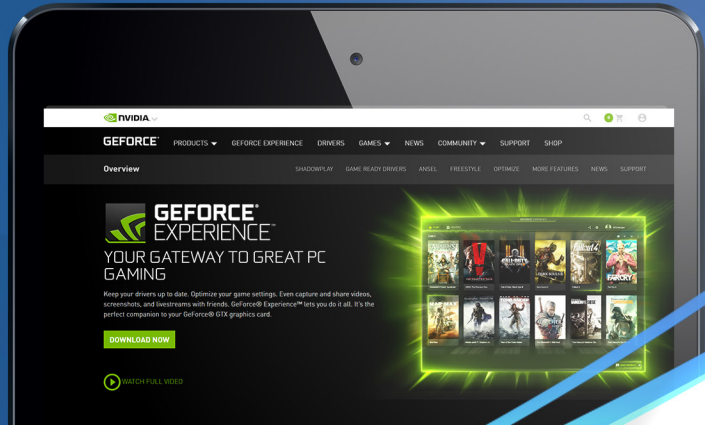


# NVIDIA

Global Customer Identity and Access Management, powered by FaunaDB



## ABOUT NVIDIA

NVIDIA's popularization of the GPU in 1999 sparked the growth of the PC gaming market, redefined modern computer graphics and revolutionized parallel computing. More recently, GPU-based deep learning ignited modern artificial intelligence – the next era of computing – with the GPU acting as the brain for computers, robots, and self-driving cars that can now perceive and understand the world around them.

## PROJECT OVERVIEW

NVIDIA creates customer-facing companion experiences for their market-leading graphics cards. These experiences allow gamers worldwide to keep their drivers up-to-date, optimize game settings, as well as capture and share videos, screenshots, and livestreams with friends. Backing these experiences is a customer identity and access management (CIAM) platform for user management.

When launching a new customer experience in 2016, NVIDIA embarked on a project to overhaul their CIAM platform to better support their global user base for that experience and to be able to scale to satisfy the demands of all new consumer experiences to be launched by the company.

## KEY REQUIREMENTS & CHALLENGES

A fundamental bottleneck to the expansion of NVIDIA's CIAM platform was its user database, the database that manages all user data within the experience. The database they were

using was cloud-based but could not scale to meet NVIDIA's increasing global needs. Also, it was heavily customized, and any changes required significant investments of time and money. Furthermore, as new consumer experiences were developed, it was hard for those systems to use the existing CIAM service within their on-premises test environments. They concluded that they needed a new CIAM platform that would:

- Scale globally in unison with consumer growth
- Scale internally with new services consuming the CIAM platform
- Deploy within multiple cloud environments (including VMWare and Amazon Web Services) for use by internal teams
- Eliminate vendor lock-in (from the existing service)
- Reduce total costs of development and operations
- Provide maximum uptime for consumer experiences with a small operations team

Correspondingly, the requirements for the new user database with the CIAM platform were just as rigorous. NVIDIA needed a database that would allow them to:

- Easily deploy and manage database clusters across multiple private data centers and public cloud environments worldwide
- Create and update data with global (distributed) transactionality
- Ensure low latency for all transactions in order to deliver the best possible end-user experience

- Deliver data with availability that satisfies NVIDIA service level agreements
- Scale or shrink clusters dynamically with changing usage patterns
- Make application queries as simple as possible to write
- Minimize the learning curve for developers using the database
- Manage service delivery and ongoing global operations with a very small team and within budget

Using these requirements, NVIDIA scored available databases, both SQL and NoSQL, and narrowed down to a list of finalists for evaluation. The final list included Microsoft SQL Server, Cassandra, and FaunaDB.

## WHY FAUNADB

FaunaDB is a relational NoSQL database that delivers data consistency, scalability, operational agility and developer flexibility in a complete, modern platform. Its architecture is optimized to deliver the safety (transactionality and reliability) of SQL systems while ensuring the scalability and flexibility of NoSQL. When evaluating the candidates for its CIAM database, NVIDIA quickly realized that only FaunaDB could satisfy all its requirements.

NVIDIA found that FaunaDB was easy to deploy across all of its cloud environments, including VMWare and Amazon AWS. Nodes could be set up, managed, and clustered without specialty hardware, additional software components, consulting engagements, or time spent developing glue code to stitch components together. With built-in replication and fault-tolerance, FaunaDB made it possible for NVIDIA to meet its desired SLAs. To say it in NVIDIA's own words: "FaunaDB just worked out of the box, with minimal operational effort."

FaunaDB's distributed transactionality made it particularly attractive. All transactions are 100% ACID, no matter the configuration of clusters and the sharding within the environment. This allowed NVIDIA to painlessly execute on its existing identity management use case, while future-proofing against a potential need for ACID transactions in subsequent projects.

Development teams found FaunaDB's query interface easy-to-use. It integrated transparently into NVIDIA's existing

programming stack with its native drivers. Queries were written in a high-level language that was more productive. The hybrid NoSQL schema combined with relational indexing ensured that queries were powerful, flexible, and a fraction of the size of their SQL or NoSQL counterparts.

FaunaDB's ability to run across public and private clouds was attractive as well as an insurance policy for unanticipated growth. It allowed NVIDIA to scale up its user experiences for big events and then back down to save costs. NVIDIA also wasn't required to standardize on any single infrastructure platform long-term, so there was no vendor lock-in.

With FaunaDB, we're able to support tens of millions of users with a small operational staff, and FaunaDB's advanced features like global replication let us maintain high availability and correctness even in the case of unexpected regional outages.

- Bill Wagner, Director of Cloud Services at NVIDIA

## RESULTS

NVIDIA launched its first experience on the new CIAM worldwide in 2016 without a hitch. Success of the initial use case was quickly followed by additional consumer-facing experiences as different teams began standardizing on the new CIAM platform. Today, FaunaDB is the core production database behind several customer-facing NVIDIA services. The key metrics are staggering:

- 100% availability of the database since the launch in 2016, despite partial failures or network partitions of underlying cloud infrastructure, through the ability to automatically route load to other physical regions of the cluster
- Increasing volumes of sustained throughput, presently at more than 30,000 requests/second (25B requests/day)
- Current footprint of 50 nodes across five global data centers, with numerous private cloud deployments for sandboxes
- Overall, one-tenth the total cost of ownership of the previous cloud-based solution