PIC
port d'informació
científica

CMS

XCache deployment: initiatives and experiences in Spain

Josep Flix [PIC/CIEMAT]

pre-GDB XCache - CERN 8th July 2019

EXCELENCIA SEVERO OCHOA

Institut de Física d'Altes Energies

IFAE

Ciemat

Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

EXCELENCIA MARÍA DE MAEZTU

Cache deployments - storage consolidation/federation?

Context: PIC Tier1 and CIEMAT Tier2, sites that are at ~10 ms latency

Several studies being done:

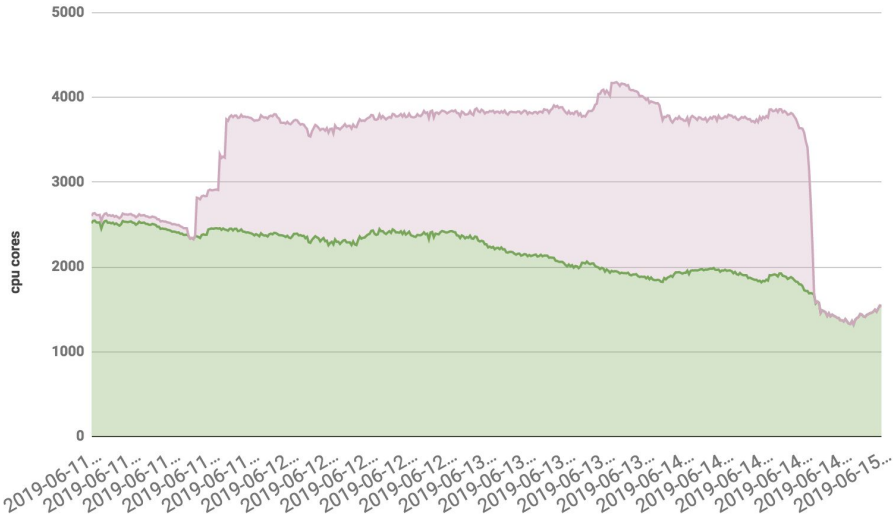
- How **latency** affects the CMS workloads?
  - Real tests with **AWS [40ms]** + sending a fraction of jobs from **PIC to CIEMAT [10ms]**, and vice versa

- How the storage systems are **utilized** in PIC and CIEMAT?
  - **Data access** studies
  - **Performance** studies
  - Which data is susceptible to be **cached** and which could be the net effect?
  - Or... could we have a **consolidated/centralized storage** and read remotely?

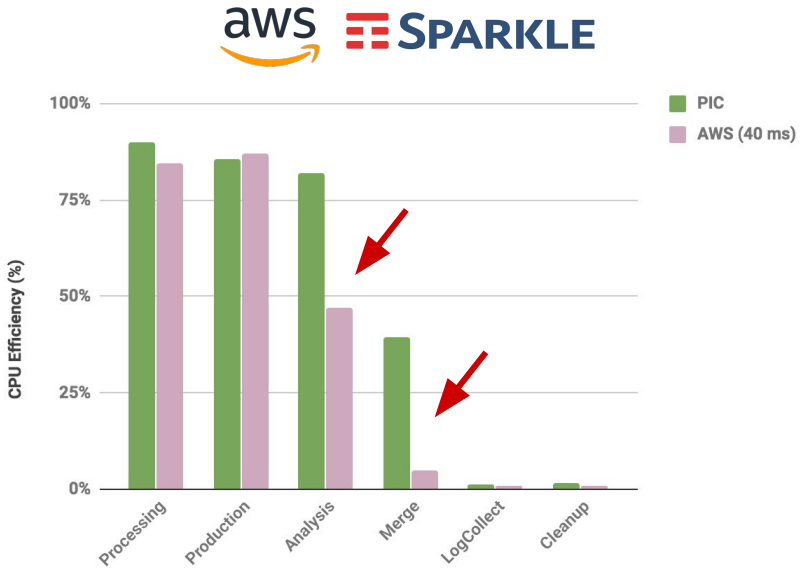
Latency effects on CMS Workloads :: AWS

We tested AWS spot instances for a week, doubling PIC compute power

Data center in Frankfurt (**~40 ms**) - used Condor-Annex



InputFiles Onsite



https://monit-grafana.cern.ch/d/nT45a1mZk/pic-dashboards-aws-tests?orgId=11&from=1560204000000&to=1560549600000&var-bin=1h&var-Input=Onsite

Latency effects on CMS Workloads :: PIC - CIEMAT

While ago we enabled **overflow** of jobs from PIC to CIEMAT and vice versa, and we deployed a regional XRootD re-director (in high availability)

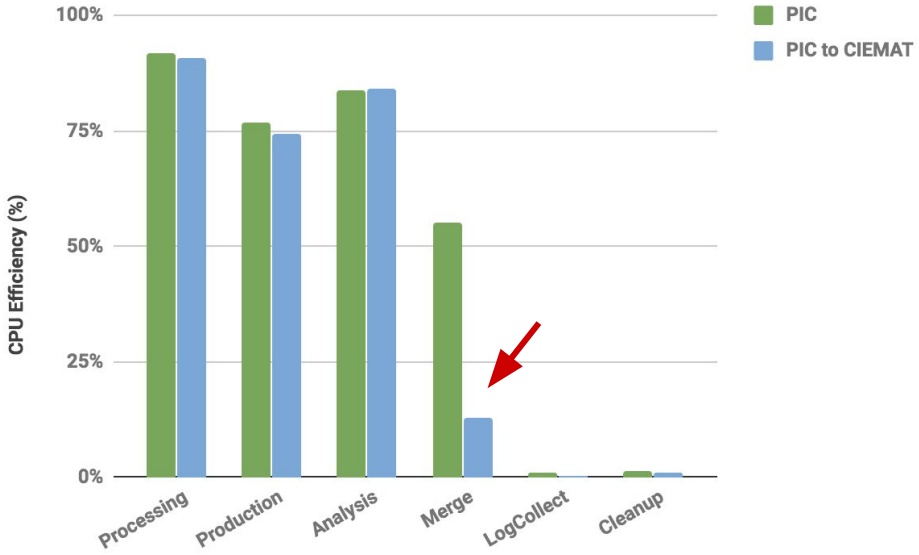
Since two months, we are **flocking** jobs from PIC to CIEMAT and vice versa, since we have HTCondor BS in both sites ➜ 80 cpu-cores available at each site **[dedicated machines, for the moment]**

Regional reads are preserved, since we have regional XRootD re-director deployed - hence we can study job degradations when running remotely

https://monit-grafana.cern.ch/d/fT9FrNRZk/pic-dashboards-pic-ciemat-re-routed-jobs?orgId=11&from=1561473412144&to=1562078212144&var-bin=$__auto_interval_bin&var-Input=All

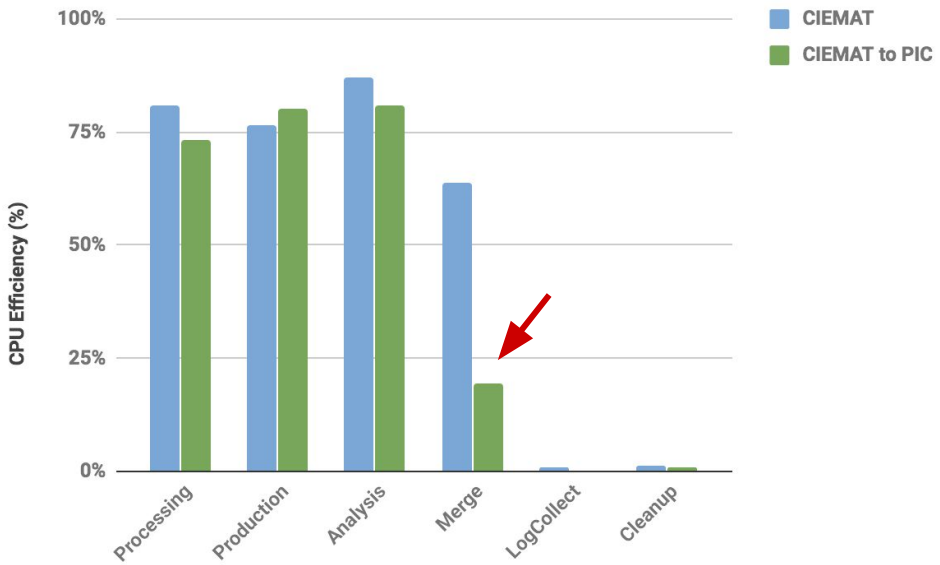
Latency effects on CMS Workloads :: PIC - CIEMAT

From 2019-06-07 to 2019-07-07



5.6 % of PIC jobs executed in CIEMAT

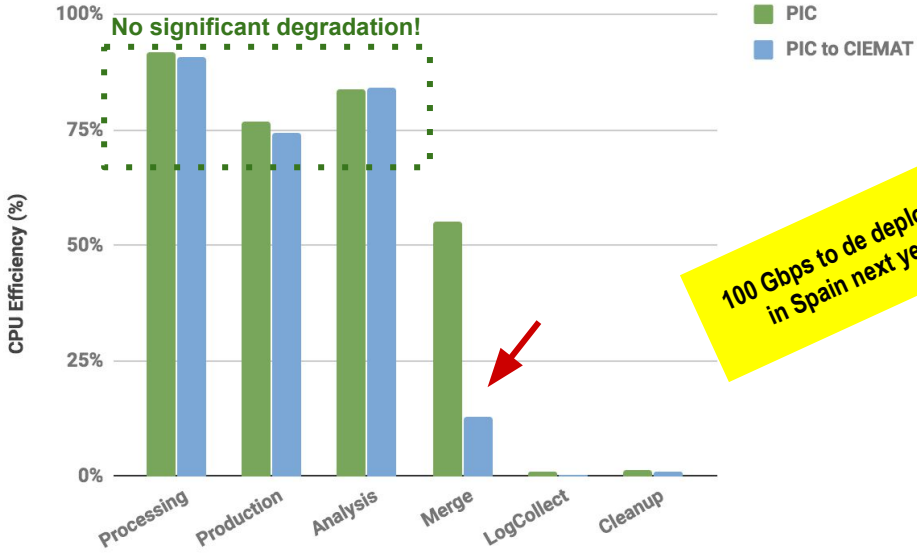
InputFiles Onsite



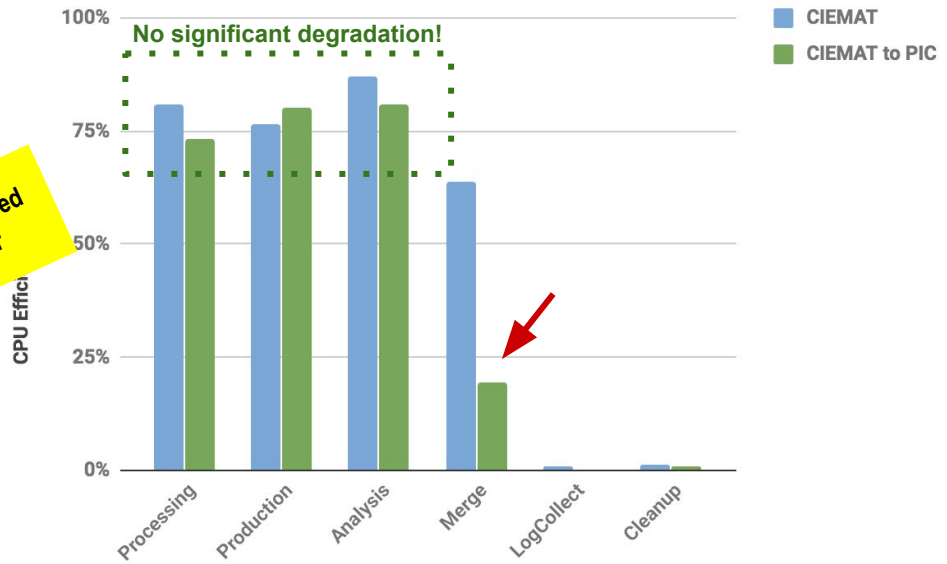
2.8 % of CIEMAT jobs executed in PIC

Latency effects on CMS Workloads :: PIC - CIEMAT

From 2019-06-07 to 2019-07-07



100 Gbps to de deployed in Spain next year

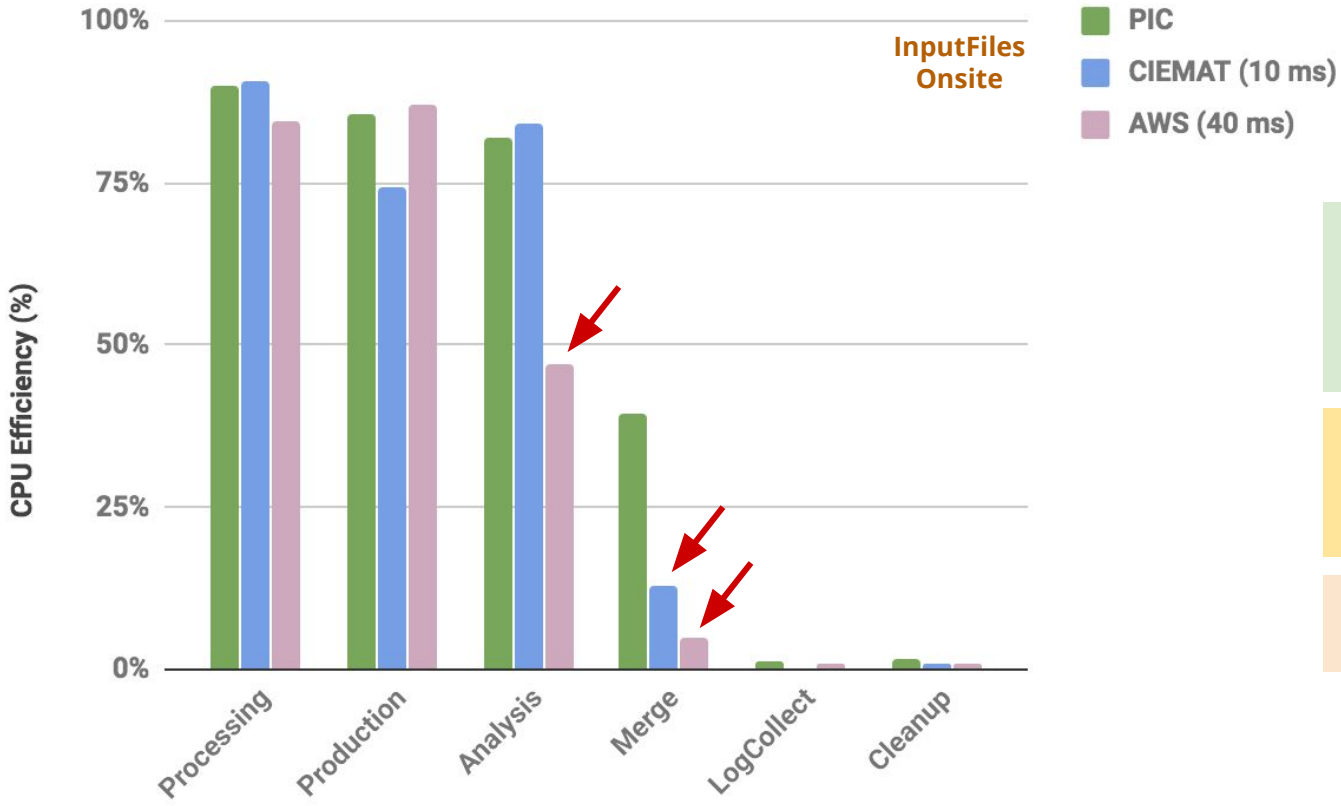


5.6 % of PIC jobs executed in CIEMAT

InputFiles Onsite

2.8 % of CIEMAT jobs executed in PIC

Latency effects on CMS Workloads :: PIC - AWS/CIEMAT



This tells us that within the region, PIC could run jobs either at PIC or CIEMAT (reading files from PIC), except merge (which should run locally always)!

This of course would cause an increase of PIC exports (stressing both for network and storage system... how much?)

At higher latencies (40 ms), analysis starts to be degraded

How the storage systems are utilized in PIC and CIEMAT?

How **data is accessed**? - are we working in the most **optimal point**? Both at PIC Tier1 and CIEMAT Tier2?

Shall we go for deployment of **caches**? For **which data**? Which are the **cost benefits**? (we can simulate based on real data accesses)

PIC and CIEMAT are close enough (10ms) - shall we aim for a data **federation** or **consolidation** of storage in the region?

➜ Concepts that are being explored atm ➜ next talk **[C. Pérez Dengra]**