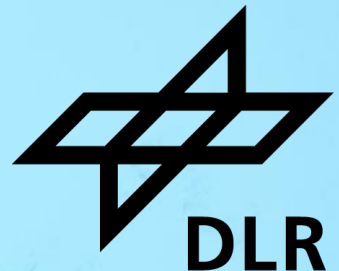


# terabyte: A "Cloud-Like" HPC System for Addressing Earth Observation Challenges

Peter Friedl and Jonas Eberle

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

23.06.2025



# terrabyte

## 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                                    |
|------------------------------|---------------------------------------------------------------------------|-------------------------------------------------------|
| <b>Performance</b>           | No virtualization overhead ✓                                              | Virtualization overhead !                             |
| <b>Elasticity</b>            | Limited & queued !                                                        | Highly elastic, scalable ✓                            |
| <b>Costs (for users)</b>     | Low if in-house use ✓                                                     | Pay-as-you-go !                                       |
| <b>Control / Sovereignty</b> | Full institutional control ✓                                              | Dependency on commercial providers (often US-based) ! |
| <b>Flexibility</b>           | Low — fixed environments, software often pre-installed by admins !        | High — users can configure their own environments ✓   |
| <b>User Experience</b>       | 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

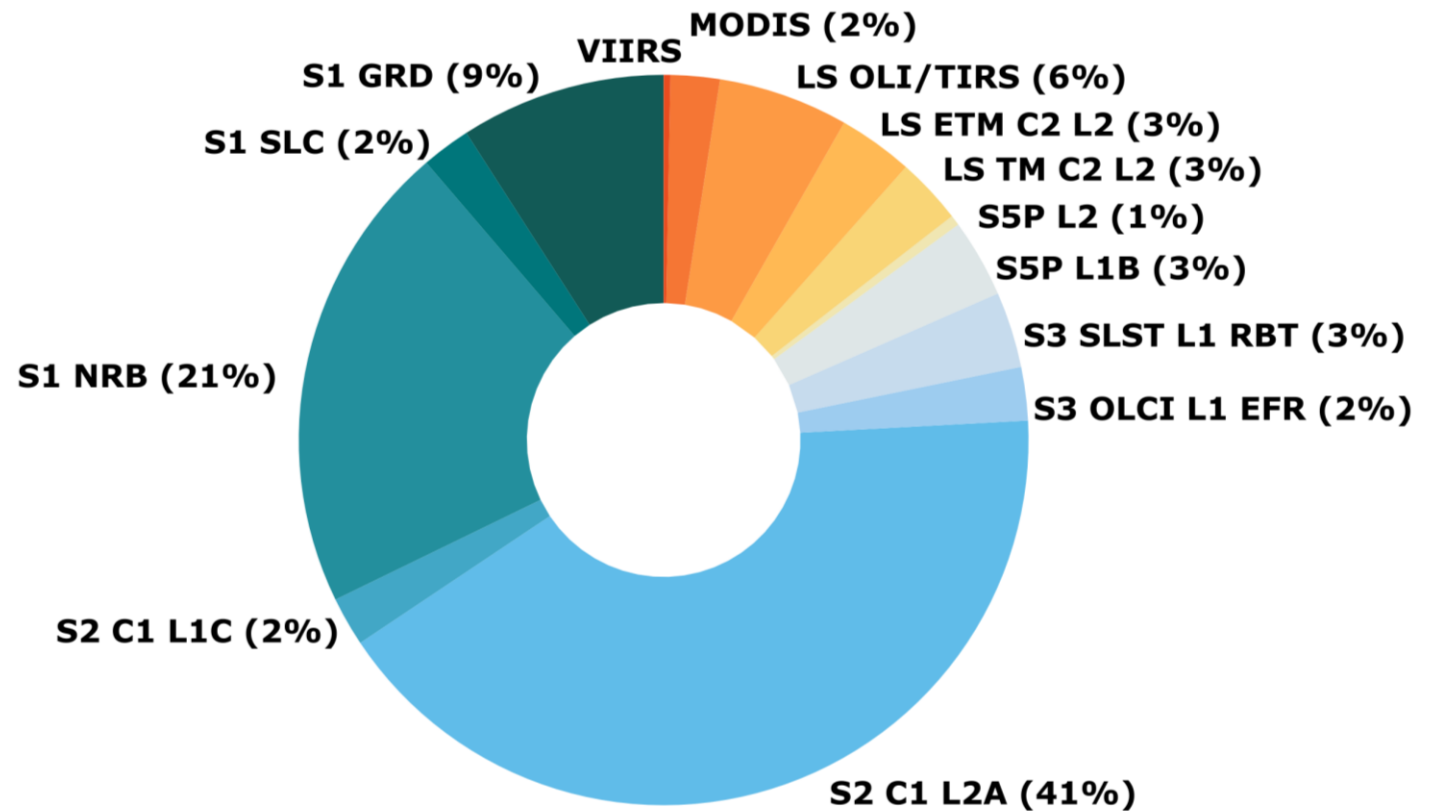
No need to download data



|                               |           |
|-------------------------------|-----------|
| 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.terabyte.lrz.de/browser>

### terabyte STAC API

[Browse](#) [Search](#)

#### Description

Curated data catalog of the terabyte platform available at <https://stac.terabyte.lrz.de/public/api>

#### Additional Resources

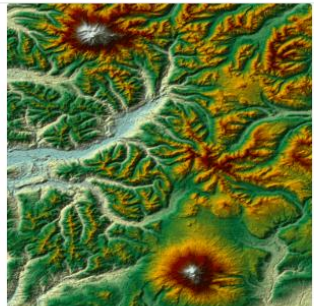
- OpenAPI service description
- OpenAPI service documentation

Catalogs **20**

[Tiles](#) [List](#)

[Ascending](#) [Descending](#)

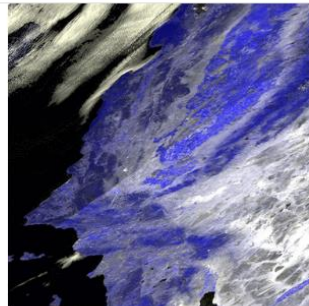
[Filter catalogs by title](#)



#### Copernicus DEM GLO-30

**GeoTiff** The Copernicus DEM is a digital surface model (DSM), which represents the surface of the Earth including buildings, infrastructure, an...

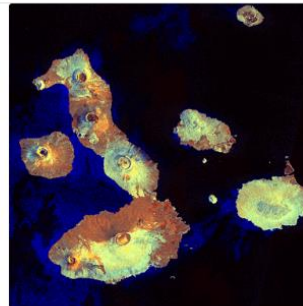
4/22/2021, 12:00:00 AM UTC



#### MOD09GQ.061 & MYD09GQ.061: MODIS Surface Reflectance Daily (250m)

**Hdf4** **Hdf** The MOD09GQ/MYD09GQ Version 6.1 product provides an estimate of the surface spectral reflectance of Terra/Aqua Moderate...

2/24/2000, 12:00:00 AM UTC - 6/10/2023, 12:00:00 AM UTC



#### Sentinel-1 SLC Level-1

This is the original Sentinel-1 SLC Level-1 dataset provided from ESA.

10/3/2014, 3:08:21 PM UTC - 6/14/2023, 3:45:08 PM UTC

```
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.terabyte.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()
```

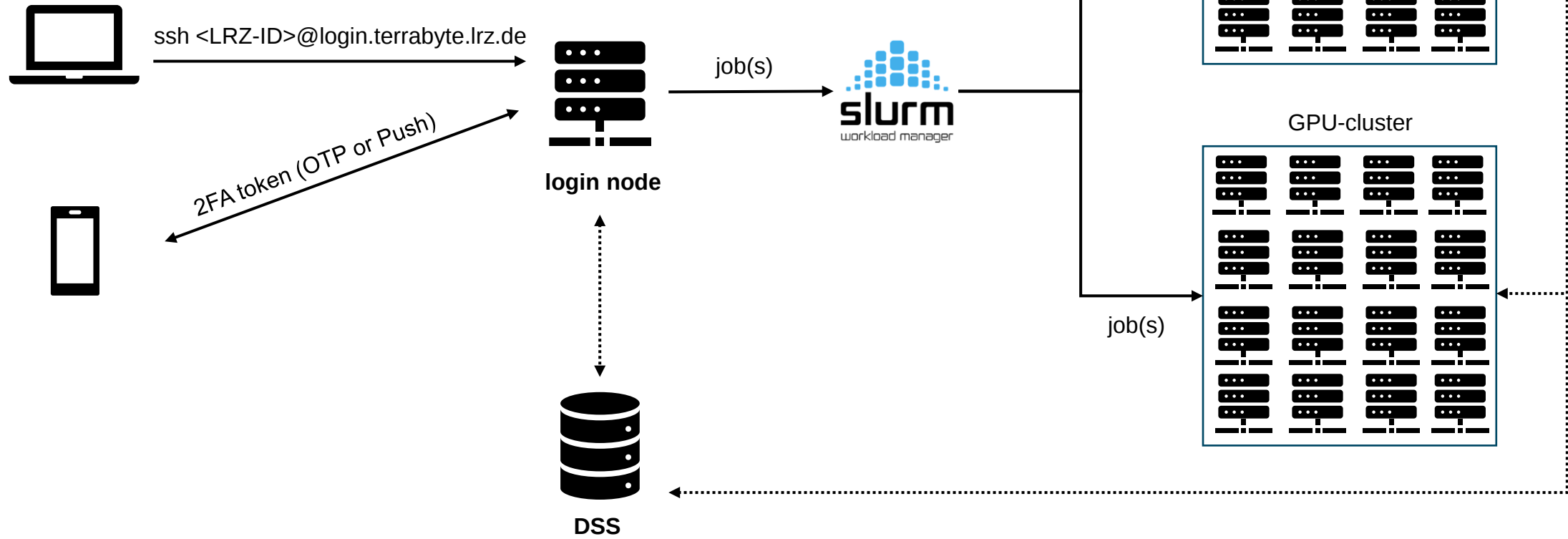


# HPC

Submit massive parallel processing jobs

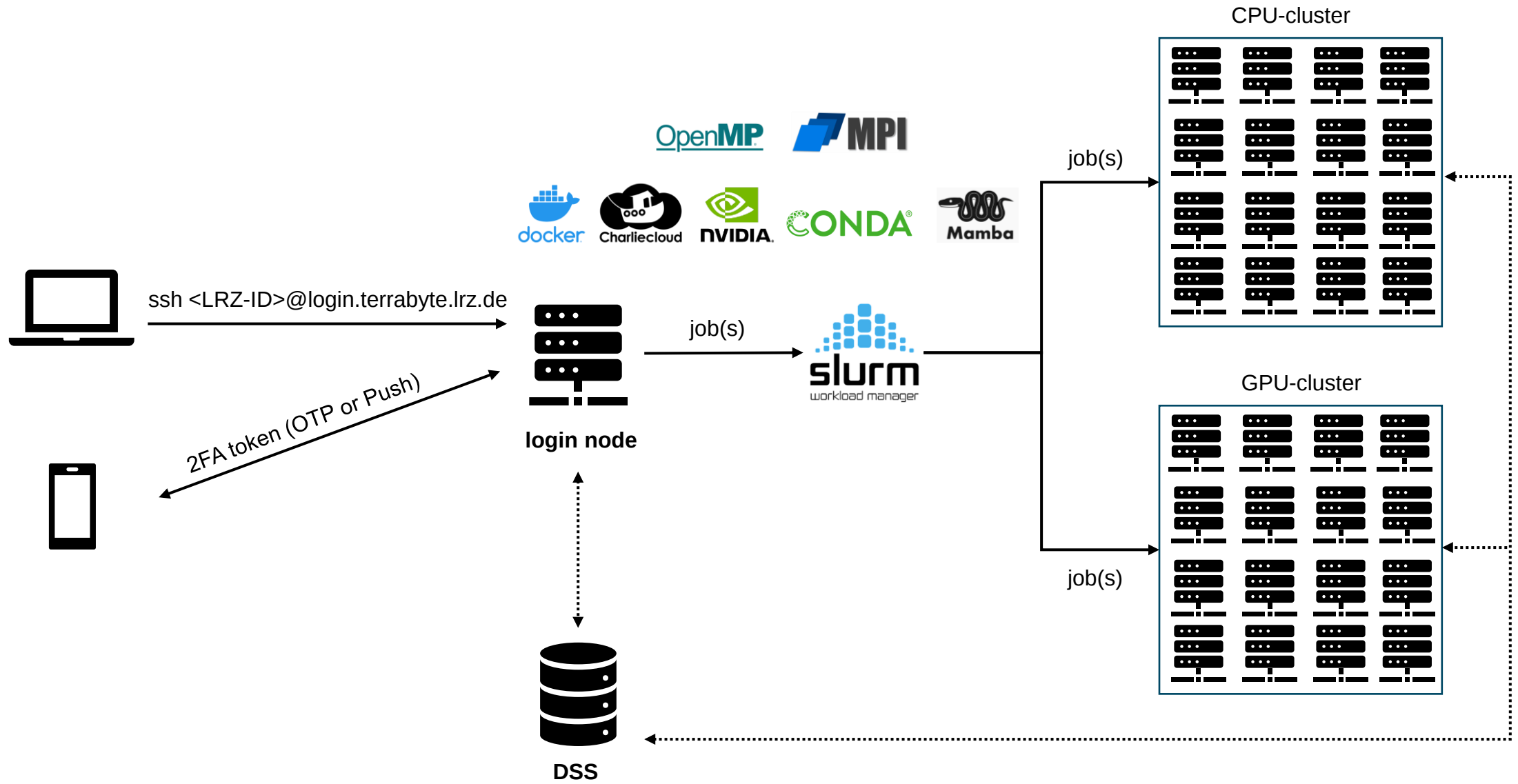


OpenMP MPI



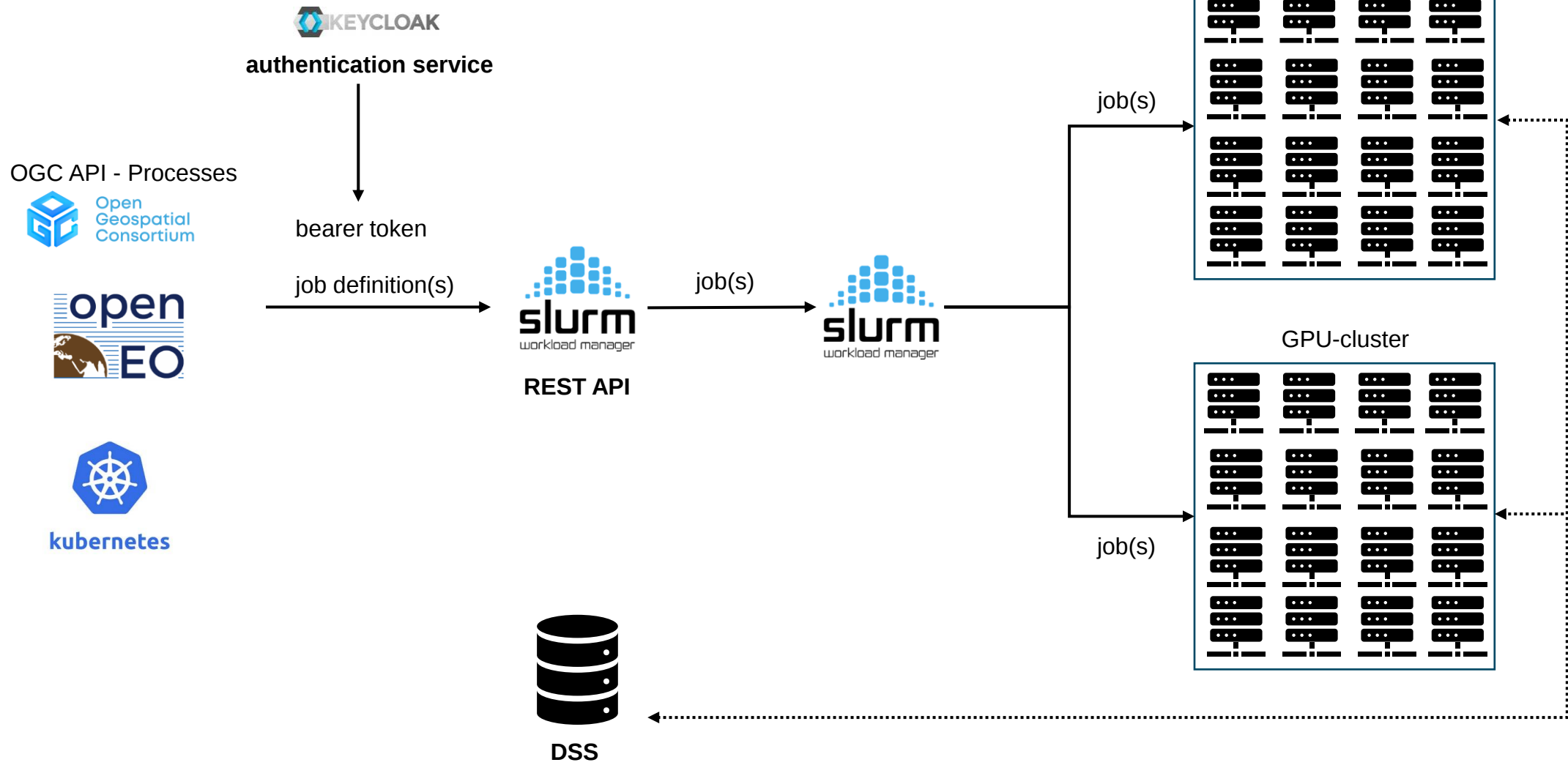
# HPC +

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.terabyte.lrz.de>



integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

## Data



Data Transfer



STAC Browser

## Desktop Apps



Container Compute



QGIS



Remote Desktop

## Development



Code Server for VS Code



Jupyter Notebook



RStudio Server



TensorBoard

## Files



Home Directory  
System Installed App

## Message of the Day

13.01.2025 :: TensorBoard

Interactive app TensorBoard is now available on Portal

13.11.2024 :: Open OnDemand Update

The Portal has been successfully upgraded to the latest version of Open OnDemand

13.11.2024 :: Container Compute

The application uses noVNC to connect to a remote desktop session within a container on a compute node via a VNC server. It provides a basic desktop environment similar to Remote Desktop (Login), along with the flexibility to run custom container applications created with Charliedcloud or Apptainer.

The Remote Desktop Compute has been decommissioned in favor of the new Container Compute.

30.10.2023 :: Portal Upgrade

The Portal has been upgraded to the latest release of Open OnDemand 3, offering a modern interface, improved performance, and enhanced compatibility with diverse HPC systems.



- ## Jupyter Notebook

version: v1.3.11

This application starts a Jupyter Lab or Jupyter Notebook server on the cluster. Users have the option to choose a Python distribution installed in their personal home directories or specify a custom Python environment from the interface.

with a custom Python distribution

micromamba/1.4.6

▼

Choose a Python environment

micromamba/1.4.6;/dss/dsstbyfs01/pn56su/pn56su-dss-0020/opt/micromamb

▼

When using Conda, kindly disable the Conda initialization script in your '~/.bashrc' file. The previously entered environments are saved in '\$HOME/.tby\_uenv.json', and you can make modifications when needed.

Enter a custom environment.

This value has higher priority than the one you selected above. Empty or whitespace entries are ignored

Select a cluster partition for your Jupyter session: *hpda2\_jupyter*

cpu@hpda2\_jupyter

▼

☐ Allow for more than one CPU node for potentially spawned processes

Number of cores per job task on a single node - CPU flavor

Serial 1 core

▼

The real memory required per node - RAM flavor

Basic 8 GB

▼

Total run time in hours

1

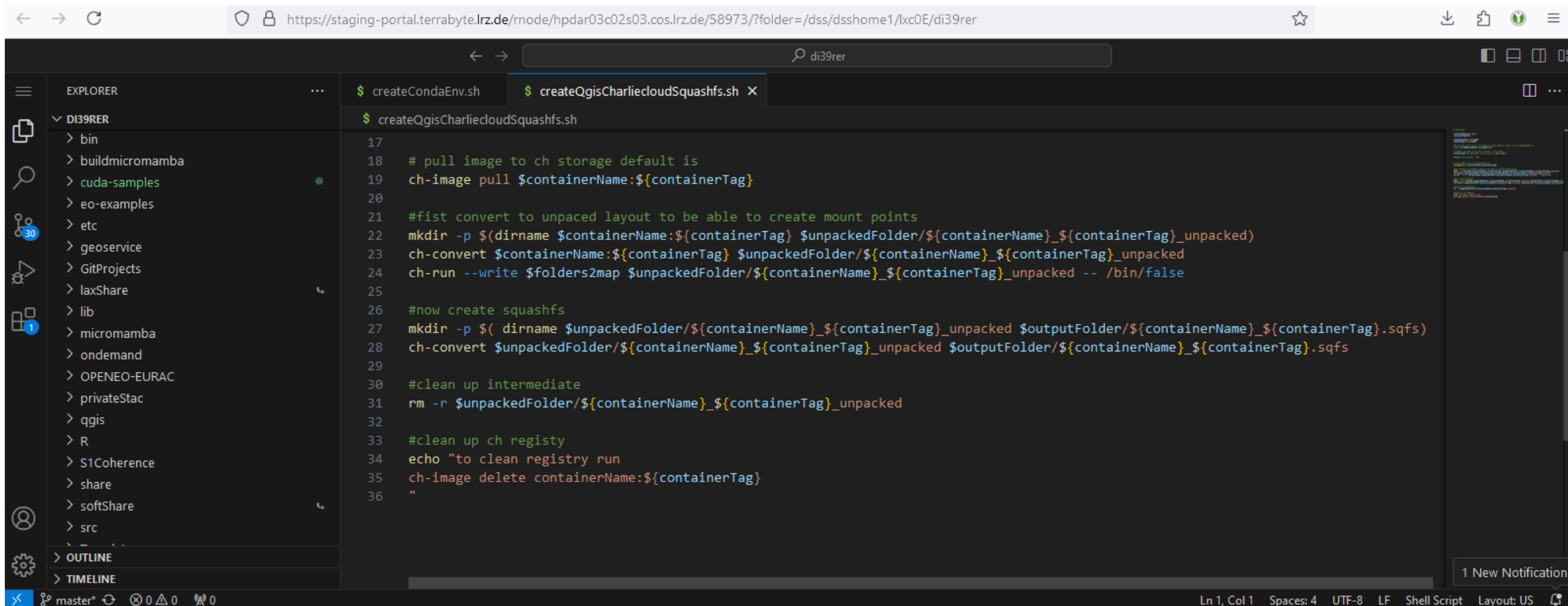
Launch

# 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



The screenshot displays a JupyterLab environment with three main panels:

- Left Panel (File Explorer):** Shows the file structure of the project, including a folder named 'Scripts' and a file named 'Dask\_HPC\_S2COG\_example.ipynb'.
- Center Panel (Code Editor):** Contains a Python script for processing satellite data. The script includes comments and code for selecting bands, computing monthly means, transforming data into a 2D array, and plotting the results. The code is as follows:

```
print("Number of scenes to process:" + str(len(lowcloud)))

Number of scenes to process:10

9. Select Bands and define computation tasks

[9]: NIR=lowcloud.sel(band=["890"])
RED=lowcloud.sel(band=["894"])

# Compute monthly mean on bands
NIR_monthly=NIR.resample(time="MS").median("time", keep_attrs=True)
RED_monthly=RED.resample(time="MS").median("time", keep_attrs=True)

# transform data into 2d-array (needed for plotting)
NIR_monthly_2d=NIR_monthly.squeeze()
RED_monthly_2d=RED_monthly.squeeze()

# Define NDVI computation task
NDVI_monthly=(NIR_monthly_2d - RED_monthly_2d) / (NIR_monthly_2d + RED_monthly_2d)

10. Plot the data

[ ]: client.shutdown()

11. export data as multiband geotiff

[ ]: # add projection
data=data.rio.write_crs(NIR_monthly.crs)
# write to geotiff
data.rio.to_raster(raster_path="/dss/dsshome1/lxc0f/di46riq/ndvi.tif", c=1)
```
- Right Panel (Dask Status Dashboard):** Provides a visual overview of the Dask cluster's status. It includes a 'Status' tab showing 'Bytes stored: 7.54 GiB' and a 'Workers' tab showing a bar chart of 'Bytes stored per worker'. The 'Tasks Processing' tab shows a Gantt chart of task execution. The 'Task Stream' tab shows a detailed view of task execution over time.

Below the Dask Status dashboard, a terminal window displays the output of the script, showing the progress of the computation and the final result of the NDVI calculation.

# User Support

Various possibilities to interact and learn



## Service Desk

⚡ Jira Service Management

<https://servicedesk.terrabyte.lrz.de>

[servicedesk@terrabyte.lrz.de](mailto:servicedesk@terrabyte.lrz.de)

## Forum



<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](mailto:peter.friedl@dlr.de)

Jonas Eberle  
[jonas.eberle@dlr.de](mailto:jonas.eberle@dlr.de)

[servicedesk@terrabyte.lrz.de](mailto:servicedesk@terrabyte.lrz.de)

