

Practical Work 5: Longest Path

Vu Dinh Anh

March 2021

Longest Path

Files: fil.py, ds2021.txt
Usage: python fil.py ds2021.txt
Expected output:

```
('ds2021/.git/logs/refs/remotes/origin/HEAD', 7)
```

Implementation

Here I used [Python Apache Beam](#) with the help of following documents:

- [Python Beam Transforms Aggregation Max](#)
- [Python Beam Examples wordcount_minimal.py](#)

Source code:

```
import sys
import os
import apache_beam as beam

from apache_beam.io import ReadFromText
from pprint import pprint

if __name__ == '__main__':
    file = sys.argv[1]
    if not os.path.isfile(file):
        print('File not found')
    else:
        with beam.Pipeline() as pipeline:
            # Reading
            lines = pipeline | ReadFromText(file)

            # Mapping
            measured_paths = (
```

```

lines
| "Measure" >> beam.Map(
    lambda line: (line, len(line.split('/')))
)

# Reducing
max_value = (
    measured_paths
    | "Max" >> beam.CombineGlobally(
        lambda paths: max(paths, key=lambda p: p[1])
    )
    | beam.Map(pprint)
)

```