DATA BACKUP AND DATA STORAGES

Iva Kavánková

Tomas Bata University in Zlin, Faculty of Applied Informatics

Nad Stranemi 4511, 760 05 Zlin

kavankova@utb.cz

Abstract

This paper focuses on data backup and archiving in the real environment of a concrete IT company engaged in software development. The introduction of the paper deals with the theoretical knowledge concerning the area of data backup and data storage. It also describes the current situation of data backup and problems with the current solution in the company. The article aims to provide suggestions for improving the current situation, including economic evaluation, to achieve efficient and, most importantly, secure data backup.

Keywords

Backup, RAID, archiving, data restore, data storage, NAS

1. Introduction

This paper is based on diploma thesis of the author [1].

Creating data backup is one of the essential activities of every company because losing data might cause catastrophic financial losses. People can backup data to prevent losing their data caused by a technical problem, robbery, fire, or natural disaster. Due to data backup, everyone can restore data and continue in work.

Data backup is still not very frequent in small companies. Usually, there are using USB flash discs, extern discs, or there is no data backup. Nowadays, servers or cloud storages represent the most popular solution. In more prominent companies, the situation is better because of financial security.

Thanks to modern technologies and the dynamic growth of backup technologies, faster solutions, automatization of processes, and many innovations could be expected. Also, the price for storing the same amount of data is decreasing. The capacity of backup medias will be dropping as well.

This paper contains an analysis of the current situation in the unnamed concrete company. The business model of the company is based on outsourced software development for other companies. The IT structure stands on a central server located in Prague, which is not ideal because HW damage can lead to the loss of all data. The main areas of general security are protection against theft or fire. Also, there are no options for employees to store their work in progress. The shared discs are the current solution for the company; however, this way is not providing sufficient capacity, and they are not used correctly for this purpose.

2. Data backup and data storages

People have been improving the process of data backup for many years. One of the first data storages was punch card. It was made of solid paper, whereas the information itself was represented by holes in concrete position. The main disadvantage is the inability to rewrite stored data. Other problems are small capacity, low baud rate, and reliability in reusing. The following storage medium was punch tape, it works on the same principle, but it was much more efficient because it was not necessary to change tapes so often. [2]

Another option is magnetic tape, which are data recorded in binary code, and different signal strength represents 0 and 1. It allows a bigger capacity than predecessors, and it is possible to rewrite the data. The problem is with sequential access to data because it might hold several hours to get the opposite side of the tape. Today it is used for archiving because the waiting did not mind too much. [2]

Floppy discs represent another generation that was created in 1967 by IBM. The capacity was 80 kB and diameter 8 inches. The most popular floppy discs were 3,5 inches big with 1,44 MB capacity. [2]

Development continued with optical discs, where the information is stored with a laser beam. Three types are now used: [3]

- **CD –** the first generation of optic discs, typically 12 cm diameter with 700 MB capacity, for reading is using a laser beam with 785 nm wavelength,
- **DVD –** the second generation of optic discs, it is possible to store data on both sides in two layers, and maximum capacity is 17,1 GB, reading is done by a laser beam with 660 nm wavelength,
- **Blu-ray Disk –** the third generation of optic discs, the technology was invented by Sony in cooperation with Philips with the more significant density of stored data and maximum is 200 GB, a laser beam with 405 nm (blue color) wavelength is used for reading.

The following table clearly describes the technical details of mentioned optic discs:

Optic disc	Wavelength for reading	Pitch
CD	780 nm	1,6 µm
DVD	650 nm	0,74 µm
Blu-ray Disc	405 nm	0,35 µm

Table 1: Comparison of optic discs, [3]

Another storage medium is USB (Universal Serial Bus) Flash disk, a small device (size is in centimeters) mainly used for data transfer between computers. Data are stored in flash memory, which allows data storage without a power supply. The newest one has a 2 TB capacity and a 5 Gbit/s baud rate. [4]

Another data storage medium is a memory stick, which also uses flash memory. Exists more types of them, but nowadays most popular are SD (Secure Digital) cards, which serve as a memory for cameras or mobile phones (capacity is up to 1 TB). [5]

Another option is HDD (Hard Disk Drive), an electromechanical device for storing and reading data with huge capacity. HW consists of the plates and reading heads; the speed of plates indicates the speed of the disc (usually 7200 rotates per minute) with a capacity is up to tens of TB. The major disorders of HDD are damaged sectors (age or extreme temperature conditions) and problems with rotations. [6]

Technology SSD (Solid State Drive) does not have rotation plates; this data is stored on the rewritable flash chips. It is possible to keep from one (SLC – Single Level Cell) to four (QLC – Quad Level Cell) bits into one cell. The main problem is the limit of rewriting cycles and higher prices. [7]

This table show comparison (advantages and disadvantages) between SSD and HDD discs.

SSD discs		HDD discs	
+	-	+	-
High speed	Limited rewriting cycles	Unlimited writing cycles	Low speed
Quiet operation	Higher price	Lower price	High noise level
Resistance to shock			Low resistance to shock
Low consumption			Higher consumption

Table 2: Comparison of SSD and HDD discs, [8]

The newest way of storing data is the cloud, which is a network of worldwide connected servers. Usually, the user pays only for cloud services that exactly used. This technology is applicable for data backup, executing applications, streaming videos, and other services. It is possible to get our files or services from anywhere and every time. [9]

Data backups are for problems with our current data. Backups are done on backup media, and the best possibility is regularly.

Backup can be done online (during normal computer operations) or offline (out of normal computer operations). It is possible to do full data backup; however, it is quite slow, because of the huge data amount. Another option is an incremental backup, where only new data are backup (from the date of full backup). The last type is differential data backup, which represents the difference between full backup and requires the full backup at the beginning of the backup cycle. [10]

RAID (Redundant Array of Inexpensive Disks) is a security method against disc failure. Security is realized by saving data on independent discs; in case of failure, the system can restore data. Level of security is given by number, for example RAID 0, RAID 1, RAID 2, RAID 0+1, RAID 1+0. RAID method can not replace data backup; it is only a supplement for bigger security of data. [11]

There are three types of network topologies – NAS (Network Attached Storage), DAS (Direct Attached Storage), SAN (Storage Attached Network). NAS is data storage connected to a local LAN network. DAS is data storage with direct access. SAN is a high-speed network of saving data, and it has a much more complicated architecture than in NAS and DAS. [12]

3. Evaluation of the current state

The company specializes in developing software for foreign customers in Europe, focusing on the security of information technology, telecommunications, logistics, mobile applications, e-shops, or information systems.

The backup center is in Prague, where are four servers. IT department is responsible for successful backups. There is the automatic backup of email via Microsoft Exchange Server for each employee; otherwise, there are backups of internal documents and project servers.

The backup center in Prague consists of four servers which include 119 virtual servers. The principle is based on storing only new data to existing ones, thus saving storage capacity. If necessary, to return to data from the past, there are using snapshot – actual state of the virtual machine.

Much software works with the servers and allows the virtualization, divided according to user level into several categories. Software VMware vSphere client enables the virtualization of servers; conversely, software Veeam Backup and Replication offers the management of the whole backup process. The cycle of storing new data (incremental backup) is set at regular intervals; it is currently performed every day at night so as not to restrict users. These backups are cyclically rewritten every four weeks (to storage space on discs). The full backup is done at night from Saturday to Sunday, where nobody is working. All processes are fully automatized, so there are eliminated human failures, and the IT department is just supervising the

whole process. The only necessary interventions are for creating a new virtual server or for defining regular backups.

The company has data backup processes on a high level; however, some improvements can be made by cloud services and adding NAS storage to every branch.

4. Design and implementation

The main goal is to improve the backup process in the company with options of cloud storages (Google Workspace, Dropbox Business, and Microsoft Azure) and NAS storage for each branch. The company is not using cloud services now; however, there are some problems with using these technologies (like cyberattacks, third party interventions), but it is still better to have other backups than just own server storage. One of the most secure cloud storage worldwide companies was chosen because the risk of bankruptcy (or other problems) is minimal.

Proposed cloud solutions:

- Google Workspace
- Dropbox Business
- Microsoft Azure

The first option is cloud storage Google Workspace, a set of cloud tools and software from Google company in the subscription form.

Picture 1: Logo of Google Workspace, [13]

This solution is complex and offers other services that the company can also use. Products are Gmail (company email with protection against spam), Meet (Videoconference), Disc (Cloud storage), Chat (Instant messaging), Calendar (shared calendars), Documents (shared text documents), Tables (shared tables), Presentations (shared presentation) and others. [14]

Google offers four different subscriptions: Business Started, Business Standard, Business Plus, and Business Enterprise. For this company, the best solution is a subscription Business Standard, which costs 9,36 € for an employee a month. If we count 1 € = 26 CZK and 278 employees, monthly costs are 67 654 CZK. [15]

In this subscription, each employee can use cloud storage 2 TB on Google Disc. It is sufficient for each employee and for storing important company data. [15]

The most significant advantage of this solution is the security and complexity of services; however, the company can not use all of them because they use mostly Microsoft products like Microsoft Teams, Outlook, and others. But in the case of using cloud services from Google, the company might decide to use other services also from them (like using Meet or Gmail). The second option is cloud storage, Dropbox business, which offers high secured cloud storage. [16]

Picture 2: Dropbox Business, [17]

This solution is only cloud storage compared to Google Workspace, and the company offers three different subscriptions: Standard, Advance, and Enterprise. For the company, the best solution is a Standard subscription, which costs 10 € per employee per month. That means monthly costs 72 280 CZK. [18] Standard subscription provides every employee 5 TB of cloud storage with 256-bit AES and SSL/TLS encryption. Also, it is possible to switch back for files up to 180 days. [18] This solution is attractive only if there are no other services suitable for the companies. The third option is the cloud platform, Microsoft Azure, using the data centers of Microsoft company. [19]

Picture 3: Microsoft Azure, [20]

There are also many complex solutions; however, cloud storage – Azure storage is the best variant for this specific task because the payment is only for the amount of stored data; storing 1 GB costs 0,049 € per month (around 1,2 CZK per month). Using 200 GB means costs 66 720 CZK per month. [21] The most significant advantage is that the company only pays for the amount of data it stores, which leads to financial savings, provided that less information is stored.

This table shows the comparison of suggested solutions:

	Google Workspace	Dropbox Business	Microsoft Azure
Capacity	2 TB	5 TB	used space
Price per month	67 654 CZK	72 280 CZK	1,2 CZK per 1 GB

Picture 4: Comparison of each suggestions

Google Workspace seems to be the best solution because of other services that the company can use. If they do not want additional services, better will be Dropbox Business because of more space.

Another improvement is NAS storage for every branch. There were chosen new and reliable products. After detailed market research, there three products were selected from companies Synology, QNAP, and Asustor.

Proposed cloud solutions:

- Synology DS3617xs
- QNAP TS-873AU-RP-4G
- Asustor Lockerstor 10 Pro-AS7110T

The first option is NAS storage Synology DS3617xs, which has slots for 12 discs with a maximum size of 168 TB. Also, it is possible to use a combination of SSD + HDD discs; the size could be 2,5'' or 3,5'' [22]. In the case of a smaller one, it is necessary to use an additional frame. It is essential to buy discs, the most reliable looks from Seagate because they have excellent reviews from real users. The concrete model is Seagate Exos 7E8 6TB, which is necessary 12 times. It is perfect to use RAID 5 or RAID 10 to increase the security of stored data. In the case of RAID 5, a capacity of 66 TB and 36 TB in the case of RAID 10. [23]

The second option is NAS storage QNAP TS-873AU-RP-4G. It has slots for 8 discs [24]. There is also necessary to buy discs, eight pieces Seagate Exos 7E8 6TB. It is better to use RAID technology – like RAID 5 (capacity 42 TB) or RAID 10 (capacity 24 TB).

The third option is Asustor Lockerstor 10 Pro-AS7110T. It is possible to use up to 10 discs. It is a very economical device with a consumption of 40 W in standby mode [25]. There is also necessary to buy ten additional discs, Seagate Exos 7E8 6TB. In the case of RAID 5, capacity is 54 TB, and in the case of RAID 10, it is 30 TB of storage space.

The following table shows the total costs of each solution. [23], [24], [25]

	Synology DS3617xs	QNAP TS-873AU-RP-4G	Asustor Lockerstor 10
Price for storage	67 673 CZK	55 123 CZK	62 131 CZK
Price for discs	62 436 CZK	41 624 CZK	52 030 CZK
Total price	130 109 CZK	96 747 CZK	114 161 CZK

Table 3: Comparison of costs

The solution from Synology is the most expensive. However financial options of the company are very open, so there are no limits. This solution has the best hardware parameters (like the most significant number of discs), so it would be the best. There will be four pieces of NAS storage (for each branch), so it is necessary to multiply four, so the total costs are 520 436 CZK. Three solutions of cloud storage and NAS storage were designed for the company. Cloud storage from Google, Google Workspace, and NAS storage Synology DS3617xs represent the best variant.

This table sum up costs of suggested solutions:

	Google Workspace	Synology DS3617xs
Price	811 848 CZK/ per year	520 436 CZK
Total price	1 333 284 CZK	

Table 4: Total costs of solutions

In total, it is 1 333 284 CZK, and this price will be increased every year by 811 848 CZK because of subscription costs. The advantages of the suggested solution are the usage of cloud services; backups are not only in one place, more extensive data security, and new and quality discs.

5. Conclusion

The paper aims at analyzed and design changes in data backup in the company. There were studied and compared options and technologies of data backup and data storage. After that, there was an analysis of the current state of the company. The design suggested three types of cloud storage (Google Workspace, Dropbox Business, and Microsoft Azure) and NAS storage (Synology DS3617xs, QNAP TS-873AU-RP-4G, and Asustor Lockerstor 10 Pro-AS7110T). Google Workspace was chosen as cloud storage because of additional services that the company might want to use. The NAS storage was chosen Synology DS3617xs because of the best hardware parameters and the number of discs. Joint cloud storage Google Workspace and NAS storage Synology DS3617xs are robust solutions for secure data backup.

References

[1] KAVÁNKOVÁ, Iva. *Data backup and data storages*. Brno, 2021. Diploma thesis. Brno University of Technology, Faculty of Business and Management . Supervisior Ing. Jiří Kříž, Ph.D. (in Czech)

[2] Koláček, M.: *Historie a současnost datových úložišť*. [Online; visited 26.01.2021]. https://www.svethardware.cz/historie-a-soucasnost-datovych-ulozist/23935 (in Czech)

[3] *Co je optický disk.* [Online; visited 03.05.2021]. https://www.sprava-site.eu/opticky-disk/ (in Czech)

[4] *USB* [Online; visited 05.05.2021]. https://cs.wikipedia.org/wiki/USB (in Czech)

[5] *Vše, co jste chtěli vědět o paměťově kartě typu SD.* [Online; visited 05.05.2021] https://www.datahelp.cz/clanky/vse-co-jste-chteli-vedet-opametove-karte-typu-sd (in Czech)

[6] *Jak funguje pevný disk.* [Online; visited 08.05.2021] https://notebook.cz/clanky/technologie/2007/HDD-technologie (in Czech)

[7] *SSD disky do serverů: jaké jsou aktuální trendy?* [Online; visited 09.05.2021] https://www.master.cz/blog/ssd-disky-do-serveru-trendy/ (in Czech)

[8] *HDD nebo SSD.* [Online; visited 09.05.2021] http://www.svetwebu.cz/?p=5691 (in Czech)

[9] *What is the cloud?* [Online; visited 06.05.2021] https://azure.microsoft.com/en-us/overview/what-is-the-cloud/

[10] *Definice a rotace záloh.* [Online; visited 01.05.2021] https://www.3s.cz/cs/odborna-sekce/detail/id/46-definice-a-rotace-zaloh (in Czech)

[11] *RAID.* [Online; visited 30.04.2021] https://www.giga-pc.cz/technicke-okenko/raid/ (in Czech)

[12] *NAS, SAN a DAS: Čím se od sebe liší datová úložiště?* [Online; visited 27.04.2021] https://www.master.cz/blog/san-nas-definice-rozdily-vyuziti/ (in Czech)

[13] Google Workspace – co, proč, jak? [Online; visited 10. 05.2021] https://www.smartzena.cz/google-workspace-co-proc-jak/ (in Czech)

[14] *Google Workspace.* [Online; visited 31.03.2021]
https://cs.wikipedia.org/wiki/Google_Workspace (in Czech)

[15] *Google Workspace.* [Online; visited 31.03.2021]
https://workspace.google.com/intl/cs/pricing.html

[16] *Dropbox.* [Online; visited 31.03.2021]
https://cs.wikipedia.org/wiki/Dropbox (in Czech)

[17] Dropbox Business Review [Online; visited 12. 05.2021]
https://www.betterbuys.com/dms/reviews/dropbox-business

[18] *Dropbox.* [Online; visited 31.03.2021]
https://www.dropbox.com/business/buy?sku=std#0

[19] *Microsoft Azure.* [Online; visited 31.03.2021]
https://cs.wikipedia.org/wiki/Microsoft_Azure (in Czech)

[20] Microsoft Azure Logo [Online; visited 10. 05.2021]
https://cs.wikipedia.org/wiki/Microsoft_Azure/media/Soubor:Microsoft_Azure_Logo.svg

[21] *Microsoft Azure.* [Online; visited 31.03.2021]
https://azure.microsoft.com/cs-cz/pricing/details/storage/

[22] *Synology DS3617xs* [Online; visited 10.05.2021]
https://www.alza.cz/synology-ds3617xs-d5561060.htm (in Czech)

[23] Seagate Exos 7E8 6TB Base FastFormat SAS. [Online; visited 10. 05.2021]
https://www.alza.cz/seagate-exos-7e8-6tb-base-fastformatsas-d5855791.htm?o=8#parametry (in Czech)

[24] *QNAP TS-873AU-RP-4G.* [Online; visited 10. 05.2021]
https://www.alza.cz/qnap-ts-873au-rp-4g-d6270241.htm (in Czech)

[25] Asustor Lockerstor 10 Pro-AS7110T. [Online; visited 10. 05.2021]
https://www.alza.cz/asustor-lockerstor-10-pro-as7110t-d6238597.htm#parametry (in Czech)