



CENTRE FOR GENOMIC SCIENCES

LI KA SHING FACULTY OF MEDICINE
THE UNIVERSITY OF HONG KONG

香港大學
李嘉誠醫學院
基因研究中心

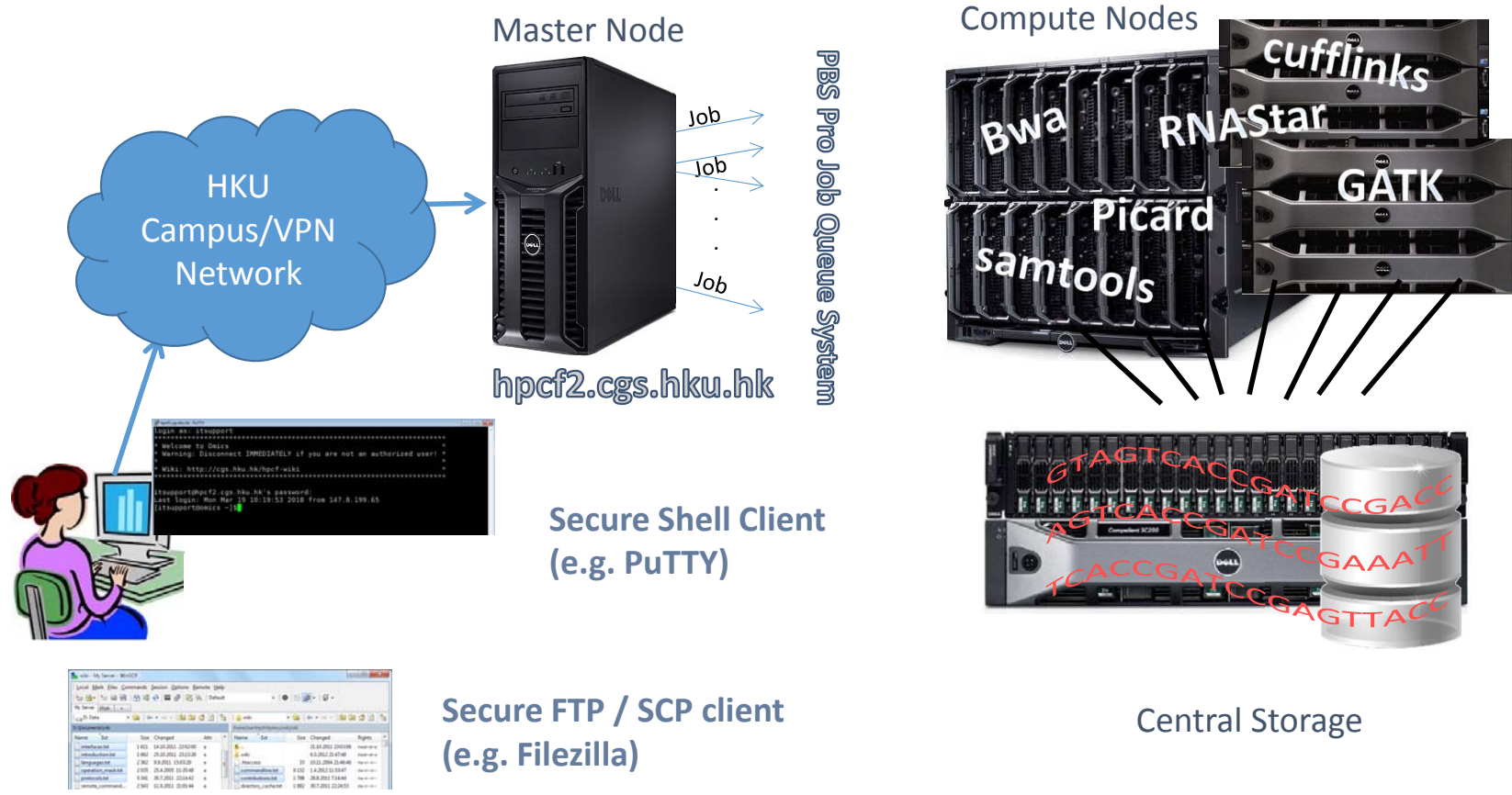
Introduction to High Performance Computing Facility (hpcf2)

27, Mar 2018 (Tue)

10:00am – 12:00pm

Seminar Room 1A, G/F

CGS HPCF2 Architecture



HPCF2 – High Performance Computing Facility

HPCF2 (2018)

- Master Node: hpcf2.cgs.hku.hk
 - Hostname: omics
- 13 (8+5) Compute Nodes
640 cores / 3.6TB RAM
- Operating System: CentOS 7
- Job Scheduler: PBS Pro v14
- Software Management: Environment Modules



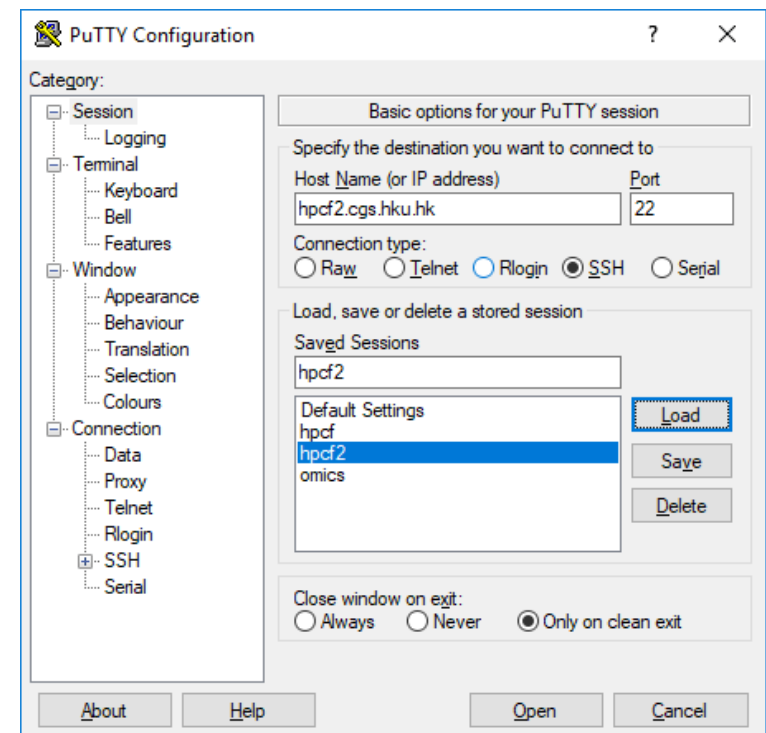
Hardware Configuration of Compute Nodes

Server	CPU Brand & Model	Cores per CPU	No. of CPU	Total Cores per server	NPs in PBS	Total Memory (GB)
hpch01	<u>Intel Xeon E5-2650 v4 2.2GHz</u>	12	2	24	48	256
hpch02	<u>Intel Xeon E5-2650 v4 2.2GHz</u>	12	2	24	48	256
hpch03	<u>Intel Xeon E5-2650 v4 2.2GHz</u>	12	2	24	48	256
hpch04	<u>Intel Xeon E5-2650 v4 2.2GHz</u>	12	2	24	48	256
hpch05	<u>Intel Xeon E5-2650 v4 2.2GHz</u>	12	2	24	48	256
hpch06	<u>Intel Xeon E5-2650 v4 2.2GHz</u>	12	2	24	48	256
hpch07	<u>Intel Xeon E5-2650 v4 2.2GHz</u>	12	2	24	48	256
hpch08	<u>Intel Xeon E5-2683 v4 2.1GHz</u>	16	2	32	64	512
			Total:	200	400	2,304

NP = number of processors

Login account

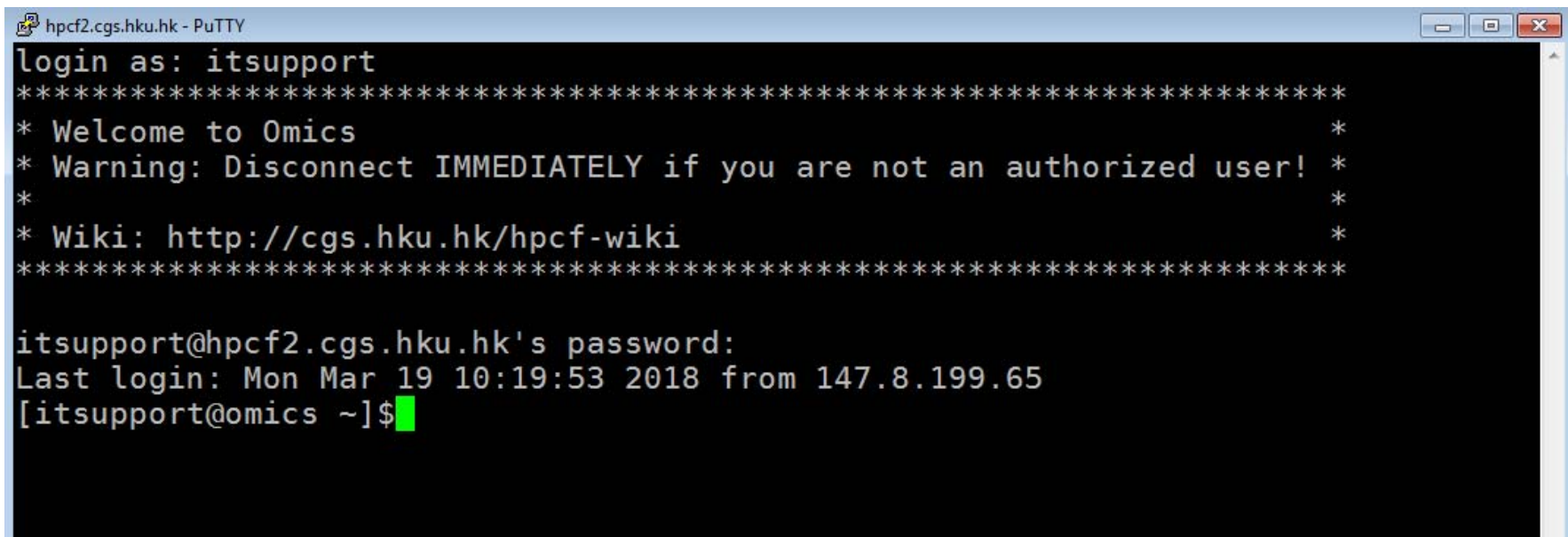
- You must have a login account to use the hpcf
 - Application form at cgs.hku.hk
- Windows PC: download and install “putty” or similar terminal application
 - <https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html>
- Mac: /Applications/Utilities/Terminal



Demo

- Login to Master Node
- Modules
- Job submission
 - Queues – different scenarios of submitting a job
- Databases available (/resources)
- Performance increase
 - GATK as example
- Installing software to personal directory

Login



```
hpcf2.cgs.hku.hk - PuTTY
login as: itsupport
*****
* Welcome to Omics *
* Warning: Disconnect IMMEDIATELY if you are not an authorized user! *
*
* Wiki: http://cgs.hku.hk/hpcf-wiki *
*****

itsupport@hpcf2.cgs.hku.hk's password:
Last login: Mon Mar 19 10:19:53 2018 from 147.8.199.65
[itsupport@omics ~]$
```

Job submission system – PBS Pro

- PBS Pro, a version of PBS (Portable Batch System) for job scheduling in HPCF.
 1. Receive job submission requests from user
 2. Check the computer resources(e.g. CPU, RAM and walltime) requested by the submitted jobs against
 - available resources
 - resource allocation policy
 3. Decide if/where/when the job is executed
 4. Monitor resource usage of jobs and terminate jobs when usage exceeds request

PBS Queues

Queue name	Max. no. of processors per job	Max. memory (GB) per job	Max. no. of job running per user	Max. no. of job queuing per user	Max. Walltime (hr)
small	2	10	18	40	6
small_ext	2	10	6	12	60
medium	12	50	12	25	24
medium_ext	12	50	6	8	60
large	12	120	3	4	84
test	24	190	1	1	1

Submit a batch job

- `qsub pbs_script`

where `pbs_script` is a text file describing the

- resource requests
- commands to be executed

PBS Script

```
#!/bin/bash
#PBS -l selected=1:ncpus=12:mem=40gb
#PBS -l walltime=48:00:00
#PBS -N MyFirstJob
#PBS -m abe
#PBS -q large

module load java/7.0_25
module load GenomeAnalysisTK/3.5
java -jar ...
```

Check job status

- Checking running jobs

qstat -ans

```
[itsupport@omics 2018-03-27]$qstat -ans
```

```
omics:
```

Job ID	Username	Queue	Jobname	SessID	NDS	TSK	Req'd Memory	Req'd Time	Elap S	Time
3954.omics	itsuppor	large	MyFirstJob	13525	1	12	40gb	48:00	R	00:01

hpch01/1*12
Job run at Mon Mar 19 at 15:36 on (hpch01:ncpus=12:mem=41943040kb)

Check job status

- Check queuing jobs

```
qstat -i
```

- Check all jobs in past 5 days (including completed jobs)

```
qstat -xan1
```

- Check details of specific job

```
qstat -xf JobID
```

Kill jobs

- Kill a job (normally works)
`qdel <jobid>`
- Force kill a job (needed for stubborn jobs)
`qdel -W force <jobid>`

qsub options for output/error file path

By default, the standard output and standard error from jobs are kept in the compute node and copied to the master node AFTER the job is complete or aborted

Default path

Standard output:

`<folder_where_qsub_is_submitted>/<jobname>.o<jobID>`


Standard error:

`<folder_where_qsub_is_submitted>/<jobname>.e<jobID>`

Specify `-e` and `-o` qsub options to override default path

e.g. `qsub -o /home/itsupport/1234_T.combined.out -e /home/itsupport/1234_T.combined.err job_script`

Genome databases

-  Some commonly used genome references / databases have been added
- /resources/

Genomes References	Database	Software Bundle Databases
Homo_sapiens	miRBase	ANNOVAR
Mus_musculus	ncRNA	CellRanger
Gallus_gallus	SILVA_DB	GATK_bundle
Equus_caballus		muTect
		oncotator



Quick demo

- **Basic use of module**

- `module avail`
- `module load cutadapt`
- `module help cutadapt`
- `module list`
- `module unload cutadapt`
- `module purge`

- **Sample script modification**

- Add module load commands/remove full path
- Log location

Software Modules: module avail

7-Zip/16.02	HTSlib/1.7	SPAdes/3.10.0	bowtie2/2.2.5
ANNOVAR/2017Jul16	KGGSeq/hg38_2018-01-09	SPAdes/3.11.1 (D)	bowtie2/2.3.4 (D)
BEDTools/2.12.0	MACS/2.0.10-2012.06.06	STAR/2.5.2a	bwa/0.6.2
BEDTools/2.17.0	MACS/2.1.0-2014.06.16	STAR/2.5.3a (D)	bwa/0.7.12
BEDTools/2.27.1 (D)	MACS/2.1.0-2015.04.20	TRF/4.07b	bwa/0.7.17 (D)
BioPerl/1.7.2	MACS/2.1.1-2016.03.09 (D)	TopHat/2.1.1	cutadapt/1.8.1
CNVnator/0.2.7	MUMmer/3.22	TrimGalore/0.4.1	cutadapt/1.15 (D)
Canu/1.5	MUMmer/3.23 (D)	TrimGalore/0.4.5 (D)	idba/1.1.3
Canu/1.6 (D)	NCBI-blast/2.2.27+	Trimmomatic/0.33	java/7.0_25
CellRanger/2.0.1	NCBI-blast/2.7.1+ (D)	Trimmomatic/0.36 (D)	java/7.0_80
CellRanger/2.1.0 (D)	Oncotator/1.9.6.1	VerifyBamID/1.1.2	java/8.0_161 (D)
DESeq2/1.10.1	PEAR/0.9.10	VerifyBamID/1.1.3 (D)	java/9.0.4
DESeq2/1.18.1 (D)	PEAR/0.9.11 (D)	XML-Parser/2.44	miniconda2/4.3.31
EBSeq/1.9.3	Perl/5.26.1	XML-Simple/2.24	muTect/1.1.4
EBSeq/1.18 (D)	Picard/2.0.1	bamUtil/1.0.13	muTect/1.1.5 (D)
FASTX-toolkit/0.0.13.2	Picard/2.17.4 (D)	bamUtil/1.0.14 (D)	plink/1.07
FASTX-toolkit/0.0.14 (D)	QIIME/1.9.1	bamtools/2.3.0	plink/1.90b5.3 (D)
FastQC/0.11.2	QIIME2/2017.12	bamtools/2.5.1 (D)	python2/2.7.14
FastQC/0.11.7 (D)	R/3.2.5	bcl2fastq/2.19	python3/3.6.4
GenomeAnalysisTK/3.5	R/3.4.3 (D)	bcl2fastq/2.20 (D)	samtools/0.1.18
GenomeAnalysisTK/3.7	RNAmer/1.2	bedGraphToBigWig/4	samtools/1.3
GenomeAnalysisTK/3.8 (D)	ROOT/5.34.36	bismark/0.14.3	samtools/1.6 (D)
HOMER/4.9	ROOT/6.12 (D)	bismark/0.19.0 (D)	strelka/1.0.15
HTSeq/0.6.1	RSEM/1.2.31	bowtie/1.0.0	strelka/2.8.4 (D)
HTSeq/0.9.1 (D)	RSEM/1.3.0 (D)	bowtie/1.2.2 (D)	tRNAscan-SE/1.3.1

```
-bash-4.2$ module list  
No modules loaded  
-bash-4.2$
```

```
-bash-4.2$ module load cutadapt  
cutadapt/1.15 is loaded  
-bash-4.2$
```

```
-bash-4.2$ module list  
  
Currently Loaded Modules:  
 1) cutadapt/1.15  
  
-bash-4.2$
```

```
-bash-4.2$ module load cutadapt/1.8.1  
cutadapt/1.15 is unloaded  
cutadapt/1.8.1 is loaded  
  
The following have been reloaded with a version change:  
 1) cutadapt/1.15 => cutadapt/1.8.1  
  
-bash-4.2$
```

```
-bash-4.2$ module load GenomeAnalysisTK
Lmod has detected the following error: Cannot load module "GenomeAnalysisTK/3.8" without these module(s) loaded:
  java/8.0_161

While processing the following module(s):
  Module fullname      Module Filename
  -----
  GenomeAnalysisTK/3.8 /software/Modules/modulefiles/GenomeAnalysisTK/3.8.lua

-bash-4.2$
```

```
-bash-4.2$ module list

Currently Loaded Modules:
  1) cutadapt/1.8.1   2) java/8.0_161   3) GenomeAnalysisTK/3.8

-bash-4.2$
```

```
-bash-4.2$ module unload cutadapt/1.8.1
cutadapt/1.8.1 is unloaded
java/8.0_161 is loaded
GenomeAnalysisTK/3.8 is loaded
-bash-4.2$ module list

Currently Loaded Modules:
  1) java/8.0_161   2) GenomeAnalysisTK/3.8

-bash-4.2$
```

Data backup

- You own the data and are responsible for your own data backup
- No centralized data backup solution but we're looking into the feasibility of it as a future service



Example Usage Email (Daily @ 8am)

jobID	username	queue	jobname	ES	End Time	cpu%	mem%	wtime%
1577.omics	cgs	small	canu_CRE3923	1	2/14/2018 11:11	50	0.38	0.39
1554.omics	cgs	medium	meryl_CRE2564	271	2/14/2018 9:51	91.79	23.25	0.49
611818.statgenpro	cgs	cgs	pbalign	0	11/8/2017 8:20	97.4	50.09	5.15
612062.statgenpro	cgs	cgs	pbalign_contig	0	11/8/2017 19:11	97.42	48.05	6.4
612061.statgenpro	cgs	cgs	pbalign_unitig	0	11/8/2017 17:59	97.78	48.04	5.68
604901.statgenpro	cgs	cgs	pbalign_contig	0	10/31/2017 1:39	96.52	60.43	2.15
605906.statgenpro	cgs	cgs	pbalign_FL	0	10/31/2017 22:11	72.92	0.27	0
604900.statgenpro	cgs	cgs	pbalign_unitig	0	10/31/2017 1:39	96.6	60.43	2.15
594845.statgenpro	cgs	large	STARlong	0	9/28/2017 15:29	13.27	65.8	0.51
594846.statgenpro	cgs	large	STARlong	271	9/28/2017 15:42	71.51	66.44	2.28
594853.statgenpro	cgs	large	STARlong	0	9/28/2017 15:49	13.56	65.8	0.5
594854.statgenpro	cgs	large	STARlong	0	9/28/2017 16:09	51.17	66.46	3.35
525667.statgenpro	cgs	cgs	insertSize	0	4/22/2017 17:21	84.29	26.05	89.65
525668.statgenpro	cgs	cgs	insertSize	0	4/22/2017 12:41	82.73	26.3	76.68
525671.statgenpro	cgs	cgs	insertSize	-11	4/22/2017 21:05	88.95	26.15	100.01
525672.statgenpro	cgs	cgs	insertSize	0	4/22/2017 8:19	83.19	26.1	64.53
525673.statgenpro	cgs	cgs	insertSize	0	4/22/2017 20:53	85.23	26.05	99.48
525674.statgenpro	cgs	cgs	insertSize	-11	4/22/2017 21:05	89.19	26.2	100.01
525675.statgenpro	cgs	cgs	insertSize	0	4/22/2017 19:50	82.44	26.2	86.13
525676.statgenpro	cgs	cgs	insertSize	0	4/22/2017 12:04	80.83	25.95	62.65
525678.statgenpro	cgs	cgs	insertSize	0	4/22/2017 15:23	71.04	25.55	19.65



CENTRE FOR GENOMIC SCIENCES

LI KA SHING FACULTY OF MEDICINE
THE UNIVERSITY OF HONG KONG

香港大學
李嘉誠醫學院
基因研究中心

Questions / Suggestions

Website: <http://cgs.hku.hk/portal/index.php/core-services/hpcf>

Wiki: <http://cgs.hku.hk/hpcf-wiki>

Email: itsupport.cgs@hku.hk