CERN

Computing Challenges for LHC

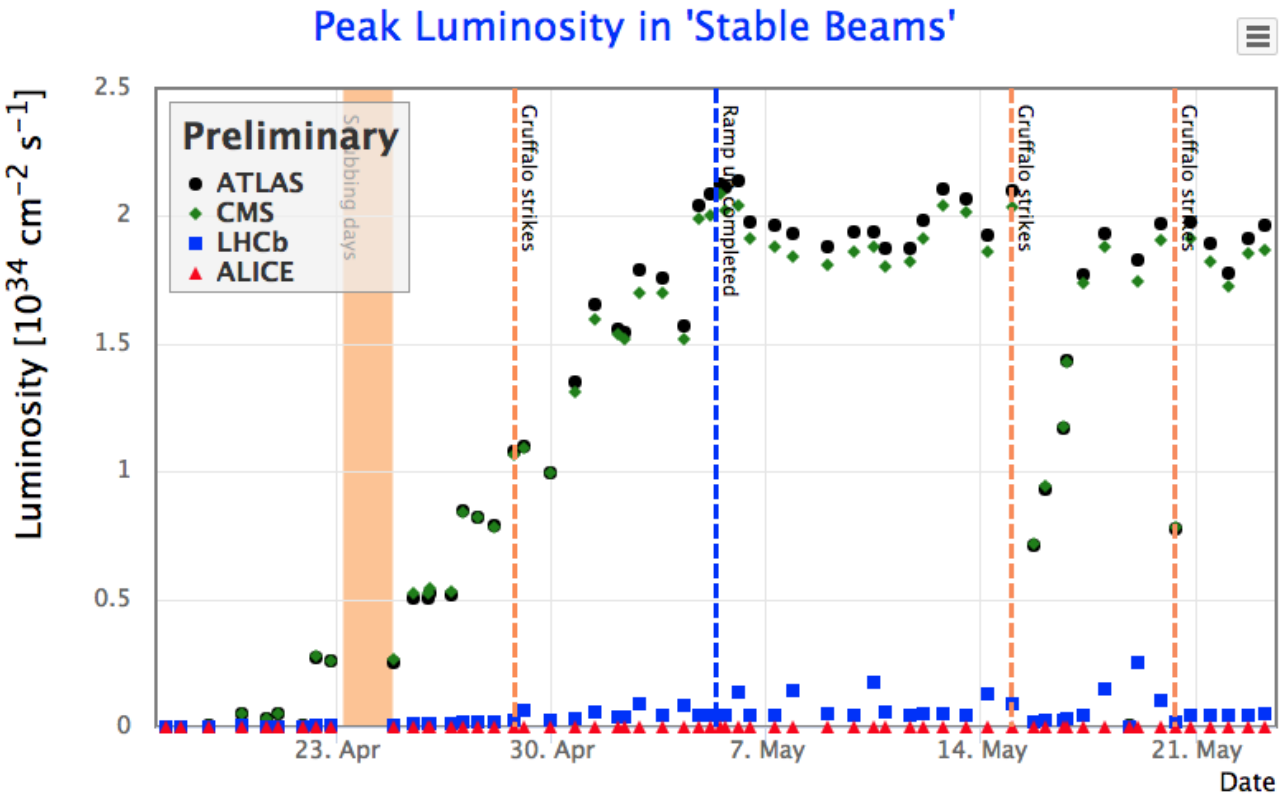
CERN Thematic School of Computing

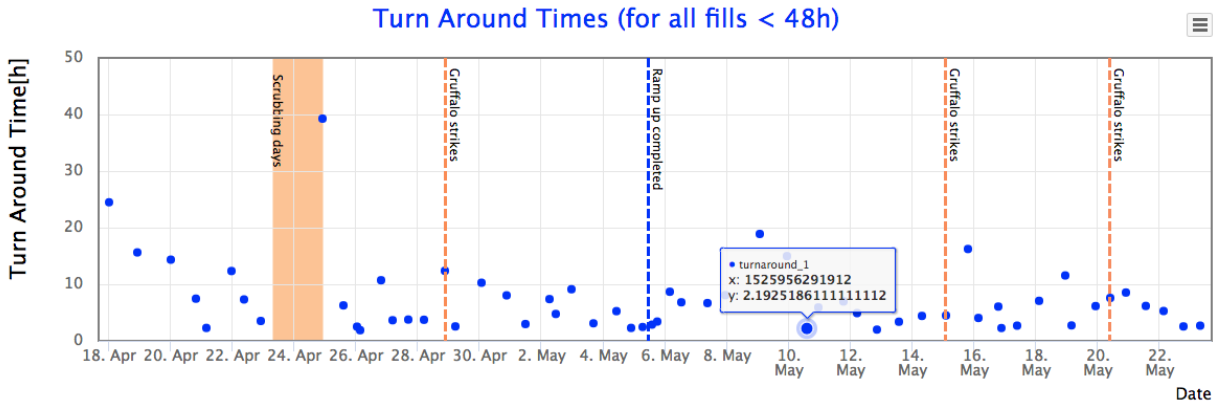
High Throughput Distributed Processing
of Future HEP Data

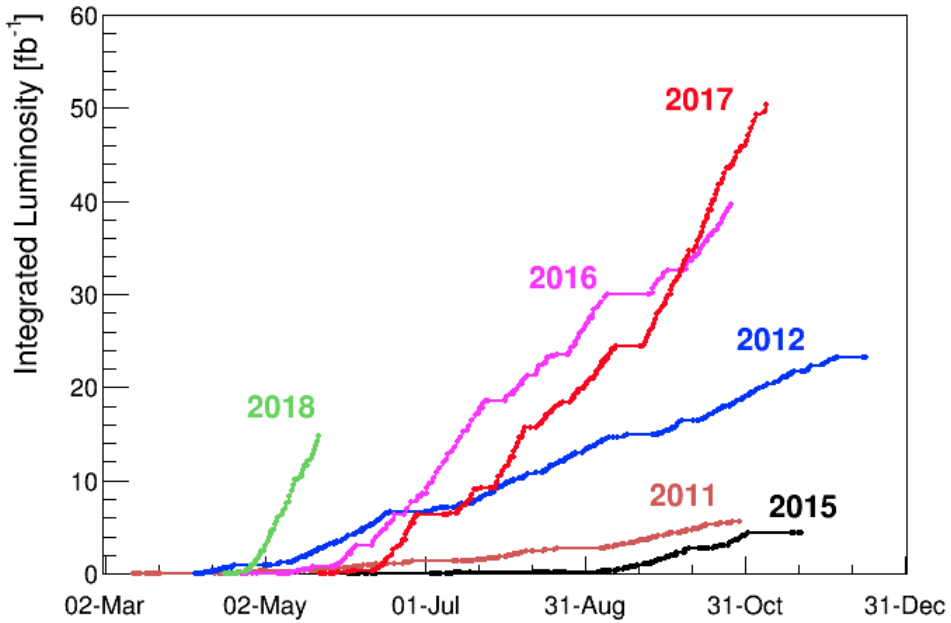
Split, Croatia

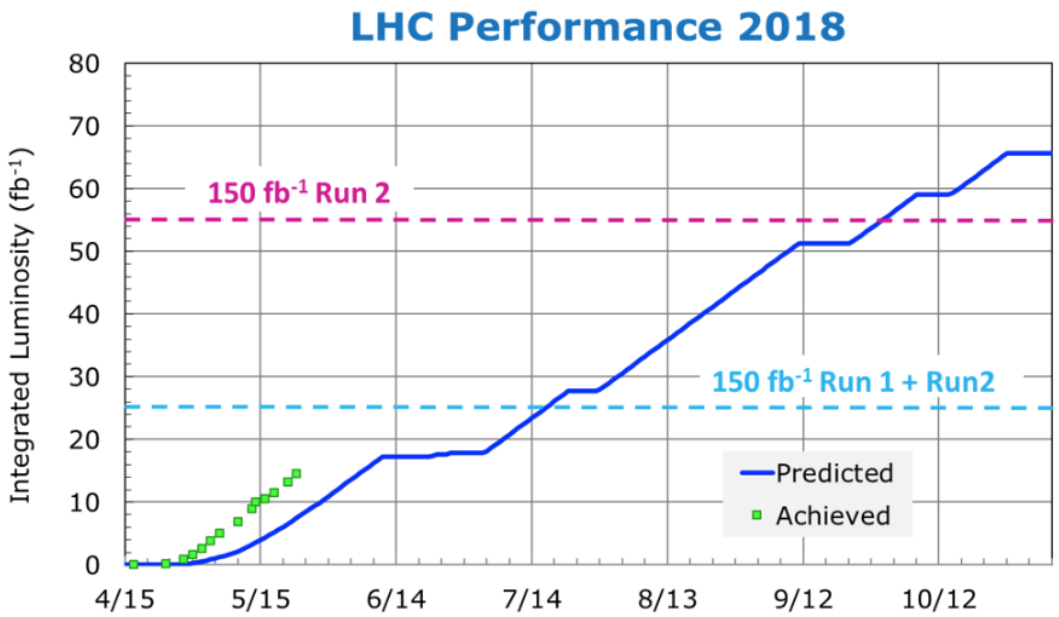
Frédéric Hemmer
CERN - IT Department

Progress - 2018



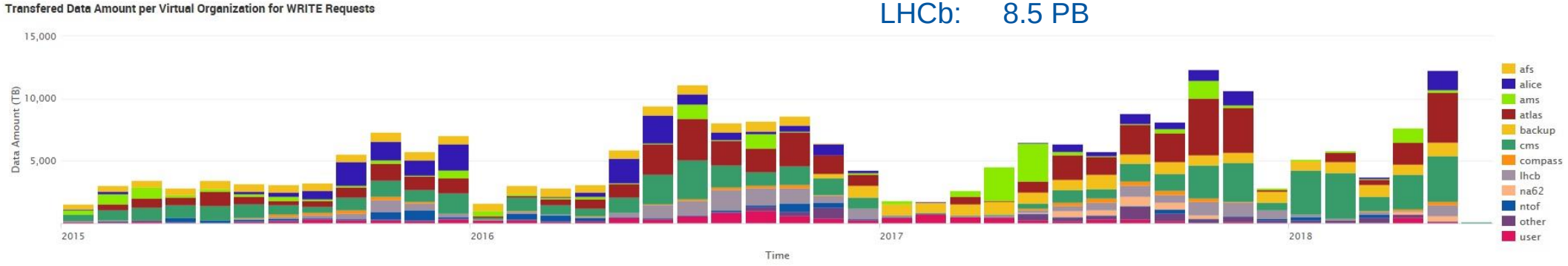


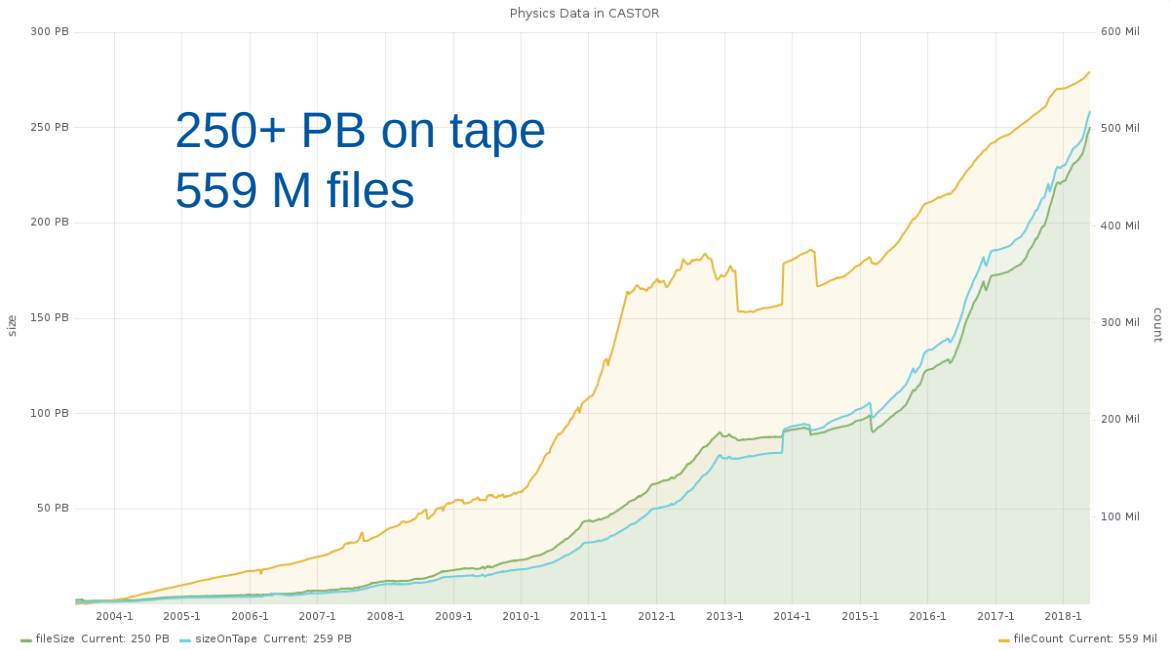




Data in 2018

ALICE: 7.6 PB
ATLAS: 17.4 PB
CMS: 16.0 PB
LHCb: 8.5 PB

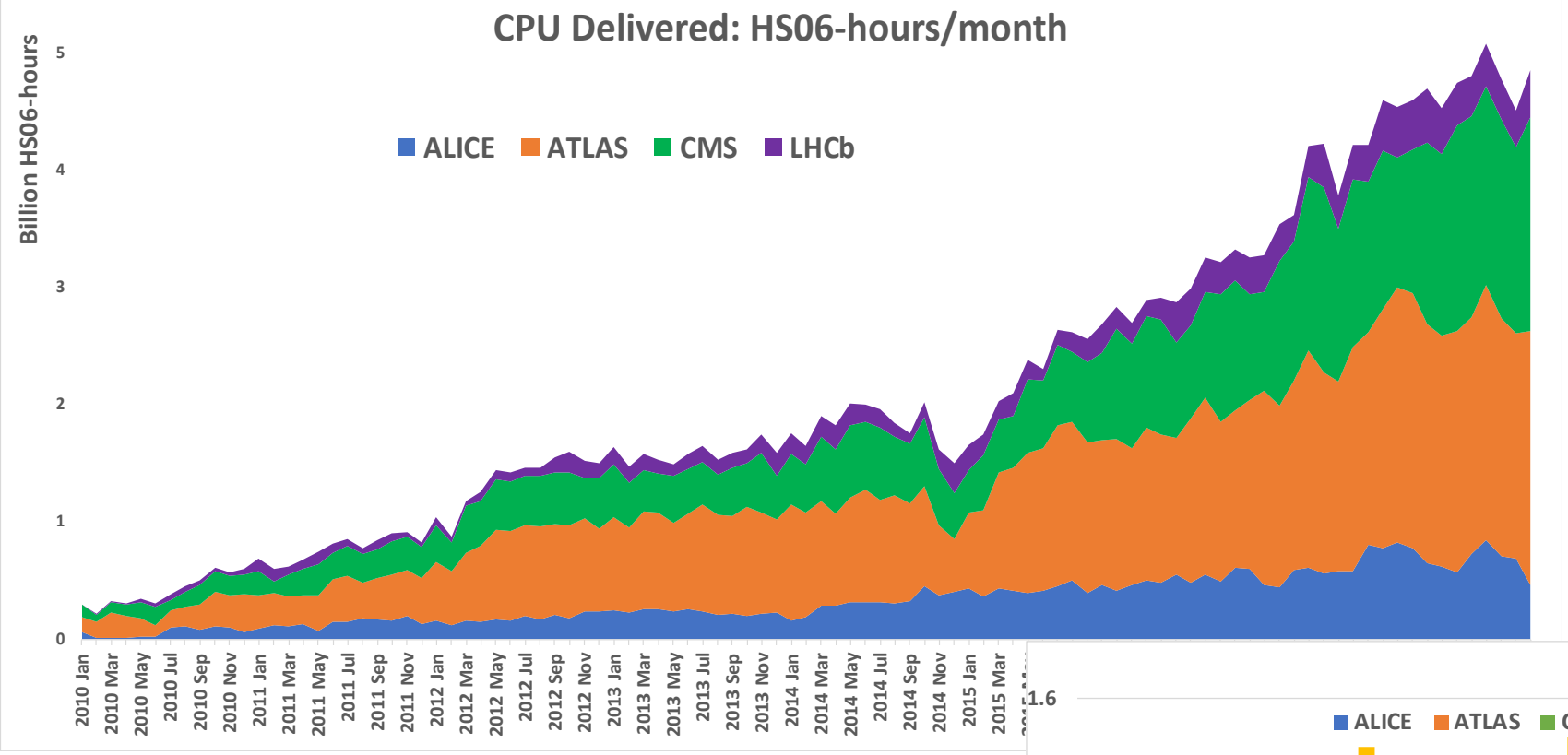


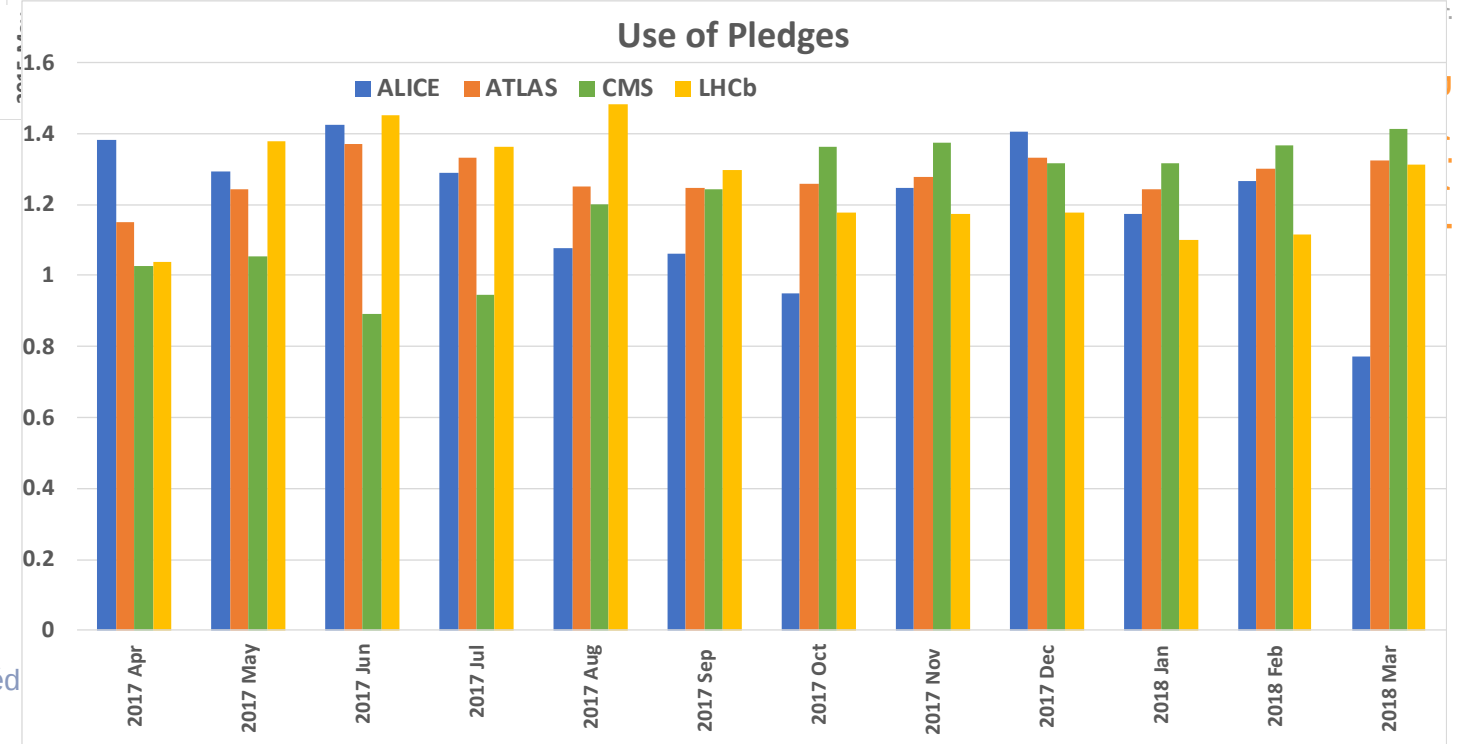


Year	LHC (PB)	Total (PB)
2016	48.3	69.8
2017	38.8	64.3
2018	***24.4!***	34.0

Document Classification: **Restricted**

CPU Delivered





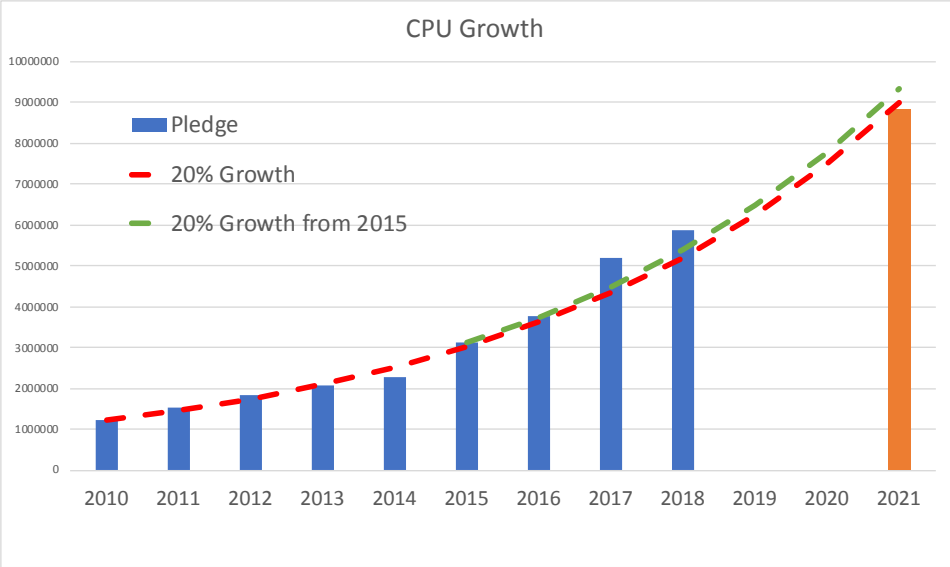
New peak: ~210 M HS06-days/month

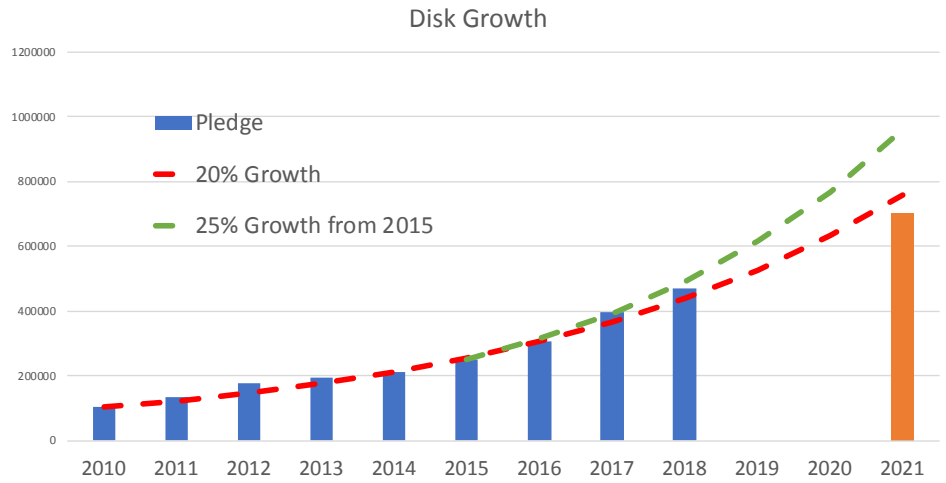
~ 685 k cores continuous

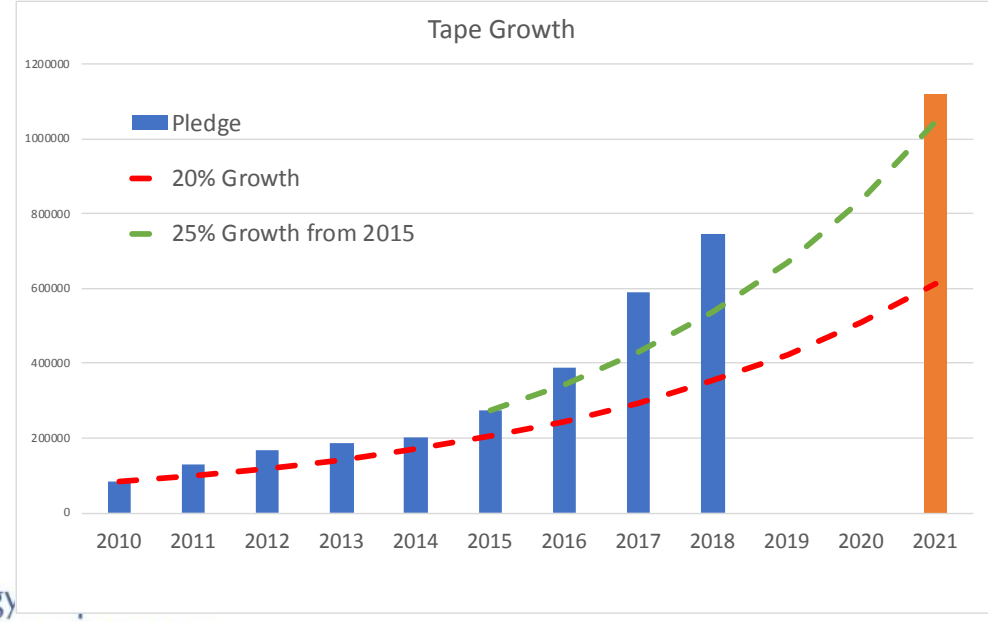
Run 3 Planning (2021-2023):

- Similar to 2018
- If the experiments luminosity level at a higher pile-up and for longer ➔
  - Potentially higher average pileup
  - Non-linear increase in CPU time
- Possibly less time between fills – more live time
- Overall the best estimate is 30% (50% conservatively) more resources needed than in 2018
  - But we have not seen 2018 yet
- For 2021: 1st year after LS2, could be only half-year live time but ramp up to optimal conditions rapidly
- Unknown:
  - Still need plans for experiment trigger rates
  - And plans for luminosity levelling

Resource evolution







- 2010-2018 – pledges
- 2021 assume 1.5 x 2018

However …

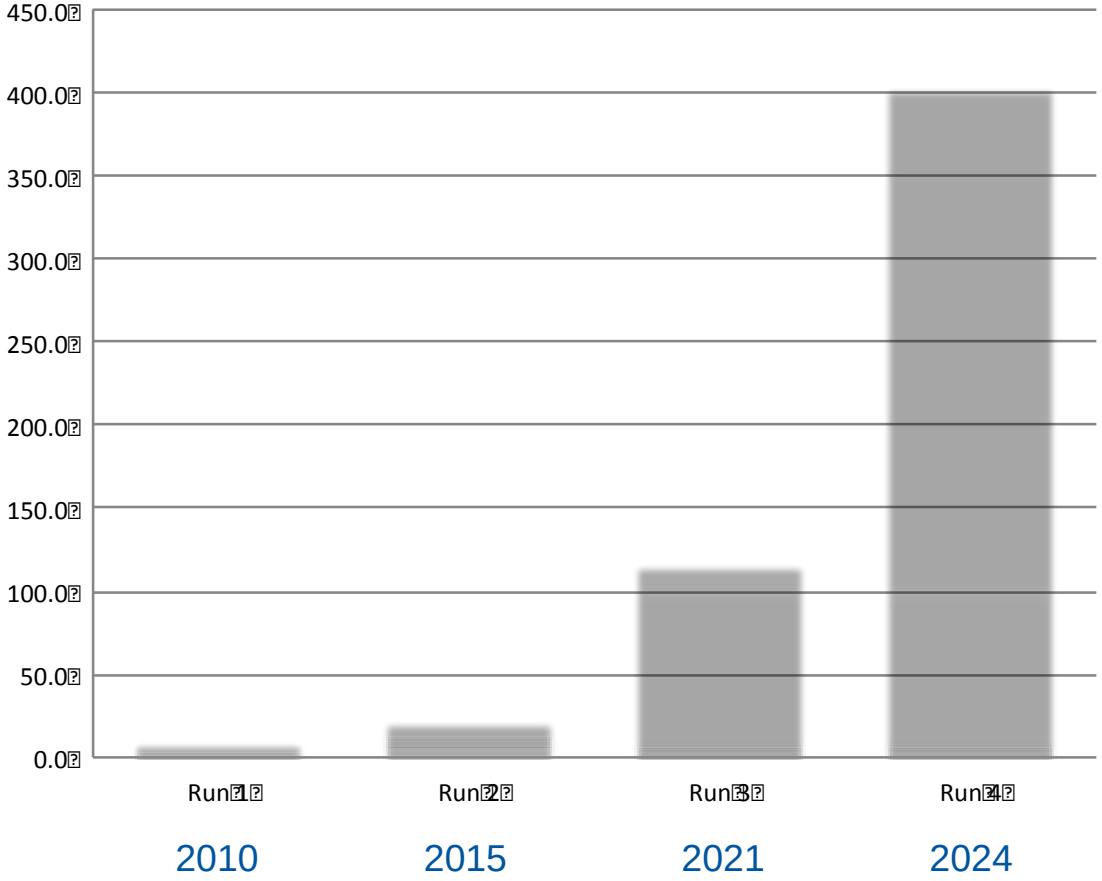
- ALICE and LHCb are upgrading during LS2, so the expectations of their needs do not follow the assumptions in the previous slides:
  - LHCb:
    - luminosity and pileup increase by factor 5.
    - Major changes in computing model result in higher trigger rate and HLT output bandwidth.
    - LHCC milestone for computing model in Q3/2018, together with engineering TDR – currently under review
  - ALICE:
    - Factor 100 increase in readout rate (50 kHz)
    - Data volume increase mitigated by online reconstruction and raw data compression in new O2 facility
    - O2 TDR is approved; summary needs are:
    - Increases in 2021 wrt 2018: CPU: 48%, disk: 74%, tape 90%

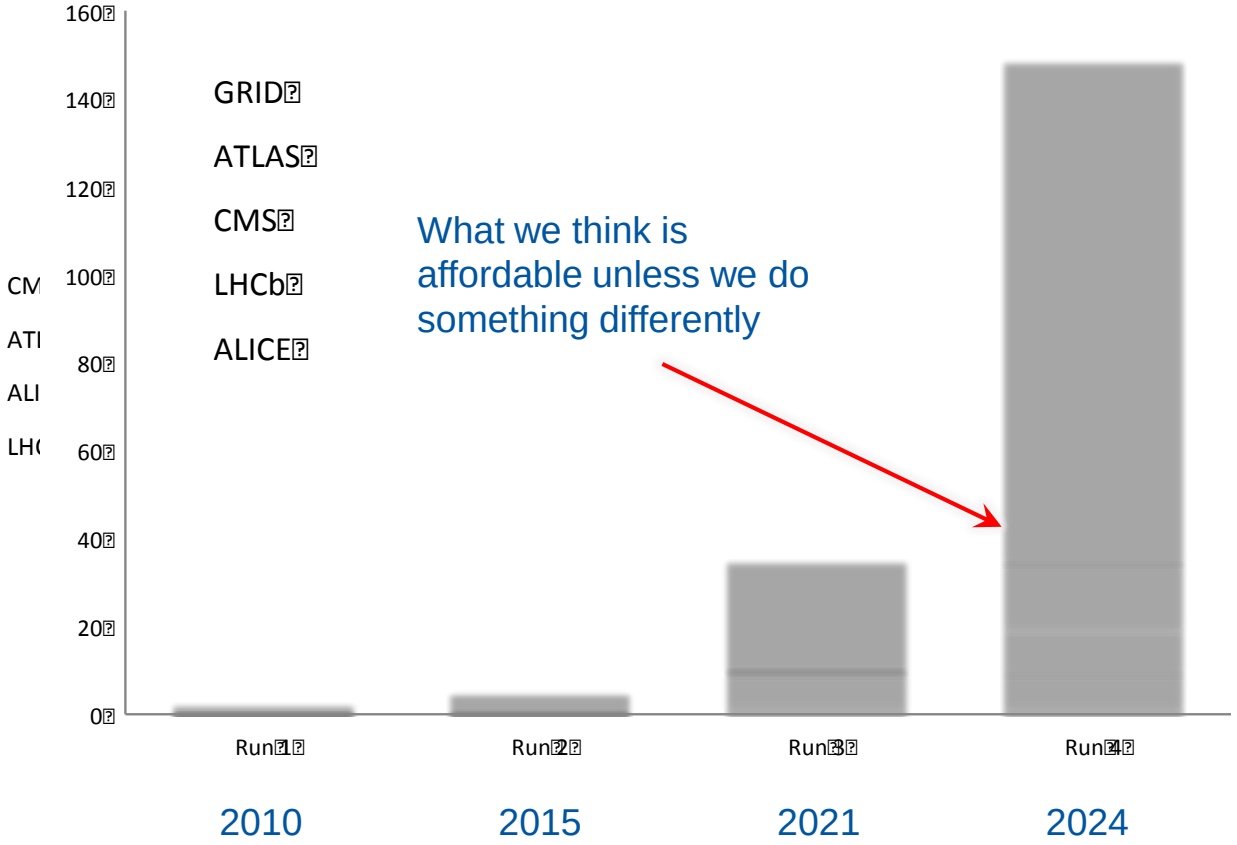
Scale of data tomorrow …

10 Year Horizon

2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 …… 2030?

Run 1 | LS1 | Run 2 | LS2 | Run 3 | LS3 | HL-LHC | FCC?





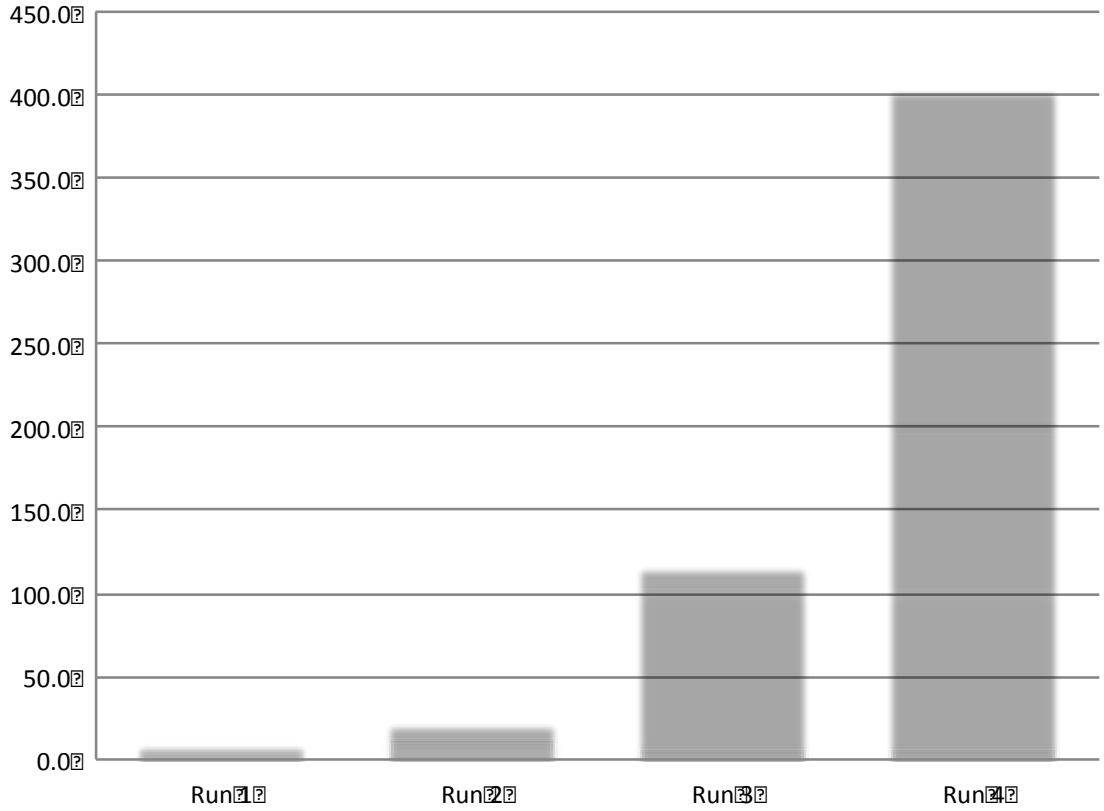
Data: ~25 PB/year → 400 PB/year

Compute: Growth > x50

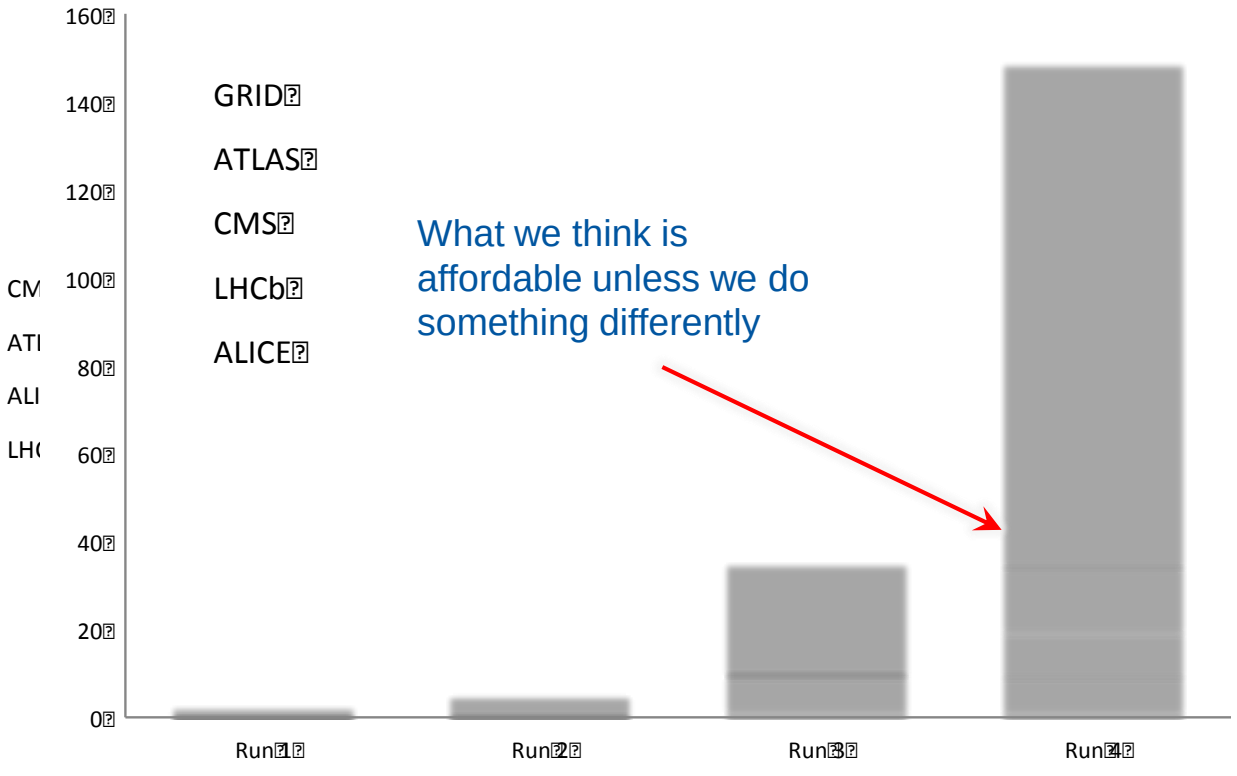
Technology revolutions are needed

Scale of data tomorrow …

10 Year Horizon



Data: ~25 PB/year → 400 PB/year

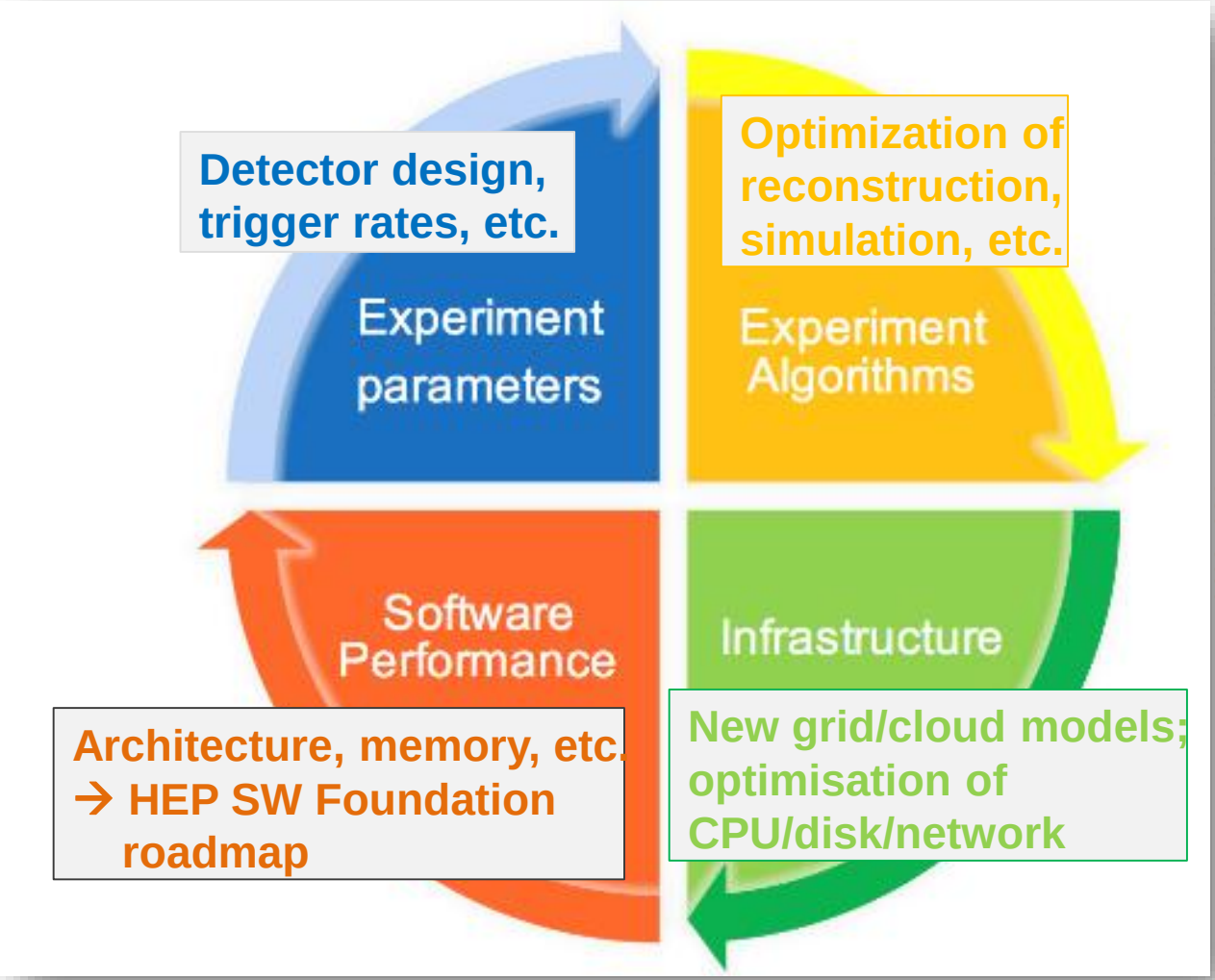


Compute: Growth > x50

Technology revolutions are needed

The HSF Community White Paper

- The HSF Community White Paper defines a roadmap for HEP software and computing R&Ds for the 2020s
  - For WLCG this is crucial in preparation for HL-LHC
  - Other research communities with computing needs at the level of WLCG will coexist on the same infrastructure
- The CWP consists of 13 work packages, each one defining a set of R&Ds
- The summary document is published:
  - https://arxiv.org/abs/1712.06982



Document Classification: Restricted

The WLCG Strategy Document

- The HL-LHC computing challenge: provide the computing capacity needed for the LHC physics program, managing the cost
- The WLCG strategy document is a specific view of the CWP, prioritizing R&Ds relevant to the HL-LHC computing challenge
- The prototyped solutions will be the foundation of the WLCG TDR for HL-LHC, planned for 2020. Timing to be re-considered?
- This is a presentation of the content of the strategy document
  - http://cern.ch/go/Tg79

WLCG Strategy - Outline

The strategy develops around five main themes …

1. Software performance
2. Algorithmic improvements / changes (e.g. generators, fast MC, reconstruction)
3. Reduction of data volumes
4. Managing operations cost
5. Optimizing hardware costs

It defines an R&D program with rough timelines, organized in sections:

- The HL-LHC challenge, hardware trends and a cost model
- Computing Models
- Experiments Software
- System Performance and Efficiency
- Data and Processing Infrastructures
- Sustainability
- Data Preservation and Reuse

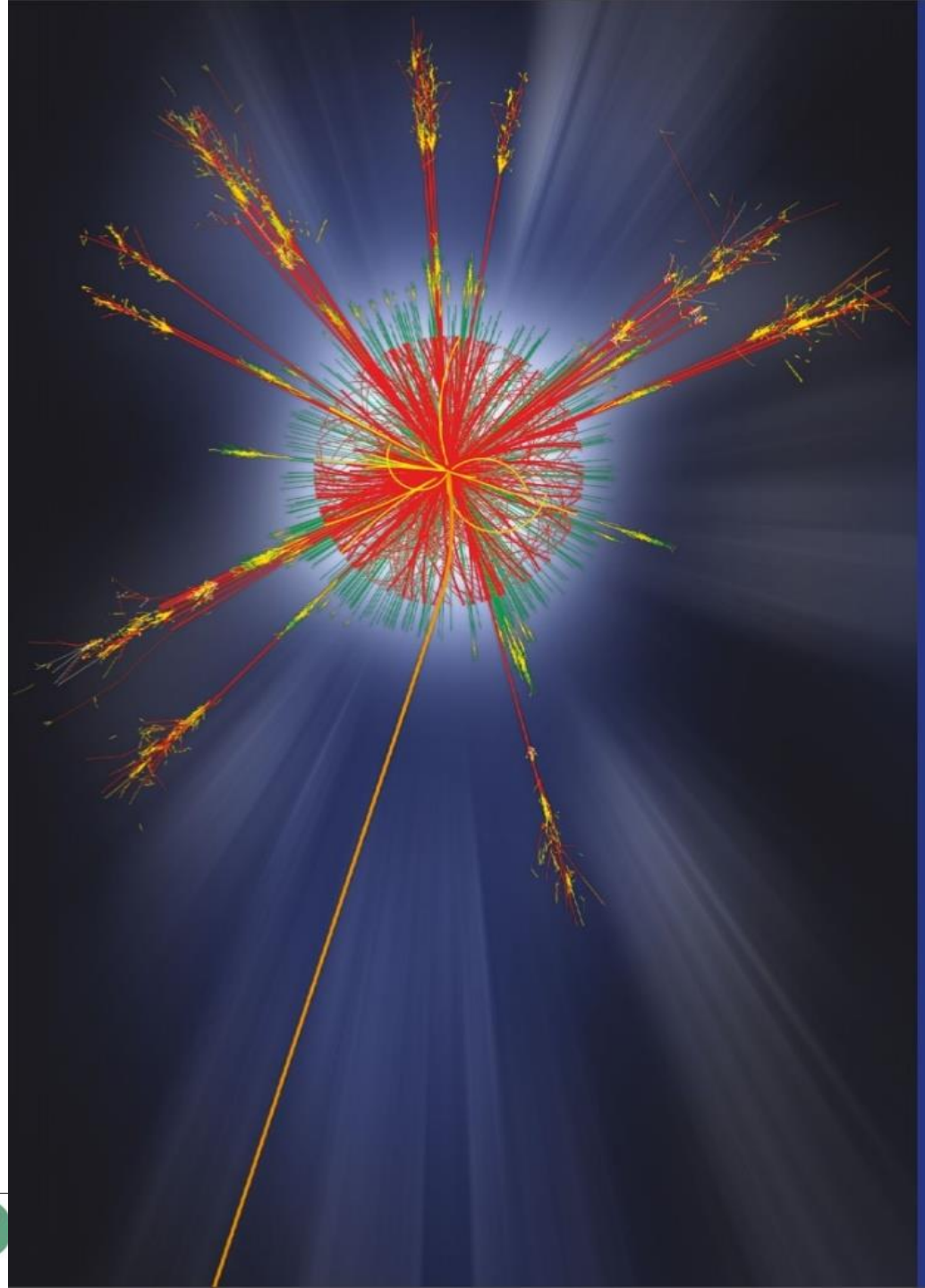
The goal is to demonstrate to the funding agencies that we are in control of the HL-LHC cost, while exploiting the full potential of the physics program

CERN openlab

WHITE PAPER

FUTURE ICT CHALLENGES IN SCIENTIFIC RESEARCH

September 2017



EXECUTIVE SUMMARY 04

INTRODUCTION 06

R&D TOPIC 1 10 DATA-CENTRE TECHNOLOGIES AND INFRASTRUCTURES

NETWORKING
- High-bandwidth links from detectors to the data centre
- Automation of network configuration and 'white-boxing'
- IoT for FRU tracking, data centre environmental monitoring...
- Integration of Wi-Fi and 5G: data security and protection

DATA-CENTRE ARCHITECTURES
- Rack disaggregation: rack-scale design
- Hierarchical storage buffers
- Software-defined infrastructure and tool-chain integration

DATA STORAGE
- Investigation of models for expansion of storage-capacity
- 'Cold storage' evolution

DATABASE TECHNOLOGIES
- Data size and rates
- Hardware evolution and consolidation
- Technologies for developer productivity
- Data-stream processing
- Time-series database workloads
- Scale-out databases and cloud resources

CLOUD INFRASTRUCTURES
- Orchestration and automation of compute provisioning
- Scalable clouds and global scientific clouds

R&D TOPIC 2 20 COMPUTING PERFORMANCE AND SOFTWARE

CODE MODERNISATION
- Storage-layer optimisations for low-latency NVRAM
- Performance-analysis tools for software
- Use of standard library facilities in C++
- Use of alternative concurrency models
- Verifying code and checking quality automatically

HETEROGENEOUS PLATFORMS AND ALTERNATIVE ARCHITECTURES
- Hybrids CPUs
- Optimising code distribution using lightweight containers

DEDICATED HARDWARE AND CO-PROCESSING SYSTEMS
- Optimising code performance using coprocessors and GPUs

R&D TOPIC 3 26 MACHINE LEARNING AND DATA ANALYTICS

DATA ACQUISITION
- Monitoring of accelerators and detectors
- Monitoring data quality
- Fast inference technology for "trigger" systems
- Anomaly detection and the search for new physics

DATA PROCESSING
- Simulation
- Jet identification and image-based event identification

BIG DATA
- Data reduction and refresh for analysis
- Optimisation of computing infrastructure

DATA ENGINEERING
- Solutions from industry, challenges and opportunities

R&D TOPIC 4 34 APPLICATIONS IN OTHER DISCIPLINES

PLATFORMS FOR OPEN COLLABORATION
- A smart data-analysis platform

LIFE SCIENCES AND MEDICAL APPLICATIONS
- Simulating biological systems in the cloud
- Large-scale analysis of genomic data
- Large-scale analysis of healthcare data

ASTROPHYSICS
- Exascale data processing at future astrophysics infrastructures

SMART EVERYTHING
- Environmental monitoring
- Traffic and mobility

ABOUT CERN OPENLAB 40

ADDITIONAL INFORMATION 42

CONTRIBUTORS 43

CONTACTS 44

Document Classification: **Restricted**

Concluding Words

- Training & Education is one of the 4 missions of CERN

- CSC Since 1970!
- 3 Different schools
  - Main, Thematic, Inverted

- Not a conference, but a summer university where experts deliver ***knowledge***
- We are again in Split, thanks to University of Split
  - Also organized the main school in 2007

- I wish you an excellent 6th tCSC in Split!

CERN School of Computing

Organised in collaboration with Tel Aviv University (TAU)

1-14 October 2018
Tel Aviv, Israel

Data storage & management technologies
Physics computing
Software engineering
Computer architecture
Computing security and networking
...and much more

Director
Sebastian Lopienski, CERN

Local Organising Committee
Erez Etzion, TAU
Yan Benhammou, TAU
Riki Cohen, TAU
Meny Raviv-Moshe, TAU

Administrative Manager
Joelma Tolomeo, CERN

Technical Manager
Nikos Kasioumis, CERN

Lecturers
François Flückiger, *CERN*
Bob Jacobsen, *UC Berkeley, USA*
Thomas Keck, *KIT, Germany*
Sebastian Lopienski, *CERN*
Andrzej Nowak, *TIK Services, Switzerland*
Alberto Pace, *CERN*
Andreas J. Peters, *CERN*
Danilo Piparo, *CERN*
Ivica Puljak, *University of Split, Croatia*
Arnulf Quadt, *Universität Göttingen, Germany*
Enric Tejedor, *CERN*
Eamonn Maguire, *Pictet AM, Switzerland*

Deadline for Application: 13 May 2018
http://cern.ch/csc

Thank you for your attention

"The task of the mind is to produce future"

Paul Valéry