

Purpose of System

The system's purpose is to give us a platform for performing memory and processor intensive bioinformatics tasks like denovo assembly and Genotype by Sequencing as well as be a location where the large amount of data can be stored.

Hardware

The system is based on a HPX XS4-2440 system from Thinkmate. It was customised to have a large amount of RAM (96 GB) and Hard drive (4TB) storage. This system has 6x16 GB of RAM for a total of 96 GB. It also has 2 x 4 TB hard drives in a dual RAID 1 configuration. There is a 2TB limit to RAID 1 partitions so the 2 hard drives have two 2TB partitions each which are mirrored to the other drive.

HPX XS4-2440

\$3,292.00 1 \$3,292.00

RAID Configuration

Supermicro SuperWorkstation 7037A-iL - 4x SATA/SAS - 6x DDR3 - 500W [SATA 3, 4 ports] No RAID (*OS) -> 1 x 500GB SATA 6.0Gb/s 7200RPM - 3.5" - Western Digital RE4 WD5003ABYZ

Supermicro SuperWorkstation 7037A-iL - 4x SATA/SAS - 6x DDR3 - 500W [SATA 3, 4 ports] RAID 1 -> 2 x 4.0TB SATA 6.0Gb/s 7200RPM - 3.5" - Western Digital RE WD4000FFYYZ

- Intel C602 Chipset - Dual 1-Gigabit Ethernet - Mid Tower - 500W Single Power Supply
- 2 x Quad-Core Intel Xeon Processor E5-2403 v2 1.80GHz 10MB Cache (80W)
- 6 x 16GB PC3-14900 1866MHz DDR3 ECC Registered DIMM
- 500GB SATA 6.0Gb/s 7200RPM - 3.5" - Western Digital RE4 WD5003ABYZ
- 2 x 4.0TB SATA 6.0Gb/s 7200RPM - 3.5" - Western Digital RE WD4000FFYYZ
- Samsung 24x DVD+-RW Dual Layer (SATA)
- NVIDIA Quadro 410 512MB DDR3 (1xDVI-DL, 1xDP)
- Logitech Keyboard K120 (USB)
- Logitech B100 Optical Mouse (USB)
- No Windows Operating System (Hardware Warranty Only, No Software Support)
- Ubuntu Linux 14.04 LTS Desktop Edition (No Media) (Community Support) (64-bit)
- Thinkmate Three Year Depot Warranty (Return System to Depot for Repair)

Barebone

Memory Technology	DDR3 ECC Reg
North Bridge	Intel C602
Form Factor	Mid-Tower
Color	Black
Memory Slots	6x 240-pin DIMM Socket
Audio	RealTek ALC889 7.1 High Definition Audio with S/PDIF header
Ethernet	Intel 82574L Dual Port Gigabit Ethernet
Power	500W AC power supply w/ PFC
External Bays	4x 3.5" Hot-swap SAS/SATA Drive Bays
Expansion Slots	1x PCI Express 3.0 x16 3x PCI Express 3.0 x8 1x PCI Express 3.0 x4 (in x8) 1x PCI-32
I/O	Nuvoton W83627
Front Panel	Power On/Off button Power LED Hard drive activity LED Network activity LED System Overheat LED
Dimensions (WxHxD)	7.6" (193mm) x 16.7" (424mm) x 20.68" (525.3mm)
Gross Weight	40 lbs (18.1 kg)
SATA 3Gbps AHCI Ports	4
SATA 3Gbps AHCI RAID Levels	0, 1, 5, 10
SATA 6Gbps AHCI Ports	2
Processor	
Product Line	Xeon E5-2400 v2
Socket	LGA1356 Socket
Clock Speed	1.80 GHz

Software

The system is running Ubuntu 14.04 LTS.

vcftools - Filters VCF files to various specifications.

samtools - SNP call program what works with SAM, BAM and outputs VCF files

bowtie2 - Aligns illumina reads to a reference genome and outputs a sam file GBS scripts have been written to automate this task.

R - Used for statistical processes. We use it for linkage analysis with the use of the onemap library.

python - Used for various scripts that analyze or change the format of files.

velvet - Used for denovo assembly of illumina reads, used with oasis

oasis - Used for denovo assembly of illumina reads, used with velvet

Tassel - Alternate GBS pipeline that doesn't require a reference sequence. Also used for Linkage Disequilibrium analysis with data from association mapping.

<http://ngsutils.org/installation/>

FastQC - Used to perform quality statistics on GBS data. Used in the GBS pipelines.

fastStructure - Used for determining the population structure of a data set for Association Analysis.

tophat/cufflinks - This software package does the alignments for RNA-Seq analysis. See <http://cole-trapnell-lab.github.io/cufflinks/>

Tassel Information

<http://www.maizegenetics.net/#!tassel/c17q9>

UNEAK GBS Pipeline

A Script to run the UNEAK GBS pipeline is located on the SVN server at /GBS_Pipelines/trunk/UNEAK_Pipeline.sh.

Linkage Disequilibrium Analysis (GLM and MLM)

There are two tutorials on the Tassel Wiki for the GLM and MLM analysis.

If DArT markers are being used for the sequence data they will need to be converted to the plink format: http://www.gwaspi.org/?page_id=145. The Genetic position in the .map file must be in sequential order. If you don't know the location of the marker then just put 1,2,3,4.... in the column.

Samtools Bowtie Pipeline

Used for aligning GBS data to a supplied reference sequence.

Script locations:

/GBS_Pipelines/trunk/BT_ST_Pipeline_SE.sh

/GBS_Pipelines/trunk/BT_ST_Pipeline_PE.sh

SSH (Terminal Access)

System is designed to be used as a headless so most access will be performed using SSH though linux SSH program or windows PuTTY.

Adding User

The adduser command should be used to create any new users on the system. Home directories are created at /mnt/data2/user_storage.

Giving User sudo privileges

```
sudo usermod -a -G sudo <user>
```

Data Locations

All data is stored on the 4TB drive that are mounted at /mnt/data1 and /mnt/data2. User storage is located at /mnt/data2/user_storage/<user name>

Reference genomes are kept in a central location due to their size:
/mnt/data1/refrence_genomes/. Older versions that were used in analysis are kept but should be compressed when not in use. (tar.gz)

Current GBS analysis files should be kept in /mnt/data1/<project_name> folders. These file get large, all not current projects should be zipped and moved to data2

User files and zipped backup files should be kept in /mnt/data2

Backup Drives

There are two external 2TB Toshiba Canvio USB3 drives connected that are used for data backup. The are mounted at /mnt/backup1 and /mnt/backup2.

****Do to a current bug with Linux USB3 handling they are not detected and mounted on boot. After a reboot of the server they USB need to be unplugged and replugged in for the drives to be recognized and both drives need to be manually mounted by root.**

Commands:

```
sudo mount /mnt/backup1  
sudo mount /mnt/backup2
```

Both drives have different purposes within the system:

/mnt/backup1 is used for manual backups of RAW GBS data. When new data is retrieved from the sequencing labs a copy should be saved in the work directory (/mnt/data1) and in /mnt/backup1/GBS_RAW/ with the barcode and genotype mapping information. Other important data can also be manually backup-ed to this drive.

/mnt/backup2 is used for automatic backups of the /mnt/data2 drive. The automatic backups are performed once per day at midnight by an rsync script that is located in the root home directory. (/root/data2_backup.sh). A log file is also present in this directory that tracks the successful backups. (/root/data2_backup_log). The script is triggered as a cron job.

SVN Server

A SVN server is running on the system. It saves the files to /mnt/data2/svn/.

SNV repositories must be created on the server using the svnadmin command

Users need to be added to /mnt/data2/svn/config/passwd-db

Checkout command: svn checkout svn://<IP>/svn/<repository>

SFTP

File access to the server is done through the sftp server. A Linux computer can use scp or FileZilla. A windows computer can use FileZilla or WinSCP.

NOHUP Usage

The nohup command is used to allow the terminal session to be closed while leaving the task processes running in the background on the server. Some programs and scripts will still error and quit due to X11 dependences.

EDIT: I recommend using tmux to manage sessions rather than nohup

Current User/Pass

Ravi Chibbar

User: ravi

Pass: W@g7NS

Monica Baga

User: monica

Pass: Bet^Ja

Aron Cory

User: aron

Pass: n#JeqP

Craig Irvine (root access)

User: cri646

Pass:

Manu

User: manu

Pass:

Sarita

User: sarita

Pass: heK#\$s