How Careful Do You Need To Be With Cloud Storage? - Bandwidth

Posted by George Crump @ 03:26:PM | Jan,28, 2011

Cloud storage continues to get treated as this extra special project that you need to be more careful with than other projects you would undertake in the data center. Our belief is that you need to be no more careful with cloud storage than any other IT project and in some cases you can be a bit more risky. In this entry we will address some of the concerns surrounding bandwidth.

In our last entry in this series we discussed the issues surrounding security and cloud storage adoption rates. The hybrid appliance (either via virtual or physical hardware) helps address this by making sure that data is encrypted throughout. The same appliances also help significantly with the bandwidth issues. In our recent testing of cloud vaulting with a very slow internet connection we are able to sync our local backup copy with our remote copy in a matter of minutes. In that time we have a complete off-site copy of our data. How can this be?

First, any time you are dealing with an existing large data set, like the type that backup will generate, most cloud vendors are going to have some sort of seed mechanism. In the case of our test a 1U appliance full of drives was sent that we populated with our full backup. We shipped the data to the cloud data center and it was seeded into their storage system. Subsequent backups only now have to send changed data via our very slow internet connection, and the process is very workable.

In primary storage applications there may not be a need for a seed step. If data is added incrementally as the application comes online most cloud transfer processes will be able to keep pace. In either case the hybrid appliance has a local high speed storage device so users and applications don't feel the impact of moving the data to the cloud.

Second, beyond the seeding process most cloud appliances will perform compression and deduplication on the data before it is sent. In most cases the combination of the two technologies can reduce the data that has to be transferred across the internet by as much as 80 to 90%. Again this process is often secondary after the initial data set has been written which means the local storage performance is not affected.

Third, appliances help resolve the issue of protocols. The communication method that most cloud appliances use is optimized for internet traffic. There are no mounting of a CIFS or NFS share and copying the data. Not surprisingly, in our testing every cloud appliance outperforms a file share copy and most outperform even an FTP copy.

The other concern about bandwidth has to do with recovery. In this area, as we discussed in our recent article "What Can You Really Do With Cloud Storage?", primary storage solutions may have an advantage. That is because on failure they typically have a better understanding of what data needs to come back. In their use cases data is recovered a file at a time as it is being accessed. So from a "seeing the data" perspective recovery can be almost instant, then there is some lag as individual files are being recovered but as we showed in our Cloud DR Test, that performance should be more than acceptable for many applications.

Backup recovery may take more work. Depending on the solution there may be a need to Fed-Ex you a recovery disk, although some solutions are learning how to leverage deduplication in the recovery process or be intelligent enough to only restore the most active data set first.

In our final entry in this series we will cover what you should do right now with cloud storage. How should you start, what precautions should you take and how should you roll the solution into production?

Track us on Twitter: http://twitter.com/storageswiss

Subscribe to our RSS feed.

George Crump is lead analyst of Storage Switzerland, an IT analyst firm focused on the storage and