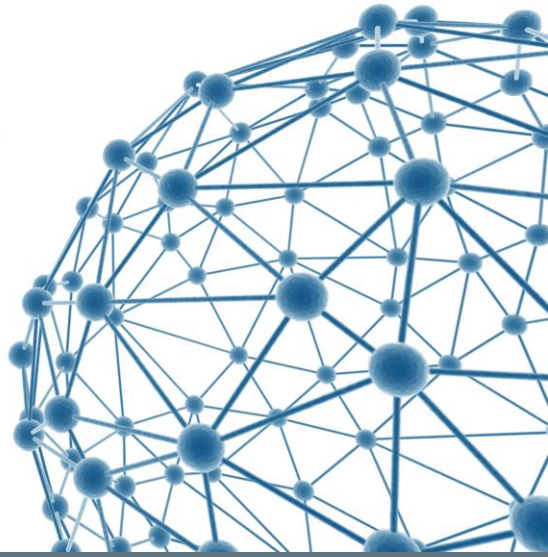


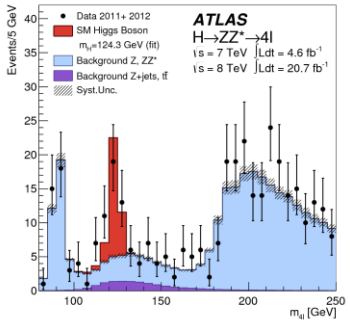

Data Management
Focus on the ATLAS experiment
(which is the one I know best)

Simone Campana, CERN

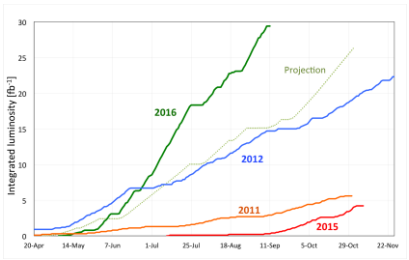


The Large Hadron Collider at CERN

Higgs discovery in Run-1



We are here: Run-2

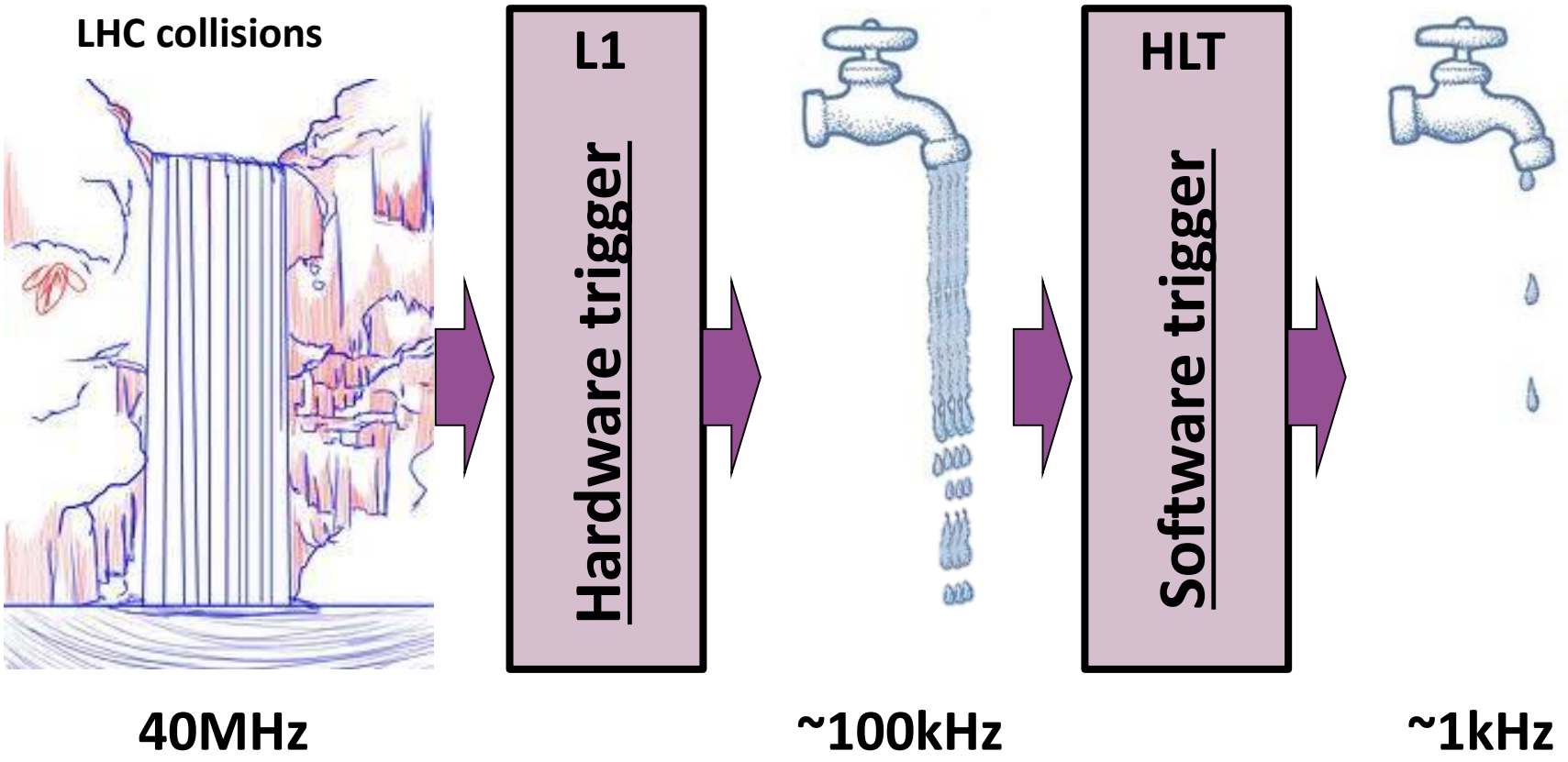


High Luminosity: the HL-LHC challenge

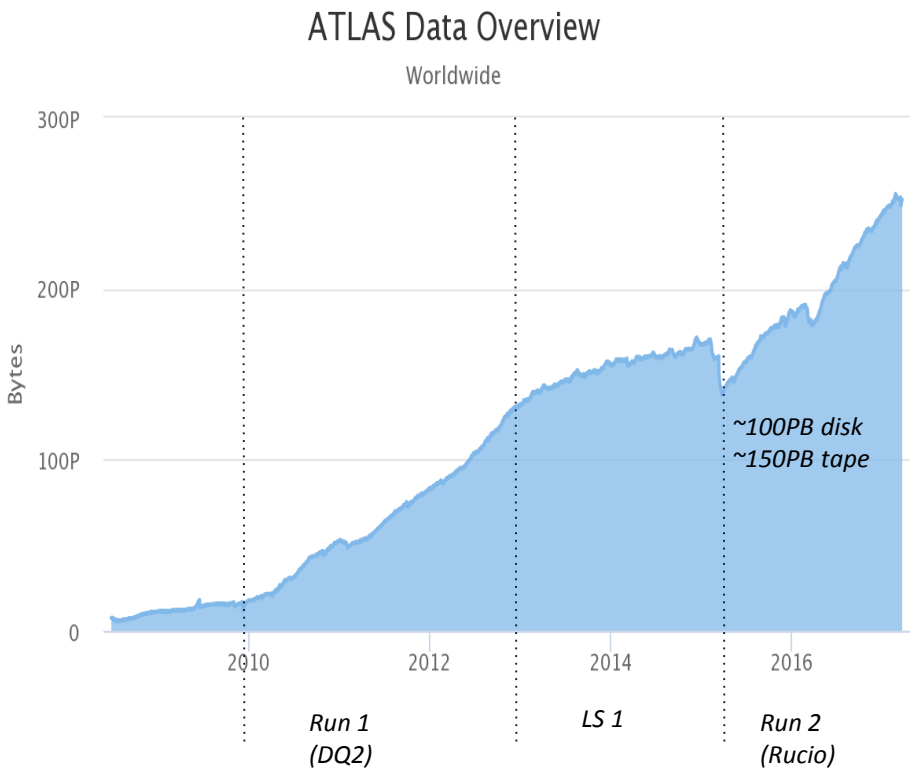
2009	Start of LHC - 2009: √s = 900 GeV
2010	Run 1: √s = 7-8 TeV, L = 2-7 x 10^{33} $cm^{-2}s^{-1}$
2011	Bunch spacing: 75/50/25 ns (25 ns tests 2011; 2012 ?) — ~25 fb^{-1}
2012	
2013	LHC shutdown to prepare for design energy and *nominal* luminosity
2014	
2015	Run 2: √s = 13-14 TeV, L = 1 x 10^{34} $cm^{-2}s^{-1}$
2016	Bunch spacing: 25 ns — >50 fb^{-1}
2017	
2018	Injector and LHC Phase-I upgrade to go to *ultimate* luminosity
2019	Run 3: √s = 14 TeV, L = 2 x 10^{34} $cm^{-2}s^{-1}$
2020	Bunch spacing: 25 ns — ~300 fb^{-1}
2021	
2022	High-luminosity LHC (*HL-LHC*), crab cavities, lumi levelling, ...
2023	
...	
2030	Run 4: √s = 14 TeV, L = 5 x 10^{34} $cm^{-2}s^{-1}$
	Bunch spacing: 25 ns — ~3000 fb^{-1}

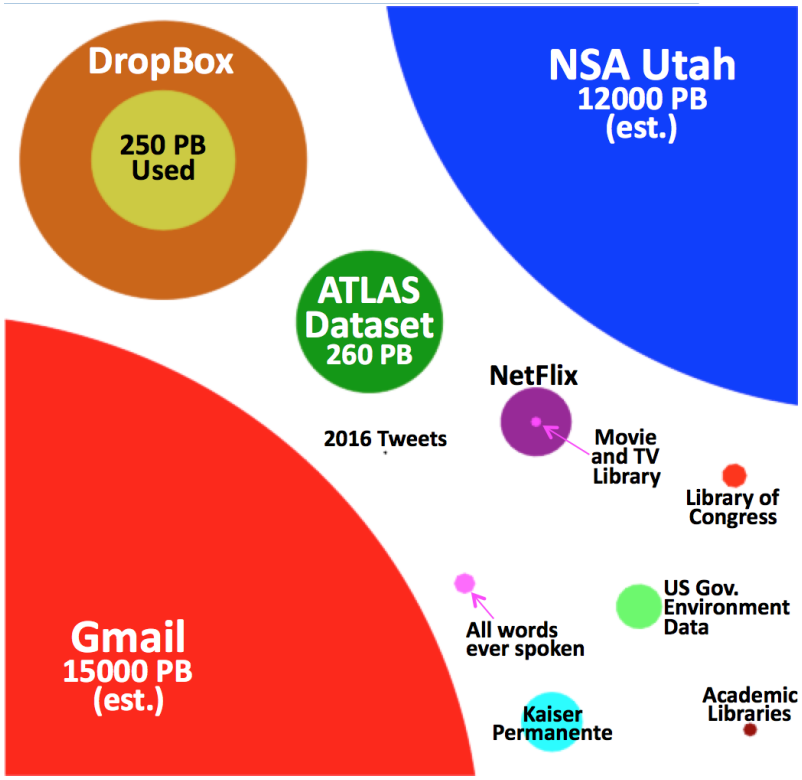
∫ L *dt*

Triggering on physics



ATLAS Distributed Data Management

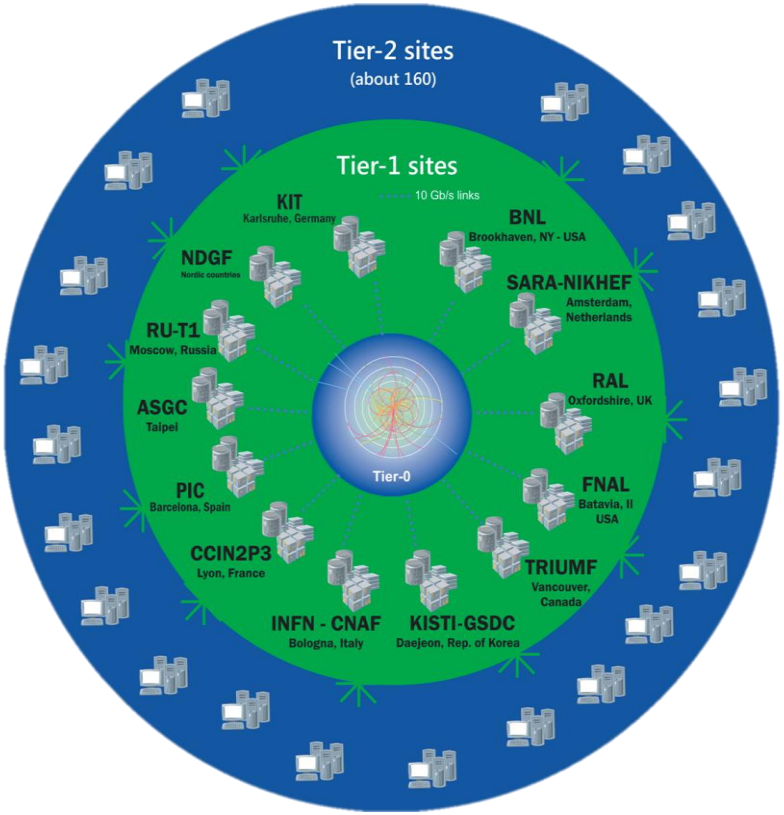




Worldwide LHC Computing Grid

- International collaboration to distribute and analyse LHC data
- Integrates computing centres worldwide that provide **computing** and **storage** resource into a single infrastructure accessible by all LHC physicists

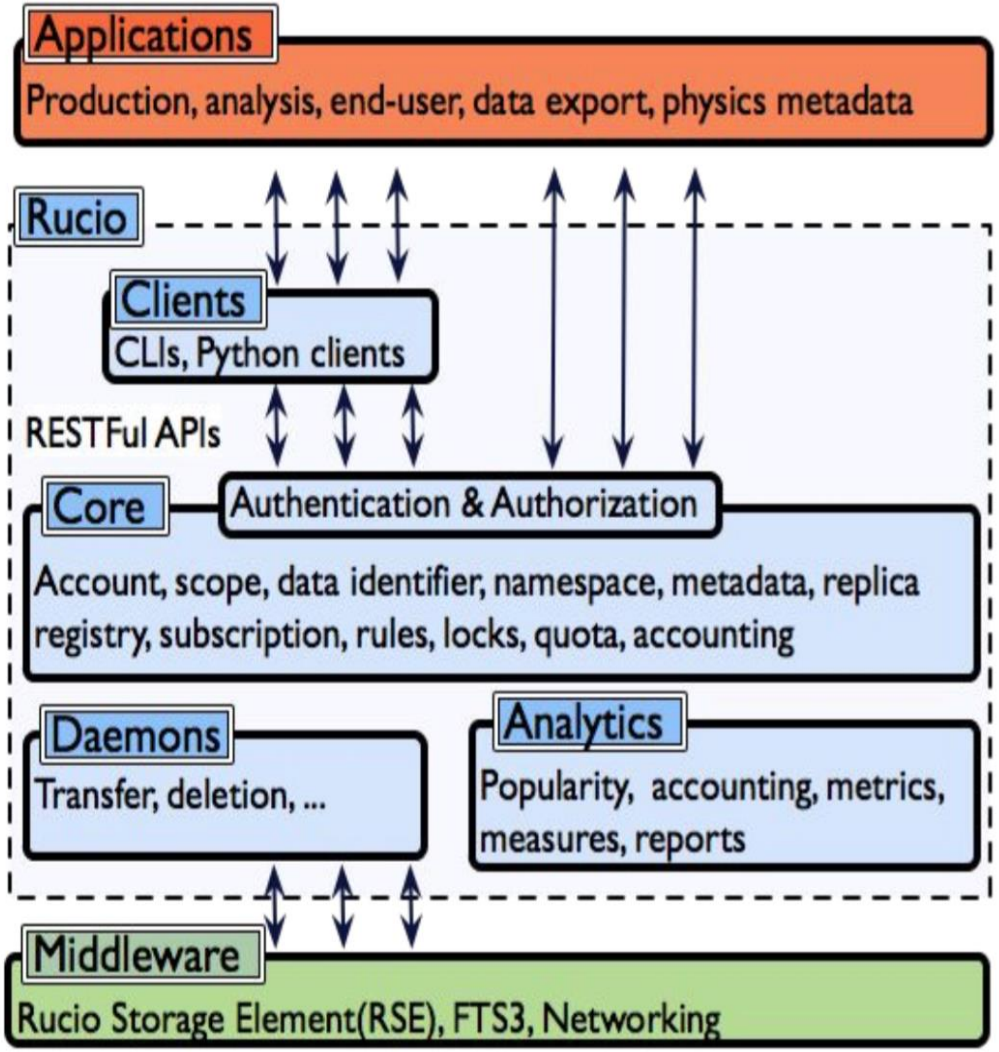
- **Tier-0 (CERN)**: data recording and archival, prompt reconstruction, calibration and distribution
- **Tier-1s**: T0 overspilling, second tape copy of detector data, more intensive tasks
- **Tier-2s:** Processing centers, being the differences with T1s increasingly blurry - more later



For **all** experiments:

- nearly 170 sites
- ~500k cores
- 200 PB of disk
- 10 Gb links and up

Data Management: Rucio architecture

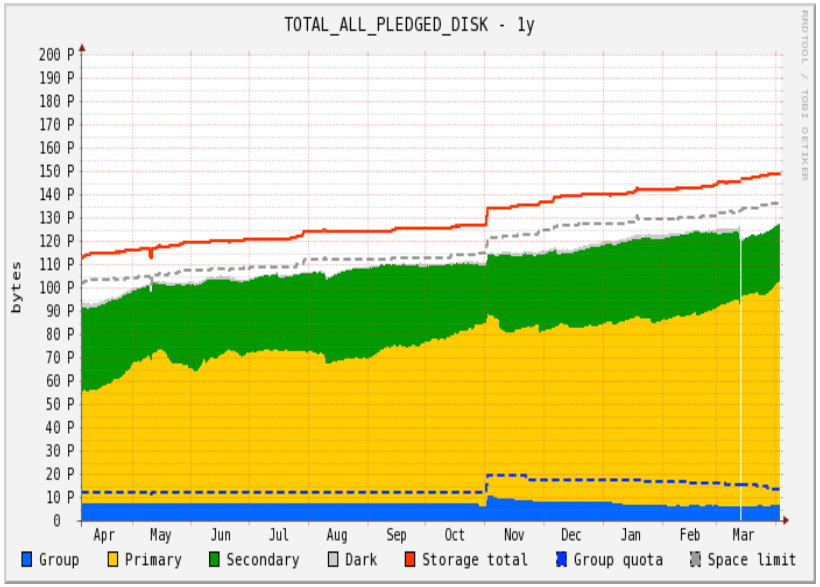


Rucio features and concepts

- Rucio accounts can be mapped to users or groups
- Namespace is partitioned by scopes (users, groups and other activities)
- Data ownership for users and groups: possibility to enable quota systems
- Replica management: rules define number of replicas and conditions on sites
- Granular data handling at file level - no external file catalogs
- Support of multiple protocols for file handling (access/copy/deletion)
  - SRM, HTTP/WebDAV, gridFTP
- Metadata storage: extensible key-value implementation
  - System-defined: size, checksum, creation time
  - Physics: number of events
  - Production: job/task that created the file

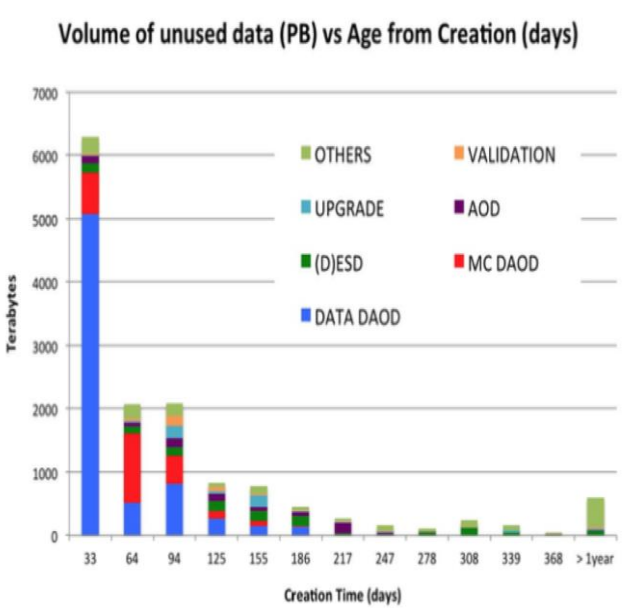
Data policies and lifecycle

- ATLAS relies on fully dynamic data replication and deletion
- Minimalistic pre-placement of only 2 replicas
- Data categories:
  - Primary (resident): base replicas guaranteed to be available on disk. Not subject to automatic clean up
  - Secondary (cache): extra replicas dynamically created and deleted according to the usage metrics
- Data rebalancing: redistribution of primary copies of popular datasets to disk resources with free space



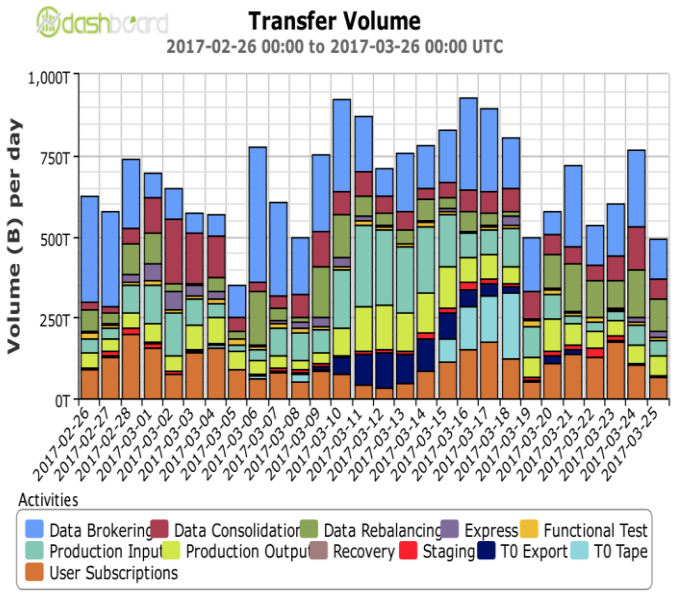
Data policies and lifecycle

- Every dataset has a lifetime set at creation
  - 6 months for Analysis inputs - fast turnaround
  - 2-3 years for Monte-Carlo simulations - expensive to regenerate
  - Infinite for RAW
- Lifetime can be extended if the data is accessed
- Expired datasets can disappear any time

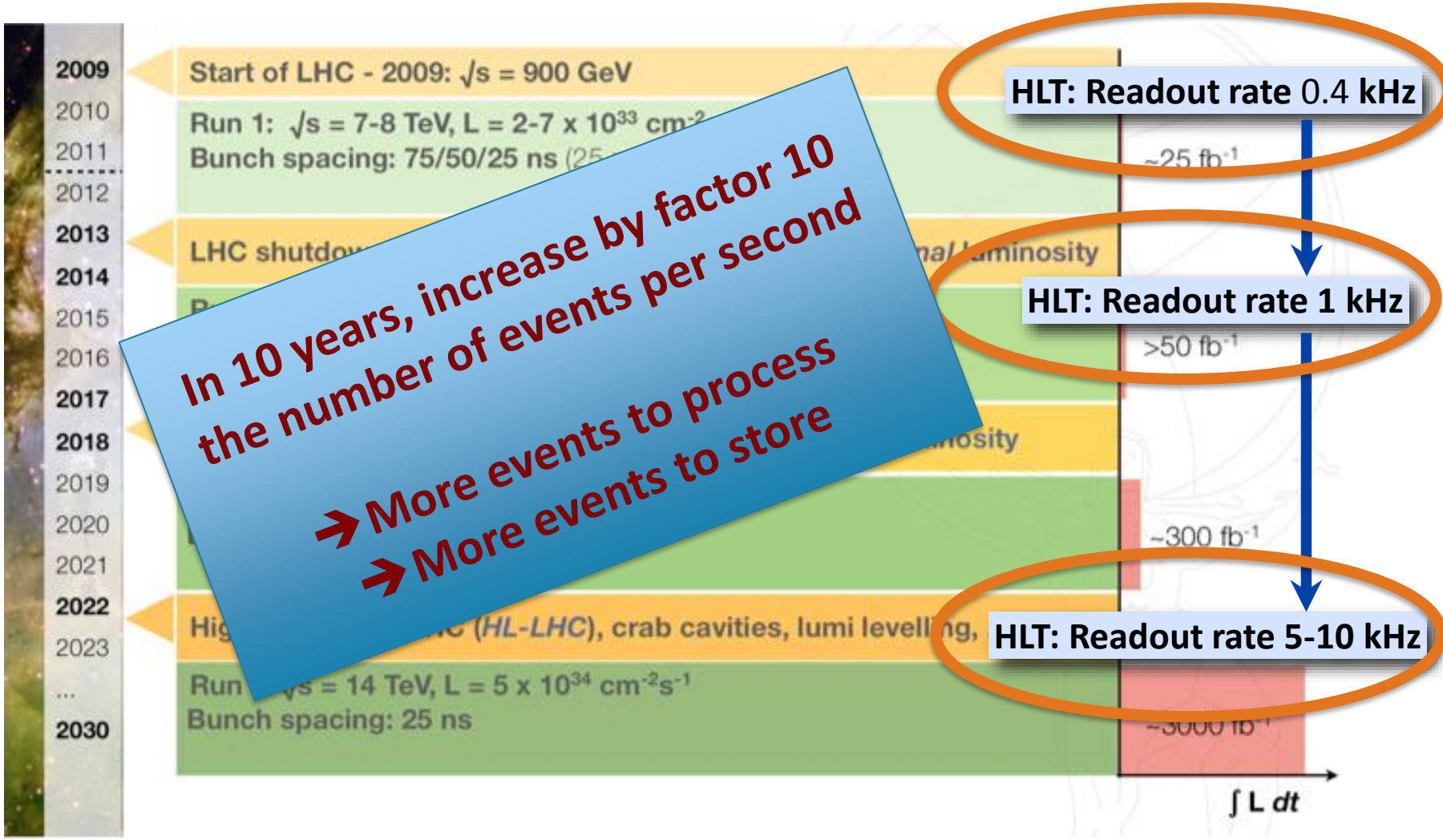


Data Management: some metrics

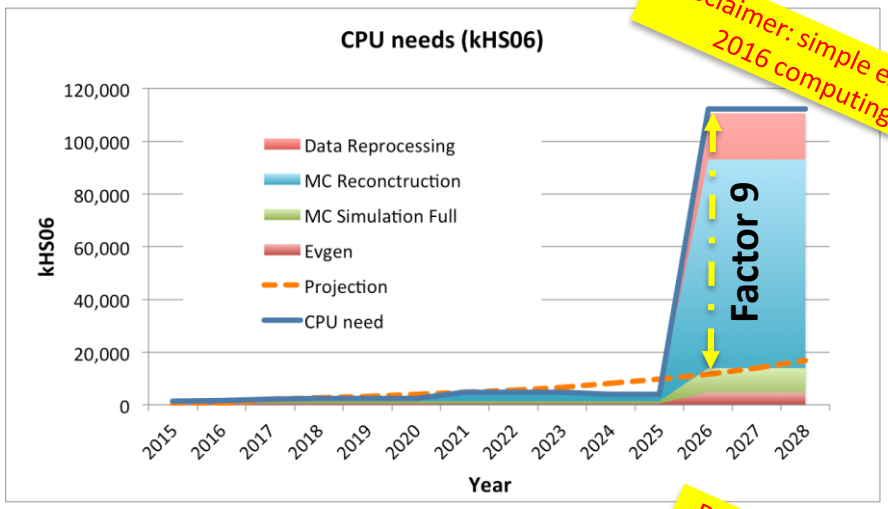
- Transfers
  - >40M files/month
  - Up to 40 PB/month
- Download
  - 150M files/month
  - 50 PB/month
- Deletion
  - 100M files/month
  - 40 PB/month



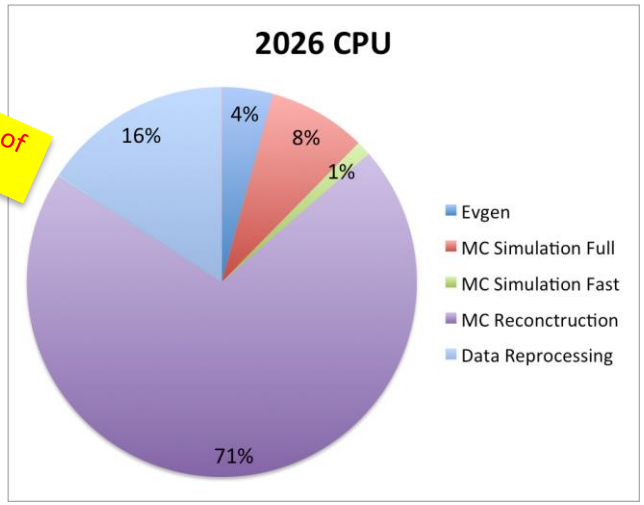
The data rate and volume challenge

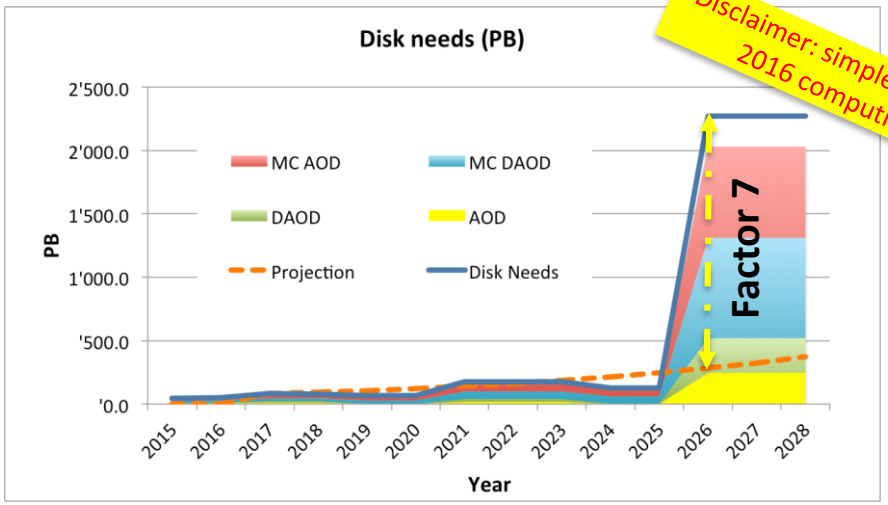


HL-LHC baseline resource needs

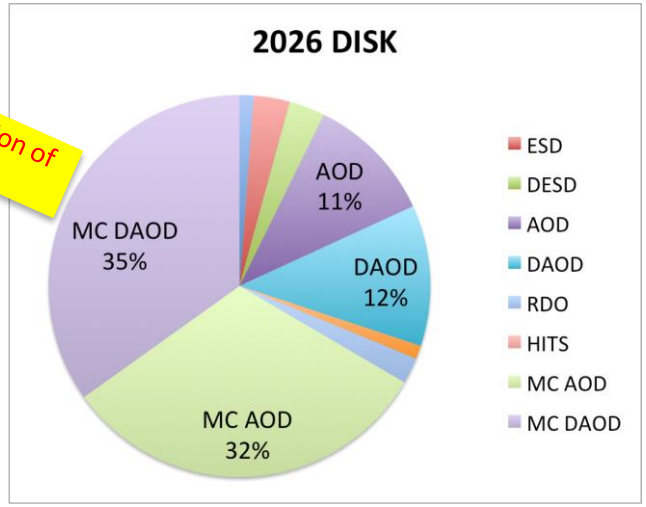


Disclaimer: simple extrapolation of 2016 computing model !

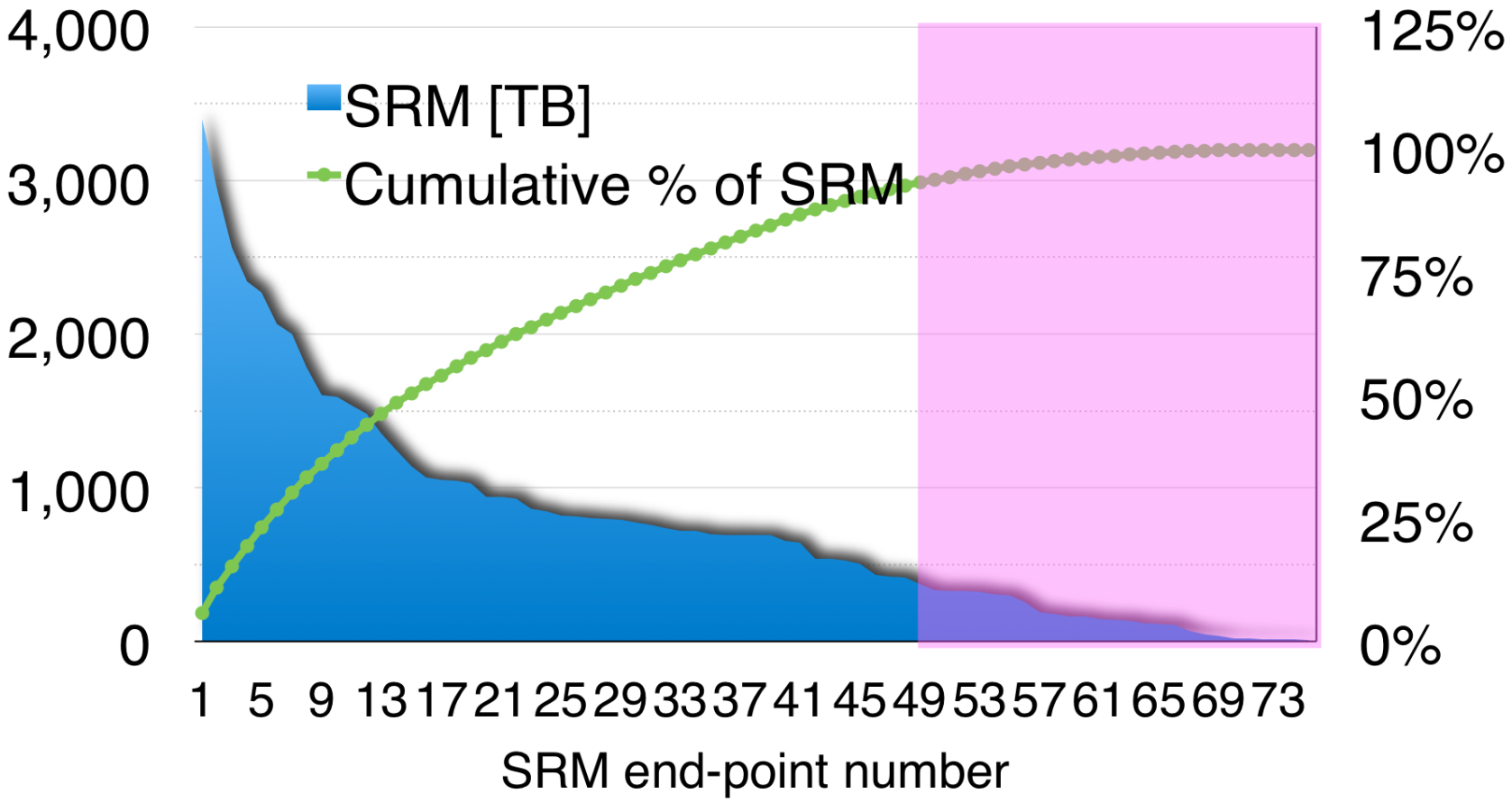




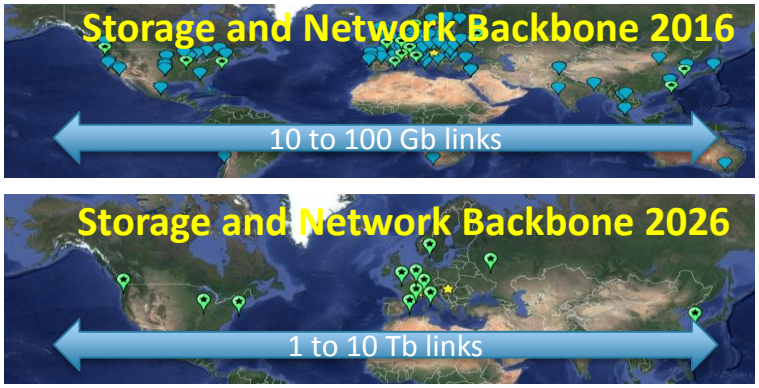
Disclaimer: simple extrapolation of 2016 computing model !

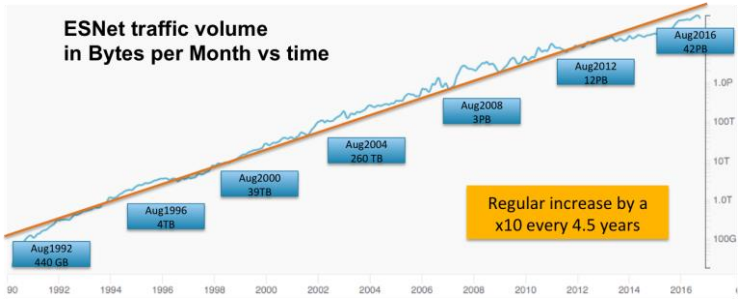


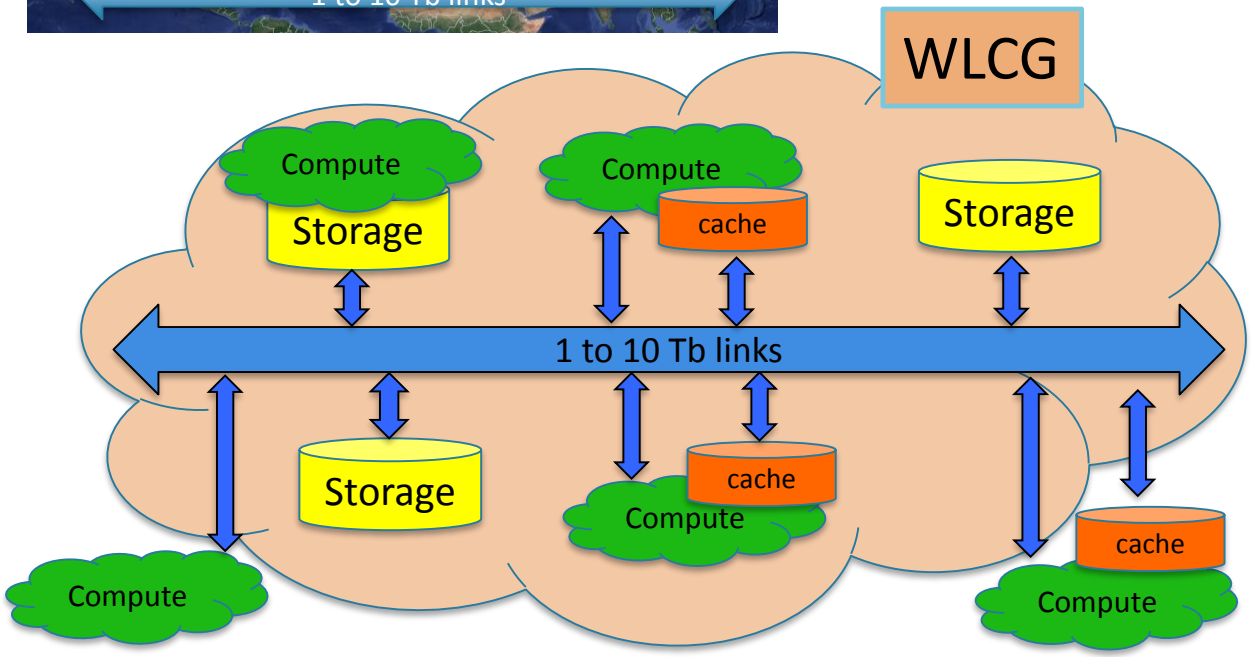
Storage Fragmentation (one year ago)



Computing infrastructure in HL-LHC







A data cloud for science

Storage and Compute loosely coupled but connected through a fast network

Heterogeneous Computing facilities (Grid/Cloud/HPC/ …) both in and outside the cloud

Consolidated Storage Endpoints