Solution Brief

OmniSci GPU Analytics Platform and NetApp ONTAP AI

Speed time to discovery with GPU-accelerated analytics

Key Benefits

Accelerate data analytics and data science
Query and visualize billions of rows of data in milliseconds for rapid time-to-business insights to gain a competitive edge.

Combine geospatial and business intelligence
Fuse large geospatial data to get real-time location intelligence. Plot and render millions of spatiotemporal points in milliseconds.

Lower TCO
Ensure maximum storage savings backed by the industry's most efficient guarantee.

Protect your data across your pipeline
Protect data in flight, in use, or at rest, and meet compliance requirements with confidence.

Start small, scale seamlessly
Start small and independently scale storage and compute as your needs change.

Simplify design and deployment
Get started faster with a validated reference architecture that detangles design complexity.

The Challenge

An explosion in IoT and big data has created opportunities for enterprises in all industries to create new value across their organizations. Enterprises that can derive more insights faster than their competition will gain the edge when it comes to optimizing operations and creating new customer touchpoints and business opportunities.

To achieve and maintain a competitive edge, today's data-driven enterprises demand analysis at the speed of thought. However, mainstream analytics tools designed to run on CPUs lack the performance needed to keep up, providing a "click-and-wait" experience that could take hours or days. This poses numerous challenges throughout the organization:

- Data scientist teams are unable to explain AI models to business leadership because of the extreme difficulty of visually exploring billion-row or large spatiotemporal datasets.
- Teams ask only questions they know will be answered because machine learning model development, including feature engineering and ongoing monitoring, is overly time consuming on large datasets.
- Teams do only what is asked of them because issuing iterative queries is enormously time consuming, hindering their ability to fully explore ideas.
- Teams run out of time on a particular problem area, which forces them to downsample, introducing risk.

The Solution

Together, NetApp, NVIDIA, and OmniSci provide a complete, integrated solution for real-time IoT analytics and machine learning. NetApp® ONTAP® AI with OmniSciDB offers extreme performance, space-saving efficiency, and integrated data protection for big data analytics and AI in a prevalidated, GPU-accelerated solution that is easy to deploy and easy to manage.

OmniSciDB is the world's fastest open-source, GPU-accelerated database. OmniSci harnesses the massive parallel processing of GPUs to deliver SQL queries across billions of records in milliseconds, making it possible to visualize IoT big data and support machine learning in real time.

NetApp ONTAP AI is a proven architecture for AI that combines NVIDIA DGX supercomputers and NetApp AFF all-flash storage. ONTAP AI enables you to simplify, accelerate, and integrate your data pipeline with a data fabric that spans from the edge to the core to the cloud.

NetApp®

omnisci
Technology partner.

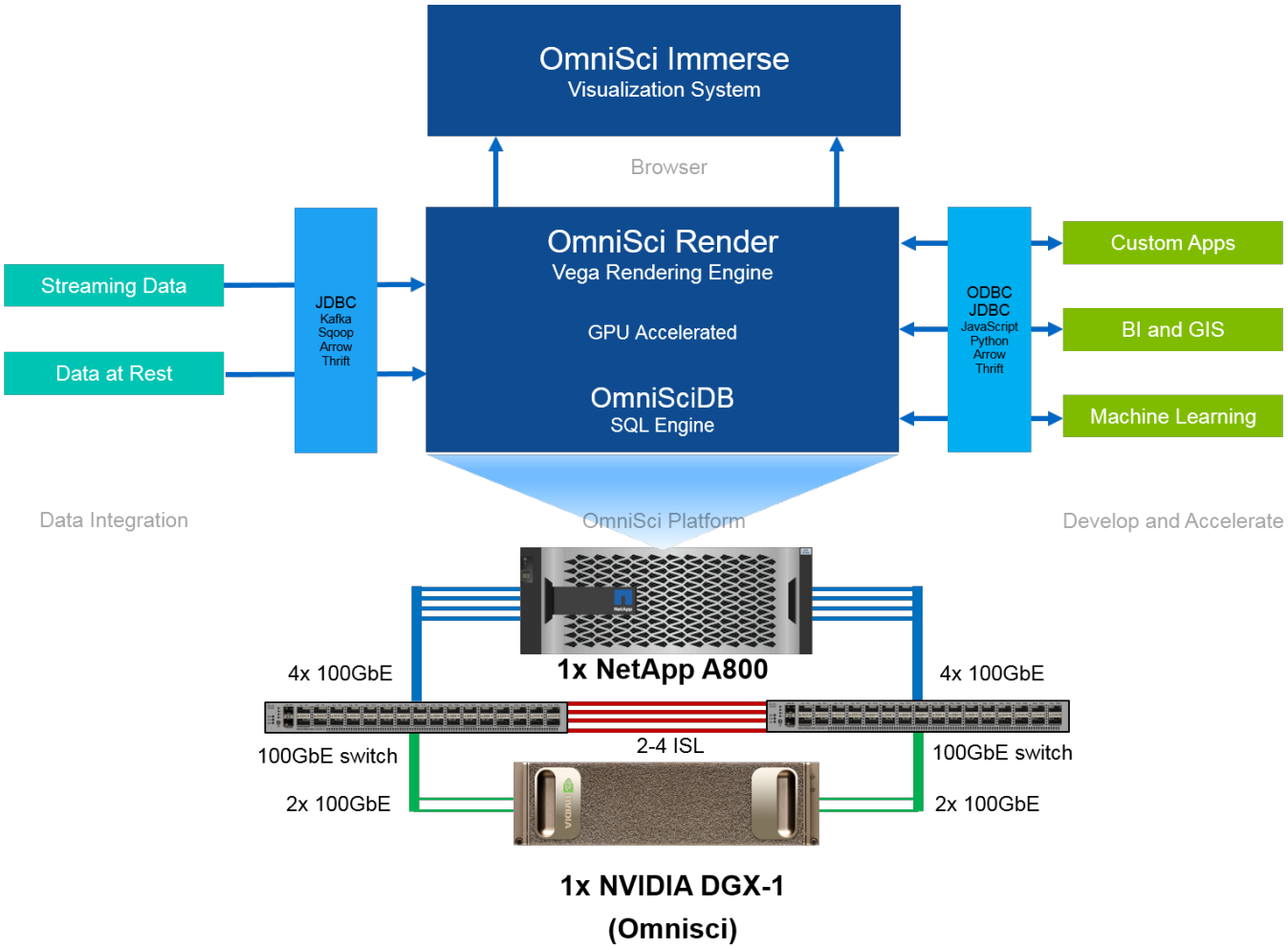

Figure 1) NetApp ONTAP AI and OmniSci solution architecture overview.

Accelerate Data Analytics and Data Science

OmniSciDB is the ideal SQL engine for IoT big data. Specifically developed to harness the parallel processing power of GPUs, OmniSciDB is capable of unprecedented ingestion speeds and can query up to billions of rows in milliseconds. With OmniSci, you can ask and answer at least 100x more questions in the same amount of time as it would take for one query on a CPU-only platform. Make faster business decisions with more efficient machine learning model design and management, including feature engineering and ongoing monitoring.

Unmask the black box of AI with interactive visual analytics at the speed of thought and real-time location intelligence powered by GPU rendering. With the OmniSci Immerse visual analytics client, you can visually explore up to millions of polygons and billions of mapped points in milliseconds—a task that is nearly impossible on large spatiotemporal datasets with CPU-based analytics tools.

With OmniSci and the ONTAP AI validated architecture, you get an infrastructure designed to enable reliable high performance from end to end. The NetApp AFF A800 array is capable of feeding data to NVIDIA DGX-1 systems up to four times faster than competing solutions.[1] A single NetApp AFF all-flash system supports throughput of 25GBps for sequential reads and 1 million IOPS for small random reads at under 500-microsecond latencies for NAS workloads, giving you plenty of performance headroom to support numerous DGX-1 servers as your requirements grow.

1. Read throughput of up to 300GBps per all-flash cluster versus 75GBps from a leading competitor.

Lower TCO

With ONTAP AI, high-density storage and compute enable you to ingest, manage, and analyze massive amounts of data in a small footprint. With NetApp technology, you don't have to trade performance for efficiency. NetApp AFF systems deliver consistent low latency under the most stressful workloads with no degradation even when data reduction technologies are turned on.

NetApp AFF all-flash arrays come standard with market-leading data reduction capabilities, including inline deduplication, compression, and compaction, to ensure maximum storage savings—all backed by the industry's most effective guarantee. We're so confident in the storage efficiency our AFF systems deliver, we'll guarantee in writing 3:1 for mixed workloads or 30:1 if you're using NetApp Snapshot™ technology.

When paired with OmniSci, ONTAP AI's inline storage efficiency feature can reduce total storage requirements by as much as 20% after volume deduplication and volume compression. Internal validation testing demonstrated a 34% reduction in OmniSci data storage, a 14% reduction in binary data, and a 49% reduction in log data.

Protect Your Data Across Your Pipeline

For high-velocity, big data workloads, there's no time for downtime. ONTAP AI offers integrated data protection features to securely control and protect your data across your data pipeline. Use the same tools to protect your data regardless of whether it is in flight, in use, or at rest, and meet compliance requirements with confidence.

NetApp Snapshot technology creates nearly instantaneous point-in-time data copies with no impact on performance and only minimal consumption of storage space. Snapshot can be paired with NetApp SnapRestore® technology to recover entire file systems or data volumes in a matter of seconds.

Start Small, Scale Seamlessly

ONTAP AI allows you to start small and grow as needed. Add compute, storage, and networking to clustered configurations without disrupting ongoing operations. Start with a 1:1 storage-to-compute configuration, and scale out as your data grows to a 1:5 configuration and beyond.

NetApp's rack-scale architecture allows organizations to start with an AFF A220 array and grow as needed to scale from hundreds of terabytes to tens of petabytes with all-flash. And with NetApp ONTAP FlexGroup, up to 20 petabytes of single namespace can handle more than 400 billion files.

Simplify Design and Deployment

The rapid pace of AI innovation makes designing an effective AI infrastructure challenging. With ONTAP AI and OmniSci, you can eliminate guesswork and get started faster with a validated reference architecture that detangles design complexity.

Strategic Use Cases

- Telecommunications: Network Reliability Analysis
- Automotive & Logistics: Vehicle Telematics Analysis
- Investment Management: Alternative Data Insights
- Oil & Gas: Well Log Analysis
- Cyber Incident Investigation
- Defense & Intelligence: GEOINT

About OmniSci

OmniSci (formerly MapD) is the pioneer in GPU-accelerated analytics, redefining speed and scale in big data querying and visualization. The OmniSci platform is used to find insights in data beyond the limits of mainstream analytics tools. Originating from research at MIT, OmniSci is a technology breakthrough, harnessing the massive parallel computing of GPUs for data analytics. Learn more about OmniSci at **www.omnisci.com**.

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize their operations. For more information, visit **www.netapp.com**. #DataDriven