

Lessons Learned Moving MAKER from HPC to the Cloud



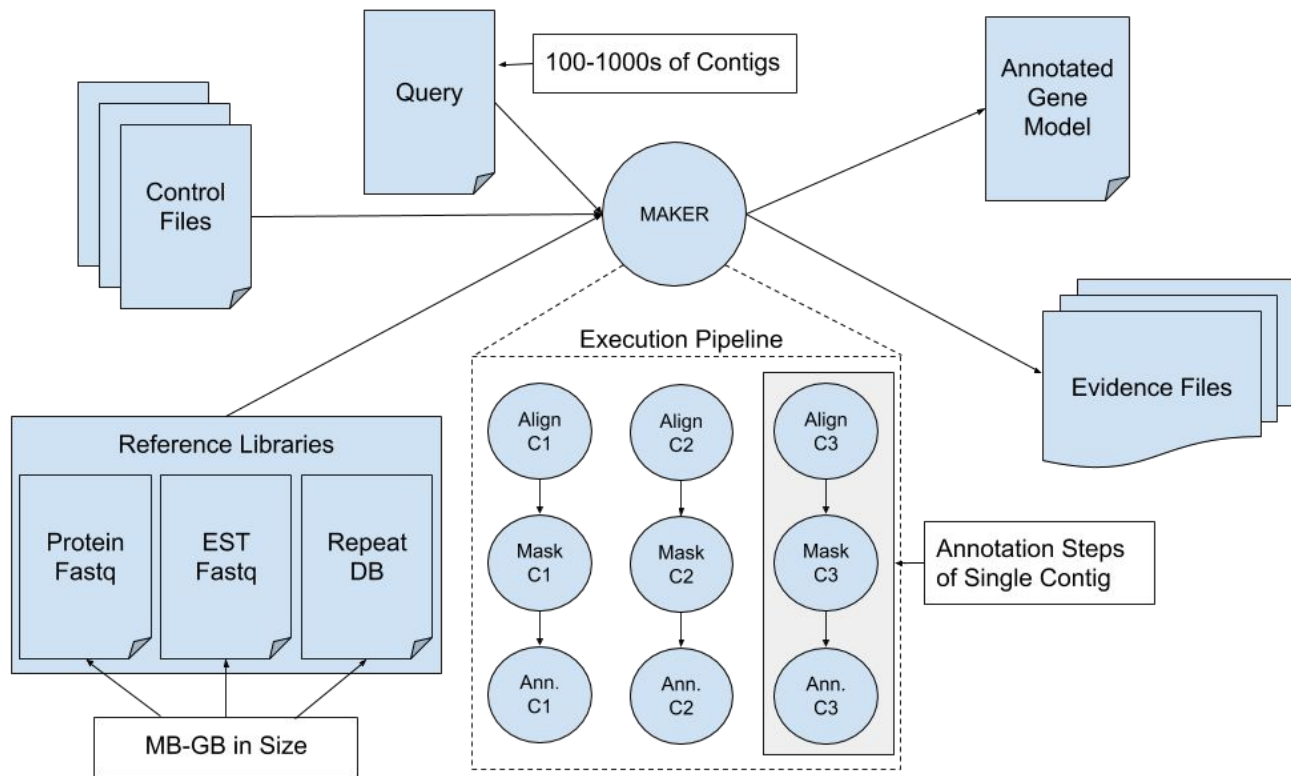
Nick Hazekamp¹, Upendra Kumar Devisetty²,
Nirav Merchant², and Douglas Thain¹

¹University of Notre Dame

²University of Arizona



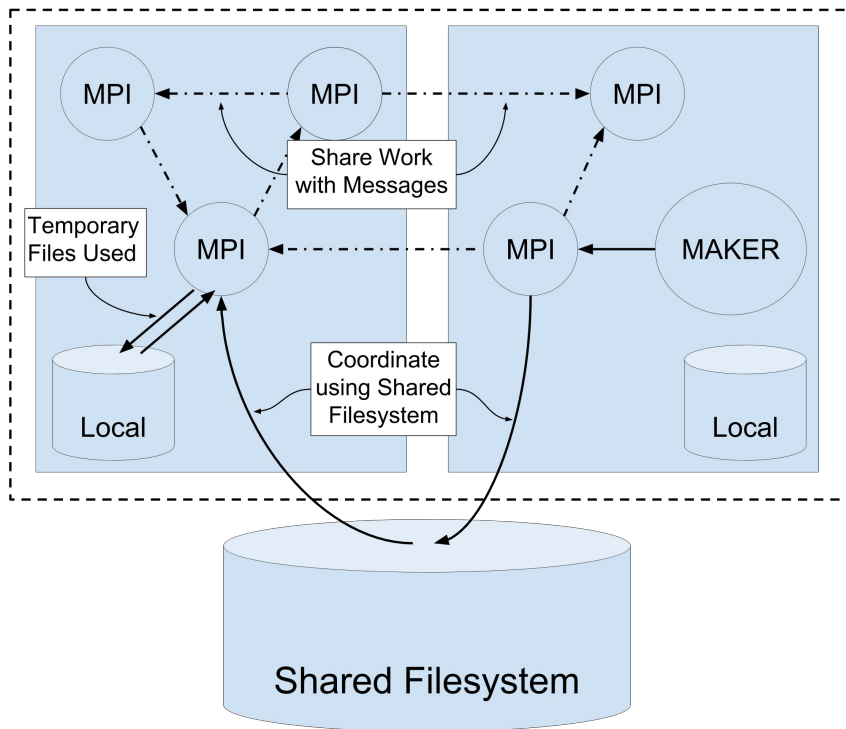
MAKER : A Bioinformatics Pipeline





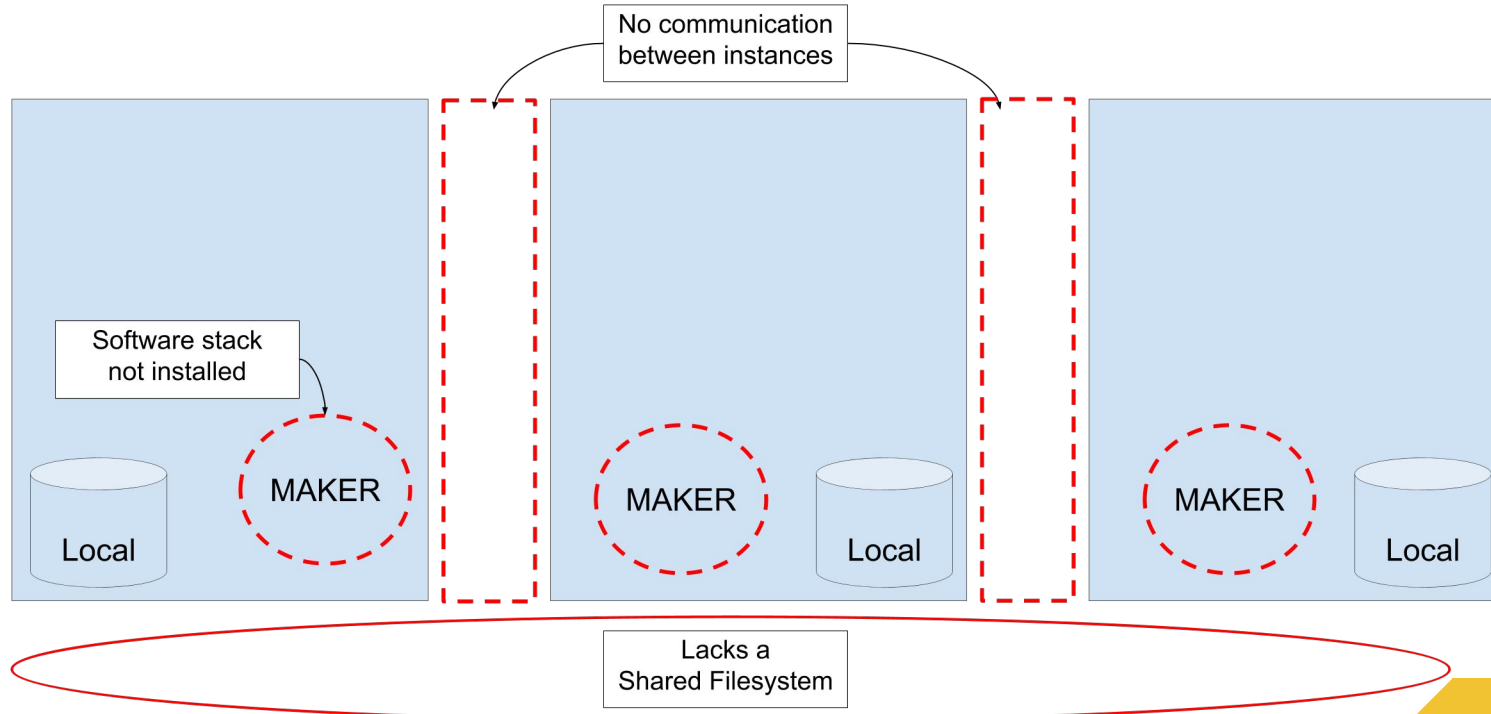
MAKER In Practice

Multi-Machine Partition

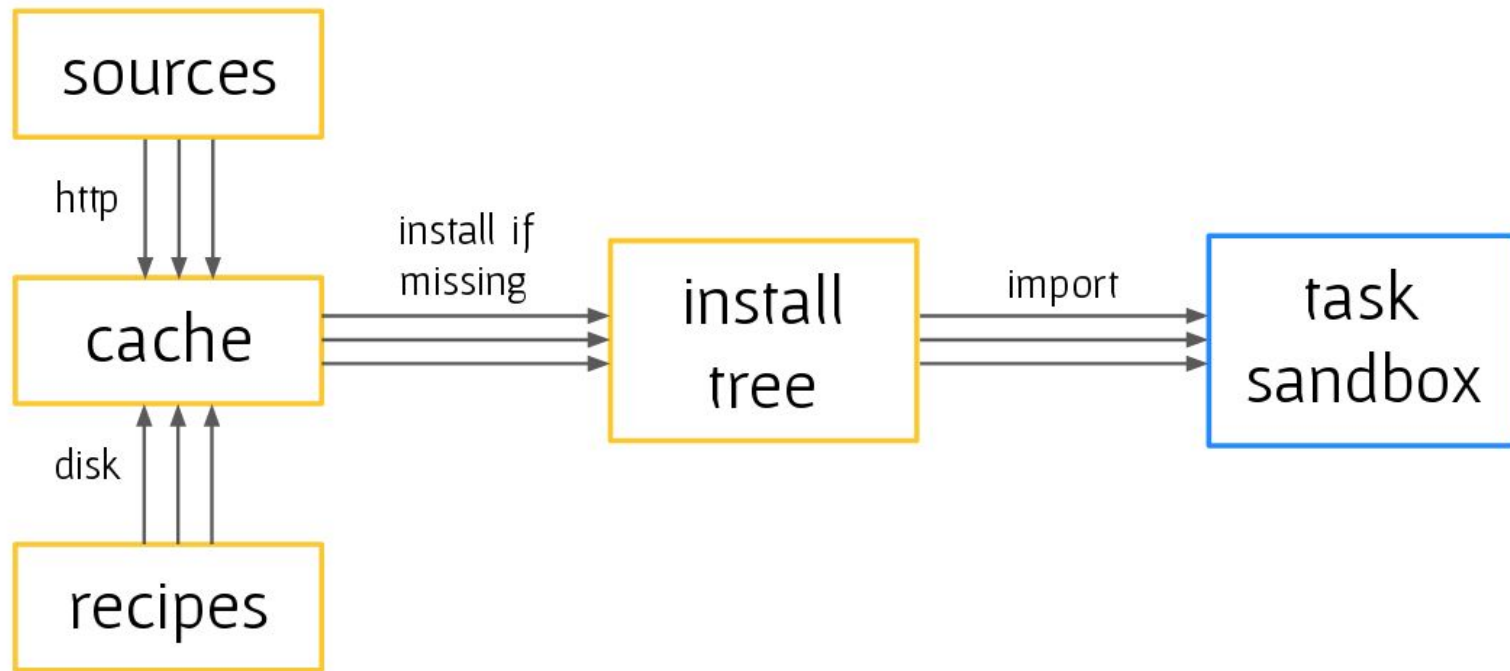




Jetstream Resources Available

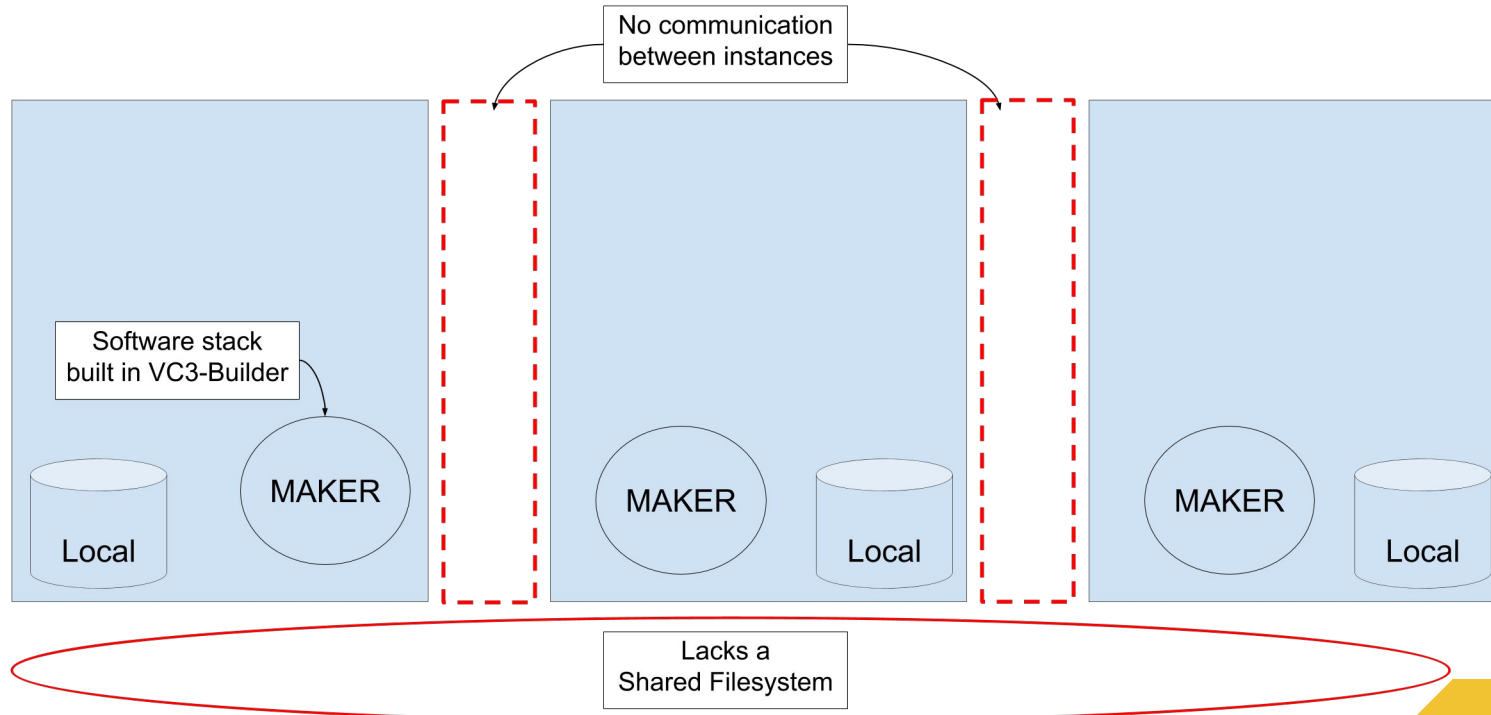








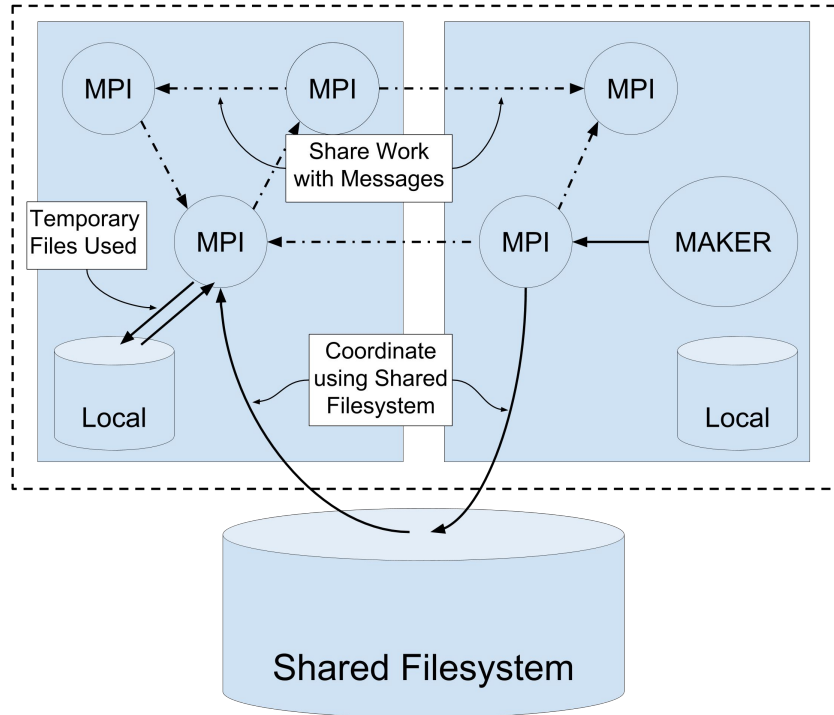
VC3-Builder Provided Software Stack





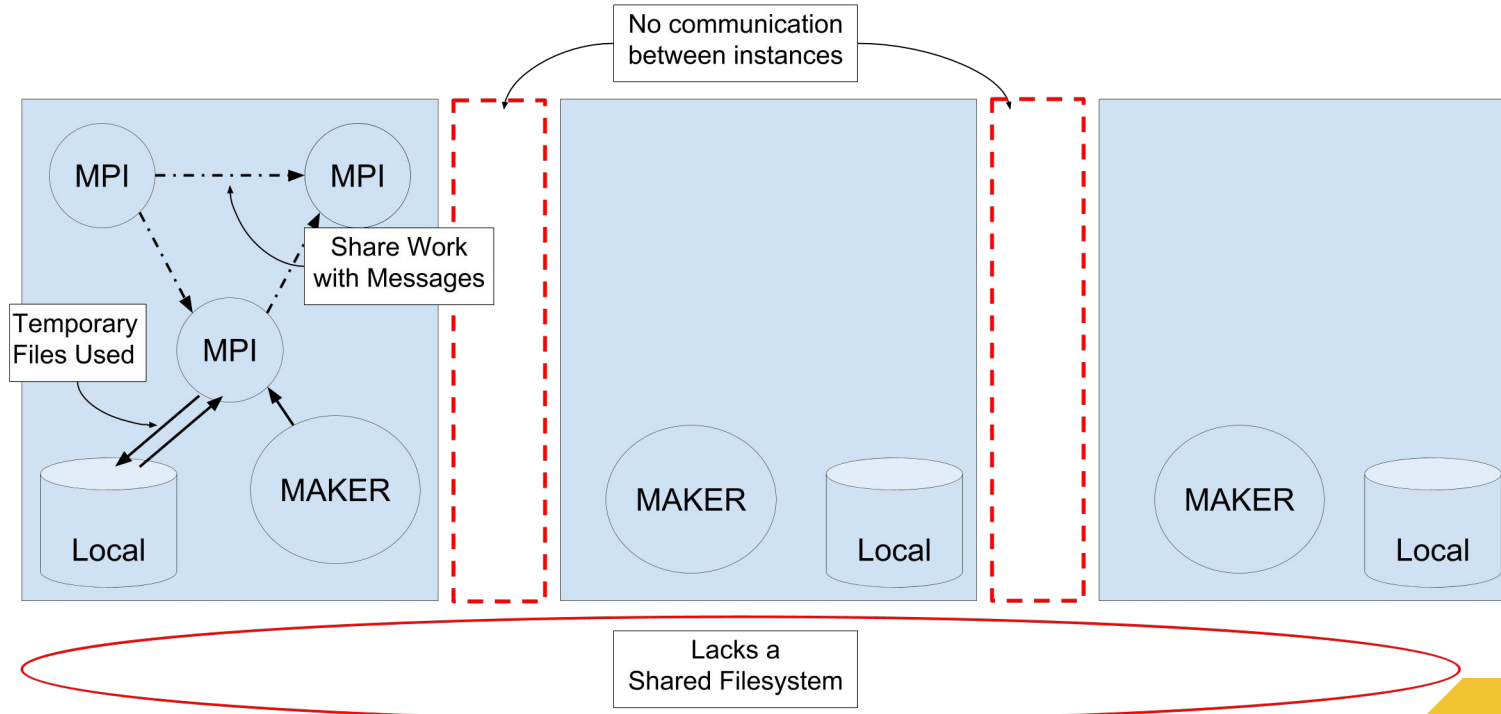
MAKER in an HPC Environment

Multi-Machine Partition



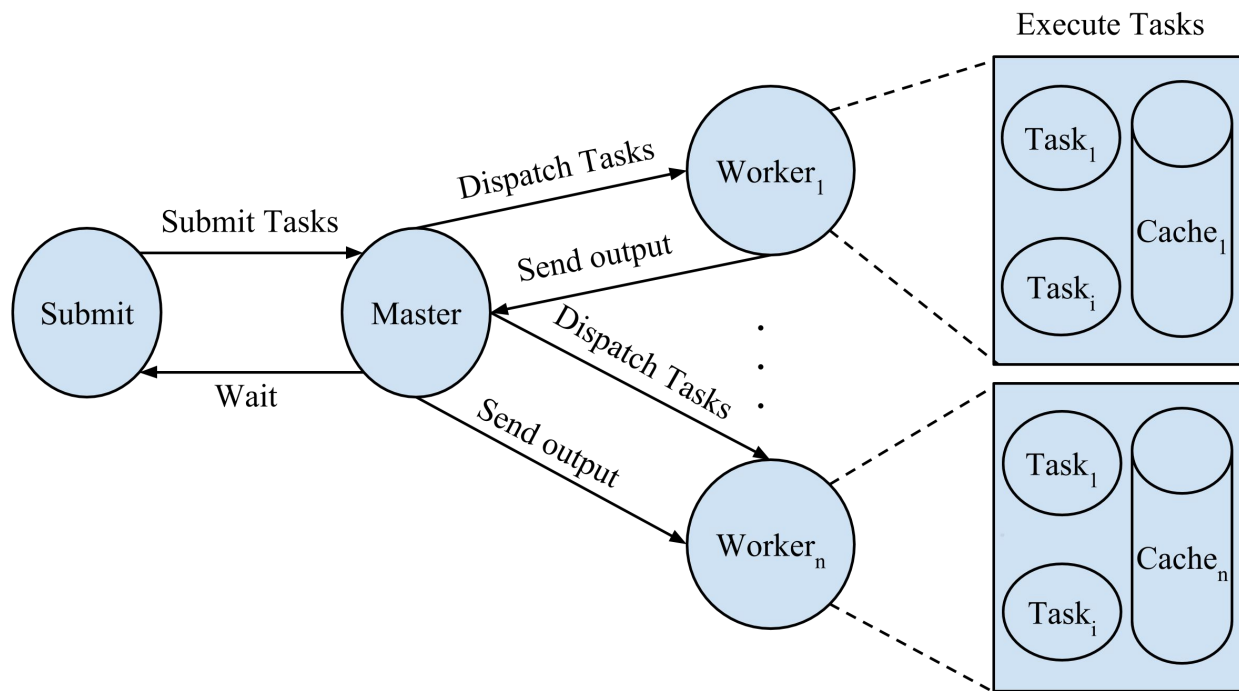


MPI Run on Single Node



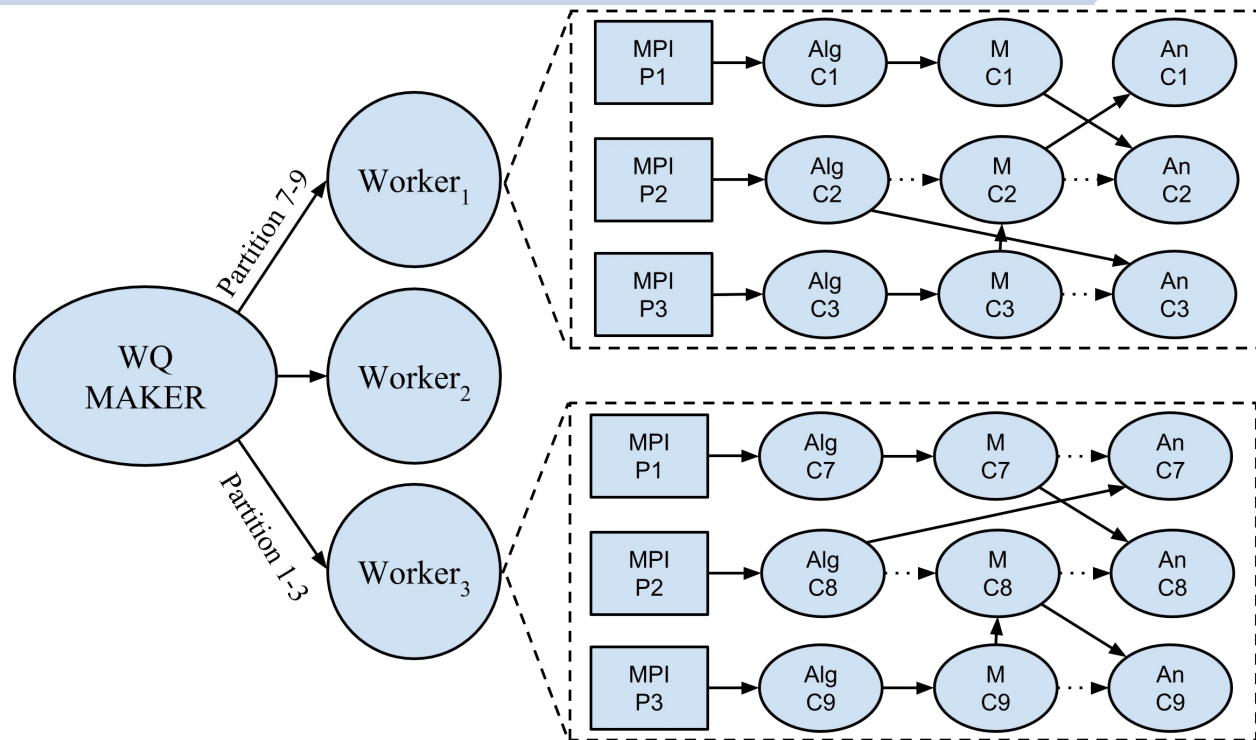


Work Queue



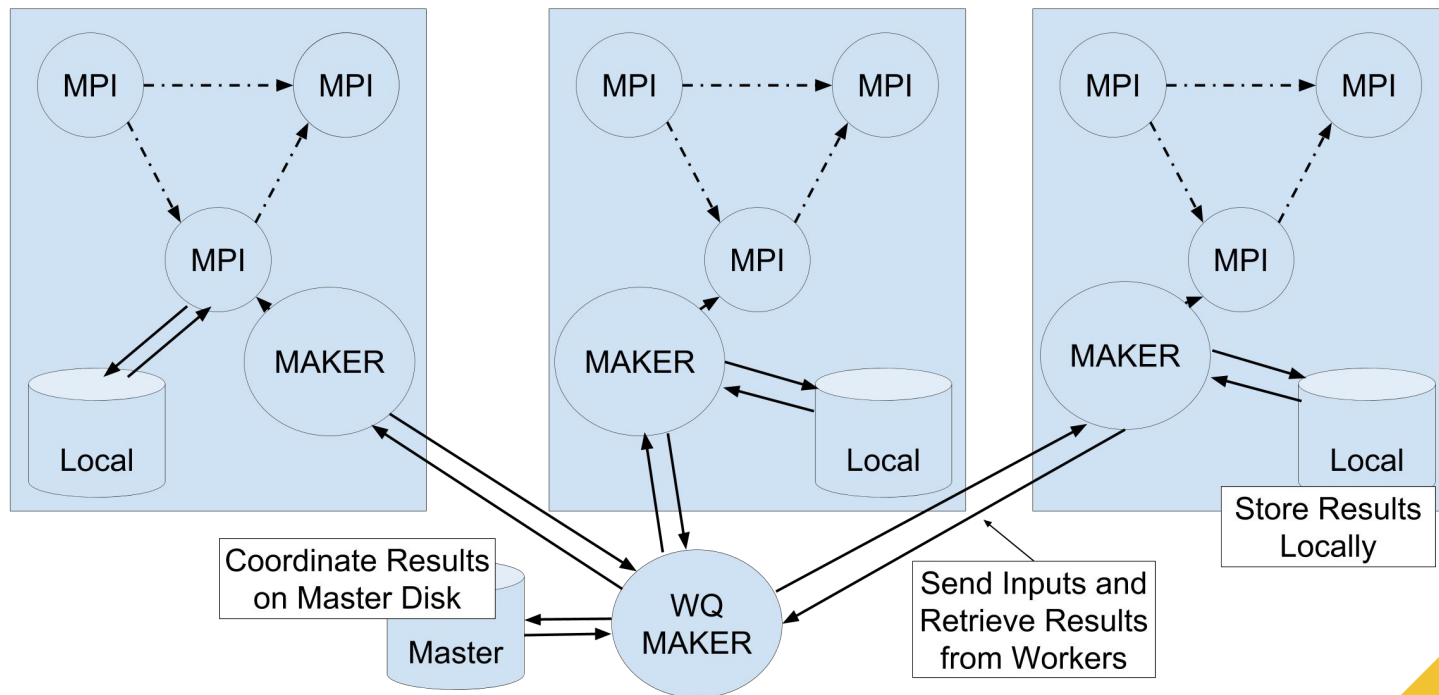


WQ-MAKER



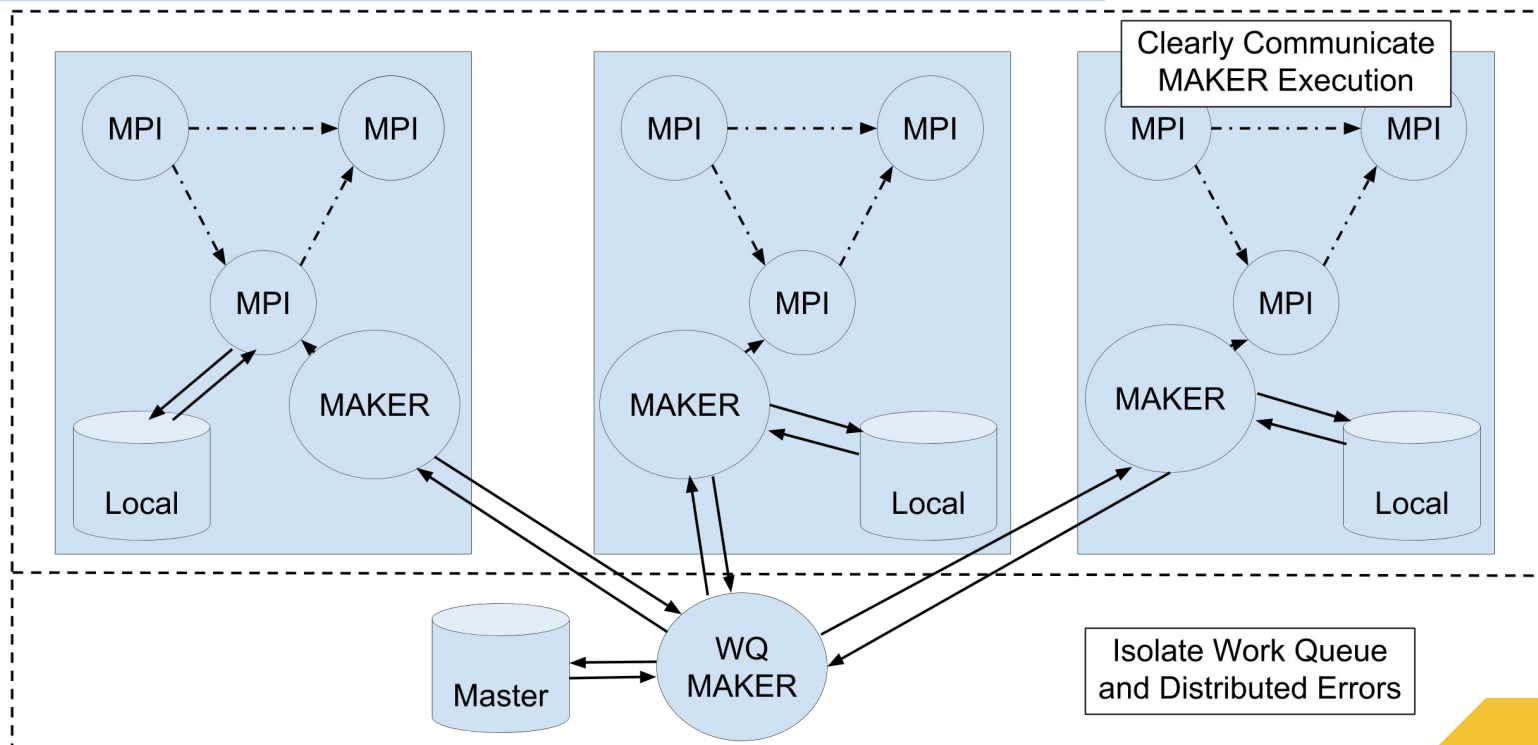


Work Queue Enabling Multi-instance Computation





Providing User Feedback - Debugging





Providing User Feedback - Debugging Examples

Level	Error
MAKER	Invalid input format
	Malformed file names containing “ ” char
	IDs too long for BLAST
Work Queue	Worker times out and disconnects
	Files fail to transfer
	No workers connecting
VC3-Builder	Environment fails to build



Providing User Feedback - Runtime

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	Idled-Out	Lost
10	12	2	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

Work Queue Status

Name: 128.120.146.4
 Project: mseparate
 Owner: jstoppel
 Version: 4.3.2-RELEASE

[More Details](#)

Name: chem0164.ucdavis.edu
 Project: reopt
 Owner: jstoppelman
 Version: 4.3.2-RELEASE

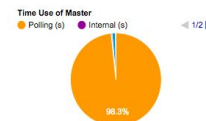
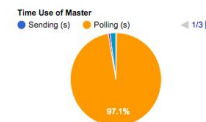
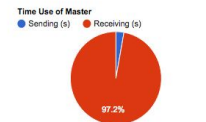
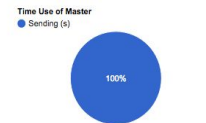
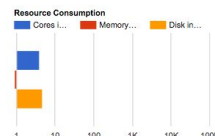
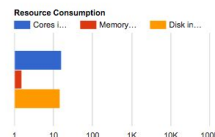
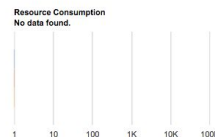
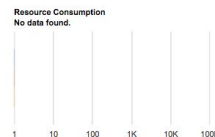
[More Details](#)

Name: earth.crc.nd.edu
 Project: lobster_rbucci_201804
 Owner: rbucci
 Version: 7.0.0 from source

[More Details](#)

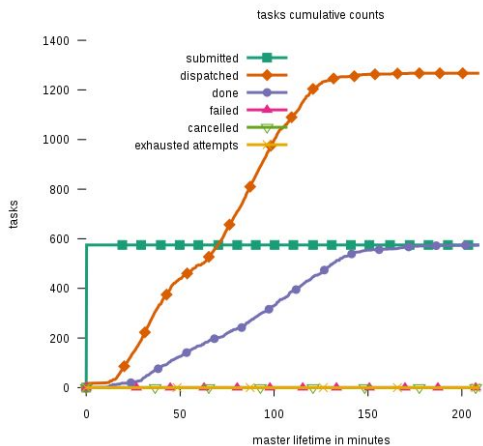
Name: earth.crc.nd.edu
 Project: lobster_rbucci_201804
 Owner: rbucci
 Version: 7.0.0 from source

[More Details](#)

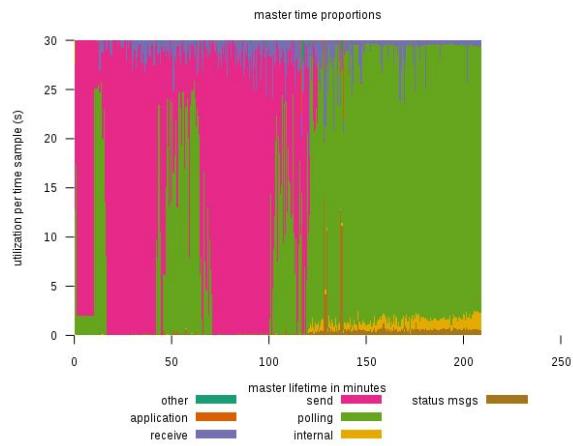




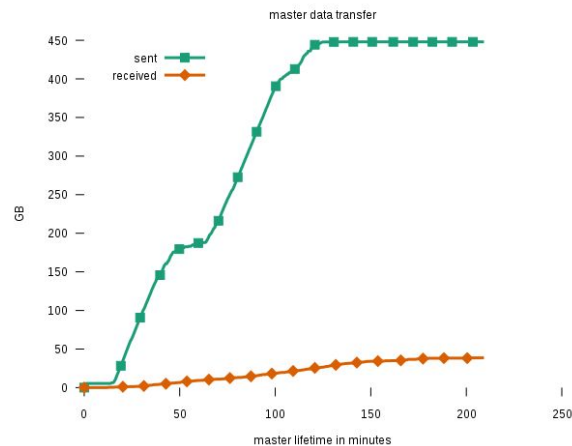
Providing User Feedback - Performance



Task Behavior



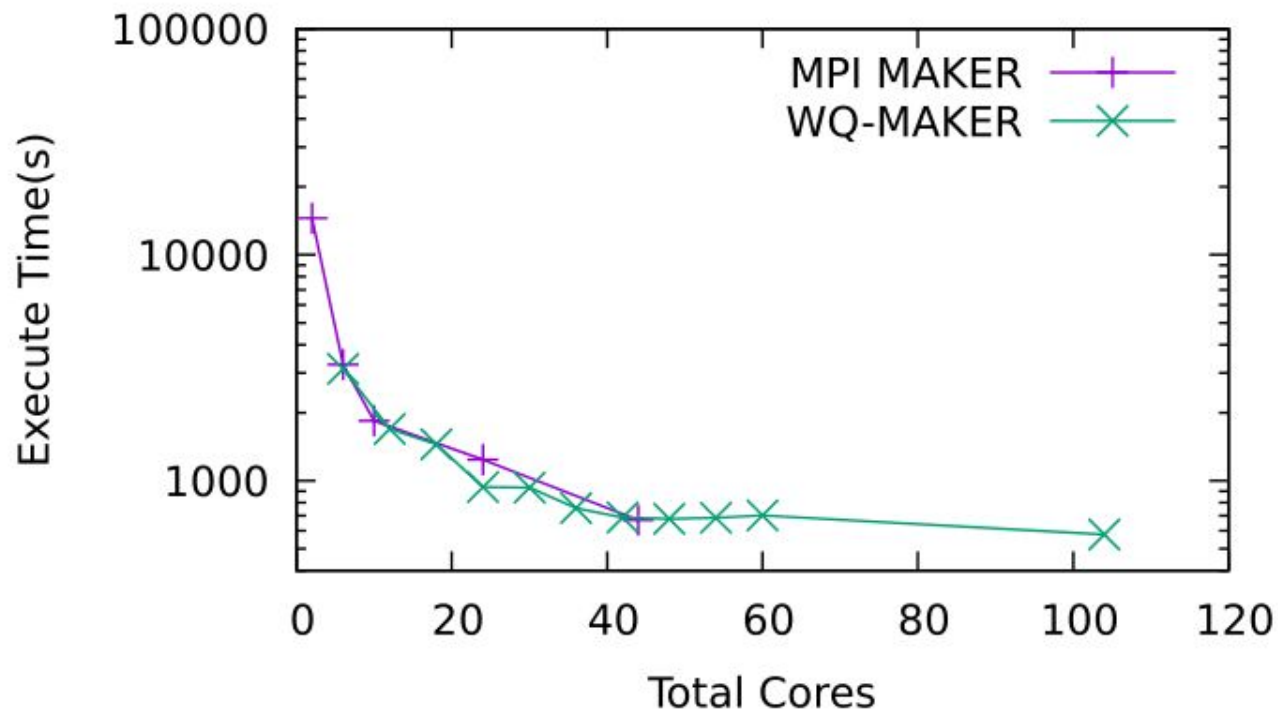
Master Time Usage



Data Transfers

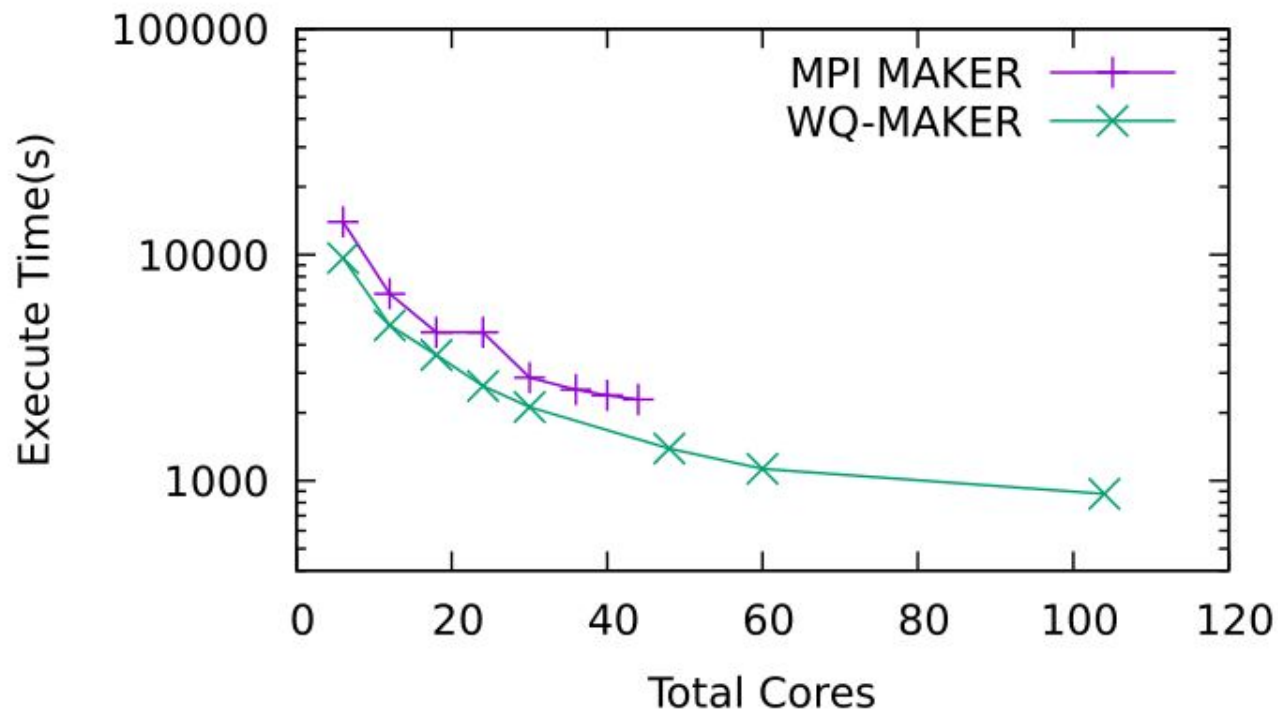


Results - Fungal Genome (41MB)



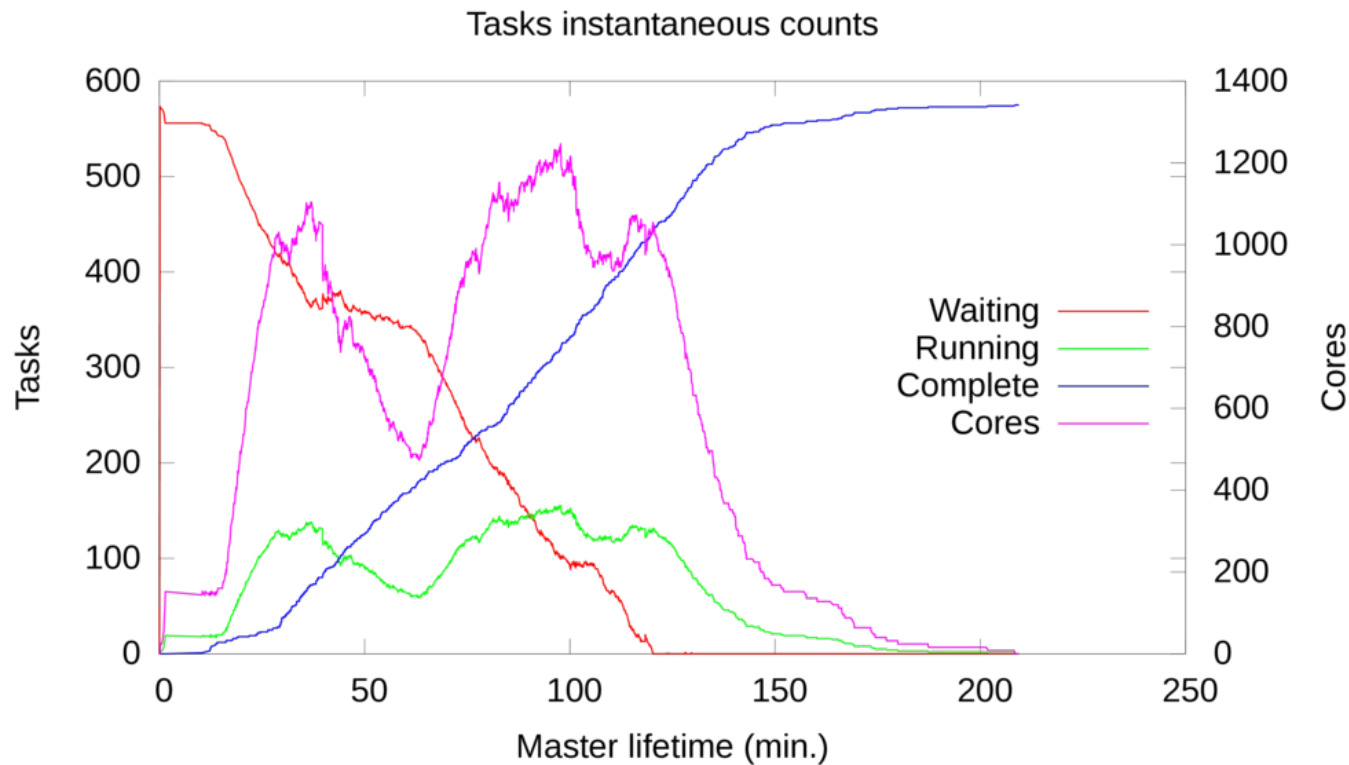


Results - Hummingbird Genome subset (900MB)





Results - Cactus Genome (1.6Gb)





Conclusions

Use a tool capable of providing consistent environment on all target platform.

Leverage multi-instance computation by coupling distributed communications and MPI execution.

Design with feedback in mind to simplify debugging information, give runtime updates, and provide performance analysis.

Questions?

VC3-Builder

VC3 Builder

NAME

vc3-builder – Deploy software environments in clusters without administrator privileges

SYNOPSIS

vc3-builder [options] --require package[:min_version[:max_version]] --require ... [-- command-and-args]

DESCRIPTION

The **vc3-builder** is a tool to manage software stacks without administrator privileges. Its primary application comes in deploying software dependencies in cloud, grid, and opportunistic computing, where deployment must be performed together with a batch job execution.

INSTALLATION

vc3-builder is a self-contained program (including the repository of dependencies recipes). If desired, it can be compiled to a truly static binary (see below).

```
curl -sO https://raw.githubusercontent.com/vc3-project/vc3-builder/master/vc3-builder
chmod 755 vc3-builder
```

or

```
git clone https://github.com/vc3-project/vc3-builder.git
cp vc3-builder/vc3-builder SOME-DESIRED-LOCATION
```

github.com/vc3-project/vc3-builder

CCTools

The Cooperative Computing Lab

[Software](#) | [Download](#) | [Manuals](#) | [Forum](#) | [Papers](#)

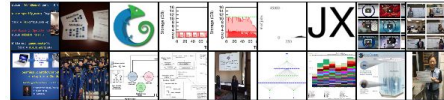
Go to the [CyVerse Container Camp 2018 Tutorial on Makeflow and Work Queue](#), March 9th!

About the CCL

We design **software** that enables our **collaborators** to easily harness **large scale distributed systems** such as clusters, clouds, and grids. We perform fundamental **computer science research** that enables new discoveries through computing in fields such as physics, chemistry, bioinformatics, biometrics, and data mining.

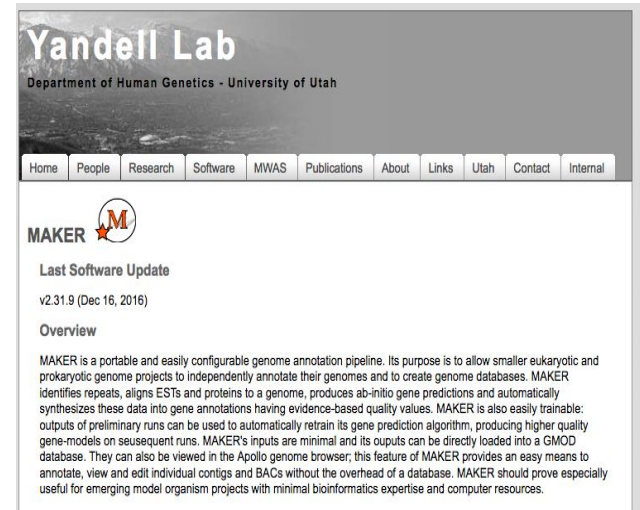
CCL News and Blog

- [CCL at CyVerse Container Camp \(12 Mar 2018\)](#)
- [Submit Your CCL Highlight \(15 Jan 2018\)](#)
- [CCL on Chameleon Cloud with ACIC \(04 Dec 2017\)](#)
- [TPDS Paper: Storage Management in Makeflow \(04 Dec 2017\)](#)
- [CCL at Supercomputing 2017 \(13 Nov 2017\)](#)
- [TPDS Paper: Job Sizing \(26 Oct 2017\)](#)
- [Makeflow Feature: JX Representation \(18 Oct 2017\)](#)
- [Announcement: CCTools 6.2.0 released \(09 Oct 2017\)](#)
- [2017 DISC Summer REU Conclusion \(30 Aug 2017\)](#)
- [\(more news\)](#)



ccl.cse.nd.edu

MAKER



yandell-lab.org/software/maker.html