Joyent

The Joyent Smart Technologies Architecture for Cloud Computing

A Joyent White Paper

Executive Summary

The term *cloud computing* encompasses many different types of services. Therefore, evaluating business needs carefully before choosing a cloud vendor is imperative. Software-, platform-, and infrastructure-as-service vendors differ not only in the type of products they offer but also in the type of architecture their cloud incorporates. This paper examines the broad architectural differences in cloud computing products, the drawbacks to more generic approaches in cloud delivery, and the Joyent philosophy of constructing cloud computing infrastructures. The paper then describes the Joyent Smart Technologies cloud architecture from server and operating system through data center and software development platform.

Contents

Introduction

In the past three years, the term *cloud computing* has become increasingly elastic. Developers, industry analysts, and customers have all stretched and morphed the definition of the term to encompass a broad range of technologies and products. While an expanding market offers businesses more choice, it also complicates the rational analysis of the underlying technologies. For this reason, companies evaluating potential cloud computing infrastructures should take a pragmatic, business-minded approach in evaluating competing cloud computing infrastructures. Put succinctly, companies need to ask, "What does this technology do for my business?"

Not All Clouds Deliver the Same Service (….nor should they)

Cloud computing can entail a simple, Web-based application available over the Internet or an array of high-powered, massively provisioned data centers capable of processing thousands of online transactions per second for millions of worldwide customers. The only constant is the cloud itself—centrally located computing resources available over the Internet or wide area network. [1] Reviewing the broad types of cloud infrastructures and their primary IT and business functions helps businesses more fully understand this bewildering span of cloud computing products and their capabilities.

Software as a Service

Among products now defined as cloud computing, software as a service (SaaS) is the most targeted in terms of functionality. Companies rent time and space from online applications as diverse as sales lead tracking software to Web-based email. Enterprise level SaaS providers deliver a wide variety of sophisticated applications such as product lifecycle management, supply chain management, and many other vertical applications. If a business needs these specific applications, the SaaS model can save them the expense of buying hardware, software,

and long-term maintenance. Drawbacks include the inability, in many cases, to fully customize the solution for individual business requirements, complexities in integrating the SaaS application with the existing business IT infrastructure, and difficulties in predicting and budgeting pay-as-you-go pricing. For many companies, however, these targeted applications work well to solve specific business problems and are not meant to replace or significantly augment corporate computing resources.

Platform as a Service

When businesses need an application development environment, platform as a service (PaaS) is an effective and efficient type of cloud computing resource to consider. Most importantly companies can produce new applications more quickly and with a greater degree of flexibility than with older development platforms. Programmers and development managers especially appreciate that the cloud provider handles all the care and maintenance of the underlying operating system, servers, storage, and application containers. PaaS is especially useful when development teams are widespread geographically or when partner companies or divisions share development efforts. Engineers can more easily share and back up a central repository of application data as well as implement tighter version control and environment variables. PaaS cloud segments for development that painstakingly mirror the actual deployment environment of the application ensure a more stable and polished finished product. Companies can build up and tear down PaaS environments on demand with no capital expenditures or long-term investment. One drawback to many PaaS implementations, however, is the tie-in to one vendor's platform and infrastructure. Customers need to ensure that the platform allows for maximum portability of applications and data. Overall, companies can save considerable capital and operating expenses using a PaaS solution. [2]

Infrastructure as a Service

The most foundational use of cloud computing is infrastructure as a service (IaaS): the rental of a complete computing platform for running applications, hosting data, or housing a companies' entire computing environment. The latter use is not as common, and the majority of companies deploying IaaS now use it as a means to expand their computing capabilities in targeted IT areas without drastically increasing capital expenditures on new hardware and software. In fact, to better serve this market, most IaaS vendors emphasize their server and storage space as opposed to providing complete data center or application services. This server rack and disk space rental model is a natural outgrowth of the industry. Some of the most prominent IaaS suppliers are e-commerce and Internet information businesses that first entered the cloud computing market as a means to recoup revenue from excess, unused hardware in their data centers. Server rack space is a useful commodity for businesses that require more computing power but want to avoid long-term capital outlays. However, many companies are discovering that rack space and virtual server rentals do not provide a complete solution, and that reliance on ever-increasing cloud-based hardware resources produces many of the same IT management issues as outright server and network hardware ownership. [3]

Limitations of Common Cloud Infrastructures

Cloud computing was initially conceived of as a way to allow businesses to unshackle themselves from the challenges associated with managing hardware while delivering applications. Unfortunately, many of the cloud computing companies have made technology decisions that don't eliminate those basic constraints – they simply virtualize the problem.

The most general issue is that hardware-centric cloud infrastructures lack focus on solving common problems in software development. Many of the cloud computing companies have a surplus of hardware due to

their existing operations, and are simply leveraging that hardware inventory as a means to accomplish additional flow of revenue. This isn't their core business model, and they have not focused their innovation on building solutions that make it easier and more efficient to deliver software. Consumers of these services should make sure that the hardware, software, and network infrastructure has been built and optimized for application delivery.

In some extreme instances, an inflexible set of hardware or virtualization parameters may determine the type of software or software development that is possible on the infrastructure, further restricting customer options and even locking in data to that specific architecture. Moves and changes to another system, platform, or architecture might be costly or extremely complex.

One of the most common issues is that the infrastructures are not truly elastic in nature. They do not have the ability to scale vertically by providing immediate access to resources beyond a subdivided physical environment. Instead they manage scale with rapid provisioning of additional virtual servers, using horizontal scale to provide a limited form of elasticity. Customers using this type of cloud computing notice that adding more processes and users requires the manual addition of more virtual CPUs and starting up new instances of the applciation, a task that their cloud vendor cannot accomplish instantly. Performance then suffers, but more importantly, the customer confronts the sprawl of more and more added resources. While these newly added servers are *virtual* machines, they are nonetheless additional resources to pay for and manage. Ironically, physical CPUs may become heavily utilized running *underutilized* virtual machines.

Many cloud customers face continual issues of sizing and capacity planning—how much disk, CPU, and RAM must they purchase? How

will new software, more users, or increased application demands affect its cloud installation?

Finally, because they use hardware simulation on a per-machine basis many cloud computing infrastructure providers are not able to take advantage of resource pooling in their billing. To maintain profitability, they effectively must pass through the underlying hardware costs and management overhead, without the advantage of resource pooling. The cloud providers should use resource pooling as a mechanism for smoothing consumption so that all complex infrastructure costs that are related but disassociated from the application business model are not transferred directly to the software provider. This is especially important for applications that receive bursts of traffic, or that are public-facing and subject to complex traffic load.

Public, Private, and Hybrid Clouds

Cloud computing offered as an IaaS solution is typically either public or private. Public IaaS deployment usually entails a number of companies sharing the same cloud hardware at the same time, with reasonable separation of data and network segments to preserve security. Private clouds are infrastructures entirely hosted in privately owned, company data centers. Virtual private clouds, which are becoming increasingly popular, are cloud resources that a vendor hosts on dedicated hardware for a customer. Hybrid cloud solutions consist of some combination of public, private, and virtual private solutions. Generally, public cloud infrastructures are the least expensive, but the lack of control and resource sharing may make it less desirable for certain types of business computing. Private clouds, while offering greater control, are a more expensive proposition as the company must purchase, house, and maintain all the hardware and software for its cloud. Many businesses find virtual private clouds the best solution. In this scenario, the cloud host maintains the hardware and software, while dedicated hosting and maintenance ensure greater security and segmentation.

The Joyent Cloud Philosophy

Many companies have moved business processes and computing resources to cloud computing in an effort to save capital and operating expenses in their data centers. [4] However, as detailed earlier, many of the problems they sought to leave behind have followed them into the cloud, including:

- Lack of application focus
- Sizing and capacity planning issues
- Low utilization of virtual resources
- No true on-demand scaling
- Unpredictable, utility-style billing

While the use of virtual resources has made cloud computing possible, Joyent believes it is not the entire answer to delivering the best business value to cloud computing customers.

Customers should spend computing dollars only on resources that help them differentiate and improve their businesses. In the vast majority of cases, the key catalyst for business improvement is software. For this reason Joyent has developed Joyent Smart Technologies to eliminate the complexities associated with managing hardware and software in cloud environments and to return the focus to business productivity and software innovation.

Using the Joyent Smart Technologies, complex capacity planning decisions are reduced dramatically, such as the required number of virtual machines, CPUs, network bandwidth, or storage capacity. Joyent Smart Technologies can dynamically allocate these resources to ensure smooth operation of the application. Even in the area of application deployment, Joyent can provide a more dynamic and responsive environment that easily accommodates growth with far fewer management decisions required. With sizing and capacity issues minimized, businesses can focus on improving applications. Joyent Smart Technologies foster innovative software development by

eliminating many complexities in the environment; therefore, software developers can more quickly deliver stable applications.

Joyent also believes that application innovation thrives best in flexible environments. For this reason, Joyent SmartDataCenter supports any flavor of virtualization. Joyent SmartMachines provide the greatest software performance and flexibility in drawing from the resource pool, but XVM virtualization is also supported to enable legacy software and operating systems. This choice gives customers a broad range of software and development options from which to choose.

Joyent flexibility is also expressed in the range of deployment options available to customers. Businesses can use pre-configured, pre-installed software packages from a growing list of third-party software partners, choose from existing software packages for Linux / Unix operating systems, or build entirely new applications with languages like Java, PhP, Python, Ruby, Rails, and more. In addition, if public cloud hosting is not acceptable, companies can license the Joyent Smart Technologies infrastructure for deployment at their site. The Joyent virtual private cloud is yet another option, offering a completely dedicated private cloud infrastructure hosted at a Joyent data center.

Joyent Smart Technologies Architecture Overview

Rather than evolving from a generic virtual machine environment, Joyent purposefully engineered its technology from the ground up to provide greater flexibility, resilience, and performance for cloud customers and their applications. This application-centric focus is key for customers in realizing greater value in cloud-based computing. Joyent Smart Technologies eliminate complexity and make dynamic adjustments for its customers through three important components: the SmartMachine, SmartDataCenter, and SmartPlatform.

SmartMachine

In a traditional network, the developer must manage the entire server environment, including all the hardware (CPU, storage, RAM) as well as the operating system. If a system needed to scale horizontally, the operator purchased more hardware. Networks accomplished vertical scaling by over provisioning—installing enough hardware to accommodate worst-case, high-traffic situations, leaving the excess hardware resources idle during other periods. Multiple machines, wasted resources, and complex management were the result.

Virtual machines, by contrast, allow the abstraction of the hardware and sharing of its resources among multiple virtual machines. If managed correctly, this can ensure a higher utilization of hardware resources. This cannot provide instantaneous vertical scaling because resources are still provided in fixed sizes to the virtual machine, however, and the developer must still manage multiple machine instances with over provisioning to create horizontal scaling. One of the complicating factors in scaling under this scenario is that the time required to boot up an operating system within a virtual machine can take minutes – so there is an ongoing need to do predictive overprovision to maintain performance under unexpected loads. A final disadvantage to virtual machine administration is that developers must also manage the operating system, creating another layer of complexity in application production and stability.

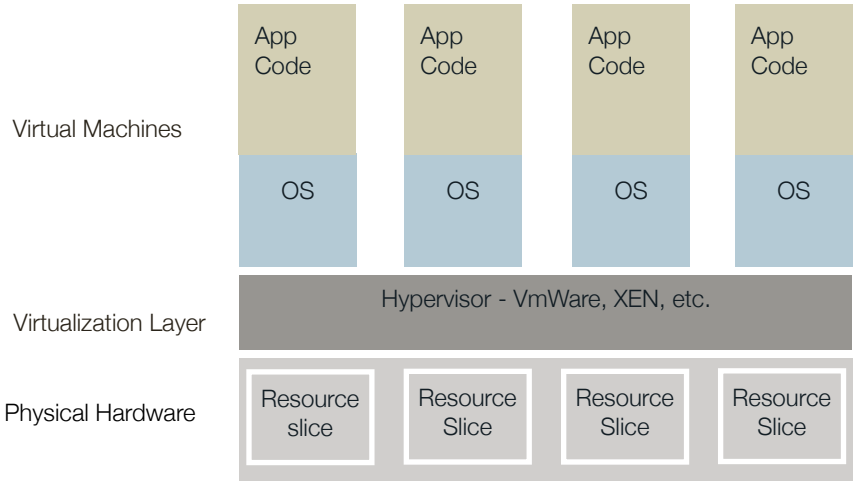

Figure 1. Architecture of traditional virtualization with Virtual Machines

The Joyent SmartMachine differs from traditional and virtual machine architectures primarily in that it goes further in abstraction of the hardware by presenting the hosted application with access to a pool of resources rather than control of a fixed resource. The SmartMachine has been designed to be very transparent to the underlying operating system, Joyent SmartOS. SmartOS uses this visibility into the SmartMachine to provide all SmartMachines with as-needed access to a large pool of available resources while still providing each SmartMachine with minimum guaranteed access to resources based on a pre-established fair share schedule. This transparency also allows the underlying operating system, Joyent SmartOS, to identify underutilized resources and use them to provide enhanced performance. In normal operating conditions, all RAM and CPU resources are either directly used by applications, or are being used by the operating system to optimize disk I/O and provide other performance enhancements to the SmartMachines.

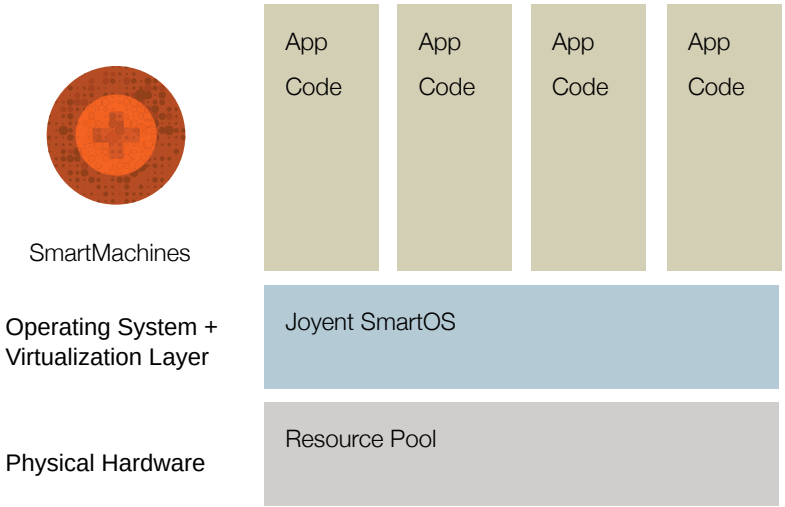

Figure 1. Architecture of SmartMachine virtualization

Furthermore, the SmartMachine's lightweight virtualization means that a highly tuned operating system (Joyent Unix) is available within the, reducing the need for the developer to manage multiple operating system images. When using a SmartMachine, the developer writes the

application and must only manage the application logic to enable horizontal scaling. This simplifies traditional hardware or virtual machine based architectures, where the developer must also create and manage multiple abstractions of physical servers with operating system and other software. Pooled hardware allows for easy and dynamic vertical scaling as well. Developers can then concentrate on application production rather than abstraction layers, operating system management, or virtual machine resources.

Because the SmartMachine is built on Joyent SmartOS, it incorporates the following core features:

Resource bursting—SmartMachines have the ability to tap the expanded pool of CPU, RAM, and network bandwidth through the Joyent SmartOS, providing needed capacity during bursts of activity or usage. Joyent SmartOS also supports more CPU cores and a larger memory footprint than other virtualization operating systems, so the resource pools will grow larger as underlying hardware is improved.

Flexible configuration— SmartMachines come preloaded with a variety of popular software environments including a vanilla Unix installation (SmartOS), development platforms (PHP, Java, Ruby, Rails), databases (MySQL, Oracle), load balancers (Zeus), and more. Customers can require variable hardware configurations, and the Joyent SmartOS OS can run XVM virtual machines to accommodate any legacy operating system in addition to its native SmartMachine.

Enhanced application performance —Joyent Unix provides a system-wide RAM cache for all disk I/O that significantly improves read and write performance for applications. Within a resource pool, any memory and CPU not directly utilized for application business logic is dedicated to providing improved performance. This approach of drawing on idle resources and providing them directly to applications is unique to Joyent's Smart Technologies.

Integrated security—The SmartOS OS also adds additional security and stability to the SmartMachine application environment. Per-machine administration is performed as root user with no ability to manipulate the operating system kernel. SmartOS isolates memory, network, and and processor isolation and blocks network reconfiguration and traffic sniffing to keep even public cloud applications completely segregated from one another. [Please see our white paper on cloud security.]

SmartDataCenter

In a multiple machine and application environment, the Joyent SmartDataCenter provides an abstraction layer for the entire data center – including compute, storage, and network resources across multiple physical and virtual locations. In addition to providing centralized management of these resources by cloud operators, SmartDataCenter provides APIs that allow applications to manage resource allocation and configuration.

SmartDataCenter's comprehensive APIs allow applications to automate horizontal scaling by adding or removing assets, detect network issues and reconfigure network topology, and perform many other complex infrastructure-related tasks that would normally require manual management.

At the network layer, the SmartDataCenter supports complete virtualization of complex, multi-device architectures, including VLANs, load balancing, VPN, and routing. SmartDataCenter also provides for the creation, provisioning, running, and termination of SmartMachines on any hardware within the SmartDataCenter. It is possible, for example, to decide to stop a process at one physical location and start it within seconds in another physical location to enhance load balancing or provide redundancy.

SmartDataCenter was designed to support a multi-tenant model for application deployment, where each tenant is provided with an

architecture of SmartMachines and network resources that are isolated from any other tenant. This multi-tenant model can be used for public hosting, as Joyent has done for it's hosting service, or it can be used to support federated, private deployment within large enterprises. Some enterprises have chosen to employ a virtual private data center configuration, reserving a dedicated set of hardware within a larger public SmartDataCenter.

SmartDataCenter can scale to millions of potential devices using a reliable message bus for management of all hardware and software in the SmartDataCenter. This message bus uses open, extensible protocols such as AMQP and XMPP that drive the largest volume communication systems currently deployed in the world.

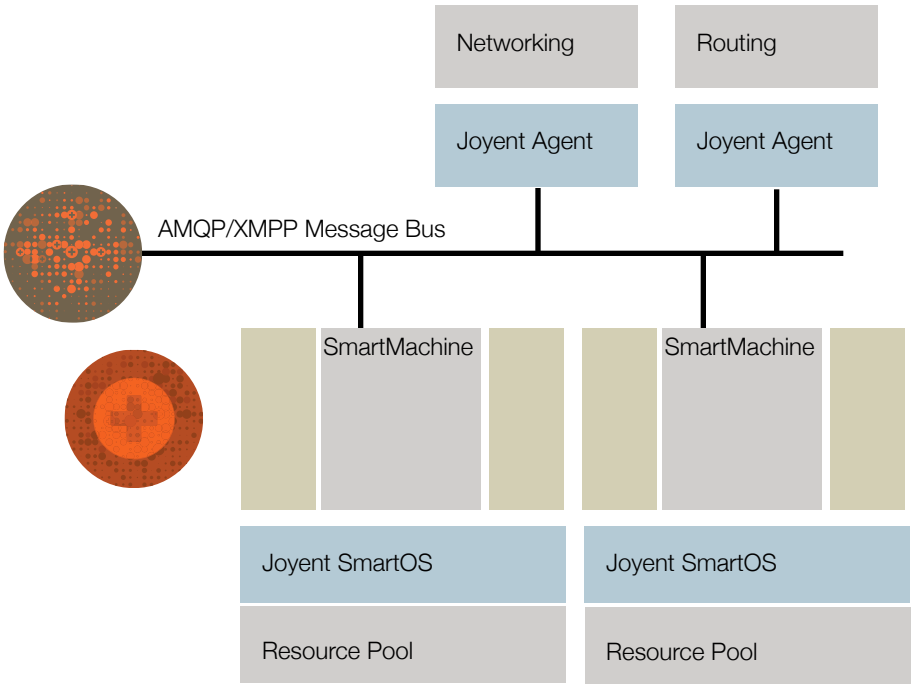

Figure 2. Architecture of SmartDataCenter

The administration of the SmartDataCenter can be automated in software, or managed through a web based administrative console. The management console can be used to monitor and manage network

assets, and provides hierarchical management where individual customers have more limited functionality.

The SmartDataCenter also incorporates the following features:

Flexible deployment—Licensing is available for private, public, or virtual private installations, and also through Joyent reseller partners, including Dell.

Centralized management—One GUI console manages the entire collection of system-wide SmartMachines, including resource allocation and OS provisioning and kernel patch management.

Extensible, resource-oriented architecture—Customers can add, remove, or introduce new capabilities and resources without disrupting the existing SmartDataCenter.

As a part of the Joyent hosting service, the SmartDataCenter has currently scaled to thousands of SmartMachinesmartMachines across multiple data centers worldwide. These applications, in turn, service millions of end user requests every day. The Joyent SmartDataCenter has run without system-wide service interruption (100 percent uptime) over six years with hardware utilization rates consistently above 70 percent.

SmartPlatform

The Joyent SmartPlatform vastly simplifies the platform-as-a-service architecture (PaaS) for cloud-based software development. Software engineers need only write JavaScript code to a single, unified SmartPlatform environment of potentially unlimited scale. The Joyent platform itself abstracts hardware and software across multiple machine instances and provides transparent horizontal scaling, unlimited programming containers, and pooled machine resources. Using the SmartPlatform, engineers do not need to deal with hardware interfaces

or operating system details, thereby freeing them to concentrate on application development.

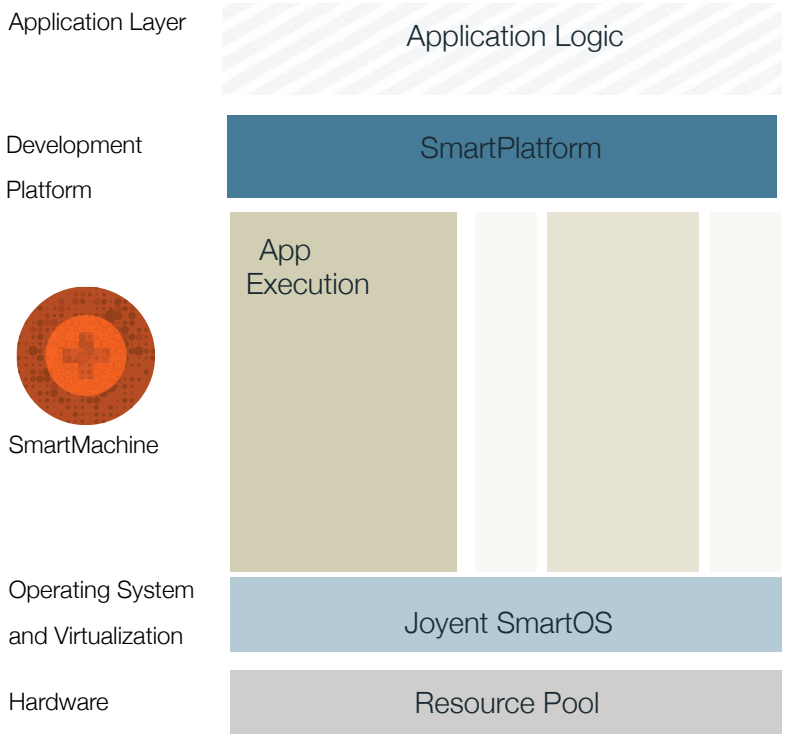

Figure 3. Application Architecture with SmartPlatform

The SmartPlatform is open source, so developers are able to deploy it to the environment of their choice and are not locked-in to any operating system or hardware environment. That said, there are numerous performance and management benefits that the other Smart Technologies provide when enabling an integrated development environment. As applications grow, Joyent SmartPlatform works with SmartDataCenter to automatically provide resources across multiple machines or even across multiple data centers, insuring better performance without recoding or recompiling. When development work is finished, companies can host the applications on Joyent SmartDataCenters or move the application to the network of their choice.

Joyent Cloud Products at Work

Joyent's Smart Technologies are the result of purposeful engineering to deliver a highly scalable, flexible, and resilient cloud computing infrastructure for business computing. As a result of these efforts, Joyent offers products that enhance applications for small businesses to global enterprises. Joyent SmartDataCenters house stand-alone, public applications as well as private, global financial processing facilities.

Joyent Application Hosting

Companies can host blogs, social networks, sales tracking, email, and other applications on Joyent public, private, or virtual private network segments and data centers. Joyent third-party partners offer many ready-to-run applications for quick installation on the Joyent cloud. Joyent services can help companies purchase or develop their own applications for use on the Joyent cloud.

Joyent PaaS

Businesses can develop applications on the Joyent SmartPlatform and quickly deploy them on the Joyent cloud. If developers need support for other virtualization environments, Joyent can provide a PaaS environment that meets their development specifications, ensuring that the finished product will perform as expected on the target platform.

Joyent IaaS

Joyent delivers public, private, or virtual private infrastructures for any sized business across multiple data centers. Joyent also licenses its Smart Technologies directly to users or through resellers to companies that wish to deploy private cloud computing facilities of their own. Resellers and VARs also license Joyent Smart Technologies to create public or private clouds for use with their customers.

Besides the cloud computing features of its SmartDataCenter, Joyent IaaS incorporates load balancing and disk caching for improved I/O, and enhanced security, billing, and reporting capabilities.

Conclusion

Before selecting a cloud computing provider, businesses should consider how they intend to use cloud computing facilities, what applications they intend to deploy in the cloud, and how the provider's cloud architecture facilitates their business goals. Whether businesses lease software, software development, or an entire computing infrastructure as a cloud-based service, they need to evaluate a potential platform for its flexibility, scalability, and resilience.

Joyent is committed to engineering and delivering high-performance cloud data centers that exceed the capabilities of common CPU and disk space rental products. Joyent believes that its Smart Technologies architecture gives companies greater choice and flexibility in deploying, developing, and growing cloud applications and infrastructure. With this application-centric focus, Joyent customers are free to concentrate on improving their business rather than managing their computing infrastructure.

References

- Knorr, Eric; Gruman, Galen. “What cloud computing really means.” Infoworld, April 7, 2008.

http://www.infoworld.com/d/cloud-computing/what-cloud-computing-really-means-031?page=0,0

- Paul, Frederick. “What PaaS means for growing companies.” Informationweek, November 18, 2008.

http://bmighty.informationweek.com/services/showArticle.jhtml?articleID=212100380

- Staten, James. “3 keys to using IaaS Public Clouds Wisely.” CIO Magazine, October 19, 2009.

http://www.cio.com/article/505318/3_Keys_to_Using_IaaS_Public_Clouds_Wisely

- Hickey, Andrew. “Study: Cloud Computing Enticing Repeat Customers.” ChannelWeb, February, 2010

http://www.crn.com/software/222600482;jsessionid=QHVTL5NGNNLMFQE1GHPSKHWATMY32JVN