

# **A workflow for implementing HTCondor Pipelines**

First Mondays: November 6, 2017

**Problem:** Building binaries for CHTC is problematic.

**Workflow:** Two ideas to alleviate the pain:

1. Build on a CentOS 7 platform that's more friendly to users.
2. Use Git to manage the environment.

**High-level view:**

Step 1: Get the binary running on pink.botany.wisc.edu. Capture the environment in a (local) git repository.

Step 2: Propagate the git repository to a "remote" server.

Step 3: Fire up an interactive HTCondor job, as usual.

Step 4: Clone the repo on the execute node and test. Push changes to the remote repo.

Step 5: Deploy as a noninteractive job, as usual.



# A workflow for implementing HTCondor Pipelines

First Mondays: November 6, 2017

**Problem:** Building binaries for CHTC is problematic.

**Workflow:** Two ideas to alleviate the pain:

1. Build on a CentOS 7 platform that's more friendly to users.
2. Use Git to manage the environment.

## Some Details:

**Step 1:** Get it running on pink.

#### Ask Steve or Henry to set up a login.

1. Connect to pink:

```
ssh pink.botany.wisc.edu
```

2. Make directory and cd into it.

3. Initialize a git repo:

```
git init
```

4. Follow the directions for building bwa:

<https://github.com/lh3/bwa>

*Hint: google "bwa github"*

5. Test the binary.

- a. Run the command:

```
./bwa
```

- b. Construct test data.

```
head -10 reference.fasta >
```

```
test/indir/reference.test.fasta
```

```
head -1000 reads.fastq > test/indir/reads.test.fastq
```

## Commit to the repo early and often!

```
git status
```

```
git add
```

```
git commit -m "description"
```

**Step 2:** Propagate to remote repository. (Actually, this is also on pink – but it could be on another server, e.g. github.com.)

1. Create a home for your remotes.

```
mkdir ~/gitRepos
```

```
cd ~/gitRepos
```

```
git --bare init runbwa.git
```

*Hint: google "git setup bare repo"*

<https://gist.github.com/joahking/780877>

**2. Connect your local git repo to this new remote:**

```
cd -      ### change to the previous dir!
git remote add origin ~/gitRepos/runbwa
git push origin master
```

**Step 3: Fire up an interactive job, as usual.**

```
condor_submit -i queue/submit/interactive.submit
```

*Hints: Use this as a starting template.*

<http://chtc.cs.wisc.edu/inter-submit.shtml>

*Use screen session to launch a job that persists.*

```
screen -S "name"
```

**Step 4 (method 1): Go to town in the interactive job.**

```
./bwa
./bwa index indir/reference.fasta
./bwa mem indir/reference.fasta indir/reads*.fastq > pe.sam
```

**Step 4 (method 2): Go to town in the interactive job.**

```
git clone pink.botany.wisc.edu:gitRepos/runbwa
```

```
./bin/bwa
./bin/bwa index test/indir/reference.fasta
```

etc.

Continue to develop your pipeline in the interactive job and then push to the remote.

```
git push origin master
```

**You can clone Steve's repo:**

```
git clone
pink.botany.wisc.edu:/home/sgoldstein/gitRepos/runbwa
```