Containerize, don’t improvise!

Importance of Containers from a Scientist’s View

Andreas Faisst

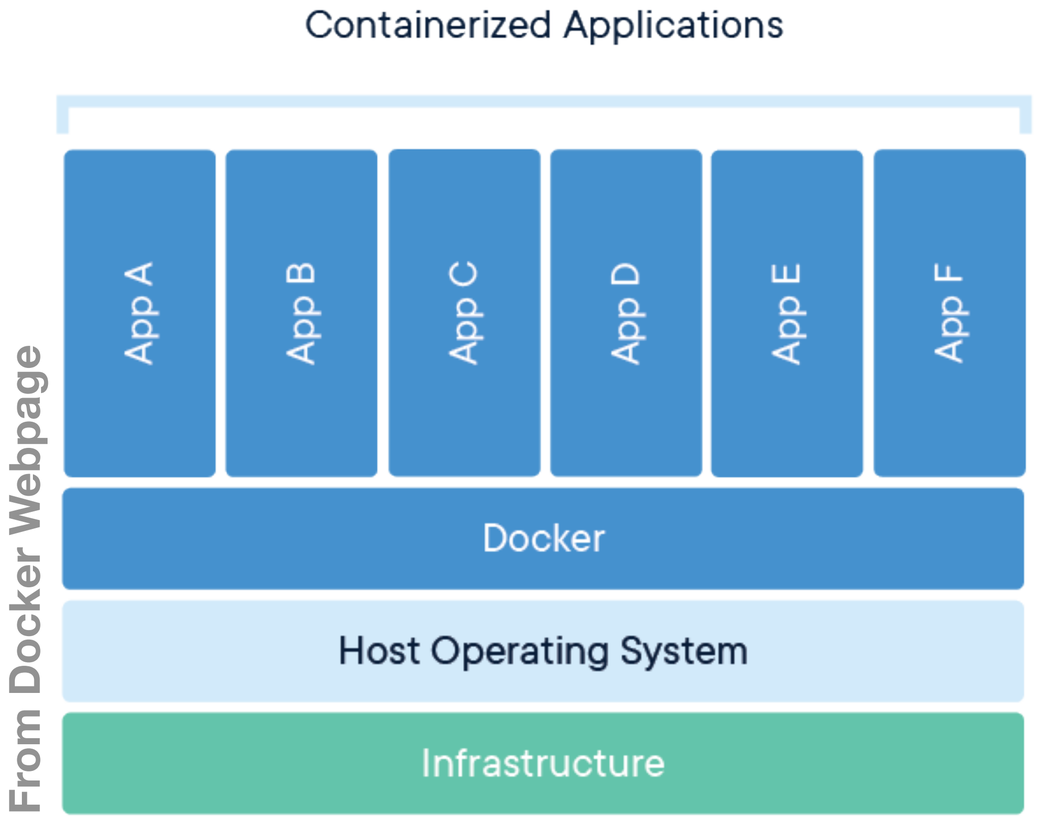
(Caltech/IPAC)

afaisst@caltech.edu

@astrofaisst

What are Containers?

- Containers host an environment with compiled applications
- They are portable and can be deployed almost anywhere

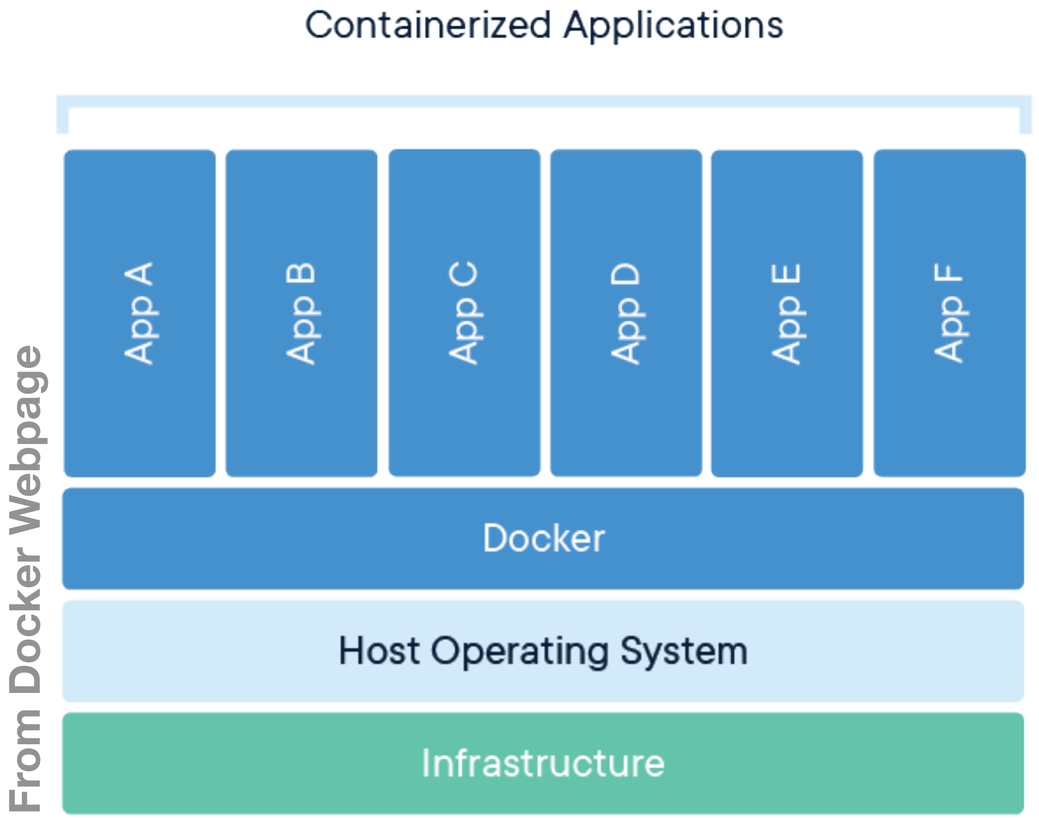


What are Containers?

- Containers host an environment with compiled applications
- They are portable and can be deployed almost anywhere
- Different types of containers and container management software (see later talks on pros and cons and details)

Kubernetes

Singularity



What Containers really are (for a Scientist)

- “A method to ease the work of lazy* scientist who do not want to compile their applications on their own (don’t improvise!).”

  *smart

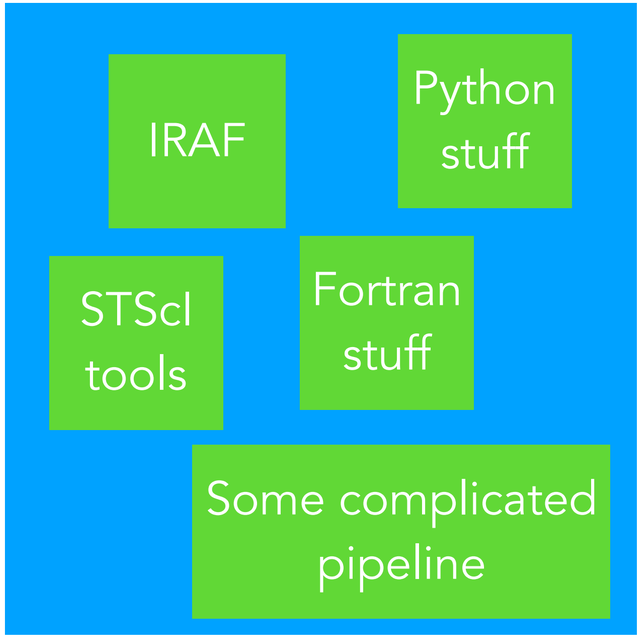
What Containers really are (for a Scientist)

- "A method to ease the work of lazy* scientist who do not want to compile their applications on their own (don't improvise!)."
  *smart
- Spend a couple of hours to install all you need in a container from scratch (do it right).
  This can include:

➡ a whole Python installation (e.g., Python 3.6)

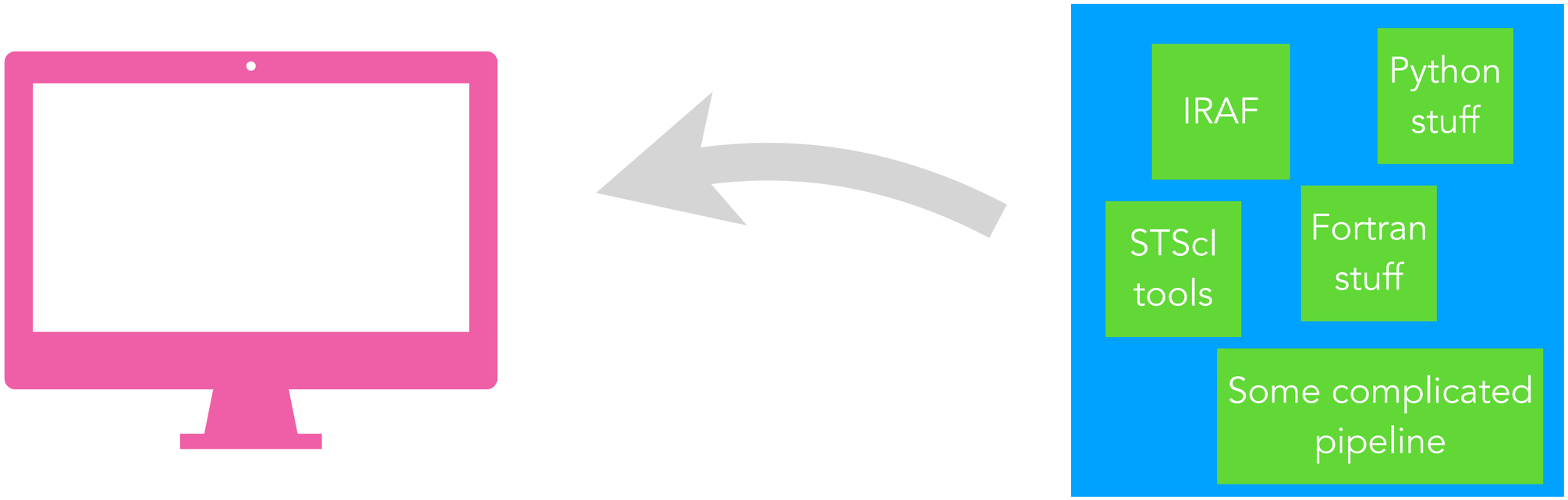
➡ various applications with dependencies

➡ data (although not recommended because you want the container image to be as light as possible)



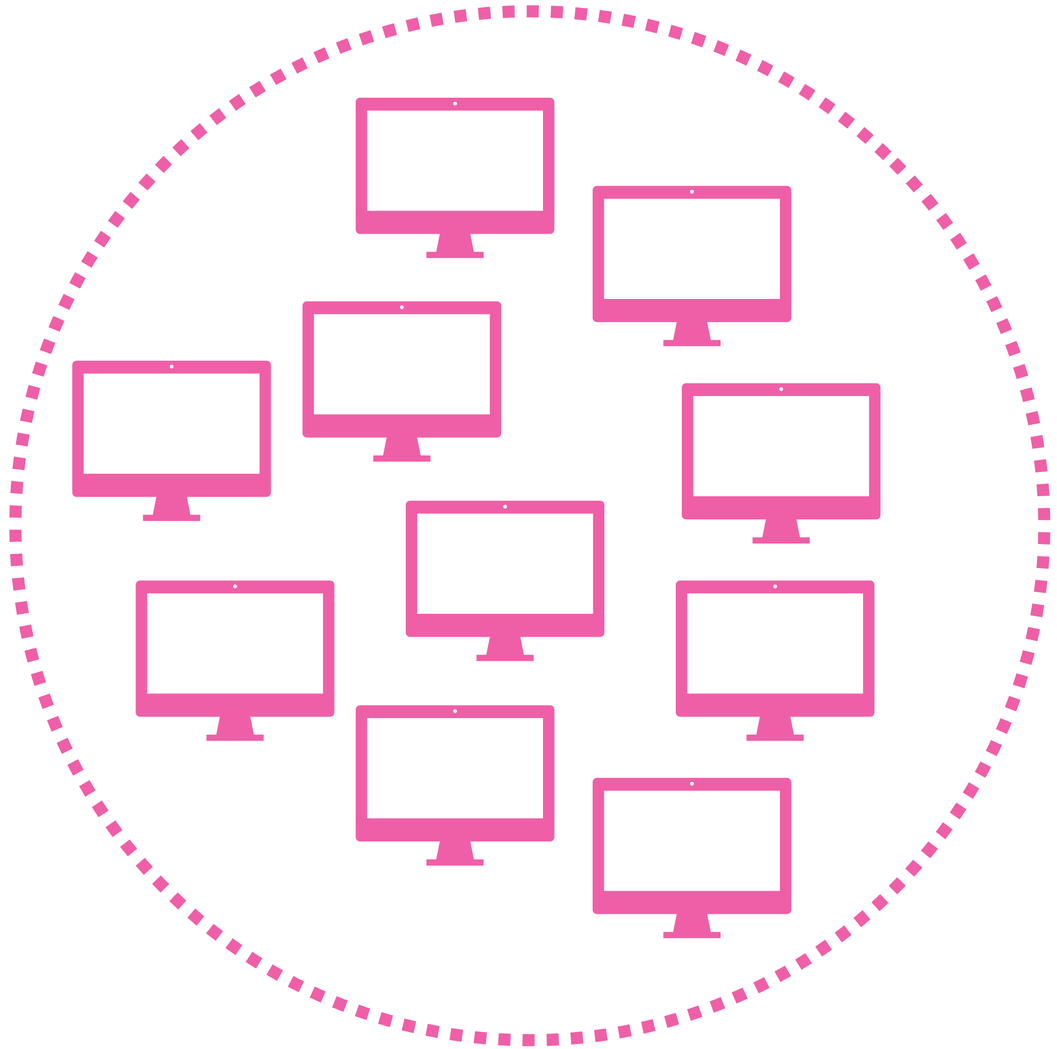
Containers are Portable

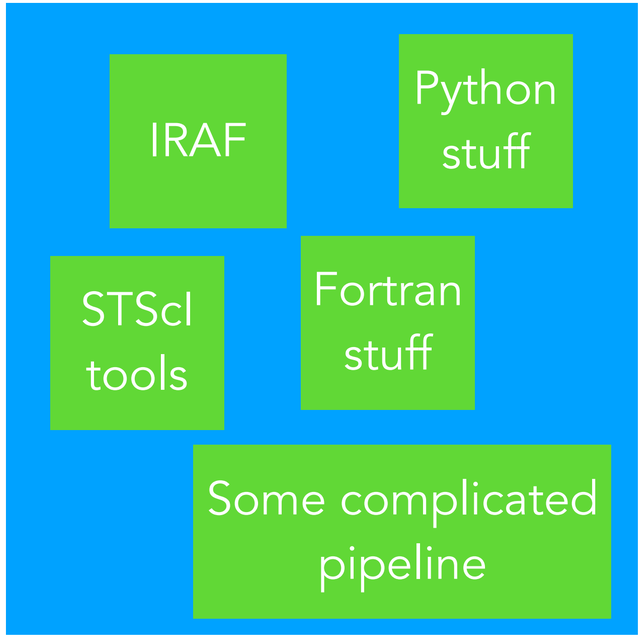
- The full container can be put on any computer (might need to match OS!)



Containers are Portable

- You can also put them in computer clusters without recompiling everything





Containers are Portable

- Or even on “cloud computing” facilities like the Amazon Web Services (AWS)

Stories of Containers

- Environments that I have used:

➡ **(Small) Docker images**

On my own desktop computer, for very specific software that is complicated to compile

➡ **Singularity environment containers**

On computing facilities of the Joint Survey Processing (JSP) project.

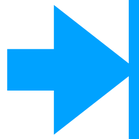
(JSP containers are created by Nathaniel S.; see also his talk).

Stories of Containers

- The syntax to “open" them is simple (even for scientists like me).

Docker:

```
> sudo docker build --tag=IMAGE
> sudo docker run --rm -it IMAGE bash
```



see **Ben’s talk** and the **13.15pm hands-on session** on creating a docker container

Singularity:

```
> singularity shell  --bind [SOME PATHS]  [IMAGE]
```

Stories of Containers

- The syntax to “open" them is simple (even for scientists).

Docker:

```
> sudo docker build --tag=IMAGE
> sudo docker run --rm -it IMAGE bash
```

see **Ben’s talk** and the **13.15pm hands-on session** on creating a docker container

Singularity:

```
> singularity shell --bind [SOME PATHS] [IMAGE]
```

Bind paths and mounts that you need

This is the container.
You don’t have to know how this was built!
Ideally pre-compiled with the help of software engineers

Stories of Containers (Joint Survey Processing)

- Example of a Singularity environment on JSP computers

  with pre-generated singularity container image “jsp_apps.sif” including all the required software!

Log in to JSP computer:

```
> ssh afaisst@vmjointproc01.ipac.caltech.edu
```

Start the container with:

```
> singularity shell    --bind /run,/stage    /singularity/jsp_apps.sif
```

Now you can run your programs without worrying about installation!

Stories of Containers (Science Platform)

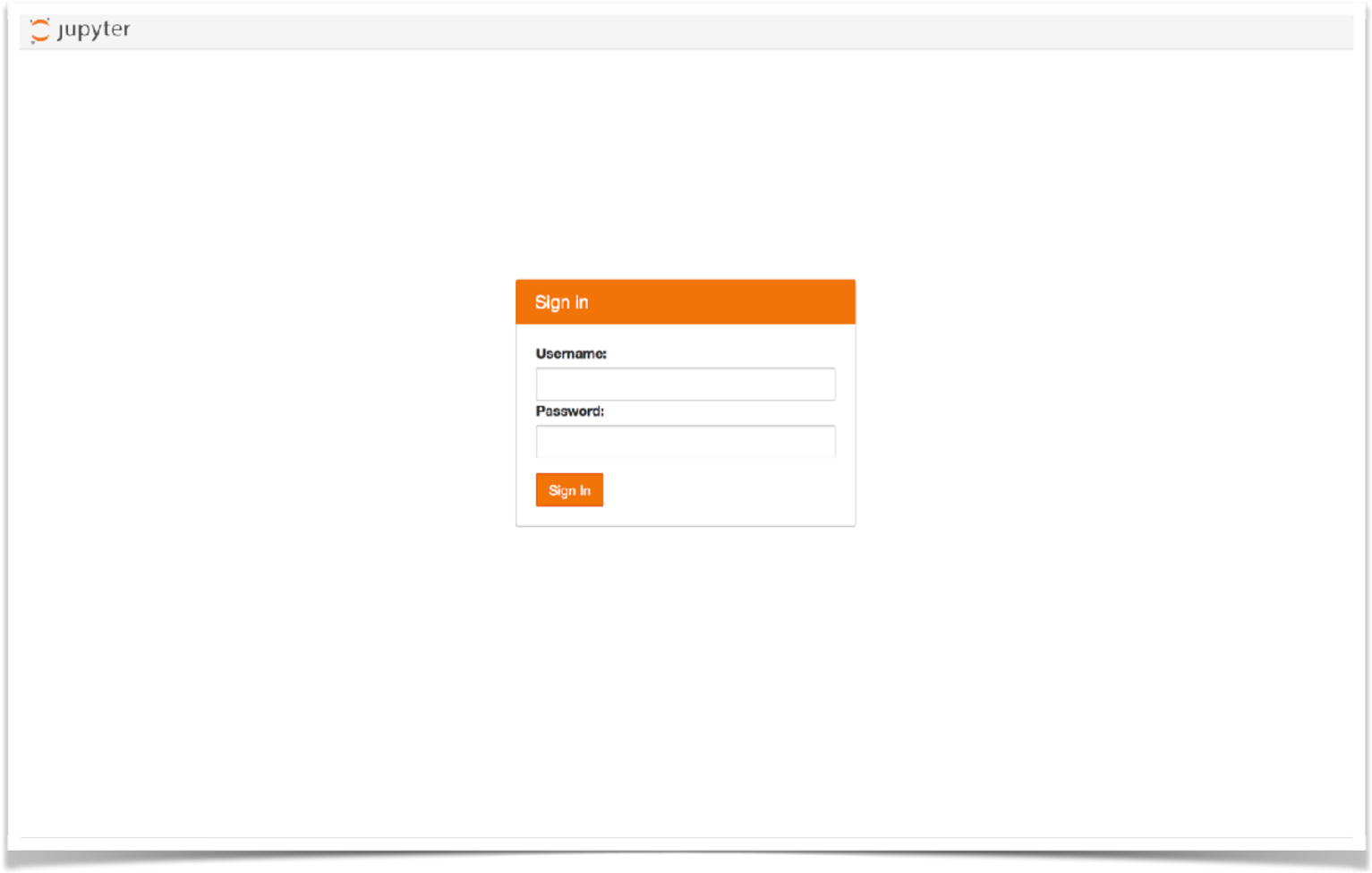
- The IPAC Science Platform (ISP) is based on containers

User can
- manipulate data close to the location where the data is stored
- avoid downloading large datasets
- avoid complicated of software installation

➡ Significantly increases the efficiency when working with very large datasets.

Stories of Containers (Science Platform)

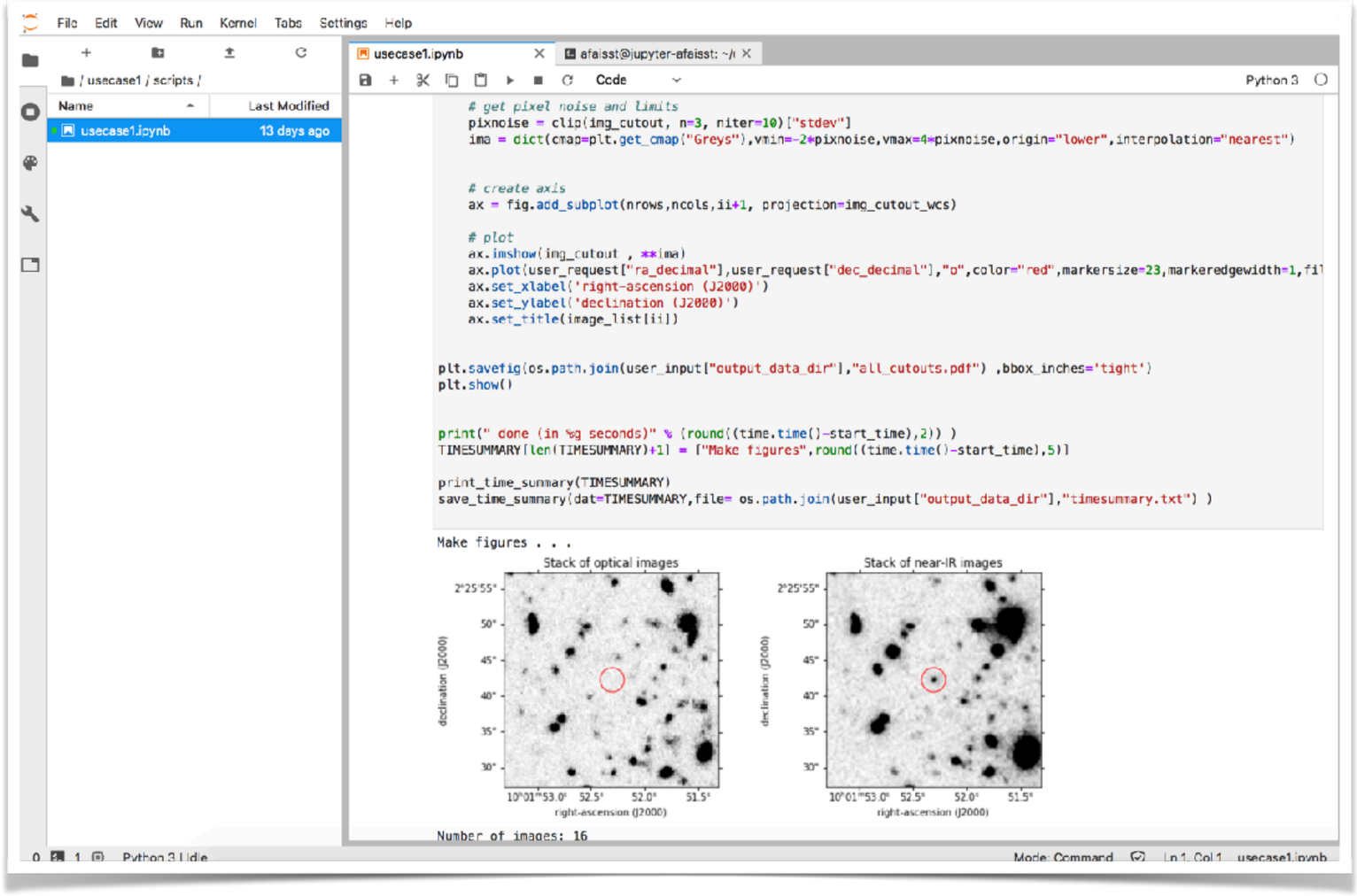
- Online login to Jupyter Hub



Stories of Containers (Science Platform)

- Starts Jupyter session (in JSP container)

➡ Computations with “instantaneous” access to data through your own machine!



Summary: Pros and Cons

Simplify the (work) life of scientists

Increase efficiency

Extend workflow to cluster/cloud computing and science platforms

Portable and reproducible

Container size affects portability (when they get too heavy!)

Have to be maintained and updated to include latest versions