Kilimanjaro Clinical Research Institute (KCRI)
Moshi, Tanzania

Patrick A. Kimu

Bioinformatician (KCRI)

MSc student in CS at UDSM

Presentation Outline

Patrick A. Kimu (Msc. CS)
Kilimanjaro Clinical Research Institute

- Background
- Our HPC Infrastructure
- HPC org… - Data flow, computation and storage
- How the HPC Works
- Current challenges
- Training Objectives - DTU
- Training Achievement

My Background

- KCRI
  - Bioinformatician:
    - To develop, maintain and upgrade Bioinformatics workflows
    - To support HPC operations
    - To perform data analysis & visualization eg. Great-life & SeqTZ projects

  - A computer science masters student
    - Developing an EWS system for diarrhea-outbreak among U5.

HPC Infrastructure

Servers’ Specs (HPC)

Turing

- CPU = 128
- Ram = 1.96T
- Disks = 8 x 14.6T, 1x931.5G, 1x349.3G, 2x894.3G,2x 2.9T.

Total turing storage size = 125.59746T

Babbage

- CPU= 48
- Ram=376GB
- Disks=238.5G, 931.5G, 24x14.6T, 93.2G, 2x447.1G

Total Babbage storage size = 352.51T

Hopper

- CPU= 96
- Ram= 755GB
- Disks= 6x14.6T, 93.2G, 238.5G, 2x2.9T

Total Babbage storage size = 93.724T

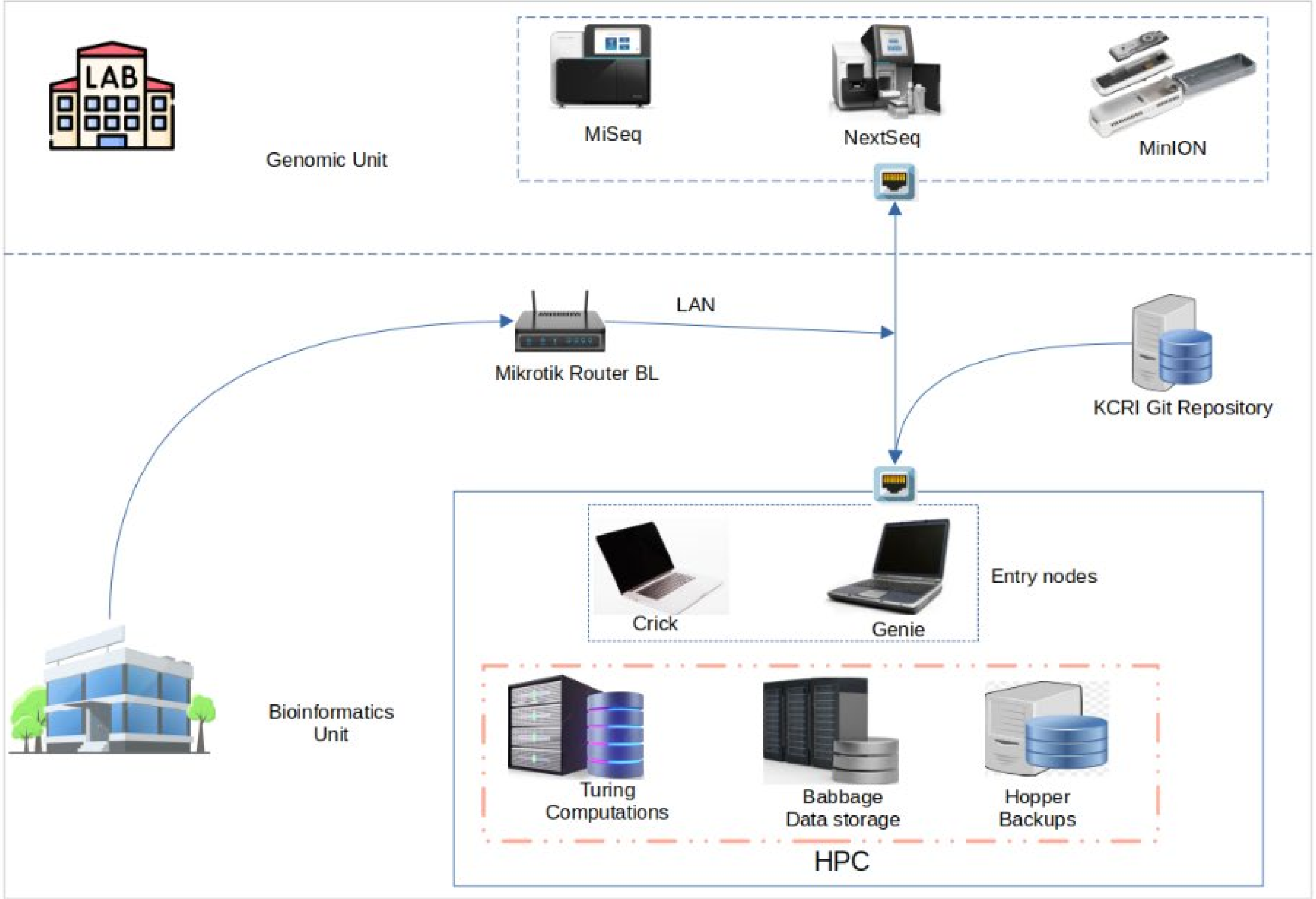
Crick

- CPU= 8
- Ram= 31.2GB
- Disks= 698.6G, 465.8G

Total crick storage size = 1164.4GB or 1.14T

Total HPC storage = 572.94T

Bioinformatics Data Flow Diagrams - review1



How KCRI-HPC Works

- **Login** → SSH is used for logging in via Crick - HPC entry node.
- **Navigating the HPC** → SSH is used to move from one server to another
- **Central directory** → The /hpc/... is shared with all the machines
- **Queuing system** → HtCondor is used to submit jobs.

Challenges

- Lack of Bioinformatics skills such as developing workflows to adapt current and changing research needs.

- Lack of ability to troubleshoot problems in HPC as a result of inadequate configuration and administration techniques. This leads to lack of flexibility and adaptability to new needs.

- Merge all project data overviews into a single overview, so that data visualization can be done.

The objectives of ongoing Training (DTU)

- To learn how to configure, maintain and upgrade HPC (Queuing systems).
- To learn bioinformatics workflows development and implementation
- To learn some system administration aspects like Server configurations especially when deploying new system into server.

eg. Reverse proxy server

Training Achievement

- **Bioinformatics aspect**:
  - Prepare workflow environments or dependencies in conda/mamba
  - to develop and implement a workflow in Snakemake.
- **HPC aspect**:
  - I am grateful of DTU local HPC team for their effort to help me learn everything they do.
  - Unfortunately,we have not done much on HPC as team don’t operate computerome (big HPC), hence they practice basic support services only.
- **Data analysis**:
  - Best data preprocessing practice and analysis

Thank you

HEKIMA NI UHURU

KCRI Kilimanjaro Clinical Research Institute
An Institution of the Good Samaritan Foundation

DANIDA

EDCTP