

OPERATING SYSTEMS & PARALLEL COMPUTING

Final Project

Parallel Computing Project with Python

Using the dataset of Google Analytics Sessions from

https://github.com/operard/opsys_parallel/tree/master/operating_systems/examples

- Use the code
“https://github.com/operard/opsys_parallel/blob/master/operating_systems/examples/cluster_analysis.py”
- And the data:
 - https://github.com/operard/opsys_parallel/blob/master/operating_systems/examples/sesion.csv.gz.001
 - https://github.com/operard/opsys_parallel/blob/master/operating_systems/examples/sesion.csv.gz.002
- to adapt the code for parallelism.

Choose the platform and the algorithm for your personal project:

- Implement the “Parallel K-means” using MPI with HPC platform
- Implement the “Parallel K-means” using PyCuda with GPU platform
- Implement the “Parallel K-means” using SparkML with Big Data Platform

When you have implemented your algorithm, I will share a big file with 140M sessions to valid the performance of your implementation.

