

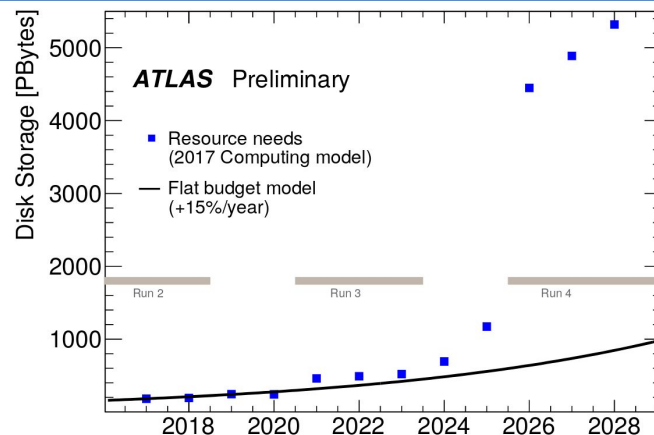
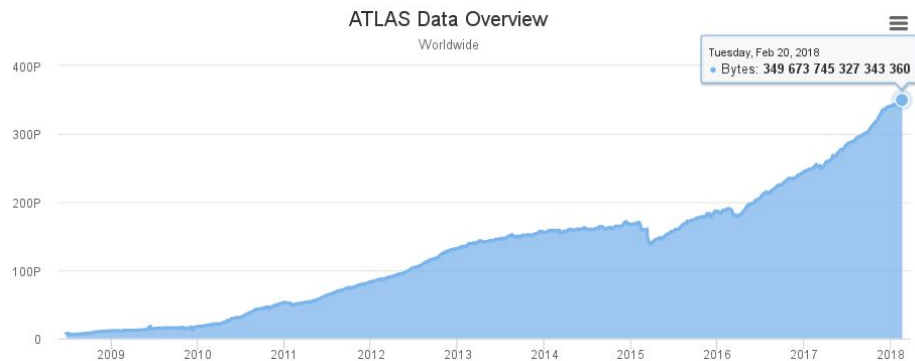


Preparing for HL-LHC data management

Mario.Lassnig@cern.ch

CERN EP-ADP-CO

The challenge



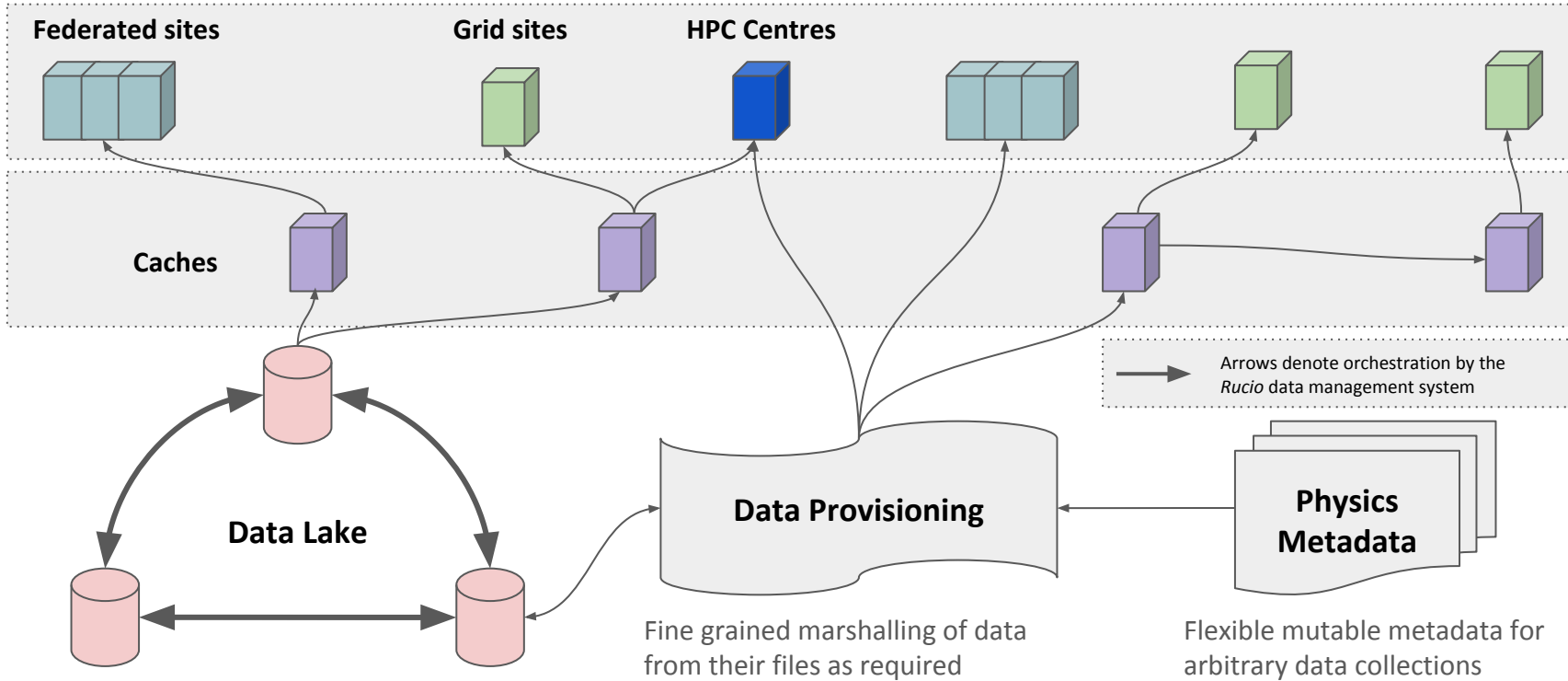
- **Major challenges that will drive data management R&D in the next 5-10 years**

- Factor 5 increase in storage requirements
- Factor 100 increase in number of objects to manage

- **Data management R&D strategy**

- Explore the usage of large reliable data centres ("lakes") with wide area cache control
- Explore fine-grained processing of data collections to better utilise heterogeneous compute and associated networks
- Stronger integration of software (file formats, new IO patterns, interactive analysis) with distributed computing

Fine-grained workflows with streaming delivery



How to move forward

- Tackling the data management challenge can only be done as a community effort
 - With many involved entities such as CERN, WLCG, HEP Software Foundation, + the experiments!
 - We expect many new developments, from cataloging, over networks, to direct IO
- Rucio, the ATLAS data management system, is getting widespread interest
 - Adopted already by AMS and Xenon1t
 - Evaluations started by experiments from HEP (e.g., CMS), atmospheric (IceCube), astronomy (e.g. LIGO), and facilities (e.g., OSG)
 - CERN EP is directly responsible for the coordination and further development of Rucio
 - Community Workshop this week! <https://indico.cern.ch/event/676472/>
- We are in the fortunate position that Rucio is ready for LHC Run-3 scale
- This gives us the chance to start with R&D now to achieve the [strategic plan](#)