RAPIDS

22.02 Release

@RAPIDSai

https://github.com/rapidsai

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

RAPIDS

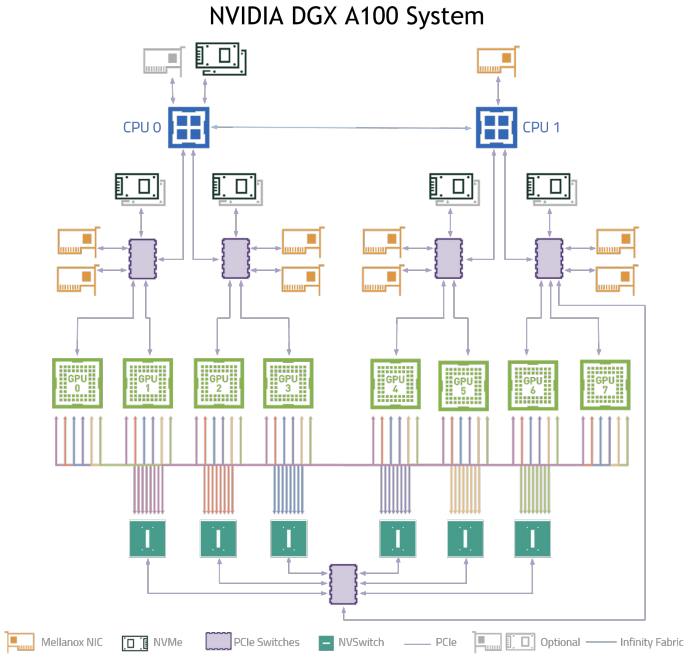
https://rapids.ai

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

cuDF

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

RAPIDS ML

- GPU-native cuML library, plus XGBoost, FIL, HPO, and more

cuGraph

- GPU graph analytics, including TSP, PageRank, and more

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

cuCIM
Computer vision & image processing primitives

...and more!

cuSpatial
Spatial analytics

node-RAPIDS
Bindings for node.js

Overview of Changes: RAPIDS 22.02 Release

- **RAPIDS** Added general support for SageMaker Studio Lab, support for Python 3.8 and 3.9, EOL support for Python 3.7; GPUDirect Storage (GDS) integration
- **RAPIDS+Dask** Improved read parquet performance when using remote storage (optimizing data transfer and caching); Improved task fusion for Dataframe aggregations and when writing parquet files
- **cuDF** Added support for decimal 128 in cuda python and Parquet reader and writer; added additional groupby method support for .corr() and .transform()
- **cuML** Expanded support for gbdt model explainability with SHAP values
- **cuGraph** added an initial release of a new Property Graph class; improved scale and performance (Louvain and Pagerank)
- **Dask-SQL** Added support for multi GPU training and inference for cuML and XGBoost models directly within SQL statements; introduced basic support for Dask's read filtering in CREATE TABLE WITH statements

cuDF Updates: Deep Dive

Release 22.02

Features added in 22.02

- Added groupby method .corr() that will compute the Pearson correlation coefficient between dataframe columns
- Added groupby method .transform() to apply aggregations to groups and broadcast the results to the group size
- Support for decimal128 in Parquet reader and writer
- Improved performance of partitioned Parquet writing
- Support for decimal128 in cudf python

Planned Upcoming Features

- String support in Numba UDFs
- Optimize compaction operations w/ cuco
- Unify cuDF UDF Interface

cuML Updates: Deep Dive

Release 22.02

Features added in 22.02

- Add the ability to compute SHAP values for a wider range of general decision tree models (XGBoost, LightGBM models, and Random Forests with categorical variables from both Scikit-learn and cuML)

Planned Upcoming Features

- New estimators like Kernel Density Estimation, Kernel Ridge Regression and others.
- Performance optimizations.
- RAPIDS-Triton backends additional features.
- Increase reuse of CUDA/C++ primitives with RAFT library

cuGraph Updates: Deep Dive

Release 22.02

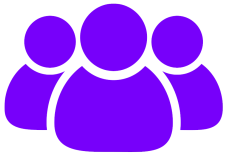
Features added in 22.02

- Improved Louvain performance and scaling
- Initial release of a new Property Graph class
- Improved doctest automation
- Scale testing: Pagerank and Louvain runs on 500+ GPUs

Planned Upcoming Features

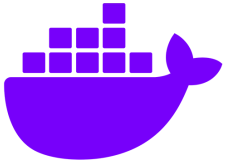
- Massive graph support, trillion edges
- cuGraph integrated with DGL
- Expanded Property Graph and new NetworkX Compatibility Module

Join the Conversation



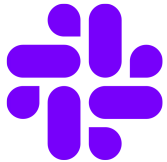
GOOGLE GROUPS

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