Abstract: Data protection is a critical aspect of all computing environments. Over the years, it has changed with the goal of protecting an enterprise’s data from device failure expanding to encompass software failure, human error, site outages and theft. Furthermore, as corporate data has become the life blood of organizations, recovery windows have shrunk and an increasing number of organizations are taking an integrated approach to data protection and high-availability to ensure minimal planned and unplanned downtime or even continuous operations.
Data Protection Strategies In Today’s Data Center

Your company is spending too much on data protection -- and yet you’re probably not actually protecting your data as well as it should be protected. Or in other words, your company could probably spend far less on data protection and actually protect its data better.

What is Data Protection?

The term *data protection* includes everything that an IT department does in order to ensure that the data that the company needs is available when it is needed -- and that it is *not* made available to those entities that should not be given access. Data protection systems include high availability (HA), backup, disaster recovery, archive and security systems. Each of these systems is attempting to give a particular business unit access to the data that it needs within a timeframe acceptable for that particular business unit.

For example, customer-facing business units need systems such as websites and databases that are always instantly accessible; downtime with such systems creates significant losses for the company. High-availability systems are designed to ensure such access. On the other end of the spectrum, consider a company’s legal department that has been given a discovery motion for all e-mails created during the last three years that match particular set of criteria. This data is important, but it only needs to be provided within a few days (or weeks, depending on the case). This is the purpose of an archive system. Between these two extremes are the backup and disaster recovery systems, whose job it is to recover data when it is damaged or lost. Finally, information security systems ensure that those who should not have access to the data are not granted such access. While there are a variety of intrusion protection and detection systems to consider, this paper will specifically address encryption, as it is often the last line of defense. If all of the intrusion protection and intrusion detection systems fail, encryption can stop an unauthorized user from actually reading or using the data they’ve stolen.

Why does it cost too much?

The reason that most companies spend too much on data protection is that they use systems that do not work together. In some cases, there are systems that could cooperate in ways that could save money and/or increase the level of data protection. There are even situations where companies are using systems that are actually working against each other, either completely ameliorating the purpose of a given system or making it cost much more than it should. In order to explain how this could be the case, this paper must first take the reader through the different risks that data protection systems are designed to mitigate. It will then explain the different types of data protection systems that are meant to mitigate these risks, while also explaining the systems that it should cooperate with or work against.
What are we protecting it from?

Most risks that plagued IT departments in the early days of computing have been completely mitigated by various forms of data protection, so it is necessary to first explain what these various risks are. It is also important to remember that each of these risks that are not mitigated can result in significant damage to a company. They can cause downtime that results in loss of current or new business. Loss of some or all customer data can result in monetary losses, significant changes in the perception of a company, and even a company going completely out of business. This is why these risks must be mitigated.

The most common risk that continues to plague IT departments is typically referred to as user error, although sometimes the "user" is often the system, network, storage, or database administrator. Humans make errors; they’re what make us human. Humans delete files that they did not intend to delete. They also accidentally enter incorrect information or instructions into an existing document or database. Perhaps they perform a global replace of one phrase with another phrase, use the wrong phrase, and then save the result in place. Perhaps they are an administrator and they accidentally drop a table in a database, delete a directory on a server, or incorrectly reconfigure a storage array.

Computer systems also sometimes do not do what they are supposed to do for various reasons. An electrical spike can cause a computer or storage system to write completely worthless data over valuable data. Storage media does not always do what it’s told either; every time a computer system writes a bit (a 1 or a 0) to storage media (disk, tape, etc.), there is a statistical chance that it will write a 1 when it meant to write a 0, or vice versa.

The next risk that must be mitigated is a minor equipment failure, such as an individual disk drive or other system component. The risk of this happening is actually quite high. There are many individual mechanical components used in IT systems not only prone to fail but also single points of failure that can take the entire system down. This is like having a weak link in a chain. However, the problems caused by minor equipment failure can almost be completely mitigated through the use of redundancy -- more on that later in this paper.

Excessive equipment failure is defined as the failure of more than one component in a single computer system, such as the failure of multiple hard drives within a single storage array. Such failures are often caused by outside factors such as electrical spikes or electrical fire localized to an individual system. A human error can also cause excessive equipment failure. Consider, for example, the data center manager who removed one too many floor tiles, causing the collapse of a portion of the floor and the resulting failure of multiple computer systems. This risk can also be mitigated through the use of redundancy, but now the entire system must be redundant -- not just individual components.
Even worse than the loss of multiple pieces of equipment is the loss of multiple computing systems or an entire data center, such as what happens in a natural disaster, large-scale fire, or even an act of terrorism. It is true that companies that suffer such losses are granted a certain amount of leeway, so damages to the company’s reputation can be minimized by publicizing the unexpected event that caused the large-scale failure. However, it is also true that companies that can survive such an event with minimal downtime or impact to customers can experience a significant increase in positive brand equity as well. Conversely, not all companies survive such events. Examples of both extremes can be found in the events of September 11, 2001. Some companies saw what was happening after the first plane hit and immediately failed over to their standby data center in another location. Other companies had their standby data center in the second tower and ceased to exist the moment the towers fell.

Finally, there are the risks of loss of control and inappropriate access. Loss of control is when information that was protected by physical and/or electronic security is no longer protected, such as when a backup tape is lost. The risk here is not so much that the data is no longer available to the company, but that it is possibly available to people outside the company. Inappropriate access is when data inside the company is accessed by hackers outside the company. The damages to a company from such an event can be quite significant, and can include a significant damage to the company’s brand equity, loss of critical intellectual property to a competitor, publishing of intellectual property on the Internet, and the direct costs of notifying and/or compensating customers whose personal data was compromised. (The overwhelming majority of US states require notification of the news media when a company loses control of a customer’s personal information.)

Data protection methods

Data protection methods can be classified into three categories: restore, recover, and overcome. Restore methods can restore your systems’ functionality after an event caused by one of the risks above, but they require a significant amount of time to do so. This is because for the affected systems to be restored back to their pre-event state a significant amount of data must be copied back before they can be used again. Just like when a politician offers to restore the glory of a city/state/country, there is an assumption that this restore will take some time. Recovery methods allow your company or business unit to continue functioning after such an event, but would require a minor disruption in service before doing so. Think of a baseball player that bobbles, but eventually catches a ball, after which someone might yell, “Nice recovery!” Finally, methods that allow you to completely overcome an event are typically the most expensive, but they would allow your company or business unit to continue functioning uninterrupted through any kind of event described in the previous section. This would be like a boat in a James Bond movie that automatically converts to a submarine when the archenemy starts shooting at the boat. The mission continues without a hitch, as Bond has overcome the event.
Restoring after an event

Backup and restore tend to be the "table stakes" of data protection because a company that cannot restore its data after a major negative event will simply go out of business. Traditionally, this functionality is provided by a backup and recovery system that does one of the following:

1. backs up to tapes that are sent off site
2. backs up to disk, copies to tapes that are sent off site
3. backs up to disk, replicates to other disks that are located off site

Regardless of which method is chosen to ensure that copies of data are stored away from the servers they are protecting, the restore method is always the same: copy significant amounts of data from the backup medium to the system to be restored. This is why it is said that very little has changed in backup and restore in multiple decades. Yes, many companies have switched from tape to disk as the initial target for backups. Yes, the industry has gotten a lot better at handling special data types such as databases, virtual servers, and other applications. But in the end, restore methods still require significant amounts of downtime to perform their function.

Besides the obvious limitation of requiring downtime for restore, there are other challenges with even the most advanced backup and recovery system. Consider, for example, the advent of disk systems with deduplication. (Deduplication is a technique that identifies and eliminates redundant blocks of data between different backup sets, significantly reducing the amount of disk needed to store backup data. This makes remote replication of backups over relatively small WAN links possible.) This technology has done quite a bit to make things much easier for today's backup administrator, but it is fundamentally incompatible with certain types of data. Applications, such as video, seismic processing and large number crunching, all create significant amounts of brand-new data every day, and will not be able to be replicated over a WAN link. Customers enabling database encryption or compression will find their deduplication system rendered ineffective due to its inability to identify any duplicate patterns to eliminate.

A business must be able to restore its data after an incident. The question for the reader is whether or not a traditional backup and restore system -- even one that is enhanced with disk -- is the best way to do that.
Recovering after an event

Recovery systems are those that allow you to recover from an event with minimal downtime. They are very different from their backup and restore counterparts, and they all have a few things in common.

The first thing they have in common is that they "backup" data incrementally throughout the day as it changes, as opposed to the backup and restore systems discussed previously that perform this operation as a batch process once a day (usually at night). This means that all of the money that would otherwise be spent on trying to complete the backup within the "backup window" can be spent elsewhere. Backing up throughout the day also creates many more points to which the data can be recovered. Where a backup & restore system would only be able to restore your data to last night’s backup (assuming it was successful), a recovery system can recover your data up to minutes or even seconds before the event occurred. This translates into much smaller amounts of lost data, which directly translates into much smaller financial losses after an event.

The second thing recovery systems have in common is that they store data in its original format. In contrast, each backup and restore system stores its data in its own backup format. There is no advantage to storing data in a different format. In fact, backup systems storing data in backup format is what classifies them as a restore system rather than a recovery system; pulling data out of the backup format and putting it back into the primary system is what takes so much time.

Because recovery systems store data in its original format, the "recovery copy" can instantly be used as the primary copy when the primary copy is damaged. Switching from the primary does require a few moments of downtime, as the server or application is instructed where it should look for its data; however, this downtime is measured in a few minutes -- not a few hours. The reader should consider the amount of money their company loses each minute a mission-critical application is down, and then imagine how much money could be saved by using a data protection system that measures its downtime in minutes rather than hours.

The types of applications that meet the descriptions in this section are continuous data protection (CDP), and near-continuous data protection (near-CDP). A CDP system replicates changes in data the moment the changes occur, and every change is stored in a log in the replicated copy. A near-CDP system takes snapshots periodically throughout the day (e.g. once an hour) and stores those snapshots and the original data in the replicated copy. A CDP system can recover your system to any point in time; a near-CDP system can recover them to the last time a snapshot was taken. CDP systems can be generic and used with any type of application, or they can be an application-specific solution that is closely integrated with that application. Near-CDP systems tend to be integrated with a storage system. This is the perfect example of how the purchase of the right storage system can significantly enhance your data protection scheme. If you have a storage system that supports snapshots and
replication, you’re halfway to having some of the most advanced recovery possibilities available.

**Overcoming an event**

What if a company could continue functioning without any interruption regardless of the type of event that takes place? No data would be lost, no customer would be told to call back later because the systems were down, and business would continue as if nothing had happened.

First consider an event that renders an entire physical campus unusable. It is possible to live the dream described in the paragraph above in such a scenario, but it does require a more comprehensive approach to data protection. This isn’t just about data; it’s not even just about servers or other systems. One must also have multiple teams of people using multiple sets of systems, each of which can take over for the other in the event of a disaster.

The less downtime you want in the event of a disaster, the higher up you must go in what technical people call “the stack". You must first, of course, have a highly available computing and storage system that synchronously replicates its data to more than one location. You must also have an application that can automatically detect that one of the locations where data is stored is no longer available and automatically begin acting on behalf of the other location.

Secondly, consider another type of "disaster." Consider the public-relations disaster of losing several backup tapes containing personal information for customers. This indeed is a disaster that can cost the company millions of dollars, unless you can overcome it. If those tapes were encrypted in the first place, they are rendered unreadable by anyone outside the organization. Current laws would not even require any type of notification, allowing the company to overcome this event as if it had never occurred, saving millions of dollars in the process.

Now consider a completely different type of event: a large electronic discovery request. The company has been sued and the plaintiff is requesting several years’ worth of e-mail. This also would be a "disaster" to most IT departments because they are not using the proper techniques for preparing data discovery. The use of the proper type of archive software, accompanied by a cost-effective medium for storing the data long-term would allow the IT department to issue a single query in order to satisfy this request. What could have become a several month process costing millions of dollars instead could simply be an afternoon’s work by a single employee. Even these types of events can be overcome as if nothing happened.

This is why a company must look at all of its data protection requirements before making architectural decisions.

---

1 A term referring to a stack of systems starting with lower level devices such as storage, followed by the computing and networking systems, with the application being at the highest level.
Saving Money

The way most companies waste money with data recovery is to use separate and incompatible tools to provide restore, recovery, and overcoming functionality. This is often referred to as the "belt and suspenders" approach. Once the company has identified which levels of functionality it wants for each application, it should look at and see if functionality provided by higher levels of data protection supersedes functionality provided by lower levels of data protection.

For example, if a company is using continuous or near-continuous data protection software, and that software meets all restore and retention requirements that the backup must meet, the company should consider not using the traditional backup system for that application. Similarly, if the solution that is providing highly available functionality also has recovery, restore, and retention functionality, does the company really need the lower-level systems for that application as well?

Another way that companies waste money is to buy data protection options that provide more functionality than is required for the application. Data required by archive and e-discovery applications, for example, do not require the random access times that disk offers, so storing them on disk for long periods of time does not make financial sense. Add that tape is actually more reliable at storing data long term and you have a very compelling case for tape-based archives.

Finally, this paper has also explained how the use of tape encryption and proper archive software can also save the company quite a bit of money if used on the right applications. And yet, the majority of companies do not use such functionality because of either fear of change or the perceived additional cost of such tools.

Summary

Companies should first consider their business requirements that drive data protection architectural decisions. Once they have done that, they should select the appropriate level of data protection for each application. Finally, they should examine all levels of data protection at the software and hardware level for each given application to look for redundancies, overprovisioning and possibilities for savings.