AMAZON S3 FOR SCIENCE GRIDS: A VIABLE SOLUTION?

Mayur Palankar and Adriana Iamnitchi
University of South Florida

Matei Ripeanu
University of British Columbia

Simson Garfinkel
Harvard University
S3 Backup – Fail-safe, Encrypted Online Backup

S3 Backup is a fail-safe, encrypted online backup solution that features aren’t available yet.

S3 Backup is powered by Amazon S3 – highly scalable infrastructure that Amazon uses to run its own global services. You ensure that your data are stored in multiple locations, safe even in the event of natural disaster or sabotage.

(yes, I said purchase. We have no debt - we owe nothing. And every SmugMug customer gets unlimited storage for life.)

Amazon S3 Reaches 5 Billion Stored Objects

Michael Arrington

April 17, 2007

45 comments »

While I was busy announcing the upcoming TechCrunch20 conference and picking fights with venture capitalists today, Dan Farber was covering the news at the Web 2.0 Expo. One of the more interesting disclosures to surface: Amazon’s Jeff Bezos announced that their S3 storage-on-demand service, which launched just thirteen months ago, surpassed 5 billion stored objects. This is up from just 800 million stored objects in July 2006.
Science Grids
- Data-intensive scientific collaborations
- Produce, analyze, and archive huge volumes of data (PetaBytes)
  - High data management and maintenance costs
  - Files are often used by groups of users and not individually

Amazon Simple Storage Service (S3)
- Novel storage ‘utility’:
  - Direct access to storage
- Self-defined performance targets:
  - Scalable, infinite data durability, 99.99% availability, fast data access
- Pay-as-you go pricing:
  - $0.15/month/GB stored and $0.10-$0.17/GB transferred

Is offloading data storage from an in-house mass storage system to S3 feasible and cost-effective?
**Approach**

- Characterize S3
  - Does it live up to its own expectations?
- Toy scenario: evaluate a representative scientific application (DZero) in this context
  - Estimate performance and costs
  - Is the functionality provided adequate?

**Outline**

- S3 architecture
- S3 performance evaluation
- Toy scenario: S3-supported DZero: cost and functionality requirements
- Lessons/suggested improvements
S3 Architecture

- Two-level namespace
  - Buckets (think directories)
    - Global names
    - Two goals: data organization and charging
  - Data objects
    - Opaque object (max 5GB)
    - Metadata (attribute-value, up to 4KB)

- Functionality
  - Simple put/get functionality
  - Limited search functionality
  - Objects are immutable, cannot be renamed

- Data access protocols
  - SOAP
  - REST
  - BitTorrent
S3 Architecture (...cont)

• Security
  – Identities
    • Assigned by S3 when initial contract is ‘signed’
  – Authentication
    • Public/private key scheme
    • But private key is generated by Amazon!
  – Access control
    • Access control lists (limited to 100 principals)
    • ACL attributes
      – FullControl
      – Read & Write (objects cannot be written)
      – ReadACL & WriteACL (for buckets or objects)
  – Auditing (pseudo)
    • S3 can provide a log record
Storage Service Requirements for Science Grids

- Data durability
- Data availability
- Access performance
- Usability
- Support for access control and privacy policies
- Low cost

Approach

- Characterize S3
  - Does it live up to its own expectations?
- Toy scenario: evaluate a representative scientific application (DZero) in this context
  - Estimate performance and costs
  - Is the functionality provided adequate?
S3 Characterization Methodology

Black-box approach using PlanetLab nodes to estimate:
- durability
- availability
- access performance
- the effect of BitTorrent on cost savings
Evaluating Amazon S3

• Durability: perfect
  – But limited scale experiment (1 year)
  – More than 137,000 requests for 10,000 files

• Availability
  – From EC2: higher than declared 99.99%
  – From over the Internet: 23 weeks of testing
    • About 15,456 access requests (every 15 mins)
    • Retry protocol, exponential back-off
      – 95.89% availability after the original access
      – 98.06% availability after the 1st retry
      – 99.01% availability after the 2nd retry
      – 99.76% availability after the 3rd retry
      – 99.94% availability after the 4th retry
      – 100% availability after the 5th retry
S3 Access Performance: Single Thread
S3 Access Performance: Multi-Thread

Amazon S3 for science grids: a viable solution?
Access Performance via BitTorrent
Approach

- Characterize S3
  - Does it live up to its own expectations?
- Toy scenario: evaluate a representative scientific application (DZero) in this context
  - Estimate performance and costs
  - Is the functionality provided adequate?
The DØ Experiment

- High-energy physics collaboration
- Traces from January ‘03 to March ‘05 (27 months)
- 375 TB data, 5.2 PB transferred
- Shared data usage: no access control
- 561 users from 70+ institutions in 18 countries
- High intensity data usage: ~550Mbps sustained access rate in DZero
- 113,062 jobs running for 973,892 hours over the period of 27 months

<table>
<thead>
<tr>
<th>Trace recording interval</th>
<th>01/2003 – 03/2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of jobs</td>
<td>113,062</td>
</tr>
<tr>
<td>Hours of computation</td>
<td>973,892</td>
</tr>
<tr>
<td>Total storage volume</td>
<td>375 TB</td>
</tr>
<tr>
<td>Total data processed</td>
<td>5.2 PB</td>
</tr>
<tr>
<td>Average data access rate</td>
<td>273 GB/hour</td>
</tr>
</tbody>
</table>
S3 Evaluation: Cost

• All data stored at S3 and processed by DZero
  – Storage cost: $691,000/year ($829,440 for S3-Europe)
  – Transfer: $335,012/year
    → $85,500/month

• Reducing transfer costs:
  – Caching:
    • 50TB cooperative cache: $43,888 per year in transfer costs
      (~10 times lower)
  – BitTorrent and distributed replicas
  – Use EC2: Replace transfer costs with $43,284/year

• Reducing storage costs:
  – Archive cold data
    – lifetime of 30% files < 24 hours, 40% < a week, 50% < a month
  – Throw away derived data
    – Distinguish between raw and derived data
Key Idea: Unbundling performance characteristics

- High availability, high durability, high performance are bundled at a single pricing point.
- Some applications need only one or two of them
  - A cache: availability and access performance
  - A backup solution (for DZero): durability and availability
- Solution: SLAs that allow the user to specify their requirements and choose pricing point.
## Unbundling Performance Characteristics

<table>
<thead>
<tr>
<th>Application class</th>
<th>Durability</th>
<th>Availability</th>
<th>High performance data access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cache</td>
<td>No</td>
<td>Depends</td>
<td>Yes</td>
</tr>
<tr>
<td>Long-term archival</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Online production</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Batch production</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The resources needed to provide high performance data access, high data availability and long data durability are different.
S3 Evaluation: Security

• Risks
  – Traditional risks with distributed storage are still a concern:
    • Permanent data loss,
    • Temporary data unavailability (DoS),
    • Loss of confidentiality
    • Malicious or erroneous data modifications
  – New risk: direct monetary loss
    • Magnified as there is no built-in solution to limit loss

• Security scheme’s big advantage: it’s simple
• ... but has limitations
  – Access control
    • Hard to use ACLs in large systems – needs at least groups (now available)
    • ACLs limited to 100 principals
  – No support for fine grained delegation
  – Implicit trust between users and the service S3
    • No support for un-repudiability
  – No tools to limit risk
Lessons: Suggested Improvements

• To lower costs: unbundle performance characteristics
  – High availability, High durability, High performance are bundled at a single pricing point.
  – Some applications need only one or two of them
  – Solution: SLAs that allow the user to specify their requirements and chose pricing point.

• To provide specific support for science collaborations
  – Better security support for complex collaborations
  – Additional functionality for better usability:
    • Metadata based searches
    • Renaming and mutating objects
  – Relax hard-coded limitations: 100 buckets, 100 users in ACL, etc.

• Lesson for application integrators: Use application-level information to reduce costs
  – Raw vs. derived data
  – Exploit usage patterns: e.g., data gets cold.

• The answer: not yet.
Thank you