

Using Lidar data in Taito Eduardo Gonzalez, CSC

CSC, 16.4.2018



CSC – Suomalainen tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon ICT-osaamiskeskus

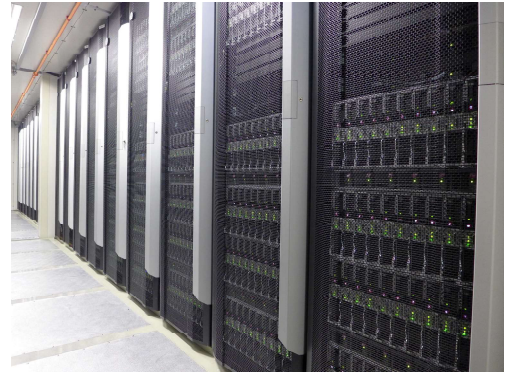
Introduction



- Why would you use Taito?
- Short introduction Taito
- Ready installed GIS (and Lidar) software
- Ready available Lidar (and other) datasets
- Short demo – Using LiDAR dataset with R in Taito
- References and contact information

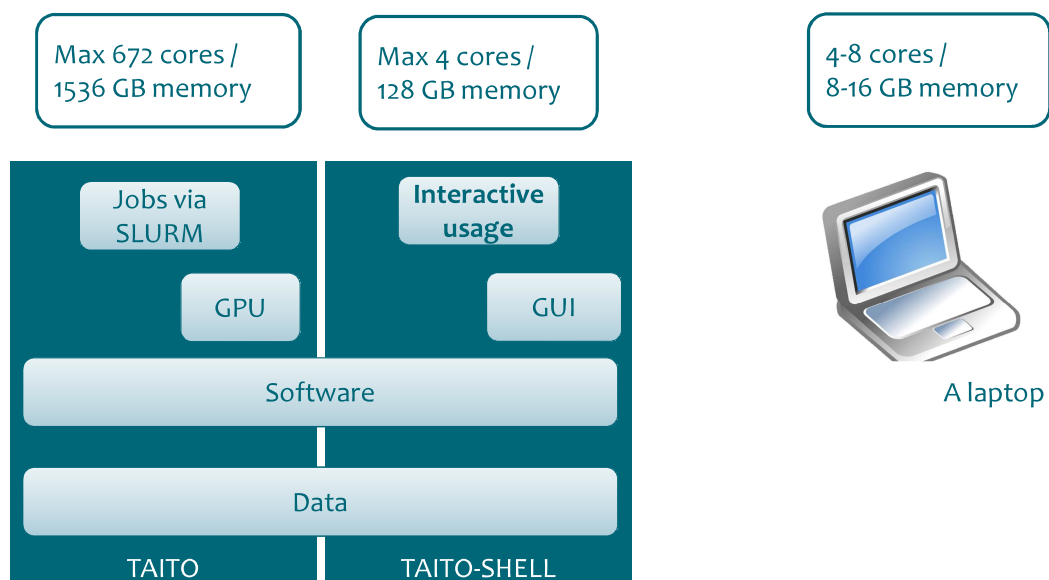
Reasons for using CSC computing resources

- Computing something takes more than 2-4 hours
- Need for more memory
- Very big datasets
- Keep your desktop computer for normal usage, do computation elsewhere
- Free for Finnish university users and state research institutes for public research
- Today we focus on **Taito**, check out other CSC's services at <https://research.csc.fi/computing-and-software>



3

What is Taito and Taito-shell?



4

Realistic expectations

- A single core of a CSC machine is about as fast as one of a basic laptop.
 - It has just a lot of them...
 - .. and more memory and faster input-output.
-
- Just running your single core script at CSC does not make it much faster.
 - For clear speed-ups you have to use several cores.
 - ... or optimize your script.

5

Say hello to Taito

```
gonzalez@taito-login3:~$ cat /etc/issue
2014-10-02: If you wish to change the project you're using for billing,
please use "newgrp groupname". You can find more information at:
http://tinyurl.com/kozfaet

2014-11-27: For jobs requiring more than 16 GB memory per core, please
use the 'hugemem' queue consisting of six 1.5 TB memory nodes with
32/40 cores each.

2017-01-13: Monthly $WRKDIR clean-up will start on Feb. 1, 2017.
Data storage info: http://bit.ly/2jDk7SP and http://bit.ly/2jLjydH

2017-01-19: Remember that memory is also billed. That is, the basis of
billing is either the number of cores or number (rounded up to the
nearest integer) of 4 GB blocks of memory, depending which is larger.
For example,
- an 8 core job with 16 GB of memory will be billed with 2 x 8 = 16 bu/h
- an 8 core job with 64 GB of memory will be billed with 2 x 16 = 32 bu/h

gonzalez@taito-login3 ~]$ echo Hello I am Taito, your friend!
Hello I am Taito, your friend!
gonzalez@taito-login3 ~]$
```

6

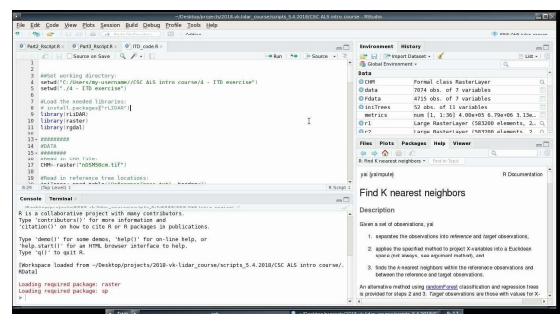
Some cons

- Using Taito is not trivial for no Linux users ... need to learn some Linux basics only
- Learning curve to get started with remote services ... not too steep though
- Need to learn a bit about batch scripts in Linux ... quite simple code
- And a bit about the SLURM platform ... just some little extra code and parameters to specify how big a computer you want to use

7

Some pros I: Software stack... AND also graphical interfaces!

- Several common GIS software already installed
 - GDAL, Proj4, TauDEM, Zonation, R packages, Python libraries, MatLab...
 - You can ask for software to be installed
 - You can also install yourself
- Also GUI tools!
 - Rstudio
 - Spyder for Python
 - QGIS, GRASS, SAGA
- Available GIS datasets (e.g. Lidar data)
- Plus the number of cores and memory resources

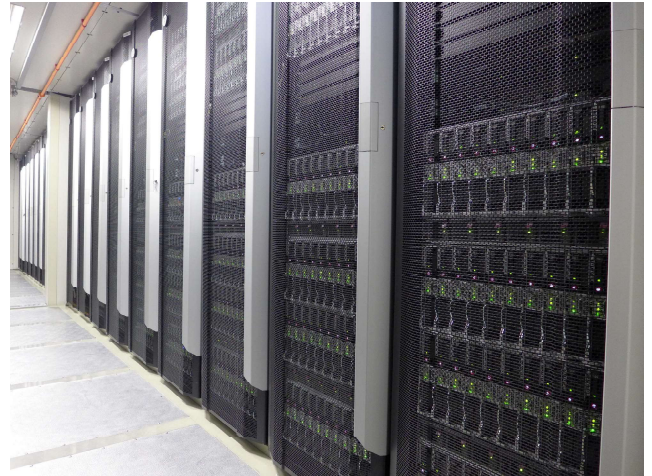


8

Taito / Taito-shell pre-installed software for GIS

- [GDAL/OGR](#)
- [GRASS GIS](#)
- [LasTools](#)
- [PDAL](#)
- [Proj.4](#)
- [Python with spatial packages](#)
- [QGIS](#)
- [R with spatial packages](#)
- [SagaGIS](#)
- [TauDEM](#)
- [Zonation](#)
- [MatLab](#)
- No: ArcGIS, PostGIS, GeoServer

You can install software also yourself



<https://research.csc.fi/software> -> Geosciences

9

Shared data area in Taito

- Hosts large commonly used datasets
- Reduces the need to transfer data to Taito
- Located at /proj/ogiiir-csc/
- All Taito users have read access.
- For data with open license
- Virtual rasters for DEMs
- If you think some other dataset should be included here, ask from servicedesk@csc.fi

https://research.csc.fi/gis_data_in_taito

MML:

- Lidar point cloud data
- Dem 2m (see virtual rasters section below)
- Dem 10m (see virtual rasters section below)

FMI

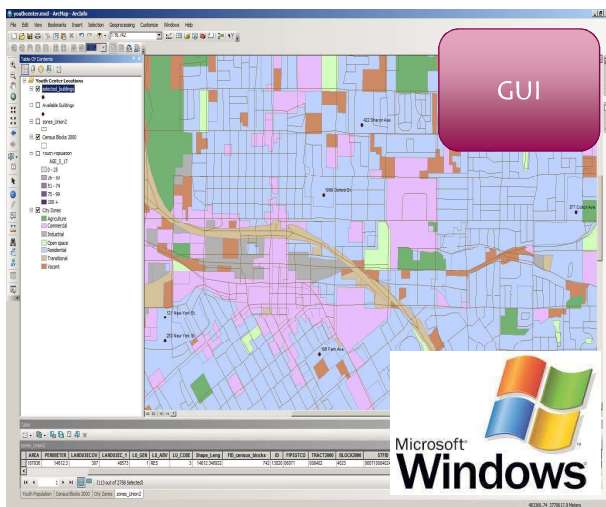
- 10km avg relative humidity
- 10km avg sea level pressure
- 10km daily max temperature
- 10km daily mean temperature
- 10km daily min temperature
- 10km daily precipitation
- 10km daily radiation
- 10km daily snow
- 10km monthly mean temperature
- 10km monthly precipitation

LUKE

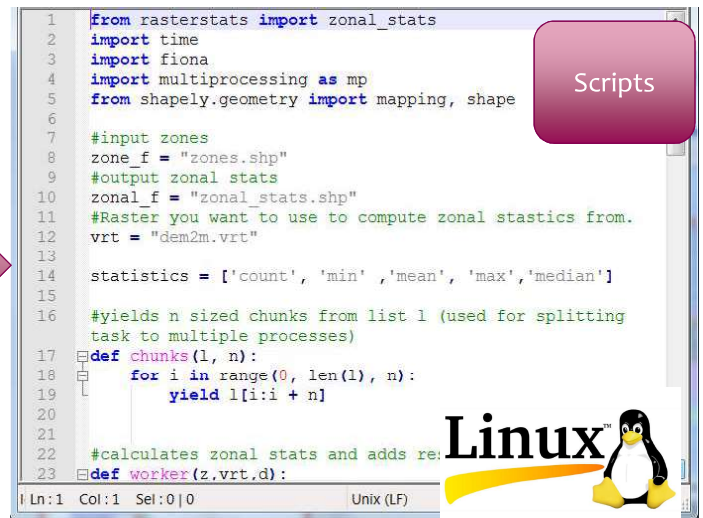
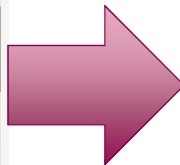
- Multi-source national forest inventory

10

The keys to geocomputing: Change in working style & Linux



GUI



Scripts

ArcGIS, QGIS, ...

R, Python, shell scripts, Matlab, ...

Example code in CSC training Github

- Examples for doing spatial analysis in CSC computing environment with:
 - Python
 - R
- Examples include also batch job scripts suitable for Taito.
- Some of the examples include samples for serial, array and parallel jobs.



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

Demo: Using Taito's Lidar datasets & R's "rlas" plackage

Topics:

- An example on how to access Taito's lidar data with R (get_lidar_files.R)
- How to use "rlas" package for basic data management (basic_rlas.R)
- Example on how to run an R script with simple batch job in Taito (batchjob_rlas_basics.sh)

The scripts are available at

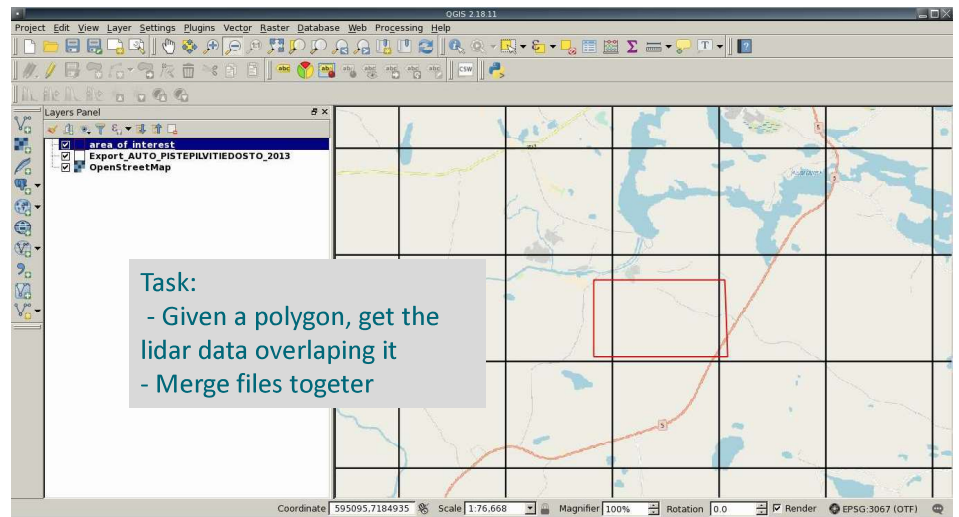
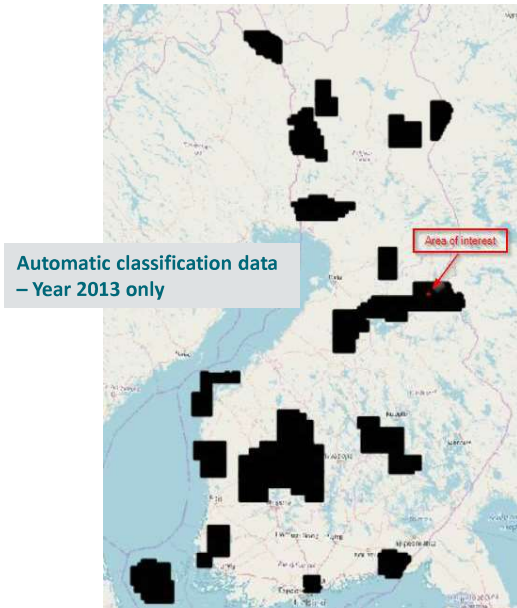
<https://github.com/csc-training/geocomputing> under the R/lidar section

Demo's overview Getting Lidar data in Finland

The screenshot shows the MML (Maanmittauslaitos) web application interface. At the top, there's a search bar and navigation options for 'Suomenkieli', 'Pääsivenskieli', and 'In English'. Below the search bar, there's a list of datasets with columns for 'Name', 'Size', and 'Last Modified'. The map view shows a geographical area with various datasets overlaid, including 'Kantalehtijako ruudakko (1)', 'Kintelehtirekisterikaartta (3)', 'Kintopistereikasteri (5)', 'Korkeusmalli (2)', 'Korkeusvyhykkeet (4)', and 'Kuntajako (3)'. The map also shows 'Lasertalustasunnitelma (1)'. The bottom part of the interface shows a 'Paikittu' section with a map of Finland and a list of datasets.

Demo's overview Lidar data in Taito

- The data (2008-2016) is ready to use in folder:
`/wrk/project_ogjir-csc/mml/laser/`



15

8.6.2017

Reference information: Getting started

- Apply for an [account and a project](#)
- Read [Taito user-manual](#), at least:
 - o [Connect to Taito](#), NoMachine
 - o Different [directories for working](#)
 - o [Module system](#) for loading software
 - o [Batch system](#) for submitting jobs
 - o [Software specific page\(s\)](#)
- Learn basic [Linux skills](#)
- Check out materials from our two days course in 2017 – [Geocomputing using CSC resources](#)

16

Reference information: Steps for running your script in Taito-shell

(0. Get yourself CSC user account)

1. Move your data and scripts to Taito (with FileZilla).
 2. Connect to Taito (with Putty or NoMachine).
 3. Open Spyder / Rstudio from GIS menu.
 4. Check which Python/ R packages do you need and if they are available in Taito.
* If needed, install it yourself or ask CSC - servicedesk@csc.fi.
 5. Fix the paths of your input/output files.
 6. Test your script in Taito-shell with some test data.
 7. Run your scripts with all data in Taito as batch job.
- (8. Rewrite your code to use several cores.)

More info and support

Geocomputing general info: <http://research.csc.fi/geocomputing>

Taito user guide: <https://research.csc.fi/taito-user-guide>

Geocomputing course materials: https://www.csc.fi/web/training/-/geocomputing_2017

Eduardo González, 0408488989

servicedesk@csc.fi