

Apache Ignite and GridGain: Adding Speed, Scalability and In-Memory Computing to SQL

SQL is the universal language of data. It's not the only data language. But it is the most widely used language for transactions, analytics and other forms of data processing. SQL is the second most widely used developer language after JavaScript.

Over the last few years, however, new business demands—from digital transformation to improving the customer experience—have overwhelmed existing SQL infrastructure. The increase in interactions through new Web and mobile apps and their underlying APIs are creative massive volumes of queries and transactions that are overloading existing databases. Improving the customer experience requires performing realtime analytics and automation during transactions and interactions, not after. Traditional data warehouses and other related tools cannot address these needs because, by design, they work on a copy of the data that is almost immediately out of data as it is extracted, transformed and loaded (ETL) from operational systems. They also don't support the new analytical approaches, from stream processing to artificial intelligence, needed for these new initiatives. The good news is that several companies have successfully implemented these new approaches to real-time analytics with the GridGain® in-memory computing platform.

GRIDGAIN: ADDING SPEED, SCALABILITY AND IN-MEMORY COMPUTING TO SQL

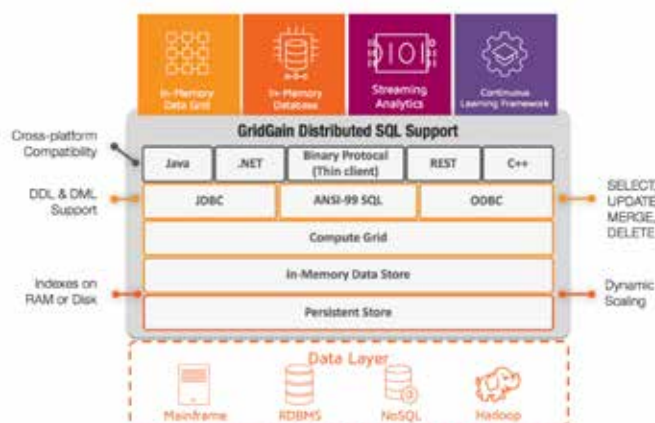
GridGain is the leading in-memory computing platform for real-time business. It is built on top of Apache® Ignite™, one of the top five Apache Software Foundation projects. GridGain Systems contributed the code that became Ignite to the Apache Software Foundation and continues to be the project's lead contributor. GridGain adds business-critical features to Ignite including enterprise-grade security, deployment, management and monitoring, and native SQL.

IMDG: OFFLOAD SQL QUERIES WITH UP TO 1000X LOWER LATENCY AND UNLIMITED HORIZONTAL SCALABILITY

Over the last decade companies have faced unprecedented challenges with speed and scalability. Companies have seen up to 1000x growth in SQL query and transaction volumes, and 50x growth in stored data during this time. They've also been asked to deliver near-real time responsiveness even as

they added more layers of systems and middleware that only increased latency.

GridGain slides in-between applications and databases as an IMDG with no rip-and-replace of the database or application. GridGain is able to accomplish this without major architectural changes because it supports ANSI-99 SQL and key-value as well as distributed 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 loads the underlying database schema and the data, and then handles the application's SQL requests via JDBC/ODBC or REST APIs. For all writes, GridGain commits transactions to the underlying database and then GridGain to ensures data consistency. GridGain then offloads all reads from the data-base, running distributed SQL in memory on its a shared nothing, scale-out architecture. The result is unlimited horizontal scalability and up to 1000x lower latency.



IMDB: SUPPORT ACID TRANSACTIONS WITH IN-MEMORY SPEED AND UNLIMITED HORIZONTAL SCALABILITY

GridGain also includes one of the few fully distributed SQL and key-value IMDBs. To use the IMDB features you can either turn on native persistence in an existing IMDG deployment and retire an underlying database, or you can create a new schema. GridGain supports:

- Distributed ANSI-99 SQL DDL and DML including SELECT, UPDATE, INSERT, MERGE and DELETE queries and CREATE and DROP table.

- Distributed SQL, including SQL collocated joins and aggregation with affinity-based partitioning to minimize network traffic.
- Distributed joins, including distributed cross-cache joins across third-party databases and the GridGain Persistent Store.
- The broadest range of in-memory distributed pessimistic, optimistic and deadlock-free ACID transactions.
- Memory-first, hybrid storage with 0-100% of data and SQL indexes in RAM, and the full data set and indexes in non-volatile storage for data durability and high availability, with all transactions first written to a write-ahead log (WAL) and then committed to disk.
- Data partitioning and replication across clusters and data centers to ensure high availability and minimize network traffic.
- Elastic scalability with automatic data rebalancing as nodes are added or removed.
- Immediate recovery on startup where any data or indexes are accessed from disk until they are loaded into RAM.
- Centralized backup and recovery, including continuous archiving, incremental and full snapshots and remote backup and recovery of any data to another cluster, on-premise or in the cloud, with a different size and topology.

GRIDGAIN: ADDING SPEED, SCALABILITY AND IN-MEMORY COMPUTING TO SQL

Once GridGain is used as an IMDG or IMDB, any of its data is immediately available for any other use; and any of its other features can be used to deliver analytics and automation during transactions or interactions. Gartner calls this approach hybrid transactional/analytical processing (HTAP). With GridGain, companies can add any new functionality or analytics using just about any language and implement HTAP. GridGain includes a compute grid that performs

real-time massively parallel processing (MPP) by distributing ANSI-99 SQL, geospatial queries, Java, .NET and C++ across a cluster, and then executing the code locally with the data in memory. This general-purpose in-memory computing enables companies to implement HTAP at any scale.

GridGain is also used to ingest, process, store and publish streaming data for large-scale, mission critical streaming 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. GridGain can ingest millions of events per second on a moderately-sized cluster. GridGain provides the broadest in-memory computing integration with Apache Spark. It can provide data to Spark using standard Spark RDD, DataFrame, HDFS and SQL APIs and collocate the data in memory with local Spark jobs across a cluster. When using DataFrames, GridGain can also accelerate SparkSQL up to 1000x by optimizing with GridGain's distributed SQL.

GridGain also provides integrated machine and deep learning as part of the GridGain Continuous Learning Framework. It enables companies to automate decisions by running machine and deep learning in real-time against petabytes of data. GridGain accomplishes this by distributing the algorithms to the data and running them locally in RAM 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 for deep learning. Developers can also implement and deploy their own algorithms across any cluster using the GridGain Compute Grid. The result is continuous learning, achieved by incrementally retraining against the latest data, to improve every decision and outcome.

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 go to 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.

© 2019 GridGain Systems. All rights reserved. This document is provided "as is". Information and views expressed in this document, including URL and other web site references, may change without notice. This document does not provide you with any legal rights to any intellectual property in any GridGain product. You may copy and use this document for your internal reference purposes. GridGain is a trademark or registered trademark of GridGain Systems, Inc. Apache, Apache Ignite, Ignite and the Apache Ignite logo are either registered trademarks or trademarks of the Apache Software Foundation in the United States and/or other countries. All other trademarks and trade names are the property of their respective owners and used here for identification purposes only.