VIŠJA STROKOVNA ŠOLA ACADEMIA

MARIBOR

IMPLEMENTACIJA SISTEMA TRUENAS ZA CELOVITO UPRAVLJANJE S PODATKI

Kandidat: David Mihalič

Vrsta študija: študent izrednega študija

Študijski program: programsko inženirstvo

Mentor predavatelj: mag. Dušan Brglez

Mentor v podjetju: Aleš Hribar Bukovič

Lektorica: Mija Čuk, univ. dipl. spl. jez.

Maribor, 2024

IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Podpisani David Mihalič sem avtor diplomskega dela z naslovom Implementacija sistema TrueNAS za celovito upravljanje s podatki, ki sem ga napisal pod mentorstvom mag. Dušana Brgleza.

S svojim podpisom zagotavljam, da:

- je predloženo delo izključno rezultat mojega dela,
- sem poskrbel, da so dela in mnenja drugih avtorjev, ki jih uporabljam v predloženi nalogi, navedena oz. citirana skladno s pravili Višje strokovne šole Academia Maribor,
- se zavedam, da je plagiatorstvo – predstavljanje tujih del oz. misli kot moje lastne kaznivo po Zakonu o avtorski in sorodnih pravicah (Uradni list RS, št. 16/07 – uradno prečiščeno besedilo, 68/08, 110/13, 56/15 in 63/16 – ZKUASP); prekršek pa podleže tudi ukrepom Višje strokovne šole Academia Maribor skladno z njenimi pravili,
- skladno z 32.a členom ZASP dovoljujem Višji strokovni šoli Academia Maribor objavo diplomskega dela na spletnem portalu šole.

Kamnik, januar 2024 Podpis študenta:

ZAHVALA

Iskreno se zahvaljujem mentorju in predavatelju mag. Dušanu Brglezu za strokovno pomoč in napotke oz. usmeritve pri nastajanju diplomskega dela, s pomočjo katerih sem lahko konkretno dosegel ustrezno in potrebno podlago za pisanje diplomskega dela.

Hvala tudi mentorju v podjetju Alešu Hribarju Bukoviču, ki mi je kot vodja oddelka omogočil lažje in hitrejše zaključevanje potrebnih aktivnosti za namen diplomskega dela z dodatno potrebnimi odsotnostmi in prilagoditvami delovnega urnika.

Zahvala gre tudi Ani, vsem bližnjim družinskim članom in prijateljem, ki so me v času celotne poti študija podpirali in ob težjih trenutkih spodbujali, da sem lahko vse študijske aktivnosti uspešno zaključil.

POVZETEK

Diplomsko delo temelji na implementaciji sistema za doseganje celovitega upravljanja s podatki, kar predstavlja različne izzive novodobnim organizacijam. Celovito upravljanje s podatki predstavlja sistem, ki ga lahko porazdelimo na osem pomembnih komponent, ki vsaka na svoj način definira svojo raven upravljanja. Diplomsko delo je razdeljeno na teoretični del, implementacijo sistema in njegovo analizo delovanja.

Teoretični del predvsem zajema vse potrebne ključne gradnike ali komponente, ki sestavljajo skupno sožitje za doseganje potrebne ravni celovitosti upravljanja s podatki. V sklopu komponent smo definirali glavne vrste arhitekture ter koncepte in metode modeliranja podatkov. Izpostavili smo načine administracije in integracije podatkov ter doseganja njihove ustrezne kakovosti. Velik poudarek je temeljil na komponenti varnosti, ki posameznikom in organizacijam predstavlja vse večji izziv pri nenehnem prilagajanju novim potencialnim varnostnim grožnjam njihovih podatkov. Vse skupaj se poveže znotraj komponente nadzora upravljanja podatkov, ki na podlagi izpostavljenih komponent določa politike in postopke upravljanja podatkov.

V drugem delu diplomskega dela smo se poglobili v celovit postopek implementacije platforme TrueNAS za doseganje zastavljenih potreb želene ravni celovitosti upravljanja s podatki. Izpostavili smo glavne informacije o izbrani platformi ter njene glavne funkcionalnosti in prednosti. Po posameznih korakih smo definirali želeno konfiguracijo sistema. Kreirali smo glavni podatkovni prostor, katerega smo porazdelili na različne nabore podatkov. Nastavili smo vse datotečne uporabniške pravice za kontroliranje dostopov do določenih podatkov, kot tudi različne aktivnosti arhiviranja in varovanja podatkov in sistema.

Analiza implementiranega sistema zajema namenske praktične primere, ki temeljijo na testiranju delovanja implementiranega sistema. Za ta namen se je pridobilo javno dostopne podatke velikih količin, katerih vsebina se ni razkrivala in s katerimi smo izvajali različne potrebne aktivnosti za preverjanje celovitosti upravljanja s podatki.

Ugotovitve diplomskega dela na podlagi teorije, implementacije sistema in njegove analize so predstavile pozitivno raven rezultatov, ki so potrdili možnost implementacije celovitega sistema za upravljanje s podatki.

Ključne besede: *podatki, upravljanje, hramba, varnost, analiza, TrueNAS.*

ABSTRACT

Implementation of the TrueNAS system for comprehensive data management

The thesis is based on the implementation of comprehensive data management system, which poses various challenges to modern organizations. Comprehensive data management is a system that can be divided into eight major components, each defining its own level of governance. The thesis is divided into a theoretical part, the implementation of the system and its operational analysis.

The theoretical part mainly covers all the necessary key constructors or components that make up the overall coexistence to achieve the required level of comprehensive data management. Within the components, we have defined the main types of architecture, and the concepts and methods od data modeling. We have outlined the ways in which data can be administered, integrated and of achieving appropriate data quality. A major focus was laid on the security component, which presents an increasing challenge for individuals and organizations to continuously adapt to new potential security threats to their data. All this linked together within the data governance component, which defines data management policies and procedures based on the outlined components.

In the second part of the thesis, we delved into the overall process of implementing the TrueNAS platform to achieve the desired level of comprehensive data management. We highlighted the main information about the selected platform and its main functionalities and advantages. Step by step, we defined the desired system configuration. We created a master data pool, which we divided into different datasets. We set up all the file user right to control access to specific data, as well as the various archiving and protection activities for data and the system.

The analysis of the implemented system includes dedicated practical examples based on the operational testing of the implemented system. For this purpose, large volumes of publicly available data, the contents of which were not disclosed, were obtained, and used to carry out the various activities necessary for verifying the comprehensive management of data.

The findings of the thesis, based on the theory itself, the implementation of the system and its analysis, presented a positive set of results that confirmed the possibility of implementing a comprehensive data management system.

Keywords: *data, management, storage, security, analysis, TrueNAS*

KAZALO VSEBINE

KAZALO SLIK

1 UVOD

Upravljanje s podatki je del vsakega posameznika, predvsem pa organizacij, ki se morajo vsakodnevno srečevati in izkoristiti velike količine podatkov. Tovrstne količine podatkov organizacijam doprinesejo pomembne vire informacij, ki jih je treba ustrezno upravljati, kar posledično terja dragocen čas za shranjevanje, arhiviranje, obdelavo ali analizo, njihovo varnost, v končni obliki pa tudi njihovo ustrezno uporabo. Tovrstno upravljanje je ključno, da podatki ne postanejo neuporabni ali poškodovani, kar lahko vodi do slabšega poslovanja organizacije in nedoseganja želenih ciljev.

Podatki se vse bolj razširjeno tretirajo kot oblika podatkovnega sredstva, katerega se uporabi za sprejemanje bolj osredotočenih poslovnih odločitev, ki temeljijo na izboljšanju ali optimizaciji poslovanja, doseganju zastavljenih poslovnih ciljev ali rezultatov in na izboljšanju konkurenčnosti na trgu. Pomanjkanje ustreznih sistemov in postopkov za upravljanje podatkov lahko privede organizacije do težav s kakovostjo podatkov, nekompatibilnosti med njimi in nedoslednosti naborov podatkov, ki potem posledično okrnijo možnosti uporabe poslovne inteligence (BI) ter drugih analitičnih aplikacij, kar lahko usmeri organizacijo tudi do napačnih odločitev in ugotovitev.

V današnjem času se veliko organizacij sooča z izzivi obvladovanja lastnih količin podatkov, ki jih niso zmožne optimalno izkoristiti sebi v prid. Rezultati tovrstnih izzivov ne izkoristijo in ne dosežejo želene ravni uporabe podatkov, zaradi nesmiselnih in neuporabnih oblik shranjevanja in uporabe podatkov znotraj različnih baz podatkov.

Upravljanje s podatki tako doprinese organizacijam neke vrste sistem, ki rešuje problem neizkoriščenih podatkov oziroma izboljša njihovo izkoriščenost in izlušči sam potencial uporabe teh podatkov. Natanko to predstavlja izziv in cilj našega raziskovalnega dela, da poskušamo implementirati celovit sistem upravljanja podatkov z uporabo izbranega operacijskega sistema, imenovanega TrueNAS. TrueNAS v osnovi prinaša operacijski sistem za strežnik NAS (angl. Network-attachhed storage), vendar omogoča veliko širšo uporabo, v katero se bomo v nadaljevanju te raziskovalne naloge poglobili, kot tudi v sam uporabljen operacijski sistem.

Upravljanje s podatki je v slovenskem jeziku zelo širok pojem, ki definira številne različne metode in koncepte za upravljanje s podatki.

V angleškem jeziku lahko z enakim prevodom razdelimo upravljanje s podatki na dva glavna aspekta:

- angl. data governance, ki predvsem določa politike in postopke ali procese, ki definirajo uporabo podatkov in kako se ti hranijo; s tem predstavlja pomembno komponento drugega aspekta;
- angl. data management, ki vse politike, postopke ali procese definirana, implementira in določa, kako se bodo ti izvajali znotraj posamezne organizacije.

V poglobljenem pogledu lahko upravljanje s podatki razdelimo na kar nekaj pomembnih delov, ki v končni fazi lahko določijo nekakšen celovit sistem. Tak sistem bomo poskušali v osnovi definirati z določenimi koncepti:

- varnost podatkov,
- odpornost podatkov in implementiranega sistema,
- preglednost in enostavna dostopnost do podatkov,
- kakovost in uporabnost podatkov,
- skrbništvo in administracija podatkov,
- analize podatkov,
- politike upravljanja s podatki.

Nekatere zgornje koncepte bomo zajeli tudi v obliki različnih hipotez, ki jih bomo skozi raziskovalno delo poskušali potrditi ali ovreči. Te bodo definirane z namenom zajeti problematiko tako na ravni ustrezno definiranih postopkov ali politik kot tudi same implementacije kasneje na praktičnem primeru.

Za tak strokovni aspekt raziskovanja se je za namen čim boljše simulacije okolja pripravil tudi enostaven fizični strežnik, ki ga bo poganjal prej omenjeni operacijski sistem. Fizično strukturo komponent strežnika oziroma našega simuliranega okolja bomo tako bolj podrobno definirali v nadaljevanju dela ali bolj podrobno na ravni implementacije okolja.

Seveda tovrstna simulacija okolja ne more zajeti vseh podjetij, saj se posamične organizacije med seboj krepko razlikujejo in sta za posamezen primer potrebna individualna analiza potreb organizacije in obseg podatkov, s katerimi organizacija upravlja. Raziskovalno delo oziroma opredeljena implementacija v nadaljevanju lahko tako zelo dobro služi kot osnova za posamezno organizacijo, ki potrebuje neke vrste sistem za upravljanje podatkov in si ga lahko zaradi enostavnih možnosti razširitve tako na strani strojne opreme kot tudi programske opreme, ali bolj natančno operacijskega sistema TrueNAS, prilagodi za svoje potrebe.

1.1 Opis področja in opredelitev problema

Raziskovanje predvsem v osnovi temelji na podatkih in kako dobro se lahko te izkoristi za določeno problematiko znotraj organizacij. Veliko podjetij ne izkoristi potenciala podatkov, ki se jih poslužuje, in lahko tako prikrajšajo uspešnost svojih poslovnih ciljev ali poslovanje. Predvsem se bomo poglobili v koncept upravljanja s podatki, ki bo zajemal:

- potrebne sisteme za doseg ustreznega rezultata,
- vse možnosti, ki nam jih celoten koncept omogoča,
- potrebe, ki jih imamo v osnovnem poslovnem okolju,
- vpeljane in definirane principe, ki bodo omogočali tovrsten koncept,
- varnost, ki jo lahko dosežemo z vpeljavo sistema za upravljanje s podatki.

Podatki so del vsakdana skoraj vsakega posameznika in lahko hitro postanejo preobilni za obvladovanje, zato lahko hitro posežemo po uporabi določenih aplikacij, programov ali fizičnih medijev za organizacijo podatkov (na primer koledarji, beležke ipd.).

V okolju organizacij pa se obsežnost podatkov krepko poveča, zato organizacija potrebuje neke vrste sistem ali več njih za ustrezno hranjenje in obvladovanje teh. To je možno obvladovati na fizični ravni z implementacijo želenih politik in postopkov, katerih se morajo zaposleni držati. Seveda lahko tak način hitro uide izpod nadzora, saj je treba sproti in fizično spremljati ustreznost doseganja definiranih politik. Za bolj celovito in brezhibno kontrolo se za tak namen implementira namenske digitalizirane sisteme in postopke, ki lahko organizaciji krepko prihranijo čas in resurse za obvladovanje vseh podatkov in njihovo upravljanje.

Tako se pojavi glavno vprašanje, ali je možno implementirati celovit sistem za upravljanje s podatki, ki bo v osnovi lahko dosegel posamezne definirane zahteve v nekem poslovnem okolju.

1.2 Namen, cilji in osnovne trditve

Namen in cilj diplomskega dela je predvsem implementirati sistem za celovito upravljanje s podatki v osnovnem poslovnem okolju. Implementiralo se bo neke vrste osnovno simulacijo poslovnega okolja, kjer bomo poskušali določiti ustrezne postopke in politike ter te implementirati na primerih za ustrezno raven varnosti podatkov, njihovo odpornost na ranljivosti kot tudi različne izpade ali napake uporabnika. Prav tako se bo poskušalo implementirati pregleden in enostaven dostop do podatkov, ki so posameznemu uporabniku na voljo za dostop, in kako se lahko podatke uporabi za namene različnih analiz. Določiti bo treba določene postopke in politike, ki bodo nadzorovale hranjenje in uporabo podatkov.

Končni izdelek bo predstavljal fizični računalnik (oziroma strežnik) prenosne narave, ki bo imel implementiran sistem TrueNAS s potrebnimi in določenimi zahtevami, kot tudi nekaterimi priporočenimi metodami (angl. best-practices) za neko osnovno poslovno okolje.

Za namen doseganja definiranega cilja se je izpostavilo pomembne trditve, ki nas bodo spremljale skozi raziskovalno delo, ki zajemajo pomembne aspekte in temeljijo na ustreznosti implementacije določenih konceptov upravljanja podatkov. Te trditve oziroma hipoteze so sledeče:

- **Hipoteza 1:** Določiti je treba ustrezne politike, ki veljajo za zbiranje, shranjevanje, obdelavo in odstranjevanje podatkov.
- **Hipoteza 2:** V primeru izgube (nezaželenega izbrisa) podatkov je te možno povrniti.
- **Hipoteza 3:** Sistem preprečuje nezaželene spremembe podatkov ob varnostnem incidentu (npr. kripto virus).
- **Hipoteza 4:** Sistem omogoča upravljanje in analizo kompleksnejših vrst ter velikih količin podatkov (Big Data – BTEC 2023).
- **Hipoteza 5:** Sistem prenese izpad enega fizičnega medija za shranjevanje podatkov (trdi disk).
- **Hipoteza 6:** Sistem omogoča varen dostop do podatkov izven lokalnega omrežja.
- **Hipoteza 7:** Sistem ne dovoljuje sprememb in dostopa do podatkov v primeru neustreznih pravic.
- **Hipoteza 8:** Sistem omogoča enostavne nadgradnje tako sistema kot strojne opreme (angl. »end-of-life«).

1.3 Predpostavke in omejitve

Glavna omejitev raziskovalnega dela bo predvsem pomanjkanje ustreznih podatkov za analizo implementirane rešitve, saj se kot podlaga raziskovalnega dela ne bo uporabilo določeno podjetje, ampak se bo simuliralo osnovno poslovno okolje. To je bila predvsem osebna odločitev z namenom širše svobode implementacije in določanja različnih politik, ki so lahko v posameznem podjetju že določene in krepko omejujejo uporabo podatkov za namen raziskovanja ali implementacije želenega sistema.

S tem bo potreben velik poudarek na pridobitvi in uporabi podatkov, ki so javno dostopni oziroma se lahko pridobi dovoljene za uporab,o ali pa bo potreben večji trud na ročnem generiranju naključnih podatkov za namen analize implementiranega sistema.

Dodatna omejitev temelji na uporabi literature za raziskovalno delo, ki bo temeljila predvsem na tuji strokovni literaturi, saj področja upravljanja s podatki domači avtorji še niso dovolj široko in podrobno opisali. S tem bo potrebno več časa za proučitev tuje strokovne literature in jo ustrezno prevesti ter uporabiti za osnovo raziskovalnega dela.

1.4 Uporabljene raziskovalne metode

Raziskovanje bo potekalo v lastni režiji z uporabo namenske programske opreme oziroma operacijskega sistema TrueNAS. Programsko opremo bo treba celovito proučiti, da lahko izpeljemo vse funkcionalnosti, ki jih sistem omogoča, in izluščiti nato vse potrebne funkcionalnosti za naš primer ter nato vse zapakirati v implementacijo želenega sistema za upravljanje s podatki.

Kot že izpostavljeno, se bo vse ostale aspekte postopkov in principov upravljanja s podatki v računalniškem sistemu proučilo predvsem v tuji strokovni literaturi.

Vse podatke, potrebne za ustrezno implementacijo in analizo sistema za upravljanje s podatki, se bo pridobilo iz javno dostopnih virov oziroma se bomo poslužili predvsem ročno generiranih naključnih podatkov, da dosežemo zadostne količine za doseg zadovoljivih rezultatov analize tovrstnega sistema.

2 UPRAVLJANJE S PODATKI

Veliko podjetij hitro spozna, da so njihovi podatki ključno poslovno orodje za doseganje potrebnih vpogledov v svoje stranke, produkte, kot tudi storitve. Privedejo lahko do ključnih inovacij in doseganja želenih strateških ciljev, ki lahko pomenijo za organizacijo pomemben korak za konkurenčnost na trgu. Ne glede na to trditev večina organizacij ne upravlja s podatki redno, kot samo poslovno orodje, s katerim lahko pridobijo višjo vrednost podatkov. Seveda je za to potreben namen oziroma zahteve, poglobljeno planiranje, medsebojna in celovita koordinacija in neke vrste zavezanost za doseg sistema znotraj organizacije (Evans & Price, 2012).

Upravljanje s podatki predstavlja implementacijo različnih načrtov, politik kot tudi namenske programske in strojne opreme za doseganje višje kontrole in varnosti ter s tem povečati vrednost podatkov znotraj njihovega življenjskega cikla. Aktivnosti in dejavnosti znotraj sistema za upravljanje s podatki segajo krepko v širino. Zajemajo tako strateške aktivnosti za namen ustreznih odločitev kot tudi tehnično implementacijo sistema. Tako pojem upravljanja s podatki potrebuje poslovni kot tudi tehnični vidik za ustrezno implementacijo in doseganje želenih vrednosti sistema (DAMA International, 2017).

Za osnovno poslovno okolje lahko izpostavimo tudi nekaj ciljev, ki bi jih želeli doseči s sistemom za upravljanje s podatki, ki jih izpostavlja organizacija DAMA International (2017):

- preglednost in razumevanje potreb po podatkih in informacijah organizacije, kot tudi njihovih deležnikov, vključno s strankami, poslovnimi partnerji in seveda zaposlenimi;
- implementirati sistem za pridobivanje, hranjenje, varnost in zagotavljanje celovitosti podatkovnih sredstev, ki so na voljo znotraj in v dosegu organizacije;
- zagotovitev kakovosti podatkov in informacij;
- doseči ustrezno zasebnost in zaupnost podatkov deležnikov organizacije;
- zaustaviti neprivilegirane ali neustrezne dostope do podatkov, njihovo manipulacijo ali uporabo;
- zagotovitev učinkovitosti uporabe podatkov za doprinos višje vrednosti organizaciji.

2.1 Glavne komponente upravljanja s podatki

Organizacija DAMA International (2017) v svoji strategiji upravljanja s podatki določa številne izstopajoče komponente, ki so posamezno zelo pomembne za doseganje celovitosti in jih je treba tudi ločeno in poglobljeno opredeliti. Te komponente so sledeče:

- **arhitektura podatkov:** definira in določa, kako se podatki hranijo, uporabljajo, spreminjajo in v začetku pridobijo;
- **modeliranje in oblikovanje podatkov:** določa poslovne koncepte in procedure, standarde imenovanja in oblike podatkov;
- **administracija podatkov:** upravljanje podatkovnih zbirk in zagotavljanje dostopnosti do podatkov;
- **kakovost podatkov:** zagotavlja, da podatki dosegajo poslovne zahteve in potrebe;
- **integracija podatkov:** združuje podatke na enem mestu v skupen enoten pogled;
- **analitika podatkov:** pridobivanje vpogledov iz podatkov;
- **varnost podatkov:** preprečevanje varnostnih incidentov in neprivilegiranih dostopov do podatkov;
- **nadzor upravljanja s podatki:** zagotavlja določno raven ustrezne uporabe podatkov znotraj organizacije.

Na tej ravni lahko izpostavimo tudi pojme poslovne inteligence, podatkovne znanosti, podatkovno skladiščenje podatkov, skrbništvo podatkov in dokumentacijo, ki definira posamezne komponente znotraj poslovnega okolja. To so lahko vse komponente upravljanja s podatki, vendar se lahko definirajo in uporabijo na nižji ravni znotraj glavnih izpostavljenih komponent.

2.2 *Arhitektura podatkov*

V osnovi arhitektura podatkov sestavlja infrastrukturo, ki povezuje na eni strani poslovno strategijo, na drugi pa strategijo podatkov v obliki tehnične izvedbe. Kot primer vzemimo primerjavo z igro šah. Če si predstavljamo, da ena figura v igri šah predstavlja organizacijo, potem arhitektura podatkov definira posamezne možne poteze na polju (Johnson, 2021).

Lahko si jo predstavljamo kot ogrodje, ki ga sestavljajo različni podatkovni modeli, politike, pravila in standardi, ki ga organizacija porabi za upravljanje in pretok podatkov znotraj poslovnega okolja. Uspešna implementacija arhitekture podatkov doprinese standardizaciji postopkov, kako se znotraj podjetja podatki hranijo, spreminjajo in pridobivajo, in pa tudi, da osebe, ki posamezne podatke potrebujejo, te tudi pridobijo (Simplilearn, 2023).

Arhitektura podatkov deluje kot osnovni načrt za določanje pretoka različnih količin podatkov znotraj organizacije in sestavlja koncept načrtovanja, ki določa strukturo podatkov za hranjenje uporabnih podatkov v organizaciji (Davoy, 2022).

Slika 1: Primer arhitekture podatkov

Vir: (https://davoy.tech/what-is-data-architecture/)

Arhitekturo podatkov lahko delimo na tri ključne komponente:

- podatkovno jezero (angl. Data Lake),
- podatkovno skladišče (angl. Data Warehouse) in
- podatkovni mart (angl. Data Mart).

Slika 2: Ponazoritev treh komponent arhitekture podatkov

Vir: (https://davoy.tech/what-is-data-architecture/)

2.2.1 Podatkovno jezero

Podatkovno jezero predstavlja osrednje skladišče, kjer se hranijo veliki ali velepodatki (angl. Big Data) v neobdelani prvotni obliki, dokler jih nekdo znotraj organizacije uporabi. Namen podatkovnih jezer je centralno hranjenje tako strukturiranih kot nestrukturiranih oblik podatkov (Rahman , 2021).

Podjetje Inden (2019) navaja kot glavne prednosti podatkovnega jezera naslednje:

- Podatki se hranijo v surovi prvotni obliki in tako ni potrebe po modeliranju ob vnosu.
- Nemoteno lahko povečamo kapaciteto z relativno majhnim stroškom.
- Hranijo se lahko podatki različnih struktur in formatov, kar nudi veliko prilagodljivost.
- Sheme se lahko definirajo po vnosu podatkov, kar ponuja fleksibilnost.

2.2.2 Podatkovno skladišče

Podatkovno skladišče predstavlja velik skupek poslovnih podatkov, ki se uporabljajo za sprejemanje odločitev znotraj organizacije. Vse količine podatkov se v podatkovno skladišče pridobivajo iz različnih virov, kot so interne aplikacije (marketing, prodaja, finance …), transakcijski sistemi (CRM, ERP …) in drugi zunanji viri podatkov. To nato predstavlja ločeno bazo podatkov, ki je neodvisna od internih aplikacij in transakcijskih sistemov ter tako ne ogroža stabilnosti sistemov (talend, 2023).

2.2.3 Podatkovni mart

V svetu, kjer dominirajo veliki podatki in analitika, podatkovni marti predstavljajo eno od prednosti za ustrezen doseg pretvorb informacij v različne poglede. Podatkovni mart je namensko usmerjena podatkovna zbirka, ki je znotraj poslovnega okolja velikokrat implementirana kot ločen segment podatkovnega skladišča. Sklop podatkov, zapisan znotraj podatkovnega marta, se po navadi navezuje na posamezno področje znotraj organizacije, kot so finance (talend, 2023).

V osnovi bi lahko rekli, da je podatkovni mart manjša oblika podatkovnega skladišča, ki je namensko usmerjeno na področje. Iz obstoječega podatkovnega skladišča lahko izpeljemo več podatkovnih martov, kot podmnožico podatkovnega skladišča.

Slika 3: Prikaz izpeljanih namenskih podatkovnih zbirk podatkovnega skladišča

Vir: (https://www.talend.com/resources/what-is-data-mart/)

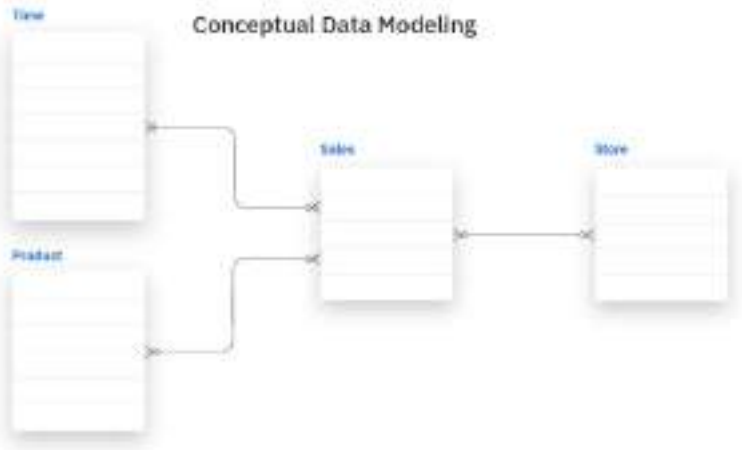
2.3 *Modeliranje in oblikovanje podatkov*

Modeliranje podatkov definira proces ustvarjanja vizualne predstavitve celotnega informacijskega sistema ali njegovih delov za namen povezovanja med podatki in strukturami. Cilj je predvsem vizualno ponazoriti vrste uporabljenih in shranjenih podatkov v sistemu, relacije med njimi in načine združevanja ter organiziranja teh različnih vrst podatkov s pripadajočimi atributi (IBM, 2023).

2.3.1 Podatkovni modeli

Podatkovne modele lahko razdelimo v tri kategorije, ki se razlikujejo glede na stopnjo abstrakcije: konceptualni podatkovni model, logični podatkovni model in fizični podatkovni model (IBM, 2023).

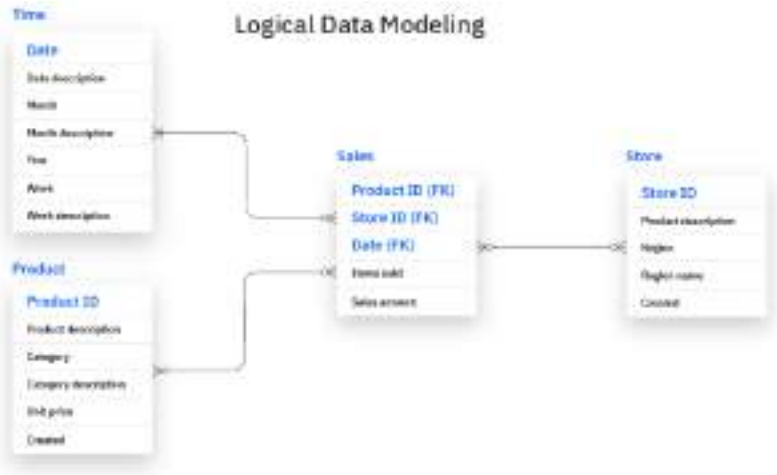
Konceptualni model je namenjen predvsem vizualizaciji osnovne slike sistema in kako bo ta organiziran.



Slika 4: Konceptualni podatkovni model

Vir: (https://www.ibm.com/topics/data-modeling)

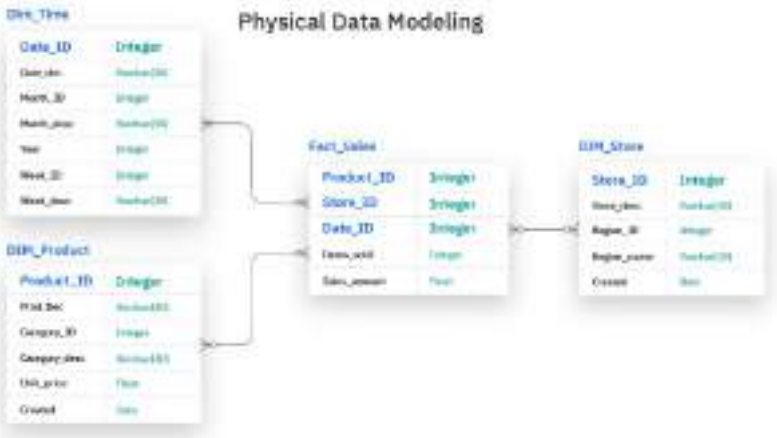
Logični model prinaša več informacij, ki prinašajo širšo sliko konceptov in relacij. Zajamejo se vsi atributi, ki vežejo vrsto podatkov, njihove določene dolžine in relacije med posameznimi atributi.



Slika 5: Logični podatkovni model

Vir: (https://www.ibm.com/topics/data-modeling)

Fizični podatkovni modeli pa nekako že ponazarjajo celotno shemo, kako bodo podatki shranjeni v bazi.



Slika 6: Fizični podatkovni model

Vir: (https://www.ibm.com/topics/data-modeling)

2.4 Administracija podatkov

Administracija podatkov predstavlja proces, v katerem se podatke nadzoruje, vzdržuje in upravlja, ki ga določa administrator podatkov. Administracija podatkov organizaciji prinaša kontrolo nad njihovimi podatkovnimi sredstvi. Vključuje logično podatkovno upravljanje, kjer se pretok podatkov lahko analizira, kreirajo se potrebni podatkovni modeli in določi ustrezne relacije znotraj njih. Administracija zajema tudi varnost podatkov in omejitev dostopa do njih (Rouse, 2016).

2.5 *Kakovost podatkov*

Zahteve za kakovost podatkov se določa v okolju organizacije v meri, da dosegajo njena določila natančnosti, veljavnosti, popolnosti in doslednosti. Ta določila se morajo določiti znotraj posamezne organizacije in se s tem lahko zelo razlikujejo med posameznimi podjetji. S tem kakovost podatkov postaja merilo, kako posamezni podatki ustrezajo določenemu namenu oziroma ali ti dosegajo definirana določila za doseganje določenih poslovnih ciljev (Suer, 2023).

Slika 7: Elementi kakovosti podatkov

Vir: (https://www.alation.com/blog/what-is-data-quality-why-is-it-important/)

Velike količine podatkov predstavljajo izziv za njihovo kakovost. Kadar upravljamo z velikimi količinami podatkov, se s tem poveča tudi količina novih informacij, ki jih je treba obdelati, kar postavi dodatne izzive pri ugotavljanju verodostojnosti podatkov (Suer, 2023).

Največji izzivi pri določanju kakovosti podatkov:

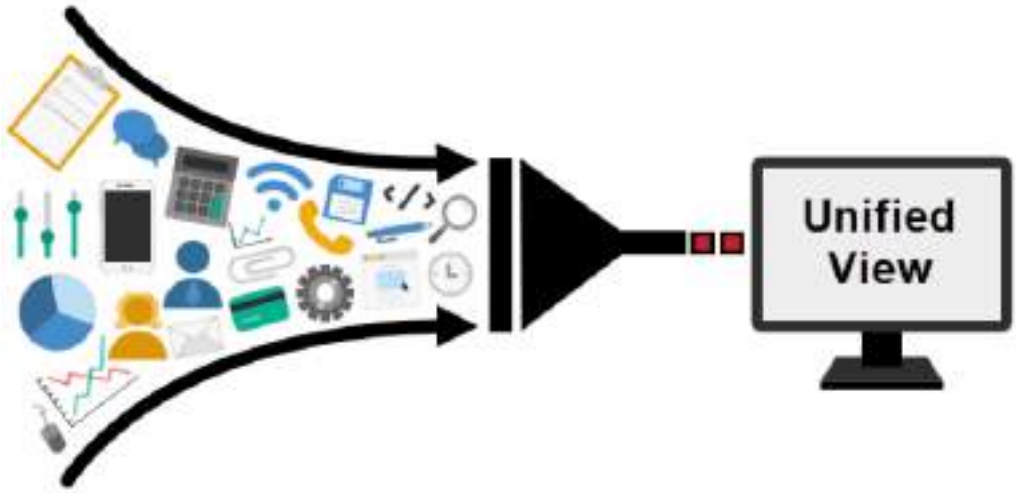
- Zakonodaja o zasebnosti in varstvu podatkov (GDPR) daje ljudem pravico do svojih osebnih podatkov, kar zahteva bistveno bolj natančne evidence strank, iz katerih so organizacije zmožne pridobiti vse informacije o posamezniku, in to nemudoma, ne da bi izpustile kateri koli del pridobljenih podatkov.
- Z implementacijo umetne inteligence in strojnega učenja v poslovno inteligenco podjetja uporabniki podatkov oteženo sledijo novim prilivom velepodatkov. Zaradi teh velikih prilivov podatkov, ki jih novo implementirane rešitve pridobivajo, se povečajo možnosti za napake in nepravilnosti kakovosti podatkov.

- Nadzor nad upravljanjem podatkov predstavlja sistem upravljanja s podatki, ki definira in upošteva notranje sklope standardov in pravilnikov, ki jih določi organizacija za zbiranje, hrambo in izmenjavo informacij. Zagotavljanje doslednosti podatkov, da ti niso zlorabljeni in so zaupanja vredni znotraj vsakega oddelka v podjetju, dosežemo z boljšim zagotovilom skladnosti s predpisi in tako zmanjšamo tveganje kazni nad organizacijo.

Visokokakovostni podatki privedejo do nižanja stroškov zaradi manjših potreb po reševanju nekakovostnih podatkov, kar preprečuje drage napake ali odločitve znotraj organizacije. Kakovostni podatki doprinesejo tudi bolj natančne analize, ki nudijo podjetju boljše poslovne odločitve (Suer, 2023).

2.6 Integracija podatkov

Proces združevanja podatkov s strani različnih virov v poenoten pogled lahko definiramo kot integracijo podatkov. Končni cilj integracije je zagotoviti uporabnikom dosleden dostop do podatkov in njihov prenos po različnih področjih in strukturah za zagotovitev vseh informacijskih potreb aplikacij in poslovnih procesov (Heavy.ai, 2022).



Slika 8: Integracija podatkov

Vir: (https://hevodata.com/learn/data-integration-vs-data-migration/)

Zahteve po integraciji velepodatkov konstanto rastejo in je ustrezna integracija podatkov ključna za obvladovanje tovrstnih količin podatkov.

Slika 9: Prednosti integracije podatkov

Vir: (https://appinventiv.com/blog/what-is-data-integration/)

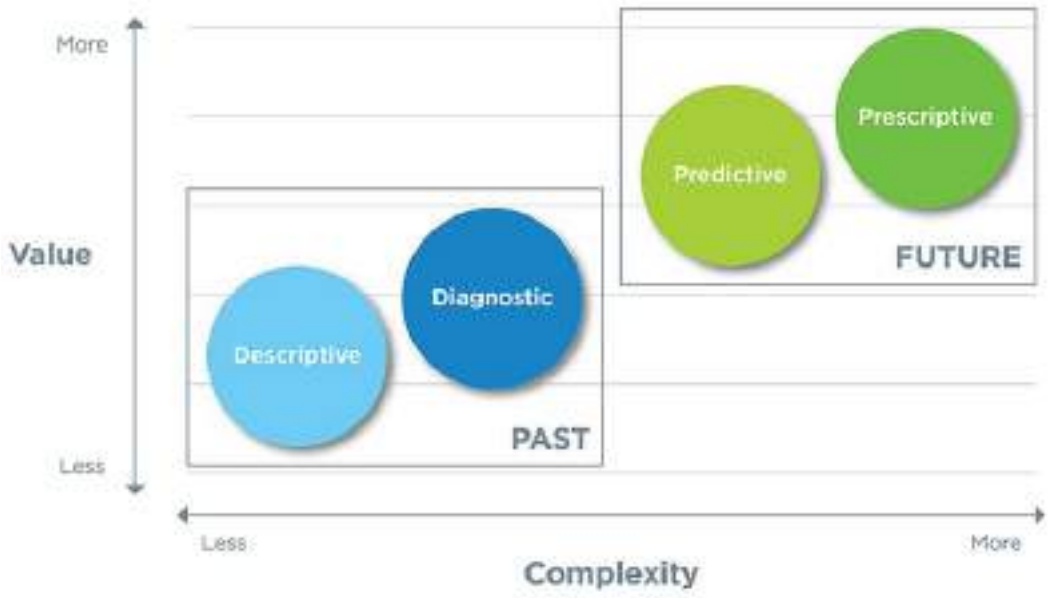
2.7 Analitika podatkov

Analitika prinaša uporabo podatkov v povezavi z različnimi orodji in tehnikami za ugotavljanje in opredelitev določenih vzorcev ali trendov, ki lahko prinašajo uporabne vpoglede v informacije, ki nudijo podporo za boljše odločitve. Primarni cilj analitike podatkov predstavlja obravnavo konkretnih vprašanj ali izzivov, ki so ključni za doseganje bolj uspešnih poslovnih izidov znotraj organizacije (CompTIA, 2023).

Prvi korak za ustrezno in smiselno analitiko podatkov je ugotoviti in določiti zahteve podatkov oziroma kako se bodo te organizacijsko definirale. Podatki so lahko ločeni po datumu nastanka, dohodkih, spolu ali demografiji podatkov. Sortirani podatki lahko predstavljajo numerične vrednosti ali pa so ločeni po določenih kategorijah. Drugi korak določa postopek pridobivanja podatkov. To se lahko izvede prek računalnikov, videonadzora, različnih spletnih virov, osebja ali pa celo okoljskih virov. V tretjem koraku je treba pridobljene podatke ustrezno integrirati in organizirati znotraj sistema, da nato lahko te uporabimo za analize. Ti podatki so lahko integrirati v Excelovih preglednicah ali znotraj druge namenske programske opreme, ki lahko sprejema statistične podatke. Zadnji korak zahteva, da podatke pred analizo ustrezno prečistimo in se s tem izognemo nepopolnosti rezultatov (kot so podvajanja, napake podatkov). Končni korak tako nudi korekcijo vseh napak pred predajo analitiku za izvedbo ustrezne analize podatkov (Frankenfield, Boyle, & Rathburn, 2023).

Kot lahko analitiko razdelimo na postopek, ki zajema več korakov, jo lahko razdelimo tudi na različne tipe analitike. CompTIA in Jake Frankenfield (2023) analitiko delita na štiri tipe:

- **Opisna analitika** (Kaj se dogaja?) opisuje dogodke, ki so se zgodili v določenem časovnem obdobju. Na podlagi trendov in vzorcev preteklih in trenutnih podatkov prikaže trenutno stanje. Odgovarja na vprašanja, vezana na pretekle in trenutne dogodke.
- **Diagnostična analitika** (Zakaj se nekaj dogaja?) se osredotoča na različne razloge, povezane z vzorci in trendi podatkov, ki so bili ugotovljeni v prejšnji fazi, kar privede do ugotovitev preteklih dejavnikov uspeha.
- **Napovedovalna analitika** (Kaj se bo verjetno zgodilo?) na podlagi tehnik napovedovanja in uporabe strojnega učenja poskuša določiti, kaj se bo verjetno zgodilo v prihodnosti, kako je zadnja hladna zima vplivala na prodajo.
- **Predpisujoča analitika** (Kaj je treba narediti?) poskuša predlagati ustrezne načine ukrepanja. Na podlagi algoritmov, poslovnih pravil in strojnega učenja sestavi namenske rešitve in predloge, kako nadaljevati za doseg želenih poslovnih izidov.



Slika 10: Pregled štirih osnovnih vrst analitike

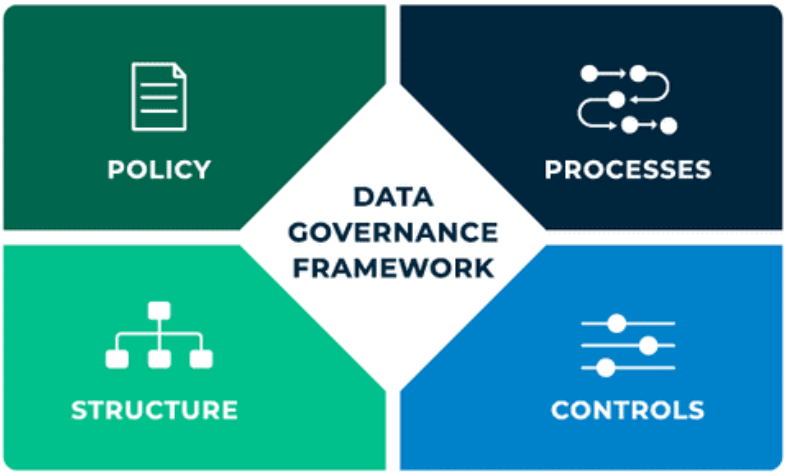
Vir: (https://www.comptia.org/content/guides/what-is-data-analytics)

Ključni del verjetnosti uspeha organizacije nudi prav analitika podatkov. Dejavnosti, kot so zbiranje, razvrstitev, analiza in predstavitev informacij, bistveno koristijo družbi. Tovrstne analitike so lahko prednost tudi malim in zagonskim podjetjem, ki iščejo konkurenčno prednost pred obstoječimi podjetji (Frankenfield, Boyle, & Rathburn, 2023).

2.8 *Nadzor upravljanja s podatki*

Nadzor ali strategija upravljanja s podatki (angl. Data Governance) je skupek procesov, vlog, politik, standardov in meril za zagotavljanje uspešne in učinkovite uporabe informacij. Vzpostavlja procese in določa odgovornosti, ki zagotovijo kakovost in varnost podatkov, ki se uporabljajo znotraj organizacije. Nadzor upravljanja s podatki definira, na podlagi katerih podatkov, s katerimi metodami in v katerih situacijah lahko posameznik izvede določene posege (talend, 2020).

Za organizacije je pomembno, da razvijejo ustrezno strategijo upravljanja s podatki, ki bo omogočila pridobivanje vpogledov, dvignila raven varnosti podatkov in ustrezno zmanjšala stroške. Strategija mora zagotavljati varno upravljanje podatkov skozi celoten proces poslovanja. Prav tako mora določiti zahteve glede shranjevanja teh podatkov, kot so beleženje zgodovine sprememb podatkov, zagotavljanje skladnosti z notranjimi predpisi in tudi vladnimi predpisi, kot je GDPR. Za ustrezno določno raven odgovornosti nad podatki in doseganje doslednih postopkov je ključnega pomena ustrezno in dobro definirana strategija upravljanja s podatki (talend, 2020).



Slika 11: Komponente nadzora upravljanja s podatki

Vir: (https://www.alteryx.com/glossary/data-governance)

Za nadzor upravljanja s podatki lahko izpeljemo štiri glavne komponente: politike nadzora upravljanja, struktura nadzora upravljanja, procesi nadzora upravljanja in kontrole nadzora upravljanja podatkov.

Na prvi pogled se lahko izpelje zelo velika primerjava z upravljanjem s podatki (angl. data management), vendar ju loči pomembna razlika. Nadzor upravljanja s podatki predstavlja strategijo, ki jo določimo za upravljanje s podatki, ko samo upravljanje s podatki določa prakse oziroma tehnični vidik implementacije definirane strategije, ki se uporabi za zaščito vrednosti podatkov. Ob definiranju strategije nadzora upravljanja s podatki določamo in vključimo ustrezne prakse upravljanja s podatki (Imperva, 2023).

Slika 12: Strategija celovitega upravljanja s podatki

Vir: (https://www.dataversity.net/data-management-vs-data-strategy-a-framework-for-business-success/)

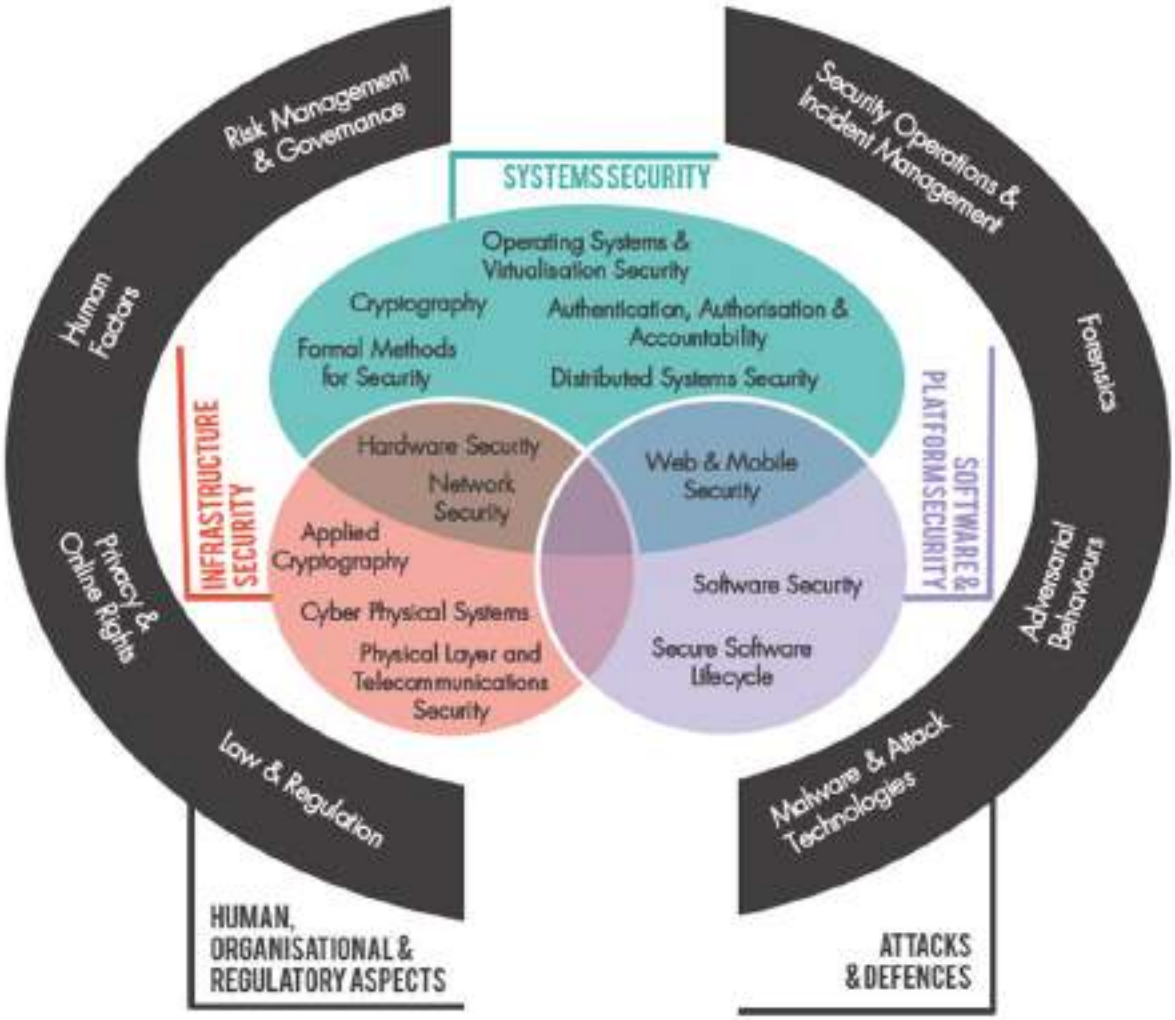
Za celovit sistem upravljanja s podatki je nujno potrebna ustrezno definirana strategija nadzora upravljanja, ki je v današnjem svetu skorajda ne smemo izpustiti, saj je bolj ali manj nujen del za implementacijo sistema upravljanja s podatki.

Slika 13: Koraki za izvedbo ustrezne strategije nadzora upravljanja

Vir: (https://www.alation.com/blog/steps-for-building-data-governance-strategy/)

2.9 *Varnost podatkov in sistema*

Distribuiran sistem v osnovi deluje na podatkih na podlagi različnih vidikov virov, distribucije, shranjevanja ali uporabe teh podatkov v določenih storitvah ali aplikacijah. Za vsak posamezen element veljajo enake klasične lastnosti CIA (zaupnost, celovitost in razpoložljivost), ki so del izbrane podatkovne verige (Rashid, in drugi, 2021).



Slika 14: Celovit pogled na kibernetsko varnost

Vir: (https://www.cybok.org/ataglance/)

Vsak sistem in s tem povezani podatki so lahko hitro del nepredvidenega varnostnega dogodka ali celo incidenta, kar lahko privede do nezaželenih tveganj sistema ali podatkov.

Varnost podatkov predstavlja skupek procesov in različnih orodij, katerih cilj je zaščititi kritična sredstva organizacije (Ekran System, 2023).

2.9.1 Načini varovanja podatkov

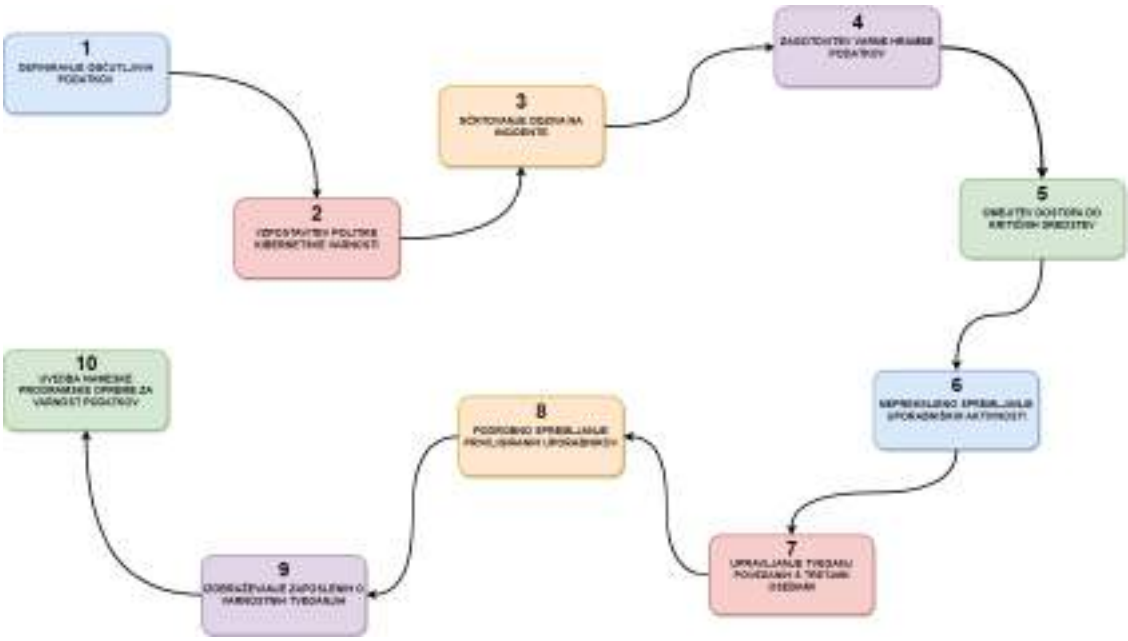
Veliko rešitev za varovanje podatkov ni kos modernim sofisticiranim načinom kibernetskih napadov, še posebej nekaterih, ki se usmerijo v arhiviranje podatkov. S tem so podjetja primorana implementirati večnivojsko kibernetsko obrambo podatkov (Kaelble, 2022).

Nekaj glavnih splošnih načinov varnosti podatkov, ki jih je v organizaciji smiselno kombinirati za najboljšo možno raven varnosti:

- **Šifriranje** s pomočjo algoritmov premeša podatke in skrije njihov pomen. S šifriranjem dosežemo, da lahko posamezne podatke prebere samo prejemnik ali oseba, ki uporablja ustrezen ključ za dešifriranje.
- **Brisanje nepotrebnih podatkov**, ko so ti obdelani in niso več potrebni za uporabo v organizaciji. S tem zmanjšamo obveznosti in možnosti vdora do podatkov.
- **Odpornost podatkov** (angl. data resilience) vključuje polno, inkrementalno varnostno kopiranje in varnostno kopiranje na podlagi sprememb kritičnih podatkov. Hranjenje kritičnih podatkov na več lokacijah prinaša boljšo odpornost in možnosti obnovitve teh podatkov v primerih kibernetskih groženj.
- **Prikrivanje podatkov** zajema postopek prikrivanja in skrivanja podatkov visokih vrednosti, na način, da se občutljive informacije prikrije z naključnimi znaki. Možnost imamo tudi nadomestitev podatkov z nizko vrednostnim varnostnim žetonom, kar nam nudi metoda tokenizacije.

Slika 15: Glavni dejavniki varnosti podatkov

Vir: (https://www.ekransystem.com/en/blog/data-security-best-practices)



Slika 16: Deset najboljših praks za implementacijo varnosti podatkov

Vir: (Lastni vir)

2.9.2 Tveganja varnosti podatkov

Ne obstaja sistem, ki bi omogočal popolno zaščito pred kibernetskimi grožnjami, neustrezno konfiguracijo ali celo notranjimi grožnjami. Kritični podatki se lahko izgubijo, poškodujejo ali pa v najslabšem primeru znajdejo v rokah napačne osebe (Ekran System, 2023).

Slika 17: Možna tveganja v organizaciji

Vir: (https://www.ekransystem.com/en/blog/data-security-best-practices)

Največja tveganja varnosti podatkov, ki jih izpostavlja Fortinet (2023):

- **Nenamerna izpostavljenost podatkov** je vzrok nenamernega ali ponesrečenega razkrivanja občutljivih podatkov zaposlenega.
- **»Phishing« napadi** se izvajajo predvsem prek e-poštnih sporočil v smiselnih oblikah, v katerih se napadalci ponazarjajo kot zaupanja vredni pošiljatelji. Večinoma vsebujejo zlonamerne povezave ali priponke, ki ob interakciji prejemnika prenesejo zlonamerno programsko opremo, kar vodi do dostopa napadalca do občutljivih podatkov (prijavni podatki uporabnikov, finančni podatki …).
- **Zlonamerna programska oprema** omogoča napadalcem z izkoriščanjem ranljivosti programske opreme ali sistemov dostop do občutljivih podatkov. Predvsem se širijo prek spletnih napadov in e-poštnih sporočil. Rezultat napada so lahko predvsem kraja podatkov, različno izsiljevanje napadalcev in poškodbe na omrežju.
- **Izsiljevalska programska oprema** predstavlja eno večjih varnostnih tveganj podatkov. Je oblika zlonamerne programske opreme, katere cilj je pridobiti dostop do določene naprave ali sistema in šifrirati vse podatke na njih. Nato napadalci zahtevajo odkupnino od njihovih žrtev z neke vrste obljubo povrnitve podatkov v ustrezno obliko ob plačilu. Nekatere izsiljevalske programske opreme imajo možnost hitre razširitve in tako lahko okužijo celotna omrežja, na katera so vezani tudi strežniki za arhiviranje podatkov, ki jih lahko virus prav tako okuži.
- **Hramba podatkov v oblaku** je povečana strategija organizacij, saj omogoča lažje sodelovanje z izmenjavo dokumentov. Z migracijo v oblak pa lahko otežimo kontrolo in varnost podatkov, kar lahko pelje do hitrejše izgube podatkov. Ker se lahko do teh podatkov dostopa praktično na vseh napravah, se lahko zgodijo hitrejša nenamerna razkrivanja podatkov nepooblaščenim osebam.

Slika 18: Vpliv varnostnih tveganj podatkov

Vir: (https://www.ekransystem.com/en/blog/data-security-best-practices)

2.9.3 Nadzor dostopa do podatkov

Nadzor dostopa do podatkov predstavlja tehniko, pri kateri urejamo dostop zaposlenih do posameznih datotek ali podatkov v organizaciji. Uporablja se načelo najmanj privilegiranega (POLP), kar pomeni, da se nadzira in upravlja pravice dostopa posameznih zaposlenih v povezavi z njihovimi vlogami v organizaciji. Na podlagi istega načela se tako tudi opredeli in omeji, do katerih podatkov ima posameznik dostop (ManageEngine, 2023).

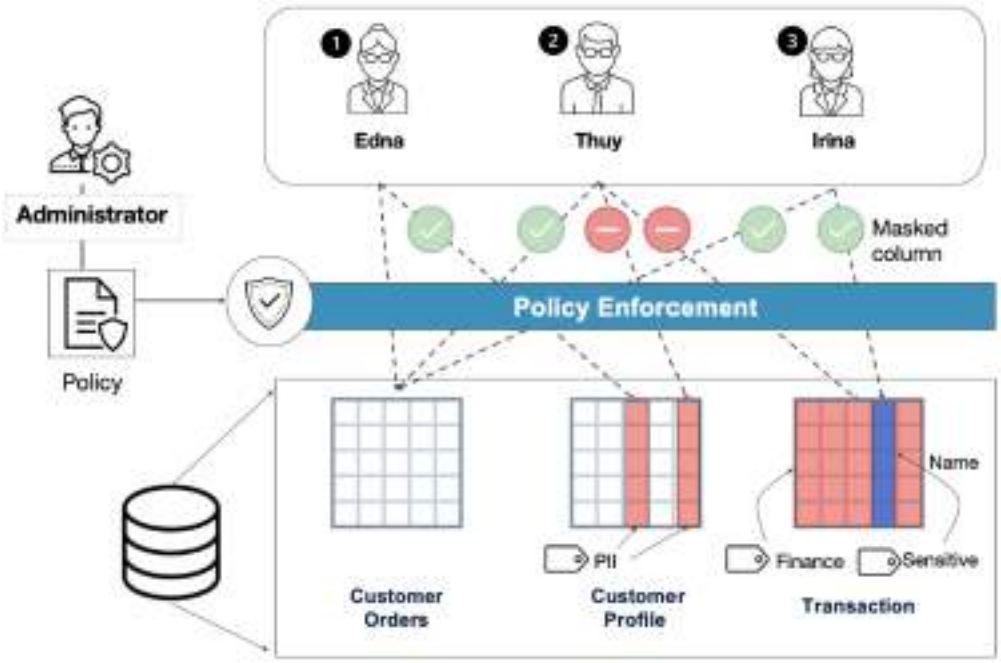
Praksa, v kateri uvajamo zanesljive politike, namenska orodja in tehnologije za omejevanje dostopa do računalniških podatkov podjetja, predstavlja temeljno orodje za varnost podatkov, s katerim želimo doseči manjša tveganja za organizacijo (Cyral, 2023).

Slika 19: Dejavnosti nadzora dostopa do podatkov

Vir: (https://cyral.com/glossary/data-access-control/)

Glavni komponenti nadzora dostopa do podatkov sta predvsem večfaktorsko avtenticiranje, kjer se vpelje več stopenj nadzora identitete uporabnika, in pa avtorizacija, ki za delovanje uporablja prej definirane politike za določanje ravni dostopa ter vrste aktivnosti, ki jih lahko posamezen uporabnik izvaja (Cyral, 2023).

Pri varnosti podatkov je nadzor dostopa ključnega pomena, da se lahko ustrezno zagotovi raven varnosti, kjer podatki ne morejo preiti v napačne roke ali pa zapustiti organizacije. V številnih podjetjih se hranijo osebni podatki njihovih strank, dokumenti, ki vsebujejo zaupne podatke, in še veliko drugih vrst občutljivih podatkov. Z uvedbo nadzora dostopa do podatkov se tako ustrezno zmanjša možnost uhajanja občutljivih podatkov.



Slika 20: Primer nadzora dostopa do podatkov

Vir: (https://docs.treasuredata.com/display/public/PD/About+Column-level+Access+Control)

Cyral (2023) izpostavlja naslednje metode za nadzor dostopa do podatkov:

- **Diskrecijski nadzor dostopa** (DAC): Lastnik sam določa, kdo lahko dostopa do njegovih podatkov na podlagi pravil, ki jih prav tako določa sam. Predstavlja metodo z najmanj omejevanja.
- **Obvezni nadzor dostopa** (MAC): Administrator razvrsti razpoložljive vire in uporabnike v posamezne stopnje tveganja in glede na njihove pristojnosti.
- **Nadzor dostopa na podlagi vlog** (RBAC): Gre za najbolj razširjen način nadzora, saj se dostop prilagaja na podlagi vlog vsakega posameznika. Posamezniki z najmanj pravicami imajo omogočen dostop samo do podatkov, ki jih nujno potrebujejo za delo.
- **Nadzor dostopa na podlagi atributov** (ABAC): Dostop se vrši na podlagi niza atributov in okoljskih pogojev (npr. čas in lokacija), ki so dodeljeni uporabniku in podatkovnim virom.
- **Nadzor dostopa na podlagi politike** (PBAC): Nadzor se vrši na podlagi vnaprej določenih pravil in politik, ki odobrijo ali zavrnejo dostop, zahtevan s strani uporabnika.

3 IMPLEMENTACIJA SISTEMA TRUENAS

3.1 Sistem TrueNAS

TrueNAS predstavlja enega najpopularnejših operacijskih sistemov, namenjenih hranjenju in upravljanju podatkov. Omogoča nične stroške na ravni programske opreme, s katero lahko dosežemo postavitev sistema za shranjevanje podatkov na profesionalni ravni. Njegova glavna prednost je odprtokodna rešitev programske opreme, ki je brez stroškov na voljo v dveh oblikah ali različicah sistema s polnim naborom funkcij. Ti dve obliki se delita na TrueNAS Core in TrueNAS Scale. Prva različica predstavlja osnovno obliko operacijskega sistema, striktno namenjenega shranjevanju podatkov, medtem ko druga, nekoliko bolj sveža rešitev, predstavlja odprtokodno obliko hiperkonvergirane infrastrukture (HCI), ki prinaša mnoge možnosti razširitev, obenem pa dodaja funkcije Linux kontejnerjev in virtualnih namizij, kar doprinese, da se lahko aplikacije izvajajo bližje podatkom (iXSystems, 2023).

Slika 21: Hiperkonvergirana infrastruktura

Vir: https://www.truenas.com/truenas-scale/

TrueNAS Scale se bo uporabil tudi v našem primeru implementacije. Tovrstni bazira na operacijskem sistemu Debian Linux in omogoča visoke možnosti razširitev, kar prinaša lažjo razširitev sistema znotraj organizacije ob višanju količin podatkov ali podatkovnih zahtev.

3.1.1 Datotečni sistem ZFS

Datotečni sistem ZFS (angl. Zettabyte File System) predstavlja osrednjo funkcionalnost oziroma arhitekturo sistema TrueNAS. Ta za shranjevanje in obdelavo podatkov uporablja standard OpenZFS, ki kot odprtokodna različica ZFS prinaša poslovne funkcije varnosti podatkov. Glavno prednost predstavlja usmerjenost v reševanje integritete in varnosti podatkov.

Z uporabo ZFS se nam odprejo naslednje funkcionalnosti:

- dosežemo lahko neomejene kapacitete in razširljivosti na vseh vrstah medijev;
- arhitektura kopiranja prek zapisov za višjo varnost podatkov in izdelavo posnetkov;
- integrirana polja RAID s podatkovnimi kontrolnimi vsotami in validacijo podatkov;
- zmogljiva transparentna kompresija podatkov in integrirana (angl. Inline) deduplikacija, ki preprečuje redundance pred ali med zapisovanjem podatkov;
- vgrajeno šifriranje na ravni diska, zbirke podatkov ali nabora podatkov;
- vgrajeno predpomnjenje branja in zapisovanja podatkov na vseh vrstah zbirk podatkov;
- omogoča samodejno zaznavanje in reševanje morebitnih nepravilnosti ali poškodovanih podatkov, ki se lahko zgodijo ob okvarah diskov ali zaradi napak na njih;
- izvajanje inkrementalnih posnetkov za obnovitev starejših različic datotek in podatkov;
- repliciramo lahko podatke na oddaljene lokacije (oblak, drugi ločen sistem …);
- omogoča integracijo različnih rešitev za varnostno kopiranje podatkov (Veeam).

3.1.2 Uporabljena strojna oprema

Za namen implementacije našega sistema se je pripravila osnovna strojna oprema za namen postavitve enostavnega strežniškega sistema. Glavne komponente strežnika so:

- procesor: Intel Core i3-10105 z vgrajenim grafičnim procesorjem,
- matična plošča: Asus B560M-C/CSM mATX B560 LGA 1200,
- RAM: Kingston FURY Renegade DDR4-3600 32GB (2 x 16GB) – brez ECC,
- trdi diski: 4 x Western Digital Red Plus 4TB (WD40EFZX),
- diski SSD: Samsung 870 EVO 250GB & HP NVMe M.2 EX900 Plus 256GB,
- napajalnik: be Quiet! SFX Power 3 300W (BN320),
- ohišje: Inter-Tech IM-1 Pocket.

Komponente se združujejo v strežnik manjše narave, ki je obenem lahko skoraj povsem prenosen, kar prinaša lažje umestitve v različne prostore kot tudi ne zaseda veliko prostora. Enostavno dostopne komponente omogočajo tudi lažje nadgradnje strojne opreme (H8).

3.2 Določitev politik in postopkov upravljanja s podatki

Vsaka organizacija ima določene podatkovne potrebe, s katerimi želi doseči zastavljene cilje, vendar morajo tovrstni podatki ustrezati definiranim potrebam, da so cilji izvedljivi. To je treba regulirati z določitvijo postopkov in politik upravljanja s podatki. V organizacijah se bi tovrstne zahteve definirale v obliki različnih pravilnikov, ko bomo za naš primer neposredno izpostavili zahteve.

Za sledečo implementacijo in uporabo sistema bo treba upoštevati naslednje:

- Vsak posameznik mora izvesti vsaj osnovno obliko avtentikacije za dostop do podatkov.
- Vsak uporabnik mora imeti ločen podatkovni prostor, do katerega ima le sam dostop.
- Vsakemu oddelku se dodeli ločen podatkovni prostor, do katerega lahko dostopajo le člani posameznega oddelka, kjer je lastnik podatkov definiran kot vodja oddelka.
- Vsi podatki znotraj posameznih oddelkov morajo biti ločeni na posamezno delovno leto, kjer so porazdeljeni na področja v ločenih mapah.
- Imenovanje posameznih datotek mora vsebovati ime datoteke in datum kreiranja.
- Vsi podatki, ki bodo del posameznih podatkovnih prostorov, morajo zagotavljati obnovitev oziroma povrnitev za vsaj dva tedna nazaj.
- Vse spremembe podatkov morajo biti sledljive za časovno obdobje dveh tednov.
- Podatki morajo biti varnostno kopirani na vsaj dveh različnih lokacijah.
- Vsi podatki morajo biti ustrezno šifrirani v primeru fizične kraje.
- Administrativni portal sistema sme biti dostopen le prek varne povezave HTTPS.
- Administrativni portal sistema mora zahtevati večfaktorsko avtentikacijo.
- Za ustreznost podatkov morajo skrbeti posamezni lastniki.
- Posamezne deljene mape morajo omogočati dostop z različnih operacijskih sistemov (poudarek predvsem na operacijskih sistemih Windows in Linux).
- Za dostop do podatkov izven omrežja (v primeru, da je to omogočeno) mora vsak uporabnik izvesti večfaktorsko avtentikacijo.

Zagotavljanje danih zahtev bo sistemu za upravljanje s podatki določalo nadzor, da lahko podatki ostanejo zavarovani in služijo njihovemu namenu, kar zajema zastavljeno hipotezo 1.

3.3 Konfiguracija sistema

TrueNAS predstavlja operacijski sistem, ki ga je treba ustrezno namestiti. Okolje omogoča namestitev znotraj virtualnega okolja kot tudi na fizično strojno opremo, kar smo tudi uporabili v našem primeru. Skozi postopek namestitve se ne bomo poglabljali, saj je ta na voljo v dokumentaciji sistema.

Nameščen operacijski sistem deluje v obliki CLI oz. ga definira vmesnik ukazne vrstice, kjer je naveden glavni podatek, ki ga potrebujemo za nadaljevanje in nekaj nastavitev. Glavni podatek predstavlja IP-naslov za dostop do grafičnega vmesnika sistema TrueNAS, kjer se bodo izvršili vsi naslednji koraki naše implementacije.

Slika 22: Vmesnik ukazne vrstice sistema TrueNAS

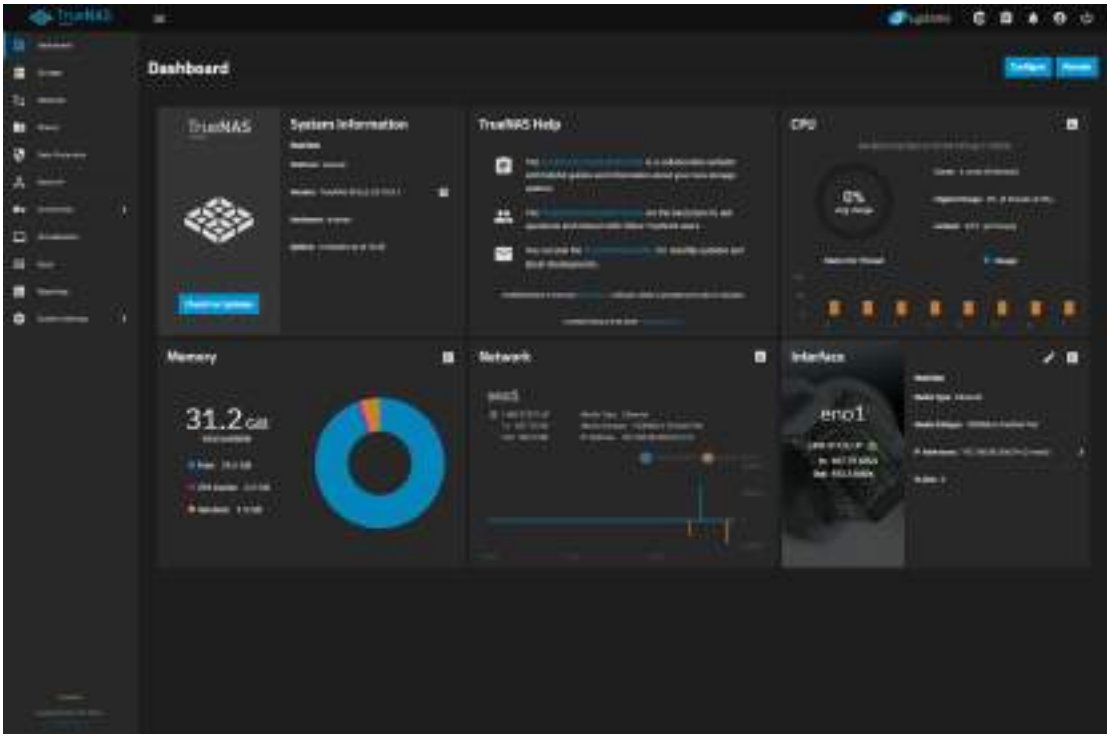
Vir: (Lastni vir)

Dostop nam v osnovi omogoča tako prek spletnega protokola HTTP kot tudi bolj varnega HTTPS, česar se bomo bolj dotaknili v predelu varnosti. Vpis spletne povezave omrežja nas popelje do našega spletnega vmesnika, kjer se zahteva vpis prijavnih podatkov, ki smo jih določili ob namestitvi.

Slika 23: Spletno prijavno okno sistema

Vir: (Lastni vir)

Prikaže se nam osnovna nadzorna plošča našega sistema, ki že v osnovi prikazuje poglede za spremljaje delovanja sistema. Prinaša pregled nad osnovnimi podatki sistema (različica, čas delovanja, ime, platforma), zasedenost delovnega pomnilnika (servisi, ZFS), uporaba procesorja (odstotek porabe, temperatura) in podatek prometa na omrežnem vmesniku sistema. Ob primarnem pogledu lahko tudi takoj enostavno preverimo za možne posodobitve sistema.



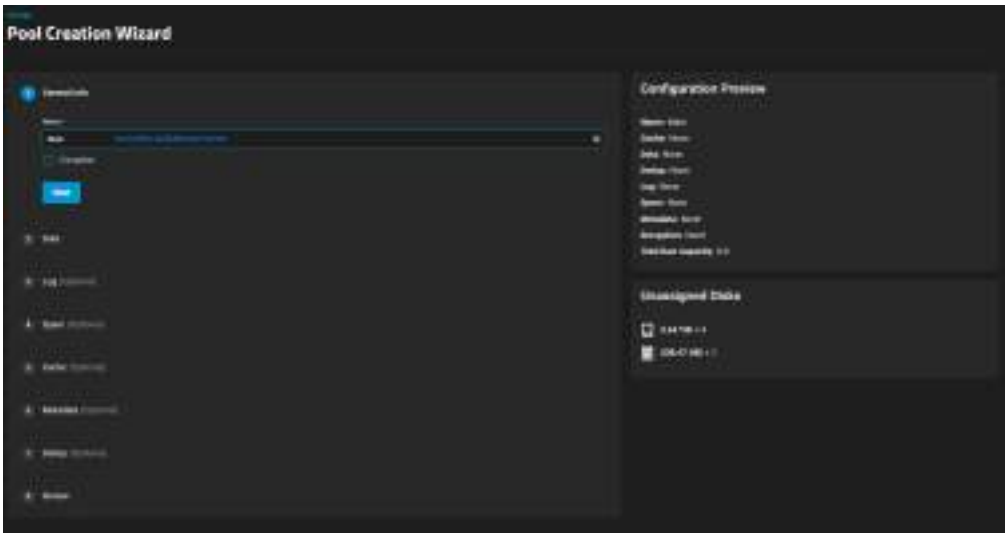
Slika 24: Osnovna nadzorna plošča sistema TrueNAS

Vir: (Lastni vir)

3.3.1 Implementacija glavnega podatkovnega prostora

Ko je sistem nameščen, dostopen, nam v taki obliki ne koristi prav nič, saj da lahko koristimo podatkovni prostor trdih diskov, moramo te najprej prikazati našemu sistemu in jih definirati na podlagi datotečnega sistema ZFS. S tem bo sistem TrueNAS pridobil kontrolo nad definiranim podatkovnim prostorom, uporabnikom pa se bo lahko tako tudi omogočil dostop do podatkovnega prostora.V stranskem meniju vmesnika se pomaknemo na pogled »Storage«, kjer bomo lahko kreirali naš glavni podatkovni prostor. Ta bo predstavljal osrednji podatkovni prostor, na podlagi katerega se bodo kreirali ločeni manjši nabori podatkov, do katerih bomo lahko nato ustrezno določali pravice dostopa.

V zgornjem desnem kotu pogleda lahko kot korak previdnosti najprej preverimo, ali so vsi naši diski vidni v sistemu, kar storimo s klikom na gumb »Disks«. Ob potrditvi, da so vsi potrebni diski vidni, se lahko kreira glavni podatkovni prostor s klikom na gumb »Create Pool«. Sistem vsebuje veliko uporabnikom prijaznih čarovnikov, ki enostavno vodijo skozi potek.



Slika 25: Čarovnik kreiranja glavnega podatkovnega prostora

Vir: (Lastni vir)

Treba je določiti ime podatkovnega prostora, postavitev podatkov v polju RAID, kjer smo za naš primer izbrali RAIDZ1, ki omogoča enojno pariteto, ki reši podatke v primeru okvare enega izmed diskov, ostali trije diski pa nam omogočajo shranjevanje podatkov. Dodatno imamo možnost implementacije diskov za namen dnevnika, predpomnilnika, deduplikcije in diska za hranjenje metapodatkov. Za naš primer smo uporabili dodaten disk NVMe SSD za implementacijo predpomnilnika našega podatkovnega prostora, kar lahko pripomore pri hitrosti dostopa do podatkov (branja) in spreminjanja teh v primeru obsežnih naborov podatkov.

Slika 26: Pogled kreiranega podatkovnega prostora

Vir: (Lastni vir)

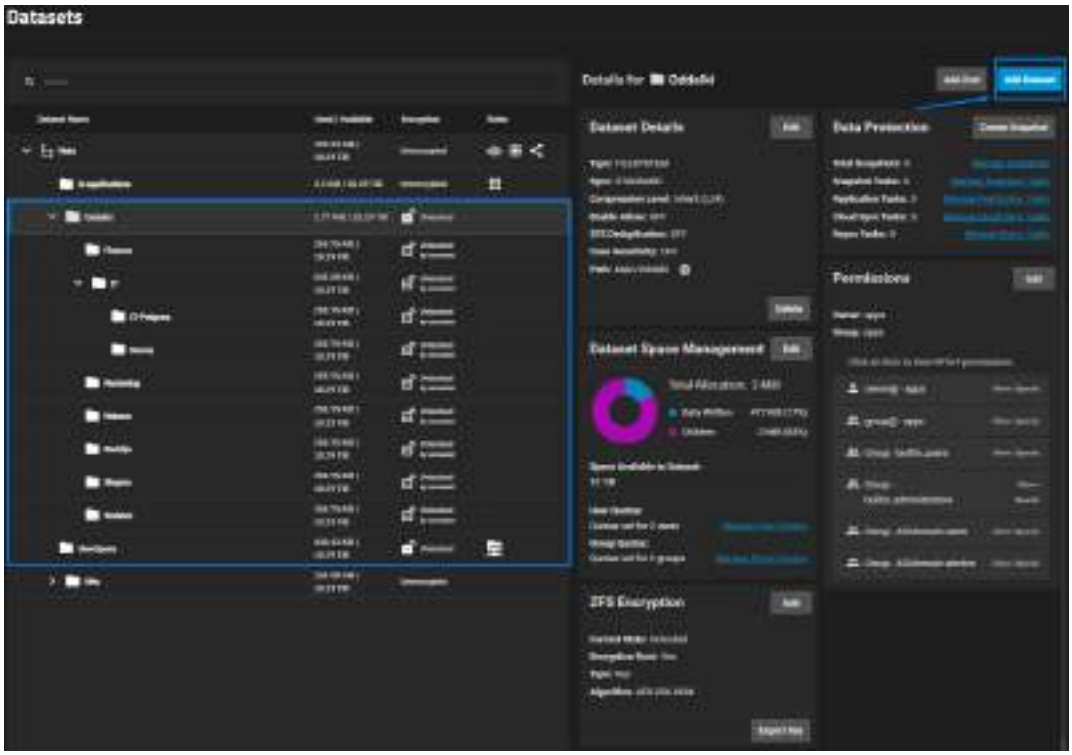
3.3.2 Priprava posameznih naborov podatkov

Definiral se je glavni podatkovni prostor, ki je relativno obširnih kapacitet, katerega je smiselno razbiti na manjše množice oziroma nabore podatkov, ki se bodo v končni fazi prikazali kot mape, ko bodo uporabniki dostopali do podatkovnega prostora. Treba se je osredotočiti na dva večja nabora:

- nabor UserSpace predstavlja podatkovni prostor vsakega posameznika, kjer bodo do svojega imenskega prostora dostopali le posamezniki in drugim uporabnikom ne bodo vidni;
- v naboru Oddelki bodo definirane skupne mape, do katerih se bo lahko dostopalo znotraj poslovnega okolja na podlagi ustreznih pravic za posamezen nabor znotraj nabora Oddelki.

Na posameznem naboru lahko določamo tudi podatkovne kvote posameznika, tako lahko omejimo preveliko kopičenje podatkov.

Ostali vidni nabori na Sliki 27 so lahko v tem koraku brezpredmetni, bodo pa pripomogli pri analizi implementacije.

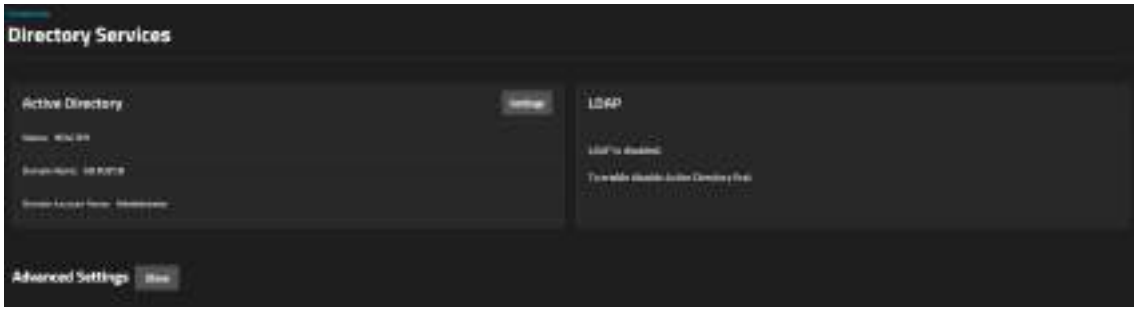


Slika 27: Definirani nabori podatkov

Vir: (Lastni vir)

3.3.3 Upravljanje uporabniških računov in določanje pravic

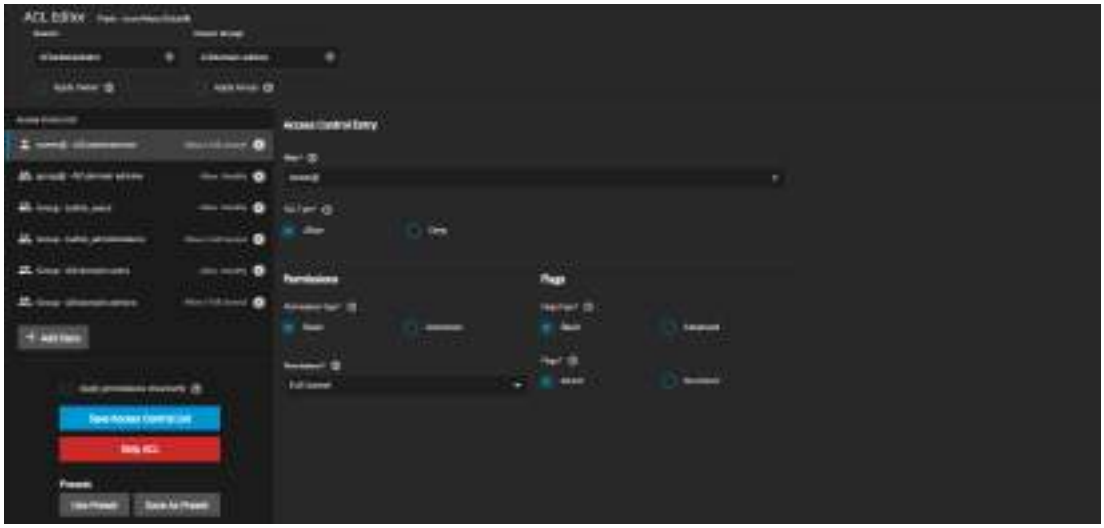
TrueNAS ima že vgrajeno funkcionalnost upravljanja z uporabniškimi računi. Te lahko kreiramo na podlagi različnih pravic in jim določamo članstvo v skupinah, katere nato na ravni naših naborov podatkov uporabimo za določanje uporabniških pravic oziroma dostopov do podatkov. Dodaten način upravljanja z uporabniškimi računi deluje na podlagi članstva v domeni prek aktivnega imenika ali izkoriščanja protokola LDAP za pridobitev uporabniških računov, ki so kreirani na domeni.



Slika 28: Povezava z aktivnim imenikom

Vir: (Lastni vir)

Na Sliki 27 je razviden predel pravic, ki jih lahko za posamezen nabor podatkov urejamo. Prikaže se nam urejevalnik ACL (angl. Access Control List) oziroma urejevalnik dostopa, prek katerega lahko na enostaven način dodajamo pravice uporabnikov, na podlagi katerih lahko nato uporabnik dostopa do posameznih naborov in glede na raven pravic samo bere podatke, te spreminja, zaganja ali pa ima popolno kontrolo nad datotekami.



Slika 29: Urejevalnik dostopa ACL

Vir: (Lastni vir)

Prek urejevalnika dostopa lahko upravljamo z lastniki podatkov, posameznimi uporabniki in celotnimi skupinami, katerim želimo omogočiti dostop. Nastavitve, ki jih določimo znotraj urejevalnika, bodo vidne na posameznih skupnih mapah v zavihku »Security«, kot primer za operacijski sistem Windows. Tako dobimo osrednji urejevalnik za upravljanje pravic posameznih naborov podatkov in tako ni treba ročno prek datotečnega strežnika urejati pravic za posamezne mape ali področja.

3.3.4 Kreiranje skupnih map

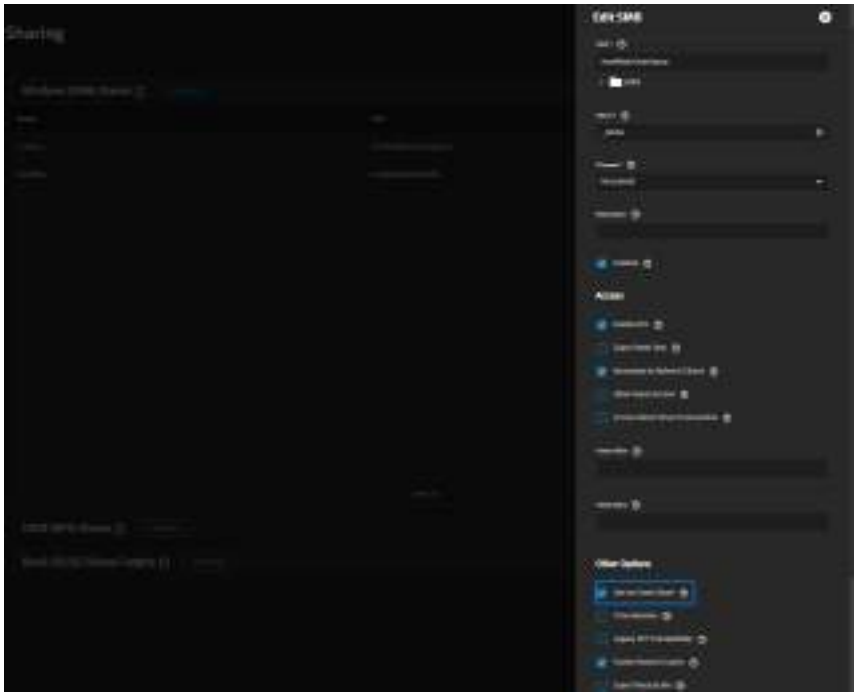
Uporabnikom je treba na nek način omogočiti dostop do podatkov, kar lahko dosežemo s kreiranjem skupnih map, prek katerih bodo uporabniki lahko dostopali do definiranih naborov podatkov.

Sistem nam pod zavihkom »Shares« omogoča kreiranje treh vrst različnih omrežnih dostopov do podatkovnega prostora sistema z uporabo različnih protokolov za učinkovito izmenjavo datotek:

- najbolj poznan in uporabljen je protokol SMB in kreiranje predvsem skupnih map za okolje Windows;
- druga možnost je kreiranje skupne mape na podlagi protokola NFS, ki je bolj razširjen v uporabi znotraj okolja UNIX;
- zadnja možnost je nekoliko drugačna in predstavlja skupni prostor iSCSI, kjer imamo možnost povezave virtualnega diska na uporabnikovem računalniku, ki se obnaša kot fizični trdi disk.

Implementiralo se bo predvsem skupne mape SMB, saj je ta protokol najbolj razširjen in enostaven za uporabo in dostopanje iz različnih operacijskih sistemov. Ker so se nabori podatkov že vnaprej definirali v drevesni strukturi, je treba kreirati skupno mapo le za najvišji nabor, ki predstavlja »Oddelki«.

Kot glavno mapo SMB pa se bo definiralo za osebne direktorije posameznikov. Imenovalo se jo bo »_Home« in bo glede na uporabniški račun kreirana podmapa, ki bo vidna samo dotičnemu uporabniku. To omogoča funkcionalnost SMB znotraj sistema, kjer v nastavitvah SMB skupne mape definiramo, da gre za domači direktorij uporabnika.



Slika 30: Konfiguracija uporabniške skupne mape

Vir: (Lastni vir)

S Slike 30 so razvidne nastavitve, kot so omejevanje glede na IP računalnika, dovoljenje dostopa gostom, omogočanje seznama za dostop (ACL), omogočanje senčnih kopij itd., kar prinaša še dodatne omejitve za posamezno skupno mapo SMB. Tako lahko izključno omogočimo dostop do skupne mape le iz enega izbranega ali več računalnikov, kjer nato ostali ne bodo morali dostopati.

Slika 31: Definirane skupne mape

Vir: (Lastni vir)

3.4 Zagotavljanje varnosti

Sistem je namenjen hranjenju različnih vrst podatkov, ki lahko segajo od nepomembnih do kritičnih vrst podatkov, katere pa je treba ustrezno zaščititi. Vendar varnost se ne začne le pri podatkih, ampak je treba korektno zaščititi tudi sisteme, na katerih se hranijo podatki.

V osnovi je pomembno zaščititi tudi fizične naprave sistema oziroma strežnik in omejiti, kdo vse ima fizičen dostop do njega. Tako omejimo tveganja kraje ali potencialnih napak sistema, če lahko do sistema dostopajo le strokovno podkovane osebe.

3.4.1 Varovanje sistema

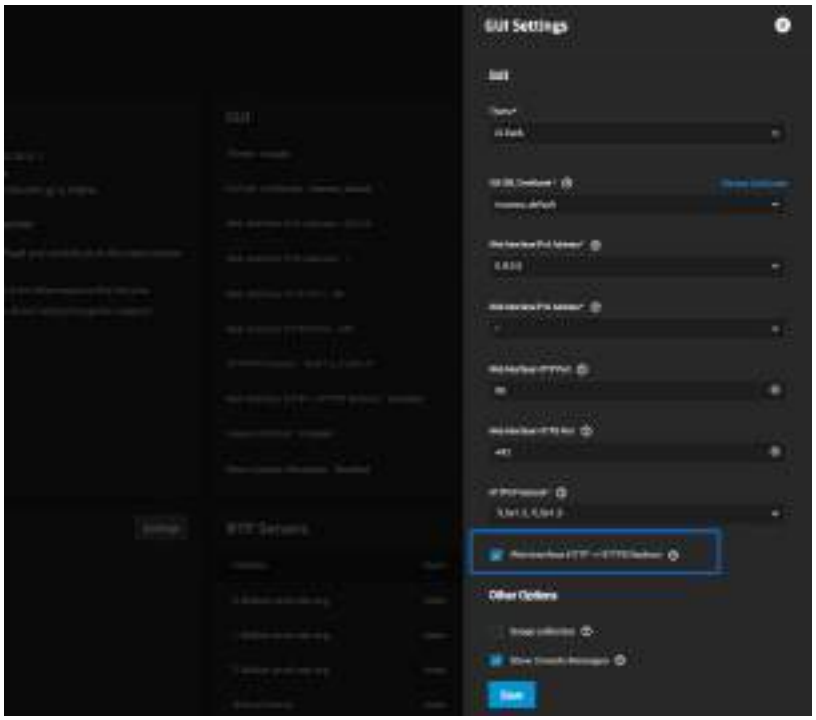
Varnost sistema se začne že pri namestiti, ko določamo uporabniško geslo administratorskega računa za dostop do sistema TrueNAS. To geslo mora vsebovati kompleksnost in korektno število znakov, saj lahko nepredviden dostop do administratorskega računa povzroči nezaželene učinke. Tveganje administratorskega računa enostavno izboljšamo s konfiguracijo večfaktorske avtentikacije, ki je že del sistema. Z uporabo večfaktorske avtentikacije omilimo potrebo po redni menjavi gesla, saj je znano, da uporabniki z vsako menjavo večinoma znižajo raven varnosti gesla.

Slika 32: Konfiguracija večfaktorske avtentikacije

Vir: (Lastni vir)

Sistem TrueNAS prav tako potrebuje omrežno povezavo za razširjeno delovanje (ni pogoj internetna povezljivost), kar dodatno razširja možna tveganja. Tovrstna tveganja se bodo omejevala z ustreznim določanjem pravic, kar bo pomenilo omejevanje dostopov znotraj omrežja.

Privzeto TrueNAS omogoča dostop tako prek povezave HTTP kot HTTPS. Ker je prva veliko bolj ranljiva, je smiselno nastaviti vsaj avtomatsko preusmeritev na HTTPS. To je enostavna možnost v glavnih nastavitvah TrueNAS pod nastavitvami grafičnega vmesnika.



Slika 33: Omogočanje preusmeritve na protokol HTTPS

Vir: (Lastni vir)

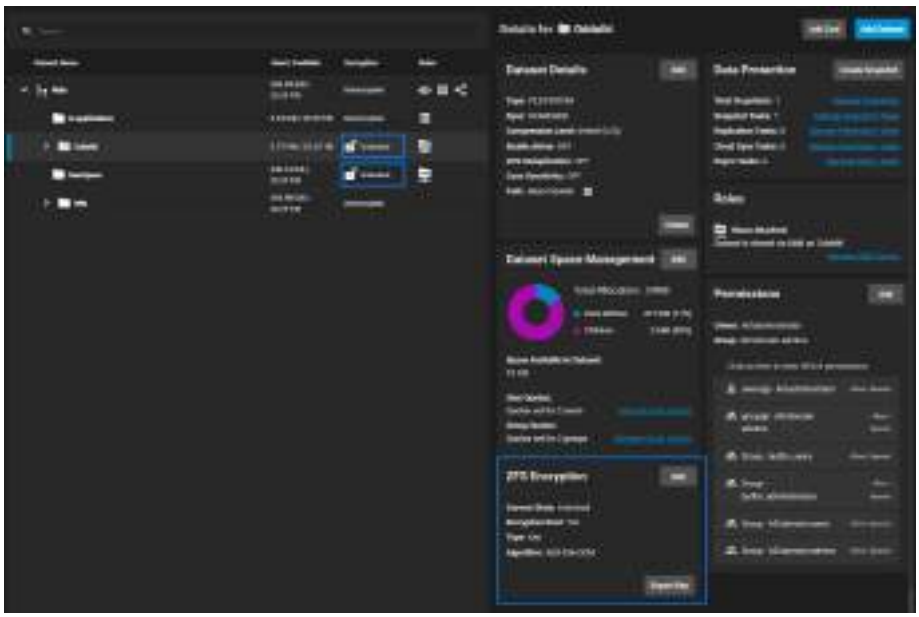
Za ustreznost implementacije HTTPS potrebujemo tudi certifikat SSL. To lahko rešimo znotraj TrueNAS, saj nam ta lahko generira certifikate oziroma že kot privzeto uporablja svojega generiranega. V primeru zunanjega dostopa do sistema pa je nujno uporabiti kvalificirano potrdilo SSL priznanega ponudnika (na primer Let's Encrypt, GoDaddy ipd.).

3.4.2 Varovanje podatkov

Podatki so ključni del takih sistemov in je treba zagotoviti kar se da visoko raven varnosti. To je seveda težje zagotavljati, saj vedno nekdo potrebuje dostop do nekaterih podatkov, kar prinaša dodatna tveganja. Osnova temelji na omejevanju dostopa do podatkov, vendar to ni dovolj. Večina si lahko predstavlja varovanje podatkov z določanjem različnih gesel na podatkih ali morda da so fizični mediji zaklenjeni v omari, vendar je varovanje podatkov veliko širši pojem in temelji predvsem na tem, kako zaščititi podatke za namen ustrezne rabe.

Zagotavljanje varnosti podatkov zajema tudi vse možnosti obnovitve podatkov ob nenamernem izbrisu. Zajema prav tako enkripcijo podatkov, ker doprinese dodatno varnost v primeru nedovoljenih dostopov ali kraje, saj brez ustreznega ključa ne morejo pridobiti podatkov.

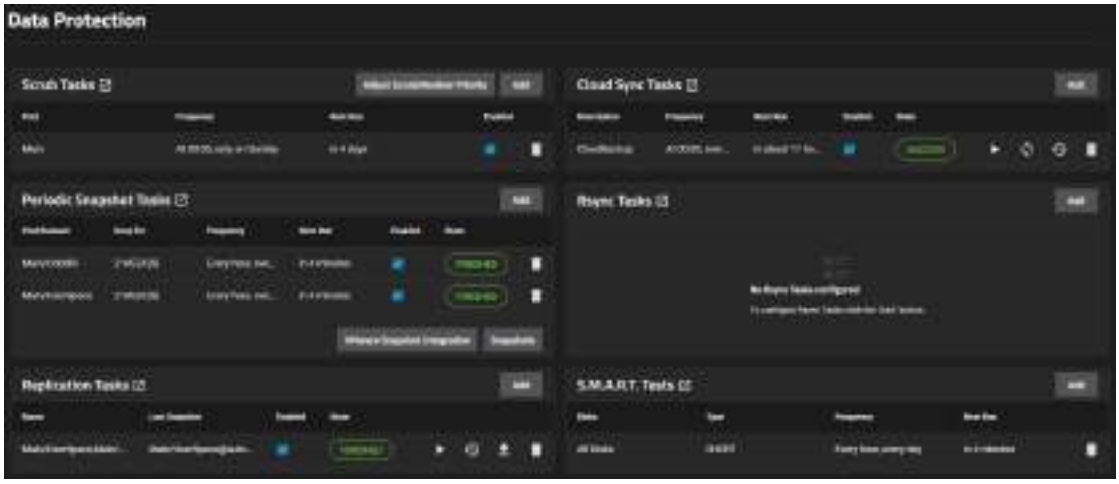
S tem namenom se je za glavne nabore podatkov implementirala tudi enkripcija, ki temelji na datotečnem sistemu ZFS.



Slika 34: Enkripcija naborov podatkov

Vir: (Lastni vir)

Sistem TrueNAS poudarja celoten pogled znotraj sistema, namenjen izključno varovanju podatkov, kjer se lahko definirajo različne aktivnosti za doseganje višje varnosti podatkov. Te aktivnosti temeljijo predvsem na različnih načinih arhiviranja, sinhronizacije, replikacije, čiščenju podatkov (ZFS) in izvajanju S.M.A.R.T. testov fizičnih diskov. Zadnji dve aktivnosti predvsem služita predčasni detekciji možnih okvar fizičnih diskov in pregledu njihove zanesljivosti, kot tudi identifikaciji težav integritete podatkov.



Slika 35: Aktivnosti varovanja podatkov v sistemu TrueNAS

Vir: (Lastni vir)

3.5 Arhiviranje ustvarjenih naborov podatkov

Arhiviranje podatkov bi lahko povezali znotraj varnosti podatkov, saj so vse aktivnosti arhiviranja povečanje stopnje varnosti podatkov. Arhiviranje nam omogoči možnosti izvajanja različnih kopij podatkov na izbrane lokacije, ki jih lahko izvajamo ob smiselnih intervalih.

3.5.1 Načini arhiviranja

TrueNAS omogoča različne načine arhiviranja, ki zaščitijo podatke v primeru neželenega izbrisa, nenamerne spremembe ali celo v primeru izsiljevalske programske opreme, ki nam kriptira naše podatke in lahko nato s pomočjo ustreznih kopij te podatke povrnemo.

Glavni načini arhiviranja znotraj TrueNAS:

- Periodični posnetki (angl. snapshots) predstavljajo opravilo, kjer se kreirajo načrtovane različice podatkovnega prostora in naborov podatkov v obliki samo za branje v danem trenutku.
- Sinhronizacija v oblaku (angl. Cloud Sync) predstavlja konfiguracijo enkratnih ali ponavljajočih se prenosov med TrueNAS in ponudnikom oblačnega prostora.
- Opravilo Rsync omogoča sinhronizacijo podatkov med dvema strežnikoma oziroma sistemoma. Opravilo rekurzivno bere po datotečnih strukturah in posodablja drugo lokacijo z vsemi novimi, spremenjenimi ali odstranjenimi datotekami. Namen opravila je predvsem olajšanje migracije na drug sistem.
- Replikacija (angl. Replication) predstavlja podobno logiko kot posnetki, le da gre tu za pravo arhiviranje. Kreira se opravilo replikacije, kjer določamo lokacijo, kam se bodo izbrani nabori podatkov replicirali in izvedli arhivsko kopijo. Se pravi, podatki se glede na konfiguracijo opravila kopirajo na drugo lokacijo, ki je na istem ali drugem sistemu.

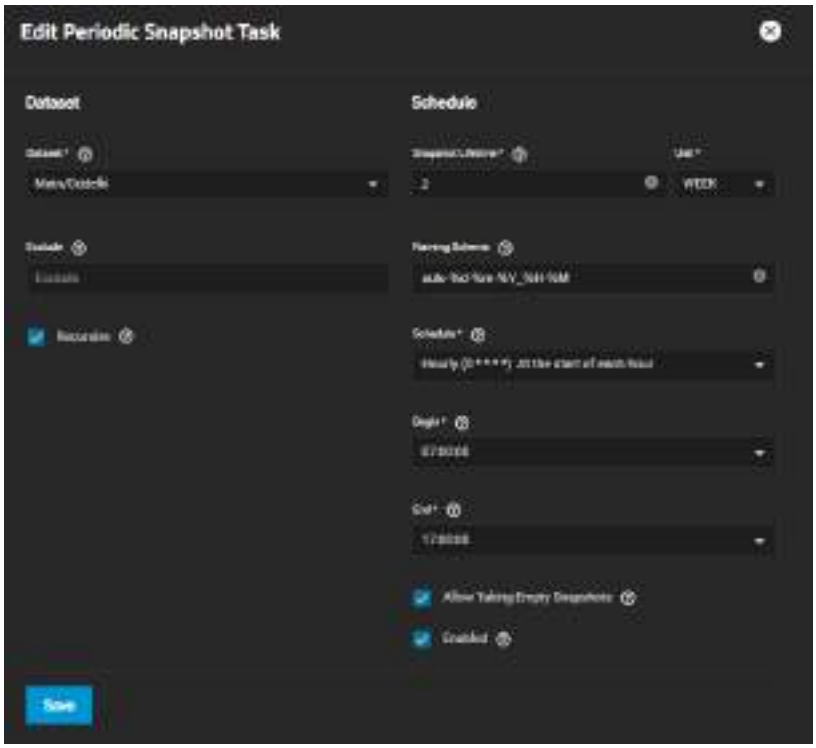
3.5.2 Implementacija arhiviranja

Dobro prakso arhiviranja predstavlja koncept 3-2-1, ki določa kreiranje treh kopij podatkov na dveh različnih fizičnih medijih in eno zunanjo lokacijo. Seveda se v modernih agilnih družbah upošteva že bolj kompleksne strategije kot 3-1-2, 3-2-2 in celo 3-2-3, kjer je poudarek na več zunanjih lokacijah v primeru potreb po obnovitvi ob katastrofah.

Seveda nič od teh strategij ni pogoj in je smiselno implementirati strategijo, ki bo najboljše razmerje med kompleksnostjo, ceno in varnostjo za posamezno podjetje oziroma implementirati najboljšo strategijo arhiviranja in varovanja podatkov, ki ga implementiran sistem omogoča.

Za implementacijo arhiviranja se bomo dotaknili treh od štirih prej izpostavljenih načinov, izpustili bomo le opravilo Rsync, saj je v trenutni obliki sistema skoraj brezpredmeten. Ostale tri bodo v sožitju dosegale relativno visoko raven varnosti podatkov implementiranega sistema.

Kot prvi korak se bosta konfigurirali dve opravili periodičnih posnetkov, in sicer na glavnih dveh naborih Oddelki in UserSpace, kjer bodo glavni uporabniški podatki. Opravili se bosta izvajali vsako uro, posnetki pa se bodo ohranjali za časovno obdobje dveh tednov.

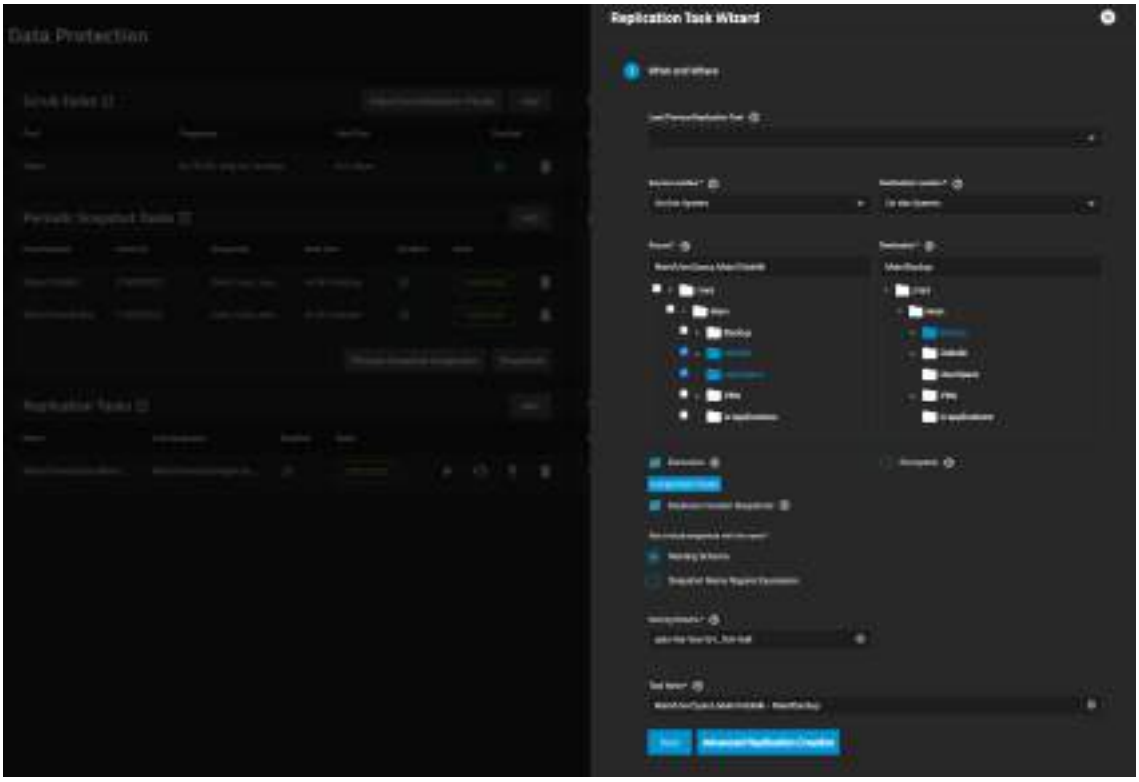


Slika 36: Kreiranje opravila periodičnih posnetkov

Vir: (Lastni vir)

Dodatno se lahko določa imenovanje in časovno obdobje, v primeru urnega izvajanja, znotraj katerega se bodo kreirali posnetki, se pravi lahko nastavimo izvajanje le znotraj delovnega časa organizacije. Ponujene periode zajemajo urno, dnevno, tedensko in mesečno izvajanje, lahko pa izberemo določanje periode po meri in povsem ročno določimo, kdaj se bodo zajemali posnetki.

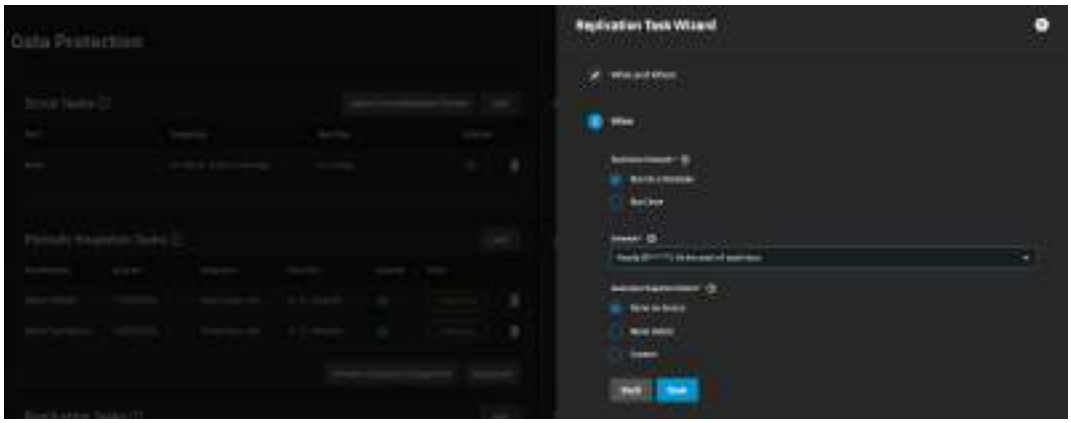
Druga stopnja arhiviranja bo konfiguracija replikacije. Ta bo služila kot dodatna bolj varna stopnja arhiviranja, kjer so bodo podatki fizično kopirali na drug določen namenski nabor podatkov. Za ta namen se je kreiral nov nabor Backup, ki je prav tako konfiguriran z enkripcijo.



Slika 37: Konfiguracija opravila replikacije

Vir: (Lastni vir)

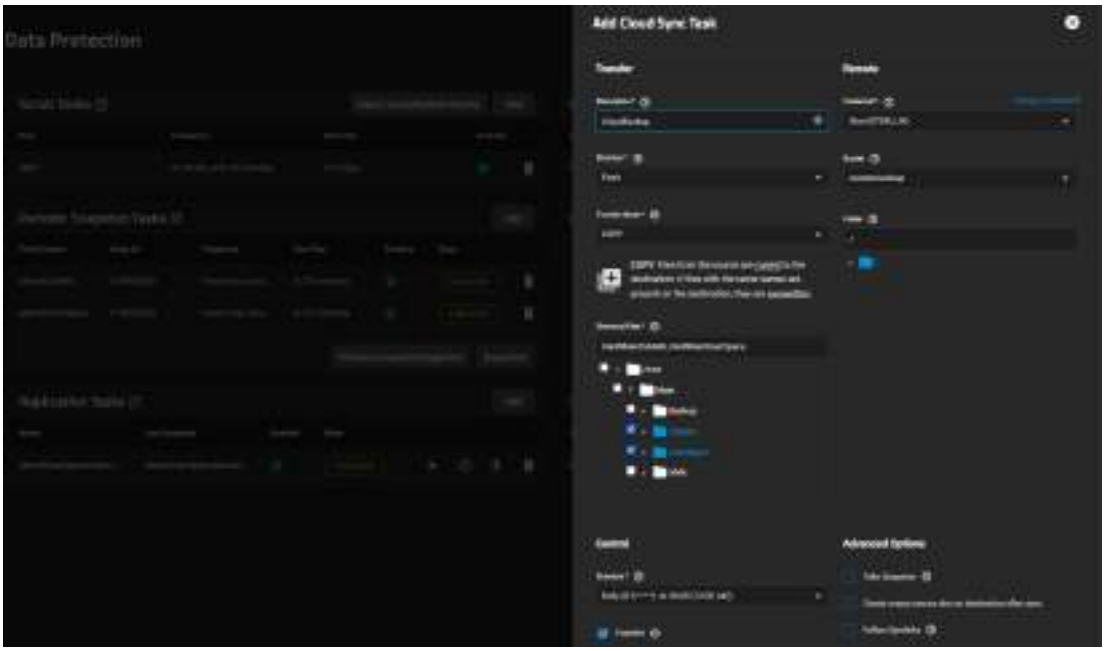
Replikacija se bo kot posnetki izvajala vsako uro, kjer bo politika hrambe podatkov določala prav tako časovno obdobje dveh tednov.



Slika 38: Konfiguracija periode izvajanja replikacije

Vir: (Lastni vir)

Zadnja aktivnost bo zajemala sinhronizacijo z oblačnim podatkovnim prostorom, kar bo predstavljalo našo zunanjo lokacijo. Ta se bo povezovala v oblačno storitev Storj, ki je rešitev proizvajalca TrueNAS. Opravilo bo definirano za izvajanje dnevno in bo enake podatke, ki smo jih zajeli pri replikaciji, sinhroniziralo v oblačni prostor, saj so izbrani nabori podatkov ključnega pomena.



Slika 39: Konfiguracija sinhronizacije v oblačni podatkovni prostor

Vir: (Lastni vir)

S tem se je implementiral relativno robusten sistem arhiviranja in varovanja podatkov. Tako bodo podatki varni pred raznovrstnimi tveganji izgube ali poškodb podatkov. Vsako uro se bo izvedlo več vrst kopij, vsakodnevno pa se bo glavno kopijo podatkov, kreirano z replikacijo, kopiralo na zunanji oblačni prostor za primere katastrof.

4 ANALIZA IMPLEMENTIRANEGA SISTEMA IN UGOTOVITVE

Implementacija sistema ni dovršena v primeru, da ne dosega zastavljenih ciljev in upošteva dogovorjenih politik in postopkov. S tem namenom je treba implementacijo ustrezno pregledati in analizirati v povezavi z realnimi testnimi primeri.

4.1 Testno okolje

Kreiralo se je enostavno domensko okolje z nekaj virtualnimi namizji različnih operacijskih sistemov. Okolje bo služilo kot podlaga za vse nadaljnje teste, s katerimi bomo preverjali ustreznost implementacije.

4.1.1 Testni sistemi

TrueNAS Scale omogoča tudi sistem virtualizacije, katerega smo za ta primer tudi izkoristili. Postavil se je domenski krmilnik Windows stražiškega okolja, kjer se je ustrezno namestila funkcionalnost aktivnega imenika. Za posamezne nabore podatkov so se definirale skupine uporabnikov za dostop do določenih področij oziroma oddelkov. Za vsako skupino uporabnikov se je definiralo po vsaj enega uporabnika.

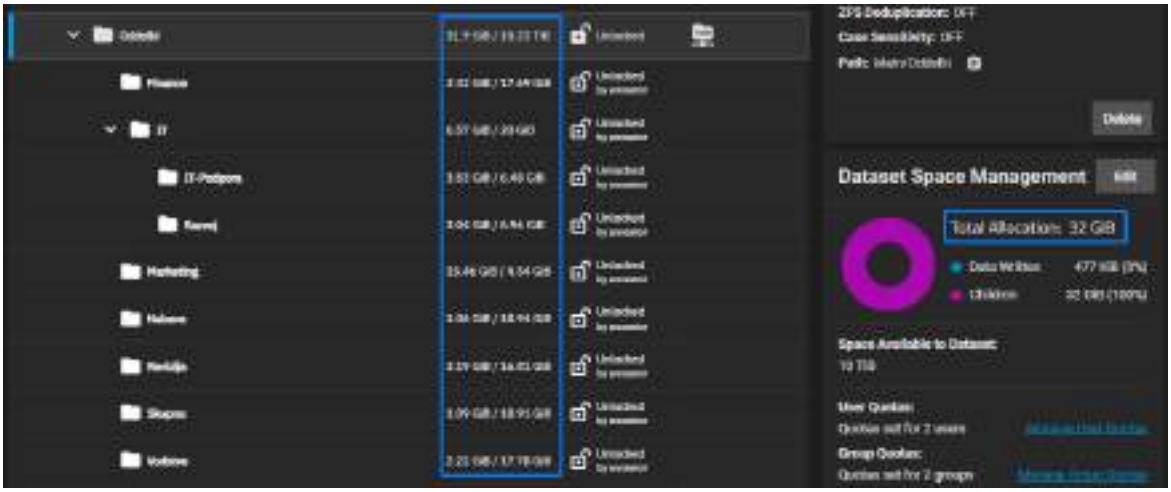
Za testiranje se je kreiralo še dve virtualni namizji, ki bosta simulirali uporabniške računalnike in dostope. Kreiralo se je en virtualni računalnik operacijskega sistema Windows 10 in en operacijskega sistema Linux Zorin. Vse skupaj definira naše simulacijsko domensko okolje.

Slika 40: Pripravljeni virtualni računalniki

Vir: (Lastni vir)

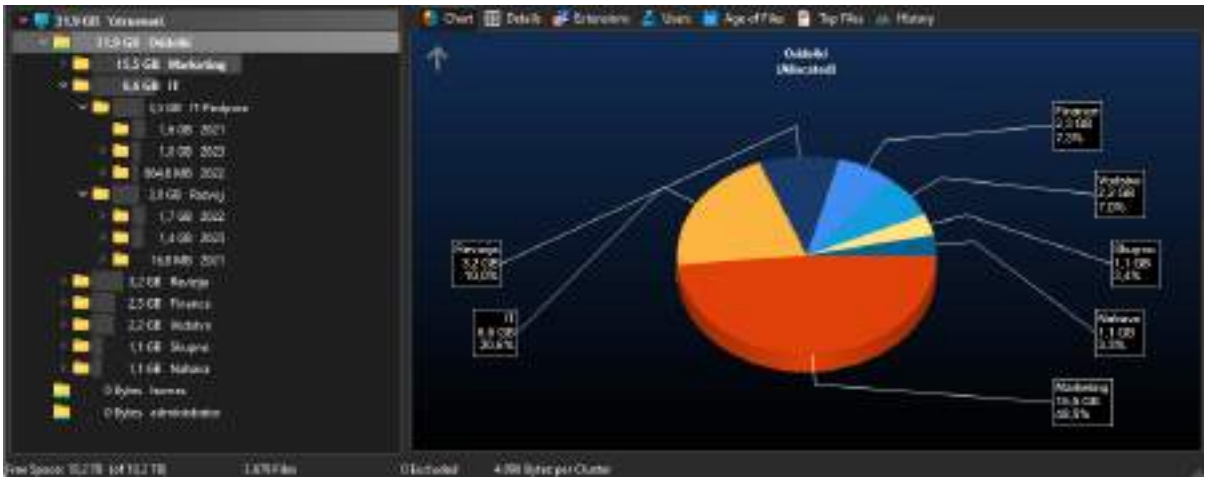
4.1.2 Priprava testnih podatkov

Sistem upravljanja s podatki je nesmiseln brez ustreznih količin podatkov. V ta namen se je prek portala Academic Torrents pridobilo večje količine podatkov različnih vrst, ki bodo odlično služile našim potrebam testiranja. Podatki, ki so na voljo na portalu, so namenjeni prav analizi in raziskovanju in so javno dostopni vsem. Vsebina podatkov se prav tako ne bo razkrivala, saj je nepomembna v taki obliki testiranja. Podatki so se razpršili naključno po definiranih naborih podatkov in tako dosežemo neko realno simulacijo okolja.



Slika 41: Pregled količine zapisanih podatkov

Vir: (Lastni vir)



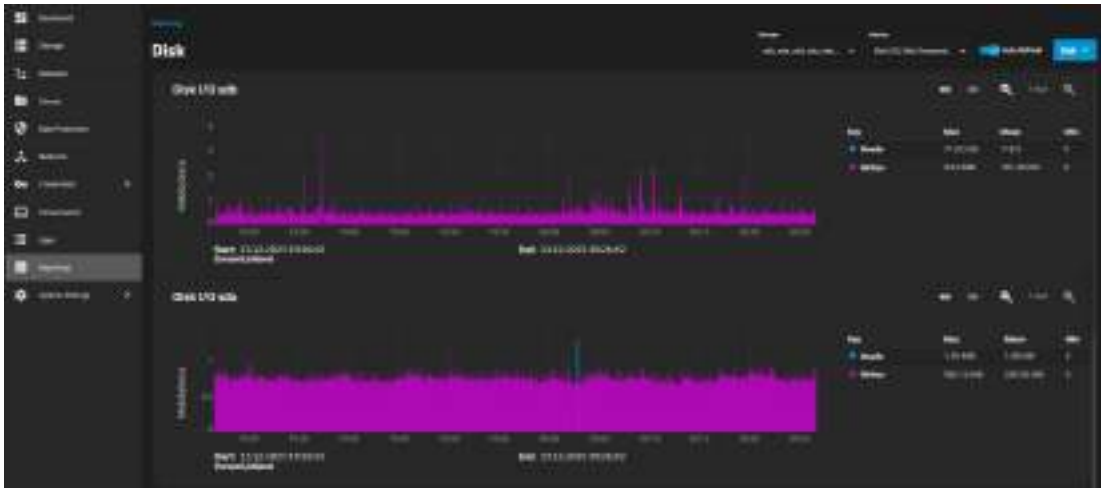
Slika 42: Podrobna analiza zasedenosti podatkovnega prostora

Vir: (Lastni vir)

Zapolnilo se je približno 32 GB podatkovnega prostora, porazdeljenega med nabori podatkov, kar predstavlja kar 3679 posamičnih datotek.

4.2 *Pregled delovanja implementiranega sistema*

Zapisovanje in branje večjih količin podatkov lahko vplivata tudi na odzivnost in delovanje sistema. Takšne obremenitve je treba spremljati, kar pa nam že v osnovi omogoča TrueNAS. Sistem ima vgrajen pogled poročanja, ki zajema glavne dele sistema, kot so procesor, delovni pomnilnik, omrežje, ZFS in vgrajeni diski.



Slika 43: Primer poročila obremenjenosti dveh diskov

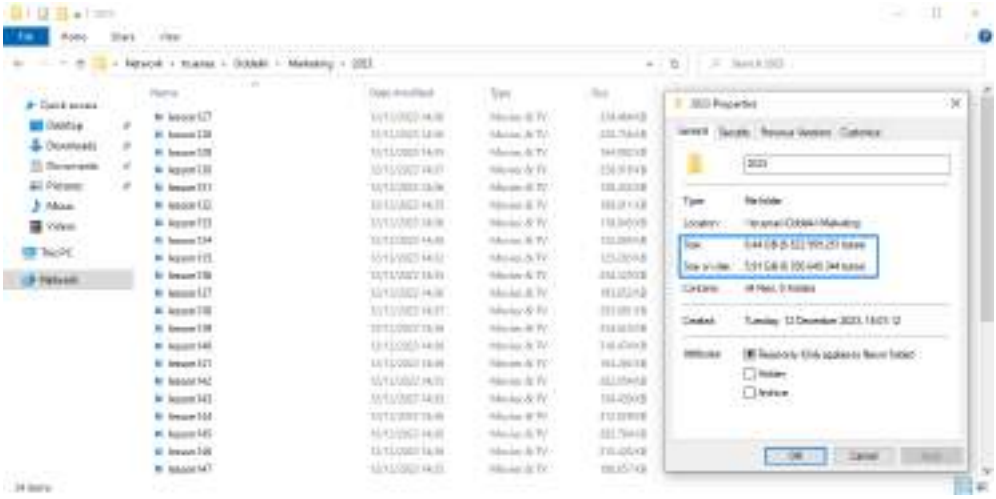
Vir: (Lastni vir)

TrueNAS omogoča tudi nameščanje aplikacij, ki dodatno razširijo funkcionalnost sistema. Na voljo je kopica različnih aplikacij, potrjenih s strani proizvajalca, lahko pa se seveda dodaja tudi druge. Na Sliki 44 je primer analitične aplikacije netdata, ki je privzeto na voljo v TrueNAS. Prikazuje glavni del nadzorne plošče ob izvajanju zapisov na podatkovni prostor. Ponuja veliko bolj podrobne informacije za analizo delovanja kot vgrajeno poročanje.

Slika 44: Analitična aplikacija netdata

Vir: (Lastni vir)

Osnovna konfiguracija naborov podatkov oziroma že samega podatkovnega prostora znotraj sistema TrueNAS privzeto omogoča kompresijo podatkov. Priporočen in privzeti algoritem kompresije predstavlja LZ4, ki predstavlja kompresijo brez izgub in je usmerjen v hitrost kompresije in dekompresije podatkov.

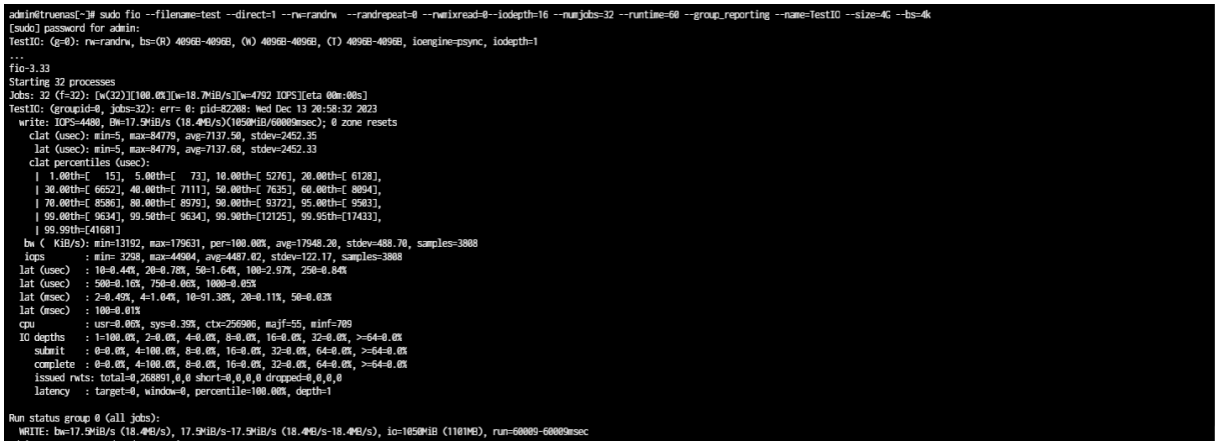


Slika 45: Prikaz kompresije LZ4 na primeru

Vir: (Lastni vir)

Pri številu 34 video datotek je uporaba kompresije prihranila okoli 530 MB podatkovnega prostora, brez izgube kakovosti zapisanih datotek.

Z uporabo Linux ukaza »fio« lahko v vgrajeni konzoli sistema TrueNAS poženemo teste zapisovanja podatkov oziroma teste I/O na ravni našega implementiranega sistema. Vrne nam okvirne najboljše podatke glede na parametre testa (velikost datotek, število datotek …).

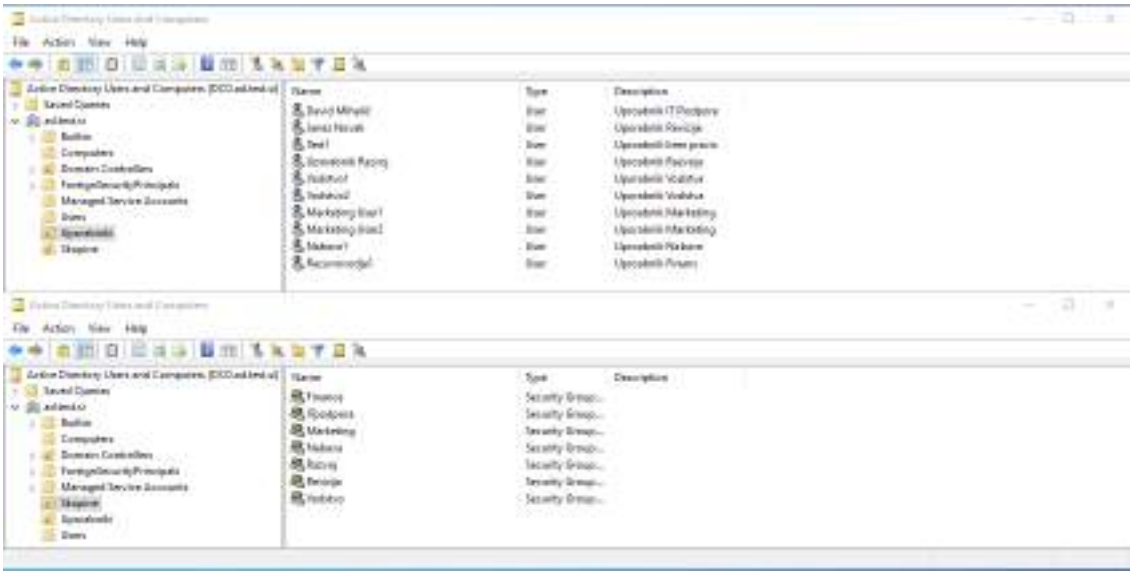


Slika 46: Primer testa I/O

Vir: (Lastni vir)

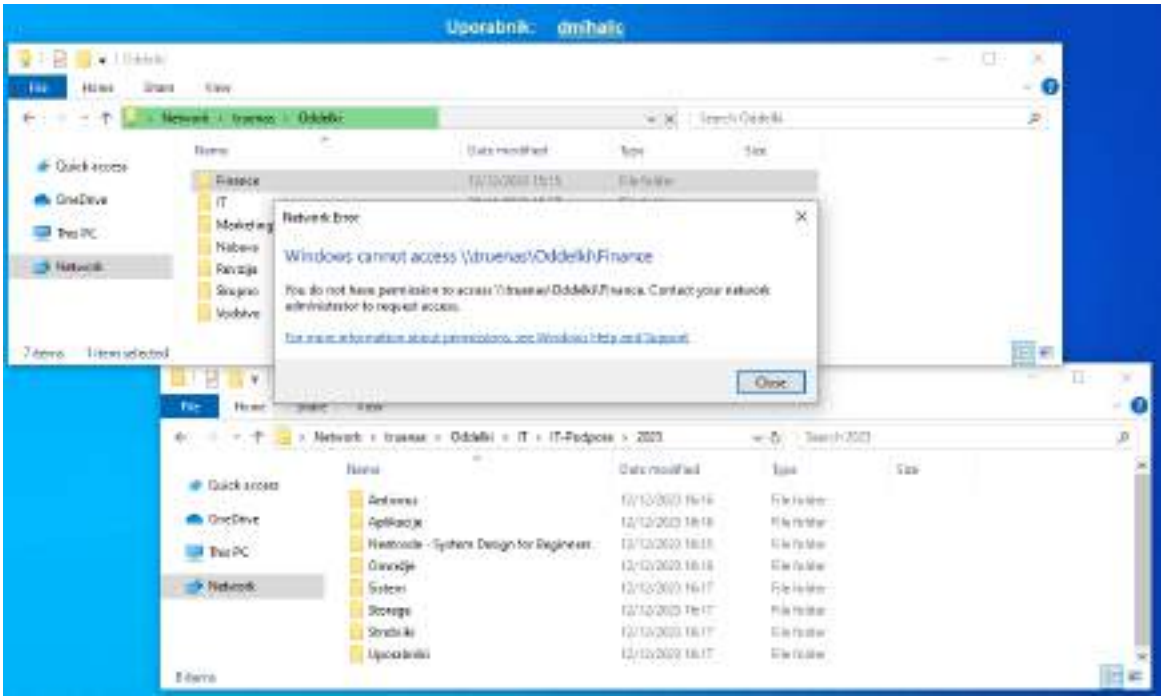
4.3 *Analiza uporabniških pravic in upravljanja s spremembami*

Določitev vseh uporabniških pravic je pomemben korak v določanju višje ravni kontrole nad podatki in odgovornosti posameznega uporabnika. Kot administratorji tovrstnih sistemov želimo ustrezno omejiti uporabnike na njihova področja, obenem pa pridobiti oziroma zagotoviti revizijsko sled sprememb podatkov.



Slika 47: Skupine in uporabniki domene

Vir: (Lastni vir)

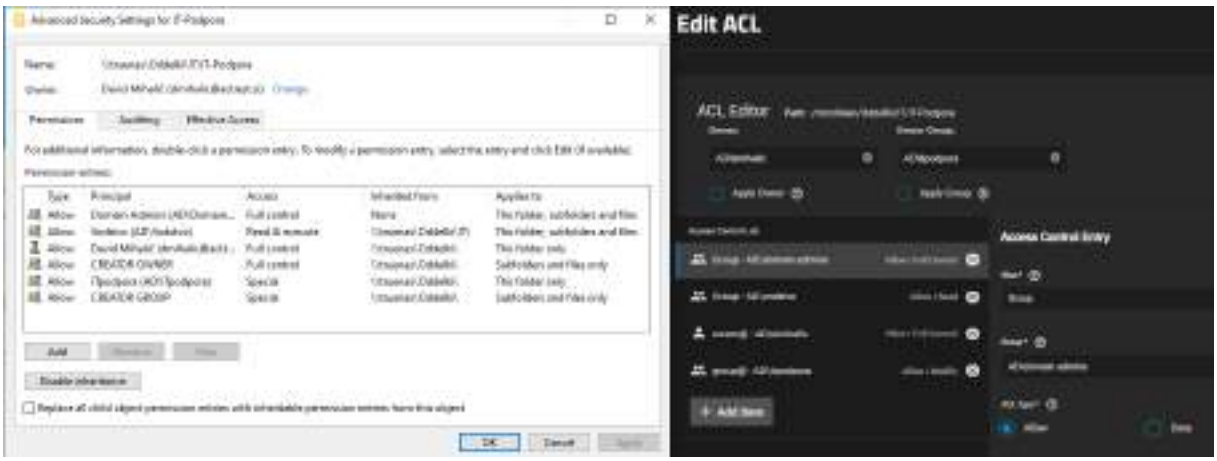


Slika 48: Primer analize ustreznosti uporabniških pravic

Vir: (Lastni vir)

Primer na Sliki 48 nakazuje akcije uporabnika »dmihalic«, kateremu smo dodelili pravice do podmape IT-Podpora. Ostali nabori podatkov ne dovoljujejo niti branja vsebine, saj uporabnik nima ustreznih pravic. Ustreznost določenih datotečnih pravic lahko izvedemo z vpogledom v lastnosti datoteke pod zavihkom varnost, kjer se nam izpiše seznam dodeljenih pravic.

Datotečne pravice na ravni operacijskega sistema primerjamo s pravicami, nastavljenimi na naboru podatkov v sistemu TrueNAS, in ugotovimo, da se pravice ujemajo identično, kar nakazuje ustrezno delovanje sistema upravljanja pravic znotraj sistema TrueNAS.

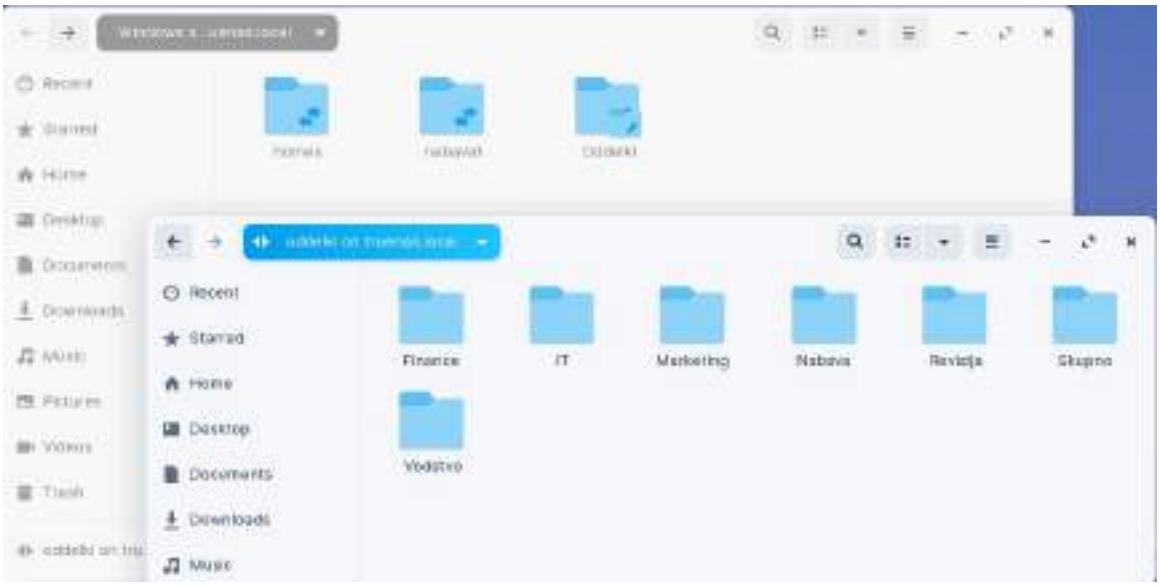


Slika 49: Primerjava datotečnih pravic Windows s pravicami nabora TrueNAS

Vir: (Lastni vir)

Znotraj sistema lahko enostavno in pregledno prek grafičnega vmesnika urejamo pravice posameznih naborov podatkov brez skrbi, da pravice ne bodo aktivne ob dostopih do podatkov. To privede do relativno solidnega upravljanja s pravicami datotek in tako doseže vse, kar smo definirali v sklopu hipoteze 7.

Enako velja za ostale operacijske sisteme, saj so večinoma zmožni operirati s protokolom SMB in dostopati do deljenih datotek. Na Sliki 50 je razviden tudi osnovni pogled ob omrežnem dostopu do našega podatkovnega strežnika TrueNAS na izbranem operacijskem sistemu Linux, avtomatsko kreirani mapi glede na uporabniško ime prijavljenega uporabnika (datoteka »homes« predstavlja identičen dostop kot mapa uporabniškega imena) in deljena mapa »Oddelki«, kjer so porazdeljene datoteke po posameznih področjih.



Slika 50: Dostop do datotek prek izbranega operacijskega sistema Linux

Vir: (Lastni vir)

Protokol SMB v okolju TrueNAS omogoča več ravni zajemanja dnevniških zapisov, ki pa so dostopni le prek konzole. Dnevniški zapisi v tej obliki so kar nepregledni in je iz teh težko razbrati potrebne podatke, zato je smiselno vzpostaviti strežnik upravljanja dnevniških zapisov, kateremu sproti posredujemo dnevniške zapise, ki jih sistem nato razčleni in poda v bolj berljivo obliko, kar pa privede do dodatnih razširitev izven našega sistema. Vseeno pa je možno razbrati različne vrste akcij nad podatki na podlagi vgrajene konzole. Na Sliki 51 je primer dnevniškega zapisa preimenovanja datoteke, kjer, če podrobno pogledamo, pridobimo podatke lokacije preimenovane datoteke, predhodno in novo imenovanje datoteke ter uporabnika, ki je akcijo preimenovanja izvedel.

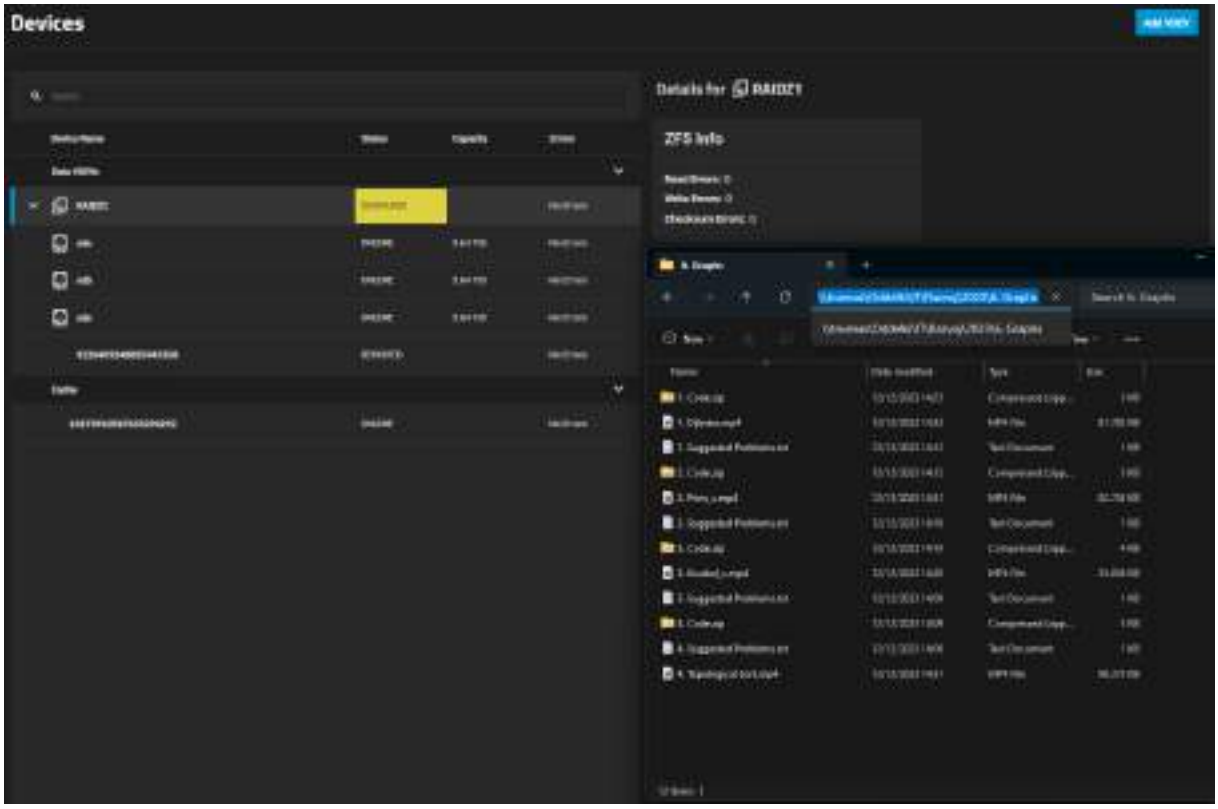
```
[2023/12/15 19:14:06.227367,  2] ../../source3/smbd/open.c:1713(open_file)
  AD\dmihalic opened file IT/IT-Podpora/2021/lesson22.mp4 read=No write=No (numopen=6)
[2023/12/15 19:14:06.229810,  3] ../../source3/smbd/smb2_trans2.c:5071(smbd_do_setfilepathinfo)
  smbd_do_setfilepathinfo: IT/IT-Podpora/2021/lesson22.mp4 (fnum 584080744) info_level=65290 totdata=110
[2023/12/15 19:14:06.230595,  3] ../../source3/smbd/smb2_reply.c:1776(rename_internals_fsp)
  rename_internals_fsp: succeeded doing rename on IT/IT-Podpora/2021/lesson22.mp4 -> IT/IT-Podpora/2021/SpremembaImenaDatoteke.mp4
```

Slika 51: Dnevniški zapisi sprememb v konzoli

Vir: (Lastni vir)

4.4 Pregled možne okvare diska

Datotečni sistem ZFS deluje v različnih oblikah namenskih polj RAID, ki v svoji konfiguraciji in načinu postavitve že omogočajo preventive ob okvari diskov. V implementaciji smo konfigurirali naš podatkovni prostor na raven RAIDZ1 polju, ki »dovoljuje« okvaro enega trdega diska in tako preprečuje nenapovedano izgubo podatkov. Sistem ima tudi vgrajene aktivnosti izvajanja različnih preventivnih testov za predčasno identificiranje možnih okvar (npr. testi S.M.A.R.T.). V primeru konfiguracije RAIDZ2 bi naš sistem imel možnost preživeti okvaro dveh diskov hkrati, vendar vsaka dodatna pariteta zelo vpliva na končno količino podatkovnega prostora.



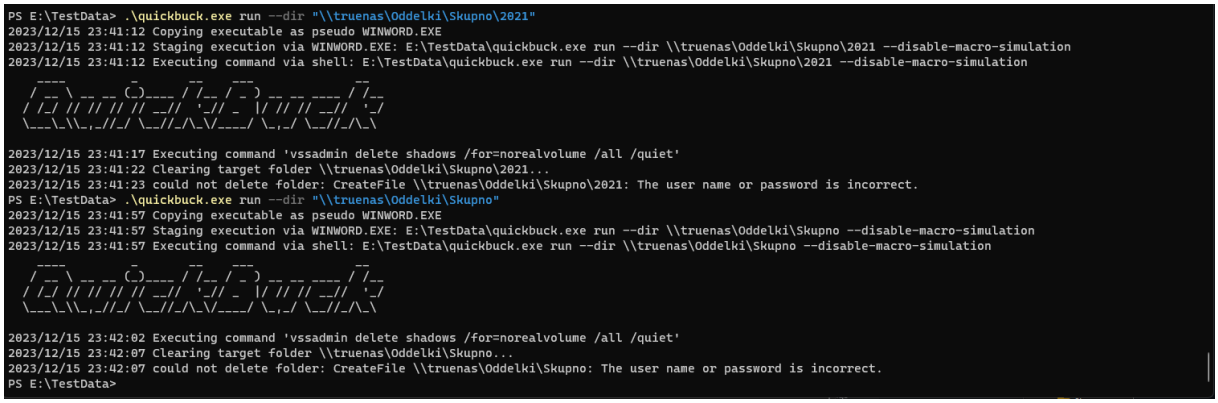
Slika 52: Simulacija izgube enega fizičnega diska polja

Vir: (Lastni vir)

Sistem zazna, da je bil eden od diskov odstranjen (simulacija s fizičnim izklopom podatkovnega kabla), kjer so podatki nemoteno dostopni. Na sistemu ni bilo zaznane nobene prekinitve, le da se je spremenilo stanje polja podatkovnega prostora. Ko zamenjamo okvarjen disk, sistem avtomatsko ponudi možnosti za popravilo stanja polja podatkovnega prostora oziroma že sam uredi stanje na prvotno. Kot rezultat testa smo obenem potrdili trditev, navedeno v hipotezi 5.

4.5 Postopki obnovitve podatkov

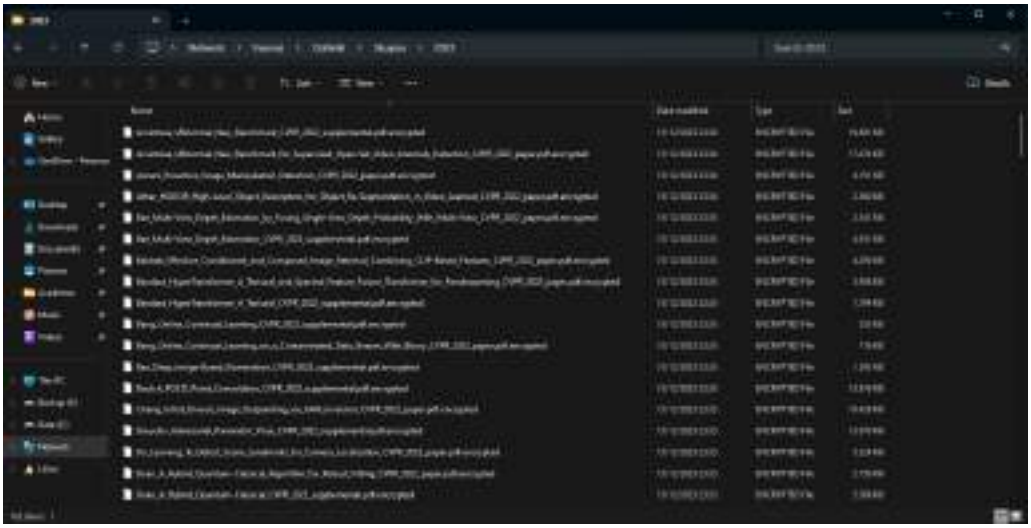
Obnavljanje podatkov predstavlja eno ključnih aktivnosti pri upravljanju podatkov, saj se nenehno lahko pripetijo neželene akcije nenamernih ali namernih izbrisov podatkov, tveganja v povezavi s kripto virusom, ki kriptira vse podatke, do katerih lahko pride, in pa ponesrečene spremembe datotek uporabnikov, kjer se želi povrniti določeno stanje datotek. Ker smo implementirali različne načine arhiviranja oziroma varovanja datotek, smo tako tovrstna tveganja krepko zmanjšali.



Slika 53: Poizkus okužbe z izsiljevalsko programsko opremo

Vir: (Lastni vir)

Simulacija izsiljevalske programske opreme ni mogla doseči nič in takoj javila napako neustreznih uporabniških pravic. Bolj sofisticirane izsiljevalske programske opreme bi verjetno naredile kaj več škode, zato bomo simulirali, da so se nam v eni izmed map kriptirale vse datoteke in te niso več berljive.

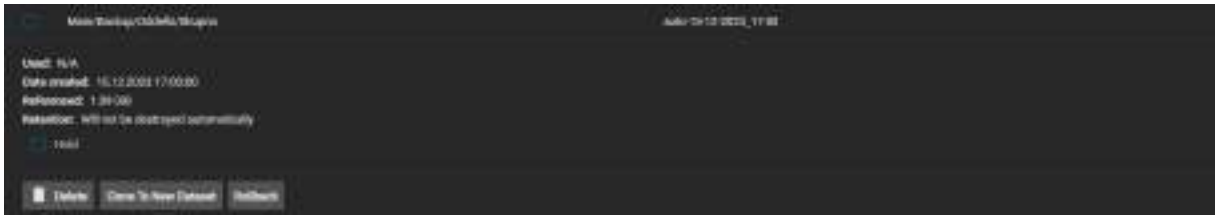


Slika 54: Kriptirane datoteke

Vir: (Lastni vir)

4.5.1 Obnovitev podatkov iz periodičnih posnetkov

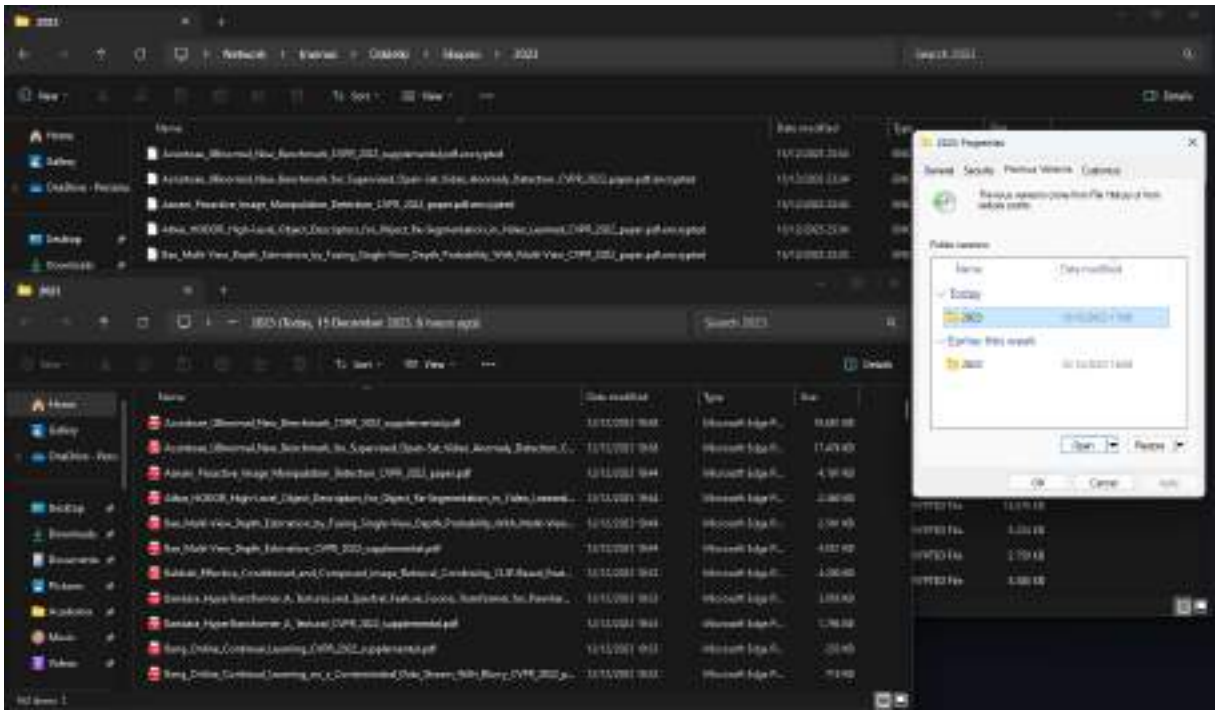
Znotraj sistema smo definirali redno enourno zajemanje posnetkov, ki beležijo posamezne spremembe. Tu imamo ponujenih več možnosti. Znotraj sistema lahko izvedemo akcijo povrnitve, podatke lahko kloniramo na ločen nabor podatkov (v primeru, da se potrebuje le neko datoteko) ali pa vse to izvedemo na ravni operacijskega sistema, kjer obstaja v lastnostih datotek v okolju Windows zavihek preteklih različic, kjer imamo enake možnosti kot v sistemu.



Slika 55: Povrnitev stanja na začetno s pomočjo sistema TrueNAS

Vir: (Lastni vir)

V okolju Windows lahko enostavno odpremo eno od predhodno zajetih različic in izberemo samo določene datoteke ali pa restavriramo celotno vsebino mape.

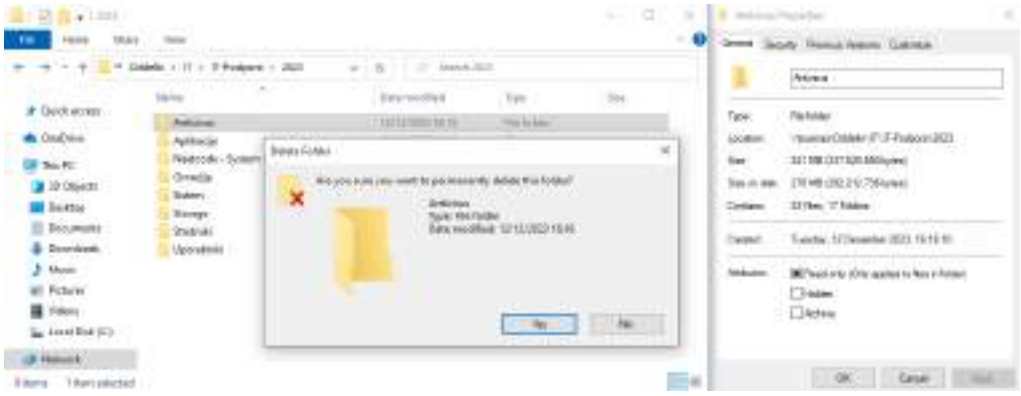


Slika 56: Povrnitev stanja s pomočjo operacijskega sistema

Vir: (Lastni vir)

4.5.2 Obnovitev podatkov iz oblačnega podatkovnega prostora

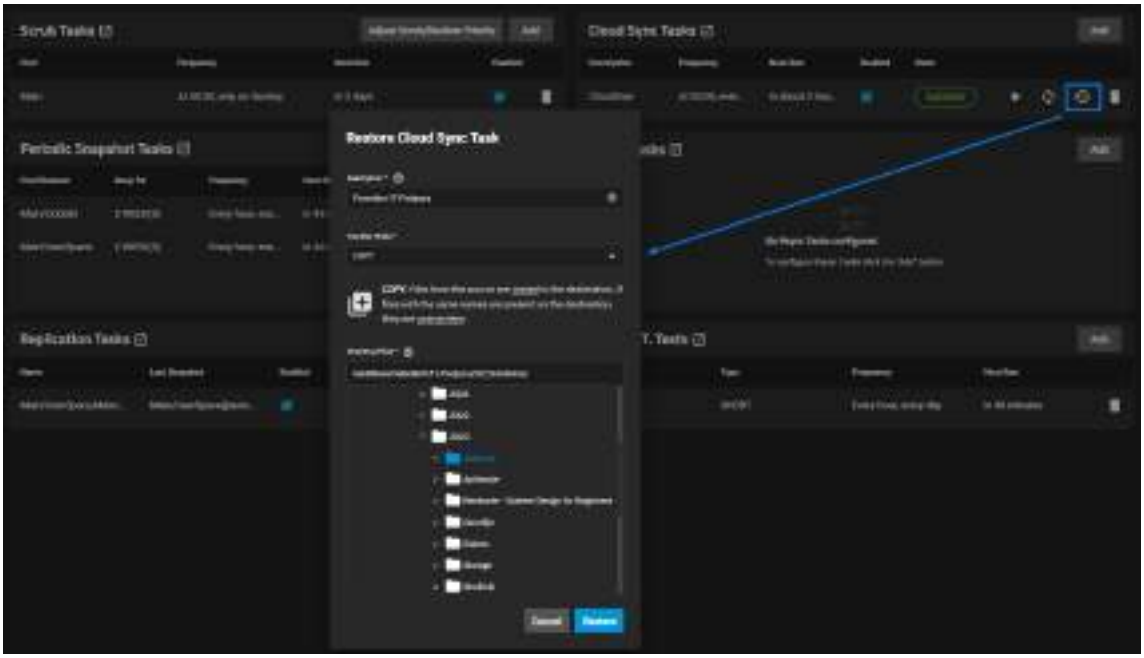
Konfiguriralo se je prav tako tudi arhiviranje na zunanjo lokacijo, ki dnevno izvaja arhivske kopije podatkov v izbrano oblačno storitev. To postane uporabno, ko celoten sistem postane nedosegljiv ali je žrtev visoko tveganega virusnega napada, ki omeji dostop do večine podatkov.



Slika 57: Izbris določene datoteke za test obnovitve

Vir: (Lastni vir)

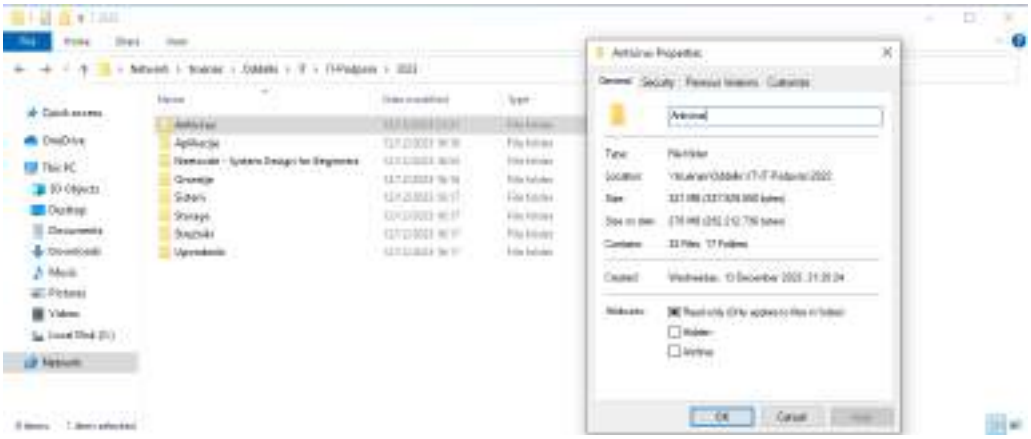
Na pogledu varnosti podatkov pričnemo postopek obnovitve za izbrano datoteko. Kreira se nov proces, ki mu moramo določiti ime in način, ki se bo uporabil za obnovitev (kopiranje ali sinhronizacija).



Slika 58: Postopek obnovitve podatkov z oblaka

Vir: (Lastni vir)

Ob končanem procesu obnovitve so ponovno vidni izbrisani podatki. Razberemo lahko tudi spremembo v kreiranem datumu datoteke, ki je vezan na datum obnovitve, in ko to primerjamo s stanjem na Sliki 57, je lepo razvidno, da gre za obnovljeno datoteko.



Slika 59: Stanje po postopku obnovitve

Vir: (Lastni vir)

Na strani oblačne storitve prav tako lahko nemoteno dostopamo do zadnjega stanja kopiranih podatkov. Do teh lahko dostopamo kjer koli, saj gre za oblačno storitev, le da za dostop do posameznih dokumentov potrebujemo ustrezno definiranega uporabnika in dodatno še geslo v obliki kombinacije različnih besed, ki doseže večnivojsko zaščito do podatkov. Tako pridobimo dodatno možnost dostopa do podatkov, ki je lahko glavni rešitelj v primeru nedostopnosti glavnega sistema.

Slika 60: Pregled podatkov znotraj oblačne storitve

Vir: (Lastni vir)

V celotnem sklopu poglavja smo obravnavali predvsem primere, katerih rezultati predvsem potrjujejo zastavljene hipoteze 2, 3 in 6.

5 KRITIČNA OCENA IN ZAKLJUČKI

Celovitost sistema temelji na različnih komponentah, ki v kombinaciji določajo ustrezno raven upravljanja podatkov. Sistem TrueNAS v povezavi s komponentami in postopki, ki jih uporablja, omogoča širok nabor primerov uporabe, ne samo v kontekstu podatkov, ampak se v določeni obliki konfiguracije lahko obnaša tudi kot celovit poslovni sistem. To je lahko ključnega pomena za nekatera manjša podjetja, ki si ne morejo zagotoviti razpršenih sistemov.

V implementacijo sistema smo se podali predvsem z namenom upravljanja podatkov, vendar za namen analize smo izkoristili dodatne funkcionalnosti sistema in definirali delujočo simulacijo poslovnega okolja.

Analize lahko hitro terjajo veliko resursov in časa, kar je težko strniti v določene smiselne teste, da se pridobi ustrezne rezultate, ki jih lahko zapakiramo v neke vrste strnjeno pregledno obliko. Podana analiza je zajela predvsem najbolj pomembne in smiselne primere, ki izpostavljajo ustreznost ali neustreznost delovanja implementiranega sistema. Vseeno lahko trdimo, da bi bilo pomembno analizo krepko razširiti in implementiran sistem soočiti z veliko bolj razširjenim naborom možnih tveganj in primerov, da bi lahko še bolj izpostavili prednosti in morda tudi slabosti implementiranega sistema.

Postopek implementacije in analiza v našem primeru nista temeljila na nobenem konkretnem primeru iz realnega sveta, kar predvsem omeji doseganje ustreznega sistema za neko poslovno okolje, saj zadevo obravnavamo bolj na podlagi teorije kot resnih primerov iz realnega poslovnega okolja. Vsekakor pa se je tako v implementaciji kot tudi v analizi poskušalo zajeti glavne temelje poslovnega okolja in pomembne vidike upravljanja podatkov, ki so povsem smiselni za osnovno poslovno okolje. Prednosti implementiranega sistema sta predvsem njegova modularnost in nadgradljivost, ki omogoča širjenje tudi izven osnovnega sistema, kar je lahko velika prednost poslovnih okolij, ki se hitro razširijo. Pričnejo lahko z osnovnim enotnim okoljskim sistemom, ki ga enostavno razširijo svojim potrebam.

Zajeta raziskava in implementacija lahko služita kot osnovno orodje za implementacijo sistema v drugih poslovnih in domačih okoljih, kjer se želi doseči vsaj osnovno raven celovitosti upravljanja s podatki.

6 SKLEP

V diplomskem delu smo obravnavali implementacijo sistema za doseganje celovitosti upravljanja s podatki. Za potrebe izvedbe tovrstnega sistema se je uporabilo sistem TrueNAS, ki predstavlja priznano namensko odprtokodno platformo za raznovrstne potrebe hranjenja in upravljanja podatkov. Diplomsko delo lahko praktično razdelimo na tri ključne segmente, ki predstavljajo teoretični del, implementacijo sistema in analizo implementacije.

Teoretični del predvsem navaja, definira in opredeljuje vse ključne gradnike ali komponente, ki so pomembni za doseganje ustrezne ravni celovitosti upravljanja s podatki. Poudarek je bil predvsem na varnosti, ki je vse večji izziv posameznikov in organizacij, ki se morajo konstantno prilagajati novim varnostnim tveganjem.

Drugi del se osredotoča na celoten postopek implementacije in konfiguracije ključnih komponent sistema TrueNAS za doseganje zastavljenih potreb za upravljanje s podatki. Navaja tudi glavne informacije o izbrani platformi in izpostavlja vse njene prednosti.

Za konec se je na različnih smiselnih primerih izvedlo analizo implementiranega sistema, na podlagi katere se je lahko izpostavila ustreznost delovanja sistema ter opredelilo rezultate za doseganje celovitosti upravljanja s podatki in ovrednotenje zastavljenih hipotez.

Glavno raziskovalno vprašanje »Ali je možno implementirati sistem za celovito upravljanje s podatki?« nas je spremljalo skozi celotno pot raziskave. Na podlagi zastavljenega vprašanja smo izpeljali tudi osem hipotez, ki so večinoma usmerjene v namen upravljanja s podatki na različnih ravneh.

Hipoteza 1, ki pravi, da je treba »določiti ustrezne politike, ki veljajo za zbiranje, shranjevanje, obdelavo in odstranjevanje podatkov«, se potrjuje že s trditvami, izpostavljenimi v teoretičnem delu naloge, kjer smo definirali komponente celovitega upravljanja s podatki. Povežemo jo lahko predvsem s komponento nadzora nad upravljanjem s podatki, ki zajema prav trditev, navedeno v hipotezi. Brez definirane komponente nadzora tako ne moremo govoriti o celovitosti sistema in je nujno potrebna za doseganje ustrezne ravni upravljanja podatkov.

Hipotezo 2, ki definira, da je »v primeru izgube (nezaželenega izbrisa) podatkov te možno povrniti«, lahko potrdimo na podlagi izvedenih testov na področju obnovitev podatkov. Implementiralo se je tri različne načine arhiviranja podatkov, ki vsak na svoj način služi namenu potreb po obnovitvi podatkov. Na primeru obnovitve podatkov iz oblačnega podatkovnega prostora se je obravnavala prav akcija izbrisa podatkov, kjer se je prikazal postopek obnovitve.

V primeru hipoteze 3, ki pravi, da »sistem preprečuje nezaželene spremembe podatkov ob varnostnem incidentu (npr. kripto virus)«, bi to lahko utemeljili prav tako na podlagi izvedene analize. Rezultati analize hipotezo 3 potrjujejo. Na podlagi analize v poglavju definicije postopkov obnovitev podatkov se je praktično izpostavil primer simulacije okužbe z izsiljevalsko programsko opremo, ki do podatkov ni uspela dostopati, za kar sta zaslužna enkripcija naborov podatkov in določanje ustreznih uporabniških pravic.

Hipotezo 4, ki definira, da »sistem omogoča upravljanje in analizo kompleksnejših vrst ter velikih količin podatkov«, lahko potrdimo na podlagi celotne implementacije in analize. Osnova za potrditev temelji že na uporabljeni platformi TrueNAS in predvsem datotečnem sistemu ZFS, ki ponuja metode omogočanja upravljanja kompleksnih in velikih količin podatkov. Dodatno pa hipotezo opredeljuje postopek analize, kjer se je implementiran sistem zapolnil z različnimi vrstami podatkov, velikih količin (angl. Big Data), katere se je nemoteno upravljalo na podlagi zastavljenih primerov analize.

Hipotezo 5, ki trdi, da »sistem prenese izpad enega fizičnega medija za shranjevanje podatkov«, na podlagi testnega primera potrdimo. Test simulacije okvare trdega diska ni imel vpliva na delovanje sistema ali dostopa do podatkov, saj se je v implementaciji sistema podatkovni prostor konfiguriralo na podlagi podatkovnega polja RAIDZ1, ki vzame en trdi disk kot pariteto in tako dovoljuje okvaro enega trdega diska.

Hipotezo 6, ki trdi, da »sistem omogoča varen dostop do podatkov izven lokalnega omrežja«, se lahko potrdi na podlagi oblike implementacije. Ker smo v sklopu implementacije definirali postopek arhiviranja podatkov izven lokalnega omrežja v oblačno podatkovno storitev, smo omogočili dostop do podatkov izven lokalnega omrežja. Dodatno smo v postopku analize izpostavili način, potreben za dostop do podatkov v oblačni storitvi, ki zahteva večnivojsko avtentikacijo, kar predstavlja višjo raven varnosti.

Hipoteza 7 navaja, da »sistem ne dovoljuje sprememb in dostopa do podatkov v primeru neustreznih pravic«. Hipoteza se potrdi na podlagi izvedenih testov analize uporabniških pravic dostopa. Iz določenih primerov uporabniškega dostopa do podatkov je bilo razvidno, da v primeru neustreznih pravic ne morejo dostopati do podatkov in njihove vsebine. Primerjava sistema upravljanja uporabniških pravic TrueNAS z datotečnimi pravicami, vidnimi znotraj naprave Windows, je pokazala, da se podatki ujemajo.

Hipotezo H8, ki trdi, da »sistem omogoča enostavne nadgradnje tako sistema kot strojne opreme (angl. »end-of-life«), lahko potrdimo na podlagi izbrane platforme za implementacijo in uporabljene strojne opreme. Kot smo izpostavili v postopku implementacije TrueNAS, že na podlagi glavne nadzorne plošče omogoča enostavno preverbo možnih novejših posodobitev sistema. V primeru strojne opreme pa so se uporabile široko dostopne potrošniške komponente, ki so enostavno zamenljive za bolj zmogljive ali novejše komponente, saj uporabljena strojna oprema predstavlja v osnovi običajen računalnik oziroma strežnik.

Največjo omejitev raziskave je predstavljala analiza implementacije sistema, saj je bilo treba pridobiti velike količine podatkov, da lahko ustrezno govorimo o upravljanju podatkov in izpostavimo zmogljivosti sistema ter izvedemo vse zastavljene korake analize. Ker diplomsko delo ne temelji na konkretnem primeru poslovnega okolja, nismo imeli podlage za uporabo določenih podatkov organizacije. Pridobitev teh podatkov je tako morala temeljiti na javno dostopnih podatkih, ki se lahko uporabijo za namene raziskav in analiz.

Ugotovitve na podlagi hipotez, implementacije sistema in njegove analize v povezavi z definiranimi komponentami upravljanja podatkov teoretičnega dela kažejo, da je sistem za celovito upravljanje s podatki možno implementirati, saj so se zajele vse posamezne komponente upravljanja s podatki vsaj v neki osnovni obliki.

VIRI, LITERATURA

CompTIA. (19. Julij 2023). *What Is Data Analytics: The Ultimate Guide.* Pridobljeno iz CompTIA: https://www.comptia.org/content/guides/what-is-data-analytics

Cyral. (2023). *Data Access Control.* Retrieved from Cyral: https://cyral.com/glossary/data-access-control/

DAMA International. (2017). *DAMA Guide to the Data Management Body of Knowledge.* Technics Publications.

Davoy. (15. November 2022). *What Is Data Architecture?* Pridobljeno iz Davoy: https://davoy.tech/what-is-data-architecture/

Ekran System. (16. Avgust 2023). *10 Data Security Best Practices: Simple Methods to Protect Your Data.* Pridobljeno iz Ekran: https://www.ekransystem.com/en/blog/data-security-best-practices

Evans, N., & Price, J. (2012). *Barriers to the Effective Deployment of Information Assets: An Executive Management Perspective.* Interdisciplinary Journal of Information, Knowledge, and Management Volume 7.

Frankenfield, J., Boyle, M. J., & Rathburn, P. (9. Avgust 2023). *Data Analytics: What It Is, How It's Used, and 4 Basic Techniques.* Pridobljeno iz Investopedia: https://www.investopedia.com/terms/d/data-analytics.asp

Heavy.ai. (2022). *Data Integration.* Pridobljeno iz Heavy.ai: https://www.heavy.ai/technical-glossary/data-integration

IBM. (2023). *What is data modeling?* Pridobljeno iz IBM: https://www.ibm.com/topics/data-modeling

Imperva. (2023). *Data Governance.* Pridobljeno iz Imperva: https://www.imperva.com/learn/data-security/data-governance/

Inden d.o.o. (2019). *Zapis podatkov v podatkovno jezero (Data Lake)*. Pridobljeno iz Inden: https://www.inden.si/zapis-podatkov-v-podatkovno-jezero-data-lake/

iXSystems. (2023). *TrueNAS SCALE Getting Started.* Pridobljeno iz TrueNAS: https://www.truenas.com/docs/scale/printview/

Johnson, J. (4. Marec 2021). *Data Architecture Explained: Components, Standards & Changing Architectures.* Pridobljeno iz BMC: https://www.bmc.com/blogs/data-architecture/

Kaelble, S. (2022). *Data Resiliency for dummies.* John Wiley & Sons, Inc.

ManageEngine. (2023). *Data access control.* Pridobljeno iz ManageEngine DataSecurity Plus: https://www.manageengine.com/data-security/what-is/data-access-control.html

Rahman , F. L. (9. Julij 2021). *Pojasnjeno IT: Kaj je podatkovno jezero?* Pridobljeno iz TechAcute: https://techacute.com/data-lake/

Rashid, A., Chivers, H., Lupu, E., Martin, A., Schneider, S., Jones, H., . . . Hellett, J. (2021). *CyBOK - The Cyber Security Body of Knowladge.* The National Cyber Security Centre.

Rouse, M. (18. Maj 2016). *Data Administration.* Pridobljeno iz Techopedia: https://www.techopedia.com/definition/28399/data-administration

Simplilearn. (15. Februar 2023). *What Is Data Architecture? Overview and Best Practices.* Pridobljeno iz Simplilearn: https://www.simplilearn.com/what-is-data-architecture-article

Suer, M. (September 2023). *What Is Data Quality and Why Is It Important?* Pridobljeno iz Alation: https://www.alation.com/blog/what-is-data-quality-why-is-it-important/

talend. (2020). *Definitive Guide to Data Governance.* talend.

talend. (2023). *What is a Data Mart?* Pridobljeno iz talend: https://www.talend.com/resources/what-is-data-mart/

talend. (2023). *What is a Data Warehouse and Why Does It Matter To Your Business?* Pridobljeno iz talend: https://www.talend.com/resources/what-is-data-warehouse/