

# Introduction to HPC



T: +27 51 401 9111 | E: [info@ufs.ac.za](mailto:info@ufs.ac.za) | [www.ufs.ac.za](http://www.ufs.ac.za)

 UFSUV |  UFSweb |  UFSweb |  ufsuv

*Inspiring excellence.  
Transforming lives.*

UNIVERSITY OF THE  
FREE STATE  
UNIVERSITEIT VAN DIE  
VRYSTAAT  
YUNIVESITHI YA  
FREISTATA



**UFS**  
INFORMATION AND  
COMMUNICATION  
TECHNOLOGY SERVICES  
(ICT SERVICES)

# Contents

- Introduction
- Components of a HPC
- Parallelization
- Scalability
- HPC resources at the UFS

# Introduction

# Instructor

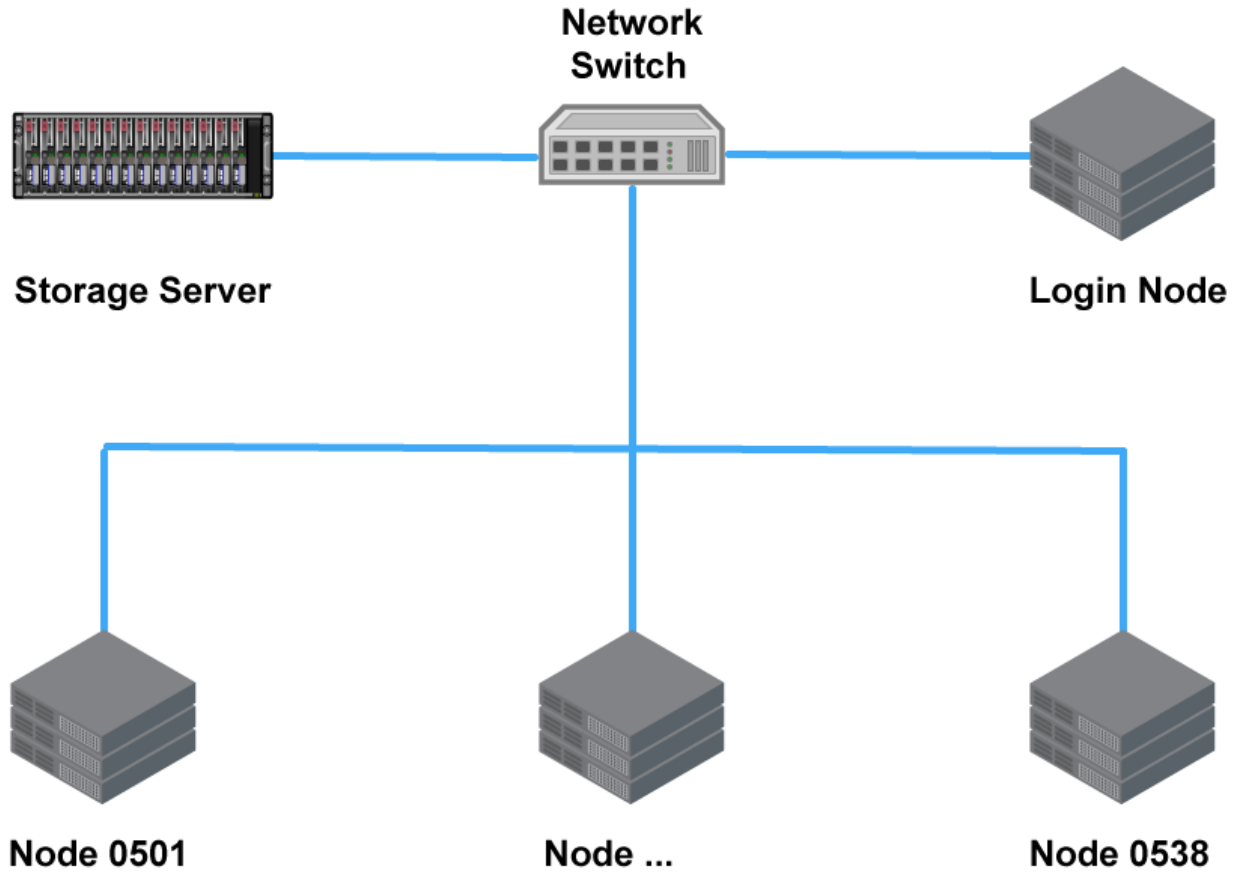
- Louis L. du Preez
- Research Specialist
  - Doctorate in Biochemistry
  - Molecular Modeling
- e-Research and High-Performance Computing
- ICT Services
- [lategandupreezl@ufs.ac.za](mailto:lategandupreezl@ufs.ac.za)

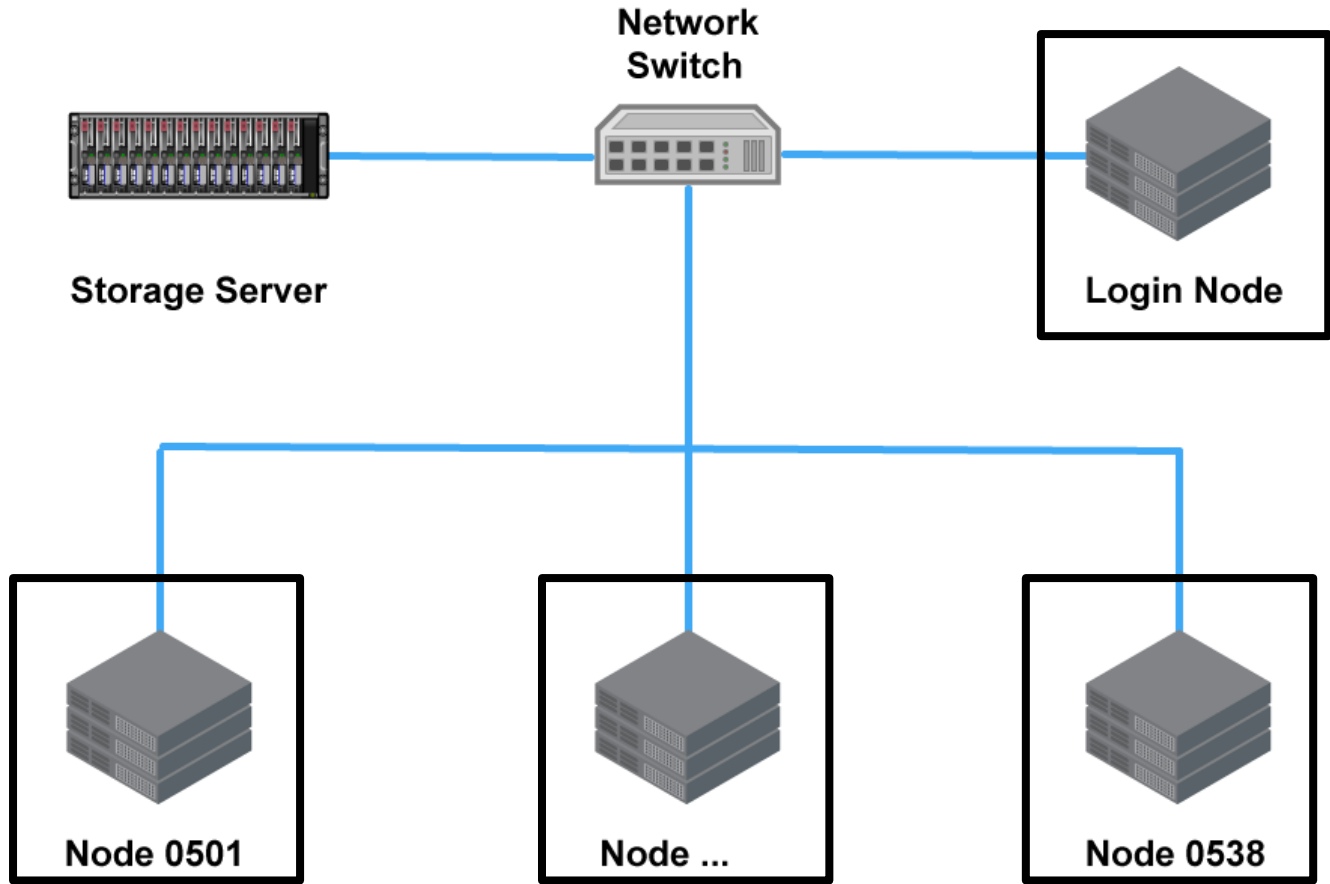


# High-Performance Computing

- Using distributed computing resources to perform computing
- Software and Hardware
- Long history (Since 1960s)
  - See <https://doi.org/10.1145/503124.503129>
- Abbreviation = HPC
- University to the Free State provides and funds HPC via ICT services to accelerate research

# Components of a HPC







# Node

- Basic constituent of an HPC
- Hardware
  - Commodity Server – EPYC/Xeon CPUs, ECC RAM, 10 gigabit LAN
  - Can contain specific accelerators: GPU, Intel MIC, etc
- Software
  - Linux as OS
  - PBS (Portable Batch System)
  - OpenMP, OpenMPI, etc

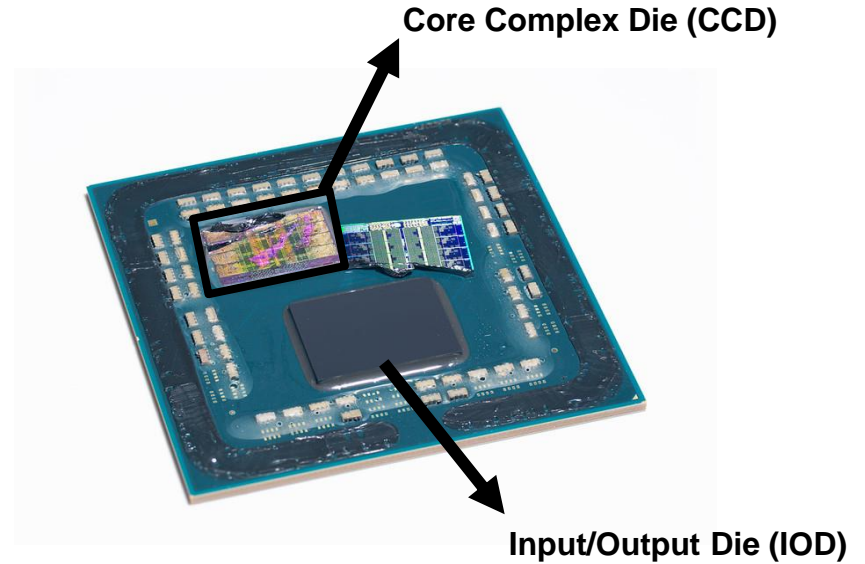
# Node Types

- Login Node
  - Provide entry point for users into the HPC environment
  - No compute activities allowed on this node!
- Compute Node
  - Performs computation
  - Contains a defined number of resources
  - UFS HPC Gen5 compute node (purchased in ~2015):
    - 32 CPU cores ( 2 x 16 core cpus ) / 64 Threads
    - 512 GB RAM

# CPUs: Processors and Cores

- Modern processors = Multi-core processors
- Physical cores
  - Independent processing units
  - Example: AMD Ryzen 5600X : 6 core / 12 thread

# AMD Ryzen 5600X layout





# CPUs: Processors and Cores

- Modern processors = Multi-core processors
- Physical cores
  - Independent processing units
  - Example: AMD Ryzen 5600X : 6 cores / 12 threads
- Threads
  - Hyper-threading (Intel) / SMT effectively splits 1 core into 2 threads
  - Effectively functions as 2 cores but share same resources > performance implications.
  - **Cores, as reported by PBS = Threads / Always double-check**

# GPUs

- Pre-2006, GPUs → special purpose accelerators built to perform specific calculations related to manipulation of 3D/2D graphical images via a fixed rendering pipeline
- Design of GPUs drifted towards more general purpose computation which allowed more flexibility
  - Nvidia releases CUDA with the G80 (Geforce 8000) in 2007
- Typical image manipulation operations → Linear Algebra → Working with matrices.
- General purpose elements allows for application of GPUs to scientific problems

## Accelerating molecular modeling applications with graphics processors

John E. Stone, James C. Phillips, Peter L. Freddolino, David J. Hardy, Leonardo G. Trabuco, Klaus Schulten



First published: 25 September 2007 | <https://doi.org/10.1002/jcc.20829> | Citations: 516



Volume 28, Issue 16

December 2007

Pages 2618-2640



Related



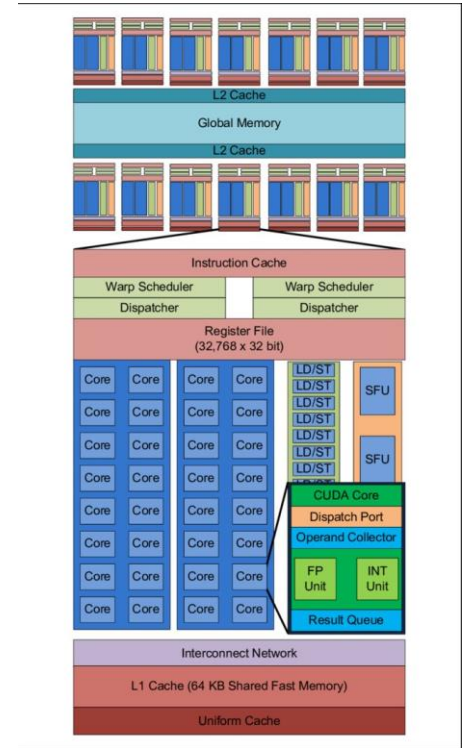
Information

Recommended



# GPUs

- GPU design philosophy → Many smaller “cores” which performs specific operation types.
- I.e. they are not comparable in capability with a traditional CPU core.
  - Streaming processors – terminology differs from vendor to vendor
- Great for computation of problems that use linear algebra – i.e. Matrices and their manipulation, etc
- Example: Machine Learning and A.I. applications



# Types of HPC infrastructures

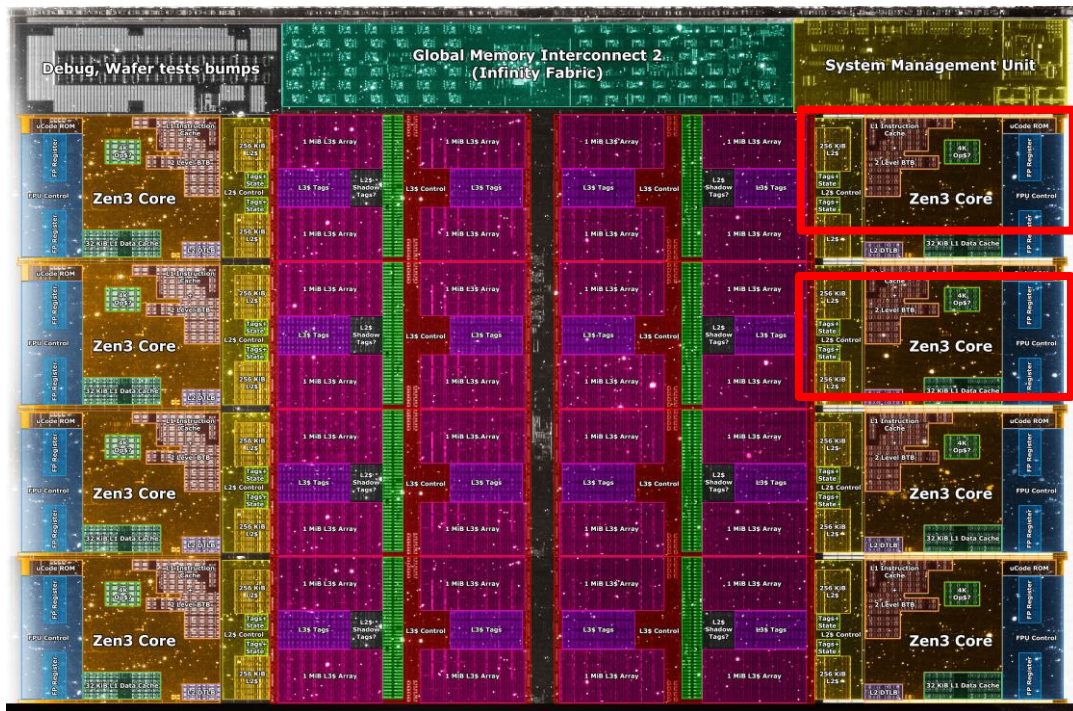
- Homogenous vs Heterogenous
- Homogenous
  - One type of compute node
  - CPU-based – was the standard
- Heterogenous
  - Mixture of types
  - CPU-only / GPU / Intel MIC
  - Recent advances in Machine Learning and A.I. > increase in GPU acceleration capability

# Parallelization

# Parallelization

- The act of performing a task or parts of a task at the same time.
- Example: Making coffee
  - You must make 10 cups of coffee
  - It takes ~ 2 minutes to make 1 cup of coffee
  - Thus –  $10 \times 2 = \sim 20$  minutes to complete the task.
- If you add another person
  - It now takes ~10 minutes.
- How did you parallelize the process? (Important)
  - You split the 10 cups of coffee task into two sub-tasks
  - 2 x (Make 5 cups of coffee)

# AMD Ryzen 5600X CCD



Execute instruction 1

At the same time\*

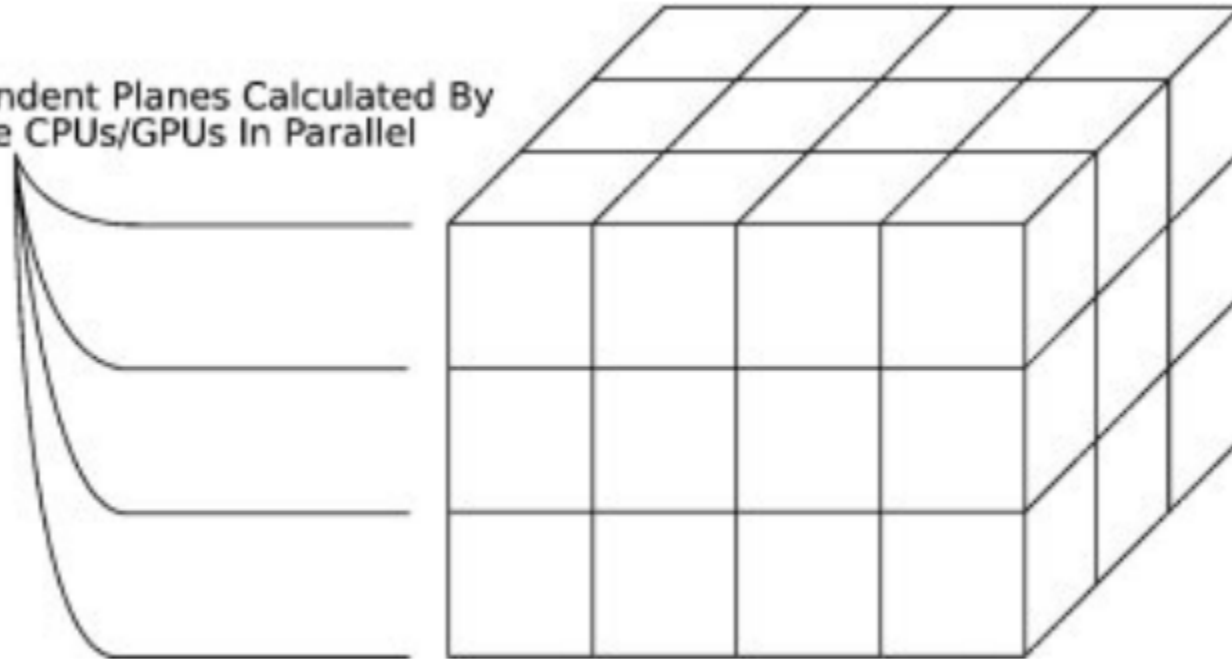
Execute instruction 2

\* Independent Instructions

# Parallelization – In HPC

- Software dependent
  - Be able to split a problem into smaller bits to distribute to the multiple processors

Independent Planes Calculated By  
Multiple CPUs/GPUs In Parallel



**Figure 2.** Decomposition of potential map into slices for parallel computation on multiple GPUs.  $190 \times 91$  mm ( $600 \times 600$  DPI).

# Parallelization – In HPC

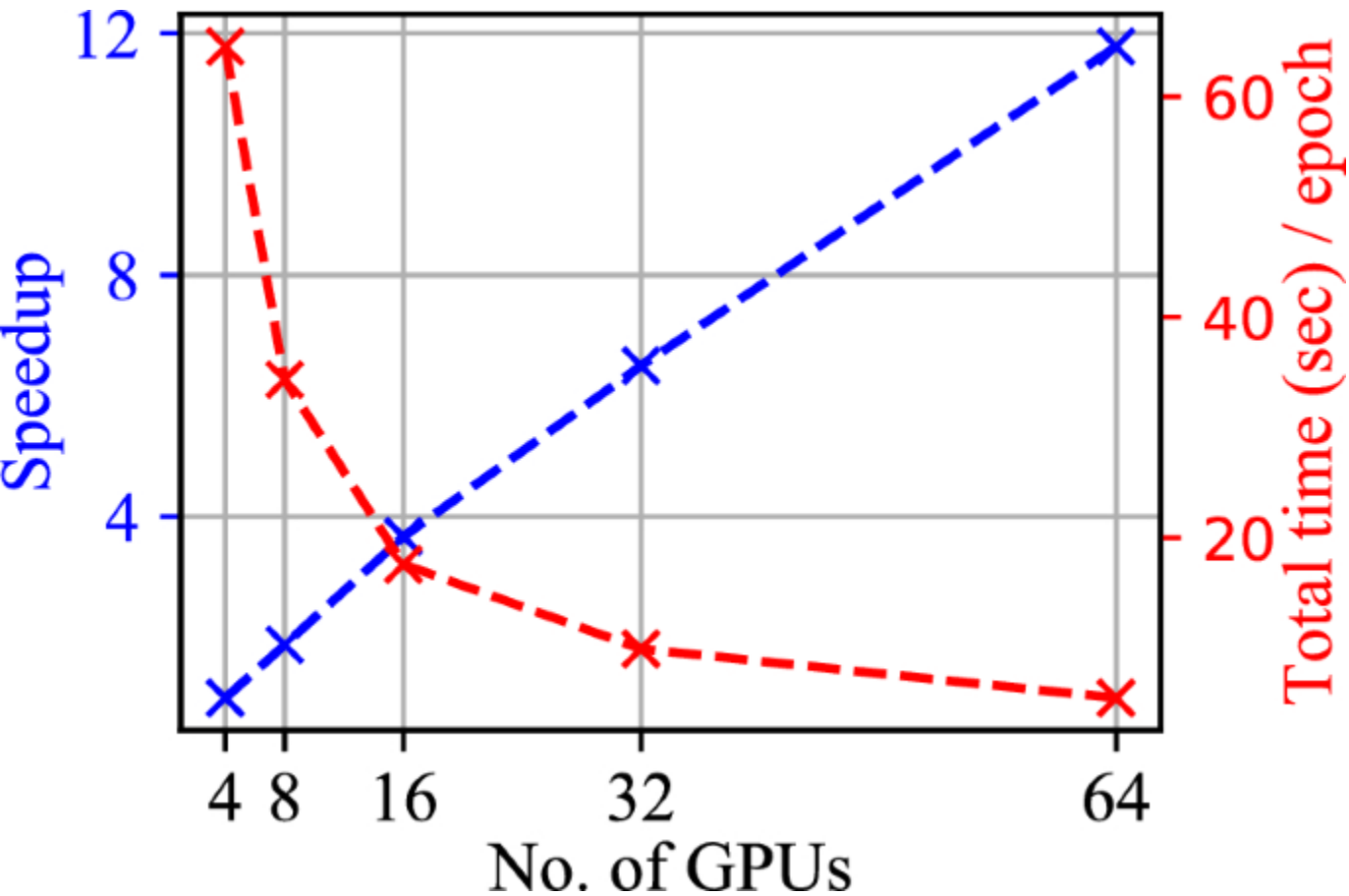
- Software dependent
  - Be able to split a problem into smaller bits to distribute to the multiple processors
  - Software uses :
    - OpenMP (threads)
    - OpenMPI/MPICH (nodes)
    - CUDA (NVIDIA GPU)
  - Nature of problem and how well the problem can be decomposed will affect scalability of the parallelization effort



# Scalability

# Scalability

- How the performance of a workload increases with the addition of computational resources.
- Back to our example of making coffee
  - Adding an additional person (node) → 20 cups of coffee are produced in half the time.
  - Example of linear scaling (2x the resources = 2x the speedup)



- Cosmology with HAL Deep Learning Cluster
- Training step for deep learning model
- Morphological classification of galaxies between spiral and elliptical classes
- Single V100 GPU → 2.1 hours
- 64 x V100 GPUs → 2.7 minutes

# Additional Resources

- UFS HPC:
  - <https://docs.hpc.ufs.ac.za>
- In-depth course on HPC (University of Iceland):
  - <https://www.youtube.com/watch?v=SH7qhC1tJmA&list=PLmJwSK7qduwVnIrlPjrfSn7QRcv3wIQj5>