Lessons Learned Moving MAKER from HPC to the Cloud

Nick Hazekamp\textsuperscript{1}, Upendra Kumar Devisetty\textsuperscript{2}, Nirav Merchant\textsuperscript{2}, and Douglas Thain\textsuperscript{1}
\textsuperscript{1}University of Notre Dame
\textsuperscript{2}University of Arizona
Jetstream Resources Available

No communication between instances

Software stack not installed

Lacks a Shared Filesystem
MAKER Dependencies
VC3-Builder Provided Software Stack

Software stack built in VC3-Builder

No communication between instances

Lacks a Shared Filesystem
MAKER in an HPC Environment

Multi-Machine Partition

- MPI
- Share Work with Messages
- Temporary Files Used
- Local
- Coordinate using Shared Filesystem
- MAKER
- Local
- Shared Filesystem
MPI Run on Single Node

- No communication between instances
- Share Work with Messages
- Temporary Files Used
- Lacks a Shared Filesystem
Work Queue

Submit Tasks

Master

Dispatch Tasks

Worker_1

Send output

Execute Tasks

Task_1

Cache_1

Worker_n

Dispatch Tasks

Task_i

Cache_n
Work Queue Enabling Multi-instance Computation

Diagram showing the setup of MPI (Message Passing Interface) processes and MAKER instances to enable multi-instance computation. The diagram illustrates the coordination of results on a master node, send of inputs and retrieval of results from workers, and storage of results locally.
Providing User Feedback - Debugging

- Isolate Work Queue and Distributed Errors
- Clearly Communicate MAKER Execution
# Providing User Feedback - Debugging Examples

<table>
<thead>
<tr>
<th>Level</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAKER</td>
<td>Invalid input format</td>
</tr>
<tr>
<td></td>
<td>Malformed file names containing “</td>
</tr>
<tr>
<td></td>
<td>IDs too long for BLAST</td>
</tr>
<tr>
<td>Work Queue</td>
<td>Worker times out and disconnects</td>
</tr>
<tr>
<td></td>
<td>Files fail to transfer</td>
</tr>
<tr>
<td></td>
<td>No workers connecting</td>
</tr>
<tr>
<td>VC3-Builder</td>
<td>Environment fails to build</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Type</th>
<th>Success</th>
<th>Failure</th>
<th>Abandon</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tasks</td>
<td>177</td>
<td>12</td>
<td>0</td>
<td>177</td>
</tr>
</tbody>
</table>

Workers:  
- Joined: 10
- Removed: 12
- Idled-Out: 2
- Lost: 2

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
Providing User Feedback - Performance

Task Behavior

Master Time Usage

Data Transfers
Results - Fungal Genome (41MB)
Results - Hummingbird Genome subset (900MB)
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

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 -L https://github.com/vc3-project/vc3-builder/releases/download/v0.1.0/vc3-builder-linux-amd64

or

git clone https://github.com/vc3-project/vc3-builder.git
cp vc3-builder/vc3-builder /PATH-TO-LOCATION

CCTools

The Cooperative Computing Lab
Software | Download | Manuals | Forum | Papers

Go to the CyVerse Container Camp 2018 Tutorial on Makefile and Work Queue, March 9th!

About the CCL
We design software that enables our collaborators to easily leverage large-scale distributed computing, 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, eco-genomics, and data science.

CCL News and Blog
- CCL at CyVerse Container Camp (12 Mar 2018)
- Submit Your CCL Proposal (15 Jan 2018)
- CCL on OpenStack Cloud with ANSYS (04 Dec 2017)
- TDG Paper: Storage Management in Makefile (04 Dec 2017)
- CCL at Supercomputing 2017 (13 Nov 2017)
- TDG Paper: Makefile (08 Oct 2017)
- Makefile Template: IX Representation (18 Oct 2017)
- Announcement: CCTools 0.3.2 released (09 Oct 2017)
- 2017 USC Summer REU Conclusions (07 Aug 2017)
- (GEMC BLOG)

ccl.cse.nd.edu

MAKER

Yandell Lab
Department of Human Genetics • University of Utah

Last Software Update
v2.31.9 (Dec 16, 2016)

Overview
MAKER is a portable and easily configurable genome annotation pipeline. Its purpose is to allow smaller suburban and prairie genome projects to independently annotate their genomes and to create genome databases. MAKER identifies repeats, aligns ESTs and proteins to a genome, produces ab initio gene predictions and automatically organizes these data into gene annotations having evidence-based quality values. MAKER is also easily portable; outputs of preliminary runs can be used to automatically refit its gene prediction algorithm, producing higher-quality gene models on subsequent runs. MAKER's inputs are minimal and its outputs can be directly loaded into a UCSC database. They can also be viewed in the ApoLogiX genome browser; this feature of MAKER provides an easy means to evaluate, view and edit individual contigs and BACs without the overhead of a database. MAKER should prove especially useful for emerging model organism projects with minimal bioinformatics expertise and computer resources.

yandell-lab.org/software/maker.html

Nicholas Hazekamp
Email: nhazekam@nd.edu