Laboratorium 7

Maisuradze Yurii

**Dask**

# Kod aplikacji:

Example.txt:

test test  
ex-librist ex-librist ex-librist  
tesasdasdas asd sadsad sad  
test exocannibalism  
exobiological  
asdsadas  
tt  
sad sad adsfsadfdsaf  
exocannibalism exocannibalism exocannibalism exocannibalism exocannibalism  
exobiological  
adsfsadfdsaf  
untightening untightening untightening untightening untightening sd untightening untightening untightening

Main.py:

import dask.bag as db  
from dask.distributed import Client  
import sys  
  
def count\_words(file\_path, word\_length):  
 client = Client()  
 bag = db.read\_text(file\_path)  
 words = bag.str.lower().str.replace("[^a-zA-Z0-9\\s]", "").str.split().flatten()  
 words = words.filter(lambda x: len(x) >= word\_length)  
 word\_counts = words.frequencies().compute()  
 sorted\_word\_counts = sorted(word\_counts, key=lambda x: x[1], reverse=True)  
 print(sorted\_word\_counts)  
 client.close()  
  
if \_\_name\_\_ == "\_\_main\_\_":  
 if len(sys.argv) != 3:  
 print("Usage: main.py <file\_path> <word\_length>")  
 sys.exit(1)  
  
 file\_path = sys.argv[1]  
 word\_length = int(sys.argv[2])  
  
 count\_words(file\_path, word\_length)

# Konfiguracja dockera

Dockerfile:

FROM python:3.9  
RUN pip install "dask[complete]"  
COPY main.py /  
ENTRYPOINT ["python", "/main.py"]  
CMD []

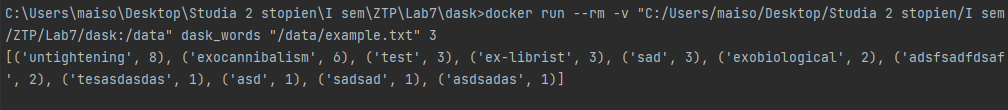
Lista komend dockera używanych do uruchomienia zadania  
main.py, Dockerfile oraz example.txt są w jednym folderze

# docker build -t dask\_words .

# docker run --rm -v "C:/Users/maiso/Desktop/Studia 2 stopien/I sem/ZTP/Lab7/dask:/data" dask\_words "/data/example.txt" <minimalna dlugość>

# Zrzuty ekranu pokazujące działanie programu

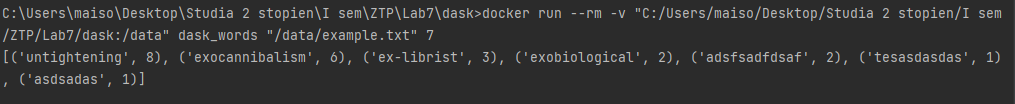
Dla długości 3:



C:\Users\maiso\Desktop\Studia 2 stopien\I sem\ZTP\Lab7\dask>docker run --rm -v "C:/Users/maiso/Desktop/Studia 2 stopien/I sem/ZTP/Lab7/dask:/data" dask\_words "/data/example.txt" 3

[('untightening', 8), ('exocannibalism', 6), ('test', 3), ('ex-librist', 3), ('sad', 3), ('exobiological', 2), ('adsfsadfdsaf', 2), ('tesasdasdas', 1), ('asd', 1), ('sadsad', 1), ('asdsadas', 1)]

Dla długości 7:



C:\Users\maiso\Desktop\Studia 2 stopien\I sem\ZTP\Lab7\dask>docker run --rm -v "C:/Users/maiso/Desktop/Studia 2 stopien/I sem/ZTP/Lab7/dask:/data" dask\_words "/data/example.txt" 7

[('untightening', 8), ('exocannibalism', 6), ('ex-librist', 3), ('exobiological', 2), ('adsfsadfdsaf', 2), ('tesasdasdas', 1), ('asdsadas', 1)]