Second-Generation Cloud Computing – IaaS Services

What it means, and why we need it now

Prepared for:

[ProfitBricks Logo]

The IaaS-Company.
Cloud Computing: Driving IT Forward

Cloud computing, particularly in the form of public infrastructure-as-a-service (IaaS), represents a major revolution in the IT industry. Of course IaaS is not just about changing or even revolutionizing the deployment, management, delivery and consumption of IT. It is about providing greater benefits to the organizations that rely on IT.

Benefits of IaaS

In just a few years, IaaS has become an essential component of the IT mix. Web-scale applications, software-as-a-service (SaaS) offerings and big data applications all depend on IaaS. Businesses, government entities and non-profits also use IaaS in ways commonly associated with traditional, in-house IT. These include resource sharing, test and development environments, production applications and sometimes even mission critical applications.

According to Neovise research and the results of many other studies, the most cited benefit by organizations using IaaS is agility, which helps them move faster and adapt to change more easily. While cost savings are not guaranteed, IaaS can often be used more efficiently than traditional IT by adjusting resource supply to match demand and by turning off services that are not in use. Organizations that use IaaS to innovate and improve their own business or operating models may have the most to gain.

How IaaS Changes Traditional IT

IaaS transfers the responsibilities of infrastructure ownership – including purchase, maintenance, support and repair as well as power, space and cooling – to a service provider. It also transforms the raw physical resources into easily consumable services that require little or no human involvement by the service provider. This much of IaaS is similar to other IT hosting models such as dedicated hosting. Yet IaaS has evolved much further, bringing additional changes in economics, ease of use, business models and technologies with it.

IaaS also changes the way IT is deployed and managed. Instead of relying on formal requests for new server, storage and network gear – and waiting for approval, shipment and set up – IaaS delivers rapid access to on-demand infrastructure through self-service interfaces and APIs. Rather than depending on multiple work orders – carried out by specialists in network engineering, server administration and storage management teams – IaaS automates administrative and operational tasks, cutting delivery times from weeks to minutes.
IaaS significantly changes the way IT is delivered and consumed. Instead of relying simply on virtualization and consolidation of physical infrastructure, IaaS pools resources and shares them between multiple tenants, enabling massive capacity for web-scale applications. Rather than cloud users committing to a fixed level of capacity on a long-term basis, IaaS offers rapid elasticity, letting users quickly adjust their level of consumption as their requirements change.

Where First-Generation Clouds Fall Short

Public IaaS is clearly a major step forward for the IT industry. Yet, as good as it already is, IaaS must get even better. Consider the shortcomings of first-generation IaaS in these categories:

**Scale**

IaaS improves scale-out computing by allowing applications to use hundreds or thousands of servers simultaneously. Unfortunately, most existing applications require design and coding changes to leverage this type of scale-out architecture. Traditional applications generally rely on scale-up architectures, yet first-generation clouds don’t support dynamic scale up of running instances. The workaround is to shut down the instance and start a new one with increased RAM and CPU. Not exactly convenient – or even feasible – at the time demand spikes. The alternative is to over-provision by configuring instances to meet peak resources requirements. The result is lower average resource utilization and higher costs.

**Performance**

Performance is often tied to scalability. While first-generation clouds excel at achieving performance through scale-out architectures, most relational databases must scale up in order to improve performance. This results, once again, in the problems typically associated with over-provisioning. First-generation clouds also tend to struggle with the level of resource sharing involved with multi-tenancy, resulting in inconsistent performance. Multiple cloud users not only share the same network, they share the same CPU cores and RAM. Noisy neighbors, or those instances consuming unexpectedly high amounts of resources, impact the consistency of performance experienced by other instances.

**Flexibility**

While first-generation IaaS is a step forward for infrastructure automation and provisioning speed, it is a step backward for flexibility. Most IaaS providers lack support for instance customization, instead relying on a fixed set of configurations from which to choose. First-generation cloud networks are even less flexible, offering little control over topologies, sub-nets, forwarding rules and IP addresses.
Integration

Other first-generation IaaS limitations are best understood from the perspective of an entire integrated data center. Cloud customers do not just seek compute power with local storage and an Internet connection. They also deploy complex, multi-tier applications that use a variety of resources across the data center, including shared storage, firewalls, load balancers and complex multi-tier networks. Unfortunately, most existing clouds present resources as a loose set of independent components, rather than as a cohesive virtual data center in which complex relationships can be modeled and deployed.

Usability

First-generation clouds offer self-service portals and APIs, providing far superior automation compared to most traditional IT environments. However, usability has made little progress – particularly when it comes to large, complex configurations. Writing code to manage IT through a cloud API is the antithesis of usability. Self-service portals for public clouds are a step in the right direction. Yet most of those fall short when used to manage large or complex IT environments due to their lack of graphical interfaces.

Components

Some first-generation limitations are best understood from the perspective of just one of the core infrastructure components: servers, networks or storage. For instance, many cloud networks suffer from inconsistent performance and may even throttle high-bandwidth users when they impact the quality of service (QoS) of other customers. Similarly, some storage area networks (SAN) used in cloud environments deliver inconsistent input/output operations per second (IOPS) and introduce unacceptable latency.

While not every first-generation IaaS offering suffers from all these shortcomings, it is time for a new class of IaaS to address these challenges holistically.

Introducing Second-Generation IaaS

Second-generation IaaS offerings generally begin with two basic advantages. First, Moore's Law continues to drives ever-improving performance at declining prices. This gives IaaS 2.0 providers access to better, faster and cheaper components. Second, IaaS 2.0 offerings are not constrained by first-generation architectures. This enables greater innovation by second-generation IaaS providers, allowing them to overcome many IaaS 1.0 challenges. New IaaS entrants are already leveraging both of these advantages.
Keep in mind that simply throwing faster hardware at any IT challenge is rarely the best solution. Success with IaaS 2.0 requires more than assembling a bunch of hardware with incremental improvements. It requires a cohesive strategy with better components that deliver even better value when used together.

What should second-generation IaaS look like? First and foremost, second-generation IaaS should address the gaps found in first-generation offerings. While not every gap must be addressed, it is important that major improvements across several gaps be made. IaaS 2.0 offerings should represent cohesive, well thought out solutions that advance the state-of-the-art in IaaS. Given the extremely competitive nature of this space, we should see a variety of approaches and plenty of innovation in second-generation offerings, including adjustments in business models.

These are just some of the important improvements that IaaS 2.0 offerings may include:

**Scale**
- Vertical scaling in addition to horizontal scaling
- Dynamic, on-the-fly instance scaling – without shutdown and restart

**Performance**
- Servers – more cores, more RAM, lower energy, optimized virtualization layers
- Storage – guaranteed IOPs, greater capacity, more control over service levels
- Network – higher throughput, lower latency, more consistent, greater isolation

**Price/Performance**
- High performance with better economics (beyond peak and consistent performance)
- Improved utilization and/or pay for actual usage

**Flexibility**
- User-defined network topologies and behaviors
- Fine-grained customization of resource attributes, configurations and service levels

**Usability**
- Graphical interfaces for designing and provisioning infrastructure and applications
- Advanced configuration management capabilities and tools

**Integration**
- End-to-end virtual data center perspective rather than just component perspectives
- Improved options for more seamless hybrid clouds

“IaaS 2.0 offerings should represent cohesive, well thought out solutions that advance the state-of-the-art in IaaS.”
ProfitBricks: A Second-Generation IaaS

As suggested earlier, it takes innovation and R&D, not just new features to deliver IaaS 2.0. The ProfitBricks founders began with this vision and have relied on more than 75 engineers and two years of development in the US and Germany to deliver it. In the process, the company has overcome a wide range of IaaS 1.0 shortcomings, leapfrogged many early competitors and produced a remarkable IaaS 2.0 solution.

"The company has overcome a wide range of IaaS 1.0 shortcomings, leapfrogged many early competitors and produced a remarkable IaaS 2.0 solution."

At the foundation of the ProfitBricks cloud is an InfiniBand network – the only one of its kind among public IaaS clouds – that delivers 80 Gbps of throughput per server. In fact, all hardware in the ProfitBricks infrastructure is equipped with dual 40Gbps InfiniBand network interface cards (NIC) and is deployed in a meshed network topology so that potential failure of any NIC, switch or router has minimal impact.

The network includes an Ethernet overlay, giving customers the performance benefits of InfiniBand while supporting all protocols that run over Ethernet. It is also architected using software-defined networking (SDN). This enables users to design and deploy their networks without the constraints imposed by most cloud networks. Hardware, rather than virtual switches, are used throughout the network, introducing only 200 nanoseconds of latency per hop.

ProfitBricks takes additional steps to help ensure customers get the performance for which they pay. Not only is network traffic isolated, the company guarantees the exclusive use of its server hardware by individual customers. Since CPU cores and RAM are not shared among multiple customers, servers are not over-subscribed and virtual server performance remains consistent.

ProfitBricks is also leading on the scalability dimension by supporting vertical scaling with on-the-fly CPU and RAM elasticity. Customers can start with one core and dynamically scale that up to 62 cores. They can also start with 1 GB of RAM and scale that all the way to 196GB – all without shutting down or restarting their instance. While the company has not yet deployed solid state drives (SSD), its shared storage solution is far from mundane. It is built with replicated Raid10 storage in two separate data center facilities, providing multiple levels of redundancy. It is also connected via InfiniBand, providing high IOPS without network constraints.

ProfitBricks has also innovated on the IaaS 1.0 business model by moving from hourly pricing to by-the-minute pricing, further enhancing the economics of per-per-use pricing. In fact, per-
minute pricing also works well with the dynamic scaling capability in the ProfitBricks cloud. Together, these capabilities let customers fine tune resource consumption and adjust corresponding spending in short time intervals.

The ProfitBricks cloud includes a unique tool called Data Center Designer (DCD) that lets customers build an entire virtual data center through a simple, intuitive drag and drop interface. Rather than simply provisioning individual infrastructure components, DCD lets users design and provision complex, end-to-end data center environments within minutes. The DCD also makes cloud computing data center design more like traditional data center design, with all of the infrastructure components and their network connections being visible and simple to modify.

The Time is Now for Second-Generation IaaS

Cloud customers are ready for second-generation IaaS now. But don’t worry about whether it is called IaaS 2.0, second-generation IaaS or something completely different. Second-generation IaaS is really about the value it provides and the challenges it overcomes relative to first-generation offerings. The ProfitBricks cloud delivers outstanding value and resolves a long list of the challenges faced by first-generation offerings.

ProfitBricks already supports more than 1,000 live customers and is an outstanding fit for customers that require high performance. These include companies handling massive numbers of database lookups, SaaS and software vendors, gaming companies and big data users. It is also a great fit for traditional applications that are not designed to run in typical cloud environments where custom sized CPU and RAM specifications are needed, horizontal scale is insufficient and/or the network requires a complex topology.

The ProfitBricks cloud is not overpowered or overpriced. In fact, it provides leading price/performance ratios and makes perfect sense for test and development environments as well as high-performance compute environments and production applications. ProfitBricks uses the phrase “Cloud Computing Accelerated.” For customers that means more than just raw performance – also indicating the speed and ease of use it provides when deploying data centers in the cloud. Given its flexibility, unique strengths and overall value, Neovise strongly recommends that all prospective IaaS customers put ProfitBricks on their short list.

“Neovise strongly recommends that all prospective IaaS customers put ProfitBricks on their short list.”
About ProfitBricks

ProfitBricks - Cloud Computing 2.0 - is the cloud infrastructure as a service (IaaS) company offering more speed and flexibility than any other cloud provider. Founded in 2010 by the previous co-founders of 1&1 AG, ProfitBricks has built the world's first, true virtual datacenter, enabling users custom defined instances with live vertical scaling and class-leading double redundant cloud storage – all with simple and transparent minute-based billing. It also developed the first graphical Data Center Designer that makes the ProfitBricks Cloud Computing service the easiest to setup and maintain. CRN Magazine recently picked ProfitBricks as the "Coolest Startup" of 2012. Headquartered in Berlin, Germany and Boston, Mass., ProfitBricks can be found online at http://www.ProfitBricks.com.

About Neovise

Based on independent research and analysis, Neovise delivers essential knowledge and guidance to cloud-related technology vendors, service providers and systems integrators, as well as business and IT organizations that purchase and use cloud-related services and technology. Our offerings include research, advisory and collateral development services that help our customers—and their customers—make optimal decisions and formulate winning strategies. Research. Analyze. Neovise.

For more information, visit www.neovise.com.