WHITE PAPER

GridGain

Improving Speed, Scalability and the Customer Experience with In-Memory Data Grids

A GridGain Systems In-Memory Computing White Paper

RE-IMAGINING SPEED AND SCALABILITY WITH IN-MEMORY COMPUTING

Improving the customer experience is arguably the biggest opportunity and threat for most companies. Amazon, eBay, Expedia, PayPal, Uber, and many other companies have proven customers want a more digital, personalized, and responsive customer experience. Their customers abandoned traditional companies with brick-and-mortar experiences to get it.

But transforming existing IT infrastructure to support digital business is hard. The adoption of new web, mobile, and other self-service channels; the addition of personalization and other automation; and the use of new types of data are pushing existing IT infrastructure and applications beyond their performance and scalability limits.

Over the last decade, the new anytime, anywhere, personalized experience has driven query and transaction volumes up 10 to 1000x. It has created 50x more data about customers, products, and interactions. It has also shrunk the response times customers expect from days or hours to seconds or less.

Many companies have focused on addressing performance and scalability challenges by scaling vertically in the short term with much more expensive hardware. But this approach is not cost effective or sustainable in the long run. System loads are growing faster than annual performance and cost improvements in hardware and show no signs of slowing down. More importantly, scaling existing systems is not enough. The customer experience must change as well.

Several companies have successfully delivered a more personalized, responsive and real-time experience. These innovators achieved this by performing what experts call hybrid transactional/analytical processing (HTAP) with in-memory computing. They performed analytics during their transactions and interactions with customers. They automated decisions to personalize the experience or cross-promote products. They proactively addressed a customer issue before it could impact a customer's satisfaction, purchasing decision, or loyalty.

These companies did not have to rip out and replace their existing applications and databases. Instead, they took a more evolutionary approach. First, they added speed and scalability by sliding an in-memory data grid (IMDG) in-between their existing applications and databases. This approach was more cost effective and provided a new in-memory layer for accessing and processing existing and new data. In-memory computing became the foundation that allowed them to combine transactions and services with real-time analytics and deliver new HTAP applications for their digital transformations and customer experience initiatives.

This white paper provides an overview of in-memory computing with a focus on IMDGs as the first foundational step in this journey. It discusses the advantages and uses of an IMDG and its role in digital transformation and improving the customer experience. It also introduces the GridGain® in-memory computing platform, and explains GridGain's IMDG and other capabilities that have helped companies add speed and scalability to their existing applications.

THE IN-MEMORY COMPUTING JOURNEY

Just about every company is going through a digital transformation. The simple reason is that customers demand a faster, more convenient customer experience anytime, anywhere on any device. And they will quickly change service providers to get it. An Aberdeen study found that companies with a strong omni-channel engagement retained 89% of their customers. Those with a weak strategy retained 33%. Digital transformation has spread beyond sales and services to all parts of a company. It has also helped improve core operations and supply chain efficiency by leveraging information from the Internet of Things (IoT), for example.

For most companies, delivering a better, more digital experience requires more real-time information about the customer, product, service or company. It also requires real-time analytics and actions through automation. Just consider some of the more common examples.

- Mobile applications have created 7x or greater interactions compared to non-mobile customers. Now customers expect an omni-channel experience that proactively engages, guides, makes recommendations or creates new personalized experiences like 1-click buying.
- IoT and all the related real-time information and context help deliver better connected products and customer experiences as applications respond to device data in real-time. Just think about Google Maps and Waze, TripAdvisor, mobile boarding passes, even basic alerting.
- Personalization during any web session requires much more historical and contextual data to be analyzed in real-time. Personalization is based on the customer's past and current behavior and on the behavior of others as a predictor of future intent.

- Service management across industries from retail sales and services to claims processing or clinical care management require gathering up-to-date information about the customer and their issues. Analytics or calculations help streamline or automate responses.
- Fraud detection requires analyzing the current behavior of each "user" and others in real-time and combining it with historical data and models to detect potential fraud on an individual or larger scale. Fraud detection increasingly uses streaming analytics or machine and deep learning to improve detection.
- Risk assessment, whether for trading, portfolio management, loans, insurance or online betting, is increasingly conducted in real-time to improve business decisions and outcomes.
- Compliance requires real-time calculations of overall financial risk for regulations like Basel III that emerged following the last financial crisis, or real-time auditing and analytics of data usage for regulations like PCI during transactions.

Implementing a digital transformation creates three major challenges for IT:

1. Delivering real-time performance and unlimited scalability for existing applications
2. Delivering a better customer experience and business outcome
3. Adopting HTAP and other new technologies over time

Delivering Real-time Performance and Unlimited Scalability

The first challenge is to handle an unprecedented growth in queries and transactions over the last decade AND deliver "real-time" end-to-end responsiveness of one second or less:

- 10-1000x growth in query and transaction volumes: No one expected this level of growth a decade ago and scaling vertically has become more expensive or impossible. The growth rate is simply too much for a single server to handle in the long run.
- 50x growth in data: The good news is that it's become more cost-effective to collect all kinds of "Big Data" about customers and operations. The bad news is that this has resulted in a 50x growth in managed data over the last decade and put a huge load on existing infrastructure.

Over time IT has added more layers of systems and middleware that only increase latency and weren't designed to handle these volumes.

Delivering a Better Customer Experience and Business Outcome

The second challenge is to improve the customer experience without requiring a big rewrite. Improving the customer experience and becoming more real-time responsive is the big opportunity for most companies. For example in retail improving the customer experience can deliver more benefit for the company than any other initiative. Retailers on average have 25% annual customer churn. A loyal customer can be twice as profitable as a normal customer. It means that even a 20% reduction in churn can double profitability over 5 to 10 years. The benefits to operations are also huge. For example, predictive maintenance, based on real-time analytics and machine learning, increased revenues by 10% for one oil company.

Improving the experience is also a big challenge. Customers want a convenient and personalized experience. They also want a better way to resolve their issues when they happen. Roughly two-thirds of all customer churn is because of poor service, not a product issue. One study found that only 4% of all customers who have a bad experience give feedback. 91% just leave without saying anything. Yet if you catch their problem and fix it in time, 70% of dissatisfied customers stay.

Improving the experience not only requires adding new channels that not only personalize and streamline the process. It also requires collecting new data and using it to detect and respond to issues as they happen, in real-time.

Adopting HTAP and Other New Technologies Over Time

Solving this problem requires a major shift in a company's focus on the customer. But it also requires adopting new technology that can collect a lot more information, and then perform analytics and decision automation during each transaction or interaction. Gartner and others call this approach hybrid transactional/analytical processing (HTAP). Whatever technology is used for HTAP, it must be able to handle the 10-1000x increase in interactions and 50x increase in data. It must also be able to perform analytics and decision automation during each transaction or interaction in real-time.

THE SOLUTION: IN-MEMORY COMPUTING

The technology many companies have chosen to improve the speed and scalability of existing and new applications is in-memory computing. These companies realized that there needed to be a new data layer in between existing appli-

cations and databases. It had to add speed and scale to existing applications. It also had to allow companies to incrementally change existing applications over time and rapidly deliver new real-time analytics and HTAP applications.

The speed comes from storing and processing data in memory rather than continually retrieving data from disk(s) before processing. While hard drive (HDD) media speeds are measured in milliseconds, RAM speeds can be measured in nanoseconds—a million times faster. The scalability comes from distributing data and computing together horizontally across a cluster of servers, or nodes, with a "shared nothing" architecture.

This means that the data is distributed in such a way that data can be processed locally on each node without having to move a lot of data over the network. The combination of a shared-nothing architecture with collocated processing is what helps achieve horizontal, linear scalability and a lower cost of ownership than other approaches.

These companies have been able to adopt in-memory computing, and in particular in-memory data grids (IMDG), to add speed and scalability to their existing applications with minimal changes to their architecture. It usually paid for itself because it removed the need to scale up existing databases by buying expensive new hardware and additional database licenses.

Even better, this approach gave them an evolutionary, lower risk path to implementing HTAP with in-memory computing over time. Their IMDG gave them a new data layer to run real-time analytics, machine and deep learning against data in memory without impacting the performance of the existing applications and databases. With each new application they added to the IMDG, they were able to access more and more of their data in memory. This gave them the ability to grow their use of HTAP to improve the customer experience and business outcomes.

Adding Speed, Scalability and Analytics at Wellington Management

Wellington Management is one of the top 20 global asset management firms in the world, with more than $1 trillion in client assets under management.

Wellington had three major challenges:

1. Its current systems were no longer scalable due to an exploding growth of financial data. It needed horizontal scalability to handle the long-term growth.
2. The 2008 financial crisis resulted in a wave of new financial regulations that resulted in more complexity and risk in existing systems.
3. Many more new and complex asset classes have been introduced in the last few years based on customer demand, and there's a big need to release new asset classes faster

Wellington's solution was to deploy its investment book of record (IBOR) on the GridGain in-memory computing platform. The Wellington IBOR serves as the single source of truth for investor positions, exposure, valuations, and performance. All trading transactions and account and back office activity flow through the IBOR in real time.

- **Horizontal Speed and Scalability**: Wellington's IBOR has unlimited horizontal scalability. It uses GridGain's SQL support to add speed and scalability by sliding in-between Oracle, its system of record, and the applications. The result is at least 10 times faster performance by adding in-memory computing on top of its Oracle database deployment.
- **Use of HTAP**: The IBOR is an HTAP system that is used by portfolio management teams for real-time position, market value, exposure, and performance analytics; by risk management teams for risk analytics and overall risk management; and by compliance teams to ensure, in real-time, that all regulatory requirements are met.

Why Wellington chose GridGain

- In-memory computing
- Horizontally scalable
- Supports distributed SQL
- ACID compliant (consistent data)
- Collocates data and computing
- Combines operational and analytical workflows (HTAP)

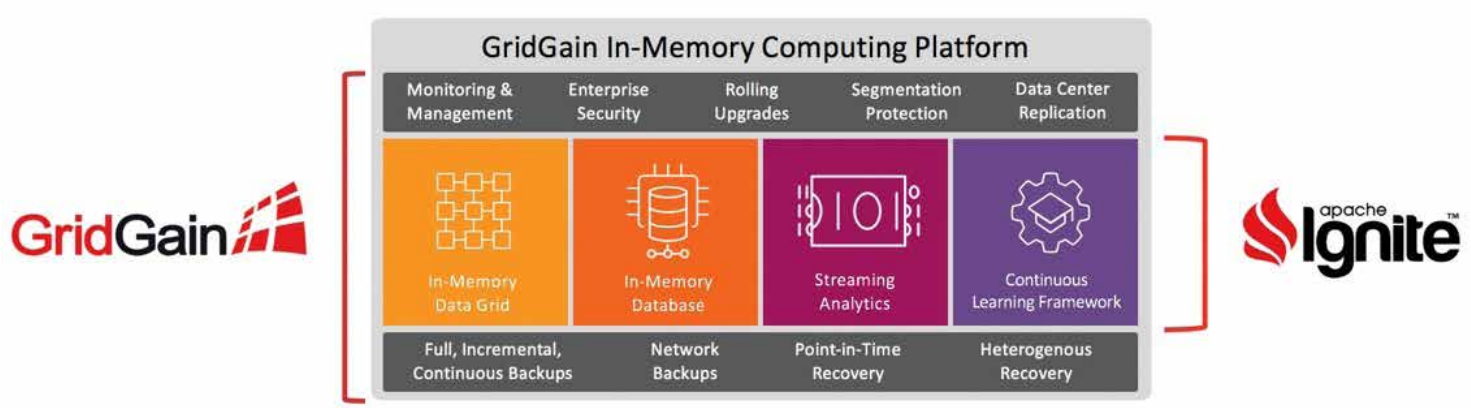

Figure 1. Apache Ignite and the GridGain In-Memory Computing Platform

APACHE IGNITE AND THE GRIDGAIN IN-MEMORY COMPUTING PLATFORM

GridGain is the leading in-memory computing platform for real-time business. It is the only enterprise-grade, commercially supported version of the Apache® Ignite™ (Ignite) open source project. GridGain includes enterprise-grade security, deployment, management, and monitoring capabilities which are not in Ignite, plus global support and services for business-critical systems. GridGain Systems contributed the code that became Ignite to the Apache Software Foundation and continues to be the project's lead contributor.

GridGain and Ignite are used by tens of thousands of companies worldwide to add in-memory speed and unlimited horizontal scalability to existing applications, and then add HTAP to support new initiatives to improve the customer experience and business outcomes. With GridGain, companies have:

- Improved speed and scalability by sliding GridGain in-between existing applications and databases as an IMDG with no rip-and-replace of the applications or databases
- Improved transactional throughput and data ingestion by leveraging GridGain as a distributed IMDB
- Improved the customer experience or business outcomes by adding HTAP that leverages real-time analytics, streaming analytics and continuous learning

GridGain customers have been able to create a new shared in-memory data foundation. This single system of record for transactions and analytics enables real-time visibility and action for their business. With each project, they have unlocked more information for use by other applications on a platform with real-time performance at peak loads and always-on availability. As a result, they can develop new projects faster, are more flexible to change, and are more responsive in ways that have improved their experiences and business outcomes.

ADDING SPEED AND SCALABILITY TO EXISTING APPLICATIONS WITH AN IMDG

One of the core GridGain capabilities and most common use cases is as an IMDG. GridGain can increase the performance and scalability of existing applications and databases by sliding in-between the application and data layer with no rip-and-replace of the database or application and without major architectural changes.

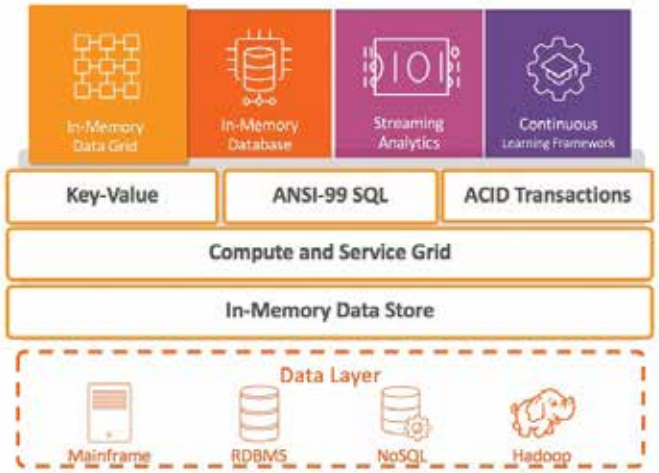

Figure 2. GridGain as an In-Memory Data Grid (IMDG)

This is because GridGain supports ANSI-99 SQL and ACID transactions. GridGain can sit on top of leading RDBMSs including IBM DB2®, Microsoft SQL Server®, MySQL®, Oracle® and Postgres® as well as NoSQL databases such as Apache Cassandra™ and MongoDB®. GridGain generates the application domain model based on the schema definition of the underlying database, loads the data, and then acts as the new data platform for the application. GridGain handles all reads and coordinates transactions with the underlying database in a way that ensures data consistency in the database and GridGain. By utilizing RAM in place of a disk-based database, GridGain lowers latency by orders of magnitude compared to traditional disk-based databases.

ADDING REAL-TIME ANALYTICS AND HTAP WITH MASSIVELY PARALLEL PROCESSING (MPP)

Once GridGain is put in place, all of the data stored in existing databases or in GridGain is now available in memory for any other use. Additional workloads are easily supported by GridGain with unlimited linear horizontal scalability for real-time analytics and HTAP.

GridGain accomplishes this by implementing a general purpose in-memory compute grid for massively parallel processing (MPP). GridGain optimizes overall performance by distributing data across a cluster of nodes, and acting as a compute grid that sends the processing to the data. This collocates data and processing across the cluster. Collocation enables parallel, in-memory processing of CPU-intensive or other resource-intensive tasks without having to fetch data over the network.

The GridGain Compute Grid is a general-purpose framework that developers can use to add their own computations for any combination of transactions, analytics, stream processing, or machine learning. Companies have used GridGain's MPP capabilities for traditional High-Performance Computing (HPC) applications as well as a host of real-time HTAP applications.

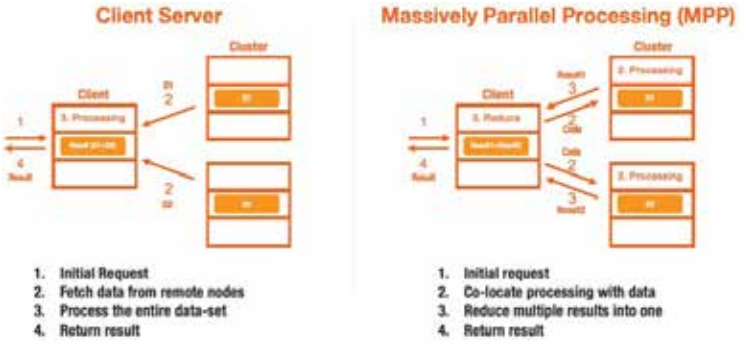

Figure 3. GridGain Compute Grid – Client Server vs. Collocated Processing (MPP)

GridGain has implemented all of its built-in computing on the GridGain Compute Grid, including GridGain distributed SQL as well as the GridGain Continuous Learning Framework for machine and deep learning. Developers can write their own real-time analytics or processing in multiple languages, including Java, .NET and C++, and then deploy their code using the Compute Grid.

Collocation is driven by user-defined data affinity, such as declaring foreign keys in SQL DDL (data definition language) when defining schema. Collocation helps ensure all data needed for processing data on each node is stored locally either as the data master or copy. This helps eliminate the network as a bottleneck by removing the need to move large data sets over the network to applications or analytics.

ADDING DEEPER INSIGHTS AND AUTOMATION WITH STREAMING ANALYTICS AND CONTINUOUS LEARNING

The capabilities GridGain supports are not just limited to real-time analytics that support transactions. GridGain is also used by the largest companies in the world to improve the customer experiences or business outcomes using streaming analytics and machine and deep learning. These companies have been able to incrementally adopt these technologies using GridGain to ingest, process, store and publish streaming data for large-scale, mission critical business applications.

GridGain is used by several of the largest banks in the world for trade processing, settlement and compliance. Telecommunications companies use it to deliver call services over telephone networks and the Internet. Retail and e-commerce vendors rely on it to deliver an improved real-time experience. And leading cloud infrastructure and SaaS vendors use it as the in-memory computing foundation of their offerings. Companies have been able to ingest and process streams with millions of events per second on a moderately-sized cluster.

GridGain is integrated and used with major streaming technologies including Apache Camel™, Kafka™, Spark™ and Storm™, Java Message Service (JMS) and MQTT to ingest, process and publish streaming data. Once loaded into the cluster, companies can leverage GridGain's built-in MPP-style libraries for concurrent data processing, including concurrent SQL queries and continuous learning. Clients can then subscribe to continuous queries which execute and identify important events as streams are processed.

Figure 5. GridGain for Stream Ingestion, Processing and Analytics

GridGain also provides the broadest in-memory computing integration with Apache Spark. The integration includes native support for Spark DataFrames, a GridGain RDD API for reading in and writing data to GridGain as mutable Spark RDDs, optimized SQL, and an in-memory implementation of HDFS with the GridGain File System (GGFS). The integration allows Spark to:

- Access all the in-memory data in GridGain, not just data streams
- Share data and state across all Spark jobs
- Take advantage of all GridGain's in-memory processing including continuous learning to train models in near real-time to improve outcomes for in-process HTAP applications

GridGain also provides the GridGain Continuous Learning Framework. It enables companies to automate decisions by adding machine and deep learning with real-time performance on petabytes of data. GridGain accomplishes this by running machine and deep learning in RAM and in place on each machine without having to move data over the network.

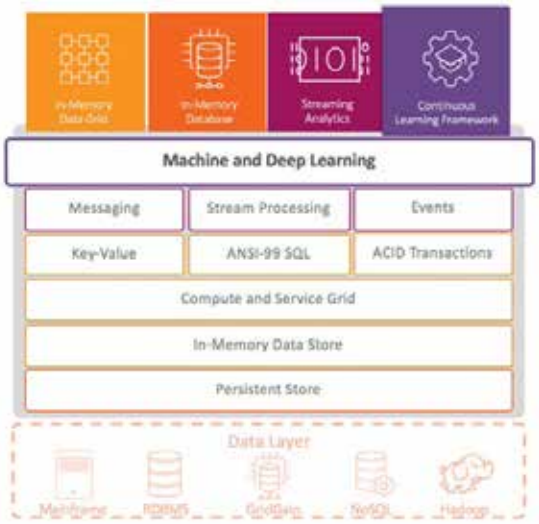

Figure 6. GridGain for Machine and Deep Learning

GridGain provides several standard machine learning algorithms optimized for MPP-style processing including linear and multi-linear regression, k-means clustering, decision trees, k-NN classification and regression. It also includes a multilayer perceptron and TensorFlow integration for deep learning. Developers can develop and deploy their own algorithms across any cluster as well as using the compute grid. The result is continuous learning that can be incrementally retrained at any time against the latest data to improve every decision and outcome.

SUMMARY

Applications and their underlying RDBMSs have been pushed beyond their architectural limits by new business needs, and new software layers. Companies must add speed, scale, agility, and new capabilities to support digital transformation and other business critical initiatives. There are many Oracle options for adding speed and scale to Oracle Database, or for replacing it—including Oracle RAC, Oracle Database In-Memory, Oracle Exadata, Oracle GoldenGate, Oracle TimesTen Classic, Oracle TimesTen Scaleout, and Oracle Coherence—and each has its place. But when the speed and scale really needs to be addressed beyond database layer, the best long term approach is in-memory computing. Not only does it add speed and scale. It unlocks data, enabling companies to be much more agile.

Contact GridGain Systems

To learn more about how GridGain can help your business, please email our sales team at sales@gridgain.com,call us at +1 (650) 241-2281 (US) or +44 (0)208 610 0666 (Europe), or complete our contact form at www.gridgain.com/contact.

About GridGain Systems

GridGain Systems is revolutionizing real-time data access and processing with the GridGain in-memory computing platform built on Apache® Ignite™. GridGain and Apache Ignite are used by tens of thousands of global enterprises in financial services, fintech, software, e-commerce, retail, online business services, healthcare, telecom and other major sectors, with a client list that includes ING, Raymond James, American Express, Societe Generale, Finastra, IHS Markit, ServiceNow, Marketo, RingCentral, American Airlines, Agilent, and UnitedHealthcare. GridGain delivers unprecedented speed and massive scalability to both legacy and greenfield applications. Deployed on a distributed cluster of commodity servers, GridGain software can reside between the application and data layers (RDBMS, NoSQL and Apache® Hadoop®), requiring no rip-and-replace of the existing databases, or it can be deployed as an in-memory transactional SQL database. GridGain is the most comprehensive in-memory computing platform for high-volume ACID transactions, real-time analytics, web-scale applications, continuous learning and hybrid transactional/analytical processing (HTAP). For more information on GridGain products and services, visit www.gridgain.com.