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:

```python
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)]
```

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)]
```

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)]