ATLAS Italia Computing

Alessandro De Salvo

20-11-2019

Organizzazione italiana di ATLAS

- **Coordinatori del Computing in Italia**
  - **Nazionali**
    - Coordinatore del Calcolo di ATLAS Italia: responsabilità politico/amministrativa
    - Coordinatore della Federazione Italiana dei Tier2: responsabilità tecnica
  - **Locali**
    - Coordinatori dei centri locali di calcolo e/o di attività nelle sedi preposte

- **Sezioni che partecipano (più o meno) attivamente al Computing in Italia o in contatto con il management italiano del calcolo**
  - Tier1 (CNAF)
  - Tier2 (LNF, MI, NA, RM1)
  - Tier3 “R&D and operations”
    - GE (Tier3 + attività varie, vedi slide dopo)
  - Tier3 “resource only”
    - LE
    - RM3
  - Tier3 “disk only”
    - TS
  - Tier3 “outsiders”
    - Grecia, Sudafrica

Organizzazione italiana di ATLAS

- **Compiti dei coordinatori del calcolo**
  - **Coordinatore del Calcolo di ATLAS Italia:**
    - **Essere il punto di riferimento della comunità nazionale per quanto riguarda le attività di calcolo. In tale ruolo deve:**
      - **Avere adeguata conoscenza dei mezzi di calcolo a disposizione della comunità Atlas Italia nei centri Tier1 e 2 e localmente nelle sedi dei gruppi.**
      - **In stretto contatto con il Responsabile Nazionale ed in collaborazione con il Coordinatore delle Analisi deve operare per armonizzare, a vantaggio di tutti, l'utilizzo delle risorse nazionali di calcolo (Tier1; Tier2), favorirne l'uso efficiente, segnalare situazioni di criticità e suggerire soluzioni per il loro superamento.**
      - **Su richiesta del Responsabile Nazionale, deve presentare periodicamente nelle riunioni di Atlas Italia un quadro complessivo dello stato del Calcolo.**
      - **Deve avere un ruolo di riferimento verso la CSN1 per quanto riguarda le attività di Calcolo.**
    - **Per svolgere queste funzioni il Coordinatore del Calcolo dovrà tenersi aggiornato sullo stato del Software & Computing in Atlas, sulle iniziative e sulle attività ad esso relative e dovrà presenziare il più possibile alle pertinenti riunioni di Atlas.**
    - **Il Coordinatore del Calcolo è il rappresentante italiano nell'International Computing Board di Atlas.**

  - **Coordinatore della Federazione Italiana dei Tier2**
    - **Proposto dal Coordinatore e dalla comunità italiana ed eletto tramite endorsement dei 2/3 degli aventi diritto al voto (con le astensioni contate come voti contrari) per la durata del mandato del Coordinatore del Calcolo.**
    - **I compiti principali del Coordinatore della Federazione Italiana dei Tier2 sono:**
      - **l'organizzazione e la supervisione delle attività tecniche inerenti la Federazione Italiana del Calcolo di ATLAS**
      - **l'interfaccia tecnico-operativa nei confronti dell'ATLAS Distributed Computing al CERN**

Partecipazione italiana alle attività di ATLAS

- **ATLAS Italia partecipa alle attività di Atlas Distributed Computing (ADC) in diversi aspetti**
  - Containers (docker, singularity) [A. De Salvo]
  - Database (Frontier, Conditions) [A. De Salvo, L. Rinaldi]
  - Harvester/Site Description (evoluzione del WFMS di ATLAS) [A. De Salvo]
  - CREM [dal 2018] [D. Barberis]
  - Installazione del software (CVMFS e distribuzione) [A. De Salvo]
  - ADC Monitoring [dal 2017] [D. Barberis]
  - ASCIG (ATLAS Software & Computing Infrastructure Group) [dal 2019] [D. Barberis]
  - Network infrastructure (LHCONE) [Tutta la federazione italiana dei T1/T2/T3 italiana]
  - Storage [G. Carlino, A. De Salvo, A. Doria, E. Vilucchi]
    - Federazioni di xrootd e HTTPD, DPM, Storage Caching, …
  - VO management [A. De Salvo, E. Vilucchi]

- **Altre attività (ex-PRIN)**
  - Cloud Computing, Object storage, etc. [A. De Salvo]
  - Hadoop (EventIndex) [D. Barberis, M. Villaplana]

- **Tier2**
  - Gestione infrastrutture, R&D, etc. [A. De Salvo, A. Doria, D. Rebatto, E. Vilucchi]

Monitoring delle performance dei siti e risorse

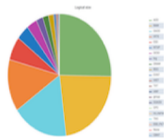
- https://adc-monitoring.web.cern.ch/

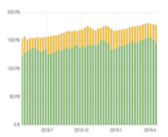
ATLAS EXPERIMENT — ADC Monitoring

Data Management


DDM Transfer Dashboard (current)


DDM Transfer Dashboard (historical)


DDM Global Accounting (snapshot)


DDM Global Accounting (historical)


DDM Site Accounting


WLCG Transfer Dashboard


FTS Transfer Dashboard


Rucio Dataset Recovery Service


Rucio Storage Monitoring

Grid Data Processing


BigPanDA Monitor


Job Accounting Dashboard


Kibana expert views


Harvester Grafana


Harvester Kibana

Tier-0 and Point 1


Tier-0 conTZole Graph


Tier-0 conTZole Tasks


Detector Operation

Sites


AGIS


Sites Downtimes


AGIS DDM Blacklisting


Hammercloud

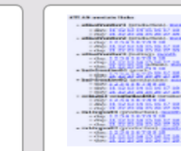

Pilot factory


SAM3 Visualization


ASAP metric

Databases


DB dashboard


Frontier AWSTAT

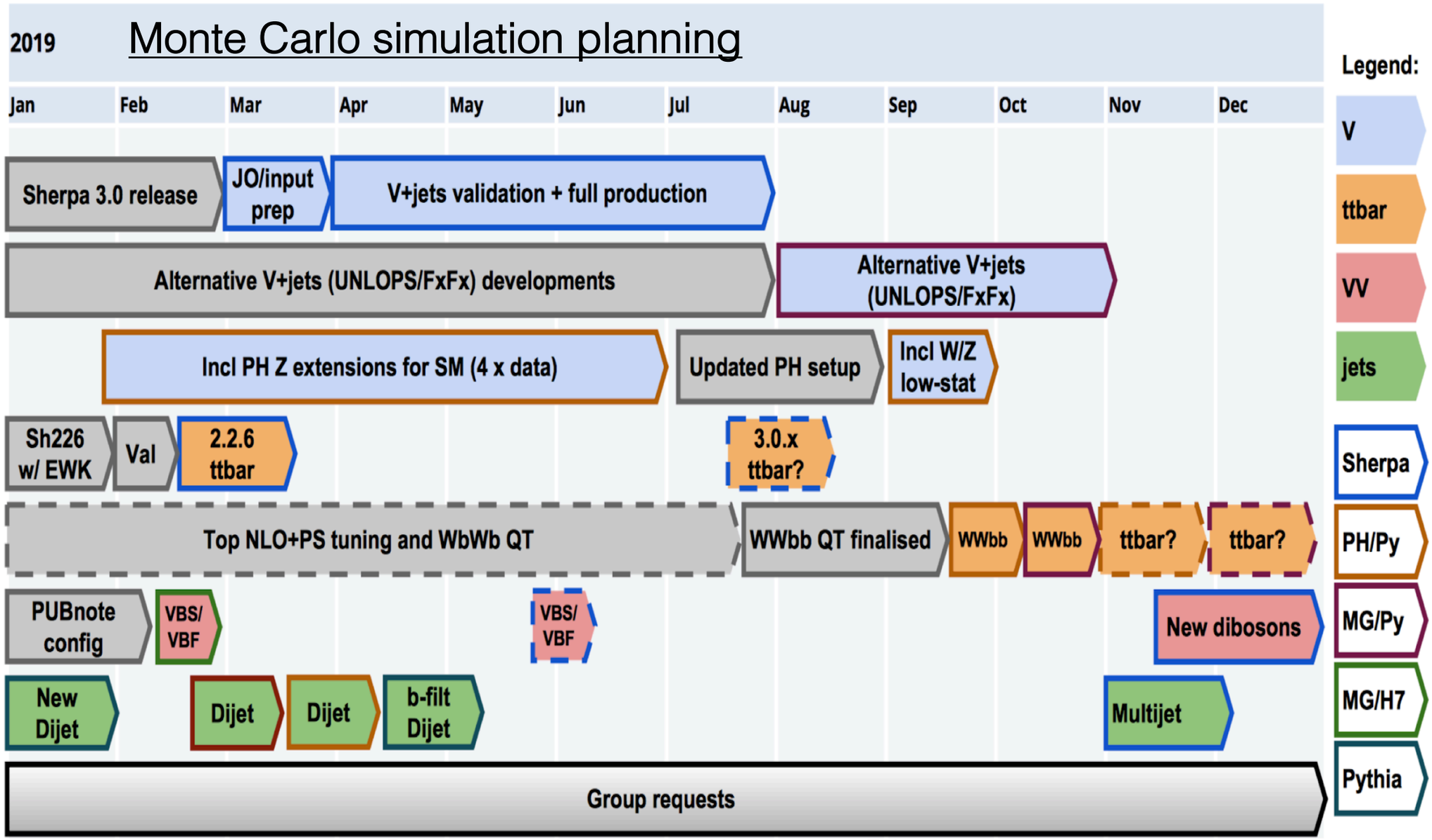

Frontier Launchpad Kibana


Frontier Squid

- **Sommario risorse**
  - https://docs.google.com/spreadsheets/d/1n3aJj05vv3vAbJ94zpgY-ebxRhWL7EWh2HsMUuLfHDA/edit#gid=852131157

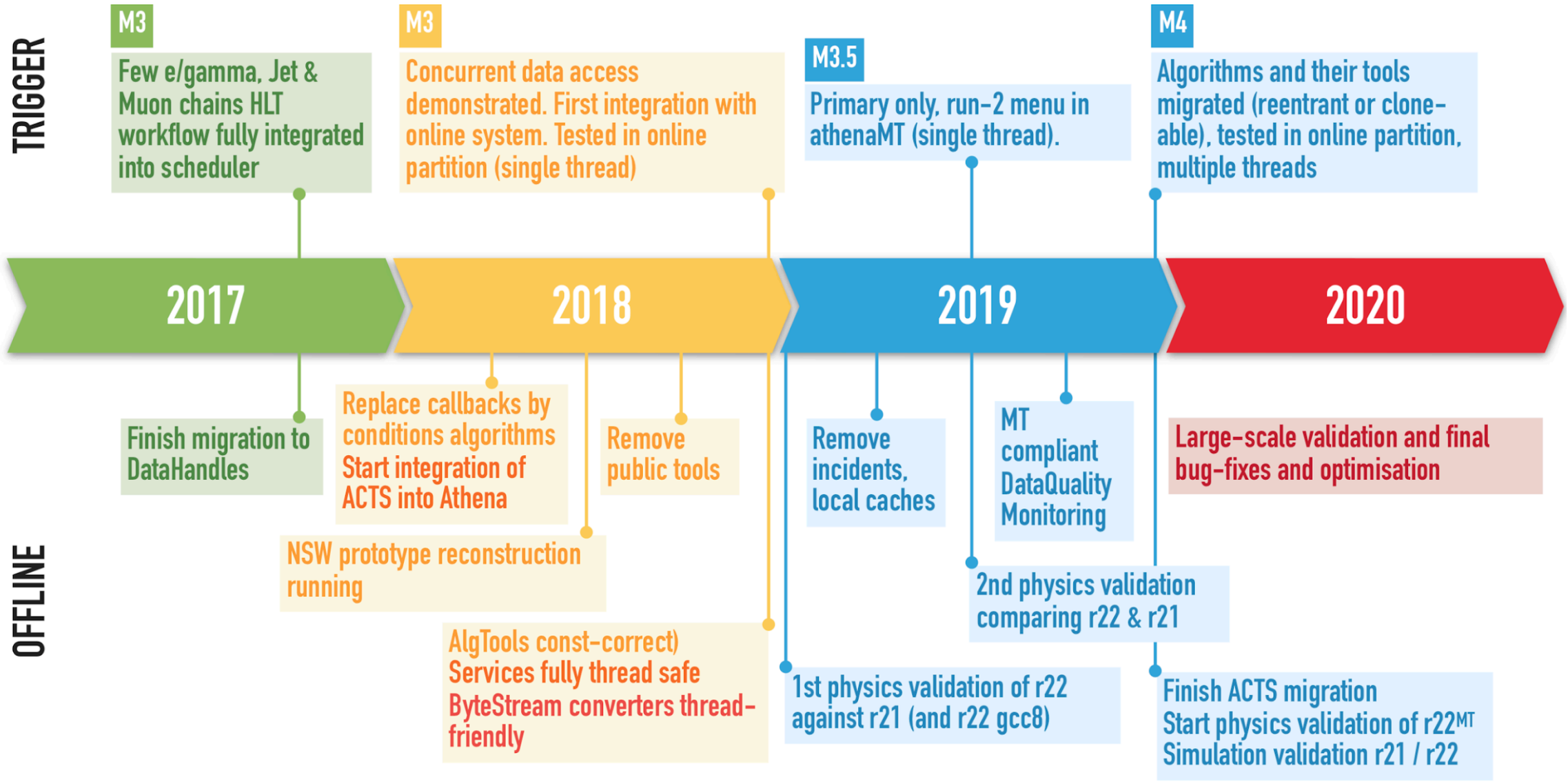
ATLAS Planning for LS2

- Data reprocessing planned for specialised cases
  - eg BPhysics and Light State stream for a fraction of 2019, Heavy Ion
- Full reprocessing of 2017-18 data under discussion
  - No compelling reason at present. 2017 reprocessing and 2018 Tier0 reconstruction are good for physics

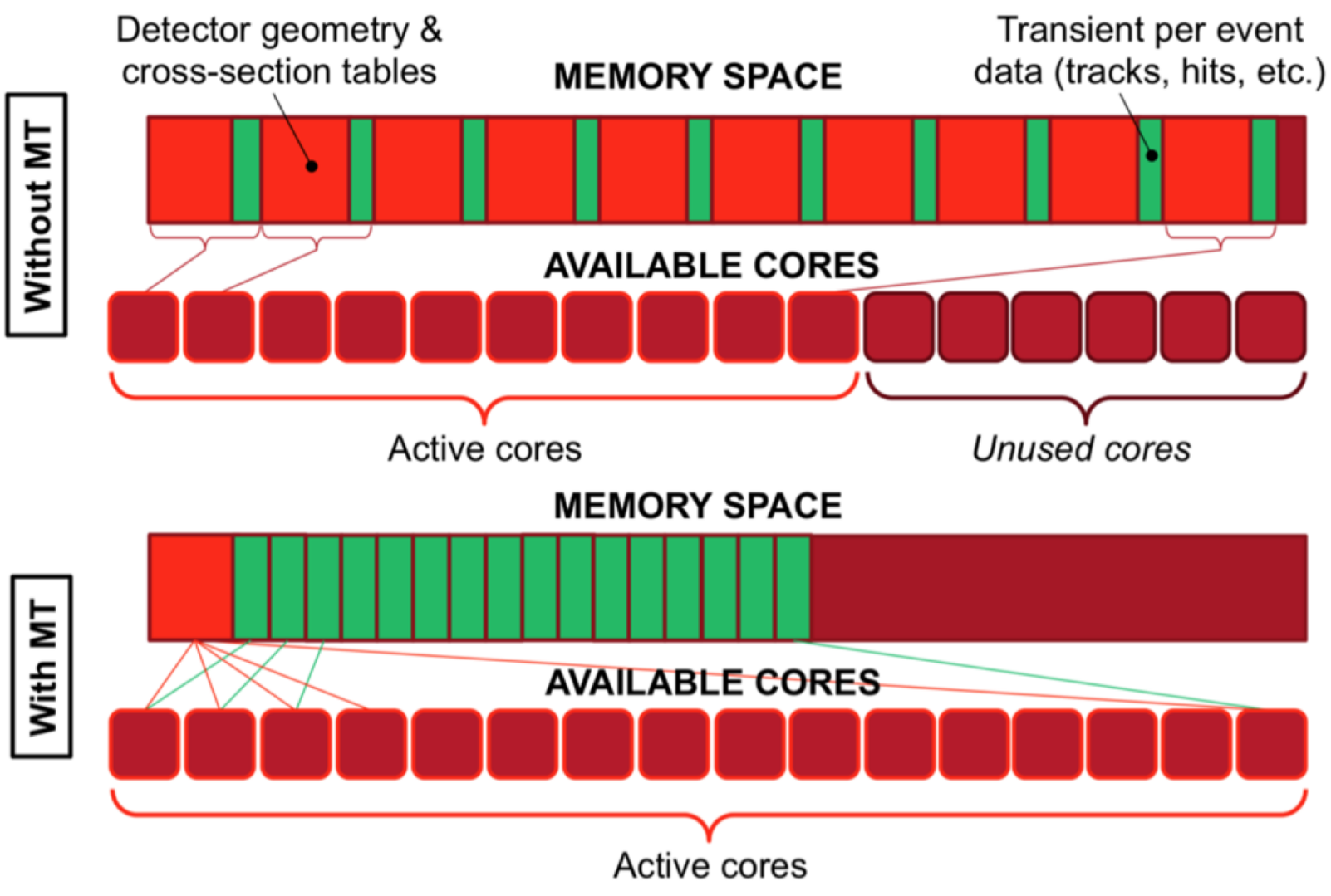


Long Shutdown 2 (LS2) ATLAS Software plans

- Updates software release plan.
  - Linked with Trigger milestones
  - More emphasis on validation
- Still understaffed. But we are seeing a reaction from the collaboration
  - Actively seeking new effort. A list of missing tasks was circulated to the collaboration



Progress towards multi-threading in ATLAS



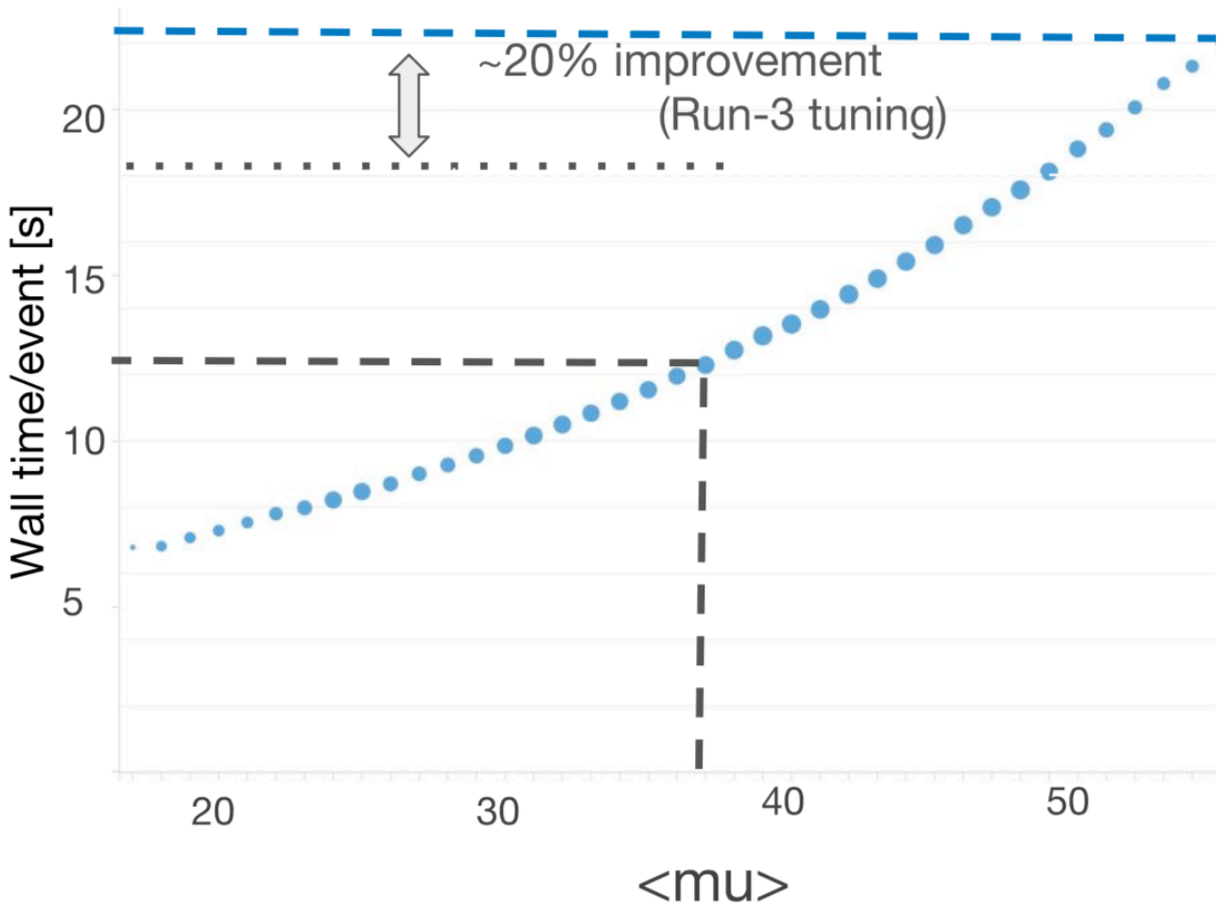
- Simulation with Geant4MT operational
  - Technical issues being sorted
- Parts of the reconstruction workflows now work in MT
  - Goal to expand the test suite validation in H2 of 2019
- Tracking migration team
  - Evolving towards ACTS
  - Closely monitored

ATLAS Computing in Run 3

- Broadly similar computing model to Run 2
  - Continue to use the highly successful nucleus-satellite model instead of the old Tier1/2
  - More HPCs coming on stream but traditional pledged and beyond-pledge grid resources still expected to provide the large bulk of the processing capacity (especially for data)
  - HLT reprocessing workflow on PanDA will stay the same as before.
    - Proven capability to spill excess Tier-0 workloads onto the wider grid when needed
    - How much availability of the HLT farm will we get for simulation?
- Disk and CPU resources are expected to increase in line with a "flat budget" (see next slides)
- Other innovations which will not affect the trigger:
  - More dynamic use of tape likely (tape carousel)
  - More operational intelligence and seamless access to different kinds of resources; new pilot mechanism

Expected ATLAS Grid resources in Run3: CPU

- No increases requested for 2020 for either disk or CPU
- Current working assumptions for Run 3
  - μ~60 during levelling, expected to be 12 hours/fill → pessimistically implies a doubling of processing time
  - Reconstruction tuning (especially: 7→8 hits in track selection) should offset this by ~20-30% so we expect to need 1.5-1.7x the current T0 CPU overall
- Should the T0 be unable to deliver this, we have the option of spilling over onto the wider grid
  - T0 processing rate should not be a limiting factor
  - BUT we need to understand the potential scale of such an eventuality to allow thorough testing: again this comes down to knowing the expected rate



- Simulation will drive the CPU requirements for the wider grid: expect overall 50% w.r.t. 2019 in 2020

Expected ATLAS Grid resources in Run3: Storage

- No increases requested for 2020 for either disk or CPU
- Disk for Run 3
  - AOD and DAOD dominates the usage; this will be restrained in Run 3 by the AMSG3 reductions (see the draft report)
    - Reminder that the trigger domain will deliver a significant portion of these savings: replacing navigation and trigger objects with decorations on offline objects in DAOD; activation of AOD_Run3_SMALL content in AOD
    - Expect to request disk increases according to "flat budget", e.g. 1.5-1.7x the Run 2 volumes; this will fully materialise ir 2022-2023
- Tape for Run 3
  - Expect to use tape more extensively in Run 3 in the "tape carousel" mechanism which will be mandated by AMSG3
  - Tape currently under-used; expect a request of around 1.5x Run 2, fully materialising in 2022-2023
- Storage is dominated by MC rather than data

ATLAS CPU, Disk, Tape for Run-3

- CPU: We aim to produce 50% of our simulation using Fast Calo Sim v2
  - Faster simulation means faster turnaround time for MC samples
  - MC statistics was an issue at the beginning of Run-2
  - ATLAS will be more agile to changes in generators, simulation, etc
  - Overall more MC will be produced
- Disk: ongoing analysis model for Run-3 study group
  - Disk growth is a concern (as highlighted by the C-RSG)
  - We aim to reduce our disk footprint for Run-3
  - Preliminary recommendation to the collaboration in February. Final report in June
    - Move towards DAOD_PHYS and DAOD_PHYSLITE
- Tape: No issues in Run-2
  - About 75% of our tape storage used. Will catch up towards the end of 2020
  - Reprocessing of MC and data running from tape
  - We will request more tape in 2021
- Resources expectations
  - Below flat-budget in 2019, No increase in 2020
  - We expect to return to flat budget increases in Run-3
    Eg (2021 resources) ~ 1.5x (2018 resources)
  - 2021 is a commissioning year, increase could be over 2021/22.
  - Still plan for a rich physics programme in 2021!

ATLAS Software deliverables for Run-3 and Run-4

Analysis software deliverables	Run 3 physics	Run 4 development
AMSG 3 recommendations implemented leading to at least 30% disk savings	**CRITICAL**	**CRITICAL**
Review of EDM and possible simplifications	**EXPEDIENT**	**EXPEDIENT**

Simulation deliverables	Run 3 physics	Run 4 development
FastCaloSimV2 fully validated and ready for production	**CRITICAL**	**CRITICAL**
FastChain complete and ready for production	**EXPEDIENT**	**CRITICAL**
Multi-threaded simulation production-ready	**EXPEDIENT**	**CRITICAL**

Conditions / databases deliverables	Run 3 physics	Run 4 development
MT-compliant in-file metadata	**EXPEDIENT**	**CRITICAL**
PyCOOL and AMI fully compliant with Python 3	**CRITICAL**	**CRITICAL**
Oracle services compliant with new licencing regime	**CRITICAL**	**CRITICAL**
COOL-REST prototype ready for production	**EXPEDIENT**	**CRITICAL**

Reconstruction deliverables	Run 3 physics	Run 4 development
Release 22 fully validated and performant in single threaded mode (AthenaMP)	**CRITICAL**	**CRITICAL**
Integration of New Small Wheels	**CRITICAL**	**CRITICAL**
Integration of new tracking software (ACTS)	**EXPEDIENT**	**CRITICAL**
Release 22 able to run in multi-threaded mode (AthenaMT)	**EXPEDIENT**	**EXPEDIENT**

Common software deliverables	Run 3 physics	Run 4 development
Offline software fully Python 3 compliant	**EXPEDIENT**	**CRITICAL**
Configuration mechanism overhauled	**CRITICAL**	**CRITICAL**
New geometry framework ready for integration for Run-4	**Not relevant**	**CRITICAL**
AthenaMT performance optimised	**EXPEDIENT**	**CRITICAL**

Evolution of WLCG

- Community White Paper
  - 1 year – bottom up review of LHC computing topics
  - 13 working groups on all aspects
  - Outlines how HEP computing could evolve to address computing challenges
  - https://arxiv.org/abs/1712.06982

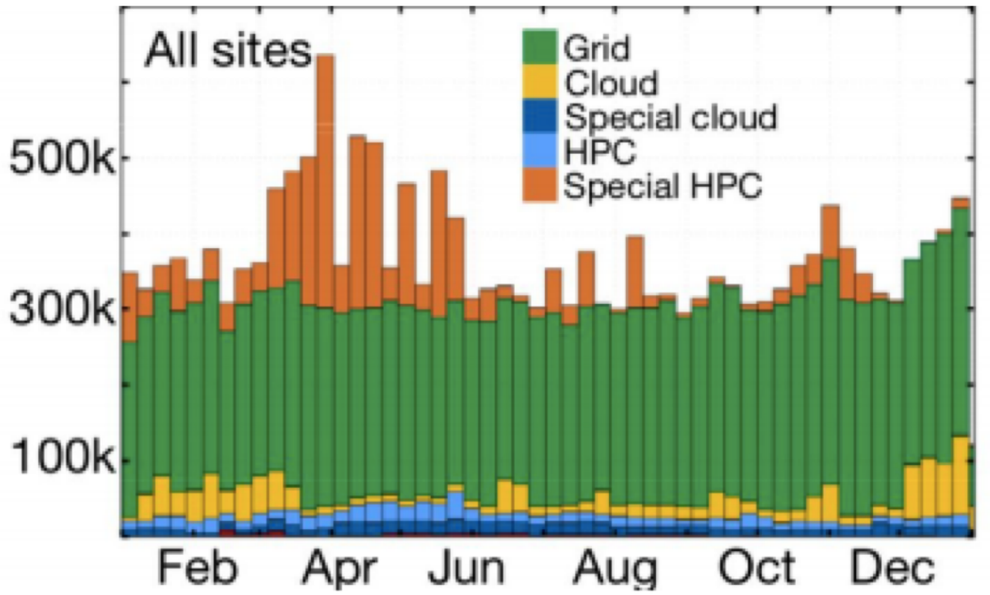
- WLCG Strategy Document
  - Prioritisation of topics in the CWP from the point of view of the HL-LHC challenges
  - Set out a number of R&D projects for the next 5 years
    - Running global system should evolve towards HL-LHC
  - http://cern.ch/go/Tg79

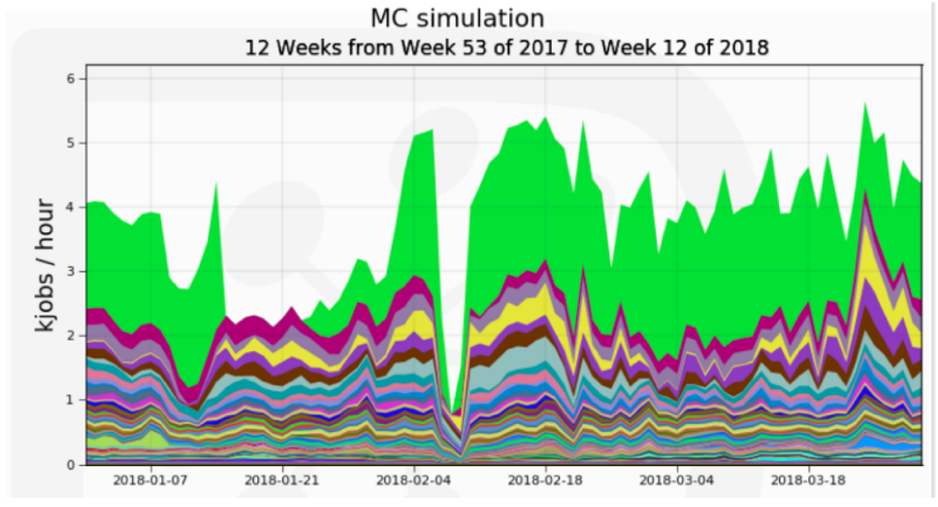
WLCG: Heterogenous computing

- Today get opportunistic use of many types of compute, in particular HPC systems, and HLT

- In future, this heterogeneity will expand; we must be able to make use of all types:

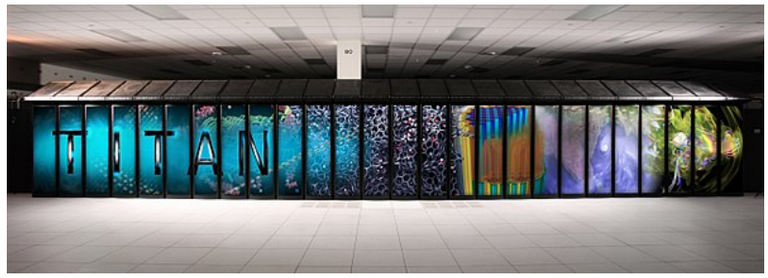
  - Non-x86 (esp GPU), HPC, clouds, HLT farms (inc FPGA?)





HPC computing in WLCG

- HPC are optimized for tightly coupled calculations, HEP applications are not designed to exploit those capabilities
  - HEP use cases require finer granularity than typical HPC applications
  - Hard to optimize: The software framework for each of the experiments is several million lines of C++ & Python and contributions from ~1000 people
- Each resource is huge but independent
  - Authorization, access, interfaces are all specific to the site
- Data access: HEP workflows often make heavy use of data and experiment specific services
  - Limitations in ingoing/outgoing access (policy) require rethinking
  - HEP data scale not suited to data distribution on an HPC
- Interfaces: Need for common interfaces for access, data handling and site services (connectivity, s/w distribution, containers, ..)
  - FNAL HEPCloud and CERN have similar approaches

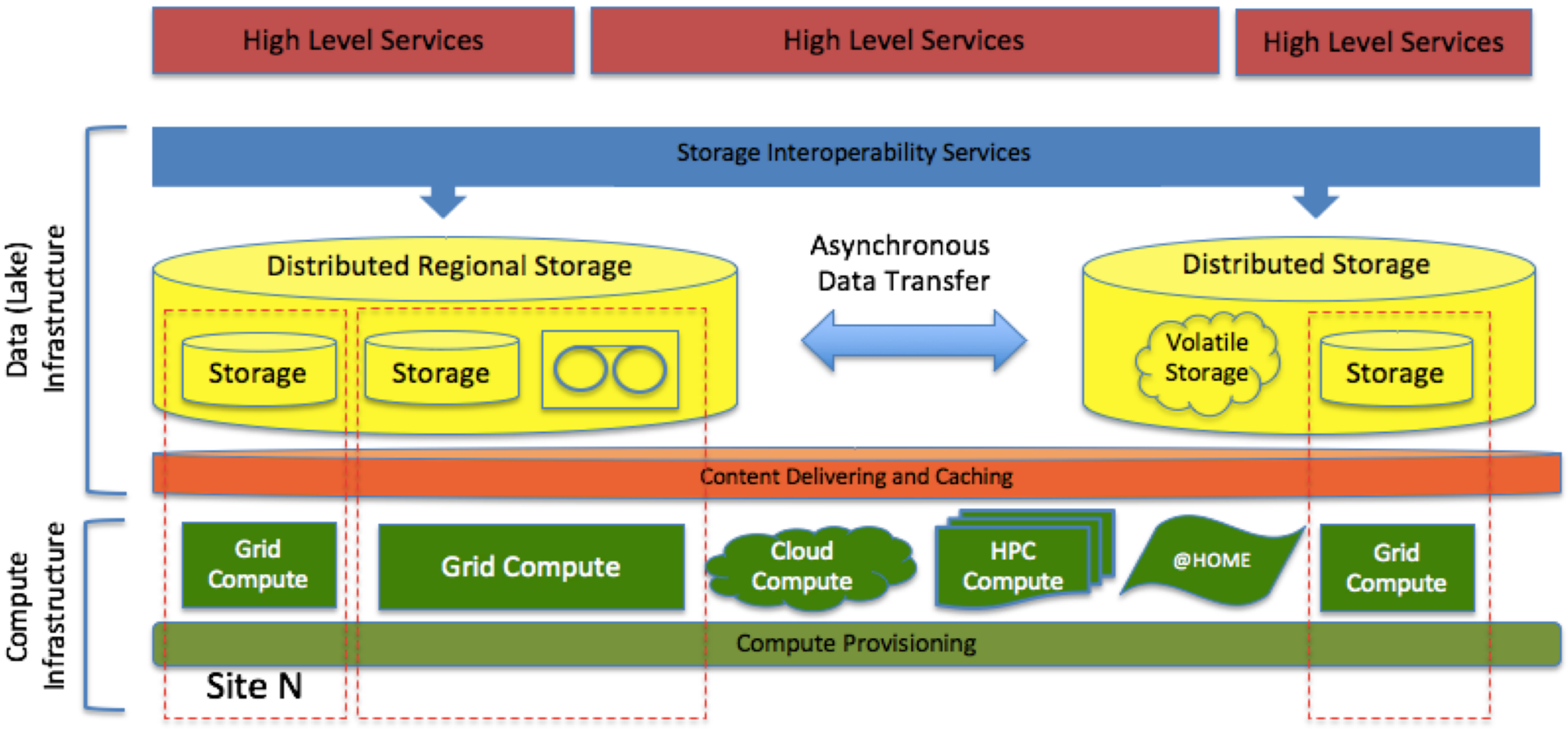


EuroHPC Joint Undertaking

BDEC2 BIG DATA AND EXTREME-SCALE COMPUTING 2

HEP engagement with DOE & NSF in USA and (together with SKA) with PRACE and EuroHPC in Europe and participating in BDEC2 workshops

WLCG: Data delivery “data lake (cloud)”



Idea is to localize bulk data in a cloud service (Tier 1’s ➔ data lake): minimize replication, assure availability

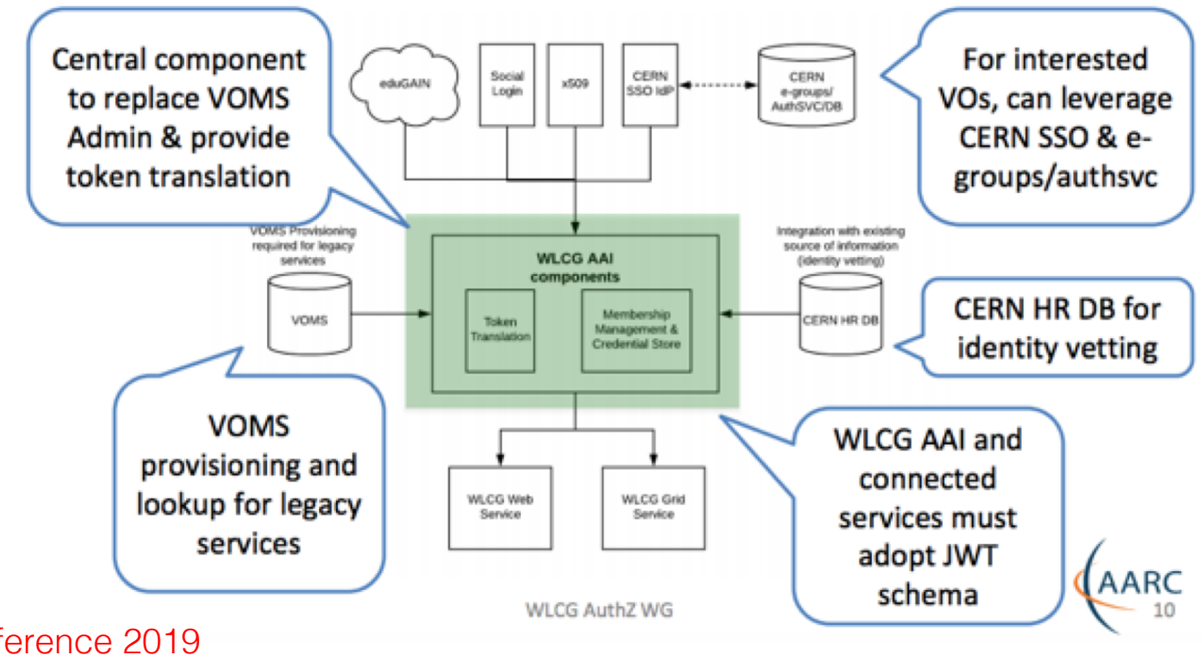
Serve data to remote (or local) compute – grid, cloud, HPC, ???

Simple caching is all that is needed at compute site

Works at national, regional, global scales

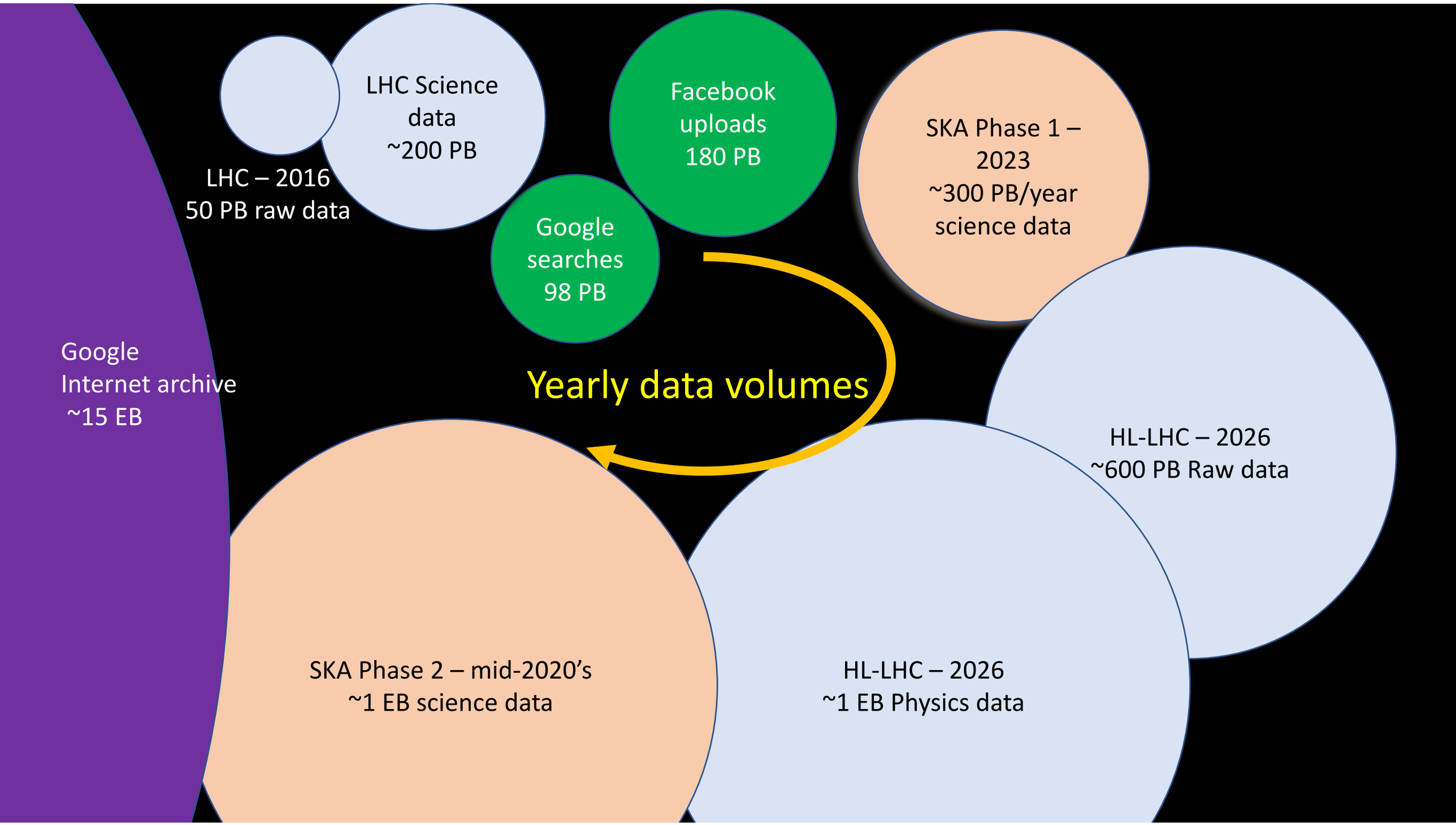
WLCG: Data management and storage

- Set of R&D projects to prototype such a data management infrastructure – and associated tools
- Aims:
  - Reduce the global cost of storage (hw and operations)
  - Enable a more effective use of existing storage
  - Be able to efficiently and scalably deliver data to large, remote, heterogenous, compute resources (LHC Tier centres or HPC, clouds, other opportunistic)
  - Build a common set of DM tools that can be used by a broad set of scientific experiments
    - Today LHC, DUNE, SKA, Belle-II, GW-3G, and others are all looking at a common set of identified tools
- Also collaboratively (LHC+SKA with GEANT) looking at underlying data transfer and network tools (replace gridftp, network protocols, etc.)
- Evolution of the AAI solutions from X.509 towards token-based systems
  - Following AARC, AARC2 models
  - In line with most modern network services



I. Bird – EGI Conference 2019

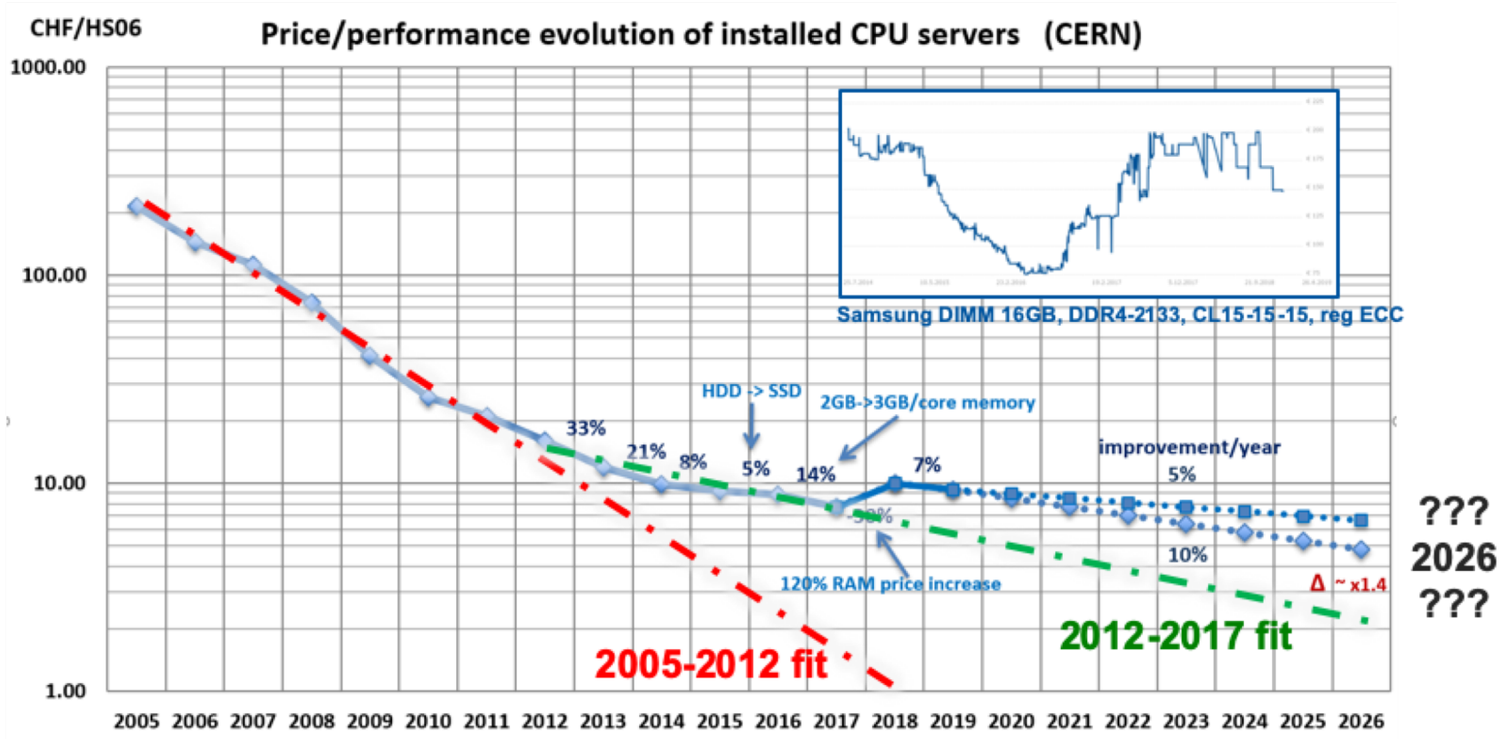
Data sizes

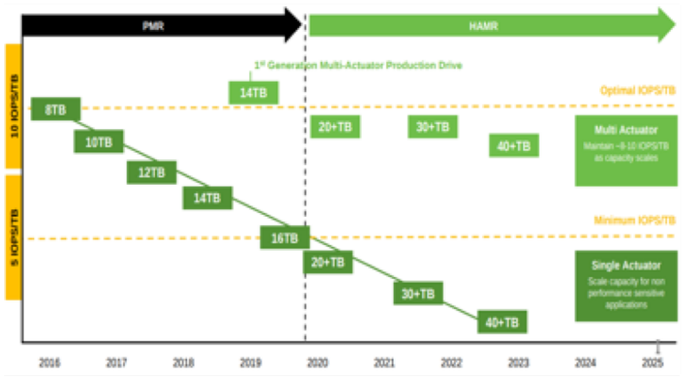

LHC – 2016
50 PB raw data
LHC Science data
~200 PB
Google searches
98 PB
Facebook uploads
180 PB
SKA Phase 1 – 2023
~300 PB/year
science data
Google
Internet archive
~15 EB
Yearly data volumes
HL-LHC – 2026
~600 PB Raw data
SKA Phase 2 – mid-2020's
~1 EB science data
HL-LHC – 2026
~1 EB Physics data

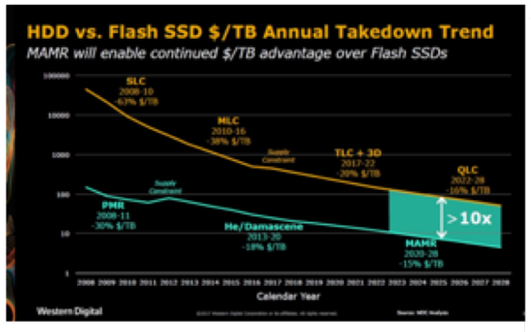
WLCG: Hardware Cost Evolution

- Previous assumptions of ~20%/year effective cost improvement for CPU & storage no longer true
- *Market-driven* rather than technology
- Science has no influence on these markets

- We have serious risks:
- Our budget outlooks are constrained (flat!)
- Risk of technologies disappearing
  - e.g. tape due to market forces
  - disk technology future not clear – costs not obvious

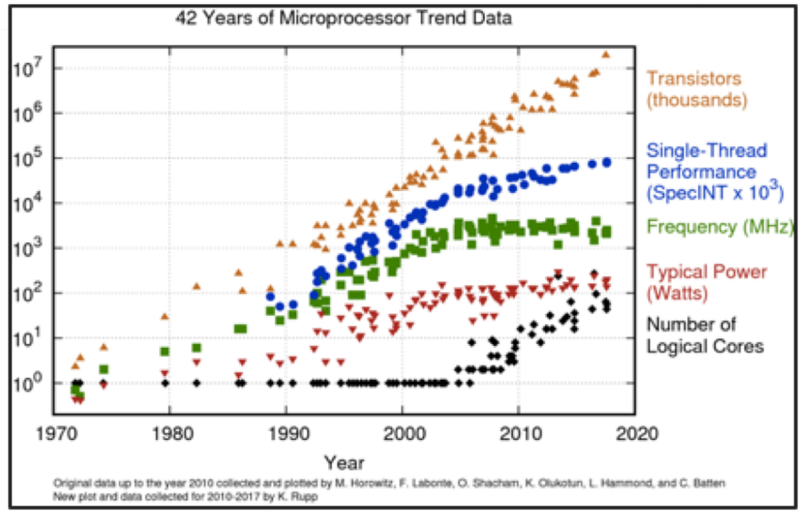






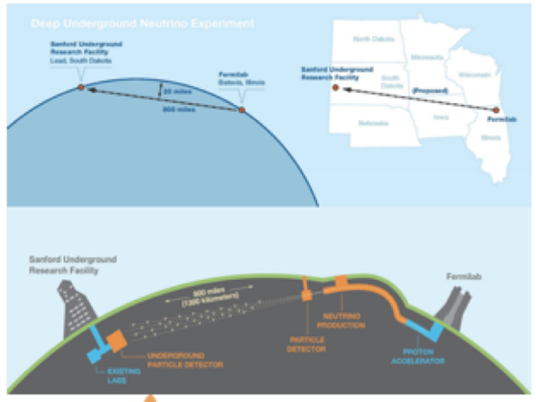
EGI, 6 May 2019

Oracle's StorageTek (StreamLine) tape library product range will be end-of-lifed, *El Reg* has learned.



WLCG Collaborations

Collaborations

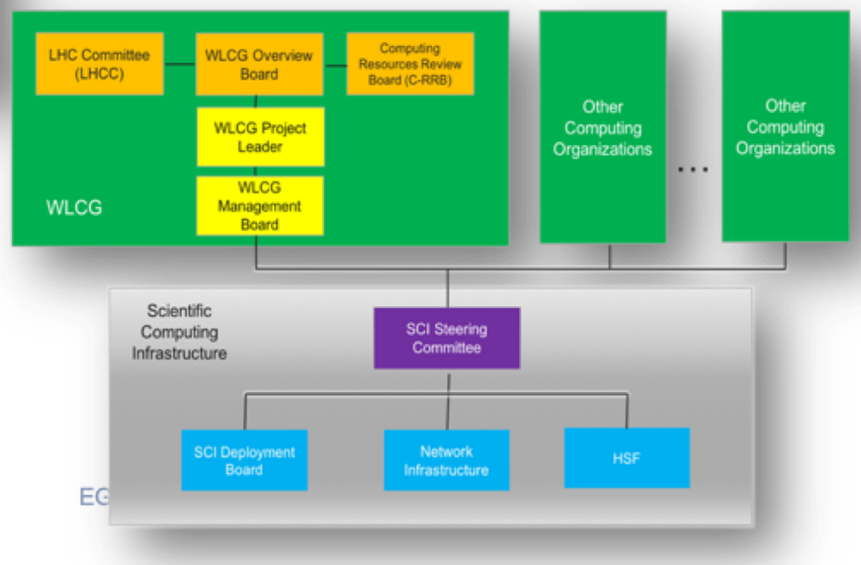


DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

DUNE will leverage the WLCG for its computing infrastructure

First formal non-LHC "associate" member of WLCG

Belle II



Astroparticle Physics European Consortium (APPEC)

APPEC

APPEC Contribution to the European Particle Physics Strategy

December 17, 2018

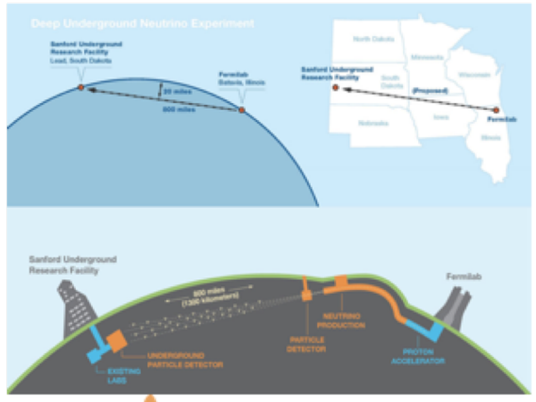
Editorial Board:
S. Katsanevas, A. Masiero, T. Montaruli, J. de Kleuver, A. Haungs

Contact Person:
T. Montaruli (APPEC Chair from Jan. 1, 2019)
Email: teresa.montaruli@unige.ch

Website : http://www.appec.org

WLCG Collaborations

Collaborations

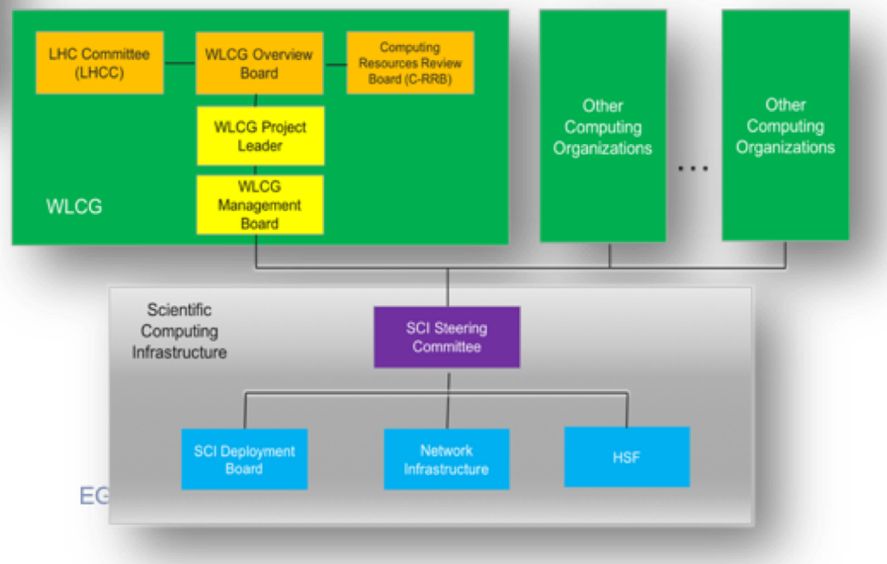


DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

DUNE will leverage the WLCG for its computing infrastructure

First formal non-LHC "associate" member of WLCG

Belle II



Astroparticle Physics European Consortium (APPEC)

APPEC

APPEC Contribution to the European Particle Physics Strategy

December 17, 2018

Editorial Board:
S. Katsanevas, A. Masiero, T. Montaruli, J. de Kleuver, A. Haungs

Contact Person:
T. Montaruli (APPEC Chair from Jan. 1, 2019)
Email: teresa.montaruli@unige.ch

Website : http://www.appec.org

WLCG Collaborations

ESFRI Science Projects

HL-LHC	SKA
FAIR	CTA
KM3Net	JIVE-ERIC
ELT	EST
EURO-VO (LSST)	EGO-VIRGO (CERN,ESO)

ESCAPE

European Science Cluster of Astronomy & Particle physics
ESFRI research infrastructures

Horizon 2020 funded project

EUROPEAN OPEN SCIENCE CLOUD

Goals:

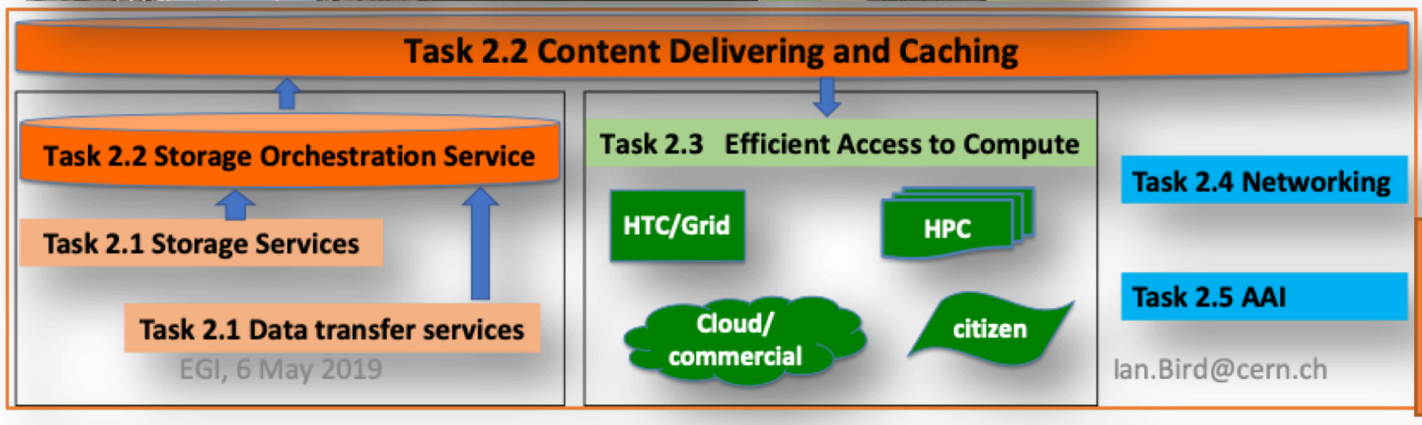
Prototype an infrastructure for the EOSC that is adapted to the Exabyte-scale needs of the large ESFRI science projects.

Ensure that the science communities drive the development of the EOSC.

Has to address *FAIR* data management, long term preservation, open access, open science, and contribute to the EOSC catalogue of services.

Work Packages

WP2 – Data Infrastructure for Open Science
WP3 – Open-source scientific Software and Service Repository
WP4 – Connecting ESFRI projects to EOSC through VO framework
WP5 – ESFRI Science Analysis Platform



Data centres (funded in WP2)
CERN, INFN, DESY, GSI, Nikhef, SURFSara, RUG, CCIN2P3, PIC, LAPP, INAF 27