

Aby zainstalować dask, wykonaj: ``pip3 install dask[complete]``.

W przypadku braku programu pip3, należy wykonać: ``sudo apt get-install python3-pip``.

Korzystając z biblioteki dask oraz poniższych danych, wygeneruj wykorzystując różne konfiguracje procesora i na podstawie 10 uruchomień wygeneruj porównanie czasowe:

Konfiguracje na których należy przetestować działanie aplikacji:

1. Konfiguracja 1 maszyny:
  - a. Typ: e2-medium
  - b. Dysk: 20GB
2. Typ maszyny:
  - a. Typ: e2-highcpu-4
  - b. Dysk: 20GB

Do wykonania:

- Qwiklabs: [Setting up a Minecraft Server on Google Compute Engine | Qwiklabs](#)
- Wygenerować tabelę (jedną), która będzie zawierać w sobie:
  - Nazwa repozytorium
  - Ilość commitów per repozytorium
  - Ilość unikalnych autorów per repozytorium

Tabela powinna być posortowana po ilości commitów

Jako wynik powinien powstać dokument:

- Opisujący sposób wykonania zadania
- Pomiar
- Wnioski

Punktacja:

- Policzenie ilości commitów i autorów jako jedna tabela: (po 2 pkt, razem 4)
- Sortowanie: 1pkt
- dwie konfiguracje: 3pkt
- Pomiar: 2pkt
- uzasadnienie: 1pkt
- Poprawna konfiguracja dask: 1pkt
- Zadanie Qwiklabs: 2pkt

Dane do pobrania

Nowe:

Aby pobrać nowe dane należy wykorzystać komendę:

`gsutil -m cp gs://pjwstk-bigdata/*.parquet .`

Stare:

- `gs://pjwstk-bigdata/0.parquet.gzip`
- `gs://pjwstk-bigdata/1.parquet.gzip`
- `gs://pjwstk-bigdata/2.parquet.gzip`
- `gs://pjwstk-bigdata/3.parquet.gzip`
- `gs://pjwstk-bigdata/4.parquet.gzip`
- `gs://pjwstk-bigdata/5.parquet.gzip`

- gs://pjwstk-bigdata/6.parquet.gzip
- gs://pjwstk-bigdata/7.parquet.gzip
- gs://pjwstk-bigdata/8.parquet.gzip
- gs://pjwstk-bigdata/9.parquet.gzip
- gs://pjwstk-bigdata/10.parquet.gzip

Pobranie danych na dysk maszyny można uzyskać komendą:  
`gsutil cp <ścieżka z początkiem gs://> <ścieżka na dysku>`

Spodziewane wyniki:

1. Ilość commitów:

Result	
repo_name	commit
piszur/vimconfig	1
DotNetAge/freezes	1
DotNetExecution/Successful	1
DottorPagliaccius/Xamarin-Custom-Controls	1
Doubledongli/Logistic-Regression	1
...	...
rperier/linux	9947
mpe/powerpc	9982
scheib/chromium	10125
chromium/chromium	10493
shenzhouzd/update	10977
[1034824 rows x 1 columns]	