WHITE PAPER

GridGain

In-Memory Computing: Leading the Fast Data Revolution

A GridGain Systems In-Memory Computing White Paper

The in-memory computing industry stands on the cusp of a fast data revolution that will greatly change the way human beings live their lives. With middleware software that allows data to be stored in RAM across a cluster of

computers and processed in parallel, data processing is now becoming possible at an unprecedented speed and scale—with costs well within reach.

Even fairly recently, in-memory computing has often required very costly hardware to obtain reasonably competitive results. For example, combining an Oracle database with millions of dollars of hardware might provide 20 million transactions per second today. In contrast, the current version of GridGain software with about $25,000 worth of hardware (ten commodity blades, each configured with about a hundred gigabytes of RAM) can provide one billion transactions per second.

This performance level—previously unattainable by even the largest organizations with the deepest pockets and now easily within reach of small businesses—is only part of the reason in-memory computing is leading the fast data revolution.

- In addition to speed, today's in-memory computing features:
- Low cost at a time of rapidly increasing demand
- Scalability and availability
- The ability to drive convergence of other data technologies
- An Open Source framework
- Robust capabilities bypassing previous limitations
- Suitability for the needs of both modern and legacy application architectures

In this article I will discuss these aspects of in-memory computing in more detail and describe how the society of the future will depend on the capabilities it provides. Over the next few years, in-memory computing technology

will be an integral part of changes that are dramatically transforming the world.

LOW COST MEETS EXPLODING DEMAND

As both a concept and a technology, in-memory has been around in some form for at least six decades, offering superior speed and real-time control compared to non-volatile data storage.

For example, the Remington Rand Univac Scientific Computer in 1956 was basically an in-memory computer. It moved data from drum storage into core memory (consisting of magnetic rings connected with wires) to provide much faster data access. An advertisement for this computer called it "[the only] commercially available digital computer capable of real-time performance… the ideal system for… simulation and for online data reduction."

This text could just as well describe today's in-memory technology. But the Univac took up a whole room and cost more than many businesses and universities could afford.

Since then, the cost of the technology—now RAM instead of rings and wires—has come down dramatically, decreasing about 30% per year for decades as performance has surged. In the mid-1990s, RAM cost about

$35 per megabyte. Today, it's less than a penny per megabyte, making in-memory computing much more accessible and popular. With the ability to achieve a billion transactions per second for just $25K—an amount easily chargeable on a credit card—the economics are now within reach of the smallest companies.

At the same time, demand has increased greatly during the past decade, with the enormous expansion of Internet scale and data growth. In 2000, the Internet was just taking off. Virtually all applications were internal applications within corporations. A huge application might have 50,000 users and work with gigabytes of data.

Today, billions of people and devices are connected to applications. Computing systems process terabytes and petabytes of data—and these systems require an infrastructure that only in-memory computing can deliver. The need for in-memory computing is likely to increase even more as categories such as cloud computing, software as a service (SaaS), and the Internet of Things (IoT) heat up, and as mobile platforms continue to expand the reach of social networking, online retail, gaming, fintech, and other applications. All of these apps need fast data access and instant response across a very large group of people or devices and in-memory computing is driving the convergence of technologies for producing that fast response in a scalable manner.

SCALABLE AND AVAILABLE

The scalability of today's in-memory technology is just as important as its speed. Clearly, moving data into RAM provides much faster data access. However, it's also important to distribute the computing—that is, to use a large cluster of commodity hardware and distribute not only the data, but also the processing workload. Sending the compute out to the data minimizes the movement of that data. The distributed computing capabilities of today's in-memory technology can provide the scale needed for both today's and tomorrow's applications.

Availability is also an important characteristic of today's in-memory technology. This technology is being used for mission-critical applications, speeding up the core business applications required to serve large numbers of customers and fast decision making. The resultant need for high availability has led to technology that is erasing the difference between memory and storage. NVDIMM (Non-Volatile Dual In-line Memory Module), for example, allows RAM to retain data even when power is lost, enabling high-availability in-memory computing. And through the use of replication and commodity hardware, redundancy can be built into any application at a very low cost.

DRIVING CONVERGENCE OF DATA TECHNOLOGIES

In-memory computing is also driving convergence of previously separate data-storage and data-processing models. Data-storage models as distinct as Hadoop and RDBMS are coming together through paradigms such as key-value stores to enable more efficient data management within a distributed model.

In the realm of data processing, in-memory computing is enabling the combining of once-separate computing architectures for OLAP and OLTP into a new model called HTAP (Hybrid Transaction/Analytical Processing). HTAP allows analytical processing of operational data without affecting the performance of the operational systems—a crucial capability for real-time analysis of that's happening with sales and inventory and supply chains.

PART OF THE OPEN SOURCE REVOLUTION

Another aspect of today's in-memory computing technology that puts it at the center of today's convergence of fast data technologies is its Open Source framework. Most of the in-memory technologies today are either being Open Sourced or are Open Source already, and Open Source is the wave of the future.

Large enterprises that initially resisted Open Source are now embracing it because it lets them create their own roadmap of the products and technologies they want to use. Google and Facebook led this trend by dropping many of the commercial products they used and creating their own infrastructure of both hardware and software.

Gartner says that by 2020, 98% of IT organizations will be running Open Source software in their mission-critical applications. The Open Source framework of today's in-memory computing technology will make it a key part of this scenario.

BREAKING THROUGH PREVIOUS LIMITATIONS

"Recent developments in in-memory computing have not only expanded its capabilities, but also erased limitations that significantly restricted the technology's uses a few years ago.

The concern about data in memory disappearing during power outages is no longer an issue, thanks to the development of non-volatile memory forms such as NVDIMM.

Previous concerns about RAM being able to store only a limited amount of data are also becoming a non-issue. Recently, Amazon introduced their X1 Instance, which lets customers rent computing instances with two terabytes of RAM for just $13.34 per hour. At such pricing levels, the economics of renting large clusters containing virtually any amount of RAM are well within reach.

MEETING THE NEEDS OF MODERN APPS AND IT

With its recent advancements, in-memory technology is now well matched to the needs of modern applications and systems. Building for the future requires the speed and the scale that this technology delivers—as well as the efficiency it provides.

A large proportion of the applications now being created are built for the cloud and will never be run on premise – almost all SaaS offerings such as Salesforce and Workday are available only in the cloud. Even Oracle and SAP are moving their solutions to cloud-only offerings. These applications and the increasing number of cloud-based businesses will depend on in-memory computing for scalability.

IT departments tasked with driving innovation while also reducing cost will find in-memory computing well suited to both goals. As a technology, in-memory computing will drive innovation by maximizing the performance of existing infrastructures and enabling solutions that were previously impossible using traditional disk-based technology. In addition, by enabling the merging of technologies such as OLAP and OLTP, it will provide substantial reductions in cost—fueling innovation while also meeting IT cost mandates.

A FUTURE BUILT ON IN-MEMORY COMPUTING

Looking ahead at technological and societal changes predicted for the next few decades, it's clear that many will require a fast data infrastructure built around in-memory computing.

Here are some of the changes expected soon or already starting to happen:

Smart homes

With increasing use of automated locks, lighting systems, thermostats, and other components in our homes, we are starting to connect these components into the cloud. Soon, all our homes will be connected into smart cities.

Smart cities

Many cities are now replacing expensive, old-fashioned lighting with lampposts that have LED lights and can be remotely controlled. These lampposts are becoming communication hubs, fitted with sensors for a wide variety of purposes—from traffic control and weather monitoring to surveillance and gunshot triangulation. With this basic infrastructure in place, cities will be able to self-adapt. They will change traffic patterns and deploy resources to various neighborhoods based on real-time detection of events such as disasters, accidents, and fires.

Megacities

Today, 50% of the world's population lives in cities (compared to about 20% in the early 1900s). In 30 years, 70% of the world's population will live in cities that are extremely large. 1These megacities will need fast, massively scalable computing to automate their functions.

Driverless traffic-controlled roadways

As self-driving cars and adaptive traffic controls become the norm, there will need to be larger, overarching controls to make sure everything is working together—including pedestrians and bicyclists, who will be part of...

Biometric people networks

Eventually, wearable devices such as Fitbits will become small enough to be embedded in people's bodies. These devices will be able to monitor everyone's health, automatically deploy drugs for sick people, and connect to larger networks to provide a type of "sixth sense" about larger events. For example, they may warn us before we cross a street that a car is about to hit us. These capabilities will need to be enabled through the cloud and fast communication links.

Smart energy grids

Current energy grids were built on the idea of having a few power plant delivering energy to many households and businesses. Now, millions of households and businesses with solar panels are becoming their own power plants, both using energy and feeding it into the system. This development calls for smart energy grids that can monitor and instantly adapt to both incoming and outgoing energy.

Cashless society

Today, selling an app in the app store might mean a delay of 60 days to get paid. Wire transfers take days, thanks

to built-in delays from aging payment clearing and settlement systems in the middle of the process. As we move towards a cashless society and real-time payments, speed and scale will become a critical requirement for success.

AI-based surveillance and security

Today, the average American will be caught on a surveillance camera 75 times per day. In London, there is one surveillance camera for every 11 people. 2As the number of surveillance cameras grow worldwide, the only way to "watch" and react to all of the footage will be through real-time automation and AI – both highly dependent on in-memory speed and scale.

Cognitive systems

Today's digital assistabts, such as "iri, are going to be more like partners in another 10 or 15 years—much like the title character in the movie Her.

Internet of flying things

According to the FAA, there will be 2.7 million commercial drones in the air by 2020 (compared to about 5,000 planes currently in the air at any given time over the US). 3Adding in hobbyist drones, there will soon be 7 million such "flying things" in our skies, and in-memory computing will be the only way to monitor and control our airspace.

Platforms, not products

We are transitioning from an age of selling stand-alone products to one of selling platforms. In the late 1970s, people bought self-contained music players such as the Sony Walkman and used them to play cassette tapes that they owned. Now, when people buy mobile phones as entertainment and communication devices, they download apps and connect to the cloud to gain most of the functionality they want.

The phone has become a connection to the platform, and much of the profit comes from third parties selling capabilities on that platform. This "sticky platform" idea is also showing up in many other industries, including gaming (i.e. Microsoft Xbox and Sony PlayStation consoles both are part of richly functional online platforms), hospitality (Airbnb), and transportation (Uber, Lyft, and so on). All of these platforms are dependent on massive scale and the ability to process data and deliver results very quickly.

Jobs for machines

In the 1990s, growth in jobs classified as ""routine" (both "routine manual" and "routine cognitive") started to flatten out as mechanization and computers made humans less necessary for these jobs. Now, even many "non-routine" jobs—both manual and cognitive—may disappear, through such technologies as deep learning, robotics, and AI. The U.S. President's Economic Report for 2016 showed an 83% chance that, in the near future, jobs paying less than $20 an hour (about $40K per year for full-time work) will disappear. The report also showed a 31% chance that jobs paying under $40 an hour (about $80K per year) will disappear. Since 87% of American workers currently earn less than $80K per year, the loss of these jobs is likely to cause great societal upheaval and bring about some type of universal basic income. With computers and automation providing many necessities, people will be able to choose whether or not they want to work. Jobs will be for machines, and living will be for people.

WHAT THE FUTURE BRINGS

The future described here is not far away, and it's a future powered by in-memory computing. Delivering speed, scalability, high availability, Open Source flexibility, and the convergence of previously separate data storage and data processing models, in-memory computing is at the core of innovation today and will transform tomorrow's world.

Contact GridGain Systems

To learn more about how GridGain can help your business, please email our sales team at sales@gridgain.com, call us at +1 (650) 241-2281 (US) or +44 (0)208 610 0666 (Europe), or complete our contact 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.