Why GPUs for Data Science?

Numerous hardware advantages

- Thousands of cores with up to ~20 TeraFlops of general purpose compute performance
- Up to 1.5 TB/s of memory bandwidth
- Hardware interconnects for up to 600 GB/s bidirectional GPU <-> GPU bandwidth
- Can scale up to 16x GPUs in a single node

Almost never run out of compute relative to memory bandwidth!
What is RAPIDS?
End-to-End GPU Accelerated Data Science

Data Preparation/ETL

GPU-accelerated ETL functions
Tracks Pandas and other common PyData APIs
Dask + UCX integration for scaling

cuDF

Analytics/ML/Graph

GPU-native cuML library, plus XGBoost, FIL, HPO, and more
GPU graph analytics, including TSP, PageRank, and more

RAPIDS ML
cuGraph

Visualization
cuxfilter

GPU-accelerated cross-filtering
pyViz integration
Plotly Dash, Bokeh, Datashader, HoloViews, hvPlot

Domain-Specific Libraries

CLX + Morpheus
Cyber log processing + anomaly detection
cuStreamz
Streaming analytics
cuSignal
Signals processing
cuSpatial
Spatial analytics
cuCIM
Computer vision & image processing primitives
node-RAPIDS
Bindings for node.js

...and more!
**Overview of Changes: RAPIDS 22.04 Release**

- **cuDF** Enabled all numpy ufuncs for DataFrames, support byte_range argument for reading text, covariance support added to GroupBy objects
- **cuML** Added 2 new kernel models, enabled interruptible execution with `Ctrl+c`, new solvers supported by ElasticNet and Lasso Regressions
- **cuGraph** Added an initial release of a new Property Graph class; improved scale and performance (Louvain and Pagerank)
- **Dask-CUDA** Added Opt-in support to track RMM Allocations, improved support for `__Array_function` and cuPy arrays with JIT-Unspill
cuDF Updates: Deep Dive

Features added in 22.04

- Enabled read text with dask_cudf using byte_range
- Enable numpy ufuncs for DataFrame
- Add ‘Spearman’ correlation method for dataframe.corr
- Add covariance for sort GroupBy (Python)
- Add libcudf strings split API that accepts regex pattern
- Implement mixed equality/conditional semi/anti joins

Planned Upcoming Features

- Support PyData and SQL rank operations
- Support zstandard codec and block GZIP compression
cuML Updates: Deep Dive
Release 22.04

Features added in 22.04
- Add Kernel Ridge Regression model
- Add Kernel Density Estimation model
- Interruptible execution with ‘Ctrl + c’
- Add Quasi-newton solver for ElasticNet and Lasso Regression.
- Include variance statistics in Target encoder

Planned upcoming features
cuGraph Updates: Deep Dive

Features added in 22.04
- Improved Louvain performance and scaling
- Initial release of a new Property Graph class
- Improved doctest automation
- Scale testing: Pagerank and Louvain runs on 1000+ GPUs

Planned Upcoming Features
- Massive graph support, trillion edges
- cuGraph integrated with DGL
- Expanded Property Graph and new NetworkX Compatibility Module
Join the Conversation

GOOGLE GROUPS
https://groups.google.com/forum/#!forum/rapidsai

DOCKER HUB
https://hub.docker.com/r/rapidsai/rapidsai

SLACK CHANNEL
https://rapids-goai.slack.com/join

STACK OVERFLOW
https://stackoverflow.com/tags/rapids
THANK YOU

@RAPIDSai

RAPIDS