

GCP TOPMed Analysis Pipeline Migration

May 11, 2019

GAC Computing

- 2009: fisher, neyman, NetApp
- 2010-2018: Compute Cluster (pearson)
 - > 1500 Silicon Mechanics compute cores
 - Isilon 12000, Nexenta, Isilon NL-410
 - GENEVA, GARNET, OLGA, TOPMed, HSCT
 - Brian & Sharon Browning, Sara Lindstroem, Tim Thornton
- 2015-2020: Cloud Computing
 - Amazon Web Services
 - Google Cloud Platform

GAC Cloud Computing (UW AWS)

- 2015. Began working on Amazon Web Services
 - Most mature cloud provider
 - Close UW/AWS relationship
 - Discounted rates
 - Local support from AWS sales & solutions architect
 - Access to AWS developers
 - UW/AWS contract in place
- Focus on analysis pipeline since most compute-intensive
 - CfnCluster (AWS Parallel Cluster) architecture based on Sun Grid Engine (SGE)
 - Switched to new HPC tools AWS Batch

GAC Cloud Computing (NHLBI AWS)

- Access to NHLBI AWS systems Summer, 2018
- Migrated work from UW AWS account
- Early adopters, issues to shake out
 - Permissions
 - Separate storage and computer accounts
 - Increasing default limits
 - Cost and performance monitoring (March, 2019)

GAC Cloud Computing (UW GCP)

- NIH STRIDES GCP announcement Summer, 2018
- Created UW account to investigate GCP
 - Google uses global identifiers for all their services
 - Long investigation into use of [user@uw.edu](#)
 - Interaction between UW & Google groups
 - NHLBI billing ID Carahsoft (difficulties assigning)
 - No UW/GCP contract
- GCP recommended Slurm Workload Manager
- Built Slurm test cluster from on-premise nodes

GAC Cloud Computing (NHLBI GCP)

- TOPMed Task Order 2 Task Area 2
- One-year project initiated 5/16/2019
 - Migrate TOPMed analysis pipeline from AWS to GCP
 - Compare in-house computing vs. GCP cloud

People

- AWS Users
 - Caitlin McHugh, Cecelia Laurie, Amarise Little, Anya Mikhaylova,
- GCP Migration
 - Dave project management
 - Roy software development
 - Robert design and build clusters
 - John testing and documentation
 - Matt and Stephanie benchmarks
 - (Google) Eric, Jonathan and others GCP support
 - (NIH) Antej to provide NIH/IT support