# Performance on multicore processors in Apache Spark MLlib

## System information

11th Gen Intel(R) Core(TM) i5-11400H @ 2.70GHz (6 cores and 12 threads). 16 GB RAM. JDK 11. Python 3.7.16.

## Information about the two datasets

| Name of dataset | Classes | Examples | Features |
|---|---|---|---|
| covtype | 2 | 581 012 | 54 |
| SUSY | 2 | 5 000 000 | 18 |

Table 1: Source: UCI repository

## Results

Before running `SVMTrainingDataframe.java` or `SVMTrainingDataframe.py` on SUSY, it's necessary to open it to replace "**-9.537825584411621094-01**" by "**-9.537825584411621094e-01**".

| | maxIter: 100. regParam: 0.1 | | | |
|---|---|---|---|---|
| | COVTYPE | | SUSY | |
| Cores | Python | Java | Python | Java |
| 1 | 10.39 | 16.08 | 191.59 | 210.12 |
| 2 | 7.85 | 12.46 | 108.76 | 121.85 |
| 4 | 6.49 | 11.14 | 69.23 | 80.13 |
| 6 | 12.28 | 14.14 | 63.02 | 67.58 |

Table 2: Elapsed time (s) by number of cores