

ADCのScience Platform (アプリケーション編)

小池美知太郎

2025/4/17

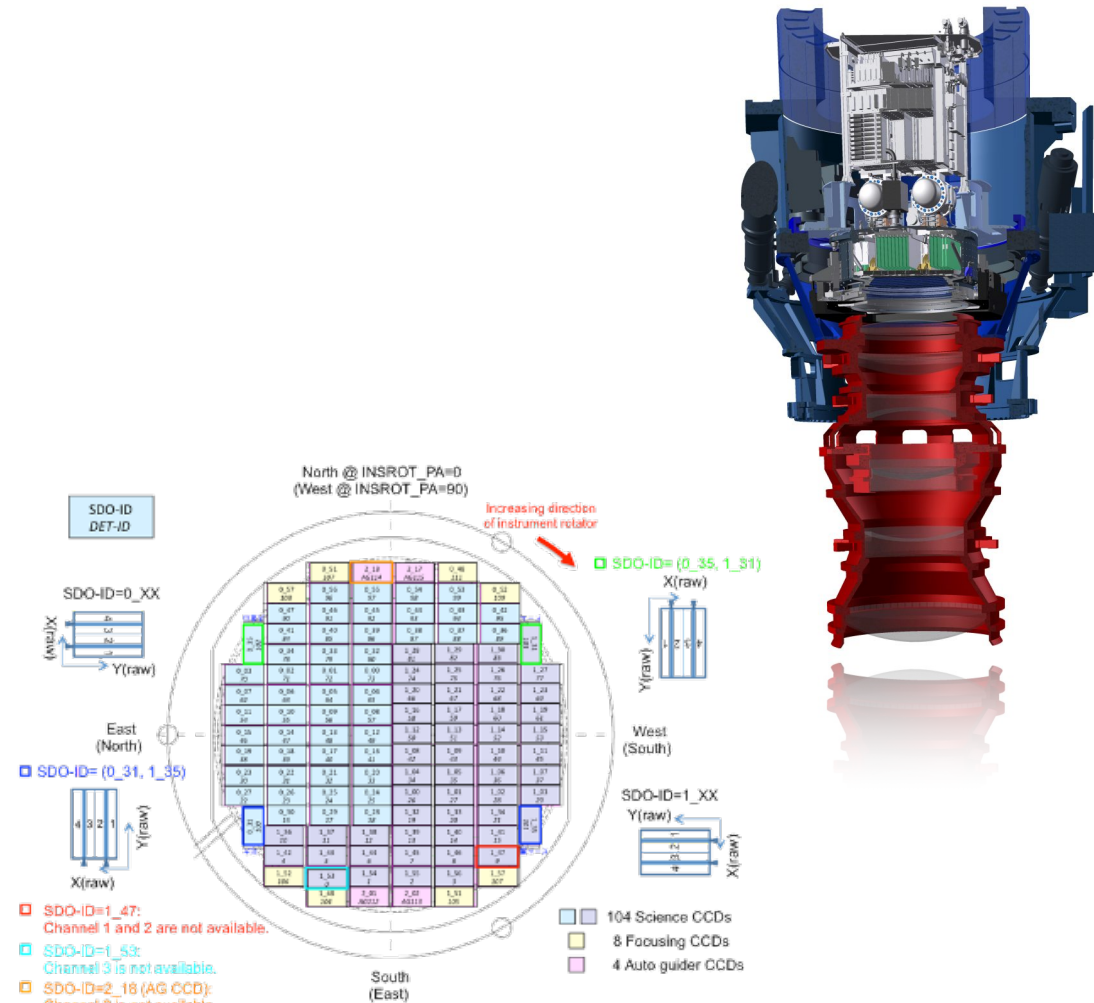
第44回天文学に関する技術シンポジウム プログラム

ADC-HSCのScience Platform (アプリケーション編)

- 目次
- 「Science Platformとは」の前にHSC SSPについて
- Science Platformとは
- Science Platformのミドルウェア・インフラ（次の森嶋さん発表）

Hyper Suprime-Cam (HSC)

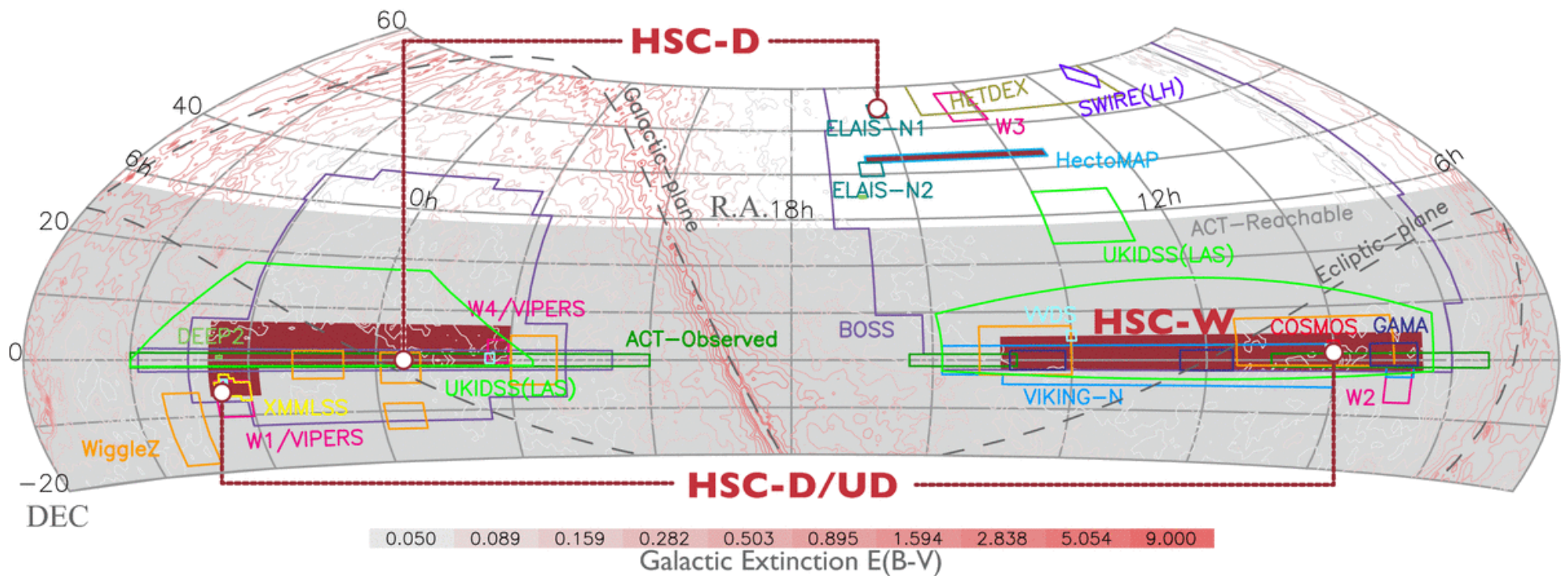
- 広視野光赤外イメージングカメラ
- 2014年から運用開始
- 視野は円形で直径 1.5 度 (面積 1.8 平方度)
- 完全空乏型 (fully depleted) CCD 計 870M pixels
 - フォーカス用などを除き 104個の CCD が円形に並ぶ
- 結像精度 0.2 秒角 (FWHM; i-band)
 - 実際の画像の PSF は 0.7秒角程度 (FWHM; i-band)



峯尾さんのADC-UMスライドを改変

HSC-SSPの観測領域

- HSC-W (Wide): 1,087 deg² (うち全バンドが目標撮像時間に達した部分 858 deg²)
- HSC-D/UD (Deep/Ultra-) 領域: 27 deg²



HSC-SSPのデータリリース

- 現在次のサービスを提供中
 - 較正済みFITSファイルそのものの取得
 - SQL による天体検索
 - 画像ビューア (hscMap)
 - 座標を指定して画像切り出し
 - その他

校正済みFITSファイルそのものの取得

DAS Search Form (PDR3)

Home | Usage

Region, Date, &c.

field map ^

Region...
Date →
Filter →
Tract →

New block

File types

☒ all

☒ frames

☒ coadds

Reruns

☒ Release PDR3

☒ U/DEEP (pdr3_dud_rev)

☒ U/DEEP (pdr3_dud [old])

☒ WIDE (pdr3_wide)

Search

Fill in the above form and click "search".

SQL による天体検索

hsc-release.mtk.nao.ac.jp/datasearch/catalog_jobs

usersrolestatic configjson configall jobssample queries

name: catalog-job 2025-04-15

release: PDR3 Citus columnar (fast / experimental)

Guess release from your SQL

SQL:

Press ctrl-enter to preview

```
1 SELECT
2   object_id
3   , ra
4   , dec
5   , i_kronflux_mag
6   , i_kronflux_magerr
7   , y_kronflux_mag
8   , y_kronflux_magerr
9   , i_kronflux_mag - y_kronflux_mag AS i_y
10  FROM
11  pdr3_dud_rev.forced
12  JOIN pdr3_dud_rev.forced2 USING (object_id)
13  WHERE
14  isprimary
15  AND boxSearch(coord, 34.0, 36.0, -5.0, -4.5)
16      /* is equivalent to
17         ra BETWEEN 34.0 AND 36.0
18         AND dec BETWEEN -5.0 AND -4.5
19         but boxSearch() is much faster
20      */
21  AND i_kronflux_flux > mag_to_flux(25.5)
22      /* is almost (*note) equivalent to
23         i_kronflux_mag < 25.5
24         but 'flux > mag_to_flux(constant)' is slightly efficient.
25         (*note: "almost" equivalent except for NaN cases)
26      */
27  LIMIT 10
28 ;
29
```

example:

function list

table list

☐ include SQL in CSV

☒ syntax check before enqueueing

☐ csv ☒ csv.gz ☐ sqlite3 ☐ fits

estimate query time

show the estimated query time and the query plan. [What is this?](#) (experimental)

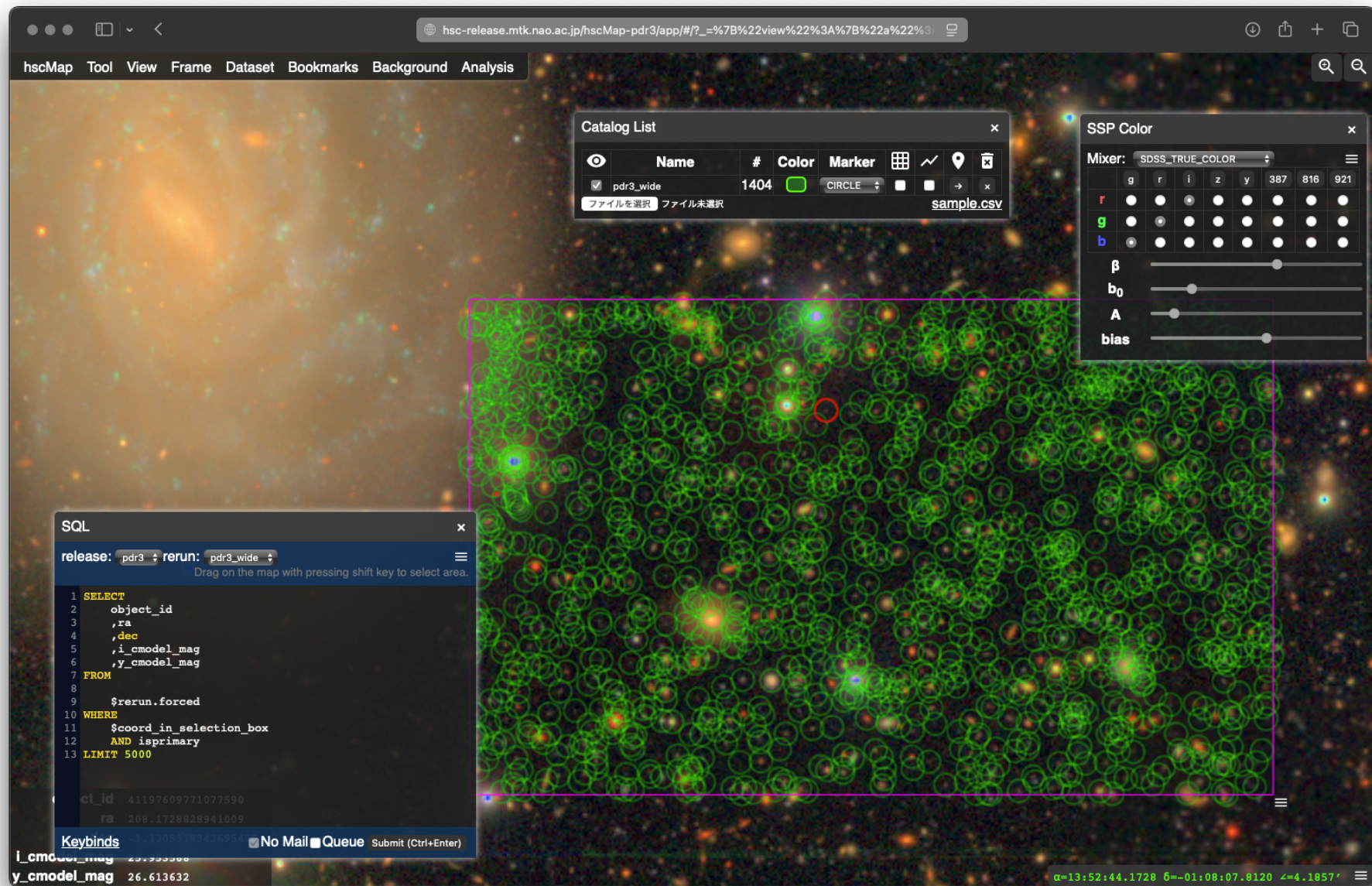
preview (timeout: 10 seconds)

enqueue (timeout: 6 hours / output limit: 10 GB)

10 records found.

object_id	ra	dec	i_kronflux_mag	i_kronflux_magerr	y_kronflux_mag	y_kronflux_magerr	i_y
37484571889013395	34.92183090781416	-4.9414931394618264	24.355623	0.031028163	23.698019	0.06213088	0.6576042
37484576183945387	34.90700964484495	-4.886137160417928	25.289902	0.10369678	24.81827	0.24527943	0.471632
37484571889010235	34.88547161757504	-4.964865895445099	25.062923	0.044685394	24.506098	0.0989095	0.55682564
37484576183968543	34.88819339101568	-4.866844742911253	21.383741	0.002041782	21.349575	0.007306132	0.034166336
37484571889011979	34.92902775635888	-4.952328160490618	23.573505	0.014584216	23.356321	0.044249568	0.21718407
37484571889012854	34.90957704485958	-4.942423462174566	24.83781	0.05341343	24.334763	0.12568435	0.50304794
37484576183972156	34.92752962129433	-4.837081977240693	24.897156	0.042264488	25.161787	0.2015411	-0.26463127
37484580478914210	34.87791448303711	-4.6656849459604555	25.464725	0.13677698	25.054802	0.35004207	0.40992355
37484576183979272	34.892485264253644	-4.775055766641277	24.761703	0.034669276	24.983252	0.15789713	-0.22154808
37484576183980570	34.948557795143245	-4.763768705737976	25.20996	0.043162633	24.82912	0.11510422	0.3808403

画像ビューワー(hscMap)



座標を指定して画像切り出し

hsc-release.mtk.nao.ac.jp/das_cutout/pdr3/

DAS Cutout (PDR1-PDR3)

From here you can take a FITS image cut out from large coadd images stored in DAS. To get point spread functions, use [PSF picker \(PDR1\)](#), [PSF picker \(PDR2\)](#), or [PSF picker \(PDR3\)](#).

There are two ways to specify a rectangular region you want:

- [Specify two diagonal corners](#)
- [Specify center & size](#)

Or you can upload a list of coordinates to get them all:

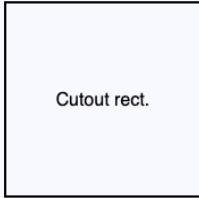
- [Upload a list of coordinates](#)

Read [the manual](#) for usage in detail.

Two diagonal corners

Cut out a rectangle that is defined by two diagonal corners: (ra_1, dec_1) and (ra_2, dec_2) .

(ra_1, dec_1)



(ra_2, dec_2)

ra₁:

dec₁:

ra₂:

dec₂:

file type: coadd

image☒

mask☐

variance☐

(Unlayered☐)

filter: i

Permit no-data area☐

tract: (can be blank)

rerun: Select...

(Find tract/rerun)

Cut out

[help?](#)

HSC-SSPのデータリリース

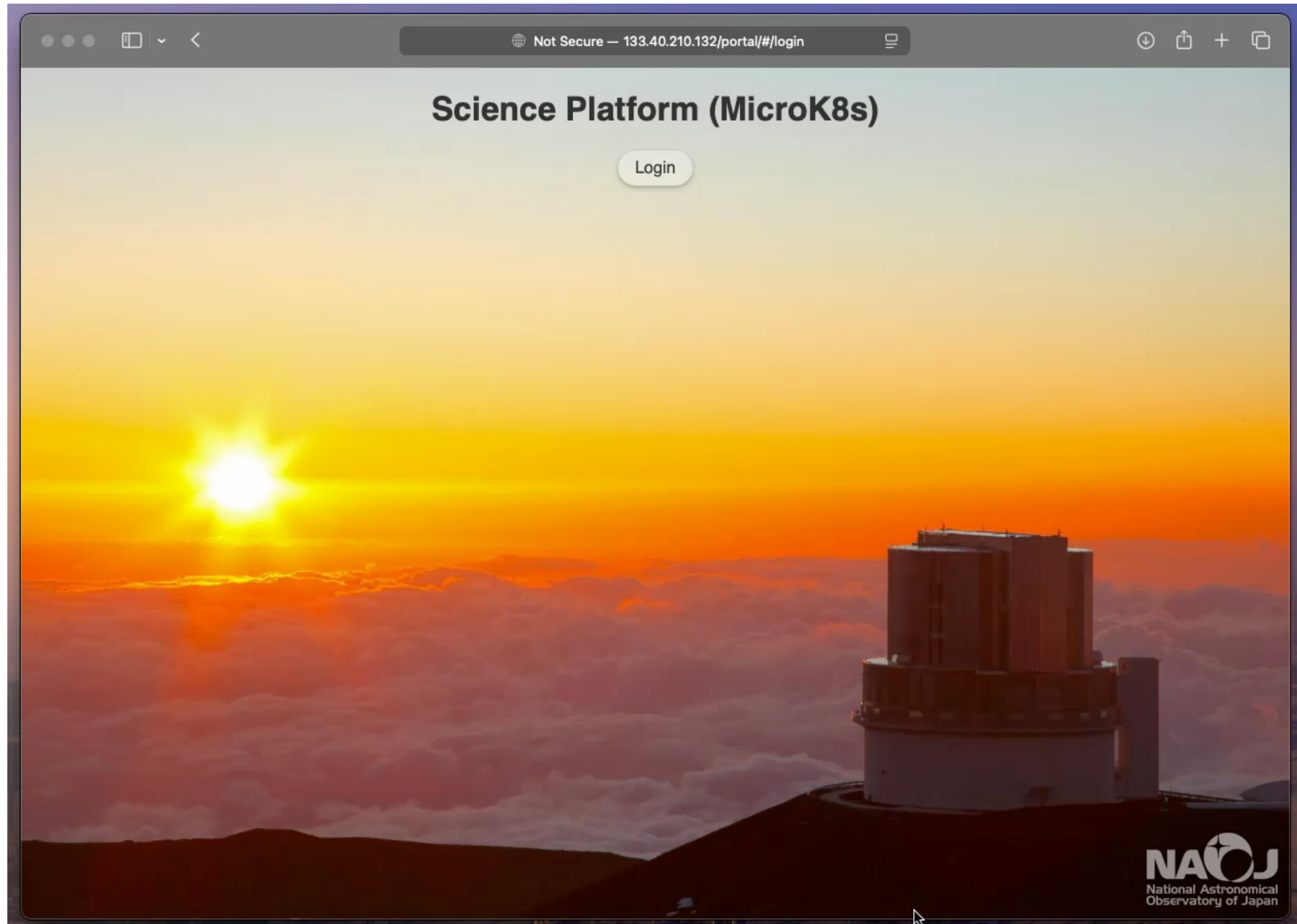
- 現在次のサービスを提供中
 - 較正済みFITSファイルそのものの取得
 - SQL による天体検索
 - 画像ビューア (hscMap)
 - 座標を指定して画像切り出し
 - その他

Science Platformとは

- 先に紹介した通りHSC-SSPのデータリリースサービスではデータにアクセスするためのいくつかのインターフェースがある
- 現在の一般的なワークフローは↑のインターフェースからユーザーが各々の環境にデータをダウンロードして解析するというもの
- この解析するための環境をデータのインターフェースと統合した形で提供しようというのがSciencePlatform
- CPU/ストレージ・解析ソフトウェア・その他データアクセス補助ソフト類

Science Platformの機能

- JupyterLab
- Nextcloud
- X11アプリケーション
- デスクトップ環境
- QuickDB
- SSHアクセス
- VSCode連携
- SSHFS連携
- 任意コンテナの実行
- Postgres, code-server, ...



JupyterLab

- WebブラウザからアクセスできるPython（など）の実行環境
- WebブラウザからアクセスできるShell
- SSPのデータは環境にマウントされている
- SSPのデータアクセスのためのmoduleがインストール済み
- hscPipeがインストール済み

Science Platform (MicroK8s)

Schema Browser

Test User

My Pods

+

Storage

102.40 MiB / 1.00 GiB

Refresh

Set Quota

File Access

Nextcloud

Nextcloud is a web-based file browser that allows you to manage files through a web interface. It offers a [PC client app](#) for syncing storage with your local disk. Additionally, it allows you to share specific directories on the internet.



You can directly access released data here. This includes data released from the HSC-SSP such as PDR1, PDR2, PDR3, and so on.

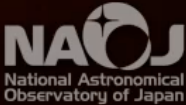
Stars Account

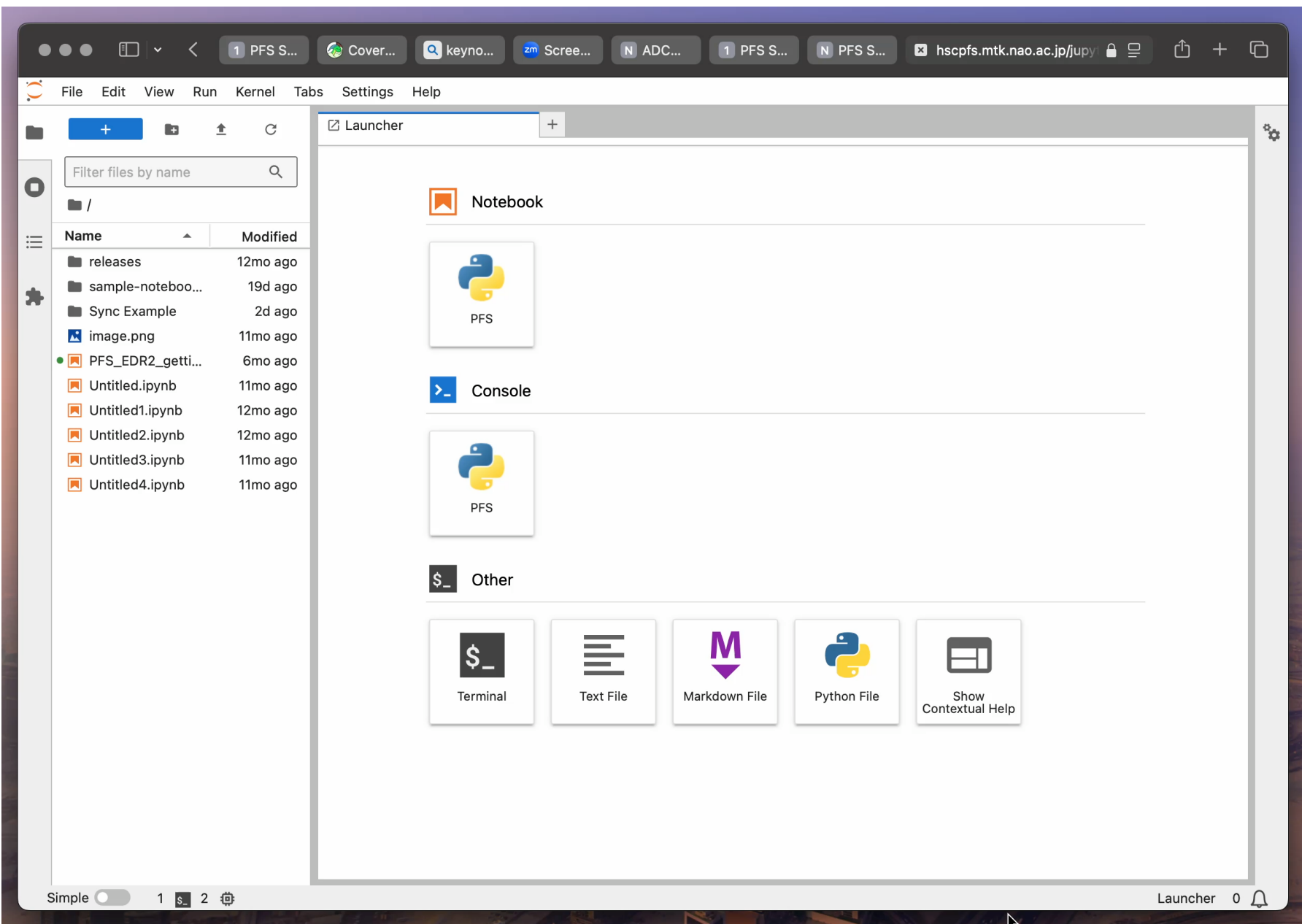
Enter you Stars account credentials to access your Stars data.

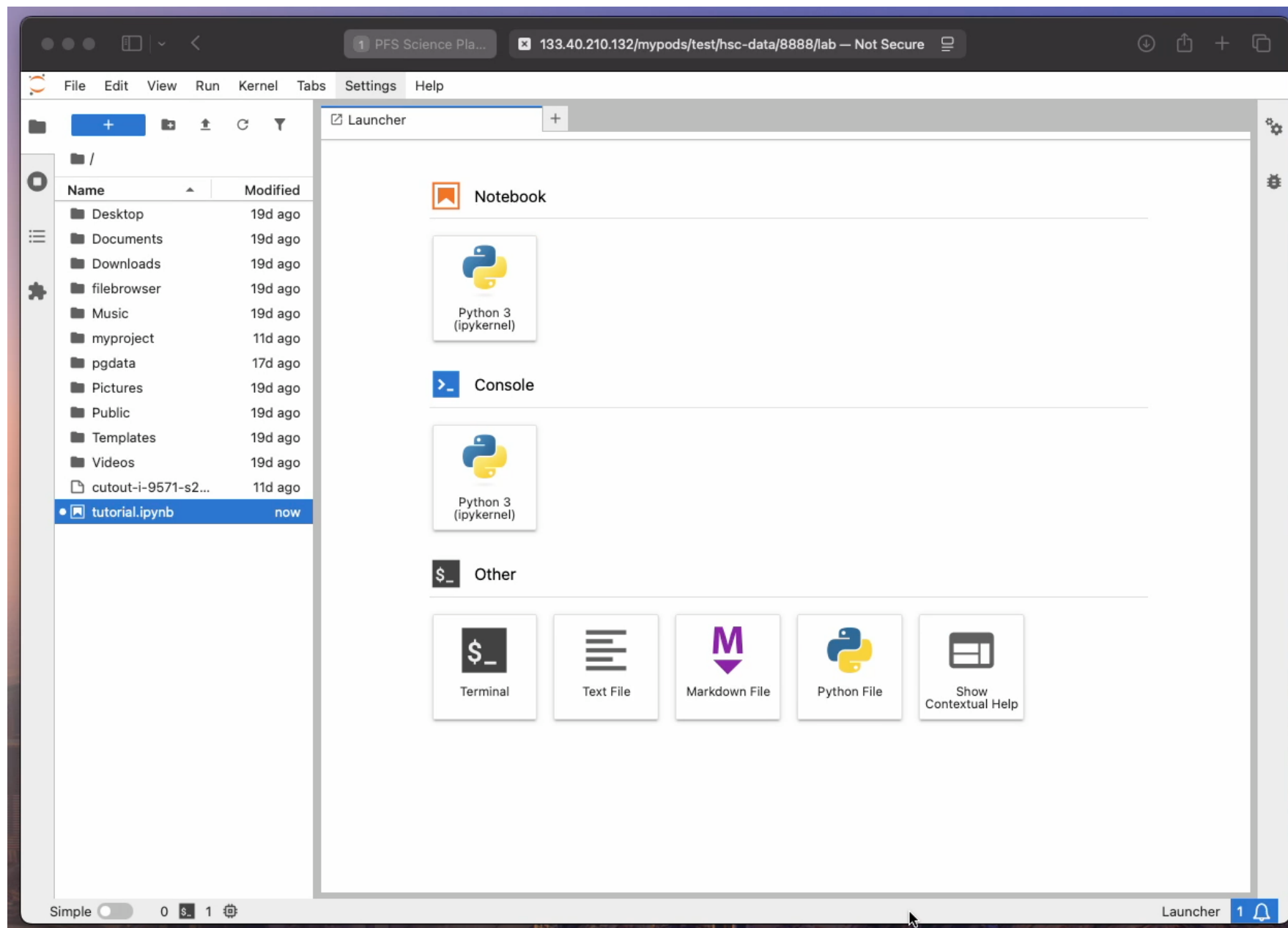
Username

Password

Connect

NAOJ
National Astronomical
Observatory of Japan





File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ sample-notebooks /

Name	Modified
hscdata.imagec...	11mo ago
hscdata.psf tuto...	11mo ago
hscmap tutorial.i...	20d ago
image.png	11mo ago
Untitled.ipynb	11mo ago

Terminal 2

hscdata.imagecutout tutori

Markdown

```
[ ]: from hscdata import imagecutout
```

Basic Usage

```
[ ]: hdul = imagecutout.getImage(216.20416023106773, -0.8893594325061295, sw='1arcmin', sh='1arcmin', filter='HSC-I')

[ ]: %matplotlib inline

from matplotlib import pyplot
import numpy

a = 10
pyplot.imshow(numpy.arcsinh(a * hdul[1].data))
```

Work with lists

```
[ ]: RA, DEC = numpy.array(list(map(float, '''
218.11728129591035 0.28466405829224695
217.6801206215572 0.2517119817197224
217.39367193096362 -0.03160953683512477
218.36453355794325 -0.11943676355077588
'''.split()))).reshape((-1, 2)).T

[ ]: for hdul in imagecutout.getImageBulk(RA, DEC, sw='1arcmin', sh='1arcmin', filter='HSC-I'):
    pyplot.imshow(numpy.arcsinh(a * hdul[1].data))
    pyplot.show()

[ ]:
```

Nextcloud

- Webベースファイルブラウザ
- Jupyterのユーザーストレージと連動
- ディレクトリ単位で他のユーザーとデータを共有したり、とあるディレクトリの内容を公開できる
- Mac, Window, Linux用のクライアントアプリがあり、これを使うとDropbox的な使い方ができる

Science Platform (MicroK8s)

Schema Browser

Test User

My Pods

✓

HSC Data

JupyterLab

Bash

http

+

Storage

102.40 MiB / 1.00 GiB

Refresh

Set Quota

File Access

Nextcloud

Nextcloud is a web-based file browser that allows you to manage files through a web interface. It offers a [PC client app](#) for syncing storage with your local disk. Additionally, it allows you to share specific directories on the internet.

You can directly access released data here. This includes data released from the HSC-SSP such as PDR1, PDR2, PDR3, and so on.

Stars Account

Enter you Stars account credentials to access your Stars data.

Username

Password

Connect

NAOJ

National Astronomical
Observatory of Japan



- All files
- Recent
- Favorites
- Shares
- Shared to Circles

Sync Example + New



No files in here

Upload some content or sync with your devices!

29.6 MB used

Files settings



Nextcloud



IMG_4393.jpeg



Stacked_107_M 42_10.....035.jpg

FileEditViewGoWindowHelp

1 PFS Science Pla...N PFS Science Pla...hscpfs.mtk.nao.ac.jp/nextcloud/apps/files/?dir=

o o o

📁

👤

🔍

🖼️

MK

📁 All files

🕒 Recent

★ Favorites

🔗 Shares

👤 Shared to Circles

🏠 Sync Example 🔗

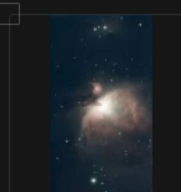
+ New

📁



IMG_4393 ⋮

📁



Stacked_107... ⋮

2 files2.1 MB

📊 29.6 MB used

⚙️ Files settings

📁

Nextcloud



IMG_4393.jpeg



Stacked_107_M 42_10.....035.jpg

X11アプリケーション

- ds9, gnuplot, emacs, ...などのx11アプリケーションにブラウザ経由でアクセス可能
- (多分) ``ssh -X`` するよりスムーズに動く

133.40.210.132/portal/ — Not Secure

PFS Science Pla...

Test User

Science Platform (MicroK8s)Schema Browser

My Pods

✓

HSC Data

JupyterLab

Bash

htop

+

Storage

102.40 MiB / 1.00 GiB

Refresh

Set Quota

File Access

Nextcloud

Nextcloud is a web-based file browser that allows you to manage files through a web interface. It offers a [PC client app](#) for syncing storage with your local disk. Additionally, it allows you to share specific directories on the internet.

You can directly access released data here. This includes data released from the HSC-SSP such as PDR1, PDR2, PDR3, and so on.

Stars Account

Enter your Stars account credentials to access your Stars data.

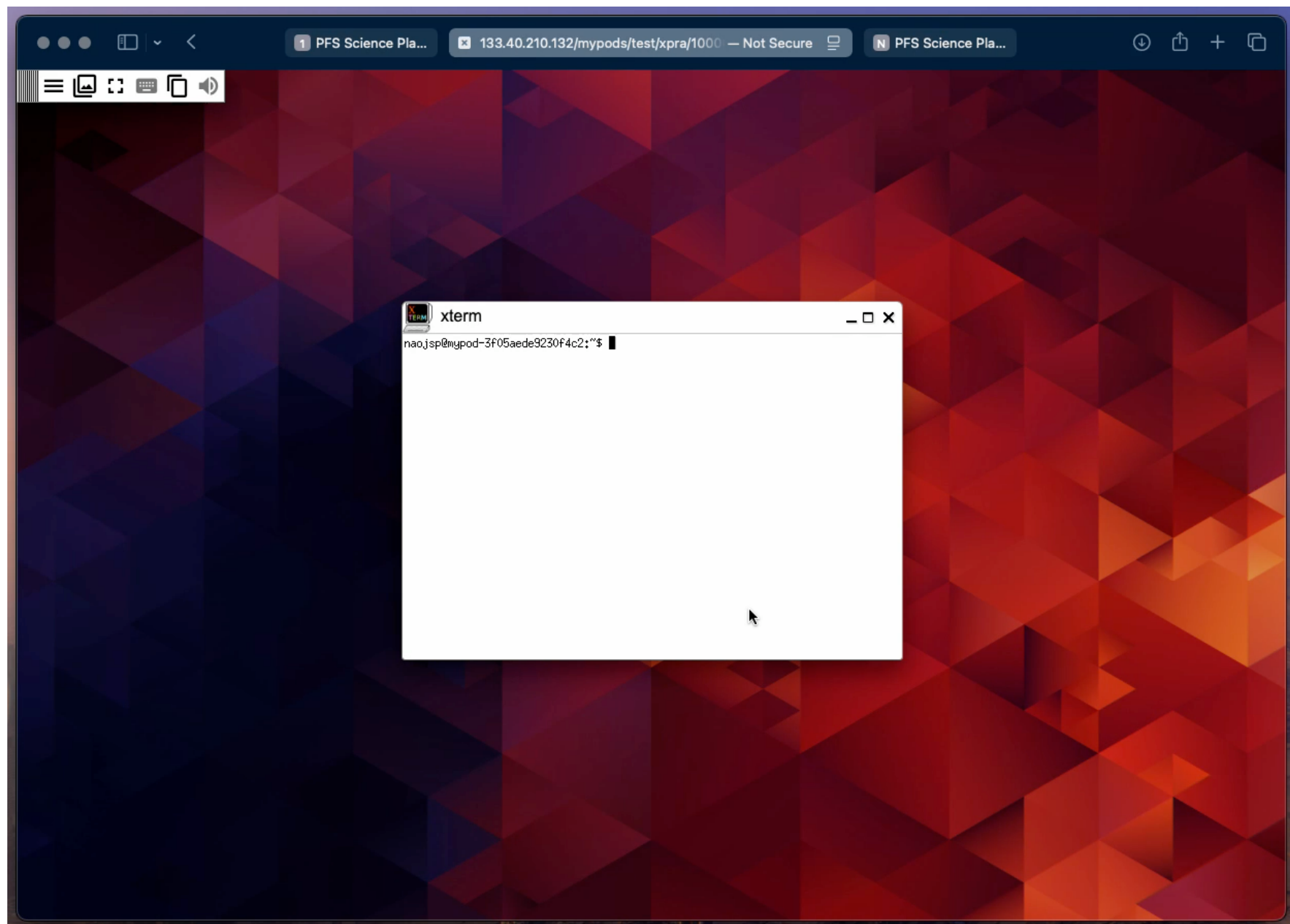
Username

Password

Connect

NAOJ

National Astronomical
Observatory of Japan



133.40.210.132/portal/ — Not Secure

mypodtestxpra...PFS Science Pla...

Test User

Science Platform (MicroK8s)Schema Browser

My Pods

✓ HSC Data

JupyterLab

Bash

htop

✓ Xpra

HTML5 client for Xpra

Storage

102.40 MiB / 1.00 GiB

Refresh

Set Quota

File Access

Nextcloud

Nextcloud is a web-based file browser that allows you to manage files through a web interface. It offers a [PC client app](#) for syncing storage with your local disk. Additionally, it allows you to share specific directories on the internet.

You can directly access released data here. This includes data released from the HSC-SSP such as PDR1, PDR2, PDR3, and so on.

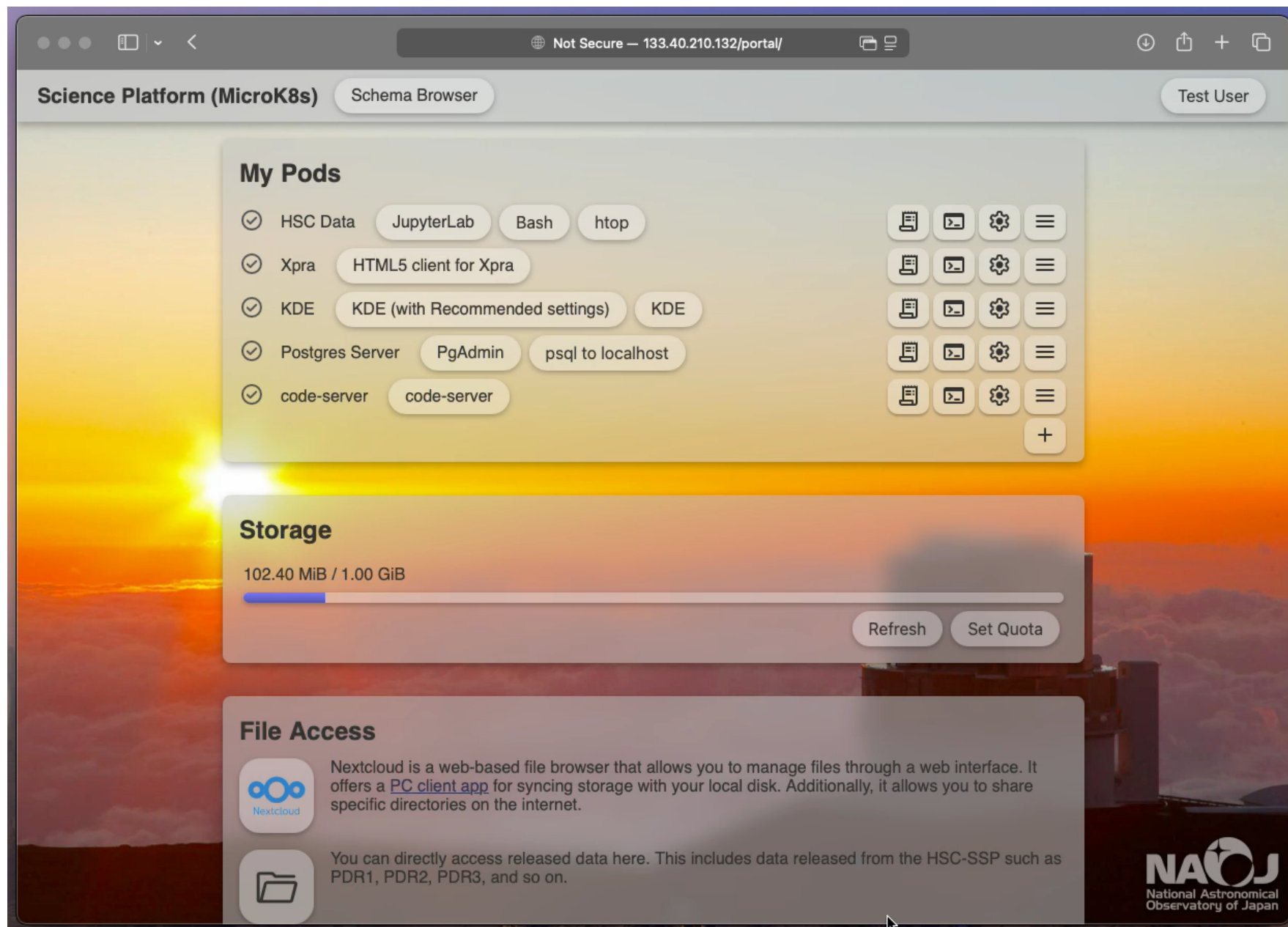
Stars Account

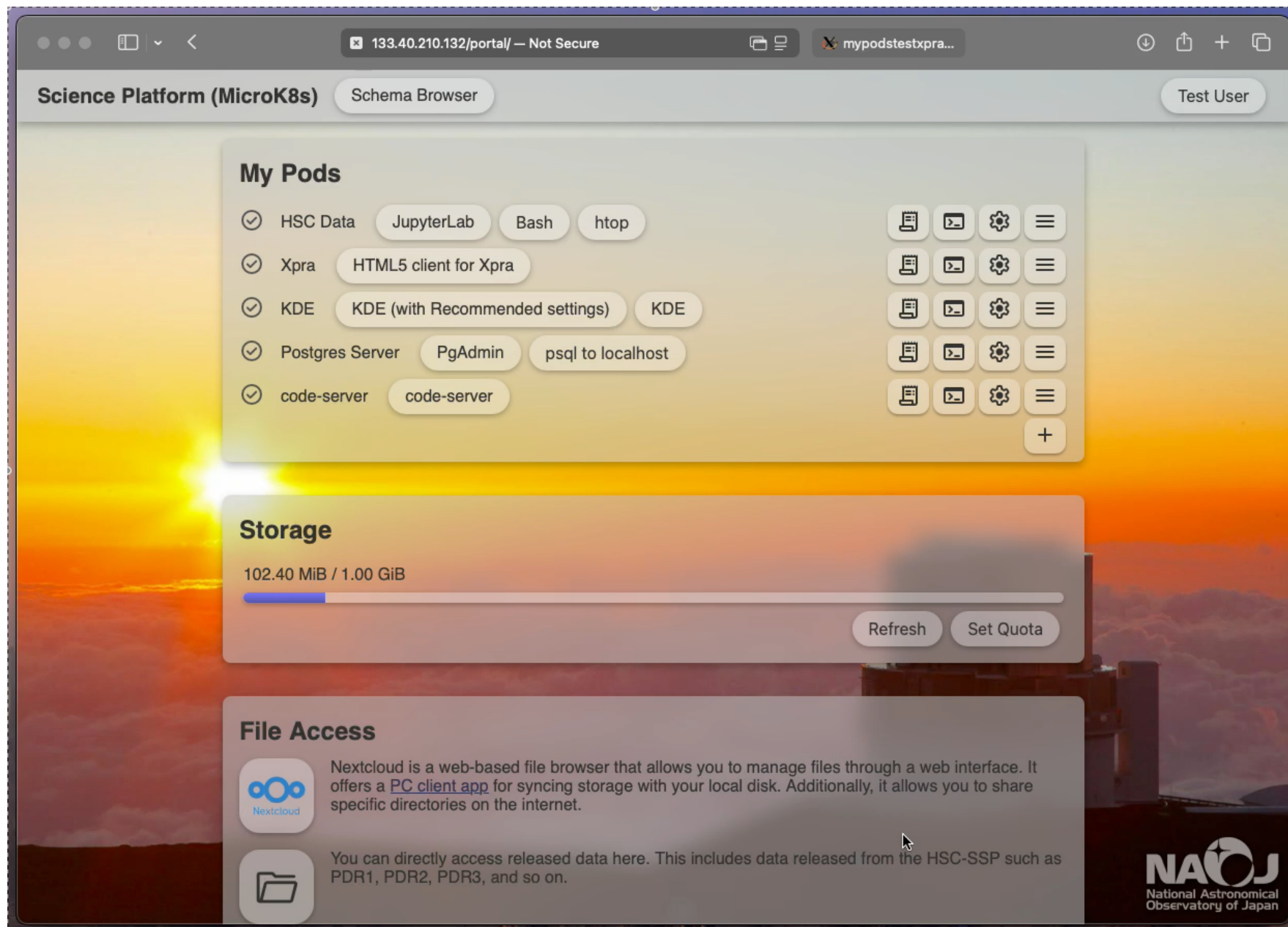
Enter your Stars account credentials to access your Stars data

NAOJ

National Astronomical Observatory of Japan

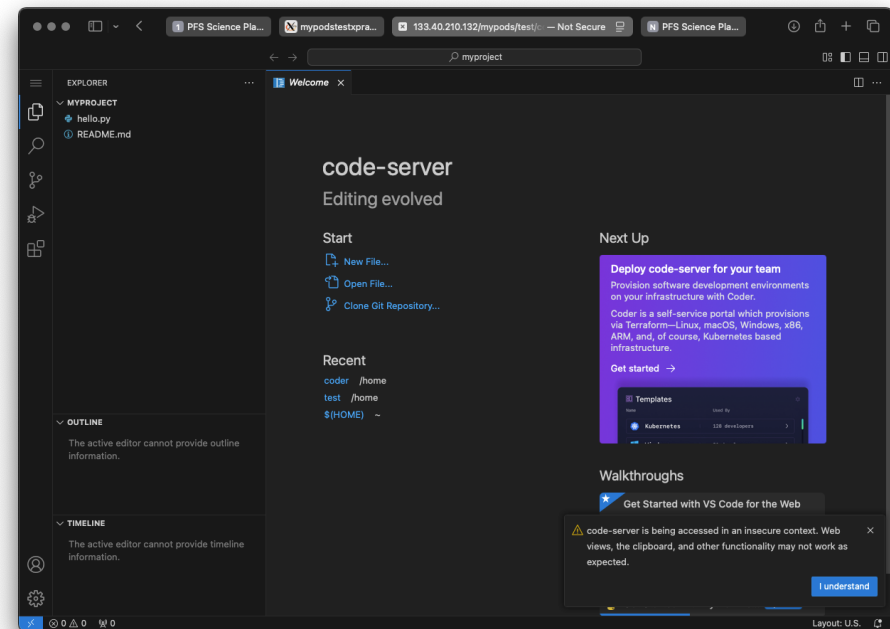
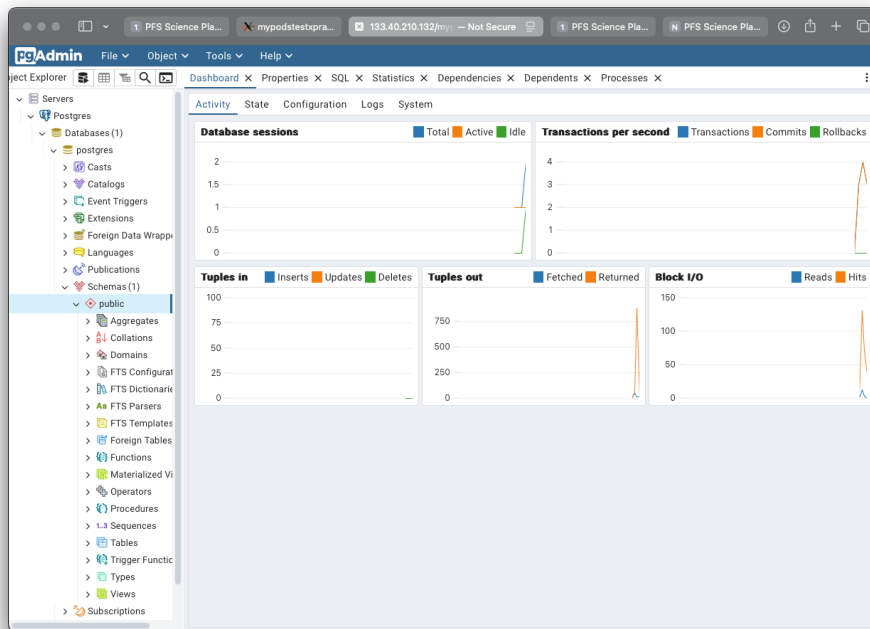
実行中の個々のPodに対するSSHアクセス





任意イメージを実行可能

- Postgres, pgAdmin, code-server(VSCodeのWeb版)などのプリセットをとりあえずサンプルとして登録してある
- ユースケースはこれから検討



QuickDB

- Jupyterから使われることを想定した、SSPの天体カタログに対して一般ユーザーが任意のPythonコードを多数のCPUで並列に実行する仕組み
- （ユーザーはMap Reduce関数をシステムに与える）
- 「似た色で近くにあるペアを探す」のようなSQLで表現するのが難しい問題に有用
- SQLサーバーからデータをダウンロードし手元で解析する手法はデータが移動するがQuickDBではデータではなくコードが移動する

QuickDB/Color-Color Diagram

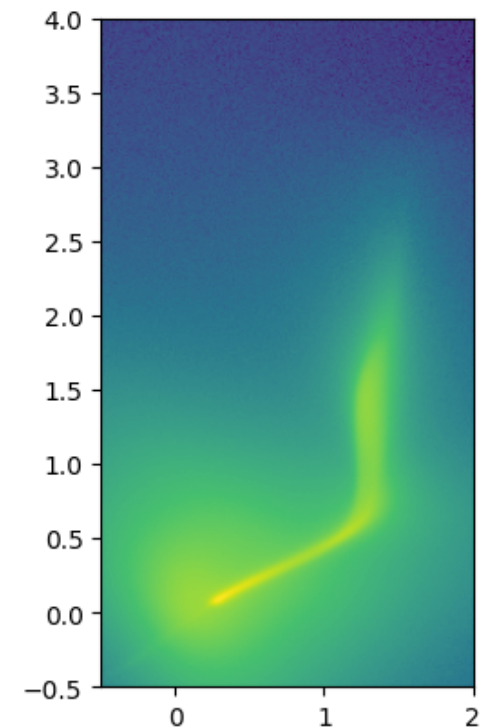
```
import numpy

def map(patch):
    where = patch('forced.i.extendedness_value') < 0.5
    patch = patch[where]
    r = patch('forced.r.psfflux_instmag')
    g = patch('forced.g.psfflux_instmag')
    i = patch('forced.i.psfflux_instmag')
    return numpy.histogram2d(
        g - r,
        r - i,
        range=((-0.5, 2), (-0.5, 4)),
        bins=(200, 400)
    )

def reduce(a, b):
    a_hist, a_xbins, a_ybins = a
    b_hist, b_xbins, b_ybins = b
    # a_xbins, b_xbinsは同じ
    # a_ybins, b_ybinsは同じ
    return a_hist + b_hist, a_xbins, a_ybins

%time res = api.mapreduce(map, reduce)

hist, xedges, yedges = res
pypplot.imshow(numpy.log(1 + hist.T), origin='lower', extent=(xedges[0], xedges[-1], yedges[0], yedges[-1]))
```



PDR3 Wideの全天体から星様の天体を選んでColor-Color Diagramを作る

Color-Color Diagram of Star-like Objects(, Not Galaxies)

map , reduce Interface

`numpy.histogram2d` returns the histogram, the edges of the bins for x, and the edges of the bins for y.

```
[ ]: import numpy

def map(patch):
    where = patch('forced.i.extendedness_value') < 0.5
    patch = patch[where]
    r = patch('forced.r.psfflux_instmag')
    g = patch('forced.g.psfflux_instmag')
    i = patch('forced.i.psfflux_instmag')
    return numpy.histogram2d(
        g - r,
        r - i,
        range=((-0.5, 2), (-0.5, 4)),
        bins=(200, 400)
    )

def reduce(a, b):
    a_hist, a_xbins, a_ybins = a
    b_hist, b_xbins, b_ybins = b
    # a_xbins, b_xbinsは同じ
    # a_ybins, b_ybinsは同じ
    return a_hist + b_hist, a_xbins, a_ybins

%time res = api.mapreduce(map, reduce)

hist, xedges, yedges = res
pyplot.imshow(numpy.log(1 + hist.T), origin='lower', extent=(xedges[0], xedges[-1], yedges[0], yedges[-1]))
```

Search for pairs of similar colors

This process is practically difficult to perform using SQL.

In such cases, the ability for users to directly execute Python code on the database offers a significant advantage.

```
[ ]: import numpy
import sys
import os
```

QuickDB/近くの似た色の天体のペアを探す

```
import numpy
import scipy.spatial

from functools import reduce
tol_mag = 0.25
match_rad = 2 # arcsec

columns = '''
sdsscentroid_flag
pixelflags_edge
pixelflags_interpolatedcenter
pixelflags_saturatedcenter
pixelflags_crcenter
pixelflags_bad
cmodel_flag
'''
columns = columns.split()

def gen_pairs(xyz, r: float=1):
    tree = scipy.spatial.cKDTree(xyz.T)
    for i, pairs in enumerate(tree.query_ball_tree(tree, numpy.deg2rad(r / 3600.))):
        for j in pairs:
            if i < j:
                yield i, j

def map(patch):
    clean = reduce(
        lambda a, b: a & b,
        (patch(f'forced.{f}.{c}') == 0 for f in 'gri' for c in columns),
        patch('forced.isprimary') & (patch('forced.i.psfflux_instmag') < 23),
    )
    patch = patch[clean]

    if patch.size == 0:
        return
    xyz = patch('forced.coord')
    pairs = numpy.array(list(gen_pairs(xyz, match_rad)))
    if len(pairs) == 0:
        return

    a_ind, b_ind = pairs.T

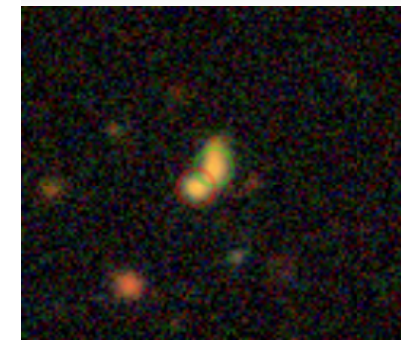
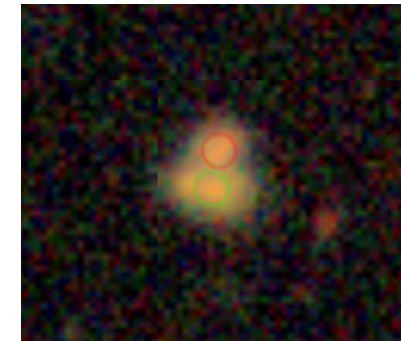
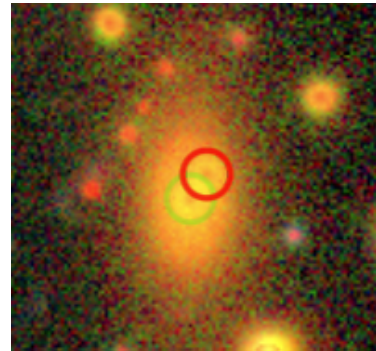
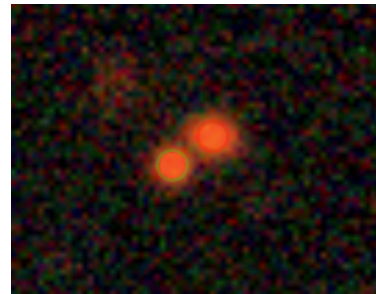
    patch_a = patch[a_ind]
    patch_b = patch[b_ind]

    g_r_a = patch_a('forced.g.psfflux_instmag') - patch_a('forced.r.psfflux_instmag')
    r_i_a = patch_a('forced.r.psfflux_instmag') - patch_a('forced.i.psfflux_instmag')
    g_r_b = patch_b('forced.g.psfflux_instmag') - patch_b('forced.r.psfflux_instmag')
    r_i_b = patch_b('forced.r.psfflux_instmag') - patch_b('forced.i.psfflux_instmag')

    ok = (numpy.abs(g_r_a - g_r_b) < tol_mag) & (numpy.abs(r_i_a - r_i_b) < tol_mag)
    ra_a = patch_a[ok]('forced.coord_ra')
    dec_a = patch_a[ok]('forced.coord_dec')
    ra_b = patch_b[ok]('forced.coord_ra')
    dec_b = patch_b[ok]('forced.coord_dec')

    return ra_a, dec_a, ra_b, dec_b

%time coords = numpy.hstack([numpy.array(coords) for coords in api.map(map, rerun='s20a_wide')])
```



Search for pairs of similar colors

This process is practically difficult to perform using SQL.

In such cases, the ability for users to directly execute Python code on the database offers a significant advantage.

```
[ ]: import numpy
import scipy.spatial

from functools import reduce
tol_mag = 0.25
match_rad = 2 # arcsec

columns = '''
sdsscentroid_flag
pixelflags_edge
pixelflags_interpolatedcenter
pixelflags_saturatedcenter
pixelflags_crcenter
pixelflags_bad
cmodel_flag
'''.split()

def gen_pairs(xyz, r: float=1):
    tree = scipy.spatial.cKDTree(xyz.T)
    for i, pairs in enumerate(tree.query_ball_tree(tree, numpy.deg2rad(r / 3600.))):
        for j in pairs:
            if i < j:
                yield i, j

def map(patch):
    clean = reduce(
        lambda a, b: a & b,
        (patch(f'forced.{f}.{c}') == 0 for f in 'gri' for c in columns),
        patch('forced.isprimary') & (patch('forced.i.psfflux_instmag') < 23),
    )
    patch = patch[clean]

    if patch.size == 0:
        return
    xyz = patch('forced.coord')
    pairs = numpy.array(list(gen_pairs(xyz, match_rad)))
    if len(pairs) == 0:
        return

    a_ind, b_ind = pairs.T
```

今後

- 2025年夏にHSC-SSP Collaborationへプレビュー
- 将来的にはPublic Releaseを目指している
- HSC-SSP以外のデータに対するアクセスも拡充予定
 - 将来のPFS、UNIONS、Rubinなど
- リリースしたら使ってみてください！

森嶋さんのインフラ・ミドルウェア編へつづく