



Containerize, don't improvise!

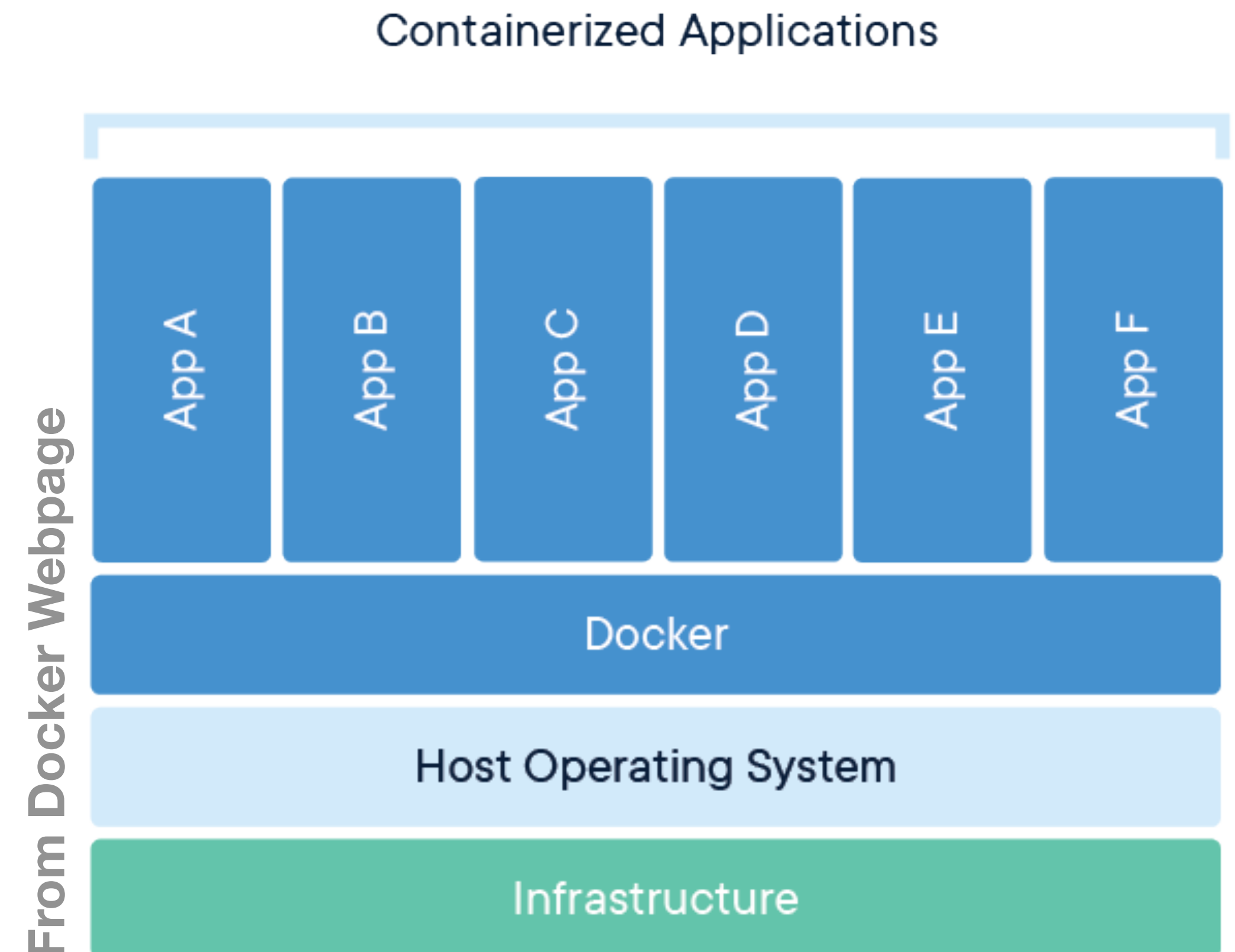
Importance of Containers from a Scientist's View

Andreas Faisst

(Caltech/IPAC)
afaisst@caltech.edu

What are Containers?

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



What are Containers?

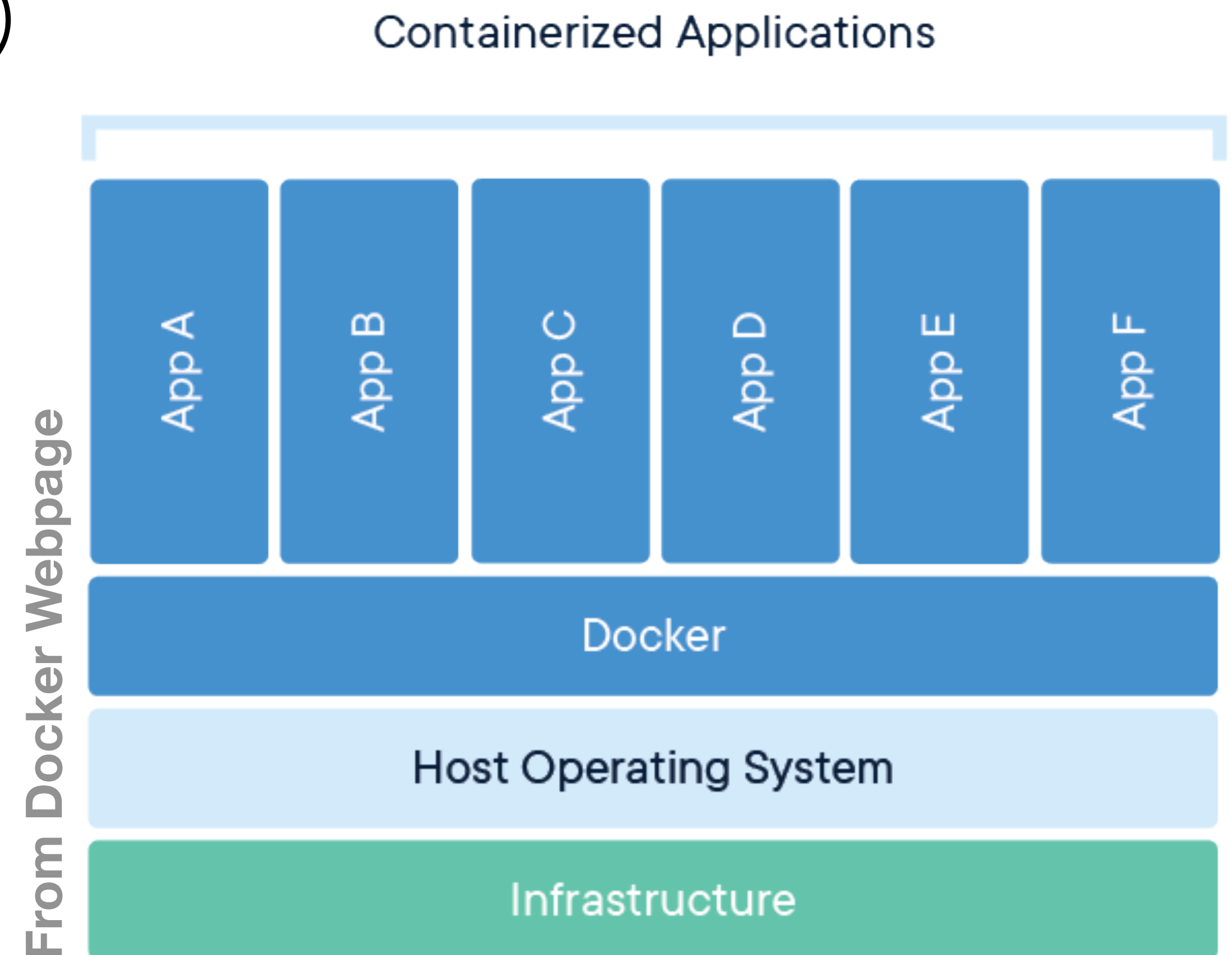
- Containers host an environment with compiled applications
- They are **portable** and can be **deployed almost anywhere**
- Different types of containers and container management software (see later talks on pros and cons and details)



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)

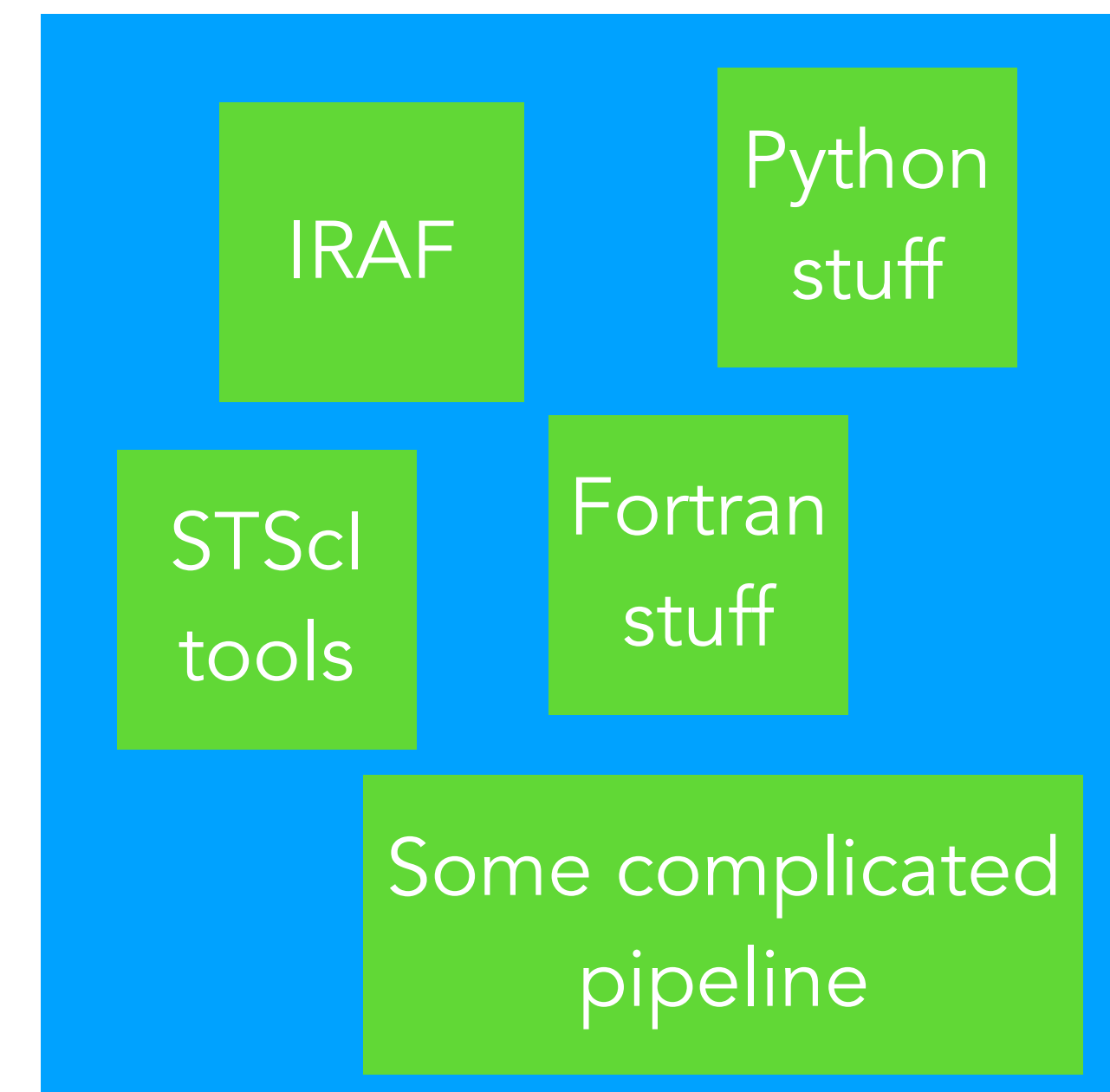
- "A method to ease the work of lazy* scientist who do not want to compile their applications on their own (don't improvise!)."

*smart

- Spend a couple of hours to install all you need in a container from scratch (do it right).

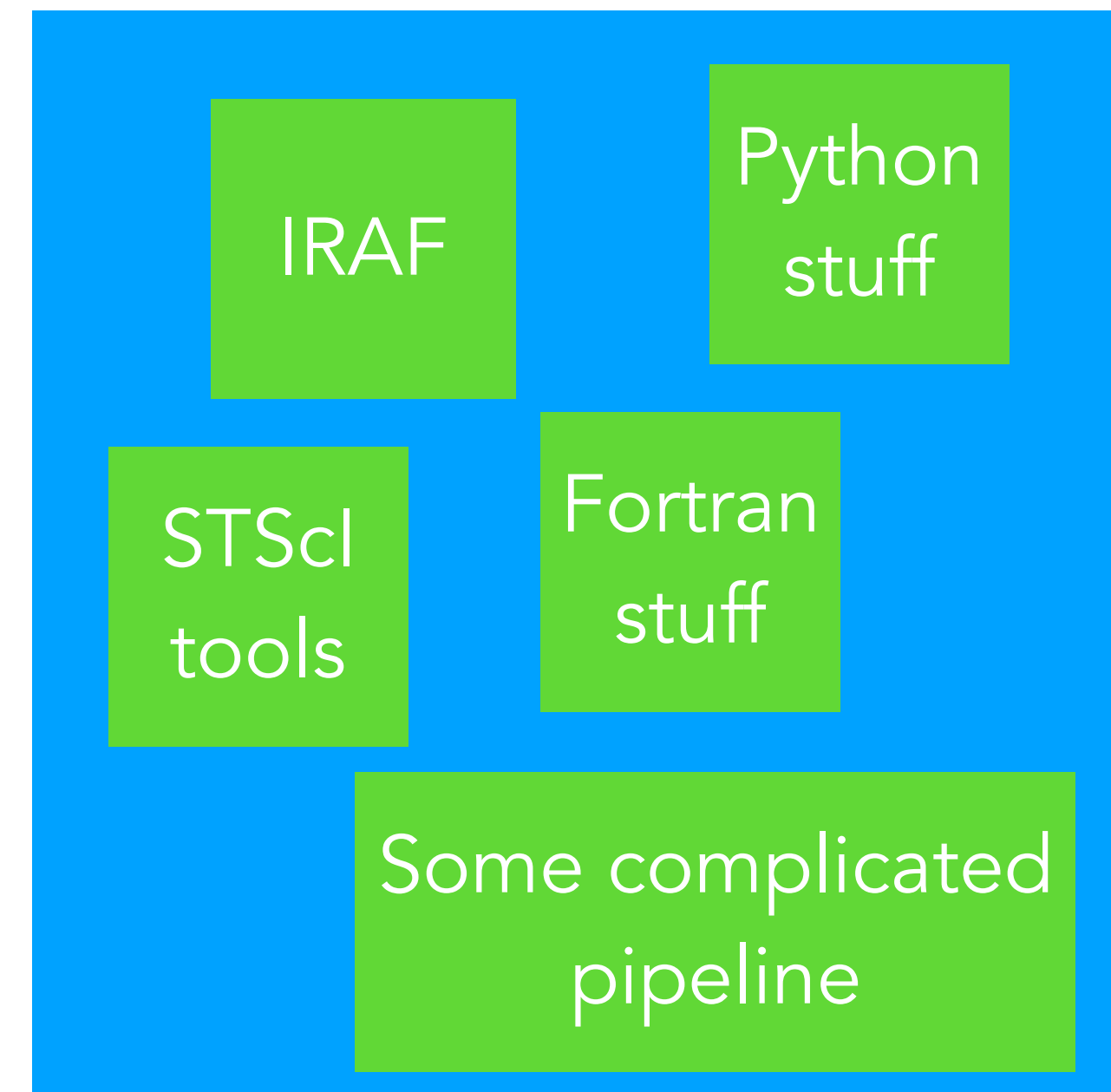
This can include:

- ➔ 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)



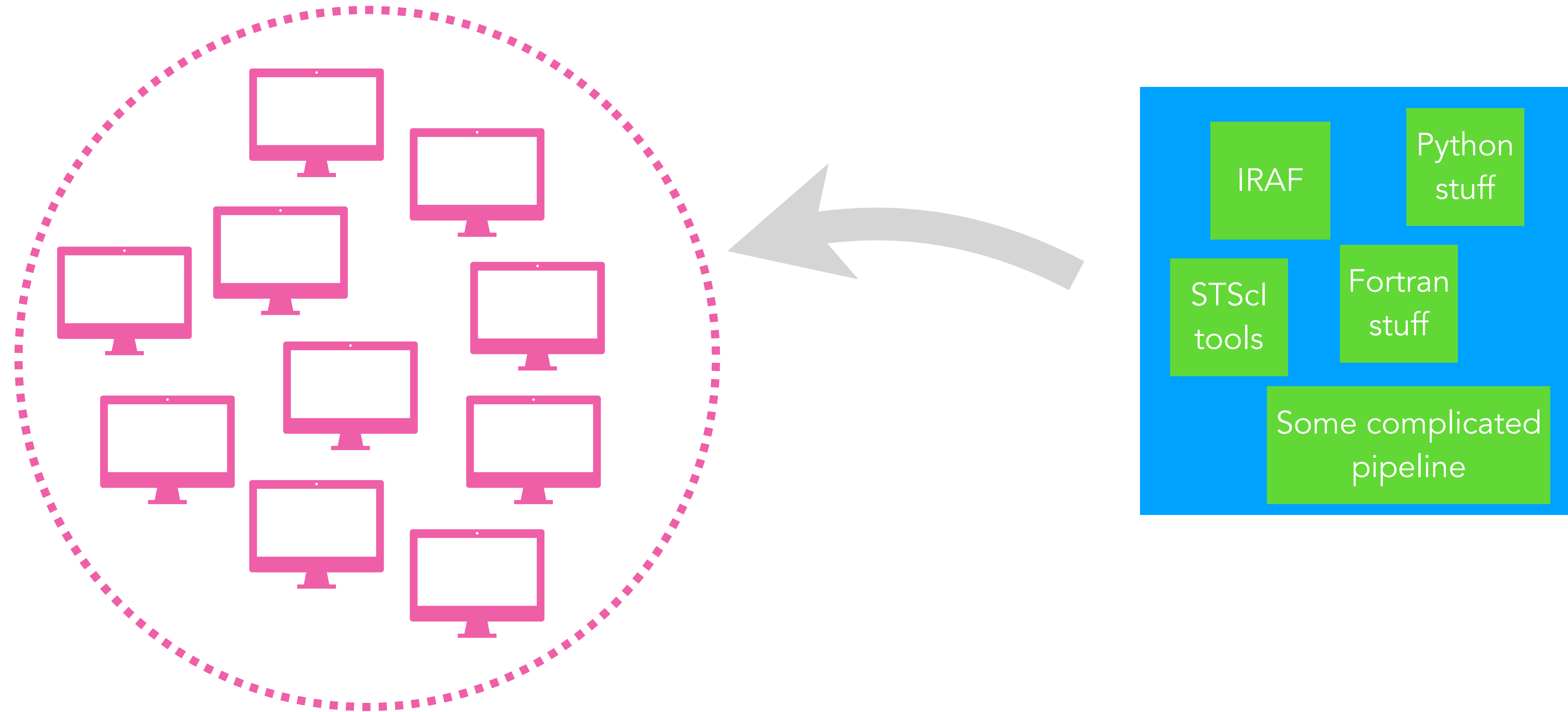
Containers are Portable

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



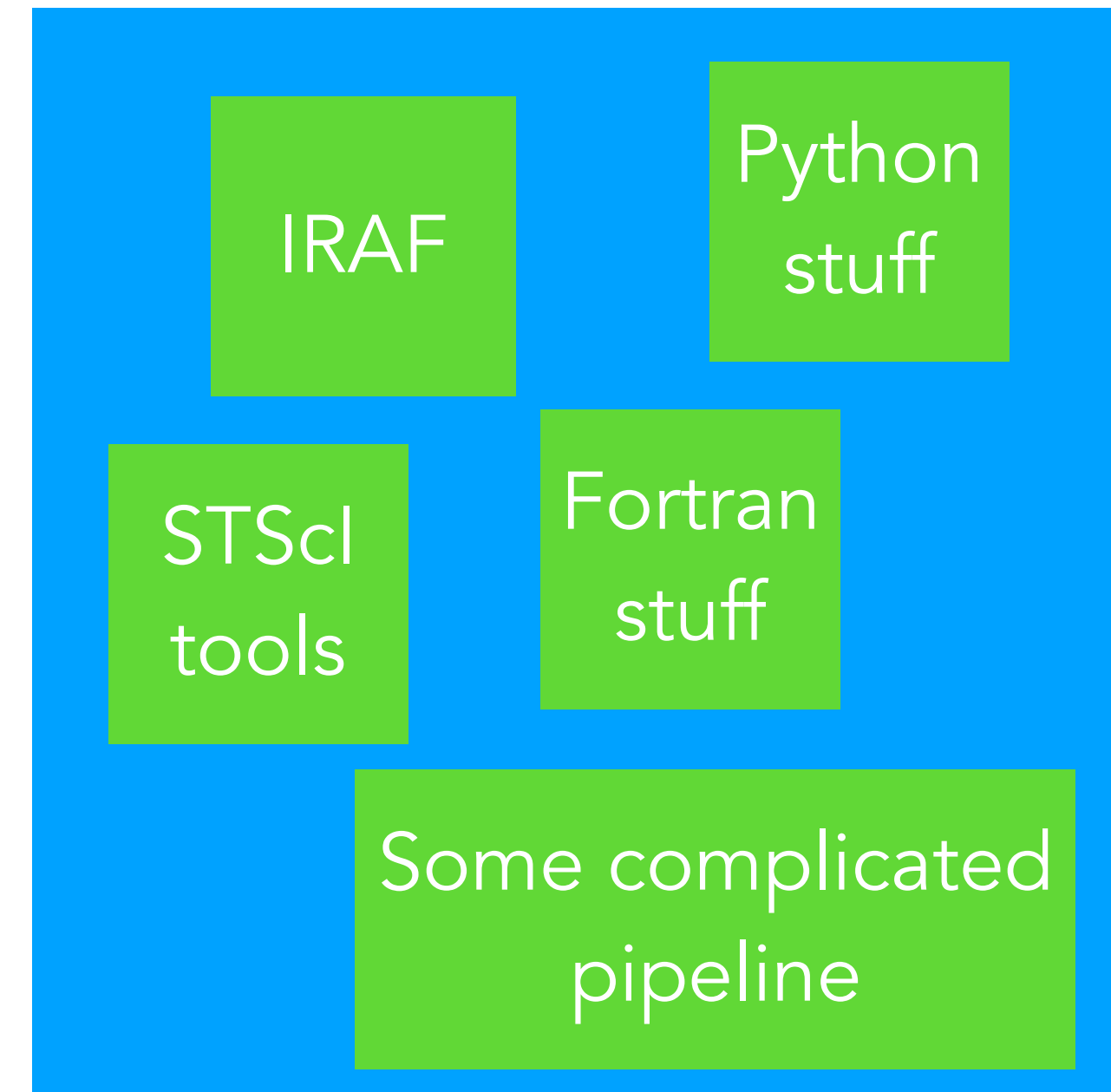
Containers are Portable

- 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**

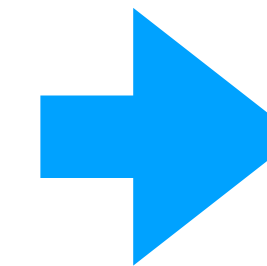
On computing facilities of the Joint Survey Processing (JSP) project.
(JSP containers are created by Nathaniel S.; see also his talk).

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:

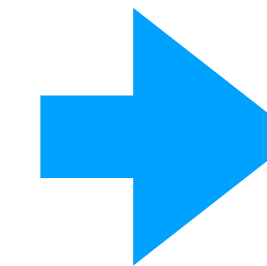
```
> singularity shell --bind [SOME PATHS] [IMAGE]
```

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

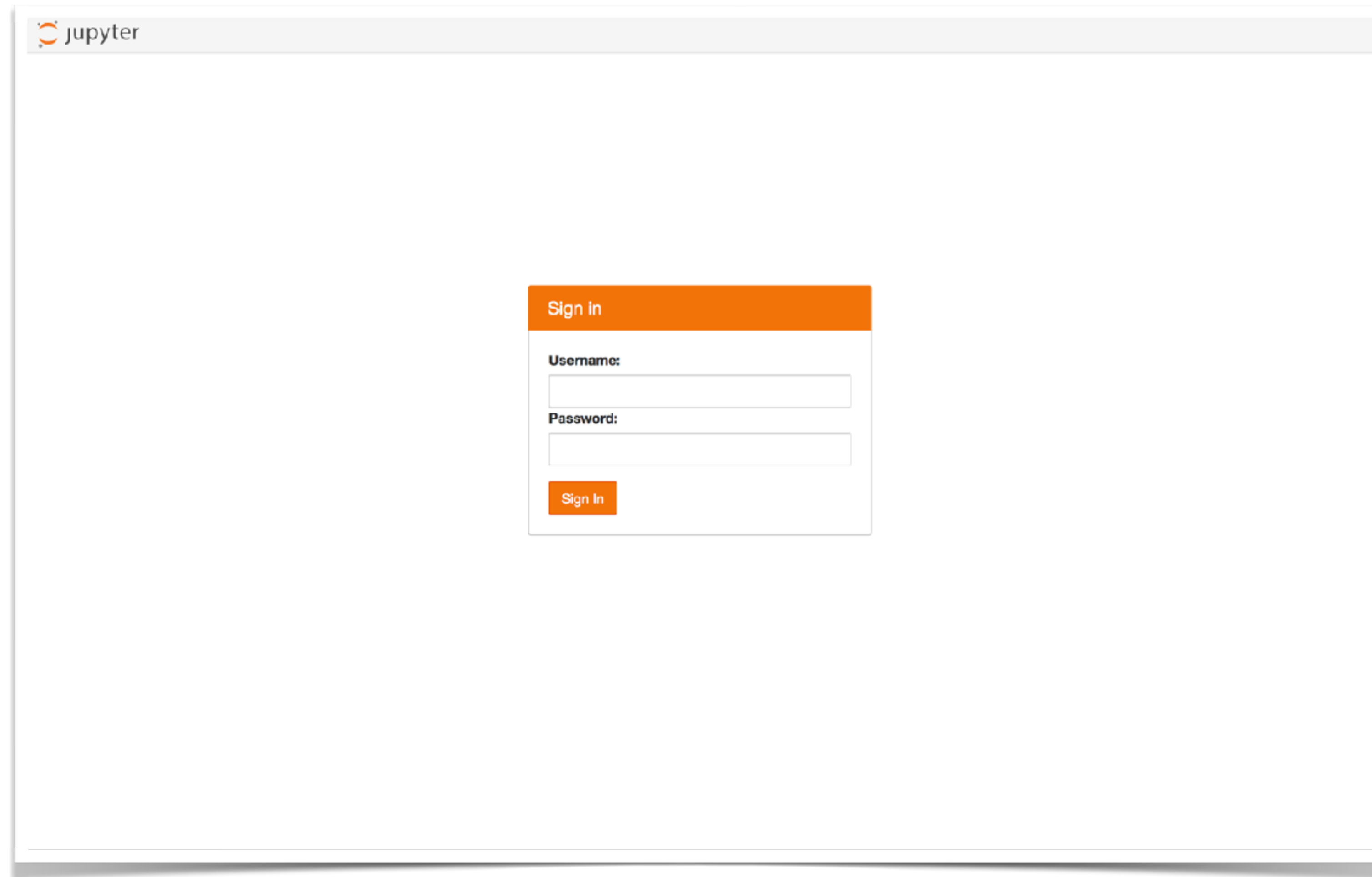
User can

- 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)

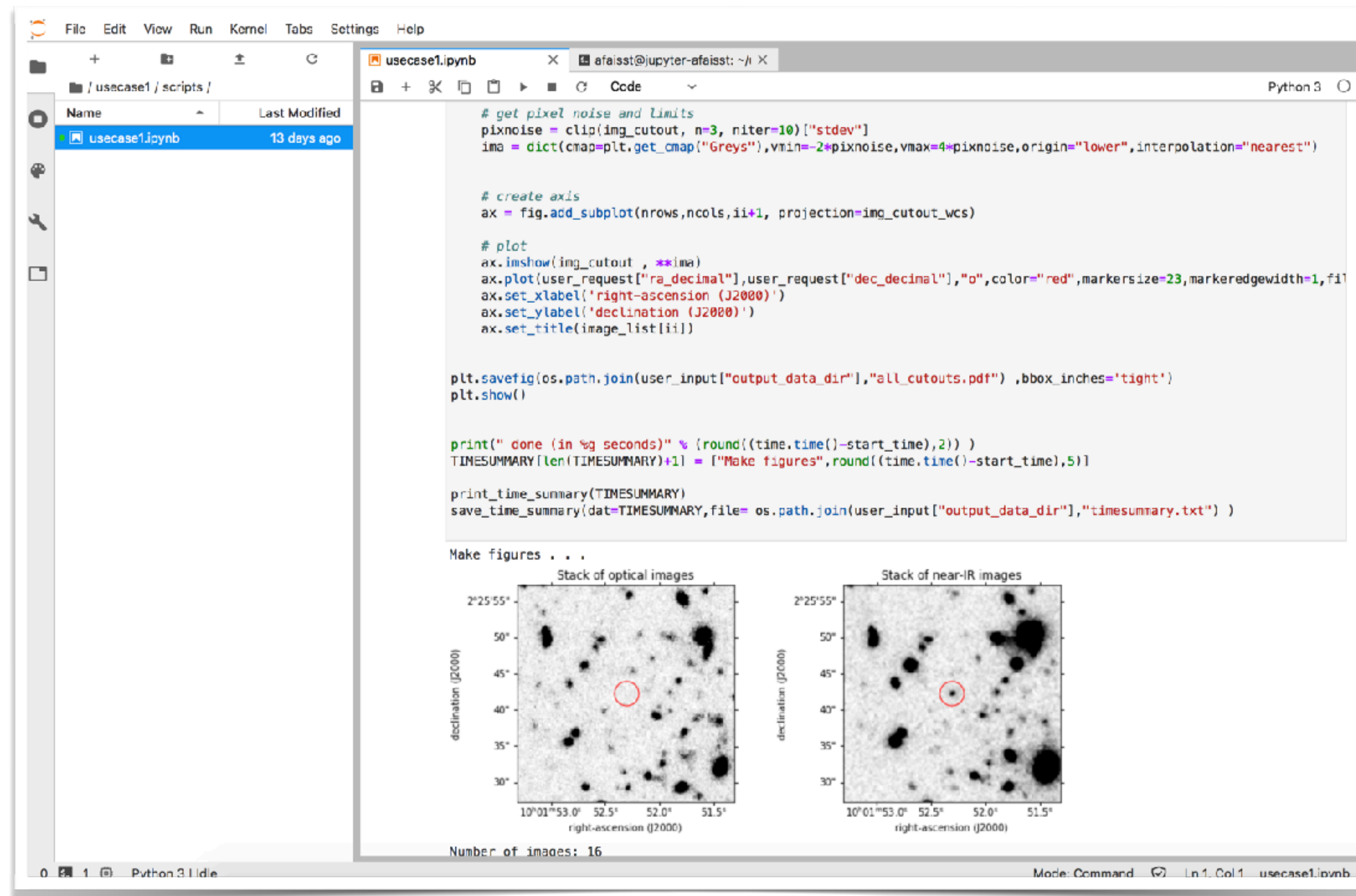
- Online login to Jupyter Hub



The image shows a screenshot of the Jupyter Hub login page. At the top left, there is a header with the Jupyter logo and the text "jupyter". In the center of the page, there is a "Sign In" form. The form has an orange header bar with the text "Sign In". Below the header, there are two input fields: "Username:" and "Password:". Below the "Password:" field, there is an orange "Sign In" button.

Stories of Containers (Science Platform)

- Starts Jupyter session (in JSP container)
 - ➔ Computations with “instantaneous” access to data through your own machine!



The screenshot displays a Jupyter Notebook environment. The left sidebar shows a file browser with a file named 'usecase1.ipynb' modified 13 days ago. The main area contains a code cell with the following Python code:

```
# get pixel noise and limits
pixnoise = clip(img_cutout, n=3, niter=10) ['stdev']
ima = dict(cmap=plt.get_cmap("Greys"), vmin=-2*pixnoise, vmax=4*pixnoise, origin="lower", interpolation="nearest")

# create axis
ax = fig.add_subplot(nrows,ncols,ii+1, projection=img_cutout_wcs)

# plot
ax.imshow(img_cutout, **ima)
ax.plot(user_request["ra_decimal"], user_request["dec_decimal"], "o", color="red", markersize=23, markeredgewidth=1, fill=False)
ax.set_xlabel('right-ascension (J2000)')
ax.set_ylabel('declination (J2000)')
ax.set_title(image_list[ii])

plt.savefig(os.path.join(user_input["output_data_dir"], "all_cutouts.pdf"), bbox_inches='tight')
plt.show()

print(" done (in %g seconds)" % (round((time.time()-start_time),2)) )
TIMESUMMARY[len(TIMESUMMARY)+1] = ["Make figures", round((time.time()-start_time),5)]

print_time_summary(TIMESUMMARY)
save_time_summary(dat=TIMESUMMARY, file= os.path.join(user_input["output_data_dir"], "timesummary.txt") )
```

Below the code, the output shows the text "Make figures . . ." followed by two side-by-side astronomical plots. The left plot is titled "Stack of optical images" and the right plot is titled "Stack of near-IR images". Both plots show a field of stars with a red circle highlighting a specific star. The axes are labeled with right-ascension (J2000) and declination (J2000). At the bottom of the plots, it says "Number of images: 16".

Summary: Pros and Cons

Simplify the
(work) life of
scientists

Increase
efficiency

Portable and
reproducible

Extend workflow to
cluster/cloud
computing and
science platforms

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

Have to be
maintained and
updated to include
latest versions