HTCF Basics

CGS_SB High Throughput Computing Facility Spring 2023

Brian Koebbe - Systems Manager Eric Martin - Operations & Systems Specialist Couch Biomedical Research Building - Room 5214 https://htcf-users.slack.com

What? Why?

A cluster is:

Set of computers (servers/nodes/hosts) that work together such that they can be viewed as a single system.

Why a cluster? Scale!

- Biological data is large and complicated.
- Many things done on a personal computer can be done in a fraction of the time on a cluster.

Brian Koebbe, Eric Martin

- Operating System Adminstrators
- HW & SW Architects / Engineers / Technicians:
 - Network
 - Storage
 - Computation
- R&D
- QA
- Support

What we'll cover

Cluster Anatomy

- Server Hardware Anatomy
- Cluster Systems Design/Layout
- How the Systems Work Together
- HTCF Specifics
 - Storage
 - Software
 - Scheduler
- Usage Example
- Q&A

- Processor
 - CPU vs Cores
- Memory
 - Memory vs Storage
- Storage
 - Local Disk
- Network







- Processor
 - CPU vs Cores
- Memory
 - Memory vs Storage
- Storage
 - Local Disk
- Network







- Processor
 - CPU = Cores
- Memory
 - Memory vs Storage
- Storage
 - Local Disk
- Network







- Processor
 - CPU = Cores
- Memory (RAM)
 - Memory vs Storage
- Storage
 - Local Disk
- Network







- Processor
 - CPU = Cores
- Memory (RAM)
 - Memory vs Storage
- Storage
 - Local Disk
- Network





 \exists





- Processor
 - CPU = Cores
- Memory (RAM)
 - Memory vs Storage
- Storage
 - Local Disk
- Network











Node Anatomy

- Processor
 - 24 Cores
- Memory (RAM)
 - 768 GB
- Storage
 - 1 Disk
- Network
 - 2 connections
 - 3 IP Addresses











Node Anatomy: 100+ Servers

.......

5

- Processor
 - 2000+ Cores
- Memory (RAM)
 - 32+TB
- Storage
 - 100 Disk
- Network
 - 200 connections
 - 300 IP Addresses

ش ف	^m ^m	ä ä	m m	شة ۲	ä ä	[*] ⁶	شة ۲	m m	شة ۲	me me
m m	m m	m m	å å	m m	å å	å å	m m ₽	å å	m m	***
m m	°™ °™	m m	å å	m m	å å	å å	m m ⊜	°™ °™	m m	** **
m m		me m	ä ä		ä ä	m m			m m	Å.
			~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~			** ** 8		** ** 8
22222										
m m	** ** 8	m m	å e	* *	å å	** ** 8	** ** 8	** ** 8	** ** 8	** ** 8
m m ₽	***	m m	å e	***	å å		** ** 8	***		***
ř.	ů ř	å å	Å Å	å å	ř.	å å	ů ř	те В	å å	Å Å
ř ř	å å	m m	å B	å å	å å	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	m m	å å	m m	å å
ř.	ř Š	شه ۲	Å B	å å	ř te	å å	å å	شه ۲	å å	شه ۲
					(月月月月月月)	PT (R) (R) (R) (R)				

## LTS Anatomy

- Processor
  - 24 Cores
- Memory (RAM)
  - 196GB
- Storage
  - 12 Disk
- Network
  - 3 connections
  - 4 IP Addresses











## LTS Anatomy: ~35 Servers

- Processor
  - 840 Cores
- Memory (RAM)
  - 6.8 TB
- Storage
  - ~420 disks
- Network
  - 105 connections
  - 140 IP Addresses





 $\Theta$ 

0 660



## HTS Anatomy

- Processor
  - 16 Cores
- Memory (RAM)
  - 256 GB
- Storage
  - 36 disks
- Network
  - 2 connections
  - 3 IP Addresses







## HTS Anatomy: 5 Servers

- Processor
  - 80 Cores
- Memory (RAM)
  - 1280 GB
- Storage
  - 180 disks
- Network
  - 10 connections
  - 15 IP addresses





HTCF



*****

°e'

**e

****

**e

*e

******

~e^

~e^

.....

<u>_</u>e^

6

"e"

*********

.....

*****

-----



****

* #

6

Second International Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second Second	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ
	888888	888888	888888	888888	888888	888888	888888
Absolution     Absolution     Absolution     Absolution       Absolution     Absolut	888888	886888	886888	*****	888888	888888	888888
absabada       absabada <td< th=""><th>888888</th><th>886888</th><th>886888</th><th></th><th></th><th>*****</th><th>888888</th></td<>	888888	886888	886888			*****	888888
SSSSSSS       SSSSSSSS       SSSSSSSS       SSSSSSSS       SSSSSSSS       SSSSSSSS       SSSSSSSS       SSSSSSSS       SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	ൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽ ൽ ൽ	ൽ ൽ ൽ	ൽൽൽ
Normalize         Normalize <t< td=""><td>888888</td><td>888888</td><td>888888</td><td>888888</td><td>888888</td><td>888888</td><td>888888</td></t<>	888888	888888	888888	888888	888888	888888	888888
	886888	888888					
absabada     absab	888888	888888					888888
SSSSSS         SSSSSSS         SSSSSSSS         SSSSSSS         SSSSSSSS         SSSSSSS         <	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ
diamonda     diamo	888888	888888	888888	888888	888888	888888	888888
Abstraction						800000	
abadada     abadada     abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada       abadada     abadada     abadada     abadada     abadada	886888	888888	888888				886888
					******	******	
				ൽ ൽ ൽ 888888	ൽൽൽ		
abashadha     abashadha     abashadha     abashadha     abashadha	888888	888888	888888	888888	888888	888888	88888
	888888	888888	886888	888888	888888	888888	886888
	or or or	<del>იში იში იში</del>		ත්ත ත්ත ත්ත	ಹಾ ಹಾ ಹಾ	ಹೆಂಹೆಂಹೆಂ	ಹಾ ಹಾ ಹಾ
$\begin{array}{c} \mathbf{u}^{\mathbf{u}} \mathbf{u} \\ \vdots \\ $	888888	888888	888888	888888	888888	888888	888888
	886888						

Nodes



LTS

HTS





## What we'll cover

- Cluster Anatomy
  - Server Hardware Anatomy
  - Cluster Systems Design/Layout
  - How the Systems Work Together

## • HTCF Specifics

- Storage
- Software
- Scheduler
- Usage Example
- Q&A

## Storage

- Faster / Slower
- Temporary / Permanent (backups)
- Larger / Smaller
- Good for small files / large files
- High / Limited accessibility
- All are shared



## HTS: /scratch

Temporary working space for jobs.

- Fastest with large reads/writes
- Slow with small reads/writes
- Size: 750TB
  - 84 million files/directories
- Speed: 10+ GB/s
- NOT backed up
- 2TB per account



# LTS: filesystem storage

/lts

For raw data/finished product data

Slower, permanent, backed up

- Accessed via login server
- Size: 4.2PB
- Speed: ~150 MB/s
- Purchased in "buckets"
  - 1TB increments
  - Recommended max 15TB
  - \$7.77/TB/month





# LTS: Object Storage

Like LTS filesystem storage but accessible via the S3 (cloud storage) protocol, not filesystem.

- Accessible
  - Internet
  - Login
  - Nodes
- Size: 4.2PB
- Speed: 300+MB R/W
- No bucket size limit
  - 1TB increments
  - \$7.77/TB/month



*****	<b>~</b> e^	*****	*****	<b>~</b> e^	<b>~</b> e [~]	*****	<b>~</b> e^	<b>~</b> e [~]	*** e**	<b>~</b> e [^]
100304	TRADE I	100304	100304	TAXABLE .	TRADES.	100304	TAXABLE .	TRADES.	10000	10000
111111		111111	111111			111111			111111	
******		******	******			******			******	
all so the	A. A.	the state	the state	A_A	A. A.	the state	A_A	A. A.	she she	A A
	******									
******		******	******			******			******	
~e~	"e"	~e~	~e~	~e^	~e^	~e~	~e^	~e^		~e^
100500	100500	100500	100500	100500	100500	100500	100500	100500	100500	10050
and the second	<b>~</b> .~	10.00	10.00	<b>~</b> _~	<b>~</b> _~	10.00	<b>~</b> _~	<b>~</b> _~	and the second	<b>~</b>
******	******	******	******	******	******	******	******	******	******	******
									-	
	- U -	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>		- U -
100500	122522	101000	101000	100500	100800	101000	100500	100800	101000	12252
the state	n. n	10.00	10.00	n. n.	n	10.00	n. n.	n	the sta	200
******	******	******	******	******	******	******	******	******	111111	
										_
~e~	"e"	~e~	~e~	~e^	~e^	~e~	~e^	~e^		~e^
100504	100500	100500	100500	100500	100500	100500	100500	100500	100500	10050
and the second second	~_~	and the second	and the second	2.0	~_~	and the second	2.0	~_~	and the second	<b>*</b> *
******	******	******	******	******	******	******	******	******	******	
a. a.	0.0	A	A	0.0	0.0	A	0.0	0.0	10.00	0.0
	100800	100000	100000	122222	100500	100000	122222	100500		111111
- m	<b>1</b>	- m -	- m -	<b>1</b> 1	<b>"</b> "	- m -	<b>1</b>	<b>"</b> "	- m -	~n^
100000	100000	Distance.	Distance.	100000	100000	Distance.	100000	100000	TRUNKS.	-
	******			******	******		******	******	******	

ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ
888888	888888	888888	000000	000000	000000	888888
ൽൽൽ	ൽൽൽ	ൽൽൽ	ൺൺൺ	ൺൺൺ	ൺൺൺ	or o
88888	88888	888888	88888			888888
			88888 88888 88888 88888 88888			
ൽൽൽ	ൽൽൽ	ൽ ൽ ൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ
		000000	888888	888888	888888	888888
			888888 886888 886888 886888	888888 888888 888888 888888 888888	888888 888888 888888 888888 888888	888888 888888 888888 888888 888888
an an an	an an an	an an an	an an an	an an an	an an an	an an an
888888	888888	888888	8888888	888888	888888	888888
<u> </u>	2 2 2	<u> </u>	000	000	000	000

## LTS: Object Storage

[janesmith@login ~]\$ s3cmd mb s3://jane1
[janesmith@login ~]\$ s3cmd ls

2023-03-01 12:34 s3://jane1

[janesmith@login ~]\$ s3cmd put somefile.fastq s3://jane1/somefile.fastq

somefile.fastq -> s3://jane1/somefile.fastq [1 of 1]
123456 of 123456 100% in 2s 51.75 kB/s done

[janesmith@login ~]\$ s3cmd put --recursive dir1 dir2 s3://jane1/somewhere/

File	'dir1/file1-1.txt'	stored a	IS	's3://jane1/somewhere/dir1/file1-1.txt'	[1	of	5]
File	'dir1/file1-2.txt'	stored a	IS	's3://jane1/somewhere/dir1/file1-2.txt'	[2	of	5]
File	'dir1/file1-3.log'	stored a	IS	's3://jane1/somewhere/dir1/file1-3.log'	[3	of	5]
File	'dir2/file2-1.bin'	stored a	IS	's3://jane1/somewhere/dir2/file2-1.bin'	[4	of	5]
File	'dir2/file2-2.txt'	stored a	IS	's3://jane1/somewhere/dir2/file2-2.txt'	[5	of	5]





ൽൽൽ	ൽ ൽ ൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൺൺ
888888	888888	888888	888888	888888	888888	888888 888888 8
ಹೆಂಹೆಂಹೆಂ	ൻ ൻ ൻ	ൻ ൻ ൻ	or o	ൽൽൽ	ൽൺൺ	
		88888	88888	888888	000000	888888
കക	க் க் க்	ക്ക്ക്	കകക	க்க்க்	க்க்க்	க்க்க்
888888	888888	888888	888888	888888	888888	888888
				***** ***** ***** *****		
ൽം ൽം ൽം	ofo ofo ofo	ಹೆಂಹೆಂಹೆಂ	or o	000 000 000	<del>იზი იზი</del> ი <del>ზი</del>	oto oto oto
		888888	888888	888888		88888
			888888 888898 888898 888898	88888 88888 88888 88888 88888	88888 88888 88888 88888 88888	
കകക	കകക	കകക	കക	കക	കക	കകക
888888 888888 888888 88888 88888 88888 8888	888888	888888	888888	888888	000000	888888 88888 8

## LTS: Reference

### /ref

# Reference space for software and read-only data sets

- Faster small reads than /scratch
- Slower writes
- Size: 4.2PB
- Speed: 3GB/300MB R/W
- Each groups begins with 1TB
  - 1TB increments
  - \$7.77/TB/month



	°	<b>^</b> 8 <b>^</b>	****	*****	<b>^</b> @ <b>^</b>	<b>^</b> 8 [^]	10 m	<b>^</b> @ <b>^</b>	<b>^</b> 8 [^]	10 m	<b>^</b> 8 <b>^</b>
		TRADE.	100500	100500	TAXABLE PARTY.	-	100500	TAXABLE PARTY.	TAXABLE PARTY.	100500	TAXABLE IN CONTRACT
1	a. a.	A A	also also	also also	A A		also also		A A	also also	
		*****	******	******	*****	******	******	******	******	******	******
1	the state	0.0	10.00	the state	202	n	the state	202	202	10,00	202
		10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
	°0°	<b>"</b> "	°0°	°0°	<b>"</b> "	<b>0</b>	°0°	<b>"</b> "	<u></u>	°0°	<b>"</b> "
		122812	100000	101000	122522	122522	100000	100800	100500	100000	121222
4	a. a.	0.0	10.00	10.00	0.0	0.0	10.00	0.0	0.0	10.00	0.0
		*****		******	******	******			******		******
1	the set	0.0	10.00	the state	202	n	10,00	202	202	10,00	202
		10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
	°	<b>^</b> 8 <b>^</b>	"e"	*****	<b>^</b> @ <b>^</b>	<b>^</b> 8 [^]	10 m	<b>^</b> @ <b>^</b>	<b>^</b> 8 [^]	10 m	<b>^</b> 8 <b>^</b>
		100511			100501	100500		100501	10050		100801
1	a. a.	A A	also also	also also	A A		also also		A A	also also	
		*****	******	******	*****	******	******	******	******	******	******
1	10 M	<b>~_</b> ~	10 at 10	100 at 100	A.A.	<b>^_</b> ^	10 at	2.0	<b>~</b> _~	10 at	<b>~</b> _~
			111111	******			******			******	
		<b>"</b> "	- m -	- B	<b>"</b> "	<b>"</b> "	- m -	<b>"</b> "	<b>1</b>	- m -	<b>"</b> "
		122832			122832	122822					

ൽൽൽ	ൽൽൽ	ൽ ൽ ൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ
888888	888888	888888	888888	888888	888888	888888
ൽൽൽ	ൻൻൻ	ൽൺൺ	ൽ ൽ ൽ	ൺൺൺ	ൺ ൺ ൺ	
	888888					
000	<u> </u>	000	000	000	000	000
888888	888888	888888	888888	888888	888888	888888
000	0 0 0	0 0 0	000	0 0 0	0 0 0	0 0 0
888888	888888	888888	888888	888888	888888	888888
80000	80000	80000	800000	000000	000000	800000
ക്ക്ക്	ക്ക്ക്	ക്ക്ക്	<b>Å</b> Å Å	ക്ക്ക്	കകക	<u>ക</u> ക ക
		888888	886888	886888	886888	886888

# Storage: /home

Personal space for shell configuration files, small scripts

- Slow
- Accessible via login, nodes
- 20GB
- Not meant to handle job workloads





## Software

- Installed and managed by labs
- Can be installed in various ways
  - Software specific instructions
  - Spack/Easybuild
  - R Libraries/Python Modules
  - Conda
  - Compiling source code
- Install Location
  - /ref/<lab>/software/
  - \$HOME (limited space, slow)



Some software installation is easy Much software installation is **hard** 

Management software (spack, conda) may help, but **no** system is foolproof.

Slack, Google, software mailing lists, forums are your friend. We are also here to help.

## Software: Notebook tools

RStudio, Jupyter Lab

- Used on the cluster for interactive manipulation and analysis typically done by batch processing.
- These tools are really web servers.
- Web browser interface requires web access (via SSL tunnels) to nodes.
- Therefore, their use can be complicated and error prone.
- Installation (of RStudio) can be even more complicated and error prone.

Research is being done to see if we can make this simpler.

## Slurm

- Resource management and job scheduling system
- Allocates requested resources
- Provides a framework for starting, executing, and monitoring work
- Arbitrates contention for resources by managing a queue of pending work





## Example Usage

Will use bowie2 to align some sequencing reads to a set of long reference sequences.

- Login
- Grab some cluster resources (CPUs, Memory) with an interactive job
- Use Spack to install bowtie2 into /ref/jslab/software (LTS)
- Download HG19 data into /ref/jslab/data (LTS)
- Make Slurm sbatch script
- Launch job
- Monitor progress
- Look at completed job stats

## Logging in



wustlkey: janesmith lab: jslab WUSTL Key Support: https://wustlkey.wustl.edu

ssh janesmith@login.htcf.wustl.edu
[janesmith@login ~]\$

## Slurm: Start Interactive Job



srun -c 8 --mem=8G -p interactive --pty /bin/bash -l

## Slurm: Start Interactive Job





#### srun -c 8 --mem=8G -p interactive --pty /bin/bash -l



## Install software if needed

[janesmith@n067 ~]\$ spack find bowtie2 ==> No package matches the query: bowtie2 [janesmith@n067 ~]\$

	ൽൽൽ 888888	ൽൽൽ 888888	ൽൽൽ 8888888	ൽൽൽ 8888888	ൽൽൽ 888888	ൽൽൽ 888888	ൽൽൽ 888888
	888888	888888	888888	888888	888888	888888	888888
	ൽം ൽം	ൽ ൽ ൽ	ൽ ൽ ൽ	ൽ ൽ ൽ	ൽൽൽ	or of	ൽ ൽ ൽ
	888888	888888	888888	888888	888888	888888	888888
		888888	888888	888888	886888	886888	888888
660 660				******	******	888888	******
	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽൽൽ	ൽ ൽ ൽ	ൽ ൽ ൽ
		888888			000000	888888	
	<u> </u>	<u> </u>	<u> </u>	000	000		000
		888888	888888	888888	888888	888888	888888
						886888	888888
	<b>იწი იწი იწი</b>	ൽൽൽ	ൽൽൽ	ൽൽൽ	or o	or o	oto oto oto
		8	88888	88888			
			886868	886888	886888	888888	886888

## Install software

[janesmith@n067 ~]\$ spack spec -I bowtie2

_____

Input spec

bowtie2

Concretized

-	bowtie2@2.4.2%gcc@8.5.0 build_system=makefile arch=linux-rocky8-x86_64
[+]	<b>^intel-tbb@2021.7.0%gcc@8.5.0</b> ~ipo+shared+tm_build_system=cmake_build_type=RelWithDebInfo_cxxstd=default_arc
[+]	<b>^cmake@3.24.3%gcc@8.5.0</b> ~doc+ncurses+ownlibs~qt build_system=generic build_type=Release arch=linux-rocky
[+]	<b>^hwloc</b> @2.8.0%gcc@8.5.0~cairo~cuda~gl~libudev+libxml2~netloc~nvml~oneapi-level-zero~opencl+pci~rocm_buil
6_64	
[+]	<b>^libpciaccess@0.16%gcc@8.5.0</b>
[+]	<b>^libtool@2.4.7%gcc@8.5.0</b> build_system=autotools arch=linux-rocky8-x86_64
[+]	<b>^m4</b> @1.4.19%gcc@8.5.0+sigsegv build_system=autotools patches=9dc5fbd,bfdffa7 arch=linux-rocł
[+]	<b>^libsigsegv@2.13%gcc@8.5.0</b>
[+]	<pre>^util-macros@1.19.3%gcc@8.5.0 build_system=autotools arch=linux-rocky8-x86_64</pre>
[+]	<b>^libxml2@2.10.1%gcc@8.5.0</b> ~python build_system=autotools arch=linux-rocky8-x86_64
[+]	<b>^perl@5.36.0%gcc@8.5.0</b> +cpanm+shared+threads build_system=generic arch=linux-rocky8-x86_64
[+]	<b>^berkeley-db</b> @18.1.40%gcc@8.5.0+cxx~docs+stl build_system=autotools patches=26090f4,b231fcc arch=linux-r
[+]	<b>^bzip2@1.0.8%gcc@8.5.0</b> ~debug~pic+shared_build_system=generic_arch=linux-rocky8-x86_64
[+]	<b>^diffutils@3.8%gcc@8.5.0</b> build_system=autotools arch=linux-rocky8-x86_64
[+]	<b>^gdbm@1.23%gcc@8.5.0</b> build_system=autotools arch=linux-rocky8-x86_64
[+]	<b>^python@3.10.8%gcc@8.5.0</b> +bz2+ctypes+dbm~debug+libxml2+lzma~nis~optimizations+pic+pyexpat+pythoncmd+readline
m=ge	neric patches=0d98e93,7d40923,f2fd060 arch=linux-rocky8-x86_64
[+]	<b>^expat@2.4.8%gcc@8.5.0+libbsd build_system=autotools</b> arch=linux-rocky8-x86_64



## Install software

[janesmith@n067 ~]\$ spack install -j \${SLURM_CPUS_ON_NODE} bowtie2

#### ==> Installing bowtie2-2.4.2-4b5rw7qdsxvipu5rhvcrxdapmsbw6cx4

- ==> No binary for bowtie2-2.4.2-4b5rw7qdsxvipu5rhvcrxdapmsbw6cx4 found: installing from source
- ==> Using cached archive: /ref/htcfadmin/software/spack/spack-0.19.1/var/spack/cache/_source-cache/archive/4c/4cc555eeeeb8ae2d47aaa1551f3f01 0e90f462865027e.zip
- ==> No patches needed for bowtie2
- ==> bowtie2: Executing phase: 'edit'
- ==> bowtie2: Executing phase: 'build'
- ==> bowtie2: Executing phase: 'install'
- ==> bowtie2: Successfully installed bowtie2-2.4.2-4b5rw7qdsxvipu5rhvcrxdapmsbw6cx4
  - Fetch: 0.23s. Build: 2m 50.25s. Total: 2m 50.48s.

[+] /ref/htcfadmin/software/spack/spack-0.19.1/opt/spack/linux-rocky8-x86_64/gcc-8.5.0/bowtie2-2.4.2-4b5rw7qdsxvipu5rhvcrxdapmsbw6cx4 [janesmith@n067 ~]\$

## Download reference sequences

[janesmith@n197 data]\$



Ø

[janesmith@n197 data]\$ wget https://genome-idx.s3.amazonaws.com/bt/hg19.zip --2023-03-09 13:32:11-- https://genome-idx.s3.amazonaws.com/bt/hg19.zip Resolving genome-idx.s3.amazonaws.com (genome-idx.s3.amazonaws.com)... 54.231.132.57, 3.5.29.224, 52.216 .50.1, ... Connecting to genome-idx.s3.amazonaws.com (genome-idx.s3.amazonaws.com)|54.231.132.57|:443... connected. HTTP request sent, awaiting response... 200 OK Length: 3694403333 (3.4G) [application/zip] Saving to: 'hg19.zip' hg19.zip 3.44G 20.0MB/s in 2m 43s 2023-03-09 13:34:54 (21.7 MB/s) - 'hg19.zip' saved [3694403333/3694403333] [janesmith@n197 data]\$ unzip hg19.zip Archive: hg19.zip inflating: hg19.1.bt2 inflating: hg19.2.bt2 inflating: hg19.3.bt2 inflating: hg19.4.bt2 inflating: hg19.rev.1.bt2 inflating: hg19.rev.2.bt2 inflating: make_hg19.sh

## Create myjob.sbatch:



#!/bin/bash

#SBATCH -c 8
#SBATCH --mem=8G

```
eval $(spack load --sh bowtie2)
```

bowtie2 -x /ref/jslab/data/hg19 -p \${SLURM_CPUS_ON_NODE} /scratch/jslab/SRR23720981.fq > results.out

## Submit job

[janesmith@login work]\$ sbatch myjob.sbatch





## Monitor job: squeue





## Check efficiency: seff



[janesmith@login ~]\$ seff 3124734 Job ID: 3124734 Cluster: htcf User/Group: janesmith/janesmith State: COMPLETED (exit code 0) Nodes: 1 Cores per node: 8 CPU Utilized: 00:40:32 CPU Efficiency: 95.30% of 00:42:32 core-walltime Job Wall-clock time: 00:05:19 Memory Utilized: 3.21 GB Memory Efficiency: 40.18% of 8.00 GB

## Expectations

- Understand the job
  - What is being done?
  - What's the bottleneck?
  - What files are being read/written?
- Ask the question "Is this polite?"
- When there's a problem, help us help you...
  - Thoroughly describe the issue
    - Job ID
    - Full path to sbatch and files in question
  - Research
    - explain what's been done thus far to find a solution

## What next

- https://htcf-users.slack.com
- We're available for lab meetings
- Future talks?

