Group 25

Assignment 3 Report

1. Design questions:
   - How does the leader process know the number of alive workers?
     The coordinator adds a new worker to the poll after accepting the connection. Therefore, all alive nodes are monitored by the poll.

   - How can we distribute work "fairly" among workers?
     The coordinator goes through the jobs list and assigns the jobs for active workers one by one. Therefore, every worker fairly receives the same workload.

   - With what messages do leader - worker communicate?
     The coordinator sends two consecutive messages to the worker for the URL length and the actual URL. The worker can read the exact URL based on URL length. After processing the work, the worker sends back the result to the coordinator.

   - How can we detect failed / crashed workers?
     The poll will throw error in case a worker is down, the coordinator just needs to handle it.

   - How do we recover when a worker fails?
     The coordinator can simply push back the incomplete jobs from failed worker to the list and distributes them to active nodes. Failure should not affect the whole process.

2. Scalability questions:
   Execution time with different number of workers can be visualized in the following graphs:

- What is the limit in scaling? (network, CPU-bound)
  The CPU can execute a maximum number of operations per cycle, we can't exceed this limitation. We can also be limited by the memory and network speed. When the workload is too little, the overhead of starting and terminating the worker can bring down the performance.

- Measure each worker's load - how is load balancing affected by scale?
  The workload of each worker decreases when more workers are added to the pool as the total workload is fairly distributed over the workers.

- Could you think of a case when the coordinator would become a bottleneck in such a system?
  The coordinator must distribute workload and collect results from the worker, the coordinator can become a bottle neck if many workers send the result at the same time and the coordinator needs time to read and check and summarize the result.