ATLAS plans for 2006: Computing System Commissioning and Service Challenge 4

Dario Barberis

CERN & Genoa University

Computing System Commissioning Goals

- We have defined the high-level goals of the Computing System Commissioning operation during 2006
  - Formerly called "DC3"
  - More a running-in of continuous operation than a stand-alone challenge
- Main aim of Computing System Commissioning will be to test the software and computing infrastructure that we will need at the beginning of 2007:
  - Calibration and alignment procedures and conditions DB
  - Full trigger chain
  - Event reconstruction and data distribution
  - Distributed access to the data for analysis
- At the end (autumn-winter 2006) we will have a working and operational system, ready to take data with cosmic rays at increasing rates

Computing System Commissioning Tests

- Sub-system tests with well-defined goals, preconditions, clients and quantifiable acceptance tests
  - Full Software Chain
  - Tier-0 Scaling
  - Calibration & Alignment
  - Trigger Chain & Monitoring
  - Distributed Data Management
  - Distributed Production (Simulation & Re-processing)
  - Physics Analysis
  - Integrated TDAQ/Offline (complete chain)
- Each sub-system is decomposed into components
  - E.g. Generators, Reconstruction (ESD creation)
- Goal is to minimize coupling between sub-systems and components and to perform focused and quantifiable tests
- Detailed planning being discussed now
  - also in relation with WLCG Service Challenge 4 schedule (see later slides)

"Realism" underway

- Updating dead material
  - Cables, services, barrel/end-cap cracks, etc.
- Define reference coordinate systems
  - GLOB=installation survey, SOL(t), BEAM(t)
- Realistic B-field map taking into account non-symmetric coil placements
  - B-field map size issues…
- Displace detector (macro)-pieces to describe their actual positions
  - E.g. EM barrel axis 2mm below beam line and solenoid axis
  - Break symmetries and degeneracy in detector descript and simulation
- Include detector "egg-shapes" if relevant
  - E.g. Tilecal elliptical shape if it is has an impact on B-field…
- Mis-align detector modules/chambers inside macro pieces
- Include chamber deformations, sagging of wires and calo plates, etc.
  - Probably at digitization/reconstruction level
- Dedicated workshops held to give coherence to these efforts

Calibration/alignment "challenge"

- Part of the overall computing system commissioning activity
  - Demonstrate the calibration 'closed loop' (iterate and improve reconstruction)
    - Athena support for conditions data reading / writing / iteration
    - Reconstruction using conditions database for all time-varying data
  - Exercise the conditions database access and distribution infrastructure
    - With COOL conditions database, realistic data volumes and routine use in reconstruction
    - In a distributed environment, with true distributed conditions DB infrastructure
  - Encourage development of subdetector calibration algorithms
    - Going on anyway, but provide collaboration-wide visibility to this work
    - Calibration done in a realistic computing environment
- Initially focussed on 'steady-state' calibration
  - Largely assuming required samples are available and can be selected
  - But also want to look at initial 2007/2008 running at low luminosity
    - Selecting events from the 'initial realistic data sample'
    - Issues of streaming – using calibration and physics data

Calib/Align – prerequisites for success

- Simulation
  - Ability to simulate a realistic, misaligned, miscalibrated detector
    - Geometry description and use of conditions DB in distributed simulation and digitisation
  - Static replication of conditions database to support this – parameters in advance
- Reconstruction
  - Use of calibration data in reconstruction; ability to handle time-varying calibration
  - Initially, static replication of conditions database – parameters in advance
  - Later, dynamic replication (rapidly propagate new constants) to support closed loop and 'limited time' exercises
- Calibration algorithms
  - Algorithms in Athena, running from standard ATLAS data (ESD, raw data?)
    - Ability to deal with substantial fractions of the whole subdetector
  - Currently focussed on subdetector studies, would be nice to exercise some 'global calibration' – E/p, spatial matching etc
- Management
  - Organisation and bookkeeping (run number ranges, production system,…)
    - How do we ensure all the conditions data for simulation is available with right IoVs?
    - What about defaults for 'private' simulations ?

Cal/Align - calibration parameters

- Subdetector parameters to be exercised (red already done for CTB)

SCT/Pixel	Alignment, dead/noisy channels, module distortions, pixel calib/thresholds
TRT	Module align., wire position, t_0, R-t, dead channels, resolution, efficiency
LAr	Electronics calibration, HV, cluster level corrections, dead material, misalignment
HEC	(Focus on energy/eta parameterisation)
TileCal	CIS calibration, cesium calibration, optimal filter coefficients
MDT	t_0, R-t, alignment corrections, temperature/field/sag/space charge corr^n
RPC	Pressure/temp, thresholds, HV/LV, currents, dead strip/efficiencies map, trig coinc.
CSC	ADC to strip charge, chamber alignment
TGC	Timing, delays, chamber alignment

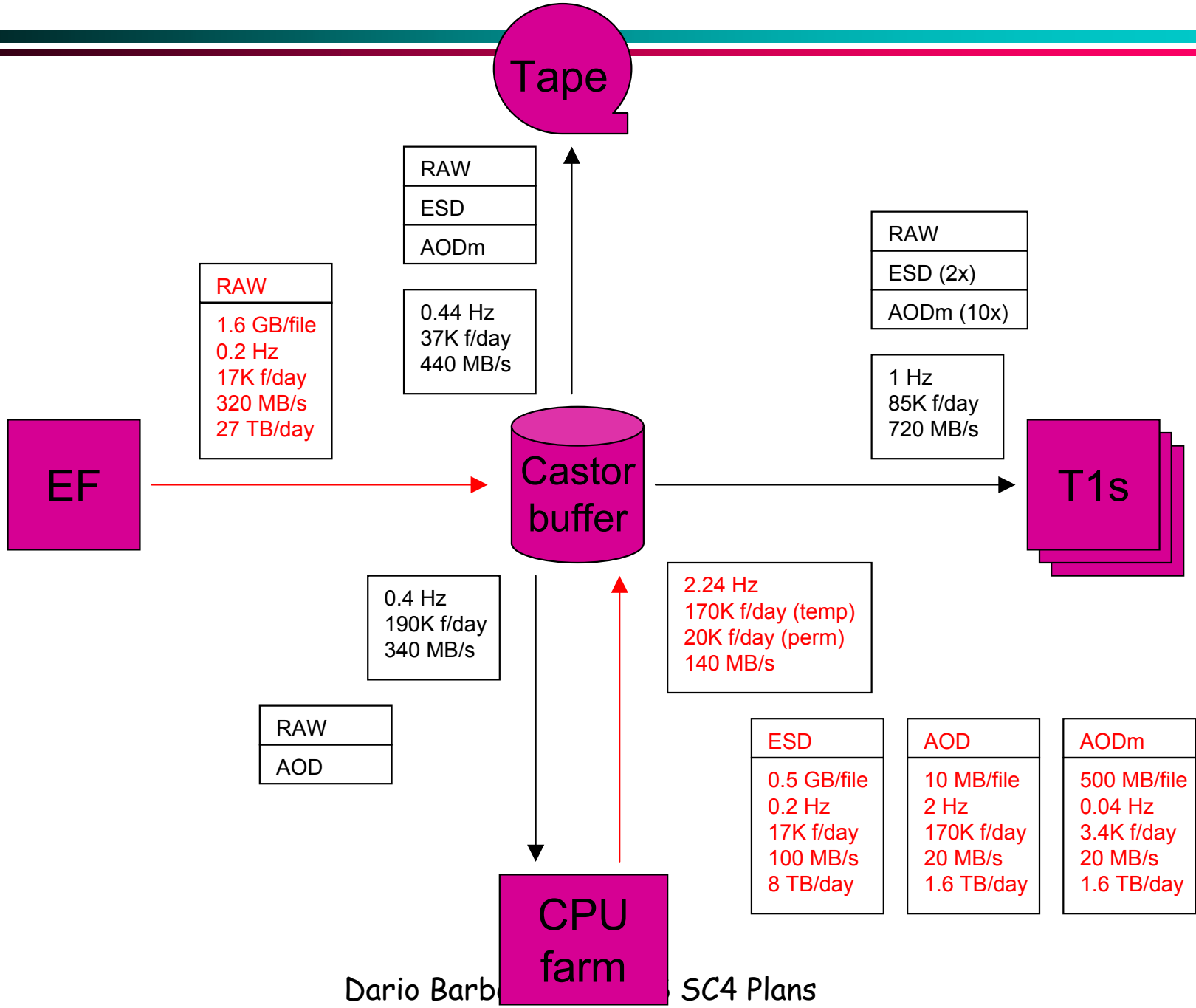
Software release plans in 2006

- End March: release 12
  - Full geometry upgrade: complete implementation of the "as-built" geometry
  - Conditions DB infrastructure in place and significant usage of COOL by subdetectors
    - Includes ROOT5 and CORAL
  - Trigger EDM in place
  - Implementation of MC Truth Task Force recommendations
  - Implementation of Event Tag working group recommendations
- End July: release 13
  - Calibration/alignment loop
  - Full schema evolution for event data
  - Support for cosmic runs in autumn 2006
- December: release 14
  - Performance optimization: CPU time, memory, physics performance
    - Including shower parameterization in the calorimeters
  - Further geometry upgrade including detector survey data
  - Full schema evolution for conditions data

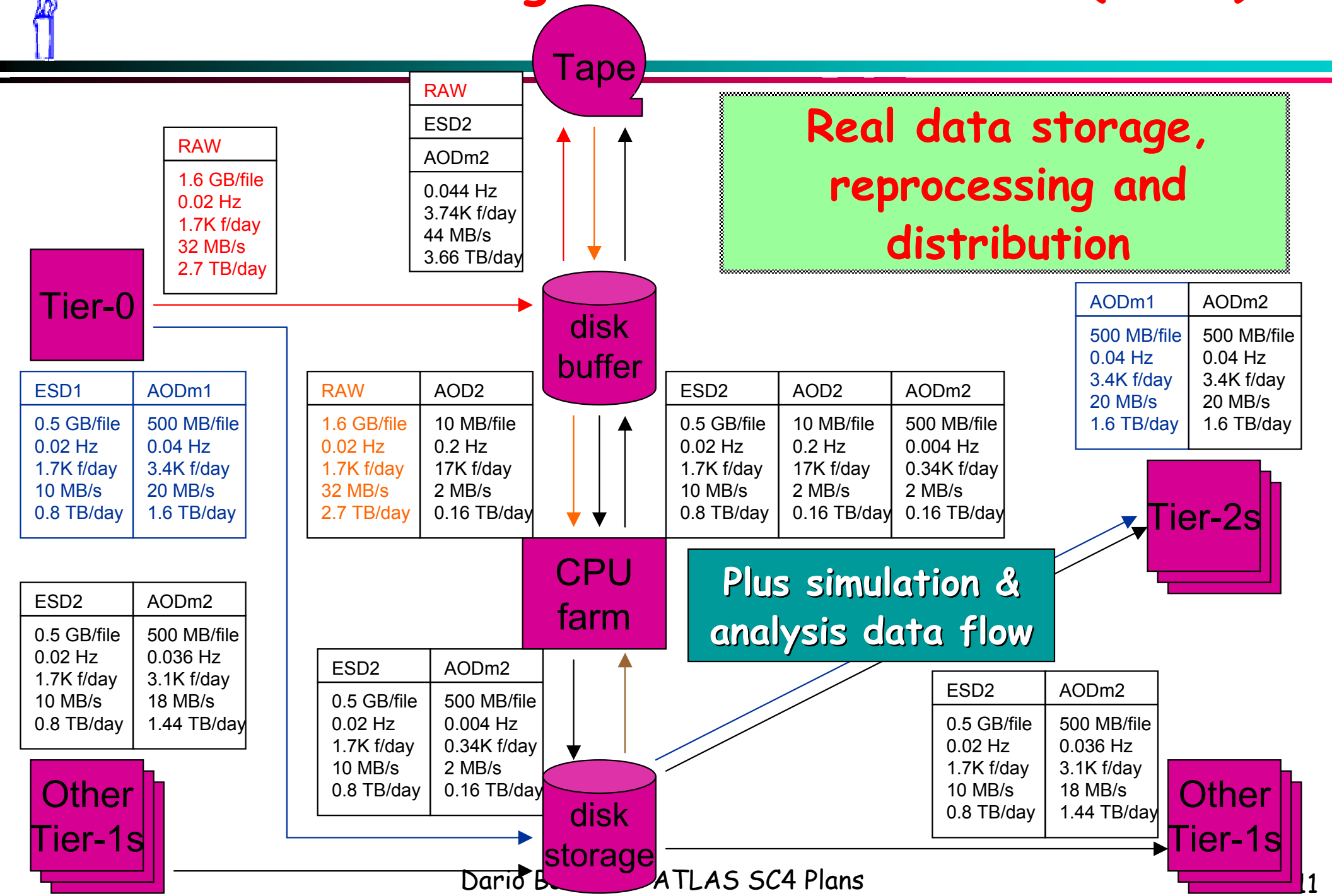
ATLAS Computing Model

- Tier-0:
  - Copy RAW data to Castor tape for archival
  - Copy RAW data to Tier-1s for storage and reprocessing
  - Run first-pass calibration/alignment (within 24 hrs)
  - Run first-pass reconstruction (within 48 hrs)
  - Distribute reconstruction output (ESDs, AODs & TAGS) to Tier-1s
- Tier-1s:
  - Store and take care of a fraction of RAW data
  - Run "slow" calibration/alignment procedures
  - Rerun reconstruction with better calib/align and/or algorithms
  - Distribute reconstruction output to Tier-2s
  - Keep current versions of ESDs and AODs on disk for analysis
- Tier-2s:
  - Run simulation
  - Keep current versions of AODs on disk for analysis

ATLAS Tier-0 Data Flow



ATLAS "average" Tier-1 Data Flow (2008)



ATLAS SC4 Tests

- Complete Tier-0 test
  - Internal data transfer from "Event Filter" farm to Castor disk pool, Castor tape, CPU farm
  - Calibration loop and handling of conditions data
    - Including distribution of conditions data to Tier-1s (and Tier-2s)
  - Transfer of RAW, ESD, AOD and TAG data to Tier-1s
  - Transfer of AOD and TAG data to Tier-2s
  - Data and dataset registration in DB (add meta-data information to meta-data DB)
- Distributed production
  - Full simulation chain run at Tier-2s (and Tier-1s)
    - Data distribution to Tier-1s, other Tier-2s and CAF
  - Reprocessing raw data at Tier-1s
    - Data distribution to other Tier-1s, Tier-2s and CAF
- Distributed analysis
  - "Random" job submission accessing data at Tier-1s (some) and Tier-2s (mostly)
  - Tests of performance of job submission, distribution and output retrieval

ATLAS SC4 Plans (1)

- Tier-0 data flow tests:
  - Phase 0: 3-4 weeks in March-April for internal Tier-0 tests
    - Explore limitations of current setup
    - Run real algorithmic code
    - Establish infrastructure for calib/align loop and conditions DB access
    - Study models for event streaming and file merging
    - Get input from SFO simulator placed at Point 1 (ATLAS pit)
    - Implement system monitoring infrastructure
  - Phase 1: last 3 weeks of June with data distribution to Tier-1s
    - Run integrated data flow tests using the SC4 infrastructure for data distribution
    - Send AODs to (at least) a few Tier-2s
    - Automatic operation for O(1 week)
    - First version of shifter's interface tools
    - Treatment of error conditions
  - Phase 2: 3-4 weeks in September-October
    - Extend data distribution to all (most) Tier-2s
    - Use 3D tools to distribute calibration data
- The ATLAS TDAQ Large Scale Test in October-November prevents further Tier-0 tests in 2006…
  - … but is not incompatible with other distributed operations

ATLAS SC4 Plans (2)

- ATLAS CSC includes continuous distributed simulation productions:
  - We will continue running distributed simulation productions all the time
    - Using all Grid computing resources we have available for ATLAS
    - The aim is to produce ~2M fully simulated (and reconstructed) events/week from April onwards, both for physics users and to build the datasets for later tests
      - We can currently manage ~1M events/week; ramping up gradually
- SC4: distributed reprocessing tests:
  - Test of the computing model using the SC4 data management infrastructure
    - Needs file transfer capabilities between Tier-1s and back to CERN CAF
    - Also distribution of conditions data to Tier-1s (3D)
    - Storage management is also an issue
  - Could use 3 weeks in July and 3 weeks in October
- SC4: distributed simulation intensive tests:
  - Once reprocessing tests are OK, we can use the same infrastructure to implement our computing model for simulation productions
    - As they would use the same setup both from our ProdSys and the SC4 side
  - First separately, then concurrently

ATLAS SC4 Plans (3)

- Distributed analysis tests:
  - "Random" job submission accessing data at Tier-1s (some) and Tier-2s (mostly)
    - Generate groups of jobs and simulate analysis job submission by users at home sites
    - Direct jobs needing only AODs as input to Tier-2s
    - Direct jobs needing ESDs or RAW as input to Tier-1s
    - Make preferential use of ESD and RAW samples available on disk at Tier-2s
    - Tests of performance of job submission, distribution and output retrieval
    - Test job priority and site policy schemes for many user groups and roles
    - Distributed data and dataset discovery and access through metadata, tags, data catalogues.
  - Need same SC4 infrastructure as needed by distributed productions
    - Storage of job outputs for private or group-level analysis may be an issue
  - Tests can be run during Q3-4 2006
    - First a couple of weeks in July-August (after distributed production tests)
    - Then another longer period of 3-4 weeks in November

Overview of requirements for SC4

- SRM ("baseline version") on all storages
- VO Box per Tier-1 and in Tier-0
- LFC server per Tier-1 and in Tier-0
- FTS server per Tier-1 and in Tier-0
- **Disk-only area on all tape systems**
  - Preferably we could have separate SRM entry points for "disk" and "tape" SEs. Otherwise a directory set as permanent ("durable"?) on disk (non-migratable).
  - Disk space is managed by DQ2.
  - Counts as *online ("disk") data* in the ATLAS Computing Model
- Ability to install FTS ATLAS VO agents on Tier-1 and Tier-0 VO Box *(see next slides)*
- Single entry point for FTS with multiple channels/servers
- Ability to deploy DQ2 services on VO Box as during SC3
- **No new requirements on the Tier-2s besides SRM SE**

Movement use cases for SC4

- EF -> Tier-0 migratable area
- Tier-0 migratable area -> Tier-1 disk
- Tier-0 migratable area -> Tier-0 tape
- Tier-1 disk -> Same Tier-1 tape
- Tier-1 disk -> Any other Tier-1 disk
- Tier-1 disk -> Related Tier-2 disk *(next slides for details)*
- Tier-2 disk -> Related Tier-1 disk (*next slides for details*)
- Not done:
  - Processing directly from tape (not in ATLAS Computing Model)
  - Automated multi-hop (no 'complex' data routing)
  - Built-in support for end-user analysis: goal is to exercise current middleware and understand its limitations (metrics)

Tiers of ATLAS

- DQ2 and ProdSys require a Tier-2 to be associated with a Tier-1
- This "virtual" association does not bring additional responsibilities to the sites, except:
  - Tier-1 is responsible for setting up and managing the FTS channel to "its" Tier-2s, as requested by ATLAS
  - Tier-2 will use the LFC server on the Tier-1 as its local catalog
- The "virtual" association is defined by ATLAS (along with the WLCG Collaboration) taking into consideration:
  - FTS channels / network connectivity
  - Available network bandwidth and storage at the Tier-1, wrt to "its" Tier-2s
  - May not be related to EGEE ROC, LCG ROC, …
- Throughout SC4, ATLAS will exercise a simple hierarchical topology (maintained 'manually')

LFC

- **The LFC server must be available at the Tier-0 and at the Tier-1s**
- Tier-2s will have their data registered at the Tier-1 LFC by the ATLAS production system
  - E.g. we manually reset LFC_HOST at the WNs
- Tier-0 LFC: contains catalog entries corresponding to data on the Tier-0 only
- Tier-1 LFC: contains catalog entries corresponding to its Tier-1 data and associated Tier-2s
- Tier-2: no LFC

Overview of FTS and VO Box

- While the VO Box is required for ATLAS (*c.f. VO Box discussion in NIKHEF end January*), new gLite 1.5 FTS provides new functionality discussed with the experiments during the FTS workshop
- ATLAS will, for LCG, slim down its VO Box agents and integrate them with FTS
  - Using what FTS calls "VO agents"
- **But FTS requires some 'area' to host these VO agents!**
- ATLAS proposes, for SC4, to use VO Boxes for this purpose
  - So ATLAS FTS servers are configured to use our VO agents hosted in our VO Box
  - Very much an ATLAS-specific effort taking advantage of gLite FTS 1.5 functionality
  - Facilitates growing of data management to many Tier2s (less ATLAS-specific deployment)
  - Temporary solution for SC4 until FTS VO agents deployment model is found
- Hence, an ATLAS VO Box will contain:
  - FTS ATLAS agents
  - And remaining DQ2 persistent services (less s/w than for SC3 as some functionality merged into FTS in the form of FTS VO agents)

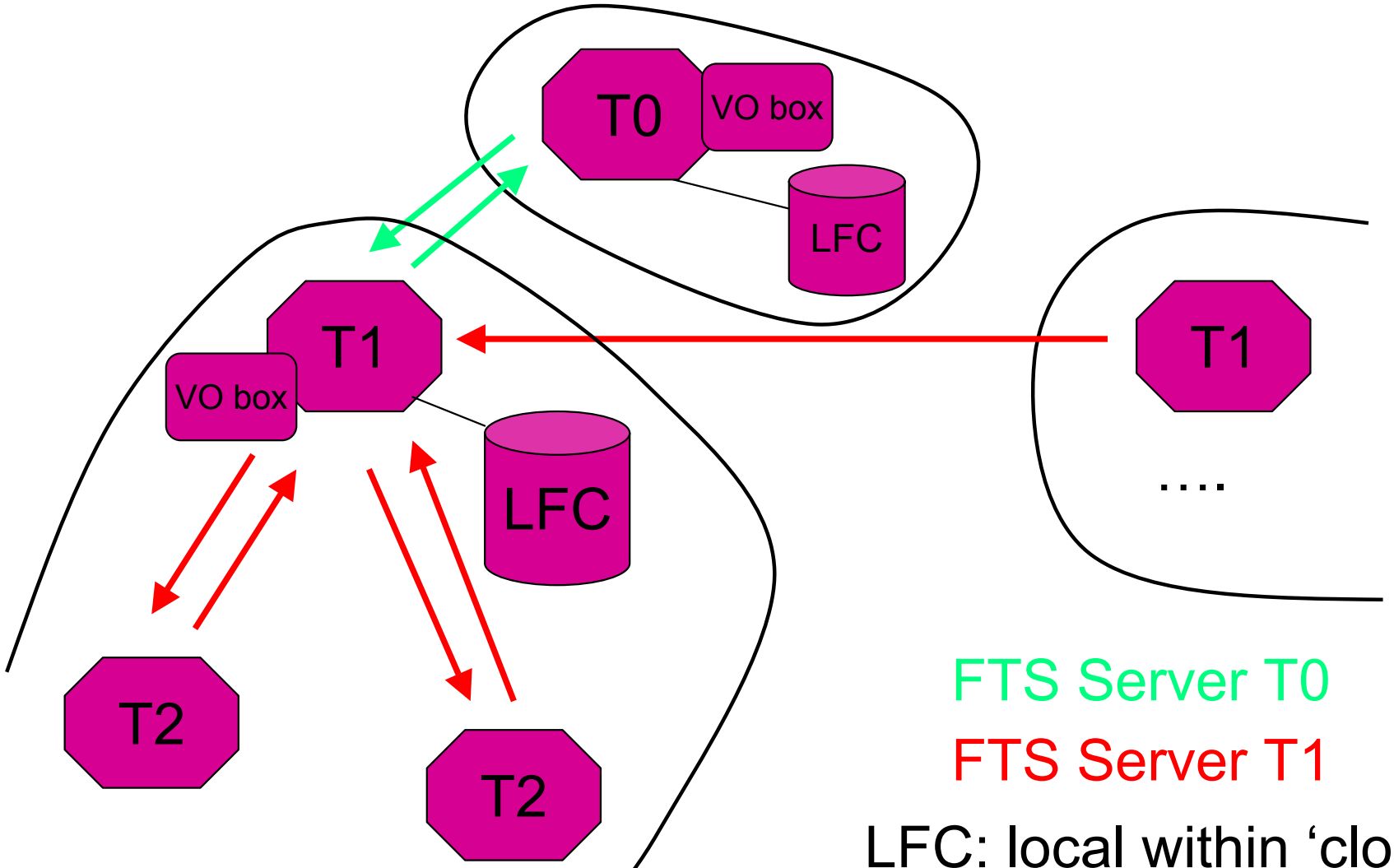
FTS integration and channels

- ATLAS will plug its DQ2 site services onto FTS, using FTS VO agents
  - VO agents will be hosted on the VO Box *(still need to work out details with LCG)*
- FTS channels must be available:
  - Tier-0 FTS server:
    - Channel from Tier-0 to all Tier-1s: used to move "Tier-0" (raw and 1st pass reconstruction data)
    - Channel from Tier-1s to Tier-0/CAF: to move e.g. AOD (CAF also acts as "Tier-2" for analysis)
  - Tier-1 FTS server:
    - Channel from all other Tier-1s to this Tier-1 (pulling data): used for DQ2 dataset subscriptions (e.g. reprocessing, or massive "organized" movement when doing Distributed Production)
    - Channel to and from this Tier-1 to all its associated Tier-2s.
    - Channel from Tier-0 to this Tier-1 (*not so relevant)*

VO Box

- *Assuming presence of same LCG VO Box component as during SC3*
  - But now supporting hosting of FTS VO Agents as well
  - Possibly adding common software (e.g *Apache, MySQL*) to default VO Box environment
- Tier-0 VO Box: contains DQ2 site services now including FTS ATLAS VO agents
- Tier-1 VO Box: contains DQ2 site services now including FTS ATLAS VO agents
- DQ2 site services will have associated SFTs for testing
- Tier-2 VO Box: not required
- *Unclear: how to package, tag and deploy FTS VO agents?*
  - *Have within the LCG VO Box a standard container for FTS VO agents?*

Tiers of ATLAS



FTS Server T0

FTS Server T1

LFC: local within 'cloud'

All SEs SRM

ATLAS SC4 Requirement (new!)

- Small testbed with (part of) CERN, a few Tier-1s and a few Tier-2s to test our distributed systems (ProdSys, DDM, DA) prior to deployment
  - It would allow testing new m/w features without disturbing other operations
  - We could also tune properly the operations on our side
  - The aim is to get to the agreed scheduled time slots with an already tested system and really use the available time for relevant scaling tests
  - This setup would not interfere with concurrent large-scale tests or data transfers run by other experiments
- A first instance of such a system would be useful already now!
  - April-May looks like a realistic request

Summary of requests

- March-April (pre-SC4): 3-4 weeks in for internal Tier-0 tests (Phase 0)
- April-May (pre-SC4): tests of distributed operations on a "small" testbed
- Last 3 weeks of June: Tier-0 test (Phase 1) with data distribution to Tier-1s
- 3 weeks in July: distributed processing tests (Part 1)
- 2 weeks in July-August: distributed analysis tests (Part 1)
- 3-4 weeks in September-October: Tier-0 test (Phase 2) with data to Tier-2s
- 3 weeks in October: distributed processing tests (Part 2)
- 3-4 weeks in November: distributed analysis tests (Part 2)