Leibniz-Institut für Astrophysik Potsdam

SciTrace

a docker based workflow with FAIR principles

Modern Science requires

Complex environments

Scientific environments (software stack) get more and more complex and less robust.
Such environments are difficult to setup and maintain. They need to be isolated.

CONTAINERIZATION → using container is not an easy task, and using it efficiently is even harder.

Modern Science requires

Complex environments

Scientific environments (software stack) get more and more complex and less robust.
Such environments are difficult to setup and maintain. They need to be isolated.

    CONTAINERIZATION → using container is not an easy task, and using it efficiently is even harder.

Large amount of data (Tera, Peta, Exa)

Most of scientific work now need to deal with large amount of data. These dataset may not be storable on local devices, and need to be stored remotely.

    CLOUD STORAGE → large numbers of Cloud Storage services and as many clients and interfaces.

Modern Science requires

Complex environments

Scientific environments (software stack) get more and more complex and less robust.
Such environments are difficult to setup and maintain. They need to be isolated.

CONTAINERIZATION → using container is not an easy task, and using it efficiently is even harder.

Large amount of data (Tera, Peta, Exa)

Most of scientific work now need to deal with large amount of data. These dataset may not be storable on local devices, and need to be stored remotely.

CLOUD STORAGE → large numbers of Cloud Storage services and as many clients and interfaces.

Complex hardware infrastructure (InfiniBand, GPUs, RAM, CPUs…)

Complex reduction, analysis or computations need High Performance Computing infrastructure and resources. Which are clearly not available on a scientists laptop.

CLOUD COMPUTING → using Kubernetes clusters requires to overcome a stiff learning curve.

SciTrace helps!

Containerization

Users provide the minimum: code, parameters and installation script

SciTrace builds efficiently your environment

Cloud Storage

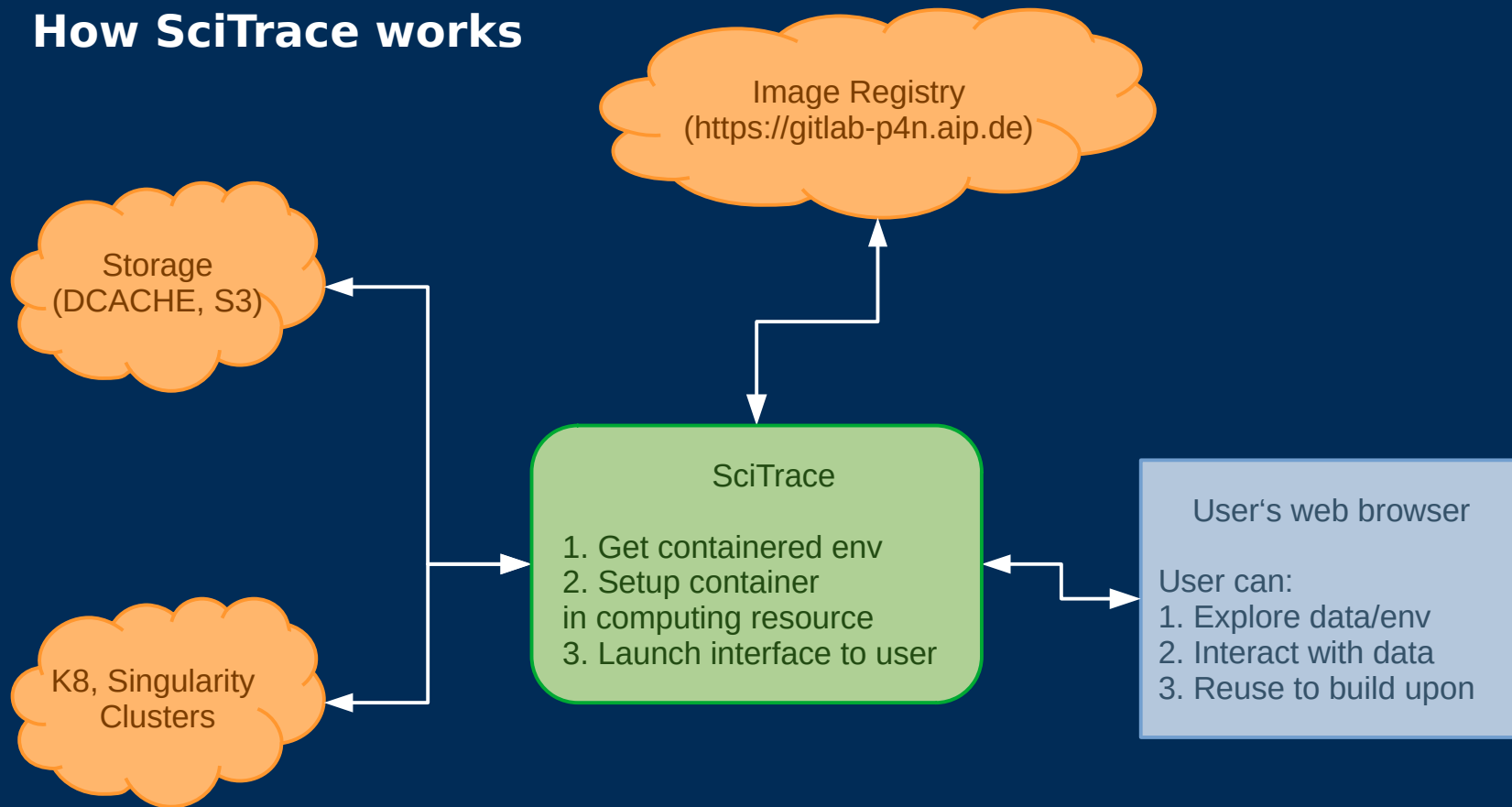Users use the preset settings of SciTrace *(p4n DCACHE, S3-aip already setup)*

SciTrace makes your data accessible to your containers transparently

Cloud Computing

Users request the resources they need

SciTrace setups your containers with best settings for your clusters

How SciTrace works

Image Registry
(https://gitlab-p4n.aip.de)

Storage
(DCACHE, S3)

K8, Singularity
Clusters

SciTrace

1. Get containered env
2. Setup container
in computing resource
3. Launch interface to user

User's web browser

User can:
1. Explore data/env
2. Interact with data
3. Reuse to build upon

SciTrace Demo