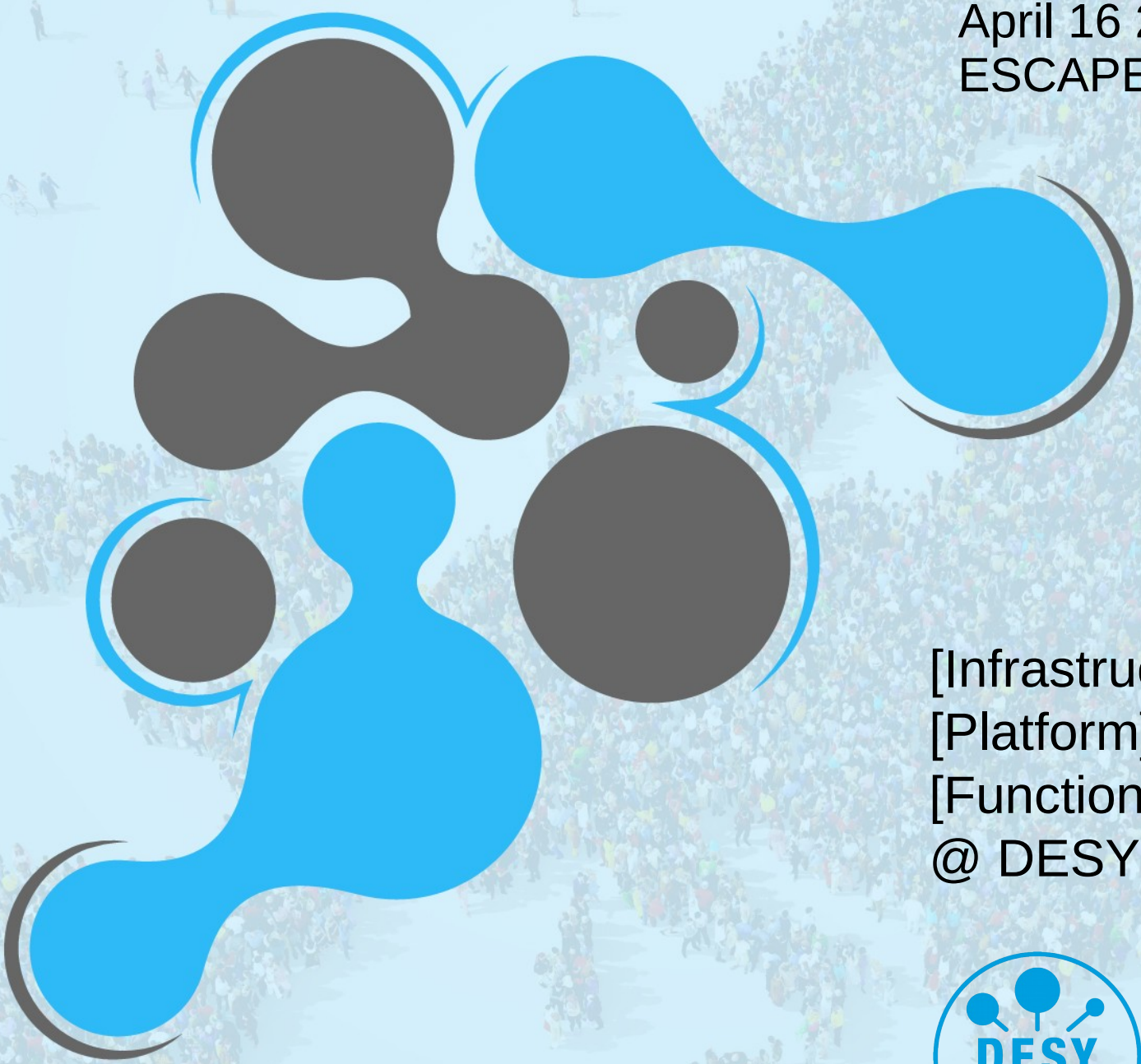


Michael Schuh, DESY.
April 16 2019
ESCAPE, Groningen

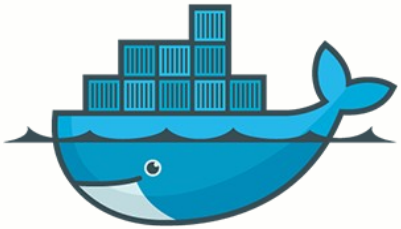
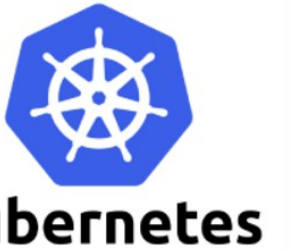
[Infrastructure]
[Platform]
[Function]-as-a-Service
@ DESY OpenStack Cloud

DESY.

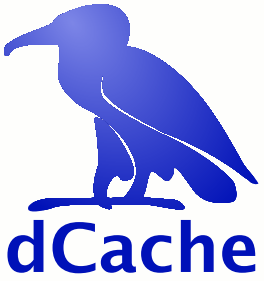
EOSCpilot
The European Open Science Cloud for Research Pilot Project
www.eoscpilot.eu

Outline

Cloud Computing
Container Orchestration
Software Defined Networking
Infrastructure as code

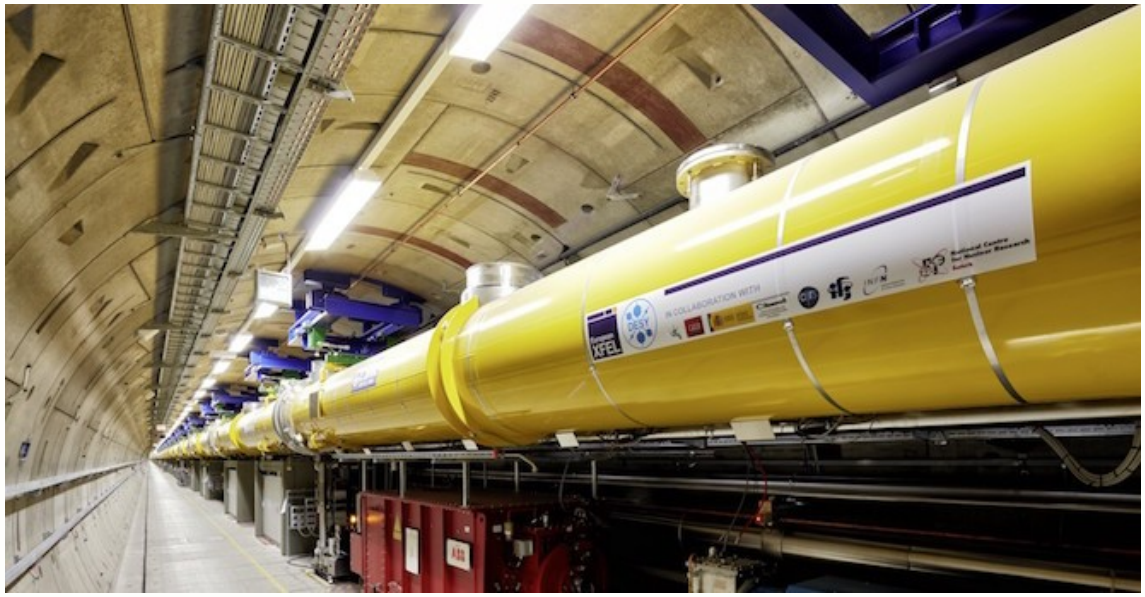
Peta-scale distributed storage
Storage events
Event stream API



Science Notebooks
Function as a service
Event-driven computing

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

DESY



DESY Deutsches **E**lektonen-**SY**nchrotron (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

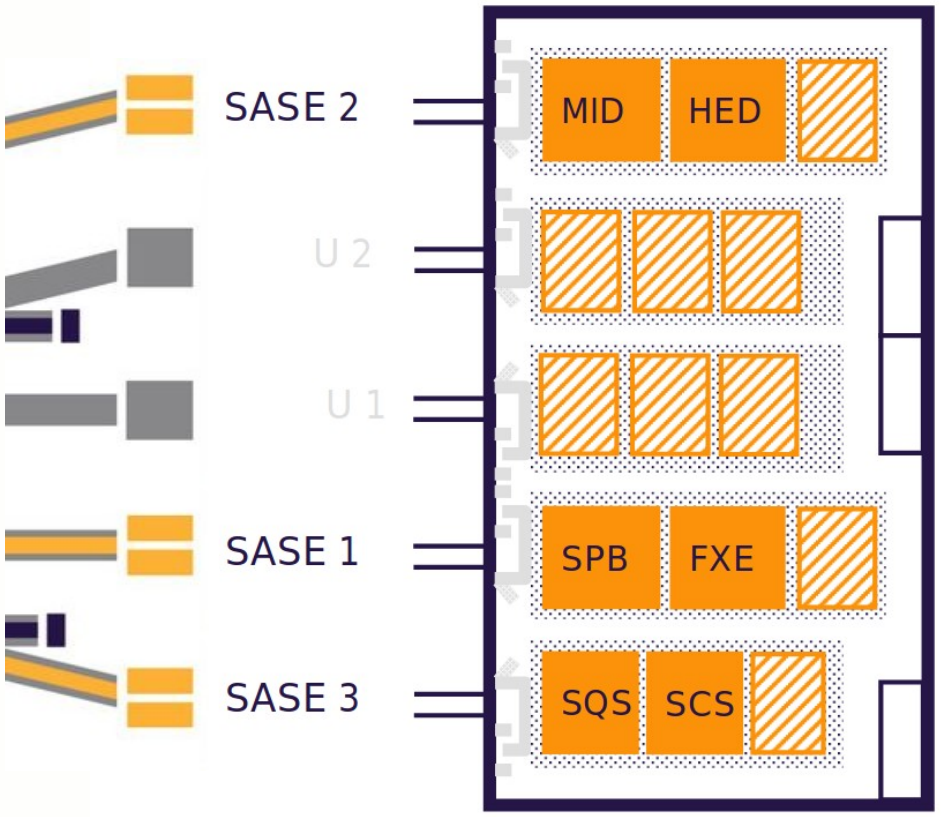
The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

DESY IT

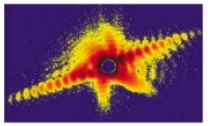
- Scientific Computing, HPC, HTC, Grid and Cloud
- LHC Tier-2 center, large scale storage and archiving
- dCache – a peta scale storage platform
- Computing and storage provider for European XFEL

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

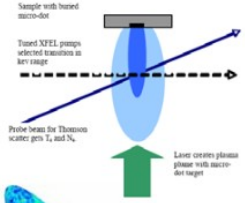
Experiments at the European XFEL



MID Materials Imaging & Dynamics

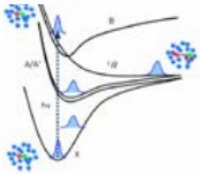


HED High Energy Density Science

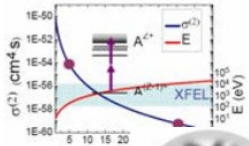


SPB Single Particle & Biomolecules

FXE Femtosecond X-ray Experiments



SQS Small Quantum Systems



SCS Spectroscopy & Coherent Scattering

More about experiments: http://www.xfel.eu

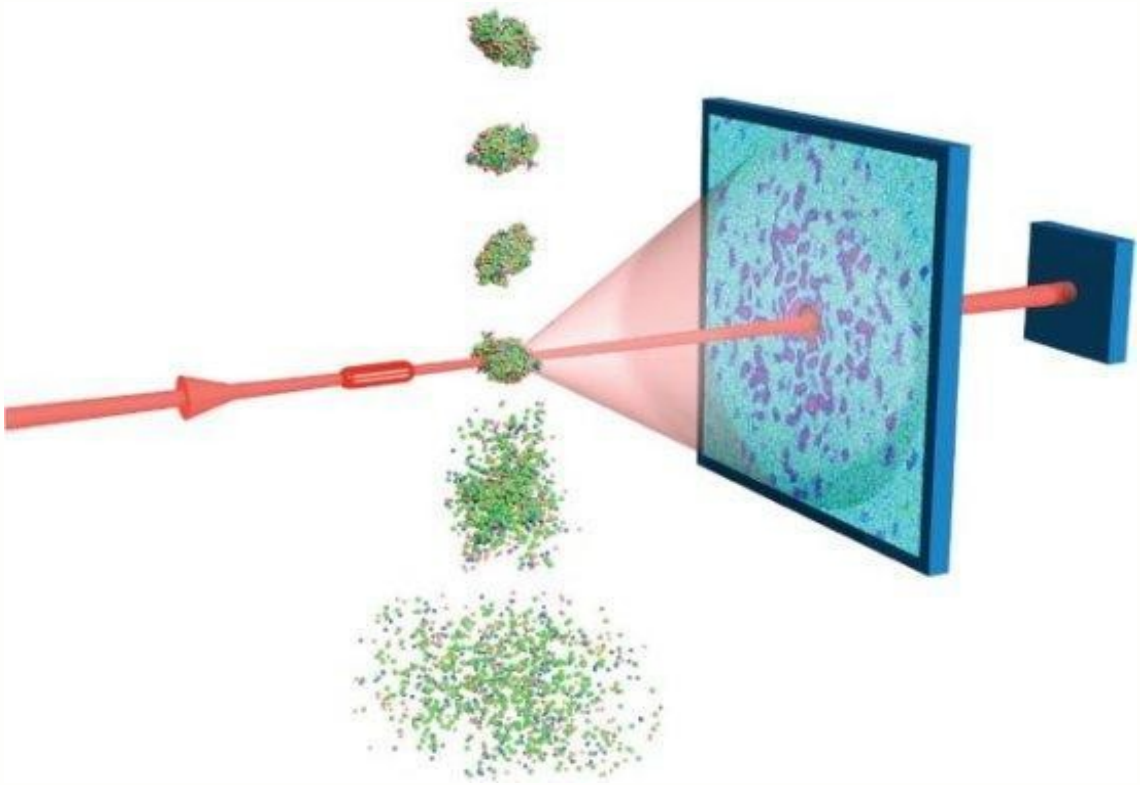
European XFEL

Experiments in physics, chemistry, materials science, biology, nanotechnology

Source: http://xfel.eu

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

Exposure time: femto-seconds (10^{-15} s)

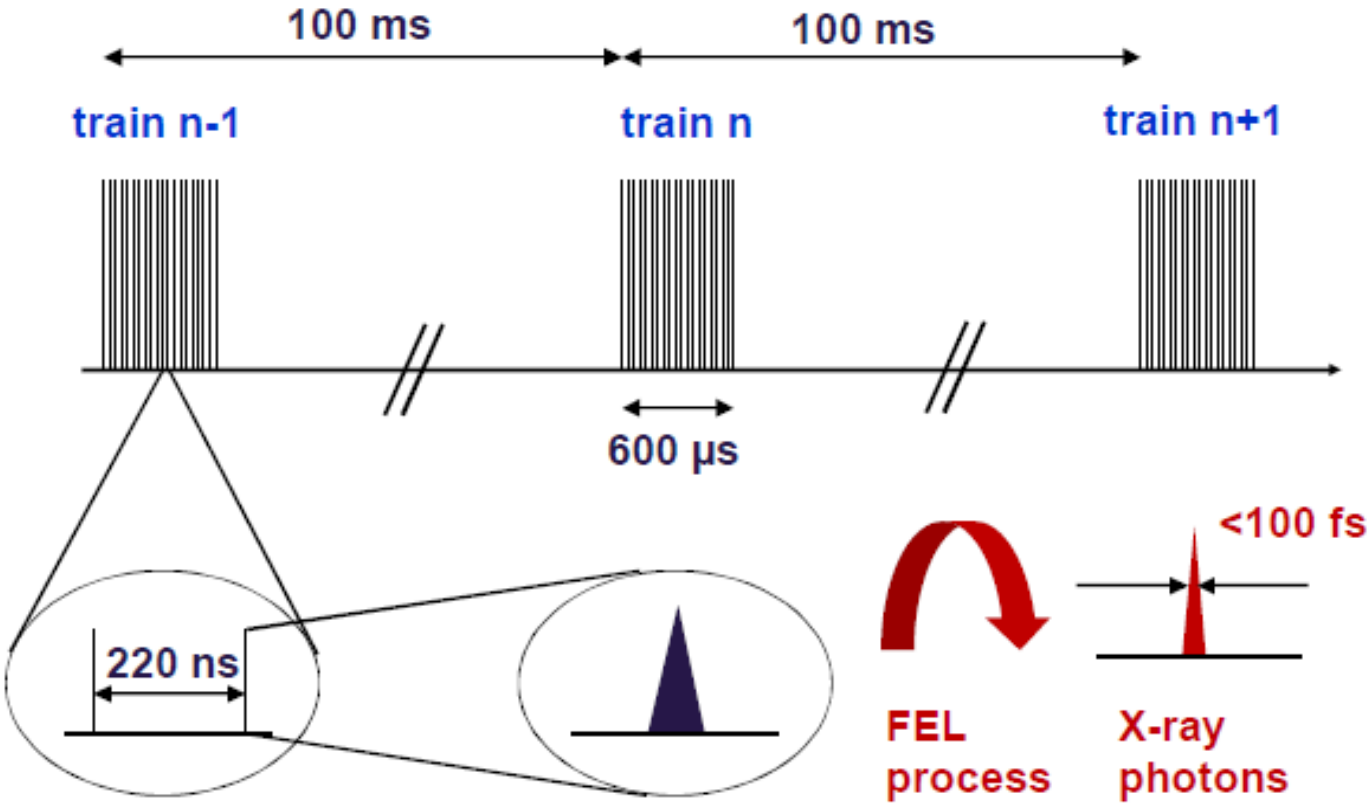


- Molecular 3D imaging at the atomic scale
- Diffraction before destruction
- Unprecedented resolution in both: time and space
- Dynamics of inner structures (not just surface) of biological objects
  - Virus
  - Cell nucleus
  - ...
- To reconstruct a 3D image, a large amount of random oriented single images have to be combined

source: https://cid.cfel.de/research/femtosecond_crystallography/

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

Data rates for the FXE instrument



Beam time per experiment: several hours

Readout in bunch structure:

- 10 Hz train rate
- 4.5 MHz puls rate
- 2700 pulses per train

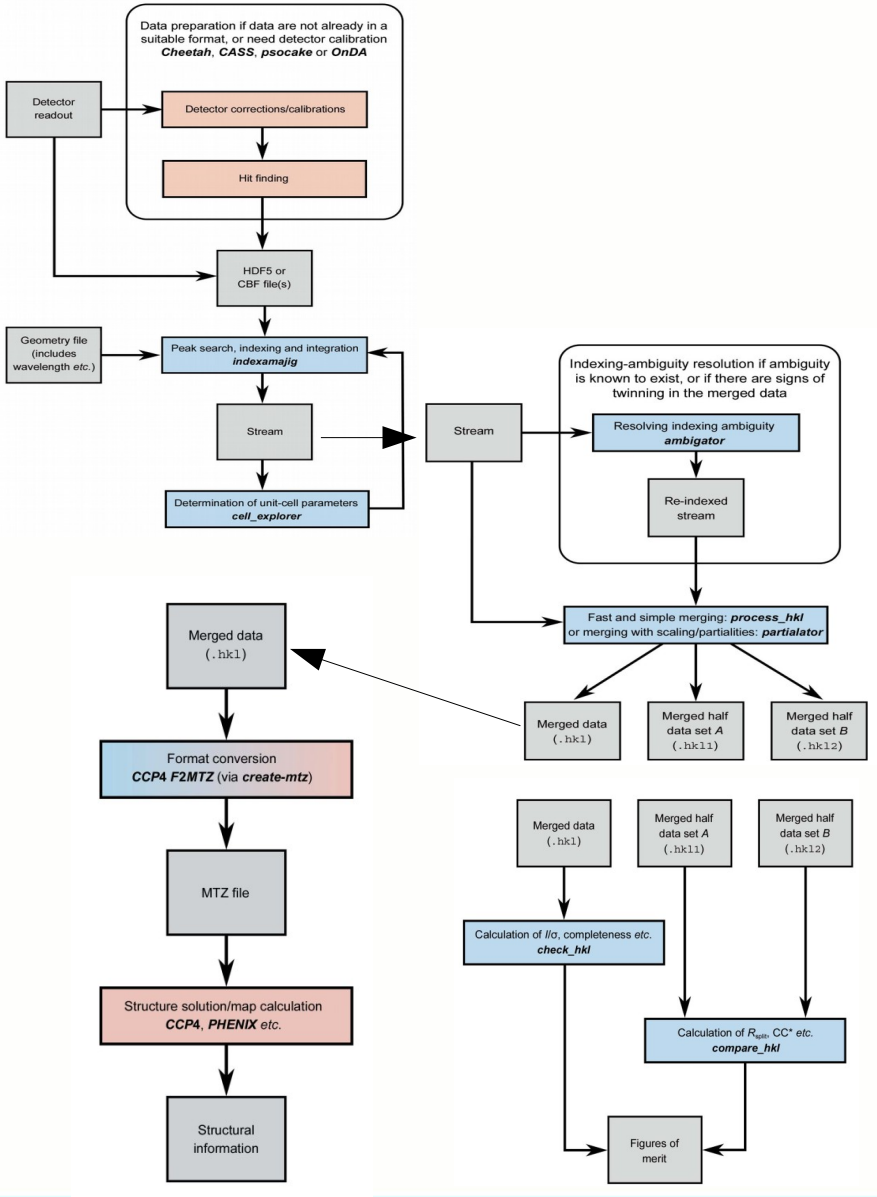
1 Mpxl detectors:
- 2MB/pulse
- 1GB/train
- 10GB/second

4 Mpxl detectors:
- 8MB/pulse
- 3GB/train
- 30GB/second

Image: http://xfel.eu

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

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

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

dCache in EOSC pilot The European Open Science Cloud for Research Pilot Project and eXtreme DataCloud

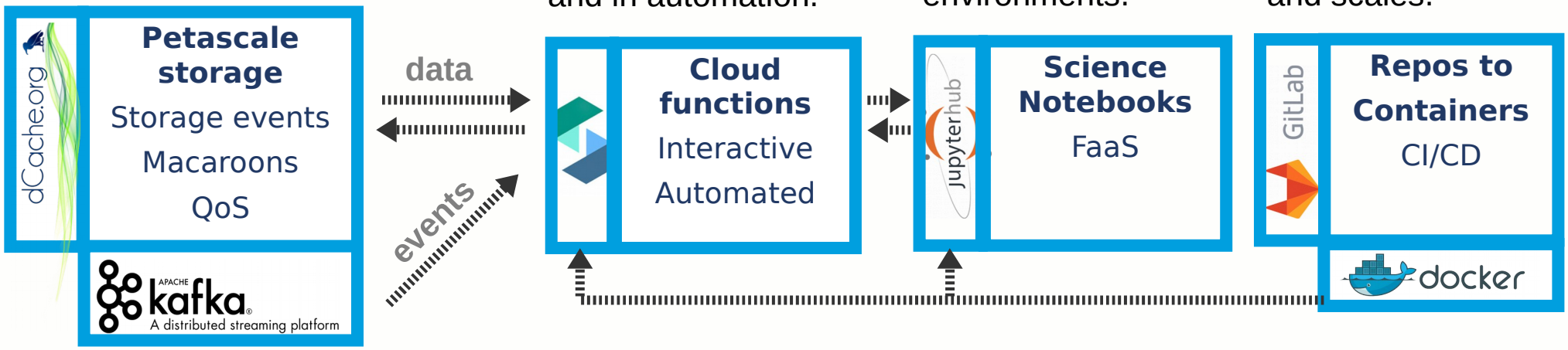
Event-driven data placement and processing

Single namespace in multi-clouds.

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

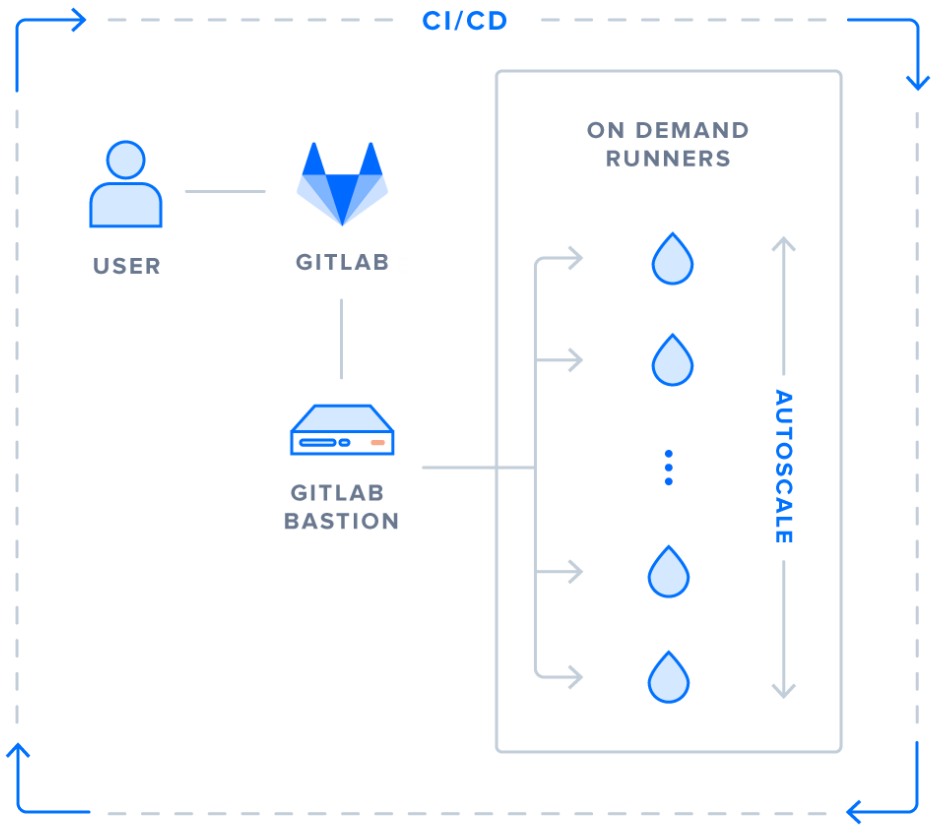
Jupyter Notebooks in user-defined environments.

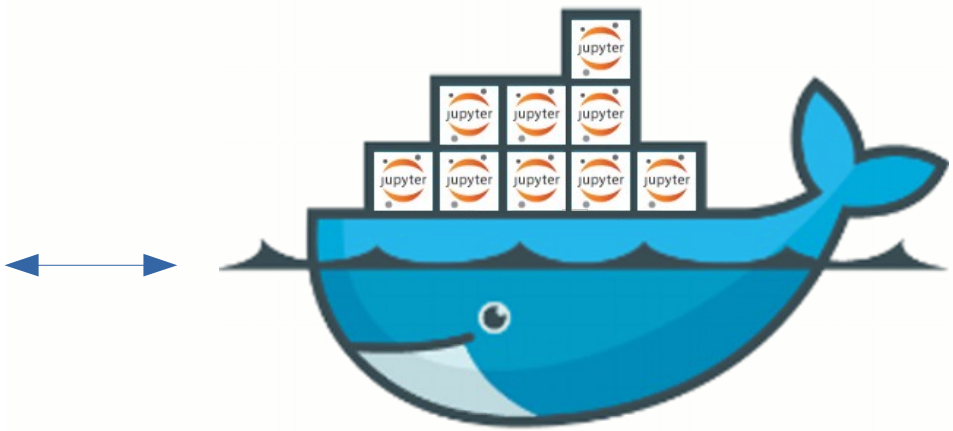
Just push code it builds, goes live and scales.



The European Open Science Cloud for Research pilot and the eXtreme DataCloud projects are funded by the European Commission, DG Research & Innovation under contract no. 739563 and 777367

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/

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

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
build	test	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

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

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
build	test	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

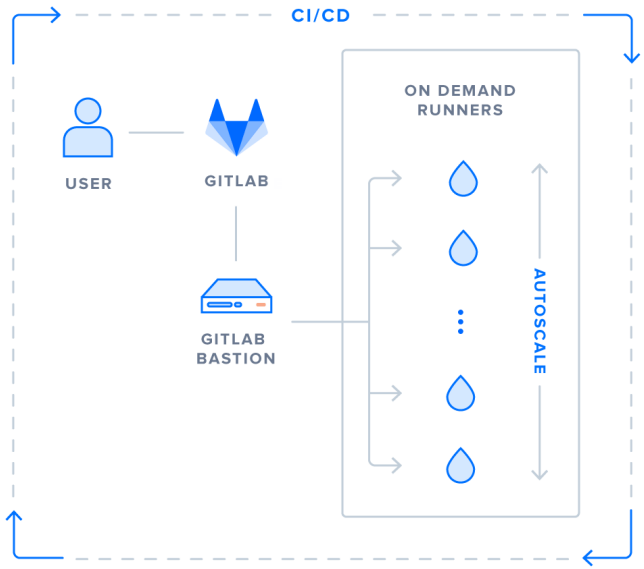
Spawner Options

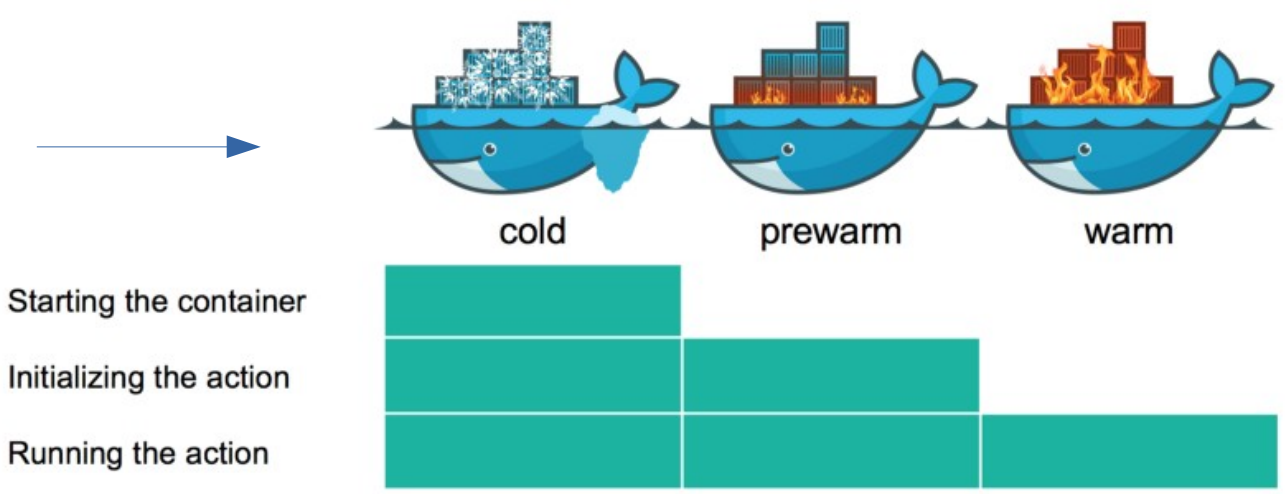
- **EOSC PaN FaaS image**
  Image with OpenWhisk client for function-as-a-service, based on the scientific python notebook: eosc-pan-git.desy.de:5555/pan/eosc-pan-wsk-on-jhub
- **Scientific Python environment**
  Default image with scipy and other utils
- **Extended Datascience environment**
  Additional 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

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

Container as a function





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

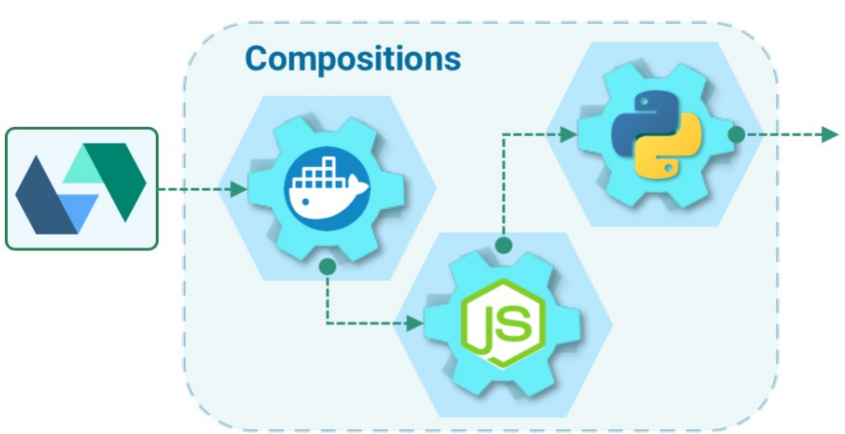
Number of services >> 1
Number of requests of a service = arbitrary function of time

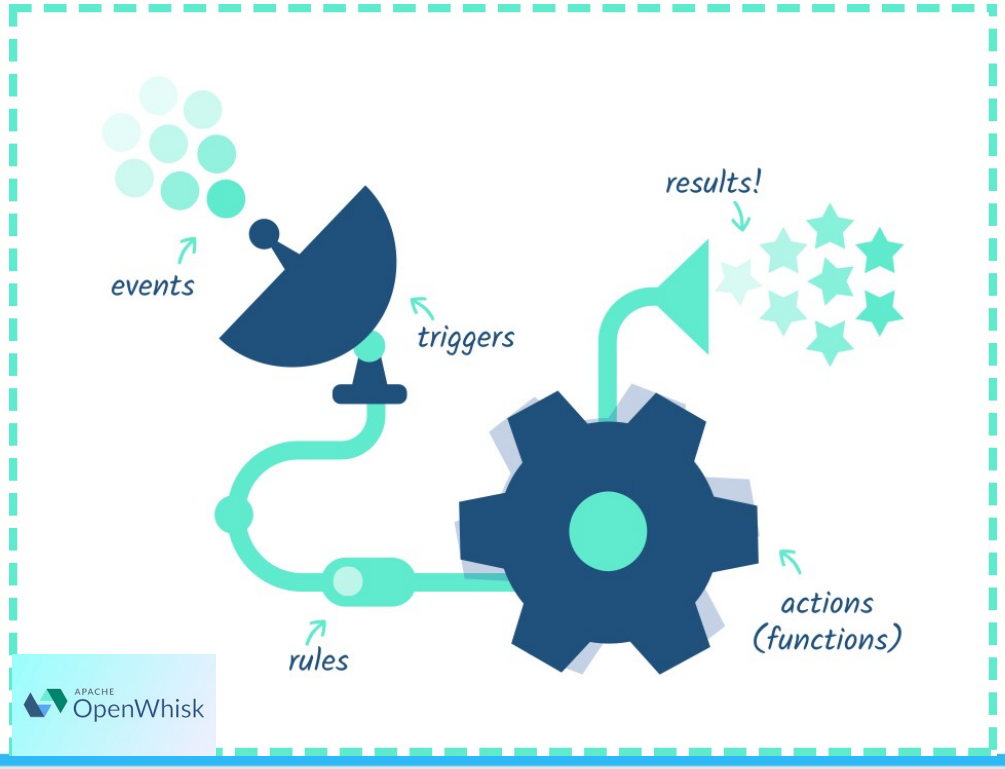
https://medium.com/openwhisk/squeezing-the-milliseconds-how-to-make-serverless-platforms-blazing-fast-aea0e9951bd0

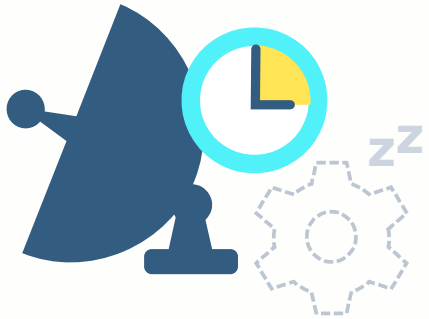
The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

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/

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

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/

Demo microservices:
https://eosc-pan-git.desy.de/pan/files-in-run
https://eosc-pan-git.desy.de/pan/hdfsee

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

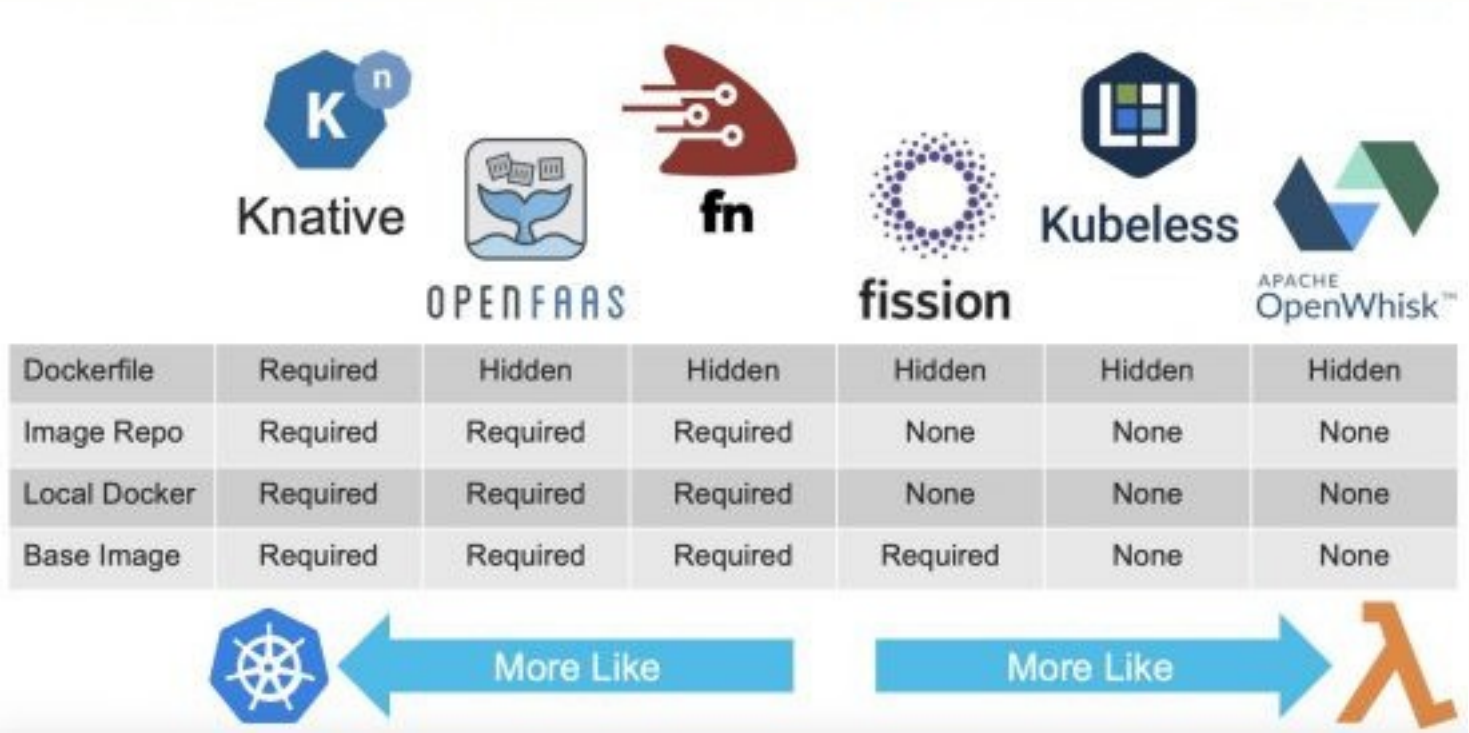
Conclusion

Portability	Cloud Orchestration templates, dockerized processes
Accessibility	Federated AAI, OIDC
Interoperability	Standard interfaces REST, JSON
User-friendlyness	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

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

FaaS on K8s: findings from fonk-apps.io

Auto-scaling microservices
Binding to event ecosystems



Dockerfile	Required	Hidden	Hidden	Hidden	Hidden	Hidden
Image Repo	Required	Required	Required	None	None	None
Local Docker	Required	Required	Required	None	None	None
Base Image	Required	Required	Required	Required	None	None

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

The European Open Science Cloud for Research pilot project is funded by the European Commission, DG Research & Innovation under contract no. 739563

Thank you for your attention!

Thanks to dCache people:
Tigran Mkrtchyan, DESY
Paul Millar, DESY
Johannes Reppin, DESY

Open Source Software:
OpenStack
Kubernetes
Docker
dCache
Kafka
Project Jupyter
GitLab
OpenWhisk
Python
Linux
…

The European Open Science Cloud for Research pilot project
is funded by the European Commission, DG Research & Innovation
under contract no. 739563