SETLabs Briefings

VOL 7 NO 7
2009

BUSINESS INNOVATION through TECHNOLOGY

CLOUD COMPUTING

Infosys®

POWERED BY INTELLECT
DRIVEN BY VALUES

Cloud Computing: Pinnacle of IT Infrastructure Democratization

Cloud Computing, in brief referring to the emergent concept of flexible access to resources on a tap, with pay-as-you-go model, over the all-pervasive network, stirs an uncanny familiarity to the popular slogan of a leading airline out of India, *Simplifly,* whose model of low cost aviation is a classic case of disruptive innovation. The notion of *simplifly* lowered the bar of access to flying, thereby democratizing air travel. On a similar note, the cloud computing concept manifests itself in myriad forms like *simpliStore* or *simpliCompute,* technically also referred to *storage as a service* or *computation as a service,* or broader terms like *platform as a service* and *infrastructure as a service*.

The democratization effect of cloud is loud and clear from the multiple SMEs, including startups, which have been early cloud adopters. Likewise, even large enterprises have begun looking at the low hanging fruits to exploit cloud, especially in areas like infrastructure rationalization, where the cost advantages of cloud are immediately visible. Notwithstanding these adoption trends, a few questions need to be answered before we can begin to see the cloud as an inflexion point in IT.

Is cost the only driver for cloud? Is there a mainstream enterprise play for cloud? Is there an innovation angle to cloud? What are the showstoppers to mainstream adoption of cloud? What applications are better suited for cloud?

In this issue, we attempt to cover the breadth and depth of issues in an attempt to provide the answers. On the democratization front, while one article discusses internal deployment of utility model via private cloud yet another article elucidates how high performance computing, usually a costly proposition, can be made available to masses for massive number crunching and simulations.

Going beyond cost, we see interesting articles on how cloud can power innovative business models like service exchanges. We also discuss the power of cloud in offering innovative solutions for consumers needing ubiquitous access to infrastructure be it to store data for wireless sensor networks or ISVs needing ubiquitous storage and compute power. On the other side, potential new service opportunities leveraging cloud are identified like infrastructure management services for cloud.

A key contribution of this issue is in the several articles that delve deeper into the several showstoppers plaguing the cloud, requiring a leap of faith on part of enterprises to go from 'adopting low hanging fruits' to 'strategic adoption.' Key showstoppers of 'interoperability' and 'standardization' are explored in depth alongside usability and security.

Overall, we hope the issue helps in 'clearing the clouds' to cloud adoption and that the word cloud is no longer a homonym, i.e., it is cloudy to adopt cloud.

Happy reading!

Srinivas Padmanabhuni PhD
Guest Editor

SETLabs Briefings

VOL 7 NO 7
2009

Trend: Cloud Computing — Transforming the IT Ecosystem 3
By Rahul Bakhshi and Deepak John
Cloud computing has a wide array of things to offer. Each stakeholder in the cloud computing space has enough reasons to smile for the goodies it brings along. The authors opine that since it is here to stay enterprises should explore the possibilities and seek the right fitment with what cloud has to offer.

Discussion: Adopting Cloud Computing: Enterprise Private Clouds 11
By Shyam Kumar Doddavula and Amit Wasudeo Gawande
An agile pay-per-use business model can prove to be cost effective for small and large firms alike and can turn a new leaf in the way business and IT usage is defined, opine the authors.

Insight: Cloud Interoperability and Standardization 19
By AV Parameswaran and Asheesh Chaddha
The authors contend that while moving the existing applications to cloud or integrating data and application to the cloud platform, an eye on standardization will lend a congenial and effortless move to the new platform.

Model: SLA Aware 'on-boarding' of Applications on the Cloud 27
By Sumit Kumar Bose PhD, Nidhi Tiwari, Anjaneyulu Pasala PhD and Srinivas Padmanabhuni PhD
While adopting cloud computing, SLA aware on-boarding of application onto cloud platforms enables a streamlined exercise, transparent deadlines and better transition, state the authors.

Platform: Can We Plug Wireless Sensor Network to Cloud? 33
By Adi Mallikarjuna Reddy V, Siva Prasad Katru and Kumar Padmanabh PhD
Virtual communities, real time data share and analysis can add a different dimension to the existing IT scenario. The authors suggest a holistic approach towards cloud through Web 2.0 technologies to usher in a new age of business.

Third Angle:Cloud – Five Minute into the First Quarter 41
Amitabh Srivastava Senior VP, Windows Azure, Microsoft in an interview with Jitendra Pal Thethi, provides clarity on defining cloud computing and discusses the key traits and future prospects of exploring this new promising IT platform.

Viewpoint: Cloud Computing Identity Management 45
By Anu Gopalakrishnan
Beyond all the silver lining that cloud holds, the author draws our attention to identity management, a growing concern in the dynamic world of virtual space. A seamless and risk-free identity management is the basis for the evolving scenario of virtual workspace, opines the author.

Opinion: Service Exchange @ Cloud 55
By Bhavin Raichura and Ashutosh Agarwal
Business on cloud can mean low investment and high returns. Services exchange on cloud can prove to be a win-win situation for all stakeholders, at either end of service offering and consuming.

Perspective: Revenue and Customer Growth for ISVs using Cloud Offerings 61
By Ajit Mhaiskar and Bhavin Raichura
Virtualization is already in vogue and ISVs have to adapt themselves to suit the changing model to be able to attract and retain customers. The authors bring up some pertinent issues related to ISVs' revenue growth and the evolving cloud space.

Research: Power in the Clouds? 69
By Sudeep Mallick PhD and Ganesan Pandurangan
High performance computing (HPC) is required for faster processing time in complex and parallel processing applications scenario. The authors focus on how the architecture of cloud computing platform can be exploited fruitfully for HPC application execution.

Spotlight: Infrastructure Management and Monitoring in the Cloud 79
By Kaustubh Janmejay Vaidya
Cloud seems to be the most alluring IT innovation to every organization today. This paper offers well-laid roadmaps to help in planning and organizing cloud adoption.

The Last Word: Cloud Computing — A Disruptive Technology 89
By Srinivas Padmanabhuni PhD

Index 93

"Cloud is designed to be available everywhere, all the time. By using redundancy and geo-replication, cloud is so designed that services be available even during hardware failures including full data center failures."

Amitabh Srivastava
Senior Vice President
Microsoft Technologies

"System integrators with a close proximity to customers have to play a crucial role in taking the benefit of cloud computing to the enterprise customers."

Raghavan Subramanian
AVP & Head – Cloud Computing CoE
Infosys Technologies Limited

SETLabs Briefings
VOL 7 NO 7
2009

Cloud Computing — Transforming the IT Ecosystem

By Rahul Bakhshi and Deepak John

Cloud computing is here to stay and promises a fresh approach to the IT ecosystem

Cloud computing has emerged at an inflexion point in the industry and our lives, where IT is all prevalent and is no longer the panacea for all industry ills. Jumping to the front seat are buzzwords like TCO, business drivers, regulatory compliance, real-time data streams, SOA, mobility, Web 2.0, etc. A quick look around shows that any company worth its salt claims to be a cloud company or at least claims to have a cloud strategy in place. The acceptance of cloud computing as a mainstream technology is gaining momentum rapidly because of a strong alignment between cloud computing and the demands of an enterprise [1].

It is interesting to note that we have all been touched by cloud computing in some way or the other, irrespective of whether or not we are aware of it. Every time we access emails through applications like Gmail and Yahoo, view content on YouTube and Flikr, or post on Facebook, we are making use of cloud computing.

Cloud computing is here to stay and Gartner hype cycle identifies cloud computing as one of the key technology triggers of our times in the 2008 hype [2].

With cloud computing sweeping across the IT and business world, the economics of this emerging world will be very different. This paper looks at the prospects that cloud computing presents to all the stakeholders in the IT ecosystem during the transformation.

THE IT ECOSYSTEM

To gather the changes that will come along with cloud computing and to better understand the way the IT ecosystem is projected to evolve, we have segmented the IT ecosystem into horizontals and verticals.

The verticals define the domain/type of services/products offered:

- **Hardware:** Infrastructure, network, storage and computing solution providers, etc.
- **Software:** Independent software vendors, value added resellers, etc.
- **Service:** Communication, media and entertainment service providers, etc.

The horizontals describe the actors in each of these verticals.

- **Enablers:** OEMs, independent software vendors, etc.
- **Delivery Agents:** Value added resellers, communication, media and entertainment service providers, last mile access suppliers, etc.
- **Consumer:** Enterprise and retail users.

The complex relationship can be best represented if we consider that the consumer utilizes services and delivery agents act as intermediaries adding value to the raw capabilities offered by the enablers [Fig. 1].

It is important to realize that the roles for the actors are not limited and the entities may span across domains and roles. With renewed interest in decoupling technology from services and the falling price of bandwidth, the cloud model of operating will drive visible growth and collaboration horizontally, vertically as well as across quadrants.

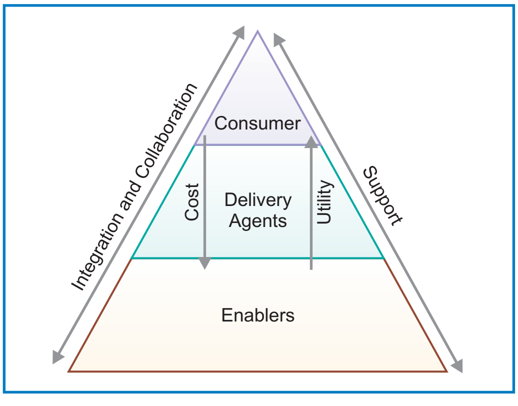

***Figure 1:** Relationships between Consumer and Enabler through Delivery Agents*
***Source:** Infosys Research*

CLOUD COMPUTING: THE RISE TO PROMINENCE

Cloud computing offers a new, better and economical way of delivering services and all the stakeholders will have to embrace the dramatic changes to exploit opportunities to avoid becoming irrelevant.

Following are some of the key trends/ changes that we expect to see:

Movement across Domains and Competencies: Technology providers are realising that in order to leverage economies of scale, it is essential to have competencies across hardware and software verticals. Nokia for example, operating in the handset and telecom infrastructure space, has boldly ventured into the social networking space with Ovi. The acquisition of US internetworking (an application service provider offering managed hosted applications like PeopleSoft and SAP) in 2006 has enabled AT&T to offer enterprise-class cloud services labelled Synaptic Hosting.

Acquisitions, although the preferred route, need not be the only route, as with all disruptive and emerging technologies, cloud computing will drive the creation of alliances spanning hardware, software and services, for instance, HP - Intel - Yahoo [3]. NetSuite, a leading vendor of on-demand enterprise services has announced its partnership with BT to deliver services via the SaaS model [4]. We also expect to see an increased clarity with standardization and interoperable open models like DTMF Incubator and Open Cloud Manifesto [5, 6].

Loss of Differentiation and Startup Power: The utility or value delivered by a product will triumph over the product itself. Hardware and software commoditization will give way to service commoditization. As George Crump from InformationWeek says "It's very hard to add a significant new capability to existing products [7]." As we move up the triangle in Figure 1, hardware and software enablers and delivery agents will collaborate and co-innovate to differentiate services.

John Foley describes start-ups as having the innate capacity to drive innovation and fill niches, while pushing down costs and driving up performance [8]. There will be a delicate shift in the balance of power from traditional enablers to start-ups that deliver ideas, the likes of 3Tera, Appirio, Coghead and Kaavo. This is reflected by the fact that venture capital interest in the 'cloud' is high. While VC funding has dried up in many areas in this recession, some 25 startups in the cloud space garnered more than $150 million in VC funding in the past year [9].

New Sales and Pricing Models: Delivery of cloud computing requires optimised infrastructure management costs and increased operational efficiencies. This will have considerable impact on the way the enablers realize revenues. The key trend to arise will be the acceptance of the subscription model (opex model) resulting in customers increasingly transforming 'my problem' into 'your problem.' With the subscription model, the enablers will require an upfront investment. However, the breakeven will be hazy depending on the volume of subscribers. The traditional delivery mechanisms (brick and mortar, media like CDs and DVDs) are on the decline and not transforming will mean sudden death, as is evident from Blockbuster's partnership with TiVo in an attempt to ward off Netflix [10].

Data/Content Driven Innovation: The application delivery platforms will become increasingly rich as they become more and more service focused [11]. For the mass, cloud computing is all about ubiquitous access to content. Thus, telecom and media companies are reinventing themselves to become on-demand solution providers, aiming to provide complete experience, as opposed to just providing services in silos. Citrix president Mark Templeton said, "Optimization of the user experience will happen in the data centre, at the edge of the network and in internet cloud, allowing IT to deliver any application to any user with the best performance, security and cost savings possible [12]." This is evident from the increased investments in server and storage consolidation. Cisco's 'medianet' suite, for one, has been built around advanced collaboration and entertainment, targeting both the business and home user.

Security and Legal Implications: It is important to realise that even if data and applications are stored and accessed remotely, the responsibility of the security and integrity of data lies with the individual. Authentication and authorization on the cloud, entangled with similar requirements offline will drive the need for interoperable (across services and devices) identity management. Single sign-on will be an area of investment. Ventures in this field include Microsoft Active Directory (within the enterprise) and OpenID. Also, issues around privacy and the way individuals exercise control over personal data stored remotely will need to be addressed.

Hardware and software services available through the cloud may span geographies and cloud providers may soon subcontract their services. All the three actors will have to understand the implications of having sensitive data on the cloud and regulatory compliances viz., SOX, HIPPA, etc., governing the same.

ROLE ENTITIES: CLOUD TRENDS

The following sections look at the consequences of the cloud for each of the roles defined in the previous section in greater detail.

The Enablers

Enablers provide resources that drive and support the creation of solutions in terms of both hardware and software that the consumer utilizes. Following are the buzz words in the enabler's arena:

Consolidation and Integration: With the markets changing rapidly, it is imperative for players to find new opportunities. Some of the recent acquisitions highlight the clear horizontal expansion across hardware and software towards services. For instance, with its purchase of Sun, Oracle has become a true cloud player with services now ranging from operating systems, programming/development platforms, ERP, CRM and other support utilities, giving Oracle an edge over its competitors and allowing it to offer the entire gamut of computing services required by any enterprise.

Examples of integration within the domain include Adobe acquiring Virtual Ubiquity - developer of online word processor; Google acquiring FeedBurner - leader in RSS services; and AT&T acquiring Ingenio - live search and commerce application provider, to name a few [13].

Ubiquity and Virtualization: The fact that the consumer would demand seamless access to content, impacts both the enablers as well as the delivery agents (providers in the software vertical, a little more than anyone else). The challenge being, developing applications that are 'portable' and offering seamless content delivery – whether on the office laptop or on the PDA while on the move or on high-definition TVs at home. This mandates higher investment in product development but does not necessarily allow a longer concept-to-market cycle.

To support the increased demand and adoption of cloud computing, the enablers are aligning their resources to provide multi-tenanted architectures, virtualization technologies along with support to highly scalable and elastic services. Virtualization technologies span platforms, resources and applications and the likes of VMware's Mobile virtualization platform are steps in that direction. In fact enterprises are already reaping benefits of this. Westar Aerospace & Defence Group has been successful in slashing their data centre size by 50% and power and cooling costs by 30% with a server virtualization solution from Dell [14].

Environmental Sustainability and Data Centres: Environmental awareness will further drive enterprises towards cloud computing as it allows considerable reduction in energy costs. Gartner estimates that over the next five years, most enterprise data centres will spend as much on energy (power and cooling) as they do on hardware infrastructure [15]. To quote VMware, "Gartner estimates that 1.2 million workloads run in VMware virtual machines, which represents an aggregate power savings of about 8.5 billion kWh—more electricity than is consumed annually in all of New England for heating, ventilation and cooling [16]." Cloud enabling technologies like virtualization and server consolidation can help enterprises reduce energy costs by as much as 80%.

Data centre consolidation will be driven by cost, space and energy savings. HP, for one, is replacing 85 data centres with just six located in America. According to IDC, America alone has more than 7000 data centres and predicts that the number of servers will grow to 15.8 million by 2010. In driving the cloud data centres, Linux complemented by open source solutions will be at the forefront. IDC expects Linux spending to boom by 21% in 2009 [17].

Cloud computing is also driving the usage of netbooks or laptops that are enhanced for mobility, compromised on computing capacity with a reduced storage capacity. Therefore, there will be an increased demand for transfer processing and storage in data centers. IDC reported that netbooks accounted for 30% of all laptop sales in Europe during the fourth quarter of 2008, with 3.6 million netbooks sold [18].

Marginalization of Fringe Players: Desktop based utilities and tools like MS Office and Norton antivirus will see a reduction in their installed user base and will ultimately be marginalized, as the same services will be available online. The traditional fringe players will have to re-invent themselves to align with the new modes of delivery, warranted by the cloud. Adobe is already providing an online version of its graphics editing program called Photoshop. Appistry is one of the more innovative companies and has recently launched the CloudIQ platform, offering enterprises the capability to port nearly any enterprise application to the cloud [19].

The Delivery Agents

Delivery agents are value added resellers of the capabilities offered by the enablers. Following are the key changes that we foresee in this domain:

Collaboration, Partner Driven Work Environments: Industry alliances are being forged and it is important for the delivery agents to weigh pros and cons before investing in the platforms. In the retail space Microsoft and Google can emerge as dominant players due to the inertia keeping consumers tied to its suite of products. Supporting them will be hardware players (a near monopoly of Intel) and virtualization providers like Citrix and VMware. The situation is complicated in the enterprise space, driven by leaders like Amazon, Oracle, IBM and Google. Cross platform compatibility and ease of migration demanded by the consumer will require the delivery agents to understand long term strategies.

Death of the System Integrators: System integrators, as we know them today, will have to take a second look at their model of operation. With the rising popularity of subscription based applications like Siebel On-Demand and SalesForce.com, the demand for customised on-premise will decrease, taking away with it the biggest market of the SIs. In the long term, IT services providers will have to increase efforts to provide end-to-end management of the IT estate (or whatever little would be left of it) or work along with the product companies to offer technical support to their customers. Once cloud computing technology reaches the critical mass, there will be an increased demand from enterprises to migrate data, applications and content to the cloud. In the short term, service providers need to ready their arsenal to deliver consulting services across technology and human resource domain.

Last Mile Connectivity: When push comes to shove, availability will triumph over utility. Internet service providers (ISPs) and last mile access supplier will have to ramp up their offerings rapidly to meet the increasing requirements of the bandwidth hungry content and applications, with fibre being the predominant technology for last mile access.

New Pricing and Delivery Models: Sales channels will also have to evolve to provide ubiquitous delivery models and the revenues are going to be long-tailed as the sales model will shift to a subscription based service, which will imply that customer retention and loyalty becomes all the more important. So all players will have to reinvent, be it the telecom operators who are shifting focus to value added services or the internet media houses that have to come up with variants of their web pages that can be accessed from mobile devices offering a consistent user experience, along with richer interactive applications to keep the customers hooked on.

Piracy: With the onset of the cloud, the users will no longer be required to download or install applications in the traditional sense. In the online world, controlled access implies that piracy will become increasingly difficult, if not impossible. Case in point being the online documentation services offered by Zoho, since there is no application that has to be installed at the users' end, there is no chance of having a pirated version of the application.

Likewise with online gaming, the problem of pirated copies of the games being spread around, resulting in millions of dollars worth of revenue loss can be curbed. OnLive is one of the pioneers in this field and has signed contracts with major video game content providers like Warner Brothers, Electronic Arts and Epic Games. What is interesting is that Nvidia, a provider of high end graphics processors and cards, primarily in the desktop segment, has welcomed the initiative of game content delivery through the cloud – a clear shift in perspectives [20].

The Consumers

Consumers are the demand side of the cloud equation and following are the trends for them:

Convergence, On-Demand: The retail customer will now, more than ever, come to expect on-demand everything - be it multimedia content, applications, gaming or storage. AMD's new campaign 'The Future is Fusion' is again reflective of the changing times. For the retail user, it is all about bringing together convergent multimedia solutions on any screen supported with advanced graphics capabilities; for the enterprise user it is delivering enhanced server and powerful virtualization capabilities [21].

Collaboration and Social Networking: Cloud based platforms like Facebook and Twitter will become destinations for collaboration, e-commerce and marketing. Enterprises are already planning to listen to the voice of the customer using such tools.

Collaboration and virtual workspace solutions will see increased investments. A key player in this space is WebEx, acquired by Cisco in 2007 for $3.2 billion – again an example of a hardware player moving to the software cloud domain. Another promising technology is IBM's Bluehouse, based on Lotus Notes. This enables employees among business partners or within the same organization to share documents and contacts, collaborate on joint project activities, host online meetings and build social-networking communities.

Back to Core Competencies: The cloud enables businesses to focus on their core competency and cloudsource the IT estate enabling the consumers to transfer risk. 'My problem' now becomes A look at an IDC study makes it clear that businesses want the cloud because of the cost benefit [22].

Decentralization of Management: The traditional view of management and governance of IT resources through standards and frameworks like ITIL, Sarbanes Oxley, HIPPA, etc., will change. As much as the technological impacts, the challenges for enterprises will also be to manage employee expectations working in a decentralised and distributed manner. Many legacy IT system integrations will break and enterprises need to clearly understand and estimate the risks of losing visibility and control over critical data.

CONCLUSION

Cloud computing promises different things to different players in the IT ecosystem. It offers a radical way of collaborating, delivering applications and content. More importantly it is here to stay. So it is easy to see why the enablers are paving the way for massive adoption of the cloud and why are the delivery agents leveraging their positions to catch the cloud demand. As for the enterprise, it reduces the TCO of the IT infrastructure while increasing agility.

It is important to realise that the complete shift to the cloud is not imminent, but enterprises will be better off with a long term vision for technology, people, information, legality and security to leverage capabilities offered by cloud computing. The delivery agents, more than any other players, need to reassess their role in enabling and delivering cloud computing to consumer for lack of innovation and not keeping pace with the growth will result in marginalization.

REFERENCES

1. Frank Gens, What User Want from IT: Speed, Relevance, Information and Innovation, IDC exchange, March 2008. Available at http://blogs.idc.com/ie/?p=141
2. Gartner Highlights 27 Technologies in the 2008 Hype Cycle for Emerging Technologies, Gartner, 2009
3. http://www.hp.com/hpinfo/newsroom/press/2008/080729xa.html
4. http://www.netsuite.com/portal/press/releases/nlpr04-22-08b.shtml
5. http://www.dmtf.org/about/cloud-incubator
6. http://blogs.zdnet.com/Hinchcliffe/?p=303
7. http://www.informationweek.com/news/software/hosted/showArticle.jhtml?articleID=210602537
8. http://www.informationweek.com/news/software/hosted/showArticle.jhtml?articleID=210602537
9. http://www.ciozone.com/index.php/Cloud-Computing/Unraveling-the-Cloud-Ecosystem.html
10. http://www.informationweek.com/cloud-computing/article/showArticle.jhtml?articleID=216300432
11. http://community.zdnet.co.uk/blog/0,1000000567,10008269o-2000458459b,00.htm
12. http://community.zdnet.co.uk/blog/0,1000000567,10008269o-2000458459b,00.htm
13. http://startup.partnerup.com/2008/01/02/2007-acquisitions-web-internet-technology/
14. http://whitepapers.techrepublic.com.com/abstract.aspx?docid=360865
15. Rakesh Kumar, Eight Critical Forces

Shape Enterprise Data Center Strategies. Available on www.gartner.com

16. http://www.vmware.com/solutions/consolidation/green/
17. http://www.economist.com/business/displayStory.cfm?story_id=11413148
18. Netbook Trends and Solid-State Technology Forecast, Consumer Behavior Report. Available at https://mr.pricegrabber.com/Netbook_Trends_and_SolidState_Technology_January_2009_CBR.pdf
19. http://gigaom.com/2009/03/09/appistry-opens-the-cloud-to-almost-all-apps/
20. http://www.edge-online.com/news/nvidia-onlive-a-net-positive-us
21. http://news.softpedia.com/news/AMD-Launches-039-The-Future-is-Fusion-039-Campaign-93711.shtml.
22. Frank Gens, IT Cloud Services User Survey, pt.3: What Users Want from Cloud Services Providers, October 2008. Available at http://blogs.idc.com/ie/?p=213

SETLabs Briefings
VOL 7 NO 7
2009

Adopting Cloud Computing: Enterprise Private Clouds

By Shyam Kumar Doddavula and Amit Wasudeo Gawande

Cost efficiency riding on the agility of cloud computing appeals enterprises the most

Cloud computing delivers IT capabilities as services-on-demand. This scalable and elastic model provides advantages like faster time-to-market, no capex and pay-per-use business model. While there are several such benefits, there are challenges in adopting public clouds because of dependency on infrastructure that is not completely controlled internally and rather shared with outsiders. Several enterprises, especially large ones that have already invested in their own infrastructure over the years are looking at setting up private clouds within their organizational boundaries to reap the benefits of cloud computing technologies leveraging such investments. This paper describes the different options available, highlighting the key advantages and challenges posed by each and the approach enterprises should be taking in adopting cloud computing with minimal risk.

WHY CLOUD COMPUTING?

Traditional infrastructure provisioning model is inefficient and does not meet the requirements of the internet era [Fig. 1]. In this system centric model, once the need for a business application is identified, its infrastructure needs are identified and a request for infrastructure is placed with the IT infrastructure team that procures and provisions the infrastructure. The application is then developed, tested and deployed on that infrastructure.

Some of the challenges with this model include –

- **Need for Large Capex:** Large investments need to be made in procuring the infrastructure for a business application. This increases the barrier for innovation as it is hard to experiment with a business idea without large investments.

- **Poor Utilization of Resources:** Application usage is not going to be constant yet the infrastructure is provisioned for peak demand, to be able to guarantee application SLAs. So, the

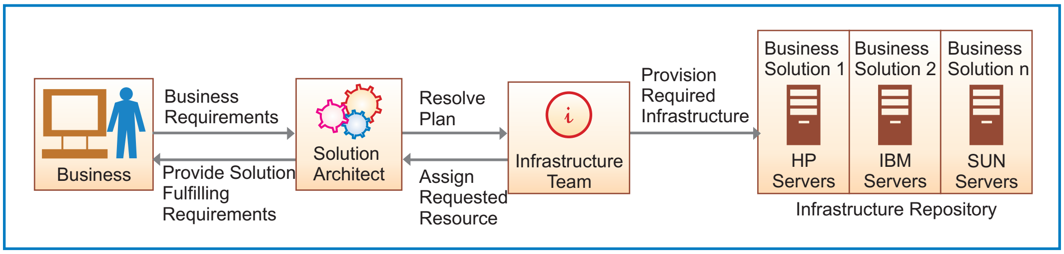

***Figure 1:** Infrastructure Provisioning: Traditional Model* ***Source:** Infosys Research*

infrastructure remains under-utilized for a major part of the time.

- **Slow Time-to-Market:** This model of procuring and provisioning infrastructure usually requires significant time and reduces the agility of an organization in creating new business solutions.

Figure 2 below provides an overview of the service centric provisioning model with cloud computing.

In the cloud computing model, IT-related capabilities are made available as services that can be provisioned on demand. There are several offerings from various vendors that enable provisioning different IT components as services, components ranging from infrastructure to platforms and applications. This is commonly referred as infrastructure-as-a-service, platform-as-a-service and software-as-a-service.

This cloud computing model offers several appealing benefits for enterprises including –

- **Faster Time-to-Market:** Enterprises can avoid the step of initial infrastructure procurement and setup, thus allowing the business solutions to be taken to market faster.

- **On-Demand Elastic Infrastructure:** Sudden spikes due to business growth, functionality additions or promotional offers can be addressed easily with

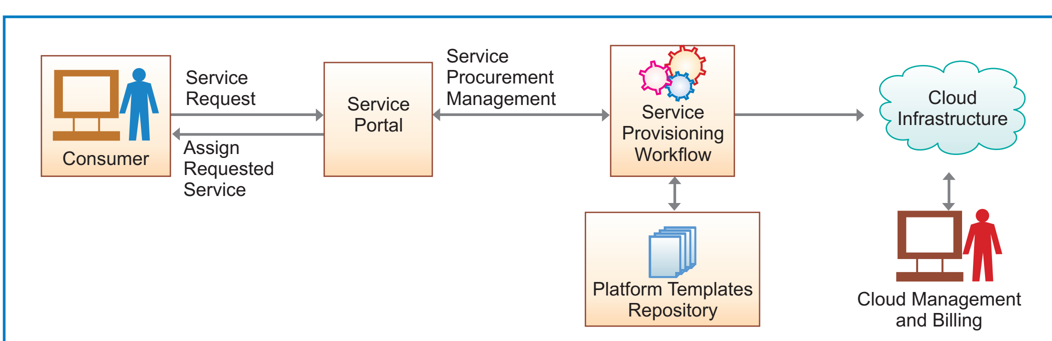

***Figure 2:** Infrastructure Provisioning: Cloud Computing Model* ***Source:** Infosys Research*

infrastructure that can be requested on demand.

- **Pay-as-Use:** Organizations can leverage the pay-as-use model of cloud computing to ensure optimum utilization of available resources.

PUBLIC CLOUDS AND CHALLENGES

Public clouds like Amazon AWS, Microsoft Azure, Google AppEngine offer infrastructure and platforms as services over the internet. In public clouds, resources and costs are shared by users who use them over the internet on pay per use model.

This model appeals especially to startups and small organizations that have not invested in hardware resources and are looking for ways to avoid the large capex involved in procuring infrastructure upfront. Even though there are several benefits like cost savings, faster time to market, etc., from this model, there are a few challenges listed below that are preventing wide scale adoption of public clouds.

- **Security:** The biggest roadblock is the potential security issues due to multi-tenant nature of public clouds. There are security and privacy concerns with sharing same physical hardware with unknown parties that need to addressed.

- **Reliability and Performance:** Performance and availability of the applications are important criteria defining the success of an enterprise's business. However, the fact that organizations lose control over IT environment and important success metrics like performance and reliability, and are dependent on factors outside the control of the IT organizations makes it dangerous for some mission critical applications.

- **Vendor Lock-in:** Cloud computing services offered by different vendors are not governed by any standards as of today. Depending on the vendor, the applications have to undergo changes to adapt to the service.

- **Leveraging Existing Investment:** Most large organizations that have already invested in their own data centers would see a need to leverage those investments as an important criterion in adopting cloud computing.

- **Corporate Governance and Auditing:** Performing governance and auditing activities with the corporate data abstracted in the public cloud poses challenges, that are yet to be addressed.

- **Maturity of the Solutions:** Some of the PaaS offering like AppEngine offer limited capabilities like only a subset of JDO API.

ENTERPRISE PRIVATE CLOUDS

In order to overcome these challenges, organizations are looking at enterprise private cloud offerings. Enterprise private cloud solutions help organizations leverage the existing IT environment and create a cloud computing platform in the private internal network. This model overcomes several challenges faced in public cloud adoption. Enterprise private clouds are seen as a natural progression of initiatives like virtualization already taken up by several organizations. Enterprise private cloud solutions add capabilities like self-service, automation and charge back over the virtualized infrastructure.

Figure 3 provides the recommended logical architecture for an enterprise private cloud.

Self Service

The private cloud solution should have a self service portal that enables users request infrastructure and platforms as a service. It should contain a service catalog that lists the categories and the services available, the associated SLAs and costs.

The service portal should enable reserving as well as requesting the services on demand.

Automation

The private cloud solution should have certain traits -

- A provisioning engine that automates the provisioning of the infrastructure
- Workflow driven with built-in approval mechanisms enabling governance
- Enable user management and integration with enterprise authentication and authorization mechanisms
- Enable enforcing enterprise policies on resource allocation through a rules engine
- Enable capturing the common deployment patterns using templates.

Self-service and automation helps reduce the time-to-market so that users can request

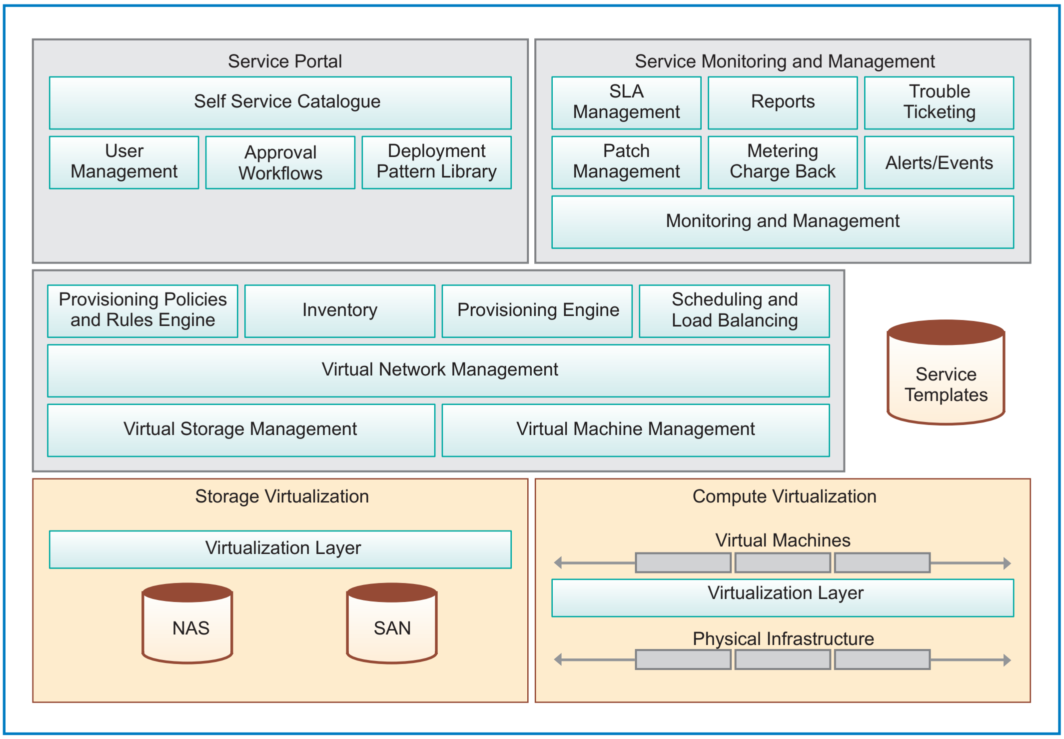

Figure 3: *Enterprise Private Cloud Architecture* ***Source:*** *Infosys Research*

for infrastructure as a service and can get it provisioned on demand.

Management and Monitoring

The private cloud solution should also have an integrated monitoring and management platform that should have the following components —

Monitoring and Management: Track various metrics at the software and infrastructure level

Metering & Chargeback: Track the usage of the various services and allow to charge back mechanisms to be plugged in

SLA Management: Enable, define and monitor SLAs for the services

Patch Management: Enable patches to be rolled out to the various software components used

Reports: Generate reports on usage, SLA adherence, etc.

Incident Management: Generate alerts when there are issues and provide ticketing mechanism to track and resolve incidents.

Virtualization

The private cloud solution should have virtualization layer that virtualize the key infrastructure components including compute, storage and network.

ENTERPRISE CLOUD: HYBRID APPROACH

Private clouds help overcome some of the challenges associated with public clouds but they are not as cost effective as public clouds since the traditional model of owning, i.e., buying and managing the infrastructure, still holds true. Not just the infrastructure, even the internal clouds are to be built and managed by the IT team. Moreover, as the underlying infrastructure is limited, it is likely to be less scalable as compared to the immensely robust and scalable infrastructure of cloud providers. The model also does not benefit from the lower upfront capital costs and less management overheads that are otherwise possible with public clouds.

So, the recommended approach is to adopt a hybrid one where both public and private clouds are used for different categories of applications. With this approach, organizations can reap the benefits of both public and private cloud models. This approach allows enterprises to adopt the public clouds partially, deploying only those services that are suitable for public clouds. The private cloud helps apply the cloud computing model internally as well. Thus the hybrid approach brings together the best in both worlds of public and private clouds.

As technology matures, there will be better options for creating such an enterprise cloud. There are already solutions available that provide abstractions over infrastructure available internally through virtualization software like vmware ESXi, Xen, HyperV and public clouds like AWS. Also there are VPN solutions available that can help create a secure network spanning infrastructure across enterprise data centers and public clouds. There are still challenges to be addressed like latency, automated routing and load balancing, end-to-end SLA management, etc., before such solutions become enterprise ready.

TYPICAL USE CASES FOR PRIVATE CLOUDS

Scenarios where there are only intermittent usages of infrastructure are ideal for cloud computing. Also, scenarios that involve sensitive data and processes, or mission critical applications are better suited for enterprise private clouds. Some of the typical use cases where enterprise private clouds can be leveraged include -

Development and Test Platforms as Services: There are studies that indicate that around 30% of the infrastructure at large enterprise is used for development and testing. These resources are not always utilized as development and testing are activities that happen occasionally. These resources can be provisioned through an enterprise private cloud so that the resources can be shared and utilized better and also the time to provision can be reduced.

Public Cloud Emulation Environments: Private clouds can be used to emulate a public cloud environment and can be used as a development and test platform while developing the applications to be deployed on the particular public cloud. The design, architecture and the actual code can be validated using the private cloud environment. Further, the same environment can also be used to test the developed applications for functionality as well as validations before it moves to the production at a public cloud. An example of this is usage of the open source Eucalyptus framework to emulate some of the Amazon AWS functionality. It can be used to create the development and test environment that emulates AWS EC2 and S3 environments.

Virtual Appliances: Private clouds can be used to create virtual appliances that leverage commodity hardware to create specialized devices like load-balancers, storage devices, etc. A machine image is created with not only the operating system but also the software stack, thus enabling creation of virtual appliances that can be provisioned on-demand.

Cloud Burst: With a computing stack that provides abstraction over the underlying cloud infrastructure and enables applications and data to reside together on both private and public clouds, when there is sudden spike in usage and the in-house private cloud environment is not able to support the requests, additional infrastructure can be provisioned from a public cloud without affecting the service quality.

High Performance Grid: Enterprise private clouds can also be used to create grid environments so that the infrastructure that would otherwise have got dedicated only for specialized grid applications can be utilized better.

TYPICAL USE CASES FOR PUBLIC CLOUDS

Some of the initial services that can be moved into public cloud are those that are not business or mission critical or do not deal with the sensitive data. Some of the typical use cases where public clouds can be leveraged include.

BPOs: Business productivity online (BPO) suite applications are one of the first applications of public clouds in enterprises. There are several vendor offerings like exchange online and Google Apps that offer messaging and collaboration software as services on subscription model that can reduce the overheads associated with maintaining such application on premise.

Data Backup and Archival: Cloud storage is cheaper and offer storage on demand. So, enterprises are looking at public cloud storage solutions for their data back up and archival needs.

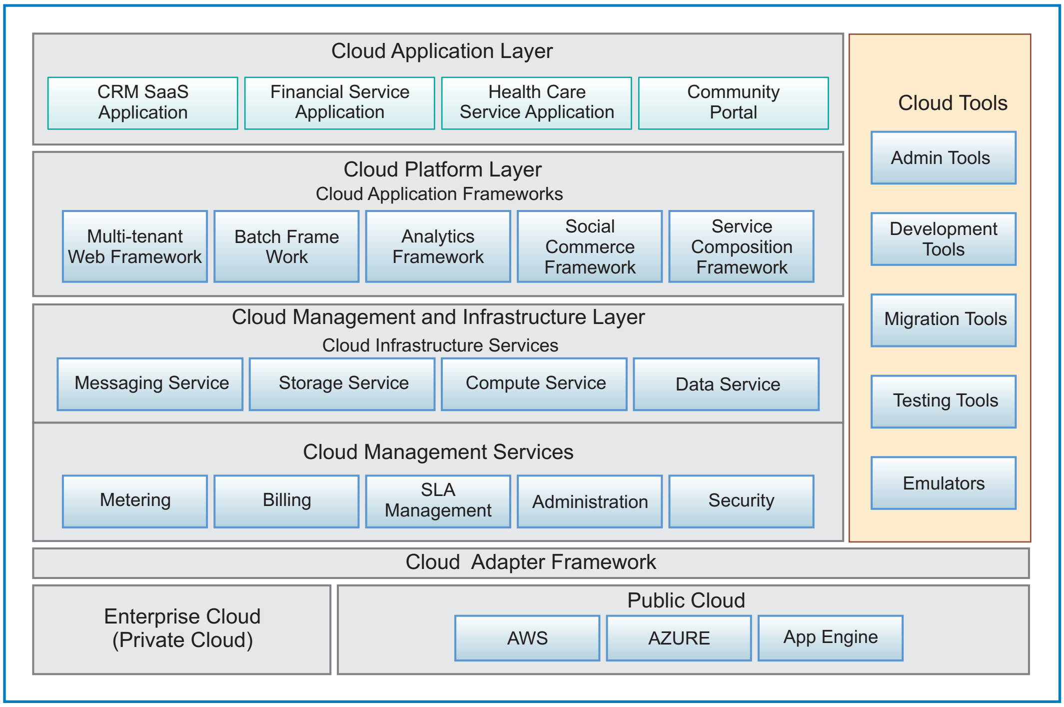

Figure 4: *Enterprise Cloud Computing Stack* ***Source:*** *Infosys Research*

Internet Content Management: Content that needs to be accessible from the internet like product literature, etc., can be stored in public cloud storage solutions. Amazon Offers S3 for storage and cloud front for CDN that increases the efficiency of delivering such content.

Organizations should build a cloud computing stack that helps them adopt this hybrid approach efficiently. The recommended cloud computing stack is shown in Figure 4 with the various layers and the various components needed for managing the cloud, developing and deploying enterprise applications and maintaining the applications using the cloud computing environment.

The cloud computing stack consists of the following layers -

Cloud Infrastructure Layer

The cloud infrastructure layer provides the core middleware capabilities like compute, storage, data stores, messaging, etc., as on-demand services. These use the infrastructure from public and private clouds and provide abstractions for the platform and application services.

Cloud Platform Layer

The cloud platform layer provides the specialized frameworks like a multi-tenant web framework for developing web based applications, analytics and batch frameworks based on MapReduce algorithms, cloud based social commerce framework, etc.

Cloud Application Layer

The cloud application layer consists of SaaS applications developed using the cloud platform services.

CONCLUSION

In the current economic climate where the expectations of efficiencies and cost savings are growing from IT organizations, enterprise private clouds provide a good opportunity to get started with cloud computing and reap the associated benefits of agility, cost savings and on-demand services while meeting the stringent enterprise security, performance and reliability requirements.

REFERENCES

1. James Staten, Deliver Cloud Benefits Inside Your Walls, Forrester, April 2009. Available at http://www.forrester.com/Research/Document/Excerpt/0,7211,54035,00.html
2. John Foley, 20 Cloud Computing Startups You Should Know, Information Week, September 2008. Available at http://www.informationweek.com/news/software/hosted/showArticle.jhtml?articleID=210602537
3. Private Cloud Computing for Enterprises: Meet the Demands of High Utilization and Rapid Change. Available at http://www.cisco.com/en/US/solutions/collateral/ns340/ns517/ns224/ns836/ns976/white_paper_c11-543729.html
4. Daniel Nurmi, Rich Wolski, Chris Grzegorczyk Graziano Obertelli, Sunil Soman, Lamia Youseff and Dmitrii Zagorodnov, The Eucalyptus Open-source Cloud-computing System, University of California, Santa Barbara. Available at http://open.eucalyptus.com/documents/nurmi_et_al-eucalyptus_open_source_cloud_computing_system-cca_2008.pdf.

SETLabs Briefings
VOL 7 NO 7
2009

Cloud Interoperability and Standardization

By A V Parameswaran and Asheesh Chaddha

Adoption of cloud rests largely on interoperabilty and standardization as they define the new age IT industry

Cloud computing can be defined as accessing third party software and services on web and paying as per usage. It facilitates scalability and virtualized resources over internet as a service providing cost effective and scalable solution to customers. Cloud computing has evolved as a disruptive technology and picked up speed in 2008 and 2009 with the presence of many vendors in cloud computing space.

With the presence of numerous vendors, the need is emerging for interoperability between clouds so that a complex and developed business application on clouds is interoperable. In this paper we provide cloud computing standards and interoperability view, examine some high level approaches for interoperability and look at important interoperability factors.

NEED FOR INTEROPERABILITY

Every new cloud service provider have their own way on how a user or cloud application interacts with their cloud leading to *cloud API propagation [1]*. This kills the cloud ecosystem by limiting cloud choice because of vendor lock-in, portability, ability to use the cloud services provided by multiple vendors including the ability to use an organization's own existing data center resources seamlessly. Business applications and data remain in cloud silos. There is a need for complex developed business applications on the clouds to be interoperable. Cloud adoption will be hampered if there is not a good way of integrating data and applications across clouds.

CLOUD COMPUTING STANDARDS AND INTEROPERABILITY VIEW

To start with, we provide a cloud computing standards and interoperability view to show some aspects/areas of interoperability and standardization in the cloud computing landscape [Fig. 1 overleaf]. When we look across the broad range of things that people consider in cloud computing, potentially hundreds of standards will be involved. The good news is that many of these standards

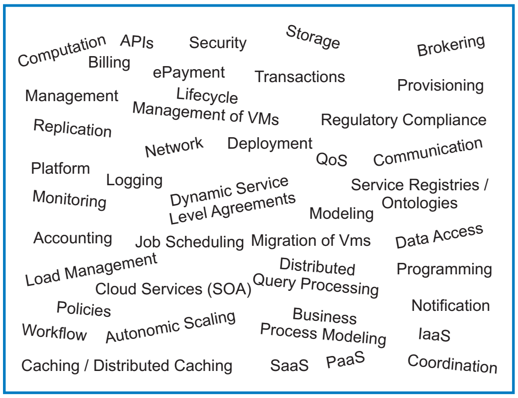

Figure 1: *Cloud Computing Interoperability View*
Source: *Infosys Research*

probably already exist. Reuse of good standards should therefore be a primary strategy. Creating a big picture view of the cloud computing landscape is therefore necessary to understand the *what, where* and *why* of standards. Once the overall view is understood, a gap analysis on the standards can then be done to see what standards we have and the standards we need.

INTEROPERABILITY APPROACHES

We discuss some of the emerging approaches for interoperability at a high level.

Approach 1: Unified Cloud Interface/Cloud Broker

Cloud computing vendors have formed a common platform — cloud computing interoperability forum (CCIF) — to address the problem of cloud interoperability and standardization [2]. The purpose of CCIF is to discuss and come up with a common cloud computing interface. CCIF is planning to come up with a *unified cloud interface* (a.k.a. cloud broker) whose features are as follows:

- Unified cloud computing is trying to unify various cloud APIs and abstract it behind an open and standardized cloud interface. Thus a key driver of the unified cloud interface (UCI) is to create an API about other APIs.
- It is a singular abstraction/programmatic point of contact that encompasses the entire infrastructure stack as well as emerging cloud centric technologies through a unified interface.
- The purpose of cloud broker is to serve as a common interface for the interaction between remote platforms, networks, systems, applications, services, identity and data.
- Having a common set of cloud definitions is an important factor that would enable vendors to exchange management information between distant cloud providers.
- The important parts of unified cloud interface (UCI) or cloud broker are a specification and a schema. The actual model descriptions are provided by the schema and the details for integration with other management models are defined by the specification.
- The unified cloud model will address both the platforms as service offerings as well as infrastructure cloud platforms. It will enable a hybrid cloud computing environment that is decentralized, extensible and secure.

Figure 2 shows a bird's eye view about the vision of the UCI project of CCIF [3, 4]. The primary goal is to come up with an abstraction layer that is agnostic to any cloud API, platform or infrastructure. The architecture comprises of layers and components with a use case described at the UCI project requirement page [5]. The architecture abstracts the usage of any cloud API and unifies them in one layer. This is done with the help of semantic web and OWL which has a pool of resources semantically understood and described. This enables the user to use these resources irrespective of whether these resources are being allocated from provider Amazon EC2 or Enomaly platform, etc. Having a unified interface with common definitions of these resources helps to do operations like allocation, de-allocation, provisioning of virtual machines or managing them through the UCI layer using the agent component. Assuming that the interface to UCI is provided to the user via a web browser or UCI cloud client, the UCI should provide a kind of a dashboard that shows the state of all allocated resources and running VMs. A component on the left side of Figure 2 is used to depict this. [6].

Approach 2: Enterprise Cloud Orchestration Platform /Orchestration layer

According to IDC, in virtually every industry, thousands of companies are trying to simplify the speed and adoption of their products and services by transforming them into cloud services. We see that the race to the cloud is accelerating [7]. The scenario that is unfolding is that there will not be just one cloud but numerous types -- private clouds and public ones. These will further get divided into general-purpose and specialized ones. Similar to the way that internet is a network of networks, InterCloud means a federation of all kinds of clouds. All these clouds will be full of applications and services. It will not be possible to use these without some type of orchestration.

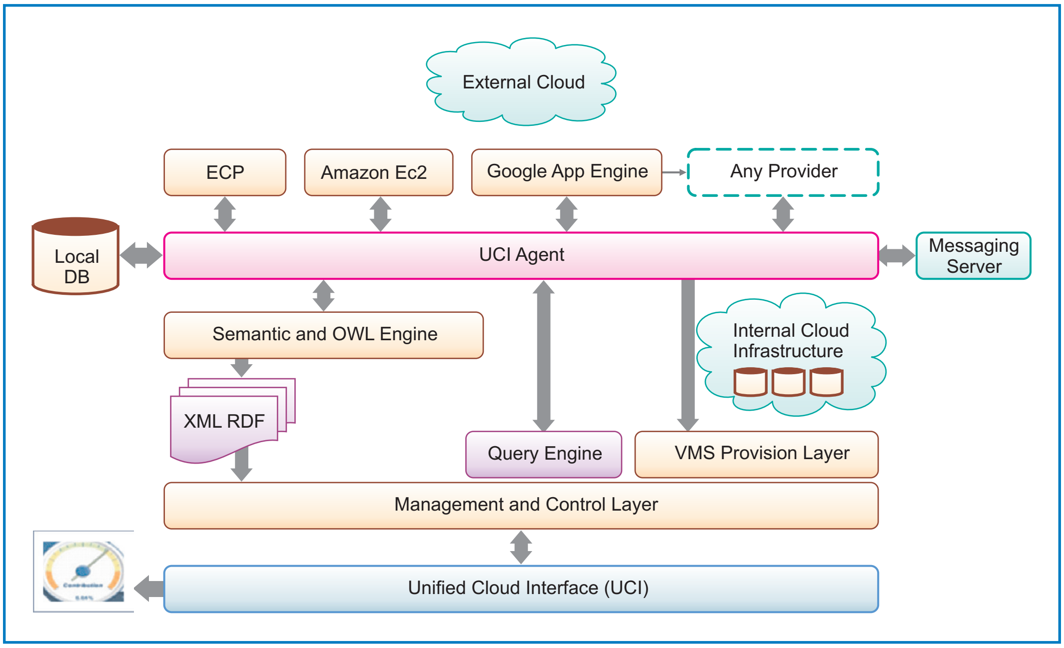

Figure 2: *UCI Architecture* ***Source:*** *www.code.google.com [8]*

The initiatives of some of the early adopters towards Cloud Orchestration are discussed below.

- Vendors like Cordys advocate the need for a layer in the cloud that provides assembly and orchestration for enterprises, which helps to deliver useful business advantages [9, 10]. Cordys delivers an enterprise cloud orchestration platform that helps enterprises to quickly adopt new ways of running their business and reaching their customers.

- Rightscale is another vendor that provides an orchestration layer/ cloud management platform. A single management platform is provided to conveniently manage multiple clouds that facilitates businesses to migrate deployments [11]. It helps businesses to manage and scale cloud deployments as well as facilitate application migration and management across multiple clouds. Similarly organizations like Suntec are looking at building an orchestration layer for billing infrastructure.

- Eli Lilly, a pharmaceuticals company uses Amazon web services and other cloud services to provide high-performance computing to hundreds of its scientists based on need. In future, it foresees the possibility of using cloud services from many different vendors and wants to avoid a scenario where Eli Lilly has to configure and manage each of those separately [12]. Eli Lilly describes the need for an intermediate orchestration layer that is in-between Eli Lilly and the various cloud services it subscribes to. This layer should be provided by another vendor and not Eli Lilly itself and should comprise of various algorithms that determine the best cloud service for a particular job based on factors like highest performance, lowest cost or other requirement. This approach will help Eli Lilly and other users to write to a single API rather than many and help to optimize service usage. Eli Lilly also sees the potential of using cloud computing for external collaboration. It is already doing some of this, but foresees that going forward, the cloud will become a point of integration between Eli Lilly and outside researchers. They have work going on at present that starts to fit into this collaborative scheme. This gives an example of how standardization needs are driven both by vendors as well as end users.

- CSC has recently announced cloud orchestration services for cloud services integration. This provides clients with features like service level management, data transparency, remote monitoring, auditing and reporting [13]. These services also provide automated arrangement, management, federation, coordination, security and operation of public, private and hybrid cloud computing environments, supporting industry-specific compliance, etc.

Figure 3 illustrates how a client can consume the services offered by more than one cloud service provider (CSP) via an orchestration layer.

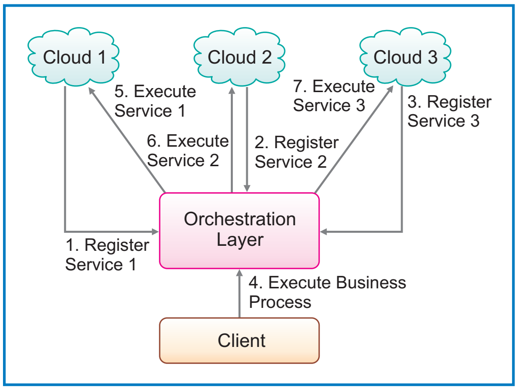

Figure 3: *Cloud Orchestration*
Source: *Infosys Research*

The features of the approach are explained below.

- Different cloud service providers can register the cloud services that they offer with the *orchestration layer*. This is similar to vendors who offer web services publishing their web services with the Universal Description, Discovery and Integration (UDDI). The orchestration layer can then dynamically select and bind to services based on criteria/ algorithms that determine the best cloud service for a particular job based on factors like highest performance, lowest cost or other requirement as specified by the client.

- Note that since the orchestration layer interacts with the cloud services offered by different vendors via different APIs, it can use user-computer interface (UCI) for interacting with different CSPs or have similar functionality built-in to be able to understand and interact with different CSPs via different APIs.

- Note that the client uses only one single API offered by the orchestration layer and thus is insulated from the different APIs offered by different CSPs.

- Figure 3 shows an example of how a client request for executing a business process (or workflow) is satisfied by the orchestration layer by invoking a sequence of three different services provided by three different CSPs.

The challenges with such an approach are discussed below.

- **Service Level Management:** Since the orchestration layer provides functionality to dynamically select and bind to services based on criteria/ algorithms that determine the best cloud service for a particular job based on highest performance, lowest cost or other requirement as specified by the client, such an approach will involve performance overhead due to runtime binding delays.

  The orchestration layer also needs to interpret client API calls and translate them suitably to invoke services provided by different CSPs. This will involve latency as well.

- **Data Volumes:** Depending on the provided service, the data volumes required to be transported across cloud services is another important factor to be considered. For certain types of services, this could be a limiting factor due to the overhead involved.

- **Platform Support:** Depending on the service, the platform support required by the service could also be a limiting factor.

- **Others:** Apart from the above, there could be other challenges like security, regulatory compliance, data transparency, etc.

IMPORTANT INTEROPERABILITY FACTORS

This section discusses the emerging scenario and other important interoperability factors from different viewpoints.

We see that there are multiple initiatives by stakeholders from industry, academia and users. This does help the problem or parts of the problem being addressed by multiple standard bodies/forums/consortiums in parallel and also provide diverse view points. But it is important for the standard bodies, vendors and users to sit together, discuss and arrive at a consensus on the standards and APIs in different areas and share information. This is all the more essential due to the duplication and overlaps among the various groups involved. The flip side of the story is that this could lead to the possibility of several standards emerging and possible lack of consensus. It is important for the standard bodies/forums/consortiums to have balanced representation of interests in order to avoid bias towards certain stakeholders' agenda.

Though initiatives like OGF's OCCI are trying to come up with standards in a quick timeframe, it takes time for standards to mature and for reference implementations to become available. Till then the users will use APIs/platforms from cloud computing vendors, whichever they feel is most suitable for their requirements. When standards emerge and these vendors want to use the services of other vendors, then they will need to use brokers/adapters for interoperability. New users however will be able to natively use the standard API. There will also be vendors developing orchestration layers to build business processes/workflows using the cloud services provided by different vendors. With some of the major vendors like Microsoft and Amazon rejecting the CCIF agenda and pursuing their own interoperability agenda, this makes standardization and consensus more difficult and could lead to multiple standards. This could lead to a scenario in the long run where multiple standards co-exist and customers using brokers/adapters for interoperability for using services from multiple cloud service providers.

It is also important to look at standards required from the perspective of different industry verticals. For example, HIPAA compliance could be important for healthcare services, SOX compliance could be important for financial services, etc. This requires active participation from different vendors and users from these verticals in standard bodies. It will also be good if different vertical specific groups are setup in order to focus and discuss the vertical specific requirements and come up with standards that are vertical specific.

Another challenge is that since there are many models of cloud computing (SaaS, PaaS, IaaS), standards are required for particular models and not just one set. There is a need to prioritize and concentrate on core set of standards to start with and then expand to other areas. It is important to note that over specification inhibits innovation. Patents and intellectual property could be a hurdle for standardization process. Unlike Sun's open cloud platform APIs, it will be interesting to see if other vendors give their cloud APIs and protocols to the community.

When applications are migrated from one cloud to another, apart from functionality, it is also important to ensure that non-functional requirements (NFRs) are satisfied as well in the new migrated environment. This requires standards for defining and exchanging meta information regarding the application between the cloud service providers to check for compliance of NFRs before actual migration of the application via VM migration. The scenario could be complex considering the fact that there could be several NFRs pertaining to security, availability, reliability, performance, scalability, etc., that requires compliance.

CONCLUSION

Interoperability and standardization have huge impact on the cloud adoption and usage and thus the industry is witnessing high amount of energy and thrust towards these from different stakeholders viz., users, vendors and standard bodies. Standardization will increase and accelerate the adoption of cloud computing as users will have a wider range of choices in cloud without vendor lock-in, portability and ability to use the cloud services provided by multiple vendors. This will also include the ability to use an organization's own existing data center resources seamlessly. Standardization further promises to help towards complexly developed business applications on the cloud to be interoperable and ensure data and application integration across clouds. It also provides business opportunities to users to choose and use services provided by many different cloud vendors based on various criteria. On the other hand it helps vendors to provide additional higher level services like orchestration, apart from normal cloud services that are needed by the users. Standardization will thus pave the way towards realizing the true potential/benefits of cloud computing.

REFERENCES

1. Cloud API Propagation and the Race to Zero (Cloud Interoperability), January 2009. Available at http://www.elasticvapor.com/2009/01/cloud-API-propagation-and-race-to-zero.html
2. The Cloud Computing Interoperability Forum. Available at http://www.cloudforum.org/
3. Unified Cloud Interface Project (UCI). Available at http://groups.google.com/group/unifiedcloud?hl=en
4. UCI Architecture. Available at http://code.google.com/p/unifiedcloud/wiki/UCI_Architecture
5. Unified Cloud Interface (UCI) Requirements. Available at http://code.google.com/p/unifiedcloud/wiki/UCI_Requirements
6. CCIF's Unified Cloud Interface Project. Available at http://code.google.com/p/unifiedcloud/
7. Forecast for Computing: Cloudy, iStockAnalyst, December 2008. Available at http://www.istockanalyst.com/article/viewiStockNews/articleid/2904589#
8. Unified Cloud. Available at http://code.google.com/p/unifiedcloud/wiki/UCI_Architecture
9. CORDYS, The Intelligent Cloud Platform. Available at http://partners.cordys.com/cordysportalpartners_com/cloud_solutions.php
10. CORDYS - Enterprise Cloud Orchestration. Available at http://www.cordys.com/cordyscms_com/enterprise_cloud_orchestration.php

11. Cloudonomics: Article RightScale Adds Amazon EC2 Europe to List of Supported Clouds. Available at http://apache.sys-con.com/node/841086
12. Eli Lilly On What's Next in Cloud Computing. Available at http://www.informationweek.com/cloud-computing/blog/archives/2009/01/whats_next_in_t.html;jsessionid=35MVLYFIRJL2GQSNDLRSKHSCJUNN2JVN?catid=cloud-computing
13. CSC Announces New Family Of Cloud Services. Available at http://www.csc.com/banking/press_releases/27609-csc_announces_new_family_of_cloud_services.

SETLabs Briefings
VOL 7 NO 7
2009

SLA Aware 'on-boarding' of Applications on the Cloud

By Sumit Kumar Bose PhD, Nidhi Tiwari, Anjaneyulu Pasala PhD and Srinivas Padmanabhuni PhD

Performance being the prime concern in the adoption of cloud, SLA aware 'on-boarding' of application can be of great help

Cloud computing is fast emerging as the next generation service delivery platform. Recent advancements in commodity server and virtualization technologies are key enablers for the interest in these platforms [1]. Cloud computing platforms hold promise for both service providers and service consumers. For service providers it is a way to minimize capacity redundancy and improve server utilization through multiplexing system resources amongst multiple customers. To service consumers, the platforms help realize the ultimate dream of capacity-on demand and pay-as-you-go concepts. To scale IT infrastructure vis-à-vis the demand for business growth is known as capacity-on-demand. Further, the consumers are not required to invest in expensive IT resources upfront as they are required to pay only for the amount of system resources they consume, known as pay-as-you-go. These are the motivating factors for the recent interest in cloud computing as a service platform.

However, performance is one of the key concerns in the possible adoption of cloud. Typically the key performance measures are average response time and throughput. These measures are a part of the service level agreements (SLA) that are legally binding agreements between service providers and consumers. There is a need, therefore, to understand the impact on an application's SLA due to its co-location with multiple other applications on the same physical host and the effect of the overheads introduced by the virtualization technologies. It is important to understand the extent to which the existing performance models can prove to be useful in addressing these issues arising out of the adoption of cloud technologies [2]. It is also required to comprehend the drawbacks of the existing models to overcome the limitations introduced by the current utility computing paradigms.

MOTIVATION FOR SLA AWARE 'ON-BOARDING'

Virtualization is the core technology behind popularity of cloud computing platforms.

Though virtualization techniques provide security and isolation guarantees, virtualization overheads and interference effects adversely affect the QoS parameters such as response time and throughput agreed upon in SLAs of applications co-hosted on the same physical box [3, 4]. However, not much research has been done to identify and understand the impact of the virtualization overheads and interference effects on these QoS parameters.

To benefit from cloud computing, enterprises are also migrating their applications from existing dedicated on-premise hosts to private/public cloud computing platforms. This migration activity is known as on-boarding. Currently, this activity is a very specialized process executed by the SMEs. This specialized process helps in identifying the system requirements of an application, based on workload experienced by the application and the client's QoS. The understanding of the system requirements helps to frame appropriate policies specific to the application and enter into service level agreements with clients. This in turn helps the service provider to manage the entire utility data-center autonomically (i.e., autonomic data-centers) without manual intervention. In this process, there is no comprehensive understanding of the system requirements of the application without precisely understanding how assured QoS of one application is affected by the co-location of another application on the same host. Service providers not only face the risk of over-provisioning during low demands but they also run the risk of under-provisioning during peak loads. Also, if the interference effects are overlooked, they face the risk of overpromising on the QoS promised in the SLA.

Further, the service providers often classify their customers into different classes such as premium, gold and silver. This classification is based on the amount of business generated from the respective customers. This often means that high net worth customers are classified into premium category. This necessitates that the service providers guarantee higher quality of service to the customers belonging to premium segment. Additionally, the service providers must have an understanding of the resource consumption pattern of different types of requests generated from such premium customers. For example, browsing interactions may not be as resource demanding as the payment interaction. These factors further increase the complexity in fulfilling the SLAs.

To make the above on-boarding activity more effective and efficient, it is important to design algorithms that can translate the application's QoS and SLA requirements to system level specifications. Further, there is a need to investigate new set of mathematical models that can accurately predict response times and throughputs even when they are co-located with other applications on the same physical box. These models should also address the virtualization overheads and consider the interference effects. We make an attempt to define a framework to address these challenges.

PERFORMANCE ENGINEERING MODELS

As shown in Figure 1, the existing performance engineering models are categorized into four classes. These are:

Single Host Operating System Models: These models deal with issues related to allocation of computing resources to multiple competing applications executing on the same server. Typically, the CPU (if the server has only one CPU) is apportioned amongst these applications on a time sharing basis.

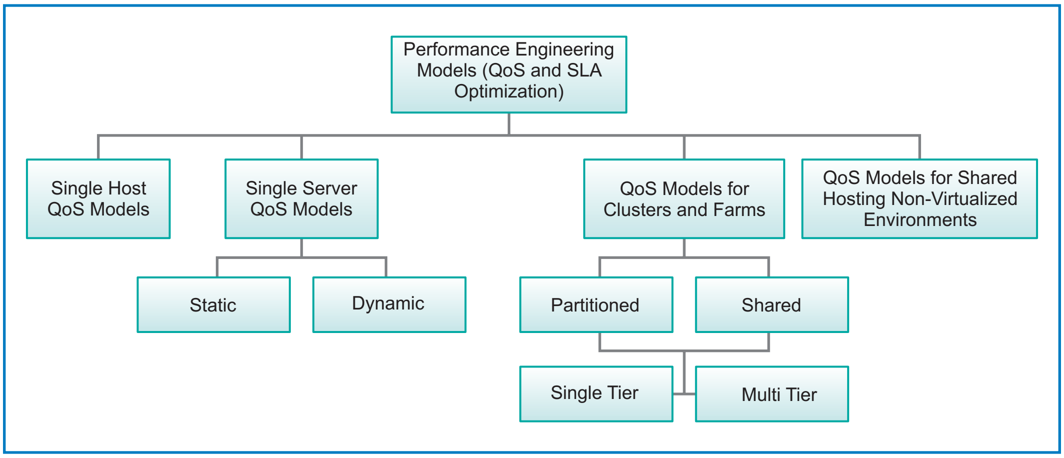

Figure 1: *Taxonomy of the Existing Performance Engineering Models*

Source: *Infosys Research*

Single Server Performance Models: These models deal with QoS/SLA issues related to servicing multiple client requests for a web-application, for instance, an e-commerce application hosted on a single server. In general, an overwhelming majority of models deal with issues related to web-servers.

Performance Models for Clusters and Farms: These models deal with resource allocation and QoS/SLA issues for scenarios where a particular tier of an application is replicated across multiple physical machines. For instance, the architecture of an e-commerce application typically consists of three tiers:

- the front-end tier for handling static web requests composed of simple HTTP (HTTPS) requests;
- the application tier for handling complex dynamic requests involving execution of java servlets, scripts and classes; and
- the database tier for handling database access requests involving lookup for non-cached data.

Performance Models for Non-virtualized Shared Hosting Environments: These models deal with resource allocation and QoS/SLA issues for scenarios where multiple applications run on single host that is a non-virtualized system.

The above performance models attempt to address questions related to capacity planning and load balancing. The models help in understanding the trade-offs of different architectural choices and aid in identifying potential bottlenecks that may degrade system performance. These models also provide performance estimates by predicting key performance metrics such as response time and throughput. However, the models assume that sufficient amount of computational resources, as needed to service requests, are available at all times. These premises do not hold true when an application is hosted on cloud platform. The very premise of a cloud platform is to make capacity available to applications on demand. The performance may degrade in times when sufficient computing resources are not made available to an application whenever the workload on the application increases. The increase/decrease in computing resource allocations to an application should be proportional to the increase/decrease in workload experienced by the application. This in essence, requires an intricate understanding of the computational resource requirements of the different components and of the various tiers of a typical three-tier application at different workloads. It is interesting to note at this point that the workload and the resource requirements are not just functions of the number of requests but also of the nature/type of requests. It is therefore pertinent to additionally gain a fine grained understanding of the resource consumption patterns of different types and classes of requests. SLA aware on-boarding of applications should take into account the above mentioned factors. Automating the SLA aware on-boarding of applications is a two step process that involves:

1. Translation of high level service level objectives into system level thresholds called SLA decomposition [5, 6]

2. Prediction of response time and throughput at different workload mixes, accounting the virtualization overheads and interference effects.

The modeling of the SLA decomposition requires capturing the relationship between the high level performance goals mentioned in the SLAs and the system goals for each application component as shown in Figure 2. The approach involves building component profiles at different workloads and for different user and request category. This requires subjecting the application to synthetic workloads for different categories. The component profiles are then suitably adjusted to reflect the overheads of the virtualization technologies being used in the cloud platforms. The resources allocated to different components are varied and detailed performance characteristics for each component are collected. The profiling technique is repeated for each category. Statistical techniques are then used to derive analytical relationship between performance metrics of a component as a function of resource allocations (CPU, memory, Network I/O, etc.). The statistical equations are suitably modified to account for the virtualization overheads depending on the type of technology used and a random variable denoting the interference effect.

Once the relationship between the resource requirements and the workload is established, it is important to predict the response time and throughput of an application in the presence of other applications on the same host. Consider an application A that is co-located with other applications B and

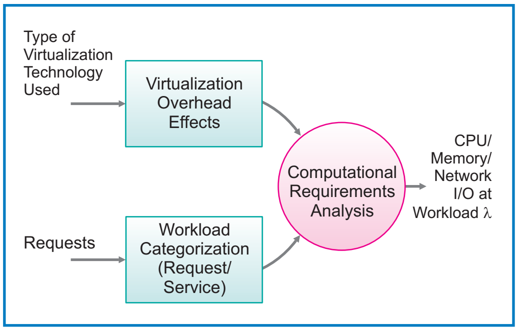

Figure 2: *SLA Decomposition Technique*
Source: *Infosys Research*

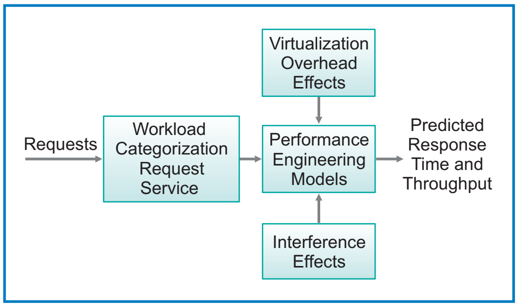

Figure 3: *Performance Prediction in the Presence of Interference Effects and Virtualization Overheads*
Source: *Infosys Research*

C on the same physical host. The resource requirements and the response time of requests and throughput of application A will be impacted by the resource consumption pattern of applications B and C co-located with it. The performance engineering model should be tweaked to be able to capture this interaction. Overview of the performance prediction in the presence of virtualization overheads and the interference effects is shown in Figure 3.

The proposed approach to SLA aware on-boarding of application onto cloud platforms has the following main steps:

1. Identifying different user and request categories of an application. It is possible to use white-box strategies where the source code is available. Black-box strategies can be employed for situations where no source code is available.

2. Subjecting the application to synthetic workloads of different categories and measuring the resource utilization of different components of the application. In essence, we build component profiles at different workload for each request category.

3. Using the SLA decomposition techniques to identify the resource requirements of different components at different workloads for each request category.

4. Establishing an analytical relationship between the resource requirements of the component and the workload.

5. Dividing the time horizon into multiple epochs. During each epoch, predict the workload and the resource requirements of the application in the next epoch. Repeat the steps from 1 to 5 or all applications that are co-located with this application.

6. Predicting the response time and throughput of an application when it is co-located with other applications on the same box. This in essence helps to account for the interference effects.

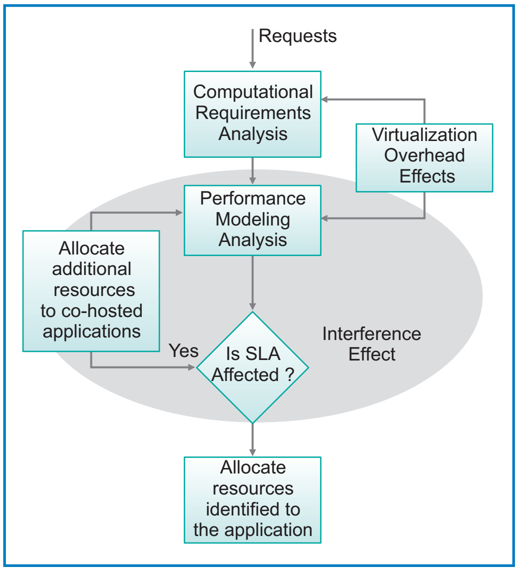

Figure 4: *Interaction between the SLA Decomposition Approach and the Performance Engineering Models*
Source: *Infosys Research*

7. Using the results of the performance testing in step 5 to revise the resource requirements of the application in step 3.

The overall interaction between the SLA decomposition technique and the performance models for identifying and quantifying the interference effect is shown in Figure 4. The approach presented helps in accounting for the interference effects while deciding the resource requirements of the applications.

CONCLUSION

SLA aware on-boarding of application is very critical for the successful adoption of cloud platforms. The need for new performance modeling techniques in this context has been explained in detail. A broad approach based on component profiling has been proposed to address the challenges associated with satisfactory performance of application on cloud platforms. The proposed approach can significantly improve the understanding of the application characteristics once deployed on cloud platforms. Additionally, it helps the service providers to provide more aggressive and practical deadlines for migrating the applications from the enterprise owned data centers to managed service provider's (MSP) data centers. The shorter schedules and elimination of the manual work further reduces the cost of operation for the service providers.

REFERENCES

1. Gartner Data Center Summit 2009, 5 – 6 October 2009, Royal Lancaster Hotel, London, UK, europe.gartner.com/ datacenter
2. S Balasamo et al., Model-based Performance Prediction in Software Development: A Survey, IEEE, Transactions on Software Engineering, 2004
3. P Barham et al., Xen and the Art of Virtualization, ACM SIGOPS Operation Systems Review, 2003
4. Y Koh et al., An Analysis of Performance Interference Effects in Virtual Environments, IEEE International Symposium on Performance Analysis, 2007
5. Y Chen et al., SLA Decomposition Translating Service Level Objectives to System Level Thresholds, 4th International Conference on Automatic Computing, 2007
6. G Jungy et al., Generating Adaptation Policies for Multi-tier Applications in Consolidated Server Environments, 5th International Conference on Autonomic Computing, 2008.

SETLabs Briefings
VOL 7 NO 7
2009

Can We Plug Wireless Sensor Network to Cloud?

By Adi Mallikarjuna Reddy V, Siva Prasad Katru and Kumar Padmanabh PhD

Cloud promises a remarkable transformation in the way people share and analyze real-time sensor data

Cloud computing is a holistic approach towards providing applications, platforms and infrastructure as an on-demand service over the internet through Web 2.0 technologies [1, 2, 3]. On the other hand, a wireless sensor network (WSN) consists of a number of tiny wireless sensor devices that have communication, computation, sensing and storage capabilities. These sensor nodes communicate with each other in an ad hoc fashion forming a WSN. They have been evolved in the past few years to enable solutions in the areas such as industrial automation, asset management, environmental monitoring, transportation business, healthcare, etc. [4].

Bringing various WSNs deployed for different applications under one roof and looking it as a single virtual WSN entity through cloud computing infrastructure is novel.

Data generated from a vast sea of sensor applications such as environmental monitoring, transportation business, healthcare, etc., is enormous. If we add this collection of sensor-derived data to various web-based virtual communities, we can have a remarkable transformation in the way we see ourselves and our planet. Some of the examples are — a virtual community of doctors monitoring patient healthcare for virus infection, portal for sharing real-time traffic information, real-time environmental data monitoring and analyzing, etc. To enable this exploration, sensor data of all types will drive a need for an increasing capability to do analysis and mining on-the-fly. However, the computational tools needed to launch this exploration can be more appropriately built from the cloud computing model rather than traditional distributed or grid approaches. Cloud computing models are designed to provide on-demand capacity for the application providers that involves three parties — the data center, the application provider and the application user vis-à-vis traditional approaches that operate on two party contracts.

Sometimes sensor data might not be of interest or sufficient to the consumers. The event of interests can be more important than raw sensor data. An event can be a simple or a composite event. Events such as temperature > 50 or humidity < 80 come under simple events. Events like fire or explosion detection which is a combination of two or more simple events come under composite event. These events are detected by considering readings from multiple sensors.

To summarize, integrating WSNs with cloud makes it easy to share and analyze real time sensor data on-the-fly. It also gives an added advantage of providing sensor data or sensor event as a service over the internet. The terms *Sensing as a Service* (SaaS) and *Sensor Event as a Service* (SEaaS) are coined to describe the process of making the sensor data and event of interests available to the consumers respectively over the cloud infrastructure.

We propose, a content-based publish/subscribe platform to utilize the ever expanding sensor data for various next generation community-centric sensing applications. This platform masks and virtualizes different WSNs and allows seamless integration of WSNs with the conventional cloud. This will shift the paradigm from the conventional sensor networks model to SEaaS sensor networks model. In this architecture - sensor, people and software are treated as individual objects that can be used to build community-centric sensing applications where people can share and analyze real time sensor data on-the-fly.

APPLICATION SCENARIOS

We consider WSNs deployed for two different applications.

Weather Monitoring and Forecasting System

Weather monitoring and forecasting system typically includes the following steps –

1. Data collection
2. Data assimilation
3. Numerical weather prediction
4. Forecast presentation [5].

Typically each weather station is equipped with sensors to sense the following parameters — wind speed/direction, relative humidity, temperature (air, water and soil), barometric pressure, precipitation, soil moisture, ambient light (visibility), sky cover and solar radiation.

The data collected from these sensors is huge in size and is difficult to maintain using the traditional database approaches. After collecting the data, assimilation process is done. The complicated equations that govern how the state of the atmosphere changes (weather forecast) with time require supercomputers to solve them.

Intelligent Transport Monitoring System

Traffic congestion has been increasing as a result of increased automobiles, urbanization, population growth and density. Congestion reduces efficiency of transport infrastructure, and increases travel time, air pollution and fuel consumption. Intelligent transport monitoring system provides basic management systems like navigation systems, traffic signal control systems, automatic number plate recognition and complex management systems like surveillance systems, systems that integrate data from other sources such as parking lot, weather, etc. [6].

Different sensors involved in this system are — inductive loops, Magneto meters, CCTV, GPS, etc. These sensors are mounted on vehicles, roads and buildings. Data available from sensors is acquired and transmitted for central fusion and processing. Predictive techniques can be developed in order to allow advanced modeling and comparison with historical baseline data. This data can be used in a wide variety of applications. Some of the applications are – vehicle classification, parking guidance and information system, collision avoidance systems, electronic toll gates and automatic road enforcement.

In the above scenarios, both the applications require storage of data and huge computational cycles. They also require analysis and mining of data to generate events. Access to this data is limited in both the cases. Integrating these WSN applications with the cloud computing infrastructure will ease the management of storage and computational resources required. It will also provide an added advantage of providing access to the application data over the internet through web.

SYSTEM MODEL

The system model depicted in Figure 1 consists of WSNs deployed for different applications, cloud infrastructure and the consumers. Consumers are those who seek services from the system. WSN consists of physical wireless sensor nodes to sense different modalities. Each sensor node is programmed with the required application. Apart from the application program, sensor node also consists of operating system components and network management components. On each sensor node, application program senses the modalities and sends back to gateway (in the cloud) directly or in multi-hop through other nodes. Routing protocol plays a vital role in managing the network topology and to accommodate the network dynamics.

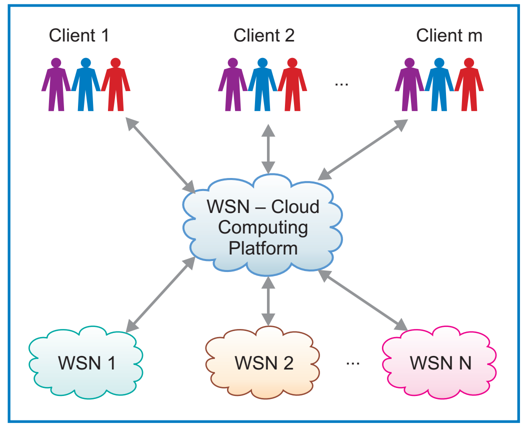

Figure 1: *System Model*
Source: *Infosys Research*

Cloud provides on-demand computational and storage resources to the consumers. It provides access to these resources through internet and comes in handy when there is a sudden requirement of resources or situations where it is not easy to assess the need in advance.

A WSN CLOUD COMPUTING PLATFORM

We propose content-based publish/subscribe platform, where the publishers are different WSNs deployed across geographical locations and subscribers are those who consume the information published. Publish/subscribe model gives an added advantage of publishers being loosely-coupled with subscribers and is scalable. The proposed platform consists of WSN virtualization manager (WSNVM), computation and storage manager (CSM), subscription registry manager (SRM), service provider (SP), metering and accounting manager (MAM) and SaaS/SEaaS application interfaces.

WSNVM masks the lower level details of each WSN cloud in terms of different platforms, sensors being used, data being generated, etc.

It also provides a unified view of different WSNs. CSM provides required computational cycles internally to process the data emanated from the sensors. It also maintains the historical sensor data of different WSNs. SRM manages the users' subscriptions and credentials. SP matches consumer interests with the sensor data and offers different disseminating mechanisms. Pricing for the offered services is calculated through MAM. SaaS/SEaS application interfaces are built using Web 2.0 technologies to access the WSN cloud platform services by clients. Figure 2 gives an overview of the components that constitute the WSN cloud platform.

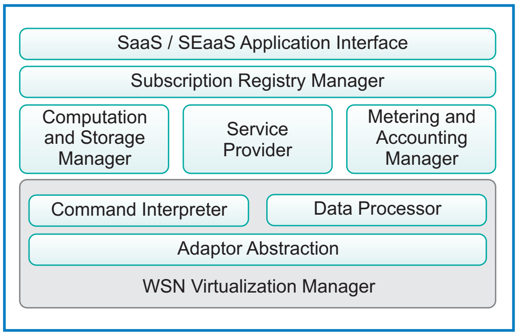

Figure 2: *WSN Cloud Computing Platform*
Source: *Infosys Research*

WSN Virtualization Manager

This component is divided into three sub-components. They are — adapter abstraction, data processing and interpretation, and command interpretation and processing.

Adapter Abstraction: This provides an abstraction to connect WSN with the gateway (gateway acts as a bridge between WSN and the server) in different ways (serial, USB and Ethernet). This abstraction is used for both communications i.e., from sensor network to gateway and vice versa. Gateway receives the raw byte stream from the communication ports and forms a raw packet out of it. This packet is queued up in a buffer for further processing.

Data Processing and Interpretation: When there is a packet available in the buffer, this component processes the packet according to the type of the packet. The packet type depends on the application being run on the platform. Processing of the packet involves extracting each field from the packet, interpreting, calibrating and applying engineering conversion formulas.

Command Interpretation and Processing: This provides reverse communication channel from the gateway to the WSN. This component is responsible for processing and interpreting various commands issued from different applications and generates the code that is understood by the sensor nodes. Reprogramming a node is also done through this component.

Computation and Storage Manager

This is same as general cloud computing infrastructure. This may not be directly related to the consumer, since he does not directly use the computation cycles or storage capacity. But, internally this module is responsible for processing and archiving the sensor data. Computation cycles are utilized internally to process the data that emanates from the sensors. Storing the sensor data will help to analyze the patterns in the data collected over a period of time. For example, weather forecasting requires solving enormous number of numerical equations over the historic data stored. Processed data records are stored in XML format.

Subscription Registry Manager

It maintains the credentials of different consumers' applications register to publisher/subscriber system for various sensor data required. For each application, registry component stores user subscriptions, sensor data and sensor event types the application is interested in. Each application is associated with a unique application ID along with the service level agreement (SLA). SLA provides basis for metering and accounting of services to be used, by covering all the attributes of the service customs. This agreement provides details concerning:

- The type of contract — limited time, long term, unlimited time, ad hoc, etc.
- The time model to be used — everyday, monday to friday, etc.
- The amount model that defines limits to the amounts of service to be provided
- Security — signatures and certificates for encryption and authentication
- Start dates and expiration dates of the contract.

Service Provider

The service provider module is divided into two sub components — analyzer component and disseminator component.

Analyzer Component: This component analyzes the incoming sensor data or event to match with user subscriptions in the SRM. If the sensor data or event matches with the interest of the subscriber, the same is handed over to the disseminator component to deliver to the appropriate users. Since the data and queries are in XML format, we use an algorithm similar to match the subscriptions of the users [7].

Disseminator Component: It receives the data or event of interest from the analyzer component and delivers the data through SaaS/SEaaS interface to the subscribed applications.

Metering and Accounting Manager

This module operates on a base assumption that all the services of the WSN cloud are contracted to the consumer via SLA mentioned above. Consumer uses signed web service requests to access the data.

Figure 3 depicts the UML sequence diagram that describes the role of MAM module in the WSN cloud platform.

- Request from the consumer to consumer web service
- Subscription registry manager checks the credentials of the service request using the supplied signature and gives a fault response (2.1) in case of unauthorized request
- Service request is sent to the MAM module, requesting it to start counting (3.1) the web service access
- The requested service is executed on the WSN cloud
- Service execution is completed
- Request message is sent to the MAM module, requesting it to stop counting the web service access

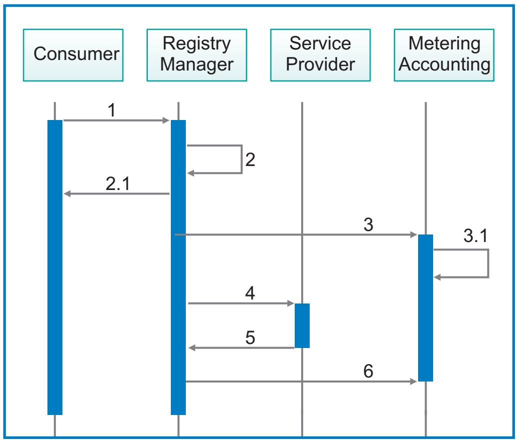

Figure 3: *Sequence Diagram*
Source: *Infosys Research*

- Message sent to the consumer indicating that the service is completed, returns the result.

SaaS/SEaaS Application Interface

The interfaces built with Web 2.0 technologies gives access to the WSN cloud platform web services. Consumers can consume the services through web services that are often referred to as internet application programming interface (IAPI). This allows the users to access the remotely hosted services over network, such as internet. Consumers can build their custom applications by weaving the required services from the WSN cloud platform.

The services are delivered to the consumers in the following ways. They are:

Continuous: As and when the requested data is available, it is sent to the consumers. The best example for this is fleet tracking with GPS sensor system. The vehicle position information is sent to the consumers continuously.

Periodic: The data is delivered to consumers at regular periodic intervals. A good example is to send across the temperature in the city at regular intervals of time to news agencies.

Event-based: The data is delivered when some event of interest occurs. This is often information deduced from the raw data such as detecting fire from temperature, humidity and light in the forest.

Query-based: Consumers can query for a specific data from the WSN cloud platform.

The two application scenarios described earlier in the paper are evaluated with the proposed WSN cloud platform. Deployed WSNs will relay the data to the gateway to which they are connected. Once the data is available to the WSN cloud platform, it takes care of the rest, right from processing to dissemination of the data (or event).

Once this system is in place, the consumers might be interested in the following services —

- Temperature of particular location/city periodically (e.g., one hour or one day)
- Weather forecast of particular location/city periodically (e.g., one hour or one day)
- Notify me when the rainfall in a particular location is above some threshold (e.g., >2cm)
- Notify me if some vehicle jumps over traffic signal
- Notify me if there is any fire event in the forest
- Notify me when particular bus reaches particular bus stop.

PLUGGING WSN INTO LEGACY CLOUD COMPUTING PLATFORMS

The proposed WSN cloud computing platform is a software platform that can be used on any of the legacy cloud computing infrastructure.

Two cases in this scenario are depicted in Figure 4. In case 1, WSN can be integrated with the legacy cloud infrastructure and the proposed software platform co-exists with the cloud management software platform such as load balancing algorithm, metering algorithms, etc. In the second case, proposed software platform co-exists with any other applications running on the cloud infrastructure as well as with the cloud management software.

Since most of the existing cloud computing platforms (hardware and software) in the market provide web services to access

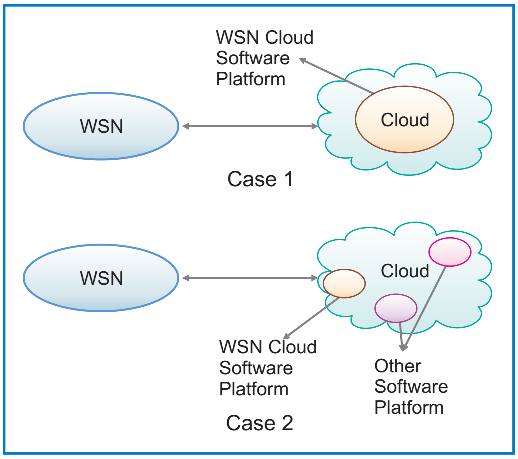

Figure 4: *WSN and Cloud Platform*
Source: *Infosys Research*

data and computing infrastructure, WSN virtualization manager uses them to store and retrieve the data from the cloud. Other services like registry, metering, service provider of the proposed platform together will run as an application instance over the existing cloud computing platform.

CONCLUSION

Cloud computing has been used as an extension of parallel processing. Coordinating various computing resources to achieve bigger task is the key of cloud computing. In wireless sensor network computing facility is available with each sensor node. Using the processed data from this intelligent sensor and using computing facility of the cloud will add another value to this domain. We believe it will shift the operational paradigm of the collaborative business process.

REFERENCES

1. J Geelan, Twenty-one Experts Define Cloud Computing, Cloud Computing Journal, January 2008. Available at http://cloudcomputing.sys-con.com/node/612375
2. L M Vaquero, L Rodero-Merino, J Caceres and M Lindner, A Break in the Clouds: Towards a Cloud Definition, ACM SIGCOMM Computer Communication Review, Vol 39 No 1, 2009. Available at http://delivery.acm.org/10.1145/1500000/1496100/p50-vaquero.pdf?key1=1496100&key2=5282660521&coll=GUIDE&dl=GUIDE&CFID=47965963&CFTOKEN=10653164
3. Weiss, Computing in the Clouds. netWorker, Vol 11 No 4, 2007
4. F Akyildiz, W Su, Y Sankarasubramaniam and E Cayirci, Wireless Sensor Networks: A Survey. Computer Networks, Vol 38 No 4, 2002
5. Weather Forecasting, Wikipedia
6. Intelligent Transportation System, Wikipedia
7. G Xu, J Ma and T Huang, A XML-based Composite Event Approach. In Proceedings of the First international Workshop on interoperability of Heterogeneous information Systems, Bremen, Germany, November 04 - 04, 2005
8. Harvard Sensor Network Testbed, MoteLab. Available at http://motelab.eecs.harvard.edu/
9. www.citysense.net
10. M M Hassan, B Song and Eui-Nam Huh, A Framework of Sensor-Cloud Integration Opportunities and Challenges. In ICUIMC '09: Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication,

New York, USA, January 2009, ACM. Available at http://delivery.acm.org/10.1145/1520000/1516350/p618-hassan.pdf?key1=1516350&key2=2312660521&coll=GUIDE&dl=GUIDE&CFID=49153557&CFTOKEN=72600633.

SETLabs Briefings
VOL 7 NO 7
2009

Cloud – Five Minutes into the First Quarter

In a discussion with Jitendra Thethi, Principal Architect, Infosys Technologies, Amitabh Srivastava, Senior Vice President, Microsoft Technologies throws light on the promise that Cloud holds for the computing industry but cautions against naïve temptations to migrate to Cloud in one go

Jitendra: Cloud computing is all new and organizations do not see a value yet in terms of leveraging the platform. Do you see a complete change in perspective and attitude from the standpoint of an IT organization? How do businesses look at the concept of considering two worlds now? How do you visualize the shift in perception, as one has to look at data in datacenters as well as data in cloud?

Amitabh: Cloud holds a lot of promise for the computing industry but the word 'cloud' is not well defined or even agreed upon. Everyone has their own definition of the word 'cloud' and sometimes hype takes over reality. So let me start with our definition of cloud and then describe our approach to provide the best value to our customers. We define cloud as a massive geo-distributed computer consisting of commodity machines, load balancers and switches that are spread across the globe. Cloud computing presents this massive geo-distributed computer as a utility service.

So, our approach to cloud computing has the following key facets:

- An operating system (OS), Windows Azure, manages this massive geo-distributed cloud computer. Our operating system based platform approach provides two benefits:
  - o First is cost. The OS efficiently owns and manages all the computing resources and also automates all management functions. This helps us drive the costs in the data center down, both capex and opex.
  - o Second is agility. Cloud is a complex environment with tens of thousands of computers operating in data centers across the globe. The OS masks the complexities by providing a rich set of abstractions that developers can use to write their cloud applications. This allows developers to focus only on their business logic and quickly take their application to market.
- Cloud is an extension of the on-premises IT. Cloud and IT are not an either-or option. Unlike some who believe that everything will move to the cloud, we believe customers should have the choice to decide what runs in their IT and what runs on the cloud. Many customers will

continue to rely on their on-premises IT for some class of applications. For example, some data has to be kept on-premises due to issues like compliance, security and privacy. Applications that require special hardware or have special connectivity and bandwidth requirements for performance reasons will continue to be on-premises. At the same time there are many workloads that will benefit from cloud. So, rather than forcing customers to pick cloud or IT, our approach is to make cloud a seamless extension of IT and let the customers decide what to run where. We will make it simpler by providing value services that allow applications to communicate securely between the cloud and on-premise IT and make it easier for services to federate ID.

- Developer's existing skills transfer to cloud. On Windows Azure, we use the same Windows programming model, so the APIs are still Win32 and the same development tools still work on cloud. Windows Azure supports all languages, and by providing command-line interfaces and REST protocols it can interface with all tools and interoperate with other platforms. Of course, there are certain aspects one needs to learn about cloud, but majority of skills simply transfer.

Easy development of new generation of applications that will span across three screens (PC, phone and TV) supported by IT and cloud. With the same Windows platform on the three screens, IT and cloud we are striving to provide a uniform and integrated experience. This will enable developers to innovate and bring new generation of applications quickly to market.

Jitendra: I do agree that the newer applications exploiting convergence, bringing different channels and serviced by a common platform is definitely one workload that can exploit cloud well. But would you tell us, what are the existing workloads in an enterprise setup that can be moved to cloud and be leveraged for cost optimization? Also, what are the additional business advantages of moving to the cloud?

Amitabh: Let us look at the characteristics of cloud.

One of the key features is elasticity. Hardware for applications are generally set up to handle peak load. For example, a service may require 1000 machines to handle peak load but on an average it only needs 30 machines. So in a traditional environment we have to provision for 1000 machines. Such applications that are elastic in nature are suitable for the cloud where one can easily add capacity on demand, only pay for what is used. So, on the cloud you will only provision the 30 machines and then provision more machines as the load increases. In addition, the machines can be returned when peak load subsides.

Another aspect is the globally distributed facet of business, where cloud is designed to be geographically distributed across continents. Cloud provides a convenient way of migrating data seamlessly across geographically distributed centers. Cloud benefits applications that are global in nature.

Enormity of scale is one of the key characteristics of cloud. Cloud is designed by using commodity machines in a highly distributed environment. If there is any application that requires massive scale, cloud is designed to handle it.

Availability is another important feature of cloud. Cloud is designed to be available everywhere, all the time. By using redundancy and geo-replication, cloud is so designed that services be available even during hardware failures including full data center failures. Our platform goes further to make services available even during updates OS and the application itself.

Many of this ultimately translates into savings in cost. For this, it is important to measure the total cost of ownership. This should include not only the hardware costs but also management and operations cost.

Jitendra: What will be the guidance to the customers who are looking at moving to the cloud? What do they need to do to be prepared to move to cloud?

Amitabh: Take a thoughtful approach. The first is to not panic and just rush into the cloud. Using my favorite American football analogy, I'd say that cloud is only 5 minutes into the first quarter. You should first try the cloud. It is important to understand the different features the cloud offers, see how you will integrate it into your environment. Then review the architecture of your application to see if your application is taking full advantage of the cloud. There is temptation to quickly take the application 'as is' to the cloud. It is like 'outsourcing you hardware' but you will not enjoy the full benefits of the cloud. Cloud provides many benefits that will lead to very substantial cost saving and give you agility in your application development, and these gains will easily make up for any initial investment you make in taking a thoughtful approach.

Jitendra: Amitabh, as you said you are building a platform that is horizontal. What in your terms defines building vertical solutions to our customers?

Amitabh: Windows Azure is a general platform that is designed to enable easy development of a wide range of applications. Our partners, ISVs, system integrated, etc., will build the various solutions. Partners, ISVs, etc., with domain knowledge in specific areas will build the verticals on our platform. We will help lower their costs and help and provide them with a rich platform that lets get to the market quickly.

Jitendra: Thanks Amitabh for your time. It has been truly wonderful talking to you and knowing your thoughts about how our customers can benefit from the Azure platform.

About the Interviewer

Jitendra Pal Thethi is a Principal Architect with Infosys and anchors presales activities for Infosys solutions and IP built on disruptive technologies in the areas of Cloud Computing, Collaboration, Data Virtualization, Call Center Optimization and Mobility. Jitendra has more than 14 years of experience in IT Industry as a Solution Architect and Technology consultant.

About the Interviewee

Amitabh holds 14 patents and has published a variety of papers. His paper on ATOM with Alan Eustace in PLDI 1994 received the Most Influential PLDI Paper Award in June 2005. He is the author of OM, ATOM and SCOOPS software systems, which have resulted in products for Digital Equipment and Texas Instruments on the Alpha and PC platforms. He led the design and development of Vulcan, a second-generation binary transformation system, at Microsoft. Vulcan is the foundation of a wide variety of tools developed at PPRC.

Amitabh earned a Bachelor's degree in Electrical Engineering from the Indian Institute of Technology, Kanpur, India and a Master's degree in Computer Science from Pennsylvania State University. He received the 2003-2004Distinguished Alumnus Award from the Indian Institute of Technology, Kanpur, and was selected as the 2004 Outstanding Engineering Alumnus at Pennsylvania State University.

SETLabs Briefings
VOL 7 NO 7
2009

Cloud Computing Identity Management

By Anu Gopalakrishnan

Online security concerns are on the rise and a robust identity management is what cloud needs now

Latest technology facilitates different service providers to unite their efforts to address a broader business space. It is possible that consumers hold multiple accounts with the service providers like e-bay, Gmail, etc. The visibility and scope of attributes for every identity has to be verified against a central trusted policy framing authority, assumed by the systems. In such a system, much is at stake if identities are not handled with extreme precaution. Such scenarios are common to high-end applications hosted on cloud computing environment. Identity management (IDM) assumes an upper hand in the whole area of cloud security. Cloud computing is an amalgamation of various technologies to meet the demands of an interdependent maze of software and services. This necessitates several IDMs, based on various technologies to inter-operate and function as one consolidated body over a cautiously shared user space. Hence IDM in clouds projects a number of new dimensions that traditional IDMs cannot meet.

Most cloud vendors have a simplified proprietary IDM solution with shortcomings that have to be understood. The challenge in this area is that there are considerable efforts towards outsourcing the IDM that gave birth to the concept of identity-as-a-service (IaaS) [1]. IaaS vendors focus on comprehensive, interoperable and quick-to-deploy solutions.

UNDERSTANDING THE NEW DIMENSIONS OF IDM IN CLOUDS

The evolution of cloud computing from numerous technological approaches and business models such as SaaS, cluster computing, high performance computing, etc., signifies that the cloud IDM can be considered as a superset of all the corresponding issues from these paradigms and many more. An IDM in cloud has to manage — control points, dynamic composite/decommissioned machines, virtual device or service identities, etc. Cloud deployments are dynamic with servers launched or terminated; IP addresses dynamically reassigned; and services started or decommissioned or re-started. So, as traditional IDM, merely managing users and services is not sufficient. When a deployment or service or machine is decommissioned, the IDM has to be informed so that future access to it is revoked. IDM should ideally store its details till it becomes active. Meanwhile access to its relevant stored data has to be monitored and granted by the defined access level for that mode as mentioned in SLA. Traditional IDM is not directly amenable for cloud computing due to these peculiarities of cloud.

Today's cloud requires dynamic governance of typical IDM issues like, provisioning/de-provisioning, synchronization, entitlement, lifecycle management, etc.

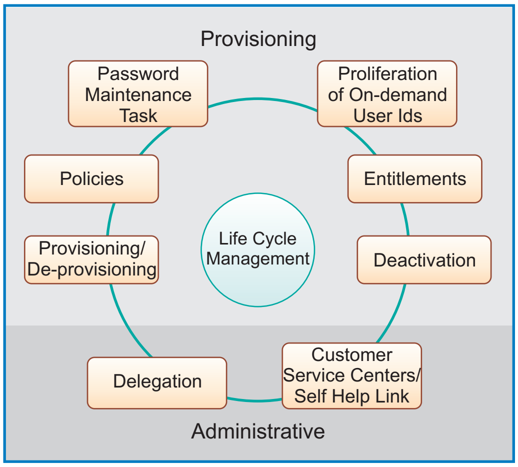

Figure 1: *The Identity Life cycle Management*
Source: *Infosys Research*

IDENTITY LIFECYCLE MANANGEMENT

Lifecycle management incorporates an integrated and comprehensive solution for managing the entire lifecycle of user identities and their associated credentials and entitlements. Functionally, it is divided into two components – the provisioning component and the administrative component. Administrative component defines delegations rules, providing self-service components to change personal details or make requests to the users. Delegation of administrative rights to local group or process-in-charge is crucial for a volatile and dynamic cloud based scenarios. Decentralizing the tasks will reduce the load on the authenticator component and also save time in making access control decisions. Figure 1 illustrates the various components of lifecycle management.

Provision and De-provisioning

In cloud, provisioning means just-in-time or on-demand provisioning and de-provisioning stands for real time de-provisioning. Just-in-time provisioning indicates the federation of user accounts without sharing prior data, based on some trust model. Service Provisioning Markup Language (SPML) provides XML based structures for representing provisioning or de-provisioning requests intended for identity lifecycle management [2]. SPML can make use of Service Administered Markup Language (SAML) assertions and facilitate a complete trust model between senders and receivers. SAML defines an XML based framework for exchanging security information for enabling SSO or identity federation regardless of the underlying architecture. OASIS Security Services is currently working on developing a SAML 2.0 profile for SPML. SAML can help SPML to establish trust and quantity, a subject against which the SPML provisioning request is targeted. This makes just-in-time provisioning and real time de-provisioning possible.

Real time de-provisioning of a user account has to synchronize instantaneously with all participating service providers. Any delay in de-provisioning could lead to security vulnerability. Some of the issues like — ways in which de-provisioning of one user affects the other federated identities in cloud are matters of judgment on the functionality of the application deployed on the cloud.

Entitlement

Entitlement refers to the set of attributes that specify the access rights and privileges of an authenticated security principal. Lack of interoperable representation of this information poses a challenge as the information needs to be exchanged among different cloud based service providers. In the absence of interoperable format, expensive and customized syntactic translation components are needed. The semantic aspect still remains to be tackled.

While some applications like SalesForce have built-in control for entitlement and authorization control for multiple attributes, others require the help of OAuth or similar such technologies [3].

Proliferation of On-demand User ID

Proliferation of on-demand user ID is a big concern in cloud computing IDM as the occurrence of multiple identities for the same user in multiple service providers' security repositories cannot be ruled out. A simple way to overcome this problem is by the adoption of OpenID mechanism [4]. OpenID works by making one primary user id as the key to authenticate a single end user with multiple service providers. However, the difficulty in this approach lies in the trust propagation and development of trusted relationships [5].

Synchronization services help expedite the roll-out and expansion of federated identity management capabilities by enabling services in cloud to federate accounts and other data necessary to build up trust relations.

CLOUD ARCHITECTURE

Cloud architecture plays an important role in choosing your IDM, SaaS or the all-in-one Platform-as-a-Service (PaaS) [6]. SaaS requires only application access, whereas PaaS will require system access (for accessing the underlying platform) as well as application access (for accessing the hosted application on the underlying platform). Both require a common IDM that can integrate well into the existing authentication mechanism. The third type of cloud architecture is Infrastructure-as-a-Service (IaaS), which is not mentioned explicitly, since the IDM requirement of PaaS and IaaS are comparable. Consider one of the most common SaaS IDM implementation using ping identity. Ping identity works by deploying the technology behind the firewall and making the identities exportable [7]. This IDM mechanism allows integration of a number of authentication mechanisms such as Microsoft Windows based authentication, LDAP authentication, CA site minder, etc. It is deployed on top of the existing authentication infrastructure and the deployment is quite efficient and fast. It uses SAML to transfer credentials. It can be perceived as a layer of abstraction over the traditional IDM that fights the challenges of IDM. This aspect of it makes this IDM architecture easy to deploy and dynamic.

PaaS is commonly defined as the delivery of a computing platform and solution stack as a service. It includes workflow capabilities for application design, application development, as well as application services such as team collaboration, web service integration, etc. PaaS IDM automatically scales up to include all these features. This is illustrated in Figure 2 overleaf.

PaaS IDM has to address various functional modules like source control, test

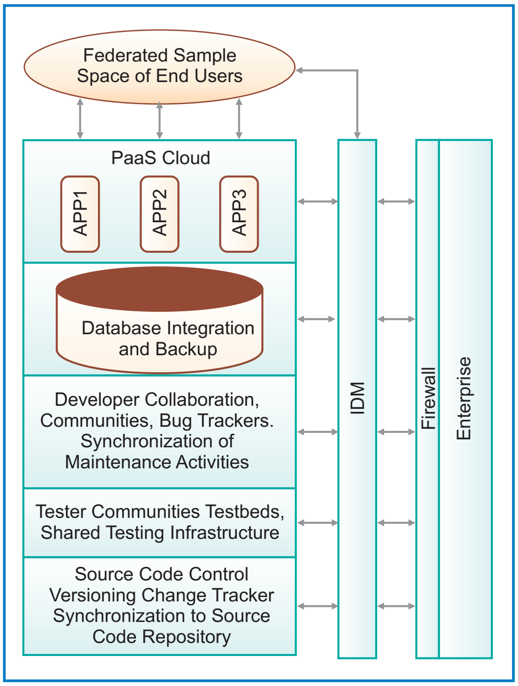

Figure 2: *PaaS IDM*
Source: *Infosys Research*

modules, development communities, etc. For the sake of simplicity, the PaaS IDM could adopt a Role-Based Access Control (RBAC) system to handle each of this and its user space. An RBAC system for source control will allot minimum set of privileges to the developer accounts and essential services, depending on the interdependency of the applications hosted on the platform. For test communities, IDM manages tester accounts, privileges, auto-run test suites and knowledge collaboration portals of the tester communities required for hosting a test bed. In case of development communities, IDM manages the collaboration of developer communities, access and privilege of each group of developer, the bug tracker system, etc. The cloud could also expect IDM to handle the database challenges, by controlling the access and synchronization with the in-premise segments. In addition to all these, IDM handles the SaaS based challenges of federated user space.

Due to vender lock-ins, the primary limitation with PaaS happens to be a fact that the complex IDM solution designed for PaaS is rendered useless while migrating to another cloud. A simple slice of IDM requirements are plotted here to illustrate the complexity of the PaaS IDM.

USER CENTRIC ACCESS CONTROL

The traditional model of application-centric access control, where each application keeps track of its collection of users and manages them, is not feasible in cloud based architectures. This is more so, because the user space maybe shared across applications that can lead to data replication, making mapping of users and their privileges a herculean task. Also, it requires the user to remember multiple accounts/passwords and maintain them. Cloud requires a user centric access control where every user request to any service provider is bundled with the user identity and entitlement information [8]. User identity will have identifiers or attributes that identity and define the user. The identity is tied to a domain, but is portable. User centric approach leaves the user with the ultimate control of their digital identities. User centric approach also implies that the system maintains a context of information for every user, in order to find how best to react to in a given situation to a given user request. It should support pseudonyms and multiple and discrete identities to protect user privacy. This can be achieved easily by using one of the open standards like OpenID or SAML.

FEDERATION OF IDENTITIES

On the internet, it is likely that each user ends up with multiple credentials and multiple access permissions across different applications provided by different service providers. These fragmented logins present a challenge to the users and service providers, in forms of synchronization of shared identities, security, etc. There is a strong need for an intrinsic identity system that is trusted across the web and within enterprises and unambiguously identifying users.

Federation of identities maintained by the multiple service providers on the cloud is very critical to cloud based service composition and application integration. An expected issue in this regard is the naming heterogeneity. Different SPs use different factors for authentication like account number, email ID, PayPal ID, etc. Also, when transactions traverse multiple tiers of service hosted in clouds, the semantics of the context of identity information has to be properly maintained, constrained and relaxed as per specific needs. Consider a complete transaction cycle for an e-bay purchase, based on PayPal account. It traverses from e-bay to supplier, through various tiers in supplier's domain to get approvals, release and shipping. Then it goes through PayPal to approve, validate, release the pay, bill the amount to the customer, etc. For each step, the federation authority decides the essential attribute of the customer to be shared with each department.

The user identity mapping in the previous environments have been one-to-one, or in other words, user ID to single user profile. In cloud architectures the mapping challenge is many-to-one, one-to-many and pseudonyms. Pseudonyms are for privacy protection details, when a user does not want his identity to be tracked as he crusades various domains.

Another issue is the trust relation setup between the service providers of the federated world. Currently it is based on policy files framed by the local authority, depending on various factors like the domain trust information automatically fed in by the trust authorities. This is not a scalable or flexible model that can meet cloud computing demands. Cloud scenarios require dynamic trust propagation and dynamic authorization.

VOLATILITY OF CLOUD RELATIONS

In a traditional model, the IDM is based on the long-term relation of a user to an organization or trust domain. In cloud, which represents the current e-commerce world, the relationships change dynamically and quickly, and the IDM has to incorporate all that. Any retrieval or cache of the volatile data has to be done cautiously. The possible damage of using old data should be studied. Like, if the user has changed his password login with old password, it should be restricted and locked in all the applications that are participating in the identity federation. Live data fetching, domain name resolution, canonicalization of the data like URL, account IDs, etc., are the challenges.

SCALABILITY

Cloud requires the ability to scale to hundreds of millions of transactions for millions of identities and thousands of connections – with short/rapid deployment cycles. Performance has to be N+1 scalable across the globe and deployments agile and quick (weeks not quarters/years). With the software today it takes ~6 months to make a single SAML/SSO connection and it doesn't address the access control and compliance issues. Open Cloud Manifesto states that clouds have to dynamically scale up and down, so that nobody needs to hoard resources to handle peak hours [9].

INTEROPERABILITY

The mass expects the cloud to provide a IDM solution that can interoperate with all existing IT systems and existing solutions as such or with minimum changes. Seamless interoperation with different kinds of authentication mechanism such as the Microsoft Windows authentication, SSO, LDAP, SAML, OPENID and OAUTH, OpenSocial, FaceBookConnect, etc., is what is expected of cloud. The syntactical barriers have to be bridged. It requires an authentication layer of abstraction to which any model of authentication can be plugged in and off dynamically.

TRANSPARENCY

Security measures assumed in the cloud must be made available to the customers to gain their trust. There is always a possibility that the cloud infrastructure is secured with respect to some requirements and the customers are looking for a different set of security. The important aspect is to see that the cloud provider meets the security requirements of the application and this can be achieved only through 100% transparency. Open Cloud Manifesto exerts stress on transparency in clouds, due the consumer's apprehensions to host their applications on a shared infrastructure, on which they do not have any control [9]. Transparency can be achieved by complete audit logging and control.

PATTERNS IN CLOUD IDM

Based on the insights gained so far three patterns in cloud IDM can be concluded. The ideal scenarios for each pattern are also mentioned.

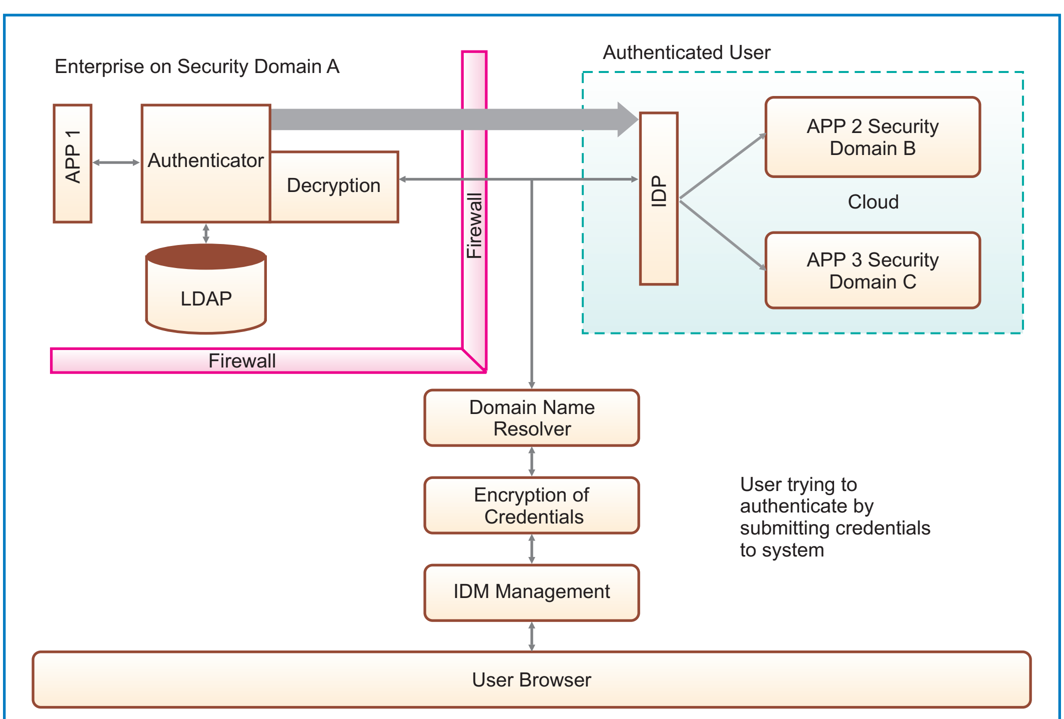

Figure 3: *Trusted IDM Pattern*

Source: *Infosys Research*

Trusted IDM Pattern

This pattern is intended for a smaller or even for a private cloud that requires security. Scalability is definitely not a feature of this cloud. But Google App Engine (appengine.google.com) that follows this pattern assures that the scalability is not a major concern at the moment as the number of requests that could be tunneled through simultaneously is quite large. The main feature of the pattern is that the authentication is always performed within the firewall. The credentials are submitted to the IDM component and it takes care of encrypting and tunneling the credentials through a secure channel to the authenticator. IDM is independent of the authentication mechanism. Hence deployment and integration is fast and efficient. Once the user is authenticated in by any authentication mechanism, then rest of the participating servers trust the user. The attributes of the user can be shared using some mechanism like SAML. Authorization can be effectively handled by XACML. A basic model of this pattern is illustrated in Figure 3 on page 50.

External IDM

This pattern is very similar to the initial pattern but for the fact that the credentials are submitted directly to the authenticator [Fig. 4]. The credentials can be collected by a different browser window, channeled by SSL. The pattern is intended for a public cloud. The IDM concentrates only on domain resolution and triggering of the authenticator to resolve the authentication. This is the architectural pattern adopted by ping identity. In ping identity, domain resolution is done by referring to a spreadsheet of valid users that is always kept updated. It can also be achieved through other mechanisms like standard domains name

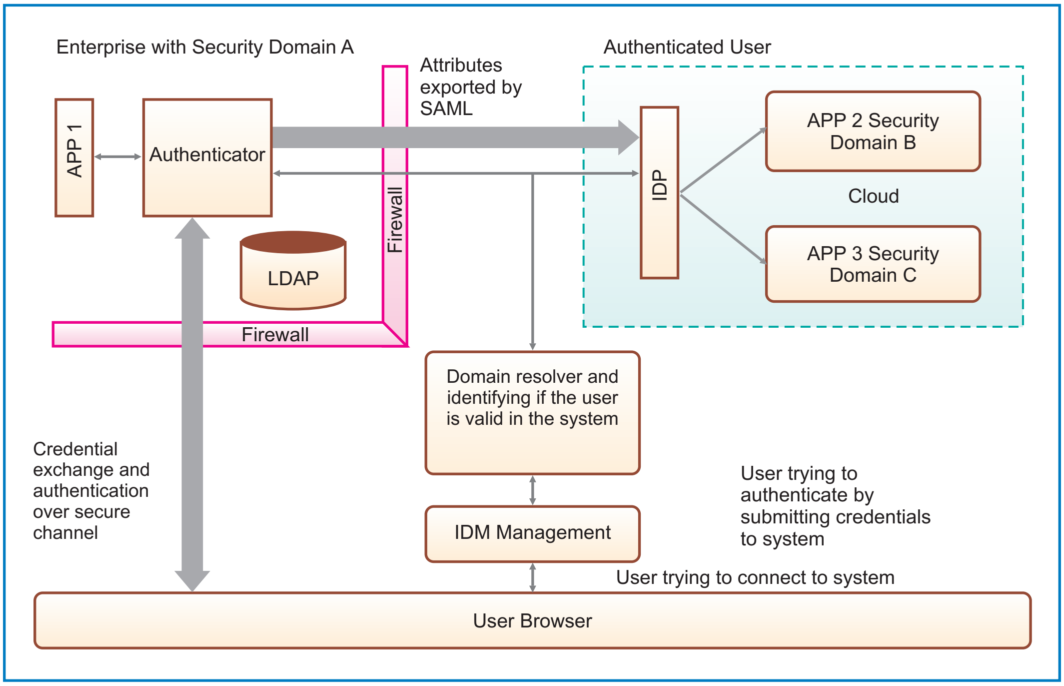

Figure 4: *External IDM* ***Source:*** *Infosys Research*

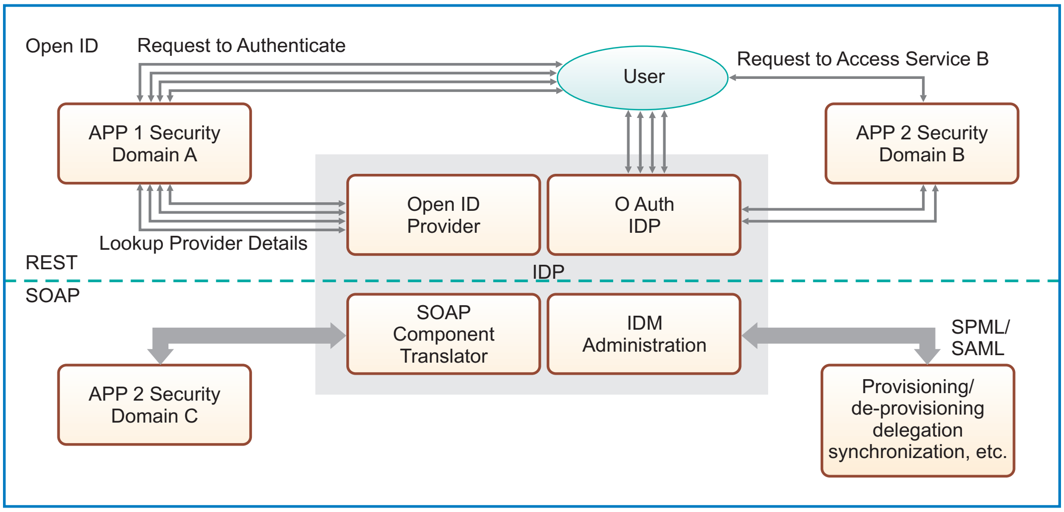

Figure 5: *Interoperable IDM* ***Source:*** *Infosys Research*

resolution, discovery or YADIS protocol, or XRDS query, etc., depending on the underlying technology used. The same drawback of pattern 1 exists in pattern 2 also. Scalability is an issue. Symplified (www.symplified.com) is vendor on cloud IDM, whose solution has close resemblance to this pattern.

Interoperable IDM Pattern

This pattern illustrates a cloud to cloud scenario, using OpenID and OAuth. The identity mechanism used, will understand and interoperate multiple identity schemes. OpenID is an open and decentralized standard for user authentication and access control, by allowing users to logon to multiple services with the same digital ID. Any service provider can authenticate the user in to the system. OAuth is again an open protocol that enables a user to grant permission to a consumer site to access a provider site without any sharing of credentials [10]. SPML is used for XML based IDM LC. This is extremely useful for an e-commerce web world where there are multiple service providers based on a common user space. The central identity system, understands all technologies used for authentication like SAML, OpenID, OAuth, etc. Let us assume that the central identity system to be collection of modules, each handling a technology, taking to a common user space and a policy database. The information is converted to different formats, depending on the technology used like OpenID, or SAML, or WS-Security and conveyed to the participating service providers [Fig. 5].

A brief comparison of the three patterns is shown in Table 1.

CONCLUSION

Of the emerging technologies cloud computing has a lot of substance. The huge set of challenges it has brought with it has to be captured and tamed to produce more benefits. Choice of IDM design for any cloud should be tailored to suit the definition of that particular cloud and open to any kind of enhancements the cloud is bound

Features	Trusted IDM Pattern	External IDM	Interoperable IDM
Security of Credentials	Very Secure	Submitted to IDP Network	Depends on Authentication Mechanism
Interoperability	Interoperable, since it is oblivious of the underlying authentication mechanism	Interoperable	Interoperable to any Authentication Mechanism and Technology
Type of cloud the pattern is best suited for	Private Cloud	Can be used in public clouds since the credentials are always submitted directly to the authenticator module and secrecy is maintained	Huge Public Clouds over Multiple Technologies
Scalability	Not Scalable Easily	Not Scalable Easily	Scalable
Speed of Deployment and Implementation	Very Fast	Fast	Speed depends on the number of technologies required
Examples of this Pattern	Google App Engine's SDC	Ping Identity	Proposed Design

Table 1: *Summary of the Patterns* **Source:** *Infosys Research*

to have in future. Essentially the design should be capable of incorporating any number of trust domains and of maintaining an effective shared user pool. As the next generation IDM IaaS, a user centric identity management is intended to be a complete all-round solution addressing all possible issues of cloud IDMs [11]. It may be the answer to the growing complexity of IDMs. The intent is to take away the complexity of IDM away from the enterprises, thereby allowing them to direct their energy and resources on their own functions, while the IaaS vendors provide the best solution or IDM based on their expertise.

REFERENCES

1. Open Cloud Manifesto, Spring 2009. Available at http://www.opencloudmanifesto.org/opencloudmanifesto1.htm
2. RSA's contribution to Cloud security guidelines. 2009. Available at http://www.cloudsecurityalliance.org/guidance
3. OAuth. Available at http://oauth.net/ OpenID Authentication 2.0 Final, 2007. Available http://openid.net/specs/openid-authentication-2_0.html
4. Illustration of OpenID based on Plaxo's use of Yahoo OpenID. Available at http://www.plaxo.com/api/openid_recipe
5. Luis M Vaquero, Luis Rodero-Merino, Juan Caceres and Maik Lindner, A Break in the Clouds: Towards a Cloud Definition, Cloud Architectures, Vol 39 No 1, Jan 2009. Available at http://delivery.acm.org/10.1145/1500000/1496100/p50-vaquero.pdf?key1=1496100&key2=0736171521&coll=GUIDE&dl=GUIDE&CFID=50720541&CFTOKEN=61415293
6. Ashish Jain, A blog on Ping Identity, Jan 12, 2009. Available on http:// itickr.com/?cat=29
7. Service Provisioning Markup Language Specification, version-1, June 2003. Available atxml.coverpages.org/PSTC-CS-SPMLCORE10.pdf

8. Christian Emig, Frank Brandt, Sebastian Kreuzer and Sebastian Abeck, Identity as a Service – Towards a Service-Oriented Identity Management Architecture, Lecture Notes in Computer Science, 2007. Available on http://www.springerlink.com/content/5865u474424qw751/.

SETLabs Briefings
VOL 7 NO 7
2009

Service Exchange @ Cloud

By Bhavin Raichura and Ashutosh Agarwal

Every stakeholder can maximize her benefits in the service exchange scenario that is powered by cloud computing

Service exchange @ cloud is a platform, where the service publishers and service subscribers can do business online for mutual benefits. It is not a new idea from business perspective. Currently, there are several players in this space like Ariba, Seekda!, webservicesX, Zuora, etc. The current trends around cloud computing and SaaS has significant impact on the traditional offerings in this area. The concept of service exchange @ cloud can be extended as an enabler of enterprise SOA implementation in private cloud scenario. It can also act as a catalyst for IT consolidation and lean IT transformation for large enterprise and government IT landscape. This discussion will focus on the value proposition of cloud computing in service exchange scenario and how it creates a win-win situation for each stakeholder.

Service exchange @ cloud provides a platform to publish web services, search pre-existing web services and subscribe and consume the published web services. Figure 1 overleaf illustrates the service exchange concept.

Service publishers can be anyone – it can be an individual developer, small or large independent software vendor (ISV) or a system integrator. Similarly, the service subscriber can be an individual, a corporate or an enterprise consuming these services over internet or through mobile devices.

The traditional web service exchanges such as Seekda! and webservicesX, provide a similar transactional platform for service publishers and service subscribers. Zuora provides value-added billing, payment and subscription management platform for such an exchange driven by publish-subscribe model [1, 2, 3].

The concept of service exchange is very extensible and can have a maturity model. For instance, once the service exchange is setup, the demand for integration platform will arise. The integration will be required for service–enterprise use case as well as service-service use case for Web 2.0 mash-up. The service exchange platform can bring the following value-propositions:

- A new revenue-channel through an eMarketplace for small/large ISVs or individual developers, along with the established business model (broker)

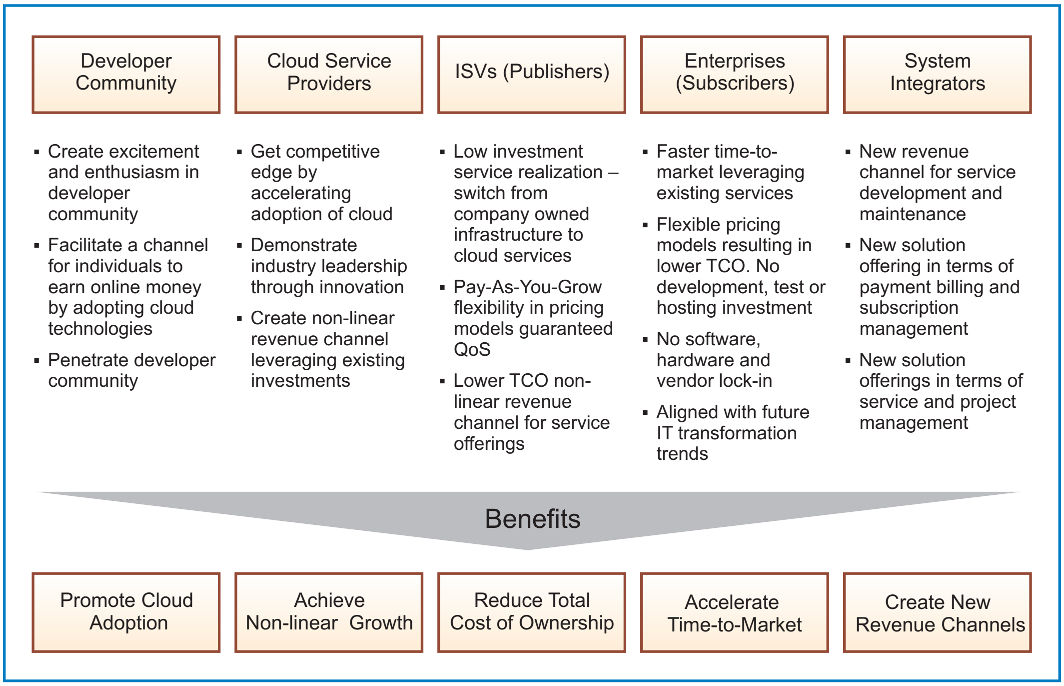

Figure 1: *Stakeholder Benefits* ***Source:*** *Infosys Research*

- Enterprise-service integration that will be required in most cases to leverage existing enterprise investments

- Service–service integration on the platform to leverage cross-service functions and provide value mash-ups.

In the course of this discussion there is a need to understand the key use cases for such a platform implementation and how cloud computing can add value to the traditional web service exchanges.

SERVICE EXCHANGE USE CASES

The use cases identified for service exchange are simplified in this paper for the need of lucidity in discussion. The actual implementation will be much more comprehensive, complex and extended. Figure 2 identifies key uses cases for service exchange implementation.

Register

The publishers and subscribers need to register to avail the services from service exchange. There will be separate registration processes for publishers and subscribers. The registration process will capture the required information, enroll the users and provide a security mechanism in terms of authentication and authorization. It will also capture the information related to payments and accounts for monetary transactions. The monetary transactions and related subscription management can also be facilitated through external service providers like Zuora.

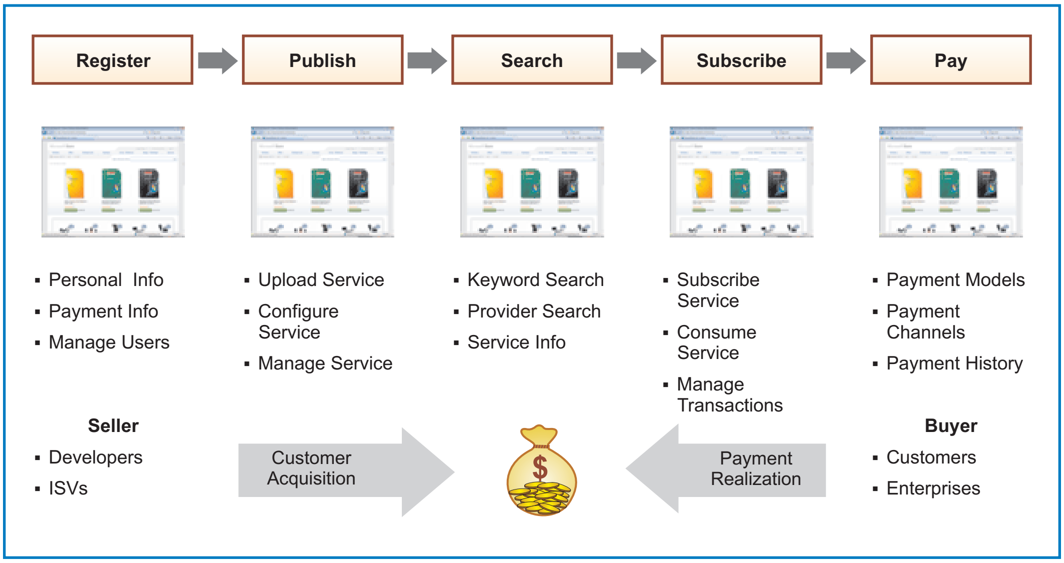

Figure 2: *Service Exchange Use Cases* ***Source:*** *Infosys Research*

Publish

After registration, the publishers can be presented with a user interface wherein they can upload the service binaries and configuration and can avail the web URL that can be used to consume the services.

The publish use case will present a publisher admin console with more comprehensive options to configure, modify, delete or suspend the service and set up the data feeds for the service. The service will also be configured to provide security for restricting the unauthorized access.

Search

The search use case will provide a basic interface to subscribers to search and identify the service they want to subscribe to. Subscribers will be able to search for the existing services through basic keyword search or by using more advanced search capabilities.

The advanced search option will include the search by service providers, technology platform and many other meta-data information. The search results will present the list of service providers that can offer the required service. It can also further help subscribers with information like rating of the service provider, rating of the service and various other service evaluation parameters along with detailed service documentation.

Also, from the perspective of revenue model it offers an opportunity for ad revenue channel through service sponsorship.

Subscribe

The subscribe use case will facilitate the subscribers to create, manage and configure services subscriptions. It will present a subscriber admin console to view, modify, configure, delete or suspend existing subscriptions. It will help subscribers to configure the security required to access and consume the services subscribed. It will also present a history view of the transactions related to the subscriptions.

Pay

The pay-per-use case addresses the monetary aspects of the service realization. It will be consumed by all — publishers, subscribers and the service exchange host. It will present the information and alerts related to payments, consolidated and comprehensive reports for financial transactions. It will also have interface with external systems for payment realization.

CLOUD COMPUTING VALUE PROPOSITION

Service exchange @ cloud has a great potential to become another success story similar to App Store, eBay or YouTube. From technology perspective, cloud computing technology brings the following value:

- Dynamically scalable infrastructure (on-demand)
- Guaranteed quality-of-service in terms of performance, scalability and availability of hosted services.

From business perspective, cloud computing brings attractive pricing models for individuals, start-ups or enterprises:

- Lower initial investment in terms of capital expenditure (capex)
- Flexible pricing and IT service models (opex).

Service exchange is comparatively an innovative business idea and there will be constraints on the budget to experiment. At the same time, huge infrastructure support is required to manage scale and quality-of-service.

Also, the business offering needs to consider various customer segments like – individuals, ISVs and enterprises - and demonstrate huge flexibility in terms of the pricing and service models.

The problem for the key decision makers to realize the business of service exchange is to balance the investment with potential growth and also having support for flexible pricing models - cloud computing simplifies this problem.

These value propositions from cloud computing facilitates and makes decision makers comfortable with the initial investment required to start an innovating offering and scale-up the infrastructure on-demand as the business grows using pay-as-you-grow pricing models.

Also, for large enterprises, consider the above benefits to existing Ariba deployments to understand how it adds value by bringing Ariba as SaaS on Cloud platform.

Although, the benefits sound interesting and promising, there are multiple challenges in realizing it, viz.,

- Lack of standardization across large players
- Lack of maturity of existing solution and service offerings
- Lack of appropriate business case and success stories to convince C-level executives and
- Lack of clarity on security, data and IP ownership in cloud based deployment scenarios.

KEY STAKEHOLDER BENEFITS

Service exchange @ cloud has something for everyone in the value-chain. Figure 3 articulates the value proposition of the cloud

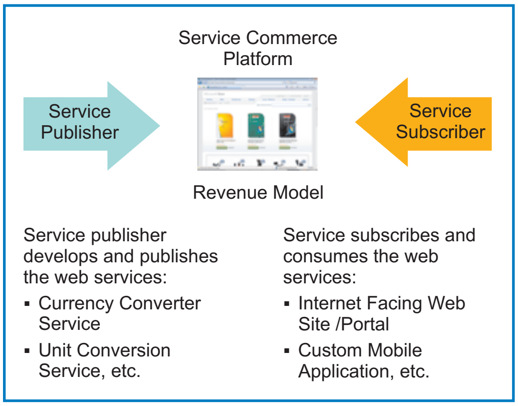

Figure 3: *Key Stackholder Benefits*
Source: *Infosys Research*

computing technology and benefits to key stakeholders. Service exchange realization can happen in multiple deployment scenarios — over internet, over private clouds or over extranet (partner network). We will articulate benefits to each stakeholder in different business scenario.

Developer Community

Service exchange @ cloud over the internet scenario provides opportunity to individual developers to develop and deploy services to earn online money. It gives a great opportunity to talented freelancers to earn money online.

Cloud Service Provider

The cloud service providers such as Microsoft, Amazon and Google can achieve competitive edge by promoting cloud adoption by driving developer community and enterprise to the proprietary service exchange.

The service exchange product offering suitable to enterprises or government for private cloud offering can open a new revenue channel for non-liner growth.

Publisher

The publishers get a low investment platform with high quality of service (QoS) services that can be consumed by enterprises in production scenarios. It creates a non-linear revenue channel for small and medium ISVs to sell their services to a large service exchange marketplace.

Service exchange also provides flexible pricing models to attract more business and offer competitive pricing. It also offers flexible investment models to facilitate pay-more-as-you-grow and start with low capex.

Subscriber

The subscribers get ready-to-use services from service exchange that can significantly influence the time-to-market new services from subscriber's perspective. It helps promoting the enterprise reuse in private cloud scenario that helps reducing the total cost of operation (TCO). The subscribers (enterprises, corporate, individuals) have multiple options of service providers, the payment and pricing models and service models to choose from and select the best-aligned for reuse. Also, all this comes without any software, hardware, vendor or investment lock-in that gives tremendous business agility for the decision makers.

System Integrator

Service exchange opens up new traditional application development and maintenance (ADM) opportunities around service development, deployment, maintenance, management, monitoring and configuration.

The innovative solution and service offering around billing, payment and subscription management can create non-linear revenue channel for system integrators for enterprise, government and other private cloud or enterprise SOA scenarios.

Large Enterprises

Large enterprises having a vision to implement enterprise SOA can benefit from enterprise wide reuse of the services through service exchange. It presents significant cost saving opportunities for capital expenditure as well as operational expenditure. It will act as a key enabler for enterprise SOA implementation.

Government

For the government IT landscape, service exchange can act as a catalyst for lean IT transformation and IT consolidation for significant cost savings and reducing TCO through private cloud realization.

CONCLUSION

Service exchange @ cloud is a highly scalable monetizing platform. Cloud service providers can and should promote the adoption of cloud offerings. The ISVs can offer various software features as services. Just as enterprises can accelerate time-to-market new services, system integrators can create new business and revenue channels and individuals can make money online. The success of such a business model is also well tested and proven as Apple App Store is to promote iPhone. Also, large enterprises and the government can conceptualize service exchange in the private cloud scenario to implement enterprise SOA while consolidating IT infrastructure to reduce TCO.

REFERENCES

1. http://seekda.com/
2. webservicesx.NET. Available at http://www.webservicex.net/WCF/default.aspx
3. Zuora: Z-Commerce Platform
4. http://www.zuora.com/products/zcommerce/
5. Bhavin Raichura and Rajat Kumar, Semantic SOA – IT Catalyst for Business Transformation, AMCIS 2007 Proceedings, AIS Electronic Library, Colorado - USA, 2007
6. Bhavin Raichura and Shaurabh Bharti, Achieve Dynamic Integration & Runtime Process Optimization using Semantic SOA, ICSD 2007, Document Research and Training Centre (DRTC), Bangalore, 2007-02
7. Bhavin Raichura and Vijay Rao, Lean IT Transformation, ebizQ.net, 2009-03. Available at http://www.ebizq.net/topics/saas/features/11121.html.

SETLabs Briefings

VOL 7 NO 7

2009

Revenue and Customer Growth for ISVs using Cloud Offerings

By Ajit Mhaiskar and Bhavin Raichura

The agility of cloud is the biggest attraction for the ISVs operating in a restricted space and budget

There are around 75,000 independent software vendors (ISVs) worldwide that drive approximately $250 billion of the software industry revenue. These ISVs produce, package, sell, deliver and update software. The market share in the ISV industry is highly skewed, wherein the top 2% ISVs garner about 80% of the industry revenue. This top 2% (about 1,700 ISVs) includes all ISVs with over $10 million in software revenue [1]. The remaining 98% of the ISVs have very limited resources in terms of ability to spend on software development, marketing, sales, software distribution and deployment. In this paper, we focus on how cloud computing offers the large number of small ISVs unique opportunities for revenue and customer growth with significantly lower capital and operating investments. We also discuss the new service offerings that small as well as large ISVs can bring to the market by leveraging cloud computing.

ISVS AND CURRENT INDUSTRY TRENDS

The existing trends in the industry like Web 2.0, social commerce, SOA, SaaS, virtualization, cloud computing and the commoditization of business intelligence provide unique opportunities to ISVs to do more with less. Table 1 overleaf shows the various opportunities available to ISVs, enabled by these recent technology trends.

Virtualization is one of the top trends in the industry today and provides important benefits to ISVs.

BENEFITS OF VIRTUALIZATION TO ISVS

Most of the ISVs today offer solutions to their customers in an on-premise model or in a hosted model. Virtualization has already become a major trend in the IT industry, resulting in ISVs and large enterprises reaping substantial benefits from adoption of virtualization technologies in their infrastructure. ISVs that have not adopted virtualization yet can certainly consider adopting it for the significant benefits it can provide.

A good example is of ICICI bank, the largest private bank in India. The bank used virtualization to consolidate 230 physical servers to just 5, running a little under 650

Technology Trend	ISV Opportunities				
	Reduce TCO	Grow Business	Improve Customer Satisfaction	Improve Agility	Competitor Differentiation
Web 2.0	Low	Medium	High	Low	High
Social Commerce	Low	High	Medium	Low	High
SOA	High	Medium	Medium	High	High
SaaS	High	High	Medium	High	Medium
Virtualization	High	Low	Medium	High	Medium
Cloud Computing	High	High	Medium	Medium	Medium
Business Intelligence	Medium	Medium	Medium	Medium	High

Table 1: *Various Opportunities Available to ISVs*
Source: *Infosys Research*

applications in their data center. This move resulted in an annual operating expense (opex) savings of over seven figures in Indian Rupees, due to higher efficiencies related to power, cooling and space. The break-even period, considering capital expenditure (capex) was about six months, with projected savings for five years of about 57 million rupees ($1.1 million) [2].

For ISVs that have already adopted virtualization, the next step is the idea that these virtual machines can be run from suitable infrastructure in any location – either within the premises of the ISVs data center or in some third party data center or somewhere on the internet, in the cloud. That is the promise of cloud computing. VMware President and CEO Diane Greene says that the evolution of virtualization began with users deploying virtual machines (VMs) for testing and development and then easing into server consolidations for production environments. The third phase was resource aggregation, with entire data centers being virtualized, followed by automation of all aggregated workloads. Cloud computing forms the final liberation phase [3].

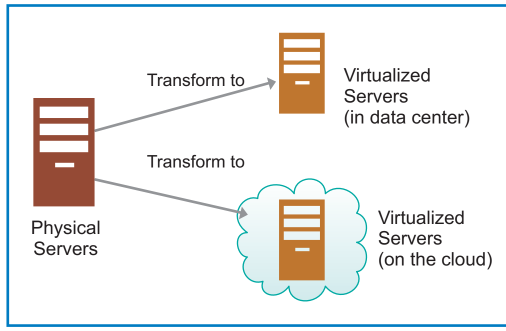

Figure 1: *Virtualization and Cloud Computing*
Source: *Infosys Research*

Almost all cloud service providers today use some form of virtualization technology to abstract the hardware underneath. Most clouds employ infrastructure software that can easily add, move or change an application with little to no manual intervention. Figure 1 shows how virtualization and cloud computing co-exist and how ISVs can transform physical servers in their data center into virtualized environments either in their data center or on the cloud or both.

THE PROMISE OF CLOUD COMPUTING

Cloud computing offers an excellent opportunity for cash-strapped ISVs to do more with less and provides them unique levers in the areas of software distribution, marketing and deployment of web-based solutions. The cloud computing technology brings together a huge amount of virtualized hardware, required software and competent IT staff to monitor these assets. The cloud computing environment and related software components are mostly fully owned, managed, supported and serviced by the cloud service provider. Gartner describes cloud computing as Infrastructure-as-a-Service [4].

The cloud computing environment can be partly dedicated (shared cloud) to a client or fully dedicated (private cloud) to a client and managed by the cloud service provider. The cloud service provider and the client can negotiate the terms for pricing, QoS, SLA and operations level agreement (OLA). Billing is done based on usage (computing based billing - $/CPU/hr or storage based billing - $/GB or data transfer based billing - $/Mbps or $/Gbps).

Companies like Amazon, SalesForce.com and Google are the pioneers in offering cloud based services. Amazon has the first mover advantage in the cloud computing area and has generated an estimated 500 million dollars from cloud offerings alone [5]. The following is a partial list of cloud offerings from different vendors –

- Amazon's elastic compute cloud (EC2) is a web service that provides resizable compute capacity in the cloud that is designed to make web-scale computing easier for developers.

- Amazon's simple storage service (S3) is an online storage web service that provides unlimited storage through a simple web services interface and has been one of the pioneers in the area of offering highly-scalable cloud based storage for a price.

- Google's AppEngine offers users the ability to build and host web applications on Google's infrastructure.

- Akamai is extending its content delivery network (CDN) to offer cloud based services. Akamai offers optimization services for cloud acceleration, cloud business continuity, cloud security, cloud applications and storage.

- Salesforce.com provides customer relationship management (CRM) solution to businesses over the internet using the SaaS model and was one of the pioneers in offering SaaS solutions.

- Facebook offers its infrastructure to developers to leverage social services.

- IBM's Blue Cloud and Microsoft Azure are the new offerings on the block. IBM recently announced LotusLive Engage, an integrated social networking and collaboration cloud service designed for businesses of all sizes.

- There are also many VC-funded startups in the area of cloud computing (Coghead, Bungee, LongJump, EngineYard, RightScale, etc).

- Virtualization solution leaders like Citrix and VMware have also presented visions of cloud infrastructures.

With big players like Microsoft, IBM and Google now entering the cloud computing and storage provider market by making big investments, the cloud services provider space is maturing fast and getting commoditized. It will be prudent for most ISVs, to desist from entering the cloud services provider market and instead focus on building new solutions around offerings from big players like Amazon, Microsoft, Google and IBM.

Table 2 overleaf shows a high-level comparison of various cloud service providers in the context of the ISV market.

Cloud Service Provider	Considerations				
	Offering Maturity	Market Adoption	ISV Focus	Platform Capabilities	Competitor Differentiation
Amazon	High	High	High	High	High
Google	Medium	Low	Medium	Medium	Medium
IBM	Low	Low	High	Medium	Medium
Microsoft	Medium	Medium	High	High	Medium
Salesforce.com	High	Medium	Medium	Medium	Medium
Startups	Medium to high	Medium	Medium	Medium	High

Table 2: *Cloud Service Providers in the context of the ISV market*
Source: *Infosys Research*

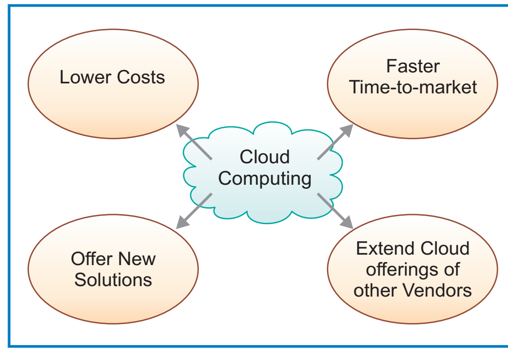

Figure 2: *Opportunities for ISVs provided by Cloud Computing*
Source: *Infosys Research*

LEVERAGING CLOUD COMPUTING FOR REVENUE AND CUSTOMER GROWTH

Cloud computing offers a great advantage to ISVs, especially the small ones, as they can now leverage the power of big data centers at low cost through the cloud service providers. This acts as a great leveler and provides plenty of freedom to innovate. There are several areas where ISVs can leverage cloud computing to get better value for money spent –

- Lower application development, solution deployment and support costs
- Faster time to market
- Offer new solutions to customers using the cloud
- Extend cloud service offerings of other vendors.

Figure 2 shows the cloud computing benefits to ISVs that help them to accelerate revenue growth and customer acquisition.

Lower Costs

Lower Application Development Costs: Most of the cloud service providers provide a rich set of tools to design, build, deliver and market cloud services. The cloud service providers also provide a powerful, scalable computing environment along with scalable storage. The cloud platform APIs allow for easy development without having to overly focus on scalability and performance aspects. A lot of development complexity and details are abstracted away by the APIs and tools provided by the cloud service providers. All this will help in lowering application development costs and providing faster time to market. However, this will also need ISVs to learn new skills in application development and will also need a significant change in mindset to deliver services using cloud infrastructure.

Lower Solution Deployment and Support Costs: With the adoption of cloud computing, most ISVs will not need to build and maintain data centers of their own. For ISVs that already have data centers of their own, cloud computing will provide additional hosting infrastructure that is highly scalable and manageable at a fraction of the cost. This will provide ISVs with easy ability to reach a significantly larger user base than what they currently support and scale quickly depending on the success of the solutions that they provide.

Faster Time-to-Market

With the help of cloud service offerings, ISV developers have to worry less about scalability and focus on aspects like solution functionality and performance of key use cases. By leveraging the ready-made services and plumbing provided by cloud service vendors, ISVs can bring their solutions to market much faster with significantly reduced investments.

ISVs should start looking for ways to quickly build, deploy and take advantage of the flexibility that cloud computing environments can bring. Vendors like IBM, Microsoft, Google and Amazon are making it easier for software developers to build solutions based on open standards that are well supported by a vast array of technical resources.

Offer New Solutions

Two interesting ways in which ISVs can leverage cloud services to offer new innovative solutions to their customers are by:

- Extending existing solutions
- Offering new web-based solutions leveraging cloud services.

Extending Existing Solutions: ISVs can extend existing solutions in various ways –

- Offer online services for existing software solutions that are currently deployed to desktops. Microsoft is promoting a similar strategy called Software + Services to offer online service extensions to its vast array of highly successful desktop-based software solutions comprising primarily of Microsoft Office and Windows.

- Offer on-demand versions of existing web-based solutions.

Cloud computing will enable ISVs to take risks with significantly lower investments in capital and operating expenditure, but still being able to scale up quickly to meet peak processing demand without over investing.

Callidus, a leader in the sales performance management (SPM) software market, spent about three years building an on-demand version of its existing products. More than one-third of the customers today use the on-demand model and this number is likely to grow to more than half the customers using it in a few years. With the addition of on-demand offerings, Callidus has been able to successfully open up the market and add several new customers by offering lower prices [6].

Offering New Web-based Solutions Leveraging Cloud Services: In general, SaaS brings business value in terms of a flexible and economical business model rather than a real technology value. Cloud computing technology complements SaaS by helping to realize these flexible business models by offering utilization-based pricing for computing and storage resources.

ISVs can build new web-based solutions, take them at the global level fast and scale them very quickly to meet global demands using cloud services. ISVs can also improve customer satisfaction by leveraging Web 2.0 and social commerce concepts in an innovative manner to offer new solutions for horizontal and vertical markets.

ISVs can offer new solutions in different verticals like manufacturing, healthcare, financial services, retail, energy management, etc., by leveraging cloud services

ISVs can also offer horizontal solutions in the areas of business intelligence and analytics, compliance, managed services, etc.

About an year and a half back, Siemens started looking at next-generation data centers and examining where unified communications (UC) fit into the picture. Gradually, the company developed a strategy to port its existing unified communications software to Amazon's Elastic Compute Cloud (EC2). The ultimate goal for Siemens is to give partners and customers a front-end portal that allows them to pick and choose the UC services needed in a flexible manner [7].

In the context of web-based ISV solutions, the challenge is to balance IT investment for a global expansion of the solution with actual growth through sales. Even for large ISVs, it is almost impossible to proactively plan scalability to enable global operations. It is also impractical to block large investments in terms of hardware, software and people while the operation size is relatively small. What is needed is a dynamic and on-demand scalability of IT assets and related services as the solution adoption grows globally. Cloud computing offers this dynamic and scalable infrastructure to facilitate quick growth in an economical manner.

Extend Cloud Service Offerings of Other Vendors

ISVs have a good opportunity to build new solutions extending existing cloud service offerings from vendors like Amazon, Google, Microsoft, IBM, etc. Some key areas where the existing vendor offerings can be extended are –

- There is a need for better tools to manage cloud deployments. Management tools that can help manage existing datacenter deployments as well as cloud deployments in an integrated manner will greatly help.

- New security, compliance and management solutions can be built to extend existing cloud service offerings

- ISVs can offer lift-and-shift services or solutions to customers who are interested in virtualizing their existing applications and putting them on the cloud.

- ISVs can also offer new kinds of managed services built around cloud service offerings from other vendors.

CHALLENGES AND RISKS ASSOCIATED WITH CLOUD SERVICES

For ISVs, cloud services are not without risks. Some of the associated risks are -

- Most of the cloud service providers today offer no guarantee of data and can also suffer occasional outages which could impact business. Few vendors allow security or process compliance audits of their cloud infrastructure.

- Most vendors today have implementations which will result in a significant vendor lock-in, even though they talk about standards compliance.

- ISVs have traditionally built hosted solutions or desktop based solutions. Making the shift from the current mindset of delivering desktop or web-based software to delivering services using utility computing will be very hard work and will require skills that most of the small ISVs do not have currently. While some ISVs will be able to take advantage of cloud services, the vast majority of ISVs will have a very difficult time making this switch.

Daryl Plummer from Gartner says that ISVs are not positioned well to become the next generation of Cloud Service Providers (CSP) or even SaaS providers. He says, some ISVs will either change their business entirely, or go out of business if cloud computing becomes the mainstream norm for delivery of systems [8].

Table 3 shows some of the key challenges for ISVs associated with cloud services.

CONCLUSION

In the tight economy prevalent today, companies are spending much less on IT and ISVs will have to take growth wherever they can find it. Cloud computing is a double edged sword which presents a significant challenge as well as an important opportunity for ISVs. ISVs offering pure-play hosting services will really struggle in fending off the big cloud vendors. ISVs offering on-premise software will be forced to innovate and build extensions to their software which uses cloud-based services. The cost of deploying software in the cloud will keep reducing at a brisk pace, potentially leading to innovative ISV offerings built around the cloud infrastructure resulting in increased competition that is very fast-moving. Cloud computing innovations

Challenge	Details
Potential Competition from Cloud Service Providers	Most ISVs will have to partner with cloud service providers like Amazon, Microsoft, Google, IBM, etc., to deliver their solutions. If these ISV services are offered as part of a bigger service offering, the ISVs will be opening doors to potential competition. Pricing will also become key as profits will have to be shared with the cloud services provider. Building a high level of trust and credibility with the cloud services provider will be critical. ISVs will also need to bring in significant differentiation in their solution offering.
Introduce New Pricing Models	ISVs will have to change their pricing model to include software, computing, storage and service price. This new model will be challenging and could impact profit margins.
Maturity of Offerings	The cloud computing technology is still maturing and many of the cloud computing offerings are not yet production ready. There are also open issues around data security, compliance, data ownership and standardization which need to be addressed.
Higher Adoption Risk	Unless the open issues around data security, compliance, data ownership and standardization are addressed, adoption of cloud services could be low. This low adoption could increase the implementation risk for ISVs. ISVs will need to display a significant amount of courage, passion and leadership to make their cloud-based offerings successful.
Extending Existing Solutions Using Cloud Services	Extending existing ISV solution offerings to the cloud will be challenging as it will involve significant enhancements and risks. Costs involved could also be significant if the existing solution is a pure desktop-based solution.
Handling Cloud Service Outage	Most cloud vendors today don't provide availability assurances and SLAs are mostly non-existent. Cloud vendors also don't allow embedding of security and management agents or monitors. Occasional outage of services from providers like Amazon, Google, IBM, Microsoft, etc., is a distinct possibility and recent outages have only provided fodder to this thought. ISVs will have to devise a plan to keep customers informed about such outages and assuage them if such outages occur.
Lack Of Geographic Coverage	With the exception of Akamai and Layered Technologies, no cloud vendor allows the placing of an application in a specific geography on the cloud. Most cloud service providers today don't have geographic coverage. Lack of geographic coverage could lead to significant performance challenges.

Table 3: *Challenges Associated with Cloud Services* ***Source:*** *Infosys Research*

happening in the industry are certainly a major point of inflection for the ISV market. ISVs that are able to innovate and navigate through these shifts will stay on to fight another day, while those who fail to innovate will perish.

REFERENCES

1. Dan Lohmeyer, How does Microsoft work with ISVs? Available at http://blogs.msdn.com/msftisvs/archive/2007/08/22/how-does-microsoft-work-with-isvs.aspx
2. Real CIO World, December 15, 2008
3. Bridget Botelho, VMware Entering Final Phase of Virtualization Evolution: Cloud Computing, IT Knowledge Exchange, May 2008. Available at http://itknowledgeexchange.techtarget.com/server-virtualization/vmware-entering-final-phase-of-virtualization-evolution-cloud-computing/
4. http://www.gartner.com/it/page.jsp?id=868812&tab=agenda
5. http://web2innovations.com/money/2008/04/22/amazon-web-services-on-its-way-to-surpass-500m-in-sales-this-year/
6. The Cloud Shift: Don't forget ISVs. Available at http://www.on-demandenterprise.com/blogs/The_Cloud_Shift_Dont_Forget_ISVs_31079814.html
7. Siemens Channel Partners: Testing Unified Communications In Amazon's Cloud, April 2009. Available at http://www.thevarguy.com/2009/04/02/siemens-channel-partners-testing-unified-communications-in-amazons-cloud/
8. Daryl Plummer, Delivering Cloud Services: ISVs - Change or Die or both! Gartner, November 2008. Available at http://blogs.gartner.com/daryl_plummer/2008/11/06/delivering-cloud-services-isvs-change-or-die-or-both/
9. http://www.forrester.com/imagesV2/uplmisc/CloudComputingWebinarSlideDeck.pdf
10. James Staten, Cloud Computing for the Enterprise, Forrester Research, February 2009. Available at http://www.forrester.com/imagesV2/uplmisc/CloudComputingWebinarSlideDeck.pdf
11. R Wang, IBM Storms Into Cloud Computing With an ISV Friendly Pricing Model, Forrester Blog, October 2008. Available at http://blogs.forrester.com/appdev/2008/10/ibm-storms-into.html
12. IBM to Deliver Software via Cloud Computing With Amazon Web Services. Available at http://www-03.ibm.com/press/us/en/pressrelease/26673.wss
13. Azure for ISVs. Available at http://www.microsoft.com/azure/isv.mspx
14. Windows World, Demystifying Cloud Computing, May, 2008.

SETLabs Briefings
VOL 7 NO 7
2009

Power in the Clouds?

By Sudeep Mallick PhD and Ganesan Pandurangan

Parallel computing and HPC workloads find their architectural options in cloud computing

High Performance Computing(HPC) discipline emerged with an aim to reduce the total execution time of an application that involves complex computations that require inordinately large amount of time to execute. It also aims at reducing the time involved in the execution of the same logic repeatedly over an inordinately large data set. HPC techniques are, by and large, based on the concepts of parallel programming. It aims at reducing the total execution time of an application by having multiple sections of it run concurrently in time in such a manner that the behavior (or functionality) of the application remains unchanged by the engineered parallelism. This is based on the assumption that the application code has some inherent parallelism that can be exploited and re-engineered.

HPC has already become critical for an enterprise's survival [1]. High throughput, low latency, huge data churn tasks such as customer analytics, risk analysis, oil and gas exploration, simulation for options pricing, drug discovery have made enterprises embrace commercial HPC techniques long back, seeking to minimize makespan and maximize the throughput of the applications.

Cloud computing has become a serious architectural option for commercial HPC applications. This is primarily due to a few critical trends and happenings that is important to appreciate – the most notable being the recent release of Amazon's Elastic MapReduce [2] accompanied by powerful auto-scaling and load balancing features. Emergence of mature public cloud platform providers and sophisticated cloud platform management solutions from the big players and adoption of virtualization technologies by most large businesses paving the way for creation of on-premise private clouds are two significant developments.

Additionally, the emergence of parallel computing frameworks such as MapReduce (popularized by Google), Microsoft's Dryad have significantly improved perception about the ease of use and effectiveness of large scale parallel computing on commodity clusters. Finally, some very encouraging benchmarking results on the performance of major public cloud platforms and use cases have come to light from the HPC scientific community who are the frontrunners in this space [3, 4, 5].

This leads us to the questions such as — what are the types of parallel problems? What are the line-of business (LOB) applications that are best suited for clouds? What is the suitability of the existing software and hardware HPC techniques on the cloud? And most importantly, how suitable cloud computing is for HPC?

WHY IS CLOUD GOOD FOR HPC?

Let us identify the aspects that make cloud computing an attractive proposition for HPC workloads.

Versatile Support for Elastic Parallel Computing Execution Environment

Cloud is attractive for HPC primarily because a well conceptualized cloud platform (public or private) provides a wide range of parallel computing options on it. As is well known, the high performance in HPC is achieved through parallel computing techniques [6]. Inherent parallelism (at bit, instruction, task levels) in target application exhibiting various degrees of parallelism (fine, coarse and embarrassingly parallel) is exploited by computation of the parallel portions on multiple processors (CPUs, multi-cores or nodes on a cluster), custom hardware platforms and accelerators (GPUs, FPGAs [1]) using different architectures and programming models (shared and distributed memory).

At a basic level, suitability of a cloud platform for an HPC application would be determined by the support provided by the platform for the execution platform requirements — CPU (speed and numbers), latency and bandwidth of memory hierarchy (cache, RAM, disk) and the network. Typically, cloud platforms provide a reasonably wide variety of compute units in terms of CPU speed, number of cores, frequency, architecture type (32, 64 bit), memory capacities and hard disk storage. Moreover, the variety is available in as many numbers as required. This makes it suitable for a wider range of existing on-premise HPC applications.

Versatile Support for Parallel Computing Styles: The availability of uni-core instances, multi-core instances and cluster of instances from the public cloud vendors makes it amenable for different architectures – shared (as in SMPs, CMPs) and distributed memory (clusters/grids) and programming models – shared (OpenMP, pThreads, etc.) and distributed address space (MPI, PVM, etc.). The existing commercial HPC applications and libraries written using these architectures and programming models can be ported on to cloud infrastructures for deriving additional benefits such as ROI and scaling.

Versatile Support for Workload Variety

Compute intensive tasks that exhibit significant data parallelism such as Monte Carlo simulation over large data sets for generating risk analysis reports in finance, BLAST searches in Life Sciences, N-body simulation, etc., can be executed on a cluster of high power CPU instances provisioned from the cloud. Monte Carlo simulations also require significant caching requirements that can be provided by distributed caching across multiple compute instances. Memory intensive tasks limited by memory latency such as dynamic programming, sparse matrix computations, etc., in many financial applications are suitable for running in instances that have higher RAM capacities and multi-core instances sharing the same physical node. For compute, memory and communication intensive tasks such as, dense linear algebra (DLA) computations as in oil and gas exploration and simulation applications that require small size message exchange can perhaps be executed on a fewer multi-core instances provisioned from the cloud rather than more number of low end uni-processor instances. The shared memory model would obviate memory latency and bandwidth issues.

For example, for tasks exhibiting coarse grained or are embarrassingly parallel, such as web page search, indexing, machine learning, etc., low speed network interconnects are not a problem and distributed memory model is appropriate for scaling. Some of these applications can experience performance gains by exploiting data parallelism on a larger number of low strength processors having a larger amount of distributed memory (total RAM across the cluster). Frameworks such as MapReduce are popular on cluster of nodes provisioned from the cloud. However, it is important to note that MapReduce is just one of the many different categories of parallel computing models [7]. Offline batch workloads where the batch data can be uploaded on the cloud storage space are very suitable for clouds.

Flexible yet Optimal Provisioning

Cloud computing infrastructure platforms coupled with dynamic provisioning features enable flexible ramp up and down of resources based on SLA requirements. Policy aware provisioning enables specification of thresholds and scenarios for resource ramp up and down to handle unexpected workload fluctuations. This enables low variation in performance and scalability in true sense. For constant workloads such as drug discovery and protein synthesis this does not matter, however for variable workloads faced by financial analytic applications this would be of great help. Amazon cloud's auto scaling feature and the provisioning and management solutions from RightScale are examples in this area. Cloud computing enables flexibility not at the cost of optimal resource allocation, but in consonance with it.

Freedom from Performance Clippers and Achievement of Better Architectural Match

The trade-off in cloud computing is between cost and performance, unlike the on-premise case where there are hard limits to the available horsepower and hence performance gains. Often parallel computing application architectures encounter bottlenecks in specific portions resulting in sub-optimal provisioning and performance. For example, the master in a master-slave configuration often becomes a bottleneck due to its centrality in the architecture, similarly certain nodes in an HPC cluster responsible for reading/writing to data sources/sinks become bottlenecks due to I/O latency and bandwidth limitations. Cloud makes possible better matching of architectural requirements.

Availability of Feature rich HPC Frameworks

The advent of the Elastic MapReduce framework by Amazon has heralded the beginning of the availability of HPC frameworks tailored to cloud computing infrastructures. Job scheduling and resource provisioning are closely tied to the topology of the cloud infrastructure and can be optimized by the cloud provider. For example, provisioning the MapReduce cluster from the same subnet or physically proximal set of hardware can result in obvious performance gains which only the cloud provider can make possible. Moreover, the cloud HPC user does not have to handle the onerous tasks of setting up clusters, provisioning adequate capacity nodes (for example, high end compute node, I/O capacity node for masters in a master-slave configuration).

Clouds for Real time Workloads

HPC workloads such as extreme transaction processing, distributed query processing, complex event processing, streaming data applications, real time analytics applications are more suitable for private clouds (in the current state of maturity of public clouds). These applications are characterized by the need for online or real time responses from high performance computation on large on-premise data, often generated in real time.

HPC Data Grid

Cloud infrastructure is appropriate for storing huge data sets for HPC computations, such as databases in BLAST searches in life sciences applications, financial market data from third party providers (such as data from Reuters Market Data System and the Reuters Tick Capture Engine, etc.) for options pricing applications, etc. Amazon's offer to host public data sets on AWS is an initiative in this direction that makes things simpler, faster and cost effective for service users. Performance of cloud can be improved in the presence of data grid middleware enabling sharing of data among the participants in the cloud. Data grids reduce the I/O – blocking calls that an application might incur when writing to files.

In memory data grid (IMDG), distributed file systems (DFS) and distributed caching strategies are the options in this area. The availability of cloud databases such as Amazon's SimpleDB, Google's BigTable, Microsoft's SQL Server Data Services, etc., that store data as key value pairs are worth exploring as the data tier of the HPC application. This enables availability of durable and pervasive data handling mechanisms across multiple compute nodes and the ability to move workloads effectively across machines.

CHALLENGES FOR HPC ON CLOUD

Cloud computing based HPC is at a nascent stage and holds great promise as indicated in the earlier section. However, there are quite a few challenges that need to be overcome henceforth.

Virtualization Related: Some of these arise due to the basic issues pertaining to virtualization and its effect on the absolute performance that can be expected as well as the variability and instability in performance. There could be unexpected performance variations when scaling to larger number of instances and cores. Another issue is the possibility of loss of performance due to the time taken in bringing up new instances as well as ramp up in virtualized infrastructure.

Cloud Management Services: Dynamic and policy based provisioning features to ensure auto-scaling and load balancing are important to ensure reliability and expected throughput of HPC workloads. The solutions in this space are still in nascent stage with many open issues. Solutions from Amazon for its own cloud and independent solutions from vendors such as RightScale [8], 3Tera would go a long way in making cloud platform (public and private) effective for HPC.

Public Cloud Related: The second category of problems arises in the case of public clouds. There are studies that indicate inordinately high latency of large size data uploads, storage costs associated with storing large amounts of basic and derived data in the cloud. Most of the current public cloud infrastructures run using high latency network and low bandwidth interconnects. HPC clusters usually require extreme low latency and high bandwidth interconnects (such as Myrinet, Infiniband) for parallel tasks that are inter-task communication intensive and I/O intensive. Another aspect is the upload of large data sets to the cloud on internet. Uploading a terabyte of data over a 1.5Mbps T1 broadband line takes more than 80 days. Hence, offline data transfer on physical disks by courier service is to be considered. This has implications in terms of security and related issues.

Benchmarking: As mentioned in an earlier section, suitability of a cloud infrastructure for a specific HPC workload will be determined by the workload characteristics and its match with the declared and observed performance of the compute infrastructure. Published performance data and benchmarking results for both uni-processor performance such as HPC Challenge and parallel computing performance benchmarks such as the NAS PB for the cloud computing infrastructure will be necessary in matching process. Without such benchmarking related inputs the cost-benefit analysis would be a faulty one.

Security: Security of large data sets imported onto public clouds from an enterprise's internal systems is definitely a concern from security point of view. Most of the Byzantine fault tolerance issues are handled by the cloud infrastructure but the application architectures also needs to account for such faults.

Transaction and RDBMS Related: Cloud storage is mostly non-relational and most of the legacy enterprise HPC applications have data hosted in RDBMS. This disconnect could lead to obvious migration and porting issues of legacy HPC applications for cloud infrastructures. This would determine suitability of a particular HPC application for cloud and could force the architects to think of innovative options. Yahoo!'s Pig, IBM's JAQL, and Facebook's Hive, MapReduce implementations such as from Greenplum and Aster Data are efforts in this direction.

HPC ARCHITECTURES FOR THE CLOUD

Analysis of various cloud providers and other participants in the cloud ecosystem, led to a representation of the cloud based HPC application as given in Figure 1 overleaf. The architecture provides many features that are common in HPC systems and are described in the following sections.

Cloud architecture can be used for HPC workloads like scientific computations and in most cases can perform at the same level of efficiency as that provided by a dedicated grid. It has to be noted that the network interconnect between the machines may not be as fast as a dedicated grid and can cause performance degradation when the nodes share a large amount of data.

CLOUD COMPUTING AND HPC WORKLOADS

The versatility of the cloud computing platform enables its mapping with a variety of HPC workload patterns [Table 1 on page 75].

Applications that have Seasonal Workloads

Scalability at low cost is the unique selling proposition for the cloud. Consumer facing applications such as retail systems that face huge demand during festive seasons are well suited for cloud architecture. Cloud infrastructure functions as a load balancer at a high level and distributes the incoming request to one of the nodes in the cloud. We classify this workload as a single job getting

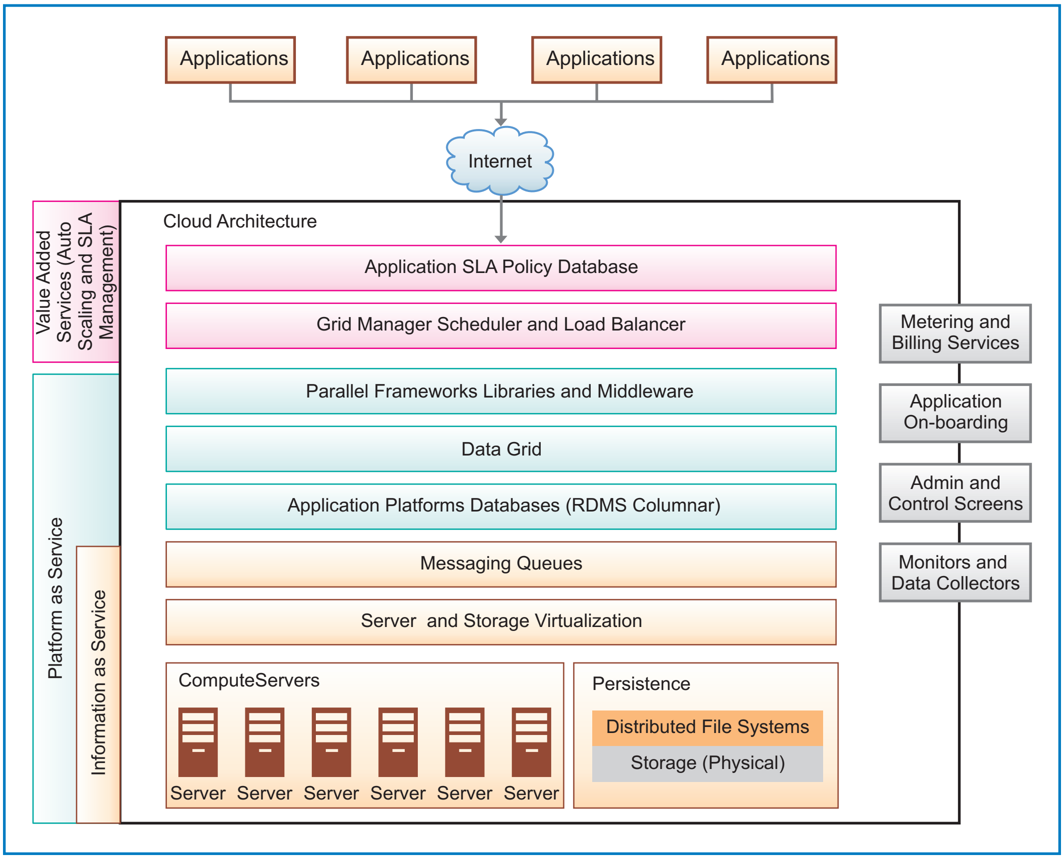

Figure 1: *Cloud Based HPC Architecture* ***Source:*** *Infosys Research*

executed in a single machine. The workload is executed by one single machine/node and hence when more jobs arrive, if more nodes are allocated, the application can easily scale and cater to the increase in demand. Such workloads are highly suitable for cloud model.

High Throughput Workloads

Applications that run against a huge data volume and that has shorter time window for execution can leverage cloud models. Such kind of workloads can be classified as single job multi machine workloads. The input can be cut in to smaller pieces and each piece can be run in one of the machine [Fig. 2 on page 76]. This effectively is data parallelism in use to get the required throughput. The performance of such workload is excellent when there is minimal data sharing between the participating nodes. Long running batch applications working on large data volumes are suitable for cloud environments.

High Performance Computing Workloads

Under this category, we have grouped application workloads as – compute intensive scientific calculations; embarrassingly parallel logics like Monte Carlo; low latency requirements for systems like trading and

Application Sterotypes	Application Characteristics	Worked Type Distribution of		Cloud Suitability	Key Factor	Example
		Job	Unit of Works			
Adaptive Systems	Systems with Seasonal Demands	Single Job run in Single Machine	1 Job corresponds to 1 Unit of Work/Unit of Work run on a Single Machine	Very High	Load Balancing (at Job Level) between the Participants	Internet facing e-commerce retail sites. tax processing system, Regulatory systems that are required to operate in response to an event
High Through-put Systems	Long Running Batch Systems	Single Job run in Multiple Machines	1 Job = Many Units of Work/ A Unit of Work run on a Single Machine. Results of Unit of Work assemble later	Very High	Data Parallelism from the Application Side Data Sharing between Machines Minimal. No Task Parallelism	Purchase order systems, updation of stock in a retail industry. Billing in telecom. Back office risk analysis batches in financial firm
High Performa-nce Data Mining	Search Engines	Single Job run in Multiple Machines	A Unit of Work is run in Many Nodes in Two Phases — Map and Reduce Phases	High	Map Reduce Algorithm Implementation - might require a Global Parallel File System	Distributed information processing, Petabyte data processing — searching for field to get a particular value
	Reduction Algorithms (Data Dependencies in Set of Records)			High		High Performance Data Analytics and mining in Telecom industry – real time information for law enforcement??-data load and retrieval on a columnar database can improve the overall throughput
High Performa-nce Computing	Compute Intensive – Scientific Calculations	Single Job run in Multiple Machines	1 Unit of Work spreads across Multiple Machine (with Varying Degrees of Data Sharing)	Medium	Performance will not be as good as a Dedicated HPC Cluster (Data + Task Parallelism)	Life science modeling for drug discovery and simulations
	Embarrassingly Parallel Logics — Monte Corlo			High	Can Leverage MPI and Open MP Libraries	Pricing application for a financial derivatives
	Low Latency Requirements for a Trading System			Still need to evolve	Machine Interconnect Speed might be Bottle	Algorithmic Trading

***Table 1:** HPC Workload Patterns* ***Source:** Infosys Research*

front office analytics. Cloud can be used for these workloads however; performance would depend on extent of data and sharing between the nodes.

For low latency applications, the current cloud architecture needs to evolve and will have to support high speed networks, have connectivity to data providers and provide infrastructure required to support complex event processing capabilities. At this juncture, these applications are suited for in house deployment.

HPC Data Analytics Frameworks

The advent of the Elastic MapReduce framework by Amazon has heralded the beginning of the availability of HPC frameworks tailored to cloud computing infrastructures. Job scheduling and resource provisioning are closely tied to the topology of the cloud infrastructure and can be optimized by the cloud provider. For example, provisioning the MapReduce cluster from the same subnet or physically proximal set of hardware can result in obvious performance gains which only the cloud provider can make

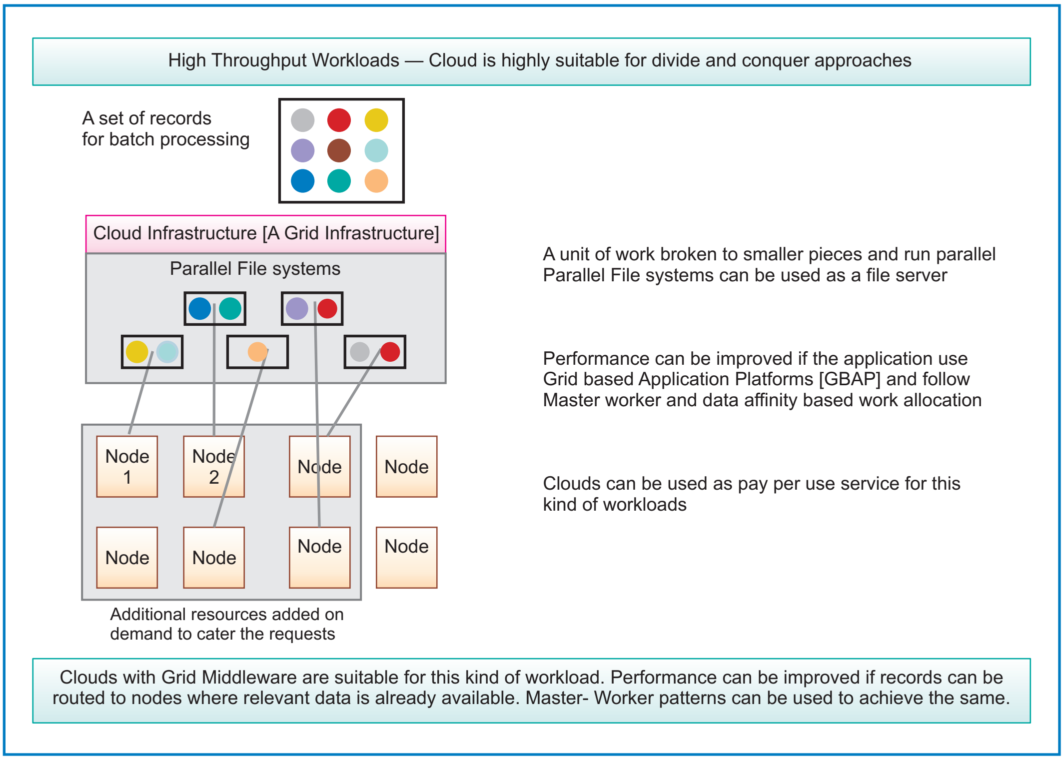

Figure 2: *Cloud for High Throughput Workload* ***Source:*** *Infosys Research*

possible [Fig. 3]. Moreover, the cloud HPC user does not have to handle the onerous tasks of setting up clusters, provisioning adequate capacity nodes (for example, high end compute node, I/O capacity node for masters in a master-slave configuration).

CONCLUSION

We foresee emergence of more mature cloud provisioning and management solutions increasing the throughput of HPC jobs through enhanced resource allocation, scheduling and reliability. Public cloud with faster inter-node interconnects, that are good for communication intensive HPC jobs, will emerge over a span of time. Specialized on-premise private clouds for HPC jobs will emerge for workloads such as extreme transaction processing, complex event processing, etc. Extensions and customization of MapReduce frameworks would emerge for different types of HPC workloads and industry verticals. Parallel computing libraries benchmarked on specific public cloud platforms would evolve. More comprehensive performance benchmarks of popular public clouds would become available. Customized HPC application stack images would become available for specific public cloud platforms easing setup of HPC applications on the cloud. Industry vertical specific SaaS HPC platforms would emerge and performance benchmarked with respect to specific public cloud platforms.

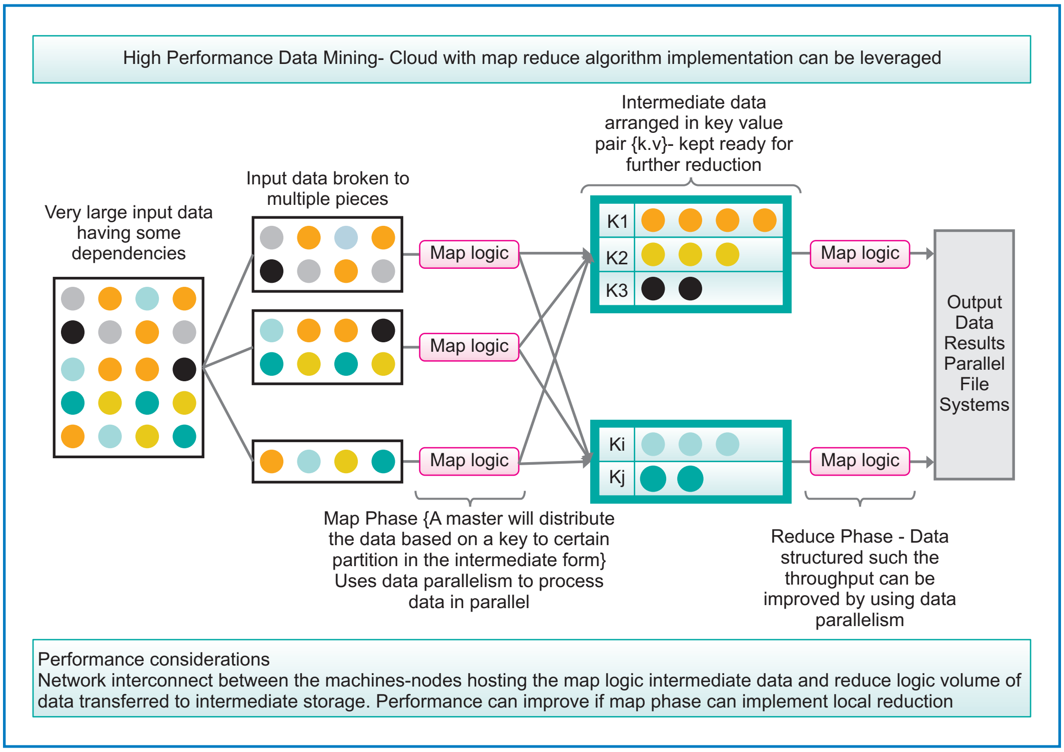

Figure 3: *Representation of Map Reduce Workloads* ***Source:*** *Infosys Research*

REFERENCES

1. Richard Walsh, HPC Directions in Financial Services, July 2008. Available at http://www.idcindia.com/events/HPC/pdf/IDC%20AP75514Q.pdf
2. Amazon Elastic Map Reduce. http://aws.amazon.com/elasticmapreduce/
3. Constantinos Evangelinos and Chris N Hill, Cloud Computing for Parallel Scientific HPC Applications: Feasibility of Running Coupled Atmosphere-Ocean Climate Models on Amazon's EC2, CCA-08 October 22–23, 2008, Chicago, IL
4. Simon Ostermann et al., An Early Performance Analysis of Cloud Computing Services for Scientific Computing, Report number PDS-2008-006, December 2008. Available at http://www.st.ewi.tudelft.nl/~iosup/PDS-2008-006.pdf
5. Michael Armbrust et al., Above the Clouds: A Berkeley View of Cloud Computing, Technical Report No UCB/EECS-2009-28. Available at http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html, February 2009, accessed May 2009
6. Krste Asanovic et al., The Landscape of Parallel Computing Research: A View from Berkeley, Technical Report No UCB/EECS-2006-183, December 2006. Available at http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-

183.html

7. Known applications of MapReduce, http://www.dbms2.com/2008/08/26/known-applications-of-mapreduce/
8. RightScale. http://www.rightscale.com/products/.

SETLabs Briefings
VOL 7 NO 7
2009

Infrastructure Management and Monitoring in the Cloud

By Kaustubh Janmejay Vaidya

Cloud computing initiative can be best driven by a stepped approach, proper planning and internal IT capability assessment

A cost optimization mandate within the business organization and the availability of cloud computing necessitates an understanding of the key aspects in managing and operating in an abstract cloud. IT management and monitoring within the local organizational periphery is more visible. What goes beyond this boundary to operate as a private cloud and further as a public cloud has its own challenges. There are issues in retaining enhanced technical visibility, monitoring and controlling, security using the right levers and tools, and on transitioning from the local-to-public-to-private level in the cloud.

The problem cannot be addressed with a single formula. It requires incremental steps within the organization that are iterative in nature, validated over a period of allotted time and those which operate in tandem with the business requirements. An attempt is made in the paper to provide an insight into working out a technical roadmap with focus on IT management and monitoring aspects while planning a cloud for the organization.

CLOUD COMPUTING FOR THE IT INFRASTRUCTURE

The cloud computing buzz has reached the nook and corner of every organization's IT arm and everyone is gearing up to get on to the bandwagon early. This technology is not relatively new, but neither has it matured for an end-to-end business functioning. There are a large number of options available in the market today and selecting the right option for one's business is a complex task due to the limited visibility of IT capabilities within the organization and also at the vendor end. There is also a mandate of effective utilization of existing IT infrastructure and avoiding further capital expenditure(capex).

Option	Type	Model	CPU	RAM	Disks	RHEL cost	VMware Support	Total Cost (USD)
1	Standalone	PowerEdge 2950	2 quad core CPU- E5410 2.33GHz 2x6MB Cache	2 GB	900 GB	Included	Not applicable	20000 approx (3 Servers)
2	Virtualized	PowerEdge 2950	2 quad core CPU- E5430 2.66GHz, 2x6MB Cache	16 GB	2700 GB	Included – 3 licenses	3 yrs	20000 Approx. (1 Server of a higher configuration)

Table 1: *Configuration of Three Standalone Servers vs Virtualized Server at Same Cost* ***Source:*** *Infosys Research*

MOVING FROM A STANDALONE TO A VIRTUALIZED INFRASTRUCTURE

For the smallest of the organizations to start, the first step is to understand the importance of moving from a standalone infrastructure to a virtualized infrastructure.

Let us consider a simple illustration on optimizing the infrastructure internally using virtualization. An organization needs infrastructure for development environment, functional testing and QA with a budget of 20000 USD for server infrastructure. At a generic level, three standalone servers of a standard configuration will be proposed for three environments.

To optimize the infrastructure, a second option of a single server with sufficient configuration can be proposed with virtualization software like VMware. Any technological change or movement should first indicate a business value.

By comparing the above indicative costs and configuration in Table 1 we note the following in Table 2.

Besides, if there are different operating systems required for two different partitions they can be accommodated on the same virtual server (e.g., Windows and Red Hat Enterprise Linux) as seen in Figure 1.

From the infrastructure monitoring

Sl. No.	Standalone Infrastructure (3 servers)	Virtualization with 1 Server
1	Limited Configuration, Scattered Capacities	Higher Configuration, Sharable Capacity
2	Limited Scalability for Servers and no scope for Processing Power Sharing between Servers	More Flexibility for Server Resource Management and Scope of Processing Power Sharing between Virtual Servers
3	Islands of Limited Configuration within Budget	Effective Higher Configuration at the Same Price within Budget
4	More Management overhead for 3 Servers	Less Management Overhead
5	More Space, Power, Cooling	Less Space, Power, Cooling
6	Green Initiatives are not served appropriately	Organizations Green Initiative Served Better

Table 2: *Comparative Analysis* ***Source:*** *Infosys Research*

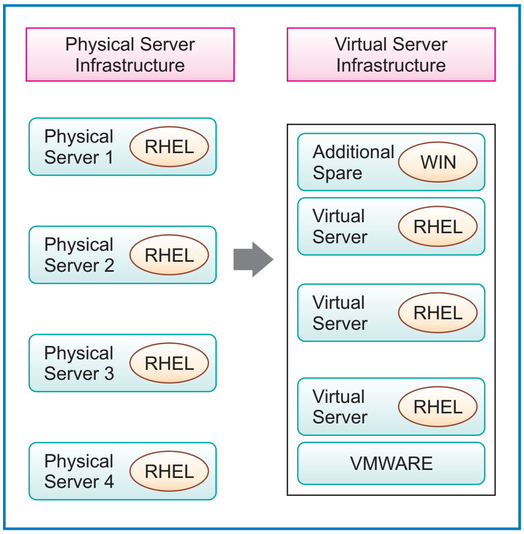

***Figure 1:** Physical Servers and Virtual Server*
***Source:** Infosys Research*

perspective, we need to monitor three server units (physical or virtual) in both the cases. However the flexibility to allocate memory to one partition on the fly is not possible in the standalone case. After monitoring in virtualized environment, if we find that the application needs more memory, we can allocate some more from the buffer capacity that is already available with us. We have the flexibility to accommodate some more environments in the same box in the future that saves cost of purchasing additional servers.

Day-to-day infrastructure management and support (backups, vendor co-ordination, OS upgrades and patching, application upgrades) is carried out by internal IT team in both the cases.

In case the infrastructure landscape is huge, the team that manages either the standalone or virtualized environment will use tools like traditional scripts or third party monitoring tools and will operate a ticketing system for resolving user support issues.

This simple example thus conveys how an IT organization can reduce infrastructure costs, optimize resources and achieve better manageability to move away from standalone dedicated infrastructure.

CREATING AN INTERNAL VIRTUALIZED INFRASTRUCTURE

A number of organizations have already made large investments in the dedicated infrastructure for multiple environments because were needed at that point in time. Now these environments with respectable numbers are either in excess or under utilized. It makes business sense to utilize the same infrastructure for new upcoming applications using the excess/spare capacities. This calls for some internal changes for unlocking these capacities by modifying the internal IT infrastructure canvas to transform it into an internal cloud, using virtualization technology.

It will help the organizations to save on new purchases and power, and thus reduce overall capex. Note that a virtual and dedicated server infrastructure may co-exist based on business criticality and organization’s overall strategy of moving to a cloud.

There are a number of prominent factors that an organization should consider and evaluate before delving into virtualization. Factors that need to be considered are –

- Business benefits in term of savings/ value delivered
- Capabilities of internal teams managing the existing infrastructure
- Flexibility in managing the virtual environment

- Groups involved in supporting the IT landscape
- Readiness to work in tandem
- Roles and responsibilities
- Policies, procedures, OLAs and SLAs affected.

They should be properly evaluated, planned and executed and need strong backing and support from the senior management. Other technological factors that will come into picture are –

- Mode of storage and data access for the application (central/NFS/SAN)
- Compatibility and interfacing of existing server commodity hardware
- Booting processes
- Network access
- Distribution and interfacing of applications
- Virtualization/cloud readiness for the application
- Ones to move and ones not to move to the virtualized infrastructure.

Due to the limited dynamic ability of the virtualized infrastructure to provision resources, there will be a significant change in the way we look at the configuration management data base (CMDB) from the perspective of application usage and updating information. Organizations should look at this process of dynamically updating CMDB. The internal ticketing process would also continue to be the same with the exception of a few variations that are a result of dynamic provisioning.

Management in the internal virtualized infrastructure would be simplified with reference to the provisioning of resources. Other factors of IT management that were done for the standalone infrastructure would be applicable here too. Rather than a localized backup, centralized backup may come into picture and will drive the organization's backup strategy.

Monitoring of the server infrastructure in case of traditional commands/scripts/third part tools would remain the same. However, the organization needs to closely monitor if there is any change in the licensing policies specific to the monitoring agents and licenses that are deployed for the virtual servers. This also holds true for all the software licenses that are installed on the virtual servers.

We note that the organization's internal virtualized infrastructure (or should we call an internal private cloud?) has a boundary and can be extended upto the limit the server resources are available in the organization. Beyond this, the IT management would face the issue of provisioning more resources or adding more servers to the internal private cloud. Thus, we understand that internal virtualized infrastructure has limitation of scaling but is initially suitable for the organization that is planning to transition in the future to private or public clouds.

CONTINUING THE JOURNEY FURTHER — THE PRIVATE CLOUD

Crossing over the boundary of the organization towards a private cloud (or call it external private cloud for location namesake) indicates that we are moving towards something that is provisioned as a service (pay-per-use) to us like a commodity like electricity or water.

It further makes sense to understand various perspectives of vendors, researchers and experts on terminologies like cloud computing, vendors of the cloud, cloud computing technology and services. Organizations also draw out key considerations before embarking on cloud services and prepare a scorecard based on the key considerations that has weight, raw score and a weighted score before deciding on a value for go, hold or no-go. The same can be considered for public cloud too but that requires greater emphasis on factors like organizational strategy, application criticality, federal norms, security aspects and compliance issues.

Private cloud can be called as a private computing facility provisioned for any organization. All the resources like server's infrastructure and cloud computing software is dedicated to the organization. Unlike the internal cloud, private cloud is extensible on demand. It has the feature of dynamic provisioning of the virtual resources. Different vendors deploy different mechanisms for quick provisioning but capacity planning aspect of the underlying systems is closely monitored. End user identity management and resource provisioning are the key factors that have impact on the security aspect [1].

One aspect to note here is that there is a separation between the person who is demanding resources and the person who is provisioning them. From an organization's perspective, although it is flexible to manage such separation, there should be an approval process set in the standard processes for dynamic provisioning of the resources. It might so happen that more resources are consumed and performance tuning aspect of the entities takes a back seat.

From infrastructure management perspective, the responsibility of managing the servers and dynamic provisioning has now shifted to the third party private vendor who manages the private cloud. The end user admin can on the fly create a server by providing the operating system, number of CPUs, memory and disk space. There are various other granular parameters that differ from vendor to vendor. The responsibility of the homegrown application tuning still lies with the IT organization whose end users work on the cloud infrastructure.

There are a few aspects like application transition to the private cloud and inter-dependency of the applications that should be thoroughly tested before transitioning. There are vendors who have a process defined as to how data should be transitioned to the cloud, scheduling the timeframes for backing up the data in the cloud, scheduling of adhoc backups and restoration options.

From infrastructure monitoring perspective, the standard tools that the organization uses like scripting can be used. Vendors also provide interfaces to display the standard monitoring parameters like CPU, disk space utilization and processes. Third party vendors and cloud vendors also are in the fray to provide the cloud performance parameters. The web service Amazon CloudWatch that tracks and monitors Amazons Elastic Compute Cloud (EC2) service provides real time monitoring by using web service APIs or command line tools for cloud resources with demand pattern parameters like resource utilization, operational performance, and CPU utilization, disk reads and writes, and network utilization [2].

Various aspects like vendor dependency, switching flexibility from one vendor to the other, business continuity and disaster recovery capabilities of the vendor, application and software support, stability, security aspects, compliance, SLA provisioning should be given consideration. There should be a process defined for integration of the vendors ticketing system with the organizations internal one in a seamless way so that call handling takes place smoothly and SLA violations get recorded and reported accurately.

FINAL DESTINATION – THE PUBLIC CLOUD

There are four key steps to cloud adoption strategy - access, validate, prepare and execute that help the CIOs integrate cloud computing with the IT strategy of a firm. Understanding the key set of activities pertaining to infrastructure, timeframes and challenges associated to move further would determine how early and successfully the organization becomes cloud ready.

Support for the organization application landscape and if they are cloud-deployable or cloud ready in the public domain should be evaluated. Here, although the technical infrastructure continues to remain like the private cloud, it is not a dedicated one but a shared one wherein the cloud space is shared by a number of organization's and their end users. The underlying infrastructure that was visible to the end user gets limited to a set of virtualized sets that can be monitored and tasks can be scheduled remotely.

Data security measures become prominent in the public cloud and necessitate data encryption and security measures to be deployed for flow of traffic across the network. All the features provided in the private cloud like creating the server on the fly, dynamic provisioning of resources, scheduling of the backups, restoring a earlier backed up snapshot are available in the public cloud also. Fluctuating and unpredictable load patterns are also factors that decide how we should scale up and down in the elastic cloud. A hybrid cloud infrastructure (mix of private and public cloud and local virtualized infrastructure) can also exist.

The journey from a standalone infrastructure to the public cloud is shown in Figure 2. Note the reduction in the periphery and the infrastructure landscape within an organization reduce as we move towards the public cloud. This surely presents a challenge to handle IT Service Management (ITSM) processes.

As per the standard ITSM support processes, the users raise a ticket in the ticketing system as illustrated in Figure 3 [3]. The ticket appropriately gets processed at the client system and is raised as an issue based on the severity level (L1, L2 or L3) in the cloud vendors ticketing system. Note that the users and the administrators access the resources in the cloud on the encrypted network as a part of security and compliance. Based on the levels, the ticket traverses through the three levels. User does not have a visibility beyond this since underlying hardware and resources provisioned are managed by the cloud vendor. Based on the dynamic provision requested, usage and the SLAs adhered as per agreed terms between the cloud vendor and the client, billing will get calculated. Resource usage statistics are visible to the administrator of the client.

From the system administrators perspective, the CMDB that holds the server configurations may change frequently (scale up or scale down) based on dynamic provisioning available. The process whether to involve the change advisory board (CAB) for every provision needs to be given a thought based on

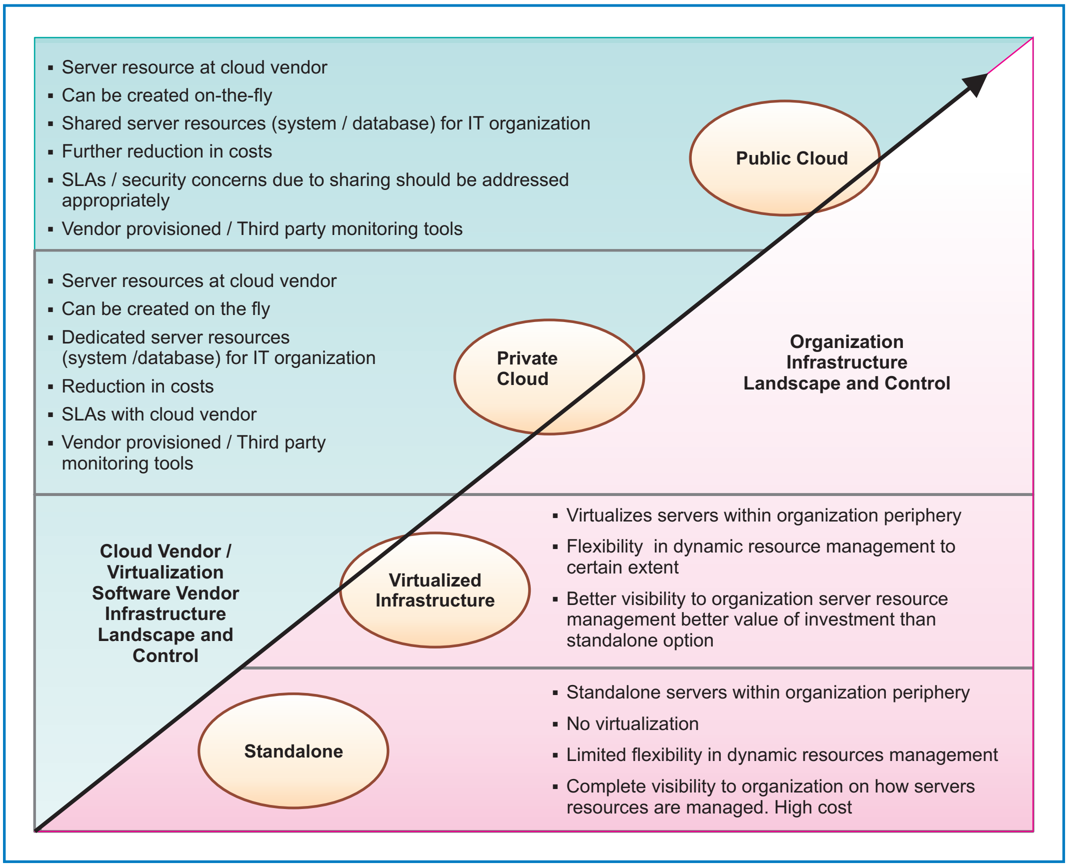

Figure 2: *Infrastructure Journey from the Standalone to the Public Cloud*

Source: *Infosys Research*

frequency and variation on higher side since it is directly linked to costs.

CONTROL IN THE CLOUD USING MONITORING TOOLS

There are internal monitoring mechanisms for cloud provisioned by the cloud vendor by a third party. However, there are a number of open source tools available which can be evaluated for usage. There are various considerations like operating systems support, support on thin clients (being lightweight) since monitoring has to be done across the network or across internet and installing them as a plug-in should not hamper the performance of the application. Some of them are Hyperic (prominently visible), Nagios, Zennos and others.

The monitoring product Hyperic HQ equips IT organizations with the ability to securely monitor internal IT infrastructure as well the cloud services [4]. HQ's ability to automatically discover, monitor and manage software services, regardless of type or location, enables organizations to easily create a unified view of the performance and health of their applications is a definite plus.

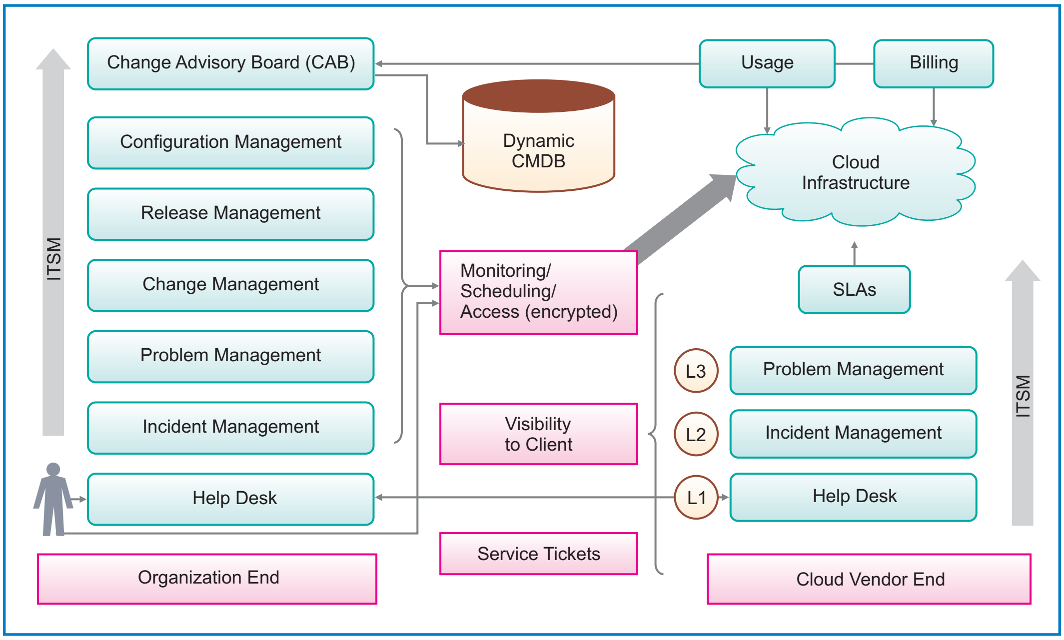

Figure 3: *Cloud Vendor and Organization ITSM processes* ***Source:*** *Infosys Research*

An open source monitoring tool Nagios is a powerful comprehensive monitoring system that enables organizations to identify and resolve IT infrastructure problems before they strike and offer visibility through web interface, alerts reporting and multi-tenant (access based view) capabilities [5].

Zennos supports full operational awareness by monitoring the entire IT infrastructure through agentless collection and control. Key features are autodiscovery, IT configuration database, alerting, fault management, availability and performance reporting and a host of other features [6].

Organization can select appropriate cloud vendor monitoring option or a third party tool or a combination of both that would serve best for comprehensive monitoring.

All the standard utilization and monitoring parameters like CPU utilization, memory utilization and disk space utilization are available in most of the products. Cloud vendors also do provide an administrator interface and the usage/billing statistics. The administrator can dynamically provision/remove the dynamic server resource created in the cloud. Clarity should be obtained on business continuity and disaster recovery provisioning, country of data storage (few countries require critical data to be stored within country), provisioning clause for switching to a different vendor.

Once the final stage is reached, the organization is free to concentrate on its core business competencies and outsource a large part of its IT operations to the cloud vendor. The internal IT hardware/license management overheads will reduce. Does this mean that the local IT organization will cease to exist? Not really. IT organization will have to prominently play a role in the following:

- Transitioning of the existing application to and from the cloud
- Ensuring data management and security aspects
- Planning and developing new application and monitoring existing applications and virtual servers
- Handle the dynamic CMDB for the cloud which now gets tied to the revenue outflow
- Handling complex interfacing handling in the cloud
- Monitoring cloud resources usage statistics and optimizing resource usage that is extensible on the fly
- Responsibility of ensuring a balance between resource demands and spending
- Handling a hybrid infrastructure mix (public, private and internal virtual infrastructure)
- Understand portability of the data being hosted and alternate plans on getting off the cloud if service provider shuts business [7].

CONCLUSION

Moving to a cloud requires small incremental steps, proper planning, willingness from teams for adoption and a very strong senior management support. From a standalone infrastructure an organization can transition to a hybrid one targeting to completely be on cloud in the future. Moving towards the cloud reduces the day-to-day IT operational management issues since they shift towards the vendors end. Yet the complexities of demand-cost management, application interfacing, security and process management and overall billing rests with the internal IT organization.

As the cloud computing adoption unfolds, an organization may also prefer to have few cloud vendors like regular IT vendors that it has today. ITSM processes of an organization should be drafted considering the dynamic set of provisioning resources and monitoring tools should be adopted as per application compatibility. Future may also necessitate interfacing between two different cloud vendors rather than operating in silos and having vendor dependency. Most of the vendors have their own offerings today and standards for cloud computing are taking shape with features that are suitable to different set of organizations. There are initiatives by organizations and groups in this direction to have defined processes and inter-operatibility between the cloud vendors and this will largely shape the adoption of the cloud in the years to come.

REFERENCES

1. Bernard Golden, Defining Private Clouds, Part One, CIO. Available at http://www.cio.com.au/article/304190/defining_private_clouds_part_one
2. Amazon Cloudwatch. Available at http://aws.amazon.com/cloudwatch/
3. IT Service Management. Available at http://www.itsm.info/ITSM.htm
4. Gartner Names Hyperic "Cool Vendor" in Latest Research Report. Available at http://www.hyperic.com/news/releases/hyperic-named-cool-vendor-by-gartner.html
5. http://www.nagios.org/about/features/
6. http://www.zenoss.com/product/network-management
7. Mario Moreira, Infrastructure - on Premises or in the Clouds? Agile Journal,

April 2009. Available at http://www.agilejournal.com/articles/columns/column-articles/1469-infrastructure-on-premises-or-in-the-clouds.

Cloud Computing — A Disruptive Technology

In a freewheeling discussion on multiple facets of cloud as a technology, business trend and other related issues Dr. Srinivas Padmanabhuni poses a set of questions to Jitendra Pal Thethi, Principal, Microsoft Technology Centre and Raghavan Subramanian, AVP and Head of Cloud Computing CoE at Infosys.

Srinivas: Raghu, what are your thoughts on cloud as a disruptive technology? How is it related to recent trends of grid, virtualization and SOA?

Raghavan: The technologies that make cloud possible have been knocking at the doors of the research community and the common man for quite some time now. Cloud is not new to scientific research where there is a need for massive computational and storage capabilities. Cloud is not new to consumer software business either given the fact that Google, Amazon, Facebook and several others have been early adopters of cloud. However what is new here are the technologies enabling these cloud capabilities now available to enterprises small and large alike. Viewed this way cloud is both evolutionary and disruptive. Evolutionary because organizations can try to aggregate their own computational, storage and network needs and use the abstraction provided by virtualization to reap cost, scalability, time-to-market and several other benefits. Disruptive because public cloud opens up the possibility to a lot of great software ideas that remained as mere ideas so far to be realized now as software and disrupt well-entrenched software. There will also be a lot of disruption on how enterprises want to develop new applications and maintain and enhance their existing applications. Infrastructure engineers will advocate IaaS, developers will push the evolution of PaaS, while business-stakeholders might be keen on SaaS. Depending on an organization's IT culture, its IT portfolio could be a mix of these different cloud delivery models. Disruption can also happen if in the enterprise world - a new application like Facebook becomes the preferred platform to write enterprise or industry specific applications. The possibilities are limitless but the hype generated is working against it.

Srinivas: Jitendra, how do you think cloud can help enterprises in contrast to the popular SMB segment?

Jitendra: The value proposition of cloud applies more to an enterprise segment than to an SMB in many different ways. Cloud certainly helps to translate the investments from capital expense to operating expense which is becoming a good financial lever for enterprises. More importantly, with a cloud based infrastructure the speed at which a solution can be brought to market is radically fast. With cloud, the lead time to get the infrastructure in place is translated from weeks to minutes that further help organizations in getting their new solutions and ideas roll out quickly. This small change is a huge catalyst to innovation in an enterprise. Organizations can quickly pilot a solution, measure its effectiveness and use it create competitive advantage.

Srinivas: What are the business opportunities from an IT services perspective?

Raghavan: Enterprises always look up to IT service providers for trusted opinions on anything new. Enterprises would like consultants to help them with their cloud strategy and implementation. In order to do this IT service providers must be able to understand an enterprise's organizational culture - the ecosystem in which it operates and its IT portfolio before adopting cloud strategy. IT service providers also need to have a firm grasp on the various cloud offerings, the potential and the limitations of the technology. Enterprises need to understand the issues one can face and acquaint themselves with limitations and workarounds to overcome them. IT service providers have the role of tempering the marketing hype and provide a balanced and unbiased view of the relevance of cloud and its adoption for an enterprise. IT service providers have the role of helping enterprises migrate relevant legacy applications and develop new applications using cloud. IT service providers must also help to identify the SaaS potential of some of the custom applications within an enterprise and help them in re-architecting a custom application into a multi-tenant, customizable, pay-as-you-go metered SaaS application.

Srinivas: Any thoughts on relative penetration of cloud in different verticals?

Jitendra: The penetration of the cloud in different verticals is truly workloads driven. Verticals like life sciences where there is a large data processing and computation involved are leveraging cloud for scaled-out architecture. Verticals like banking and financial industries are coming toward approaching cloud from a cost structure and resource optimization perspective. In the area of manufacturing and retail the emphasis is more towards leveraging the publically available infrastructure and scenarios of partner integration.

Srinivas: What have been the typical business drivers considering the early adopters we have worked with?

Raghavan: Once again, the term cloud is broad as it includes public, private, community and hybrid models of cloud delivery. From the early adopters of public cloud we see a few patterns emerge. Testing and other environments that are not required through the year are being created on a need basis on the public cloud. One-off computational needs, like data crunching, archiving, etc., are also happening on public clouds. Some of the new applications with less emphasis on security are getting developed in a public cloud. On the private cloud front it is primarily to do with virtualization of data centers. So the applications are being moved away from deployment models where they used to run on dedicated hardware. This further means that independent software vendors (ISV) support for virtualization is becoming an important factor for the future of ISVs. The business drivers for the public cloud are time-to-market, pay-as-you use, less in-house staff, opex instead of capex, standardization of IT and the ability for IT department to embrace an imminent future paradigm among various other things. For private or internal cloud the business drivers are very similar to virtualization i.e., abstraction of hardware resources, elastic scalability (limited), improved utilization of hardware, etc.

Srinivas: What are the business and technology inhibitors/showstoppers when it comes to cloud adoption?

Jitendra: Most of the business challenges are associated with the lack of trust on an externalized infrastructure. Data privacy, security and compliance are pressures which do not allow use of a cloud infrastructure crossing global boundaries. Ambiguity on SLA definition and lack of end-to-end SLAs for business services do not leave businesses to take a calculated risk to grab the cost advantages. Lastly, the penalties agreed upon by the cloud providers for SLA breach is disproportionate to business loss caused due to a possible outage.

From a technology standpoint, the heterogeneity of managing different cloud vendors lead to integration challenges and management overheads. Today, only a handful of software licenses from vendors are available on a pay-by-use model and hence are not available as part of cloud provider. Third, many cloud providers impose new models of development like use of column database rather than relational database that further creates a lot of ambiguity and design problems. Lastly, the migration from an on-premises application to cloud is not simple with the existence of multiple vendors that make the overall migration cost and time high and therefore not much practicable.

Srinivas: What are your thoughts on cloud standardization and interoperability?

Raghavan: Let us look at the standardization needs at the IaaS layer, since PaaS and SaaS bring a different set of problems to the picture. IaaS needs a standard to ensure that an application and its dependencies can be extracted and abstracted from its current running environment and deployed on to any target virtual machine, private or public. This problem arises because cloud-players are providing point-solutions that address this problem only in the narrow context of their self-interest. Open virtualization format (OVF) is trying to address this problem through standards, while companies like AppZero are addressing this through their technologies (Virtual Application Appliance). The second problem comes from the proprietary storage formats used by cloud-players that would mean either vendor lock-in or the need to write data extraction logic for vendor specific storage technologies. This problem has to be solved either through standards or by innovative technologies.

Srinivas: What are the key takeways for systems integrators?

Raghavan: System integrators with a close proximity to customers have to play a crucial role in taking the benefit of cloud computing to the enterprise customers. However their approach to cloud computing will determine if they continue to retain their proximity to the enterprises. Let me illustrate this point with one example. Consider SLAs. System integrators have to decide whether they will bundle the infra-offering and front the SLAs to provide a single-point-of-contact for customers or whether they partner with infra-players and carve out different realms of SLAs thereby leading the end-customer to deal with multiple players. There are multiple ways in which system integrators can retain and increase their proximity with their customers by making software (whether it is running on IaaS or PaaS infrastructure or if it were already consumed as SaaS) the focus of the interaction with the enterprise customers. There are new models like process-as-a-service (PraaS) where process services are bundled along with software to provide a higher level of abstraction to customers.

About the Discussants

Jitendra Pal Thethi *is a Principal Architect with Infosys and anchors presales activities for Infosys solutions and IP built on disruptive technologies in the areas of Cloud Computing, Collaboration, Data Virtualization, Call Center Optimization and Mobility.*

Jitendra has over fourteen years of experience in IT industry as a Solution Architect and Technology consultant. He can be reached at Jitendra_Thethi@ infosys.com.

Raghavan Subramanian *is an Associate Vice President and the head of Infosys Cloud Computing center-of-excellence.*

Raghu was in the recent past with the Infosys SaaS unit where he was product manager for SaaS social media platform and program manager for Infynity, the Infosys Social Network.

He advises many clients in the areas of information security, Enterprise Application Integration and Object-oriented programming.

He has co-authored two patent applications, has written several white-papers and articles in peer-reviewed journals and magazines.

Srinivas Padmanabhuni PhD *is a Principal Research Scientist, leading the Software Engineering and Distributed Computing Labs in SETLabs, Infosys. He has several publications to his credit including a book, several book chapters and conference publications. His research interests include distributed computing, global software development, service oriented software engineering and enterprise architecture. He can be reached at srinivas_p@ infosys.com.*

Index

Application Development and Maintenance, also ADM 59
Agents 4, 6-7, 9, 67, 82
    Delivery 4, 6-7, 9
    Management 67
    Monitoring 82
Business Productivity Online, also BPO 16
Change Advisory Board, also CAB 84, 86
Cloud Computing Interoperability Forum, also CCIF 20-21, 24-25
Component 21, 30, 36-37, 46, 51
    Administrative 46
    Agent 21
    Analyzer 37
    Application 30
    Authenticator 46
    Disseminator 37
    IDM 51
    Processes 36
    Profiles 30
    Provisioning 46
    Registry 37
Configuration Management Database, also CMDB 82, 84, 86-87
Content Delivery Network, also CDN 17, 61
Customer Relationship Management, also CRM 6, 17, 63
Data 5-7, 13, 17, 19, 23, 25, 28, 32-35, 41, 43, 62, 64, 66-67, 84, 91
    Application 35
    Assimilation 34
    Center 5-7, 13, 19, 25, 28, 32-33, 41, 43, 62, 64, 66
    Collection 34
    Encryption 84
    Monitoring 33
    Ownership 67
    Privacy 91
    Security 67
    Service 17
    Volumes 23
Dense Linear Algebra, also DLA 70
Elastic Compute Cloud, also EC2 16, 21, 25, 63, 66, 77, 83
Extensible Messaging and Presence Protocol, also XMPP 20
Extension Protocol, also XEP 20
Identity Management, also IDM 45-53, 83
Independent Software Vendor, also ISV 43, 55-68, 91
Infrastructure as a Service, also IaaS 20, 24, 45, 47, 53, 89, 91-92
IT Service Management, also ITSM 84, 86-87
Layer 14-15, 17-18, 21-24, 50
    Abstraction 21
    Authentication 50
    Cloud Application 17-18
    Cloud Infrastructure 17-18
    Cloud Platform 17-18
    Orchestration 21-24
    UCI 21
    Virtualization 14-15
Line of Business, also LOB 70
Manager 35-37, 39
    Metering and Accounting, also MAM 35-37
    Computation and Storage 35-36
    Subscription Registry 35-37
    WSN Virtualization 35-36, 39
Non-functional Requirements, also NFR 25
OpenID 5, 47-48, 50, 52-53

Platform as a Service, also PaaS 13, 20, 24, 47-48, 89, 91-92
Quality of service, also QoS 20, 28-29, 56, 59, 63
Sales Performance Management, also SPM 65
Sensor Event as a Service, also SEaaS 34-38
Service Administered Markup Language, also SAML 46-52
Service Provisioning Markup Language, also SPML 46, 52-53
Service Level Agreement, also SLA 15, 27, 32, 37, 46, 63, 67, 71, 74, 83
Software as a Service, also SaaS 4, 17, 20, 24, 34-38, 45, 47, 48, 55, 58, 60-63, 65, 67, 76, 89-92
Total Cost of Ownership, also TCO 3, 9, 56, 59-60, 62
Unified Cloud Interface, also UCI 20-21, 25
Wireless Sensor Network, also WSN 33-39

SETLabs Briefings

BUSINESS INNOVATION through TECHNOLOGY

How to Reach Us:
Email:
SETLabsBriefings@infosys.com
Phone:
+91-40-67048455
Post:
SETLabs Briefings,
B-19, Infosys Technologies Ltd.
Electronics City, Hosur Road,
Bangalore 560100, India

Subscription:
setlabsbriefings@infosys.com

Rights, Permission, Licensing and Reprints:
praveen_malla@infosys.com

Editorial Office: *SETLabs Briefings, B-19, Infosys Technologies Ltd.*
Electronics City, Hosur Road, Bangalore 560100, India
Email: SetlabsBriefings@infosys.com http://www.infosys.com/setlabs-briefings

SETLabs Briefings is a journal published by Infosys' Software Engineering & Technology Labs (SETLabs) with the objective of offering fresh perspectives on boardroom business technology. The publication aims at becoming the most sought after source for thought leading, strategic and experiential insights on business technology management.

SETLabs is an important part of Infosys' commitment to leadership in innovation using technology. SETLabs anticipates and assesses the evolution of technology and its impact on businesses and enables Infosys to constantly synthesize what it learns and catalyze technology enabled business transformation and thus assume leadership in providing best of breed solutions to clients across the globe. This is achieved through research supported by state-of-the-art labs and collaboration with industry leaders.

Infosys Technologies Ltd (NASDAQ: INFY) defines, designs and delivers IT-enabled business solutions that help Global 2000 companies win in a flat world. These solutions focus on providing strategic differentiation and operational superiority to clients. Infosys creates these solutions for its clients by leveraging its domain and business expertise along with a complete range of services. With Infosys, clients are assured of a transparent business partner, world-class processes, speed of execution and the power to stretch their IT budget by leveraging the Global Delivery Model that Infosys pioneered. To find out how Infosys can help businesses achieve competitive advantage, visit www.infosys.com or send an email to infosys@infosys.com

NOTES

Authors featured in this issue

ADI MALLIKARJUNA REDDY V
Adi Mallikarjuna Reddy V is a Junior Research Associate in SETLabs, Infosys. He can be contacted at Adi_Vanteddu@infosys.com.

AJIT MHAISKAR
Ajit Mhaiskar is a Principal Technology Architect with the Manufacturing business unit of Infosys. He can be reached at Ajit_Mhaiskar@infosys.com.

AMIT WASUDEO GAWANDE
Amit Wasudeo Gawande is a Technology Lead at SETLabs, Infosys. He can be contacted at Amit_Gawande@infosys.com.

ANJANEYULU PASALA
Anjaneyulu Pasala PhD is a Senior Research Scientist at SETLabs, Infosys. He can be reached at Anjaneyulu_Pasala@infosys.com.

ANU GOPALAKRISHNAN
Anu Gopalakrishnan was a Product Technical Architect in SETLabs, Infosys.

ASHEESH CHADDHA
Asheesh Chaddha is a Project Manager with Performance Engineering and Enhancement practice, Infosys. He can be reached at Asheesh_Chaddha@infosys.com.

ASHUTOSH AGARWAL
Ashutosh Agarwal is a Senior Project Manager and certified PMP with Infosys. He can be contacted at ashutosha@infosys.com.

A V PARAMESWARAN
A V Parameswaran is a Senior Technology Architect with Infosys Performance Engineering and Enhancement practice. He can be reached at Parameswaran_AV@infosys.com

BHAVIN JAYANTILAL RAICHURA
Bhavin Jayantilal Raichura is a Senior Technology Architect with the Manufacturing unit of Infosys. He can be contacted at Bhavin_Raichura@infosys.com.

DEEPAK JOHN
Deepak John is a CRM Consultant with the Enterprise Solutions business unit at Infosys. He has a keen interest in emerging technologies and can be reached at deepak_john@infosys.com.

GANESAN PANDURANGAN
Ganesan Pandurangan is a Senior Technology Architect working with System Integration unit of Infosys. He can be contacted at Ganesan_Pandurangan@infosys.com.

KAUSTUBH VAIDYA
Kaustubh Vaidya is a Project Manager for shared database services with the Infrastructure Management Services unit of Infosys. He can be reached at Kaustubh_Vaidya@infosys.com.

KUMAR PADMANABH
Kumar Padmanabh PhD is a Research Scientist and leads the wireless sensor networking (WSN) research group at SETLabs, Infosys. He can be contacted at Kumar_Padmanabh@infosys.com

NIDHI TIWARI
Nidhi Tiwari is a Senior Technical Architect with SETLabs, Infosys. She can be reached at nidhi_tiwari@infosys.com.

RAHUL BAKHSHI
Rahul Bakhshi is a Consultant with the Enterprise Solutions business unit of Infosys. He can be contacted at rahulsantosh_b@infosys.com.

SHYAM KUMAR DODDAVULA
Shyam Kumar Doddavula is a Principal Technology Architect at SETLabs, Infosys. He can be reached at shyamkumar_d@infosys.com.

SIVA PRASAD KATRU
Siva Prasad Katru is a Junior Research Associate in SETLabs. He can be contacted at SivaPrasad_Katru@infosys.com.

SRINIVAS PADMANABHUNI
Srinivas Padmanabhuni PhD is a Principal Research Scientist in SETLabs, Infosys. He can be reached at srinivas_p@infosys.com.

SUDEEP MALLICK
Sudeep Mallick PhD is a Principal Research Scientist with Infosys SETLabs. He can be contacted at sudeepm@infosys.com.

SUMIT KUMAR BOSE
Sumit Kumar Bose PhD is a Senior Research Associate with Distributed Computing Lab of SETLabs. He can be reached at Sumit_Bose@infosys.com.

Subu Goparaju
Vice President
and Head of SETLabs

"At SETLabs, we constantly look for opportunities to leverage technology while creating and implementing innovative business solutions for our clients. As part of this quest, we develop engineering methodologies that help Infosys implement these solutions right first time and every time."

For information on obtaining additional copies, reprinting or translating articles, and all other correspondence, please contact:

Telephone : +91-40-67048455

Email: SetlabsBriefings@infosys.com

Infosys®

POWERED BY INTELLECT
DRIVEN BY VALUES