CERN

HEP Computing Outlook
(for the next 10 years)

Tony Cass (but extensively reusing material prepared by Ian Bird, WLCG Project Leader)

2nd Asia Tier Center Forum

1st December 2016

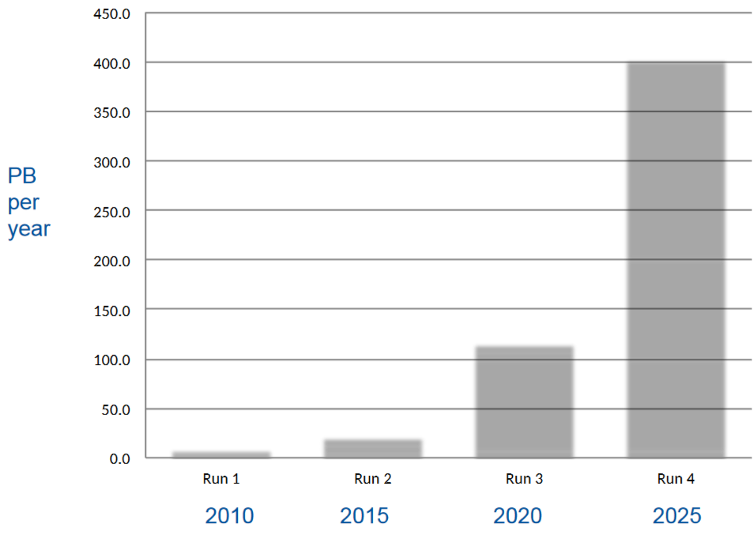
Money

Data

Big Science

Networks have to follow LHC data growth

Slide courtesy of Tony Wildish, NERSC/JGI



Expecting to record 400PB/year in Run4

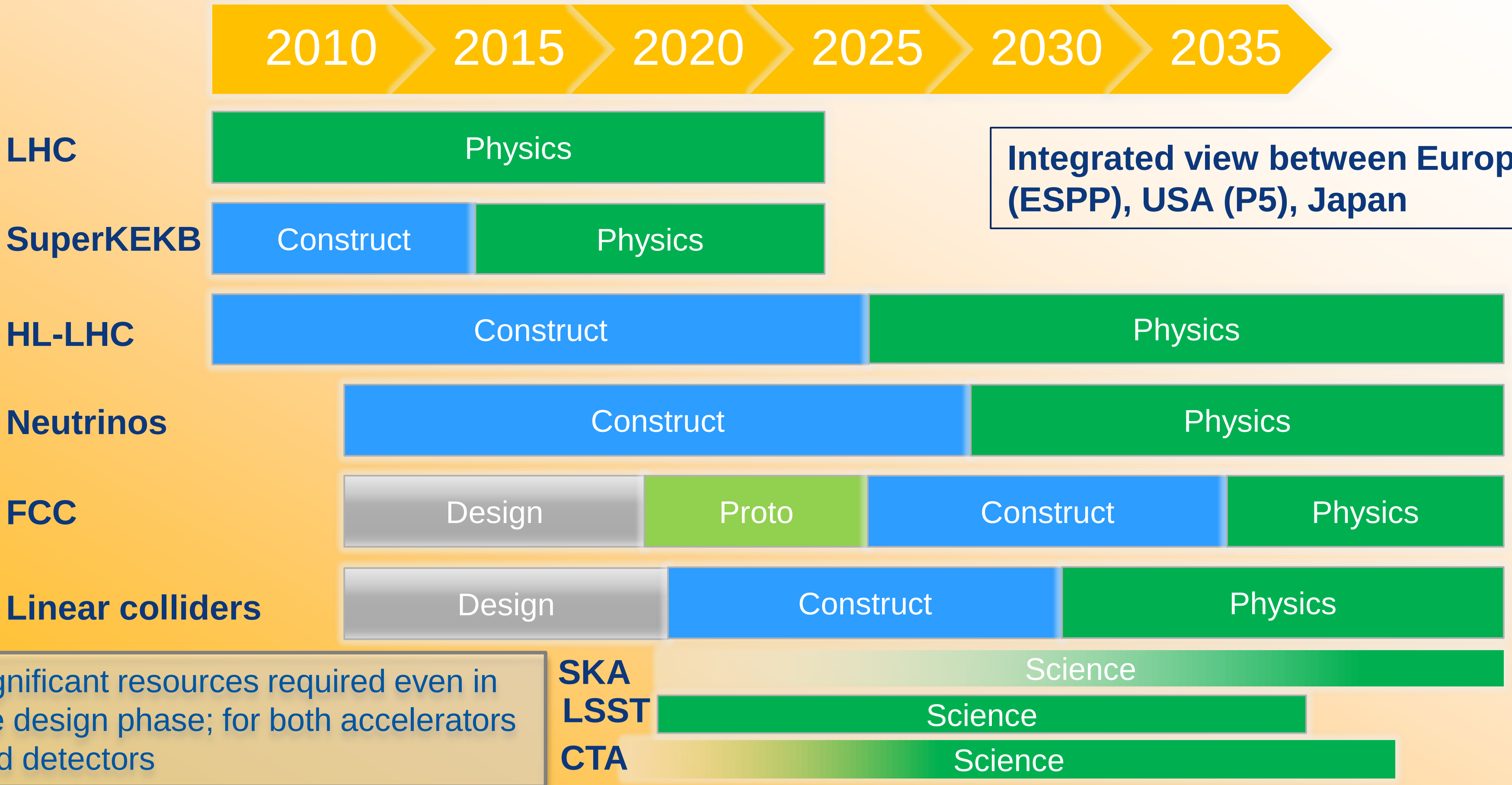
CMS
ATLAS
ALICE
LHCb

Computing needs expected to be around 50x current levels, if budget available

Networks must grow accordingly

CERN | Information Technology Department

HEP Facility timescale



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

Significant resources required even in the design phase; for both accelerators and detectors

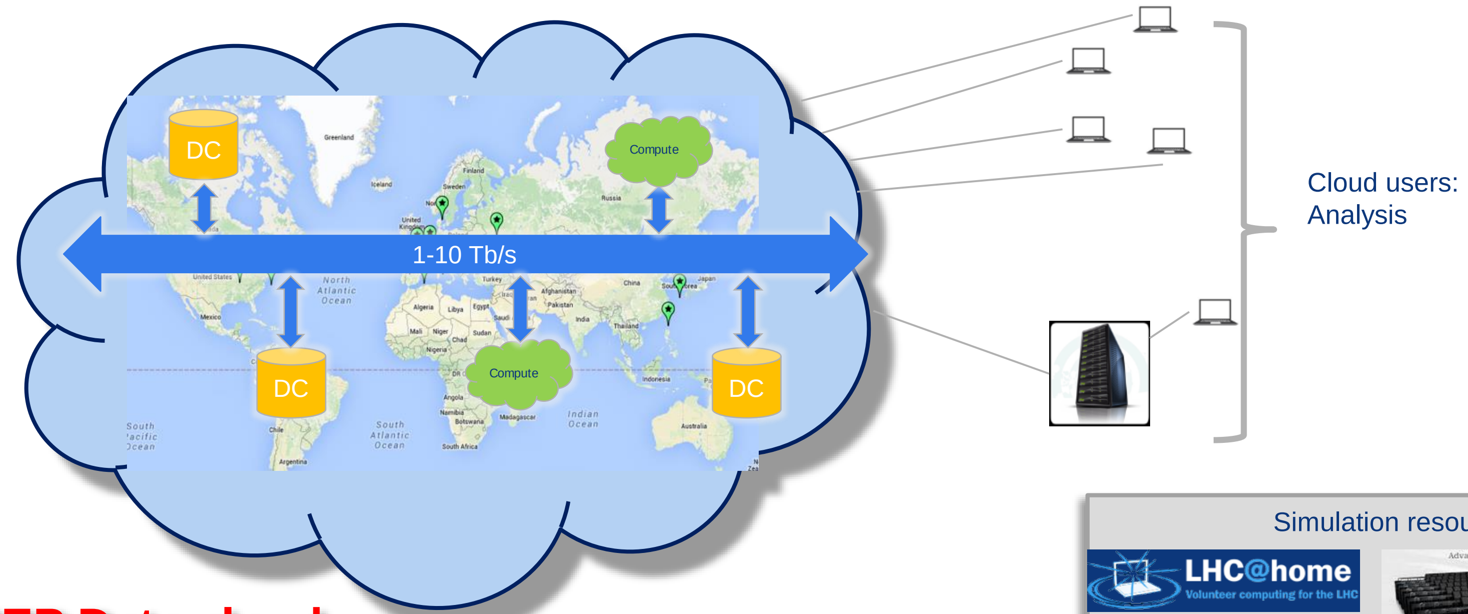
HEP Computing over 10 years

- HL-LHC is the major challenge
  - But also: neutrinos, Belle-II, development for linear collider experiments, FCC, etc.
    - And HENP facilities will also host Astro/Astroparticle experiments
- WLCG is a global collaboration, and successful in supporting LHC computing
  - Many lessons learned:
    - Technology, sociology, funding
- Can we imagine a HEP-wide computing environment?
  - Building on what we have, evolving and providing tools and infrastructure services to be used or adapted to future needs of HENP

Key Components

- A general infrastructure and services (data stores, compute facilities, networking, etc.);
  - Associated services like AAA, security, base monitoring, operational support
  - Needs to be capable of supporting different computing models and agile to technology changes
- "Software": full stack from workflow and data management tools to application level; Common tools, libraries, etc.;
  - BUT: a set of optional tools, contributed, developed, maintained, by the community;
  - Common R&D and support tools (technology tracking, software tooling);
  - This is essentially what the HSF is mandated to do.
- A Steering group structure to organize and evolve the above – mandated by e.g. ICFA
  - Constituted from major global HENP facilities, experiments, observers (A-P)
- Experiment/facility-specific (optional) processes
  - e.g. WLCG for the LHC resource management, etc.

Possible Model for future HEP computing infrastructure



HEP Data cloud

Storage and compute

Infrastructure – 2

- Networks
  - LHCOne – global HENP (and A-P) overlay network (LHCOPN is LHC-specific),
  - Address the need for access to commercial clouds via NREN's
  - How to manage a HEP data cloud network (multi-Tb interconnect, SDN)
- AAA services
  - Federated identities, authorization mechanisms, accounting
  - Policies
- Operational services
  - Support, incident response, monitoring, etc.

A “data cloud” for HEP

- Few – O(5-10) - large centres
  - Multi-Tb private (SDN) network between them
  - Treat as a single “virtual data centre”
    - Policy replicates data inside for security and performance
    - Think of RAID across data centres
  - Store all of the “AOD” data here; Do not replicate data to global physics institutes (major cost)
- Pluggable compute capacity:
  - HEP resources at these centres & other large centres
  - Commercial compute
- Model allows commercial data centres
  - For storage – enough redundancy that a commercial centre could unplug
  - For compute
  - Relies on networking and GEANT/Esnet etc. connections to commercial entities, policy
- Users access data in this cloud remotely
  - Eventually download “ntuples” – or equivalent
  - All organised processing is done in this model
- Enables new analysis models: all data can be seen as colocated
  - Get away from the “event-loop” → queries, machine-learning, etc.

This idea has been discussed in the WLCG community (e.g. see I. Fisk CHEP plenary)

- Hybrid model:
  - HEP-resources at a level we guarantee to fill → cost-effective
  - Commercial resources for “elasticity”
- Needs new funding models

Beyond HEP

- This virtual data cloud model may be very interesting for other sciences
    - E.g. SKA Regional Centres
    - Works also for DUNE, Future facility development, others
  - Can provide resiliency and long term preservation capabilities
- Integrating commercial resources
  - Requires (potentially) significant changes in funding models
  - Can we actually procure commercial resources at large-enough scale to get economy?
    - HNSciCloud as a proof-of-principle of joint procurement
    - Can we purchase from the largest cloud vendors? Politics?
    - Real cost-efficiency and elasticity requires a “spot-market” price
  - How do we arrange performant and secure network connections to commercial resources?

CERN WLCG

Summary

- HEP has dominated global "Big Science" for (at least) the past 10 years—with significant funding linked to the start of the LHC.
- The world will look very different in 2026.
- HEP needs to be able to adapt to an environment where we may not be the dominant force
  - Exploit commercial and opportunistic resources, not expect to have our own
  - Act as a peer in a global Big Science collaboration
- Whatever happens, network connectivity and infrastructure will be key to ensure access to resources whether they are national, regional or global.