



## NVIDIA AND BLAZINGDB GPU-ACCELERATED DATA WAREHOUSE

For today's enterprises, data growth is exponential. But current technologies are failing to grow at the same rate, leading to an analytics bottleneck. Organizations are looking to increase revenue and efficiency by merging disparate data sets across the business. Data driven decisions are also requiring more individuals to step into analytical roles. But the only tools that handle organization-wide information require high-level coding skills, and are very complicated to implement institutionally. A simpler, more scalable solution is needed that can help customers structure appropriate processes, software, and hardware to handle this monumental increase in data.

## Industry Challenges

- > **ETL is Difficult** - Legacy databases require a unique ingest process—either manual or automated—for every dataset it needs to analyze. The variety and scale of data sources make these processes prohibitively expensive for many data sets, locking up valuable insights in enterprise data lakes.
- > **Scale Out is Expensive** - Distributed computing tools, such as MapReduce, require highly skilled engineering professionals. Massively parallel processing (MPP) databases are easier to use, but when they need to scale, engineering must transfer it to a larger cluster, purge old data, or engage in costly tuning.
- > **Existing Infrastructure Hinders Productivity** - Legacy MPP databases run queries on terabyte data sets in minutes to hours. As query times increase, the productivity of analysts, reporting systems, and operational processes is significantly reduced.

## Integrated Solution


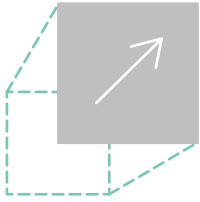
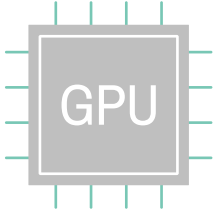
BlazingDB is a high-performance SQL analytics engine powered by on NVIDIA® DGX™ systems and GPUs. It runs at the computational scale of supercomputer clusters, helping enterprise organizations to run incredibly fast SQL on their enterprise data lakes and perform analytics directly on raw files.

BlazingDB is fully integrated with high-performance, open-source data pipelines. By harnessing the power of GPUs, BlazingDB can query Apache Parquet files—which is a columnar storage format available to any project in the Hadoop ecosystem—in minutes. It also integrates with Apache Arrow, a columnar in-memory analytics layer designed to accelerate big data. This allows BlazingDB to hand data off directly in memory to popular data pipeline technologies.

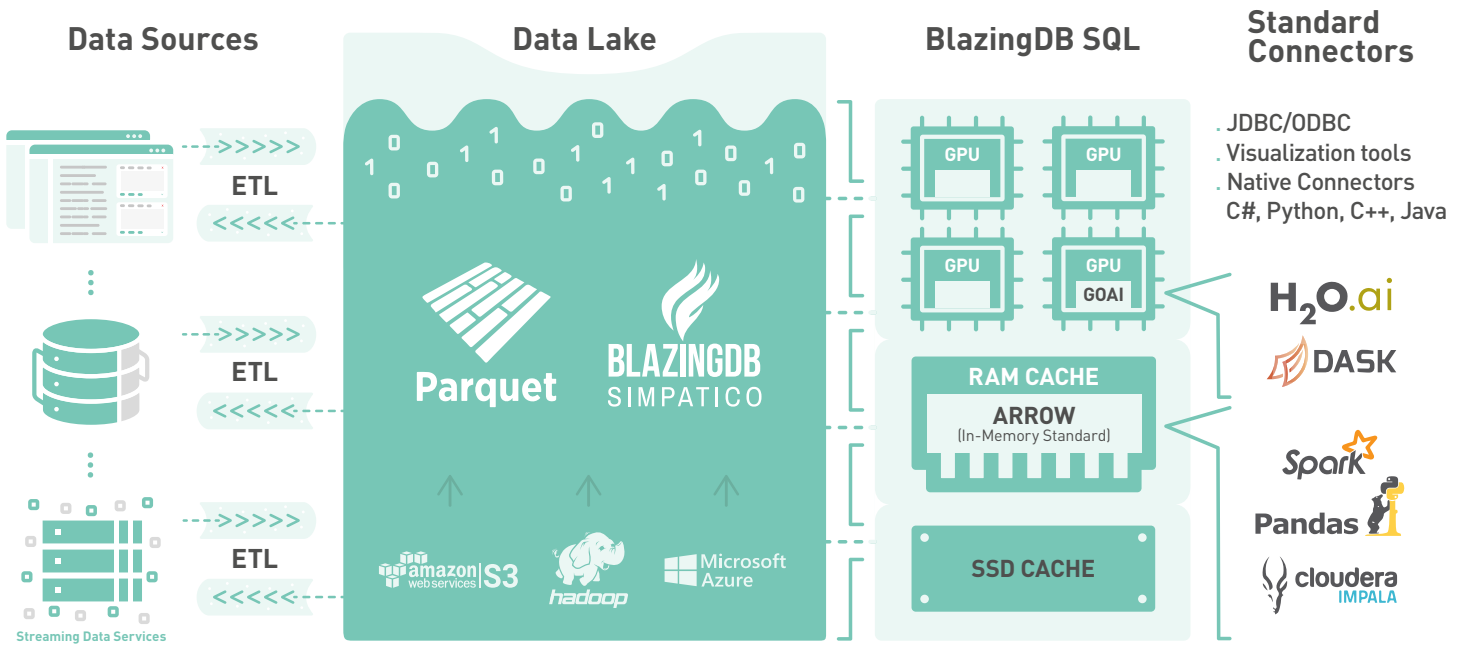
Finally, as a member of the GPU Open Analytics Initiative (GOAI), BlazingDB is fully integrating the GPU Data Frame (GDF). The GDF enables other GPU-accelerated technologies, such as deep learning, visualization, and advanced analytics, to keep the data on GPUs and seamlessly integrate with BlazingDB, using high performance connectors.

Customers can now seamlessly integrate high-performance, GPU-powered SQL analytics into their current data pipeline architecture. This delivers substantive productivity gains without adding complexity to an already complicated IT landscape.

# Together, NVIDIA and BlazingDB Deliver

SEAMLESS DATA PIPELINE INTEGRATION	HIGHLY ELASTIC SCALE OUT	BLAZINGLY FAST SQL
 <p>Full integration with high-performance, open-source and popular commercial data pipeline technologies.</p> <ul style="list-style-type: none"> <li>&gt; Direct queries to Apache Parquet and BlazingDB's Simpatico (GPU-compressed distributed file format) off Hadoop and Cloud Data Lakes</li> <li>&gt; Apache Arrow integration</li> <li>&gt; Member of GOAI</li> </ul>	 <p>Highly elastic infrastructure with very accessible SQL analytics language.</p> <ul style="list-style-type: none"> <li>&gt; One Model, the workforce analytics platform, experienced 30% reduction in Redshift spend month over month</li> <li>&gt; Simple SQL query with no re-training needed</li> <li>&gt; Self-configuration for full optimization and easy administration</li> </ul>	 <p>The ability to process terabytes of data on GPU-accelerated, state-of-the-art query execution engine in seconds.</p> <ul style="list-style-type: none"> <li>&gt; One Model ran 150 queries that showed a speedup of 2.5X - 10X against Redshift</li> <li>&gt; Pharma customer increased multi-terabyte query speeds by 20X at no additional cost</li> </ul>

## BlazingDB Marketecture



## Industry Insights

Customers use the massively parallel compute power of GPUs to provide higher throughput for compute-intensive workloads and achieve significant performance gains, without the hidden cost of scale-out architecture. This can result in dramatic cost savings.

FINANCIAL SERVICES	PHARMA	RETAIL
Run huge analyses across billions of transactions to identify which customers are best suited for increased credit lines, using simple, distributed, and scalable SQL data warehouse.	Query billions of data points in subseconds in a SQL data warehouse to identify high risk factors of certain diseases across exponentially growing genetic data.	Run pricing and profit calculations to deliver the right products at the right price at the right time, leveraging large scale, distributed SQL query languages.

## Recommended Infrastructure

NVIDIA GPUs are available in servers, DGX Systems, 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.

<b>TESLA</b> SERVERS IN EVERY SHAPE AND SIZE		<b>DGX SYSTEMS</b> AI TOOLS FOR INSTANT PRODUCTIVITY		<b>CLOUD EVERYWHERE</b>	
      		   <small>an IBM Company</small>			

## Find Out More

NVIDIA Accelerated Analytics - Helping customers effectively analyze, visualize, and unleash the power of AI to transform their digital business into an AI enterprise.

Website: [www.nvidia.com/analytics](http://www.nvidia.com/analytics)

Contact: [dgxanalytics@nvidia.com](mailto:dgxanalytics@nvidia.com)

Partner Webpage: [www.nvidia.com/dgx-apps](http://www.nvidia.com/dgx-apps)

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

Blog: [blogs.nvidia.com](http://blogs.nvidia.com)

BlazingDB - Helping enterprise organizations reach the limits of speed and scale of high-performance data warehouses through a next-generation SQL database on GPUs.

Website: [www.blazingdb.com](http://www.blazingdb.com)

Contact: [info@blazingdb.com](mailto:info@blazingdb.com)

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

Blog: [www.blog.blazingdb.com](http://www.blog.blazingdb.com)