



# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

## FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

## ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

# ZÁLOHOVÁNÍ DAT A DATOVÁ ÚLOŽIŠTĚ

DATA BACKUP AND DATA STORAGES

## BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

## AUTOR PRÁCE

AUTHOR

Milan Olšina

## VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Kříž, Ph.D.

BRNO 2021

# Zadání bakalářské práce

Ústav: Ústav informatiky  
Student: **Milan Olšina**  
Studijní program: Systémové inženýrství a informatika  
Studijní obor: Manažerská informatika  
Vedoucí práce: **Ing. Jiří Kříž, Ph.D.**  
Akademický rok: 2020/21

Ředitel ústavu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává bakalářskou práci s názvem:

## Zálohování dat a datová úložiště

### Charakteristika problematiky úkolu:

Úvod  
Cíle práce, metody a postupy zpracování  
Teoretická východiska práce  
Analýza současného stavu  
Vlastní návrhy řešení  
Závěr  
Seznam použité literatury  
Přílohy

### Cíle, kterých má být dosaženo:

Cílem práce je analyzovat systém zálohování dat z digitalizačního oddělení organizace. Na základě analýzy současného stavu a konkrétních požadavků organizace poté navrhnout úpravy systému včetně plánu na zakomponování nového hardwaru.

### Základní literární prameny:

CUBR, Ladislav. Dlouhodobá ochrana digitálních dokumentů. Praha: Národní knihovna České republiky, 2010. ISBN 978-80-7050-588-5.

DEMBOWSKI, Klaus. Mistrovství v hardware. Brno: Computer Press, 2009. ISBN 978-80-251-2310-2.

LACKO, Luboslav. Osobní cloud pro domácí podnikání a malé firmy. Brno: Computer Press, 2012. ISBN 978-80-251-3744-4.

PECINOVSKÝ, Josef. Archivace a komprimace dat. Praha: Grada Publishing a.s., 2003. ISBN 80-24-0659-8.

POŽÁR, Josef. Manažerská informatika. Plzeň: Aleš Čeněk, 2010. ISBN 978-80-7380-276-9.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2020/21

V Brně dne 28.2.2021

L. S.

-----  
Mgr. Veronika Novotná, Ph.D.  
ředitel

-----  
doc. Ing. Vojtěch Bartoš, Ph.D.  
děkan

## **Abstrakt**

Tato bakalářská práce se zaměřuje na posouzení současného stavu zálohování dat a datových úložišť v prostředí digitalizačního pracoviště Vědecké knihovny v Olomouci. Na základě provedené analýzy a požadavků organizace je vytvořen vlastní návrh řešení, který upravuje systém zálohování dat včetně navržení nového hardwaru.

## **Klíčová slova**

data, zálohování, úložiště, pevný disk, RAID

## **Abstract**

This bachelor's thesis focuses on assessing the current state of data backup and data storage of the Olomouc Research Library's digitisation workplace. A proposal for a solution is based on the conducted analysis and the requirements of the organisation. The solution recommends modification of the data backup system including the design of new hardware.

## **Keywords**

data, backup, storage, hard disk drive, RAID

### **Bibliografická citace**

OLŠINA, Milan. *Zálohování dat a datová úložiště* [online]. Brno, 2021 [cit. 2021-05-12]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/135476>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav informatiky. Vedoucí práce Jiří Kříž.

### **Čestné prohlášení**

Prohlašuji, že předložená bakalářská práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 12. května 2021

.....

*podpis autora*

## **Poděkování**

Tímto bych rád poděkoval především Ing. Jiřímu Křížovi, Ph.D. za vedení mé bakalářské práce. Dále bych rád poděkoval zaměstnancům Vědecké knihovny v Olomouci za poskytnutí podkladů pro vypracování této práce.

# OBSAH

ÚVOD .....	11
CÍLE PRÁCE, METODY A POSTUPY ZPRACOVÁNÍ .....	12
1 TEORETICKÁ VÝCHODISKA PRÁCE .....	13
1.1 Data .....	13
1.2 Zálohování dat.....	13
1.2.1 Archivace dat .....	14
1.3 Typy záloh.....	14
1.3.1 Úplná záloha .....	14
1.3.2 Přírůstková záloha.....	14
1.3.3 Rozdílová záloha.....	14
1.4 Úložná média .....	15
1.4.1 Magnetické pásky .....	16
1.4.2 Optické disky .....	17
1.4.3 USB flash disky .....	17
1.4.4 Pevné disky .....	18
1.4.5 Solid-state drive .....	19
1.4.6 Cloudová úložiště .....	20
1.5 Topologie síťových úložišť .....	21
1.5.1 Direct-attached storage (DAS).....	21
1.5.2 Network-attached storage (NAS).....	21
1.5.3 Storage area network (SAN).....	22
1.6 Systém RAID .....	22
1.6.1 RAID 0.....	23
1.6.2 RAID 1.....	23
1.6.3 RAID 2.....	24

1.6.4	RAID 3.....	24
1.6.5	RAID 4.....	25
1.6.6	RAID 5.....	26
1.6.7	RAID 6.....	27
1.6.8	Kombinace .....	28
1.7	Hot plug, hot swap a hot spare .....	28
2	ANALÝZA SOUČASNÉHO STAVU .....	30
2.1	Popis organizace.....	30
2.1.1	Organizační struktura.....	31
2.1.2	Digitalizační pracoviště .....	31
2.2	Hardware .....	32
2.2.1	Skenery .....	32
2.2.2	Pracovní stanice .....	32
2.2.3	Servery .....	33
2.2.4	Úložiště .....	33
2.2.5	Síťová infrastruktura .....	34
2.2.6	Ostatní HW .....	34
2.3	Software .....	35
2.4	Proces zálohování.....	37
2.5	Shrnutí.....	38
3	VLASTNÍ NÁVRHY ŘEŠENÍ .....	39
3.1	Nový hardware .....	39
3.1.1	Síťová infrastruktura.....	40
3.1.2	Pracovní stanice .....	40
3.1.3	Servery .....	41
3.1.4	Disková pole .....	42

3.2	Zálohovací plán.....	45
3.3	Zhodnocení navrhovaného řešení .....	47
	ZÁVĚR .....	49
	SEZNAM POUŽITÝCH ZDROJŮ .....	50
	SEZNAM POUŽITÝCH OBRÁZKŮ .....	53
	SEZNAM POUŽITÝCH TABULEK.....	54

## ÚVOD

V dnešní době jsou data pravděpodobně nejcennějším informačním aktivem, ať už se jedná o data firem a organizací nebo pouze o data jednotlivců. Zároveň jejich objem neustále roste a lze očekávat, že tento trend bude v budoucnu pokračovat. To klade velké nároky na bezpečné uložení všech těchto dat, která jsou mnohdy klíčová pro fungování informačních systémů.

Ztráta důležitých dat může vzniknout z mnoha důvodů. Zápavy, požáry, selhání hardwarových komponentů, lidské chyby nebo dokonce úmyslná vymazání představují jen krátký výčet možných hrozeb. Je zřejmé, že data je potřeba proti těmto hrozbám chránit. Nejčastějším způsobem takovéto ochrany je zálohování dat.

Zálohováním se rozumí činnost zkopírování dat z primárního úložiště na záložní médium. Toto médium by mělo splňovat určitá kritéria jako například doba životnosti nebo fyzicky oddělené uložení od původních dat. V případě jakéhokoliv problému s původními daty je poté použita tato záložní kopie pro uvedení datové základny do původního stavu. Zálohování by se mělo týkat nejen dat organizací, ale i dat jednotlivců. Stále však existují zástupci obou jmenovaných skupin, kteří tento proces podceňují většinou až do doby, kdy o svá cenná data nenávratně přijdou.

## **CÍLE PRÁCE, METODY A POSTUPY ZPRACOVÁNÍ**

Cílem práce je analyzovat systém zálohování dat z digitalizačního oddělení organizace. Na základě analýzy současného stavu a konkrétních požadavků organizace poté navrhnout úpravy systému včetně plánu na zakomponování nového hardwaru.

Metody a postupy, které jsou použity při zpracování bakalářské práce, zahrnují studium odborné literatury a relevantních elektronických zdrojů, na jejichž základě jsou zpracována teoretická východiska práce, a konzultace se zaměstnanci organizace. Ty mají za cíl získat všechny potřebné informace pro vyhotovení analýzy současného stavu prostředí digitalizačního pracoviště. Tyto kapitoly společně s požadavky organizace poslouží jako podklady pro návrh vlastního řešení.

# 1 TEORETICKÁ VÝCHODISKA PRÁCE

Tato kapitola bakalářské práce je věnována teoretickým východiskům. Ta budou sloužit jako podklady pro následující části práce.

## 1.1 Data

Nejdříve je nutné definovat samotný pojem **data**. Existuje mnoho různých definic a pohledů na data. Pro tuto práci bude použita následující definice dat: „*Data jsou vhodným způsobem vyjádřené zprávy, které vypovídají o světě a jsou tudíž srozumitelné pro příjemce, jímž může však být nejen člověk, ale i technický prostředek*“ (1).

Jako data lze tedy označit čísla, slova, obrázky a mnoho dalších podobných údajů, které byly určitým systematickým způsobem zaznamenány na konkrétní médium. Interpretací a zpracováním těchto dat, jakožto základního materiálu, lze získat **informace**, které pro nás jakožto pro příjemce přinášejí užitek (1).

Data jsou vnímána jako ta nejcennější součást téměř každého počítače, případně informačního systému. V porovnání s poškozením hardwaru či softwaru je částečná nebo úplná ztráta dat mnohdy mnohem závažnějším problémem ať už pro jednotlivce nebo celou organizaci. Ztráta dat může být přitom způsobena celou řadou událostí jako například selhání hardwaru, krádež, úmyslné vymazání, přírodní katastrofa a podobně. Je tedy nutné jejich ochraně věnovat značnou pozornost (2).

## 1.2 Zálohování dat

Nejčastějším způsobem ochrany proti ztrátě dat je **zálohování**. Jedná se o proces vytváření kopie dat. Tato kopie by měla zajistit snadné a rychlé obnovení dat do původního stavu v případě potřeby. Předmětem zálohování jsou především operativní data, která podléhají častým změnám a jsou nutná k běžnému chodu organizace. Původní data zůstávají i po vytvoření kopie na svém místě a je možné s nimi dále pracovat (2).

### **1.2.1 Archivace dat**

**Archivací** dat se rozumí dlouhodobé uložení dat za účelem případného pozdějšího využití. Ukládaná data již prošla zpracováním a není potřeba je v průběhu času modifikovat. Primárním cílem archivace je bezpečné uložení těchto dat v dlouhodobém časovém horizontu. Na rozdíl od zálohování zde není kladen tak velký důraz na rychlou obnovu těchto dat (2).

## **1.3 Typy záloh**

Pro zmenšení objemu zálohovaných dat se používá několik základních typů záloh. Nejznámějšími jsou úplná, přírůstková a rozdílová. V praxi se tyto základní typy používají společně v různých kombinacích.

### **1.3.1 Úplná záloha**

Úplná záloha je nejjednodušším typem zálohy. Její princip spočívá v kopírování celé vybrané množiny dat na záložní médium a označení příslušných souborů jako zálohovaných pomocí vynulování archivního atributu. Výhodou této zálohy je především snadná obnova dat, jelikož každá úplná záloha je samostatná a nezávislá na ostatních zálohách. Nevýhody tohoto typu zálohy jsou vázány na fakt, že se zálohují všechna vybraná data bez ohledu na to, zda byla od poslední zálohy změněna, což klade velké nároky na úložiště i na čas vytváření zálohy (3).

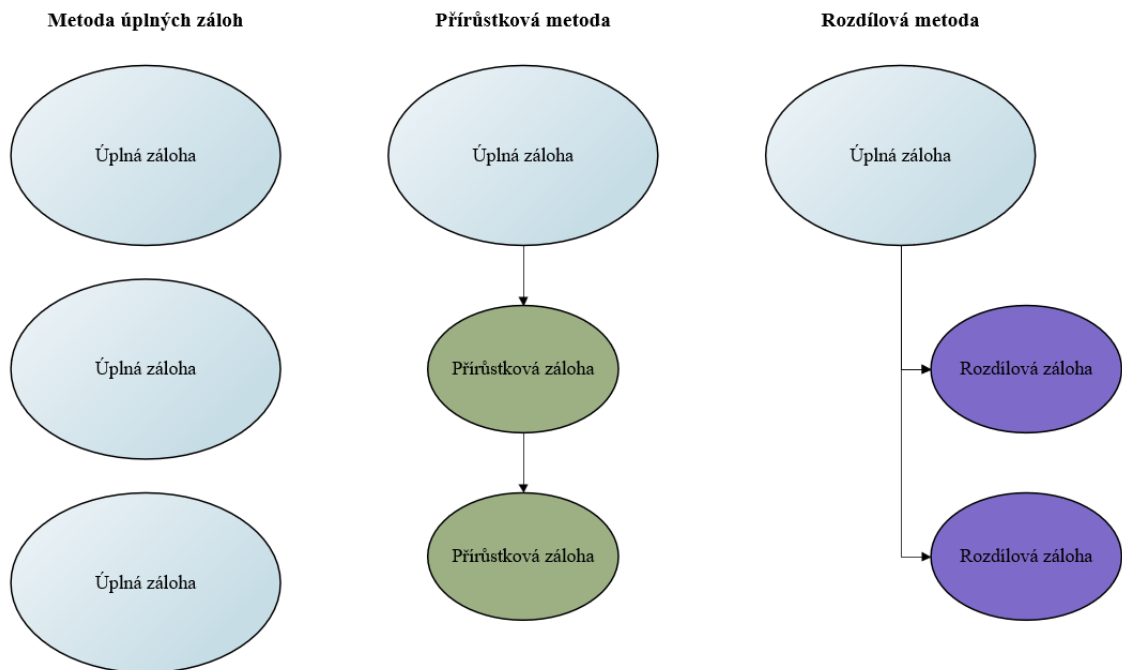
### **1.3.2 Přírůstková záloha**

Přírůstková metoda, také někdy zvaná inkrementální, zálohuje pouze ta data, která byla změněna od poslední úplné nebo přírůstkové zálohy a označuje je za zálohované. Pracuje tedy s výrazně menším objemem dat než záloha úplná a díky tomu je její provedení o mnoho rychlejší. Hlavní nevýhodou je obtížnější obnova dat, kdy je potřeba nejdříve obnovit úplnou zálohu a poté postupně všechny na ni navazující zálohy přírůstkové (3).

### **1.3.3 Rozdílová záloha**

Rozdílová či diferenční metoda zálohuje data změněná od poslední úplné zálohy s tím rozdílem, že tato data neoznačí jako zálohovaná. Další vytvořené rozdílové zálohy jsou tím pádem na sobě zcela nezávislé. Objem zálohovaných dat je tedy menší než u úplné

zálohy, ale při opakovaném použití větší než u přírůstkové. Pro obnovení je potřeba mít k dispozici poslední úplnou zálohu a poslední rozdílovou zálohu, starší rozdílové zálohy nejsou potřebné (3).



**Obrázek č. 1: Zjednodušené schéma typů záloh**  
(Zdroj: Vlastní zpracování)

## 1.4 Úložná média

Tato podkapitola je věnovaná popisu jednotlivých úložných médií používaných při zálohování. Záloha by vždy měla probíhat na místo oddělené od klasického úložiště pro pracovní data. Úložiště sloužící k zálohování by mělo být používáno pouze k této činnosti. Nelze jednoznačně říci, které médium je pro účely zálohování nejvhodnější, jelikož každé nabízí specifické výhody a nevýhody. Volba vhodného typu závisí na množství faktorů jako například objem a kvalita dat, dostupnost, potřeby změn dat, doba uchování zálohy a podobně (4).

### 1.4.1 Magnetické pásky

Magnetické pásky jsou jedním z nejstarších médií pro ukládání dat, ale jejich použití je stále velmi časté. Existuje několik různých technologií pásek, nejznámější a nejpoužívanější je v současnosti technologie LTO (Linear Tape-Open) vyvinutá jako otevřený formát konsorciem firem HP, IBM a Quantum. Samotná páska je namotána na cívce uvnitř kazety, která se zavádí přímo do páskových knihoven (5).

Hlavní výhodou magnetických pásek je vysoká kapacita pro ukládání dat. V současnosti se pohybuje v řádu jednotek či nižších desítek TB. Díky neustálému vývoji se však tyto kapacity stále zvyšují, přičemž lze očekávat zdvojnásobení kapacity každé dva až tři roky. Další podstatnou výhodou je jejich dlouhá životnost až několik desítek let, která je řádově vyšší než u ostatních běžně používaných médií. Cena jednotlivých pásek je s přihlédnutím na poskytovanou kapacitu velmi nízká, ale pořízení příslušné páskové knihovny může být velmi nákladnou investicí. Znatelnou nevýhodou pásek je pomalý přístup k datům, který je dán jejich fyzickou konstrukcí (6).



Obrázek č. 2: Magnetická páska typu LTO s kazetou  
(Zdroj: 6)

## 1.4.2 Optické disky

Optické disky jsou plastové kotouče o průměru 12 cm. Zápís a čtení dat z nich probíhá za pomoci laserového světla. Existují varianty pouze pro čtení (ROM), pouze pro jeden zápis (R) a pro opakovaný zápis (RW). Z několika typů optických disků jsou jejich nejznámějšími variantami CD, DVD a Blue-ray, který však není s předchozími typy kompatibilní díky použití laserového paprsku jiné vlnové délky. Kapacity závisí na konkrétním typu disku a pohybují se v rozmezí 650-900 MB u CD, 4,7-17 GB pro DVD a 25-128 GB v případě Blue-ray disků. Mechanikou pro práci s CD a DVD dnes disponuje většina počítačů, Blue-ray disky se z důvodu nutnosti speciální mechaniky používají spíše v oblasti distribuce filmů a konzolových her. Hlavní nevýhodu použití optických disků v oblasti zálohování dat je tak jejich nízká kapacita a nízká životnost (7).



**Obrázek č. 3: Optické disky-DVD**  
(Zdroj: Vlastní zpracování)

## 1.4.3 USB flash disky

USB flash disky jsou dnes velmi populární metodou přenosu dat mezi počítači. Používají polovodičovou paměť typu flash, která neobsahuje žádné pohyblivé součástky. Připojují se přes univerzální sběrnici USB, kterou je dnes vybaven každý počítač a také většina ostatních elektronických zařízení. Dříve pracovaly s rozhraním USB 2.0, dnes se dostává do popředí jeho nástupce USB 3.0, jenž dovoluje dosáhnout vyšších přenosových rychlostí. Kapacity se pohybují v širokém rozmezí od několika GB až po několik TB.

Mezi nevýhody tohoto typu média lze zařadit relativně krátkou životnost při intenzivním používání a poměrně vysoká cena u USB flash disků s vysokou kapacitou. Nutno také dodat, že díky velikosti a stylu používání existuje značné riziko ztráty či odcizení média (7).



**Obrázek č. 4: USB flash diský**  
(Zdroj: Vlastní zpracování)

#### **1.4.4 Pevné disky**

Pevný disk (HDD, hard disk drive) je v současnosti nejběžnějším datovým úložištěm. Jedná se o elektromechanické zařízení, které k zápisu a čtení dat používá změny magnetického toku. Skládá se z několika nad sebou umístěných kovových kotoučů, tzv. ploten, hlavy pro čtení a zápis dat, válcového šroubu sloužícímu pro otáčení ploten a řídicí elektroniky. Nejčastěji se používají disky o velikosti 3,5 palce, popřípadě 2,5 palce. Připojení disků k počítači, serveru nebo diskovému poli se realizuje nejčastěji pomocí rozhraní SATA (Serial ATA) nebo SAS (Serial Attached SCSI). Kapacity dnešních disků se pohybují v řádu několika TB, přístupová doba činí pouze několik milisekund. Výhodou pevných disků je jejich nízká cena a dobrá dostupnost. V případě diskových polí tvořených větším počtem disků lze zajistit i určitou ochranu uložených dat, například pomocí systému RAID. Pevné disky disponují řadou nevýhod, mezi které patří vysoká spotřeba energie, hlučnost a přirozené opotřebení součástí dané poměrně složitou konstrukcí s pohyblivými částmi. Jsou také citlivé na otřesy a náchylné

k mechanickému poškození. To může vést k výskytu chyb, případně selhání celého disku (7).



Obrázek č. 5: Pevný disk  
(Zdroj: 8)

#### 1.4.5 Solid-state drive

Solid-state drive, v češtině nejčastěji označován jako SSD disk, je zařízení pro ukládání dat, které je přímým konkurentem klasických pevných disků. Data se u tohoto typu disku ukládají na polovodičovou paměť typu flash. Díky tomu jsou SSD disky konstrukčně jednodušší než pevné disky, jelikož neobsahují žádné pohyblivé části. Díky tomu se také mohou vyrábět v menších velikostech, nejčastější je 2,5 palce. K připojení využívají rozhraní SATA, PCI Express nebo M.2. Hlavní výhodou SSD disků je jejich přístupová a přenosová rychlost, která je díky jejich konstrukci mnohdy řádově vyšší než u pevných disků. Díky tomu se těší stále větší oblibě v řadách uživatelů. Z hlediska kapacity SSD většinou mírně zaostávají. Některé typy sice dosahují kapacity několika TB podobně jako pevné disky, ale většina zůstává nejčastěji na 500 GB. Cena je při přepočtu na uložená data také vyšší. Hlavní nevýhodou SSD disků je relativně nízká životnost daná omezeným počtem zápisů na flash paměť (9).



**Obrázek č. 6: SSD**  
(Zdroj: 8)

#### **1.4.6 Cloudová úložiště**

Pro ukládání dat lze kromě klasických fyzických médií použít i cloudových služeb. Cloud computing lze definovat jako „metodu poskytování IT ve formě služby, přičemž zákazník platí jen za to, co právě využívá“ (10).

Službou může být prakticky jakýkoliv prostředek IT jako například aplikace, úložiště, operační systémy, servery a podobně (10).

Výhodou cloudového řešení je především rychlá a snadná implementace, na rozdíl od realizování srovnatelného řešení za pomoci vlastní infrastruktury. S tím souvisí také eliminace podstatné části nákladů vynaložených na správu a údržbu vlastního hardwaru a softwaru (10).

Hlavní nevýhodou cloudu je závislost na stabilním internetovém připojení. V případě jeho selhání není možné k poskytovaným službám přistupovat vůbec. Další nevýhodou je závislost na poskytovateli služeb, který může například navýšit jejich cenu. Existuje i krajní možnost krachu poskytovatele, ale ta je v případě velkých, dlouho působících IT společností značně nepravděpodobná. Některá cloudová řešení poskytují oproti desktopovým méně funkcí nebo mají méně komfortní uživatelské rozhraní. Poslední, ale zásadní věcí, kterou je třeba zvážit, je legislativa. Problémy můžou nastat především

v případě, kdy jsou poskytovatel a konzument služby z různých zemí s jinými právními normami (10).

## **1.5 Topologie síťových úložišť**

Pro ukládání dat v prostředí organizace, kde existuje potřeba vícenásobného přístupu ke stejným datům, se používají úložiště zapojená do místní sítě, nejčastěji se přitom jedná o disková pole, tj. speciální zařízení obsahující několik klasických pevných disků. Pro servery i klientské stanice představuje diskové pole jeden celek (jeden disk). Existuje několik způsobů, jak diskové pole připojit do sítě. Volba konkrétního řešení je vždy závislá na požadavcích konkrétní organizace (11).

### **1.5.1 Direct-attached storage (DAS)**

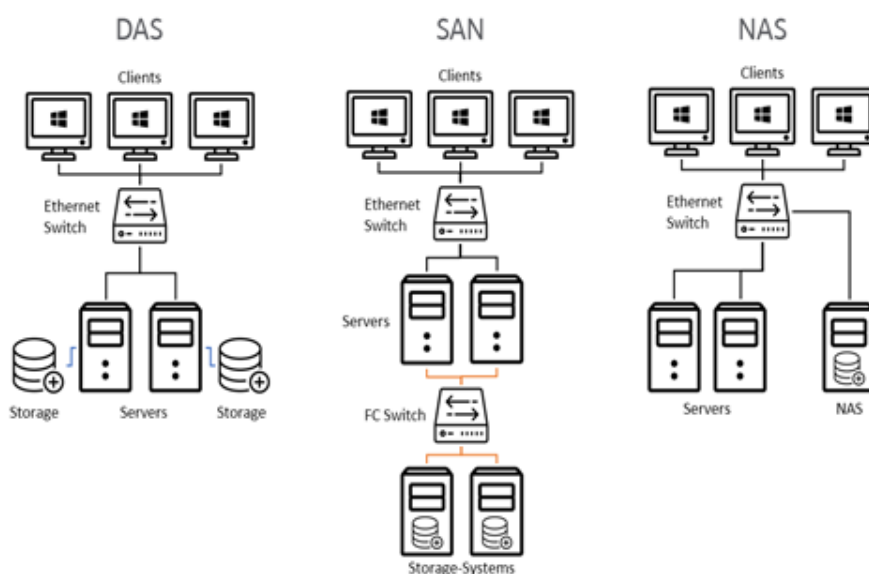
U tohoto typu zapojení je diskové pole přímo propojené se serverem, případně server samotný disponuje několika pevnými disky pro účely ukládání dat. V obou případech platí, že přístup k datům probíhá přes server. To s sebou přináší jak výhody, tak i nevýhody. Výhodou je relativně rychlý přístup k datům, nevýhodou je složitější sdílení dat s ostatními zařízeními v síti. Také je zde možnost zpomalení jiných serverových funkcí při přenosu dat (12).

### **1.5.2 Network-attached storage (NAS)**

Jedná se o typ zapojení, při kterém je diskové pole připojeno do místní sítě za účelem sdílení dat mezi polem a servery nebo klientskými stanicemi. Ty vnímají pole jako jeden logický disk. Data jsou sdílena na úrovni jednotlivých souborů pomocí protokolů jako je NFS, SMB nebo AFP. Oddělení úložiště dat od serveru představuje výhodu ve formě menších nároků na výkon serveru. Na druhou stranu nutnost přenosu dat přes síť klade zvýšené nároky na síťovou infrastrukturu, především při použití nestavových protokolů. Z pohledu ceny zařízení a nákladů na provoz je NAS ve většině případů velmi efektivním řešením pro uložení dat (12).

### 1.5.3 Storage area network (SAN)

Tento způsob zapojení umísťuje úložiště do speciální datové sítě, která je oddělená od běžné LAN sítě, a spojuje je se servery. Tato spojení bývají často redundantní kvůli zajištění dostupnosti. Data z úložiště jsou sdílena na úrovni bloků dat, nikoliv souborů. Samotnou síť lze realizovat buďto pomocí klasických ethernet switchů a kabelů nebo za pomoci speciálních optických switchů a infrastruktury (Fibre Channel) v případě požadavků na vysoký výkon a rychlost. Toto řešení je ovšem značně nákladné (12).



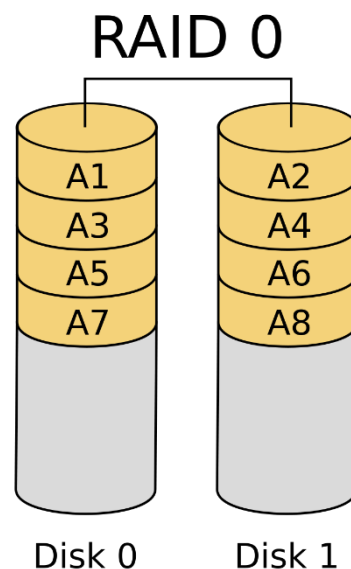
Obrázek č. 7: Schéma DAS, SAN a NAS  
(Zdroj: 13)

### 1.6 Systém RAID

RAID (zkratka z anglického Redundant Array of Independent Disks) je systém pro ochranu dat diskového pole v případě selhání pevného disku. Část kapacity pole se vyčlení pro ukládání redundantních dat, díky kterým je možné v případě potřeby obnovit soubory uložené na discích. Různá úroveň zabezpečení je značena čísly (nejčastější jsou úrovně RAID 0, RAID 1, RAID 5 a RAID 6). RAID nelze považovat za zálohu dat, jelikož nesplňuje podmínky zálohy jako odlišné fyzické uložení nebo možnost obnovy starší verze. Proto je vždy třeba zahrnout disková pole se systémem RAID do plánu záloh (7, 14).

### 1.6.1 RAID 0

V případě, že je pole v zapojení RAID 0, se nejedná o skutečný RAID, jelikož data nejsou ukládána redundantně. Existují dvě možnosti realizace: lineární ukládání a prokládání. První možnost zahrnuje ukládání dat na disky postupně. Až se zaplní první, přejde se na další v pořadí. Při selhání jednoho z disků budou dotčeny jen soubory na tomto konkrétním disku. Druhou metodou je prokládání (anglicky striping), při kterém jsou data ukládána na disky střídavě. Tato metoda tak neposkytuje žádnou ochranu dat, pouze zvyšuje rychlost čtení a zápisu. V případě výpadku jednoho z disků existuje velká pravděpodobnost, že budou poškozeny všechny uložené soubory. Kapacita pro uložená data je v obou případech rovna součtu kapacit všech použitých disků (7, 14).

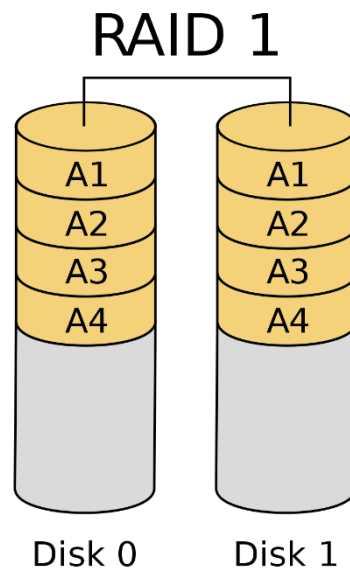


Obrázek č. 8: Schéma RAID 0, metoda prokládání  
(Zdroj: 14)

### 1.6.2 RAID 1

Uspořádání RAID 1 poskytuje velmi jednoduchou, ale poměrně efektivní ochranu uložených dat. Data jsou ukládána současně na dva (popř. i více) disky. Tomuto systému se říká zrcadlení (anglicky mirroring). Pokud jeden disk selže, je možné ihned pracovat s kopií dat na druhém disku. Rizikem zůstává selhání řadiče disků, při kterém přestanou fungovat oba disky. To je řešeno současným použitím dvou řadičů (duplexing). Rychlost zápisu je o něco pomalejší z důvodu ukládání na více disků, rychlost čtení je naopak

o něco vyšší. Nevýhodou tohoto zapojení je fakt, že kapacita celého pole je poloviční oproti součtu kapacit jednotlivých disků z důvodu dvojnásobného zápisu (7, 14).



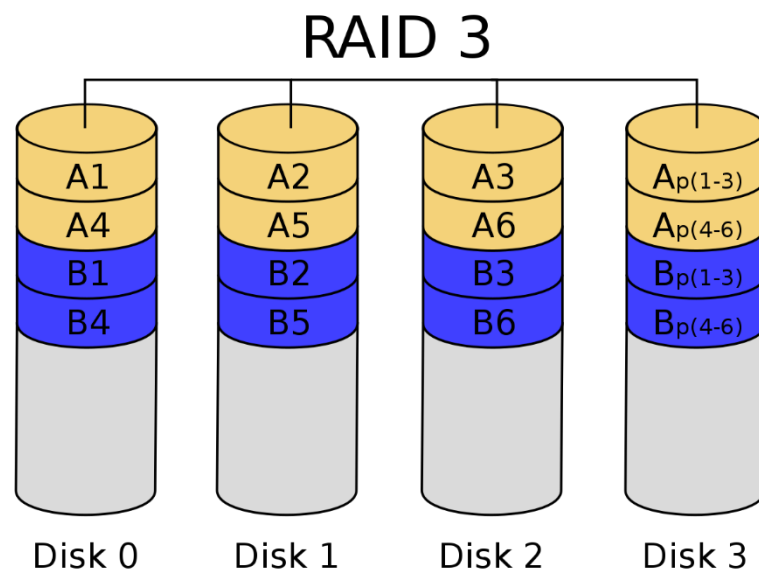
**Obrázek č. 9: Schéma RAID 1**  
(Zdroj: 14)

### 1.6.3 RAID 2

Při zapojení RAID 2 jsou data ukládána na disky prokládáním po bitech a pro každý uložený byte je spočítán kontrolní kód pro opravu chyb (tzv. Hammingův kód), který je zapsán na kontrolní disk. Tento způsob zapojení se dnes v praxi téměř nevyužívá (7, 14).

### 1.6.4 RAID 3

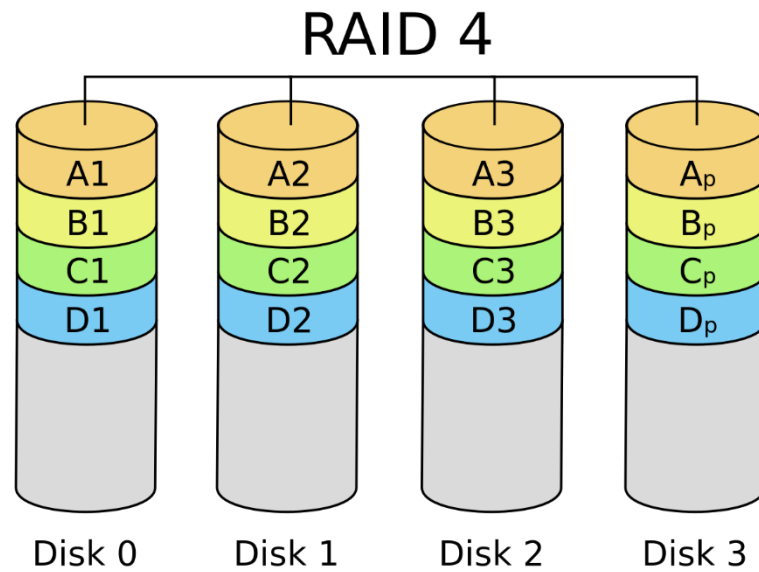
U RAID 3 je místo kódu pro opravu chyb použit paritní bit. K použití jsou nutné minimálně 3 disky, přičemž na 2 jsou zapisována data prokládáním po bytech a poslední disk je použit pro paritní bit. Toto zapojení je odolné proti výpadku jednoho disku. Pokud selže jeden z datových disků, lze data dopočítat pomocí parity. V případě selhání paritního disku jsou data na ostatních discích neporušena. Kapacitu lze spočítat jako součet kapacit datových disků (7, 14).



Obrázek č. 10: Schéma RAID 3  
(Zdroj: 14)

### 1.6.5 RAID 4

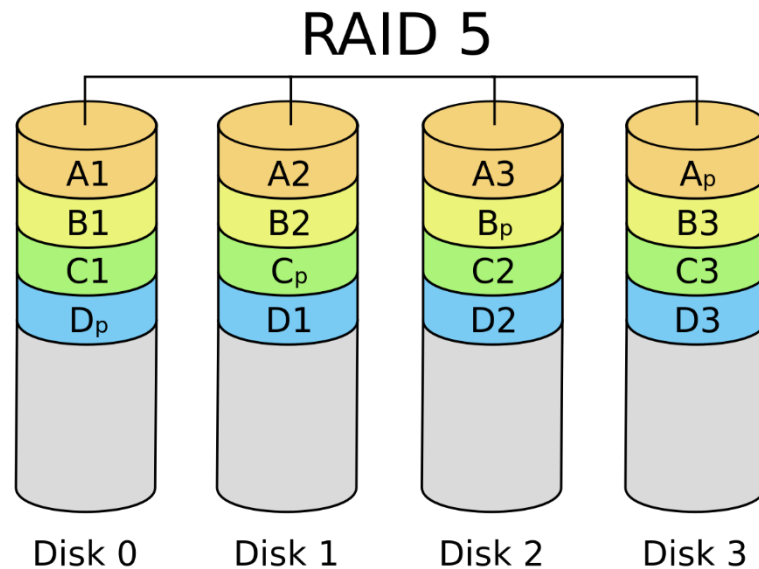
Jedná se o velmi podobné zapojení jako RAID 3 s tím rozdílem, že data jsou prokládána po blocích (sektorech), nikoliv po bitech nebo bytech jako v předchozích verzích. To přináší výhodu v možnosti provádět více operací čtení zároveň. Nevýhodou je fakt, že při zápisu, který může navíc probíhat jen jeden v daném čase, je nutné navíc zapsat informace na paritní disk. Opět je zapotřebí minimálně 3 disků, zapojení je odolné proti výpadku jednoho libovolného disku a kapacita je dána počtem datových disků (7).



Obrázek č. 11: Schéma RAID 4  
(Zdroj: 14)

### 1.6.6 RAID 5

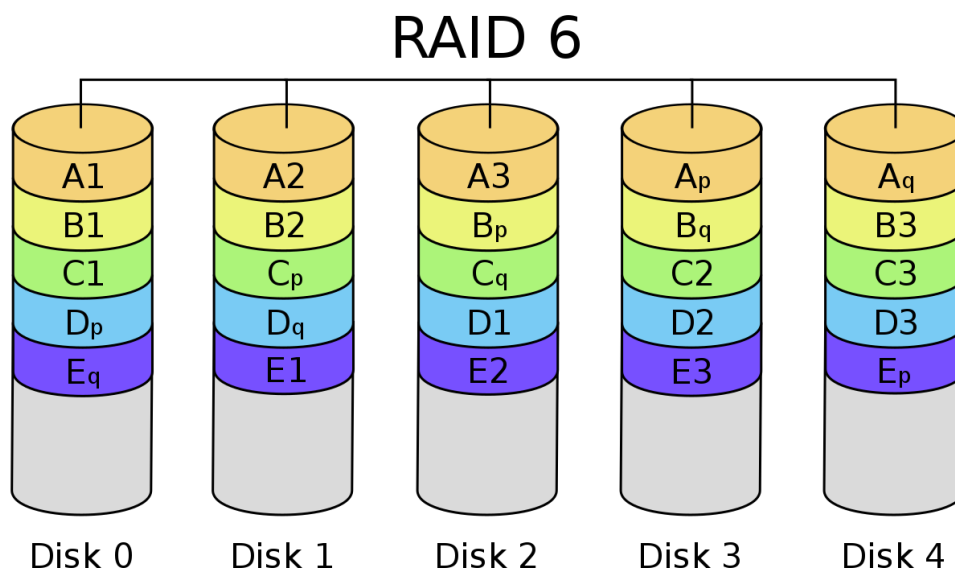
RAID 5 na rozdíl od všech předchozích konfigurací nepoužívá jeden disk pro ukládání informací o paritě, protože ty se zapisují střídavě na všechny disky mezi bloky dat. Data jsou stejně jako v RAID 4 prokládána po blocích. K fungování je potřeba alespoň 3 disků, přičemž zapojení je odolné proti výpadku jednoho libovolného disku. Čtení je rychlejší díky možnosti číst zároveň z více disků, zápis je zpomalen nutností dopočítat informace o paritě (7, 14).



Obrázek č. 12: Schéma RAID 5  
(Zdroj: 14)

### 1.6.7 RAID 6

RAID 6 je podobným řešením jako RAID 5 s tím rozdílem, že používá dvojitou paritu. Každá z nich je počítána jiným způsobem a uloženy jsou společně s daty prokládáním na všechny dostupné disky. Výhodou tohoto způsobu je odolnost vůči selhání dvou disků. Minimální počet disků je 4, ale kapacita je při použití 4 disků z důvodu nutnosti uložení informací o paritě pouze poloviční. Je tedy vhodné použít větší počet disků (například při použití 6 disků je kapacita pro data  $\frac{2}{3}$  celkové kapacity). Rychlost čtení z disků je podobná jako u RAID 5 (lze číst z více disků zároveň), ale zápis je z důvodu výpočtu dvou parit pomalejší (14).



Obrázek č. 13: Schéma RAID 6  
(Zdroj: 14)

### 1.6.8 Kombinace

Někdy se v praxi používají i kombinace základních typů RAID. Nejčastějšími jsou RAID 0+1, RAID 1+0, RAID 5+0, RAID 6+0 nebo RAID 10+0. Existují však i další možnosti. Hlavním důvodem vytváření kombinací je zvýšení redundance dat a tím i bezpečnosti, popřípadě rychlosti čtení a zápisu do celého pole (14).

### 1.7 Hot plug, hot swap a hot spare

Jedná se o pojmy související s disky a diskovými poli. Nemusí se však vztahovat jen k nim, ale třeba i k periferním zařízením, zdrojům a podobně. Popisují následující vlastnosti:

- **hot plug** – značí schopnost do zařízení přidat další komponent bez nutnosti přerušení činnosti celého systému (tzv. za běhu),
- **hot swap** – schopnost do zařízení nejen přidat, ale i z něj odstranit komponenty (např. vadné disky v diskových polích) bez nutnosti ukončení činnosti systému (15),
- **hot spare** – jedná se o komponent, který je k zařízení připojen, ale aktivním se stane až v případě nutnosti (např. disk v poli, na který se začne zapisovat v případě poruchy jiného disku) (16).

Pojmy hot plug a hot swap se v některých případech zaměňují nebo jsou považovány za rovnocenné s tím, že většina takto označených zařízení mají funkcionalitu odpovídající definici hot swap. Záleží ovšem vždy na konkrétním typu zařízení.

## 2 ANALÝZA SOUČASNÉHO STAVU

V této kapitole bakalářské práce bude popsán současný stav Vědecké knihovny v Olomouci, konkrétně jejího digitalizačního pracoviště. Analýza se bude týkat především používaného hardwaru, softwaru a procesu zálohování dat. V závěru kapitoly budou zhodnoceny výhody a nevýhody současného řešení.

### 2.1 Popis organizace

Vědecká knihovna v Olomouci byla založena v roce 1566 jako součást jezuitské akademie. Ta byla o 7 let později povýšena na univerzitu a knihovna se stala univerzitní knihovnou. Knihovní fond se rozšiřoval až do období třicetileté války, kdy během okupace Olomouce švédskými vojsky byla většina knih odvezena jako válečná kořist. Po třicetileté válce se knihovna začala opět doplňovat jak nákupy, tak četnými dary. Roku 1704 byl pořízen svazkový katalog, který se dodnes zachoval a dle něj knihovna čítala celkem asi 4300 svazků. Koncem 18. století byla univerzita včetně knihovny zestátněna a na krátkou dobu přestěhována do Brna. Poté však byla navracena zpět do Olomouce. V roce 1807 získala knihovna právo povinného výtisku pro území Moravy, díky čemuž dnes disponuje nejbohatší sbírkou moravik v ČR. Zrušení univerzity v roce 1860 znamenalo osamostatnění knihovny a také její přejmenování na Studijní knihovnu v Olomouci. Na počátku 20. století byla knihovna přestěhována do budovy bývalé teologické fakulty. V roce 1935 bylo ministerstvem školství odebráno právo povinného výtisku a o rok později byla knihovna opět přestěhována, tentokrát do budovy bývalé banky na Bezručově ulici, kde působí dosud. Po druhé světové válce byla v Olomouci obnovena univerzita a název knihovny se opět změnil na Univerzitní, zůstala ale samostatnou institucí a roku 1947 jí bylo navraceno právo povinného výtisku pro celé území tehdejší Československé republiky. V šedesátých letech se zřizovatelem knihovny stal Krajský národní výbor v Ostravě a její název byl změněn na Státní vědeckou knihovnu. Po roce 1989 byla knihovna převedena pod Ministerstvo kultury (17).

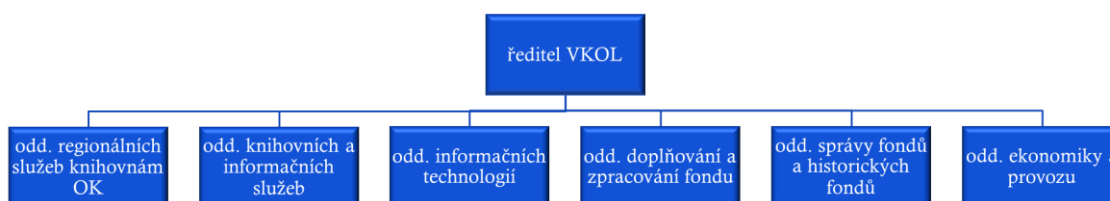
Od roku 2001 je knihovna pod oficiálním názvem Vědecká knihovna v Olomouci příspěvkovou organizací Olomouckého kraje. Jedná se o druhou nejstarší a třetí největší knihovnu v ČR (18). Knihovní fond čítá celkem přes 2 300 000 svazků. Na jeho velikosti

má značný podíl právo povinného výtisku, kdy je vydavatel povinen odevzdat knihovně jeden výtisk z každého vydání neperiodické publikace (19).

Knihovna disponuje několika budovami, které využívá ke své činnosti. Jde o hlavní budovu a budovu ředitelství a studoven na ulici Bezručova v centru Olomouce, sklad knih na ulici Holická, depozitář v městské části Hejčín a budova na ulici Ostružnická, která slouží převážně k pronájmu. Knihovna spravuje také Červený kostel, který dříve sloužil jako sklad knih.

### 2.1.1 Organizační struktura

V současné době je v knihovně zaměstnáno 95 stálých zaměstnanců. Knihovna je rozdělena na 6 oddělení a v jejím čele stojí ředitel jmenovaný zřizovatelem. Každé oddělení má v organizačním řádu knihovny přidělenou náplň práce. Organizační schéma znázorňuje obrázek níže.



Obrázek č. 14: Schéma organizační struktury  
(Zdroj: Vlastní zpracování)

### 2.1.2 Digitalizační pracoviště

Digitalizační pracoviště knihovny se nachází v budově depozitáře na ulici Tomkova v městské části Hejčín společně s knihařskou a restaurátorskou dílnou. Organizačně spadá pod oddělení informačních technologií.

Hlavní náplní práce je digitalizace historických dokumentů, převážně těch vytištěných na kyselém papíře, který postupem času degradují. Jedná se především o noviny, mapy a ostatní historické tisky.

Jelikož předmětem práce je zálohování dat z digitalizačního pracoviště, nikoliv celé knihovny, budou veškeré další části této analýzy orientovány pouze na digitalizační pracoviště.

## 2.2 Hardware

Tato podkapitola je věnována specifikaci hardwarového vybavení na digitalizačním pracovišti.

### 2.2.1 Skenery

Pracoviště v současné době disponuje třemi skenery. Prvním je velkoformátový skener **i2s SupraScan Quartz HD A0**, který se převážně používá ke skenování rozsáhlých svazků novin a velkoformátových dokumentů. Dále se zde nachází ruční skener **i2s Copibook HD600**, jenž slouží ke skenování knih do formátu A2. Posledním je automatický skener **4DigitalBooks DL-mini-i**. Jeho využití je poněkud omezené, jelikož není vhodným pro starší dokumenty. Všechny skenery produkují převážně obrázky ve formátu TIFF, které poté procházejí dalším zpracováním.

### 2.2.2 Pracovní stanice

Zaměstnanci pracoviště mají k dispozici pracovní stanice od společnosti Dell. Konkrétně se jedná o sestavy **Dell Optiplex 7010 MT** Jejich specifikace jsou následující:

- procesor: Intel Core i7-3770 3.40 GHz,
- operační paměť: 16 GB DDR3,
- grafická karta: AMD Radeon HD 7470 1 GB,
- pevné disky: 2 TB + 1 TB,
- optická mechanika: DVD+/-RW,
- ethernet: 1× gigabit port.

Celkem se zde nachází 6 stanic pro zaměstnance pracoviště a brigádníky, 4 z nich byly pořízeny v roce 2013 a zbylé 2 v roce 2016. Jedná se o jiný typ s podobnými parametry. Primárně jsou stanice využívány ke zpracování skenů, jejich doplňování o metadata a ostatní potřebné činnosti.

Dále jsou použity 3 stanice v serverovně. Jedna byla předělána na úložiště pro skenery, kam se ukládají naskenované obrázkové soubory ve formátu TIFF, disponující 15 TB v zapojení RAID 6. Z této kapacity je momentálně využito 12 TB, přírůstek za rok činí průměrně 5 TB. Další stanice jsou použity jako firewall a jako validátor vyprodukovaných datových balíčků.

### 2.2.3 Servery

V serverovně se nachází 2 fyzické servery. Konkrétně jde o typy **Dell PowerEdge R520 TPM** a **Dell PowerEdge R420 TPM**. Jejich specifikace jsou shrnuty v následující tabulce.

**Tabulka č. 1: Parametry serverů**  
(Zdroj: Vlastní zpracování dle interní dokumentace)

Parametr	PowerEdge R520 TPM	PowerEdge R420 TPM
Procesor	2 × Intel Xeon E5-2407 2.20 GHz	2 × Intel Xeon E5-2407 2.20 GHz
Operační paměť	2 × 8 GB RDIMM	2 × 8 GB RDIMM
Řadič	PERC H710	PERC H310
Pevné disky pro OS	2 × 600 GB 3,5" hot-plug SAS (600 GB RAID 1)	2 × 300 GB 3,5" hot-plug SAS (300 GB RAID 1)
Pevné disky pro úložiště	6 × 3 TB 3,5" hot-plug SAS (11 TB RAID 6)	---
Ethernet	Broadcom 5720 Dual Port 1 Gb	Broadcom 5720 Dual Port 1 Gb
Výška	2 U	1 U

Servery byly zakoupeny v roce 2013 Olomouckým krajem a převedeny do majetku knihovny v roce 2014. Úložiště na serveru PowerEdge R520 je v současnosti obsazeno 10 TB dat. Aplikace běžící na serverech jsou popsány v kapitole věnované softwaru.

### 2.2.4 Úložiště

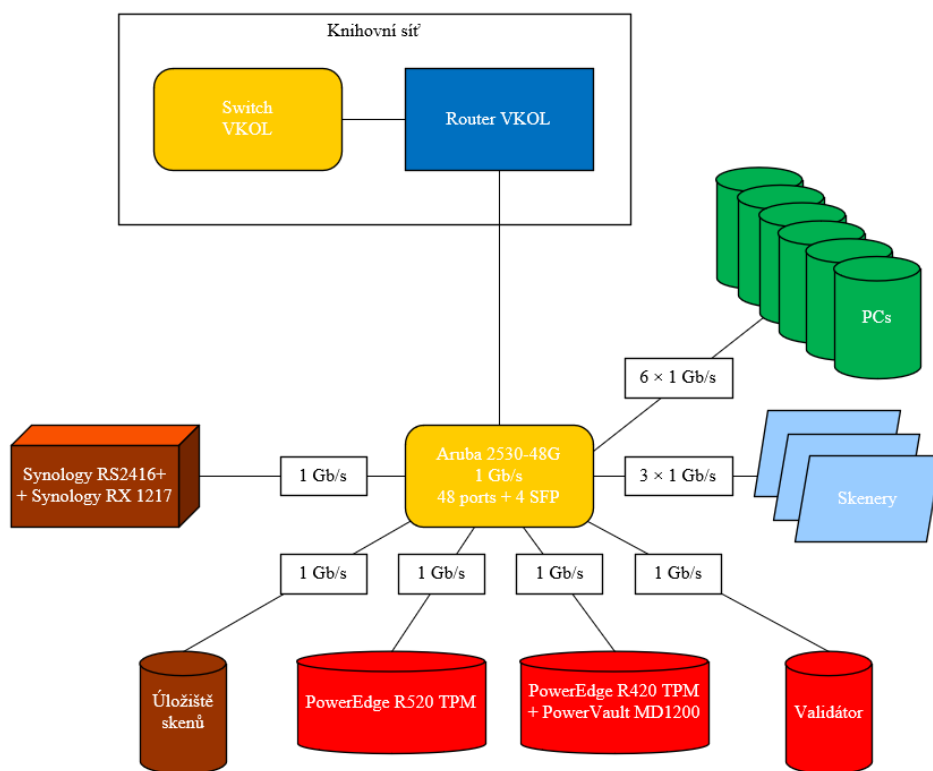
K dispozici je několik datových úložišť. Prvním je diskové pole **Dell PowerVault MD1200**, které je přímo spojené se serverem PowerEdge R420 TPM. Je osazeno deseti 3,5palcovými SAS disky o kapacitě 3 TB. Celkem tak poskytuje 25 TB při zapojení RAID 5. V současné době je z této kapacity využito 21 TB a roční přírůstek se pohybuje od 1 do 5 TB podle momentálního vytížení pracoviště.

Dalším místem pro data je síťové úložiště NAS **Synology RS2416+** a k němu připojená expanzivní jednotka **Synology RX1217**. Každý z těchto modelů v sobě obsahuje dvanáct 3,5palcových SATA disků o kapacitě 4 TB. Při zapojení RAID 6 je tedy k dispozici

2 × 40 TB. Toto úložiště slouží pro zálohy dat z pracoviště. Zaplněno je cca 90 % celkové kapacity s tím, že roční přírůstek dat se v závislosti na činnosti pracoviště pohybuje od 5 do 15 TB.

## 2.2.5 Síťová infrastruktura

V budově je zavedena kabeláž kategorie 5 s šířkou pásma 100 MHz umožňující přenosovou rychlost až 1 Gb/s. Propojení jednotlivých zařízení v síti zajišťuje switch **Aruba 2530-48G**, který disponuje 48 gigabitovými porty a 4 SFP porty. V budově je také zapojeno několik přístupových bodů pro Wi-Fi připojení. Zapojení všech prvků do sítě je znázorněno na obrázku.



**Obrázek č. 15: Zjednodušené schéma sítě**  
(Zdroj: Vlastní zpracování dle interní dokumentace)

## 2.2.6 Ostatní HW

Uložení všech zařízení v serverovně zabezpečuje stojanový skříňový datový rozvaděč se zástavnou výškou 42 U. Napájení zajišťují zdroje nepřerušovaného napětí (UPS) **APC Smart**.

## 2.3 Software

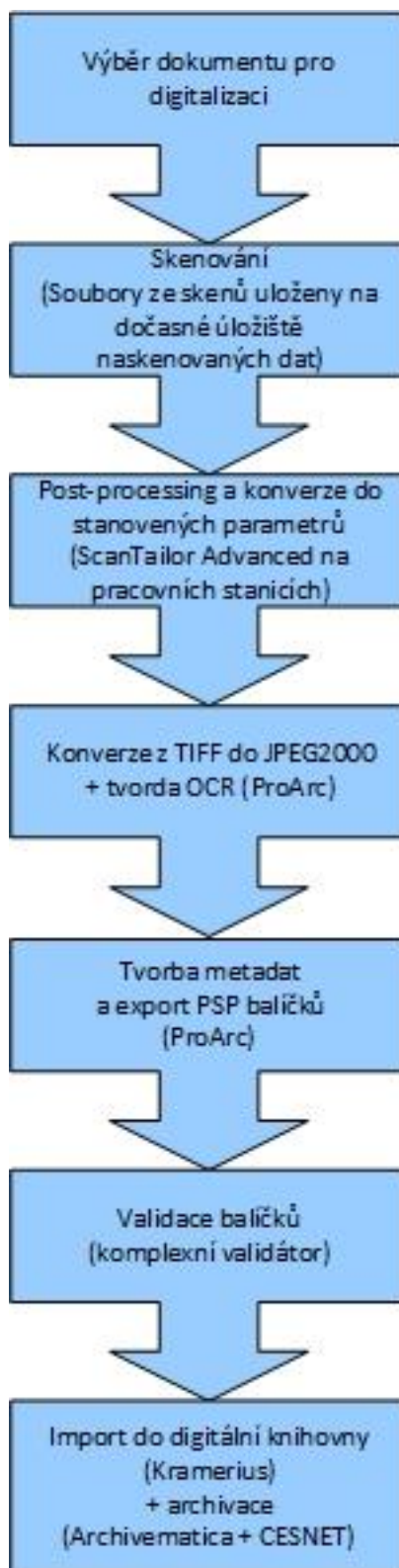
Tři pracovní stanice disponují operačním systémem **Windows 10**, zbytek stanic používá systém **Linux**. Operační systémy serverů jsou **CentOS** a **Debian**.

Na post-processing (rovnání, ořezávání) je používán program **ScanTailor Advanced** nainstalovaný na pracovních stanicích.

Pro tvorbu metadat a export PSP balíčků (producer submission package) slouží systém **ProArc**, jehož instance je zavedena na serveru PowerEdge R520 TPM. Hotové digitalizované dokumenty se zveřejňují pomocí digitální knihovny **Kramerius**, která běží na serveru PowerEdge R420 TPM s daty uloženými na přímo připojeném úložišti.

PSP balíčky vyexportované z ProArcu před zveřejněním v systému Kramerius procházejí kontrolou (validací) v **komplexním validátoru**. Jedná se o aplikaci Národní knihovny ČR, která kontroluje balíčky podle platných standardů Národní digitální knihovny (20).

Schéma celého procesu digitalizace je znázorněno následujícím obrázkem.



**Obrázek č. 16: Zjednodušené schéma procesu digitalizace**  
(Zdroj: Vlastní zpracování dle interní dokumentace)

K zálohování a archivaci dat je v rámci pracoviště použito několika nástrojů. **Clonezilla** je souhrnné rozhraní pro vytváření celkových záloh pomocí obrazů disků. Druhým nástrojem je zálohovací program **BorgBackup**. Oba výše zmíněné nástroje jsou open-source softwarem. Určitá data jsou navíc archivována v systému pro digitální archivaci **Archivematica**, který je zprovozněn na virtuálním serveru krajského úřadu Olomouckého kraje, a také v síti CESNET.

## 2.4 Proces zálohování

Digitální dokumenty, které úspěšně prošly celým procesem digitalizace, jsou zpřístupněny v systému Kramerius. Zároveň se z těchto dokumentů vytváří archivní balíčky, které jsou posílány na krajské úložiště aplikace Archivematica a také do sítě CESNET. Data uložená na diskovém poli serveru Kramerius tedy nejsou z důvodu velkého objemu (desítky TB) přímo zálohována. Při havárii serveru Kramerius by obnova dat byla značně komplikovaná a zabrala by časové období v řádu týdnů z důvodu složitého získávání dat z kraje a CESNETu.

Data ze systému ProArc jsou z části zálohována pomocí nástroje Clonezilla na úložiště Synology. Zároveň jsou některá data programem BorgBackup posílána do sítě CESNET. Data ze skenů, která vyžadují další zpracování v rámci procesu digitalizace, jsou uložena na dočasné úložiště a zálohují se prostřednictvím BorgBackup na CESNET.

Záloha dat z pracovních stanic je prováděna na úložiště Synology. Použitý nástroj je závislý na operačním systému konkrétního počítače. Na stanice s Windows 10 je použita Clonezilla, stanice s linuxovým řešením se zálohují přes BorgBackup.

Zálohy neprobíhají podle pevně daného plánu. Důvodem je nestálost potřeby konkrétních dat pro chod a práci pracoviště. Obvykle se intervaly záloh pohybují od 1 × týdně až po 1 × měsíčně. Objemy jednotlivých záloh jsou relativně velké. Pracovní stanice se průměrně pohybují kolem 500 GB, zálohy dočasného úložiště a ProArcu v jednotkách TB podle momentální vytíženosti.

## 2.5 Shrnutí

Celkově lze konstatovat, že zálohování dat na digitalizačním pracovišti neprobíhá v optimálních podmínkách. Chybí koncepce nejen pro intervaly zálohování, ale i pro určení, která data se vlastně mají zálohovat. Některá data by byla v případě havárie velmi těžko obnovitelná (např. data z Krameria).

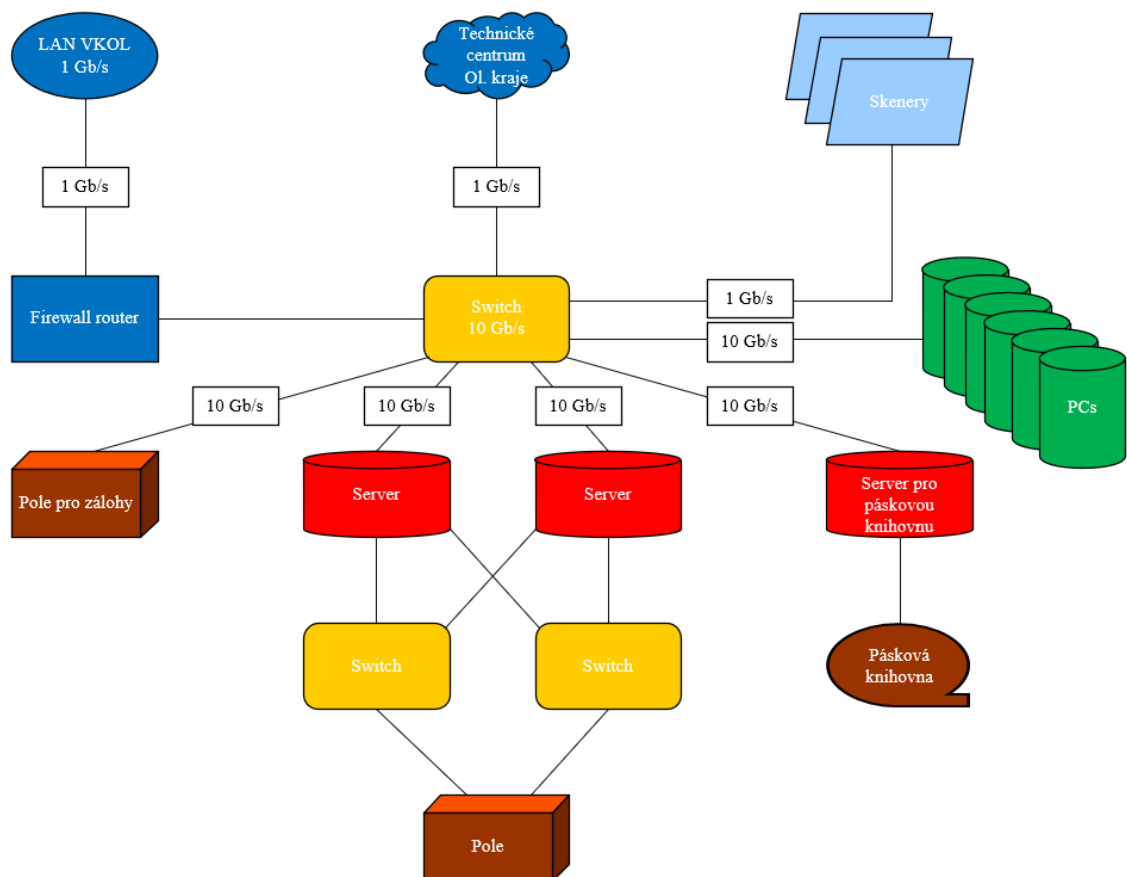
Dalším nedostatkem je stáří hardwarových komponentů. Kromě diskových polí Synology už všem zařízením vypršela podpora. V ideálním případě by mělo dojít k hardwarové obnově financované z dotačních programů pro projekty digitalizace. Také je dobré uvažovat o posílení síťové infrastruktury, jelikož stávající rychlosti nejsou při objemu dat produkovaných na pracovišti vhodné.

### 3 VLASTNÍ NÁVRHY ŘEŠENÍ

Tato kapitola bakalářské práce je zaměřena na popis jednotlivých možností řešení vedoucích k zefektivnění zálohovacího procesu digitalizačního pracoviště organizace. Tato navrhovaná řešení vycházejí z teoretických podkladů práce a také provedené analýzy současného stavu.

#### 3.1 Nový hardware

Původní plán počítal s výměnou serverů a přidáním nového diskového pole. Vzhledem k stáří původních hardwarových komponentů pracoviště bylo rozhodnuto, že budou veškerá stávající zařízení nahrazena novými. Tato obnova bude zahrnovat i síťovou infrastrukturu. Předpokládané zapojení znázorňuje obrázek níže.



Obrázek č. 17: Předběžné schéma plánovaného zapojení  
(Zdroj: Vlastní zpracování dle interní dokumentace)

### 3.1.1 Síťová infrastruktura

Prvním krokem modernizace pracoviště by mělo být posílení síťové infrastruktury. Především zapojení serverů a diskových polí by mělo být realizováno platformou umožňující přenosovou rychlost 10 Gb/s. Toho lze na úrovni horizontálního vedení docílit použitím metalických kabelů kategorie 6A s frekvenčním rozsahem 500 MHz. U tohoto typu kabelů je potřeba klást zvýšený důraz na jejich správné vedení, dodržovat minimální poloměry ohybu a odstup od silového vedení. Páteřní sekce bude realizována pomocí optických kabelů. Bude také potřeba pořídit nový switch s dostatečnou propustností. Tato práce se výběru konkrétních síťových prvků podrobněji nevěnuje.

### 3.1.2 Pracovní stanice

Nové pracovní stanice by měly splňovat dva požadavky. Prvním je operační paměť alespoň 16 GB, která je nutná pro každodenní práci se soubory skenů, druhým je potom SSD disk o kapacitě 500 GB a slot pro sekundární HDD. Zároveň je požadováno připojení pracovních stanic do sítě 10 Gb/s. Tomu musí odpovídat i síťové karty. V plánu je pořídit celkem 6 nových stanic.

Po průzkumu trhu byla vybrána pracovní stanice **HP Z2 Tower G5**, jejíž parametry splňují nastavené požadavky. Konkrétní specifikace jednotlivých komponentů stanice shrnuje následující tabulka.

**Tabulka č. 2: Specifikace pracovní stanice**  
(Zdroj: Vlastní zpracování dle stránek výrobce)

Parametr	HP Z2 Tower G5
Procesor	Intel Core i7 10700 (8 core, 2,9 GHz, 16 MB cache)
Operační paměť	32 GB (2 × 16 GB + 2 volné DIMM sloty) DDR4 3200 MHz
Grafická karta	NVIDIA Quadro P1000 4 GB
Disk	512 GB SSD + 4 TB SATA 3,5" HDD
Ethernet	Intel X550-T2 10GbE



**Obrázek č. 18: Pracovní stanice HP Z2 Tower G5**  
(Zdroj: 21)

Při této konkrétní konfiguraci by cena jedné stanice činila 2 300 amerických dolarů (dle stránek výrobce). Při přepočtu na české koruny kurzem ČNB ze dne 14. 4. 2021 (1 USD = 21,677 Kč) je výsledná cena 49 857 Kč.

### **3.1.3 Servery**

Požadavky na nové servery jsou následující:

- dva šestnáctijádrové procesory,
- 64 GB operační paměti,
- síťová karta pro 10 Gb/s.

Opět bude potřeba pořídit 2 servery, pro aplikace Kramerius a ProArc.

Byl vybrán server **HPE ProLiant DL380 Gen10 Server**. Specifikace konkrétního vybraného modelu jsou uvedeny v tabulce.

### Tabulka č. 3: Specifikace serveru

(Zdroj: Vlastní zpracování dle stránek výrobce)

Parametr	HPE ProLiant DL380 Gen10 Server
Procesor	2 × Intel Xeon Gold 5218 (16 core, 2,3 GHz, 22 MB cache)
Operační paměť	256 GB (8 × 32 GB) DDR4 2933 Hz
Řadič	HPE Smart Array P408i-a
Disky pro OS	2 × 400 GB SAS 12G SSD
Ethernet	HPE Ethernet 10Gb 2 port BASE-T BCM57416 Adapter
Výška	2 U



Obrázek č. 19: HPE ProLiant DL380 Gen10 Server

(Zdroj: 22)

Cena serveru této konfigurace by se pohybovala kolem 28 000 amerických dolarů (dle stránek výrobce), v závislosti na dalších volitelných prvcích. Při přepočtu na české koruny kurzem ČNB ze dne 14. 4. 2021 (1 USD = 21,677 Kč) je výsledná cena 606 956 Kč za jeden server.

#### 3.1.4 Disková pole

Objem dat digitalizačního pracoviště dosahuje již dnes hodnot v řádu desítek TB, proto je nutné vybrat takové úložiště, které pokryje všechny stávající nároky a zároveň bude disponovat dostatečně velkým prostorem i pro budoucí rozšíření. Produkce dat na pracovišti totiž bude i do budoucna pokračovat v rostoucím trendu.

V plánu je pořízení jednoho pole pro primární úložiště dat serverů a jednoho dedikovaného pro zálohy (viz obrázek číslo 17). Na základě současných kapacit a ročních přírůstků bylo spočítáno, že první pole by mělo disponovat kapacitou minimálně 70 TB a pole pro zálohy alespoň 130 TB (zaokrouhloeno na desítky TB nahoru). Takto

dimenzovaná úložiště by měla pokrýt nejen současné, ale i budoucí (v horizontu 5 let) potřeby pracoviště.

Z výše zmíněných důvodů bylo po průzkumu trhu vybráno diskové pole **QNAP TS-2477XU-RP-2700-16G**, které bude sloužit pro zálohy dat. Jeho vybrané specifikace jsou následující:

- 24 pozic pro 3,5" nebo 2,5" SATA III pevné disky nebo 2,5" SATA SSD,
- podpora RAID 0, RAID 1, RAID 5, RAID 6, RAID 10, RAID 50, RAID 60,
- osmijádrový procesor AMD Ryzen 7 2700 3,2 GHz,
- 16 GB DDR4 RAM,
- 2 × 10 Gb/s Ethernet SFP+ port a 2 × 1 Gb/s Ethernet RJ45 port,
- rack provedení, výška 4U,
- podpora hot swap.



**Obrázek č. 20: Diskové pole QNAP TS-2477XU-RP-2700-16G**  
(Zdroj: 23)

Jako pole pro běžné ukládání dat pracoviště bylo vybráno **QNAP TS-1277XU-RP-2600-8G** s následujícími specifikacemi:

- 12 pozic pro 3,5" nebo 2,5" SATA III pevné disky nebo 2,5" SATA SSD,
- podpora RAID 0, RAID 1, RAID 5, RAID 6, RAID 10, RAID 50, RAID 60,
- šestijádrový procesor AMD Ryzen 5 2600 3,4 GHz,
- 8 GB DDR4 RAM,
- 2 × 10 Gb/s Ethernet SFP+ port a 2 × 1 Gb/s Ethernet RJ45 port,
- rack provedení, výška 2U,
- podpora hot swap.



**Obrázek č. 21: Diskové pole QNAP TS-1277XU-RP-2600-8G**  
(Zdroj: 24)

Pole je nutné osadit disky. Pro tento účel byl vybrán konkrétní typ disku **Seagate Exos X10 (ST10000NM0086)**, který je kompatibilní s oběma poli. Tento disk disponuje kapacitou 10 TB, rozhraním SATA 6 Gb/s (SATA III), rychlostí 7200 otáček za minutu a střední dobou mezi poruchami 2,5 milionu hodin.



**Obrázek č. 22: Disk Seagate Exos X10**  
(Zdroj: 25)

Pole budou osazena disky takto: do QNAP TS-1277XU-RP-2600-8G bude vloženo celkem 10 disků a dva sloty budou ponechány volné pro případné budoucí rozšíření. Z 10 disků jich bude 9 zapojeno do RAID 6 a jeden ponechán jako hot spare. Celkem tak bude pole poskytovat 70 TB využitelného prostoru.

QNAP TS-2477XU-RP-2700-16G bude osazeno 20 disky, z nichž 18 bude zapojeno do RAID 60 a 2 budou ponechány jako hot spare. Kapacita pole tak bude činit 140 TB. RAID 60 je dvojúrovňové zapojení tvořené prokládáním několika podřazených logických polí typu RAID 6. Tento typ zapojení je odolný proti výpadku dvou disků v každém podřazeném poli a je velmi efektivní při použití většího počtu disků. Při tomto zapojení bude možné v budoucnu rozšířit pole až na 200 TB (za předpokladu, že budou opět použity 10TB disky a nebude vyhrazen hot spare).

Celkem tedy bude třeba pořídit 30 disků.

**Tabulka č. 4: Přehled cen polí a disků**  
(Zdroj: Vlastní zpracování, ceny dle Alza.cz)

<b>Položka</b>	<b>Cena bez DPH</b>	<b>Cena včetně DPH</b>
QNAP TS-2477XU-RP-2700-16G	131 810 Kč	159 490 Kč
QNAP TS-1277XU-RP-2600-8G	87 347 Kč	105 690 Kč
Seagate Exos X10	7 182 Kč	8 690 Kč

Podrobnější rozpis cen jednotlivých položek včetně jejich počtů je uveden v kapitole *3.3 Zhodnocení navrhovaného řešení.*

## **3.2 Zálohovací plán**

Na nových serverech budou opět běžet aplikace Kramerius a ProArc. Data, se kterými tyto servery budou pracovat, budou uložena na vyhrazeném poli. Zálohy těchto dat budou poté probíhat na druhé dostupné diskové pole určené pouze pro zálohy. Na to se budou ukládat i zálohy pracovních stanic.

Zároveň se stále budou z hotových digitalizovaných dokumentů zveřejňovaných v systému Kramerius vytvářet příslušné archivní balíčky, které budou zálohovány na krajské úložiště aplikace Archivematica a také do sítě CESNET. Nově přibude i zálohování pomocí páskové knihovny.

Nyní je potřeba stanovit, jak často se budou jednotlivé zálohy vytvářet, jaký typ zálohy použít a jak dlouho je uchovávat. Všechny tyto informace, včetně předpokládaného objemu jednotlivých záloh, jsou uvedeny v následující tabulce.

**Tabulka č. 5: Zálohovací plán**  
(Zdroj: Vlastní zpracování)

Odkud	Kam	Typ	Opakování / četnost	Předpokládaný objem	Délka uchování
<b>Primární pole</b>	Zálohovací pole	Úplná	na začátku provozu	cca 50 TB	stále
<b>Primární pole</b>	Zálohovací pole	Rozdílová	1 × týdně	Desítky až stovky GB	1 měsíc
<b>Primární pole</b>	Zálohovací pole	Součtová	1 × denně	Jednotky GB	1 týden
<b>Pracovní stanice</b>	Zálohovací pole	Úplná	1 × měsíčně	Stovky GB	2 měsíce
<b>Pracovní stanice</b>	Zálohovací pole	Rozdílová	1 × týdně	Jednotky GB	2 týdny
<b>Pracovní stanice</b>	Zálohovací pole	Součtová	1 × denně	Jednotky GB	1 týden
<b>Primární pole</b>	CESNET	Archivní balíčky Kramerius	dle potřeby	Jednotky GB	stále
<b>Primární pole</b>	Archivematica	Archivní balíčky Kramerius	dle potřeby	Jednotky GB	stále
<b>Primární pole</b>	Pásková knihovna	Archivní balíčky Kramerius	dle potřeby	Jednotky GB	stále

Diskové pole bude zálohováno nejdříve úplnou zálohou (obrazem), poté každý týden rozdílově a každý pracovní den součtově. Tímto způsobem bude možné v případě havárie pole obnovit data do podoby, jakou měla v den před havárií. Zároveň díky použití kombinace rozdílových a součtových záloh nebudou zálohy trvat příliš dlouho a budou zabírat méně místa než při použití úplných záloh.

Pracovní stanice budou zálohovány jednak úplně každý měsíc, rozdílově každý týden a součtově každý pracovní den. Tak bude zajištěna ochrana pracovních dat na nich uložených. V případě nižšího využití některých konkrétních stanic lze přistoupit pouze k úplným zálohám jednou za měsíc.

Frekvence tvoření archivních balíčků ze systému Kramerius záleží na aktuální situaci na pracovišti, proto je v plánu uvedeno, že zálohy těchto balíčků do příslušných systémů budou probíhat dle potřeby, nejlépe hned v den jejich vytvoření.

Tento zálohovací plán je vytvořen s cílem vyvážit objem uložených záloh a délku jejich případné obnovy. Na rozdíl od současného stavu odpovídá jasně na otázky co, kam a jak často zálohovat i s přihlédnutím na doby uchování těchto záloh.

Zároveň s plánem záloh by měl být sestaven i **plán obnovy**, který do detailu popíše, jak v případě havárie obnovit ztracená data. Zahrnovat by měl všechny zdroje dat (v tomto případě diskové pole a pracovní stanice). Také by měl být sepsán na takové úrovni, aby obnovu podle něj mohl provést libovolný zaměstnanec bez ohledu na jeho znalost této problematiky. Sestavení plánu obnovy v takovém detailu je závislé na spoustě konkrétních faktorů, ať už z hlediska HW, SW, sítě apod., proto se jím tato práce více do hloubky zabývat nebude. Pouze doporučuje jeho sestavení jako smysluplný a žádoucí krok, který do značné míry eliminuje riziko závislosti obnovy na zkušenostech a znalostech konkrétních lidí v organizaci.

Dalším navrženým opatřením je realizování **zkušební obnovy** zálohovaných dat. Účelem tohoto procesu je ověření toho, zda se dají data bez problémů obnovit. Někdy totiž může dojít k porušení zálohovaných dat, která se potom z části nebo vůbec nedají obnovit. Je lepší na tuto skutečnost přijít v rámci zkoušky než při skutečné havárii. Zkušební obnovu lze provést například na nějaké staré nepotřebné zařízení.

### **3.3 Zhodnocení navrhovaného řešení**

Následující tabulka shrnuje všechny položky nově navrženého hardwaru, jejich počet a ceny. Jelikož se tato práce nevěnovala výběru konkrétních položek síťové infrastruktury a skenerů, je tato částka pouze dílčí, výsledná částka zahrnující veškeré součásti nově nakupovaného systému bude mnohem vyšší. Je nutné také počítat s cenou za práci.

**Tabulka č. 6: Přehled cen veškerého HW**

(Zdroj: Vlastní zpracování, ceny dle Alza.cz a stránek výrobců)

<b>Položka</b>	<b>Počet ks</b>	<b>Cena za ks bez DPH</b>	<b>Cena za ks s DPH</b>	<b>Cena celkem bez DPH</b>	<b>Cena celkem s DPH</b>
HP Z2 Tower G5	6	49 857 Kč	60 327 Kč	299 142 Kč	361 962 Kč
HPE ProLiant DL380 Gen10 Server	2	606 956 Kč	734 417 Kč	1 213 912 Kč	1 468 834 Kč
QNAP TS-2477XU-RP-2700-16G	1	131 810 Kč	159 490 Kč	131 810 Kč	159 490 Kč
QNAP TS-1277XU-RP-2600-8G	1	87 347 Kč	105 690 Kč	87 347 Kč	105 690 Kč
Seagate Exos X10	30	7 182 Kč	8 690 Kč	215 460 Kč	260 706 Kč
<b>Celkem</b>	-----			<b>1 947 671 Kč</b>	<b>2 356 682 Kč</b>

Financování takto náročné investice nelze realizovat pouze z rozpočtu knihovny, ale je třeba použít i jiné zdroje. Jednou z možností je například využití dotací z Evropské unie určených pro projekty spojené s rozvojem digitalizace.

Hodnocení nově navrženého zálohovacího plánu není zcela jednoznačné jako v případě návrhu hardwaru. Jeho zavedení, průběh a kontrola si vyžádá určité náklady, především v podobě mezd příslušných pracovníků. Jeho přínos pro organizaci je ovšem nesporný, neboť minimalizuje riziko ztráty klíčových dat, jejichž obnovení by při stávajícím stavu bylo sice proveditelné, ale nesmírně časově a pracovníčně náročné. Obdobně lze hodnotit i plán obnovy a realizaci zkušebních obnov.

## ZÁVĚR

Tato bakalářská práce se věnovala zálohování dat a datovým úložištím. Hlavním cílem práce bylo analyzovat systém zálohování dat v prostředí digitalizačního pracoviště Vědecké knihovny v Olomouci. Na základě této analýzy a požadavků organizace byly poté zpracovány vlastní návrhy řešení upravující stávající systém. Práce je rozdělena na 3 hlavní části: teoretická východiska, analýza současného stavu a vlastní návrhy řešení.

V první části práce věnované teoretickým východiskům jsou vymezeny základní pojmy vztahující se k dané problematice. Popsány jsou zde pojmy data, zálohování a archivace, dále jednotlivé typy záloh, konkrétní používaná úložná média, topologie síťových úložišť, systém RAID a související pojmy. Zdrojem pro tuto část byla odborná literatura a relevantní internetové zdroje.

Druhá část práce je zaměřena na detailní analýzu současného stavu prostředí digitalizačního pracoviště VKOL. Nejprve je stručně popsána historie a současnost knihovny včetně její organizační struktury a začlenění digitalizačního pracoviště. Poté je popsán používaný hardware zahrnující skenery, servery, pracovní stanice a disková pole. Dále je pozornost věnována softwaru, síťové infrastruktuře, samotnému procesu digitalizace a také procesu zálohování. Na závěr je současný stav stručně zhodnocen.

Poslední kapitola práce zahrnuje konkrétní návrhy řešení, které vychází jak z provedené analýzy, tak z konkrétních požadavků organizace. Navržen je nový hardware pro pracoviště a nový plán zálohování dat. Hardware zahrnuje jak disková pole a disky, tak i pracovní stanice a servery. Plán záloh zahrnuje konkrétní zdroje a cíle dat, typy záloh, jejich přepokládaný objem a délku uchování. Dále je doporučeno sestavení plánu obnovy a pravidelné provádění zkušební obnovy dat. V závěru kapitoly je provedeno zhodnocení návrhu.

Hlavní cíl této práce byl splněn. Návrhy předložené v této práci, zpracované na základě provedené analýzy a požadavků organizace, mohou být použity pro úpravy stávajícího procesu zálohování a výběr nového hardwaru pro potřeby pracoviště.

## SEZNAM POUŽITÝCH ZDROJŮ

- (1) POŽÁR, Josef. *Manažerská informatika*. Plzeň: Aleš Čeněk, 2010. ISBN 978-80-7380-276-9.
- (2) PECINOVSKÝ, Josef. *Archivace a komprimace dat*. Praha: Grada Publishing a.s., 2003. ISBN 80-247-0659-8.
- (3) ONDRÁK, Viktor, Petr SEDLÁK a Vladimír MAZÁLEK. *Problematika ISMS v manažerské informatice*. Brno: Akademické nakladatelství CERM, 2013. ISBN 978-80-7204-872-4.
- (4) Zálohování. *Jak na Internet* [online]. Praha: CZ.NIC, © 2021 [cit. 2021-03-02]. Dostupné z: <https://www.jaknainternet.cz/page/1180/zalohovani/>
- (5) Proč se stále k zálohování používají pásky? *VAHAL - komplexní dodávky hardware a software* [online]. Prostějov: VAHAL, © 2021 [cit. 2021-03-03]. Dostupné z: <https://www.vahal.cz/o-firme/clanky/proc-se-stale-k-zalohovani-pouzivaji-pasky.html>
- (6) Why the Future of Data Storage is (Still) Magnetic Tape. *IEEE Spectrum* [online]. New York: IEEE, © 2021 [cit. 2021-03-03]. Dostupné z: <https://spectrum.ieee.org/computing/hardware/why-the-future-of-data-storage-is-still-magnetic-tape>
- (7) DEMBOWSKI, Klaus. *Mistrovství v hardware*. Brno: Computer Press, 2009. ISBN 978-80-251-2310-2.
- (8) SSD vs. HDD. *Western Digital Obchod* [online]. San Jose: Western Digital Corporation, © 2021 [cit. 2021-03-18]. Dostupné z: <https://shop.westerndigital.com/cs-cz/solutions/ssd-vs-hdd>
- (9) SSD vs. HDD: What's the Difference? *PCMag* [online]. New York: Ziff Davis, © 1996-2021 [cit. 2021-03-05]. Dostupné z: <https://www.pcmag.com/news/ssd-vs-hdd-whats-the-difference>
- (10) LACKO, Ľuboslav. *Osobní cloud pro domácí podnikání a malé firmy*. Brno: Computer Press, 2012. ISBN 978-80-251-3744-4.

- (11) Co je diskové pole. *Správa sítě - slovník pojmů* [online]. Praha: Aira GROUP, © 2016 [cit. 2021-03-18]. Dostupné z: <https://www.sprava-site.eu/diskove-pole/>
- (12) Úložné systémy - NAS vs. SAN. *Datacentrum WEDOS* [online]. Hluboká nad Vltavou: WEDOS Internet, © 2021 [cit. 2021-03-18]. Dostupné z: <https://datacentrum.wedos.com/a/78/ulozne-systemy-nas-vs-san.html>
- (13) Storage-Grundlagen: DAS, SAN und NAS im Überblick. *Thomas-Krenn.AG* [online]. Freyung: Thomas-Krenn.AG, © 2002-2021 [cit. 2021-03-19]. Dostupné z: <https://www.thomas-krenn.com/de/tkmag/allgemein/storage-grundlagen-das-san-und-nas-im-ueberblick/>
- (14) Co to je RAID a k čemu slouží? *GIGA PC* [online]. Praha: GIGA PC, © 2021 [cit. 2021-03-19]. Dostupné z: <https://www.giga-pc.cz/technicke-okenko/raid/>
- (15) Hot Plugging – Hot Swapping. *Webopedia: Tech Dictionary for Students, Educators & IT Professionals* [online]. Nashville: TechnologyAdvice, © 2021 [cit. 2021-03-22]. Dostupné z: <https://www.webopedia.com/definitions/hot-plugging/>
- (16) Hot Spare. *Techopedia* [online]. Edmonton: Techopedia, © 2021 [cit. 2021-03-22]. Dostupné z: <https://www.techopedia.com/definition/3655/hot-spare>
- (17) Z historie knihovny. *Vědecká knihovna v Olomouci* [online]. Olomouc: Vědecká knihovna v Olomouci, ©2008-2021 [cit. 2021-02-26]. Dostupné z: <https://www.vkol.cz/z-historie-knihovny-1>
- (18) O Knihovně. *Vědecká knihovna v Olomouci* [online]. Olomouc: Vědecká knihovna v Olomouci, ©2008-2021 [cit. 2021-01-21]. Dostupné z: <https://www.vkol.cz/o-knihovne>
- (19) Zákon č. 37/1995 Sb., o neperiodických publikacích ze dne 8. února 1995.
- (20) Komplexní validátor. *Novinky — Národní digitální knihovna* [online]. Praha: Národní knihovna ČR, 2020 [cit. 2021-02-26]. Dostupné z: <https://old.ndk.cz/archivace/komplexni-validator>

- (21) HP Z2 Tower G5 Workstation. *HP® Official Site* [online]. Palo Alto: HP Development Company, © 2021 [cit. 2021-05-04]. Dostupné z: <https://h20195.www2.hp.com/v2/GetPDF.aspx/c06728373.pdf>
- (22) HPE ProLiant DL380 Gen10 server. *Hewlett Packard Enterprise (HPE)* [online]. Houston: Hewlett Packard Enterprise Development, © 2021 [cit. 2021-05-04]. Dostupné z: <https://buy.hpe.com/us/en/servers/proliant-dl-servers/proliant-dl300-servers/proliant-dl380-server/hpe-proliant-dl380-gen10-server/p/1010026818>
- (23) TS-2477XU-RP. *QNAP* [online]. New Taipei City: QNAP Systems, © 2021 [cit. 2021-05-04]. Dostupné z: <https://www.qnap.com/cs-cz/product/ts-2477xu-rp>
- (24) TS-1277XU-RP. *QNAP* [online]. New Taipei City: QNAP Systems, © 2021 [cit. 2021-05-04]. Dostupné z: <https://www.qnap.com/cs-cz/product/ts-1277xu-rp>
- (25) Exos X10. *Seagate US* [online]. Fremont: Seagate Technology, © 2021 [cit. 2021-05-04]. Dostupné z: [https://www.seagate.com/files/www-content/datasheets/pdfs/exos-x-10DS1948-1-1709US-en\\_US.pdf](https://www.seagate.com/files/www-content/datasheets/pdfs/exos-x-10DS1948-1-1709US-en_US.pdf)

## SEZNAM POUŽITÝCH OBRÁZKŮ

Obrázek č. 1: Zjednodušené schéma typů záloh.....	15
Obrázek č. 2: Magnetická páska typu LTO s kazetou .....	16
Obrázek č. 3: Optické disky-DVD.....	17
Obrázek č. 4: USB flash disky.....	18
Obrázek č. 5: Pevný disk .....	19
Obrázek č. 6: SSD.....	20
Obrázek č. 7: Schéma DAS, SAN a NAS.....	22
Obrázek č. 8: Schéma RAID 0, metoda prokládání.....	23
Obrázek č. 9: Schéma RAID 1.....	24
Obrázek č. 10: Schéma RAID 3.....	25
Obrázek č. 11: Schéma RAID 4.....	26
Obrázek č. 12: Schéma RAID 5.....	27
Obrázek č. 13: Schéma RAID 6.....	28
Obrázek č. 14: Schéma organizační struktury .....	31
Obrázek č. 15: Zjednodušené schéma sítě .....	34
Obrázek č. 16: Zjednodušené schéma procesu digitalizace.....	36
Obrázek č. 17: Předběžné schéma plánovaného zapojení .....	39
Obrázek č. 18: Pracovní stanice HP Z2 Tower G5.....	41
Obrázek č. 19: HPE ProLiant DL380 Gen10 Server .....	42
Obrázek č. 20: Diskové pole QNAP TS-2477XU-RP-2700-16G .....	43
Obrázek č. 21: Diskové pole QNAP TS-1277XU-RP-2600-8G .....	44
Obrázek č. 22: Disk Seagate Exos X10 .....	44

## SEZNAM POUŽITÝCH TABULEK

Tabulka č. 1: Parametry serverů .....	33
Tabulka č. 2: Specifikace pracovní stanice.....	40
Tabulka č. 3: Specifikace serveru .....	42
Tabulka č. 4: Přehled cen polí a disků .....	45
Tabulka č. 5: Zálohovací plán.....	46
Tabulka č. 6: Přehled cen veškerého HW .....	48