Sunet Drive

An Academic Toolbox with Jupyter Notebooks

Micke Nordin

Richard Freitag (presenter)

Magnus Andersson

Co-funded by the European Union

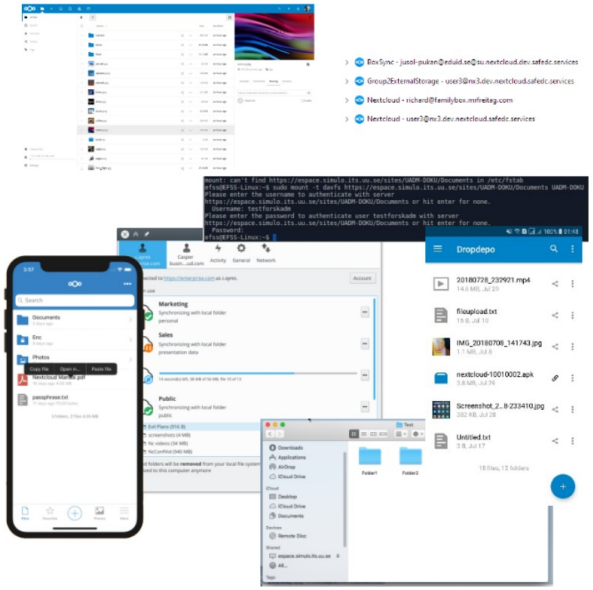
GÉANT

Sunet Drive – An Academic Toolbox

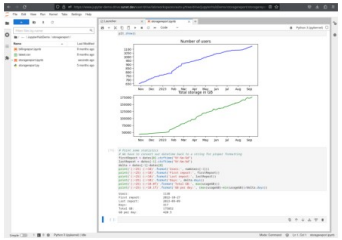
Primary data sources

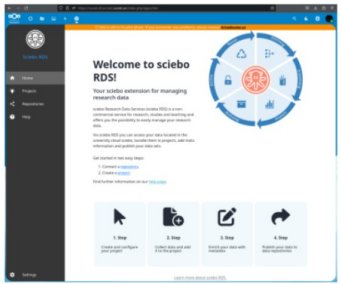
- UV-Vis
- IVIUM
- ...
- PIA
- IPCE

Secure Storage



Analyze and Publish





General – Sunet Drive and JupyterHub

Sunet Drive

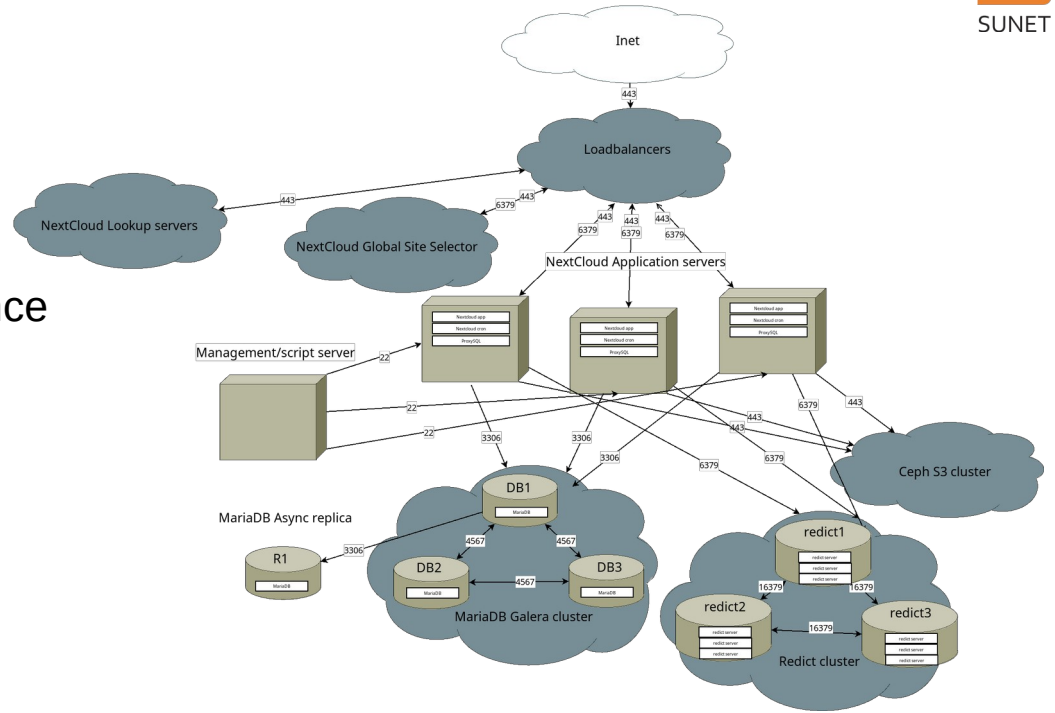
- Based on Nextcloud
- Storage in Ceph S3
- Currently stores ~450 TB of mostly research data
- 20 Universities – growing rapidly
- Upcoming features is:
  - Process research data – crunch the numbers
  - Publish research
  - MFA protected zones

JupyterHub

- a Hub (tornado process) that is the heart of JupyterHub
- a configurable http proxy (node-http-proxy) that receives the requests from the client's browser
- multiple single-user Jupyter notebook servers (Python/IPython/tornado) that are monitored by Spawners
- an authentication class that manages how users can access the system

Nextcloud Architecture

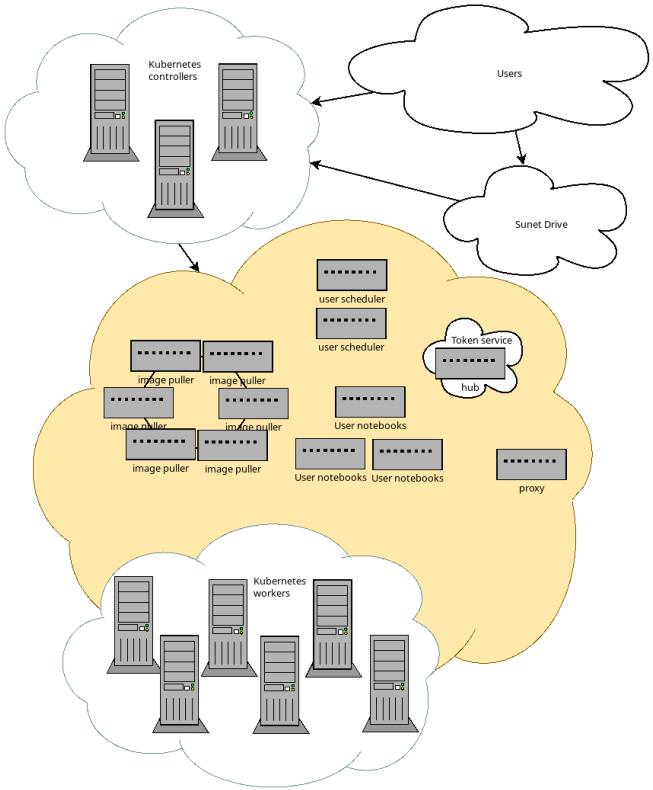
- Load-balanced Sunet Drive Frontend Servers*
- MariaDB Galera cluster as database
- Redict Cluster for improved performance
- All storage in Ceph S3 storage cluster



* https://github.com/SUNET/sarimner-frontend

Jupyter Architecture

- Load-balanced controllers*
- Scalable by adding more workers
- All storage synced to openstack cinder



* https://github.com/SUNET/sarimner-frontend

Jupyter Integration – How?

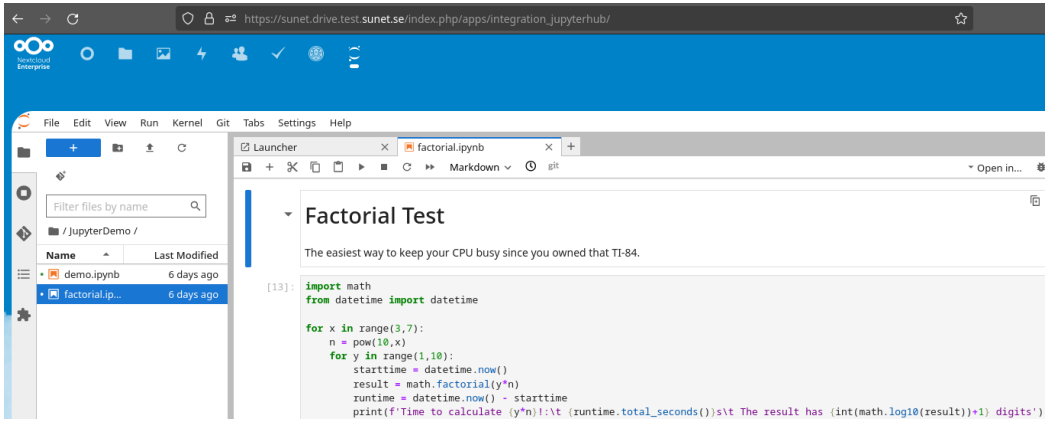
- JupyterHub in Kubernetes*
- Nextcloud app for easy access**
- Integration via OpenID Connect**
- Implemented sync via Oauth tokens**

* https://z2jh.jupyter.org/en/stable/

** https://github.com/sunet/nextcloud-jupyter

JupyterHub Integration in Nextcloud?

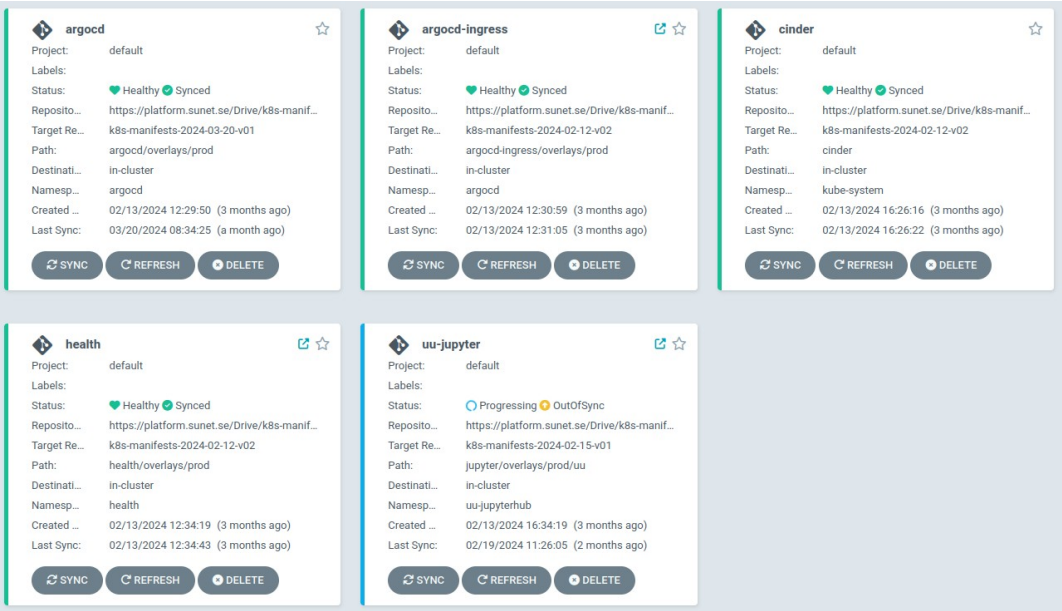
- Integrated in Sunet Drive
- Users can individually execute notebooks
- Users get 10GB “local” Jupyter Storage
- Storage is continuously synchronized between Sunet Drive and JupyterHub
- Other use-cases are being evaluated



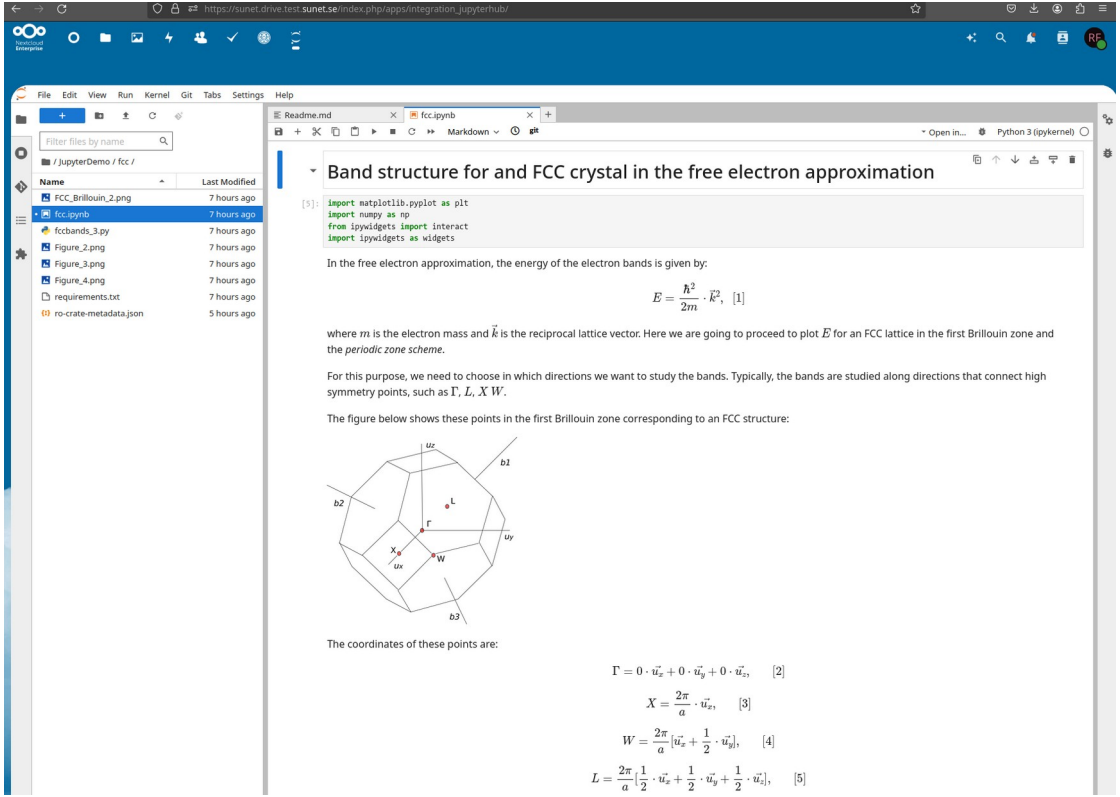
Jupyter Deployment and Scalability

60 concurrent users

- 6GB of RAM --> 100MB per user
- 20-25% CPU load --> 1-2% CPU per user per vCPU
- CPU overprovisioning show no significant scheduling penalties
- 16vCPU/32GB could host 100-150 idling jupyter-servers, with 30-40 concurrent users per kubernetes-worker
- Memory seems to be the biggest concern



Demo


https://sunet.drive.test.sunet.se/index.php/apps/integration_jupyterhub/
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ JupyterDemo / fcc /
Name
Last Modified
FCC_Brillouin_2.png 7 hours ago
fcc.ipynb 7 hours ago
fccbands_3.py 7 hours ago
Figure_2.png 7 hours ago
Figure_3.png 7 hours ago
Figure_4.png 7 hours ago
requirements.txt 7 hours ago
ro-crate-metadata.json 5 hours ago
Readme.md
fcc.ipynb
Markdown
Open in...
Python 3 (ipykernel)
Band structure for and FCC crystal in the free electron approximation
[5]:
import matplotlib.pyplot as plt
import numpy as np
from ipywidgets import interact
import ipywidgets as widgets
In the free electron approximation, the energy of the electron bands is given by:
$$E = \frac{\hbar^2}{2m} \cdot \vec{k}^2, \quad [1]$$
where m is the electron mass and $\vec{k}$ is the reciprocal lattice vector. Here we are going to proceed to plot E for an FCC lattice in the first Brillouin zone and the periodic zone scheme.
For this purpose, we need to choose in which directions we want to study the bands. Typically, the bands are studied along directions that connect high symmetry points, such as Γ, L, X W.
The figure below shows these points in the first Brillouin zone corresponding to an FCC structure:
uz
b1
b2
L
Γ
X
W
ux
uy
b3
The coordinates of these points are:
$$\Gamma = 0 \cdot \vec{u_x} + 0 \cdot \vec{u_y} + 0 \cdot \vec{u_z}, \quad [2]$$
$$X = \frac{2\pi}{a} \cdot \vec{u_x}, \quad [3]$$
$$W = \frac{2\pi}{a}[\vec{u_x} + \frac{1}{2} \cdot \vec{u_y}], \quad [4]$$
$$L = \frac{2\pi}{a}[\frac{1}{2} \cdot \vec{u_x} + \frac{1}{2} \cdot \vec{u_y} + \frac{1}{2} \cdot \vec{u_z}], \quad [5]$$

Thank you

Any questions?

kano@sunet.se

freitag@sunet.se (presenter)

mandersson@sunet.se

Co-funded by the European Union

GÉANT

tnc24

RENDEZVOUS À RENNES

Rennes, France | 10-14 JUNE 2024