A Compiler Toolchain for Workflows

Peter Bui, University of Notre Dame
CCL Workshop, June 2012
Abstractions for Software

All-Pairs (Regular Graph)

Makeflow (Irregular Graph)

Work Queue (Dynamic Graph)

```plaintext
while( more work to do )
{
    foreach work unit {
        t = create_task();
        submit_task(t);
    }
    t = wait_for_task();
    process_result(t);
}
```
Makeflow

- Makeflow is a workflow engine for executing large complex workflows on clusters, clouds, and grids.

- Makeflow is similar to traditional Make.
  - Supports restricted syntax for clarity.
  - Uses transaction journal to store provenance.

- Makeflow does not require a distributed filesystem.

http://www.nd.edu/~ccl/software/software/makeflow
Scientific Workflows
Compiler Toolchain For Workflows

- Applying techniques from traditional compilers and programming languages to the construction of distributed workflows.
  - **Compiler**: generating workflows from a high-level programming language.
  - **Linkers**: packaging and modifying workflows.
  - **Profilers**: analyzing and monitoring workflows.
  - **Run-time**: executing and managing workflows.
Features Coming Soon in 3.6.0!

- Local variables
- Nested Makeflow command
- Enhanced provenance
- Garbage collection
- Linking workflows
Local Variables

- Add **lexical scoping** to Makeflow script to allow for variables local to task.

**uname.disc**: /bin/uname
   ```
   @BATCH_OPTIONS=requirements = MachineGroup == "disc"
   /bin/uname -a > uname.disc
   ```

**uname.ccl**: /bin/uname
   ```
   @BATCH_OPTIONS=requirements = MachineGroup == "ccl"
   /bin/uname -a > uname.ccl
   ```

**uname.gh**: /bin/uname
   ```
   @BATCH_OPTIONS=requirements = MachineGroup == "gh"
   /bin/uname -a > uname.gh
   ```
Nested Makeflow Command

- Add **MAKEFLOW** command to allow Makeflow to recognize when a task is a nested workflow.

```
MAKEFLOW "<dag_path>" "<dag_work_dir>" "<dag_wrapper>"
```
Enhanced Provenance

- In addition to recording workflow events, the transaction journal now also:
  - Embeds DAG.
  - Records Workflow Status.
  - Includes Debugging Symbols.
log.path = Makeflow.makeflowlog
log.starts = 1336672272.29
log.failures =
log.abortions =
log.completions = 1336672272.6
log.elapsed_time = 0.303680896759
log.percent_completed = 100.0
log.average_tasks_per_second = 59.2727438311
log.current_tasks_per_second = 0
log.estimated_time_left = None
log.state = completed
log.finished = True
log.goodput = 0.57969045639
log.badput = 0
log.nodes.waiting = 0
log.nodes.running = 0
log.nodes.completed = 18
log.nodes.failed = 0
log.nodes.aborted = 0
log.nodes.retried = 0
log.nodes.total = 18
Makeflow Monitor

Makeflow: nostash.1.makeflowlog
  Status: Completed
  Time Started: 09:39:46
  Time Elapsed: 32:12
  Average Tasks/Minute: 3104.44
  Tasks: Waiting: 0, Running: 0, Completed: 100000, Failed: 0,
         Aborted: 0, Retried: 0, Total: 100000

Makeflow: nostash.2.makeflowlog
  Status: [=======================================] 75.02%
  Time Started: 10:12:08
  Time Elapsed: 25:42
  Average Tasks/Minute: 2917.43
  Current Tasks/Minute: 2853.65
  Estimated Time Left: 08:45
  Tasks: Waiting: 24968, Running: 15, Completed: 75017, Failed: 0,
         Aborted: 0, Retried: 0, Total: 100000
Makeflow Report

- Statistical report of workflow that includes:
  - Workflow Summary.
  - Tasks Summary, Profiling, Histogram.
  - Symbols Summary, Profiling.
Implemented garbage collection methods to handle temporary intermediate files.

![Graph comparing execution time across different filesystems and collection methods.]
Linking Workflows

● Application Linker

○ Convert absolute paths to relative paths.
○ Copy or symlink data and executables to sandbox.
○ Automatically starch executables.

● Workflow Linker
Summary

● Makeflow allows users to construct large data intensive scientific workflows.
  ○ Currently in use by CCL and collaborators at the University of Notre Dame.
  ○ **Ongoing work** to improve Makeflow and add new features.
  ○ Growing **ecosystem** of tools for generating, managing, executing, and analyzing workflows.
Questions?

- Are these features useful to you?
- What new features would you like to see?
- How can we help you accomplish your research goals?

http://www.nd.edu/~ccl/software/makeflow