



IVOA - science platforms

Moving code to the data

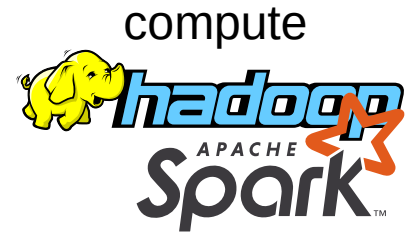
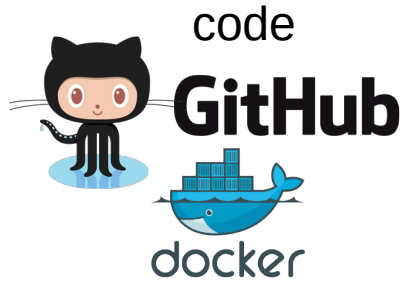
Dave Morris

IVOA interop
Tucson, Arizona
November 2023

The problem

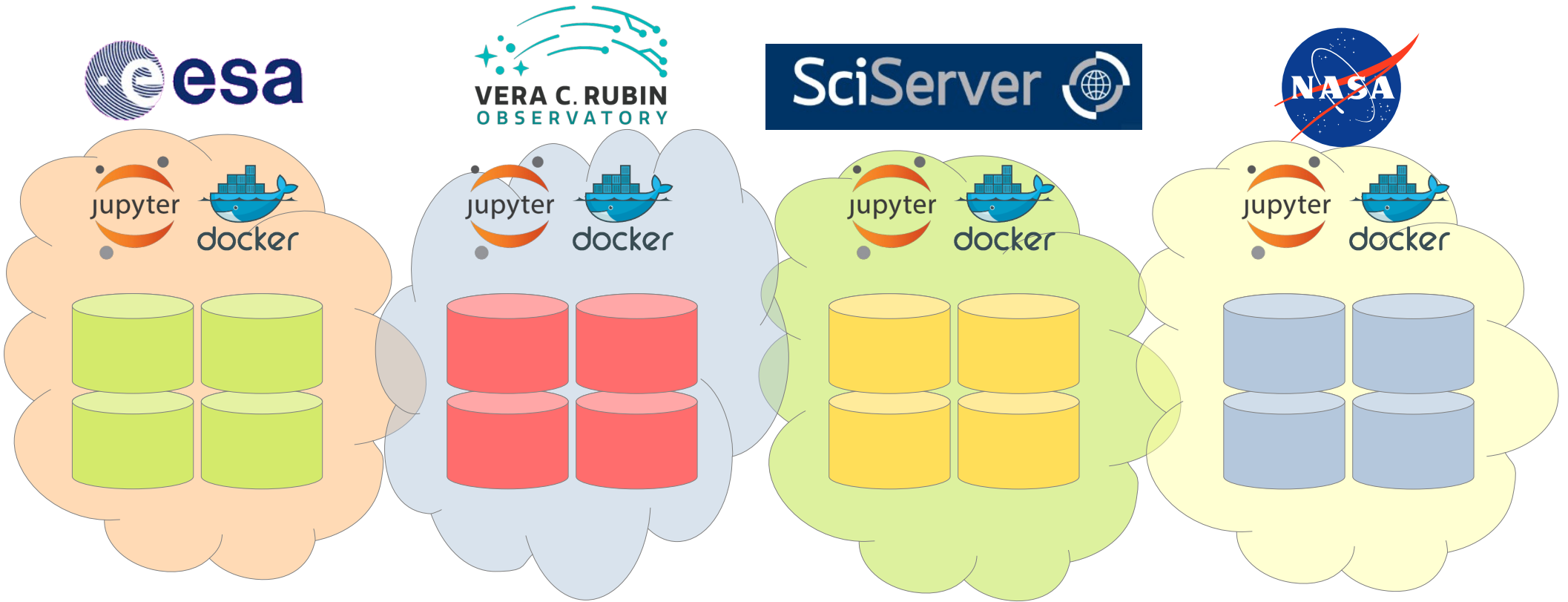
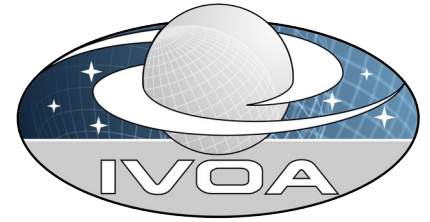
Heterogeneous code, data and compute.

Everyone is slightly different.



Current state

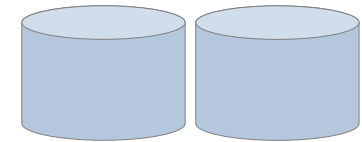
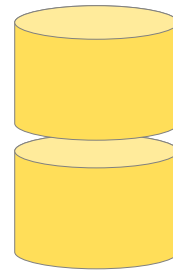
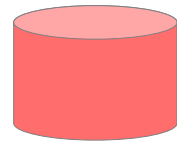
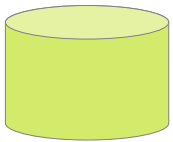
Cloud platforms, cloud datasets, cloud compute



IVOA interop
Tucson, Arizona
November 2023

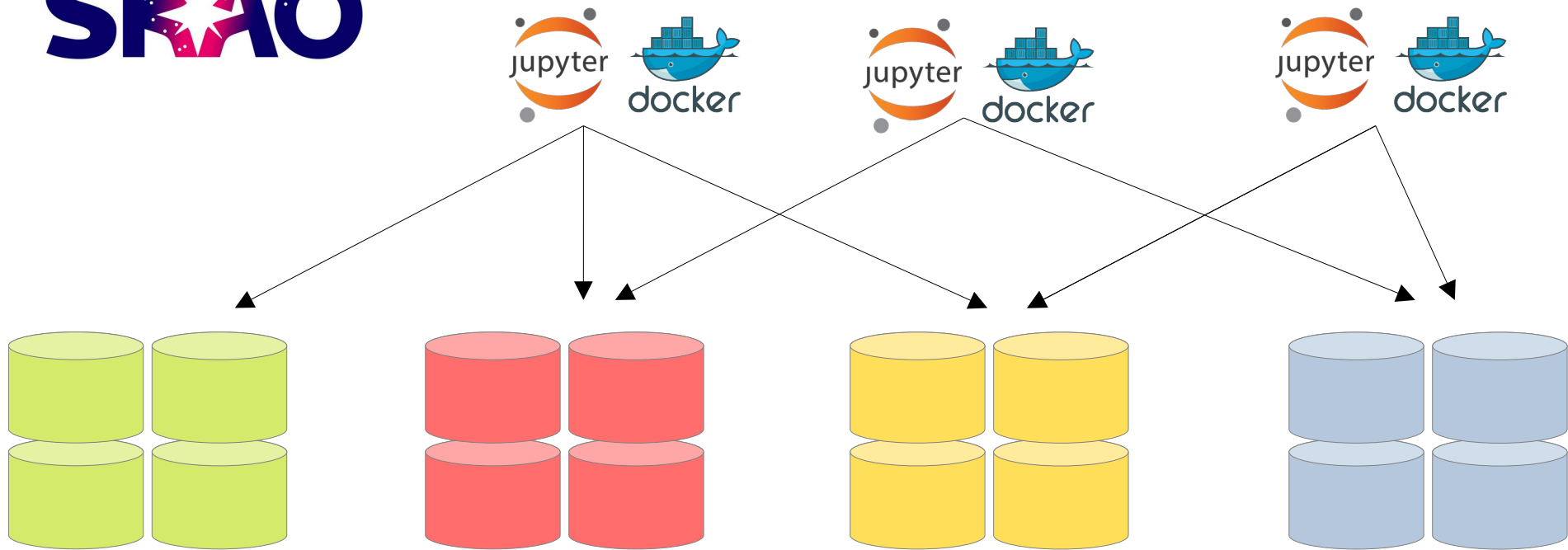
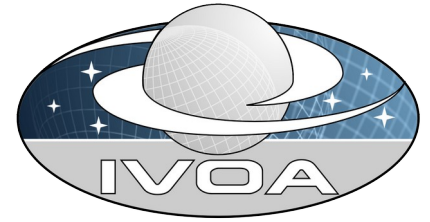
Current state

Smaller platforms, local datasets, local compute



Current state

Distributed data, distributed compute





Moving code to another platform

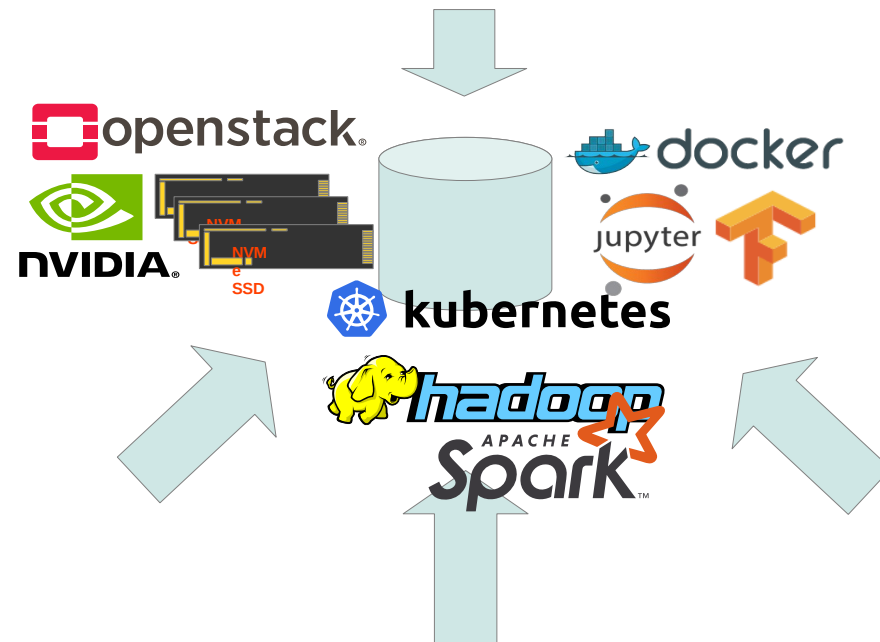
Asking the right question

Harvest all the metadata into a central registry.

That registry would need to understand the details of all the technologies.

Including authentication and authorization.

Find all the platforms that **<this> user** can run **<this>** task with **<this>** data.





IVOA ExecutionPlanner

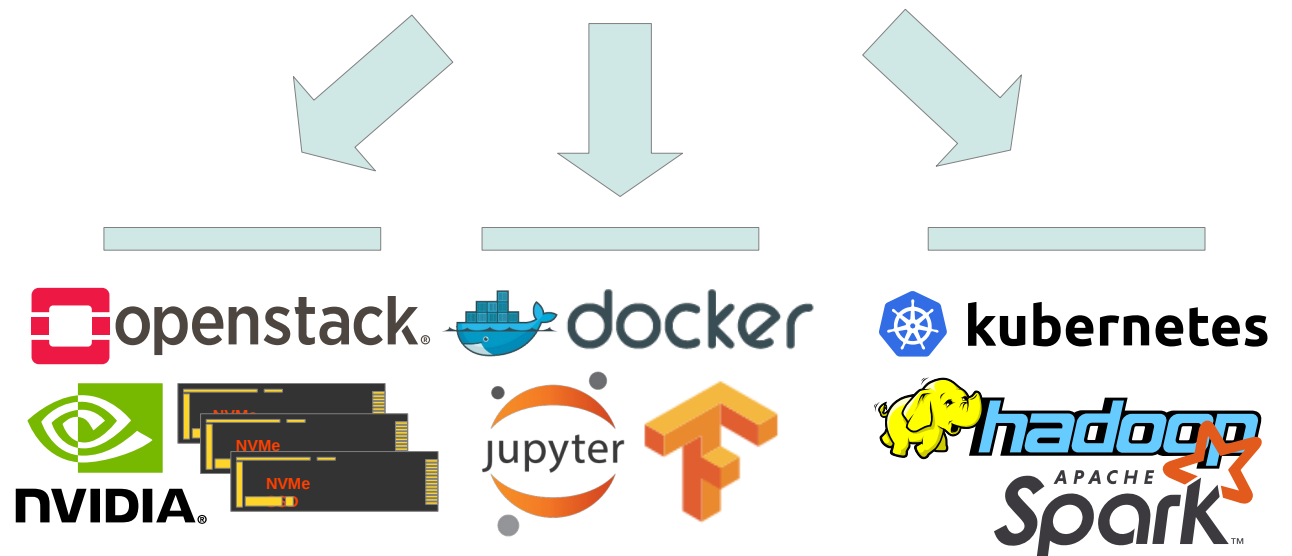
Asking the right question

Alternatively we can delegate the question to the individual platforms

Each platform only needs to understand the technologies it provides.

If a platform doesn't understand the question, it can just say no.

Can your platform run **<this>** task ?



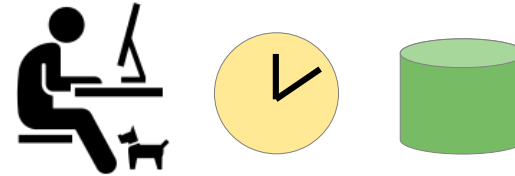
IVOA ExecutionPlanner

Asking the right question

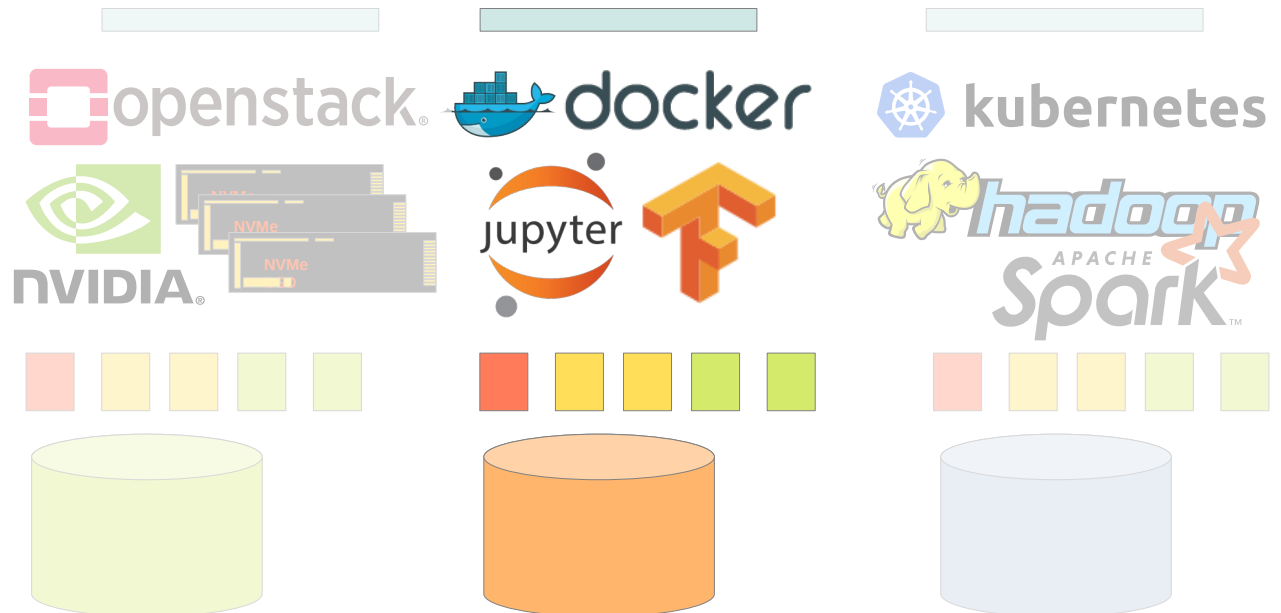
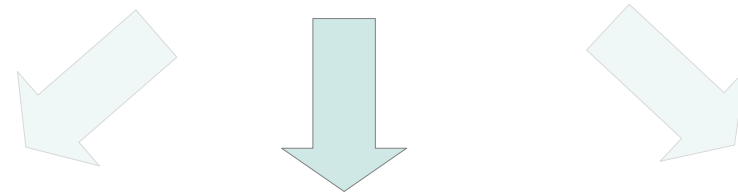
Each platform only needs to know about its own capabilities.

If we add data, then a platform can take account how 'close' the data is.

If a platform can't access the data, it can just say no.



<when> can <I> run <this> task with <this> data on your platform ?



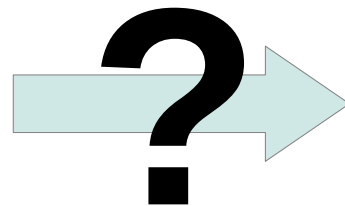
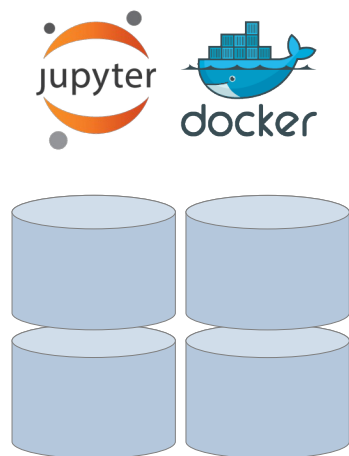


Why would we need this ?

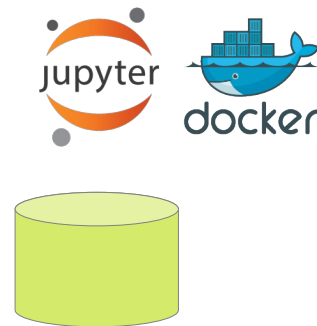
Our our platform is the primary source for our data.

Our users don't need to use other platforms.

BIG NAME PLATFORM



Another platform



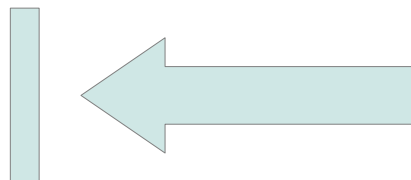
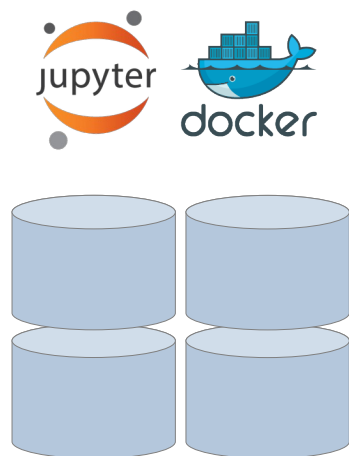


Why would we need this

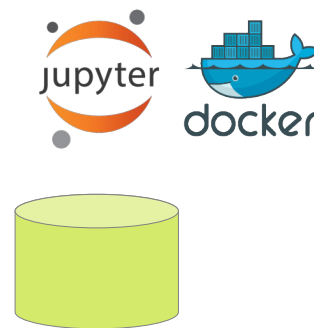
ExecutionPlanner is not for your uses to escape.

ExecutionPlanner provides another route for users to access your platform.

BIG NAME PLATFORM



Another platform



Another client



IVOA ExecutionPlanner



Initial working draft

Work in progress

We need your input

Your use cases



*International
Virtual
Observatory
Alliance*

IVOA Execution Planner Version 1.0

IVOA Working Draft 2023-07-01

Working Group
GWS

This version

<https://www.ivoa.net/documents/ExecutionPlanner/20230701>

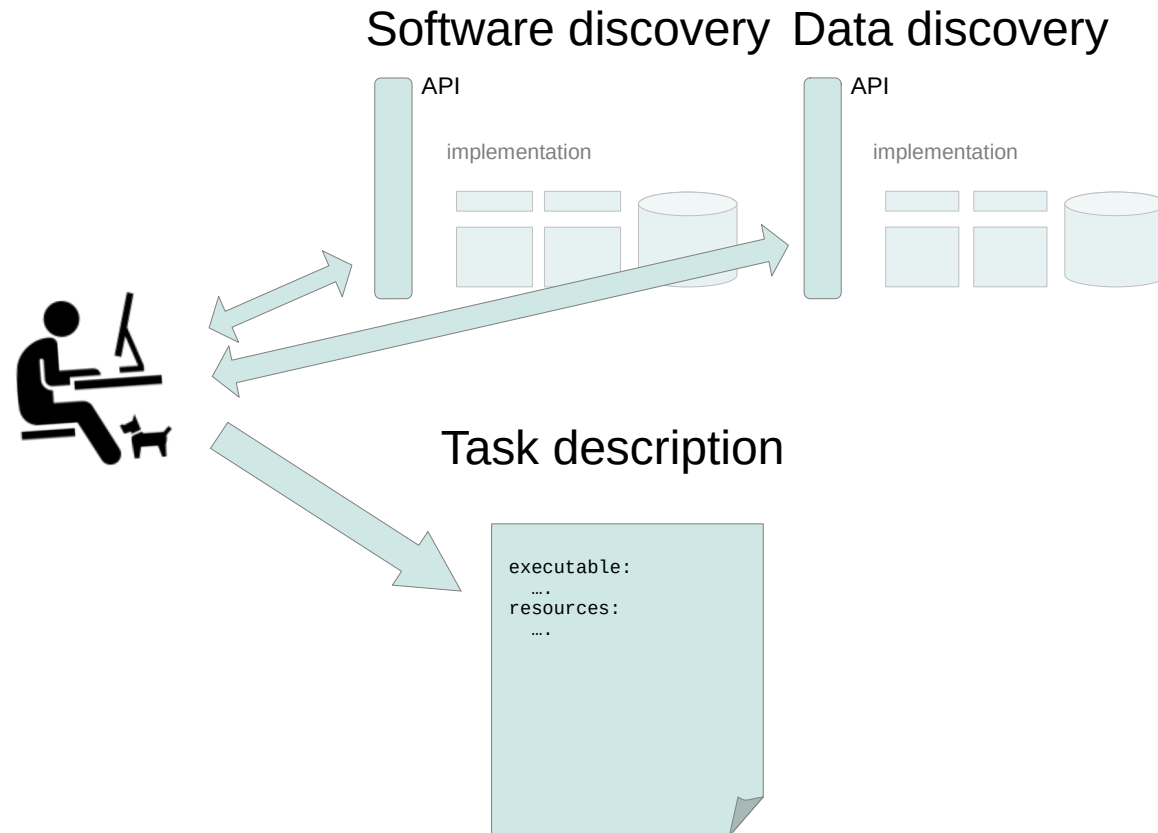
Latest version

<https://www.ivoa.net/documents/ExecutionPlanner>

<https://github.com/ivoa-std/ExecutionPlanner>

IVOA interop
Tucson, Arizona
November 2023

Step #1 Discovery, select the data and software





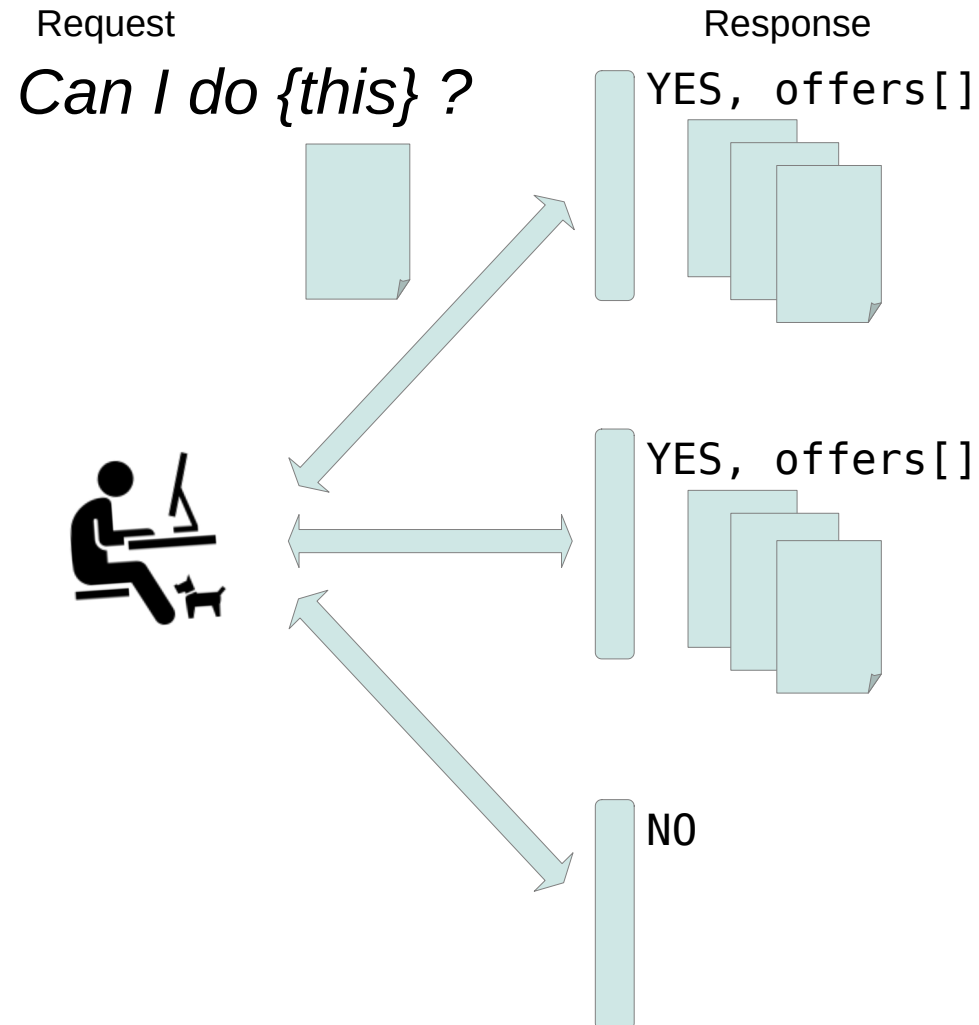
Step #2 Query, can I do this ?

Execution Planner API

Query one or more Execution Planner services.

Each Execution planner replies with one or more offers.

The offers will include details of the available resources and estimated cost.

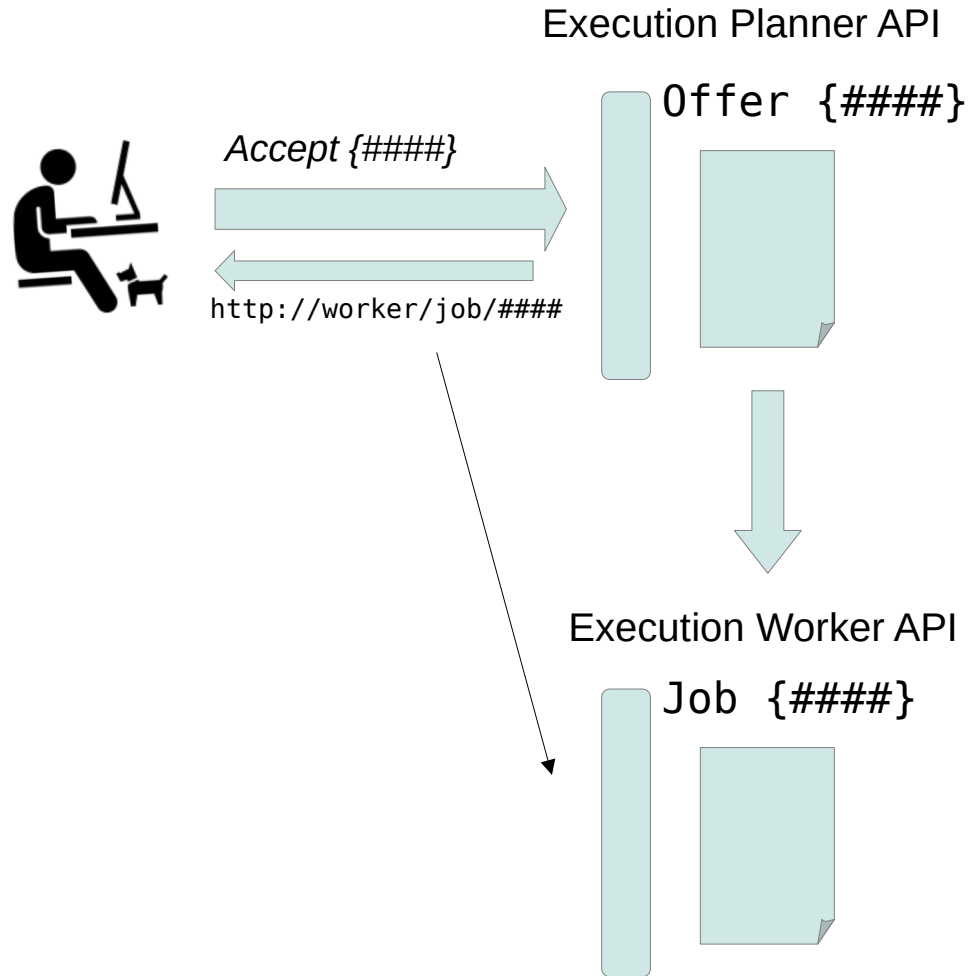




Step #3 Accept an offer

Chose the offer that best fits your requirements.

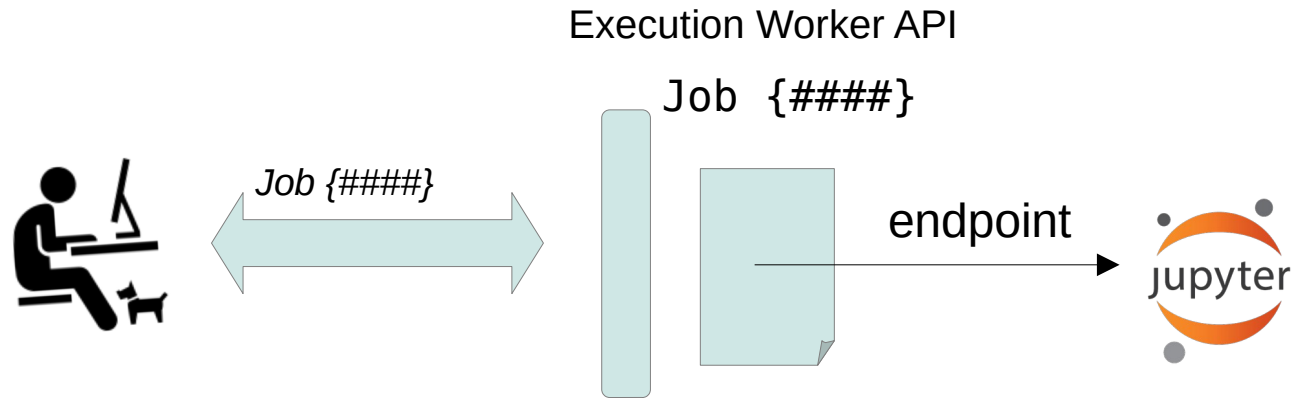
Accepting an offer initiates a job in an Execution Worker service.



Step #4 Monitor the job



User can monitor the job status via the Execution Worker API



PENDING

The job has been created, but the resources have not be prepared yet.

SETUP

The service is setting up the resources needed to execute the job.

READY

The job is ready and waiting to start.

RUNNING

The job is running.

TEARDOWN

The job has finished and the service is clearing up the resources.

FAILED

COMPLETED

CANCELLED



The data model

Blocks for executable and resources that can be extended by type.

Described in YAML, but intended to be serialization agnostic.

```
# Request to run a Jupyter notebook.
request:

# Details of the executable.
executable:

# A URI identifying the type of executable.
type: "https://... /types/jupyter-notebook"

# The details, specific to a Jupyter notebook.
spec:

notebook: "https://.../example.jpnb"
requirements: "https://.../requirements.txt"
```




The data model

Everything is optional.

Leave the notebook blank and you just get a Jupyter session.

```
# Request to run a Jupyter session.  
request:  
  
# Details of the executable.  
executable:  
  
# A URI identifying the type of executable.  
type: "https://... /types/jupyter-notebook"  
  
# The details, specific to a Jupyter notebook.  
spec:  
  
notebook:  
requirements: "https://.../requirements.txt"
```



The data model

Request and offer are based on properties and metrics.

ANY properties and metrics.

What we have:

We have properties for Jupyter notebooks and Docker containers.

We have properties for compute resources, cpu cores, memory and storage.

Still todo :

Properties to refer to and describe data sets.

Properties to measure data proximity.(network cost).

Properties to measure data access (latency and bandwidth).

IVOA ExecutionPlanner



Initial working draft

Work in progress

We need your input

Your use cases



*International
Virtual
Observatory
Alliance*

IVOA Execution Planner Version 1.0

IVOA Working Draft 2023-07-01

Working Group
GWS

This version

<https://www.ivoa.net/documents/ExecutionPlanner/20230701>

Latest version

<https://www.ivoa.net/documents/ExecutionPlanner>

<https://github.com/ivoa-std/ExecutionPlanner>

IVOA interop
Tucson, Arizona
November 2023