Is Your Cloud Preparation Up to Snuff? Part 1

RFG Exchange: RFG is conducting a series of roundtable discussions about cloud critical success factors with gurus from user and vendor organizations. Enterprise users across diverse industry verticals are grappling with how best to leverage emerging cloud and container platforms so that they deliver the required flexibility, scalability, and security while still enabling cost efficiency and performance. In this edited transcript, we hone in on the challenges associated with various aspects of DevOps. All of "The Rounds" videos will be available on the new RFG Exchange Web site (www.RFGex.com), YouTube, and SoundCloud. Joining our host, Cal Braunstein, CEO and executive director of research at Robert Frances Group, are:

- Joe McKenna, former VP distributed hosting architecture and engineering at Fidelity Investments,
- Carlos Pereira, Distinguished Systems Engineer, worldwide DC and cloud, Cisco Systems,
- Diogenes Rettori, OpenShift Product Manager at Red Hat, and
- Jeff Schaefer, Senior Technical Staff Member, Chief Security Architect, cloud managed services at IBM.

RFG: When companies construct their business cases I've seen two scenarios arise. One is the view where line of business individuals or IT developers state they're going to the cloud, and usually they're thinking public cloud. Here they have in mind a business case that is tied to how much more rapidly they can get product out and drive a revenue stream or business loyalty. The other scenario is when it gets thrown over the transom to the IT operations staff. IT operations usually does not address the potential revenue stream; instead it thinks about hard dollar cost, and it struggles building a cost-based business case. Which way is right?

Carlos Pereira: I would start a little further back from the business case and look at the business side of the house. Companies are experiencing a lot of transformation. Organizations are getting more competitive, more agile, and evolving toward a digitized economy. This is the business transformation that's going on. These companies need to evolve their operational model towards a user consumption model. Cloud is just one operational model that allows them to consume faster and be more agile in support of the whole business case.

If you look at it from the developers view as they go to the cloud, it's a faster rate of change than the IT legacy environment. Here you run up against the risk assessment challenges that hinder or prevent them from going. And that remains true even if you evolved towards DevOps with the goal of being more agile.

The real business case is more than just as a view of how many dollars the firm can save or spend on a particular cloud. Instead it is how the company evolves its business transactions towards what it needs to go towards a more digital world, or towards a paperless world, or other new or different business models.

Diogenes Rettori: For me it's related also to the type of workloads. What justifies running on the public cloud, and what justifies running on the private cloud? Ideally I see a private cloud where you have workloads of different characteristics to the point where you can even attain 95 percent utilization on your private cloud. This is achieved because you have guiding principles and policies that drive how you define the characteristics of your workload, enabling the workloads

to be scaled up and down or disposed of. Properly selected workloads allow for a higher consumption of your internal resources and can justify either public or private clouds.

Jeff Schaefer: It comes down to what you need to run and can run in the cloud. If you have highly regulated, highly industry-driven workloads that can't run in a public place, you may not have choice. But it may not come down to that. Use of a cloud would get you some agility and some improvement in the time-to-market that you don't have with traditional infrastructure.

Joe McKenna: When we were trying to do cost comparisons and evaluating rank relative to our peers in delivering our services – compute, storage, and network – we compared ourselves not only in the financial industry firms but also to cloud providers. When you look at cloud providers, you have to look beyond the cost of compute and the cost of the amount of storage you're using. There's the network bandwidth and the storage bandwidth that you're starting to use. There are other things depending on the application that that goes back to Diogenes' point of what workloads are deployed out there. Maybe something that's extremely high I/O, whether it is it network or storage, might not be cheaper to put out there. It might be cheaper to build the infrastructure and deliver those services internally. It really comes down to understanding workloads, understanding the costs, and the structures, and the different pieces.

Diogenes Rettori: I can't stress enough the importance of using hybrid cloud environment to avoid lock-in. You should design for hybrid cloud when you build your guiding principles so that you force the design to support it from the start. One needs to design for workloads that can be moved across clouds. One should not design for lock-in with a vendor- or a provider-specific technology that will require a painful change to adapt to another vendor [sometime in the future].

Joe McKenna: We had heard some numbers in a conversation recently about the approximate $2.9 billion being spent quarterly to build out Google Cloud. A ridiculous amount of money that they're spending, and the amount of infrastructure they're putting into a cloud compared to a big enterprise. There's just no way to compete at that level of hardware being available for variable usage to bring up or down. When you get to that off-premise cloud, there's got to be some finite definition as to how far as your going to go. Now is it going to be for all my workloads at peak? Or am I going to save all my critical workloads at peak plus some buffer, and use good workload management to shut down the noncritical workloads kind of like we did, and still do, on the mainframe? When it gets busy on the mainframe, we shut down all of the developers and anything else needed and repurpose those MIPS to do the business. Similarly, one would shut down all the non-essential work and then run just the critical business applications on the private cloud.

If you're bursting... there's a gentlemen that came up with the term data gravity, and I said it's not quite right it's data anchor cause I'm bursting out and dragging the anchor with me as I'm trying to burst. I'm staging that data outside to burst out if I need it, or I'm going to take a huge performance hit. These types of approaches I am still not comfortable with. I'd love to hear some of the other feedback from the others as far as what they do.

Diogenes Rettori: I try to think of it from the perspective of the application that is consuming the resources. Everything that we do is to enable an application that represents business value to be achieved. Taken from the perspective of the application, workloads should be created in a way that allows for disposability, and allows for the ability to move across different environments. Although most of the workloads that are out there have not been created with this flexibility in mind. It's going to take some time for the cloud-native or even cloud ready workloads to be present that will avoid lock-in. These types of applications will enable a higher consumption of resources that you are paying for. Essentially you want to use 98% of the process resources. Anything that you don't use is money left on the table.

Ideally you should always be at peak with that because not being at peak means that there is money that you are paying for that you're not using. You can only be at peak if the workloads that you have that allow for such distribution and if the underlying technologies support it. For example, use of methodologies and procedures in packaging models like microservices that offer this capability and flexibility, but it's going to take a long time.

Jeff Schaefer: It's a little bit of misnomer that folks don't have the ability or the tools don't exist to understand your workload characteristics and capacity planning. I think it's out there. You can understand at least to a better extent than you could maybe five years ago what your workload planning should look like, and you can plan a little bit better. I'm not saying that's 100%.

Carlos Pereira: I have a forward-looking view about this. If I look at [non-SaaS] clouds – whether on-premise or public – you expose infrastructure to be consumed. When I talk about capacity planning or basically capacity augmentation, you want to look at how much of the workloads one puts in these boxes. One of the things that I believe will become the norm in the upcoming years is that you won't make those calculations anymore. You're going to use machine learning to predict this for you.

Why I'm saying this? Because you have a lot of existing legacy supporting existing businesses. In each case you have to have visibility into those applications to define how independent or dependent those application spaces are. At the same time you have a bunch of cloud ready applications on the cloud. All the pieces together – non-cloud and cloud-native – run a full microservices system. You need to have cohesiveness, or else how can you do capacity planning? The only way you do that is to have full visibility, and when you have full visibility you can define the dependencies. That is going to be a machine learning exercise. Why? Because it will be able, because of scale to predict what your capacity planning would be. Which if you go back to 10 years ago, we basically had an idea of our capacity using capacity planning tools designed for particular environments.

Today you have a vast availability of potential resources for developers. I would think it's becoming digital and more accessible. The technology empowers that. So capacity planning would be better, but with clouds is thought of differently. Cisco is developing, IBM is developing, as well as other emerging players with these new approaches to capacity planning. It's just a normal evolution. From that sense, capacity planning is and will depend upon full visibility of all aspects of the environment, and will also help predict or provide hints on how you should proceed.

RFG POV: The key is the operational business model that the enterprise intends to execute. The IT business case and the selection of the cloud environment and workloads must be designed to support the business model. IT executives should work with their line of business peers to develop the operations model and the supporting IT business case.

Additional relevant research and consulting services are available. Interested readers should contact Client Services to arrange further discussion or interview with Mr. Cal Braunstein, CEO and Executive Director of Research.