

# Pipeline II

## Pipeline II

Pipeline II is an upgrade of the existing pipeline server. It likely to be a somewhat scaled back version of our earlier plans for [Pipeline NG](#). It should use a database schema similar to the current pipeline, but extended to handle new requirements.

### Documentation

- [Pipeline II User's Guide](#)
- [Pipeline II Frequently Asked Questions \(FAQ\)](#)
- [Pipeline User's Group](#)
- [Pipeline II Administrators Guide](#)
- [Pipeline II Release Notes](#)
- [Pipeline II Recent Activity](#)
- [Development notes](#)

### Requirements

- Task scheduling should be more flexible than current linear chain
  - Should support parallel execution of tasks
  - Should allow dependency chain to be more general than the input file requirements
  - Should support parallel sub-tasks, with number of sub-tasks defined at runtime
  - Perhaps support conditions based on external dependencies
- Should allow for remote submission of jobs
  - Perhaps using GRID batch submission component, or Gluster specific batch submission system
  - Will need to generalize current system (e.g. get rid of absolute paths)
- Support reprocessing of data without redefining task
  - Need way to mark Done task as "ReRunnable"
  - Need to support multiple versions of output files
- Ability to Prioritize tasks
- Ability to work with "disk space allocator"
- Would be nice to set parameters (env vars) in task description
- Would be nice to be able to pass in parameters in "createJob"
- Ability to suspend tasks
- Ability to kill tasks
- Ability to throttle job submission (ie max number of jobs in queue)
- Ability to map absolute path names to FTP path names (site specific)
- Would be nice to remove need for "wrapper scripts"
- Ability to specify batch options (but portability problems)

For more details see [Talks at Developers Workshop](#)