Leveraging Cloud in the Rapidly Evolving Workplace of the Future

Tom Chefalas, Jim Doran, Lorraine M. Herger, Andrzej Kochut, Steve Mastrianni, Charles Schulz, Dennis G Shea
IBM Corporation, IBM Research
T.J. Watson, Research Center, Yorktown Heights, NY, USA
{chefalas, jrdoran, herger, akochut, stevemas, dgshea, cschulz}@us.ibm.com

***Abstract*—Rapid advances in cloud computing have made it possible to replace individual working environments with centralized and consolidated workplaces. These workplaces bring many advantages over traditional self-managed environments, such as improved security and compliance, and rapid on-boarding. In this paper we describe the implementation of a desktop cloud at IBM Research.**

Keywords: cloud, desktop cloud, virtual desktop,

I. INTRODUCTION

In this paper we address how a team at IBM is leveraging the rapid advances in cloud computing, end user security and the proliferation of smart mobile devices, to support the rapidly evolving workplace of the future. We describe an approach and solution that securely extends the capabilities of desktop cloud computing to the appropriate segments of our enterprise user community. We will describe one of the user scenarios we have successfully demonstrated, monitored and deployed to the IBM Research community - the rapid on-boarding of our summer student population on the Research Compute Cloud (RC2) this past summer. This pilot demonstrated a significant transformation on how a temporary/transient population (student interns) can be on-boarded into an enterprise, with the end result of having their work environment and productivity tool set available and usable upon the first day of employment (most within an hour). This greatly streamlined the process with respect to the deployment, installation and configuration of the I/T environment, which can be replicated for larger groups going forward.

The structure of the paper is as follows: We first summarize the benefits of leveraging desktop virtualization to provision individual work environments - both the benefits to end user productivity and satisfaction, as well as the benefits to the organization in streamlining IT process and reducing the enterprise's cost to deploy and manage the workforce I/T environment. Next, we discuss how we leverage the desktop cloud to securely provide access to the end user's 'workplace' and data from a wide variety of devices and locations. Additionally, we outline the key challenges to provide this environment in a secure and efficient fashion. Next, we provide details on how we constructed such a cloud platform in the T.J. Watson Research center which we refer to as the Research Compute Cloud (RC2) and some statistics on usage, performance and user satisfaction. We also focus on one of the desktop cloud enterprise workloads that we are deploying on our RC2 [1] research platform for various segments of our IBM population. [IBM Research has built and deployed the Research Compute Cloud (RC2), an environment that is a living laboratory and also a production environment that offers cloud computing services to a broad, worldwide employee community]. Finally, we discuss some of the challenges of providing secure mobile access to our desktop cloud for a large enterprise population and conclude the paper with the future vision and potential applications of this technology and approach, as well as the business case analysis which makes this model attractive [2]. As the workforce of the enterprise has evolved over the past several years, we have come to realize that 'one size fits all', the common approach of workplace services over the last 10+ years, is no longer relevant or applicable to meeting the workforce needs of the next generation of the enterprise. We will discuss our experiences in detail in the closing section of our paper.

I/T operations have evolved over the years in large enterprise organizations. IT staff, who early on had to manage a few large data centers and a limited number of departmental machines, over time became challenged to manage at least one IT device per user, to currently must manage a constellation of smart devices per user - from phones, tablets and laptops, to high end workstations. The true costs of end-to-end lifecycle management of information technology, - purchase, deployment, maintenance, replacement, sunset, which had been previously been the purview of the centralized CIO shop, has now become more widely distributed throughout the enterprise or campus. Consequently, in today's enterprise desktop environment, non-hardware and software costs

978-1-4577-1591-4/11/$26.00 ©2011 IEEE

dominate the total cost of deploying and managing the end users workplace environment [3,4]. Maintenance tasks such as application deployments and updates, patch management, and security fixes, have all become a much more daunting challenge, given the number of end user devices the IT staff have to manage and the increasing costs in labor. Moreover, and perhaps less visible and easy to track, the utilization of hardware resources in the distributed desktop model is very low, in contrast to the increasing cost per device in areas such as desk side and call center support as the proliferation of device type increases. We see this trend as increasing over time – rapid technological changes and easy access to the 'latest and greatest' in the consumer markets is now a mindset that has invaded the enterprise, with the expectation by employees that the enterprise can support whatever is available in consumer markets. Managed environments, which abstract the enterprise environment from the physical device, thus allowing for more diversity in the end user platform have shown promise in accommodating the consumer model within the enterprise. From pilots and deployments that IBM has had with several school systems the total cost of ownership to deploy a managed end user workstation environment have seen significant cost savings, for example with the Pikes County School system, cost savings over a five year span were projected to be as great as 50% [5].

II. Vision and Strategy – A Commentary

The demographic of the IBM workforce has changed dramatically over the last decade. Ten years ago, most IBMer's worked in an office, within the United States, Western Europe or Japan, sat at a desk most days with a single fixed phone, and accomplished most of their work with a desktop system, collaborating with colleagues who occupied offices nearby. The profile of the IBM employee is quite different today. IBM now counts approximately 500k systems on its network on any given work day, with only about 5% of these being fixed desktops, and the remainder being laptops, smart phones or tablets. 50% of all employees work remotely, and rarely see the inside of an IBM facility. IBMer's are distributed across all time zones, and countries, with matrixed, worldwide teams as the 'new normal'.

The workforce itself has also dramatically changed over this last decade. Greater than 50% of employees have been with IBM less than five years. The high turnover rate in the employee population means that there must be a quick and efficient way to provision employees, and to quickly de-provision, as well. Workers now have an expectation of flexibility that previously did not exist – working any place, anytime, anywhere. Cell phones, smart phones and social media have become ubiquitous in society as well as in the enterprise, fostering the need for constant and unlimited connectivity and communication. The line between work life and personal life has not just blurred – it has disappeared; and therefore, so have the lines between enterprise work environments and personal computing environments. This melding of computing environments presents a challenge to the management of enterprise assets. How does the enterprise secure and protect its critical assets from 'leaking' outside its domain? How do enterprise owners ensure that SPI (sensitive personal information) is not compromised? How does the enterprise ensure that its customer's data is not inadvertently exposed onto the Internet? And, how does the enterprise protect itself from the latest virus, malware, bot, etc.? These are all very real and serious threats. Each day we read about the most recent security breach of a major corporation or government agency. These breaches are not only an embarrassment to the firm, but can also entail serious fines or generate lawsuits.

How can a corporation guard its assets, yet not overly constrain its employees with unnecessary, and often costly, policies, or processes? This is not a simple question to answer, since the answer is often not a 'one size fits all' solution. Within IBM we are developing a user segmentation approach for determining the workplace needs of the employee population. This approach requires analyzing the needs of employee segments, and characterizing employees' needs with a set of use cases, based on job role. The use cases are then analyzed to determine technology requirements and to develop technology deployment plans. Currently, from the known uses cases, IBM has defined thirteen IT personas and the associated IT requirements for these personas. Within these personas we have identified a set of users who have a need for what has been named 'safer platforms'. These platforms are virtual, or cloud based desktops that provide heightened protection for special job roles. These job roles have the following characteristics:

- Privileged access to client and/or IBM data
- System level access to critical systems
- Support of multiple client systems
- Require a persistent desktop and a customized environment.

Accommodating these users in a flexible manner – access to sensitive data, to client information, applications, etc., during one part of the day, and at lunch, or the end of the day, allowing for access to personal data, websites, bill paying, etc., -on the same device – is the goal. Currently, it is necessary in many cases to have multiple connections, or multiple devices. This model is neither sustainable nor cost effective. In effect, it nearly doubles the number of devices per person (for particular segments of the population), and with this doubling comes all of the costs of device management, software licenses, security,

compliance, etc. Clearly, we must move to a more streamlined, flexible and converged environment.

The desktop cloud can provide the user the environment to 'mix and match' these types of activities in a secure way, and to do it with a single device. The architecture and implementation allows for the separation of enterprise work from personal applications and reduces the risk of horizontal spread of malware across environments. It also improves employee productivity and morale, since employees can do their personal tasks in between work activities. Again, as the line between work-personal life has blurred, this accommodation is critical to develop a cultural environment that takes employee needs into consideration.

Another advantage of the virtualized environment is management of heterogeneous device types across the enterprise. In an environment of 500k users, it is nearly impossible to mandate, control, or test every user device. However, by virtualizing the enterprise client, the client becomes agnostic to the underlying platform and device. This strategy reduces the complexity of monitoring, management, and application compatibility. The centralized repository of the images also insures disaster recovery and business continuity.

Virtualized environments, due to the arrival of mature cloud technology, are now a real possibility for large segments of employee populations in the enterprise. In the past, due to technical issues such as network latency, management issues, server and storage costs, virtual machine immaturity, and software primitiveness, virtualization could not be considered a viable option. However, with the maturity of these foundational technologies, the changing workforce, and the increased focus of enterprises on security and compliance, virtualized environments have now, once again, become 'front and center' with CIO's in all industries.

III. Results From An Initial On-boarding of Our New Workforce – IBM Research Summer Interns

As we stated previously, the demographic of the IBM workplace has changed dramatically over the last decade. A good forecaster of what is to come is to see how our population of research summer interns at T.J. Watson Research Center, Yorktown Heights assessed the pros and cons of their access to the Research Desktop Cloud. Overall approximately 70% of the research interns were able to login in and get access to the IBM set of productivity tools during their initial one hour on-boarding session. This might have been expected by them, but in years past, students spent hours or days, installing and configuring the appropriate settings to have their workplace configured for their use. Given the resources and capability provisioned by the research desktop cloud, approximately 50% worked most of their day in that environment. This is significant since the interns also access this resource from a high end notebook. In the future, as their client becomes a light device such as a tablet or smartphone, with access to drive a larger screen, this is a good indicator that this model of workplace computing will evolve. The student employees also will be using their client device to access a wide variety of compute intensive resources provisioned from various cloud resources.

This summer rollout was the first time we had 100% of our summer interns use the desktop cloud to access their productivity tools such as corporate email, Bluepages (IBM's corporate directory) and Sametime (instant messaging). Some of the key pros in addition to rapid onboarding were: 1. preserve the state of the workplace and access from many devices and a diverse set of locations. 2. the ability to leverage a large amount of computing resources from a 'lightweight' device (and if the job required, access significant additional compute resources from the cloud). But just as there were pros, some challenges were encountered. A key reported con was network latency when working remotely. Working on-site in an IBM network environment (wired or wireless) was satisfactory, but some of the local hotels and other places where the interns connected remotely introduced a delay that caused usability concerns. Our Research colleagues from IBM Almaden performed further investigation and recommend enhancements that are currently being evaluated for incorporation with our next wave of users. Overall, given this was the first summer for full usage by our summer intern population, we were encouraged that this model of computing will evolve into a significant computational environment suitable to support the next generation of enterprise users. As the world user population moves from desktop or high-end note book computers to lighter weight mobile devices, the desktop cloud computing model will be ready to accommodate the every day user demands.

IV. Secure Workplace

The internet and the web have become ubiquitous in the workplace. They are essential components of a productive and efficient work environment. But they have also opened up many new security risks to the enterprise and have engendered new laws, standards, and requirements for protecting the security and privacy of proprietary and personal data. Add to this the trend toward more types of mobile and personally owned devices which are replacing or augmenting the traditional workstation, and you have a very challenging environment in which to manage exposures and risks.

The traditional methods of securing the enterprise network and all devices connected to that network are fast becoming debilitating. We can no longer expect employees to limit their mobile access to a single device type or even devices from a single manufacturer or running a common operating environment. This space is developing too fast and if an enterprise limits itself to a single device or manufacturer they risk being left behind if they are not on the right 'horse' at the moment. Also devices are very personal and possibly task specific. No device can be optimized for every possible use case. The only practical alternative is to manage the security of the information instead of managing the security of all of the devices.

Several important concepts come into play here. The first is to classify the data. The second is to assess the risk of the activity that is being requested based on the classification of the data. And the third is to manage where functions are performed so that inappropriate activities can be prevented.

The goal here is to keep data in the appropriate environment where its security can be adequately maintained. The cloud provides an important tool in this endeavor. By judiciously moving the handling of data from devices to the cloud we limit the exposures to loss or misuse of that data. The cloud represents a controlled environment where the enterprise can establish and maintain standards for access to, handling of, and communication of sensitive data. At the same time the cloud can enable a wide range of devices to access or interact with that data without actually transmitting the data to the device.

This can be achieved through several related but distinct approaches all based on the single concept of performing the handling of the data in the cloud and managing the interactions with and control of the application from the device. Virtual desktop, virtual applications, and streamed OS or applications are types of personal computing where the important data can be managed and secured in the Cloud.

Virtual Desktop is a very illustrative example. In this case all of the data and also the applications run exclusively in a virtual machine which resides in the datacenter or on the cloud. Using one of several remote presence protocols (RDP, ICA, and VNC are examples), the user can interact with the desktop from virtually anywhere and any device, assured that the data is never transferred out of the controlled environment. Even if the user's device were to be stolen no data is stored on the device so all of the data remains secure in the cloud. By carefully managing a cloud based desktop environment the Enterprise can easily enable a broad range of mobile or personal devices so that the productivity of the employees is maximized.

V. Stateless Virtual Desktops

Early generation virtual desktop deployments focused on providing a dedicated virtual machine to each end user. Each virtual desktop was a monolithic desktop PC that contained the operating system, applications and user personality. This approach to on-boarding legacy PC users is a good initial approach for transforming a physical desktop environment to a virtual desktop environment, but it does not overcome the challenges of managing the virtual desktop images and optimizing the resource utilization of the cloud. In order to overcome these challenges, a new model is appearing in which stateless virtual desktops are paired with 'OnDemand' Applications and User Personalities. The pairing of the 'OnDemand' user personality occurs at the point in time in which a user connects to a virtual desktop. As a result, the underlying virtual desktop and image are separated from the user personality. This improves the manageability of the desktop and image because the virtual machine instance can evolve independently of the user personality. For example, patch management can occur to one common shared stateless desktop image versus each dedicated virtual desktop. In addition, the user personality can be assigned to any available stateless desktop, which improves resource utilization.

The end user of a virtual desktop expects to retain the "personal" desktop model of the PC world. The monolithic virtual desktop model preserves this behavior because each user has his own separate desktop instance with his own set of applications, own configuration and own data. This monolithic model retains the "personal" attribute that PC end users are accustomed to having. Unfortunately, this approach is inefficient in managing storage and managing the core desktop image within a cloud. By introducing stateless virtual desktops with 'OnDemand' user personalities, we significantly reduce storage consumption by having a core image shared by multiple users and each user only consumes storage pertaining to his own personality. In addition to the 'OnDemand' coupling of the user personality with the core image, an 'OnDemand' coupling of an application can occur at the time of use. These 'OnDemand' methods for the user personality and applications transforms the monolithic virtual desktop into a dynamic assembly of a virtual desktop that fulfills the end user's requirement for his "own" desktop and the requirement to reduce storage consumption and improve manageability within the desktop

VI. CLOUD.CONNECTION BROKER

The functionality of the connection broker can be divided into two major groups: end user related and system administration related. The first group is focused on providing end users with secure, flexible and efficient access to desktop cloud. The latter is about optimizing and simplifying administration tasks to allow cost effective management of large numbers of virtual desktop instances.

In order to support its key functions connection broker exposes a secure portal which is a single point of access to virtual desktops and other assets, such as published applications. Users authenticate with the portal and then are presented with the list of assets they can use. Upon selecting one of the virtual desktops, the user is connected to it and can start working with the instance.

To achieve the above functionality a comprehensive data model is required. First, it represents customer organizational units that should be entitled to use the desktop cloud. Each organization can have multiple user groups which are intended to represent sets of users with similar business needs resulting in similar needs for the IT environment. Another critical piece of information in connection broker data model is the inventory of cloud images that are suitable to be used as virtual desktops. This catalog is a subset of all images that the underlying cloud image library contains. In addition, each registered image contains metadata specific to the desktop cloud, such as details of remoting protocol (e.g., RDP or VNC) to be used when connecting to it. The next important concept in the data model is the virtual desktop pool – a set of virtual desktop instances. The key characteristics of the pool are type (elastic or static), size, and image on which it is based. The elastic pool can expand or contract based on the number of connected users. This implies that the user state is separated from the image and therefore can be overlaid on top of the base virtual desktop instance. The pool size defines the maximum number of virtual desktops that can be provisioned within the pool. The image determines the content of the virtual instances in the pool - all virtual desktop instances are created based on this image therefore making them homogenous at the time of creation.

Another critical element of the data model is the entitlement and assignment system. User groups are entitled to specific desktop pools. The entitlement does not create actual assignment of a user to a desktop instance; it simply denotes that users in the group are allowed to own instances from a given pool. An entitled user can obtain a desktop assignment either via the administrator or using the “Self-service” feature. “Self-service” allows users to list a set of desktop pools they are entitled to and request assignment of one of the desktops in one of the pools. This function significantly reduces management costs of the system, reducing administrative workload.

One more key function of the connection broker is to perform workload dependent optimization which is specific to desktop cloud. For example, the decision about increasing or decreasing the size of the pool should be based both on the current size of the connected user population and the expected number of users in the near future. More advanced scheduling mechanisms can be used to further optimize the size of the pool and reduce the cost of the system.

Connection broker also gathers detailed usage information. It allows profiling user access patterns as well as applications used, leveraging the information to further optimize the system. For example, a detected pattern in which users connect to the system can be used to plan the optimal size of the pools at each period during the day or week. Knowledge of applications used together with usage durations and frequencies can be leveraged to optimize multiple aspects of desktop cloud, for example, license usage tracking, shaping the content of the future image releases, as well as decisions how to deliver each of the applications.

VII. THE MOBILE WORKFORCE: ENABLEMENT FROM THE CLOUD

Today's enterprise workforce is increasingly diverse and flexible in terms of how they use technology. Depending on the role, an employee might require access to the company's critical information while on the road. This poses significant challenges for the IT environment, from flexibility of access to secure enforcement of end point policies. A good example of the mobile workforce employee is the sales representatives who spends most of their time traveling and yet needs up to date access to customer information which can only be accessed via the enterprise network. The traditional approach to facilitating this type of access is to issue a laptop to the mobile workforce employee, where the data can be locally stored. This allows full access to the required data, but creates significant security exposure - all of the data stored on the local disk of the laptop can be potentially compromised if the laptop is lost or stolen. The physical dimensions of the device can also be cumbersome for a mobile employee. With the advent of cloud computing and widespread adoption of smart phones, desktop function can be provided in a more secure, efficient and lightweight manner. The desktop cloud allows end users to access their desktops remotely from either a traditional PC or from a mobile device. A key advantage is that the user's data and applications are no longer tied to a

particular device, operating environment, or location. Instead, users can connect from the most convenient device-type, given their physical location and network connectivity and still perform required business functions. Moreover, the data being accessed remains in a secure data center with only screen updates sent to the client. Therefore, even if the end user device is lost or stolen, the critical business data is not exposed. Another beneficial aspect of the desktop cloud is its ability to rapidly provision new virtual desktops. Instead of going thru a costly and time consuming process of provisioning traditional desktops, virtual desktops can be created in minutes due to compute cloud provisioning automation. Such a feature can be beneficial in an emergency response scenario. For instance, an insurance company, faced with the need to process claims after a large scale natural disaster, may need to rapidly hire additional employees and immediately deploy them in the field to assess damages. In such a case the company needs to provide controlled access to its data and applications to the newly hired users. Provisioning traditional laptops for them would be very costly and time consuming. Instead, the company can provide them with virtual desktops in the cloud almost immediately and at the same time tightly control data access, both the type of data and the time period over which to allow access. For instance, in an emergency, temporary employees may play many roles, and need access to varying data sets. Using a desktop cloud model, it is straightforward to manage and update access controls so employees only have access to the data they need to have for any particular task, and the access is revoked when no longer needed.

In order to meet all of the previously mentioned requirements, the research community has been working on both security and performance aspects of cloud computing. An example of such work for desktop clouds are [6] and [7] which focus on optimizing desktop cloud user experience as well as overall system efficiency. As mentioned in the section on our summer intern pilot, the development of our connection broker, and the stateless virtual desktops, we are integrating these new maturing technologies into our environment to assist the various employee segments to become more productive in a secure manner, regardless of device or location.

VIII. CLOUD COMPUTING ENABLEMENT: RC2

Cloud computing can be viewed as a platform hosting applications and services that is being driven by significant trends, such as: 1. the broad shift to new Internet-based business models and Web 2.0 applications; 2. the growth in connected mobile devices, real-time data streams, search operations, collaboration, social networking, and consumer-generated 'big data'; 3. increasing complexity, management, and energy costs which drive requirements for efficiencies in high asset utilization and datacenter consolidation.

Cloud computing provides computing as a utility. Just as electric companies provide electricity when and where needed, cloud computing vendors dynamically provision, configure, and de-provision information technology (IT) capability as needed, transparently and seamlessly. This allows IT consumers to focus on their specific problem and not on the computing resources it requires.

The acquisition and rollout of physical servers is capital intensive and can be time consuming as well. An important feature of RC2 is the ability to support provisioning initiated on a self service basis, allowing users to request and receive compute resources on-demand.

RC2 constitutes an environment in which we deal with many of the thorny issues presented to enterprise I/T computing. For example, how does one fund and recover cost associated with creating a shared utility like computing model? In RC2 we implemented a strategy to recover the total cost of ownership involved in creating, running, and supporting a cloud computing environment, which is important in order to sustain and grow the needs of a compute utility. It is also important to shape the behavior and expectations of those who consume the infrastructure resource. Chargeback objectives seek to capture and account for the total cost of ownership for RC2.

RC2 aims to provide IBM Research with a platform where exploratory technologies can be rapidly introduced and evaluated with minimal disruption. This requirement calls for a componentized, extensible cloud architecture.

The primary architecture for RC2 provisioning is made up of 3 primary types of server resources:

Gateway

The target of the gateway is to provide routing to the private network in the routable network option.

Compute nodes

The compute nodes are the physical machines, managed as hypervisors. The target of the compute nodes is provisioning VMs. Hypervisors are running different client-side management applications:

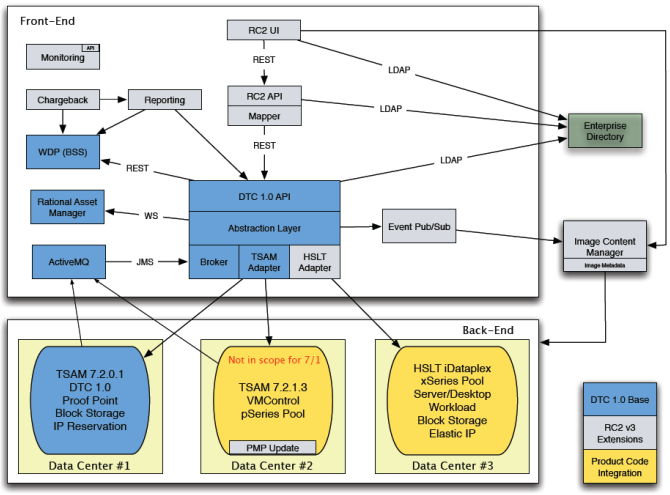
- e-Jabber
- zookeeper
- libvirt

Storage Nodes

The storage nodes are running as a physical machines with no virtualization. The target of the storage nodes of the architecture is to cover all storage needs – for VMs and storage on demand. In the current deployment the storage is provided by iSCSI that is very optimal from management point of view, because it shares the existing network infrastructure. From the management and control aspect the storage nodes are running e-Jabber and zookeeper clients.

The following attached references architectures depict both a modular management system, coupled with a horizontal scaleout scheme, which provide large cloud infrastructure growth as demands increase.

Figure 1: Virtual Desktop Architectural View



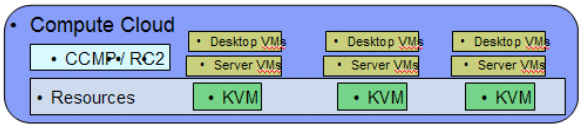
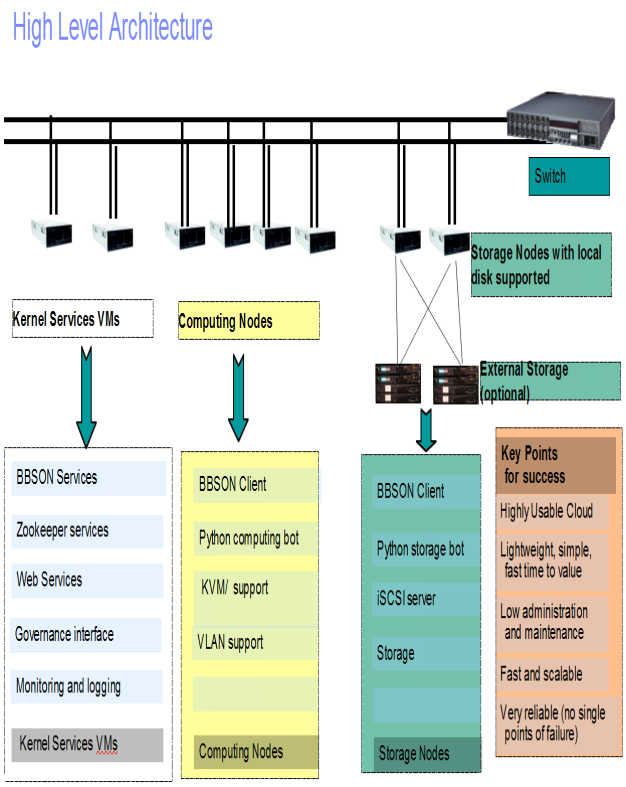


Figure 2: RC2 Cloud Resource Scale out



IX. Conclusion

Cloud computing and virtualization are approaching a maturity in the enterprise which makes them usable by many segments of the population – both the mobile workforce, and those who work in a fixed setting,- though not for the same reasons. The technology provides many attractive features in terms of management – image management, security, monitoring, provisioning, etc. – all can be simplified by the use of the cloud environment. Challenges still remain, of course, with areas such as latency, and offline access. User expectations must also be managed, so that there is no sense of 'take away' when an employee moves to a virtual environment. However, the many attractive features of the desktop cloud make it inevitable that it will remain an IT priority for most firms as they move to the workforce and workplace of the future.

References

[1] RC2-A Living Lab for Cloud Computing, Ryu, Zhang, et. al., Proceedings of LISA '10: 24th Large Installation System Administration Conference, Proceedings, pages 201-208.

[2] A Review of Cloud Business Models and Sustainability, Chang, Wills, De Roure, 2010 IEEE 3rd International Conference on Cloud Computing, Proceedings, pages 43-50.

[3] Desktop Total Cost of Ownership: 2011 Update, Gartner RAS Core Research Note G00208726, Federic Troni, Mark Margevicius, Michael Silver, Nov 16, 2010

[4] MarketScope for Hosted Virtual Desktop Services William Maurer, Lilian Dutra, Richard T. Matlus, Gartner #G00175055, June10th, 2010

[5] Desktops-As-A-Service Elongates The PC Refresh Cycle: A Pike County Schools Case Study, Natalie Lambert with Ben Echols and Robert Whiteley, Forrester Research, Inc. January 21, 2009

[6] Interactive Resource-Intensive Applications Made Easy, H. Andrés Lagar-cavilla, Niraj Tolia, Eyal De Lara , M. Satyanarayanan, In Proceedings of Middleware 2007

[7] Managing Responsiveness of Virtual Desktops using Passive Monitoring, Rajdeep Bhowmik, Andrzej Kochut, Kirk Beaty, IEEE/IFIP International Symposium on Integrated Network Management (IM), Long Island, New York, USA, June 1-5, 2009