Advanced LaTeX and HPC

Automating research data typesetting pipelines

Ben Keene
Office of Research Cyberinfrastructure
University of Central Florida

October 23, 2025

What is this workshop about?

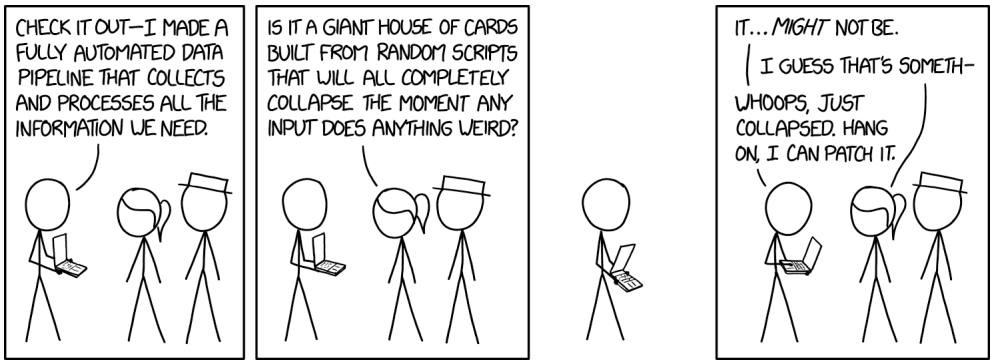


Figure: *xkcd 2054: "Data Pipeline"*

Itinerary

I want you to leave this workshop with a better understanding of HPC workflows for research.

HPC
High-performance computing, GPUs, scheduling. How is it different from regular computing?

ACCESS
NSF-funded, free resource.

Workflows
When is it worth it to automate?

Example
Put it in practice on a cluster.

Who am I?

- PhD student, math dept., ABD
- Research Technology Engineer (Facilitator) @ UCF RCI

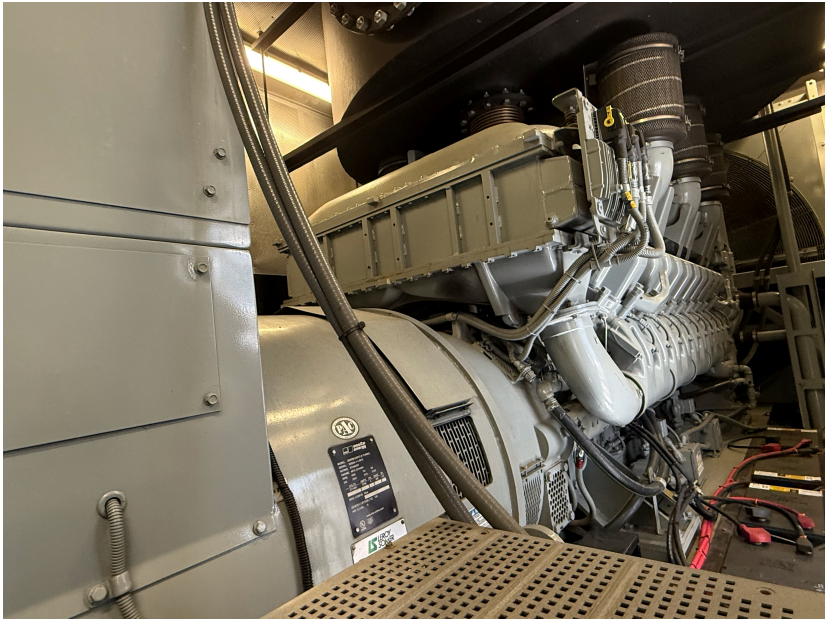
Research Area: Scientific computing, machine learning

HPC

High-performance computing, in the research setting, generally refers to the use of a cluster of powerful computers to solve a computationally intensive problem.

- Scheduled
- Linux based
- Shell scripting and Python are useful


HiPerGator AI
HiPerGator AI
HiPerGator



ACCESS

`https://access-ci.org/`

Free HPC resources for US-based researchers (and PhD students working on their dissertation).

Workflows

How are you going to *wrangle* the machine(s)/systems you will use for **your** research workflow?

Probably something like

run some niche software $\rightarrow$ get some output $\rightarrow$ copy over to LaTeX document

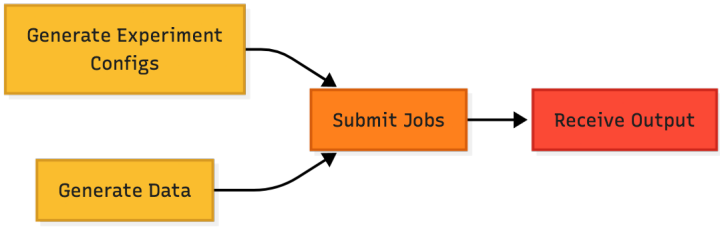
Ideally:

submit job on HPC $\rightarrow$ produce a `.tex` file $\rightarrow$ import in main doc

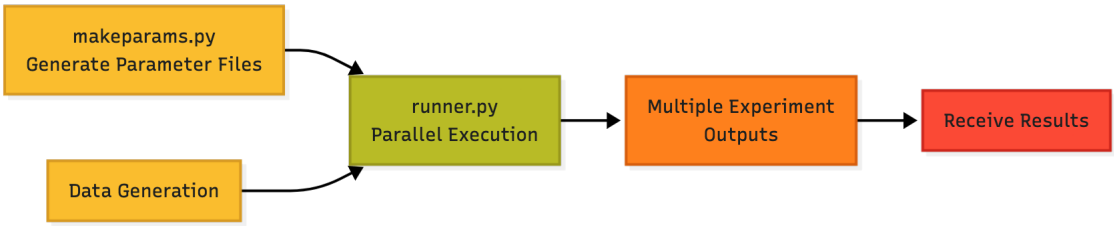
1. Naive Local Workflow

2. Naive HPC Workflow

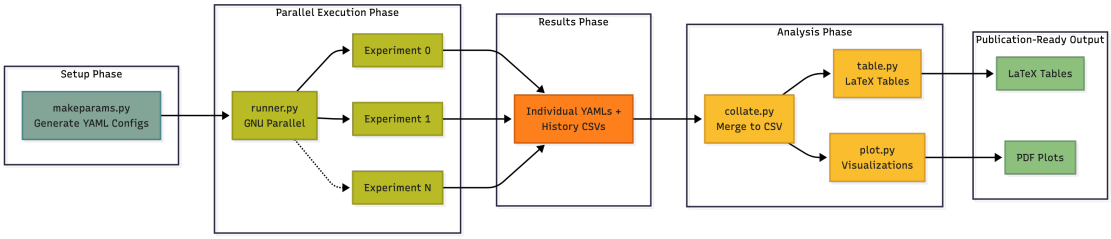
3. Better HPC - Separated Concerns



4. Advanced HPC - Parallel Execution



5. Complete Pipeline - Your Final Workflow



Background

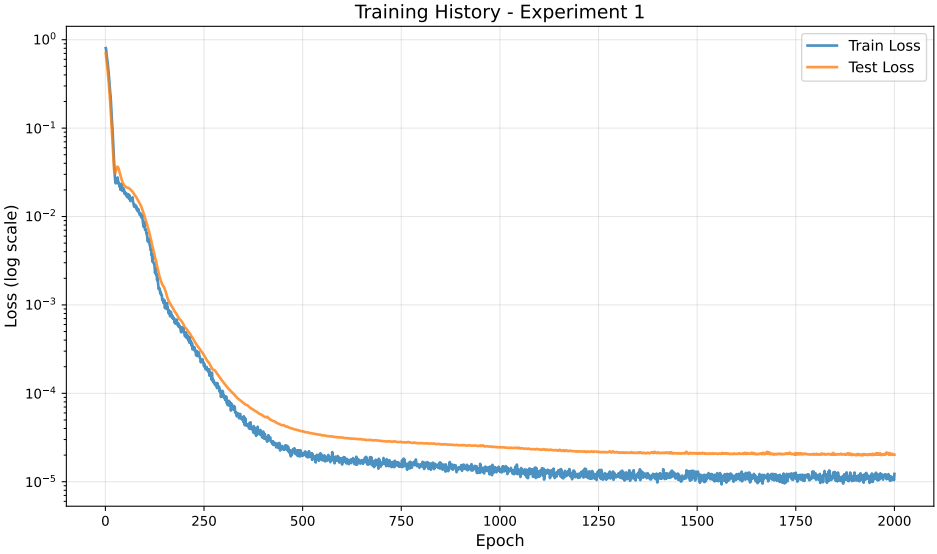
We will train FCNNs (very small and basic neural networks) to approximate the function

$$f(x) = e^{-x^2/2}, x \in [-1, 1]$$

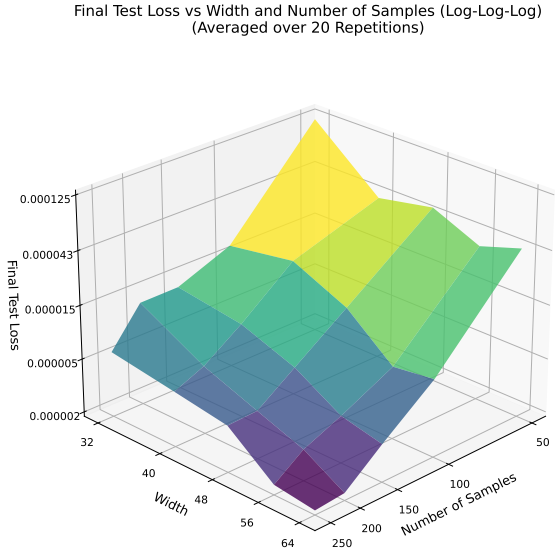
Loosely: We provide the NN input data $x_1, x_2, \ldots, x_n$ and calculate how far off it is from $e^{-x_1^2/2}, e^{-x_2^2/2}, \ldots, e^{-x_n^2/2}$.

The weights of the neural network will be updated to minimize the "loss" (difference between the network and $f(x)$).

Training dynamics



Test loss against parameters



Config Comparison

Table: Top 10 configurations by final test loss

Width	Depth	Samples	Final Test Loss	Final Train Loss	Best Test Loss	Time (s)
64	2	200	1.423e-07	2.074e-07	1.086e-07	11.02
56	2	150	1.706e-07	1.014e-07	1.683e-07	8.61
56	2	250	1.899e-07	7.337e-08	1.280e-07	12.88
64	2	150	1.922e-07	3.139e-07	1.510e-07	8.44
40	2	250	1.973e-07	1.969e-07	1.682e-07	5.49
56	2	250	2.264e-07	1.417e-07	1.393e-07	12.87
64	2	250	2.715e-07	8.749e-07	1.107e-07	12.14
32	2	250	3.064e-07	1.038e-06	2.577e-07	4.93
48	2	250	3.288e-07	4.706e-07	3.288e-07	12.20
56	2	250	3.389e-07	5.689e-07	1.986e-07	12.57

Table: Power-law exponents for test loss vs width and num_samples. Row headers show starting width w, column headers show starting num_samples n. Each cell (w, n) contains two values: the width exponent (calculated from w to next width at fixed n), and the num_samples exponent (calculated from n to next num_samples at fixed w).

w	50	100	150	200
32	-0.31, 0.13	-4.51, -3.09	0.31, 0.16	-1.02, -3.91
40	3.29, -1.23	4.64, -0.43	-1.42, -0.87	0.94, -2.59
48	-7.27, -0.87	-0.93, -3.16	3.31, 0.63	-0.99, -0.34
56	4.24, 0.54	1.19, -1.55	1.20, -1.68	4.68, -3.53

Thoughts and Tips

Dataframes

- `pandas.DataFrame.to_latex`
  Much better than by hand.
- Useful for data analysis, amazing for compiling data.

Containers

- Apptainer is Docker for HPC.
- Define a base Linux image, set up your environment, and load the container at the start of Slurm jobs.

Example

https://github.com/benkeene/hpc-and-latex

Thank you!

Thank you for attending, we will email you a copy of the slides, which contain everything you need.

Please feel free to create an ACCESS account and request to join our workshop project by emailing `ResearchIT@ucf.edu`. Make sure in the subject line you write "HPC and LaTeX Workshop ACCESS". We will let you use this project for a month to play around with the workshop materials.