



CERN

European Organization for Nuclear Research  
Organisation Européenne pour la Recherche Nucléaire



# EOS/LST 2010

# CERN

## IT/ATLAS

Presenter: Andreas Joachim Peters

Acknowledgements for participation, help, contributions & discussions to IT-DSS & ES Group, XROOT project, the ATLAS LST Members and Stephen Gowdy

## GDB 12.1.2011

[project-eos@cern.ch](mailto:project-eos@cern.ch)

CERN IT Department  
CH-1211 Genève 23  
Switzerland  
[www.cern.ch/it](http://www.cern.ch/it)



# Demonstrator Goals

IT

ATLAS

develop and evaluate potential of EOS storage prototype towards

- efficiency
- tunable reliability
- lower operational costs

Reliability  
Flexibility  
Efficiency  
Simplicity

integrate as grid site into standard framework

gain operational experience & evaluate efficiency running

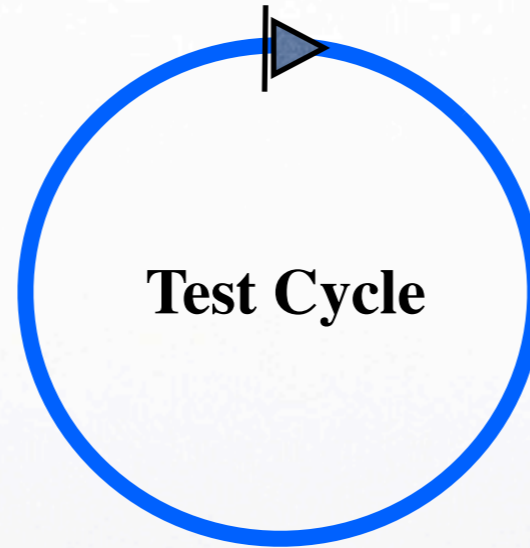
- production jobs
- analysis jobs



# Phases & Coordination

## 1 Preparation and pre-testing

- Storage set up and configuration
- ATLAS queues set up and configuration. Uploading test data.
- DDM functional test
- HammerCloud test (submission analysis jobs)
- Storage configuration tuning (if needed)



## 2 Tuning

- Set up ATLAS Grid site
- Upload ATLAS data, conditions & meta-data
- Run large scale HammerCloud test for several days
- Final tuning

## 3 Test Running

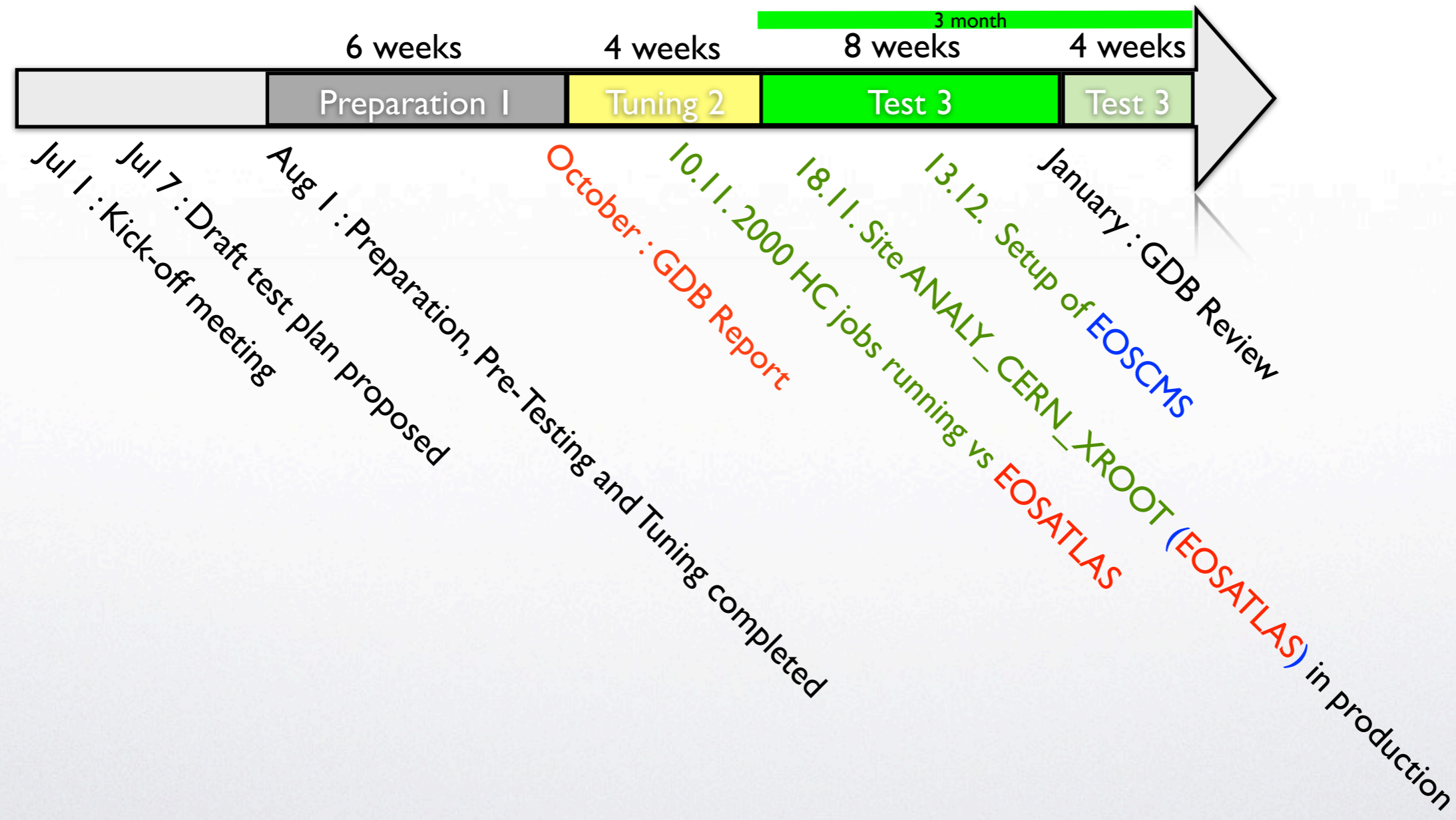
- Declare site as 'ATLAS production and analysis' grid site, allow production and analysis jobs brokering to the site

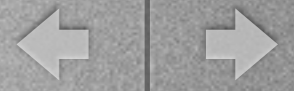
## Operations and Coordination

- Overall coordination will be done by IT-DSS Group Leader and ADC Coordinator
- LST2010 day-by-day coordination will be performed by 2 coordinators (one from IT-DSS, one from ADC)
  - Daily meetings during Preparatory and Tuning steps
- ATLAS Computing shifts will follow site issues during Production step. The procedure may be different from the other sites
  - Wash-up common meeting after each phase completion



# Timescale & Milestones





# EOSATLAS Usage

- Used as grid site in production by ATLAS with 500 job slots and 33 disk server [since 18.11.2010]
  - **Dataset Import** (2 replicas)
    - via gridFTP-Gateway/SRM-BestMan
    - ~60% preload, ~40% dynamically loaded by PD2P (TI-TI algorithm was tuned using EOS)
  - **Job Input DB Releases** (6 replicas)
    - via xrdcp download
  - **Job Input Data**
    - via xrdcp download
  - **Job Output Data** (3 replicas)
    - via SRM/gridFTP

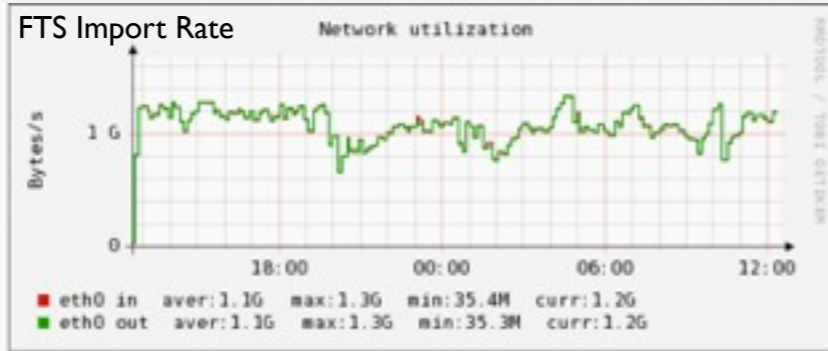


# EOSATLAS Usage

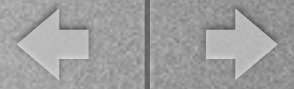
1.11.2010-05.01.2011 66 days

File Opened Read	Data Read	File Opened Write	Data Written	Logical Space	Inodes max used
1.9 M	2.12 PB	3.2 M 1.6 M*	3.55 PB 1.77 PB*	650 + 40 TB	0.9M Files 31k Directories

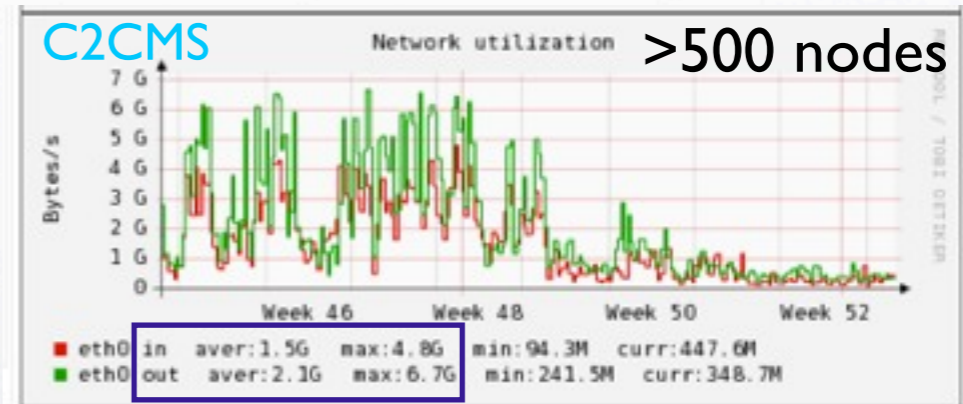
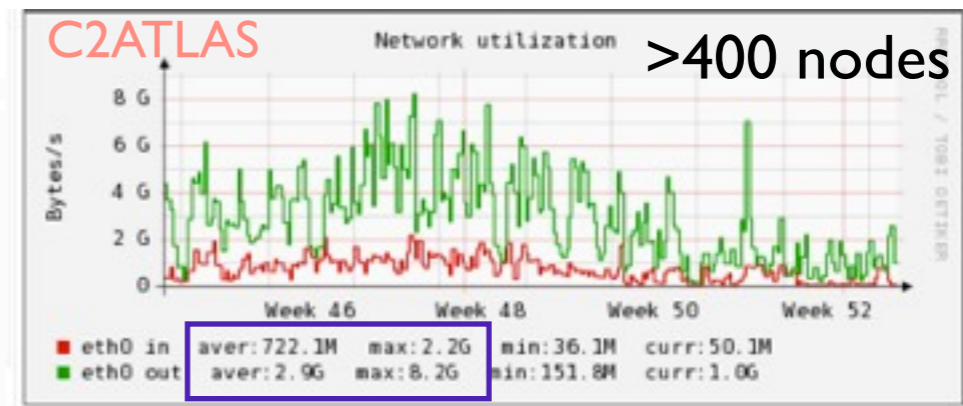
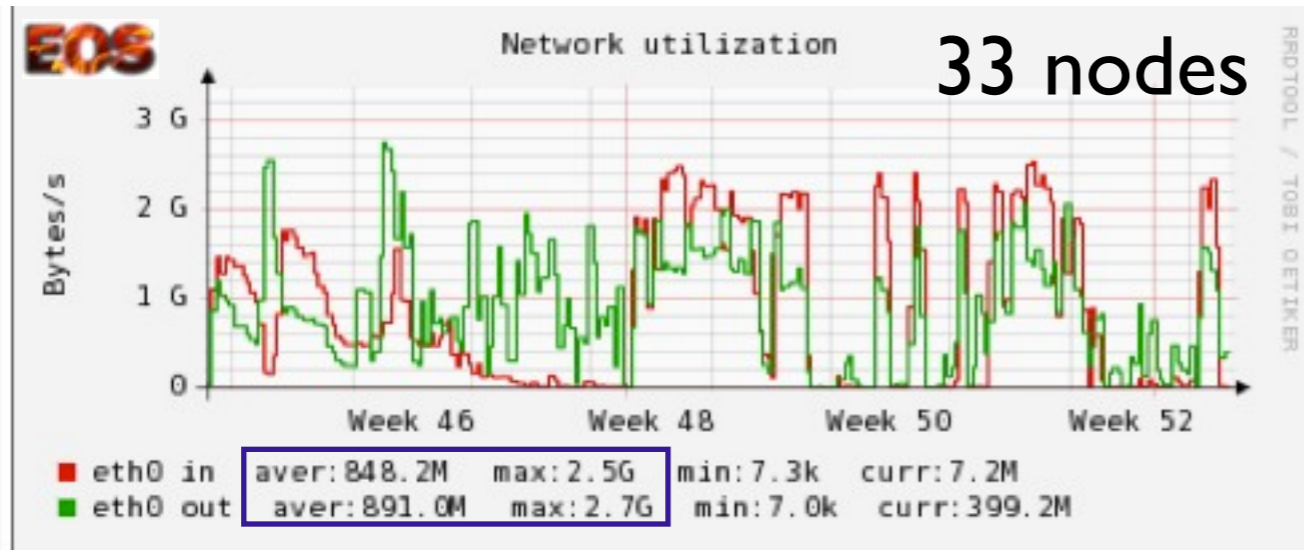
\* incl. replication factor 2



830k FTS transfers  
nominal import at **1.1-1.3 GB/s**



# EOSATLAS Usage



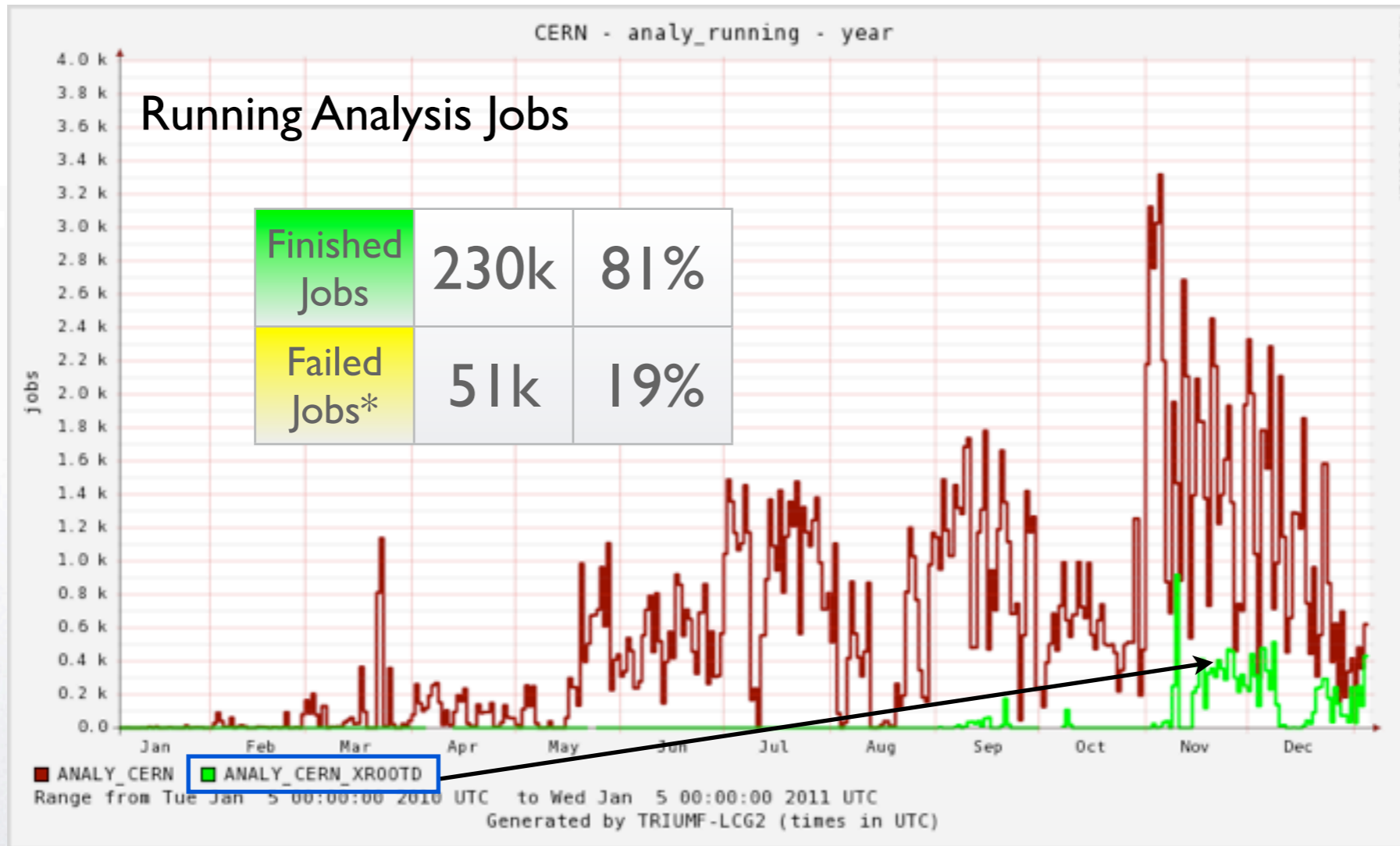
## Scale Comparison

EOS	C2ATLAS	C2CMS	C2ALL
Nodes	8%	7%	2%
IO Read	30%	42%	11%
IO Write	110%	56%	28%

EOS Server well tested:  
avg. running at 25% of available IO bandwidth



# EOSATLAS Analysis



\* failed jobs dominated by application/user induced errors





# Handling of Operational Problems & Hardware Failures

## GGUS

5 Tickets found

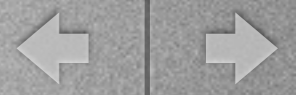
Ticket-ID	Type	Concerned VO	Notified Site	Resp. Unit	Status	Date	Last Update	Info
<a href="#">65826</a>	Team	atlas	CERN-PROD	ROC_CERN	in progress	2011-01-03	2011-01-05 13:49	failed transfers from PIC_SCRATCHDISK to CERN-PROD...
<a href="#">64939</a>	Team	atlas	CERN-PROD	ROC_CERN	verified	2010-12-03	2010-12-09 17:00	CERN-PROD_EOSDATADISK transfer failures
<a href="#">64247</a>	Team	atlas	CERN-PROD	ROC_CERN	verified	2010-11-15	2010-11-17 07:48	many file transfers failing from NDGF to CERN-PROD...
<a href="#">63926</a>		atlas	NDGF-T1	NGI_NDGF	verified	2010-11-04	2010-12-05 17:08	gridftp transfers NDGF_T1_DATADISK to CERN-PROD_EO...
<a href="#">63901</a>	Team	atlas	CERN-PROD	ROC_CERN	solved	2010-11-04	2010-11-30 09:13	Transfer errors for CERN-PROD_EOSDATADISK

SRM & gridFTP ports had to be opened in firewall for external user and NDGF access

ENOSPACE: FTS transfers scheduled on full or unavailable disks

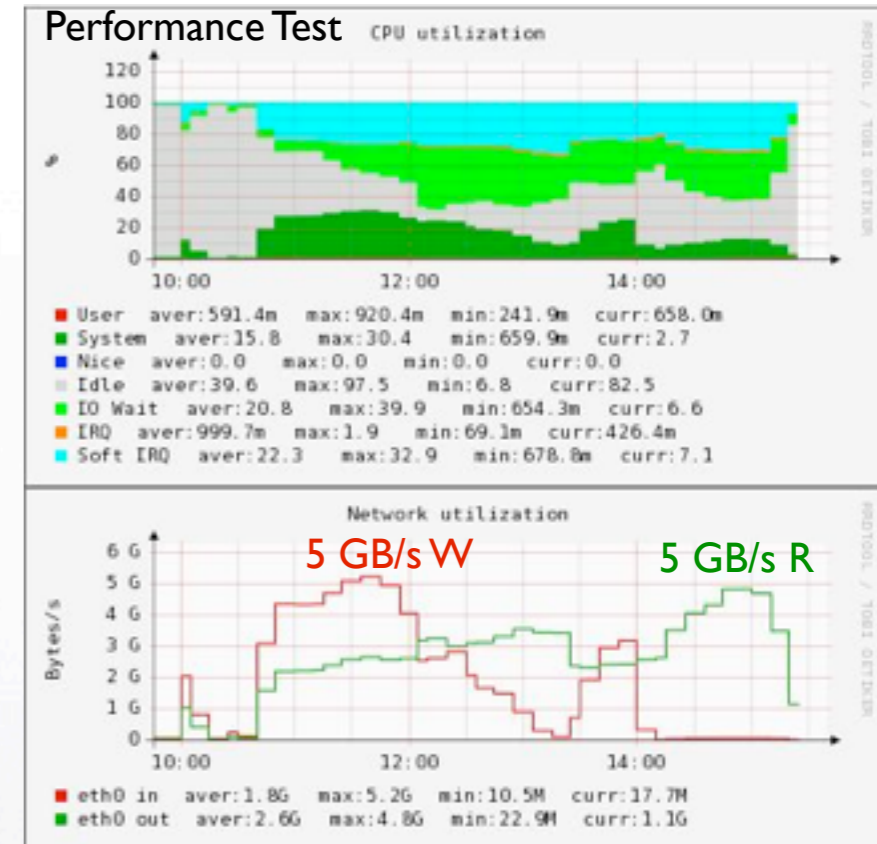
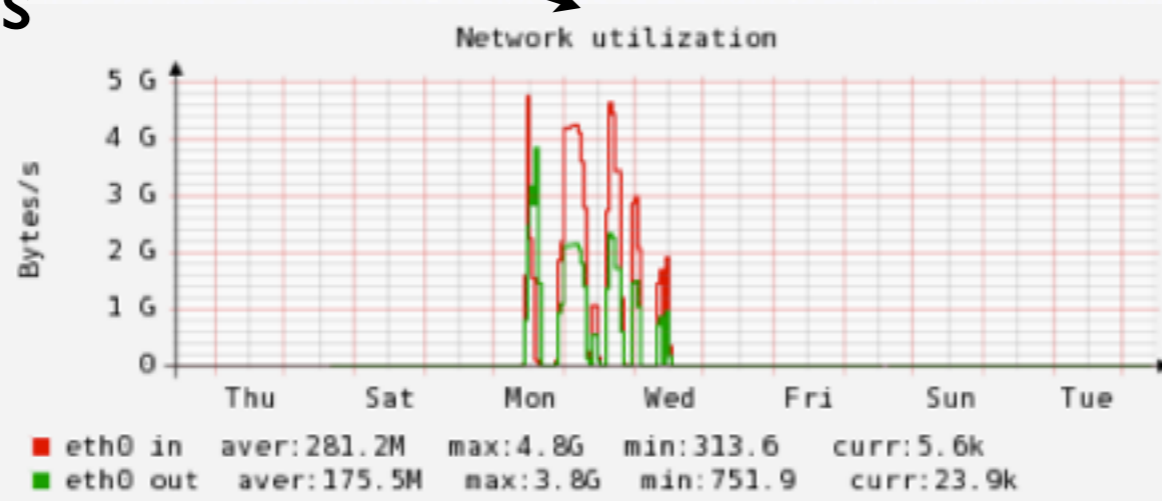
## HW Failures

25 hardware failures tracked (~20 disks to be exchanged, few backplane and one power supply failure).



# EOS Setup for CMS

- 10 x 10 Gbit Server 420 TB
  - xroot only access (no SRM)
- Import from C2CMS via xrd3cp
  - 120 TB (x2)
  - 36k files



Tests have just started ...



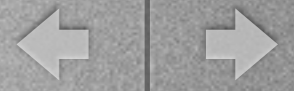
# ATLAS Feedback

- ATLAS expressed satisfaction about stability, efficiency & performance of EOS as grid site storage
  - besides peculiarity of CERN cloud no extra configuration/patches
    - 'standard' protocols SRM, gridFTP & xroot + X509 authentication
  - high-speed wide area import from world wide distributed sources at 1.1 GB/s with internal realtime replica
    - EOS subscribed automatically to 'hot' datasets worldwide via PD2P
- ATLAS asked to extend test to run two more month with PD2P subscription



## Next Steps, Enhancements & future directions

- **GDB/MB feedback**
- **Evaluate analysis with remote access**
  - via remote access protocol (xroot)
  - via POSIX and FUSE mount in job sandbox - requires lxbatch deployment  
(Buff. Cache/Kernel read ahead)
- **Capture initial operational experience and re-iterate on EOS software architecture & implementation**



# Mid-/Longterm Future

- CERN CC has a lot of existing capacity which can be exploited with adapted software on client & server side (experiment usage/framework & storage system)
  - In the grid environment experiments use only a reduced common subset of storage functionality
  - Aim to significantly increase the efficient utilization of existing resources
  - Tape should be linked close to EOS as a big space provider
  - Investigate other deployment models which may lead to further HW consolidation and TCO savings
    - disk server participates as batch node
    - batch node participates as disk cache



# Summary

- EOS Demonstrator is running successfully since two months in a production environment under high load
- Valuable input collected during demonstrator test about functional and operational requirements & needs for the future
- EOS has shown some of its potential to achieve high quality service with increased resource efficiency
  - monitoring to further quantify its benefits has to be put in place
  - ATLAS use case as grid site storage does not exploit fully its capabilities (namespace performance, quota ...)
- Propose to continue the already existing close collaboration with experiments