JOURNAL OF
Software Technology
October 2011 Vol.14 No.4

Cloud Computing

DACS

http://journal.thedacs.com

http://iac.dtic.mil/dacs

TECH VIEWS

By John Dingman, Editor

There has been so much published about cloud computing in the last couple of years it would seem difficult to find new things to say about it. However, the model is changing so rapidly there is indeed plenty to say about it, and there will be for years to come. As you will read in "Cloud Computing – How Easy Is It", the environment is evolving so fast that even porting a single application to the cloud becomes a challenge of keeping up with the software releases.

Much of the current literature on cloud computing addresses security. This is certainly a valid concern, especially for anyone thinking about 'releasing' their private data 'into the void', which requires a great deal of confidence in what security measures are in place. Larry Clinton writes in "One Side Now" that a 2011 study found 62% of security experts had "little or no faith" in cloud security. Ominous. But behind this statistic may be the force that will motivate corporations and researchers to find answers. Kaus Phaltankar introduces Compliance as a Service (CaaS) in his article as a solution for both providers and customers.

Entirely new disciplines are being created as a result of the cloud. Joe Weinman's article on "Cloudonomics" describes his newly created model for quantification of cloud services. For anyone who first might like an introduction to cloud computing, Arlene Minkiewicz does a wonderfully eloquent job in her poetically titled "Cloud Nine, Are We There Yet?"

Cloud Nine, Are we there yet?

By Arlene Minkiewicz

AN INTRODUCTION TO CLOUD COMPUTING AND A LOOK AT WHETHER THE PROMISES OF CLOUD COMPUTING ARE REAL OR JUST VAPOR.

In 1961 at the MIT Centennial, John McCarthy opined "if computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility…. the computer utility could become the basis of a new and important industry" [1]. In 2006, Amazon Web Services was launched providing computing on a utility basis. Since that time the notion of cloud computing has been emerging and evolving.

Cloud computing is a paradigm that makes the notion of utility computing a reality. Instead of Information Technology (IT) organizations investing in all of the hardware, software and infrastructure necessary to meet their business needs, cloud computing makes access to hardware, software and infrastructure available through the internet, generally utilizing a pay for use model. Basically cloud computing allows an organization to adopt a different economic model for meeting IT needs by reducing capital investments and increasing operational investments, a model which is likely to offer cost savings to many organizations.

There is still a great deal of hype around cloud computing, as many vendors have their marketing engines further into the clouds then their technology supports. Despite this Gartner predicts that by 2012 one in five businesses will not own its own IT assets. [2]. In late 2010 the Office of Management and Budget (OMB) under direction from the White House told federal agencies that starting in 2012 they are expected to consider cloud first "whenever a secure, reliable, cost-effective cloud option exists." [3]

There are certainly many reasons why an organization would consider moving at least some of their IT functions into the cloud. In addition to potential cost savings the cloud offers the possibility of increased availability, easier collaboration, lower capital costs, scalability and virtualization. There are of course concerns as well. The technology is still relatively immature with no definitive set of standards for interface or compliance with regulations. Businesses lose hands on control of their IT resources with little recourse if their IT vendor shuts down or goes out of business. Additionally, there are security and data privacy concerns. There is also the fact that not all ventures into the cloud will be cost effective for the business.

This article introduces the concept of cloud computing and discusses the potential benefits for a business as well as those things which could be barriers to adoption. It examines the types of applications where cloud computing is an efficient cost effective solution and the types of applications where its use could be problematic or costly. Several examples of successful cloud implementations are presented and discussed.

Cloud Computing

"Cloud computing embraces cyber-infrastructure and builds upon decades of research in virtualization, distributed computing, grid computing, and more recently networking, web and software services." [4] In other words, although the term cloud computing is relatively new, the concepts and technologies behind cloud computing have been emerging and evolving for some time. Consumers of cloud computing access hardware, software and networking capabilities from third party providers.

So what is "the cloud" anyway? The cloud refers to the resources and applications that are available on the Internet or other network via any device that connects to the internet or other network. The term cloud originates from the diagrams that are often used to portray the reaches and capabilities of the Internet. According to the National Institute of Standards and Technology (NIST), cloud computing delivers five key features; on-demand self-service, ubiquitous network access, location independent resource pooling, rapid elasticity and measured service [8].

Cloud computing providers offer Internet connected servers which house applications and can store data. They also provide needed capabilities such as virtualization, grid management, database, and communications infrastructure. Usage monitoring and billing mechanisms are also required. Through virtualization, multiple customers can access the same

piece of physical hardware through separate server instances – allowing them to share hardware resources while isolating their operations and data from one another. The process for scaling and provisioning to meet changing customer demand is usually fully automated through software designed for that purpose. Application Program Interfaces (APIs) or web services provide control and access for the cloud.

There are four types of clouds discussed in the literature. Some consider only public clouds to be cloud computing but there are many instances where cloud computing technology is being applied to other types of clouds as well:

- **Public cloud** – Available to any user of the Internet willing to meet the terms and conditions of the cloud service providers. The public cloud is owned by the organizations selling cloud services
- **Private cloud** – Cloud computing infrastructure and technologies are maintained and operated for a specific organization, department or agency. The private cloud is owned by the organization, department or agency that utilizes it. It may be run by them or a 3rd party organization. A private cloud makes sense to bring some cloud computing benefits to an organization that for security or legal reasons cannot have their data in "the cloud"
- **Community cloud** – Cloud infrastructure that is established and maintained where several organizations, departments or agencies have similar concerns, security requirements, or compliance requirements.
- **Hybrid cloud** – There are multiple interpretations for what specifically constitutes a hybrid cloud but it clearly denotes the combined use of multiple types of clouds linked together through unique interfaces to allow organizations to optimize their use of the cloud without exposing themselves to potential risks of a public or community cloud.

In addition to there being various types of clouds, there are also several types of cloud computing offerings described in the literature:

- **Software as a Service (SaaS)** – Applications that are accessed via the cloud. End users access commercially available software applications remotely through the internet. Typical examples of SaaS include collaboration, project management, document management, social networking, customer relationship management (CRM) and Human Resource (HR) applications.
- **Infrastructure as a Service (IaaS)** – Computer Infrastructure is accessed via the cloud. Rather than purchasing, provisioning and maintaining servers, data center and network equipment, end users utilize computer infrastructure, generally through a platform virtualization environment, through the Internet. IaaS is usually purchased on a utility computing basis where the user only pays for the resources that they utilize such as processing by the hour or storage by the day. Typical examples of IaaS include backup and recovery, storage, content delivery networks, service management and computation.
- **Platform as a Service (PaaS)** – Development platform is accessed via the cloud. End user has access to the hardware, software and infrastructure necessary to develop or test applications. Typical examples of PaaS include database, development and testing, and business intelligence environments

Benefits

One could easily see the value cloud computing might bring to organizations – particularly small to medium enterprises (SMEs) that may not have the capital to invest in the IT infrastructure that might take their business to the next level. Some benefits will be also be attractive to larger organizations although there may be additional barriers to their adoption.

One oft cited benefit of cloud computing is the cost savings that can be offered to organizations that chose to take advantage of cloud services. Potential cost savings come from several sources. Certainly an organization can expect to save money on both hardware and software. As less hardware is being used, maintenance costs on existing hardware will reduce and less new hardware will be acquired. Software license and maintenance fees will be eliminated. These benefits are particularly important in organizations where needs for hardware and/or software peak and ebb over time. Instead of resourcing to peak needs, they have the opportunity of only paying for what they actually use. This will not only offer savings for hardware and software but also has the potential to reduce an organizations need for space, power consumption and IT staff. According to James Staten of Forrester, "Most enterprise data centers are using less than 50 percent of the total capacity of their resources."[5]

Because cloud technology provides for automatic or near automatic scaling and provisioning, the cloud can add agility to an organization as their needs for IT change. Reliability and availability may be significantly improved because large, well

established cloud providers have equipment and redundancy built into their offerings. Dynamic provisioning makes it possible for cloud providers to offer Infrastructure on Demand (IOD). Portability is increased because IT resources are no longer tied to a location, data and applications are available wherever the Internet is available.

Cloud providers and consumers benefit from the fact that software is located in one central location. Updates and repairs are accomplished easily and can be delivered to all clients as soon as they are available with no installation dramas on the end user's side.

Finally, if all of the worlds computing can be done with fewer servers running at any given time, the world benefits through a reduced carbon footprint. A study recently conducted by Microsoft, Accenture and WSP Environment and Energy found that for large deployments (10,000 users) energy use and carbon emissions could be reduced by more than 30% while for small deployments (100 users) the improvement could be as much as 90%.[6]

Risks and Challenges

Clearly there are many reasons an organization might consider transcending to the clouds. There are of course also reasons that the cloud may not be the right option for every organization or for every type of application.

Security is a key concern of many potential cloud customers. Certain organizations and agencies may never be comfortable with housing specific categories of data outside of their own walls. Additional, organizations, agencies and/or governments may have regulations as to physical locations where certain types of data are allowed to be stored. Data stored in the cloud could be anywhere. The good news is that security concerns in the cloud are being proactively addressed by the federal government in anticipation of the cloud first initiative mentioned earlier. In 2009 the GSA's cloud office established working groups on both security and standards and in 2010 they launched a government wide security certification and accreditation process for solutions in the cloud. Lockheed Martin conducted a survey in 2010 where they found that the more people know about the cloud, the less concerned they are with security in the cloud. [7]

Some have cited reliability as a concern. This may seem contradictory to the availability benefit noted earlier, but surely not all cloud providers have the same level of redundancy built into their systems. Additionally, cloud providers tend to co-locate their servers in one or few locations – taking advantage of favorable real estate and power costs- so a major power outage could have a significant effect on service availability.

There are instances where cloud computing may not be a cost effective solution, especially for organizations with large amounts of data and intensive data processing requirements. Potential cost savings could be overshadowed by the cost associated with the high bandwidth required to process all that data.

Another risk to consider when evaluating cloud based solutions has to do with the loss of control that occurs when you opt for the cloud. There are several aspects of this loss of control to think about. IT staff lose a level of control because they are no longer free to design platforms for specific business needs and they can't change technology on a whim. Another concern relates to the portability of cloud based solutions. Because of the relative immaturity of the technology, standards have yet to be established and cloud APIs are proprietary. If a cloud provider were to go out of business, or if the service they deliver deteriorates, there may be no quick exit strategy for customers who need to switch to another cloud provider.

Cost Drivers

As previously mentioned, one potential benefit of cloud computing is cost savings. And there is compelling evidence that migration to the cloud is definitely worth it for many computing needs of many organizations and agencies. When evaluating cloud solutions it is important to look at the cost of migration as well as the costs of operation once transition to the cloud is complete. What follows are some factors to consider when evaluating potential migrations to the cloud.

Clearly, the complexity of the migration is a significant factor in its overall cost. Complex migrations take more time and effort than simple ones. The nature of the capabilities being migrated as well as the volume of data involved in the migration will significantly impact the cost of transition to the cloud.

Cost benefit is realized when equipment can be eliminated. Cloud solutions that result in an organizations elimination of servers and other equipment will realize more cost benefit. The more capacity that is moved to the cloud, the more an organization can lower their maintenance, acquisition and IT staff costs – increasing the cost benefit of cloud migration.

Another factor is the efficiency of the organization's current operation. If an organization has already achieved high

utilization of their resources through optimization, load balancing or constant high volume operations, they may find less benefit than an organization that resources for peak but normally operates way below peak.

The type of cloud involved may drive cost as well. The potential for cost savings is greatest with the public cloud because there are many customers using the same resources allowing providers to pass savings along to their customers. Community and private clouds have more limited audiences, making them likely to be more expensive per user. Clearly the extent of a private or community cloud can influence the amount of cost savings they can offer to end users.

Requirements for security can also drive costs of cloud migration and on-going operation. Storage and transmission of secure data requires both digital and physical safe guards which naturally come with a price.

Additionally, there are organizational and cultural considerations when migrating to the cloud. Marketing, training and education are important when introducing any new tool or process to an organization in order to combat cultural resistance. New policies, standards and software license agreements (SLAs) may need to be developed, deployed and implemented.

Examples and Findings

While there are many reasons for organizations to consider the cloud, there are also many factors to evaluate before making the leap. It seems that the cloud offers huge potential benefits for smaller organization while larger organizations and government agencies are likely to find that some hybrid solution combining public and private cloud concepts will be their best option.

In light of the cloud first directive, several agencies of the US Federal government have already begun to deploy solutions in the clouds with some relatively impressive successes. By targeting capabilities that are currently expensive or inefficient and well suited to the cloud is a good strategy for cloud migration. A few examples follow.

Los Alamos National Laboratories wanted to roll out an infrastructure on demand architecture to facilitate quick rollout of new projects and eliminate other delays. Because of the nature of their work, security was a significant concern so they decided to go with a private cloud. They created a cloud using Microsoft© SharePoint for cloud workflows and integration point, VMware vCloud Director to manage and operate the cloud, and VMware vShield to provide security. With this architecture they have been able to provision a server, an activity that used to take 30 days, in under 30 minutes. As they are now using virtualization, they have been able to eliminate physical hardware reducing maintenance costs, power and electronic waste. According to Anil Karemel, IT Manager at Los Alamos, they expect their eventual savings to be $1.3 million annually. No data was available on the cost of the migration. [9]

The Defense Information Systems Agency, recognized that implementation of new software and systems at the DoD was expensive, time consuming and was being conducted in an environment less than conducive to cross collaborate and ubiquitous delivery. To address this, DISA created Forge.mil which provides tools and services for rapid development, testing and deployment of software to the entire DoD. Cloud provider CollabNet provides a software development platform that facilitates reuse and collaboration for Forge.mil's 5000 users. DISA estimates that Forge.mil saves between $200,000 and $500,000 per project.

The US Federal Government's website, USA.gov provides users with information about benefits, grants, jobs, taxes, health, voting, technology, and other information useful to the citizenry. Naturally access to USA.gov varies dramatically as conditions in the country and the world change with spikes in traffic around natural disasters, national elections, etc. The General Services Administration decided to move USA.gov to the Terremark's Enterprise Cloud service. In doing this they found that site upgrade time went from nine months to one day and monthly down time moved from two hours to near zero. The cost to operate the legacy USA.gov operations was $2.35 million annually plus personnel costs of $350,000. The move to the cloud resulted in a total annual cost of $650,000 resulting in a 72% cost savings.

More examples of federal migrations to the cloud can be found in [10].

Conclusions

The cloud computing paradigm offers organizations and federal agencies an alternative to meeting all their IT needs with in-house resources. Cloud computing consumers use the Internet (or other network) to run applications and store data on servers that could be anywhere in the world. The ability to access computing power on a utility basis offers the consumer the opportunity to save costs since costs are shared

among all of provider's users. It also allows them to reduce or remove internally acquired and maintained software and hardware – reducing costs for acquisition, maintenance and potentially IT staff. Cloud servers are generally utilized at a much higher utilization than in house computers because they are shared by so many users, increasing the productivity of the cloud providers and eliminating waste on the users side. Cloud computing providers offer virtualization and provisioning capabilities that may increase significantly the efficiencies with which solutions are available to the end user.

While there are many benefits to cloud computing, there are concerns as well. Security continues to be an issue, driving some organizations to less cost effective private clouds. There are also risks associated with the fact that the technology is still immature and the APIs being used are proprietary, making portability between providers problematic. End users must be able to deal with some loss of control of their IT environment.

The US Federal government agencies have been challenged by the OMB and President Obama to start thinking "cloud first" whenever and wherever it makes sense. And many federal agencies have already started to address that challenge with noticeable improvements in cost and productivity. While there is still immaturity and imperfections in the cloud solutions that are available, there is definitely reason to believe that cloud computing can help both industry and the government do more with less.

References

[1] http://en.wikipedia.org/wiki/Utility_computing

[2] Guseva; "Gartner: Top Technology Predictions for 2010 and Beyond" , CMS Wire, January 2010, available at (http://www.cmswire.com/cms/enterprise-20/gartner-top-technology-predictions-for-2010-and-beyond-006390.php) (retrieved 3/2011)

[3] Samson, "Feds take cloud-first approach to IT", InfoWorld Tech Watch, December 08,2010, available at http://www.infoworld.com/t/cloud-computing/feds-take-cloud-first-approach-it-829 (retrieved 3/2011)

[4] Vouk, M.A., "Cloud computing – Issues, Research and Implementations", 30th International Conference on Information Technology Interfaces (ITI 2008), Croatia, 2008

[5] Leavitt Communications, "Is Cloud Computing Really Ready for Prime Time?", available at http://www.leavcom.com/ieee_jan09.htm (retrieved 4/2011)

[6] "Cloud Computing Study for Microsoft shows dramatic reduction in carbon emissions", WSP Environment & Energy, Nov 2010, available at http://www.wspenvironmental.com/newsroom/news-2/view/cloud-computing-study-for-microsoft-shows-dramatic-reduction-in-carbon-emissions-235 (retrieved 4/2011)

[7] The Download, Cloud Computing Research Study, An 1105 Government Information Group Research Study, Lockheed Martin, 2010, available at http://download.1105media.com/GIG/Custom/2011PDFS/CloudComputing/CloudComputingLM.pdf (retrieved 4/2011)

[8] Wyld, D., "Moving to the Cloud: An Introduction to Cloud Computing in Government, IBM Center for the Business of Government E-Government Series, 2009

[9] Robb,D, "Building a Private Cloud at Los Alamos", IT Enterprise Planet,com, Sept 22, 2010, available at http://www.enterpriseitplanet.com/article.php/3904821 (Retrieved 4/2011)

[10] Kundra, V., "State of Public Sector Cloud Computing", May 2010, available at http://www.info.apps.gov/sites/default/files/StateOfCloudComputingReport-FINALv3_508.pdf (retrieved 4/2011)

About the Author

Arlene F. Minkiewicz is the Chief Scientist at PRICE Systems, LLC with over 27 years of experience at PRICE building cost models. She leads the cost research activity for TruePlanning, the suite of cost estimating products that PRICE provides. She is a software measurement expert dedicated to finding creative solutions focused on helping make software development professionals successful. She is widely published and speaks frequently on software related topics.

Cloudonomics: A Rigorous Approach to Cloud Benefit Quantification

By Joe Weinman

RATHER THAN BEING A HAZY CONCEPT, THE CLOUD CAN BE AXIOMATICALLY DEFINED AND RIGOROUSLY ANALYZED THROUGH A NEW APPROACH CALLED CLOUDONOMICS

Cloud computing and cloud services consistently place at the top of surveys ranking IT trends and CIO interests. This is largely due to the position of the cloud at the nexus of macro trends such as social media, the Internet, Web 2.0, and mobility and broadband wireless, as well as the opportunity for benefits such as reduced cost and enhanced agility.

Traditional approaches to assessing cloud benefits largely fall into two categories: vague yet enticing words, such as "agility," and empirical data from case studies, which may or may not apply to the general case.

Cloudonomics—a term and discipline founded by the author (Weinman, 2008)—seeks to provide a rigorous foundation based on calculus, statistics, trigonometry, system dynamics, economics, and computational complexity theory, which can be used to interpret empirical results. We will provide an overview of these results together with references to more detailed analyses.

Defining the Cloud, from an Economic Viewpoint

Many definitions of Cloud Computing exist. Perhaps the most widely accepted is the one developed by the National Institute of Standards and Technology, now stable at version 15 (Mell and Grance, 2011):

> "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.
>
> This cloud model promotes availability and is composed of five essential characteristics: …on-demand, self service … broad network access … resource pooling … rapid elasticity…[and] measured service."

This is an excellent, broadly applicable definition. From an economic viewpoint, however, we can use a semantically equivalent mnemonic—CLOUD (Chan, 2009)—which can help surface economic benefits. A CLOUD is a service that has the following attributes:

- **Common Infrastructure**—i.e., pooled, standardized resources, with benefits generated by statistical multiplexing.
- **Location-independence**—i.e., ubiquitous availability meeting performance requirements, with benefits deriving from latency reduction and user experience enhancement.
- **Online connectivity**—an enabler of other attributes ensuring service access. Costs and performance impacts of network architectures can be quantified using traditional methods.
- **Utility pricing**—i.e., usage-sensitive or pay-per-use pricing, with benefits applying in environments with variable demand levels.
- **on-Demand Resources**—i.e., scalable, elastic resources provisioned and deprovisioned without delay or costs associated with change.

We shall overview results concerning these benefits and additional related topics. The results are often counterintuitive. Various layers—Infrastructure as a Service, Platform as a Service, and Software as a Service—all have different benefit drivers. Here, we shall focus on Infrastructure as a Service, which is a foundation for many other benefits. After all, a salient difference between Platform Services and Service-Oriented Architectures and Integrated Development Environments ultimately comes down to Infrastructure resources, and a salient different between licensed software and SaaS ultimately resides in infrastructure costs and flexibility, including pricing and elasticity. Thus, we shall focus on infrastructure.

The Value of Common Infrastructure

What is the value of consolidating demands from independent sources into a common pool rather than partitioning them?

The traditional answer—"economies of scale"—certainly has some validity. Overhead costs can be reduced, and buyer power enhanced through volume purchasing.

However, another key value of consolidation is what might be called the "statistics of scale" (Weinman, 2008). Under the right conditions, multiplexing demand can generate benefits in terms of higher utilization and thus lower cost per delivered resource—with unutilized resource costs factored in—than unconsolidated workloads, for infrastructure built to peak requirements. For infrastructure built to less than peak, demand multiplexing can reduce the unserved demand, reducing a penalty function associated with that unserved demand, which may represent either loss of revenue or a Service-Level agreement violation payout.

A useful—if imperfect—measure of "smoothness" or "flatness" is the coefficient of variation c_v (not to be confused with either the variance σ^2 nor the correlation coefficient. This coefficient is defined as the non-negative ratio of the standard deviation σ to the absolute value of the mean $|\mu|$. The larger the mean for a given standard deviation, or the smaller the standard deviation for a given mean, the "smoother" the curve is.

This smoothness is important, because a facility with fixed assets servicing highly variable demand will achieve lower utilization than a similar one servicing relatively flat demand. To put it another way, one with low utilization has excess assets, whose cost—whether leasing or depreciation—must be carried by revenue-generating ones.

With that as background, the beauty of the cloud comes into focus: multiplexing demand from multiple sources *may* reduce the coefficient of variation (Weinman, 2011d).

Specifically, let $X_1, X_2, ...X_n$ be n independent random variables with identical standard deviation σ and positive mean μ and thus each with coefficient of variation $c_v(X)$. Note that they need not have the same distribution: one may be Normal, one may be exponential, and so forth.

Since under these conditions the mean of the sum is the sum of the means, the mean of the aggregate demand $X_1 + X_2 + \ldots X_n$ is $n\mu$. Since the variance of the sum is the sum of the variances, the variance of the aggregate demand is $n\sigma^2$ and therefore the standard deviation is $\sqrt{n} \times \sigma$. Thus, the coefficient of variation is $\frac{\sqrt{n}\sigma}{n\mu} = \frac{\sigma}{\sqrt{n}\mu} = \frac{1}{\sqrt{n}} c_v(X)$. In other words, adding n independent demands together reduces the coefficient of variation to $1/\sqrt{n}$ of its unaggregated value.

Thus, as n grows larger, the penalty function associated with insufficient or excess resources grows relatively smaller.

Importantly, it does not take an enormous number of such demands to approximate "perfection." Aggregation of 100 workloads will be within 10% of the penalty associated with an infinitely large cloud provider, and aggregation of 400 workloads will be within 5%.

It must be noted however, that the assumption of workload independence is a key one. There are two other possibilities worth considering. One is that workloads are not independent, but are negatively correlated or even complementary. If the two demands are X and $1-X$, say, the sum is of course merely the "random" variable "1," which has a 0 standard deviation. Such a scenario is not that farfetched: appropriate selection of customer segments can lead to a virtuous situation. In the early days of AC electric power, Samuel Insull targeted consumers who needed lighting in the morning and at night, trolley operators, whose peak electricity use was at rush hour, and factories, thus generating relatively flat aggregate demand [Carr, 2008].

The other possibility is that of perfectly correlated demand. Specifically, if each of n demands is characterized by X, then the aggregate demand is nX and the variance of the sum is $n^2\sigma^2(X)$. Thus the standard deviation is $n\sigma(X)$ and the mean is $n\mu(X)$. The coefficient of variation of the aggregate is unhelpfully $\frac{n\sigma(X)}{n\mu(X)} = \frac{\sigma(X)}{\mu(X)} = c_v(X)$. In other words, the coefficient of variation of the aggregate is the same as any of its components.

A weaker condition, where we merely have at least one simultaneous peak, is equally problematic from the perspective of attempting to increase utilization and thus derive favorable economics.

Two lessons may be drawn from this. First, contrary to the proposition that only a few large cloud providers will survive, these statistical arguments suggest that for correlated demand or simultaneous peaks, midsize providers don't generate much

benefit relative to "private" implementations, but then again neither do large ones, and for independent demands a midsize provider can achieve statistical economies that are quite close to that of an infinitely large provider.

Second, if a "community cloud" is intended to aggregate demand from correlated components, it will not generate any benefits due to the statistics of scale, although it certainly may generate economies of scale, should there be any. The data on whether there truly are economies of scale for large cloud providers relative to large enterprises is mixed. After all, today's ultramegadatacenter-based cloud providers use the same pods that are available to any enterprise or the government, and probably not at a substantially different discount. Large cloud providers may have benefits in terms of locating near cheap power, but this isn't an economy of scale, it is, well, an economy of locating near cheap power, equally available to anyone else. While early entrants have advantages in automation, these differences are being eroded as 3rd parties offer management and administration, virtualization, provisioning, billing, portal, and other software on either an open source or competitive cost basis.

The Value of Location-Independence

At the dawn of the computing age, since the mountain of equipment—relays, dials, vacuum tubes, etc.—would not come to the users, the users would come to the mountain. Fast forward half a century, and the equipment, or at least applications, services, and content, do come to the user over global networks. An important attribute of today's world is this ubiquity and availability, regardless of whether one is using wired, wireless, converged, or satellite networks.

However, emerging applications and functions are increasingly latency sensitive. If latency constraints are not met, there is economic loss. Consider the case of word or phrase suggestions incrementally provided with each keystroke. If AJAX (Asynchronous Javascript and XML) processing, including a network round trip to a server, is delayed sufficiently, an entire word will be typed just as the first set of responses to the first keypress begins to arrive. This has no value.

Latency can also impact other application contexts, from ruining the flow of natural conversation and thus hindering collaboration when delays cross 200 milliseconds, to reducing revenue in eCommerce and online search applications [Hamilton, 2009].

To make things worse, for many applications it is not just the latency of a single round trip request-response transaction that matters, but the latency of multiple round trips, for example, as numerous objects such as images are fetched to load a web page.

Moreover those latencies add up. A wait of 300 milliseconds or 3 seconds for a web page to load is not a lot, until you multiply it by thousands of knowledge or contact center workers, each with hundreds of transactions per day, or consider the importance of timeliness under battlefield conditions.

As a result, the single instance datacenter is not suited for these types of tasks. And, more than simply an outbound stream which may be buffered and/or delayed without substantial impact, these are interactive and real-time services.

Given human response times in the tens and low hundreds of milliseconds, the circumference of the Earth, and the speed of light in fiber (only 124 miles per millisecond), supporting a global user base requires a dispersed services architecture. While there are many thorny issues of coordination, consistency, availability, partition-tolerance, we will address a simple one: the investment implications of such distributed architectures, especially given the potential of common infrastructure.

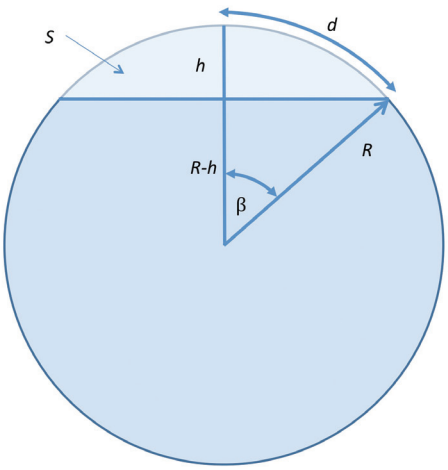
Latency is strongly, but not perfectly, correlated with distance. The reason for the imperfect correlation has to do with routing anomalies in wireline networks, specifics of router hops and optical-electronic-optical conversions, and so forth, before congestion or link outages even enter the picture. However, given the strong correlation, we can use distance as a proxy for latency. On a plane, both worst-case and expected latency are proportional to the radius of the circle centered on the service node. Consequently, while there are variations due to whether the coverage strategy is circle packing (like cannon balls—with gaps in-between) or circle covering (full coverage but with overlaps), the area covered is proportional to the radius and the number of service nodes (Weinman, 2011).

For n service nodes on a plane, the area A covered depends on the radius r related to the latency/distance, and a constant of proportionality k that depends on the packing/covering strategy. Thus, $A = kn\pi r^2$. Therefore, if the area A is a constant we can rearrange terms to realize that $r \propto 1/\sqrt{n}$.

This simple equation gives us a $\sqrt{n}$ term but for geometric rather than statistical reasons. Similar economic characteristics

ensue: it doesn't take many nodes to make rapid initial gains, but then there are rapidly diminishing returns: getting worst-case global network round-trip latency from 160 milliseconds to 80 or 40 or 20 takes only a handful or a couple of dozen nodes, but after that, thousands or millions of nodes will only result in microsecond or nanosecond improvements.

To be precise, the Earth is not a plane. At best, it approximates a sphere. Consequently, if we are to devise a formula, we need to consider packings or coverings not of circles, but of "spherical caps" (like baseball caps, but without rims). A useful formula is that the surface area S of a spherical cap is proportional to its height: $S = 2\pi Rh$. We can thus calculate the surface area of the cap in terms of its angular radius β, since $S = 2\pi R^2(1-\cos(\beta))$. If we double this angular radius—which is equivalent to doubling the worst-case distance and thus latency along the surface of the sphere, we clearly have a surface area of $2\pi R^2(1-\cos(2\beta))$, rewriteable as $2\pi R^2(1-\cos^2(\beta)+\sin^2(\beta))$.



To help understand this, suppose we are placing service nodes. If we have just one and place it, say, at the North Pole, the worst case distance is the antipodal point: the South Pole. If we have two service nodes, optimal placement would be, in effect, the North Pole and the South Pole, and the worst case distance is that to a point on the equator, say, Quito. By increasing the number of service nodes from 1 to 2, we have decreased the distance by ½.

In other words, rather than the $r \propto 1/\sqrt{n}$ law, we have $r \propto 1/n$ rule when n is 1 or 2. However, as n increases we get closer and closer to $r \propto 1/\sqrt{n}$ as the surface of the sphere increasingly approximates a plane locally.

Given that $r \propto 1/\sqrt{n}$ is exactly correct on a plane and increasingly correct on a sphere, especially at the scale of continents or countries, it tells us that the diminishing returns due to the inverse square root make private investment increasingly difficult. An enterprise or government would be investing more and more capital in service nodes to get less and less improvement in response time.

However, depending on the application profile, the cloud can alter the balance. Specifically, if the cost is based on pay-per-use, having say, 1 node with 10,000 users or 10,000 nodes each with 1 user will cost the same. In practice, there will be some tradeoffs. For example, shorter transaction routes reduce aggregate network usage requirements on a bandwidth-mile basis and thus save either on (customers') network costs or (providers') infrastructure investments. On the other hand, maintaining multiple copies of a non-trivial application takes up storage space which costs money or can incur additional license fees. User data that partitions cleanly is cost-insensitive to division, but excessively mobile users may incur data transport costs that are nontrivial. In short, both public cloud and private implementations have application-dependent characteristics regarding user experience and network, storage, and processing costs.

The Value of Utility Pricing

Conventional wisdom suggests that cloud services must be cheaper due to immense economies of scale. However, empirical data offer a mixed and nuanced view (McKinsey, 2009; Harms and Yamartino, 2010). Briefly, economies of scale certainly may exist for large service providers. However, the question is not whether large service providers exhibit economies of scale, but whether these cost economies are sufficiently advantaged relative to enterprise or government scales to overcome the margin, SG&A, and uncollectables cost element *disadvantages* that any sustainable, rational, commercial service must of necessity have, and whether the total package exhibits any competitive net *price* advantage. Moreover, cloud technology is a moving target: any current advantages due to, say, proprietary automation or provisioning technology may not be sustainable in the long term as vendors arise that offer them to all players including small cloud providers, enterprises, and governments.

Some people believe that if cloud services are more costly on a unit basis then they should be avoided: after all, why pay more? However, this misses the cloud value proposition generally, including specifically the value of utility pricing.

The complete value proposition of pay-per-use pricing includes the benefit that, regardless of unit cost, these resources are paid for only if used, in contradistinction to owned, dedicated resources.

In everyday life, one often pays more for utility pricing. A midsize car may be financed, leased, or depreciated for roughly $300 per month or $10 per day. That same car from a rental car service might cost $45 per day, even after allowing for insurance and car washes and so forth. One does *not* shy away from "overpaying," because in fact, one is not overpaying, one is saving money relative to owning that car but only using it for one or two days.

Consequently, some rules about cloud value are clear (Weinman, 2011c). If the unit cost of the cloud is in fact cheaper than the unit cost of an owned resource, then the cloud should always be the solution. In effect, if I can rent a car for $5 a day, and owning it costs $10, I should use a rental for both short term and long term requirements.

If the cloud costs the same and resource demand is flat, both strategies cost the same. Where it gets interesting is when the cloud costs more, but the resource demand is variable.

Let the demand for resources be denoted by $D(t)$, where $0 \le t \le T$. Let the peak demand be $P = \max(D(t))$ over that period and the average demand be $A = \mu(D(t))$ over the same period. Let the utility premium—the ratio of the cloud unit cost to the owned and dedicated unit cost—be U. For the rental car example above, U would be 4.5 (=$45/$10). Let the baseline cost for owned resources be B. It should be noted that linear usage-sensitive pricing implies that we only care about A, not P for the utility priced resources, since the total cost of meeting demand $D(t)$ is $A \times U \times B \times T$, which is the value of the definite integral $\int_0^T U \times B \times D(t)\,dt$ For an owned solution that must meet peak resource requirements, the total cost is $P \times B \times T$. For the utility-priced resources to be a lower total cost solution, we must simply ensure that $A \times U \times B \times T < P \times B \times T$. The B and the T drop out, leading us to the insight that the cloud is cheaper when $(P/A) > U$. In English, this means that if the cloud's unit cost is U times that of a dedicated resource, but the Peak-to-Average ratio is higher than that, a pure cloud solution will be less expensive.

It's beyond the scope of this article, but it can be shown that for most variable demand situations, a hybrid of dedicated and owned resources is cost-optimal. This matches our intuitive optimization strategy: use owned resources such as homes and owned cars for long duration usage, use rented ones such as hotels or rental cars for short ones. The key factor turns out to be the percentage of time within the planning period that a given resource level is needed. If the percent of time that a given demand is needed is greater than the inverse of the utility premium, we can take that tranche of demand and serve it more cheaply via ownership. If it is less than the inverse of the utility premium, it may be served more cheaply via pay-per-use rental. Finally, if it is equal, there is a breakeven zone where either approach will work well.

Think of it this way: you probably need a car every day for your commute to work: owning is cheaper than renting. A spare car that you need only one day a year when your car is serviced is best met through a rental car. If your child is home from college 3 months out of the year, and rentals cost 4 times as much, the cost is equal, regardless of the approach you use. If you have 3 kids in college, whether you rent 0, 1, 2, or 3 cars won't matter.

In practice (Weinman, 2009), demands often are highly spiky. In commercial contexts, news stories, marketing promotions, special events, product launches, Internet flash floods and the like drive traffic spikes, enhancing the value of pay-per-use pricing. In government applications there also are spikes, e.g., due to tax filing, egovernment initiatives, and the like. For defense and intelligence applications, there are often tradeoffs between resource requirements and time limits on effective use of information. For example, an initiative might require extremely fast turnaround—say, processing hundreds of hours of drone video—to provide actionable intelligence. This creates a demand spike with national security implications.

One important cost component that we have not addressed yet is the cost of the network in hybrid solutions. If a hybrid solution *would* be optimal, but a fixed price $20 million network is required to make it work, the cost equations clearly will suffer. If the network is *variably* priced, that will shift where the breakeven point is for the hybrid. In the college student example, either a daily bus fare or a bus pass to get to and from the rental car location would alter the attractiveness of the rental car option, as would accounting for the value of time spent.

The Value of on-Demand Services

One other key attribute of the cloud as we've defined it is on-demand resources or services based on those resources. Briefly, the value of "on-demand" arises from avoiding both

excess resources and insufficient resources (Weinman, 2011f, Islam, S. et al., 2011). Excess resources are costly due to the weighted average cost of capital used to acquire the resources, or the opportunity cost of the capital not being productively employed elsewhere. Then there is the risk of obsolescence, requiring premature write-offs, the risk of loss, or the cost to ensure those resources against loss. They require floor space, and often require power and cooling.

Conversely, insufficient resources mean lost revenue, poor customer experience, loss of brand equity due to poor customer experiences, and in defense applications, inability to support the mission. These costs may have nonlinearities.

Let us briefly consider symmetric linear penalty costs associated with insufficient and excess resources. In effect, if the demand is D and the resources are R, the penalty cost is proportional to $|D-R|$. If they are time varying, the penalty cost is proportional to $\int |D(t)-R(t)|\,dt$. This will make analysis more tractable.

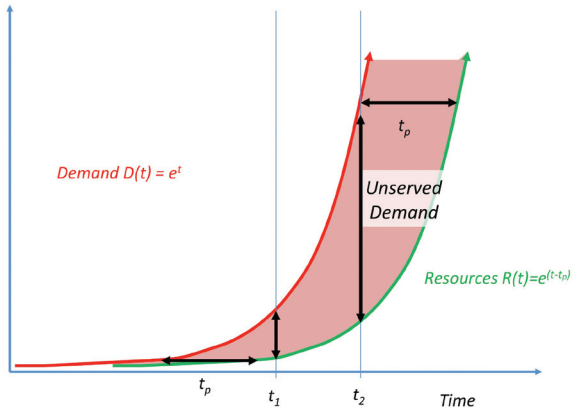
If demand is flat, on-demand resources are unnecessary.

If demand grows linearly and predictably, even long provisioning intervals are acceptable: one merely need order additional resources on a regular schedule.

Conversely, if demand declines linearly and predictably, if resources can be deprovisioned without extra costs (from packing, restocking, auction, etc.), on-demand is not necessary.

It is when demand is unpredictable and/or non-linear that issues arise with traditional resourcing and on-demand really shines. For example, let demand in a given time period be uniformly distributed from say, 0 to 1. If resources are set to 0, the penalty will always be due to insufficient resources and the expected value of the difference will be 1/3. If resources are set to 1, the penalty will always be due to excess resources and again, the expected value of the difference will be 1/3. However, if the resources are set at the midpoint, the expected value of the penalty will only be $\frac{1}{6}=\frac{1}{2}\times\frac{1}{3}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{3}\times\frac{1}{2}$. Of course, if the right resources are available at the right time, the penalty is zero. If we scale this model up to a uniform distribution on $[0,P]$, and assess the penalty over time T with a penalty cost of c, it is clear that the total penalty for fixed resources, even if we can set the level optimally, is $\frac{1}{6}P\times T\times c$. If the penalty function is linear but asymmetric, for example, the cost of unmet demand is higher than the cost of unused resources, the optimal fixed strategy tilts in favor of a "better safe than sorry" approach of excess resources, with an optimal point being distribution-dependent.

Space does not permit detailing the calculations I've done elsewhere (Weinman, 2011f), but we will highlight conclusions. When the demand function is exponential, i.e., $D(t)=e^t$, any fixed provisioning interval that attempts to provision in accordance with the current demand level (i.e., there is no forecasting) will fall exponentially further behind. This is because if the fixed provisioning interval is k we set resources $R(t)=e^{t-k}$, and thus the difference is $D(t)-R(t)=e^t-e^{t-k}=e^t\left(1-e^{-k}\right)=k_1e^t$, where $k_1=1-e^{-k}$, and thus the penalty cost is $c\times k_1e^t$, in other words, grows exponentially over time as well.



Exponential Growth with Continuous Monitoring And Non-Zero Provisioning Interval

The Value of Online Connectivity

Connectivity is an enabler of the prior value drivers. After all, on-demand resources can't be shared unless various customers in various locations can access them. Without sharing, pay-per-use pricing is economically unattractive for service providers to offer.

Connectivity then, has values and costs. When examining connectivity the costs may be clear: dollars per Gigabyte transferred or the capital costs of routers or optical facilities. The value is harder to quantify, since it often is an externality. A good approach is to consider the marginal cost, if any, of connectivity, and use it to offset the benefits associated with specific other pure cloud or hybrid cloud approaches. Often, this will be easy: an ecommerce spike can be probability-adjusted, the revenue associated with it estimated, and the marginal benefits associated with partial or total cloud solutions thus quantified. Against this must be deducted

any marginal network costs to assess the value of alternate approaches.

Behavioral Cloudonomics

We will now turn to some ancillary considerations in cloud economics. A basic assumption above is that neoclassical economics, expected utility theory, and quantitative finance rule decision-making. If, say, the expected, i.e., probability-adjusted, net present value of choice A costs less than that of choice B, we should select it.

A more recent approach to economic decision making is behavioral economics (Ariely, 2008, Lehrer, 2010) which admits that humans do not always make purely rational and quantitative decisions, but that often various cognitive biases and heuristics play a role, as well as "bounded rationality," the limits of people to calculate appropriate answers to decisions. In this field, numerous experiments have been conducted, books and scholarly papers written, and Nobel prizes awarded. We can't do the topic justice here.

However, we note that a number of behavioral economic favors impact the cloud—what may be called Behavioral Cloudonomics (Weinman, 2010). For example, perhaps the most well known result is that of "loss aversion:" people generally get less satisfaction from gaining a dollar than they feel pain from losing one. As a result of this and other factors, customers may recognize the financial advantage of pay-per-use, but avoid it due to what has been called a "flat-rate" bias (Lambrecht and Skiera, 2006). For example, the fear of an unexpected large monthly cell phone bill can lead one to overpay for services by signing up for a flat-rate plan even though measured service—i.e., utility pricing—would be cheaper due to normal light usage. Such biases could negatively impact cloud adoption.

On the other hand, they can also work to the cloud's advantage. Dan Ariely and his colleagues at MIT have conducted experiments regarding the special attraction of "free" (Shampaner and Ariely). Again, rational behavior is subsumed by irrational: subjects typically select a free ten dollar gift certificate (a ten dollar value) over paying seven dollars for a twenty dollar gift certificate (a thirteen dollar value). The lack of upfront investment in using the cloud is thus extremely attractive as it aligns with this particular bias.

Computational Complexity

Bounded rationality is not only an attribute of humans, but a very real factor in numerous decision problems due to computational complexity.

We have argued above that service providers can achieve extremely high utilization, while customers can benefit from leveraging dispersed resources. While generally true, the problem of satisfying demand of varying levels out of a distributed pool of resources constrained by say, distance, turns out to be computationally intractable. Elsewhere (Weinman, 2011b) I've proven that CLOUD COMPUTING DEMAND SATISFIABLITY is strongly NP-complete, i.e., not currently believed to be perfectly solvable in a useful amount of time, based on a transformation of BOOLEAN 3-SATISFIABILITY.

The implications are that there will always be some "friction" or "entropy" in the system. In turn this means that even if there is *exactly* the right aggregate capacity in a distributed cloud system, there may be no way to effectively determine in useful time the right assignment of capacity to demand. Even the mere act of determining service node locations has also been shown to be intractable (Megiddo and Tamir, 1983).

We have discussed the benefits of consolidation, i.e., demand aggregation into pooled resources. We have also discussed the benefits of dispersion in user experience enhancement. Unfortunately, we can't have both: we can choose to optimize the statistics of scale by building fewer consolidated facilities, and we can choose to optimize user experience by building more, dispersed facilities. The computational complexity results tell us that not only are there tradeoffs, but that determining an optimal tradeoff is intractable.

Axiomatic Foundations

While the topic of foundations arguably should be discussed first, we have left it until the end, because now their usefulness will be clearer. At one extreme, one might tie cloud computing to specific implementations of specific releases of specific hardware and software products. However, from the analyses above, it should be clear that the cloud, or at least a C.L.O.U.D., can be considered as an abstract concept.

As such, it can admit to an axiomatic formulation. Briefly, one can define (Weinman, 2011a) a cloud structure—in the spirit of a finite state automaton or Petri Net—as a structure that is a 6-tuple $\left(\mathbb{S},\mathbb{T},G,Q,\delta,q_0\right)$, where $\mathbb{S}$ is metric space, i.e., a set of points with a distance metric, $\mathbb{T}$ is time which is a measure space, i.e., a set of periods with a suitably-behaved duration measure; $G=(V,E)$ is an oriented graph; Q is a set of states, q_0 is an initial state; and δ is a transition function which may be deterministic or non-deterministic, and possibly recursively defined. Each state in Q combines assignments of resource capacity and demand, resource allocations, node

location, and pricing. Such a rich state definition captures the multiple interrelationships that may exist in the real world: capacity relative to demand drives pricing in accordance with price elasticity of demand and capacity increases to exploit market needs, pricing and resource location drive allocations as users seek the cheapest and nearest services, and allocation patterns lead to capacity planning for new resource levels. This definition is abstract and rigorously built on the foundations of mathematics: metric spaces, measure spaces, axiomatic set theory, σ -algebras, vector spaces, and the like.

A cloud can then be axiomatically defined as an entity with this structure that meets the five cloud axioms we've specified above, also suitably formalized.

As a result, we can consider a wide variety of clouds: compute clouds, hotel clouds, rental car clouds, restaurant clouds, and the like. Hamburger chains and hotel chains have numerous abstract structural similarities to the large cloud computing providers, after all.

Many of the results above are of direct importance to the business value of cloud computing, but, critically, are of general and broad applicability.

Summary

There are many important considerations as consumers, enterprises, and civilian, defense and intelligence agencies evaluate using various cloud patterns for some or all of their computing needs. In assessing those options, financial considerations are an important component. While exact cost figures for alternatives are important, a broad understanding of the theoretical underpinnings of the multiple values that cloud computing can provide are helpful.

In what might be called General Cloudonomics, we can see that these principles apply not only to computing, but any business with pay-per-use pricing for shared resources ubiquitously accessed over a network. Moreover, these general principles can help us navigate the current cloud world, which may be described as energetic positioning of a technology and business model that has clear benefits that have been proven across a variety of domains. Appropriate hybrids of owned, dedicated resources and on-demand, pay-per-use resources can minimize cost while maximizing flexibility and elasticity, although tradeoffs will always exist based on which performance parameter is prioritized. Moreover, not only are there tradeoffs, but limits to optimization due to the computational intractability of a number of problems revolving around geographically dispersed, networked architectures with various constraints on connectivity and capacity.

About the Author

Photo Credit: Bob Rannells, Bob Rannells Photography

Joe Weinman leads Communications, Media, and Entertainment for Hewlett-Packard's Business Solutions organization, with a team spanning the Americas, EMEA, and the Asia Pacific regions. He is known in the cloud computing community as the founder of Cloudonomics, a rigorous, multidisciplinary analytical approach leveraging economics, behavioral economics, statistics, calculus, computational complexity theory, simulation, and system dynamics to characterize the sometimes counterintuitive multi-dimensional business, financial, and user experience benefits of cloud computing and similar on-demand, pay-per-use business models. He is also the author of Cloudonomics, John Wiley & Sons, available Spring 2012.

He is a frequent global keynote speaker, a prolific inventor that has been awarded 14 U.S. and international patents, and an author who contributes to numerous on-line and print publications. He has held a variety of executive positions of increasing responsibility beginning with research and development at AT&T Bell Laboratories through his current position. He received a Bachelor of Science and Master of Science in Computer Science from Cornell University and the University of Wisconsin - Madison respectively, and completed Executive Education at the International Institute for Management Development in Lausanne, Switzerland. Email: joeweinman@{hp,hotmail,gmail}. Twitter: @joeweinman.

References

Ariely, D., 2008. *Predictably Irrational: The Hidden Forces that Shape Our Decisions*. HarperCollins.

Carr, N., *The Big Switch: Rewiring the World from Edison to Google*, W. W. Norton & Co., 2008.

Chan, T., 2009. Full Interview: AT&T's Joe Weinman. GreenTelecomLive. http://www.greentelecomlive.com/2009/03/16/full-interview-att%E2%80%99s-joe-weinman/

Hamilton, J., 2009. The Cost of Latency. Perspectives. http://perspectives.mvdirona.com/2009/10/31/TheCostOfLatency.aspx .

Harms, R., and Yamartino, M., 2010. The Economics of the Cloud. http://www.microsoft.com/presspass/presskits/cloud/docs/The-Economics-of-the-Cloud.pdf.

Islam, S., Lee, K., Fekete, A., Liu, A. How a Consumer Can Measure Elasticity for Cloud Platforms. University of Sydney, Technical Report 680. http://sydney.edu.au/engineering/it/research/tr/tr680.pdf .

Lambrecht, A., and B. Skiera, 2006. Paying too much and being happy about it: Existence, causes, and consequences of tariff-choice biases. *Journal of Marketing Research*. vol. 43, May 2006, pp. 212-223.

Lehrer, J., 2010. *How We Decide*. Mariner Books.

McKinsey & Co., 2009. Clearing the Air on Cloud Computing. http://uptimeinstitute.org/content/view/353/319

Megiddo, N., and Tamir, A., 1983. New Results on the Complexity of p-Center Problems. *SIAM Journal of Computing*, Vol. 12, No. 4, November, 1983.

Mell, P. and Grance, T., 2011. The NIST Definition of Cloud Computing (DRAFT). NIST Special Publication 800-145 (DRAFT). http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf.

Shampaner, K., and Ariely, D. Zero as a special price: the true value of free products. http://web.mit.edu/ariely/www/MIT/Papers/zero.pdf

Weinman, J., 2008. The 10 Laws of Cloudonomics. http://gigaom.com/2008/09/07/the-10-laws-of-cloudonomics/

Weinman, J., 2009. Peaking Through the Clouds. http://gigaom.com/2009/06/25/peaking-through-the-clouds/

Weinman, J., 2010. Lazy, Hazy, Crazy: The 10 Laws of Behavioral Cloudonomics. http://gigaom.com/2010/06/06/lazy-hazy-crazy-the-10-laws-of-behavioral-cloudonomics/

Weinman, J. 2011. As Time Goes By: The Law of Cloud Response Time. Working Paper. http://www.joeweinman.com/papers.htm.

Weinman, J. 2011a. Axiomatic Cloud Theory. Working Paper. http://www.joeweinman.com/papers.htm.

Weinman, J. 2011b. Cloud Computing is NP-complete. Working Paper. http://www.joeweinman.com/papers.htm.

Weinman, J. 2011c. Mathematical Proof of the Inevitability of Cloud Computing. Working Paper. http://www.joeweinman.com/papers.htm.

Weinman, J. 2011d. Smooth Operator: The Value of Demand Aggregation. Working Paper. http://www.joeweinman.com/papers.htm.

Weinman, J. 2011e. The Market for 'Melons': Quantity Uncertainty and the Market Mechanism. Working Paper. http://www.joeweinman.com/papers.htm.

Weinman, J. 2011f. Time is Money: The Value of On-Demand. Working Paper. http://www.joeweinman.com/papers.htm.

we like your feedback

At the DACS we are always pleased to hear from our journal readers. We are very interested in your suggestions, compliments, complaints, or questions. Please visit our website ***http://journal.thedacs.com****, and fill out the survey form. If you provide us with your contact information, we will be able to reach you to answer any questions.*

The DACS Gold Practice Initiative:

- Promotes effective selection/use of software acquisition & development practices
- Defines essential activities/benefits of each practice
- Considers the environment in which each practice is used
- Addresses the timeliness of practice benefits
- Recognizes interrelationships between practices that influence success or failure
- Contains quantitative and qualitative information
- A continually evolving resource for the DoD, Government, Industry and Academia
- Free to use/free to join

Learn More About the DACS Gold Practice Initiative:
http://www.goldpractices.com

Current Gold Practices:

- Acquisition Process Improvement
- Architecture-First Approach
- Assess Reuse Risks and Costs
- Binary Quality Gates at the Inch-Pebble Level
- Capture Artifacts in Rigorous, Model-Based Notation
- Commercial Specifications and Standards/Open Systems
- Defect Tracking Against Quality Targets
- Develop and Maintain a Life Cycle Business Case
- Ensure Interoperability
- Formal Inspections
- Formal Risk Management
- Goal-Question-Metric Approach
- Integrated product and Process Development
- Metrics-Based Scheduling
- Model-Based Testing
- Plan for Technology Insertion
- Requirements Management
- Requirements Trade-Off/Negotiations
- Statistical Process Control
- Track Earned Value

THE DACS GOLD PRACTICE INITIATIVE

DACS

The Data & Analysis Center for Software

100 Seymour Road
Utica, NY 13502
http://www.thedacs.com

Cloud Computing – How Easy is it?

By Thomas Kwasniewski

A SIMPLE EXPERIMENT OF PORTING A SMALL APPLICATION TO THE CLOUD TEACHES US THAT THERE ARE MANY POTENTIAL PITFALLS TO SUCCESS AND EXPERIENCE IS EVERYTHING

Most articles on cloud computing discuss security, policies, economies of scale, or other high level topics. For a more pragmatic approach, we decided to simply port an existing application to the cloud and report on what the experience taught us.

Selecting the Application

Finding an existing, simple application that would benefit from the power of the cloud architecture was the first step. Fortunately, an associate nearby was using a spreadsheet application to perform Monte Carlo simulations for physics-based modeling and analysis. The application was reaching its limit in the number of simulations that could feasibly be run given the time required to complete them. The simulations, being run over a weekend, were producing results on a set of 30,000 simulations and seemed to limit out at that number.

The application being used was Microsoft Excel, with Visual Basic for Applications (VBA) code used to execute the algorithms in the scientific model and to randomize the variables used in the Monte Carlo simulations. It was very compute-intensive, not I/O bound, and thus an excellent candidate for migration to the cloud.

Choosing the Cloud Platform

Using a spreadsheet for the Monte Carlo simulation had been proven somewhat difficult. It was not very configurable, the dataset was embedded, the variables and formulas were scattered on one worksheet, the results returned on other sheets, and sorting was a separate, user-initiated action. Even if it were possible to push the existing spreadsheet up to the cloud and have it execute via a virtual machine environment, and then access it via a user interface, it was not an ideal solution. A standalone application migrated to a platform-as-a-service (PaaS) would likely be a much more logical and robust approach. An application where the user interface, data sets, and domain logic were separated and loosely coupled would be a superior approach. This approach would give the most flexibility to adapt to further iterations of the project. If the model changed, or the need for separate models with different algorithms and business logic arose, it would be simple to respond by adding a new module. The decision, then, was made to rewrite the spreadsheet as a standalone Microsoft Windows application and port this to the PaaS cloud.

Selecting the Vendor

This step and the rewriting of the Excel spreadsheet to a standalone application went hand in hand. Driving factors that were immediately apparent:

- The existing algorithms were written in VBA. Customer projects can vary greatly and the complex math and science underlying them can be daunting. For this and future projects, it was important to be able to port the existing formulas as-is, without rewriting them, and in such a way that would not require expert-level understanding of the theory behind them.
- The Monte Carlo simulation randomize functions were also written in VBA
- The precision and format of the results should exactly match the existing spreadsheet to prove the results from the cloud application were correct.

The above factors suggested a port to VB.net as the most prudent solution. Excel and VB.Net, both being Microsoft products, led to the decision to use the Microsoft cloud product, Windows Azure. Also, Windows Azure offers a free demo capability which allowed us to test this with minimal cost. A free cloud trial, including Tools and an SDK, made sense for this project, as it would for other organizations venturing into the cloud for the first time.

A quick tour of the Windows Azure web site confirmed that their platform did support Visual Basic.net along with C#, C++, PHP, Ruby, Python and Java. At this point, due to the factors listed above, the Windows Azure platform and PaaS as the type of cloud platform were clearly the best choices.

Throughout the process more research was performed and knowledge gained about Windows Azure as well as other cloud

vendors and platforms. Several helpful books for this project included *The Cloud at Your Service* (Rosenberg & Mateos, 2010), *The Cloud Computing Bible* (Sosinsky, 2011), *Azure In Action* (Hay & Prince, 2010) and *Programming Windows Azure* (Krishnam, 2010). This reading proved invaluable in helping gain additional insight and knowledge that couldn't be gathered simply by browsing vendor web sites.

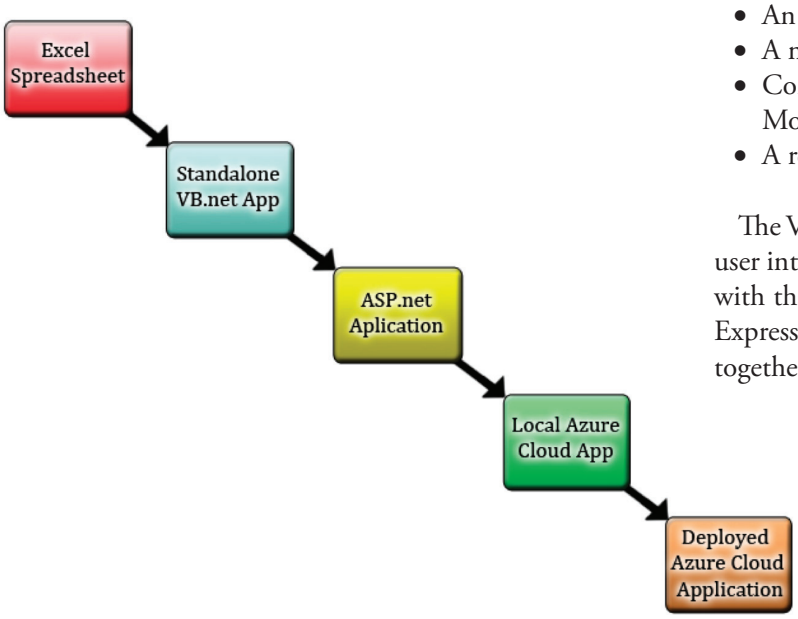


Figure 1: Major Application Migration Development Phases

Figure 1 depicts the overall process that was taken to migrate the application.

Rewrite the Application for the Cloud

The first step in migrating the application, was to create a standalone VB.net application that would duplicate the main functionality of the existing Microsoft Excel spreadsheet. Since VB.net was chosen as the language, Microsoft's Integrated Development Environment (IDE) was needed to develop the application.

The logical components of the Excel spreadsheet were divided into building blocks that would need to be created. They were:

- A Graphical User Interface, including:
  - Input parameters for the model
  - Browse controls for input/output data files
  - Function list box
  - Status box to display intermediate and final results
  - Progress bar
  - Run/Stop buttons
- A File I/O section
- An array to hold data set in memory for fast access
- A math section to duplicate formulas
- Control Logic, including the randomize functions for Monte Carlo simulations
- A results and sort routine

The VB.net environment is extremely strong in its graphical user interface (GUI) capability, prompting the choice to start with that aspect of the design first. The Visual Studio 2010 Express IDE worked flawlessly and the user interface came together quickly (Figure 2).

Thanks in part to the IntelliSense feature of the IDE coding the File I/O, the general control logic and the randomize functions were easy. Next it was necessary to tackle the math functions. Since the Excel spreadsheet used VBA, however, porting them to VB.net was almost seamless. The result was a working VB.net application that produced the same results as the original spreadsheet.

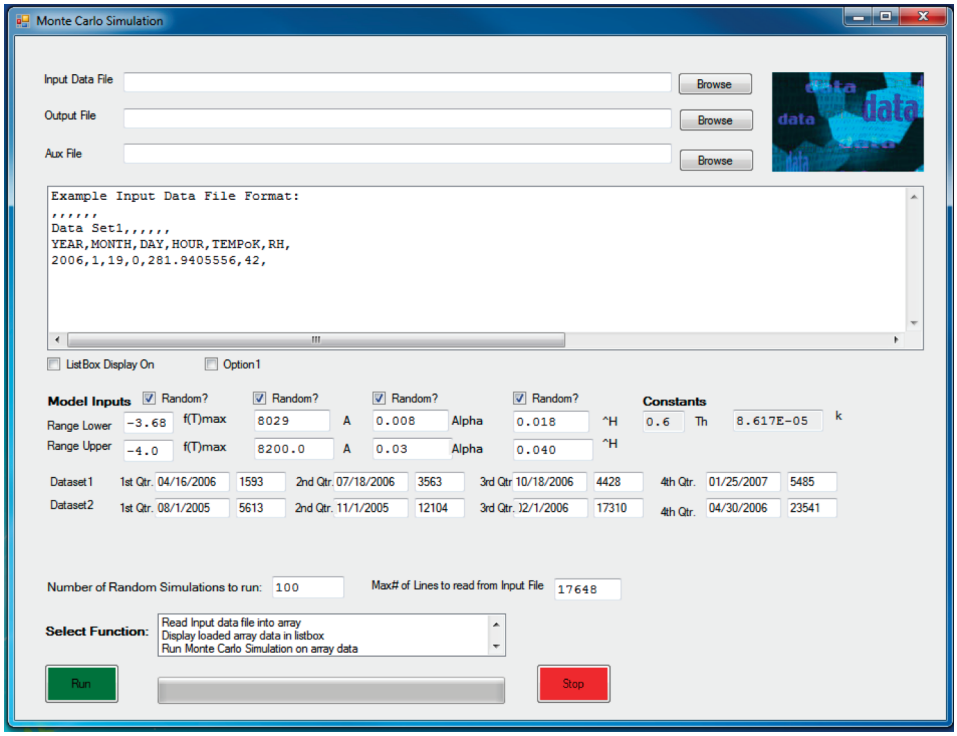


Figure 2: GUI of VB.net App in Visual Studio 2010 Express IDE

Porting the VB.net App to the Cloud

Preliminary research into Windows Azure stated that it supported VB.net and so it was assumed that it should be ready to migrate. As it turns out, supporting a language is one thing; its host environment is another. The cloud infrastructure is quite different from a desktop PC or server and therefore the existing application needed to be adapted.

As described in Introducing Windows Azure (Chappell, 2009), "To interact with the application, the user relies on a single web role instance." In this scenario, the user communicates through a web browser to a web role, which in turn communicates with the worker role (the background processing) via message queues. It turned out that the standalone VB.net application needed to be split into three loosely coupled distinct parts. They would be:

1. A web role to handle the user interface. The ASP.net application would be running under the Internet Information Services (IIS) web server.

2. A worker role with multiple instances. The VB.net module would be running under Windows Server 2008.

3. A class module with the actual business logic. The VB.net class module would be imported by the worker role.

The final architecture showing these parts is shown in Figure 3. The next step was to move the standalone VB.net application to the ASP.net application.

VB.net to ASP.net

After completing the VB.net to ASP.net port, the web role GUI would have to be written, and the domain logic would be completed in the form of a VB.net class module, leaving only the worker role/web role inter-communication code. That was the plan at the beginning of this phase, although lessons were learned as it progressed.

The basic idea was to duplicate the existing VB.net GUI functionality in an ASP.net web form. Wherever feasible, the input fields, labels, controls, names etc. had to be kept exactly the same. Also the decision was made to use the vendor's IDE, Visual Studio 2010, throughout the project, and not to code part of the project with a text editor, web design tool, or any other tool. Since the end goal was to develop a working ASP.net front end for a Windows Azure application, the supplied default templates were used initially.

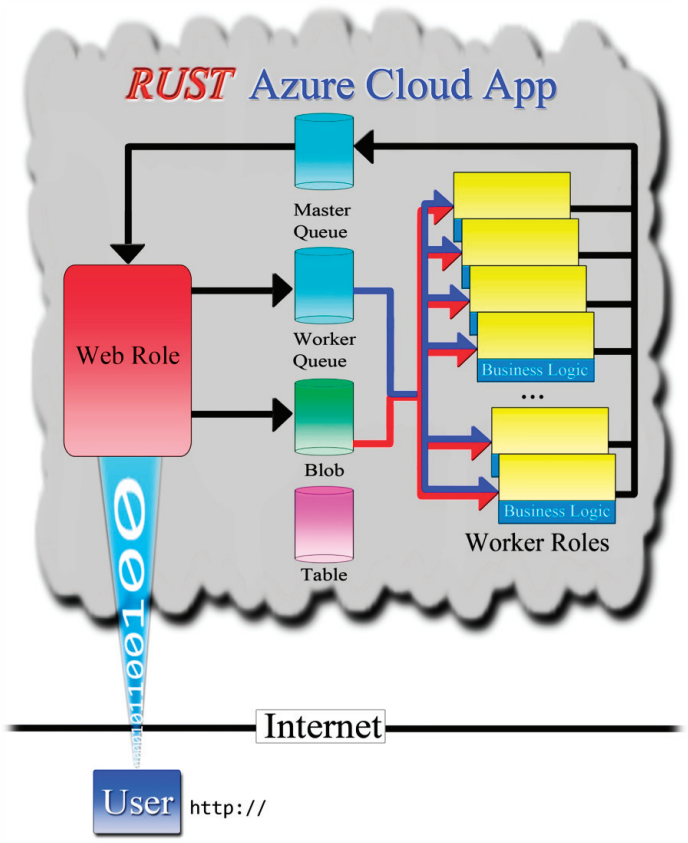


Figure 3: Topography of the Migrated Cloud App

Immediately a problem was encountered, or at the very least a decision needed to be made before getting started. Somewhat confusingly, Visual Studio offers two ways to create an ASP.net application, a web project or a web site. Based on (MacDonald, Mabbutt, & Freeman, 2010), the simpler approach was chosen to use the project-less web site. At first glance, creating a new web site with Visual Studio 2010 produced a template with these main items:

- About.aspx
- Default.aspx
- A master page, site.master
- A style sheet, site.css

Using the tool's "split view" between source and design views revealed a clean style, header, title and, most importantly, a built-in navigation bar. The idea behind the site.master was to give an ASP.net application a consistent look and feel between pages with code inherited from one central source file. Once again, the code functionality was separated as much as possible in order to aid in troubleshooting when deploying to the cloud - i.e. the default web page without any cloud specific code should come up first. The built-

in navigation code was a significant time saver and, with some minor adjustments to the style sheet, the web site was personalized with a clean look and feel.

Developing the ASP.net web site in the same language as the standalone model, and using the same IDE used for the standalone VB.net application made the task straightforward. Since this was a web application, the method of accessing the data file that the simulation ran against needed to be changed from system I/O to a file upload. However, the built-in server file upload control and example code made this task easy. As stated previously, this migration step involved splitting the domain logic Monte Carlo Simulation into a separate VB.net Class module. It was evident that the Windows Azure application would require using some type of API for the web role/worker role inter-communication, but for this interim step, a simple import statement was used to emulate that functionality.

After completing the above steps, the project was ready to be tested. The Visual Studio 2010 IDE has a built-in web server to test an ASP.net application before actual deployment, and that is what was used to verify that the ASP.net web site would produce the same results as the standalone application. Initial verification was successful (Figure 4). At this point, the project was ready for the actual migration to Windows Azure.

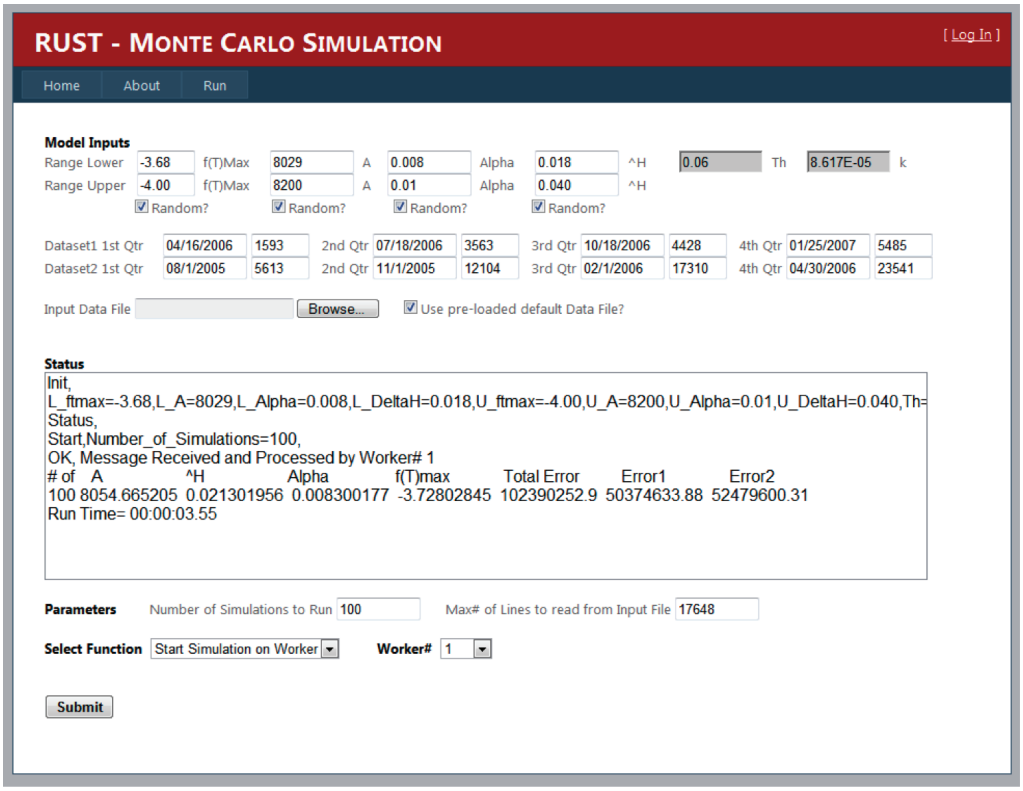


Figure 4: ASP.net Monte Carlo Simulation Tool

ASP.net to Windows Azure

Setting up the Windows Azure Development Environment

Finally the project was at the stage where it was able to begin operating on the Windows Azure platform. One issue encountered with the development platform was that during the March/April 2011 timeframe of this project the Azure SDK changed from Version 1.3 to 1.4 and the installation method changed as well. A much simpler, all-in-one installation of SDK and Tools became available in April 2011. In fact, it changed while an installation of the tools was already in progress.

The next task was to start filling in the Windows Azure Project template with the previously created ASP.net code, including the VB.net Class module. Using code from the hands-on labs of the Windows Azure platform Training Kit, a first cut of the messaging code for the web role/worker role communication was designed and coded. In the Windows Azure platform the web roles and worker roles are loosely coupled and communicate asynchronously with messages stored in storage queues. Basically, the logic of calling a VB.net function or subfunction to initiate an action was replaced with writing a message to a storage queue. After the code for this mechanism was completed, the Azure application was ready to be compiled, built, and then executed in the Azure Emulator on the local machine.

For the initial running of the web role and a skeleton of the worker role in the Azure emulator, the actual domain logic (Monte Carlo simulation code) was disconnected to enable visualization of the running Azure cloud application. The web role displayed in the browser and the output log of the Azure emulator showed one instance of the worker role up and running.

Debug and Refine the Message Queues

Again using the example code from labs of the Windows Azure platform Training Kit and various articles on Azure message techniques, the web role/worker

role communication logic was designed. All of the examples found to this point, however, used a single message queue with one-way communication from the web role (the master), to the worker role (the slave). Surprisingly, it was even stated that these queues were one-way in nature and that another process should be used to communicate back to the sender (Hay & Prince, 2010). This one-way communication was not appropriate for this particular application because the web role needed to know when the operations it gave to the worker roles were complete and what the results of those operations were.

A design using two queues was decided upon, one queue for in-bound messages (master queue) and one for out-bound (worker queue). Each message written to a queue had its own unique message ID and each worker role had its own role instance ID. It was now clear that a system could be designed to logically process the items of work. Searching the Internet revealed one simplistic example, written in C#, that used two message queues in this manner (Sawaya, 2010). Incorporating concepts from this example, the design worked with a single role instance. Next, the application was rebuilt with multiple worker role instances and was successfully able to exchange messages in a logical fashion.

Rewrite File Access for Cloud Storage

Now that there was a system in place where the web and worker roles could communicate with each other, the difference between cloud storage and local file access had to be addressed. The Monte Carlo simulation used a comma-delimited text data file consisting of approximately 17,000 lines (512Kb in size) that was loaded into an array for processing. The simulation iterated through the array one row at a time with one set of randomized variables. Successive iterations would yield another random set of variables. However, the worker role could not access the local file system as the standalone application could, hence the need to rewrite this code section for the cloud. Two options were considered to address this rewrite:

- Embed the data file as a static VB resource
- Upload the file to the web role and then write it to BLOB storage, in which:
  - The web role would send a message to worker role with BLOB ID
  - The worker role would retrieve file from BLOB storage

The second approach would be the most flexible for future implementations that had different data files. Note that these were only "possibilities". One Azure publication warned of "gotchas", such as the inability of the Azure emulator to limit local functionality to what was available in the actual cloud environment (Krishnan, 2010). In other words, it was possible to write code to access the local file system or send emails with standard methods and they would function fine while running in the local emulator but they would fail in the Cloud.

Therefore, both approaches were coded and, by default, an attempt was made to load the data file as a VB resource. An operator initiated action could start the second method of uploading the data file to the web role and using BLOB storage to store and retrieve it. The Lab examples were helpful; however they were only manipulating binary files. Searching the MSDN proved to be the answer as the BLOB does support text file storage and by using the appropriate methods and properties, the code was written. After debugging and testing both methods, all could be tied together.

Polishing the Application for the Cloud

As stated previously, this was a scientific application running Monte Carlo simulations on a dataset and producing a single final result. Therefore, a typical scenario from an operational viewpoint would be to initialize n-worker-roles with different sets of parameters, and then have them start long-running simulations. Interim status and final results would be obtained by the web role polling the worker queue for messages and presenting them in a status list box. This status display was accomplished by selecting the Status command from the function list and clicking on the Submit button. However with 10 to 20 worker roles, or more, running, the status update needed to be automated.

A better approach would be to poll the worker message queue at a timed interval and repaint only the status list box. Visual Studio 2010 proved invaluable, as this was accomplished using a combination of three ASP.NET AJAX server controls: the ScriptManager control, the UpdatePanel control, and the Timer control. This function was made optional to give the user explicit control when needed.

Testing the Completed Azure App in the Local Emulator

The Windows Azure platform emulator, running under Visual Studio, has the capability of testing locally in two stages. The first is to test an application with local code and local storage; and then the second with local code and actual cloud storage. The first stage was completed without any notable issues.

In order to test code in the local Azure emulator using the cloud storage account, it was necessary to reconfigure both

the web role and worker role configuration, specifically the properties on the Settings tab (when using Visual Studio 2010). The *DataConnectionString* and *Diagnostics.ConnectionString* settings needed to be changed from "*UseDevlopmentStorage = True*" to the storage credentials just obtained.

With the re-configuration complete and the application rebuilt and restarted, the Azure application launched and ran. Since the latency of the cloud storage was known to impact performance, the initial concern was only with testing to ensure the application still functioned properly.

Deploying the Application to the Cloud

The final step in this test project was to actually move the application to the Azure cloud platform. Having done a thorough job of testing the application in the local cloud emulator, confidence was high that the objective was about to be achieved of harnessing the power of the cloud.

The Visual Studio 2010 *Create Service Package Only* option was used to publish the Azure application and upload the Azure application files via the Windows Azure management portal. The application was deployed to the Azure staging area which presented a Globally Unique Identifier-type of URL to access the cloud application.

A mistake, but also a lesson learned, was that the application was configured for nine worker roles. The deployment time was directly proportional to the number of instances configured. Until the deployment is solid, it is more expedient to configure the minimum number of roles needed to do a preliminary test of an application.

The moment when the web role and nine worker roles turned from busy to ready, and finally ***green***, was a significant milestone. When the application *url* in the Cloud was clicked on, however, a long-spinning browser activity icon was followed by an error stating:

"Server Error - Unknown Error, Cannot display error details from a Remote Server".

The default Web page, which didn't have any Azure specific code, did not appear.

The remoteness of the cloud was evident. The application was known to work fine in the local Azure emulator and, in fact, the ASP.net application was successfully running on an external IIS Web server, so the problem was not immediately apparent. This was a difficult problem, since there wasn't a detailed error message to explain the situation. At the end of the day a decision was made to temporarily undeploy.

After searching the Internet, the conclusion was reached that the problem was really IIS and ASP.net centric. A fix was found that would enable a detailed error message to be produced. The web.config had to be modified to allow a remote server to display a detailed error message.

The next day the application was rebuilt, redeployed, and tried again. After several tries, it was properly configured, finally producing the real error message:

"Default.aspx cannot be found or does not exist".

After researching the problem and searching ASP.net/ IIS issues, the answer was found - in the three ASP.net files copied from the ASP.net application to the Azure application, "CodeFile" needed to be changed to "CodeBehind". The root cause was related to how existing, working, ASP.net files are added to an Azure project.

Application Type	Number of Simulations	Time in Minutes:Seconds
Excel Spreadsheet	5000	27:29
VB standalone	5000	08:05
ASP.net	5000	02:55
Azure emulator	5000	04:23
Azure Cloud, Extra Small 1 Worker	5000	03:30
Azure Cloud, Extra Small 5 Workers	5 x 5000 = 25,000	03:25
Azure Cloud, Extra Small 9 Workers	9 x 5000 = 45,000	03:15
Azure Cloud, Extra Small 9 Workers	9 x 25000 = 225,000	16:17
Azure Cloud, Extra Small 9 Workers	9 x 50,000 = 450,000	34:52

Table 1: Performance Metrics

At last, the web page finally launched. Local testing in the Azure emulator paid off as the application worked properly and as intended. Its performance, as shown in Table 1, was as expected - each worker role subsequently added performed their simulations in the same amount of time. The more workers that were added, the more work that was accomplished.

Conclusions and Lessons Learned

The Cloud Really Does Work

As can be seen from Table 1, the Cloud performs as advertised. Using this case study cloud application with 9 worker roles, 225,000 Monte Carlo simulations were performed in a little over 16 minutes, compared to the Excel spreadsheet which took over 20 hours to perform. In test cases, the Windows Azure cloud performed in a linear fashion, with each worker role taking between 3:15 and 3:25 to execute 5000 simulations with random variables. The limit on the trial version was a total of twenty (20) instances, but this can be greatly expanded depending on the subscription plan.

Applications Require a Sizeable Rewrite for the Cloud

Given that almost every migration scenario is different, and this case study only migrated to one cloud platform - a PaaS type, the underlying cloud architecture will dictate that other cloud vendor platforms will require applications to be rewritten to some degree. Golden states; "once they find out how difficult it is to move an application to an external cloud, their enthusiasm dwindles." (Golden, 2009).

Cloud Platforms are Changing and are Immature

During the timeframe of the migration, from March through May of 2011, the Windows Azure SDK changed from version 1.3 to 1.4. Also, the Windows Azure management portal changed and that made the deployment section of the books as well as the online documentation obsolete. These changes made it difficult for the author building a Windows Azure project for the first time.

This article is condensed from a DACS Technical Report entitled "Cloud Computing in the Goverment", which can be downloaded for free from: www.thedacs.com/techs/abstract/518136.

About the Author

Tom Kwasniewski is an accomplished Software Engineer with over thirty years in both hardware and software systems design and programming. He has worked extensively with firmware, diagnostics, communication protocols, client-server models, databases, web applications, and cloud computing applications. His professional experience includes work in the computer manufacturing arena, nuclear power industry, and defense industry and is currently a software engineer with Quanterion Solutions supporting the DACS. Tom is a Magna Cum Laude graduate of SUNYIT with BS in Computer Science.

References

Golden, 2009, Golden, B. (2009, January 22), T*he Case Against Cloud Computing*, Retrieved May 02, 2011, from CIO: http://www.cio.com/article/print/477473

Hay & Prince, 2010, Hay, C., & Prince, B. H., *Azure in Action*, Greenwich, Manning Publications

Krishnan, 2010, Krishnan, S., *Programming Windows Azure*, Sebastopol, O'Reilly Media

MacDonald, Mabbutt, & Freeman, 2010, MacDonald, M., Mabbutt, D., & Freeman, A., *Pro ASP.NET 4 in VB 2010*, New York, Apress.

Pugh, E., Kwasniewski, T., *Cloud Computing in the Government*, Data & Analysis Center for Software, 2011, http://www.thedacs.com/techs/abstract/518136

Rosenberg & Mateos, 2010 Rosenberg, J., Mateos, A., *The Cloud at Your Service*, Greenwich, Manning Publications

Sawaya, 2010, Sawaya, G. (2010, May 17), *AzureHelloWorld*. Retrieved April 12, 2011, from AzureHelloWorld-Uv-wiki: http://www.cs.utah.edu/formal_verification/mediawiki/index.php/AzureHelloWorAz

Sosinsky, 2011, Sosinsky, B., *The Cloud Computing Bible*, Wiley Publishing

The Data & Analysis Center for Software

Online Learning Center

Technical Training On Demand

Cost effective way for organizations to provide continuous learning opportunities for their employees

Accessible

Classes available 24/7/52
From:

- Home
- Office
- Travel

Accreditation Certification

- CEUs granted
- IACET accreditation supported

Affordable

- Pay one annual subscription fee ($475)
- Take as many classes as you want

Comprehensive

- 450+ classes
- 12,000 topics
- Programming & Web Development
- Course Catalog

Latest Technologies

- Java XML
- Oracle
- Server technologies
- .NET Framework

Flexible

- Fit course work into your schedule
- Self Paced
- Refresh Knowledge
- Study during lunch or after work

To view the catalog visit:
www.thedacs.com/training
For details call: 1.800.214.7921

Automating Cloud Security Authorizations

By Kaus Phaltankar

SECURITY IS A MAJOR CONCERN FOR CLOUD COMPUTING. THE REQUIREMENTS AND CHALLENGES OF A POTENTIAL SOLUTION MODEL ARE EXAMINED

This article describes the challenges and solutions for Security Authorization in a cloud based services environment. We describe Compliance as a Service (CaaS) that supports various service models as well as the deployment models in cloud-based environment. This makes the solution elastic, metered and cost effective, all characteristics of a cloud based offering. The CaaS solution addresses the complex Security Certification and Authorization (C&A) needs of Cloud based Systems. The solution should addresses various regulations such as FISMA, HIPAA as well as incorporates Information Assurance (IA) frameworks like the NIST 800-53rev3, DIACAP DODI 8500.2, FedRAMP, CNSS 1253 as well as ISO 27001

The CaaS solution can be used in private, public, hybrid or a community cloud deployment model. It fully implements the six-step Risk Management Framework (RMF) described in NIST Special Publication (SP) 800-37Rev.1. The CaaS solution can be embedded in the cloud by a Cloud Service Provider (CSP) or used by an Cloud Services Customer (CSC) for private cloud or hybrid cloud-based Systems, to certify and maintain the Security Authorization of the Information System.

Security Certification Scope of Cloud Computing Infrastructure in C&A

"Cloud computing has been defined by NIST as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or cloud provider interaction. Cloud computing technologies can be implemented in a wide variety of architectures, under different service and deployment models, and can coexist with other technologies and software design approaches. The security challenges cloud computing presents, however, are formidable, especially for public or hybrid clouds whose infrastructure and computational resources are part or fully owned by an outside party that sells those services to the general public.", NIST-Draft-SP800-144

The decision to embrace cloud-computing technology is primarily a risk-based decision, not a technology based decision for scalability and performance. Though, a CSC may transfer the risk of cloud computing to the CSP, CSC cannot transfer the accountability for maintaining the security of its System. The line of demarcation for the security responsibility is based

Security Authorization Challenges in the Cloud

Cloud computing is not a single capability, but a collection of essential characteristics that are manifested through various types of technology deployment and service models. The NIST definition of cloud computing, with three service models; Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) and four deployment models are shown in figure below:

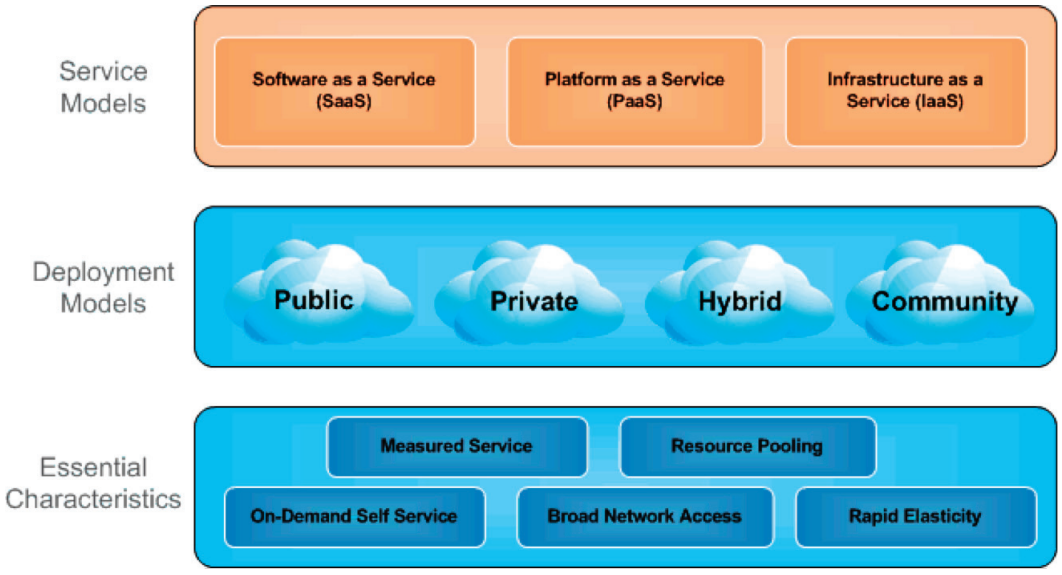


Figure 1: NIST definitions

on the services provided by the provider, e.g. a SaaS provider is responsible for the application and the entire underlying platform and the infrastructure. If the SaaS provider has obtained the underlying platform and the infrastructure as a service, the SaaS provider shall demand and inherit the security and performance SLAs from the PaaS and IaaS service providers.

FISMA mandated Security Certification and Authorization

The Federal Information Security Management Act (FISMA) of 2002 mandates each federal agency to implement a comprehensive information security program for its systems. The security programs mandated by FISMA are intended to identify and quantify threats to assets based on risk analysis as per the requirements of FIPS 199 and FIPS 200. The risk-based approach mandated by FIPS 199, categorizes each system using the key attributes of Confidentiality, Integrity and Availability. The security controls implemented on the assets are then evaluated using various the Information Assurance (IA) control frameworks such as NIST 800-53Rev.3 (mandated by FIPS 200), Federal Risk and Authorization Management Program (FedRAMP) or for DoD Systems DODI 8500.2 (DIACAP). The NIST Risk Management Framework as applied on an Information System is shown in Figure 2:

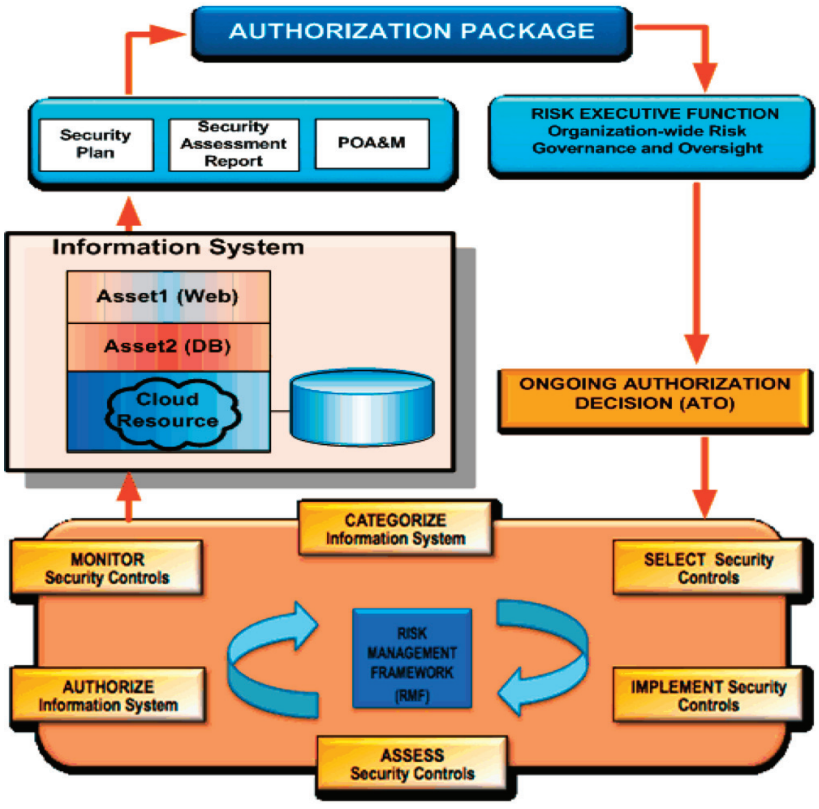


Figure 2: NIST SP800-37Rev.1 Risk Management Framework (RMF)

The biggest challenge for agencies using cloud-based solution today is to truly understand what it means to conduct C&A on a System where the System boundary and System assets are not static. In Figure 3, the Risk Management Framework (RMF) has been applied on the Information System with three assets. The Asset 1 and Asset 2 are well defined (static), while Asset 3 is a dynamic Cloud-based asset. This combination of assets creates a 'hybrid' Information System. As a Cloud based asset, the asset scope for Asset 3, is likely to change based on the performance or business continuity requirements of the client. e.g. scaling of number of Virtual Machines (VMs) or storage units. In such scenarios, how does one define the System boundary for Security Authorization and meet the requirements of Continuous Security Authorization per NIST RMF?

The line of demarcation of security responsibility based on the service model creates additional challenges for the certification process. Though, each CSP is likely to provide certification documentation for their own component, the customer of the Cloud Service has to maintain their own C&A documentation and needs to gather information from each of the underlying CSPs. The CSP provided information is then combined with the Information System specific controls into a comprehensive Certification Package, which is then submitted to the agency DAA for Authorization. All these challenges make the traditional approach to C&A non-workable in Cloud-based Systems.

In the world of tighter budgets, agencies need to automate the C&A process and deploy a solution that can address the needs of traditional systems as well as any form of cloud based systems. The automation solution needs to provide a high degree of scalability, elasticity and measurability based on a usage-based pricing model for cost effectiveness.

Solution for cloud security authorization

The CaaS solution for Cloud Security Authorization enables agencies to meet the regulatory requirements of C&A Automation and Continuous Monitoring. The solution addresses the requirements of a stand-alone, cloud-based or a hybrid System. It fully addresses the challenges laid out earlier.

Let's take the example of a 'hybrid' Information System shown in Figure 2, which utilizes say, the SaaS services of a Cloud Service Provider (Asset

3). The client Information System Owner relies on the SaaS provider to provide the security of the cloud service component (Asset 3) and inherit its controls as a component of the overall hybrid Information System. On his part, the SaaS Cloud Service Provider (CSP) has responsibility for all cloud components and shares the responsibility for the interconnection controls with the client utilizing the Provider cloud services.

The CaaS solution described next utilizes following key definitions: Systems and Subsystems, Type Authorization, Common Controls and Inheritance.

System and Subsystems

According to NIST 800-37 Rev.1, the best way to address the security certification of a 'complex' System is to decompose the System into more manageable 'Subsystems'. It further states that, *"Treating an information system as multiple Subsystems, each with its own Subsystem boundary, facilitates a more targeted application of security controls to achieve adequate security and a more cost-effective risk management process"*. The Cloud Service layers and their components form the Subsystems.

'Type' Authorization

The Cloud Subsystems are certified using a 'Type' certification. A Type certification is applied to a system consisting of a common set of hardware, software, and firmware whose instances may be deployed at multiple locations. In such cases, the IA controls can be tested at one location and the certification and authorization (C&A) of this instance can be applied to identical instances deployed at other locations. The Type certification requires that the installation environment must be described in detail, including any interconnection controls with other systems or the datacenter. The Type authorization allows the Cloud environment to scale from single VM to 1000s of VMs across multiple installation sites, meeting the elasticity requirements of the cloud services. The Type certification can be applied to any cloud service type, IaaS, PaaS, or SaaS (such as in Figure 3).

Common Controls

The NIST RMF defines three types of Security Controls applied on an Information System: a. System-Specific controls; b. Common controls; and c. Hybrid controls (i.e. controls that have both System-specific and common control characteristics). It is ideal for organization to identify and implement as many 'Common controls' as possible. The Common controls enable a cost-effective and consistent application of information security controls and help simplify the risk management process.

Inheritance

When Common controls are used to support documentation of controls for a specific Information System, they are referred by that specific System as 'inherited controls'. In the case of Cloud Computing infrastructure, each layer of the service can inherit controls from a lower layer as shown in Figure 3.

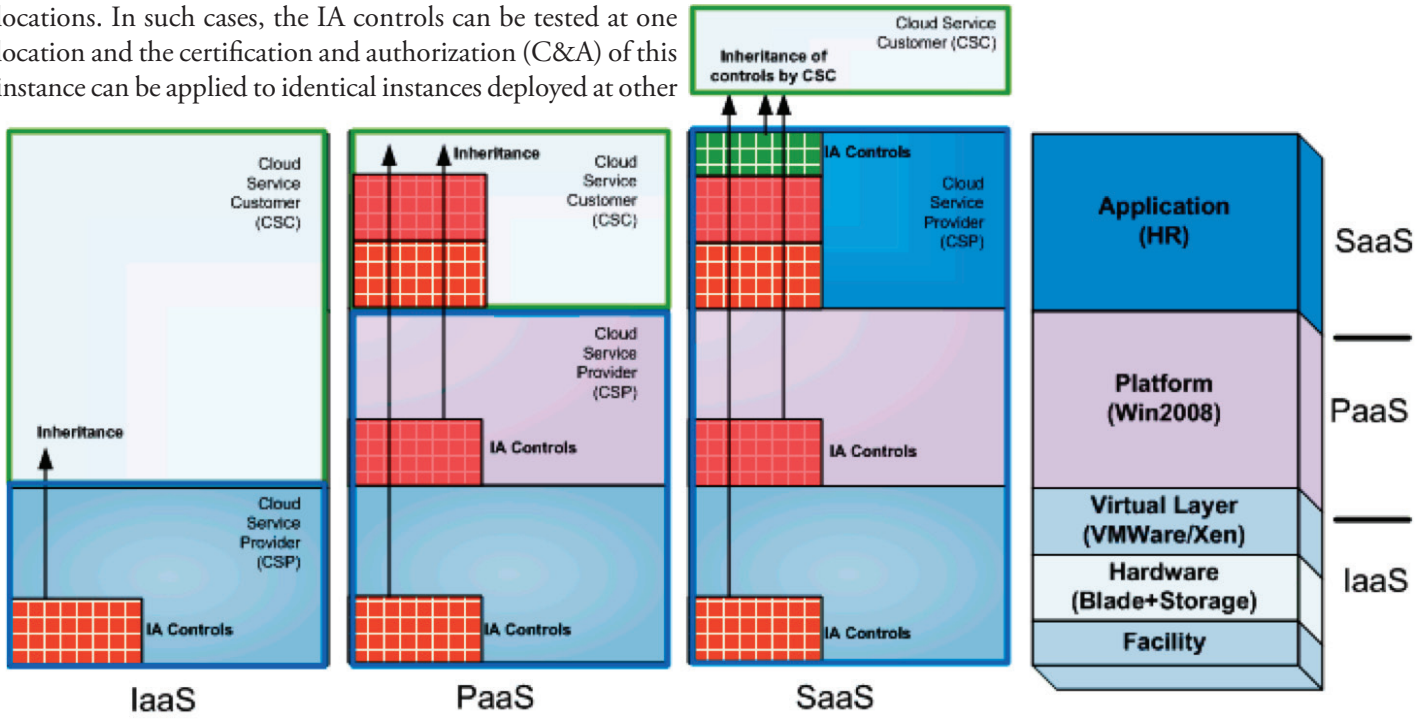


Figure 3: Cloud Services based on inheritance

In Figure 3, in the IaaS Service instance, Cloud Service Customer (CSC) inherits controls from the IaaS service from the Cloud Service Provider (CSP); while in the case of PaaS service instance, CSC inherits controls from both the IaaS and the PaaS service layers. Each 'service' specific controls are documented as a set of Common Control Profiles (CCPs) and can be inherited by Systems and Subsystems for authorization of a standalone or a hybrid Cloud-based System.

The System Security Plan (SSP) for the Information System documents the decomposition of the complex Information System into Subsystems. The SSP notes the type of certification, for example, 'Type' Certification applied to the Subsystems with description on each Subsystem's asset boundary. It also notes the list of Common Controls based on the CCPs that will be 'inherited' from an underlying Subsystem or an external System with applicable interconnection controls and Service Level Agreements.

Compliance as a Service for Cloud Computing

The CaaS solution is embedded within the Cloud infrastructure and offered to the customers of the Cloud as a C&A automation and Compliance Monitoring Service. The CaaS, provides a highly scalable and cost effective solution for both the Provider Systems as well as the Customer Systems. For customer of SaaS services, the Cloud Service Provider, has the responsibility for the security of all the layers and can inherit applicable controls from the underlying layers. The inherited Common Controls and SaaS specific controls are combined to create the C&A certification package by the Cloud Service Provider including the SSP, Security Assessment Review (SAR) or the System Test and Evaluation (ST&E) documents plus other artifacts for consumption by the customer. We take a detailed look at the SaaS service provider package based on the key definitions for CaaS described earlier.

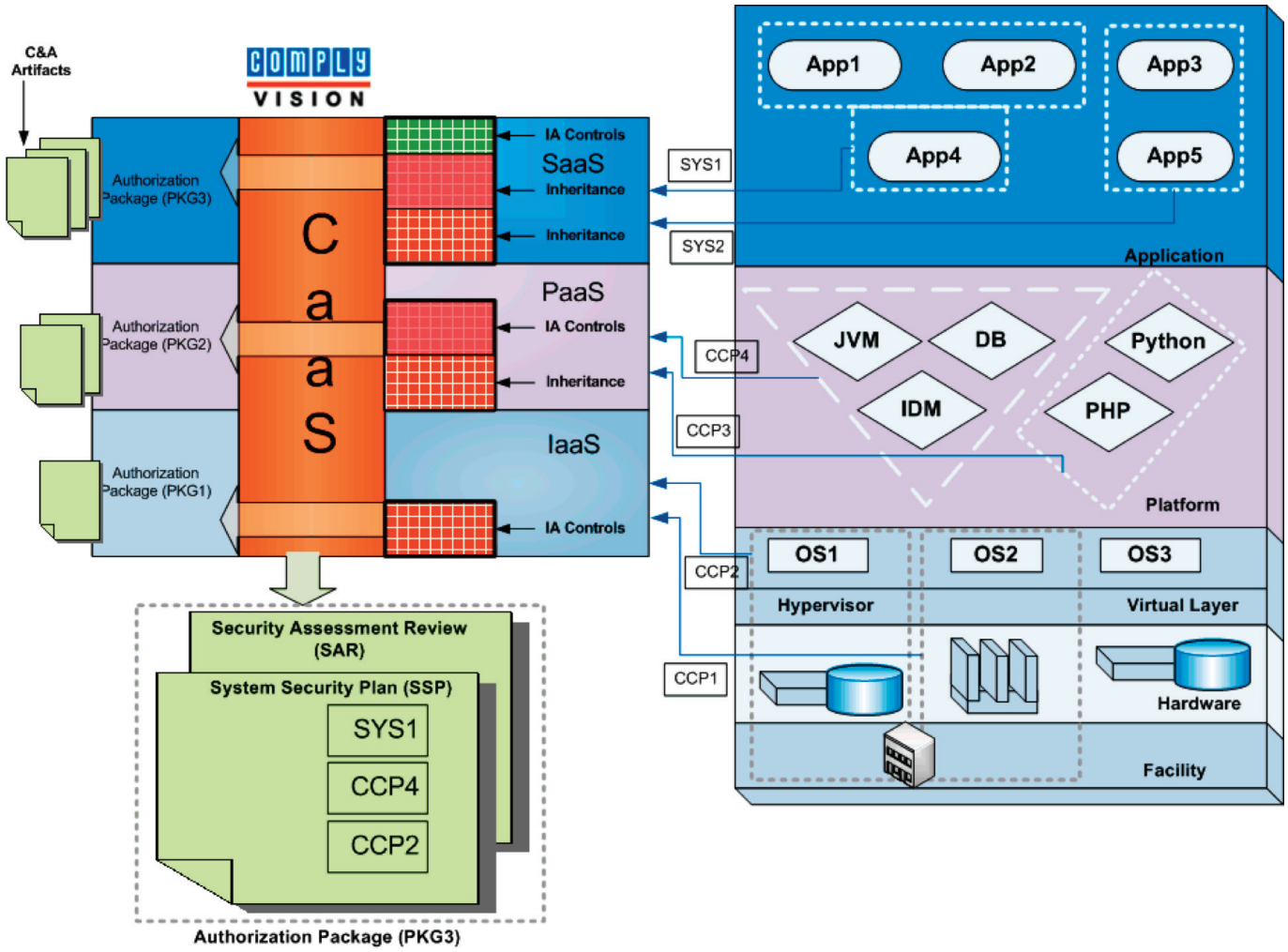


Figure 4: CaaS - C&A Package for Cloud-based SaaS

In Figure 4, the customer of SaaS services desires FISMA Certification package, PKG3. The customer, a US government agency, is subscribing to the SaaS based Human Resources application consisting of three application components, App1, App2 and App3.

The complex Human Resources SaaS can be broken into three Sub-Systems, matching each service layer. The SaaS service relies on the PaaS service components shown in the triangle with Java Virtual Machine (JVM), MySQL Database and local Identity Management System within a Linux Platform. The Linux platform is running on the IaaS components of hypervisor, rack-based server and associated storage provided by a Network Attached Storage or Storage Area Network, in a hosted facility. Each underlying IaaS and PaaS layer can have their own package, such as PKG1 and PKG2, based on a 'Type' certification and available to customers for direct consumption or by inheritance.

The combined Security Certification package, PKG3, with all the associated artifacts such as the SSP, the SAR is generated by CaaS component and is made available by the provider to the agency for the agency DAA authorization.

Compliance as a Service for Hybrid Cloud-based Systems

The CaaS solution can also be used to certify and authorize hybrid systems of an organization. It is anticipated that this will be predominant model in the early and possibly later stages of Cloud adoption. An example of this setup is given below:

In our example, the SaaS service certification package (PKG3), is incorporated into the System authorization package for the Information System located at the agency. An independent application of the NIST RMF is shown on the agency's own Information System on the right. As shown in Figure 5, the CaaS solution provides support for multiple

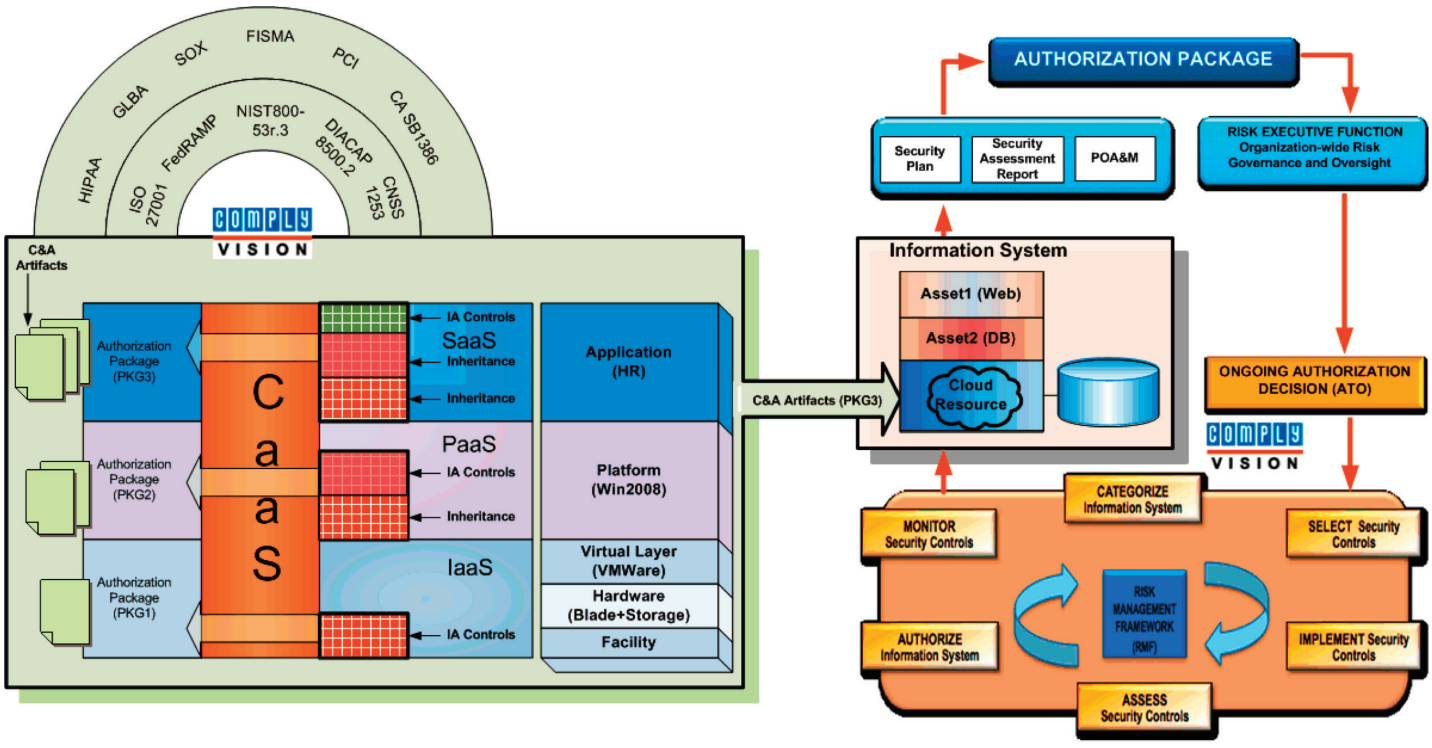


Figure 5: CaaS with a Hybrid System implementation

The Controls applicable to each component of the layers are grouped together in a layer specific CCP. The upper layer, as discussed in the previous section, can inherit the CCPs of the lower layers. In our example, the SaaS service provider inherits CCP2 from IaaS, CCP4 from PaaS services and documents application specific controls for the SaaS layer in the SYS1 profile.

Federal and State regulations as well as multiple Information Assurance frameworks from NIST to DIACAP to ISO.

The initial C&A of a System is the first step in getting the Authority To Operate (ATO), but the on-going Continuous Authorization (CA) is only possible with Continuous Monitoring of all the applicable controls. The Continuous

Monitoring requirement has been mandated by the Office of Management and Budget (OMB) and described in NIST SP800-37 Rev.1 as well as NIST SP800-137. Next, we look into the requirements and solution for Continuous Authorization.

Continuous Monitoring

The current FISMA law, and the changes it is undergoing, recognizes the interconnected nature of the Internet and agency networks. The OMB is advocating new security reporting requirements that demonstrate compliance, while emphasizing risk based compliance analysis through *Continuous Monitoring*. These requirements have to be met in the existing systems, as well as new systems being procured by an agency.

The organizations must still maintain formal authorizations and acceptance of risk but may leverage results of continuous monitoring assessments to support the ongoing ATO. The initial ATO and ongoing Continuous Monitoring are required for newly procured systems as well as for continued operation of an existing system. Continuous Monitoring encompasses everything from monitoring changes to the System asset components, Situational Awareness data from assets, to conducting ports and protocol analysis using vulnerability analysis tools and keeping the system related Plan of Action and Milestones (POA&M) updated. It also includes policy monitoring and documentation updates for annual or significant change related re-certifications.

NIST SP 800-37 Rev1 – Risk Management Framework

NIST 800-37 Rev.1 (Chapter 3, P.36) says, "*Organizations may choose to eliminate the authorization termination date, if the continuous monitoring program is sufficiently robust to provide the authorizing official with the needed information to conduct ongoing risk determination and risk acceptance activities with regard to the security state of the information system and the ongoing effectiveness of security controls employed within and inherited by the system*".

The Continuous Monitoring of a system requires compliance with three key requirements:

1. Change and Configuration Management of Assets
2. Monitoring of Security Controls using Automated Tools
3. Documentation Updates and Reporting

Change and Configuration Management of Assets

The Configuration Management controls within NIST SP 800-53rev2/3 address the needs of change and configuration management of a system with the goal of enabling and maintaining security while managing the risks on an on-going basis.

Monitoring of Security Controls Using Automated Tools

The objective of monitoring of security controls is to determine if the controls implemented or inherited by the system continue to be effective over time as the system undergoes changes. This encompasses ALL controls and not a subset. The Continuous Monitoring of controls requires monitoring of each control with varying frequencies based on:

- Control volatility – The more volatile controls need to be monitored more frequently
- Organization and system risk tolerance
- Current threat information that might affect the system

The Continuous Monitoring strategy needs to specify the monitoring frequencies along with the reporting frequency and the details required for reporting. As the size and complexity of today's system increases, it is hard to gather the required details and the frequency desired manually. The process of monitoring and reporting has to be automated using tools that provide situational awareness data in support of:

- Risk based decisions
- Evaluation of 'on-going' authorization
- Asset and configuration management to identify changes and their impact on security posture
- Reporting the system security status at any point in time

Some of the technical and operational controls lend themselves to automation much more easily than others. These controls can be evaluated and monitored using vulnerability assessment tools or event management and alerting systems that provide situational data based on logs and Simple Network Management Protocol (SNMP) alerts.

The supplemental guidance for the Continuous Monitoring control (CA-7) from NIST 800-53 Rev.3 states that, "*A continuous monitoring program allows an organization to maintain the security authorization of an information system over time in a highly dynamic environment of operation with changing threats, vulnerabilities, technologies, and missions/business processes. Continuous monitoring of security controls using*

automated support tools facilitates near real-time risk management and promotes organizational situational awareness with regard to the security state of the information system. The implementation of a continuous monitoring program results in ongoing updates to the security plan, the security assessment report, and the plan of action and milestones - the three principal documents in the security authorization package. A rigorous and well executed continuous monitoring program significantly reduces the level of effort required for the reauthorization of the information system. Continuous monitoring activities are scaled in accordance with the impact level of the information system."

The Situational Awareness data from an asset needs to be mapped to the System that includes this asset in its system definition and the Information Assurance control that may have been affected. This provides data to the system owner and the DAA on evaluating the risk to the Information System. The remediation action may then be documented as a task or a POA&M item.

Documentation Updates and Reporting

The critical documents such as the SSP, SAR or the ST&E report and the POA&M, need to be updated and kept current as per the Continuous Monitoring process. These three key documents and the supporting artifacts are required for any authorizing official to conduct their evaluation of risk, and granting of continued authorization. The POA&M reports are especially critical, as they need to be made available to OMB upon request or at least quarterly.

There is a significant amount of effort involved in keeping these detailed information documents updated and reflect the current state accurately. Just as the asset information becomes potentially stale within twenty-four hours of documentation, the SSP and SAR documents are equally prone to become stale in a short order. An automation tool is the only way to keep these extensive (sometime 200-300 pages) documents updated on an on-going basis.

The SSPs need to reflect changes to the System and its assets based on Change and Configuration Management process as well as any updates to the control implementation, while the SAR/ST&E documents need to reflect the testing and validation of the controls to identify any risks introduced due to these changes. Any new risk identified or an impact on the existing POA&M item needs to be reflected and reported to the authorizing official and OMB.

About the Author

Kaus Phaltankar, as a Subject Matter Expert in Cyber Security, brings more than 15 years of Internet, network engineering, and operations experience from leading corporations including Hewlett Packard, MCI and Citicorp. As the founder of multiple companies including ViewTrust Technology and NetPlexus Corporation, Kaus has been a pioneer in developing innovative technologies to meet challenges related to security and compliance requirements of commercial and federal enterprises. He has implemented network and security solutions for Fortune 100 clients globally.

Kaus holds Masters in Telecommunications and Computer Science from The George Washington University, Washington DC; is a Certified Information Systems Auditor (CISA) and Certified Information Systems Security Professional (CISSP).

Two Great Reliability Solutions from the RIAC & DACS.

System Reliability Toolkit

RIAC

The RIAC/DACS System Reliability Toolkit provides technical guidance in all aspects of system reliability, allowing the user to understand and implement techniques to ensure that system and product designs exhibit satisfactory hardware, software and human reliability, and to minimize the inherent risks associated with deficiencies in system reliability.

To purchase, please contact:
The Reliability Information Analysis Center
100 Seymour Road
Suite C-101
Utica, NY 13502
1-877-363-RIAC
http://theRIAC.org

The DACS Software Reliability Sourcebook

DACS
Data & Analysis Center for Software

This first edition of the DACS Software Reliability Sourcebook provides a concise resource for information about the practical application of software reliability technology and techniques. The Sourcebook is divided into nine major sections, plus seven supporting appendices.

To purchase, please contact:
The Data & Analysis Center for Software
100 Seymour Road
Suite C-102
Utica, NY 13502
1-800-214-7921
http://thedacs.com

One Side Now: The Need to Adopt a Business Systems Approach to Cloud Security

By Larry Clinton

THE VIEW OF CLOUD COMPUTING, FOCUSING ON SECURITY AND EFFICIENCY, FROM A BUSINESS PERSPECTIVE

Joni Mitchell's beautiful lament from forty-five years ago is timely today as an apt description of the unspoken truth of many enterprise managers who have put their corporate IP and other data "in the cloud."

> *"I've looked at clouds from both sides now. Fom up and down and still somehow, Its clouds illusions I recall, I really don't know clouds, at all."*
>
> *"Both Sides Now" Joni Mitchell*

What could be a better explanation for the stunning finding of the 2011 PricewaterhouseCoopers Global Information Security Survey which found that 62% of the information security experts polled had "little or no faith in the security of the cloud" – including 49% who had already put their information there? [1]

One major reason for this remarkable finding maybe that right now many enterprises are not looking at cloud computing from "both sides," but rather only the "up" side of such a deployment and not fully appreciating the "down" side.

The result is that cloud deployments may provide the illusion that they are as "solutions" to an enterprise's IT issues. They are not. Cloud deployments, which in many cases may be wise and even necessary options, are tactics that must be understood and analyzed from a full business systems perspective. In short, we need to look at clouds from both sides, now.

The Digital Imperative to Reduce Security

What could prompt presumably reasonable and competent corporate managers to knowingly place their corporate data in a place where they have little or no faith in its security?

Money.

Of course making money is the primary job of most enterprises. Moreover, the pressure on enterprises to be efficient and profitable is only magnified by the increasingly competitive world economy.

By now virtually every enterprise of even moderate size has plowed into its business plan the efficiencies and growth opportunities related to digitalization such as improved product and personnel tracking, remote workforces or web based marketing. However, many organizations still fail to account for the downside of the digital revolution – cyber security.

In fact, despite seemingly continuous reports of more, and more severe, cyber attacks recent surveys have documented that many – perhaps most – enterprises have been deferring or reducing their investments in cyber security in recent years. [2]

The tradeoff between efficiency and economy benefits of digitalization at the cost of security considerations is not a phenomenon confined to the emergence of cloud computing.

For example, deploying unified communications (UC) platforms such as the Voice over Internet Protocol (VoIP) yield substantial cost savings but "while unified communications offer a compelling business case, the strength of the UC solutions in leveraging the internet is also vulnerability. Not only are UC solutions exposed to the security vulnerabilities and risk that the Internet presents, but the availability and relative youth of UC solutions encouraged malicious actors to develop and launch new types of attacks." [3]

In addition business strategies that optimize customer intimacy and supply chains require companies to connect to vendor and customer networks. While tighter integration with business partners provide clear business benefits, it also means the ability to defend against attacks depends on your partner's or customer's security capabilities and policies.

As CIO Magazine reported when analyzing the 8th annual PricewaterhouseCoopers Global survey "customers want to spend their money on-line and use more fancy apps to do it…So you have to guard against vulnerabilities attackers can exploit to steal your customer's private data and core assets…. Increasingly complex business relationships are forcing you to give outsiders access to your internal systems. You need protection from an attack against a business partner that might spill over to your network. [4]

A similar issue arises with respect to cloud computing. By now the efficiency and economy benefits of cloud options are well known. You pay only for what you need, potential cost savings within your own IT department, computer time takes the path of long distance telephone service (free computer time for everyone), new business models emerge and venture capital assumptions change since there is less capital upfront costs for computing!

Just like VoIP a few years ago and the ongoing extension of IT supply chains, cloud computing has emerged as one of the hottest developments in information technology, largely driven by perceived economic benefits ranging from cost savings and efficiencies. [5] And like the VoIP deployment and extended network relationships, security may be undermined because of competitive pressures driving these cost efficient strategies.

The security issues inherent in moving to the cloud have also been abundantly debated – at least in IT circles. As with any large issue broad generalities are questionable. At least for some, such as many small and mid-sized firms that were not spending substantially on their own security, movement to the cloud may well enhance security by providing them access to systems and personnel they could not previously afford.

Regardless of the pro-security arguments of cloud advocates, the current consensus is to be wary of security in the cloud. Although cloud services may seem fairly straight forward to the user, they are actually fairly complex relationships not only between the client and the vendor, but possibly several different vendors. Applications that had previously been managed from behind a corporate firewall may now be exposed over the Internet and out of the control of the data owner. Indeed, determining exactly where your data is may be quite difficult in a cloud configuration raising new and as yet unanswered problems not only for data owners but for regulators and law enforcement.

So called "insider threats" are also more challenging since rogue subscribers can buy their way into the cloud and launch their attacks with a level of system access that would have been prevented in traditional models. This possibility leads cloud providers to resist sharing details about their security and privacy procedures thus making it more difficult to receive the the security assurances that an organization might require from a traditional IT provider.

Perhaps it is not surprising that even the US federal government driven by its own increasingly demanding financial requirements, has announced a "cloud first" policy targeting a full quarter of federal IT spending – 20 billion dollars – for migrating to cloud computing solutions while acknowledging, but not resolving, the security issues. [6]

Approaching Cloud Security from a Business Systems Perspective

One of the biggest conundrums facing cyber security policy makers is the puzzling question as to why the issue is not self correcting. If it is true that enterprises are losing, or at least apparently at risk for losing, so much due to poor cyber security, why are they not increasing their investment in security practices and technologies sufficient to solve the problem?

A number of recent articles have attempted to address this conundrum pointing out that there is a comparatively low value placed on security by consumers. [7] The interconnected nature of cyber systems dislocates the harms for vulnerabilities from the source of the vulnerability [8] that some attacks, such as those perpetrated by state-sponsored groups may be unresponsive to commercial economic concerns [9] as well as other variables.

However, one argument that may have particular relevance to the case of cloud computing is that many enterprises are structured on a 20th century model that is inconsistent with horizontal nature of modern cyber based enterprises. These advocates, such as the Internet Security Alliance [10] and the American National Standards Institute, suggest that these structural flaws may lead organizations to underestimate the true economic nature of the cyber threat because they see cyber security as primarily an "IT" issue wherein the appropriate metrics for measuring the effect of cyber failures are things like "downtime" and IT repair costs. In many cases downtime and repair costs are minimal compared to the the real economic threat due to loss of corporate IP and brand loyalty and other factors not generally considered as part of the "IT" security conversation.

Moreover, the single biggest category of cyber attacks are not hackers breaking in from the outside, but insiders who may have access to the technological controls. As a result human resource management may be as important as software upgrades in enhancing cyber security, but such countermeasures may well be underappreciated in a security model presumed to be primarily technical. In addition legal, purchasing, finance and procurement departments may place cyber security at lower priority in completing their jobs since it is "IT's problem" and thus may unwittingly undermine otherwise sound policies.

In truth, cyber security is not really an "IT" issue, but rather an enterprise wide risk management issue that needs to be managed not by the CIO or CISO (although obviously they need to be central to the discussion) but by a cross organizational team, complete with a cross organizational budget. In addition to representation from IT and finance, this team needs to include contributions from the HR, communications, legal/compliance as well as risk management departments. Moreover, the group needs to be headed by someone with cross organizational authority such as the CFO, the CRO or even CEO, or at minimum someone reporting directly to these senior – enterprise wide – officers.

Research conducted by ANSI and ISA over a two year period illustrates that these disparate portions of the organization not only have substantial impact on the strength of the cyber system they all use, but have clearly different perspectives on the issue and roles to play in analyzing it. Only if the full organizational system which uses the system is involved in the enterprise protection strategy with a full understanding of the issues will appropriate solutions be realized [11].

In a recent article, Claude R. Baudoin illustrates the problem created when the business units drive to adopt cloud strategies in order to achieve their narrow, though valid, departmental goals, and the complications this creates for organizational governance:

"While IT itself may not fully understand these architecture issues, there are definitely things that the business overlooks when they read about how easily they can now procure services in the cloud…. and the very organization that should arbitrate between the business, IT, and the suppliers may lack the knowledge to do so well…The executives should help IT 'sit at the adults table' in order to help govern the entire enterprise, not just because IT holds the critical asset—information. For example, issues with serious legal and financial impact such as data resiliency, can become uncontrollable if cloud sourcing is done improperly. Therefore, the CFO and the legal counsel cannot just tell the CIO to go away and make it happen. Instead this should be a collegial management effort." [12]

The ISA-ANSI model goes beyond even Baudoin's suggestion that IT needs to be part of an isolated discussion arguing that the issues generated by adopting a cloud option need to be considered on a enterprise wide system basis. The strategic model laid out in The Financial Management of Cyber Risk [10] suggests a six step process:

- Cross-departmental officers "own the problem"
- Appoint a cross-departmental cyber risk team
- Meet regularly
- Develop and maintain a cross-departmental cyber risk magement plan
- Develop and adopt a total cyber risk budget
- Implement, analyze, test, and review feedback

Happily there has been a substantial increase in the number of firms that are beginning to look at these issues in following the enterprise wide model. In 2008 a Carnegie Mellon study reported that only 17% of enterprises have a cross organizational cyber risk team. However when that same question was asked in 2010 the results jumped up to 65%. [13] And the 2011 Global Information Security Survey noted increases in the number of CFOs and COOs taking leadership roles in cyber security (up from 11% to 15%) [14]. However even with these gains, about a third of studied enterprises lack a cross organizational security team and the vast majority are still not led by senior managers with financial impact to justify broader security investment. Only when we look at clouds from both sides – indeed all sides – will we see through clouds illusions and make decisions on an enterprise wide risk management basis.

About the Author

Larry Clinton is President of the Internet Security Alliance (ISA). ISA is a multi-sector trade association with membership from virtually every one of the designated critical industry sectors. The mission of the ISA is to combine advanced technology with economics and public policy to create a sustainable system of cyber security.

Mr. Clinton is known for his ability to take the complicated issues in this space and explain them clearly to a wide range of audiences – professional, policy makers and the general public. He has been featured in mass media such as USA Today, the PBS News Hour, the Morning Show on CBS, Fox News, CNN's Situation Room, C-SPAN, and CNBC. He has also authored numerous professional journal articles on cyber security. This year he has published articles in the Cutter IT Journal, the Journal of Strategic Security and the Journal of Software Technology.

Mr. Clinton is regularly called upon to testify before both the U.S. House and Senate. In 2008 ISA published its Cyber Security Social Contract which is both the first and last source cited in the Executive Summary of President Obama's Cyber Space Policy Review, which also cited more than a dozen ISA's white papers –far more than any other source.

The ISA's pro-market, anti regulatory approach to cyber security is outlined in its numerous publications including the ISA Cyber Security Social Contract and the Financial Management of Cyber Security which were written by the ISA Board of Directors and edited by Mr. Clinton.

References

[1] PricewaterhouseCoopers, Respected, but still restrained, findings from the 2011 Global State of Information Security Survey, 2011.

[2] PricewaterhouseCoopers, The Global State of Information Security, 2008. Center for Strategic & International Studies, In the Crossfire: Critical Infrastructure in the Age of Cyber War, 2010.

McAfee, Unsecured Economies: Protecting Vital Information, January 2009.

[3] Internet Security Alliance, Navigating Compliance and Security for Unified Communication, 2009 p21.

[4] "Business partners with Shoddy Security; Cloud Providers with Dubious Risk Controls; What's a CIO to do?" CIO Magazine October 2010

[5] Yoo, Christopher S., Cloud Computing: Architectural and Policy Implications, (Technology Policy Institute), January 2011 p6.

[6] Vivek Kundra "Federal Cloud Computing Strategy" February 8, 2011

[7] Friedman, Allan, "Economic and Policy Frameworks for Cybersecurity Risks, Center for Technology Innovation at Brookings, July 21, 2011.

[8] The Chapter entitled Cyber Security Social Contract by Larry Clinton will appear in Scott Jasper, ed. Conflict and Cooperation in the Commons, forthcoming from Georgetown University Press (2012).

[9] Clinton, Larry, A Relationship on the Rocks: Industry-Government Partnership for Cyber Defense, Journal of Strategic Security, Volume 4, #2, Summer 2011.

Clinton, Larry, A Theory to Guide US Cyber Security Policy, Cutter IT Journal, May 2011.

[10] American National Standards Institute and the Internet Security Alliance, The Financial Impact of Cyber Risk: 50 Questions every CFO Should Ask, 2008.

[11] American National Standards Institute and Internet Security Alliance, "The Financial Impact of Cyber Risk: An Implementation Framework for CFO's", 2010.

[12] Baudoin, Claude R., How the cloud Impacts IT Governance, Cutter IT Journal Vol 24 # 7 July 2011

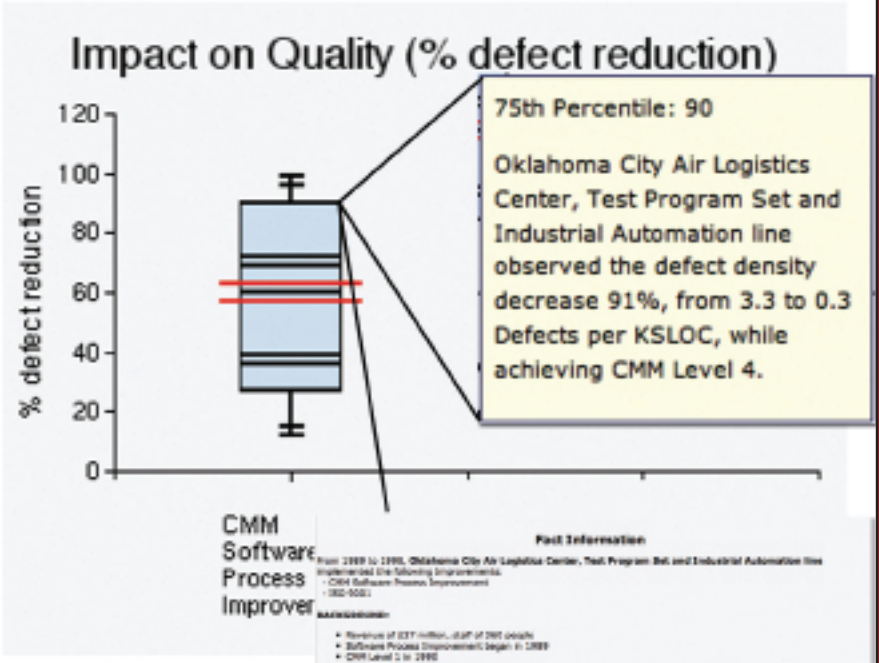
[13] Westby, Jody, Governance of Enterprise Security: CyLab 2010 Report, Carnegie Mellon CyLab, June 15, 2010.

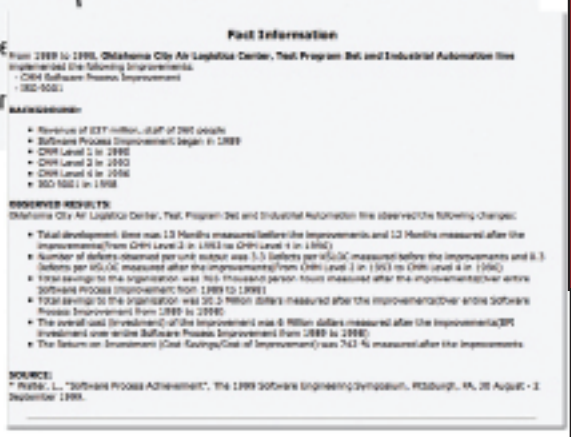
Westby, Jody & Power, Richard, Governance of Enterprise Security: CyLab 2008 Report, Carnegie Mellon CyLab, December 1, 2008.

[14] PricewaterhouseCoopers, Respected-but still restrained, findings from the 2011 Global State of Information Security Survey, 2011.

ARE YOU GETTING THE MAX FROM YOUR SOFTWARE INVESTMENT?

The DACS ROI Dashboard





Access the DACS ROI Dashboard!

http://www.thedacs.com/databases/roi/

Technologies Covered:

- SEI/CMM/CMMI
- SEI Team Software Process (TSP)
- SEI Personal Software Process (PSP)
- Inspections
- Reuse
- Cleanroom

And Many More!

Graphs Showing Impact of Software Technologies on:

- ROI
- Productivity
- Quality

Summarizes Facts from Open Literature

The Data & Analysis Center for Software
P.O. Box 1400
Rome, NY 13442-1400
http://www.thedacs.com

Introducing...

...the newest member of the Quanterion Automated Reliability Toolkit family...

QuART ER
Enhancing Reliability
Quanterion Automated Reliability Toolkit

››› With over two dozen new or improved tools!

		QuART	QuART PRO	QuART ER
		$99	$189	$399
General Knowledge	Reliability Advisor	No	Yes	Yes
	Reliability Program Cost	No	Yes	Yes +
	Reliability Imp rovement Cost	No	Yes	Yes
	Warranty Cost	No	Yes	Yes
	Statistical Distributions	No	Yes	Yes +
	Checklists	No	Yes	Yes
	Definitions	Yes	Yes	Yes
	Acronyms	Yes	Yes	Yes
Concept	Reliability Potential	Yes	Yes	Yes
	Reliability Allocation	No	No	Yes
	Reliability Tailoring	Yes	Yes	Yes
	Reliability Approach Assessment	No	No	Yes
Design Analysis	Derating	Yes	Yes	Yes
	Thermal Design	No	Yes	Yes
	Failure Modes	No	Yes	Yes +
	FMECA Worksheets	No	No	Yes
	Material Durability Improvement	No	No	Yes
	Parts Count Analysis	Yes	Yes	Yes
	Software Reliability Prediction	No	No	Yes
	Redundancy Modeling	Yes	Yes	Yes +
	Reliability Adjustment Factors	Yes	Yes	Yes
	Interference Stress Strength Analysis	No	No	Yes
	Availability Calculator	No	No	Yes
Testing	Benefit of Reliability	No	No	Yes
	Reliability Growth Testing - Duane Method	Yes	Yes	Yes
	Reliability Growth Testing - AMSAA-Crow Method	No	No	Yes
	Reliability Growth Testing - Crow Extended Method	No	No	Yes
	Reliability Demonstration Test Equal Risk RDT	Yes	Yes	Yes
	Reliability Demonstration Test Variable Risk RDT	No	No	Yes
	Reliability Test Cost	Yes	Yes	Yes
	Accelerated Reliability Test	Yes	Yes	Yes +
	WeiBayes Substantiation Testing	No	No	Yes
	Test Results / Confidence Exponential Distribution	Yes	Yes	Yes
	Test Results / Confidence Binomial Distribution	No	Yes	Yes
Production	Design of Experiments 2 Factor DOE	No	No	Yes
	Design of Experiments 3 Factor DOE	No	No	Yes
	Design of Experiments 4 Factor DOE	Yes	Yes	Yes
	HASS - ESS	No	Yes	Yes
	HASA	No	Yes	Yes
	Sampling Plans	No	No	Yes
	Process Capability	No	No	Yes
Field	Weibull Analysis	Yes	Yes	Yes +
	Optimal Replacement Interval	No	No	Yes
	Sparing Analysis (Graphical)	Yes	Yes	Yes
	Spares Analysis (Tabular)	Yes	Yes	Yes +

+ indicates improved features or functions

Download for only $399

QUANTERION
SOLUTIONS INCORPORATED

quanterion.com
(315)-732-0097/(877) 808-0097

Affordable tools for improving reliability...

Quanterion Solutions in a team member of the operation of the Reliability Information Analysis Center (RIAC)

This is a paid advertisement.

Article Submission Policy

The DACS Journal is a theme-based quarterly journal. In the past DACS has typically solicited specific authors to participate in developing each theme, but we recognize that it is not possible for us to know about all the experts, programs, and work being done and we may be missing some important contributions. In 2009 DACS adopted a policy of accepting articles submitted by the software professional community for consideration.

DACS will review articles and assist candidate authors in creating the final draft if the article is selected for publication. Note that DACS does not pay for articles published. Note also that submittal of an article constitutes a transfer of ownership from the author to DACS with DACS holding the copyright.

Although the DACS Journal is theme-based, we do not limit the content of the issue strictly to that theme. If you submit an article that DACS deems to be worthy of sharing with the community, DACS will find a way to get it published. However, we cannot guarantee publication within a fixed time frame in that situation. Consult the theme selection page and the Author Guidelines located on the Journal web site (see **https://journal.thedacs.com/**) for further details.

To submit material (or ask questions) contact **news-editor@thedacs.com**

Recent themes include:

- Earned Value
- Software Testing
- Project Management
- Model Driven Development
- Software Quality and Reliability
- Cyber Security

ABOUT THE JOURNAL OF SOFTWARE TECHNOLOGY

DACS JOURNAL EDITORIAL BOARD

Distribution Statement:
Unclassified and Unlimited

DACS
100 Seymour Road
Utica, NY 13502-1348
Phone: 800-214-7921
Fax: 315-732-3261
E-mail: news-editor@thedacs.com
URL: http://www.thedacs.com/

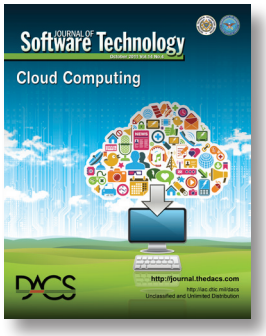
ADVERTISEMENTS

The DACS Journal is now accepting advertisements for future newsletters. In addition to being seen by the thousands of people who subscribe to a paper copy, an electronic version is available at the DACS website, exposing your product, organization, or service to hundreds of thousands of additional eyes every month.

For rates and layout information contact: news-editor@thedacs.com

COVER DESIGN

Shelley Howard
Graphic Designer
Quanterion Solutions, DACS



ARTICLE REPRODUCTION

Images and information presented in these articles may be reproduced as long as the following message is noted:

"This article was originally published in the DACS Journal of Software Technology, Vol.14, No.4 October 2011."
Requests for copies of the referenced journal may be submitted to the following address:

Data & Analysis Center for Software
100 Seymour Road
Utica, NY 13502-1348

Phone: 800-214-7921
Fax: 315-732-3261
E-mail: news-editor@thedacs.com
An archive of past newsletters is available at **https://journal.thedacs.com.** In addition to this print message, we ask that you notify DACS regarding any document that references any article appearing in the *DACS Journal.*

ABOUT THIS PUBLICATION

The DACS Journal of Software Technology is published quarterly by the Data & Analysis Center for Software (DACS). The DACS is a DoD sponsored Information Analysis Center (IAC), administratively managed by the Defense Technical Information Center (DTIC). The DACS is technically managed by Air Force Research Laboratory, Rome, NY and operated by Quanterion Solutions Incorporated.

Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the DACS. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the DACS, and shall not be used for advertising or product endorsement purposes.

Data & Analysis Center for Software
100 Seymour Road
Suite C-102
Utica, NY 13502

Return Service Requested

Journal of Software Technology 14-4 October 2011: Cloud Computing

IN THIS ISSUE