WHITE PAPER

GridGain

A Roadmap for In-Memory Computing: Current and Emerging Use Cases

A GridGain Systems In-Memory Computing White Paper

As businesses everywhere struggle to keep up with the ever-increasing need for speed and scale—and strategically leverage the massive amounts of data driving their organizations—more and more of them are turning to in-memory computing (IMC). IMC takes combines important lessons from caching, Hadoop and high-performance computing (HPC). It moves data into RAM for speed and distributes data and computing across clusters of low-cost computers and cloud infrastructure for scale. This high-performance architecture is rapidly becoming the method of choice for today's real-time applications that are focused on big data, fast data, streaming analytics or machine and deep learning.

The in-memory computing revolution is far from over, as demand continues to spur further innovation in in-memory computing technologies. This white paper examines:

- Factors contributing to increased demand for IMC
- Common IMC use cases today
- Emerging use cases on the horizon
- How Apache® Ignite™ and GridGain® address these use cases

WHY DEMAND FOR IN-MEMORY COMPUTING (IMC) KEEPS GROWING

An ever-increasing need to deliver better real-time experiences to consumers and track and respond to issues faster across business processes drives the move towards IMC. The success of Amazon, eBay, Expedia, PayPal, Uber, and many other companies proves that customers want a more digital, personalized, and responsive real-time customer experience. Customers abandoned traditional companies with slower brick-and-mortar experiences to get it. Many other companies have demonstrated how real-time analytics and automation improve core business operations. Companies have adopted IMC to add the speed, scale and computing that existing IT infrastructure cannot deliver in support of these new initiatives.

Heightened performance expectations

A recent Computer Business Review Magazine survey of 200 UK-based CEOs found that ninety percent of the CEOs believed there is a significant gap between the performance level that consumers expect and what IT can deliver. These results echo similar concerns among CEOs in the US.

Consumers' growing expectations for real-time performance are driving business expectations similarly upward. One-click shopping has become the golden standard in online retail. Consumers get very frustrated by processes that take too long, as well as applications with poor performance. As users become accustomed to real-time response in consumer apps such as social media and online gaming, they expect the business technologies they use within enterprise and public-sector organizations to deliver the same level of performance. They expect real-time, immediate response—a level once available only for mission-critical applications. All too often, businesses do not deliver this expectation, which leads to customer attrition. If a web page fails to load in three or four seconds, they will abandon a shopping cart and go to a different website.

Many organizations turn to IMC technology as standalone technology or as options embedded in their existing data platforms to address the gap between performance and customer expectations. The best way to improve application performance is putting the data in RAM, as close to the consuming application as possible. In the last few years, operational databases (which started off mostly as disk-persisted) have increasingly added in-memory options to speed up transaction processing and support analytical insight. Platforms for event and stream processing are using in-memory technology to rapidly ingest, analyze, and filter data on the fly. Many analytic databases and data warehouses rely on in-memory technology to quickly run complicated queries on large data sets. In addition, some companies use in-memory technology to sit on top of Hadoop (which is more batch-oriented than interactive) so that they can more rapidly access data from their data lakes (made up of Hadoop clusters).

Need for scale

The performance gap is not just about speed. It is about speed and scale together. Over the last decade, the new model of an anytime, anywhere, personalized experience requires sub-second response times every time. It also requires support for a huge increase in interactions as more people move online and do more searches on products and services before buying. The average query and transaction volumes for some companies have grown up to 1000 times over the last decade.

IMC technology doesn't just help with speed. It addresses both speed and scale together. The best IMC technologies move data into RAM for speed, and cluster low-cost hardware together for horizontal scale using an approach like Hadoop. Like Hadoop, IMC technologies also move the code and processing to the data without requiring large data sets to be moved over a network. This is how IMC achieves linear horizontal scalability for a host of different workloads, and new real-time uses of data.

Digital business, IoT, and other initiatives require support for new types of computing. IMC benefits include the real-time analytics needed during transactions and interactions. Some examples are:

- Personalization of each Web page based on preferences and browsing history
- Cross-promotion of other products such as the recommended products you see on Amazon while shopping
- Ad re-targeting that follows you to other sites following an abandoned purchase

Gartner and others call this approach hybrid transactional/analytical processing (HTAP). IMC also supports real-time stream processing, analytics and automation, which is used for a host of IoT functions—from autonomous driving to the preventative maintenance of complex machinery. IMC is also well-suited for machine and deep learning, in part because horizontal scaling on commodity infrastructure provides a more cost-effective alternative to dealing with massive data sets than traditional database technology.

Explosive data growth

The increase in scale is not just about the volume of queries or transactions. The amount of managed data has gown 50 times over the last decade, and it is not expected to slow down anytime soon. Transactional data, which used to be 80 percent of all data managed, has been dwarfed by the growth in information about customers, products and interactions from Web sites, third-party social networks, mobile devices and the Internet of Things (IoT). This data comes from a host of new sources, including:

- **Web traffic**. There is much more data than the basic analytics from Google Analytics. Capturing, tracking, even replaying every customer interaction on a Web site has been possible for nearly two decades now. This information has been used to segment customers, identify top usability issues, even retarget customers after they abandoned their shopping carts. But it can easily become petabytes of information if stored for months or years.
- **Social data**. Facebook, LinkedIn, Twitter, Instagram, etc. hold a treasure trove of information about personal preferences and relationships. This information is extremely valuable to businesses. In financial services, for example, those relationships help identify cross-selling opportunities for bank accounts, college savings funds, and life insurance.
- **Mobile data**. Current mobile data traffic is expected to grow at a compound annual growth rate of 39 percent and reach over 100 exabytes per month by 2023 (with three-quarters of the traffic being video-related). Consumers activity, location, and preferences information is invaluable to companies or government agencies trying to sell to or provide better services to customers or citizens.
- **IoT**. In 2019, the number of IoT devices is expected to exceed eight billion, more than the number of people worldwide. If you count the number of smartphone and other "non-IoT" connected devices, the number is approaching 20 billion. The generated amount of information and the uses it can be put to is staggering, especially for more complex devices. The average connected car can generate four TB of data per hour of driving. When the anti-lock brake system is deployed at low temperatures, for example, Volvo connected cars send data to the Swedish Transport Authority and alerts authorities to frozen roads that need attention. Current Formula One cars have about 300 sensors. A train manufacturer seeking to improve braking efficiency employs about 400 sensors per train—a number likely to increase to 4,000 over the next five years. With hundreds or thousands of sensors in a vehicle taking readings every millisecond, the output quickly adds up to a massive volume of data.

Ingesting and analyzing all the data from apps and devices is already overwhelming. The data industry is racing to keep up with these data volumes, even as it must deal with another daunting challenge: processing the data quickly enough to keep up with today's expectations of real-time response.

Better, cheaper, faster IMC technology

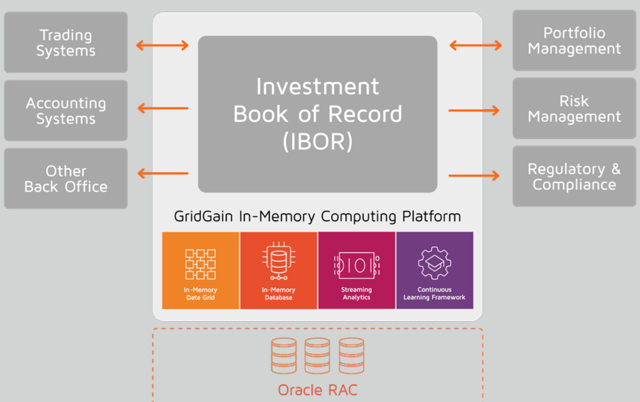
With the ever-increasing need for speed and scale that can support exploding data volumes for both existing and new types of real-time computing, companies have turned to IMC technologies. IMC has been attractive in part because of its use of more modern processing built on horizontally scalability commodity infrastructure. This adoption has led to a virtuous cycle of continued growth, adoption and innovation. Further adoption has in turn fueled the use of IMC for new use cases. This leads to new entrants into the IMC market, which in turn furthers more innovation around IMC technologies and wider adoption within companies. The innovation has also helped make IMC technologies even more affordable.

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 is 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)

CURRENT USE CASES OF IMC TECHNOLOGIES

Many companies started by adopting in-memory data grid (IMDG) technology. IMDGs—and more comprehensive in-memory computing platforms that include IMDGs, such as Apache® Ignite™ and its enterprise version, GridGain—are one of the fastest-growing technology segments along with Hadoop.

Adding speed and scale to existing applications. The first, most common use case for IMDGs has been to improve the speed and scalability of existing applications. These companies realized a need for a new data layer in between existing applications and existing databases. This layer 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.

IMDGs provide a convenient way to speed up query performance without having to migrate or extensively rework the underlying database. They accomplish this feat by sliding in between existing applications and databases and adding in-memory speed and horizontal scale on top of existing databases that are mostly limited to vertical scalability. The IMDGs act as a read-and-write-through in-memory caching and computing layer. The data is distributed in such a way that it 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 for scaling existing applications. This is especially true for IMDGs. IMDGs can be added with minimal changes to existing applications because they support existing ways of connecting to applications and databases such as SQL. Adding an IMDG often pays for itself because it removes the need to buy expensive new hardware and additional database licenses to scale databases vertically. Companies who implement IMC technologies such as Ignite or GridGain find that they can process queries up to 1,000 times faster than before, and meet any required query scalability needs.

Even better, this approach gave early adopters 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 allows growth in HTAP use to improve their customers' experience and business outcomes.

Building real-time analytics and HTAP. Another common project for IMC technologies is real-time analytics. The challenge for analytics in many ETL architectures is the delay in performing the Extracts, then the Transforms to cleanse and merge data, and then the final Loads into a data warehouse. Even with a more real-time approach, such as Change Data Capture (CDC) that incrementally captures changes from the write-ahead logs of databases, there was usually a significant lag between the operational data and the data warehouse that results in out-of-date visibility into the current state of the business. Real-time solutions such as Teradata, SAP Hana, or Oracle Exalytics, which all accelerated ETL with hardware, could end up costing hundreds of thousands of dollars per terabyte of data and hitting scalability limits by the time they approached a petabyte in size (depending on the deployment).

Rather than continue with separate online transactional processing (OLTP) and online analytical processing (OLAP), companies have begun to implement both together as HTAP with IMC. Once an IMDG is used to add speed and scale to applications, for example, more nodes are added in the same cluster to handle other workloads such as for analytics. Data across applications are merged in real-time for specific analytic needs. As operational data is updated within the applications and the IMDG, it is automatically updated in the other nodes for analytics.

There are also cases where the analytics is separate, such as for high-performance computing (HPC) tasks in financial services, or to accelerate certain types of reporting. In this case, the data is bulk-loaded or streamed into a cluster and compute tasks performed either in bulk or incrementally during stream processing. There are two main advantages to this architecture. First, the data can be ingested horizontally across nodes for scale. Second, the data processing can be architected to minimize network traffic during the analytics for very large data sets that cannot be moved over the network fast enough.

At some point, holding this much data in RAM is either too expensive or inefficient, especially when only a subset of the data is used for analytics. Some IMC platforms add in-memory database (IMDB) capabilities for holding and persisting new types of data or existing data sets and transactional throughput rates that are too large for traditional scale-up databases. Nonvolatile memory is used to store the entire dataset and any indexes, while RAM is used more as a cache to hold the most frequently used subset. This type of architecture is very well suited for data sets where only a small subset of the data, such as the current operational data, is frequently used. The IMDB optimizes different types of computing by running different computing tasks on separate nodes with different data sets as part of the same cluster.

Delivering new APIs and applications. The other most common project for IMC technologies is API management. In many ways, API management is related to adding speed and scale to existing applications, because digital business initiatives and self-service use cases are two of the most common reason existing applications become overloaded.

Increasing Innovation at a Leading Global Bank

A leading global bank was struggling to increase their pace of innovation, both to service changing customer needs and to move beyond traditional banking with new services and business models. Part of the challenge was that core banking services resided on a mainframe that was hard to change. They were also unable to continue to scale their infrastructure and handle increasing loads. Mobile traffic was growing 25 percent per year, and each mobile user interacted seven times more than a Web user.

Using GridGain, the bank added an in-memory computing layer in front of their mainframe and end-to-end middleware that helped open up and share previously siloed data and functionality across channels. GridGain not only easily scaled to offload from the mainframe and handle the growth. It helped cut end-to-end latency to less than 100ms. It also helped the bank rapidly create new APIs for consumers and third parties and be first to market with new services for PSD2, SEPA and instant payments.

APIs are at the heart of the challenges with both speed and scale because they are the main access point for digital business initiatives. Ideally there is a single, common API used to deliver the same functionality across different channels and a seamless omnichannel experience. The challenge with each API is that it needs to be part of a sub-second response time, every time. If there is an application accessing the API, and perhaps a mobile application as well, then the end-to-end latency of the API might be in the low hundreds of milliseconds. It is usually impossible for an API to respond this quickly without IMC technologies. There are usually several layers of APIs that access middleware (that in turn access several applications) that then access their underlying databases. Then the results must come back over the network, and often be merged and processed even more before a response is returned to the end user application.

The best solution is to bring the data to the API. Many companies have implemented an IMDG where each API has a node collocated with each API to bring all the data needed by the API into local memory for speed. They often bundle the API and node together in a Docker container managed using Kubernetes. This architecture provides elastic scalability. IMC technologies provide a host of options for lowering transaction latency as well.

EMERGING USE CASES FOR IN-MEMORY COMPUTING

Adding speed and scale to existing applications, real-time analytics and APIs are common use cases across the thousands of companies using IMC technologies today. But other use cases have already emerged as companies leverage the data in memory for new uses or leverage their IMC technology for new projects.

Today's technology supports all these use cases and is proven to work. But several of these use cases are still evolving, and the technology is changing. Based on more recent usage, you can expect to see IMC technologies used broadly for the following:

- First class support for in-memory SQL
- IMC as a system of record
- Hybrid storage models for very large datasets
- IMC for artificial intelligence (AI)
- Broad adoption of on-volatile memory (NVM) for IMC

First class support for in-memory SQL. About ten years ago, many in the data-processing community started to move towards NoSQL and a host of other new query languages for different use cases. In recent years, however, the community has recognized the value of SQL and embraced it once again. Not only is SQL the most popular data query language, it is the second-most popular developer language after JavaScript. It is also hardened and mature, with several advantages for searching and querying data.

The ability to index data in SQL provides a significant advantage with respect to searching. When there is no key to the data, without SQL or indexes, searching for a specific piece of data requires a full scan—which is significantly slower. Also, as use cases of in-memory computing become more and more complex, users need robust querying capabilities such as complex aggregation functions and joins. No one wants to struggle through a learning curve or complicated setups to create queries in a proprietary language when all that is needed is one line of SQL code.

For these reasons, first class SQL support (in addition to NoSQL, streaming, and other data-access mechanisms) is a necessity for all serious IMC vendors and projects. Apache Ignite and GridGain have a very strong level of support through the ability to perform distributed processing of ANSI SQL-99 compatible SQL across data stored in memory and on disk. Other popular projects, including Spark and Hadoop, have led to the growth of a whole cottage industry of SQL support. IMC vendors who do not yet offer SQL support are likely to get on board soon, to better handle the new, more complex use cases that are becoming prevalent.

IMC as a system of record. The increased affordability and capabilities of in-memory technology, as mentioned earlier, have made memory-centric architectures the dominant paradigm for today's and tomorrow's data systems. Many businesses (including over 50 percent of GridGain's clients) use in-memory computing platforms as their systems of record and the authoritative data source for business-critical records, as they would in a database. Those who implemented IMC technologies for new applications, APIs, and analytics effectively implemented it to provide a new real-time system of record. Existing databases are used primarily as backup devices once IMC technologies are deployed.

Contributing to this development is the ability of more and more in-memory data platforms to perform hybrid transactional and analytical processing (HTAP). These platforms, which include Apache Ignite and GridGain, can handle operations and transactions as well as analytics—so there is no need for a business to have separate databases for operational and analytical purposes.

As more and more IMC platforms adopt SQL, it will be yet another reason they will be used as an IMDB and a main real-time system of record.

Hybrid storage models for very large datasets. Once a far-off goal, the idea of storing large databases entirely in memory has become increasingly attainable, but what are the practical limitations? How large can such databases be? Can an in-memory solution support petabytes of data?

Three or four years ago, such a scenario would have seemed out of reach, due to the cost of RAM and the impractical cluster size that would be needed (dozens of computers per cluster). However, technology keeps evolving. With products such as the Fujitsu M10 server (with 64 terabytes of RAM), or cloud instances from Amazon, Google, and Microsoft (with terabytes of RAM), even petabyte IMC clusters are attainable.

Of course, a petabyte cluster would still be millions of dollars. While they do already exist, and will eventually be needed by most companies, petabyte clusters are more than what most companies need right now. They will need generate a positive return on investment (ROI).

The Apache Ignite development community (including GridGain) developed a hybrid storage model with a universal interface to all storage media (RAM, flash, or disk, plus NVM in the future). Users do not need to know details of where their data is stored. They can use the same full, unified API (including both SQL and key-value access) to address the data. Apache Ignite or GridGain moves data to the fastest layer available, as needed, while keeping the complete data set persisted in non-volatile storage.

This hybrid storage model gives businesses the flexibility to easily adjust storage strategy and capacity—as well as data-processing performance—without changing data-access mechanisms. To increase the speed of systems, users simply add RAM to their clusters. To add capacity at a lower cost, they can add flash memory or disk space.

While this hybrid storage model may sound simple, developing it took years due to the architectural differences between RAM, flash, and disk. Most of the algorithms within Apache Ignite and GridGain have been optimized for RAM, which is byte-addressable, whereas flash and disk are block-addressable. To read one byte of data from disk or flash, the entire block of data must be read. This requires extra time and leads to a large difference in latency compared to RAM. Accessing a byte of data in RAM takes nanoseconds, compared to milliseconds with flash and disk. Reconciling the latency differences among these storage media was a significant challenge—which explains why very few in-memory computing solutions have successfully addressed this problem, apart from Apache Ignite and GridGain.

Businesses looking toward petabytes of data in the future can prepare for that future by choosing an in-memory solution with a hybrid storage model. The hybrid storage feature in Apache Ignite and GridGain is available now, and developers of other in-memory solutions are working on similar features because they know that hybrid storage is the logical next wave for in-memory computing.

IMC for artificial intelligence (AI). Among the exciting new use cases for in-memory computing are the growing number of artificial intelligence projects involving machine learning (ML)—that is, computers learning from data, rather than through programmed instructions—and deep learning (DL), a type of ML that uses multi-layered neural networks. While ML has been around in some form for over 50 years, only recently has the technology to apply it at high speed to very large data sets become relatively accessible.

Machine learning on small, dense data sets—typically data sets that fit within the storage of a single computer—is essentially a solved problem. Libraries in the Python, Java, and Scala community address this problem. Projects such as Spark solve it very efficiently within one computer's memory. However, applying the statistical algorithms of machine learning and deep learning on large, sparse data sets is a much more complex challenge.

Machine learning and deep learning on large, sparse data sets requires a data management system that can store terabytes or petabytes of data and perform fast parallel computations. These are tasks for which in-memory computing platforms such as Apache Ignite and GridGain are ideally suited. By storing data in RAM (which is much faster than disk storage) across clusters of computers, and by sending computations directly to the data for parallel processing, these IMC platforms can process the data with the maximum speed and performance allowed by the hardware.

Apache Ignite and GridGain provide additional support for in-memory machine learning and deep learning, and are already used by early adopters. Beyond a host of pre-built algorithms, companies can also write their own algorithms, or use TensorFlow. Ignite and GridGain push the algorithms to the data and perform model training in place against any data set which might otherwise take hours to move over a network, and run the algorithms with linear horizontal scale. In-memory support for AI is crucial for applying ML and DL algorithms to tens of terabytes of streaming data in the many future continuous learning use cases.

GridGain

Broad adoption of on-volatile memory (NVM) for IMC. As in-memory technology continues to evolve, the most important new development is the emergence of NVM, which will have an enormous impact on data storage over the next decade.

The idea of NVM is very simple. DRAM (Dynamic Random Access Memory), the type of RAM currently used in computer systems, loses all the in-memory data if the power goes out. NVM fixes this problem, retaining its data during a power loss. This development removes the need for similar fault-tolerance on a software level.

Early versions of non-volatile memory were a simple and clunky combination of normal DRAM, flash memory, and a battery to power the copying of compressed data from RAM into flash when the system lost power. Today, we have much more advanced technology in the form of 3D XPoint such as Intel Optane. 3D XPoint features a transistor-free, byte-addressable architecture based on a stackable, cross-gridded data-access array that its makers liken to a three-dimensional checkerboard. This technology can be used both for flash or flash-like devices and for RAM that inserts into the deep sockets in a computer's motherboard.

Industry experts predict increasing adoption of the NVM technology demonstrated in 3D XPoint over the next three to five years. Intel Optane has already been released to accelerate SSD performance, and Intel Optane DC Persistent Memory, which plugs directly into DIMM memory slots, is becoming generally available in 2019. A decade from now, NVM such as Intel Optane will probably be the prevalent model of storing data on computer systems for performance. Disk and flash systems might be relegated to legacy systems and backup scenarios.

As the necessary infrastructure emerges, the adoption of non-volatile memory will bring fundamental changes to in-memory computing. No longer will in-memory computing systems require disk-based storage for fault tolerance. Instead, with NVM retaining data through power outages, IMC systems can be 100 percent in-memory, entirely composed of memory-based data storage.

To adapt to this major shift in data strategies, forward-looking providers of IMC platforms are making NVM support a high priority. GridGain currently supports NVM, including Intel Optane, within a hybrid storage model that can easily scale up and evolve with the changing storage paradigm.

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.

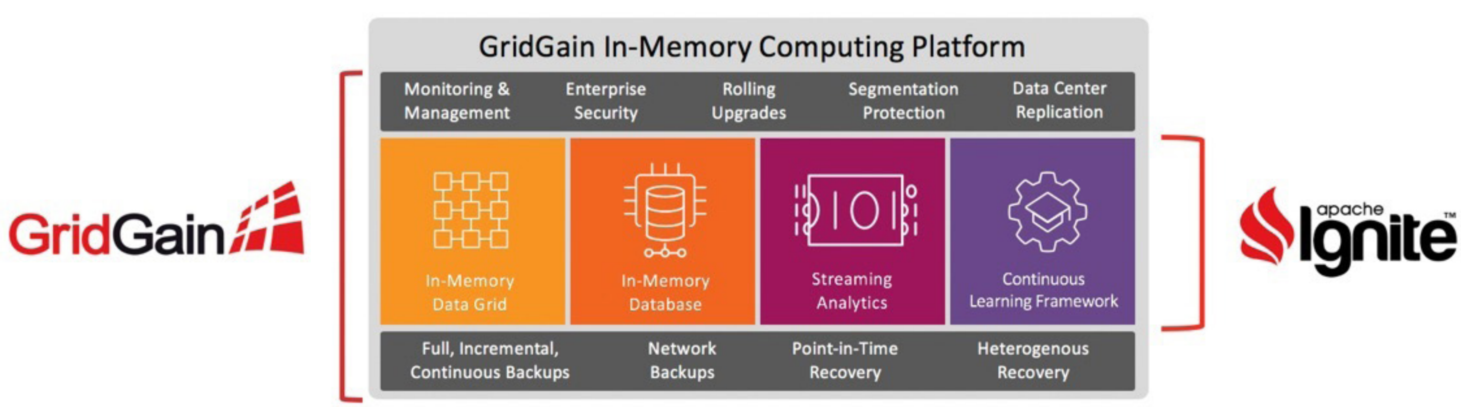


Figure 1. Apache Ignite and the GridGain In-Memory Computing Platform

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

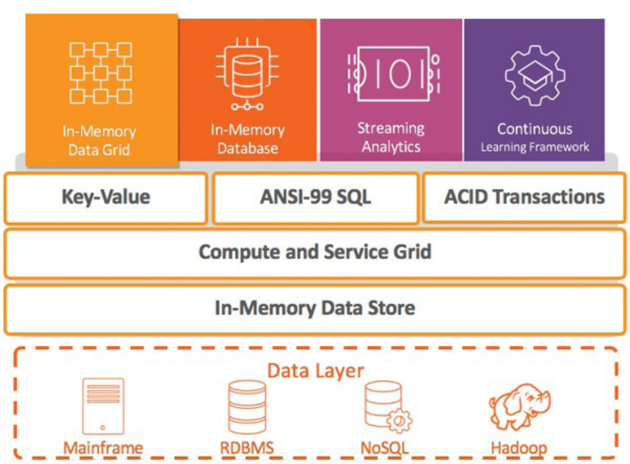


Figure 2. GridGain as an In-Memory Data Grid (IMDG)

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.

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.

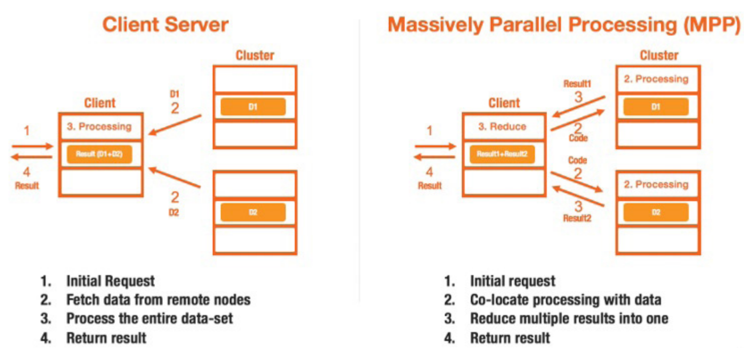


Figure 3. GridGain Compute Grid – Client Server vs Collocated Processing (MPP)

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.

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.

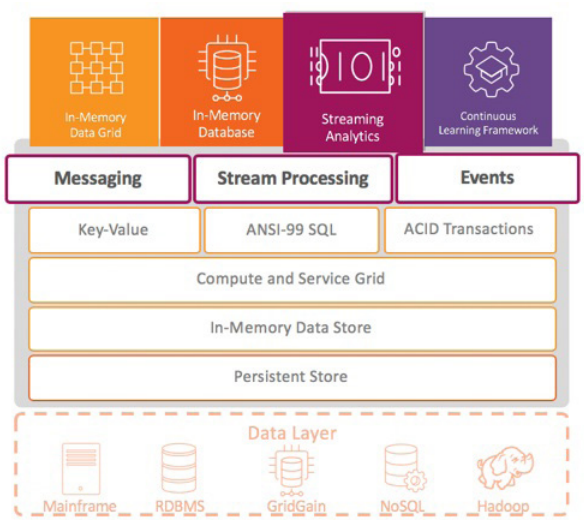


Figure 4. GridGain for Stream Ingestion, Processing and Analytics

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.

GridGain also provides the broadest in-memory computing integration with Apache Spark. The integration includes

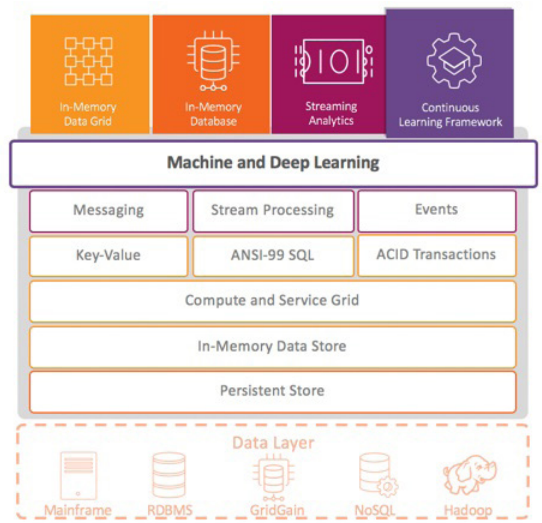


Figure 5. GridGain for Machine and Deep Learning

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.

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.

THE FUTURE OF IN-MEMORY COMPUTING

As businesses cope with an explosion of data and users who increasingly expect real-time performance, many have turned toward newly affordable, increasingly full-featured in-memory data solutions. As a result, in-memory computing platforms are becoming the systems of record for a growing number of organizations, allowing them to keep most or all their data in memory for the fastest possible performance—and providing them with a flexible platform which is ready to support the coming waves of in-memory innovation.

The in-memory platforms of the future will embrace these trends and go further. Not only will they offer the key capabilities that database users expect, such as SQL support, they will also provide a bridge to emerging use cases, such as machine learning and deep learning, and transformative new storage technologies, such as non-volatile memory. In other words, they will follow in the footsteps of Apache Ignite and GridGain. With full ANSI SQL-99 support, extensions for ML and DL, NVM support, and a hybrid storage model that enables flexible expansion to support petabytes of data, Apache Ignite and GridGain are leading the way toward a vibrant in-memory future.

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 for 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.