Glacier Unveiled

Andrew Woods
DuraSpace
PASIG May, 2013
The Basics

Glacier is:
- An Amazon (AWS) cloud storage service
- Designed for low-access, long-retention data
- Elastic
- Encrypted
Integration Options

• Direct to Glacier (Java, .NET, REST, SOAP)
  – Vaults and Archives
    • Up to 1000 Vaults
    • Up to 40-TB per Archive
  – Support for range requests (1-MB chunks)

• Glacier via S3
  – Lifecycle policies
Open Questions

- Durability... where do the eleven 9s come from?
- What is the back-end technology?
- What server-side bit-integrity processes exist?
Cost Factors

<table>
<thead>
<tr>
<th></th>
<th>Storage $/GB/month</th>
<th>Transfer $/GB</th>
<th>Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3</td>
<td>$.095 to $.055</td>
<td>$.12 to $.05</td>
<td>N/A</td>
</tr>
<tr>
<td>RRS</td>
<td>$.076 to $.037</td>
<td>$.12 to $.05</td>
<td>N/A</td>
</tr>
<tr>
<td>Glacier</td>
<td>$.010</td>
<td>$.12 to $.05</td>
<td>??</td>
</tr>
</tbody>
</table>

- **Request Operations**
  - $.00005/request = $1 for 20,000 requests

- **Deletion Policy**
Transfer Costs

- Same for S3 and Glacier
- Tiered pricing

<table>
<thead>
<tr>
<th>Range</th>
<th>Cost/GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 10-TB</td>
<td>$0.12</td>
</tr>
<tr>
<td>10 to 40-TB</td>
<td>$0.09</td>
</tr>
<tr>
<td>40 to 100-TB</td>
<td>$0.07</td>
</tr>
<tr>
<td>100 to 350-TB</td>
<td>$0.05</td>
</tr>
</tbody>
</table>

- In lowest tier, all content can be read for the cost of annual storage
- In the highest tier, all content can be read for approx. ½ the cost of annual storage
Data Retrieval Costs

• Variables
  - S: Total amount of data stored (GB)
  - R: Total amount of data retrieved (GB)
  - ΔT: Time period over which data is retrieved (hrs)
• PRR: Peak-Retrieval-Rate
  - Greatest amount of data retrieved in any hour over the course of a month (based on when retrieval was initiated, spanning 4 hours)
• PBRR: Peak-Billable-Retrieval-Rate
  - PRR minus Free daily allowance divided by hours of retrieval
• Cost = PBRR * Retrieval fee * hrs/month
• Cost = (R/ΔT – (S * .05 / 30 / min[ΔT,24hrs])) * $.01/GB * 720hrs/month
Glacier Retrieval Cost...

...Not So Tricky:

\[
\frac{R}{\Delta T} - \\
\left( S \times 0.05 / 30 / \min[\Delta T, 24\text{hrs}] \right) \\
\times \$0.01/\text{GB} \\
\times 720\ \text{hrs/month}
\]
Disaster Recovery

100-TB in 3 Days

<table>
<thead>
<tr>
<th>Storage</th>
<th>Retrieval</th>
<th>Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$92,300</td>
<td>$12,200</td>
<td>$8,300</td>
</tr>
<tr>
<td>$8,300</td>
<td>$10,000</td>
<td>$8,300</td>
</tr>
</tbody>
</table>

Annual Cost

- S3:
  - Storage: $92,300
  - Transfer: $8,300
- Glacier:
  - Storage: $12,200
  - Retrieval: $10,000
  - Transfer: $8,300
Data Retrieval Results

• Assumptions
  – All content is to be retrieve (R = S)
  – Retrievals exceed 24 hours
• At what value of $\Delta T$ does cost = $S \times \$.12/\text{GB}$?
  – All content can be retrieved in...
    • 3 days for the cost of annual storage
    • 1 month for $\sim1/10^{\text{th}}$ the cost of annual storage
    • 3 months for $\sim1/50^{\text{th}}$ the cost of annual storage
    • 6 months for $\sim1/100^{\text{th}}$ the cost of annual storage
Use Cases

• Disaster Recovery
  – Budget 3x annual storage cost for ~3day full data recovery
    • 1x annual storage cost for actual storage
    • 1x annual storage cost for retrieval fees
    • 1x annual storage cost for data transfer

• Bit-Integrity Auditing
  – Schedule quarterly or greater iteration cycle
  – Plan for annual cost to be limited to the storage cost plus EC2 instance for checksum generation

• Access Copy
  – Probably not an ideal use case
Questions?

Thank you

awoods@duraspace.org