GridGain

Introducing the GridGain
In-Memory Computing Platform

A GridGain Systems In-Memory Computing White Paper

OVERVIEW

Companies pursuing digital transformation and omnichannel customer experience initiatives often find that their existing IT infrastructure is unable to support the speed and scalability these initiatives demand. The need to provide a more personalized, real-time end user experience has forced companies to transform batch-based processes that take days or hours into real-time, automated processes that take seconds or less. Companies that have adopted web, mobile, social and IoT technologies to support their transformation have experienced 10-1000x increases in queries and transactions. These initiatives have created a host of new data sources about the end users and business that have resulted in a 50x explosion in data volumes over the last decade. How can companies support 10-1000x increases in the volume of queries and transactions, leverage 50x as much data to make better decisions, and do everything that used to take hours or days in seconds or fractions of a second?

The answer for many companies has been in-memory computing. In-memory computing offers speed and scalability for existing and new applications. The speed comes from storing and processing data in memory rather than continually retrieving data from disk before processing. While hard drive (HDD) media speeds are measured in milliseconds, RAM speeds can be measured in nanoseconds, a million times faster. Scalability comes from distributing data and computing together horizontally across a cluster of servers.

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™ 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 Apache Ignite are used by thousands of companies worldwide to add in-memory speed and unlimited horizontal scalability to: existing or new OLTP or OLAP applications; new or existing hybrid transactional/analytical processing (HTAP) applications; streaming analytics; and continuous learning use cases involving machine or deep learning. With GridGain, companies have:

- Added speed and scalability by sliding GridGain in-between existing applications and databases as an in-memory data grid (IMDG) with no rip-and-replace of the applications or databases

- Developed high volume, low latency HTAP applications with unlimited scalability

- Leveraged streaming analytics and real-time processing including machine and deep learning to automate and improve business decisions
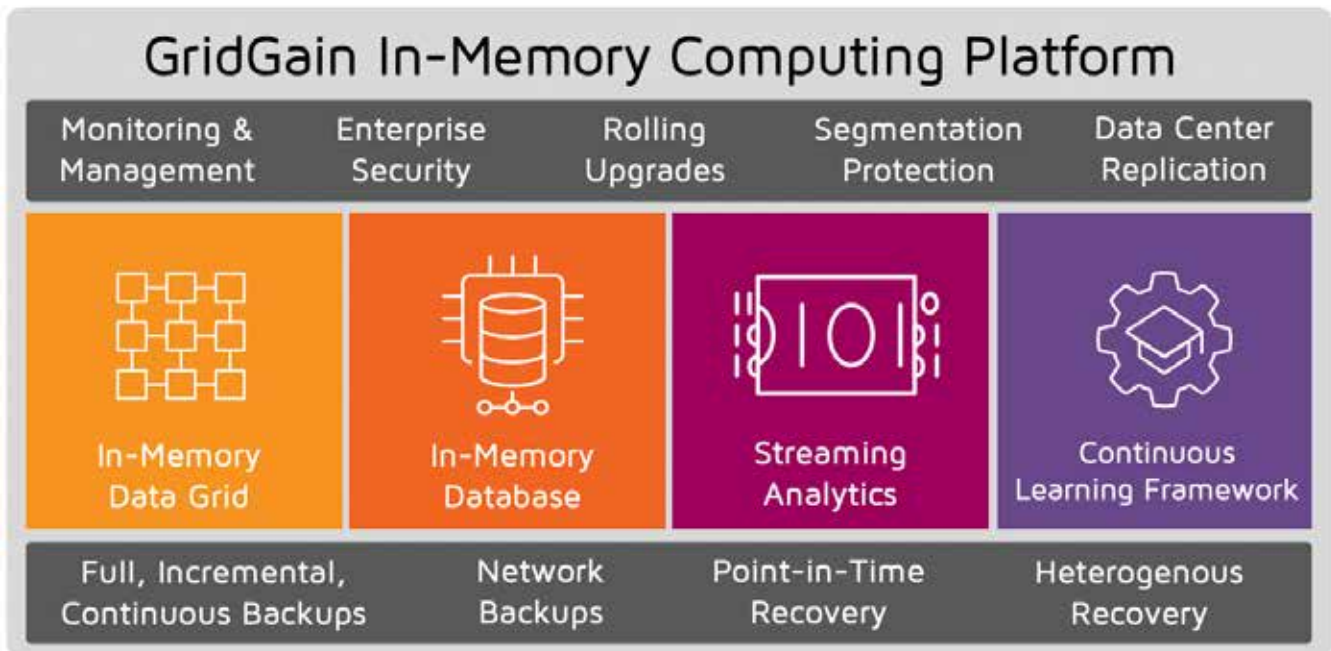

Figure 1. The GridGain In-Memory Computing Platform

GridGain customers have created a new shared in-memory data foundation. This single system of record for transactions and analytics enables real-time action and visibility for their business. With each project, they have opened up more data for other applications on an in-memory computing platform with real-time performance at peak loads and always-on availability. As a result, they have been able to develop new projects faster and at a lower cost, be more flexible to change, and be more responsive in ways that have improved their end user experiences and business outcomes.

GRIDGAIN SOFTWARE EDITIONS

GridGain product editions include different capabilities and levels of support to fit specific needs. They are subscription-based products included with GridGain Support.

The GridGain Community Edition is a binary build of Apache Ignite created by GridGain Systems for companies that want to run Apache Ignite in a production environment. It includes optional LGPL dependencies, such as Hibernate L2 cache integration and Geospatial Indexing. It benefits from ongoing QA testing by GridGain Systems engineers and contains bug fixes which have not yet been released in the Apache Ignite code base. It is suitable for small-scale deployments which do not require support for multiple data centers, enhanced management and monitoring capabilities, or enterprise-grade security.

The GridGain Enterprise Edition is for companies that plan to run GridGain as an IMDG in production. It includes all of the features of Apache Ignite plus enterprise-grade features including data center replication, enterprise-grade security,

rolling upgrades, expanded management and monitoring capabilities, and more. The Enterprise Edition is extensively tested by GridGain Systems and is recommended for use in large-scale or mission-critical production deployments or environments with heightened security requirements.

The GridGain Ultimate Edition is for companies that plan to run GridGain as an in-memory database (IMDB) in production. It enables users to put the GridGain Persistent Store into production with confidence. The Ultimate Edition includes all the features of the GridGain Enterprise Edition plus centralized backup and recovery management. You can perform full, incremental and continuous backups locally or across a network, and point-in-time recovery as well as heterogeneous recovery where you can restore a cluster to any location on premise or in the cloud with a different size and topology.

GRIDGAIN 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.
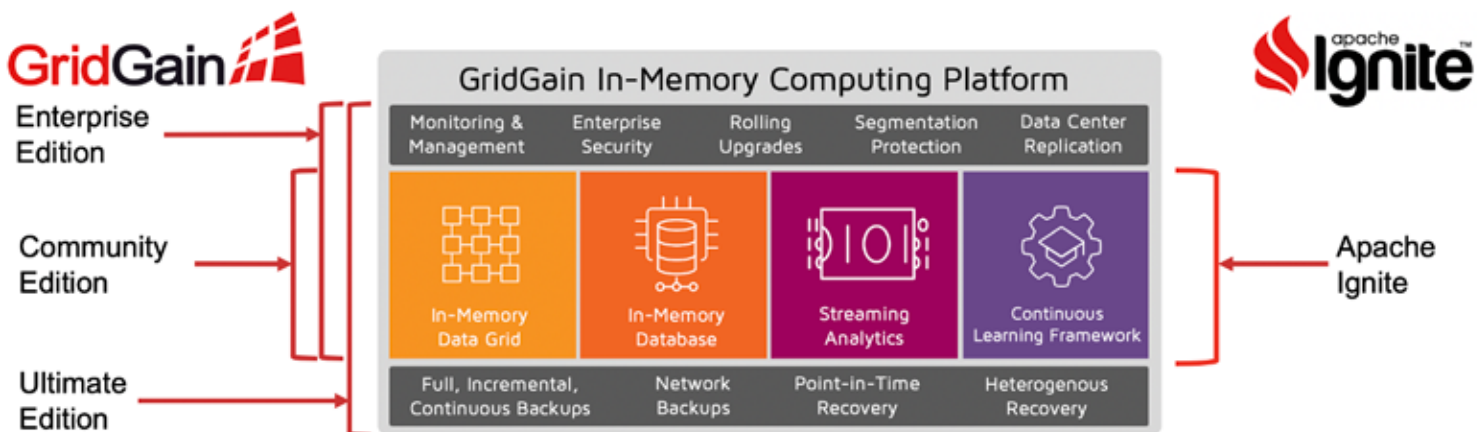

Figure 2. Apache Ignite and GridGain Editions

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.

SUPPORT LEVELS	PRODUCTS			
	APACHE IGNITE	GRIDGAIN COMMUNITY EDITION (CE)	GRIDGAIN ENTERPRISE EDITION (EE)	GRIDGAIN ULTIMATE EDITION (UE)
Basic	●	●		
Standard		●	●	
Premium			●	●

Figure 3. GridGain Support Levels

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 4. GridGain Support Level Comparison

GridGain

APACHE IGNITE AND GRIDGAIN ARCHITECTURE

Memory-Centric Storage

GridGain provides a distributed in-memory data store that delivers in-memory speed and unlimited read and write scalability to applications. It is a distributed, in-memory SQL and key-value store that supports any kind of structured, semi-structured and unstructured data. GridGain's unlimited horizontal scalability comes from a shared-nothing, node-based cluster architecture. Each node delivers low latency and predictable access times using an in-memory first architecture that stores data in off-heap RAM by default.

Third Party Persistence and the GridGain Persistent Store

GridGain can bring together almost any data into memory and deliver unlimited read scalability on top of third party databases. GridGain can sit as an IMDG on top of all popular RDBMSs such as IBM DB2®, Microsoft SQL Server®, MySQL®, Oracle® and PostgreSQL®. It also works with NoSQL databases such as Apache Cassandra™ or MongoDB® and with Apache Hadoop™. GridGain also provides its own built-in persistence with the GridGain Persistent Store, a distributed in-memory database (IMDB). Its performance for high volume, low latency transactions and data ingestion exceeds the read and write performance and scalability of traditional databases. GridGain can sit on top of all these databases at the same time as an IMDG and coordinate transactions in-memory with the underlying databases to ensure data is never lost.

ANSI-99 compliant SQL

Unlike other in-memory data grid technologies, GridGain supports high performance, low latency ANSI-99 compliant distributed SQL. GridGain is the only vendor that supports SQL DDL and DML for real-time or batch queries and transactions, for any data spread across any combination of third party databases or the GridGain Persistent Store. This enables companies to use their existing SQL assets and SQL skillsets with in-memory computing instead of having to rewrite applications or replace databases.

Support for ACID Transactions

GridGain also has the broadest support for distributed ACID transactions. GridGain's integrated SQL and ACID transaction support makes it the only technology that can add speed and scalability by sliding in-between SQL-based applications and RDBMSs and preserving the use of SQL. GridGain inter-cepts all SQL queries and transactions. This architecture has offloaded all queries from existing databases, lowered SQL query times 10-1000x and delivered unlimited horizontal read scalability. It has given companies a way to handle the increased loads without having to rip out and replace existing applications and databases. It has also provided a way to easily migrate data at any time from existing databases to GridGain Persistent Store as an IMDB to improve transaction and write performance.

Massively Parallel Processing (MPP)

GridGain also eliminates another common bottleneck associated with Big Data; the network. Data has gotten so big that just moving it takes minutes or hours. With GridGain, data can be distributed to nodes based on data affinity declarations that help collocate processing with data. This improves performance and scalability by reducing data transmission over the network. GridGain provides the broadest implementation of distributed MPP algorithms to enable collocated processing. It supports the parallelization of distributed computations based on Fork/Join and extensively uses distributed parallel computations internally. GridGain's distributed SQL and key-value APIs, MapReduce, compute grid, service grid, streaming analytics and machine learning capabilities all leverage MPP. GridGain also exposes APIs so developers can add and distribute their own user-defined MPP functionality using C++, Java, or .NET.

The combination of GridGain's shared-nothing architecture and MPP delivers unlimited horizontal, linear scalability. Companies can scale out data and MPP horizontally using any combination of servers or cloud infrastructure. This approach pays for itself by providing a much more cost-effective alternative than scaling existing databases or applications that can often only be scaled vertically with expensive hardware.

In-Memory Support for Third-Party Technologies

GridGain enables in-memory computing to be used for just about any type of project by providing the broadest third-party technology support. This includes support for all leading RDBMSs, for NoSQL databases including Cassandra and MongoDB, and the broadest integration with Apache Spark™ and Apache Hadoop. When used with these technologies, GridGain transparently adds in-memory access to any data with the speed, scalability and power of GridGain.

GridGain

Unified API

GridGain makes data accessible anywhere as SQL or key-values through a unified API. You can securely access any data in any cluster regardless of its deployment configuration. It is implemented across:

- SQL (ODBC and JDBC)
- REST, .NET, C++, Java/Scala/Groovy and SSL/TLS
- A binary protocol that supports lightweight clients for any language

Linear Horizontal Scalability

GridGain and Apache Ignite are Java-based in-memory computing platforms. They provide one of the most sophisticated horizontal scale-out and clustering architectures available. The architecture delivers elastic scalability: nodes can be dynamically added to, removed from, or upgraded in a running cluster with zero downtime. Nodes can automatically discover each other across LANs, WANs and any cloud infrastructure.

All nodes in a GridGain cluster are equal and can play any logical role. The node flexibility is achieved by using a Service Provider Interface (SPI) design for every internal component. The SPI model makes GridGain fully customizable and pluggable. This not only enables tremendous configurability of the system. It gives companies the flexibility to support existing and future server infrastructure on premises or in the cloud.

Deployable Anywhere

GridGain is easy to use and deploy anywhere on commodity servers. It can be deployed on premises, in a private cloud using Kubernetes®, Docker® or other technologies, on public cloud platforms such as Amazon Web Services (AWS®), Microsoft Azure® or Google Cloud Platform®, or on a hybrid private-public cloud.

HOW COMPANIES USE GRIDGAIN

GridGain is typically used to:

- Add speed and scalability to existing applications
- Build new, modern, highly performant and scalable transactional and/or analytical applications
- Build streaming analytics applications, often with Apache Spark, Apache Kafka™ and other streaming technologies
- Add continuous machine and deep learning to applications to improve decision automation

Companies start in any one of these areas. Over time, as

GridGain is used with more projects, it becomes a common in-memory data access layer that can support the data, performance and scalability needs for any new workload:

- Data services and other APIs that help deliver an omni-channel digital business
- New customer-facing applications that support new products and services
- Real-time analytics that help improve operational visibility or compliance
- Streaming analytics, machine and deep learning that help improve the customer experience and business outcomes

GridGain helps deliver these types of projects faster while giving companies a foundation for a more real-time, responsive digital business model and the ability to be more flexible to change.

IN-MEMORY DATA GRID (IMDG) FOR ADDING SPEED AND SCALABILITY TO APPLICATIONS

A core GridGain capability and most common use case 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 no major architectural changes. GridGain supports all common RDBMSs including IBM DB2, Microsoft SQL Server, MySQL, Oracle and PostgreSQL, NoSQL databases such as Cassandra and MongoDB, and Hadoop.
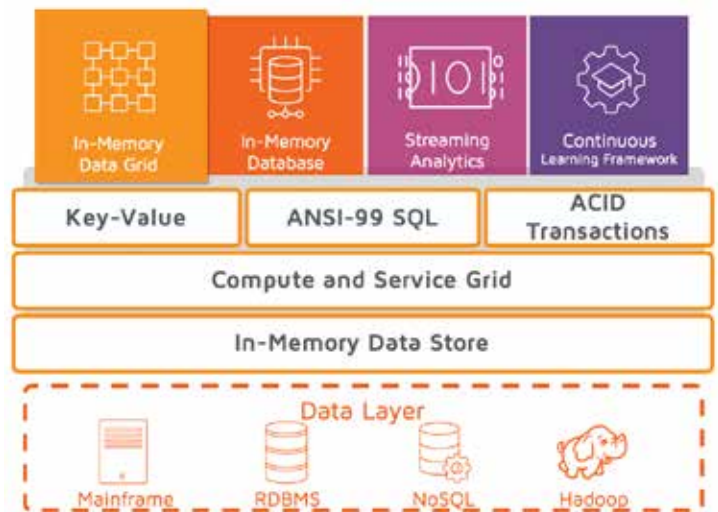

Figure 5. GridGain as an In-Memory Data Grid (IMDG)

GridGain generates the application domain model based on the schema definition of the underlying database. It then loads the data and acts as the data platform for the application. GridGain handles all reads and coordinates transactions

with the underlying database in a way that ensures data consistency in both the database and GridGain. By utilizing RAM in place of disk, GridGain lowers latency by orders of magnitude compared to traditional disk-based databases.

The primary benefits and capabilities of the Ignite IMDG include:

• ANSI SQL-99 support including DML and DDL

• ACID transaction support

• In-memory performance orders of magnitude faster than disk-based RDBMSs

• Distributed in-memory caching that offloads queries from the existing database

• Elastic scalability to handle up to petabytes of in-memory data

• Distributed in-memory queue and other data structures

• Web session clustering

• Hibernate L2 cache integration

• Tiered off-heap storage

• Deadlock-free transactions for fast in-memory transaction processing

• JCache (JSR 107), Memcached and Redis client APIs that simplify migration from existing caches

HYBRID IN-MEMORY DATABASE (IMDB) FOR HIGH VOLUME, LOW LATENCY TRANSACTIONS AND DATA INGESTION

A GridGain cluster can also be used as a distributed, transactional IMDB to support high volume, low latency transactions and data ingestion, or for low cost storage.

The GridGain IMDB combines distributed, horizontally scalable ANSI-99 SQL and ACID transactions with the GridGain Persistent Store. It supports all SQL, DDL and DML commands including SELECT, UPDATE, INSERT, MERGE and DELETE queries and CREATE and DROP table. GridGain parallelizes commands whenever possible, such as distributed SQL joins. It allows for cross-cache joins across the entire cluster, which includes joins between data persisted in third party databases and the GridGain Persistent Store. It also allows companies to put 0-100% of data in RAM for the best combination of performance and cost.

The in-memory distributed SQL capabilities allow developers, administrators and analysts to interact with the GridGain platform using standard SQL commands through JDBC or ODBC or natively developed APIs across other languages as well.

Figure 6. GridGain as an IMDB
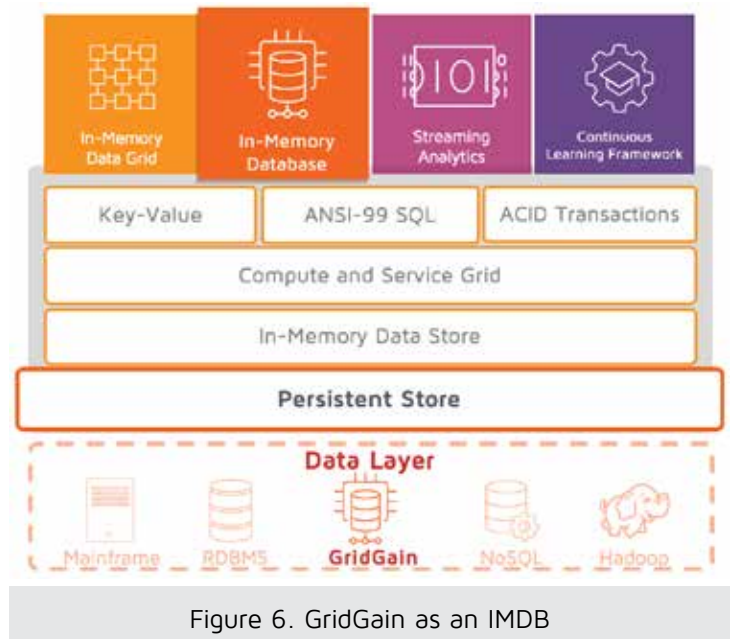
The primary capabilities of the GridGain's hybrid IMDB include:

• ANSI SQL-99 compliance

• ACID transactions support

• Full support for SQL DML including SELECT, UPDATE, INSERT, MERGE and DELETE

• Support for DDL commands including CREATE and DROP table

• Support for distributed SQL joins, including cross-cache joins across the entire cluster

• SQL support through JDBC and ODBC without custom coding

• Geospatial support

• Hybrid memory support for RAM, HDD, SSD/Flash, 3D XPoint and other storage technologies

• Support for maintaining 0-100% of data in RAM with the full data set in non-volatile storage

• Immediate availability on restart without having to wait for RAM warmup

GridGain

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.

Figure 7. GridGain for Stream Ingestion, Processing, and Analytics

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 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.

CONTINUOUS LEARNING FRAMEWORK FOR MACHINE LEARNING AND DEEP LEARNING

Ignite also provides Ignite Machine Learning (ML), MPP-style machine learning and deep learning with real-time performance on petabytes of data. Ignite provides several standard machine learning algorithms optimized for collocated processing including linear and multi-linear regression, k-means clustering, decision trees, k-NN classification and regression. It also includes a multilayer perceptron for deep learning along with TensorFlow integration. Developers can develop and deploy their own algorithms across any cluster as well as using the compute grid.
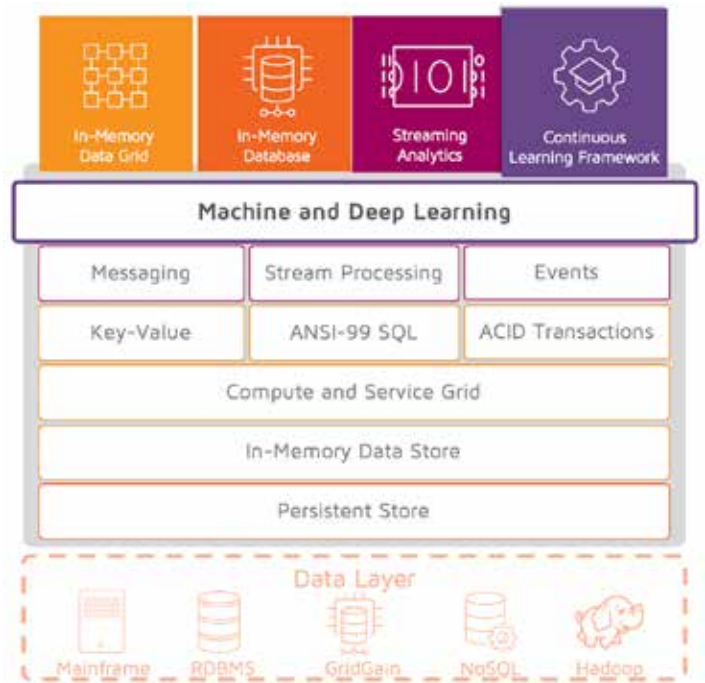

Figure 8. GridGain for Machine Learning and Deep Learning

GridGain

CORE FEATURES OF GRIDGAIN COMMUNITY EDITION (AND APACHE IGNITE)

The GridGain Community Edition, the enterprise-ready and supported version of Apache Ignite, combines a core set of capabilities that support many use cases in one integrated platform (see Figure 1.) GridGain includes a high performance in-memory data store combined with support for both third-party persistence and the GridGain Persistent Store. It offers a broad range of MPP capabilities that include distributed SQL, a general-purpose compute grid, a service grid for micro-services, stream processing and analytics, and machine and deep learning. It provides integration with third-party technologies to add in-memory computing to a host of use cases. It also exposes a unified API across languages that enables nearly any type of client to access these capabilities.

In-Memory Data Store

GridGain provides an in-memory data store where each node in a GridGain cluster by default stores data in RAM. The data is kept in off-heap storage to ensure low latency and consistent access times. The system is multi-model, with the ability to support structured, semi-structured and unstructured data. The data is accessible via SQL using JDBC/ODBC drivers or APIs, or using key-value APIs. GridGain provides a durable memory architecture where in-memory data can be mapped to leading third-party databases or to the GridGain Persistent Store to save data within the same cluster. This support enables companies to bring almost any data into RAM and create a single in-memory data layer across existing and new systems.

Third-Party Database and Native Persistence Support

GridGain can be deployed as an IMDG on top of all leading relational databases (including IBM DB2, Microsoft SQL Server, MySQL, Oracle and PostgreSQL), NoSQL databases (such as Cassandra or MongoDB), and Hadoop and use these databases for persistence. When used with a third-party database, GridGain holds the most up-to-date version of the data in-memory. Whenever a transaction occurs, GridGain passes it to the underlying database and, upon successful completion, updates the data in-memory. This enables GridGain to offload all reads from databases, increasing speed and providing existing systems a lot more room for growth.

GridGain also offers the GridGain Persistent Store that allows GridGain to be used as a distributed hybrid IMDB. It is a multi-model database that combines ANSI-99 compliant SQL with key-value API and ACID transaction support. It is also a hybrid memory database that supports any combination of RAM, HDD, SSD/Flash, 3D XPoint and other storage technologies.

GridGain Persistent Store lets you choose the best combination of performance and cost for each situation. With GridGain Persistent Store you can have 0-100% of your data and SQL indexes in RAM across the cluster while the full data set and indexes are stored in non-volatile storage to provide durability and availability. All data is stored and treated the same way. GridGain uses write-ahead logs for transactions and incremental snapshots optimized to ensure in-memory speed, data consistency and recoverability.

GridGain Persistent Store also provides immediate recovery in the case of a failure. You do not need to wait for data to be loaded in RAM after a cluster restart. All operations use data stored on disk until that data is loaded into RAM. Immediate recovery is one of the many features that have helped companies ensure high availability.

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.
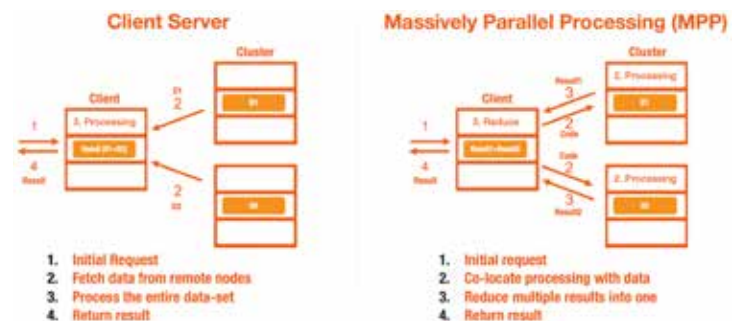

Figure 9. Ignite Compute Grid – Client Server vs Collocated Processing (MPP)

© 2019 GridGain Systems, Inc.

GridGain

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.

ACID Transaction Support

GridGain provides user-tunable ACID transaction support. You can define whether the system enforces strict or eventual consistency by using pessimistic or optimistic transactions. Strict consistency with pessimistic transactions supports high-value transaction use cases where consistency is more important than speed. Eventual consistency using optimistic transactions can be valuable for use cases where speed is paramount and the potential later rollback or correction of some transactions would be acceptable.

Service Grid

GridGain provides a service grid to deploy and scale microservices 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.
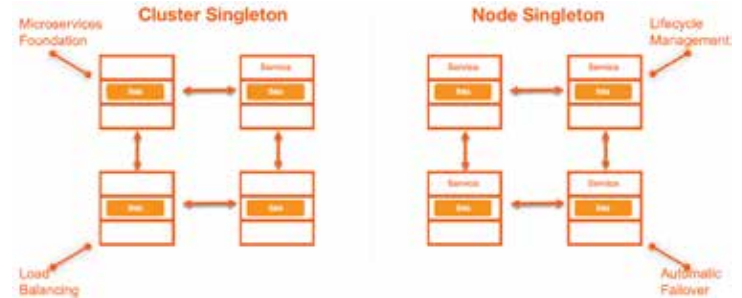

Figure 10. GridGain Service Grid – Singleton vs Cluster Configurations

Distributed Messaging and Events

GridGain provides high performance, cluster-wide messaging functionality to exchange data via publish- subscribe and direct point-to-point communication models. The primary capabilities of distributed messaging include support for topic-based publish-subscribe models and direct point-to-point communication, a pluggable communication transport layer, support for message ordering and cluster-aware message listener auto-deployment.

The distributed events functionality in GridGain allows applications to receive notifications about cache events occurring in a distributed grid environment. Developers can use this functionality to be notified about the execution of remote tasks or any cache data changes within the cluster. In GridGain, event notifications can be grouped together and sent in batches and/or timely intervals. Batching notifications help attain high cache performance and low latency.

The main capabilities of Distributed Events in GridGain include the ability to subscribe to local and remote listeners, enable and disable any event, set local and remote filters for fine-grained control over notifications, and batch notifications automatically to improve performance.

Unified API and Third-Party APIs

The GridGain unified API supports a wide variety of common protocols for the application layer to access the same data using SQL or a key-value API. Supported protocols include: SQL (via JDBC/ODBC and REST API), REST, C++, .NET, MapReduce, Java/Scala/Groovy, SSL/TLS, and a binary protocol that supports lightweight clients for any language.

GridGain

GridGain enables companies to either implement or replace caching technologies using industry standard or third-party APIs:

- JCache (JSR 107) implementation that leverages the full power of Ignite for in-memory data caching, collocated processing and pluggable persistence.

- Hibernate second-level (L2) cache that can be made visible to all sessions and supports clustering and transactions

- Memcached-compliant API for Java, PHP, Python, Ruby and other Memcached clients

- Redis-compatible API that Redis clients can use to store and retrieve data from Ignite

GridGain also allows for most of the data structures from the java.util.concurrent framework to be used in a distributed fashion. For example, you can take java.util.concurrent.BlockingDeque and add to it on one node and poll it from another node. Or you could have a distributed Primary Key generator, which would guarantee uniqueness on all nodes.

Apache Cassandra Integration

Apache Cassandra, like Apache Ignite, supports high volume write performance with linear scalability and high availability using a scale-out architecture. Cassandra is very powerful for certain use cases where high volume write performance is important but has its limitations. Cassandra only guarantees eventual consistency. This means queries, such as lookups for inventory availability, are not immediately up-to-date, so it's generally not used for transactional applications where real-time accuracy is important.

Cassandra can also be efficient for certain types of pre-defined queries. However, it lacks the query flexibility people expect from SQL, such as the ability to do any join, aggregation, grouping or ad-hoc query. Cassandra requires that the data is modeled such that each pre-defined query results in one row retrieval. This pre-planning requires knowledge of the required queries before modeling the data. Each new business request for data can result in the need to either change the Cassandra model design or do the analytics outside of Cassandra.

When used as an IMDG with Cassandra, GridGain improves read performance and adds ANSI-99 compliant SQL support to run any queries, including joins, aggregations and groupings. GridGain can also be used in place of Cassandra for new applications requiring any combination of high volume read and write performance, SQL and query flexibility, or support for immediate transactional consistency. Ignite also supports many additional types of projects. It can accelerate existing applications with no rip-and-replace, deliver lower latency through in-memory computing, and improve application scalability through MPP-style collocated processing.

Streaming Analytics with Apache Spark

Apache Spark is an open source fast and general-purpose engine for large-scale data processing of event-driven streaming data in memory. GridGain is the ideal underlying in-memory data management technology for Apache Spark because of its in-memory support for managing stored "data at rest" and ingesting and processing streaming "data in motion".
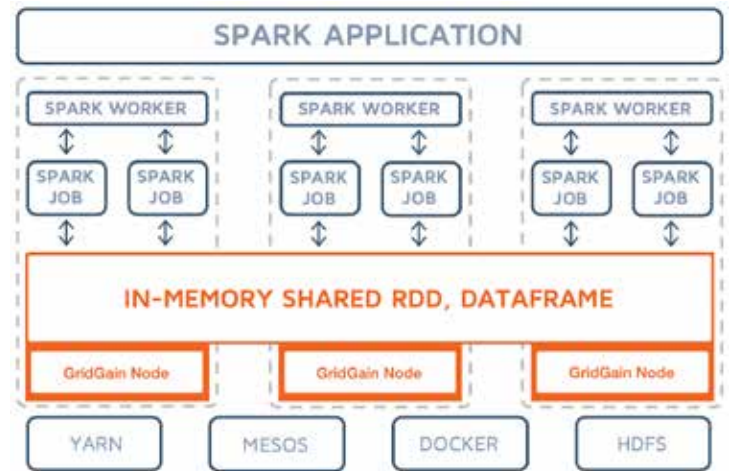

Figure 11. GridGain and Apache Spark

GridGain provides the broadest in-memory computing integration with Apache Spark. Native support for Spark Data-Frames allows Spark developers to access data from and save data to GridGain to share both data and state across Spark jobs. The GridGain RDD API lets developers read from and write to GridGain caches as mutable RDDs, unlike existing immutable Spark RDDs. Both the RDD and DataFrame support make GridGain caches accessible locally in RAM inside Spark processes executing Spark jobs.

GridGain also integrates its distributed SQL into SparkSQL plans. This allows Spark to take advantage of the advanced indexing and MPP-style distributed joins in GridGain. The combination can improve Spark SQL query performance by as much as 1000x.

The GridGain File System (GGFS) provides in-memory access via HDFS. Spark developers are able to leverage all of GridGain's in-memory storage and processing capabilities including machine 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.
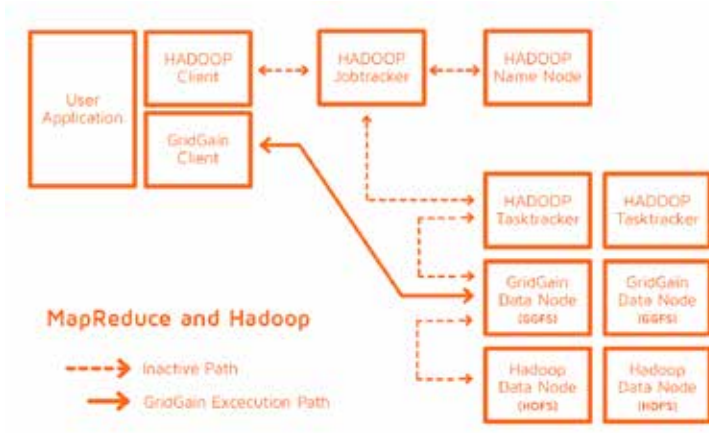

Figure 12. Ignite File System (IGFS) for Hadoop Acceleration

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.

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.

Distributed In-Memory File System

One of the unique capabilities of GridGain is a file system interface to its in-memory data called the GridGain File System (GGFS). GGFS delivers similar functionality to Hadoop HDFS, including the ability to create a fully functional file system in memory. GGFS is at the core of the GridGain In-Memory Accelerator for Hadoop and can be plugged into any Hadoop or Spark environment.

The data from each file is split on separate data blocks and stored in cache. Developers can access the data in each file with a standard Java streaming API. For each part of the file a developer can calculate an affinity and process the file's content on corresponding nodes to avoid unnecessary networking.

WHAT GRIDGAIN® ADDS TO APACHE IGNITE

The GridGain Enterprise and Ultimate Editions include all of the features in Apache Ignite and the GridGain Community Edition plus additional integration, security, deployment, monitoring, network segmentation protection and optimization, data center management, and high availability capabilities. The GridGain Enterprise Edition includes all of these features for mission critical in-memory data grid use cases. For those using GridGain as an in-memory database leveraging the GrdiGain Persistent Store, the Ultimate Edition includes all of the Enterprise Edition features plus advanced backup and recovery capabilities.

MANAGEMENT AND MONITORING

GridGain Web Console, the GUI-based Management and Monitoring tool, provides a unified operations, management and monitoring system for GridGain deployments. GridGain Web Console provides management and monitoring views into all aspects of GridGain operations. This includes HPC, Data Grid, Streaming, and Hadoop acceleration via standard dashboards, advanced charting of performance metrics, and grid health (telemetry) views, among many other features.

ENTERPRISE GRADE SECURITY

The GridGain Enterprise Edition includes enterprise-grade Security that provides extensible and customizable authentication and security capabilities. It includes both a Grid Authentication SPI and a Grid Secure Session SPI to satisfy a variety of security requirements.

NETWORK SEGMENTATION PROTECTION

Network Segmentation Protection detects any network disruption within the grid to manage transactional data grids during a 'split brain' scenario. The options for handling these network occurrences are fully configurable to help ensure the best approach to recover from different types of network-related issues.

ROLLING PRODUCTION UPDATES

The Rolling Production Updates feature enables you to co-deploy multiple versions of GridGain and allow them to co-exist as you roll out new versions. This prevents downtime when performing software upgrades.

DATA CENTER REPLICATION

GridGain reliably replicates data on a per-cache basis across two or more regions connected by wide area networks. This allows geographically remote data centers to maintain consistent views of data. With GridGain reliability and predictability, Data Center Replication ensures business continuity and can be used as part of a disaster recovery plan. Data Center Replication integrates with your application so that caches marked for replication are automatically synchronized across the WAN link. Predictable performance includes:

• Asynchronous data replication to maximize performance

• Fault tolerance

• Conflict resolution when needed

• Scaling (at the local cluster, and in the number of WAN-linked clusters)

• Multiple types of topologies supported

• Brokerless (no JMS broker needed for replication)

ORACLE GOLDENGATE INTEGRATION

The Oracle GoldenGate integration in the GridGain Enterprise and Ultimate Editions provides real-time data integration and replication into a GridGain cluster from different environments. When configured, the GridGain in-memory computing platform will automatically receive updates from the connected source database, converting the data from a database relational model to cache objects.

CENTRALIZED BACKUP AND RECOVERY MANAGEMENT

The GridGain Ultimate Edition provides centralized backup and recovery using either the GridGain Web Console or Snapshot Command Line Tool. You can perform, schedule and manage backups, and then recover to any point in time on any cluster using a combination of full and incremental snapshots with continuous archiving. This includes the ability to backup remotely using network backups, and then (re) deploy to a different cluster of any size anywhere on premise or in the cloud. You can also deploy backups to support testing in development, quality assurance (QA) and staging environments.

FULL, INCREMENTAL AND CONTINUOUS BACKUPS

Within the GridGain Web Console or Snapshot Command Line Tool you can centrally perform or schedule full and incremental snapshots across a distributed cluster. You can then use them as backup and restore points for later recovery. You can also use continuous archives of write-ahead log (WAL) files to backup down to each transaction. The combination of full and incremental snapshots with continuous archiving helps ensure data is never lost.

The GridGain Ultimate Edition minimizes the administrative and hardware costs of backup and recovery by providing the ability to schedule full and incremental snapshot creation and deletion, as well as the movement of snapshots to another (remote) lower cost location.

NETWORK BACKUPS

With the GridGain Ultimate Edition, snapshots do not need to be stored locally on the same cluster machines used to handle the operational load. They can also be managed and stored remotely on premise or in the cloud. When combined with the remote storage of continuous archives, this capability helps ensure a cluster can be quickly recovered even if an entire data center disappears.

POINT-IN-TIME RECOVERY

You can quickly restore a GridGain Ultimate Edition cluster to any point in time through the combination of full and incremental backups with continuous archiving. Point-in-time recovery can be used to restore a system up to any change without having to manually resubmit or replay existing transactions that occurred following a full or incremental snapshot. Continuous archiving helps spread network loads to minimize peak network traffic. It also allows recovery to be more granular and up-to-date, which helps reduce overall downtime needed to restore a cluster to a current, valid state.

HETEROGENEOUS RECOVERY

The GridGain Ultimate Edition also allows you to restore an existing cluster to another location, on premise or in the cloud, with a different size and topology. GridGain already allows you to dynamically add nodes to a cluster for scalability, and create a hybrid cluster across any collection of nodes or data centers on premise or in the cloud. Heterogeneous Recovery enables you to rapidly bring up a different size cluster the moment an existing cluster goes down or bring a new cluster up so that you can take an existing cluster down. This helps reduce downtime and increase availability.

SOLUTION	APACHE IGNITE	COMMUNITY EDITION	ENTERPRISE EDITION	ULTIMATE EDITION
In-Memory Data Grid	●	●	●	●
In-Memory Database	●	●	●	●
Streaming Analytics	●	●	●	●
Continuous Learning	●	●	●	●
In-Memory Compute Grid	●	●	●	●
In-Memory Service Grid	●	●	●	●
In-Memory Hadoop Acceleration	●	●	●	●
Distributed In-Memory File System	●	●	●	●
Advanced Clustering	●	●	●	●
Distributed Messaging	●	●	●	●
Distributed Events	●	●	●	●
Distributed Data Structures	●	●	●	●
JCache (JSR-107) Support	●	●	●	●
Auto-Loading of SQL Data	●	●	●	●
Dynamic Cache Creation	●	●	●	●
Java/.NET/C++ Interoperability	●	●	●	●
Portable Objects	●	●	●	●
Security Updates		●	●	●
Maintenance Releases, Patches, Support		●	●	●
Centralized Management & Monitoring (GUI)			●	●
Enterprise-Grade Security			●	●
Network Segmentation Protection			●	●
Rolling Production Updates			●	●
Data Center Replication			●	●
Centralized Backup and Recovery Management				●
Full, Incremental and Continuous Backups				●
Point-in-Time Recovery				●
Network Snapshots				●
Heterogeneous Recovery				●

Figure 13. Comparison of Apache Ignite and GridGain Editions

GridGain

SUMMARY

GridGain is an In-Memory Computing Platform that many companies use as shared data and processing infrastructure across projects to deliver in-memory speed and unlimited scalability for transactions, analytics, hybrid transactional and analytical processing (HTAP) and streaming analytics. It provides speed and scale by sliding between existing application and data layers as an in-memory data grid (IMDG) with no rip-and-replace of the underlying database. GridGain enables companies to deliver high volume, low latency transactions and analytics. It also simplifies streaming and analytics by acting as a shared data store and compute engine with real-time stream ingestion, processing, streaming analytics and continuous learning.

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 the form at www.gridgain.com/contact to have us 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.