Proposal for Open Repositories 2013: 24x7

Title: Amazon Glacier: Why to use it, When to use it, and What it will cost you.
Author: Bill Branan
Affiliation: DuraSpace

Proposal

It is a problem with which we are all familiar: Digital content needs to be stored safely somewhere, but with limited staff time, storage hardware, and financing, a good solution can be elusive. Enter cloud computing - in particular, Amazon's recently announced Glacier offering. Glacier provides an offsite, durable, and flexible storage solution, all for the low price of $0.01 per gigabyte. At first glance, this might appear to be a perfect fit for your storage needs, and it may be, but using Glacier also poses some dangers which need to be understood before you start the data transfer.

Certainly Glacier provides an excellent option if data storage were the only concern. Glacier is part of the Amazon Web Services suite of tools, so it is being offered by the current market leader in the cloud computing space. Amazon's reputation suggests, and analysts as well as early adopters agree, that Glacier is well engineered to handle storing vast amounts of content in a way that is secure, durable, and cheap. In fact, according to Amazon, Glacier is designed to provide an average annual durability of 99.999999999% through the use of replication, integrity checking, and automatic self-healing. These kinds of features often do not come cheap, but Glacier offers storage for 1 cent per gigabyte.

The obvious question, at this point, is: "So what's the catch?" There are actually two substantial "catches" with Glacier, the first being that data stored in Glacier cannot be immediately accessed. Unlike most other cloud storage services, Glacier requires that you request that your data be made available to you, then wait some amount of time before being able to retrieve that content. This time delay is said to be 3 to 5 hours on average, but no maximum time period is provided. If you need to be able to access your content quickly, such as to restore a production service or application in the event of an emergency, this delay would likely be unacceptable. The second "catch", and what is likely to be the primary concern for

most users, is that the pricing for the retrieval of content can vary widely depending on how much content is requested at one time. Amazon provides a very small allocation of free content retrievals, but once this threshold is passed, the price can rise dramatically. This presentation will illustrate just how much a content retrieval could cost, and what methods should be taken to keep those costs as low as possible.

Considering that there are both significant benefits as well as substantial drawbacks to Glacier, when is Glacier a good idea? There are certain use cases which lend themselves well to the Glacier model, and many others which do not. This presentation will discuss those use cases, and attempt to provide guidelines which can be used to determine when and where Glacier is a good fit.