CompF4 Analysis Facilities Summary

Mark Neubauer, Oksana Shadura, Gordon Watts

Scope/Definition of the Topic

Analysis facilities

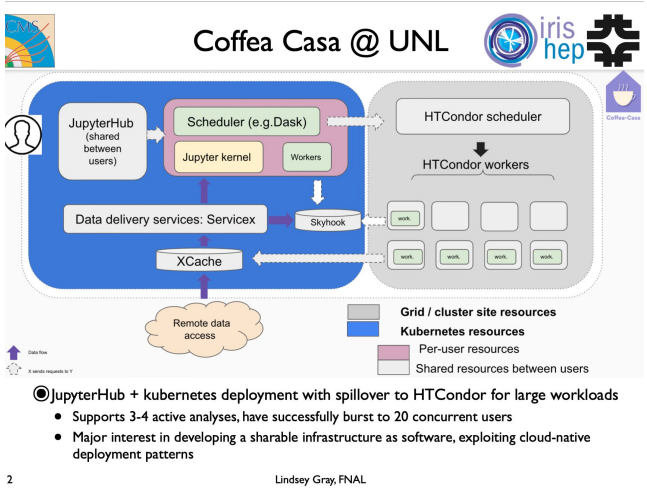
The infrastructure and services that provide integrated data, software and computational resources to execute one or more elements of an analysis workflow. These resources are shared among members of a virtual organization and supported by that organization.

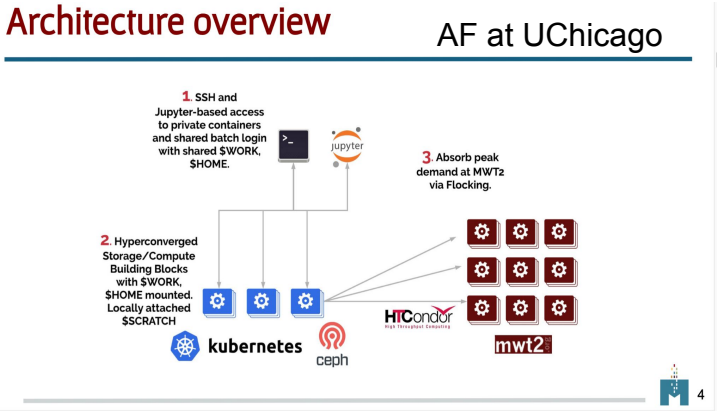
Community Input/Relevant Whitepapers

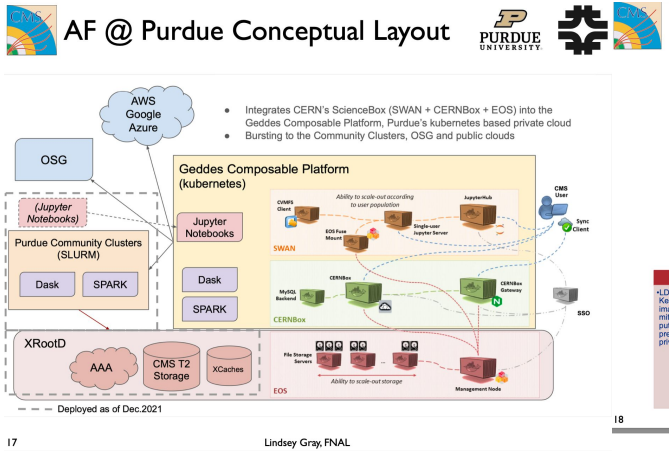
- Doug Benjamin, Kenneth Bloom, Brian Bockelman, Lincoln Bryant, et al. ”Analysis Facilities for HL-LHC“, arXiv:2203.08010 [hep-ex] (pdf).
- Kevin Lannon, Paul Brenner, Mike Hildreth, Kenyi Hurtado Anampa, Alan Malta Rodrigues, Kelci Mohrman, Doug Thain, Benjamin Tovar. ”Analysis Cyberinfrastructure: Challenges and Opportunities“, arXiv:2203.08811 [physics.data-an] (pdf).
- Maria Acosta Flechas, Garhan Attebury, Kenneth Bloom, et al. ”Collaborative Computing Support for Analysis Facilities Exploiting Software as Infrastructure Techniques”, arXiv:2203.10161 [physics.data-an] (pdf).
- And relevant LOI submitted before.

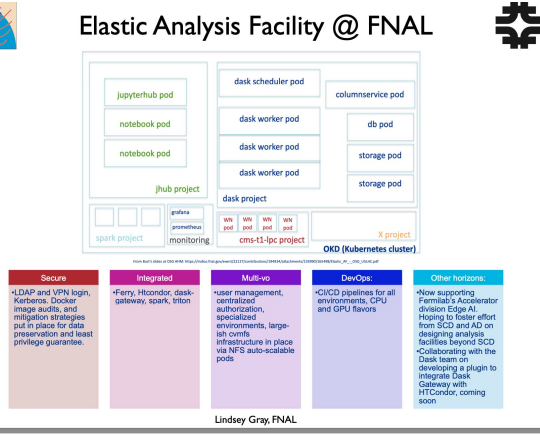
Analysis Facilities

- Existing traditional user facilities (CERN lxplus/lxbatch, LPC, BNL, SLAC and others, mostly DOE managed)
- New exciting analysis facilities prototypes









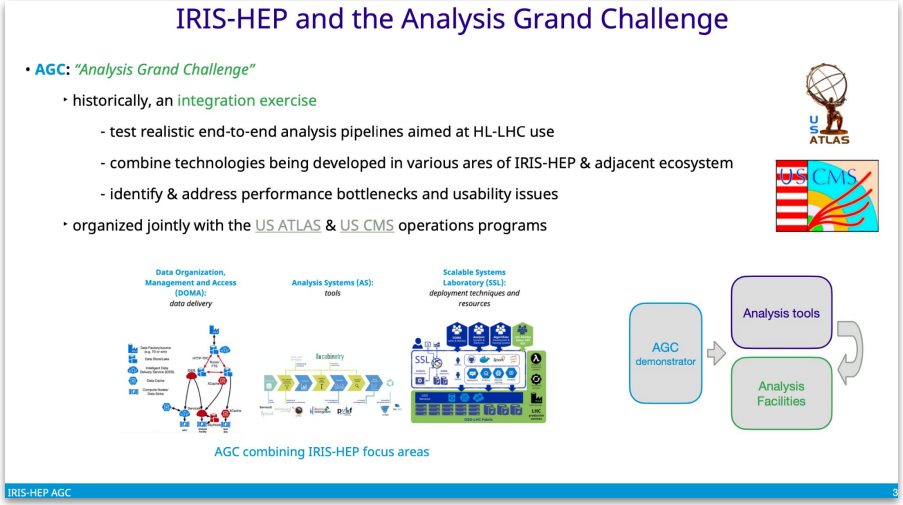
Analysis Cyberinfrastructure: Challenges and Opportunities

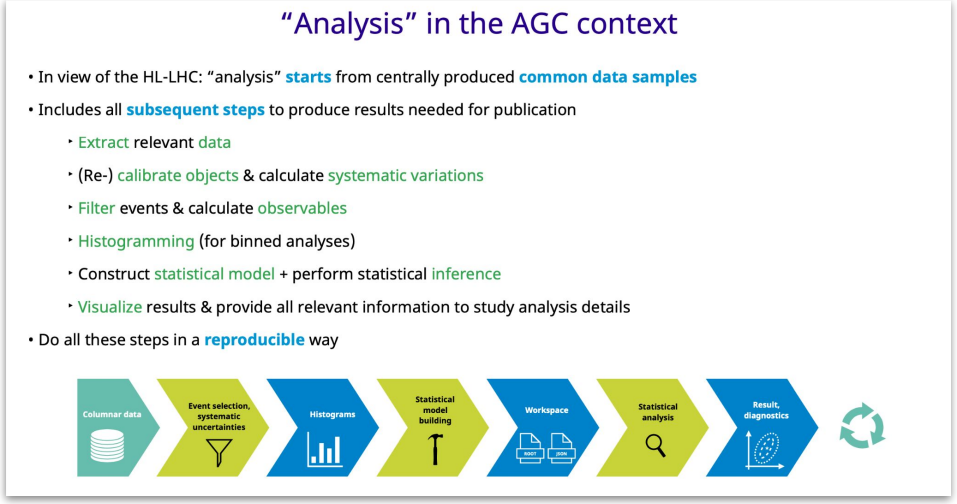
Main points to take in account while building Analysis Facility Infrastructure

- Interoperability between different analysis facilities
- Identity management integration in Analysis Facilities
- Integration of Data organisation, Management and Access
- Resource sharing and scaling
- Sharing environment between users

Analysis Facilities Challenges for Energy Frontier and Other Frontiers

- For the Snowmass Energy Frontier, the HL-LHC presents significant challenges for the analysis community.
  - One of possible solutions is to develop better integration strategy of software components for analyzing the data as well as the deployment of the analysis software at analysis facilities, such as Analysis Grand Challenge

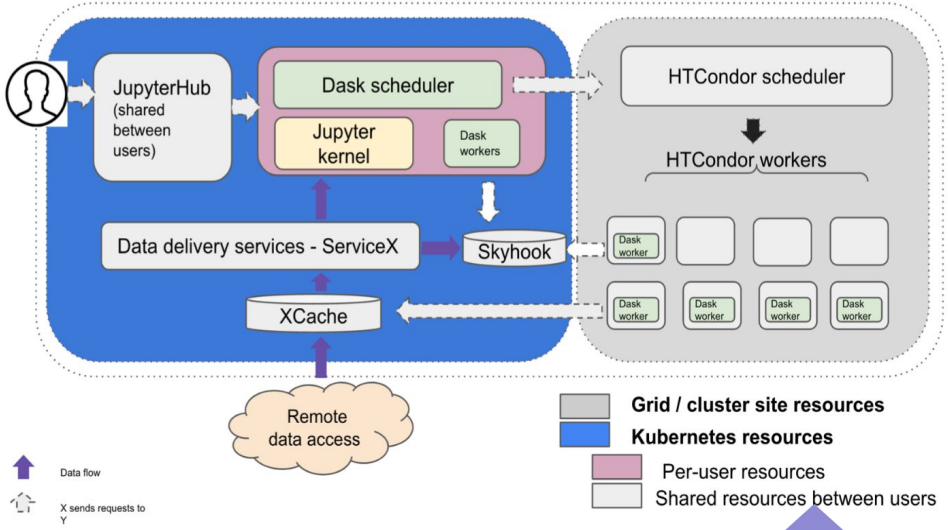




A.Held,O.Shadura, "AGC", ICHEP 2022

Analysis Grand Challenge components:

tools and facilities where AGC will be executed



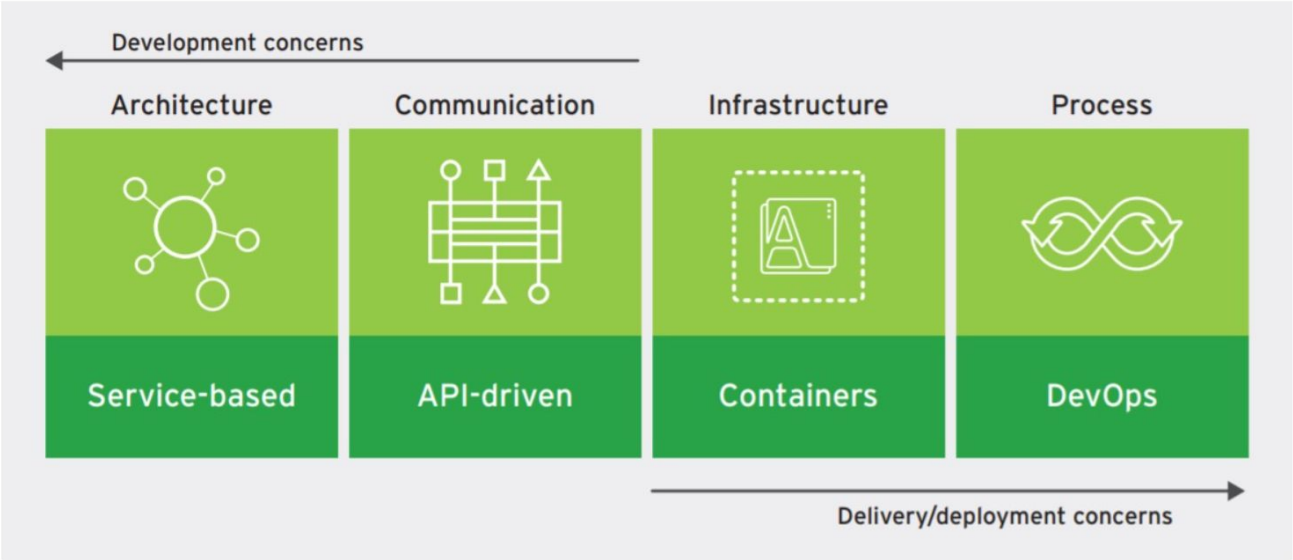
AGC analysis tools
(or any other)

Coffea-casa Analysis Facility
(or other facilities)

Coffea-Casa

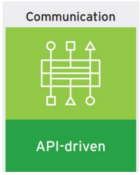
A.Held,O.Shadura, “AGC”, ICHEP 2022

Leveraging Cloud Native for designing, building and operating analysis facilities

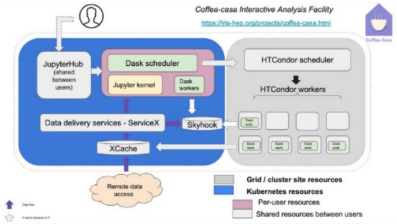


Cloud-native technologies empower organizations to build and run scalable applications in modern, dynamic environments *such as public, private, and hybrid clouds. Containers, service meshes, microservices, immutable infrastructure, and declarative APIs exemplify this approach.*

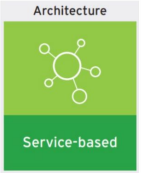
Leveraging Cloud Native: API-driven Communication



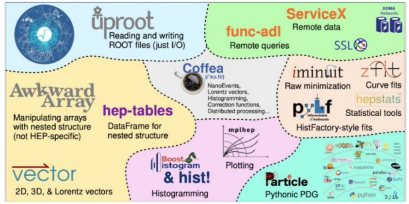
- APIs connect different applications or services together.
- Encourages application architectures that are highly distributed and modular.
- Prevents obsolescence
- Provides easy integration paths with existing systems



Leveraging Cloud Native: Service-based Architecture



- Diverse analysis requires diverse technology ecosystems
- Break up monolithic data structures and tightly coupled applications into loosely coupled and independently deployable smaller components, or services
- Modernize and optimize traditionally complex analysis workflows



Maria Acosta, **Snowmass CompF4 Topical Group Workshop**

Leveraging Cloud Native: Container Infrastructure

Containers provide flexibility, portability and isolation without the additional overhead of virtual machines

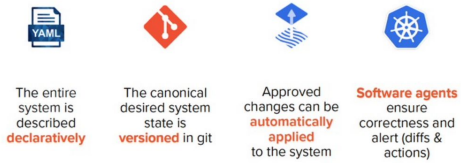
Container orchestration defines tools that automate the deployment, management, scaling, networking, and availability of container-based applications.

- Provides a unified declarative description and configuration language.
- Implements service discovery, load balancing, automated rollouts and rollbacks, plus other features key to providing stable services.
- Multiple types of container orchestration systems, institution needs to assess requirements and needs

argo GitHub HARBOR HELM okd

Leveraging Cloud Native: DevOps/GitOps processes

- Takes advantage of a declarative system to manage the configuration and operations of every element of the platform, from the infrastructure through to the applications.
- Provides observability and control - ensuring that the platform is reliable and operable.



Principles of GitOps: https://gitops-community.github.io/kit/

Recommendations/Research Directions

- We suggest **to work on prototyping of Analysis Facilities together with analysis software developers**, **resource providers and analysis facility architects** using modern techniques, such as exploring concepts:
  - ”Infrastructure as Code";
  - the integration of “federated identities";
  - to facilitate the preservation of user environments and many others.

- We also recommend that Analysis facilities be made interoperable, allowing users to navigate seamlessly from one Analysis Facility to another, and easily extensible, to accommodate future needs without disruptions.