

CUDF - GPU DATAFRAMES

Rostislav Stoyanov
26321

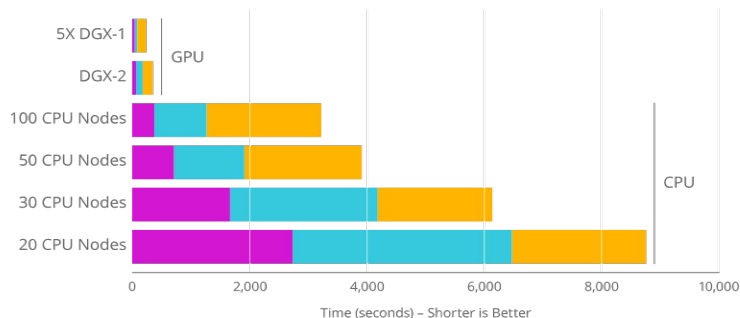
RAPIDS

- Data science framework
- Execute end-to-end pipelines entirely on GPU
- Familiar to Python libraries

RAPIDS

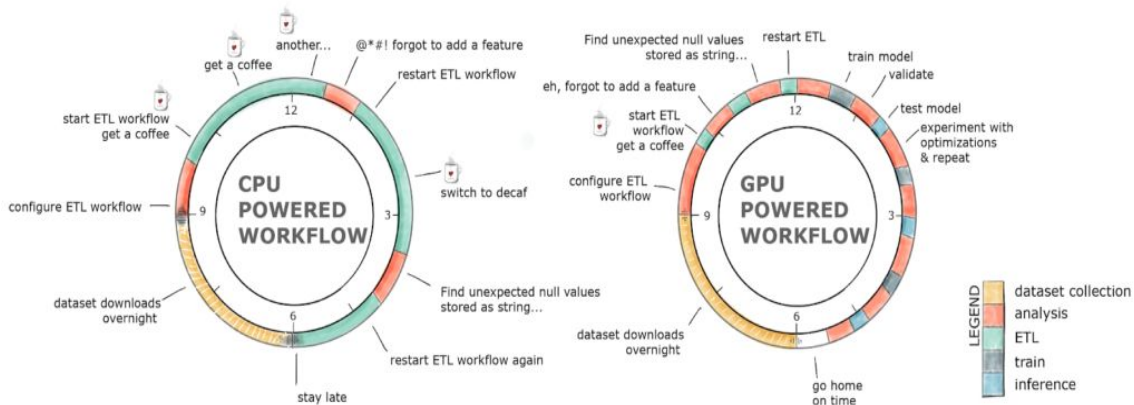


RAPIDS(cont)



- Utilize CUDA primitives
- Extends GPU Open Analytics Initiative (GOAI) work
- Based on Apache Arrow

DAY IN THE LIFE OF A DATA SCIENTIST



Source:

<https://developer.nvidia.com/blog/gpu-accelerated-analytics-rapids/>

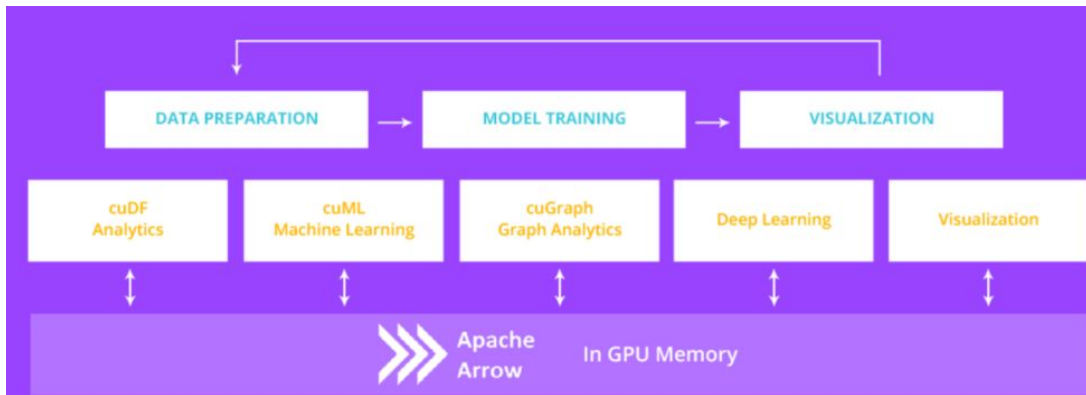
Applications

- Accelerated Simulation of Air Pollution
- Monitoring the Decline of Air Pollution Across the Globe During the COVID-19 Pandemic



Libraries

- cuDF - pandas-like dataframe manipulation library
- cuML - a collection of machine learning libraries that will provide GPU versions of algorithms available in scikit-learn
- cuGraph - a NetworkX-like accelerated graph analytics library
- NVTabular



cuDF

- cuDF is a GPU DataFrame library for loading, joining, aggregating, filtering, and otherwise manipulating data
- Provides a pandas-like API that will be familiar to data engineers & data scientists, so they can use it to easily accelerate their workflows without going into the details of CUDA programming.



The background is a solid dark blue. In the top right corner, there is a decorative pattern of triangles in various shades of blue, including dark blue, medium blue, and light blue.

DEMO