



IVOA - science platforms

Moving code to the data

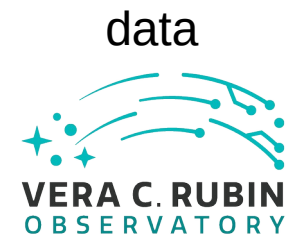
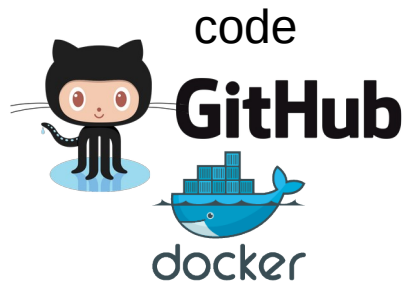
Dave Morris
Institute for Astronomy
Edinburgh



The problem

Heterogeneous code, data and compute.

Everyone is slightly different.



FAIR



Finding data sets

#1 Science criteria for things I'm interested in

....

Finding analysis code

#1 Science criteria for things I want to do

....

Finding execution platforms

....

FAIR



Finding data sets

- #1 Science criteria for things I'm interested in
- #2 In data formats I can use
- #2 At locations I can access

Finding analysis code

- #1 Science criteria for things I want to do
- #2 In packages I can use

Finding execution platforms

- #2 That can run the code I want
- #2 That can access the data I want
- #2 That are available when I need them

FAIR

Finding data sets



#1 Science criteria for things I'm interested in

IVOA Registry

Needs extending to cover datasets as stand alone things

Needs to reference to the data object itself rather than a data access service

#2 In data formats I can use

#2 At locations I can access

IVOA VOSpace

Supports immutable data

Supports multiple formats of the same data

Supports data in external 'cloud' storage – e.g. S3 and Rucio

Needs extending to handle location and proximity

FAIR

Finding analysis code



#1 Science criteria for things I want to do

OSSR made a start, but doesn't have a vocabulary

We need to develop a data model and vocabulary

#2 In packages I can use

IVOA ExecutionPlanner

Metadata schema for describing executable packages

The compute hardware it requires

The software environment it requires

FAIR

Finding execution platforms

#2 That can run the code I want

#2 That can access the data I want

#2 That are available when I need them

IVOA ExecutionPlanner

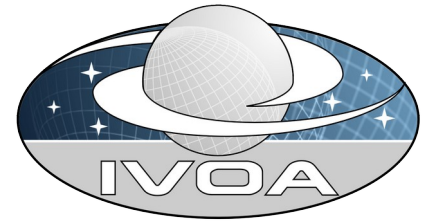
Metadata schema for describing executable tasks

- The compute hardware it requires

- The software environment it requires

- The external data it requires

- The time range it is required





IVOA ExecutionPlanner

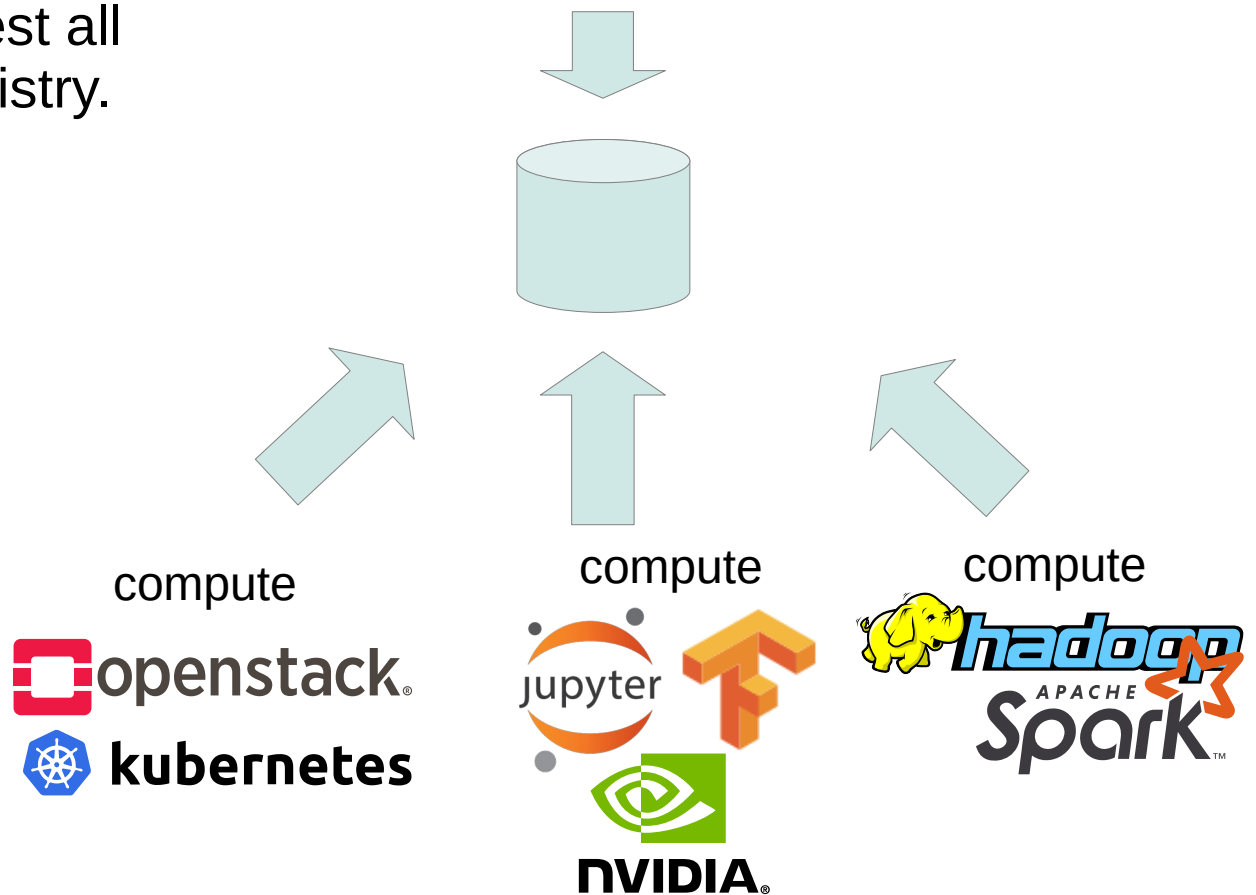
Asking the right question

In theory, it is possible to harvest all the metadata into a central registry.

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

Which places a lot of complexity in a single entity.

Find all the platforms that can run **<this>** task





IVOA ExecutionPlanner

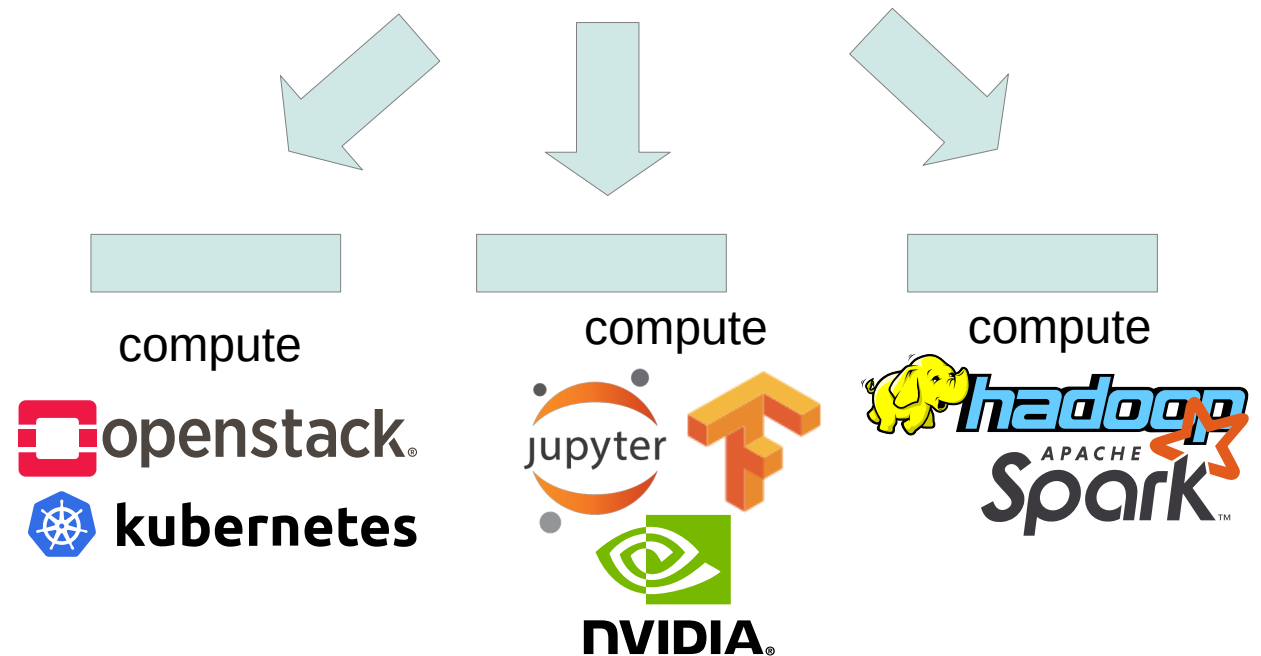
Asking the right question

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 ?



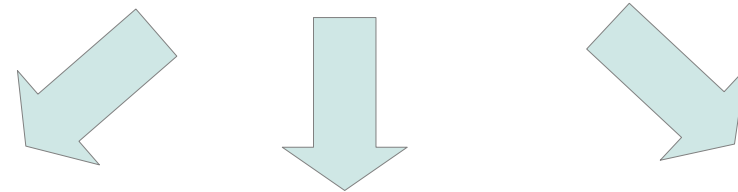


IVOA ExecutionPlanner

Asking the right question



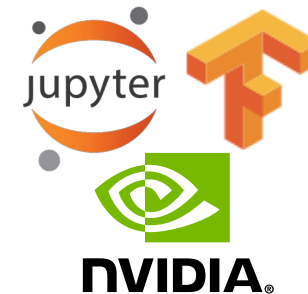
Can **<I>** run **<this>** task on your platform ?



compute



compute



compute



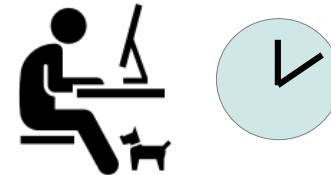
If we add the user's identity, then the platforms can apply access controls.

The platforms can apply different quotas to different users.

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

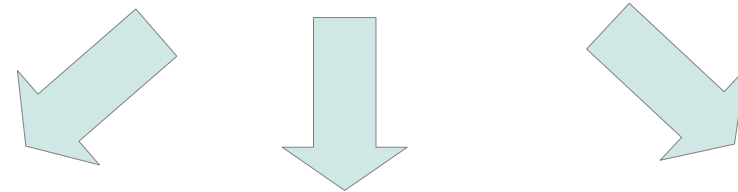
IVOA ExecutionPlanner

Asking the right question



If we add time, then a platform can take account of how busy it is.

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



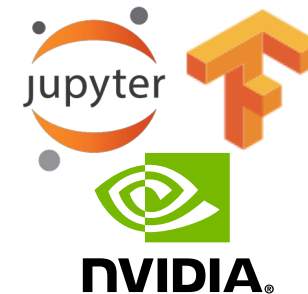
The platforms can offer a range of time slots for the user to choose.



compute



compute



compute



If a platform doesn't have any time available, it can just say no.

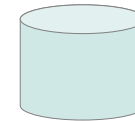
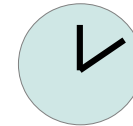
IVOA ExecutionPlanner

Asking the right question

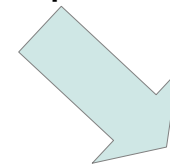
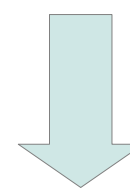
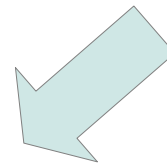
If we add data, then a platform can take account how 'close' the data is.

The platforms can quote different start-up times depending on how long it will take to stage the data.

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



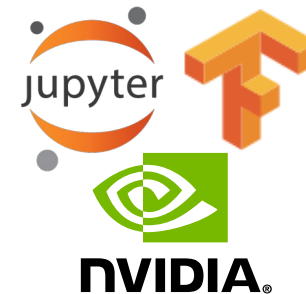
<when> can <I> run <this> task
with <that> data on your platform ?



compute



compute



compute



FAIR

We have ideas and plans for most of these



Finding data sets

- ✓ #1 Science criteria for things I'm interested in
- #2 In data formats I can use
- #2 At locations I can access

Finding analysis code

We (IVOA) don't know how to do this (yet).

- #1 Science criteria for things I want to do**
- #2 In packages I can use

Finding execution platforms

- ✓ #2 That can run the code I want
- #2 That can access the data I want
- #2 That are available when I need them



Thank you, and any questions ?

Dave Morris
Institute for Astronomy
Edinburgh
<dmr@roe.ac.uk>

<https://ivoa.net/>

<https://github.com/ivoa/ExecutionPlannerNote/>

