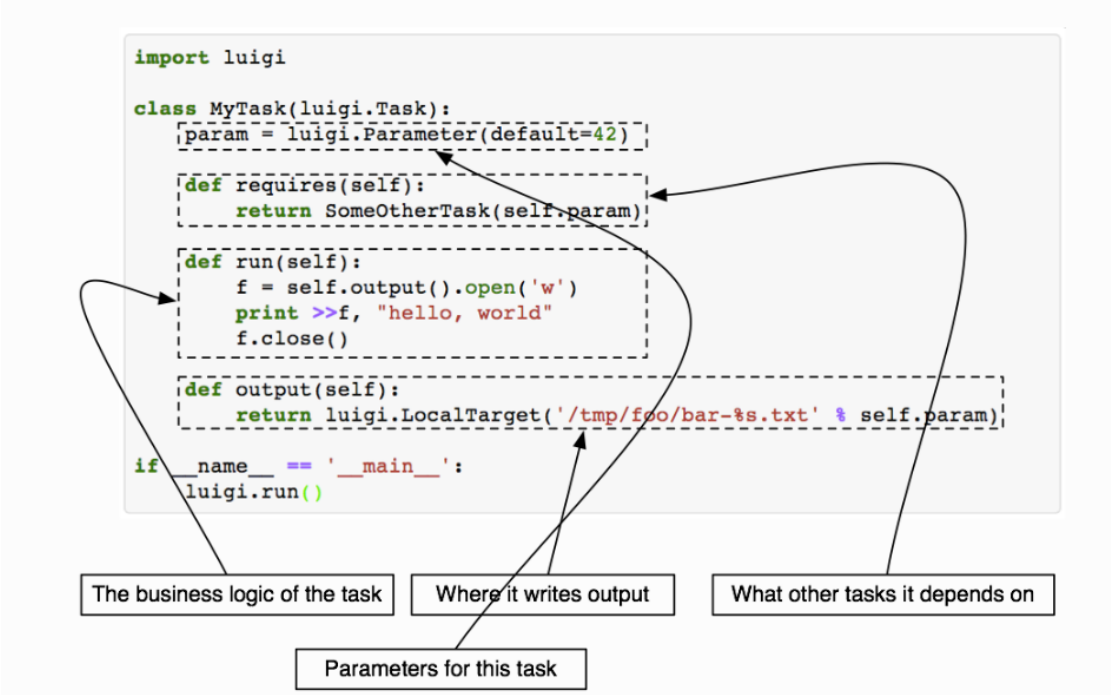
Luigi

Luigi is a Python package that helps you build complex pipelines of batch jobs. It handles dependency resolution, workflow management, visualization, handling failures, command line integration, and much more.

There are two fundamental building blocks of Luigi - the Task class and the Target class. Both are abstract classes and expect a few methods to be implemented. In addition to those two concepts, the parameter class is an important concept that governs how a Task is run.

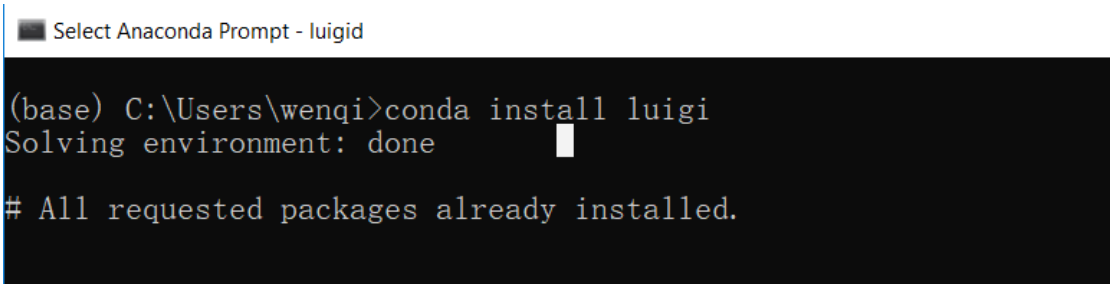
The Target class corresponds to a file on a disk, a file on HDFS or some kind of a checkpoint, like an entry in a database. Actually, the only method that Targets have to implement is the existsmethod which returns True if and only if the Target exists.Tasks are where the execution takes place. Tasks depend on each other and output targets. Here is a example of a Task class:



Here is a demo of how to use it:

1. Install luigi package

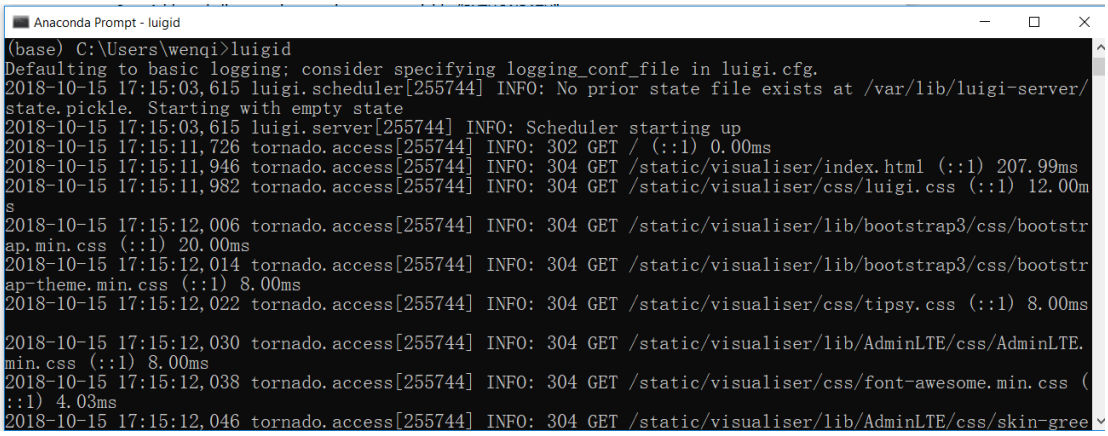
Input *conda install luigi* in anaconda prompt

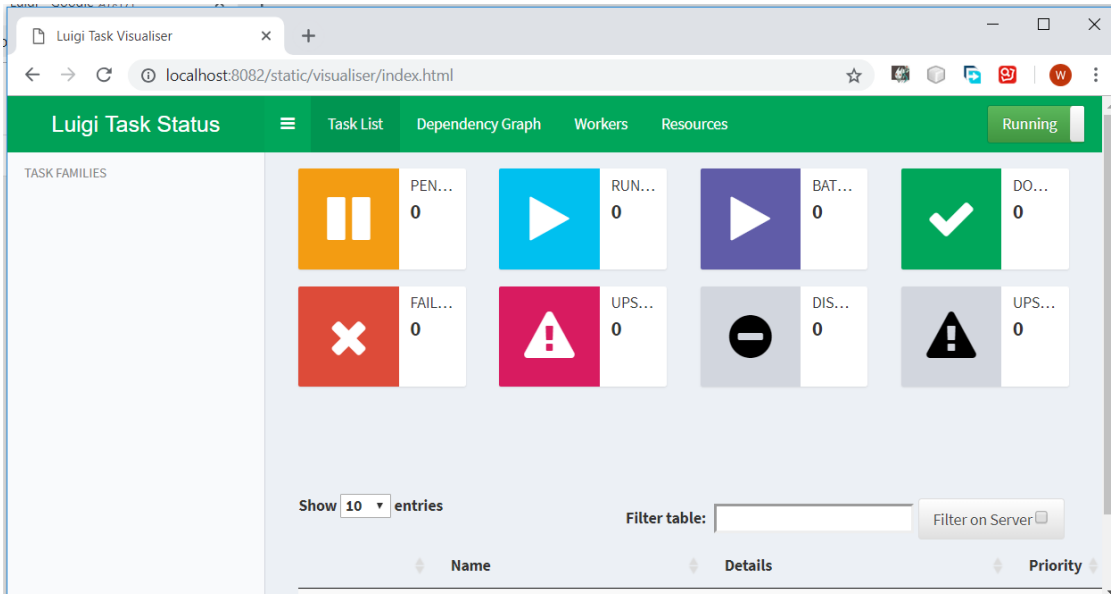


2. Add work directory into environment variable “PYTHONPATH”

3. Check the visualizer page

Input *luigid* in anaconda prompt, then check visualizer page in http://localhost:8082





4. Create a python file named “example.py”

```
import time
import luigi
from luigi.local_target import LocalTarget
class RunAllTasks(luigi.Task):
    def requires(self):
        for i in range(10):
            yield RunExampleTask(i)
    def run(self):
        with self.output().open('w') as f:
            f.write('All done!')
    def output(self):
        return LocalTarget('tmp/RunAllTasks.txt')
class RunExampleTask(luigi.Task):
    number = luigi.IntParameter()
    def run(self):
```

```
            time.sleep(self.number)

            with self.output().open('w') as f:
                f.write("This is task # {}".format(self.number))
        def output(self):
            return LocalTarget('tmp/runExampleTask_{}.txt'.format(self.number))
if __name__ == "__main__":
        luigi.run()
```

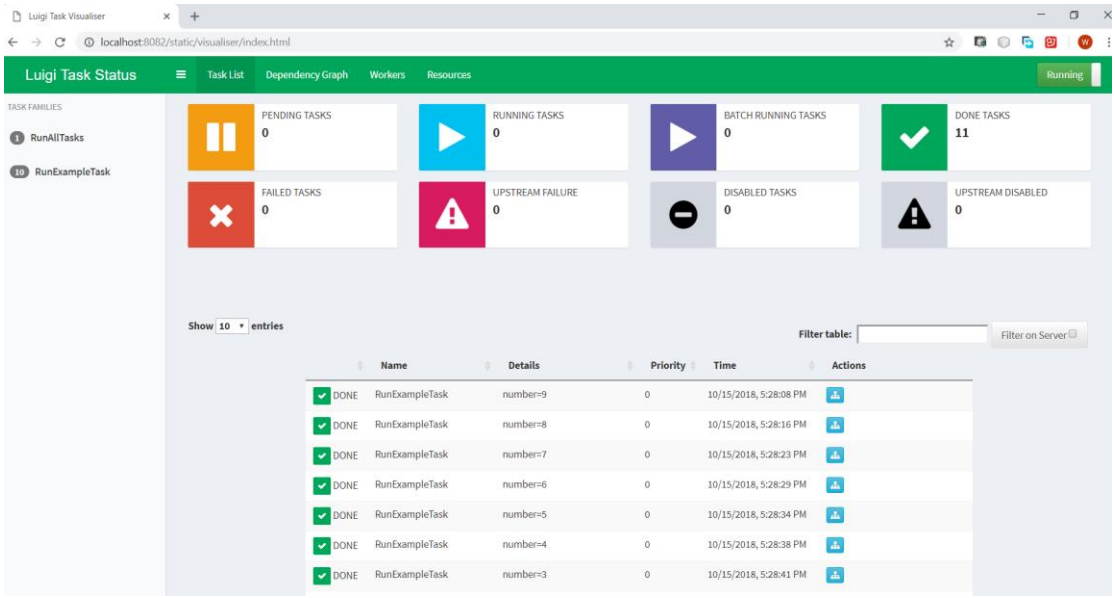
5. Change directory to your work directory, then run the task

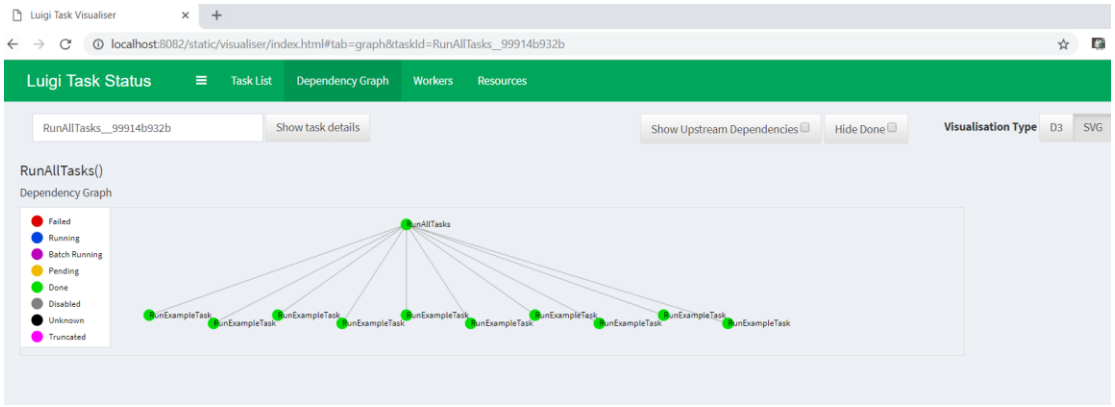
Anaconda Prompt

```
(base) C:\Users\wenqi>cd C:\Users\wenqi\Desktop\7390\luigi

(base) C:\Users\wenqi\Desktop\7390\luigi>luigi --module example RunAllTasks
```

6. We can check the task status in visualizer page





Reference：

Luigi documentation :https://luigi.readthedocs.io/en/stable/

Luigi github: https://github.com/spotify/luigi