

# MAKER as a Service: Moving HPC applications to Jetstream Cloud

Nicholas Hazekamp<sup>1</sup>, Upendra Kumar Devisetty<sup>2</sup>, Nirav Merchant<sup>3</sup>, and Douglas Thain<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering,  
University of Notre Dame, Email: {nhazekam,dthain}@nd.edu

<sup>2</sup>CyVerse, The University of Arizona, Email: {upendra}@cyverse.org

<sup>3</sup>The University of Arizona, Email: {nirav}@email.arizona.edu



UNIVERSITY OF  
NOTRE DAME

College of Engineering



## Abstract

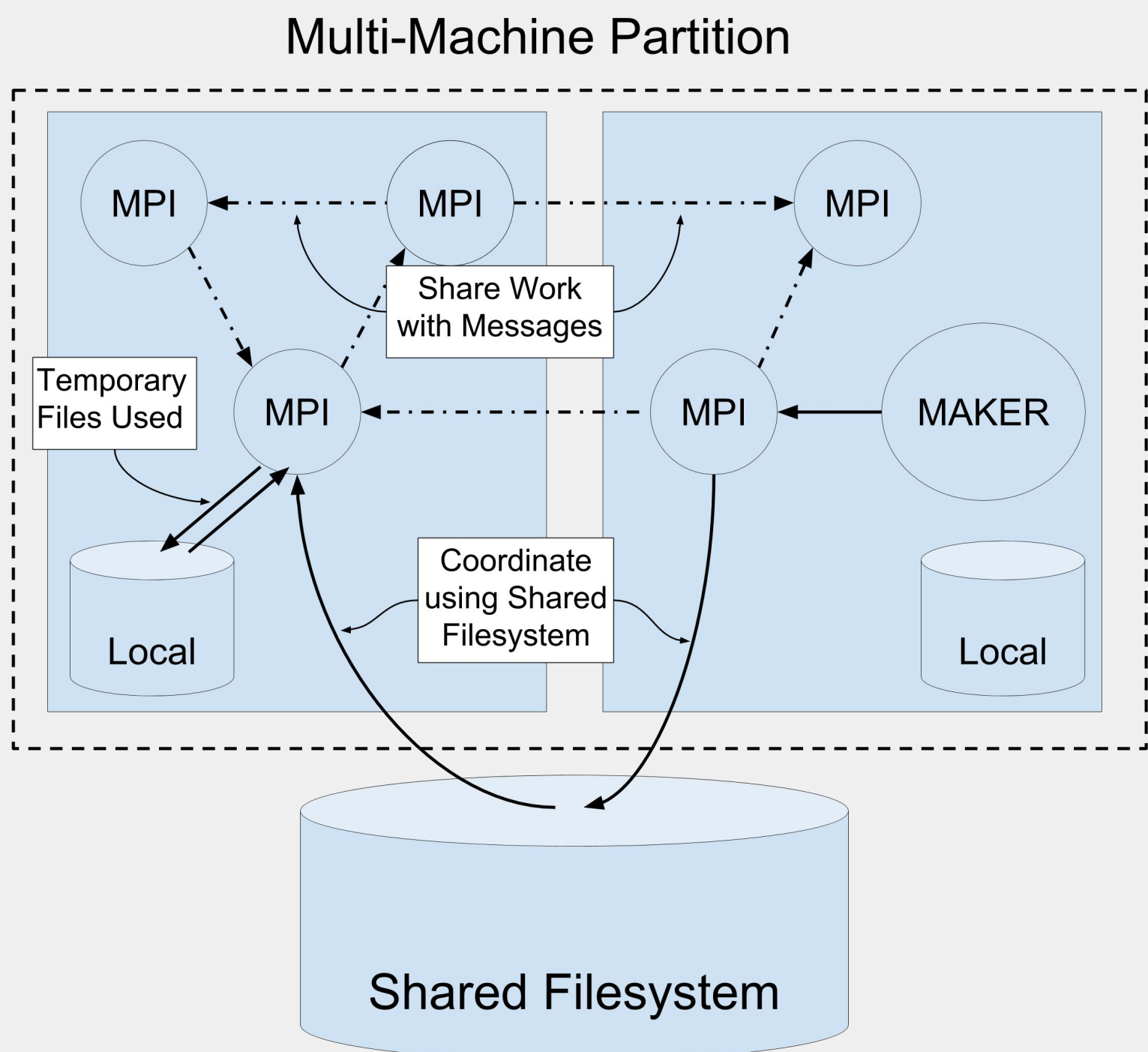
As cloud resources become more common as an execution platform, the need to transition applications between HPC and the cloud becomes a necessity. However, because of the complex setup and system specific demands of these applications, transition is difficult and may not scale as desired. Jetstream is 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.

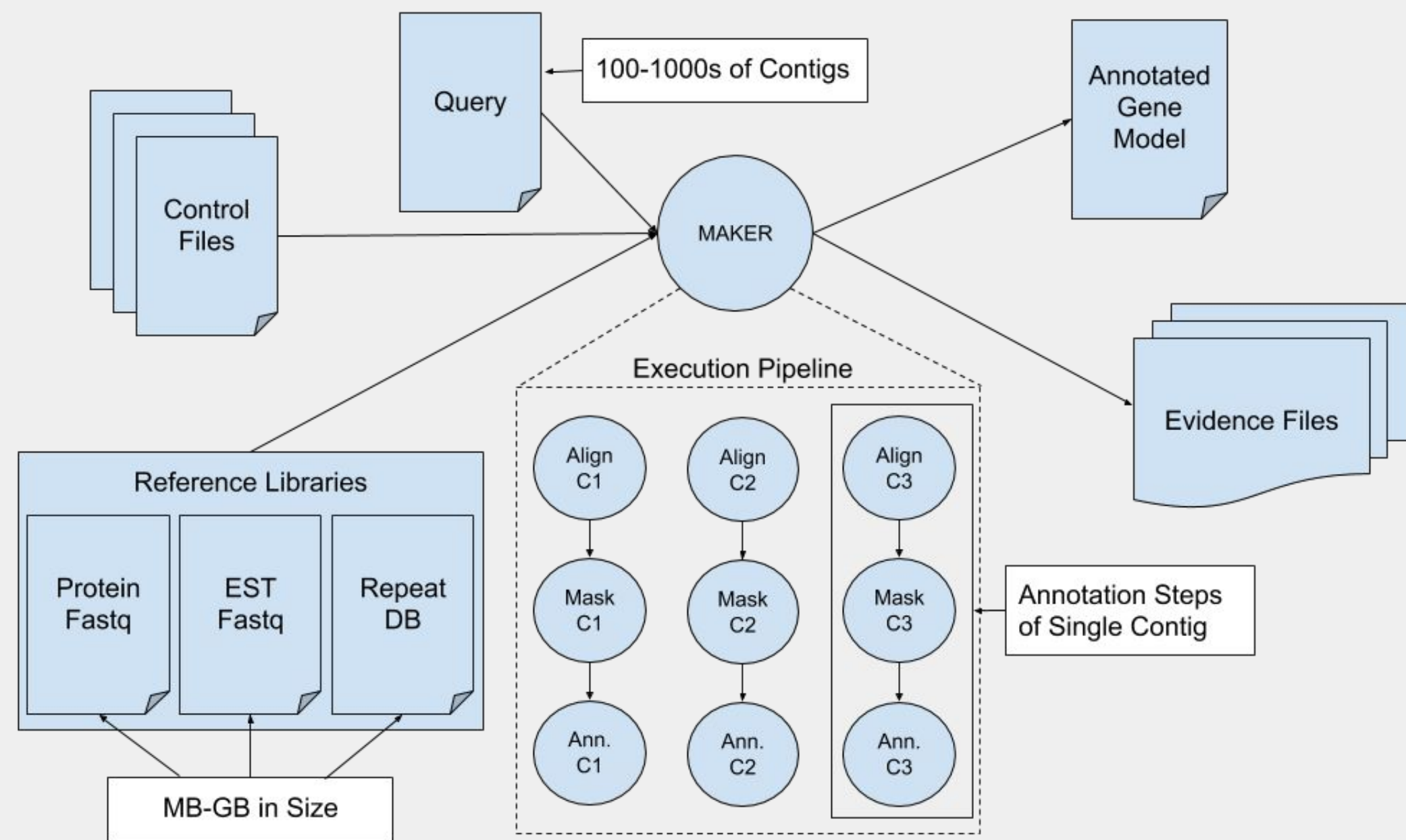
## Challenges

- 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.

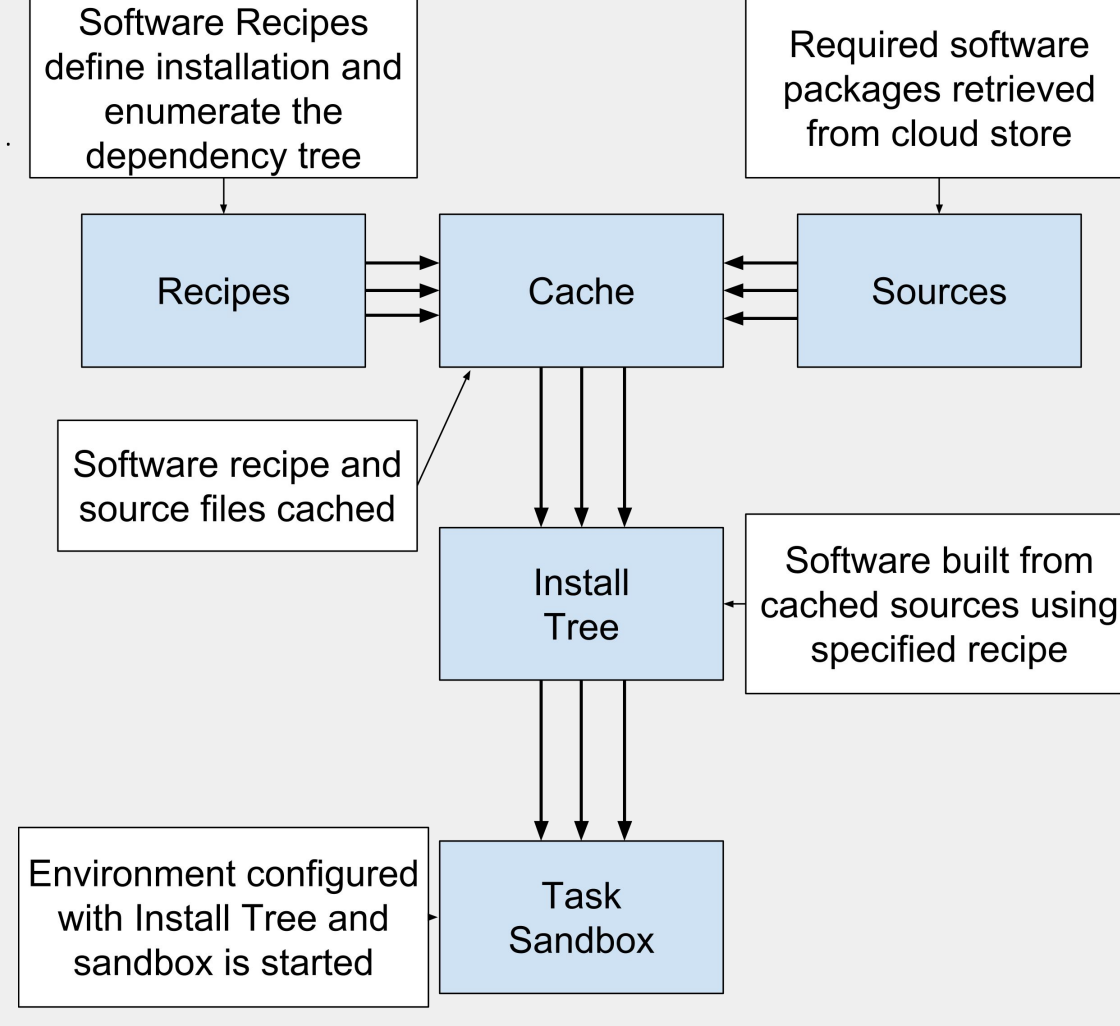
### MAKER Standard Installation



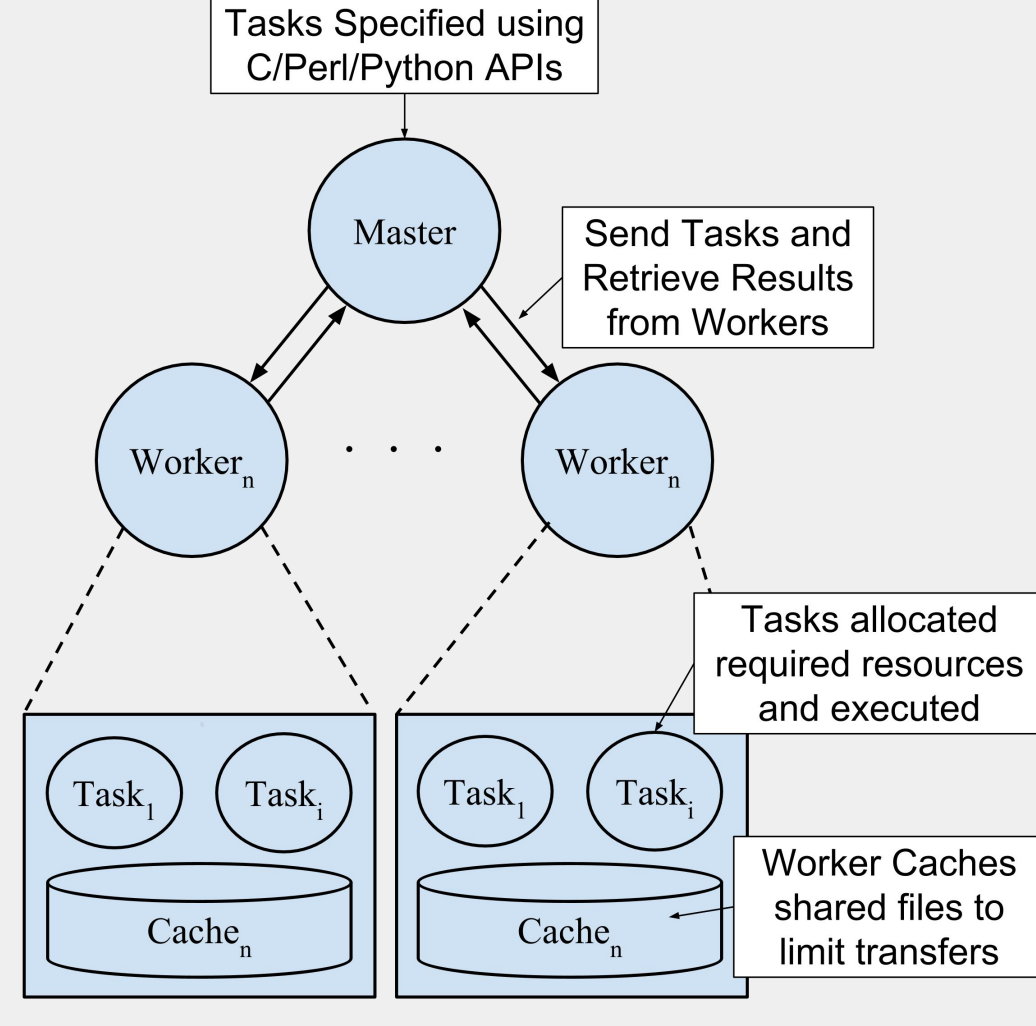
## MAKER - Bioinformatic Pipeline



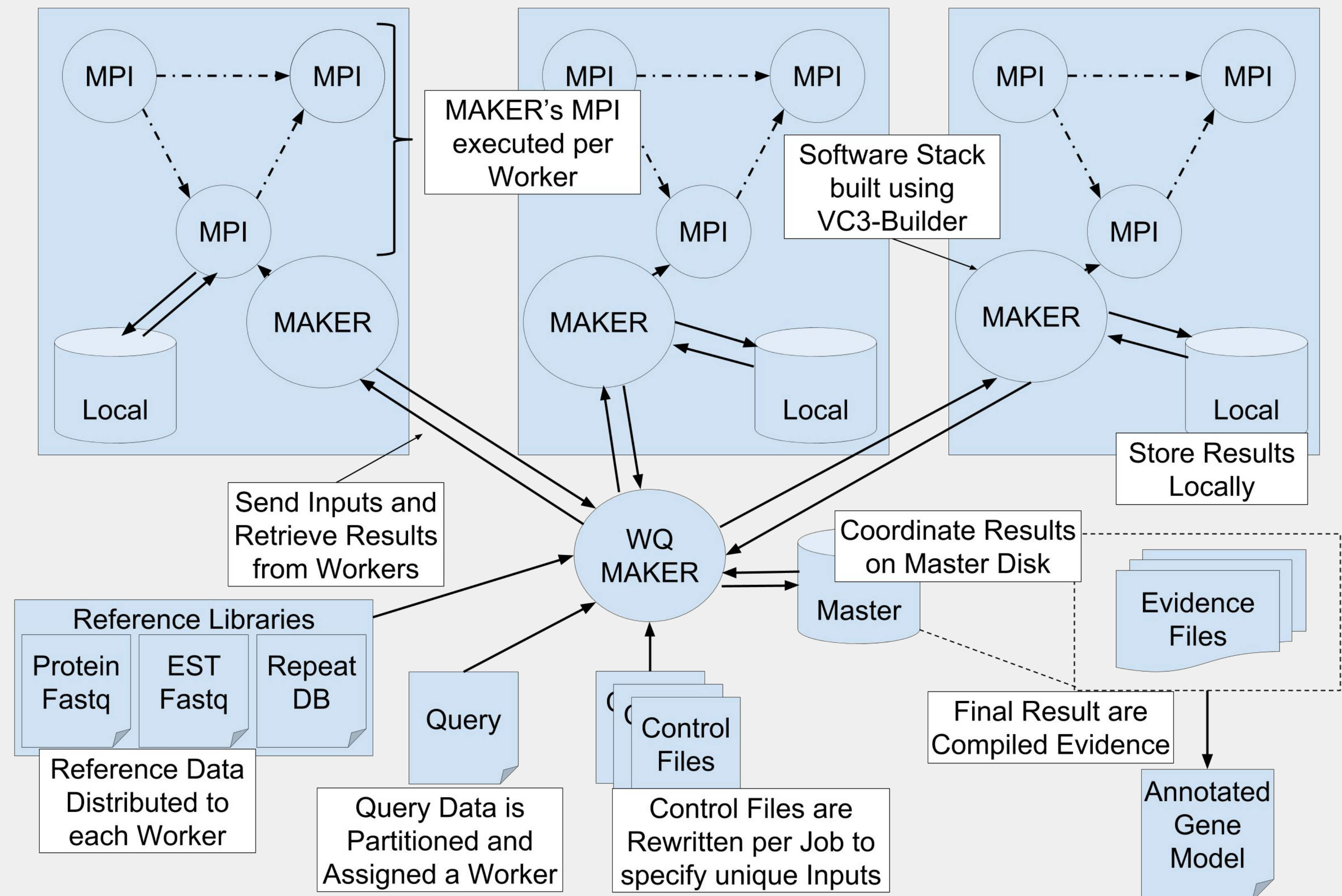
## VC3-Builder



## Work Queue

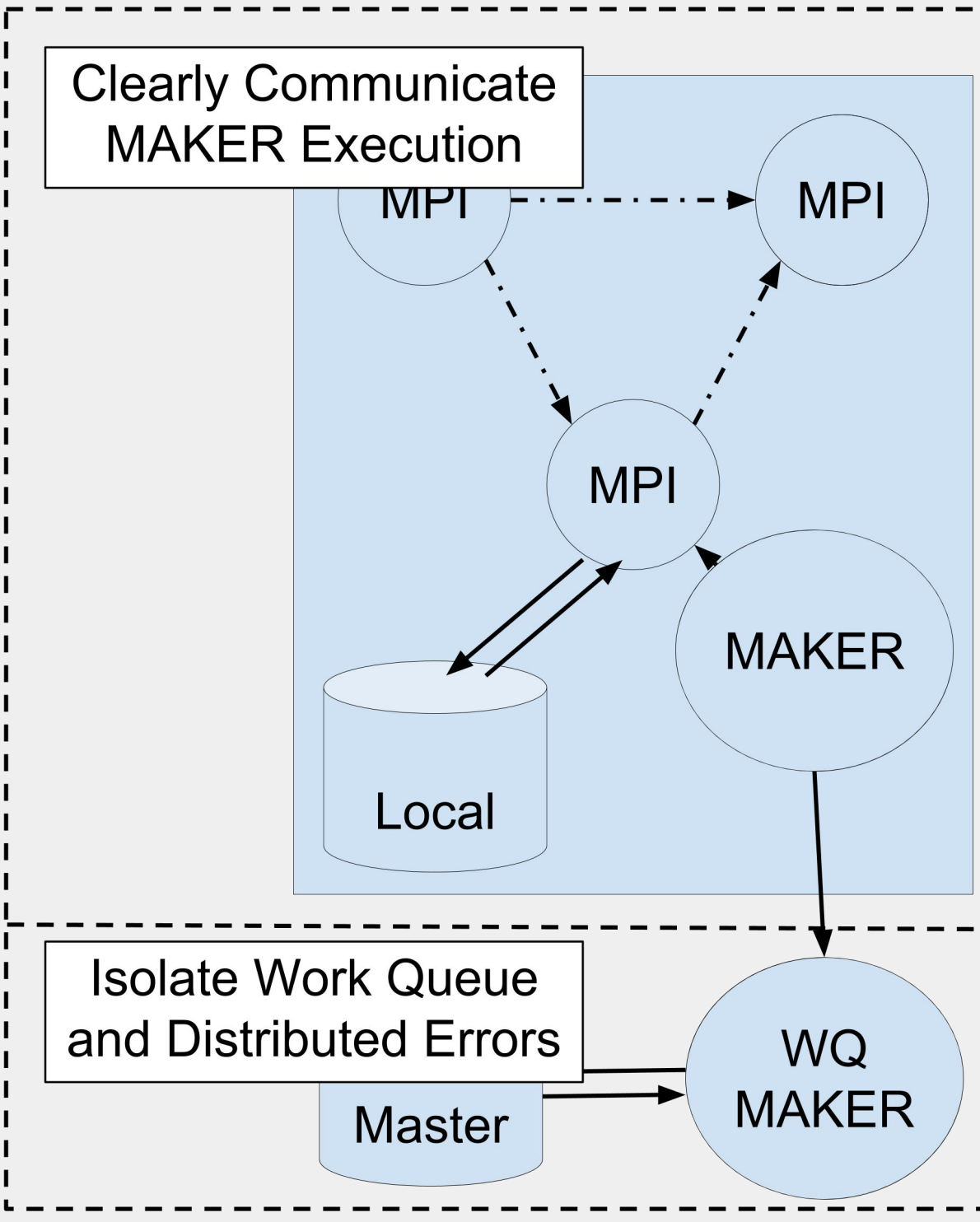


## WQ-MAKER System Design



## User Feedback

### Debugging



### Runtime Feedback

Sat Sep 23 15:24:25 2017 :: File Fungal annotated :: 177  
Sat Sep 23 15:24:25 2017 :: Failed to annotate :: None

Type	Success	Failure	Abandon	Total
Tasks	177	12	0	177

Workers:	Joined	Removed	Idle-Out	Lost
	10	12	2	0

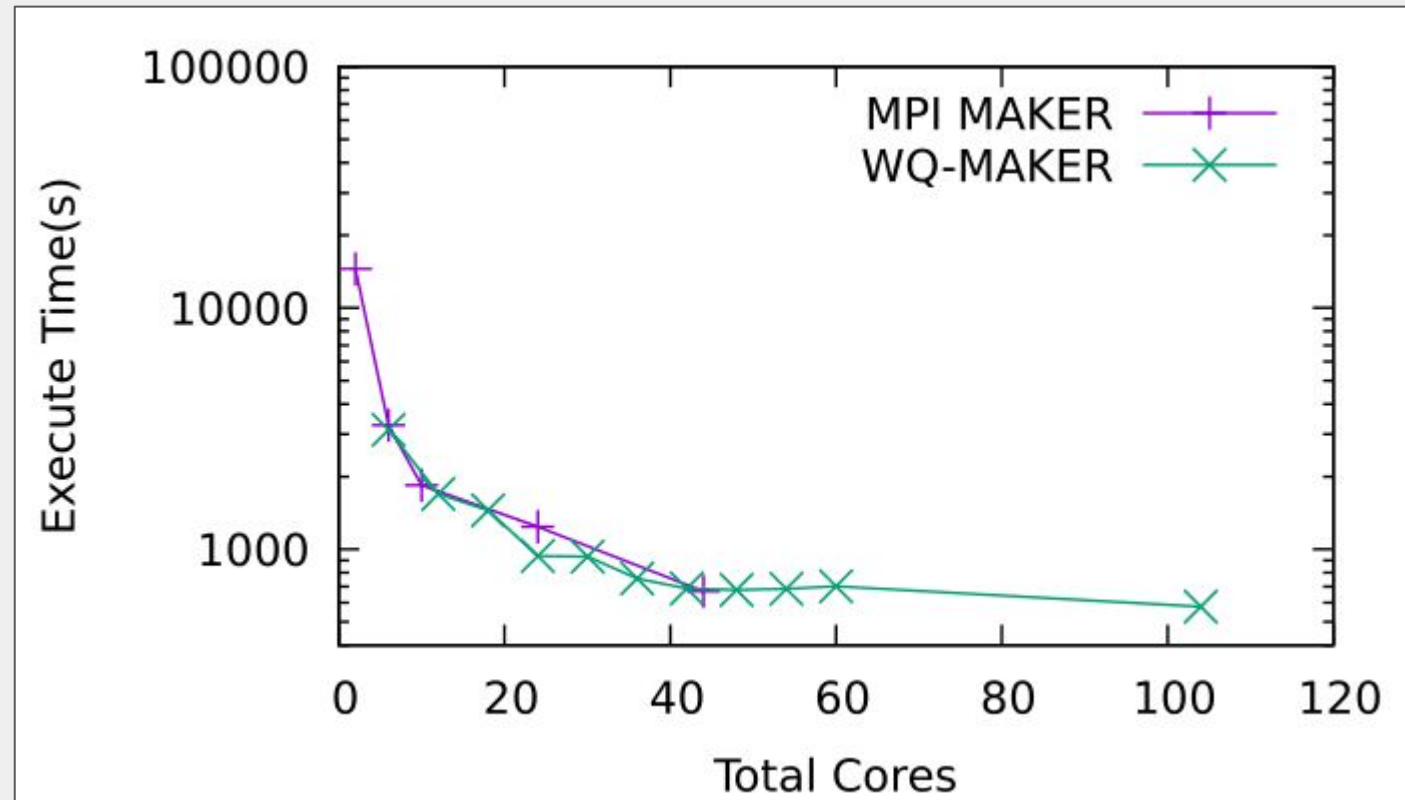
  

Work Queue Wall Time: 1d 18:02:15  
Cumulative Task Wall Time: 5d 21:40:58  
Cumulative Task Good Execute Time: 5d 8:33:32  
Work Queue Send Time: 0d 0:11:18  
Work Queue Receive Time: 0d 0:58:28

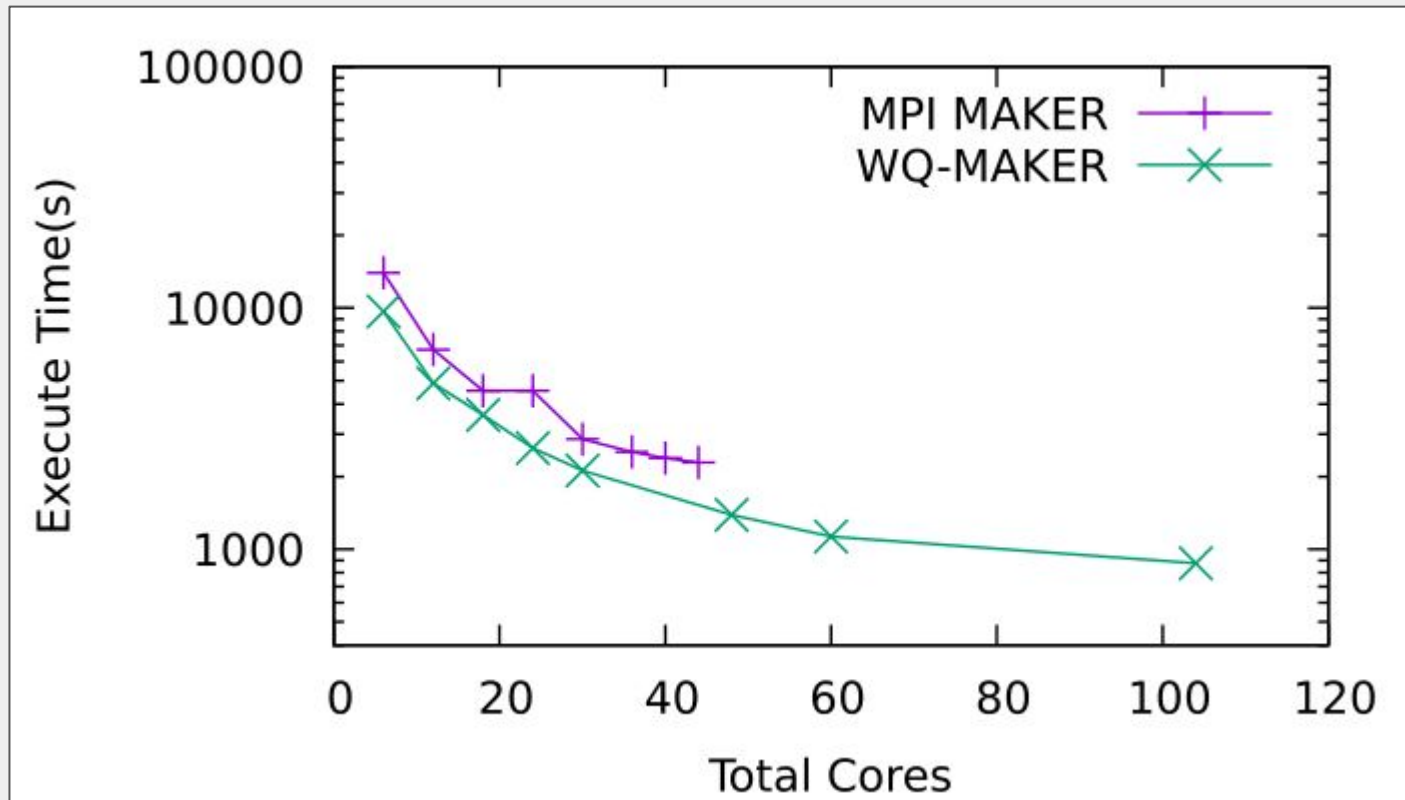
### Where you can find this work:

CCL/CCTools (Includes Work Queue)	ccl.cse.nd.edu
MAKER	yandell-lab.org/software/maker.html
VC3-Builder	github.com/vc3-project/vc3-builder

## Evaluation Results

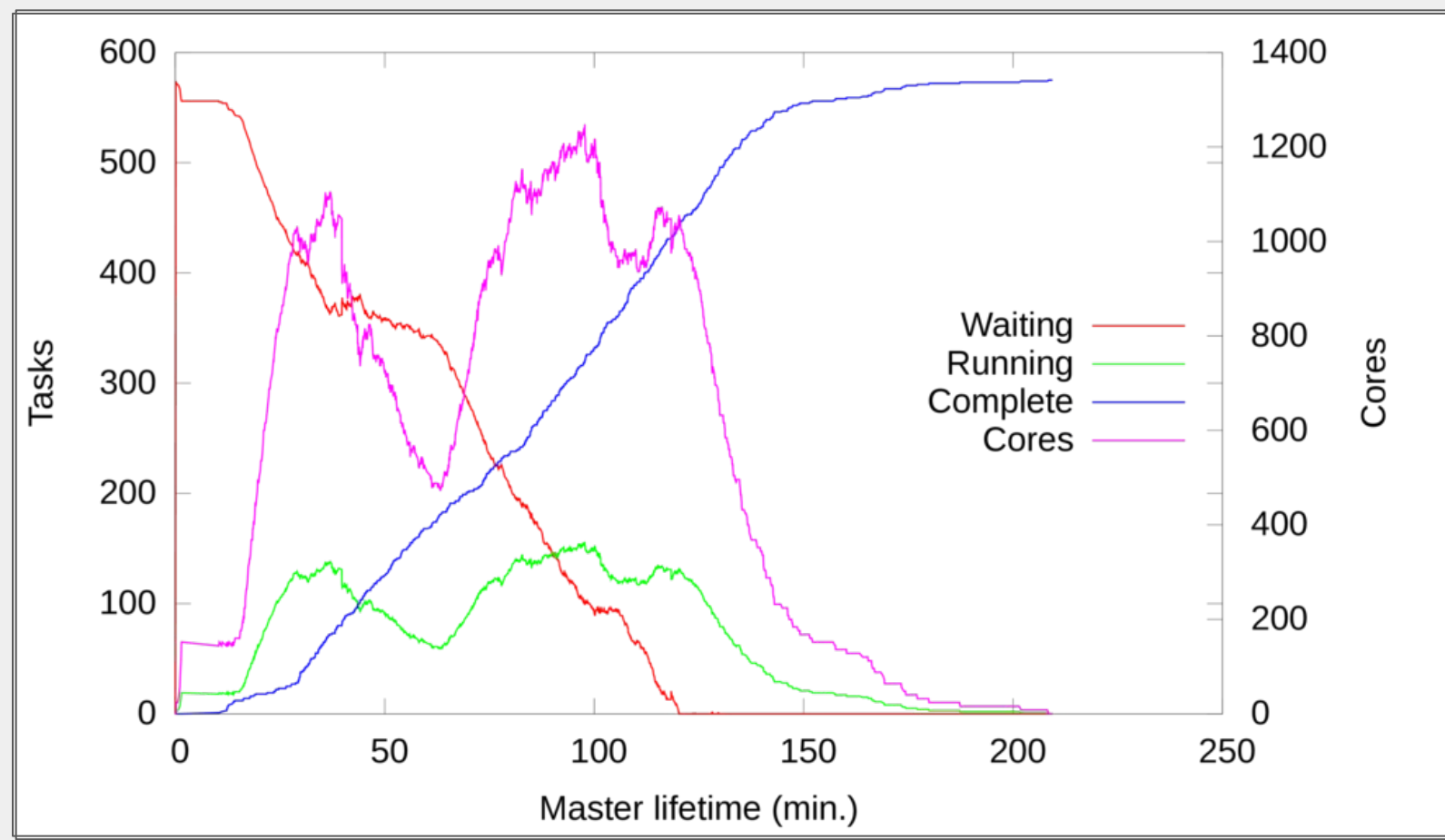


Comparison of execution time using MPI and WQ-MAKER on variable number of cores, evaluating Fungal Genome(41MB)



Comparison of execution time using MPI and WQ-MAKER on variable number of cores, evaluating subset of Hummingbird Genome(900MB)

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.



	Execution Time	Cores	Total CPU Hrs	Speedup
MAKER*	52 days	1	---	---
MPI MAKER	57 hrs	24	1368	---
WQ-MAKER	3.36 hrs	80-1344	1725.5	15.8

## 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.