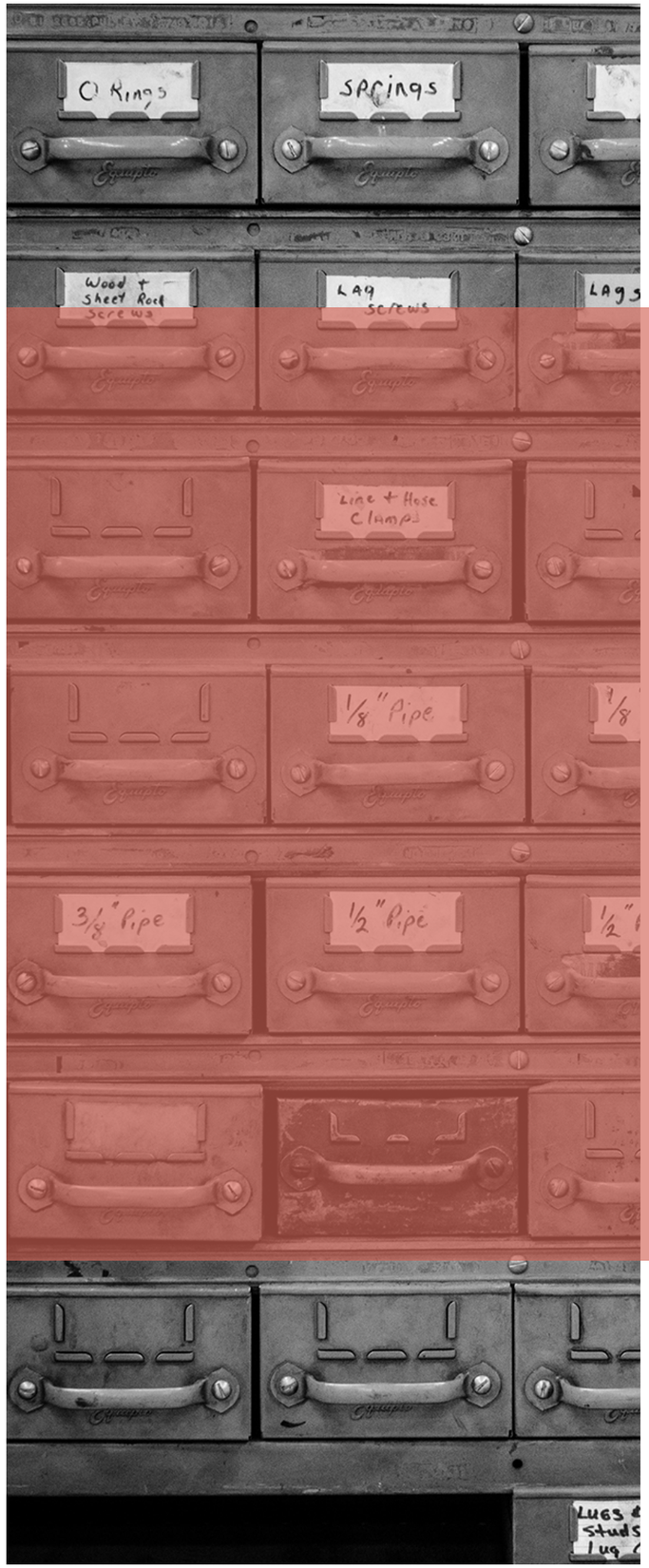
DATA PROTECTION 101

What You Need to Know About Ensuring the Security and Integrity of Your Data



Contents

Data Protection is a Big Deal

How many times have you heard or read the following statement? Data is the lifeblood of every organization. If that's true, why are so many companies lackadaisical regarding their data protection strategies? The reasons are many, from lack of support from the C-suite to simply not understanding what data protection is or its importance. Nonetheless, the need for a solid data protection strategy continues to intensify.

Every day, an estimated 2.5 quintillion bytes of data is generated — and there are no signs of things slowing down. Data comes from an increasingly vast and diverse array of sources: digital photos and videos; sensors that gather everything from climate information to the number of steps taken per day; posts to social media sites; cell phone GPS signals and much more.

Complicating the matter is the constant barrage of threats ranging from malware to human error — and the plethora of tactics to combat them. It's hard to stay on top of data protection best practices, much less the most recent advances. Even the sheer number of relevant terms can be overwhelming. Is a backup and recovery plan the same as a disaster recovery plan? Is disaster recovery the same as business continuity? What's the different between backup and replication? Where do things like encryption and deduplication fit in?

In this eBook, we'll answer these questions and more.

The Language of Data Protection

A discussion on data protection first requires a review of some of the key terminology. While this list isn't fully comprehensive, it does provide some of the basic terms you should know in order to understand data protection and to develop and implement a data protection plan.

Backup is a component of data protection and refers to the process of periodically saving data in a secure on - or off-site location and bringing it back when it is needed.

Business Continuity (BC) refers the processes and procedures that ensure a business can continue operations during and after a disruptive event.

Business Impact Analysis is a procedure used to collect information on a wide range of areas from recovery assumptions and critical business processes to inter-dependencies and critical staff that is then analyzed to assess the impact a disruptive event may have.

Data Archiving entails moving data that is no longer actively used to a separate storage device for long-term retention.

Data Protection is the process of safeguarding data from corruption and/or loss. It includes both the operational backup of data and business continuity (BC) and disaster recovery (DR).

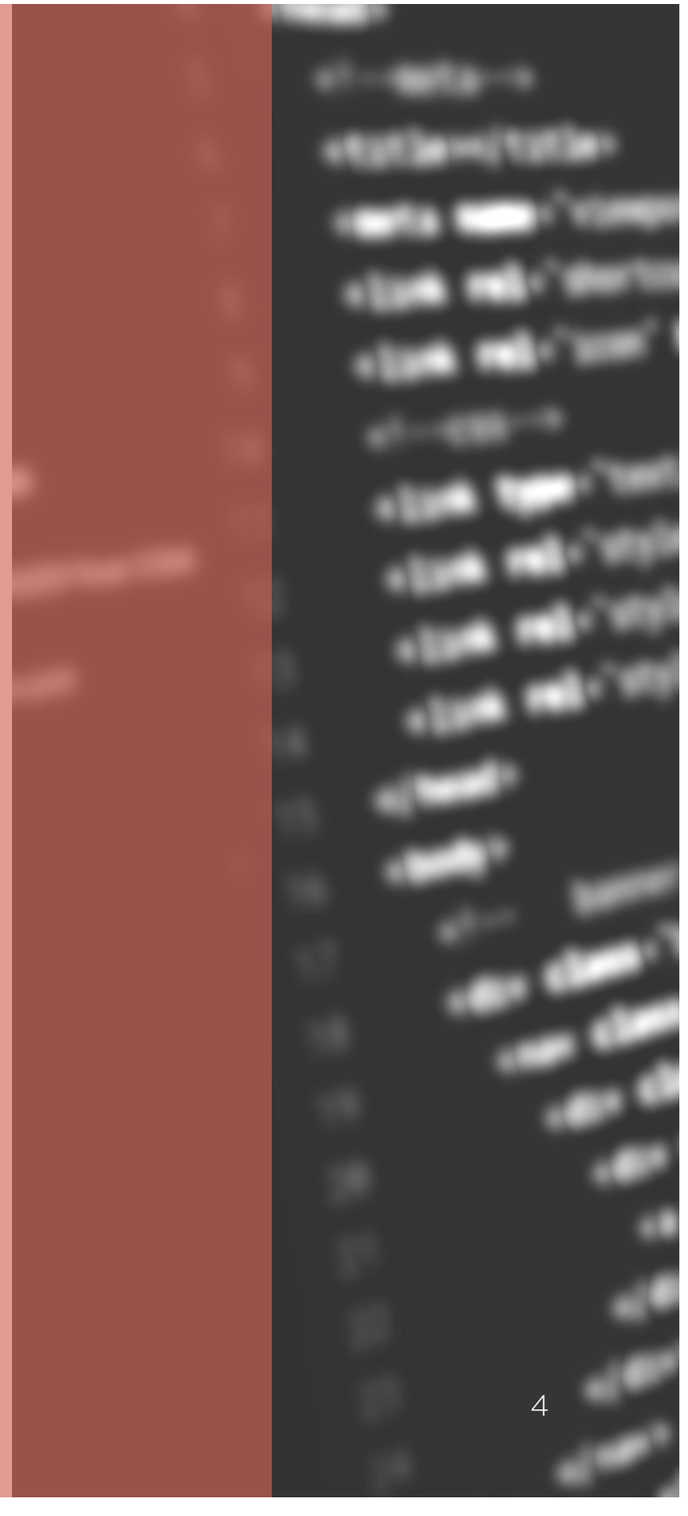
Disaster Recovery (DR) is the process for replicating your entire IT environment, including data, and then making it available after a disruptive event when your primary environment is unavailable.

Mission-critical is used to describe data that is essential to the functioning of your business and its processes.

Recovery Point Objective (RPO) is the point in time to which a firm must recover data as defined by the organization. The RPO dictates which replication method will be required.

Replication refers to copying data from one location to another either immediately or with a short time delay. It's similar to backup in that both create a secondary copy of data. The difference is that with replication, any data corruption or user file deletion is immediately or very quickly replicated to the secondary copy. That makes it ineffective for backup.

Recovery Time Objective (RTO) is the duration of time and service level within which a business process must be restored after a disruption in order to avoid unacceptable losses. RTO begins when a disaster hits and does not end until all systems are up and running.



BC and DR: The Subtle Differences

Much of the confusion over data protection starts with the BC/DR aspect — or rather the difference between the two concepts. BC enables a business to continue their operations while DR refers to recovering from a disaster. Both involve data and are about not letting an event disrupt business operations any longer than necessary. However, BC is business-centric while DR is data-centric.

BC entails a comprehensive strategy that will allow a business to function during and after a disruptive event occurs. Note the word "during," because you want — and need — your business to continue operating normally at all times.

Since most businesses rely heavily on data for their day-to-day operations, a BC plan typically requires data to be accessible with little or no downtime when a disruption occurs. For example, if a power outage takes out your on-site production servers, your data and applications "fail over" to a system located elsewhere. If all goes as planned, end users won't even know a disruptive event took place.

BC also goes beyond data recovery. It focuses on what data and applications are most important for keeping a business running. All data and applications are important, but the most essential get priority.

DR is a subset of BC, and focuses on getting essential data and systems up and running after a disruptive event. The emphasis here is on "after" as you want to access your data after a disruptive event even if your primary environment is unavailable.

As with BC, some data is more important than the rest for DR, so it gets recovered first. The speed at which it needs to be recovered is also important. Some data may need to be immediately recovered after a disruptive event; other data can wait days or even weeks. How quickly data is needed will influence the choice of replication and recovery tactics.

Data Backup

A backup is a copy of data, and is used as a safeguard against unexpected data loss and application errors. If you were to lose your original data, you can use the backup to make it available again. This differs from data archival, which is used to protect information that isn't needed for day-to-day business but must be retained for a long period — often for compliance purposes.

Backups are scheduled and entail making copies of data files at intervals. It could be hours or days between the copies, depending on your business requirements. The data copies are saved to a physical hard drive, tape, disk or to a virtual tape library (VTL), which are kept offsite. The techniques used depend on the type of data you're backing up and how convenient you want the recovery process to be. The basic types of backups include:

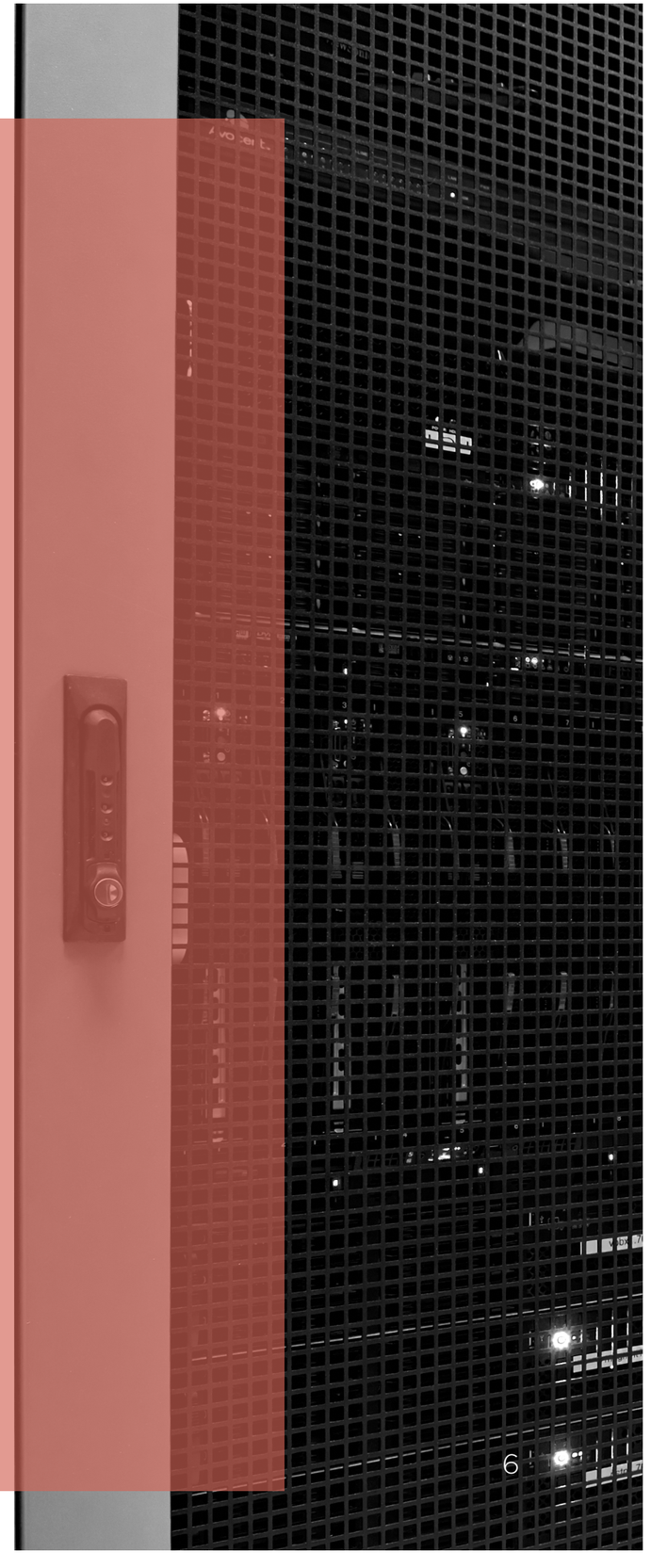
- Normal/full backups
- Copy backups
- Differential backups
- Incremental backups

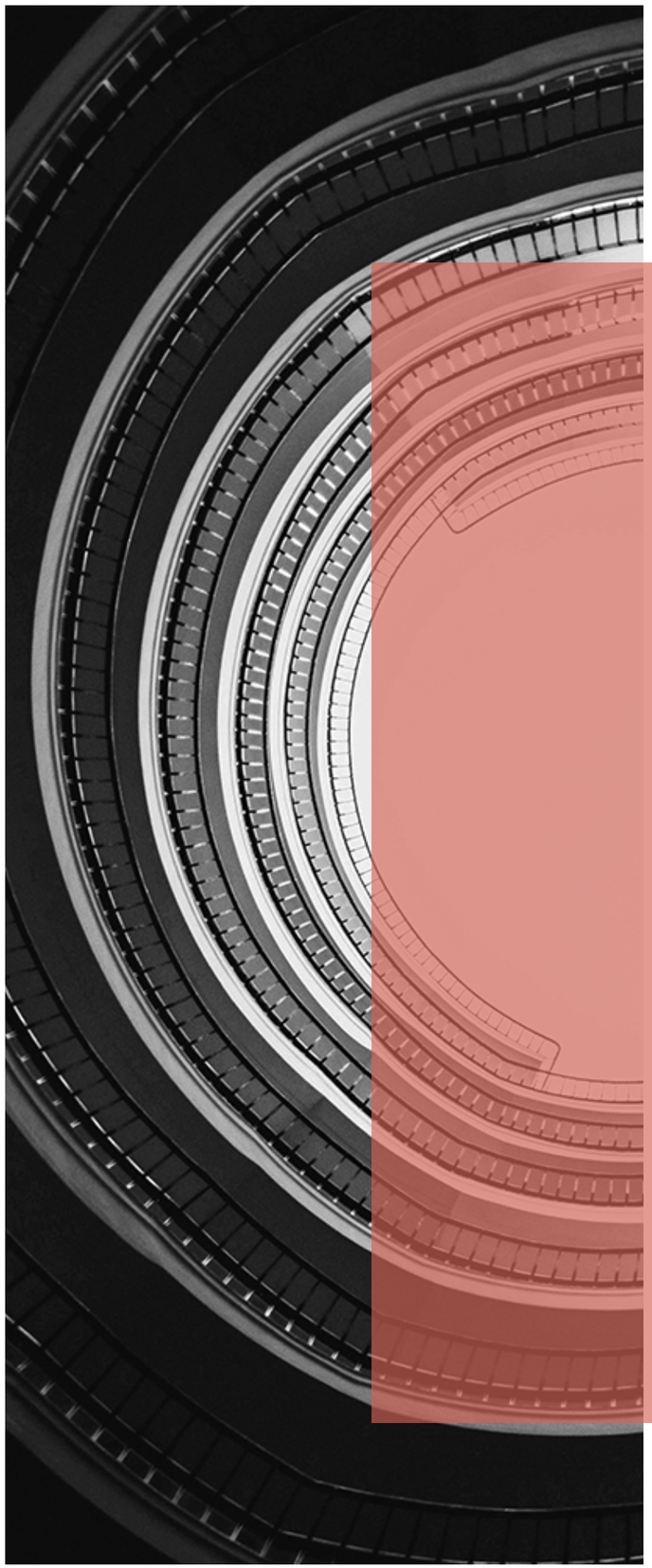
While you can do your own backups, many organizations hand off the responsibility to third-party companies. These vendors can back up the data on their hardware or on customer-provided hardware; rotate tapes and manage customer's backup library; replicate data to alternate sites as needed; and handle other aspects of the process as needed. Whether you go the do-it-yourself route or work with a vendor, here are some options to consider:

Disks or tape. Tape backups are fairly inexpensive these days. However, they entail slower backup and recovery times, and require managing the physical tapes. Tapes also can degrade; when you go to restore your data, the tapes may not be readable. Hard disks offer a faster backup and recovery process than tape, and include additional benefits such as deduplication and data compression.

Hybrid. Data is backed up on a local device and in a secure offsite data center or the cloud for redundancy. You always have a secure local copy of your data as well as the offsite copy. Your systems are backed up to the local device first, so you don't have to worry about the replication to the data center or cloud affecting the performance of your systems or Internet connection.

Direct-to-cloud. With this option, you bypass the local device and send your data directly to the cloud. You're still backing up your data to a remote data center, but without the local copy on site. Depending on your Internet speeds and specs of your machines, these backups could take much longer.





Data Backup Optimization

Backup isn't a "take-it-as-it-is" kind of thing. There are various ways to manipulate data to optimize the process. Among the benefits of doing so are improved backup and recovery speeds, data security and reduced bandwidth requirements. Data manipulation tactics include:

Compression. Various schemes can be employed to shrink the source data so it uses less storage space. The decrease in data size also generates faster transfer speeds. Compression is frequently a built-in feature of tape drive hardware.

Deduplication. There is potential for redundancy when multiple similar systems are backed up to the same destination storage device. Deduplication eliminates multiple copies of the same thing. It can be applied at the file level, on raw blocks of data, on a server before any data moves to the backup media or at the target storage device.

Duplication. Backups can be duplicated to a second set of storage media. This allows for rearranging the backup images to speed up recovery, or to have a second copy at a different site or on another storage medium.

Encryption. Encryption uses algorithms to change the form of data to protect it during backup and recovery. Only those holding an encryption key can read the data. Data can be encrypted while at rest and/or in transit. Encryption is a CPU-intensive process, so it slows down backup.

Multiplexing. When there are more data systems to be backed up than there are destination storage devices, using a single device that allows for simultaneous backups can be more efficient.

Staging. Backups can be copied to a staging disk right before being copied to tape. This helps avoid problems in matching the speed of the final destination device with the source.

Data Recovery

Fact. Optimizing and backing up data does you little good if you can't recover it when you need it — and in the form you need it.

Each type of data has its own requirements for recovery point objective (RPO) and recovery time objective (RTO). Keep in mind that it takes time to copy data back to the production system. The techniques used to optimize backups, like deduplication and compression, can add to the recovery time needed.

That's why it is a good idea to employ a tiered storage scheme for your backups. This enables you to direct the various types of data to specific kinds of storage media, depending on how quickly you'll need to retrieve the information.

In most cases, all the data has to be recovered. There are data protection products that perform changed block recoveries. These products examine data that is already in place on production storage and recover only the data needed to set the clock back for an application. This assumes the primary storage system is up and running and that data corruption didn't cause the failure.

If the storage system fails or a changed block recovery is not available, in-place recovery may be an option. This allows backup data to be moved to a live state on the backup appliance, eliminating the need to move data across a network. It can also help protect against server or storage system failure.

In general, backup and recovery strategies are reserved for data that you can do without for 24 hours or more. If you need recover data much quicker, you may need to employ replication in addition to backup, or explore the use of a disaster recovery as a service (DRaaS) solution.

Data Replication

Data replication is similar to data backup in that it copies and moves data to another location. However, it's typically done in real time or near-real time rather than at periodic intervals. Replicating all of an organization's data in real time would be prohibitively expensive, so it's typically reserved for only the most essential data needed to keep a business up and running when a disaster or other disruptive events occurs.

As is the case with data backup, recovery time objectives (RTO) and recovery point objectives (RPO) are important. How long can you be without critical workloads and how much data loss can your business reasonably absorb?

Replication is usually performed outside your operating system, in the cloud or on virtual machines. Because a copy of all your mission-critical data is there, you can "failover" and migrate production seamlessly to cloud. There's no need for wait on IT for backup tapes to be pulled. RPO can be as quick as 10 seconds after production.

There are four common types of replication:

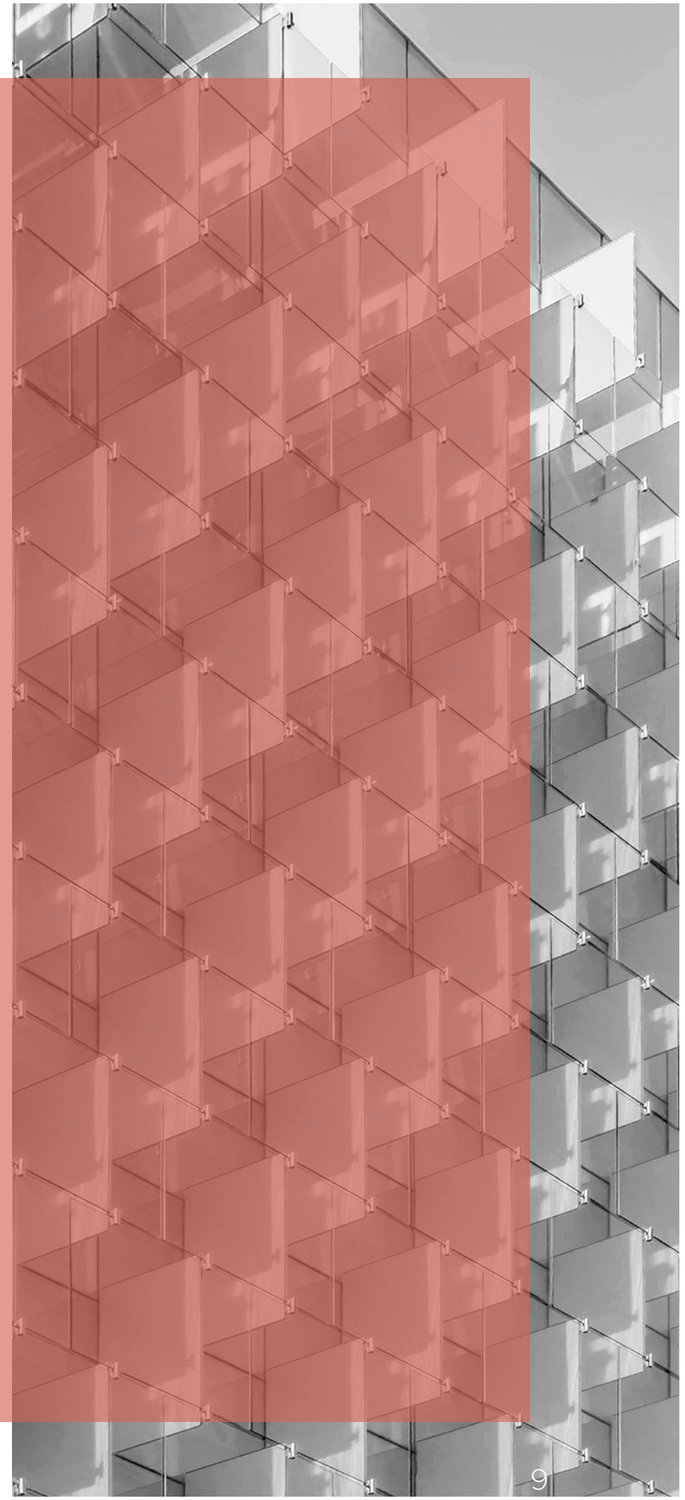
Asynchronous: Data is first written to primary storage and then copied to secondary storage. It can be done in near real-time or scheduled.

Synchronous: Data is written to secondary storage at the same time it's written to primary storage.

Real-Time: All new data and changes are captured as they occur, and transferred to the secondary device, either synchronously or asynchronously.

Point-in-Time: New and changed data is transferred to the secondary device on a periodic or scheduled basis, and therefore this type of replication can only occur asynchronously.

Continuous Data: Rather than scheduled replication that happens a few times a day, continuous data protection is just that: continuous. This results in even less potential data loss in the event of a failure.



The Case for Replication and Backup

Replication and backup work differently. Replication creates a copy of your data in real time or near real time, and and can better provide a more up-to-date copy of your data than a backup. It is for keeping mission-critical production operational so your business can continue running during a disaster or failure. It's important for BC. However, very few companies can afford to replicate all their data. It's also not fool-proof because it produces a duplicate. That means it copies every change, even if the change was a virus.

Backup creates a copy of your source data or changes to it, and provides you with access to versions that you can roll back to as necessary. Replication needs to be combined with another type of technology to create recovery points to roll back to if required. And since you are probably not replicating all of your data, you will eventually need backups to put everything back to where it was before the disruptive event. Plus, backup provides a compelling solution for creating, managing and accessing data on a day-to-day basis, as well as help for DR.

However, replication offers a significantly faster recovery time. It may take considerably longer to recover a whole site from backups compared with from a replication copy.

Bottom line: replication and backup both have their own advantages and disadvantages. As such, it makes sense to incorporate both into an overall data protection strategy.

The Backup and Recovery Plan

The following are some of the considerations to take into account when developing a backup and recovery strategy — and determining whether additional methods such as replication should be incorporated:

- How important is the data on your systems? Do you need to back it all up or is there some that won't matter if it's lost? If specific data must be backed up, when and for how long? What would happen if you couldn't access it right away?

- What type of information does the data contain? Who needs to use or access it? Is it subject to regulatory requirements of any kind? Does it need to be kept private? Does it need to be kept for a specific period of time? How often will you need to access it? The answers to the last question will help you determine if data archival may be required.

- How often does the data change? The frequency of change can affect your decision on how often the data should be backed up. For example, data that changes daily should be backed up daily.

- How quickly do you need to recover the data? Do you need it right now to keep business going? If that's the case, replication may need to be part of your strategy.

- Can you go without it for a day or maybe even a week? Is there some you need quicker than the rest? Recovery time is an important factor in creating a backup plan.

- Do you have the equipment to perform backups? If you're doing your own backup, you'll need hardware. If you don't already own it, you'll need to purchase it. For some companies, outsourcing may be more cost effective than incurring capital expenses.

- Who will be responsible for the backup and recovery plan? You'll need someone reliable to handle the tasks, as well as someone to back that person up. If you don't have the staff or resources in place, out-sourcing may be a good option.

- What is the best time to schedule backups? Scheduling backups during off-peak hours will speed the backup process but that's not always possible. Carefully plan when key system data can be backed up. This may also affect your use of automation for the backup process.

- Do you need to store backups off-site? In almost all cases, the answer is yes. The bigger question may will the off-site location be one you manage or belong to a vendor? Will it be a physical data center or the cloud?

Additional Considerations

There are several other issues that need to be taken into account as well, such as these:

Automation. Setting schedules of automatic backups and other recovery-related tasks will ensure you remove the risks that come with manual processes.

Authentication. Using an authentication mechanism is a good way to prevent the backup scheme from being used for unauthorized activity.

Backup Validation. Backup validation is used to test the validity of backups to ensure data can be recoverd.

Chain of Trust. Removable, physical storage media should only be handled by trusted individuals. Establishing a chain of trusted individuals and/or vendors is essential for ensuring data security and integrity.

Cloud Options. Back and recovery plans should utilize multiple methods of data storage for maximum flexibility and redundancy. There are several cloud-based options that are scalable, cost-effective and won't be impacted by any sort of on-premises disaster. The cloud can make recovery simple and convenient since you can access data through any Internet connection.

Costs. Everything has a price, whether it's do-it-yourself backups or outsourcing to managed services provider. In addition, all types of storage media have a finite capacity and an associated cost. Matching storage capacity with the backup needs is an important part of a backup and recovery plan.

Data Retention. Regulations and policies often dictate how long you need to keep backups. But you don't want to keep them longer than necessary to avoid unwanted liability and sub-optimal use of storage media.

Data Security. Backup and recovery isn't just about ensuring access. It's also about restricting unauthorized access and preventing data theft or corruption. What security processes and procedures will you incorporate?

Monitoring. For added security and peace of mind, you may opt to have a third-party monitor your backups. Some monitoring services include the collection of historical metadata that can be used for project data growth, locating redundant primary storage capacity and identifying reclaimable backup capacity.

Network Bandwidth. Distributed backup systems can be affected by limited network bandwidth.

3-2-1 Best Practice

An additional consideration to keep in mind as you develop a data protection strategy, whether you will implement it on your own or work with a third-party company, is the use of best practices. As is the case with technology in general, best practices for implementing and managing various technologies and processes change rapidly. One best practice that remains consistent in terms of backup and recovery solutions is the “3-2-1 rule”.

The 3-2-1 rule refers to:

- Keeping at least **three** copies of your data, including the original copy and at least two backups.

- Keeping your backed-up data on **two** different storage types.

- Keeping at least **one** copy of the data offsite.

It’s a simple practice, but one that helps ensure you’ll have a copy of your data no matter what happens. Keeping at least three copies of your data helps prevent the loss of your only copy of your data. Keeping those copies on different storage types makes it less likely than any kind of disaster or failure would destroy all data copies. Keeping at least one copy offsite, preferably in geographically diverse location, helps ensure there is no single point of failure and that your data is safe from disasters that are unlikely to affect two separate locations.

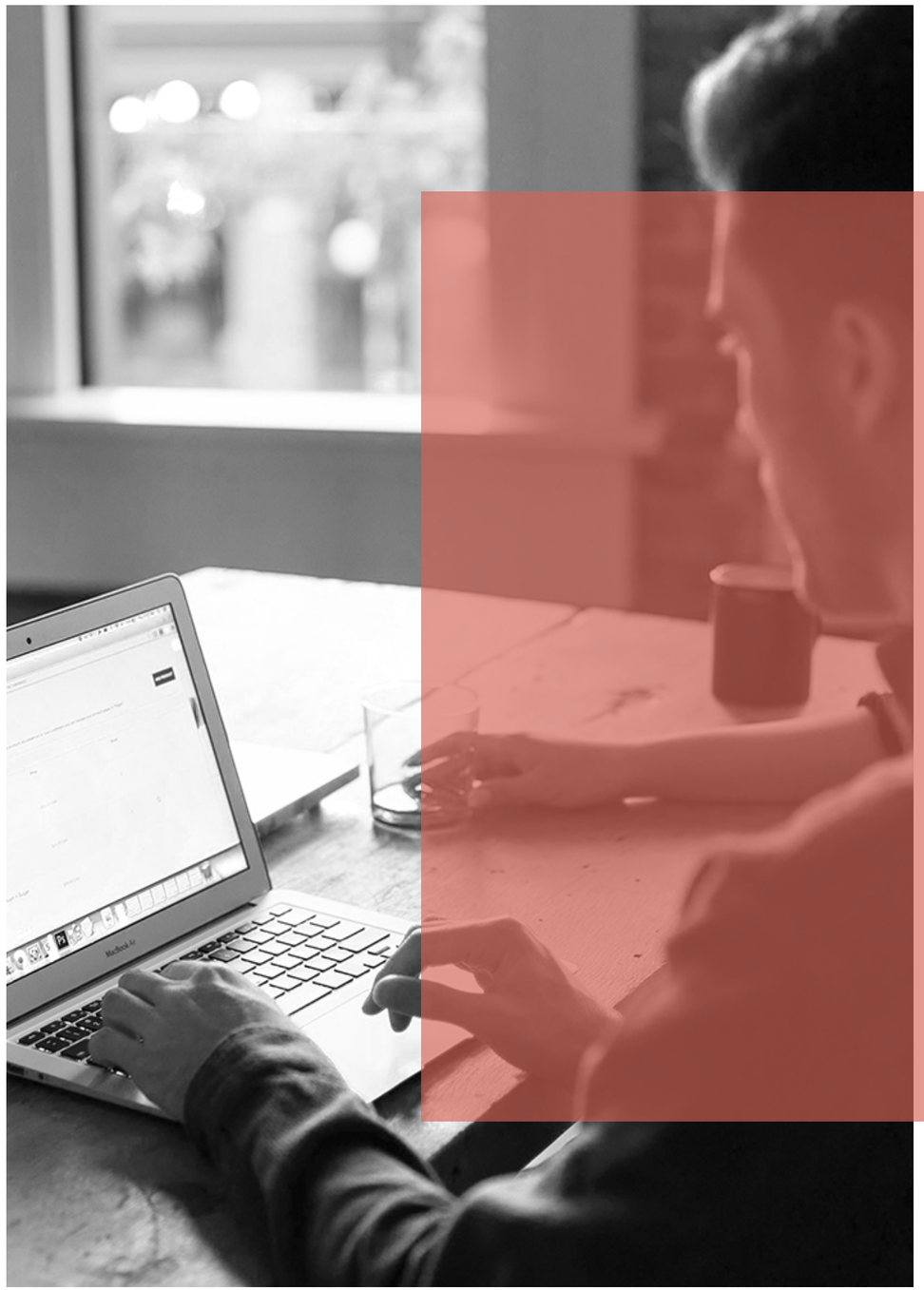
Pick a Data Protection Partner

If this all sounds like a lot to digest, it is — and there is much more to it that isn't covered in this eBook. Data protection is a multi-faceted, ever-changing business need, and it requires a commitment in terms of time, resources and continuous learning. Not surprisingly, many companies find outsourcing all or part of their data protection and/or disaster recovery needs makes good business sense. The key is to pick the right provider.

Choose one that is willing to learn about your company's unique needs and challenges rather than one just interested in selling an off-the-shelf, packaged data protection solution. There are lots of data protection products and services out there. One may even be just right for you. Or, a combination of data protection tactics and technologies may better meet your needs. Go with a vendor that makes meeting your specific needs a priority.

Ideally, you want to work with a vendor that has a history of backup and recovery success — and customers willing to vouch for the results the vendor delivered for them. The best third-party companies to handle data protection are also willing to back their solutions with service level agreements and around-the-clock technical support. They stay at the forefront of technological advances, and they understand that data protection isn't just a technology need. It's a business requirement.

Even from the technology perspective, data protection isn't just about one aspect of your IT operations. That's why if you can find a vendor or companies that have partnered to deliver integrated solutions, all the better. For example, if you're interested in moving mission-critical data to the cloud, look for a cloud services provider that also offers data protection. Even the most secure cloud environments can't protect your data when it's moving to and from the cloud.



Conclusion

Data is integral to business operations and success. The key to a solid data protection strategy is minimizing the risk to data by getting it off-site, offline and out-of-reach — but ensuring it can be reached when needed and in the form it's needed.

A well-designed plan also minimizes the risks to your business, reduces costs, increases compliance, and helps improve overall business service levels. You can go it alone, or you can engage a third-party company to help you develop and implement the solutions best suited to your data protection needs.

To learn more about data protection, including backup and recovery, visit www.USSIGNAL.com