Disaster Recovery as a Cloud Service: Economic Benefits and Deployment Challenges

Tim Wood, Emmanuel Cecchet, KK Ramakrishnan*, Prashant Shenoy, Kobus van der Merwe*, and Arun Venkataramani

UMass Amherst and AT&T*
Disasters happen

Disasters are expensive
Data Center Disasters

- Disasters cause expensive application downtime

- Truck crash shuts down Amazon EC2 site (May 2010)
- Lightning strikes EC2 data center (May 2009)
- Squirrels bring down NASDAQ exchange (1987 and 1994)

Need plans and systems in place to recover from disasters
Disaster Recovery

- Use DR services to prevent lengthy service disruptions
- Long distance data backups + failover mechanism
  - Periodically replicate state
  - Switch to backup site after disaster

Can the cloud reduce the cost of DR and improve the level of service?
DR Metrics

- DR Goal: minimize data loss, downtime, and cost

- Recovery Point Objective (RPO)
  - Amount of tolerable data loss

- Recovery Time Objective (RTO)
  - Acceptable system downtime

We focus on RPO and RTO > 0
Why DR Fits in the Cloud

**Customer:** pay-as-you-go and elasticity
- “Normal” case is cheap (need few resources to make backups)
  - Lower cost for a given RPO
- Can rapidly scale up resources after disaster is detected
  - Cloud’s virtualized infrastructure reduces RTO
  - Can allow for **business continuity**

**Provider:** High degree of multiplexing
- Customers will not all fail at once
- Can offer extra services like disaster detection

Is the cloud an economical platform for DR today? What additional features are needed?
**DR on Demand**

- **Warm Backup Site**
  - Cheaply synchronize state during normal operation
  - Obtain additional DR resources on demand after failure
  - Short delay to provision and initialize applications

---

Diagram showing the flow of applications from the Enterprise DC to the DR Cloud, with backup state and DR Server highlighted.
Cost Analysis Scenario

- Compare the cost of DR in Colocation center to Cloud.
- Colo case pays for servers and space at all times.
- Cloud DR only pays for resources as they are used.
- Case 1: RUBiS ebay-like multi-tier web application
  - 3 web front ends
  - 1 database server
- Only database state is replicated.
Cost Analysis: Colocation vs Cloud

- **Normal Case**
  - Resources needed to replicate DB state

- **Post-Disaster**
  - Resources needed to run all application components

<table>
<thead>
<tr>
<th></th>
<th>Normal Case</th>
<th>Post-Disaster</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Servers</strong></td>
<td>colo = 4 servers</td>
<td>colo = 4 servers</td>
</tr>
<tr>
<td></td>
<td>cloud = 1 VMs</td>
<td>cloud = 5 VMs</td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td>5 GB/day</td>
<td>180 GB/day</td>
</tr>
<tr>
<td><strong>Colocation:</strong></td>
<td>$28.04/day</td>
<td>$66.01/day</td>
</tr>
<tr>
<td><strong>Cloud:</strong></td>
<td>$3.80/day</td>
<td>$52.03/day</td>
</tr>
</tbody>
</table>

- **99% Uptime cost (3 days of disaster per year)**
  - Colo: **$10,373 per year**
  - Cloud: **$1,562 per year**
RPO vs Cost Tradeoff

- Case 2: Data Warehouse
  - Post-disaster twice as expensive with Cloud
    - Cloud charges premium for high powered VM instance
    - Cloud still cheaper overall due to lower normal case costs
- Cloud allows tradeoff between RPO and cost
  - Only pay for DR server during periodic backups in cloud

![Cost vs RPO Graph]

Colo center pays server and space costs regardless of RPO!
Cost Analysis Summary

- Benefits of cloud computing depend on:
  - Type of resources required to run application
  - Variation between normal mode and post-disaster costs
  - RPO and RTO requirements
  - Likelihood of disaster

Cloud has greatest benefit when post disaster cost much **higher** than normal mode
Provider Challenges

- **Revenue Maximization**
  - Mainly makes income from storage in “normal” case
  - But must pay for servers and keep them available
  - Can use pricing mechanism such as **spot instances**
    - Rent resources but be able to quickly reclaim for DR
  - Rent **priority resources** at higher cost that are guaranteed to be available

- **Correlated Failures**
  - Large disasters could affect many customers simultaneously
  - Cloud provider must
    - Use a risk model to decide how many resources to own for DR
    - Spread out customers to minimize impact of correlated failures
More DR Challenges

• **Planning**
  • Use models to help understand tradeoff between cost and RPO/RTO for a given application and workload

• **Efficient state replication**
  • Minimize the bandwidth and cloud server costs in the normal case

• **Post Disaster Failover**
  • Enable business continuity by minimizing recovery time
  • Automated/virtualized cloud infrastructure can lower RTO
Summary

• Cloud based Disaster Recovery
  • Can substantially reduce cost for customer
    • Particularly when server cost varies before/after disaster
  • Provides flexible tradeoff between cost and RPO
  • Can lower recovery time, enable business continuity
  • Provider must handle correlated failures

• Open challenges
  • How many resources must provider reserve for DR?
  • How to seamlessly transfer network connections?
  • How to fail back to primary site after disaster passes?

twood@cs.umass.edu
## Cost Details

<table>
<thead>
<tr>
<th>RUBiS</th>
<th>Public Cloud</th>
<th></th>
<th>Colocation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Replication</td>
<td>Failover</td>
<td>Replication</td>
<td>Failover</td>
</tr>
<tr>
<td>Servers</td>
<td>$2.04</td>
<td>$32.64</td>
<td>$26.88</td>
<td>$26.88</td>
</tr>
<tr>
<td>Network</td>
<td>$0.54</td>
<td>$18.00</td>
<td>$1.16</td>
<td>$39.14</td>
</tr>
<tr>
<td>Storage</td>
<td>$1.22</td>
<td>$1.39</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total per day</td>
<td>$3.80</td>
<td>$52.03</td>
<td>$28.04</td>
<td>$66.01</td>
</tr>
<tr>
<td>Total per year</td>
<td>$1,386</td>
<td>$18,992</td>
<td>$10,234</td>
<td>$24,095</td>
</tr>
<tr>
<td>99% uptime cost</td>
<td>$1,562 per year</td>
<td></td>
<td>$10,373 per year</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Warehouse</th>
<th>Public Cloud</th>
<th></th>
<th>Colocation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Replication</td>
<td>Failover</td>
<td>Replication</td>
<td>Failover</td>
</tr>
<tr>
<td>Servers</td>
<td>$4.08</td>
<td>$12.00</td>
<td>$8.51</td>
<td>$8.51</td>
</tr>
<tr>
<td>Network</td>
<td>$0.10</td>
<td>$0.12</td>
<td>$0.22</td>
<td>$0.26</td>
</tr>
<tr>
<td>Storage</td>
<td>$3.50</td>
<td>$3.92</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total per day</td>
<td>$7.68</td>
<td>$16.04</td>
<td>$8.73</td>
<td>$8.77</td>
</tr>
<tr>
<td>Total per year</td>
<td>$2,802</td>
<td>$5,853</td>
<td>$3,186</td>
<td>$3,202</td>
</tr>
<tr>
<td>99% uptime cost</td>
<td>$2,832 per year</td>
<td></td>
<td>$3,186 per year</td>
<td></td>
</tr>
</tbody>
</table>
Enabling Business Continuity

• Business continuity allows applications to keep working after a disaster
  • Crucial for critical business/government services

• Virtualized cloud infrastructure can lower RTO
  • Automates VM creation and cloning
  • Cloud can also help with disaster detection

• Many remaining challenges
  • How to ensure application is revived in a consistent/correct state?
  • How to redirect traffic to failover site?
DR Requirements

- Recovery Point Objective (RPO)
  - Amount of tolerable data loss
- Recovery Time Objective (RTO)
  - Acceptable system downtime
- Performance
  - Impact on normal operation and after recovery
- Consistency
  - Correctness of application data and outputs
- Geographic Separation
  - DR site should not be affected by same disaster
What is the cloud good for?

• Cloud platforms are best for users who have variable needs over time
  • Customers only pay for what they use
  • Providers get economy of scale and can multiplex resources for many customers

• Applications well matched for the cloud:
  • Web sites with growing or variable demand
  • Infrequent compute intensive jobs (monthly payroll)

• and...
  • Disaster recovery!