

Datascience

01 - Introduction



Dávid Visontai

ELTE, Physics of Complex Systems Department

2022.02.08.

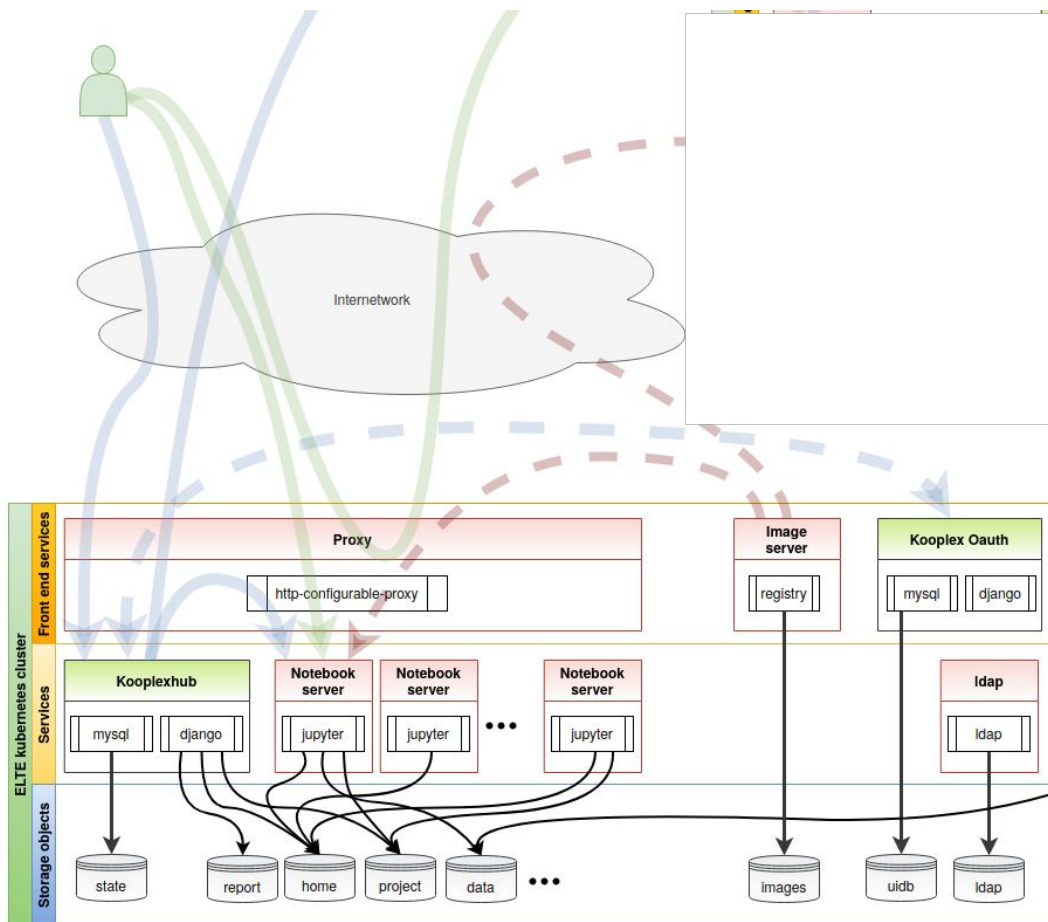
Schedule

- Technical information for this course
- Datascience and physics
- Basics of programming in python
- Using numpy, scipy and simpy for linear algebra
- Basics of plotting and creating figures
- Working with timeseries

The platform can be accessed at

<https://k8plex-edu.elte.hu>

Kooplex containers



- **Docker** containers
Virtualization
- **Docker** images
e.g. *hub.docker.com*
- **Kubernetes**
Orchestration

Virtualization engines:

- Docker
- Podman
- Nvidia enroot
- Apptainer
- ...

→ KISS principle

When do we generate data?

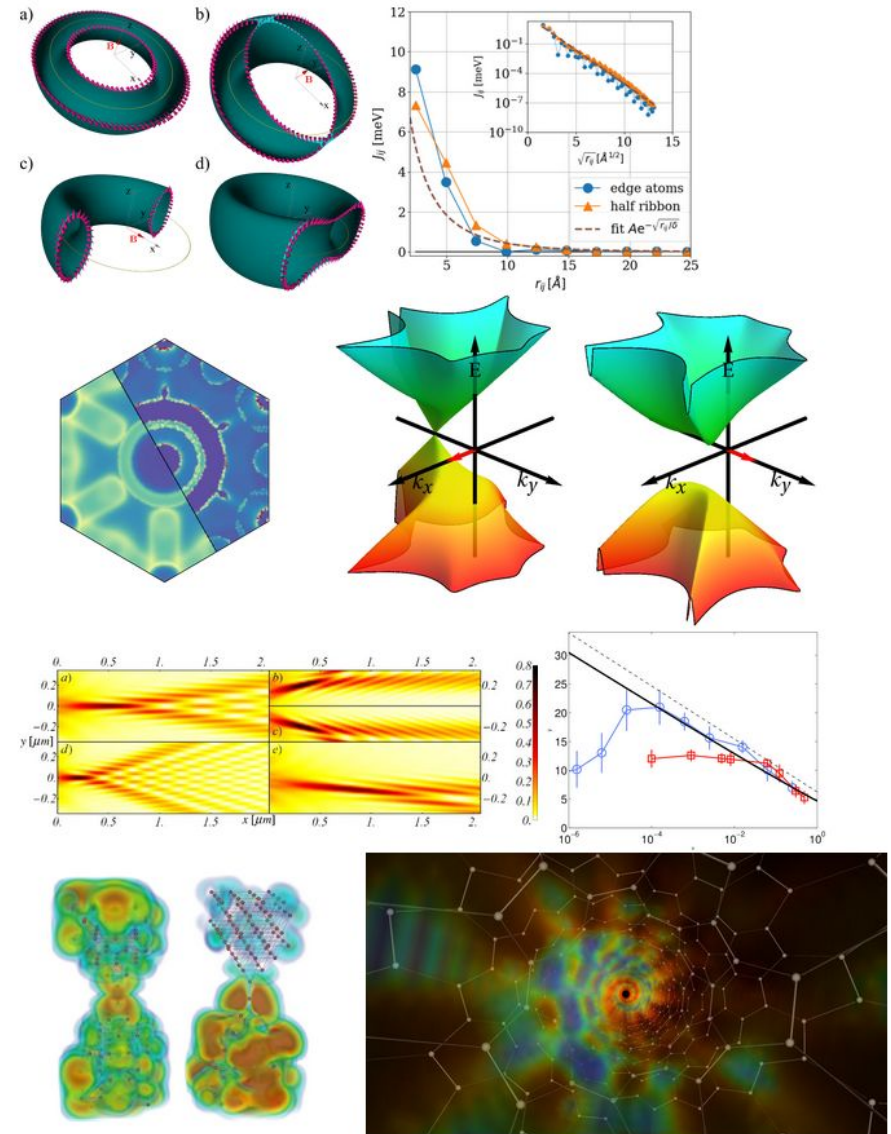
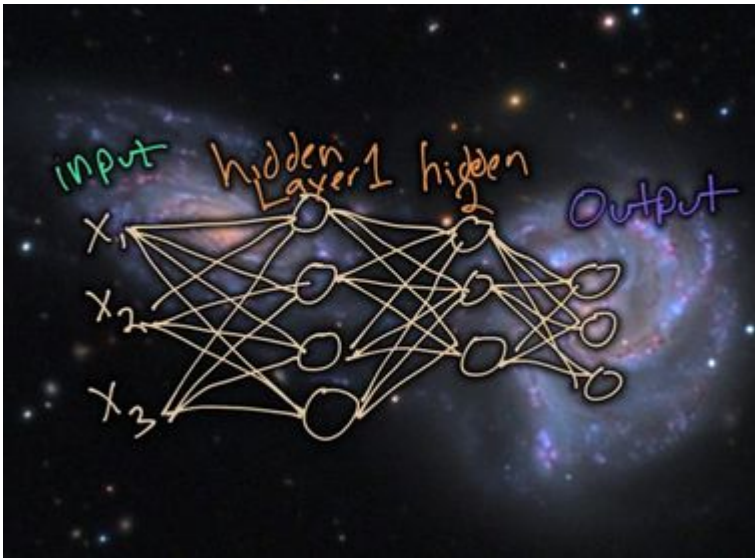
- In experiments (raw data, analysed data)
- In simulations (output of softwares)

What do we do with data?

- Store it (database, file format)
- Analyse it (clean, filter, transform, reduce)
- Share it to collaborators (structure, logic)
- Support published results with data (provenance)

Datascience and physics

reproducible-science
genetics
Python
machine-learning
RNA-seq
NGS
big-data
cancer
aging



Public datasharing sites

- <https://osf.io/>
- <https://datadryad.org/stash>
- <https://figshare.com/>
- <https://www.re3data.org/>

Basics of plotting and creating figures

- Matplotlib - creates images, not scalable, but still very popular

Javascript based libraries:

- Bokeh - <https://docs.bokeh.org/en/latest/>
- Plotly - <https://plot.ly/>

Higher level plotting


- Holoviews - <http://holoviews.org/>

Working with timeseries

First task: /workdir/01-timeseries-waterlevel

[1] <https://waterdata.usgs.gov/nwis>

Search for Sites With Data

Current Conditions	Sites with real-time or recent surface-water, groundwater, or water-quality data.
Site Information	Descriptive site information for all sites with links to all available water data for individual sites.
	Map of all sites with links to all available water data for individual sites.

Frequent Searches By Data Category

Surface Water	Water flow and levels in streams and lakes.
Groundwater	Water levels in wells.
Water Quality	Chemical and physical data for streams, lakes, springs, wells and other sites.
Water Use	Water use information.