CLOUDERA

Accelerate data engineering and data analytics in the enterprise

NVIDIA GPU ACCELERATION ON CLOUDERA DATA PLATFORM

Solution Highlights

- Expand AI use cases with a complete production ML toolkit enabled by NVIDIA computing
- Generate models that produce highly accurate data and insights
- trusted by the business Operate a fully secure ML environment that can meet evolving requirements
- Reduces ML training time and the frequency of model deployment from days to minutes

"With Cloudera Data Platform Powered by NVIDIA, enterprises will be able to seamlessly accelerate data analytics on critical applications like Spark 3.0 without any code changes. These breakthroughs will enable companies to analyze data in real time to gain the intelligence needed to navigate evolving customer demands."
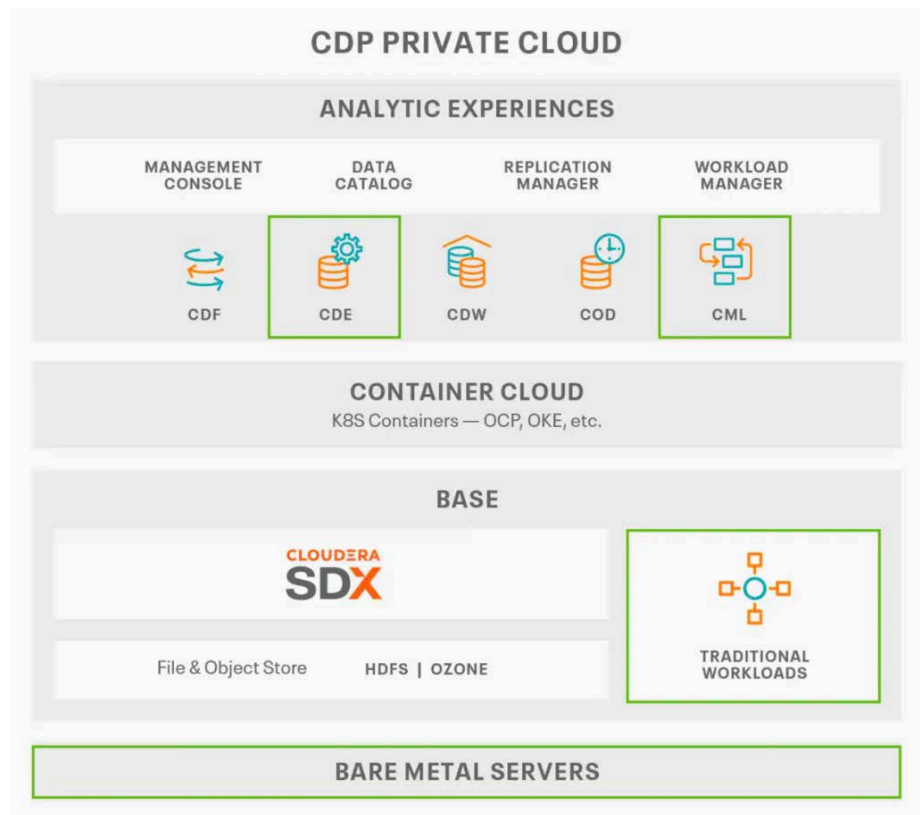
Manuvir Das, Head of Enterprise Computing, NVIDIA

Why Cloudera & NVIDIA

Today, data processing and data engineering has become the world's largest computing segment. Modest improvements in the accuracy of analytics models translate into billions to the bottom line. To build the best models, data scientists toil to train, evaluate, iterate, and retrain for highly accurate results and performant models. With RAPIDS on the Cloudera Data Platform (CDP), processes that took days now take minutes, making it easier and faster to build and deploy value generating models.

Optimized For Spark 3.0 Workloads

Enterprises can easily leverage GPU-accelerated Apache Spark 3.0 on CDP to remove bottlenecks and quickly improve performance - significantly improving time to insight and the return on investment for data-driven enterprises.

CDP PRIVATE CLOUD — ANALYTIC EXPERIENCES: MANAGEMENT CONSOLE, DATA CATALOG, REPLICATION MANAGER, WORKLOAD MANAGER. CDF, CDE, CDW, COD, CML. CONTAINER CLOUD — K8S Containers — OCP, OKE, etc. BASE — CLOUDERA SDX, File & Object Store, HDFS | OZONE, TRADITIONAL WORKLOADS. BARE METAL SERVERS.

NVIDIA Accelerated

Easily deploy end-to-end data science pipelines on Cloudera Data Platform running on NVIDIA accelerated infrastructure to improve your data-driven operations.

CLOUDERA DATA PLATFORM

- Accelerate your data science toolchain with minimal code changes and deliver hassle-free scaling on NVIDIA-Certified systems
- Increase data science productivity with visibility, security, and governance of the complete data science lifecycle

PROOF POINTS

- GPU-computing runs Apache Spark jobs in CDP 10x faster than legacy Hadoop CPU-based infrastructure and 5x faster over modern CPU-based servers
- Including both the cost of accelerated hardware and Cloudera's software premium for GPUs, achieve better than 3x cost-relative benefit

NVIDIA-CERTIFIED SYSTEMS

- Cost-effective NVIDIA infrastructure empowers IT teams to an accelerated CDP solution for intuitive, self-service ML—now and into the future.
- Accelerated CDP Starter Solutions available to enable enterprises to confidently deploy scalable hardware and software solutions that securely and optimally run accelerated workloads.

Learn more at nvidia.com

About Cloudera
At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at: cloudera.com

Joint Solution Overview

Running data science workloads on an accelerated Cloudera Data greatly improves time to value by enabling data scientists to collaborate in a single unified platform that is all inclusive for powering any AI use case. With the latest release, accelerated Apache Spark 3.0 workloads now run seamlessly on CDP. With GPU acceleration, data science teams can leverage purpose-built tooling for agile experimentation, data analytics and machine learning 10x faster and at lower cost.

Future Proof Architecture

Cost-effective NVIDIA infrastructure empowers IT teams to deliver an accelerated CDP solution for intuitive, self-service ML — now and into the future. NVIDIA-Certified servers are available from leading OEM server vendors. For companies looking to jumpstart their AI journey, Accelerated CDP Starter Solutions are available to confidently deploy scalable hardware and software solutions that securely and optimally run accelerated workloads.

Joint Solution Benefits

NVIDIA and Cloudera have tested and benchmarked workloads across a wide range of infrastructure configurations and boiled it down to two simple recommendations:

- For companies buying servers dedicated for running Apache Spark for data analytics and ETL in CDP, a CDP-READY configuration comprised of four NVIDIA-Certified servers with two NVIDIA A30 GPUs per server offers over five times the performance at less than 50% incremental cost relative when compared to modern CPU-only alternatives.

- For companies buying servers for running not just Apache Spark but also machine learning in CDP, or if these servers may be used for other AI-related applications during their lifetime, upgrade to an AI-READY configuration comprised of four NVIDIA-Certified servers with one NVIDIA A100 GPU per server offers over eight times the performance at less than 50% incremental cost relative when compared to modern CPU-only alternatives. And these numbers are just the Apache Spark benchmarks; acceleration on ML and AI training is even more significant.

Next Steps

NVIDIA-Certified Systems on CDP dramatically boosts the performance of your end-to-end analytics workflows, speeding up the time to value while reducing overall cost. NVIDIA-Certified servers are available from leading OEM server vendors and come preconfigured, tested, and validated to work for a variety of AI-related workloads—including Apache Spark. Ready to start your AI journey? Accelerated CDP Starter Solutions are the easiest path to deploy GPU-accelerated CDP.

CLOUDERA

Cloudera Inc. 5470 Great America Pkwy, Santa Clara, CA 95054 USA cloudera.com