



White Paper

# GPU Database: A Disruptive Analytics Technology

*By Janak Kalaria, ICF*

## Executive Summary/Overview

With the volume, variety, and velocity of data increasing at a tremendous rate, legacy data analytics applications are falling short to meet the business needs of leading organizations. Legacy analytics technologies not only fail to provide optimum performance, they are too expensive to process and analyze large datasets.

GPU (graphics processing unit) databases, given their massively parallel processing capabilities, is disrupting the data analytics landscape by providing high performance computing in a cost-effective manner. Emerging technologies (such as machine learning and the Internet of Things) stand to greatly benefit in the dawn of this technology.



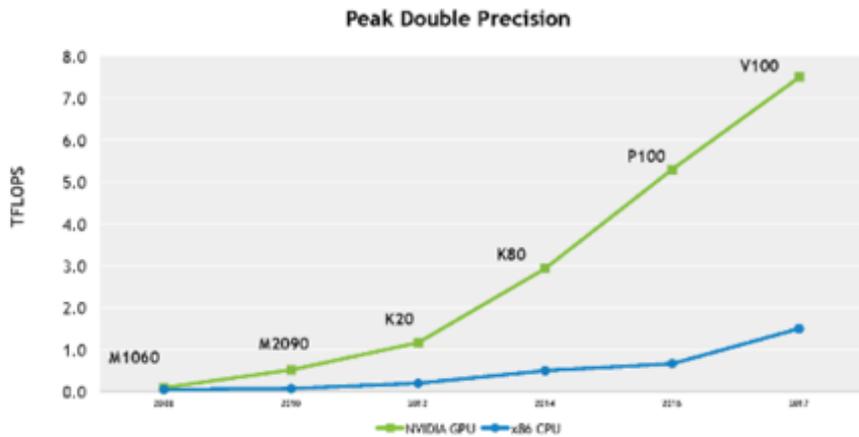
## Current Data analytics challenges

- **Performance issues affect business users' productivity:** Outdated systems struggle to ingest and query data simultaneously, thereby making it difficult to process live streaming data. Because the legacy/enterprise data warehouse and visualization platforms cannot scale with the growing volume, variety and velocity of data, users experiences delays in critical business decisions and allow key opportunities to slip away.
- **Labor/Time required to pre-process data:** Organizing, modeling, and pre-processing data requires a significant time window to integrate with legacy business intelligence platforms. This causes many data analytics projects to either fail or go over budget.
- **CPU-based platforms drive up infrastructure cost while scaling:** Current data analytics platforms are limited by compute power/capacity provided by the CPUs (Central Processing Units). Deploying several CPUs in large clusters can be cost prohibitive for all but a handful of organizations when processing large data sets.
- **Solution maintenance complexity threatens the viability of analytics solutions:** Frequent changes are often required of data models, data pre-processing scripts, hardware and software optimizations to ensure optimum system performance. Hiring and retaining staff with the necessary skillsets (e.g. data modeling, data pre-processing, data visualization development, etc.) can be costly.

## GPU Databases: A Disruptive Analytics Technology

GPU databases excel in handling large data sets. While they were once primarily used for processing graphics, over time both the programmability and processing power of the GPU have evolved to suit applications requiring high computation power. Due to their massively parallel processing capabilities, GPUs are capable of processing data up to 100 times faster than configurations containing only CPUs as illustrated in Exhibit 1. The GPU's small and efficient cores are designed to perform similar, repeated instructions in parallel -- thereby making it ideal for expediting the processing-intensive workloads required in today's data analytics applications.

## EXHIBIT 1: COMPUTE/PROCESSING PERFORMANCE COMPARISON BETWEEN CPUS AND GPUS

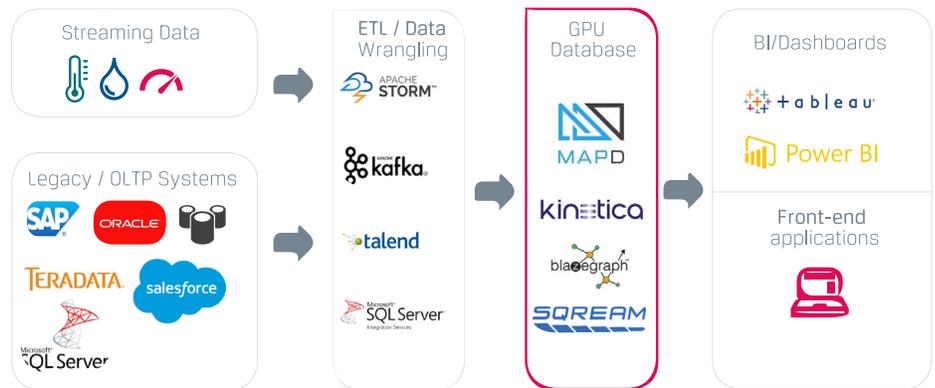


Source: NVIDIA

GPU databases offer significant benefits compared to legacy data analytics applications:

- Powered by GPU's parallel processing capabilities, databases such as MapD are able to process and analyze millions of rows of data in merely seconds.
- GPU databases don't need data to be organized in cubes, unlike legacy data analytics applications. This greatly reduces the time and labor to pre-process the data prior to analysis.
- GPU databases can be easily scaled both up and out to increase capacity and performance while being cost effective. Scaling up is done by adding more or faster GPUs, whereas scaling out entails adding more servers in a cluster.
- GPU databases are easily accessible using the familiar SQL syntax, in which existing staff can ramp up on using GPU databases with little to no training.
- GPU databases have an open architecture, making them easy to interoperate and integrate with a wide variety of existing applications and platforms as illustrated in Exhibit 2.

## EXHIBIT 2: GPU DATABASES' OPEN ARCHITECTURE FACILITATE EASY INTEGRATION WITH EXISTING APPLICATIONS/PLATFORMS



Source: ICF

## Application of GPU Databases

In addition to the benefits above, here are some practical applications of GPU:

- Machine Learning and Deep Learning:** Machine learning (ML) and Deep Learning (DL) have emerged as viable technologies for helping organizations discover actionable insights in data and, more importantly, make predictions about future or otherwise unknown events with a high degree of confidence. For example, machine learning can predict which motor vehicle drivers are most likely to be involved in an accident based on driving behavior. A key component of machine learning is training the ML model, which involves uncovering critical correlations and hidden insights in the data which dictates/informs future predictions. Model training is the most resource-intensive step and thereby the biggest potential bottleneck in ML. GPU databases provide the high-performance computing environment for model training by leveraging the massively parallel processing capabilities and unifying data with compute and model management to facilitate data exploration at any scale.
- The Internet of Things (IoT):** Live data can provide high value, but only if processed in real time. Without the processing capacity required to ingest and analyze live data streams, organizations risk missing opportunities by not being able to make key business decisions in a timely manner. For example, it's crucial for organizations to monitor customer information from various sources to monitor and analyze buying behavior in real time. IoT provides ample opportunities to harness actionable insights from connected devices and make these devices operate more effectively and intelligently. GPU databases provide the processing power and capacity needed to take full advantage of the IoT phenomenon. A GPU database can ingest, analyze, and take appropriate action on streaming data in real time.

**About ICF**

ICF (NASDAQ:ICFI) is a global consulting services company with over 5,000 specialized experts, but we are not your typical consultants. At ICF, business analysts and policy specialists work together with digital strategists, data scientists and creatives. We combine unmatched industry expertise with cutting-edge engagement capabilities to help organizations solve their most complex challenges. Since 1969, public and private sector clients have worked with ICF to navigate change and shape the future. Learn more at [icf.com](http://icf.com).

Any views or opinions expressed in this white paper are solely those of the author(s) and do not necessarily represent those of ICF. This white paper is provided for informational purposes only and the contents are subject to change without notice. No contractual obligations are formed directly or indirectly by this document. ICF MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

No part of this document may be reproduced or transmitted in any form, or by any means (electronic, mechanical, or otherwise), for any purpose without prior written permission.

ICF and ICF INTERNATIONAL are registered trademarks of ICF and/or its affiliates. Other names may be trademarks of their respective owners.

- **Geospatial analytics:** Large volumes of data available from mobile sources like vehicles and smartphones provide opportunities to derive actionable insights from geospatial aspects. For example, shipping and logistic companies can improve efficiency of carrier routes by analyzing the data streaming from mobile scanning devices. Leveraging GPU's roots in graphics processing, we can effectively run geospatial algorithms on large datasets in real time and provide results in a map/geospatial based graphics -- displayed instantly on ordinary web browsers. A GPU database also enables the ingestion, analysis, and results rendering on a single platform without the need to move data among different layers or technologies to achieve the expected outcome.

**About the Author**

**Janak Kalaria** is a Principal with ICF and has more than a decade of experience architecting and developing data driven analytical systems. He has led projects ranging from needs assessment and solution selection to full-scale implementation and deployment. He has directed projects involving information management, data science, cloud computing, big data analytics, and decision support systems for a number of U.S. federal agencies, the European Commission, and commercial organizations.

For more information, contact:

**Janak Kalaria**

[janak.kalaria@icf.com](mailto:janak.kalaria@icf.com) + 1.703.934.3934

 [facebook.com/ThisIsICF/](https://facebook.com/ThisIsICF/)

 [twitter.com/ICF](https://twitter.com/ICF)

 [youtube.com/icfinternational](https://youtube.com/icfinternational)

 [plus.google.com/+icfinternational](https://plus.google.com/+icfinternational)

 [linkedin.com/company/icf-international](https://linkedin.com/company/icf-international)

 [instagram.com/thisisicf/](https://instagram.com/thisisicf/)

