Framework Luigi
para ETL com Python

Prof. Dr. Diego Bruno

Education Tech Lead na DIO
Doutor em Robótica e *Machine Learning* pelo ICMC-USP

Framework de ETLs

Prof. Dr. Diego Bruno

Machine Learning

Framework

Luigi é um framework de execução criado pelo Spotify que cria pipelines de dados em Python. É um pacote Python (2.7, 3.6, 3.7 testado) que ajuda a construir pipelines complexos de trabalhos em lote. Ele lida com resolução de dependências, gerenciamento de fluxo de trabalho, visualização, tratamento de falhas, integração de linha de comando e muito mais.

Framework

Luigi é um framework de execução criado pelo ***Spotify*** que cria pipelines de dados em Python. Em tese, é um pacote Python (2.7, 3.6, 3.7 testado) que ajuda a construir pipelines complexos de trabalhos em lote. Ele lida com resolução de dependências, gerenciamento de fluxo de trabalho, visualização, tratamento de falhas, integração de linha de comando e muito mais.

Luigi

Tópicos

***Target*:** Em palavras simples, um alvo contém a saída de uma tarefa. Um destino pode ser um local (por exemplo: um arquivo), (MySQL etc);

Tópicos

***Task*:** - Tarefa é algo onde o trabalho real ocorre. Uma tarefa pode ser independente ou dependente. O exemplo de uma tarefa dependente é despejar os dados em um arquivo ou banco de dados. Antes de carregar os dados, os dados devem estar lá por qualquer meio (*scraping*, API, etc). Cada tarefa é representada como uma classe Python que contém certas funções-membro obrigatórias. Uma função de tarefa contém os seguintes métodos:

Tópicos

***Task*:**

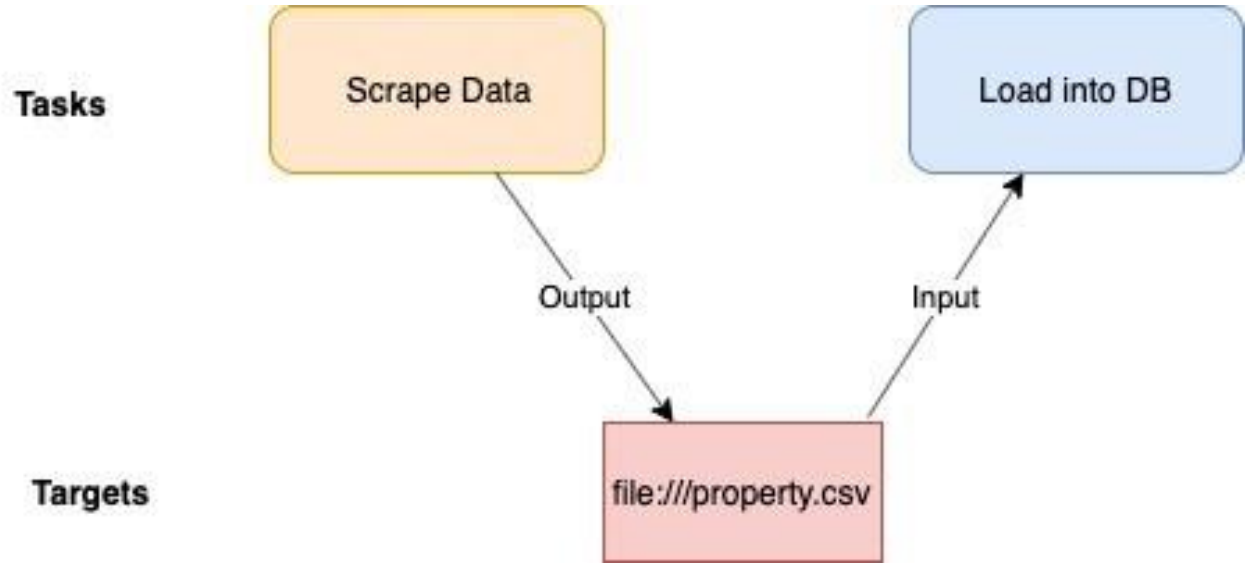
require():- Esta função membro da classe *task* contém todas as instâncias de tarefas que devem ser executadas antes da tarefa atual.

output():- Este método contém o destino onde a saída da tarefa será armazenada. Isso pode conter um ou mais objetos de destino.

run():- Este método contém a lógica real para executar uma tarefa.

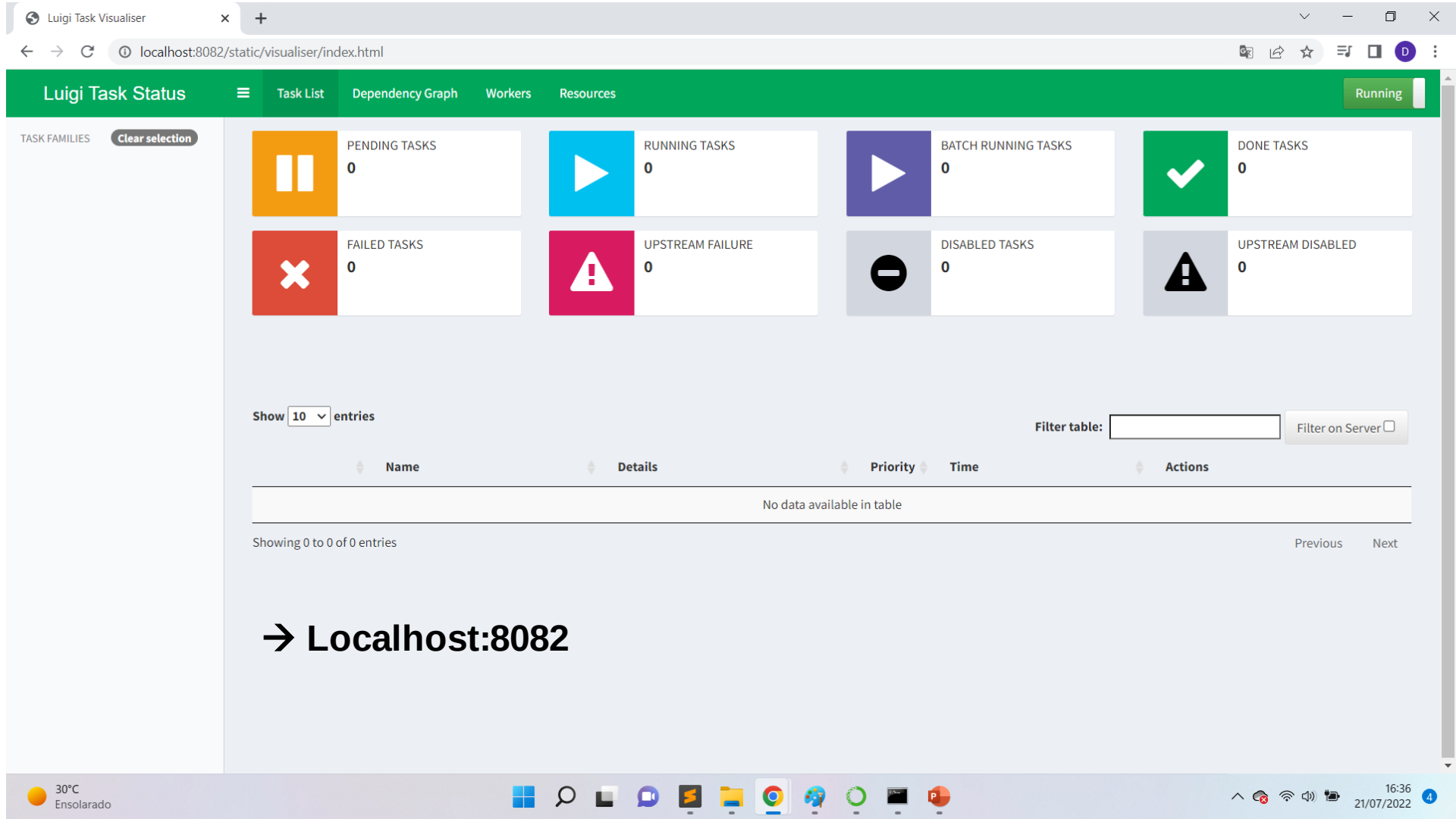
Tópicos

A representação do processo será algo como abaixo:



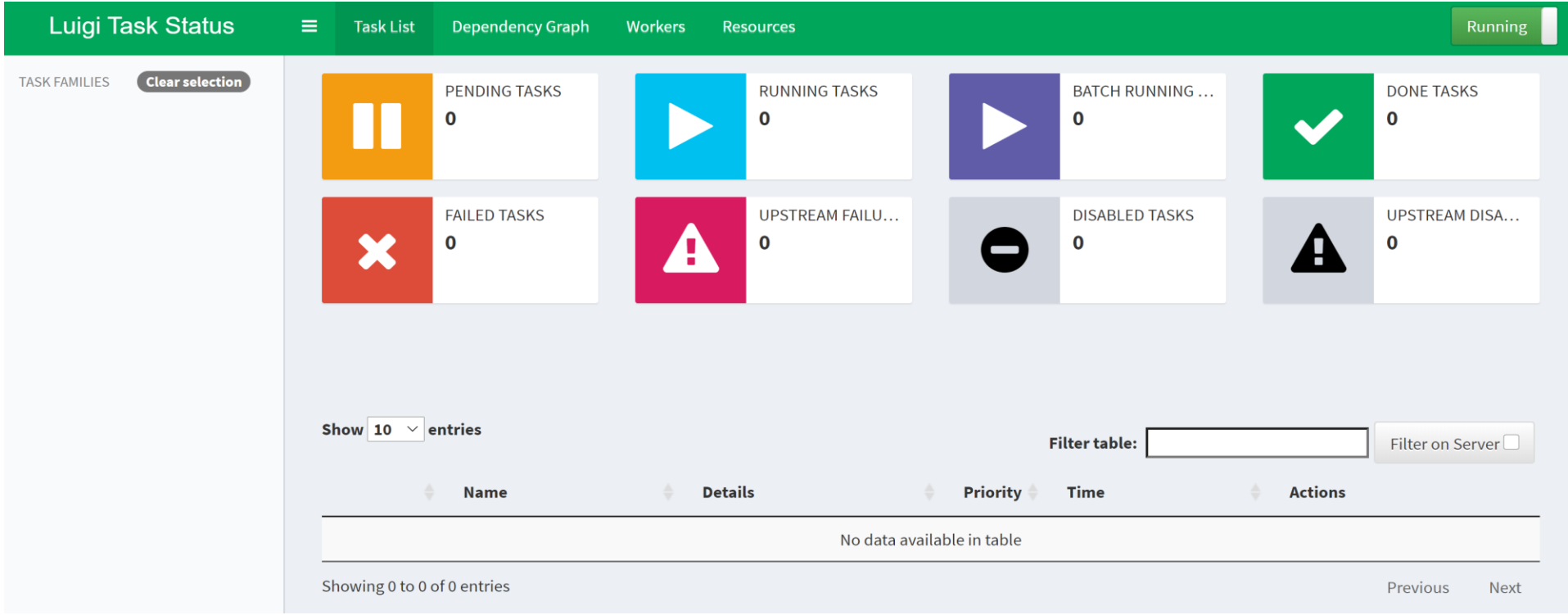
C:\Windows\system32\cmd.exe - luigid

```
TL-Python-Luigi) C:\Users\diego>luidid
uidid' não é reconhecido como um comando interno
 externo, um programa operável ou um arquivo em lotes.

TL-Python-Luigi) C:\Users\diego>luigid
22-07-21 16:23:35,144 luigi[13052] INFO: logging configured by default settings
22-07-21 16:23:35,146 luigi.scheduler[13052] INFO: No prior state file exists at /var/lib/luigi-server/state.pickle
Starting with empty state
22-07-21 16:23:35,156 luigi.server[13052] INFO: Scheduler starting up
22-07-21 16:24:34,950 tornado.access[13052] INFO: 302 GET / (::1) 1.00ms
22-07-21 16:24:35,086 tornado.access[13052] INFO: 200 GET /static/visualiser/index.html (::1) 132.56ms
22-07-21 16:24:35,126 tornado.access[13052] INFO: 200 GET /static/visualiser/css/luigi.css (::1) 9.54ms
22-07-21 16:24:35,143 tornado.access[13052] INFO: 200 GET /static/visualiser/lib/jquery-1.10.0.min.js (::1) 14.00ms

22-07-21 16:24:35,150 tornado.access[13052] INFO: 200 GET /static/visualiser/lib/AdminLTE/css/skin-green-light.min.
s (::1) 5.99ms
22-07-21 16:24:35,165 tornado.access[13052] INFO: 200 GET /static/visualiser/lib/bootstrap3/css/bootstrap.min.css (
1) 13.06ms
22-07-21 16:24:35,171 tornado.access[13052] INFO: 200 GET /static/visualiser/css/font-awesome.min.css (::1) 18.70ms

22-07-21 16:24:35,175 tornado.access[13052] INFO: 200 GET /static/visualiser/css/tipsy.css (::1) 22.73ms
22-07-21 16:24:35,181 tornado.access[13052] INFO: 200 GET /static/visualiser/lib/bootstrap3/css/bootstrap-theme.min
ss (::1) 28.93ms
22-07-21 16:24:35,188 tornado.access[13052] INFO: 200 GET /static/visualiser/lib/AdminLTE/css/AdminLTE.min.css (::1
34.25ms
22-07-21 16:24:35,195 tornado.access[13052] INFO: 200 GET /static/visualiser/lib/datatables/css/jquery.dataTables.m
```


Luigi Task Visualiser
localhost:8082/static/visualiser/index.html
Luigi Task Status
Task List
Dependency Graph
Workers
Resources
Running
TASK FAMILIES
Clear selection
PENDING TASKS
0
RUNNING TASKS
0
BATCH RUNNING TASKS
0
DONE TASKS
0
FAILED TASKS
0
UPSTREAM FAILURE
0
DISABLED TASKS
0
UPSTREAM DISABLED
0
Show 10 entries
Filter table:
Filter on Server
Name
Details
Priority
Time
Actions
No data available in table
Showing 0 to 0 of 0 entries
Previous
Next
→ Localhost:8082
30°C
Ensolarado
16:36
21/07/2022

Framework

O *framewok Luigi* tem suporte para trabalho de forma gráfica



Obrigado!

Prof. Dr. Diego Bruno

Machine Learning