PIC Report - J. Flix [on behalf of PIC team]

Hosted by

PIC
port d'informació
científica

HEPiX Autumn-Fall 2018 / Barcelona

08-12 October 2018

PIC in numbers

October 2018

CPU: 108 kHS06
Disk: 9.2 PB
Tape: 25.0 PB

Spanish WLCG Tier-1 centre → **~85% of resources**
→ Offers 5% of Tier1 data processing for CERN's LHC detectors ATLAS, CMS and LHCb

ATLAS Tier-2 and **ATLAS data analysis facility** → **~10% of resources**

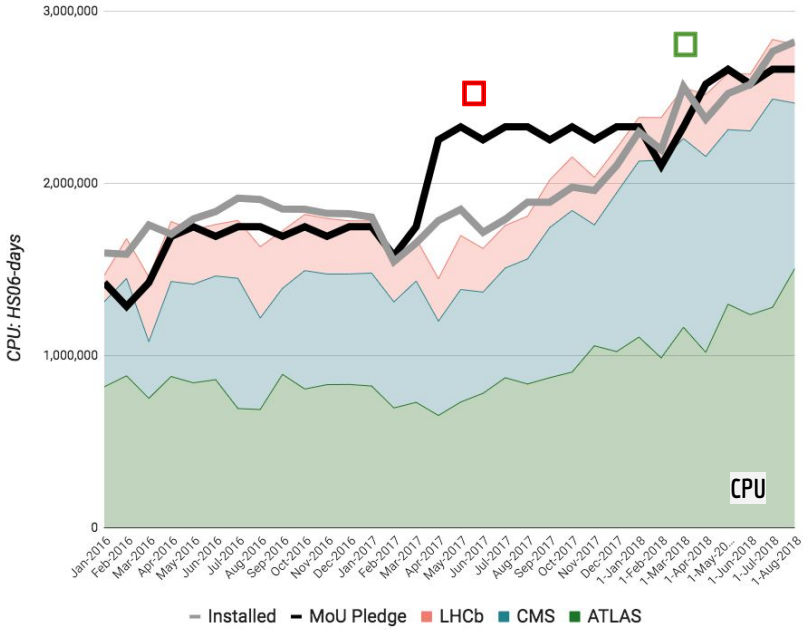
T2K [neutrinos], **MAGIC** and **CTA** [gamma-ray astronomy], **PAU** and **EUCLID** [cosmology], **VIP** [instrumentation]

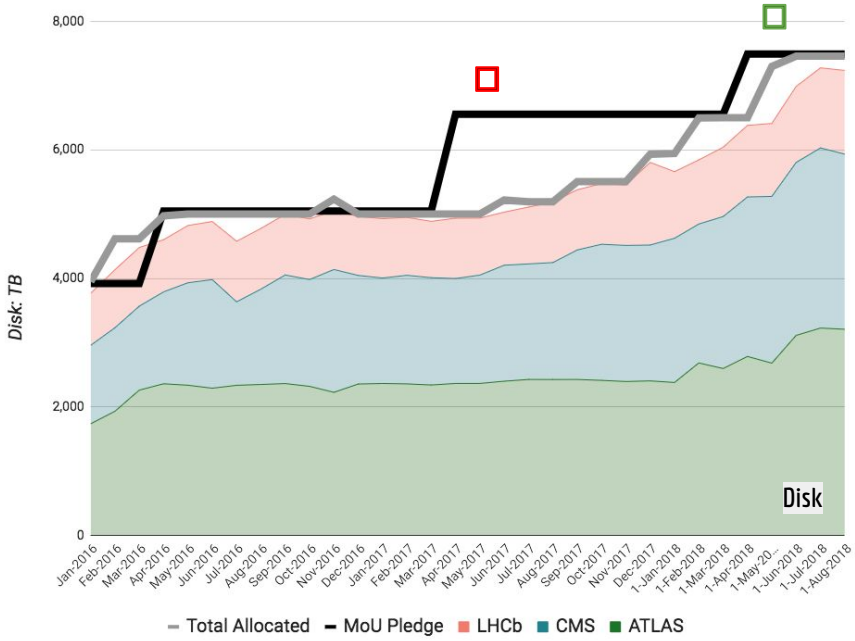
(some) news at a Glance

- ~60% of PIC farm managed by **HTCondor**
- Lots of **microcode updates** (Spectre and Meltdown)
  - Done with Variant 1, 2, 3a. Working on L1TF
- **dCache** v.3.2.18 in production → migrating to 4.2 before Summer ‘19
- **Enstore** 6.1.0-5 in production
  - LT04 tapes are history now (fully migrated to new technology)
  - CMS massive deletions, repacks and recovering of tape space - ATLAS Tape Stress tests along the summer
- The new v**irtualization platform** works really smooth with oVirt 3.6
- **Containers**: testing Rancher 2.0 on RancherOS
- **ELK** running in v6.4.1 (latest) & new **Grafana** views in place
- More (new) resources being added to the **Hadoop cluster**
- **HNSciCloud** activities → see J. Casals talk Wed. 10th at 9am [link]
- **Network incidents** [under investigation]
- **IPv6** in production: some issues observed

PIC Tier1 purchases

- Late availability of funding **delayed deployment** of the 2017 Tier1 pledges for CPU & Disk
- The 2018 Tier1 pledges installed ~in time
  - ~10 kHS06, the needed T10KT2 tapes + 1 drive T10KD, and 900 TB disk

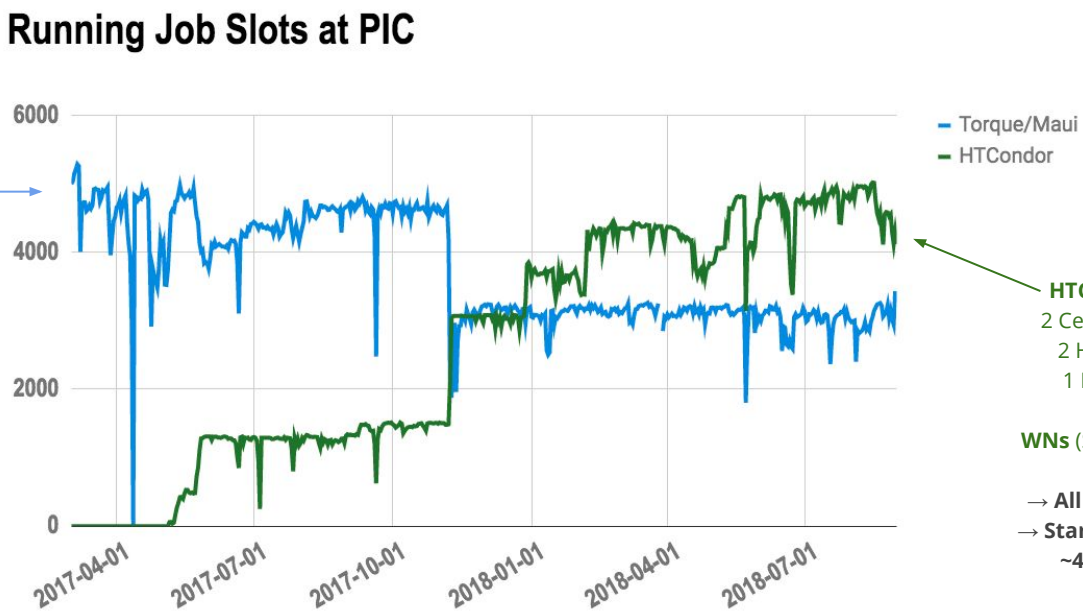




HTCondor migration at PIC

Torque-4.2.10/Maui + MCFloat
Tier-1, Tier-2 residual use
Cosmology projects
local jobs, MPI, small projects

WNs (SL6, ~3300 slots, 40k HS06)



HTCondor v8.6.12 (WNs v8.6.10).
2 Central Manager en HA (CentOS7)
2 HTCondor-CEs Tier-1 (v3.1.2-1)
1 HTCondor-CE Tier-2 (v3.1.2-1)

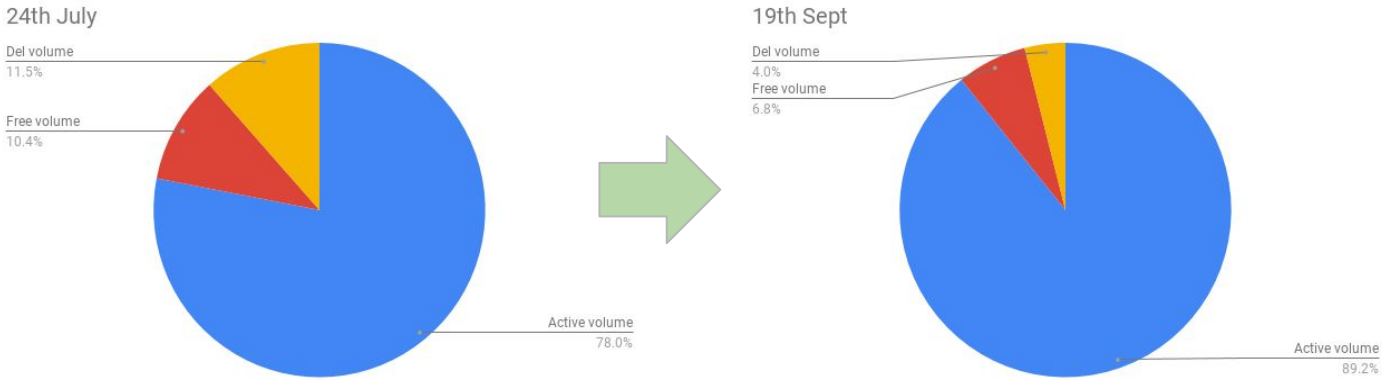
WNs (SL6, approx 5100 slots, 70k HS06)

→ All new WNs go to HTCondor pool
→ Starting migration WNs to CentOS7
~4% gain wrt HS06 seen in SLC6

Starting to move non-LHC projects and users to HTCondor pool

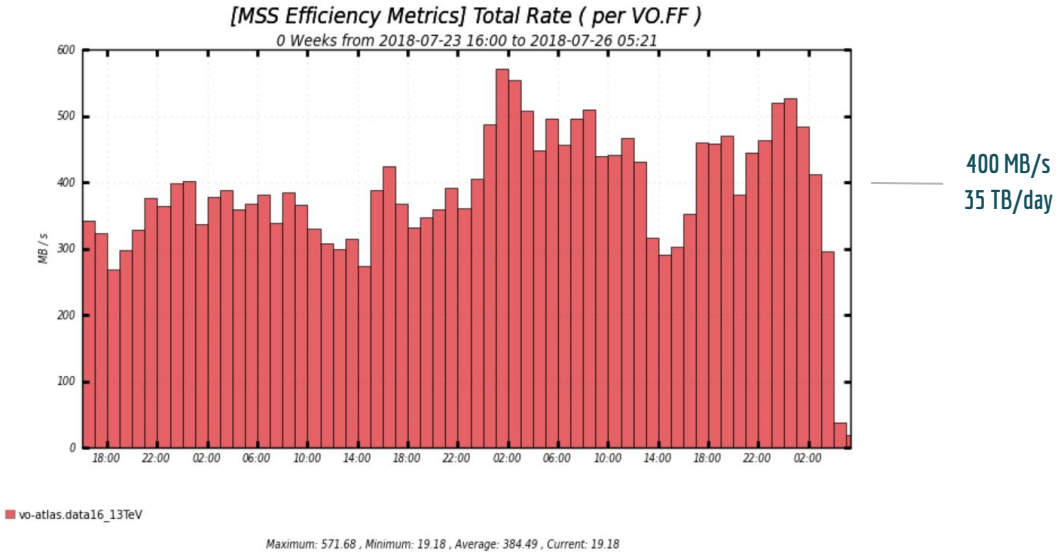
CMS tape deletions / repacks / compaction

- CMS **massive data deletion** by the end of May in PIC (1.6 PB)
  - We ran a repack campaign and recovered 1.8 PB
- Compaction of **small File Families** [i.e. CMS primary datasets with small number of files]
  - Sub-optimal use on ~100 TK10C cartridges (avg. usage ~25%)
  - Generic File Families introduced and we run a compaction of this data onto new schema
    - ~40% of tapes were recovered completely
    - Average active space on tapes went from 78% to 89%
    - The number of File Families reduced from 86 to 56



Tape Stress Tests: ATLAS Carousel model

- Reading from tape 150-200TB of AOD data (file size 2-3GB)
  - **Test #1: 20th July** → no competing activities but 5K bulk requests and 1 single dataset used
  - **Test #2: 20th August** → competing activities (ATLAS writes and CMS reads/writes) but >5k requests and 3 datasets used
  - Clearly, different running conditions → site comparisons **will be shown at the BoF session**

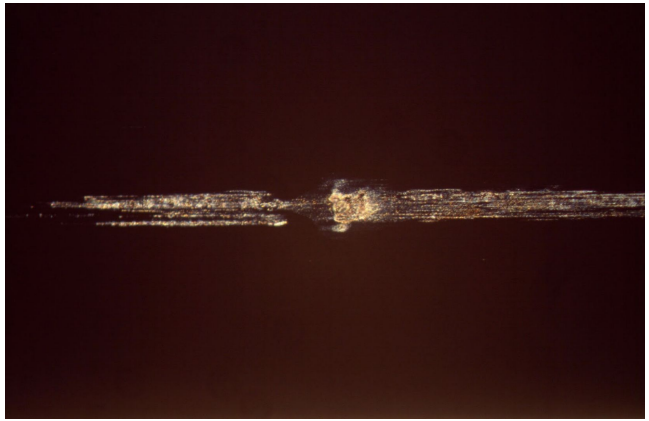


Tape Stress Tests: ATLAS Carousel model

- Reading from tape 150-200TB of AOD data (file size 2-3GB)
  - **Test #1: 20th July** → no competing activities but 5K bulk requests and 1 single dataset used
  - **Test #2: 20th August** → competing activities (ATLAS writes and CMS reads/writes) but >5k requests and 3 datasets used
  - Clearly, different running conditions → site comparisons **will be shown at the BoF session**
- **[main]** Lessons learnt from PIC
  - Retune max active restores per recall pool
  - Retune Srm Manager-related params to handle pinning requests
  - Checksum on the fly for pool to pool transfers
- **[other]** findings
  - Some FTS issues (see **INC1751084**) - FTS timeouts (see **INC1772899**)
  - Communication issues between FTS and DCache (e.g. issues with the GFTP performance marker)
  - Understand the role of Rucio vs FTS in the management of the requests submitted to site (e.g. to cap bulk requests per activity)
  - Acknowledge limits dCache (at least 3.2) functionalities (e.g. lack of faireshare per-data activities, rebalance of the available space on the pools, performance transfer pool-to-pool, etc...

Problem with a TK10C tape drive roller

- In August 2018, we detected read errors (5 files) on a **T10KC** cartridge
  - We were able to read the problematic files using a T10KD drive
- The cartridge was sent to Colorado (Oracle Data Recover) and they found that the problem supposedly had to do with a **roller damage in a drive**
- Each T10KC drive has 5 rollers - dumps sent to Oracle. They **replaced 1 drive** (main suspect) **and an additional one** (preventive) → PIC has 8 T10KC drives

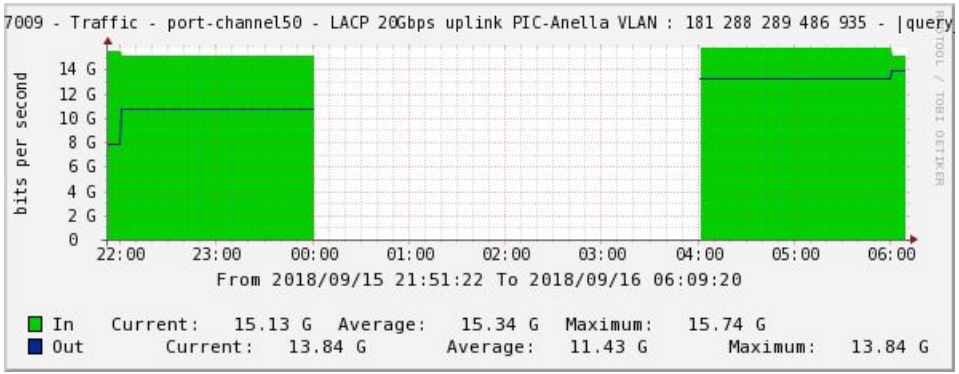


Ipv6

- PIC WNs (80%) and Storage in **dual-stack**
  - Data transfers between those WNs and Storage should go over IPv6
- **Last month:** 17.3 kTransfers → ~40% of transfers went under IPv4 **(!) Why???**
  - By protocol: 72% Gftp / 24% XRootd / 4% NFS
  - By experiment: 68% ATLAS / 26% CMS / 1% LHCb / ...
- Checks:
  - Xrdcp and gfal works fine in IPv6 (without specifying anything)
    - But, what happens to those clients brought by old software releases? **(!)**
  - Globus-url-copy needs -ipv6 as an option to force IPv6 **(!)**
  - What happen to ROOT reads via XRootD? **(!)**
    - TFile.Open("xroot://....") is done via IPv4 or IPv6?
- We need to **investigate further** the stage-in and stage-out procedures for the LHC experiments at PIC (dominant VOs), protocols, and client versions used

Network incidents 1/2

- 3 major incidents (**PIC blackouts**) last months:
  - 3rd August 00 - 08 **[8 hours]**
  - 16th September 00 - 04 **[4 hours]**
  - 28th September 13:45 - 14:15 **[30 minutes]**
- In all of the cases, one disk pool hangs, and starts a cascade of events:
  - The switch that handles all of the disk pools (Dell S4048-ON) turns mad
  - The aggregated 80 Gbps port from Dell S4048-ON to core router goes down
  - Then, the core router aggregated connection to Firewalls goes down



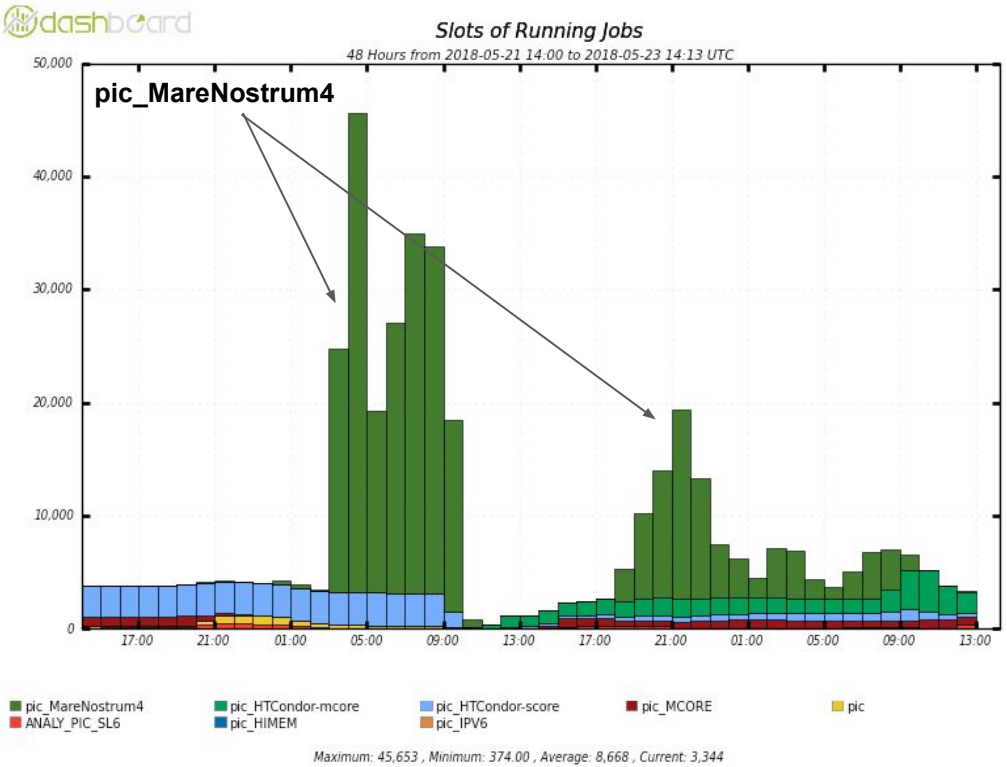
Network incidents 2/2

- In all of the occasions, the **disk pool originating** the cascade was a SuperMicro server (latest purchases) with network cards Intel(R) 82599:
  - These pools are 145 TB/server - we have lot of pools of this type deployed in PIC
  - The fact that they originated these can be for the simple fact that they are the majority (!)
  - But, they run CentOS 7.5 and we need to re-compile the network driver (ixgbe) for each new kernel that is deployed on the servers - that makes them 'different' wrt other pools
- We opened a case with DELL and NEXUS support. Some **actions taken**:
  - DELL protocol Spanning-Tree rstp used, while core router configured with rapid-pvst+. Apparently this might cause incompatibilities → DELL switch upgraded to support rapid-pvst+
  - FlowControl deactivated (LACP) and we are preparing an upgrade of the core router (Nexus)
  - VLAN500 was trunk connected to DELL. This is the VLAN of the DMZ, hence connected to Firewalls. This has been deconfigured to avoid future network loops in the DELL switch

Network incidents 2/2

- In all of the occasions, the **disk pool originating** the cascade was a SuperMicro server (latest purchases) with network cards Intel(R) 82599:
  - These pools are 145 TB/server - we have lot of pools of this type deployed in PIC
  - The fact that they originated these can be for the simple fact that they are the majority (!)
  - But, they run CentOS 7.5 and we need to re-compile the network driver (ixgbe) for each new kernel that is deployed on the servers - that makes them 'different' wrt other pools
- We opened a case with DELL and NEXUS support. Some **actions taken**:
  - DELL protocol Spanning-Tree rstp used, while core router configured with rapid-pvst+. Apparently this might cause incompatibilities → DELL switch upgraded to support rapid-pvst+
  - FlowControl deactivated (LACP) and we are preparing an upgrade of the core router (Nexus)
  - VLAN500 was trunk connected to DELL. This is the VLAN of the DMZ, hence connected to Firewalls. This has been deconfigured to avoid future network loops in the DELL switch

End-to-end integration of BSC CPU [ATLAS]



April 2018 Tests on the MareNostrum HPC in joint collaboration with IFIC Tier2 site

More than 40k-cores used using an ARC-CE hosted at PIC to run ATLAS simulations

Data async. transferred to PIC and registered into ATLAS DM system

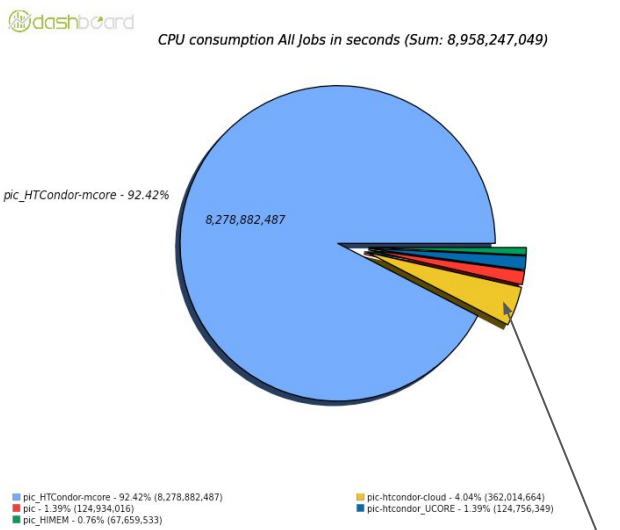
⟶ Exploiting grants of ~100 khours
⟶ Next phase approved (via PRACE)

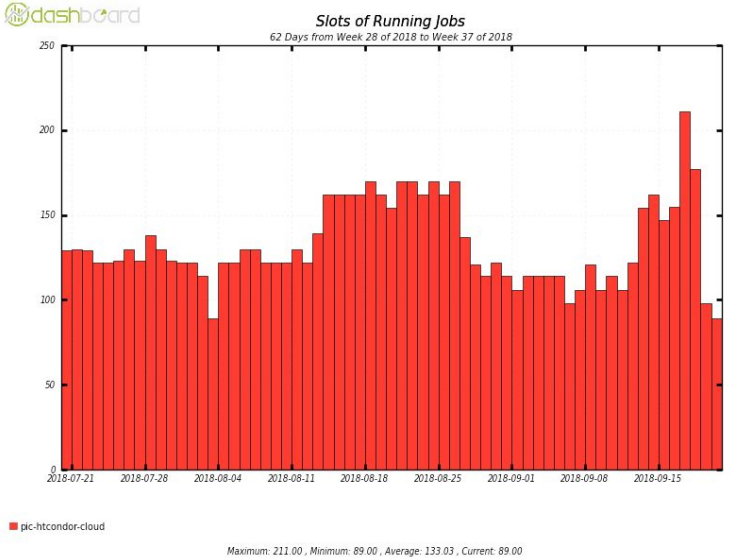
End-to-end integration of Cloud CPU

ATLAS Example →

HNSciCloud
OTC compute nodes integrated into HTCondor

CMS sending production jobs since September 2018





HNSciCloud
-4% ATLAS CPU at PIC since mid July 2018

Cores used by ATLAS jobs running in the cloud from 20 July - 20 September 2018 sent from PIC

→ more details @ J. Casals talk Wed. 10th at 9am [link]

Future activities Tier1

- Joint project with **HTCondor development team** to exploit HPC resources without network connectivity in the compute nodes
    - Necessary to run CMS pilot jobs on compute nodes with no network connectivity
- Participation in EU ESCAPE project in **data lakes R&D**
- Build a **disk caching system** between PIC & CIEMAT (Madrid CMS Tier2) sites
    - PhD student at PIC working on this
- **Federate** PIC and CIEMAT local **batch systems**

Gràcies!

Preguntes?