FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION (ITC)

UNIVERSITY OF TWENTE.

USING FOSS TO DEVELOP AND OPERATE A GEOSPATIAL COMPUTING PLATFORM

dr. ing. Serkan Girgin MSc
s.girgin@utwente.nl

20 October 2021

Center of Expertise in Big Geodata Science (CRIB) is a horizontal facility established in **March 2020** to **enable** the better use of **geospatial cloud computing and big data technologies** in education, research, and institutional strengthening activities at **ITC**.

Mission

Collect, develop, and share **operational know-how** on cloud computing and big data technologies to solve large-scale geospatial problems.

Vision

Position UT/ITC as a globally renowned center of excellence in **geospatial cloud computing and big data** science.

https://itc.nl/big-geodata

Motivation

Not all research problems require **cloud computing** and **big data** technologies!

- ITC is heterogeneous with respect to interests and needs, and for some people the topic **is not and will not be** *interesting.*
- Even if there is *no apparent need or interest*, it is **still important** to have at least *a basic understanding* of the topic since it is **becoming** *a key component* of the geo-information and EO domain.
- **This is an institutional priority.**

Big Data Needs Assessment

- Key findings of the **status quo analysis** and **user needs assessment***
  - **Information** on big data technology (BDT) should be actively communicated to the staff and students
  - **Proficiency** of the staff and students should be improved
  - Easy-to-use **computing infrastructure** should be made available
  - **Research projects** should be enhanced and improved with BDT
  - BDT know-how should be transferred to **alumni and partners**

Modern **computing infrastructure** is necessary not only for big geospatial data analysis, but also for **geospatial data analysis** in general.

UT does not provide a common computing infrastructure
ITC did not have a common (geospatial) computing infrastructure
ITC **departments** have their own computing solutions

- Operation and management practices differ
- Usually managed by staff who have other primary roles
- Usually access is restricted

* Girgin, S. (2020) Big Geodata at ITC: Status Quo and Roadmap

User Groups

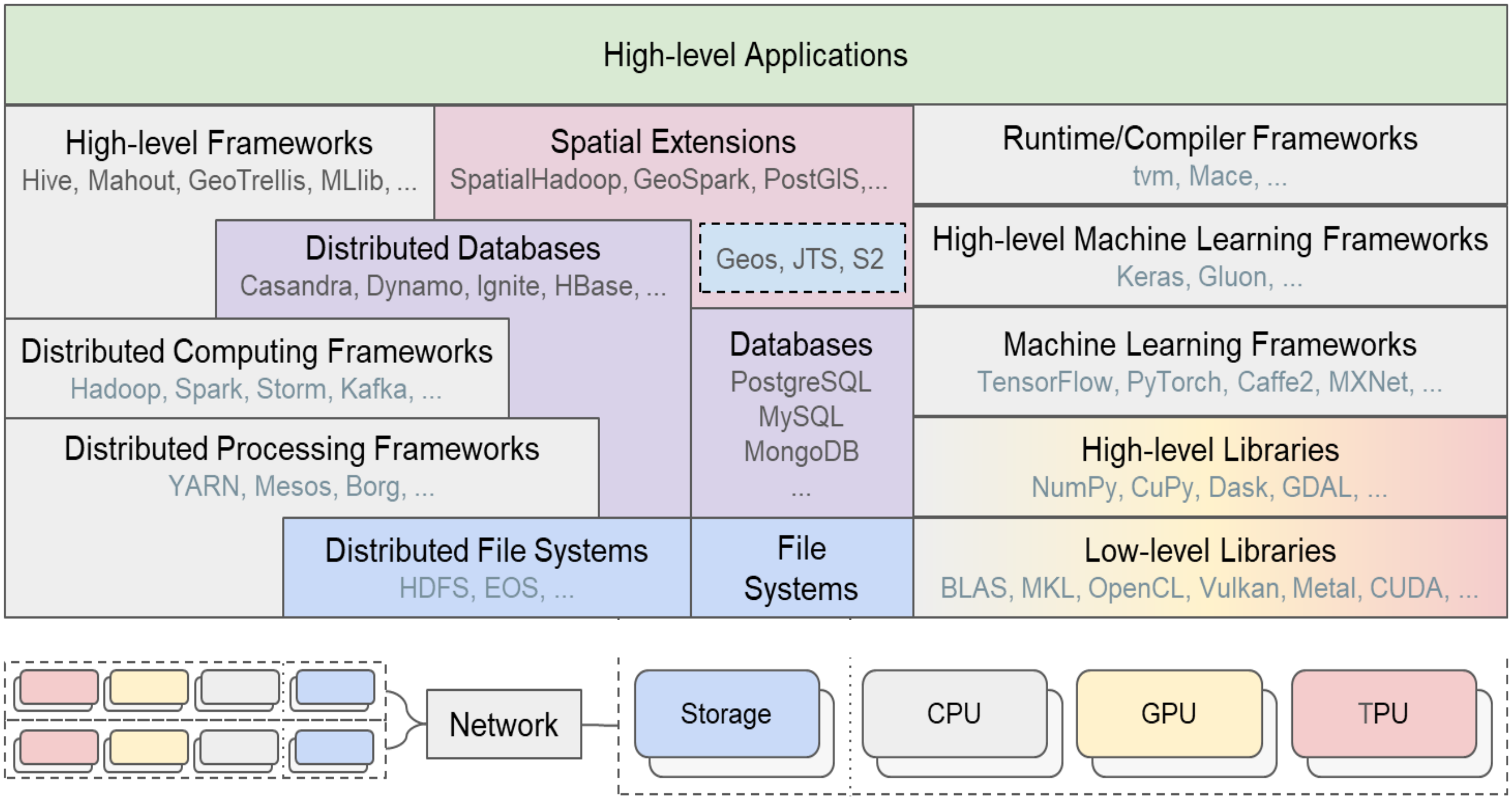
User Group	Needs	Actions Required
Everybody	• Basic information	• Basic training • Periodic updates
Interested students	• Detailed education • Access to infrastructure	• Specialized courses • Infrastructure
Interested staff	• Detailed information with case studies • Guidance for use • Access to infrastructure	• Technology-specific training • Hands-on practices • Support for use • Infrastructure
Inexperienced staff	• Detailed information for problem solving • Guidance for implementation • Access to infrastructure	• Tool-specific training • Hands-on practices • Problem-specific support • Infrastructure
Experienced staff	• Advanced training for early adaption and efficiency • Access to advanced infrastructure	• Tool-specific advanced training • Technology-specific support • Advanced infrastructure

ITC Geospatial Computing Platform

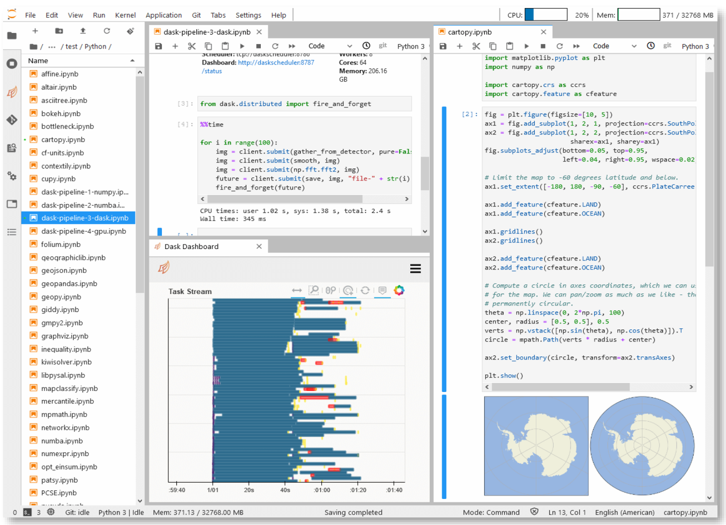
- Designed to serve primary activities identified by the needs assessment:
  - **Self learning**
  - **Exploratory research**
  - **Education**
- Design criteria

Highly available	24/7, no queue
Ready to use	Pre-installed software
User friendly	Interactive user interface
GPU enabled	GPU for each user
Distributed-computing friendly	Computing cluster
Low cost	Feasible investment

Target Big Geospatial Data Stack



Innovative Solution



NVIDIA Jetson AGX Xavier Cluster

JupyterHub on Docker Swarm

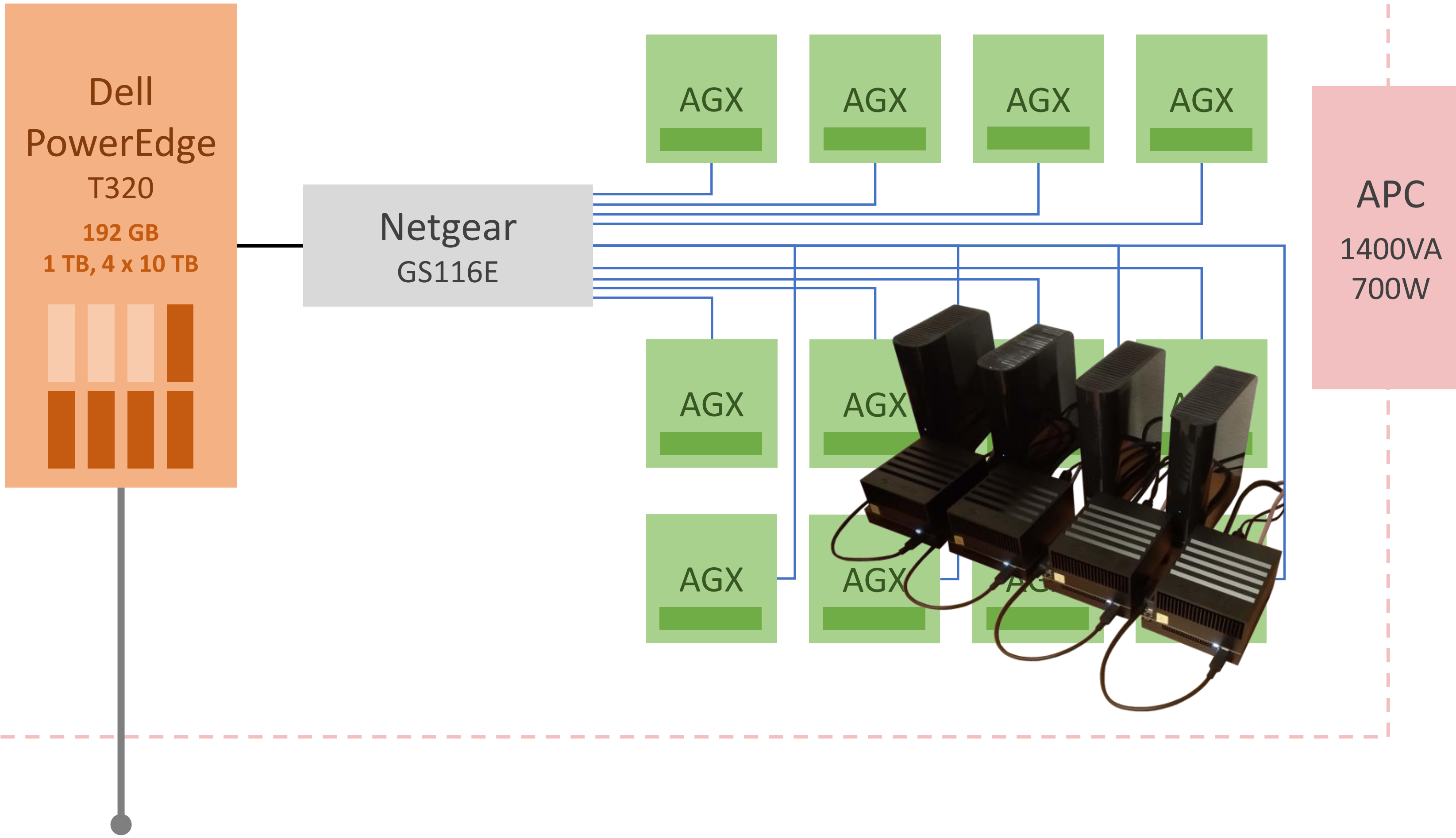
NVIDIA Jetson AGX Xavier

- **8-core CPU**
  (NVIDIA Carmel **ARMv8.2**, 2.26GHz, **NVIDIA L4T**)
- **512-core GPU**
  (**Volta Architecture** with 64 Tensor Cores)
- **32GB memory**
  (256-bit LPDDR4x, 2133MHz, 137GB/s, **Unified**)
- **32GB storage**
  (eMMC 5.1)
- **Dual Deep Learning Accelerator***
- **Vision Accelerator***
- **4x** 4Kp60 video encoder
  (H.264/H.265)
- **2x** 8Kp30 / **6x** 4Kp60 video decoder
  (H.265)
- **Gigabit Ethernet**
  (RJ45)
- **500 GB / 1 TB M.2 NVMe SSD**
  (Samsung EVO 970 Plus, 3GB/s)

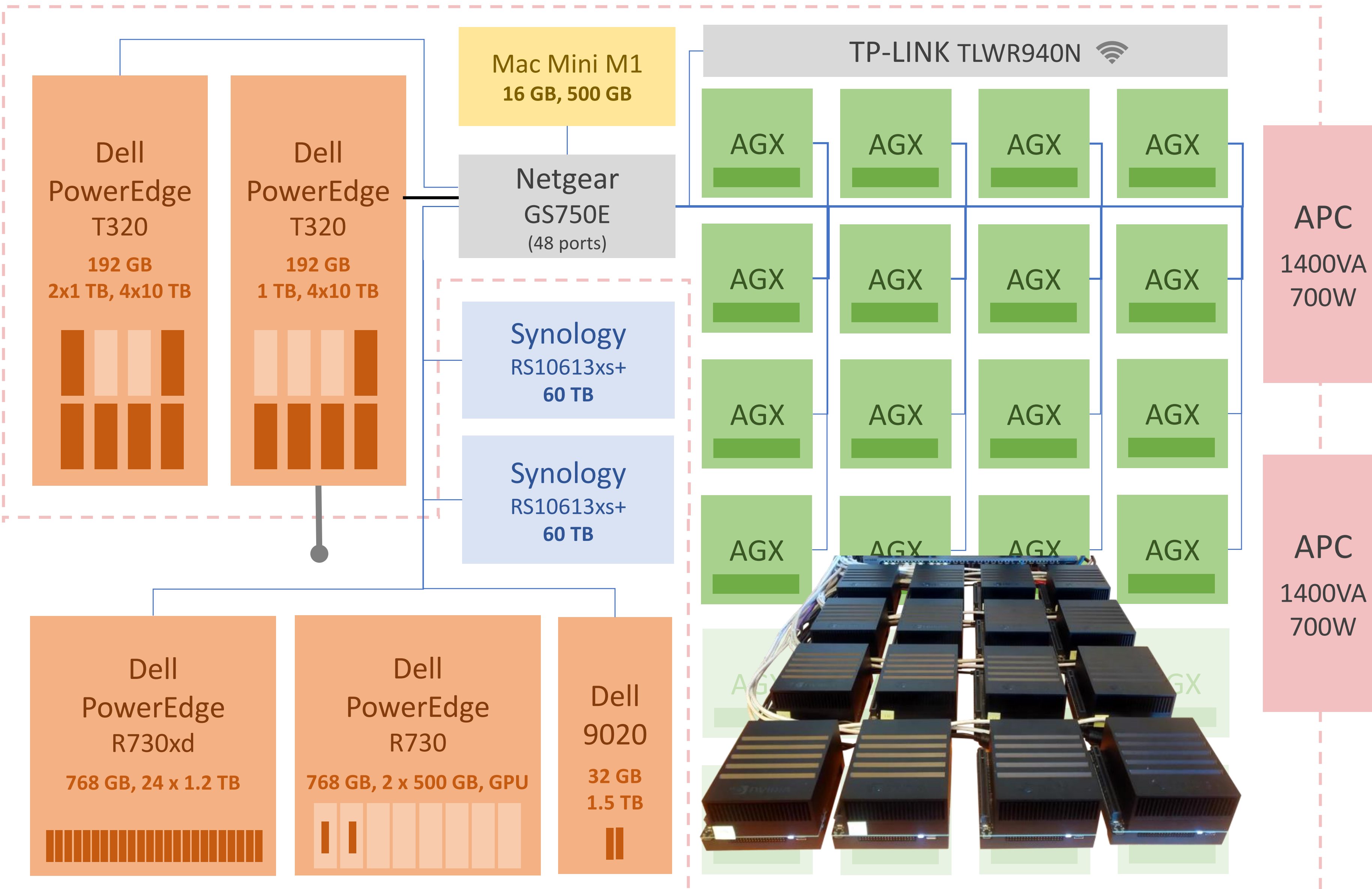
ⓘ https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit
ⓘ https://elinux.org/Jetson_AGX_Xavier

Computing Cluster

We upgrade and repurpose **idle** resources and make them available on the platform for **common use**.



Computing Infrastructure



Current Usage

- Operational since **January 2021**
- **555*** registered users
- **50*** shared workspaces for projects and courses
- **5-25*** concurrent users at a time
- Provided approximately **37,500*** hours of multi-core/GPU computation
- Overall, quite **positive feedback** from a wide-range of use cases
  (4.61/5.00 according to the user survey)
- Other **UT units** (e.g., DCC, BDSI) are interested in having similar platforms
- LISA decided to build a similar platform for **UT-wide use**
  Co-developed by CRIB, available at https://jupyter.utsp.utwente.nl (VPN)
- 4TUResearchData, TU Delft, FAO are interested to **have similar platforms**

* As of 20 October 2021

Platform as a Service

https://crib.utwente.nl

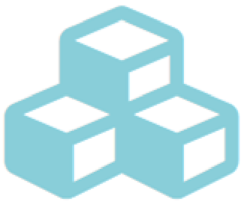
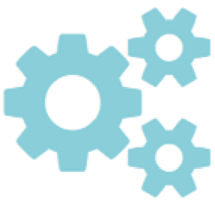
- Based on **open-source** software (Ubuntu, Docker, JupyterHub, JupyterLab, …)
- Accessible through a **web browser** (No software installation or VPN are required)
- **No registration** is required (Login with UT credentials)
- Each user has an individual and isolated **working environment**
- Each user has access to all available* **unit resources**, including **GPU**
- Each user has access to all available* **cluster resources**
- **Replicated storage** with minimum two copies (Hardware failure protection)
- **Distributed storage** for big data processing (HDFS)
- Automatically scales and **balances workload** among the units
- Low **energy footprint** (10-30W per unit)

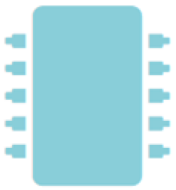
* Resource availability depends on resource usage of other active user

Key Features

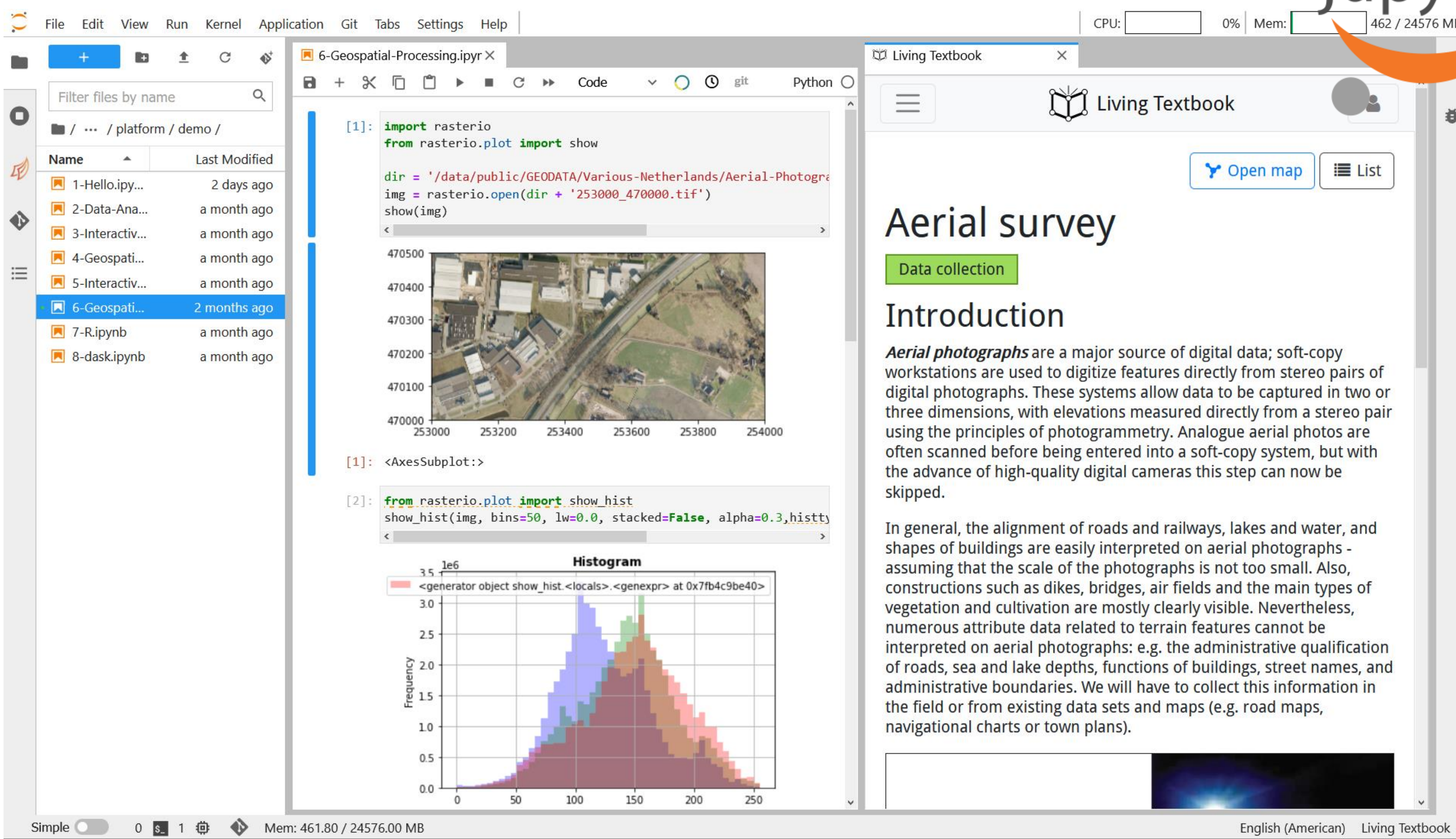
https://crib.utwente.nl

- **Interactive notebook**, **terminal** and **remote desktop** access are available
- Multiple **interactive languages** are supported (Python, R, Julia, Octave, Go, …)
- **Up-to-date** and **optimized** software packages are **ready to use** (No setup required)
- Users can install **additional** packages (e.g., Python, R packages)
- Different **architectures** and **OS-specific applications** are supported (e.g., Windows)
- Distributed computing clusters are **ready to use** (Dask, Apache Spark)
- **Public** assets are shared by all users (e.g., OSM Planet Data)
- **Shared workspaces** allow assets to be shared by selected users
- Access can be granted to **external users**
- **User support** is available
- Provided and maintained by **CRIB** at no extra cost (i.e., free PaaS)



Interactive Access



ⓘ JupyterLab documentation: https://jupyterlab.readthedocs.io/en/stable/

Available Software

https://crib.utwente.nl

hadoop | Apache Spark | DASK | NVIDIA CUDA | OpenCL | Open MPI | OpenMP

TensorFlow | Keras | PyTorch | mxnet | scikit learn | SCIKIT-CUDA | XGBoost | RAPIDS

Coming Soon!

GDAL | PROJ | Cartopy | GeoPandas | GeoPy | Iris | PySAL | R

NumPy | Numba | pandas | xarray | NetworkX | igraph | ONNX RUNTIME | TensorRT

OpenCV | scikit-image | FFMPEG | ImageMagick | GraphicsMagick | matplotlib | FFTW | Apache ARROW

… and many more!

ⓘ Complete lists of more than **750+ Python** and **350+ R** packages are available at `/public/platform`

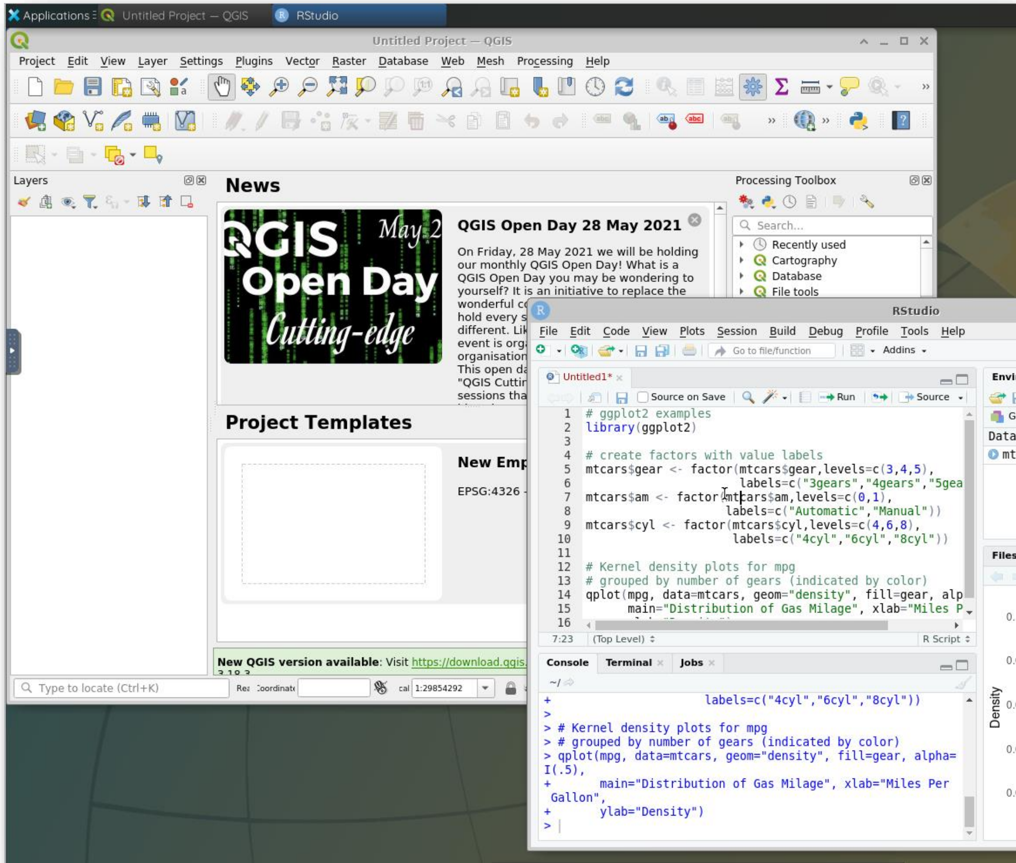
Available Desktop Applications

- QGIS
- GRASS GIS
- SAGA GIS
- SNAP
- **ILWIS 3***
- **ILWIS 4***
- VS Code
- PyCharm
- R Studio
- Netlogo
- GNU Octave
- **MATLAB***
- Glueviz
- Orange Data Mining
- Firefox

TigerVNC

noVNC

WINEHQ



Available Services

https://crib.utwente.nl

GeoServer
Open source server for sharing geospatial data

MapServer
Open source platform for publishing spatial data

PostgreSQL
Open source relational database

MariaDB
Open source relational database

GeoNode
Open source geospatial content management system

Dataverse
Open source research data repository software

Gitea
Open source lightweight code hosting solution

Open Data Kit
Open source platform to collect data quickly, accurately, offline, and at scale

Support Center

https://crib.utwente.nl/support/

UNIVERSITY OF TWENTE. | CRIB

Support Center Home | Knowledgebase | Open a New Ticket | Check Ticket Status

Search our knowledge base [Search]

[Open a New Ticket]

[Check Ticket Status]

Welcome to the CRIB Support Center!

In order to streamline support requests and better serve you, we utilize a support ticket system. Every support request is assigned a unique ticket number which you can use to track the progress and responses online. For your reference we provide complete archives and history of all your support requests.

Quick Access

- Report a Problem
- Shared Workspace Request
- Course Workspace Registration with Canvas Integration
- External Account Request
- Account Removal Request
- Account Transfer Request
- Software Request
- Dataset Request
- Database Request

Featured Questions

How can I access to the platform?

Is it secure?

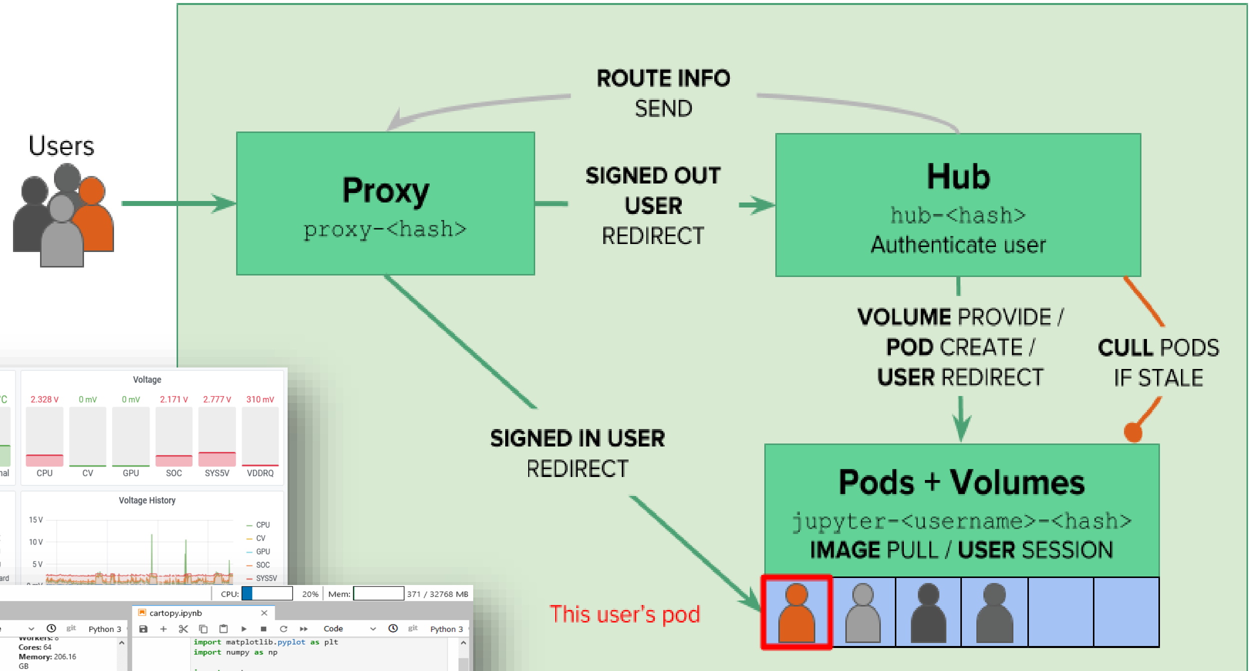
How can I use the platform?

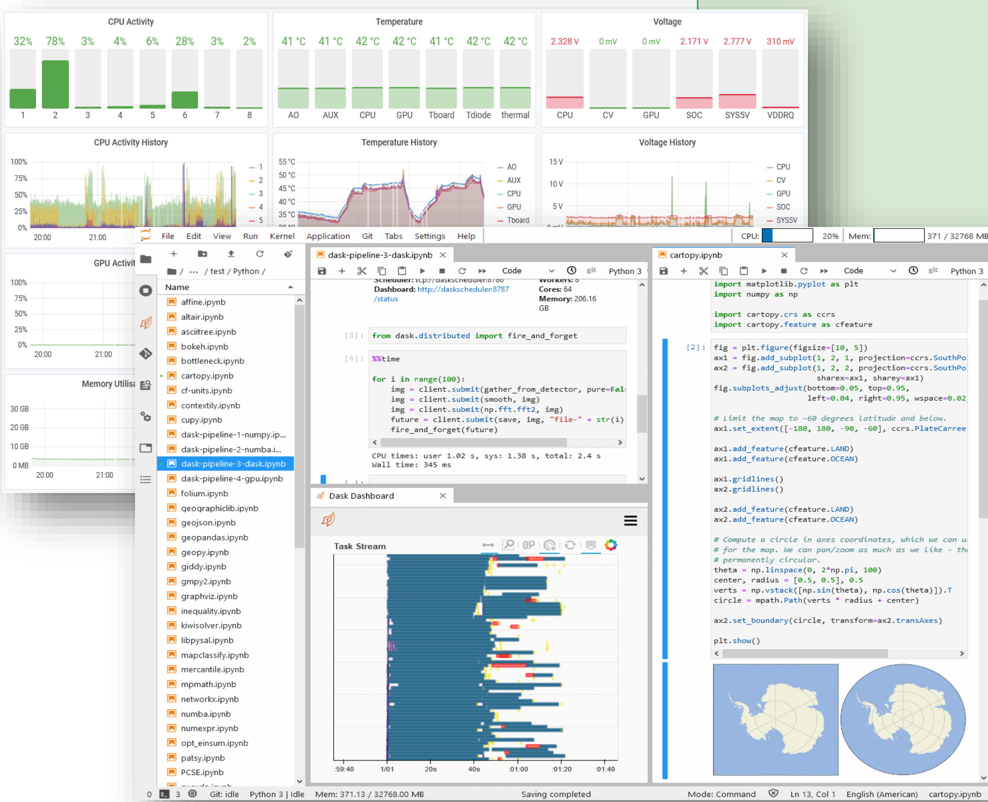
Which programming languages are supported on the platform?

Which libraries and packages are supported by the platform?

osTicket

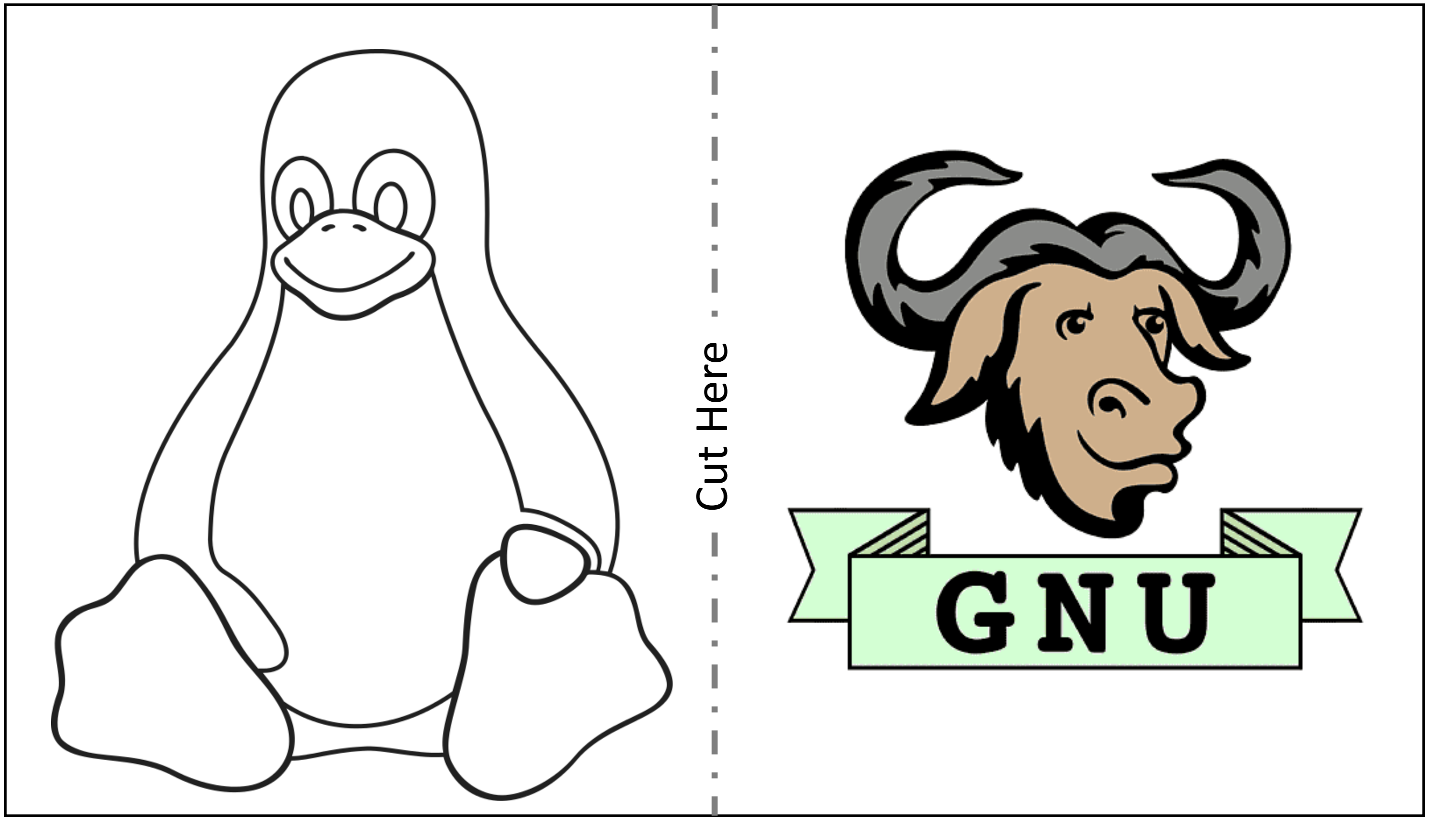
User Access


Users
ROUTE INFO
SEND
Proxy
proxy-<hash>
SIGNED OUT
USER
REDIRECT
Hub
hub-<hash>
Authenticate user
VOLUME PROVIDE /
POD CREATE /
USER REDIRECT
CULL PODS
IF STALE
SIGNED IN USER
REDIRECT
Pods + Volumes
jupyter-<username>-<hash>
IMAGE PULL / USER SESSION
This user's pod

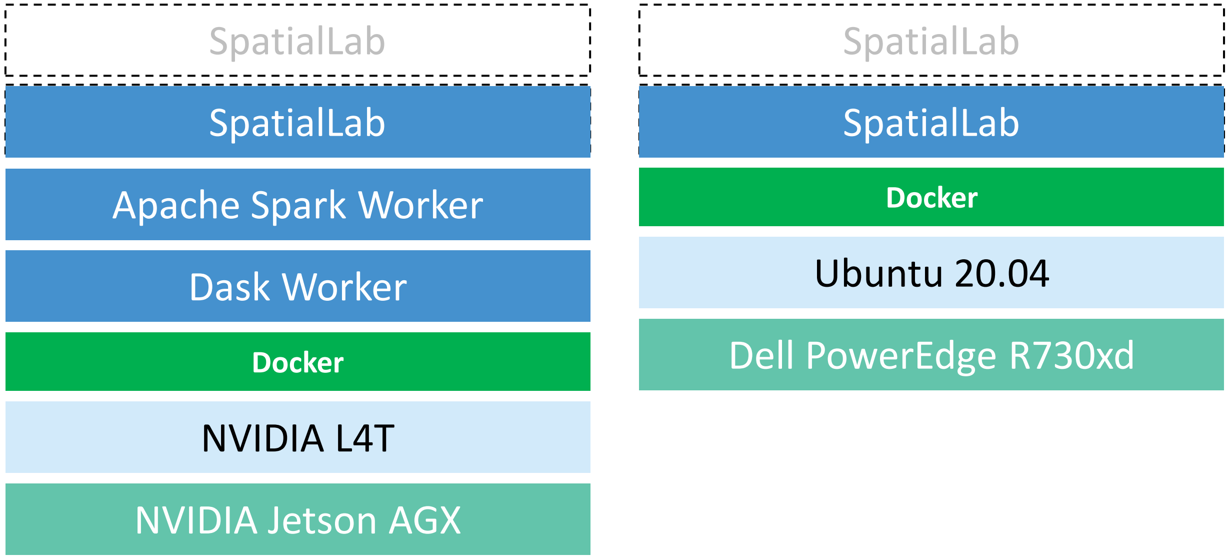


Computing Environment

Container Image (common, volatile) **Binded Volumes** (personal, permanent)



Container Orchestration



- **Minimum impact** on the host units
- **Custom-built** container images for better performance
- **SpatialLab** image: Ubuntu 20.04, All software
  - NVIDIA Jetson AGX : **21.0 GB**
  - Intel x86-64 : **42.9 GB**
  - Intel x86-64 GPU : **53.9 GB**

ⓘ Docker Swarm mode overview: https://docs.docker.com/engine/swarm/

Operation and Maintenance

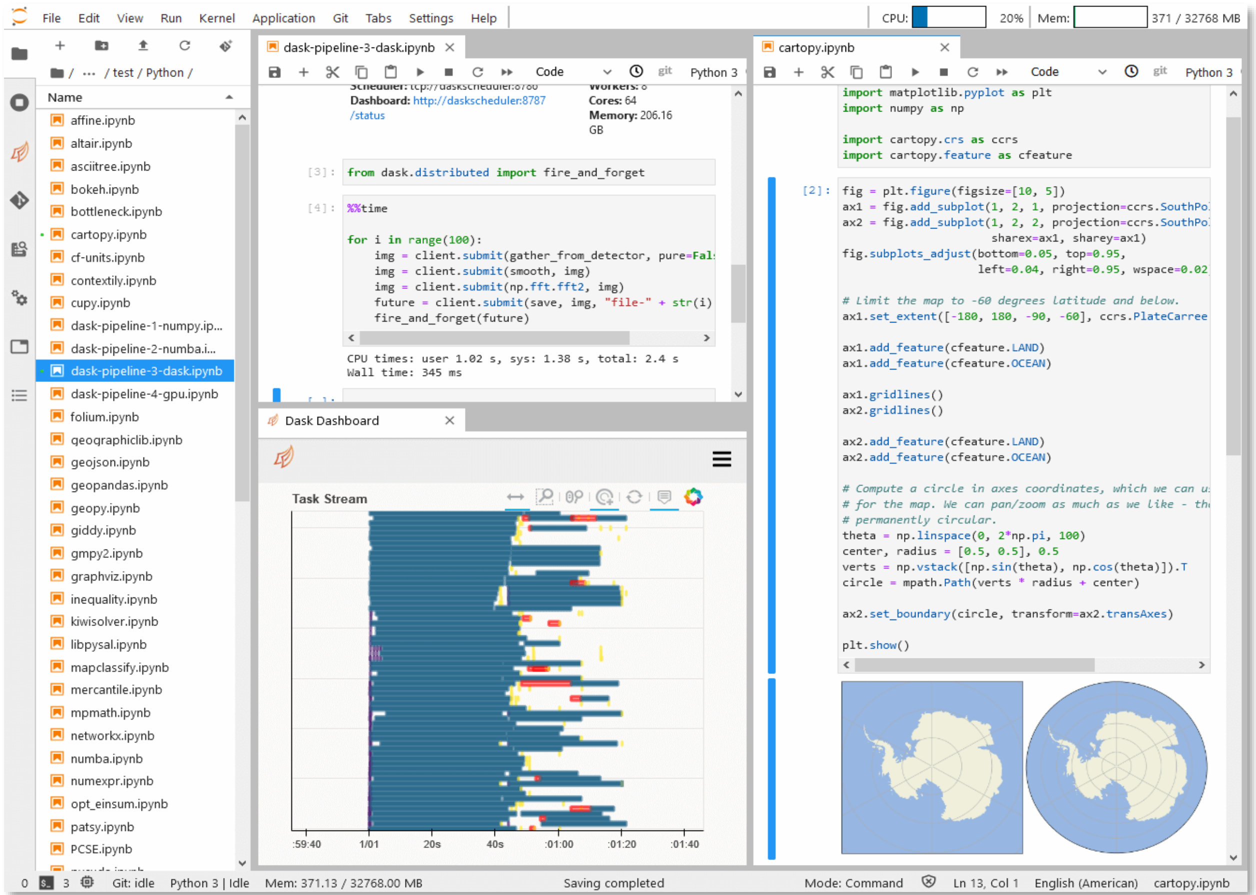
- On-demand and regular (monthly) **rolling updates** (2 restarts in 5 months)
- **Identical** working environment for ARM64 and Intel x86-64 units
- Automated shared workspaces for the **departments**
- Automated shared workspaces for **Canvas courses**
- Automated **notification** to newcomers
- **Daily** check for malicious threads
- **Continuous** resource and performance monitoring

Major Challenges

- Major upgrades
  - Python 3.6 → 3.8
  - JupterLab 2.x → 3.x
- Major issues
  - Kernel crash with GPU
  - Slow-down under heavy use
- Major requests
  - (Ana)Conda
  - More computing accounts
  - Direct access

Quick Demo

https://crib.utwente.nl



ⓘ Available on the platform at `public/platform/demo`

Contact

https://crib.utwente.nl

https://itc.nl/big-geodata

crib-itc@utwente.nl
s.girgin@utwente.nl

@BigGeodata

Big Geodata Newsletter