

**Česká zemědělská univerzita v Praze**

**Provozně ekonomická fakulta**

**Katedra informačních technologií**



## **Bakalářská práce**

**Možnosti zálohování a archivace dat s cílem návržení  
strategie u firmy Rohlik Group**

**Lukáš Mevald**

© 2024 ČZU v Praze



# ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Provozně ekonomická fakulta

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Lukáš Mevald

Informatika

Název práce

**Možnosti zálohování a archivace dat s cílem navržení strategie u firmy Rohlik Group**

Název anglicky

**Options for data backup and archiving with the goal of designing a strategy at Rohlik Group**

---

### Cíle práce

Cílem práce je navrhnout několik variant řešení zálohování a archivace dat na základě parametrů bezpečnosti, ceny a efektivity.

Cílem teoretické části práce je rozbor a zvážení možností zálohování dat, jak fyzicky pomocí HDD a SSD, tak online pomocí NAS a cloud s ohledy na prostředky, jakými jsou cena, místo, objem dat, rychlost a bezpečnost. V praktické části budou analyzovány archivační a zálohovací strategie firmy Rohlik Group. Jejich strategie bude porovnána s daty získanými z teoretické části.

### Metodika

- Provedení podrobné rešerše strategií archivace a zálohování dat jak fyzické, tak online.
- Shrnutí poznatků a určení výhod a nevýhod určitých strategií.
- Zkoumání a analýza použitých strategií u firmy Rohlik Group.
- Porovnání obou strategií.
- Navržení nové strategie na základě rešerše a porovnání strategií.
- V závěru bude vyhodnocena optimální strategie.

**Doporučený rozsah práce**

35-45s.

**Klíčová slova**

Data, Bezpečnost, Cena, Cloud, NAS , Mechanické úložiště, Flash úložiště , Strategie

---

**Doporučené zdroje informací**

LEIXNER, Miroslav. *PC – zálohování a archivace dat*. Praha: Grada, 1993. ISBN 80-85424-73-8.

MURUGESAN, San; BOJANOVA, Irena; IEEE COMPUTER SOCIETY. *Encyclopedia of cloud computing*. Chichester: IEEE Press, 2016. ISBN 978-1-118-82197-8.

NÁDENÍČEK, David. *Zálohování podnikových dat*. Zlín: Univerzita Tomáše Bati ve Zlíně, 2022, 52 s.

Dostupné také z: <http://hdl.handle.net/10563/51177>. Univerzita Tomáše Bati ve Zlíně. Fakulta aplikované informatiky, Ústav počítačových a komunikačních systémů. Vedoucí práce Králík, Lukáš.

NELSON, S. *Pro Data Backup and Recovery*. Apress, 2011. ISBN 9781430226628

PETR, Jiříček. *Záloha dat na optické disky-implemtace technologie RAID*. 2023. Bachelor's Thesis. České vysoké učení technické v Praze. Vypočetní a informační centrum.

---

**Předběžný termín obhajoby**

2023/24 LS – PEF

**Vedoucí práce**

Ing. Martin Havránek, Ph.D.

**Garantující pracoviště**

Katedra informačních technologií

---

Elektronicky schváleno dne 4. 9. 2023

**doc. Ing. Jiří Vaněk, Ph.D.**

Vedoucí katedry

---

Elektronicky schváleno dne 3. 11. 2023

**doc. Ing. Tomáš Šubrt, Ph.D.**

Děkan

V Praze dne 27. 02. 2024

### **Čestné prohlášení**

Prohlašuji, že svou bakalářskou práci "Možnosti zálohování a archivace dat s cílem navržení strategie u firmy Rohlik Group" jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autor uvedené bakalářské práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 15.3.2024

---

## **Poděkování**

Rád bych touto cestou poděkoval Ing. Martinu Havránkovi, Ph.D. za odbornou pomoc a poskytnutí cenných rad při vedení této práce.

# **Možnosti zálohování a archivace dat s cílem navržení strategie u firmy Rohlik Group**

## Abstrakt

Tato bakalářská práce se zaměřuje na možnosti, kterými se zálohují a archivují data ve firmě. Hlavní problematikou teoretické části je seznámení se s tematikou zálohování a archivace dat ve firemním prostředí s důrazem na technologie záloh.

Hlavní problematikou praktické části je analýza zálohování a archivace dat u firmy Rohlik Group s následným cílem navržení optimalizace pro jejich strategii na základě části teoretické.

**Klíčová slova:** Data, Bezpečnost, Cena, Cloud, NAS, Mechanické úložiště, Flash úložiště, Strategie

# **Options for data backup and archiving with the goal of designing a strategy at Rohlik Group**

## **Abstract**

This bachelor's thesis focuses on the options used to back up and archive data in a company. The main issue of the theoretical part is to get acquainted with the topic of data backup and archiving in a corporate environment with emphasis on backup technologies. The main issue of the practical part is the analysis of data backup and archiving in Rohlik Group company with the subsequent aim of proposing an optimization for their strategy based on the theoretical part.

**Keywords:** Data, Security, Price, Cloud, NAS , Mechanical Storage, Flash Storage , Strategy



# Obsah

<b>1 Úvod .....</b>	<b>11</b>
<b>2 Cíl práce a metodika.....</b>	<b>12</b>
2.1 Cíl práce.....	12
2.2 Metodika.....	12
<b>3 Teoretická východiska .....</b>	<b>13</b>
3.1 Data .....	13
3.2 Zálohování dat .....	13
3.3 Archivace dat .....	13
3.4 Důvod zálohování .....	14
3.5 Důvody ztráty dat.....	15
3.6 Datová úložiště.....	16
3.6.1 Magnetické pásky .....	17
3.6.2 Optické disky .....	17
3.6.3 Pevná a SSD disková úložiště.....	18
3.6.4 NAS úložiště .....	21
3.6.5 Cloud .....	22
3.7 Disková pole RAID.....	22
3.7.1 RAID 0 .....	23
3.7.2 RAID 1 .....	24
3.7.3 RAID 5 .....	24
3.7.4 RAID 6 .....	25
3.7.5 RAID 10 .....	25
3.8 Způsob zálohování .....	25
3.8.1 Zálohování plné .....	26
3.8.2 Zálohování postupné (inkrementální) .....	26
3.8.3 Zálohování rozdílové (diferenciální).....	27
3.9 Rotační zálohy .....	27
3.9.1 Round Robin .....	27
3.9.2 Grandfather-father-son .....	28
3.9.3 Tower of Hanoi .....	28
3.10 Vytváření záloh.....	29
3.10.1 Manuální kopírování .....	29
3.10.2 Záloha pomocí specializovaných zálohovacích programů.....	29
3.11 Software pro zálohování.....	29
3.11.1 7-zip.....	30

3.11.2	BackupPC.....	30
3.11.3	Bacula.....	30
3.11.4	Duplicati.....	31
<b>4</b>	<b>Vlastní práce.....</b>	<b>32</b>
4.1	Firma Rohlik Group.....	32
4.1.1	Vlivy na ztrátu dat ve firmě Rohlik Group .....	32
4.2	Analýza současného zálohovacího řešení firmy Rohlik Group .....	33
4.2.1	Použitá strategie.....	33
4.2.2	Nástroje zálohování a vysvětlení pojmů .....	34
4.2.3	Zálohovací média.....	36
4.2.4	Souhrn současné strategie .....	37
4.3	Návrh optimalizace řešení zálohování .....	38
<b>5</b>	<b>Zhodnocení výsledků analýzy .....</b>	<b>43</b>
<b>6</b>	<b>Závěr .....</b>	<b>44</b>
<b>7</b>	<b>Seznam použitých zdrojů .....</b>	<b>46</b>
<b>8</b>	<b>Seznam obrázků a tabulek .....</b>	<b>50</b>
8.1	Seznam obrázků.....	50
8.2	Seznam tabulek.....	50

# 1 Úvod

V dnešní digitální éře, se jednotlivci, firmy a podniky stále více spoléhají na svá elektronická data ať už pro svůj osobní život či běh firmy. Ne vždy však tyto osoby či firmy myslí na to, jak tato data zabezpečit oproti jejich ztrátě. V dnešní době ztráta dat může znamenat skoro až nevyčíslitelnou hodnotu ať už monetární či emocionální. Možností, jak o tato data v dnešní době přijít, je nespočet. Ať už by se jednalo o lidskou chybu či nějaký útok. Řešením nebo aspoň pokusem o prevenci ztráty dat je zálohování. Avšak s ohledem na různé možnosti firem či jednotlivců je nutno zvážit správnou strategii zálohování, která nejvíce vyhovuje právě dané firmě či osobě. Nutné je zvážit možnosti finanční, hardwarové a softwarové pro zvolení nejlepší strategie kompatibilní se všemi požadavky. Může se také stát, že pouhá záloha dat nebude stačit. Dále je nutné dbát na nařízení ze strany státu, který může vyžadovat dostupnost dat i po dobu několika let. Pro tato data je zapotřebí využití principu archivace.

Náplní této bakalářské práce je právě řešení strategií a způsobů, jak data zálohovat a archivovat. Nejdříve se práce zabývá tím, co to jsou vůbec data, poté jde více do důvodů, proč data zálohovat a jaké jsou důvody jejich ztráty. Poté už se práce věnuje konkrétním hardwarovým či cloudovým zařízením, na která se dají data zálohovat, kde pak práce navazuje na technologie umožňující kvalitní a spolehlivou zálohu či archivaci dat z pohledu softwarového řešení. Ve finální části se práce zaměřuje na analýzu současné zálohovací strategie firmy Rohlik Group a její následné optimalizaci.

## **2 Cíl práce a metodika**

### **2.1 Cíl práce**

Cílem práce je navrhnout optimální řešení zálohování a archivace dat ve firmě Rohlik Group na základě parametrů bezpečnosti, ceny a efektivity. Cílem teoretické části je rozbor strategií zálohování dat, jak fyzických, tak online s ohledy na parametry, kterými jsou cena, místo, objem dat, rychlost a zabezpečení. V praktické části bude analýza zálohovací a archivační strategie firmy Rohlik Group, kde poté bude porovnána s daty z teoretické části a bude navržena optimalizace strategie.

### **2.2 Metodika**

V teoretické části bude provedena podrobná rešerše strategií zálohování a archivace dat jak fyzicky, tak online ze zdrojů jak knižních, tak webových. Poznatky z různých strategií budou shrnuty a určeny jak výhody a dále nevýhody. V praktické části bude provedena analýza zálohovacích strategií firmy Rohlik Group, kde následně budou výsledky porovnány s rešerší. Na základě porovnání bude vyhodnocena a navržena optimalizace stávající strategie.

## **3 Teoretická východiska**

### **3.1 Data**

Data jsou v informatice definována jako reprezentace faktů, konceptů nebo instrukcí ve formě, která se dá použít v komunikaci, interpretaci nebo zpracování pomocí člověka či počítače.

Jsou reprezentována pomocí charakterů jako je například abeceda, číslice anebo speciální znaky. (1)

### **3.2 Zálohování dat**

K procesu zálohování dat lze přistupovat jako ke způsobu, který umožňuje vytvoření kopií informací na záložním datovém úložišti, a to mimo hlavní datové úložiště, kde jsou informace obvykle uloženy. Tímto krokem se získává kopie, která slouží jako ochrana pro případy, kdy by mohlo dojít k poškození nebo ztrátě původních dat. Díky této záloze je možné původní informace rychle a se skoro nulovými ztrátami obnovit. Proces zálohování zahrnuje zálohování nejnovějších informací, které jsou během další zálohy aktualizovány nebo nahrazeny, aby byla současná záloha vždy nejnovější. (2)

Aby bylo riziko ztráty dat co nejmenší, je nutné vybrat správnou strategii zálohování a pravidelně ji využívat a data zálohovat. Pro firmy je také důležité vybrat správné časové období pro provedení těchto záloh, aby nenarušily její běžný chod.

### **3.3 Archivace dat**

Archivování dat je proces přesouvání dat, která už nejsou aktivní na oddělené pevné úložiště pro dlouhodobé zachování. Archivní data jsou starší data, která pořád zůstávají důležitá pro firmu či organizaci a musí být zachována pro budoucí reference anebo dokumenty potřebné uchovat dle zákona. (3) Archivní data jsou zapotřebí uchovávat na bezpečném místě a sama archivovaná data nejsou zapotřebí aktualizovat na rozdíl od dat záložních. Při archivaci velkého množství dat je stále zapotřebí určitá úroveň údržby, protože

i nejvíce spolehlivé uložení není imunní oproti času a jím způsobené ztrátě dat ve tvaru bitové hniloby. (4)

Archivace dat by se neměla zaměřovat se zálohováním dat. Oboje sice uchovávají data, ale jejich účely jsou odlišné. Datové archívy fungují jako repositář pro data, ke kterým vzácně přistupujeme, ale pořád jsou dostupná. Zálohy jsou součástí systému pro obnovování dat v případě korupce či zničení. Záložní data se většinou skládají z důležitých informací, které musí být rychle obnoveny při ztrátě či smazání. (3)

### 3.4 Důvod zálohování

Jedním z hlavních důvodů, proč zálohovat data jak pro jedince či firmu je to, že bez určitých dat se pracovat nedá. Váš počítač může mít nejmodernější hardwarové vybavení, ale to se na rozdíl od dat vždy dá koupit, zároveň můžete mít koupený nejlepší software, ale ten se skoro vždy dá za pomoci klíčů a aktivací stáhnout znovu. S daty je problém u jednotlivce ten, že mohou mít jak paměťovou, tak sentimentální hodnotu, a proto je jejich cena skoro až nevyčísitelná. To samé platí pro firmu, která může vše nahradit až na jejich ztracená data, která mohou obsahovat vše od know-how po uživatelská data.

V současnosti se zvyšují právní požadavky pro oblast zálohování dat. Je nutno tedy pečlivě zálohovat a archivovat data týkající se finančních či úředních dokumentů. Pokud dojde ke ztrátě těchto pro firmu a stát důležitých dat, mohou na firmy dopadnout značné finanční sankce a pokuty nebo dokonce i tresty. (5)

Díky kvalitnímu zálohování dat se můžeme vyvarovat následujícím problémům:

#### **Ztrátě dat**

Firmy dnes sbírají a produkují velké množství dat, včetně dat zákazníků, dat zaměstnanců a systémová data. Ztráta těchto dat může mít potencionálně ničivé dopady na firmu. (6)

#### **Ochrana proti kyberútokům**

Kyberútoky se pořád vyvíjejí a žádná firma není ochráněna proti jejich rizikům. (7) Implementováním datových záloh se může minimalizovat odstávka systému a ochránit data oproti úniku. (8)

### **Zachování věrnosti zákazníků**

Ztráty dat vedou ke ztrátě důvěry zákazníků. Být firmou, která ztrácí data, je velmi škodlivé pro její pověst na trhu, obzvláště pokud se jedná o data zákazníků. (7) Kvalitní záloha dat pomůže jak vám, tak vaší firmě, zaměstnancům, zákazníkům, také vám pomůže s vaší důvěryhodností a reputací. (9)

### **Dodržování předpisů**

V závislosti na vašem odvětví může být ze zákona požadováno, abyste zavedli zálohování a obnovování dat, abyste vyhověli předpisům a ochránili citlivé informace. (8)

### **Příprava na nepředvídatelné situace**

Ke ztrátě dat může dojít nenadále a bez jakéhokoliv varování. Ztráta dat může vzniknout kvůli lidem či strojům, a proto je důležité být připraven s funkční zálohou dat, aby se firma mohla rychle vzpamatovat.

## **3.5 Důvody ztráty dat**

Vznik ztráty dat může nastat několika způsoby:

### **Lidská chyba**

Lidskou chybou může být smazání omylem či přepsání dat nebo upuštění počítače či disku. Také je zde možnost, že člověk poškodí data záměrně. (10)

### **Chyba hardwaru**

Asi nejčastějším důvodem ztráty dat je chyba či poškození hardwaru z důvodu teplot, špatného skladování, špatná ventilace a použití mimo určené parametry. (11) Selhání disku stářím či při špatném zacházení může také poškodit data.

### **Poškození softwaru**

Žádný software není skvělý, také se u něho mohou vyskytovat chyby, které mohou poškodit vaše data anebo mohou být napadena hackery. Vývojáři softwarů mají snahu tyto chyby a slabiny opravovat, ale ztráty dat se pořád stávají. (11)

### **Kybernetická nelegální činnost**

V případě počítačových virů či jiných škodlivých programů, kterými jsou například většina druhů malwarů, může také dojít k ztrátě dat. (11) Napadení virem či malwarem se může stát lidskou chybou či využití slabiny v používaném softwaru, kde může dojít k odcizení dat či zablokování dat a poté vyžadování tzv. výkupného pro znovu odblokování daných dat. V nejhorších případech se ale jedná o útok, který má data zničit a tím poškodit danou firmu.

### **Krádež**

Když zloděj či jiný zločinný element odcizí buď počítač s citlivými daty anebo jiná přenosná zařízení.

### **Přírodní katastrofy**

Požár, záplavy či jiné přírodní katastrofy mohou způsobit ztrátu dat, přičemž 40 až 60 % malých podniků po katastrofě už nikdy neotevře. (10)

### **Výpadky a přepětí sítě**

Výpadky proudu a přepětí mohou vést ke ztrátě dat, zejména pokud jsou data v průběhu ukládání nebo nejsou řádně uložena či zálohována. (12)

## **3.6 Datová úložiště**

Pro správnou zálohu a archivaci dat je nutno vybrat dobré a kvalitní úložiště dle vašich parametrů. Každé úložiště má jiné parametry, kterými jsou kapacita, rychlost čtení zápisu, životnosti dat, jak dané úložiště přistupuje k datům a zda zařízení bude fungovat i bez připojení. Úložiště, které je vhodné pro dlouhodobou archivaci dat ale má pomalý zápis,



bude výhodnější právě pro archivace, a ne rutinní zálohování. Proto je tedy nutno zvolit to správné zařízení, podle toho, co firma potřebuje a v jaké cenové hladině se pohybuje, protože cena je také důležitým faktorem.

### 3.6.1 Magnetické pásky

Magnetické pásky jsou jedním z nejstarších způsobů ukládání dat, ale i přes jejich stáří jsou v dnešní době stále využívány pro archivaci dat kvůli jejich kvalitám, mezi které náleží:

- Cena – Jejich cena je jednou z nejmenších mezi ostatními datovými úložišti.
- Vysoká kapacita – Jedna páska může uložit velký objem dat, přičemž jsou zde i takové, které mohou uložit až několik terabajtů dat. (13)
- Výdrž – Magnetické pásky mají výdrž až 30 let a jsou mnohem méně náchylné k fyzickému poškození na rozdíl od ostatních zařízení, jako například pevný disk. (14)
- Přenosnost – Pásky jsou jednoduše přenosné, a proto jsou spolehlivou volbou pro ukládání mimo hlavní budovy v případě havárie. (14)
- Offline záloha – Pásky mohou fungovat offline a mohou poskytovat verze důležitých souborů, které jsou téměř imunní proti kyberútokům. (15)
- Dlouhá životnost dat – Integrita dat uložených na páskách se postupem času výrazně nezhoršuje, což z nich dělá spolehlivou volbu pro dlouhodobé ukládání dat. (15)

Navzdory zlepšování diskových a cloudových technologií tak páskové zálohování zůstává klíčovou součástí infrastruktury mnoha moderních firem. Přibližně 80 % podniků zahrnuje pásky do svých zálohovacích systémů a zájem o páskové zálohovací technologie v posledních letech pořád roste. (16)

### 3.6.2 Optické disky

Optické disky jsou také starší možností pomocí, kterými lze zálohovat a archivovat data. Ve své době byly velice populární hned z několika důvodů. Kvůli jejich přenosnosti, jednoduchosti v přečtení i zápisu dat, protože skoro každý firemní i osobní počítač měl čtečku pro tato zařízení, která i umožňovala zápis neboli tzv. vypálení dat na tento disk.

Jejich cena byla v té době lákavou alternativou k pevnému disku. V dnešní době se ale toto médium nepoužívá.

Optických disků je hned několik druhů. Prvním z nich je CD (Kompaktní disk), který umožňoval úložnou kapacitu cca 700 MB dat, dále zde bylo DVD (Digitální video disk), které mělo úložní kapacitu 4,7GB v jedné vrstvě a 8,5 GB ve dvou vrstvách. Nakonec je zde Blu-ray, který nám umožňuje uložit 25 GB v jedné vrstvě a 50-60 GB ve dvou je tu zde i Blu-ray Disk eXtra large s kapacitou až 128 GB. Jak už bylo zmíněno, disky se na zálohu dat už moc nepoužívají, ale stále i když už mizivě se používají na prodej fyzických kopií filmů a her. (17)

Výhody zálohy na optický disk:

- Výdrž – stejně jako magnetické pásky tak i optické disky mají menší šanci na to, že ztratí data kvůli toku času než pevné disky.
- Cena – Optické disky jsou poměrně jednoduché a levné na výrobu.
- Nevýhody zálohy dat na optický disk:
- Limitovaná kapacita – Optické disky mají velice limitovanou kapacitu oproti pevným diskům.
- Náchylnost k poškození – Optické disky mohou být poškozeny mnoha způsoby, kterými je například: přímý svit slunce, velké teplo a škrábance.
- Pomalá rychlost přenosu či čtení dat – Získávání dat z optického disku může být pomalejší než z pevných disků. (18)

### 3.6.3 Pevná a SSD disková úložiště

Pevný disk je jedním z nejvíce rozšířeným médiem pro ukládání dat. Disky se rozdělují hned podle několika parametrů. Disky mohou být interní, tedy uvnitř počítačového systému, a tedy určeny k pouze k ukládání dat z daného systému. Opakem je disk externí, který je přizpůsoben použití mimo interní komponenty počítače, proto se data na něm dají lehce přenášet mezi ostatními počítačovými jednotkami. Je pouze potřeba ho připojit do určitého portu, ve většině případech je to USB. Výhodou ukládání dat na externí disk je prevence proti odcizení dat či jinému zneprístupnění dat tím, že daný disk schováme do bezpečného

úložiště, například trezoru. Při přenosu externích disků také musíme dbát na fyzické nárazy, které by mohly daný disk poškodit.

Podle stylu ukládání dat rozdělujeme disky na HDD (Hard disk drive) a SSD (Solid state drive).

HDD – Je typ disku, kterému se přezdívá mechanický, protože jak vyplývá z názvu, u disku se používají mechanické součástky na ukládání dat. Díky tomu jsou nevolatilní, a proto zachovávají data, i když nejsou připojeny do elektrické sítě na rozdíl od paměti RAM. Důvodem, proč je tento disk nazýván mechanický, spočívá ve způsobu, jak disk ukládá data. Uvnitř disku se nachází hned několik mechanických součástí, které provádějí právě ono ukládání dat. Místo, kam se ukládají data, se nazývají plotny, které jsou magnetické a při spuštění se začnou rychle otáčet, a proto někdy můžeme u starších disků slyšet zvuky, protože se právě jedná o ony plotny, které se začínají točit. Dále se v disku nachází ručička se zapisovací a čtecí hlavičkou, která se pohybuje po plotnách a buď data zapisuje nebo čte. Tyto součástky nemají velkou odolnost, a právě proto je vnějšek disku sestaven z pevných materiálů a pevně zapečetěn tak aby se do disku nedalo jednoduše dostat. Přestože disk je takto ochráněn, bychom se měli vyvarovat přílišných nárazů anebo silným magnetům. (19) Pevný disk se zdá jako skoro až ideální zařízení pro zálohu a archivaci, ale jako každé zařízení má svoje výhody a nevýhody.

Výhody pevných disků:

- Kapacita a cena – V dnešní době jsou pevné disky s velkou kapacitou až několika terabajtů velice dostupné jak po cenové, tak množství stránce.
- Možnost zápisu dat – Pevné disky umožňují až skoro nekonečnou možnost přepisu dat.
- Jednoduchost – Zálohu na pevný disk lze provést velice rychle a bez nutnosti vyznání se v oboru.
- Rychlost – Rychlost pevných disků je větší než jak už v již dříve zmiňovaném optickém disku,
- Možnost použít i bez internetu – K datům se dostaneme, i když je například výpadek sítě.

Nevýhody pevných disků:

- Výdrž – I když lze pevný disk nekonečně přepisovat, tak tu jsou pořád jeho mechanické součástky, které jsou náchylné k poruše, která může vyústit ke ztrátě dat.
- Náchylnost ke většině už dříve zmiňovaných problémů, kterými jsou přírodní katastrofy, napadení viry a selhání mechanických součástí vlivem času či špatné manipulaci.
- Převážení či skladování – Pevné disky i ve svojí nejmenší formě pořád zabírají více prostoru a jsou fyzicky těžší než třeba už zmiňované optické disky, a proto tedy je pro ně potřeba více prostoru, který jim musí být přizpůsoben proto, aby nedošlo k žádné ztrátě dat vlivem špatného skladování. (20)

Je nutno zmínit, že jednou z dalších nevýhod pevného disku je to, že i když jsou data zapsána magneticky a teoreticky, mají nekonečnou výdržnost, je zpravidla dobré minimálně jednou za několik let disk zapojit a data zkontrolovat. (21)

SSD – Solid state disky jsou nástupci HDD a časem se stávají dostupnější, ale pořád nejsou na stejné cenové úrovni jako HDD. SSD na rozdíl od HDD nemají žádné mechanické součástky a tím pádem jsou odolnější, rychlejší a tišší než HDD. SSD ukládá data pomocí tzv. flash-based memory, která způsobuje onu větší rychlost. Data na SSD se ukládají do série NAND flash čipů, které jsou organizované buď do mřížky, bloku nebo řádků. Tyto čipy mohou zachovat data i bez zapojení do sítě. Čtení a zápis na SSD je rychlejší než na HDD. Problém však nastává v situaci, kdy data potřebujeme přepsat. Mazání dat z SSD lze provádět pouze na úrovni bloku kvůli velkému napětí, které zatěžuje buňky na nižších úrovních. Řadiče SSD jsou mnohem sofistikovanější než jejich protějšky u HDD. Typicky mají SSD 4 nebo 8 kanálů, které fungují jako potrubí pro neustálý přenos dat na rozdíl od HDD, které mají jenom jeden kanál. Tato technologie je jedním z hlavních důvodů, proč je SSD rychlejší než HDD. (22) I přes všechny výhody SSD, kterými jsou rychlost zápisu, větší odolnost proti nárazům, imunita oproti poškození mechanických součástí a tichá až téměř neslyšitelná práce, jsou SSD nevhodnou volbou pro zálohování a archivaci dat hned z několika důvodů. Jednou z největších negativ pro jakoukoli firmu, která by se rozhodla zálohovat či archivovat data na SSD, je jejich cena. Dalším nedostatkem SSD je jejich omezený počet zápisů dat, který je zároveň spojený s jejich menší celkovou kapacitou.

V neposlední řadě je nutno zmínit, že v situaci, kdy se poškodí data na SSD, tak obnova daných dat je mnohem těžší a dražší než u HDD. (23)

#### 3.6.4 NAS úložiště

NAS neboli Network-Attached Storage (datové úložiště na síti) je, jak už název vypovídá, dedikované datové úložiště, které umožňuje uživatelům a jejich zařízením přístup k datům, která jsou uložena na centrální diskové jednotce.

Zařízení NAS jsou specializované počítače pro podporu velkého úložiště a přístupu do sítě. Jsou složeny z několika hlavních komponentů: pevné disky, síťové protokoly, jednoduchý operační systém a síťové rozhraní. NAS se připojí na síť pomocí routeru a poté umožní přístup k datům z jakéhokoliv zařízení, které je připojeno k síti.

NAS může být využit jak k záloze dat a k jejich archivaci následujícími způsoby: Pro archivaci může být NAS využit tím, že vytvoříme aktivní datový archív pro dlouhodobé ukládání dat. NAS zároveň může podporovat cloudové technologie.

Na zálohu dat může být NAS nastaven, aby vytvářel automatické zálohy dat, dovoluje také aby fungoval jako souborový server, tiskový server, a dokonce i jako záložní systém v případě havárie. (24)

Je nutno podotknout, že i když je NAS úžasný způsob pro ukládání velkého množství dat a umožnění jejich dostupnosti kdekoliv, NAS není přizpůsoben pro zálohování sám o sobě, a právě pro zaručení bezpečnosti a záloze dat se používají ostatní způsoby, kterými jsou například disková pole. O diskových polích se v dalších kapitole. Teď je nutno říct, že při výběru diskového pole pro váš NAS nelze použít nějakou předlohu, ale musíte se opravdu rozmyslet, co je pro vaše účely nejlepší. Jednoduchým důkazem rozmanitosti názorů, co pro jaký NAS můžeme využít, může být (25), který říká, že RAID 6 má lepší výkonnost a protekci než RAID 5 v některých případech. Dokumentace od (26) naznačuje, že RAID 5 nabízí podobný výkon jako RAID 0 s výhodou větší protekce dat než RAID 6.

### 3.6.5 Cloud

Jedná se o síť vzájemně propojených počítačů či serverů, které mohou posloužit k úschově a sdílení dat nebo používání aplikací a výpočetního výkonu. Data nejsou fyzicky uložena na uživatelském počítači, ale jsou uložena právě v cloudu, tedy na úložišti poskytovatele, ke kterému má uživatel přístup s pomocí internetu. Cloud může být využit jak k archivaci, tak k záloze dat, protože data na cloudu jsou uložena na více místech v několika kopiích. (27) Fyzické poškození dat je u cloudu skoro až minimální, ale zase je zde nebezpečí internetových hrozeb ve tvaru kyberútoků. Cloudové služby by avšak měly být velice zabezpečené a zašifrované, což by mělo tato data chránit před neoprávněným přístupem. Avšak největší problémem s cloudovou technologií je ten, že pro jeho operaci je potřeba nepřetržité a kvalitní připojení k internetu. (28)

### 3.7 Disková pole RAID

RAID je zkratkou anglických slov Redundant Array of Independent Disks, což v překladu znamená Vícenásobné Diskové Pole Nezávislých Disků. RAID představuje způsob, který umožňuje distribuci dat mezi více diskových úložišť, která jsou spojena do jediného systému, kde vytvářejí tzv. iluzi jediného disku. Toto spojení disků může být provedeno pomocí softwaru, kde za rozdělení dat mezi disky odpovídá operační systém či jiný specializovaný software, nebo hardwarově, s využitím diskového řadiče k řízení RAID nastavení. Obě tyto metody spojení přinášejí specifické nevýhody. Softwarové spojení může vést k vyšší zátěži procesoru daného zařízení a potenciálně zpomalit jeho funkce, mezitím hardwarové řešení, který nezatěžuje procesor, bývá finančně náročnější. (29)

Dále RAID nabízí zvýšenou ochranu dat proti riziku selhání nebo poškození jednotlivých disků. V případě výpadku některého z disků v poli se ono pole dostane do tzv. degradovaného stavu, ve kterém je jeho výkon typicky nižší, avšak stále jsou všechna uložená data k dispozici. Poté, co se vymění havarovaný disk za nový a začlení se zpět do pole, začne tzv. rekonstrukce pole, při které dochází k dopočítávání a zapisování chybějících údajů na nový disk. Po dokončení rekonstrukce je RAID opět plně funkční a začne se opět synchronizovat. V některých případech, kdy se v poli nachází tzv. rezervní disk, dojde automaticky k rekonstrukci dat.

Je důležité zmínit, že ačkoliv technologie RAID přináší zvýšenou ochranu dat v případě vyřazení či poškození disku z provozu, tak není toto řešení samo o sobě složitou zálohovací strategií. Poskytuje ochranu dat proti jejich vymazání nebo umožňuje jejich obnovu do předchozího stavu.

Technologie RAID existuje v různých provedeních a nastaveních, označovaných jako úrovně, které jsou znázorněny pomocí číslic. Každá úroveň nabízí jiný způsob zápisu a distribuce dat mezi disky. Mezitím, co se některé úrovně zaměřují na rychlost daného disku, jiné upřednostňují větší důraz na ochranu dat před chybami, někdy je dokonce možné tyto funkce kombinovat. (30) Pro účely zálohování se obvykle doporučují úrovně RAID 1, RAID 5, RAID 6 a RAID 10, přičemž zejména poslední tři jmenované úrovně nabízí největší výhody. (5)

### 3.7.1 RAID 0

RAID pole 0 ve skutečnosti vlastně není reálným polem RAID, protože neobsahuje žádné redundandní informace, a proto neposkytuje datům žádnou ochranu. Když se porouchá člen tohoto pole, tak to znamená ztrátu dat. Jednotlivá zařízení v tomto poli jsou spojena do logického celku a vytvářejí tak kapacitu součtu všech daných členů. Toto spojení může být dvěma způsoby. Prvním ze způsobů je zřetězení a druhým způsobem je prokládání.

#### **Zřetězení**

Pomocí zřetězení budou vaše data postupně ukládána na disky v poli tak, že jakmile se zaplní první, začne se ukládat na druhý, pokud máte třetí, tak na třetí a tak dále až do vyčerpání počtů disků. Výhodami tohoto typu je zvětšení kapacity přidáváním dalších disků a skutečnost toho, že při výpadku některého ze členů je možnost, že některé soubory budou nedotčené.

## **Prokládání**

Prokládáním jsou data na disky ukládána cyklicky. Prostor na diskových jednotkách je rozdělen na části pevně dané velikosti a zápis nebo čtení delšího úseku dat probíhá z více disků. Při poškození disku tak není pravděpodobné, že by nějaký soubor zůstal nepoškozen. Avšak prokládání může zrychlit čtení i zápis větších bloků dat, protože je možné zároveň číst nebo zapisovat jeden blok z jednoho disku a následující blok z disku jiného. Zrychlení čtení by mělo být teoreticky menší než v případě RAID 1, avšak při reálném použití je čtení i zápis v poli RAID 0 výrazně rychlejší než v RAID 1. (30)

### **3.7.2 RAID 1**

RAID 1 též zvaný mirroring nebo v českém jazyce zrcadlení. Tento typ RAID je poměrně jednoduchá, ale efektivní ochrana dat. Tato ochrana se provádí zrcadlením obsahu disků tím způsobem, že se jejich obsah současně zaznamenává na dva disky. V případě toho, že by jeden disk havaroval, či byl poškozen, se pracuje s kopií, která je ihned k dispozici. Je zde šance že rychlost zápisu na disky v RAID 1 bude pomalejší kvůli právě onomu zápisu na více disků. Jedinou nevýhodou RAID 1 je skutečnost, že je zapotřebí dvojnásobná disková kapacita. (30)

### **3.7.3 RAID 5**

Minimální počet diskových jednotek pro RAID 5 jsou 3, přičemž kapacitu jednoho členu budou zabírat samoopravné kódy, které jsou uloženy na členech střídavě. Výhodou je, že lze využít paralelního přístupu k datům, protože delší úsek dat je rozprostřen mezi více disků, takže čtení je rychlejší. Nevýhodou RAID 5 je pomalejší zápis kvůli nutnosti výpočtu samoopravného kódu, díky tomu je ale odolný proti výpadku jednoho disku. (30)

Jak už bylo zmíněno, RAID 5 není nejlepší volbou pro systémy, které zapisují veliké objemy dat, kvůli dopadu na výkonost spojeného s psaním paritních dat. Když už některý disk selže, znovuvytvoření pole RAID 5 může být také velice časově náročné. (29)



### 3.7.4 RAID 6

RAID 6 je obdoba RAID 5, avšak na rozdíl od RAID 5 se zde používají dva paritní disky, přičemž je na každém z nich parita vypočtena jiným způsobem. Opět jsou zde kvůli přetížení paritních disků jsou paritní data uložena střídavě na všech discích. Výhodou je odolnost proti výpadkům dvou disků, přičemž rychlost čtení je porovnatelná s RAID 5, ale na rozdíl od RAID 5 je zápis pomalejší právě kvůli výpočtu dvou sad paritních informací. Minimální počet disků pro sestavení RAID 6 jsou čtyři disky. V minimální konfiguraci se však RAID 6 příliš nepoužívá z důvodu malé výsledné kapacity, protože kapacita pole je poloviční, a tedy porovnatelná jako v konfiguraci RAID 1 neboli zrcadlení dvou párů disků. Při zrcadlení navíc není potřeba počítat dvě sady paritních informací a je tedy mnohem rychlejší při zápisu a nepotřebuje ani vysoký výpočetní výkon. Proto je také doporučováno používat RAID 6 u pěti a více disků. (30)

### 3.7.5 RAID 10

RAID 10 můžeme interpretovat jako RAID 1 + 0, je nutno ale nezaměnit za RAID 0+1, protože na rozdíl od něho se RAID 1+0 odlišuje tím, že se data nejdříve zrcadlí a poté se tato pole vloží do dalšího diskového pole typu RAID 0 pro větší zrychlení přenosu dat. Maximální počet pevných disků, které mohou selhat bez jakýkoliv následků, je jeden v každém poli. Tento typ pole se často používá u hodně vytížených databázových aplikací. Nemusí se zde počítat paritní data, čímž se vše zrychluje, popřípadě i zlevňuje, pokud je zapotřebí. (30)

## 3.8 Způsob zálohování

Při zálohování je nejen nutno zvážit, na jakém úložišti se bude ukládat, ale také jakým způsobem. Nejznámější způsoby, které se volí při zálohování dat, jsou plná záloha, přírůstková neboli inkrementální a rozdílová. Každý způsob spočívá v jiném přístupu ke tvorbě dané zálohy, má jinou velikost objemu dat a probíhá jinou rychlostí. Při rozhodování, jaký způsob zálohování se chce použít, je nutno si uvědomit, jak moc často dochází ke změně dat a tím pádem nutnosti k jejich zálohování. V nejlepším případě by si měla firma

rozhodnout, v jaký čas zálohu provádět tak, aby minimalizovala ztrátu dat, ale zároveň nezatížila firemní systémy tak, aby byly nepoužitelné. Vzhledem k tomu, že jednotlivé druhy záloh mají různé nároky na využití hardwaru, je nutno přemýšlet o tom, jestli zaměstnanec či uživatel bude pracovat s daným systémem či nikoliv. Zvolený způsob zálohování by tak měl odrážet potřeby jak zákazníka, tak i dané firmy, což se ale často bez kompromisu neobejde.

### 3.8.1 **Zálohování plné**

Jak už vyplývá z jeho označení, tento typ zálohování zahrnuje vytvoření kompletní zálohy neboli kopie všech dat systému, který byl zahrnut v dané záloze. To znamená, že se zálohují i data, která od posledního zálohování nebyla změněna. Hlavní výhodou této metody je její jednoduchost, což umožňuje v případě potřeby snadné obnovení systému pomocí naposledy uložené zálohy. Nevýhodou této zálohy je čas potřebný pro provedení plné zálohy, ale i pro její import, který je vysoký. Další nevýhodou je vysoký nárok na kapacitu úložných médií. (31)

### 3.8.2 **Zálohování postupné (inkrementální)**

Inkrementální zálohování je proces, při kterém se zálohují pouze ta data, která byla od poslední úplné nebo inkrementální zálohy změněna. Tím se eliminuje opakované zálohování dat, která nebyla od poslední zálohy nijak změněna. Aby bylo inkrementální zálohování efektivní, je důležité ho propojit s úplným zálohováním. Úplné zálohování poskytuje základní kopii všech dat, na která pak navazuje inkrementální zálohování. Ačkoliv je pro maximální bezpečnost doporučeno pravidelně provádět i ty zálohy úplné, hlavní předností inkrementálního zálohování je jeho efektivita a úspora času a místa, jak na discích, tak na serverech v podobě využití menšího počtu disků na rozdíl od úplného zálohování. Nevýhodou je doba importu zálohy, která nejdříve probíhá importováním plné zálohy a následně importem všech dalších přírůstkových záloh. (31)

### 3.8.3 Zálohování rozdílové (diferenciální)

Zálohování rozdílové neboli diferenciální spočívá v tom, že zálohuje všechny soubory, které byly pozmeněny od poslední zálohy. Podobně jako u inkrementální zálohy je nutno nejdříve provést zálohování úplné. Výhodou rozdílové zálohy je její snazší správa na rozdíl od té inkrementální, protože při obnovení dat je zapotřebí obnovit pouze plnou zálohu a nejnovější diferenciální zálohu. Nevýhodou diferenciální zálohy je to, že každá diferenciální záloha obsahuje všechny změny, ke kterým došlo od poslední plné zálohy, což způsobuje neustálý růst velikosti zálohy. V důsledku toho je záložní cyklus kratší než u inkrementálních záloh. (31)

## 3.9 Rotační zálohy

Rotační metody zálohování se používají při zálohování velkého množství dat. Zároveň také přispívají k rovnoměrnému využití úložných médií. Jako u ostatních zálohovacích strategií je i zde hned několik možností, které si můžete vybrat.

### 3.9.1 Round Robin

Round Robin můžeme pojmenovat jako nejjednodušší rotační schéma. Round Robin funguje takovým způsobem, že je pro každý pracovní den vyhrazeno jedno úložné zařízení. Každý den se poté provede na dané úložné médium plnou zálohu všech dat určených pro zálohování. Toto rotační schéma nám umožňuje obnovu dat s maximálním posunem zpět o jeden týden. Toto schéma je vhodné pro využití v menších společnostech a například s využitím NAS, které může fungovat jako primární úložiště dat. Zároveň je toto řešení vhodné v těch podnicích, kde není problém provádět plnou zálohu každý den, a zároveň je časový posun o týden zpět dostačující. (32)

### 3.9.2 Grandfather-father-son

Strategie zálohování známá jako "Dědeček-Otec-Syn" náleží mezi nejčastěji používané strategie. Tato strategie spočívá v použití sestav médií pro denní (syn), týdenní (otec) a měsíční (dědeček) zálohování, přičemž může být použito jedno nebo více médií současně, a klade důraz na inkrementální zálohy. Principy fungování této metody můžeme popsat následovně: Ze začátku se využívají denní sady médií (Syn), na které se každodenně, až na poslední den týdne, použijí inkrementální zálohy. Tyto sady jsou pak v následujícím týdnu přepsány novými zálohami. Týdenní sady médií (Otec) se používají pro plné zálohy, které se provádějí na konci každého pracovního týdne, a jsou přepisovány po měsíci. Nakonec, měsíční sady médií (Dědeček), obsahující tři sady, které jsou aktualizovány na konci každého měsíce a po určité době, většinou po třech nebo více měsících, jsou opět přepsány. (32)

### 3.9.3 Tower of Hanoi

Strategie zálohování inspirovaná hrou věže z Hanoje, jejíž kořeny sahají do Číny, se opírá o strategii přesouvání pěti disků mezi třemi tyčemi s cílem minimalizovat počet kroků potřebných k úspěšnému přesunu.

Při aplikaci této strategie na zálohování se obvykle používá pět sestav médií, označených například jako A, B, C, D a E. Počáteční záloha je uložena na sestavu A, přičemž další zálohy se provádějí každý druhý den. Sestava B začíná s první zálohou následující den po A a poté pokračuje v intervalu každé čtyři dny, čímž se vyhýbá krytí se zálohou na sestavě A. Záloha na sestavě C začne prvním dnem, kdy nejsou naplánovány žádné zálohy na ostatních sestavách A a B, s opakováním každých osm dní. Pro sestavy D a E je první záloha naplánována na první den, kdy nedochází k žádným zálohám na předchozích sestavách, s opakovacím cyklem šestnácti dnů. Tato metoda také nabízí flexibilitu v možnosti rozšiřování cyklu za pomoci přidání dalších médií. (32)

### 3.10 Vytváření záloh

Zálohování vždy musí provést uživatel sám nebo za pomoci některého z mnoha nástrojů. Nástroje mohou mít mnoho forem, například už vestavěné v operačním systému nebo software přímo specializovaný pro zálohování. (2)

#### 3.10.1 Manuální kopírování

Záloha pomocí manuálního kopírování je nejjednodušším způsobem, jak zálohovat naše data. Uživatel si sám určí, jaká data potřebuje zálohovat a poté je manuálně překopíruje na jím už určené úložné médium. V případě, že se jedná o jednorázové či velmi málo časté zálohování menšího počtu souborů, je tento způsob zálohování nejjednodušší a nejvíce dostupnou variantou pro zálohu. Nevýhodou této metody je fakt, že ve větším měřítku je nevhodná, nepraktická a velice náročná pro uživatele i hardware.

#### 3.10.2 Záloha pomocí specializovaných zálohovacích programů

Zálohování lze rovněž realizovat za pomoci specializovaných softwarů navržených právě pro tento účel. V porovnání s manuálním zálohováním pomocí kopírování má uživatel možnost využít software, který nejenže provede zálohu, ale dokonce umožní nastavení pravidelného a automatického zálohování. Příkladem takového softwaru může sloužit aplikace pro zálohování, která je dostupná pro počítače s operačním systémem Windows. (2)

### 3.11 Software pro zálohování

Zálohovací softwary jsou speciálně určeny pro profesionální tvorbu záloh. Tyto softwary nám poskytují nástroje, které mohou zmenšovat velikosti záloh pomocí tzv. komprese či usnadnit nebo i dokonce plně zautomatizovat celý proces zálohování. Trh se zálohovacími softwary je velký a uživatel by si proto měl vybrat takový software, který nejvíce vyhovuje potřebám uživatele. Příkladem právě takových programů, které může firma

využít: Jedním z parametrů, podle kterých jsou dané programy vybrány je ten fakt, že to jsou tzv. open source programy. Open source programy jsou ty programy, které mají volně přístupný zdrojový kód ať už pro kontrolu věrohodnosti softwaru či pro jeho změnění přesně pro vaše účely. Dalším parametrem bylo, aby mohl být program používán na více operačních systémech. Jedním z posledních parametrů je to, aby byl program určen pro využití ve firmě.

### 3.11.1 7-zip

Nejdříve se musí začít s programem 7-zip, který neprovádí zálohy sám o sobě, ale funguje jako mezikrok pro zmenšení velikosti dat, které by bylo nutno zálohovat. 7-zip je open source program, který je určen pro tzv. kompresování dat. Tato funkce může drasticky snížit objem dat, který by poté zabíral méně místa na úložných médiích. (33)

### 3.11.2 BackupPC

BackupPC je j také open source program, který se zaměřuje pro zálohování dat pro podniky. Do jeho předností se mohou zařadit možnost zálohy souborů a složek z libovolného počítače s podporou několika operačních systémů. Je zde také možnost nastavení plánu zálohování. Program také dokáže detekovat duplikáty a zbavit se jich. Program také podporuje všechny již zmíněné druhy záloh jako je celá, inkrementální a diferencální. Další výhodou programu je poskytování centrální správy pro zálohy a zálohované počítače. (34)

### 3.11.3 Bacula

Bacula je také open source program určený pro firemní zálohy. Umožňuje provedení záloh na různých operačních systémech. Bacula také podporuje zálohy přes protokoly, jako FTP, SFTP a NFS. Bacula také podporuje detekci duplikátů a kompresi a zároveň také podporuje používání šifrování pro ochranu dat před neoprávněným přístupem. Program také poskytuje centrální zprávu záloh a zálohovaných počítačů, umožňuje vytvářet a spouštět zálohovací plány či spravovat stavy záloh. Jednou z výhod Baculy je možnost funkci programu rozšířit pomocí modulů, které poskytují další funkce. Mezi tyto funkce může být zahrnuta záloha emailů a databází či využití virtuálních strojů. (35)

#### 3.11.4 Duplicati

Jako poslední je program Duplicati. Duplicati je opět open source program zaměřený na zálohování dat. Stejně jako programy před ním, program umí zálohovat data z různých operačních systémů. Možnost zálohy přes protokoly. Rozeznávání duplikátů a šifrování. Opět je zde přítomna centrální zpráva záloh. Program má také možnost zasílání emailů v případě chyby při zálohování. Duplicati má také moduly které mohou rozšířit jeho funkce. Jednou z výhod programu Duplicati je podpora zálohy do cloudových služeb, kterými mohou být například Google cloud storage, Amazon S3 nebo Microsoft azure, což umožní přístup k zálohám téměř odkudkoliv. (36)

## 4 Vlastní práce

Praktická část práce se zaměřuje na procesy zálohování a archivace dat v prostředí firmy Rohlik Group. V prvních kapitolách se práce bude zaměřovat na analýzu firmy Rohlik Group. Na prostředí, ve kterém operuje, u jakých dat se provádějí zálohy. Další kapitola se bude zaměřovat na aplikované řešení záloh. Cílem bude zjistit, jakým způsobem záloha probíhá. V další kapitole praktické části se práce bude věnovat návržení nové či vylepšení už stávající zálohovací strategie.

### 4.1 Firma Rohlik Group

Firma Rohlik Group operuje online supermarket, kde nabízí široký výběr potravin a dalších produktů, které doručuje zákazníkům až do jejich domova. Působí v několika zemích Evropy, ale cílem této práce bude analýza právě té české části firmy. Vzhledem ke skutečnosti, že se jedná o online obchod s jak webovým, tak i aplikačním prostředím, tak firma pracuje s velkými objemy dat. Firma jako Rohlik musí pracovat s velkým rozsahem dat, kterými jsou například data jako údaje o produktech, dostupnost a ceny, zároveň musí i pracovat s tzv. citlivými daty zákazníků, kterými jsou údaje jako věk, adresa a zabezpečená hesla. Dále také musí pracovat s daty potřebnými pro chod firmy, kterými mohou být data ohledně faktur a účetnictví, která je potřeba dokonce zachovávat více než několik let kvůli našemu stanovenému zákonu, a proto většina či spíše všechna data se musí zálohovat a archivovat.

#### 4.1.1 Vlivy na ztrátu dat ve firmě Rohlik Group

Jak už bylo zmíněno v teoretické části, každá firma se pokouší o to, aby se co nejúspěšněji vyvarovala ztrátě dat jak už od vlivů externích, kterými jsou přírodní katastrofy či hackerské útoky, tak i po interní vlivy typu nespokojení či špatně proškolení zaměstnanci, kteří chtějí firmě nějak uškodit. I samotná firma Rohlik Group se nevyhnula těmto problémům, které mohly vyústit ve ztrátu dat. Například v roce 2016, kdy kvůli výpadkům serverů společnosti Google nefungovaly i části firmy Rohlik Group, které se na tyto služby spoléhaly. (37) Toto však nebyl jediný incident, kdy mohlo dojít ke ztrátě dat. V roce 2018



firma čelila velkému hackerskému útoku, který dokonce vyřadil i jeden ze serverů. (38) Samozřejmě je nutno chránit nejen samotná data, ale i místa, kde jsou data uložena a tuto funkci dodržuje právě jedno z datových center firmy Rohlik Group. Tím je právě datové centrum Meteor v pražském Karlíně. I když je provoz tohoto datacentra spíše na ústupu a v dnešní době se zde nacházejí už spíše jenom NAS, které využívají pevné disky na uchovávání dat. I přes tuto skutečnost je toto centrum chráněno před vnějšími vlivy, aby se ani tato data nemohla ztratit. Do budovy mohou pouze osoby, které projdou přes recepci či jsou vlastníci přístupové karty. Objekt má nepřetržitou ostrahu a v budově jsou čidla, která detekují kouř. Zároveň je objekt hlídán kamerovým systémem.

## 4.2 Analýza současného zálohovacího řešení firmy Rohlik Group

Tato kapitola se zabývá analýzou současného způsobu zálohování dat. Nejdříve je nutno popsat, jakým způsobem se zálohují data. Dále na jakých médiích se data uchovávají, a nakonec jaké nástroje jsou využity při záloze.

### 4.2.1 Použitá strategie

Firma Rohlik Group pro důležitá data a jejich zálohy primárně využívá cloud, jmenovitě GCP neboli Google Cloud Platformu, minoritně také využívá Microsoft Azure, na kterém jsou uloženy dokumenty v rámci Office 365 a nakonec také i AWS, což jsou Amazon Web Services na kterém jsou zase uloženy data k machine learningu. Firma se zálohami spoléhá na cloud, kde se data zálohují v rámci nabízené služby, příkladem tedy mohou být 30 dní snapshotů, databází a binlogy pro každou operaci taktéž 30 dní zpět. Pro nejvíce kritická data vznikají ještě geolokačně oddělené zálohy v pražských data centrech, kde se zálohuje replikačním způsobem. Jak při přístupu, tak i při záloze na cloud se používá S3 protokol, primárně tedy při přístupu na platformu Google Cloud. U dat, která se vyhodnotí jako „méněcenná“ či „nepotřebná“ nebo nejsou využívána s velikou intenzitou, tak firma Rohlik Group využije službu Google Cloud platformy s názvem Glacier, kde se uchovávají například archivační data na rotačních discích či magnetických páskách. Firma Rohlik Group tedy provádí zálohy pomocí BaaS neboli zálohování jako služby, kde operují s daty

pomocí API, které je vystaveno od společnosti, která jim tuto službu poskytuje. Ze softwarů, které firma Rohlik Group využívá na správu dat, je nutno zmínit Veeam, obzvláště pro správu virtuálních počítačů. Nutno je také zmínit open source software CEPH, kde je ovšem jeho používání spíše na ústupu. Z pohledu hardwaru tak servery firmy Rohlik Group jsou vybaveny SSD disky od firmy Samsung a nejnovější či hodně používaná data jsou uložena na SSD discích s NVMe technologií opět od firmy Samsung. Pevné disky se v serverech firmy Rohlik Group se téměř nevyužívají až na výjimky, jako byl už výše zmíněný NAS, ve kterém se pořád ještě využívají. Co se týče zálohování dat na zařízeních propůjčených do osobního vlastnictví tak firma spoléhá, že uživatelé nemají velice důležitá data na těchto zařízeních. Jsou k tomu tedy aktivně instruováni, aby se předešlo problémům ztráty dat.

#### 4.2.2 Nástroje zálohování a vysvětlení pojmů

Firma Rohlik Group používá GCP neboli Google Cloud platform pro uchování produkčních dat a infrastruktury. Zároveň tuto platformu ale využívá pro zálohování dat především pomocí S3 protokolu. Google Cloud platform je služba nabízená společností Google, která nabízí veliké množství funkcí. Nás ovšem bude zajímat pouze funkce ukládání a zálohování dat. Jedná se také o jednu ze tří největších Cloud platform na trhu. Do této skupiny také spadá výše zmíněný Microsoft Azure a Amazon web services, které firma Rohlik Group také využívá, ale ne v takové míře jako Google Cloud platform. Důvodem, proč si firma Rohlik Group vybrala právě Google Cloud platform, jsou výhody této platformy, které spočívají ve škálovatelnosti této platformy pro potřeby firmy, na spolehlivosti a bezpečnosti firmy Google, a především v ceně pro tuto platformu. Google si cenu účtuje podle lokality a využitých Gigabitů dat za měsíc. Firma Rohlik Group tedy platí za standardní úložiště 0,020 dolarů. Dále je zde možnost využití Nearline a Coldline úložiště, které stojí 0,010 dolarů a 0,004 dolaru. Poslední je archivní úložiště, které stojí 0,0012 dolarů. Všechna tato data jsou ukládána na serveru nacházejícím se v Belgii v lokalitě europe-west 1. (39)

Dále firma využívá S3 protokol, což je služba vyvinutá společností Amazon a od jeho vzniku se S3 protokol stal standardem pro cloudová úložiště a je implementován i v jiných cloudových službách, včetně Google Cloud platformě, kde je právě využíván firmou Rohlik Group. Dále S3 protokol v platformě Google Cloud nabízí možnost ukládání a načítání dat

přímo z cloudu pomocí speciálního aplikačního rozhraní, které všechno ovládání usnadňuje. V neposlední řadě také vysokou dostupností a zabezpečeností dat uložených v GCP dále umožňují nastavení automatického mazání dat po uplynutí určité doby a zároveň také umožňují uchovávání více verzí určitých dat pro snadné obnovení.

Firma Rohlik Group také využívá pro dlouhodobé ukládání dat službu Glacier. Služba Glacier je zpoplatněna ještě nižšími cenami než standartní úložní služby, ale za to nabízí nižší přístupnost k datům. Načtení dat ze služby Glacier může trvat v rámci minut až hodin na rozdíl od okamžitého přístupu standartních úložišť. Zároveň se zde platí poplatky za načtení dat ze služby Glacier. Zde je nutno zmínit, že sama služba Glacier využívá HDD neboli pevné disky pro ukládání dat, kde je předpokládána vyšší frekvence využívání. Zároveň také služba využívá magnetické pásky přesně u těch dat, kde je předpokládána méně častá přístupnost. Firma Rohlik Group využívá tedy služby Glacier k archivaci dat z důvodu její ideálnosti pro tento úkol.

Firma Rohlik Group také využívá software Veeam od stejnojmenné firmy. Zde je však nutno podotknout, že se u tohoto softwaru platí za licenci neboli, že tento produkt je placený. Software Veeam se specializuje na zálohování a obnovu dat z jakéhokoli prostředí, ať už by se jednalo o fyzické, virtuální či cloudové. Software Veeam nabízí spoustu funkcí mezi, kterými jsou:

- Zálohování pomocí tzv. obrazů, kde je možnost zálohovat celé virtuální stroje (Virtual machine) nebo i fyzické servery, přičemž se zachová datová konzistence a usnadní datová obnova.
- Software Veeam umožňuje rychlou obnovu dat na různá zařízení, včetně originálního zařízení, jiného virtuálního stroje či cloudu.
- Software také nabízí replikaci dat na vzdálený server v případě poruch či katastrof.
- Nabízí také ochranu dat ve formách deduplikace, komprese, šifrování a ověřování dat.
- Veeam také nabízí snadnou správu dat pomocí centralizované konzole.

Firma Rohlik Group využívá software Veeam především pro jeho podporu práce s virtuálními stroji. (40)

Další software, který firma Rohlik Group využívá, je software se jménem CEPH. CEPH je open-source software, který slouží k distribuci a ukládání dat zároveň také umožňuje jejich správu a ochranu na daném hardwaru. Software CEPH se skládá z několika důležitých komponent, které spolupracují, aby data byla co nejvíce dostupná. CEPH má tzv. monitory, které sledují stavy clusterů a zajišťují jejich dostupnost. Software také ukládá metadata o objektech uložených v clusteru. Součástí programu je také OSD (Object Storage Daemons), který ukládá data objektů na disky v clusteru a RADOS (Reliable Autonomic Distributed Object Storage) neboli virtuální objektové úložiště, které umožňuje přistupovat k datům uchovávaných v clusteru. Software CEPH obsahuje spoustu výhod, kterými jsou již výše zmiňovaná dostupnost dat, škálovatelnost při potřebě operovat s velkými počty dat, spolehlivost při ochraně dat a flexibilita v podpoře různých typů úložišť. (41)

#### 4.2.3 Zálohovací média

Jak už bylo zmíněno v kapitole o používané strategii u firmy Rohlik Group, tak firma ve svých vlastních serverech využívá především SSD disky s technologií NVMe od firmy Samsung. Servery samy o sobě jsou vybaveny open-source operačním systémem Linux, který se v dnešní době stal standardem pro využívání na serverech díky svým kvalitám a možnostem ho modifikovat. Pro zjištění, jaké přesně disky využívá firma Rohlik Group, je nejdříve potřeba stáhnout program nvme-cli (42), poté se do konzole napíše příkaz nvme list, který vypíše informace, jako název disku a jeho kapacitu.

Díky těmto informacím bylo zjištěno, že u hlavního serveru firma využívá 16 disků Samsung mzql27t6hbla-00a07. Tento disk od firmy Samsung byl vytvořen specificky pro firmu a podniky s využitím v serverech. Jedná se o SSD disk, který využívá technologii NVMe s následujícími parametry:

- Kapacita daného disku je 7,68 terabajtů.
- Disk lze připojit pomocí Peripheral Component Interconnect Express (PCIe) neboli Expresní mezipřipojení periferních komponent 4.0 a je zpětně kompatibilní i s předchozími verzemi PCIe až do 1.0.
- Sekvenční rychlost zápisu až do 4000 MB/s a sekvenční rychlost čtení až do 6700 MB/s.

- Náhodné zapisování dat až 200 000 IOPS (input output za vteřinu) a náhodné čtení dat až 1 100 000 IOPS.
- Disk také podporuje šifrování typu AES 256 pro ochranu dat.
- Disk také disponuje ochranou proti náhlým výpadkům proudu. (43)

Průměrná hodnota tohoto disku se na českém trhu v době psaní mé práce pohybuje okolo 25 000 tisíc korun českých. Z těchto informací můžeme vypočítat, že hlavní server firmy Rohlik Group bez slev stojí přibližně 400 000 tisíc korun českých.

Dále bylo zjištěno, že ostatní servery firmy Rohlik Group využívají jinou variantu SSD disku s NVMe technologií, který nese jméno Samsung mzql23t8hcls-00a07. Tento disk byl také vytvořen pro používání ve firmách a podnicích, a především v jejich serverech. Jeho parametry jsou následující:

- kapacita disku je 3,84 terabajtů
- disk lze opět připojit pomocí Peripheral Component Interconnect Express (PCIe) neboli expresní mezipřipojení periferních komponent 4.0 a je zpětně kompatibilní i s předchozími verzemi PCIe až do 1.0
- sekvenční rychlost zápisu až do 4100 MB/s a sekvenční rychlost čtení až do 6900 MB/s
- náhodné zapisování dat až 180 000 IOPS (input output za vteřinu) a náhodné čtení dat až 1 000 000 IOPS
- disk také podporuje šifrování typu AES 256 pro ochranu dat
- disk také disponuje ochranou proti náhlým výpadkům proudu (44)

Průměrná hodnota tohoto disku se na českém trhu v době psaní mé práce pohybuje okolo 14 000 tisíc korun českých. Z listu informací, které bylo získáno pomocí příkazu nvme list, se může spočítat, že průměrný počet těchto disků v serveru je 14. Díky těmto informacím je možno vypočítat, že ostatní servery bez slev stojí firmu Rohlik Group přibližně 196 000 tisíc korun.

#### 4.2.4 Souhrn současné strategie

Současná strategie zálohování a archivace dat firmy Rohlik Group je vysoce závislá na cloudovém řešení cizích společností, kterými jsou Google, Amazon a Microsoft. Dále je zde závislost na jediném datacentru firmy Google, které je lokalizované v Belgii. Firma také ve

své strategii používá velice kvalitní software pro zálohování s názvem Veeam, který i přes jeho cenu je jeden z nejlepších na trhu. Firma také využívá open-source software CEPH, který je spíše na ústupu. Na závěr firma ve svých serverech využívá kvalitní SSD disky s NVMe technologií od velice známé a spolehlivé firmy Samsung. Modely disků jsou nové a díky garanci od Samsungu je možno konstatovat, že výměna není potřebná.

### 4.3 Návrh optimalizace řešení zálohování

Prvním návrhem pro zlepšení strategie zálohování a archivace dat u firmy Rohlik Group je zbavit se závislosti na cloudových řešeních. Ideálním řešením je zde vytvoření či pronajmutí serverů ať už ve firmě či v jiné lokalitě, kde by byla kompletní záloha produkčních dat neboli dat nezbytných k chodu firmy. Tento server by byl vybaven SSD disky s technologií NVMe od firmy Samsung, konkrétněji modelem mzql27t6hbla-00a07, který by byl vhodný pro tento účel díky své kapacitě spolehlivosti a rychlosti. Pořizovací a sestavovací cena tohoto serveru by zpočátku byla velkým nákladem pro firmu Rohlik Group, ale jeho provoz a údržba by později stála pouze 25 000 tisíc korun českých měsíčně. Tento server by byl kritický v případě kompletního selhání datacentra od společnosti Google a umožňoval by firmě Rohlik Group operovat nepřetržitě.

Dalším doporučením je ve formě vylepšení cloudové struktury pro firmu Rohlik Group. Návrh spočívá ve využití dvou cloudových datacenter společnosti Google. V tomto návrhu bude firma Rohlik Group platit více za služby Google Cloud platform, avšak výhodou toho řešení bude duplikace dat ve dvou datových zónách neboli data i funkčnost firmy Rohlik Group bude zachována i v případě selhání jednoho z datacenter firmy Google. Ideálním datacentrem pro tuto variantu je datacentrum ve Frankfurtu, což je Google Cloud platform zóna europe-west3. Důvodem vybrání tohoto datacentra je především jeho blízkost k hlavnímu sídlu firmy Rohlik Group. Jedinou už zmiňovanou nevýhodou je, že ceny za stejné služby jsou u zóny europe-west3 vyšší než u zóny europ-west1.

Další možností optimalizace zálohovací strategie firmy Rohlik Group je zjištění, jestli je opravdu nynější zálohovací software Veeam nejvíce vhodným pro požadavky firmy Rohlik Group anebo jestli by nemohl být nahrazen softwarem jiným, který je lepší či levnější ve funkcích právě potřebných pro firmu Rohlik Group. Při rozhodování o výběru nového či

o ponechání starého softwaru bylo rozhodnuto pomocí vícekritériální analýzy variant. Přesněji bylo využito vícekritériální analýzy variant způsobem bodovacím a bodovacím s vahami. Pro tuto metodu bylo nutno zvolit několik parametrů, podle kterých se softwary hodnotily. Zvolené parametry byly následující:

- Úspora peněžních nákladů – kritérium, které znázorňuje ušetření peněžních nákladů v korunách českých, pokud bude zvolen daný software.
- Kompatibilita s ostatními operačními systémy – kritérium, které vyjadřuje schopnost daného softwaru pracovat i v prostředí jiných operačních systémů.
- Doba použitelnosti licencí – kritérium vyjadřuje dobu, po kterou je licence platná po zakoupení či jiném způsobu získání. Maximální hodnotu zde představuje licence neomezená.
- Pochopitelnost uživatelského rozhraní – Toto kritérium představuje, jak moc je daný software pochopitelný bez jakéhokoli předešlého používání.
- Funkce – kritérium, ve kterém je zahrnuto kvalita a množství základních funkcí, které má software v sobě od základu.
- Podpora Cloudových technologií – kritérium, které znázorňuje, jak daný software dokáže v základu pracovat s Cloudem a jeho technologiemi.
- Dokumentace – kritérium vypovídající o existenci, kvalitě či obsáhlosti dokumentaci neboli popisu funkcí u daného softwaru.
- Zabezpečení – kritérium zabezpečení vyjadřuje, jak moc se daný software stará o či poskytuje zabezpečení dat, se kterými pracuje.
- Dodatečné funkce – kritérium dodatečných funkcích znázorňuje možnost získání dodatečných funkcí do daného softwaru, ať už za pomoci zakoupení modulů či stažení volně přístupných modulů pro vylepšení.

Kritéria byla změněna do maximalizačního tvaru. Kritéria jsou bodově ohodnocená na škále od jedné [1] do deseti [10], kde číslo deset představuje nejlepší ohodnocení kritéria a číslo 1 naopak nejhorší ohodnocení. Váhová ohodnocení kritérií byla určena na základě konzultace s pracovníky firmy Rohlík Group a autora uvážení na základě analýzy požadavků strategie zálohovací strategie firmy Rohlík Group. První krok vícekritériální analýzy variant nalezneme v (Tabulka 1) a (Tabulka 2). Výsledek vícekritériální analýzy variant nalezneme v (Tabulka 3).

Tabulka 1 První část vícekriteriální analýzy variant pro software na zálohování

	Úspora peněžních nákladů (Kč)	Kompatibilita s ostatními OS	Doba použitelnosti licencí	Pochopitelnost uživatelského rozhraní
BackupPC	8	7	10	6
Bacula	9	9	10	3
Duplicati	7	10	8	7
Veeam	1	10	7	10
CEPH	7	3	10	1
Kritérium	MAX	MAX	MAX	MAX
Váhy	0,145	0,109	0,127	0,073

Zdroj vlastní zpracování

Tabulka 2 Druhá část vícekriteriální analýzy variant pro software na zálohování

	Funkce	Podpora Cloudových technologií	Dokumentace	Zabezpečení	Dodatečné funkce
BackupPC	4	4	7	6	7
Bacula	8	4	9	9	8
Duplicati	9	10	5	8	8
Veeam	10	10	10	10	9
CEPH	8	7	7	10	9
Kritérium	MAX	MAX	MAX	MAX	MAX
Váhy	0,091	0,182	0,073	0,145	0,055

Zdroj vlastní zpracování



Tabulka 3 Výsledky vícekriteriální analýzy dat pro software na zálohování

	Metoda bodovací	Metoda bodovací s vahami
BackupPC	59	6,49
Bacula	69	7,64
Duplicati	72	8,24
Veeam	77	8,25
CEPH	62	7,15

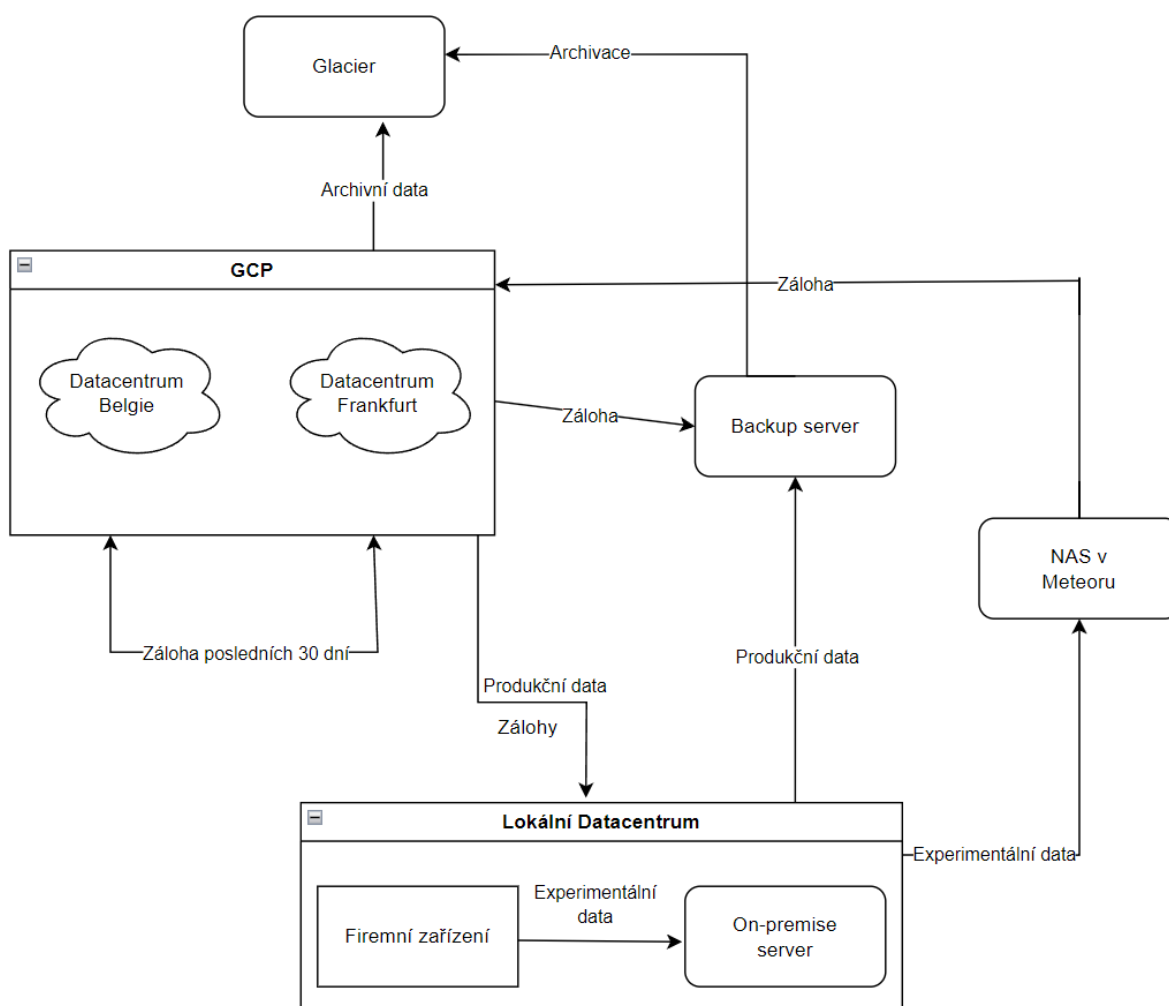
Zdroj vlastní zpracování

Podle výsledků vícekriteriální analýzy variant (Tabulka 3) lze vyhodnotit, že stávající software, který firma Rohlik Group využívá, je adekvátní a odvádí kvalitní práci a za daných podmínek a vah kritérií ho není doporučeno měnit za software jiný.

Pokud by se, avšak situace změnila a byly by změněny preference vah u firmy Rohlik Group, je zde možno doporučení ať už pro firmu Rohlik Group či pro jinou firmu ve stejném oboru zvážit možnost využití druhého nejlepšího softwaru, kterým je software Duplicati. Výhodou softwaru Duplicati je, že v jeho základní podobě se jedná o open-source neplacený produkt, který se jeho funkcemi vyrovná i některým placeným softwarům. V jeho pro placené verzi pro podniky tyto kvality nabývají ještě lepších rozměrů s jediným problémem, kterým je, že i přes všechny tyto kvality se nevyrovná softwaru jako je Veeam i přes jeho několikanásobně větší cenu.

Na (Obrázek 1) je vyobrazen tok dat firmy Rohlik Group, kdyby byla využita navržená optimalizace. V GCP, které využívá dvě datacentra – Belgie a Frankfurt je uložena infrastruktura, produkční data a zálohy. Tato data mají dostupnost až 30 dní zpět, poté jsou smazána, pokud nejsou určena pro archivaci, v tomto případě budou uložena do služby GCP Glacier. Záloha z GCP je také replikačně ukládána na backup server. S daty z GCP pracuje lokální datacentrum, kde vznikají experimentální data, pod kterými si můžeme představit různé výzkumy, analýzy trhu, experimentální funkce. Tato experimentální data jsou uložena na on premise serveru nebo v NAS v Meteoru, který tato data zálohuje do GCP. Produkční data se také posílají na backup server, který tato data uchovává a pokud jsou data určena data k archivaci, tak je opět archivuje do GCP Glacier.

Obrázek 1 Schéma toku dat firmy Rohlik Group při použití optimalizované strategie



Zdroj vlastní zpracování

## 5 Zhodnocení výsledků analýzy

Hlavním výstupem práce bylo provedení důkladné analýzy současné strategie zálohování a archivace dat ve firmě Rohlik Group, která provozuje internetový obchod rohlík.cz a pracuje s velkým objemem dat na denní bázi. Součástí analýzy byl popis firmy a seznámení, s jakými daty pracuje, dále zde byl čtenář seznámen o způsobu, jak firma chrání svá data jak před vnějšími, tak vnitřními vlivy.

Díky hlavní části analýzy bylo zjištěno, že firma Rohlik Group pro zálohování svých dat především využívá cloudových technologií od firem jako je Google, Amazon a Microsoft. Především tedy využívá GCP neboli Google Cloud platforma, na kterou se pomocí replikačního způsobu zálohování zapisovala data. Firma také využívá datacentra v České republice a v analýze bylo jmenovitě zmíněno datacentrum Meteor v pražském Karlíně, kde firma ještě využívá NAS neboli Network attached storage. Pomocí poznatků z teoretické části autor vyhodnotil, že příliš velká závislost na cloudových technologiích by mohla v důsledku jejich výpadku negativně ohrozit jak chod firmy, tak její data. Prvním návrhem pro optimalizaci zálohovací strategie tedy bylo vyřešení této závislosti. Doporučením bylo pořízení nového výkonného a vícekapacitního serveru, na který by se zálohovalo opět replikační metodou. Server by zde byl v případě selhání cloudových služeb od společnosti Google a aby umožnil jak bezpečnost dat, tak i chod společnosti a jejího internetového obchodu. Další možností optimalizace bylo zvýšení používání cloudových služeb GCP a využitím místo jednoho datacentra dvě. V tomto případě v situaci, kdy jedno datacentrum selže, firma pořád může pracovat s druhým. Naopak předností firmy byl její výběr zálohovacího softwaru Veeam. Jeho nadřazenost před ostatními softwary potvrdila vícekriteriální analýza variant, která software Veeam vyhodnotila jako nejlepší pro současné potřeby firmy Rohlik Group.

Hlavním výstupem byla doporučení, která se snažila pokrýt všechny slabé části současné zálohovací strategie firmy Rohlik Group, které by mohly nastat a vyústit ve ztrátu dat. Na rozdíl od současné strategie autorovi optimalizace, eliminuje problémy v případě výpadku hlavních cloudových služeb a zároveň nastoluje otázku, jestli má firma dále pokračovat v používání jejího současného zálohovacího softwaru.

## 6 Závěr

Účelem této bakalářské práce bylo vypracovat novou či optimalizovanou strategii zálohování u firmy Rohlik Group, díky analýze její současné zálohovací strategie. Dílčími cíli bylo zjištění a popsání, jakými způsoby je řešena problematika zálohování a archivace podle literární rešerše a porovnat ji s analýzou současné strategie firmy Rohlik Group.

Teoretická část byla především zaměřena na seznámení čtenáře s problematikou zálohování a archivací dat. Byly zde čtenářovi přiblíženy termíny jako data a zároveň i procesy zálohy a archivace dat s uvedenými důvody, proč je dobré data zálohovat či archivovat. Uvedené zde byly také problémy, které mohou vyústit ve ztrátu dat nebo při špatné či nedostatečné záloze dat. Dále zde byly vysvětlena datová úložiště, na která se data ukládají a jaké mají využití při záloze dat. Práce dále navázala na problematiku diskových polí RAID, která zjednodušují operace s datovým úložištěm pro jak osobní, tak firemní účely v řešení problematiky zálohování dat. Dále byly dopodrobna vysvětleny způsoby, kterými se data mohou zálohovat. V tomto případě bylo zmíněno zálohování plné, inkrementální a diferenciální s návazností na popis využití rotačních záloh. Předposledním vysvětlovaným tématem byla vysvětlena problematika vytváření záloh, ať už manuálně či za využití nástrojů. Poslední kapitoly se věnovaly nástrojům, softwarům, kterých je možno využít jak v osobním, tak firemním prostředí pro zálohování či archivaci dat.

V praktické části bylo před analýzou provedeno seznámení čtenáře s firmou Rohlik Group, které následně vyústilo v popis vlivů, které by mohly mít vliv na ztrátu dat i s uvedenými incidenty, kdy k nim došlo. Zároveň zde byla také uvedena opatření, která provádí firma Rohlik Group, aby se těmto vlivům a incidentům vyvarovala. Následně byla provedena podrobná analýza strategie zálohování dat, ve které bylo popsáno, jaké technologie a jaké principy jsou využívány v dané strategii. V následující kapitole byly všechny tyto technologie a principy dopodrobna vysvětleny. Na tuto kapitolu navazoval pečlivý popis fyzických úložných médií, které se při strategiích využívají. Všechny tyto poznatky ve shrnutí umožnily posoudit, že i přes kvalitní část strategii zálohování firmy Rohlik Group, se v ní nacházejí i nedostatky. Na základě vypracované teoretické části bylo možno navrhnout nového a optimalizovaného řešení strategie pro zálohu a archivaci dat pro firmu Rohlik Group. Tato nová a optimalizovaná strategie zahrnovala větší využití on-premise či jiných lokálních serverů nebo zvolení jiné strategie při práci s cloudovými

platformami. Zároveň práce vyhodnocovala vhodnost využitých programů používaných při zálohování dat.

Autor doufá, že výsledek bakalářské práce bude přínosný pro firmu Rohlik Group, ale také doufá, že bude mít přínos i pro ostatní firmy stejného typu, které například nemají takové peněžní možnosti, ale stále budou moci využít poznatky z této práce.

## 7 Seznam použitých zdrojů

1. TUTORIALSPPOINT. Computer – Data and Information [online]. [cit. 2024-03-13]. Dostupné z: [https://www.tutorialspoint.com/computer\\_fundamentals/computer\\_data.htm](https://www.tutorialspoint.com/computer_fundamentals/computer_data.htm)
2. PECINOVSKÝ, Josef. Archivace a komprimace dat. Praha: Grada Publishing, 2003. ISBN 80-247-0659-8.
3. YU, Johnny a Brien POSEY. TECHTARGET. What is data archiving? [online]. [cit. 2024-03-13]. Dostupné z: <https://www.techtargget.com/searchdatabackup/definition/data-archiving>
4. DRUVA. What is a data archiving system? [online]. [cit. 2024-03-13]. Dostupné z: <https://www.druva.com/glossary/what-is-data-archiving-definition-and-related-faqs>
5. NELSON, Steven. Pro Data Backup and Recovery. Berkeley (California): Apress, 2011. ISBN 978-1-4302-2662-8.
6. BADER, Sarah. REWIND. Why Data Backup is Important for Your Business [online]. 2022 [cit. 2024-03-13]. Dostupné z: <https://rewind.com/blog/data-backup-important-for-business/>
7. BLOGS.MANAGEENGINE. Six reasons why data backups are crucial for your business [online]. 2019 [cit. 2024-03-13]. Dostupné z: <https://blogs.manageengine.com/corporate/general/2019/01/23/six-reasons-why-data-backups-are-crucial-for-your-business.html>
8. KIRBTECH. Why Data Backup is Important for Your Business [online]. [cit. 2024-03-13]. Dostupné z: <https://kirbtech.com/importance-of-data-backup-for-business/>
9. HENRY, Dave. VERITAS. 6 Reasons You Must Have Data Backup and Recovery For Your Company [online]. 2022 [cit. 2024-03-13]. Dostupné z: <https://www.veritas.com/blogs/6-reasons-you-must-have-data-backup-and-recovery-for-your-company>
10. ROCK, Tracy. INVENIOIT. Top 7 Causes of Data Loss and How to Combat Them [online]. 2022 [cit. 2024-03-13]. Dostupné z: <https://invenioit.com/continuity/top-causes-data-loss/>
11. TERAMIND. Common Causes of Data Loss Threatening Businesses Today [online]. 2022 [cit. 2024-03-14]. Dostupné z: <https://www.teramind.co/blog/causes-of-data-loss/>

12. INVESTOPEDIA. Data Loss: What it is, How it Works, Common Causes [online]. 2022 [cit. 2024-03-14]. Dostupné z: <https://www.investopedia.com/terms/d/data-loss.asp>
13. RECORDNATIONS. Why Do Businesses Still Use Backup Tapes? [online]. [cit. 2024-03-14]. Dostupné z: <https://www.recordnations.com/articles/backup-tapes/>
14. CROCKETT, Paul. TECHTARGET. What is magnetic tape storage? [online]. [cit. 2024-03-14]. Dostupné z: <https://www.techtarget.com/searchdatabackup/definition/magnetic-tape>
15. HU, Charlotte. POPSCI. Why a tech giant is embracing old-school magnetic tape storage [online]. 2022 [cit. 2024-03-14]. Dostupné z: <https://www.popsci.com/technology/ibm-diamondback-tape-library/>
16. NAKIVO. Magnetic Tape Backup in 2023 [online]. 2023 [cit. 2024-03-14]. Dostupné z: <https://www.nakivo.com/blog/tape-backup-overview/>
17. VÍTEK, Jan. SVETHARDWARE. Zálohování a archivace dat: jaké jsou možnosti? [online]. 2016 [cit. 2024-03-14]. Dostupné z: <https://www.svethardware.cz/zalohovani-a-archivace-dat-jake-jsou-moznosti/43212-2>
18. SHELDON, Robert a Rodney BROWN. TECHTARGET. What is optical storage ? [online]. [cit. 2024-03-14]. Dostupné z: <https://www.techtarget.com/searchstorage/definition/optical-storage>
19. CRUCIAL. What is a hard disk drive (HDD)? [online]. [cit. 2024-03-14]. Dostupné z: <https://www.crucial.com/articles/pc-builders/what-is-a-hard-drive>
20. Z-DATABASE. HDD Backup [online]. [cit. 2024-03-14]. Dostupné z: <https://z-database.de/en/hdd-backup/>
21. SMITH, Francis. HARDDRIVETALK. How long can an SSD store data without power? [online]. [cit. 2024-03-14]. Dostupné z: <https://harddrivetalk.com/how-long-can-an-ssd-store-data-without-power/>
22. DATARECOVERYSPECIALISTS. How do SSDs Store Data? [online]. [cit. 2024-03-14]. Dostupné z: <http://www.datarecoveryspecialists.co.uk/blog/how-do-ssds-store-data>
23. COBB, Mike. DRIVESAVERS DATARECOVERY. Pros and Cons of Solid State Drives (SSD's) [online]. 2021 [cit. 2024-03-14]. Dostupné z: <https://drivesaversdatarecovery.com/pros-and-cons-of-solid-state-drives-ssds/>

24. BIGELOW, Stephen J. TECHTARGET. What is network-attached storage (NAS)? [online]. [cit. 2024-03-14]. Dostupné z: <https://www.techtarget.com/searchstorage/definition/network-attached-storage>
25. FRUHLINGER, Josh a Ann BEDNARZ. NETWORKWORLD. What is NAS (network-attached storage) and how does it work? [online]. 2021 [cit. 2024-03-14]. Dostupné z: <https://www.networkworld.com/article/965220/what-is-nas-network-attached-storage-and-how-does-it-work.html>
26. SEAGATE. RAID modes - New Network Volume Wizard [online]. [cit. 2024-03-14]. Dostupné z: <https://www.seagate.com/manuals/network-storage/business-storage-nas-os/raid-modes/network-volume-wizard/>
27. ALGOTECH. Co je cloud a jak funguje? [online]. [cit. 2024-03-14]. Dostupné z: <https://www.algotech.cz/novinky/co-je-cloud-a-jak-funguje>
28. DRBOHLAVOVÁ, Tereza. ORANGEACADEMY. Co je to Cloud: Proč ho využívat a jak funguje? [online]. 2023 [cit. 2024-03-14]. Dostupné z: <https://orangeacademy.cz/clanky/co-je-to-cloud/>
29. GILLIS, Alexander S. TECHTARGET. What is RAID ? [online]. [cit. 2024-03-14]. Dostupné z: <https://www.techtarget.com/searchstorage/definition/RAID>
30. GIGA-PC. RAID [online]. 2019 [cit. 2024-03-14]. Dostupné z: <https://www.giga-pc.cz/technicke-okenko/raid/>
31. HIESS, Michael. ANEXIA. THE 3 BEST BACKUP STRATEGIES FOR YOUR DATA BACKUP [online]. 2021 [cit. 2024-03-14]. Dostupné z: <https://anexia.com/blog/en/the-3-best-backup-strategies-for-your-data-backup/>
32. 3S. Definition and Backup Rotation [online]. 2012 [cit. 2024-03-14]. Dostupné z: <https://www.3s.cz/en/specialized-section/detail/id/46-definition-and-backup-rotation>
33. 7-ZIP. 7-zip [online]. [cit. 2024-03-14]. Dostupné z: <https://www.7-zip.org/>
34. BACKUPPC. Backuppc [online]. 2020 [cit. 2024-03-14]. Dostupné z: <https://backuppc.github.io/backuppc/>
35. BACULA. Bacula [online]. [cit. 2024-03-14]. Dostupné z: <https://www.bacula.org/>
36. DUPLICATI. Duplicati 2.0 [online]. [cit. 2024-03-14]. Dostupné z: <https://www.duplicati.com/>



37. Podrobnosti k aktuálnímu výpadku. [online]. 2016 [cit. 2024-03-14]. Dostupné z: Facebook, <https://www.facebook.com/damerohlik/posts/podrobnosti-k-aktu%C3%A1ln%C3%ADmu-v%C3%BDpadkunejen%C5%BEe-nefunguje-navigace-k-z%C3%A1kazn%C3%ADk%C5%AFm-kter%C3%A1-sp/1814741308768523/>
38. KOČÍ, Mirek. PCTUNING. Rohlik.cz čelil v pátek dosud největšímu hackerskému útoku [online]. 2018 [cit. 2024-03-14]. Dostupné z: <https://pctuning.cz/article/rohlik-cz-celil-v-patek-dosud-nejvetsimu-hackerskemu-utoku>
39. GOOGLE. Cloud Storage pricing [online]. [cit. 2024-03-14]. Dostupné z: <https://cloud.google.com/storage/pricing#europe>
40. VEEAM. Software zajišťující moderní ochranu dat, zálohování a obnovení [online]. [cit. 2024-03-14]. Dostupné z: <https://www.veeam.com/cz>
41. CEPH. The Future of Storage™ [online]. [cit. 2024-03-14]. Dostupné z: <https://ceph.io/en/>
42. GITHUB. Nvme-cli [online]. 2024 [cit. 2024-03-14]. Dostupné z: <https://github.com/linux-nvme/nvme-cli>
43. SAMSUNG. PM9A3 NVMe® U.2 7.68TB [online]. [cit. 2024-03-14]. Dostupné z: <https://www.samsung.com/us/business/computing/memory-storage/enterprise-solid-state-drives/pm9a3-nvme-u-2-ssd-7-6tb-mz-ql27t600/>
44. SAMSUNG. MZQL23T8HCLS-00A07 [online]. [cit. 2024-03-14]. Dostupné z: <https://semiconductor.samsung.com/ssd/datacenter-ssd/pm9a3/mzql23t8hcls-00a07/>

## **8 Seznam obrázků a tabulek**

### **8.1 Seznam obrázků**

Obrázek 1 Schéma toku dat firmy Rohlik Group při použití optimalizované strategie.....42

### **8.2 Seznam tabulek**

Tabulka 1 První část vícekriteriální analýzy variant pro software na zálohování.....40

Tabulka 2 Druhá část vícekriteriální analýzy variant pro software na zálohování.....40

Tabulka 3 Výsledky vícekriteriální analýzy dat pro software na zálohování .....41