



# Powering Financial Fraud Prevention with In-Memory Computing

## A GridGain Systems In-Memory Computing White Paper

January 2017

## Contents

Powering Financial Fraud Prevention with In-Memory Computing .....	2
Where Financial Fraud Occurs .....	2
Evolving Techniques for Fraud Prevention .....	3
Technologies Used for Fast Data Analysis .....	3
The Move to In-Memory Grid Computing: Faster, Better ROI.....	5
Financial Institutions Using In-Memory Computing .....	5
GridGain Systems: A Leader in In-Memory Computing .....	6
Meeting the Challenges of Real-Time Fraud Prevention .....	7
Contact GridGain Systems .....	8
About GridGain Systems .....	8

## Powering Financial Fraud Prevention with In-Memory Computing

Financial fraud is now a multi-billion-dollar business and growing rapidly, with Juniper Research predicting that online fraud alone will climb from \$10.7 billion in 2015 to 25.6 billion in 2020. Failure to detect and prevent fraud can harm the reputations of financial firms and reduce confidence in the industry as a whole.

Protecting their customers from fraud and protecting themselves from fraud-related losses are high priorities for financial institutions. However, fraud prevention is not a simple task, and firms must tackle it simultaneously with other crucial tasks such as ensuring regulatory compliance. To accomplish these data-intensive tasks in a timely manner, financial firms need solutions that are flexible, scalable, reliable, and fast enough to analyze extremely large datasets in real-time.

Fortunately, today's in-memory technologies provide powerful tools for combatting fraud – tools that perform complex processing, modeling, and analysis of big data in real-time. This white paper will discuss what financial fraud is, how firms are addressing the problem, and why in-memory computing technologies such as the GridGain in-memory computing platform are perfectly suited to the task of detecting and stopping fraud wherever it occurs.

### Where Financial Fraud Occurs

The term “financial fraud” encompasses a wide variety of illegal practices. Financial fraud typically shows up in the following arenas:

- **Checks** that are written by unauthorized institutions or officers
- **Credit cards** that are stolen and used illegally
- **Mortgages** that are illegally manipulated
- **Corporate financial statements** that are changed or illegally manipulated
- **Securities** that are traded using illegal techniques
- **Payments** that are requested fraudulently or rerouted to improper destinations
- **Identity theft** in which thieves steal financial information or impersonate others in order to make money
- **Forgery** of documents, signatures, banknotes, or works of art to produce financial gain
- **Computerized banking** and computer-based financial transactions employed improperly to produce financial gain
- **Tax evasion** in which corporations or people avoid paying taxes that they owe

The Financial Fraud Research Center at Stanford University estimates that Americans currently lose \$50 billion dollars a year to fraudulent practices such as these. The magnitude of this loss provides strong incentives for financial firms to find more effective techniques for fraud prevention.

## Evolving Techniques for Fraud Prevention

Traditional approaches to identifying fraud relied on manual verification and analysis. These approaches involved people auditing transactions directly, looking at who was providing passwords and other information. There were few apparent ways to automate this process. However, technology has progressed to the point where firms are now able to collect much more information and employ sophisticated techniques to automatically process and analyze data to detect fraud as it happens.

The techniques that banks and financial services firms currently use to detect fraud include the following:

- **Statistical and multi-channel analysis:** Calculating parameters (such as averages and performance metrics), linking together data from multiple sources (channels), and analyzing correlations between different data measurements to find patterns that help with fraud detection
- **Models and probability distributions:** Calculating models and probability distributions that predict how financial data will behave, so actual data can be compared against predictions and variations can be flagged as potential indications of fraud
- **User profiles:** Computing and maintaining user profiles associating personal data with specific users to help identify atypical behavior or attributes that may be fraudulent or indicate identify theft
- **Real-time algorithmic analysis:** Using algorithms to identify and validate user actions as they occur, in real-time
- **Data clustering and classification:** Analyzing known patterns and profiles and classifying them for use in algorithms and models – essentially creating a data repository for fraud detection
- **Artificial intelligence and machine learning:** Using machine learning techniques such as neural networks to refine automated fraud detection, reducing false positives (false alarms) and improving behavior-based predictions for current transactions and users

Performing these knowledge-intensive activities in real-time on extremely large datasets requires high performance and highly scalable technologies, as the next section discusses.

## Technologies Used for Fast Data Analysis

To apply the processing and analysis techniques needed for fraud detection in real-time or near real-time on a large scale, financial firms combine these techniques with technologies that can provide fast data analytics in a high intensity, transactional environment.

These technologies include the following:

**Big Data.** The first step in using financial data for fraud detection is to prepare data and make it available for analysis. Big data technologies provide ways to organize large datasets into multiple pools and connect them in real-time for immediate fraud detection and additional analysis.

**Apache™ Hadoop® with MapReduce.** Stopping fraudulent transactions in large datasets in real-time requires speed and efficiency. The average speed of executing a transaction may be only milliseconds, and within those single-digit milliseconds, the processing system must analyze the transaction, validate it, and check all available data pools without affecting the performance of processing the transaction. Hadoop with MapReduce is designed to help in situations exactly like these. It organizes hierarchical data to improve performance, allowing quick conclusions as to whether a transaction should be stopped.

**Complex Event Processing (CEP) with data streaming.** This technology, used in many financial institutions today, involves looking at multiple streams of incoming data and using artificial intelligence (AI) to identify meaningful events, such as potential fraud. It uses neural networks and other AI paradigms to decide how incoming data elements affect the behavior of the system as a whole as transactions are processed.

**Near real-time systems.** Trying to solve fraud issues after they occur is an expensive strategy and it poses risks to a company's reputation. A much better approach is to stop those transactions while they are happening. This approach requires extremely fast and efficient processing so financial firms are turning to near real-time systems.

**Data partitioning and parallel processing clusters.** When there are many transactions coming in at the same time, lining them up one by one to check them for fraud is not an option. To operate in real-time and maintain acceptable performance, the system must include multiple processors operating on the data simultaneously – that is, clusters of connected computers processing the data in parallel. It is also important to have the distribution available on the clusters to process those transactions regardless of where they occur, while maintaining data consistency. A system with data partitioning and parallel processing clusters is essential to meet these needs.

**Scalable data architecture.** We are operating in the world of constantly growing data. Large financial institutions are experiencing 20 to 30 percent data growth year over year, and they cannot risk running out of space. They must be able to add more storage while not losing performance, which means they need a scalable data architecture.

**In-memory computing.** Combatting fraud is an analytically intense process that uses performance-hungry models and it must be performed in the fastest possible way: using in-memory computing. Because in-memory computing involves keeping data in RAM for extremely fast access, with no disk-related slowdowns, it is faster than any other storage-based computing method.

In the next sections, we will discuss how in-memory solutions such as the GridGain in-memory computing platform have evolved to be fast, affordable, and comprehensive in their ability to combine all of the technologies listed above.

## The Move to In-Memory Grid Computing: Faster, Better ROI

For applications that require heavy analytics and real-time (or near real-time) transaction processing of hundreds or even millions of transactions per second, the market is now moving from disk to in-memory computing. The reasons for this trend involve both performance and Return On Investment (ROI).

**1000x Faster.** The move from disk to memory is a key factor in improving performance. However, simply moving to memory is not sufficient to guarantee the extremely high memory processing speeds needed at the enterprise level. Enterprise-level speed requires cluster computing, with multiple machines performing analyses at the same time, and parallel distribution of data. These capabilities are important for providing high availability, disaster recovery, and concurrency across systems – and they are all provided in the GridGain in-memory computing platform. Clients who have implemented the GridGain In-Memory Data Fabric to detect and prevent fraud in their transactions have found that they can process those transactions about 1000 times faster.

**10x ROI Improvement.** The cost of memory has dropped roughly 30% per year since the 1960s, so memory has become much more affordable in recent years. While it may still be slightly more expensive than disk, the performance is so much better that it improves ROI significantly. Clients who have implemented the GridGain in-memory computing platform have seen a tenfold or more improvement in their ROI.

With these substantial improvements in speed and ROI, it is not surprising that many financial institutions are turning to the GridGain in-memory computing platform for big-data applications such as fraud detection and prevention.

## Financial Institutions Using In-Memory Computing

Financial institutions use GridGain for a variety of fraud detection use cases involving high-volume transaction processing and big-data analytics, such as checking for compliance with anti-money-laundering (AML) and “know your customer” (KYC) regulations, looking for market manipulation, or monitoring other regulated areas. They are using complex event processing for real-time or near real-time customer views and analysis of positions, so they require ultra-low latency in real-time or near real-time data processing and analytics.

Banks who have implemented the GridGain in-memory computing platform – including Barclays, Citi, Sberbank, and others – are seeing a measurable difference at the transactional level. They no longer need to export the data to another system for analysis and approval (or disapproval) of the transaction. That model too often involved performance degradation and post-transaction processing delays, with

clients unable to complete transactions until all required steps were performed. In contrast, because the GridGain in-memory computing platform does most of the required computing in a distributed and in-memory fashion, it can process transactions with no noticeable slowdown to the clients. Because GridGain verifies the integrity of each transaction before allowing it to go through, the result is a safer environment for clients.

**Customer Case Study: Sberbank.** One of the most noteworthy GridGain Systems financial services customers is Sberbank, the largest bank in Russia and the third largest in Europe. The company had 130 million customers and had begun to struggle to validate transactions in real-time due to increasingly high volumes. Traffic was becoming too substantial, and traditional legacy systems were struggling to keep up with the transaction processing without slowing down transactions and creating user dissatisfaction. The need existed to be scalable and transact in real-time but this simply was not possible with legacy systems.

Sberbank analyzed more than ten potential solutions from vendors in the in-memory computing space and found that the GridGain in-memory computing platform provided the best performance results, allowing the bank to significantly improve performance. With GridGain, the company was able to generate one billion transactions per second in a test environment using only 10 Dell® blades with a combined memory of one terabyte. This system cost about \$25,000, which is a significant reduction compared to the days when using in-memory technology cost millions of dollars.

The GridGain in-memory computing platform also provided several other important capabilities that Sberbank needed, including machine-learning and analytics, scalability, ease of deployment, hardware independence of cluster components, and a rigorous level of transactional consistency. Of particular importance was the ability to conduct integrity checking and rollback on financial transactions. Sberbank could not find that level of consistency with other in-memory computing solutions.

In a [January 2016 article in RBC](#), Herman Gref, the CEO of Sberbank, said that the bank selected the GridGain Systems technology to build “a platform that will enable the bank to introduce new products within hours, not weeks.” He went on to state that the GridGain in-memory computing platform enables Sberbank to provide “unlimited performance and very high reliability” while being “much cheaper” than the technology used previously. Sberbank is using GridGain’s in-memory computing platform to implement capabilities such as “machine learning, flexible pricing, and artificial intelligence”, that could not be provided by the other vendors evaluated – a group that included Oracle®, IBM® and others.

## GridGain Systems: A Leader in In-Memory Computing

With companies facing tremendous data growth and the need for real-time and near real-time fraud detection, demand for the GridGain in-memory computing platform is growing dramatically. This comprehensive platform contains a complete feature set that surpasses the capabilities of mere in-memory databases, making it well suited to financial use cases involving machine learning, risk analysis, real-time analytics, complex event processing, and other capabilities oriented toward financial fraud prevention.

As a complete in-memory computing platform, GridGain helps 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 financial services companies to build less expensive systems that perform hundreds of times better. Sberbank, Barclay's and Citi realized such benefits with the GridGain in-memory computing platform.

Clients enjoy the following:

- **A 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, Compute Grid, SQL Grid, Streaming Grid and 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 so 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 Systems Support](#), available to [GridGain Professional Edition](#) and [GridGain Enterprise Edition](#) users, includes rolling updates, faster availability of all releases and patches, and 24/7 enterprise-level support.

## Meeting the Challenges of Real-Time Fraud Prevention

As financial institutions and other companies are inundated with ever-increasing amounts of data to process and analyze for potential fraud, they are looking for high performance and highly scalable ways

do so in real-time and near real-time in order to stop fraud before it affects their finances and reputations. Fortunately, in-memory computing solutions can now provide the level of performance and scale these companies need. The GridGain in-memory computing platform offers a scalable, comprehensive, and affordable solution – an elegant and efficient way to stop fraud in its tracks.

## Contact GridGain Systems

To learn more about how GridGain In-Memory Data Fabric can help your business, please email our sales team at [sales@gridgain.com](mailto:sales@gridgain.com), call us at +1 (650) 241-2281 (US) or +44 (0) 7775 835 770 (Europe), or complete our [contact form](#) to have us contact you.

## About GridGain Systems

GridGain Systems is revolutionizing real-time data access and processing by offering enterprise-grade in-memory computing solutions built on Apache Ignite. GridGain solutions are used by global enterprises in financial, software, ecommerce, retail, online business services, healthcare, telecom and other major sectors. GridGain solutions connect data stores (SQL, NoSQL, and Apache Hadoop) with cloud-scale applications and enable massive data throughput and ultra-low latencies across a scalable, distributed cluster of commodity servers. GridGain is the most comprehensive, enterprise-grade in-memory computing platform for high volume ACID transactions, real-time analytics, and hybrid transactional/analytical processing. For more information, visit [gridgain.com](http://gridgain.com).

### **COPYRIGHT AND TRADEMARK INFORMATION**

© 2017 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. Windows, .NET, Microsoft Azure, Azure, and C# are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. AWS and Amazon Web Services are either registered trademarks or trademarks of Amazon Web Services, Inc. JEE and Java are either registered trademarks or trademarks of SUN Microsystems and/or Oracle Corporation in the United States and/or other countries. Apache, Apache Ignite, Ignite, the Apache Ignite logo, Apache Spark, Spark, Apache Hadoop, Hadoop, Apache Cassandra, and Cassandra 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.