

The Advanced Virgo Computing Model (CM): progresses in the last year

Pia Astone

Data Analysis Coordinator of the Virgo collaboration

Introduction: Gravitational Wave Detector's Network

- With two large interferometers, Advanced Virgo, near Pisa, and HF-GEO, near Hannover, Europe is at the forefront of research on gravitational waves. NSF supported the construction of two Advanced LIGO (aLIGO) detectors in US.
- LIGO Scientific Community (two aLIGO and GEO) and the Virgo Collaboration have a common goal: the detection of GW and their use as an astronomical tool.
- Data taking at reduced sensitivity will start in 2015 (aLIGO only, commissioning for AdV). In 2016 we will have a 6 months run. Planned sensitivity will be reached in 2018 (with a 9 months run). 1 year run from 2019.
- This network is going to expand with the Japanese detector KAGRA (2018) and Indigo in India (2020).

Introduction: Gravitational Wave Data and Analysis

- Each detector will produce $O(15\text{MB/s})$ of data, continuously during science runs, with duration of few months up to one year or more.
- Data are distributed to various CCs (Virgo data at CNAF and CCIN2P3) where they are analyzed.
- Several smaller facilities receive and analyze a fraction of these data.
- Three main kinds of analysis:
 - In-time analysis for detector characterization over hundreds of channels;
 - Low-latency searches: scientific data of all the detectors of the network transferred and analyzed within $O(10)$ minute or less, to produce triggers for EM and neutrino follow-ups;
 - Off-line searches, often computationally very demanding (sometimes computationally bounded)

Introduction: some considerations

- We need to face the problem of providing storage resources for permanent data archiving, fast data access and large computing resources for analyses.
- A robust data distribution and access framework and of transparent access to geographically distributed computing resources is a crucial achievement for the collaboration
- In the advanced detectors era the computing burden will increase due to the increase in sensitivity, run time and in the number of detectors.
- Moreover, the analysis of GW data in conjunction with those of telescopes and neutrino detectors, will increase the number of users and of scientific projects devoted to data exploitation.

The basic organization of computing

- EGO site at Cascina, where the detector is, hosts the Tier-0
- The detector “primary data” are distributed to Tier-1s, CNAF and CCIN2P3 (one full copy to each) with a maximum latency of 1 day
- CNAF, CCIN2P3 and LIGO clusters are the main places where offline analysis are done. Follows an important comment on the use of LIGO resources (slide 9)
- The EGO farm is fully dedicated to data production, commissioning, detector characterization and “low-latency” scientific searches (for which we need to release fast triggers to our partners for EM (electromagnetic) follow-up)
- We have recently been asked to investigate on the possible usage of other external computing resources, in addition to the national computing centers. We are investigating on a possible role for Wigner (Hungary is in Virgo). But not only (resources from Holland ,Poland)

Management for Computing

- Historically, the Data Analysis coordinator was also the Computing Coordinator
- Since last October the organization has changed and we have a Computing Coordinator (Gergely Debreczeni). Very positive and important for the collaboration.
- Computing Model has been discussed with the “External Computing Committee” (chair: Manuel Delfino. CNAF and CCIN2P3 directors part of the committee). We will do the same for the Implementation Plan. Many useful discussions have been held with the committee
- To help and ease the technical discussions with our external CCs we have now a committee : CTCC (Computing Technical Coordination Committee). This is the forum which we use for joint discussions (Luca Dell’Agnello is here for CNAF, Rachid Lemnari for CCIN2P3)

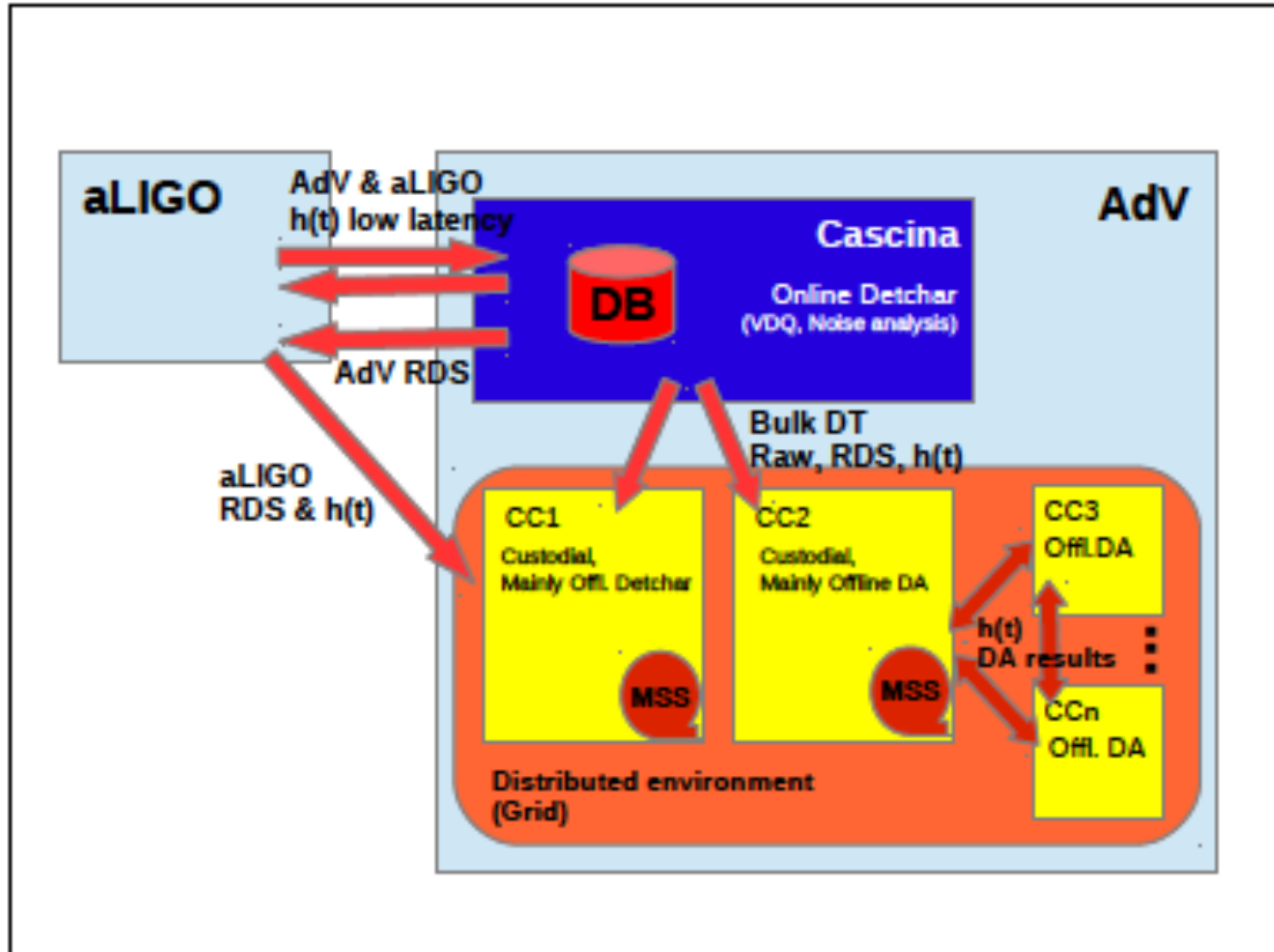
INDEX of the CM

- I. Computing and DA (‘ ‘ Data analysis’) Workflows.
- II. Data Model.
- III. Data management, distribution and access.
- IV. Software description and management. This section contains details on milestones and responsibilities for each project.
- V. Computing Facilities resource requirements (Storage and CPU foreseen needs for next years).

We have then prepared the Implementation Plan (IP), where technical solutions are presented (we use Redmine, a project management application). It has also Milestone Summary Tables and Manpower Tables, for the activities in the CM.

And we have written the Management Plan, with the procedures to update the CM and the IP

Data workflow for Scientific analysis and Detector characterization (Fig. 1 of the CM)



The aLIGO Computing re-organization

- The NSF has mandated a thorough review of LIGO and LSC computing requirements for the ADE. The first step in this process was a review held on 8 May at NSF headquarters to present the Advanced LIGO construction project's proposal for acquiring hardware in preparation for the first science run near the end of 2015.
- The presentation also covered the later years 2016-2018, which included plans for expanding CBC searches to include spinning, non-aligned BH-NS and BH-BH systems. The projected computing requirements for these systems, using the full projected sensitivities of the ADE instruments, were deemed extremely large in view of anticipated NSF resources during this epoch.
- The review committee found that plans for the first science run were sufficiently modest to be justified. However it recommended that before any larger scale procurement or shared resource requests are approved, the LSC and the Laboratory must pay greater attention to optimizing and streamlining search code performance and performing a broader set of benchmark tests of search codes on platform technologies that were not presented during the review.
- The NSF has requested that LIGO Laboratory undertake a series of actions in the next 3 - 4 months to address the review committee's recommendations.

Data	CNAF [TB]	CCIN2P3 [TB]
Raw data	745	745
AdV RDS	11	11
LIGO RDS	22	22
Trend data	1.5	1.5
Minute trend data	0.25	0.25
AdV h(t) and status flags	3	3
MDC h(t)	9	9
Calibration output	1	1
Omicron triggers	–	4
DQ veto production data	–	negligible
Spectrogram data	–	1
MonitoringWeb data	–	0.8
DQ developments data	–	0.5
DQ segment	negligible	negligible
NoEMi data	12	–
BURST	15	3
CBC	4.5	0.5
CW	25	–
STOCHASTIC	–	3.6
Total	849.5	802.8


Storage @CCs

(tapes and cache disks)

850 TB/yr CNAF

803 TB/yr CCIN2P3

 Transferred from Cascina

 Produced at the CCs

We plan to store on disks only the last run data or the last year. Thus only the need for tapes storage increments as data are produced

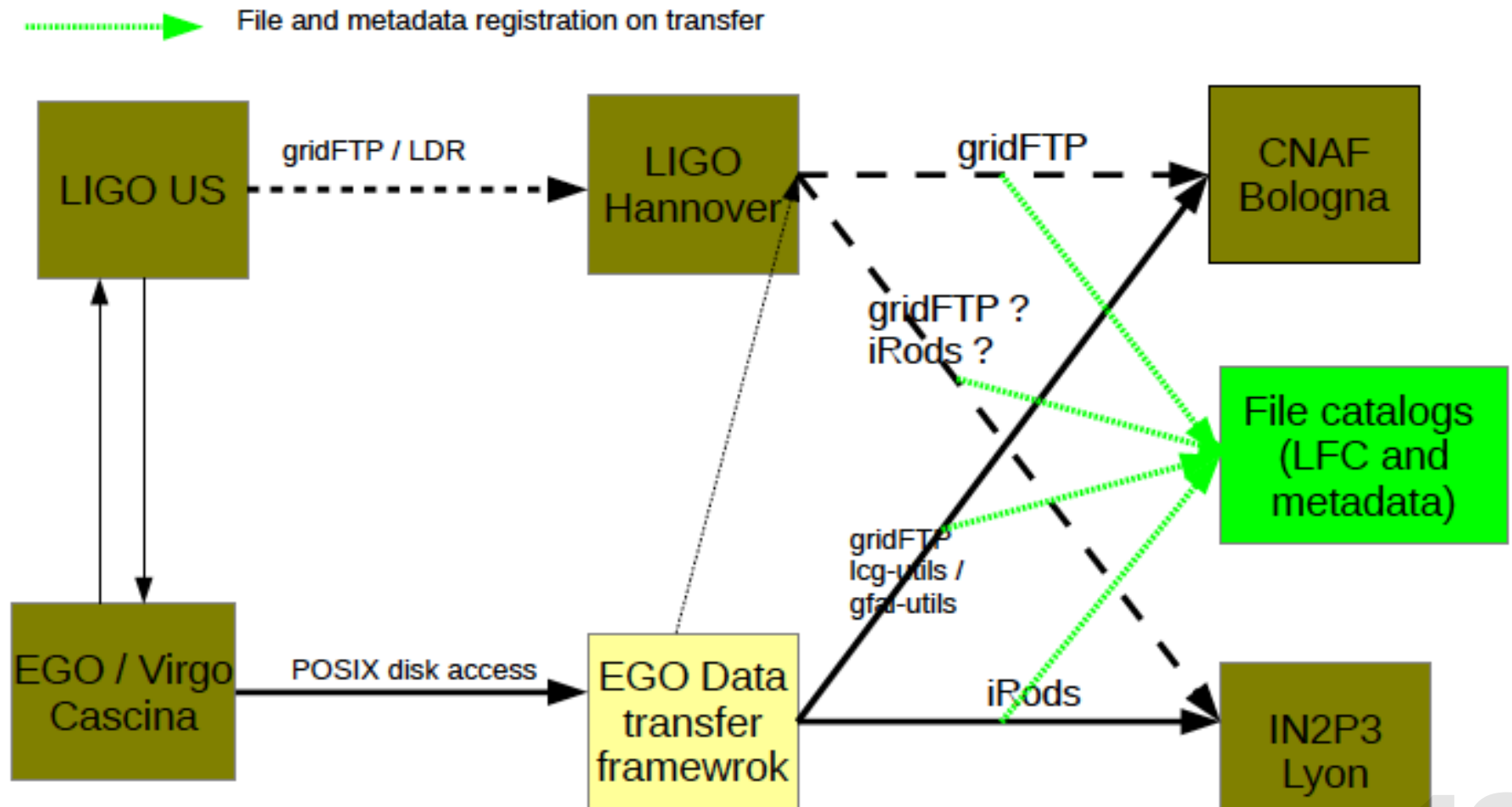
10

Data Transfer

- EGO IT Department developed a DT framework, which is able to use multiple backend.
- Choice of backend depends on the destination site - sites give maximal support if we use their preferred tool for DT:
 - iRods for CCIN2P3
 - gridFTP for CNAF
- After transfer, the physical location of the files (site dependent) and their metadata are registered in a file catalog which is able to answer to typical queries of the analysis pipelines (gps time, channel, interferometer, data type, etc.).

Type of catalog to be discussed, should if possible be aLIGO (LDR) compatible. In any case a simple LFC catalog will also be available for jobs for registering various datasets not part of the bulk data transfer (intermediate results, preprocessed data, etc..)
- Local data access is also needed,
 - native POSIX access is available at CNAF
 - CCIN2P3 provides POSIX access after simply sourcing XrootD environment

Data Transfer Model



- ➡ Low latency data transfer (memory copy) managed by CM (Max latency: minutes)
- ➡ Multi-backend bulk data transfer managed by EGO framework (Max latency: 1 day)
- - - ➡ (Possibly) 3rd party bulk data transfer managed by EGO framework (Max latency: 1 day)
- - - - ➡ 3rd party copy control channel, no data flow
- ➡ Ligo data transfer, not Virgo responsibility

Data Transfer Flux

- 1) AdV data transferred from Cascina to CCs (2-4 TB/day)
- 2) LIGO RDS and $h(t)$ data transferred from one aLIGO cluster to CCs. 60 GB/day
- 3) AdV scientific data –which we call $h(t)$ - are transferred from Cascina to aLIGO and aLIGO $h(t)$ are transferred from aLIGO to Cascina following different rules, defined to guarantee the low-latency workflows for these searches (few seconds !).

These are $O(8)$ GB/day from Cascina to aLIGO;

$O(16)$ GB/day from aLIGO to Cascina (2 detectors)

Computing facilities resource requirements

Part V of the CM

- At regime, the storage needed for 1 year of data (including raw) will be $< \sim 1\text{PB}$;

Pipeline needs in kHS06 power	local	GRID/CLOUD
Detchar Data Quality	1	–
Detchar Noise studies	1	1 ?
BURST	negl	3
CBC	–	33+
CW	–	60+
STOCHASTIC	negl.	negl.
TOTAL	2 ?	97+

Units here are kHS06(power).

The energy kHS06.day needed for 1 yr is power*365

Used here:
1 core= 10 HSE06

Table 3: Summary Table: Estimation of the computing needed locally in the CCs and under GRID/CLOUD at a regime situation (2018+), under certain hypotheses on the parameter space covered. Units are power in kHS06. The “+” indicates that this is the minimal request, with more resources we could cover a wider parameter space

GWTools - the C++/OpenCL based
GravitationalWave data analysis Toolkit - is
an algorithm library aimed to bring the immense
computing power of emerging manycore
architectures - such as GPUs, APUs and
many-core CPUs - to the service of gravitational
wave research.

GWTools is a general algorithm library intended
to provide modular building blocks
for various application targeting the computationally
challenging components of GW
data analysis pipelines.

Web:

<http://www.gwtools.org>

Distributed Job Submission framework: requirements (IP, section 6)

- Should be future-proof, i.e. it has to have a development and maintenance roadmap ensured for at least the following 5 years.
- Should be easy to use for an average user and pipeline developer.
- Should be compatible with any data transfer system to be developed for bulk data transfer
- Should allow various data access mechanisms available on the centers where it enables the execution of the jobs
- Has to have a strict authentication and authorization system in place
- Should be able to handle scientific, relational workflows
- Should be able to directly use or to submit to the biggest european computing infrastructure to the EMI middleware.
- It should be possible to port Various LIGO pipelines without or with minor modifications.
- Should enable job execution, monitoring, logging, etc..
- Should not be specific to any of the target execution site but flexible enough to be extended to any (or many) future - currently unseen or unexpected - computing architecture.
- Should be easy to operate, maintain

Now using or testing: EMI Grid, Pegasus, Dirac

Cloud resources, custom images

As cloud resources are becoming more and more important Virgo has to be prepared to use them.

- ◆ User space software installation and maintenance is not always easy in Grid world.
- ◆ Creating custom images which contains the preferred OS, software packages and Virgo environment could make life easier in many cases.
- ◆ These images then can be submitted to Cloud resources with no problem, Virgo users will find a uniform environment independently where the job is running

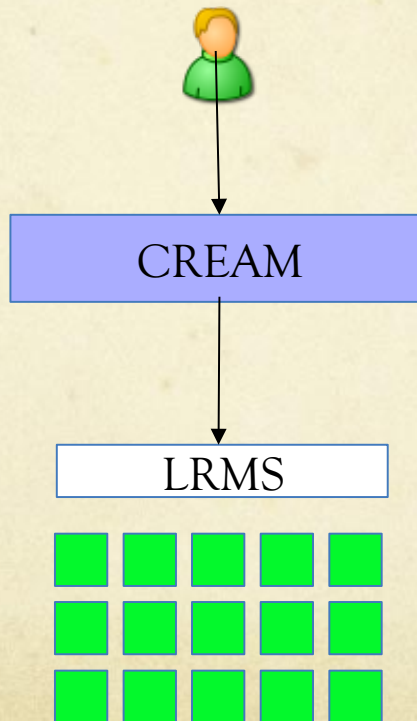
Virgo is evaluating which computational model to adopt in order to face the challenges set up by the new increased and more complex data analysis model. We are preparing some test images.

Lisa Zangrando, INFN PD, is working on some use cases (to begin: Burst searches) for Virgo

The CREAM-CE

CREAM-CE is a Grid component developed by INFN-Padova

- It is a Grid interface towards computational resources handled by a local batch system (e.g. LSF, Torque, SLURM, etc)
- Deployed in several Grid production sites (WLCG and EGI infrastructures)

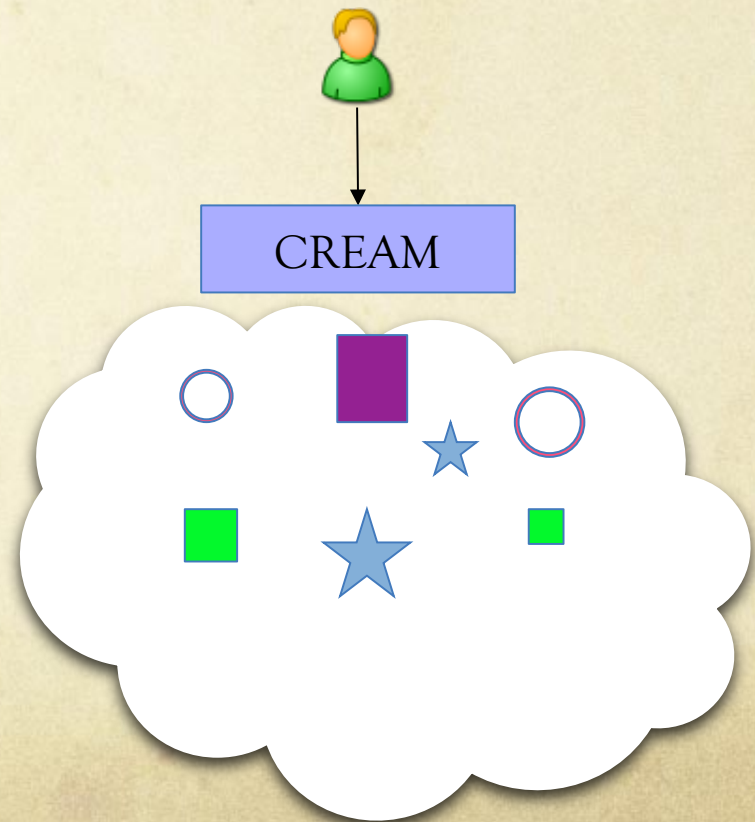


CREAM: GRID & CLOUD integration

Today the CREAM architecture is evolving for enabling the GRID & Cloud integration. It will be a Grid interface towards computational CLOUD resources handled by CLOUD IaaS framework (i.e. OpenStack)

- Nothing changes for the Grid users (no updates of the WS interfaces are foreseen)
- Resources (virtualized) will be allocated and then destroyed dynamically by CREAM as needed
- Grid user can describe which kind of resources she needs (e.g. cpu, ram, os, storage, software, etc) wherein its job will be executed
- User can even upload its own VM image (i.e. customized job execution environment)

See tomorrow's presentation
“Scheduling in Openstack”

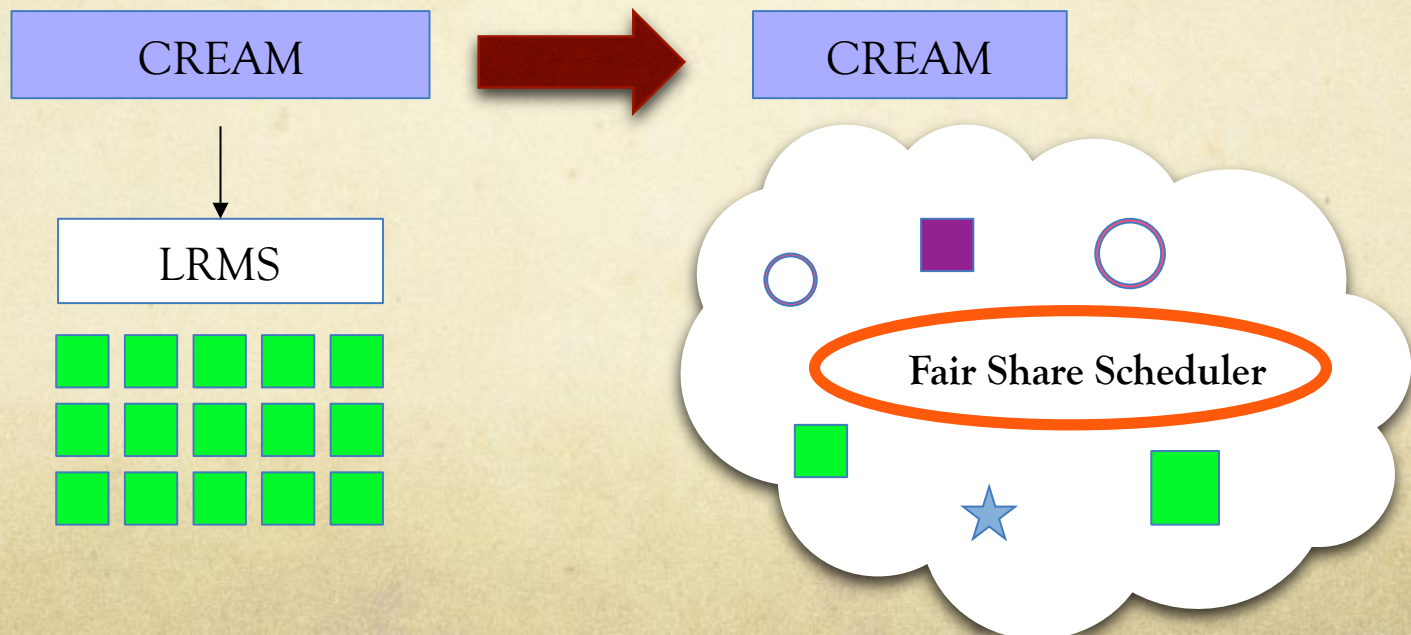


CREAM: GRID & CLOUD integration

The batch system is not needed anymore

- Fair-share scheduling strategy and request “queuing” mechanism provided by the new scheduler (FairShareScheduler) for OpenStack-Havana developed by INFN-PD
- it solves the OpenStack issue occurring whenever the resources are fully allocated
 - ◆ enhanced request management and resource allocation
 - ◆ support of different QoS levels (i.e. priority)
 - ◆ enforcement of specific policies based, for example, on the effective resources usage

See tomorrow's presentation “Scheduling in Openstack”



Alternatives, backups, extensions I

Might be interesting to check...if we had the manpower

There are many valid scientific case for high scale, compute intensive (not data intensive), important but not time critical analysis.

These analysis can make use of „spare computing resources” very cheaply. For such a resource a good example is Amazon's Spot Price offer [1], which gives a

- High Frequency Intel Xeon E5-2670 (Sandy Bridge) Processors*
- SSD-based instance storage for fast I/O performance
- Balance of compute, memory, and network resources

These resources can be much cheaper than any pre-allocated or on-demand grid resources, as such they are ideal for well understood, tested and optimized scientific pipelines which needs compute power.

[1] <http://aws.amazon.com/ec2/pricing>

Alternatives, backups, extensions II

Might be interesting to check...if we had the manpower

Mobile devices are dominating the world in number (not yet in compute power, but...)

- ◆ There are quite some very simple algorithm with high arithmetic density which could be easily ported to various mobile OSes such as Android
- ◆ „Compute-while-on-plug” style volunteer (or paid) contributions could have manifold benefits starting from Outreach, Public Relations to scientific results.
- ◆ An Einstein @ Phone style framework is under consideration but requires lot of manpower to develop.

- We aim at enhancing the role of Europe in the development of Computing and Data Analysis strategies for the search of GW.
- This is a fundamental step in view of setting-up a world-wide network of GW search communities.
- Main items:
 - ❑ Data distribution within GW community; setup of a common infrastructure to distribute and catalogue ITF data and scientific metadata;
 - ❑ Data remote access;
 - ❑ Transparent access to computing resources (common job submission and monitoring framework);
 - ❑ Software automatic distribution and installation at analysis sites;
 - ❑ Open data; potentially of interest to many outside GW community (astronomy/astrophysics, geology,...);
 - ❑ Security issues;
 - ❑ Data preservation;

BACKUP slides

External Computing Centers (CC)

○ JECC (Joint External Computing Committee)

The agreements with External Computing Centers are defined at the level of the Virgo and EGO managements, through discussions and meetings of the JECC group. JECC is composed of the collaboration Spokesperson, the EGO Director, the head of the EGO IT Department, the data analysis coordinator (DAC), the computing coordinator, the two Virgo reference persons in each external CC and by one person appointed by the directors of each CC.

External Computing Centers (CC)

- CTCC (Computing Technical Coordination Committee) All the technical discussions, aimed to help the collaboration to find good solutions, common to the different CCs when applicable, are done within the CTCC group. Asked by ECC, STAC and approved by the Council (July 2013).
- Composed by the DAC chair, the Computing Coordinator, the head of EGO IT Dept., and one or more experts representative of each external CC, nominated by the Directors (actually: Luca Dell' Agnello, Rachid Lemrani).
- We have set up a mailing list, ctcc@ego-gw.it, open to the members and to expert members of the collaboration. It has proven to be a very efficient and rapid method to address questions.
- It is important to try to exploit common solutions at the different CCs and/or to share different approaches they might want to suggest to the collaboration

Towards the CLOUD paradigm

Today the Scientific Computing is moving towards the CLOUD paradigm

- GRID is a consolidated technology but suffering of too much limitations
 - static resource allocation and partitioning, homogeneous environment for job execution, etc
- CLOUD is a recent technology not yet enough mature for replacing definitely GRID
 - its paradigm assumes all resources are unlimited
 - This assumption is too strong: not realistic for small/medium CLOUD infrastructures
 - In Openstack if none resource is available, the user request will fail and forgotten
 - The scant provisioning of CLOUD providers may imply:
the need to manage efficiently their limited set of resources
definition of different QoS (Quality of Services) levels in order to favor the requests of high-quality users at the expense of others less privileged or the enforcement of specific policies based, for example, on the effective resources usage.

No one open-source IaaS vendor (e.g. Openstack, OpenNebula) is addressing such issues
Virgo is evaluating which computational model to adopt in order to face the challenges set up by the new increased and more complex Virgo's data analysis model

Data access

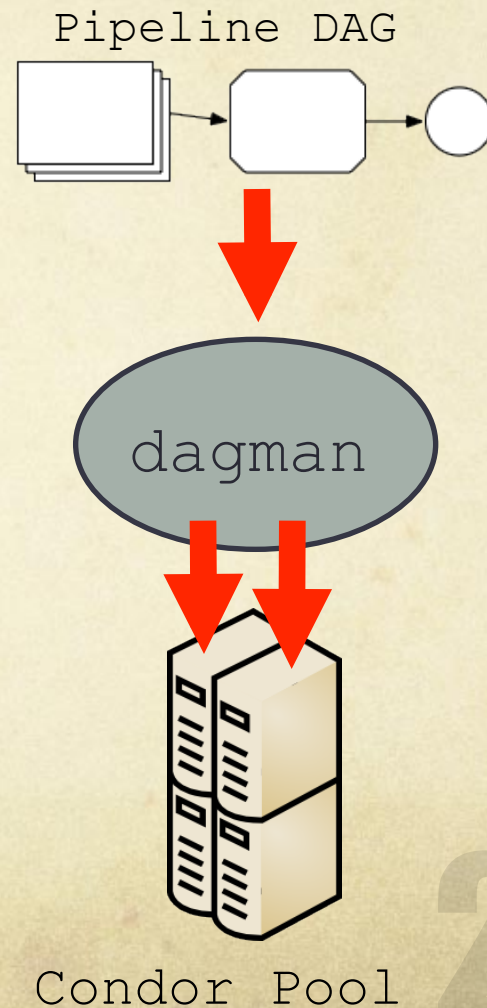
- ◆ Seems not to be a problem, except the fact that there is a quite big discrepancy (40%) between a typical job's wall clock time and CPU time which is caused by the slow local I/O. This is probably due to staging or other tape/disk bottlenecks which we need to solve.
- ◆ Local data access
 - λ Native POSIX access is available at CNAF
 - λ CCIN2P3 provides POSIX access after simply sourcing XrootD environment
- ◆ Remote data access. Very similarly
 - λ CNAF provides remote gridFTP access to data
 - λ CCIN2P3 provides remote iRods access

Moving pipelines (CBC group) to run at Virgo CCs

Current CBC pipelines are designed to run on a condor-based cluster (LIGO DataGrid).

The pipeline is described by a DAG, showing which jobs need to be executed in which order.

condor_dagman sends jobs to a Condor Pool, a collection of compute nodes with access to data, executables etc



Moving pipelines

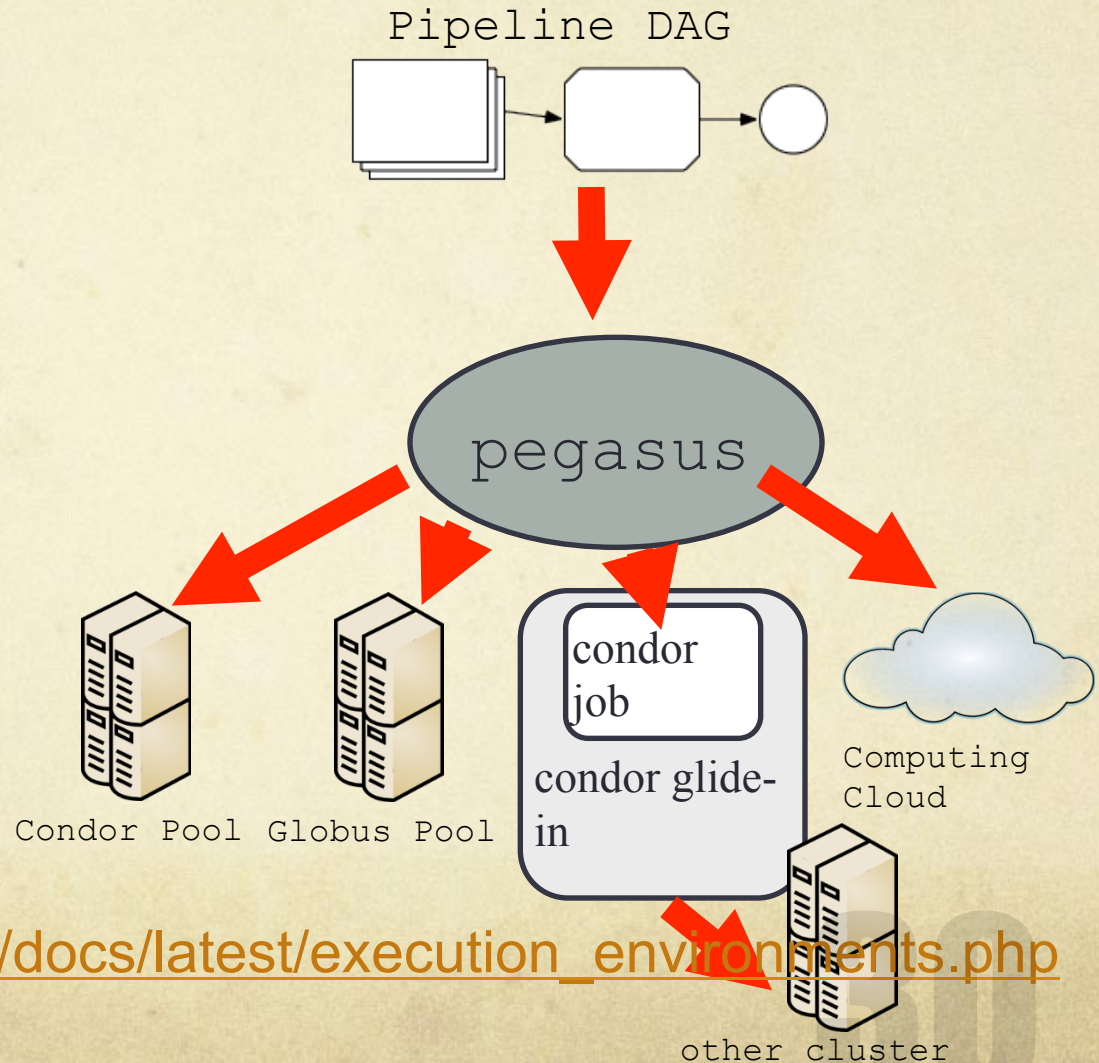
Pegasus acts as a middle layer that can send its jobs to different targets

Condor pools

Globus clusters

Other clusters using condor glide-in

Cloud computing (e.g. Amazon EC2)



http://pegasus.isi.edu/wms/docs/latest/execution_environments.php

for more details

28/05/2014

Moving pipelines – status for CBC

Started with LALInference parameter estimation pipeline, as it is simpler than ihope (search pipeline) and with TIGER, for GR tests

- ✓ **M**odify pipeline to create Pegasus DAX as well as Condor DAG
- ✓ **E**xecute DAX on LIGO DataGrid condor pool (local access to data, executables)
- **E**xecute DAX on LIGO DataGrid without using local resources (ligo_datafind etc)
- **E**xecute DAX inside a condor glide-in on CNAF cluster.