PIC
port d'informació
científica

ASPERA

# Astro-particle Physics
# and
# Computing

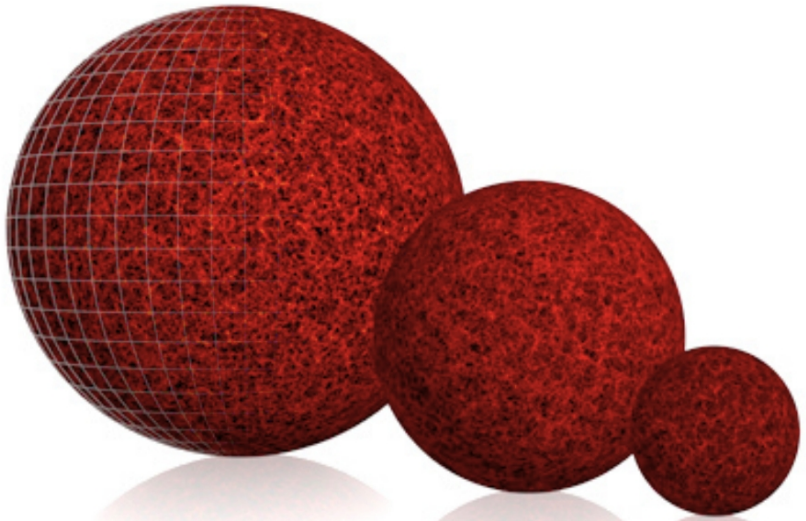
Manuel Delfino
Director, Port d'Informació Científica (PIC)

*Presented at the ASPERA Roadmap Event,*
*Collège de France, Paris, November 21, 2011*

# Conclusions

- Computing and Data Processing are essential parts of Astro-particle Physics.
- Experiments will take data for decades and their computing must have continuing support and planning with technological renewal.
- Requirements are already large and they are growing.
- Costs are not negligible and must be planned for. “Computing Models” are an essential tool.
- A lot can be learnt from High Energy Physics, particularly from LHC. Not everything is the same, however.
- Data centers are in quantum transition: Portals + “Industrial” Clouds (resource efficient mega-centers)

# Workshops: Third one at Albert Einstein Institute, Hannover, May 3-4, 2012



## Computing and Astroparticle Physics

2nd ASPERA Workshop – Barcelona, Spain – May 30-31, 2011

**Models for Computing and Data Processing Pipelines**

**Technologies for Data Processing and Computing**

Organized by:

PIC port d'informació científica

Call for Participation, including proposals for Short Presentations, open until May 15, 2011. Please apply at:

http://indico.cern.ch//conferenceDisplay.py?confId=134280

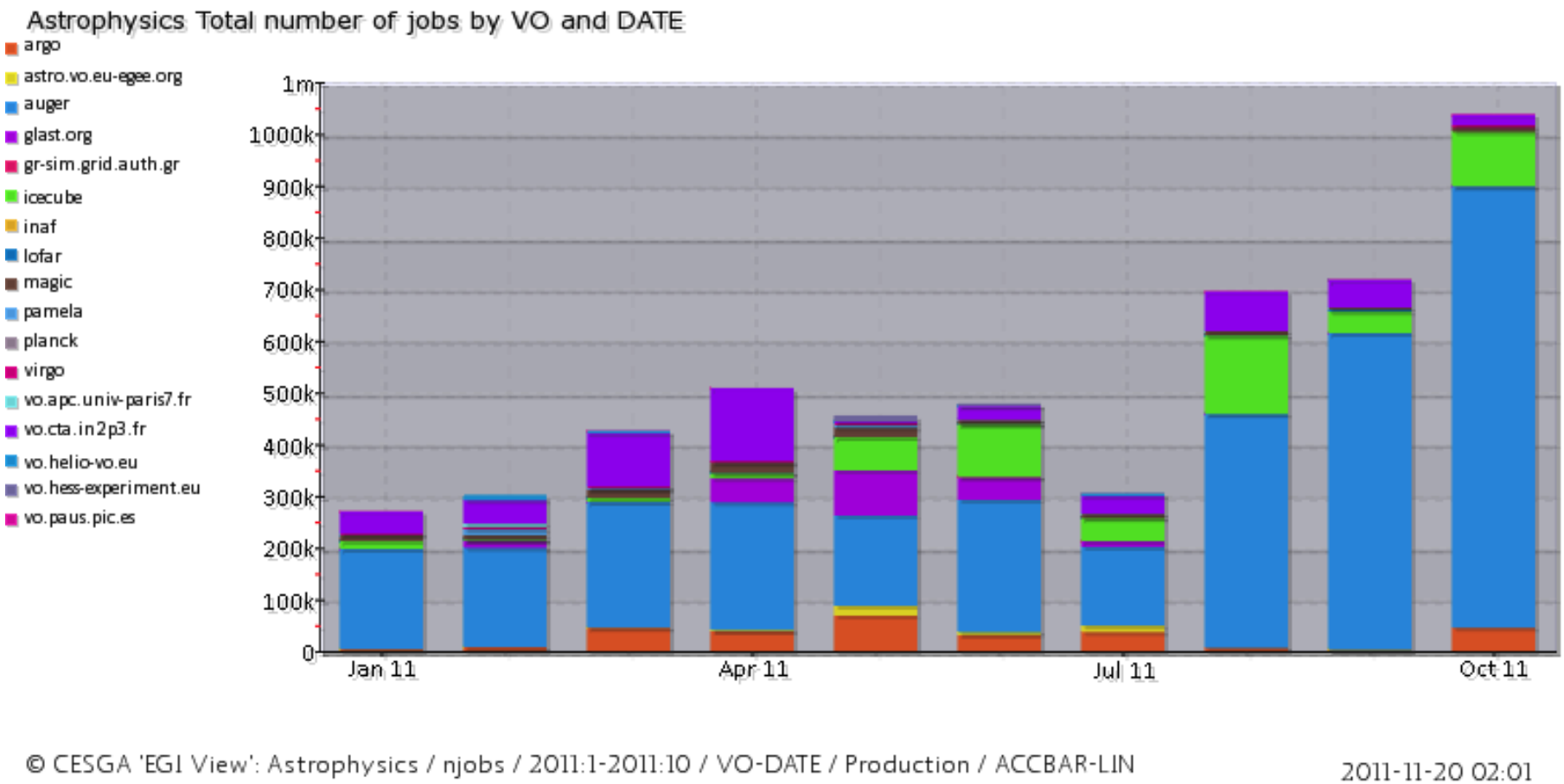
ASPERA
Astroparticle Physics for Europe
www.aspera-eu.org

Illustration credit: MICE project http://www.ice.cat/mice
IEEC/ICE, UAB Campus, Bellaterra, Spain

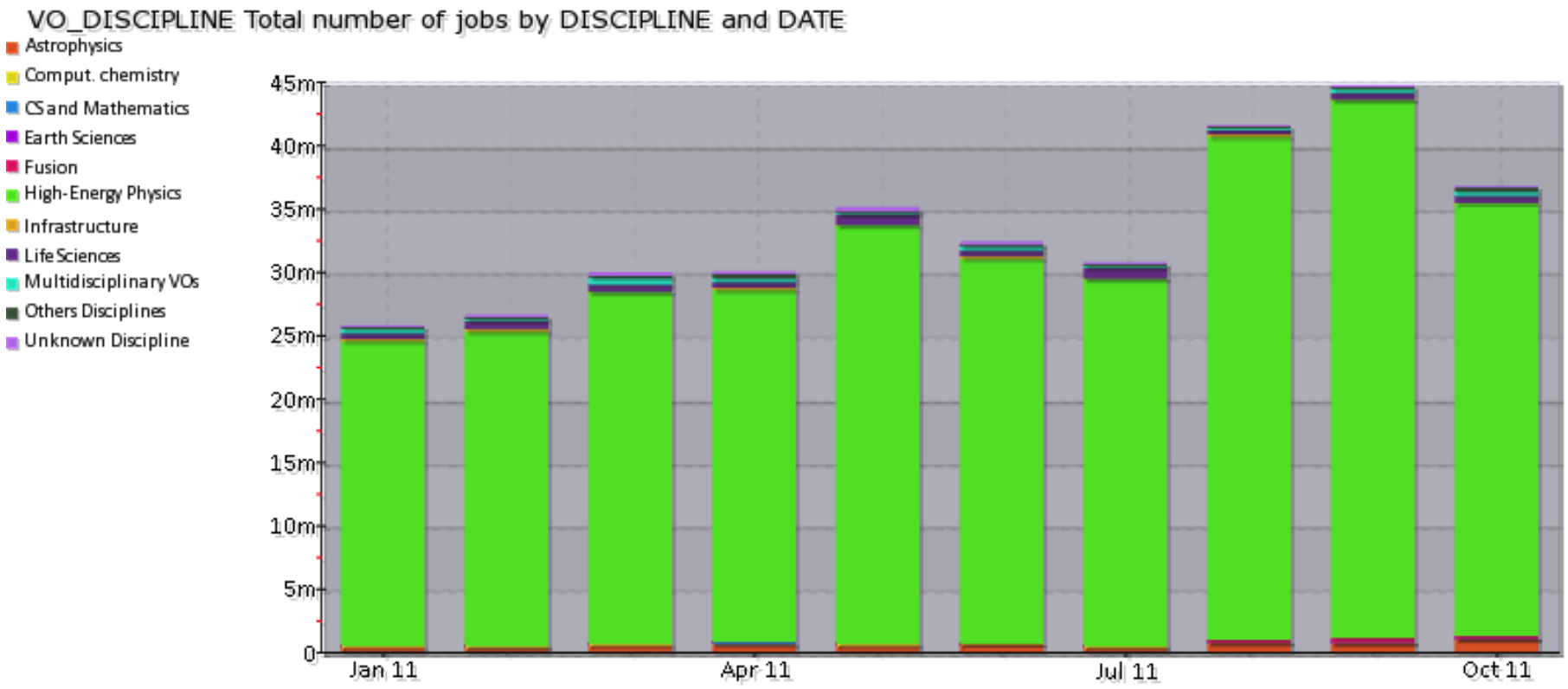
# Distributed Computing and Data Processing (1)

- Distribution important from funding perspective
- Worldwide LHC Computing Grid has taught us:
  - Data can be shipped between centers quickly. Research networks are a very reliable component.
  - Data Processing and HEP simulation tasks can be efficiently shared between many sites.
  - Computing and Data Processing capacity can be efficiently shared between many sites.
  - Wider distribution requires a more homogeneous system environment or a simpler one (or both) to keep things manageable.
  - A smallish number of sites contributes most of the resources. Cost/capacity much higher at small sites.

# Use of EGI/OSF/WLCG Grid (1)



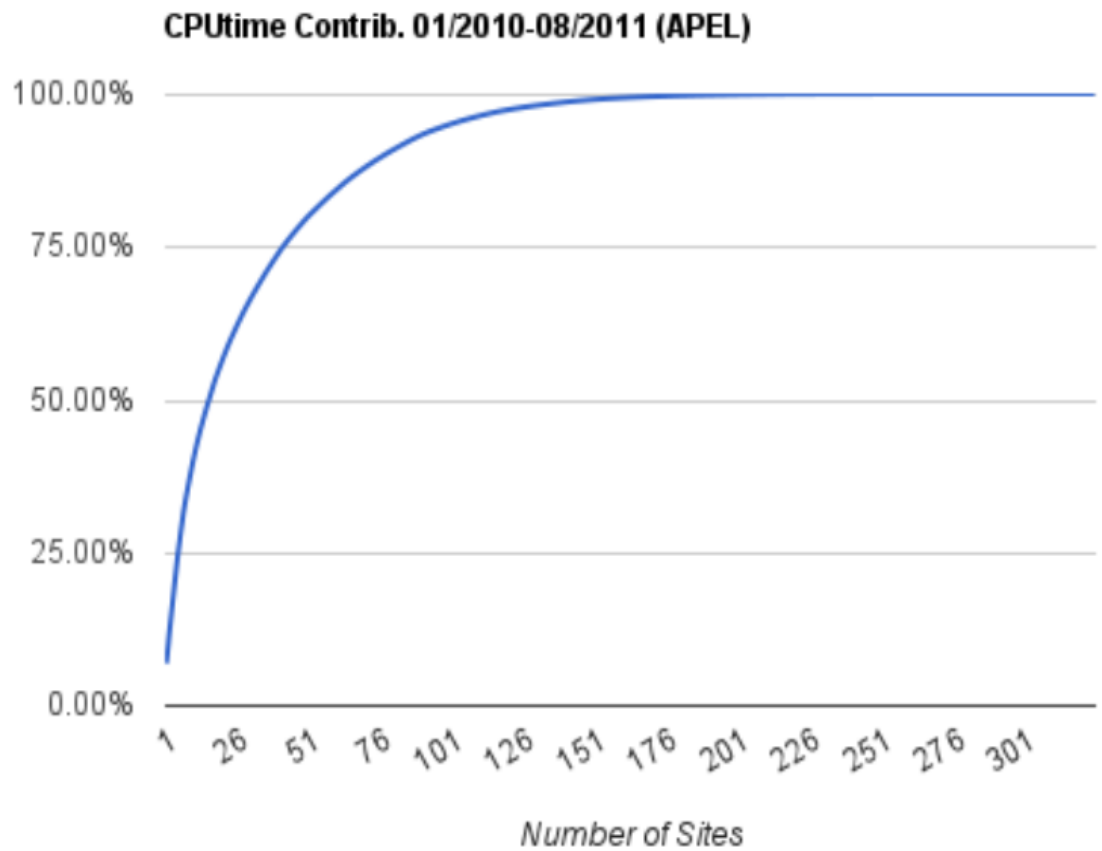
# Use of EGI/OSF/WLCG Grid (2)




© CESGA 'EGI View': VDISCIPLINE / njobs / 2011:1-2011:10 / DISCIPLINE-DATE / Production / ACCBAR-LIN

2011-11-20 02:01

# Cumulative contribution by sites
## (1.5 years of Grid accounting data)




- Large part of the workload goes to small fraction of large sites (50% done in 16/325 sites)
- Large part of the operational effort may be goes to small sites

*(reported by G. Merino, PIC/CIEMAT, at the EGI Technical Forum, Lyon, Sept. 21, 2011)*

# Distributed Computing and Data Processing (2)

- Grid, Cloud and Portals
  - The Grid was layered over existing Computer Clusters. Therefore it cannot go deep into operating systems and security infrastructure.
  - Requirement to include a very large number of sites with “skin-deep” system results in large human costs.
  - Cloud is the “industrialization” of grid-like computing. Emphasis is on small number of very large sites in order to achieve economy, efficiency and ecology.
  - Portals are much less of a buzzword, but are probably more important. They will be the Users Window into the data processing cloud.
  - Portals are also the key to “public” data access.

# Examples of Portals

To download a MAGIC fits file, right-click on the corresponding "FITS" link and choose "Save As". To open the file, you will need a fits viewer, like fv.

Please consider all content of these files to be PRELIMINARY!!

**Search the database by the source name, article title, reference or filter by the year.**

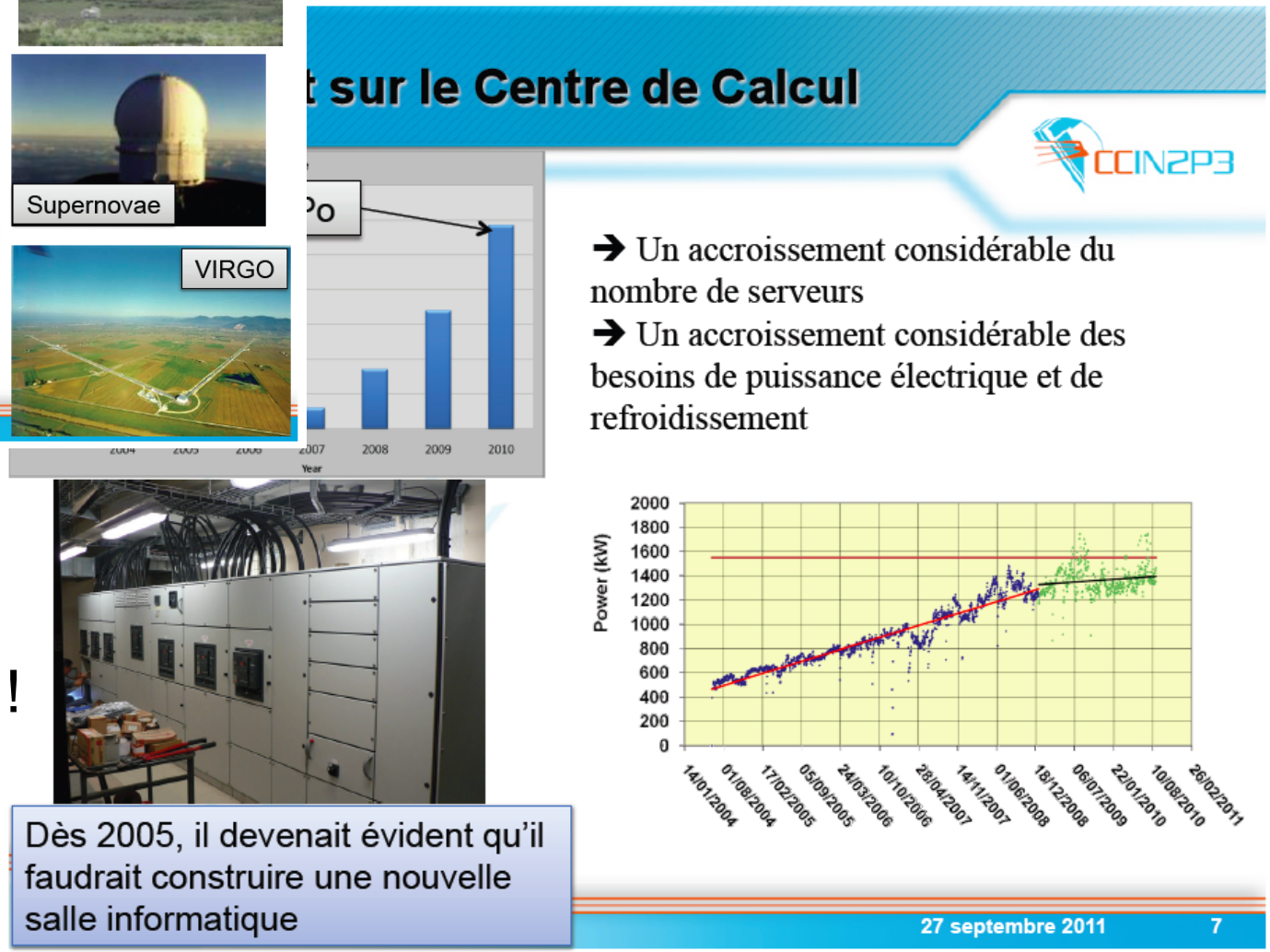
Reset

| Source | Article | Year | Reference | Download |
|---|---|---|---|---|
| Crab Pulsar | Observations of the Crab Pulsar between 25 and 100 GeV with the MAGIC I Telescope | 2011 | (J.Aleksić et al., ApJ 742, 43) | FITS |
| Scorpius X-1 | A search for very-high energy gamma-ray emission from Scorpius X-1 with the MAGIC telescopes | 2011 | (J.Aleksić et al., ApJ Lett. 735, L5) | FITS |
| Segue 1 | Searches for Dark Matter annihilation signatures in the Segue 1 satellite galaxy with the MAGIC-I telescope | 2011 | (J.Aleksić et al., JCAP 06, 035) | FITS |
| PKS 1222+21 | MAGIC discovery of VHE Emission from the FSRQ PKS 1222+21 | 2011 | (J.Aleksić et al., ApJ Lett. 730, L8) | FITS |
| 3C 66A | Observations of the Blazar 3C 66A with the Magic Telescopes in Stereoscopic Mode | 2011 | (J.Aleksić et al., ApJ 726, 58) | FITS |
| IC 310 | Detection of very high energy gamma-ray emission from the Perseus cluster head-tail galaxy IC 310 by the MAGIC Telescopes | 2010 | (J.Aleksić et al., ApJ Lett. 723, L207) | FITS |
| 1FGL J1954.3+2836, 1FGL J1958.6+2845 | M | | ) | FITS |
| Cygnus X-3 | M | | | FITS |

# Distributed Computing and Data Processing (3)

- More on Clouds, Virtualization, IaaS and Energy Efficiency
  - Emerging area, partly covered in workshops, will be included in Hannover workshop.
  - Virtualization in principle allows decoupling between hardware infrastructure and application frameworks: But beware of data-intensive and HPC applications.
  - Virtualization and Clouds go together. Clouds are all about industrial virtual computing and Infrastructure as Service.
  - Industrialization also driven by Energy Efficiency needs:
    - Big HEP Data Center was <1 MW until recently.
    - CERN upgrading to 3.5 MW.
    - 2011 Largest Supercomputer Japan K 13 MW
    - Amazon 2008 Oregon Data Center 10 MW
    - Largest UK Data Center Infrastructure 80 MW
  - Must invest in collaboration with Cloud and IaaS Industry

# CC-IN2P3 inauguration of new computer room (D. Boutigny)




Note:
Growth driver is
Astro-particle Physics !

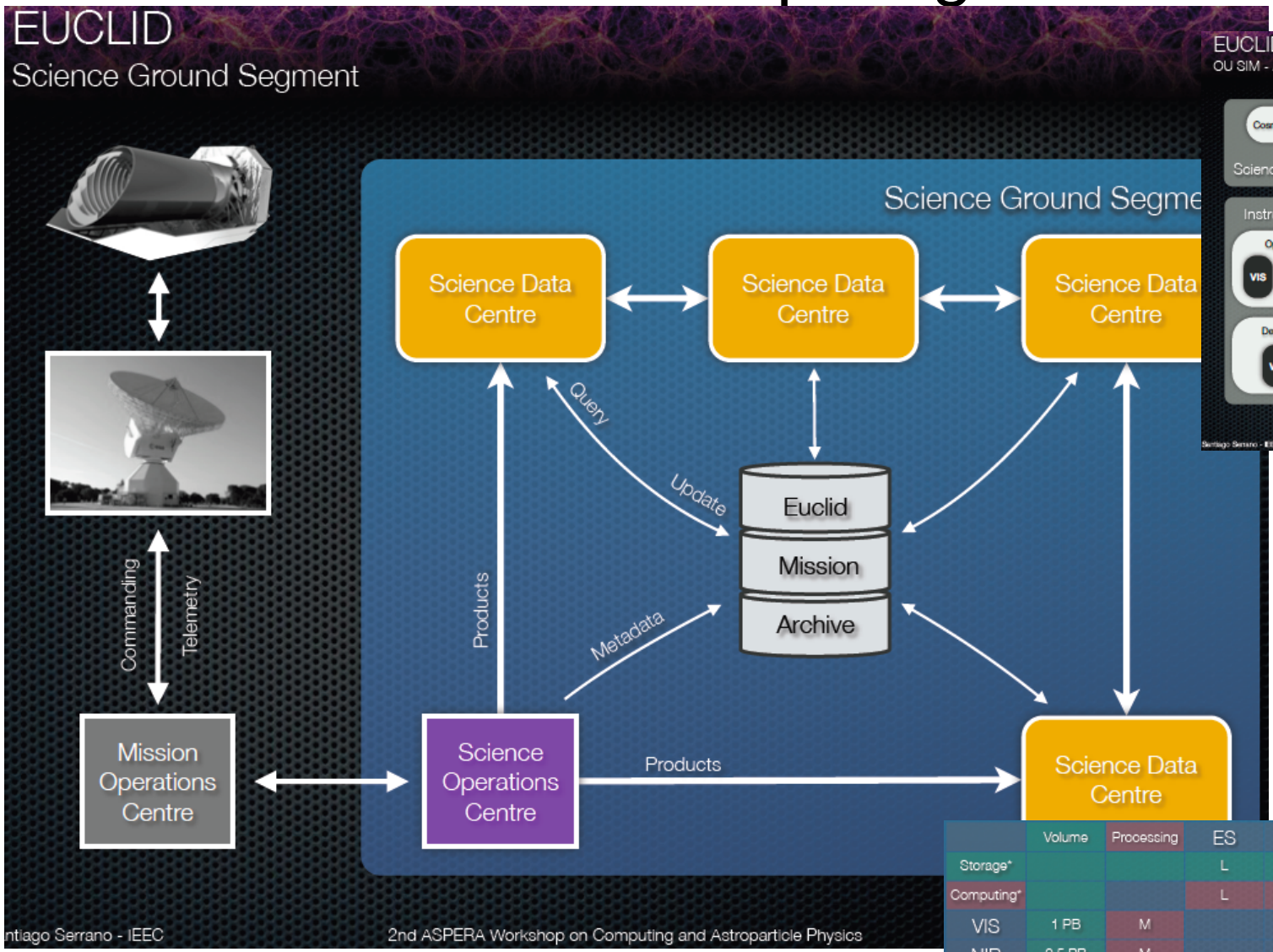
# High-Performance vs High Throughput (HPC vs HTC) Computing

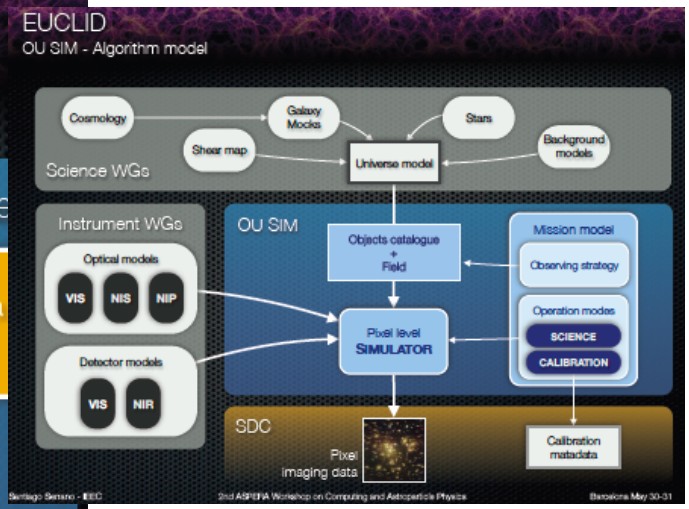
- HTC use by ApP is sufficiently similar to HEP that centers like LHC Tier1s can extend service reliably and with economies of scale.
- ApP uses additional technologies:
  - Supercomputers (e.g. for Cosmology simulations)
  - GPU-accelerated clusters (e.g. Gravitational waves)
  - Databases loaded with large scientific datasets
- Need to invest in:
  - Improved data access and transfer HPC ↔ HTC
  - GPU and many-core programming techniques
  - Efficient traversal and replication of databases

# Computing Models as tools for Financial and Technical Planning (1)

- Key elements of a Computing Model:
  - DATA: Raw and Derived data types, sizes and volumes
  - WORKFLOWS: Reconstruction, Analysis, Simulation
- USER ANALYSIS: One of the most difficult parts to model. Need to make some assumptions.
- Need a procedure to manage the requirements changes as much as the Computing Model itself.
  - Solid accounting infrastructure that enables regular reporting.
  - Formal procedure for periodically feedback requirement changes into the capacity planning/funding.
- One Computing Model per project ? Or one per “area” of ApP ?
  - Opportunities for re-use and economy, particularly software.

# Example: Work on EUCLID Computing Model

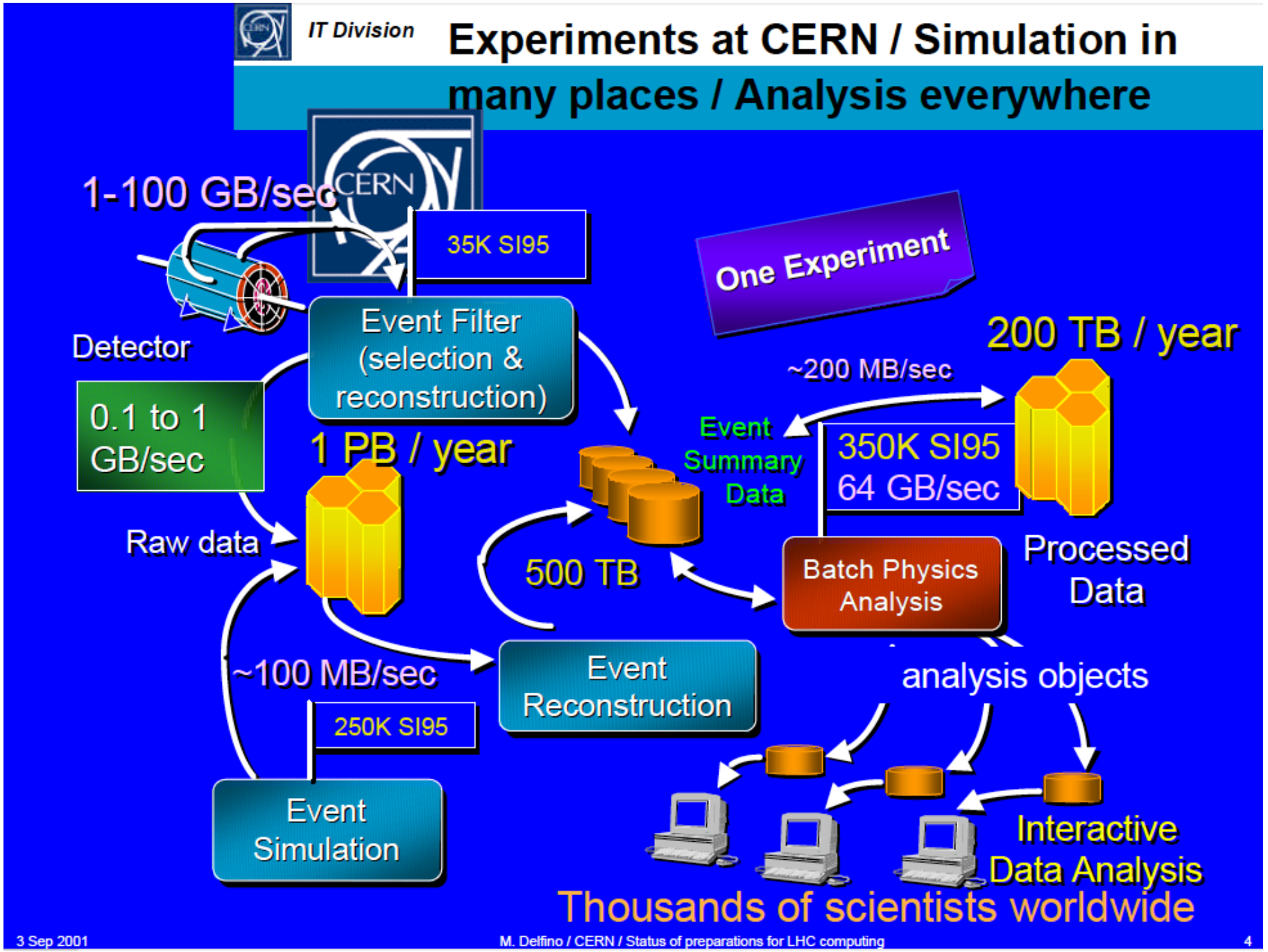



| | Volume | Processing | ES | ITA | UK | D | NL | FR | CH |
|---|---|---|---|---|---|---|---|---|---|
| Storage* | | | L | L | XL | XL | L | XL | L |
| Computing* | | | L | L | M | M | L | L | L |
| VIS | 1 PB | M | | | | | | | |
| NIR | 0.6 PB | M | | | | | | | |
| SIR | 60 TB | S | | | | | | | |
| EXT | 10 PB | L | | | | | | | |
| SIM | TBD | TBD | | | | | | | |
| MER | 13 PB | L | | | | | | | |
| SPE | 0.4 PB | M | | | | | | | |

OU to SDC MATRIX (TBC)

# Presentation at CHEP 2001 about LHC Computing



# Computing Models as tools for Financial and Technical Planning (2)

- Some personally chosen quotes ...
- for projects:
  - Plans are worthless. Planning is essential (Dwight D. Eisenhower)
  - You can't build a reputation on what you are going to do (Henry Ford)
- for Funding Agencies:
  - A thousand policemen directing traffic cannot tell you why you come or where you go (T.S. Elliott)
  - Those who plan do better than those who do not plan even though they rarely stick to their plan. (Winston Churchill)
- for all of us:
  - It is not the strongest of the species that survive, not the most intelligent, but the one most responsive to change. (Charles Darwin)
  - If you can dream it, you can do it. (Walt Disney)

# Conclusions

- Computing and Data Processing are essential parts of Astro-particle Physics.
- Experiments will take data for decades and their computing must have continuing support and planning with technological renewal.
- Requirements are already large and they are growing.
- Costs are not negligible and must be planned for. “Computing Models” are an essential tool.
- A lot can be learnt from High Energy Physics, particularly from LHC. Not everything is the same, however.
- Data centers are in quantum transition: Portals + “Industrial” Clouds (resource efficient mega-centers)