Response of the ATLAS Spanish Tier2 for the collisions collected in the first run at LHC

Presenter: Santiago González de la Hoz
IFIC (Instituto de Física Corpuscular), Centro mixto CSIC-Universitat de València

[1]GONZALEZ, S.,[1]SALT, J., [1]AMOROS, G, [1]FERNANDEZ, A., [1]KACI, M., [1]LAMAS, A., [1]OLIVER, E., [1]SANCHEZ, J., [1]VILLAPLANA, M., [2]BORREGO, C., [2]CAMPOS, M., [2]NADAL, [2]J., OSUNA, C., [2]PACHECO, A., [3]DEL PESO, J., [3]PARDO, J., [3]GILA, M., [4]ESPINAL, X.

[1]IFIC-Valencia, [2]IFAE-Barcelona, [3]UAM-Madrid, [4]PIC-Barcelona

Outline

- **Introduction**
- **Operation of the Tier2 during the LHC beam events**
  - Resources
  - Operation procedure
- **Data management of the collision events**
  - Data Distribution Policy
  - Distribution in the storage elements (space tokens) & access to the data
- **The simulated physics events productions**
  - The ATLAS distributed production system
  - Simulated events production at ES-ATLAS-T2
  - Contribution of Iberian cloud to the EGEE and ATLAS
- **Distributed Analysis and Local Analysis Facilities**
  - From Raw data to the final analysis phase
- **Conclusions**

Introduction

- **Portugal and Spain**
  - forming the **South Western Europe** Regional Operations Centre (ROC) within EGEE. Iberian Cloud
  - are collaborating with the **ATLAS computing effort**:
    - storage capacity
    - computing power
  - have **an ATLAS federated Tier2** infrastructure
- **ATLAS federated Tier2 in Spain consists of the three sites:**
  - **IFAE-Barcelona**
  - **UAM-Madrid**
  - **IFIC-Valencia** (Tier2 activities coordinator)
- **The Tier1 is at PIC-Barcelona**

Operation of ES-ATLAS-T2

- **Resources**
  - During the 1st LHC collisions the ES-T2 provided the hardware resources fulfilling the ATLAS requirements of the MoU.

	ES-ATLAS-T2	
	Pledge'09	Provided *(Nov-Dec)*
CPU (HEP-SPEC06)	5396	**5341**
Disk (TB)	656	**750**

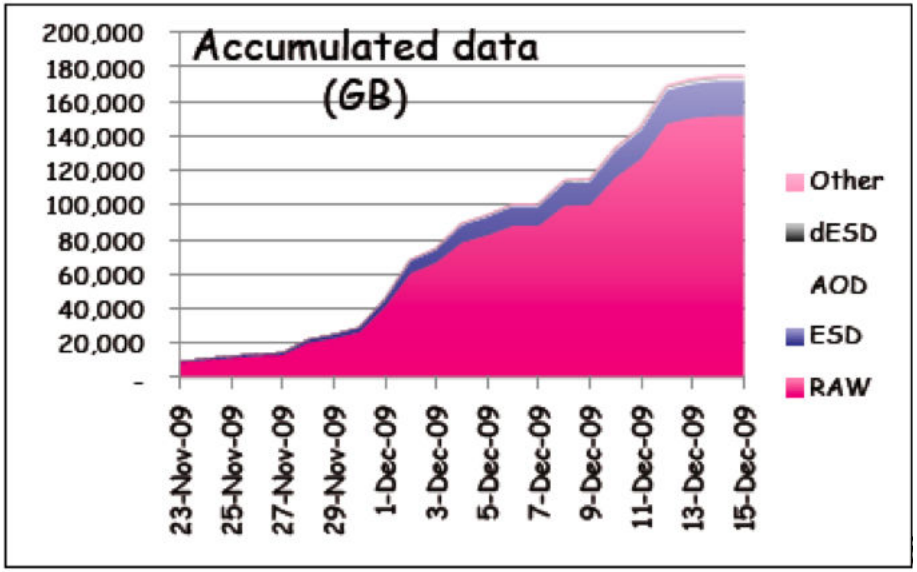
  - Disk space is managed **by** **two distributed storage systems**:
    - dCache at IFAE and UAM
    - Lustre at IFIC
  - WNs have 2 GB of RAM per CPU core (to run the highly demanding ATLAS production jobs)
  - In addition each site provides users **with a Tier3 infrastructure** for data analysis:
    - a part is based on Grid architecture (batch system)
    - another part being a standard computing cluster (interactive analysis)

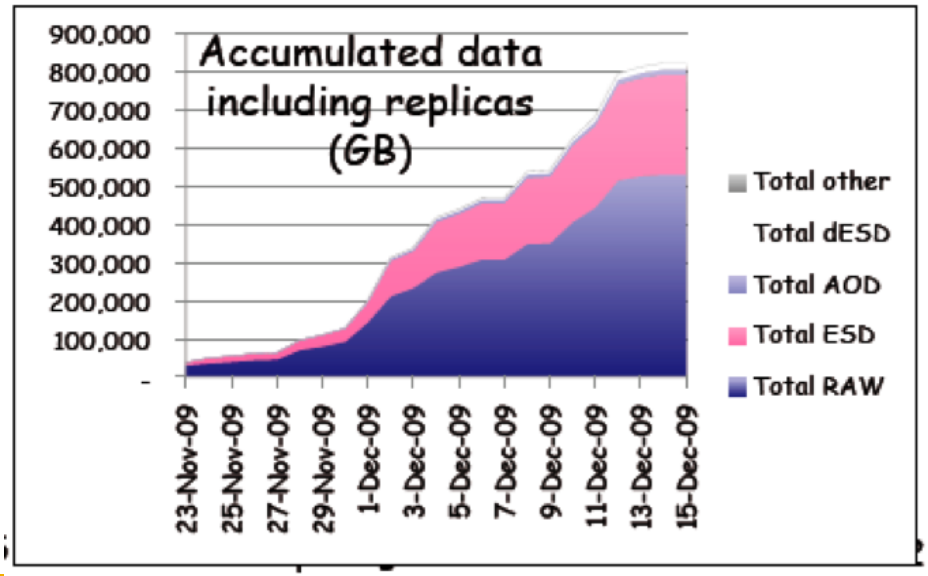
Operation of ES-ATLAS-T2

- **Operation procedure (The Usual one)**
  - Tools for automatic installation&configuration (large number of machines)
    - To install and configure the operating system, Grid middleware and storage system.
      - *QUATTOR - tool used at ES-T2 (UAM participating in its development).*
      - *ATLAS software is installed submitting a Grid job using a centralized procedure from the ATLAS software team.*
  - Monitoring (to assure the needed reliability of the T2 centre)
    - To detect any possible failure of the services or of the hardware
    - Two levels of monitoring:
      - *Global LHC Grid (Service Availability Monitoring).*
      - *Internal to the site using tools like Nagios, Ganglia or Cacti.*
  - Management of updates (to avoid to jeopardize the centre availability)
    - Every site has a pre-production cluster where software updates are tested.
  - Fair share (submit jobs according to their priority)
    - CPU usage fairly shared between MC and analysis jobs
      - *The % assigned to each role is configured on the scheduler (Maui)*
      - *It is used by the batch queue system (Torque)*

Data Management of the collision events

- **Data distribution policy (during 2009 beam run period)**
  - Raw data were transferred from CERN to the 10 Tier1s (keeping 3 replicas of them).
  - Reconstructed data ESD/AOD/dESD were distributed to Tier1s and Tier2s, retaining 10 (13) replicas.
    - Large number of replicas due to the low luminosity of early data
  - A large fraction of these data has been recorded on disk
    - To perform analyses for the understanding of the detector response





Data Management of the collision events

- **Distribution in the storage element (spacetokens)**
  - Storage in ATLAS is performed by using space tokens
    - ATLASDATADISK is dedicated for real data
    - ATLASMCDISK is populated with Monte Carlo production
    - In addition, space tokens are reserved for physics groups and users

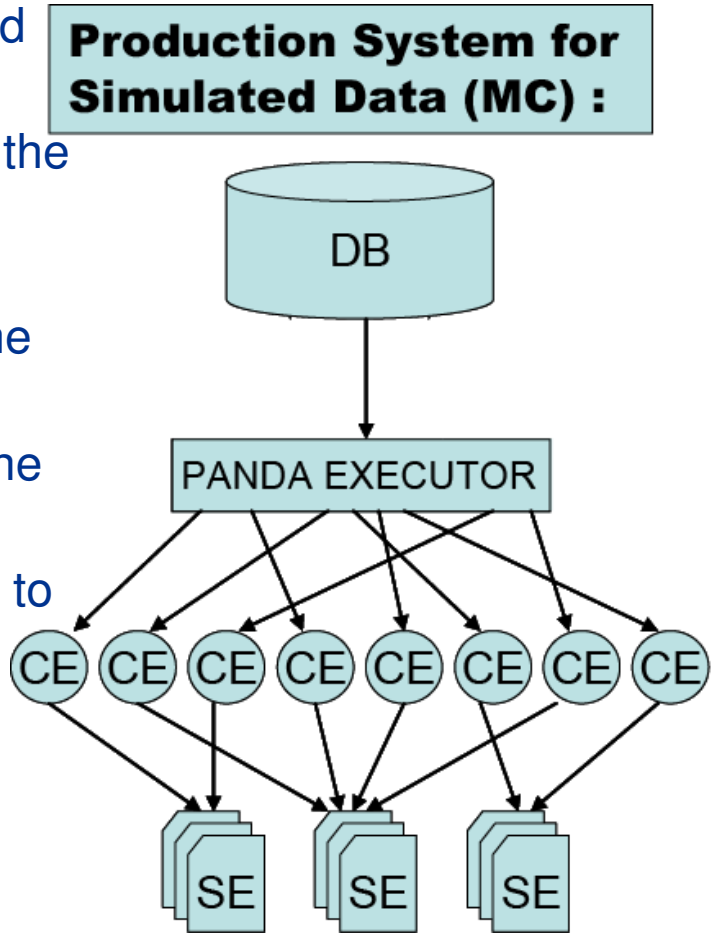
Data distribution in ES-T2 on February 2010

Site	SpaceToken	Total GB	Used GB	Free GB	%
UAM	ATLASDATADISK	55879	12935	42944	23
	ATLASMCDISK	55879	35348	20530	63
IFAE	ATLASDATADISK	69849	16491	53357	23
	ATLASMCDISK	93132	30492	62640	32
IFIC	ATLASDATADISK	114612	42285	72327	36
	ATLASMCDISK	143953	69214	74739	48

- **Access to the data**
  - Users can submit a job to the Grid, the job will be running at the site where the data are located.
  - Users can request to replicate some data to the Tier2 site they belong to, in order to access the data in a native mode.

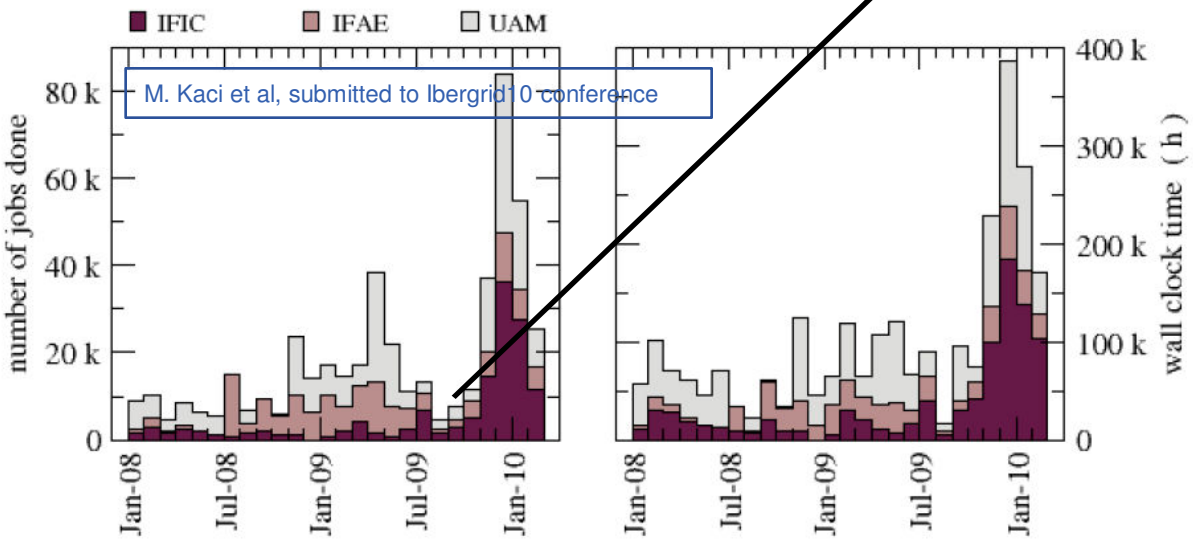
The simulated events production

- **The ATLAS distributed production system (based on Pilot jobs since January 2008)**
  - A database (DB) where jobs to be run are defined as well as their run-time status.
  - An executor (PANDA) which takes the jobs from the DB and manages sending them to the ATLAS computing resources, using pilots jobs.
    - Check the correct environment for running the jobs
    - Have the ability to report free resources on the cluster they are running.
    - Together with DDM transfer the needed data to the site before running the job
  - A distributed data management (DDM) system which stores the produced data on the adequate storages resources and register them into the defined catalogues.



The simulated events production

- **Simulated events production at ES-ATLAS-T2**
  - T2-ES has been actively contributing to this production effort.
  - An increase is observed for the few last months (massive activity took place in ATLAS).
  - The five last month:
    - 33k jobs ran at IFAE
    - 95k jobs ran at IFIC
    - 85k jobs ran at UAM

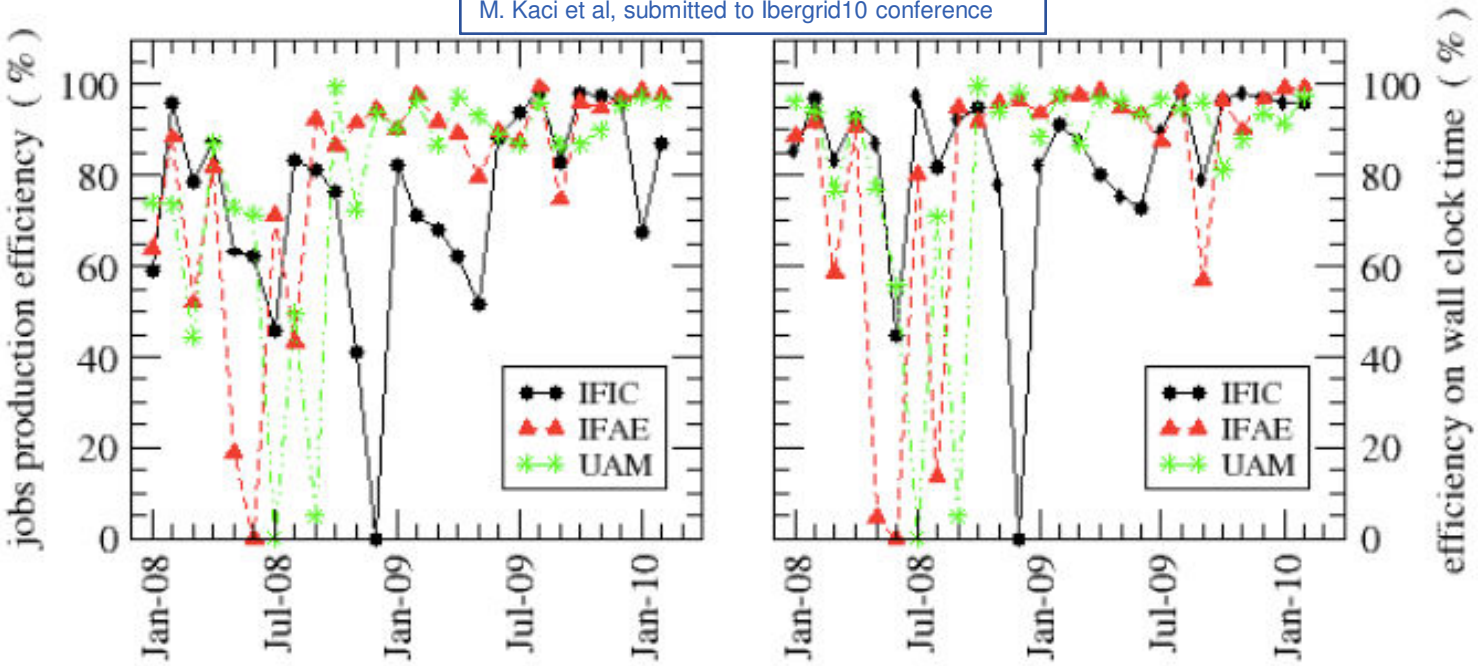


- For some months, T2-ES is not contributing because of:
  - errors related to the storage element (missing file, read/write failures).
  - errors due to bad installation of the ATLAS software.
  - downtime needed for upgrades.

The simulated events production

- **Simulated events production at ES-ATLAS-T2**
  - Job efficiency = *(n#jobs successfully done)/(n#jobs total submitted)*
  - Wall clock times efficiency = *(wall clock time spent to run successful jobs)/(total wall clock time used by all the submitted jobs)*
  - Jobs efficiency is around or more than 90%
  - Wall clock time efficiency ~ 95%

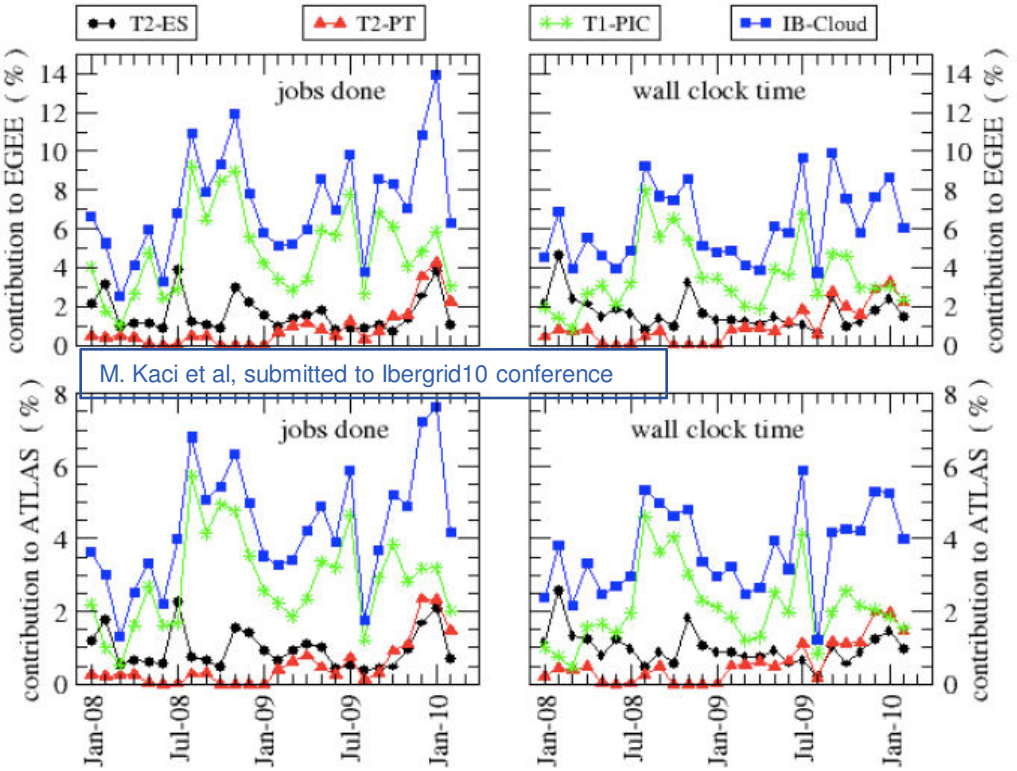
M. Kaci et al, submitted to Ibergrid10 conference



The simulated events production

- **Contribution of IB-Cloud to EGEE and ATLAS**

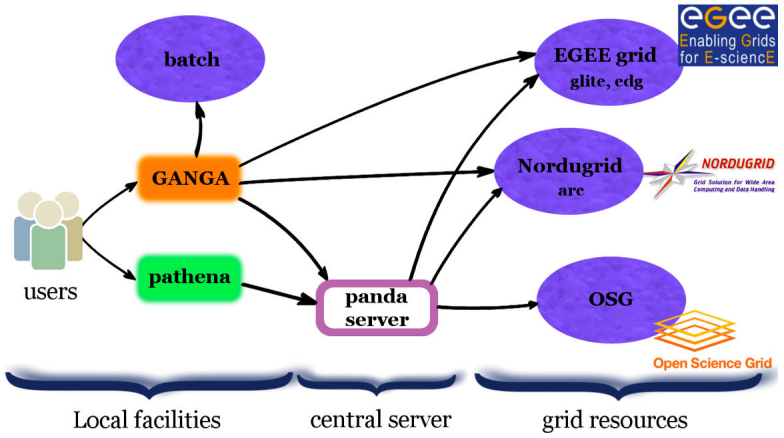
  - 12-14% in EGEE and 8% in ATLAS for the jobs done.
  - 10% in EGEE and 6% on ATLAS for the wall clock.
  - Jobs successfully run during 2008, 2009 and first two months in 2010. Efficiencies are in parenthesis.
  - An increase of the simulated events production activity together with an improvement of the measured efficiencies.



Number of jobs	2008	2009	2010 (*)
IB-Cloud	5.1×10^5 (75.3%)	1.5×10^6 (86.8%)	3.5×10^5 (91.4%)
EGEE	7.2×10^6 (74.0%)	1.9×10^7 (86.3%)	3.8×10^6 (88.7%)
ATLAS	1.2×10^7 (76.2%)	3.1×10^7 (86.3%)	6.2×10^6 (89.7%)
IB-Cloud/EGEE	7.2 %	7.6 %	9.2 %
IB-Cloud/ATLAS	4.2 %	4.8 %	4.6 %

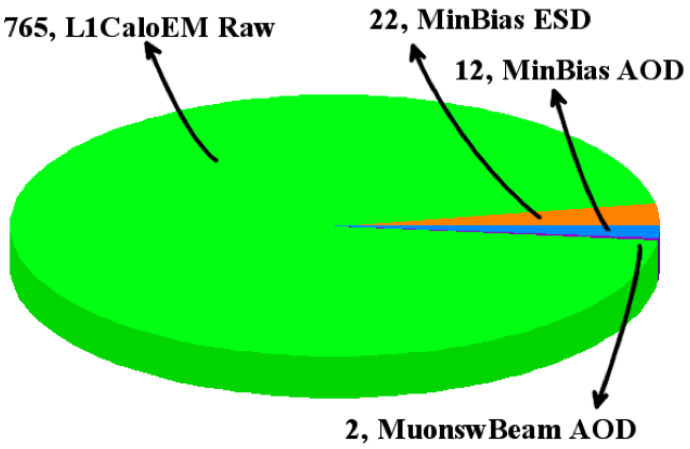
Distributed Analysis & Local Facility

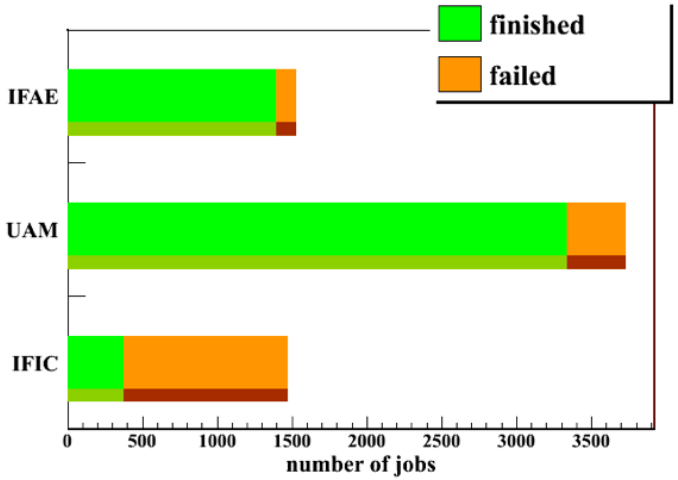
- **Analysis on collision events performed on the ES-T2 resources fall in two categories:**
  - Detector and Combined performance
  - Physics Analysis
- **Typical Analysis**
  - First stage is to submit a job to the Grid (Distributed analysis)
    - The output files will be stored in the corresponding Tier2 storage
  - Later the output files can be retrieved for a further analysis on the local facilities of the institute.
- **ATLAS distributed analysis in the ES-T2**
  - Two front-ends were used: Ganga and PanDA
    - Differ in the way they interact with the user and the Grid resources



Distributed Analysis & Local Facility

- **Ganga has been intensively used by the ATLAS community on ES-T2**
  - More than 800 jobs submitted via Ganga to the ES-T2 from first collisions to March 2010.
  - User is analyzing RAW, ESD and AOD data
- **Jobs are submitted to PanDa via a simple python client**
  - Number of analysis jobs since first collisions to March 2010 allocated by Panda in the ES-T2.
  - More than 5000 jobs submitted via Panda



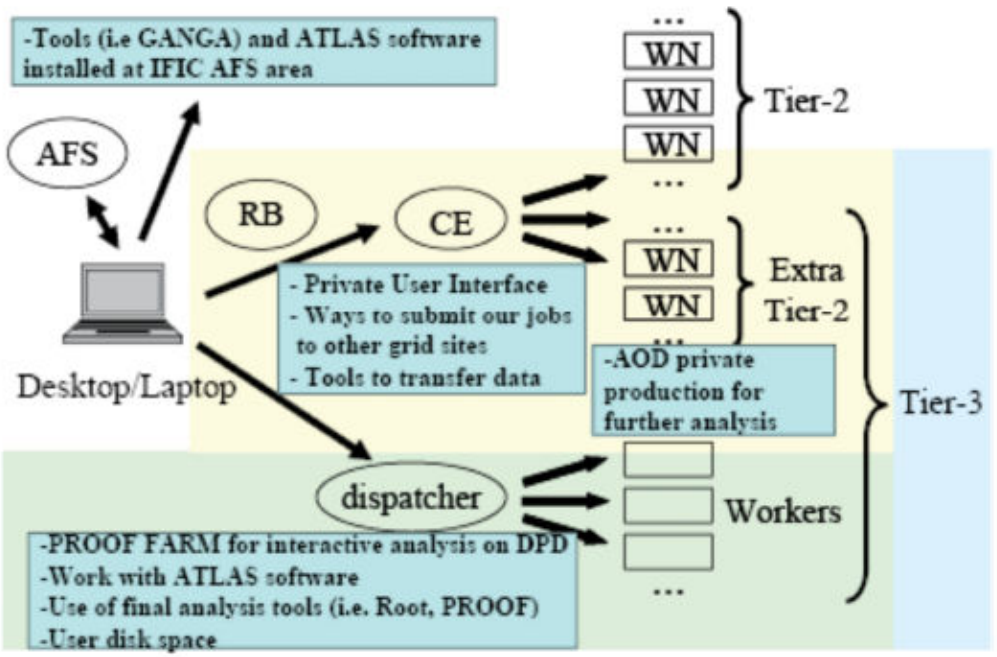


Distributed Analysis & Local Facility

- **Analysis at local facilities**
  - The outcome for the analysis jobs at the Tier2 using Distributed Analysis is a dataset of n-tuples (DPD) stored in the SE.
  - Next step is carried out at the local facility:
    - Retrieving the dataset generated with two different procedures:
      - *Request a subscription to the ATLAS DDM system to replicate the dataset to the LOCALGROUPDISK area (space token at the Tier3 facility of the Institute)*
      - *Use the ATLAS DDM client tools to download the dataset to the local disk space*
  - Every site (UAM, IFAE and IFIC) has a local facility (Tier3) providing CPU and disk resources for the last phase of the analysis:
    - Some User Interfaces used to perform interactive analysis on the final datasets produced at the distributed analysis
    - A local batch farm to provide additional computing power for analysis
    - A PROOF farm for parallel processing of root analysis jobs.

Distributed Analysis & Local Facility

- **Tier3 Facility at IFIC**
- **From the raw data to the plot**
  - Following the ATLAS Computing Model, only analysis jobs which read data on disk can be submitted to the Grid.



- Raw data are recorded on the tape systems of the Tier1s.
- Raw data are reconstructed resulting ESD and AOD data distributed to the Tier2s disks.
- In the first step of the analysis (Distributed analysis), AOD are processed in order to have filtered versions (dAOD, DPD, ntuples) with interest events for your institute site.
- After having these DPDs/ntuples, one starts the core analysis (local facility-Tier3) with the most common data analysis framework (ROOT) to plot the final results

Conclusions

- **It has been shown that ES-T2 responded very efficiently at the different stages, from collecting data to final experiments results:**
  - receiving and storing the produced data, thanks to the high availability of its sites and the reliable services provided,
  - providing easy and quick access to these data to users,
  - contributing continuously and actively to the simulated physics events production,
  - providing the required distributed analysis tools to allow users to use the data and produce experimental results, and,
  - setting up infrastructure for the final analysis stages (Tier3s) managed by the Tier2 personnel.