

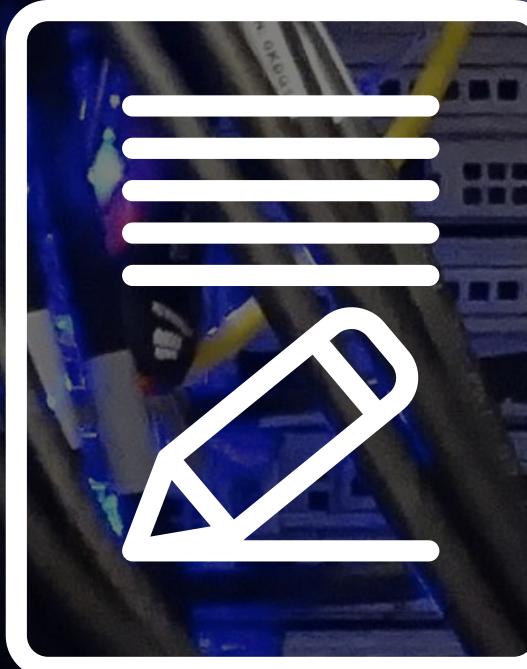


PennState
Institute for Computational
and Data Sciences

Introduction to High-Performance Computing

By Jason C. Nucciarone

Table of Contents



1. What is ICDS?
2. Download example code
3. Foundations of HPC
4. The Software Stack
5. MPI
6. Numba
7. Dask
8. Conclusion

What is ICDS?

- Founded in 2011
 - Originally named ICDS
- Campus Research Community
- Provide various computing services
 - Seed Grants for new research projects
 - Faculty Co-Hires + startup packages
 - Cyberinfrastructure
- Manage the Roar, HPRC, and Roar Collab computing clusters
 - 25 petabytes of data storage
 - 36,500 computing cores
 - 890 gigaflops of performance

Who am I?



Jason C. Nucciarone
High-Performance Computing Software
Engineer/Consultant

Download the example code



[https://github.com/
NucciTheBoss/intro
_to_hpc_hackpsu](https://github.com/NucciTheBoss/intro_to_hpc_hackpsu)

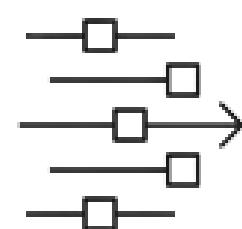
Foundations of HPC

High-Performance Computers (supercomputers)
do what your personal computer cannot

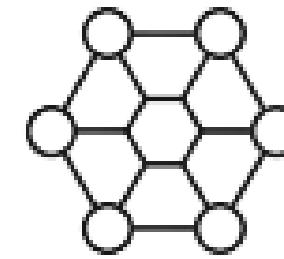


Supercomputers are much better at using
all available resources for running jobs

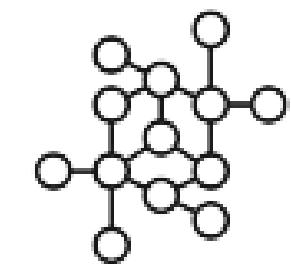
- Jobs -> Computational Tasks
- When there are too many jobs, you need a scheduler
- HPC systems usually have one of four designs



Parallel



Cluster

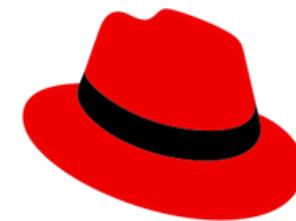


Distributed

- Used for a wide variety of use cases
 - Cancer research, protein analysis, materials development

Software Stack

Operating Systems



Red Hat



ubuntu

CentOS

Applications

Lmod



ANACONDA

[A]
apptainer.org

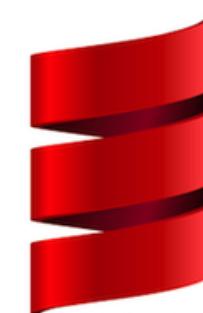
Programming Languages



julia



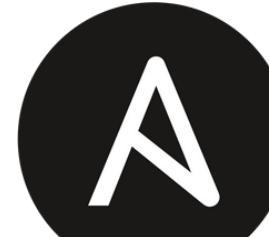
python



Scala



Management Software



ANSIBLE

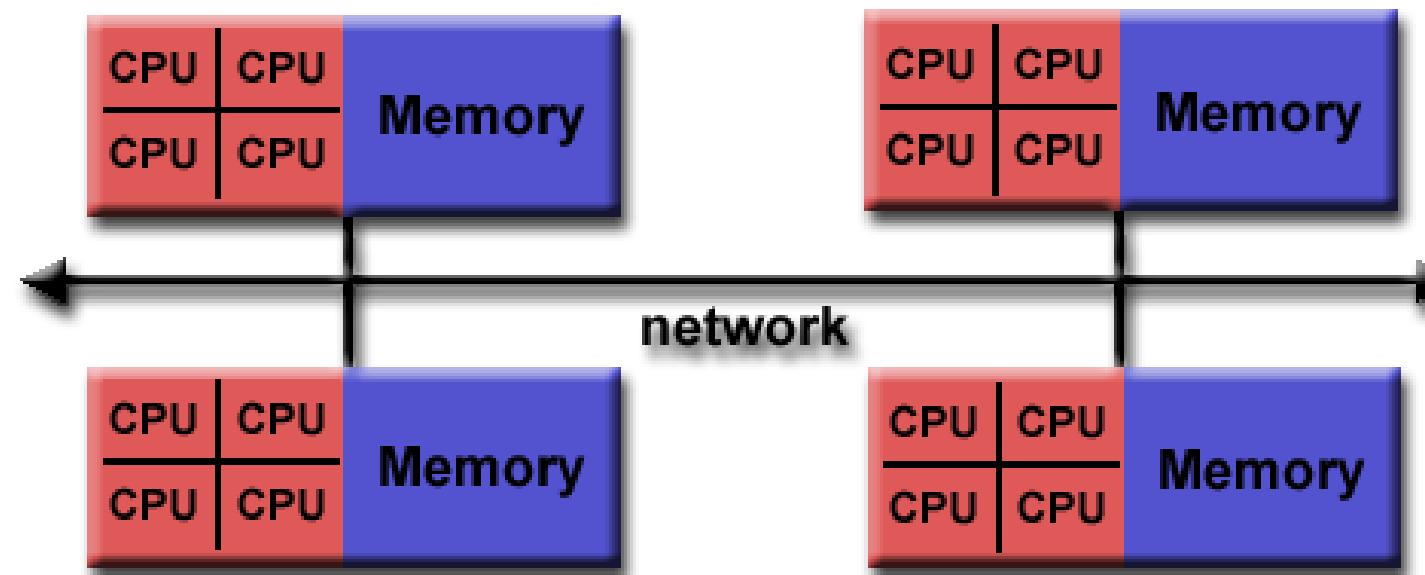


Grafana



Message Passing Interface

- Message Passing Interface is not a library, but a standard for what message passing libraries should be
 - Many implementations:
mpich, openmpi, LLNL MPI



- **OpenMPI**, is an open-source implementation of Message Passing Interface
 - Can write MPI-enabled code using the ***mpi4py*** Python package



Example Time



Numba



- Numba is a high-performance compiler
 - "Python but faster"
- Translates eligible Python blocks directly into machine code
 - Do not need to spend time compiling Python line by line
- Optimized for running Python code on Nvidia GPUs and experimental support for AMD GPUs

Numba is a Just-in-Time compiler for Python

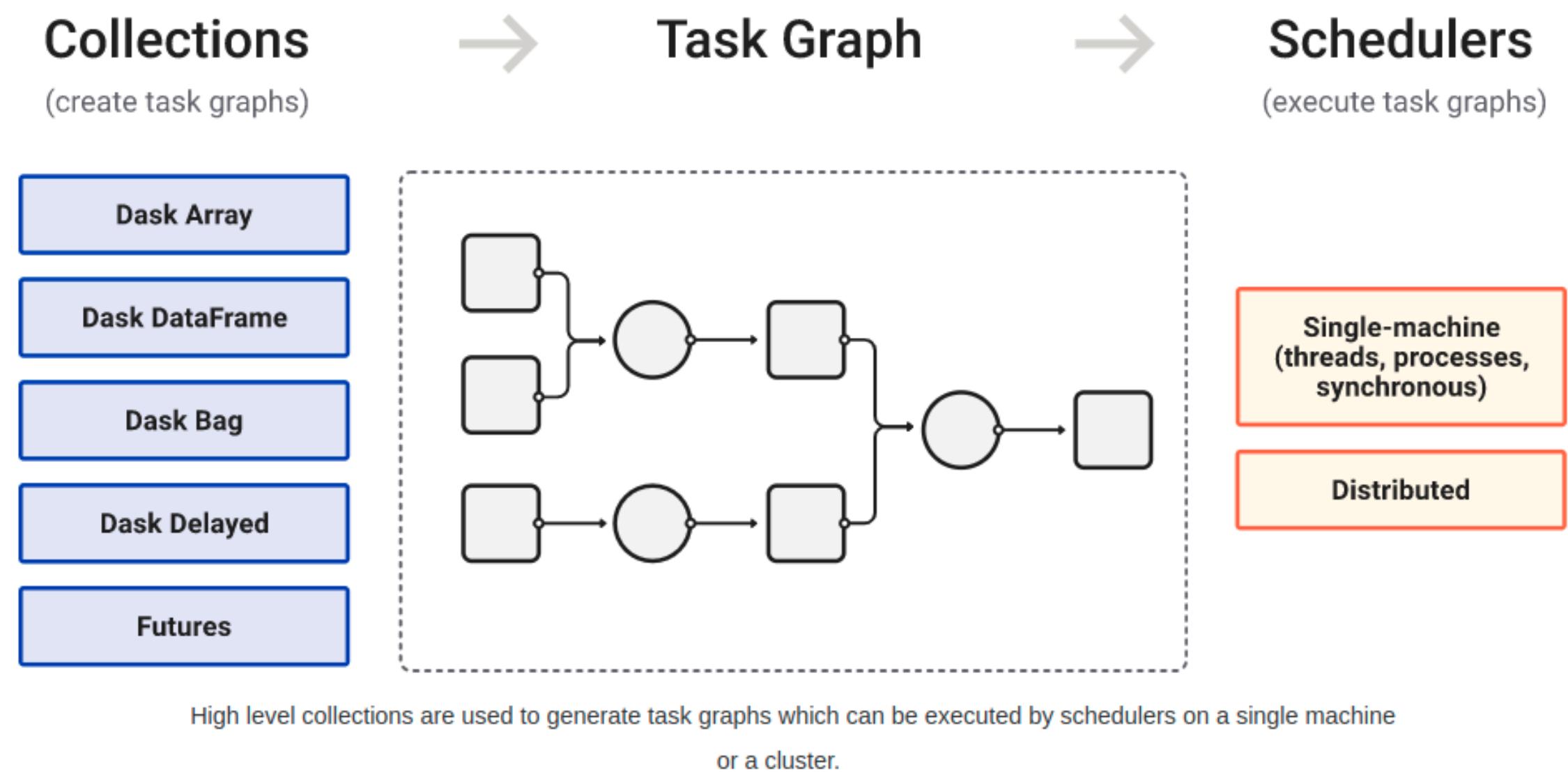
Rather than compiling Python ahead of time, it compiles while the program is running

Example Time

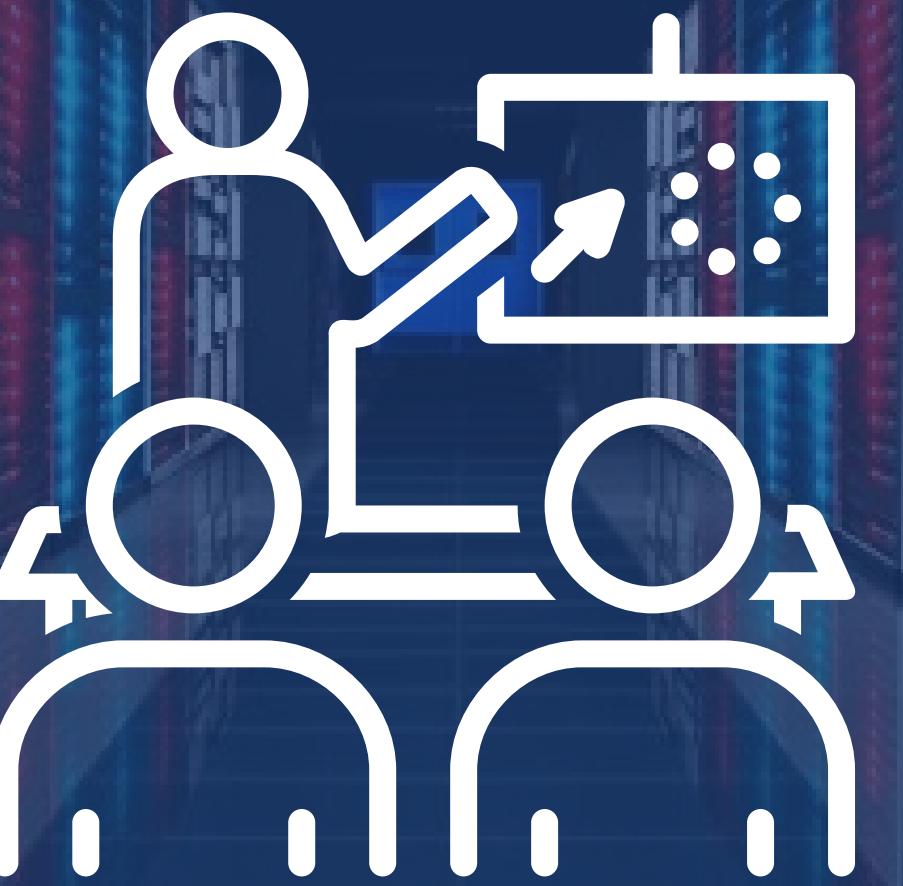


Dask

- Dask is a flexible library for parallel computing in Python
- Composed of two parts
 - **Dynamic Task Scheduling**
 - **"Big Data" Collections**
- Big Data Collections work by loading in chunks of data rather than the whole dataset



Example Time



A photograph of a server room with multiple rows of server racks. The racks are dark-colored and filled with various electronic components. The room has a high ceiling with visible structural beams and lighting fixtures. A large, semi-transparent blue rectangular overlay covers the entire image, containing the text "Questions?".

Questions?