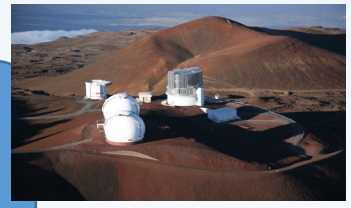
FY2023 - Science User Support for Subaru Telescope Observing Data at Astronomy Data Center (ADC)

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



Subaru Telescope

Subaru Raw Data Archive for observers (STARS)

Hilo/Hawaii

Subaru Raw Data Archive for observers Mirror (MASTARS)

Mitaka/Japan

Domestic data

Multi-band data

Public Raw Data Archive (SMOKA)

Data Analysis Platforms (MDAS & LSC)

JVO

HSC-SSP Science Archive

Users

1. Raw Data Archive – This fiscal year

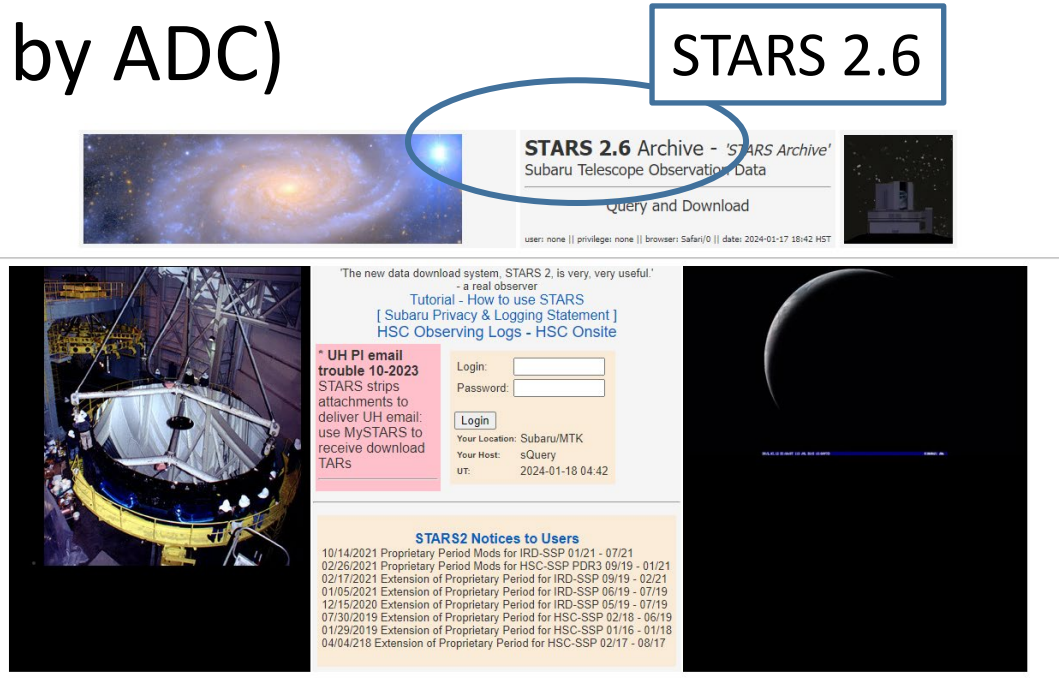
- Computer System Update for renewal of rental contract every 5-6 yrs (a.k.a Replacement) both in Hilo and Mitaka sites
- Status
  - Procurement & contract – done
  - Hilo) Transition is almost done (by May 2023) – See Kody's report
  - Mitaka) Designing detailed config. & Preparation of system migration
- 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/

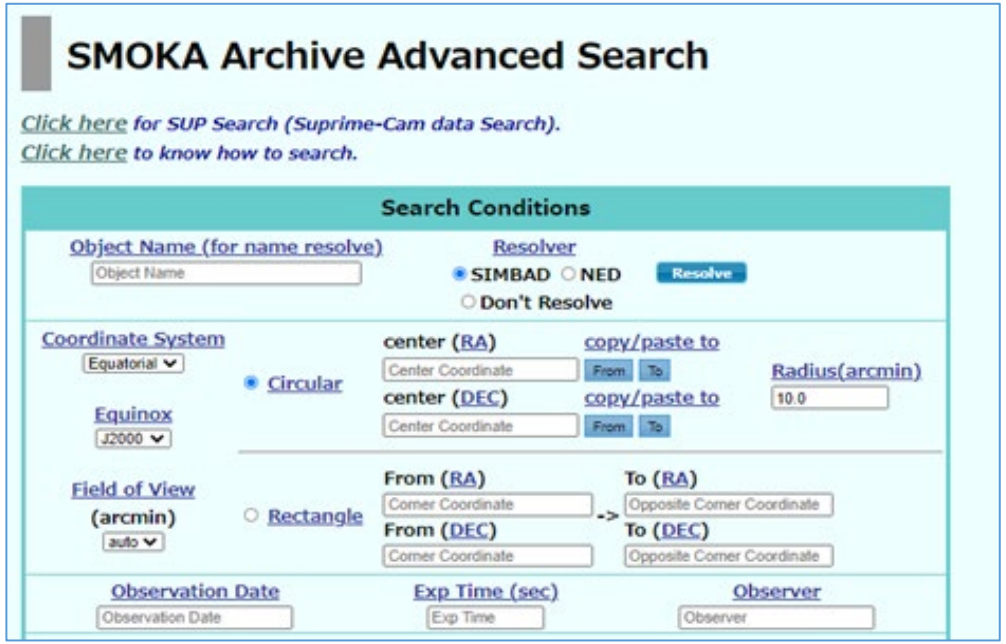
- STARS (Hilo)
  - Code change (Python 2 -> 3) – done
  - System migration to STN6-VM DB-query servers + new disks – done
  - Coordination for various interfaces with Mitaka archives – ongoing
  - Some new features – efficient handling units of HSC exposures etc - done
- MASTARS (Mitaka; system being 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



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

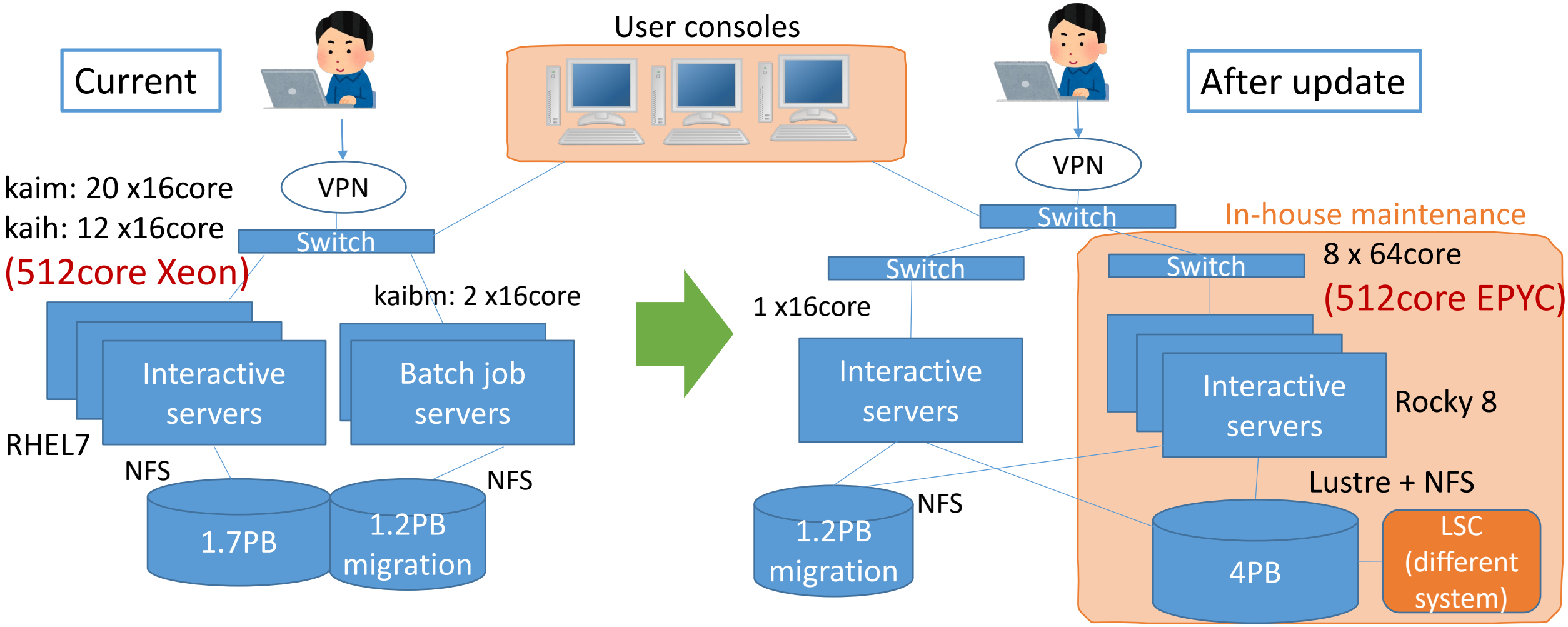
2. Data Processing Platform – This fiscal year

- 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^{th} year) – discussing storage update
  - Adjusting Queue Configuration etc

2-1. MDAS

https://www.adc.nao.ac.jp/MDAS/mdas_e.html

System update in FY2024 (<~ July)



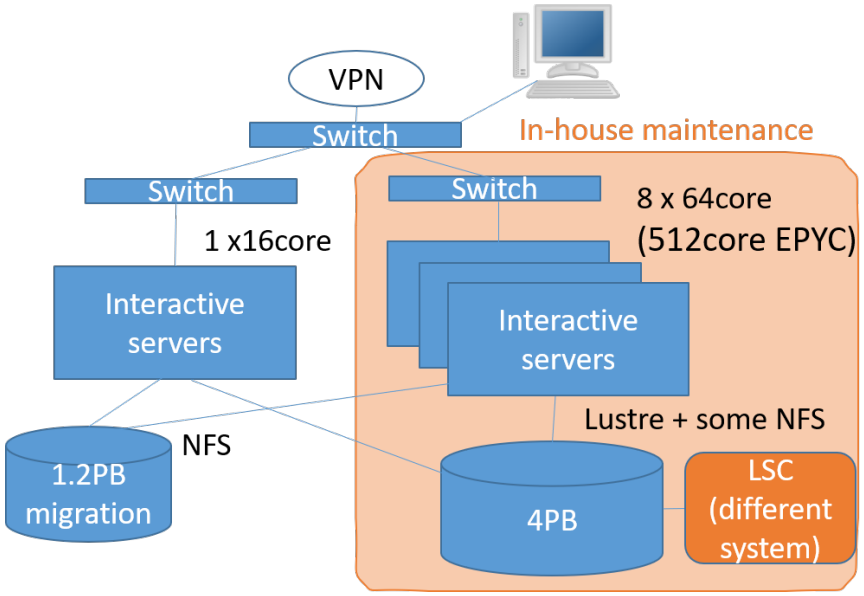
We will maintain the same scale of CPU and memory with larger storage area

2-1. MDAS

https://www.adc.nao.ac.jp/MDAS/mdas_e.html

- 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 5th 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

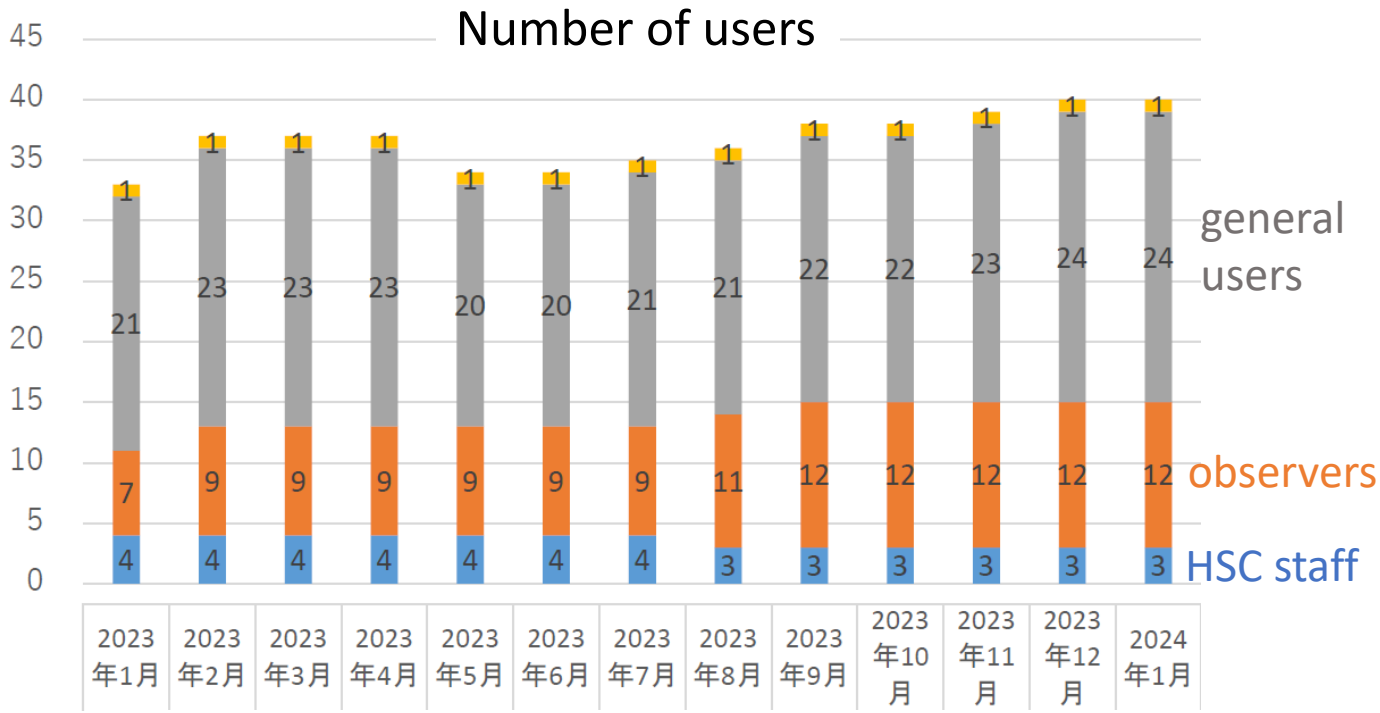
Your cooperation & comments appreciated

2296 core / 40 nodes on 5PB GPFS

LSC System Usage

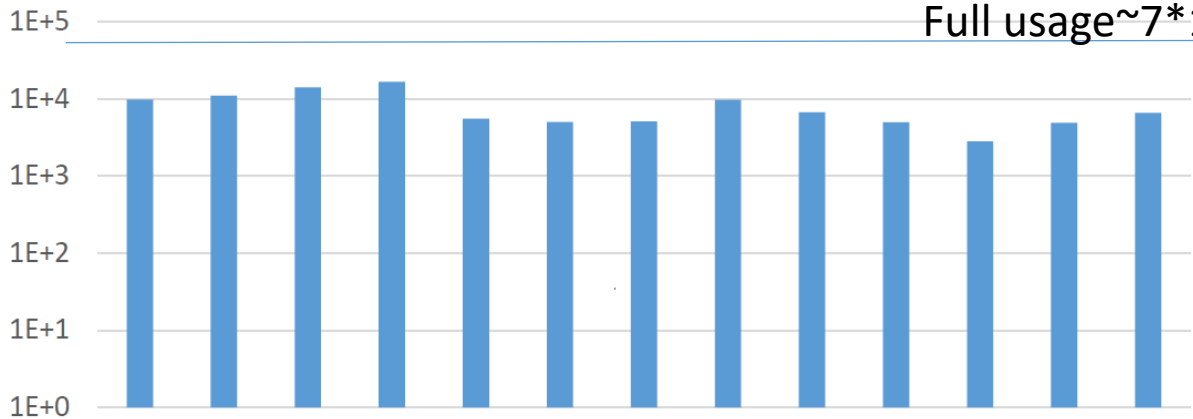
See the uploaded slides later

- 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



Total CPU time [core*days] in each year-month

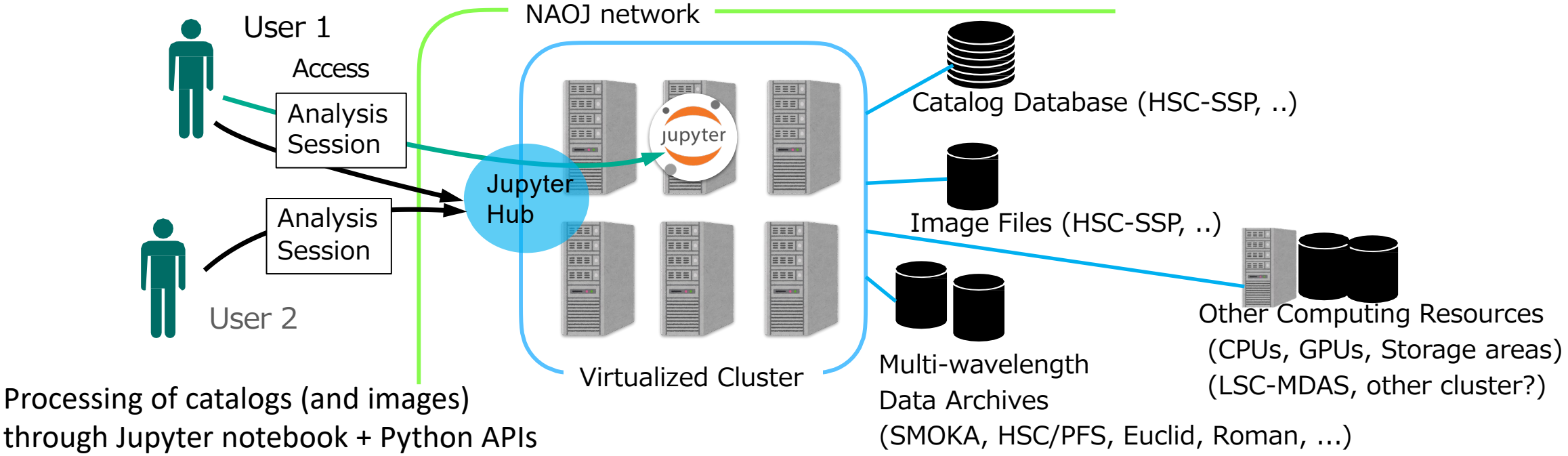
Full usage~7*10^4



	2022年12月	2023年1月	2023年2月	2023年3月	2023年4月	2023年5月	2023年6月	2023年7月	2023年8月	2023年9月	2023年10月	2023年11月	2023年12月
all	9,813	11,068	14,222	16,753	5,546	5,023	5,134	9,809	6,709	4,992	2,836	4,932	6,623

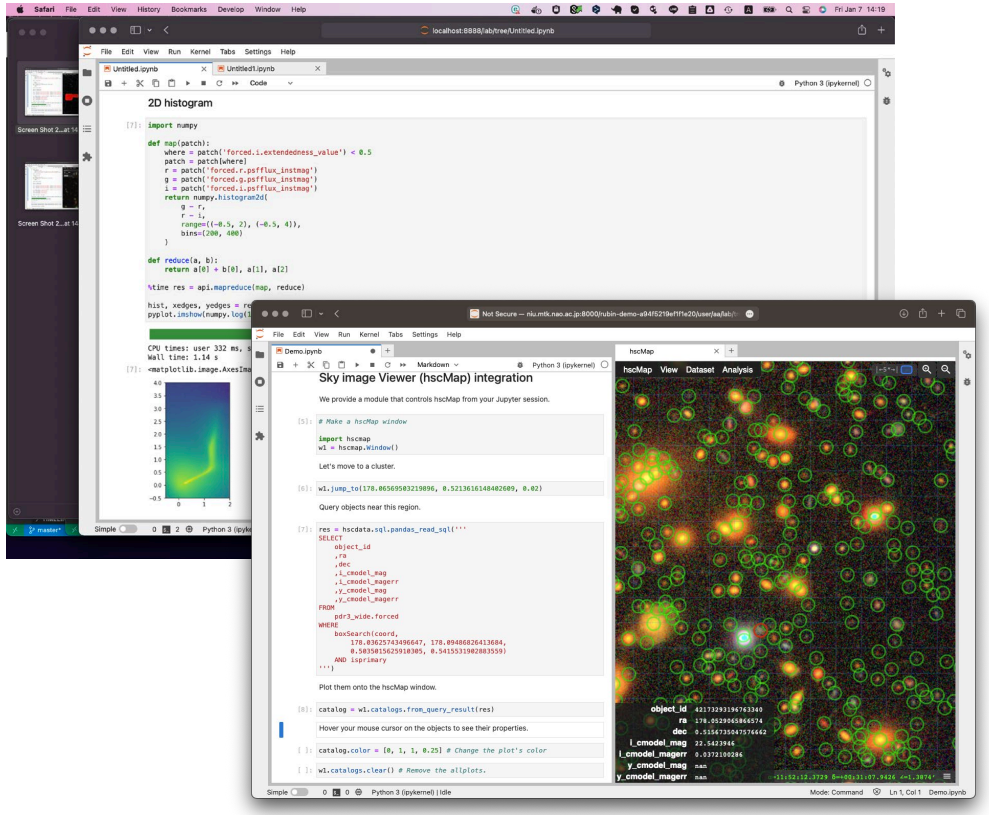
3. Science Products: Science Platform

- 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..

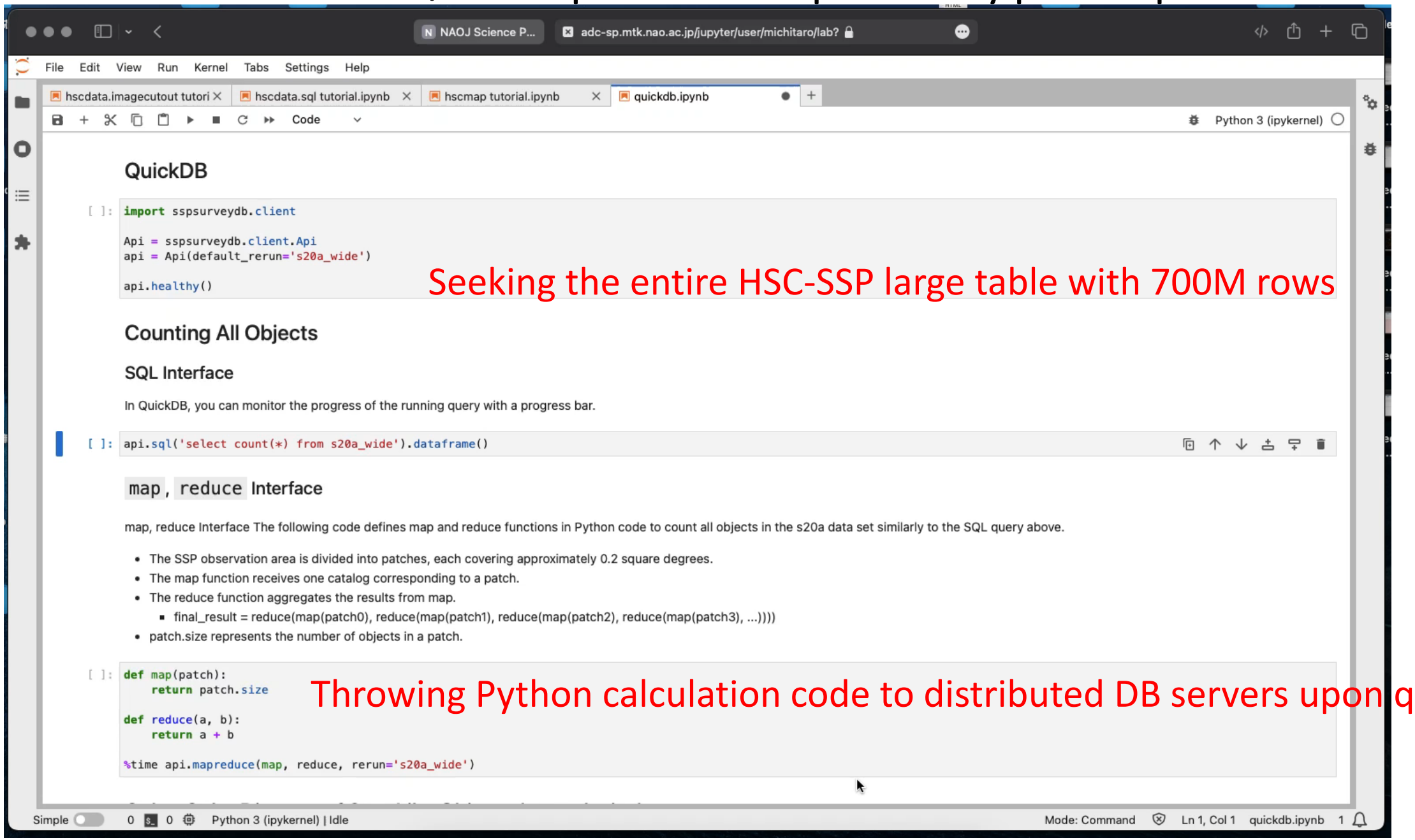


Prototype of HSC Science Platform

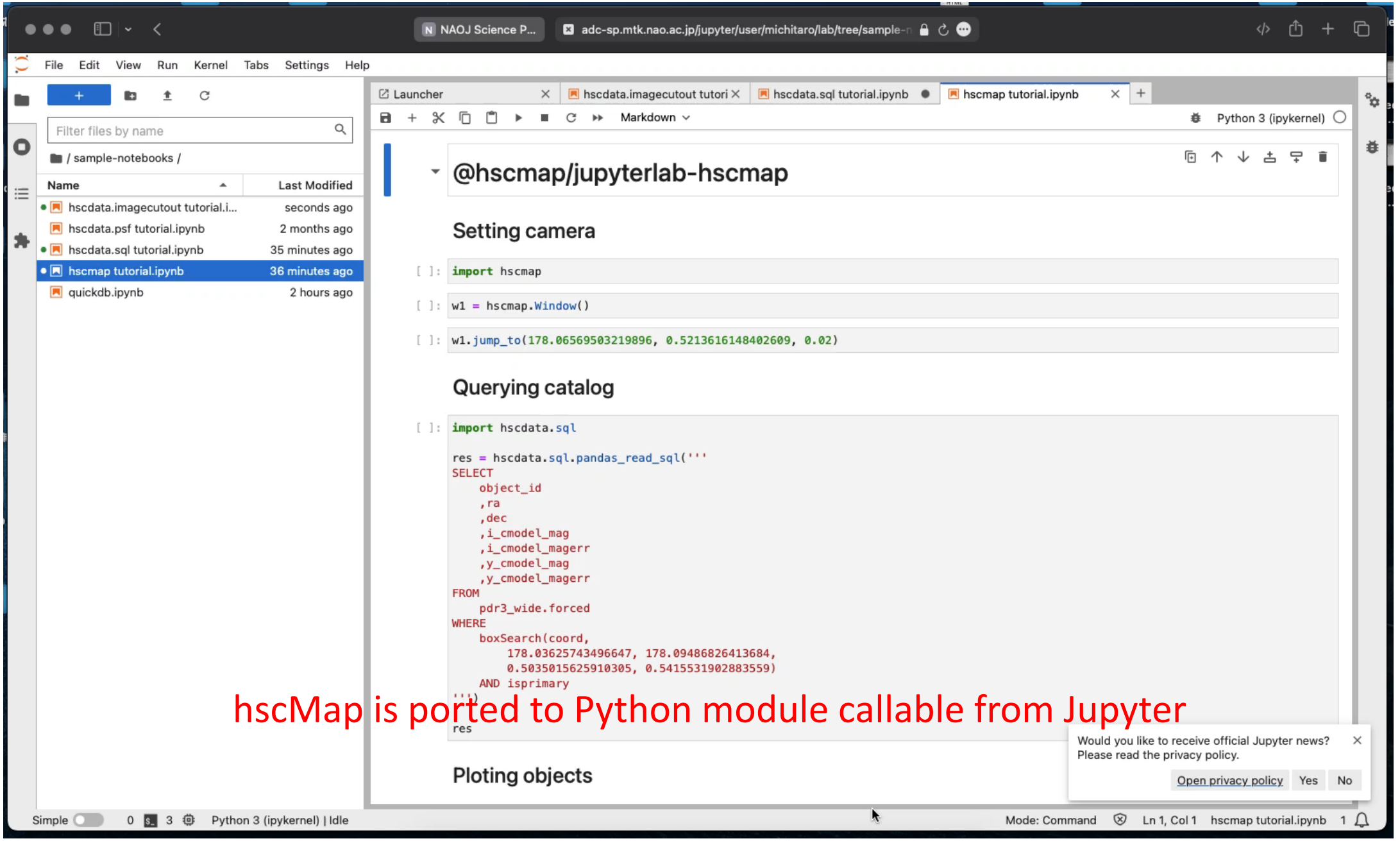
- Prototyping SP on the HSC data release PDR3 as extension func.
- ADC Team Tentative target plan
  - First prototype -2023.4 – done – being updated
  - Internal review by a few experts
    2024.2 – being updated
  - Preview by SSP collaboration 2024Summer
  - Design for PDR (and public data) this year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



Demonstration / Snapshot of prototype - quickdb



Demonstration / Snapshot of prototype - hscMap



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
  - LSC 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