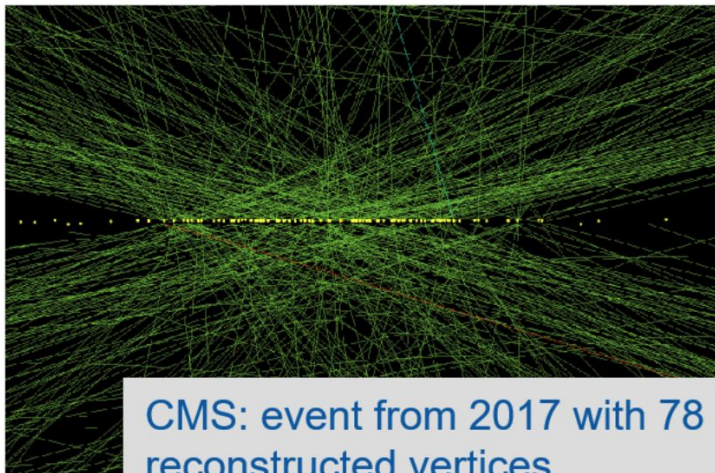# HL-LHC Analysis Model and Tools
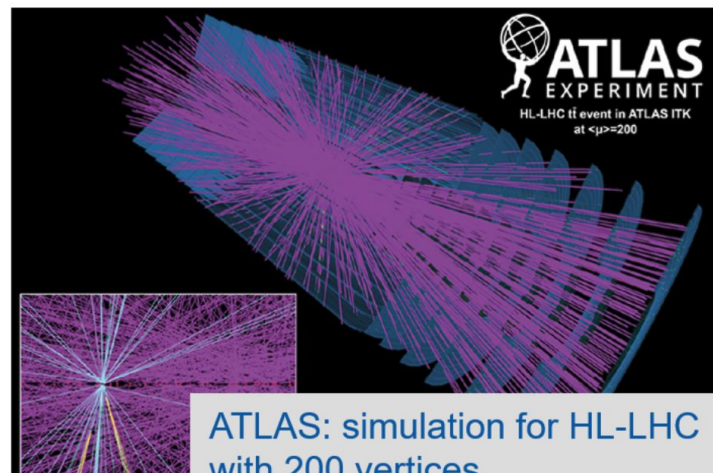
Ian Bird, Simone Campana, Aristeidis Fkiaras

CERN, Geneva, Switzerland

Groningen, 17 April 2019
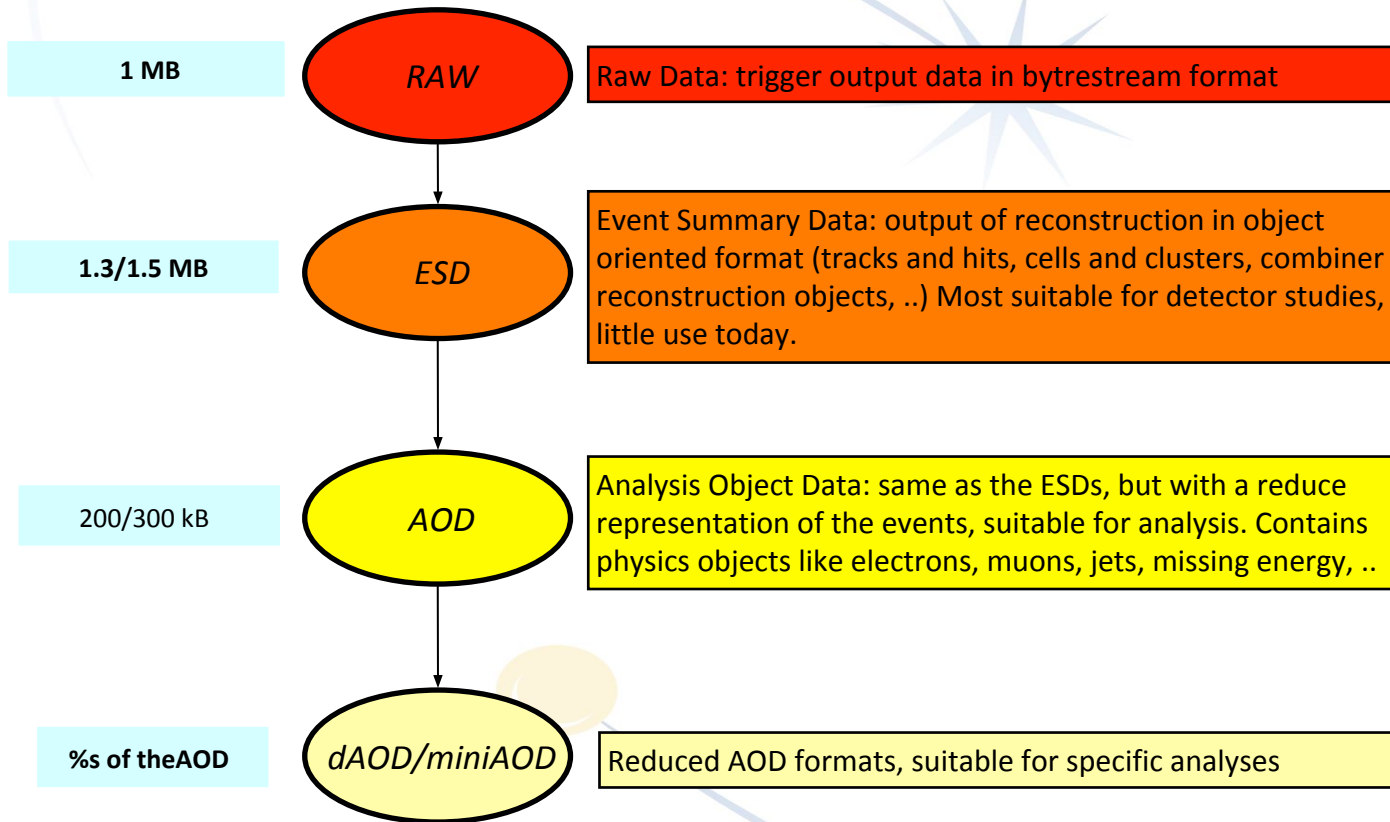
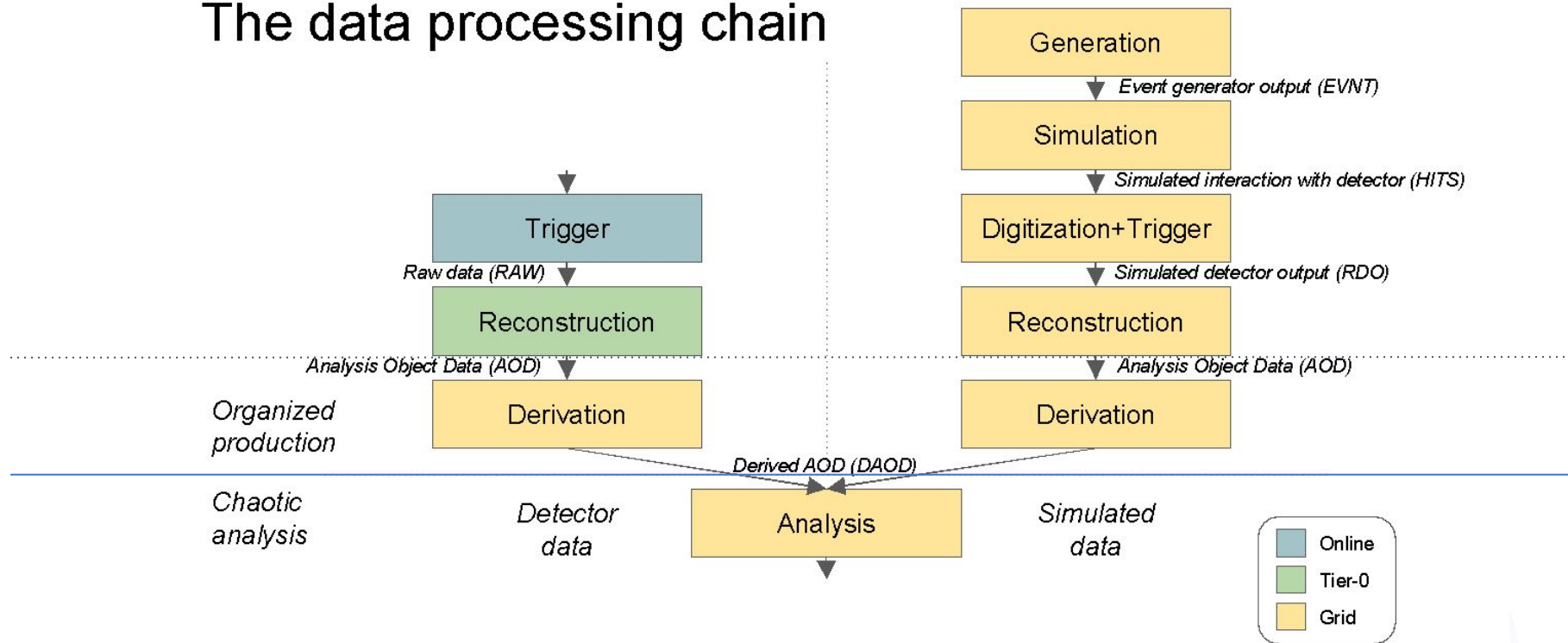CMS: event from 2017 with 78 reconstructed vertices

ATLAS: simulation for HL-LHC with 200 vertices

Data Model

1 MB	*RAW*	Raw Data: trigger output data in bytrestream format
1.3/1.5 MB	*ESD*	Event Summary Data: output of reconstruction in object oriented format (tracks and hits, cells and clusters, combiner reconstruction objects, ..) Most suitable for detector studies, little use today.
200/300 kB	*AOD*	Analysis Object Data: same as the ESDs, but with a reduce representation of the events, suitable for analysis. Contains physics objects like electrons, muons, jets, missing energy, ..
%s of theAOD	*dAOD/miniAOD*	Reduced AOD formats, suitable for specific analyses

Aristeidis.Fkiaras@cern.ch

The data processing chain

The data management Stack

- High Level Data management System
  - Organizes (catalog) and manages (transfer, delete, lookup) the data

- Middleware: implements functionalities in a common layer

- Storage: different systems based on different technologies

- Protocols:
  - Storage Manager Protocol (SRM)
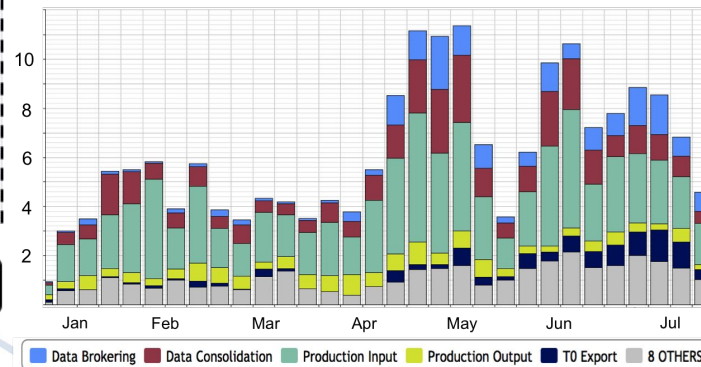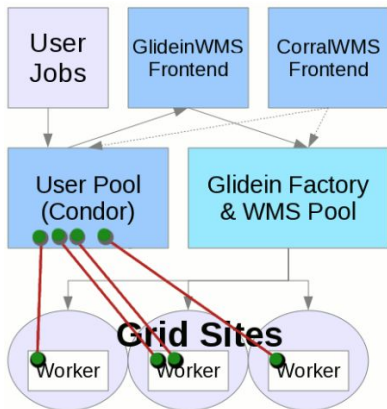 - Transfer and Access protocols (HTTP/xrootd/gridFTP/..)

Data Management System

Middleware (e.g. FTS)

SRM

HTTP
xrootD
gridFTP

Storage

SRM

HTTP
xrootD
gridFTP

Storage
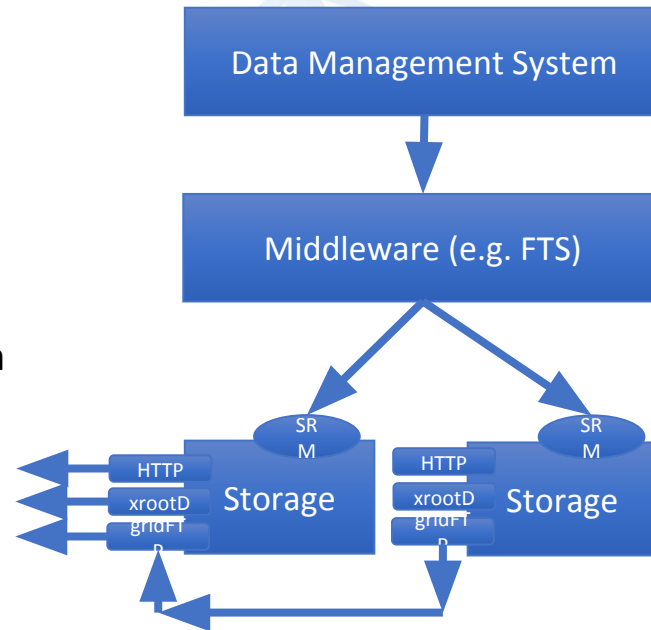
Distributed Data Management: Rucio
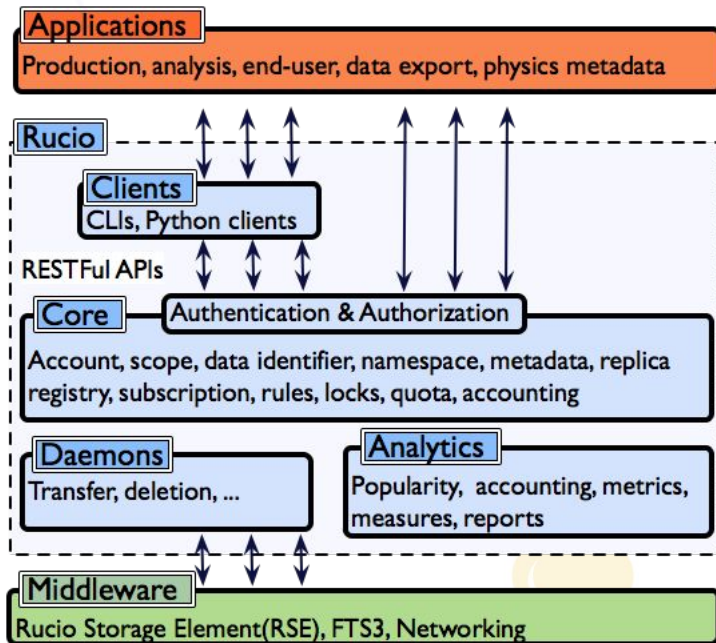
- Organize data in a hierarchical structure (datasets)
- Enable data lookup, access, replication, deletion
- Ensure data integrity and consistency

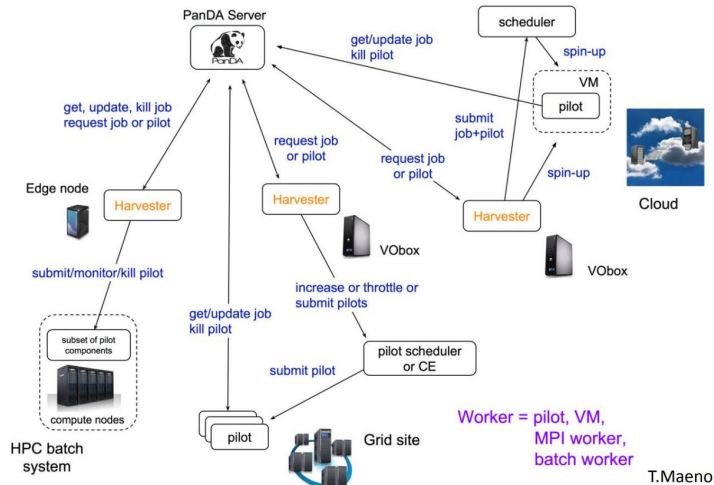Data transfer volume per week by activity (PB)
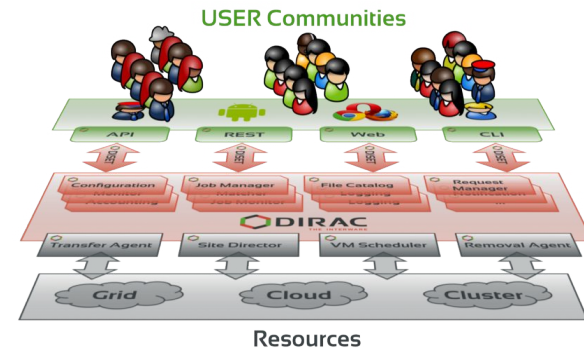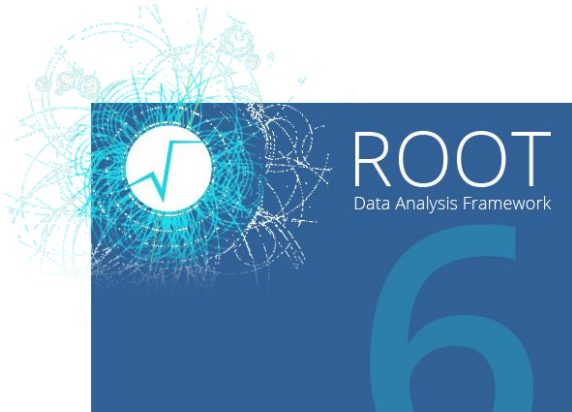
Workflow Management Systems

GlideinWMS

Panda

Dirac

Aristeidis.Fkiaras@cern.ch

Chaotic Analysis and Analysis Preservation

17 April 2019
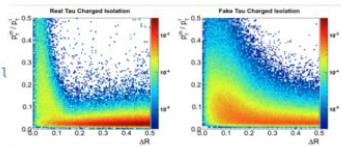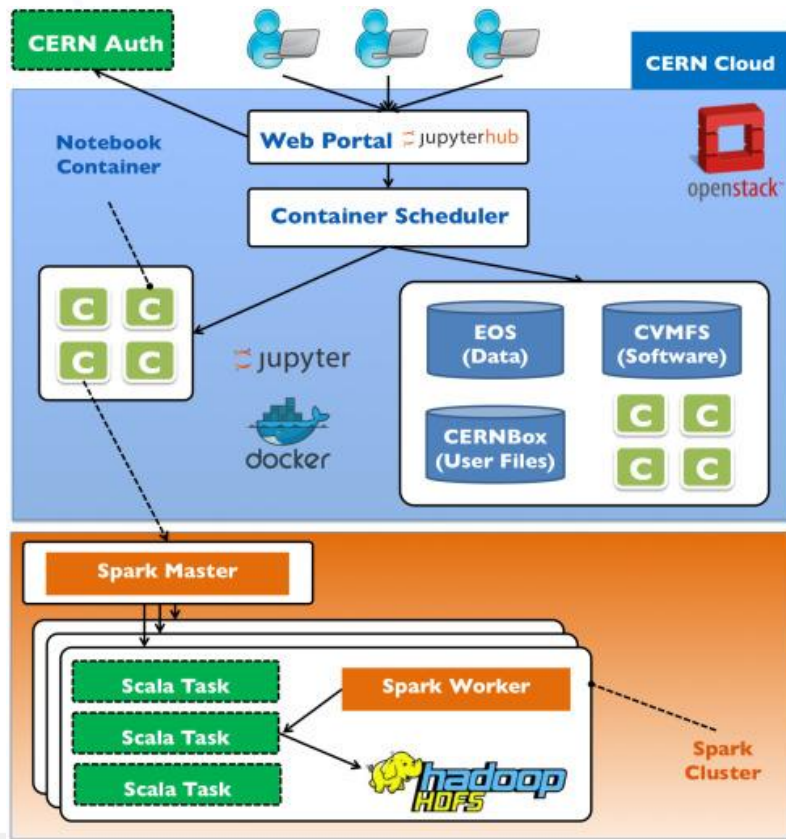
Aristeidis.Fkiaras@cern.ch

ROOT

- Modular scientific software toolkit for big data processing, statistical analysis, visualisation and storage

- C++, Python bindings

SWAN

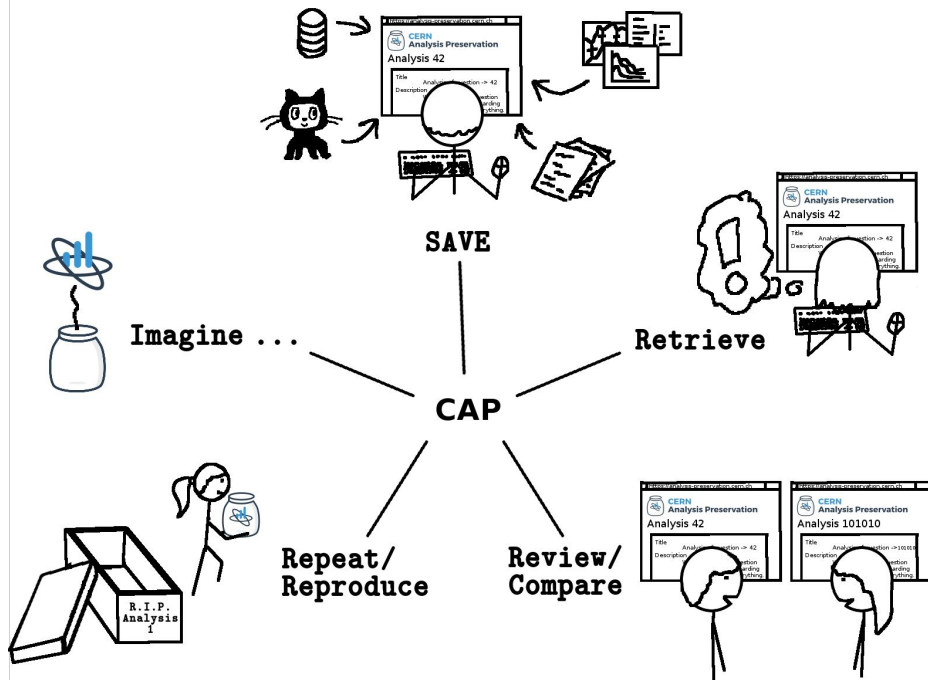- Web-based analysis facility via notebooks
- Access experiments' and user data in the CERN cloud

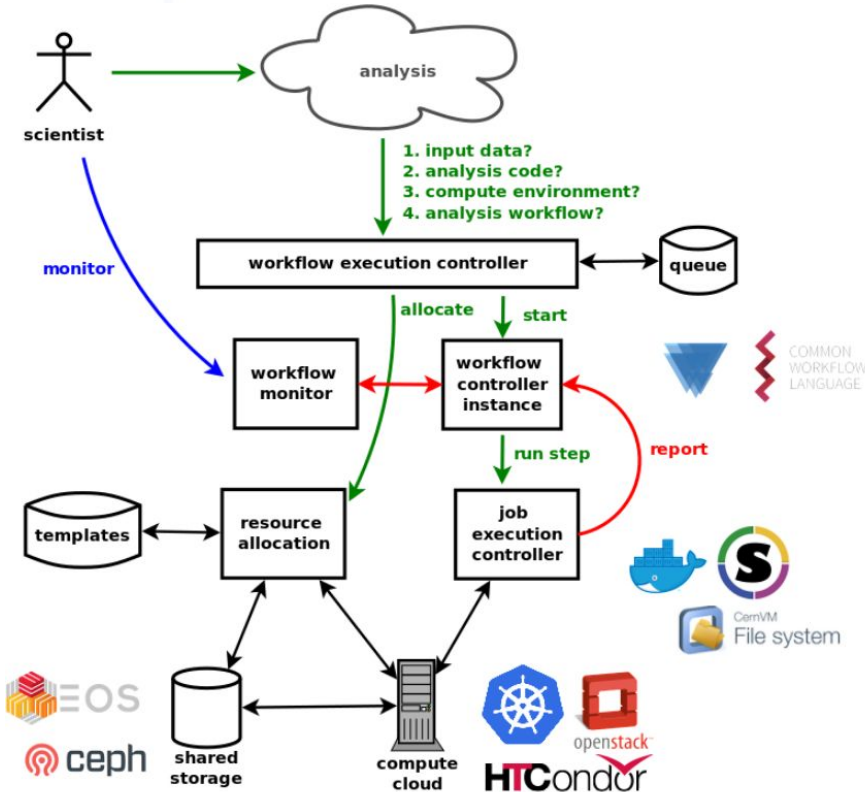17 April 2019 Aristeidis.Fkiaras@cern.ch

CAP (CERN Analysis Preservation)

- Preserve information relevant to analysis so that it remains understandable and reusable

- Collaboration aware Access Control List

Reana

- Instantiate computational workflows on remote clouds
- Rerun analyses with modified input data, parameters or code

Thank you!

Comments?

Questions?