



ATLAS SC3 report and SC4 outline

Dario Barberis
CERN & Genoa University

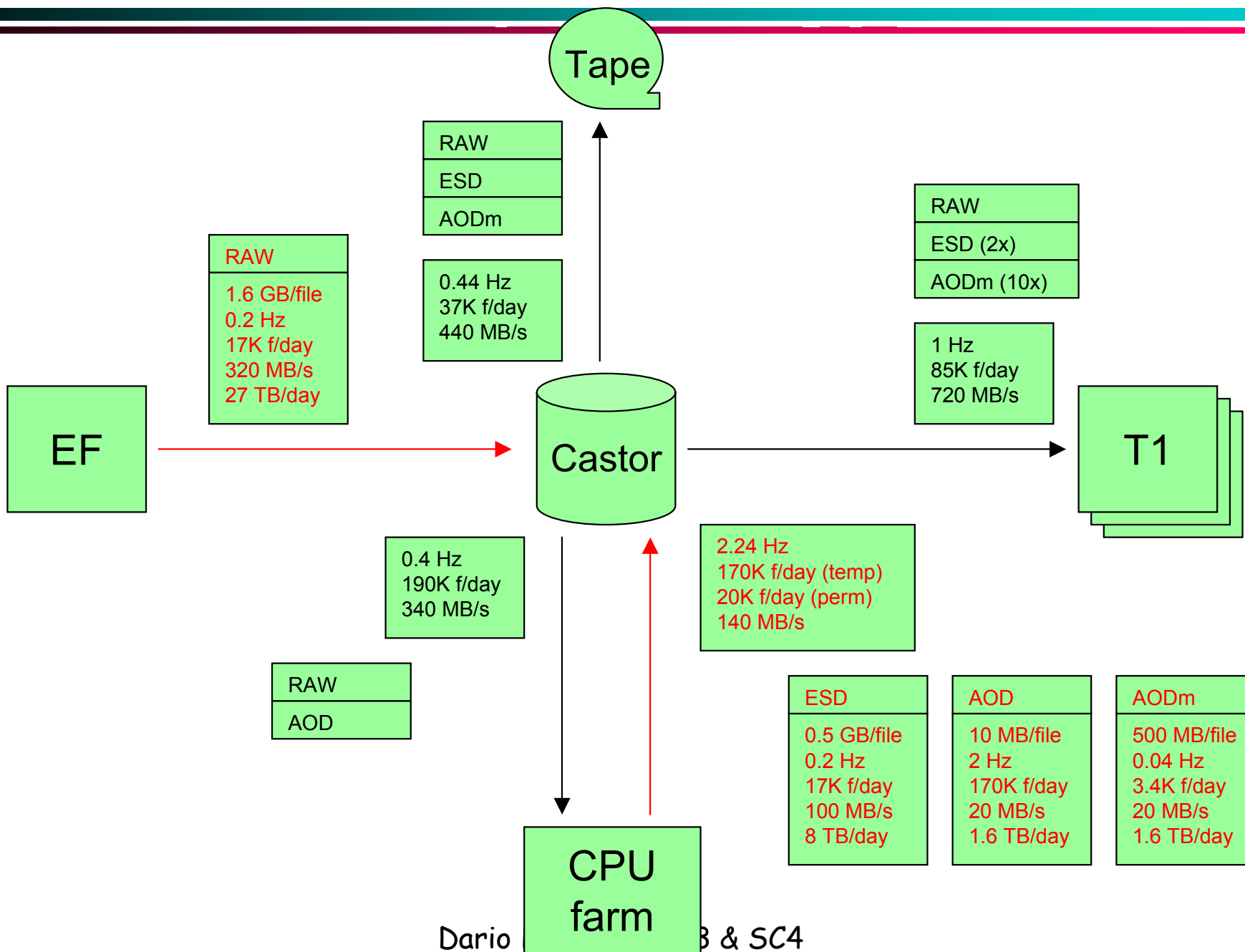
ATLAS SC3 original plans

- Complete tests of Tier-0 operations
 - Including data distribution to Tier-1s and Tier-2s
 - Following the Computing Model as described in the C-TDR (June'05)
- Distributed production exercise
 - Simulation run at Tier-1s and Tier-2s
 - Output data stored at Tier-1s
- Timescales:
 - Low-rate tests in September-October
 - Main Tier-0 runs in November
 - Distributed production continuing "forever" using the established service infrastructure

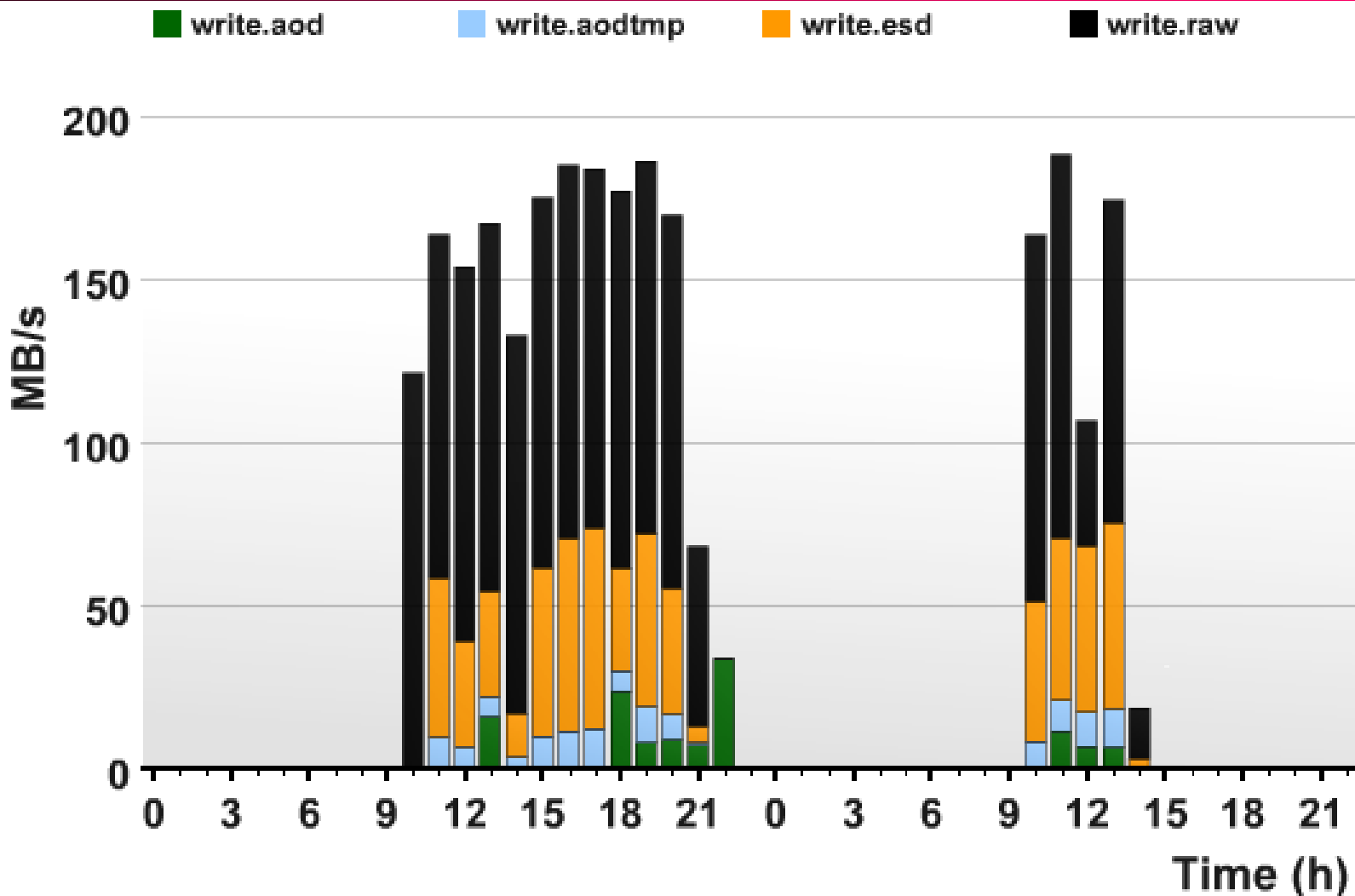
ATLAS SC3 re-scoping

- Reality check at end of summer 2005:
 - FTS channels established only for Tier-0→Tier-1 traffic
 - No FTS available for data distribution to Tier-2s
 - FTS not available for distributed production
 - Instabilities in Castor (since solved) and in data management and network services at many sites
 - ATLAS new ProdSys not 100% ready
- Rescoping of ATLAS SC3:
 - Drop the distributed production part from SC3
 - It is run now continuously as a separate operation
 - Factorize the 2 parts of Tier-0 operations:
 - Internal Tier-0 data flow
 - Data transfer from Tier-0 to Tier-1s

ATLAS Tier-0 data flow



Castor writing rates (19-20 Dec.)



- EF farm → Castor ("write.raw")
- reco farm → Castor ("write.esd" + "write.aodtmp" for reco jobs, "write.aod" for AOD-merging jobs)

ATLAS SC3 Tier-0 tests

- 20 December 2005:
 - Tier-0 operation test reached:
 - 30% of full ATLAS data rate (320 MB/s) from event filter to CASTOR
 - 50% (of 340 MB/s) from CASTOR to Tier-0 CPU farm
 - 50% (of 140 MB/s) from Tier-0 CPU farm to CASTOR
- We had 16 CASTOR servers in December, we'll reach 48 soon for the rerun of the Tier-0 operation test in Jan. 2006 (starting today).

LCG Service Challenge 3

- The T0-T1 part of the service challenges are a test of the Grid middleware
 - FTS: reliable file transfer service
 - SRM: unified mass storage interface
 - LFC: file replica catalog
- LCG runs a 'throughput' phase, then lets experiments try themselves with their own software on top
 - Since these are 'simple' generic components, we add complex ATLAS functionality
 - Notion of datasets and controlled movement
- We started officially on 1st Nov
 - The aim was to ramp up week by week 2%, 5%, 10%, 20% of the operational data flow (320MB/s EF -> Tier 0, 720 MB/s combined Tier 0 -> Tier 1)

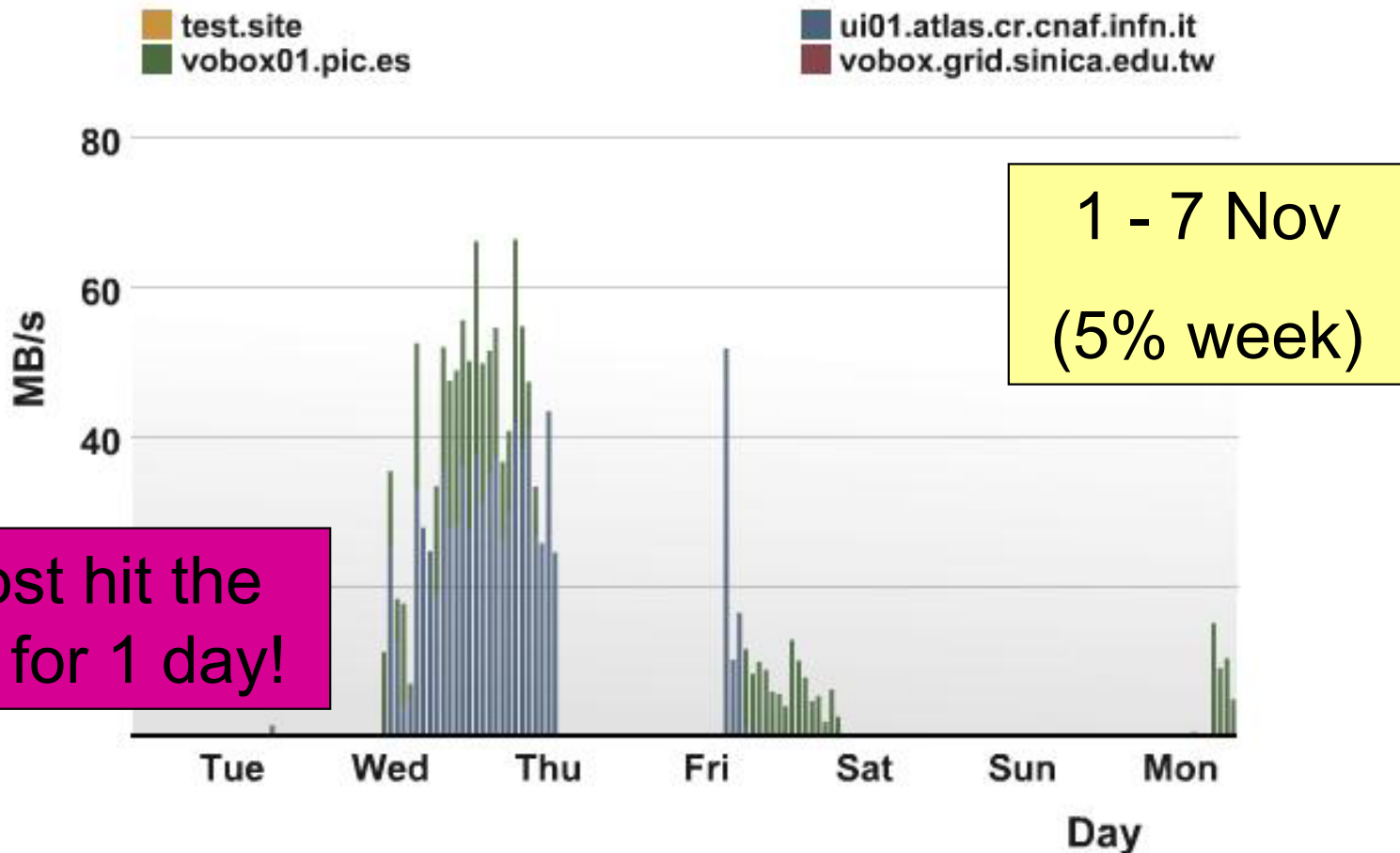
Problems...

- There are 10 Tier 1 sites - as of today we have shipped data to 9 of them
 - It takes a lot of time and effort to set up the components necessary for each site
 - Usually human errors and inefficiencies (on our part and others)
- The sites' components are not in general reliable
 - Storages are inaccessible, disks fill up, grid certificates expire, hardware/software upgrades etc etc...
- Castor @ CERN gave some problems...
- But the Grid middleware from LCG is stable and gives good enough performance
- In summary the data transfer needs constant babysitting and we're a long way from fully functional Grids able to handle our data rates.
- For details see
 - <https://uimon.cern.ch/twiki/bin/view/Atlas/DDMServiceChallenge3Progress>

Some results

Transfers CERN - Tier 1 centres in the last week
Average throughput per hour

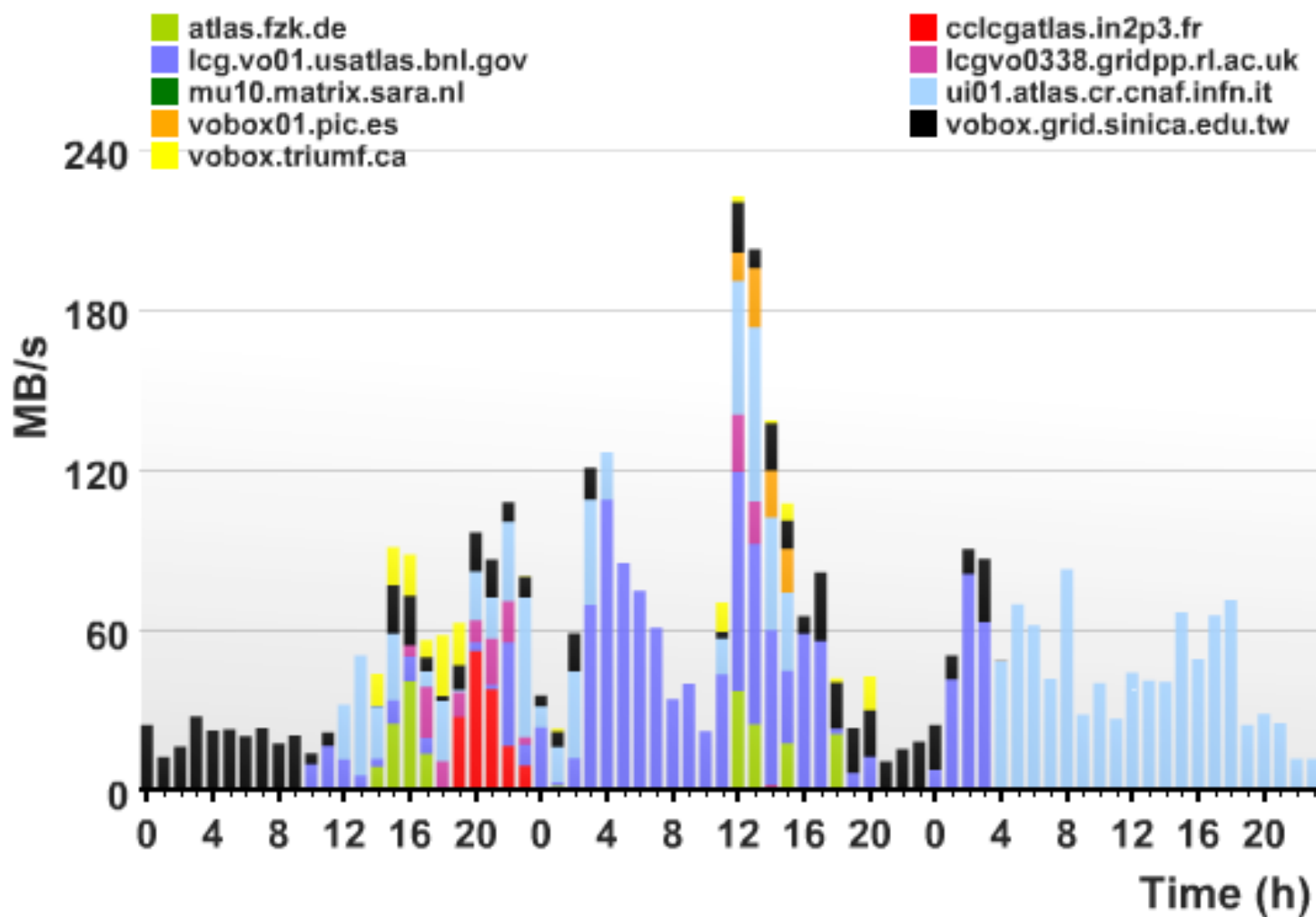
The current time at CERN is 20:42:43 Mon 07 Nov 2005



Peak Throughput 220 MB/s on 16/12

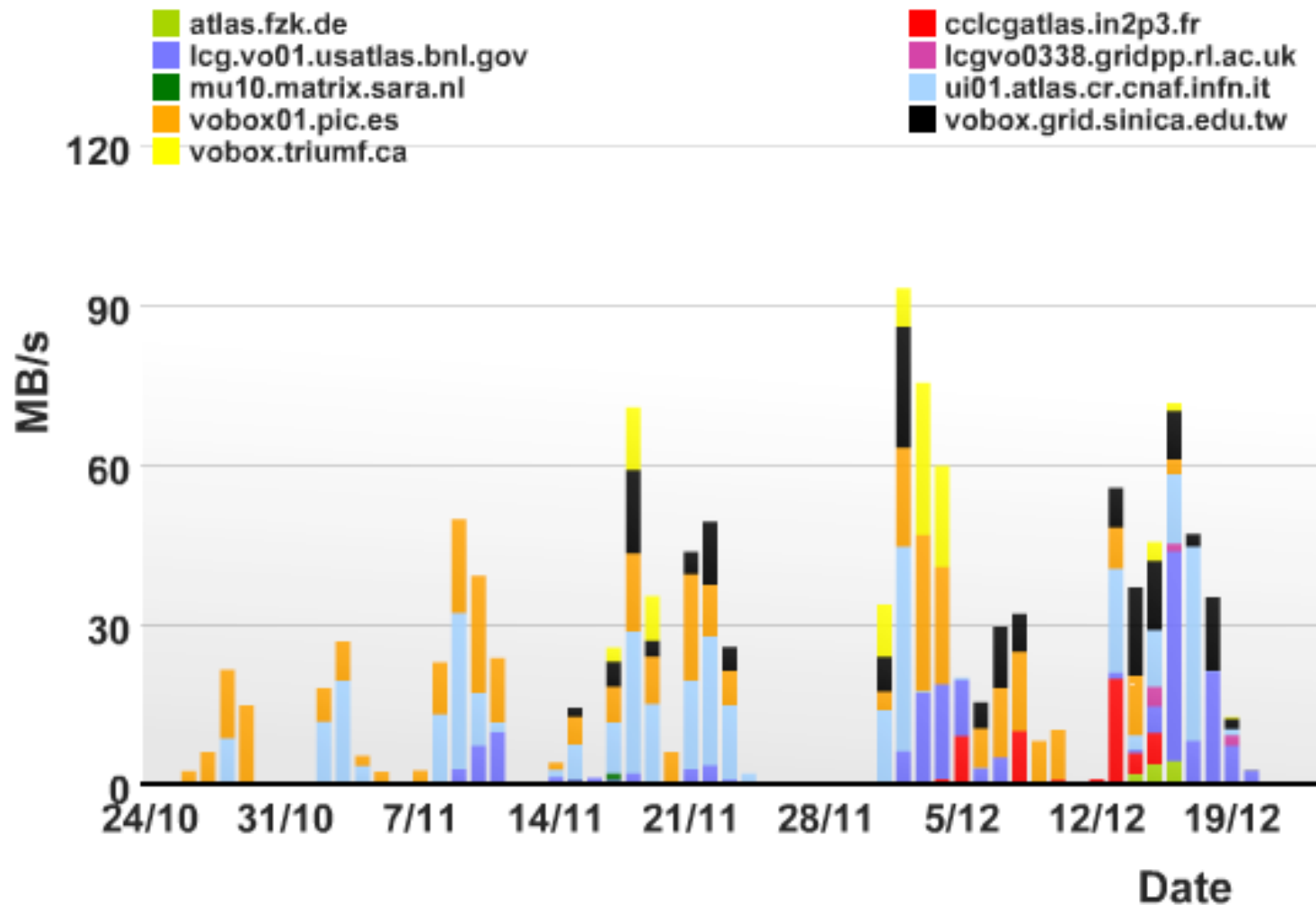
Data throughput from 15/12/2005 0:00 to 18/12/2005 0:00

Tier 0 - Tier 1 Transfers



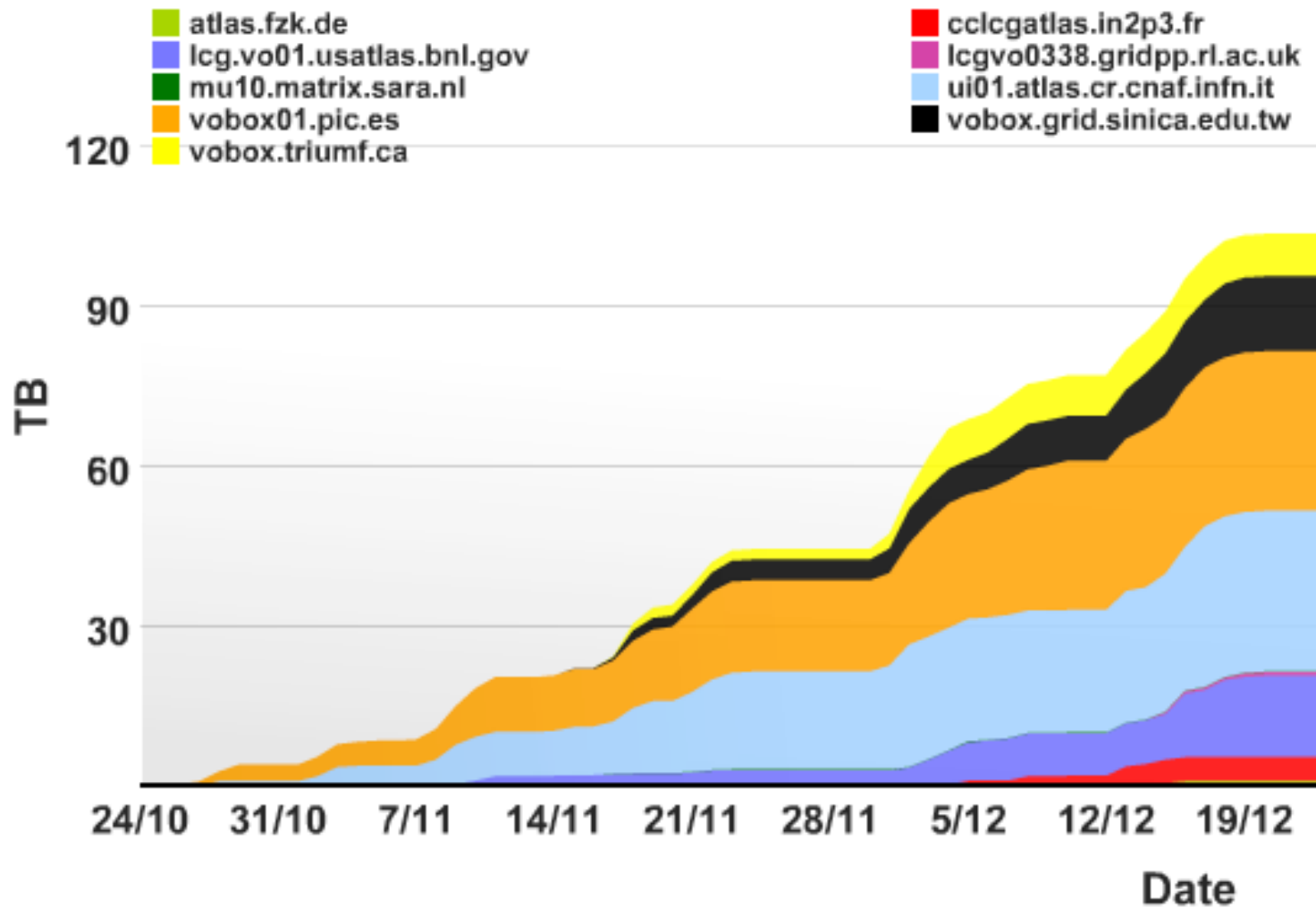
Overall results

Transfers CERN - Tier 1 centres Average throughput per day



Overall Results

Transfers CERN - Tier 1 centres Total cumulative data transferred



Summary of Tier 1 sites (in order we used them)

- CNAF - VOBOX deployment
- PIC - VOBOX deployment
- BNL - VOBOX deployment
- ASCC - VOBOX deployment
- Triumpf - used for testing central deployment model
- SARA - used for testing central deployment model
- Lyon - VOBOX (managed by Lyon personnel)
- FZK - used for testing central deployment model
- RAL - used for testing central deployment model
- NorduGrid - interaction with DDM unclear..

Progress...

- The plan earlier in 2005 was to keep ramping up to 100% but this turned out not to be possible with the current Grid T0-T1 infrastructure (limited to ~200 MB/s)
 - We achieved 220 MB/s (for one hour) but the maximum sustained for one day was 90 MB/s
- Our DQ2 software seems to work as required
 - Most problems with integration of "Grid" and "storage" middleware (srm-dCache; srm-Castor) at the sites.
- We meet throughput targets at various points
 - But not consistently sustained
 - Other constraints on team running the exercise meant periods where nothing ran
- We realise communication with sites is not ideal!
 - We are trying to improve...

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
 - 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
- 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

- These tests run at first independently, then concurrently
- Timescale between April and October 2006
 - Tier-0 tests in blocks of increasing scope
- Every test must be preceded by a period of "service preparation" tests
 - Checking that the underlying infrastructure is working properly

ATLAS SC4 Requirements

- Active FTS channels between all sites
 - FTS useable by anyone with a valid certificate
- SRM "baseline WG version" deployed everywhere
- Disk-only areas at all SEs
 - No sudden migration of files to tape
- Agreed (and secure) way to deploy experiment's services
- Full implementation of VOMS groups and policies for job submission and data management
- Queues with different priorities for production and analysis jobs
- Stability of middleware and service infrastructure
 - "production quality"
- Services run as services
 - No sudden maintenance or upgrade interruptions