Status and Perspective of INFN scientific Computing

A. Zoccoli
INFN and University of Bologna

EU-T0 Meeting Bologna February 8 2018
OUTLINE

- Organization and current status
- Future challenges and requests
- Guidelines and actions for future developments
1 National Computing Center (CNAF in Bologna) hosting the WLCG Tier-1, where is concentrated the major part of the personnel

9 Tier-2 centers on the Italian territory sitting mainly in University Physics Departments

C3S “Comitato di Coordinamento attività Calcolo Scientifico INFN”. Mandate: make proposals for research and developments for scientific computing, including the development of a shared computing infrastructure, to support researchers primarily from INFN, but also from different research institutions.
Tier 1:
• 200 kHS06 CPU power equivalent to 20000 cores
• 19 PB disk space
• 57 PB tape library

10 Tier 2:
• 240 kHS06 CPU power equivalent to 24000 cores
• 18 PB disk space

Network provided by GARR Consortium (10-100 Gbps), under upgrading

Major part of the computing centers are multidisciplinary (HEP, astroparticle, neutrino, etc.)
Resources distribution

Italian contribution 8%

In the world → WLCG

Half of the resources @ Tier1

In Italy

Very effective infrastructure for LHC data analysis!
Computing for theoretical physics

Currently exploiting CINECA, the national HPC computing center.

Use of INFN resources @ CINECA from 2012

Agreement INFN – CINECA valid for 3 years:
use of 6% MARCONI + 15 Mcorehours
GALILEO (~1.4 Pflops)
## Costs & manpower

### Budget

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost (M€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU, disk and tape</td>
<td>4</td>
</tr>
<tr>
<td>HPC@CINECA</td>
<td>0.5</td>
</tr>
<tr>
<td>Electric power</td>
<td>2</td>
</tr>
<tr>
<td>Network(@GARR)</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total/year</strong></td>
<td><strong>11.5</strong></td>
</tr>
</tbody>
</table>

### Personnel

<table>
<thead>
<tr>
<th>Infrastructure</th>
<th>Middleware and software</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 FTE</td>
<td>30 FTE</td>
</tr>
</tbody>
</table>
Future needs
By assuming:
• Trigger rate 5-10 kHz, factor10 wrt today
• Expected Pileup ~140-200 ⇒ complexity of events increased by a factor 10

→ Back on the envelope evaluation of resources needs = 100x wrt today !!
Expected mitigation effects:
- Technological evolution for CPU and storage: +20%/y performances → 6x in 10 years
- Computing Models Optimization (CPU: exploitation of parallel architectures) and of the trigger scheme (raw data vs AOD)

Current estimated gap:

**CPU:** 6x (CMS) e 4x (ATLAS)

**Disk space:** 4x (CMS) e 7x (ATLAS)
Future Astroparticle experiments (SKA, CTA, Euclid, …) will produce unprecedented amount of data.
Huge resources demand resulting in requests of HPC and Big Data management will come from many different research fields in the next years:

- HEP and astroparticle physics
- Human brain
- Biology
- Personalized medicine & genomics
- Weather predictions
- Climate change studies
  - Material studies
  - Smart manufacturing & Industry 4.0
- IoT
- SmartC

Impact on private sector
Future developments
Guidelines & Actions

• Evolution of the Middleware
• Evolution of the WLCG infrastructure
• Cooperation and synergy with the European context (ESOC + EuroHPC)
• Optimization of the INFN e-infrastructure for HEP and other disciplines
• Integration of the Italian HTC and HPC infrastructure
• Search a new location for the Tier-1
• Test beds for the use of the INFN e-infrastructure from private companies
INFN is deeply involved in the transition to CLOUD for example with the H2020 project: INDIGO-DataCloud: INtegrating Distributed data Infrastructures for Global ExplOitation (11.3 M€) 26 partners PI D. Salomoni (INFN)
The Horizon2020 eXtreme DataCloud – XDC project aims at developing scalable technologies for federating storage resources and managing data in highly distributed computing environments, as required by the most demanding, data intensive research experiments in Europe and worldwide.
Optimization of the INFN e-infrastructure for HEP and other disciplines

- INFN started a survey of its computing infrastructure in order to evaluate the status, the sustainability and the evolution.

- INFN signed a MOU started a collaboration with INAF to expand and exploit a common infrastructure. First action is hosting data and provide computing resources to the CTA experiment.

- INFN started a collaboration with ASI to host data of the Copernicus and CosmoSkyMed satellites and to provide computing resources to the relevant interested communities.

<table>
<thead>
<tr>
<th>30 e-Infrastructures</th>
<th>Big (T1+T2+CS+TS)</th>
<th>Small</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>95% Resources</td>
<td>5% Resources</td>
<td></td>
</tr>
</tbody>
</table>
The goal: test a flexible use of the infrastructure by using remote resources at Tier-1, by using in a transparent way CPU at Tier-2 Recas in Bari.

Performed tests also on small scale with private cloud providers: ARUBA e AZURE (microsoft)
INFN started a project with CINECA in order to integrate them to provide services to:

- Institutional basic and applied research
- Proof of concept and innovation for private organizations and industries
Search for a new location for the Tier-1 ECMWF center will be hosted in Bologna from 2019 in the Tecnopolo area.

The goal: provide a new location for the INFN Tier-1 to take into account future expansion.

Possibility to host in the area also:
- INFN Tier-1
- Cineca computing center

Already allocated 40 M€ for the Italian government to refurbish the area. Looking for extra budget for IFN & CINECA
Goal: build an European Data Infrastructure (with HPC and Big Data resources) based on the existing e-infrastructures for research to provide services also for Public and Private Partnership

Partners: France, Spain, Luxemburg, Italy

Test beds with involvements of private companies:

Space, Smart Manifacturing, Materials, Personalized medicine

Modular e-infrastructure to be realized at European, national or regional level, which may serve different fields
The presentation is based on the work performed by the C3S Committee.

Thanks to

all the people helped in the preparation of the slides: D. Lucchesi, T. Boccali, S. Campana, L. Cosmai, D. Salomoni, P. Vicini …….
Conclusions

- In the next years we have to cope with an unprecedented amount of data coming from many different fields (not only HEP)
- INFN e-infrastructure is in a transition phase
- We are actively exploring the most suitable solutions for our future
Backup slides
## External funds

<table>
<thead>
<tr>
<th>Anno</th>
<th>Tipo</th>
<th>Progetto</th>
<th>Budget Totale (M€)</th>
<th>Budget INFN (M€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>H2020</td>
<td>INDIGO-DataCloud</td>
<td>11.1</td>
<td>2.1</td>
</tr>
<tr>
<td>2015</td>
<td>FISR</td>
<td>High performance data network</td>
<td>12.5</td>
<td>12.5</td>
</tr>
<tr>
<td>2016</td>
<td>H2020</td>
<td>ExaNeSt</td>
<td>8</td>
<td>0.7</td>
</tr>
<tr>
<td>2017</td>
<td>H2020</td>
<td>EOSCpilot</td>
<td>10</td>
<td>0.3</td>
</tr>
<tr>
<td>2017</td>
<td>H2020</td>
<td>EOSC-HUB</td>
<td>30.0</td>
<td>1.8</td>
</tr>
<tr>
<td>2017</td>
<td>H2020</td>
<td>Extreme DataCloud</td>
<td>3.1</td>
<td>0.6</td>
</tr>
<tr>
<td>2017</td>
<td>H2020</td>
<td>Deep Hybrid DataCloud</td>
<td>3.0</td>
<td>0.4</td>
</tr>
<tr>
<td>2017</td>
<td>H2020</td>
<td>EuroEXA</td>
<td>20</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Totale</strong></td>
<td></td>
<td></td>
<td><strong>97.7</strong></td>
<td><strong>19.1</strong></td>
</tr>
<tr>
<td>2018</td>
<td>MIUR</td>
<td>PON</td>
<td>15</td>
<td>?</td>
</tr>
</tbody>
</table>
Toward a Cloud based system

EOC-pilot
The European Open Science Cloud for Research Pilot Project

Definition of the EOSC governance
Realization of demonstrators
PI: UK INFN: D. Salomoni

EOSC-hub

Data and services for EOSC
PI: EGI INFN: L. Gaido

XDC
extreme DataCloud

Development of technologies for storage resources federation and for data management
PI: INFN D. Cesini

DEEP HybridDataCloud
Designing and Enabling E-Infrastructures for intensive Processing in a Hybrid DataCloud

Services and support for intensive computing for different disciplines
PI: CSIC (Spain) INFN: G. Donvito
INFN projects toward HPC exascale

**European Exascale System Interconnection Network & Storage**
PI: Foundation for Research & Technology, GR
INFN: P. Vicini

**Co-designed Innovation and System for Resilient Exascale Computing in Europe: From Applications to Silicon**
PI: Institute of communication and computer systems, GR
INFN: P. Vicini

**Acquisition and exploitation of many core "next gen" for INFN HPC e HTC infrastructures**
P. Vicini et al.
Computational theoretical physics at INFN: status and perspectives (2018-2020)


* Conveners

(Dated: April 26, 2017)

We present the status of computational theoretical physics at INFN, the results obtained by its research groups active in this field and their research programs for the next three years. Computational theoretical physics, besides its own importance, is a powerful tool in understanding present and future experiments. A continued support of INFN to computational theoretical physics is crucial to remain competitive in this sector. We assess the high performance computing resources needed to undertake the research programs outlined for the next three years.

<table>
<thead>
<tr>
<th></th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGT: hadron physics</td>
<td></td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>LGT: QGP and BSM</td>
<td>207</td>
<td>432</td>
<td>648</td>
</tr>
<tr>
<td>LGT: flavor physics</td>
<td>117</td>
<td>234</td>
<td>387</td>
</tr>
<tr>
<td>Colliders Phenomenology</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>General Relativity</td>
<td>142</td>
<td>182</td>
<td>227</td>
</tr>
<tr>
<td>Cosmology and Astroparticle Physics</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Nuclear Theory</td>
<td>18</td>
<td>27</td>
<td>36</td>
</tr>
<tr>
<td>Fluid Dynamics</td>
<td>50</td>
<td>80</td>
<td>110</td>
</tr>
<tr>
<td>Quantitative Biology</td>
<td>9</td>
<td>18</td>
<td>27</td>
</tr>
<tr>
<td>Disordered systems</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Condensed matter</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Grand Total (M core-h)</td>
<td>607</td>
<td>1097</td>
<td>1638</td>
</tr>
<tr>
<td>Grand Total (Eq. Pflops)</td>
<td>4.6</td>
<td>8.4</td>
<td>12.5</td>
</tr>
</tbody>
</table>
Possible evolution of the WLCG infrastructure

First proposal from CERN

Fast network (∼Tb/s)

Computing Centers, with CPU and disk space

Data Centers (DC): host the major part of tape and disk space with small CPU (core of the infrastructure)

Commercial Clouds and HPC centers when available (elastic model)
Cooperation and synergies with the European context

- In 2015 launch of the European Open Science Cloud (EOSC) initiative to provide an e-infrastructure for all European research communities.
- In 2016 signature of the EuroHPC agreement to provide a 10 years European strategy toward HPC exascale machines.