Supercharge eCommerce with In-Memory Computing

A GridGain Systems In-Memory Computing White Paper

The growth of eCommerce over the past 20 years has been nothing short of breathtaking. It continues to grow at an astonishing pace, pushing innovations in technology, logistics, and marketing to fascinating new levels. Amazon, which controls roughly 46 percent of all eCommerce, sold a billion items during the 2016 holiday season. This equates to processing 40 million orders—500 per second—on Black Friday alone and shipping 18 million packages that day with the help of 200,000 seasonal workers.

Government statistics show that eCommerce only represents roughly nine percent of all U.S. retail sales so there's plenty of room for growth. However, retailers trying to grow their online businesses face numerous challenges in highly competitive markets. Customer retention and revenue growth depend on how fast users can find, configure, and buy the item they want, perhaps with a push from real-time personalization that creates a unique shopping experience for each shopper. Site speed at scale is crucial, and high availability is a critical factor in providing consistent performance no matter how heavy the user load becomes on Black Friday or any other day.

With consumers wiling to walk away within seconds, the technological underpinnings of any online store or other customer-facing product must be impeccably designed, built and maintained for peak performance.

NO ONE STARTS FROM SCRATCH

Unfortunately, most eCommerce sites—even Amazon—no longer get to start from scratch. Online stores are built over time, layer by layer, with multiple engines built on multiple technology stacks, perhaps tapping into many different back ends. Quick fixes designed to speed up individual problem spots may only exacerbate problems as overlapping technology stacks grow even more tangled.

Add to that challenge the need to build for performance at scale. After all, retailers want to be ready to welcome five or ten or a hundred times as many shoppers next year and beyond. At the same time, retailers demand high-speed transaction and real-time big data analytics from the same platform. That platform must also provide high availability features such as load balancing and failover.

One way to evolve eCommerce technology to make it as fast, available, and scalable as possible is to move to an in-memory computing platform designed to address all the issues that eCommerce developers face.

IN-MEMORY COMPUTING: THE RIGHT CHOICE NOW

In-memory technology has been around for decades, ever since developers realized that working with data stored in memory was faster than working with data stored on disk. Today, that speed means everything. Digital transformation has pushed companies closer to customers who demand real-time interactions, and the amounts of Internet traffic, connected devices, and data—all of which require instant analysis—continue to grow. It's estimated that total global data will grow from eight zettabytes in 2015 to 35 or more zettabytes in 2020.

Luckily, the cost of memory continues to tumble—down 30 percent annually over time—so moving to in-memory platforms has gotten more cost-effective with each passing year. According to Gartner, the in-memory technology market will grow to $10 billion by the end of 2019, representing 22 percent compound annual growth over time.

The combination of these two trends means that the time is right for eCommerce businesses to leverage the improved performance and scale provided by in-memory computing.

APACHE® IGNITE™ TACKLES ECOMMERCE ISSUES

The growth of the in-memory computing market includes in-memory databases, in-memory analytics, in-memory data grids, in-memory app servers, in-memory messaging, event stream processing, and analytic servers. The GridGain and Apache Ignite in-memory computing platforms provide complete support for all of these components.

Apache Ignite:

- Supports data caching, massive parallel processing, in-memory SQL, streaming, and much more.

- Slides in between the existing application and data layers easily, with minimal changes to code.

- Offers ACID-compliant transactions as well as analytics support.

- Works with all popular RDBMS, NoSQL, and Hadoop databases and offers a Unified API with support for a wide range of languages.

- Can be deployed on premise, in the cloud, or in hybrid environments.

- Can be configured for high availability with automatic load balancing and failover.

GridGain

- Remains highly customizable for a multitude of uses as it handles computing and transacting on large-scale data sets in real time.

The GridGain and Apache Ignite in-memory computing platforms run everything in memory to yield performance and scalability gains that address the main concerns of eCommerce developers the most.

ECOMMERCE TAKES ADVANTAGE OF IN-MEMORY BENEFITS

GridGain has customers in a wide range of eCommerce arenas—everything from fitness and travel to Internet of Things and consumer technology. In a 2016 survey, GridGain asked them about their in-memory computing usage. The top three answers, each at or above 60 percent, were database caching, application scaling, and high-speed transactions. Also on the list: Spark acceleration, real-time streaming analysis, and database scaling. Clients said they most frequently use these features in private clouds, on premise, in AWS®, or in Microsoft Azure®.

GridGain eCommerce customers use many languages and protocols to access their data. In the survey, Java was the clear leader, with SQL, Scala, .NET, and MapReduce following behind. They also responded that the most common data stores they to use with GridGain are, in order, PostgreSQL®, Microsoft SQL Server®, MySQL®, MongoDB®, and Apache Cassandra®.

It's interesting to understand how GridGain's eCommerce customers choose to deploy the in-memory computing platform. A powerful starting point is the In-Memory Data Grid, inserted between the application and database layers to cache the disk-based data from the RDBMS, NoSQL, or Hadoop databases in RAM, delivering a performance increase up to 1,000 times. This in-memory key value store can replicate and partition data caches across multiple nodes and deliver elastic on-demand scalability to add new nodes. Distributed in-memory transactions can also be ACID-compliant. The data grid offers support for all popular RDBMS, with read-through and write-through and support for write behind. Setup remains completely flexible to address any unique use case.

Compute Grid

GridGain provides a compute grid which enables parallel, in-memory processing of CPU-intensive or other resource-intensive tasks. It can be used for any High-Performance Computing (HPC) applications that leverage Massively Parallel Processing (MPP). The compute grid helps optimize overall

cluster performance by collocating processing with data to optimize data processing and minimize network traffic. The system includes a comprehensive library of functions that includes machine and deep learning. Developers can develop and distribute their own code for any combination of transactions, analytics, stream processing or machine learning using Java, .NET or C++. They can also leverage data affinity with collocated processing to achieve linear scalability as data sets grow.
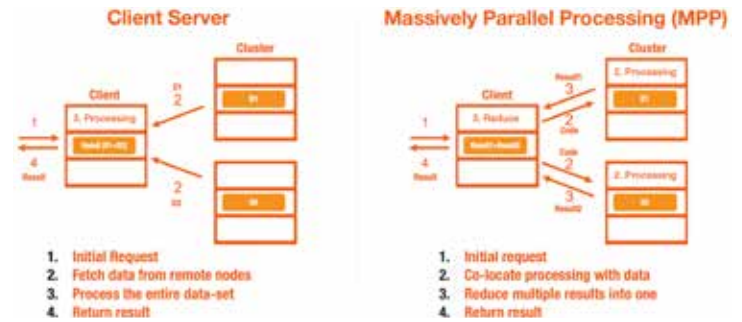

Figure 1. Ignite Compute Grid – Client Server vs Collocated Processing (MPP)

The primary capabilities of the compute grid include:

- Zero code (peer-class loading) deployment
- Dynamic clustering
- Fork-Join and MapReduce processing
- Distributed closure execution
- Load balancing and fault tolerance
- Distributed messaging and events
- Linear scalability
- Standard Java ExecutorService support
- Collocated processing support for multiple languages including Java, .NET and C++
- Distributed SQL

On top of this MPP architecture, Ignite provides ANSI-99 compliant, horizontally scalable distributed SQL. It supports all SQL, DDL and DML commands including SELECT, UPDATE, INSERT, MERGE and DELETE queries and CREATE and DROP table. Ignite supports distributed SQL joins. It allows for cross-cache joins across the entire cluster, which includes joins between data persisted in third party databases and Ignite's native persistence.

All processing, including SQL, is architected to collocate data and processing in a way that minimizes data movement across the network. Administrators can declare affinity keys such as foreign keys in DDL to partition data across the cluster. Distributed SQL joins are optimized with MPP techniques to take advantage of multi-table partitioning and replication.

This helps ensure joins can happen with data locally on each node. Ignite can perform distributed SQL as real-time or batch across a single cluster that spans a host of third-party databases with SQL and NoSQL data.

Service Grid

GridGain provides a service grid to deploy and scale micro-services across the cluster for digital business and other initiatives. It allows users to control how many instances of their service are deployed on each cluster – as a cluster singleton, node singleton, or as multiple instances across the cluster. The service grid guarantees continuous availability of all deployed services in case of node failures, including guaranteeing a single cluster or node singleton, or load balancing with multiple instances across the cluster.
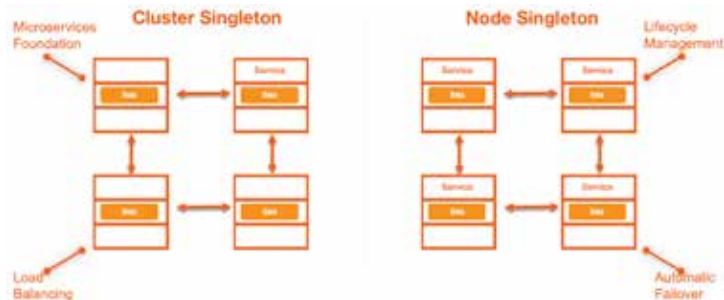

Figure 2. GridGain Service Grid – Singleton vs Cluster Configurations

Stream Ingestion, Data Management, Processing and Real-time Analytics for Streaming Analytics

GridGain is used by the largest companies in the world to ingest, process, store and publish streaming data for large-scale, mission critical business applications. It is used by several of the largest banks in the world for trade processing, settlement and compliance; by telecommunications companies to deliver call services over telephone networks and the Internet; by retailers and e-commerce vendors to deliver an improved real-time experience; and by leading cloud infrastructure and SaaS vendors 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 3. 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). When deployed together, Spark can access all of the in-memory data in GridGain, not just data streams; share data and state across all Spark jobs; and take advantage of all of GridGain's in-memory loading and processing capabilities including continuous learning to train models in near real-time to improve outcomes for in-process HTAP applications..

In-Memory Hadoop Acceleration

The GridGain accelerator for Hadoop enhances existing Hadoop environments by enabling fast data processing using the tools and technology your organization is already using today.

In-Memory Hadoop Acceleration in GridGain is based on the industry's first dual-mode, high-performance in-memory file system that is 100% compatible with Hadoop HDFS and an in-memory optimized MapReduce implementation. In-memory HDFS and in-memory MapReduce provide easy to use extensions to disk-based HDFS and traditional MapReduce.

GridGain

This plug-and-play feature requires minimal to no integration. It works with open source Hadoop or any commercial version of Hadoop, including Cloudera®, HortonWorks®, MapR®, Intel®, AWS, as well as any other Hadoop 1.x or Hadoop 2.x distribution.
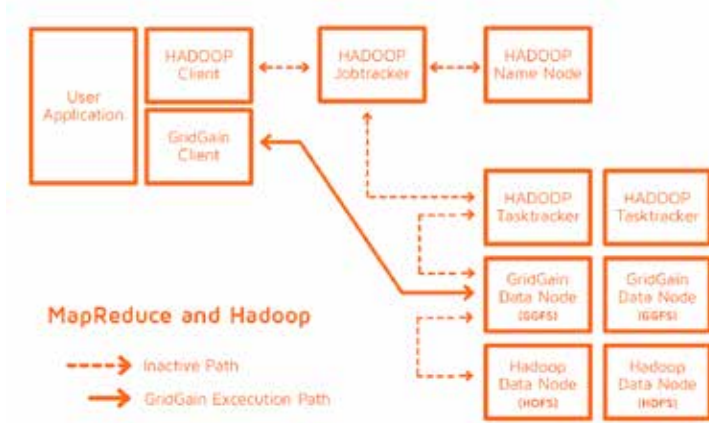

Figure 4. Ignite File System (IGFS) for Hadoop Acceleration

With all these features and tools enabled and running optimally, eCommerce sites can vastly improve their customer experiences by:

• Reacting to user choices in real time

• Delivering location and behavior-based recommendations

• Storing real-time user session data

• Serving dynamic content with sub-millisecond latency

• Customizing pricing on the fly

The package is made complete by GridGain's Web Console, a monitoring and management tool that shows the performance of individual nodes, the performance of clusters, the performance of services, and even the performance of individual SQL operations.

DEPLOYING IN-MEMORY COMPUTING PLATFORMS TO IMPROVE ECOMMERCE

The world of eCommerce offers several good examples of how GridGain's in-memory computing has improved performance and functionality for online retailers. The bottom line: high-speed transactions create customer satisfaction and revenue growth.

Real-time personalization and user session management: Experticity is an online environment where 650 lifestyle and consumer brands representing more than 4 million products get connected to brand advocates. Its developers needed to be able to react to user choices in real time and generate instant location and behavior-based recommendations. That

also meant Experticity would need to store session data in real time as well to bridge between brands and consumers, helping brands quickly improve their content to foster budding client relationships, push promotions, and focus on cross-selling and upselling.

To serve dynamic content, eCommerce developers and architects use in-memory systems to orchestrate the collection of different page content, such as text and graphics elements, from different data stores.

This is then combined with pricing and inventory databases. Ultimately, this information is integrated into a transactional shopping cart where inventory, shipping, payment, verification, and fraud detection can all happen in real time. Working with all of this in-memory means that despite an increasing number of data sources or data volume, performance and analytics stay consistently fast.

Real-time analytics: In this example, big data meets fast data across many different datasets and data structures. Real-time analytics are crucial for enabling user session analysis, generating behavior-based recommendations, making location-based offers, identifying trending items, and delivering spot promotions.

One GridGain customer needed to accelerate its existing NoSQL database, add SQL support, and create an ad-hoc query capability across two terabytes of data that it wanted to cache in memory. By moving to an in-memory computing platform it accomplished all those goals within its existing homegrown applications, vastly improve its performance, and finally gain the ability to run thousands of ad-hoc real-time reports.

Caching product catalogs: This vitally important task matters to any online retailer that wants to make a powerful impact on shoppers. To create a personalized shopping experience, a site must pull many different objects, images, text, and even video from different product catalogs and serve them all up very quickly. Getting that done requires intelligent caching across multiple data structures, different locations, and different back ends. Rule-based and customizable pre-sorting and pre-fetch is also needed. Those are precisely the features that the GridGain's In-Memory Data Grid offers.

Bringing it all together: An online consumer electronics store was in search of an open-source in-memory solution that could deliver analytics faster than its current Oracle Coherence-based systems. In their testing, GridGain was twice as fast as Oracle and demonstrated better ANSI SQL-

GridGain

99 support, resiliency and fault tolerance, and transactional consistency. The online electronics retailer ultimately re-engineered its store to include the in-memory computing platform between its database and its application code.

GRIDGAIN SYSTEMS: A LEADER IN IN-MEMORY COMPUTING

The GridGain in-memory computing platform helps eCommerce users consolidate onto a single high-performance and highly scalable big-data solution for transactions and analytics, resulting in lowered TCO. Advanced SQL functionality and API-based support for common programming languages enable rapid deployment. This, coupled with the rapidly decreasing cost of memory, boosts ROI for in-memory computing initiatives, enabling online retailers to build less expensive systems that perform thousands of times better. Any eCommerce company wrestling with the need for high-performance transactions and/or analytics across large data volumes can benefit from the GridGain in-memory computing platform.

Clients enjoy the following:

A Unified High-performance Architecture

The GridGain in-memory computing platform consists of multiple grids connected by a clustered in-memory file system. The In-Memory Data Grid, In-Memory Compute Grid, In-Memory SQL Grid and In-Memory Service Grid are interconnected. Computations occur as close as possible to the data used in the computation. Additional features such as high throughput, low latency, load balancing, caching, in-memory indexing, streaming, Hadoop acceleration and other performance improvements are crucial to success in real-time modeling, processing, and analytics.

Scalability

The GridGain in-memory computing platform excels in terms of scalability, allowing companies to add cluster nodes and memory in real-time with automatic data rebalancing. As a hardware-agnostic solution, clients can choose their preferred hardware for scaling up.

Full SQL Support

GridGain is ANSI SQL-99 compliant and the In-Memory SQL Grid supports DML users can leverage their existing SQL code using the GridGain JDBC and ODBC APIs. For users with existing code bases which are not based on SQL, they can leverage their existing code through supported APIs for Java, NET, C++, and more.

High Availability

The GridGain in-memory computing platform provides essential high availability features such as data-center replication, automatic failover, fault tolerance, and quick recovery on an enterprise-level scale.

Transaction Processing

The GridGain in-memory computing platform supports ACID-compliant transactions in a number of user-configurable modes.

Security Features

The GridGain in-memory computing platform supports authentication, authorization, multiple encryption levels, tracing, and auditing.

Open Source Framework

GridGain is based on Apache® Ignite™, a popular open source project with many contributors that has been tested globally. GridGain Systems was the original creator of the code contributed to the Apache Software Foundation that became Apache Ignite and fully supports the technology behind Apache Ignite. The GridGain Enterprise Edition extends the features in Apache Ignite to provide enterprise-level capabilities and services, such as additional security, data center replication, auditing mechanisms, a GUI for management and monitoring, network segmentation, and a recoverable local store.

Production Support

GridGain is the only company to provide commercial support for Apache Ignite. GridGain Basic Support for Apache Ignite and the GridGain Community Edition includes timely access to professional support via web or email. The team can help troubleshoot performance or reliability issues and suggest workarounds or patches, if necessary. A two-hour initial consultation allows our support team to understand your current environment for more effective support in the future. The consultation helps identify issues and improve the performance or reliability of your deployment.

Standard Support is for companies deploying the GridGain Enterprise Edition or the GridGain Community Edition in production. With 24x7 support hours and web, email and phone access, Standard Support is perfect for ongoing production deployments. An annual license to the GridGain Enterprise or Community Edition is available with the subscription.

Premium Support is for companies deploying the GridGain Enterprise Edition or the GridGain Ultimate Edition for mission-critical applications. Premium Support is available 24x7 with the fastest initial response time, more named support contacts than Standard Support, and web, email and phone access. Premium Support is available with a license to the GridGain Enterprise or Ultimate Edition.

PERFORMANCE, SCALABILITY, PERSONALIZATION…AND HAPPY CUSTOMERS

An in-memory computing platform can deliver a high-performance experience featuring real-time personalization, content personalization, high-speed transactions, user-session tracking, and real-time analytics across a variety of platforms. With that toolkit in hand, eCommerce developers can find a path to a future of sales growth and success.

SOLUTION	UNSUPPORTED	BASIC SUPPORT	STANDARD SUPPORT	PREMIUM SUPPORT
Software	Ignite/Community	Ignite/Community	Community/Enterprise	Enterprise/Ultimate
Troubleshoot Performance or Reliability Issues		●	●	●
Identify Bugs and Product Limitations & Workarounds		●	●	●
Support Hours		9x5	24/7	24/7
Initial Response Time		4 Hours	2 Hours	1 Hour
Unlimited Support Incidents		●	●	●
Named Technical Contacts		2	3	4
Initial Consultation Support		2 Hours		
Maintenance Releases			●	●
Hot Bug Fixes			●	●
Enhanced Features (Enterprise & Ultimate Editions)			●	●
Custom Contracts			●	●
Login Support to Customer Environment			●	●
Support Channels		Web/Email	Web/Email/Phone	Web/Email/Phone

Figure 5. GridGain Support Level Comparison

GridGain

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 fill out the form at www.gridgain.com/contact and we will contact you.

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.