Il modello di calcolo di Advanced Virgo (AdV)

Pia Astone, Virgo Data Analysis Coordinator

CCR workshop. Genova 2013/05/30

AdV will begin data taking in early 2016.

- 2016/2017: 6 months;
- 2017/2018: 9 months
- 2019 ->: 1 yr

When the detector takes data the Duty Cycle is on the average ~ 80%

AdV is part of a network of many detectors

→ need to transfer, archive, analyze in our CCs also (some) aLIGO data . aLIGO will begin data taking in 2015

We are preparing the CM now

(and the Implementation Plan, where technical solutions are detailed)

→ we need tight interactions with CCs

- Computing and DA (``Data analysis") Workflows
- Data Model
- Data management, distribution and access
- Software description and management. Here FTEs and milestones for the projects will be detailed.
- Computing Facilities resource requirements

(Storage, CPU needs: preliminary estimations ready)

Red = done; blue = to be completed

- AdV has a hierarchical model for data production and distribution.
- Cascina (EGO) hosts the Tier-0.
- ITF ``Primary data” are distributed to Tier-1s: CNAF and CCIN2P3 (one full copy in each CC), maximum latency of 1 day.
- CNAF, CCIN2P3 and LIGO clusters are the main places where the offline science analyses are done.

Workflows (Part I of the CM)

AdV data analysis activities can be summarized into three main categories, for which we have detailed the workflows in the CM:

- Commissioning and calibration;
- Detector characterization - ``detchar”: data quality, noise studies;
- Scientific analysis divided into 4 main groups: Burst, CBC, CW, Stochastic searches.

Workflows @ Tier-0 (EGO, Cascina)

- The farm is fully dedicated to data production, commissioning, detector characterization and ``low-latency” g.w. searches.
- Detector characterization work gives important support to commissioning and scientific analysis:
- ``Online” (very small latency, of a few seconds) or

 ``in-time” (latencies up to a few hours, but well defined)
- The only scientific analyses which will run in Cascina are the ``low-latency” g.w. searches, where we need to release fast triggers to partners of the astronomical community for follow-up.

Data management, distribution, access at Cascina Part III.

At Cascina, we need a data management system able to:

- Store the data on the main storage system for local access (a circular buffer O(400 TB/yr)
- Backup selected data for crash recovery;
- Transfer to the Tier-1s, following given requirements;
- Guarantee the I/O needs of all the workflows running there.

The new farm @ EGO (Cascina)

- We are now running benchmarks to compare different baseline OS or different virtualization configurations. Tests on the storage and network I/O will follow.
- A preliminary collection of needs for all the DA workflows @ Cascina (detchar, low-latency science searches) gives approximately the need for a power of O(2 kHSE06).
- The idea is to have a multiprocessor farm with virtual machines installed dinamically when needed, in order to optimize the use of resources. And, an extension of the virtualization, will be an interface to Cloud:
  - In view of a possible extension to external IaaS (Infrastucture as a Service) cloud resources, the new farm and the network should be "Cloud-ready" since the beginning

Basic model for data transfer (DT) and access (DA) at the Tier-1s

- We would like to have the same DT and DA procedures at CNAF and CCIN2P3.

- For this reason the CM supports the use ``GRID tools”
(currently LCG tools, but we would try to be ready for any evolution of the grid), which would allow us to use existing codes/procedures
- We need anyhow to guarantee to users also the interactive access (also only on virtual machines) and the usage of local batch systems.
- For CNAF this is not a problem because files transferred via SRM are also accessible via POSIX; at IN2P3 in order to maintain the actual Xrootd access we should install a layer ("Xrootd door") between SRM and Xrootd.

Storage @Ccs

(tapes and cache disks)

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

850 TB/yr CNAF
803 TB/yr CCIN2P3

Transferred from Cascina

Produced at the CCs

Need to better organize the usage of tapes, to reduce the data access times

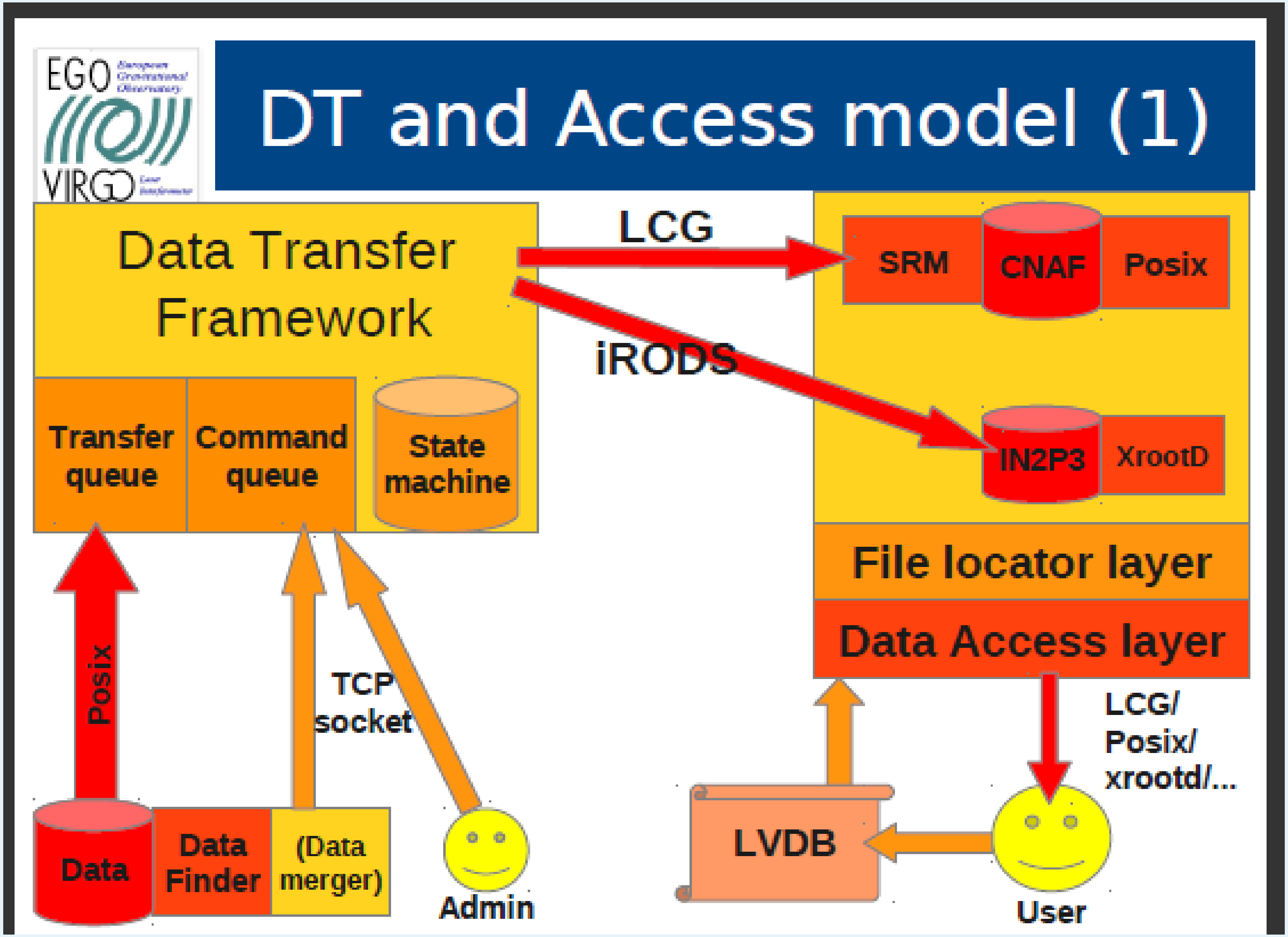
Data Transfer. Part III of the CM

- 1) AdV data are transferred from Cascina to CCs (~=2 TB/day)
- 2) LIGO RDS and h(t) data are transferred from one LIGO cluster to CCs (now testing from Hannover to CNAF). 60 GB/day
- 3) AdV h(t) are transferred from Cascina to aLIGO and aLIGO h(t) are transferred from LIGO to Cascina following different rules, defined to guarantee the low-latency workflows for these searches (few seconds !).

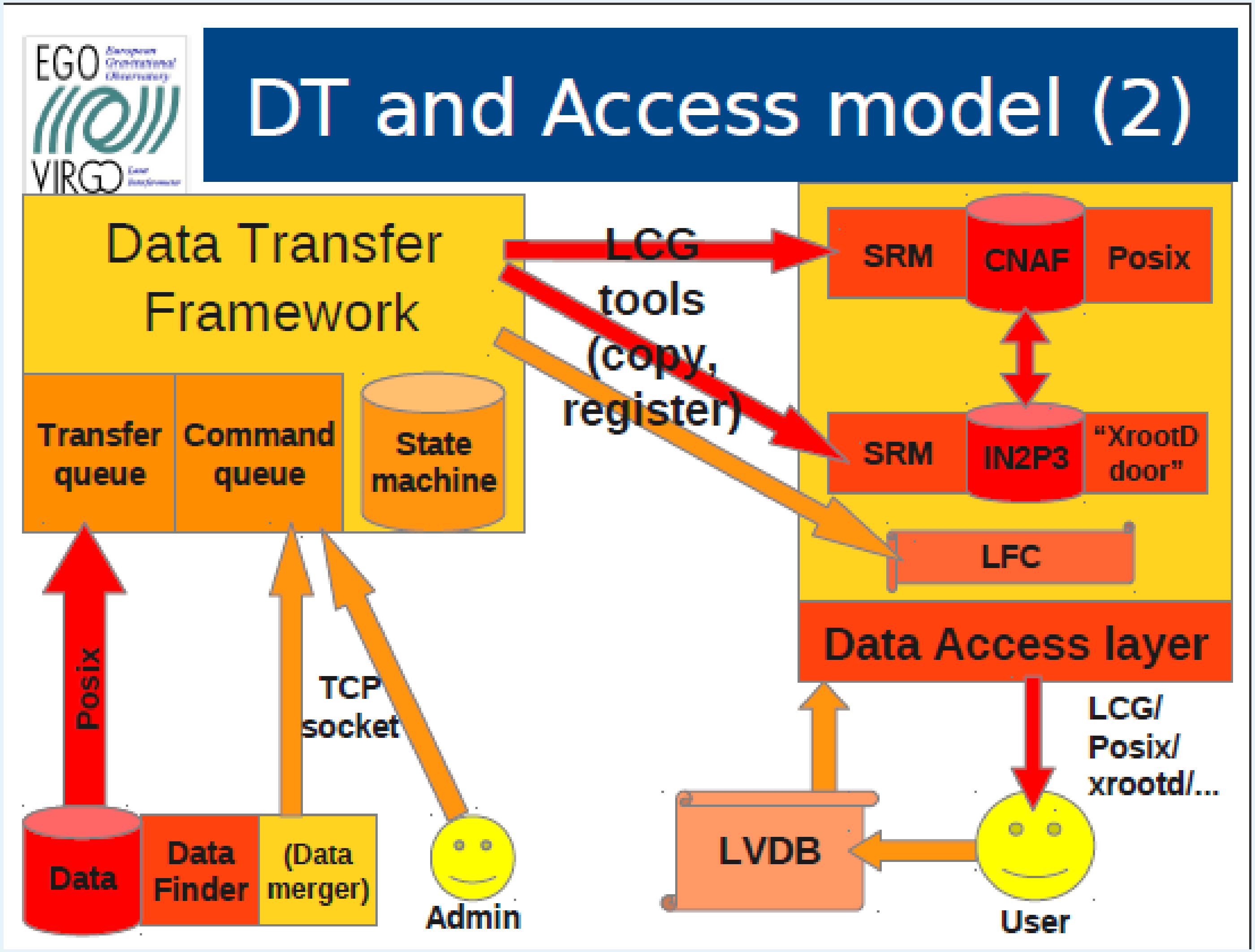
  These are ~ 8 GB/day from Cascina to aLIGO;

  ~16 GB/day from aLIGO to CNAF and Lyon

First possible scenario (Data here might either be AdV or aLIGO)



Second scenario (Data here might either be AdV or aLIGO).
This is my prefferred scenario, as it maximizes the use of common tools



Data Access (local and remote) @ CCs

- Among the data in the table, the ones for which we have the need of parallel fast Data Access are those transferred from Cascina or from LIGO clusters, which have dimensions of the order of (1-1.6) GB each.
- We will use DataBases, for bookkeping information (LVDB) and for the files locations (if possible, lfc). Need to understand scalability and syncronization issues.
- And we will organize a ``file locator service”.
- We wonder if an alternative for remote DA might be possible with ``Storage Federations”, at least for some use case: e.g. only scientific analysis using the basic g.w. Channel (3 TB/year/detector)

Computing needs at the CCs. Preliminary estimations

Pipeline	Cores per 1 year	kHSE06	kHSE06.day
CW All-Sky τ min = 10000 yr	330	3.3	1200
CW All-Sky τ min = 1000 yr	33000	330	120000
CBC main analysis Ihope + gwtools	300	3.0	1100
CBC Parameter Estimation and GR tests	1400-2800	14-28	5000-10000
Burst cWB offiline	300	3.0	1100

1 year of data at duty cycle of 100%. Analyzed in one year.
For reference: LHC @ CNAF has used a power of 88 kHSE06

As presented in the table, we have some analysis which are high computationally demanding, and in some cases we need to limit the scientific goal for computational problems.

We are actually running under GRID (CW searches), or preparing the software to be able to run under GRID (CBC and Burst searches). Tests @ CNAF for the highest demanding pipelines, where the workflows were tighten to the LSC DATA GRID and an important work for the porting is needed.

What to do to be ready by 2015/2016 ? Ready to run even very high demanding pipelines @ our CCs ?

We are exploiting two solutions:

1) prepare some GPU based libraries and codes
2) test the DIRAC solution for the GRID/CLOUD

Dirac: Distributed Infrastructure with Remote Agent Control

- We would like to test the Dirac solution- suggested to us by CNAF colleagues- for the codes which already ran under GRID @ CNAF.

- We are not expert of Dirac (nor of CLOUD). We would need support or help from people at the CCs and/or from colleagues from other experiments who have or will do the same

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

GWTools delivers remarkable speed-up factors for various use cases. Depending on the algorithmic composition of the application. Typical acceleration factors range from 10x to 150x.

Examples:

1. Fast Fourier Transform: x150
2. Various vector operations: x100
3. Template generation: x130
4. Trigger clustering: x12
5. Maximum finding: x25
6. PDE integration: x4 – x50

Exact numbers strongly depend on the CPU and GPU used for comparison. Numbers above are shown for single threaded single core CPU.

Conclusions

We are now preparing the AdV Computing model.

We have experience from the past runs of the Virgo Detector, but a huge work is needed to re-organize the model to fulfill the new needs of the Advanced Detector.

We would benefit a lot from tight interactions with CCs colleagues, to work as much as possible to projects which will be supported in the future (GRID/CLOUD. Storage federations..)

BACKUP slides

Data access model. Section 4.1

- ``Locator Database”: a unified catalog of the data distributed among a variety of resources.
- ``Data bookkeeping Database”: a catalog of information on the data, like data quality.

A ``Locator Service” will then serve these information to users and software applications

Preliminary estimation of CPU needs at the Ccs.

- 1 yr of data, analyzed in one year (divided by 365 gives the power);
- 1 core = 10 HSE06
- CW estimations based on real benchmarks we have done, and not considering job failures
- Burst and CBC estimations based on extrapolations of analyses which ran @ LIGO clusters.

name	Cores	kHSE06	kHSE06.day
CNAF	13500 (real: 13800)	135	49270
LHC@CNAF	8800	88	32000
E@H Distributed CW search	17000	170	62000

Data Model (part II. Of the CM)
This is a summary table:

Data	Data flow [GB/day]	Buffer length in Cascina [year]	Buffer space in Cascina [TB]	Offline storage/year [TB]
IGWD data	6200	(0.008-3)	420	795
Calibration, DS	85.7	permanent	31	1
Detchar: DQ	11.7	1	4.3	4.3
Detchar: Noise	45.3	(0.1-1)	12.9	12
Science analysis	142	1	0.5	51.6
Total	6485	(0.008-permanent)	468.7	863.9

Table 2.18: Summary Table for all data. Offline storage space does not include multiple copies of the data.

Table 2.18

Data access model.

- Local Data Access: besides other the general considerations, we need to define the criteria to store the data on tape in the CCs, such that the recalling from tape will be fast enough for our procedures. The requirement in the CM is very simple: ``optimize the use of data long-term archiving system and the use of the disk caching system”
- Farther on the horizon: the Remote Data Access. We want to have Remote Data Access in place by ADE. One possible solution is to organize a ``Data Streaming Service”, which will serve channel files taken from the raw, RDS or h(t) frames. This service will make use of the ``Locator Service”. Technical solutions again need to be exploited.