



NVIDIA AND ANACONDA GPU-ACCELERATED OPEN DATA SCIENCE PLATFORM

We're in the midst of a perfect storm in which open-source, next-generation hardware and the unrelenting growth of data have unleashed new opportunities to change the world. Traditional, slow CPU infrastructures are no longer meeting customers' demands. To take data science to the next level and unlock the potential value in these rich, new data sources, we need to empower data science teams with the latest and most innovative data science methods. These include AI and machine learning, as well as new compute resources that can crunch more data faster.

INTEGRATED SOLUTION

Anaconda is the largest data science platform with more than 6 million users. The joint solution—powered by NVIDIA® DGX™ Systems and NVIDIA GPUs—enables data science teams and accelerates time-to-value by connecting data, analytics, and computation. Anaconda, NVIDIA, and the data science community have collaborated to democratize Python by establishing a new common data framework, called GPU Data Frame. This framework eliminates data transfers between CPUs and GPUs to deliver huge performance gains for data science workloads. By moving the compute to the data and eliminating data latency, Anaconda and NVIDIA let users easily deploy AI apps into production with hyperspeed GPU-accelerated Python, the fastest growing data science language.

INDUSTRY CHALLENGES

Open-source chaos: Inability to manage open source in an enterprise environment

- > The rate of change in open source presents unique challenges to enterprise IT governance and compliance policies and regulations.
- > The breadth and depth of open-source dependency graphs needed by demanding high-level analytics are rapidly increasing, making it more difficult to manage workflows.




Hardware limiting innovation: Inability to easily build data and computationally intensive data science models

- > Data movement, especially on big data, introduces latencies as copying data from CPUs to GPUs is limited by the available memory and network speed that curtails the use of innovative data science methods.

Operational inefficiencies: Inability to easily build data and computationally intensive data science models

- > Open-source isn't ready for enterprise production environments.
- > Deploying data science in production requires additional provisioning and coding that data science teams aren't equipped for.

Together, NVIDIA and Anaconda Deliver

ACCESSIBLE PYTHON	GPU-ACCELERATED PYTHON	PRODUCTION-READY PYTHON
 <p>Democratize Python to make it available for all data scientists.</p> <p>Anaconda has been downloaded over 30 million times with over 6 million users and is used for AI and machine learning data science workloads. These include TensorFlow, Theano, Keras, Caffe, Neon, Lasagne, NLTK, and spaCY, plus over 1,000 more Python packages.</p> <p>560,000 GPU developers ensure that GPU-accelerated Python is accessible everywhere.</p>	 <p>Achieve unparalleled acceleration with Python on GPUs for data science workloads powered by AI.</p> <p>GPUs have 4,000+ cores per device in some cases, versus 16 to 32 cores in typical CPU-based devices. This translates into tangible hardware savings of approximately 90 percent on average.</p> <p>NVIDIA NVLink™ lets users scale multi-GPU systems. It's a high-bandwidth, energy-efficient interconnect that allows data sharing at rates 5X to 12X faster than the traditional PCIe interconnects.</p>	 <p>Move beyond ad hoc analysis to AI-driven knowledge and insights with production-ready applications for business-critical functions.</p> <p>Deploy AI applications using Python language with a few clicks.</p> <p>Customers on average experience a 10X or more speedup with NVIDIA GPUs.</p>

Industry Insights

CATEGORIES	USE CASES
<p>Ad Tech</p>	<ul style="list-style-type: none"> > Ad bidding: Train real-time ad-bidding models that evaluate billions of impressions to identify target segments, predict likely response rates, and determine optimized ad-bidding strategies. > Cross-selling recommendations: Evaluate large quantities of customer purchase and behavioral data to create contextually relevant product recommendations and drive top-line revenue. > Multi-channel attribution: Determine marketing channel effectiveness to optimize spend and activities by evaluating customer transactions and behaviors over time.

<p>Financial Services</p>	<ul style="list-style-type: none"> > Fraud detection: Filter data and identify patterns of unusual behavior with machine learning as quickly as possible to minimize losses due to fraud. > Seeking alpha: Seek new sources of alpha within huge stock and equity trade data to instantly respond to changing market conditions. > Risk management: Identify leading indicators of operational or credit risk to minimize or prevent losses.
<p>Gaming</p>	<ul style="list-style-type: none"> > Cohort analysis: Use player events and engagement data to segment players based on behavioral patterns to adjust in-game promotions. > Multi-channel attribution: Discover media performance across multiple advertising channels and use it to tailor in-game onboarding, offer incentives, and deliver relevant content. > Fraud detection: Prevent fraudulent traffic from being attributed to your paid channels with IP filtering and distribution modeling that detect anomalies.
<p>Government</p>	<ul style="list-style-type: none"> > Intrusion detection: Identify an attack that's currently in progress. > Forensic analysis: Understand the impact of an attack and prevent future intrusions. > Fraud collusion: Evaluate longer periods of complex behavior to identify networks of collusion between actors and predict potential future attacks.
<p>Insurance</p>	<ul style="list-style-type: none"> > Claims fraud: Model claims data to predict changes in type and frequency of insurance claims due to external factors. > Telematics: Monitor customers' driving behavior and actions to create incentives that influence behavior and drive down claims. > Health risk modelling: Blend actuarial data with rich, new data sources—including IoT, environmental factors, and the latest medical research—to assess patient risk levels.
<p>Manufacturing</p>	<ul style="list-style-type: none"> > Yield optimization: Improve production processes through discovery of factors that correlate with higher defect rates. > Product design recommendations: Prescribe product designs by evaluating past designs, customer usage data, and target market criteria. > Demand sensing: Monitor and collect demand signals from point-of-sale to prevent stock outages by automatically triggering replenishment.

Recommended Infrastructure

NVIDIA GPUs are available in servers, supercomputers, and cloud platforms around the world. You can now get end-to-end accelerated analytics solutions powered by NVIDIA GPUs with supporting software technologies and support from NVIDIA experts.

NVIDIA® TESLA® SERVERS IN EVERY SHAPE AND SIZE		DGX SYSTEMS THE ESSENTIAL AI TOOLS FOR INSTANT PRODUCTIVITY		CLOUD EVERYWHERE	
      				   	

Find Out More

GPU-accelerated analytics help customers effectively analyze, visualize, and unleash the power of AI to transform their digital business into an AI enterprise.

Website: www.nvidia.com/analytics

Contact: dgxanalytics@nvidia.com

Partner webpage: www.nvidia.com/dgx-apps

Twitter: [@NvidiaAI](https://twitter.com/NvidiaAI)

Blog: blogs.nvidia.com

Anaconda is the world's most popular Python data science platform.

Website: www.anaconda.com/enterprise

Contact: sales@anaconda.com

Twitter: [@anacondaic](https://twitter.com/anacondaic)

Blog: www.anaconda.com/blog/developer-blog/