GridION Mk1

Site Installation and Device Operation Requirements

V1 July 2019

GridION IT Requirements

Checklist

This checklist represents the minimal requirements for installing the GridION in your institution. For full explanation of requirements please continue to the document.

Item/Setup Required	Reason	Provided?
1x 1 Gbp/s RJ45 port with DHCP service running	Connection to IT infrastructure and internet	
1x 1 Gbp/s Ethernet cable	Connection to IT infrastructure and internet	
HTTPS/port 443 TCP access to AWS eu-west-1 IP ranges listed at http://docs.aws.amazon.com/general/latest/gr/aws-ip-ranges.html	Telemetry feedback, EPI2ME analysis	
HTTPS/port 443 and HTTP/port 80 TCP access to Linode-hosted IP addresses 178.79.175.200 and 96.126.99.215 for software distribution	Software updates	
USB mouse	Device control	
USB keyboard	Device control	
DisplayPort compatible monitor	Device control	
Power supply capable of delivering 650 W	GridION power	
Storage: *Sufficient infrastructure for required storage option	Long-term data storage	

*Size of storage required will depend on use case. Please refer to the information below for storage guidelines

GridION Device

Overview

The GridION^{TM} is a benchtop device for nanopore sequencing designed to run and analyse up to five MinION or Flongle flow cells or combination of. It is ideal for labs with multiple projects that need the advantages of nanopore sequencing:

- Simple library preparation
- Real-time analysis
- Biological insights from long reads

In addition the GridION also enables users to offer nanopore sequencing as a service when certified.

The GridION benefits from the inclusion of on-board compute which permits data acquisition, analysis and feedback, basecalling, data streaming and device control, all without placing any additional burden on existing IT infrastructure.

 All device control, basecalling, analysis and orchestration on the GridION is carried out by pre-installed custom software created by Oxford Nanopore Technologies. The default data analysis workflow when using the GridION X5 is as below:

Figure 1: Default data analysis workflow of the GridION device

Specifications

The GridION is designed around a simple user interface on top of cutting-edge custom electronics providing real-time analysis solutions:

Component	Specification
Operating System	- Ubuntu 16.04 running off Intel CPU - Customers are advised to keep up to date with all software and security patches
Storage	- 4 TB internal SSD
GPU cards	- 1x Nvidia Volta GV100
Memory	- 64 GB RAM
Size and weight	- H220 x W 365x D 370 mm - 11 kg
Environmental ranges	- Designed to sequence at +18 to +25°C*

*Functional range of electronics -5 to +40°C

Site Requirements

Installing the GridION in your institution is similar to installing any new computer. Requirements for the device are as follows:

Component	Requirements

1x RJ45 Port	- IP address via either DHCP service OR statically assigned
	- TCP running on port 80 HTTP and port 443 HTTPS
	- Requires perimeter firewall permissions:
	   o To access AWS eu-west-1 IP ranges listed at https://ip-ranges.amazonaws.com/ip-ranges.json for telemetry feedback and EPI2ME analysis
	   o To 178.79.175.200 and 96.126.99.215 for software updates
	- Connected to local storage infrastructure

1x Power supply	- 1x country specific C13 cable included with device
	   o Maximum power draw 650W
	   o Maximum current 6 A
	   o Supply voltage 100 – 240 VAC (50/60 Hz)

1x Monitor	- Compatible with DisplayPort connection
1x Keyboard	- Compatible with USB connection
1x Mouse	- Compatible with USB connection

Networking explanations

The two requirements for firewall permissions are:

1. Access to the AWS eu-west-1 region for telemetry feedback and use of the EPI2ME platform through port 443
2. Access to the Linode-hosted IP addresses 178.79.175.200 and 96.126.99.215 for software distribution

Telemetry

Telemetry information is collected by MinKNOW during sequencing runs as per the Terms and Conditions to allow monitoring of device performance and enable remote troubleshooting. Some of this information is obtained from free-form text entry fields, therefore no personally identifiable information should be included. No sequence data is collected.

The EPI2ME platform is hosted within AWS and provides cloud-based analysis solutions for multiple applications. Use of this platform entails users uploading sequence data in .fastq format from the EPI2ME Agent, which is then processed through defined pipelines within the EPI2ME Portal. Downloads from EPI2ME are either in Data+Telemetry or Telemetry form. Telemetry information is used to populate reports within the EPI2ME portal.

Software updates

The IP address from which you receive software updates will depend upon your geographical location. Updates are performed through apt on the terminal, therefore outbound-only access is required. Users are notified of the availability of software updates through the community and full instructions for updating are provided on each release note
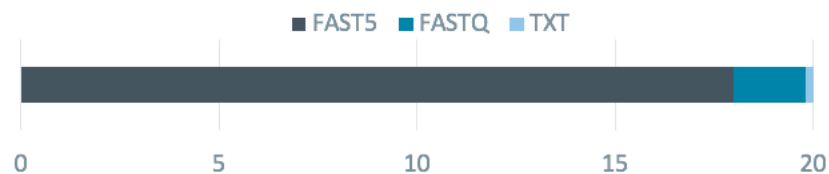
Storage

File types

Nanopore sequencing data is stored in two file types,.fast5 and fastq. Basecalling summary information is stored in a sequencing_summary.txt file:

- .fast5 is a customized file format based upon the .hdf5 file type, which is designed to contain all information needed for analyzing nanopore sequencing data, including raw signal data, and tracking it back to its source. As default each .fast5 file will contain 4000 reads although this can be configured when starting a run. The size of a .fast5 file size will vary dependent on number of reads contained and the length of DNA/RNA sequenced. See the table below for estimates of storage space required for a typical sequencing run

- .fastq is a universal text-based sequence storage format, generated when the nanopore signal data is basecalled and containing both the sequence of DNA/RNA and its quality scores. By default the device saves up to 4000 sequences in one .fastq file. The size of a .fastq file will vary

dependent on number of reads contained and the length of sequence. See the table below for estimates of storage space required for a typical sequencing run

- sequencing_summary.txt contains metadata about all basecalled reads from an individual run. Information includes read id, sequence length, per-read qscore, duration etc. The size of a sequence summary file will depend on the number of reads sequenced

Typically 1 Gbase of sequence data takes up approximately 20 Gbytes of storage. This typically comprises of 90% fast5 files, 9% fastq files and 1% sequence summary file.

Example file sizes below are based on different throughputs from an individual flow cell, with a run saving both fast5 and fastq files with a read N50 of 25Kb. The GridION can run 5 up to 5 flow cells simultaneously.

Output (Gbases)	Fast5 storage (Gbytes)	Fastq storage (Gbytes)	Fastq + fast5 storage (Gbytes)
10	160	20	180
15	240	30	270
30	480	60	540

As an experiment progresses .fast5 files are produced for all reads. If basecalling is chosen these reads are utilized by the onboard software (more information below) to generate sequence data which is then stored in .fastq files.

Long-term storage

The GridION has sufficient SSD disk space for multiple runs to be carried out, storing both .fast5 and .fastq data. However it is imperative this data store is cleared regularly in order to prevent successive runs from terminating due to lack of storage space. For this a site must provide storage to transfer data off the device.

The GridION runs on Ubuntu and is able to mount multiple filesystem types. We recommend storage presented as NFS or CIFS. The form (and volume) of data to be stored will depend on customer requirements:

- Storing .fast5 files with raw read data in will permit re-basecalling of data when new algorithms are released by Oxford Nanopore. In such cases new releases of basecallers have enabled significant improvements in basecalling accuracy of existing datasets through re-basecalling.

Further, selected Oxford Nanopore and third party tools utilize the raw signal information contained within the fast5 to extract additional information from the raw signal e.g modified bases calling, reference-guided SNP calling or polishing of data

- Retaining just .fastq files will allow use of standard downstream analysis tools utilizing the DNA/RNA sequence, but no further sequence data can be generated when improvements in basecalling become available

Included Software

Oxford Nanopore Technologies build and provide numerous software types involved in acquisition, orchestration and analysis:

MinKNOW

MinKNOW carries out several core tasks:

- Device control, including run parameter selection
- Data acquisition
- Real-time analysis and feedback
- Data streaming
- Basecalling (through integrated Guppy)

The MinKNOW software carries out several core tasks: data acquisition, real-time analysis and feedback, basecalling, data streaming, controlling the device, and ensuring that the platform chemistry is performing correctly to run the samples. MinKNOW takes the raw data and converts it into reads by recognition of the distinctive change in current that occurs when a DNA strand enters and leaves the pore. MinKNOW then basecalls the reads, and writes out the data into .fast5 or .fastq files.

Guppy

Guppy is a data processing toolkit that contains the Oxford Nanopore Technologies' basecalling algorithms, and several bioinformatic post-processing features. It is run from the command line but hs a low entry level for use. A selection of configuration files allow basecalling of DNA and RNA libraries, made with Oxford Nanopore Technologies current sequencing kits, in a varied range of flow cells.

If rebasecalling off the GridION it is possible to use an external GPU to basecall with Guppy, and this will dramatically increase basecalling speed over using CPU. Guppy currently works with only NVIDIA GPUs, with the following specific models supported:

NVIDIA Volta GV100

NVIDIA Tesla V100

NVIDIA GTX1080Ti

NVIDIA Jetson TX2

It is possible to use other NVIDIA GPUs for basecalling, but support for those is limited.

EPI2ME

The EPI2ME platform is a cloud-based data analysis service offering a number of workflows for end-to-end analysis of nanopore data. The basecalled read files from a sequencing run can be uploaded to the EPI2ME platform via another piece of software called the Desktop Agent. The user selects a workflow in the Agent such as 16S alignment or Barcoding, and the Agent transfers the read files one by one into the cloud for real-time analysis. The analysis results are generated and updated in real-time in the form of a report in the EPI2ME platform.

Frequently asked questions

Can I use an uninterruptible power supply (UPS) with the GridION?

Yes you can use a UPS however we are unable to recommend a specific product

Do you have any recommendations for how to move data off the GridION in real time during a sequencing run ?

Currently we recommend moving the data off the box using rsync run hourly through crontab. For further details please email support@nanoporetech.com

Change log

Date	Version	Changes made
9Jul 2018	V1	Initial version