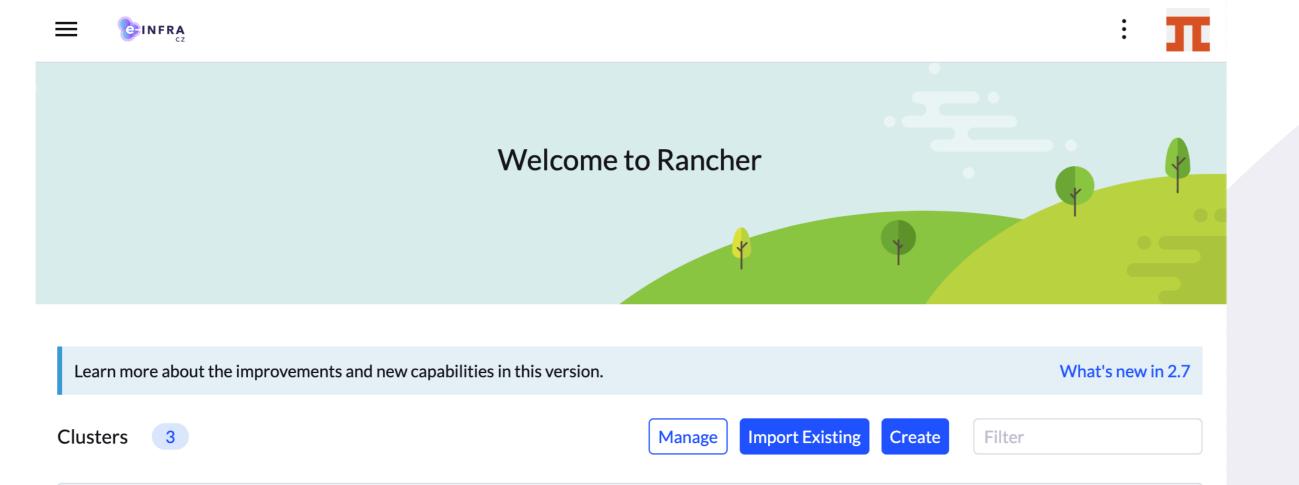
Easy Deployment of Dask and Jupyter Notebooks on Managed Kubernetes



6/16/23 3:38 PM



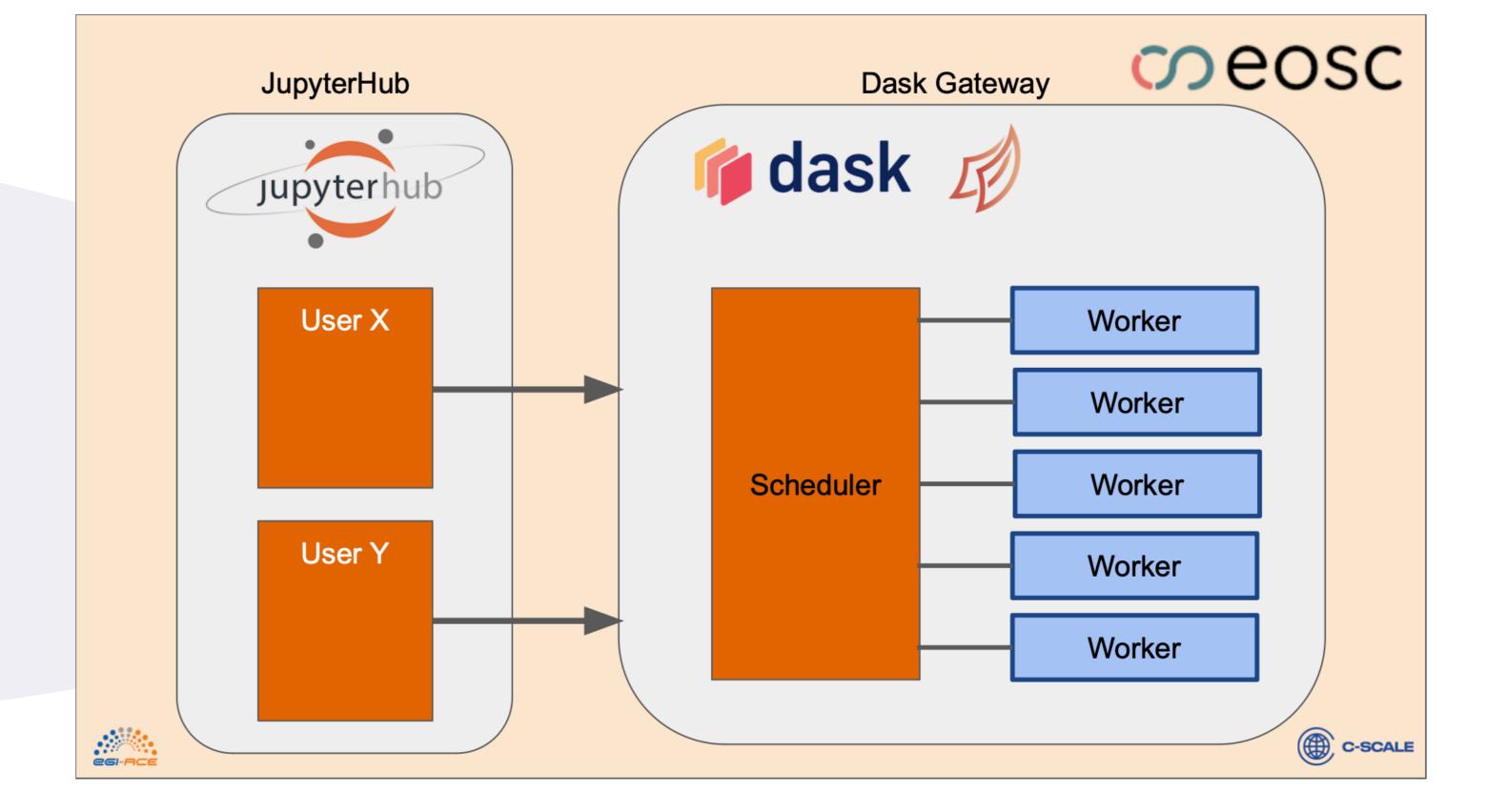
Managed Kubernetes

- Fully managed platform with integrated services
- Let's Encrypt certificates
- Automatic DNS registration
- Storage integration
- Docker container registry

State 🗘	Name 🗘	Provider 🗘	Kubernetes Version	CPU 🗘	Memory 🗘	Pods 🗘
Active	kuba-cluster	Imported RKE2	v1.24.13+rke2r1	2112 cores	17 TiB	2083/5280
Active	kubh-cluster	Imported RKE2	v1.26.5+rke2r1	768 cores	2.95 TiB	160/960
Active	rancher	Local RKE2	v1.25.10+rke2r1	24 cores	94 GiB	111/330

- Minimal infrastructure knowledge required
- Focus on application
- Need to accept shared infrastructure rules





Operated by CERIT-SC, ICS MUNI, Docs: https://docs.cerit.io

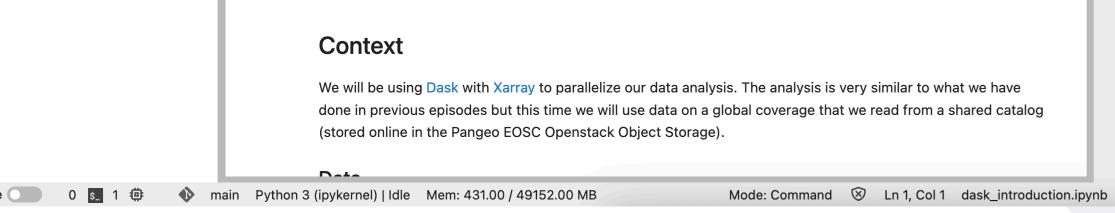
Successful Case

PANGEO is a global community initiative for Big Geoscience Data with active developers, scientists, and users. PANGEO requested training infrastructure for Jupyter Notebooks and Dask computing clusters. It was implemented on the managed Kubernetes platform.

			dask_i	introdu	ı Jup	oyterLab	×	+								~
÷	\rightarrow C	e (🔒 das	kgate	way.vn	n.fedcl	oud.eu	u/use	r/lh	ejtmanek/lab/tree/tutorial/pangeo101/dask_introduction.ipynb	Û	☆	•	• [:
$\mathbf{\hat{C}}$	File E	dit	View	Run	Kerne	el Git	Tabs	s Se	ettir	ngs Help						
	+		Đ	±	C	•\$⁺	토 da	ask_in	tro	duction.ipynb \times +						°0
	Filter	files k	by nam	6		٩	8	+ 3	X	🗇 📋 🕨 🔳 C 🕨 Markdown 🗸 🗮 🕚 git	3	🛚 Pytho	on 3 (ip	ykern	el) 🔿	L .
0	🖿 / tut	torial ,	/ pange	eo101 /						Parallel computing with Dask	F	^ ↓	, †	Ŧ	Î	¢
Ŵ	Name		▲		st Modi months											
◆★	 das data taly visu xarr 	a_diso y.geoj ualizat	co json ti	2 2 2	months months months months	ago ago ago				Authors & Contributors Authors • Tina Odaka, Ifremer (France), @tinaok • Pier Lorenzo Marasco, Ispra (Italy), @pl-marasco Contributors • Anne Fouilloux, University of Oslo (Norway), @annefou • Guillaume Eynard-Bontemps, CNES (France), @guillaumeeb						
										 Overview Questions What is Dask? How can I parallelize my data analysis with Dask? Objectives Learn about Dask Learn about Dask Gateway, Dask Client, Scheduler, Workers Understand out-of-core and speed-up limitations 						

Benefits and successful story

- Scientists benefit from a collaborative platform to do research in a consistent and scalable way.
- Scientists focus on science and don't need to worry about managing the underlying platform: Kubernetes cluster, container registry, DaskHub deployment, etc.
- The platform fosters reproducible open science with state of the art tools. Resulting notebooks can be easily shared and reproduced by others.
- More than 100 researchers from across more than 10 countries have benefited from this deployment to learn how to include the



Pangeo tools into their research without having to worry about the technical details related to deployment and management of the platform.

Lukáš Hejtmánek^{1,2}, Adrián Rošinec^{1,2}, Sebastian Luna-Valero³

1. Masaryk University

2. CESNET

3. EGI Foundation

containers@e-infra.cz

MASARYK UNIVERSITY

cesnet ******







muni.cz/go/containers