Capacity Management for Cloud Computing

Chris Molloy
Distinguished Engineer
Member, IBM Academy of Technology

October 2009
Is a cloud like touching an elephant?
Gartner defines cloud computing as “a style of computing where massively scalable (and elastic) IT-related capabilities are provided as a service to multiple customers using Internet technologies.”

Source: Gartner 10/13/08
Cloud Computing definitions are varying, but a common set of attributes can be identified.

IT Customers:
- Flexible pricing
- Outsourced, on demand provisioning
- Unlimited scaling
- SW developer platform
- Flexible

Press:
- Pay by consumption
- Lower costs
- On demand provisioning
- Grid and SaaS combination
- Efficient infrastructure
- Simple and easy

IT Analysts:
- Variable pricing
- No long term commitments
- Hosted, on demand provisioning
- Massive, elastic scaling
- Standard Internet technology
- Abstracted infrastructure
- Service-oriented

Financial Analysts:
- Utility pricing
- Multi-core chips
- Hosted, a-a-s provisioning
- Parallel, on demand processing
- Scalable
- Virtualized, efficient infrastructure
- Flexible

Common Attributes of Clouds
- Flexible pricing
- Elastic scaling
- Rapid provisioning
- Advanced virtualization

Frequently Cited Examples
- Amazon Compute and Storage Services
- Google App Services
- Salesforce App Exchange

Source: IBM Corporate Strategy analysis of MI, PR, AR and VCG compilations
Cloud Computing – A flexible, efficient delivery model for business process, application and IT infrastructure services

**Private Cloud**
- Client owned and managed
- Access limited to client and its partner network
- Drives efficiency, standardization and best practices while retaining greater customization and control

**Public Cloud**
- Service provider owned and managed
- Access by subscription
- Delivers select set of standardized business process, application and/or infrastructure services on a flexible price per use basis

**Cloud Services**
- Customization
- Efficiency
- Security and Privacy
- Availability

**Cloud Computing Model**
- Standardization
- Capital Preservation
- Flexibility
- Time to Deploy
Why Clouds? New Economics!

Cloud Computing Benefits
- Reduced costs
- Rapid innovation
- Faster time-to-market
- Acceleration of competitiveness and differentiation
- Improved customer service

Economies of Op-Ex

Virtualization

Economies of Cap-Ex

Private clouds

Public clouds
Decide the right mix for your enterprise

**Delivery Models**

- Off Premises
  - Shared
  - Dedicated
- On Premises
  - Utility
  - Off Premises

**Managed Operations**

**Financial Models**

- On Premises
  - Fixed
- Mixed
- Variable
- Public Cloud Services
- Private Cloud Services
Capacity Managers Need to Understand Cloud Workloads to Assess Impact on Performance

...as workloads are migrating accordingly. For example:

- Mission Critical
- Packaged Apps
- High Compliancy

- Test Systems
- Pre-production
- Developer Platform

- Variable Storage
- Software as a Service
- Web Hosting
A Spectrum of Potential Web Platform Service Providers*

* Illustration only, not a complete list of vendors. Some providers may provide services in other areas than those shown. Not all providers have the same level of support for cloud/Web architectures.

Source: Gartner Group
Many “*-as-a-Service” models are being discussed with 3 major categories are emerging.

- **Infrastructure-as-a-Service**
  - Shared virtualized, dynamic provisioning
  - Servers, Networking, Data Center Fabric, Storage

- **Platform-as-a-Service**
  - Development Platform as a Service
  - Middleware, Web 2.0 Application Runtime, Java Runtime

- **Software-as-a-Service**
  - Collaboration, Industry Applications
  - CRM/ERP/HR, Business Processes

Examples:
- IBM
- Adaptive Infrastructure
- Google Code
- Amazon Web Services
- Windows Azure
- Salesforce
- LotusLive

*Adapted from IBM's Cloud Computing on Demand.
Summary: What is Cloud Computing?

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

An infrastructure management and services delivery methodology
- Cloud computing is a way of **managing** large numbers of highly **virtualized resources** such that, from a management perspective, they resemble a single large resource. This can then be used to deliver services with **elastic scaling**.
Obstacles to Cloud Computing

Greatest Concerns Surrounding Cloud Adoption at Your Company

- Security: 45%
- Integration with existing systems: 26%
- Loss of control over data: 26%
- Availability concerns: 25%
- Performance issues: 24%
- IT governance issues: 19%
- Regulatory/compliance concerns: 19%
- Dissatisfaction with vendor offerings/pricing: 12%
- Ability to bring systems back in-house: 11%
- Lack of customization opportunities: 11%
- Measuring ROI: 11%
- Not sure: 7%
- Other: 6%

*Respondents selected up to three criteria.
SOURCE: CIO Research
## Drivers and Obstacles to Cloud Computing

<table>
<thead>
<tr>
<th>Drivers for Cloud Computing Adoption</th>
<th>Obstacles to Cloud Computing Adoption</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology Advancement</strong></td>
<td><strong>Adoption</strong></td>
</tr>
<tr>
<td>Web 2.0</td>
<td>Availability of Cloud Services</td>
</tr>
<tr>
<td>Mashups</td>
<td>Data Lock-In</td>
</tr>
<tr>
<td>Virtualization</td>
<td>Data Confidentiality &amp; Auditability</td>
</tr>
<tr>
<td>Increased bandwidth and processing power</td>
<td></td>
</tr>
<tr>
<td>Reduce fixed costs</td>
<td></td>
</tr>
<tr>
<td>Improve utilization</td>
<td>Growth</td>
</tr>
<tr>
<td>Increase responsiveness</td>
<td>Data Transfer Bottlenecks</td>
</tr>
<tr>
<td>Optimize investments</td>
<td>Performance Unpredictability</td>
</tr>
<tr>
<td>Analytics</td>
<td>Scalable Storage</td>
</tr>
<tr>
<td>Compute intensive applications</td>
<td>Bugs in Large Distributed Systems</td>
</tr>
<tr>
<td>Storage intensive applications</td>
<td>Scaling Quickly</td>
</tr>
<tr>
<td><strong>Demand for IT improvement</strong></td>
<td><strong>Policy &amp; Business</strong></td>
</tr>
<tr>
<td><strong>Business Need for New Applications</strong></td>
<td><strong>Software Licensing</strong></td>
</tr>
<tr>
<td>Optimization</td>
<td></td>
</tr>
<tr>
<td>Analytics</td>
<td></td>
</tr>
<tr>
<td>Compute intensive applications</td>
<td></td>
</tr>
<tr>
<td>Storage intensive applications</td>
<td></td>
</tr>
</tbody>
</table>

Sources:
- IBM Corporate Strategy & Market Intelligence.
Capacity planners need to incorporate Internet based bandwidth, latency, and jitter as part of their arrival rate, distribution, and service time metrics into the capacity plan.

- **Yellow**
  - Asynchronous = 2000 km
- **Green**
  - Real-time Tolerant = 400 km
- **Red**
  - Real-time Intolerant = 50 km
The Capacity Management Process

**Inputs**
- Technology
- SLAs, SLRs and Service Catalogue
- Business Plans and Strategy
- IS, IT Plans and Strategy
- Business requirements and volumes
- Operational schedules
- Deployment and Development plans and programmes
- Forward Schedule of Change
- Incidents & Problems
- Service reviews
- SLA breaches
- Financial Plans
- Budgets

**Sub-processes**

**Business Capacity Management:**
- trend, forecast, model, prototype, size and document future business requirements

**Service Capacity Management:**
- monitor, analyse, tune and report on service performance, establish baselines and profiles of use of services, manage demand for services

**Resource Capacity Management:**
- monitor, analyse, run and report on the utilisation of components, establish baselines and profiles of use of components

**Outputs**
- Capacity Plan
- Capacity Database
- Baselines and profiles
- Thresholds and alarms
- Capacity reports (regular, ad hoc and exception)
- SLA and SLR recommendations
- Costing and charging recommendations
- Proactive changes and service improvements
- Revised operational schedule
- Effectiveness reviews
- Audit reports

Version 3 Update: Component verses Resource Capacity Management
Cloud Implications on ITIL Capacity Management

<table>
<thead>
<tr>
<th>Sub-Process</th>
<th>Infrastructure as a Service</th>
<th>Platform as a Service</th>
<th>Software as a Service</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amazon</td>
<td>Azure</td>
<td>Sales Force</td>
</tr>
<tr>
<td>Business Capacity Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Capacity Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component Capacity Management</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Considerations
- Performance Service Levels
- Workload Analysis
- Processor and I/O Analysis
- Training
- Processes and Procedures
- Tools
The Future - Dynamic Load Sharing Between Clouds
Will You Be Ready for the Capacity Management Impact?

Off-load Dev, workload to public cloud

Move work to follow the moon

To make room for Bubble Workload (e.g. Year End)

Prepare for Disaster
Conclusions from White Paper

1. The industry analysts have taken different positions on what customers are going to use clouds, what they are going to use them for and what the cloud adoption rates will be.

2. Cloud computing implementations have both common attributes and individual attributes that distinguish one cloud from another. The capacity manager should be aware of these differences and incorporate these differences into their operations.

3. The positive economic impact of cloud computing has increased how appealing it is to companies, especially in the current economic situation.

4. Multiple cloud computing models are emerging.

5. The sub-processes of ITIL capacity management easily align to the cloud computing models that are emerging.

6. The same tools, processes, and procedures that are used for non-cloud workload in each of the capacity management sub-processes can be modified to support cloud workload with a minimal amount of effort.

7. Customers who use public clouds which are accessed over the Internet should understand the network implications to the interactive and non-interactive workloads that they put in the cloud.

8. Cloud computing is in its infancy, and capacity planners will have to respond to technology changes in cloud computing for the next several years.
SEE BLUE. THINK GREEN.

Overall:  [www.ibm.com/green](http://www.ibm.com/green)
For CIOs:  [www.ibm.com/cio](http://www.ibm.com/cio)
For facilities managers:  [www.ibm.com/services/siteandfacilities](http://www.ibm.com/services/siteandfacilities)