ATLAS Tier-0 Scaling Tests

Luc GOOSSENS
Armin NAIRZ
Gilbert POULARD
Miguel BRANCO
David CAMERON

CERN, Geneva/Switzerland

PH/ATC Group

Overview of this Talk

- Introduction: Tier-0 Functions
- Tier-0 Architecture
- Tier-0 Scaling Tests: Scope, Scale and Initial Plans
- Test Setups and Results
- Conclusions
- Future Plans

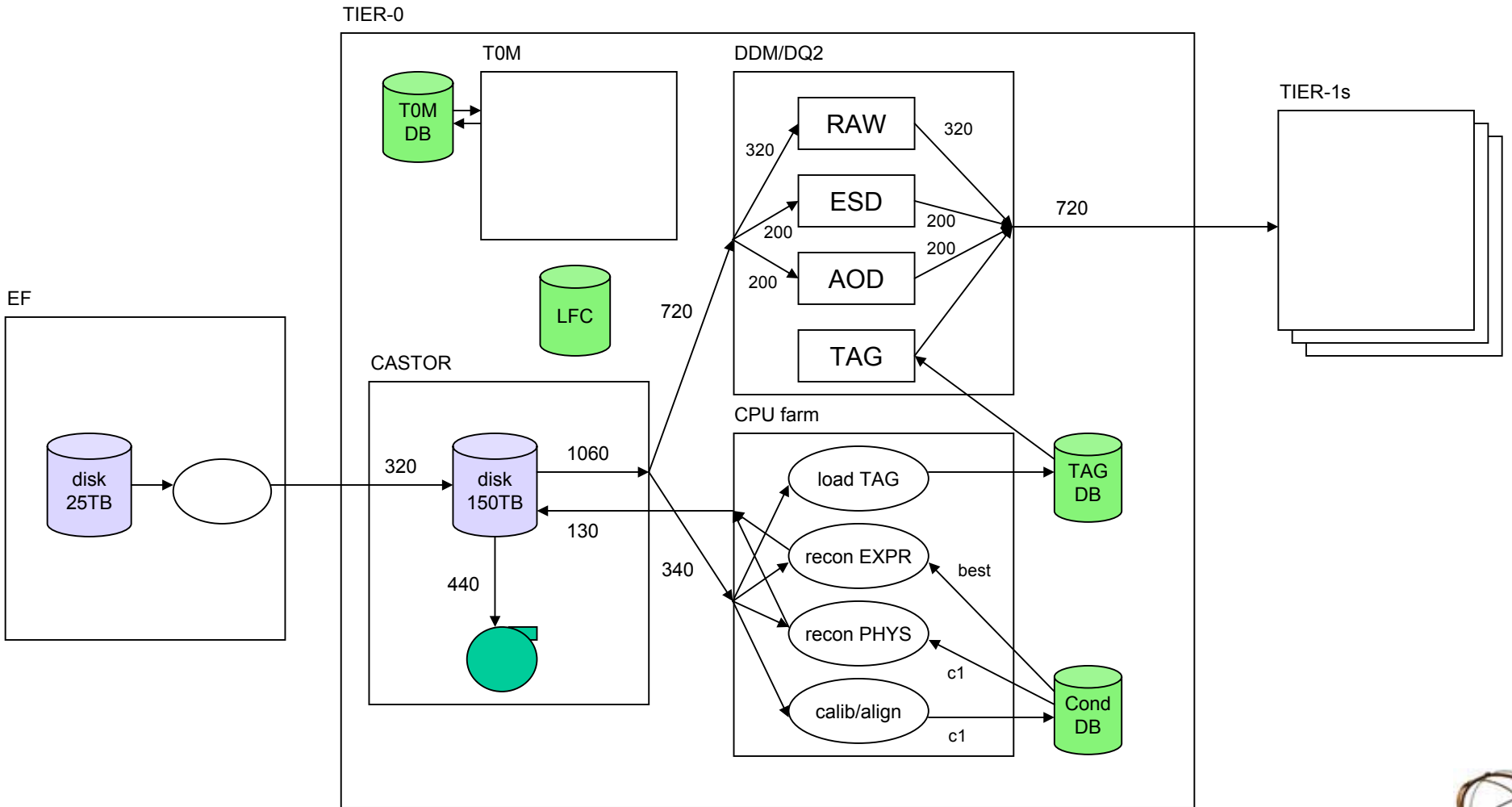
- Introduction: Tier-0 Functions
- Tier-0 Architecture
- Tier-0 Scaling Tests: Scope, Scale and Initial Plans
- Test Setups and Results
- Conclusions
- Future Plans

Tier-0 Functions

- calibration and alignment processing
- first pass ESD (Event Summary Data)/AOD (Analysis Object Data) and TAG production (= reconstruction)
- archiving of RAW data and first pass ESD/AOD/TAG on tape
- distribution of RAW data and first pass ESD/AOD/TAG to Tier-1s
  - one copy of RAW
  - two copies of ESD
  - 10 copies of AOD/TAG

- Introduction: Tier-0 Functions
- Tier-0 Architecture
- Tier-0 Scaling Tests: Scope, Scale and Initial Plans
- Test Setups and Results
- Conclusions
- Future Plans

Tier-0 Architecture



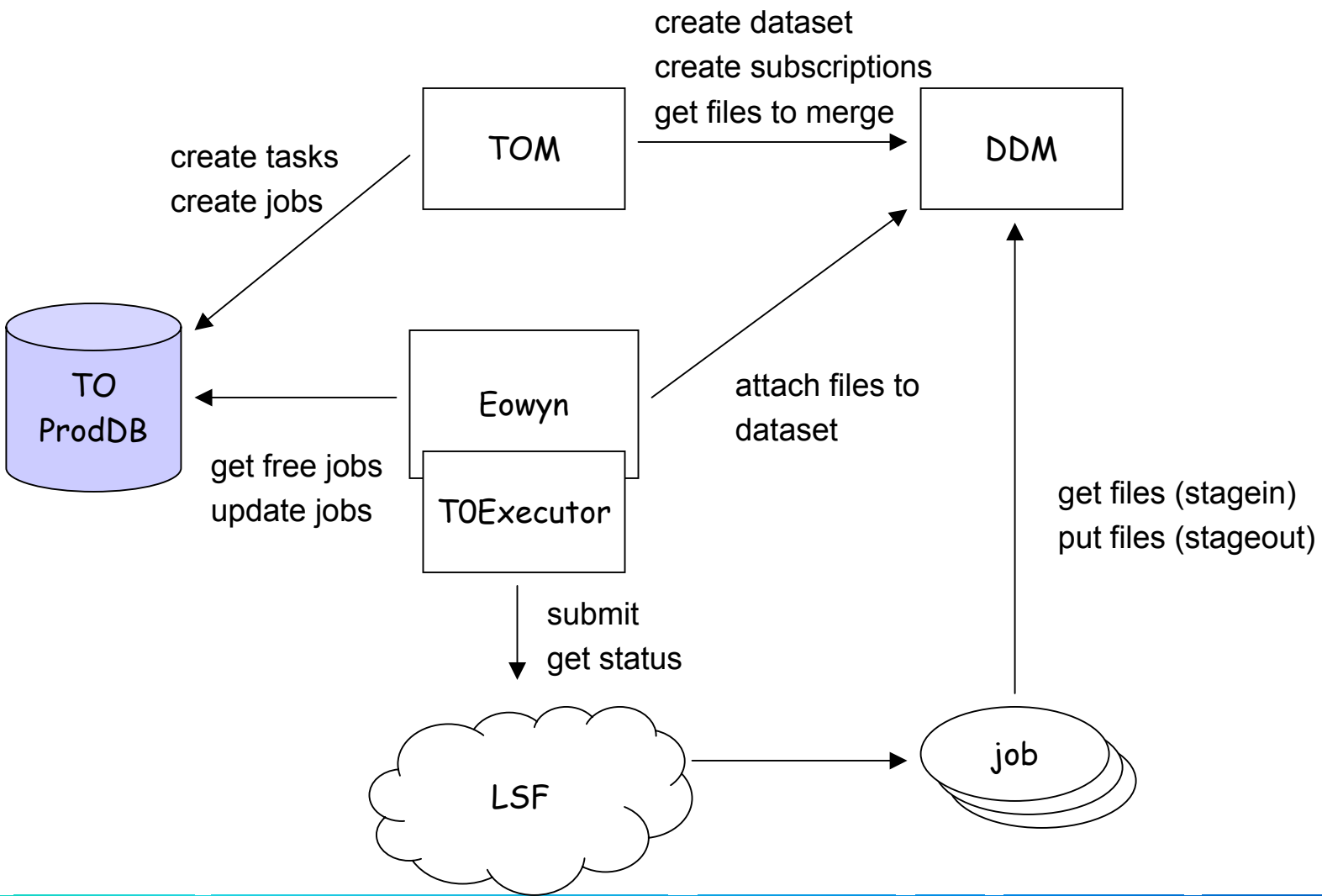
Tier-0 Architecture

- Event Filter (EF) writes RAW data into Castor
- Tier-0 Manager (T0M) defines and orchestrates all Tier-0 jobs
  - data driven
  - reconstruction jobs, AOD merging jobs, TAG uploading jobs, calibration & alignment jobs, ...
  - jobs run on LSF batch system at CERN
- T0M enters Tier-0→Tier-1 export tasks into instance of ATLAS distributed data management system (DQ2)

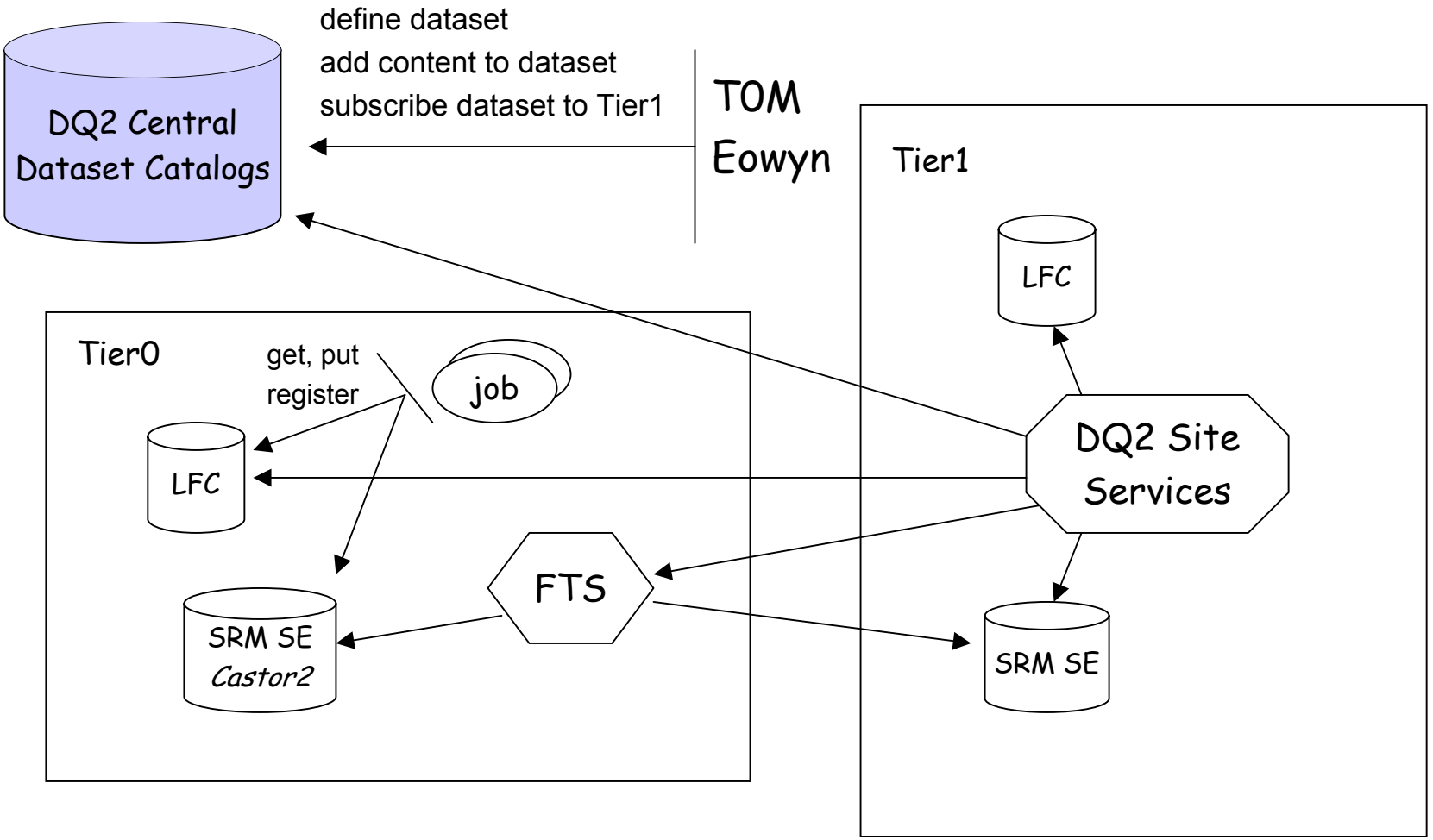
TOM Architecture

- based on ATLAS Production System (ProdSys)
  - ProdSys is well-tested and reliable/mature
- TOM only needs to define jobs based on the data (as recorded in DQ2), ProdSys takes care of running the jobs on LSF and interacting with DQ2 (registration of output files in catalogues)
- ProdSys architecture features a 'facility' neutral supervisor component and 'facility' specific executors
  - standard supervisor Eowyn was used
  - custom Tier-0 executor was written interfacing to LSF but also exploiting the peculiarities of the Tier-0 setup

TOM Architecture



Distributed Data Management



- Introduction: Tier-0 Functions
- Tier-0 Architecture
- Tier-0 Scaling Tests: Scope, Scale and Initial Plans
- Test Setups and Results
- Conclusions
- Future Plans

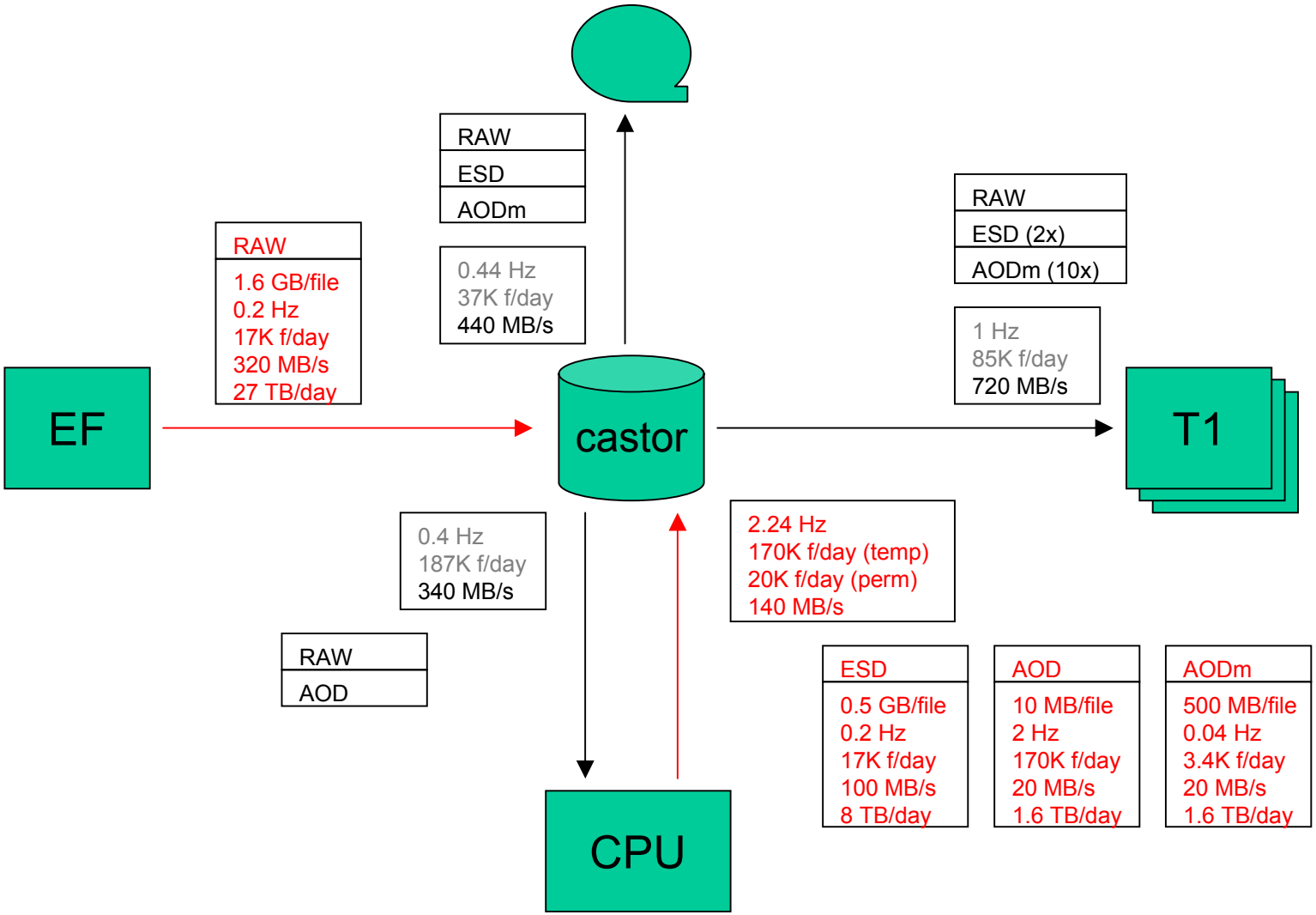
Scope of the Tier-0 Scaling Test

- calibration & alignment processing not included yet
- CondDB and TagDB streams only if exercising the actual access patterns
  - in the end not exercised
- ... leaving :
  - EF writing into Castor
  - ESD/AOD production on reco farm
  - archiving to tape
  - export to Tier-1s of RAW/ESD/AOD
- the goal was to test as much as possible, as realistic as possible
- mainly data-flow/infrastructure test (no physics value)

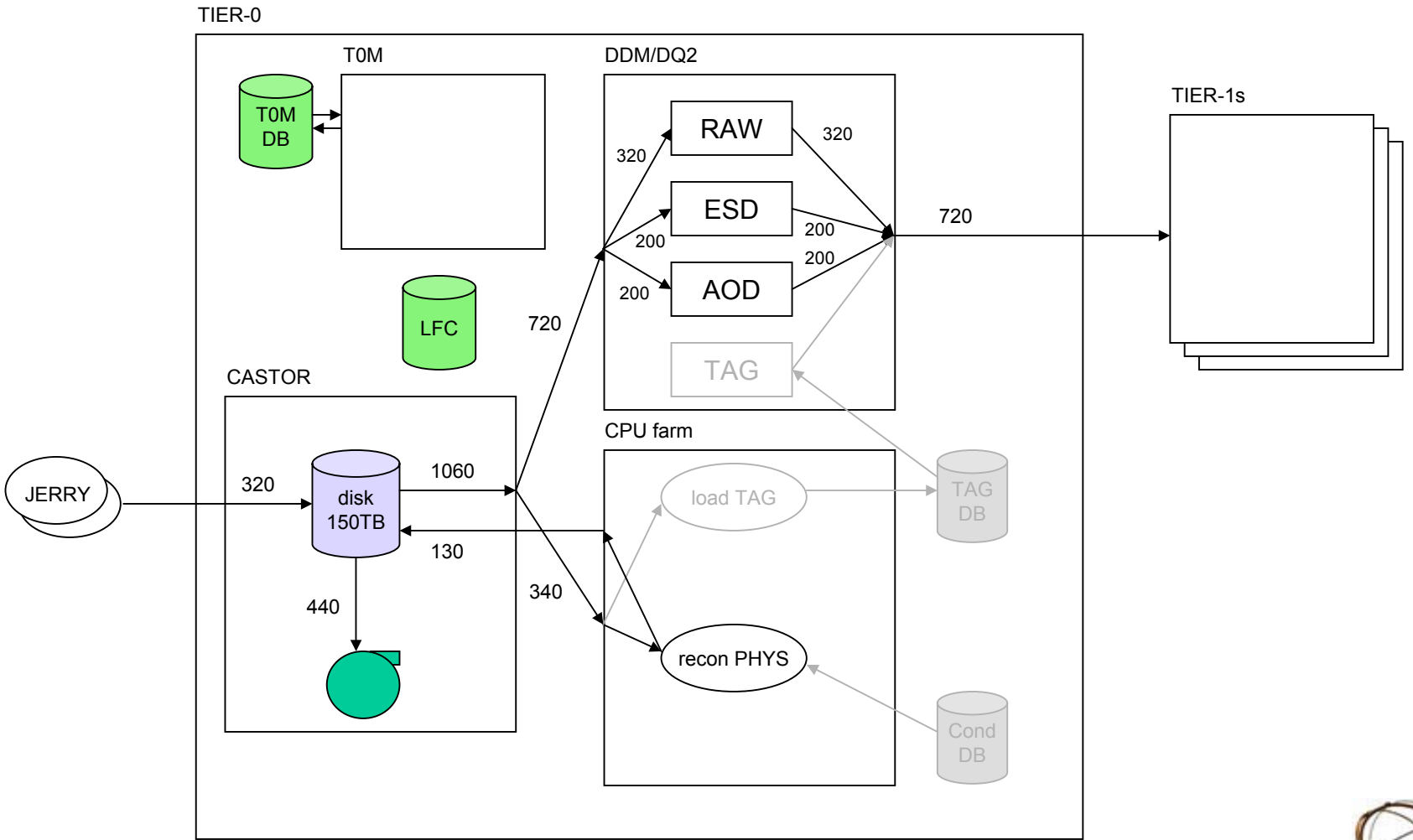
Scale (Nominal)

- O(10K) jobs per day
- O(10K) permanent files per day
  - RAW, ESD and merged AOD
- O(100K) temporary files
  - unmerged AOD
- disk writing 460 MB/s
- disk reading 1500 MB/s
- tape writing 440 MB/s
- approx 3000 recon jobs in parallel for real recon

Scale (Nominal)



Scale (Nominal)



Strategy

- modular approach with stub/fake replacement components for the ones missing or having reached their max platform
  - EF→ Castor stream e.g. was replaced with Jerry jobs running on LSF producing RAW files at controllable rate
    - generates comparable writing load on Castor
  - planned also to have fake Tier-1
    - generating comparable reading load on Castor
    - not obvious, so not implemented (yet)
- replacing real recon with sleep variant
  - generating the same amount of IO but consuming no CPU
    - can run 8 jobs per CPU instead of just one
  - allows to trade sleep time vs. number of parallel jobs
    - introduces artificial limit on number of parallel readers/writers
    - also test degradation for 3000 (fake) job case !!
- export to Tier-1s: merged with LCG Service Challenge 3 (SC3)

Initial Plan

- start in week 42 (Oct) at 1%
- ramp up gradually to 100% in week 51 (Dec)
- weekly targets: 1%, 2%, 5%, 10%, 10%, 20%, 20%, 50%, 50%, 100%

- Introduction: Tier-0 Functions
- Tier-0 Architecture
- Tier-0 Scaling Tests: Scope, Scale and Initial Plans
- Test Setups and Results
  - Oct-Dec 2005 Test
  - Jan 2006 Test
- Conclusions
- Future Plans

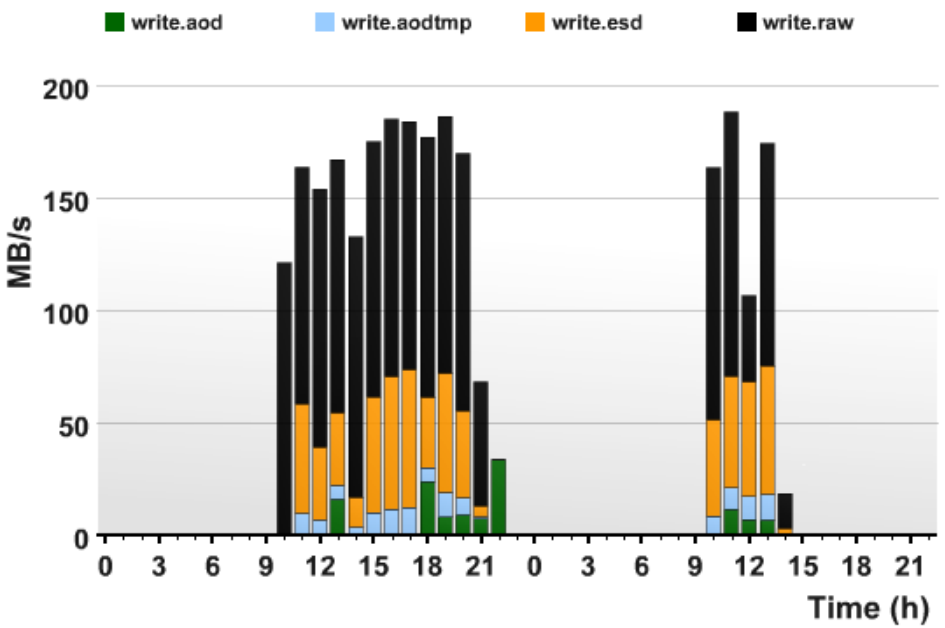
Oct-Dec 2005 Test

- results
  - implementation and integration of all necessary Tier-0 software components
    - Eowyn + Tier-0 Executor + DDM (DQ2)
    - Jerry (EF emulation)
    - TOM (with functionality to handle RAW/recon jobs and AOD merging jobs)
  - week 21 Nov: successful tests running 'real' Athena reconstruction
    - reached sustained rate of 60 jobs/hour on our 60-machines farm
    - corresponds to 25 MB/s Castor→farm (8%), 10 MB/s farm→Castor (7%)
  - week 5 Dec: reached 273 MB/s aggregated read/write inside Tier-0 (>30%)
  - week 12 Dec: reached peak rate of 220 MB/s (>30%) in Tier-0→Tier-1 export
    - up to 7 Tier-1s involved simultaneously, 8 Tier-1s involved in total

Oct-Dec 2005 Test

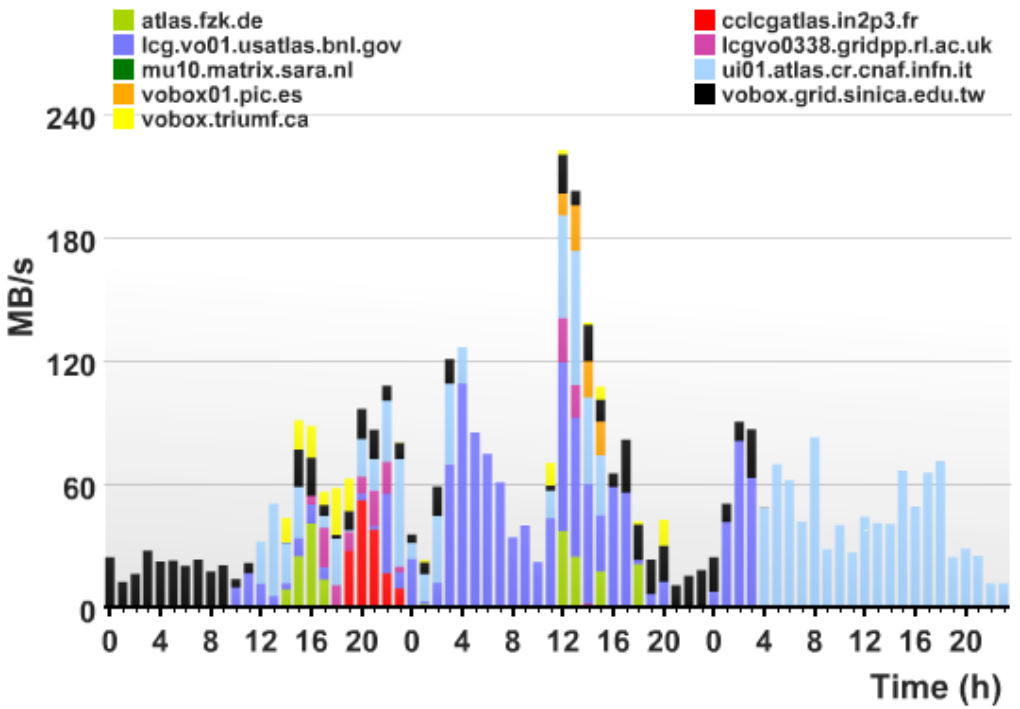
Data throughput from 19/12/2005 0:00 to 20/12/2005 22:00

Tier 0 Internal Transfers



Data throughput from 15/12/2005 0:00 to 18/12/2005 0:00

Tier 0 - Tier 1 Transfers



Oct-Dec 2005 Test

- **hardware evolution**
  - CPUs:
    - started with 10 LSF nodes (first testing, Oct)
    - gradually ramped up to finally 10 (EF farm) + 120 (reco farm) LSF nodes (last weeks of Dec)
  - disk servers: ramp-up from 4 → 8
  - tape writing dropped out at an early stage of the exercise
- **main problems encountered**
  - insufficient network connectivity
    - made us stop actually before week 51 (limits reached)
  - bug in Castor2 concerning (premature) file purging
  - slow deployment of DQ2 onto Tier-1s
  - occasional instabilities on storages at the Tier-1s
    - difficult to sustain rates
  - difficulty to monitor Grid transfers
- **intermediate conclusions and measures**
  - easily reached limits of hardware configuration
  - negotiated additional test time in January, after major hardware upgrades

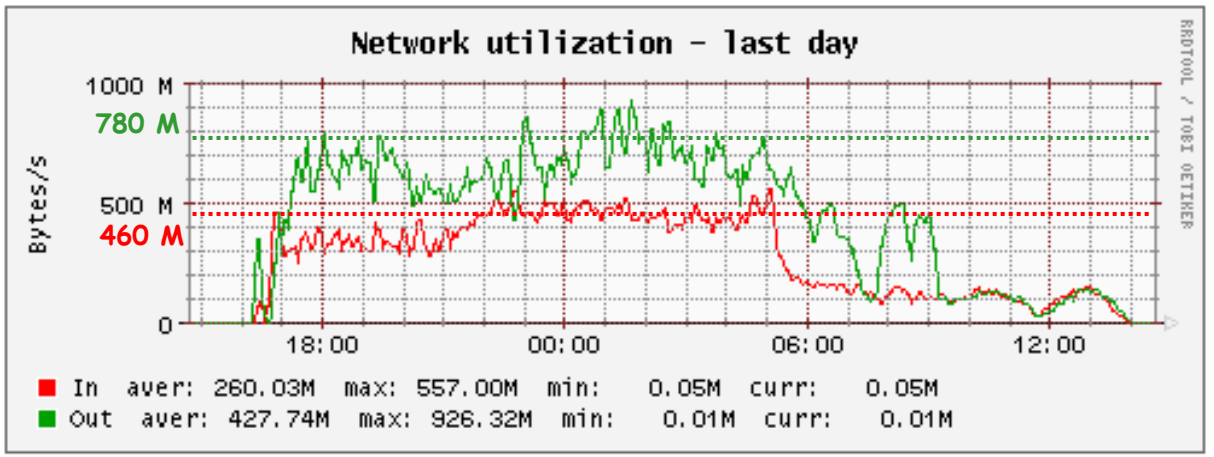
Jan 2006 Test

- **hardware configuration**
  - clusters of ~20 LSF machines to emulate Event Filter (SFO), ~200 LSF machines (reco farm)
  - all connected to the same 10-Gb switch
  - initially 48 disk servers, 24 tape servers; finally 24 disk servers, ~12 tape servers
  - no export to Tier-1s
    - LCG Service Challenge 3 re-run in parallel !
- **run characteristics**
  - jobs required writing/reading to/from disk, writing to tape
    - no recall from tape
  - Jerry, TOM, Eowyn/TOExecutor run from CERN lxplus machines (using shared file system AFS)
  - number of parallel jobs: from 100-200 (week 2 - week 3) to up to 3000 (~15 per node; week 4)
    - latter needed special queue and node configurations
  - number of parallel streams:
    - writing to disk: 8-15 (EF), ~40 (reco; but large fluctuations)
    - reading from disk: ~40 (reco; but large fluctuations)

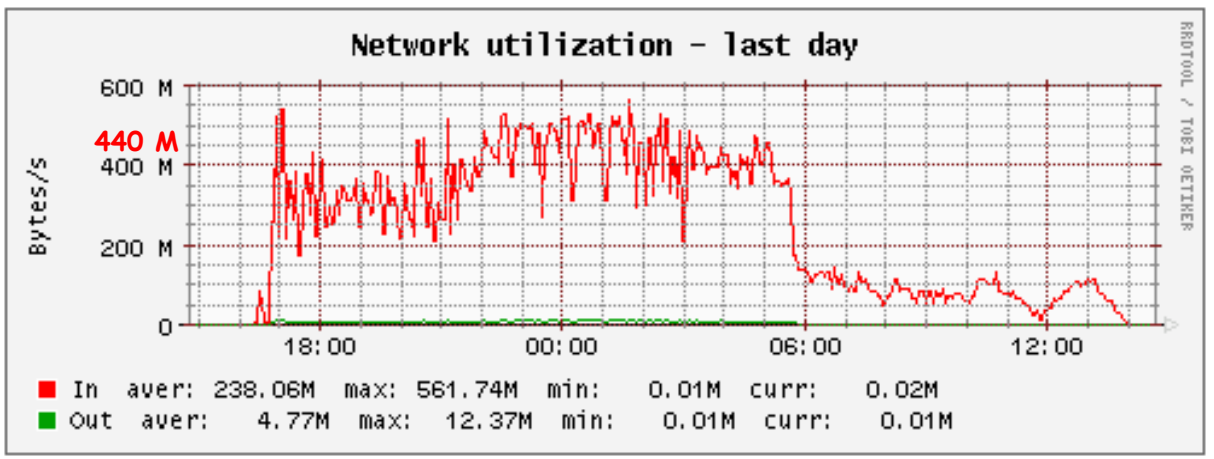
Jan 2006 Test

READING (nom. rate: 780 MB/s)
- Disk → WN
- Disk → Tape

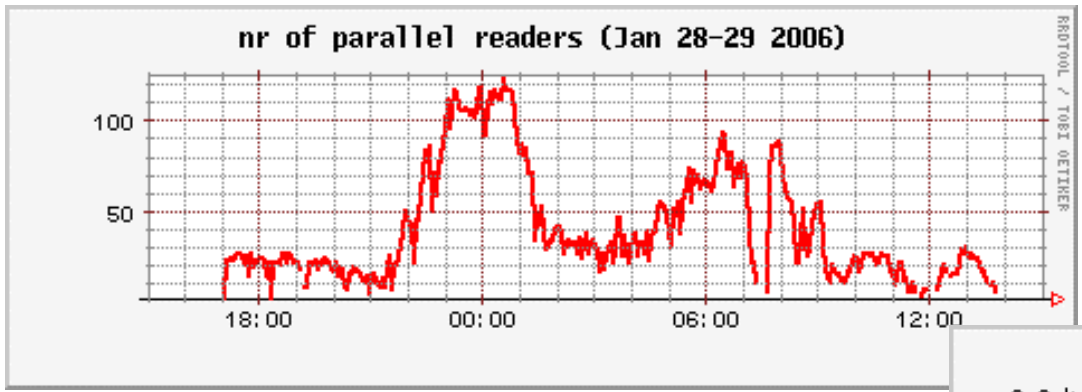
WRITING (nom. rate: 460 MB/s)
- SFO → Disk
- WN → Disk

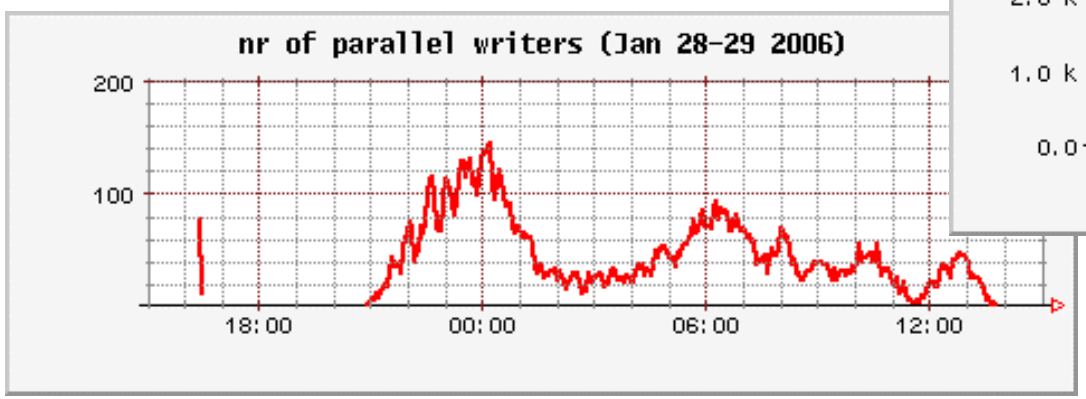


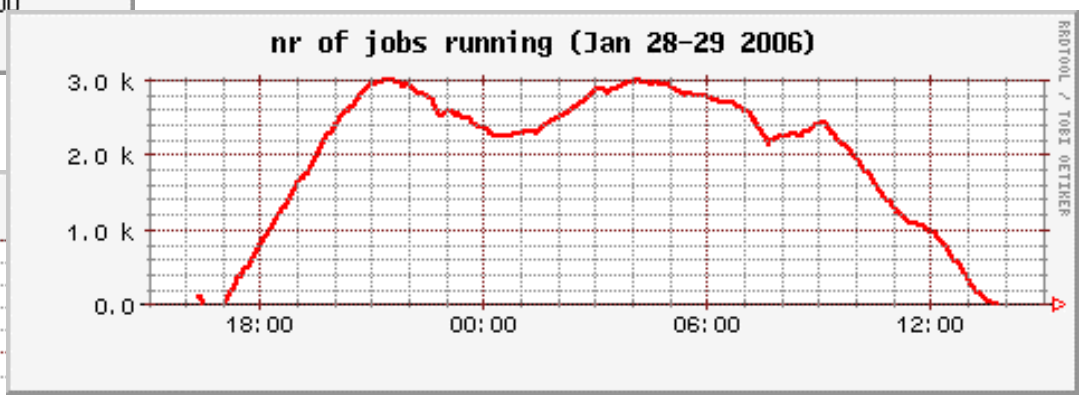
WRITING (nom. rate: 440 MB/s)
- Disk → Tape



Jan 2006 Test







Example snapshots from (post-run) monitoring

- Introduction: Tier-0 Functions
- Tier-0 Architecture
- Tier-0 Scaling Tests: Scope, Scale and Initial Plans
- Test Setups and Results
- **Conclusions**
- Future Plans

Main Problems Encountered

- **several iterations necessary to get our (unusual) LSF configuration correct**
- **LSF system could not manage our load of 3000 jobs**
  - latency/response time got too long to sustain required submission rate and total number of jobs
- **several iterations necessary to get the Castor2 configuration correct**
  - bug in file purging algorithm
  - problems with load balancing strategy
  - changed hardware several times
  - additional complication of insufficient number of tapes available
- **pre-xmas test was dominated by insufficient network connectivity**
  - took a long time (too long) to figure this out
- **too limited monitoring both Castor2 side and ATLAS side**
- **multiple problems in sustaining transfer rates to Tier-1s**
  - unstable storage
  - several iterations to get required ATLAS configuration

Conclusions

- in the end we reached our goal for this first test: nominal rates with nominal jobs for Tier-0 internal data flows
  - major success ! and above expectations !
  - this is without non-negligible 720 MB/s Tier-0→Tier-1 export
- however, large fraction of the time smth was wrong/not working/...
  - we depend on many things: LSF, Castor2, Grid middleware, AFS, ...
  - will require major improvement on this for next round
- many problems were identified and addressed
  - network and Castor2 should be OK now
  - LSF still needs some work
    - dedicated Tier-0 LSF master ?

- Introduction: Tier-0 Functions
- Tier-0 Architecture
- Tier-0 Scaling Tests: Scope, Scale and Initial Plans
- Test Setups and Results
- Conclusions
- Future Plans

Future Plans

- T0-2006/0 test
  - 1 week beginning of April (to be confirmed)
  - intermediate test
- T0-2006/1 test
  - 3 weeks end of June (to be confirmed)
  - goals:
    - SFO (simulator) in ATLAS pit and pit → Castor link
    - realistic CondDB and TagDB dataflows
    - comprehensive real-time monitoring in place
      - » towards control room operations
    - automatic and continuous operation for O(week)
    - exercise induced error scenarios
- T0-2006/2 test
  - 3 weeks in September (to be confirmed)
  - consolidation of June test