

Docker

Contain your Enthusiasm

Michael Chambers, Justin Payne
and some slides from Steve Tsang

Welcome to Dependency Hell!

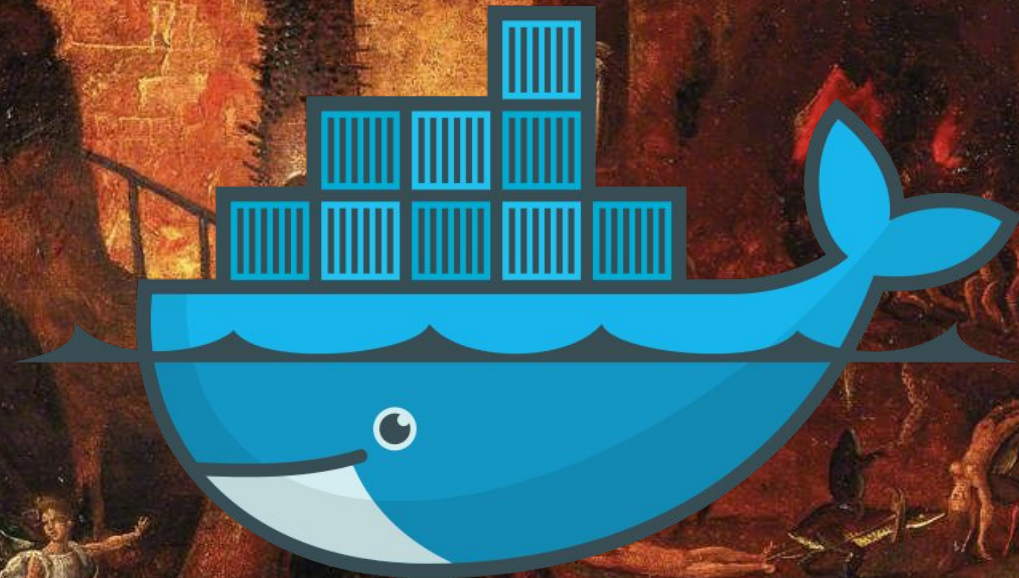
Step 2 - Executable Software Dependencies

You should have the following software installed before using the SNP Pipeline.

Software	Tested Version	Description
Bowtie2	2.3.4.1	A tool for aligning reads to long reference sequences
SMALT	0.7.6	A tool for aligning reads to long reference sequences
SAMtools	1.8	Utilities for manipulating alignments in the SAM format
Picard	2.18.4	A set of tools for manipulating sequence data
GATK	3.8-1-0	Variant discovery and genotyping tools
VarScan	2.3.9	A tool to detect variants in NGS data
tabix	1.8	A generic indexer for tab-delimited genome position files
bgzip	1.8	Part of the tabix package, bgzip is a block compression utility
BcfTools	1.8	Utilities for variant calling and manipulating VCFs and BCFs

Note: the versions above are tested and known to work together. Other versions may also

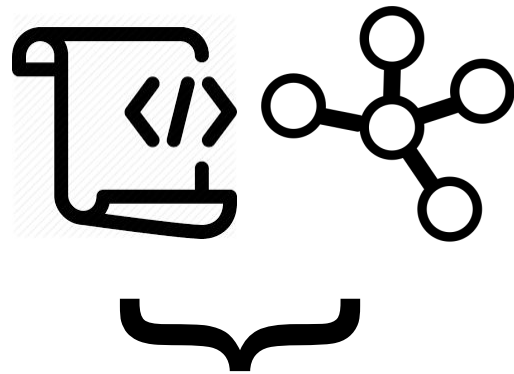
JK!



Images and Containers

Containers are execution environments that are:

- Spun off of images
- Versioned and history-aware
- Built according to a recipe (the Dockerfile)
- Isolated(...) from the host environment
- Portable and shareable



INPUT -----> OUTPUT

Images and Containers - Basic Docker Commands

```
$ docker pull
```

```
$ docker images
```

```
$ docker run <ur_container>
```

```
$ docker run -it <ur_container>
```

```
$ docker container prune
```

Images - Gettin' em

```
$ docker pull crashfrog/snp-pipeline:latest
```

```
justin.payne@hackathon6:~$ docker pull crashfrog/snp-pipeline:latest
latest: Pulling from crashfrog/snp-pipeline
22ecafbbcc4a: Pull complete
580435e0a086: Pull complete
8321ffd10031: Pull complete
08b8f28a13c2: Pull complete
2b401702069a: Pull complete
a3ed95caeb02: Pull complete
eae027dcdc0e: Pull complete
9f92b353dafc: Pull complete
1536e43d28f2: Downloading [----->] 351.2MB/553.6MB
25aacc2428f2: Download complete
759c3a41e312: Download complete
272d2820b6fd: Download complete
22a96ccf86b1: Download complete
307c53996492: Download complete
fc6f87094ae9: Download complete
4e24140ed23d: Download complete
9bde0673352c: Download complete
c4a22c1ea5e2: Download complete
d3eb52ff61db: Download complete
9c9029a2a2c4: Downloading [----->] 19.88MB/137.8MB
becd74bddf6d: Waiting
60786f360e31: Pulling fs layer
```

Images - Listin' em

```
# view all ur docker images  
$ docker images
```


Images - Runnin' em (making a container)

```
$ docker run crashfrog/snp-pipeline:latest -h
```

```
justin.payne@hackathon6:~$ docker run crashfrog/snp-pipeline:latest -h
usage: cfsan_snp_pipeline run [-h] [-f] [-m MODE] [-c FILE] [-Q grid|torque]
                             [-o DIR] (-s DIR | -S FILE) [-v 0..5]
                             [--version]
                             referenceFile
```

Run the SNP Pipeline on a specified data set.

positional arguments:

referenceFile	Relative or absolute path to the reference fasta file
---------------	---

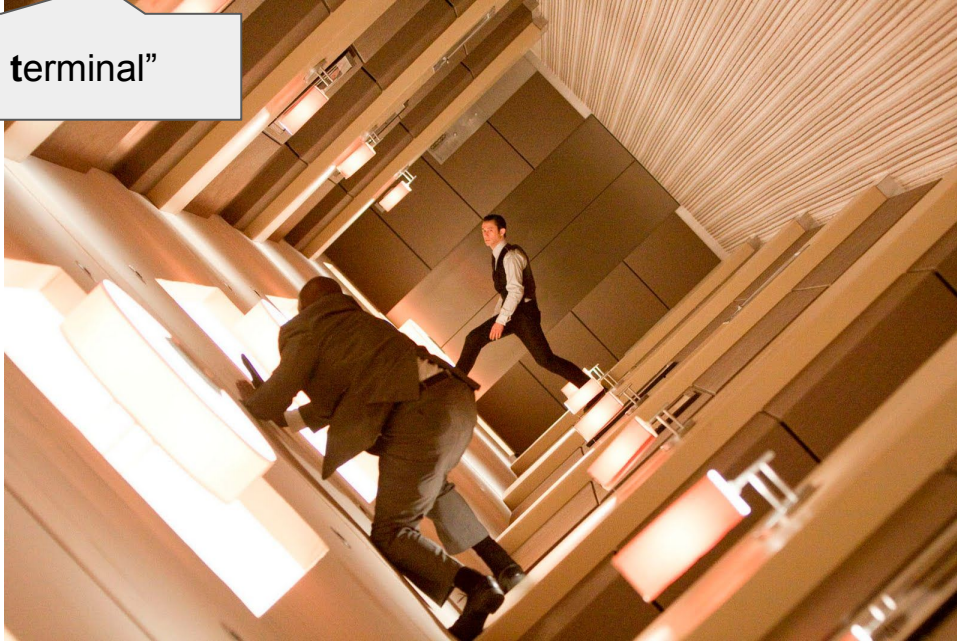
optional arguments:

-h, --help	show this help message and exit
-f, --force	Force processing even when result files already exist and are newer than inputs (default: False)
-m MODE, --mirror MODE	Create a mirror copy of the reference directory and all the sample directories. Use this option to avoid

Containers - Gettin' in em

```
$ docker run -it crashfrog/snp-pipeline:latest
```

“interactive, bind to **terminal**”



Containers - Prunein' em

removes all stopped containers
\$ docker containers prune



Images - The Dockerfile

```
Dockerfile
1 FROM edb/base:1.2
2
3 LABEL maintainer="jonathan.greene@fda.hhs.gov"
4
5 # WORKDIR /tmp
6
7 RUN \
8 yum update -y && \
9 yum install -y wget aws-cli tar bzip2 gzip file && \
10 wget https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86_64.sh && \
11 bash ./Miniconda3-latest-Linux-x86_64.sh -b -p /usr/bin/miniconda3 && \
12 rm -rf ./Miniconda3-latest-Linux-x86_64.sh
13
14 ENV PATH=/usr/bin/miniconda3/bin:$PATH
15
16 RUN \
17 conda install -y -c bioconda -c conda-forge mlst && \
18 pip install -r edb_tools/requirements.txt
19
20 ADD runner.sh /tmp/
21 ADD generate_json.py /tmp/
22
23 ENV EDB_API=https://7v1hjvhai3.execute-api.us-east-1.amazonaws.com/development/db-action
24
25 # ENTRYPOINT /tmp/runner.sh
```

Images - Buildn' em

```
# creates a new image from a Dockerfile  
$ docker build -t name_image path/to/Dockerfile
```


Some Helpful Resources

- [Docker Quick Guide](#)
- [Anatomy of a Dockerfile](#)
- [Docker Cheat Sheet \(.pdf\)](#)
- [Don't forget about Singularity!](#)

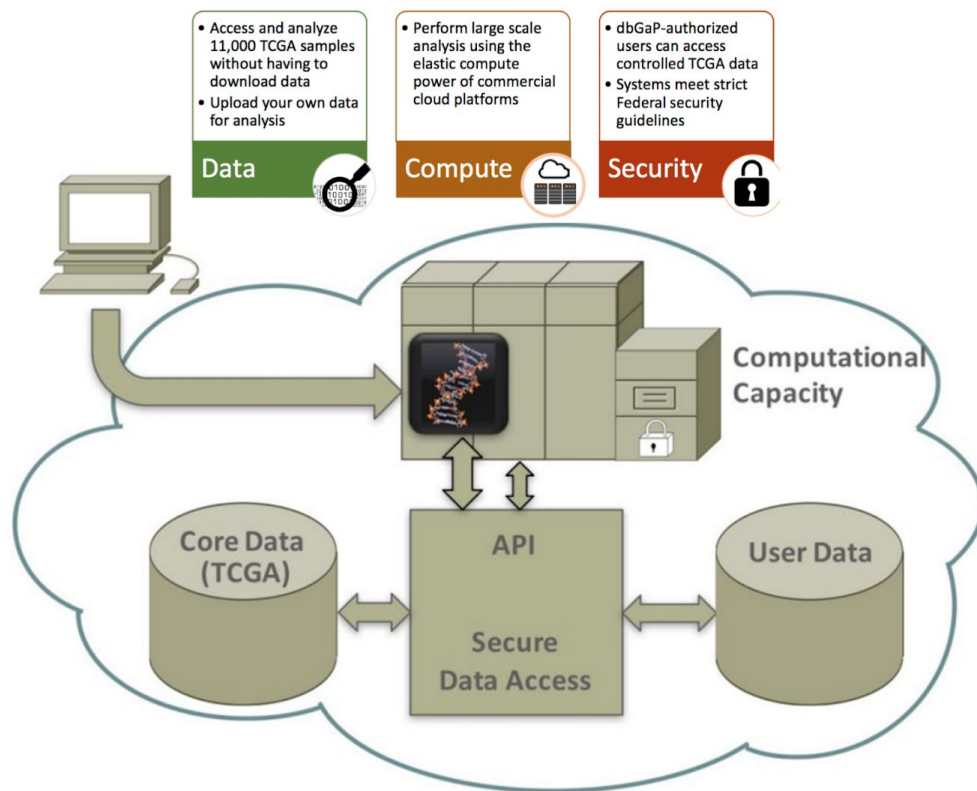
Cloud Resources Concept: Co-located Compute & Data



docker



COMMON
WORKFLOW
LANGUAGE

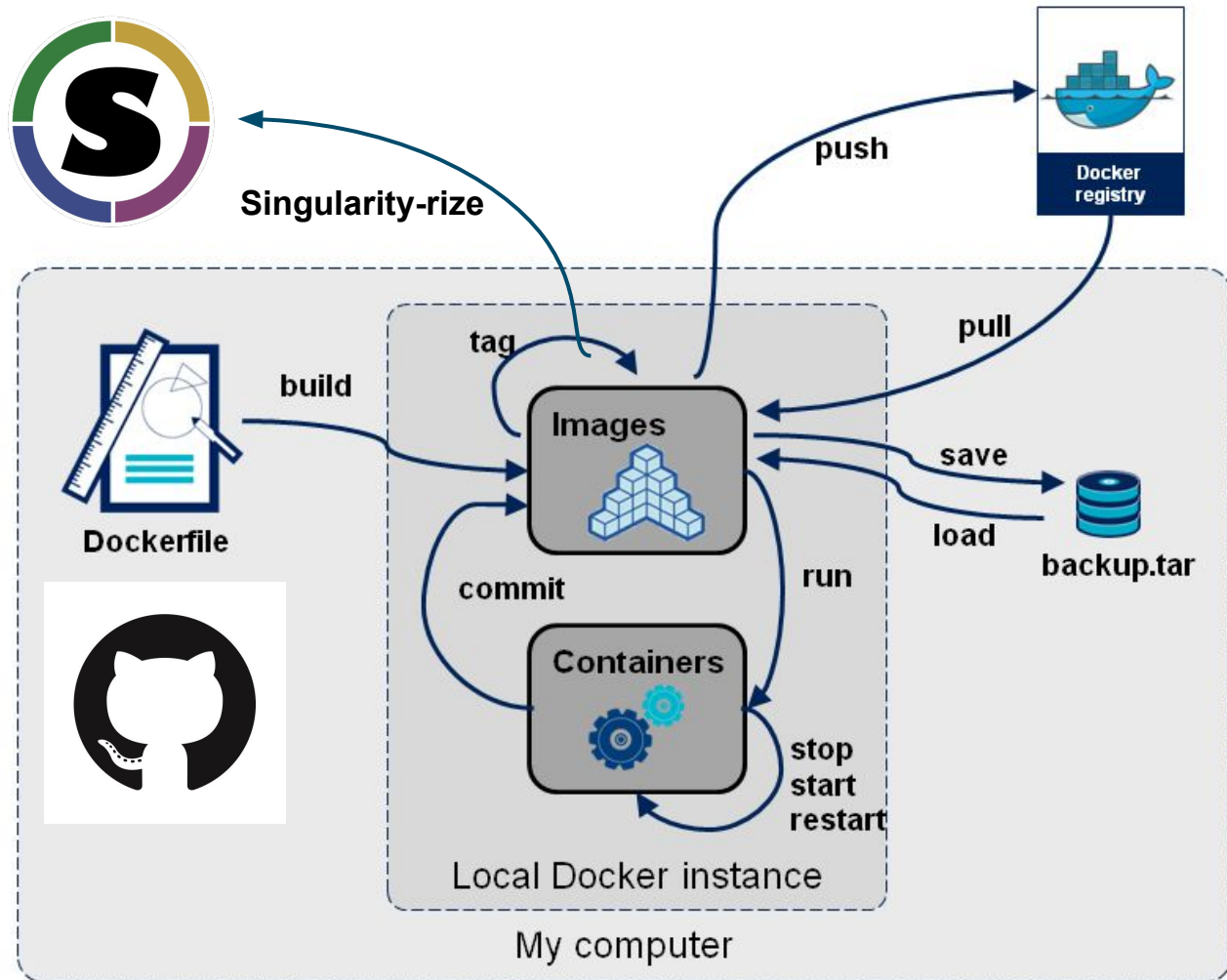


Google Cloud Platform



Democratize access to NCI-generated genomic & related data
Provide cost-effective computational capacity for the cancer research community

Container Concept



Sharing Docker-based Tools

The screenshot shows the Dockstore web interface. The top navigation bar includes links for Tools, Workflows, Search, and Documentation, along with user-specific links for My Tools and My Workflows. The user 'stevetsa' is logged in. A search bar is prominently displayed. On the left, a sidebar contains filters for Search, Entry Type, Language, Author, Workflow: Organization, and Labels. The 'Labels' filter is expanded, showing a list of labels with 'nci' selected and circled in red. The main content area shows a search result for 'nci' with a 'Share' button and a search bar containing 'the Labels is nci'. Below this, there are tabs for 'Browse Tools' and 'Browse Workflows'. The 'Browse Tools' tab is active, displaying a list of tools. The tools are listed in a table with columns for Name, Stars, Author, Format, and Project Links. The tools shown are 'Metaphlan-ISBCGC-Dockstore', 'Metaphlan-SBCGC', and 'Metaphlan-WDL'. The 'nci' label is circled in red in the sidebar, and the search bar in the main content area also contains 'nci'.

Search

Search: the Labels is nci

Browse Tools

Browse Workflows

```
# run it locally with the Dockstore CLI
$ dockstore tool launch --entry quay.io/collaboratory/dockstore-tool-bamstats:1.25-6_1.0 --json Dockstore.json
# run WDLs directly on the FireCloud
```

Show 10 entries

Name	Stars	Author	Format	Project Links
Metaphlan-ISBCGC-Dockstore		Steve Tsang	CWL	GitHub
Metaphlan-SBCGC		Steve Tsang	CWL	GitHub
Metaphlan-WDL		n/a	WDL	GitHub

Showing 1 to 3 of 3 entries

Previous 1 Next

