Special Focus: Implementing Big Data in the Cloud

Presented to Government Big Data Symposium

Kevin L. Jackson
Vice President and General Manager, NJVC Cloud Services

March 5, 2013
Agenda

- Why use cloud computing for big data?
- What’s wrong with using a single cloud services provider (CSP)?
- What makes CSPs different from one another?
- Why should you consider using a cloud services broker (CSB)?
Why Consider the Cloud for Big Data?

- Ubiquitous Access to Infrastructure
- Scalable
- Long-Term Financial Commitment Avoidance
- Capital Expenditure Minimization
- Shared Infrastructure
- Procurement Agility
Why Consider Multiple CSPs?

<table>
<thead>
<tr>
<th>Provider Offering</th>
<th>Description</th>
<th>Small Package Estimated Monthly Price</th>
<th>Medium Package Estimated Monthly Price</th>
<th>Large Package Estimated Monthly Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Web Services™</td>
<td>On demand commodity cloud provider. Enterprise support available at additional cost</td>
<td>$2,697</td>
<td>$9,568</td>
<td>$34,515</td>
</tr>
<tr>
<td>GovCloud Amazon Web Services™</td>
<td>On demand cloud designed to allow US government agencies and contractors to move sensitive workloads into the cloud and meet specific regulatory and compliance requirements</td>
<td>$3,132</td>
<td>$10,693</td>
<td>$38,329</td>
</tr>
<tr>
<td>SAVVIS Essential</td>
<td>Entry-level enterprise cloud - lowest cost of entry: optimized for development</td>
<td>$2,006</td>
<td>$6,667</td>
<td>$24,413</td>
</tr>
<tr>
<td>SAVVIS Balanced</td>
<td>Higher-level security &amp; SLA: optimized for eCommerce &amp; Web Hosting</td>
<td>$3,423</td>
<td>$10,138</td>
<td>$34,904</td>
</tr>
<tr>
<td>SAVVIS Premier</td>
<td>Highest level of security, redundancy, availability &amp; performance for mission critical apps</td>
<td>$7,615</td>
<td>$18,605</td>
<td>$80,575</td>
</tr>
</tbody>
</table>
# CSP Specialization – Commodity vs. Enterprise

## Assessed Applications

<table>
<thead>
<tr>
<th>Application Group</th>
<th>Readiness</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Intelligence</td>
<td>86.3%</td>
<td>35% -</td>
</tr>
<tr>
<td>Data Storage</td>
<td>88.0%</td>
<td>30% - 50%</td>
</tr>
<tr>
<td>CRM</td>
<td>94.0%</td>
<td>55.00000000000</td>
</tr>
<tr>
<td>Database</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Diagrams

- **Benefit vs. Readiness**
  - Business Intelligence
  - Data Storage

- **Infrastructure**
  - Commodity
  - Enterprise
  - Public Cloud
  - Private Cloud
  - CRM
  - Business Intelligence
  - Data Storage
CSP Specialization – Marketplace Priorities

⭐ Commercial vs. Government
⭐ Commodity vs. Enterprise
⭐ Security
⭐ Quality of Service
⭐ Cost Sensitivity
## Capacity Requirements

### Compute

<table>
<thead>
<tr>
<th>Type</th>
<th>CPU</th>
<th>Cores</th>
<th>Processor (GHz)</th>
<th>Memory (GB)</th>
<th>Storage (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web</td>
<td>10</td>
<td>4</td>
<td>3.00</td>
<td>10</td>
<td>500</td>
</tr>
<tr>
<td>DB Servers</td>
<td>111</td>
<td>4</td>
<td>3.111</td>
<td>10</td>
<td>5/11</td>
</tr>
<tr>
<td>VPN</td>
<td>10</td>
<td>4</td>
<td>3.00</td>
<td>10</td>
<td>500</td>
</tr>
</tbody>
</table>

### Operating System
- Windows
- Solaris
- Linux RedHat
- Linux Ubuntu
- Linux CentOS

### Storage
- Additional Storage: 1000 GB

### Network
- Bandwidth (Mbps): 2

### Specify Cloud Capacity

<table>
<thead>
<tr>
<th>Component</th>
<th>Order</th>
<th>Recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>500</td>
<td>480.00</td>
</tr>
<tr>
<td>Memory (GB)</td>
<td>400</td>
<td>400.00</td>
</tr>
<tr>
<td>Storage (GB)</td>
<td>20000</td>
<td>21,234.00</td>
</tr>
<tr>
<td>VMCs</td>
<td>31</td>
<td>63.00</td>
</tr>
<tr>
<td>VPUUs</td>
<td>150</td>
<td>209.00</td>
</tr>
<tr>
<td>Bandwidth (Mbps)</td>
<td>2</td>
<td>2.00</td>
</tr>
</tbody>
</table>
## Technical Compatibility

<table>
<thead>
<tr>
<th>Feature</th>
<th>Terremark</th>
<th>Savvis</th>
<th>Cisco/EMC</th>
<th>Amazon</th>
<th>Go Grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Private Cloud</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>External Private Cloud</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Enterprise Class</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commodity Class</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Hybrid with Hosting</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Delivery</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SAN Storage</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large Data Transfer</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VLAN Firewall</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dedicated Network</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Security Requirements

<table>
<thead>
<tr>
<th>Security</th>
<th>Server Tier Firewall</th>
<th>IDS/IPS</th>
<th>Virtual Perimeter Firewall and Routing between VLANs</th>
<th>Enterprise Grade Virtual Perimeter Firewall</th>
<th>Additional Cost</th>
<th>Enterprise Grade Virtual Perimeter Firewall</th>
<th>Additional Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Server Tier Firewall</td>
<td>IDS/IPS</td>
<td>Virtual Perimeter Firewall and Routing between VLANs</td>
<td>Enterprise Grade Virtual Perimeter Firewall</td>
<td>Additional Cost</td>
<td>Enterprise Grade Virtual Perimeter Firewall</td>
<td>Additional Cost</td>
</tr>
<tr>
<td></td>
<td>99.5% Annual Uptime only for access to VM in particular region. Service credit up to 10%. Certifications: SOC 1/SSAE 16/SSAE 18, FISMA Moderate, PCI DSS Level 1, ISO/IEC 27001, International Traffic In Arms Compliance, FIPS 140-2, HIPAA Support, Cloud Security Alliance (CSA)</td>
<td>100% infrastructure 99.9% end-to-end availability of customer assets. Service Credit up to 30%.</td>
<td>99.9% end-to-end availability of customer assets. Service Credit up to 60%.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Managing Multiple CSPs with an Internal CSB

Attributes of Internal CSB

✱ Perceived as core responsibility
✱ Responsible for all CSB roles
✱ Triggered by having multiple CSPs, and role applies to both public and private cloud services

Who will be responsible for end-to-end delivery of cloud services?

Answer: 80% of respondents say it should be the "CIO" or "Designated IT Manager"
Managing Multiple CSPs with an External CSB

CSB Value

Without a CSB

- Leakage Reduces Cloud Benefits
- 8% Savings Realized
- 57% Savings Promised

- Physical Capacity
- Unused Capacity
- Operations
- Utility
- Arch. Inefficiencies
- Vendor Lock-in
- Billing Complexity
- VM Sprawl

$90,000
$80,000
$70,000
$60,000
$50,000
$40,000
$30,000
$20,000
$10,000
$0
Application Specifics Drive Deployment Options

- Storage Requirements
- Network Latency
- Geographic Restrictions
- Access to Social Media Data
- Legal Restrictions
- Big Data Security
- Data Privacy
- Data Providence
Reducing Cost Via Spot Markets
Virtustream SpotCloud

“SpotCloud: Buy cloud servers, and instantly get the broadest choice of precise geo-targeted capacity at the best price.”
Reducing Cost Via Spot Markets
Amazon Spot Instances

- Name your own price for Amazon EC2 computing capacity
- Spot price varies in real time based on supply and demand
- Can provide the most cost-effective option for obtaining compute capacity
- Can significantly lower computing costs for time-flexible, interruption-tolerant tasks
- May be able to simultaneously accelerate the computational task and reduce overall cost (i.e. distributed, fault-tolerant tasks like Web crawling or Monte Carlo applications)
NCOIC GCC Infrastructure:
Jan 2013
A leading global power generation company found that usage of public cloud models for a subset of workloads would drive 30+% reduction in overall infrastructure costs.

### IT Infrastructure Annual Cost
Indexed US$, Year 1 Baseline\(^1\) = 100

1. Transition costs are spread over 2 years; retained costs excluded.

- Nearly 30% improvement in cost offered by public cloud-enabled solution
- Public cloud only used for ~25% of workloads
- Modest use of public cloud to drive utilization appears to have substantial benefit
- Efficiencies do not reflect likely price declines in public cloud services

\(^1\) Source: Everest Group
Summary

- Why use cloud computing for big data?
  - Reduce cost, improve agility and flexibility

- What’s wrong with using a single CSP?
  - All CSPs aren’t created equal

- What makes CSPs different from one another?
  - Specialization, capabilities and industry/mission relevant experience

- Why should you consider CSB?
  - Increased savings, improved agility and flexibility, and mission alignment, and avoidance of CSP lock in
Thank You

Kevin L. Jackson  
Vice President and General Manager,  
Cloud Services  
NJVC, LLC  
703.335.0830  
kevin.jackson@njvc.com

http://www.cloudcuity.com  
http://www.njvc.com  
http://govcloud.ulitzer.com