

# Glacier Unveiled

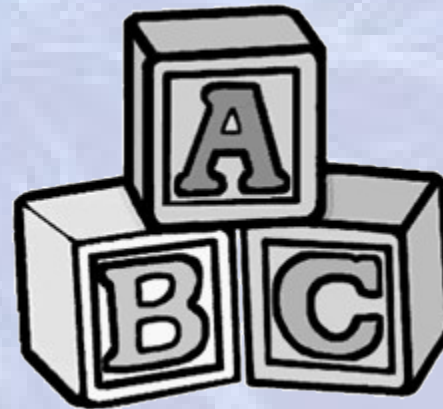
Andrew Woods  
DuraSpace  
PASIG May, 2013



# The Basics

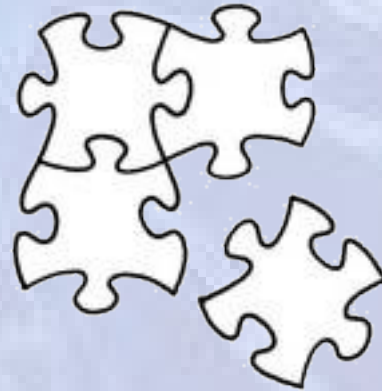
Glacier is:

- 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

	Storage \$/GB/month	Transfer \$/GB	Retrieval
S3	\$.095 to \$.055	\$.12 to \$.05	N/A
RRS	\$.076 to \$.037	\$.12 to \$.05	N/A
Glacier	<b>\$.010</b>	<b>\$.12 to \$.05</b>	<b>??</b>

- Request Operations
  - \$.00005/request = \$1 for 20,000 requests
- Deletion Policy



# Transfer Costs

- Same for S3 and Glacier
- Tiered pricing

0 to 10-TB	\$0.12 / GB
10 to 40-TB	\$0.09 / GB
40 to 100-TB	\$0.07 / GB
100 to 350-TB	\$0.05 / GB

- 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.  $\frac{1}{2}$  the cost of annual storage



# Data Retrieval Costs

- Variables
  - S: Total amount of data **stored** (GB)
  - R: Total amount of data **retrieved** (GB)
  - $\Delta T$ : 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 =  $(R/\Delta T - (S * .05 / 30 / \min[\Delta T, 24\text{hrs}]))) * \$.01/\text{GB} * 720\text{hrs/month}$



# Glacier Retrieval Cost...

...Not So Tricky:

$(R / \Delta T -$

$(S * .05 / 30 / \min[\Delta T, 24\text{hrs}]))$

\* \$.01/GB

\* 720 hrs/month





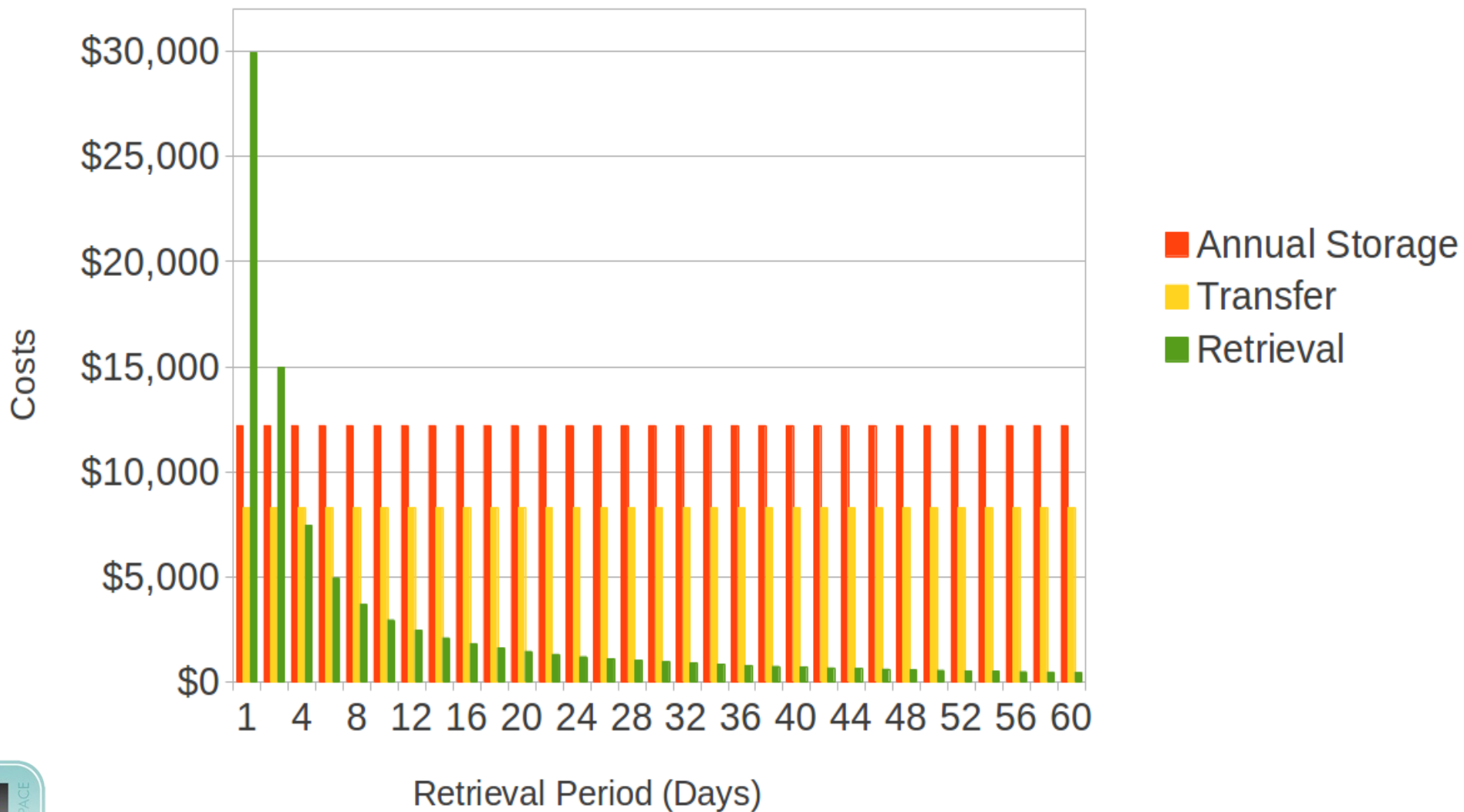
# Disaster Recovery

100-TB in 3 Days



# Full Restore (100-TB)

## Relative Costs Per Retrieval Period



# Data Retrieval Results

- Assumptions
  - All content is to be retrieve ( $R = S$ )
  - Retrievals exceed 24 hours
- At what value of  $\Delta T$  does  $\text{cost} = S * \$0.12/\text{GB}$ ?
  - All content can be retrieved in...
    - 3 days for the cost of annual storage
    - 1 month for  $\sim 1/10^{\text{th}}$  the cost of annual storage
    - 3 months for  $\sim 1/50^{\text{th}}$  the cost of annual storage
    - 6 months for  $\sim 1/100^{\text{th}}$  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](mailto:awoods@duraspace.org)

