SAY GOODBYE TO DIY DATA CENTERS
AN INFRASTRUCTURE-AS-A-SERVICE INTENSIVE

Through this year-long series of whitepapers and webinars, independent analyst Ben Kepes will be building a cloud computing curriculum designed for technologists and non-technical users alike. The mission is to build widespread knowledge about the cloud revolution and encourage discussion about the cloud's benefits for businesses of all sizes. Read more CloudU whitepapers and register for upcoming webinars at www.rackspace.com/cloud/cloudu

Diversity Limited

Sponsored By: rackspace HOSTING

Diversity Limited

Table of Contents

Executive Summary 2

A Virtual Tour of a Data Center 3

The Vision and Economics of IaaS 4

The Building Blocks of IaaS 6

- Compute 6
  - *The Foundations of Cloud Compute: Hypervisors and Virtualization* 7
  - *The Choice Is Yours: Operating Systems in the Cloud* 7
  - *The Choice Is Yours: RAM* 7
  - *The Choice Is Yours: Disk* 8
- Storage 9
  - *Cloud Storage* 9
  - *Content Delivery Networks* 9
- Network 10
  - *Switching and Routing* 10
  - *Domain Name System* 11
  - *Load Balancing* 11
- Database 12
- Monitoring and Autoscale 13
  - *Monitoring* 13
  - *Autoscale* 13

Case Study: IaaS Helps KarmaCRM Start-Up 14

Different Delivery Models for IaaS 15

- Public Cloud 15
- Private Cloud 15
- Virtual Private Clouds 16
- Cloud Bursting 16

Summary 17

Appendix: Key Components of Data Center Reliability 18

- Physical Security 18
- Precision Environment 18
- Conditioned Power 18
- Core Routing Equipment 19
- Network Technicians 19

About Diversity Analysis 20

About Rackspace 21

About the author 21

- Ben Kepes 21

Executive Summary

As IT increases in complexity each year, many small and medium businesses have little or no understanding of what a move to Cloud Computing really means in terms of underlying technology. While we contend that one of the core benefits of the Cloud is that it enables users to forget about technology and concentrate on their core business, this lack of information can create doubts about the value of moving to the Cloud.

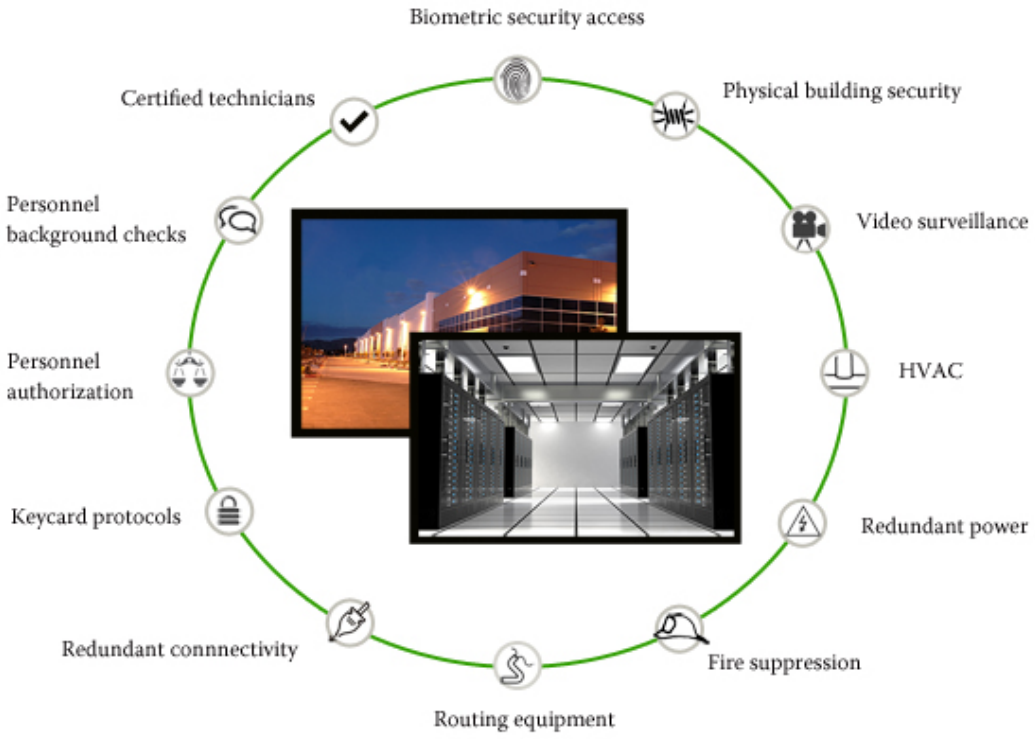
Many of these fears have been created by traditional vendors wishing to cast doubt on the ability of Cloud Computing to provide a truly robust and secure level of service. To this end we have written this whitepaper to give a moderately technical look at what goes into an Infrastructure as a Service offering.

In this report we start by taking a virtual tour of a modern data center to explain the robust, and expensive, features they contain. We then look at some specific services that make up Infrastructure as a Service (IaaS) offerings, services which enable end users to leverage the benefits of world-class data centers without the expense and complexity of maintenance. Finally, a variety of IaaS delivery models are discussed, from public cloud, to private and virtual private clouds, to cloud busting.

A Virtual Tour of a Data Center

Data centers are exceedingly complex and expensive projects to build and maintain. The level of sophistication of a modern data center is far in excess of what most standalone organizations could afford to build. From physical security to multiple redundant power supplies, modern data centers leave no stone unturned. It is important for users of IaaS to have a basic understanding of what goes into a modern data center in order to understand the financial impact of following an on-premises strategy, especially when the goal is uptime and security that matches the major Cloud Computing providers.

The diagram below details the multiple features that go into building a modern data center.



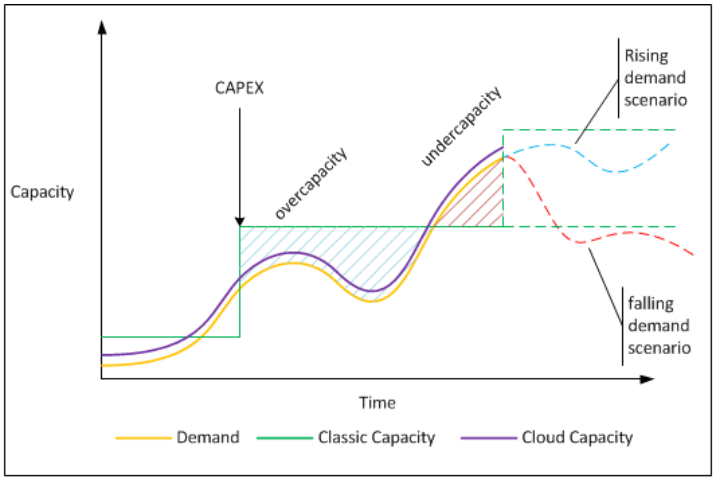
A detailed list of some of the important features of a modern data center are available in the appendix to this paper. The main point is that these myriad features together create infrastructure that is secure and reliable, but also complex and expensive for an individual business to maintain. With an overview of how a data center is built, we will now look at the economics that make IaaS so persuasive.

The Vision and Economics of IaaS

The vision of IaaS is that all the benefits of world-class data centers are available, on demand, to end users using a pay-as-you go model. IaaS vendors aim to offer customers the ability to acquire every aspect of infrastructure, as a service, offering a shift from massive capital expenditure and inflexible design decisions to IT that is far more flexible, elastic and modular. This model provides significant benefits, many of which are created via the economics of Cloud Computing.

Previously we published a CloudU report [1] that took an in-depth look at the economics of Cloud Computing. As described in detail in *Cloudonomics: the Economics of Cloud Computing*[2], under traditional IT models, organizations built their own IT infrastructure. This introduced two specific problems, the first of which relates to capacity planning. The capacity planning problem was created because of the fact that organizations had to spend significant time estimating potential peak loads for their data center and purchasing sufficient hardware to handle these loads. As the diagram below illustrates, this either left them with excess capacity that was under-utilized or, even worse, left them with demand beyond the capacity of their infrastructure which led to service degradation.

The purple line, closely following actual demand, shows the benefits of Cloud Computing infrastructure to scale up and down readily along with actual demand.



Capacity vs. Utilization curves [3]

The second economic problem created by utilizing internal infrastructure is that it required large-scale capital expenditure. This tends to significantly extend the decision-making process as capital is limited and any asset purchases need to be highly justified.

One of the core tenets of Cloud Computing is that it is a recurring expenditure model and, as we detailed in a previous report [4], is much like telephone or electricity expenditure in that it is accounted for as a standard operating expense. Just as organizations don't expect to have to invest capital in providing themselves with the ongoing use of telephone lines, this expectation holds true for IT infrastructure also.

With an understanding of the economics that make Cloud infrastructure so compelling, we will move on to an explanation of the different services that typically make up an IaaS offering.

The Building Blocks of IaaS

In a previous CloudU report [5], IaaS was defined as

> *a way of delivering Cloud Computing infrastructure – servers, storage, network and operating systems – as an on-demand service. Rather than purchasing servers, software, data center space or network equipment, clients instead buy those resources as a fully outsourced service on demand*

IaaS is essentially a set of building blocks that end users leverage to accomplish their IT goals. While the lines between IaaS, PaaS (Platform as a Service) and SaaS (Software as a Service) are blurring [6], typically IaaS is considered to be made of up

- Compute
- Storage
- Network
- Database
- Monitoring and Autoscale

While it is possible to purchase a complete IaaS offering without thinking about the component parts that go to make it up, adopting a more modular approach will help to ensure a solution that is tailored for the unique requirements of the organization. To achieve this aim requires an understanding of these discrete parts. Let's look at each in turn.

Compute

Compute is the most fundamental aspect of IaaS; it is the ability to make use of physical servers lying in a data center somewhere, on demand. The primary aspects of compute are really not that much different from the features of servers that many of us are familiar with: operating systems, RAM and disk. There are two components, however, that between them make up the one key difference between compute in the Cloud and compute in a traditional IT environment: the hypervisor and virtualization. And while the typical end user will never interact with the hypervisor, it is useful to begin the discussion of compute with a brief overview of this powerful software.

The Foundations of Cloud Compute: Hypervisors and Virtualization

A hypervisor is a software component that allows multiple operating systems to run concurrently on a single server. The hypervisor is the tool that allows virtualization to occur. As we detailed in the first CloudU whitepaper [7];

> *Virtualization was developed to overcome the limitations of physical hardware as it enables multiple pseudo-servers to be run on one physical device. This division of a single physical server into multiple "virtual" servers containing multiple sets of segregated data is the backbone of Cloud Computing, as it allows for far greater flexibility and resource utilization.*

When multiple users are using the same physical server, it is the hypervisor that guarantees that RAM and disk options selected at server provisioning are respected and no single customer monopolizes server resources. With this basic understanding of virtualization, let's look at the components of compute that are important to end users.

The Choice Is Yours: Operating Systems in the Cloud

When moving to the Cloud, one of the most basic questions that vendors will ask is which operating system do you want to use. Typically, most Cloud Computing vendors will offer a wide variety of operating systems to choose from, including multiple distributions of Windows and Linux. Since different operating systems are better for some tasks than others, choosing an operating system is a matter of personal preference based on the application being run. If a specific distribution of, say, Linux is a requirement for an important business application you are running, check with Cloud Computing vendors to ensure that it is available.

The Choice Is Yours: RAM

Frequently, Cloud Computing vendors will talk about the size of a slice. Not a pizza reference, this jargon speaks to the amount of random-access memory (RAM) included with a virtual server. Is it 256MB or 16 GB? These quantities of RAM are the "size" of the "slice" or virtual server. Like choosing an operating system, choosing the amount of RAM for a server should be based on the application being run. A complex SQL database might, for instance, require more RAM than a single static webpage. In general, one should pick the smallest server size capable of running the job. In the case of heavier than expected demand, the beauty of Cloud computing is that additional identical servers can be added instantly to handle the load.

The Choice Is Yours: Disk

When you bought your last computer, it might have come with 2 or 4 GB of RAM, and 160 GB of storage. This storage is commonly referred to as disk. It is the amount of storage that is available locally on the computer, without utilizing an external storage mechanism like an external hard drive. When purchasing a virtual server from a Cloud Computing vendor, a certain amount of disk is also available. When loading applications onto the server, this disk is consumed, as it is when storing things like documents in the case of a file server. Like choosing an Operating System and RAM, consideration should be given to storage needs when deciding how much disk is required.

While storage on the local server disk is considered a part of compute, there is a lot more to storage in the Cloud and that is where we will turn now.

Storage

IaaS storage has several components that work together to ensure safe, reliable and readily accessible data. There are also mechanisms related to the delivery of that data that are worth exploring.

Cloud Storage

Using Cloud storage, data is stored on multiple virtual servers among large storage pools provided by Cloud Computing vendors. Using Cloud storage, customers can expect that their data may physically span across multiple servers but that redundancy and failover will ensure that, in the event of a failure, data will continue to exist via automatic processes at the Cloud vendor's end.

Storage, apart from disk discussed above, can either be in block-based protocols or, alternatively, adopt a file-based approach. These two approaches can be summarized as follows;

Block-level storage is analogous to an external hard drive attached to a PC. In this case however the hard drive is virtual and it is mounted against a server instance. This instance then utilises the block as an individual drive for storing data that may be used for many applications – file storage and database storage being two examples. Block-level storage is extremely flexible, but is more complex to administer than file-level storage

File-level storage by contrast is used to store individual files where it is not necessary to mount a storage "block" directly to the virtual server. File-level storage tends to be simpler than block-level storage and is useful as a simple, bulk store for raw files such as images, videos, music files, JavaScript and CSS

In addition. reputable Cloud storage vendors allow customers to access their storage via an Application Programming Interface (API) meaning that processes such as backups, replication, and disaster recovery can be automated without resorting to manual processes.

Content Delivery Networks

Storing an object is one thing, but delivering it around the world at lightning fast speed is another. While not always a part of a Cloud Computing provider's capabilities, a Content Delivery Network (CDN) is a capability that end users delivering a lot of images, videos and even JavaScript or CSS for websites should seriously consider. A Content Delivery Network is a system of geographically

dispersed servers containing multiple copies of data. These servers are placed at various points in a network across the globe so as to minimize the distance-to-data for particular customers. With a move to offsite infrastructure, a CDN helps to reduce any speed impact caused by increased distance between users and the data they wish to access.

To illustrate the value of a CDN take, for example, the case of an ecommerce website hosted in the United States with most website visitors also located in the United States. The time that it takes the product catalogue images to load is probably quite fast, since the server and those accessing it are both located in the same general geography. What would happen, however, if the products suddenly became popular in China, or South Africa? The time to download the product images from the US based server would increase for these global visitors. CDN's solve this problem by taking image files and putting them on servers all over the world. Then, when a request for that images comes in, the CDN serves it from the location closest to the requestor. The webmaster need only upload the image to one place, their server, and the CDN takes care of the rest, distributing the file all over the world for optimal performance. By utilizing a CDN, Cloud vendors solve some of the issues created by a move to centralized infrastructure.

While storage and delivery are a critically important part of Cloud Computing, it is the next section, the network itself, which ensures information travels between points quickly and efficiently.

Network

Under traditional models, organizations had to invest not only in the hardware and software to run their data centers, but also in a significant level of network components to tie the servers together, and tie those combined servers to the Internet. In this section we look at the component parts that make up the network aspects of IaaS.

Switching and Routing

Switching and routing are two critical components of data transfer. Switching refers to the grouping of all data - regardless of its content, type or inherent structure - into individual "packets" which are then queued, ready to be sent over the network.

Routing, on the other hand, refers to the process of determining which particular

path along the network individual packets will be transferred.

IaaS vendors invest significant sums in buying the most advanced switching and routing technology available, in order to maximize the throughput and quality of data transfer across their networks. For an individual organization, the cost and stress of worrying about such technicalities is counter-productive and expensive.

Domain Name System

Domain Name System or DNS refers to the hierarchical naming system that is used to identify every entity connected to the Internet. All resources on the Internet are assigned a DNS and these codes are translated to addresses that are meaningful to users. DNS is analogous to a phone book in that assigns a particular name (for example www.diversity.net.nz) with a numerical identifier (210.48.108.35).

Cloud DNS services give organizations the ability to programmatically manage all aspects of DNS, for example the creation, updating, and deletion of different kinds of DNS records. DNS is a very important piece of a complete Cloud stack.

Load Balancing

Load balancing, as the name implies, is a technique that allows workloads to be distributed across multiple resources. The aim of load balancing is to optimize utilization and throughput while reducing the response time. Setting up a load balanced environment is possible using only code and Cloud servers but it can be more complex than many businesses want to take on. For this reason, some IaaS providers offer load balancing as a service (LBaaS). While the details differ between vendors, LBaaS gives one the ability to share workloads between various servers. Some vendors even offer the ability to set up specific types of load balancing algorithms particular to customer requirements. Since so much variation exists in the market for this service, with some vendors not offering it at all, it is important for prospective users of IaaS to understand their needs and ensure they're signing up to a service that meets their expectations.

We have seen how compute, storage and network work together to provide computing resources, but there is another common building block that we have not yet examined: the database.

Database

Cloud databases provide a readily scalable and easily configured method for organizations to create, store, and access their data. As with other parts of Cloud Computing, this model alleviates the need for organizations to purchase expensive hardware and software, deal with software upgrades, and hire professionals for administrative and maintenance tasks which are taken over by the service provider.

Database as a Service gives the ability to create an instance of a database that can be used as yet another building block of infrastructure. By combining a database with other requisite modules (load balancers, object stores, etc.), organizations can create an infrastructure "bundle" that is specifically tailored for their needs.

With all of these components in place, Cloud infrastructure relies on advanced monitoring and the ability to scale automatically to meet the ever-changing demand profiles of users. It is these two areas that we will now discuss.

Monitoring and Autoscale

IaaS frequently includes monitoring and autoscaling capabilities to give visibility over system performance, and ensure that performance meets demand.

Monitoring

Many organizations fear a move to Cloud Computing because they perceive it will leave them ignorant of the performance of their infrastructure. To address these fears, most reputable Cloud Computing vendors include some level of monitoring as part of their core offering, so that end users can assess uptime, performance, throughput and other relevant measures. For more sophisticated needs, SaaS-based monitoring solutions exist which give even finer control over Cloud monitoring.

Autoscale

Given that Cloud Computing providers are utilizing massive data centers, they have the ability to provide computing resources to individual customers practically without limits. In order to enjoy this compelling quality, however, IaaS needs to be designed to provide automatic scaling, within parameters that individual users can set. For example, one user may want to scale their application regardless ofhow many resources it is utilizing, while another may wish to set limits as to how far their infrastructure can scale. Tools to automate this scaling are quickly becoming a core feature of IaaS offerings.

A modern data center can be seen as a highly complex amalgam of multiple component parts covering different areas:

- Compute
- Storage
- Network
- Database
- Monitoring and Autoscale

By combining services across all of these areas, organizations can obtain an IaaS setup uniquely tailored to their individual needs. There are multiple methods for delivering IaaS, however, and it is to these methods we now look.

Case Study: IaaS Helps KarmaCRM Start-Up

KarmaCRM [8] is a customer relationship management application for SMBs. A recent startup, Karma had no legacy technology choices to take into account when determining its infrastructure set-up. This freedom gave them the ability to cherry-pick the very best options for the situation.

Founded by John Paul Narowski, a veteran of other CRM vendors, KarmaCRM makes extensive use of Cloud Computing for all parts of its operation. Narowski points out that by utilizing Cloud Computing, KarmaCRM can

> *focus on developing and growing our company. By leveraging cloud offerings we gain access to a wealth of services that we couldn't afford to create or manage in-house. In the IT world, it's far too easy to fall into the trap of doing everything in-house, and waste your IT talent managing mail servers instead of growing your business.*

KarmaCRM uses a number of different IaaS building blocks including compute, storage, database, and monitoring to run their business. By using infrastructure elements provided by the third parties, Karma has managed to build a service level far in excess of what it would have otherwise been able to afford. Particular services that Karma utilizes include:

- **Compute:** Virtual Linux servers for applications, database, and staging environments
- **Storage:** File-level Cloud storage for daily server backups
- **Monitoring:** Cloud monitoring to cover CPU/Memory usage as well as automate server provisioning

One of the benefits that Narowski sees from using IaaS is the ability to have redundant solutions. KarmaCRM is able to obtain a fully redundant infrastructure set-up with world-class levels of security solutions and the most current solutions available, all for a few dollars a month.

As a new business, Narowski also enjoys the fact that the more technical parts of infrastructure, such as load balancing and DNS, are taken care of by their Cloud Computing vendor.

Different Delivery Models for IaaS

Having examined the common building blocks of IaaS, we can look now at several different approaches towards delivering infrastructure as a service. These approaches can be divided into public, private and hybrid infrastructure. It is important for organizations to understand the difference between these three approaches.

Public Cloud

Public Cloud is considered infrastructure that consists of shared resources, deployed on a self-service basis over the Internet. The benefit of Public Cloud is that organizations are taking advantage of the highest levels of efficiency as Public Cloud pools and averages a huge number of different users, all with varying usage patterns. This creates a smoothing effect that results in the highest levels of efficiency as infrastructure costs are spread out among many users. Another benefit of Public Cloud is that it creates almost limitless scaling opportunities. Public Cloud vendors have such massive pools of resources that there are few situations in which Public Clouds cannot handle the scaling requirements that users may have.

Private Cloud

By contrast, Private Cloud is infrastructure that emulates Cloud Computing but does so on a private network. Where a Private Cloud approach is taken, organizations may use either their own hardware or the hardware of a third party. However, in both cases the hardware is dedicated to their own use and there is no sharing of infrastructure between users. It has been argued [9] that Private Clouds remove the very attributes that make Cloud Computing compelling – sharing, multi-tenancy and massive scalability. Many vendors are simply pushing organizations to traditional virtualization products, and calling it Cloud. While it is true that virtualization brings significant benefits to organizations, these benefits are much less than those bought by true Cloud Computing in many cases.

The third category of IaaS is Hybrid Clouds. Hybrid Clouds fall into two distinct categories.

Virtual Private Clouds

Virtual Private Clouds are created when Public Cloud vendors fence a part of their own Cloud infrastructure to provide specifically for a single customer. This virtual private approach has several attendant benefits – it answers some of the concerns that organizations raise with regards to sharing hardware with other organizations. At the same time, it still offers the ability to scale automatically and at very short notice.

Cloud Bursting

Cloud bursting can be seen as a combination of traditional dedicated hosting alongside Public Cloud networks. Using Cloud bursting, an organization will continue to use their own infrastructure for their regular needs, while any peak events will be automatically passed through to the Public Cloud provider. Cloud bursting brings several real benefits: it allows organizations to continue utilizing infrastructure they already own, while realizing some of the scale benefits of the Cloud.

While Cloud bursting introduces some issue in terms of management and governance, a number of tools now exist to automate the connection and management of a combined on-premises/external Cloud approach towards infrastructure.

Summary

Modern data centers are incredibly complex operations that include high technology in all parts of their operations – from physical environment to data storage to network delivery. Building a modern data center is an expensive project and one that is difficult to justify while mature Cloud Computing alternatives exist.

IaaS is a term that covers a range of services, which should be thought of as individual building blocks – while it is possible for organizations to purchase IaaS services as a complete offering, many will have a need to pick specific offerings depending on their needs and usage patterns.

While IaaS certainly removes significant maintenance and administration from end users, it still requires those users to have an understanding of the component parts of a data center so that they can make informed decisions when sourcing and assessing alternatives.

Appendix: Key Components of Data Center Reliability

Physical Security

- **Security of the building** - Keycard protocols, biometric scanning protocols and round-the-clock interior and exterior surveillance should be a standard monitoring procedure for data centers
- **Authorization of personnel** - Only authorized data center personnel should be granted access credentials to data centers
- **Background checking** - Every potential data center employee should undergo multiple and thorough background security checks before they're hired

Precision Environment

- HVAC (Heating Ventilation Air Conditioning) systems should have redundancy built in. This ensures that a duplicate system immediately comes online should there be an HVAC system failure.
- Data centers should include systems to regularly circulate and filter supply air
- Data centers should have advanced fire suppression systems which not only detect and extinguish fires, but provide cutover to alternative infrastructure in the event of a fire

Conditioned Power

- Data centers should have power systems that are designed to run uninterrupted via Uninterruptible Power Systems (UPSs) and backup generators
- These emergency power systems should have their own redundant systems to ensure supply in the event of a backup system failure

Core Routing Equipment

- Data centers should utilize routing equipment with fully redundant systems
- Data centers should have multiple fibre optic connections to multiple service providers

Network Technicians

- All technicians working within the data center should be certified by a third party certification organization up to, or above, the level at which they work
- Technicians should be available 24x&x365 in case there is an emergency requiring the attention of skilled IT professionals.

About Diversity Analysis

Diversity Analysis is a broad spectrum consultancy specializing in SaaS, Cloud Computing and business strategy. Our research focuses on the trends in these areas with greater emphasis on technology, business strategies, mergers and acquisitions. The extensive experience of our analysts in the field and our closer interactions with both vendors and users of these technologies puts us in a unique position to understand their perspectives perfectly and, also, to offer our analysis to match their needs. Our analysts take a deep dive into the latest technological developments in the above mentioned areas. This, in turn, helps our clients stay ahead of the competition by taking advantage of these newer technologies and, also, by understanding any pitfalls they have to avoid.

Our Offerings: We offer both analysis and consultancy in the areas related to SaaS and Cloud Computing. Our focus is on technology, business strategy, mergers and acquisitions. Our methodology is structured as follows:

- Research Alerts
- Research Briefings
- Whitepapers
- Case Studies

We also participate in various conferences and are available for vendor briefings through Telephone and/or Voice Over IP.

About Rackspace

Rackspace Hosting is the world's leading specialist in hosting and Cloud Computing. The San Antonio-based company provides Fanatical Support® to its customers, across a portfolio of IT services, including Managed Hosting and Cloud Computing. Rackspace is also the founder of OpenStack™, an open source Cloud platform with broad industry support, designed to offer Cloud consumers greater choice. For more information, visit www.rackspace.com.

About the author

Ben Kepes

Ben is the founder and managing director of Diversity Limited, a consultancy specializing in Cloud Computing/SaaS, Collaboration, Business strategy and user-centric design. More information on Ben and Diversity Limited can be found at http://diversity.net.nz

1. http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Cloudonomics-The_Economics_of_Cloud_Computing.pdf

2. http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Cloudonomics-The_Economics_of_Cloud_Computing.pdf

3. http://www.chades.net/

4. http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Revolution_Not_Evolution-Whitepaper.pdf

5. Understanding the Cloud Computing Stack http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Understanding-the-Cloud-Computing-Stack.pdf

6. http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Understanding-the-Cloud-Computing-Stack.pdf

7. http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Revolution_Not_Evolution-Whitepaper.pdf

8. http://www.karmacrm.com/

9. http://www.zdnet.com/blog/saas/when-will-the-crowd-turn-against-private-Cloud/973