CENTER OF EXPERTISE IN BIG GEODATA SCIENCE

UNIVERSITY OF TWENTE.

ITC GEOSPATIAL COMPUTING PLATFORM

UT DCC THEMATIC SESSION ON CLOUD COMPUTING

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

18 April 2021

Center of Expertise in Big Geodata Science (CRIB) is a *horizontal facility* that **enables** the better use of **big geodata technology** in *education, research, and institutional strengthening* activities at **ITC**

Mission

Collect, develop, and share **operational know-how** on big data technology to solve large-scale geospatial problems

Vision

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

https://itc.nl/big-geodata

Cloud computing is the on-demand availability of computer system resources, especially **data storage** and **computing power**, *without direct active management* by the user

Cloud Computing Services

- **Software** as a Service (**SaaS**)
  [On-demand software]
  - Provider supplies the infrastructure and platforms that run the applications
  - User uses provided applications through an interface

- **Platform** as a service (**PaaS**)
  - Provider supplies the infrastructure, services, and tools that allow the user to deploy applications
  - User deploys applications and alters settings of the application-hosting environment

- **Infrastructure** as a service (**IaaS**)
  [On-demand hardware]
  - Provider supplies the infrastructure
  - User deploys and run arbitrary software, including OS

- **Function** as a service (**FaaS**)

Currently we are using one!

- R Studio Cloud
- Matlab Online
- Authorea
- …

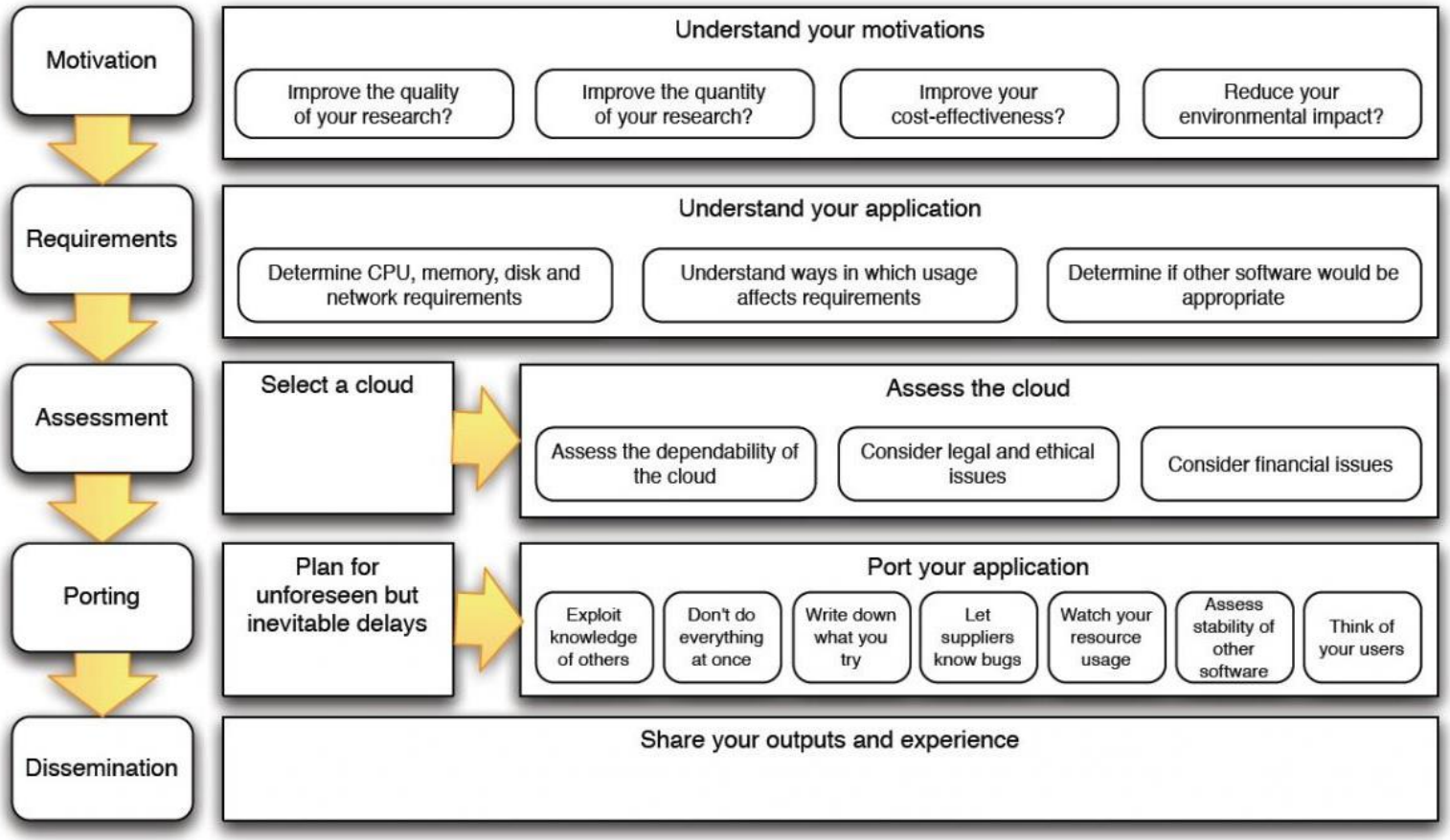
e.g. **ITC Computing Platform**

- Google Colab
- Amazon SageMaker
- Azure ML Studio
- ...

e.g. **LISA VRE**

- Microsoft Azure
- Amazon AWS
- Google Cloud
- …

Moving to the Cloud



Source: [Best practice for using cloud in research (Hong et al., 2018)]

ITC Geospatial Computing Platform

- Operational since **January 2021**
- Currently serves **235** registered users with **6-18** concurrent users at a time
- Designed to serve primary activities identified by a *comprehensive* user needs assessment:
  - **Self-learning**
  - **Exploratory research**
  - **Education**
- Provides **highly-available, easy-to-use** environment with good performance
  - **User-friendly** interface for data analysis and visualization
  - **Ready-to-use** scientific and geospatial analysis software
  - **Parallel and distributed computing** by using high-level frameworks
  - Computing by using special processing units (e.g., **GPU**)

Resources

- **16 x NVIDIA Jetson AGX Xavier** computing units (128 cores, 512 GB)
  - **8-core** CPU (NVIDIA Carmel ARMv8.2, 2.26 GHz)
  - 512-core **GPU** (Volta architecture with 64 Tensor Cores)
  - **32GB** memory (DDR4x, 137 GB/s)
  - 500 GB – 1 TB local storage (NVMe SSD, 3 GB/s)
- **Big data** computing unit
  - 2 x 8-core CPU (Intel Xeon E5-2640, 32 threads, 2.60 GHz)
  - **24 TB** local storage (20 x 1.2 TB 2.5” 10K SAS 12 GB/s HDD, RAID 20+2)
  - **768 GB** memory
- **Hub** server (6-core, 192 GB)
- **200 TB** storage (0.2 PB)

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

Architecture

https://crib.utwente.nl

- Based on **open-source** software (Ubuntu, Docker, JupyterHub, …)
- Accessible through a **web browser** (No software installation is 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)
- Low **energy footprint** (10-30W per unit)

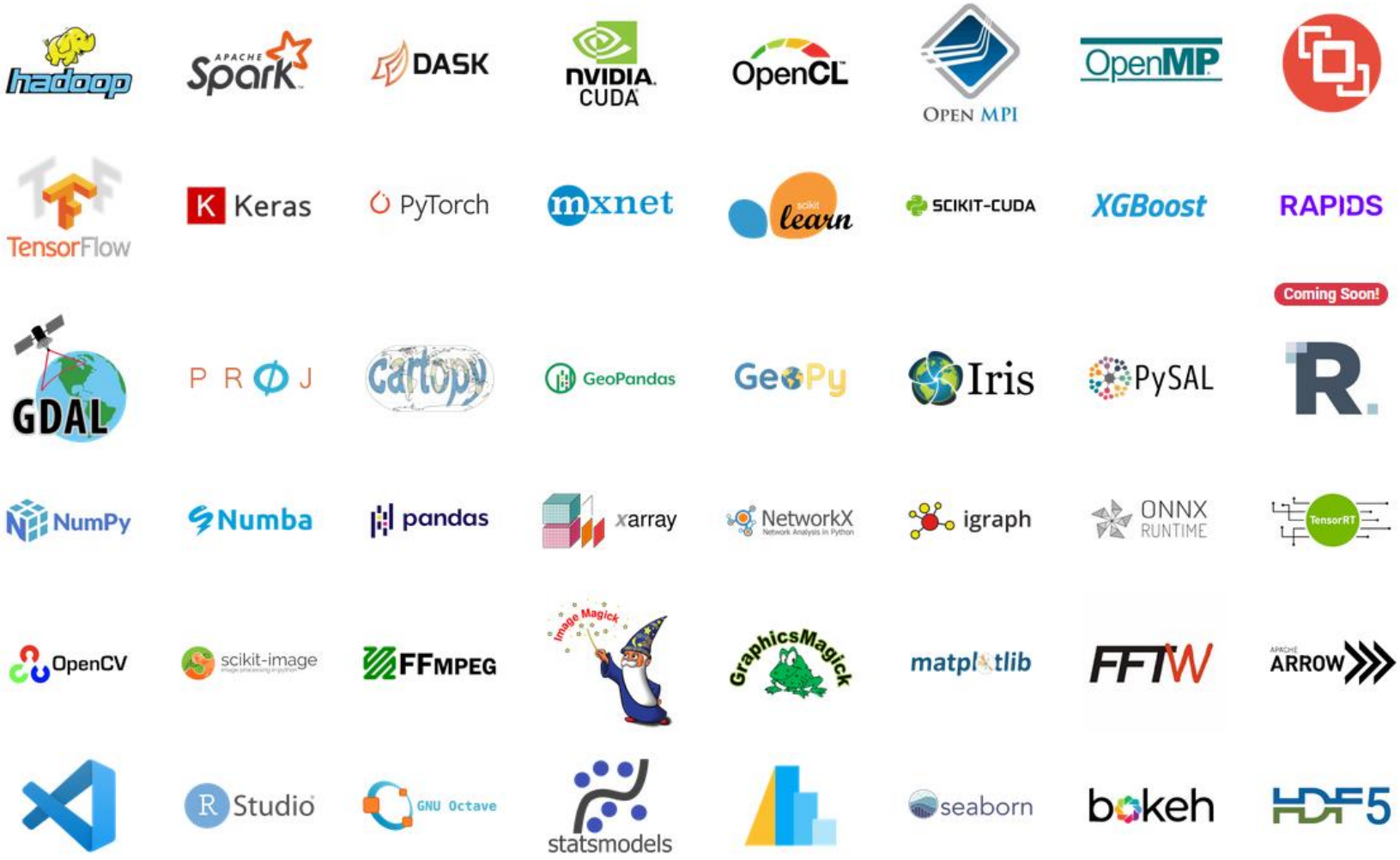
Key Features

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)
- Distributed computing clusters are **ready to use** (Dask, Apache Spark)
- **Public** assets are shared by all users
- **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)

Available Software

https://crib.utwente.nl



and hundreds more…

Additional 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

In cooperation with ITC Research Data Team

Incubating! - BETA

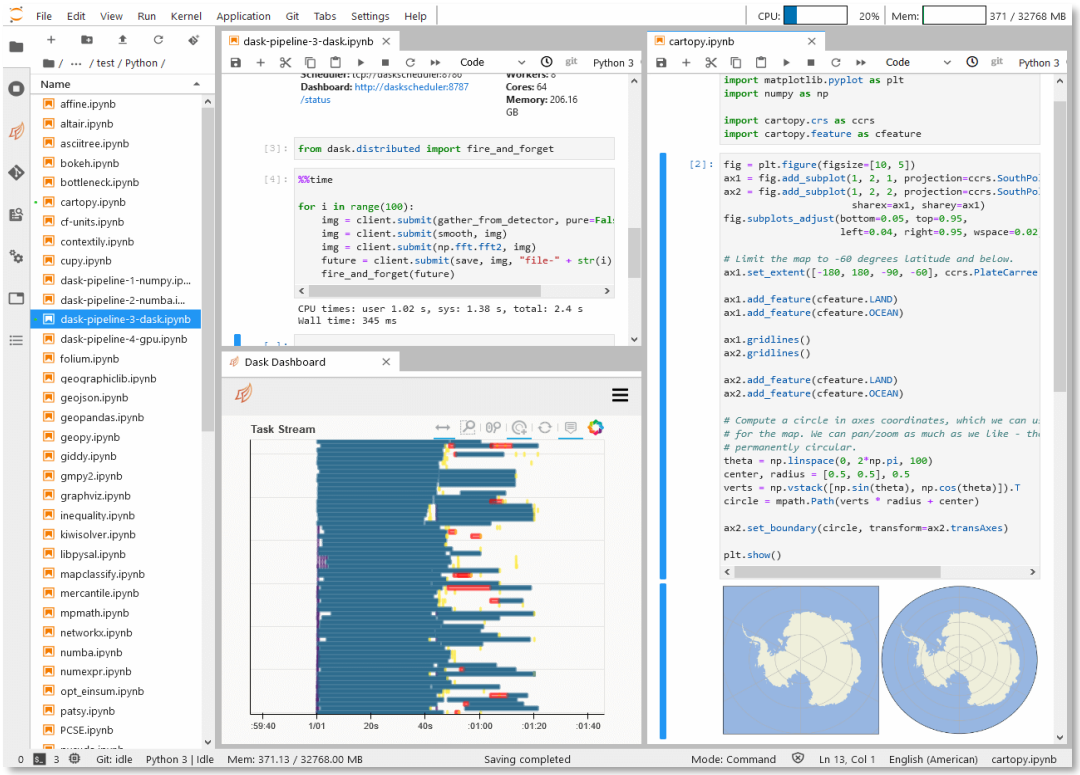
Potential Use Cases

https://crib.utwente.nl

- Education
  - Computation platform for **courses** (Shared course workspaces)
- Research
  - **M.Sc.** / Ph.D. thesis studies
  - Collaborative (big) data analysis and visualization
  - Strengthen **project proposals** (Reduced budget needs for small projects, e.g., 50-100K EUR)
- Capacity Development
  - **Self-learning** (Cloud computing, distributed computing, GPU computing, Machine Learning, …)
  - Computation platform for **training activities** (e.g., workshops)

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

@BigGeodata

Big Geodata Newsletter