RAPIDS

0.19 Release

@RAPIDSai

https://github.com/rapidsai

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

RAPIDS

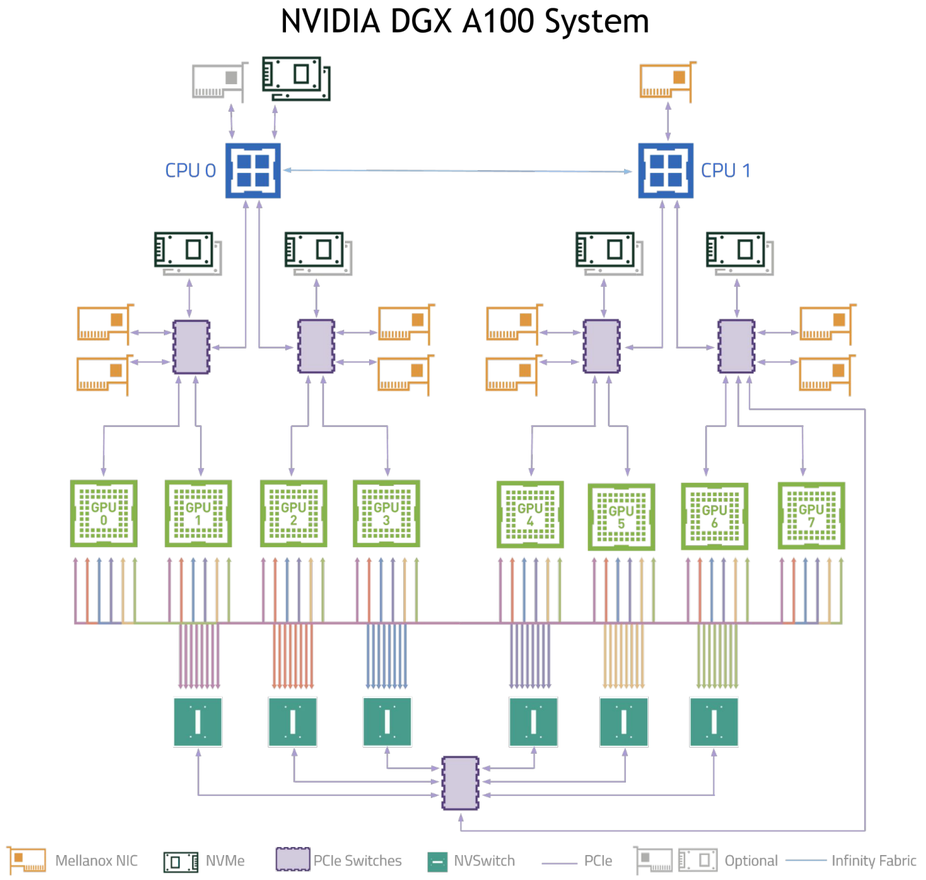
https://rapids.ai

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

cuML

- GPU-native ML library, including 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 0.19 Release

- **RAPIDS** CUDA 11.2 now supported by all RAPIDS libraries; initial release of the new cuCIM library
- **RAPIDS+Dask** Improved performance and memory spilling (JIT_UNSPILL), added capability to log spilling, improved UCX Debugging and Documentation, UCX 1.9 Support, RAPIDS Memory Manager(RMM) logging with Dask-CUDA
- **cuDF** Support for fixed-point decimal types in Python; more groupby and rolling window aggregations; support for list type operations in Python; expanded dictionary type operations in C++;
- **cuML** Scikit-learn-compatible preprocessing, Single Linkage Hierarchical Clustering Algorithm; SHAP explainability; improved Random Forest classification, improvements to forest inference, DBSCAN, kNN
- **XGBoost** 1.4.0 ships with 0.19, including improvements to Dask integration and prediction functions
- **cuGraph** new Random Walk; RMAT data generator; continued improving graph primitives for performance, work started on supporting multiple seeds for BFS, SSSP, and Egonet
- **CLX** Sensitive information detection workflow and training script, crypto mining and GPU malware detection and training script, host introspection workflow and feedback prototype

cuDF Updates: Deep Dive
Release 0.19

Features added in 0.19

- Decimal data type is now supported for *joins*, *read_parquet*, and column comparison functions in Python
- *Unique* and *sort* functions for *groupby* aggregation are now available
- Support for nested types such as *lists* and *structs* in Python and a Medium blog to elaborate it
- Enhanced support for dictionary data types in C++

Planned Upcoming Features

- *Cumulative operations* for groupby
- Conditional Joins
- ORC GDS Support
- Decimal Type Support for ORC and CSV

cuML Updates: Deep Dive

Release 0.19

Features added in 0.19

- Scikit-learn compatible preprocessing - now no longer experimental - 10+ preprocessing methods
- SHAP explainability - also ready for production - explain predictions of any cuML or sklearn model
- New *Random Forest* backend for classification models - better performance and accuracy
- New *Single Linkage Hierarchical Clustering Algorithm*
- *Logistic Regression* accepts *sample_weight* parameter
- *predict_proba* function is now available for *XGBoost-style* models in *Forest Inference Library (FIL)*
- New distance metrics for *Approximate Nearest Neighbors (ANN)*
- cuML integrated into *AutoGluon*

Planned Upcoming Features

- New *Random Forest* backend will be expanded to support regression
- *Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm*
- Fast Fourier Transform (FFT) accelerated *t-Stochastic Neighborhood Embedding (t-SNE)*

cuGraph Updates: Deep Dive

Release 0.19

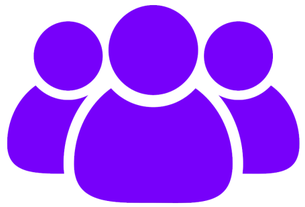
Features added in 0.19

- New sampling algorithm, *Random walk*
- Improved performance of graph primitives on graphs with widely varying vertex degrees
- *Recursive Matrix graph data generator*
- Enhance graph partitioning scheme
- Enhance multi-node multi-gpu *Louvain*

Planned Upcoming Features

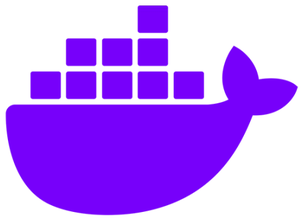
- *Breadth First Search* with depth limit functionality
- Multi-Node Multi-GPU *Weakly* connected components
- Batch *Random Walk*
- *Breadth First Search* using multiple sources in a graph and in multiple graphs
- *Egonet* extractor using multiple sources

Join the Conversation



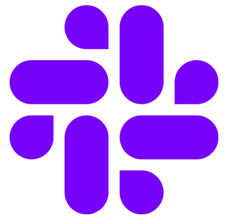
GOOGLE GROUPS

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