Services, Semantics, and Cloud

Tharam Dillon, Chen Wu, Elizabeth Chang
Digital Ecosystems and Business Intelligence Institute
Curtin University of Technology, Perth, Australia
14th December 2009

SOCA 2009

Gartner’s Hyper Curve 2009
Gartner’s top 10 technology areas in 2009

- Gartner identified Top 10 Technology areas for 2009
  - **Virtualization**
  - **Cloud Computing**
  - Servers: Beyond Blades
  - Web-Oriented Architectures
  - Enterprise Mashups
  - Specialized Systems
  - Social Software and Social Networking
  - Unified Communications
  - Business Intelligence
  - **Green Information Technology**

Google Trend: Cloud Computing vs. SOA
What is Cloud (Scope)?

- Service
- Platform
- Utility

What is Cloud (Definition)?

- A Cloud is a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resource(s) based on service-level agreements established through negotiation between the service provider and consumers
  
  (Buyya et al. 2009)

- Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services
  
  (Armbrust et al. 2009)

- So simply put
  
  - **Cloud** = Datacenter (Hardware + Software)
  
  - **Infrastructure** (hardware and middleware)
  
  - **Platform** (System software)

  - Cloud computing = **SaaS** + **Cloud**
Layered-architecture of Cloud

- **Applications**  (Layer 5)
- **Mashup**  (Layer 4)
- **SaaS**  (Layer 3)
- **PaaS**  (Layer 2)
- **IaaS & DaaS**  (Layer 1)
- **Virtualization**  (Layer 0)

Amazon Cloud Realization

- **Simple Queue Service**
  - Send & receive messages asynchronously in the cloud
- **EC2 Service (Compute Power)**
  - Store & manage data in the cloud
- **Simple Storage Service (S3)**
  - Store large data on S3 for online access
- **ESS Storage Service**
- **Simple DB Service**

[Link](http://www.telkite.com/images/amazon.gif)
SalesForce Cloud Realization

Key characteristics in eyes of Cloud users

- **Elasticity**
  - No up-front commitment and contract – pay as you go
  - You use it whenever you want, and let it go once you finish
  - Scale up and down – computing on-demand
  - Scale horizontally – various services on-demand
  - Infinite, Immediate, and Invisible computing resources

- **Service Level Agreement and QoS**
  - Negotiation

- **Robust and reliable**

- **Ubiquitous access**

- **Unawareness of which resources being used**

- **Availability**
Key enabling technologies

- Virtualization
  - Virtual Machines
  - Virtual Clusters
  - Virtual Network overlays
- Parallelization
  - MapReduce
- Web services
  - WSDL, SOAP, and REST
- Distributed and Unlimited Storage
  - Google MegaStore/BigTable
  - Amazon EBS/SimpleTable
- Networking
  - VPN (for Private Cloud)

Cloud can benefit Service Computing research

- Service development on Cloud
  - Google AppEngine
  - Global Software Development with Cloud Platforms
- Service testing on Cloud
  - Cloud-Testing@CloudIntelligence
- Service deployment on Cloud (Service Cloud)
  - Amazon Machine Image
- Service integration/mashup/composition on Cloud
  - Yahoo Pipes
- Business process outsourcing through Cloud
Service can benefit Cloud Computing research

- **WS-* may provide insight for cloud computing**
- Service Description for Cloud Service
  - WSDL is currently being used
  - More description languages will be used
- Service Discovery for Cloud Service
  - Various service discovery models can be leveraged
  - More support for SLA, QoS, security (through various WS-*)
- Service Composition for Cloud Service
  - Holy-grail of service-oriented computing
  - A great deal of research in this area can be applied
- Service Management for Cloud Service
  - SOA governance principles can be adapted

SOA

- Architectural style
  - standard interfaces
  - decoupled coupling
- A high level of abstraction: Integration/Process-oriented
- What’s missing in SOA for SMEs?
  - No standard computation models (Computing-oriented)
    - How to run my services with minimum cost?
    - How to scale up/down my services applications?
  - How to maximize the utilization of my IT services in order to support my business services
- In our previous work
  - We propose the notion of SoftGrid, which
  - add services into computation models (SoftGrid)
  - Similar to OGSA, but add the notion of Space
Cloud and SOA

- Distributed computing paradigms
- How to see these two
  - Are they at the same technical/business level?
  - Do they aim to achieve the same goal?
  - Can they be employed at the same time?
    - If so, how?
- We believe that Cloud provides a practical approach for SMEs to accelerate and consolidate the realization of SOA through the facilitation of
  - Service development
  - Service testing
  - Service deployment
  - Service composition/mashup
  - Business process outsourcing

Cloud and Grid

- Similarity
  - Resource virtualization
- Differences
  - Grid emphasizes the "resource sharing" to form a virtual organization
    - Cloud is often owned by a single physical organization, who allocates resources to different running instances
  - Grid wants to provide the maximum computing capacity for a huge task through resource sharing
    - Cloud wants to suffice as many small-to-medium tasks as possible based on users’ real-time requirements
  - Grid trades re-usability for (scientific) high performance computing
    - Cloud is pulled by business requirements in the first place
  - Grid asks for maximum computing
    - Cloud is after on-demand computing – Scale up and down
Grid Computing
- Federation
- Single task
- Maximum capacity
- Virtual organization

Cloud Computing
- Logically Centralized
- Multi tasks, multi-tenants
- Overall capacity
- Physical organization

Cloud and High Performance Computing

- Scientific Applications vs. Business Applications
  - State dependency
  - Data dependency

- The current Cloud is not geared for HPC
  - Not yet matured enough for HPC
  - Like Cluster computing, Cloud infrastructure focuses on enhancing the overall system performance as a whole
  - HPC aims to enhance the performance of a specific scientific application using resources across multiple organizations

- The key difference lies in **elasticity**
  - For cluster computing, the capacity is often **fixed**
  - For cloud computing, we often do not know a-prior how many processors do we need
Public Cloud

- IaaS
  - INFRASTRUCTURE AS A SERVICE

- PaaS
  - PLATFORM AS A SERVICE
  - "Every day more than 3,000 businesses sign up for Google Apps and move to the cloud"

- SaaS
  - SOFTWARE AS A SERVICE

More on SaaS – Multi-tenancy

- **Level 1:**
  - Ad-Hoc/Custom

- **Level 2:**
  - Configurable

- **Level 3:**
  - Configurable
  - Multi-Tenant-Efficient

- **Level 4:**
  - Scalable,
  - Configurable
  - Multi-Tenant-Efficient

Source: Microsoft MSDN Architecture Center
Google SaaS family

Apple SaaS Cloud – iTune University

- A section of Apple’s iTunes Store
- Designed for higher education to store and disseminate
  - institutional content,
  - course work, or
  - other multimedia materials
Private Cloud

- Why do we need private cloud?
  - Security, Privacy & Trust
  - Data transfer cost
  - Fixed Cost vs. Variable Cost
  - Everything is under control

- Amazon Private Cloud

Hybrid Cloud

- Virtual Machine vs. Virtual Infrastructure

Sotomayor et al. 2008
Cloud: Outside-in

- Services
  - computational power
  - storage
  - business applications
  - XaaS

- Interface
  - SOAP and WSDL
  - RESTful

- Resource representation
  - Semantics for Cloud services and parameters
  - Dependencies

Cloud: Inside-out

- Virtualization
  - Virtualization saves Microsoft customers on average half a million per year

- Resource Management

- Metadata and Semantics

- Parallelization
  - Optimal parallelisation models for different types of applications/domain or programming styles
  - Internode communication

- Lifecycle management

- Testing, Monitoring, Diagnose, and Verification
Cloud: Socio-Technical and Economical Impact

- Controllable vs. Uncontrollable
- Multi-tenant data architecture
  - Security
  - Privacy
  - Authentication
- Costing
  - basic load vs. peak load
  - Data transfer cost
- What types of applications are more suitable to be moved to cloud?
  - Website
  - CRM Systems
  - etc.

Moving to the cloud – What to move?

Q: Current and future level use of cloud services in your organization? (1=none, 5=widespread)

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Current</th>
<th>In 3 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT Management Apps</td>
<td>26.2%</td>
<td>39.3%</td>
</tr>
<tr>
<td>Collaborative Apps</td>
<td>25.4%</td>
<td>46.3%</td>
</tr>
<tr>
<td>Personal Apps</td>
<td>25.0%</td>
<td>36.1%</td>
</tr>
<tr>
<td>Business Apps</td>
<td>32.4%</td>
<td>34.0%</td>
</tr>
<tr>
<td>App Development/Deployment</td>
<td>16.0%</td>
<td>25.9%</td>
</tr>
<tr>
<td>Server Capacity</td>
<td>15.6%</td>
<td>28.7%</td>
</tr>
<tr>
<td>Storage Capacity</td>
<td>15.5%</td>
<td>31.5%</td>
</tr>
</tbody>
</table>

Source: IDC Enterprise Panel, August 2008 n=244
US SaaS Market Composition & Penetration

- Collaboration, CRM, and Other account for 70%
- Other includes: core BI & analytics; ALM, PPM
- Penetration (percent of US SaaS market) vs. Saturation (SaaS as percent of all spending)

SaaS Apps by App type

Approx 12% of US CRM Apps
Approx 2% of US ERM Apps
Approx 5% of US Other Apps
Approx 22% of US Collaborative Apps
Approx 5% of US SCM Apps
Approx 4% of US Engineering Apps
Approx 1.5% of US O&M Apps (incl B/OSS)

Source: IDC SaaS Adoption Survey, November 2008. n= 200 US-based IT Director and above, and 200 LOB Director and above.

Moving to the cloud – Cost of moving

- **Amazon Public Cloud Cost**

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Memory (GB)</th>
<th>Compute Units</th>
<th>Storage (GB)</th>
<th>Platform</th>
<th>Linux CPU/Hr</th>
<th>Windows CPU/Hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>1.7GB</td>
<td>4</td>
<td>160GB</td>
<td>32-bit</td>
<td>$0.10</td>
<td>$0.125</td>
</tr>
<tr>
<td>Large</td>
<td>7.5GB</td>
<td>4</td>
<td>850GB</td>
<td>64-bit</td>
<td>$0.40</td>
<td>$0.50</td>
</tr>
<tr>
<td>Extra Large</td>
<td>15GB</td>
<td>8</td>
<td>1690GB</td>
<td>64-bit</td>
<td>$0.80</td>
<td>$1.00</td>
</tr>
<tr>
<td>High CPU, Medium</td>
<td>1.7GB</td>
<td>5</td>
<td>350GB</td>
<td>32-bit</td>
<td>$0.20</td>
<td>$0.30</td>
</tr>
<tr>
<td>High CPU, Large</td>
<td>7GB</td>
<td>20</td>
<td>1690GB</td>
<td>64-bit</td>
<td>$0.80</td>
<td>$1.20</td>
</tr>
</tbody>
</table>

- **Costing a plan**
  - Fixed cost (if using hybrid Cloud)
  - Variable cost (Compute units)
  - Data Transfer cost
    - Application should always be close to data
  - Long term value of CPUs

Walker 2009, IEEE Computer, Special Issue.
Open problems

- Decomposition
  - Data dependency makes it extremely complicated
  - Top-level decomposition to lower-level ones
  - How to estimate complexity of existing algorithms?

- Costing Optimization
  - Long-term costs of cloud development

- Vulnerability issues
  - Security of providers access to user data
  - Trustworthiness of providers (Amazon warrants 99.95% uptime)
  - Security of data placement
  - Security of data representation
  - Recoverability of data
  - Tracking of illicit activities on the cloud
  - "Vendor" lock-in

Some directions for SOA researchers

- Service Description and Discovery
- Service Level Agreement
- Security and Data privacy
- Service Composition/Integration
- Emergent Specification
Services and Cloud in Australia

- Cloud Computing at Peta-scale working group (since 2008)
- Australian Research Collaboration Service (2009 – 2013, $ 97 million)
- Cooperative Research Centres (CRC)
  - A proposal on “Cloud-based Applications”

Cloud in the future

- Three possible scenarios (Nelson 2009)
  - Many cloud
  - Hazy cloud
  - Open cloud

Thank you for your attention!