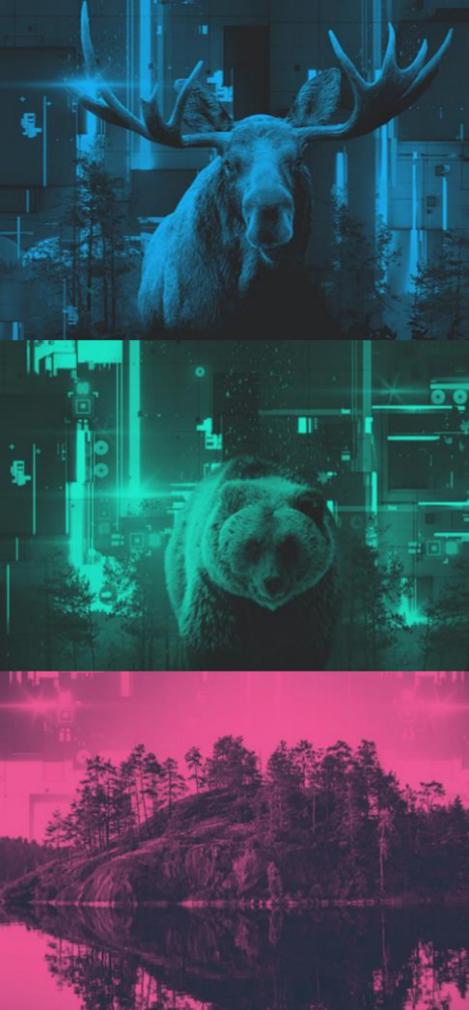# Workshop: Point cloud research in CSC computing environment

Zoom, 07.02.2022

Kylli Ek, Katri Tegel, Samantha Wittke
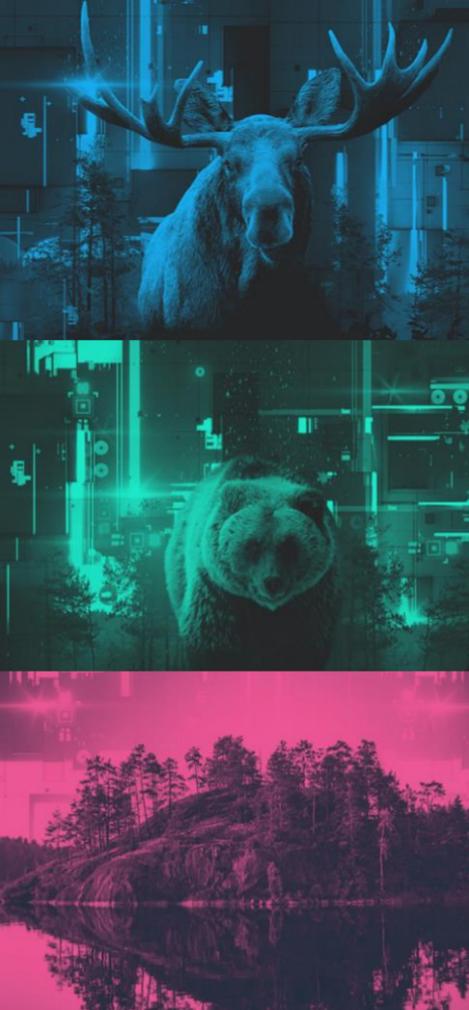
CSC

ICT Solutions for Brilliant Minds

Workshop timetable

12:00-13:30

- Introduction to point cloud data analyis in CSC computing environments
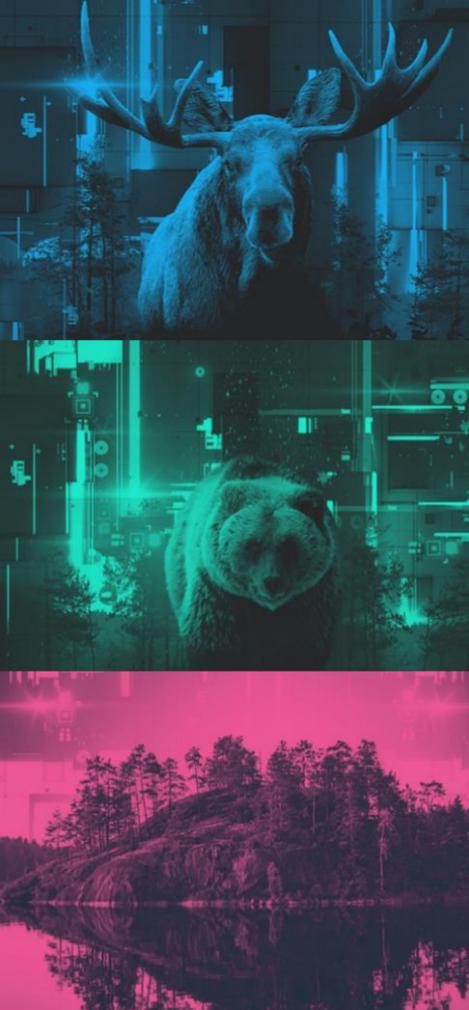
- Research project presentations

13:30-13.45 Break

13:45-15:30 Discussion about tools, data and good practices of point cloud analysis in CSC computing environments

Participants

- 52 registered participants

- Universities: Aalto, Helsinki, Jyväskylä, Oulu, Tampere, UEF

- Research institutes: FMI, LUKE, NLS/FGI, SYKE, VATT

- Companies: Arbonaut, Terrasolid, NTT Data

CSC

- Non-profit state enterprise with special tasks owned by Finnish state (70%) and higher education institutions (30%)

- Headquaters in Keilaniemi, Espoo

- Side offices and supercomputers in Kajaani

- 500+ employees

- GIS: Paituli, GIS courses, universities ArcGIS consortium
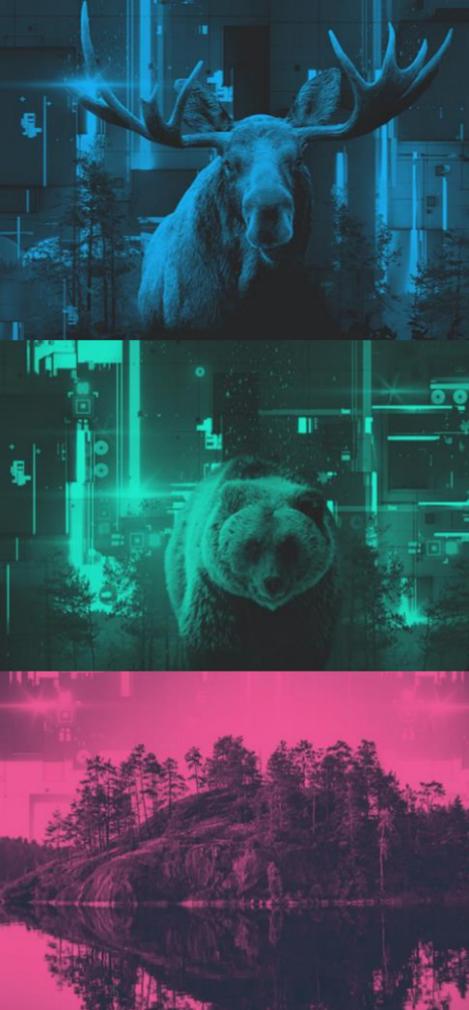
Why using CSC resources?

- Long computing time (> 2 hours)

- High memory usage (> 8 GB)

- Large datasets (> 50 GB)

- Outsource computations, keep own computer free

- Software readily available

- Computing support

-> Puhti

- Server needs -> cPouta

- Course computers (same setup) -> Notebooks

- It's free for open science and teaching for Finnish university and state research institute users

Computing resources for you

	Puhti HPC	cPouta cloud
System	Supercomputer	Virtual machine cloud
Software	Pre-installed software + user-installed software	User-installed software
Data	Main Finnish datasets	-
Use cases	Run demanding analyses with numerous CPUs or GPUs	Setup your own virtual machine and environment
Max per job / VM / container	**4000** CPUs / **80** GPUs **1500** GB memory	**48** CPUs / **4** GPUs **240** GB memory

Average laptop:
**4-8 CPUs** /
8-32 GB memory

PUHTI SUPERCOMUPUTER

Supercomputers in Kajaani

- **Puhti** (national HPC, CSC):
    - 27 280 CPU cores and 240 GPU cores
    - **Wide stack of geospatial software**

- **Mahti** (national HPC, CSC):
    - 89 856 CPU cores and 96 GPU cores
    - No geospatial software

- **LUMI** (EuroHPC, international LUMI consortium)
    - 98 304 CPU cores and **10 240 GPU** cores
    - Software stack unclear
    - 20% for companies
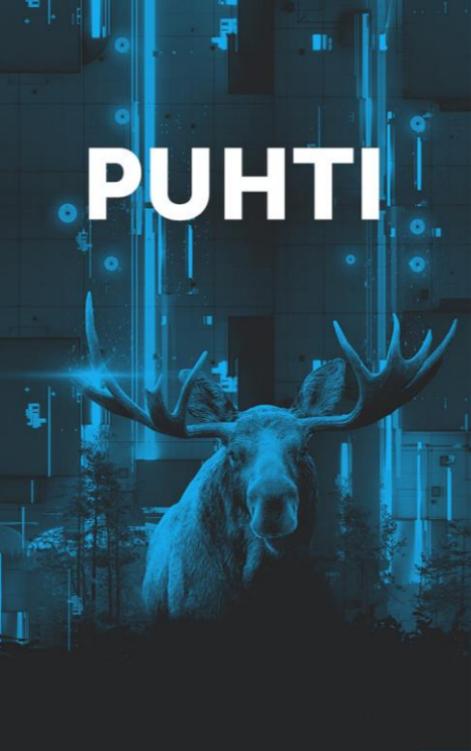   - CPU side in use, GPU coming later this year

Key to geocomputing in HPC: GUI → Scripts
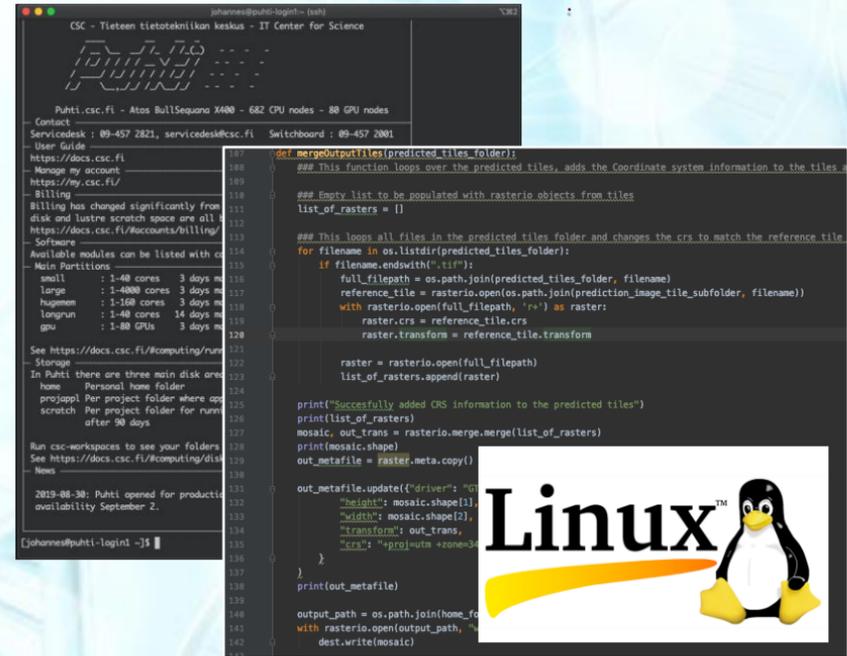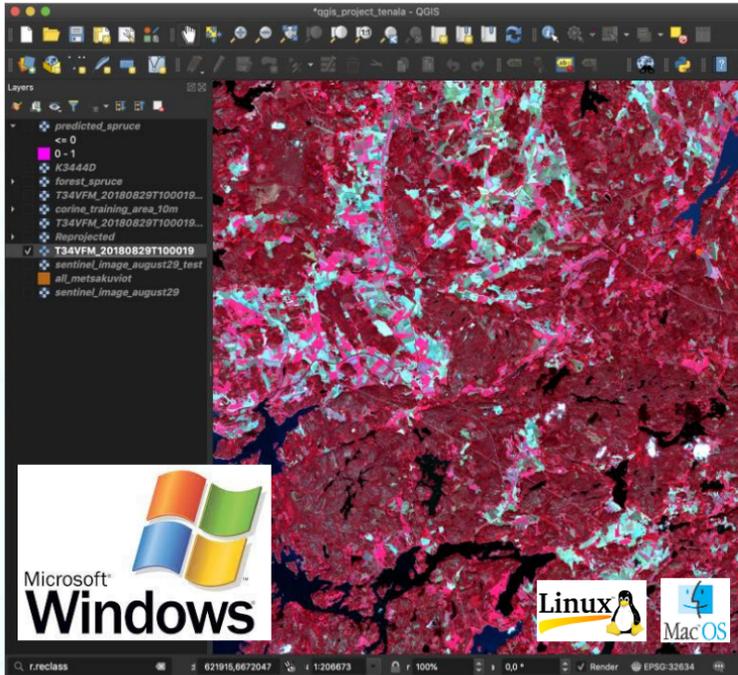
Graphical user interfaces: ArcGIS, QGIS,...
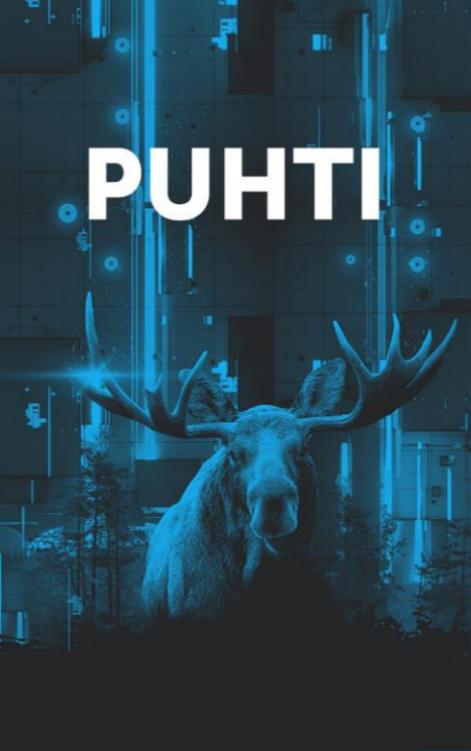
Scripts: Python, R, shell, Matlab, ...

Keep it real

- Single core of Puhti = ~speed of laptop

- But: Puhti has **many cores**

- … and **more memory** and **faster input-output (I/O)**

  - Running single core scripts on Puhti does not make them faster

  - Speedup: multi-core parallel processes

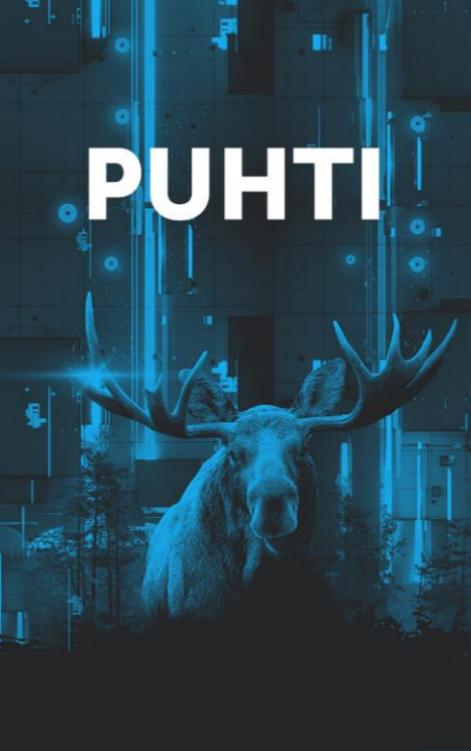 - … or optimize your script

Geospatial Software in Puhti

- ArcGIS Python API
- FORCE & SPLITS
- GDAL/OGR
- GRASS GIS
- **LasTools**
- **MatLab /** Octave
- Mapnik
- OpenDroneMap
- Orfeo Toolbox
- **PCL**
- **PDAL**
- **Python geospatial packages**
- **QGIS,** for viewing point clouds

- **R geospatial packages**
- SagaGIS
- SNAP, Sen2cor
- WhiteboxTools
- Zonation

- **Something missing?**
  - Ask us :)
  servicedesk@csc.fi

Matlab in Puhti

- **Matlab parallel server (MPS),**
  - This requires MATLAB on your own computer with license from your home organization's license server for:
    - parallel computing toolbox
    - all other toolboxes you want to use (for example Lidar)
  - CSC has serverside license
  - Lidar toolbox can be used

- **Interactive Matlab** in Puhti web interface (no Lidar toolbox)
- **Matlab Jupyter** in Puhti web interface (no Lidar toolbox), coming soon
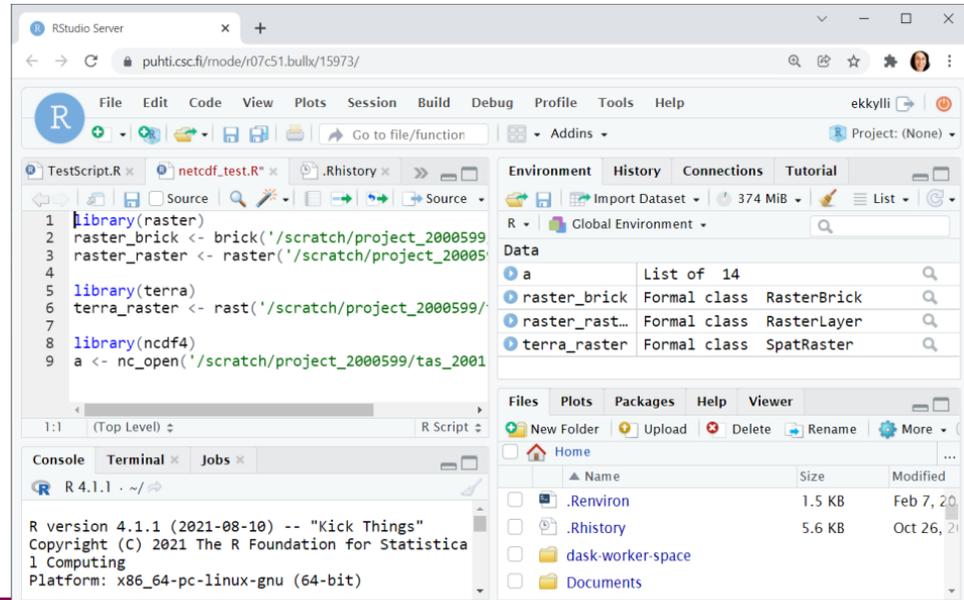
https://docs.csc.fi/apps/matlab
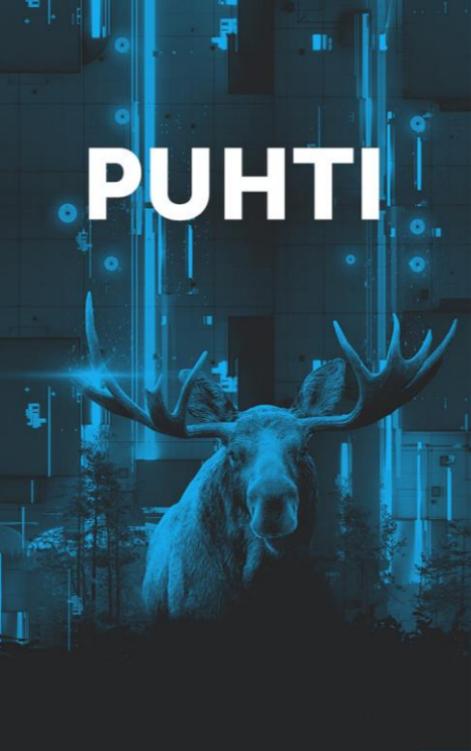
Lastools in Puhti

- By CSC installed only the open source tools

- Instructions for installing (very easy unzip) and using closed source (and licensed) .exe tools

- https://docs.csc.fi/apps/lastools/

Puhti web interface

- Graphical interface to Puhti

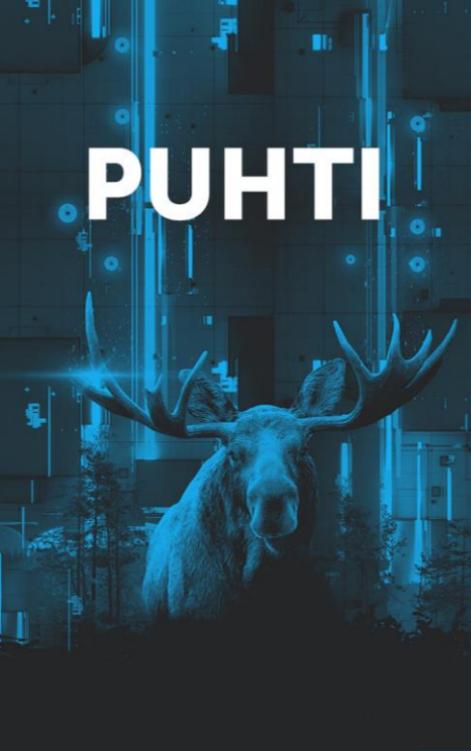- Used with web-browser
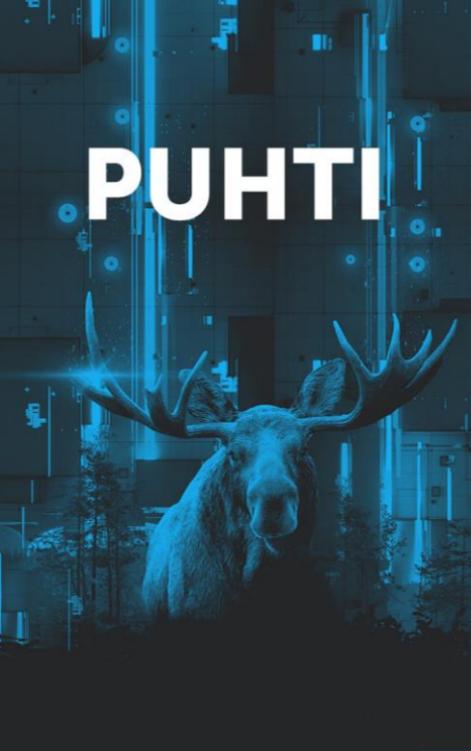
- Jupyter Notebooks, Rstudio, QGIS, Matlab

10/01/2021

Parallel computing with geospatial software

- **Some tools have at least some support built-in**
  - lidR, OpenDroneMap, Whiteboxtools (?)
  - SAGA GIS, GRASS GIS (some parts)

- **Parallel libraries for Python**
  - dask
  - multiprocessing
  - Joblib
  - …

- **Parallel libraries for R**
  - future
  - snow
  - foreach
  - …

Geospatial data in Puhti

geo portti | Finnish Geospatial Research and Education Hub

- Hosts large commonly used datasets with open license
- Removes transfer bottleneck
- Located at: **/appl/data/geo/**
- All Puhti users have **read** access
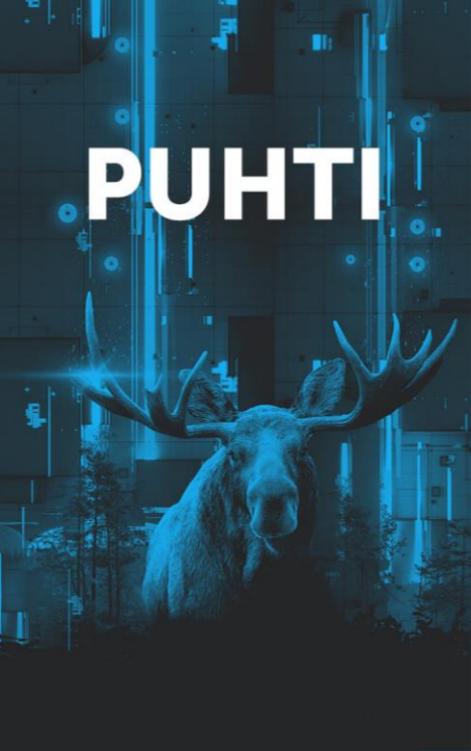
Datasets available (~ **13TB**)
- Paituli data +
    - SYKE open datasets
    - LUKE Multi-source national forest inventory
    - NLS Virtual rasters for DEMs
    - Sentinel and Landsat mosaics

https://docs.csc.fi/data/datasets/spatial-data-in-csc-computing-env/

Want more? - contact us at **servicedesk@csc.fi**

- Finnish Digital and Population Data Services Agency
- Finnish Food Agency
- Finnish Meteorological Institute (FMI)
- Finnish Transport Infrastructure Agency, Digiroad
- Institute for the Languages of Finland (KOTUS)
- Karelia University of Applied Sciences
- Latuviitta
- National Land Survey (MML)
- Natural resource institute Finland (LUKE)
- Statistics Finland
- University of Helsinki, Digital Geography Lab

Example scripts

https://github.com/csc-training/geocomputing

- Batch job files
- Parallel examples for **R** and **Python**
    - Different parallelization libraries
    - Array jobs as well as parallel jobs
- Examples for **Allas** data transfers with **R** and **Python**
- Sentinel image download example (python)
- **SNAP** array job example

Getting started with Puhti

Read Puhti documentation
- Geocomputing: https://research.csc.fi/geocomputing
- GIS software specific pages: https://docs.csc.fi/apps/#geosciences

Come to a course
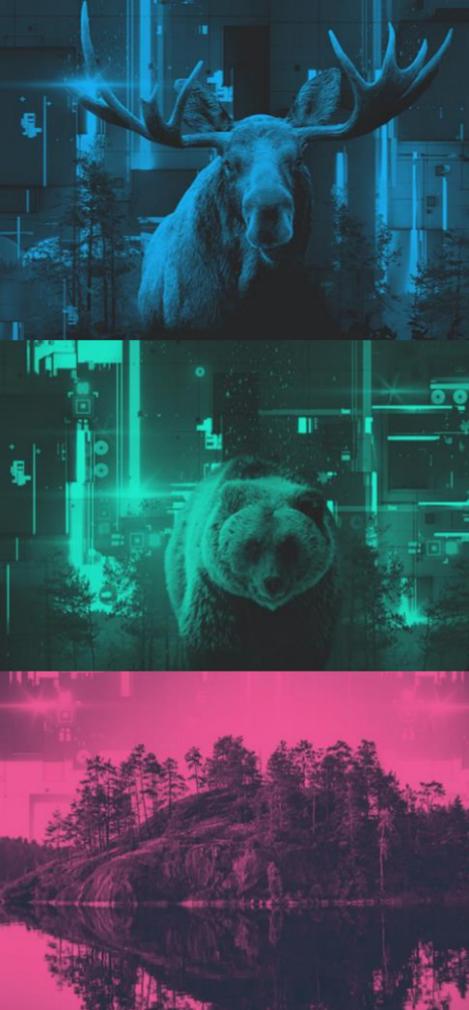- Introduction to Python GIS, 7-10.3.2022
- Using CSC HPC environment efficiently, 23-25.3.2022
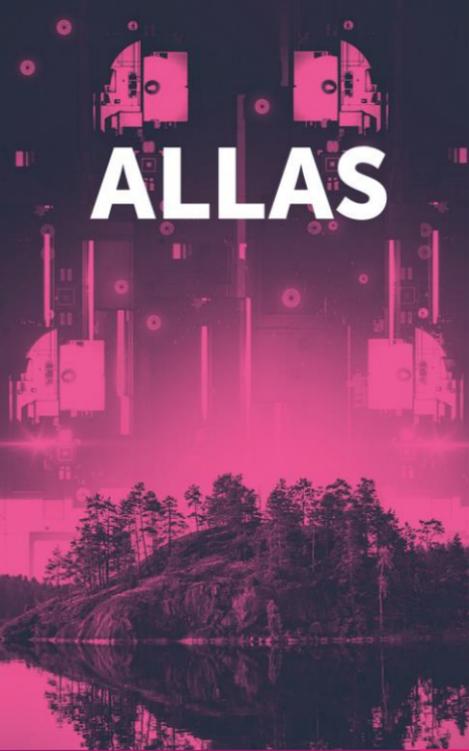
Contact CSC
- servicedesk@csc.fi

Other options than Puhti

cPouta / ePouta

- Virtual Machines (VM) for project use, CPU and GPU

- **Server software:**
    - API: GeoServer, MapServer
    - Web-maps: OpenLayers, Leaflet
    - Databases: PostGIS

- **Speficic use cases:**
    - Node-locked licensed software, Agisoft MetaShape
    - Installations that require root access

- **Setup from scratch:**
    - Operating System (any Linux distribution)
    - Firewall, network, IP etc
    - Software
    - Data

- ePouta for sensitive data → more strict network rules

Allas object storage

- Central data service for all academic projects

- Up to 200 Tb of space for a project

- Access from Puhti, cPouta, personal computer or any other place

- GDAL has good support for reading directly from object storage

- Option to make data public

- See also: CSC webinar - **Allas and Geospatial data https://www.youtube.com/watch?v=mnFXe2-dJ_g**

Lidar data analysis platform for governmental organizations

- Ordered by National Survey of Finland, developed at CSC

- **Mainly for NLS new 5p/m2 lidar data analysis**

- High security, requires strong authentication

- Development has just started

- For end-users something hopefully available at end of 2022

- No access for university users
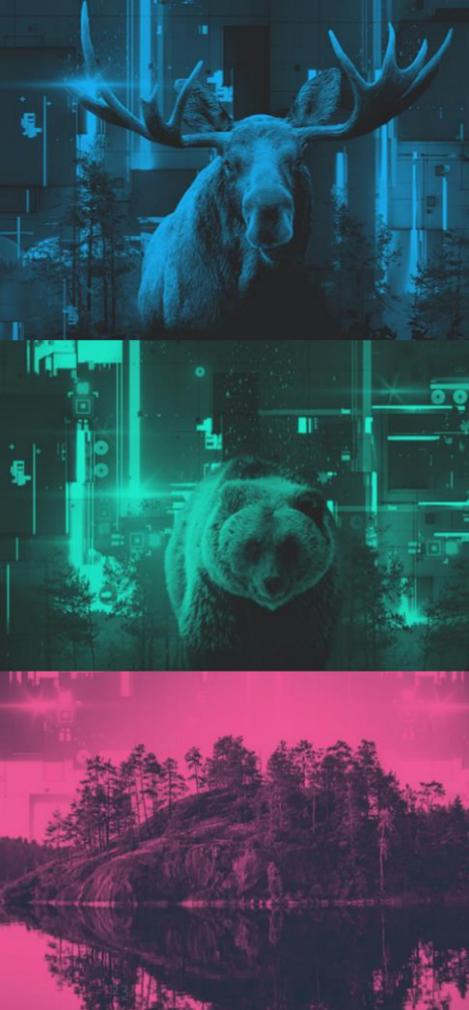
Summary

- **Puhti is an excellent tool if you**
    - Need more computing power
    - Don't want to run long analyses on your personal computer
    - Have **a lot** of data
    - Are using eg open Lidar data provided by NLS
    - Are willing to use scripts for your work
    - Have some basic Linux skills
    - Are willing to learn to use a HPC
- **cPouta is an excellent tool if you**
    - Want to use databases and web-/mapservices
    - Have Linux server administration skills