Welcome to our Viking wiki pages.

We have tried to include as much information about Viking within these pages to help you access and utilise this fantastic resource. If however you find your question has not been answered or you are having issues with Viking please email itsupport@york.ac.uk where someone from our team will be in touch.

To see what queues we have available check here.

Use our visual tool Ganglia to see how busy Viking is.

Introduction

Viking is a large Linux compute cluster aimed at users who require a platform for development and the execution of small or large compute jobs.

Viking is a multidisciplinary facility, supporting a broad spectrum of research needs, free at the point of use to all University of York researchers. Viking is as much a facility for learning and exploring possibilities as it is a facility for running well-established high-performance computing workloads. In this light, we encourage users from all Faculties, backgrounds and levels of ability to consider Viking when thinking about how computing might support their research.

Instructions on accessing the system from Linux and Windows are here Accessing the Servers.
What is a cluster?

A cluster consists of many (100's - 1,000's) rack mounted computers called nodes. It is accessed via login nodes.

Why use a cluster?

- you don't want to tie up your own computer for many hours or days
- you want to run many programs (jobs) at the same time
- you want to use parallelism to obtain your results more quickly
- you need to access more resources (memory, disc space) than is available on your own computer

Limitations of a cluster

Cluster are not the answer to all research computing problems. Some of the limitations are:

- they cannot run Microsoft Windows programs
- they are not suitable for persistent services (web servers, databases)
- jobs that run for many months require special attention

Hardware Infrastructure (September 2018)

Cluster Configuration

- 137 standard nodes
  - 2 cpus per node
  - Intel Xeon 6138 20-core 2.0 GHz (40 cores)
  - 192 GB RAM
- 33 high memory nodes
  - 2 cpus per node
- Intel Xeon 6138 20-core 2.0 GHz (40 cores)
- 384 GB RAM

2 large compute nodes
- 4 cpus per node
- Intel Xeon 6130 16-core 2.1 GHz (64 cores)
- 768 GB RAM

1 very large compute node
- 4 cpus per node
- Intel Xeon Platinum 8160 24-core 2.1GHz (96 cores)
- 1.5 TB RAM

2 GPU nodes
- 2 cpus per node
- Intel Xeon 6138 20-core 2.0 GHz (40 cores)
- 384 GB RAM
- 4 x NVIDIA Tesla V100 32GB SXM2

2 login nodes
- 2 cpus per node
- Intel Xeon 6138 20-core 2.0 GHz (40 cores)
- 192 GB RAM

High speed interconnect
- Mellanox EDR 100 Gb Infiniband with 2:1 over-subscription

High performance filestore
- Lustre filestore
  - 2,556 TB usable capacity
  - sustained 12 GB/sec write performance
- 48 TB of NVME-backed burst-buffer filesystem
  - sustained 18 GB/sec write performance

**Slurm Workload Manager Queues**

<table>
<thead>
<tr>
<th>Queue name</th>
<th>Maximum job time</th>
<th>Nodes available</th>
<th>Total Cores Available</th>
<th>Nodes Config - Cores - Memory</th>
</tr>
</thead>
</table>

In addition to the above there are a number of departmental queues that have a range of job times some up to 80 days in length.

Please note: you should not specify the queue in your job script, the grid engine will select the most appropriate queue for your job.

Please contact andrew.smith@york.ac.uk if you have any comments, or suggestions, on the configuration of the queues.

**Software Infrastructure**


Grid Engine - Slurm Workload Manger 18.08.4 ([https://slurm.schedmd.com/](https://slurm.schedmd.com/))

**Installed Software**

To be updated

**Pictures of Viking**
How the Grid Engine Works - move????