Interprocess message specifications

V 1.0

The major revision number of this document should match the major revision number of the associated flowchart.

NOTE: As of Feb 1, 2014, we've switched around our job/task terminology so that it (hopefully) makes more sense. **A *Job* is composed of 6k *Tasks*. Users submit *Jobs*; worker nodes process *Tasks*.** This change has been made in both this document and the accompanying diagram document.

Job Queue Message Specification

```
{
  "JobID" : "My Job ID",
  "operations" : {},
  "??" : "??"
}
```

- *Job ID* is a UUID, generated by the job-initiating application, which will be used to identify this job from start to finish. All tasks, results, and reports generated from this Job will be tagged with this ID.
- *operations* is a hash such as we provide to the management ops validator

What else should go in here?

- Something providing field geometry, etc. ?
- A *Final Action*, such as a url to post report to? Email address? Something indicating where results are sent or how the client is notified that the task is done ?

Report Gen Queue Message Specification

```
{
  "Job ID" : "My Job ID",
  "Task IDs" : ["TaskID_0","TaskID_1","TaskID_...","TaskID_5999"] ,
  "Final Action" : ""
}
```

- *Job ID* is the same Job ID as in the Job Queue. This is the name of the queue that report handler's subprocess for this job will poll for task results
- *Task IDs* is an array of unique task ids, each of which is a UUID generated by the GIS Query Node
- *Final Action* (or something similar). See comment in Job Queue section

Task Queue Message Specification

```
---
EZQ:
  "result_queue_name" : "My Job ID"
...
{
  "Task ID" : "The UUID for this task",
  "mmp360 input data" : {}
}
```

- The initial block between --- and … is a YAML block containing EZQ directives. *result_queue_name* tells EZQ to place the results of this task into the named queue. This name should be set to the Job ID, as that is the name of the results queue created by GIS Query Node for this job
- *Task ID* is a UUID matching one of those in the Task IDs array of a Report Gen Queue message
- *mmp360 input data* is whatever data structure is needed for a single mmp360 Worker Node to process this task. Depending on the size of this structure, we may need to have GIS Query Node write out the task input details to S3

and place “bucket” and “key” directives outside of the userdata object. In that case, the *mmp360 input data* field should be removed from this specification.

Results Queue Message Specification

```
{
  "bucket" : "Name of an S3 bucket",
  "key" : "Key of the file containing the full results of this task",
  "Task ID" : "The UUID for this task"
}
```

- *bucket* is the name of the S3 bucket that workers are told to store their results in. That bucket name is set at initial configuration of the worker node AMI. We repeat it here so that EZQ can handle fetching the data automatically. If it turns out that the result data from a worker will always be fairly small (<256k), we can skip caching results in S3 and simply place them directly in the userdata portion of the message.
- *key* is the name of the specific file containing this task's results
- *Task ID* is the task id associated with these results

pf_msg Specification

```
"push_file: bucket,filename"
```

This message is placed on STDOUT by the Pre-Grid program, indicating a that the file *filename* placed on local storage should be uploaded to the S3 bucket *bucket*.

em_msg Specification

```
"enqueue_msg: {...message...}"
```

This message is placed on STDOUT by the Pre-Grid program, indicating the JSON message *{...message...}* should be placed into the Task Queue.