Analysis Facilities

Un bref survol

L. Duflot

Introduction

- HSF a récemment créé le groupe Analysis Facilities Forum dédié à l'implémentation des AF
  - Solutions techniques
  - Liens avec l'environnement data center, cluster, T3, T2, ….
- Je me concentre sur cet aspect des choses mais dans HSF il y a les activités software / framework depuis plus longtemps
  - Data Analysis Working Group
  - Data Analysis Ecosystem Workshops mai 2022
- Dans la pratique, l'implémentation matérielle doit être optimisée pour les solutions software et in fine pour répondre aux besoins des utilisateurs
- Mon but : faire une introductions aux développements récents sur le sujet (basée sur le kick-off meeting et réunions suivantes du groupe)
- *Mes opinions personnelles – italique + vert*

What is an Analysis Facility?

"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."

People

- Dedicated support staff
- Maintenance and development

Services

- Access to experimental data
- Storage space for per-group or per-user data
- Access to significant computing resources

Hardware

- CPUs and disks
- Growing need for GPUs

Software

- ROOT
- Python-based ecosystem
- Interactivity

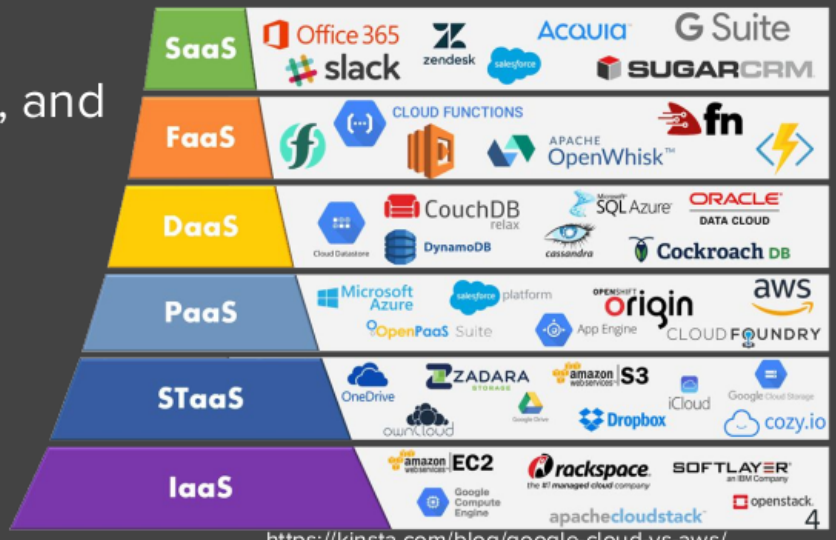
9

A. Forti

Why is this important now?

Technologies evolution

- New technology developments in DOMA areas
  - Development of innovative data delivery services
  - Advanced caching techniques
- Switching to use bearer token eg. SciTokens authorisation
- Integrating different types of resources/"aaS" available on demand: public and private clouds, and Kubernetes (k8s) with increased analysis containerization
- Evaluation of successful AF characteristics



https://kinsta.com/blog/google-cloud-vs-aws/

*DOMA: Data Organization, Management and Access

A. Forti

Why is this important now?

New analysis techniques

- Development of new analysis workflows for more efficient analysis
  - Columnar analysis instead of traditional event loop approach
  - Workflow management tools
  - Increased use of alternative hardware
- Increased interoperability between ROOT and Python data-science tools
- Integration and adoption of industry tools in HEP analysis frameworks
  - Including machine/deep learning algorithms and sophisticated analytics engines like Apache Spark, DASK
- New user interfaces for interactive analysis (e.g. Jupyter notebooks)
- Analysis preservation - REANA platform

5

A. Forti

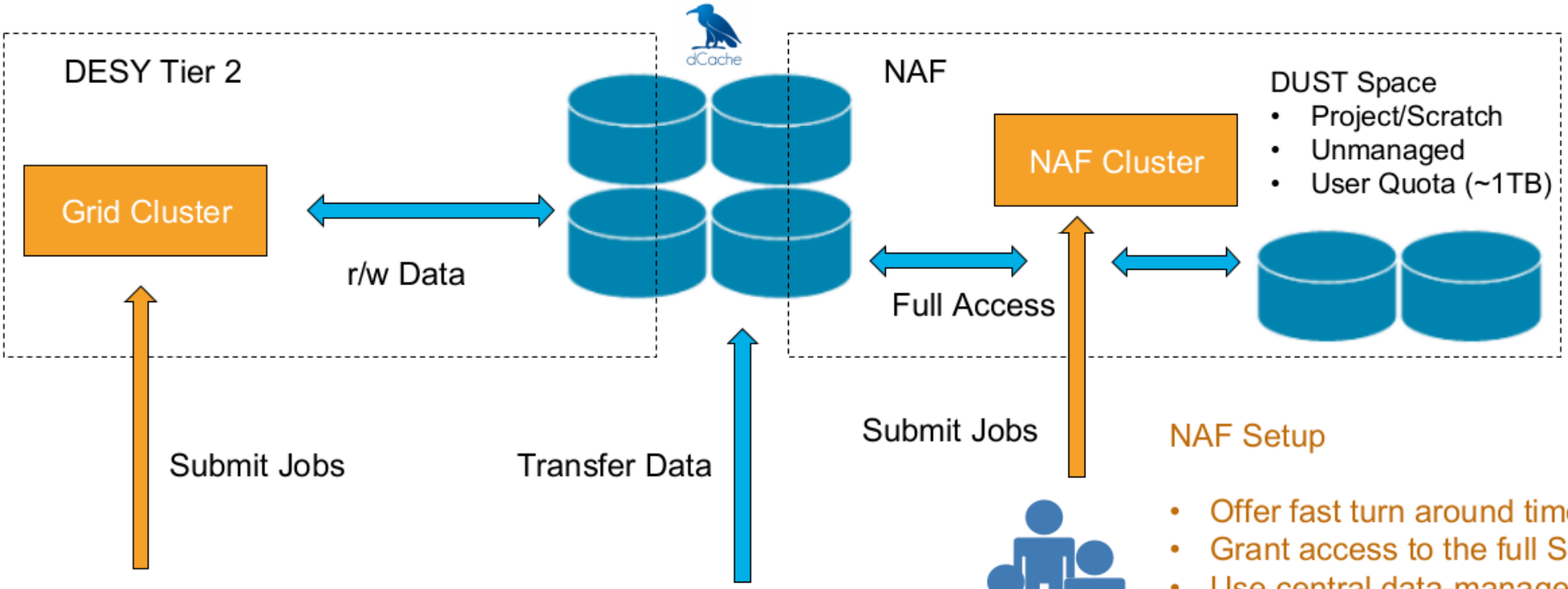
Depuis longtemps….

- Cela fait longtemps qu'il a des AF, elles sont traditionnellement basées sur
  - Login sur des front-ends
  - Service de stockage
  - Batch system
- Il existe encore de nombeuses AF basées sur ce principe :
  - lxplus, CC-IN2P3 AF, NAF @ DESY, BNL, SLAC, …. beaucoup de clusters de labos / universités
  - Peuvent différer par l'accès au stockage
    - Purement local ou accès à des SE Grille comme les LOCALGROUPDISK pour ATLAS
  - Parfois un front-end via Jupyter notebook
- *Perso : même si les étudiants qui arrivent sont plus familiers avec les Jupyter notebook, je n'y vois pas un vrai avantage par rapport au login sur un front-end*

Paradigm: HEP Analyses are Data Driven

As Underlying Principle of the NAF

- Almost all HEP data analyses require access to large amounts of data



NAF Setup

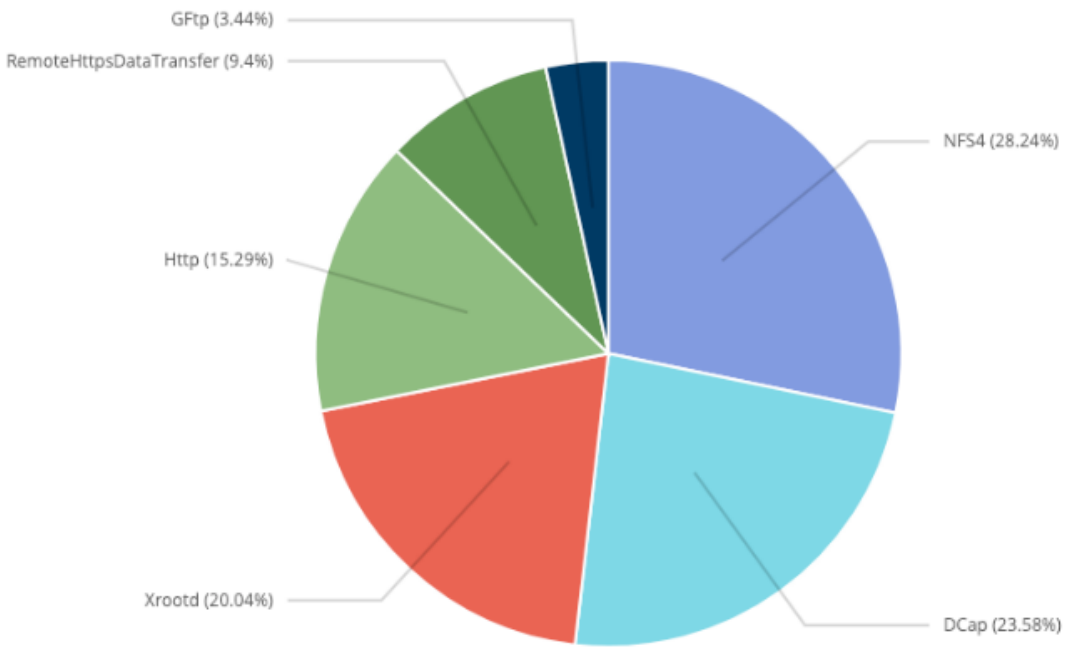
- Offer fast turn around times
- Grant access to the full SEs
- Use central data-management of experiments to store data at DESY

C. Voß

Data Access

Protocol and Patterns

- Support all commonly available protocols for HEP
  - WebDav, XrootD & gsi-FTP, dCap as legacy
- Rely heavily on POSIX-like access
  - Storage infrastructure based on dCache and GPFS
  - Scalable NFS architecture across the NAF
  - Belle II group uses NFS almost to 100%
  - ATLAS users read (Rucio managed) LocalGroupDisk via NFS
  - Smaller groups use tools supporting only local files or industry formats like S3 or Hadoop
- Observed issues with NFS in the past, i.e. hanging clients (esp. SL6) → fixed with help of developers
- NFS makes testing of new tools very easy



C. Voß

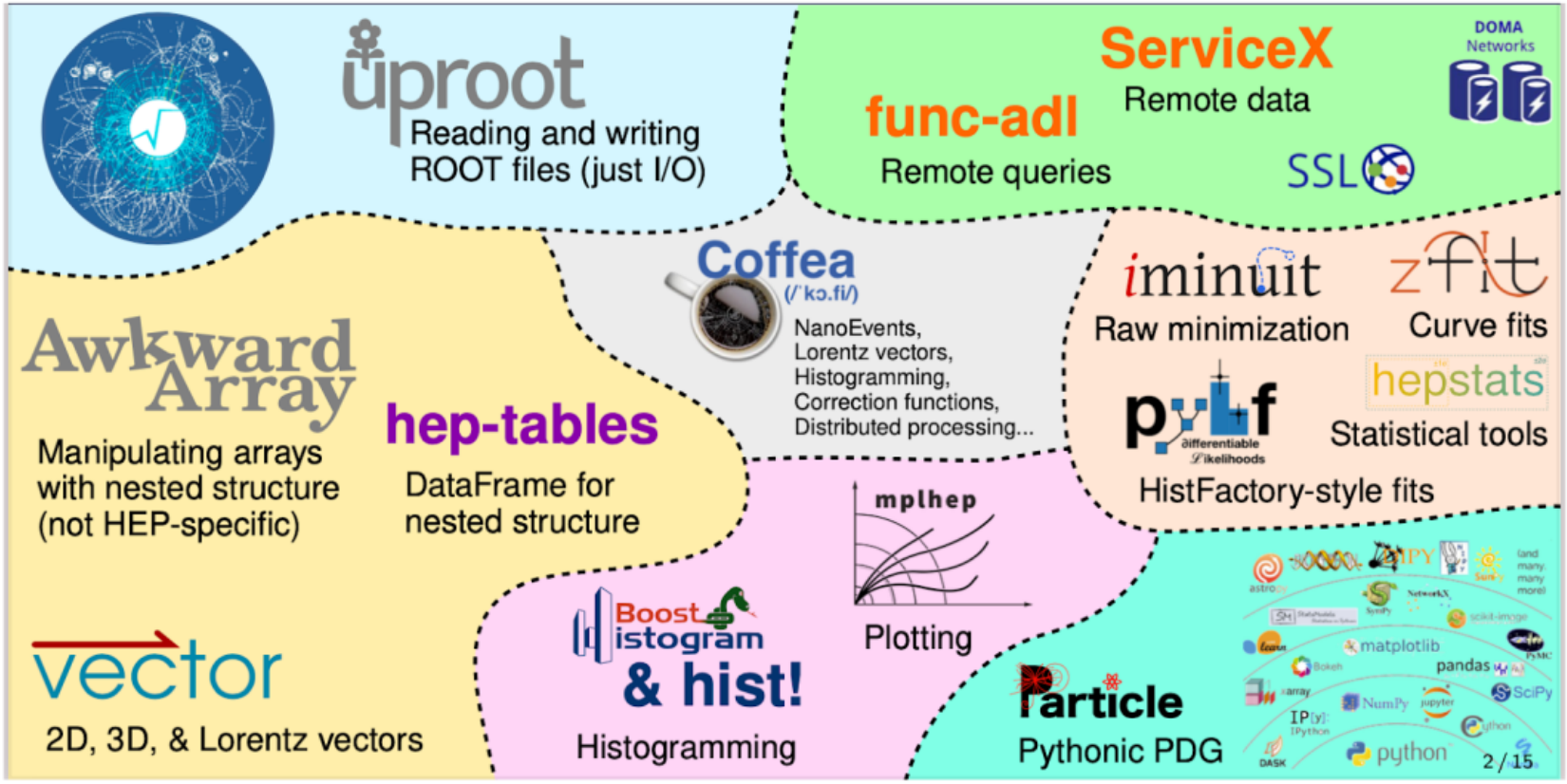
Un pas de plus...

- CERN SWAN :
  - Jupyter + EOS/CERNBOX + batch ou spark
- INFN :
  - JupyterHub / JupyterLab + Dask (vers batch HTC, HPC, T2s) + Xcache
- BNL :
  - Jupyterhub + batch (HTcondor pour HTC, slurm pour HPC)
  - OKD : REANA testbed, serviceX

Évolution de l'écosystème

- Utilisation de plus en plus grande des outils de ML, souvent via python et ses solutions d'analyse
  - ML, dask, scipy, etc...
- Une fraction des physiciens attirés par
  - les écosystèmes d'analyses python
  - Analyse « colonne par colonne » au lieu « d'événement par événement »
  - Description de l'analyse avec le framework qui gère le processing plutôt que d'écrire explicitement une boucle sur les événements
  - *Perso : deux derniers points n'ont pas été démontrés « à l'échelle » encore moins que ce sont de meilleures solutions que les nôtres*
- Projet US IRIS-HEP
- Une vision de quasi-interactivité
  - *Perso : irréaliste*

New Facilities for New Tools



- Generally: AFs provide a curated substrate upon which to easily deploy these (stacks of) tools at scale
  - The exact way in which this service is provided is currently filled with opinion, but there is a large degree of convergent evolution
  - We'll enumerate all the efforts and the directions under study at present
  - Each effort, while similar, does have different foci
- While not in the box above - RDataFrame is within the scope of all AFs discussed in this talk

3

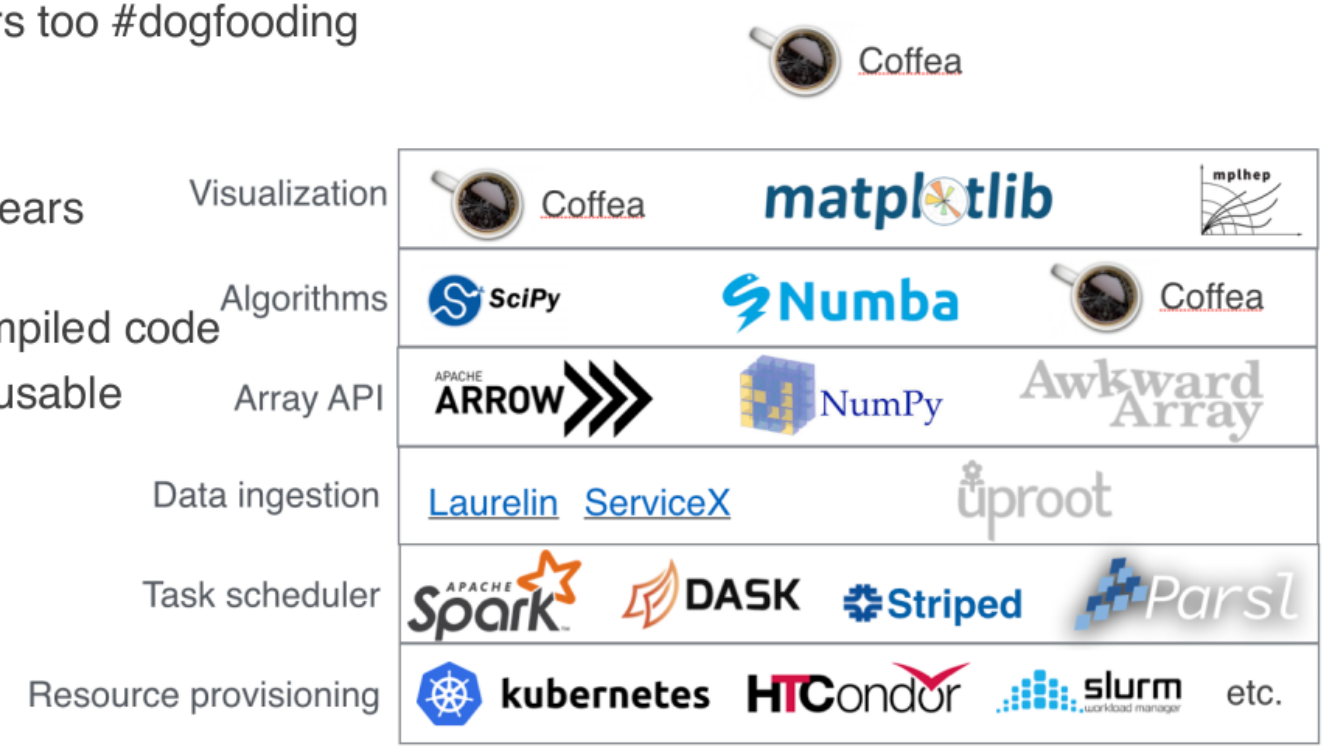
Lindsey Gray, FNAL

What's Coffea?

- A package in the scientific python ecosystem
  - `$ pip install coffea`
- A user interface for columnar analysis
  - With missing pieces of the stack filled in
- A minimum viable product
  - We are data analyzers too #dogfooding
- A really strong glue

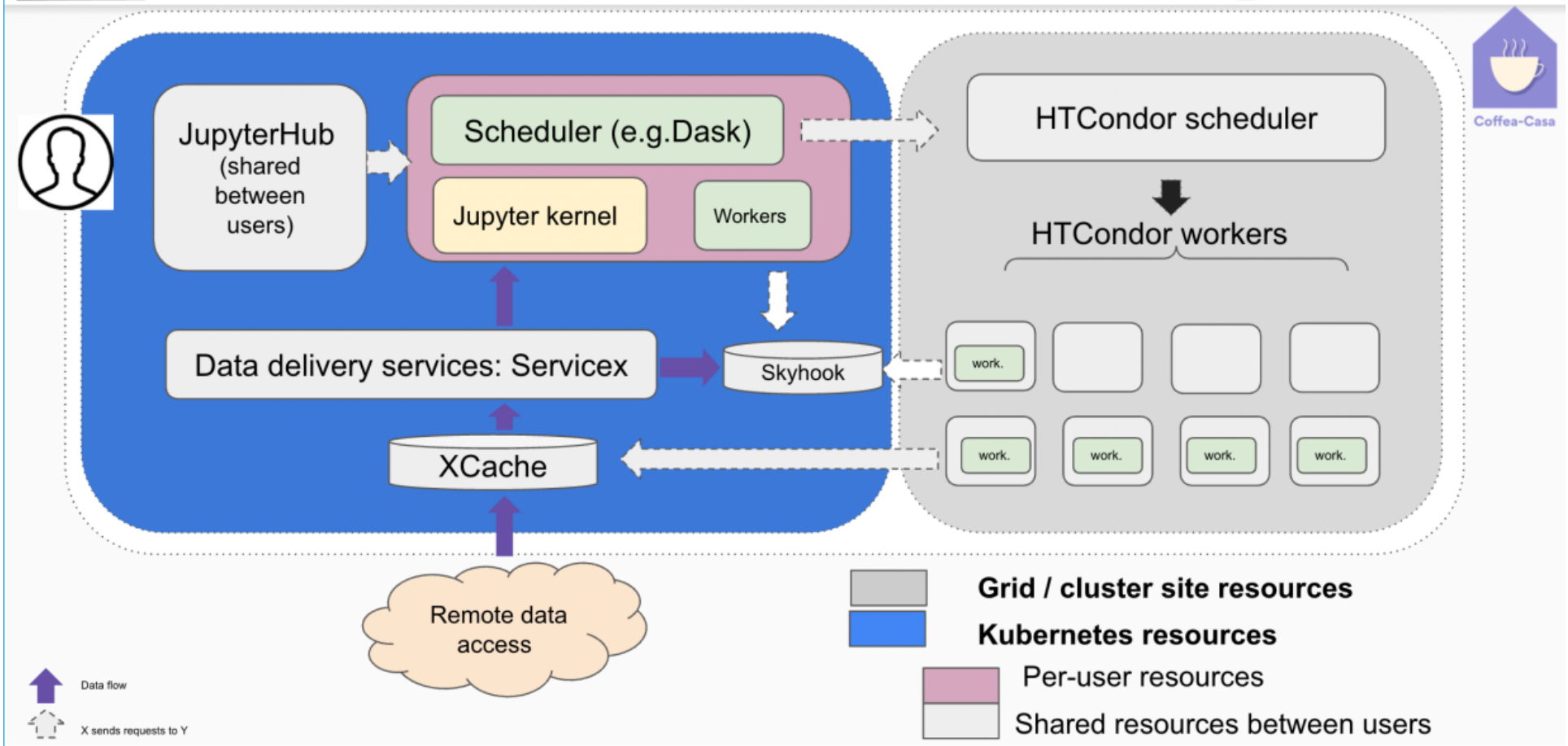
- Going strong for four years

- Roughly as fast as compiled code
  - Significantly more reusable



Lindsey Gray, FNAL

Coffea Casa @ UNL



- JupyterHub + kubernetes deployment with spillover to HTCondor for large workloads
  - Supports 3-4 active analyses, have successfully burst to 20 concurrent users
  - Major interest in developing a sharable infrastructure as software, exploiting cloud-native deployment patterns

Infra similaire mise en place à UChicago

12 Lindsey Gray, FNAL

« new gen AF »

- Coffea-casa @ UNL et U-Chicago, Elastic AF @ FNAL
  - Solutions construites au dessus de kubernetes (services, compute), débordement sur cluster locaux (partagé avec la grille ou standalone, HPC, etc.)
  - Xcache pour les accès distant
  - ServiceX comme service de transformation de données vers un format « colonne » adapté aux outils (en particulier python)
  - Scheduling par dask (par ex)
  - UI via Jupyter

Résumé

- Beaucoup de buzz autour des nouveaux concepts d'AF, de l'utilisation de kubernetes pour construire des services « dynamiques » et des outils software d'analyse « colonne », en bonne partie soutenu par IRIS-HEP
- Cette présentation n'est pas faite pour dire qu'il faut faire la même chose mais
  - Présenter les idées, se préparer à ce qui peut émerger
  - Voir s'il y a des physiciens qui s'intéresse au concepts, la demande doit venir d'eux
-

Backup

Glossary

- DOMA: Data Organization, Management and Access
- WLCG: Worldwide LHC Computing Grid
- HTTP: Hypertext Transfer Protocol. HTTP is the protocol used to transfer data over the web. A typical flow over HTTP involves a client machine making a request to a server, which then sends a response message.
- HTTPS: Hypertext Transfer Protocol Secure - the secure version of HTTP used for secure communication over a network
- SciTokens: The SciTokens project builds a federated ecosystem for authorization on distributed scientific computing infrastructures.
- IAM: Identity and Access Management
- OIDC: OpenID Connect. An authentication protocol which verifies user identity when trying to access a protected HTTPs end point.
- aaS: “as a Service”. Eg. PaaS = Platforms as a Service, SaaS = Software as a Service
- Kubernetes: (k8s) is an open source container orchestration platform that automates many of the manual processes involved in deploying, managing, and scaling containerized applications.
- Apache Spark: Apache Spark is an open-source unified analytics engine for large-scale data processing.
- Dask: flexible library for parallel computing in Python. Similar to Apache Spark but integrates with existing Python tools.
- Ray: Ray is a high-performance distributed execution framework targeted at large-scale machine learning and reinforcement learning applications
- REANA: reusable and reproducible research data analysis platform
- HSF: HEP Software Foundation
- ServiceX: ServiceX is a data extraction and delivery delivery service
- XCache: cache-based data approaches to increase efficiency of CPU use (via reduced latency) and network (reduce WAN traffic)
- Data lake: storage service geographically distributed across large data centers connected by fast network with low latency. Alternative to running jobs at site where files are located.

8

A. Forti

ServiceX

- « ServiceX is a data extraction and delivery delivery service. Users provide a dataset identifier and a selection statement that specifies filters and columns. ServiceX brings up parallel workers to open the files in the dataset and uses experiment approved frameworks to extract the data and store it in popular columnar file formats for easy analysis using familiar tooling. »

Skyhook Data Management

- SkyHookDM « is an extension of the Ceph open source distributed storage system for the scalable storage of tables and for offloading common data management operations on them, including selection, projection, aggregation, and indexing, as well as user-defined functions »