The background is a deep space scene. At the bottom, the curved horizon of Earth is visible, showing a thin blue atmosphere and a bright yellow-orange glow from the sun or a star on the left. A prominent, bright orange arc curves across the upper half of the frame. The rest of the sky is dark blue and black, filled with numerous stars and the faint, hazy structure of the Milky Way galaxy.

National Space Data Lab

Tor Björn Minde
RISE

SPACE INNOVATION FORUM

4 - 5 September 2019

Background and purpose

- Need for a common resource for development and analysis on data from space
- A platform for development of services based on open code will enable innovation
- FN, ESA and EU say it is a prerequisite to fulfill the goal of Agenda 2030.

Nationellt Rymddatalabb will be a **national knowledge and data hub** and work on earth observation data and for the development of AI-based analysis of data, generated in space systems.

The purpose is to

- Increase the use of data from space for the development of society and industry
- Start up a **data lab for space data** that will become a tool for different users

Project partners

Project partners



AI INNOVATION of Sweden



Reference group



AI INNOVATION of Sweden

Swedish National Space Agency

- Support the development of the Swedish space sector
- Promote the development of space related technology and knowledge with special focus on transportation, communication, environment and climate
- Work together with other actors to increase the interest in natural sciences and technology, especially among young people

AI Innovation of Sweden

Public funding



Corporate



Medium sized companies & Consulting



SMEs & Startups



Academia & Research Institutes



Lifelong Learning



Public organizations



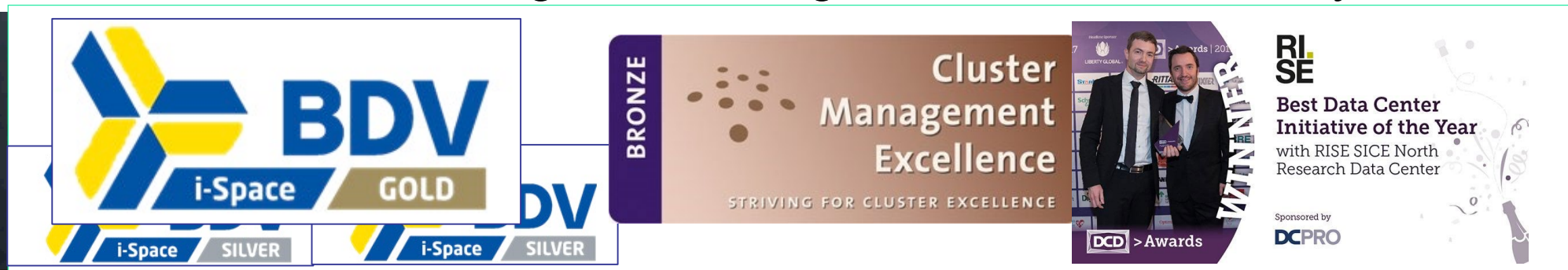
LTU – AI and Space Technology

- Top university for space engineering (+master program)
 - On the list of Forbes *top 30 under 30*
 - Campus in Luleå and Kiruna
- Top university in *Applied AI*
 - Hundreds of highly cited AI-related journal publications
 - 3 LTU AI-researchers in IVA's 100 list of innovators
 - Leaders in various machine learning competitions
 - Sweden's only NVIDIA DLI Ambassador
- Synergies with the newly established AI innovation hub www.ltu.ai



ICE by RISE

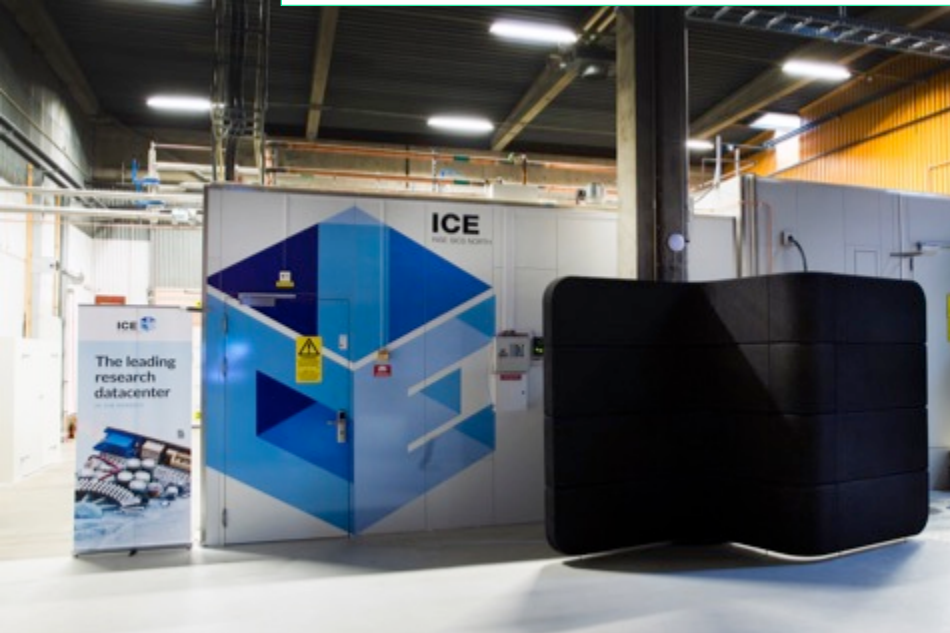
A full-scale research datacenter and test environment with the purpose to increase knowledge and strengthen the AI & DC ecosystems.



BDV
i-Space **GOLD**
i-Space **SILVER**

BRONZE
Cluster Management Excellence
STRIVING FOR CLUSTER EXCELLENCE
DCD Awards

RISE
Best Data Center Initiative of the Year
with RISE SICE North Research Data Center
Sponsored by DC PRO



- 2000 physical servers
- 250 kW
- 200 TB RAM
- Upto 10 petabyte storage
- 50 000 cores
- 240 GPUs
- 1,1 M cuda cores
- 12,5 petaflops
- HDFS clusters
- OpenStack ECC
- OCP servers

Together with LTU



Data driven innovation

We will perform	Nationellt Rymddatalabb Impacts
<ul style="list-style-type: none">• Pilot projects, including end user involvement, to develop data driven innovation based on EO, GNSS and onboard operational data• Collaboration with Swedish Governmental bodies to develop services that they need for decision making and environmental monitoring.• Cooperation with Tillväxtverket and publication of challenges on their platform for open and data driven innovation	<ul style="list-style-type: none">• Common arena for end users and developers• Increased competitiveness for Swedish actors• Focus on Space data and the opportunities arising• Resource for development of Space data algorithms

Organization of Activities

We will perform	Nationellt Rymddatalabb Impacts
<ul style="list-style-type: none">• 4 hackathons/workshops with researchers, data scientists, and developers with the aim to develop data driven innovation. Example: Kartdagarna• 2 open developer events for data driven innovation together with other data labs, researchers, clusters (i.e. Skogstyrelsen)• 2 seminars with new actors (the use of space data)• 2 workshops with partners to develop services• 2 developer events for education on tools and data monitoring	<ul style="list-style-type: none">• become a knowledge hub• create methodology for the use of space data that leads to digital development and innovation• gain Swedish industrial competence and ability• gender equality• increase attractiveness in collaboration with global partners

Open Data Cube

Technical Solutions and Tools



Rymdstyrelsen
Swedish National Space Agency



AI INNOVATION of Sweden 

Open Data Cube

Open Data Cube (ODC) is an open source Geospatial Data Management and Analysis Software project <https://www.opendatacube.org/>. It is mostly written in the Python programming language.

Used to organize the data into an efficient database structure. Data can be queried by

- Satellite mission (Sentinel-1 / 2 / 3)
- Processing level (1C, 2A)
- Geographic coordinate (longitude, latitude)
- Time interval
- Spectral bands

Sentinel-2: Data Products

Sentinel-2 provides two main data products

- **Level-1C:** Top-of-atmosphere reflectances in cartographic geometry. Products are image tiles of 100 km²
- **Level-2A:** Surface reflectances in cartographic geometry. Can be used directly in downstream applications without the need for further processing.

Sentinel-2: Images

- Satellite image products are cropped into tiles.
- Some tiles are only partially filled with valid data.
- Each spectral band has its own image file.



Sentinel-2: Spectrums

Provides high-resolution optical imaging for land services e.g.

- Imagery of vegetation, soil and water cover
- Inland waterways and coastal areas

Multi-spectral data with 13 bands in the visible, near infrared, and short wave infrared. Revisits every 5 days.

Sentinel-2 Bands	Central Wavelength (μm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

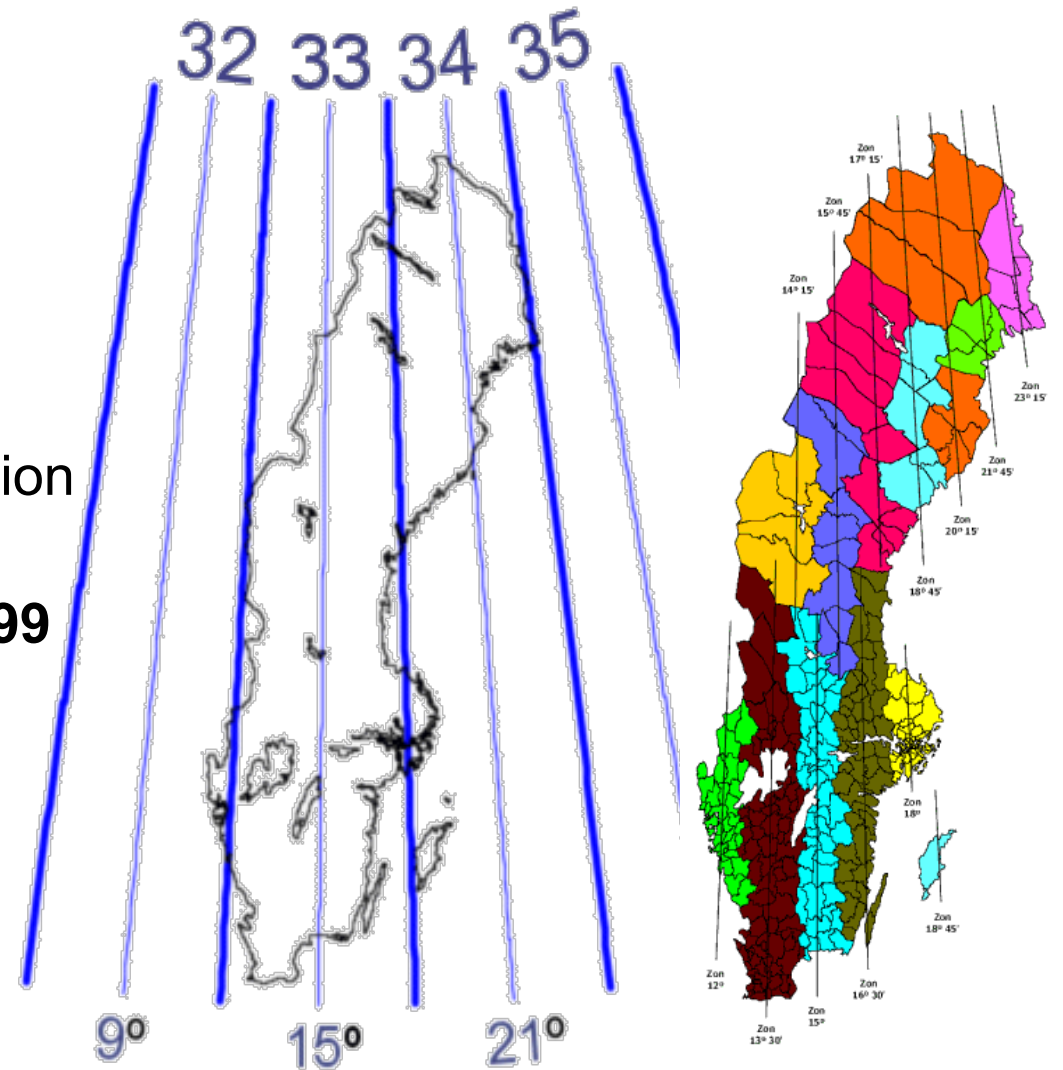
Open Data Cube: Ingesting

During ingestion the Sentinel **JPEG** images are extracted and converted to the **NetCDF** format.

Sentinel images use the combined UTM projection and WGS84 ellipsoid coordinate system.

Lantmäteriet recommends using the **SWEREF 99 TM** coordinate system for national applications.

Ingested data can be automatically transformed to other coordinate systems and stored in analysis ready format. **Storage space** is the only limitation.



Open Data Cube: User interaction

ODC provides a Python Application Programming Interface (API) for querying data. Various third party tools can be used for processing, e.g.

- **NumPy** - Provides large, multi-dimensional arrays and matrices and high-level mathematical functions
- **Matplotlib** - Image plotting library
- **Pandas** - Data analysis toolkit
- **Shapely** - Set-theoretic analysis and manipulation of planar features (points, lines, polygons)

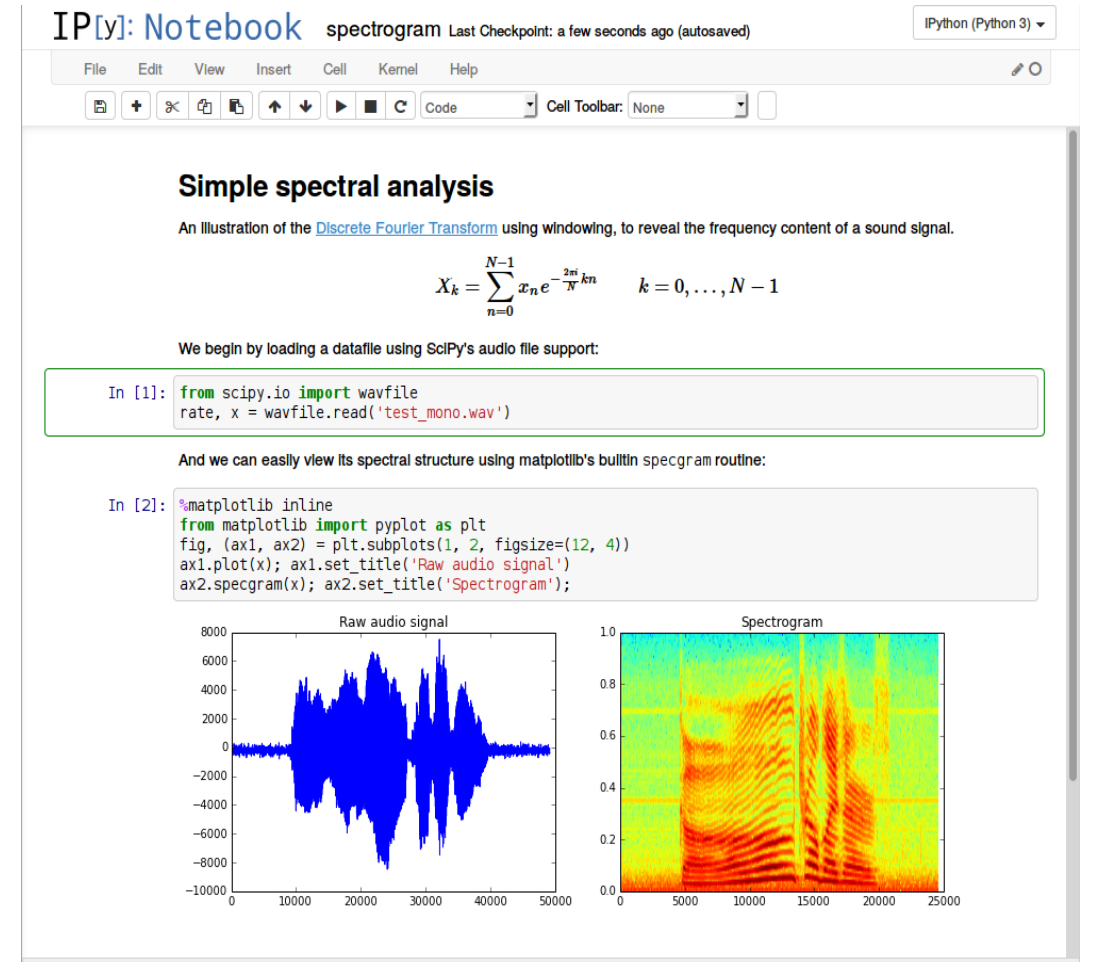
Python code for image processing can be written in normal **text files**, command line **interpreter**, or through **Jupyter Notebook**.

Jupyter Notebook

Jupyter Notebook is a web-based interactive computational environment. Some of its features are

- Interactive read-eval-print loop Python **code execution**
- Display **plots** directly on the web page
- Code **documentation** through Markdown language

Multiple users environment supported through **JupyterHub**.



The screenshot shows a Jupyter Notebook titled "spectrogram" with a Python 3 kernel. The notebook content includes:

- Simple spectral analysis**: An illustration of the Discrete Fourier Transform using windowing to reveal the frequency content of a sound signal.
- Equation**:
$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \dots, N-1$$
- Code Cell 1**:

```
In [1]: from scipy.io import wavfile
rate, x = wavfile.read('test_mono.wav')
```
- Code Cell 2**:

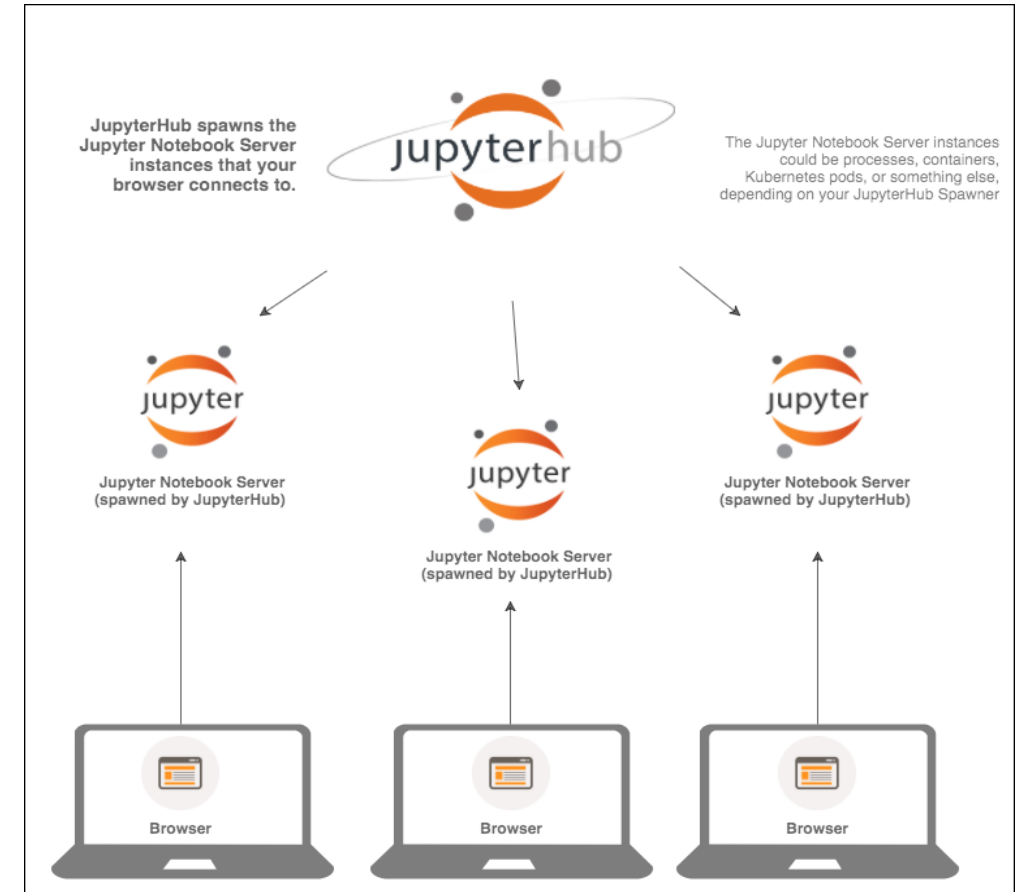
```
In [2]: %matplotlib inline
from matplotlib import pyplot as plt
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
ax1.plot(x); ax1.set_title('Raw audio signal')
ax2.specgram(x); ax2.set_title('Spectrogram');
```
- Plots**: Two side-by-side plots. The left plot, titled "Raw audio signal", shows a blue waveform of the audio signal over time (0 to 50,000 samples). The right plot, titled "Spectrogram", shows a heatmap of the signal's frequency content over time (0 to 25,000 samples).

JupyterHub

Multi-user server which supports spawning and managing multiple Jupyter Notebook instances.

Users have access to **shared data** as well as **private data**.

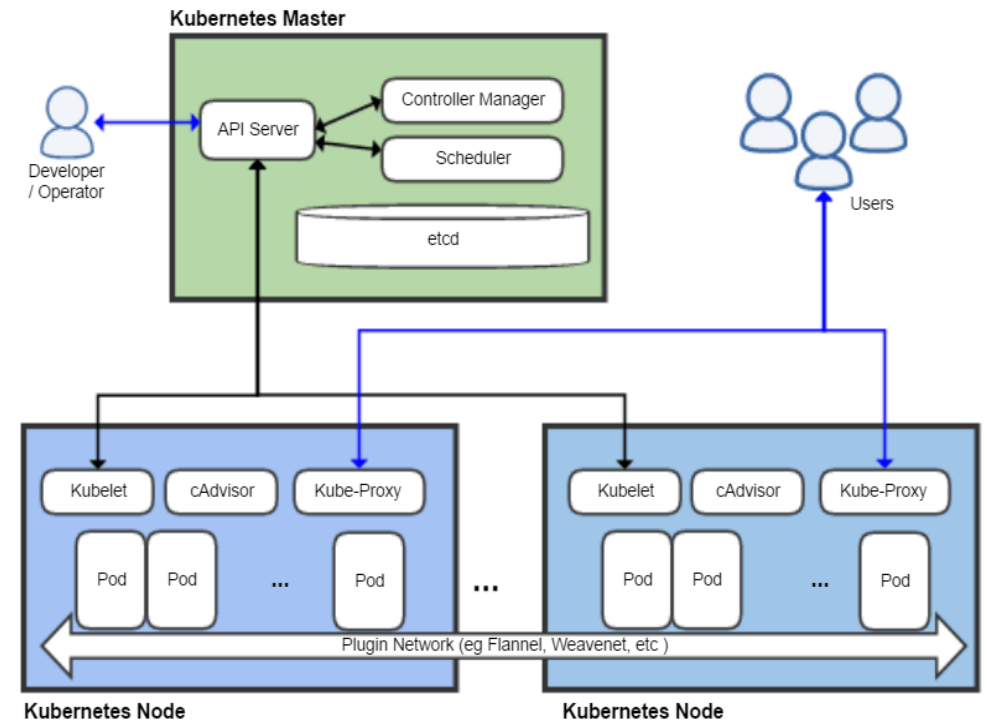
Image analysis requires a lot of processing power and memory. Computational resources can be managed by **Kubernetes**.



Kubernetes

Kubernetes is "a platform for automating deployment, scaling, and operations of application containers across clusters of hosts"

- **Containers** are groups of applications isolated from the main operating system
- **Pods** are groups of containers guaranteed to run on the same node
- **Nodes** are (physical) machines where pods are run
- **Services** are sets of pods that work together, such as web servers using shared databases

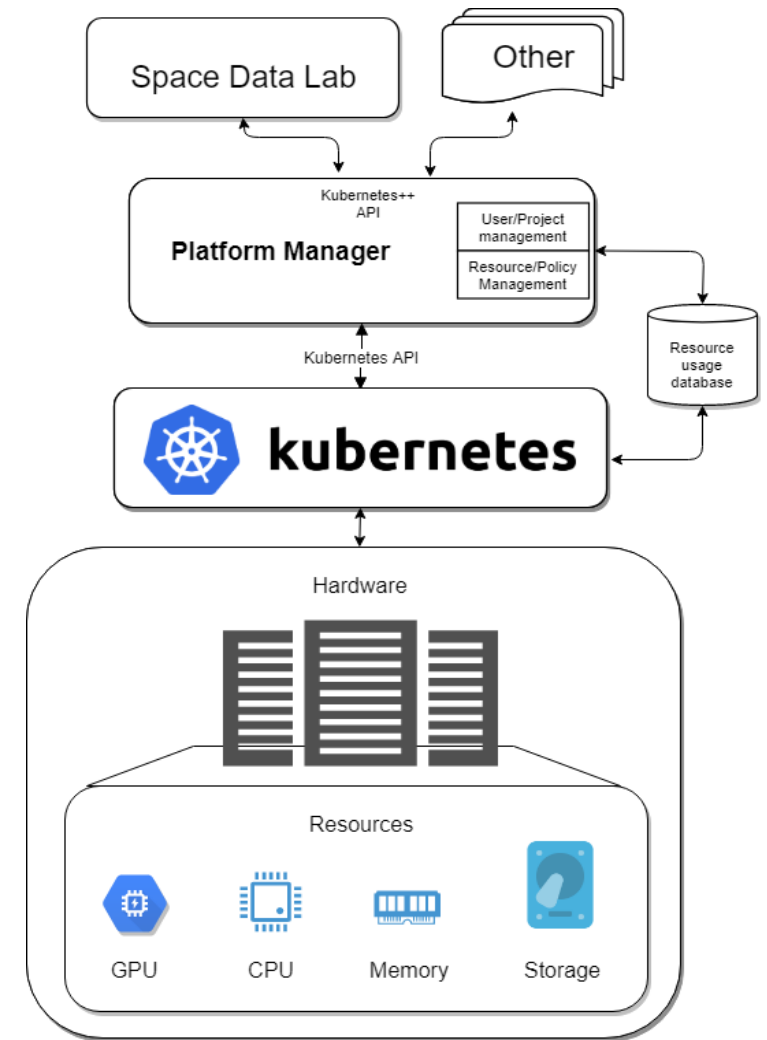


Space Data Lab on ICE

JupyterHub on Kubernetes

The ODC project provides pre-made Docker **containers** which can be run on Kubernetes.

This allows efficient distribution of computational resources for multi-user JupyterHub notebooks. Each user can be automatically assigned a server to work with, or possibly multiple servers, for heavy image analysis workloads.

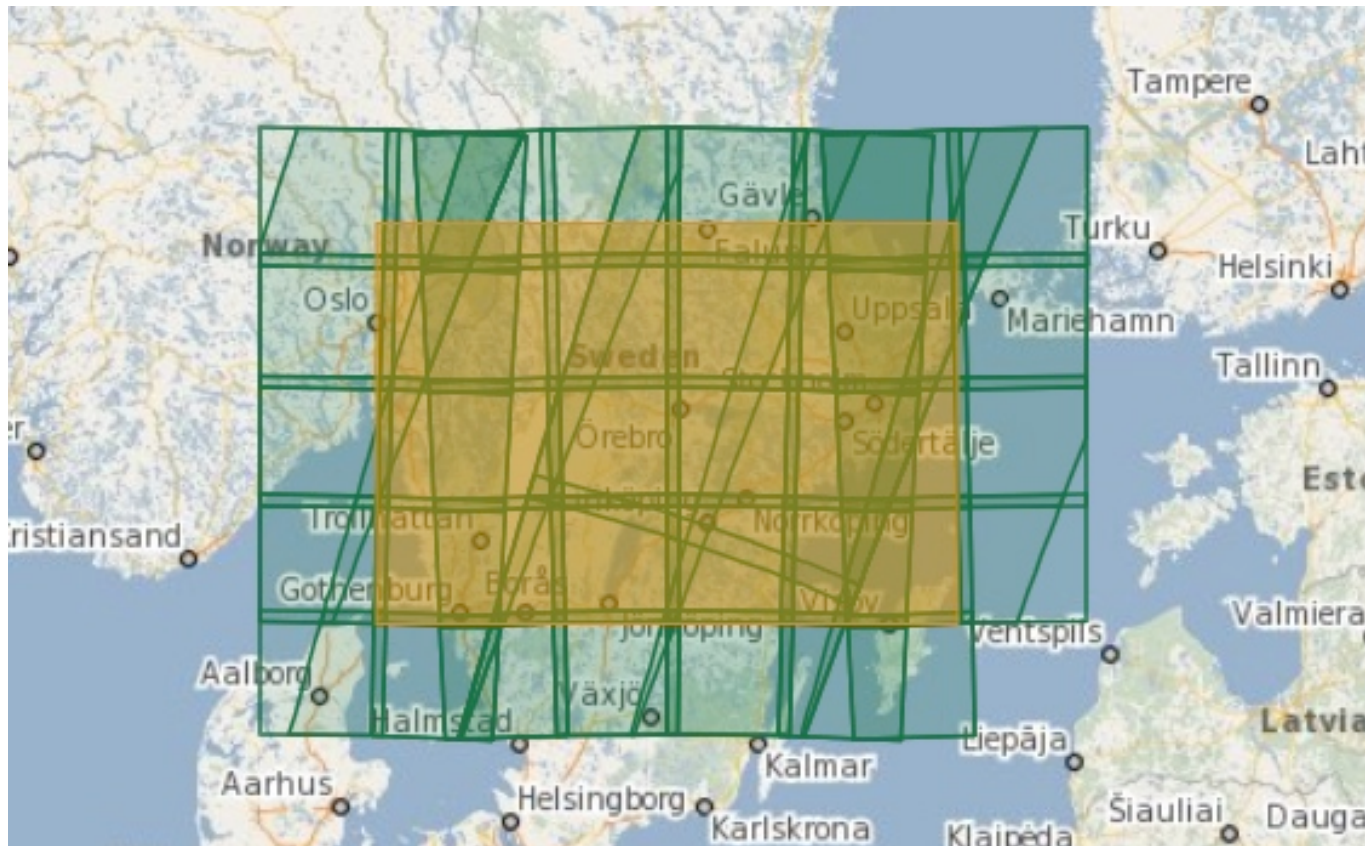


Current status

- Sentinel 2 Level 1C data from 16/12 2016, takes 3.6 TB. New data ~0.4 TB per month.
- Sentinel 2 Level 2B data from 26/3 2018 to now, takes 7.3 TB. Latest images are downloaded as they become available, ~0.5 TB per month.
- Experimental setup for Open Data Cube + JupyterHub on Kubernetes:
- Shared database for satellite images (PostgreSQL).
- One admin node can download and add new images.
- Multiple user nodes can read the database. Work notebooks and other data stored on persistent drives. Users environments are isolated from each other.



Current status



Green is the satellites passing by the area

Yellow is area of downloaded data