MAKER as a Service: Moving HPC applications to Jetstream Cloud
Nicholas Hazeckam\textsuperscript{1}, Upendra Kumar Devisetty\textsuperscript{2}, Nirav Merchant\textsuperscript{3}, and Douglas Thain\textsuperscript{1}
\textsuperscript{1}Department of Computer Science and Engineering, University of Notre Dame, Email: {nhazeckam,dthain}@nd.edu
\textsuperscript{2}CyVerse, The University of Arizona, Email: {upendr}@cyverse.org
\textsuperscript{3}The University of Arizona, Email: {nirav}@email.arizona.edu

Abstract
As cloud resources become more common as an execution platform, the need to transition applications between HPC and the cloud becomes a necessity. However, because of the complex setup and system specific demands of these applications, transition is difficult and may not scale as desired. Jetstream is a NSF funded cloud service that is aiming to provide these services for users in a dynamically allocated nature.

In this work we look at three key areas to focus on when transitioning between resources: providing a portable reproducible environment, scaling between local and remote resources, and using feedback to the user for informing configuration and runtime decisions. Building on the MAKER bioinformatic application, we have deployed Work Queue-MAKER on the Jetstream cloud platform, helping to annotate over 30 genomes and accelerating performance from days to hours and weeks to days.

Challenges
- Differences between HPC and Cloud platforms make it difficult to provide consistent versioning and configuration of software.
- Cloud instances lack a shared filesystem, limiting many multi-machine MPI applications.
- Complexity inherent in MPI applications and cloud platforms combine to make debugging and performance analysis difficult.

MAKER Standard Installation

MAKER - Bioinformatic Pipeline

VC3-Builder

Work Queue

Evaluation Results

Comparison of execution time using MPI and WQ-MAKER on variable number of cores, evaluating Fungal Genome(41MB)

This graph shows the number of executing tasks and used cores overlayed with the total available and completed tasks. This was executed on a Condor Cluster with each instance build MAKER with VC3 on available machines.

Conclusion
Deploying MAKER on Jetstream has helped us to better understand reliable deployment, scalability, and what is useful feedback. We generated a consistent build using machine images built from VC3-Builder. We provided dynamic scalability by using Work Queue to distribute partitions, and we ran existing MPI parallelism at the worker level. WQ-MAKER users leveraged runtime data to understand performance, improve scaling, and diagnose errors. WQ-MAKER has been used for over 30 genomes and can accelerate annotation time using a variety of resources.