

# Introduction to Pandas & Python in Distributed Environments

Stanislav Böhm  
IT4Innovations

[stanislav.bohm@vsb.cz](mailto:stanislav.bohm@vsb.cz)

[github.com/spirali](https://github.com/spirali)

9:00 - 10:30

Introduction to Pandas

10:45 - 12:00

Python in Distributed Environments

9:00 - 10:30

Introduction to Pandas

10:45 - 12:00

Python in Distributed Environments

Why

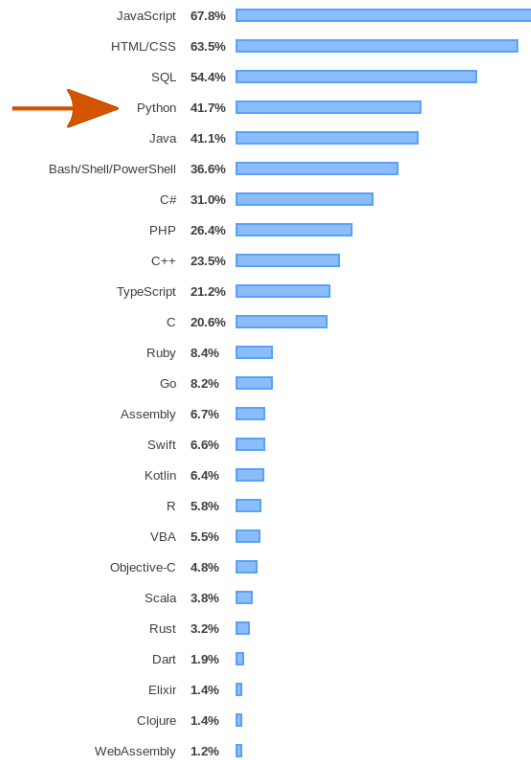


?

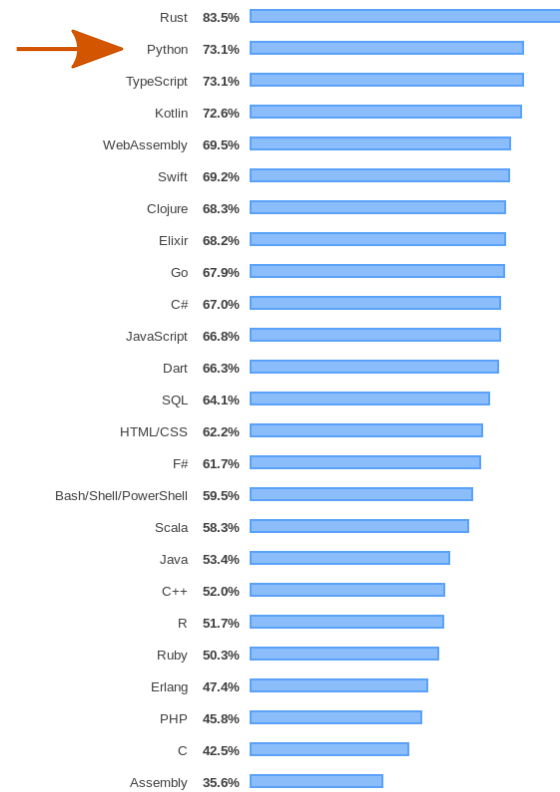


# Survey 2019

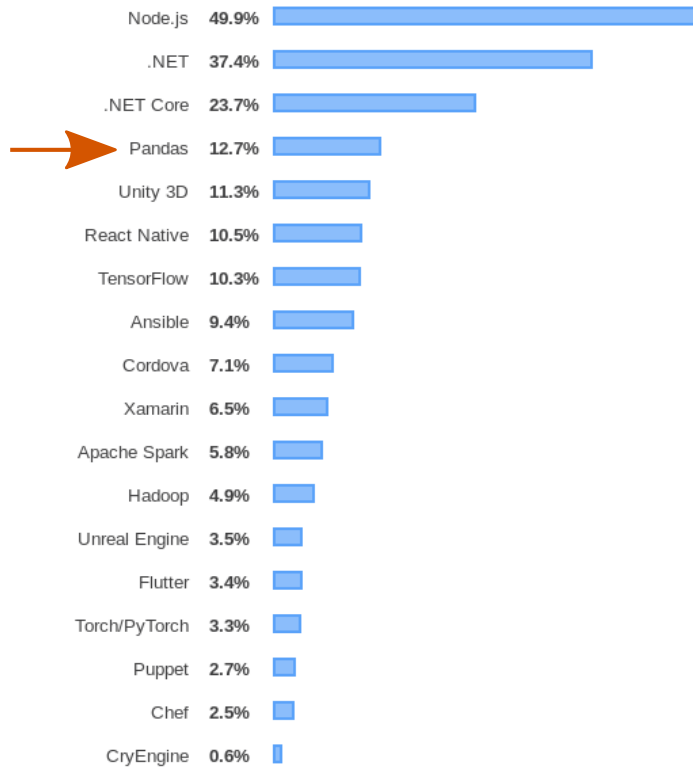
## Most **used** programming languages



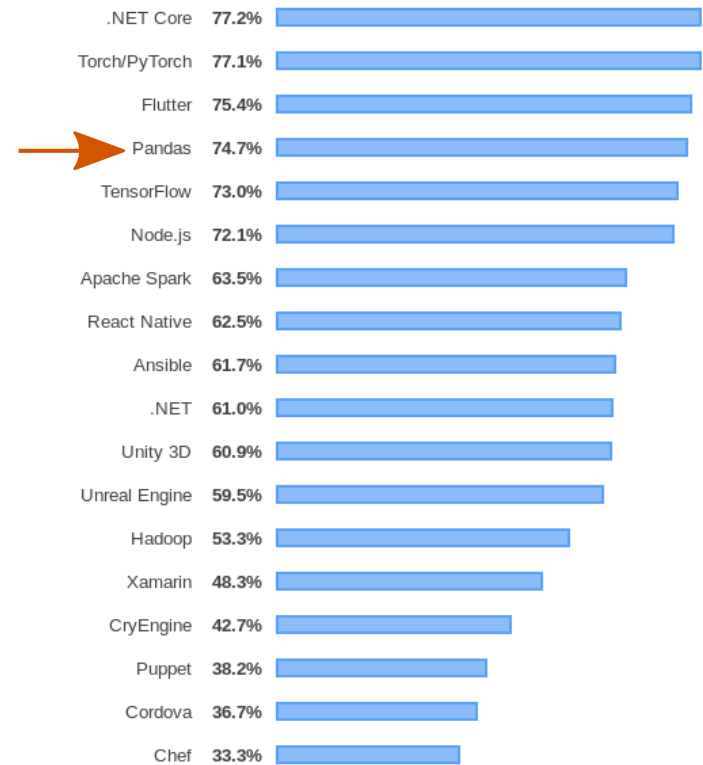
## Most **loved** programming languages



Most **used** frameworks/libraries/tools  
(without programming languages and web frameworks)



Most **loved** frameworks/libraries/tools  
(without programming languages and web frameworks)



# Hands-on

<http://hydra.vsb.cz>

(not `https`)

9:00 - 10:30

Introduction to Pandas

10:45 - 12:00

Python in Distributed Environments



*CPU bound* vs *Memory bound*

Salomon: *128GB* per node

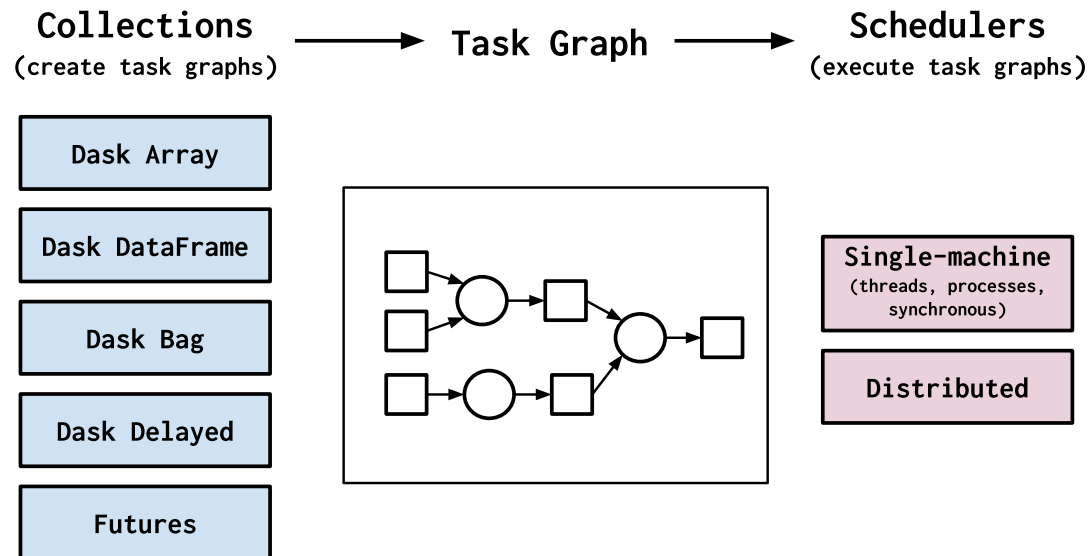
Barbora: *196GB* per node



<https://github.com/vaexio/vaex>



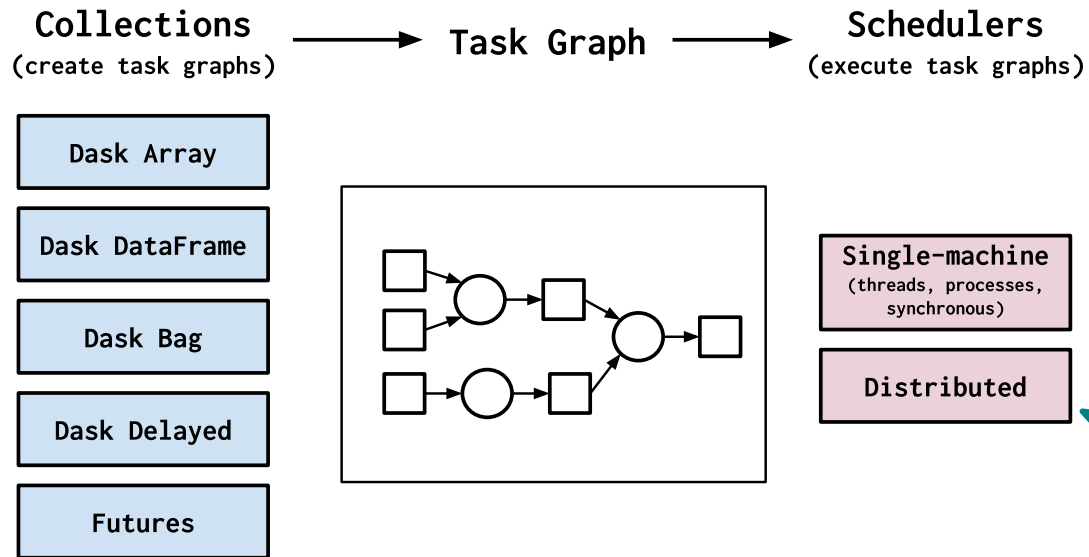
<https://dask.org/>



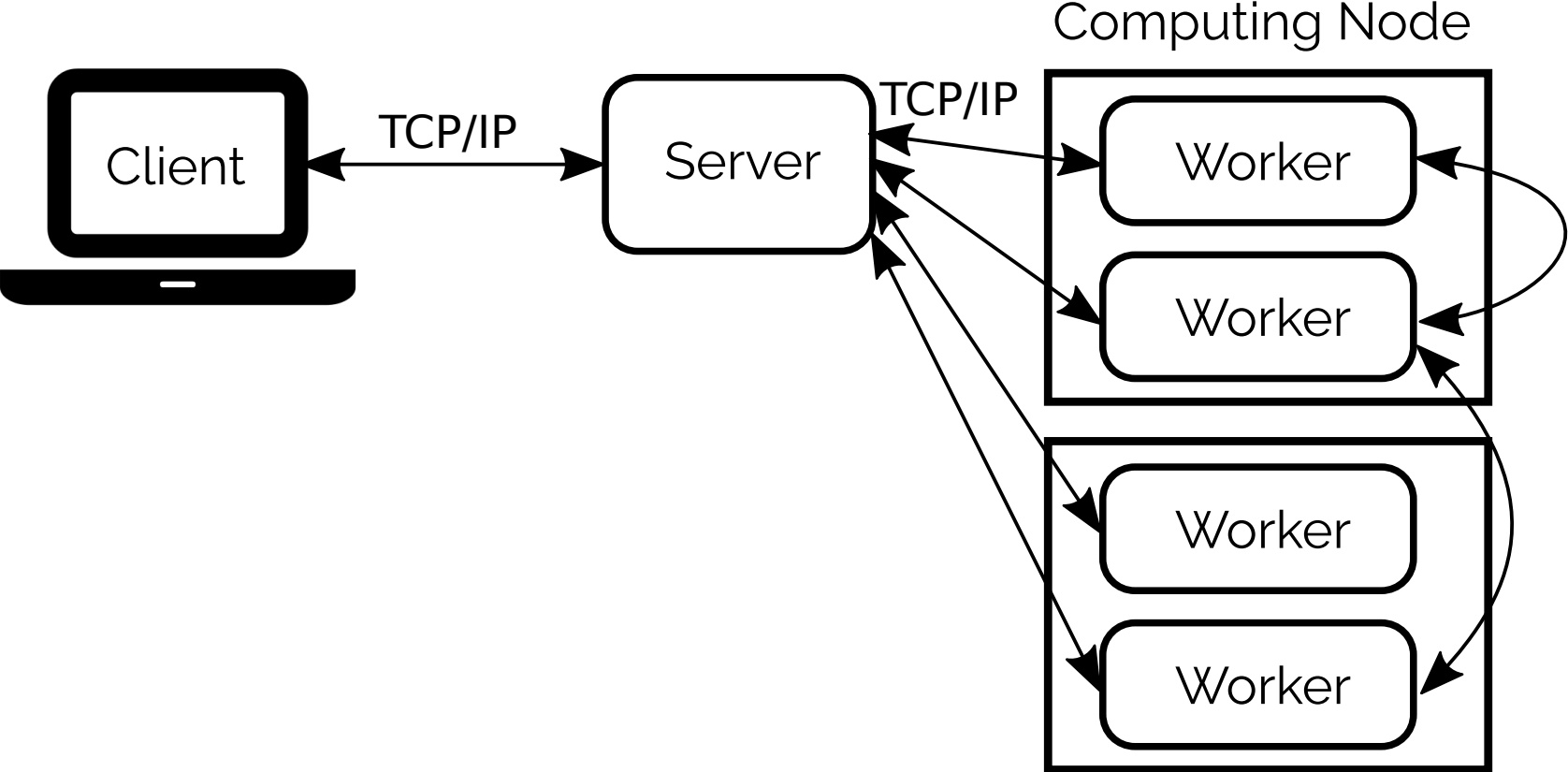
[source: docs.dask.org]

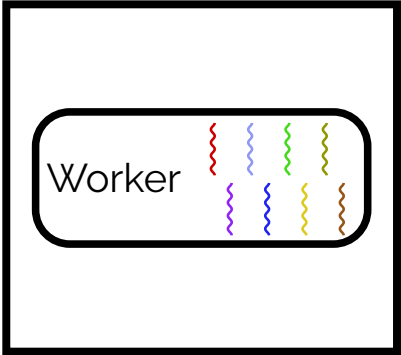


# DASK

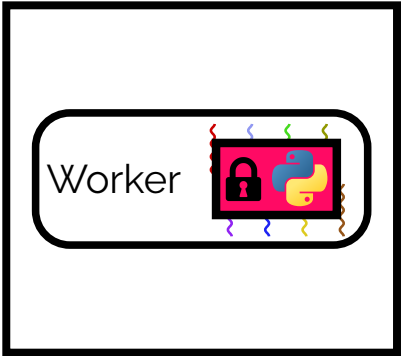


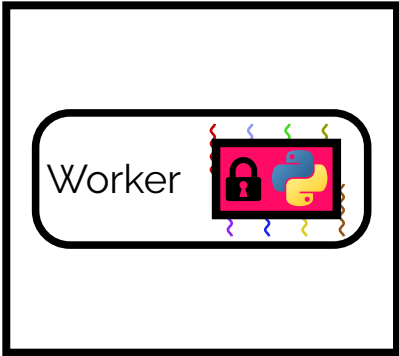
[source: docs.dask.org]



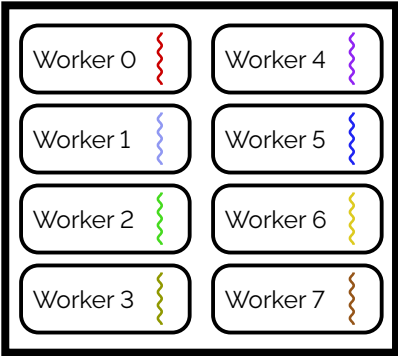


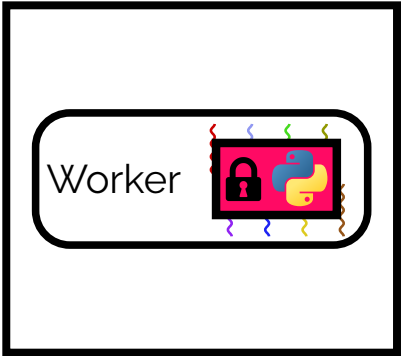




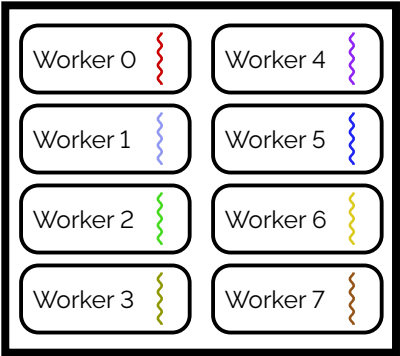


or

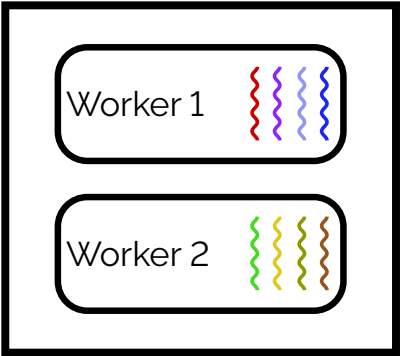




or



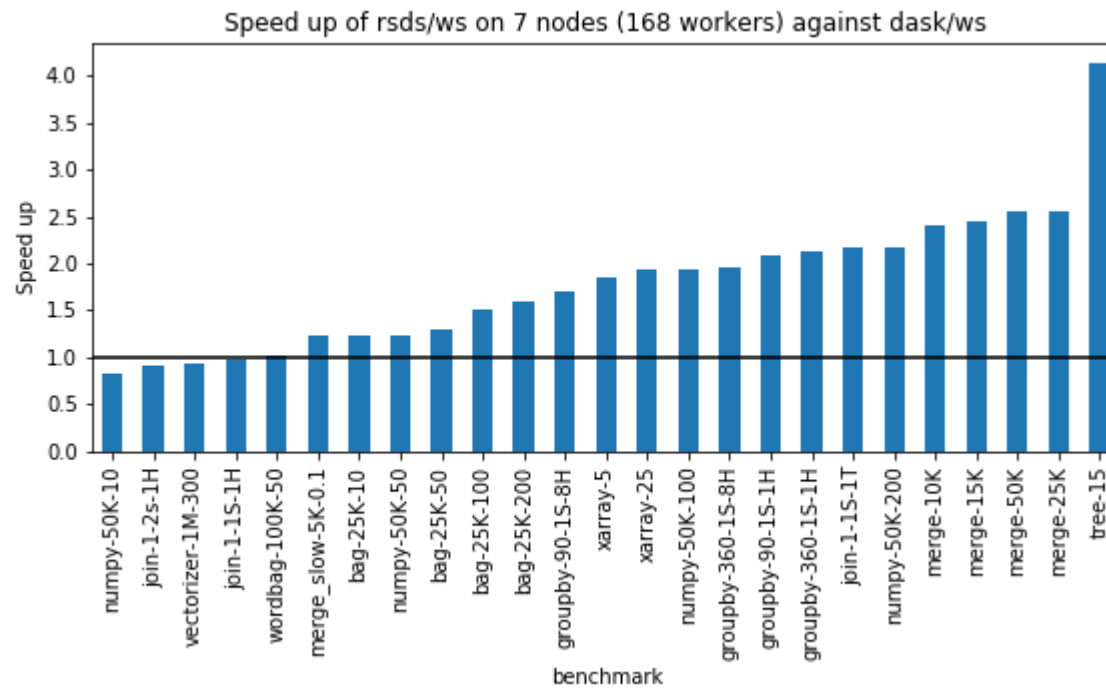
or



?

# RSDS

<https://github.com/It4innovations/rsds>





<https://github.com/ray-project/ray>



<https://modin.readthedocs.io/en/latest/>



<https://spark.apache.org/>

# ExaQUTE

Exascale Quantification of Uncertainties for  
Technology and Science Simulation

```
@exaquite.task()  
@exaquite.mpi(n_processes=10)  
def do_some_computation(source_data, parameters):  
    ...
```