VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ
FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY
INSTITUTE OF INFORMATICS

ZÁLOHOVÁNÍ A DATOVÁ ÚLOŽIŠTĚ

BACKUP AND FILE SYSTEMS

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

Michal Gabriel

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. Jiří Kříž, Ph.D.

BRNO 2022

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA PODNIKATELSKÁ

Zadání bakalářské práce

Ústav:	Ústav informatiky
Student:	**Michal Gabriel**
Vedoucí práce:	**Ing. Jiří Kříž, Ph.D.**
Akademický rok:	2021/22
Studijní program:	Manažerská informatika

Garant studijního programu 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 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 vypracování návrhu zálohování a datového úložiště na souborovém systému ZFS pro pokladní, účetní software a jiné soubory.

Základní literární prameny:

FONG, Yinfung a Stephen MANLEY. Efficient true image recovery of data from full, differential, and incremental backups. 2004. Spojené státy americké. US007251749. Uděleno 31.07.2007. Zapsáno 12.02.2004

CHUNLU WANG, YUANYUAN WU, ZHANYE WANG a TAO XU. ISCSI-based data protection system for virtual machine. Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC) [online]. IEEE, 2013, 2013, , 2085-2089 [cit. 2021-11-17]. ISBN 978-1-4799-2565-0. Dostupné z: doi:10.1109/MEC.2013.6885394

LOESER, Henrik. Manage Your Cloud Object Storage Data with the MinIO Client and rclone: Simple access to your S3-based data on IBM Cloud [online]. 18.11.2021, , 1 [cit. 2021-11-23]. Dostupné z: https://www.ibm.com/cloud/blog/manage-your-cloud-object-storage-data-with-the-minio-client-and-rclone

RUGGIERO, Paul a Matthew A. HECKATHORN. Data Backup Options. United States Computer Emergency Readiness Team [online]. USA [cit. 2021-11-24]. Dostupné z: https://us-cert.cisa.gov/sites/default/files/publications/data_backup_options.pdf

VUPPALA, Sai Pranav. Focus: Backing Up Your Data with UrBackup. Open Source for You [online]. India, New Dehli, 2021, (2021) [cit. 2021-11-23]. ISSN 2456-4885. Dostupné z: https://www.proquest.com/docview/2584750273/fulltext/ACDC86846AEC4D30PQ/9

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2021/22

V Brně dne 28.2.2022

L. S.

Ing. Jiří Kříž, Ph.D. garant	doc. Ing. Vojtěch Bartoš, Ph.D. děkan

Abstrakt

V této bakalářské práce je návrh funkčního řešení zálohování a datového uložiště na souborovém systému ZFS, které je alternativou k dlouho používanému RAID řešení.

Abstract

In this bachelor‘s thesis is design for functional backup solution and data storage on ZFS filesystem, which is alternative/similar to RAID solution which was used yearly ago.

Klíčová slova

ZFS, RAID, ARC, L2ARC, ZEDLET, URBACKUP, RAIDZ, RAIDZ1, RAIDZ2, UPS

Key words

ZFS, RAID, ARC, L2ARC, ZEDLET, URBACKUP, RAIDZ, RAIDZ1, RAIDZ2, UPS

Bibliografická citace

GABRIEL, Michal. *Zálohování a datová úložiště*. Brno, 2022. Dostupné také z: https://www.vutbr.cz/studenti/zav-prace/detail/143765. 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 11. května 2022

…….……

podpis studenta

Poděkování

Rád bych poděkoval vedoucímu práce Ing. Jiřímu Křížovi, Ph.D. za cenné rady při vykonávání této práce, také bych rád poděkoval rodině za podporu při studiu, a také bych rád poděkoval všem vyučujícím, kteří mně provázeli studiem na fakultě podnikatelské. Dále bych rád poděkoval oponentovi Ing. Rudolfu Čejkovi práce za cenné rady a připomínky k této práci.

OBSAH

ÚVOD

Tématem této práce bude návrh řešení pro zálohování a datové uložiště pro konkrétní implementaci ve společnosti.

V aktuální době jsou v téměř každé domácnosti nebo ve společnostech jisté prvky digitálních technologií. S rostoucím množstvím dat a důležitostí soukromých nebo vnitrofiremních dat je moudré si tato data zabezpečit proti jejich ztrátě. Typickým příkladem těchto dat mohou být fotografie nebo obchodní data společnosti. V případě, že tomu firemní politika vyhovuje nebo je na této věci zájem jednotlivce, tak je třeba tato data vhodným způsobem ochránit.

VYMEZENÍ PROBLÉMU A CÍLE PRÁCE

Cílem této práce je vypracování návrhu zálohování a datového uložiště na souborovém systému ZFS pro pokladní, účetní software a jiné soubory.

V první kapitole popíši společnost, její předmět podnikání. Zmíním se také o požadavcích investora, která budou klíčová pro návrh systému.

V teoretické části zmíním témata, kterými se v práci budu dále zabývat.

Výsledkem bude návrh projektu, který bude navržen s ohledem na funkčnost a bezpečnost.

1 TEORETICKÁ VÝCHODISKA PRÁCE

V oboru informačních technologií je mnoho funkcí a mnoho výrazů, které mají své zkratky. Proto rád zmíním alespoň část z těch, se kterými budu pracovat.

1.1 RAID

Raid je metoda redundance dat. Funguje na principu, když se nad několika disky, v závislosti na konkrétním druhu, vytvoří virtuální blokové zařízení. Na tomto blokovém zařízení lze vytvořit následný souborový systém. Jednotlivé metody RAID kromě JBOD mají za podmínku, že pro celkovou kapacitu se počítá s násobkem velikosti nejmenšího disku. Pro datová pole RAID se nejčastěji používají řadiče podporující RAID na hardwarové úrovni. Výrobcem těchto řadičů jsou například společnosti 3Ware, Adaptec, LSI. Nevýhoda RAID pole je ne vždy úspěšná přenositelnost pole napříč řadiči různých výrobců. Z toho důvodu v případě nefunkčního řadiče je správce odkázán v ideálním případě na řadič stejného výrobce. (15)

1.1.1 **BBU Cache – Backup Battery Unit**

Jedná se o baterii, která je zapojena propojovacím kabelem do řadiče. Jde o zálohování elektrické energie k uložišti mezipaměti z paměti RAM při zápisu směrem do disků. V případě výpadku elektrické energie je tato cache schopna dočasně uchovat tuto mezipaměť. Toto dočasné uchování dat je kriticky důležité například v databázi proběhlých platebních transakcí. (16)

1.1.2 **JBOD – Just Bunch Of Drives**

„Jen hromada disků“ – Virtuální souborový systém, který můžeme složit z různých disků, různých kapacit. Při použití disku prvního 40GB a druhého 120GB, nám vznikne pole o velikosti 160GB. Nevýhodou JBOD je, že nevyužívá řadu funkcí RAID technologie. (15)

1.1.3 **RAID-0**

Stripping, neboli prokládání, jako v předchozím případě použijeme dva disky první 40GB a druhý 120GB, tak nám vznikne 80GB pole. Bloky jsou ukládány střídavě na oba dva disky. V případě více disků, je každý další blok ukládán na následující disk. (15)

1.1.4 **RAID-1**

Mirroring nebo zrcadlení, je typ kdy se každý blok ukládá na oba dva disky. V případě více než dvou se ukládá na každý z nich. Jednoduše řečeno lze hovořit o dvou kopiích dat. Výhodou RAID-1 je jednoduchá obnova dat, není potřeba dopočítávat data z kontrolních součtů jako v případě RAID-5 nebo RAID-6, což v případě většího datového pole výrazně zkrátí čas obnovy. Minimálně musíme použít 2 disky. (15)

1.1.5 **RAID-5**

Pro tento typ RAIDu je minimální počet disků 3. Nícméně při třech discích nejsme chráněni proti ztrátě dat. Proto jako optimální minimální počet se uvažují 4 disky. Data jsou ukládána postupně napříč všemi disky, přitom kontrolní součet se ukládá taky napříč všemi disky. A to z důvodu, že v případě výpadku jednoho z disků (závada, vadný disk,..) je systém schopen tyto chybějící data dopočítat na základě kontrolních součtů ze zbylých disků. Typickým příkladem je výměna disku. Nicméně není možné vyměnit současně více než jeden disk. V tom případě by systém nebyl schopen dopočítat chybějící bloky. Z tohoto důvodu se prodlužuje doba obnovy v případě, že správce potřebuje výměnit více disků, při končící životnosti disků. (15)

1.1.6 **RAID-6**

Jedná se vpodstatě o totéž jako RAID-5, nicméně kontrolní bloky jsou uloženy celkem na dvou discích ze všech. Z toho důvodu tento typ chrání proti ztrátě dat i v případě výpadku dvou disků současně. Výsledkem, je tedy vyšší míra ochrany, nebo relativně delší čas pro obnovu dat. Příkladem by bylo, jestliže při výměně jednoho disku a následnému dopočtu chybějících bloků přestane být dostupný druhý disk, tak datové pole zůstane stále konzistentní. Minimálním počtem disků jsou 4. Nicméně doporučeným minimem je 5 disků. (15)

1.1.7 **RAID-10**

Je nástavba dvou různých RAID polí. Prvně jsou vytvořena dvě samostatná zrcadla, tedy 2xRAID1, a tato dvě pole jsou postavena do pole RAID0, tedy tato pole jsou proložena. Minimálním počtem disků jsou 4. Výhodou tohoto druhu je nízká doba obnovy chybějících bloků a také vysoké rychlosti čtení dat z celkového pole. Typické použití mohou být virtualizace nebo databázové platformy. (15)

1.2 ZFS

Zettabyte file systém je konkrétní typ souborového systému, který má podobnou funkci jako RAID. Nejedná se o HW řešení datového pole, nýbrž o softwarové. Namísto RAID řadiče se zde používají HBA řadiče. HBA řadič propojuje disky zapojené do serveru přímo na PCI sběrnici na základní desce. Výhodou je rychlejší přístup k diskům a teoreticky vyšší datová propustnost. Nevýhodou je vyšší zátěž na výpočetní výkon daného serveru. ZFS je velmi náročné na alokaci RAM, a to z důvodu držení metadat souborů v paměti RAM, nicméně to záleží na konkrétní konfiguraci datového pole. Další výhodou i nevýhodou může být vysoká škálovatelnost prostřednictvím spousty konfigurovatelných parametrů ZFS, které mohou mít signifikantní roli pro rychlost datového pole a integritu dat. Rovněž v případě použití ZFS „obcházíme“ tzv. optimální konfigurace serverových sestav, kdy nejsme schopni bez pomocných nástrojů sledovat stav disků na fyzické úrovni tak jako v případě RAID řadiče. (3)

1.2.1 **PLP – Power loss protection**

Jedná se o vlastnost, kterou mohou podporovat disky. Její účel je podobný jako BBU Cache, tedy dočasně uchovat data v kondenzátoru který je umístěn v každém disku. Chrání malé množství dat v mezipaměti proti ztrátě. (3)

1.2.2 **Ashift**

Ashift značí jaká bude velikost sektoru datového pole. Nastavená hodnota ashift značí exponent 2^n bitů. Máme li pole složeno z disků které mají fyzické sektory o velikosti 512bajtů, tak použijeme 2^9 bitů (= 512B). Zatímco pro disky s 4 kilobajtovými sektory použijeme hodnotu ashift=12, tedy 2^12 bitů. V případě neodpovídající hodnoty ashift můžeme sledovat nižší výkon uložiště nebo i rychlejší opotřebení např. SSD disků. (14, str. 14)

1.2.3 **Recordsize**

V souborovém oddílu taky můžeme nastavit maximální velikost bloku, jak budou data ukládána. Při použití databází je vhodnější velikost ukládaného bloku menší než například pro filmy nebo souborové obrazy. V případě, že tuto velikost nastavíme špatně, tak můžeme sledovat výkonostní dopad celkového uložiště. Hlavním dopadem může být ztráta I/O operací anebo datového toku. Většina databázových systémů využívá 8KB bloky, proto tedy jedná-li se převážně o databázové použití, tak nastavíme recordsize na 8KB. Zatímco užíváme li datové pole jako úložiště větších souborů, tak recordsize nastavíme vyšší. Ideálním se zdá rozmezí 128KB-1MB. (2)

1.2.4 **Komprese dat**

Pro snížení obsazenosti uložiště můžeme využít kompresi dat, která při správném nastavení nemá markantní dopad na výkon datového pole. V případě, že zapneme kompresi, tak můžeme využít z následujících komprimačních metod: Gzip, Lzjb, lz4, Zle. Aktuální kompresní poměr můžeme sledovat prostřednictvím příkazu zfs get compressratio. (2)

1.2.5 **Snapshot**

Jedna z funkcí ZFS je možnost vytvoření aktuálního obrazu datového pole. Její předností je možnost nahlédnout a číst již změněné soubory nebo obnovit původní strukturu dat na základě obrazu. V tom případě by se jednalo o funkci rollback. Velikost obrazu je dána změnou souborů v datovém poli od vytvoření obrazu. Zprvu je tedy velikost obrazu téměř nulová, postupem času jak se soubory v datovém poli mění, narůstá také obsazená kapacita tímto obrazem. Obraz může být jak nejvyšší úroveň datového pole, tak se také může jednat i o jednotlivá podpole. Tento obraz je vytvořen takřka okamžitě po zadání daného příkazu. Je v režimu pouze pro čtení. (3)

1.2.6 **CoW – Copy-on-write**

V případě zápisu nebo změny dat se tento zápis provádí na kopii daného konkrétního bloku, který je logicky umístěn na jiném místě než ten původní. Teprve až je tato změna zapsána na disk, tak se změní metadata k danému souboru o pozici bloku. Výhodou je vyšší zabezpečení proti ztrátě původních dat, např. v případě ztráty elektrického napájení. Hlavním využitím CoW techniky je možnost vytváření snapshotů souborového systému. (4)

1.2.7 **TXG – Transaction Group**

TXG představuje prostor v paměti RAM, kde se dočasně ukládají požadavky k zápisu na disk. Tato transakční skupina provede zápis na disk každých 5 sekund ve výchozím nastavení. Tento interval zápisu dat je z důvodu možného skládání dat do větších bloků, a tím pádem nižšího zatížení I/O operací pro disky. Interval provedení zápisu lze upravit parametrem. (2)

1.2.8 **ECC – Error Correction Code**

ECC je technologie pamětí RAM pro dopočet chybného bitu v paměťovém sektoru za použití paritního bitu. Funguje na principu jednoho až dvou přidaných čipů na paměťovém modulu ve kterých probíhá dopočet. V případě ZFS souborového systému využijeme schopnosti ECC v ARC cache. (3)

1.2.9 ARC – Adaptive Replacement Cache

V dokumentaci ZFS najdeme konfiguraci pod názvem Primarycache. Jedná se o načtení uživatelských dat nebo metadat nebo obojího do paměti RAM. Využitím může být třeba opakované dotazování na konkrétní soubory, např. databáze, kdy se tato data nečtou z disku, ale podávají se přímo z RAM, což ve výsledku vede k vyšší datové propustnosti, nižší zátěži I/O operací disků a výrazně nižší latenci. Dle využití se dělí na dvě různé kategorie na MFU (Most frequently used) a MRU (Most recently used). Tyto parametry včetně celkového využití ARC lze monitorovat prostřednictvím příkazu *arc-summary.py*. (3)

Základní poučkou bývá, že na 1TB hrubé kapacity odpovídá 1GB paměti RAM. Mimo to samozřejmě musí být navíc také jistá volná paměť RAM pro operační systém a ostatní aplikace. (3)

1.2.10 L2ARC

L2ARC je mezistupeň mezi pamětí RAM a disky. Slouží jako read cache pro náhodný přístup, nikoliv tedy pro sekvenční čtení. Pro L2ARC nejsou nikterak velké nároky na kvalitní nebo rychlý disk. Jako vhodné se tedy jeví použít levné SSD s nízkou latencí. V případě chyby dat v této mezipaměti se data znovu načtou z datového pole. (3)

1.2.11 ZIL/SLOG

Slouží jako mezipaměť určená pro zápis. Do této mezipaměti se ukládají pouze synchronní zápisy. Pro tuto mezipaměť je třeba vhodných disků, které jsou rychlé, mají nízkou latenci a mají funkci PLP. V případě, že nemají funkci PLP, tak v případě ztráty napájení, nebo vadného disku plnícího funkci L2ARC, pravděpodobně přijdeme o 5 sekund dat zápisu v případě výchozí konfigurace TXG. Jako optimální disk se jeví SSD Intel Optane do slotu PCI-E, který má dlouhou životnost v podání TerrabyteWritten (=TBW), a zároveň nízkou latenci a vysokou propustnost. V případě, že pro tuto mezipaměť nepoužijeme samostatné uložiště, jedná se o ZIL. Pokud použijeme samostatné uložiště, např. SSD disk, tak mluvíme o této mezipaměti jako SLOG. (3)

1.2.12 Synchronní a asynchronní zápis

V případě ZFS jsou všechny zápisy dat v asynchronním režimu, tedy potvrzení o zápisu jsou vrácena okamžitě, zatímco pro synchronní data je třeba v aplikaci nebo službě vyvolat zápis ZFS funkci sync(). V tom případě potvrzení o zápisu se vrátí až teprve, když jsou

data úspěšně zapsána na disk. Synchronní zápis lze trvale deaktivovat a to parametrem sync=off. (3)

1.2.13 Deduplikace

Jedna z vlastností ZFS je také možnost deduplikovat data a tím snížit obsazený prostor na úložišti. Tedy, jestliže je na datovém poli uloženo více stejných souborů, tak v případě zapnuté deduplikace se podruhé tento soubor nezapíše na disk, vytvoří se na něj pouze odkaz v deduplikační tabulce, která je načtena v paměti RAM. Tato funkce je velmi náročná na kapacitu paměti. Základní poučka je, že na 1TB datového pole je třeba 5GB paměti RAM. (3)

1.2.14 Zed monitoring - ZFS Event Daemon

Služba sledující dění v modulu ZFS, tzv. kernel události. Typickým příkladem může být zaslání informace o odpojení pole nebo disku nebo o ztrátě komunikace s daným zařízením. Následně tyto informace zasílá do služby syslog, která je následně ukládá do souboru. Další předností je, že tato služba může také informace o změnách stavů disků zasílat i e-mailem, což je vesměs nezbytný prvek pro dohled nad celým softwarovým řešením ZFS. (14, str. 47)

1.2.15 Provázanost ZFS na jiné služby

ZFS obsahuje implementace podpory sdílení pole přes Sambu, NFS. Také nejsou překážky v použití ZFS pole do iSCSI protokolu. (3)

1.2.16 ZFS scrub a resilver

Termín resilver je analogií pojmu Rebuild u RAID řadiče. Jedná se o kontrolní přepočet dat, zdali jsou správná, a pokud ne, tak dopočte chybějící kousky dat, nebo v případě výměny disku rovnou data celého disku. Používá se příkazem zfs scrub, který prvně kontroluje data. Jestliže je třeba vyměnit disk, tak při výměně disku se chybějící data dopočítávají v procesu Resilver. (3)

1.2.17 Šifrování dat v ZFS

V závislosti na použitých datech, může být třeba zabezpečit data např. proti odcizení. Typickým příkladem mohou být strategické dokumenty podniku. Jiným příkladem mohou být osobní údaje osob. ZFS má implementované šifrování dat v případě, že jej tak nastavíme. Šifrování aktivujeme parametrem encryption=on. Následně data v tomto oddílu jsou šifrovány prostřednictvím metody AES a to v délce klíče 128,192 nebo 256 bitů. (1)

1.3 Ceph

Je škálovatelné softwarově definované datové uložiště optimalizované pro použití napříč několika servery v klastru. Jednoznačnou výhodou je možnost použití logického uložiště jako nástavby napříč několika servery. (13)

1.3.1 Objektové úložiště

Aplikace programované v jazyce C, C++, Java, Python, Ruby a PHP mohou přistupovat k uložišti přímým přístupem k RADOS rozhraní prostřednictvím knihovny Librados. RADOS je logické uložiště napříč několika servery. Další metodou je přístup přes RESTFUL rozhraní, které přistupuje přes RADOS gateway. Objektové uložiště je směrováno na aplikace. (13)

1.3.2 Blokové úložiště

Pro stanice nebo pracovní stanice je třeba využít jiného přístupu než objektového. RBD blokové úložiště se rozumí jako vytvoření virtuálního blokového zařízení, na kterém si daná stanice může vytvořit vlastní souborový systém. Na dané stanici RBD uvidíme jako samostatný fyzický disk, i když ve stanici tento disk není. Funguje tedy na podobném principu jako iSCSI protokol. Typickým využitím jsou virtuální disky pro virtualizační servery. (13)

1.3.3 Datové úložiště

Pro POSIXové systémy je možné použít připojení na CEPH uložiště prostřednictvím knihovny Fuse na následný souborový systém CephFs. (13)

1.4 UPS – Uninterruptible power supply

V případě potřeby zajištění provozu proti elektrickým výpadkům je třeba využít zařízení UPS, které jsou schopny pokrýt výpadek energie, ať úmyslný nebo neúmyslný. UPS ve většině případů zálohuje elektrickou energii v řádu pár desítek minut. V případě větších datacenter se může jednat i o řády hodin až pár desítek hodin. Cílem je zajistit zákazníkům anebo službám maximální dostupnost služeb, popř. dat. Největším výrobcem UPS zařízení je společnost APC. (12)

UPS zařízení se dělí na několik kategorií dle jejich funkčnosti a vnitřního zapojení.

1.4.1 Standby UPS

Jedná se o nejčastěji používané UPS v domácím použití. Schéma funguje tak, že ze zdroje se, jak nabíjí baterie, tak v nezávislé větvi proudí elektrická energie do spotřebičů přes přepěťovou ochranu a elektrický filtr. V případě výpadku elektrické energie UPS vyhodnotí nový stav a přepne elektrický proud z baterií do spotřebičů. (12)

1.4.2 Line Interactive UPS

Lineární UPS jsou nejčastěji používané v menších korporacích. Spotřebiče jsou napájeny z elektrické sítě přes odpojovač a následný AC/DC invertor. Z tohoto invertoru je taktéž nabíjena DC baterie. V případě výpadků elektrické energie UPS rozepne odpojovač a napájí spotřebiče z baterie. (12)

1.4.3 Online UPS s dvojitou konverzí

Dvojitá konverze v UPS je často používaná ve výkonech na 10kVA. Dle schématu je baterie umístěna mezi dvěma AC/DC invertory. Z prvního se dobíjí jak baterie, tak i přes druhý DC/AC invertor spotřebiče. Výhodou je, že spotřebiče jsou galvanicky oddělené od elektrické sítě a tudíž chráněny proti krátkodobým přepětím. Většinou tyto druhy UPS obsahují také přepínač, kdy spotřebiče přepojíme přímo na elektrickou síť a tím pádem je vnitřní elektronika UPS odpojena. Tato možnost je užitečná pro výměnu baterií online. (12)

1.5 iSCSI

iSCSI je protokol pro sdílení datového úložiště prostřednictvím IP protokolu. Server sdílí iSCSI target, celý logický disk (LUN) jako blokové zařízení. Z tohoto důvodu iSCSI klient (initiator) tento disk vidí jako fyzicky osazený, přičemž tento disk v daném klientovi ani není fyzicky zapojen. Uplatněním může být použití ve virtualizačních serverech. (5)

1.6 Cloud

Pod pojmem cloud můžeme chápat vzdálené uložiště na nějaké platformě. Typickým příkladem je datové úložiště připojené k webovému rozhraní. Z open-source řešení jsou nejčastějšími

zástupci Nextcloud a Opencloud. V případě ostatních variant největšími „hráči“ jsou Google, Microsoft, Amazon.

Cloud uložiště lze využít i pro zálohování dat. Výhodou cloudu je jednoduchý přístup k datům odkudkoliv, ochrana proti interním incidentům jako například výpadek elektrického napájení. Dále je možným uplatněním použití cloudu jako úložiště pro zálohy, kde sice může být problém kapacita linky, ale na druhou stranu je subjekt vyvarován některým rizikům.

Značnou nevýhodou může být případná nedůvěra. Konkrétně, zda-li osoba důvěřuje provozovateli cloudu, že jsou data v bezpečí. Například proti odcizení a poškození dat, riziko náhlé nedostupnosti dat vlivem výpadku spojení. Vidina šifrování v cloudu, je také o důvěře, zda provozovatel pouze neuvádí, že jsou data zabezpečena, nicméně třeba v pozadí může mít k datům přístup, což může mít dopad na interní data společnosti, případně i na osobní údaje.

1.7 Zabbix monitoring

V zájmu zachování integrity dat nebo sledování vytíženosti serveru anebo jiných důvodů je rozumné na serveru využít monitoringu stavu ZFS polí a stavu disků. A to hlavně z důvodu, aby případný administrátor věděl, že je třeba vyměnit disk v serveru nebo například, že je nedostatek paměti RAM pro ARC cache. V návrhu řešení mé bakalářské práce budu využívat výstupů Zabbix. Monitorování zařízení Zabbixem funguje prostřednictvím SNMP, ICMP a TCP kontrol. Zabbix používá dvouvrstvou architekturu klient-server. (11)

1.8 Samba

Samba je služba-protokol pro sdílení souborů. Mimo sdílení souborů, umí také plnit funkci doménového kontroléru, alternativou k Active Directory od Microsoft serveru. Nicméně nemá plnohodnotné funkce Microsoft Active Directory, což může být problematické. (17, str. 17)

V rámci mého projektu budu prostřednictvím Samba sdílení dat konfigurovat práva odděleně pro každého užiavtele prostřednictvím FACL. Sdílení souborů, je tedy dle konfigurace nezávislé na konfiguraci POSIXových práv. (17, str. 178)

1.9 Přístupová práva

V rámci společnosti je třeba zřídit řízení přístupu k datům, a to z důvodu, aby neoprávněné osoby neměly přístup k některým datům.

1.9.1 POSIX ACL

V případě UNIXových operačních systémů využíváme POSIXových práv v rozsahu 000-777 případně i se čtvrtou číslicí pro zadání atributů. První číslice značí vlastníka souboru, druhá číslice značí skupinu k souboru, a třetí číslice značí ostatní uživatele. Jednotlivé číslice nabývají hodnot 0-7. V bitovém vyjádření je to 2^3, kde se jedná o kombinaci práv pro čtení, zápis a spouštění, tedy RWX práva. Dále v POSIX ACL můžeme nastavit i takzvaný „sticky“ bit. Prostřednictvím tohoto bitu jsme schopni omezit právo smazat jednotlivou složku pro uživatele, kteří mají k této složce právo zápisu nebo pro super-user uživatele. V případě nastavování práv uživatelům a skupinám můžeme využít jak slovního vyjadření, tak i číselného vyjádření téhož - UUID nebo GUID. Tato práva nastavujeme prostřednictvím příkazu chmod. Aktuální nastavení práv k souboru nebo složce můžeme číst například příkazem ls s parametrem -l. (10)

1.9.2 FACL

V případě rozsáhlejších konfigurací oprávnění využijeme podrobnější FACL nastavení přístupu. Výhodou FACL je možnost nastavovat práva pro vícero uživatelů samostatně. Tato práva nastavujeme příkazem setfacl, čtení aktuální konfigurace práv čteme příkazem getfacl. (10)

1.10 Zálohování

V případě ochrany dat před ztrátou, a tím pádem i vzniklou škodou, je třeba do systému zahrnout zálohování dat. Pod pojmem zálohování rozumíme stav, kdy data jsou umístěná na jiném fyzickém zařízení v jiné lokalitě, např. budově, než jsou operační data, neboli běžně používaná. Tato běžně používaná data jsou k dispozici oprávněným uživatelům takřka ihned, tedy je nízká prodleva na požadavek přístupu. Pod touto nízkou prodlevou si můžeme představit jako ukazatel latenci disku nebo diskového pole, která se pohybuje ve většině případů v řádu milisekund. V případě nefunkčnosti datového pole nebo disku, a tedy nepřístupnosti k operačním datům, je třeba tato data obnovit v co nejvyšší možné míře. Jde tedy o snahu minimalizovat ztrátu jednotlivých souborů, které mohou mít vysoký dopad jak na společnost, tak na jednotlivce.

V případě obnovy dat ze zálohy se přístupová doba razantně prodlužuje, kdy už jenom obnova může trvat i několik hodin.

V této bakalářské práci se budu věnovat zálohovacímu řešením UrBackup, a to z důvodu, že používání programu je zdarma pod licencí GNU-AGPL. Existují take lepší a výkonnější řešení od jiných společností. Nicméně užívání těchto jiných programů je mnohem dražší

1.10.1 **UrBackup**

Hlavní předností programu UrBackup je možnost použití aplikací klient i server jak na UNIX operačních systémech, tak i na systémech Microsoft Windows. Zálohy jsou možné provádět on-site i off-site. (8) Výhodou aplikace UrBackup je podpora souborového systému ZFS a tedy i využití ZFS funkcí jako je komprese, deduplikace a obrazy (snapshots). (8)

1.10.2 **On-Site záloha**

V případě, že chceme zálohovat na lokální uložiště, nebo v rámci jedné datové sítě, tak mluvíme o zálohování On-Site. Jedná se o zálohu v rámci objektu. (9)

1.10.3 **Off-Site záloha**

Naopak oproti on-site zálohám, v případě off-site rozumíme tak, že zálohu provedeme na místo jiné, než kde jsou fyzicky umístěna. Typickým příkladem může být záloha na cloudové uložiště, nebo na datové pásky v jiném objektu nebo městě. Hlavní nevýhodou off-site zálohy je to, že při přenosu dat se může neoprávněná osoba pokusit odchytit tento datový tok. Pro off-site zálohy do cloudového uložiště můžeme použít například Backblaze, Google Drive, Amazon S3. (9)

1.10.4 **Obrazová záloha**

Pod pojmem „obrazová záloha“ rozumíme takovou zálohu, kdy zálohovací program vytvoří aktuálni kompletní obraz logického oddílu s tím, že v případě obnovy dat se obnovuje tento obraz jako celek. Některé zálohovací programy jsou schopny tento obraz otevřít jako katalog, kde si uživatel může z celého obrazu extrahovat konkrétní hledaný soubor. (8)

1.10.5 **Souborová záloha**

Uživatel nebo osoba spravující více počítačů může také dle vlastních preferencí navolit konkrétní soubory, které se mají zálohovat. Značnou výhodou je nižší nárok na kapacitu uložiště určeného pro zálohy. (8)

1.10.6 **Plná záloha**

Plná záloha značí, že se vždy zálohuje celá předem nakonfigurovaná množina dat, například celý oddíl. (7)

1.10.7 **Inkrementální záloha**

Naproti tomu inkrementální záloha funguje na principu zálohy příbytku dat oproti předchozí inkrementální záloze nebo po plné záloze. Pro obnovení dat z inkrementálních záloh, je třeba mít kompletní a neporušený řetězec inkrementálních záloh. (7)

1.10.8 **Diferenciální záloha**

V případě diferenciální zálohy je oproti inkrementální rozdíl v tom, že se tato záloha provede jako přírustek dat oproti poslední plné záloze. Nejedná se tedy o zálohu vztaženou oproti poslední inkrementální záloze. Pro obnovení dat s použitím diferenciálních záloh nám stačí poslední úplná záloha, plus jakákoliv dostupná diferenciální záloha. (7)

1.10.9 **Rclone**

Rclone je aplikace pro správu dat na cloudovém uložišti. Její hlavními přednostmi je možnost použít cloudové uložiště jako virtuální složku v souborovém systému, další velmi užitečnou funkcí je šifrování na straně stanice. Ke cloudovému uložišti se připojujeme prostřednictvím HMAC přihlašovacích údajů. (6) Šifrování funguje na principu zašifrování souboru a jeho metadat na lokálním PC a následně odeslání do cloudového uložiště. V případě stáhnutí se tento soubor prvně stáhne a následně na lokálním zařízení se teprve dešifruje. Nevýhodou tohoto šifrovacího procesu může být vyšší zátěž na výpočetní výkon lokálního zařízení, což je ale vyváženo ochranou dat.

2 ANALÝZA SOUČASNÉHO STAVU

V této kapitole obeznámím s aktuálním stavem firmy, včetně krátkého popisu oboru podnikání, pro který bude výsledný návrh řešení brán jako optimální.

2.1 Popis firmy

Společnost Gamitech, s.r.o. byla založena v březnu roku 2016. Oborem podnikání je gastronomie. V daném objektu je několik datových zařízení. Všechna tato zařízení je třeba zajistit proti ztrátě informací, a to zejména obchodní data v provozovně a účetní data běžného účetního období.

2.2 Firemní objekt

Datová síť je v bytovém, třípatrovém domě se sedlovou střechou. V objektu jsou zřízena pravidla přístupu do jednotlivých oblastí.

2.3 Personální zastoupení firmy

Ve firmě v době před pandemií SARS-CoV-2 bylo zaměstnáno několik osob na pozici servírka/číšník. Jejich náplní práce je obsluha hostů. Jedná se tedy o roznos nápojů a jídel, mytí znečištěného nádobí, přijímání plateb od zákazníků a veškeré další činnosti související s profesí číšník/servírka. Součástí jejich popisu práce je také zadávat průběžně obchodní data do elektronické pokladny. V praxi tedy provádějí pravidelně úkony jako objednávky, platby, přesuny objednávek a stavů, případně rezervace, a jiné nahodilé situace. Z pohledu datové sítě nemají tito provozní zaměstnanci pravomoc zasahovat do toku dat v počítačové síti. Jediné, kdy se s danou sítí aktivně svévolně dostávají do styku, je veřejný WiFi přístupový bod. Na elektronické pokladně je softwarově zabezpečen přístup uživatelským omezením použití výhradně na jednu aplikaci s nemožností jejího ukončení. Druhá forma jistého zabezpečení je použití neprivilegovaného uživatele v systému Windows 10.

2.4 Současné a budoucí potřeby investora

V současnosti si je investor vědom jistých nedostatků. Investor by rád inovoval datové úložiště. Do budoucna je požadavek na zavedení škálovatelného datového úložiště s prvky ochran proti napěťovým výpadkům a tedy následné ztrátě dat. Investor by do budoucna rád zavedl robustnější datový systém, který by sloužil i jako hlavní úložiště pro virtualizační server a jeho jednotlivé virtuální stroje. V době do 10 let by investor rád nainstaloval alternativní zdroj energie formou

fotovoltaického systému na místní elektrickou rozvodnout síť. Vyžaduje možnost napojení FVE systému do datové sítě a dále na další systémy prostřednictvím RS485 nebo Modbus kvůli monitoringu a ovládání. Jako další prvek by se rád investor v příštích letech v objektu věnoval automatizaci jak na úrovni mechanické (MAR), tak i softwarové. Jako důvod mechanické automatizace uvádí, že by tato automatizace zjednodušila procesy inventarizace zásob. Jako další budoucí potřeby investor uvádí implementaci měřících zařízení energetických zdrojů jako elektrická energie, zemní plyn a voda s možností ukládaní dat do grafů a následného vyhodnocování dat.

2.5 Hardwarové vybavení

Veškerá stávající datová uložiště fungují čistě jako společné – sdílené v okruhu opravněných osob. Všechny komponenty včetně datového uložiště jsou propojeny počítačovou sítí.

V současné době je datové uložiště zprovozněno na desktopovém zařízení s operačním systémem Microsoft Windows, s dvoujádrovým procesorem Intel Core2Duo E6750, 4GB paměti RAM a RAID řadičem ASR-3405 s připojenou BBU. V počítačové skříni jsou osazeny 3 rotační disky 3,5 palce. Jejich skladba je následující: Jeden disk 512GB je vyhrazený pro operační systém Microsoft Windows a dva disky 1TB v režimu RAID-5 zapojené do řadiče, které slouží jako sdílené uložiště dat. V rámci společnosti je toto úložiště využíváno pro klonování databázových souborů z pokladního softwaru, účetní databáze, dále jako úložiště osobních souborů. Řadič ASR-3405 je optimální pro mírné rozšíření až na čtyři disky. Systémový disk je zapojen přímo na základní desku počítače, čímž je vyřazena závislost na funkčnosti diskového řadiče. Průměrná okamžitá spotřeba tohoto zařízení činí 70W.

Sdílení souborů je nastaveno prostřednictvím Windows sdílení souborů.

Ve stávající infrastruktuře nejsou zakomponovány prvky pravidelného zálohování. Cloud je používán pouze zřídka jednotlivci na vzdálené sdílení souborů prostřednictvím URL adresy.

Druhé zařízení je zřízené jako privátní autoritativní DNS server o změřené průměrné okamžité spotřebě ve výši 100W zřízený na platformě FreeBSD, který obsluhuje přibližně 20 osob.

Tab. 1: Datová infrastruktura

Poř. Číslo	Název	Přibližná elektrická spotřeba [W]	Celkový počet disků	Skladba disků	Operační systém
1	Pokladní kasa POS	40	1		Windows 10
2	DNS	100	1		FreeBSD
3	FileServer	70	3	1+2*1	Windows 7

2.6 Používaný software

V počítačové síti je na většině stanic operační systém Microsoft Windows 10.

Dalšími síťovými aplikacemi jsou:

- Pohoda od firmy Stormware ve variantě SQL síťové edice. SQL databáze běží na systému Microsoft SQL 2012
- Pokladní software Conto od firmy Consulta. Většina operací v tomto softwaru je pouze lokální v režimu klient-server. Například synchronizování veškerých účetních dat se zapisuje do SQL serveru Firebird.

Investor by rád následně tyto systémy svázal a utvořil tedy jistý celek pro skladovou evidenci a docházky s možností pravidelné automatizované konverze obchodních dat do účetního programu.

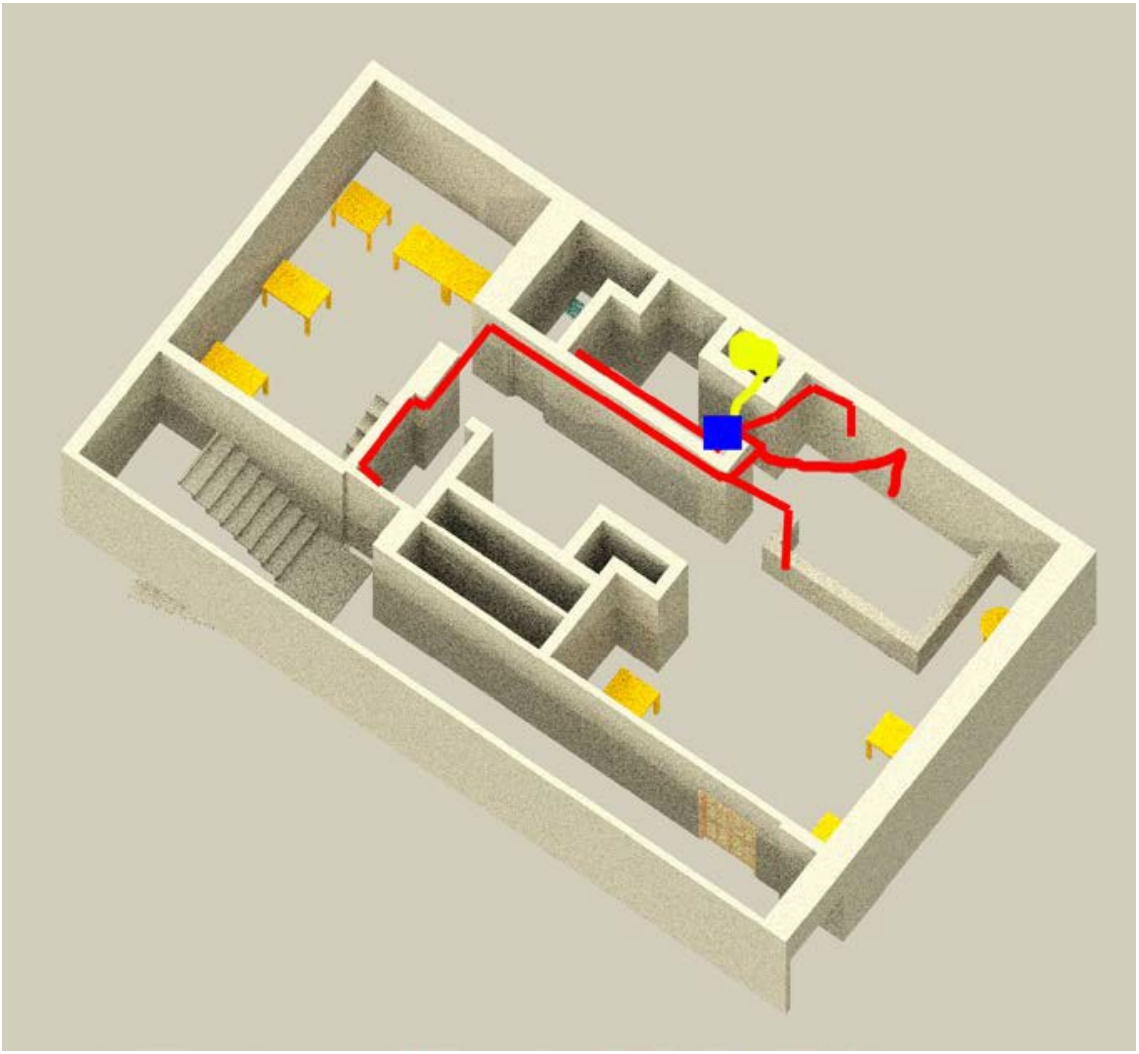
2.7 Subjektivní zhodnocení současného stavu

Stávající datové úložiště neobsahuje prvky pravidelných záloh jak sdíleného uložiště, tak zálohování jednotlivých pracovních stanic. Systém neobsahuje oznamování závad v systému, tedy systém neupozorní, zda-li je jeden ze dvou disků vadný nebo ne, což může vést až k fatálním následkům. Momentální zálohování nejdůležitějších dat probíhá formou kopírování souborů skrze intranet na disk jiného počítače. Ve stávajícím systému nejsou zřízeny žádné kvóty, žádné složitější struktury. V případě výpadku napájení (ať vadou spínaného DC zdroje, nebo přívodu elektrické energie) je možná ztráta dat, jelikož v systému uložení dat není zakomponován prvek záložního zdroje elektrické energie UPS. Data uložená na discích v jednotlivých zařízeních nejsou chráněna proti fyzickému přístupu. Jednotlivá uložiště nejsou šifrována.

2.8 Komunikační infrastruktura

V místě návrhu je v celém objektu použité výhradně metalické zapojení sítě a to zejména z finančního hlediska.

Část fyzického rozložení kabeláže je uvedeno na následujícím náčrtu. Náčrt je vyobrazením prostor společnosti. Vyznačené linky na obrázku 1 jsou vedeny volně a chaoticky ve stropním podhledu. Červeně je značená horizontální kabeláž, modře je značeno uložení routeru od společnosti Mikrotik a switche D-Link volně loženého na polici, žlutou barvou je značená vertikální kabeláž.



Obr. 1: Vizualizace datové kabeláže prostor společnosti [Vlastní zpracování]

Součástí stávajícího stavu není prvek pro monitoring síťového provozu.

2.8.1 **Telefonní infrastruktura**

V objektu je dále funkční analogová telefonní infrastruktura s telefonní ústřednou od společnosti 2N s napojením na externí ISDN bránu. Do budoucna je v plánech zřídit digitální telefonní síť se systémem Asterisk a nahrazení stávajících analogových telefonních zařízení. Jako nástupce mají být použity IP telefony Avaya se standardem IEEE 802.3af s podporou protokolu H323. Jediná analogová pobočka, která není v plánu měnit, je telefonní vrátník. Proto by nový systém obsahoval jednu PSTN kartu. Digitální telefonní ústřednu by tvořil systém asterisk a gatekeeper pro obsluhu telefonů Avaya by plnila aplikace Gnugk.

2.9 Programové vybavení

Součástí systému jsou některé aplikace, které je třeba zachovat. Hlavním důvodem je provoz v provozovně, kde servírky a číšníci jsou obeznámeni s obsluhou počítačového softwaru Consulta Conto. Druhou aplikací, která je potřeba zachovat je účetní program Stormware Pohoda za účelem podvojného účetnictví s již existujícími účetními záznamy.

2.9.1 Pokladní systém Conto

Jedná se o prodejní místo s dvouvrstvou architekturou. Podle zakoupené licence databázový server může být jak síťové verze, tak lokální verze. Tento pokladní server byl zakoupen jako lokální instalace. Proto tedy serverová databáze Firebird je nainstalována přímo na pokladním zařízení. Toto řešení znamená nevýhodu. V případě závady na počítači je pravděpodobné, že dojde k úplné ztrátě dat, protože databáze není síťová. Na druhou stranu výhodou je, že pokladní systém je nezávislý na stavu počítačové sítě. Tato nezávislost měla hlavní rozhodovací roli pro pořízení lokální instalace.

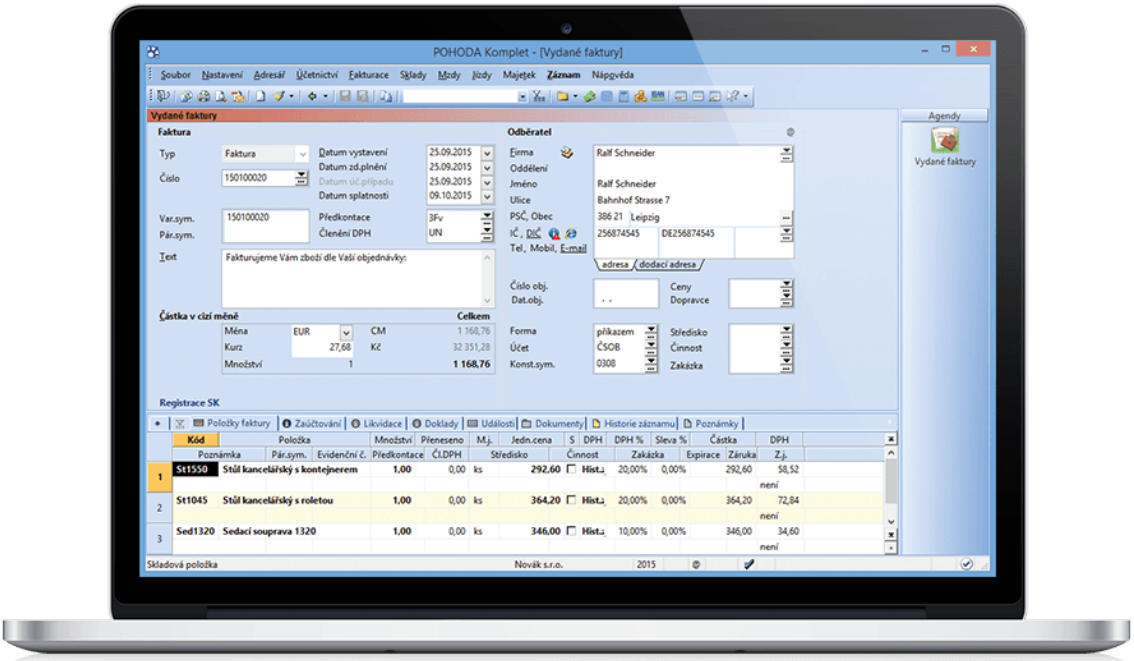
Obr. 2: Pokladní systém Conto (18)

2.9.2 Účetní program Pohoda

Software Pohoda má také možnost si zvolit lokální licenci nebo síťovou. Druhou možností výběru je použití MDB souborové databáze nebo SQL databáze. V případě, že je zakoupena SQL verze, tak jako databáze je použit Microsoft SQL Server v edici Express, který je součástí této licence programu Pohoda., jako výčet uvedu následující funkce, které Pohoda umí:

- daňová evidence
- podvojné účetnictví
- fakturace
- objednávky
- pokladna
- banka
- majetek
- kniha jízd
- skladová evidence
- personalistika

Účetní program Pohoda je potřebné zachovat, protože je v něm průběžně prováděno zapisování dokladů a zpracovávání výkazů.



Obr. 3: Stormware Pohoda (19)

2.10 Chybějící informační systémy

V systému dále chybí jakékoliv prvky centrálního sběru záznamů ze zařízení všech druhů. V případě, že kterákoliv aplikace, služba nebo zařízení začne vykazovat chyby, nebo se stane nedostupným nebo nefunkčním, tak provozovatel není obeznámen s touto změnou stavu. V horším případě ani není schopen zjistit a opravit příčinu závady. Proto bych navrhnul použití sběrného mechanismu a následného centralizovaného zpracování v ELK systému (Elastic, Logstash, Kibana)

Objekt dále neobsahuje žádné bezpečnostně testovací systémy. Investor tedy není s to sledovat zda jsou jeho systémy bezpečné či nikoliv. Ve stávajícím stavu by byla škoda zjištěna, pokud vůbec, formou domněnky při vyřazení služeb. Proto bych navrhnul na úrovni telekomunikační infrastruktury integrovat prvky IDS/IPS a na úrovni aplikační tester hrozeb v závislosti na aktuální mezinárodní databázi hrozeb CVE. Jako rozumné se mi zdá systém Nessus, nebo jednotlivé doplňky do systému ELK.

3 VLASTNÍ NÁVRHY ŘEŠENÍ

Pro svoji práci jsem si vybral jako server pro datové uložiště HP Proliant DL380e gen8 14LFF o velikosti 2U. Tento server může být osazen ze přední strany až 12ti 3,5 palcovými disky a ze zadní strany až dvěma 3,5 palcovými disky. V tomto serveru je osazen backplane podporující SAS komunikační protokol. Server je běžně dodáván s RAID řadičem HP P420 a integrovaným řadičem na základní desce HP B120i. Pro naše použití ale využijeme HBA řadiče HP H220, který je vhodný pro ZFS. Dále na základní desce serveru se nachází slot pro SD kartu a USB slot. Tyto dva sloty využijeme pro zavaděč linuxového jádra a následného načtení knihovny ZFS.

Z důvodů vysokých nároků na paměť do serveru navrhuji osadit ze začátku 12 kusy 4 GB modulů LV DIMM 1600MHz katalogové číslo HP 713981-B21. V případě potřeby je tato kapacita možná rozšířit až na 384 GB. Server bude osazen dvěma procesory Intel Xeon E5-2450L se základním taktem 1,8GHz.

Jelikož server bude plnit také roli virtualizačního serveru, tak bude osazen dvěma NVMe disky Kingston DC1000B 240 GB s funkcí PLP jako úložiště pro virtuální zařízení. Výhodou těchto disků Samsung je vysoké IOPS, v poměru ke své ceně dobrá životnost a 5 let záruka. Tyto dva NVMe disky budou připojeny do PCI-E slotu prostřednictvím adaptéru.

Jako operační systém jsem vybral Proxmox a to z důvodu, že hlavní účel serveru bude virtualizace. Další nespornou výhodou je, že Proxmox je opensource řešení a tudíž je zdarma. Při instalaci jsem vybral jako souborový systém pro root oddíl ZFS v režimu zrcadlení.

3.1 Příprava zaváděcího oddílu Grub

Jelikož server HP Proliant DL380e Gen8 neumí zavádět operační systém přímo ze zařízení se souborovým systémem ZFS, tak je třeba připravit linuxový zavaděč pro načtení linuxového jádra s podporou ZFS. Jako zavaděč použiji Grub. Tento zavaděč bude nahrán na SD kartě se souborovým systémem Ext4. Abychom mohli načíst linuxový root oddíl, tak je třeba upravit soubor grub.cfg kam připíši řádek „insmod zfs“ následně příkazem „update-grub“ aktualizuji novou konfiguraci. Tento zavaděč včetně nové konfigurace blokově zkopíruji na vnitřní USB disk pro případ, že by SD karta selhala. To provedeme příkazem „dd if=/dev/sdp of=/dev/sdm“

3.2 Vytvoření datových polí

Pole s názvem rpool bude pro root oddíl skládající se ze dvou levných SSD disků 240 GB pro operační systém Debian s nástavbou Proxmox v režimu zrcadlení. Pole s názvem nvme-pool jako blokové uložiště pro virtuální počítače, tank1 jako sdílené uložiště prostřednictvím NFS a Samba pro pracovní data, tank2 bude sloužit jako prostor pro zálohy. Cloud pole je samostatně určeno pro aplikační službu Nextcloud.

Jednotlivá pole vytvářím přímo na základě unikátního identifikátoru disků založeného na seriovém čísle disku, které můžeme zjistit prostřednictvím aplikace Smartmontools například příkazem smartctl -a /dev/sdX. Unikátní identifikátor jsem zvolil z důvodu možné záměny aliasů sdX při výměně více disků současně, nebo pří nevhodném rozpoznání aliasu při načítaní operačního systému, což by mohlo mít za následek nekonzistentnost dat v jednotlivých datových polích.

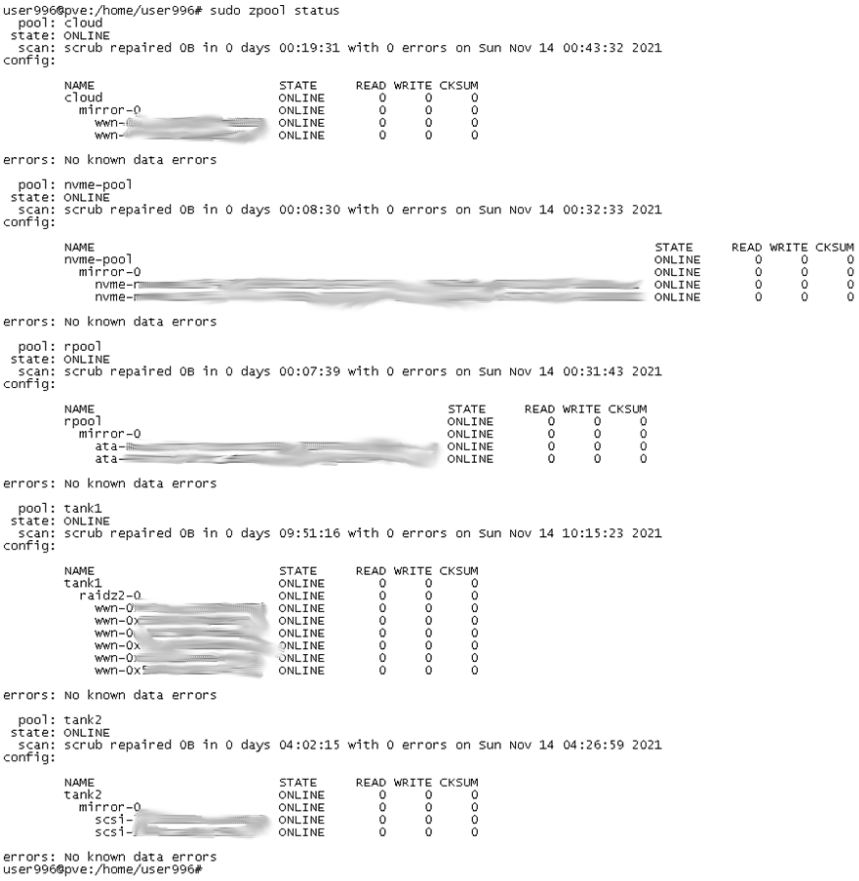
Jednotlivá datová pole vytvořím příkazem následujícím postupem.

```
# zpool create tank raidz2 c1t0d0 c2t0d0 c3t0d0 c4t0d0 c5t0d0
# zpool status -v tank
  pool: tank
 state: ONLINE
 scrub: none requested
config:

        NAME         STATE     READ WRITE CKSUM
        tank         ONLINE       0     0     0
          raidz2-0   ONLINE       0     0     0
            c1t0d0   ONLINE       0     0     0
            c2t0d0   ONLINE       0     0     0
            c3t0d0   ONLINE       0     0     0
            c4t0d0   ONLINE       0     0     0
            c5t0d0   ONLINE       0     0     0

errors: No known data errors
```

Obr. 4: Postup vytvoření ZFS pole (20)



Obr. 5: Stavy ZFS polí a jednotlivých disků [Vlastní zpracování]

Soupis stavu všech datových polí zjistíme příkazem „zpool status“. Ve výše uvedených datových polích vidíme, že všechny disky jsou v pořádku a že poslední kontrola dat z 14. listopadu proběhla bez chyb.

Na obrázku č. 13 jsou znázorněny všechny parametry datasetu rpool. Většina parametrů je výchozích. Zapnul jsem kompresi s algoritmem LZ4, vypnul jsem ukládání dat přístupů k souboru (=atime) a explicitně jsem nastavil synchronní zápis na výchozí hodnotu tedy standard. V praxi to znamená, že systém si sám vyhodnotí, kdy provede zápis synchronně a kdy asynchronně.

3.2.1 **Výpočet deduplikačního poměru**

Po naplnění pole daty zjistíme příkazem „zdb -S *pool*“ jaký by byl poměr k deduplikaci a také ukáže aktuální kompresní poměr jaký naše data mají.

Pro každé pole jsem tedy zjistil poměry k deduplikaci a také pro aktuální kompresi. Tím lze zjistit, zda by zapnutá deduplikace přinesla nějaký užitek za cenu spotřebované paměti RAM. U

datových polí cloud a nvme-pool jsou komprese vypnuté, u ostatních je použita komprimační metoda LZ4.

Tab. 2: Poměry deduplikace a komprese [Vlastní zpracování]

Název pole	Hrubá kapacita	Využití [%]	Dedup ratio	Compress ratio
cloud	2TB	7	2.22	1.00
rpool	240GB	56	1.04	1.41
nvme-pool	240GB	71	1.39	1.00
tank1	12TB	80	1.07	1.03
tank2	4TB	76	1.10	1.11

Z poměru deduplikace lze vyčíst, že užitek v snížení duplicit by byl u datového pole cloud a mírný také pro pole nvme-pool.

NAME	PROPERTY	VALUE	SOURCE
rpool	type	filesystem	-
rpool	creation	Tue Nov 23 15:19 2021	-
rpool	used	93.1G	-
rpool	available	41.6G	-
rpool	referenced	104K	-
rpool	compressratio	1.38x	-
rpool	mounted	yes	-
rpool	quota	none	default
rpool	reservation	none	default
rpool	recordsize	128K	default
rpool	mountpoint	/rpool	default
rpool	sharenfs	off	default
rpool	checksum	on	default
rpool	compression	lz4	local
rpool	atime	off	local
rpool	devices	on	default
rpool	exec	on	default
rpool	setuid	on	default
rpool	readonly	off	default
rpool	zoned	off	default
rpool	snapdir	hidden	default
rpool	aclinherit	restricted	default
rpool	createtxg	1	-
rpool	canmount	on	default
rpool	xattr	on	default
rpool	copies	1	default
rpool	version	5	-
rpool	utf8only	off	-
rpool	normalization	none	-
rpool	casesensitivity	sensitive	-
rpool	vscan	off	default
rpool	nbmand	off	default
rpool	sharesmb	off	default
rpool	refquota	none	default
rpool	refreservation	none	default
rpool	guid	3460066999470856580	-
rpool	primarycache	all	default
rpool	secondarycache	all	default
rpool	usedbysnapshots	0B	-
rpool	usedbydataset	104K	-
rpool	usedbychildren	93.1G	-
rpool	usedbyrefreservation	0B	-
rpool	logbias	latency	default
rpool	objsetid	54	-
rpool	dedup	off	default
rpool	mlslabel	none	default
rpool	sync	standard	local
rpool	dnodesize	legacy	default
rpool	refcompressratio	1.00x	-
rpool	written	104K	-
rpool	logicalused	128G	-
rpool	logicalreferenced	46K	-
rpool	volmode	default	default
rpool	filesystem_limit	none	default
rpool	snapshot_limit	none	default
rpool	filesystem_count	none	default
rpool	snapshot_count	none	default
rpool	snapdev	hidden	default
rpool	acltype	off	default
rpool	context	none	default
rpool	fscontext	none	default
rpool	defcontext	none	default
rpool	rootcontext	none	default
rpool	relatime	off	default
rpool	redundant_metadata	all	default
rpool	overlay	off	default
rpool	encryption	off	default
rpool	keylocation	none	default
rpool	keyformat	none	default
rpool	pbkdf2iters	0	default
rpool	special_small_blocks	0	default

Obr. 6: Parametry ZFS datasetu rpool [Vlastní zpracování]

3.2.2 **Údržba datových polí**

V případě, že bychom potřebovali vyměnit například vadný disk, tak můžeme použít příkaz „zfs replace pool device new-device“.

V mém návrhu řešení využívám automatické kontroly konzistence dat prostřednictvím příkazu „zfs scrub“ iniciovaném aplikací Cron pravidelně každou druhou neděli v měsíci.

Za účelem přehlednosti disků z přední strany serveru na každém HDD rámečku osadím štítek s identifikačními údaji jednotlivého disku ve formátu „Kapacita, Výrobce, poslední 4 znaky sériového čísla“. Příklad vypadá následovně „2TB TOSH F9XA“.

Pro větší přehlednost na datový rozvaděč umístím informační list k jednotlivým diskům v následujícím formátu. Hlavní důvod je lepší orientace pro čtení SMART dat dle knihovny cciss.

Tab. 3: Informační list disků [Vlastní zpracování]

SW pozice	SN	Značka	By-id alias	Kapacita	HW řádek	HW sloupec	Unix alias	ZFS pole
Cciss,1	ABCD	Seagate Ironwolf	Scsi-0xA…	2TB	2	4	sde	tank2

3.3 Zálohovací služba UrBackup

Pro zálohování klientských stanic v mém návrhu použiji software UrBackup. Zálohy provádím komprimované ve formátu .vhdx. Procesy zálohování jsou automatizované. V případě souborových záloh jsem nastavil inkrementální zálohy po 5ti dnech, a úplné po 30ti dnech. V případě záloh celých svazků jsem nastavil inkrementální zálohy na každý týden a úplné jednou za dva měsíce

Název počítače	Online	Naposledy spatřeno	Nejnovější souborová záloha	Nejnovější záloha celých svazků
PC1	Ano	22.01.22 02:06	03.01.22 00:07	02.01.22 19:10
PC2	Ne	07.01.22 14:10	Nikdy	07.01.22 13:31
PC3	Ne	16.01.22 22:15	Nikdy	Nikdy
PC4	Ne	15.01.22 16:41	Nikdy	08.01.22 01:01
PC5	Ne	03.01.22 16:08	Nikdy	03.01.22 02:09

Obr. 7: Soupis zařízení v zálohovacím softwaru UrBackup [Vlastní zpracování]

Jednotlivé zálohy můžeme spravovat pomocí webového rozhraní. V případě souborových záloh je možné konkrétní zálohu katalogově procházet. V případě záloh celých svazků tato možnost není.

Souborové zálohy

Čas zálohy	Přírůstková	Velikost
22.01.22 02:12	Ano	236.01 MB
03.01.22 00:07	Ano	197 MB
02.01.22 19:03	Ne	235.01 MB

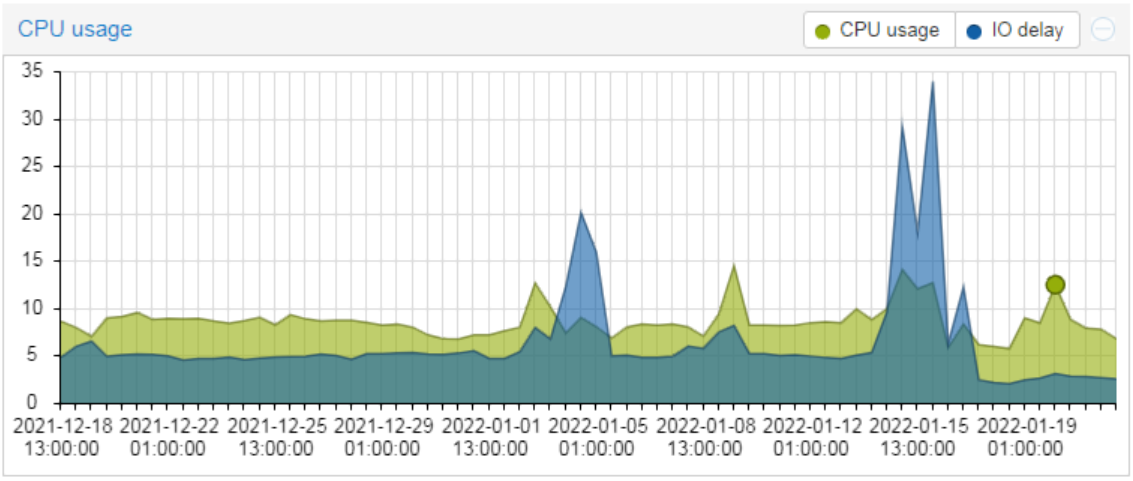
Zálohy celých svazků

Čas zálohy	Svazek	Přírůstková	Velikost
02.01.22 19:10	C:	Ne	106.79 GB
02.01.22 19:10	SYSVOL	Ne	4.47 MB

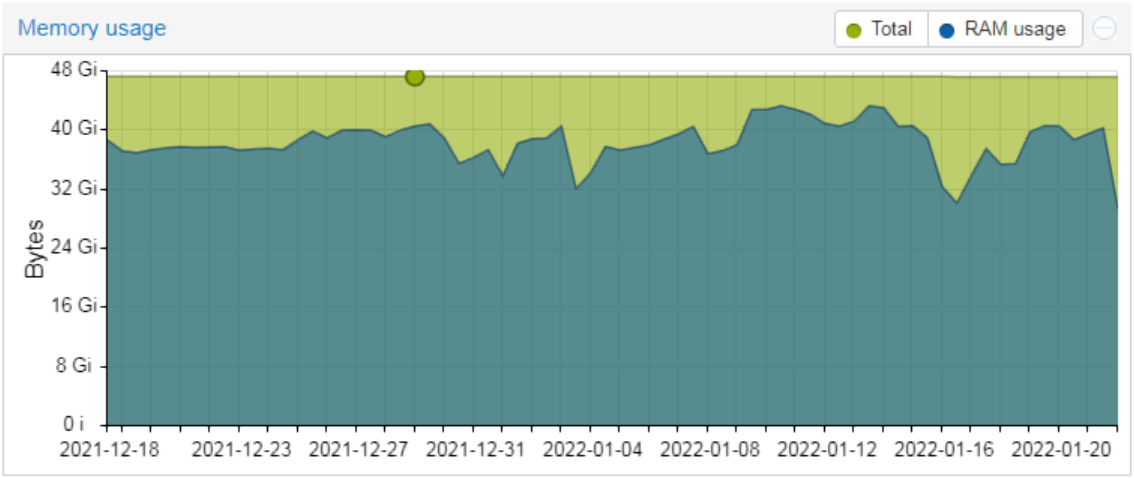
Obr. 8: Výčet jednotlivých záloh vybraného zařízení [Vlastní zpracování]

3.4 Statistiky využití

Ze statistik virtualizačního serveru je možné vidět, že reálné využití procesorů serveru za posledních 30 dní je nízké. Obsazená paměť RAM se jeví jako vysoká, nicméně ve virtuálních strojích je celková přiřazená paměť nevyužita ze sta procent.



Obr. 9: Využití CPU a IO zpoždění z Proxmox UI [Vlastní zpracování]

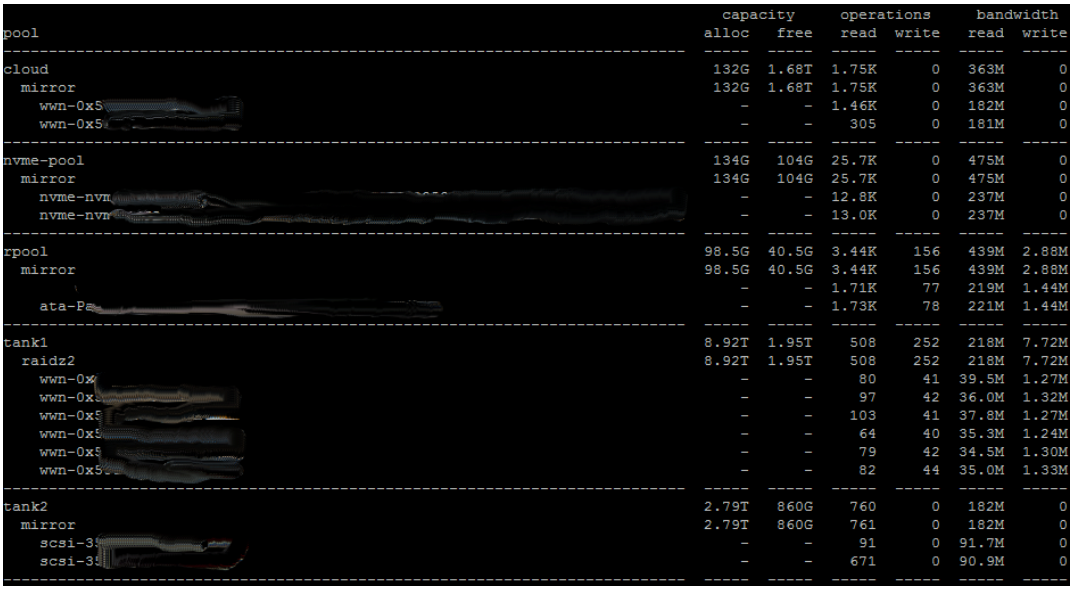


Obr. 10:Využití paměti RAM z Proxmox UI [Vlastní zpracování]

Jako příklad také uvádím rozdíly hodnot čtení a zápisu, jak ve formě datového toku za jednu sekundu, tak i počtu vstupně výstupních operací za jednu sekundu. Na obrázku č. 6 je vyobrazen datový provoz v období bez zátěže. Naproti tomu obrázek č. 7 ukazuje jak jsou datová pole vytížena v případě kontroly konzistence dat. Vysoká špička IO zpoždění je způsobena úlohou scrub, která se provádí pravidelně každý měsíc.

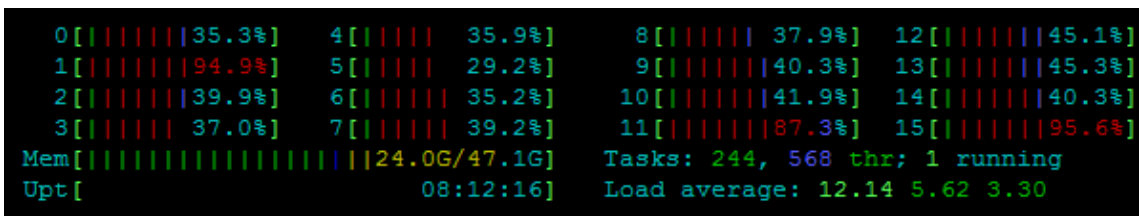
pool	capacity alloc	capacity free	operations read	operations write	bandwidth read	bandwidth write
cloud	132G	1.68T	0	0	1.21K	1.25K
mirror	132G	1.68T	0	0	1.21K	1.25K
wwn-0x5	-	-	0	0	542	641
wwn-0x5	-	-	0	0	698	641
nvme-pool	134G	104G	9	62	133K	1.12M
mirror	134G	104G	9	62	133K	1.12M
nvme-nvme.1	-	-	4	30	66.2K	572K
nvme-nvm	-	-	4	31	67.1K	572K
rpool	98.5G	40.5G	3	38	122K	1.02M
mirror	98.5G	40.5G	3	38	122K	1.02M
sdk3	-	-	1	18	61.4K	523K
ata-P	-	-	1	19	60.7K	523K
tank1	8.92T	1.95T	15	285	1.75M	10.4M
raidz2	8.92T	1.95T	15	285	1.75M	10.4M
wwn-0x5	-	-	2	46	362K	1.74M
wwn-0x5	-	-	2	49	263K	1.74M
wwn-0x5	-	-	2	46	299K	1.74M
wwn-0x5	-	-	2	48	359K	1.74M
wwn-0x5	-	-	2	46	234K	1.74M
wwn-	-	-	2	46	273K	1.74M
tank2	2.79T	860G	9	0	464K	15.7K
mirror	2.79T	860G	9	0	464K	15.7K
scsi-	-	-	4	0	230K	7.85K
scsi	-	-	4	0	234K	7.85K

Obr. 11: Zpool iostat -v 1 - klidový stav [Vlastní zpracování]



pool	capacity alloc	capacity free	operations read	operations write	bandwidth read	bandwidth write
cloud	132G	1.68T	1.75K	0	363M	0
mirror	132G	1.68T	1.75K	0	363M	0
wwn-0x5	-	-	1.46K	0	182M	0
wwn-0x5	-	-	305	0	181M	0
nvme-pool	134G	104G	25.7K	0	475M	0
mirror	134G	104G	25.7K	0	475M	0
nvme-nvm	-	-	12.8K	0	237M	0
nvme-nvm	-	-	13.0K	0	237M	0
rpool	98.5G	40.5G	3.44K	156	439M	2.88M
mirror	98.5G	40.5G	3.44K	156	439M	2.88M
	-	-	1.71K	77	219M	1.44M
ata-Pa	-	-	1.73K	78	221M	1.44M
tank1	8.92T	1.95T	508	252	218M	7.72M
raidz2	8.92T	1.95T	508	252	218M	7.72M
wwn-0x	-	-	80	41	39.5M	1.27M
wwn-0x5	-	-	97	42	36.0M	1.32M
wwn-0x5	-	-	103	41	37.8M	1.27M
wwn-0x5	-	-	64	40	35.3M	1.24M
wwn-0x5	-	-	79	42	34.5M	1.30M
wwn-0x5	-	-	82	44	35.0M	1.33M
tank2	2.79T	860G	760	0	182M	0
mirror	2.79T	860G	761	0	182M	0
scsi-3	-	-	91	0	91.7M	0
scsi-3	-	-	671	0	90.9M	0

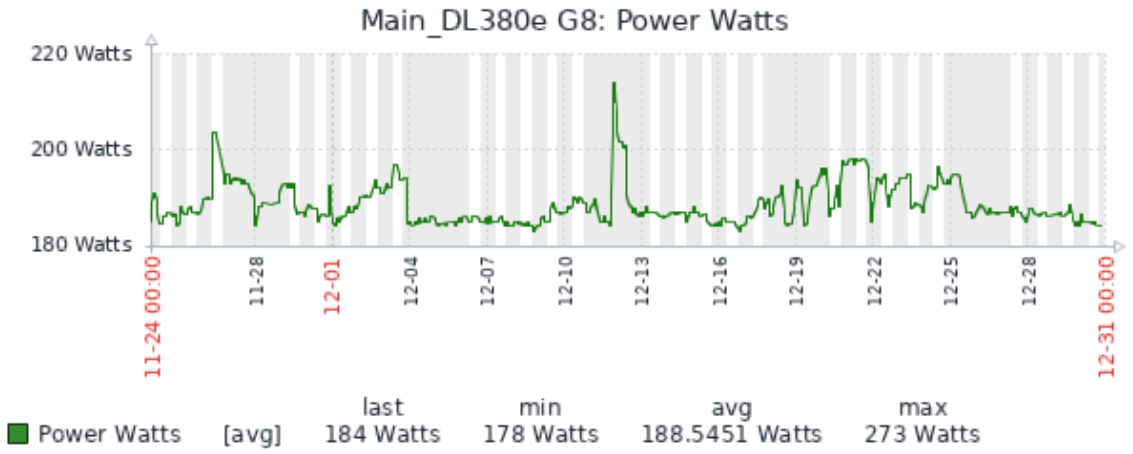
Obr. 12: Zpool iostat -v 1 - ZFS scrub proces [Vlastní zpracování]



Obr. 13: Využití výpočetního výkonu při ZFS scrub všech polí [Vlastní zpracování]

Jednotlivým virtuálním zařízením jsem nastavil maximální možné využití IOPS a datového toku. Hlavním důvodem je předejití stavu, kdy by jedno virtuální zařízení využilo 100% výkonu disků a nezbyl tak volný výpočetní výkon pro ostatní.

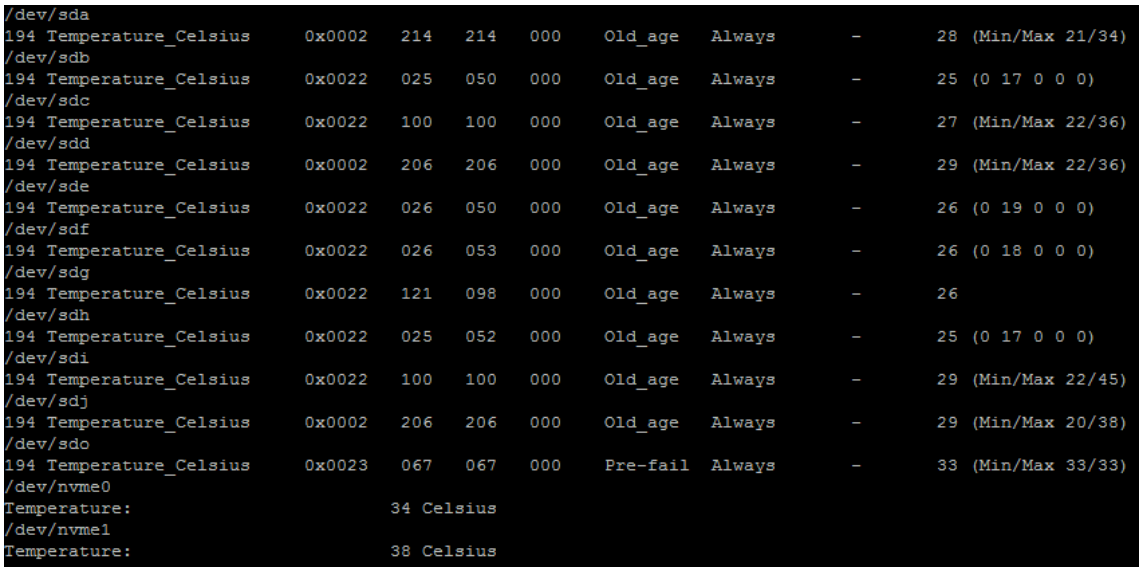
Na obrázku č. 10 je v Zabbixu vyobrazena okamžitá spotřeba serveru za poslední týden prosince 2021. Hodnoty jsou čteny přes SNMP protokol. K měření je použit samostatný měřič spotřeby.



Obr. 14: Graf spotřeby serveru v Zabbixu [Vlastní zpracování]

Z důvodu možných výpadků elektrické energie použiji UPS od společnosti APC v rackovém provedení Smart-UPS 1500VA. Aktuální stav UPSky je také monitorován Zabbixem a následné události jsou obratem oznamovány prostřednictvím mailové zprávy. Aktuální stav UPSky současně také monitoruje server prostřednictvím aplikace Apcupsd. V případě, že nastane výpadek elektrické energie z distribuční soustavy, tak server započne automaticky po 15ti minutách a zároveň s maximálně zbývajícími 15ti minutami možného provozu UPSky vypínací sekvenci, kdy korektním způsobem vypne jak všechny virtuální stroje, tak nakonec i sebe sama.

Pro dlouhou životnost disků je taky potřebné udržovat disky v rozumných teplotách. Tyto teploty jsou rovněž monitorovány a v případě výkyvu je správce informován. Níže na obrázku lze vidět jak takový výstup prostřednictvím jednoduchého skriptu může vypadat.



Obr. 15: Teploty disků ze SMART [Vlastní zpracování]

Informovanost správce serveru zprostředkovávám prostřednictvím automatizovaných emailů ze Zabbixu na vlastní mailový server, který bude prostřednictvím Push notifikací tyto emaily pravidelně stahovat do mobilního zařízení. Hlavní účel vlastního mailového serveru spatřuji v prevenci zahlcení SMTP protokolu častými zprávami do sítě typu Internet, což by mohlo následně vést k záznamu v SMTP blacklistu.

Jak může takový informační mail vypadat přikládám na následujícím obrázku:

Problem started at 13:26:43 on 2021.12.31
Problem name: More than 95% used on
dataset nvme-pool/subvol-1
proxmox
Host: proxmox
Severity: High
Operational data: 15.45 GB, 558.4 MB
Original problem ID: 427789

Obr. 16: Zabbix notifikace [Vlastní zpracování]

ARC cache jsem v mém návrhu zvolil s rozmezím 12-19,6GB. Systém si poté sám optimalizuje aktuální obsazenost. V případě nedostatku volné celkové paměti RAM systém ARC cache uvolní. Dále z obrázku č. 11 můžeme vidět vysokou efektivitu ARC cache v řádku Cache hit ratio s hodnotou 98,3%. Tento údaj nám zároveň dává informaci, že by systém a datová uložiště dokázali, efektivně využít i více než mnou nastavených maximálních 19,6GB.

```
ARC status:                                                      HEALTHY
        Memory throttle count:                                         0

ARC size (current):                                    99.5 %   19.5 GiB
        Target size (adaptive):                       100.0 %   19.6 GiB
        Min size (hard limit):                         61.2 %   12.0 GiB
        Max size (high water):                            1:1   19.6 GiB
        Most Frequently Used (MFU) cache size:         44.2 %    7.8 GiB
        Most Recently Used (MRU) cache size:           55.8 %    9.8 GiB
        Metadata cache size (hard limit):              75.0 %   14.7 GiB
        Metadata cache size (current):                 21.5 %    3.2 GiB
        Dnode cache size (hard limit):                 10.0 %    1.5 GiB
        Dnode cache size (current):                    46.7 %  704.2 MiB

ARC hash breakdown:
        Elements max:                                            3.9M
        Elements current:                              58.9 %    2.3M
        Collisions:                                            231.3M
        Chain max:                                                7
        Chains:                                                267.9k

ARC misc:
        Deleted:                                                16.0M
        Mutex misses:                                            2.0k
        Eviction skips:                                        119.4k

ARC total accesses (hits + misses):                              36.9G
        Cache hit ratio:                               98.3 %    36.3G
        Cache miss ratio:                               1.7 %   630.1M
        Actual hit ratio (MFU + MRU hits):             98.3 %    36.3G
        Data demand efficiency:                        95.8 %    15.0G
        Data prefetch efficiency:                      61.1 %     6.9M

Cache hits by cache type:
        Most frequently used (MFU):                    98.6 %    35.8G
        Most recently used (MRU):                       1.4 %   490.6M
        Most frequently used (MFU) ghost:             < 0.1 %     2.2M
        Most recently used (MRU) ghost:               < 0.1 %     6.6M

Cache hits by data type:
        Demand data:                                   39.6 %    14.3G
        Demand prefetch data:                         < 0.1 %     4.2M
        Demand metadata:                               60.4 %    21.9G
        Demand prefetch metadata:                     < 0.1 %     9.5M

Cache misses by data type:
        Demand data:                                   99.3 %   625.4M
        Demand prefetch data:                           0.4 %     2.7M
        Demand metadata:                                0.1 %   768.3k
        Demand prefetch metadata:                       0.2 %     1.2M

DMU prefetch efficiency:                                         3.8G
        Hit ratio:                                      1.0 %    36.3M
        Miss ratio:                                    99.0 %     3.7G

L2ARC not detected, skipping section
```

Obr. 17: Souhrn využití paměti ARC [Vlastní zpracování]

Záloha ZFS datového pole je prováděna manuálně na samostatný disk, který je běžně uložený na jiném místě. Tuto zálohu provádím jednou za 3 měsíce.

K tomuto využívám aplikaci Syncoid, která dokáže naplno využít výhod ZFS, včetně snapshotů. Konkrétně tedy spuštěním programu syncoid se provedou obrazy souborového systému včetně jeho podpolí a tato data se obratem zálohují na cílový disk. Výhodou také je, že syncoid podporuje inkrementální zálohy těchto obrazů.

Manuální metodu zálohování jsem zvolil z důvodu, jestliže by data na ZFS datovém poli byla vadná, tak by se v případě automatické zálohy prostřednictvím skriptu záloha přepsala těmito vadnými daty. Z důvodu ochrany dat proto navrhuji dělat zálohy manuální.

```
user996@pve:~# syncoid tank1/Jamendo cloud/test
INFO: Sending oldest full snapshot tank1/Jamendo@production220114 (~ 94.5 MB) to new target filesystem:
94.8MiB 0:00:00 [ 107MiB/s] [=============================================================>] 100%
INFO: Updating new target filesystem with incremental tank1/Jamendo@production220114 ... syncoid_pve_2022-01-
22:20:32:58-GMT01:00 (~ 5 KB):
2.13KiB 0:00:02 [ 826 B/s] [=========================>                                      ] 40%
```

Obr. 18: Záloha datasetu prostřednictvím Syncoid [Vlastní zpracování]

Jako cloud řešení jsem zvolil open-source platformu Nextcloud. Jeho hlavním účelem bude synchronizace dokumentů, fotografií a ostatních souborů z různých zařízení na server.

3.5 Metodika měření výkonu datového pole nvme-pool

Pro měření výkonu byla použita aplikace fio-plot která využívá aplikaci fio. Měření bylo prováděno na hostu s operačním systémem Proxmox v 7.1-10, linuxovém jádře 5.13.19-2 a knihovně ZFS ve verzi 2.1.2. Použitá verze fio je 2.25.

Měření jsem prováděl prostřednictvím následujícího příkazu: „bench-fio -d /nvme-pool/fio -t directory -s 2G -b 4k 4M --iodepth 1 32 --numjobs 1 16 --mode whereis{read|write|randread|randwrite} --extra-opts log_avg_msec=1000 -o /nvme-pool/bench_fio/ --time-based“

Pro měření bylo použito datového pole zrcadleného ze dvou NVMe disků Transcend TS256GMTE110S zaplněného ze 42%, kdy výrobce uvádí, že jednotlivý disk má dosahovat maximálních rychlostí až 1600MB/s pro čtení a až 1100MB/s pro zápis. Dále výrobce uvádí, že při použití 4KB bloků má disk dosahovat až 90 tis. IOPS pro čtení a až 250 tis. IOPS pro zápis. Pro měření výrobce uvádí, že použil aplikace IOmeter a CrystalDiskMark.

Jelikož je ARC cache nakonfigurována na hodnotu 12-19GB, tak provedu měření o celkovém objemu dat 32GB na každý průchod, aby se do jisté míry snížil vliv mezipaměti na výsledky měření. Na datovém poli je nastavena koprimace dat v režimu LZ4.

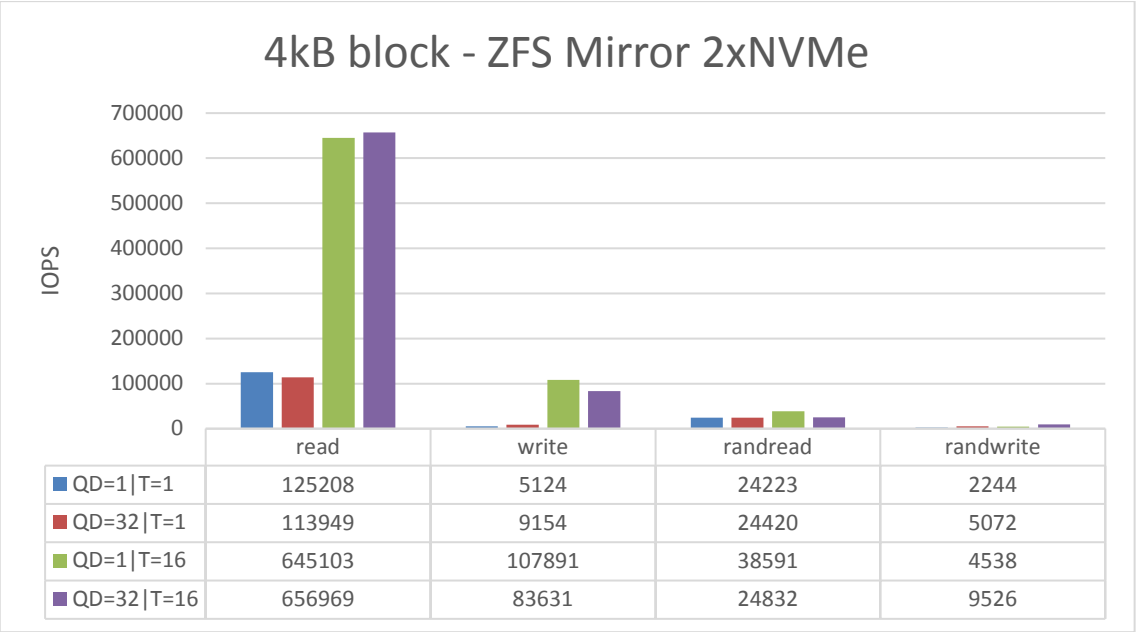
3.5.1 **Výsledky měření**

Následující grafy jsou vyobrazením střední hodnoty z měření o délce času v hodnotě 60 sekund. V případě měřených 4kB bloků je hlavním sledovaným parametrem IOPS. Čím je vyšší hodnota IOPS, tím lepší reálný výkon aplikace, např. při použití databáze. U 4MB bloků zase hodnotíme maximální možný datový tok. Typickým příkladem použití může být kopírování nebo zpracovávání větších souborů, kdy služba nebo aplikace nemusí zpracovávat jednotlivá metadata souboru, ale zařizuje přenos nebo jinak interaguje přímo pouze s daty.

V případě velikostí bloků 4KB lze výsledky z obrázku 16 a 17 interpretovat tak, že datové pole dosahuje nejvyšších hodnot při sekvenčním čtení a za použití 16ti vláken. Jelikož hodnoty sekvenčního čtení v případě 16ti vláknového přístupu převyšují dvojnásobek výrobcem uváděného maxima, lze předpokládat že hodnoty jsou ovlivněny prostřednictvím ARC mezipaměť v paměti RAM. Při náhodném čtení jsou měřené hodnoty mnohem nižší, než výrobce uvádí pro jednotlivý disk. Hodnoty náhodného čtení jsou výrazně nižší, než v případě sekvenčního čtení. V tomto případě bych hodnotil hodnoty jako přiměřeně akceptovatelné. V případě zápisu jsou hodnoty velmi nízké.

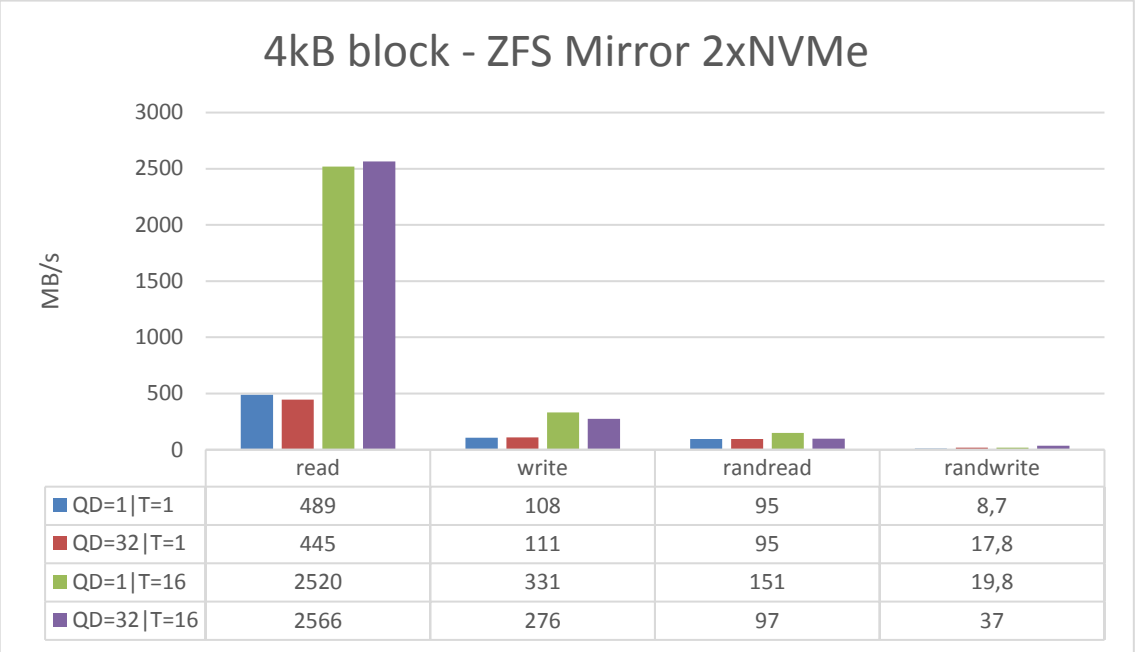
Tyto hodnoty jsou zapříčiněny tím, že recordsize datasetu nvme-pool je 128kB a mnou prováděné měření má velikost bloku 4kB. Z toho vyplývá, že velikost bloku je 1/32 nastaveného recordsize. Ve výsledku při takovémto zápisu se zapíše námi požadovaných 4kB dat a zbytek disk musí souborový systém musí doplnit do 128kB. Znamená to tedy, že námi zjištěné výsledky měření jsou přibližně 1/32 toho jaké údaje by byly při optimální konfiguraci recordsize. Při recordsize=4kB by měly tedy výsledky těchto hodnot zápisů dosahovat přibližně 32 násobku.

Jelikož je toto datové pole využíváno výhradně pro virtuální stroje, které jsou optimalizované pro 128kB bloky, tak konfiguraci recordsize ponechávám na hodnotě 128kB.



	read	write	randread	randwrite
QD=1|T=1	125208	5124	24223	2244
QD=32|T=1	113949	9154	24420	5072
QD=1|T=16	645103	107891	38591	4538
QD=32|T=16	656969	83631	24832	9526

Obr. 19: Měření IOPS - 4kB bloky [Vlastní zpracování]

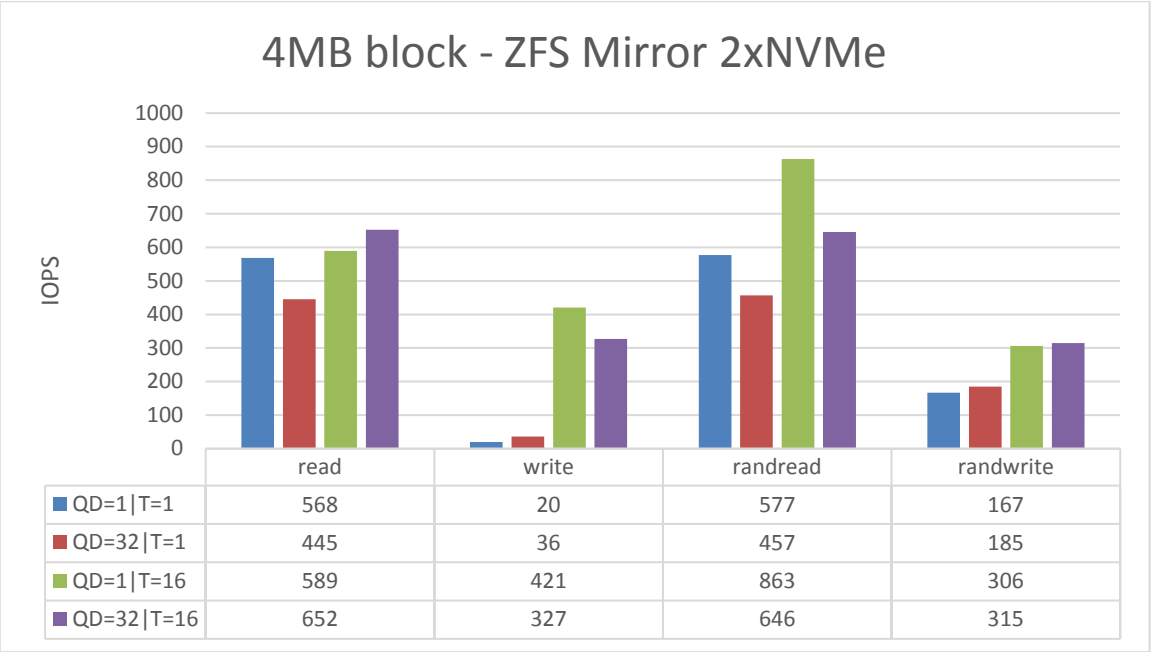


	read	write	randread	randwrite
QD=1|T=1	489	108	95	8,7
QD=32|T=1	445	111	95	17,8
QD=1|T=16	2520	331	151	19,8
QD=32|T=16	2566	276	97	37

Obr. 20: Měření datového toku - 4kB bloky [Vlastní zpracování]

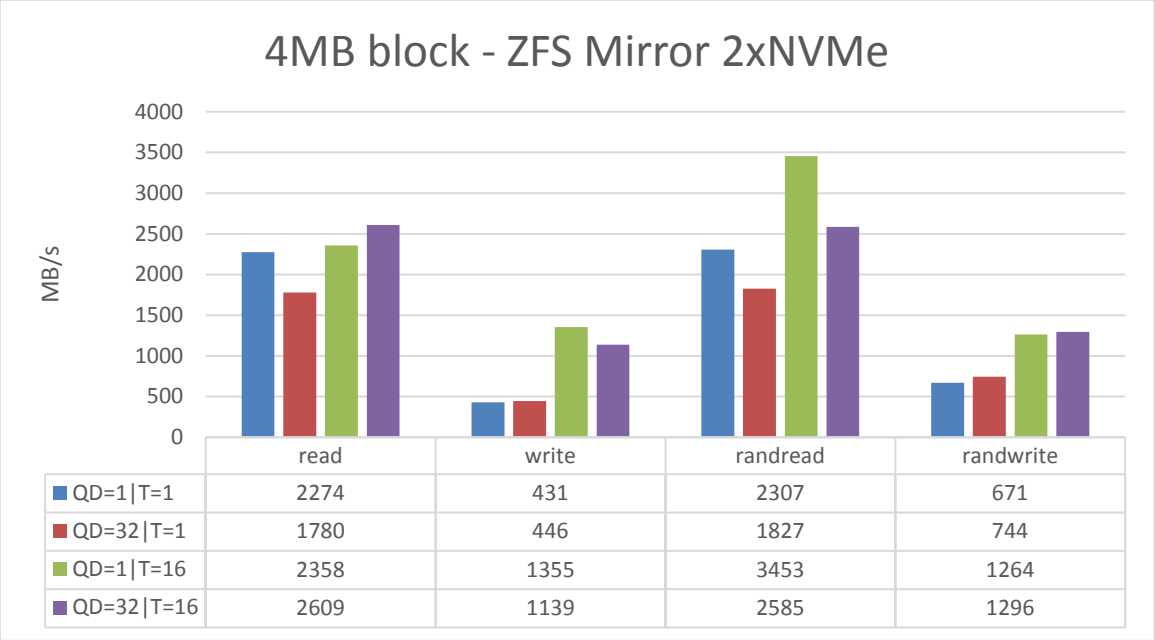
Při měření 4MB bloků, na obrázku 19 a 20, již dosahujeme lepších výsledků, protože nejsme ovlivněni menším měřeným blokem než je recordsize datasetu. Změřené hodnoty čtení by

teoreticky měly dosahovat až dvojnásobku výrobcem uváděných rychlostí. Změřené hodnoty dvou disků jsou oproti maximu výrobcem uvedené na jeden disk v hodnotách 105-203%. Tyto hodnoty hodnotím jako velmi dobré. Nicméně stále z výsledků měření nejsme schopni vyčíst nakolik má podíl ARC cache na výsledných hodnotách. V případě zápisu by výsledné hodnoty měly dosahovat maximálně jedno-násobku výrobcem uváděných hodnot z důvodu zvoleného režimu zrcadlení. V tomto případě jsou výsledky taky dobré. Dále je z výsledků patrné, že vyšších rychlostí se dosahuje při více-vláknovém použití. Pro účely virtualizace jsou tyto výsledky optimální. Pro jedno vláknovou aplikaci bychom tedy mohli sledovat nižší výkon pro zápis.



	read	write	randread	randwrite
QD=1|T=1	568	20	577	167
QD=32|T=1	445	36	457	185
QD=1|T=16	589	421	863	306
QD=32|T=16	652	327	646	315

Obr. 21: Měření IOPS - 4MB bloky [Vlastní zpracování]



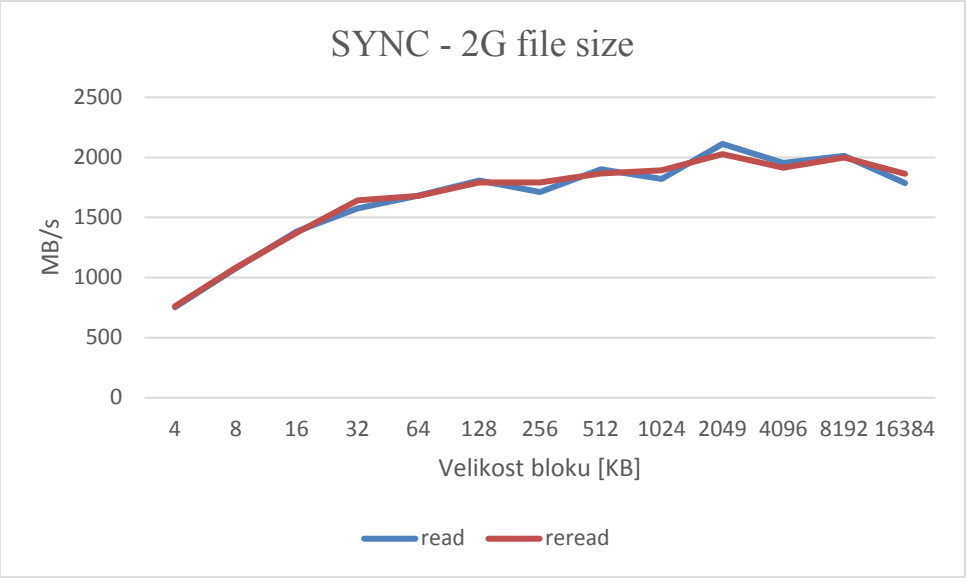
	read	write	randread	randwrite
QD=1|T=1	2274	431	2307	671
QD=32|T=1	1780	446	1827	744
QD=1|T=16	2358	1355	3453	1264
QD=32|T=16	2609	1139	2585	1296

Obr. 22: Měření datového toku - 4MB bloky [Vlastní zpracování]

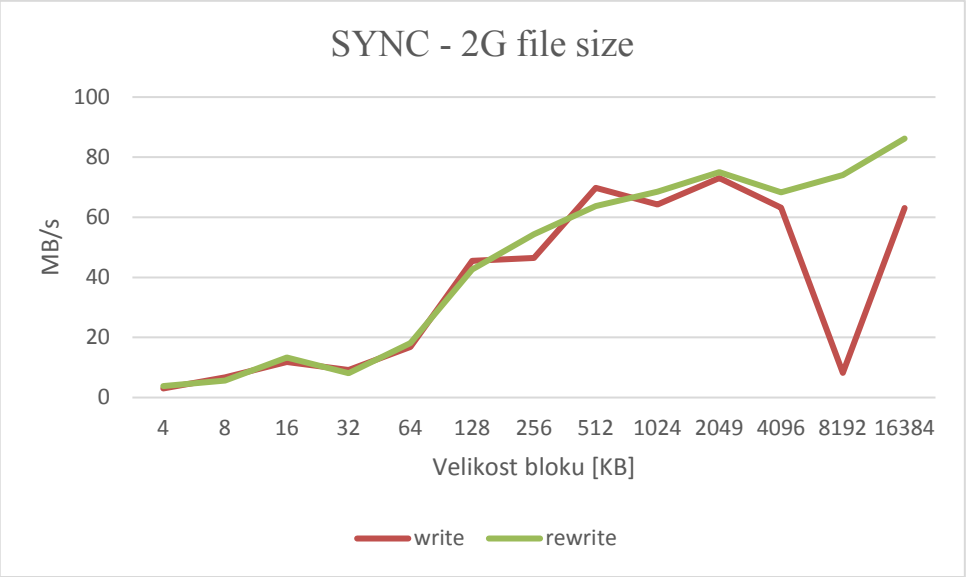
3.6 Vliv velikosti bloku na rychlost

Druhým měřením jsem chtěl vyobrazit rozdíly rychlostních hodnot v závislosti na tom jaká je zvolená velikost bloku. Prostřednictvím aplikace Iozone jsem změřil hodnoty synchronního čtení a zápisu, tak abych v maximální možné míře změřil hodnoty disků a nikoliv mezipaměti ARC.

Měření aplikací Iozone bylo provedeno s následujícími paramatery: *„iozone -a {-o} -s 2G -i 0 -i 1 -p -f /nvme-pool/iozone/test“*. Synchronní měření bylo prováděno parametrem „-o“ a asynchronní bez tohoto paramatru.

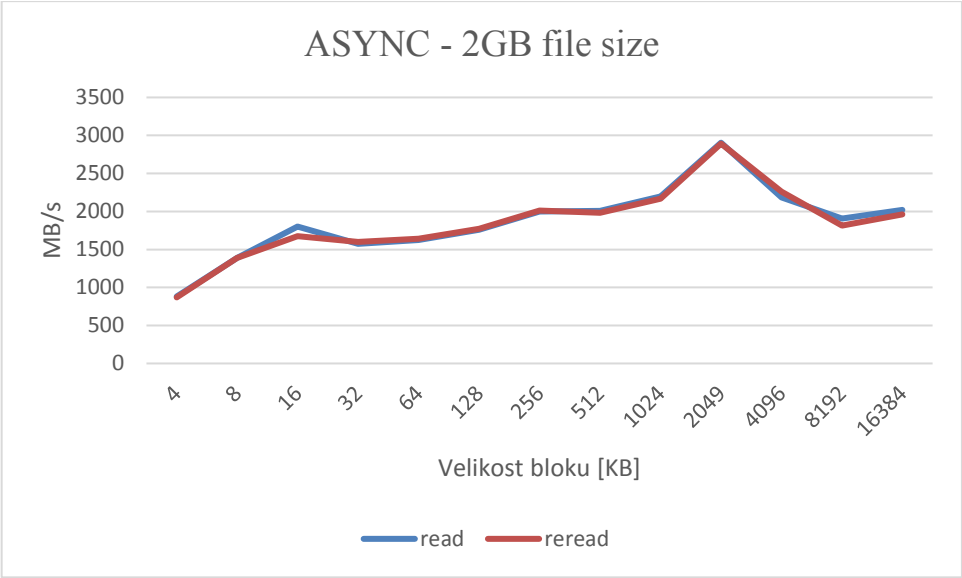


Obr. 23: Iozone rychlost čtení v závislosti na velikosti bloku - synchronní [Vlastní zpracování]

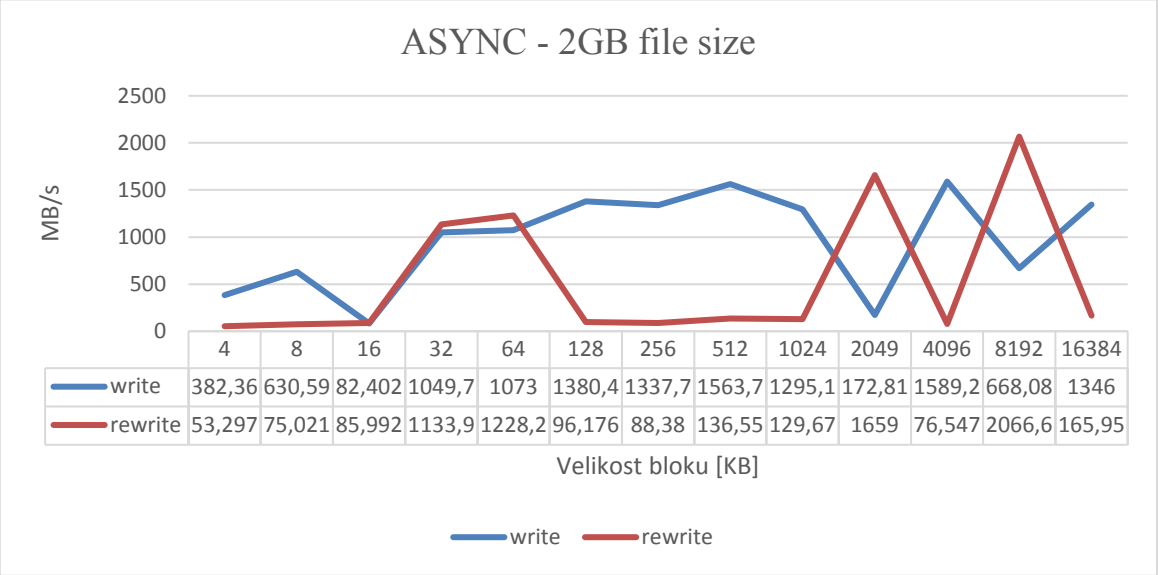


Obr. 24: Iozone rychlost zápisu v závislosti na velikosti bloku - synchronní [Vlastní zpracování]

Z důvodu úplnosti jsem změřil jak synchronní, tak i asynchronní čtení a zápis dat, aby byl vidět rozdíl hodnot, na který má vliv ARC cache.



Obr. 25: Iozone rychlost čtení v závislosti na velikosti bloku - asynchronní [Vlastní zpracování]



Obr. 26: Iozone rychlost zápisu v závislosti na velikosti bloku - asynchronní [Vlastní zpracování]

Z výše uvedených výsledků je možné vidět při asynchronním měření navýšení rychlosti čtení a to hlavně v oblasti 2MB velikosti bloku o 43% na 2,9GB/s namísto 2GB/s v případě synchronního měření. Výrazné navýšení lze vidět také na obrázku 26 asynchronního zápisu. Nicméně je zajímavé, že v případě asynchronního zápisu se vyskytují již velké výkyvy hodnot.

Pro vyšší hodnoty datového toku synchronního zápisu by bylo vhodné do systému připojit zařízení typu ZIL/SLOG.

ZÁVĚR

V rámci této bakalářské práce je na základě zadání návrh modelu pro zálohování a datové uložiště ve společnosti Gamitech, s.r.o. V analýze současného stavu je popsán původní stav společnosti, dále technického a softwarového vybavení společnosti. V návrhu vlastního řešení jsem navrhnul funkční model datového uložiště na systému ZFS a zálohovacího řešení UrBackup. Jako výsledek jsem provedl sadu měření, pro zjištění zda by mohl tento návrh být nasazen do funkčního provozu.

SEZNAM POUŽITÝCH ZDROJŮ

(1) ZFS file systems: Encryption. *Oracle* [online]. USA: Oracle, 2012 [cit. 2021-11-10]. Dostupné z: https://docs.oracle.com/cd/E23824_01/html/821-1448/gkkih.html

(2) JUDE, Allan a Michael W. LUCAS. Sysadmin: Tuning OpenZFS. *Sysadmin* [online]. 2016, 11.2016, , 41-4 [cit. 2021-11-10]. Dostupné z: https://www.usenix.org/system/files/login/articles/login_winter16_09_jude.pdf

(3) DEVAINE, Max. Stavba a zkušenosti se ZFS storage. *ZFS* [online]. 2020, 18.5.2020, **2020**(1), 6 [cit. 2021-10-29]. Dostupné z: https://www.abclinuxu.cz/blog/Max_Devaine/2019/9/zfs-stavba-zkusenosti-se-zfs-storage

(4) TOPONCE, Aaron. ZFS Administration: Copy-on-write. *ZFS* [online]. 2012, 14.12.2012, **2012**, 1 [cit. 2021-11-10]. Dostupné z: https://pthree.org/2012/12/14/zfs-administration-part-ix-copy-on-write/

(5) CHUNLU WANG, YUANYUAN WU, ZHANYE WANG a TAO XU. ISCSI-based data protection system for virtual machine. *Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)* [online]. IEEE, 2013, 2013, , 2085-2089 [cit. 2021-11-17]. ISBN 978-1-4799-2565-0. Dostupné z: doi:10.1109/MEC.2013.6885394

(6) LOESER, Henrik. *Manage Your Cloud Object Storage Data with the MinIO Client and rclone: Simple access to your S3-based data on IBM Cloud* [online]. 18.11.2021, , 1 [cit. 2021-11-23]. Dostupné z: https://www.ibm.com/cloud/blog/manage-your-cloud-object-storage-data-with-the-minio-client-and-rclone

(7) FONG, Yinfung a Stephen MANLEY. *Efficient true image recovery of data from full, differential, and incremental backups*. 2004. Spojené státy americké. US007251749. Uděleno 31.07.2007. Zapsáno 12.02.2004

(8) VUPPALA, Sai Pranav. Focus: Backing Up Your Data with UrBackup. *Open Source for You* [online]. India, New Dehli, 2021, (2021) [cit. 2021-11-23]. ISSN 2456-4885. Dostupné z: https://www.proquest.com/docview/2584750273/fulltext/ACDC86846AEC4D30PQ/9

(9) RUGGIERO, Paul a Matthew A. HECKATHORN. Data Backup Options. *United States Computer Emergency Readiness Team* [online]. USA [cit. 2021-11-24]. Dostupné z: https://us-cert.cisa.gov/sites/default/files/publications/data_backup_options.pdf

(10) SHARMA, Shashank. File access with ACLs. *Linux Format* [online]. Bath, UK: Future Publishing, 2019, **2019**(246), 62-63 [cit. 2021-11-24]. Dostupné z: https://www.proquest.com/magazines/file-access-with-acls/docview/2166932601/se-2?accountid=17115

(11) MEHROTRA, Neetesh. Admin - Overview: An Introduction to ZABBIX. *Open Source for You* [online]. New Dehli, 2017, 01.03.2017, , 3 [cit. 2021-11-24]. Dostupné z: https://www.proquest.com/magazines/admin-overview-introduction-zabbix/docview/1877865423/se-2?accountid=17115

(12) RASMUSSEN, Neil. White Paper #1. *The Different Types of UPS Systems* [online]. 2004, , 9 [cit. 2021-11-24]. Dostupné z: https://www.apcdistributors.com/white-papers/Power/WP-1%20The%20Different%20Types%20of%20UPS%20Systems.pdf

(13) BORGES, G, S CROSBY a L BOLAND. CephFS: a new generation storage platform for Australian high energy physics. *Journal of Physics: Conference Series*. 2017, **898**. ISSN 1742-6588. Dostupné z: doi:10.1088/1742-6596/898/6/062015

(14) WOJSŁAW, Damian. *Introducing ZFS on Linux: Understand the Basics of Storage with ZFS*. Berkeley, CA: Apress, 2017. ISBN 978-1-4842-3305-4.

(15) KHAWATREH, Saleh a El-Omari NIDHAL. *RAID-based Storage Systems*. 2018. International Journal of Computer Applications, 7 s. 180. Dostupné z: https://www.researchgate.net/publication/344324442_RAID-based_Storage_Systems

(16) FISCHER, Werner. Battery Backup Unit (BBU/BBM) Maintenance for RAID Controllers. *Thomas Krenn* [online]. 2013 [cit. 2021-11-24]. Dostupné z: https://www.thomas-krenn.com/en/wiki/Battery_Backup_Unit_(BBU/BBM)_Maintenance_for_RAID_Controllers

(17) VERNOOIJ, Jelmer R., John H. TERPSTRA a Gerald CARTER. *The Official Samba-3 HOWTO and Reference Guide* [online]. 2005, 589 s. [cit. 2021-11-24]. Dostupné z: https://www.researchgate.net/publication/228392680_The_official_Samba-3_HOWTO_and_reference_guide

(18) Pokladní systém Conto. *Consulta* [online]. Vyškov, ©2022 [cit. 2022-03-08]. Dostupné z: http://www.consulta.cz/editor/image/vkladane_soubory/pic/contoscr/contomain_ori.jpg

(19) Účetní program Pohoda. *Stormware* [online]. Jihlava, ©2022 [cit. 2022-03-08]. Dostupné z: https://www.stormware.cz/image/page/responsive/pohoda/ImgScreenPOHODA.png

(20) Creating a ZFS Storage Pool [online]. @2022 [cit. 2022-03-08]. Dostupné z: https://docs.oracle.com/cd/E19253-01/819-5461/gaynr/index.html

SEZNAM OBRÁZKŮ

SEZNAM TABULEK