Commissioning Distributed Analysis at the CMS Tier-2 Centers

Alessandra Fanfani, Giuseppe Codispoti, Claudio Grandi, University and INFN-Bologna, Italy; **Daniele Spiga,** CERN, Switzerland;**José Hernandez**, CIEMAT, Spain; **Sanjay Padhi, Frank Würthwein**, University of California San Diego, US; **Vincenzo Miccio,**INFN-CNAF,Italy**; Fabio Farina,** INFN-Milano,Italy; **Mattia Cinquilli**, INFN-Perugia, Italy; **Tibor Kurca**,IPNL Lyon,France; **Nicola DeFilippis,** LLR-Ecole Polytechnique,France; **Sergey Kalinin** , RWTH,Germany; **Haiying Xu** University of Purdue, US

Abstract

CMS has identified the distributed Tier-2 sites as the primary location for physics analysis. There is a specialized analysis cluster at CERN, but it represents approximately 15% of the total computing available to analysis users. The more than 40 Tier-2s on 4 continents will provide analysis computing and user storage resources for the vast majority of physicists in CMS. The CMS estimate is that each Tier-2 will be able to support on average 40 people and the global number of analysis jobs per day is between 100k and 200k depending on the data volume and individual activity. Commissioning a distributed analysis system of this scale in terms of distribution and number of expected users is a unique challenge.

The CMS Tier-2 analysis commissioning activities and user experience are reported. The 4 steps deployed during the Common Computing Readiness Challenge that drove the level of activity and participation to an unprecedented scale in CMS will be presented. The dedicated commissioning tests employed to prepare the next generation of CMS analysis server are summarized. Additionally, the experience from users and the level of adoption of the tools in the collaboration is presented.

CCRC08 analysis activities

During CCRC08 **[389]** various analysis exercises were performed to gain an overall understanding of the performance and readiness of the Tier-2 sites for CMS data analysis.

Phase 0: preparation

Tier-2 sites get a copy of some datasets to analyze, grant users write access on storage and declare their CPU/disk resources.

Phase 1: controlled job submission

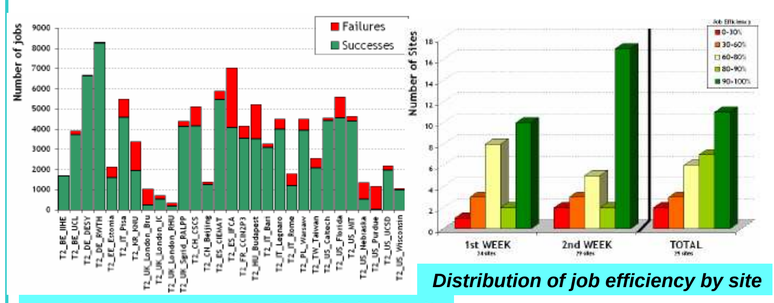
For two weeks, run centrally organized workflows with two types of activities:

•Site performance measurement

- Different jobs with increasing complexity:
  - long-running CPU intensive jobs with moderate I/O
  - long-running I/O intensive jobs
  - short-running jobs with local stage out of O(10MB) file
- try to run at as many sites as possible: up to 38 sites involved across EGEE, OSG, and Nordugrid.
- Very mixed error rates, ranging from less than 1% at many sites to up to 50% at a few sites due to catastrophic storage failures. Overall success rate ranged from 92-99% for these exercises based on more than 100,000 jobs submitted.

•Simulation of physics groups workflows

- Mimic some realistic activity of a physics group:
  - submit via CRAB server analysis-like jobs running for about 4 hours with remote stageout of a O(20MB) file to selected Tier-2 sites
  - about 105000 jobs were submitted in 2 weeks
- Focus on sites that are considered commissioned: up to 29 sites.
- Most failures were problems accessing the input data (0.1%-10%) and staging out remotely the output (due to old grid clients and promptly fixed). During the second week, sites with efficieny above 90% significantly increased.



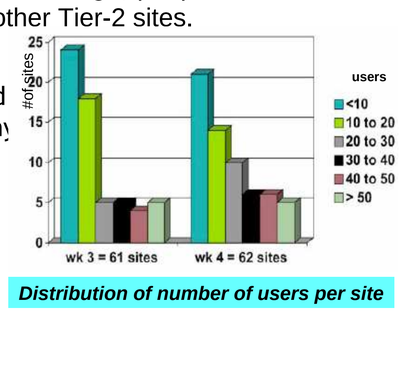
Number of successful and failed jobs by site

Distribution of job efficiency by site

Phase 2: chaotic job submission

During the last ten days of the challenge, people at Tier-2 sites were encouraged to submit to other Tier-2 sites.

This activity was clearly visible in the CMS Dashboard showing lots of users at many sites.



Distribution of number of users per site

Phase 3

•"Stop watch" exercise

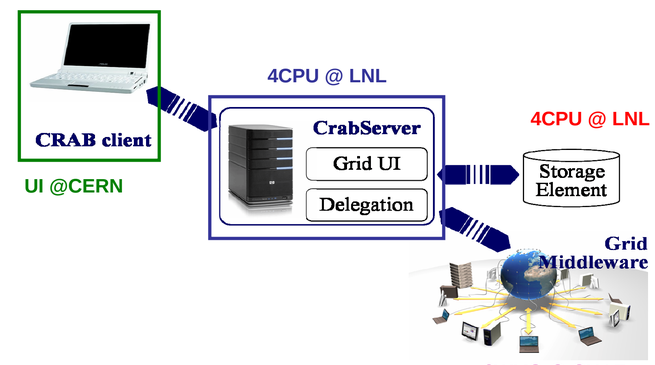
Each Tier-2 measures the total latency of its operation from dataset download from to completion of the analysis jobs. Very different results across sites: from few hours to days, dominated by the time to complete the dataset transfer and job competition at site.

•"Local scope" DBS exercise

It consisted in publishing user data produced at Tier-2 sites in a private DBS and run analysis jobs on them. It was successfully demonstrated in the context of a simulation exercise for the Electroweak analysis group.

Commissioning tests of CRAB server

CRAB server **[77]** is an Analysis server developed for CMS distributed analysis. During CCRC08 the aim was testing the readiness of sites, by filling their resources, rather than a scale test of the CRAB server.
To test the actual server scalability and reliability up to the expected CMS operational rates a dedicated test environment was set up:



CRAB server services and test setup

Phase 1: Single user phase

A single user performed a controlled job submission pattern:

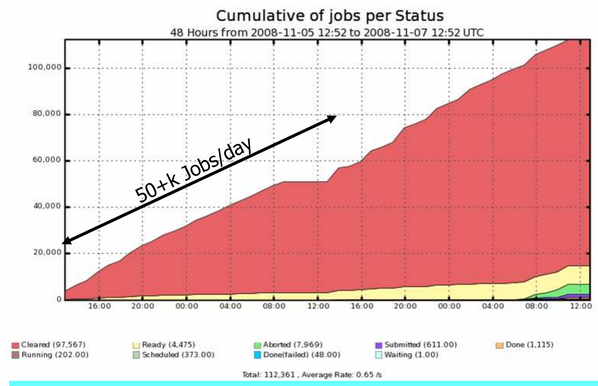
- Constant rate of 500-600 jobs every 20min
- Peaks of 1-2K jobs every 5h
  - Short-running jobs, not reading a dataset and without stage-out

• More than 200k jobs in less than a week: above 40 Kjobs/day
• More than 50 sites used
•The CRAB server was able to cope with that rate with no indication of reaching a breaking point.

Destination Sites Distribution of ~200Kjobs in Phase 1

Phase 2: Multi user environment phase

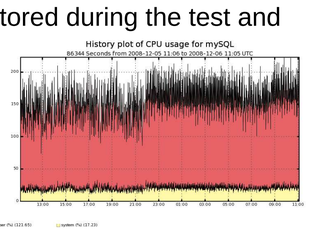
Emulation of a multi-user environment using 12 user certificates each with different submission pattern.
• Submission rate was increased to above 50,000 jobs/day



Cumulative distribution of jobs submitted to CRAB server during the multi-user test phase

• No CRAB server misbehaviour identified due to the multi-user environment
• Some limitations in components handling user's output were identified and taken into account in the development cycle

The CRAB server services were monitored during the test and the breakdown of CPU load usage is



Monitor of the CPU load for MySQL

- 2 CPUs for accessing the underlying database (MySQL)
- ~ 1.5 CPUs for handling users in/output sandbox(GridFTP)
- ~ 1 CPU for all the CRAB server components

outlining the need of at least a 4 CPU machine

Reference analysis usage:

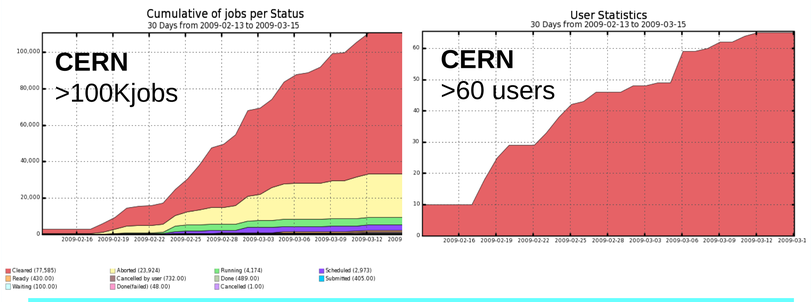
- Currently the whole CMS analysis: 30K jobs/day
- CMS Computing model target: 100-200Kjobs/day

Some CRAB server instances deployed at different sites to serve physic's group activities and regional community can cope with analysis needs.

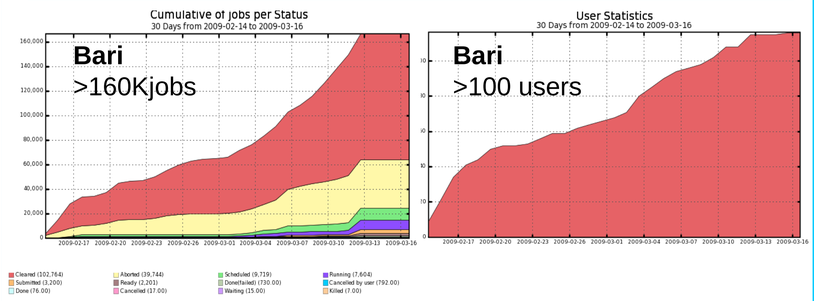
CRAB server Deployment and usage

CRAB server instances have been deployed since few months in several countries:

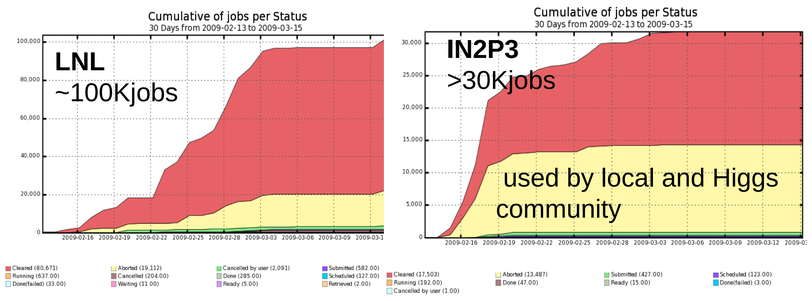
- 4 instances in production (CERN, Italy, France) open to worldwide distributed CMS users or local communities or specific physics group
- 3 test instances (Diego US, Purdue US, Germany)



Cumulative jobs (left) and users (right) of CERN server during last month



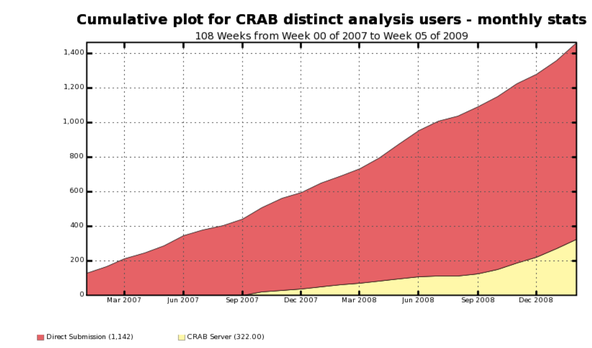
Cumulative jobs (left) and users (right) of Bari server during last month



Cumulative jobs of Legnaro (left) and IN2P3 (right) server during last month

With CRAB server deployment:

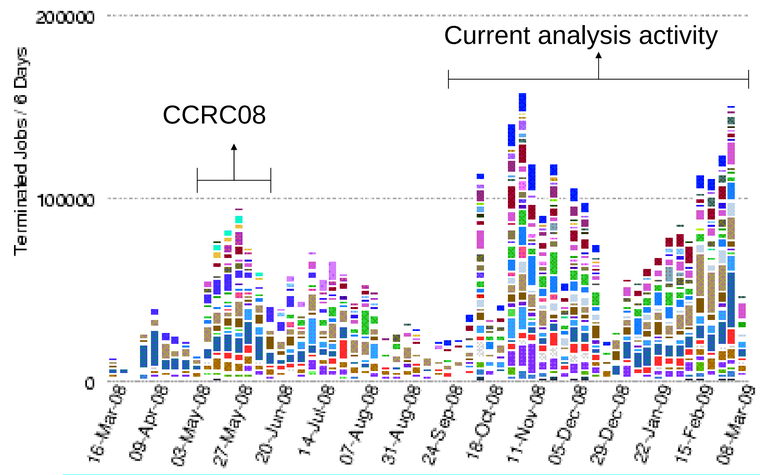
- spontaneous increase of users switching to use CRAB server
- >20% users currently use CRAB server



Cumulative number of distinct users per month since 2007 , with an increase of CRAB server usage

Analysis activity at Tier-2s

More than 10Milion analysis jobs were submitted during last year:
• CCRC08 drove the level of activity to an unprecedented scale in T2s
• Users analysis has now surpassed that scale



Number of analysis jobs by Tier2s during last year from CMS dashboard (each color refers to a site)

Analysis job efficiency is roughly 60%. Most of the failures faced by the users are ([207]) :

- remote stageout issues and user application errors
- jobs aborted by the grid, often due to site problems
- few % of failures reading data at site

Conclusions

Support of distributed data analysis is very challenging because of its diversity in dimensions. During CCRC08 several analysis exercises were performed to test the performance and readiness of the Tier-2 sites for data analysis to an unprecedented scale. To reach an high automation level in user analysis an Analysis server has been developed. Dedicated tests have demonstrated the CRAB server reliability and scalability up to 50Kjobs/day. Some CRAB server have been deployed and are in production to serve worldwide distributed CMS users or local communities. The level of adoption of CRAB server is increasing and this will give indications about further automation and users needs. Current analysis activity at T2s has already superseded the scale of the CCRC08 challenge.

[389] D.Bonacorsi,L. Bauerdick "CMS results from Computing Challenges and Commissioning of the computing infrastructure"
[77] D.Spiga "Automatization of User Analysis Workflow in CMS", **[207]** J.Letts et al."CMS Analysis Operations"