ESCAPE WP5 IDA Update

26-10-2020

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

James Collinson

ESCAPE WP5 Progress Meeting

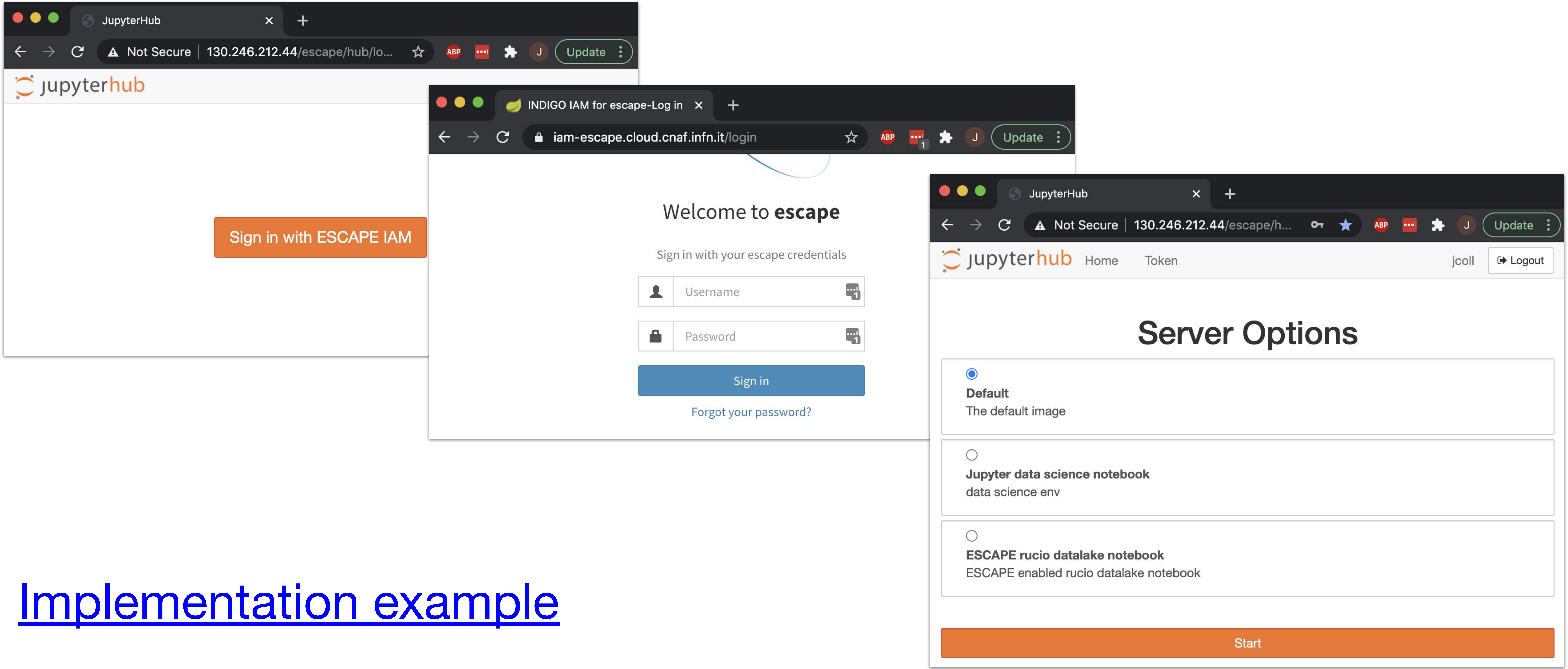
Overview

- Expanding the JupyterHub platform
  - Authentication
  - Developing SKA data science use case
  - Integrating Rucio
- Investigation: Query workflows from JupyterHub server

JupyterHub prototype: recap

- JupyterHub instance set up ‘from scratch’
  - Helm-based deployment running on Kubernetes cluster
  - Dynamic persistent storage using NFS
  - Hosted at STFC Cloud (RAL)
- Links:
  - Instance: http://130.246.212.44/escape
  - Scripts/docs: https://github.com/rohinijoshi06/jupyterhub-on-k8s

JupyterHub: Authentication



JupyterHub

Not Secure | 130.246.212.44/escape/hub/lo...

Update

jupyterhub

Sign in with ESCAPE IAM

INDIGO IAM for escape-Log in

iam-escape.cloud.cnaf.infn.it/login

Update

Welcome to **escape**

Sign in with your escape credentials

Username

Password

Sign in

Forgot your password?

JupyterHub

Not Secure | 130.246.212.44/escape/h...

Update

jupyterhub Home Token jcoll Logout

Server Options

● **Default**
The default image

○ **Jupyter data science notebook**
data science env

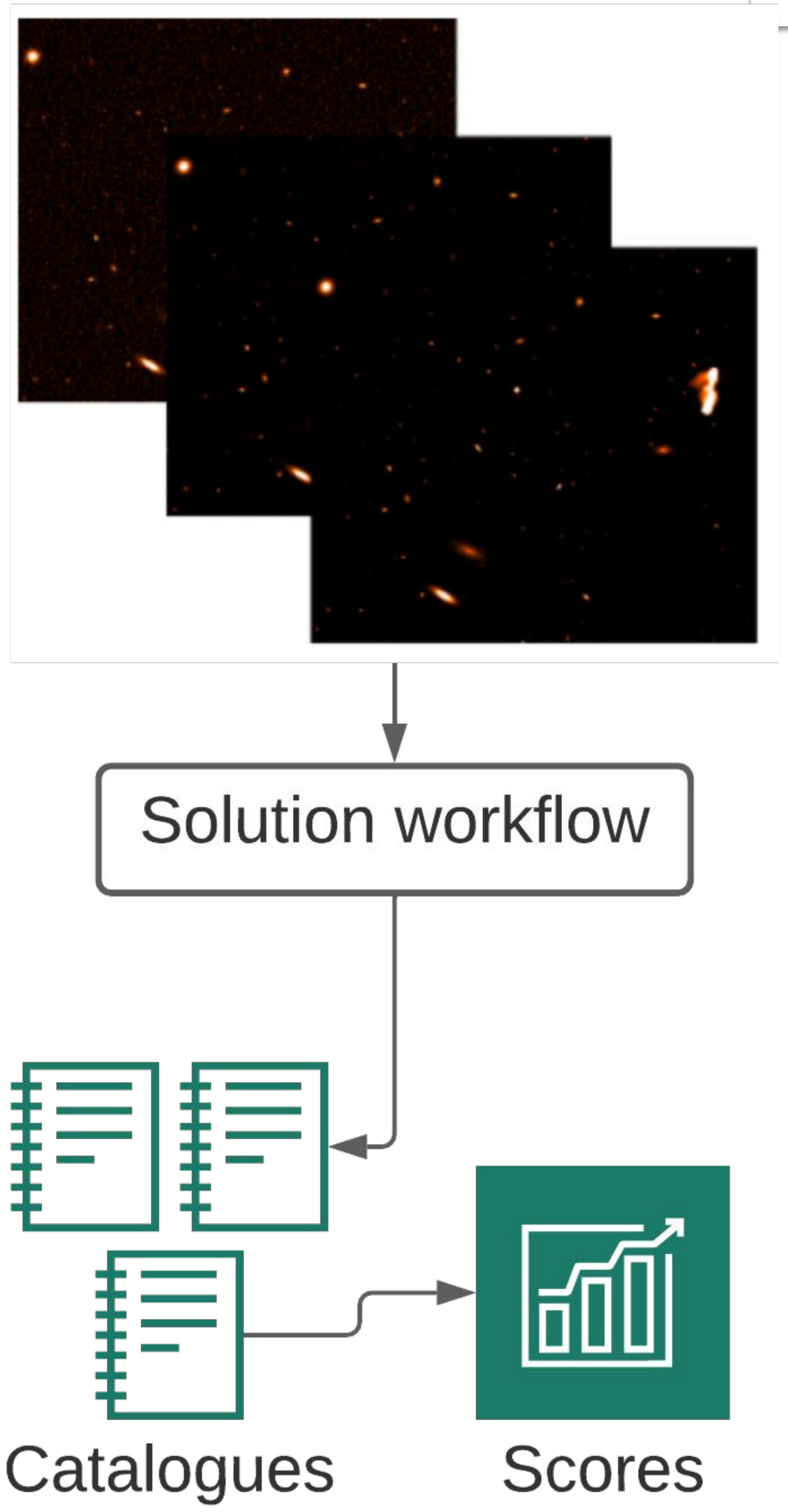
○ **ESCAPE rucio datalake notebook**
ESCAPE enabled rucio datalake notebook

Start

[Implementation example]

JupyterHub: SKA analysis environment

- Science Data Challenge 1
  - Build expertise in community
  - Simulated radio images in 3 bands (560, 1400, 9200 MHz)
  - Source finding and classification
- Workflows lacked reproducibility



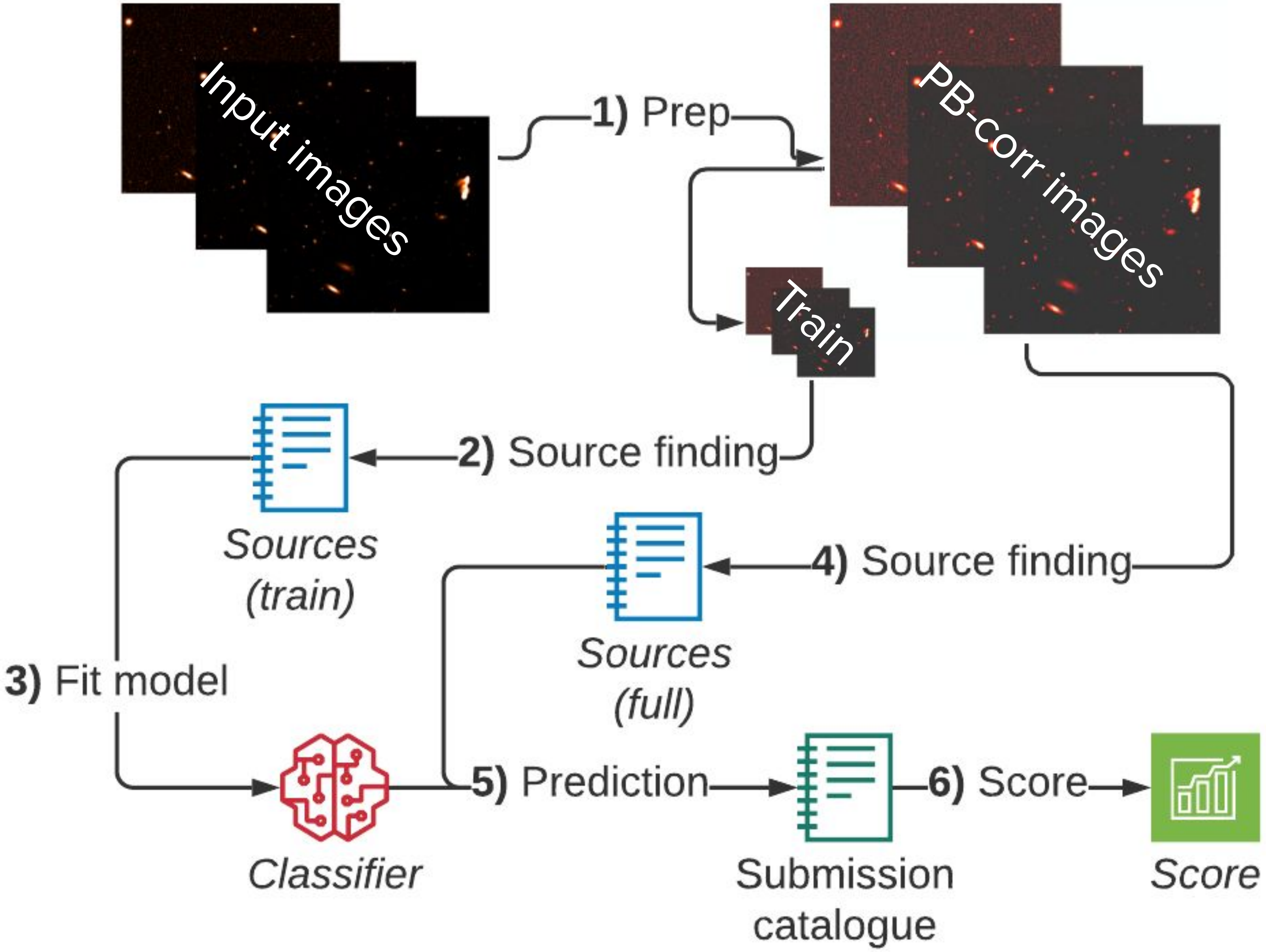
Solution workflow

Catalogues

Scores

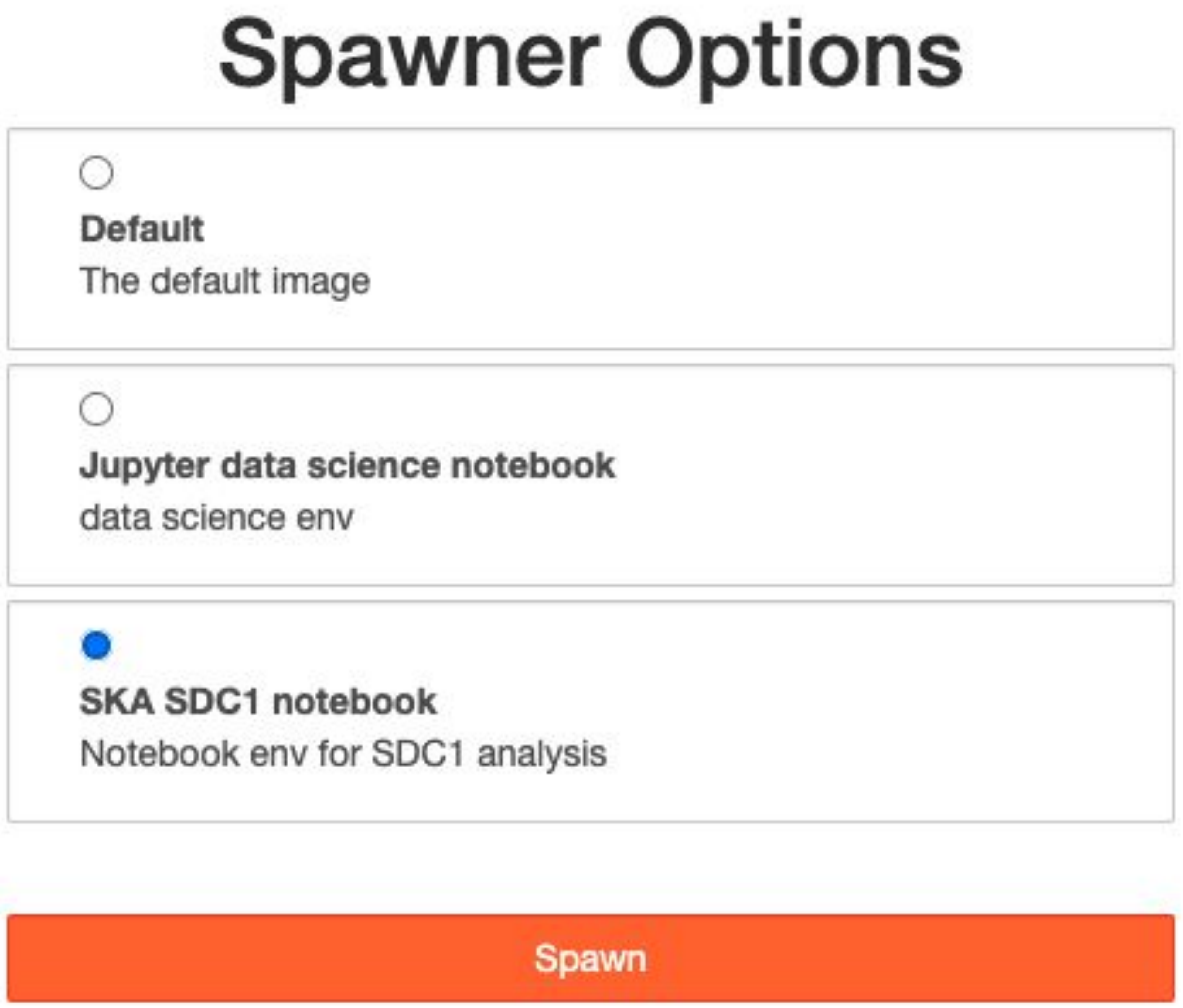
JupyterHub: SKA analysis environment

- Solution:
  - PyBDSF source finding
  - ML classification
- Output:
  - Environment with tools
  - Python wrappers with API
  - Solution script (adaptable)
  - GitLab



JupyterHub: SKA analysis environment

- SDC1 env on JupyterHub
- Built [Docker image]() on top of jupyter/base-notebook
  - Unit tests pass
- Next steps:
  - Scale up instance provision
  - Integration testing
  - Provide example notebook



Spawner Options

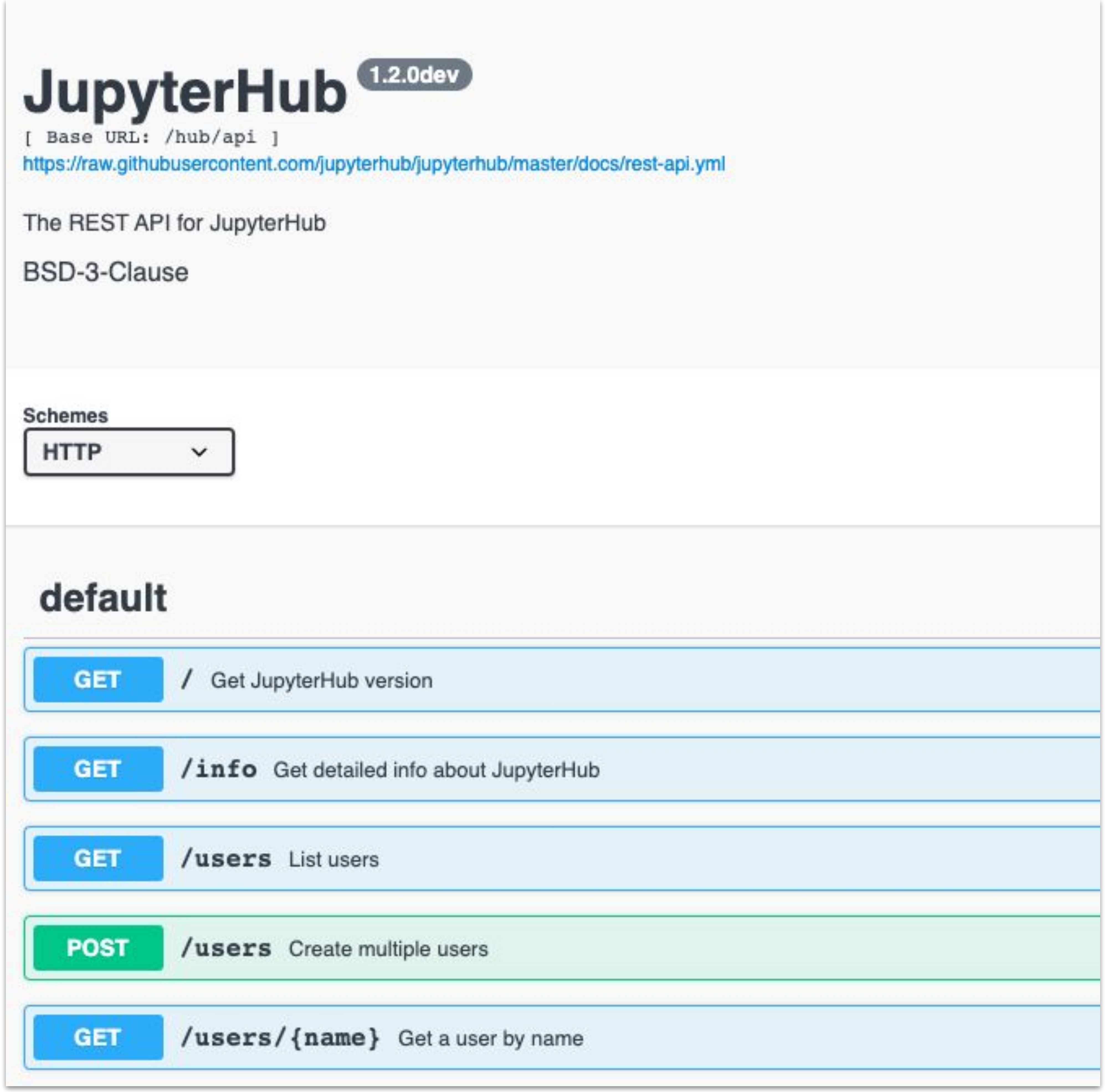
○ **Default**
The default image

○ **Jupyter data science notebook**
data science env

● **SKA SDC1 notebook**
Notebook env for SDC1 analysis

Spike: Querying JupyterHub workflows

- Spike for IDA WG
  - Query JupyterHub REST API for workflows
- More difficult than hoped
  - Expand functionality of API?
  - Consistent ESAP env?
  - Query via DockerHub?



JupyterHub 1.2.0dev

[Base URL: /hub/api]

https://raw.githubusercontent.com/jupyterhub/jupyterhub/master/docs/rest-api.yml

The REST API for JupyterHub

BSD-3-Clause

Schemes

HTTP

default

GET / Get JupyterHub version

GET /info Get detailed info about JupyterHub

GET /users List users

POST /users Create multiple users

GET /users/{name} Get a user by name

Next steps

- JupyterHub instance towards future science data challenge viability
    - PoC with SDC1 - starting point for SDC2 participants
    - Integrate workflows, staging of data, IVOA interfaces
- Overlap with work for WP2, WP3 and WP4
- Upcoming: roadmap planning for PI9 (Dec 20 - Feb 21)
    - We welcome suggestions as we develop our feature roadmap