

Migrating High Performance Computing to the Cloud for a Bioinformatics Core *- A Pilot Test*

Jingzhi Zhu & Charlie Whittaker

Bioinformatics and Computing Core

The Koch Institute at MIT

What we do in computing

HPC
Clusters



> 50 labs

Storage
Servers



> 50 labs/cores

Application
Servers



> 10 labs/cores

Computing challenges

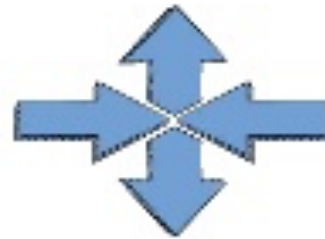
- Storage capacity
- Long job queue
- Aging equipment
- Maintenance efforts
- No billing model

Cloud computing

- Instant Access



- Scalability



- Transparent cost

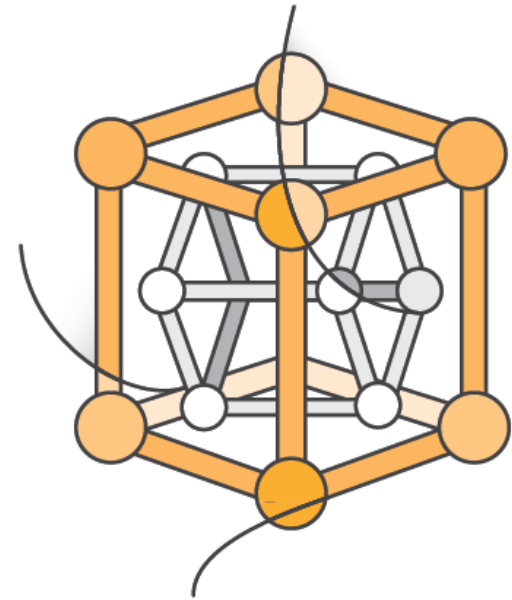


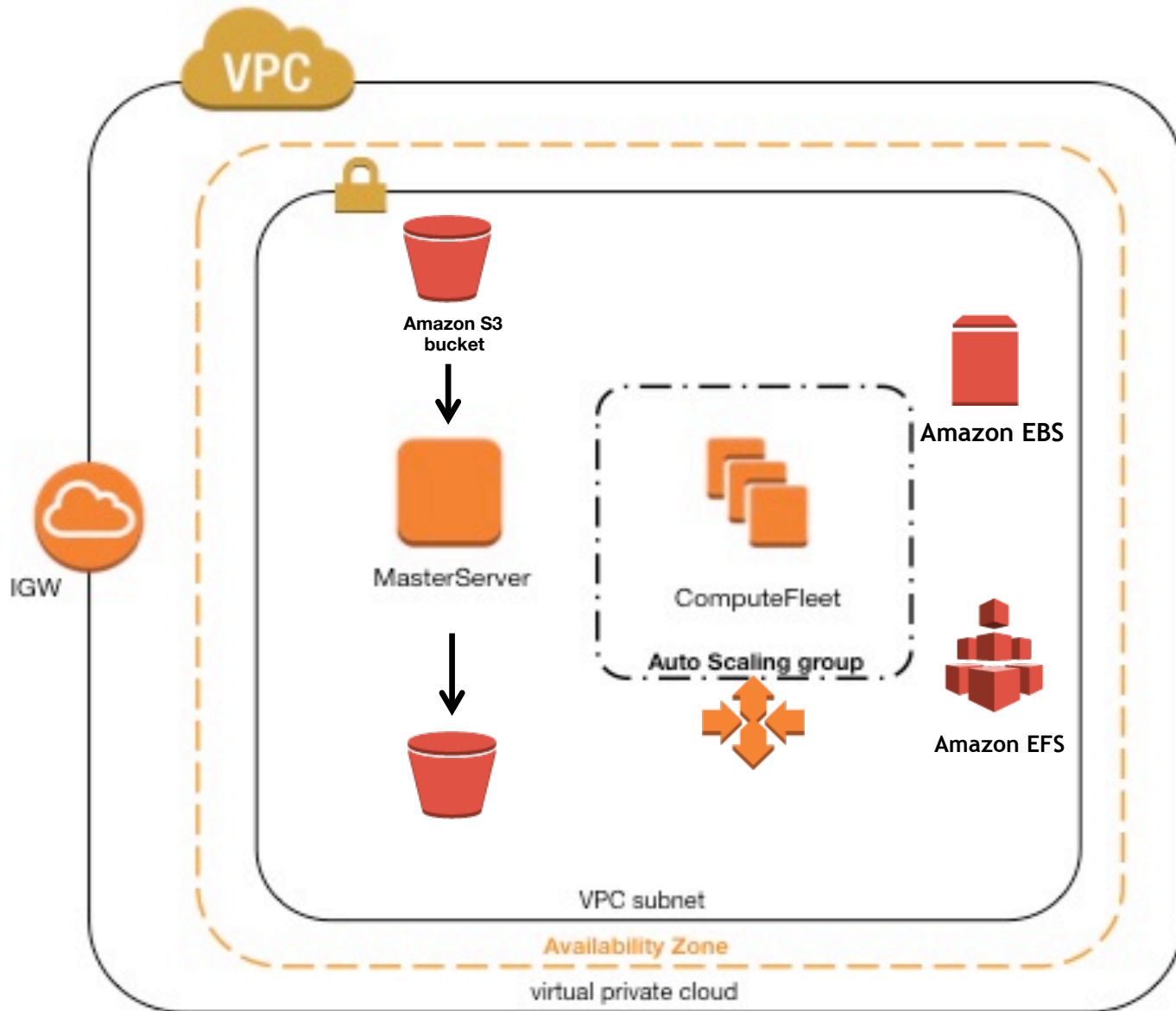
Why choose Amazon Web Services (AWS)

- MIT participates AWS service provided under Internet2
- MIT provides high-speed connection link to AWS

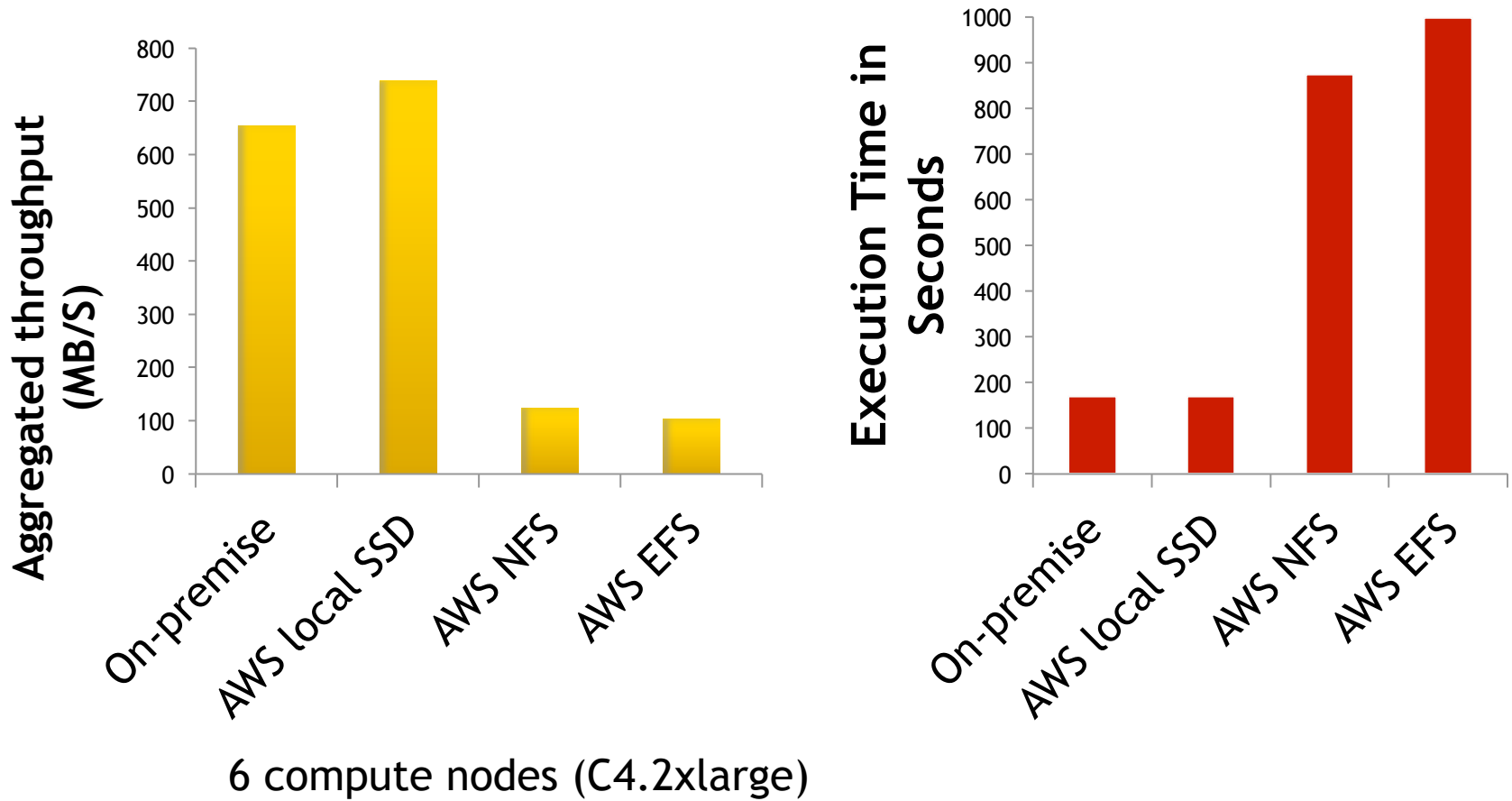
Instant HPC: CfnCluster

- Cloud formation cluster (CfnCluster)
- Framework to build and manage High Performance Computing (HPC) clusters on AWS
- Quick start in 10 min
- Elasticity using auto scaling

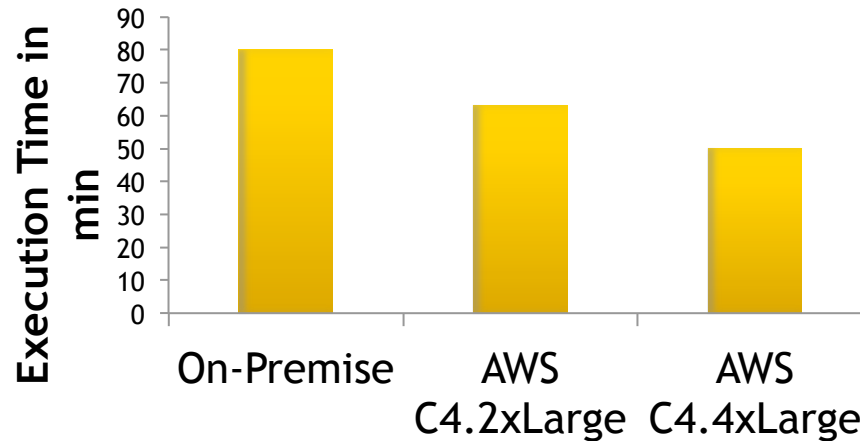




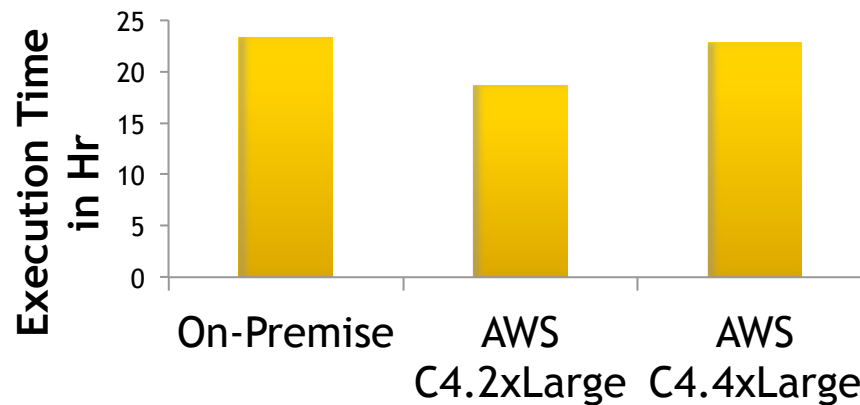
IO performance test using fio



Performance test: bwa

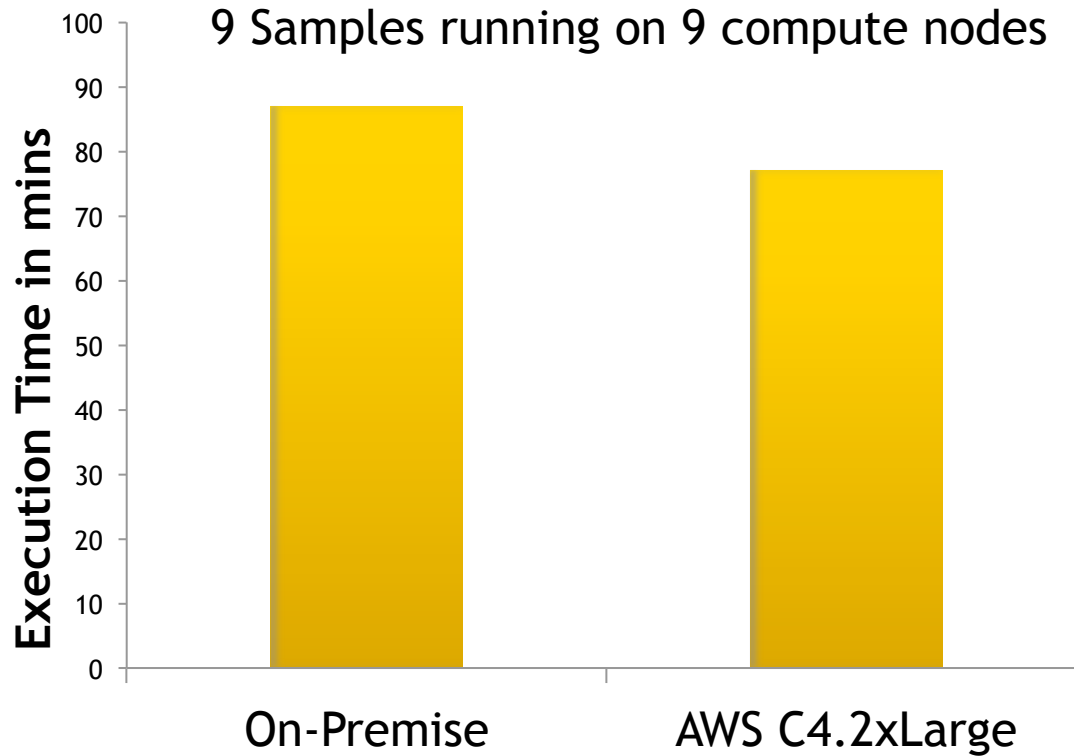


bwa aln: CPU bound



bwa sampe: IO bound

RNA-Seq Pipeline Test: rsem

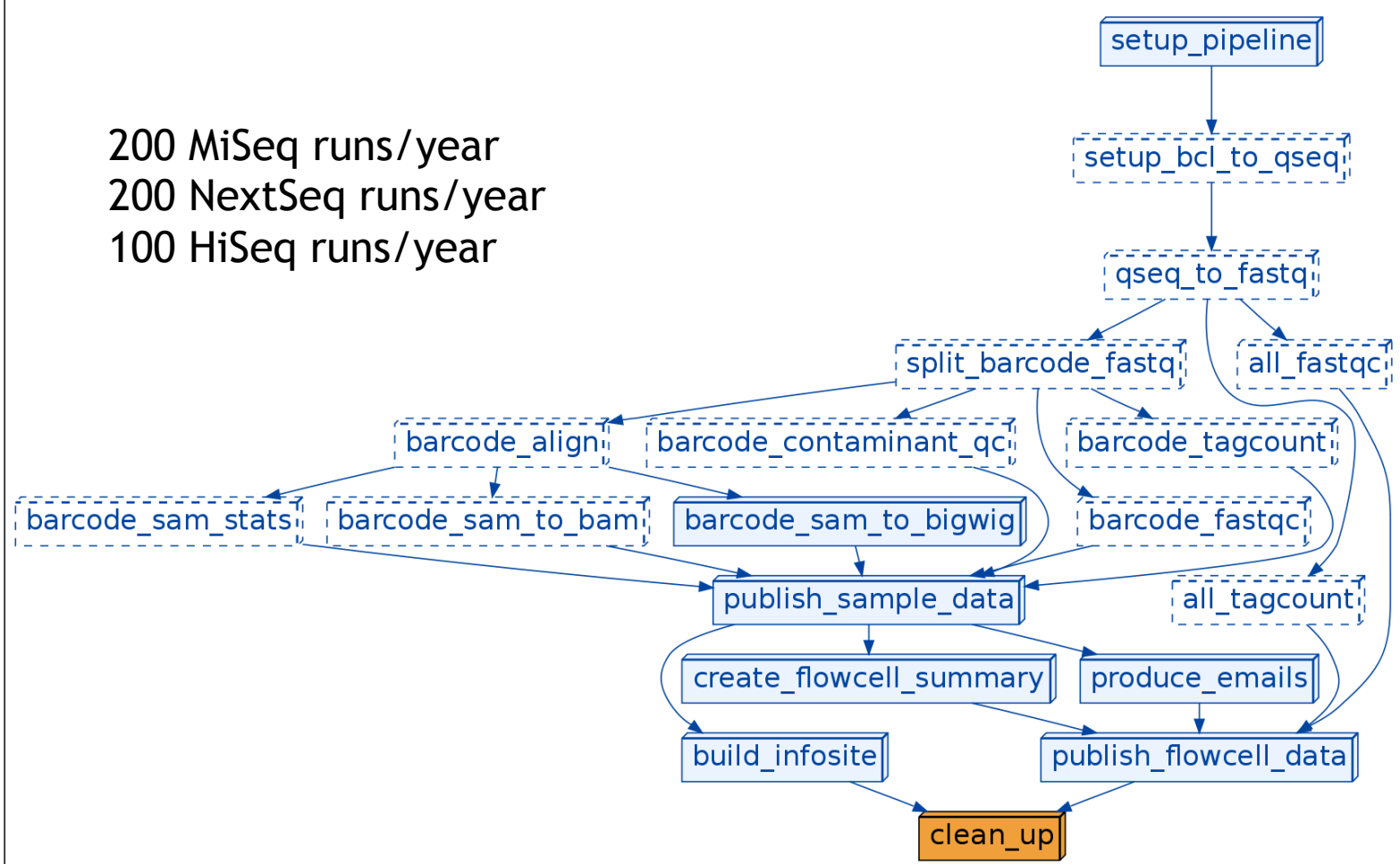


Total Project Cost on AWS: **\$10**

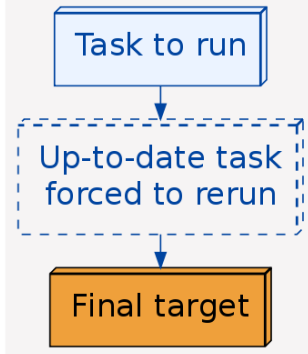
Genomics Core Illumina Pipeline

Pipeline:

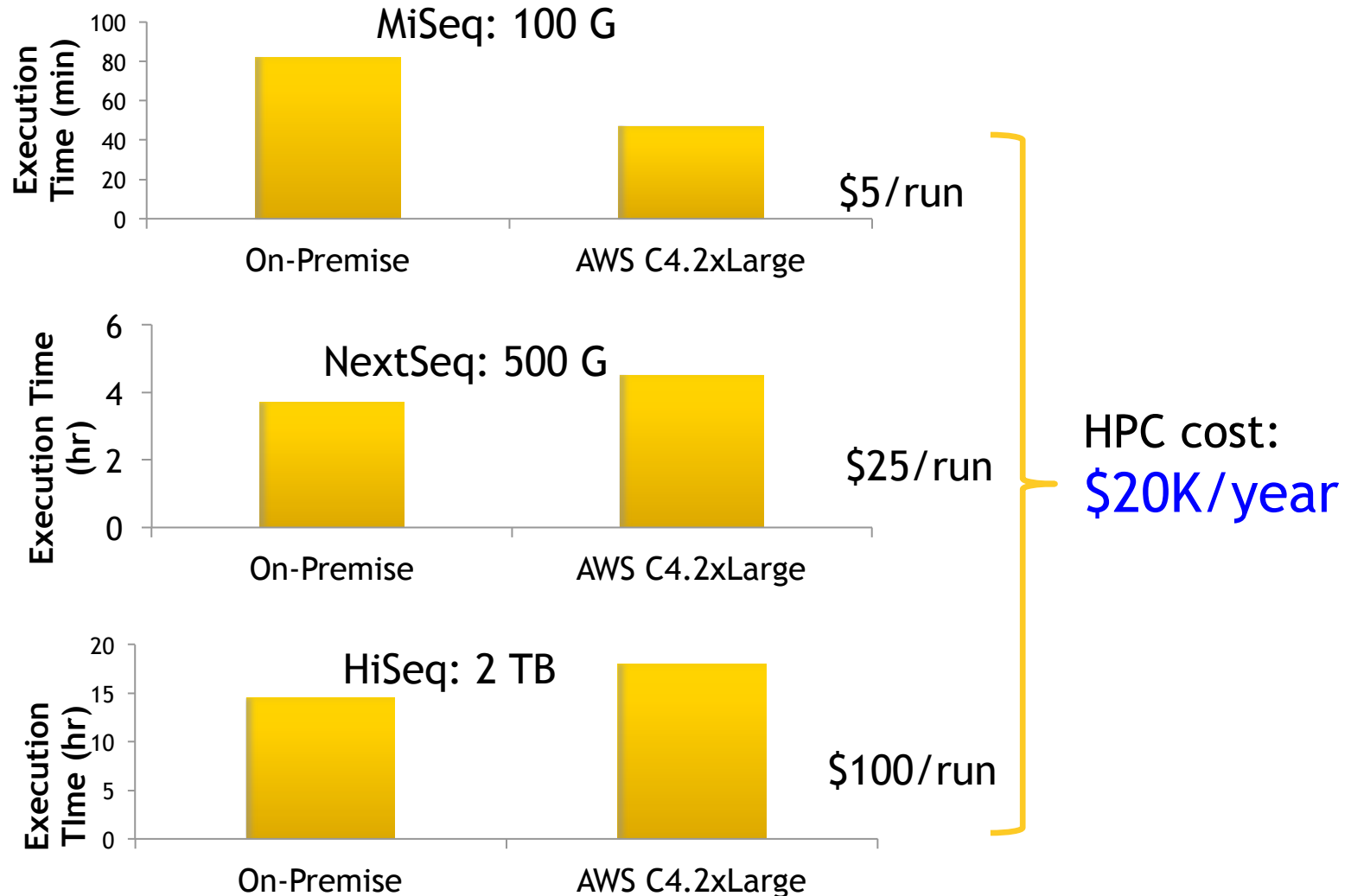
200 MiSeq runs/year
 200 NextSeq runs/year
 100 HiSeq runs/year



Key:



Performance Test: Illumina Pipeline



Cost Estimation

- On-premises HPC: \$20K-30K/year (not including MIT-provided infrastructure such as room, power, cooling and network)
- On-premises storage: \$50K/year
- Get wall clock from SGE history (`$qacct -d 365`)
- AWS: 300,000 hr/year x \$0.4/hr = \$120K/year
- Parallel factor: assume 2-3 jobs running simultaneously on one node on average, HPC cost on AWS could be \$40K-60K/year
- Spot instances (saving cost but may lose productivity)
- AWS S3 storage cost: \$360/TB/year

Thoughts on Cloud Computing

- Solve on-premises computing challenges (capacity planning, job queue, aging equipment, maintenance and billing)
- User training
- Need expert knowledge of AWS
- Hybrid environment

Summary

- Cloud computing provided us expanded computing capacity
- AWS HPC performance comparable to our on-premises hardware