

# Embedding Docker

Seattle SnowHack2019 at Fred Hutch (2/4-6)

Team Members:

Aakash Sur

Dev Nambi

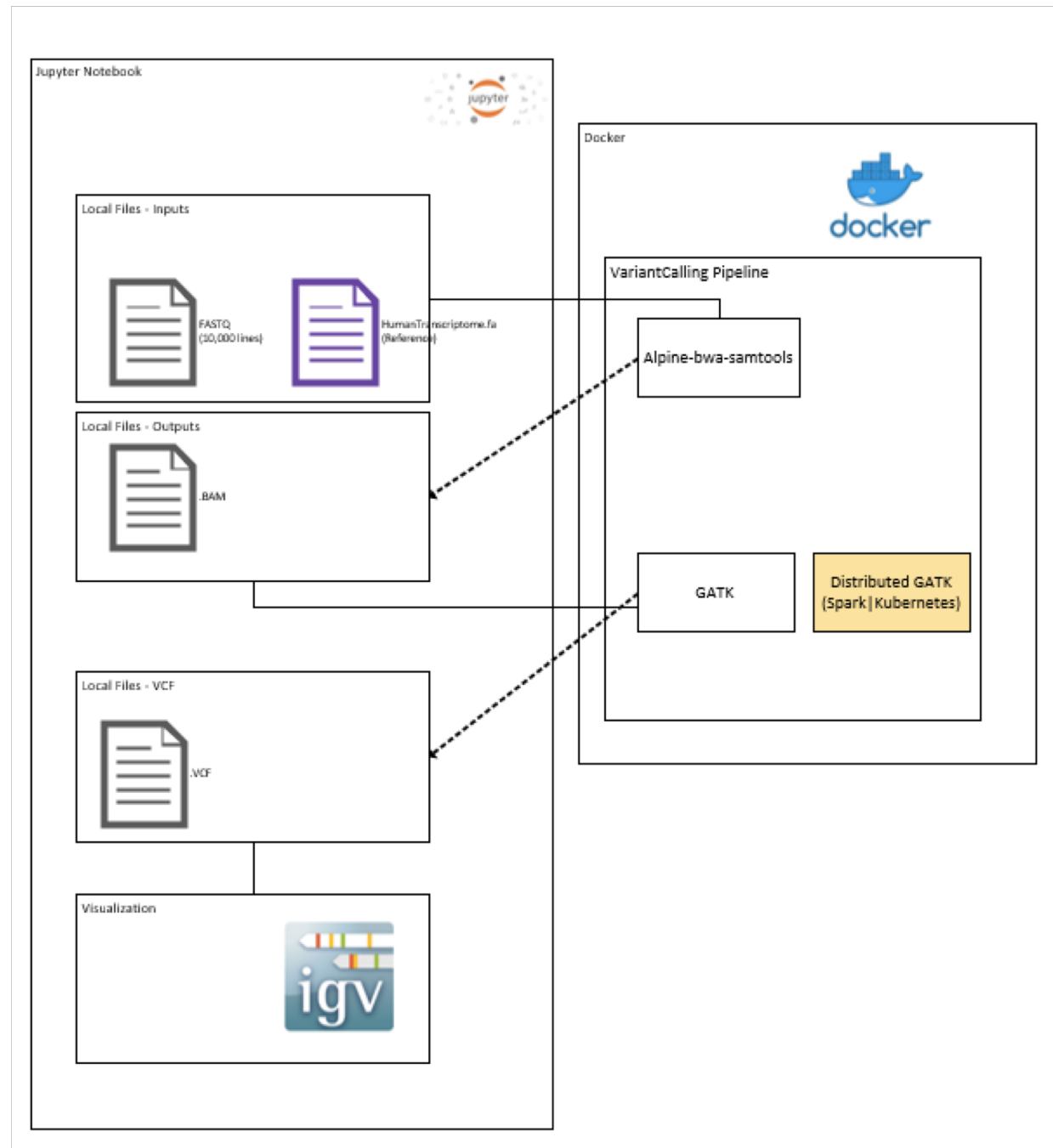
Ling-Hong Hung

Xingzhi Niu

Vera Diaz

Ka Yee Yeung

Wes Lloyd



# Step 1: Sample Data

Showing 1 to 50 of 90 samples

« Previous Next »

| Sample                  | Sex    | Population          | GRCh38 | Phase 3 | Phase 1 | Platinum pedigree | Structural variation | Gambian variation | 90 Han Chinese | Geuvadis |
|-------------------------|--------|---------------------|--------|---------|---------|-------------------|----------------------|-------------------|----------------|----------|
| <a href="#">HG00479</a> | Female | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00513</a> | Female | <a href="#">CHS</a> | •      | •       | •       | •                 |                      | •                 |                |          |
| <a href="#">HG00525</a> | Female | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00537</a> | Female | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00556</a> | Male   | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00436</a> | Male   | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00443</a> | Female | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00448</a> | Male   | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00683</a> | Male   | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00690</a> | Female | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |
| <a href="#">HG00619</a> | Male   | <a href="#">CHS</a> | •      | •       | •       |                   |                      | •                 |                |          |

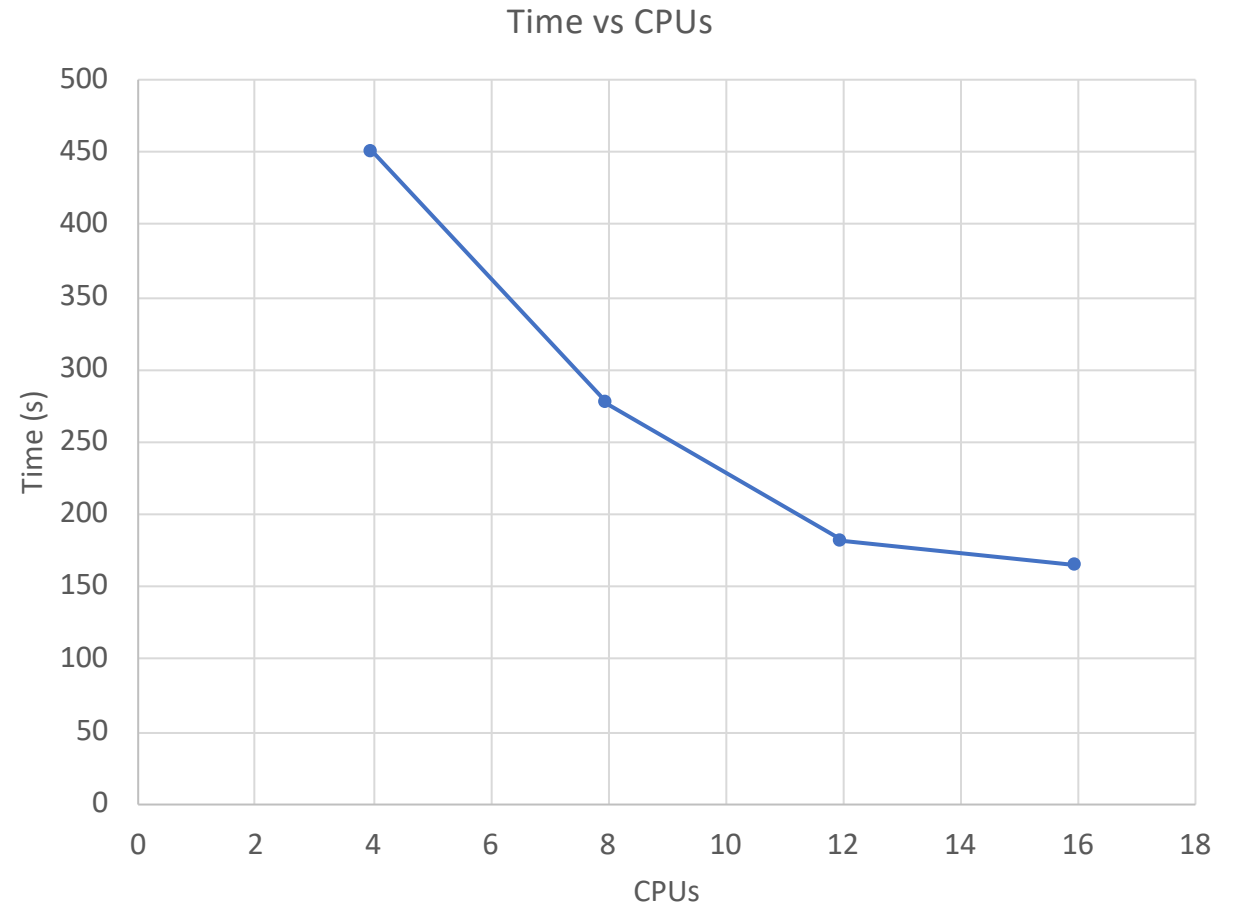
Individual HG00479  
First 1 million reads  
from ERR1044750  
sequencing run.

# Step 2: Alignment

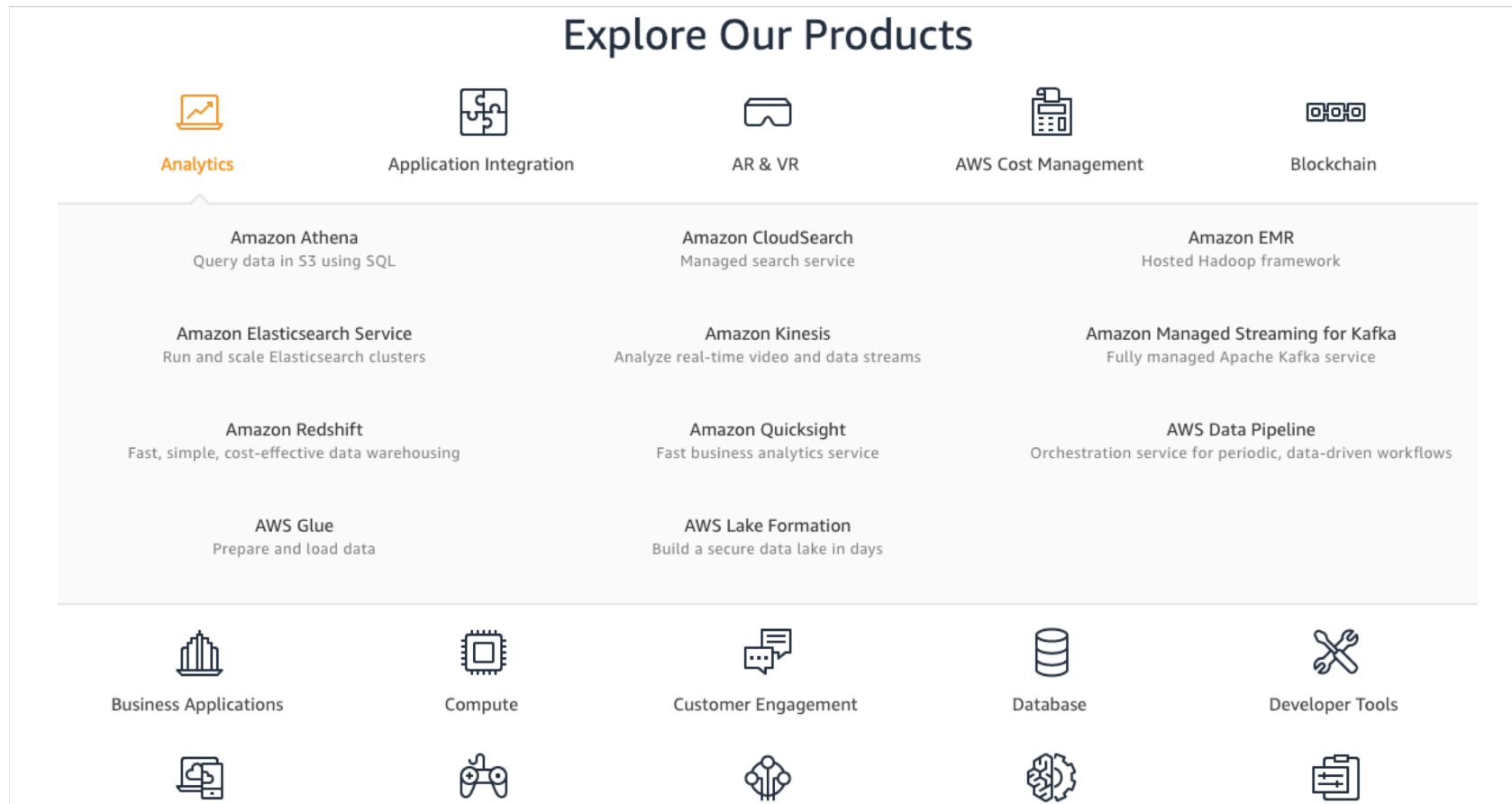
- BWA + Samtools Image
  - Ultra-lightweight, only 6mb for BWA and 50mb for samtools
- Parallelization
  - Multi-thread
  - Multi-container
  - AWS Batch

# Multi-Container Speed Up

- 1 million reads, 4 threads: 538.536 sec
- 500k reads, 4 threads: 306.468 sec
- 500k reads, 4 threads: 319.482 sec
- Amdahl's Law



# Lessons from AWS Batch



# Lessons from AWS Batch

- Setup CLI credentials
- Publish to Elastic Container Registry
- Create Job Queue
- Create Job Definition
- Create Job Environment
- Create Job Roles in IAM
- Create Job
- Create S3 Access Token
- ...

# Step 3: Variant Calling

- Running GATK
- Inside a container
- Inside a spark container locally
- On a Elastic Map Reduce Spark cluster
  - Stuck in the AWS Jungle!
  - Permissions, Ports, Credentials!