



STATO DELLE ATTIVITA BORSE INFN SUL CALCOLO

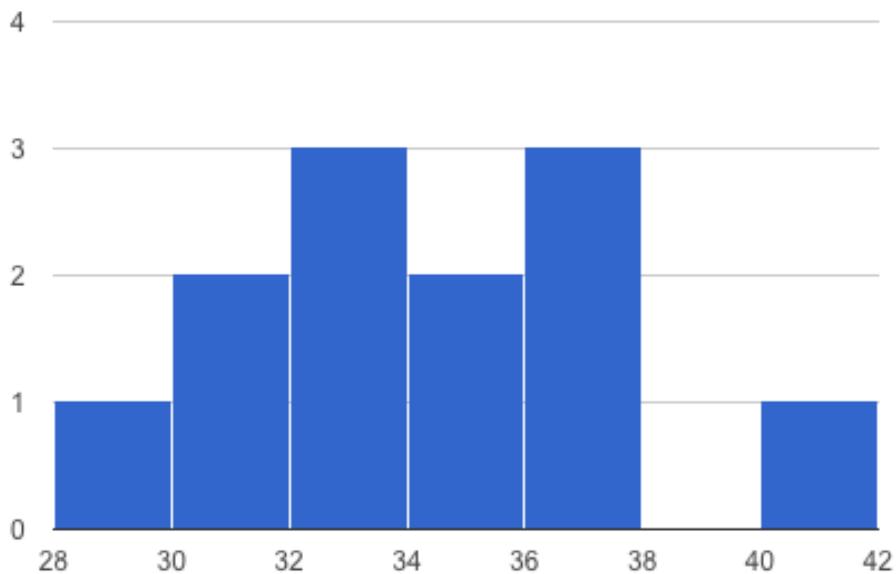
Stefano Bagnasco | INFN Torino

QUALCHE DETTAGLIO

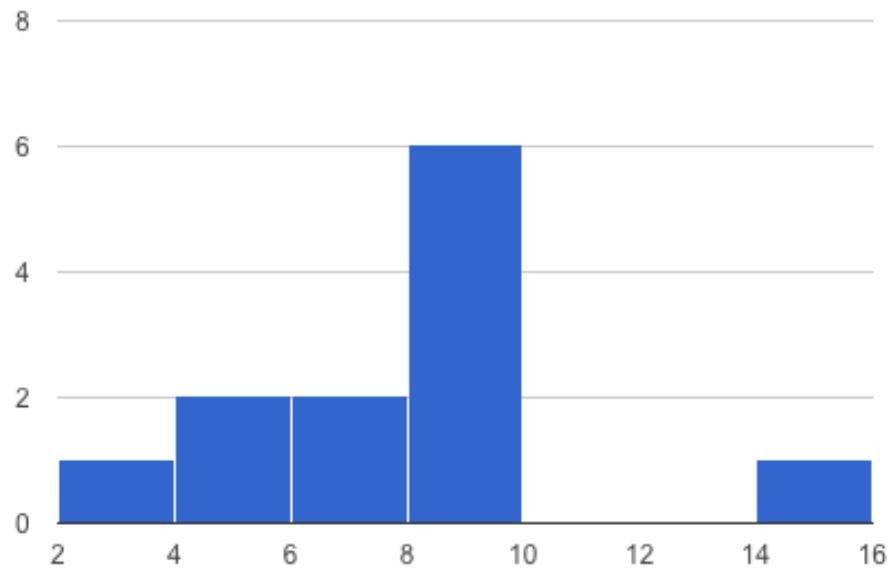
- **Concorso per titoli** n. 18827/2017
“Conferimento di n. **12** borse di studio post-lauream della durata di un anno e rinnovabili per un ulteriore anno, da attribuire a laureati in Informatica, Fisica e Ingegneria, per soggiorni di studio e ricerca presso le Sezioni, Laboratori Nazionali e Centri dell’INFN”
- Domande pervenute n. **34**
- Candidati ammessi n. **32** (di cui n. **1** comunica di voler rinunciare alla partecipazione al concorso)
- Risultano idonei n. **23** candidati e n. **12** vincitori
- Atti approvati dalla GE il 12/05 (delibera n. 11371)
- Prese di servizio un po' alla volta (l'ultima in aprile), qualche defezione e ripescaggio

AGE DISTRIBUTION

Età anagrafica (vincitori)

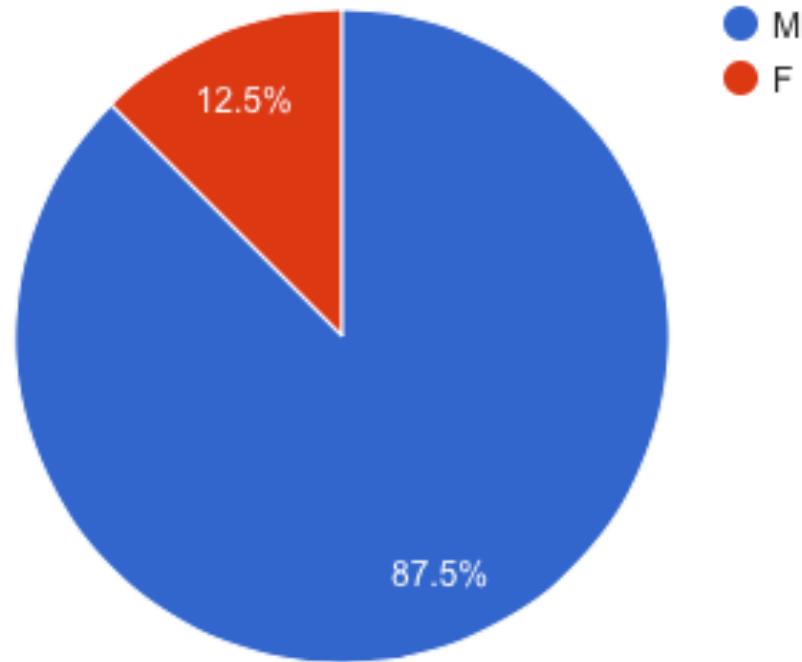


Anzianità di laurea (vincitori)

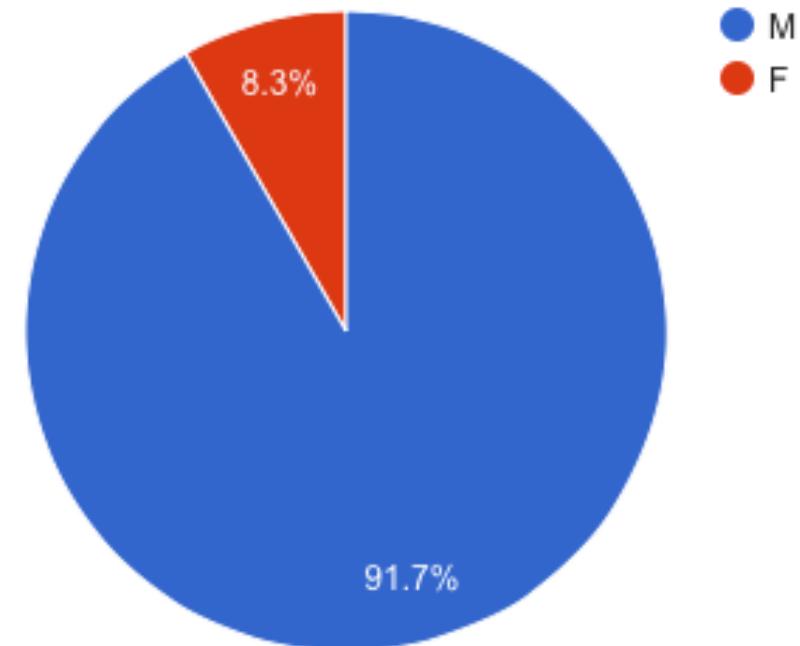


Concezio.Bozzi@fe.infn.it

Genere (ammessi)



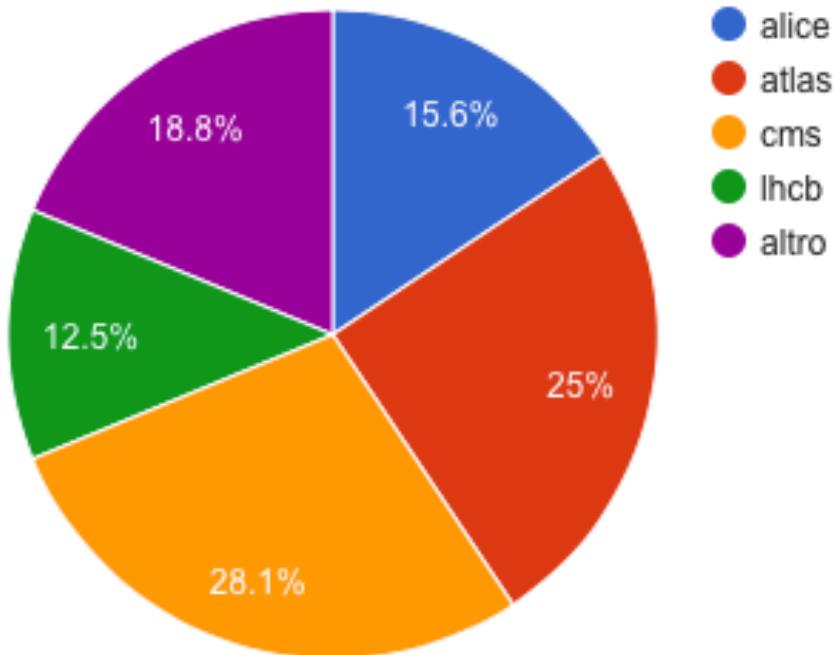
Genere (vincitori)



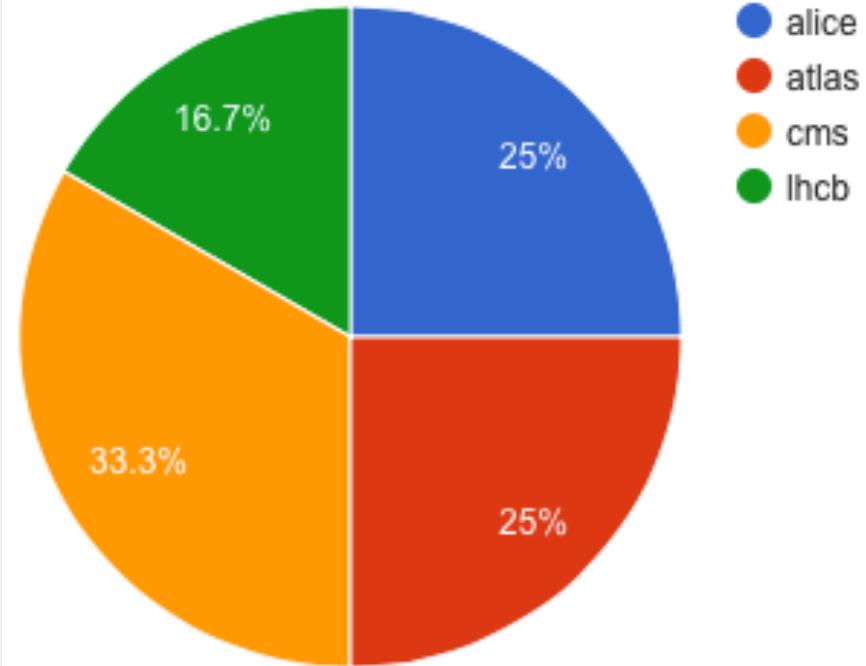
Concezio.Bozzi@fe.infn.it

ESPERIMENTO

Esperimento (ammessi)



Esperimento (vincitori)

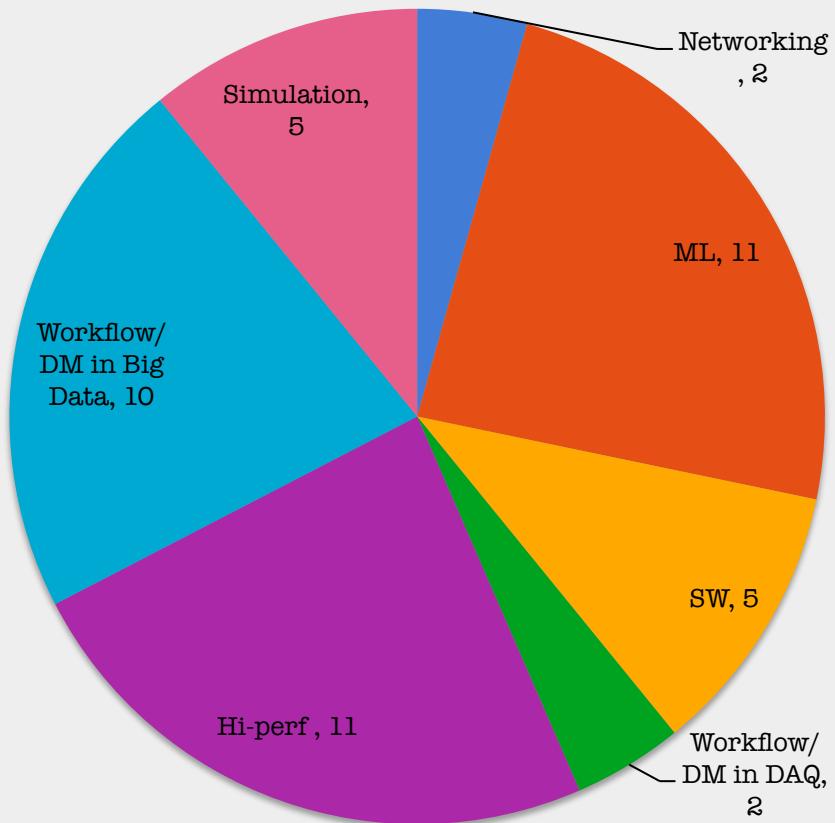


Concezio.Bozzi@fe.infn.it

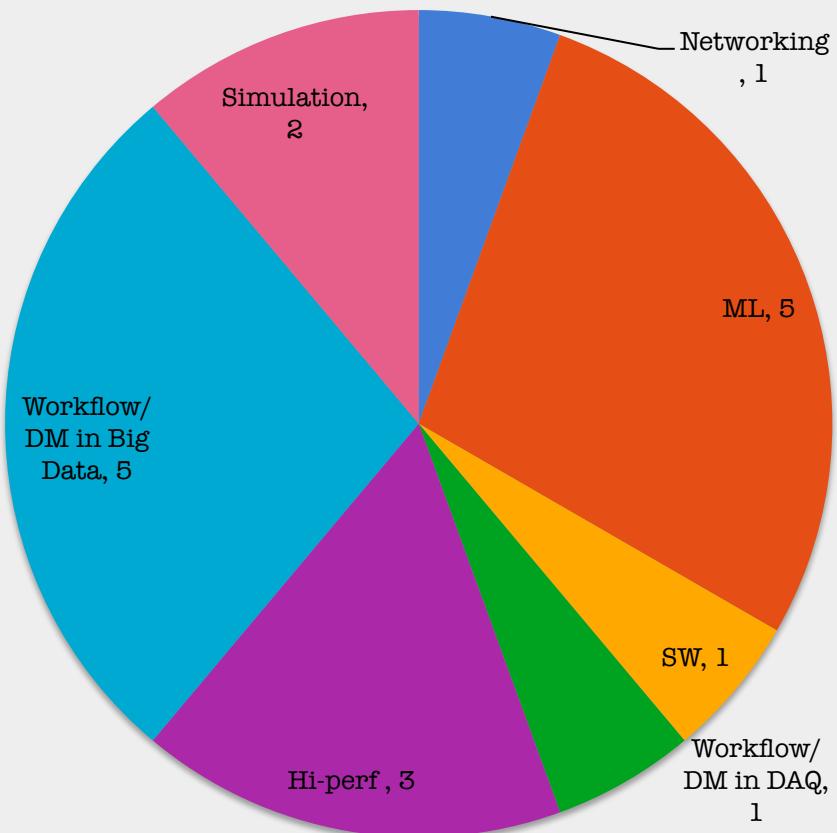
1. **Novel networking techniques** for data acquisition, trigger and data processing
2. **Machine and deep learning** for HEP in reconstruction, selection and analysis algorithms
3. **New technologies in software**: architecture dependent optimizations, task based frameworks, data acquisition systems
4. **Innovative Workflow and Data Management** solutions for trigger-less data acquisition
5. **High performance data analysis and algorithms**: trigger and reconstruction algorithms, use of hardware accelerators
6. **Innovative Workflow and Data Management** solutions for **Large Scale science**: large datasets, large workloads, heterogeneous resources
7. **Evolution of simulation frameworks**: handling of generation tools, development of fast parametrized detector simulati

TEMI DI RICERCA

Ammessi



Vincitori



Concezio.Bozzi@fe.infn.it

MAY THE FOURTH BE WITH YOU!

INFN AND THE FUTURE OF SCIENTIFIC COMPUTING EPISODE 1: THE HPC OPPORTUNITY

MAY 4TH, 2018



- 51 partecipanti
- Focus su Quantum Computing (talk di IBM, Microsoft, CERN) e HPC
- 4 talk sulle attività più avviate
- il resto lightning per presentarsi
- Di seguito solo alcuni esempi...

Using Hadoop ecosystem tools for distributed datacenters and the ALICE O2 Monitoring

Modular Stack solution for ALICE O2 monitoring

ALICE O2 Facility:

- 268 First Level Processors
- 1500 Event Processing Nodes

Requirements:

- Capable of handling O2 monitoring traffic – 600 kHz
- Scalable >> 600 kHz
- Low latency
- Compatible with CentOS 7
- Open Source, well documented, actively maintained and supported by developers
- Impose low storage size per measurement

Machine Learning as a Service



CHALLENGES

- Reconstruction
- Analysis
- Trigger
- Data quality
- Detector monitoring
- Computing operations
- Monte Carlo tuning
- ...

REQUIREMENTS

- Workflow definition
- Results reproducibility
- Multi-tenancy (scheduling, authentication...)
- Parallel execution and scaling
- Data handling
- Ease of use and management
- ...

IMPLEMENTATION

- Lightweight virtualization
- Modularity
- Flexibility
- Heterogeneous back-end infrastructures
- ...

LEVERAGING

Existing OpenSource software (mature and maintained)



INDIGO-DataCloud products

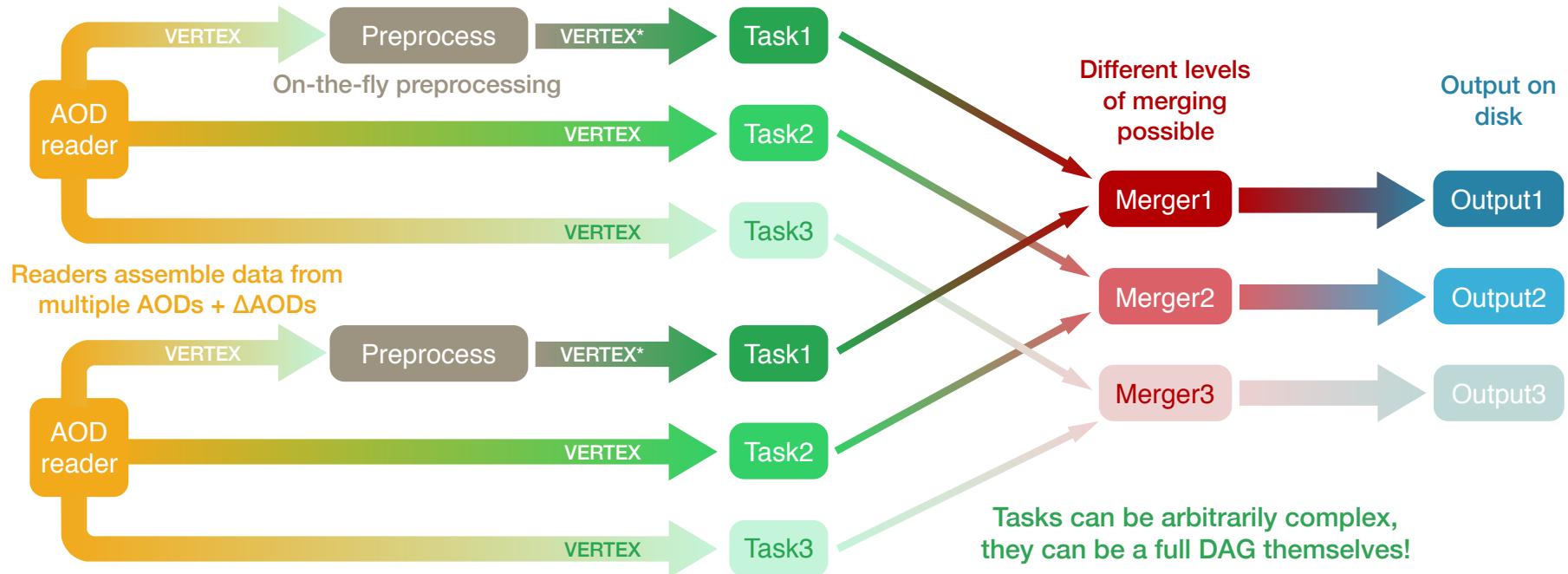


DEEP Hybrid DataCloud products

Workflow handling: the Data Processing Layer

File reading is expensive: readers unzip once and dispatch, tasks subscribe to data

ALICE Run3 software is based on the Data Processing Layer: independent processes exchanging data over the network using message queues. Using it for analysis too

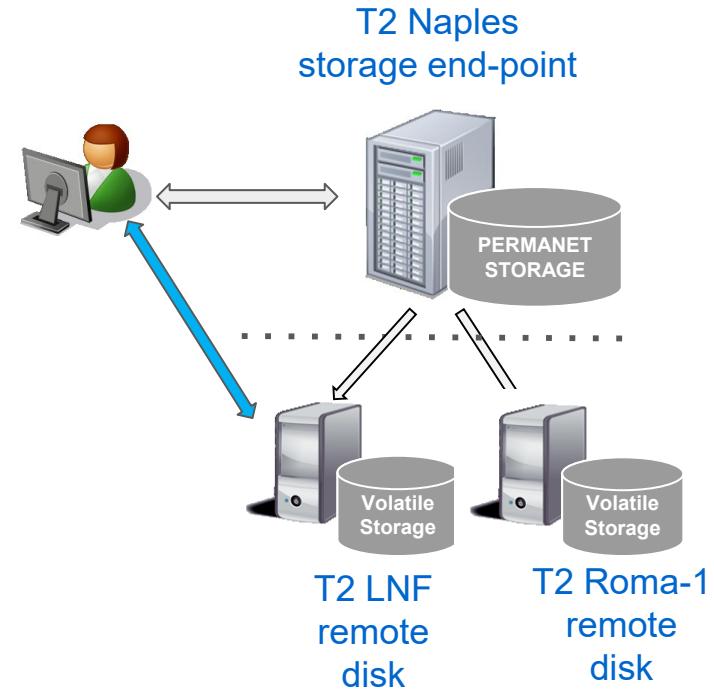


Our implementation

- The Disk Pool Manager (DPM) is a data management solution widely used within ATLAS, in particular in three Italian Tier2.
- The latest versions of DPM are used in our implementation, that offer the possibility to manage volatiles pools to be used as caches.

By exploiting the fast connections between sites, we are deploying a first testbed among Naples, Frascati and Roma-1 using DPM. The aim is to study and develop a configuration in which a primary site represents a single entry point for the entire archiving system and each site can use its storage as permanent storage or as local cache.

Using a cache system the local site administrators can be dispensed from managing a complete storage system. The site became transparent for the central operations of the experiment.



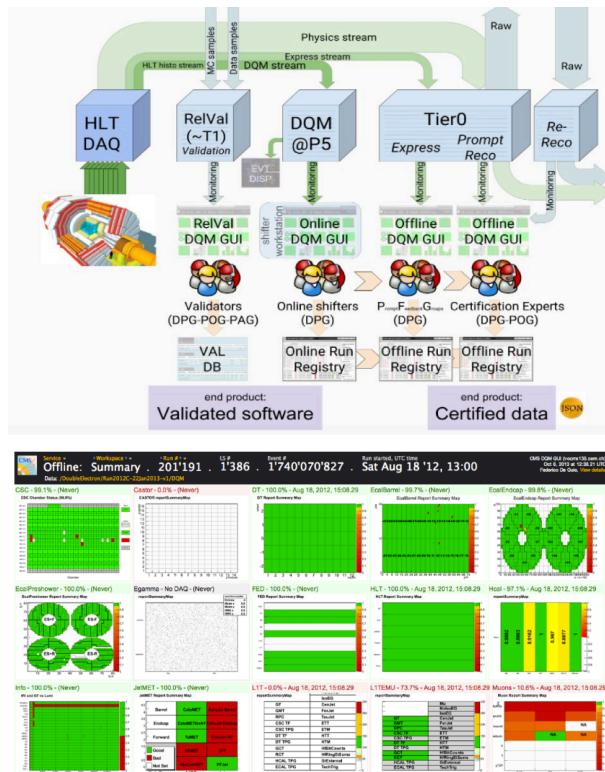
A cache system to support several scenarios

- **Leverage national networking to reduce total maintained storage resources**
- **“Data-lake” approach:**
 - Interposing cache on top of a central custodial site
- **Opportunistic Computing:** to bring **not pledged resources** in the computational model
 - From the experiment point of view: to integrate cloud based resources with **zero effort**
- **Dynamic Site Extension:**
 - **Peak of usage** or more in general buying **external cloud resources**:
 - see activities like: Aruba, Microsoft Azure, HNSci project etc

NOTE: The presented activity lives within the CMS Data Management project

CMS DATA QUALITY MONITORING (DQM) SYSTEM

- A critical asset to guarantee a high-quality data for physics analyses (online and offline)
- Online DQM assess data goodness and identifies emerging problems in the detector
 - Data with poor quality is flagged by eyeballing DQM GUI and comparing a set of histograms to a reference good sample
- Problems with current strategy:
 - Delay: human intervention and tests require collecting sufficient statistics
 - Volume budget: amount of quantities a human can process in a finite time period
 - Human driven decision process: alarms based on shifter judgment
 - Changing running conditions: reference samples change over time
 - Manpower: the effort to train a shifter and maintain instructions



- Abbiamo un pool di persone che daranno sicuramente un contributo significativo al calcolo degli esperimenti a LHC
- Competenze ed esperienze abbastanza diversificate
 - Docker, mesos, event index, big data infrastructure, caching systems, monitoring, fast simulation, event building, deep learning, machine learning, GPGPU, FPGA
- Dobbiamo fare in modo che le loro attività facciano crescere il calcolo scientifico INFN
- Ed estendere l'esperienza oltre LHC

GRAZIE!



Domande?