Cloud Redux

**Mahesh H. Dodani**, IBM, U.S.A.

1   BACK TO CLOUD BASICS

"The thing I love about my iPhone is that it's not a piece of technology I think about but rather, it's the way interact with it to get what I want done. It has its quirks, but it works…for millions of people.

The point here is that Cloud is very much like the iPhone. As Sir James (Urquhart) says "Cloud isn't a technology, it's an operational model." Just like the iPhone.

Cloud is still relatively immature and it doesn't have all the things I want or need yet (and probably never will) but it will get to the point where its maturity and the inclusion of capabilities (such as better security, interoperability, more openness, etc.) will smooth its adoption even further and I won't feel like we're settling anymore…until the next version shows up on shelves." – Cloud Maturity: Just Like the iPhone, There's An App For That… Chris Hoff Blog

As cloud computing matures, enterprises are realizing the importance of looking at cloud computing as an architectural platform that enables their IT to be aligned with their business and respond with speed to address the changing business requirements. As I have discussed in my previous articles on cloud, IBM views cloud computing as a new consumption and delivery model inspired by consumer internet services. IBM's definition of cloud computing is driven by five key delivery characteristics:

- On-demand self-service, which focuses on delivering IT services driven by user requests.

- Ubiquitous network access, which focuses on delivering IT services anytime, anywhere and through user-chosen devices.

- Location independent resource pooling, which focuses on delivering IT services through any resource pool that is capable of running the underlying workload and delivering on the required service-level agreements.

- Rapid elasticity, which focuses on delivering IT services through resource pools that can expand and contract based on the requirements of the underlying workload and the usage characteristics.
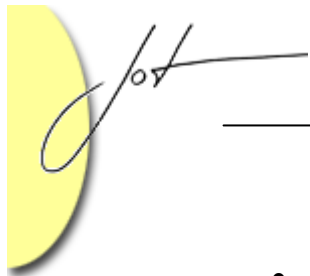
- Flexible pricing models, which focuses on delivering IT services that can be metered for usage and charged for (if needed) through pricing models including subscription, usage pricing (pay for use), etc.

In essence, the cloud computing definition below covers the capabilities needed to deliver on the characteristics above:

- Cloud computing is an emerging style of IT delivery in which applications, data, and IT resources are rapidly provisioned and provided as standardized offerings to users over the web in a flexible pricing model.

- Cloud computing is a way of managing large numbers of highly virtualized resources such that, from a management perspective, they resemble a single large resource.

  This can then be used to deliver services with elastic scaling. Enabling the focus on cloud computing as an architectural platform for IT service delivery, Figure 1 shows the separation of concerns among the service requestor, service provider and the service creator, and shows the capabilities required to deliver the different types of cloud services – infrastructure, platform and applications.
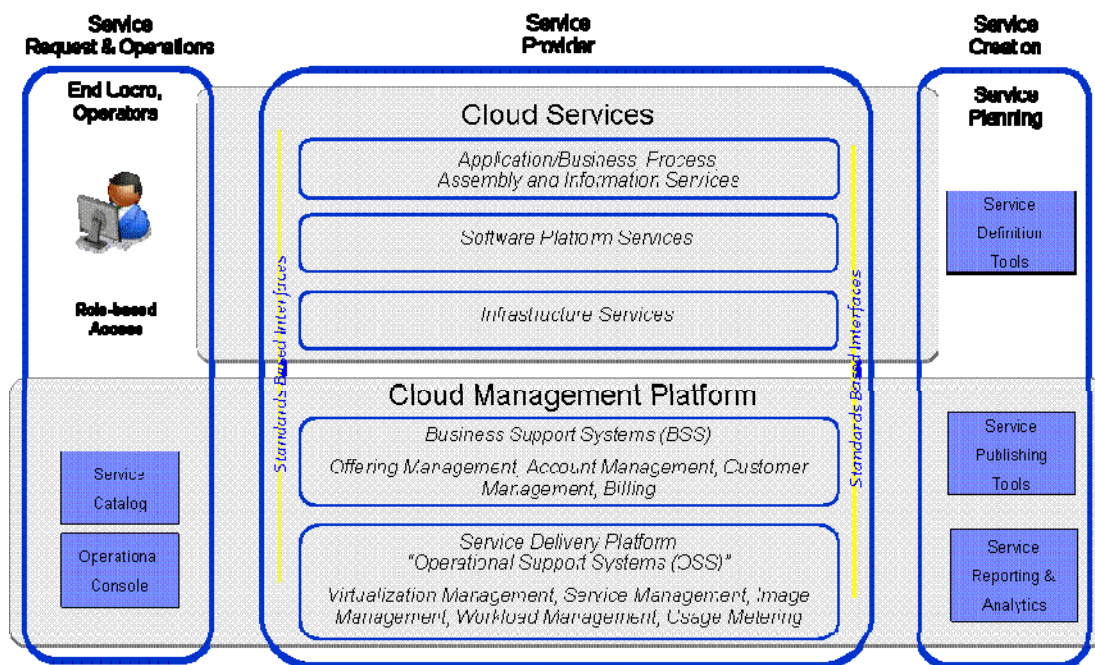

Figure 1: Cloud Architectural Model

Note the definition of a common cloud management platform that delivers the business support systems and operational support systems needed to deliver the different types of cloud services. The sophistication of these BSS and OSS capabilities depend on the level of characteristics needed to deliver the cloud services. For example, to support flexible

pricing models, a public cloud service provider would need all of the BSS capabilities along with the OSS metering capability. On the other hand, an enterprise that has chargeback mechanisms in place, will need the BSS billing capability along with the OSS metering capability.

Another component of the maturing of cloud manifests itself in the models for delivering cloud services. Though the initial focus of the cloud capabilities were in supporting private and pubic clouds, the realization that enterprises will have to manage multiple cloud delivery models has shifted the focus on capabilities to support such a mixed delivery environment as shown in Figure 2. In particular, capabilities are needed to support such an inter-connected cloud delivery environment including the support for federated identity; an "enterprise service bus" capability to allow the different service providers to work together and to handle service interconnections, flow and mediations; an event infrastructure to manage the events that are generated as cloud services are instantiated and used; and a data pipe that allows (potential large) workloads and data to flow between service providers.
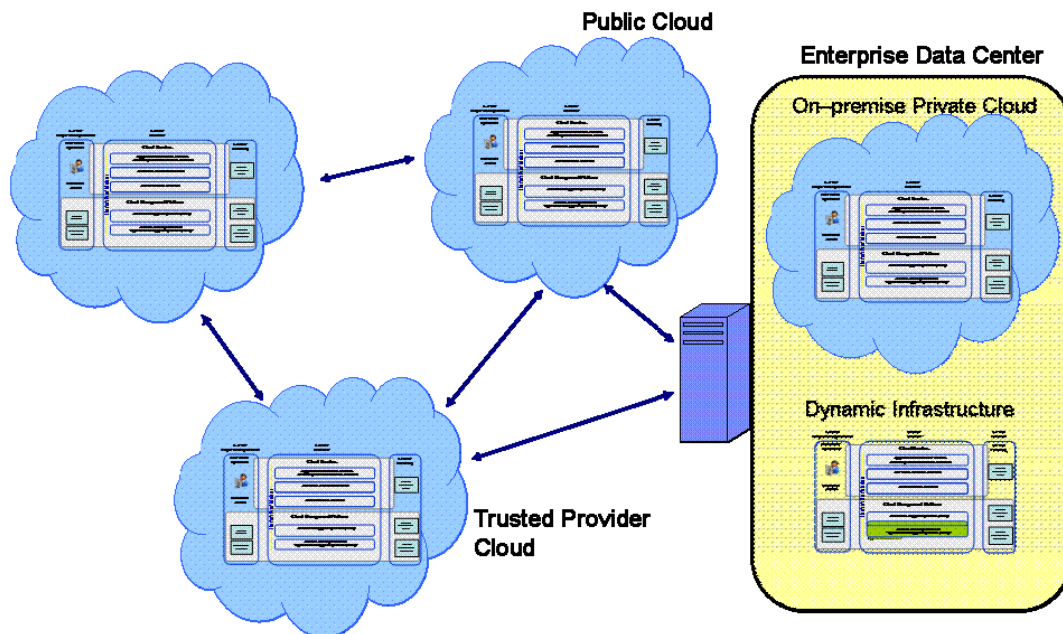

Figure 2: Cloud Inter-connected Delivery Models

The final aspect of the maturity relates to measuring the value delivered by cloud computing. The initial focus on this value was on the leverage that was possible from changing how the infrastructure delivered IT services. This infrastructure leverage was from the virtualization of resources that primarily drove lower capital requirements, and the increase in utilization of resources that drove the key cloud characteristic of elastic scaling. The current focus of cloud value is the leverage that is possible from driving efficiencies in the labor associated with IT delivery. In particular, automation of repetitive management tasks allows for administrator efficiencies that allows them to
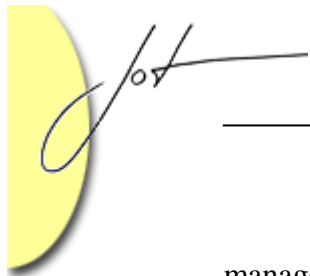
manage more services and/or focus on providing higher-value services. The focus on workload standardization allows for user and administrator efficiencies as this facilitates self-service access and management of services, and the ability to deliver services for use faster.

2   FOCUSING ON WORKLOADS

Workloads have become critical in the cloud journey. As we have discussed in earlier articles, workload analysis is an important step in determining what can run effectively in a cloud environment. It is important to understand the characteristics of the workloads in the context of a cloud delivery to determine its suitablity to be delivered as a cloud service. We want to avoid workloads where risk and migration cost may be too high, including those that are database intensive, have complex transaction processing, packaged application (e.g. ERP) workloads, and highly regulated workloads. We want to focus on workloads which can be standardized and take advantage of running as a cloud service, including web infrastructure applications, collaboration infrastructure, development and test workloads, and high performance computing workloads. Finally, we should consider new workloads that have been made possible by cloud including high-volume low cost analytics, collaborative business networks, and industry focused "smart" applications. The analysis focused on determining which workloads were appropriate and not appropriate to be delivered as cloud services based on experiences, as discussed below.

Through our current experiences, we find the following workloads as feasible for current public cloud capabilities:

- Development, test and pre-production systems.
- Mature packaged offerings, like e-mail and collaboration.
- Batch processing jobs with limited security requirements.
- Isolated workloads where latency between components in the application is not an issue.
- Infrastructure as a service, including compute resources and storage as a service.
- Managed services for the IT infrastructure, including backup and restore, virus scans, etc.

Of course, our experiences have also shown that we should (with the current maturity of cloud) avoid the following workloads:

- Workloads which depend on sensitive data normally restricted to the enterprise, e.g. employee information, or health care records.
- Workloads composed of multiple, co-dependent services, such as high throughput online transaction processing

- Workloads requiring a high level of auditability, accountability, e.g. those subject to Sarbanes-Oxley.

- Workloads based on 3rd party software which does not have a virtualization or cloud aware licensing strategy.

- Workloads requiring detailed chargeback or utilization measurement as required for capacity planning or departmental level billing.

- Workloads requiring extensive customization.

What has changed is the approach that we now take on analyzing workloads for their suitability for cloud delivery based on an analysis of the gain from moving the workload to be delivered as a cloud service along with the pain from moving the workload to cloud. This analysis, generically, is shown in Figure 3 below to show how a service provider can determine which workloads to focus on. Note the classification of the workloads into different types (e.g. "content-centric" or "database-centric" architectures) which provides additional information in determining the difficulty of moving these workload types into the cloud. Underlying this workload analysis are the characteristics to be considered, including standardization (which provides for higher gain from the cloud), and enterprise policies around security and data (which cloud lead to a higher pain from moving to the cloud.)
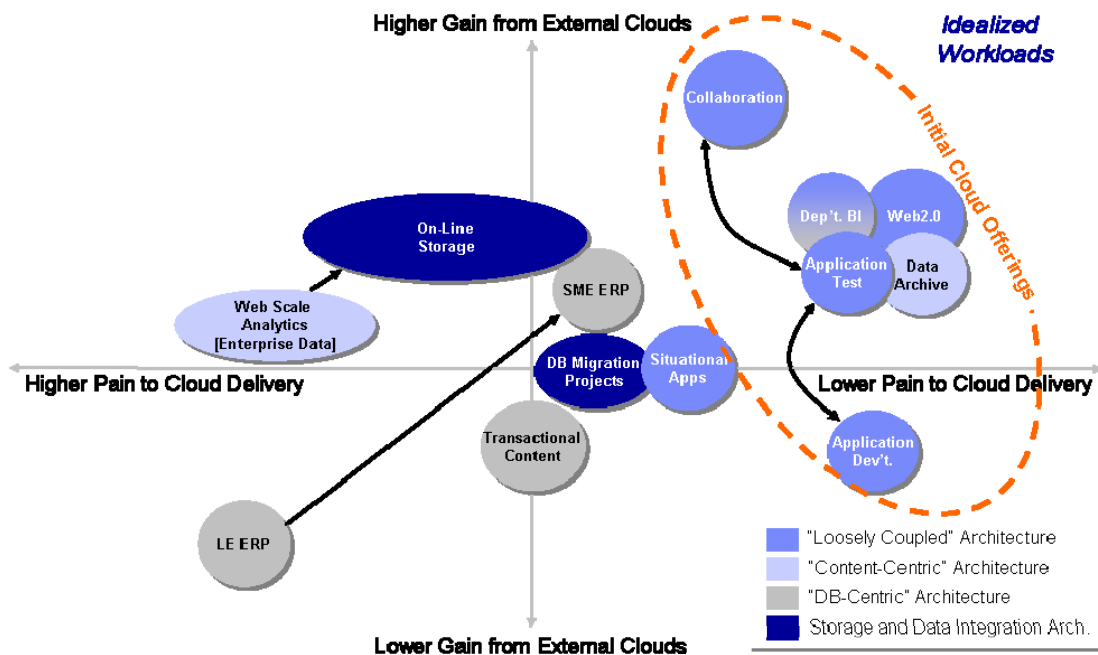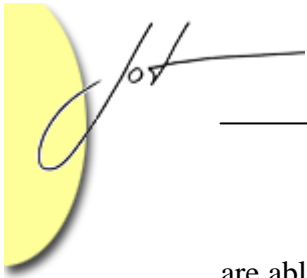

Figure 3: Workload Analysis

In addition to this workload analysis, the maturing of the cloud allows service providers to focus on delivering workload-specific cloud capabilities. These workload focused cloud capabilities can deliver fine tuned support for the workloads and ensure that they

are able to get the most value from cloud delivery. This support includes the ability to use these services in a hybrid mode.

As an example, consider the workload specific cloud offerings from IBM as shown in Figure 4 below. In essence, for every workload (e.g. development and test), IBM provides three different types of offerings – standardized workload services that are delivered on the public IBM cloud, private workload services that can be built and/or managed by IBM, and a "workload services in a box" appliance that offers a pre-built, pre-integrated system that is ready for use to deliver workload specific services. In order to take advantage of these workload specific capabilities, enterprises will need to understand their workload requirements and design a cloud solution to address the requirements that may involve one or more of the offerings. To showcase the use of all the offerings, consider the scenario where an enterprises':

- Developers can use the public developer services provided by IBM Smart Business Development and Test on the IBM Cloud

- Developers and testers can do unit/functional testing using IBM Cloudburst, which provides a "cloud in a box" environment targeting test workloads.

- Testers can do build, verification and production testing using private cloud services implemented and deployed (either in the enterprise datacenters or managed by IBM) using IBM Smart Business Test Cloud.
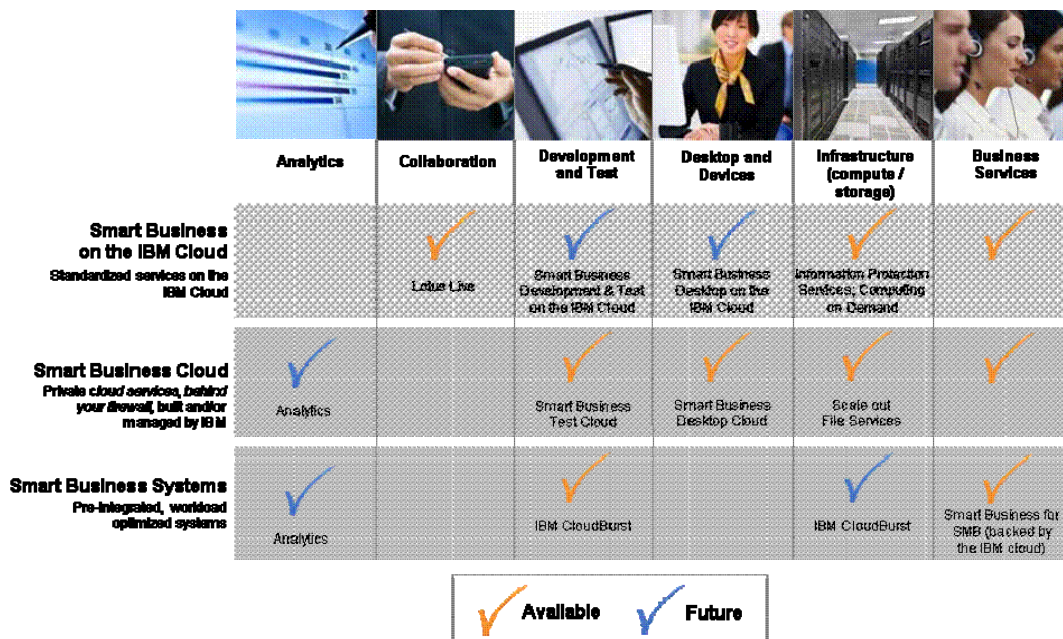

Figure 4: Workload Specific Cloud Capabilities

3 GETTING YOUR HEAD IN THE CLOUDS, AGAIN!

It is important to re-calibrate your cloud journey to keep pace with and take advantage of the capabilities that are being delivered by a maturing IT service delivery platform. Besides what we have discussed above, it is important also to keep abreast of the changes that are happening in cloud standards – as these will play a critical role in the inter-connected cloud. To gain better insight into the standards that are needed, consider the requirements for security and data protection in the cloud shown in Figure 5.
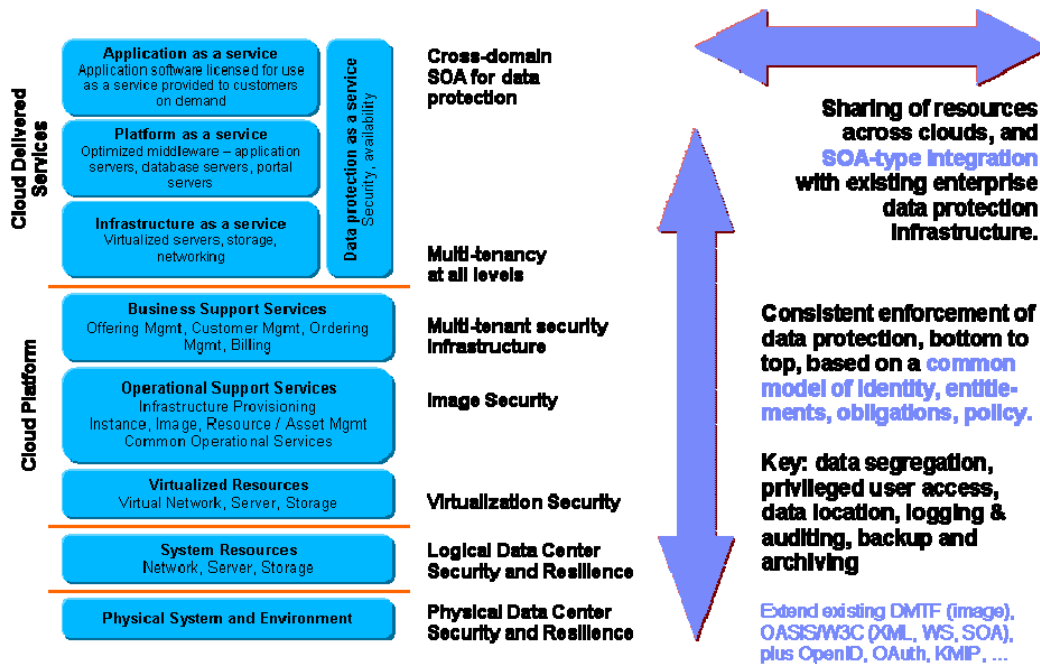
Figure 5: Security and Data Protection Concerns

Standards are needed not only to support the additional concerns that cloud brings to the IT service delivery platform stack – including security for virtualization, images, workloads, and multi-tenancy, but also to support the movement of these workloads across service providers. As we move workloads between service providers in the interconnected cloud, standards are needed to ensure that the elements that are needed for workloads to move between these providers are in place. Standards are needed to cover authentication and authorization, access, data segregation, service management, metering and billing, etc. To further exacerbate the problem of having enough standards, there are multiple standards bodies working on extending their existing standards to cover cloud as well as developing new standards to integrate their established standards with cloud. Recently, there has been the establishment of the open cloud manifesto with the intent of setting the core set of principles towards an open cloud – where established standards will allow different cloud service providers to interact and inter-connect effectively.
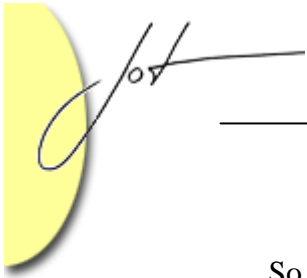
So, remember to keep your ears close to the cloud and ensure you are keeping track of the changes occuring as cloud matures and are able to leverage the emerging capabilities, standards, and value. In the meantime, I will continue our cloud journey in the next article by taking a deeper dive into the architectural model and the underlying capabilities you will need to harness in your cloud journey.

About the author

Mahesh Dodani is a software architect at IBM focusing on Cloud Computing. His primary interests are in enabling communities of practitioners to design and build solutions that address complex business needs and deliver value. He can be reached at dodani@us.ibm.com.