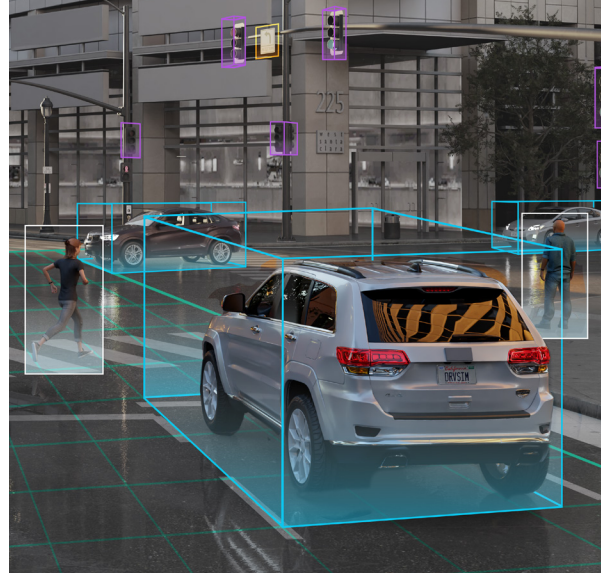




NVIDIA RAPIDS Accelerator for Apache Spark

Faster results and lower cost for data analytics and machine learning in the public sector.



Government agencies work with extremely large, often sensitive datasets, extracting insights to protect the public health, facilitate commerce, and deliver other vital services.

Many enterprises use Apache Spark for key data operations, such as ingesting raw data into data lakes, analyzing business processes, loading data into data warehouses, and data preprocessing at the start of machine learning pipelines. An estimated **80 percent of the Fortune 500 use Apache Spark 3.**

However, growing workloads are now being constrained by slow, CPU-based infrastructure. To keep up with critical data operations, government agencies have had to rewrite software code to optimize use of current infrastructure or increase capital expenditures to build more data center capacity. As the need for data analytics grows, Gartner® projects government spending on data center systems will grow at a compound annual growth rate (CAGR) of 7.4 percent to reach \$37.4 billion in 2027¹.

Slow processing costs time and money—and creates a larger carbon footprint. Meanwhile, data processing workloads are becoming increasingly critical to basic government functions and delivery of citizen services.

The Solution: Accelerate Apache Spark

The **NVIDIA RAPIDS™ Accelerator for Apache Spark** takes advantage of NVIDIA GPUs to accelerate Apache Spark workloads without code changes. It operates as a plug-in to popular Apache Spark platforms. The RAPIDS Accelerator speeds up relevant Spark operations while allowing others to continue running on the CPU. As a result, processing time can be accelerated up to 5X or more, allowing the same work to be completed with 4X less infrastructure cost.

5X

Faster Execution Time

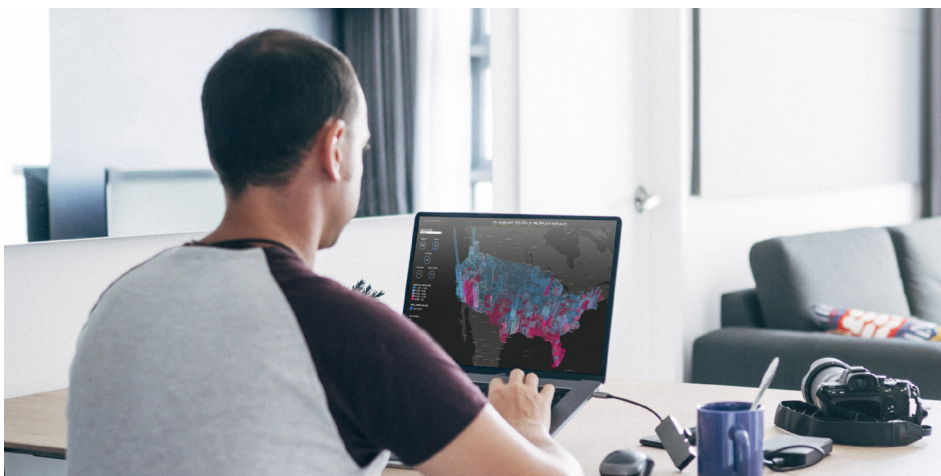
- > Move data in and out of data lakes more quickly
- > Take advantage of faster analytics
- > Accelerate AI pipelines

4X

Lower Costs

- > Save on cloud usage costs
- > Reduce power consumption and carbon footprint

¹ Gartner, Forecast: Enterprise IT Spending for the Government and Education Markets, Worldwide, 2021-2027, 4Q23 Update, By Daniel Snyder, Marlena Brown, Kelly Calhoun Williams, Irma Fabular, Inna Agamirzian, Rika Narisawa, James Ingham, February 2 2024.

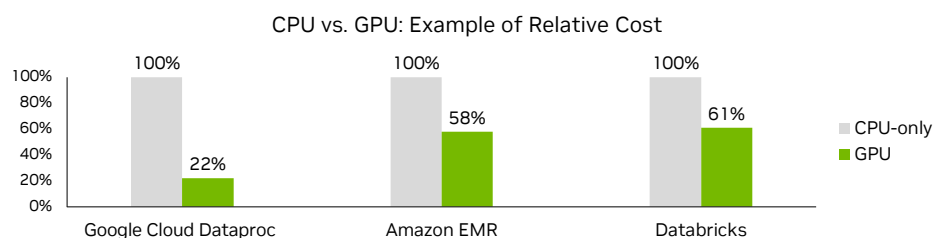


Who Needs the RAPIDS Accelerator?

All public sector organizations with growing data operations can benefit by accelerating Apache Spark workloads on GPUs. Accelerating data analysis allows enterprises to achieve faster time to results, be more cost-effective, and get more done with the same investment.

Quickly Analyze Workloads for Cost and Performance

The RAPIDS Accelerator includes the **Accelerated Spark Analysis Tool**, which analyzes workloads and provides savings estimates. This makes it easy to determine which workloads would benefit most from GPU acceleration.



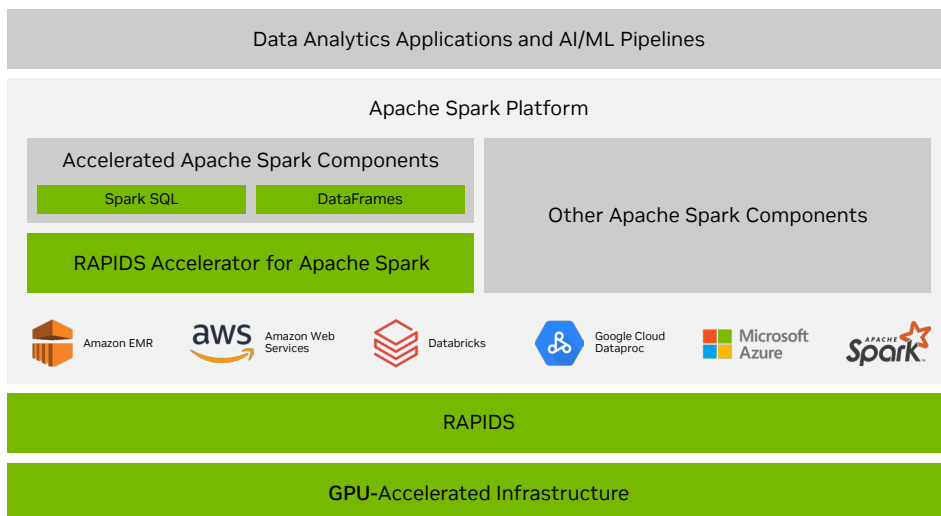
Based on NDS benchmark.
Does not include NVIDIA AI Enterprise license cost.

How Does It Work?

The RAPIDS Accelerator supports data analytics workflows for GPU-accelerated infrastructure. By operating as a plug-in to Apache Spark platforms, it accelerates workflows without requiring code changes. It can work with Spark standalone, YARN clusters, and Kubernetes clusters. Additionally, advanced features such as adaptive query execution for SQL and dynamic partitioning pruning ensure consistent, accelerated performance.

Government Spotlight

To overcome data analysis challenges, the **IRS began using the NVIDIA RAPIDS Accelerator** for Apache Spark with Cloudera. With Cloudera running on NVIDIA GPUs, workloads immediately ran up to 5X faster with no code changes. When the IRS team ran new code optimized for Apache Spark and graph analysis on GPUs in a distributed Spark cluster, they experienced a remarkable speedup of 20X. This yielded a 50 percent reduction in infrastructure costs for data science and data engineering workflows.



NVIDIA AI Enterprise Software Platform

The RAPIDS Accelerator is available with NVIDIA AI Enterprise, an end-to-end, secure, cloud-native AI software platform optimized to accelerate enterprises with industry-proven AI. NVIDIA AI Enterprise delivers validation and integration for NVIDIA AI open-source software, access to AI solution workflows to speed time to production, certifications to deploy AI everywhere, and enterprise-grade support, security, and API stability while mitigating the potential risks of open-source software. NVIDIA AI Enterprise is optimized and tested for production deployments on all major Spark platforms.

NVIDIA Enterprise Support

NVIDIA AI Enterprise includes **NVIDIA Enterprise Business-Standard Support**. With direct access to NVIDIA experts to support implementation and optimization, defined service-level agreements, and priority security notifications and fixes, this service ensures that projects stay on track.

Ready to Get Started?

To learn more about GPU-Accelerated Apache Spark, visit: nvidia.com/spark-tool

To learn about NVIDIA public sector customers and solutions, visit: nvidia.com/public-sector