Understanding Cloud Storage Costs

http://datasys.cs.iit.edu/

Illinois Institute of Technology

Computer Science Department
Data-Intensive Distributed Systems Laboratory

Motivation

Cloud Storage is a new paradigm for saving the user data in large scales on a remote location. Recently some of the cloud pioneer companies as well as many startups have provided online storage of mass scales to different level of users. The users of these systems vary from huge tech and non-tech companies to the small start-ups and individual users. Unfortunately the pricing of these products have not yet adapted to the budget limit all of the potential users. There is a big need for lower cost storage system that can be used by the individual users.

Considering the gap between the raw storage price and the price that is offered by the current online storage providers, there is a big opportunity to provide a cloud storage system that focuses on normal end-users instead of big companies. The cost estimation for a data center node with large data capacity shows promising results for the data cost. The cost of having storage in different platforms in different sizes are compared with price of personal storage device over a three year time span that is believed to be the average lifetime of hard drive .

Comparing the number of the potential users of such a system with traditional users of online storage shows that there is a good opportunity in this area to provide such a system. A storage system that provides services to this type of users can be easier in some aspects of the service. Such a system can make some assumptions on some features of the cloud storage such as lower availability rate and lower bandwidth. Normal users mostly use the cloud storage for their backup data. That means they don't need to access their data so often, and they will also do not need to have a parallel access to their data from the different locations at the same time.

New storage providers have the opportunity to use new technologies to provide cost effective service. One of the big challenges in online storage systems is reliability and fault tolerance. Storage providers need to provide 3x replication for the data. That means they need to use 3 times more than the size of data in order to provide a reliable system. Using dispersed storage systems, the new providers can save on their resources providing the same reliability by only using 1.2x replication. This is a customizable feature in dispersed storage systems and can be adjusted according to the importance of the data. That mean the reliability can be used as an on-demand feature that can be set by the data owner.
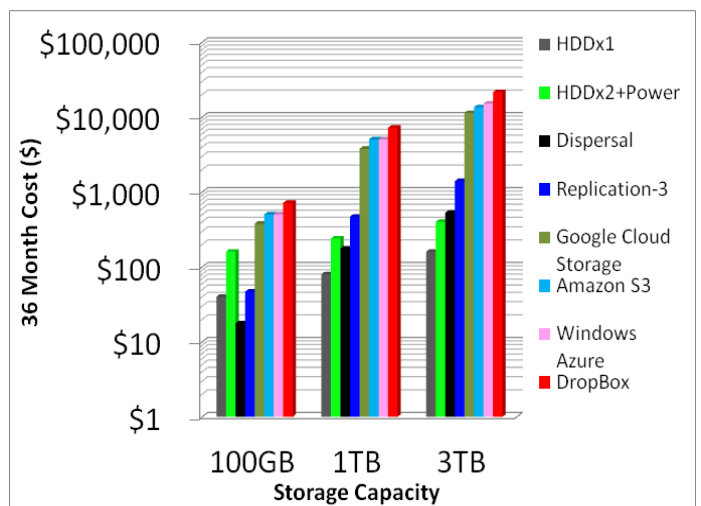
Results

The comparison between the cost estimation of dispersed storage system and the price of different cloud storage providers are very promising. The dispersed storage provides the most cost effective price comparing to other systems. As the storage size gets bigger the price of dispersed storage gets closer to the raw storage price. Comparing to the price of cloud providers, the price of dispersed storage is similar to Amazon S3's price. As the storage size gets bigger and reaches to 1 Petabytes, the price of the dispersed storage system gets very cheaper than the Amazon S3 which is the cheapest provider in that order of size. At the storage capacity of 10 PB, the dispersed storage costs around 10X cheaper than the Amazon S3 cost. This gap shows that there is a good opportunity for new investors to compete with the current pioneer vendors of cloud storage.