ATLAS Database Software Status

Database Readiness Workshop

CERN, Geneva, Switzerland

February 6, 2006

Alexandre Vaniachine (ANL)

Outline

Part 1

- ATLAS response to Workshop Goals

Part 2

- ATLAS response to Workshop Focus including:
  - The definition of concrete detector data models
  - Refined estimates for volume and access patterns
  - Connection between online and offline
- See also reports from ATLAS Tier 1 sites

Part 1: Workshop Goals

Review the status of the database project milestones on the experiment side

ATLAS CSC Goals

- 2006 is the year of ATLAS CSC
- CSC goals includes calibration and alignment procedures and conditions

WLCG SC4 Workshop - 12 February 2006

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

Calibration & Alignment

- Included in CSC acceptance tests

WLCG SC4 Workshop - 12 February 2006

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)

ARGONNE NATIONAL LABORATORY

ATLAS Plans for WLCG SC4

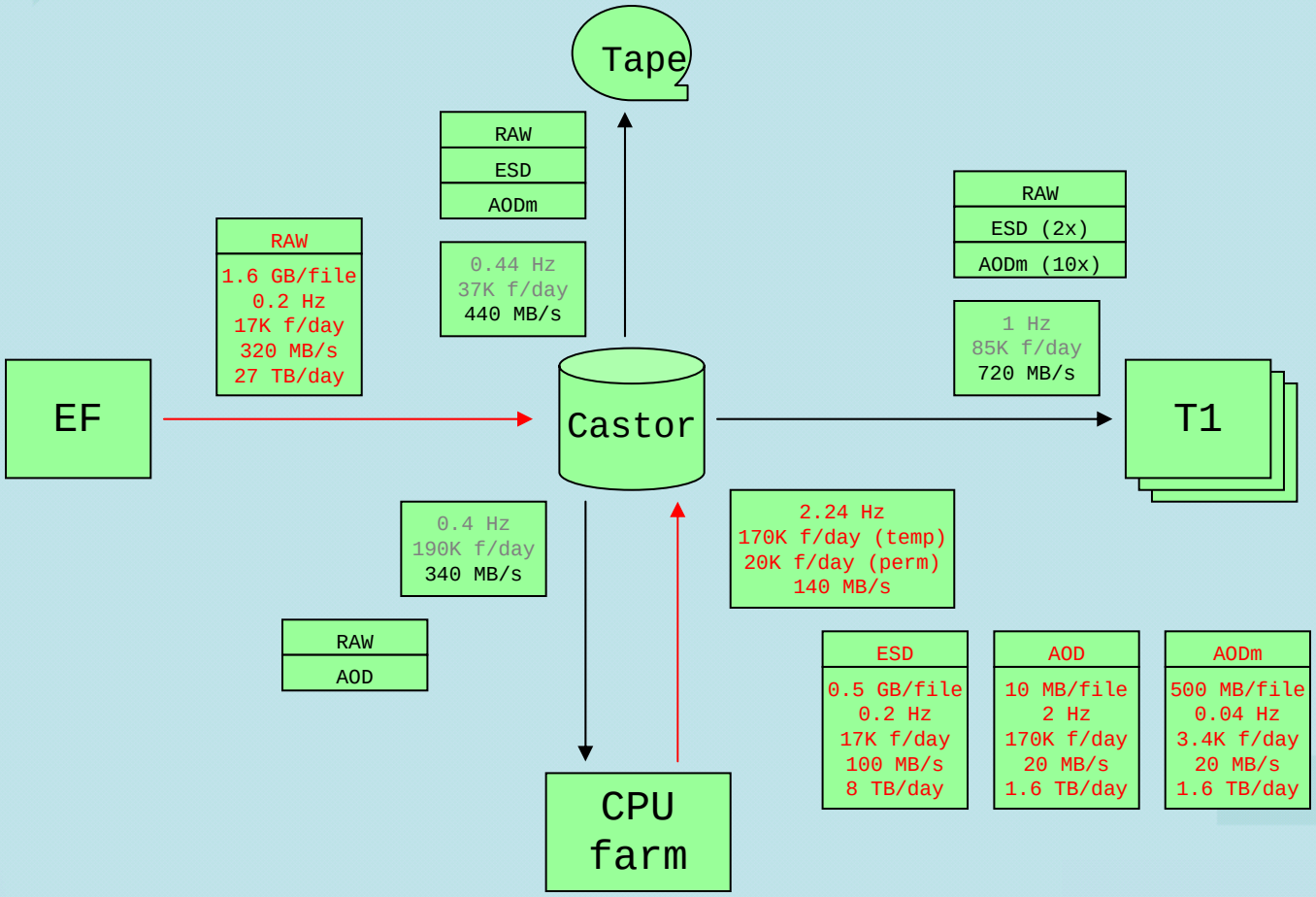
- Discussions of SC4 plans at the ATLAS CMB/SPMB level and below in preparation for the SC4 Workshop at CHEP06

Databases in ATLAS **SC4 Coordination Group** are under ***Zhongliang Ren*** who takes care of the coordination of other operations activities

ATLAS plans for 2006
Computing System Commissioning
Service Challenge 4

Dario Barberis
CERN & Genoa University

SC4: Complete Tier-0 Test



Calibration Loop Exercise

- Included in the Complete SC4 Tier-0 Test

WLCG SC4 Workshop - 12 February 2006

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)

Loaded SC4 Schedule

- All three phases of the SC4 Tier-0 tests include Calib & Align
- By September we need production-level LCG3D

WLCG SC4 Workshop - 12 February 2006

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…

ARGONNE NATIONAL LABORATORY

Part 2: Workshop Focus

To what extend has the development of the main applications (conditions data and other apps to be tested as part of SC4) been completed?

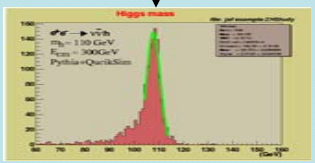
Calibration/alignment "challenge"

- 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 ?

Calib/Align in the Data Flow

L1
L2
Event Filter
RAW
GeomDB
CondDB
File Catalogue
Reconstruction
ESD
AOD Production
AOD
TAG
Future
DPD
Physics Analysis
Event Selection

Event Generation
SIM
Detector Simulation
Pileup
Digitization
Trigger Simulation
SIM

Artwork by Stefan Stonjek

Calibration Loop Streams

- Expected upstream data flow for the calibration data:
  - Not all calibrations will be done at Tier-0
  - Some will be done even at the Tier-2 level
  - The sub-detector Calibration Centers will:
    - receive calibrations event data streams
    - upload calibrations data
- Calibrations are on the critical path:
  - Must have calibrations to proceed with Tier-0 reconstruction
- Discussions of the Calibration Centers operations are in progress
  - Is the upstream data flow planned in other LHC experiments?

LCG3D Replication Tests

Once Oracle streams are established between Tier-0 and Tier-1:

- ATLAS can replicate the 'small' databases such as the Geometry DB
- Start to look at COOL replication, which will involve more data
  - a lot depends on the subdetectors:
    - how they implement their conditions data

Growth in DB Volumes

- Major ATLAS efforts towards realistic conditions results in database volumes growth

WLCG SC4 Workshop - 12 February 2006

"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

ARGONNE NATIONAL LABORATORY

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

Data Volume Limits

- Despite the Calib/Align progress ATLAS does not have yet the exact quantitative data volume numbers at this point
- These are still rather unknown as the sub-detectors continue to develop their calibration models and calibration frequencies
- We expect that the volume will be dominated by the sub-detector with the largest number of channels – i.e. pixels
- Thus we watch closely the ongoing investigation of the benefits of the TOT calibration
  - The Time-Over-Threshold calibration is done on the pixel-level
  - The amount of data per one TOT calibration may approach 1 GB
- But don't be 'scared' – in ATLAS we will store such large calibration volumes outside of relational databases :
  - In the POOL files referenced from COOL IOV database
  - We are already doing with the ~100 MB of calorimeter calibration data

TAG Replication in SC4

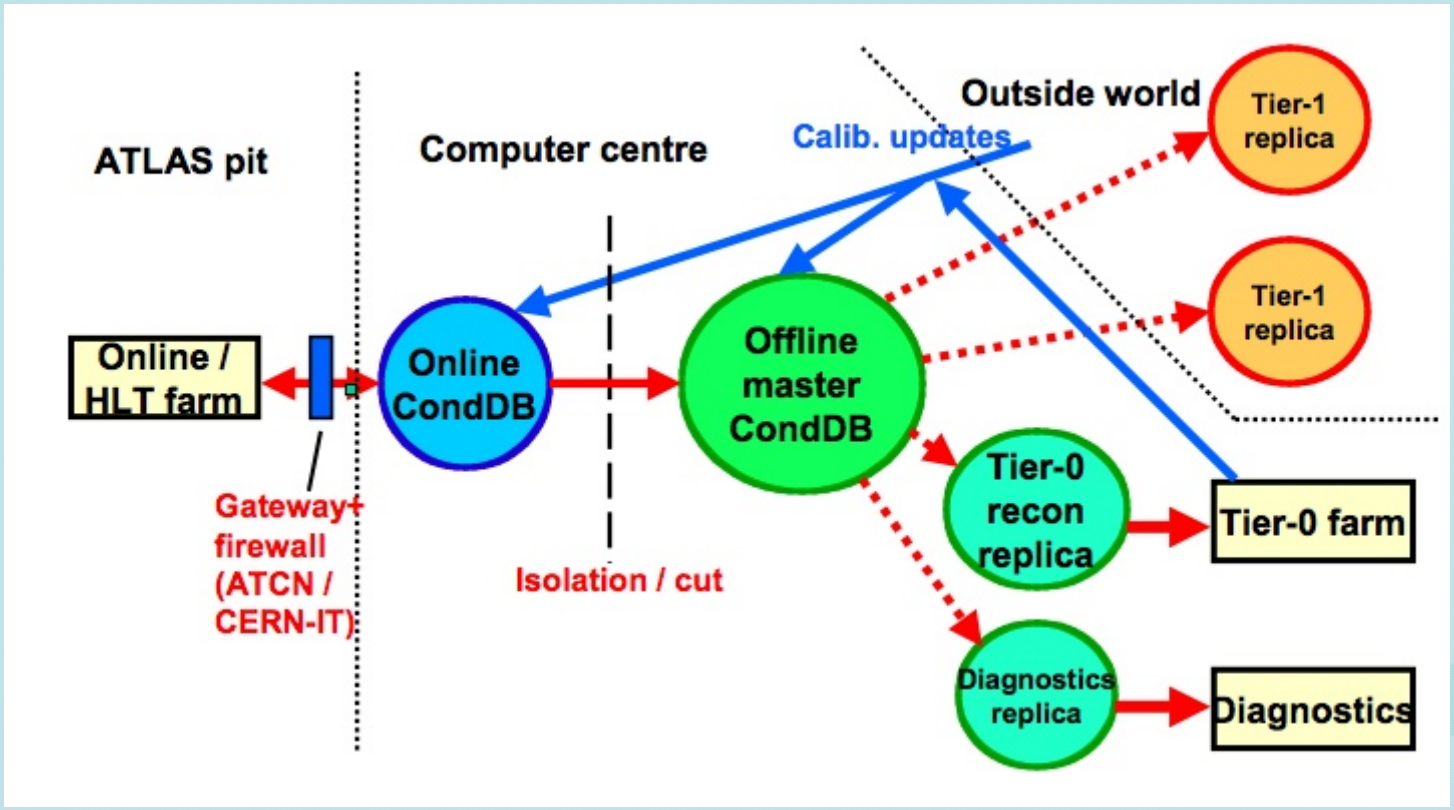
- An explicit part of the Tier-0 test
  - Loading TAGs into the relational database at CERN
  - Replicating it using Oracle streams from Tier-0 to Tier-1s and to Tier-2s
- Also as an independent test, using TAG files that are already available generated
- TAG will have a lot more data than ATLAS will use in COOL replication in the near future

TAG Replication Schedule

- The Event Store Workshop for ATLAS developers is scheduled for February 27 - March 1 at CERN
- Among all Event Store issues the Workshop is allocating half the time to the metadata topics in general and to the event-level metadata - the TAG database - in particular
  - Including TAG database replication
- TAG production at the Tier-0 and replication to Tier-1s is expected in May
  - The details need to be negotiated with the Tier-0 team
- The Workshop will finalize TAG database replication milestones as a part of the Tier-0 SC4 test

Online-Offline Connection

- A well-documented schema was reported at the last LCG3D Workshop



Artwork by Richard Hawkings

Online Activities

- TriggerDB is the DB application for the full trigger system
- It has a rather involved schema to describe the selection
  - also has a RAL-based interface software to read out this data
- For easy population we have a GUI and python scripts
- Now integrating the read-out software with offline/online
- TDAQ has a separate scalability testing schedule outside of SC4:
  - For the TDAQ LVL1 Trigger the data volume is small (<1 MB), the number of clients is around 20
    - Test of the performance are to be done in the Spring
  - For the TDAQ HLT the information is much bigger and the number of clients is of the order of thousands
- A program of work is beginning now, with the goal of having a fully functional system for the large scale tests at the end of 2006

TDAQ Large Scale Test

- The Large Scale ATLAS TDAQ Test will use significant CERN resources
- Thus, the Large Scale TDAQ Test in will follow the completion of Tier-0 tests and will start in October-November of 2006
- All TDAQ LVL1/HLT performance tests are done with the help of ATLAS database project team (in coordination through Richard)

COOL Status in Offline

- COOL software fully integrated into Athena from Summer 2005
  - Integration consistent with old Lisbon MySQL implementation,
    - Subdetector clients can migrate adiabatically

Data model:

- small data payloads directly 'inline' in COOL
- for large payloads, COOL stores:
  - POOL-token refs to data in POOL files or
  - POOL object-relational (only in prototype form now)
- Functionality and performance testing ongoing in both online and offline environments

COOL Usage in Offline

Currently small amounts of conditions data in COOL

- Production: uses Lisbon ConditionsDB (MySQL)
  - Primarily conditions data from 2004 Combined Test Beam (CTB)
  - Full migration to COOL is underway:
    - CTB data being moved to phase out Lisbon by Summer 06
- Commissioning: conditions data from subdetector
  - Already started
  - Going to COOL not Lisbon

COOL usage grows in simulation and reconstruction now :

- COOL data volumes and use will increase significantly

Software Next Steps

Before SC4 production phase:

- Migrate to CORAL-based COOL
  - Especially for indirection/replica/failover
- Handle schema evolution of existing data
- Need HVS or equivalent for COOL tag management

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

Deployment Next Steps

- Need in production:
  - 'Static' replication of COOL conditions data from February 2006
  - Dynamic replication from June 2006
- Customers:
  - Ongoing support for commissioning data
  - Major simulation production from April 2006
  - Major reconstruction production from July 2006
    - Including calibration
- Closed loop cycles:
  - reconstruct, improve calibration, re-reconstruct

Replication Strategies

- COOL API-level copy from Oracle->SQLite
  - for static replicas - works
- Explore Oracle streams for COOL Tier-0 -> Tier-1
  - as soon as 3D ready
- Need dynamic COOL API-level copy (updating replicas)
  - from June 2006
- Evaluate Frontier-based replication as soon as possible
  - New ATLAS manpower (US student) available February 6
- Experience with different solutions in first half of 2006 will guide replication choices at various Tiers for calibration-closed loop cycles in second half of 2006

Cache Consistency

- We appreciate CERN studies of the cache consistency

We plan to do practical testing with Frontier to determine what our FroNTier caching methods will be with COOL:

- Payloads are cached and the IOV lookups are not:
  - The payload lookups will carry an identifier for the particular version/object being retrieved
  - There are use cases when the IOV lookups are not worth caching because they will rarely be repeated
  - Or the IOV lookups could only be cached in certain circumstances
    - e.g. in HLT caching IOV lookups could be very useful
- In that way we would avoid the cache consistency problems
  - as long as we do not reuse these identifiers
    - which we do not in the COOL model

Ideas for Cache Managing

- To solve the FroNTier cache invalidation issue we will manage cache consistency in a 'controlled' production environment with well-defined 'database content' releases:
  - For example, we wouldn't launch the distributed reconstruction of a particular dataset (e.g. a particular period of data, or a set of MC samples) until the appropriate conditions DB data are available
  - Before launching, we could invalidate ALL of our caches, and let them repopulate themselves as data is requested
- By construction, we would not then be expecting this data to change 'under our feet' during a reconstruction pass
- The same applies in the HLT environment, where we could invalidate caches before each run start
- But this obviously does not apply in chaotic end user environments, and for individuals developing new calibrations

These are just ideas at this stage - they will need to be tested

Conclusions

2006 in ATLAS is a year of the Computing System Commissioning (CSC)

- COOL commissioning is a part of CSC
- ATLAS Tier 0 tests (part of WLCG SC4) include
  - Growing ATLAS database applications: (1) Realistic Conditions DB: Geometry DB, Magnetic Field, etc; (2) Calibration and alignment; (3) TAG database
  - A production quality system using 3D tools is scheduled for September
    - By then, we need to have made choices regarding: Frontier vs replicas, the proper use of Oracle streams, caching strategies etc.
    - We are interested in testing 3D tools as soon as they are ready
- Database project milestones after completion of Tier 0 tests in September:
  - Database infrastructure for TDAQ scalability test
    - scheduled for October-November
  - Deployment of database infrastructure to support large scale distributed production (following FroNTier evaluation in ATLAS)