Secretaría de Estado de Investigación, Desarrollo e Innovación

MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD

Notes on Computing for the LHC

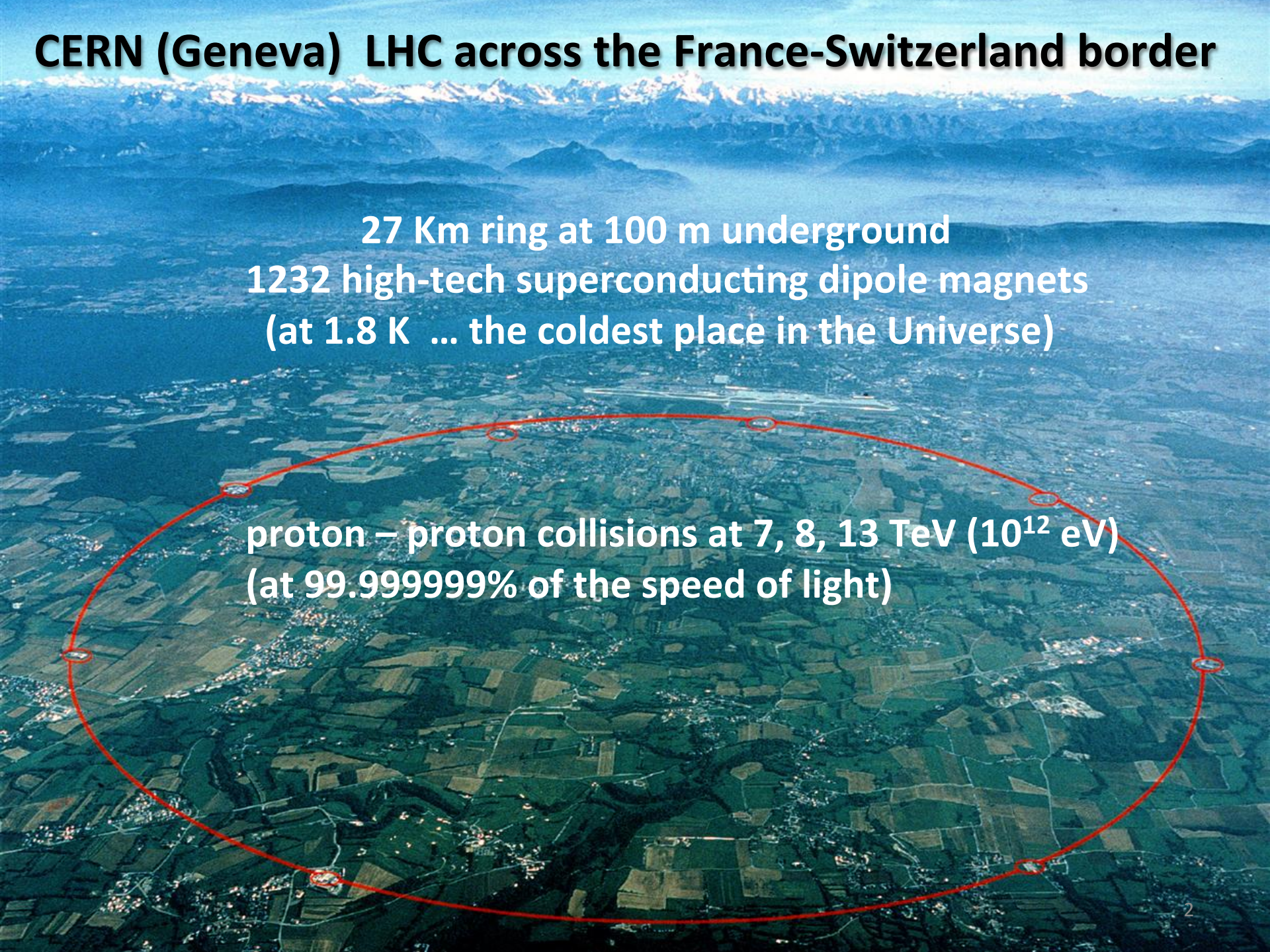
M. Martínez

(Manager Spanish HEP Program)

Portuguese – Spanish Workshop on Research and Education Networks and E-Science

Salamanca, Spain, 7th Mary 2018

CERN (Geneva) LHC across the France-Switzerland border

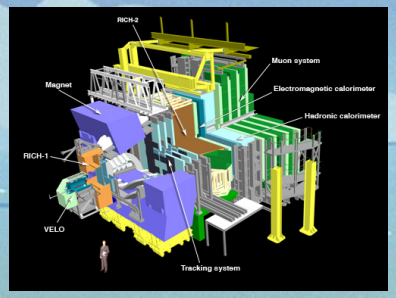


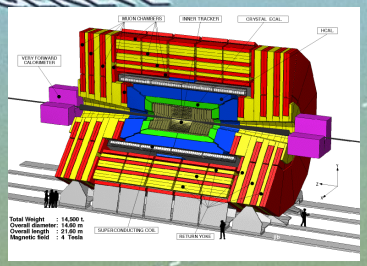
27 Km ring at 100 m underground
1232 high-tech superconducting dipole magnets
(at 1.8 K … the coldest place in the Universe)

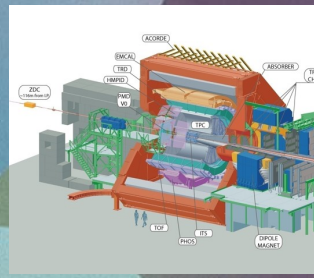
proton – proton collisions at 7, 8, 13 TeV (10^{12} eV)
(at 99.999999% of the speed of light)

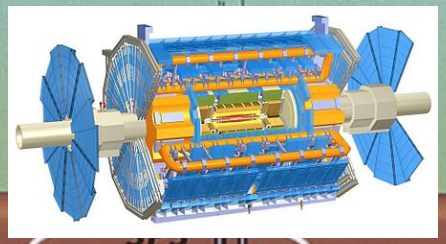
Overall view of the LHC experiments.

Operations for 20 years
About 10^{16} collisions (every 25 ns)
About 40 PB / year of data









LHC - B
Point 8

CERN

ATLAS
Point 1

ALICE
Point 2

CMS
Point 5

TI 8

TI 2

LHC - B

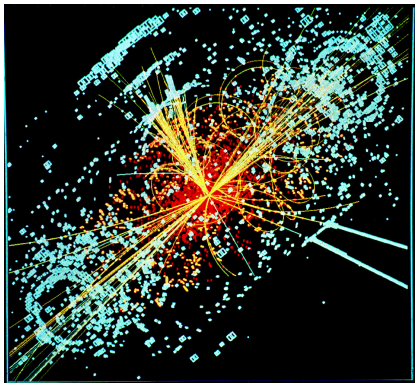
ATLAS

ALICE

CMS

LEP/LHC

Outline



- Introduction to WLCG
- WLCG @ Spain
- Cloud Computing

- Towards HL-LHC
- Role of HPC
- Final notes

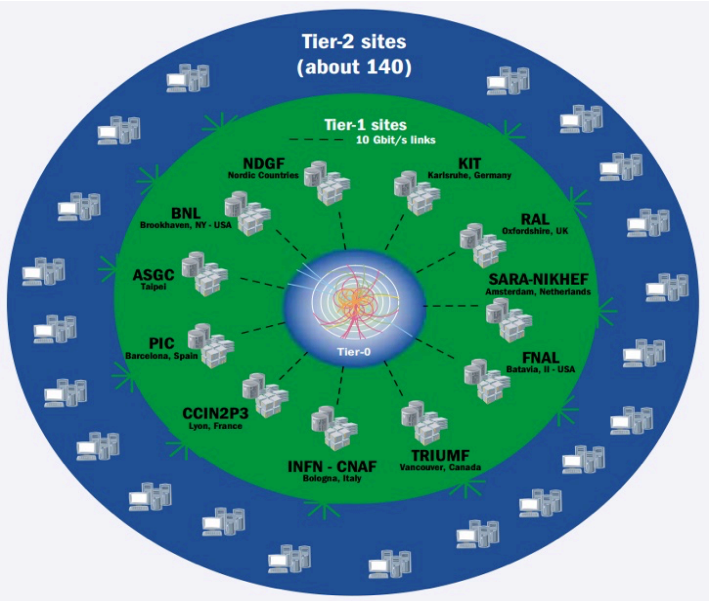
Thanks to all the WLCG-Spain community for their help in preparing this talk

The Worldwide LHC Computing Grid

Worldwide distributed computing infrastructure for data-intensive processing of LHC data

Tier structure with:

Tier0 @ CERN
Tier1s, 13 worldwide (1 in Spain @ PIC)
More than 140 Tier2s (6 in Spain)
+ Many small analysis clusters (Tier3s)



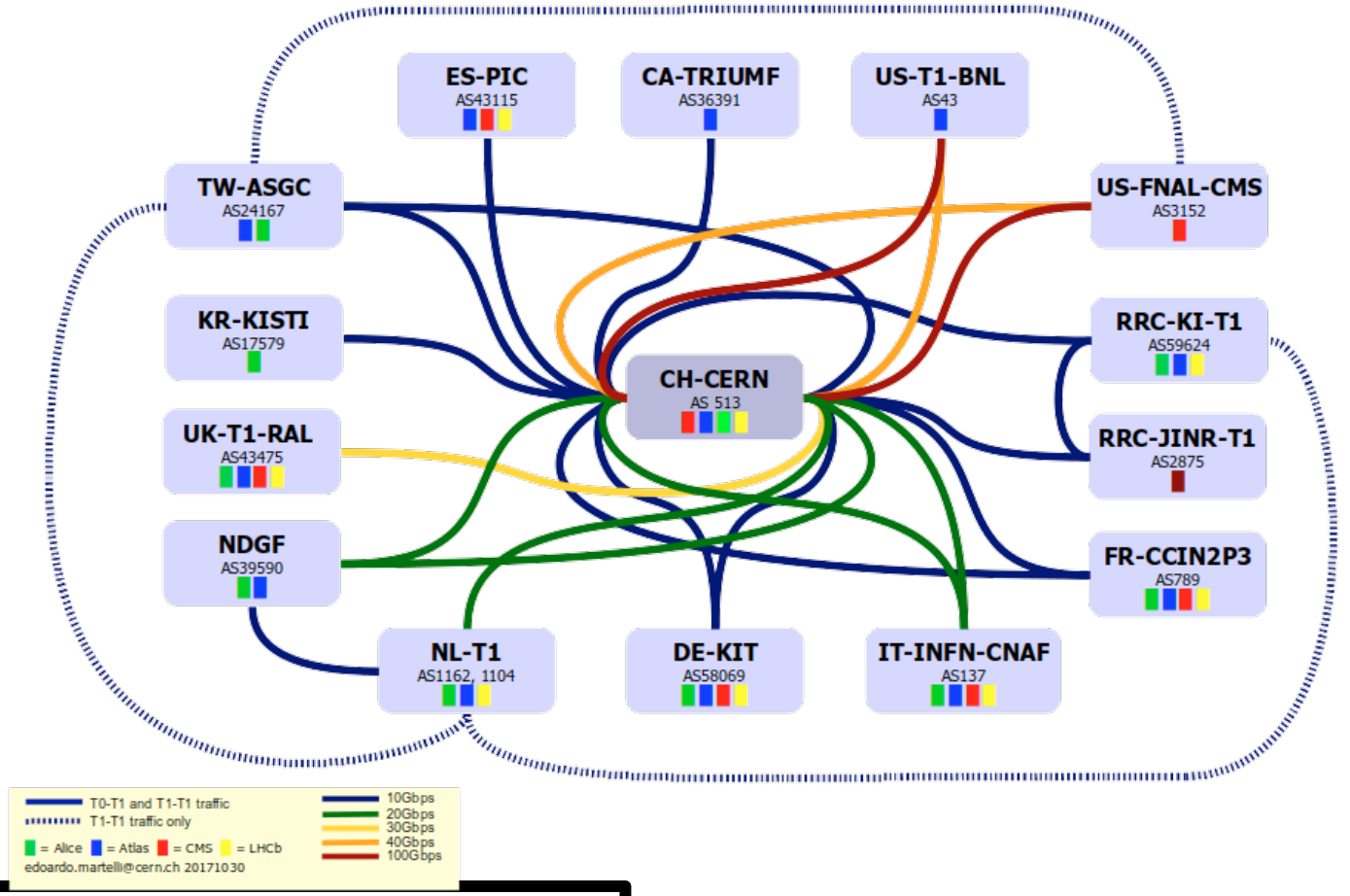
WLCG has been a tremendous success and is the result of ~20 years of development . These days typically

About 0.7M cores + 1EB storage (disk + tapes)
More than 30 GBytes/s (WAN traffic between sites)

The original Tier1-Tier2 architecture getting washed out

LHC Optical Private Network

Optical fibres working at 10-100 Gbps connect CERN to each of the 13 Tier1 centres around the world, the high-bandwidth **LHC Optical Private Network** (LHCOPN)

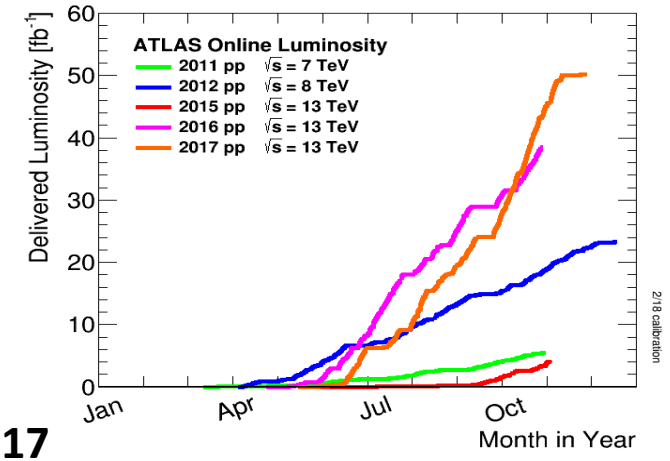


- Optical private network with **10/100 Gbps WAN links** between centers
- ~Tbps LAN bandwidth between compute and storage nodes at centers

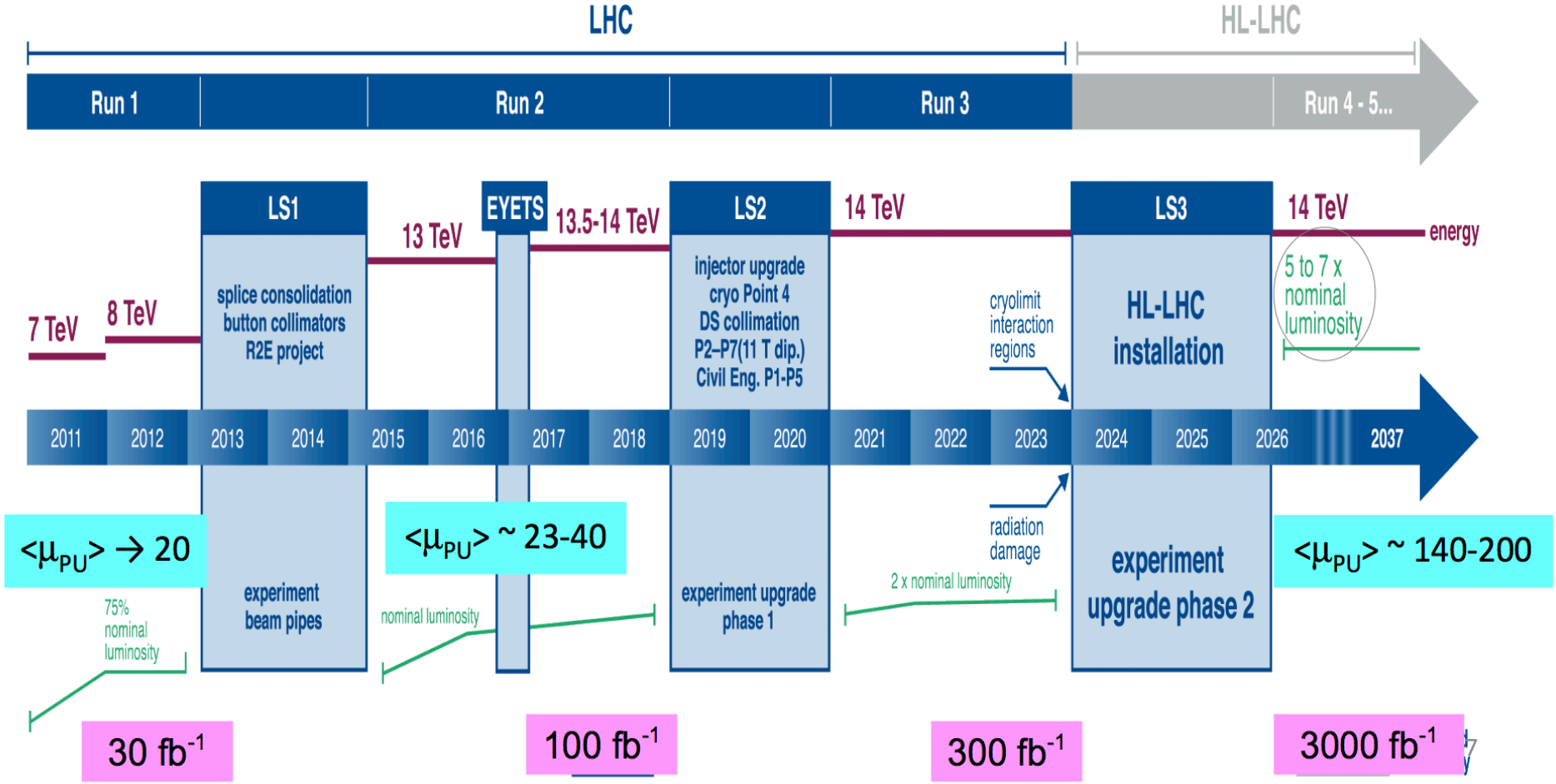
Tier2 centers in Spain connected to Tier1 via LHCONE provided by REDIris (10-20 Gbps)

LHC schedule

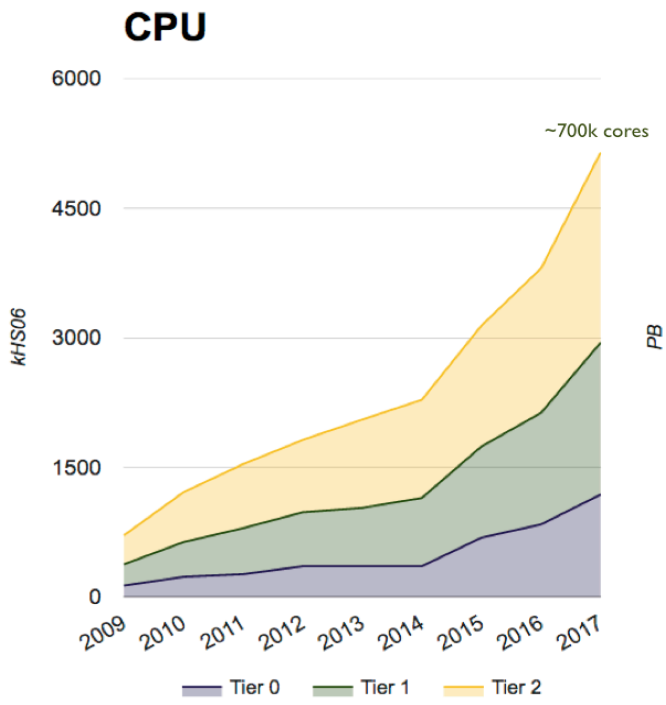
The improved performance of the LHC during run I+II and the accumulated data sample translated into larger computing needs over the years

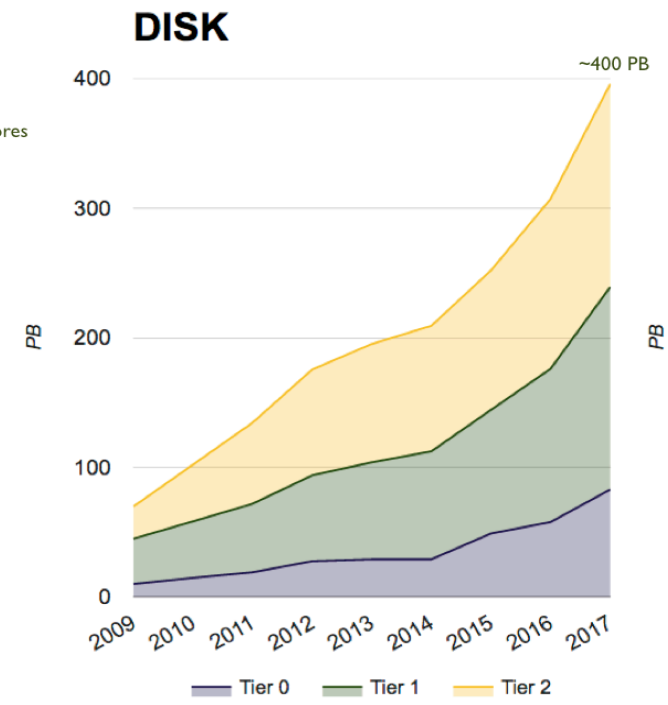


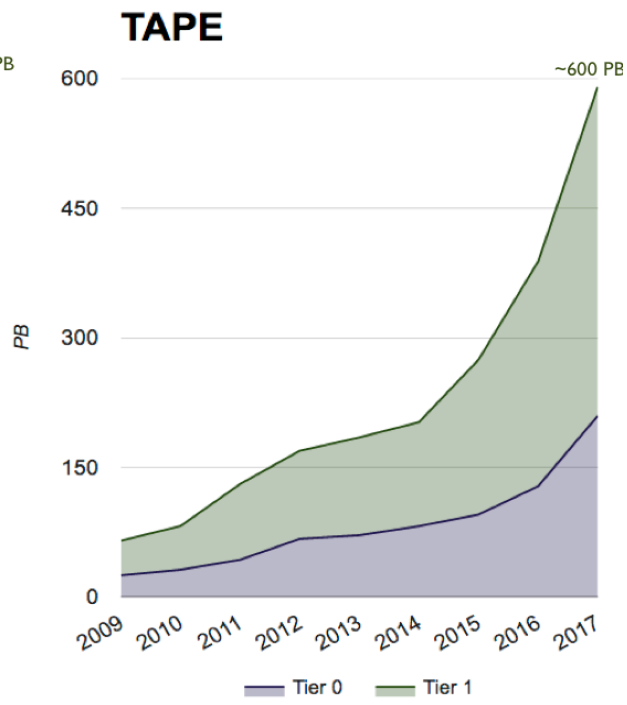
2018 resources needed are 20%-25% more than those in 2017



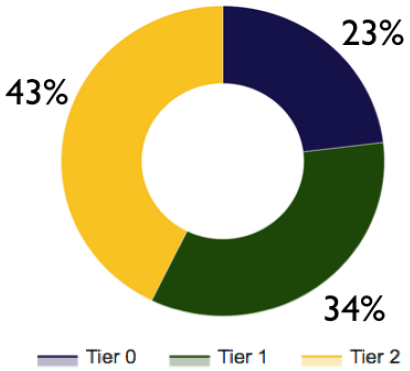
Worldwide LHC Computing Grid (WLCG)

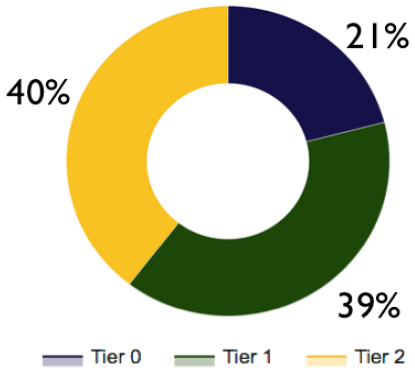


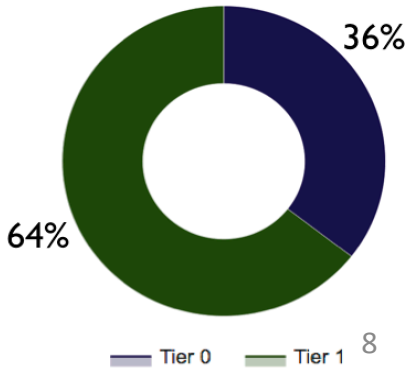




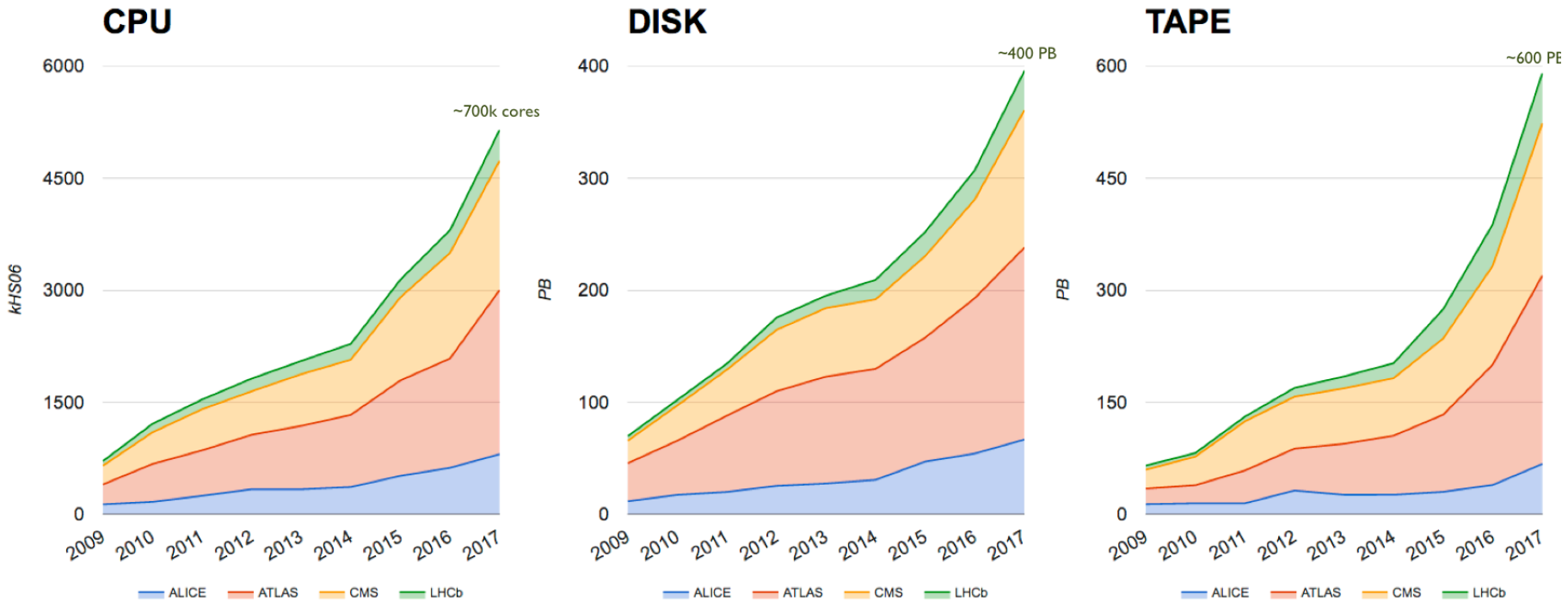
TODAY







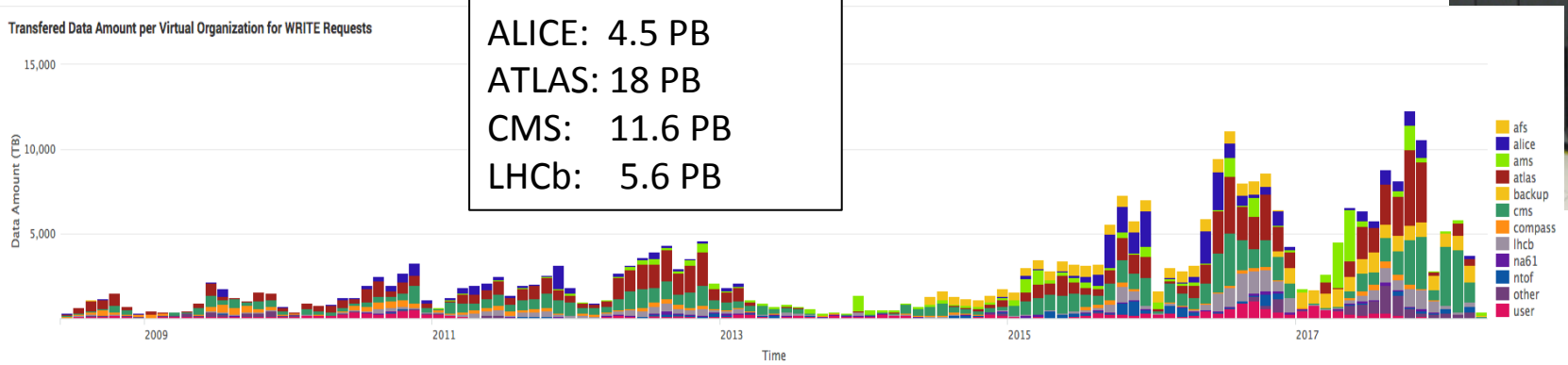
Worldwide LHC Computing Grid (WLCG)

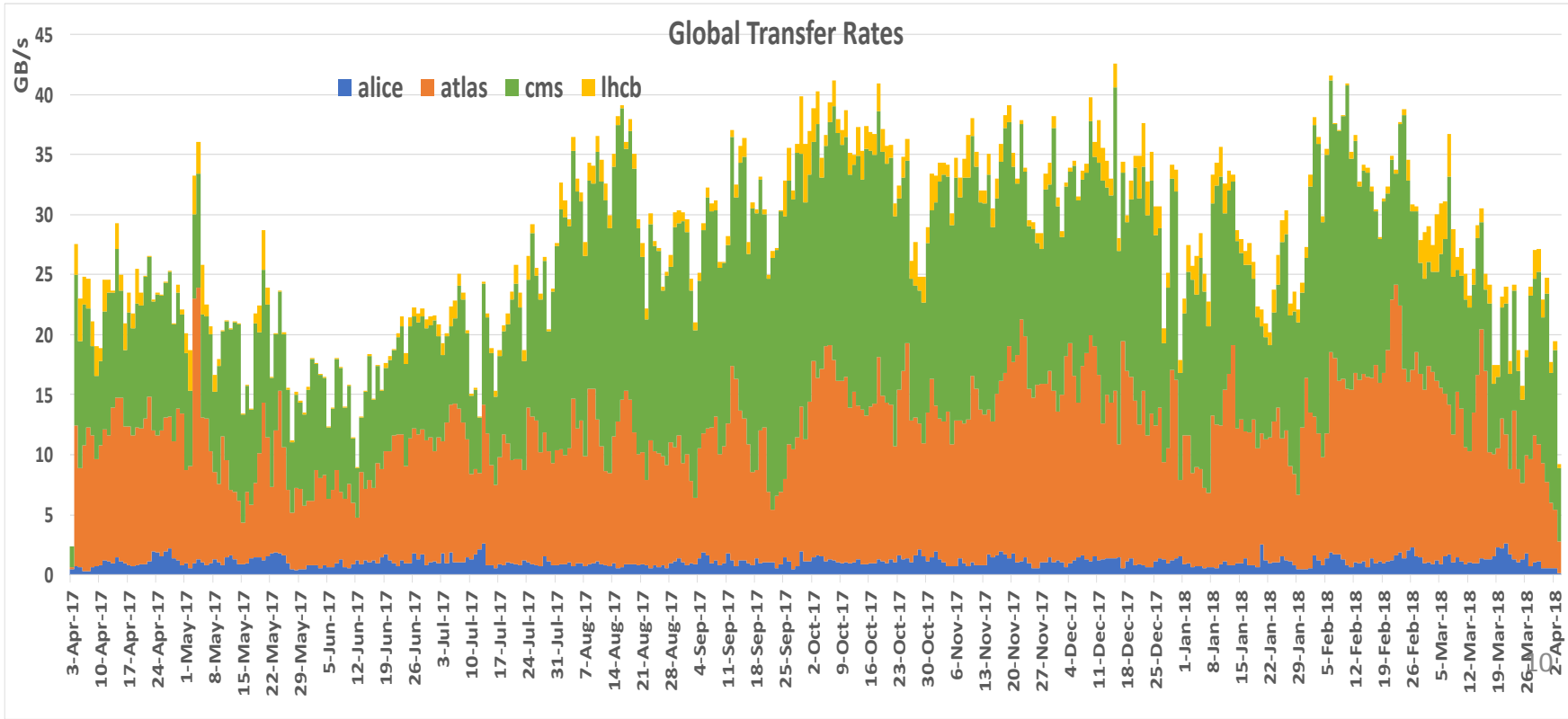


- Large variety of workflows
  - CPU-intensive: simulation (~50%)
  - I/O-intensive: data reconstruction (~25%) and analysis (~25%)
- High availability (24x7 in big centers) and reliability (>95%)
  - Serving a worldwide community of ~10k users, continuous ingest of data
  - Fast response to service interruptions
  - Large asociated operation costs…

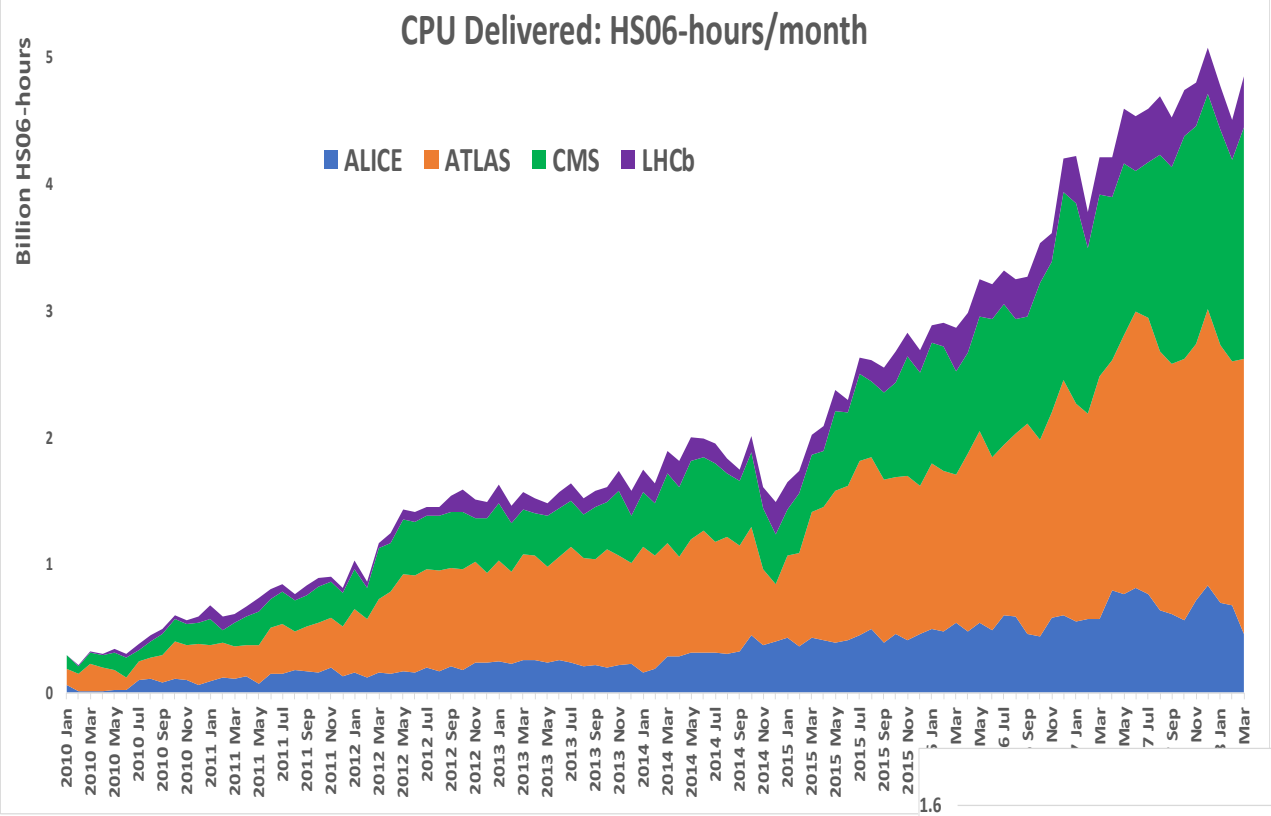
Data

2017: 40 PB
ALICE: 4.5 PB
ATLAS: 18 PB
CMS: 11.6 PB
LHCb: 5.6 PB



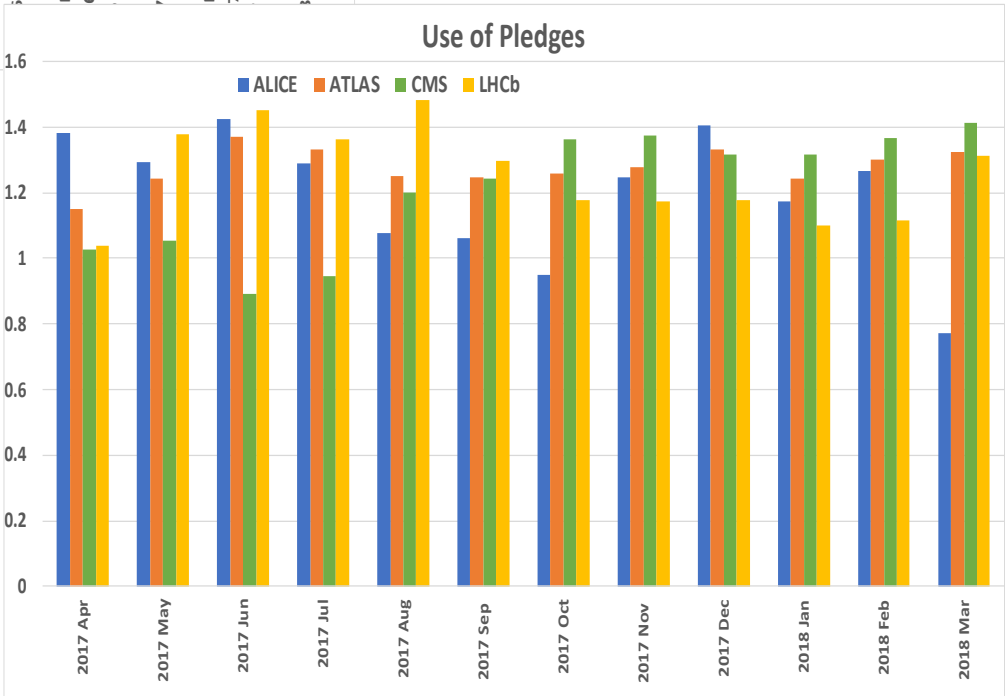


CPU Delivered



New peak: ~210 M HS06-days/month
~ 685 k cores continuous

1 core ~ 10 HS06
→ 500M CPU hours/month

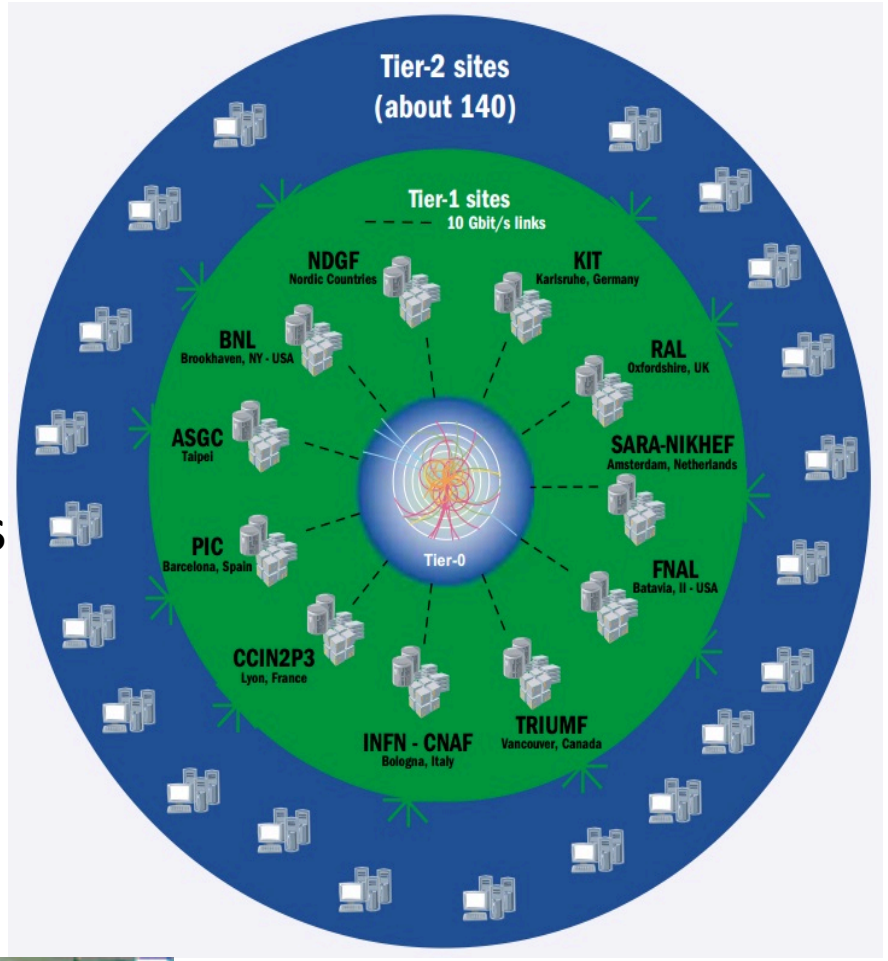


WLCG@Spain

Tier-0 (CERN): (15%)
Tier-1 (13 centers): (40%)
Tier-2 (~140 centers): (45%)

About 7500 cores (~93 kHS06)
8.3 PB of Disk
25 PB Magnetic Tape

Provides 5.0% of Tier1 Resources for ATLAS, CMS
Provides 6.5% of Tier1 Resources for LHCb

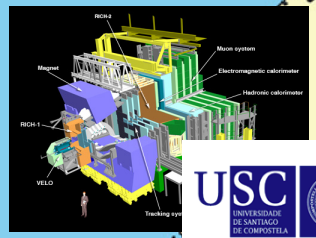


One of the Tier 1 LHC Centers at the UAB Campus, Barcelona (T2 + T3 parts integrated)

Institut de Física d'Altes Energies

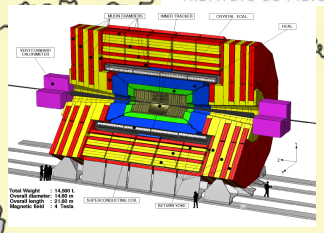
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

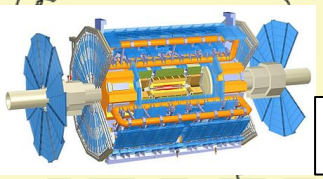
ATLAS/CMS/LHCB Tiers Map



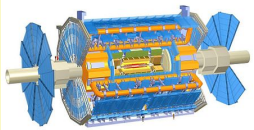
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

IFCA Instituto de Física de Cantabria





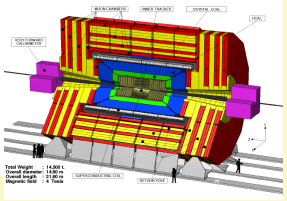
T1 center



UAM UNIVERSIDAD AUTÓNOMA DE MADRID

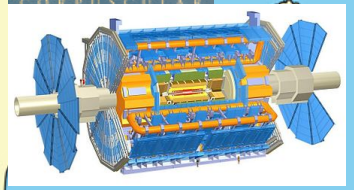
IFAE

Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas



Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

IFIC INSTITUT DE FÍSICA CORPUSCULAR



Tier2 centers in Spain connected to Tier1 via LHCONE provided by REDIris (10-20 Gbps)

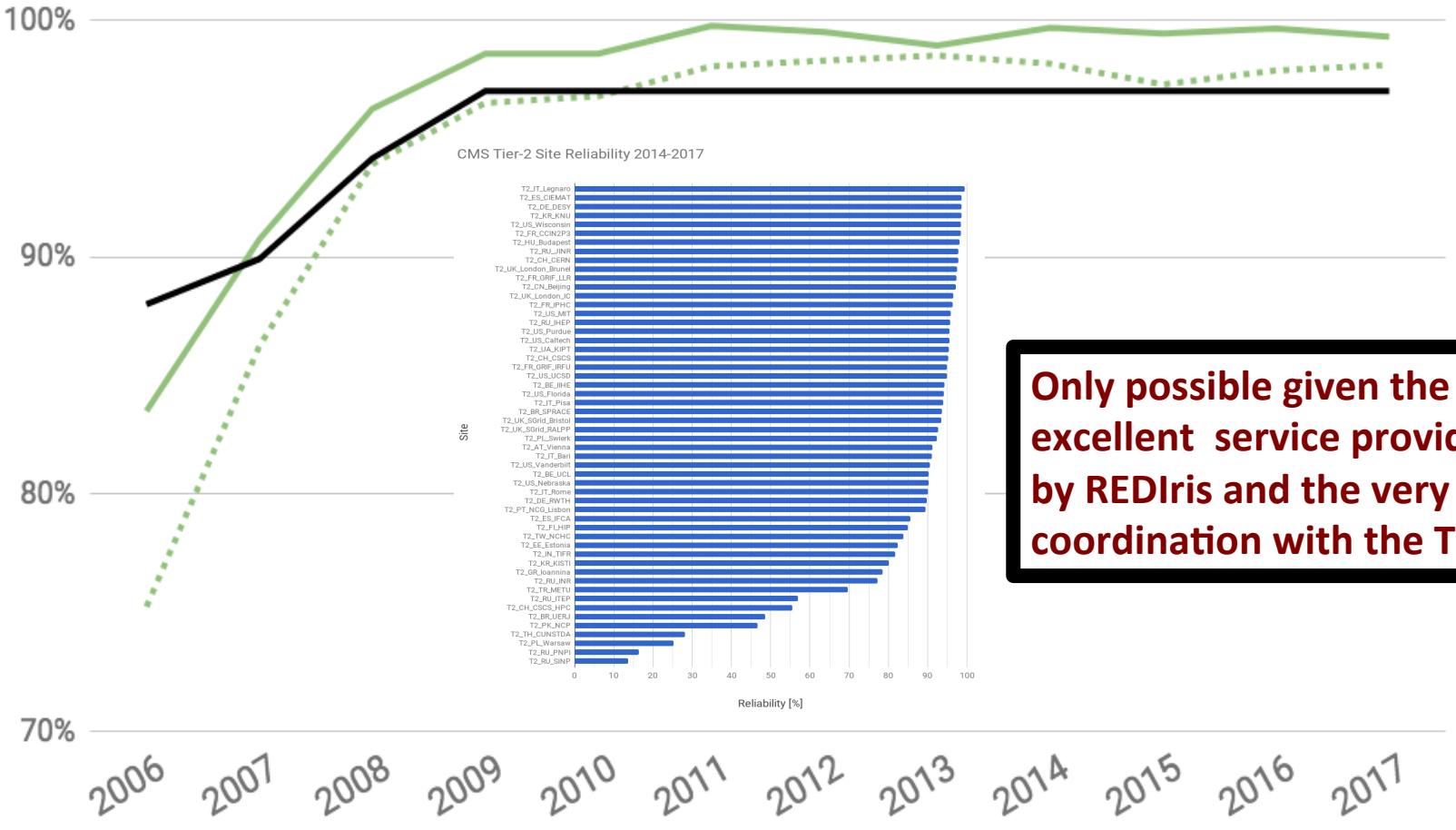
WLCG centers are these days among the largest generators of network traffic in Spain

Need for dedicated network connections for LHC Tiers and HPCs in the near future ?

WLCG @ Spain

Fiabilidad PIC Tier-1

About 5% of WLCG



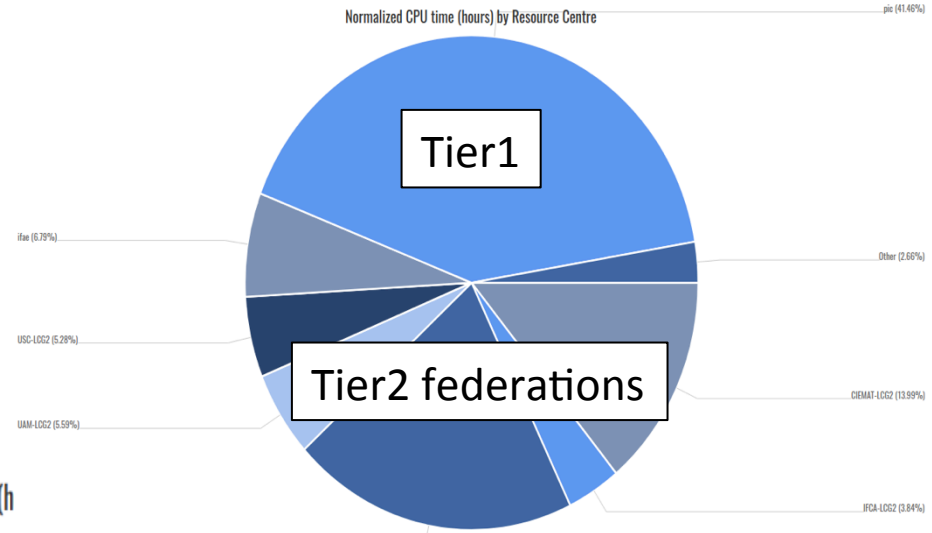
Only possible given the excellent service provided by REDIris and the very good coordination with the Tier centers

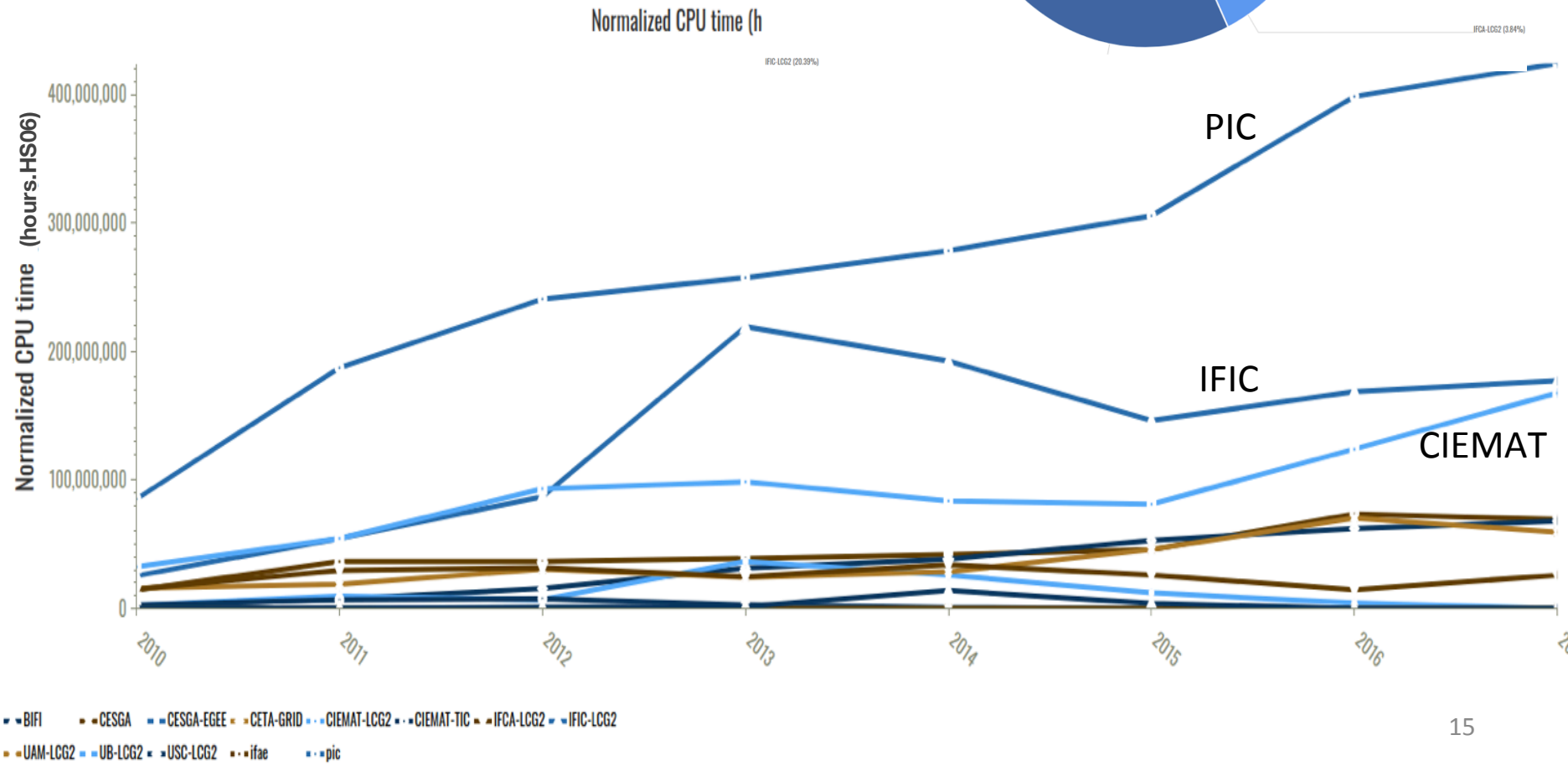
Consistent 24/7 excellent performance during a full decade

Spanish Tier1 and Tier2 centers at the top-3 worldwide in terms of performance

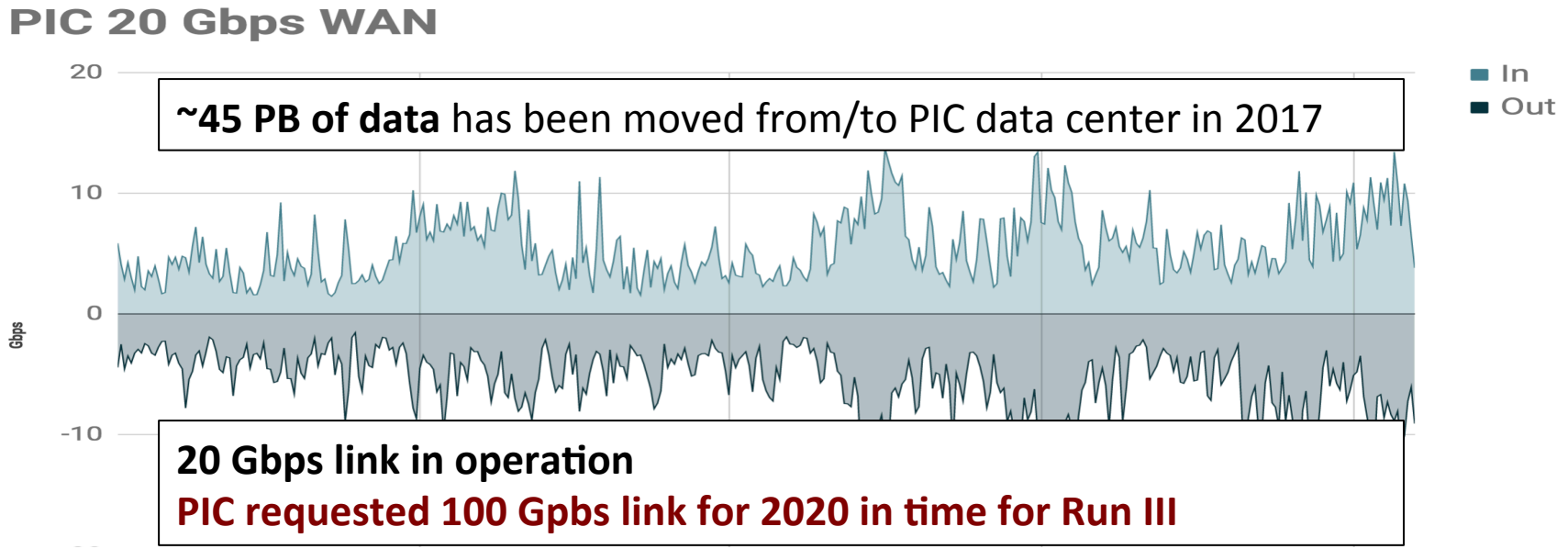
WLCG @ Spain

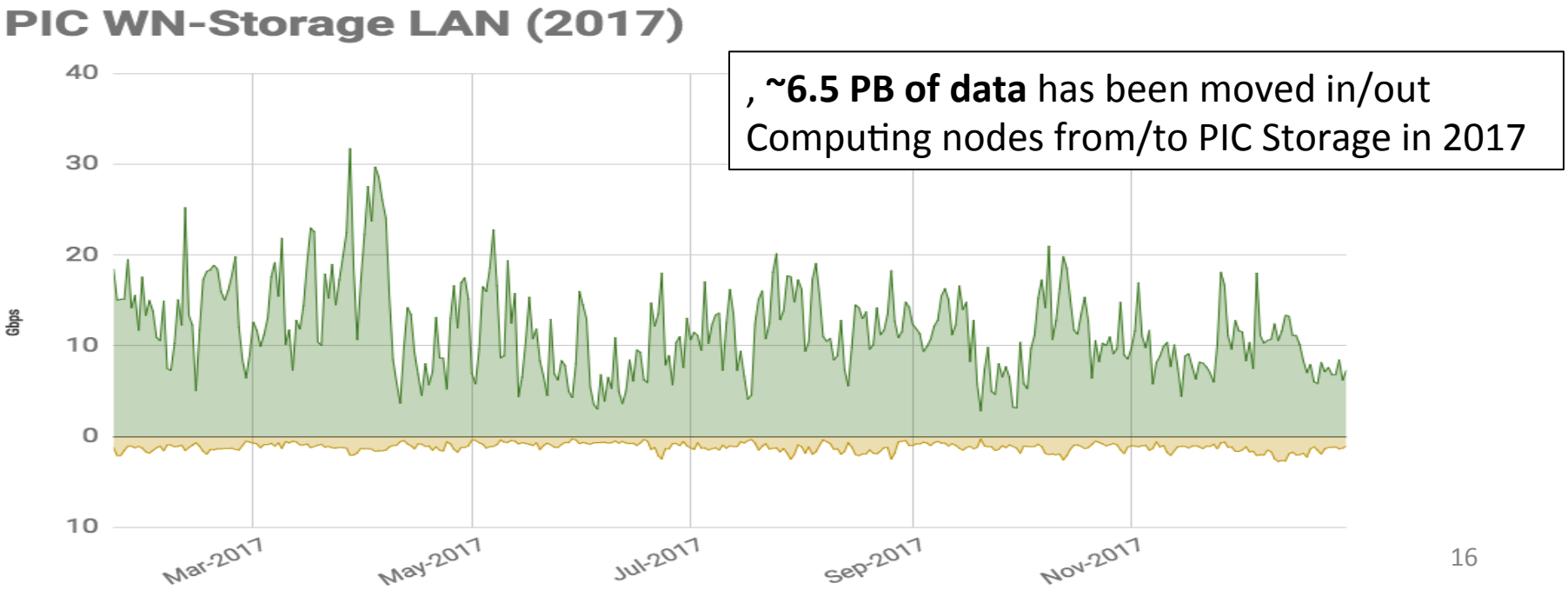
~500M CPU hours in 2010-2017



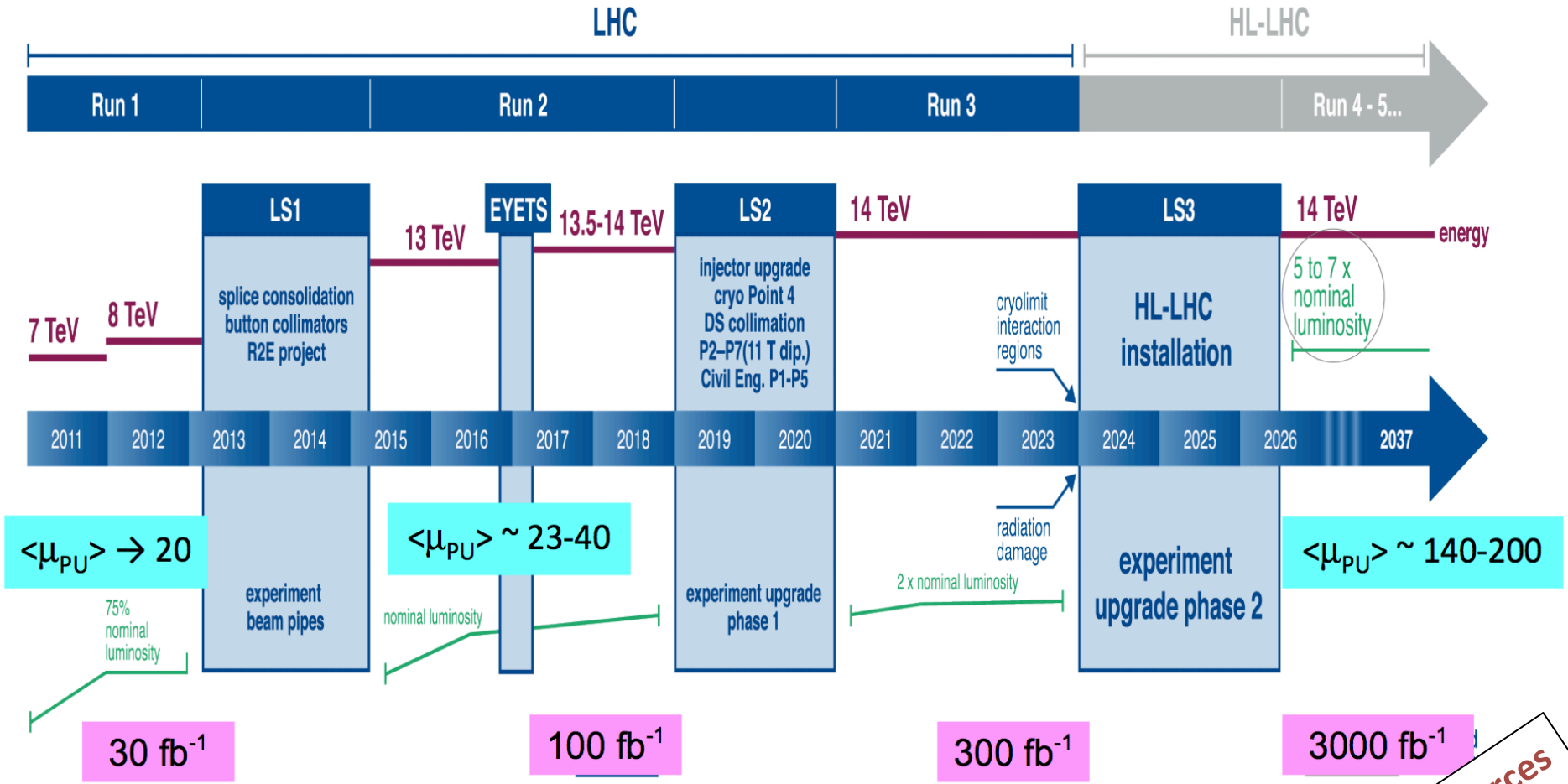


WLCG @ Spain


PIC 20 Gbps WAN
20
10
0
-10
Gbps
In
Out
~45 PB of data has been moved from/to PIC data center in 2017
20 Gbps link in operation
PIC requested 100 Gpbs link for 2020 in time for Run III


PIC WN-Storage LAN (2017)
40
30
20
10
0
10
Gbps
Mar-2017
May-2017
Jul-2017
Sep-2017
Nov-2017
, ~6.5 PB of data has been moved in/out
Computing nodes from/to PIC Storage in 2017

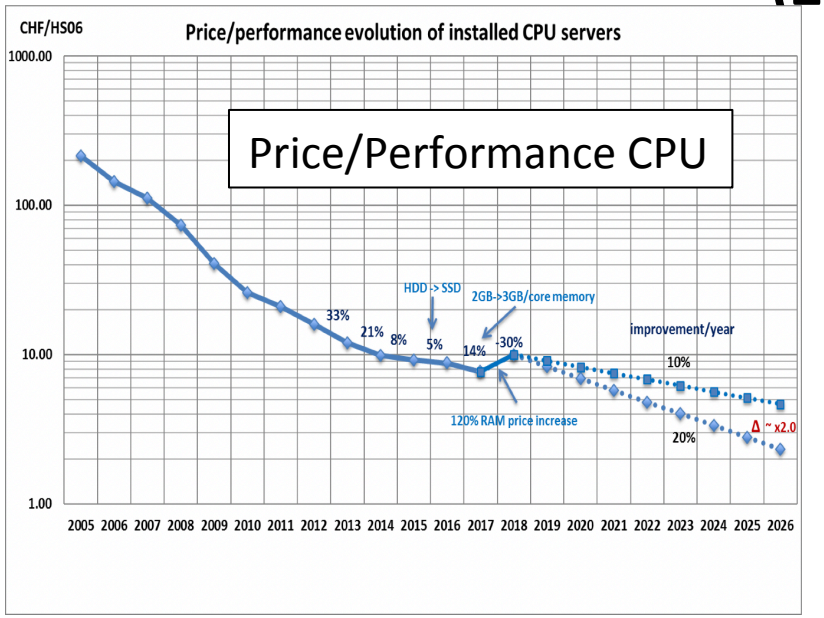
LHC Schedule

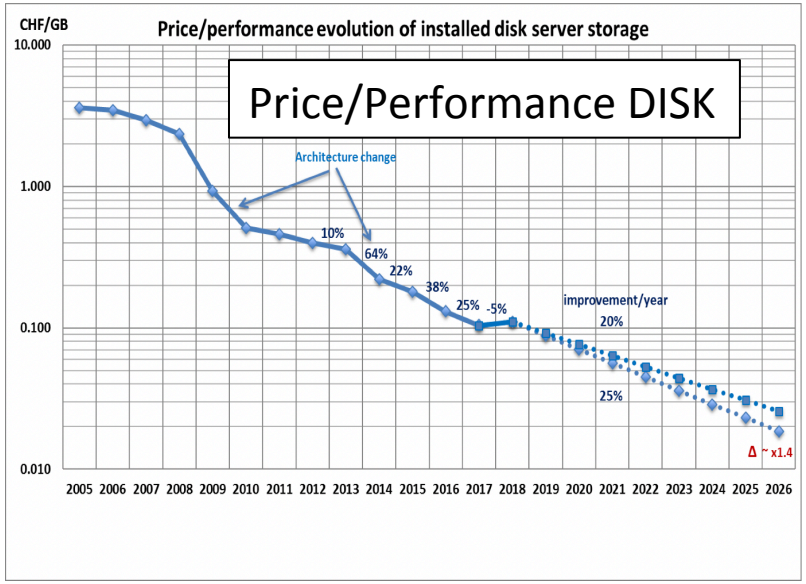


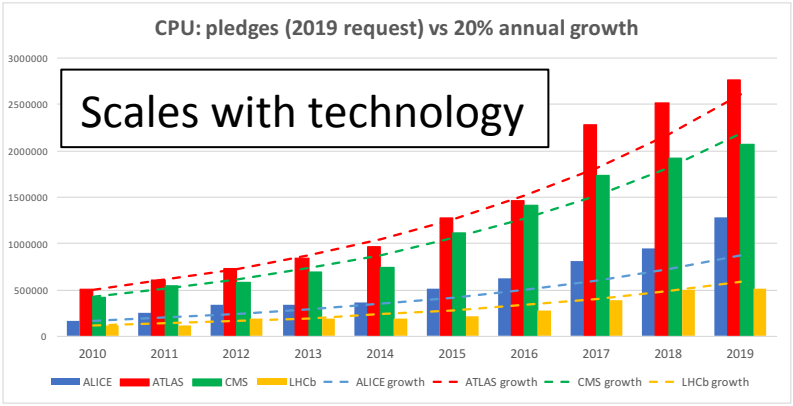
Computing for the LHC runs under the assumption of flat budget
Relying on a 20% increase each year of performance for same $$
Improvements in the experiment side (Software/Data Management)
Opportunistic use of voluntary resources
Use of Cloud and HPC resources
→ **Altogether will allow the experiments to go ONLY across Run III**

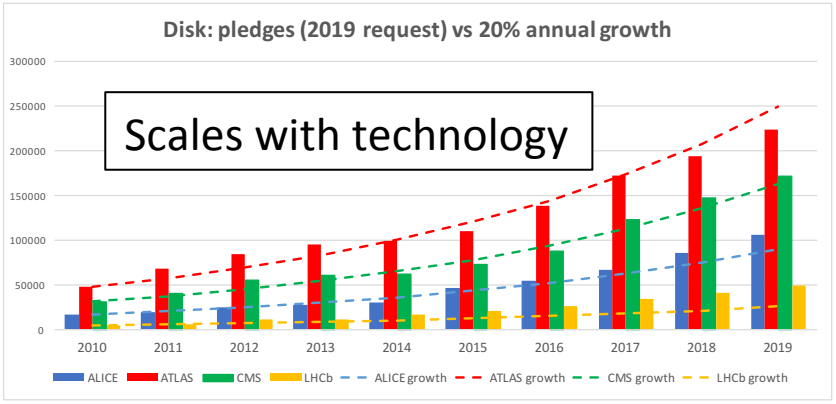
To Find Computing Resources for HL-LHC is a problem !

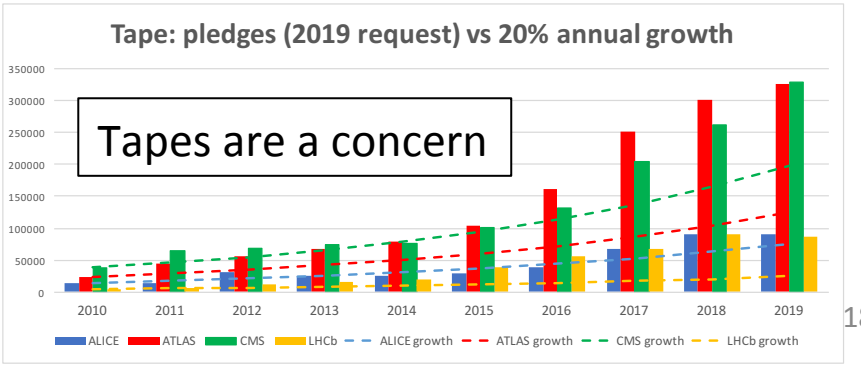
End Run II and preparation for Run III
(2018→2026)



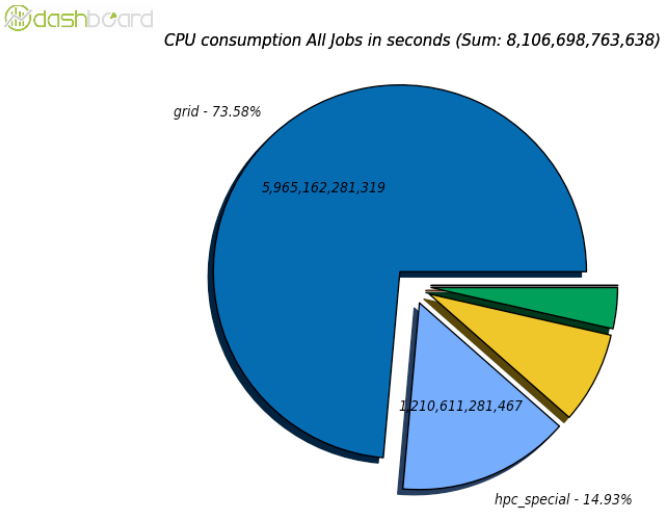






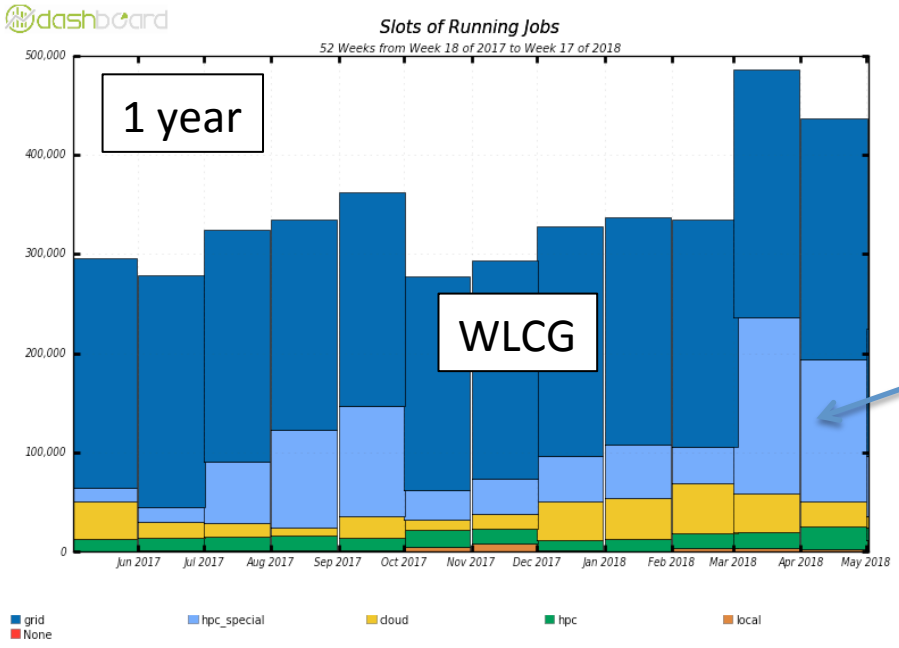


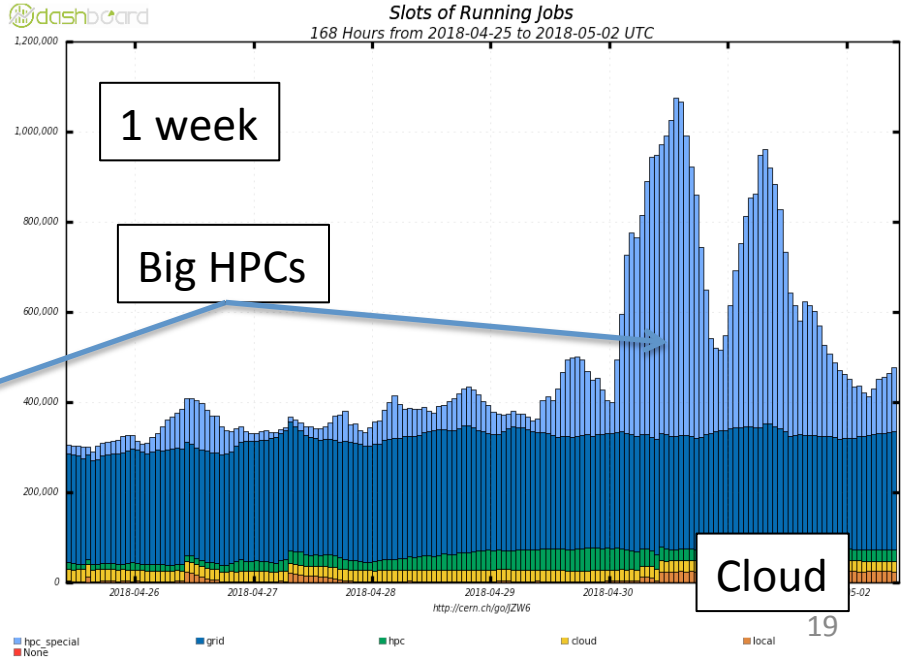
Using Clouds & HPC



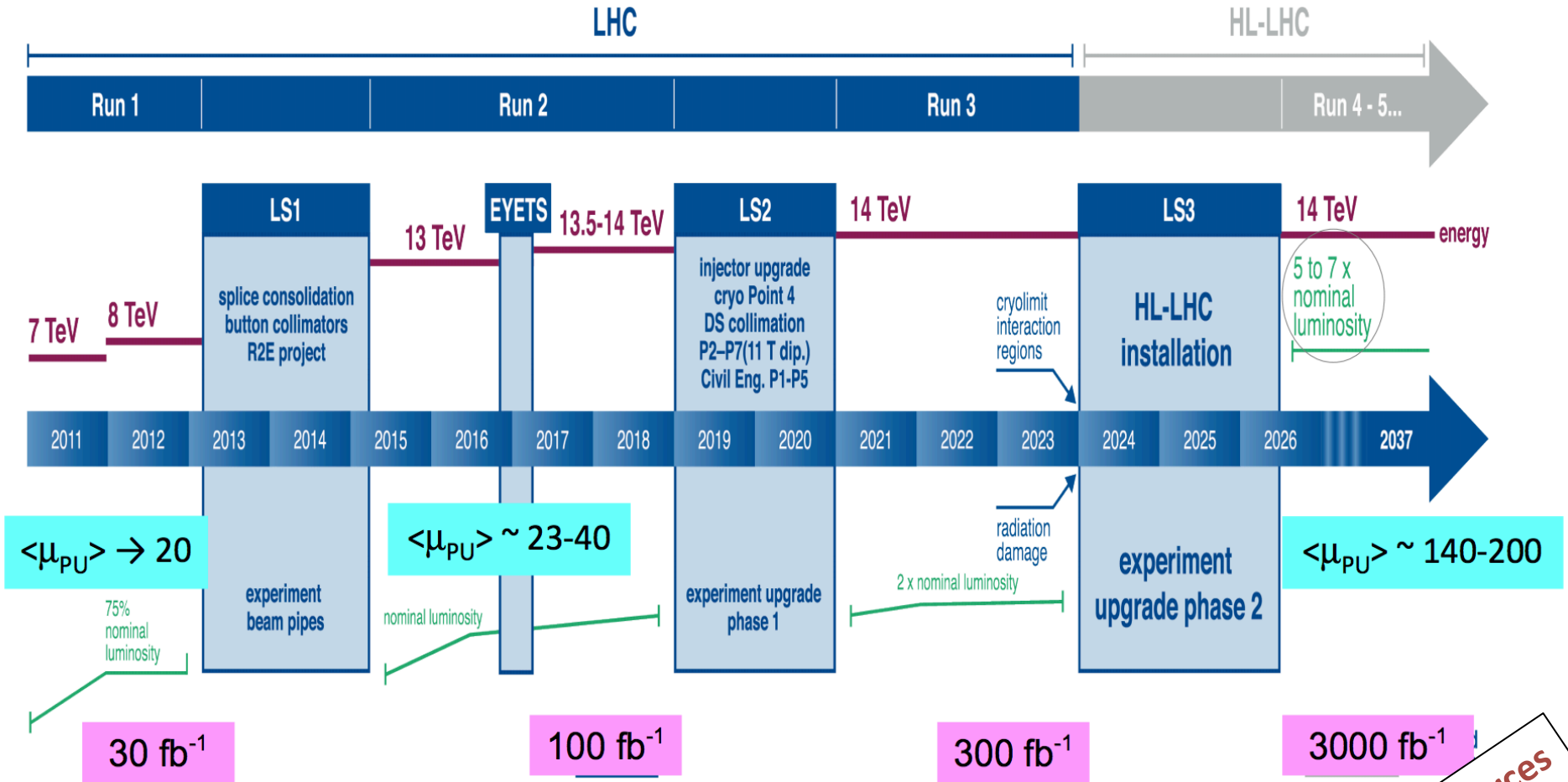
Experiments at the LHC are already very active in using opportunistic resources from HPCs and Cloud

These days in the ATLAS experiment
¼ of the resources come from HPCs + Cloud





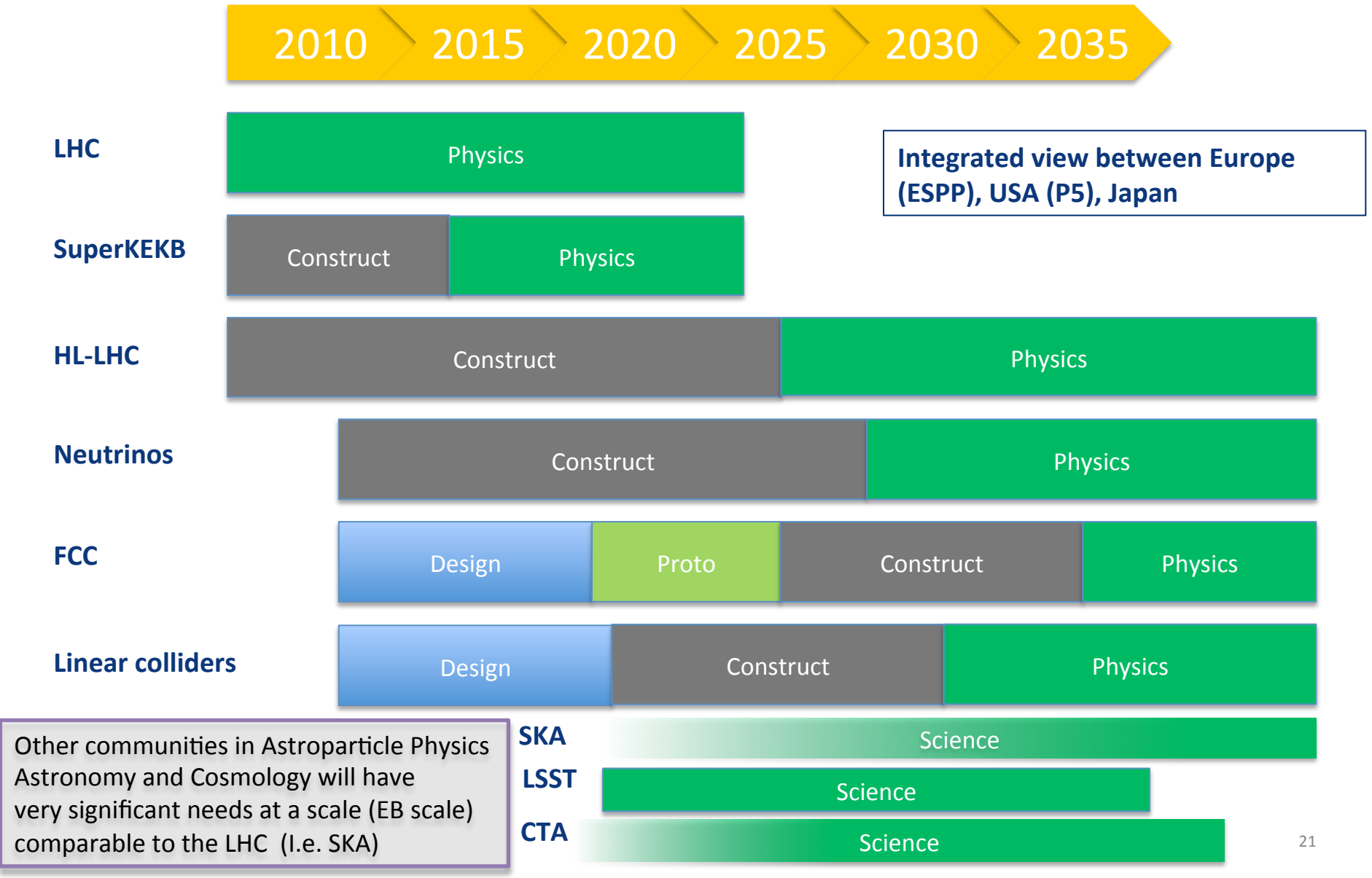
LHC Schedule



Computing for the LHC runs under the assumption of flat budget
Relying on a 20% increase each year of performance for same $$
Improvements in the experiment side (Software/Data Management)
Opportunistic use of voluntary resources
Use of Cloud and HPC resources
→ **Altogether will allow the experiments to go ONLY across Run III**

To Find Computing Resources for HL-LHC is a problem !

HL-LHC is not alone (Synergies)



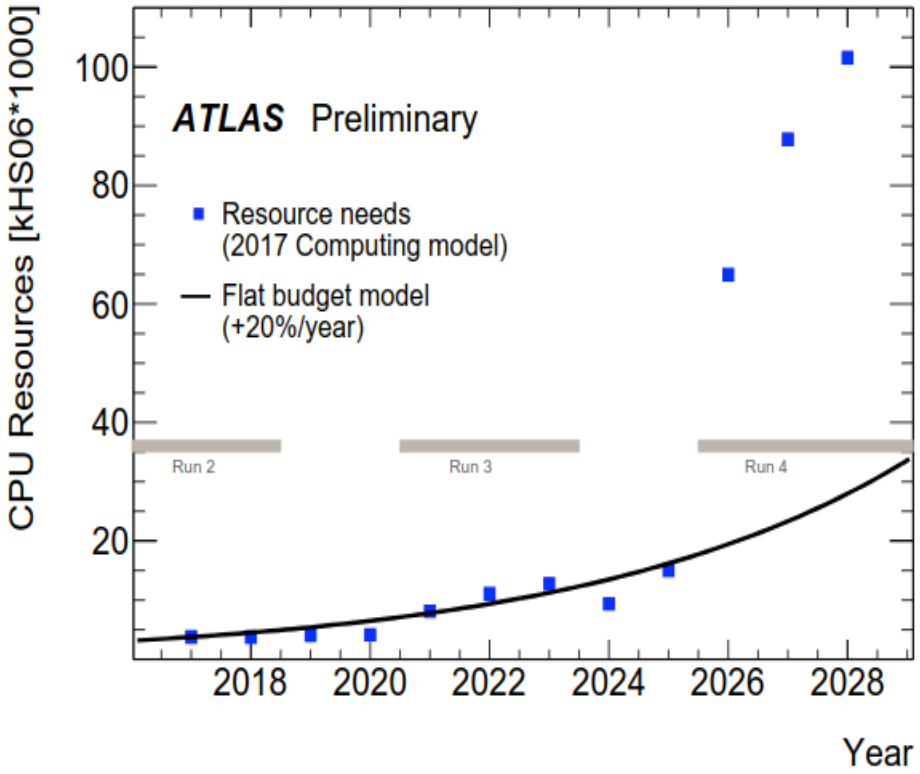
Integrated view between Europe (ESPP), USA (P5), Japan

Other communities in Astroparticle Physics Astronomy and Cosmology will have very significant needs at a scale (EB scale) comparable to the LHC (I.e. SKA)

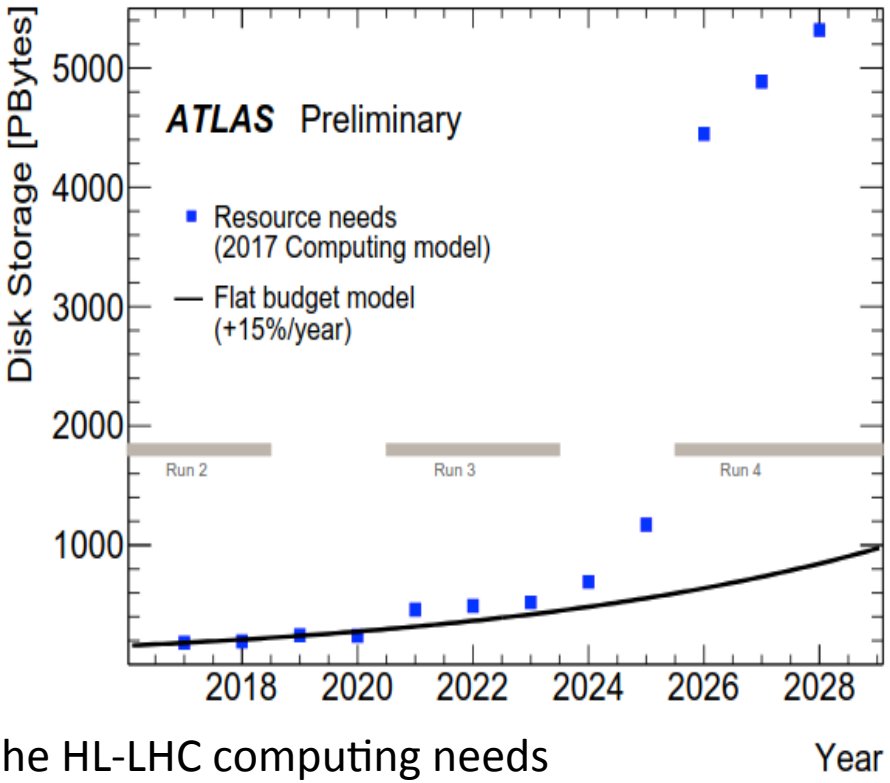
HL-LHC Challenge

By 2026 CMS+ATLAS: 10 EB, 15M cores
(Spain: 300 PBs, 500k cores)

CPU: deficit by a factor 3



DISK: deficit by a factor 5



Current WLCG model will not scale to face the HL-LHC computing needs

- → Factor 10 data rates and complexity
- → Factor 20-25 in CPU and storage
- → Technological evolution will not be sufficient

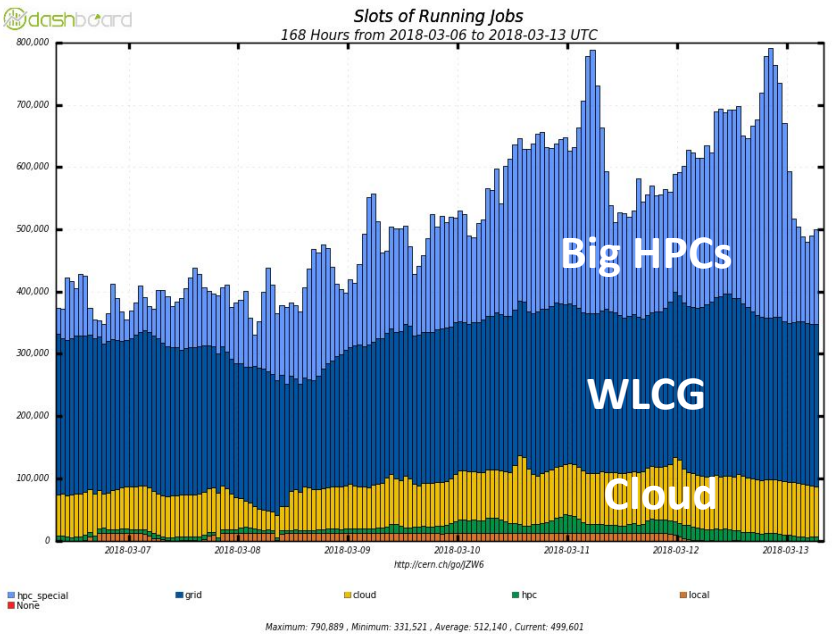
→Changing computing model is a must !

https://arxiv.org/abs/1712.06982

Computing @ HL-LHC

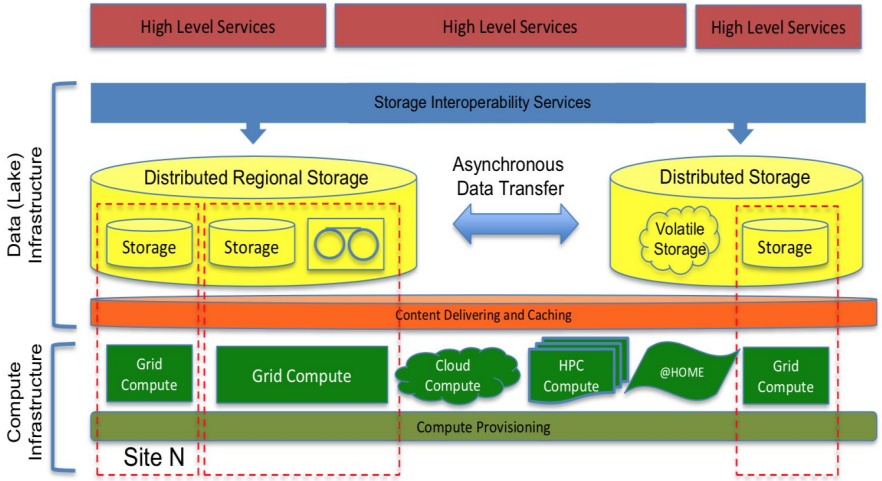
Scaling CPU (not the biggest problem)

- Room for further optimization in simulation and reconstruction software in the experiments
- Leveraging on co-processors/GPUs
- More opportunistic computing
  - HPC centers
  - Clouds



Scaling Storage (the big problem)

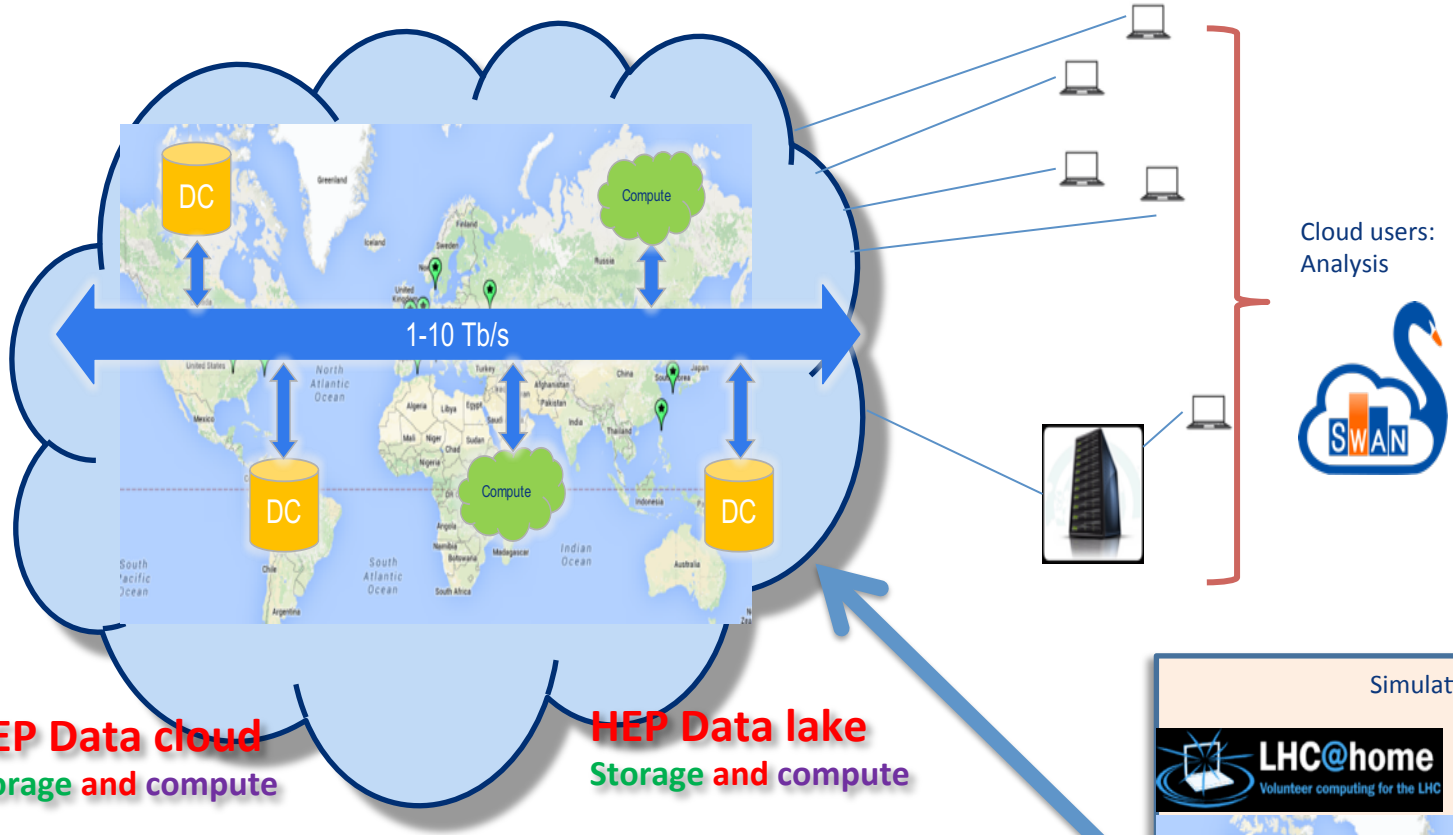
- No much room left on data format
- No much room left in the # of replicas
- Massive use of tape does not look like a solution (too slow, geographically limited, etc…)

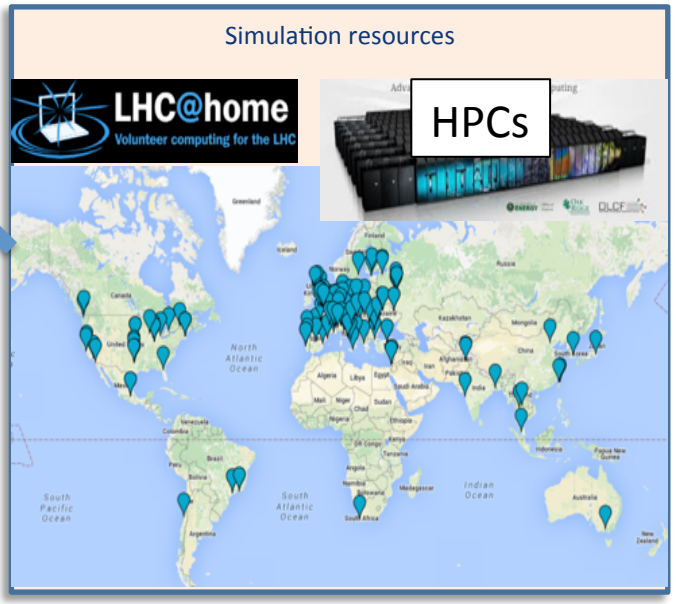


Call for new solutions → Data Lake ?

- We can count on ~100x bandwidth growth by HL-LHC (rely on network strength)
- Integrated consolidation of distributed storage (and compute) facilities, leveraging high-bandwidth networks (Data streaming to the client)

Model for Computing @ HL-LHC





- **Build a data cloud**
  - **Large centers connected with multi-Tb networks**
- **Pluggable compute capacity**
  - HPC, Cloud providers
  - For simulation (CPU-intensive)
- **Remote user analysis**

Notes on Computing for HL-LHC

- LHC computing model (WLCG) evolving towards less # of centers and more powerful (MWatts) ….model being put forward by CERN..
  - LHC Data Cloud from large reliable centers connected with Terabit networks (data backbone)
  - Rest of sites would consume data and deposit simulation outputs in the data backbone ..

- Facilitate the integration of resources from HPC, commercial clouds, volunteer computing….again consuming and depositing data to the backbone.

→ Portugal & Spain should actively explore possible solutions to be part of this selected club → Federation of centers inside Spain/Portugal, relationship with local HPCs, etc..

Notes on Enabling Technologies

- **Wide Area Network**
  - LHCOPN private network (T0, T1) and LHCONE overlay network
    - Some countries migrating to 100 Gpbs (e.g. US)
    - **Plan in RedIris to deploy 100 Gbps infrastructure?**
  - **High data traffic between centers**
    - Data distribution and replication, O(PB/day)
    - WAN data access

      (application running in one center read data from a remote center)

- Distributed high throughput computing (HTC) technologies
  - **Workload management system** to join distributed CPU resources in a common pool and schedule and match available slots with workloads
  - **Data federation** to enable data streaming through the WAN from storage to applications
  - **Authentication and authorization** system for secure access to services

Notes on HPCs for LHC

- **Clearly HPC resources will be part of the HL-LHC computing solution**
- Well suited for simulation (CPU-intensive, >50% needs in LHC)
- Other (data-intensive) workflows much more difficult to run at scale
  - >1 MB/s/core => 10 Gbps/1000 cores WAN
- **HTC requirements for HPC**
  - Need to connect to an automated workflow management system
  - External connectivity from compute nodes (connection to external WMS, data I/O)
  - Access to network from applications running at HPCs
  - Virtualization platform to execute applications
  - Edge services (workload entry point, caches, mount application software repository)
- **Substantial integration effort needed**
  - Integration underway in other countries
  - **Technical collaboration between HPC and HTC communities**
  - **LHC experiments delegate integration to local HTC communities**
  - Demonstrate running in production at scale before accepting HPC pledges from countries
  - Model for resource allocation

Summary/Final notes

- **WLCG has been a model of success for providing computing resources to the LHC during the last decade**
- **WLCG model will not scale as needed to face the HL-LHC period (2026 →)**
- **Synergies with other large-scale scientific adventures**
- **R&D on new model for HL-LHC computing**
  - **Cluster of large centers connected with multi-Tb networks**
  - **Relying on powerful networks (RedIris a main player)**
  - **Active involvement of HPC centers & Clouds**
  - **Will require close coordination of WLCG and HPC worlds**
  - **Will probably require enhanced/devoted network connections for Tiers and HPCs**
- **A Portugal/Spain active collaboration in facing the next 10-15 years of computing challenges for big science is a key for success**