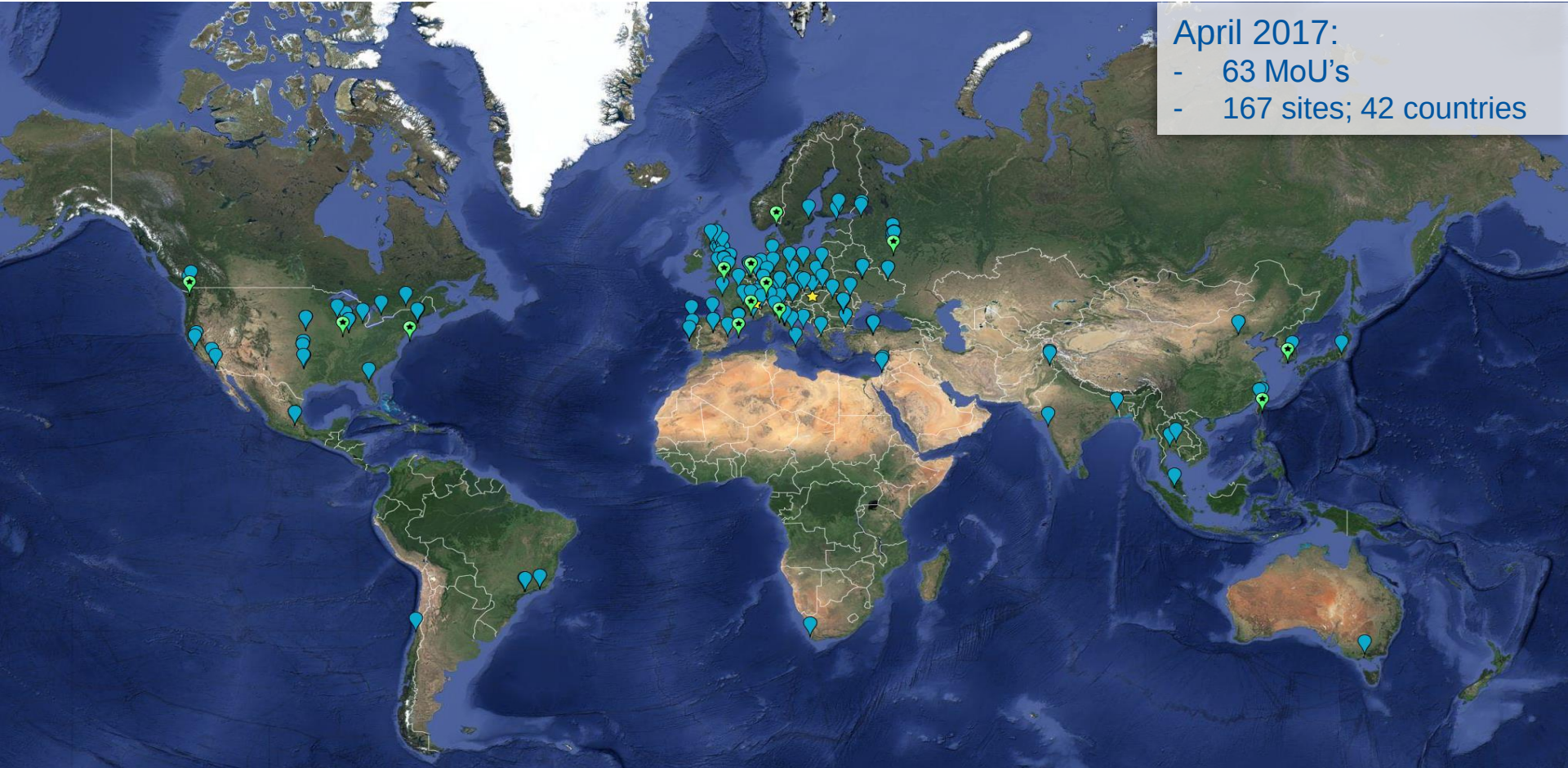
WLCG Workshop – Introduction

Ian Collier, on behalf of Ian Bird

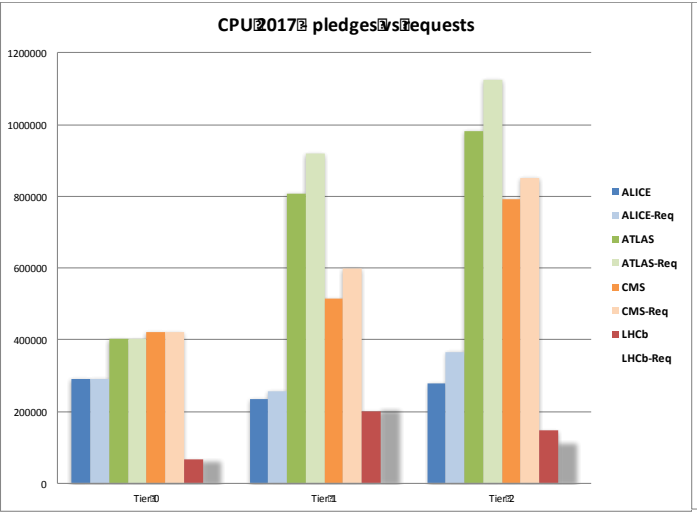
WLCG Workshop

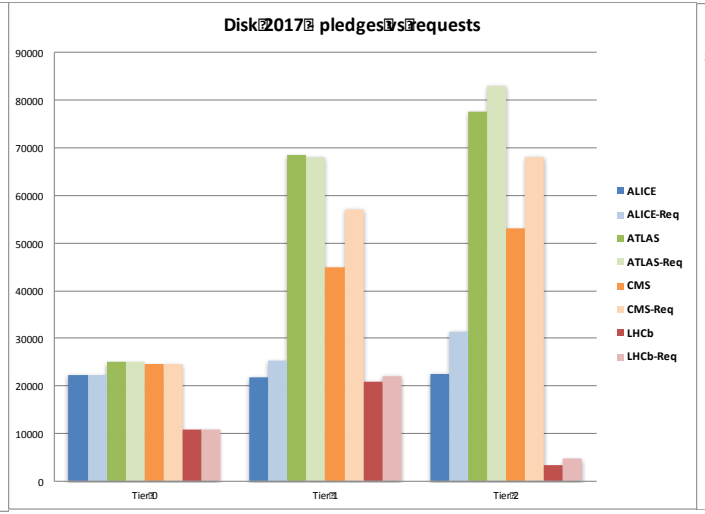
Manchester, 19th June 2017

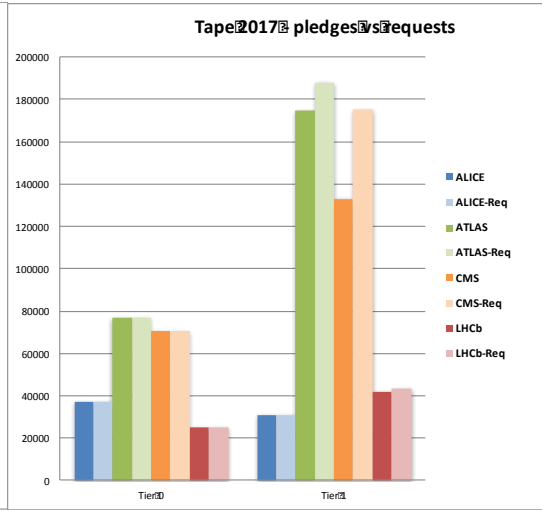
WLCG Collaboration



2017 Pledge situation



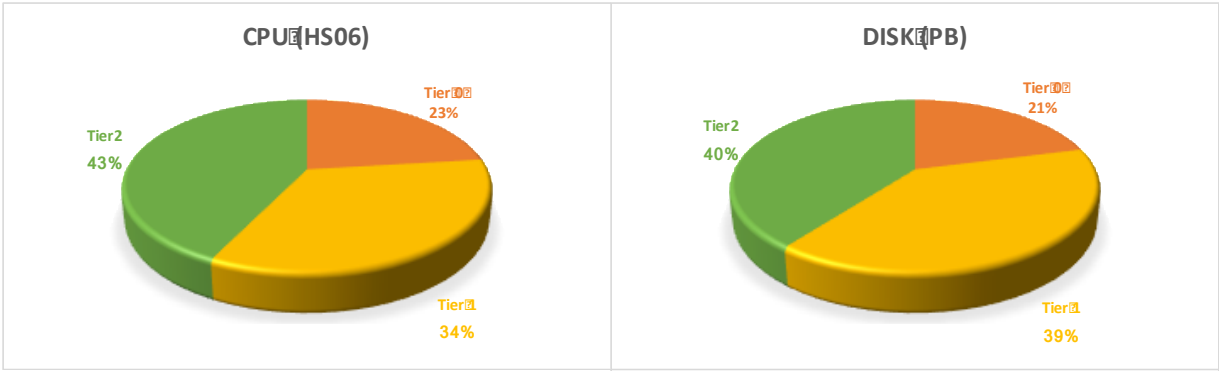




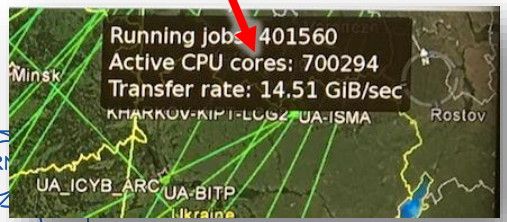
Not all is deployed yet for 2017 – a few delays
Full resources expected by end of this month

[Light colours: Request; darker colours: pledge]

Pledged resources 2017



- 5.2 M HS06
  - >500 K cores (if bought today)
  - Actually many more



- 985 PB Storage
  - 395 PB disk
  - 590 PB tape

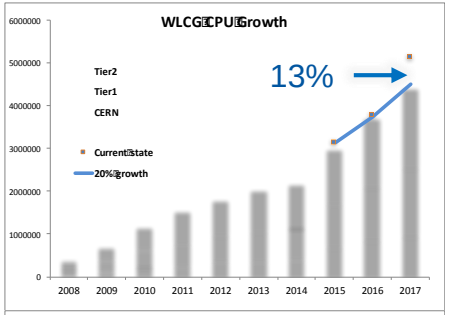
LHC Performance

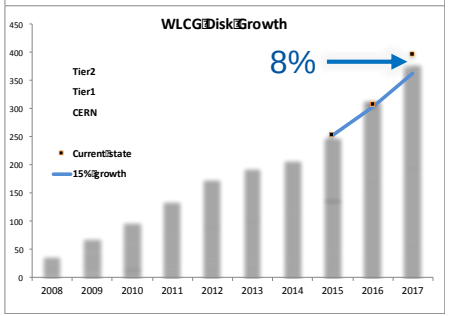
- In 2016, the LHC availability (live time) was much greater than anticipated, leading to some 40% more data generated than planned
- This had implications for resource needs in 2016, and in 2017 assuming equally high availability and the increased luminosity
- At the October RRB, (some) funding agencies agreed that they would help on a “best effort” basis with more resources, but pledges would not be increased
  - LHCC proposed to review the mitigation measures the experiments and WLCG had taken to minimise the additional requests
  - Really mandated that we remain within a flat budget “no matter what” (my phrasing)

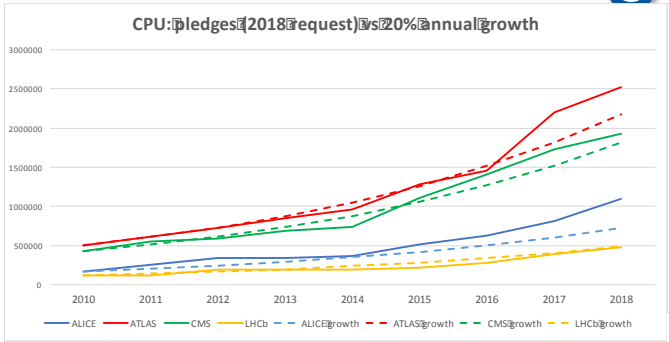
Mitigation measures reviewed by LHCC

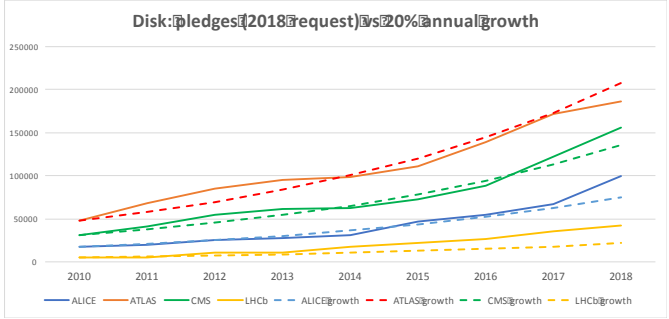
- In February the LHCC reviewed the measures taken by the experiments to mitigate the shortfall in resources relative to the exceptional LHC performance
- Concluded that: (CERN-LHCC-2017-004)
  - "The **LHCC congratulates** the LCG and experiments on the successful implementation of mitigation measures to cope with the increased data load. "
  - *"The **LHCC notes** that the margins to reduce the resource usage in the short term without impact on physics have been exhausted. "*

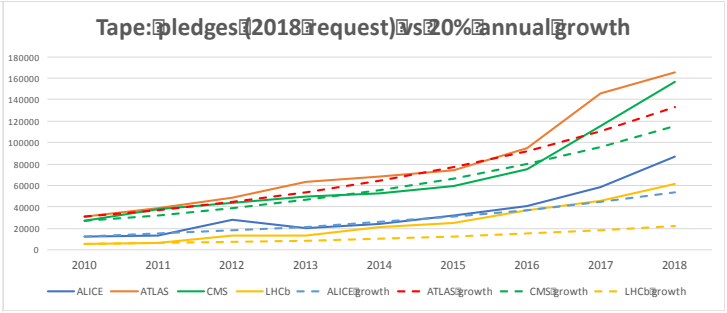
Comments on flat budgets









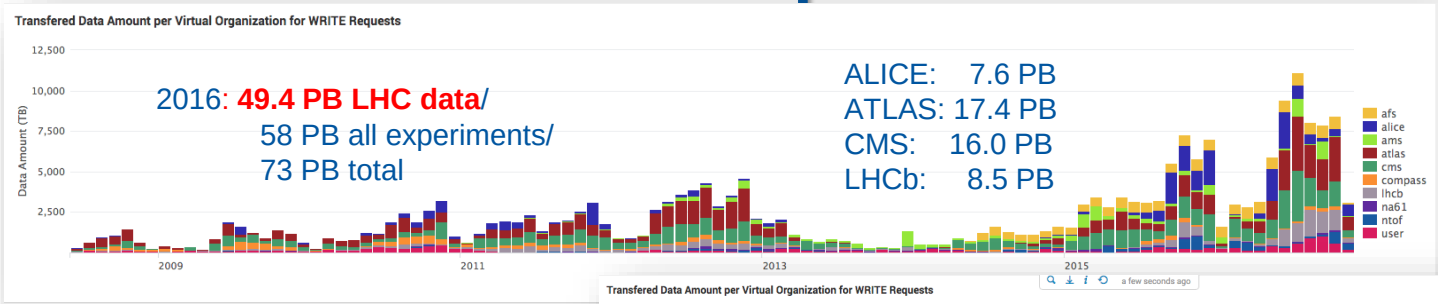


Extrapolations from 2010:

- Ignore no investment in 2013,14
- Deviations from “flat budget” are generally not enormous, and are corrected
- Jump in 2017 – LHC performance
- Tape needs still increase

- We need to clarify what is meant by flat budgets:
  - We assume: constant budget/investment even in long shutdown years
  - This did not happen in LS1

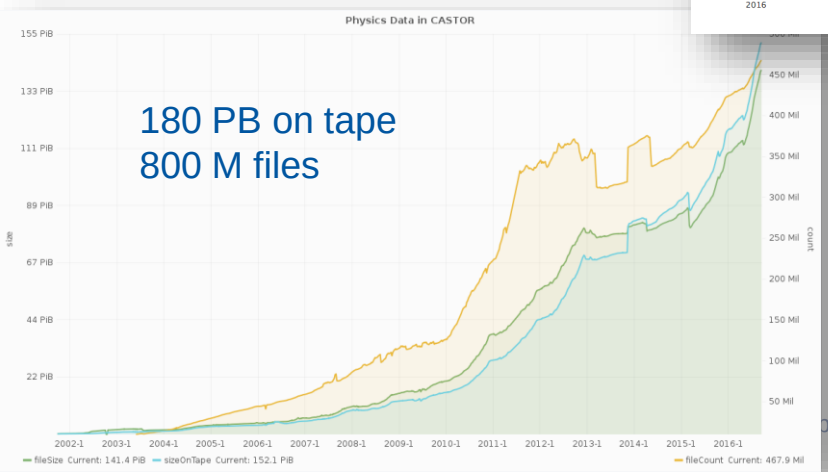
Data in 2016 - updated



2016: **49.4 PB LHC data**/
58 PB all experiments/
73 PB total

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

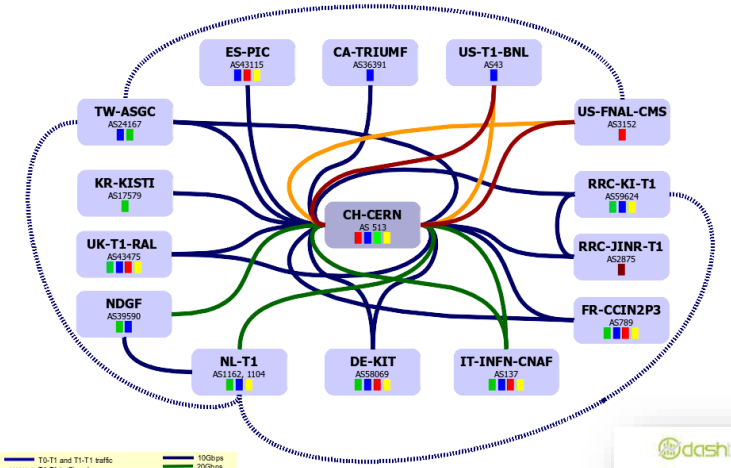
11 PB in July

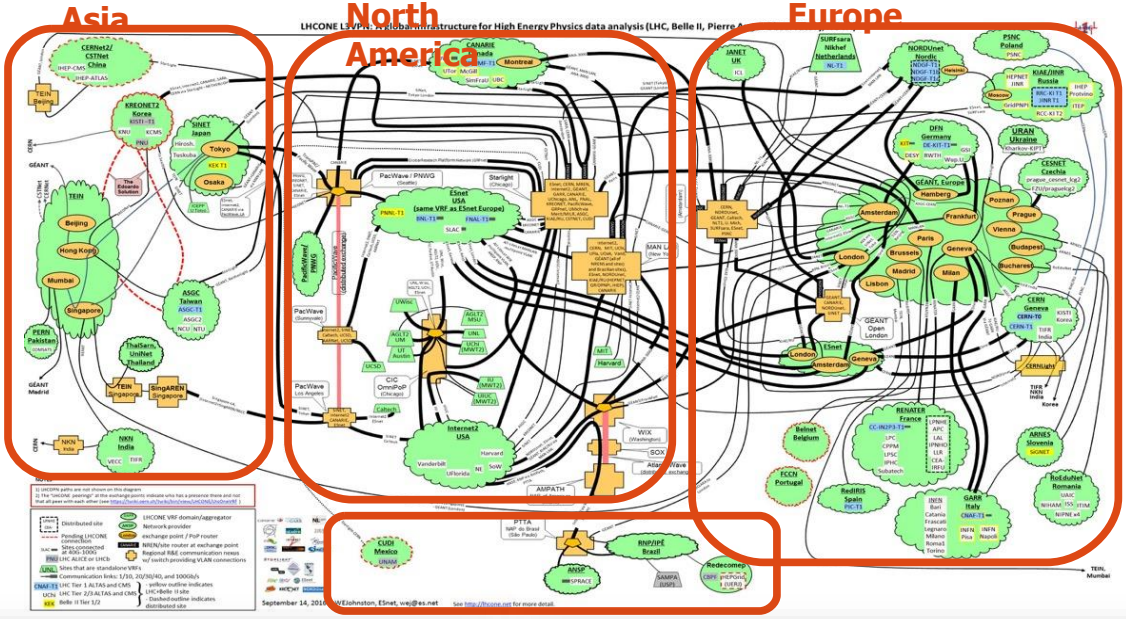


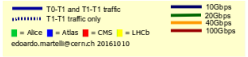
180 PB on tape
800 M files

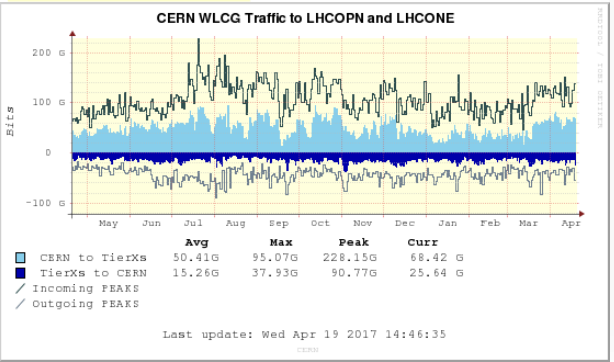
Data transfers

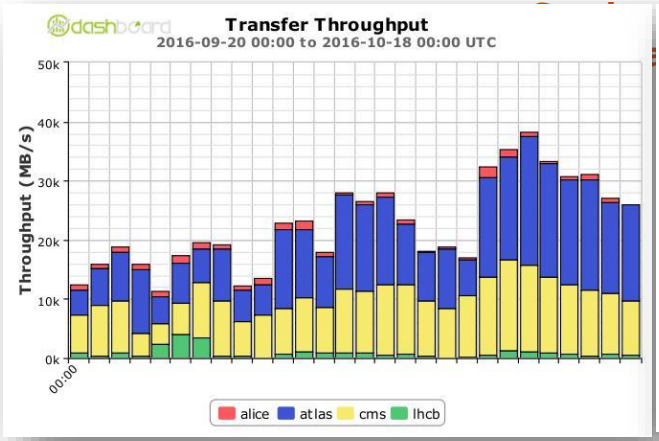
LHCOPN

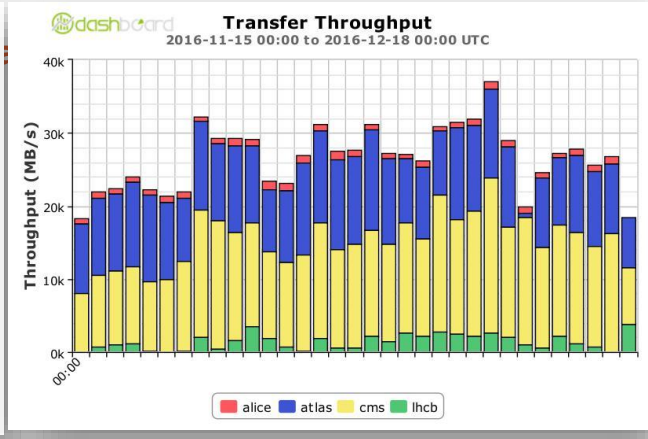




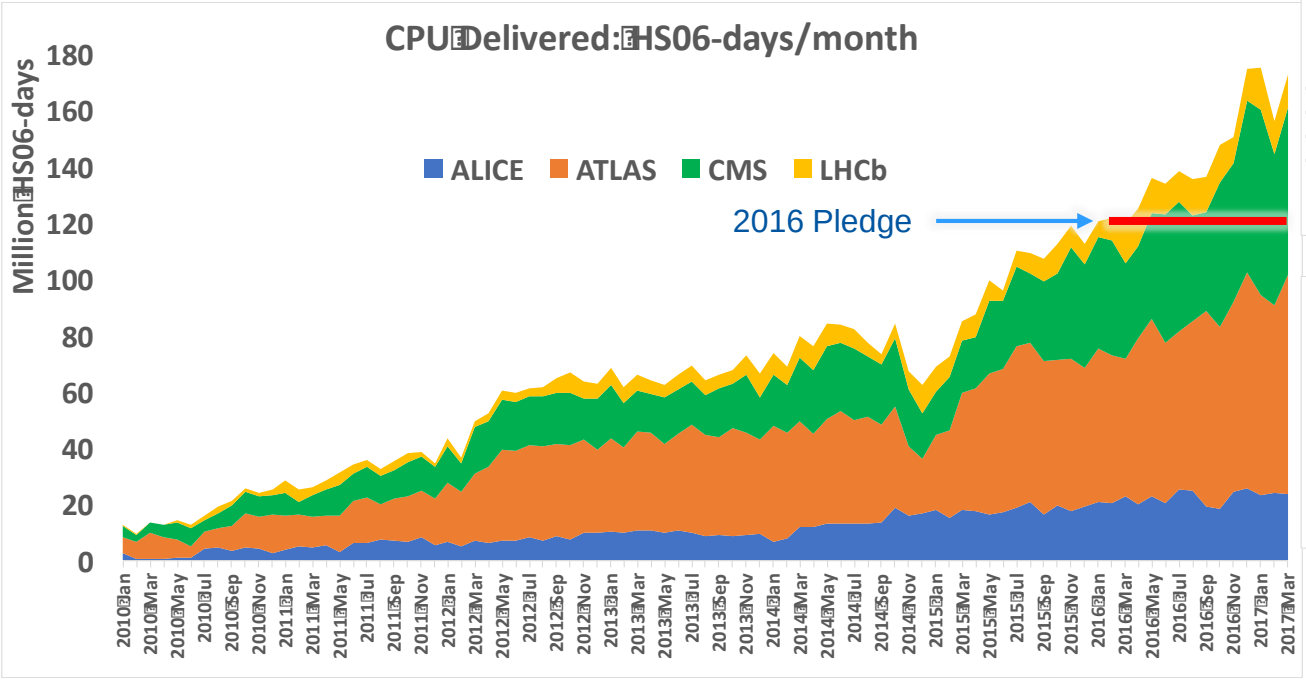


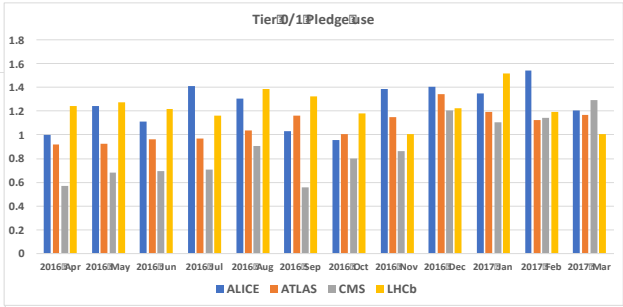


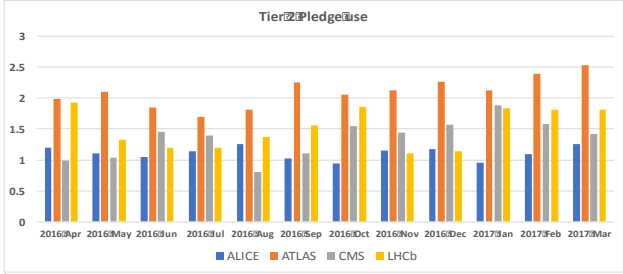




CPU Delivered







New peak: ~180 M HS06-days/month

~ 600 k cores continuous

Resource pledging process

NB. This is modified (by RRB) wrt the MoU ideas

- In year n:
  - C-RSG review in Spring to confirm requests for year n+1
    - Needed as procurements at this scale take ~1 year
  - C-RSG review in Autumn – 1st look at requests for year n+2
    - Often also ”adjustments” requested for year n+1
      - But this is too late to affect (most) procurements
    - Also FA’s confirm pledges for year n+1
- Initially had a 3-5 year outlook, but this is impractical:
  - Requests difficult to foresee that far ahead (LHC conditions, schedule, etc. – usually not confirmed until Chamonix of the running year)
  - Budgets mostly not known on that timescale: FA’s do not discuss budget outlook
- For Run 2; in 2013 we made an outlook for 2015, 2016, 2017

Community White Paper

- Mentioned at previous RRB
- Goal to have a Community White Paper (CWP) on overall strategy & roadmap for software/computing for HL-LHC
  - Deliverable of an NSF-funded pre-project
  - Also takes account of Belle-II, ILC, neutrinos, etc.
- To be delivered by summer 2017
- Kick-off workshop held in San Diego 23-26 Jan
- Final workshop next week in Annecy
- Will be used as input for the LHCC report later this year, developing roadmap towards TDR for HL-LHC computing in 2020

HL-LHC Computing TDR

- Agreed with LHCC to produce TDR for HL-LHC computing in 2020
- In 2017 we will provide a document describing the roadmap to the TDR (strategy document)
  - Using the CWP as input
  - Describing potential new computing models
  - Defining prototyping and R&D work that will be needed
- The TDR will not be the end – technology evolution in 6-7 years will be significant, cannot afford not to follow it
- NB. Very different situation from the original TDR –
  - we have a working and well-understood system that must continue to operate and evolve into the HL-LHC computing programme

Strategy document in 2017

- Describe the HL-LHC computing challenge given what we currently understand
  - Running conditions, trigger rates, event complexity, based on reasonable extrapolations of today's computing models
  - This will be a snapshot of a (yearly?) update of these numbers
- Describe the potential computing models and how they could change the cost and/or physics output
  - Necessarily at a high level
- Cost models
  - Appropriate metrics, balance/trade-off between CPU, storage, network etc
- State-of-the-art understanding of evolution of technology
  - 2-3 years is already difficult to predict; 10 years is impossible (even for the technology companies)
- Set out what we see as R&D areas, and potential prototyping activities or demonstrators:
  - Goals, metrics, resources, plans
- The HSF CWP will provide the basis of this

Technical topics

- Computing models
  - Different scenarios
  - Use of in-house, commercial, dedicated architectures, HPC, opportunistic, etc. resources
- Technology “choices” – may not be a choice but market-driven
- Data management and data access layer
  - End-to-end performance considerations; models of data delivery, event streaming, etc.
- Networking
- Resource provisioning layer
- Workload management layer
- Analysis facilities – how will analysis be done – traditional vs ”query” vs ML, …
- These above lead to ideas about facilities and how they may look
- The stated (and agreed) intention in the CWP discussion is to make these components as common and non-experiment specific as possible
  - Clarify what really needs to be specific
- The CWP will provide the details of progress and R&D roadmaps in many key areas

What a 2020 TDR may contain

- Broad expectations of costs of computing – based on expected evolution of the models
  - But 2020 is still 6 years before Run 4 – a lot will change and we must not be too prescriptive
  - Rather have to show evolution goes towards maintaining a constant cost (or not!)
- Updated requirements for 1st years of HL-LHC
  - To be regularly updated
- Updated technology expectations
  - Snapshot as understood in 2020
- Firmer ideas of computing models based on the prototyping work
  - Roles of online, Tier 0, other facilities
  - Bulk data management, processing, analysis models, simulation
  - Roadmaps for R&D that is still required
- Data preservation – how to use Run 1,2,3 data
- A lot of details will not affect the cost significantly, and are part of the operating and evolving service
- Updates of key CWP strategic areas

Scientific Computing Forum

- Initiative of the CERN Directorate
  - At the request of the Council to have more “informal” interaction on strategic topics
- Have held 2 meetings (https://indico.cern.ch/category/9249)
  - February and May 2017
  - Membership not yet settled – not only member states
- Discussions
  - First meeting – strategy paper, reflecting at high level some of the ideas for long term computing evolution (for WLCG)
  - Second meeting – Relationship of CERN and WLCG with SKA; input from several countries on their scientific computing strategies

(Aside) Globus

- NSF has announced end of support for open source Globus toolkit, from end 2017
  - I have been in touch with NSF to ask about support for LHC – they recognize the problem
    - No feedback yet
  - What will OSG and EGI do?
- Fall-back – WLCG takes relevant packages and maintains them
  - gsi, gridftp, myproxy
  - And perhaps eventually replaces them

Conclusions

- Run 2 in 2016 delivered 50 PB of new data, following exceptional performance of the LHC
  - Continued to set new performance records in all areas
- WLCG infrastructure continued to be even more active in the EYETS
- 2017/18 look to be challenging in terms of resource availability, esp if LHC meets expected luminosities, availability
- Activity (& engagement) is ramping up to look at evolution of the computing models for the future