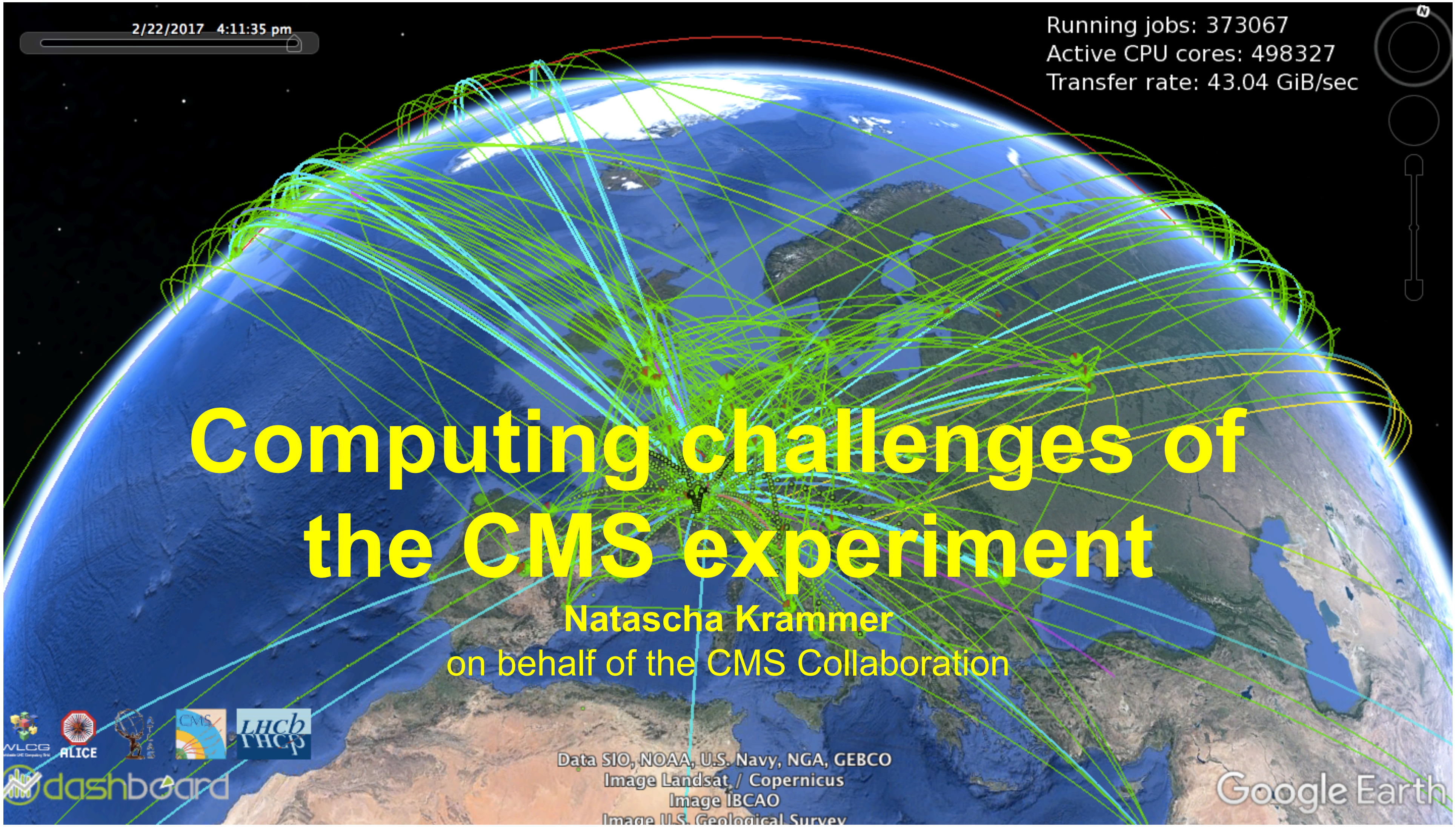


Computing challenges of the CMS experiment

Natascha Krammer
on behalf of the CMS Collaboration

27th February 2017

Outline

- Present and future computing needs in
  - LHC Run 2 and HL-LHC

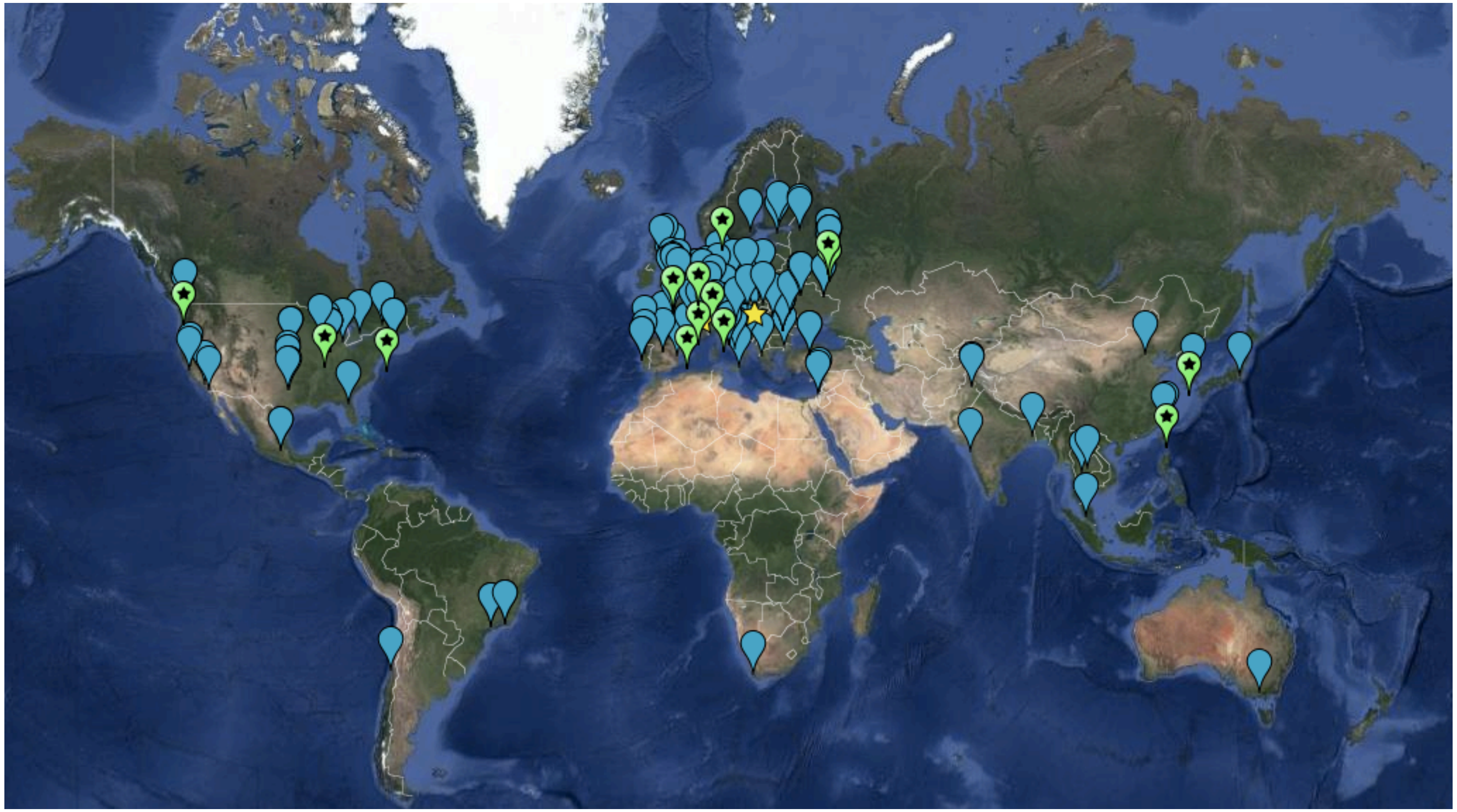
LHC RUN 2 HiLumi HL-LHC PROJECT

- How to full-fill the computing needs
  - cloud computing, High-Performance Computing

Google Cloud Platform amazon webservices™ HELIX NEBULA THESCIENCECLOUD

WLCG collaboration

- WLCG landscape
  - Over 170 computing centers in 42 countries
  - CPU: ~350.000 of today's fastest cores (3.8 million HEPSpec06)
  - Storage: Disk: 310 PB, Tape: 390 PB



LHC Run 2 computing needs

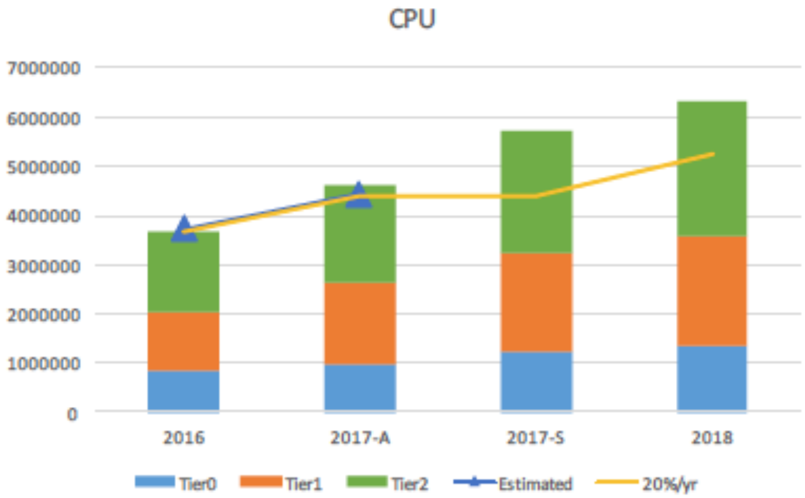
- LHC Run 2 performance is above expectations
  - all factors driving computing have increased above expected levels
  - p-p collisions

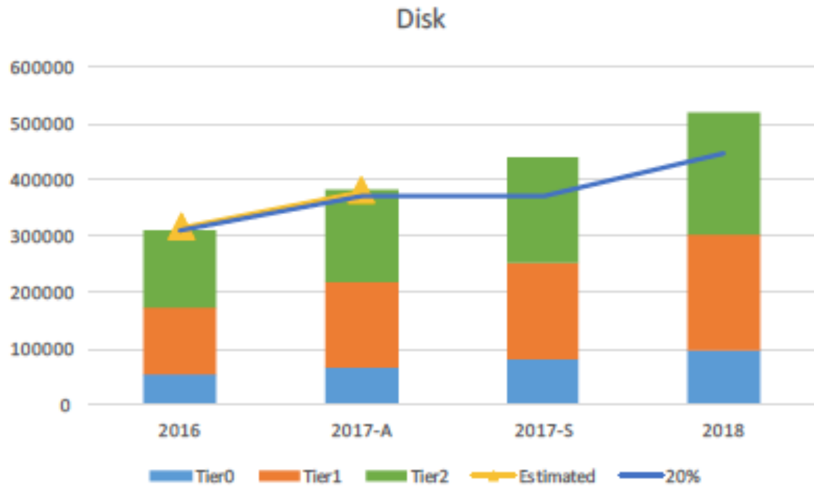
	2016 exp.	2016	2017 exp.	2018 exp.
Peak Luminosity (10^{34} $cm^{-2}s^{-1}$)	1.0	1.5	1.7-1.9	1.7-1.9
Integr. Luminosity (fb^{-1})	25	40	~45	~45

- For 2016, the available resources were sufficient
  - More tapes at CERN have been needed
- Analysis for 2017, 2018
  - Expectations are increased, the requirements are ~20% above previous estimates

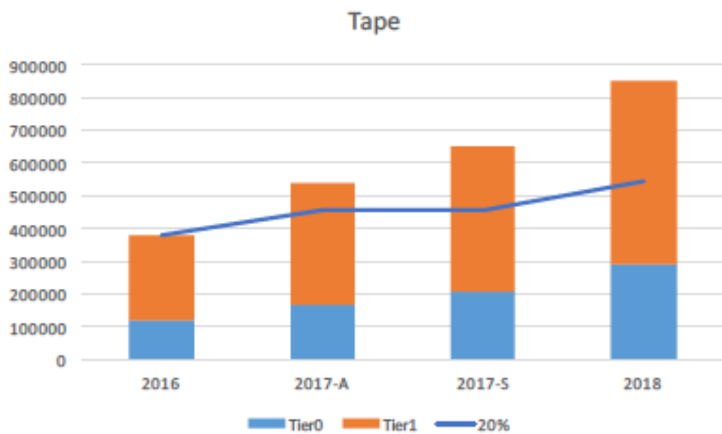
LHC Run 2 computing needs

- WLCG estimates made in 2014 for LHC Run 2 up to 2018
- Growth of 20%/year starting in 2016 (“flat budget”)



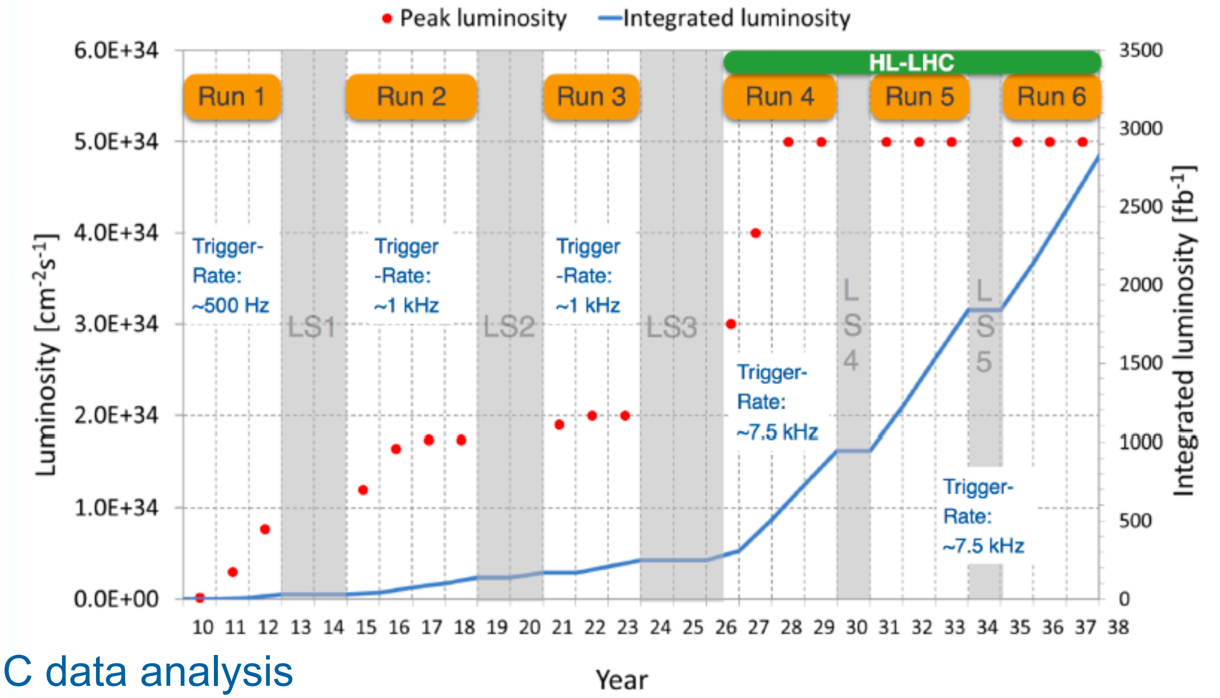


Ian Bird, CWP Workshop, Jan. 2017



- 2016➔2018
  - CPU 3.8➔6.3 (million HEPSpec06)
  - Disk 310 PB➔520 PB
  - Tape 390 PB➔850 PB

HL-LHC computing

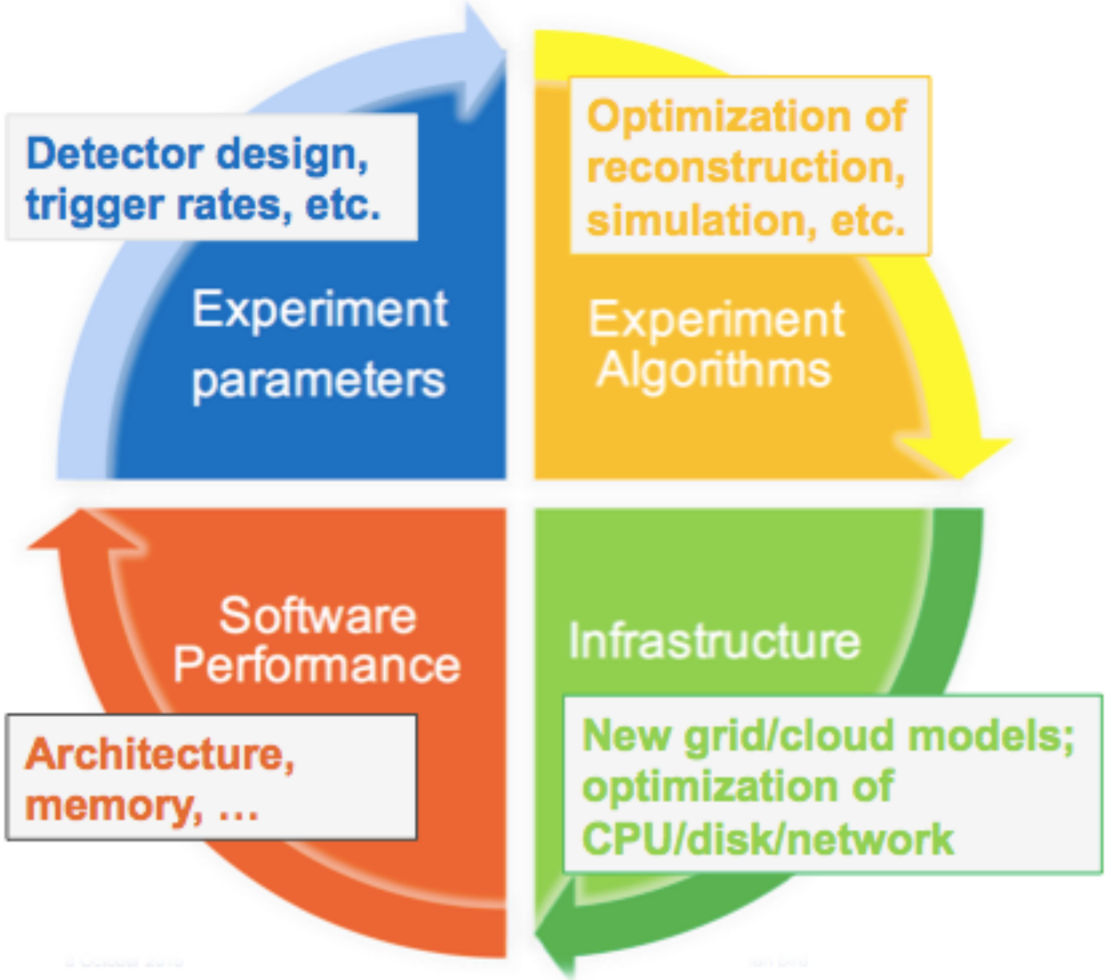


Oliver Gutsche, CHEP 2016

- HL-LHC data analysis
  - To extract physics results requires to handle/analyze a lot more data!
  - Tests started with new technologies
    - "Big Data" technology (new toolkits and systems to support analysis of datasets in industry)
    - Cloud Computing
    - High Performance Computing (HPC)
  - Educates our community to use industry-based technologies
  - Use tools developed in larger communities reaching outside of our field

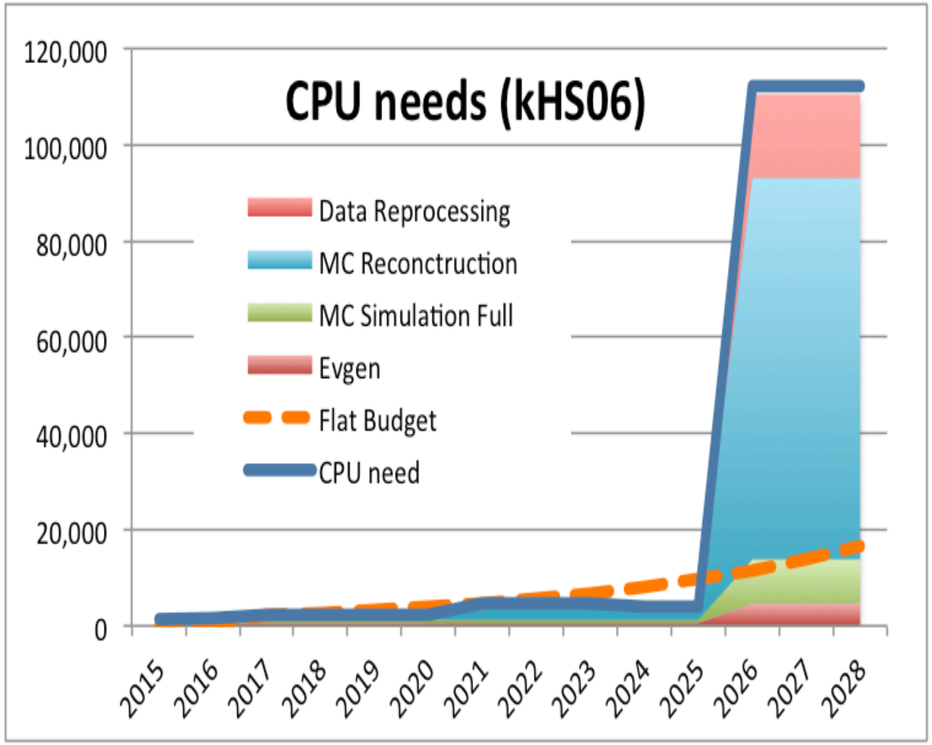
HL-LHC computing cost parameters

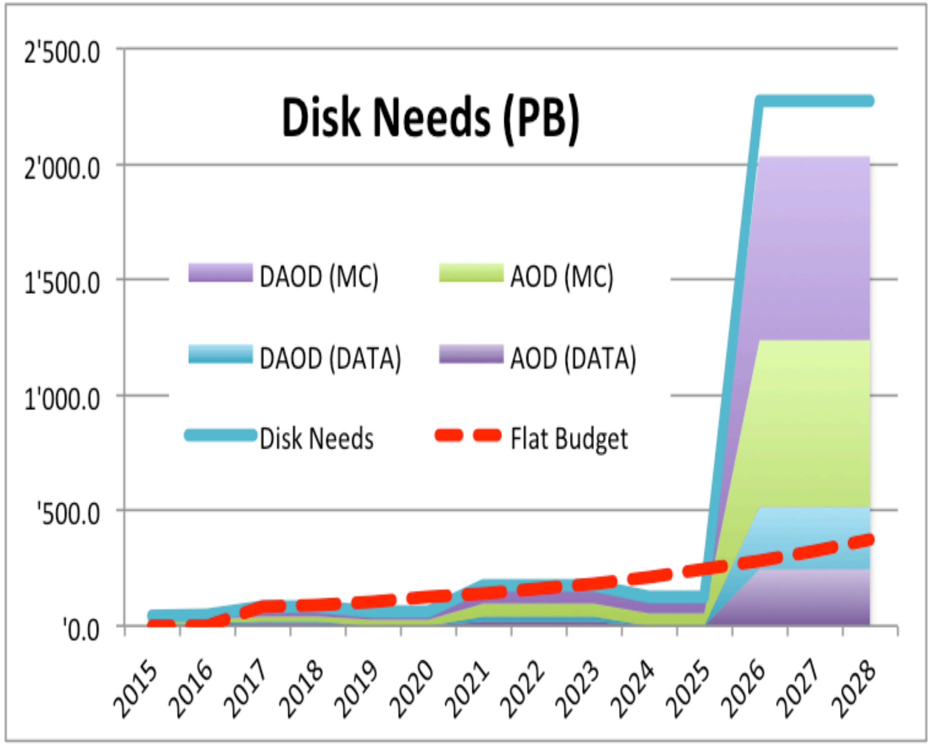
- Future computing cost drivers



Ian Bird, CWP Workshop, Jan. 2017

➢ Initial studies on computing needs for HL-LHC



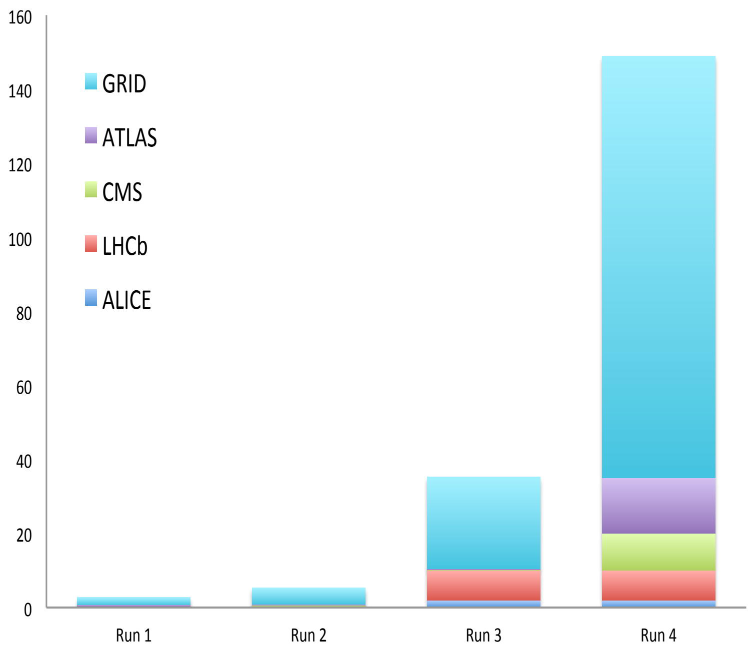


Frédéric Hemmer CERN School of Computing, Aug. 2016

HS06….HEPSpec 2006 based on SpecInt 2006

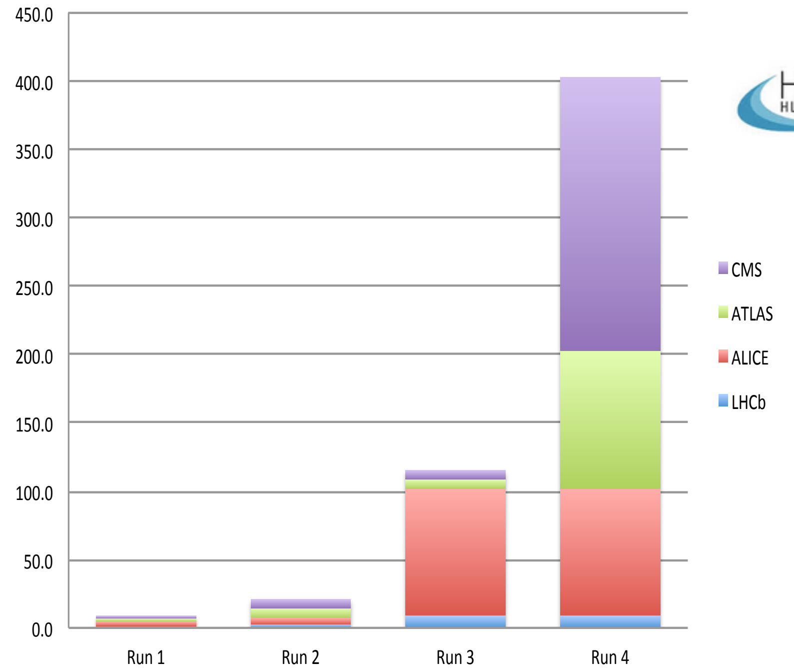
HL-LHC computing needs

Computing: Growth > x 50



2010 2015 2021 2025 (HL-LHC)

Data: ~25 PB/year → 400 PB/year, x 16

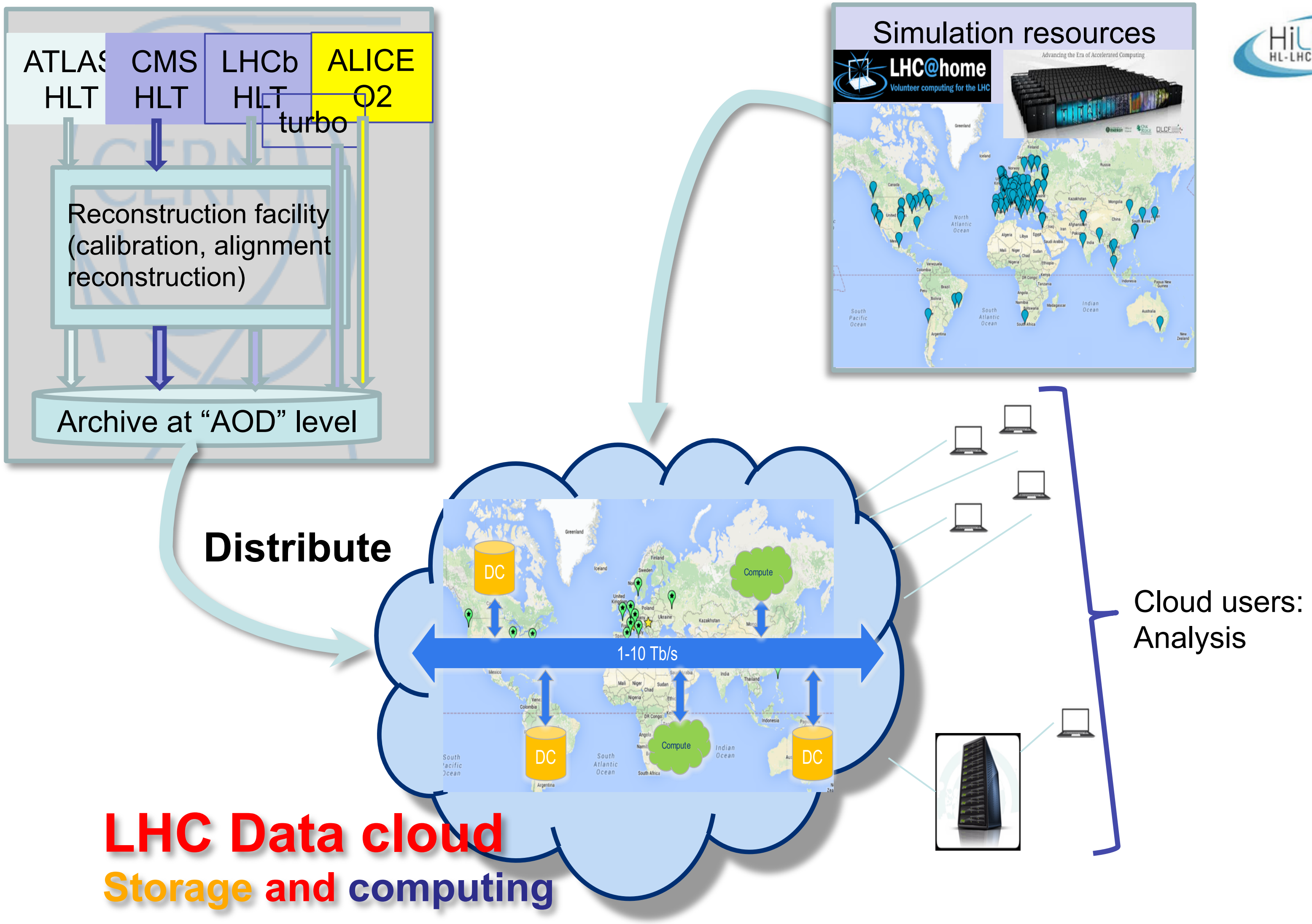


2010 2015 2021 2025 (HL-LHC)

Frédéric Hemmer ,CERN School of Computing, Aug. 2016

- Simple model based on today's computing models, but with operating parameters (pile-up, trigger rates, etc.)
- Technology at ~20%/year will bring x 6-10 in 10-11 years
- At least x 10 above is realistic to expect from technology with reasonable constant cost

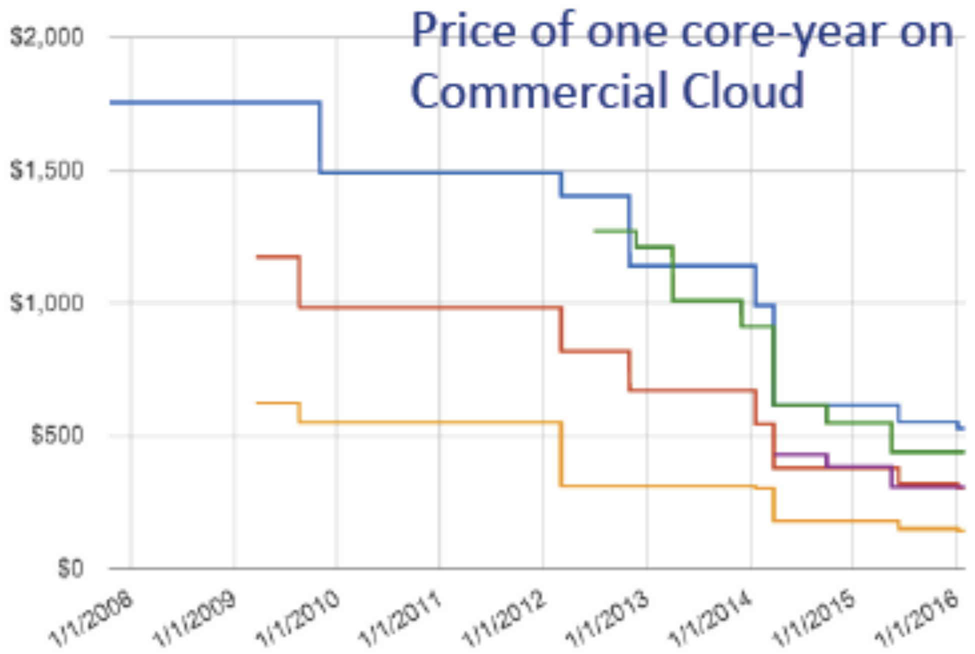
Possible Model for future HEP computing



Frédéric Hemmer, CERN School of Computing, Aug. 2016

HL-LHC computing needs

- HL-LHC computing needs will be > 50 x current capacity
- Commercial clouds can provide increased capacities for decreased costs compared to the past
- Elasticity usage needed – Usage not steady-state



Burt Holzman, CHEP 2016

Computing Resource Providers

- Classes of Resource Providers

GRID

Virtual Organizations (VOs)
→ Pledges

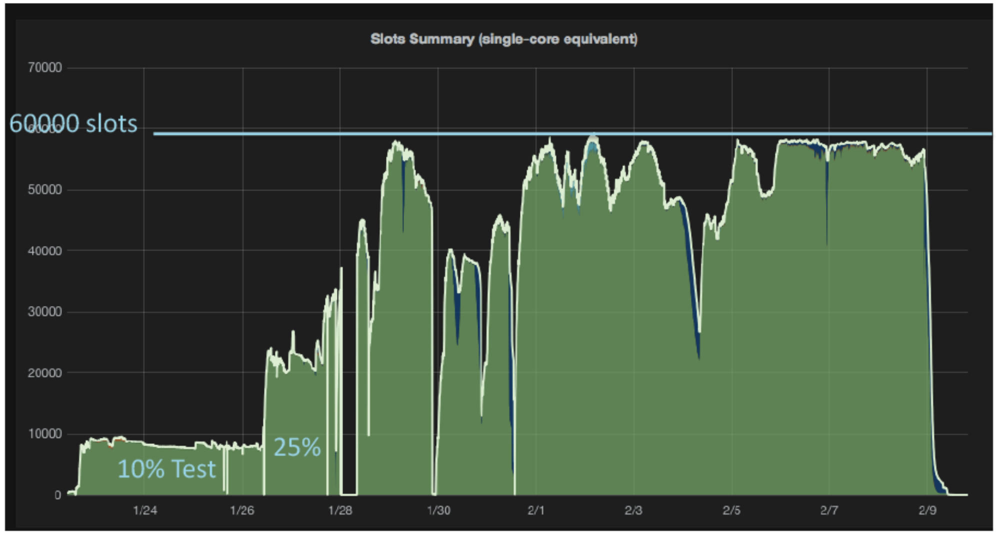
Cloud

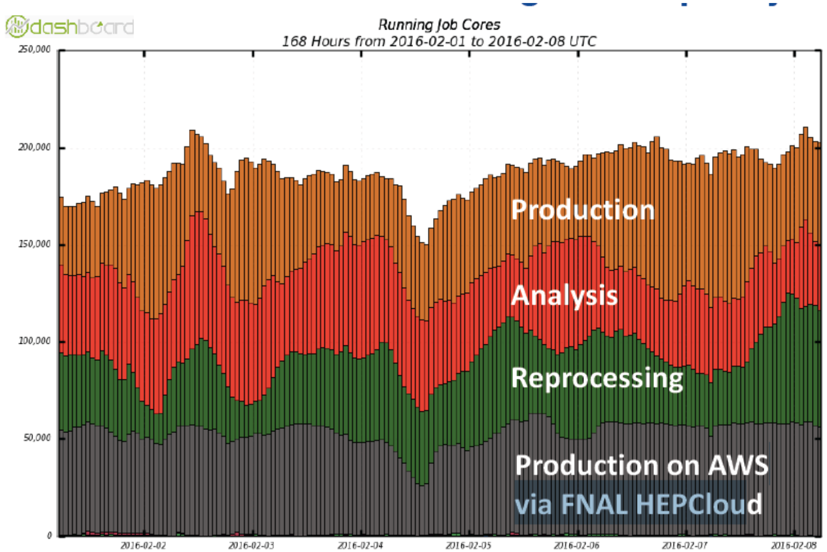
Community Clouds and Commercial Clouds
→ Rented resources

HPC

Researcher granted access to High-Performance Computing installations
→ Exploit opportunistic resources

- Fermilab HEPCloud and CMS use case (2016) - Amazon Web Services (AWS)
  - Reaching ~60k slots
  - 25% of CMS global capacity
  - CMS simulation
    - 2.9 million jobs, 15.1 million wall hours
    - 518 million events generated



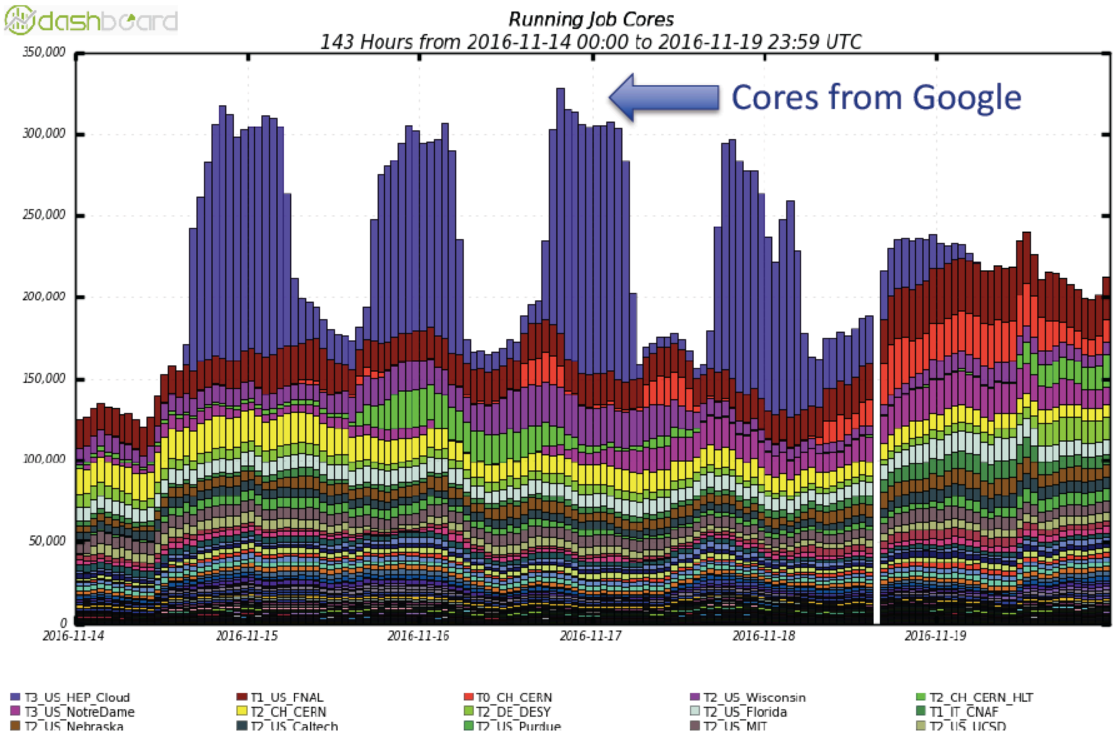


Oliver Gutsche, HSF CWP, Jan. 2017

Fermilab HEPCloud / Google

- Fermilab HEPCloud and CMS use case (2016) – Google Cloud Platform
  - Double the size of global CMS computing resources
  - Aiming to generate 1 billion events in 48 hours
  - 730.000 simulation jobs, 6.35 million wall hours used
  - 205 million physics events generated, yielding 82 TB of data

Google Cloud Platform



Oliver Gutsche, HSF CWP, Jan. 2017

Fermilab HEPCloud / AWS

- On-premises vs. cloud cost comparison – Amazon Web Services (AWS)
  - Average cost per core-hour
    - On-premises resource: 0.9 cents per core-hour (includes power, cooling, staff)
    - Off-premises at AWS: 1.4 cents per core-hour (ranged up to 3 cents per core-hour at smaller scale)
  - Benchmarks - specialized (“ttbar”) benchmark focused on HEP workflows
    - On-premises: 0.163 ttbar/s (higher = better)
    - Off-premises: 0.158 ttbar/s

Computing performance roughly equivalent

Cloud costs higher – but approaching equivalence

Helix Nebula Science Cloud

- Helix Nebula Science Cloud
  - European hybrid cloud platform that will support high-performance, data-intensive scientific computing
  - for end-users from many research communities: High-energy physics, astronomy, life sciences,…
- sponsored by 10 of Europe's leading public research organizations and co-funded by the European Commission (H2020). Procurers: CERN, CNRS, DESY, EMBL-EBI, ESRF, IFAE, INFN, KIT, SURFSara, STFC
  - Funds (> 1.6 MEuro), manpower, use-cases with applications & data, in-house IT resources
- November 2016 the 4 winning consortia for the Helix Nebula Science Cloud have been presented

HELIX NEBULA
THESCIENCECLOUD

Ian Bird, Helge Meinhard, CWP Workshop, Jan. 2017

Summary

- Challenges on HL-LHC computing
  - HEP computing much more capacity is needed
  - New computing models and more efficient software have to be developed
  - Additional resources are needed – cloud computing, High-Performance computing
- Commercial cloud resources
  - Tests performed by Fermilab and CMS on Amazon Web Services and Google Cloud Platform
  - Proven capability to execute efficiently both data intensive and CPU intensive workflows
  - Cloud resources are much more competitive in terms of cost than in the past
  - Potentially an interesting resource to supplement to the existing resources