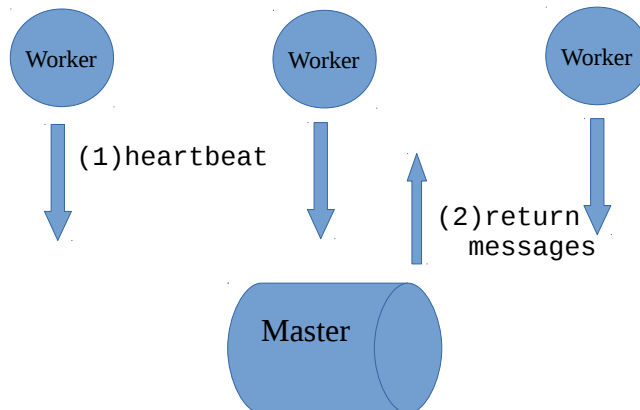


# Design Patterns

## (1) heartbeat



This is the workflow for heartbeats. Heartbeats messages are initiated by workers. The master just received heartbeats messages passively, and attach task information in return messages to inform workers what to do.

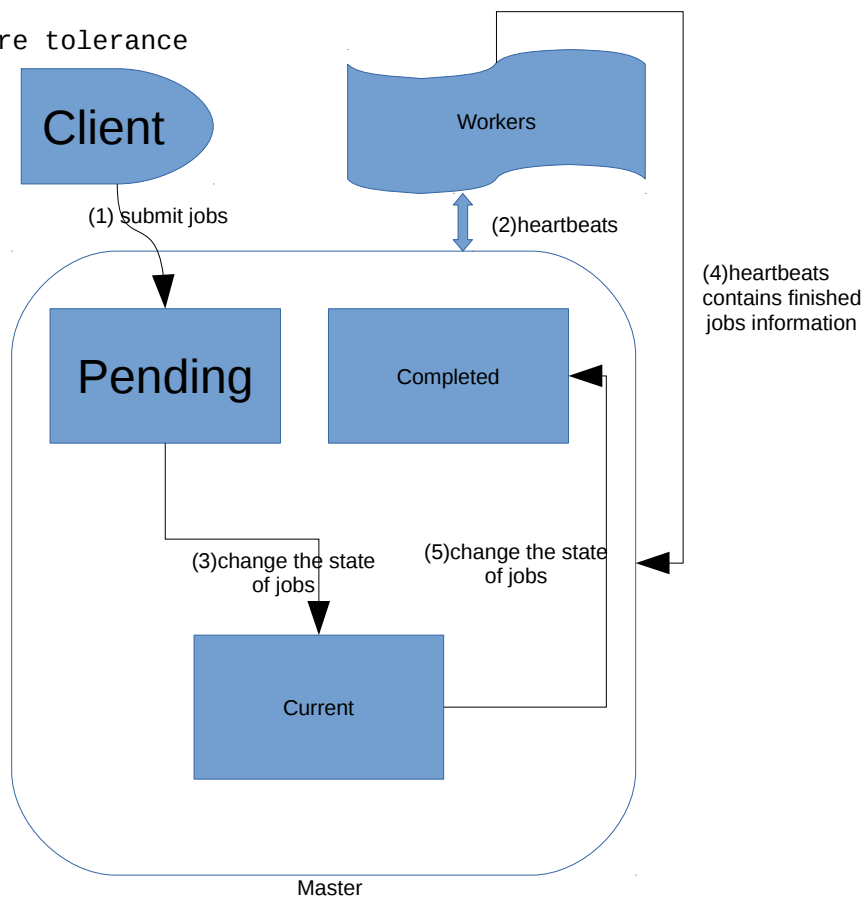
The format of task information sent to workers is:

`job_name, input_file, start_index, chunk_size, number of reducers`

## (2) detect worker failure

The master maintains a dictionary called **workers\_info** to track all workers information. Occasionally, the master will examine the time of last received heartbeat of all workers. If  $\text{current\_time} - \text{last\_received\_time}$  is greater than **TIMEOUT**, the master will remove the worker from **workers\_info**.

## (3) failure tolerance



Above is how the state of jobs changes. The master keeps tracking the state of jobs in order to recover a job when a worker is down. The state of jobs is stored in data structures as below:

```
completed_map_jobs = set()
current_map_jobs = set()
pending_map_jobs = set()

pending_reduce_jobs = set()
completed_reduce_jobs = set()
```

## Experimental Results

Below is the result of both sequential execution and distributed execution after running sorting operation on a 20K text file. There are two workers for distributed execution.

```
[yhuang103@bass06 Default-project2]$ time /opt/python2.7/bin/python2.7 mr_seq.py Sorting 20 1 sort.txt Sort
----- 0 -----
real    0m13.052s
user    0m0.454s
sys     0m0.471s
```

```
[yhuang103@bass06 Default-project2]$ time /opt/python2.7/bin/python2.7 client.py Sorting sort.txt 20 1 Sort
real    0m6.189s
user    0m0.124s
sys     0m0.046s
[yhuang103@bass06 Default-project2]$
```

This is the result of another experiment after running WordCount task on a 71MB text file. There are two workers for distributed execution.

```
[yhuang103@bass06 Default-project2]$ time /opt/python2.7/bin/python2.7 mr_seq.py WordCount 20000000 4 test.txt WordCount
real    1m52.953s
user    0m53.697s
sys     0m2.227s
[yhuang103@bass06 Default-project2]$
```

```
[yhuang103@bass06 Default-project2]$ time /opt/python2.7/bin/python2.7 client.py WordCount test.txt 20000000 4 WordCount
real    0m24.227s
user    0m0.141s
sys     0m0.045s
[yhuang103@bass06 Default-project2]$
```