

Step09 at IFIC

Javier Sánchez
Alejandro Lamas
Santiago González

IFIC – Institut de Física Corpuscular
Centro mixto CSIC –Universitat de València



Step09 activities

<https://twiki.cern.ch/twiki/bin/view/Atlas/Step09>

- **Dates**

- May 25 - 31 Setup
- **June 2 - 14 Running** (N.B. 1 June is a holiday at CERN)

- **Tier-2 Activities**

- Tier-2s will participate in **MC simulation** and in **user analysis** components of the challenge.
 - **User analysis** in the EGEE clouds will happen through both **WMS job submission** and through **pilot jobs**.
 - supporting **PanDA** analysis, to have analysis queues (named **ANALY_MYSITE**)
 - To facilitate this Tier-2s should configure their batch systems to accept the **/atlas/Role=pilot** VOMS role and set up the following shares:

| <u>Activity</u> | <u>VOMS Role</u> | <u>ATLAS Batch System Share</u> |
|---------------------------|------------------------|---------------------------------|
| Production | /atlas/Role=production | 50% |
| Analysis (pilot based) | /atlas/Role=pilot | 25% |
| Analysis (WMS submission) | /atlas | 25% |

- **Data Distribution**

- Total data volume to distribute to Tier-2s is **112TB** during the 2 week test. This will be distributed through the "functional test" framework according to the shares here (**50% IFIC, 25% IFAE, 25% UAM**):
 - http://atladcops.cern.ch:8000/drmon/ftmon_TiersInfo.html

- **Tier-1**

- Tier-1s should ensure they define file families for this project path in their tape system so that tapes can be effectively recycled at the end of the exercise. (N.B. this project will produce data in both the **ATLASDATATAPE** and **ATLASMCTAPE** areas.)

Step09 feedback

<https://twiki.cern.ch/twiki/bin/view/Atlas/Step09Feedback#T2>

- Main information requested for sites providing feedback to help with the evaluation of STEP09. Sites should of course feel free to be creative and report anything they think is interesting or useful.
- Tier2
 - **Data Distribution**
 - See https://twiki.cern.ch/twiki/bin/view/Atlas/Step09Logbook#So_long_and_thanks_for_all_the_d for sites who had problems (next slide)
 - Did your site have **problems with getting the large volume of data** distributed during STEP09?
 - Was your infrastructure adequate?
 - Did you have enough headroom to cope with backlogs quickly?
 - Did other activities, specifically user analysis, interfere?

Step09 feedback

| Site | Completed Datasets | Missing Datasets | Comments |
|---------------------|--------------------|------------------|---|
| ALBERTA-LCG2 | 3000 | 394 | SAN failure 12 June (ongoing) disabled all write access to dcache pools |
| DESY-HH | 6413 | 362 | Backlog from FZK <code>gridftp</code> problems last week, then catching up well but disk filled up on Saturday |
| GRIF-LAL | 2214 | 1042 | Disk filled up - was removed from SS |
| IFAE | 3273 | 117 | Disk filled up |
| IFIC-LCG2 | 6168 | 550 | Disk filled up |
| INFN-T1 | 13602 | 65 | Outstanding issue with transfer timeouts from TRIUMF |
| RO-02-NIPNE | 559 | 790 | Taken offline because of transfer errors |
| RO-07-NIPNE | 1029 | 313 | Taken offline because of transfer errors |
| RRC-KI | 6061 | 652 | SRM reliability was poor in the final catchup period |
| RU-PROTVINO-IHEP | 1416 | 564 | IHEP currently has only 100 Mb/sec link. Low transfer efficiency over STEP. |
| SFU-LCG2 | 9483 | 140 | Good throughput (800 Mbps) but gap in activity of around 6 hours seen 11 June ; could have helped complete subscriptions |
| TR-10-ULAKBIM | 1566 | 1161 | Misconfigured FTS. Some very strange FTS errors seen during final catchup weekend |
| UKI-SOUTHGRID-RALPP | 3360 | 34 | Getting <code>TRANSFER_MARKERS_TIMEOUT</code> and <code>gridftp_copy_wait: Connection timed out</code> on FTS channel, throttles throughput |
| WISC | 4488 | 12 | |

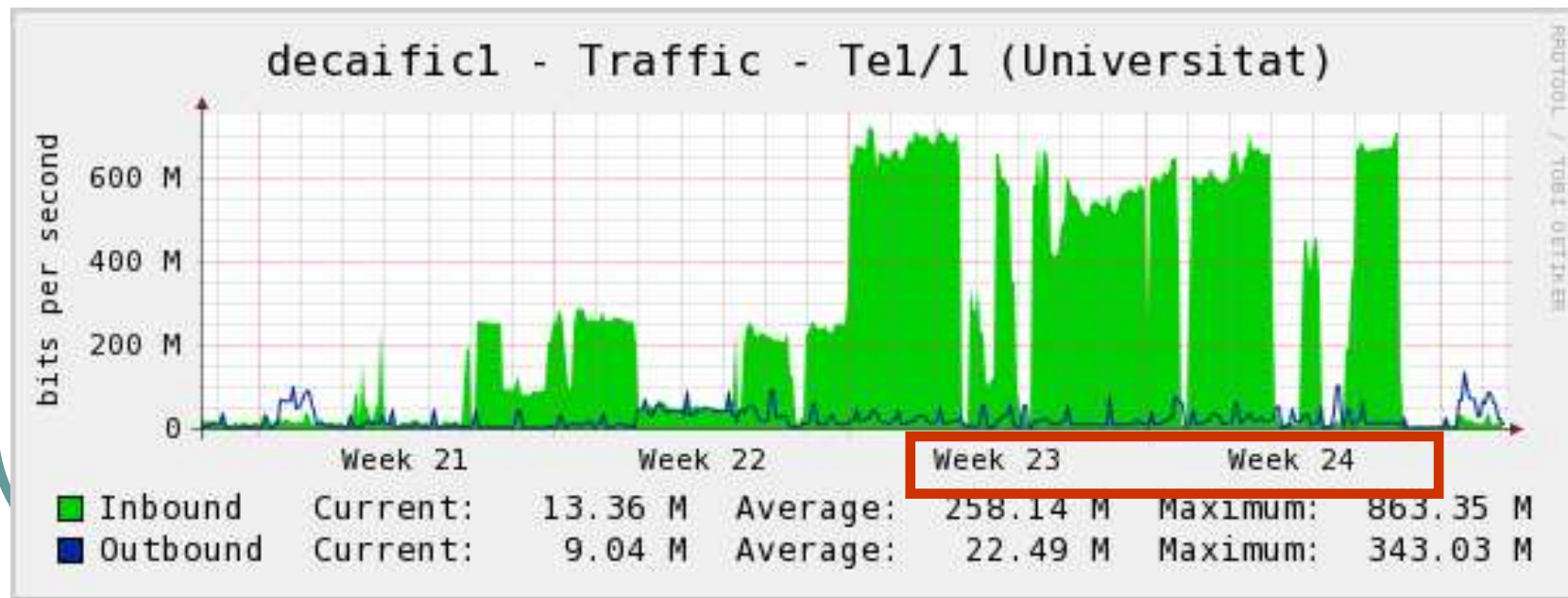
Step09 feedback

- **Workload**

- **N.B.** Panda **pilot based analysis** used your site's ANALY queue local mover to copy data to the worker node; the **WMS based analysis** used a mixture of the file stager (also a copy to worker node method) and **local LAN protocol** (known as DQ2_LOCAL, resolved to dcap, rfio or xroot). Because of problems with the analyses producing too much output the file stager analyses were switched off in week 2.
- **Number of jobs run each day for production, WMS analysis and pilot analysis.**
- **Efficiencies** in the batch system for each type of job.
- The **throughput** (running jobs x job efficiency) **of analysis jobs** is a useful metric of the capacity of your site to perform analysis
- Do you know what the **limitations are in your site infrastructure** in delivering successful analysis results (network, i/o on servers or WNs, etc.).
- **Information about fairshares:**
 - The request was for https://twiki.cern.ch/twiki/bin/view/Atlas/Step09#EGEE_Tier_2_resources
 - However, we think there were significant problems with delivering enough analysis pilots in some clouds (<http://tinyurl.com/nqnbm9>, slide 8) and many sites introduced caps to manage the analysis during the STEP challenge, which was more useful than letting things run wild. This probably overwhelmed any fairshare effects, but if you believe there were problems you might wish to comment.
- What specific tests would help you to improve analysis performance over the summer?
- A **description of your network layout and disk servers** might be useful for other sites.

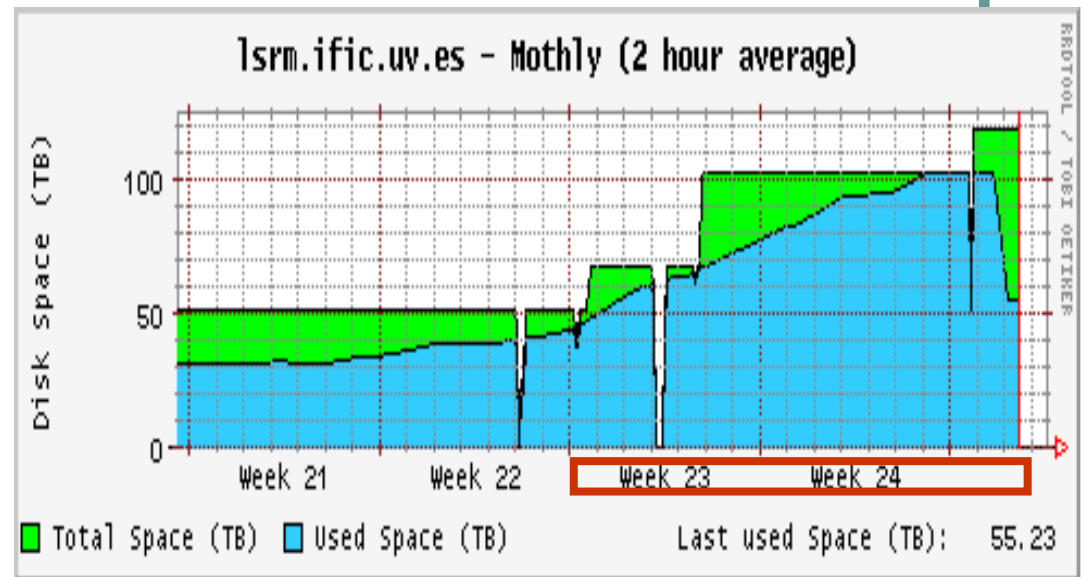
Data Transfer

- Data distribution in the week previous to the step09 exercise showed a bottleneck in the gridftp server.
- A new machine had been deployed on Sunday May 31st. Transfer bandwidth increased from 200 Mbps to 600 Mbps.
- During the exercise (weeks 23rd and 24th) data transfer rate was steady during long periods. No problems were observed in the network infrastructure neither local or remote.



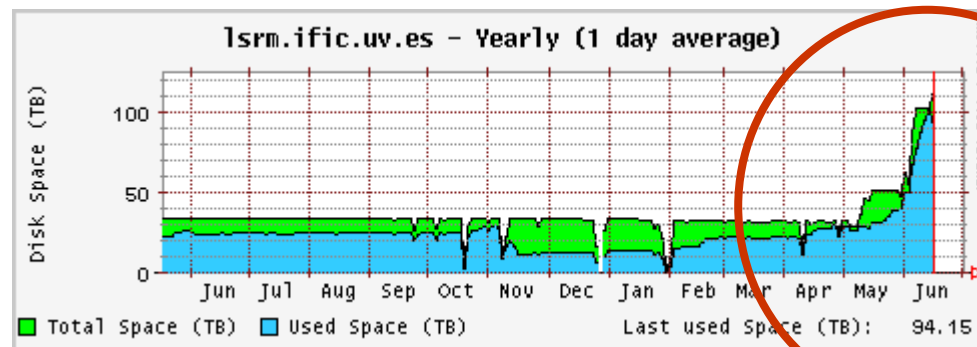
Data Transfer

- From Jun3 1st to 14th data size transferred to our SE is:
 - 57 TBytes incoming (112TB total, 50% to IFIC ~56 TB)
 - 1 TBytes outgoing
 - distributed in:
 - 44 TBytes /lustre/ific.uv.es/grid/atlas/atlasdatadisk/step09
 - 8.4 GBytes /lustre/ific.uv.es/grid/atlas/atlasproddisk/step09
- During the exercise **disk went full two times**.
 - **More space was added** on the fly with the deployment of two new disk servers (17 TB + 34 TB) **to reach 102 TBytes** total capacity installed.
 - Space tokens were updated to reflect the used and available space but:
 - **ATLAS tools seems to ignore the values and continue transfers even when no space available was published for a specific token.**



Data Transfer

- Data occupancy during this exercise increased exponentially.
 - A **quick reaction** and the flexibility to **add more servers easily to LUSTRE** file system lead us to absorb the data without major problems.



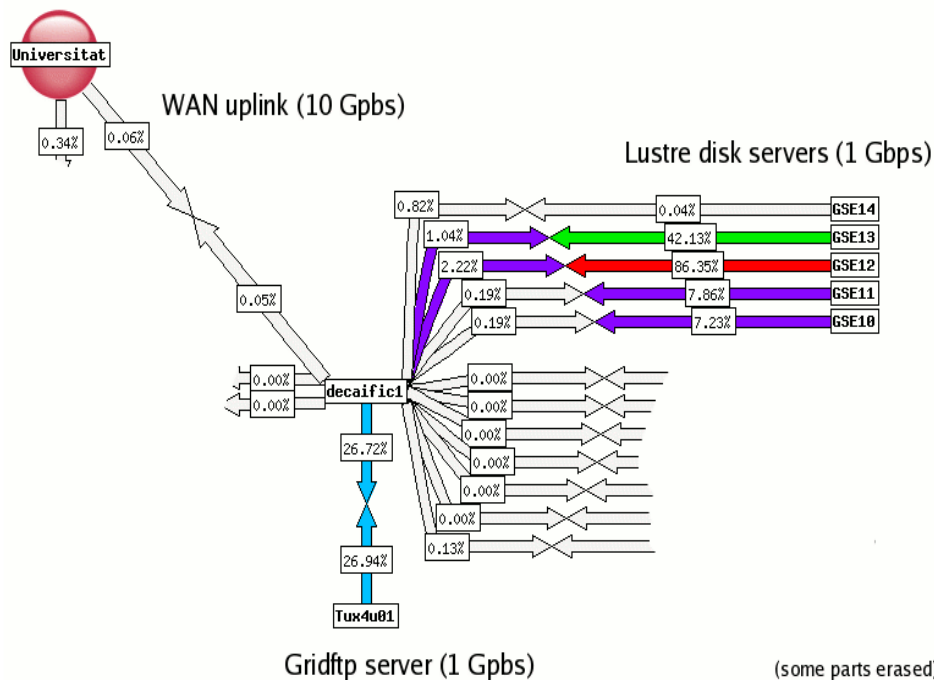
Data Transfer

- Worker nodes accessed the data using local posix I/O (read only) and gridftp protocols depending on the jobs policy.
 - WNs and disk servers are connected to Cisco switch at 1 Gpbs.
 - In the present configuration:
 - WNs share 1 Gpbs in groups of 8
 - disk servers share a 1 Gpbs in groups of 2.
 - This will change in the near future to have a dedicated 1 Gpbs for every WN and a 2 Gpbs for every disk server.

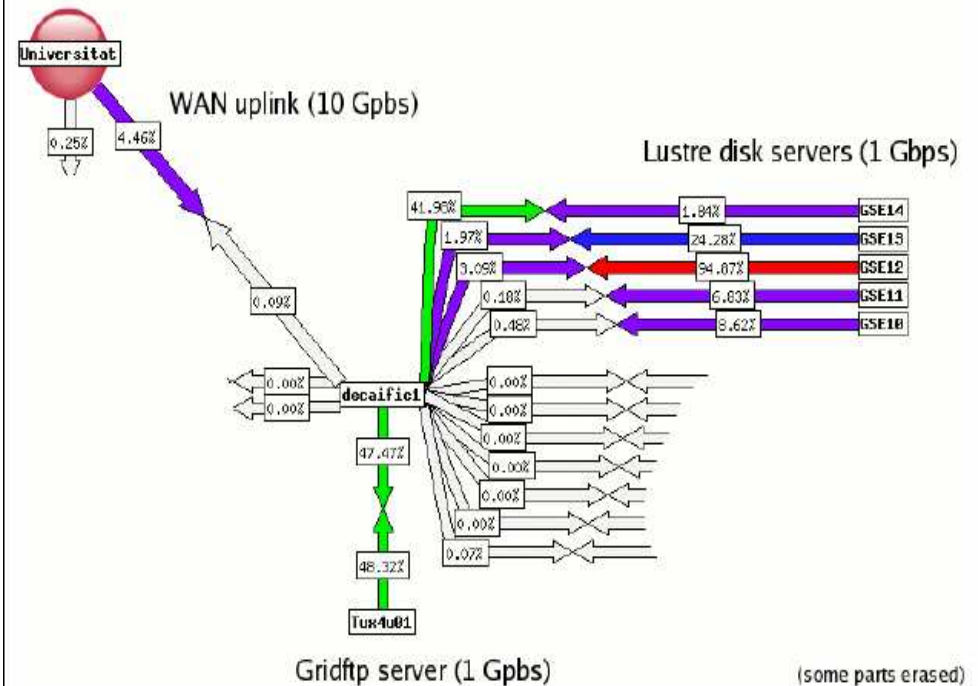
Data Transfer

- High load have been observed sometimes but performance was reasonable

Bandwidth graph. LAN Posix I/O (Lustre) + GridFTP



Bandwidth graph. LAN Posix I/O (Lustre) + WAN GridFTP



Job processing

- Pilot role was added in the VOMS mapping
 - but no proper scheduling policy could be setup in time, so user jobs coming from panda and WMS were treated equally.
 - In order to achieve this, a new account pool had to be setup but it was not easy with the present configuration tool (*quattor*) so we decided not to split the mappings.

- Fair share at June 8th:

| Group | % | Target |
|------------------|-------|--------|
| Atlas users | 43.23 | 50 |
| Atlas production | 56.49 | 50 |

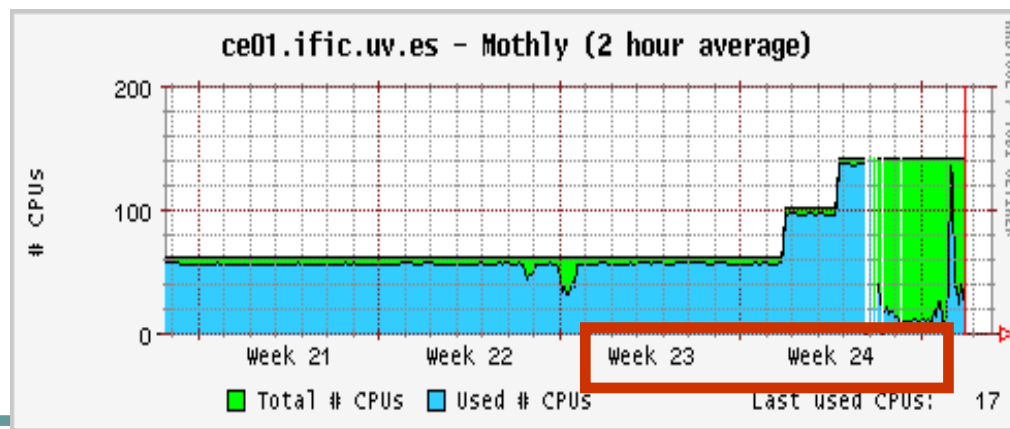
- Fair share at June 16th:

| Group | % | Target |
|------------------|-------|--------|
| Atlas user | 26.34 | 50 |
| Atlas production | 73.13 | 50 |

- This unbalance is under investigation.

Job processing

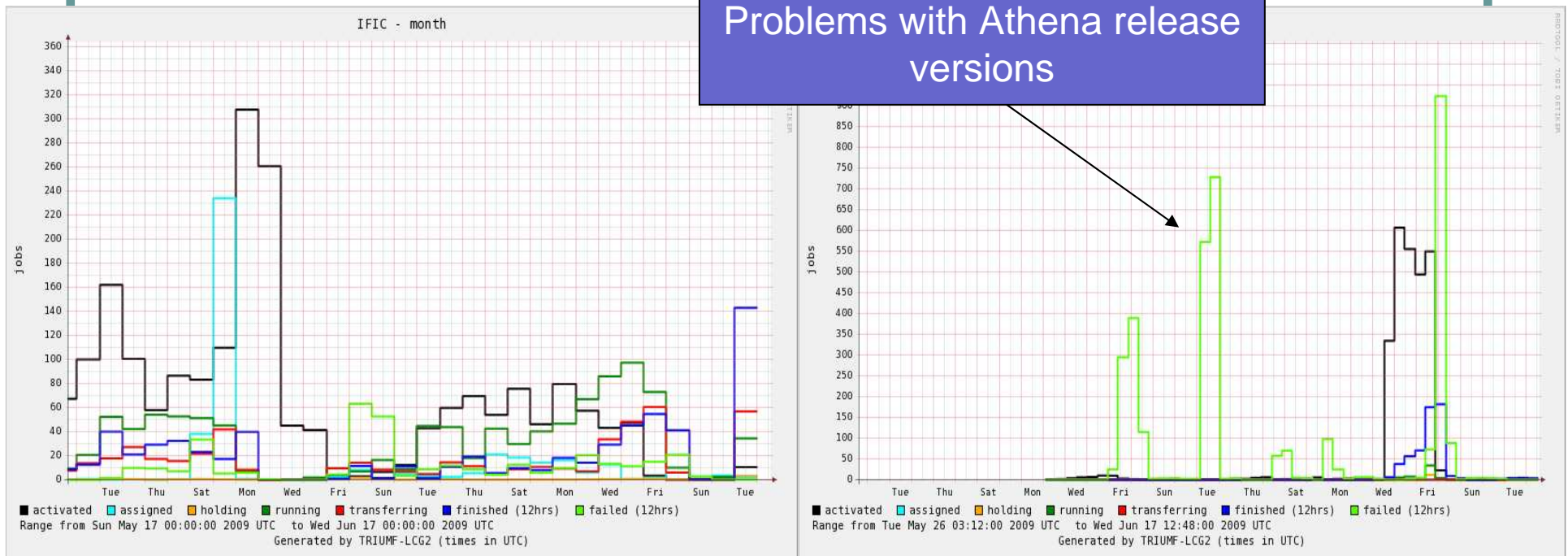
- Occupancy was high and we added 40 + 40 cores more in the last week (24).
- We had some instabilities in the CE during last Sunday.
 - We think that the number of cores we added stressed the CPU power at the CE (Pentium D CPU 3.20GHz). An upgrade is foreseen in the coming weeks (hardware is already bought).



Job processing

- Job statistics from ANALY and IFIC queue from 17th May to 17th June.

Problems with Athena release versions



Job processing

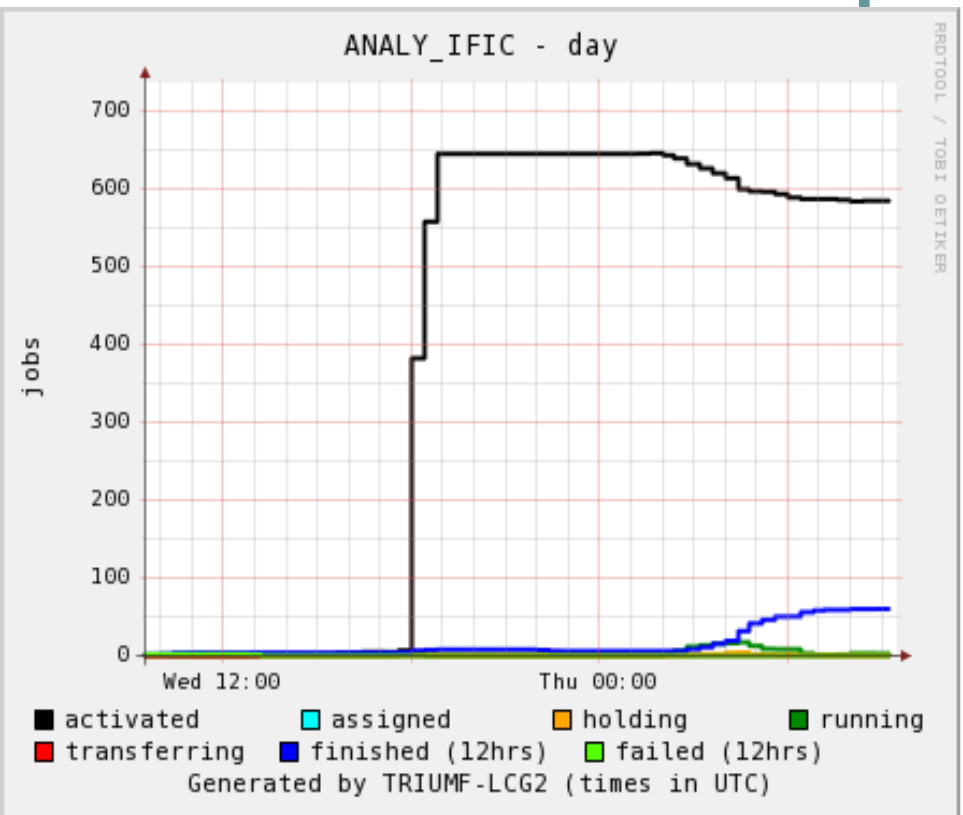
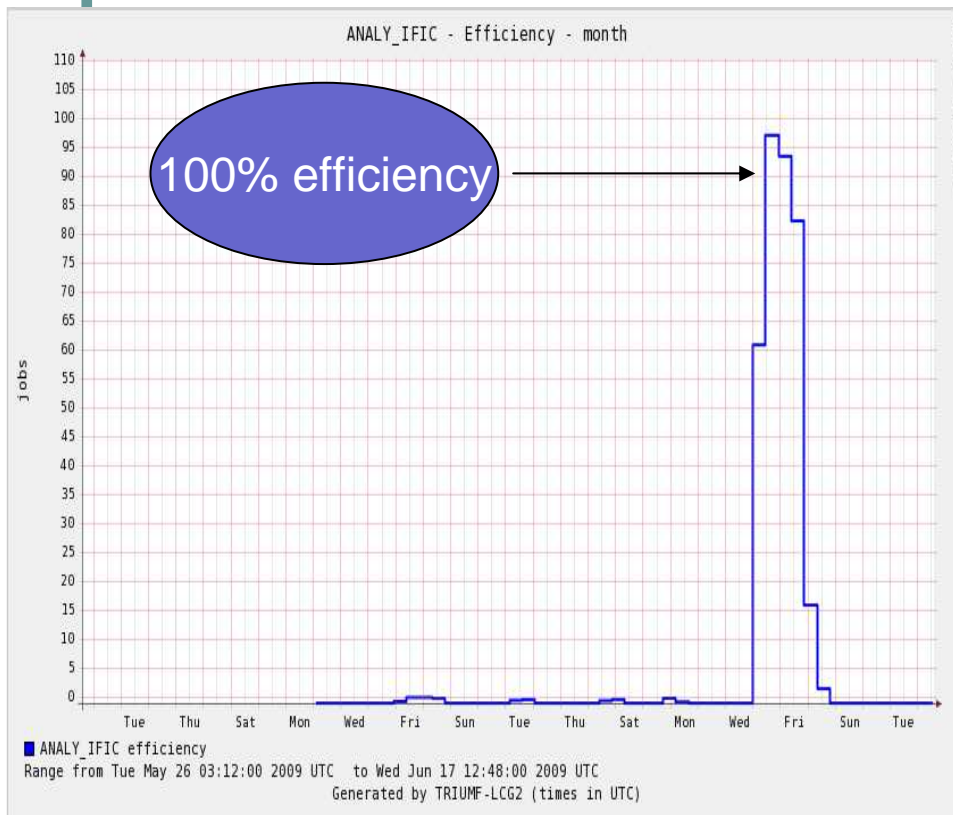
- In the first week we had **a lot of problems with panda jobs** trying to use an **non existing Athena version** at our site.
 - The **release** was **installed** during the **last week**.
 - The problem to install the version seems not to be related with our setup but with the procedure itself (installation bug).
 - For that reason, our **efficiency for pilot jobs was around 21% (ANALY_IFIC)** while for **WMS jobs (IFIC-LCG2_MCDISK)** was **around 82%** as is shown in next table (17th June).

| ES | SUBMITTED | RUNNING | COMPLETED | FAILED | Efficiency | #Files Expected | #Files Processed |
|------------------|-----------|---------|-----------|--------|------------|-----------------|------------------|
| IFIC-LCG2_MCDISK | 62 | 96 | 1130 | 299 | 0.82 | 8105 | 8105 |
| ANALY_IFIC | 443 | 109 | 455 | 1691 | 0.21 | 1506 | ? |

Statistics taken from: <http://gangarobot.cern.ch/st/status.html>

Job processing

- After installing the Athena versions used for panda jobs our **pilot job efficiency** raised to **17% in a few minutes** being around **100% in the last days**.

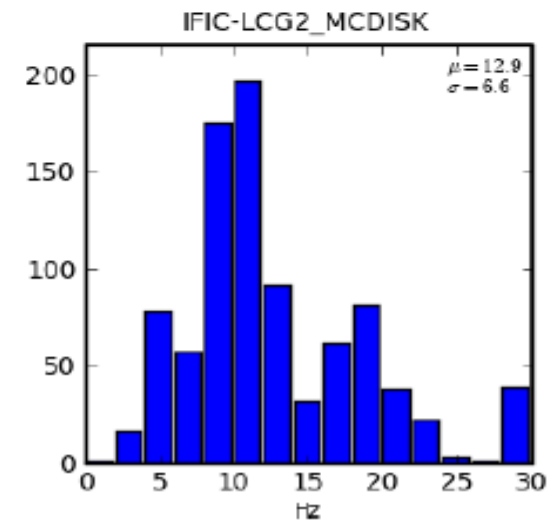
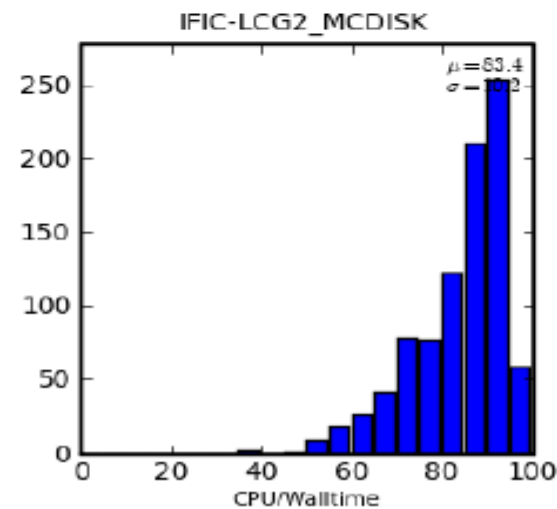
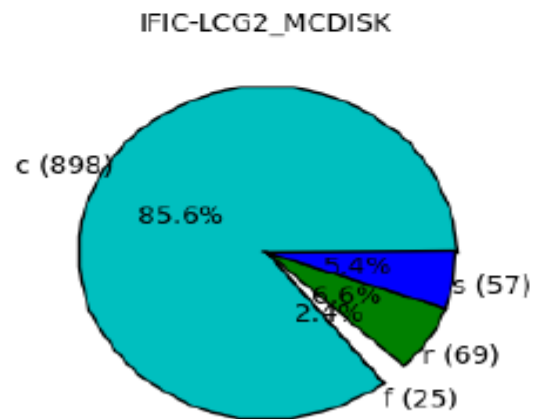
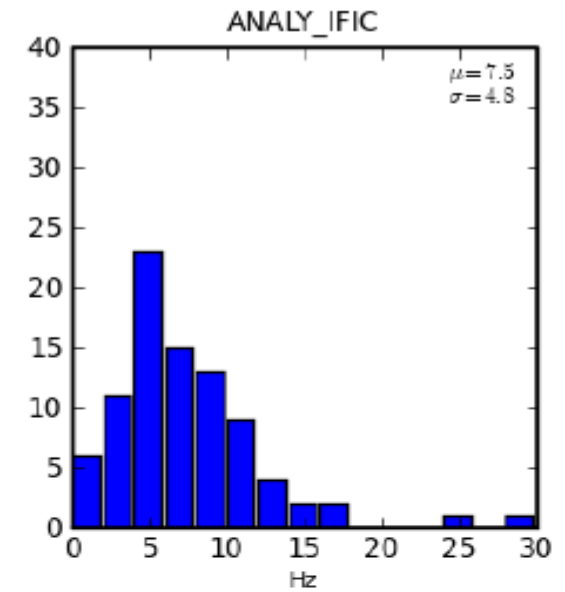
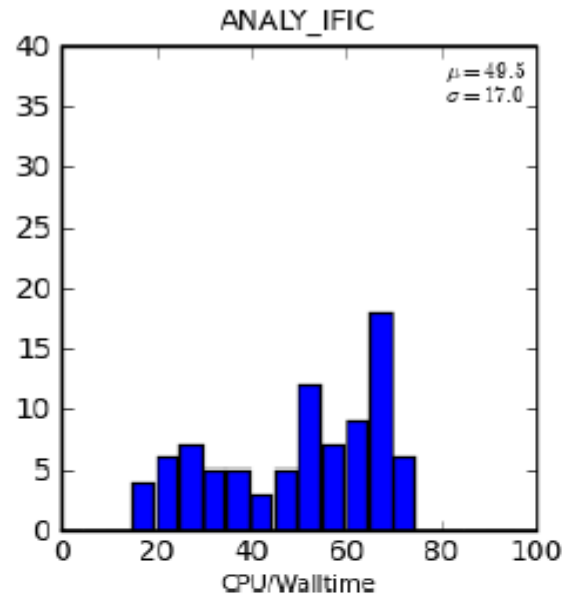
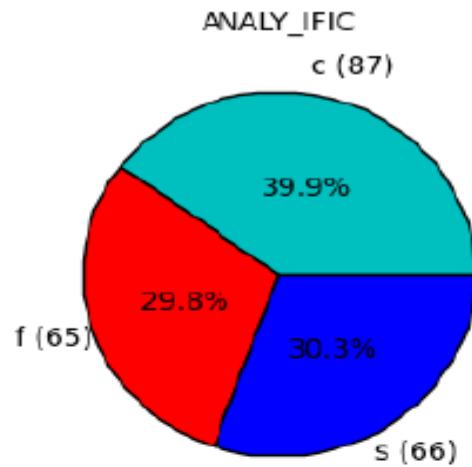


Job processing

- Next figures (next slide) are showing:
 - IFIC efficiency
 - CPU Wall/time
 - CPU Percent Utilization
 - Events per second (Hz)
 - This is the event rate for the Athena execution only, i.e. the denominator is the time from Athena start to finish.
- Overall efficiency and details statistics for the Distributed Analysis test at IFIC:

| Site | Job eff. | CPU/Wall | Events (Hz) | SW setup (s) | Input/Output (s) |
|----------------------|----------|----------|-------------|--------------|------------------|
| ANALY_IFIC | 21% | 49.5% | 8 | 290 | 980(82) |
| IFIC- LCG2_MCDISK | 82% | 83.4% | 13 | 380 | 11(33) |

Job processing



Job processing

- Influence on the **input data protocol** on **CPU/Wall time efficiency** using natural storage system protocol was **tested in previous distributed analysis tests**.
- Sites with **POSIX-like access to the file system** (Lustre) **perform better without triggering copies** in the background using lcg tools (file stager).
 - The same results was found in this test for WMS jobs **but not for PANDA pilot jobs**, where the CPU/Wall time efficiency was reduced from 83.4% to 49.5 %.
 - Similar behavior is observed in other sites using Lustre (like LIP-Lisbon), therefore it would be nice to investigate **why using pilot jobs the CPU Wall time efficiency is reduced in a factor 2 for a (our) site**.
 - **BECAUSE:**
 - Panda pilot based analysis used your site's ANALY queue local mover to copy data to the worker node; the WMS based analysis used a mixture of the file stager (also a copy to worker node method) and local LAN protocol (known as DQ2_LOCAL, resolved to dcap, rfio or xroot).

Conclusions

- **Data transfer rate was steady during long periods.**
 - No problems were observed in the network infrastructure neither local nor remote.
- **Disk went full two times.**
 - More space was added to reached 102 TB
 - Quick reaction and flexibility to add more servers easily to Lustre file system
 - Space Tokens were updated but ATLAS tools seems to ignore the values.
- **Our Worker Nodes accessed to the data using local posix I/O (read only) and gridftp protocols depending on the jobs policy.**
 - High load have been observed some times but performance was reasonable:
 - In the present configuration WNs share a 1 Gbps in group of 8 while disk servers share a 1 Gbps in group of 2. **THIS WILL CHANGE SOON.**
- **Pilot role was added in the VOMS mapping but not in the proper scheduling policy.**
 - New account pool had to be setup but not easy with quattor.
 - The fair share between analysis and production was 50% in the first week but 70% production and 30% analysis the second week.
 - Under investigation !!!

Conclusions

- **WN occupancy was high**
 - We added 40 + 40 cores in the last week
 - Instabilities in our CE probably because the high number of cores stressed the CPU's CE.
 - Upgrade in coming weeks!!
- **Problems with pilot jobs in the first week**
 - Non existing Athena release version in our site.
 - The problem to install the version seems not to be related with our setup but with the procedure itself (installation bug).
 - Release installed as soon as possible
 - Efficiency for pilot jobs was 21% while for WMS jobs was 82%.
 - After installing the release pilot job efficiency in last days was 100%.
- **CPU/Wall time for WMS jobs (~82%) is twice for pilot jobs (~21%).**
 - This is because panda pilot based analysis used your site's ANALY queue local mover to copy data to the worker node; the WMS based analysis used a mixture of the file stager (also a copy to worker node method) and local LAN protocol (known as DQ2_LOCAL, resolved to dcap, rfio or xroot).
- **ATLAS Post Mortem**
 - There will be an ATLAS internal post mortem on the afternoon of 1st July (40-R-C10 is booked - preliminary agenda soon)
 - N.B. The WLCG post mortem is 9-10 July, <http://indico.cern.ch/conferenceDisplay.py?confId=56580>