

Michael Schuh, DESY
Mai 28 2019
PaNOSC WP6

[Infrastructure]
[Platform]
[Function]-as-a-Service
@ DESY OpenStack Cloud



EOSCpilot
The European Open Science
Cloud for Research Pilot Project
www.eoscipilot.eu



Cloud Computing
Container Orchestration
Software Defined Networking
Infrastructure as code



kubernetes

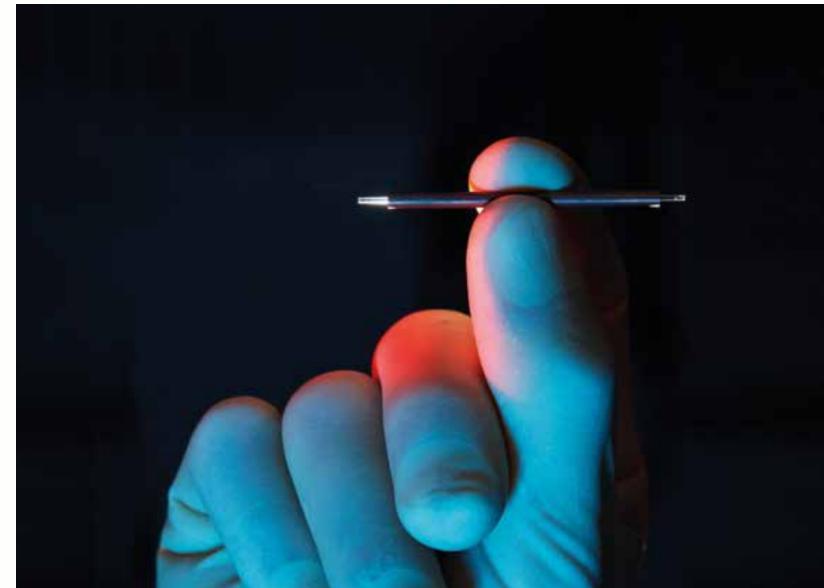
 **openstack.**[®]

Science Notebooks
Function as a service
Event-driven computing



Peta-scale distributed storage
Storage events
Event stream API





DESY Deutsches Elektronen-SYnchrotron (German Electron-Synchrotron)

- Physics with Photons, Free Electron Lasers
- Accelerator technologies
- Experimental particle physics
- Astroparticle physics

Images: http://www.desy.de/femto_eng/index_eng.html

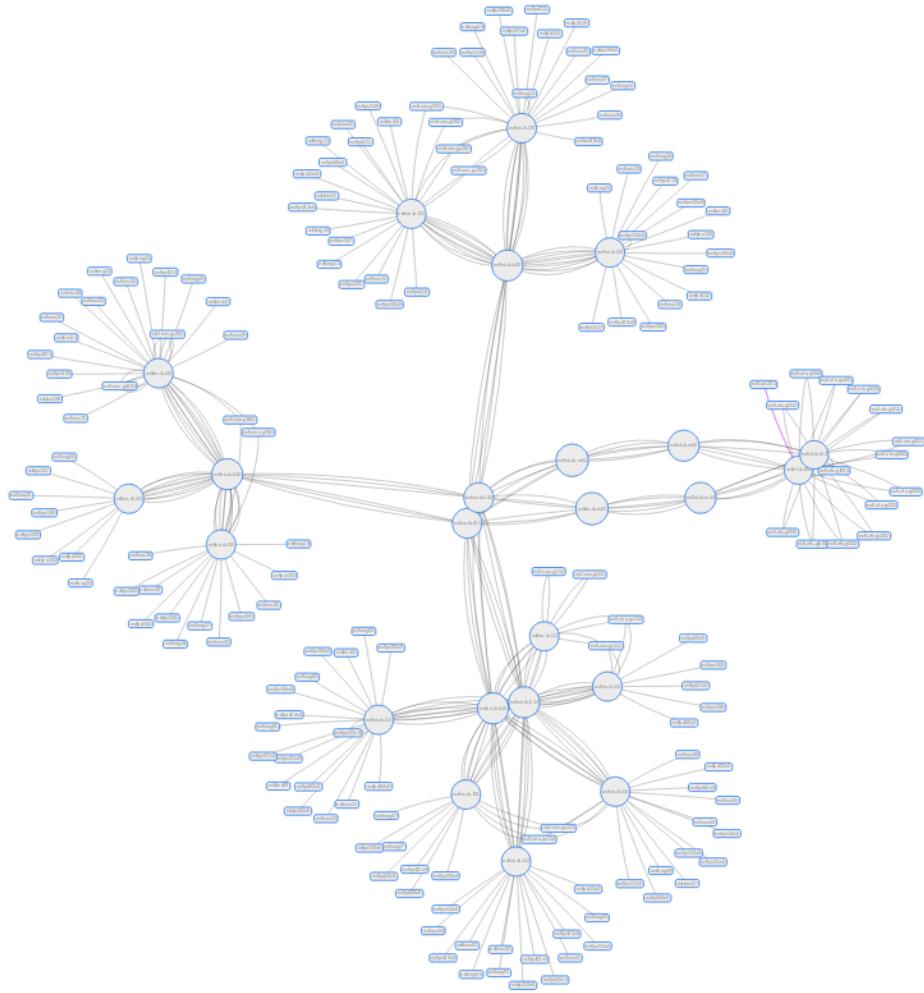


- Scientific Computing, HPC, HTC, Grid and Cloud
- LHC Tier-2 center, large scale storage and archiving
- dCache – a peta scale storage platform
- Computing and storage provider for European XFEL

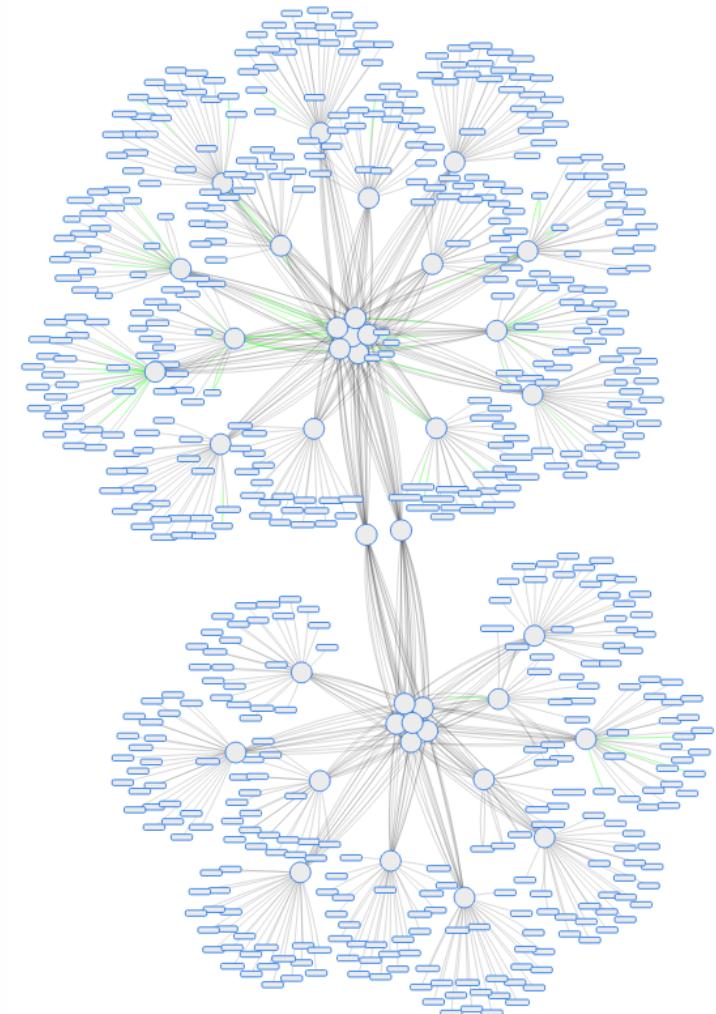




Compute infrastructure, Infiniband fabrics



Online / Offline Compute Clusters
~28,000 CPUs, 130 GPUs, ~250 TB RAM
Infiniband fabrics





Maxwell Jupyter Job Options

Maxwell partitions

shared node on Jupyter partition



Choice of GPU

none



Note: For partitions without GPUs (or choice of GPUs) the GPU selection will be set to 'none'

Job duration

8 hours



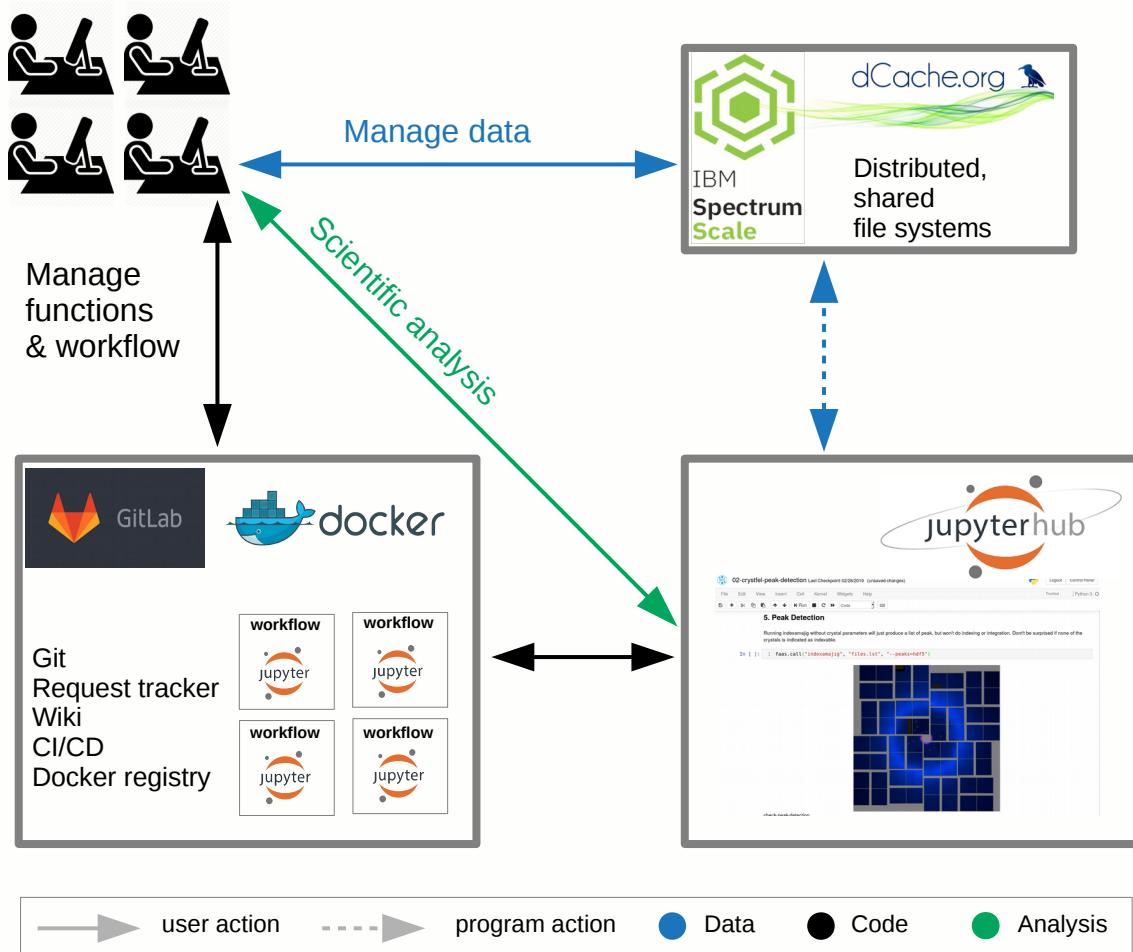
Note: on the shared Jupyter partition (jhub) the time limit is always 7 days!

Current Status					
Partition	# nodes	# avail	# GPUs avail	# P100 avail	# V100 avail
jhub	11	11	0	0	0
maxwell	61	8	0	0	0
maxgpu	19	16	16	5	9
all	327	143	0	0	0
allgpu	88	35	35	21	9
upex	183	117	14	14	0

Spawn

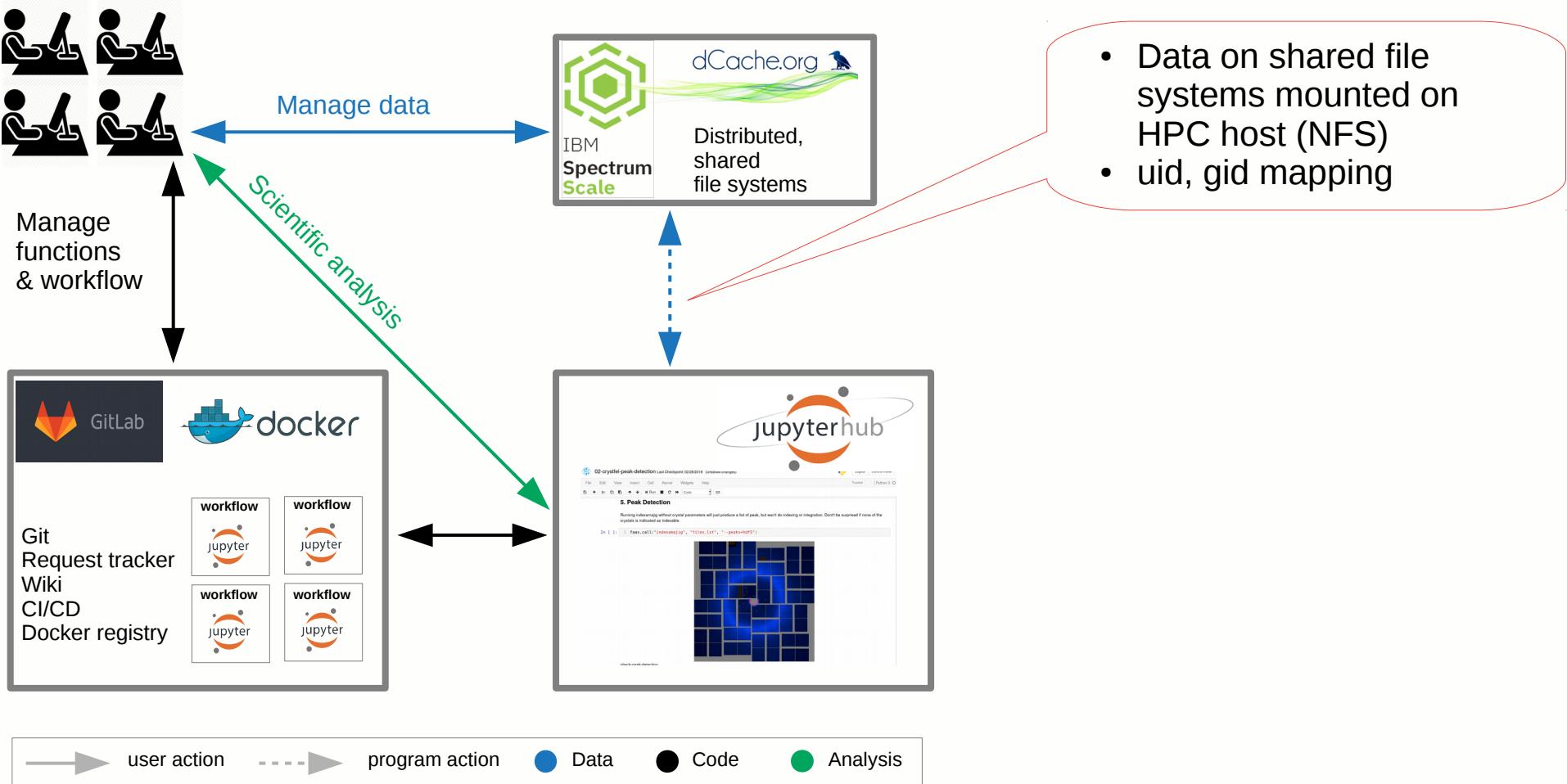


Interactive HPC with Jupyter Notebooks



Contact: eosc-pan-info@desy.de
Icons: flaticon.com (freepik/prettyicon)

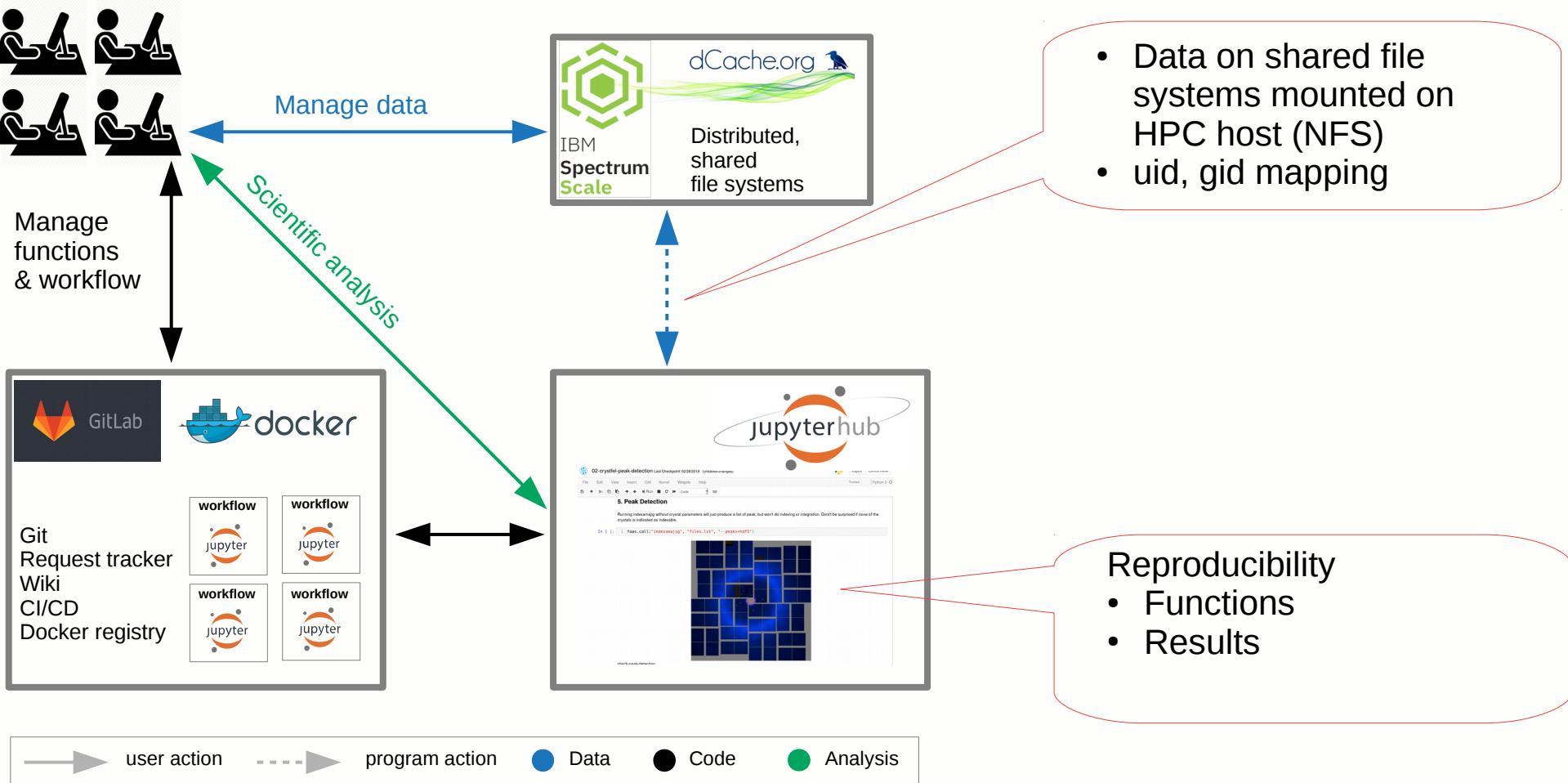
Jupyter Notebooks on HPC cluster



Contact: eosc-pan-info@desy.de
Icons: flaticon.com (freepik/prettyicon)



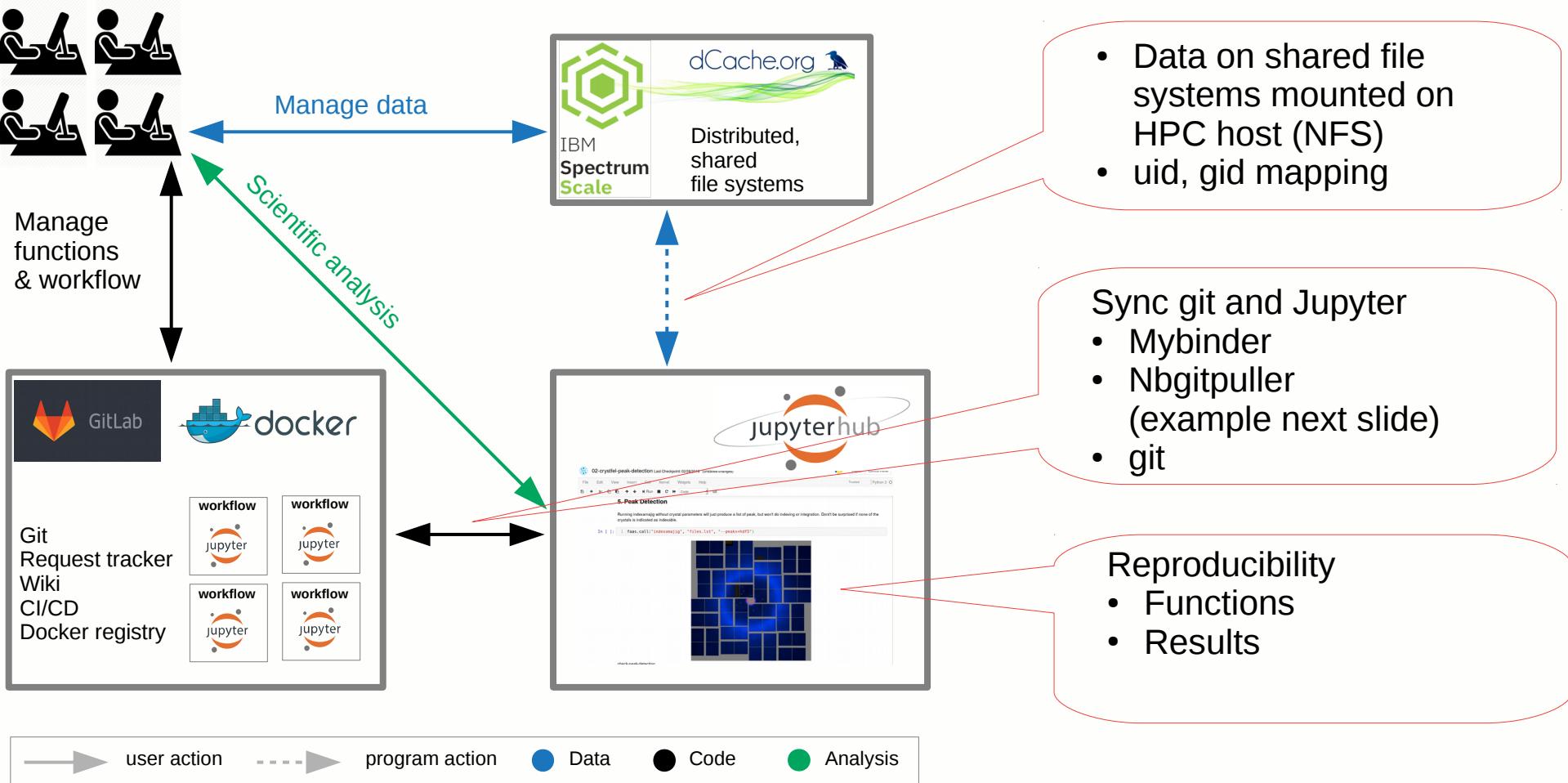
Jupyter Notebooks on HPC cluster



Contact: eosc-pan-info@desy.de
Icons: flaticon.com (freepik/prettyicon)



Jupyter Notebooks on HPC cluster



Contact: eosc-pan-info@desy.de
Icons: flaticon.com (freetipk/prettycon)



Jupyter Notebooks: working with git

nbgitpuller link generator

Use the following form to create your own [nbgitpuller](#) links.

Default Launch from Canvas

<https://eosc-pan-jhub.desy.de/hub/user-redirect/git-pull?repo=https%3A%2F%2Fgithub.com%2Ffangohr%2Fintroduction-to-python-for-com&file=index.ipynb>

JupyterHub URL

<https://eosc-pan-jhub.desy.de/>

The JupyterHub to send users to. [nbgitpuller](#) must be installed in this hub.

Git Repository URL

<https://github.com/fangohr/introduction-to-python-for-com>

branch

master

File to open

index.ipynb

This file or directory from within the repo will open when user clicks the link.

Application to Open

Classic Jupyter Notebook

JupyterLab

Link created with

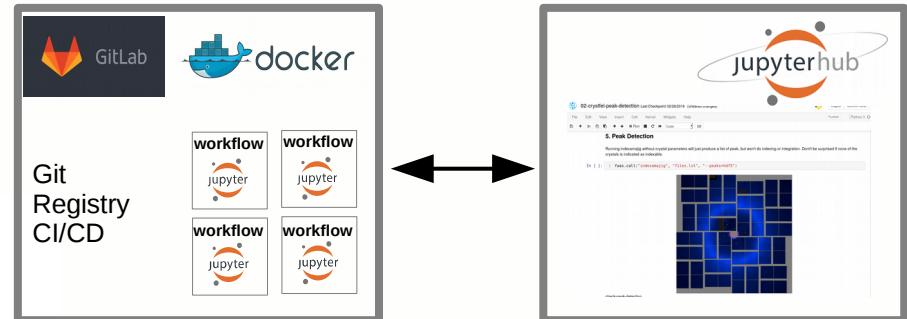
<https://jupyterhub.github.io/nbgitpuller/link.html>

Pulls git repo to Jupyter Server (need account)



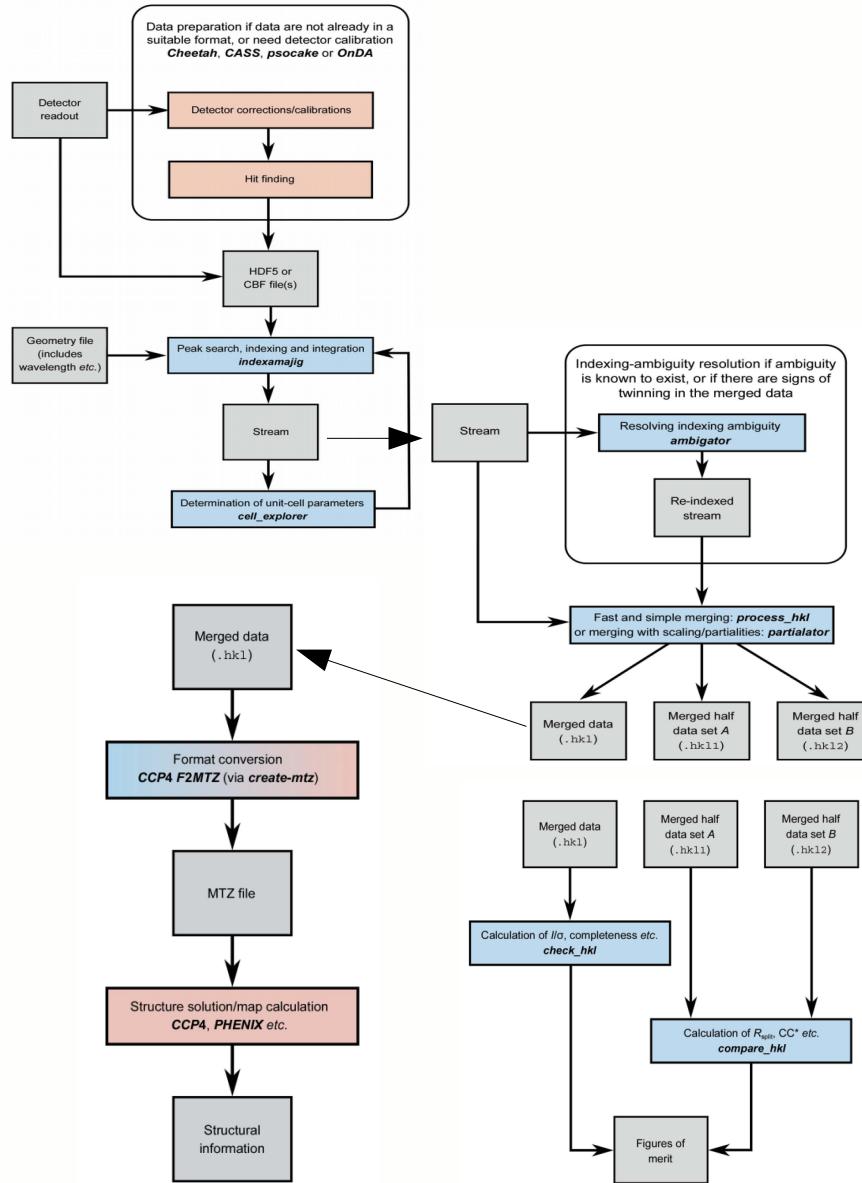
IntroductionToPython

Needs some improvements, but already works with current setup.





Data processing with the CrystFEL framework



Processing serial crystallography data with CrystFEL: a step-by-step guide

<http://journals.iucr.org/d/issues/2019/02/00/ba5291/index.html>
<https://www.desy.de/~twhite/crystfel>

Microservices:

„CrystFEL is a suite of software comprising 15 core programs: [...] CrystFEL is primarily a command-line-driven piece of software, with some exceptions [...].”

Reproducibility challenge:

„In addition to the core programs, the CrystFEL package contains a repository of scripts which are intended to be copied to the working directory and customised to suit the individual situation.”

Image: <http://journals.iucr.org/d/issues/2019/02/00/ba5291/index.html>



dCache in EOSC and eXtreme DataCloud

Single namespace
in multi-clouds.

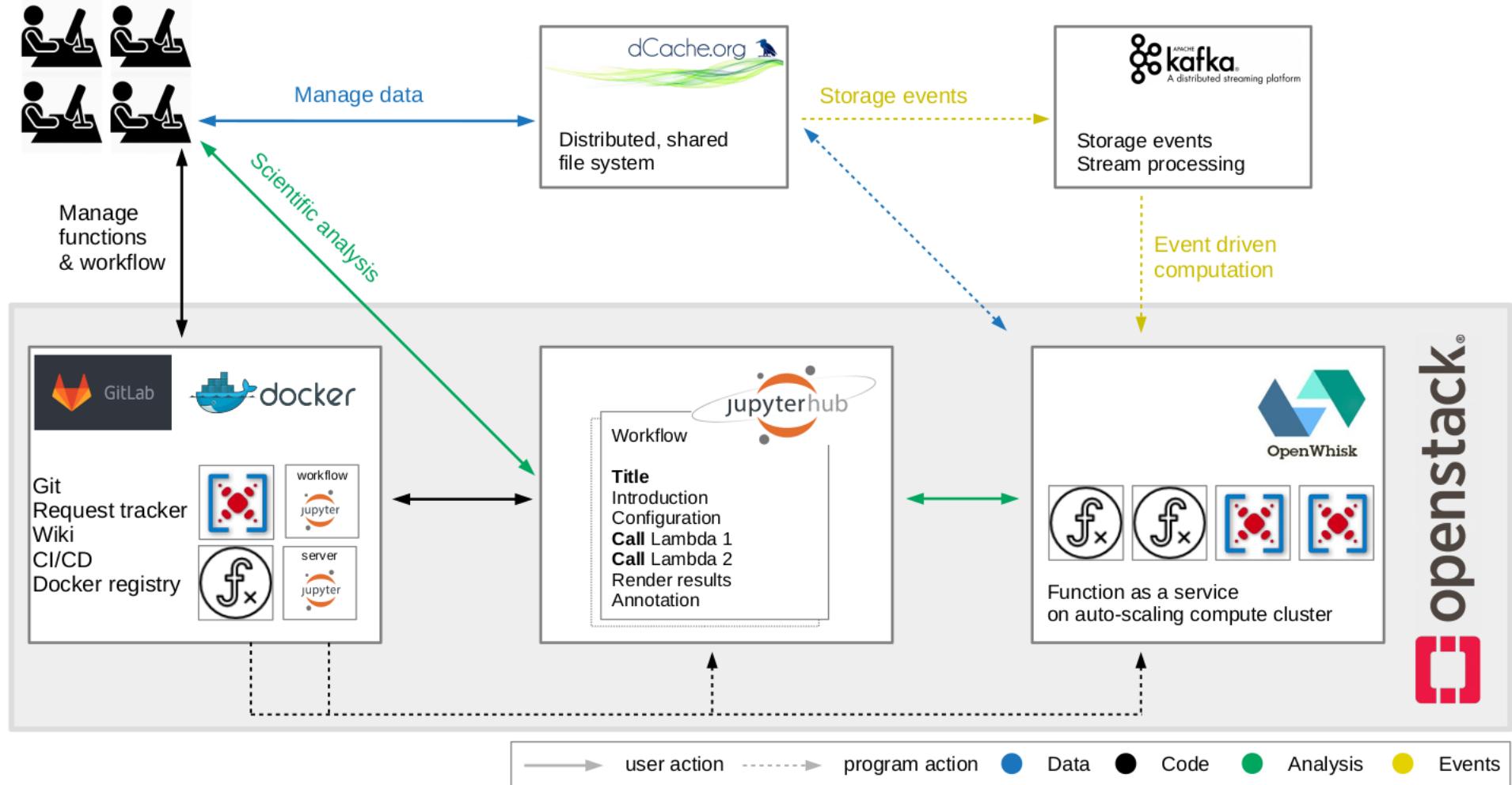


Function-as-a-Service
in Science Notebooks
and in automation.

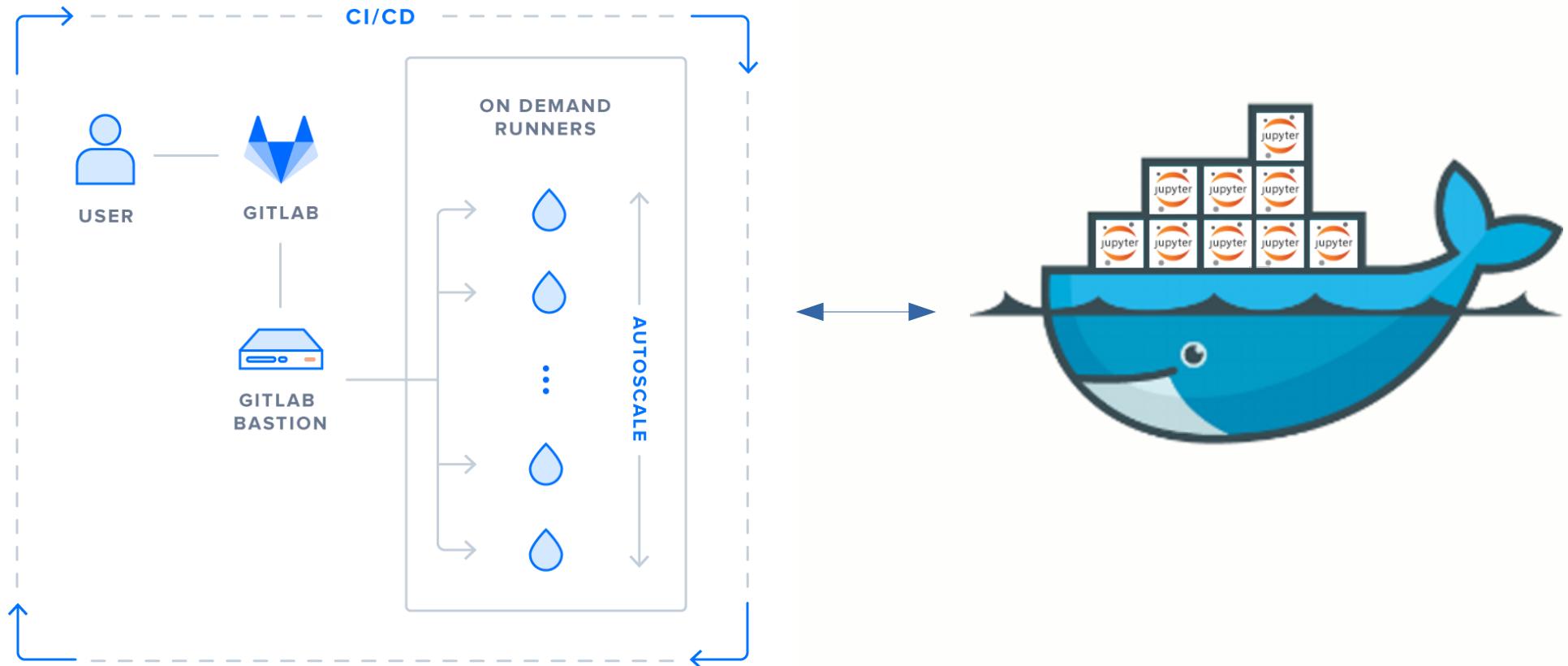
Jupyter Notebooks
in user-defined
environments.

Just push code
it builds, goes live
and scales.

Analysis and automation platform



CI/CD for user defined software stacks



- HEAT templates for OpenStack deployment
- Creates and destroys VMs with Docker-Machine using OpenStack driver
- Users define their environments for build and tests as docker containers

Image CI/CD: <https://about.gitlab.com/2018/06/19/autoscale-continuous-deployment-gitlab-runner-digital-ocean/>



User environments in Jupyter Hub

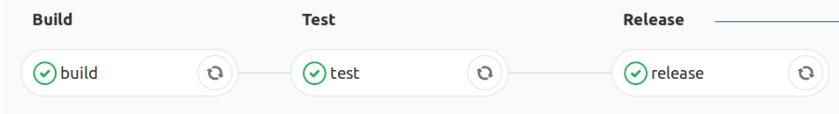
Add requirements.txt

① 4 jobs from [master](#) in 6 minutes and 59 seconds (queued for 3 seconds)

latest

3d73cf79 ... ↗

Pipeline Jobs 4



pan > eosc-pan-wsk-on-jhub > **Container Registry**

Container Registry

With the Docker Container Registry integrated into GitLab, every project can have its own space to store its Docker images.

Learn more about [Container Registry](#).

pan/eosc-pan-wsk-on-jhub ↗

Tag	Tag ID	Size	Created
latest ↗	a1557b47c	1.55 GiB	1 week ago
master ↗	a1557b47c	1.55 GiB	1 week ago



User environments in Jupyter Hub

Add requirements.txt

4 jobs from [master](#) in 6 minutes and 59 seconds (queued for 3 seconds)

[latest](#)

[3d73cf79](#) ... ↻

Pipeline Jobs 4

Build → Test → Release

```
graph LR; Build((Build)) --> Test((Test)); Test --> Release((Release));
```

pan > eosc-pan-wsk-on-jhub > Container Registry

Container Registry

With the Docker Container Registry integrated into GitLab, every project can have its own space to store its Docker images.

Learn more about [Container Registry](#).

pan/eosc-pan-wsk-on-jhub ↻

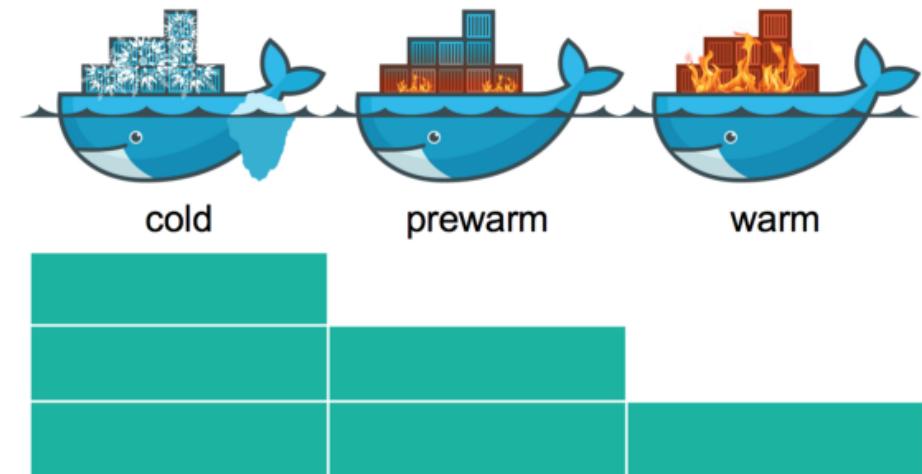
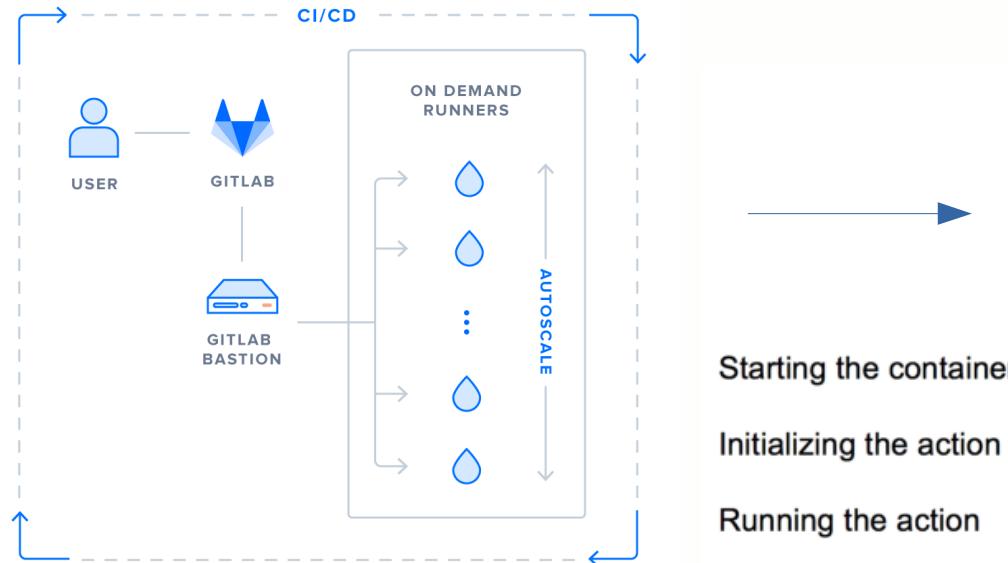
Tag	Tag ID	Size	Created
latest	a1557b47c	1.55 GiB	1 week ago
master	a1557b47c	1.55 GiB	1 week ago

Spawner Options

- EOSC PaN FaaS image**
Image with OpenWhisk client for function-as-a-service, based on the scientific python notebook: eosc-pan-git.desy.de:5555/pan/eosc-pan-wsk-on-jhub
- Scientific Python environment**
Default image with scipy and other utils
- Extended DataScience environment**
Additional packages for Python, R, and Julia.
- Spark environment**
The Jupyter Stacks spark image!
- nbgitpuller-dev**
We are working on importing your projects from shared links -dev.

Spawn

Container as a function



Cloud functions: No infrastructure management by the user
Efficient scaling per-function, rapid provisioning

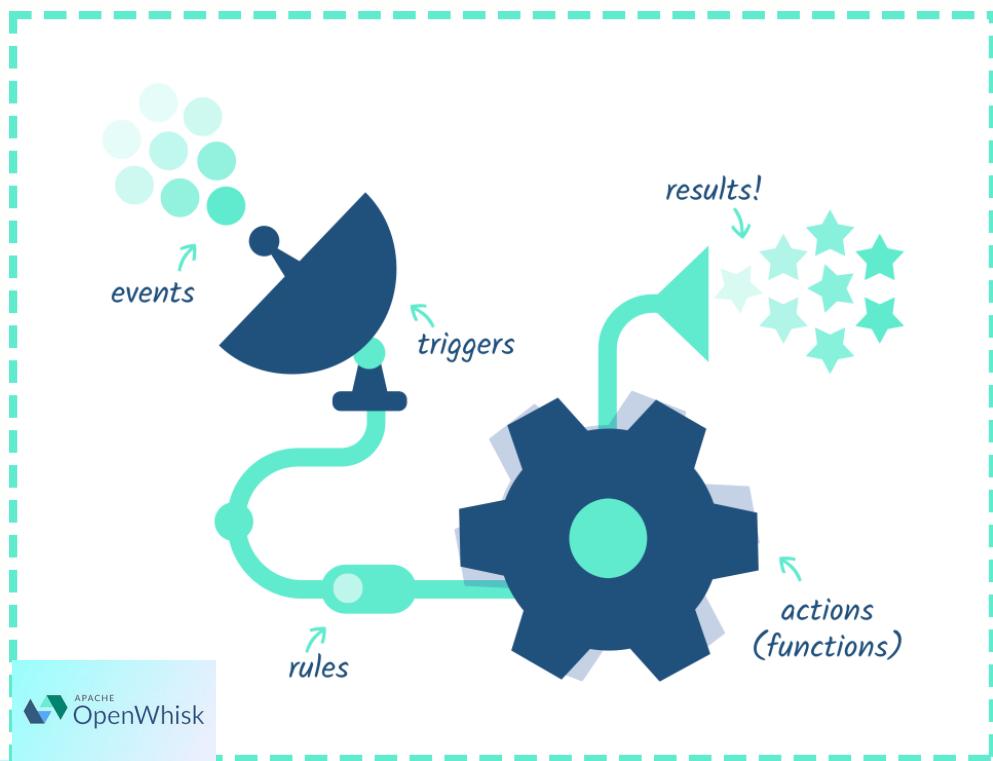
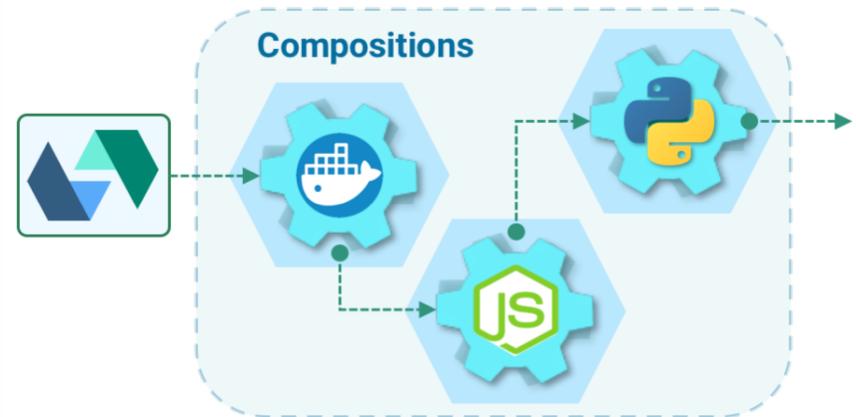
Number of services $>> 1$

Number of requests of a service = arbitrary function of time

<https://medium.com/openwhisk/squeezing-the-milliseconds-how-to-make-serverless-platforms-blazing-fast-aea0e9951bd0>

Event driven computation

- Execute code in response to events
 - Incoming data in dCache
 - Data re-staged from tape
 - New messages in dedicated queues
 - Produce derived data
 - Extract metadata
 - Manage data locality



Images: <https://openwhisk.apache.org/>



Live demonstration: FaaS in Science Notebooks

Demo Jupyter Notebooks:

<https://eosc-pan-git.desy.de/pan/dcache-event-demo>

Demo Jupyter Server:

<https://eosc-pan-git.desy.de/pan/eosc-pan-wsk-on-jhub>

Demo Jupyter Hub:

<https://eosc-pan-jhub.desy.de>

Demo data:

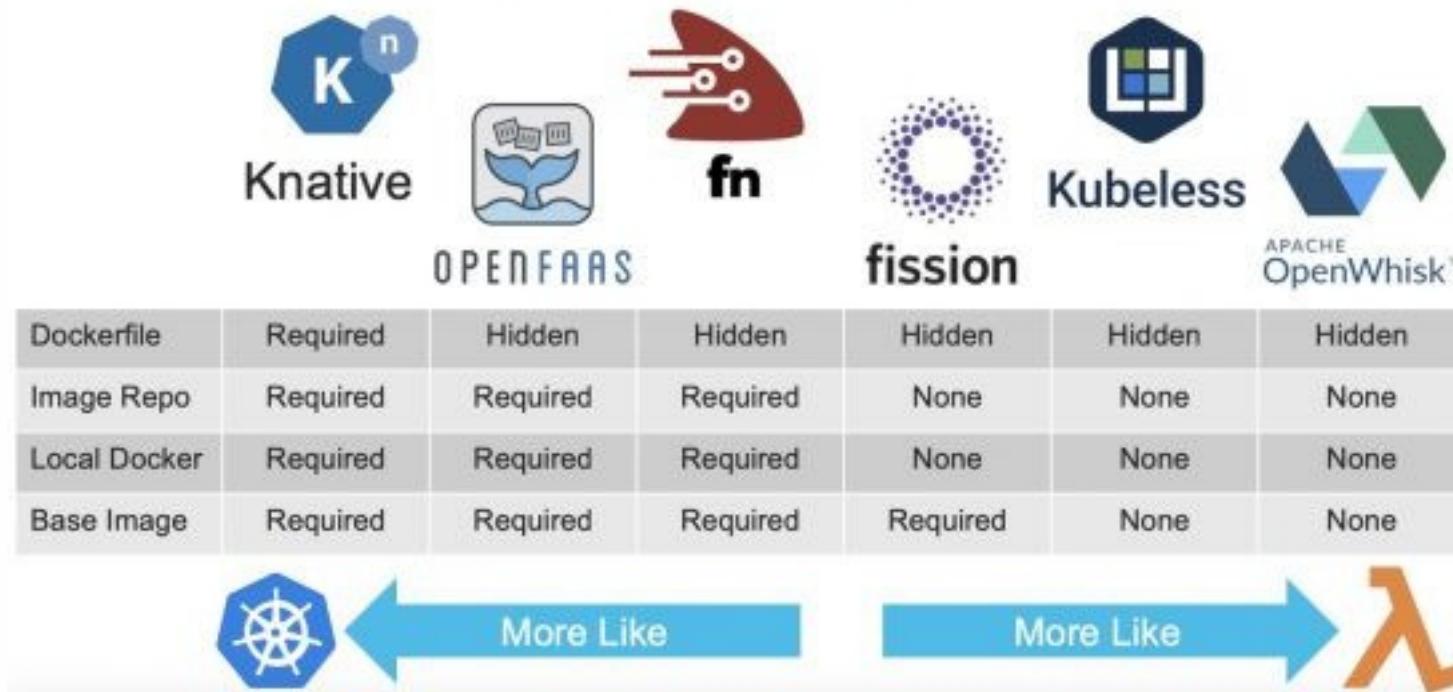
<https://dcache-xdc.desy.de:443/Demos/EOSC/cxidb-21/>



Portability	Cloud Orchestration templates, dockerized processes
Accessibility	Federated AAI, OIDC
Interoperability	Standard interfaces REST, JSON
User-friendliness	Server-hidden Industry leading projects: GitLab, Jupyter Notebooks
Reproducibility	Everything in version control and CI/CD hashsums/pids for data, infrastructure-as-code .. and also for deployed functions and publications
Scalability	Auto-scaling Vms and Containers on OpenStack + Kubernetes

FaaS on K8s: findings from fonk-apps.io

Auto-scaling microservices
Binding to event ecosystems



Functions run as K8s pods
Orchestrate source-to-container builds
Routing and managing traffic

Functions run as single containers
Containers for language runtimes
Only add function code

Source: <https://blogs.cisco.com/cloud/examining-the-faas-on-k8s-market>

Thank you for your attention!

Thanks for contributions:

Tigran Mkrtchyan, DESY
Paul Millar, DESY
Johannes Reppin, DESY

Open Source Software:

OpenStack
Kubernetes
Docker
dCache
Kafka
Project Jupyter
GitLab
OpenWhisk
Python
Linux
...

