Cloudera and NVIDIA Accelerate Data Analytics and AI in the Cloud

SANTA CLARA, Calif., April 12, 2021 — Cloudera, (NYSE: CLDR), the enterprise data cloud company, today announced that Cloudera Data Platform (CDP) will integrate NVIDIA RAPIDS-accelerated Apache Spark 3.0 libraries, intended to enable enterprises to accelerate data pipelines and improve performance of data and ML workflows. With the release earlier this year of Applied ML Prototypes (AMPs) in CDP and the power of NVIDIA GPU hardware, customers like the IRS and XYZ can access packaged ML use cases, "but also accelerate data processing and model training at a lower cost across any on-premises, public cloud or hybrid cloud deployment," according to the companies.

Enterprise data engineers are utilizing data sets on a magnitude and scale never seen before, such as transforming supply chain models, responding to increased levels of fraud, or developing new product lines. For data scientists, the bottlenecks created by massive amounts of data directly impact the cost and speed at which companies can train and operate models across the organization. Cloudera and NVIDIA's integration is expected to give enterprises the ability to quickly respond to emerging and ongoing business challenges and deliver insightful analytics.

"We need to be able to make accurate decisions at speed utilizing vast swathes of data. That challenge is ever-evolving as data volumes and velocities continue to increase," said Joe Ansaldi, IRS/Research Applied Analytics & Statistics Division (RAAS)/Technical Branch Chief. "The Cloudera and NVIDIA integration will empower us to use data-driven insights to power mission-critical use cases such as fraud detection. We are currently implementing this integration, and are already seeing over three times speed improvements for our data engineering and data science workflows."

For every company struggling with massive data sets, an open-source GPU-accelerated data science pipeline means the difference between being able to train models or never being able to do them at all. Such a pipeline can directly empower an organization's ability to transform using artificial intelligence. GPU-accelerated Apache Spark 3 runs seamlessly on CDP, allowing organizations to support their HPC, use cases, and data science needs – from research to production – with a secure, scalable, and open platform for machine learning.

"At a time when speed is everything, businesses are relying on the power of data more than they ever have. Our partnership with NVIDIA will give customers the rocket fuel they need to better understand their data and realize the true transformational potential of AI," said Arun Murthy, Chief Product Officer, Cloudera. "CDP analytic experiences are purpose-built to enable data specialists to confidently navigate the storm of both exponential data growth and siloed data analytics, operating across multiple public and private clouds. Deepening our existing integration with NVIDIA is a natural next step for us. Our customers will be able to maintain the competitive edge they already have by using our enterprise data cloud services."

"Cloudera and NVIDIA's collaboration over the years has helped companies around the world make better data-driven decisions, faster," said NAME, ROLE, NVIDIA. "By leveraging NVIDIA's RAPIDS capabilities, data scientists and engineers working in Cloudera Data Platform will be able to deliver analytics and machine learning models built on massive amounts of data. This integration will accelerate their machine learning and data engineering processing power to improve operational accuracy and deliver faster, more simplified iteration workflows at scale."

The public cloud implementation of NVIDIA RAPIDS-accelerated Apache Spark 3.0 libraries is now GA and the on-premise product partnership will be GA this summer.

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.