Data Backup and Recovery

Contents

Data Backup and Recovery

Executive Summary

“Data is a precious thing and will last longer than the systems themselves.” Tim Berners-Lee, inventor of the World Wide Web. A Framework for Web Science (Foundations and Trends)

The word "precious" implies both value and fragility. For any organisation, the data it holds is valuable and critical to its operation. But it is also fragile: a simple mistake or software bug can wipe it all out with disastrous consequences.

Backing up data can protect it. If one copy is deleted or corrupted, it can be replaced with a backup. However, having a second copy will not get the system running if there is a complete hardware failure. That is the difference between **Backup** and **Disaster Recovery (DR)**. Backups are purely about data; DR is concerned with the process of restoring a complete system after an outage.

When considering a supplier for a backup/DR solution contact your Local Authority and/or the appropriate NEN’s member who are both likely to have trusted suppliers whose quality of service has been validated.

The effect of a major loss of data can be catastrophic for an organisation. A study by the British Chambers of Commerce found that 93% of businesses that suffer a data loss of more than ten days filed for bankruptcy within a year. While a school cannot go out of business the costs of a data loss can be extremely onerous.

Data loss can result from hardware failure, software corruption, human error or malicious intent in the form of viruses and malware. Ransomware is one particular form of malware and can affect all organisations, including schools.

The design of an appropriate backup regime requires knowledge of:

1. what can cause a system problem,
2. what data needs to be backed up,
3. where that data is,
4. each element's relative importance, and
5. where the backups should be stored for easy restoration if or when required?

What can cause a problem?

The first point of potential failure to consider is a hardware fault of some kind. Recovering from a failure of this type should just be a matter of replacing the equipment and restoring any lost data from backup.

Server and client computers are susceptible to software bugs in applications and operating systems which could corrupt data. While data stored in the Cloud is less likely to be subject to application or operating system bugs systemic failures can still occur. In schools many pupils leave at the end of the academic year and their data is deleted - a simple mistake could cause important data to be erased.

Cybersecurity incidents can cause extensive data loss and disruption. After an incident, local, always-connected backups may also have been corrupted, so it is essential to have remote, offline backups, with the ability to roll back in time to safe copies of the corrupted files.

Data Audit

A data audit covers what data must be backed up, where it is, how often it needs to be backed up, and where to store the backups. The knowledge gained through this process enables an appropriate backup strategy to be designed.

Schools' data can be broken down into three areas: business data (management, staff, and pupil files), applications and licences, and configurations. Further, we can divide data into Core (essential/critical) and non-Core. When carrying out the data audit, this division between Core and non-Core can help decide what needs to be backed up and how frequently.

The audit should also identify where data is stored. File server(s), MIS server, e-mail server, and web server(s) will cover most of the locally held data but other locations may need to be considered, including online storage (GoogleDrive, Dropbox, etc.).

If Virtual Machines are used instead of "real" servers or whole server images are saved then all the required configurations will be in the backup. If neither is done, restoration becomes much more complex as operating system, applications (with all required configurations), and data will all need restoring.

Backup Strategy

The data audit will have identified what data needs to be backed up, where it is located, and each item's relative importance allowing a suitable strategy to be designed. Of particular importance for many schools is making sure that any third party supplier supports backing up the various Google offerings and/or Microsoft's Office365 in addition to other cloud services.

A typical backup regime is predicated on a clearly time-defined incident: files created before a specific, recent time are clean and can safely be restored. But not all incidents are so well defined (a ransomware attack, for example). For these incidents it should be possible to restore older file versions. This will involve a greater loss of data but is preferable to its total loss.

Finally, once the backup regime is in place, run tests regularly with all staff who may be required to restore files from backup so that they become familiar with the process.

Data Backup and Recovery

Introduction

“Data is a precious thing and will last longer than the systems themselves.” Tim Berners-Lee, inventor of the World Wide Web. A Framework for Web Science (Foundations and Trends)

In the Tim Berners-Lee quote above, the word "precious" implies both value and fragility. For any organisation, the data it holds is critical to its operation. It may be a vast database of user profiles for targeting adverts or a simple e-mail list for a local charity. However, in both cases, that data is the organisation's lynchpin.

But it is also fragile: a simple mistake or software bug can wipe it all out in a split second with disastrous consequences.

Making backups of the data can protect it. If one copy is deleted or corrupted, it can be replaced with one of the backups. However, what if the data is on a server that has succumbed to a complete failure: it cannot be rebooted; nothing can save it? Then what? Just having a second copy of the data will not get the system up and running again.

That is the essential difference between simple **Backups** and **Disaster Recovery (DR)**. Backups are purely about the data; DR is concerned with the process of restoring a complete system (hardware, applications, and data) after an outage.

The rest of this guidance note is mainly concerned with Backups, but some elements of DR are also included. It does consider, for example, how whole servers (or virtual machines) can be backed up and how applications and configurations should be handled. Both fall more naturally into the DR area as these elements are not usually considered “data” but are essential for recovery after a major outage.

When considering a supplier for your chosen backup/DR solution, contact your Local Authority and/or the appropriate NEN member who are both likely to have trusted suppliers they have worked with and validated the quality of their services.

Overview

Data is critical to any organisation's smooth running, be it a business, charity, or school.

The effect of a major loss of data can be catastrophic for an organisation. A study by the British Chambers of Commerce found that 93% of businesses that suffer a data loss of more than ten days filed for bankruptcy within a year. A DTI/Price Waterhouse Coopers report found that 70% of small firms went out of business within a year after a major data loss. While a school cannot go out of business - their pupils still need to be taught - the costs of a data loss can be extremely onerous. There are the financial costs of restoring the data, time lost, the general disruption to pupils' education, and reputational damage.

Data loss can result from the usual vagaries of all IT systems (hardware failure, software corruption, human error) to malicious intent in the form of viruses and malware. Ransomware is one particular form of malware that seems to be on the rise and affects all organisations, including schools.

As with most things in life, prevention is better than cure. Having a robust and comprehensive backup regime in place is the best protection against data loss and the subsequent high restoration costs after a major failure or system compromise.

The design of an appropriate backup regime will require knowledge of

1. what can cause a system problem,
2. what data needs to be backed up,
3. where that data is,
4. each element's relative importance, and
5. where the backups should be stored for easy restoration if or when required.

Why Backup?

Hardware Failure

The first point of potential failure to consider is a hardware fault of some kind. It is almost inevitable that, at some point, a critical piece of hardware will fail through general wear and tear before being replaced as part of general network maintenance. Recovering from a failure of this type should be just a matter of replacing the equipment and restoring any lost data from backup.

Precisely which data needs to be restored to complete a full recovery needs careful consideration when planning what to backup. For example, if a firewall fails, it will be necessary to reinstate all the pre-existing rules on the replacement. Similarly, if a web server fails, it is not enough to only restore the website's content: the webserver software's configuration, and possibly an associated database, will also need to be restored from backup.

A common point of failure in all computers is the hard disks. Being mechanical, with many moving parts working to very fine tolerances and often running 24/7 in servers, they wear out and eventually fail. Server disks are usually monitored, and steps can be taken to replace them before they fail: but even then, sudden failures, for example, a power outage, can damage the disc.

SSD drives are less likely to fail suddenly as they have no moving parts but, being a new technology, statistics comparing SSD and HDD failure rates are somewhat limited (but see *Backblaze Drive Stats for Q1 2021*).

Software updates or bugs can affect both HDDs and SSDs by, for example, overwriting exiting data or corrupting the partition table.

And then, there is the human element: deleting data by mistake or reformatting the wrong disk on a server is always a possibility. A good backup can turn a disaster into a minor inconvenience.

Software Issues

Data loss can also occur due to software problems. Minor bugs may, for example, cause a database to be corrupted. Servers and client computers are both susceptible to this type of problem, as are applications and operating systems. While bugs can surface at any time, they are most likely to occur when upgrading from one version to another. Critical data should always be backed up before performing a software upgrade so that it can be easily reverted to a previous, stable version.

Human error can also play a part. In schools, for example, at the end of the academic year, many pupils leave, and their data is deleted: home directories wiped, e-mail accounts removed. A simple error here could delete important examination material.

While data stored in the Cloud is less likely to be subject to application or operating system bugs, systemic failures can still occur, and data may still become corrupted. And being in the Cloud is no protection against human error.

Cybersecurity problems

The third category of event that may lead to data loss is cybersecurity incidents. These can be external threats, like viruses and other malware, to ransomware attacks, which are becoming more frequent, as highlighted in the TES (Sept 2021) and the NCSC. When a school is subject to a cybersecurity incident that corrupts or encrypts essential data, backups are the only route to restoration.

Everything possible should be done to prevent a successful attack, but you only need one successful intrusion to cause havoc.

After a cyber security incident, local backups are also likely to have been corrupted, so it is essential to have remote, offline backups, with the ability to roll back in time to safe copies of the corrupt files to minimise any data loss.

A good place to start when considering what processes need to be in place to secure your network is the NEN's three Cyber Security guidance documents:

- Cybersecurity Guidance for schools
- Cybersecurity Checklist
- Cybersecurity What if?

These include links to many other resources, particularly to the NCSC that was involved in their creation.

Data Audit

Even after taking all the precautions you can to protect yourself against data loss (by, for example, implementing the NCSC’s 10-Steps), the worst may still happen.

In order to design an appropriate backup strategy, you will need to carry out a data audit covering what data must be backed up, where it is, how often it needs to be backed up, and where to store the backups themselves.

What?

Schools’ data can be broken down into three areas: what can loosely be described as business data (see below), applications and licences, and configurations. Further, we can divide data into Core (essential/critical) and non-Core. Non-Core data can either easily be restored without a backup or is “low-value” - i.e. does not impact the school’s running or its core educational objectives. For example, pupils’ work required for examination purposes would be a Core item, whereas KS1-3 student files can be lost with little impact on their education and so would be considered non-Core.

When carrying out the data audit, this division between Core and non-Core can help decide what needs to be backed up and how frequently. Core data will need to be backup up as frequently as practically possible, whereas non-Core data can have a longer backup interval accepting that some may be lost if a restore from backup is ever required.

Business Data:

To a large extent, the other categories (Applications and configurations) can be restored "by-hand" if necessary. It will take time and be very disruptive, but it can be done. Business data (using the term very broadly) is valuable and will often be impossible to restore without a backup. The restoration of parental or financial data, for example, will be fraught with difficulties and reputational damage.

What is "Business data" when considering schools? Again there are three basic categories: management, staff, and pupil files.

Management data will include, for example, staff personnel records, finance, suppliers details, development plans, meeting minutes, medical records, pupil records (including medical data and other private information). These are all required so that the school can continue its primary purpose of educating its students.

Staff data includes files created by the staff for teaching purposes: teaching plans, syllabi, homework sheets, work submitted by students for marking, for example. The majority of these files will probably be saved in the users' home directories on central file servers, but other, possibly shared, locations will also have to be considered.

Pupil data will mainly be in their home directory, but other locations (e.g. a class directory or “cloud” directory) may also exist.

Applications:

Schools use a wide range of applications both for administration and education. Some of this software will be online and accessed via a web browser. If local systems are corrupted, then these resources may become unavailable if, for example, passwords are deleted and cannot be restored. Local software applications may need to be reinstalled (on client machines or servers): not only must the program code be available in some form, but licences may need to be produced before it can be reinstalled.

Most apps are downloaded as software installers (rather than being supplied on CD) and will need to be restored from backup or re-downloaded before the application can be reinstalled.

In some schools, standard client images are used to rebuild a computer after a failure: these will need to be backup up.

Configurations:

All computers (desktops, laptops, tablets, servers, routers, etc.) have to be configured: from a unique IP address to complex firewall rules, from a local application's settings to an apache webserver's multi-site configuration. These will have to be restored in the event of a cyber attack and need to be available in a backup.

For servers, the simplest method is to back up an image of the server or use virtual machines which can be backup up in their entirety. Whatever process is used, it is essential to audit these configurations and plan for their backup and restoration.

For client computers (including laptops, tablets, etc.), rather than backing up individual machines, it is generally easier to maintain standard builds which can be used to re-image a damaged or infected computer. If policies are in place that prevent data from being stored locally or that automatically synchronise data with central servers, restoring from an image is straightforward and avoids any data loss.

Where?

Having identified what data needs to be backed up, the other part of the audit is to list where the data is stored.

Business Data:

The most obvious locations are all the local servers: file server(s), MIS server, e-mail server, web server(s), amongst others. While these servers will cover most of the locally held business data (as defined above), other locations may need to be considered. Can students save files to desktops or laptops? If so, is that the only copy or are they synchronised to some other location - local file-server, GoogleDrive, Dropbox, etc. Whether individual client devices need to be backup up separately will depend on the answer to that question.

Ideally, no data should be stored on these devices so that if there are problems, they can be rebuilt with a standard, clean image without loss of data.

Hinted at in the previous paragraph is consideration of any cloud services used. These may range from a small number of simple Dropbox accounts to every student and staff member having space on GoogleDrive or a similar online service. Many schools' have replaced local MIS servers with an online service where data is not held locally but “in the cloud”. Similarly, there may be online educational platforms where pupil progress-tracking data is stored online. With all remote services, it is critical to understand (1) what backup guarantees the supplier provides and (2), can the data be exported or backed up using a third party product so that a backup can be held separate from the supplier's system?

Applications:

In the main, client applications themselves do not need backing up in the traditional sense: most can either be re-downloadable or come on CD. If licences have been retained (either in paper form or backed up electronically), reinstallation should not be a problem.

The same will apply to applications running on servers (Databases, web servers, etc.) but with the added complication of the required configuration files (see below).

If virtual-machine servers are employed, backing up becomes easier as VM systems provide a backup command which creates a compressed image of the VM. This image can then be saved to a separate device: a NAS server, for example, or to some remote backup location.

Configurations:

The issue of configuration files goes away if either (1) VMs are used instead of “real” servers or (2) a whole server image is created and saved so that a corrupted server can be wiped and re-imaged complete with all its configuration files. If neither is done, restoration becomes more complex as operating system, applications (with all required configurations), and data will need restoring separately: a potentially long and complicated process.

Backup Strategy

The data audit will have identified what data needs to be backed up, where it is located, and each item’s relative importance. A strategy can now be developed to define the frequency of backups and where they will be stored.

When considering backup frequency, there may be different strategies for different data: static or low-value data (i.e. non-Core) may be backed up each week, whereas dynamic or valuable data (i.e. Core) must be backed up more frequently. For example, a finance database may be replicated, providing instantaneous failover and "dumped" for off-site storage every few hours. This part of the backup strategy is site-specific and can only be decided at the local level. When determining how often to back up data, always consider its value and how often it changes. Valuable data only requires very infrequent backup if it rarely changes; less valuable data may still need a nightly backup if it changes very rapidly.

The "where" part of the strategy is more consistent over all sites: the tried and trusted 3-2-1 strategy is universally accepted as the gold standard, although how it should be implemented has altered as the available technologies have changed. There are copious references to the 3-2-1 strategy - google "3-2-1 backup strategy" for a plethora of references - but the classic definition defines the key idea:

(a) you keep 3 copies of your data: one live data and two backups;
(b) on 2 different media (disk and tape);
(c) with 1 copy off-site.

The term 3-2-1 was originally coined by US photographer Peter Krogh while writing a book about digital asset management (see here for more background). With the advent of Cloud technologies and ubiquity of disk-based storage, the 2-media rule looks outdated and "off-site" is not as simple now as taking tapes home!

So what is the modern version of 3-2-1? Firstly, (a) is as before: 3 copies of the data - one live and two backups.

Rule (b) is probably not (or, at least, less) relevant now. Tape still has a place but mainly for archival purposes rather than true backup. Similarly, CD/DVD can be used for archiving but writing and reading times are too slow for backup/restore for all but the smallest backup tasks.

It is (c) where the main issues arise as, although a Cloud data store is essentially “off-site”, in many cases, it is not genuinely disconnected from the live data. In the original definition, the primary purpose of “off-site” was to create an “air-gap” between the live and backup data.

An added complication is that, in current usage, data stored in the Cloud is often the live data. So how should that be backed up? In the cases of data stored in the Cloud, “off-site” can be taken to mean “with another provider”. One should not consider data stored in the Cloud as being backed up by the provider. They may indeed have backups, but (a) they are not likely to be available on a file-by-file basis to the customer, and (b) will be using the same networking infrastructure and so may be compromised in the same way as the live data.

For data in a Cloud service, it is best to use a third-party company (RedStore, Barracuda, for example) that can directly back up the live data to their own cloud network. This avoids the necessity of downloading the Cloud data onto a local data store for backup and at the same time provides a software-only solution, i.e. there are no local servers to manage. Of particular importance for many schools in this regard is making sure that any third party supplier supports backing up the various Google offering and/or Microsoft's Office365. (See the Links section below).

The final element in the backup strategy - having identified the data, its value, its location, and where its backups should go - is the question "how often?".

In any discussion of business continuity and disaster recovery (BCDR), two acronyms are sure to be discussed. RPO (Recovery Point Objective) and RTO (Recovery Time Objective). These are both important and related concepts that help design an effective backup strategy: one that minimises the loss of data to an acceptable level. These are usually defined in business terms but apply just as well to any organisation with substantial amounts of data to manage.

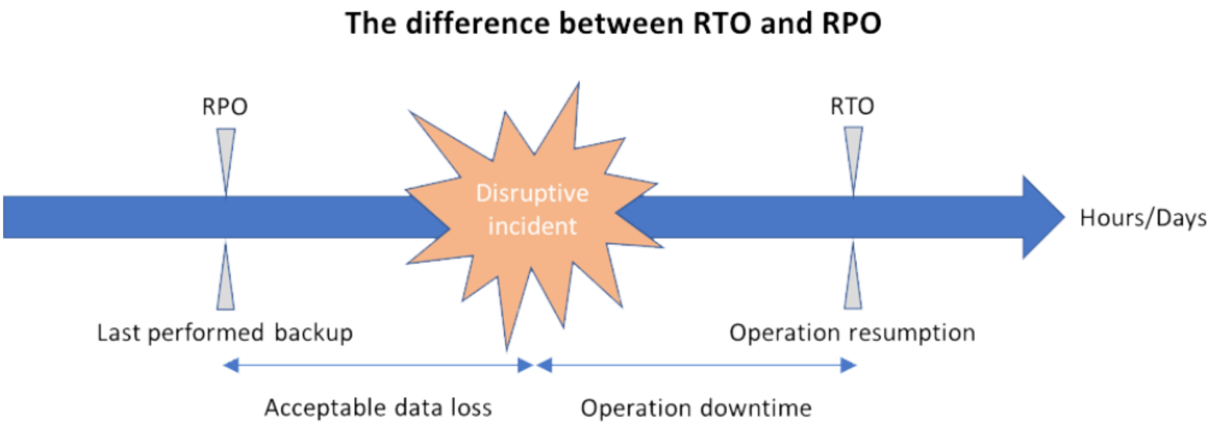
The ***Recovery Time Objective*** *is the targeted duration of time and a service level within which a business process must be restored after a disaster (or disruption) in order to avoid unacceptable consequences associated with a break in business continuity.*

The ***Recovery Point Objective*** *(RPO) is the maximum targeted period in which data might be lost due to a major incident.*

Put more simply, RTO is a target time for a service to be restored. For example, if your school web server fails: how long can this be unavailable without causing unacceptable consequences? 1 hour, 2, 24? What about the school's MIS system? This is a critical service, so maybe the RTO for this should be 0-1 hours. The RTO is not a system-wide number but one that will depend on the specific data under consideration. Once the RTO has been defined, it will indicate the backup schedule required to meet this target.

Similarly, the RPO is simply a target for the age of data we are prepared to lose. For example, setting an RPO of eight hours on student medical records would mean that you are prepared to lose (or re-enter) any changes made less than eight hours before the incident.

In other words, the RPO is looking backwards from the incident at what data we are prepared to lose, while the RTO is looking forward from the incident to when the data will be fully restored.



Source: https://advisera.com/27001academy/knowledgebase/what-is-the-difference-between-recovery-time-objective-rto-and-recovery-point-objective-rpo/

By considering the relative importance and changeability of the various classes of data identified in the data audit, appropriate RTO/RPO values can be assigned, and a backup strategy designed that meets them.

In general, an RTO/RPO design is predicated on a clearly time-defined incident: files created before a specific (and recent) time are clean and can safely be restored. But not all incidents are so well defined. A ransomware attack, for example, may have been corrupting files days or weeks before finally triggering the encryption and demand for payment. Or a virus may have been circulating for some time before being noticed. In these cases, simple RTO/RPO backups will consist of already corrupted files.

In these severe incidents, a successful restore requires that files older than a typical maximum RPO of 24 hours can be recovered. This is usually achieved in one of two ways.

(1) By creating snapshots of the backed-up data at specific points in time (daily, weekly) and retaining them for longer periods (8 days and 10 weeks, for example).

(2) By recording the differences between backups for a specified period (3 months, for example), any changes can be undone prior to being restored.

In both cases restoring a file to its position further back in time will mean a more significant loss of data with all that that implies, but it will be better than losing all your data in the case of a ransomware incident.

Testing

Once the data audit has been completed and the backup regime designed and implemented, everything should be tested. Have the backups been created? Are they stored in the correct place? Are the timestamps correct? Are they about the right size? These questions are straightforward to answer and provide good evidence that everything is being backed up correctly.

The more complex task is to test, and become familiar with, the restoration process. Clearly, it is not a good idea to test the backup on live data - if the restore does not work for any reason, then you may corrupt the original! Set up a server specifically to practice restoring individual files and directories.

The aim of testing is to become familiar with the process and check the validity of the backup by comparing the restored files with the originals.

Restoring whole virtual machines and hardware servers can be checked in the same way using test servers. One crucial point to consider when restoring either a VM or a real server is that it will have the same IP address as the original: this will either need to be changed before it is booted or disconnected from the network to avoid errors.

Testing should be carried out regularly with all technical staff who may be required to restore files from backup.

Summary

Data is critical to the smooth running of any organisation, be it a commercial business, a charity, or a school.

The importance of backing up data cannot be overstated: a severe incident resulting in substantial data loss would cripple even the largest company. For schools, it could mean students losing examination projects, parental financial details being lost or stolen, etc.

Loss of data can occur in many ways: simple human error, bugs in software, hardware failure, cybersecurity failures (virus, ransomware, etc.). In each case. the route back to a fully functioning system with little or no data loss is a robust and efficient backup process.

The design of an appropriate backup regime requires that all the data held is audited so that these critical questions can be answered.

- What data do we hold?
- Where is it?
- How often does it change?
- How important is it?
- How long can we afford to be without it?

Remember that data stored in the Cloud (e.g. Office365) is not backed up. Special arrangements need to be set up to back up cloud data: usually, this will be via a third-party provider who can directly backup cloud data to their system without running through the local network (cloud-to-cloud backup).

Once this information has been gathered, the required RPO and RTO for each class of data can be set. Some consideration should also be given to the possibility of restoring files that are older than the agreed RPO.

While restoring older versions of files is far from ideal and will result in more data loss, it may be necessary where a cyber security incident has gone undetected for some time, and the backup files have also become corrupted or infected.

With all the information in place, an efficient and effective backup strategy can be designed, implemented, and regularly tested.

Useful Links

NEN:

NEN members. https://nen.gov.uk/about-us/providers/. [Accessed 6 January 2022].

NEN Cybersecurity guidance:

NEN Cyber Security Guidance for Schools. https://nen.gov.uk/advice/nen-cyber-security-guidance-for-schools/. [Accessed 6 January 2022].

Cybersecurity Checklist. https://nen.gov.uk/advice/cybersecurity-checklist/. [Accessed 6 January 2022].

Cybersecurity – What If?. https://nen.gov.uk/advice/cybersecurity-what-if/. [Accessed 6 January 2022].

Ransomware and Schools:

Alert: Further ransomware attacks on the UK education sector by cybercriminals. https://www.ncsc.gov.uk/news/alert-targeted-ransomware-attacks-on-uk-education-sector. [Accessed 6 January 2022].

Cyberattacks on schools: the facts. https://www.tes.com/magazine/article/cyberattacks-schools-facts. [Accessed 6 January 2022].

Backup and DR:

The 3-2-1 backup rule: Has Cloud made it obsolete?. https://www.computerweekly.com/feature/The-3-2-1-backup-rule-Has-cloud-made-it-obsolete. [Accessed 6 January 2022].

For secure data backup, here's how to do the 3-2-1 rule right. https://www.networkworld.com/article/3527303/for-secure-data-backup-here-s-how-to-do-the-3-2-1-rule-right.html. [Accessed 6 January 2022].

What is the difference between Recovery Time Objective (RTO) and Recovery Point Objective (RPO)?. https://advisera.com/27001academy/knowledgebase/what-is-the-difference-between-recovery-time-objective-rto-and-recovery-point-objective-rpo/. [Accessed 6 January 2022].

HDD v SDD statistics:

Backblaze Drive Stats for Q1 2021. https://www.backblaze.com/blog/backblaze-hard-drive-stats-q1-2021/. [Accessed 6 January 2022].

Office365:

Do you really need to back up Microsoft 365?. https://www.druva.com/blog/do-you-really-need-to-backup-office-365/. [Accessed 6 January 2022].

Best MS Office 365 Backup Solutions Comparison. https://afi.ai/blog/ms-office-365-backup-solutions. [Accessed 6 January 2022].

Microsoft 365 Backup. https://www.redstor.com/solutions/microsoft-365-backup/. [Accessed 6 January 2022].

Google:

How to Backup Google Drive: A Step by Step Guide. https://spinbackup.com/blog/how-to-backup-google-drive-step-by-step-guide/. [Accessed 6 January 2022].

1304 - How to manage your Google Classroom backups. https://support.redstor.com/hc/en-gb/articles/360013002777-1304-How-to-manage-your-Google-Classroom-backups. [Accessed 6 January 2022].

Restoring with Google Vault: A Step-by-Step Guide. https://www.backupify.com/blog/restoring-with-google-vault-a-step-by-step-guide. [Accessed 6 January 2022].