

# FY2023 - Science User Support for Subaru Telescope Observing Data at Astronomy Data Center (ADC)

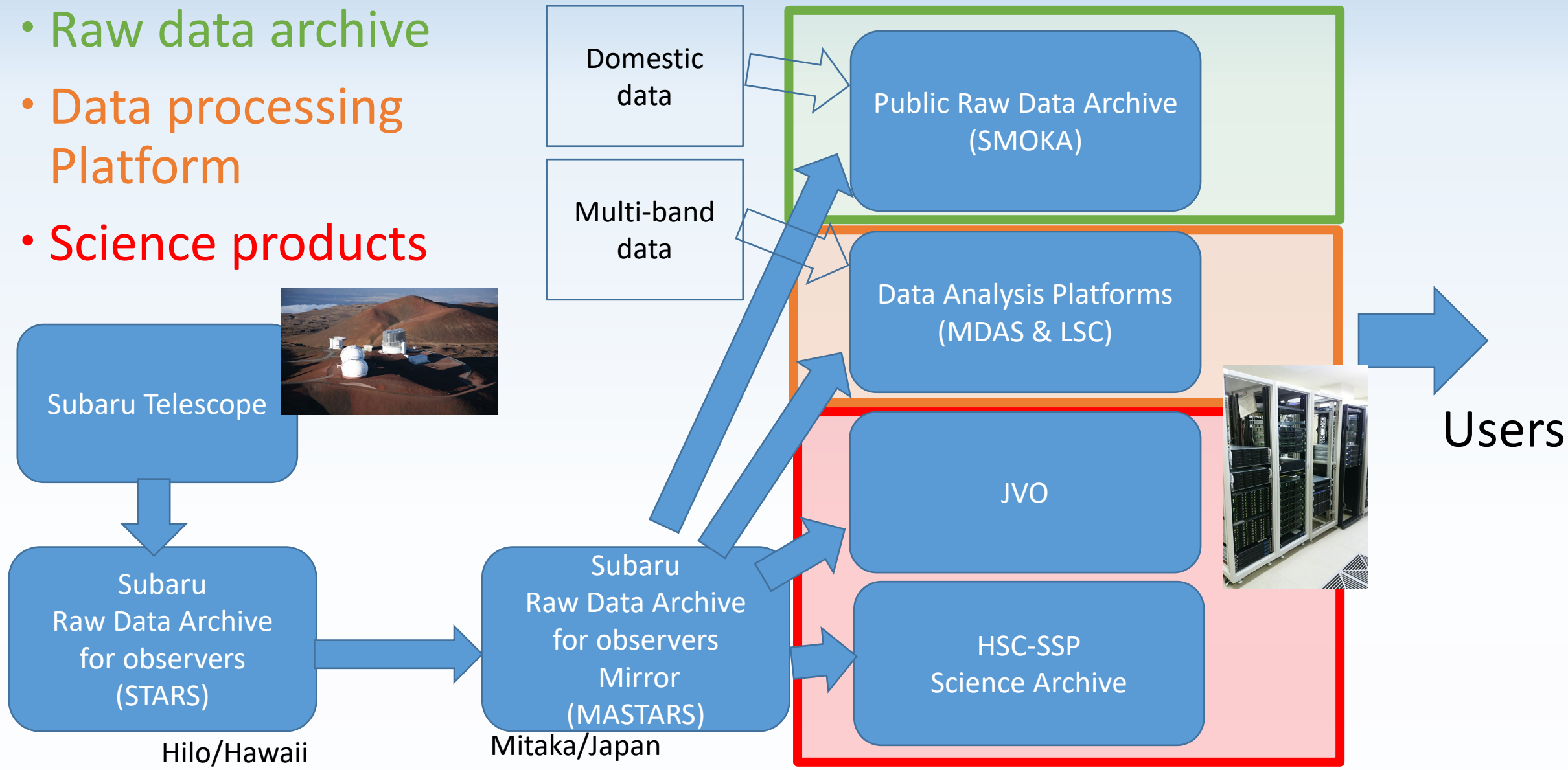
Hisanori Furusawa

for ADC/Subaru open-use support team

1/24/2024 Subaru Users Meeting FY2023

# Data Flow and Services for Subaru Data Sciences

- Raw data archive
- Data processing Platform
- Science products



# 1. Raw Data Archive – This fiscal year

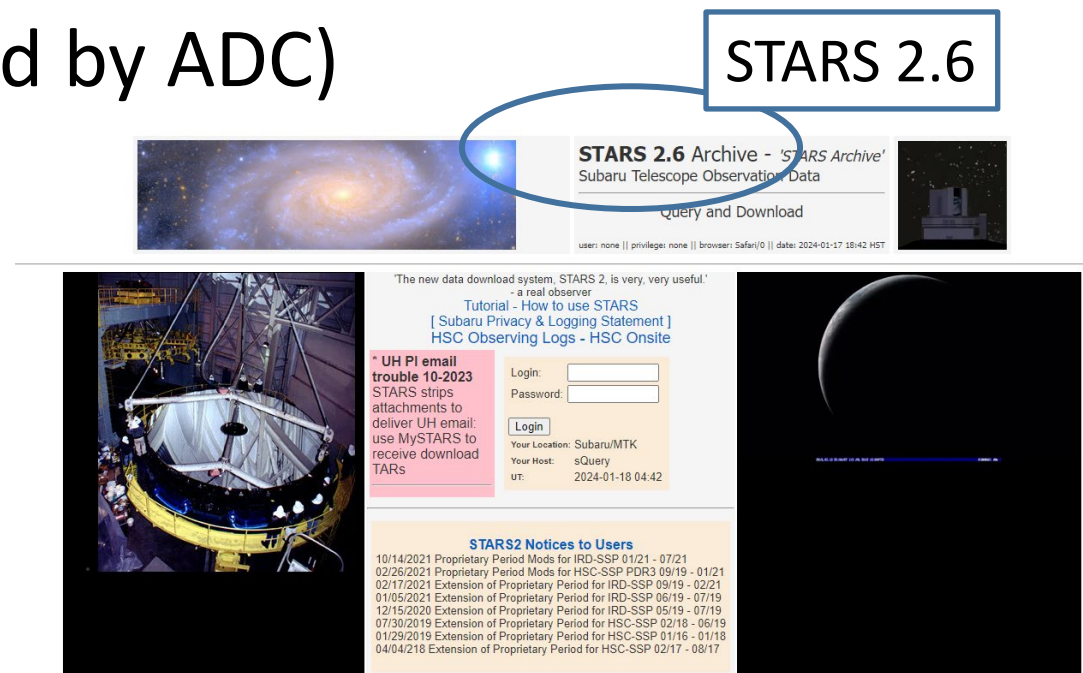
- **Computer System Update** for renewal of rental contract every 5-6 yrs (a.k.a **Replacement**) both in Hilo and Mitaka sites
- **Status**
  - Procurement & contract – done
  - Hilo) Transition is almost done (by May 2023) – See Kody's report
  - Mitaka) Designing detailed config. & Preparation of system migration
- **Timeline (Mitaka)**
  - August 2023 – intermediate transition : some of the components removed
  - July 2024 – New contract starts; completely moved to new systems
  - Transition will happen during the next fiscal year - slightly different from the past updates - Trying to make the interruptions minimal.

# 1-1. STARS/MASTARS

<https://stars.naoj.org/>

- STARS (Hilo)
  - Code change (Python 2 -> 3) – done
  - System migration to STN6-VM DB-query servers + new disks – done
  - **Coordination** for various interfaces with Mitaka archives – ongoing
  - Some **new features** – efficient handling units of HSC exposures etc - done
- MASTARS (Mitaka; system being maint.d by ADC)
  - Preparing for system **migration**
  - **Possible short down times in May-June**
- Ongoing / future work
  - Accepting new instruments
    - PFS, SCExAO modules etc
  - The next-generation archive system

STARS 2.6



# 1-2. SMOKA (Public data archive)

- See Uchiyama-san's report (this morning)
- Also [System Migration](#) for Replacement (w/ not-small changes)
- Trying to [maintain major archival functions](#)
- Discussions and Preparation
  - SCExAO data (FastPDI, MEC,)
  - [PFS](#) data
    - Relating data sets necessary for data analysis
  - Connection with the processing platform
  - Mechanism for allowing remote storage

The screenshot shows the 'SMOKA Archive Advanced Search' web interface. It features a teal header with the title and two links: 'Click here for SUP Search (Suprime-Cam data Search)' and 'Click here to know how to search.' Below the header is a search form with several sections: 'Object Name (for name resolve)' with a text input and a 'Resolver' section with radio buttons for 'SIMBAD' (selected), 'NED', and 'Don't Resolve', plus a 'Resolve' button. The 'Coordinate System' section has a dropdown for 'Equatorial' and a radio button for 'Circular' (selected). The 'Field of View (arcmin)' section has a dropdown for 'auto'. The 'Observation Date' section has a text input. The 'Exp Time (sec)' section has a text input. The 'Observer' section has a text input. The 'center (RA)' and 'center (DEC)' sections have text inputs and 'copy/paste to' buttons. The 'Radius(arcmin)' section has a text input with the value '10.0'. The 'From (RA)' and 'To (RA)' sections have text inputs and 'Opposite Corner Coordinate' labels. The 'From (DEC)' and 'To (DEC)' sections have text inputs and 'Opposite Corner Coordinate' labels.

<https://smoka.nao.ac.jp/>

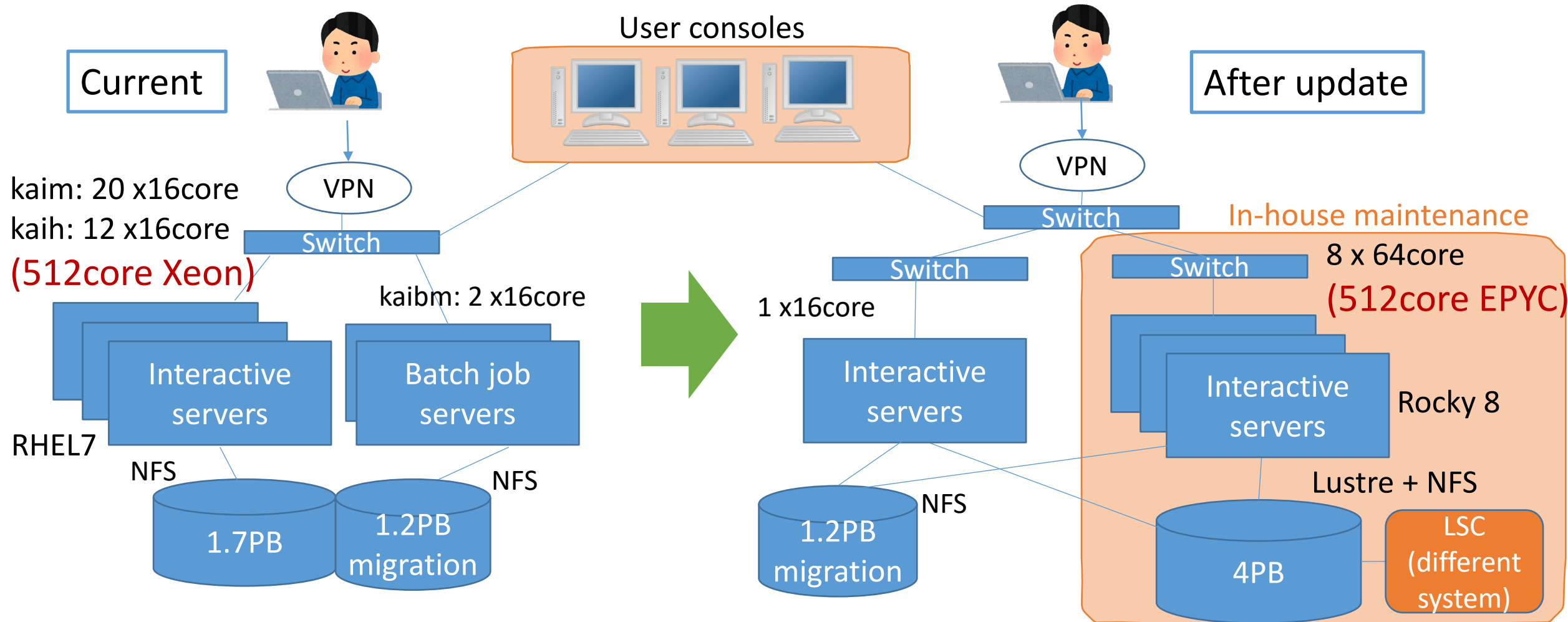
## 2. Data Processing Platform – This fiscal year

- MDAS (Multi-wavelength Data Analysis System)
  - Role) General interactive data analysis and drafting papers etc
  - Also experiencing System Migration
  - New system will largely depend on in-house components
- Large-Scale data analysis system, or PC Cluster (LSC)
  - Role) Intensive non-interactive data analysis (HSC) over batch job system
  - Trying to cope with system aging (5<sup>th</sup> year) – discussing storage update
  - Adjusting Queue Configuration etc

# 2-1. MDAS

[https://www.adc.nao.ac.jp/MDAS/mdas\\_e.html](https://www.adc.nao.ac.jp/MDAS/mdas_e.html)

## System update in FY2024 (<~ July)



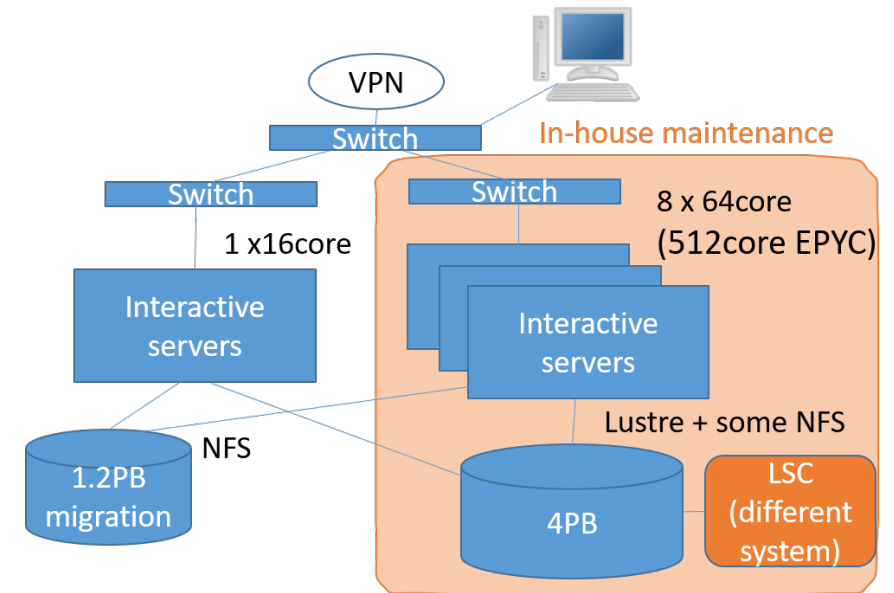
We will maintain the same scale of CPU and memory with larger storage area

## 2-1. MDAS

[https://www.adc.nao.ac.jp/MDAS/mdas\\_e.html](https://www.adc.nao.ac.jp/MDAS/mdas_e.html)

- Timeline
  - Delivery of the in-house system – Feb 2024
  - Migration Feb – Jun 2024
  - Stable operation with new system (rental + in-house) – Jul 2024
- Related events
  - Renewal of accounts as usual – Apr 2024
  - Will set a ~2month period for securing your necessary files (~May-Jun 2024)
  - Will expect some short downtime

Your cooperation appreciated





## 2-2. LSC updates

- Continue to be cooperated by ADC-Subaru
- Discussion for maintaining the LSC function
  - GPFS 5PB in the 5<sup>th</sup> year with support contract expiring
  - Coping with shortage of GPFS capacity, with cooperation by users
  - Preparing external disks (Lustre), 1.5PB (last year) + 2PB (this year)
  - Maintaining existing computing nodes with OS being updated
- Adjustment of resource usages
  - Prioritized HSC-SSP jobs until next Spring (~1500core)
  - Tuned queue config. to prevent ql jobs from being suspended forever by qm jobs

2296 core / 40 nodes on 5PB GPFS

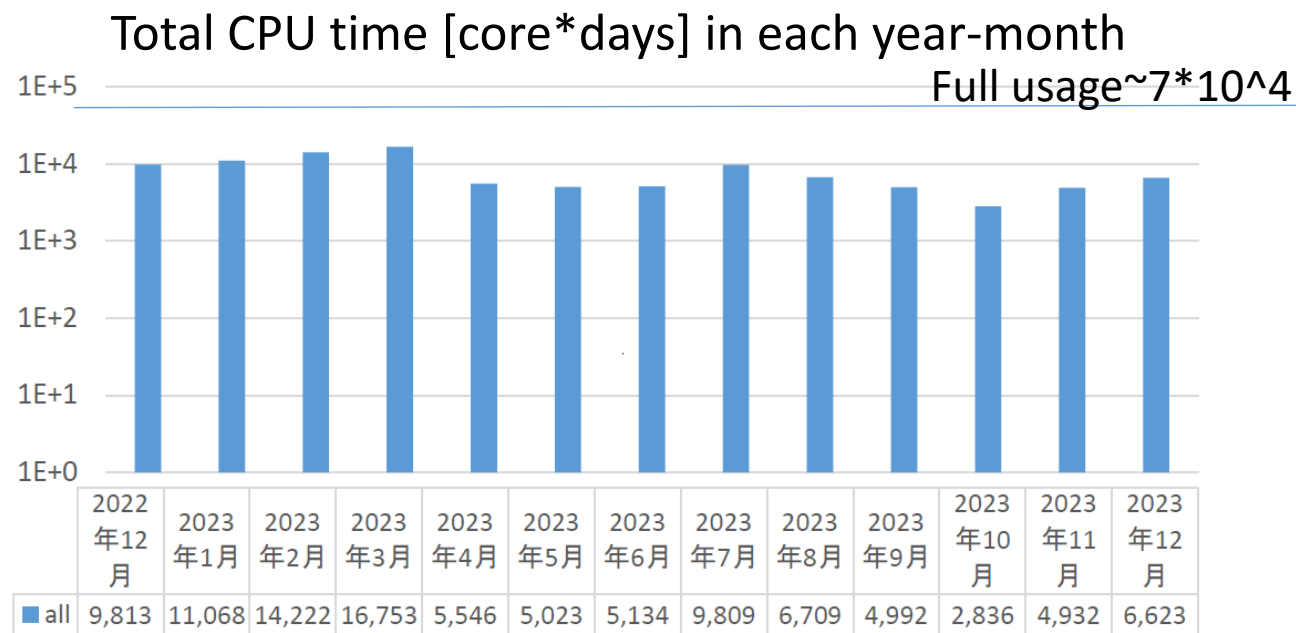
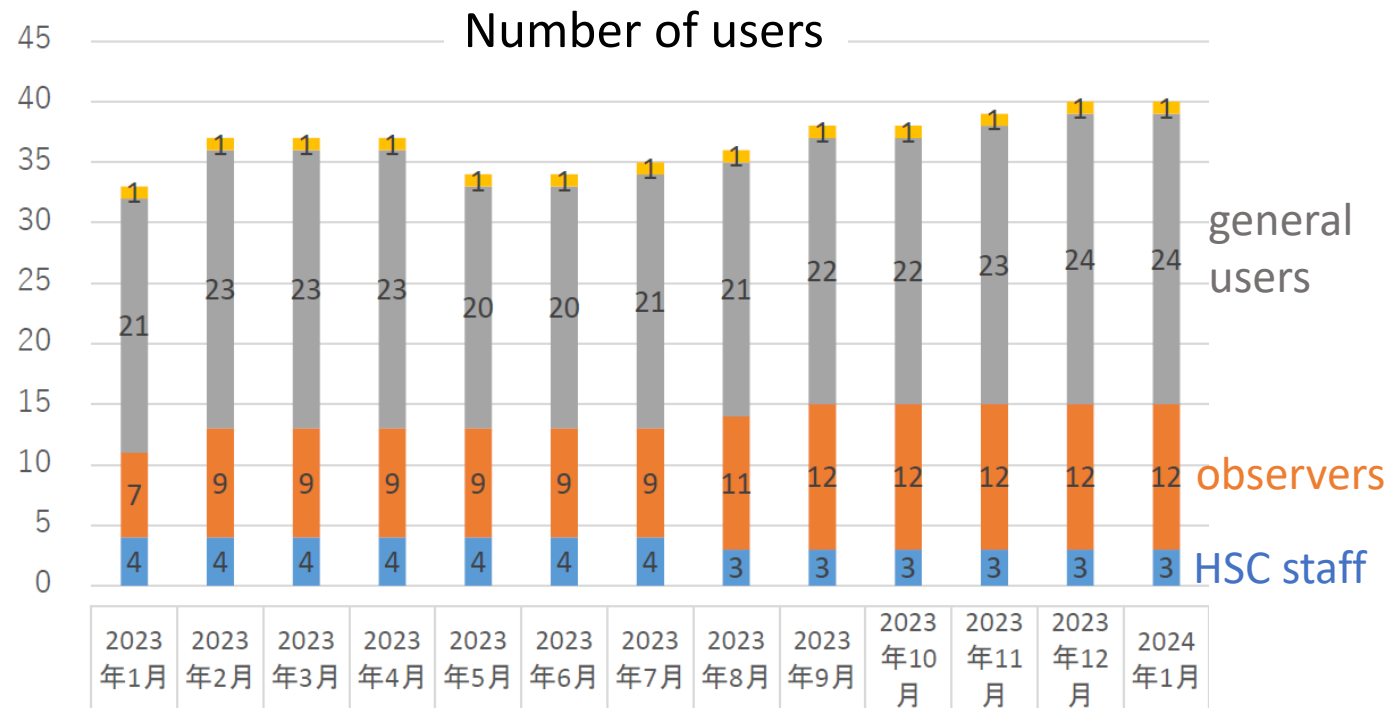
Your cooperation & comments appreciated



# LSC System Usage

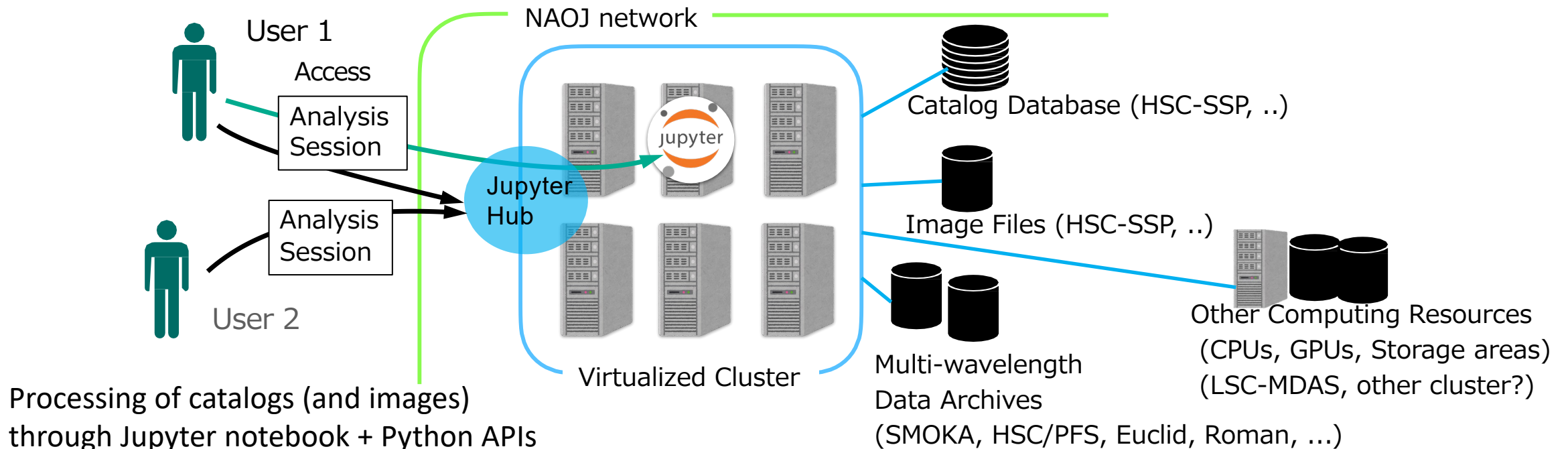
See the uploaded slides later

- User (increased)
  - 40 users
  - 12 observers & 24 general users
- CPU occupation (same as FY23)
  - PBS jobs only (without HTConodr) included in this calculation
  - >~20% CPU time on average
  - >60-70% in peak
  - Partly depends on HSC-SSP tasks
- Storage usage
  - 4.4PB / 4.5PB (98%)
  - Removed old HSC-SSP products
  - New SSP processing underway



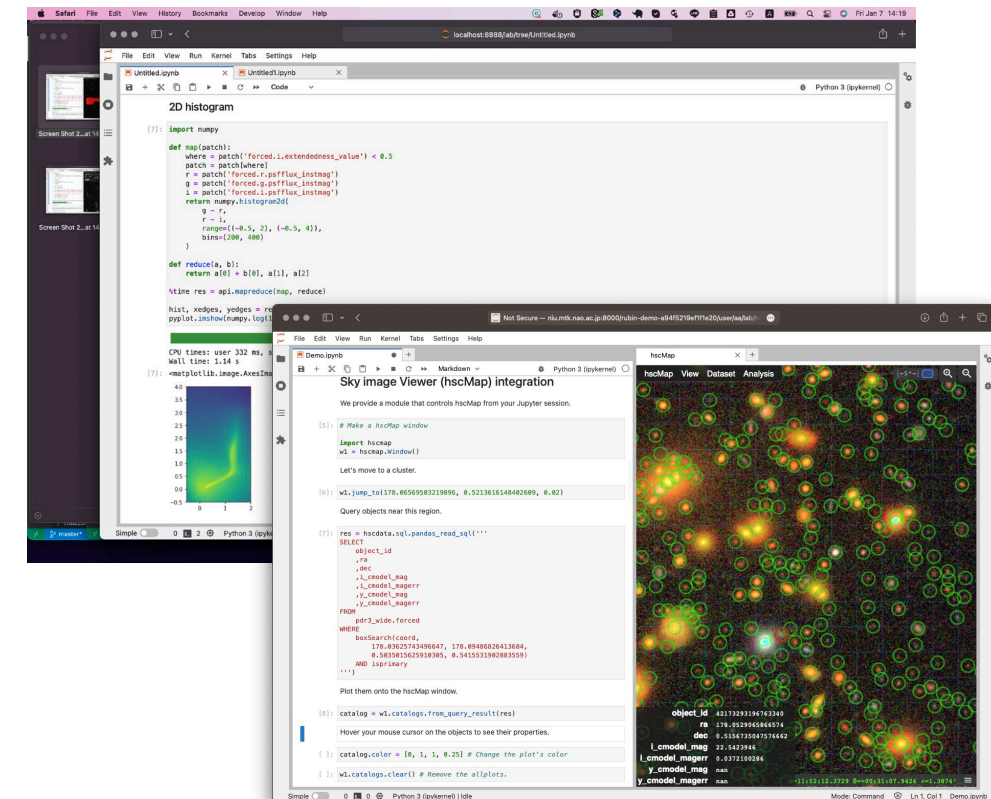
# 3. Science Products: Science Platform

- Developing a JupyterHub-based data analysis platform
  - efficient analysis over the existing products from remote
  - efficient use of computing resources
- ADC+Subaru coworking to implement services to HSC and PFS sharing the software design
- SP will also be applied to Rubin Japanese data access center, & Euclid, etc..



# Prototype of HSC Science Platform

- Prototyping SP on the HSC data release PDR3 as extension func.
- ADC Team Tentative target plan
  - First prototype -2023.4 – done – being updated
  - Internal review by a few experts
  - 2024.2 – being updated
- Preview by SSP collaboration 2024Summer
- Design for PDR (and public data) this year
- Serving PFS engineering data in 2024.2



# Demonstration / Snapshot of prototype - quickdb

QuickDB

```
[ ]: import sspsurveydb.client

Api = sspsurveydb.client.Api
api = Api(default_rerun='s20a_wide')

api.healthy()
```

Seeking the entire HSC-SSP large table with 700M rows

Counting All Objects

SQL Interface

In QuickDB, you can monitor the progress of the running query with a progress bar.

```
[ ]: api.sql('select count(*) from s20a_wide').dataframe()
```

map, reduce Interface

map, reduce Interface The following code defines map and reduce functions in Python code to count all objects in the s20a data set similarly to the SQL query above.

- The SSP observation area is divided into patches, each covering approximately 0.2 square degrees.
- The map function receives one catalog corresponding to a patch.
- The reduce function aggregates the results from map.
  - `final_result = reduce(map(patch0), reduce(map(patch1), reduce(map(patch2), reduce(map(patch3), ...))))`
- `patch.size` represents the number of objects in a patch.

```
[ ]: def map(patch):
    return patch.size

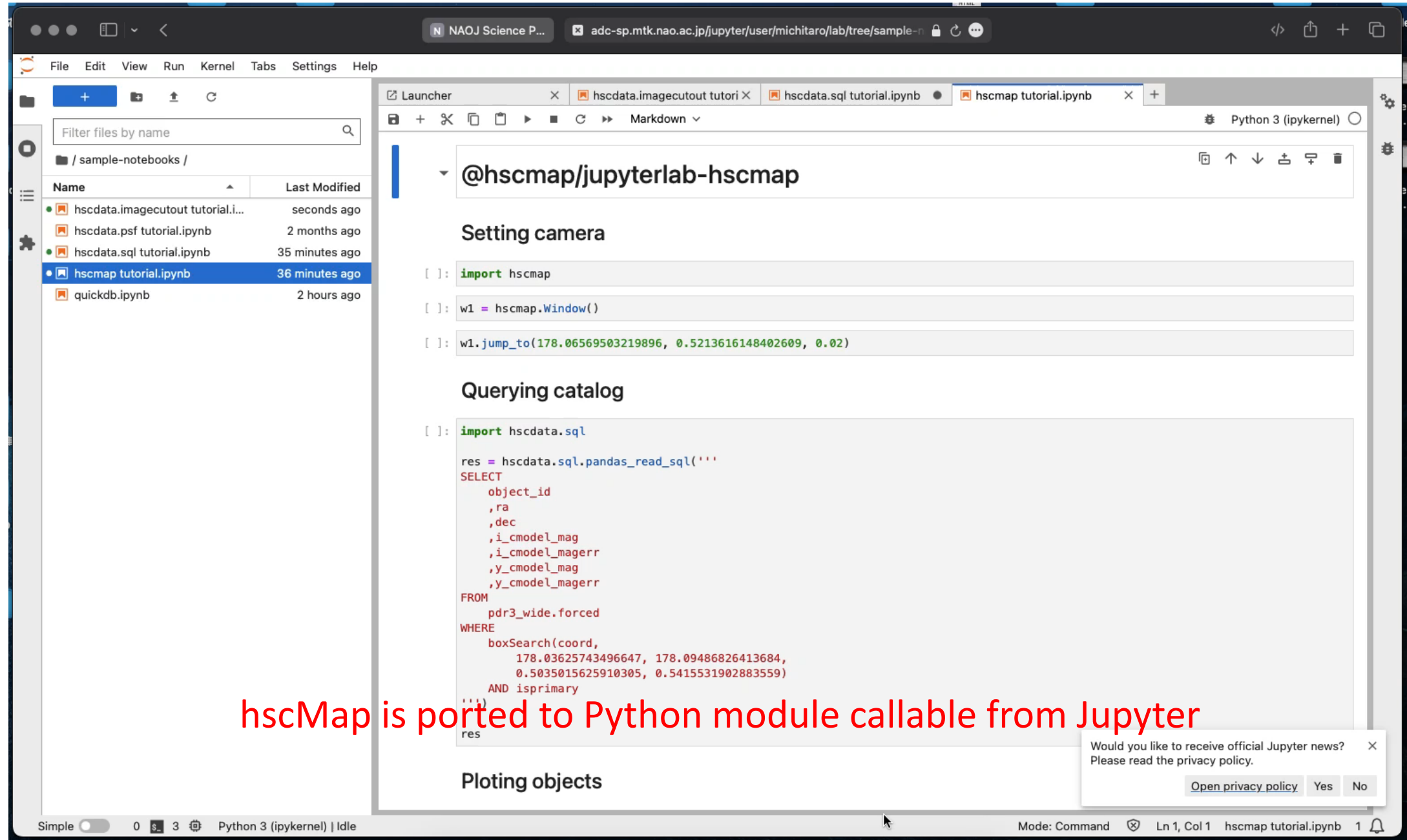
def reduce(a, b):
    return a + b

%time api.mapreduce(map, reduce, rerun='s20a_wide')
```

Throwing Python calculation code to distributed DB servers upon query

Simple 0 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 quickdb.ipynb 1

# Demonstration / Snapshot of prototype - hscMap



The screenshot displays a JupyterLab environment with a file browser on the left and a notebook editor on the right. The file browser shows a directory named `/sample-notebooks/` containing several files, with `hscmap tutorial.ipynb` selected. The notebook editor shows the `@hscmap/jupyterlab-hscmap` notebook, which is currently in `Markdown` mode. The notebook content includes sections for setting a camera, querying a catalog, and plotting objects. The status bar at the bottom indicates the current mode is `Command` and the file is `hscmap tutorial.ipynb`.

Launcher

hscdata.imagecutout tutorial X hscdata.sql tutorial.ipynb hscmap tutorial.ipynb +

Python 3 (ipykernel)

Filter files by name

/sample-notebooks/

Name	Last Modified
hscdata.imagecutout tutorial.i...	seconds ago
hscdata.psf tutorial.ipynb	2 months ago
hscdata.sql tutorial.ipynb	35 minutes ago
hscmap tutorial.ipynb	36 minutes ago
quickdb.ipynb	2 hours ago

### @hscmap/jupyterlab-hscmap

#### Setting camera

```
[ ]: import hscmap
```

```
[ ]: w1 = hscmap.Window()
```

```
[ ]: w1.jump_to(178.06569503219896, 0.5213616148402609, 0.02)
```

#### Querying catalog

```
[ ]: import hscdata.sql
```

```
res = hscdata.sql.pandas_read_sql('''
```

```
SELECT
```

```
    object_id
```

```
    ,ra
```

```
    ,dec
```

```
    ,i_model_mag
```

```
    ,i_model_magerr
```

```
    ,y_model_mag
```

```
    ,y_model_magerr
```

```
FROM
```

```
    pdr3_wide.forced
```

```
WHERE
```

```
    boxSearch(coord,
```

```
        178.03625743496647, 178.09486826413684,
```

```
        0.5035015625910305, 0.5415531902883559)
```

```
    AND isprimary
```

```
''')
```

```
res
```

#### Plotting objects

hscMap is ported to Python module callable from Jupyter

Would you like to receive official Jupyter news?  
Please read the privacy policy.

[Open privacy policy](#) Yes No

Simple 0 3 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 hscmap tutorial.ipynb 1



# Summary

- ADC offers/cooperates services of **raw data archive**, **data processing platform**, and **science products** for the Subaru community
  - STARS <https://stars.naoj.org/>
  - SMOKA <https://smoka.nao.ac.jp/index.jsp>
  - MDAS [https://www.adc.nao.ac.jp/MDAS/mdas\\_e.html](https://www.adc.nao.ac.jp/MDAS/mdas_e.html)
  - LSC [https://www.adc.nao.ac.jp/LSC/lsc\\_e.html](https://www.adc.nao.ac.jp/LSC/lsc_e.html)
  - JVO <http://jvo.nao.ac.jp/index-e.html>
  - HSC <https://hsc-release.mtk.nao.ac.jp/doc/>
- Experiencing **big computer updates** with rental contract renewal
- Continuing **HSC data production and data services**, being transitioned to long-term operation at ADC
- Developing **science platform for HSC, PFS, extended to Rubin to utilize existing and future data products and computing resources**
- Discussing **new-generation data archives**