

# SDSC Summer Institute 2022

***Title: Python for HPC***

*Instructor: Mahidhar Tatineni*

*Date: 09:45AM-12:00 (PT), August 3, 2022*

*Slack support channel: main-room*

*Slack general support: help-desk*



# Outline

- Introduction to Jupyter notebooks and JupyterLab
- Single-node Python code optimization with numba
- Dask tutorial: overlap functions, introduction to dask array, distributed scheduler
- Dask array in-depth tutorial for multi-core, out-of-core, multi-node computing

# Outline

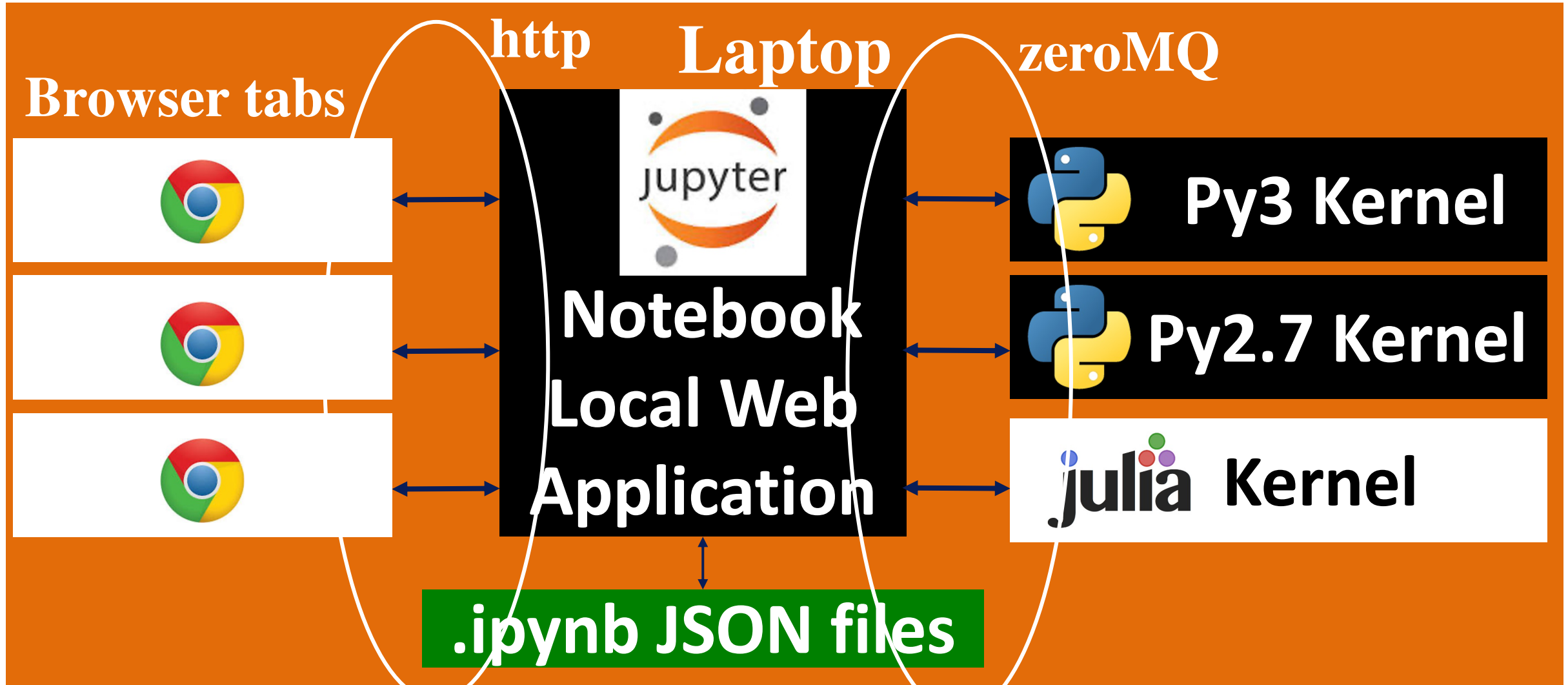
- **Introduction to Jupyter notebooks and JupyterLab**
- **Single-node Python code optimization with numba**
- **Dask tutorial: overlap functions, introduction to dask array, distributed scheduler**
- **Dask array in-depth tutorial for multi-core, out-of-core, multi-node computing**

# Overview of Jupyter notebooks and JupyterLab

- Browser based interactive console
- Supports multiple sessions in browser tabs
- Each session has a Kernel executing computation
- Saved in JSON format
- LIGO notebook examples for interactive data analysis of gravitational waves from black holes merging:

<http://beta.mybinder.org/repo/losc-tutorial/LOSC> Event tutorial

# Jupyter notebook local





# Jupyter notebook remote

Laptop



https +  
password

Jupyter  
Notebook  
Web  
Application

.ipynb JSON files

Server



Py3 Kernel



Py2.7 Kernel



julia Kernel

# Clone workshop repository

**ssh into Expanse with training account**

**git clone <https://github.com/sdsc/sdsc-summer-institute-2022>  
cd sdsc-summer-institute-2022**

# Launch notebook job

- Change to python HPC directory and launch job:  
`cd 4.2a_python_for_hpc`  
`bash launch_jupyter_singularity.sh`  
(note: this is using galyleo)
- Check your job status with:  
`queue -u $USER`
- Open browser on your laptop and connect to URL



# Outline

- Introduction to Jupyter notebooks and JupyterLab
- **Single-node Python code optimization with numba**
- Dask tutorial: overlap functions, introduction to dask array, distributed scheduler
- Dask array in-depth tutorial for multi-core, out-of-core, multi-node computing

# Numba: JIT compiler for Python

- Based on LLVM (compiler infrastructure behind clang, Apple's C++ compiler)
- Turns Python code into machine code on-the-fly

# Synchronization with content on YouTube (Andrea Zonca's Channel)

- Terminal on Expanse with training account
- Browser window or phone with Youtube videos:  
<https://bit.ly/pythonhpc2021> (also linked from repo)
- #main-room Slack open for questions
- Start to watch the first video (Introduction) [you can speed up to 1.25x]
- Reconvene for questions in 15 min

# Numba

- Watch numba 0 : basics, 10 minutes

<https://www.youtube.com/watch?v=-aUkLZmrasA>

- Watch numba 1 : numpy, 10 minutes

<https://www.youtube.com/watch?v=ET372Rq1i8I>

# Numba

- Watch numba 2: threads, 10 minutes

<https://www.youtube.com/watch?v=Tfaoy6x2CJg>

- Watch numba 3: groupby pixels, 10 minutes

<https://www.youtube.com/watch?v=4VxHd2qwkro>

# Dask Tutorial

- Dask 1 delayed: 15 minutes

[https://www.youtube.com/watch?v=oaUwrw\\_WDAI](https://www.youtube.com/watch?v=oaUwrw_WDAI)

- Break: 5 minutes

# Outline

- Introduction to Jupyter notebooks and JupyterLab
- Single-node Python code optimization with numba
- Dask tutorial: overlap functions, introduction to dask array, distributed scheduler
- Dask array in-depth tutorial for multi-core, out-of-core, multi-node computing



# Dask Tutorial

- **Dask 3 Arrays: 20 minutes**

<https://www.youtube.com/watch?v=5hH--5EuBek>

- **Dask 5 Distributed: 20 minutes**

<https://www.youtube.com/watch?v=NEHxLjMed7I>

# Outline

- Introduction to Jupyter notebooks and JupyterLab
- Single-node Python code optimization with numba
- Dask tutorial: overlap functions, introduction to dask array, distributed scheduler
- Dask array in-depth tutorial for multi-core, out-of-core, multi-node computing

# Dask array

- Watch dask array 0-2: 15 minutes

[https://www.youtube.com/watch?v=2\\_dbnm6nCk](https://www.youtube.com/watch?v=2_dbnm6nCk)

**Note: We are not doing the multi-node example (setup needed) but the video is available.**