Protecting Against Data Loss

A look at the backup practices of web hosting companies
and what you can do to help safeguard your data.

Table of Contents:

Section One
Introduction:
Do not assume the availability of server backups

In today's web hosting environment, with hosting plans often boasting thousands of gigabytes of storage, it is important to consider and understand how these plans are being backed up to protect against data loss and corruption. It is not enough to assume that a web hosting company is making backups on a regular basis, as you do not want to be asking these questions after data loss has occurred. In addition to hosting companies performing reliable backups, it is important that end users also keep their own local backups in the event the host backup is unavailable. It is also important to note that some hosts will not restore data that was lost because of an end user action (accidental deletion or overwriting of files by the client).

This whitepaper will examine the various methods for backups that exist today, and outline a "good practice" backup strategy for both hosts and end users. The whitepaper will also include some good questions you should ask your hosting provider to fully understand how backups are handled.

Section Two
Is RAID a viable backup solution?
The difference between redundancy and backups

A few years ago RAID based systems were the exception in the web hosting industry. They were very expensive, and their reliability was suspect in many cases. Today, RAID should be considered a standard for any web hosting environment. However, it is very important to note that RAID is not a valid backup method. It is only a tool for increasing reliability in keeping your data available and online. RAID is susceptible to data corruption, deletion, or malicious activity just like any single drive system.

What is RAID? RAID is an acronym that stands for “Redundant Array of Inexpensive Disks”. It is a system whereby multiple disks are combined into a single (sometimes larger) capacity storage device that can also offer increased reliability in many configurations. The main types of RAID you’ll see in hosting environments are:

RAID-1 - Typically 2 drives with mirrored data that provide increased reliability to allow for the failure of one drive, while also increasing performance on reads from the disks.

	Drive 1	Drive 2
Block 1	1	1
Block 2	2	2
Block 3	3	3

RAID-5 - Typically 3 or more drives using striping and parity to offer the storage capacity of 2 drives while allowing for the failure of one drive.

Section Two Continued…

	Drive 1	Drive 2	Drive 3
Block 1	1	2	P
Block 2	3	P	4
Block 3	P	5	6

RAID-10 - Typically 4 or more drives using both mirroring and striping to provide the storage capacity of 2 drives while offering performance increases and being able to sustain the failure of one drive.

	Drive 1	Drive 2	Drive 3	Drive 4
Block 1	1	1	2	2
Block 2	3	3	4	4
Block 3	5	5	6	6

Some hosting companies will tout their RAID systems as a backup method, but this is a dangerous system to rely on for data integrity. All it takes for the complete loss of data is for the RAID array to fail (hardware RAID card fails, 2 or more drives fail at the same time, electrical fault damages multiple drives in the array) and all the data stored in the array is gone. Partial data loss can occur if some data becomes corrupt, as this corruption will permeate to all the drives in the array, rendering the data unusable. A compromised server could allow hackers to destroy the data just as if it was stored on a single drive. Even an error from a web hosting employee could cause this data to be lost or corrupt. RAID does not protect your data from these types of events, as the primary purpose of RAID is to provide increased reliability by redundantly storing the live data on the server.

RAID is an essential tool for increasing reliability and performance, but it should not be considered a backup method to safeguard your data.

Section Three
Types of backup methods:
Drive based local and remote options

There are many different types of backups that can be performed at the server level. Ten years ago tape based backups were a standard for many hosting organizations. Tape backups could store large amounts of data, and the tapes could be removed and stored securely in another physical location for disaster recovery purposes. These days, tapes are just too slow for both backing up and especially for restoring large amounts of data. Most tape systems are also sequential read/write only, meaning data can only be restored in specific chunks, and getting to those segments can be a lengthy process. Testing the physical integrity of tapes is also problematic for many systems.

In today's hosting world, disk to disk backups have become the standard for data backups. Disks are faster and large enough to handle backing up large amounts of data quickly and reliably, while allowing for faster restores. However, just as there are many different versions of RAID, there are many different types of disk to disk backups and each has its own advantages and disadvantages.

The most basic setup for a disk to disk backup is to have a second or multiple hard drives attached to the same server that is used for holding backups of the primary drive(s). This allows for faster backups as there is no networking latency to slow things down. There are a number of software based solutions that determine how these backups are actually created and managed (and too many to discuss here), but most systems take either a full backup of all data at regular intervals (daily, weekly, monthly), or use a modified incremental backup system to store multiple backups over a period of time. If the backups are stored properly, they can allow for a full

Section Three Continued…

recovery of all data and the restore of specific data from any number of intervals.

Local disk to disk backups do have some disadvantages. Overall system load can be higher when the backup runs, as the server must manage both sets of drives during this period. Also, since the backup system is attached to the server, electrical faults can affect the integrity of the backups as well as the live data. With local backups, a compromised server could allow a hacker to wipe out both the live data and backup data at one time.

Another form of disk to disk backups is the use of a remote server in the same datacenter that is used to store backups of one server or multiple servers. This eliminates the problems with a compromised server losing both its live and backup data, and can lower overall system load on the primary server during backups. Also, if the remote backup server is configured so that it can access the server it's backing up, but not vice versa, it eliminates the vulnerability of the main server being compromised and all live and backup data being deleted.

Since the remote server is located in the datacenter, backup speeds will be higher on a LAN, resulting in less time to make backups. These remote backup servers can also employ a RAID-10 setup to increase performance and reliability of the backup server. Remote backups in the same datacenter are susceptible to a datacenter wide event such as an electrical fault or natural disaster that could wipe out both the live and backup data.

Similar to the remote disk to disk backup in the same datacenter is a remote backup to an off-site server that is in a separate physical

Section Three Continued…

location. This type of backup avoids potential issues with electrical faults or natural disasters destroying both the live and backup data at the same time. It is a great method to add for disaster recovery planning, and can be an emergency method for moving servers quickly to another datacenter if the need arises. Since the backup server is now in a separate physical location, this can introduce slowdowns in the speed with which a backup can be completed, so this must be accounted for when designing such a system to compliment local backups and remote datacenter backups.

Section Four
Problems with the use of large drives / arrays: Large hosting plans can affect your uptime

For budget hosts, utilizing less costly hardware and cheaper bandwidth providers is a necessity to be able to offer the low prices for hosting. Unfortunately, this can cause some serious issues for their clients when components fail. Using less robust hard drives (IDE vs. SCSI/SAS, no RAID, etc…) can result in more hard drive failures over time. When a hard drive fails data can be lost, requiring restoring from backups. This can result in hours of downtime as well as the loss of recent emails, orders, and other critical data. The length of a drive restoration from backup increases substantially with the size of the drive. As a result, the downtime incurred from a drive failure and subsequent restore is generally longer with budget hosts that utilize very large drives and offer hosting plans with enormous amounts of disk space.

Some budget hosts also may not be maintaining proper backups, so a hard drive failure could cause the loss of all data on the server, which could be disastrous for many websites. Other hardware failures can be more common for lower quality hardware such as those affecting switches, UPS units, power supplies, and RAM. In any case where inferior hardware is implemented, it will be more likely to fail over time resulting in downtime and loss of revenue for the client.

Section Five
The cost to implement an effective backup system: Why budget hosting can lead to backup deficiencies

Even with the price of drives dropping as technology improves, the cost of an effective backup solution is significant. Once you take into account the drives, equipment, software, and most importantly human management of such a complex backup system, you will quickly realize that the cost to manage all of this far exceeds some of the budget offerings from hosts trumpeting hundreds of GB's of storage for unbelievably low monthly charges.

Let's take a look at a robust setup and backup system and see what's involved:

- Single server solution using enterprise level 146 GB SAS drives in a RAID-10 array (4 drives)
- Backup drive (300 GB SAS) locally attached – 1 drive
- Software to manage local backups including daily backups with 7 day incremental backups available
- Remote backup server in same datacenter utilizing 750 GB enterprise SATA II drives in a RAID-10 array (4 drives)
- Bare metal restore software running on remote backup server to manage hourly backups and the ability to one click restore a server from empty drives in the event of a failure of the main server.
- Remote off-site backup server for disaster recovery utilizing 750 GB enterprise SATA II drives in a RAID-10 array (4 drives)

The above setup would require a minimum of 13 hard drives (5 of those being expensive SAS drives), 2 servers, 3 hardware RAID controllers, and multiple software solutions working together to

Section Five Continued…

provide storage space for only 292 GB of data. If a plan offers each customer 500 GB of storage space for only $9.99 per month, you can quickly see the host must be cutting corners somewhere with regards to the backup and retention of client data.

Section Six
Questions to ask your host:
Fully understand your host's backup procedures

There are a few questions you should ask your hosting company with regards to backups and data integrity:

- How often are backups made?
- How long are backups saved for recovery purposes?
- Where are the backups stored?
- Can I request a file be restored from a backup if needed? How far back?
- Do you have a disaster recovery plan? How would data be restored?

Section Seven
Steps you can take to protect your data:
Self reliance is an important safety net

Even if your hosting company has a bulletproof backup system in place, you should *always* make your own backups of your data just in case. It sounds simple, but there are a myriad of items to consider including database data, passwords, e-commerce data, statistics, email settings, and many other variables.

At the very least, every file you upload to your hosting company should be backed up on your local computer. This will allow you to rebuild your web pages in the event data is lost by your host. If your website has scripts and programs that create images, text files, and webpages, you should regularly download these files as well.

If your website makes use of a database, you should setup a regular schedule to download your database data (SQL dump) to your local computer as well. If the database server is corrupted on the hosting end, you can restore from your last SQL dump file rather quickly. Most hosts have a web based tool available that allows you to download your database data into an SQL dump file.

For e-commerce sites, you will want to download order data, sales data, and other pertinent data that is valuable to your organization. This will allow you to either re-create the data if it's lost online, or at least have an archive of data to reference orders and sales.

These local backups are often your only recourse to restore your site if a host ever loses your data and cannot restore it from their backups. If the data has any importance, you should make sure you have a system in place to perform regular backups. There is software that can help automate this, or ask your host for suggestions.

Overview of the LexiConn Backup System

LexiConn's approach to backing up client data is a multi-faceted system that takes into account the ideas presented in this whitepaper. The process has redundancies built-in to ensure data can be restored under a number of different disaster scenarios, while minimizing the impact these backups take on live servers.

Backup processes include:

- LexiConn utilizes hardware RAID-1 with SAS hard drives for reliability of data. Although not a backup method, it does help increase overall uptime to allow for the failure of one hard drive without causing a server to crash.
- All shared hosting servers utilize a local SAS backup drive that takes daily backups of all files, and stores these backups going back five days. This allows a file to be restored from any day going back 5 days, and allows a full restore from the local drive if needed. A local drive restore would only be a few hours in most cases.
- A remote backup server utilizing RAID-10 with four hard drives is used along with enterprise class software to take 3 hour backups of shared servers. These 3 hour backups can be used to perform a bare metal restore of a server in the event the data on that server was lost.
- A remote backup server located in a physically separate datacenter in another part of the country utilizing RAID-10 with four hard drives is used to take daily backups of all shared servers. These backups can be used for disaster recovery purposes in the event the datacenter where the servers are located was damaged or destroyed.

Summary:

The importance of server/site backups is never fully appreciated until disaster strikes and it is discovered that the only available backup is corrupt, out of date, or just plain non-existent. Failure to understand and confirm your host's backup policies and complement them with your own local backup procedures puts your business at great risk.

It is also essential to understand that in the event of data loss and drive failure, the size of your host's packages (and in most cases the corresponding size of their server hard drives and RAID arrays) can affect the speed in which a restore can take place and increase the time that your website is unavailable. While you may not be utilizing the 10 gigabytes of space that comes with the package, many other clients on the same system may be and that in turn will negatively affect how quickly a restore can be completed. In that sense, selecting the largest possible hosting package with an inexpensive provider may not work to your benefit, especially if you have no intention of using the majority of the web space allocated.

Lastly, implementing a redundant backup system that relies on local and remote backups is not a small undertaking, incurs significant costs, and requires regular monitoring. Backups provide the type of insurance that an online enterprise must have, so always be diligent in choosing a web host that will provide the type of backup safety net required to avoid the consequences of potential data loss.

Since 1995, LexiConn has been recognized within the hosting industry and by clients alike, as providing top of the line web hosting services and ecommerce solutions. Our experience over the years has taught us what is required to maintain this high level of service, which has solidified our reputation and performed a key role in our clients' success.

We would welcome you to contact us regarding your hosting and ecommerce needs as we believe you will be able to quickly identify the areas in which we can help your business succeed. We value the success of our clients and work personally with each one, which is another thing that you will not find in the budget hosting industry.

About LexiConn:

LexiConn is a privately held company located in Colchester, CT. Founded in 1996, LexiConn currently hosts over 4500 web sites in over 40 countries. The company provides shared and managed server hosting, ecommerce and database solutions, as well as domain registration services. More information is available at www.lexiconn.com.

LexiConn Internet Services, Inc.
63 Norwich Avenue, Suite 102
Colchester, CT 06415

Toll Free: 800-817-6881
Fax: 860-537-9675
Email: sales@lexiconn.com