

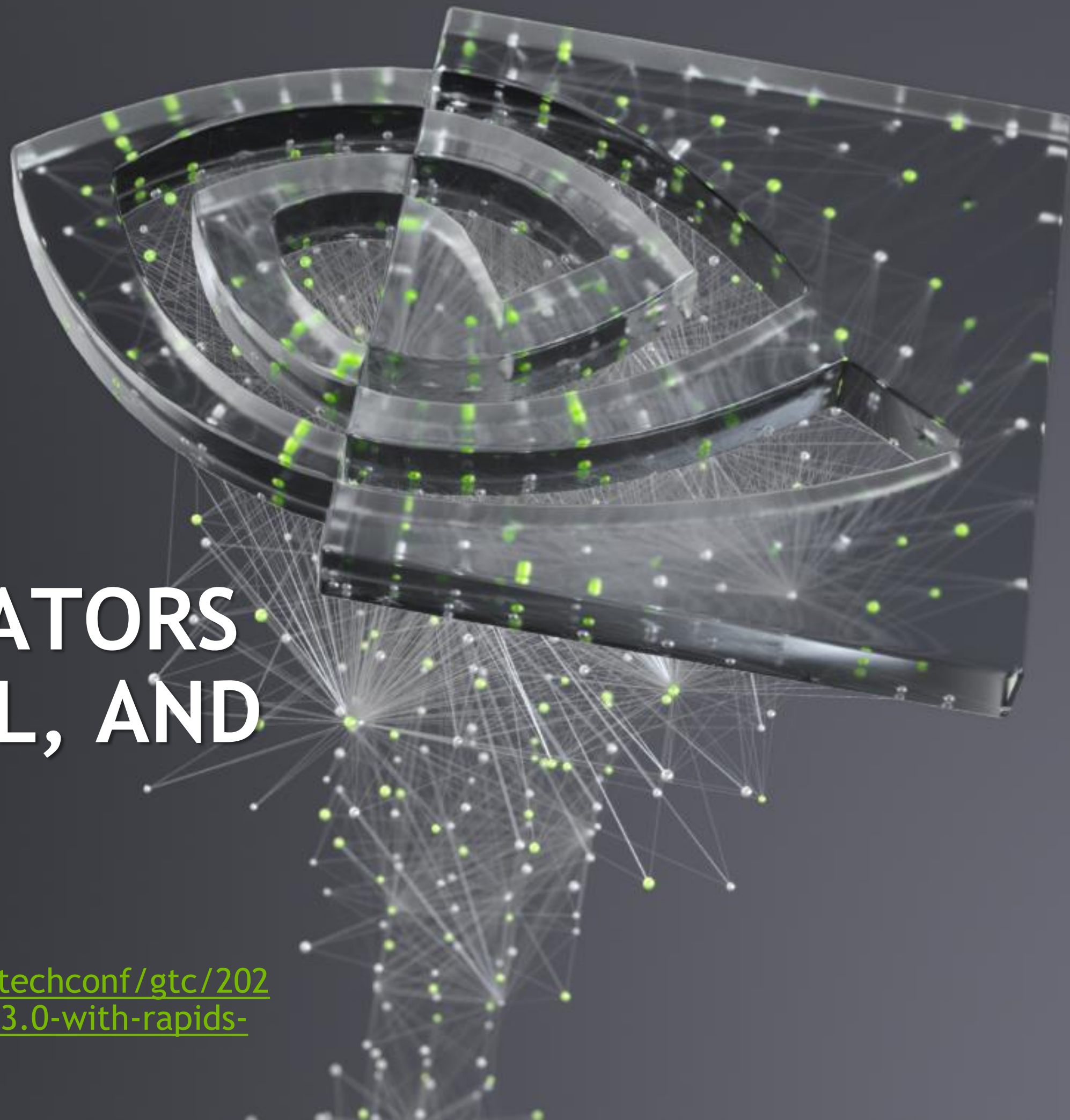


nvidia

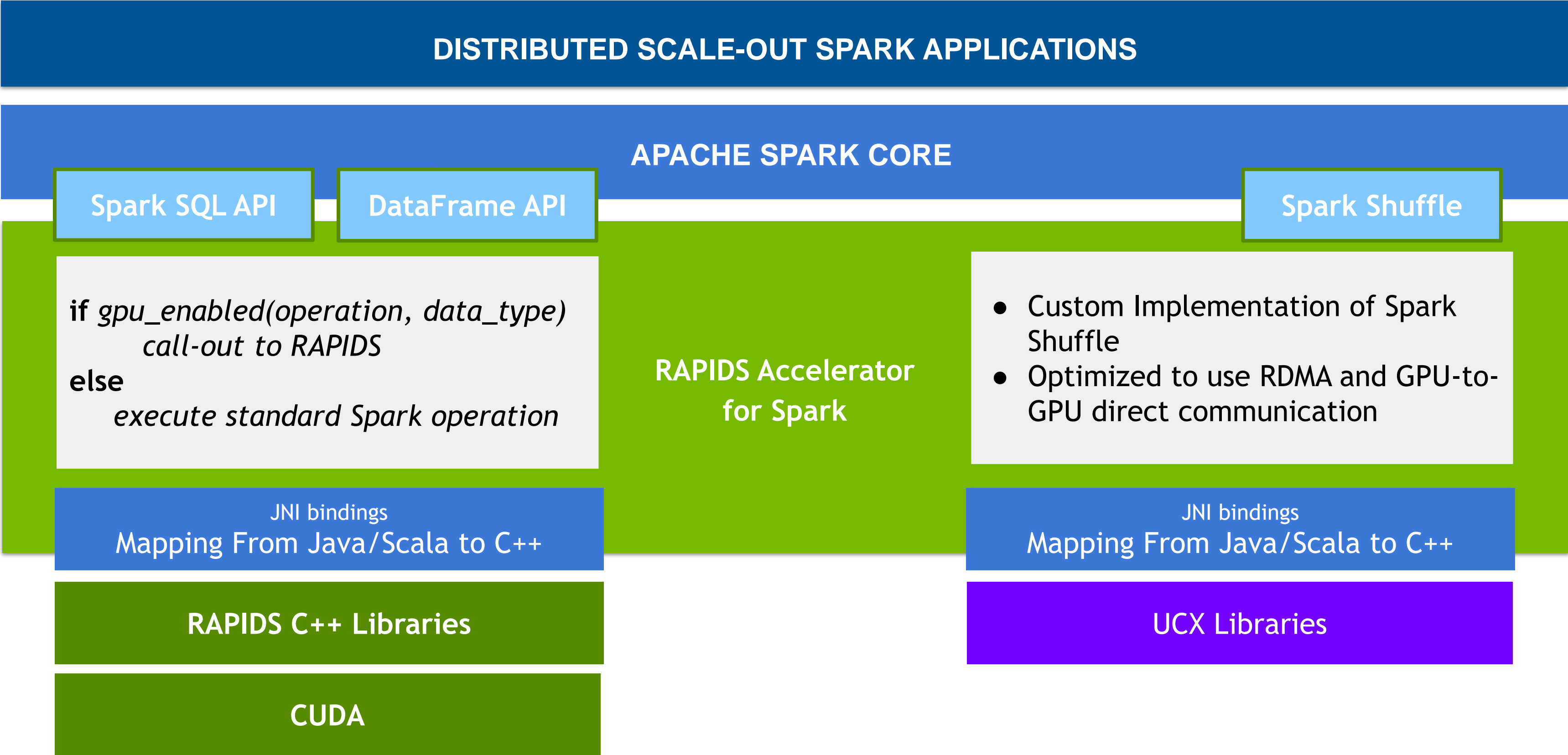
UTILIZING ACCELERATORS TO SPEEDUP ETL, ML, AND DL APPLICATIONS

Full Deck From GTC Can Be Found Here:

<https://developer.download.nvidia.com/video/gputechconf/gtc/2020/presentations/s22674-accelerating-apache-spark-3.0-with-rapids-and-gpus.pdf>



RAPIDS ACCELERATOR FOR APACHE SPARK (PLUGIN)

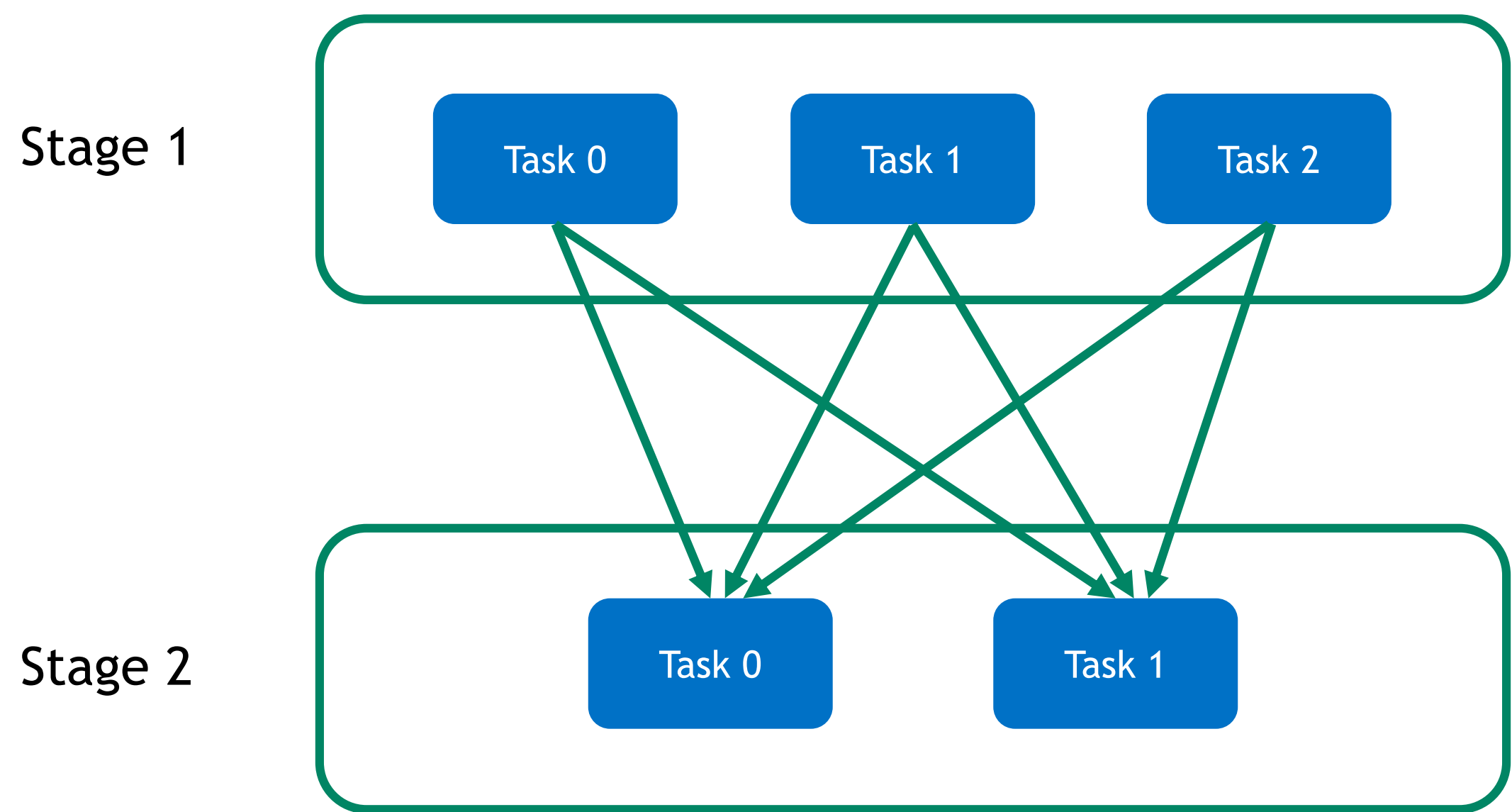




ACCELERATED SHUFFLE

SPARK SHUFFLE

Data Exchange Between Stages



UCX LIBRARY

Unified Communication X

Abstracts communication transports

Selects best available route(s) between endpoints

TCP, RDMA, Shared Memory, GPU

Zero-copy GPU memory transfers over RDMA

RDMA requires network support (IB or RoCE)

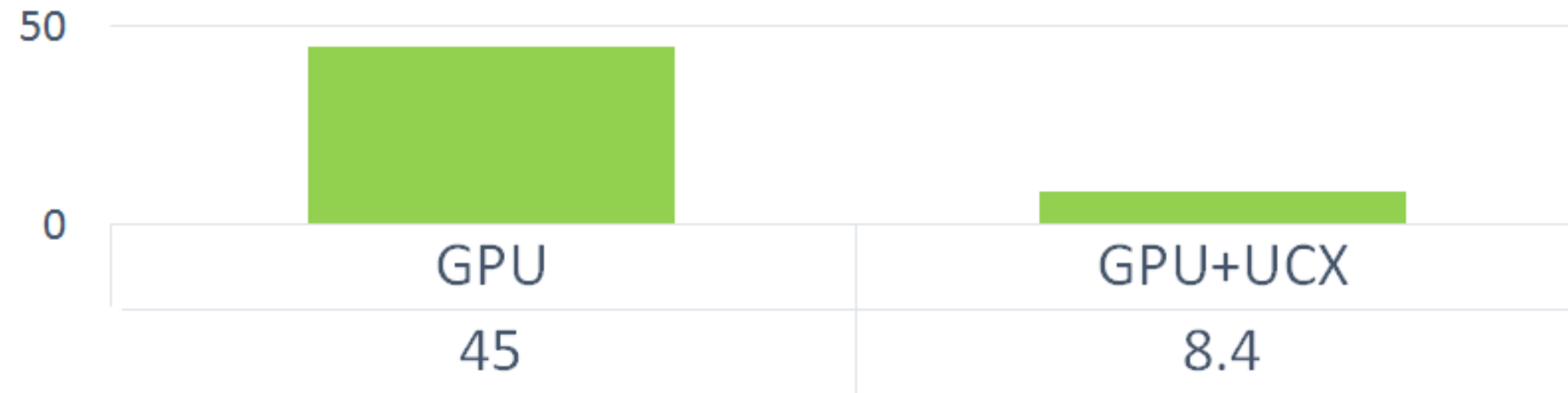
<http://openucx.org>



ACCELERATED SHUFFLE RESULTS

Inventory Pricing Query

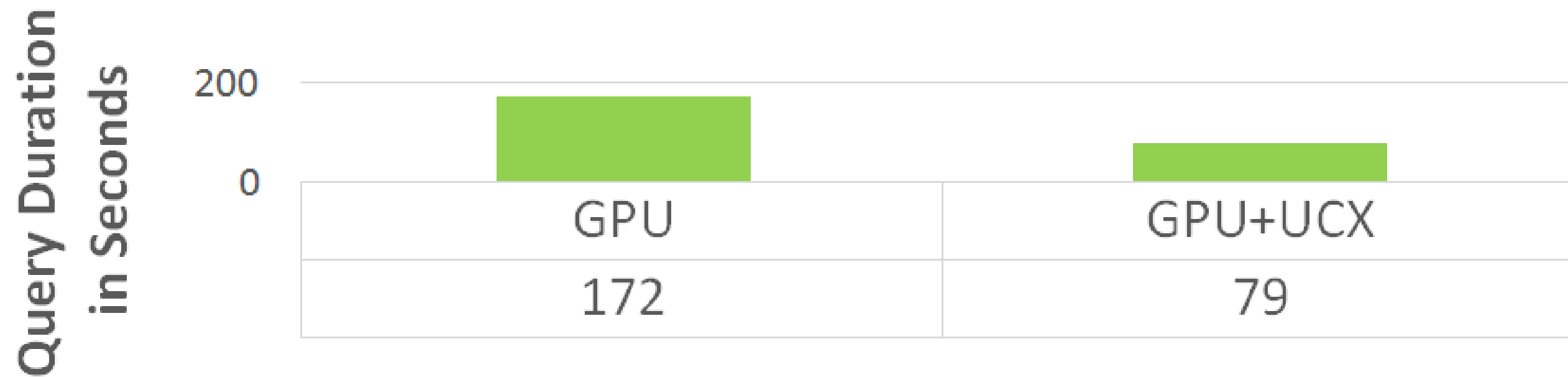
Query Duration
in Seconds



UCX for Apache Spark Enables >5X Reduction in Time for Inventory Pricing Queries

ACCELERATED SHUFFLE RESULTS

ETL for Logistical Regression Model



UCX for Apache Spark Enables >2X Reduction in Query Duration

