Towards a DataLab as a service

Andrés Heredia Canales (heredia@ifca.unican.es)
Aida Palacio Hoz (aidaph@ifca.unican.es)

(cc) Álvaro López García

Institute of Physics of Cantabria (IFCA) (CSIC-UC)

EGI CONF 2023
Poznań, 21st June, 2023

Motivation

- Each researcher deploys their own platform for analytics
- Lack of time to install the environments: libraries, services and large amount of data
- Need experience to compile some tools due to a hardware and software version dependency
- Becomes difficult to manage by the administrator

Our DataLabaaS solution: hot spots

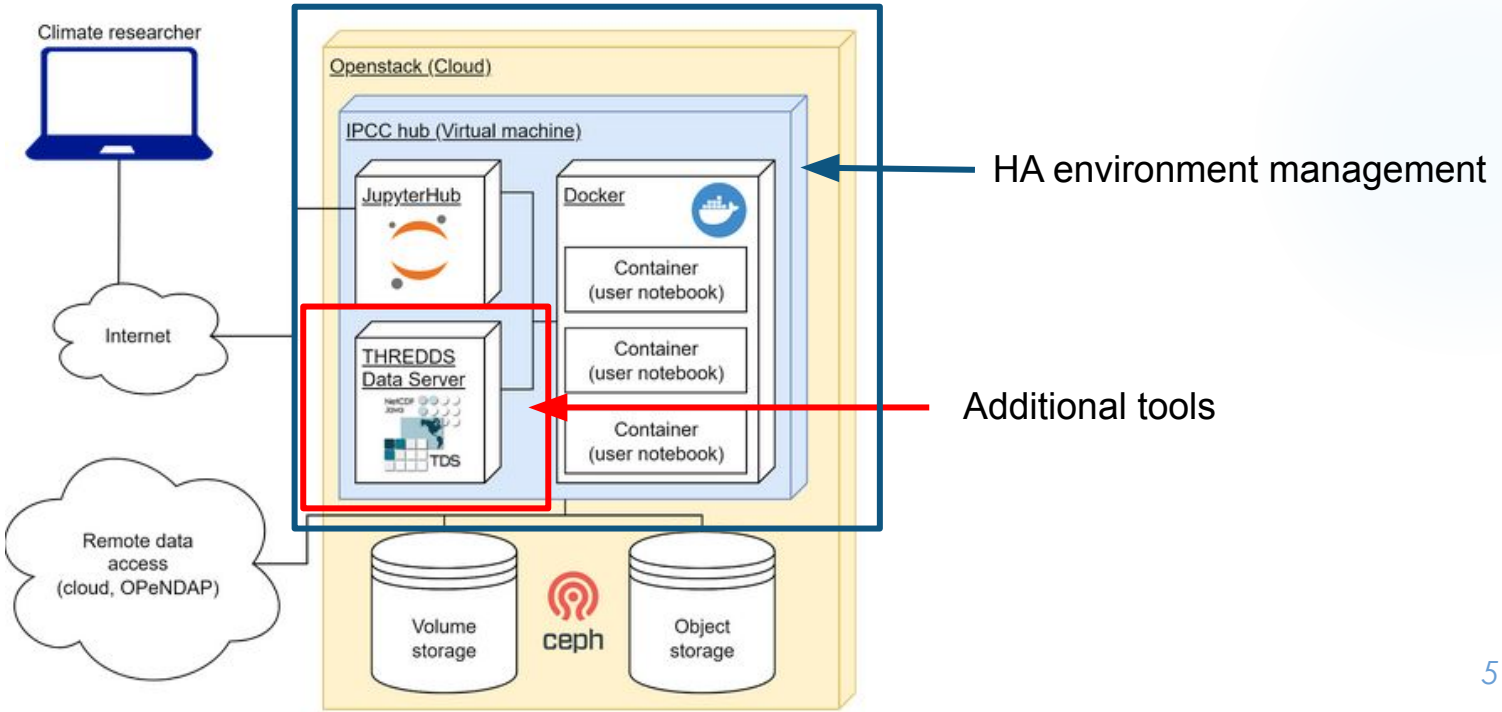
- Merge multiple user-managed platforms for analytics in a **all-in-one platform**
- **High-Availability** platform that provides **computing resources on demand** through a k8s cluster
- **Customizable** user analytics environments
  - Default analytics tools: Jupyter, analytics libraries, etc.
  - **(plus)** Attach additional services
  - **(plus)** Automatic download and access to large amount of data

Examples of use cases I

- Dummy instances
- Intrusion Detection System data
  - Setup of additional services automatically
  - Customize the data access

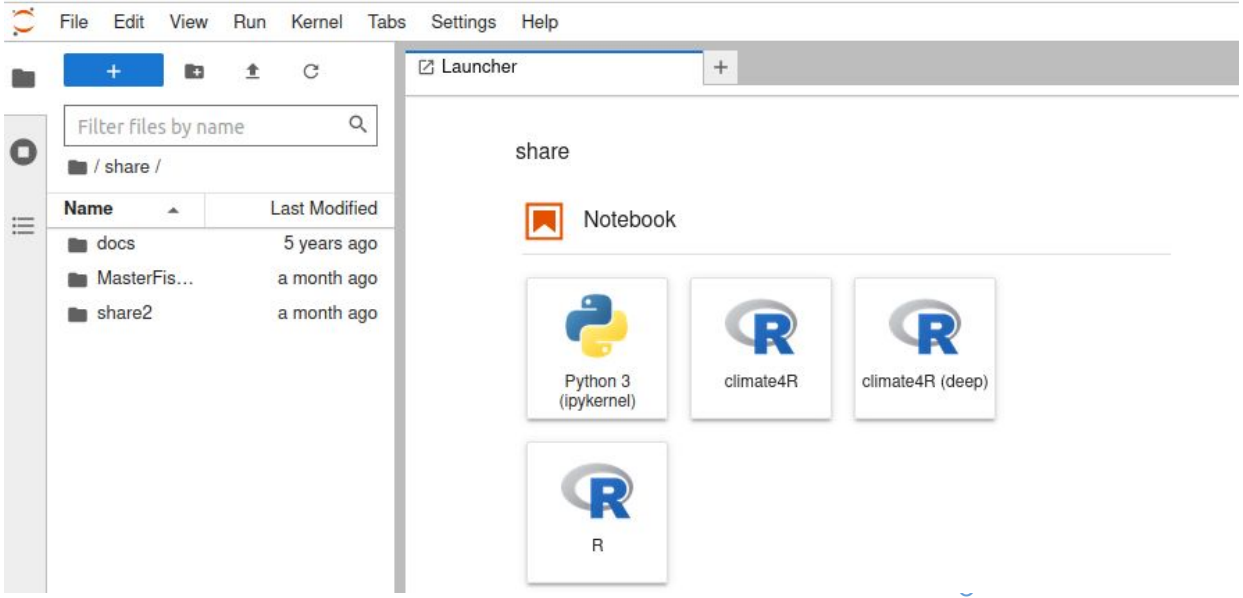
Examples of use cases II

- The Intergovernmental Panel on Climate Change (IPCC)

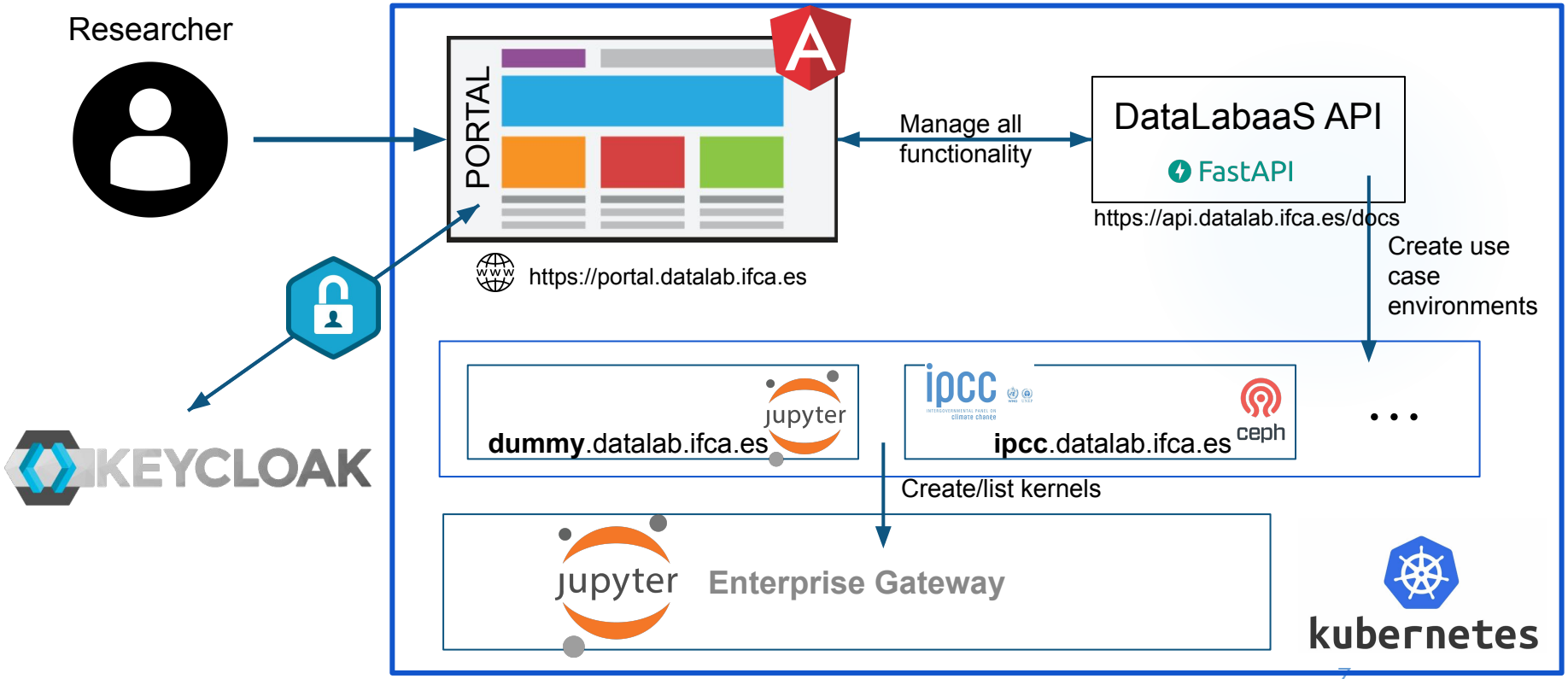


Examples of use cases III

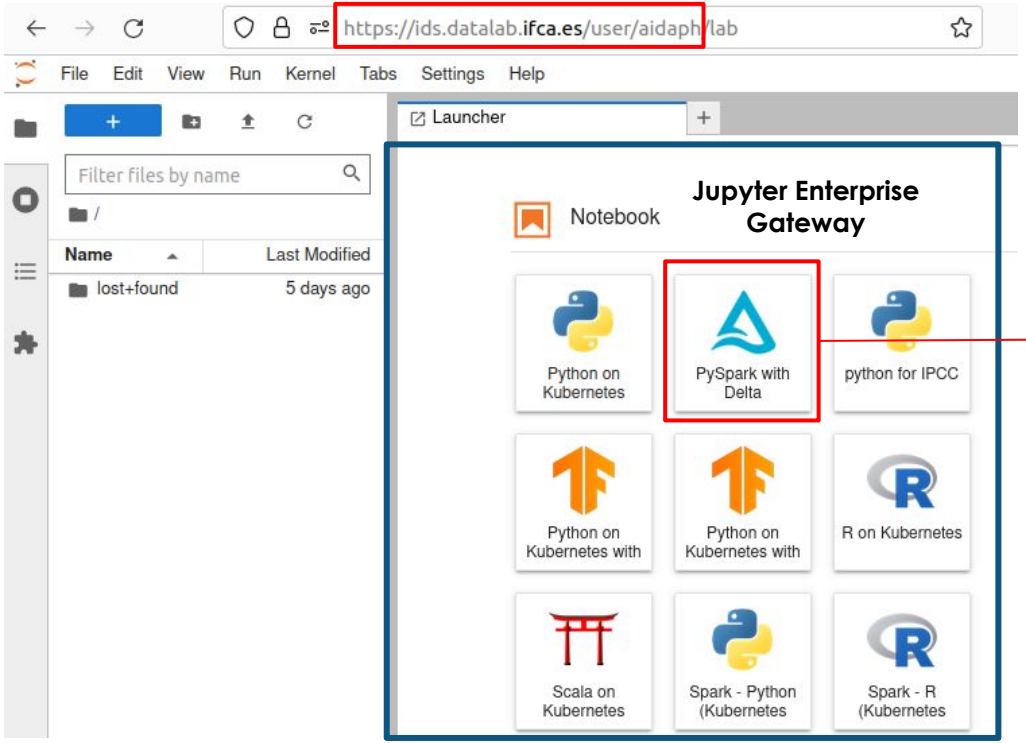
- Master courses - jupyterhub instance based on containers
  - All the containers and kernels running in HA instead of in a single instance



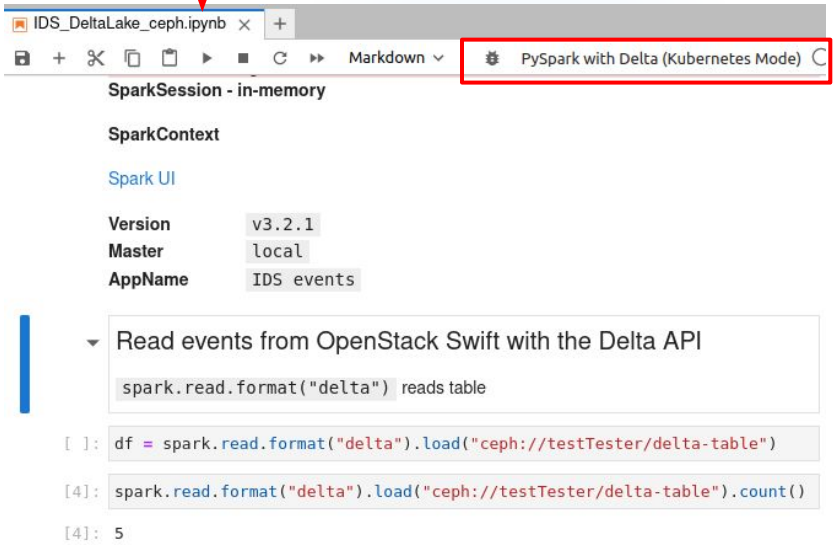
DataLabaaS architecture



Create new Jupyter Notebook (JEG + K8s)



Create new pod in k8s



Future plans

- Support multiple AAI (EGI CheckIn, IFCA SSO, etc)
- Add customizable services to the list of supported ones
- Integrate this service into EOSC

Thanks for your attention!

Towards a DataLab as a service

Andrés Heredia Canales (heredia@ifca.unican.es)
Aida Palacio Hoz (aidaph@ifca.unican.es)

(cc) Álvaro López García

Institute of Physics of Cantabria (IFCA) (CSIC-UC)

EGI CONF 2023
Poznań, 21st June, 2023