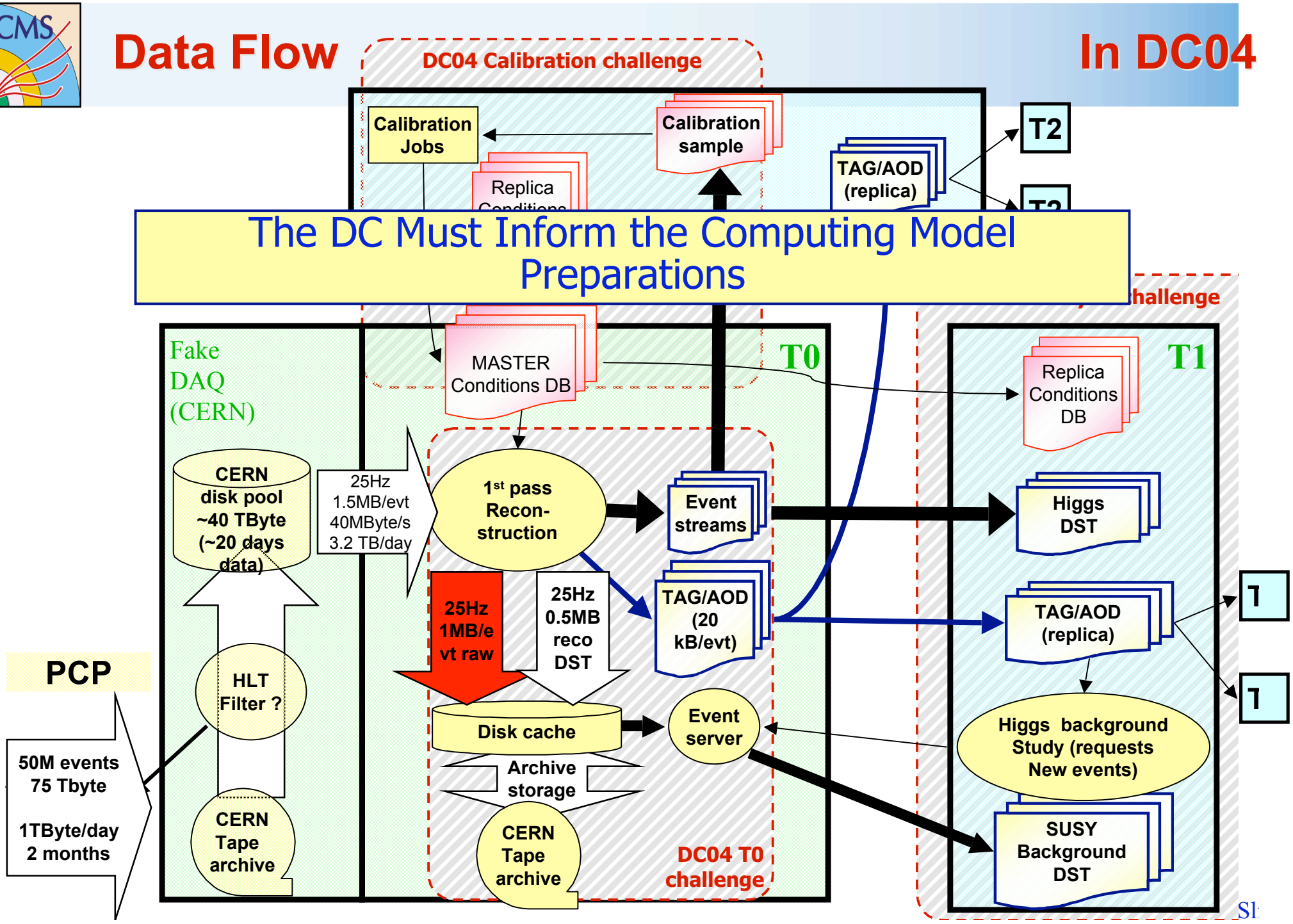
Update on CMS DC04

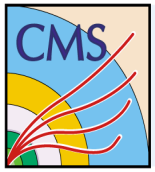
David Stickland
CMS Core Software and Computing

Schedule: PCP, DC04, C-TDR …



Data Flow In DC04





Strawman Model I. Guiding Principles

- Data access is a much harder problem than CPU access.
  - Avoid bottlenecks, distribute data widely, quickly.
- Require a balance between the common a-priori goals of the experiment and the individual goals of its collaborators
  - The experiment must be able to partition resources according to policy
- No dead-time can be introduced to the data acquisition by the offline system
  - All potential points of blockage must have "relief valves" in place.
  - The Tier-0 must keep up in real time with the DAQ. Latencies must be no more than of order 6-8 hours.
- Tier-1 centers are largely resources of the experiment as a whole
  - Data intensive tasks need to run at Tier-1 centers
- Tier-2 centers are focused more at geographic and/or physics groupings
  - It must be possible to replicate modest sized data sets to the Tier-2's in a timely way.

Building a Strawman Model

1. Raw data is "streamed" from the online system at 100 MB/s.
    1. The streams may not be the final ones
2. Raw data is sent to MSS at the Tier-0
    1. A second copy of the raw data is sent offsite
3. First pass reco of (some of) the raw data at the Tier-0
    1. Some first pass reconstruction may be carried out away from CERN
4. The Tier-0/1 first pass keeps up with the DAQ rate
    1. Tier-0 is available outside the LHC running period for rerunning etc.
5. The DST may be further/differently streamed wrt DAQ Streams.
    1. Some (10%?) event duplication is allowed
6. Calibration "DST's" are sent to the Tier-1/2 responsible for the processing
7. The full DST is kept at the Ti□er-0 and at each Tier-1
8. The full TAG (selection data) is stored at the Tier-0-1 and-2
9. Scheduled Analysis passes on DST/TAG data are run at the Tier-1's
10. Tier-2 centers are the point of access for most user analysis/ physics preparation.
11. …

DC04 Status Today

- Pre-Challenge Phase (MC Gen, Simu, Digitization)
  - Generation/Simulation steps going very well

	Requested	Completed
CMSIM G3	52M	48M
OSCAR G4	16M	0 (But started now)
Not Yet assigned	7M (Probably OSCAR)	

  - N.B. Of this
    - 1.5M with LCG0 (~40kSI2Kmonths)
    - 2.3M with USCMS/MOP (~50kSI2k months)
- Digitization Step getting ready
  - Complicated. May only have about ~30M Digitized by Feb 1
- Final schedule for DC04 could slip by ~1 month (March/April)

DC04 Scales at T0,1,2

- Tier-0
  - Reconstruction and DST production at CERN
    - 75TB Input Data (25TB Input buffer?)
    - 180kSI2k.month =400 CPU @24 hour operation (@500SI2k/CPU)
    - 25TB Output data
    - 1-2TB/Day Data Distribution from CERN to sum of T1 centers
- Tier-1
  - Assume all (except CERN) "CMS" Tier-1's participate
    - CNAF, FNAL, Lyon, Karlsruhe, RAL
  - Share the T0 output DST between them (~5-10TB each?)
    - 200GB/day transfer from CERN (per T1)
    - (Possibly stream ~1TB Raw-Data to Lyon to host full EGamma dataset?)
  - Perform scheduled analysis group "production".
    - ~100kSI2k.month total = ~50 CPU per T1 (24 hrs/30 days)
- Tier-2
  - Assume about 5-8 T2:
    - 2 US, 1UK, 2-3 Italian, 1 Spanish, + ?
    - Store some of TAG data at each T2 (500GB? 1TB?)
    - Estimate 20CPU at each center for 1 month

DC04 Tasks At T1 and T2 (under discussion)

- "Most" T1s participate to Analysis Group Scheduled Productions
- One T1 and Two T2 do pseudo-calibration
  - Analyzing calibration DST's, exercising round-trip for calibration back to T0
- Two T1 and Two T2 exercise LCG RB/RLS tools to prepare and submit jobs, accumulate results running over DST at one or both T1 centers.
- One T1 and Two T2 centers exercise LCG tools (GFAL/POOL/RLS for job preparation and execution. (runtime file access from WAN/MSS)
- 1-2 T2 centers exercise Tag processing, defining new collections constructing deep-copies at T1 and exporting new collections ba to T2
- ....

POOL/RLS/GFAL/SRB etc.

- CMS has invested heavily in POOL and is very pleased with the progress
  - POOL is a completely vital part of the CMS data challenge
- CMS has invested heavily in the POOL file-catalog tools to work with the EDG/RLS through POOL
  - Actual RLS hidden from CMS by POOL-fc
  - No objection to any other RLS backend to POOL, but we don't want to se it except via POOL.
- SRB has been very useful for file transfers, MSS transparency etc in the PCP
  - Probable we start with RLS and SRB interoperating (according to their functions)
  - Try federated MCAT (at least two sites) to improve uptime
  - Need to understand complimentarily and/or transition with GFAL
    - In which CMS has also been testing and is pleased with the progress
    - (N.B. We want the posix access, kernel module on each WN)