Cutter
IT Journal

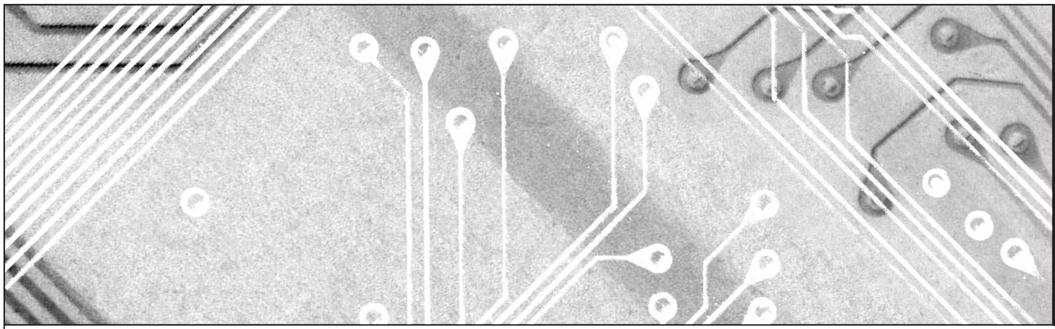
The Journal of
Information Technology Management

Vol. 28, No. 10
October 2015

> "IaaS is here to stay in many organizations, especially smaller firms and startups, and it will be the dominant form of corporate IT infrastructure in the coming years."
>
> — Vince Kellen,
> Guest Editor

IaaS: Ready for Liftoff?

CUTTER
CONSORTIUM



Cutter IT Journal

About Cutter IT Journal

Part of Cutter Consortium's mission is to foster debate and dialogue on the business technology issues challenging enterprises today, helping organizations leverage IT for competitive advantage and business success. Cutter's philosophy is that most of the issues that managers face are complex enough to merit examination that goes beyond simple pronouncements. Founded in 1987 as *American Programmer* by Ed Yourdon, *Cutter IT Journal* is one of Cutter's key venues for debate.

The monthly *Cutter IT Journal* and its companion *Cutter IT Advisor* offer a variety of perspectives on the issues you're dealing with today. Armed with opinion, data, and advice, you'll be able to make the best decisions, employ the best practices, and choose the right strategies for your organization.

Unlike academic journals, *Cutter IT Journal* doesn't water down or delay its coverage of timely issues with lengthy peer reviews. Each month, our expert Guest Editor delivers articles by internationally known IT practitioners that include case studies, research findings, and experience-based opinion on the IT topics enterprises face today — not issues you were dealing with six months ago, or those that are so esoteric you might not ever need to learn from others' experiences. No other journal brings together so many cutting-edge thinkers or lets them speak so bluntly.

Cutter IT Journal subscribers consider the *Journal* a "consultancy in print" and liken each month's issue to the impassioned debates they participate in at the end of a day at a conference.

Every facet of IT — application integration, security, portfolio management, and testing, to name a few — plays a role in the success or failure of your organization's IT efforts. Only *Cutter IT Journal* and *Cutter IT Advisor* deliver a comprehensive treatment of these critical issues and help you make informed decisions about the strategies that can improve IT's performance.

Cutter IT Journal is unique in that it is written by IT professionals — people like you who face the same challenges and are under the same pressures to get the job done. *Cutter IT Journal* brings you frank, honest accounts of what works, what doesn't, and why.

Put your IT concerns in a business context. Discover the best ways to pitch new ideas to executive management. Ensure the success of your IT organization in an economy that encourages outsourcing and intense international competition. Avoid the common pitfalls and work smarter while under tighter constraints. You'll learn how to do all this and more when you subscribe to *Cutter IT Journal*.

Cutter IT Journal®

Editor Emeritus: Ed Yourdon
Publisher: Karen Fine Coburn
Group Publisher: Chris Generali
Managing Editor: Karen Pasley
Production Editor: Linda M. Dias
Client Services: service@cutter.com

Cutter IT Journal® is published 12 times a year by Cutter Information LLC, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA (Tel: +1 781 648 8700; Fax: +1 781 648 8707; Email: citjeditorial@cutter.com; Website: www.cutter.com; Twitter: @cuttertweets; Facebook: Cutter Consortium). Print ISSN: 1522-7383; online/electronic ISSN: 1554-5946.

Subscription rates are US $485 a year in North America, US $585 elsewhere, payable to Cutter Information LLC. Reprints, bulk purchases, past issues, and multiple subscription and site license rates are available on request.

Opening Statement

by Vince Kellen, Guest Editor

The uncertain future of cloud computing seems to have stabilized among IT leaders with the acceptance of infrastructure as a service (IaaS). IaaS is here to stay in many organizations, especially smaller firms and startups, and it will be the dominant form of corporate IT infrastructure in the coming years. With the breaking of Moore's law (at least for the foreseeable future and with regard to silicon-based computing), IaaS may provide everyday computing cost benefits thanks to the efficiencies large-scale dedicated vendors can provide.

But many organizations, most notably the larger and more conservative companies, are still on the fence about moving their infrastructure to an IaaS model. IaaS can be deployed in different models, including on-premises and off-premises, managed by a third party or managed internally, and/or using private or public cloud environments.

Which model(s) should an organization adopt? How do firms know if moving their hardware, software, servers, storage, and other infrastructure components to a third-party provider is right for them? Should they consider IaaS only for temporary or experimental workloads? Also worth considering and planning for are the technical/security risks, scalability, and legal/contract issues that are critical to a successful IaaS platform deployment. In this issue of *Cutter IT Journal*, our authors share their insights on the issues organizations should contemplate before moving to IaaS.

IN THIS ISSUE

We begin the issue with an article by Cutter Senior Consultant James Mitchell and his colleague Frank Khan Sullivan who offer CIOs and their organizations advice on running a successful cloud RFP. Using a dining analogy to demonstrate why a multi-vendor approach provides more flexibility, reliability, and lower costs, they ask, "Are you worried that even if you went to that single restaurant that had everything you wanted on the menu, the quality might degrade over time?... No single cloud vendor is good at everything." Observing that cloud's cost-cutting promise depends on sharing, they counsel organizations to specify their requirements "in the middle of what the market offers," as overly specific requirements and the attendant customization will drive up costs. They also suggest that working with a cloud financial broker can help organizations maximize their vendor choices and get the best price overall.

From such high-level concerns, we turn our attention to the cloud's underpinnings in Lukasz Paciorkowski's article on cloud-native design. While he acknowledges that the subject of building and operating cloud-native applications "is a very technical discussion," he argues that "it can have a profound impact on the business and operational model, [and thus] it should not be ignored by anybody who is serious about capturing the value of cloud computing." After exploring the techniques, platforms, and powerful new tools that make this design approach possible, Paciorkowski outlines the business benefits of cloud-native design and discusses the characteristics that might make an organization a good — or not so good — fit for this emerging solution.

In our next article, Saman Michael Far takes us from theory to practice in his case study of IaaS implementation at FINRA, an independent nongovernmental organization that monitors and regulates US securities trading. Seeking to reduce its infrastructure spending, automate its production support, and increase its analytics capability, FINRA found that moving to a virtual private cloud using Amazon Web Services (AWS) and open source platforms was the best way to achieve its objectives. Far discusses the reasoning behind this decision, as well as many issues organizations face when moving to the cloud, such as security, business and architectural concerns, disaster recovery planning, and impacts on culture.

Next, Annie C. Bai reminds us that "information security can never be fully outsourced." As tempting as it might be to consider security, privacy, and data integrity someone else's problems once you move to IaaS, the unfortunate truth is that all these things are still your organization's responsibility. After all, she notes, "it is your business that will suffer the consequences of any interruption of access or any flaws in your data integrity." Bai gives recommendations for addressing these issues in legal agreements with your provider, so that if outages or data breaches do occur, the vendor will provide redress. “There will be some unforeseen hiccups with IaaS,” Bai notes, “but their impact will be mitigated if your operations plans are in place and your legal protections are in order.”

IaaS is being adopted rapidly, and the complexity within IaaS requires careful thinking, planning, and implementing.

Looking ahead as IaaS and cloud technologies advance, in our final article I identify seven threats in cloud computing that are likely to have adverse impacts. These are:

1. The behavioral inertia in vendors and companies around adopting the new cloud economic model
2. Struggles for dominance between cloud providers, resellers, and end-using companies regarding contract terms
3. A lack of imagination and planning regarding significant potential IaaS market failure or other black swan events
4. Companies providing no credible threat of defection, thus allowing vendor lock-in and lack of price-performance competitiveness
5. Poor understanding by companies regarding how their workloads actually consume IaaS resources, preventing companies from extracting full value
6. Continued merger activity in the market, which can reduce supply chain diversity and competitiveness
7. Difficult considerations regarding countries that insist data for their companies cannot leave their borders, also known as data sovereignty concerns

I believe that, collectively, CIOs and their companies will have more to say about how the IaaS market unfolds and that CIOs have a responsibility to shed light on these issues.

CONCLUSION

Taken together, the articles in this edition of *Cutter IT Journal* show clearly that IaaS is being adopted rapidly, and the complexity within IaaS requires careful thinking, planning, and implementing. If there is one thing to glean from this issue, it is that not one of our authors describes technical inadequacies in the IaaS market as a showstopper. Instead, our authors’ eyes are on the pricing, contractual, organizational, implementation, and high-level risk management topics. CIOs and their companies have technical options so long as they can see their way through these issues.

Vince Kellen is a Fellow of Cutter’s Business Technology & Digital Transformation Strategies and Data Analytics & Digital Technologies practices and a member of the Cutter Business Technology Council. Dr. Kellen brings a combination of entrepreneurial, academic, and IT strategy experience to his role. He is currently Senior Vice Provost, Analytics and Technologies at the University of Kentucky, one of the top public research institutions and academic medical centers in the US. Dr. Kellen’s areas of responsibility include institutional research and analytics, enterprise software, research computing, academic technology, IT infrastructure, and cloud services, supporting the university’s 17 colleges, 30,000 students, and 14,000 faculty and staff. He has also served as a partner with strategy consulting firms, where he helped Fortune *500 and midsized companies with business and IT strategies, IT organizational development, customer experience management, CRM, and data warehousing and analytics.*

Dr. Kellen currently teaches at the University of Kentucky’s College of Communications on IT and strategy and previously taught at DePaul University’s College for Computing and Digital Media. He is a recipient of the coveted CIO *magazine Top 100 award in 2007 and 2014, was one of* InformationWeek*’s Chiefs of the Year in 2013, received the* Computerworld *Honors Laureate award in 2013, and was one of four recipients selected globally for Dell’s 2012 Transformational CIO award. Dr. Kellen has served on advisory boards to top IT companies, including Dell, SAP, Microsoft, AT&T, and Apple, and cohosts (with IT futurist Thornton May) a regional Kentucky CIO forum called The CIO Practicum. He is the author of four books on database technology and author of more than 250 articles and presentations on IT and business strategy topics. Dr. Kellen earned his PhD in computer science (human-computer interaction), an MS degree in e-commerce, and a BA degree in communications, all from DePaul University. He can be reached at vkellen@cutter.com.*

Cutter Webinars: register now at www.cutter.com/events

The Unseen Contract: Your Greatest Outsourcing Friend or Foe

Webinar with Dr. Sara Cullen, Senior Consultant, Cutter Consortium

Date:	Wednesday, 18 November 2015
Time:	6:00 pm - 7:00 pm EST
Location:	At your desk — just dial in!
Duration:	60 minutes
Fee:	Complimentary
Register:	www.cutter.com/events

We spend a tremendous amount of time getting an outsourcing contract right; assuming that if it is thorough and airtight, we'll get what we are expecting. But the most important contract isn't the written one. Another contract, an arguably more crucial one, is the unseen contract. This is a psychological contract — the aspects that are believed to be understood despite not being written down and possibly not even discussed. The unseen contract has a direct effect on how successful our outsourcing arrangements will be.

In this webinar, Cutter Senior Consultant Dr. Sara Cullen of Melbourne University will explain how the unseen contract reveals itself, why it matters so much, how it can derail outsourcing, and most important, how you can harness it to deliver superior value.

Darkitecture — Why It's Important

Webinar with Balaji Prasad, Expert Contributor, Cutter Consortium

Date:	2 December 2015
Time:	5:00 pm - 6:00 pm EST
Location:	At your desk — just dial in!
Duration:	60 minutes
Fee:	Complimentary
Register:	www.cutter.com/events

The term "shadow IT" is a commonly used phrase to indicate that IT investments and expenses are incurred without the IT organization's involvement or concurrence. Shadow IT is symptomatic of how enterprises really operate. There are processes, systems, and structures that are visible on the surface, and these work well for the most part. However, there are activities, investments, people, and decisions that are outside of the well-delineated boundaries of an organizational function.

There is a similar phenomenon of things that are visible but set in the backdrop of a much larger expanse of dark matter. If we have an architecture that outlines the visible enterprise, there is a "darkitecture" that we cannot describe, but that exists nevertheless.

Spend an hour exploring the potential that lies beyond the edge of your visible architecture. Sign up now for this complimentary webinar.

What Should a CIO Consider When Running a Cloud RFP?

by James Mitchell and Frank Khan Sullivan

As the cloud market has grown up, CIOs are now in a position to rely on a modern, robust ecosystem of cloud computing vendors to deliver efficiencies and cost savings at scale. To do this, however, the role of the CIO demands an understanding of how to build and subsequently manage that supply chain when deciding to procure infrastructure as a service. This article discusses how the CIO, ahead of launching an RFP, should prepare for the challenge of coordinating multiple, competing IaaS service offerings to establish the type of mature supply chain more commonly found in established utilities markets.

THE SHOPPING LIST IS DEAD; LONG LIVE THE SHOPPING LIST

Over the last decade, many enterprise IT departments resisted calls to fully outsource their function to large systems integrators. Enterprise IT is now being coaxed into considering a less extreme form of IT outsourcing — either a full migration of certain applications to the public cloud, or a gradual transition via a hybrid cloud model.

Within the next 10 years, enterprise IT will adopt a utility consumption model in order to retain inhouse control without the financial and support burden of inhouse delivery. They will take a multi-vendor approach to avoid vendor lock-in and maintain efficient pricing. This all means that enterprise IT procurement will require a radically different approach to selecting the best vendors. RFP tendering approaches designed for highly specific physical hardware are just not suitable for the procurement of intangible, perishable, shared utility services.

CIOs must assess their organization’s readiness to adopt such a multi-vendor cloud procurement approach. Creating a robust supply chain for on-demand IT services will not be as straightforward as simply using services made readily available in established utility markets, like power. Regulators are just starting to wield their influence to encourage major cloud providers to cooperate in offering a resilient utility service of the quality offered by established utilities such as electricity and natural gas.

AVOID LOCK-IN BY MAKING ONE CHANGE TO YOUR PROCUREMENT PROCESS

Most workloads can be run in the cloud, and the capabilities of IaaS providers have evolved to a point where CIOs are forced to make a choice: invest in their own aging data centers or invest in building a modern ecosystem of suppliers to deliver the same capabilities. Some clouds offer a vertically integrated and deep set of products, but the savvy CIO is wary of vendor lock-in. This is precisely why CIOs who lead the organizational change to adopt cloud computing must understand supply chain management.

The two most common approaches — choosing everything from a single vendor or simply chasing the lowest price — do not fully price in risk or optimize for it. Let us take the classic example of availability. Is it better to rely on a single provider, Alpha, that provides 99.99% uptime (which amounts to 52 minutes of downtime a year and is expensive to achieve), or to seamlessly switch between two providers, Beta and Gamma, that each provide 99% uptime (which amounts to 87 hours of downtime a year and is considerably cheaper to achieve)?

TWO VENDORS, TWO THROATS, TWO HANDS

When a CIO takes the view that having a single supplier is easier because there is only “one throat to choke,” they should question who is choking whom. Using multiple cloud service providers increases the flexibility and resilience of your supply chain as a whole. That one cloud provider that promised you the moon is unlikely to deliver top marks in all performance categories, including availability, because if your organization has already contractually committed to five years of usage in order to “unlock” discounts in price, there is no real incentive to go beyond minimum service levels. After the CIO’s procurement team has short-listed prospective suppliers based on technical ability, the procurement focus should turn to the costs and practicalities of combining multiple suppliers that together can deliver a high-quality service at a reasonable price point.

Adopting a multi-vendor approach also mitigates risks relating to capacity shortages, single supplier risk, and issues over quality or poor vendor performance. By defining which workloads are most readily portable, an organization is better placed to avoid lock-in wherever possible. Ahead of issuing an RFP, a CIO might also consider benchmarking suppliers based on their specialism(s) in order to identify areas of collaboration between suppliers and assess pricing according to their combined merit.

COMBINING SUPPLIERS OFFERS LOWER RISK, LOWER COST, AND BETTER SPECIALIZATION

Coming back to our example of availability, it is more cost-effective, for exactly the same risk of system outage, to rely on the combination of Beta and Gamma. If we can assume their outages will be independent events, we would multiply 1% by 1% to get the same 0.01% system outage risk as for Alpha — the single, more expensive supplier. The key question is whether the much cheaper capacity bookings at both Beta and Gamma are still less expensive than Alpha's premium service offering once you add in the cost of building the capability to switch seamlessly between them.

Cloud supplier pricing is incredibly complex and packaged as incredibly simple. It is often the case with the unsuspecting cloud buyer that the suppliers with the simplest pricing are also the most expensive. But that's changing as cloud buyers who don't have the internal skills to create an effective benchmark for vendor pricing comparison turn to brokers who do. IaaS isn't a commodity, but it should be priced like one in order to make evaluating proposals fairer and more transparent.

HAVE YOUR CLOUD AND EAT IT, TOO

The most common problem we see in RFPs for cloud services is the reuse of tendering materials that were designed for the outright purchase of physical objects. Problems that arise include buyers being unnecessarily specific in the definition of their requirements, buyers trying to impose what they would do in their small data center to an exascale cloud provider servicing thousands of diverse clients, and many others. The best way to spot these problematic approaches is by way of analogy.

Choose from a Defined Menu

Imagine you have to feed a family of fussy eaters. You start off by letting each person choose exactly what they want to eat, and naturally everybody chooses something different. Nightmare!

It quickly becomes unfeasible to meet the exact requirements of each family member, so you have to compromise. As a family, the way you save money on your food costs is to make sure the component ingredients and a limited set of tools necessary to prepare and serve them are the same across all meals. You tend not to have different perishable ingredients for every meal, and you don't invest in a brand-new pan every time you want to boil an egg.

The secret to feeding your family on a budget is then the same for delivering IT services in the enterprise … try to get everyone to choose from a carefully selected menu!

If an organization issues an RFP that demands a service or a contract that is too far from the market norm, then the benefits of sharing are degraded, driving up the cost to deliver.

It Is Cheaper Because You Are *Sharing*

There is a fundamental difference between procuring the ingredients and capabilities to deliver a service that is solely for your own use versus procuring an on-demand shared utility service such as restaurant food or cloud services.

The attraction of accessing "the cloud" is the ability to leverage the benefits of sharing, whether that be between different types of users within a single organization (private cloud), or between different organizations (community and public cloud). However, these benefits come at a cost relating to the buyer's ability to define the exact specification of what is bought, and the terms on which that purchase is made.

If an organization fails to recognize this, and issues an RFP that demands a service or a contract that is too far from the market norm, then the benefits of sharing are degraded, driving up the cost to deliver.

Should You Try That New Restaurant?

To extend the restaurant analogy, would you commit your family to eating at a single restaurant for a year without trying it out? Probably not.

It makes sense to encourage representatives of each cloud usage case to try out a few different services ahead of the procurement and solicit their feedback. It is also important to assign a trial budget in order to sample the actual menu, not just the free appetizers, in order to understand the pricing and payment experience, which can vary dramatically between clouds.

As with restaurants, it can be unclear what level of support or service you can expect to be included in the price, and whether they will accept your preferred payment method. Ask your would-be cloud consumers, and their related stakeholders in legal, finance, and other departments, to take a taste test with prospective suppliers before you issue the RFP.

> **Deciding whether to lock in heavily discounted future prices versus keeping your switching options open requires access to the cloud pricing history and the analytical skills to know what to do with it.**

Should You Ask About the Specials?

It is one thing to look at the menu, but there are always questions for the waiter, and if you happen to be allergic to nuts, you're better off focusing your questions on how the meal is prepared than worrying about whether service charges are included.

Restaurant meals are not a truly fungible commodity, which is why we rely on the likes of Zagat and Michelin to help us compare quality. Cloud is also not a true commodity, and so similar brokerage models exist and are useful.

Keeping Your Options Open

Does the prospect of listening to your family complain about going to the same restaurant every day for a year not appeal to you? Are you worried that even if you went to that single restaurant that had everything you wanted on the menu, the quality might degrade over time? What if the restaurant opposite, the one with three Michelin stars, drops its prices to be cheaper than your restaurant?

The same applies to choosing a single cloud provider. No single cloud vendor is good at everything, despite what a vendor may claim; and what may seem great now may only be the flavor of the month. Why would you commit to using the same provider for a year?

IMPLEMENTING THE MULTI-SUPPLIER APPROACH TO CLOUD

By this point we hope we have persuaded you that taking advantage of multiple, competing service offerings is the most cost-effective way to procure IaaS. So how should you go about it?

Who Should Lead This Research?

Who better to lead the assessment of prospective suppliers and go on to manage vendor performance than the people who run your own data center? When someone from your own data center operations team joins you on a visit to assess a supplier, they're the ones likely to ask all the penetrating questions about redundancy, security, cooling, and disaster recovery. The logical progression for technical and operational roles is to transition to vendor performance management roles because the same subject matter experts who write your RFP are also the ones most qualified to assess their counterparts on the supplier side. Price is irrelevant if the service simply doesn't work for you. However, the CIO must also deploy a new set of financial and risk analysis skills in the RFP process in order to complement the technical side.

Should You Go for the Discounts?

In cloud provisioning, a one-year deal often represents a 50% discount on hourly or monthly contract rates. A three-year deal can be 70% cheaper. The discount is to entice more customers to commit to future spending so that the cloud provider can de-risk their up-front CAPEX investment.

Deciding whether to lock in heavily discounted future prices versus keeping your switching options open requires access to the cloud pricing history and the analytical skills to know what to do with it. This is the role of a financial broker (full disclosure: this is the authors' specialty). The more switching options you can show your financial broker, the better the deal you will secure.

The Role of Financial and Technical Brokers

Using a technical broker or a cloud marketplace is one great way to access those switching options. Think of it like choosing to frequent a food court every day instead of sticking to a particular restaurant. You're committing to paying the parking charges to use the marketplace, but that should be cheap considering the savings you'll reap from having multiple vendors fiercely competing for your business.

One downside to the cloud marketplace approach is that you get multiple receipts for your aggregated consumption, which can be a real burden for whoever controls the budget. Cloud services are priced and invoiced in a way that makes fair comparison of value for money a real challenge. When pricing is oversimplified, it is often true that pricing is more expensive than necessary. If not, the vendor is unlikely to remain the cheapest for long due to pricing inefficiencies. Look for services that provide unified billing, either as part of the marketplace offering, or as a standalone service — your finance team will thank you! Basically, it's the financial broker's job to save the buyer money and support the RFP process to ensure the buyer has enough choice.

A More Pragmatic Approach to RFPs

Your RFP should specify your benchmark requirements, which ought to be set, if at all possible, in the middle of what the market offers. This should be defined as broadly as possible and not include any arbitrary specifications (e.g., "Do you really require champagne, sir, which can only come from the Champagne region of France, or will any form of sparkling wine suffice?"). You must understand that vendors bundle the cost of adaptation into the overall cost, which makes it hard to work out the premium added for exactly meeting your specific requirements.

This will require some research for each characteristic that you specify. Your research will quickly uncover that vendors offer services that deliver more or less than your benchmark requirements. In your RFP, you must make clear what happens when the service is above or below your benchmarks. You can avoid significant problems if you have a strong focus on vendor performance management, open communication with your supplier(s), and a willingness by both parties to engage in corrective measures outlined in a service-level agreement (SLA), such as remuneration in the form of a discount or, as preferred by vendors, a service credit. The best possible outcome is achieved through swift, open dialogue. Worst-case scenario, before a cloud buyer cancels the contract or takes legal action, arbitration has proven to be a very successful method for conflict resolution. If an SLA is clear on how conflict will be handled from the outset, and if arbitration is included as an option, a CIO may have a better way to deal with performance issues in the contract than resorting to protracted legal action and time-consuming tasks when moving to a new supplier.

Be pragmatic. Expect that what you ask for in your RFP will not exactly match what vendors are offering. If you insist that vendors meet your precise specifications, be prepared to accept a price increase.

Your RFP should specify your benchmark requirements, which ought to be set, if at all possible, in the middle of what the market offers.

FOUR QUESTIONS TO CONSIDER IN ORDER TO RUN A BETTER TENDER

1. How will you inform a large audience of vendors that you have a commercial requirement, and how will you handle it if vendors choose not to respond to your RFP due to a minor mismatch between their service description and what you have specified?
2. What if vendors add in the cost of unnecessary adaptations to their service, thus driving up the price? Can somebody from your data center team act as a liaison with your procurement team to decide what are acceptable limits for customization or price?
3. What if the differences don't get discussed, and vendors say that the service meets your requirements even though it doesn't? Whose responsibility is it to independently verify a vendor performance issue that is not made clear in an SLA?
4. Does your contract specify arbitration as a way of working around the differences short of terminating the contract (with all the ensuing disruption that implies)?

When you want to procure the hardware and capabilities to run your own IT infrastructure, you can specify any details you choose. This works because you are buying those ingredients and cooking utensils for life. The problem is that you cannot cook at home on the scale offered by a global restaurant chain, particularly not for short-term consumption. A very high level of specificity does not work well when you are renting something for a short period of time; the vendor will want to repurpose what it sells to you for the next customer. This limits your technical customization. The vendor also wants to minimize its cost of sale and operating costs associated with the transaction, which is smaller due to its shorter commitment period. This limits the amount of financial and contractual customization that can be done.

THE CIO IS FAST BECOMING A SUPPLY CHAIN MANAGER

A CIO must consider who will be involved in testing potential providers before sending RFPs, how the data center team's roles may change, how to appropriately benchmark and compare supplier pricing, and the extent to which customization can be minimized in order to maximize choice. Consider the input of a cloud broker when choosing to run a tender that allows multiple vendors to join a robust, modern, and resilient supply chain.

James Mitchell is a Senior Consultant with Cutter Consortium's Business Technology & Digital Transformation Strategies practice and CEO of Strategic Blue, a financial cloud brokerage firm offering cloud computing, financial, and analytical services more commonly seen in the commoditized energy markets. For the past five years, Dr. Mitchell has been pioneering various financial cloud brokerage concepts that are now gaining widespread acceptance and adoption based on his background as an originator of structured physical commodity transactions at Morgan Stanley.

Featured in books and reports as the leading authority on how the IT market will change because of the move to a utility delivery model, he is a frequent speaker at events and often successfully predicts how the cloud market will develop over the next five years. For instance, Dr. Mitchell correctly predicted the price coupling currently observed in the cloud IaaS market. His expertise in the pricing of pseudo-commodities is highly relevant for any organization looking to buy, sell, resell, or trade cloud services at scale. He has a doctorate in DNA nanostructures from Oxford University and a first-class master's degree in experimental and theoretical physics from Cambridge University. He can be reached at jmitchell@cutter.com.

Frank Khan Sullivan is the VP Marketing of Strategic Blue. His work focuses on increasing the understanding of financial brokerage concepts in the cloud computing market. He helps large organizations understand the value of adopting cloud computing and applying advanced price analysis techniques to find cost savings and efficiencies. Before joining Strategic Blue, Mr. Sullivan worked with leading technology companies in the software and cloud computing industry. He is based in London, UK, and can be reached at frank@strategic-blue.com.

Cloud-Native Design — What Has Changed?

by Lukasz Paciorkowski

Cl1oud-related discussions are happening everywhere. One topic that is both interesting and has a lot of potential is how to build and operate the new applications tailored for the cloud environment. To a great extent, it is a very technical discussion, but as it can have a profound impact on the business and operational model, it should not be ignored by anybody serious about capturing the value of cloud computing. In this article, I give an overview of how to design, build, and operate a new breed of cloud-native applications. I will also look at how cloud-native design positively impacts a business, bringing new and crucial capabilities to your organization.

CLOUD-NATIVE DESIGN? WHAT'S THAT?

Like any other information technology, cloud requires a specific approach from developers. In order to benefit from the features that cloud computing offers us, it is important to adapt to the technology and its specifics. There are different terms used to describe how well (or how poorly) an application is prepared to run in the cloud. The most common are cloud-centric, cloud-enabled, and cloud-native, which are used in many contexts and very often interchangeably. As these are not official definitions, I have taken the liberty of putting them on one scale so the clear "progression toward cloud nirvana" can be seen (see Figure 1).

Two questions arise instantly:

1. Do we want to move toward cloud-native design, and is that our ultimate aim?
2. How do we design (or redesign) our application and organization so they fit the cloud world?

The first question I will try to answer later in the article. The second one points toward some kind of standard, pattern, or method for designing cloud applications. Thankfully, some smart people have already explored cloud design patterns and shared their thoughts publicly. There are two major "manifestos" that can guide developers and help them to achieve the status of "cloud wizards":

- The Twelve-Factor App[1] is a 12-point methodology that ensures your application will be suited for the cloud. It focuses on functional requirements that every application should meet in order to be fit for the cloud environment.

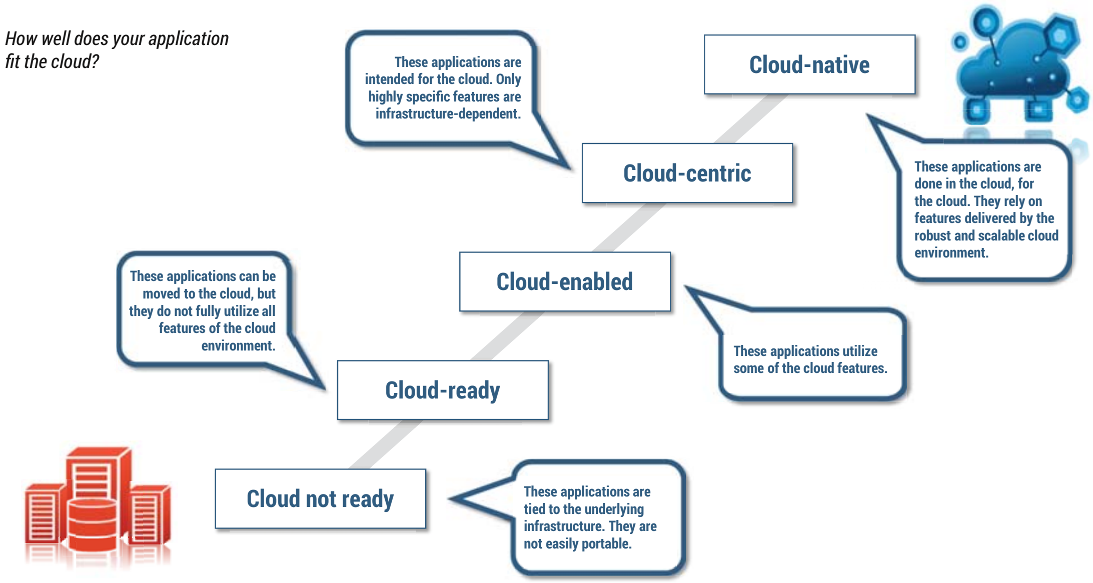

How well does your application fit the cloud?
Cloud-native
These applications are done in the cloud, for the cloud. They rely on features delivered by the robust and scalable cloud environment.
Cloud-centric
These applications are intended for the cloud. Only highly specific features are infrastructure-dependent.
Cloud-enabled
These applications utilize some of the cloud features.
Cloud-ready
These applications can be moved to the cloud, but they do not fully utilize all features of the cloud environment.
Cloud not ready
These applications are tied to the underlying infrastructure. They are not easily portable.

Figure 1 — Levels of "cloudification."

- The Reactive Manifesto[2] offers four very crisp points on the qualities every cloud application should possess. It can be compared to the generic nonfunctional requirements.

There are many other guides, words of advice, analyses, and architectural patterns that describe different aspects of planning, designing, and coding cloud-centric applications. Some of them embody specifics of the given cloud platform, programming language, or type of application. I therefore encourage everybody to explore relevant materials in depth.

MICROSERVICES: A PARTICULAR ARCHITECTURAL PATTERN

Scalability, elasticity, and adaptability — these are the terms commonly associated with the cloud, and so should they be with your cloud application. One way to achieve such flexibility is to decompose your design into many independent parts: namely, microservices (see Figure 2).

A microservice is an architectural pattern in which we try to isolate the smallest possible application functionality within the independent module. We also want to make sure that each and every one of those modules is elastic and can scale horizontally so it never becomes a bottleneck for the performance of the entire system. Thus, microservices should be stateless, message-driven, and responsive. If one instance goes down, a different one will take over.

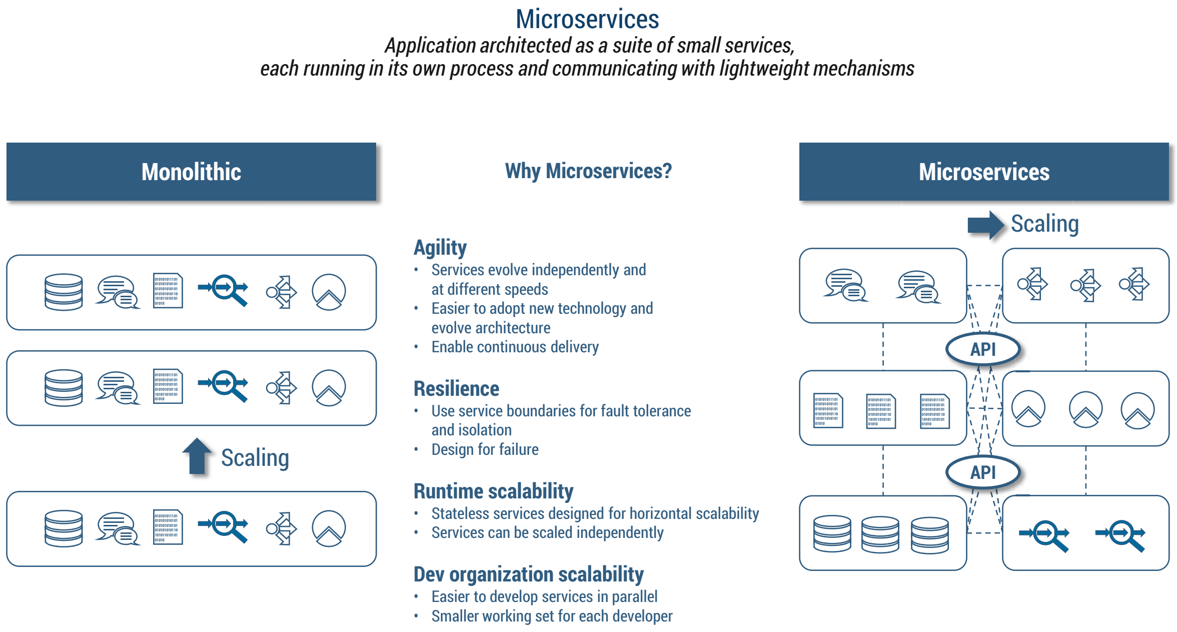
Another important feature of microservices is how they deal with errors. Overall, we could discuss two opposite approaches to handling the unexpected. One states that you should capture and appropriately react to each and every error your application can throw at you. You should focus on error resilience, error handling, and error management. Nothing can be a surprise to your application! The second is to embrace the errors. Let them flow! It's better to fail gracefully and restart quickly than try to handle every possible situation. Cloud-native design (and microservices) embraces the latter approach.

A LIGHTWEIGHT CONNECTIVITY

Along with microservices, cloud-native design promotes the use of application programming interfaces (APIs). Although this idea is not new, its popularity is growing rapidly along with the broad adoption of different kinds of platforms. APIs help to keep the architecture modular, where different modules can be developed independently at their own pace. Platforms like OpenStack, Cloud Foundry, Bluemix, Facebook, Twitter, and many others expose their functions via APIs to enable partners and third-party developers to add functionality on top of them.

There are two main reasons why APIs are important for cloud developers:

1. You will need to consume APIs exposed by the cloud platforms you will most probably use.



Microservices

Application architected as a suite of small services, each running in its own process and communicating with lightweight mechanisms

Agility
- Services evolve independently and at different speeds
- Easier to adopt new technology and evolve architecture
- Enable continuous delivery

Resilience
- Use service boundaries for fault tolerance and isolation
- Design for failure

Runtime scalability
- Stateless services designed for horizontal scalability
- Services can be scaled independently

Dev organization scalability
- Easier to develop services in parallel
- Smaller working set for each developer

Figure 2 — Monolithic architecture vs microservices.

2. You will want to expose the functions of your microservices via APIs so you can easily connect them with other modules and services.

If you have had an opportunity to work with SOA, you will definitely notice the similarities between the two concepts. Figure 3 outlines the most visible differences, showing APIs as a lightweight version of the SOA approach.

The API concept also has a profound impact on business. It enables organizations to expose their services and data so that they are very easy to consume. The API economy aims at creating a partner ecosystem of organizations and people who use your data or services in a way that brings additional value to the end customer.

FORGET ABOUT HARDWARE

The rise of software developers can be traced back to the 1950s, when Fortran was created. It was the first widely adopted programming language that allowed for the decoupling of hardware and software through the use of high-level instructions. Unfortunately, even then programmers had to take into account the specifics of the underlying IT infrastructure. Virtual machines and script language interpreters helped to abstract the programs even more, but at the end of the day, programming was always dependent on the infrastructure preparation. Installing and configuring the server, network, and storage were necessary prerequisites. It took time, and it was a distraction from the real aim of programming, which is coding the business functionality.

With cloud computing, the need for hardware platform–specific knowledge is fading away. Today programmers can focus on ... programming. PaaS brings the promise of an always ready, always available, and always up-to-date software development environment. This can be achieved by introducing an abstraction layer between the hardware and software. One of the most popular and fastest-growing open source platforms serving this purpose is OpenStack.

OpenStack is a modular standard dealing with different domains associated with IT infrastructure. It handles the setup and configuration of storage, networks, and computation nodes. Thanks to a broad development ecosystem, it supports most of the hardware components available on the market. Its standardized API allows programmers to dynamically manipulate and configure the technical environment according to the requirements, workload levels, and other runtime parameters. OpenStack is a platform where hardware adjusts to the software, and the concept of “software-defined everything” becomes a reality. No more screwdrivers nor low-level console commands. Everything can be managed using standardized high-level API calls.

Another standard that makes things easier for developers is Cloud Foundry (see Figure 4). It is an open source platform that supports the full application lifecycle starting from initial development to testing, deployment, and operation. Its built-in features made it a natural match for DevOps[3] and continuous delivery. In addition, Cloud Foundry enables application autoscaling, logging, and version control; supports distributed

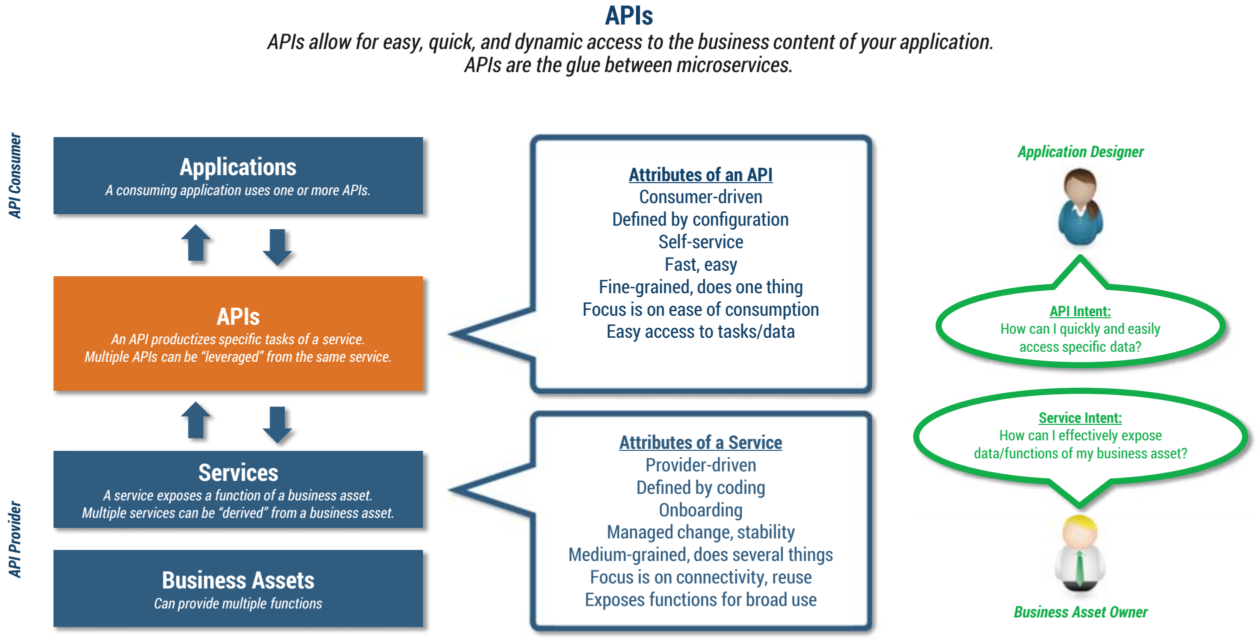

APIs
APIs allow for easy, quick, and dynamic access to the business content of your application.
APIs are the glue between microservices.
API Consumer
Applications
A consuming application uses one or more APIs.
APIs
An API productizes specific tasks of a service.
Multiple APIs can be "leveraged" from the same service.
Services
A service exposes a function of a business asset.
Multiple services can be "derived" from a business asset.
Business Assets
Can provide multiple functions
API Provider
Attributes of an API
Consumer-driven
Defined by configuration
Self-service
Fast, easy
Fine-grained, does one thing
Focus is on ease of consumption
Easy access to tasks/data
Attributes of a Service
Provider-driven
Defined by coding
Onboarding
Managed change, stability
Medium-grained, does several things
Focus is on connectivity, reuse
Exposes functions for broad use
Application Designer
API Intent:
How can I quickly and easily access specific data?
Service Intent:
How can I effectively expose data/functions of my business asset?
Business Asset Owner

Figure 3 — APIs as a lightweight version of SOA.

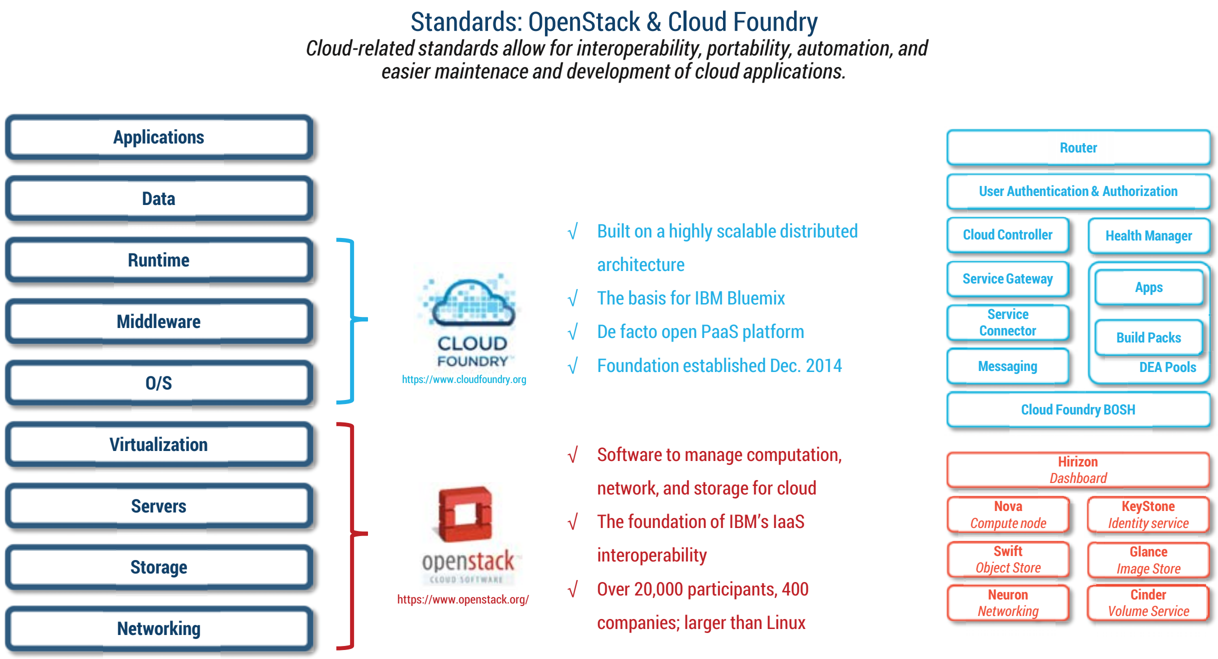

Standards: OpenStack & Cloud Foundry
Cloud-related standards allow for interoperability, portability, automation, and easier maintenance and development of cloud applications.
Applications
Data
Runtime
Middleware
O/S
Virtualization
Servers
Storage
Networking
CLOUD FOUNDRY
https://www.cloudfoundry.org
√ Built on a highly scalable distributed architecture
√ The basis for IBM Bluemix
√ De facto open PaaS platform
√ Foundation established Dec. 2014
openstack
CLOUD SOFTWARE
https://www.openstack.org/
√ Software to manage computation, network, and storage for cloud
√ The foundation of IBM's IaaS interoperability
√ Over 20,000 participants, 400 companies; larger than Linux
Router
User Authentication & Authorization
Cloud Controller
Health Manager
Service Gateway
Apps
Service Connector
Build Packs
Messaging
DEA Pools
Cloud Foundry BOSH
Hirizon
Dashboard
Nova
Compute node
KeyStone
Identity service
Swift
Object Store
Glance
Image Store
Neuron
Networking
Cinder
Volume Service

Figure 4 — The OpenStack and Cloud Foundry standards.

filesystems; and delivers many other out-of-the-box features. In short, this platform is designed with the cloud environment in mind, and it is a perfect fit for cloud-native apps. As an example, the IBM Bluemix PaaS is built on top of Cloud Foundry, adding many ready-to-use technical and business services that support analytics, Internet of Things, mobile, and other domains.

NEW, POWERFUL TOOLS

With the advent of cloud-native design, new tools, languages, and concepts appear. New (and not so new) languages like Node.js, Ruby, and Go are used to build lightweight, very responsive, and RESTful applications. Very often they deliver ready libraries for the most common operations, connectivity, security, and logging so developers do not have to reinvent the wheel and can focus on coding essential business functionality. Other standard components, like databases, are exposed as a service as well. Installing and configuring a database is no longer required in the cloud world. One very popular DBaaS platform is IBM Cloudant, which is based on the Apache CouchDB open project. Not only is it available through the Internet, exposing RESTful APIs for data manipulation, it is also a NoSQL database. This means there is no specific schema definition needed, and changes to the data structure can be done on the fly. There are a few types of NoSQL databases, such as graph, map, or document databases. Cloudant uses the JSON standard to store the data as documents. It is perfect for distributed, highly scalable, and responsive Web-based applications. Together with the Node.js language, which also supports JSON natively, they make a perfect match for cloud-native apps.

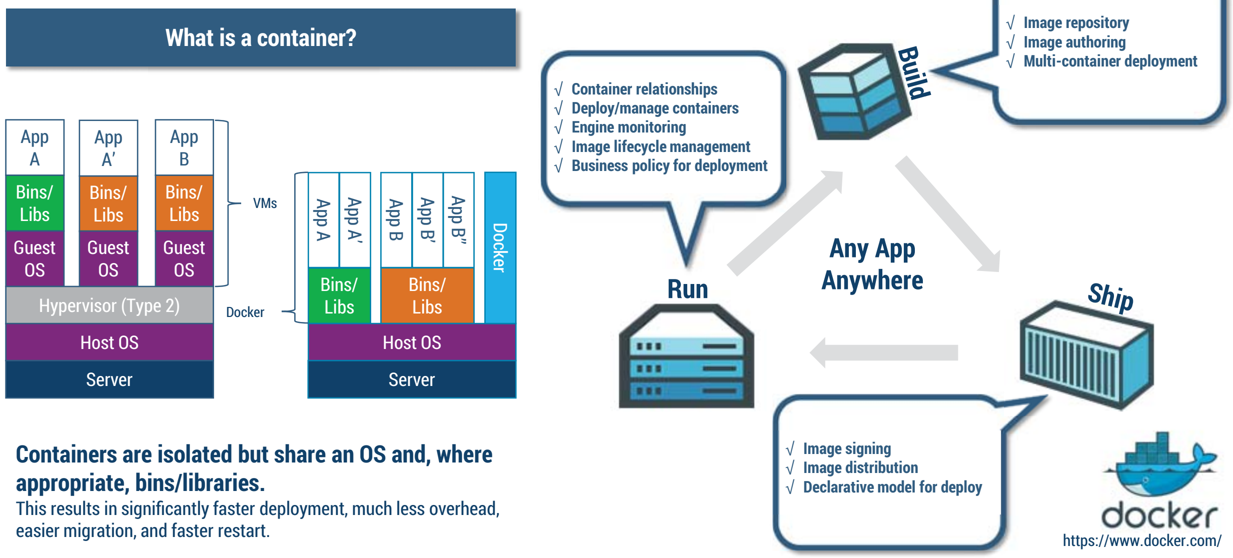
Another technology that is gaining great traction and popularity within the developer community is containers. Docker, a specific implementation of this concept, is already widely used and supported by most of the cloud platforms (see Figure 5). A container allows programmers to package their application together with everything necessary to run it: code, system libraries, and tools. This package (called a "container") can be deployed on any other system, ensuring that it will work as designed. Containers are often considered a lightweight version of the virtual machine.

On top of new languages, databases, and containers, there is a growing portfolio of tools delivered only in the form of public cloud services. We can see many examples of such offerings in the area of advanced analytics. There are numerous reasons why these services are not sold in an on-premises model, such as:

- The high entry cost for the underlying infrastructure makes it commercially unattractive if it is not sold as a public service.
- A public service allows better control over who is using state-of-the-art solutions and how.
- A public service offers better versioning control, thus allowing for instant bug fixing.

Containers: Docker

Containers are a relatively new concept that allow developers to conveniently package their developments and "ship" them to any destination.



Containers are isolated but share an OS and, where appropriate, bins/libraries.
This results in significantly faster deployment, much less overhead, easier migration, and faster restart.

Figure 5 — How containers work (Docker example).

Introducing cloud-native design and opening your software products to the public (or at least an off-premises) cloud enables your business with a new set of tools that otherwise would be not available.

WHY SHOULD WE CHANGE THE WAY WE WORK?

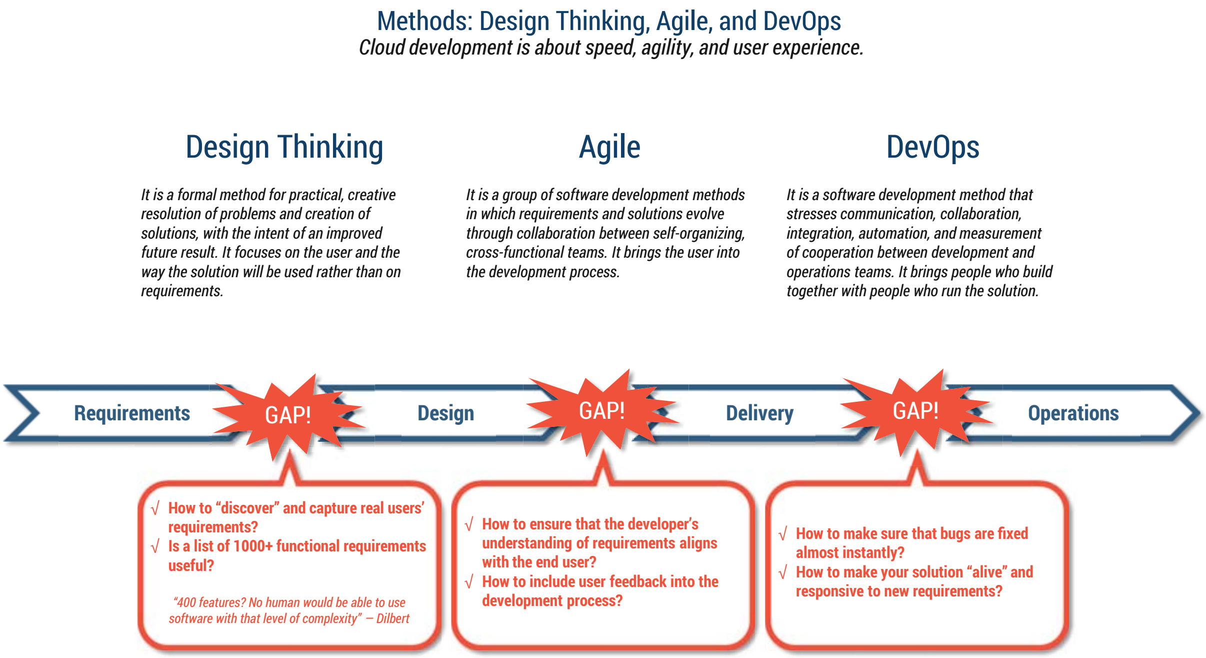
On top of all the changes in software engineering enforced by cloud technologies come changes in techniques, approaches, and our mental attitude toward application design, development, and operation. One very important feature of the "new" way of working is user centricity. Mobile apps development taught us that people will simply not use poorly designed or nonresponsive applications. Therefore, everything we do should focus on user experience. Users as well as business departments expect that the new breed of cloud-native applications will evolve, quickly adjusting to the new requirements. New features and fixes of known bugs should be delivered instantly without impacting business continuity. These expectations require changes in how designers, developers, and administrators work and collaborate (see Figure 6).

Much has been written about Agile and the benefits of using delivery methods like Scrum. But in order to use cloud to its full potential, it is crucial to look at the entire application lifecycle. Agile methods address the design and delivery stages well, but the requirements gathering/planning phase and moving from deployment to maintenance/operations are equally important.

As cloud-native design focuses on user centricity and speed of delivery, appropriate methods and practices should be implemented to address the commonly known issues.

The traditional approach to the planning and requirements gathering phase focuses on the functional features of the future software. It does not give much insight into how this application will be used by the end users. As a result, design (and, later on, development) is driven by many not always correct assumptions. This is where design thinking can help. By focusing on the user and how the software will be used, architects and designers can get better insights into what is really needed. Instead of asking what features a given application should have, design thinking tries to discover what the end user wants to achieve by using the application. IT professionals can then design the application guided by a combination of these insights and business and usability best practices. This approach enables them to deliver a delightful user experience, which solves the specific business problem in a way that the users could not have envisioned themselves.

Another deficiency that prevents organizations from leveraging the cloud's speed of delivery appears after software is put into production. The team that knows the app the best is the operations team — they work with it every day, they handle the errors and complaints from the end users, and they know what could be done better. Unfortunately, many times this team is separated from the development team; communication therefore



Methods: Design Thinking, Agile, and DevOps

Cloud development is about speed, agility, and user experience.

Design Thinking

It is a formal method for practical, creative resolution of problems and creation of solutions, with the intent of an improved future result. It focuses on the user and the way the solution will be used rather than on requirements.

Agile

It is a group of software development methods in which requirements and solutions evolve through collaboration between self-organizing, cross-functional teams. It brings the user into the development process.

DevOps

It is a software development method that stresses communication, collaboration, integration, automation, and measurement of cooperation between development and operations teams. It brings people who build together with people who run the solution.

√ How to "discover" and capture real users' requirements?
√ Is a list of 1000+ functional requirements useful?

"400 features? No human would be able to use software with that level of complexity" – Dilbert

√ How to ensure that the developer's understanding of requirements aligns with the end user?
√ How to include user feedback into the development process?

√ How to make sure that bugs are fixed almost instantly?
√ How to make your solution "alive" and responsive to new requirements?

Figure 6 — Design thinking, Agile, and DevOps are good ways to leverage cloud-native design.

takes more time and requires unnecessary overhead. Misunderstandings and unclear requirements further impede the delivery of new fixes and versions. DevOps addresses this gap by bringing development and operations practices together, thereby enabling continuous delivery. In the DevOps world, adjustments and enhancements are introduced as often as needed; the quarterly release strategy becomes obsolete. The most famous example of a working DevOps practice is Netflix, where changes to the production platform are introduced as many as 100 times per day.

In summary, cloud-native design requires changes not only in the tools, design, and platforms we use, but also in the way we operate our software solutions. Design thinking, Agile, and DevOps are not the only ways to leverage cloud-native design, but they are definitely good places to start.

THE PROMISE OF CLOUD-NATIVE DESIGN

Cloud-native design holds a lot of promise. Despite the effort needed to set up the tools, methods, and organizational practice around this approach, more and more organizations are following the successful implementations proven by Netflix and Amazon. But this is not a purely technology-driven trend. Adjusting all parts of your organization to work in a new way will enable your business to gain agility, which is crucial in a digital world. Below I discuss the four most important benefits that come from cloud-native design.

Flexibility

Using the microservices architectural pattern together with the proper implementation of DevOps practices should increase the speed at which your product can evolve. Consequently, innovation can be delivered faster. In a digital world, the ability to rapidly enhance, change, and improve your product is crucial. Being first is a very successful strategy. By decoupling the functions of your product, it is possible to introduce multiple development paths with different speeds and release strategies. Integration based on APIs can help to reduce dependencies between modules, applications, and the front and back ends.

Scalability

Digital business does not have geographical boundaries. It is safe to assume that for every business that can be digitized, there is a whole world of opportunities. Rolling out your solution to new countries has never been easier. Cloud-native design helps to prepare for rapid growth of data and your user base; high availability and theoretically unlimited scalability are two of the most important benefits of the cloud. The pay-as-you-grow model makes CAPEX investments obsolete. Thanks to the IaaS, PaaS, and SaaS cloud models, your organization does not have to become asset-intensive. Instead, you can balance your spending on technology with the current size of your business.

Portability

In a fully hybrid cloud environment, your cloud-native product can work on any infrastructure from any vendor anyplace in the world. It is impossible to predict which vendor will come up with better and cheaper technology. Therefore, staying vendor-neutral is a very good strategy. By using widely adopted standards (such as OpenStack, Cloud Foundry, and containers), it is possible to dynamically manage deployment of your workloads. Cloud-bursting techniques allow you to expand your on-premises cloud with the capacity delivered by off-premises platforms.

Automation/Self-Service

The cloud environment is based on standards, reuse, and automation. Those characteristics allow for competitive pricing, which reduces the manual labor needed to set up, configure, and manage IT infrastructure and platforms. On top of automation capabilities, self-service portals are an important part of the overall cloud-native environment. Thanks to them, developers can create necessary environments by themselves — and it takes only seconds. A well-designed application can scale itself, adjusting automatically to the current demand and the number of users. Both automation and self-service have a profound impact on the costs of operations and administration. They represent a significant business benefit of moving toward cloud-native design.

A new IBM Institute of Business Value report[4] gives a good overview of how important cloud-related technologies are for the condition and competitiveness of your business. In a digitized economy, technology becomes a core part of the business and is the source of competitive advantage. Introducing cloud-native design helps to capitalize on the market change driven by massive digitization.

IS CLOUD-NATIVE DESIGN A GOOD FIT?

Looking at the full spectrum of platforms, tools, techniques, and methods associated with cloud-native design, the obvious question arises: is cloud-native design the best fit for your organization? As always, there is no one right answer. Your decision should be based on many factors. Also, adapting to the new way of working, new tools, and new platforms can be done in many different ways. Moving to a cloud-native design practice will require time and most probably will generate additional costs. It requires careful consideration as to whether the future benefits, flexibility, and scalability of your product make it worth the effort.

Purpose and Audience

One of the most important factors in deciding for or against cloud-native design is the purpose and audience of your application. If your product is a mission-critical system (e.g., integrated flight system, core banking system), it might be counterproductive to use the cloud-native design paradigm. Such systems are not easily changed, and most of the time they run on dedicated hardware; therefore, the benefits of the cloud might not apply here. There are other more suitable architectural patterns you might use instead. Consider, too, who will be using your application. It is common knowledge that the mass end user is far more demanding regarding the design and responsiveness of an application than the corporate user. The latter is often happy with a simple windowed app as long as he or she does not have to work with black-screen, DOS-like applications. Although with Generation Y this attitude is changing, for corporate, transaction-heavy applications, cloud-native design might be overkill.

Introducing cloud-native design helps to capitalize on the market change driven by massive digitization.

Size and Scope

Size matters. There is not much point in designing a sophisticated microservices architecture with highly scalable characteristics if you know that the number of end users for the app will be roughly 100. Likewise, if the planned application will be very limited in scope and not exposed as a service, and if there is no strategy in place to increase its functionality in the foreseeable future, there is also no reason to decouple functional modules. Microservices require additional effort to embed common functions in every instance, and most of the time supporting services are necessary (such as dispatchers, proxies, and load balancers). For small applications, full-blown cloud-native design might not make economic sense.

Team Culture and Maturity

Cloud-native design requires both technological and organizational change. It might be increasingly difficult to push for “doing things differently” in a mature team with years of experience and well-defined procedures and processes. Sometimes the effort needed to change and persuade the team members of the benefits of cloud-native design might come at a price that is simply too high. On the other hand, a certain team maturity is necessary to maintain productivity and avoid wasteful practices. Cloud-native design might bring a lot of tools and practices that will not work for your organization and only cause an increase in the workload. Mature team members will be able to recognize the harmful practices and correct them.

> **The learn-as-you-go method almost always ensures that a new organizational culture will emerge.**

Time and Costs

Moving to cloud-native design requires time and money. It is probably not the best idea to change your design and delivery methods while in the middle of a project with an aggressive schedule. That said, it is also very risky to kick off a pure transformation project, as those tend to be very long, and organizations need to wait a significant amount of time to see tangible results. Regarding cloud-native design as an architectural style, it is not always beneficial to rewrite existing applications in accordance with the new paradigm. Organizations should carefully analyze whether it would be more cost-effective to modify existing apps or create new ones from scratch. Both time and costs will depend also on the adoption approach your organization chooses, as I discuss in the next section.

APPROACHES FOR ADOPTING CLOUD-NATIVE DESIGN

Should you decide that introducing cloud-native design is the right strategy for your organization, there are a variety of ways to go about it.

All or Nothing

This scenario assumes a binary decision. You either stick to your well-known, optimized, and procedural way of designing and operating, or you move completely to the new way of doing things. This approach fits best those organizations that are unable to keep up with market demands any longer. If you are under paralyzing pressure from the competition and always-growing, always-unfulfilled demands from your end users, it might be time to revolutionize your entire delivery model. Changing your tools and methods, redesigning your product, and adjusting how your organization operates might be the only way to survive.

The all-or-nothing scenario requires an immense amount of focus on organizational change management. Most of the time, it also requires a fundamental change of the organization’s culture. You can be sure of resistance from all sides. (All this in the name of the greater good.) In big organizations, such transformation projects can easily take two to three years before producing the first noticeable results.

Learn As You Go

This scenario aims at introducing incremental improvements — evolution rather than revolution. It is most suitable for organizations that noticed the change in the market early enough. There should be a consensus among teams and management that, in order to continue growing, things need to change. Setting up a group of coaches and “architects of change” is a good practice. They will show others the way, but the learning and change will come organically from the team members.

Trainings and enablement sessions are crucial to building awareness among employees. Improving the delivery pipeline by gradually moving toward DevOps practices at the same time that you decouple parts of your products will bring the flexibility. This, in turn, will increase the responsiveness of your delivery organization. The learn-as-you-go method almost always ensures that a new organizational culture will emerge.

Proof of Concept

This scenario relies on the introduction of “two-speed IT.” On the one hand, you keep things as they were and focus on sustainability and further optimization. On the other hand, you create a small unit, department, or team that works according to the new cloud-native rules. This approach is most appropriate for medium and big companies where the all-or-nothing method is too risky and the learn-as-you-go method will take too long.

Treat the effort like a managed revolution — starting small, collecting lessons learned, adjusting the approach, and finally rolling out the change to the entire organization. A recent article by researchers at McKinsey[5] offers advice on successfully implementing a two-speed IT operating model.

GOING NATIVE

Cloud-native design, with all its associated methods and practices, is definitely an approach worth considering. Although the topic itself touches a lot of technical aspects, the driving arguments behind adopting this "new way of working" are coming from the business. In the cloud context, technology is an enabler that helps organizations achieve better results in a shorter time with smaller investments. In a digitized economy, business agility is driven by technology, and cloud-native design helps to align technical capabilities with the market strategy.

ENDNOTES

[1] The Twelve-Factor App (http://12factor.net).

[2] The Reactive Manifesto (www.reactivemanifesto.org).

[3] Sharma, Sanjeev. "Adopting DevOps for Continuous Innovation." IBM developerWorks, 2 June 2014 (www.ibm.com/developerworks/library/d-devops-continuous-innovation).

[4] Brown, Doug, et al. "New Technology, New Mindset: Strategic IT Infrastructure to Compete in the Digital Economy." IBM Institute for Business Value, 2015.

[5] Bossert, Oliver, Martin Harrysson, and Roger Roberts. "Organizing for Digital Acceleration: Making a Two-Speed IT Operating Model Work." McKinsey & Company, October 2015 (http://goo.gl/Yiis2t).

Lukasz Paciorkowski is an IBM Cloud Advisor and a member of the IBM Academy of Technology leadership team. His background is in enterprise architecture and business transformation. Currently, Mr. Paciorkowski is focusing on digital business transformation and new business models enabled by the cloud. He is also helping European customers find the best path forward with cloud adoption. He can be reached at l.paciorkowski@gmail.com.

The Reasons for (and Benefits of) Moving to IaaS:
A Case Study at FINRA

by Saman Michael Far

The Financial Industry Regulatory Authority (FINRA), the largest independent regulator for all securities firms doing business in the US, is moving its technology platform to the Amazon Web Services (AWS) cloud and open source platforms.

We conceived the move three years ago during a review of our systems that resulted in the decision to fundamentally rebuild our market regulation platform on the cloud, and to do so using open source platforms. The program has been underway for close to two years, and 70% of systems are currently operating in the cloud.

In describing our experiences, I will begin by outlining our objectives for moving to the cloud and how these resulted in choosing a virtual private cloud using a large-scale cloud provider (AWS) rather than building our own private cloud. I will also discuss how we addressed several concerns that companies considering a migration to the cloud often face, including security, the balance of business and architectural concerns, DevOps requirements, disaster recovery, and implications for our culture.

WHAT DOES FINRA DO?

FINRA is dedicated to investor protection and market integrity through effective and efficient regulation and complementary compliance and technology-based services. FINRA touches virtually every aspect of the securities business, from registering and educating all industry participants to examining securities firms, writing rules, enforcing those rules and federal securities laws, and informing and educating the investing public. In addition, FINRA provides surveillance and other regulatory services for equities and options markets, as well as trade reporting and other industry utilities. FINRA also administers the largest dispute resolution forum for investors and firms.

Most relevant to our cloud initiative, FINRA is responsible for regulating 99% of equities and 70% of options trading in US securities markets. The market regulation function within FINRA receives market-data feeds that can exceed 75 billion records per day and processes this data, creating multi-petabyte data sets and searching for wrongdoing by market participants.

For example, in the case of equities, the data is received from the various US stock exchanges, broker-dealers, alternative trading systems known as “dark pools,” and industry organizations. It is then normalized and integrated to create a multi-node graph for each order on the US markets. These graphs can vary in size from several nodes to millions of nodes, as buy and sell orders are routed around the country in search of the best transaction price, also called the execution price.

After creating a complex picture of the state of the markets at every moment in time, surveillance algorithms scan the data for fraud and market manipulation. Alerts are generated, and analysts examine behavior patterns in the marketplace by querying the multi-petabyte data sets to home in on suspicious behavior in the markets.

The people and technologies needed to accomplish this represent the majority of the organization’s IT footprint. These very high volumes of data come with challenges. Market volumes are steadily increasing and can be volatile. For example, it is not unusual to experience peak market volumes three times larger than the average. Exchanges are dynamic and evolving, regulations are continually being enhanced, and new rules are being created. Simultaneously, new products are being introduced that create new potential targets for wrongdoers. And during all of this, market manipulators themselves are continuously innovating.

BEFORE THE CLOUD

Until recently, FINRA’s data center environment was similar to that of many other companies. Due to our big data processing needs, we have made extensive use of EMC Greenplum and IBM Netezza data-processing appliances, along with various NAS and SAN storage systems used for holding final data and as jump points for data movement. These are combined with various proprietary large-scale ETL tools and significant adoption of Linux and Oracle and are accompanied by some .NET and SQL Server environments. Most operations were housed in a primary data center, and a backup data center was maintained in a sufficiently distant location.

OBJECTIVES FOR MOVING TO THE CLOUD AND OPEN SOURCE

Two principles guided our effort to move the market regulation systems to a new platform. The first was a decision to move to the AWS virtual private cloud platform, and the second was to use open source technologies to totally update our systems. The migration plan itself was designed to accomplish three broad objectives:

1. Decrease our infrastructure spending in order to redirect expenditure to data analytics
2. Improve productivity, reliability, and efficiency by increasing automation of production support tasks
3. Increase the business value through improved accessibility to data and data analytics with burst access to unbounded commodity storage and computing power

Our choice of open source platforms was driven by a desire to harness large clusters of commodity hardware on the cloud rather than maintain exotic data-processing appliances, and to increase execution flexibility in the face of a rapidly evolving and fragmented big data tools market.

We believed — and subsequent experiences have confirmed — that by going the open source route and using platforms such as Hadoop, HBase, and Hive, we would avoid being overcommitted to a single vendor, benefit from the large community that is contributing to the advancement of these tools, and, perhaps most importantly, develop inhouse expertise in the cutting-edge tools best suited for our data needs. That last objective has allowed us to evaluate and contribute to other emerging technologies, adopting new tools with relatively little disruption.

VIRTUAL PRIVATE CLOUD VS. PRIVATE CLOUD

Our decision to move to a virtual private cloud using AWS instead of our own private cloud was made early in the process and was a natural outcome of our objectives, none of which would have been possible with a private cloud solution. This solution allowed us to decrease our infrastructure spending (Objective #1) by:

- Provisioning for an average load and dynamically expanding our computing resources to handle peak loads, rather than maintaining a fixed infrastructure cost base dictated by peak loads
- Purchasing resources at the time of need instead of incurring capital outlay six to nine months in advance
- Taking advantage of Moore’s Law cost efficiencies as new hardware emerged rather than waiting for the typical three-year depreciation cycle on purchased hardware to expire before exploiting new and more cost-effective generations of hardware

> **Our decision to move to a virtual private cloud using AWS was a natural outcome of our objectives, none of which would have been possible with a private cloud solution.**

Objective #2, the automation of production support, requires a highly scripted, API-driven platform layer over the hardware infrastructure. We explored the option of developing this layer internally and decided against it for several key reasons. First, there would be a high level of investment required in the middleware, which would divert funds away from our development of technology to support core business objectives. Second, we would not be able to bring to bear the same level of resources as a company with a broad customer base. Third, the gap between our custom middleware and cloud providers’ PaaS offerings would surely increase over time, making a homemade middleware solution progressively less viable. There are a host of third-party middleware solutions oriented toward bringing automated cloud platform–like functionality to a private data center. These options were rejected because fundamentally they required a private data center and would offset the commodity infrastructure savings goals of Objective #1.

Objective #3, which was to provide business analytics through innovative uses of commodity hardware resources, was naturally suited to a virtual private cloud solution hosted by a large-scale infrastructure provider like AWS. The economies of scale provided by this solution, coupled with the ease and cost effectiveness of rapid and temporary provisioning, eliminated the private cloud option.

By the completion of the migration program, all significant market regulation systems will have been migrated to AWS. A hybrid environment would entail increased complexity and cost without any tangible benefits. Of course, during the migration program, by definition we have been operating in a hybrid environment with some applications having transitioned while others are still in the process of changing.

> **The scale of operations in AWS allows for approaches to security that would be impractical in an enterprise private data center.**

ADDRESSING THE SECURITY CONCERNS

We performed an exhaustive analysis of cloud security as part of our planning. Rather than evaluate AWS cloud security against a theoretical ideal case, we took the practical approach of comparing AWS security against what FINRA can actually achieve in our private data centers. The analysis concluded that cloud security exceeds our private data center capabilities.

Any Internet-connected data center, whether privately built, colocated, or cloud built, requires best practices security safeguards such as intrusion detection and malware scanning. These are our responsibilities regardless of whether we are in a traditional data center or in the cloud. With this understanding, we turned our attention to the commonly raised security concerns surrounding cloud-based infrastructures. These are rooted in two issues:

1. Multi-tenancy risk

2. Insider threats

At the core of the multi-tenancy risk is the concern that hardware resources are virtualized and one is unaware of other parties running on the same virtualized hardware. The issue is whether a party could bypass the various security safeguards of the virtualization software and gain access to your data. This concern is mitigated by two factors. First, the sheer scale of a large cloud provider that dynamically allocates workload across hundreds of thousands of machines provides a high degree of anonymity. If you don't know who your neighbors are, they don't know who you are either, and it is extremely improbable that they can find you. Second, we chose to encrypt all data in the cloud, whether at rest or in motion. This combination of factors effectively mitigates any practical multi-tenancy risk.

The risk of an insider threat at the cloud provider is analogous to the same threat in a private data center. We found this risk to be significantly lower with a cloud provider than in a private data center due to the former's scale of operations. To begin with, the data is striped over tens of thousands of disks in tens of data centers, so it is simply not possible for a cloud provider employee to remove a hard drive belonging to a particular company. Furthermore, higher-level access to data by insiders in an infrastructure team is much more complex due to the separation of duties that can be achieved when operating at the scale of a cloud provider. Thus, the barrier to coordinated collusion is much greater than in an enterprise-level data center. When combined with the data encryption mentioned earlier, we concluded that our risk of an insider threat is lower with a cloud provider than in a private data center.

Other mechanisms offered by the AWS cloud in particular provide us with greater security than we could achieve in a private data center. For example, software-defined networks let us effectively use and manage micro-segmentation, with firewall groups that allow an application server to access only one database server.

The general theme in these findings is that the scale of operations in AWS allows for approaches to security that would be impractical in an enterprise private data center. These approaches range from increased separation of duties to a level of investment required in security R&D and infrastructure that is not practical unless amortized over a large number of enterprises.

INCREASES IN BUSINESS VALUE

Increasing business value as a result of reimplementing our systems on the cloud was a goal from the outset. We believed this was necessary in order to provide a concrete basis on which to base architectural decisions.

One of the initial business-oriented goals was to provide rapid end-user analytics by utilizing open source big data platforms, particularly Hadoop and HBase. As an example, a commonly used system at FINRA responds to user queries and assembles complex graphs of securities trades across multiple execution venues by querying petabyte-scale data. The incumbent system in the private data center provided a response time of between 20 minutes and 4 hours for commonly executed queries, with times varying according to the complexity of the trading graph and query parameters. This system was reimplemented early in the program using HBase and harnessing massive compute clusters to provide a new system, which reduced query times to between a sub-second and 90 seconds for similar queries.

With successes such as this, our goals for business benefits have broadened to make the program into a joint technology and business effort.

THE SELECTION OF OPEN SOURCE TECHNOLOGIES

FINRA's decision to move from its proprietary platforms to an open source one and from an on-premises environment to the cloud was driven by the following factors:

- Moving a legacy cost basis to the cloud makes little sense.
- Functionality for application servers, relational databases, and ETL has become commoditized and is ripe for use of open source.
- In a market for big data platforms that is highly fragmented and rapidly evolving with new technologies and no clear winners, open source provides the greatest agility and ability to both move to new technologies and take advantage of platform innovations as they emerge.

Our private data center environment had the typical enterprise mix of Oracle and SQL Server databases. We made the architectural decision to use the Postgres and MySQL open source databases in conjunction with Amazon's Relational Database Service (RDS) and immediately benefited from the scalability and multiple Availability Zone resiliency provided, along with the elimination of our system database administrator burden. Within RDS, we chose Postgres for large-scale systems and MySQL as an option for storing small application states. We allowed for deviations from these choices with justification and permission at the senior VP level, but interestingly, to date no teams have made a request for deviation.

The choice of open source in the big data arena was also accompanied by a choice to build inhouse core engineering competence in big data platforms. This skills development was coupled with strategic partnerships with key big data platform support vendors, including Cloudera, Pentaho, and AWS.

In all of these cases, we found that the open source decision was met with enthusiasm by inhouse development staff and was generally viewed as a way of enhancing and updating their technical skills.

THE ROLE OF DEVOPS

Prior to embarking on this program, FINRA had a fairly mature DevOps capability through the automation of builds and software deployment, along with a very exhaustive regression test suite for key systems.

These capabilities have been a necessary cornerstone of our cloud program. Our approach to operating system patching illustrates the importance of DevOps in this context. OS patching occurs as part of the build cycle instead of through the traditional private data center approach of applying patches to groups of machines. In the new model, the following steps are taken during the build and deploy pipeline: application code is built, an OS image is built with the latest patches and security updates, the two are combined into a single package, and the regression tests run on this image. Upon successful completion of the regression suite, the package is deployed to machines that are themselves deployed dynamically as part of clusters, autoscale environments, and static configurations.

In this setting, DevOps automation is used to eliminate the costs and unreliability of manual and semi-manual deployment processes. Perhaps more importantly, critical operating security updates can be applied on a continuous basis as part of an automated pipeline.

Extensive automation and scripting of our AWS application stack have further paved the way for eliminating much of the manual compliance checking that occurs in a traditional data center. Capacity reports, configuration checks, policy enforcement, monitoring for exceptional access, and other commonly performed administrative tasks are automated.

DISASTER RECOVERY

The traditional model of maintaining separate data centers for production and disaster recovery is superseded by the multiple Availability Zone facilities of a cloud provider like AWS. The multitude of data centers available in a local geographic area, together with redundant power grids, different flood plains, and redundant emergency fuel supplies, introduce a new and more reliable model for disaster recovery than the more limited minimum-distance, two–data center approach.

As part of rearchitecting our applications, we specified that systems would be brought up in arbitrary, rotating Availability Zones and data centers during normal operation, thus ensuring fault tolerance for disaster recovery purposes. We also chose to utilize the US West region of AWS as a backup in the event of nation-crippling disasters, however remote the possibility.

CULTURE SHIFT

As part of our shift to cloud and open source platforms, we chose to introduce a number of culture changes. Early in the process, we decided to make the cloud migration a rallying cry for the technology organization. Specifically, we challenged senior technology staff regarding the fundamentals of what our systems did and how well they served the business. This resulted in key changes in the way we addressed the fundamentals of our multi-petabyte, big data problem. In this process, new high-potential technology leaders were identified and elevated in the organization. The hiring and staffing effort that accompanied this effort also provided an opportunity to further reshape the technology profile of the company.

Another culture shift has been related to infrastructure support and operational staff. There has been a clear reduced need for traditional operations and infrastructure staff due to the use of infrastructure and platform as a service. We have capitalized on the availability of API-controllable infrastructure and platforms by further automating production support, operational, monitoring, and reporting tasks with the goal of eliminating all manual work in this area and repurposing operations staff to script-writing DevOps roles.

Within software development teams, there has been a drive toward further emphasis on regression test suites. This has led to additional blurring of the tester and developer roles, with test suites being written by team members who have the same skills as those writing software features. I expect that the tester role will continue to be blurred with the developer role and lose its distinction in the near future.

Perhaps the most important culture shift has been in the profile of developers we seek to attract. With the combination of cloud, open source, and big data platforms, we require and hire the same profile of developer that product companies are seeking. Similarly, there is greater focus on software development acumen in managers along with the traditional managerial skills and business domain knowledge demanded in most IT environments.

Within this context, activities such as hack-a-thons — usually associated with technology product companies — have taken a prominent role in staff retention and career development. Projecting forward, we foresee a trend to place more emphasis on college graduate hiring in combination with the more experienced middle and senior managers.

CONCLUSION

As of this writing, we have completed 22 months of a 30-month program to rearchitect FINRA's market regulation portfolio in the AWS cloud. Approximately 70% of the systems are in production on the cloud, and the program has been a success by any measure. This success has now turned our attention to migrating the remainder of our portfolio to the cloud and utilizing the processing power, storage, and flexibility of the cloud to further our analytic capabilities in areas of pattern recognition, machine learning, and other data analysis capabilities.

For other enterprises contemplating such an effort, I would summarize three key considerations from our journey. First, it is important to gain hands-on knowledge about the cloud early on by assigning a group of motivated and highly competent programmers to develop prototypes and proofs of concept. This allows myths to be debunked and subsequent analysis to be grounded in practical experience and based on fact. Second, to meet our objectives, we found it necessary to rearchitect our systems and fully utilize cloud and open source functionality. Third, and perhaps most importantly, we chose to focus on consistent delivery of significantly increased business value. This focus gave teams a concrete basis on which to make architectural tradeoffs and framed the project as an enterprise project with business support rather than as an infrastructure upgrade.

While these are clearly not essential elements for all companies contemplating cloud migration, failure to incorporate these factors could well result in a lost opportunity to move an organization to the next level of capability.

Saman Michael Far is Senior VP of Technology at FINRA, responsible for development of technologies to monitor financial markets, discover wrongdoing, and provide analytic visibility into over 60 billion market events received per day. Previously at FINRA, he was responsible for developing surveillance and regulation technology for the regulation of financial firms. Prior to joining FINRA, Mr. Far worked with venture capital firms to assess over 100 venture-funded startups and, where required, manage the rebuilding of teams, productization of core technology, and maximization of total addressable market and valuations. In the 1990s, he was one of four key people who led Open Text Corporation from an early stage search engine pioneer through a successful NASDAQ IPO. Mr. Far has a bachelor's degree in engineering science and a master's degree in electrical engineering/computer science, focused on distributed real-time data processing, from the University of Western Ontario. He can be reached at saman.far@finra.org.

Adopting IaaS: The Legal and Security Issues You Can’t Ignore

by Annie C. Bai

Cloud computing is trending because of the benefits of cost reduction, scalability, control, and flexibility. The economic incentives to purchase your computing infrastructure from specialized cloud providers (i.e., IaaS) are swiftly realized. Your business can immediately shed the capital and operating costs of paying for privately owned equipment and the staff to manage and maintain that equipment. The risks of IaaS, in comparison, are more nebulous and long-term, and they are harder to predict because the factors contributing to these risks are out of your hands. Data security is now the province of your IaaS provider; data integrity is also highly impacted by your provider; and data legality is affected by international regulations and politics. It is always a tricky situation when your cost-benefit analysis necessitates weighing matters that do not compare well — not so much apples to oranges, but apples to pernicious vine seeds that may or may not sprout and overtake your apple trees.

The issues detailed below can and should be addressed prior to implementing an IaaS product, and to whatever extent possible, by your legal agreements with your provider.

SECURITY IS STILL AN ISSUE

First off, it is important to understand that information security can never be fully outsourced. No matter what you pay for, administration of IaaS resources must be undertaken with care. This care should be commensurate with the sensitivity of the data that is being stored remotely. Another consideration is whether you are using a public cloud, private cloud, or a hybrid situation. Deploying and administering your infrastructure is still your responsibility. Permissions must be properly configured, keys must be managed, applications must be developed and updated with security in mind, and employees must be trained to avoid spearphishing and social engineering attempts on their credentials.

It is tempting to consider an IaaS package to be the “one and done” and get back to work. Someone else is doing the work now, right? No: that is like ditching your car and hiring a chauffeur ... without doing a reference check on the driver or giving him directions. How is he going to get you where you need to go? With any valuable resource, you must tailor it to your needs and manage it well. In the eyes of your customers and the law, you are still responsible for the security of your data, so do your due diligence before choosing a provider. Ask potential vendors probing questions about their security certifications, their policies and risk-control processes, their technical mechanisms, and whether they undergo external audits. These are considered industry best practices. Your business should absorb this information and match it up to your internal capacity to carefully manage the IaaS product. Security companies such as Symantec have detailed the myriad assumptions and mistakes that administrators make in adopting an IaaS platform.[1] This process is known as performing a risk assessment, which organizations should always undertake when making an important business decision.

> **It is important to understand that information security can never be fully outsourced.**

DATA INTEGRITY IS ALSO STILL AN ISSUE

Network access, data integrity, and data availability must be considered as you shift your proprietary information into the cloud. Cloud providers have specialized expertise in providing computing infrastructure, yes, but that does not mean that their service is not susceptible to breakdowns or interruptions. Ultimately, it is your business that will suffer the consequences of any interruption of access or any flaws in your data integrity.

From the business angle, you should be prepared with a business continuity and disaster recovery plan. From the legal angle, you should ensure that your service-level agreement (SLA) specifies, in terms that you find acceptable, how and when the provider will provide its services, address interruptions, and offer redress when there are disruptions to service. Not only will you need access to your corporate data on demand, but your customers and clients may have the right to demand access to or amendment of it, or the right to know that it is maintained with integrity. In the US, healthcare patients have a right to request their medical records,[2] students and their parents have the right to view educational records,[3] and so on. Your ease of access to customer information is integral to compliance with the many consumer privacy protection laws.

> **I can hardly think of a commercial situation these days that does not invoke problems around privacy and confidentiality.**

Conversely, you do not want the IaaS provider to have free access to your data. The service agreement should explicitly specify "hands off" the data that the provider stores for you, unless access to it is required in order to provide you services. There should be no opportunity for the provider to monetize or otherwise use your data. There will be some unforeseen hiccups with IaaS, but their impact will be mitigated if your operations plans are in place and your legal protections are in order.

LOCATION IS NOW A BIG ISSUE

Before you moved your computing infrastructure to the cloud, it may have taken a lot of effort to maintain and store said equipment, but you certainly knew where it was — in your server closet, room, or whatnot. But when it comes to IaaS, where exactly are your computing resources located, and where is your data located? These questions, in turn, trigger a panoply of other questions with regard to data privacy regulations.

The Internet and "the cloud" seem to promise a uniform situation with universal availability from any location. In reality, there are myriad political and legal factors that inconveniently pull back the veil on this fiction. The location of your data may cause it to be subject to international data protection laws and even foreign government surveillance. Since Edward Snowden revealed the US National Security Agency's (NSA) surveillance activities, many foreign countries are no longer accommodating of the interests of American companies in transferring data across nation-state lines. In the US, we view data privacy as a consumer and market-based issue. For example, we are fairly comfortable with the tradeoff between allowing our personal information and behavior to be leveraged in return for the free Internet and free apps on our smartphones. In many other parts of the world, privacy is a more visceral issue and is labeled a fundamental human right. From religious texts invoking privacy to the experience of Nazi use of data to create population identification systems, there is a deep and troubled history with piercing the shield of personal privacy.

I provide this historical background to underscore the contemporary challenges for companies with international customers (which includes nearly every Internet commerce business). Until October 2015, 4,000 US companies operated under a Department of Commerce program with the European Union that permitted them to transfer the personal data of EU citizens across national lines. As a consequence of the furor over the revelations of global NSA surveillance, that agreement was just invalidated, without concrete plans to either replace it or enforce the current gap in the law. Countries throughout the rest of the world are closely watching the US-EU standoff and have voiced their own complaints and concerns about international data transfers. This situation adds another layer of risk and uncertainty for any company seeking to store its data in the cloud. If you have no business need for your data to cross the Atlantic, then make sure that you voice that preference in your contract and make provision for extracting it whenever you need to.

WHEN IS PRIVACY NOT AN ISSUE?

I can hardly think of a commercial situation these days that does not invoke problems around privacy and confidentiality. The US does not have an overarching data privacy regime but instead regulates the matter via many sector-specific laws. There are data privacy regulations that protect Americans as healthcare patients, financial and credit customers, students, drivers, video consumers, telemarketing targets, and so on. Broader laws exist in the area of data breaches, and even broader authority lies with the US Federal Trade Commission, which can investigate and fine companies for privacy infractions in the name of fair trade practices.

In our cyber landscape, data breaches are a constant. An IaaS vendor should be expected to satisfy certain confidentiality and security standards for protecting your systems and their contents. It should also notify you about security incidents that *may* have compromised your data. In the unfortunate instance of a breach of information or other harm caused to your business operations, your contract should apportion the responsibility and liability between you and your provider. These provisions will affect your ability to respond to and investigate breaches in an era where timeliness is expected by regulators, the media, and consumers.

Furthermore, the distributed service model works because IaaS providers colocate users' virtual machines and provision their resources to maximize customer usage. This is an efficiency, but also a potential problem. When the customer's machine is virtual, but the physical server is actual, guess which reality trumps the other when security is breached? The consequence of sharing the same machine is that colocated data may be susceptible to leakages to other clients located on the same server. Recently, researchers found a way to steal Amazon's secret encryption keys in a side-channel attack on a colocated client. This hacking took place in a lab environment and hence is still a hypothetical threat. As similar threats have been exposed over the years, reputable providers have addressed them. Nonetheless, it demonstrates why it is critical that your cloud provider be a reliable major vendor.

In sum, legal protections can serve as a safety net to cover any gaps in the security of the actual resource. Legal provisions and security best practices should work in concert. One without the other is insufficient in the area of IaaS products, because this is neither an established area of legal doctrine nor of business standards. In addition to the matters detailed in this article, there are business and sector-specific legal issues with regard to intellectual property, trade secrets, foreign direct investment, and corporate governance when you move to an IaaS paradigm. To reliably protect your business interests, you must take matters into your own hands. Work with your legal and technical support teams to ensure that you are comprehensively assessing the risks of moving to IaaS and designing sound administration practices for managing this valuable resource. Remember, security can never be wholly outsourced!

ENDNOTES

[1] Wueest, Candid, Mario Ballano Barcena, and Laura O'Brien. "Mistakes in the IaaS Cloud Could Put Your Data at Risk." Symantec Corporation, 2015 (http://goo.gl/cRdWUm).

[2] Under the Health Information Portability and Accountability Act (HIPAA), patients have the right to access and amend their medical records. This applies to business associates who provide contracted-for services to healthcare providers. See: "Right to Access Medical Records." US Department of Health and Human Services (HHS) (www.hhs.gov/ocr/privacy/hipaa/faq/right_to_access_medical_records).

[3] Under the Family Education Rights and Privacy Act (FERPA), eligible students have the right to access and amend their personal information. See: "Family Educational Rights and Privacy Act (FERPA)." US Department of Education (www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html).

Annie C. Bai, CIPP/US, CIPM, is a graduate of NYU School of Law. Ms. Bai is Privacy Counsel at Single Stop, a national anti-poverty nonprofit and speaks on modern privacy law for New Directions for Attorneys at Pace Law School. She consults with for-profit and not-for-profits on privacy and data security on privacy audits in a variety of industry sectors. Ms. Bai is a member of the Technology Operations Committee for Per Scholas, an IT workforce training nonprofit. She can be reached at anniecbai@me.com; Twitter: @AnnieCBai.

Darkness on the Edge of Cloud

by Vince Kellen

A lot has changed in a few years.

When I talked about cloud three years back, I got frownie-faces from my peers. Skeptical looks that belied a deeper-seated fear or trepidation, probably having more to do with their internal image of what a CIO should be than the promise or peril in the new technology.

Now, enthusiasm runs ebulliently through the vendor community, animating the animal spirits and spurring on entrepreneurs in search of profits and glory. Cloud has been elevated to high strategy on the billionaire chess board. Mergers and acquisitions are abuzz. Amazon, armed with an overly energetic workforce, gets hypercompetitive in all ways good and ill, supplanting Oracle as one of our most vociferous vendors and perhaps the new alpha predator. Numerous smaller vendors — tiny even in the aggregate, compared with Amazon's might — are quickly learning the new cloud lingo, differentiating themselves from Amazon and contemplating symmetrical and asymmetrical warfare. Today it's Everyone vs. Amazon.

In the CIO office, hybrid cloud is not avoided and is now assumed. CIOs are no longer fearing it, but trying to rope it in. The corporate data center business is officially in decline. Legal staff are taking a closer look at all the agreements the business people previously signed willy-nilly. Like a formerly thrill-seeking teenager who has sworn off having secret parties while her parents are away, business users may be getting over their glee in the power of their corporate credit cards and starting to work more maturely with central authorities in cloud contracting.

Amidst all this market kerfuffle, my teams are currently immersed in a slightly provocative and full-throated effort to vacate our data center. We have been looking around to see how the marketplace has shifted, where it has not, and, more importantly, what threats we are likely to face as we all go to the cloud. We have found seven of them.

THE 7 BASIC THREATS IN CLOUD COMPUTING

1. Sticky Mental Models and Behavioral Inertia

Despite the vendor enthusiasm for enterprise-class infrastructure as a service (IaaS suitable for corporations big and small), we have discovered that a fair number of these vendors, even the leaders, keep applying old managed services and colocation contract thinking. For me, the key difference between cloud infrastructure and all other forms of procurable infrastructure is that in cloud infrastructure, you pay based on actual consumption of CPU, memory, storage, and networking use, and in much smaller increments. The fundamental algebra in the pricing is based on virtual servers, virtual storage, and virtual networks. In old managed services and colocation deals, the basic economic unit is square footage in rack space.

The two pricing models are very different. In a square-foot rack space pricing model, the customer is responsible for all unused cycles, unused network usage, unused memory, and unused storage. In a true cloud contract, the vendor is responsible for managing the vacancy rate. While this sounds hard on the vendor, it isn't. Vacancy rates in corporate data centers are very high. A cloud vendor can resell that vacancy and collect more profit per square foot as each server is potentially maximally used. Moreover, companies all vary in their usage over time. The more heterogeneous and larger the customer base, the more predictable aggregate growth and consumption patterns become.

The problem is that vendor salespeople and architects can't seem to get colocation thinking out of their heads. We find ourselves constantly reminding vendors to pursue Amazon-like pricing models. Even after the contract marriage ceremony between us and the vendor, we still need to remind them to be faithful to their contract promises as we walk down the aisle. If anything defines the competitive landscape in the cloud, it is exactly this inertia. The only way Amazon's competitors can match

Amazon's price-performance ratio is to maximize their infrastructure usage. The only way consuming companies can break through to new levels of IT price performance is to shatter these sticky mental models and ensure behavioral compliance from everyone each step of the way. Change is hard even when you would think profit motives align. For all parties, the real "switching cost" to a new model is the effort to learn the new cloud dance.

2. Contract War

Enterprise IaaS is different from what we know as "public cloud" in one respect. Enterprise contracts have terms and conditions altered to address the needs of the client. Typical areas include exactly what risks and liabilities each party will accept and which risks and liability they cannot accept, which national or state laws must be respected, in which country or state parties can sue, and contract language that might have to address custom technical configurations or integrations.

In addition to these contract issues, the large public cloud providers — Google, Amazon, and Microsoft — all engage resellers to handle transactions with companies. For companies wishing to use the big three in their portfolio of cloud options, reseller agreements become important. Amazon may have specific terms and conditions that it wishes to ensure are present in all agreements its resellers make. Right now we are detecting fights between Amazon and its resellers as to whose agreement reigns supreme. In many cases, Amazon has the right to terminate reseller agreements on short notice. If services have been prepaid, it might not be clear if the end customer is out money and workloads should a reseller be terminated. Furthermore, each of the big three can decide to manage their reseller channel differently, giving resellers more or less discretion on what additional contract terms the resellers can offer.

Unsurprisingly, this new value chain has sprouted, looking very much like value chains in other industries with all the attendant power struggles between players. Unsuspecting client companies may become collateral damage along the way. I urge everyone to read all these agreements and contracts carefully, or hire someone with the technical, financial, and legal expertise to help. Yes, there are vendors very willing to part fools from their money.

3. Black Swan Events

While on the surface Amazon looks like a juggernaut poised to be a monopoly in the IaaS market, its experience with diverse and complicated needs from midsized to large corporations is rather thin compared with the long-time data center outsourcing firms of yore. From my perspective, the enterprise IaaS market is a jump ball. It is not clear anyone will dominate. However, should Amazon or Microsoft or Google become a near monopoly in this market, it raises a question: will they be too big to fail? While people often scoff at the idea of one of these firms having a drastic financial or technical failure, given events in the last couple decades, this question deserves a good hard look. It is very likely a sizeable percentage of the worldwide economy will be dependent on this new infrastructure.

For all parties, the real "switching cost" to a new model is the effort to learn the new cloud dance.

If software products based on IaaS become complex and depend on many other vendors and firms using the same or different IaaS providers, a failure of one of these firms could cascade to all of its customers and so on. While today the IaaS supply chain is probably very flat, meaning each software vendor does not depend on another firm also dependent on IaaS, the future holds no such guarantees. Since the supply chain in IaaS can be obscured from view, it is impossible to know what the interdependencies may be. Such murky network failures were at the heart of the 2008 financial meltdown, as firms were unsure which of their financial products had hidden risk in them.

Other disturbing scenarios assuredly lurk, or so my nightly dreams tell me. We need people now who excel in these scenarios to counsel companies. With so much blue sky thinking abounding, we need black swan[1] experts more than ever.

4. No Credible Threat of Defection

Great salespeople instinctively know when customers are locked in — and can exploit the situation when they are. Don't fool yourself that this calculation does not cross the minds of your favorite vendors. They would not be worth their weight if they hadn't honed this basic instinct. What firms have to do is invest enough in diversity in the IaaS supply chain to maintain a credible threat of defection. Unfortunately, the current cloud contracts tend to charge you nothing to place data in their flytrap IaaS environment; they only charge you when you take the data out. These costs can be significant if not exorbitant. Like Hotel California, you can check in anytime you like, but you can never leave.

What firms have to do is invest enough in diversity in the IaaS supply chain to maintain a credible threat of defection.

The good news is that there are a few IaaS providers that have attractive network egress pricing. Over time, I suspect and hope that we will see different pricing models to serve different customer needs. However, this won't happen as long as CIOs are incorrigibly risk-averse and keep securing long-term, single-source contracts with no credible threat of defection.

Compared to other industries, the IT supply chain is overly simplistic. Moore's law has aided this process, providing us with new crops of CPUs and storage options that deliver a lot more for a lot less. This, too, is ending. But if CIOs don't establish enterprise architectures and associated procurement vehicles that provide possible escape routes, the new enterprise IaaS will look very much like the old managed service colocation contract.

5. Poor Workload Resource Consumption Vector Analysis

If one carefully measures the actual RAM, CPU, storage, and network usage an application actually consumes and then compares this to the resource bundle pricing formula Amazon or other enterprise IaaS providers utilize, you will notice glaring mismatches. The application may consume less CPU or RAM than the vendor offering provides. The "vacancy rate" (percent of unused resources) for the application might be still too high. While the IaaS vendor may be giving what looks like an excellent price, customers may be paying two to three times what they should.

What can be done about this? This is a hard problem to solve. Vendors cannot resell your unused resources unless they can define the contours of the container precisely around your workload parameters. Doing this would require a level of granularity in pricing (metering) and in technical isolation that may not be feasible. Nonetheless, Amazon leads the way with a variety of package "sizes" (instances) to choose from. What companies need to do is measure their workloads carefully and select the right resource consumption package and pricing model from the vendor.

Companies will also have problems judging whether the services the contract says they are receiving are actually being delivered. In industries that move physical goods, independent verification of the meter or scale being used to measure the commodity is common. Perhaps down the road we may see independent metering companies, probably looking like cloud contract auditors, which will require providers to open their books and show clearly that the service described is accurately measured.

My hunch is that this pricing flexibility will be a promising area of development for competitive IaaS providers. Can virtual offerings be dynamically reconfigurable to constantly seek the best match between the pricing vector and the workload consumption vector? At least one curious mind wants to know.

6. Continued Merger Activity

Oh my! In the space of a few months, a good piece of the emerging enterprise cloud market quickly aggregated. Some time ago, IBM bought SoftLayer, Cisco bought Metacloud, EMC bought Virtustream, and Dell bought Enstratius. Just recently Dell saw everyone's bet and raised them, oh, $67 billion by buying EMC. Game on!

While aggregation does provide some scale advantage, the removal of choice in the market is concerning. Most big IT vendors do not like to compete on price day in and day out. With infrastructure cloud moving to a commodity and Amazon setting the competitive table, the aggregation occurring is designed to help the acquiring firm reduce competition and get scale. An opposing force to this consolidation pressure is midsized competitors that can cleverly manage their customer workloads and still achieve great price-performance ratios. A looming question is whether this opposing force is enough to ensure diversity and competitiveness in the supply chain long term. Will the market get dominated by four or fewer players providing 80% of the supply? If so, I don't think this development will be good for consuming companies.

7. Data Sovereignty Concerns

Perhaps the most intriguing of trends is the notion that countries are taking control of their cyber infrastructures and treating them as national assets worthy of control and defense. For multinational companies, this poses a challenge in that workloads will need to be carved up along national lines and, more specifically, around the data in those workloads. This, of course, provides opportunity for cloud providers to add services ensuring that workloads reside in different regions.

But what happens in a state of war? What does a global telecommunications and data infrastructure look like when the economic foundations of most counties are supported by international flows of data and a future infrastructure is demarcated by national boundaries with data "immigration officers" at each border? For now, the more vexing data sovereignty issues are contained and pesky, not virulent. But what of tomorrow?

If this data sovereignty pressure continues or grows, I suspect more interesting network and workload designs that package security with the data will be much more prevalent. Parents: tell your kids to go into IT security!

WHAT WE DO NEXT WILL DETERMINE THE OUTCOMES

An old maxim is that no individual determines market pricing; the market does that. Likewise, no individual company, buyer, or seller in this new value chain will fully shape the final outcomes. To a large extent, the powerful forces of fear and greed, collectively expressed one deal at a time, will shape this market space. Together, what CIOs and their companies do next will determine the outcome.

Perhaps the critical questions for CIOs are these:

- Do you believe a change in IT supply chain management is occurring and is going to affect you?
- Does a diverse, competitive supply chain in IT make a difference for your firm?
- Do alternative models, not yet discussed adequately here or in the industry press, exist?
- Is your company positioned to take advantage of these shifting market forces?
- What will you do next?

From my perspective, it looks as if companies are wanting things that can only be found in the darkness on the edge of cloud. Our job is to shed more light on these things.

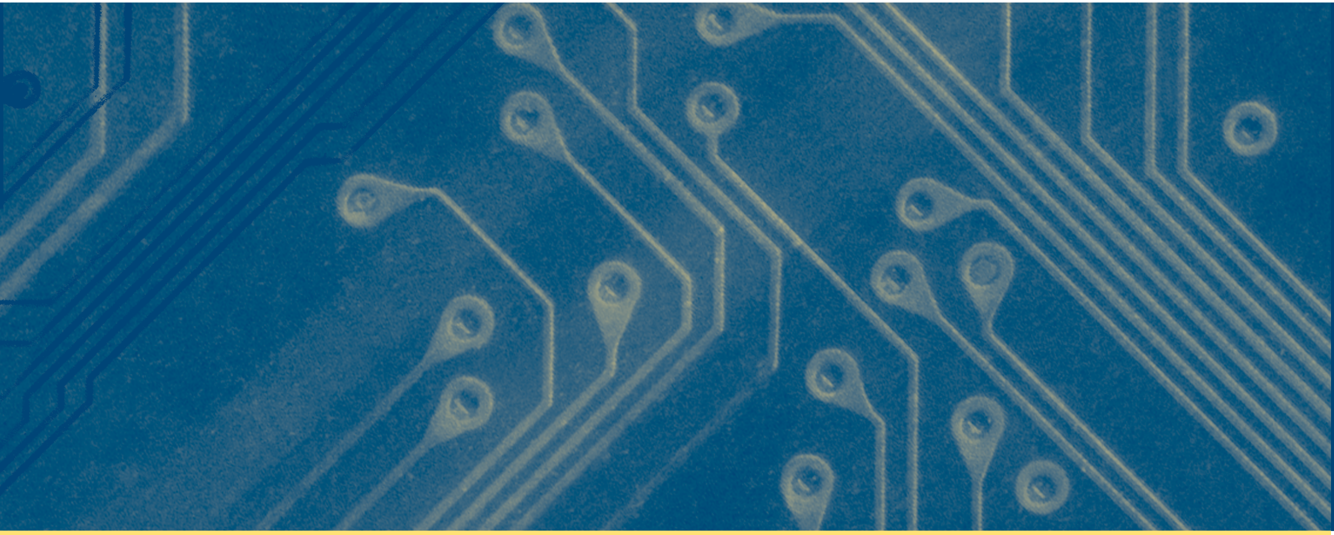
ENDNOTE

[1]"Black swan theory" (Wikipedia).

Vince Kellen is a Fellow of Cutter's Business Technology & Digital Transformation Strategies and Data Analytics & Digital Technologies practices and a member of the Cutter Business Technology Council. Dr. Kellen brings a combination of entrepreneurial, academic, and IT strategy experience to his role. He is currently Senior Vice Provost, Analytics and Technologies at the University of Kentucky, one of the top public research institutions and academic medical centers in the US. Dr. Kellen's areas of responsibility include institutional research and analytics, enterprise software, research computing, academic technology, IT infrastructure, and cloud services, supporting the university's 17 colleges, 30,000 students, and 14,000 faculty and staff. He has also served as a partner with strategy consulting firms, where he helped Fortune *500 and midsized companies with business and IT strategies, IT organizational development, customer experience management, CRM, and data warehousing and analytics.*

Dr. Kellen currently teaches at the University of Kentucky's College of Communications on IT and strategy and previously taught at DePaul University's College for Computing and Digital Media. He is a recipient of the coveted CIO *magazine Top 100 award in 2007 and 2014, was one of* InformationWeek*'s Chiefs of the Year in 2013, received the* Computerworld *Honors Laureate award in 2013, and was one of four recipients selected globally for Dell's 2012 Transformational CIO award. Dr. Kellen has served on advisory boards to top IT companies, including Dell, SAP, Microsoft, AT&T, and Apple, and cohosts (with IT futurist Thornton May) a regional Kentucky CIO forum called The CIO Practicum. He is the author of four books on database technology and author of more than 250 articles and presentations on IT and business strategy topics. Dr. Kellen earned his PhD in computer science (human-computer interaction), an MS degree in e-commerce, and a BA degree in communications, all from DePaul University. He can be reached at vkellen@cutter.com.*

Cutter
IT Journal



About Cutter Consortium

Cutter Consortium is a truly unique IT advisory firm, comprising a group of more than 100 internationally recognized experts who have come together to offer content, consulting, and training to our clients. These experts are committed to delivering top-level, critical, and objective advice. They have done, and are doing, groundbreaking work in organizations worldwide, helping companies deal with issues in the core areas of software development and Agile project management, enterprise architecture, business technology trends and strategies, enterprise risk management, metrics, and sourcing.

Cutter offers a different value proposition than other IT research firms: We give you Access to the Experts. You get practitioners' points of view, derived from hands-on experience with the same critical issues you are facing, not the perspective of a desk-bound analyst who can only make predictions and observations on what's happening in the marketplace. With Cutter Consortium, you get the best practices and lessons learned from the world's leading experts, experts who are implementing these techniques at companies like yours right now.

Cutter's clients are able to tap into its expertise in a variety of formats, including content via online advisory services and journals, mentoring, workshops, training, and consulting. And by customizing our information products and training/consulting services, you get the solutions you need, while staying within your budget.

Cutter Consortium's philosophy is that there is no single right solution for all enterprises, or all departments within one enterprise, or even all projects within a department. Cutter believes that the complexity of the business technology issues confronting corporations today demands multiple detailed perspectives from which a company can view its opportunities and risks in order to make the right strategic and tactical decisions. The simplistic pronouncements other analyst firms make do not take into account the unique situation of each organization. This is another reason to present the several sides to each issue: to enable clients to determine the course of action that best fits their unique situation.

For more information, contact Cutter Consortium at +1 781 648 8700 or sales@cutter.com.

The Cutter Business Technology Council

The Cutter Business Technology Council was established by Cutter Consortium to help spot emerging trends in IT, digital technology, and the marketplace. Its members are IT specialists whose ideas have become important building blocks of today's wide-band, digitally connected, global economy. This brain trust includes:

- Rob Austin
- Ron Blitstein
- Tom DeMarco
- Lynne Ellyn
- Israel Gat
- Vince Kellen
- Tim Lister
- Lou Mazzucchelli
- Ken Orr
- Robert D. Scott