DLI Accelerated Data Science Teaching Kit

Lecture 21.1 - RAPIDS Benefits

Benefits of RAPIDS

Easy integration and familiarity

- Easy end-to-end accelerated analytics using GPUs
- Connecting data practitioners to High Performance Computing
- Familiar syntax for most data scientists
- Integrated with several data science frameworks like Apache Spark, Numba, etc. along with Deep Learning frameworks like Pytorch, Tensorflow, etc.
- Overcomes communication bottlenecks with the use of frameworks like UCX-py

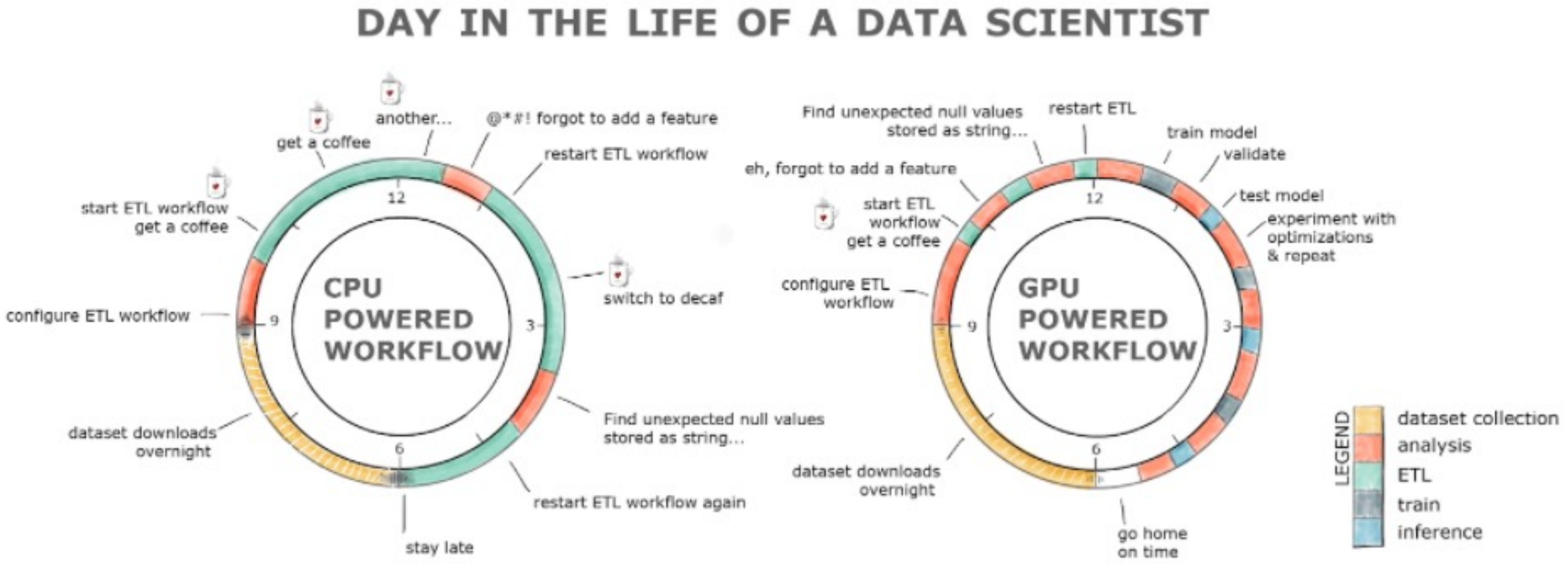
Benefits of RAPIDS

Run anywhere at scale

- Great scalability: a single workstation to multi-GPU servers to multi-node clusters
- Provides a platform to scale up and out with the help of other libraries like Dask
- Run anywhere: Cloud or on-premise environment
- Faster data access with less data movement
- Bridging the gap between compute resources and existing frameworks

Benefits of RAPIDS

Faster and saves time



Shorter ETL workflows save time

Image credit: NVIDIA

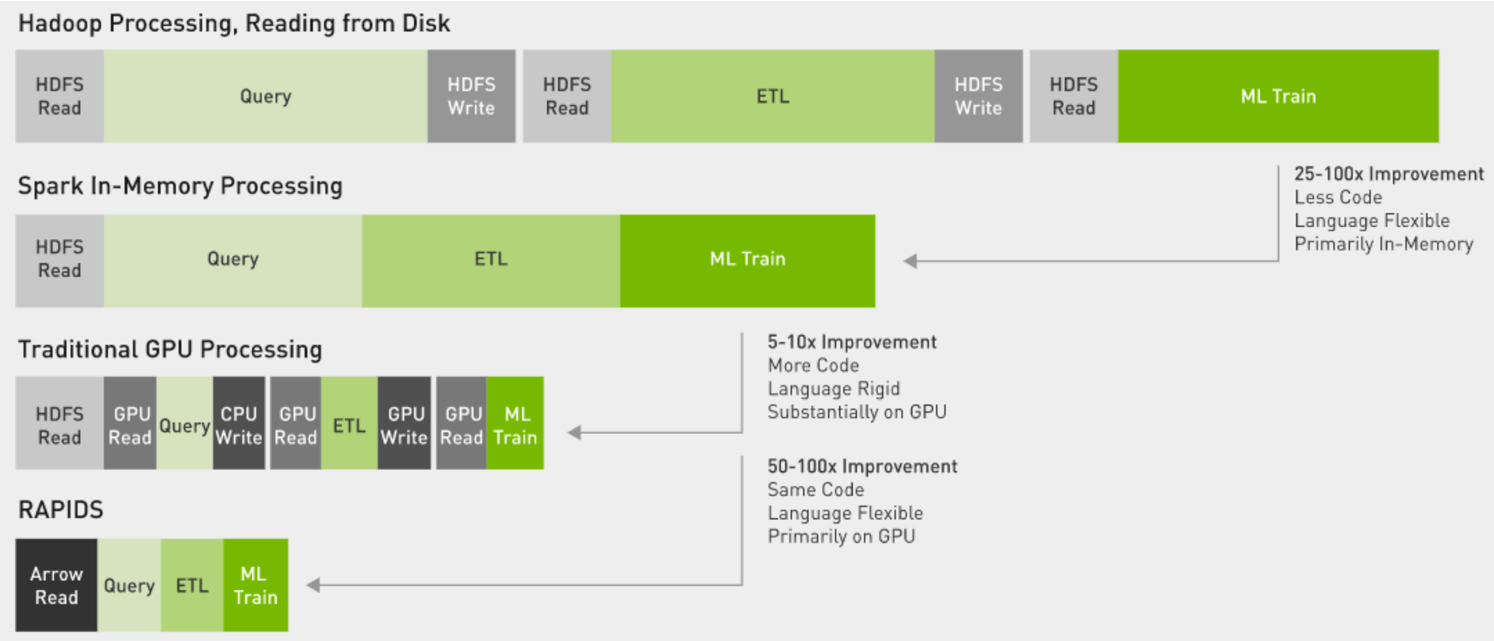
Speed Comparison

Introduction

- Data exploration is a vital part of Data Science Workflows:
    - Understanding, cleaning, manipulating data
    - Consistent data types, formats and filling in gaps
    - Reiterating to add features
- Pandas: Functions developed with vectorized operations for top-speed computations
- But they are CPU-constrained, therefore, tasks are very time consuming
- RAPIDS: Much faster Extract, Transform and Load (ETL) tasks

Speed Comparison

Data Processing Evolution



Hadoop Processing, Reading from Disk

HDFS Read | Query | HDFS Write | HDFS Read | ETL | HDFS Write | HDFS Read | ML Train

Spark In-Memory Processing

HDFS Read | Query | ETL | ML Train

25-100x Improvement
Less Code
Language Flexible
Primarily In-Memory

Traditional GPU Processing

HDFS Read | GPU Read | Query | CPU Write | GPU Read | ETL | GPU Write | GPU Read | ML Train

5-10x Improvement
More Code
Language Rigid
Substantially on GPU

RAPIDS

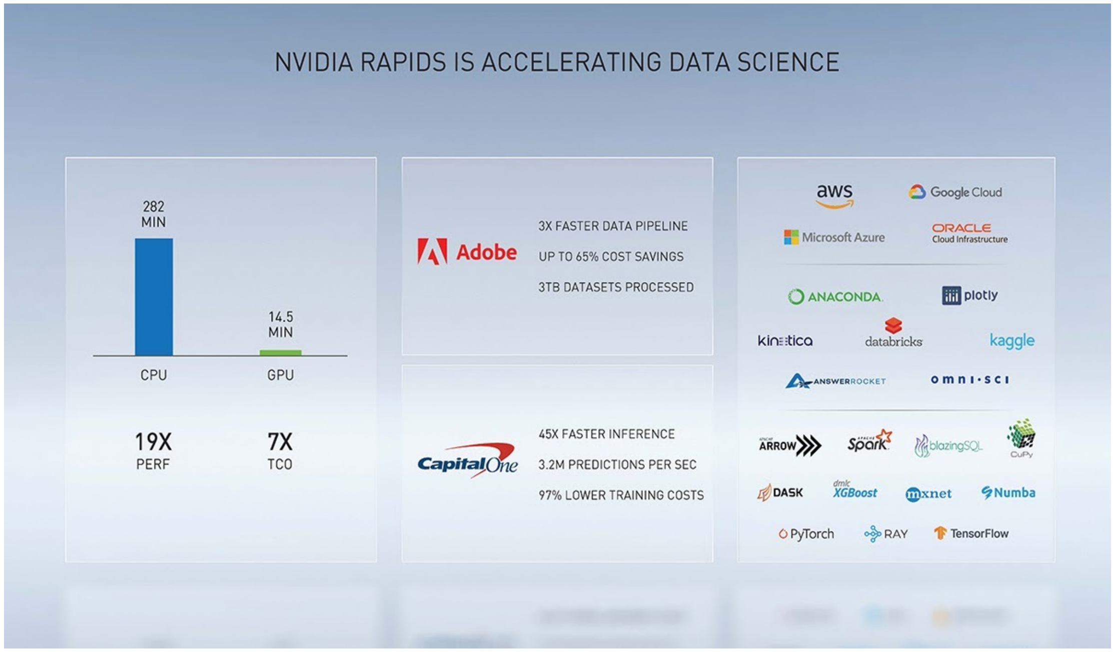
Arrow Read | Query | ETL | ML Train

50-100x Improvement
Same Code
Language Flexible
Primarily on GPU

Overview of data processing with different frameworks

Image credit: NVIDIA

Speed Comparison



NVIDIA RAPIDS IS ACCELERATING DATA SCIENCE

19X PERF

7X TCO

Adobe

3X FASTER DATA PIPELINE

UP TO 65% COST SAVINGS

3TB DATASETS PROCESSED

Capital One

45X FASTER INFERENCE

3.2M PREDICTIONS PER SEC

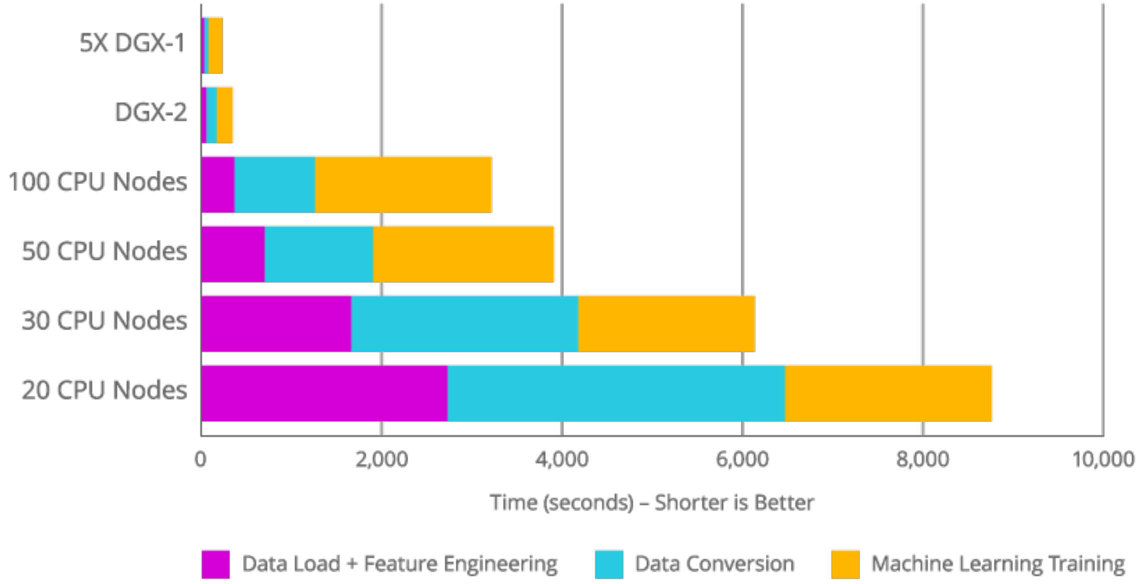
97% LOWER TRAINING COSTS

20x faster on GPUs and 7x more cost-effective than top CPU baseline

Image credit: NVIDIA

Speed Comparison

Example: Fannie Mae loan performance Dataset: 400 GB data in memory



Results for a complete ETL (manipulating DataFrames and training a gradient boosted decision tree model on the GPU using XGBoost)

Image credit: NVIDIA

NVIDIA DEEP LEARNING INSTITUTE | Georgia Tech | PRAIRIE VIEW A&M UNIVERSITY

DLI Accelerated Data Science Teaching Kit

Thank You