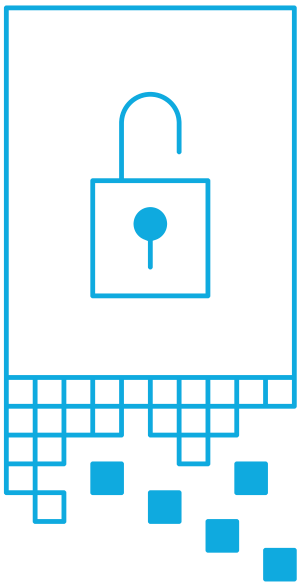
Data protection 101 whitepaper

# What you need to know about data security and integrity

Every day, an estimated 2.5 quintillion bytes of data are 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.

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.



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 are 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 difference between backup and replication? Where do things like encryption and deduplication fit in?

**Call:** 1-866-244-7474
**Visit:** business.shaw.ca/cloud

**Follow us:**

Shaw Data Centre & Cloud Solutions
Powered by Flexential

## The language of data protection

A discussion on data protection first requires a review of some of the key terminology. While this isn't a comprehensive list, it does provide some of the basic terms you should know 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** refers to 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 interdependencies and critical staff that is then analyzed to assess the potential impact of a disruptive event.

**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 and disaster recovery.

**Disaster recovery** is the process of replicating identified parts or all your 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. Mission-critical is used to describe data that is essential to the functioning of your business and its processes.

**Recovery point objective** is the point in time to which a firm must recover data as defined by the organization. The recovery point objective dictates which replication method is required.

**Recovery time objective** is the duration of time within which a business process must be restored after a disruption to avoid unacceptable losses. Recovery time objective begins when a disaster hits and does not end until all systems are operational.

**Replication** refers to copying data from one location to another either immediately or with a short time delay. It's like 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. In the case when data corruption is replicated to the secondary copy, this makes replication ineffective for backup.

## Business continuity and disaster recovery: The subtle differences

Much of the confusion over data protection starts with business continuity/disaster recovery — or rather the difference between the two concepts. Business continuity enables a business to continue operations while disaster recovery refers to recovering from a disaster. Both involve data and are about not letting an event disrupt business operations any longer than necessary.

Business continuity 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 always.

Since most businesses rely heavily on data for their day-to-day operations, a business continuity plan typically requires the accessibility of data 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.

Business continuity 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. In most cases, IT leadership works with identified business leaders to identify, classify and prioritize which business services and applications are critical.

Disaster recovery is a subset of business continuity 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 save your data somewhere so that you can recover it after the event. Disaster recovery can be as simple as backing your data up remotely even if you can't access it right away.

As with business continuity, some data is more important than the rest for disaster recovery, so it gets recovered first. The speed of recovery is also important. Some data may need recovered immediately after a disruptive event; other data can wait days or even weeks. How quickly data is needed will influence the choice of backup and recovery tactics.

## Data backup

A backup is a copy of data and is used as a safeguard against unexpected data loss or 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 is retained for a long period — often for compliance purposes. Backups are scheduled and entail making copies of data files at intervals, such as hours or days, depending on your business requirements. The data copies are saved to a physical hard drive, tape, disk or to a virtual tape library, and kept offsite. The appropriate techniques depend on the type of data you back up and the desired convenience level of the recovery process.

Backups are performed at periodic intervals, while replication is done in real-time or near real-time. The basic types of backups include:

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

While you can do your own backups, many organizations hand off the responsibility to third-party companies. The 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 inexpensive. However, they entail slower backup and recovery times, and require managing the physical tapes. 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 a copy is moved to 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.

## Data backup optimization

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

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. Consider the following manipulation tactics:

**Compression** – Various schemes can shrink the source data so it uses less storage space. Compression is frequently a built-in feature of tape drive hardware.

**Deduplication** – There is potential for redundancy when you back up multiple, similar systems to the same destination storage device. Deduplication eliminates multiple copies of the same thing. It is 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 are 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. Encryption is a CPU-intensive process, so it slows down backup.

**Multiplexing** – When there are more data systems to back up than there are destination storage devices, use a single device that allows for simultaneous backups for efficiency. Staging – Copy backups to a staging disk before copying to tape. This helps avoid problems in matching the speed of the destination device with the source.

## Data recovery

Each type of data has its own requirements for recovery point objective and recovery time objective. 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 general, reserve backup and recovery strategies for data that you can do without for 24 hours or more. If you need to 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 solution.

## Data replication

Data replication is like data backup in that it copies and moves data to another location or a parallel storage attached network in the same location. However, it's typically done in real time or near-real time rather than at periodic intervals. Replicating all an organization's data in real time is 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 and recovery point objectives are important. How long can you operate 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 to wait on IT to pull backup tapes.

There are four common types of replication:

- Asynchronous: Data written or changed on the primary storage device is sent to the secondary storage device, but the timeliness and/or successful completion does not impact the primary storage device or the writing applications.
- Synchronous: Data written or changed on a primary storage device is copied to a secondary device at the same time, block by block. If the data isn't committed to the secondary storage device, it doesn't get committed to the primary device, causing an error in the writing application.
- 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.

An alternative to replication is the use of continuous data protection. Rather than scheduled replication happening 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. When considering continuous data protection replication methods, take wide area network speed into consideration to ensure there is enough bandwidth.

## 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 is more likely to provide a more up-to-date copy of your data than a backup. It is reserved for keeping mission-critical production operational so your business can continue running during a disaster or failure. It's important for business continuity. 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. Combine replication with continuous data protection or another type of technology to create recovery points to roll back to if required. And since you probably don't replicate all of your data, you will eventually need backups to put everything back to where it was before the disruptive event. However, replication offers a much faster recovery time. It may take considerably longer to recover a whole site from backups compared with from a replication copy.

## The backup and recovery plan

Take the following considerations into account when developing a backup and recovery strategy — and determine whether you should incorporate additional methods such as replication:

- 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 needs 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? Is it private? Should you store it 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 is required.
- How often does the data change? The frequency of change can affect your decision on how often to back up the data. 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, make replication 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 is more cost effective than incurring capital expenses.
- Who is 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, outsourcing is 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 to back up key system data. 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 is will you or a vendor manage the off-site location? Will it reside in a physical data center or the cloud?

## Create a solid data protection strategy

The bottom line is that every organization will have different requirements for a data protection plan to meet its specific needs and preferences. As such, it makes sense to evaluate technologies. Incorporate the right mix to design a solid data protection strategy as the need for data protection continues to intensify.