terrabyte: A "Cloud-Like" HPC System for Addressing Earth Observation Challenges

Peter Friedl and Jonas Eberle

German Aerospace Center (DLR) – German Remote Sensing Data Center (DFD)

23.06.2025

terrabyte

Earth Observation Center

DLR

terrabyte
How it started

- Cooperation between DLR and Leibniz Supercomputing Center (LRZ)
  - DLR EOC: Software Services, Data management & access, User support, EO research
  - LRZ: Hosting and Operations of IT infrastructure, IT research
- Launched in June 2023

terrabyte
A “cloud like” HPC System

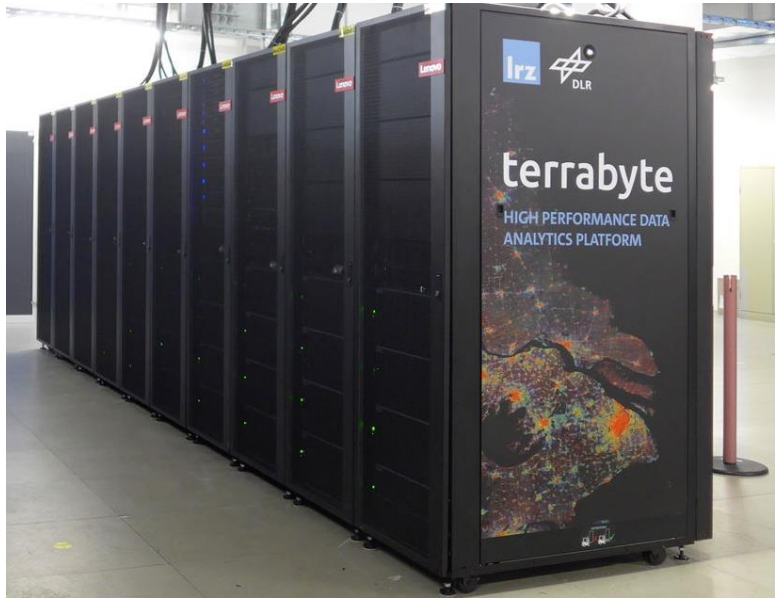
Aspect	(Classic) HPC	(Commercial) Cloud
Performance	No virtualization overhead ✓	Virtualization overhead !
Elasticity	Limited & queued !	Highly elastic, scalable ✓
Costs (for users)	Low if in-house use ✓	Pay-as-you-go !
Control / Sovereignty	Full institutional control ✓	Dependency on commercial providers (often US-based) !
Flexibility	Low — fixed environments, software often pre-installed by admins !	High — users can configure their own environments ✓
User Experience	Technical barrier: SSH-only access, command-line tools, batch-only jobs !	Generally easy — web UI, self-service provisioning ✓

Terrabyte aims at combining the performance of HPC with the usability and flexibility of cloud systems!

Hardware
Fast and powerful resources

Online-Storage (DSS)
- 53 PB net
- 6 PB user space
- GPFS Filesystem with 360 Gbit/s Infiniband

Compute Cluster
- 272 CPU-Nodes with ~44.000 Intel Xeon Platinum CPU vCores und 273 TB RAM
- 48 Nvidia A 100 GPU-Nodes with 188 GPUs and 47 TB RAM + 14 TB GPU RAM

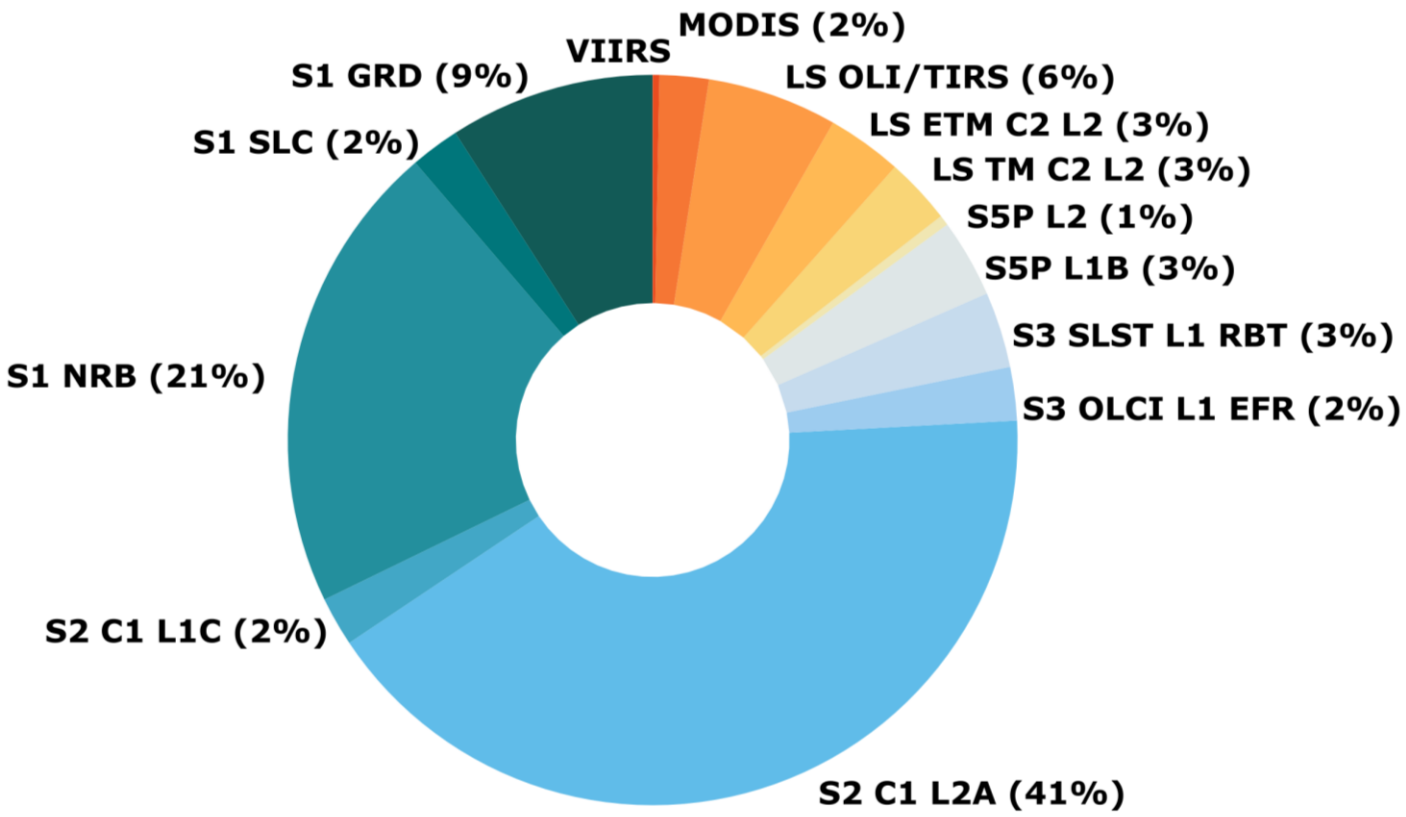


Network connection between DLR-satellite data archive (60 PB) and LRZ
- 10 Gbit/s, upgrade to 100 Gbit/s possible

Internet connection between LRZ and CDSE, USGS, NASA
- max. 100 Gbit/s, depending on server and traffic

Data Availability
No need to download data

Sentinel-1 GRD	4 177 TB
Sentinel-1 SLC**	1 000 TB
Sentinel-1 NRB*	9 607 TB
Sentinel-2 Collection 1 L1C**	1 000 TB
Sentinel-2 Collection 1 L2A*	19 043 TB
Sentinel-2 L2A MAJA Europe	700 TB
Sentinel-3 OLCI L1 EFR	1 085 TB
Sentinel-3 SLST L1 RBT	1 543 TB
Sentinel-5P L1B	1 595 TB
Sentinel-5P L2	210 TB
Landsat TM C2 L2	1 350 TB
Landsat ETM C2 L2	1 525 TB
Landsat OLI/TIRS C2 L2	2 646 TB
MODIS	1 000 TB
VIIRS	135 TB



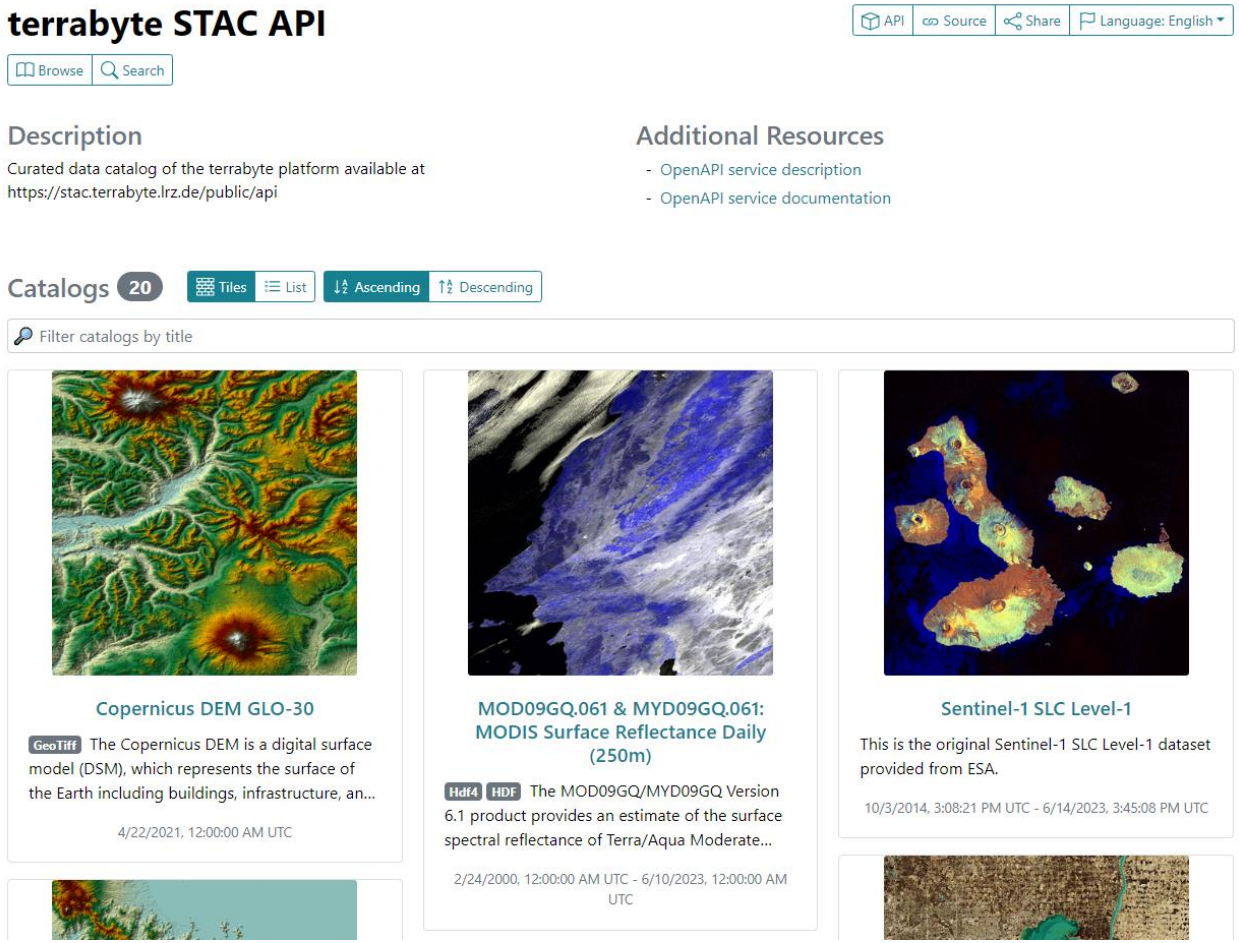
* not yet completely available

** on-demand capacity

… more to come (e.g., PlanetFusion, PlanetScope, TSX, EnMAP, DESIS)

Data Availability
Data access via STAC

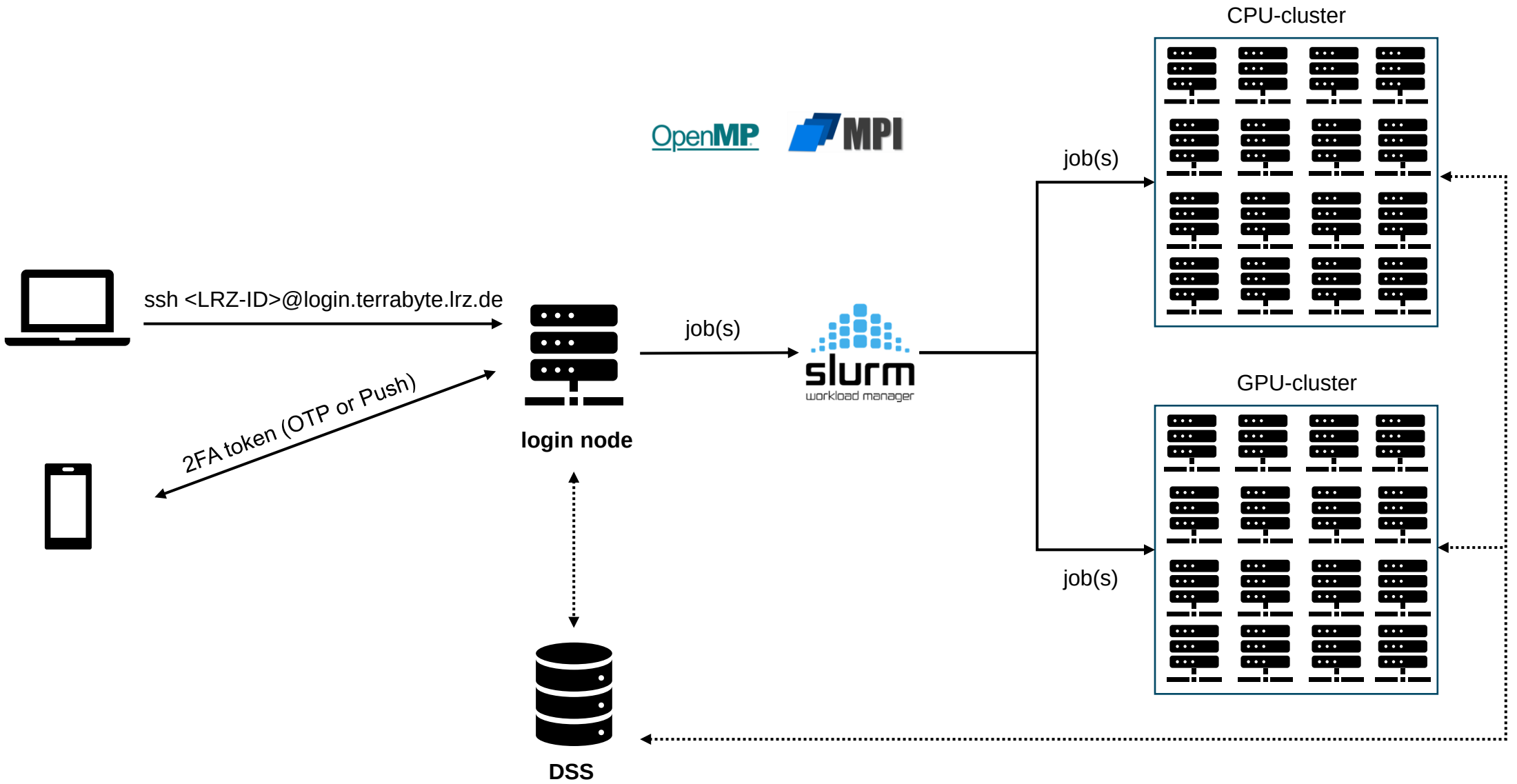
https://stac.terrabyte.lrz.de/browser



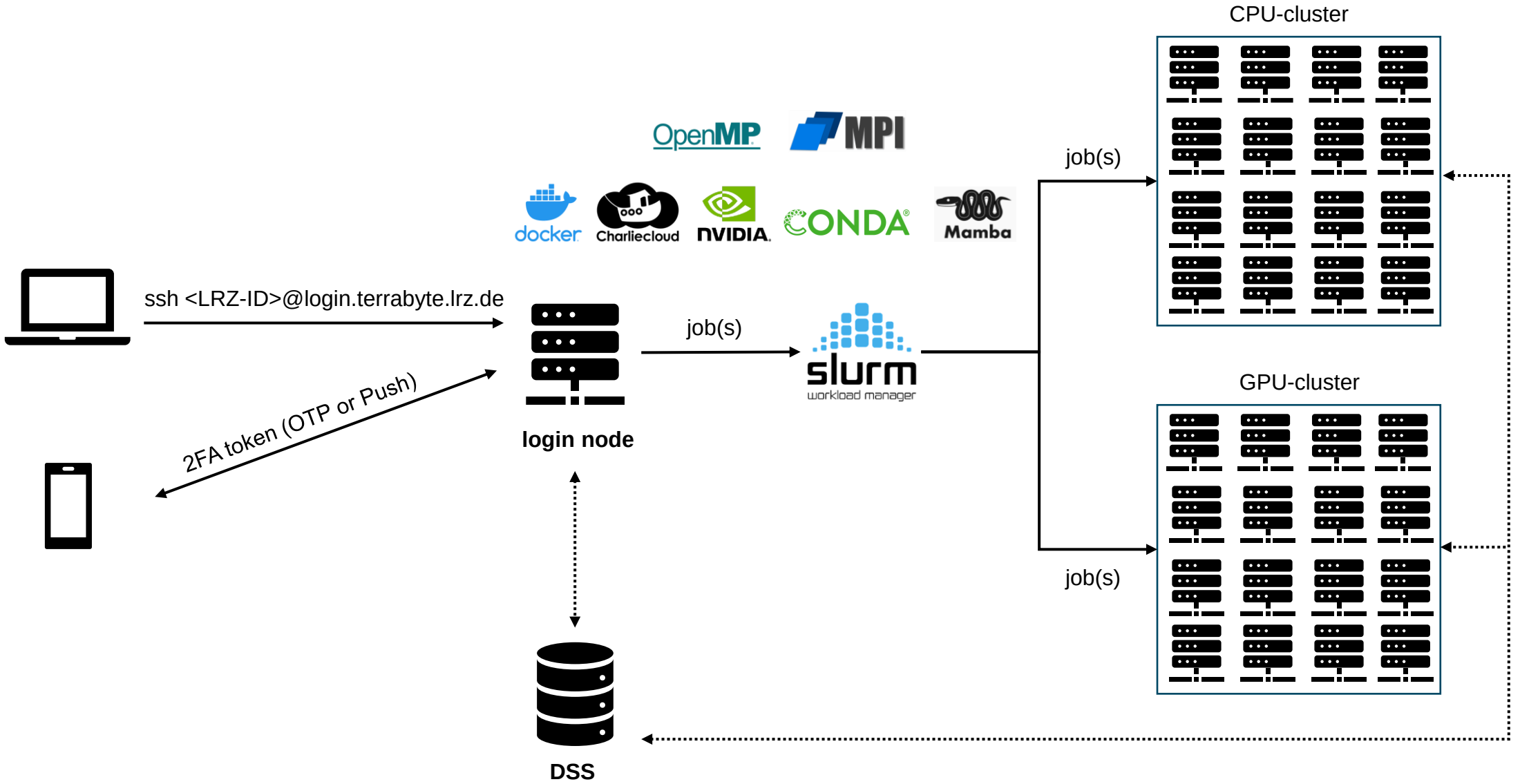
https://stac.terrabyte.lrz.de/public/api

```python
import json
import os
from pystac_client import Client

lon1=29.327168
lat1=-11.745695
lon2=40.445137
lat2=0.985788

outdir="/dss/dsshome1/lxc0F/di46riq/Scripts/S1_ARD_SNAP/" # dir where to put scene list
indir="/dss/dsstbyfs01/pn56su/pn56su-dss-0008/Sentinel-1/GRD/." #main directory of input data

# find COP-DEM tiles that cover scene
# STAC API root URL
URL = 'https://stac.terrabyte.lrz.de/public/api'
# custom headers
headers = []
cat = Client.open(URL, headers=headers)

# limit sets the # of items per page so we can see multiple pages getting fetched
search = cat.search(
    collections = "sentinel-1-grd",
    bbox=[lon1, lat1, lon2, lat2],
    datetime="2020-01-01/2020-03-31",
    query={
        "sar:instrument_mode": {
            "eq": "IW"}
    }
)

# PySTAC ItemCollection
items = search.item_collection()
```

HPC
Submit massive parallel processing jobs

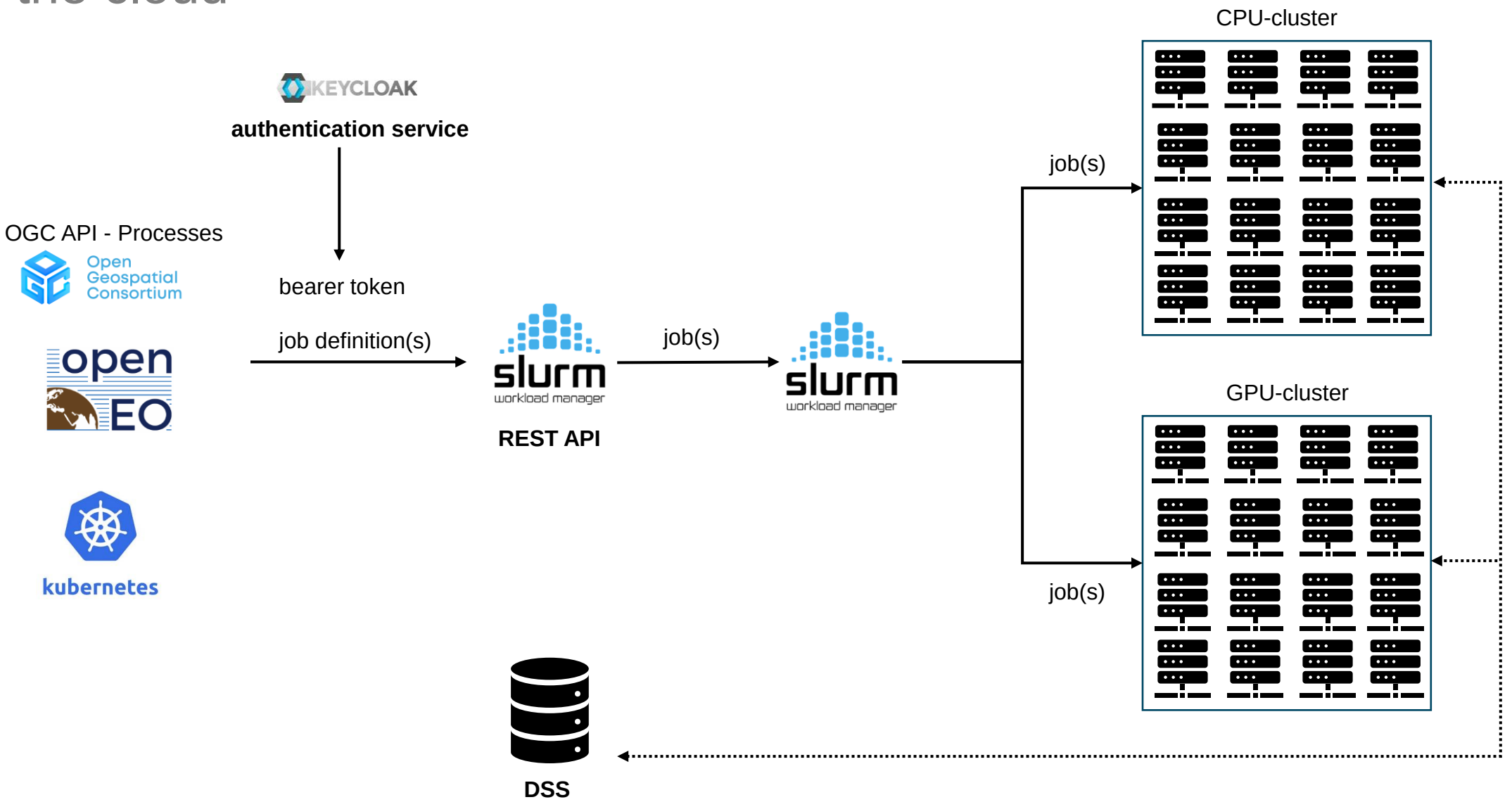


HPC +
Allow users to bring their software environments



HPC + Cloud hybrid

Interact with other web processing services in the cloud

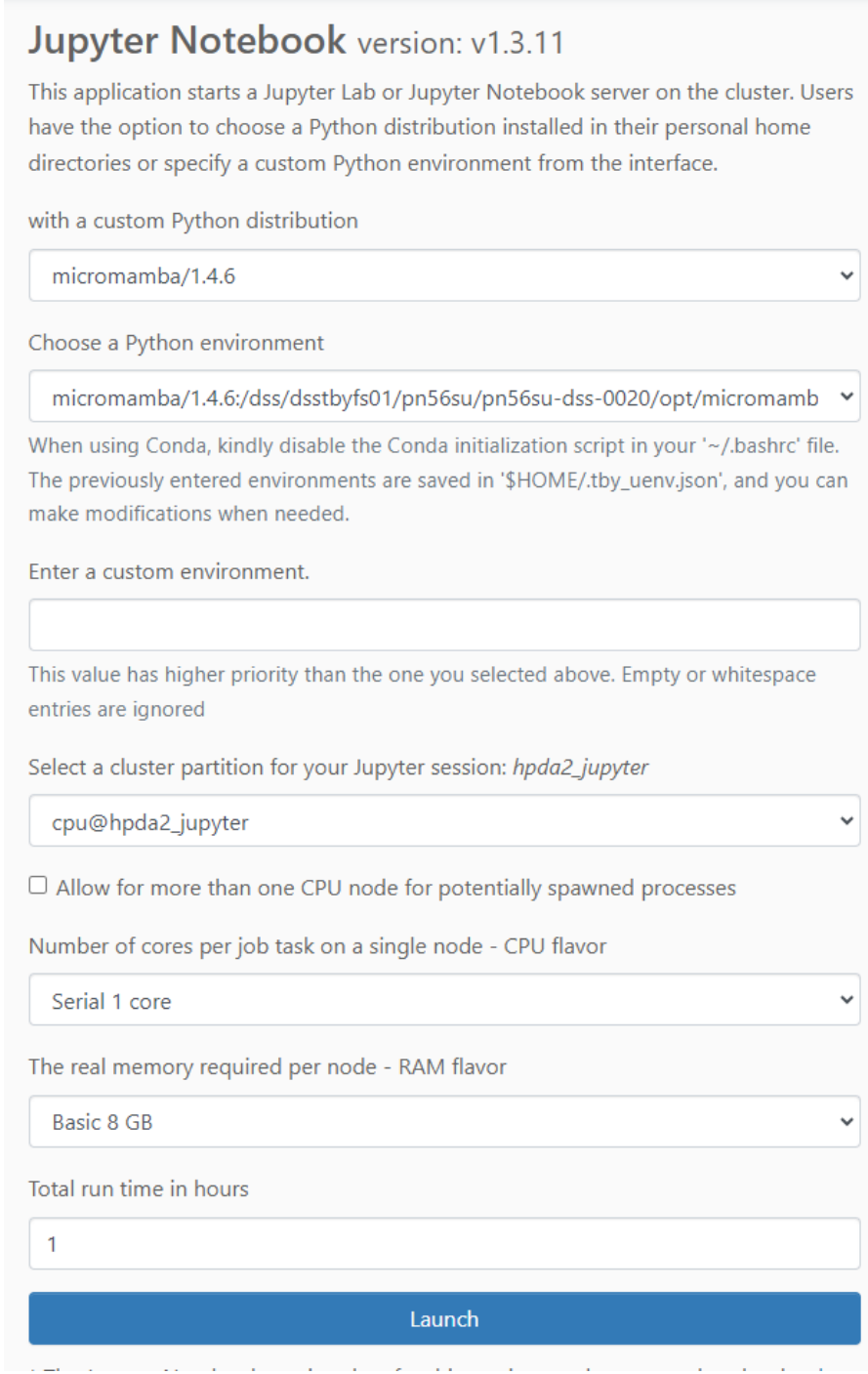


Compute Portal
GUI-based access to software apps

https://portal.terrabyte.lrz.de

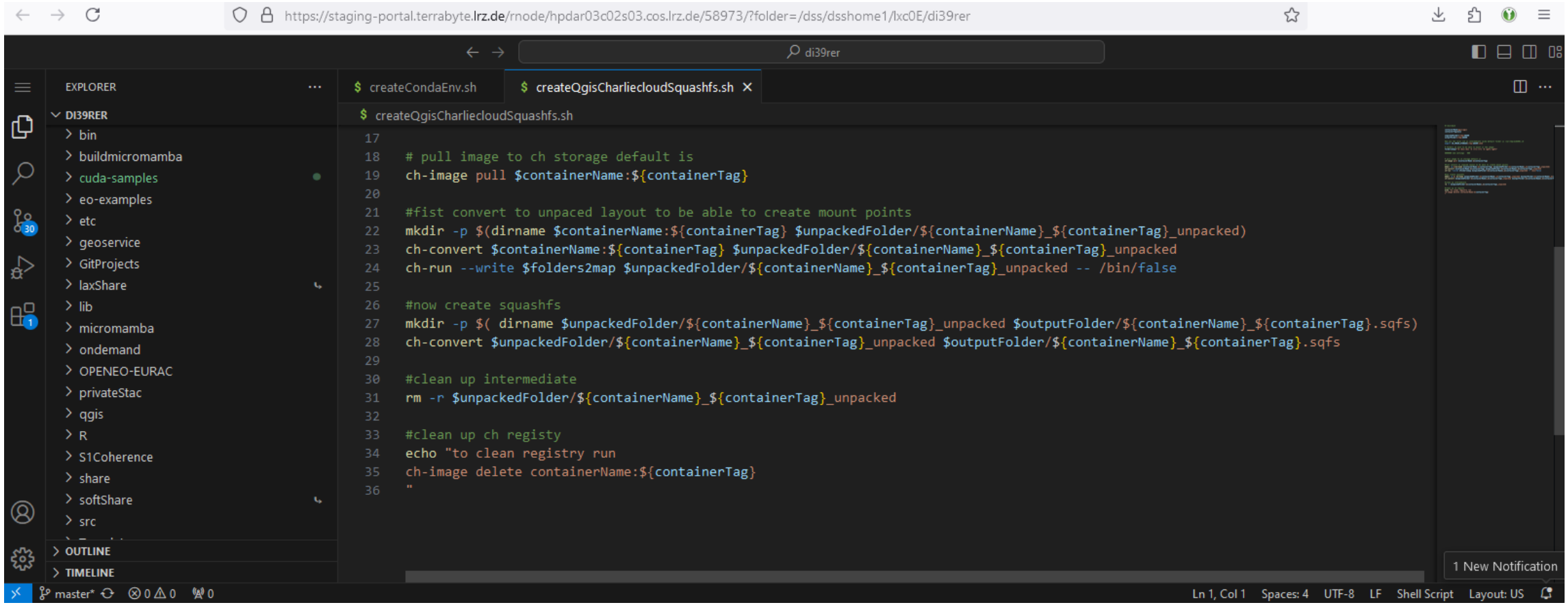
Compute Portal
Example: Jupyter Notebooks

- Pre-defined Python environments
- Users can create their own environments
  - Micromamba
  - Conda
  - Charliecloud container images
- Users can select computing requirements (CPUs, GPUs, RAM)
- Notebooks launched on terrabyte HPC with fast access to data

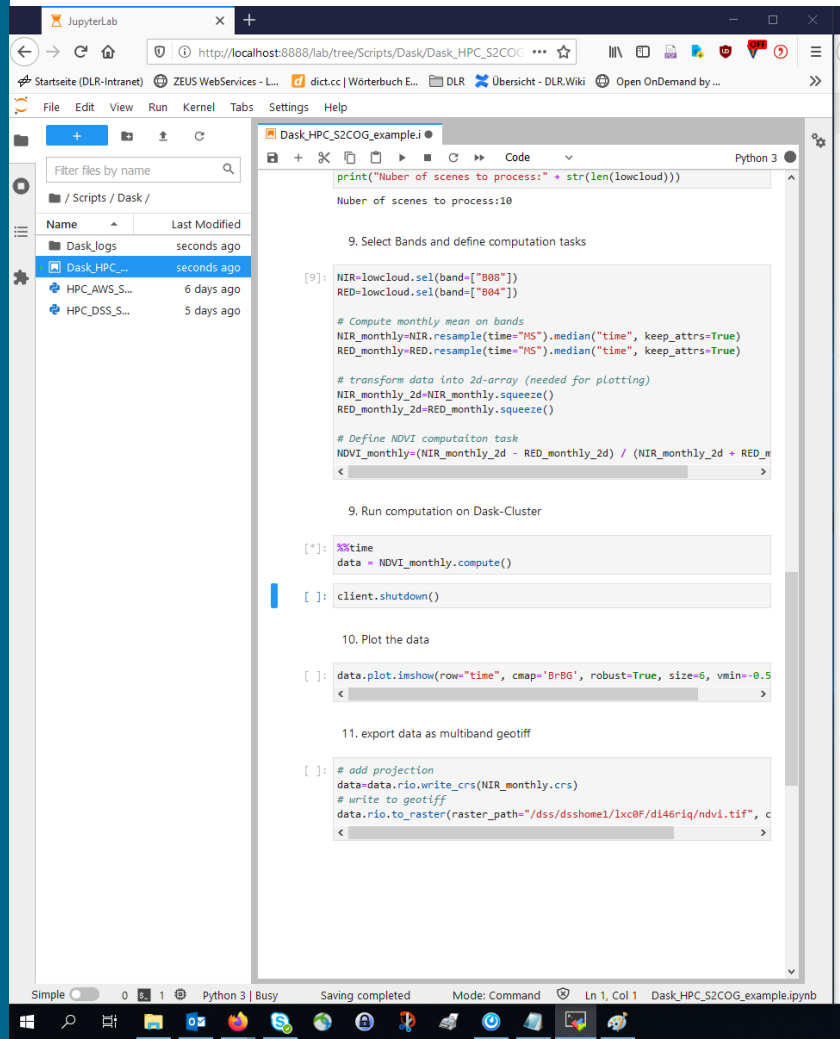


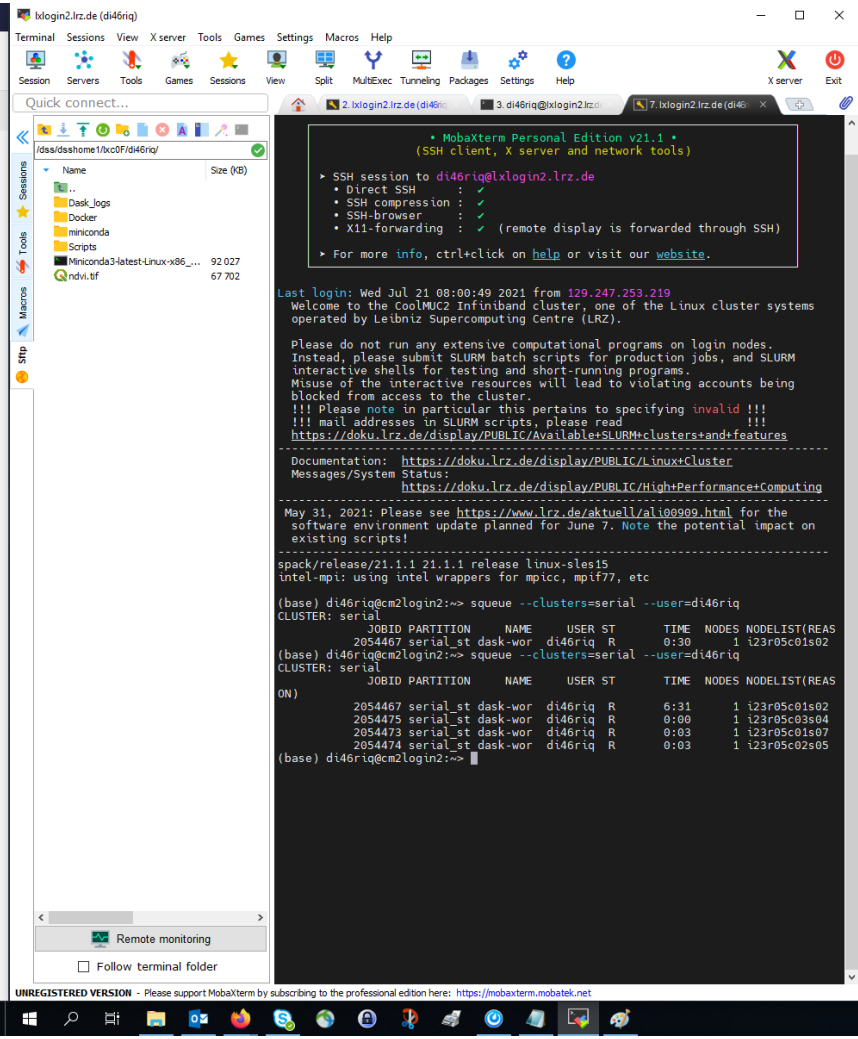
Compute Portal
Example: VS Code Code Server

- VS Code launched in Browser with direct access to Code in terrabyte



Compute Portal
Parallel computation with xarray - dask





User Support
Various possibilities to interact and learn

Service Desk

Jira Service Management

https://servicedesk.terrabyte.lrz.de

servicedesk@terrabyte.lrz.de

Forum

Discourse

https://support.terrabyte.lrz.de

Documentation

https://docs.terrabyte.lrz.de

Workshops & Cafés

Example Notebooks + Python libraries

https://github.com/DLR-terrabyte

What is next?

- Interoperable algorithms (e.g., OpenEO, OGC EO Application Packages)
  - e.g., sometimes requested by ESA project funding
  - Supports open science principles
- User-specific interfaces
  - On-demand processing (e.g., S2-MAJA, Sentinel-1 NRB)
  - User-defined processing workflows based on HPC
  - User-defined data access & visualization (e.g., Google Earth Engine like, OpenEO)
- Kubernetes cluster
  - terrabyte services run as scalable services in the cloud
  - Users can run services on Kubernetes
- Alignment with ESA's Earth Observation Exploitation Platform Common Architecture (EOEPCA+)
  - Design and development of resource & data management and data cube services (together with EOX IT Services, EOFarm, Development Seed)
  - Contribute HPC processing use cases (OpenEO, OGC API Processes)

How do you get access?

- As a DLR-member you can create an account within minutes and get instant access
- Non-DLR-members can get access, if they are in a collaborative scientific project with DLR

Want to try terrabyte?

Collaborate with our colleagues! ☺

Peter Friedl
peter.friedl@dlr.de

Jonas Eberle
jonas.eberle@dlr.de

servicedesk@terrabyte.lrz.de