insideBIGDATA

InsideBIGDATA Guide to

Hyperscaling Your SaaS Infrastructure

by Daniel D. Gutierrez

BROUGHT TO YOU BY GridGain

Introduction:
Performance and Scale Challenges of Today's SaaS Businesses

Software as a Service (SaaS) is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over the Internet. Instead of companies installing software on their own servers, known as the on premises distribution model, application software providers host the software in the cloud and charge customers according to the time they spend using the software, or based on a monthly or annual fee. SaaS is becoming increasingly popular, and as the industry develops, more and more companies are dropping older business models in favor of this rapidly evolving methodology.

Some of the benefits of the SaaS model are that it is easier to administrate, all users have the same version of the software because updates and patch management will be done automatically, and plus collaboration is a lot easier. In addition, SaaS serves to grant global accessibility, making remote work models easier, which reduces costs and improves work performance.

From productivity apps and CRM application suites to software services which manage cloud apps and deployments and even enable the creation of hybrid clouds, SaaS is extremely broad and runs the gamut.

Contents

Types of SaaS

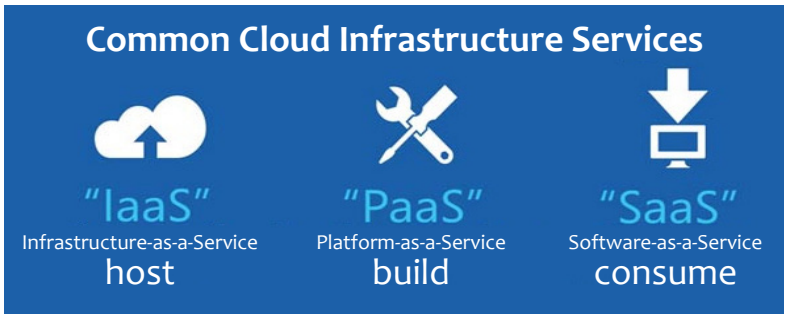
There are a number of different types of SaaS businesses and their consumption models of middleware/development tools software. The three most common cloud infrastructure services are: *Software as a Service* (SaaS), *Platform as a Service* (PaaS), and *Infrastructure as a Service* (IaaS). There is also a special type of PaaS called *private PaaS*.

SaaS represents the largest cloud market and is still growing quickly. SaaS uses the web to deliver applications that are managed by a third-party vendor and whose interface is accessed on the clients' side. Most SaaS applications can be run directly from a web browser without any downloads or installations required, although some require small plugins. Examples of well-known SaaS providers include Salesforce, Netsuite, Workday, Google Gmail, Microsoft 365, Citrix GoToMeeting, and Cisco WebEx.

PaaS is used for applications, and other development, while providing cloud components to software. Developers experience real gains with PaaS, specifically a framework they can build upon to develop or customize applications. PaaS makes the development, testing, and deployment of applications quick, simple, and cost-effective. Enterprise PaaS provides line-of-business software developers a self-service portal for managing computing infrastructure from centralized IT operations, the platforms installed on top of the hardware. Examples of PaaS providers include Amazon Web Service (AWS), Microsoft Azure, and Google Compute Engine (GCE).

The enterprise PaaS can be delivered through a hybrid model that uses both public IaaS and on-premise infrastructure or as a pure private PaaS that only uses the latter. A private PaaS enables enterprise developers to leverage all the benefits of a public PaaS to deploy, manage, and monitor applications, while meeting the security and privacy requirements an enterprise demands.

IaaS is a self-service model for accessing, monitoring, and managing remote data center infrastructures such as compute (virtualized or bare metal), storage, networking, and networking services (e.g. firewalls). Instead of having to purchase hardware outright, users can purchase IaaS based on a consumption model, similar to electricity or other utility billing. Examples of IaaS providers include IBM, AT&T, HP, Cisco Systems, Comcast, EMC, and Microsoft (Windows Azure Infrastructure Services).



Three common cloud infrastructure services: Iaas, Paas, and SaaS

SaaS Performance and Scalability

In order to be ready for any level of growth (which translates to business success), SaaS systems need to be able to *scale*, i.e. grow capacity on the back end to match customer demand without any major changes to the architecture of the application.

Scalability means something quite different in a SaaS environment than for traditional, behind-the-firewall software. Prasad Jogalekar and Murray Woodside's definition of SaaS scalability from their IEEE paper "Evaluating the Scalability of Distributed Systems" is very pertinent.

> "Scalability means not just the ability to operate, but to operate efficiently and with adequate quality of service, over the given range of configurations."

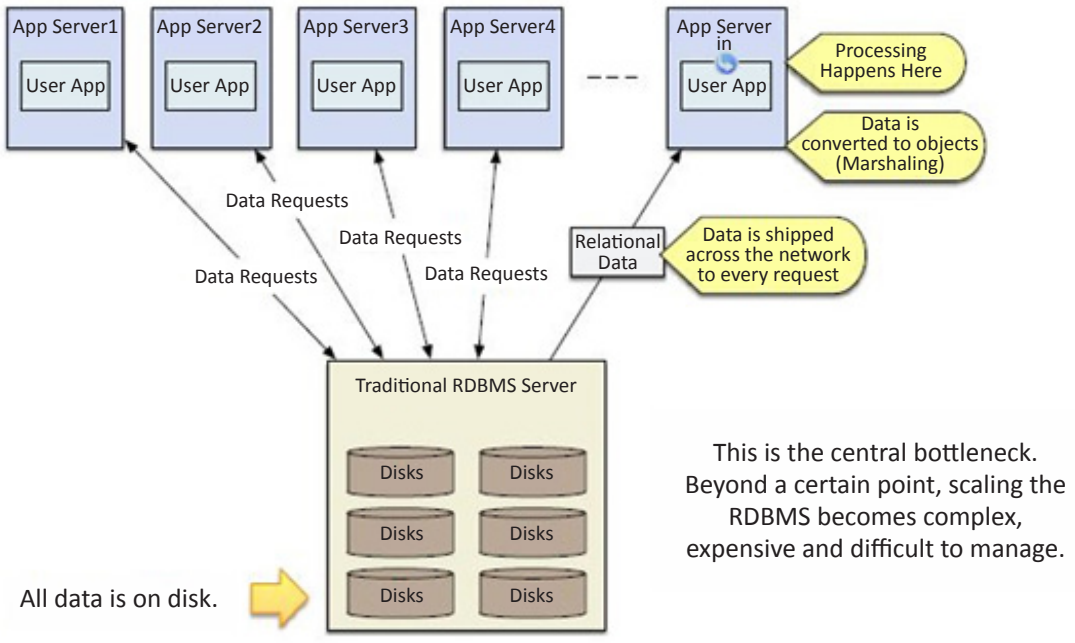
The typical SaaS business experiences challenges in building a high-performance, hyperscale service. *Hyperscale* computing is a distributed computing environment in which the volume of data and the demand for certain types of workloads can increase exponentially yet still be accommodated quickly in a cost-effective manner. In contrast to building a SaaS business 5-10 years ago, today you need to make sure the software can scale to millions of users. As a result, the growing demands of big data, software-as-a-service, mobile computing, Internet scaling along with falling RAM prices have spawned a range of new applications accessing and processing data in memory.

In-memory computing (IMC) is viewed as a solution to scalability issues for both "born-in-the-cloud" SaaS offerings, as well as ISV's converting their software into a SaaS offering. Clearly, not all SaaS companies are running in the Amazon cloud, and "elastic" from day one. Many companies built their own infrastructure—some running cloud VMs, some not. They typically add customers by adding more hardware, but when you get to a certain scale, this model is no longer adequate. There are many SaaS companies with similar themes for how their businesses evolved over time. They grew their architecture to a certain point by buying hardware, solving performance problems as they came up, but once at a certain level they need to entrust a more holistic approach toward performance and scalability.

SaaS companies just building out their own application-hosting environment need to figure out what cloud stack to adopt, e.g. the OpenStack cloud operating system that controls large pools of compute, storage, and networking resources. An important question to ask is whether you have the scale and performance needed to make your application a widely successful SaaS offering?

Some very compelling SaaS applications are now hitting the market, all with a common need for performance and scale. The goal of this guide is to educate SaaS companies in how IMC can make a big difference in the architecture of their technology. The guide shall be the authoritative cookbook to capitalize on IMC as the vehicle whenever anyone searches "How do I scale my SaaS business?" and endeavor to meet the SLA for customers.

Traditional Architecture



Traditional architecture where performance and scalability is addressed by adding hardware

In-Memory Computing for SaaS Applications

IMC is a clear winning technology for SaaS companies, and it's important to fully understand the nuances of the technology. For a deeper examination of IMC technology, please download the "insideBIGDATA Guide to In-Memory Computing" sponsored by GridGain. The guide includes: Introduction to IMC, The Business Case for IMC, Types of IMC, and Performance Benchmark.

One leading-edge industry using IMC to achieve top performance and scale is financial services where companies operate in a constant state of flux and where immediacy is critical. Historically, the financial services industry has some of the highest computing demands compared to other sectors due to the importance of low latency exchanges. In a very real and direct sense, speed translates to profit and latency translates to loss. As a result, the industry has been one of the earliest adopters of IMC technology, which offers data processing that is orders of magnitude faster and more scalable than traditional methods. Unlike disk-based processing, in which data is stored predominantly on disk and data is taken to the computation — a time-consuming, resource-intensive process — IMC stores the data mostly in the RAM of a cluster of computers and brings the computation to your data, yielding superior performance on commodity hardware.

Here is a small sample of how financial services firms currently use IMC:

- **Real-time trade reporting –** IMC provides the power necessary to remain competitive in trading, and any endeavor that relates to it, including support, modeling, testing, back-testing and re-pricing.
- **Post-trading processing for 24/7 global trading –** As trading has become almost universally global, financial services firms now must close their books daily. In order to achieve same-day reconciliation and straight-through processing (STP), companies rely on IMC technology to deliver the post-processing performance required. IMC is called upon to manage the extremely large amount of data at the required speed, enabling financial services firms to engage in 24/7 global trading.
- **Risk analysis –** In large part, financial services involve an exercise in risk management. Practically everything that a financial services company does, somehow directly relates to risk, as companies constantly must assess their exposure based on current strategies. The questions that must be asked are unfathomably complicated, and answering them in an actionable timeline and with accuracy requires the advanced computing capabilities that only IMC technology provides.
- **Fraud detection –** Financial institutions have been using in-memory computing for fraud detection for quite some time. The real-time processing demands for fraud detection mean that delays of seconds, even milliseconds, can't be tolerated. Acceptable performance requires real-time data access for ultra-fast processing, superior reactive response and proactive planning. Only IMC can provide that level of performance.

An IMC solution like GridGain's offers a broad pallet of capabilities that form a robust platform for SaaS companies.

For firms trying to meet the extreme demands of financial services computing, the question is not whether to leverage IMC technologies, but rather how to best leverage it. It is a good idea for SaaS companies to strive for those same advantages.

IMC appeals to SaaS companies due to the amount of performance and scale it can deliver. Furthermore, IMC fits well into a multi-tenant, cloud, SaaS environment. With the hybrid SaaS model, vendors offer both a hosted SaaS version of their application and an on-premise version. An IMC solution like the GridGain In-Memory Data Fabric can be deployed for either version and the deployment looks the same. For a SaaS company having customers who want their own data segregated and capacity dedicated to them, an IMC solution like GridGain's offers a broad pallet of capabilities that form a robust platform for SaaS companies.

Benefits to SaaS Companies when Adopting an In-Memory Data Fabric

In this section we'll explore the important benefits for SaaS companies when adopting an In-Memory Data Fabric.

Addressing one scenario, some SaaS companies having silo deployments and separate deployments per customer may be facing performance issues at a smaller scale which were solved adequately by some in-house solutions like doing application level caching. These same companies now may want to advance up to a SaaS infrastructure and start adding customers based on an on-demand model and scale out to hundreds or thousands of grid nodes — at that point no in-house solution would be acceptable. In this case, turning to in-memory solutions like in-memory data fabrics, and even data grids is a logical choice.

There's a need to go beyond the box in order to work on large data sets and perform computations in parallel on a cluster. This situation dictates the need to use IMC strategically (as with GridGain's In-Memory Data Fabric).

Many SaaS companies are aware of the benefits of caching, and some are using local application level caches. But these same companies now must act due to an expanding customer base, and the need to handle increasing data capacities. It is unlikely that all the data can be loaded into memory on one box, and even if it were possible, one box can't make calculations in a reasonable period of time. There's a need to go beyond the box in order to work on large data sets and perform computations in parallel on a cluster. This situation dictates the need to use IMC strategically (as with GridGain's In-Memory Data Fabric).

Many SaaS companies might be at a critical transition-point concerning scalability – how can the business move to the next generation environment? It turns out that an In-Memory Data Fabric is the logical choice as a technology solution to ease this transition.

At a high level, these companies build silos by customer or by functionality and now want to start breaking those down to have a common platform for everybody and eliminate data redundancy. An In-Memory Data Fabric provides the ability to add customers easily. For many SaaS companies, adding new customers is a new installation — new boxes, new software. If they move to the cloud, adding a new customer might be as simple as creating a new account on the system, or if the system has too many customers, then it might involve starting a new image on Amazon, or a new image on VMware.

One important driver for moving from silo deployments to the SaaS model is the ability to quickly grow and quickly add customers. Products like the GridGain In-Memory Data Fabric, with scale-out capabilities, enable this kind of growth. Many SaaS companies might be at a critical transition-point concerning scalability — how can the business move to the next generation environment? It turns out that an In-Memory Data Fabric is the logical choice as a technology solution to ease this transition.

Advantages of an In-Memory Data Fabric over Alternative Technologies

There are a number of choices available to a SaaS company that are designed to address expanded performance and scalability requirements, and it is important to carefully examine each option to ensure that the selected solution is going to provide a long-term path toward viability in a rapidly expanding customer adoption climate. Here is a short list of available alternatives:

- Application level caching.
- NoSQL (Not only SQL) solutions that are non-relational, distributed, open source, and horizontally scalable. Beginning in 2009, the data structures used by NoSQL databases (e.g. key-value, graph, or document) differ from those used in relational databases, making some operations faster with NoSQL. NoSQL databases are increasingly used in big data and real-time web applications.
- In-memory SQL databases are database management systems that primarily rely on main memory for data storage rather than a disk storage system. In-memory databases are faster than disk optimized databases since they eliminate the seek time when querying the data.

Even after considering the above technologies, SaaS vendors are wisely asking themselves what's beyond the traditional relational database — and may arrive at the conclusion that they should use application level caching, or an in-memory database. However, a concrete understanding of the pros and cons is required.

Application level caching to improve performance is often deployed as a stop-gap measure before considering an option with more long ranging effects. Application performance and throughput can be improved in many ways by incorporating a caching strategy within the product. But ultimately, this is a limited approach and the issue of performance and scalability will need to be revisited at some point.

An important realization about the in-memory database route is that it requires replacing an existing enterprise-class relational database like Oracle or SQL Server — a challenging proposition indeed. Teams of DBAs would be out of work, resulting in significant pushback for such an initiative. The choice of new technology should be the path of least resistance, but replacing an enterprise database is the path of "most" resistance.

An in-memory data fabric like GridGain's allows the company to keep their database since it is a software layer that sits between the application and the existing database. Companies can keep their reporting tools and other tools, and they can also decouple the performance of the application from the performance level of those tools without having to replace them.

Other IMC vendors might have several different products that can be cobbled together. GridGain, on the other hand, has a single product with one set of APIs, one set of docs, and is just one vendor relationship to manage. GridGain offers one integrated solution that allows the vendor's application to pull data in memory for faster access, be able to do compute in a distributed fashion against the data, and do SQL queries across diverse data sets including structured, semi-structured and unstructured data.

Load	• How is your web application responding to different simulated loads?
Stress	• How stable is your web application beyond the normal load conditions?
Scalability	• How long can your web application scale up and scale out in terms of user loads and transactions?
Reliability	• How long can your web application sustain good response levels under expected load conditions?
Volume	• How is your web application responding under different volumes of data?

Important questions surrounding performance and scale for any SaaS company

Customer Decision Points for Selecting IMC

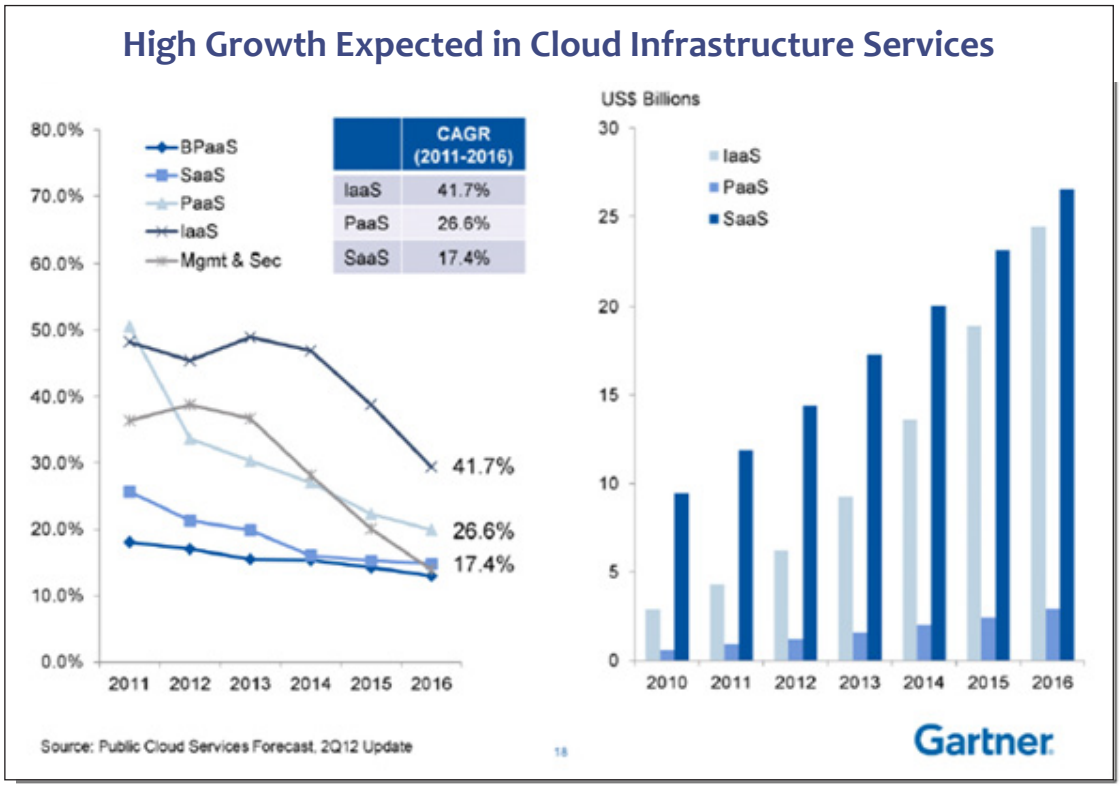
As the SaaS industry continues to evolve at a rapid pace, the technology upon which businesses in this space are based requires an increasing degree of performance and scalability in order to address the needs of a growing customer base. Many of these companies are choosing IMC to resolve these mission critical requirements. In this section, we'll focus on common reasons IMC is chosen for SaaS enablement. We'll do this by identifying the needs of specific market segments within the SaaS community to see how IMC technology can solve key requirements for continued growth.

As one example, consider a SaaS company in the financial services industry, and the pressures it has in supporting a growing user base with significant demands for performance and scalability. The provider could look at individual IMC solutions both on the open source and commercial front, but ultimately the benefit of finding many of the solutions it needs all under one roof, helped to refine their selection. In many cases, companies like this already know IMC was the direction in which they should be headed.

At the same time, another SaaS company offering services for the transportation industry may already be on the IMC track, possibly using local application caching technology as a first step, but without distributed grid technology. As before, companies like this already know IMC was a proven strategic direction.

Another class of online business might involve a new orientation of social media application where the goal is ephemeral messaging. These applications have a real need to work entirely in memory; by their very nature they don't want to put anything on disk. These application providers were cloud based from the beginning, but need a solution that had robust capabilities around data management and consistency.

The bottom line is IMC is not just for high-end applications like financial services. Even applications from traditional SaaS vendors, in very different verticals, can engage IMC and see considerable gains. The specific industry of a SaaS company doesn't matter when trying to build a SaaS application — IMC is a respected architecture that should be given serious consideration.



Compound Annual Growth Rate (CAGR) and rising revenue for cloud infrastructure services including SaaS presents an opportunity for IMC to help sustain continued growth

GridGain Solutions for SaaS Enablement

This guide has presented a number of technical and business considerations for SaaS companies facing the need to embrace new solutions for enhanced performance and scalability. In this section, we'll tie these ideas together by focusing on the *GridGain In-Memory Data Fabric* and how its breadth of features represent a sound foundation for building a robust SaaS architecture. The GridGain In-Memory Data Fabric is built on top of an open source, incubating Apache project — Apache Ignite™ (incubating) — and is designed as an enterprise-grade data access and processing solution on-premises and in the cloud for today's world of Fast Data.

Running in any public, private or hybrid cloud environment, the GridGain In-Memory Data Fabric enables definitive performance and scale for any Java, .Net or C++ application, including the most mission-critical workloads. Whether you are a "born-in-the-cloud" SaaS provider determined to accelerate time to market for your next hyperscale application, an ISV looking to add "as a service" capabilities to your existing software solutions, or an enterprise delivering internal, on-premises software as a service to your growing user base, chances are you'll have the following concerns:

- Quickly and seamlessly adding users to your application
- Meeting or exceeding aggressive SLA provisions
- Improving application performance by orders of magnitude
- Seamlessly scaling up or scaling out on cost-effective commodity hardware
- Providing secure, multi-tenant access to your data
- Minimizing the amount of rework to application code and databases you are already invested in

GridGain In-Memory Fabric

The GridGain In-Memory Data Fabric provides a unique feature set designed to help you migrate legacy software or build brand new cloud applications for a SaaS world that delivers highest performance, including for mission-critical or real-time applications, and ultimate scalability:

- **Highly distributed architecture** designed from the ground up for in-memory processing and/or storing of data, enabling any application to rapidly scale up and support dramatic increases in number of users and/or transactions handled.
- **100x higher throughput and/or lower latencies** for high-performance applications with native support for structured, semi-structured or unstructured data sources, avoiding risky and costly "rip-and-replace" of existing RDBMS, NoSQL or Hadoop infrastructures, and allowing to freely cross-query between these data sets using standard SQL syntax.
- **Built-in security and multi-tenancy,** ensuring that multiple tenants or applications can run independently and highly securely on the GridGain In-Memory Data Fabric while efficiently sharing precious high-performance resources.
- **Extensive cross-platform support** for C++, Java, Scala, .NET and REST, ensuring that the vast majority of your new or existing applications can be easily SaaS-enabled with minimal amount of work.

PaaS for SaaS Providers

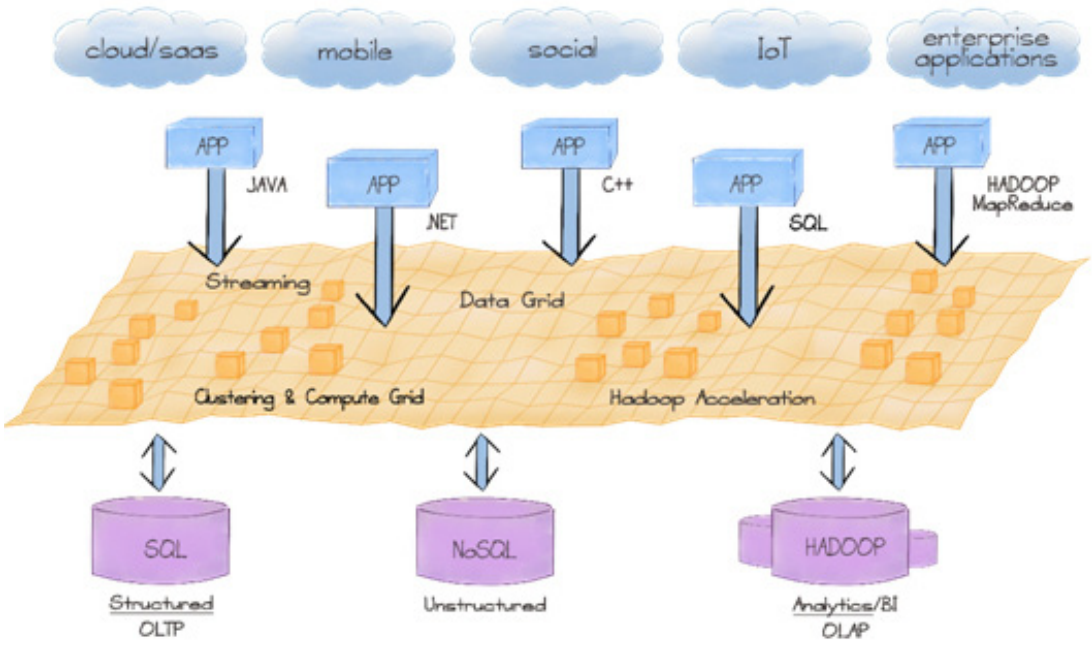
The GridGain In-Memory Data Fabric is a proven software solution which delivers unprecedented speed and unlimited scale to accelerate the growth of a SaaS business and enables high-performance transactions, real-time streaming and fast analytics in a single, comprehensive data access and processing layer. Here is a short list of motivations for choosing GridGain for your SaaS business:

- If your plan is to run your high-performance SaaS business in your own cloud infrastructure, or if you are looking for ways to optimize performance and reach of your SaaS offerings on the infrastructure of your cloud provider of choice.
- If you are a cloud provider looking to include PaaS capabilities that allow your customers to rapidly migrate or build their high-performance, hyperscale SaaS offerings to your platform
- If you need an easy, flexible and virtually unlimited way to scale different types of analytical, transactional or hybrid applications in a multi-tenant environment.

SaaS for the Real-Time Enterprise

For IT organizations wishing to transform into a service provider for their enterprise, creating the ability to rapidly and flexibly deliver internal software as a service to critical lines of business, the GridGain In-Memory Data Fabric offers a strategic approach to in-memory computing that delivers performance, scale and comprehensive capabilities far above and beyond traditional disk-based or even in-memory-enhanced databases, data grids or other in-memory-based point solutions. It offers a secure, highly available and manageable data environment that allows companies large and small to process full ACID transactions and generate valuable insights from real-time, interactive and batch queries.

GridGain In-Memory Data Fabric



The GridGain In-Memory Data Fabric architecture is uniquely able to support a growing SaaS business

DATA FABRIC FEATURE: In-Memory Data Grid

With its In-Memory Data Fabric, GridGain offers industry leading data grid functionality characterized by the fact that data are stored in-memory as opposed to traditional DBMS software that utilizes disk as the primary storage mechanism. By utilizing system memory rather than spinning disk, data grids are typically orders of magnitude faster than traditional DBMS systems. The GridGain data grid feature supports standard SQL for querying in-memory data including support for distributed SQL joins.

As streaming data never ends, an application must be able to provide a size limit or a time boundary on how far back each request or each query should go.

In direct response to the typical SaaS company's requirements for unlimited growth, GridGain's data grid feature contains an impressive feature set including advanced security, fault tolerance, topology resolutions, load balancing, collision resolutions, connected jobs, local node storage and much more. In a clustered in-memory solution like GridGain's, the collection of all individual node memory can be used as a single, expansive "fabric" of virtually connected memory. Large data sets can be effectively partitioned across all nodes for high-end scalability, and computations can be intelligently parallelized for optimal processing speed.

At a fundamental level, GridGain enables promoting data up from residing in slow mechanical storage systems to fast memory. GridGain's data grid feature solves many critical SaaS pain-points at once:

- Performance
- Scalability
- High availability
- Data consistency and reliability
- Detailed insight and management

DATA FABRIC FEATURE: In-Memory Compute Grid

Besides the data grid capability, the GridGain In-Memory Data Fabric also includes an In-Memory Compute Grid, which provides the means for parallel, in-memory processing of CPU-intensive or other resource-intensive tasks, including traditional High Performance Computing (HPC) and Massively Parallel Processing (MPP).

DATA FABRIC FEATURE: Real-time Streaming

To address the needs of many Saas applications for which traditional processing methods and disk-based storages, like databases or file systems, fall short — the GridGain In-Memory Data Fabric offers stream-processing capabilities. In-memory streaming combines both event workflow and *Complex Event Processing* (CEP) capabilities fully integrated in one product.

Processing of market feeds, electronic trading by many financial companies on Wall Street, security and fraud detection, real-time sales lead management — all these applications produce large amounts of data at very fast rates and require appropriate infrastructure capable of processing data in real-time without blockages.

One of the most common use cases for stream processing is the ability to control and properly pipeline distributed events workflow. As events are coming into the system at high rates, the processing of events is split into multiple stages and each stage has to be properly routed within a cluster for processing.

One of the key features of many CEP systems is the ability to control the scope of operations on streamed data. As streaming data never ends, an application must be able to provide a size limit or a time boundary on how far back each request or each query should go.

DATA FABRIC FEATURE: Hadoop Acceleration

With its In-Memory Data Fabric, GridGain offers Hadoop acceleration as well as a standalone In-Memory Accelerator for Hadoop built on top of the In-Memory Data Fabric, which expand the benefits of IMC to the Hadoop world by enabling enterprises to achieve unmatched performance and scale with their existing MapReduce and HIVE applications. All this is possible without requiring any code change to the native MapReduce, HDFS and YARN environment. This result is especially attractive for SaaS providers running analytics in a Hadoop distribution, a major milestone and turning point for Hadoop in the past year.

Prior to this offering, running IMC in an existing Hadoop environment required code changes to the application, reducing organization's ability to quickly derive the full performance benefits of an in-memory architecture. The In-Memory Accelerator for Hadoop allows for true plug and play deployment, meaning that within minutes of download, developers can deliver up to 10x performance improvement on their MapReduce applications.

GridGain's Hadoop acceleration is based on dual-mode, high-performance in-memory file system that is 100% compatible with Hadoop HDFS — and an in-memory optimized MapReduce implementation. GridGain's in-memory MapReduce effectively parallelizes the processing of in-memory data stored in GGFS. It eliminates the overhead associated with job tracker and task trackers in a standard Hadoop architecture while providing low-latency, HPC-style distributed processing.

By enabling SaaS companies to more readily leverage in-memory performance at scale for their Hadoop clusters, GridGain is extending the benefits of its proven IMC platform to a larger enterprise community.

Summary

In this Guide we have made the case for enabling SaaS companies to meet hyperscale requirements by taking advantage of IMC technology. The needs of the SaaS software distribution model are unique from other classes of software and the Guide has focused on the benefits for adopting IMC as a solution to these needs.

The Guide includes the following definitions, pain-points, and solutions centered on the coupling of SaaS and IMC technologies:

- **Performance and scale challenges of today's SaaS Businesses –** a discussion of the types of SaaS platforms — SaaS, PaaS, and IaaS, SaaS performance and scalability requirements, and the drawbacks of traditional SaaS architecture used to address scalability by adding hardware.
- **IMC for SaaS applications –** showing how the financial services industry has successfully adopted IMC, how IMC fits well into a multi-tenant, cloud, and SaaS environment, how vendors using the hybrid SaaS model to offer both a hosted SaaS version of their application and an on-premises version can use an IMC solution deployed for either version with the deployment looking the same.
- **Benefits to SaaS companies when adopting IMC –** growing SaaS companies need solutions beyond application level caches where new technologies like IMC can help evolve past the commonly used silo approach in order to quickly grow and quickly add customers.
- **Advantages of in-memory data fabrics over alternative technologies –** in-memory data fabrics offer distinct advantages over competing technologies like NoSQL and in-memory databases to address the needs of increased performance and scalability. A solution like GridGain's won't force you to replace your database. Rather than a collection of point solutions, the GridGain In-Memory Data Fabric is a cohesive product with a broad feature-set.
- **Customer decision points for selecting IMC –** a focus on common reasons IMC is chosen for SaaS enablement.
- **GridGain solutions for SaaS enablement –** overview of the industry's premiere IMC solution to address the needs of growing SaaS businesses.