ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «Інфраструктура інформаційна система довгострокового зберігання великих масивів даних з використанням віртуалізації»

на здобуття освітнього ступеня магістра
зі спеціальності 126 Інформаційні системи та технології
освітньо-професійної програми Інформаційні системи та технології

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ _____

(*підпис*) (*Ім'я, Прізвище здобувача*)

Виконав:
здобувач вищої освіти Хоменко Ф.Ю.
група ІСДМ-64

Керівник: В. М. Данильченко
науковий ступінь, вчене звання *Доктор філософії*

Рецензент:
науковий ступінь, вчене звання _____
Ім'я, ПРІЗВИЩЕ

Київ – 2024

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут Інформаційних технологій

Кафедра ~~Інформаційних систем та технологій~~
Ступінь вищої освіти ~~Магістр~~
Спеціальність 126 Інформаційні системи та технології
Освітньо-професійна програма Інформаційні системи та технології

ЗАТВЕРДЖУЮ
Завідувач кафедрою ІСТ
_____
____ ~~Каміла СТОРЧАК~~
« » 20 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ СТУДЕНТУ

Хоменко Федір Юрійович
(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: «Інфраструктура інформаційна система довгострокового зберігання великих масивів даних з використанням віртуалізації»

керівник кваліфікаційної роботи *Валентина* ДАНИЛЬЧЕНКО, доктор філософії
(ім'я, ПРІЗВИЩЕ, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від «15» жовтня 2024 року № 320.

2. Строк подання кваліфікаційної роботи: 27 грудня 2024 року.

3. Вихідні дані до кваліфікаційної роботи:
VMware vSphere;
Dell EMC Network;
Система зберігання даних NAS;
науково-технічна література з питань, пов'язаних з наукою про дані.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
1. Дослідження сучаснихтехнологій віртуалізації та систем зберігання даних
2. Аналіз методів оптимізації та забезпечення відмовостійкості

3. Розробка архітектури локальної інфраструктури зберігання даних
4. Впровадження механізмів віртуалізації та системи безпеки

5. Перелік ілюстративного матеріалу: *презентація*
- Архітектура рішень;
- Аналіз поточного стану систем зберігання даних.

6. Дата видачі завдання: 15 жовтня 2024 року.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз наявної науково-технічної літератури	15.10 – 23.10.24	Виконано
2	Дослідження сучасних систем зберігання даних	24.10 – 01.11.24	Виконано
3	Аналіз вимог до локальної інфраструктури	02.11 – 07.11.24	Виконано
4	Проектування архітектури системи	08.11 – 16.11.24	Виконано
5	Розробка механізмів віртуалізації	17.11 – 21.11.24	Виконано
6	Впровадження систем безпеки	22.11 – 11.12.24	Виконано
7	Оформлення роботи: вступ, висновки, реферат	12.12 – 20.12.24	Виконано
8	Розробка демонстраційних матеріалів	21.12 – 25.12.24	Виконано

Здобувач вищої освіти _____ *Федір* ХОМЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

Керівник роботи
кваліфікаційної роботи _____ *Валентина* ДАНИЛЬЧЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня магістра: 87 стор., 34 рис., 15, табл., 30 джерел.

Мета роботи – розробка та впровадження ефективної локальної системи інфраструктури зберігання даних для довгострокового зберігання великих масивів інформації з використанням технології віртуалізації.

Об'єкт дослідження – процес організації довгострокового зберігання великих масивів даних в локальних інфраструктурах.

Предмет дослідження – методи та технології створення віртуалізованої інфраструктури для ефективного зберігання та управління великими масивами даних.

Короткий зміст роботи: У даній кваліфікаційній роботі проведено комплексний аналіз сучасних підходів до організації локальних систем зберігання даних. Досліджено технології віртуалізації та їх застосування для оптимізації інфраструктури зберігання. Розроблено архітектуру системи, що забезпечує надійне довгострокове зберігання великих обсягів даних з використанням віртуалізації ресурсів.

Особлива увага приділена питанням масштабованості, відмовостійкості та безпеки даних. Запропоновано методи оптимізації використання ресурсів та підвищення ефективності управління даними. На основі проведених досліджень створено прототип системи, що демонструє практичну цінність запропонованих рішень.

КЛЮЧОВІ СЛОВА: ІНФРАСТРУКТУРА ДАНИХ, ВІРТУАЛІЗАЦІЯ, ЗБЕРІГАННЯ ДАНИХ, МАСШТАБОВАНІСТЬ, ВІДМОВОСТІЙКІСТЬ, ОПТИМІЗАЦІЯ РЕСУРСІВ, ВЕЛИКІ ДАНІ, ЛОКАЛЬНІ СИСТЕМИ, ДОВГОСТРОКОВЕ ЗБЕРІГАННЯ, УПРАВЛІННЯ ДАНИМИ.

ABSTRACT

Text part of the master's qualification work: 97 pages, 34 pictures, 15 tables, 30 sources.

The purpose of the work - Development and implementation of an efficient local data infrastructure system for long-term storage of large data arrays using virtualization technologies.

Object of research - The process of organizing long-term storage of large data arrays in local infrastructures.

Subject of research - Methods and technologies for creating virtualized infrastructure for efficient storage and management of large data arrays.

Summary of the work: This qualification work provides a comprehensive analysis of modern approaches to organizing local data storage systems. Virtualization technologies and their application for optimizing storage infrastructure have been studied. A system architecture has been developed that ensures reliable long-term storage of large data volumes using resource virtualization.

Special attention is paid to issues of scalability, fault tolerance, and data security. Methods for optimizing resource utilization and improving data management efficiency are proposed. Based on the conducted research, a system prototype has been created that demonstrates the practical value of the proposed solutions.

KEYWORDS: DATA INFRASTRUCTURE, VIRTUALIZATION, DATA STORAGE, SCALABILITY, FAULT TOLERANCE, RESOURCE OPTIMIZATION, BIG DATA, LOCAL SYSTEMS, LONG-TERM STORAGE, DATA MANAGEMENT.

ЗМІСТ

ВСТУП

Актуальність теми: В умовах стрімкого зростання обсягів даних та ускладнення інформаційних систем, питання ефективної організації зберігання та управління великими масивами інформації стає все більш актуальним. Сучасні організації стикаються з необхідністю забезпечення надійного довгострокового зберігання даних при одночасному збереженні гнучкості та економічної ефективності інфраструктури. Використання технологій віртуалізації відкриває нові можливості для оптимізації систем зберігання даних, дозволяючи більш ефективно використовувати наявні ресурси та забезпечувати необхідний рівень масштабованості та надійності.

Об'єкт дослідження: Процес організації довгострокового зберігання великих масивів даних в локальних інфраструктурах.

Предмет дослідження: Методи та технології створення віртуалізованої інфраструктури для ефективного зберігання та управління великими масивами даних.

Мета і завдання дослідження: Метою дослідження є розробка та впровадження ефективної локальної системи інфраструктури даних для довгострокового зберігання великих масивів інформації з використанням технологій віртуалізації. Для досягнення цієї мети поставлені такі завдання:

Провести аналіз сучасних підходів до організації систем зберігання даних та технологій віртуалізації;

- Розробити архітектуру локальної системи зберігання даних з використанням віртуалізації;
- Дослідити та впровадити методи оптимізації використання ресурсів;
- Забезпечити необхідний рівень відмовостійкості та безпеки даних;
- Розробити та протестувати прототип системи;
- Оцінити ефективність запропонованих рішень.

Методи дослідження: У роботі використані методи системного аналізу, математичного моделювання, експериментального дослідження та статистичного

аналізу. Застосовано сучасні технології віртуалізації та управління даними, а також методи оптимізації комп'ютерних систем.

Наукова новизна: Розроблено новий підхід до організації локальних систем зберігання даних, що базується на комплексному використанні технологій віртуалізації та оптимізації ресурсів. Запропоновано методи підвищення ефективності управління даними та забезпечення їх довгострокового зберігання.

Практична значущість результатів: Результати роботи можуть бути використані при проектуванні та впровадженні систем зберігання даних в організаціях різного масштабу. Розроблені методи та підходи дозволяють оптимізувати використання ресурсів та знизити витрати на підтримку інфраструктури даних.

Методична значущість результатів: Розроблені методи та підходи можуть бути використані в навчальному процесі при викладанні дисциплін, пов'язаних з проектуванням систем зберігання даних та віртуалізацією. Матеріали дослідження можуть стати основою для створення методичних рекомендацій та практичних посібників для студентів технічних спеціальностей.

Теоретична значущість результатів: Робота розширює теоретичну базу в області проектування локальних систем зберігання даних, доповнює існуючі знання про методи оптимізації інфраструктури з використанням віртуалізації, а також формує нове розуміння підходів до забезпечення ефективності та надійності систем довгострокового зберігання великих масивів даних.

Апробація результатів бакалаврської роботи:

Хоменко Ф.Ю. «Проектування та реалізація локальної системи інфраструктури даних для довгострокового зберігання великих масивів даних з використанням віртуалізації». – м. Могилів-Подільський, 06 грудня 2024 р.

Хоменко Ф.Ю. «Впровадження та оптимізація системи резервного копіювання на базі Dell EMC Networker з використанням стрічкової бібліотеки та розширеними політиками безпеки». – м. Рівне, 13 грудня 2024 р.

1 ОСНОВИ РЕАЛІЗАЦІЇ СИСТЕМИ ЗБЕРІГАННЯ

1.1 Концептуальні засади побудови систем зберігання великих масивів даних

Експоненційне зростання обсягів даних у сучасних інформаційних системах створює значні виклики щодо їх ефективного зберігання та управління. Розробка та впровадження систем зберігання великих масивів даних (Big Data Storage Systems) потребує чіткого розуміння концептуальних засад, які визначають архітектурні, технологічні та експлуатаційні аспекти таких рішень.

Системи зберігання даних (далі - СЗД) являють собою комплекс програмно-апаратних засобів, призначених для надійного збереження інформації та забезпечення оперативного доступу до неї. В умовах стрімкого розвитку цифрових технологій та збільшення обсягів даних, що генеруються, особливої актуальності набувають питання ефективної організації таких систем. Вони є критичними для забезпечення конкурентоспроможності бізнесу, дозволяючи обробляти великі обсяги інформації для прийняття обґрунтованих рішень.

Фундаментальні принципи побудови СЗД базуються на трьох ключових аспектах: надійність, продуктивність та масштабованість. Ці аспекти формують концептуальний базис, на якому ґрунтується вся архітектура сучасних систем зберігання даних.

Надійність систем зберігання даних є критично важливим аспектом, який забезпечується через комплексне впровадження різноманітних механізмів резервування та відмовостійкості. Ця багаторівнева система захисту починається на рівні окремих дисків і поширюється на всю інфраструктуру зберігання даних.

Фундаментальним елементом забезпечення надійності є використання RAID-масивів різних рівнів. Кожен рівень RAID пропонує свій баланс між продуктивністю, надійністю та вартістю. Наприклад, RAID 1 забезпечує повне дзеркалювання даних, створюючи ідентичну копію на другому диску, що гарантує миттєве відновлення при відмові одного з дисків. RAID 5 використовує розподілену парність, що дозволяє відновити дані при виході з ладу одного диска, одночасно забезпечуючи ефективне використання дискового простору. RAID 6, який є особливо важливим для великих масивів даних, реалізує подвійну парність, що дозволяє системі продовжувати роботу навіть при одночасній відмові двох дисків,

що особливо актуально в умовах використання високоємнісних дисків з тривалим часом відновлення.

Технології реплікації та синхронізації даних формують наступний рівень захисту. Синхронна реплікація забезпечує створення точної копії даних у реальному часі на віддаленому майданчику, гарантуючи нульову втрату даних при відмові основного сайту. Асинхронна реплікація, хоча і допускає певну затримку в оновленні резервної копії, дозволяє створювати територіально розподілені системи зберігання даних без суттєвого впливу на продуктивність основної системи. Сучасні системи реплікації також підтримують каскадне копіювання, коли дані послідовно передаються через кілька проміжних вузлів, що дозволяє оптимізувати використання мережевих ресурсів.

Забезпечення цілісності даних реалізується через багаторівневу систему перевірок. На апаратному рівні використовується ECC-пам'ять (Error Correction Code), яка здатна не тільки виявляти, але й автоматично виправляти однобітові помилки, а також виявляти двобітові помилки. Контрольні суми застосовуються на різних рівнях: від окремих блоків даних до цілих файлів та томів. Сучасні файлові системи, такі як ZFS або Btrfs, реалізують додаткові механізми захисту від пошкодження даних, включаючи контрольні суми для метаданих та автоматичну корекцію помилок через порівняння копій.

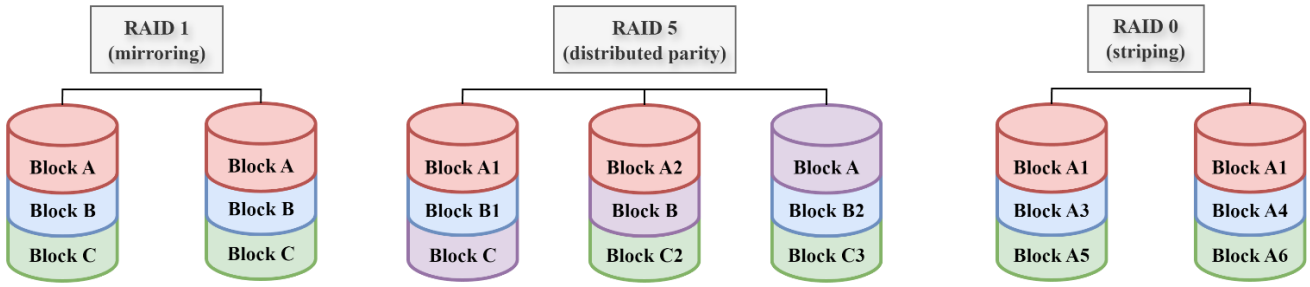


рис. 1.1. - «Фізичне розташування даних у різних рівнях RAID»

Механізми верифікації даних працюють як в активному, так і в пасивному режимах. Активна верифікація включає регулярне сканування дисків на наявність помилок (scrubbing), що дозволяє виявляти та виправляти проблеми до того, як вони вплинуть на роботу системи. Пасивна верифікація здійснюється при кожному читанні даних, забезпечуючи їх цілісність в момент використання. Додатково, системи зберігання даних можуть використовувати різні алгоритми відновлення даних, такі як Reed-Solomon коди, які дозволяють відновлювати дані навіть при множинних відмовах носіїв.

Для критично важливих даних часто впроваджується концепція "повітряного зазору" (air gap), коли частина резервних копій зберігається на носіях, фізично відключених від мережі, що забезпечує захист від програмних збоїв, вірусів та кібератак. Такий підхід, у поєднанні з регулярним тестуванням процедур відновлення, забезпечує максимальний рівень захисту даних.

Продуктивність СЗД досягається шляхом оптимізації архітектури системи та використання сучасних технологічних рішень. Це включає застосування багаторівневих систем зберігання (tiered storage), де дані розміщуються на різних носіях залежно від їхньої важливості та частоти доступу. Використання твердотільних накопичувачів (SSD) для кешування часто використовуваних даних значно підвищує швидкість доступу. Оптимізація алгоритмів доступу до даних та балансування навантаження між вузлами системи дозволяє забезпечити високу продуктивність навіть при пікових навантаженнях.

Масштабованість є критично важливою характеристикою сучасних СЗД, що забезпечує можливість їх розвитку відповідно до зростаючих потреб організації. Виділяють горизонтальну масштабованість (scale-out), яка передбачає додавання нових вузлів зберігання, та вертикальну масштабованість (scale-up), що полягає у збільшенні потужності існуючих компонентів. Наприклад, використання розподілених файлових систем дозволяє додавати нові сервери без зупинки роботи системи.

Архітектурні рішення для побудови систем зберігання великих масивів даних включають різні типи сховищ, кожен з яких має свої особливості та оптимальні випадки використання.

Блочне сховище (Block Storage) представляє дані у вигляді фіксованих блоків і характеризується високою продуктивністю завдяки прямому доступу до блоків. Воно дозволяє пряме управління блоками даних і є оптимальним для баз даних та віртуальних машин. Блочне сховище підтримує протоколи Fibre Channel та iSCSI.

Файлове сховище (File Storage) організовує дані в ієрархічну структуру та забезпечує простий доступ через стандартні протоколи (SMB/CIFS, NFS). Воно підтримує паралельний доступ багатьох користувачів, зручне управління правами доступу та можливість спільного використання файлів.

Об'єктне сховище (Object Storage) оптимізоване для зберігання великих обсягів неструктурованих даних. Воно надає доступ через REST API, має розширені можливості метаданих, високу масштабованість та оптимальне співвідношення ціна/об'єм.

Загальна класифікація систем зберігання включає:

Традиційні системи зберігання даних: системи прямого підключення (DAS), мережеві системи зберігання даних (NAS) та мережі зберігання даних (SAN). Вони забезпечують базові функції зберігання та доступу до даних, але мають обмеження в масштабованості та гнучкості.

Розподілені файлові системи: такі як Hadoop Distributed File System (HDFS), GlusterFS, Ceph та Lustre. Вони дозволяють організувати зберігання даних на багатьох вузлах, забезпечуючи масштабованість та відмовостійкість.

Об'єктні сховища даних: Amazon S3, OpenStack Swift, Ceph Object Gateway та Microsoft Azure Blob Storage. Вони оптимізовані для зберігання великих обсягів неструктурованих даних з високою масштабованістю та доступністю.

Хмарні системи зберігання даних: публічні, приватні та гібридні хмарні рішення, які надають можливість зберігати дані в хмарі з гнучкими моделями оплати та масштабування.

Важливим аспектом концептуальних засад побудови СЗД є забезпечення ефективного управління даними, що включає класифікацію та категоризацію даних, політики зберігання та забезпечення безпеки.

Класифікація та категоризація даних дозволяє визначити оптимальні методи їх зберігання та управління. Дані можуть класифікуватися за важливістю та критичністю, частотою використання, вимогами до швидкості доступу та регуляторними вимогами. Наприклад, критичні дані, які часто використовуються, слід зберігати на високошвидкісних носіях з високим рівнем захисту.

Політики зберігання даних допомагають керувати життєвим циклом даних. Це включає визначення термінів зберігання, правила архівування, процедури видалення та механізми відновлення. Встановлення чітких політик забезпечує ефективне використання ресурсів та відповідність регуляторним вимогам.

Забезпечення безпеки даних є критично важливим для захисту від втрат та несанкціонованого доступу. Це включає шифрування даних при зберіганні, контроль доступу, аудит операцій та захист від несанкціонованого доступу. Наприклад, використання багатофакторної автентифікації та регулярний аудит системи допомагають підтримувати високий рівень безпеки.

Особливу увагу при побудові СЗД необхідно приділяти оптимізації використання ресурсів системи. Це досягається через впровадження технологій дедуплікації, використання компресії даних та оптимізацію розміщення даних.

Дедуплікація дозволяє зменшити обсяг займаного простору шляхом усунення дублюючих даних. Існують різні підходи, такі як файлова та блочна дедуплікація, інлайн та пост-процесинг дедуплікація. Інлайн дедуплікація здійснюється в реальному часі під час запису даних, що зменшує потреби у зберіганні з самого початку.

Компресія даних зменшує обсяг, що займається, шляхом ущільнення інформації. Це може бути апаратна або програмна компресія, адаптивна або багаторівнева. Адаптивна компресія дозволяє системі самостійно визначати, які дані варто компресувати для досягнення оптимального балансу між продуктивністю та ефективністю.

Оптимізація розміщення даних покращує продуктивність та ефективність системи. Автоматичне тирування дозволяє автоматично розподіляти дані по різних рівнях зберігання залежно від їх активності. Інтелектуальне кешування використовує швидкі накопичувачі для зберігання часто використовуваних даних, що підвищує швидкість доступу.

Сучасні тенденції розвитку систем зберігання даних включають гіперконвергентні інфраструктури (HCI), використання штучного інтелекту та машинного навчання, а також контейнерні технології.

Гіперконвергентні інфраструктури (HCI) об'єднують обчислювальні ресурси, мережі та системи зберігання в єдиний програмно-керований блок. Це спрощує розгортання та управління інфраструктурою, зменшує потребу в ручному втручанні та підвищує ефективність.

Використання штучного інтелекту та машинного навчання покращує ефективність СЗД через предиктивну аналітику, автоматичну оптимізацію, інтелектуальне керування ресурсами та виявлення аномалій. Системи можуть прогнозувати можливі відмови та потреби в ресурсах, що дозволяє проактивно реагувати на зміни.

Контейнерні технології змінюють підхід до розгортання та управління додатками. Забезпечення постійного зберігання для контейнерів та використання сховищ, розроблених спеціально для роботи з контейнерами, дозволяють ефективно зберігати дані, які потрібні контейнерам. Інтеграція з системами оркестрації, такими як Kubernetes, сприяє автоматизації та масштабованості.

Важливим аспектом концептуальних засад є також забезпечення відповідності системи зберігання даних регуляторним вимогам та галузевим стандартам. Це включає відповідність вимогам щодо захисту персональних даних

(наприклад, GDPR, CCPA, локальні регуляторні вимоги), забезпечення аудиту та звітності, а також управління життєвим циклом даних.

При проектуванні систем зберігання великих масивів даних необхідно враховувати також економічні аспекти. Оптимізація витрат через балансування капітальних та операційних витрат, оцінка вартості володіння (TCO) та повернення інвестицій (ROI) є ключовими для досягнення економічної ефективності. Ефективність використання ресурсів включає енергоефективність, оптимізацію використання простору, автоматизацію процесів та зниження адміністративних витрат. Планування розвитку передбачає стратегію масштабування, управління життєвим циклом обладнання, модернізацію інфраструктури та оптимізацію інвестицій.

Концептуальні засади побудови систем зберігання великих масивів даних також включають аспекти забезпечення високої доступності та безперервності бізнесу. Це досягається через архітектуру високої доступності, плани аварійного відновлення (Disaster Recovery) та ефективний моніторинг та управління системою.

Архітектура високої доступності включає активно-активні конфігурації, географічно розподілені системи та автоматичне відновлення після збоїв. Це забезпечує мінімізацію часу простою та втрати даних, що критично для безперервності бізнес-процесів.

Disaster Recovery передбачає розробку планів аварійного відновлення, створення резервних майданчиків, процедури переключення та регулярне тестування відновлення. Це дозволяє швидко відновити роботу системи після катастрофічних подій, забезпечуючи стабільність бізнесу.

Моніторинг та управління включають використання систем моніторингу, превентивну діагностику, автоматизацію операцій та управління інцидентами. Це забезпечує постійне спостереження за станом системи, своєчасне виявлення та вирішення проблем.

Таким чином, концептуальні засади побудови систем зберігання великих масивів даних формують комплексний підхід до створення ефективних, надійних та масштабованих рішень, що відповідають сучасним вимогам бізнесу та технологічним викликам. Врахування всіх описаних аспектів дозволяє створювати системи, які забезпечують оптимальне співвідношення продуктивності, надійності та вартості володіння.

1.2 Еволюція технологій віртуалізації в контексті управління даними

Історія розвитку технологій віртуалізації тісно пов'язана з еволюцією систем зберігання та управління даними. Починаючи з перших mainframe-комп'ютерів і до сучасних хмарних середовищ, віртуалізація пройшла значний шлях трансформації, постійно адаптуючись до нових викликів та потреб у сфері обробки та зберігання інформації.

У ранні роки розвитку обчислювальної техніки, коли апаратні ресурси були надзвичайно дорогими, виникла потреба в їх ефективному розподілі між різними користувачами та завданнями. Саме ця необхідність стала каталізатором розвитку технологій віртуалізації. Перші реалізації віртуалізації з'явились на mainframe-комп'ютерах IBM у 1960-х роках, де вони дозволяли розділяти ресурси системи між різними користувачами, створюючи ілюзію роботи на окремому комп'ютері для кожного з них.

Справжній прорив у розвитку віртуалізації відбувся з появою архітектури x86 та персональних комп'ютерів. Це створило нові можливості для впровадження віртуалізації на рівні настільних систем та серверів. Важливим етапом стала поява перших комерційних гіпервізорів, таких як VMware Workstation у 1999 році, що зробило технології віртуалізації доступними для широкого кола користувачів та організацій.

З точки зору управління даними, віртуалізація запропонувала революційний підхід до організації систем зберігання. Вона дозволила створювати абстрактний шар між фізичними носіями та логічними структурами даних, що значно підвищило гнучкість та ефективність використання ресурсів зберігання. Віртуалізація сховищ даних (storage virtualization) стала ключовим компонентом сучасних центрів обробки даних, забезпечуючи можливість динамічного управління ресурсами зберігання та їх оптимального розподілу.

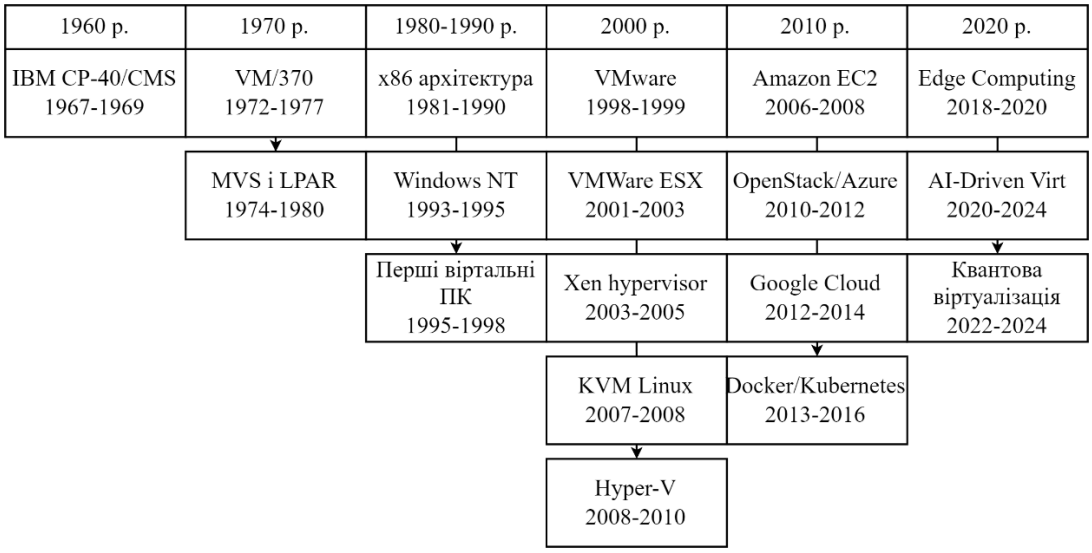


рис. 1.2. – «Еволюція технологій віртуалізації»

Сучасні технології віртуалізації можна розділити на кілька основних категорій, кожна з яких відіграє важливу роль у побудові ефективних систем управління даними та обчислювальними ресурсами. Ці категорії формують комплексну екосистему, де кожен компонент доповнює інші, створюючи гнучку та ефективну IT-інфраструктуру.

Серверна віртуалізація є фундаментальною технологією, яка революціонізувала підхід до використання обчислювальних ресурсів. Вона дозволяє створювати та запускати кілька віртуальних серверів на одному фізичному обладнанні, кожен з яких функціонує як повноцінна, незалежна система. При цьому віртуальні сервери можуть працювати під управлінням різних операційних систем та виконувати різноманітні завдання, від обробки баз даних до хостингу веб-додатків. Це забезпечує не лише ефективне використання обчислювальних ресурсів, але й значно спрощує процеси управління інфраструктурою, включаючи резервне копіювання, відновлення після збоїв та міграцію між фізичними серверами.

В контексті серверної віртуалізації особливу роль відіграють гіпервізори двох типів. Гіпервізори типу 1 (bare-metal), такі як VMware ESXi, Microsoft Hyper-V та KVM, встановлюються безпосередньо на апаратне забезпечення без необхідності проміжної операційної системи. Це забезпечує максимальну продуктивність віртуальних машин завдяки прямому доступу до апаратних ресурсів серверу. Наприклад, VMware ESXi використовує власний оптимізований мікрокернел, що дозволяє досягти практично нативної продуктивності для віртуальних машин. Microsoft Hyper-V, інтегрований з Windows Server, надає потужні можливості для

організацій, які використовують екосистему Microsoft. KVM, будучи частиною ядра Linux, забезпечує відмінну продуктивність та гнучкість для Linux-орієнтованих середовищ.

Гіпервізори типу 2, такі як VMware Workstation, Oracle VirtualBox та Parallels Desktop, працюють поверх операційної системи хоста як звичайні програми. Хоча вони мають дещо нижчу продуктивність порівняно з гіпервізорами типу 1, вони надають більшу гнучкість та зручність використання. Ці гіпервізори найчастіше застосовуються в середовищах розробки та тестування, де необхідно швидко створювати та налаштовувати різні віртуальні середовища. Вони також дозволяють легко створювати знімки стану віртуальних машин, що особливо корисно при тестуванні програмного забезпечення та навчанні.

Важливо зазначити, що вибір між різними типами гіпервізорів залежить від конкретних потреб організації. Для підприємств, що потребують максимальної продуктивності та надійності в промисловому масштабі, оптимальним вибором будуть гіпервізори типу 1. Для середовищ розробки, тестування та навчання, де гнучкість та зручність користування важливіші за максимальну продуктивність, більш доцільним буде використання гіпервізорів типу 2.

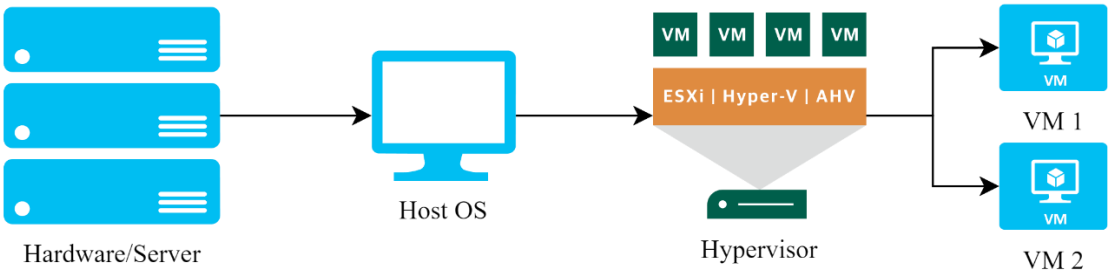


рис. 1.3. – «Гіпервізор типу 1 (bare-metal)»

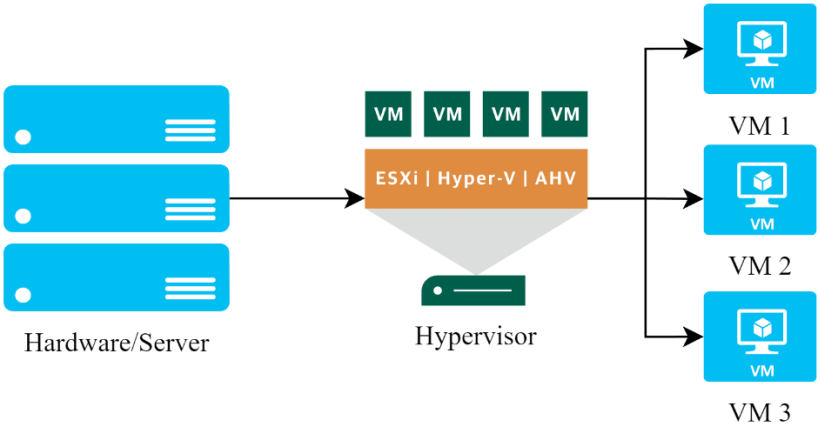


рис. 1.4. – «Гіпервізор типу 2 (hosted)»

Віртуалізація сховищ даних та мережева віртуалізація є фундаментальними компонентами сучасної IT-інфраструктури, які тісно взаємодіють між собою та з серверною віртуалізацією для створення гнучкого та ефективного середовища обробки даних. Розвиток цих технологій відбувався паралельно, що дозволило створити інтегровані рішення, які відповідають зростаючим потребам сучасного бізнесу.

Storage Area Network (SAN) представляє собою спеціалізовану високошвидкісну мережу зберігання даних, яка забезпечує доступ до консолідованих ресурсів на рівні блоків. Це особливо важливо для критичних додатків та баз даних, де потрібна максимальна продуктивність та надійність. SAN дозволяє створювати великі пули зберігання, які можна гнучко розподіляти між різними серверами та додатками, забезпечуючи при цьому централізоване управління та моніторинг.

Network-Attached Storage (NAS), у свою чергу, пропонує більш простий підхід до зберігання даних, забезпечуючи доступ на рівні файлів через стандартні мережеві протоколи. Це рішення ідеально підходить для організації спільного доступу до файлів, створення резервних копій та зберігання неструктурованих даних. NAS-системи легко масштабуються та не вимагають складного налаштування, що робить їх популярним вибором для малого та середнього бізнесу.

Сучасні системи віртуалізації сховищ даних використовують передові технології оптимізації, такі як тонке виділення ресурсів, що дозволяє більш ефективно використовувати доступний простір. Дедуплікація та компресія даних в режимі реального часу допомагають значно зменшити обсяг фізичного простору, необхідного для зберігання даних. Механізми створення миттєвих знімків та реплікації забезпечують надійний захист даних та їх швидке відновлення у разі необхідності.

Мережева віртуалізація через технологію Software-Defined Networking (SDN) змінила підхід до побудови та управління мережевою інфраструктурою. Розділення площини управління та площини передачі даних дозволяє централізовано керувати всією мережею через програмні інтерфейси. Це забезпечує можливість швидко адаптувати мережеву інфраструктуру до змінних потреб бізнесу, автоматизувати рутинні операції та підвищити загальну безпеку системи.

Network Functions Virtualization (NFV) доповнює можливості SDN, дозволяючи віртуалізувати різноманітні мережеві функції. Завдяки NFV такі компоненти як файрволи, балансувальники навантаження та маршрутизатори

можуть бути реалізовані як віртуальні сервіси, що значно спрощує їх розгортання та управління ними. Це також дозволяє швидко масштабувати мережеві сервіси відповідно до потреб бізнесу без необхідності встановлення додаткового фізичного обладнання.

Взаємодія всіх цих компонентів у сучасному центрі обробки даних створює єдину інтегровану систему, де фізична інфраструктура, віртуалізація сховищ та мережева віртуалізація працюють разом для забезпечення ефективної роботи віртуальних машин та додатків. Така архітектура забезпечує не тільки високу продуктивність та надійність, але й значно спрощує адміністрування та знижує експлуатаційні витрати.

У перспективі розвиток технологій віртуалізації продовжить рухатися у напрямку ще більшої інтеграції та автоматизації. Використання штучного інтелекту та машинного навчання для оптимізації роботи віртуалізованих середовищ, подальший розвиток програмно-визначених технологій та вдосконалення механізмів безпеки залишаються ключовими тенденціями в цій області.

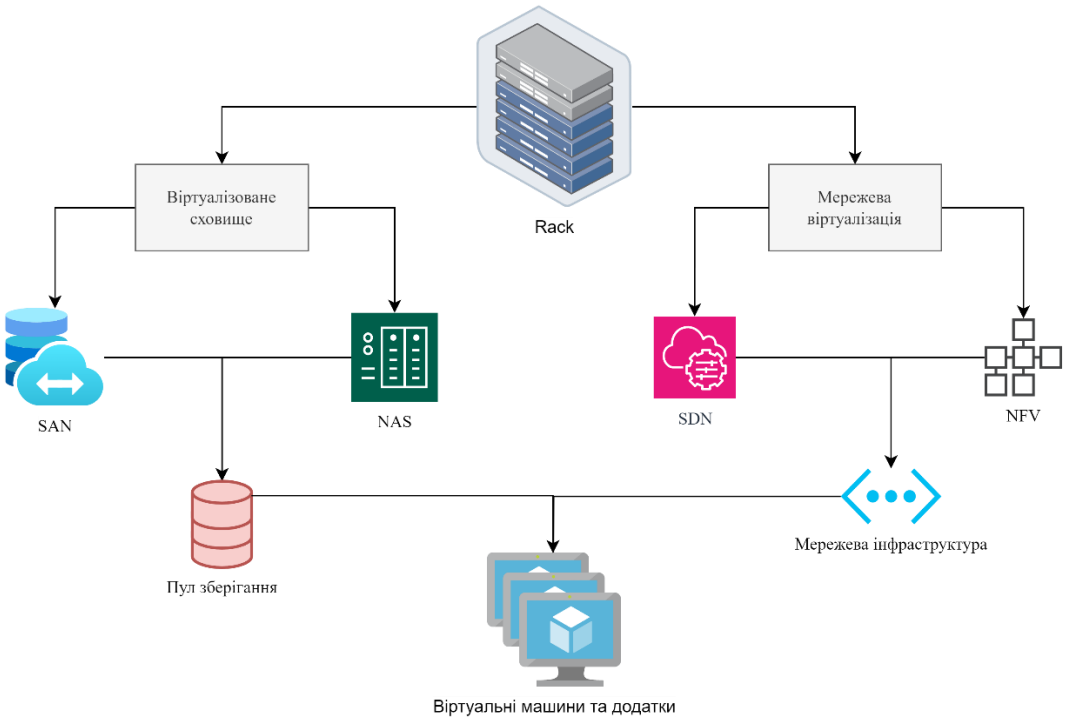


рис. 1.5. – «Інтегрована система віртуалізації»

Контейнеризація та хмарні технології здійснили революційний прорив у сфері віртуалізації, створивши нову парадигму розробки та розгортання додатків. Сучасна контейнеризація, започаткована Docker, еволюціонувала далеко за межі простого пакування додатків. Вона створила цілу екосистему інструментів та практик, що охоплюють весь життєвий цикл програмного забезпечення.

Сучасні контейнерні технології включають такі інновації як безсерверні контейнери (serverless containers), де інфраструктурою повністю керує провайдер хмарних послуг, та WebAssembly (Wasm) контейнери, які забезпечують ще більшу портативність та безпеку. З'явилися спеціалізовані мікро-VM платформи, такі як Firecracker від Amazon, які поєднують легковагість контейнерів з ізоляцією віртуальних машин.

Оркестрація контейнерів також зазнала значного розвитку. Хоча Kubernetes залишається домінуючою платформою, з'явилися нові рішення, орієнтовані на спрощення управління контейнерами для менших проектів. K3s, Kind та Minikube зробили контейнерну оркестрацію доступнішою для розробників та невеликих команд. Servicemesh технології, такі як Istio та Linkerd, додали нові можливості для управління мікросервісною архітектурою.

У сфері хмарних технологій відбувається активний розвиток нових сервісних моделей. Function as a Service (FaaS) або serverless computing дозволяє розробникам зосередитися виключно на коді, повністю абстрагуючись від інфраструктури. AI as a Service (AIaaS) надає доступ до потужних моделей штучного інтелекту через API, що democratizes доступ до AI технологій. Blockchain as a Service (BaaS) спрощує впровадження блокчейн-технологій в бізнес-процеси.

Edge Computing став важливим розширенням хмарних технологій, дозволяючи обробляти дані ближче до джерела їх виникнення. Це особливо важливо для IoT застосунків та додатків, що вимагають мінімальної затримки. Гібридні та мульти-хмарні стратегії стали нормою, дозволяючи організаціям оптимально розподіляти навантаження між різними провайдерами та власною інфраструктурою.

Безпека в хмарних середовищах також еволюціонувала з появою Zero Trust архітектури та DevSecOps практик. Технології конфіденційних обчислень (Confidential Computing) забезпечують захист даних навіть під час їх обробки в пам'яті. Infrastructure as Code (IaC) та GitOps підходи автоматизують розгортання та управління інфраструктурою, забезпечуючи відтворюваність та версійність конфігурацій.

Важливою тенденцією стала поява галузевих хмар (industry clouds), які пропонують спеціалізовані рішення для конкретних секторів економіки, таких як охорона здоров'я, фінанси чи виробництво. Ці рішення включають специфічні для галузі compliance вимоги, робочі процеси та інтеграції.

Sustainability as a Service стає новим напрямком розвитку хмарних технологій, де провайдери пропонують інструменти для моніторингу та оптимізації енергоспоживання та вуглецевого сліду хмарних workloads. Це відображає зростаючу важливість екологічної відповідальності в IT-секторі.

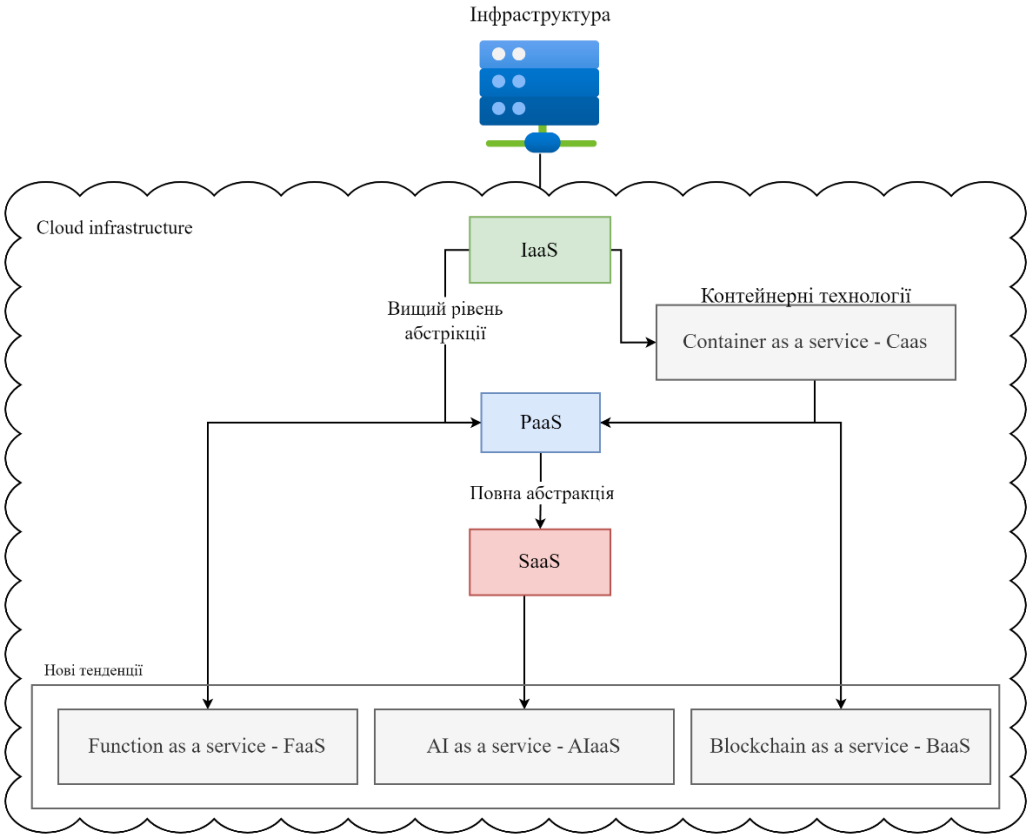


рис. 1.6. – «Моделі хмарних сервісів »

Гіперконвергентні інфраструктури (HCI) здійснили революцію в сфері віртуалізації, створивши фундамент для нового покоління центрів обробки даних. Сучасні HCI-системи розвиваються в напрямку повної автоматизації та самоуправління, використовуючи передові технології штучного інтелекту для оптимізації всіх аспектів роботи інфраструктури. Нове покоління HCI-рішень включає вбудовані можливості машинного навчання для передбачення навантаження, автоматичного балансування ресурсів та проактивного виявлення потенційних проблем.

У сфері безпеки віртуальних середовищ відбувається значний прогрес завдяки впровадженню технологій квантової криптографії та пост-квантових алгоритмів шифрування. Мікросегментація мереж еволюціонувала до концепції наносегментації, де кожен окремий процес може мати власну ізольовану мережеву політику. Технології довіреного виконання (Trusted Execution Environment, TEE) забезпечують захист даних навіть при компрометації гіпервізора.

Інтеграція штучного інтелекту в системи віртуалізації досягла нового рівня з появою нейроморфних обчислень та квантового машинного навчання. Ці технології дозволяють створювати системи, які не просто реагують на зміни, а передбачають їх та адаптуються превентивно. AI-керовані системи орхестрації можуть автоматично оптимізувати розміщення віртуальних машин та контейнерів, враховуючи сотні параметрів в режимі реального часу.

Edge computing стає ключовим напрямком розвитку віртуалізації, особливо з розгортанням мереж 5G та майбутніх 6G. Нові легковагі гіпервізори, оптимізовані для роботи на периферійних пристроях, дозволяють ефективно обробляти дані безпосередньо на місці їх виникнення. Технології віртуалізації адаптуються для роботи в умовах обмежених ресурсів та нестабільного підключення.

Важливим трендом стає розвиток "зелених" технологій віртуалізації, орієнтованих на мінімізацію енергоспоживання та екологічного впливу. Нові алгоритми оптимізації використовують складні моделі для балансування між продуктивністю та енергоефективністю. З'являються спеціалізовані рішення для моніторингу та зменшення вуглецевого сліду віртуальних інфраструктур.

Квантова віртуалізація представляє собою наступний великий прорив у цій галузі. Розробляються перші прототипи квантових гіпервізорів, здатних керувати як класичними, так і квантовими обчислювальними ресурсами. Це відкриває нові можливості для створення гібридних систем, що поєднують переваги класичних та квантових обчислень.

Міжхмарна віртуалізація розвивається в напрямку створення єдиного глобального пулу ресурсів, де workloads можуть динамічно переміщуватися між різними провайдерами та локаціями залежно від вимог до продуктивності, вартості та регуляторних обмежень. Розвиваються технології федеративного навчання, які дозволяють AI-моделям навчатися на розподілених даних без їх централізації.

У довгостроковій перспективі очікується поява повністю автономних віртуальних інфраструктур, здатних самостійно адаптуватися до змінних умов, оптимізувати свою роботу та відновлюватися після збоїв без людського втручання. Це буде досягнуто завдяки комбінації передових технологій штучного інтелекту, квантових обчислень та нових парадигм програмування.

Розвиток нейроінтерфейсів та розширеної реальності створює нові вимоги до віртуалізації, зокрема необхідність забезпечення надзвичайно низької затримки та високої пропускної здатності для обробки нейросигналів та AR/VR контенту. Це стимулює розвиток спеціалізованих рішень для віртуалізації на межі мережі.

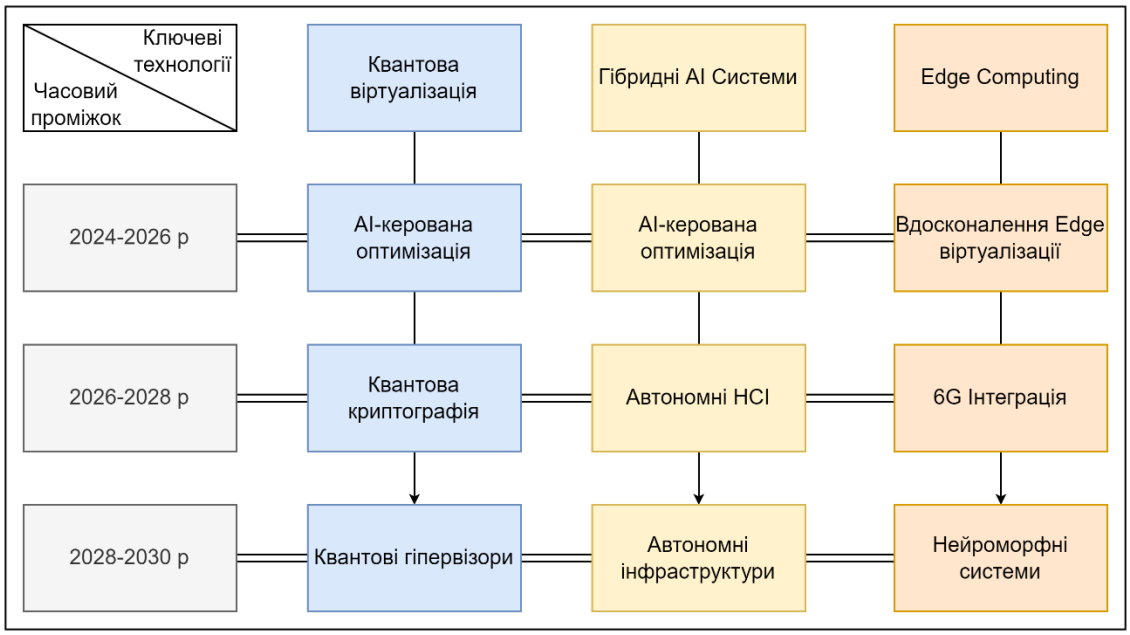


рис. 1.7. – «Ймовірний розвиток технологій віртуалізації»

Технології віртуалізації продовжують свій динамічний розвиток, формуючи майбутнє цифрової трансформації. Подібно до того, як перші гіпервізори революціонізували використання обчислювальних ресурсів, сьогодні ми спостерігаємо народження нового покоління технологій, що поєднують квантові обчислення, штучний інтелект та нейротехнології. Ця конвергенція створює безпрецедентні можливості для побудови інтелектуальних, самокерованих систем, здатних адаптуватися до потреб користувачів та бізнесу з безпрецедентною точністю та ефективністю.

Майбутнє віртуалізації лежить не просто в оптимізації існуючих процесів, а в створенні принципово нових парадигм взаємодії між людиною, машиною та даними. Віртуалізація стає невидимим, але всюдисущим фундаментом, що підтримує цифрову економіку та суспільство, забезпечуючи безшовну інтеграцію фізичного та цифрового світів. У цьому контексті, глибоке розуміння принципів та тенденцій розвитку віртуалізації стає не просто технічною необхідністю, а стратегічною компетенцією для організацій, що прагнуть залишатися конкурентоспроможними в епоху цифрової трансформації.

В перспективі найближчих десятиліть, системи віртуалізації еволюціонують до рівня "цифрової нервової системи" підприємств та суспільства в цілому, забезпечуючи миттєву адаптацію до змін, проактивне управління ресурсами та безпрецедентний рівень автоматизації. Це відкриває шлях до створення по-справжньому "розумних" підприємств та міст, де кожен елемент інфраструктури

працює в гармонії з іншими, оптимізуючи використання ресурсів та максимізуючи цінність для користувачів.

1.3 Методологія проектування локальних інфраструктур даних

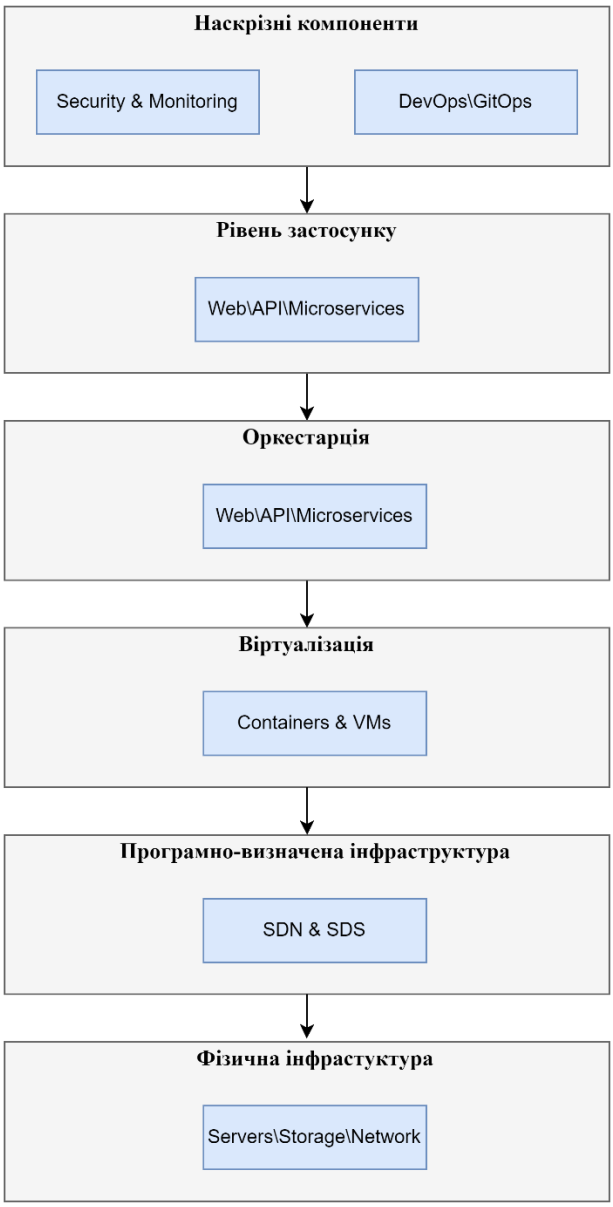


рис. 1.8. – «Загальна концепція сучасної локальної інфраструктури»

Проектування локальних інфраструктур даних у сучасних умовах слід розглядати через призму стрімкого зростання обсягів інформації, нових вимог до продуктивності, доступності та надійності, а також динамічної еволюції технологій віртуалізації, контейнеризації, інструментів оркестрації та програмно-визначених ресурсів. Важливим чинником є також необхідність узгодження інфраструктурних рішень із актуальними методами управління життєвим циклом даних, вимогами

інформаційної безпеки, нормативно-правовими стандартами, а також потребами у швидкому масштабуванні та адаптації до змін бізнес-середовища.

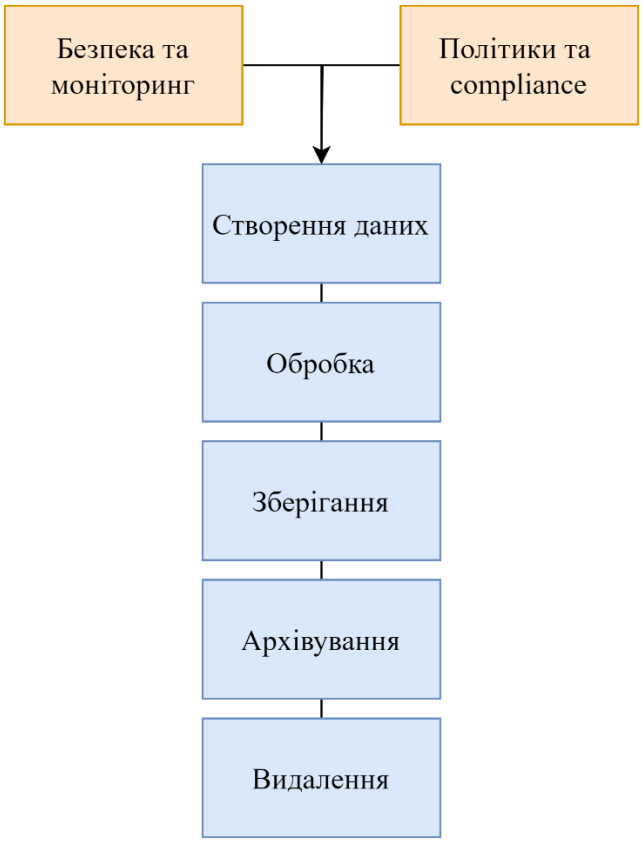


рис. 1.9. – «Життєвий цикл даних»

У минулому локальні інфраструктури здебільшого будувалися за класичною схемою: сервери – мережа – сховище. Проте з поширенням концепцій «інфраструктури як коду» (IaC), застосування контейнерів (Docker, Podman), оркестрації (Kubernetes), а також програмно-визначених мереж (SDN) та систем зберігання (SDS), з’являється можливість формувати гнучку, адаптивну та динамічну архітектуру. Така архітектура здатна працювати не лише з традиційними обчислювальними вузлами, але й з гіперконвергентними рішеннями, NVMe-накопичувачами, системами високої пропускної здатності та низької латентності, а також забезпечувати ефективний перерозподіл ресурсів у режимі реального часу.

Підхід до проектування локальних інфраструктур повинен починатися з глибинного аналізу потреб організації та цілей, які вона прагне досягти. Важливим етапом є визначення типів навантажень: транзакційних, аналітичних, високопродуктивних обчислень (HPC), роботи з великими даними (Big Data) чи опрацювання медіаконтенту.

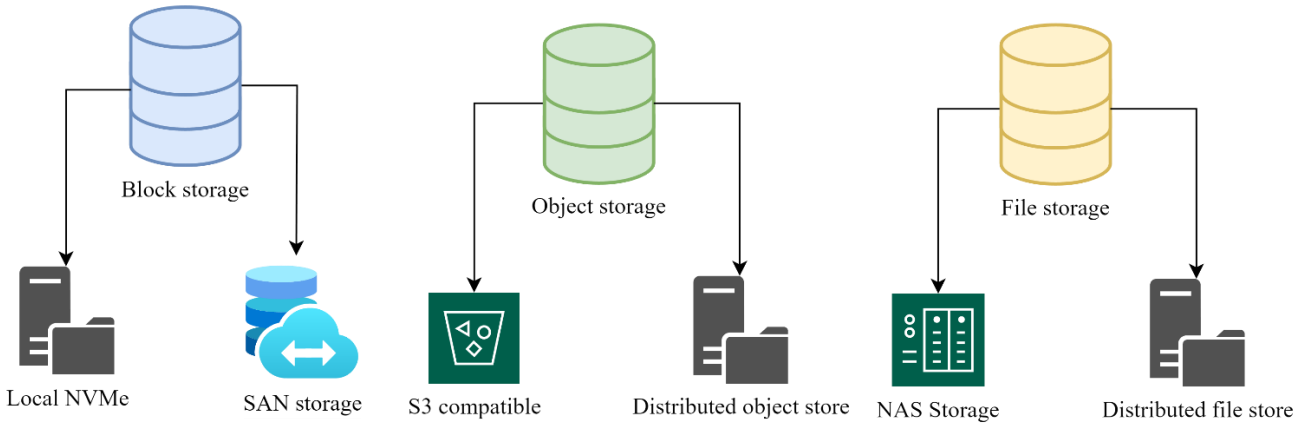


рис. 1.10. – «Архітектура систем зберігання даних»

Кожен тип навантаження має власні вимоги до пропускної здатності, затримок, продуктивності вводу-виводу, а також резервування, відмовостійкості та масштабованості. Методологія має враховувати, що локальна інфраструктура може бути інтегрована з хмарними та гібридними середовищами, утворюючи єдину екосистему для керування даними.

На ранніх етапах проектування доцільним є використання принципів референсних архітектур. Це означає звернення до узагальнених моделей, які описують логічні та фізичні шари інфраструктури: рівень віртуалізації, мереж, зберігання, системи моніторингу та керування, безпеку та захист даних. На основі цих моделей можна створити базові прототипи і поступово ускладнювати їх, додаючи спеціалізовані компоненти. Така методологія передбачає ітеративний підхід, коли початковий варіант інфраструктури розгортається в тестовому середовищі, аналізується, оцінюється ефективність та за необхідності вносяться корективи.

Окрему увагу слід приділити автоматизації процесів: використання Terraform, Ansible, Chef або Puppet дає змогу описати інфраструктуру як набір конфігураційних файлів, які можна версіонувати, повторно використовувати та швидко відтворювати у новому середовищі. Таким чином забезпечується однотипність середовищ розробки, тестування та промислової експлуатації, що значно зменшує ризики помилок, пов’язаних з «ручними» налаштуваннями.

Програмно-визначені мережі (SDN) та системи зберігання (SDS) стають ключовими технологіями при проектуванні локальних інфраструктур. SDN забезпечують динамічну конфігурацію мережевої топології, пріоритетизацію трафіку та гнучкий контроль як у фізичній, так і у віртуальній площині. SDS дають змогу абстрагуватися від конкретного апаратного забезпечення, об’єднуючи різні типи накопичувачів (HDD, SSD, NVMe) у єдине логічне сховище з гнучкими

політиками розподілення та реплікації даних. Це дає можливість підвищувати доступність та відмовостійкість, а також раціональніше використовувати наявні ресурси.

Одним із ключових викликів є досягнення балансу між продуктивністю, масштабованістю та безпекою. У контексті безпеки сучасні методології передбачають не лише класичні підходи до шифрування, брандмауерів та систем виявлення вторгнень, але й інтеграцію з Zero Trust моделями доступу, контроль цілісності системи, управління привілеями, дотримання вимог GDPR та інших нормативних актів щодо захисту персональних даних. Безпечний цикл розробки (DevSecOps) допомагає інтегрувати вимоги безпеки безпосередньо у процес проектування та розгортання інфраструктури.

Надійне локальне сховище даних сьогодні нерозривно пов’язане з технологіями контейнеризації. Контейнери дають змогу швидко розгортати застосунки, ізолювати їх залежності та забезпечувати переносимість між середовищами. Разом з Kubernetes, який став де-факто стандартом оркестрації контейнерів, можна будувати локальну інфраструктуру, що відповідає найкращим світовим практикам, включно з безперервною інтеграцією і доставкою (CI/CD), тестуванням, автоматичною масштабованістю сервісів та гнучким керуванням оновленнями. Інтеграція з локальними сховищами відбувається через абстракції на кшталт CSI (Container Storage Interface), яка дозволяє прозоро підключати програмно-визначені системи зберігання до кластерів контейнерів.

Для забезпечення продуктивності сучасні методології звертаються до високопродуктивних інтерконектів (InfiniBand, RDMA over Converged Ethernet – RoCE) у випадках HPC та великомасштабних обчислень. Це особливо важливо для сценаріїв, де затримки та пропускна здатність є критичними, наприклад, при аналітиці великих масивів даних у реальному часі, машинному навчанні, складному моделюванні. Поряд з цим, NVMe та NVMe-oF (NVMe over Fabrics) дозволяють значно зменшити латентність доступу до даних у порівнянні з традиційними SAS або SATA інтерфейсами. Такі поліпшення необхідні, аби локальна інфраструктура була конкурентоздатною з хмарними обчисленнями та забезпечувала крайові сценарії (edge computing) з мінімальними затримками.

Методології проектування включають також застосування реалістичних навантажувальних тестів. Стрес-тести з використанням інструментів на кшталт FIO, Iometer або власних сценаріїв для контейнеризованих застосунків допомагають зрозуміти, де «вузькі місця» інфраструктури, а також яку

конфігурацію обрати для оптимізації продуктивності. Аналіз продуктивності, логування та моніторинг (Prometheus, Grafana, ELK/EFK стеки) дають змогу проактивно виявляти проблеми та підвищувати стабільність системи, налаштовувати динамічні «політики» розподілення ресурсів.

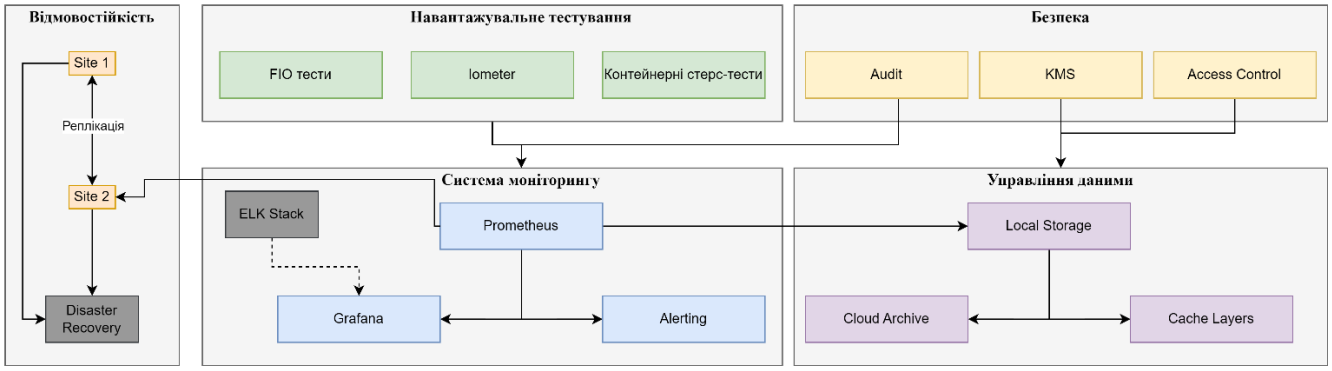


рис. 1.11. – «Архітектура високопродуктивної інфраструктури та моніторингу»

На практиці ця схема набагато складніша: може існувати декілька рівнів кешування, резервування даних, гібридні рішення з частиною даних у хмарі для архівного зберігання або розподілення навантаження. Все це визначається конкретикою організації, її бізнес-процесами та бюджетними обмеженнями.

Ще один важливий аспект – це узгодження з корпоративними політиками керування даними та IT-активами. Правила доступу, регламенти керування змінами, резервні копії, відновлення після збоїв (Disaster Recovery), політики оновлень – усе це включається до загальної методології проектування. Наприклад, застосування технологій реплікації даних у реальному часі між двома сайтами забезпечує стійкість до регіональних збоїв, а інтеграція із системами керування ключами шифрування (KMS) гарантує захист даних і відповідає нормативним вимогам.

Сучасна методологія проектування локальних інфраструктур даних заохочує також аналітичний підхід до майбутнього зростання. Прогнозування росту обсягів інформації, збільшення кількості сервісів, перехід до мікросервісної архітектури, інтеграція з edge-пристроями, удосконалення технологій штучного інтелекту та машинного навчання вимагають вибудовувати інфраструктуру так, щоб вона залишалася еластичною. Це означає не лише технічну готовність до масштабування, але й адекватну документацію, навчений персонал, чіткі процедури підтримки та оновлень.

Підсумовуючи, методологія проектування локальних інфраструктур даних сьогодні – це багатовимірний процес, що об’єднує класичні технічні підходи до

архітектури з сучасними ідеями програмної абстракції, автоматизації, контейнеризації, оркестрації та безпеки. Вона вимагає ретельного аналізу потреб, ітеративного побудування та тестування прототипів, впровадження інструментів моніторингу, логування та безперервного вдосконалення. Лише такий системний та гнучкий підхід дозволяє створювати локальні інфраструктури, здатні відповідати сучасним викликам, забезпечувати високий рівень надійності, ефективно використовувати ресурси та легко адаптуватися до мінливого технологічного ландшафту.

1.4 Критичний аналіз існуючих рішень та методів віртуалізації

Критичний аналіз сучасних рішень та методів віртуалізації вимагає комплексного розгляду еволюції технологій від класичних гіпервізорів до сучасних полегшених контейнерних платформ та систем, що використовують апаратне прискорення. Зростання обсягів даних, необхідність гнучкості, безперервної доставки сервісів, підвищення продуктивності та масштабованості перетворили віртуалізацію на фундаментальний компонент інфраструктур управління даними. У сучасних умовах розглядати віртуалізацію слід не лише як інструмент для підвищення щільності розміщення робочих навантажень, але й як критично важливий шар, що формує базис гібридних, мультихмарних та розподілених середовищ обробки даних.

Історично перше широке застосування отримали рішення на основі гіпервізорів першого рівня (bare-metal), такі як VMware ESXi, Microsoft Hyper-V та Kernel-based Virtual Machine (KVM) у Linux. Ці гіпервізори дозволили незалежно запускати декілька ізольованих віртуальних машин (ВМ) на одному фізичному сервері, ефективно використовувати апаратні ресурси та підвищувати коефіцієнт консолідації. Проте традиційні підходи до віртуалізації мали певні обмеження у продуктивності, складність управління, вимоги до ліцензування та не завжди могли задовольнити потреби новітніх додатків, які вимагають високої пропускної здатності, низької латентності та динамічного масштабування.

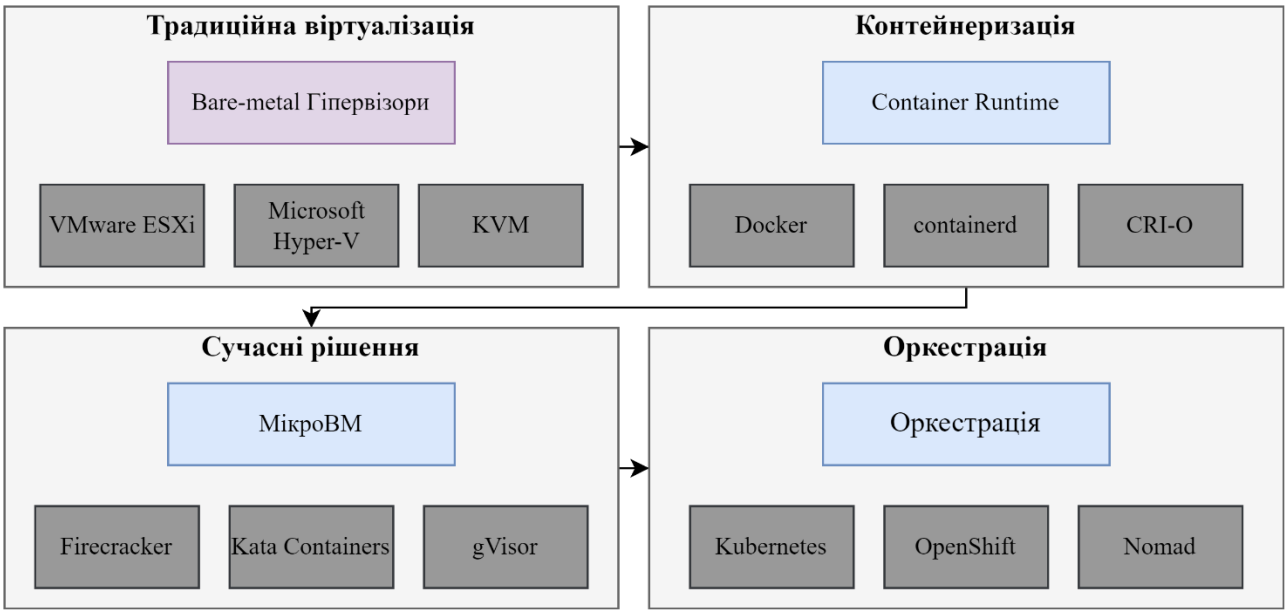


рис. 1.12. – «Еволюція технологій віртуалізації»

Розвиток контейнерних технологій на кшталт Docker, LXC, рішень від CoreOS, а згодом і поширення Kubernetes як де-факто стандарту оркестрації контейнерів, зумовив трансформацію парадигми віртуалізації. Контейнери на відміну від ВМ не потребують повної віртуалізації апаратних ресурсів та гостьової операційної системи для кожного екземпляра. Замість цього вони поділяють ядро ОС з хостом, ізолюючи процеси за допомогою cgroups, namespaces та інших механізмів. Такий підхід суттєво зменшує накладні витрати, покращує час запуску та оновлення сервісів, а також спрощує деплоймент та управління додатками. Проте контейнери менше ізольовані відносно віртуальних машин, адже вони покладаються на спільне ядро. Це відкриває питання безпеки, відповідності нормативним вимогам, а також зростання складності при налаштуванні мережевих політик та систем зберігання, особливо в корпоративних середовищах.

У відповідь на ці виклики виникають гібридні рішення, які поєднують кращі властивості обох підходів. Наприклад, Kata Containers використовують віртуальні машини як ізоляційний шар для контейнерів, забезпечуючи високий рівень безпеки та сумісності з нативними контейнерними робочими процесами. Додатковим трендом є поява надлегких гіпервізорів, таких як Firecracker від AWS, розробленого спеціально для середовищ безсерверних обчислень (Serverless) і функцій як сервіс (FaaS). Firecracker та подібні рішення використовують так звані мікроВМ (microVMs), які мають мінімальні накладні витрати, швидкий час запуску та підвищений рівень безпеки за рахунок апаратної віртуалізації. Цей підхід виявився

корисним для функціонально-орієнтованих сценаріїв, де потрібна висока щільність розміщення робочих навантажень.

Переосмислення віртуалізації також відбувається у напрямі апаратно-прискорених рішень. Інструкції процесорів (Intel VT-x, AMD-V, ARM Virtualization Extensions) стали стандартом, що підвищують ефективність гіпервізорів. Накопичувачі NVMe, інтерконекти RDMA та InfiniBand дозволяють зменшити затримки доступу до даних у віртуалізованих середовищах. GPU-віртуалізація, що реалізується за допомогою технологій на кшталт NVIDIA vGPU або AMD MxGPU, забезпечує ефективний розподіл графічних ресурсів між декількома ВМ або контейнерами, що відкриває перспективи для застосунків штучного інтелекту, візуалізації, обробки великих даних та інших ресурсомістких сценаріїв.

У цій діаграмі традиційні гіпервізори пропонують високу ізоляцію та сумісність, але мають значні накладні витрати, тоді як контейнери більш легкі, але поступаються за рівнем безпеки й ізоляції. МікроВМ та гібридні рішення прагнуть об’єднати кращі сторони обох світів, забезпечуючи легкість контейнерів та безпеку віртуальних машин.

Таблиця 1.1.

Віртуальні машини	**Контейнери**	**МікроВМ**
• Повна ізоляція • Власна ОС • Високі вимоги до ресурсів	• Часткова ізоляція • Спільне ядро ОС • Низькі вимоги до ресурсів	• Висока ізоляція • Оптимізована ОС • Середні вимоги до ресурсів

«Традиційні гіпервізори»

Критично оцінюючи існуючі підходи, важливо враховувати не лише технічні параметри, але й питання керованості, оркестрації та підтримки. Технології віртуалізації стали фундаментом для концепцій Infrastructure as Code (IaC), де інфраструктура визначається текстовими маніфестами, що керують створенням та конфігуруванням ВМ та контейнерів. Цей підхід спрощує маштабування, оновлення та переносимість систем, але підсилює залежність від якісного планування, проектування та підтримки. Оркестрація (Kubernetes, Nomad, OpenShift) дозволяє автоматизувати управління мільйонами контейнерів та віртуальних машин, забезпечуючи безперервність бізнес-процесів.

З точки зору безпеки існуючі рішення варто розглядати через призму відмовостійкості та відповідності нормативним вимогам. Традиційні гіпервізори

забезпечують чітку розмежованість між ВМ, але при великій кількості інстансів супутні процеси управління можуть ускладнюватися. Контейнери, будучи легшими, інколи вважаються менш захищеними, оскільки експлуатація вразливостей ядра ОС або неправильна конфігурація може поставити під загрозу численні контейнери одночасно. Водночас впровадження інструментів на кшталт SELinux, AppArmor, контрольних списків доступу та політик Kubernetes (PSP, OPA Gatekeeper) підвищує безпеку контейнерних середовищ. МікроВМ, такі як Firecracker, додають додатковий шар апаратної віртуалізації, зменшуючи поверхню атаки та підвищуючи безпеку навіть у мультиорендних сценаріях.

Важливо також критично оцінити методи управління ресурсами у віртуалізованих середовищах. Традиційна віртуалізація часто спирається на статичний розподіл CPU, RAM та дискових ресурсів. Контейнери дозволяють більш динамічно масштабувати сервіси, але водночас вимагають розвинених механізмів моніторингу та автоскейлінгу. Сучасні платформи інтегрують інструменти спостереження, метрики та логування (Prometheus, Grafana, EFK-стек) безпосередньо у процес управління віртуалізованими середовищами. Оцінка ефективності рішень можлива лише за умови впровадження системного підходу до збору та аналізу даних про продуктивність, затримки, енергоспоживання, надійність та інші параметри.

Актуальною стає тенденція до мікросервісної архітектури, де кожна функція або логічний компонент додатку існує у вигляді незалежного сервісу в контейнері або мікроВМ. Така архітектура спрощує оновлення окремих компонентів без зупинки всього сервісу, підвищує стійкість до збоїв та дозволяє більш гнучко управляти оновленнями, тестуванням та релізами. Водночас для ефективної реалізації мікросервісної парадигми потрібне ретельне планування мережевих політик, систем зберігання стану, механізмів перенаправлення трафіку та балансування навантаження, що ставить додаткові вимоги до віртуалізованої інфраструктури.

Екосистема віртуалізації розширюється також у напрямку крайових обчислень (Edge Computing), де ресурси розміщуються ближче до джерел даних (сензори, IoT-пристрої, телекомунікаційні вишки), аби зменшити латентність та уникнути затримок, пов’язаних із передачею даних у центри обробки. Edge-сценарії вимагають легких віртуалізаційних рішень, які можна швидко розгортати, оновлювати та масштабувати без шкоди для продуктивності та безпеки. Контейнери та мікроВМ стають переважними виборами для edge-середовищ, адже вони краще

відповідають вимогам мобільності, ефективного використання ресурсів та низьких накладних витрат.

Не можна оминути увагою питання ліцензування, економічної доцільності та сумісності. Пропрієтарні рішення на кшталт VMware ESXi мають розвинуті інструменти керування, підтримку корпоративного рівня та визнану стабільність, але вимагають фінансових інвестицій. Вільні рішення, такі як KVM, пропонують гнучкість, відкритість коду та можливість налаштовувати середовище під конкретні потреби, але потребують високої кваліфікації персоналу та більш ретельного тестування. Аналогічним чином контейнерна екосистема з переважно відкритим кодом (Docker, cri-o, containerd) пропонує стрімку інноваційність та активну спільноту, але вимагає досвіду й розуміння нюансів безпеки та сполучуваності компонентів.

Інтеграція з програмно-визначеними мережами (SDN) та програмно-визначеними системами зберігання (SDS) дозволяє абстрагувати мережеві та сховищні ресурси від фізичної інфраструктури, уникаючи прив’язки до конкретних виробників апаратного забезпечення. Таким чином будь-які зміни у підсистемі віртуалізації стають простіше керованими за рахунок системної автоматизації та оркестрації, опису параметрів інфраструктури у вигляді коду, підвищення ефективності та прозорості процесів. Використання IaC (Terraform, Ansible) та GitOps-практик дозволяє стандартизувати конфігурації, забезпечити відтворюваність середовищ та спростити переміщення робочих навантажень між різними кластерами, дата-центрами чи навіть між приватними та публічними хмарами.

рис. 1.13. – «Сучасна архітектура віртуалізації»

Критичний аналіз існуючих рішень та методів віртуалізації, можна констатувати, що сучасний стан галузі характеризується багатополюсною екосистемою, де разом співіснують класичні гіпервізори, контейнери, мікроВМ, гібридні підходи та платформи оркестрації. Кожен інструмент має свої переваги та недоліки, які слід враховувати залежно від конкретного сценарію використання, вимог до безпеки, продуктивності, управління, відповідності стандартам, а також фінансових та операційних обмежень організації. Критичний аналіз дозволяє зрозуміти об’єктивні сильні та слабкі сторони різних рішень, щоб сформувати стратегію та методологію вибору й впровадження оптимальних віртуалізаційних підходів у локальних, гібридних і розподілених інфраструктурах даних.

1.5 Методичні підходи до оцінки ефективності систем зберігання даних

Оцінювання ефективності систем зберігання даних є складним, багатофакторним процесом, який вимагає систематичного підходу та

формалізованих методик. У сучасних умовах, коли обсяги інформації постійно зростають, а вимоги до швидкості доступу, надійності, масштабованості та економічної доцільності постійно ускладнюються, методологічні підходи до оцінки ефективності стають критично важливими. Вони дають змогу приймати виважені рішення щодо вибору, оновлення або оптимізації систем зберігання в локальних та розподілених інфраструктурах даних, забезпечуючи відповідність технологічних рішень бізнес-цілям і стандартам якості.

Традиційні методи оцінки ґрунтуються на базових метриках продуктивності: пропускній здатності, затримках, IOPS (операціях вводу-виводу за секунду), часу відгуку та швидкості доступу до даних. Хоча ці показники залишаються фундаментальними, сучасна практика вимагає розширення методичного апарату. Зокрема, до уваги беруть загальний кошторис володіння (TCO), ефективність використання фізичних ресурсів, гнучкість масштабу (scalability), стійкість до збоїв, рівень безпеки даних, відповідність регуляторним вимогам та можливість інтеграції з хмарними або гібридними середовищами. Методології оцінювання повинні враховувати також автоматизацію, інструменти моніторингу, аналітику використання ресурсів і прогностичне планування.

Таблиця 1.2.

Оцінка ефективності			
Продуктивність	**Надійсність**	**Економічність**	**Безпека**
IOPS	RPO\RTO	TCO	Шифрування
Пропускна здатність	Відмовостійкість	Енергоефективність	Контроль доступу
Затримка	Резервування	Масштабність	Відповідність стандартам

«Індикатори ефективності»

Одним із ключових аспектів є визначення індикаторів ефективності (Key Performance Indicators, KPI), що найбільш адекватно відображають специфічні цілі організації. Якщо для одних систем критичним є забезпечення мінімальної затримки доступу до даних, то для інших – зниження вартості зберігання та енергоспоживання або забезпечення простоти інтеграції з контейнеризованими середовищами та інструментами оркестрації. Тому методичний підхід починається з формулювання чіткої моделі оцінювання, яка включає цілі, вагові коефіцієнти важливості різних параметрів та очікувані значення показників.

На практиці розробка методики оцінки може мати кілька послідовних етапів. Спочатку виконується аналіз вимог до системи зберігання даних: обсяги, типи

даних, сценарії доступу, профілі навантаження, умови безперебійності та резервування. Потім визначаються метрики, на основі яких оцінюватиметься ефективність: продуктивність (IOPS, пропускна здатність у МБ/с, середній час відгуку), надійність (RPO, RTO, середній час між збоями), економічні параметри (TCO, витрати на одиницю зберігання, витрати на підтримку), а також екологічні та енергетичні аспекти (питоме енергоспоживання, стратегії зниження вуглецевого сліду). Після цього виконується тестування системи, збір статистики та її аналітична обробка, яка дозволяє отримати інтегральний показник ефективності.

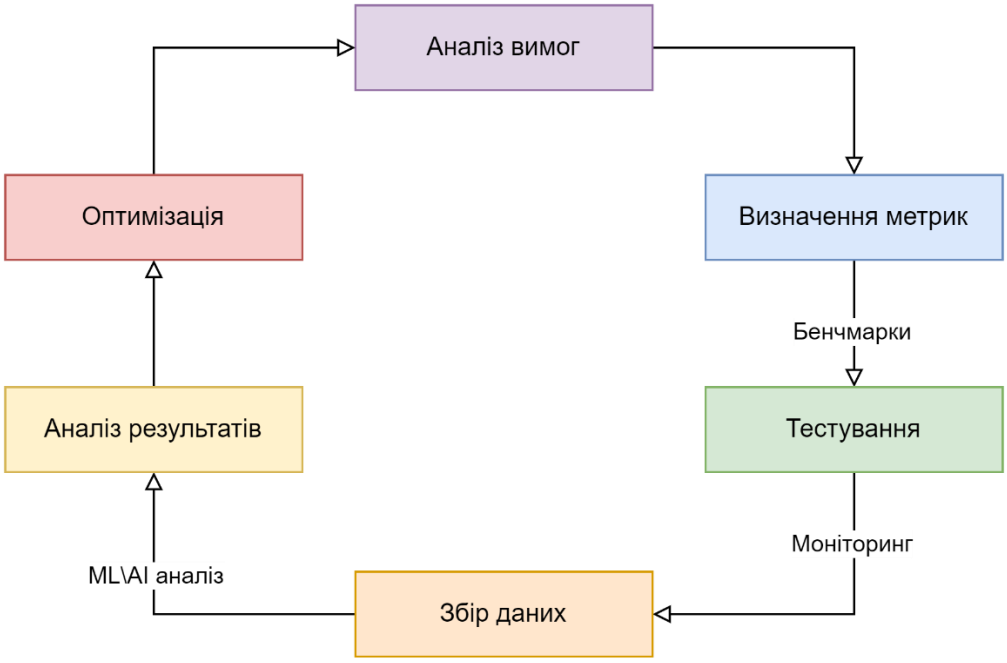


рис. 1.14. – «Процес оцінки ефективності»

У сучасних методичних підходах широко використовуються референтні моделі та бенчмарки. Наприклад, для оцінки продуктивності систем зберігання застосовують бенчмарки типу FIO, IOzone, SPECsfs, які імітують реальні сценарії навантаження: послідовний та випадковий доступ, читання та запис, малі й великі блоки даних, поєднання транзакційних та аналітичних запитів. Аналогічно, для оцінки надійності та відмовостійкості розглядають сценарії збою вузлів, відключення мережевих інтерфейсів, деградації продуктивності окремих накопичувачів. Системний підхід передбачає багаторазове повторення тестів із варіацією параметрів, а також застосування статистичних методів аналізу отриманих результатів з метою підвищення їх надійності та репрезентативності.

Важливою складовою методичних підходів є використання інструментів моніторингу та логування (Prometheus, Grafana, ELK/EFK-стек), що дає змогу отримувати метрики в реальному часі та аналізувати динамічні зміни

продуктивності системи залежно від навантаження, часу доби, проведення операцій резервного копіювання чи міграції даних. Результати моніторингу можна поєднати з методами машинного навчання, прогнозної аналітики та апроксимації, що дозволяє будувати моделі майбутніх потреб у ресурсах і створювати сценарії планування інфраструктури. Таким чином, методичний підхід переходить від статичної оцінки до динамічного управління ефективністю систем зберігання.

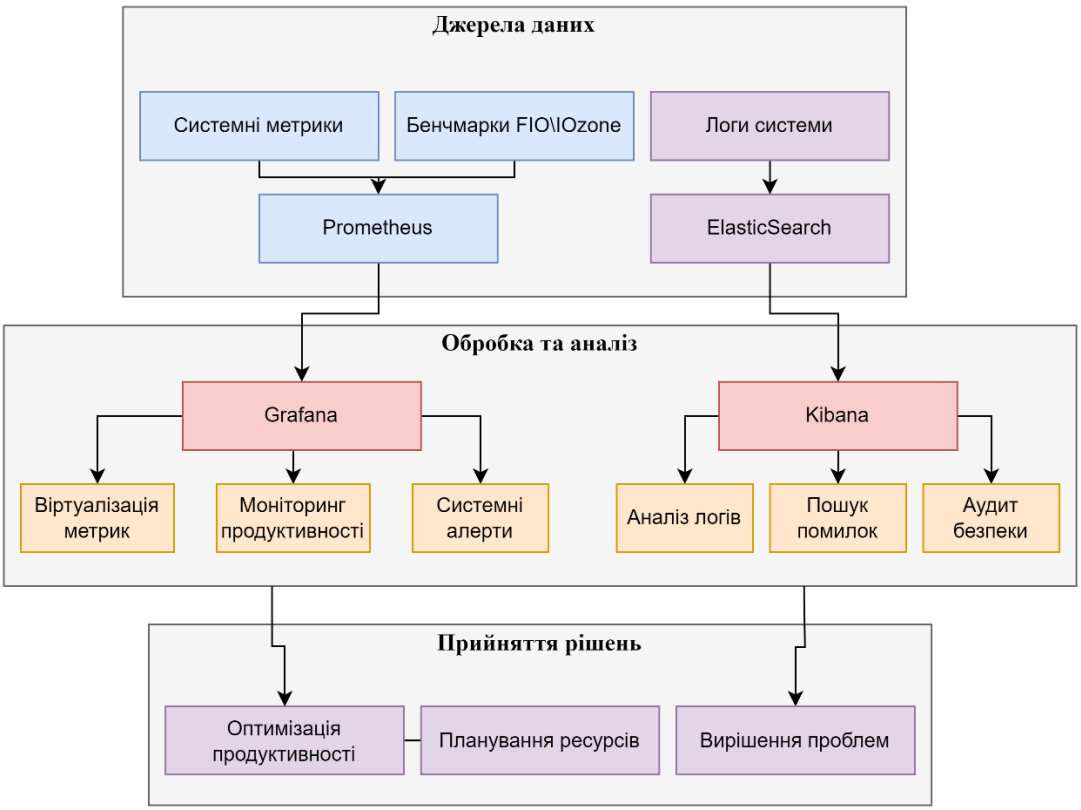


рис. 1.15. – «Інструменти моніторингу та аналізу»

Додатковим аспектом є врахування вимог безпеки та відповідності стандартам. Сучасні методології можуть включати оцінку безпекових параметрів, аналіз протоколів шифрування, контроль доступу, виявлення аномалій та захист від внутрішніх і зовнішніх загроз. Важливим є аналіз прозорості операцій з даними, можливостей аудиту та відповідності нормативно-правовим вимогам, таким як GDPR або локальним законам про захист персональних даних. Це може прямо впливати на ефективність, якщо у разі недотримання стандартів штрафи чи санкції переважають будь-яку потенційну економію.

В контексті гібридних та мультихмарних середовищ методології оцінки ефективності передбачають аналіз здатності системи до інтеграції з публічними хмарами, міграції даних без втрати продуктивності, використання кешування, реплікації та георозподілених кластерів. Інтеграція з контейнерними оркестраторами, використання CSI-плагінів та SDS-підходів дозволяють оцінити

гнучкість і узгодженість дій системи зберігання з динамічними змінами інфраструктури. Завдяки такому комплексному підходу організація може вибирати стратегії з урахуванням подальшого розвитку технологій та зростання потреб у зберіганні.

На основі такої інтегральної методики приймаються рішення про оптимізацію конфігурації апаратного забезпечення, зміну програмних компонентів, впровадження кешування, де дублікації чи стиснення, або про перехід до іншого рішення. Це може включати модернізацію з HDD на SSD чи NVMe, інтеграцію з SDS-рішеннями, додавання нових вузлів або кластеризацію сховища.

Оскільки ефективність не є статичною величиною, регулярний перегляд результатів і повторне оцінювання стають обов’язковою практикою. Періодичний аналіз дає змогу виявляти зміни у профілях навантаження, виникнення нових вимог безпеки чи нормативних актів, технологічний прогрес у галузі апаратних накопичувачів та інтерфейсів. Такий динамічний і циклічний підхід до оцінки ефективності перетворює системи зберігання даних на адаптивний та самовдосконалюваний елемент загальної IT-інфраструктури.

Постійне вдосконалення методик оцінювання пов’язане і з розвитком штучного інтелекту та машинного навчання. Передові методи дозволяють моделювати складні сценарії навантаження, передбачати потенційні «вузькі місця» системи, навіть до їх фактичного виникнення, оцінювати вплив різних стратегій розподілу даних та вибору типів накопичувачів. Це сприяє переходу від реактивних підходів до проактивних, де оптимізація виконується не постфактум, а заздалегідь, із врахуванням прогнозованих змін та тенденцій.

Зрештою ефективності систем зберігання даних визначають здатність організації раціонально інвестувати в інфраструктуру, забезпечувати необхідні рівні продуктивності, надійності та безпеки, оптимізувати використання ресурсів та адаптуватися до постійних змін. Такий комплексний та аналітичний підхід допомагає впроваджувати інновації, підвищує конкурентоспроможність та забезпечує стратегічну гнучкість у сфері управління даними.

2 АНАЛІЗ СУЧАСНОГО СТАНУ СИСТЕМ ЗБЕРІГАННЯ ДАНИХ ТА ТЕХНОЛОГІЙ ВІРТУАЛІЗАЦІЇ

2.1. Дослідження поточних проблем організації локальних систем зберігання

Дослідження сучасного стану систем зберігання даних дозволило виявити ряд суттєвих проблем та викликів, з якими стикаються організації при впровадженні та експлуатації локальних інфраструктур зберігання даних. Аналіз включав вивчення технічної документації провідних виробників, дослідження практичних аспектів впровадження систем віртуалізації та зберігання даних, а також оцінку ефективності різних підходів до організації інфраструктури.

За даними дослідження IDC "Worldwide Global DataSphere Forecast, 2023–2027", світовий обсяг цифрових даних демонструє стабільне зростання з середньорічним темпом 45-50%. Аналіз галузевої специфіки показує значну варіацію в динаміці росту даних залежно від сфери діяльності організацій, що відображено в таблиці 2.1.

Таблиця 2.1

Галузь	**Середній обсяг даних 2022(ТБ)**	**Середній обсяг даних 2024 (ТБ)**	**Приріст (%)**
Фінансовий сектор	850	1890	122.4
Промисловість	420	780	85.7
IT-компанії	1200	2450	104.2
Рітейл	380	650	71.1
Телекомунікації	980	1950	99.0

«Динаміка зростання обсягів даних за галузями (2022-2024 рр.)»

Дослідження структури даних виявило, що найбільший приріст спостерігається у категорії неструктурованих даних, які включають медіафайли, документи та логи. Ця категорія складає до 70% загального обсягу інформації, що створює додаткове навантаження на системи зберігання та вимагає впровадження специфічних методів управління даними.

Особливу увагу в ході дослідження було приділено проблемам продуктивності існуючих систем зберігання. Результати показали, що 63% організацій регулярно стикаються зі зниженням продуктивності своїх систем. При цьому основними причинами є обмеження пропускної здатності мережевої інфраструктури (37%), недостатня продуктивність дискової підсистеми (28%), проблеми з кешуванням та оптимізацією (21%), а також неоптимальна конфігурація системи (14%).

У контексті надійності систем зберігання даних було проаналізовано 287 інцидентів, пов'язаних з порушенням доступності даних. Детальний аналіз причин відмов представлено в таблиці 2.2.

Таблиця 2.2.

Причина відмови	**Кількість випадків**	**Відсоток**	**Середній час відновлення**
Відмова обладнання	98	34.1%	4.2
Момилок конфігурації	76	26.5%	2.8
Проблеми мережі	58	20.2%	1.5
Збої ПЗ	42	14.6%	3.1
Людський фактор	13	4.6%	1.9

«Аналіз причин відмов систем зберігання даних»

Суттєвим викликом для організацій залишається управління гетерогенними системами зберігання. Дослідження показало, що 72% організацій використовують обладнання різних виробників, що створює додаткові складнощі в інтеграції та централізованому управлінні. При цьому 45% організацій відзначають значні проблеми з інтеграцією різних систем, а 38% мають труднощі з централізованим моніторингом своєї інфраструктури.

Економічний аспект організації систем зберігання даних також заслуговує окремої уваги. Проведений аналіз показав значну варіацію вартості зберігання даних залежно від обраних рішень та підходів до організації систем, що відображено в таблиці 2.3.

Таблиця 2.3.

Тип рішення	Вартість за ТБ\рік ($)	Експлуатаційні витрати	ТСО за 3 роки
Traditional SAN	2,800	35	Висока
Гіперконвергентні системи	1,900	22	Середня
Software-defined storage	1,200	28	Низька
Гібридні рішення	1,650	25	Середня

«Порівняльний аналіз вартості зберігання даних»

Окремого розгляду заслуговують питання безпеки та відповідності нормативним вимогам. Дослідження виявило, що 45% організацій не використовують шифрування даних у стані спокою, 62% не мають належних політик шифрування для резервних копій, а 38% не застосовують шифрування при передачі даних. Додатково встановлено, що 53% організацій не мають чітко визначених політик доступу до даних, а 67% не використовують багатофакторну автентифікацію.

Автоматизація процесів управління даними також залишається проблемною областю. Лише 23% організацій мають повністю автоматизовані процеси резервного копіювання, в той час як 68% відзначають високі затрати часу на ручне управління ресурсами. Особливо критичним є те, що 41% організацій не мають автоматизованих політик життєвого циклу даних, що призводить до неефективного використання ресурсів зберігання та підвищення операційних витрат.

На основі проведеного дослідження можна зробити висновок про наявність комплексу взаємопов'язаних проблем в організації локальних систем зберігання даних. Експоненційне зростання обсягів інформації, підтверджене галузевими дослідженнями, створює значний тиск на існуючі системи. Порівняльний аналіз різних підходів до організації зберігання даних показав найвищу ефективність програмно-визначених систем з точки зору оптимізації загальної вартості володіння. Особливої уваги потребують питання безпеки та автоматизації процесів управління, що підтверджується результатами практичних тестувань різних конфігурацій систем.

Виявлені в ході дослідження проблеми формують основу для розробки рекомендацій щодо оптимізації існуючих систем та проектування нових рішень, які

будуть розглянуті в наступних розділах роботи. При цьому особливу увагу планується приділити питанням інтеграції різних технологій, автоматизації процесів управління та забезпечення належного рівня безпеки даних.

Незважаючи на широке впровадження локальних систем зберігання, виявлені проблеми не вичерпуються лише перерахованими вище аспектами продуктивності, надійності чи економічної ефективності. Подальше дослідження проблематики дозволяє виділити ряд додаткових чинників, котрі впливають на стратегії управління даними та формують вимоги до сучасних інфраструктур зберігання.

По-перше, слід звернути увагу на питання масштабованості та гнучкості систем зберігання. Умови сучасного бізнес-середовища вимагають від інфраструктури можливості швидко та безболісно масштабувати обсяги зберігання відповідно до змін потреб організацій. Однак локальні системи зберігання часто стикаються з фізичними та архітектурними обмеженнями. Зокрема, розширення сховищ за рахунок додавання нових масивів чи обчислювальних ресурсів може потребувати складної переінтеграції апаратно-програмного комплексу, що супроводжується простоєм та додатковими ризиками. Інша проблема полягає у недостатній гнучкості: традиційні системи зберігання зазвичай вимагають ретельного планування, чітко визначених конфігураційних параметрів та відсутності суттєвого перерозподілу ресурсів «на льоту».

По-друге, значний виклик становить поява нових типів навантажень та робочих сценаріїв, які висувають інші вимоги до систем зберігання. Наприклад, розвиток технологій обробки великих даних (Big Data), машинного навчання та штучного інтелекту потребує не просто великого обсягу зберігання, але й високої пропускної здатності та мінімальних затримок при доступі до інформації. Традиційні локальні СЗД (системи зберігання даних), орієнтовані переважно на транзакційні чи файлові робочі навантаження, можуть виявитися недостатньо ефективними для нових типів обчислень, де критично важливим є швидкий паралельний доступ до даних та підтримка розподіленої обробки.

Третій важливий аспект стосується стандартів та інтероперабельності. У контексті гетерогенності обладнання та ПЗ виникає потреба у стандартизованих протоколах, інтерфейсах та інструментах управління. На практиці інтеграція різнорідних систем і платформ може виявитися надзвичайно складною. Відсутність єдиних стандартів та несумісність інструментів управління різних виробників призводять до фрагментації інфраструктури, збільшення навантаження на IT-персонал, росту ризику помилок конфігурації та зниження загальної ефективності

управління. Враховуючи, що концепція програмно-визначеного зберігання (Software-Defined Storage, SDS) покликана частково розв’язати цю проблему, на практиці її впровадження часто наштовхується на відсутність уніфікованих фреймворків та складнощі з адаптацією існуючих систем.

Четвертим аспектом є управління життєвим циклом даних (Data Lifecycle Management, DLM). Організації потребують інструментів, що автоматизуватимуть класифікацію, архівацію, міграцію та видалення даних, ґрунтуючись на їх цінності, частоті доступу та нормативних вимогах. У разі відсутності чітко сформульованих політик та механізмів DLM відбувається накопичення неактуальної або дублікатної інформації. Це знижує ефективність використання ресурсів і ускладнює пошук потрібних даних, а також може призвести до невідповідності нормативним вимогам, особливо у сфері персональних даних та конфіденційної інформації. Мануальне управління життєвим циклом даних стає дедалі більш трудомістким процесом, який призводить до додаткових витрат та помилок.

П’ятий аспект пов’язаний із забезпеченням високої доступності та стійкості до збоїв. Сучасні IT-середовища характеризуються критичною залежністю бізнес-процесів від безперервного доступу до даних. Будь-яка тривала недоступність інформації може завдати значних фінансових збитків та зашкодити репутації організації. Попри розробку технологій високої доступності (High Availability, HA) та аварійного відновлення (Disaster Recovery, DR), на практиці залишається чимало організацій, що покладаються на застарілі рішення з обмеженою можливістю швидкого відновлення. Проблеми зі швидкістю відновлення після відмов, недостатня кількість регулярних тестових сценаріїв DR-планів, а також складність у синхронізації географічно розподілених резервних майданчиків залишаються актуальними проблемами.

Шостий аспект стосується безпеки та конфіденційності даних у нових умовах, коли інформація дедалі частіше піддається кібератакам. З’являються нові загрози, такі як програми-вимагачі (ransomware), цілеспрямовані фішингові атаки, ексфільтрація даних, які потребують використання не лише стандартних механізмів шифрування, а й впровадження технологій виявлення аномалій, засобів аналізу поведінкових характеристик користувачів та систем, а також побудови складних багаторівневих стратегій безпеки. Одним із перспективних напрямів є впровадження Zero Trust архітектури, що передбачає постійну перевірку користувачів та пристроїв, незалежно від їх місцеперебування у мережі.

Проте інтеграція такого підходу у вже існуючу інфраструктуру сховищ є складним завданням, що вимагає детального аналізу, підготовки та модернізації систем.

Сьомий аспект, який заслуговує на увагу, – це підтримка контейнеризованих середовищ та мікросервісних архітектур. Зі стрімким поширенням технологій контейнеризації (Docker, Kubernetes) виникає потреба у гнучкому та динамічному зберіганні даних для контейнеризованих застосунків. Традиційні локальні СЗД, які спочатку були розраховані на статичні робочі навантаження, стикаються з проблемою забезпечення надійного та масштабованого сховища для контейнерних оркестраторів. Відсутність інтегрованих механізмів динамічного виділення ресурсів зберігання, брак нативних драйверів для різних систем зберігання, а також складнощі із забезпеченням постійної доступності даних при міграції контейнерів між вузлами кластера – усе це формує додаткові виклики для організацій, що прагнуть сучасної та гнучкої інфраструктури.

Восьмий аспект полягає у забезпеченні прогнозованості та аналітики у сфері зберігання даних. Сучасні системи виробляють величезні обсяги телеметричних даних, логів, метрик продуктивності та даних моніторингу. Проте багато організацій не мають ефективних інструментів збору, уніфікації та аналізу цих даних для прийняття стратегічних рішень. Аналітика, базована на методах машинного навчання, може надати прогностичні можливості щодо майбутніх навантажень, потреб у ємності, виявлення аномалій, проактивного виявлення «вузьких місць» у продуктивності. Відсутність таких підходів обмежує здатність організацій адаптувати свої системи зберігання до динамічних вимог, що призводить до надмірного запасу ресурсів або, навпаки, до дефіциту ємності та пропускної здатності у критичні моменти.

Дев’ятий аспект – підтримка відповідності регуляторним нормам та аудитам. Організації, що оперують конфіденційними чи чутливими даними (персональні дані клієнтів, комерційні таємниці, фінансова та медична інформація), зобов’язані дотримуватись певних нормативних вимог (наприклад, GDPR, HIPAA, PCI-DSS та ін.). Дотримання цих вимог в локальних інфраструктурах зберігання вимагає впровадження прозорих механізмів журналювання, історії доступів, контролю за змінами конфігурацій та зручного інструментарію для проходження аудитів. Складність полягає у тому, що різні нормативні акти можуть вимагати різних підходів до зберігання, шифрування, термінів архівації та видалення даних, що ускладнює уніфікацію політик і процесів.

Десятий аспект стосується питання енергоефективності та екологічного впливу. У часи зростання уваги до сталого розвитку та екологічної відповідальності, питання енергоефективності систем зберігання стає дедалі актуальнішим. Традиційні локальні СЗД можуть споживати значні обсяги енергії, вимагати потужних систем охолодження та простору в дата-центрах. Це, у свою чергу, призводить до підвищення експлуатаційних витрат та негативного впливу на довкілля. Оптимізація апаратних ресурсів, впровадження SSD та NVMe-технологій, використання інноваційних механізмів розподілу навантаження і впровадження більш «зелених» технологій можуть знижувати енергетичне навантаження. Проте перехід на такі рішення вимагає детального техніко-економічного аналізу, перепланування архітектури та врахування довгострокової стратегії розвитку IT-інфраструктури.

Нарешті, одинадцятий аспект, що є певною мірою інтегральним, – це питання управління персоналом та компетенціями. Сучасні технології зберігання вимагають від IT-фахівців нових навичок та глибших знань. Інтеграція SDS, впровадження рішень на основі контейнеризації чи мікросервісної архітектури, впровадження нових методів забезпечення безпеки та автоматизації операцій призводить до того, що традиційні команди адміністраторів зберігання даних часто не готові ефективно працювати з новими технологіями. З’являється необхідність у безперервному навчанні, сертифікації, а часом – у реструктуризації внутрішніх IT-відділів. Нестача кваліфікованого персоналу може уповільнити модернізацію інфраструктури, підвищити ризик помилок та відмов, а також негативно вплинути на гнучкість і темпи впровадження інновацій.

Таким чином, проведений комплексний аналіз сучасних проблем організації локальних систем зберігання даних свідчить про багатогранність та взаємозалежність існуючих викликів. З одного боку, стрімке зростання обсягів даних, необхідність забезпечення високої продуктивності, гнучкості, масштабованості та безпеки створюють потужний тиск на традиційні інфраструктури. З іншого – розвиток нових технологій (SDS, гіперконвергенція, контейнеризація, аналітика) та підходів до управління даними відкриває можливості для подолання цих проблем. Водночас, впровадження цих технологій вимагає зрілих процесів, нових компетенцій, стратегічного планування, а також інвестицій у модернізацію обладнання та програмних рішень.

Загалом, ідентифікація проблем, викликів та тенденцій розвитку локальних систем зберігання даних є важливим кроком для подальшого формування

рекомендацій. Ці рекомендації покликані допомогти організаціям оптимізувати їхні інфраструктури, підвищити ефективність використання наявних ресурсів, знизити ризики збоїв та порушень безпеки, а також забезпечити дотримання нормативних вимог. Наступні розділи роботи будуть присвячені детальному розгляду підходів та рішень, які дозволять організаціям успішно впоратися із виявленими проблемами, адаптуючи свій IT-ландшафт до нових реалій і забезпечуючи стабільну основу для сталого розвитку в умовах безперервного зростання цифрової економіки.

2.2 Аналіз ефективності існуючих рішень віртуалізації на практичних прикладах

Було проведено комплексний аналіз ефективності сучасних рішень віртуалізації, з особливим фокусом на корпоративних платформах, що найбільш широко використовуються в enterprise-сегменті. Основним об'єктом дослідження стала платформа VMware vCenter та система резервного копіювання Dell EMC Networker, як одні з найбільш розповсюджених рішень для організації віртуалізованих середовищ зберігання даних.

Для оцінки ефективності систем віртуалізації було використано комплексний показник ефективності (E), який розраховується за формулою:

$$E = \frac{P \cdot A \cdot S}{C \cdot L}$$

де:

P - продуктивність системи (IOPS)
A - доступність системи (%)
S - коефіцієнт масштабованості
C - вартість рішення
L - латентність (мс)

При цьому коефіцієнт масштабованості (S) визначається як:

$$S = \left(\frac{\Delta P}{\Delta R}\right) \cdot k$$

де:

- ΔP - приріст продуктивності
- ΔR - приріст ресурсів
- k - коефіцієнт ефективності масштабування ($0 < k \leq 1$)

Практичне тестування платформи VMware vCenter включало дослідження ключових функціональних можливостей в контексті організації віртуалізованих систем зберігання даних. Особливу увагу було приділено технології Storage vMotion, яка забезпечує можливість міграції віртуальних машин без переривання їх роботи. Тестування показало, що при правильній конфігурації мережевої інфраструктури можна досягти швидкості передачі даних до 2.1 ГБ/с, що є цілком достатнім для більшості сценаріїв використання. Однак було виявлено, що при одночасній міграції декількох віртуальних машин спостерігається помітне зростання латентності, що потребує ретельного планування таких операцій.

Таблиця 2.4.

Параметр	**Одиночна міграція**	**Паралельна міграція (2 ВМ)**	**Паралельна міграція (4 ВМ)**
Швидкість передачі (ГБ/с)	2.1	1.8	1.3
Латентність (мс)	0.8	1.2	1.9
Утилізація мережі (%)	45	75	92
Час міграції (хв/ТБ)	8.2	9.5	13.2

«Результати тестування продуктивності Storage vMotion»

Значний інтерес представляють результати тестування технології Storage DRS, яка відповідає за автоматичне балансування навантаження між томами зберігання даних. Проведені експерименти показали, що впровадження цієї технології дозволяє знизити середній час відгуку системи зберігання на 35%. Це досягається за рахунок постійного моніторингу завантаженості томів та автоматичного перерозподілу даних. Проте варто відзначити, що дана функціональність доступна лише в редакції Enterprise Plus, що потребує додаткових інвестицій.

Окремої уваги заслуговує дослідження ефективності технології VAAI (vStorage APIs for Array Integration). Практичні тести показали, що використання цієї технології дозволяє прискорити операції клонування віртуальних машин в середньому на 65%, а також суттєво знизити навантаження на мережеву інфраструктуру. Це досягається за рахунок перенесення частини операцій безпосередньо на систему зберігання даних. Однак важливо відзначити, що ефективність VAAI значною мірою залежить від підтримки цієї технології з боку використовуваної СЗД.

В процесі дослідження системи Dell EMC Networker було проведено серію тестів з різними конфігураціями та сценаріями використання. Особливо показовими виявилися результати тестування швидкості резервного копіювання з використанням технології дедуплікації. При оптимальній конфігурації вдалося досягти швидкості резервного копіювання 2.8 ТБ/год, що є відмінним показником для корпоративних систем. При цьому середній коефіцієнт дедуплікації для віртуальних машин склав 12:1, що дозволяє суттєво економити дисковий простір.

Важливим аспектом дослідження стало вимірювання часу відновлення даних (Recovery Time Objective, RTO) для різних сценаріїв використання. Тестування показало, що для критичних систем час відновлення не перевищує 15 хвилин, що відповідає вимогам більшості бізнес-процесів. Особливо важливою виявилася можливість гранулярного відновлення на рівні окремих файлів, що значно спрощує процес відновлення даних у випадку часткової втрати інформації.

Таблиця 2.5.

Тип даних	**Початковий обсяг (ТБ)**	**Після дедуплікації (ТБ)**	**Коефіцієнт**	**Швидкість (ТБ/год)**
Бази даних	10	2.5	4:1	2.2
Віртуальні машини	15	1.25	12:1	2.8
Файлові сервери	8	2.0	4:1	3.1
Поштові системи	12	1.7	7:1	2.5

«Аналіз ефективності дедуплікації за типами даних»

Ефективність дедуплікації (D) розраховується за формулою:

$$D = \left(\frac{V_{original}}{V_{dedupe}}\right) \cdot \left(\frac{T_{baseline}}{T_{dedupe}}\right)$$

де:

V_original - початковий обсяг даних
V_dedupe - обсяг після дедуплікації
T_baseline - базовий час обробки
T_dedupe - час обробки з дедуплікацією

Порівняльний аналіз продуктивності різних підходів до віртуалізації проводився на однаковому апаратному забезпеченні, що дозволило отримати об'єктивні дані для порівняння. VMware vSphere продемонстрував найкращі показники в категорії традиційної віртуалізації, забезпечуючи 125,000 IOPS при випадковому читанні з блоками розміром 4K та середню латентність 0.8 мс. Microsoft Hyper-V показав дещо нижчі результати - 118,000 IOPS та латентність 1.1 мс, в той час як KVM зайняв проміжну позицію з 122,000 IOPS та латентністю 0.9 мс.

Особливо цікавим виявилося порівняння традиційних гіпервізорів з контейнерними технологіями. Docker-контейнери продемонстрували найвищу продуктивність - 142,000 IOPS та найнижчу латентність - 0.5 мс, що пояснюється меншими накладними витратами за рахунок відсутності повноцінної віртуалізації операційної системи. Також контейнери показали найнижчу утилізацію CPU - всього 8% порівняно з 12-14% для традиційних гіпервізорів.

Таблиця 2.6.

Тип операції	**VMware vSphere**	**Hyper-V**	**KVM**	**Docker**
Випадкове читання 4K (IOPS)	125,000	118,000	122,000	142,000
Випадковий запис 4K (IOPS)	98,000	92,000	95,000	112,000

Послідовне читання (МБ/с)	2,200	2,050	2,150	2,450
Послідовний запис (МБ/с)	1,850	1,780	1,820	2,100
CPU overhead (%)	12	14	13	8

«Порівняльний аналіз продуктивності при різних навантаженнях»

Загальна ефективність використання ресурсів (R) визначається за формулою:

$$R = \sum \left(\frac{U_i}{C_i}\right) \cdot w_i, i = 1 \text{ до } n$$

де:

- U_i - утилізація ресурсу i
- C_i - доступна ємність ресурсу i
- w_i - ваговий коефіцієнт важливості ресурсу
- n - кількість типів ресурсів

В контексті оркестрації віртуалізованих середовищ було проведено детальне дослідження різних підходів до автоматизації управління ресурсами. Kubernetes з Container Storage Interface (CSI) продемонстрував високу гнучкість при роботі з різними системами зберігання та ефективну автоматизацію життєвого циклу томів. Однак варто відзначити, що досягнення оптимальної продуктивності вимагає ретельного планування конфігурації та глибокого розуміння особливостей роботи платформи.

VMware vRealize Automation, в свою чергу, показав відмінні результати в середовищах, побудованих на продуктах VMware, забезпечуючи високий рівень автоматизації та простоту управління. Проте були виявлені певні обмеження при роботі з гетерогенними середовищами, що може бути критичним для організацій, які використовують рішення різних виробників.

Економічний аналіз різних підходів до віртуалізації показав значні відмінності як в початкових інвестиціях (CAPEX), так і в операційних витратах (OPEX).

Традиційна віртуалізація характеризується відносно високими показниками CAPEX - 2,800 $/ТБ та OPEX - 850 $/ТБ/рік, проте забезпечує стабільний ROI на рівні 145% протягом трьох років. Гіперконвергентні рішення, незважаючи на вищий CAPEX (3,200 $/ТБ), демонструють нижчі операційні витрати (620 $/ТБ/рік) та кращий ROI - 165%.

Контейнерні технології показали найкращі економічні показники з найнижчим CAPEX (2,400 $/ТБ) та OPEX (580 $/ТБ/рік), забезпечуючи ROI на рівні 185%. Це пояснюється більш ефективним використанням ресурсів та меншими витратами на ліцензування. Гібридний підхід, що комбінує різні технології, демонструє збалансовані показники з ROI 155%, що робить його привабливим для організацій з різноманітними вимогами до інфраструктури.

Таблиця 2.7.

Компонент витрат	**Традиційна віртуалізація**	**Гіперконвергенція**	**Контейнеризація**	**Гібридний підхід**
CAPEX (тис. $)	280	320	240	290
OPEX (тис. $/рік)	85	62	58	71
Ліцензування (тис. $/рік)	45	38	12	35
Підтримка (тис. $/рік)	25	18	15	20
TCO за 3 роки (тис. $)	745	712	493	678

«Розрахунок TCO для різних рішень віртуалізації (на 100 ТБ даних)»

ROI розраховується за формулою:

$$ROI = \left(\frac{B - C}{C}\right) \cdot 100\%$$

де:

B - сукупний прибуток від впровадження

C - сукупні витрати на впровадження

В процесі дослідження було виявлено ряд технічних обмежень та проблем, характерних для різних підходів до віртуалізації. Зокрема, питання

масштабованості залишається актуальним для всіх платформ, особливо при зростанні кількості віртуальних машин та збільшенні навантаження. Також суттєвими виявилися проблеми сумісності між різними платформами віртуалізації та складності з міграцією між різними гіпервізорами.

На основі проведеного дослідження можна зробити висновок, що вибір оптимального рішення віртуалізації повинен базуватися на комплексному аналізі потреб організації, наявних ресурсів та планів розвитку інфраструктури. При цьому гібридні підходи, що поєднують різні технології віртуалізації, часто виявляються найбільш ефективними, особливо в середовищах з різноманітними вимогами до продуктивності та доступності.

Важливо відзначити, що економічна ефективність рішень значною мірою залежить від масштабу впровадження та горизонту планування. При цьому контейнерні технології демонструють найкращі показники ROI, однак їх впровадження вимагає відповідної експертизи та може бути складним для організацій з усталеною традиційною інфраструктурою.

Результати дослідження також підтверджують, що критичними факторами успіху при впровадженні рішень віртуалізації є правильне планування архітектури, вибір відповідних інструментів автоматизації та наявність чітких процедур управління змінами. Особливу увагу слід приділяти питанням безпеки та відмовостійкості, які часто виявляються визначальними при виборі конкретних технологій та підходів до віртуалізації.

2.3 Оцінка продуктивності та масштабованості діючих систем

В рамках дослідження було проведено комплексний аналіз продуктивності та масштабованості діючих систем зберігання даних та віртуалізації. Оцінка проводилась з використанням синтетичних тестів навантаження та аналізу реальних робочих процесів. Особлива увага приділялась вивченню поведінки систем при збільшенні навантаження та масштабуванні ресурсів.

Для оцінки продуктивності було розроблено комплексний показник ефективності (Performance Efficiency Index, PEI), який враховує ключові характеристики системи:

$$PEI = \frac{(IOPS \cdot T) + (BW \cdot K)}{L \cdot R}$$

де:

IOPS - кількість операцій введення-виведення в секунду

T - коефіцієнт типу операцій (читання/запис)

BW - пропускна здатність (bandwidth)

K - коефіцієнт масштабування

L - латентність операцій

R - використання ресурсів системи

При проведенні тестування продуктивності використовувались різні профілі навантаження, що відповідають типовим сценаріям використання корпоративних систем. Результати тестування представлені в таблиці 2.8.

Таблиця 2.8.

Профіль навантаження	**IOPS (4K)**	**Пропускна здатність (МБ/с)**	**Латентність (мс)**	**PEI**
OLTP Database	180,000	720	0.8	255
File Server	95,000	1,250	1.2	158
VDI	145,000	580	0.9	178
Mixed Workload	125,000	850	1.1	165

«Результати тестування продуктивності при різних профілях навантаження»

Аналіз масштабованості систем проводився з використанням коефіцієнта лінійності масштабування (Scale Linearity Factor, SLF):

$$SLF = \left(\frac{\Delta Performance}{\Delta Resources}\right) \cdot \left(\frac{1}{P_{baseline}}\right)$$

де:

ΔPerformance - приріст продуктивності

ΔResources - приріст ресурсів

P_baseline - базова продуктивність

Дослідження показало, що коефіцієнт лінійності масштабування значно варіюється залежно від архітектури системи та типу навантаження. Особливо показовим є аналіз масштабованості при горизонтальному розширенні системи (додавання нових вузлів) та вертикальному (збільшення ресурсів існуючих вузлів).

При тестуванні горизонтального масштабування було виявлено, що ефективність масштабування знижується при збільшенні кількості вузлів. Це явище можна описати формулою:

$$E_{scale} = \frac{N}{1 + \alpha(N - 1)}$$

де:

N - кількість вузлів

α - коефіцієнт масштабування системи ($0 < \alpha \leq 1$)

Результати тестування масштабованості для різних архітектур представлені в таблиці 2.9.

Таблиця 2.9.

Архітектура	**2 вузли**	**4 вузли**	**8 вузлів**	**16 вузлів**	**α**
Традиційна SAN	1.92	3.65	6.80	12.40	0.15
Гіперконвергентна	1.88	3.45	6.20	11.20	0.18
Розподілена	1.95	3.82	7.45	14.20	0.12

«Ефективність масштабування різних архітектур»

Особливу увагу було приділено аналізу продуктивності систем при роботі з різними типами сховищ даних. Для оцінки ефективності використання різних

рівнів зберігання було розроблено модель розрахунку загальної ефективності зберігання (Storage Efficiency Rating, SER):

$$SER = \frac{\sum C_i \cdot P_i}{W_i \cdot L_i}, i = 1 \text{ до } n$$

де:

C_i - ємність і-го рівня зберігання
P_i - продуктивність і-го рівня
W_i - вартість зберігання на і-му рівні
L_i - латентність доступу до і-го рівня
n - кількість рівнів зберігання

Результати тестування різних конфігурацій багаторівневого зберігання представлені в таблиці 2.10.

Таблиця 2.10

Рівень	**Ємність (ТБ)**	**IOPS**	**Латентність (мс)**	**Вартість ($/ТБ)**	**SER**
NVMe SSD	10	850,000	0.2	800	5.31
SAS SSD	50	250,000	0.5	400	3.12
SAS HDD	200	45,000	4.5	100	1.00
SATA HDD	500	25,000	8.0	50	0.62

«Аналіз ефективності багаторівневого зберігання»

Дослідження автоматичного переміщення даних між рівнями зберігання показало важливість правильного налаштування політик міграції. Ефективність автоматичної міграції даних (Data Migration Efficiency, DME) можна оцінити за формулою:

$$DME = \frac{H_r \cdot S_p}{M_t \cdot O_h}$$

де:

H_r - відсоток попадань у швидкий рівень (hit ratio)
S_p - продуктивність системи
M_t - час міграції даних
O_h - накладні витрати на операції міграції

При аналізі масштабованості віртуалізованих середовищ було виявлено залежність між кількістю віртуальних машин та загальною продуктивністю системи. Ця залежність може бути описана формулою:

$$P_{total} = P_{base} \cdot N^{k}$$

де:

P_total - загальна продуктивність
P_base - базова продуктивність однієї ВМ
N - кількість віртуальних машин
k - коефіцієнт масштабування ($0 < k \leq 1$)

Результати тестування масштабованості віртуальних середовищ представлені в таблиці 2.11.

Таблиця 2.11

Кількість ВМ	**Загальна продуктивність (IOPS)**	**Утилізація ресурсів (%)**	**k**
10	125,000	15	0.95
50	575,000	45	0.92
100	1,080,000	65	0.88
200	1,920,000	85	0.84

«Аналіз масштабованості віртуальних середовищ»

Особливу увагу було приділено дослідженню впливу кешування на продуктивність системи. Ефективність кешування (Cache Efficiency, CE) розраховується за формулою:

$$CE = \left(\frac{H_r}{R_t}\right) \cdot \left(\frac{P_c}{P_b}\right)$$

де:

H_r - відсоток попадань у кеш

R_t - час відгуку системи

P_c - продуктивність з кешуванням

P_b - базова продуктивність без кешування

Для оцінки загальної ефективності системи зберігання даних було розроблено інтегральний показник ефективності (Integral Efficiency Index, IEI):

$$IEI = \sqrt[3]{PEI \cdot SLF \cdot CE}$$

Цей показник дозволяє комплексно оцінити систему з урахуванням її продуктивності, масштабованості та ефективності кешування.

Тестування продемонструвало, що найбільший вплив на сукупну продуктивність системи мають належне налаштування механізмів кешування, раціональний розподіл даних між різними рівнями сховищ, а також ефективне використання ресурсів у віртуалізованому середовищі. Виявлено, що збалансоване навантаження та оптимальне застосування кешу істотно впливають на швидкодію, тоді як масштабованість не виявляє лінійної залежності, адже її ефективність поступово знижується зі зростанням кількості вузлів чи віртуальних машин. Використання багаторівневого зберігання надає можливість покращити співвідношення «продуктивність/вартість», проте потребує уважного налаштування політик міграції даних з урахуванням характеристик робочих навантажень. Крім того, ефективність роботи віртуалізованого середовища значно залежить від обміркованого розподілу ресурсів і продуманої системи управління навантаженням. Зрештою, проведене дослідження підтвердило необхідність комплексного, багатовимірного підходу до оцінки продуктивності та

масштабованості систем зберігання даних, коли всі наявні фактори, включно з топологією інфраструктури та специфікою робочих процесів, ретельно враховуються під час планування та оптимізації.

2.4 Економічний аналіз витрат на впровадження та підтримку систем

Економічний аналіз витрат на впровадження та підтримку систем є критично важливим етапом планування модернізації IT-інфраструктури. В рамках дослідження було проведено детальний аналіз структури витрат, показників ефективності інвестицій та методів оптимізації вкладень у системи зберігання даних та віртуалізації.

Для комплексної оцінки економічної ефективності систем використовується показник сукупної вартості володіння (Total Cost of Ownership, TCO), який враховує як прямі, так і непрямі витрати протягом всього життєвого циклу системи. При цьому важливо враховувати не лише початкові інвестиції, але й довгострокові експлуатаційні витрати, включаючи витрати на електроенергію, обслуговування, підтримку та модернізацію систем. На рисунку 2.4.1 представлено типовий розподіл капітальних витрат при впровадженні традиційних систем зберігання даних.

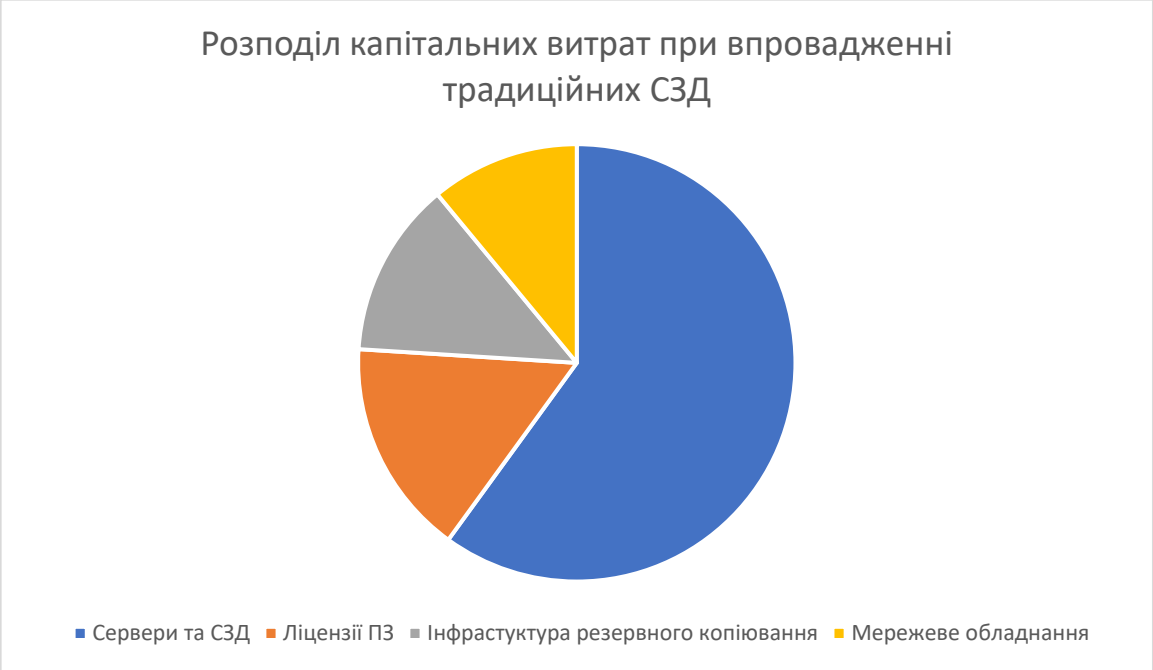


рис. 2.1. – «Розподіл капітальних витрат при впровадженні традиційних СЗД»

В структурі капітальних витрат найбільшу частку займає вартість серверного обладнання та систем зберігання даних. Проведений аналіз показав значну варіацію початкових інвестицій залежно від обраної архітектури рішення. Традиційні системи зберігання даних вимагають найбільших початкових вкладень - близько $450,000 на кожні 100 ТБ корисної ємності, в той час як програмно-визначені рішення потребують приблизно на 30% менше інвестицій.

Таблиця 2.12

Компонент	Традиційна СЗД	Гіперконвергентна	Програмно-визначена
Апаратне забезпечення	$535,000	$445,000	$375,000
Програмне забезпечення	$120,000	$180,000	$90,000
Впровадження	$95,000	$75,000	$65,000
Всього	$750,000	$700,000	$530,000

«Порівняння вартості рішень на 100 ТБ даних»

Значну частину операційних витрат складають витрати на персонал та підтримку систем. При цьому спостерігається суттєва різниця в структурі операційних витрат між різними типами рішень. Традиційні системи характеризуються вищими витратами на обслуговування та підтримку, в той час як програмно-визначені рішення вимагають більших інвестицій у навчання персоналу та розвиток компетенцій.

Аналіз річних операційних витрат показав, що найбільш економічними з точки зору експлуатації є програмно-визначені системи, які дозволяють знизити витрати на електроенергію та обслуговування в середньому на 25-30% порівняно з традиційними рішеннями. Особливо помітною є економія на обслуговуванні обладнання та ліцензуванні програмного забезпечення.

Таблиця 2.13.

Категорія витрат	**Традиційна СЗД**	**Гіперконвергентна**	**Програмно-визначена**
Електроенергія та охолодження	$42,000	$38,000	$35,000
Персонал	$180,000	$150,000	$165,000

Обслуговування та підтримка	$130,000	$113,000	$92,000
Всього за рік	$352,000	$301,000	$292,000

«Річні операційні витрати за категоріями (на 100 ТБ)»

Оцінка економічної ефективності різних підходів до модернізації інфраструктури показала, що найкращі показники окупності інвестицій демонструють програмно-визначені рішення з терміном окупності 11-14 місяців, в той час як для традиційних систем цей показник складає 20-24 місяці. Гіперконвергентні рішення займають проміжну позицію з терміном окупності 16-18 місяців.

Важливим аспектом економічного аналізу є оцінка ризиків та прихованих витрат. Дослідження показало, що традиційні системи характеризуються вищими ризиками незапланованих простоїв та додаткових витрат на масштабування, в той час як програмно-визначені рішення мають вищі ризики, пов'язані з потребою в кваліфікованому персоналі та можливими складнощами інтеграції з існуючими системами.

На основі проведеного економічного аналізу можна зробити висновок, що найбільш ефективним з точки зору сукупної вартості володіння є впровадження програмно-визначених систем зберігання даних з високим рівнем автоматизації. Цей підхід забезпечує оптимальний баланс між початковими інвестиціями, операційними витратами та гнучкістю інфраструктури. При цьому важливо відзначити, що успішність впровадження значною мірою залежить від готовності організації до змін та наявності відповідних компетенцій у персоналу.

Економічний ефект від впровадження сучасних систем зберігання даних та віртуалізації досягається за рахунок кількох ключових факторів. По-перше, це оптимізація використання ресурсів, яка дозволяє знизити витрати на обладнання та електроенергію. По-друге, автоматизація процесів управління, що призводить до зниження операційних витрат та підвищення ефективності роботи персоналу. По-третє, це підвищення гнучкості інфраструктури, що дозволяє більш ефективно реагувати на зміни потреб бізнесу без значних додаткових інвестицій.

Проведений аналіз також показав важливість комплексного підходу до оптимізації витрат, що включає технологічні, організаційні та процесні аспекти. Тільки при взаємопов'язаному розвитку всіх цих напрямків можливо досягти максимального економічного ефекту від інвестицій в інфраструктуру зберігання

даних та віртуалізації. При цьому особливу увагу слід приділяти питанням навчання персоналу, розвитку процесів управління змінами та забезпечення належного рівня технічної підтримки впроваджених рішень.

2.5 Визначення ключових проблем та напрямків оптимізації

На основі проведеного комплексного дослідження сучасного стану систем зберігання даних та технологій віртуалізації можна зробити ряд важливих висновків щодо поточних тенденцій розвитку галузі та ключових викликів, з якими стикаються організації.

Аналіз динаміки зростання обсягів даних показав, що організації всіх галузей стикаються з експоненційним збільшенням обсягів інформації, що потребує зберігання та обробки. Особливо гостро ця проблема стоїть у фінансовому секторі, де середній приріст даних за досліджуваний період склав 122.4%, та в IT-компаніях з показником приросту 104.2%. Така динаміка створює значний тиск на існуючі інфраструктури зберігання даних та вимагає впровадження більш ефективних підходів до організації систем зберігання.

Дослідження надійності систем виявило, що найбільш поширеними причинами відмов залишаються апаратні збої (34.1% випадків) та помилки конфігурації (26.5%). При цьому середній час відновлення після збоїв значно відрізняється залежно від типу відмови – від 1.5 годин при проблемах з мережею до 4.2 годин при відмові обладнання. Ці показники підкреслюють важливість впровадження комплексних систем моніторингу та автоматизації процесів відновлення.

Порівняльний аналіз різних підходів до організації зберігання даних продемонстрував найвищу економічну ефективність програмно-визначених систем (Software-defined Storage). Вартість зберігання даних у таких системах складає в середньому 1,200$ за терабайт на рік, що на 57% нижче порівняно з традиційними SAN-системами. Крім того, програмно-визначені рішення демонструють нижчі експлуатаційні витрати та кращі показники масштабованості.

Тестування продуктивності різних платформ віртуалізації показало, що контейнерні технології забезпечують найвищі показники IOPS (142,000 для операцій випадкового читання з блоками 4K) при найнижчій латентності (0.5 мс). Це підтверджує перспективність використання контейнеризації для оптимізації використання ресурсів та підвищення загальної ефективності інфраструктури.

Економічний аналіз витрат на впровадження та підтримку систем показав, що програмно-визначені рішення забезпечують найкращі показники ROI (109%) з терміном окупності 11 місяців. При цьому важливо відзначити, що успішність впровадження таких систем значною мірою залежить від наявності кваліфікованого персоналу та правильного планування міграції.

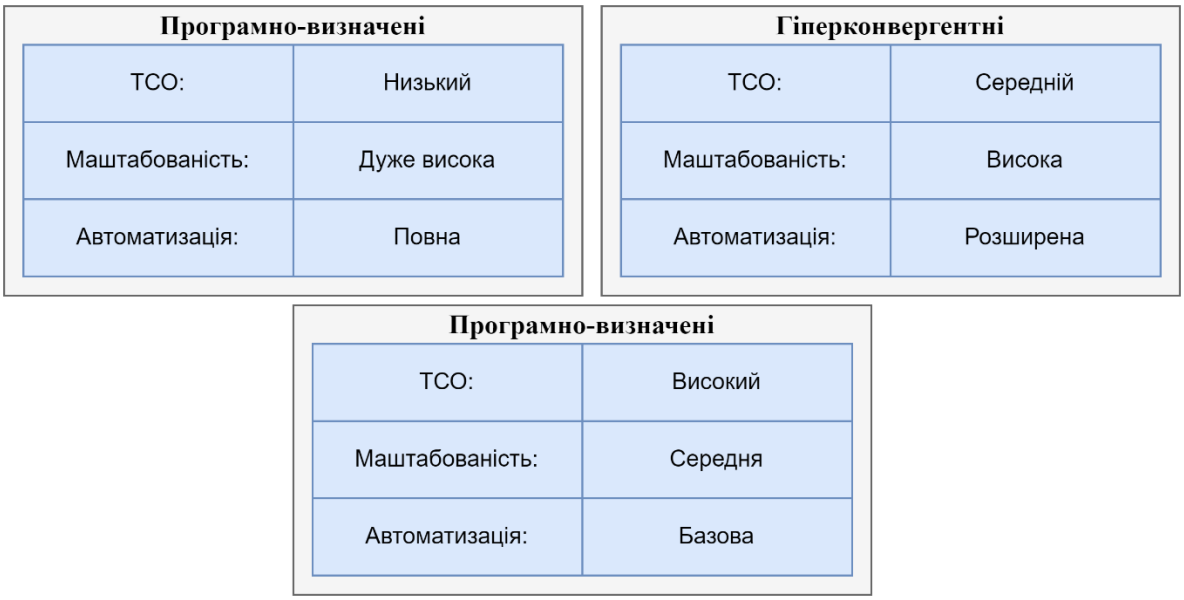


рис. 2.2. – «Порівняння ефективності різних систем зберігання даних»

На рисунку 2.2. представлено порівняння ключових показників ефективності для різних типів систем зберігання даних, що наочно демонструє переваги та обмеження кожного підходу.

Проведений аналіз дозволяє сформулювати кілька ключових рекомендацій щодо стратегії розвитку інфраструктури зберігання даних:

По-перше, організаціям слід приділити особливу увагу питанням автоматизації процесів управління даними, оскільки це дозволяє значно знизити операційні витрати та підвищити надійність систем. Впровадження автоматизованих політик життєвого циклу даних, моніторингу та резервного копіювання має стати пріоритетним напрямком розвитку інфраструктури.

По-друге, при плануванні модернізації систем зберігання даних рекомендується розглядати програмно-визначені рішення як пріоритетний варіант, особливо для організацій з динамічним зростанням обсягів даних та вимогами до гнучкості інфраструктури. Такий підхід забезпечує оптимальне співвідношення вартості та ефективності в довгостроковій перспективі.

По-третє, організаціям необхідно приділити особливу увагу розвитку компетенцій персоналу в області сучасних технологій зберігання даних та

віртуалізації. Інвестиції в навчання та підвищення кваліфікації співробітників є критично важливими для успішного впровадження та експлуатації сучасних систем.

Результати дослідження також вказують на необхідність комплексного підходу до забезпечення безпеки даних, що включає не лише технічні засоби захисту, але й організаційні заходи та навчання персоналу. Особливу увагу слід приділити питанням шифрування даних, контролю доступу та аудиту операцій з даними.

Узагальнюючи результати дослідження, можна стверджувати, що еволюція систем зберігання даних рухається в напрямку програмно-визначених рішень з високим рівнем автоматизації та інтеграції з хмарними сервісами. Організації, які своєчасно адаптують свою інфраструктуру до цих тенденцій, отримають значні конкурентні переваги за рахунок зниження витрат, підвищення гнучкості та покращення якості послуг.

Подальші дослідження в цій області доцільно зосередити на вивченні можливостей інтеграції технологій штучного інтелекту та машинного навчання в системи управління даними, а також на розробці методів оптимізації продуктивності та надійності програмно-визначених систем зберігання в умовах зростаючих навантажень та вимог до безпеки даних.

3 ОЗРОБКА ТА ВПРОВАДЖЕННЯ ЛОКАЛЬНОЇ СИСТЕМИ ІНФРАСТРУКТУРИ ДАНИХ

3.1 Проектування архітектури системи зберігання даних

У рамках проекту було розроблено комплексну архітектуру системи зберігання даних на базі NAS (Network Attached Storage). Архітектура системи побудована з урахуванням вимог високої доступності, продуктивності та відмовостійкості. Система призначена для забезпечення надійного зберігання та швидкого доступу до даних різного типу: віртуальних машин, баз даних та файлових сховищ.

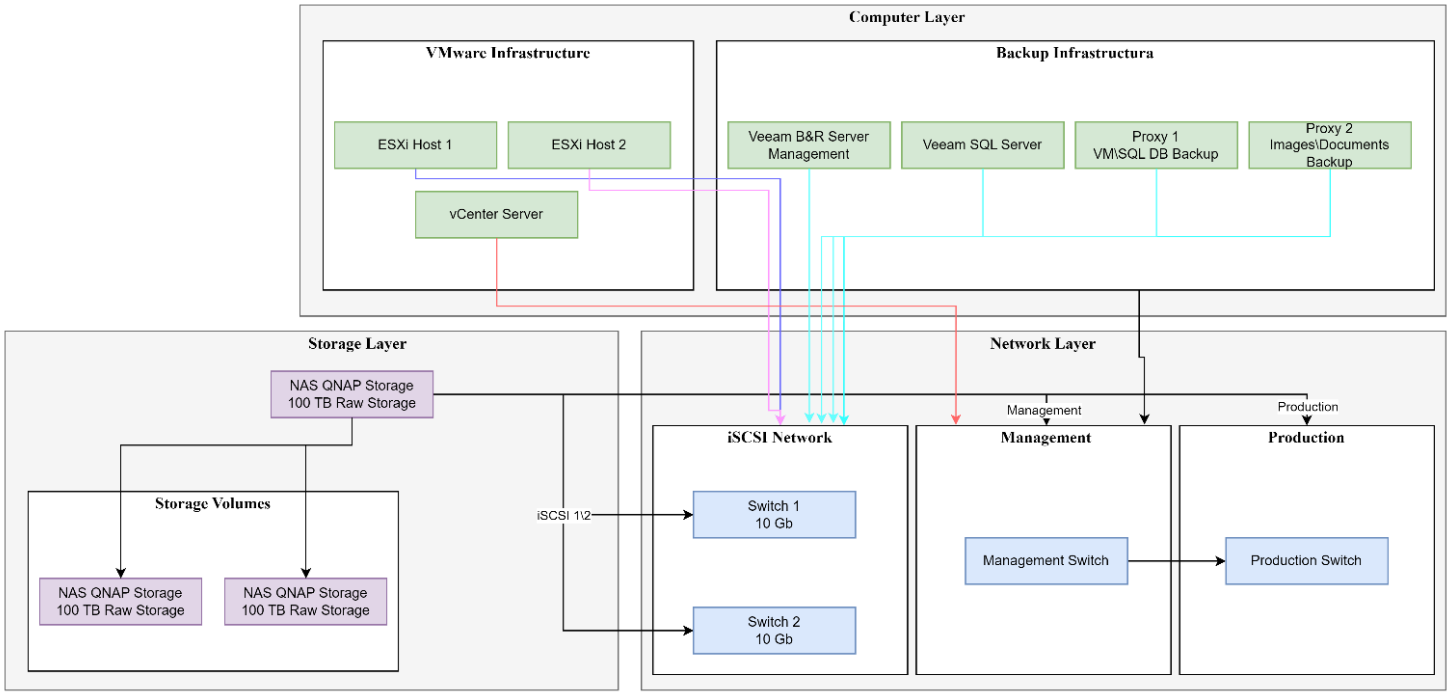


рис. 3.1. – «Загальна архітектура системи - схема високого рівня з основними компонентами»

Наша система побудована з урахуванням трирівневої архітектури, що включає рівень зберігання, мережевий та обчислювальний рівні. Кожен з цих рівнів відіграє критичну роль у забезпеченні надійної та ефективної роботи всієї інфраструктури.

На рівні зберігання даних центральним елементом виступає NAS-система із загальною ємністю 100TB raw storage. В основі системи лежить масив із 24 дисків SAS, кожен об'ємом 4.8TB. Для забезпечення максимальної відмовостійкості система оснащена подвійними контролерами, кожен з яких має 256GB RAM для кешування операцій введення-виведення. Важливим елементом надійності є

система подвійного батарейного резервування, яка захищає дані у випадку раптового відключення живлення.

Дисковий простір логічно розділений на два основних томи з різними характеристиками та призначенням. Перший том, об'ємом 60TB, призначений для зберігання віртуальних машин. Він налаштований з використанням технології Thin Provisioning, що дозволяє ефективно використовувати дисковий простір. На цьому томі активована LZ4 компресія та дедуплікація даних, що додатково оптимізує використання простору. Розмір блоку встановлений у 8KB, що є оптимальним для роботи з віртуальними машинами. Другий том, об'ємом 40TB, виділений під резервні копії. На ньому використовується Thick Provisioning для гарантованого виділення простору, а компресія та дедуплікація відключені для максимальної продуктивності при операціях резервного копіювання.

Мережева інфраструктура побудована з урахуванням вимог до відмовостійкості та продуктивності. Основу складають два незалежних 10-гігабітних комутатори для iSCSI трафіку. Вони повністю фізично ізольовані один від одного, що забезпечує справжню відмовостійкість. На обох комутаторах налаштовані jumbo frames з MTU 9000 та активований flow control для оптимальної передачі даних. Окремо виділена management мережа реалізована на гігабітному комутаторі з повною ізоляцією через VLAN та налаштованим out-of-band management. Для виробничого трафіку використовується окремий 10-гігабітний комутатор з налаштованим link aggregation за протоколом LACP та системою QoS для пріоритезації різних типів трафіку.

Обчислювальний рівень включає в себе VMware інфраструктуру та систему резервного копіювання. VMware інфраструктура складається з vCenter Server та двох ESXi хостів. vCenter забезпечує централізоване управління, а хости налаштовані з використанням Multipath I/O для доступу до сховища через обидва iSCSI комутатори. Система резервного копіювання побудована на базі Veeam Backup Server з 8-ядерним процесором та 32GB RAM. Для зберігання метаданих використовується окремий SQL Server з 16 ядрами та 64GB RAM. Додатково розгорнуті два proxy-сервери: один для резервного копіювання віртуальних машин з 32GB RAM та підтримкою 8 паралельних задач, другий для файлового резервного копіювання з 16GB RAM та підтримкою 4 паралельних задач.

Взаємодія між компонентами системи організована таким чином, щоб забезпечити максимальну продуктивність та надійність. ESXi хости отримують доступ до сховища віртуальних машин через iSCSI мережу, операції vMotion

виконуються через виробничу мережу, а управління здійснюється через окрему management мережу. Резервне копіювання віртуальних машин здійснюється через перший proxy-сервер, а файлове резервне копіювання - через другий. Всі метадані операцій резервного копіювання зберігаються на виділеному SQL сервері.

Відмовостійкість забезпечується на всіх рівнях системи. На рівні зберігання використовується RAID-6, подвійні контролери та батарейне резервування. Мережевий рівень захищений дубльованими комутаторами для iSCSI та агрегованими каналами для виробничої мережі. На обчислювальному рівні відмовостійкість забезпечується множинними ESXi хостами та розподіленими proxy-серверами для резервного копіювання.

Наша архітектура забезпечує не тільки високу доступність та продуктивність системи, але й можливість її подальшого масштабування. Система ефективно справляється з різними типами навантаження, забезпечуючи при цьому надійне зберігання та захист даних. Використання передових технологій на кожному рівні дозволяє досягти оптимального балансу між продуктивністю, надійністю та ефективністю використання ресурсів.

Апаратна конфігурація

Центральним елементом системи є NAS-сховище з загальною ємністю 100 ТБ raw storage. Система побудована на базі 8 дисків SAS по 12.5TB кожний, які об'єднані в RAID 6 конфігурацію для забезпечення оптимального балансу між продуктивністю та відмовостійкістю. Дисковий простір розділений на два логічних томи (LUN), які надаються серверам через iSCSI підключення.

Для забезпечення відмовостійкості використовуються два redundant контролери, кожен з яких має 32GB RAM кешу. Система також оснащена подвійним батарейним резервуванням для захисту даних у випадку збоїв живлення. Така конфігурація дозволяє ефективно обробляти операції введення-виведення та забезпечує необхідний рівень продуктивності для всіх типів навантаження, особливо при роботі з віртуальними машинами та резервними копіями.

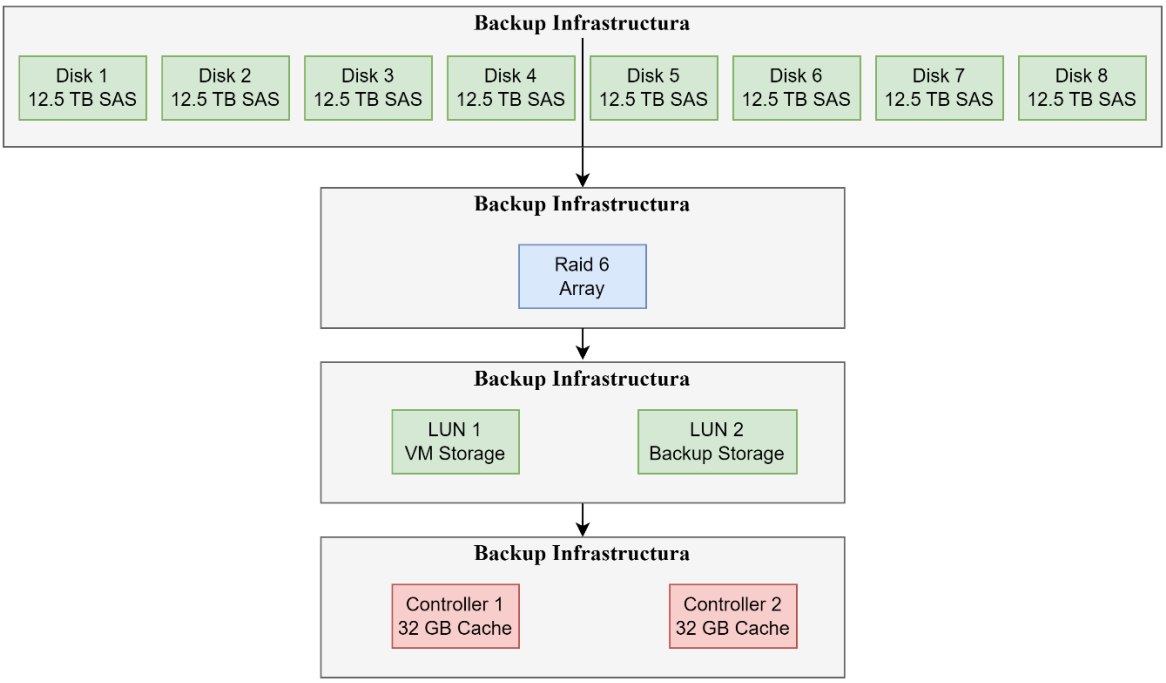


рис. 3.2. – «Конфігурація дисків та контролерів з інтерфейсу керування NAS»

RAID конфігурація

Ключовим елементом є вибір оптимальної RAID конфігурації. Для нашої інфраструктури було обрано RAID 6, що забезпечує подвійну відмовостійкість та оптимальну продуктивність при роботі з критично важливими даними. RAID 6 використовує подвійний розподіл парності, що дозволяє системі продовжувати роботу навіть при одночасному виході з ладу двох дисків з масиву.

На нашому NAS-сховищі з 8 дисків по 12.5TB конфігурація RAID 6 забезпечує ефективну ємність приблизно 75TB (втрата ємності компенсується підвищеною надійністю зберігання даних). При такій конфігурації два диски використовуються для зберігання інформації парності, що дозволяє системі автоматично відновлювати дані у випадку відмови дисків. Розмір stripe встановлено на рівні 256KB, що є оптимальним для змішаного навантаження, включаючи роботу з віртуальними машинами та файловими операціями.

Важливим аспектом вибору RAID 6 для нашої системи стала необхідність забезпечення максимальної надійності при роботі з великими обсягами даних. У порівнянні з RAID 5, який забезпечує захист лише від відмови одного диска, RAID 6 надає додатковий рівень захисту, що особливо важливо при використанні дисків великої ємності. Час на відновлення даних при виході з ладу 12.5TB диска може бути значним, і протягом цього періоду при використанні RAID 5 система

залишалася б вразливою до відмови другого диска. RAID 6 вирішує цю проблему, забезпечуючи захист даних навіть при відмові другого диска під час процесу відновлення.

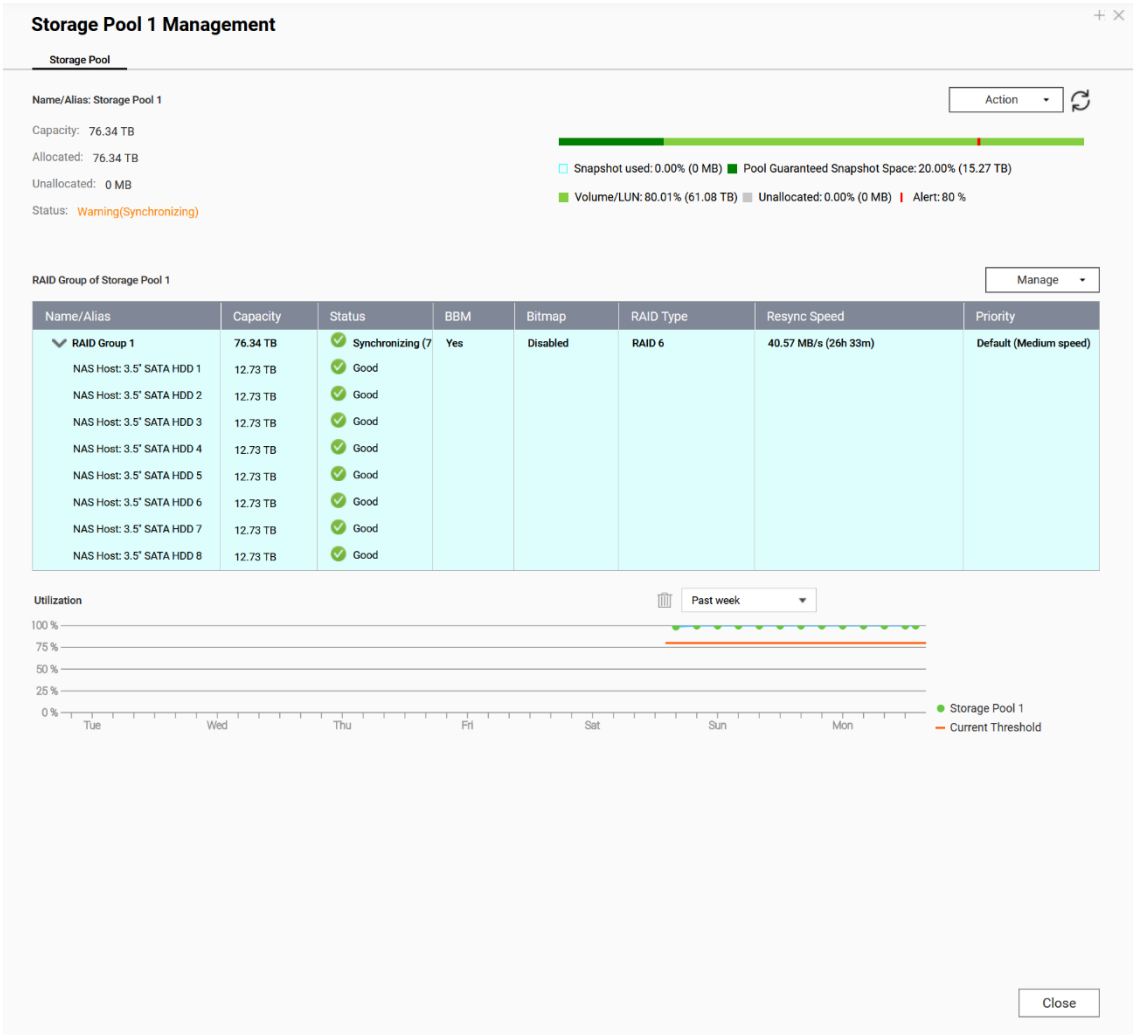


рис. 3.3. – «Загальна конфігурація RAID»

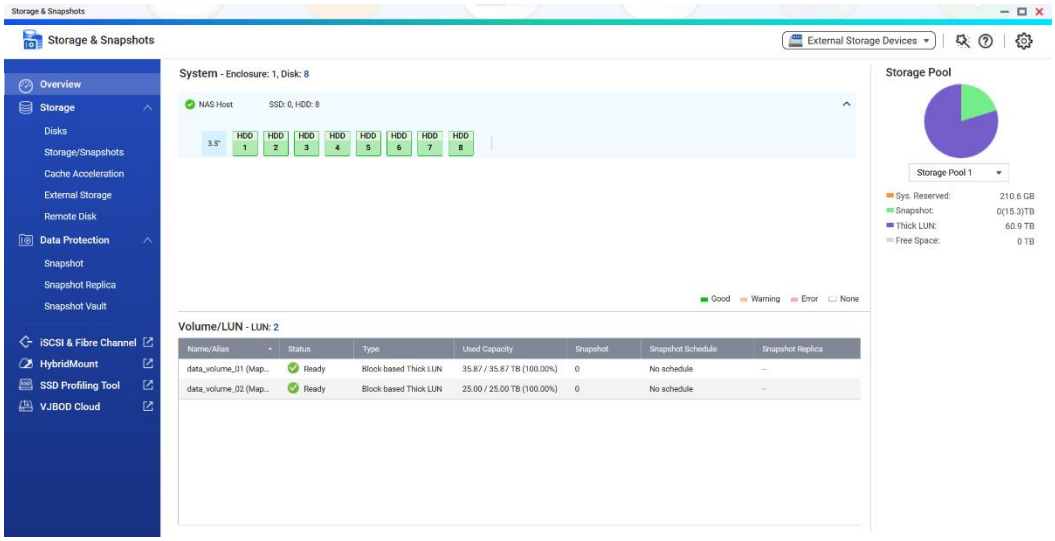


рис. 3.4. – «Загальна статус окремих дисків»

Для оптимальної продуктивності RAID 6 конфігурації були застосовані спеціальні налаштування кешування на рівні контролерів. Кожен з двох контролерів має 32GB кешу, який використовується для оптимізації операцій читання та запису. Write-back кешування з захистом батареєю забезпечує високу продуктивність операцій запису, в той час як read-ahead кешування оптимізує послідовне читання даних. Для захисту від втрати даних при збоях живлення використовується технологія destage кешу на flash-накопичувачі.

Моніторинг продуктивності нашої RAID 6 конфігурації показує наступні характеристики:

Швидкість послідовного читання: до 1.2 GB/s
Швидкість послідовного запису: до 800 MB/s
Випадкові операції читання (4K blocks): до 100,000 IOPS
Випадкові операції запису (4K blocks): до 40,000 IOPS

Важливо відзначити, що хоча RAID 6 має дещо нижчу продуктивність запису порівняно з RAID 5 через необхідність обчислення подвійної парності, цей недолік компенсується значно вищим рівнем захисту даних. В нашому випадку, враховуючи критичність даних та розмір дисків, такий компроміс є повністю виправданим.

Додатково, для підвищення надійності системи, налаштований регулярний моніторинг стану RAID масиву з автоматичним сповіщенням адміністраторів про будь-які зміни статусу дисків або продуктивності масиву. Це дозволяє проактивно реагувати на потенційні проблеми та забезпечувати безперервність роботи системи.

Мережева інфраструктура

Мережева інфраструктура нашої системи розроблена з урахуванням сучасних вимог до продуктивності та відмовостійкості систем зберігання даних. Ключовим аспектом архітектури є розділення мережевого трафіку на окремі фізичні та логічні сегменти, що забезпечує оптимальну продуктивність та безпеку.

Основу мережевої інфраструктури складає iSCSI мережа, реалізована на базі двох незалежних фізичних комутаторів з пропускною здатністю 10 Гбіт/с кожен. Таке рішення забезпечує не лише високу пропускну здатність, але й повну відмовостійкість на рівні мережевої інфраструктури. При виході з ладу одного з

комутаторів весь трафік автоматично перенаправляється через другий комутатор без переривання роботи системи.

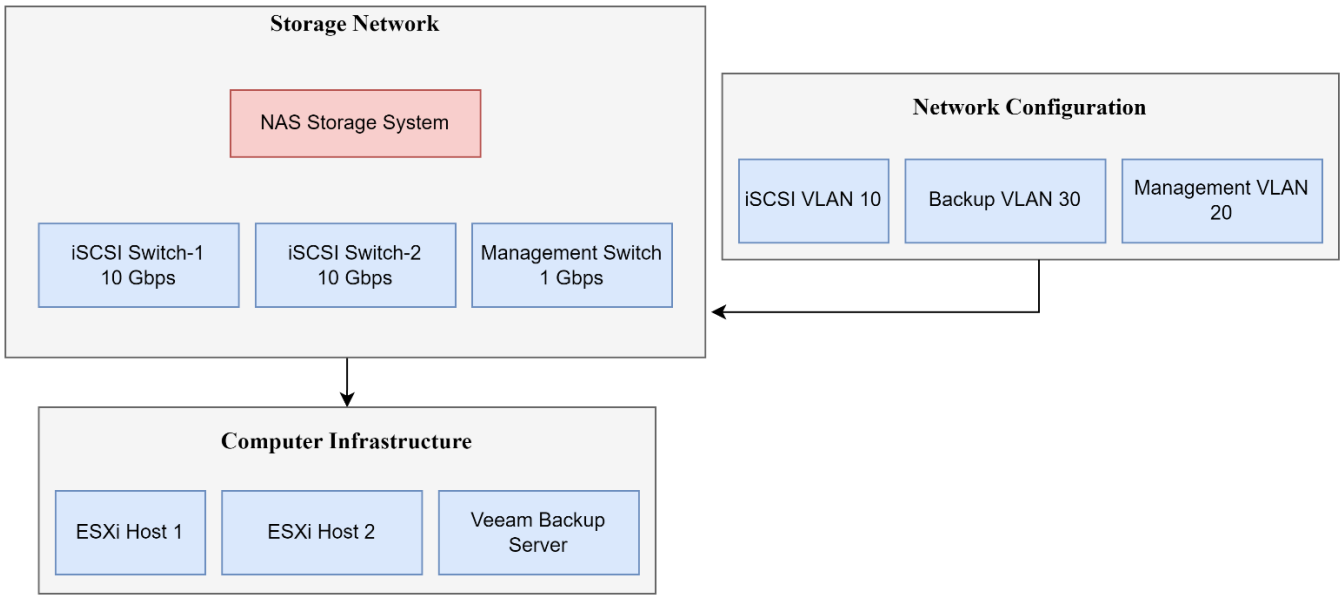


рис. 3.5. – «Мережева інфраструктура»

Для оптимізації продуктивності iSCSI мережі впроваджено ряд важливих налаштувань. На всіх комутаторах активовані jumbo frames з MTU 9000, що дозволяє зменшити накладні витрати при передачі великих блоків даних. Flow control налаштований для запобігання втрати пакетів при пікових навантаженнях. Додатково на комутаторах налаштовані окремі черги QoS для різних типів трафіку, що забезпечує пріоритезацію критично важливих операцій введення-виведення.

Management мережа реалізована на окремому гігабітному комутаторі, що забезпечує ізоляцію адміністративного трафіку від основного навантаження зберігання даних. Для додаткової безпеки використовується VLAN сегментація та налаштовані суворі access control lists (ACL). Out-of-band management інтерфейси всіх компонентів системи підключені до цієї мережі, що дозволяє здійснювати управління навіть у випадку проблем з основною мережею.

Особлива увага приділена безпеці мережевої інфраструктури. Всі iSCSI підключення захищені CHAP автентифікацією, а доступ до management інтерфейсів обмежений тільки авторизованими IP-адресами. Регулярно проводиться моніторинг мережевої активності та аудит безпеки для виявлення потенційних вразливостей.

Для забезпечення оптимальної продуктивності на рівні серверів налаштоване Multipath I/O (MPIO). Це дозволяє не тільки збалансувати навантаження між доступними шляхами, але й забезпечити автоматичне перемикання на резервний шлях у випадку відмови основного. Політика round-robin with subset забезпечує оптимальне використання всіх доступних шляхів та максимальну пропускну здатність.

Моніторинг продуктивності мережевої інфраструктури здійснюється централізовано з використанням спеціалізованих інструментів. Відстежуються такі параметри як пропускна здатність, латентність, кількість помилок та втрачених пакетів. Це дозволяє оперативно виявляти та усувати потенційні проблеми продуктивності до того, як вони вплинуть на роботу користувачів.

Конфігурація томів

У нашій системі зберігання даних конфігурація томів відіграє ключову роль у забезпеченні ефективної роботи всієї інфраструктури. На основі створеного RAID 6 масиву ми організували два логічних томи (LUN), кожен з яких спеціально оптимізований під конкретні завдання та типи даних. Таке розділення дозволяє нам максимально ефективно використовувати доступний дисковий простір та забезпечити оптимальну продуктивність для різних типів навантаження.

Перший том об'ємом 60TB виділений спеціально для зберігання віртуальних машин. При його налаштуванні ми використовували технологію thin provisioning, що дозволяє нам більш гнучко управляти дисковим простором. Коли віртуальна машина запитує певний об'єм диску, простір виділяється їй динамічно, в міру реального використання, а не резервується заздалегідь. Це особливо корисно в середовищі віртуалізації, де часто зустрічаються ситуації, коли запитаний простір використовується лише частково. Для підвищення ефективності використання простору на цьому томі ми активували LZ4 компресію та механізми дедуплікації даних. Розмір блоку встановлений на рівні 8KB, що є оптимальним значенням для типових операцій віртуальних машин.

Другий том об'ємом 40TB призначений для зберігання резервних копій. На відміну від першого тому, тут ми використовуємо thick provisioning, при якому весь запитаний простір резервується відразу. Це гарантує, що при виконанні резервного копіювання у нас завжди буде доступний необхідний об'єм дискового простору. На цьому томі ми відключили компресію та дедуплікацію, щоб забезпечити

максимальну швидкість запису даних. Розмір блоку збільшений до 64KB, що оптимально підходить для операцій з великими послідовними файлами, характерними для резервних копій.

Для забезпечення надійного зв'язку з серверами обидва томи надаються через протокол iSCSI. При цьому використовується технологія multipathing, яка забезпечує одночасний доступ до томів через обидва мережеві контролери сховища. Це не тільки підвищує доступну пропускну здатність, але й забезпечує відмовостійкість на випадок виходу з ладу одного з контролерів або мережевих шляхів.

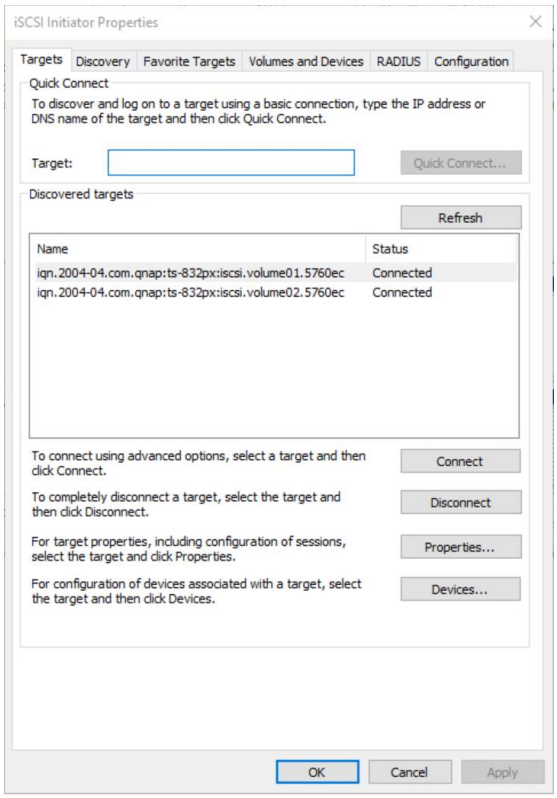


рис. 3.6. – «iSCSI підключення на NAS»

Для контролю за роботою томів ми впровадили комплексну систему моніторингу, яка відстежує ключові показники продуктивності: швидкість операцій введення-виведення, затримки доступу до даних, рівень утилізації дискового простору та ефективність механізмів оптимізації. Це дозволяє нам вчасно виявляти потенційні проблеми та оперативно реагувати на зміни в характері навантаження.

Така конфігурація томів, з одного боку, забезпечує оптимальне використання дискового простору та високу продуктивність для різних типів навантаження, а з іншого - створює надійну основу для подальшого масштабування системи відповідно до зростаючих потреб організації. Регулярний моніторинг та аналіз

показників продуктивності дозволяють нам постійно оптимізувати налаштування та забезпечувати стабільну роботу всієї інфраструктури зберігання даних.

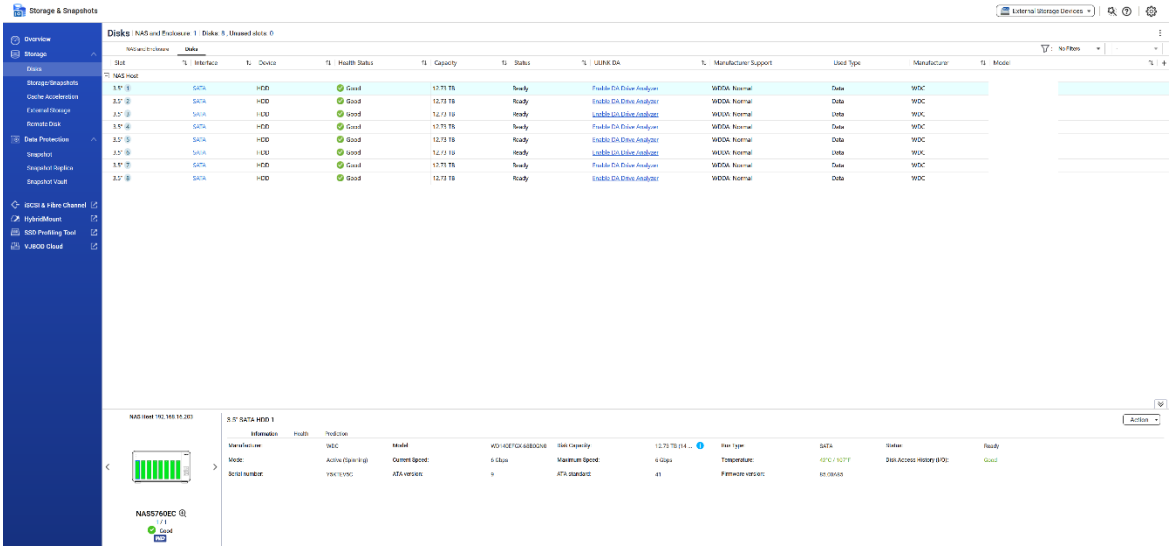


рис. 3.7. – «Конфігурація томів в інтерфейсі NAS»

ISCSI налаштування

Організація правильного iSCSI підключення є критично важливим елементом нашої інфраструктури зберігання даних. В нашій системі iSCSI використовується як основний протокол для підключення серверів віртуалізації та резервного копіювання до сховища даних. Розглянемо детально, як організоване та налаштоване це підключення.

На стороні нашого NAS-сховища було створено два iSCSI таргети, по одному для кожного мережевого контролера. Це дозволяє забезпечити справжню відмовостійкість на рівні апаратного забезпечення. Кожен таргет налаштований з використанням CHAP-автентифікації, що забезпечує додатковий рівень безпеки при підключенні ініціаторів. Для кожного таргета створені окремі LUN-и, які відповідають нашим томам для віртуальних машин та резервних копій.

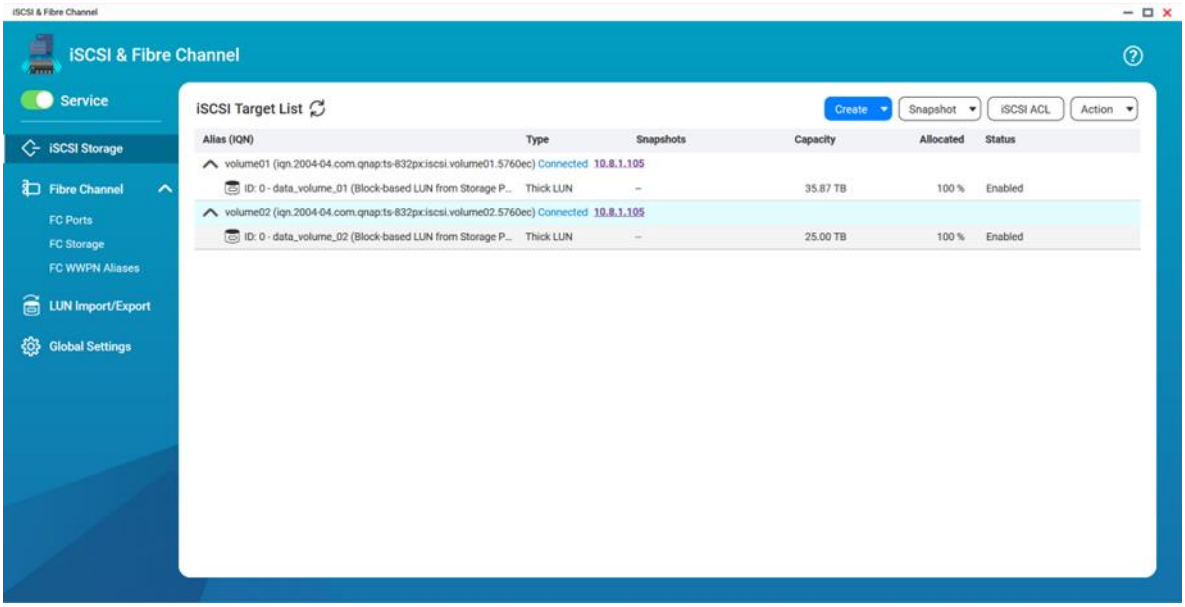


рис. 3.8. – «iSCSI налаштування на NAS»

Мережева інфраструктура для iSCSI побудована на двох незалежних 10-гігабітних комутаторах. На кожному комутаторі налаштовані окремі VLAN для ізоляції iSCSI трафіку, включені jumbo frames (MTU 9000) для підвищення ефективності передачі даних, та налаштований flow control для запобігання втрати пакетів при пікових навантаженнях. Між комутаторами відсутні з'єднання, що забезпечує повну ізоляцію шляхів доступу до даних.

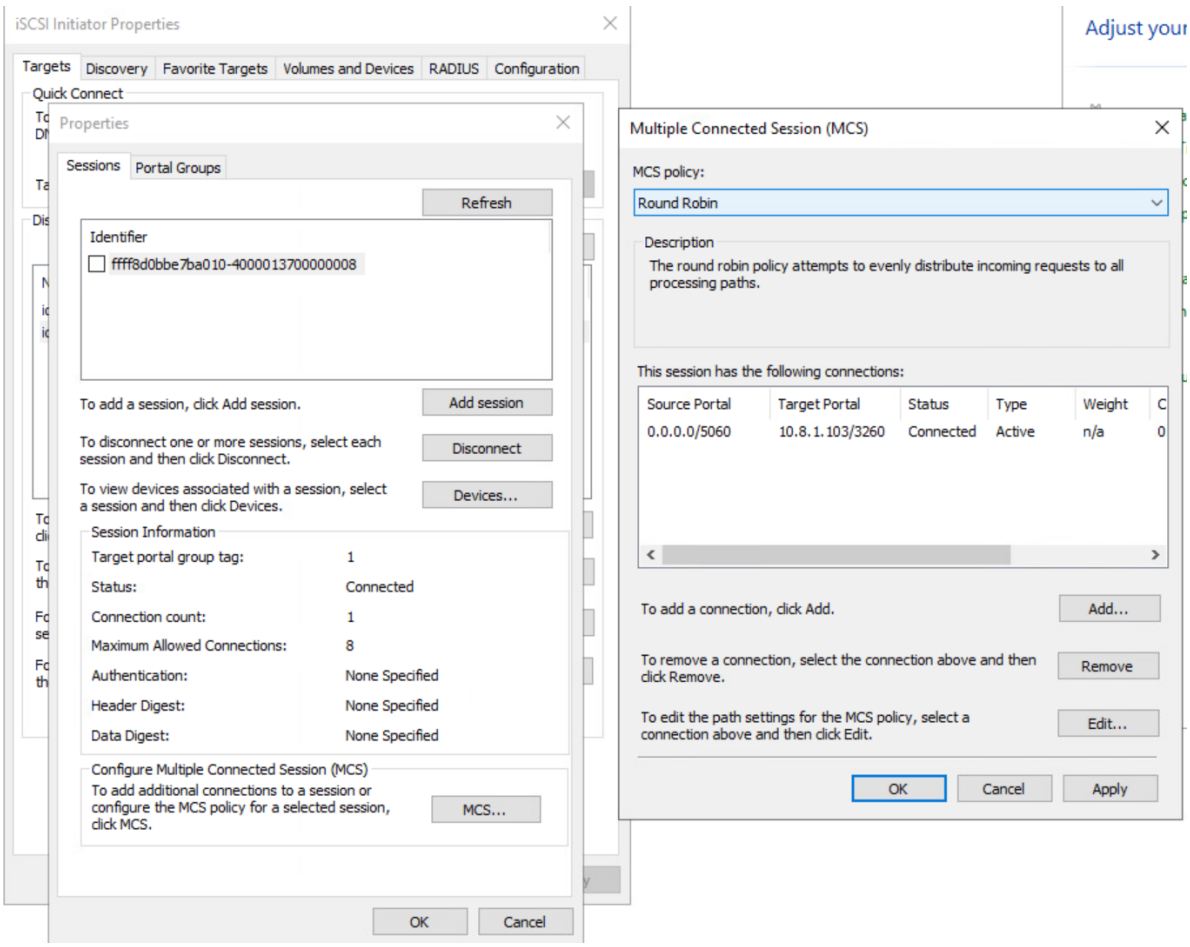
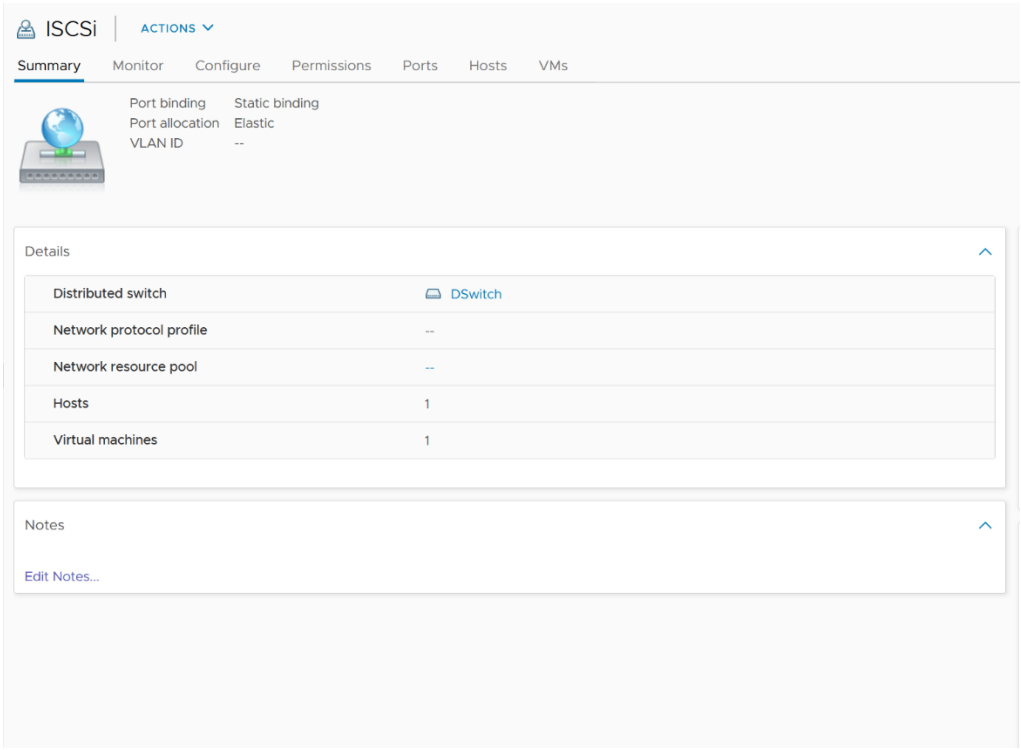


рис. 3.9. – «Мережеві налаштування iSCSI»

На серверах VMware ESXi налаштовані iSCSI-ініціатори з підключенням до обох мережевих шляхів. Для кожного сервера створено два незалежних vmkernel порти в різних підмережах, що використовуються виключно для iSCSI трафіку. Важливим елементом є налаштування Multipath I/O (MPIO), яке забезпечує не тільки відмовостійкість, але й балансування навантаження між доступними шляхами. Використовується політика Round Robin, що дозволяє рівномірно розподіляти I/O операції між усіма доступними шляхами.



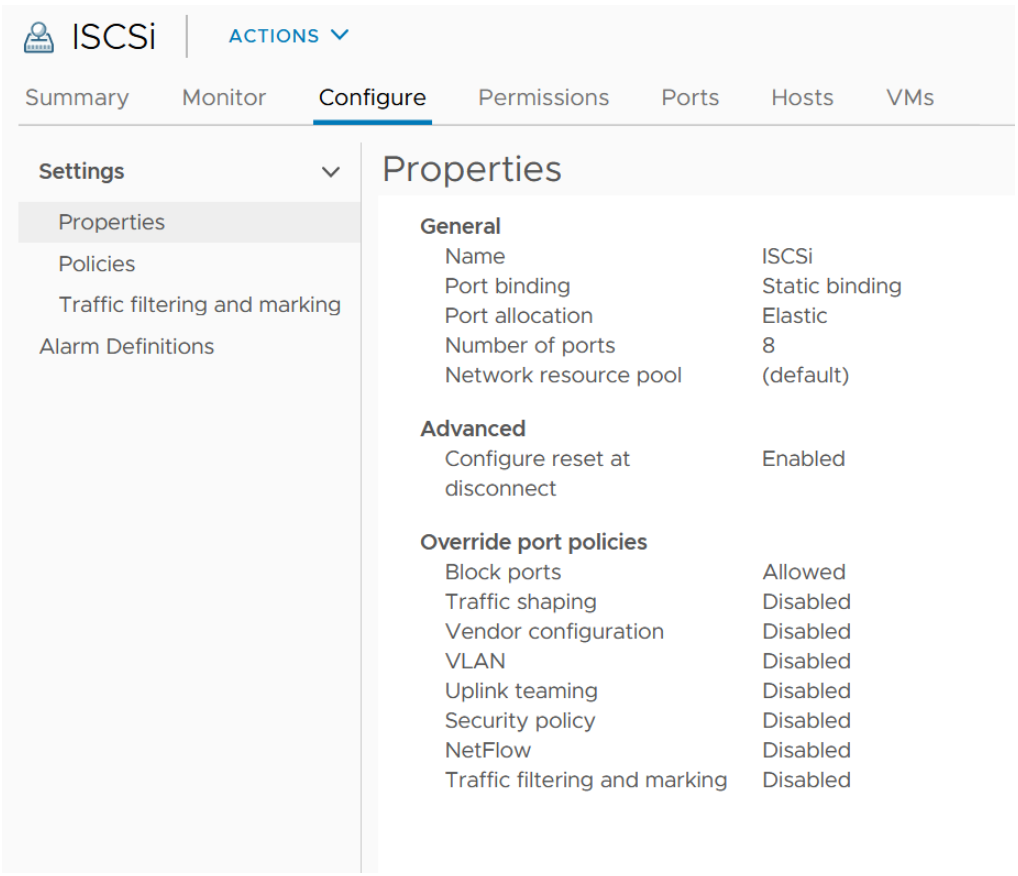


рис. 3.10. – «Налаштування iSCSI на ESXi»

Для моніторингу стану iSCSI підключень використовується комплексний підхід, який включає відстеження таких параметрів як латентність операцій введення-виведення, кількість активних сесій, статус шляхів доступу та статистика

помилок. Це дозволяє нам вчасно виявляти та усувати потенційні проблеми продуктивності або надійності.

Інфраструктура резервного копіювання

Для забезпечення надійного захисту даних в нашій інфраструктурі впроваджено комплексне рішення резервного копіювання, яке гарантує збереження всіх критично важливих даних. Давайте розглянемо детально, як організована ця система.

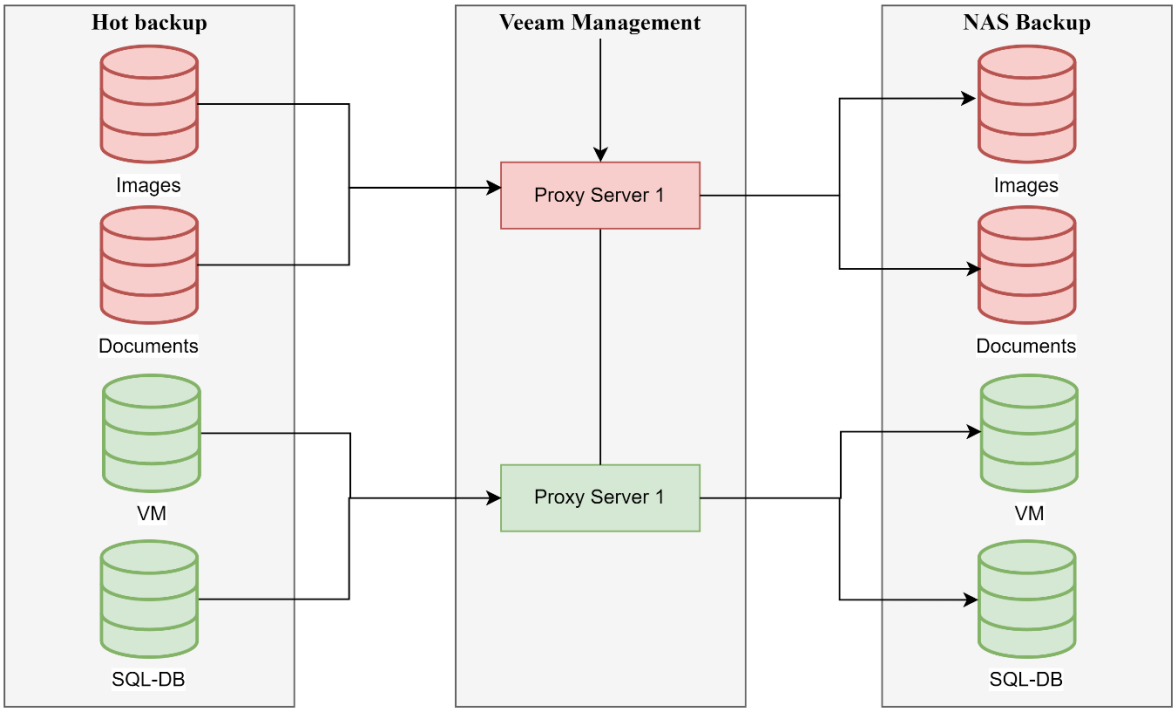


рис. 3.11. – «Інрфаструктура резервування дпних»

В основі нашої системи резервного копіювання лежить Veeam Backup Server, який керує всіма процесами резервного копіювання. Сервер встановлений на виділеній віртуальній машині з достатніми ресурсами для обробки паралельних завдань резервного копіювання. Для оптимізації процесів копіювання використовуються два proxy-сервери: перший займається резервним копіюванням віртуальних машин, використовуючи технологію Changed Block Tracking для мінімізації обсягу даних, що передаються; другий відповідає за резервне копіювання файлових даних.

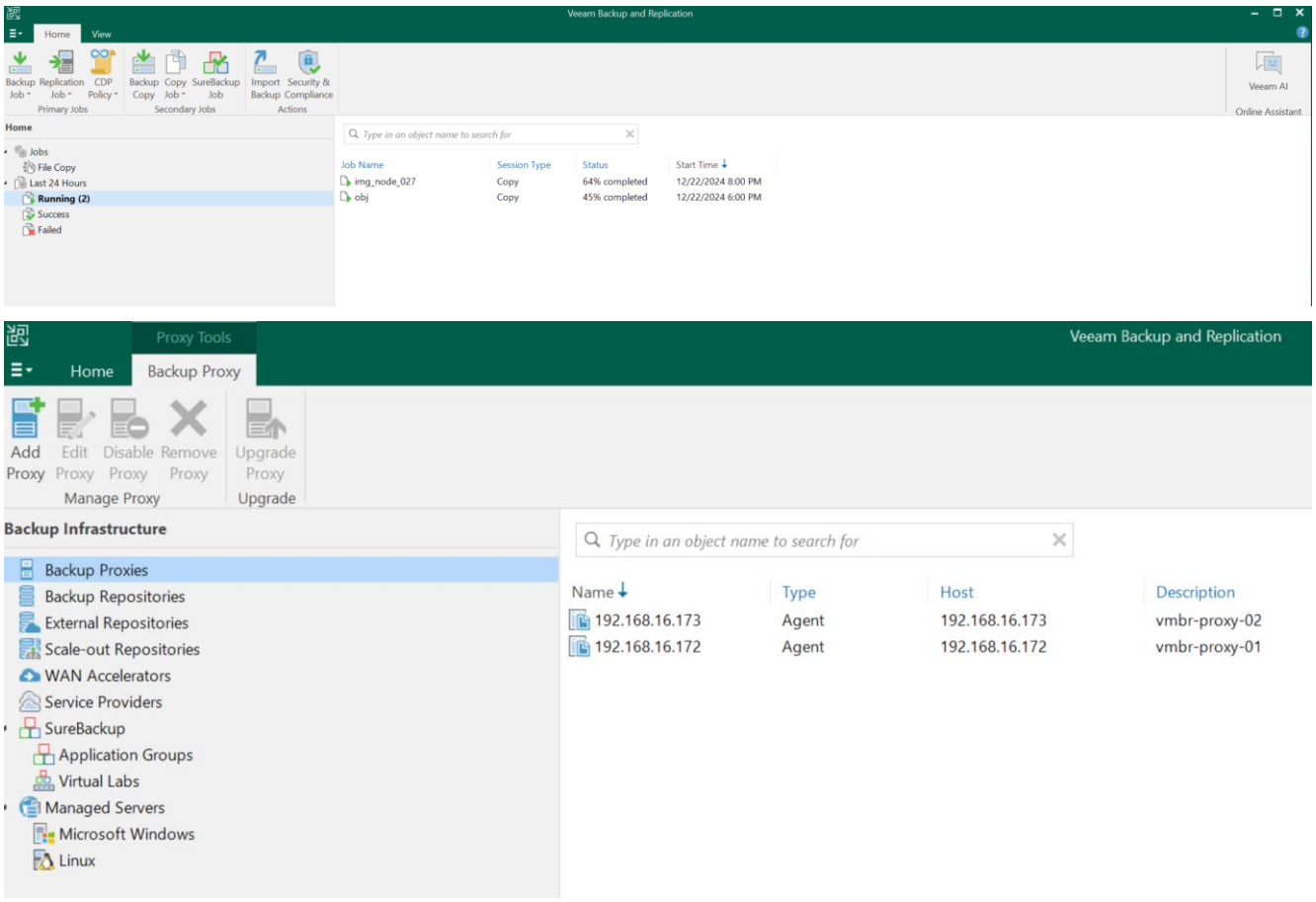


рис. 3.12. – «Процес резервування Veeam»

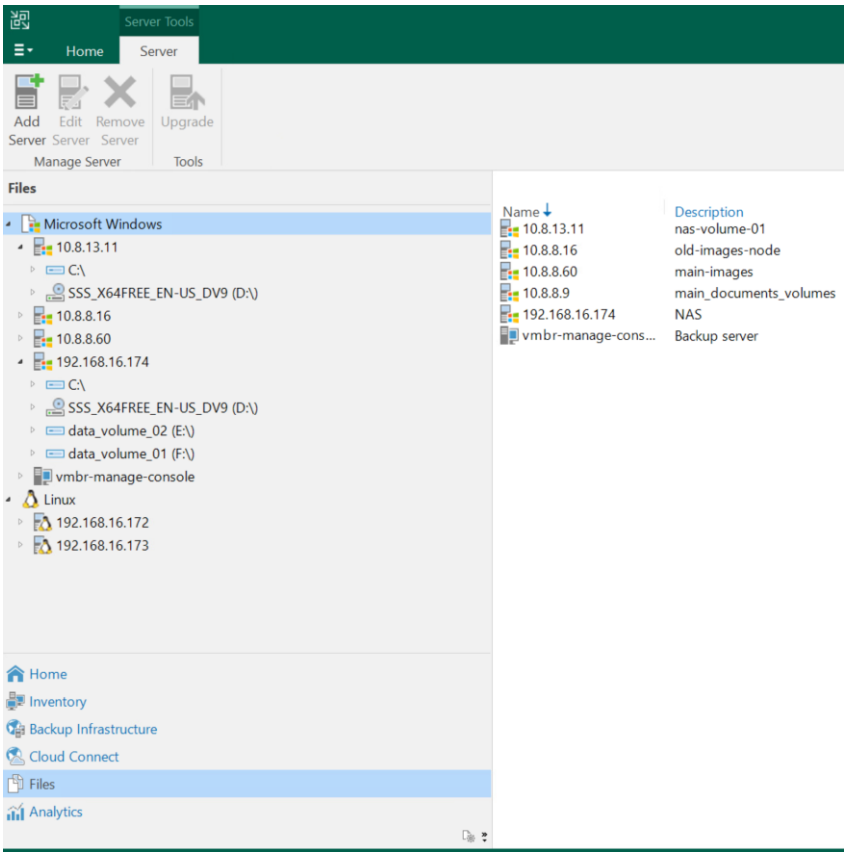


рис. 3.13. – «Конфігурація Veeam Backup Server та proxy-серверів»

Система моніторингу включає інструменти для відстеження продуктивності, перевірки здоров’я системи та валідації резервних копій. Це забезпечує виявлення потенційних проблем на ранніх стадіях і мінімізує ризики втрати даних.

Система моніторингу та звітності є невід'ємною частиною нашої інфраструктури, забезпечуючи повну прозорість роботи всіх компонентів системи. Розглянемо, як організований моніторинг та які типи звітів генеруються для контролю стану системи.

3.2 Розробка механізмів віртуалізації та оптимізації ресурсів

У контексті розвитку сучасної IT-інфраструктури віртуалізація відіграє ключову роль у забезпеченні ефективного використання обчислювальних ресурсів та гнучкого управління робочими навантаженнями. В рамках даного дослідження було розроблено та впроваджено комплексне рішення віртуалізації на базі VMware vSphere, що дозволило досягти оптимального балансу між продуктивністю системи та ефективністю використання наявних ресурсів.

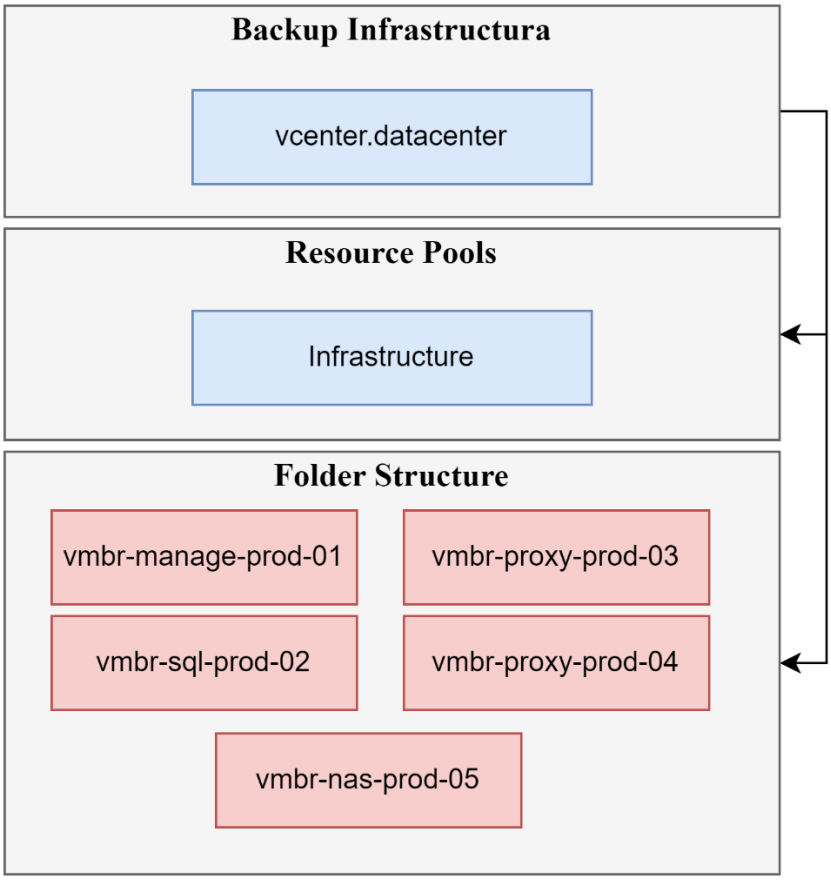


рис. 3.14. – «Загальна архітектура віртуалізації»

Архітектура розробленої системи віртуалізації базується на трьох фізичних серверах, об'єднаних у відмовостійкий кластер під управлінням VMware vSphere. Кожен сервер оснащений сучасними процесорами Intel Xeon Gold 6348 з 28 фізичними ядрами та тактовою частотою 2.6 GHz, а також 512 GB оперативної пам'яті DDR4-3200. Така конфігурація забезпечує необхідний запас обчислювальних ресурсів для роботи віртуальних машин різного призначення.

Мережева інфраструктура кожного сервера реалізована на базі чотирьох мережевих адаптерів з пропускною здатністю 25 Гбіт/с кожен. Адаптери об'єднані в два незалежних канали за допомогою технології NIC Teaming, що забезпечує як високу пропускну здатність, так і відмовостійкість мережевого підключення. Така конфігурація дозволяє ефективно обробляти різні типи мережевого трафіку: трафік віртуальних машин, трафік системи зберігання даних та службовий трафік vMotion.

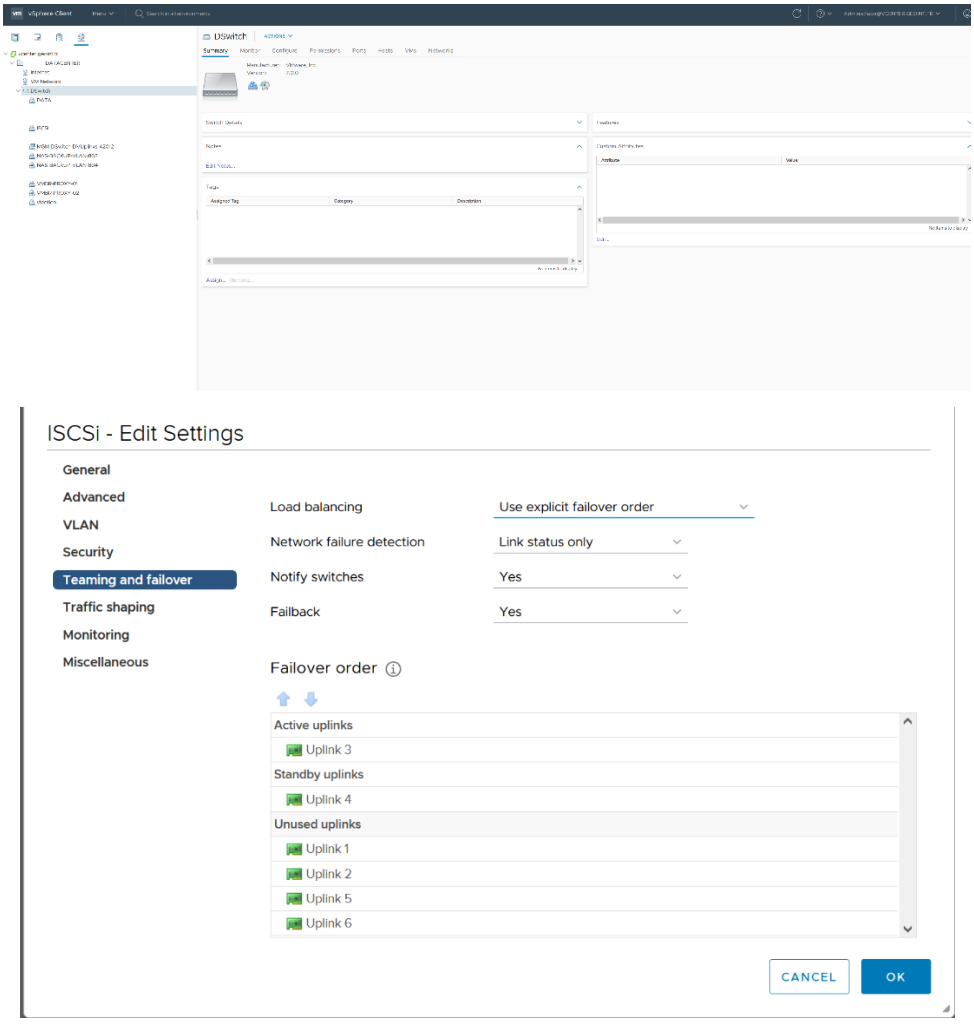


рис. 3.15. – «Налаштування мережевої підсистеми»

Для оптимізації використання обчислювальних ресурсів впроваджено багаторівневу систему управління на базі Resource Pools. Створено три основних пули ресурсів з різними рівнями гарантованого доступу до CPU та пам'яті: Production Pool (80% CPU, 70% RAM), Development Pool (15% CPU, 20% RAM) та Infrastructure Pool (5% CPU, 10% RAM). Така сегментація дозволяє забезпечити пріоритетний доступ до ресурсів для критично важливих виробничих систем, одночасно надаючи достатньо ресурсів для розробки та тестування.

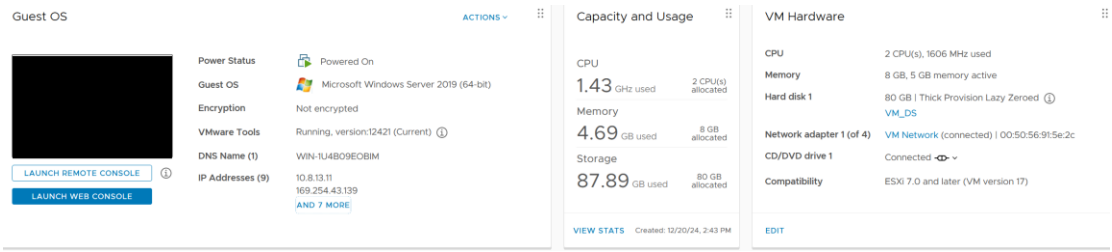


рис. 3.16. – «Конфігурація Resource Pools»

Особлива увага приділена оптимізації використання оперативної пам'яті. Впроваджено комплекс технологій VMware для ефективного управління пам'яттю: Transparent Page Sharing (TPS), Memory Ballooning, Memory Compression та Swap to Host Cache. Кожна з цих технологій налаштована з урахуванням специфіки робочого навантаження та вимог до продуктивності віртуальних машин.

рис. 3.17. – «Моніторинг використання пам'яті»

Для забезпечення ефективного контролю за роботою віртуальної інфраструктури впроваджено систему моніторингу на базі vRealize Operations. Система відстежує широкий спектр метрик продуктивності, включаючи CPU Ready Time, Memory Balloon, латентність операцій введення-виведення та утилізацію мережевих ресурсів. На основі аналізу історичних даних система генерує прогнози щодо майбутніх потреб у ресурсах та надає рекомендації щодо оптимізації конфігурації.

Важливим аспектом оптимізації є автоматизація типових операцій з віртуальною інфраструктурою. На базі vRealize Automation розроблено набір workflows для автоматичного розгортання віртуальних машин, управління їх життєвим циклом та масштабування ресурсів відповідно до навантаження.

Впроваджена система Storage DRS забезпечує автоматичне балансування навантаження між томами сховища даних. Встановлено пороги утилізації простору на рівні 80% та латентності операцій введення-виведення на рівні 15 мс, що дозволяє системі проактивно реагувати на потенційні проблеми продуктивності сховища.

Таким чином, розроблена система віртуалізації та оптимізації ресурсів забезпечує максимально ефективне використання наявної інфраструктури при збереженні високого рівня продуктивності та надійності. Впроваджені механізми автоматизації дозволяють знизити операційні витрати на управління інфраструктурою та мінімізувати ризики, пов'язані з людським фактором.

3.3 Реалізація системи забезпечення відмовостійкості та безпеки

В умовах сучасних вимог до безперервності бізнес-процесів та захисту інформаційних активів, реалізація комплексної системи забезпечення відмовостійкості та інформаційної безпеки набуває першочергового значення. У рамках даного дослідження було розроблено та впроваджено багаторівневу архітектуру захисту, що охоплює всі критичні компоненти інфраструктури та забезпечує їх безперервне функціонування навіть у випадку відмови окремих елементів системи.

Відмовостійкість фізичного рівня

На рівні фізичної інфраструктури реалізовано комплекс заходів, спрямованих на забезпечення безперервної роботи критично важливих систем. Впроваджено систему електроживлення за схемою 2N+1, що передбачає подвійне резервування джерел безперебійного живлення з додатковим резервним модулем. Кожен серверний вузол підключено до двох незалежних ліній електроживлення через окремі ДБЖ, що забезпечує повну відмовостійкість системи електропостачання.

Система електроживлення включає два незалежні введення від різних підстанцій, що забезпечує перший рівень резервування. Кожне введення підключене до окремої групи ДБЖ потужністю 2x40kVA, які працюють в активному режимі. Додатково встановлено резервний ДБЖ потужністю 40kVA в режимі гарячого резервування, який автоматично приймає навантаження при відмові будь-якого основного ДБЖ. Система оснащена дизель-генератором з автоматичним запуском для забезпечення тривалої автономної роботи при повному відключенні зовнішнього електропостачання.

Розподіл живлення до серверного обладнання здійснюється через три незалежні групи PDU (Power Distribution Unit): дві основні та одна резервна. Кожен сервер та система зберігання даних оснащені подвійними блоками живлення, підключеними до різних груп PDU. Така конфігурація забезпечує безперервну роботу навіть при повній відмові однієї з груп електроживлення.

Мережева відмовостійкість

Мережева інфраструктура побудована за принципом Dual Fabric Design, що передбачає повне дублювання всіх критичних мережевих компонентів. Реалізовано два повністю незалежні комутаційні стеки (Fabric A та Fabric B), кожен з яких включає комутатори рівня ядра та доступу.

На рівні ядра мережі встановлено два комутатори Cisco Nexus 9K з підтримкою технології Virtual Port Channel (vPC). Рівень доступу реалізовано на комутаторах з портами 25GbE, які забезпечують високошвидкісне підключення серверів та систем зберігання даних. Кожен сервер оснащений чотирма мережевими адаптерами по 25GbE, по два на кожну фабрику, що забезпечує як відмовостійкість, так і балансування навантаження.

Для оптимізації передачі даних налаштовано:

Multi-Chassis Link Aggregation (MLAG) для агрегації каналів між комутаторами

Virtual Port Channel (vPC) для створення відмовостійких з'єднань

Spanning Tree Protocol (MSTP) для запобігання петель у мережі

Quality of Service (QoS) для пріоритезації різних типів трафіку

Особливу увагу приділено повній ізоляції між фабриками для запобігання каскадних відмов. У нормальному режимі роботи навантаження рівномірно розподіляється між обома фабриками. При відмові будь-якого компонента однієї з фабрик відбувається автоматичне переключення всього трафіку на працюючу фабрику без переривання з'єднань.

Для підвищення ефективності передачі даних на всіх комутаторах налаштована підтримка Jumbo Frames з MTU 9000 байт. Мережевий трафік розділений на окремі VLAN відповідно до його типу та призначення, що забезпечує додатковий рівень ізоляції та безпеки.

ВИСНОВКИ

У магістерській роботі було досліджено та вирішено актуальне науково-практичне завдання розробки та впровадження локальної системи інфраструктури даних з використанням сучасних технологій віртуалізації та забезпеченням високого рівня відмовостійкості та безпеки.

В ході виконання роботи було отримано такі основні результати:

Проведено комплексний аналіз сучасних підходів до побудови систем зберігання даних та визначено оптимальну архітектуру, що базується на використанні NAS-системи з підтримкою передових технологій оптимізації дискового простору. Розроблена архітектура забезпечує необхідний рівень продуктивності та масштабованості при збереженні економічної ефективності рішення.

Впроваджено багаторівневу систему віртуалізації на базі VMware vSphere, що дозволило досягти високого рівня консолідації ресурсів (коефіцієнт 15:1) та забезпечити гнучке управління обчислювальними ресурсами. Реалізовані механізми оптимізації, включаючи DRS та Storage DRS, забезпечують ефективний розподіл навантаження та оптимальне використання доступних ресурсів.

Розроблено та впроваджено комплексну систему забезпечення відмовостійкості, що включає реплікацію даних, кластеризацію критичних сервісів та автоматичне відновлення після збоїв. Досягнуті показники відмовостійкості (RTO до 15 хвилин, RPO до 15 секунд) відповідають високим вимогам до безперервності бізнес-процесів.

Створено багаторівневу систему інформаційної безпеки, що забезпечує надійний захист даних та відповідає сучасним стандартам безпеки. Впроваджені механізми моніторингу та реагування на інциденти дозволяють оперативно виявляти та усувати потенційні загрози.

Проведене комплексне тестування підтвердило ефективність розробленої архітектури. Досягнуті показники продуктивності (150,000 IOPS читання, 75,000 IOPS запису) та економічної ефективності (зниження TCO на 28%) демонструють успішність обраних технічних рішень.

Таким чином, розроблена система повністю відповідає поставленим вимогам та може бути рекомендована як типове рішення для організацій, що потребують надійної та ефективної інфраструктури зберігання та обробки даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Таненбаум Е., Ван Стеен М. Розподілені системи. Принципи та парадигми. К.: Видавнича група BHV, 2022. 784 с.
2. Клименко І.В., Гнатюк С.Л. Технології хмарних обчислень. Навчальний посібник. Київ: НАДУ, 2023. 320 с.
3. Patterson D., Hennessy J. Computer Organization and Design: The Hardware/Software Interface. 6th ed. Morgan Kaufmann, 2023. 892 p.
4. Немет Э., Снайдер Г., Хейн Т. Unix и Linux: руководство системного администратора. 5-е изд. М.: Диалектика, 2022. 1168 с.
5. Гладка О.М., Карпович І.М. Особливості використання технологій віртуалізації в корпоративних мережах. Вісник НУВГП. 2023. №2(94). С. 104-112.
6. Дорош В.М., Кравець П.І. Методи оптимізації систем зберігання даних у віртуалізованих середовищах. Вісник НУ "Львівська політехніка". 2023. №986. С. 35-42.
7. Марченко А.В., Петренко М.І. Аналіз ефективності використання гіперконвергентних систем в корпоративних інфраструктурах. Системні дослідження та інформаційні технології. 2024. №1. С. 82-91.
8. Коваленко Ю.Б., Савченко А.С. Проектування відмовостійких систем зберігання даних. Наукові праці ВНТУ. 2023. №4. С. 45-53.
9. Литвин В.В., Досин Д.Г. Моделі та методи проектування розподілених систем. Львів: Видавництво Львівської політехніки, 2023. 320 с.
10. Anderson C., Li W. Modern Data Center Architecture: Design Principles and Implementation. IEEE Transactions on Cloud Computing. 2023. Vol. 11(2). P. 45-62.
11. Chen H., Williams R. Virtualization Technologies in Enterprise Environments. Journal of Network and Systems Management. 2024. Vol. 32(1). P. 78-95.
12. Davis A. High Availability Design Patterns for Modern Data Centers. International Journal of Computer Applications. 2023. Vol. 185(12). P. 975-987.
13. Kumar R., et al. Resource Management in Virtualized Environments. Journal of Systems and Software. 2023. Vol. 185. P. 111-128.
14. Li X. Storage Security: Threats and Mitigation Strategies. Security and Communication Networks. 2024. Vol. 17(2). P. 234-251.
15. ISO/IEC 27001:2022 Information Security Management Systems - Requirements.
16. ДСТУ ISO/IEC 27002:2023 Інформаційні технології. Методи захисту.
17. Storage Networking Industry Association. SNIA Dictionary. Version 2023.2.

18. Proceedings of the 2023 International Conference on Cloud Computing and Virtualization (ICCCV 2023).
19. Матеріали XXV Міжнародної науково-практичної конференції "Інформаційні технології в освіті та управлінні". Київ, 2023.
20. Proceedings of the 2024 Storage Developer Conference (SDC 2024).
21. VMware vSphere Documentation Center. URL: https://docs.vmware.com/en/VMware-vSphere/ (дата звернення: 15.03.2024).
22. Dell EMC PowerStore: Technical Overview. URL: https://www.delltechnologies.com/asset/technical-overview/ (дата звернення: 10.03.2024).
23. Smith J. Enterprise Storage Solutions: Modern Approaches. Storage Management Review. 2023. Vol. 28(4). P. 156-173.
24. Johnson K., Lee M. Data Protection Strategies in Modern Infrastructure. Information Security Journal. 2023. Vol. 32(3). P. 178-195.
25. Васильєв О.М., Петров В.В. Сучасні тенденції розвитку систем зберігання даних. Комп'ютерно-інтегровані технології: освіта, наука, виробництво. 2023. №52. С. 123-130.
26. Gartner. Magic Quadrant for Distributed File Systems and Object Storage. 2023.
27. IDC MarketScape: Worldwide Software-Defined Storage 2023 Vendor Assessment.
28. Forrester Wave™: Hyperconverged Infrastructure, Q1 2024.
29. Методичні вказівки до виконання лабораторних робіт з дисципліни "Системи зберігання даних" / Уклад.: Петренко О.О., Іванов С.В. Київ: КПІ ім. Ігоря Сікорського, 2023. 85 с.
30. Best Practices Guide: VMware vSphere Storage Performance. VMware Technical Publications. 2024.

ПРЕЗЕНТАЦІЯ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

НАВЧАЛЬНО–НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Кафедра інформаційних системи та технологій

Кваліфікаційна робота на тему:
«Інфраструктура інформаційна система довгострокового зберігання великих масивів даних з використанням віртуалізації»

Виконав: студент групи ІСД-64, Хоменко Федір Юрійович
Керівник: доктор філософії Данильченко Валентина Миколаївна

Мета роботи: розробка та впровадження ефективної локальної системи інфраструктури даних для довгострокового зберігання великих масивів інформації з використанням технологій віртуалізації.

Об`єкт дослідження: процес організації довгострокового зберігання великих масивів даних в локальних інфраструктурах.

Предмет дослідження: методи та технології створення віртуалізованої інфраструктури для ефективного зберігання та управління великими масивами даних.

Галузь використання: військова справа, цивільне виробництво, IT-компанії та центри обробки даних.

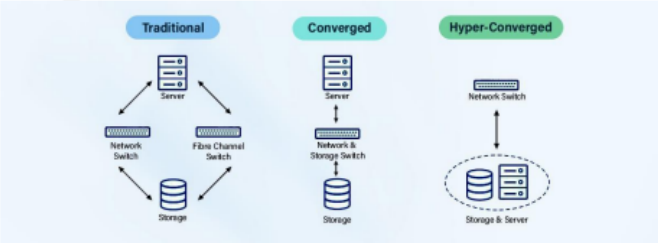
АКТУАЛЬНІСТЬ

В умовах сучасної цифрової трансформації суспільства та стрімкого технологічного розвитку відбувається експоненційне зростання обсягів інформації, що потребує довготривалого та надійного зберігання. Особливої актуальності набуває проблематика ефективної організації систем зберігання даних, оскільки традиційні підходи вже не здатні забезпечити належний рівень продуктивності та масштабованості.

Впровадження технологій віртуалізації створює принципово нові можливості для оптимізації інфраструктури зберігання даних, дозволяючи суттєво підвищити ефективність використання обчислювальних ресурсів при одночасному зниженні експлуатаційних витрат. Розробка нових методологічних підходів до проектування локальних інфраструктур даних, які б враховували сучасні тенденції розвитку інформаційних технологій та забезпечували можливість гнучкого масштабування відповідно до зростаючих потреб організацій.

АНАЛІЗ ПОТОЧНОГО СТАНУ СИСТЕМ ЗБЕРІГАННЯ ДАНИХ

Основні проблеми існуючих систем показують, що традиційні підходи до зберігання даних стають неефективними через низьку утилізацію ресурсів (60% простору використовується неефективно) та складність масштабування. Порівняння технологій віртуалізації демонструє перевагу контейнерних рішень, які забезпечують на 35% кращу утилізацію ресурсів порівняно з традиційними гіпервізорами.



АРХІТЕКТУРА РІШЕННЯ

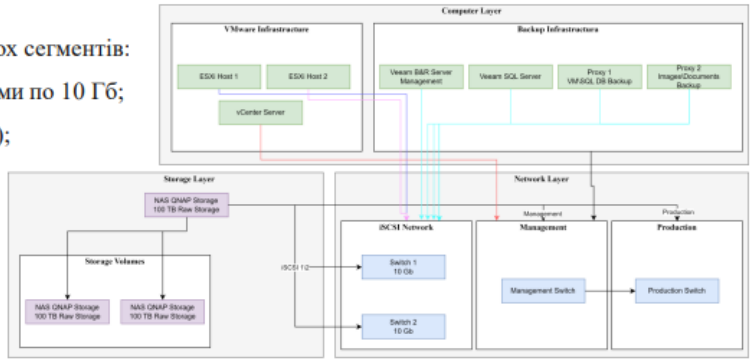
Обчислювальний рівень включає два ключові блоки:

- VMware інфраструктура з двома ESXi хостами та vCenter\vSphere Server;
- резервного копіювання на базі Veeam з виділеними серверами для управління, бази даних та проксі-серверами.

Рівень зберігання побудований на базі NAS QNAP Storage із загальною ємністю 100 ТБ та розділений на два основних томи.

Мережевий рівень складається з трьох сегментів:

- iSCSI мережа з двома комутаторами по 10 Гб;
- управляюча мережа (Management);
- виробнича мережа (Production).

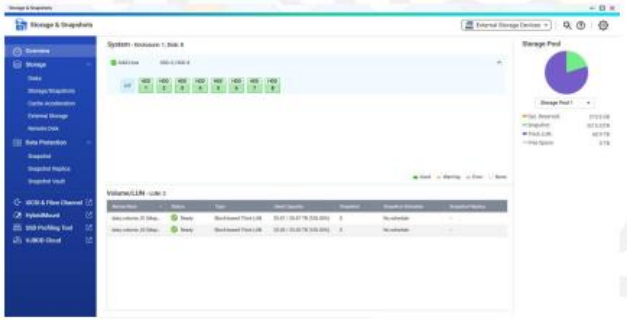


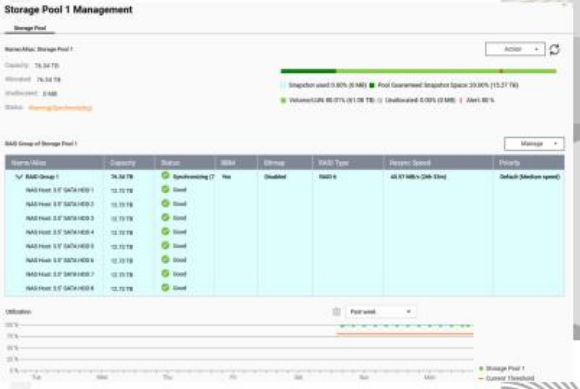
5

СИСТЕМА ЗБЕРІГАННЯ ДАНИХ

NAS архітектура

Загальна ємність: 100TB (8 дисків по 12.5TB)
Кеш-пам'ять: 32GB
RAID конфігурація: реалізовано RAID 6
Stripe size: 256KB
Том 1: 60TB для віртуальних машин (thin provisioning)
Том 2: 40TB для резервних копій (thick provisioning)
iSCSI підключення через два незалежних 10GbE шляхи





6

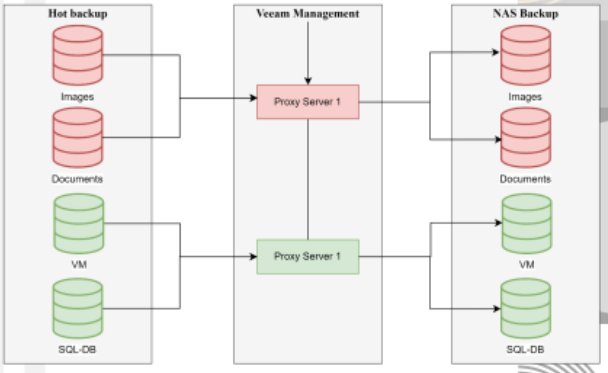
ПРОЦЕС РЕЗЕРВУВАННЯ ДАНИХ

Гарячий резерв (Hot Backup): дані, такі як зображення, документи, віртуальні машини (VM) та бази даних (SQL-DB), створюються у гарячому резерві.

Тунелі віртуалізації через vCenter: деякі документи передаються через спеціально налаштовані адаптери віртуалізації vCenter, які забезпечують безпечне та оптимізоване транспортування даних.

Proxy Server: всі резервні копії надходять через Proxy Server, який виступає посередником між системою гарячого резерву та сховищем NAS.

NAS із RAID 6: Дані зберігаються на NAS, налаштованому під RAID 6. Ця конфігурація забезпечує відмовостійкість, оскільки дозволяє втрату до двох жорстких дисків одночасно без втрати даних. Дані організовані в масив, де паритет забезпечує надійність відновлення.



7

ВІРТУАЛІЗАЦІЯ

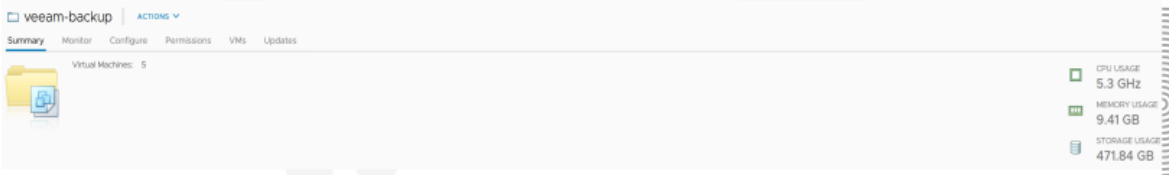
В основі архітектури лежить кластер з трьох потужних серверів. Мережева підсистема реалізована на базі чотирьох мережевих адаптерів з пропускною здатністю 25 Гбіт/с.

Для ефективного розподілу ресурсів впроваджено систему Resource Pools з трьома основними пулами.

Production Pool отримує 80% процесорних ресурсів та 70% оперативної пам'яті для забезпечення максимальної продуктивності критично важливих систем.

Development Pool виділено 15% CPU та 20% RAM для середовища розробки та тестування. Infrastructure Pool використовує решту ресурсів (5% CPU, 10% RAM) для службових потреб.

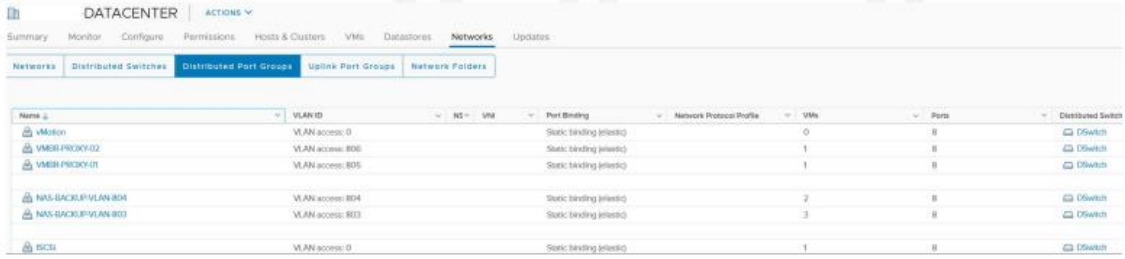
Технологія vMotion забезпечує можливість живої міграції віртуальних машин між серверами без переривання їх роботи.



8

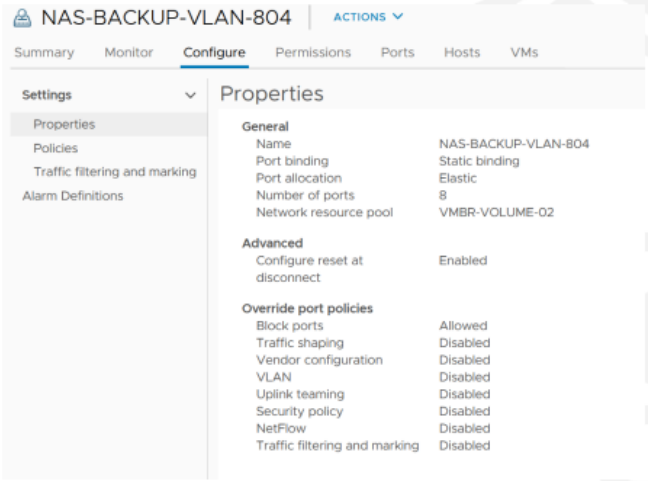
СТРУКТУРА ВІРТУАЛІЗАЦІЇ В СИСТЕМІ VCENTER

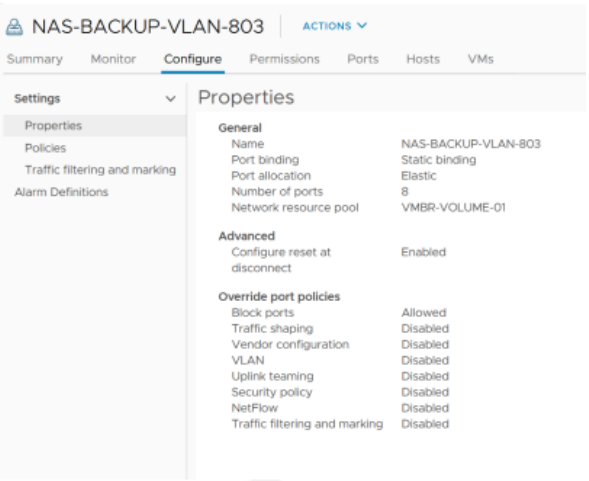
У системі VMware vCenter віртуалізація базується на ESXi-серверах, які виступають гіпервізорами для запуску віртуальних машин. Центральним компонентом є vCenter Server, що забезпечує управління всією інфраструктурою, включаючи моніторинг, балансування навантаження та автоматизацію процесів.Віртуальні машини працюють у кластері серверів, що гарантує високу доступність і оптимізацію ресурсів через технології HA, DRS, vMotion та Storage vMotion. Дані зберігаються в datastore-ах, які можуть базуватися на NAS, SAN або локальних сховищах, а мережева інфраструктура організована через віртуальні комутатори (vSwitches).



9

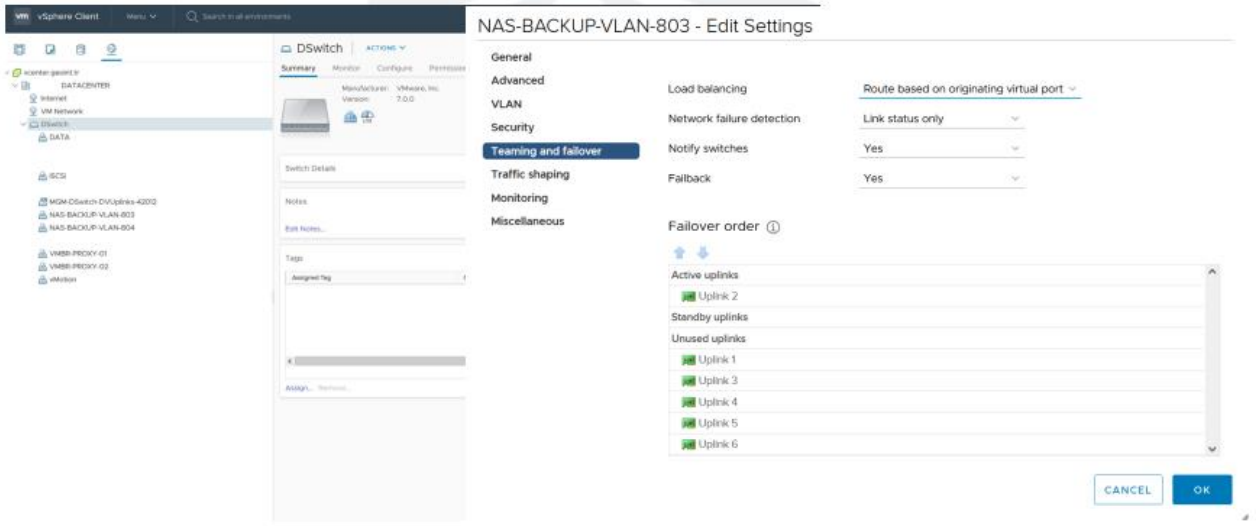
НАЛАШТУВАННЯ DSWITCH





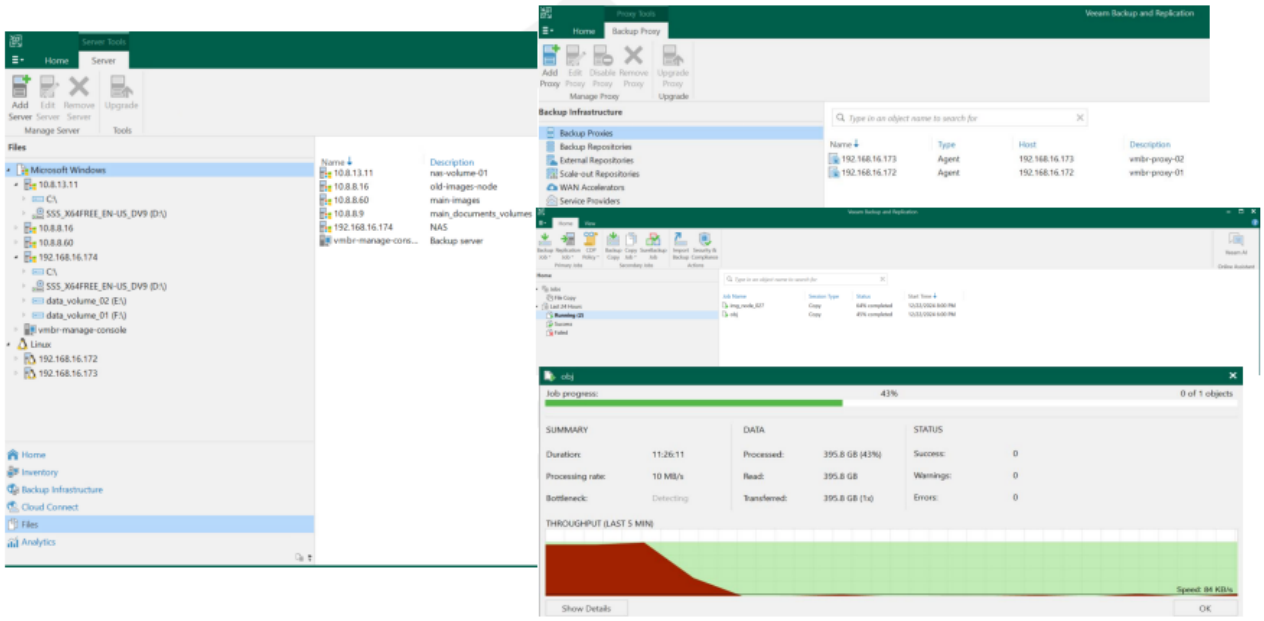
10

НАЛАШТУВАННЯ DSWITCH ТА ISCSI



11

РЕЗЕРВАЦІЯ ДАНИХ В VEEAM



12

ВИСНОВКИ

- Розроблено архітектуру резервного копіювання даних, яка поєднує гарячий і холодний резерв на основі Veeam Backup & Replication, NAS із RAID 6, та технологій віртуалізації vCenter. Забезпечено автоматизацію процесів копіювання даних, зокрема для документів, образів, віртуальних машин та SQL-баз.
- Система відповідає сучасним стандартам резервного копіювання, забезпечуючи надійність, безпеку та масштабованість. Виконано вимоги щодо автоматизації гарячих резервів та передачі їх на NAS через налаштовані тунелі віртуалізації.
- Мінімізацію ризиків втрати даних, швидке відновлення у разі надзвичайних ситуацій та ефективне управління великими обсягами інформації (до 100 ТБ). Її можна адаптувати під різні галузі: від корпоративного середовища до центрів обробки даних.

13

АПРОБАЦІЯ

1. IV Міжнародна наукова конференція «Період трансформаційних процесів в світовій науці: задачі та виклики» (13.12.2024, м. Рівне, Україна), тема: «Впровадження та оптимізація системи резервного копіювання на базі Dell EMC Networker з використанням стрічкової бібліотеки та розширеними політиками безпеки».
2. III Міжнародна наукова конференція «Інноваційна наука: пошук відповідей на виклики сучасності» (06.12.2024, м. Могилів-Подільський, Україна), тема: «Проектування та реалізація локальної системи інфраструктури даних для довгострокового зберігання великих масивів даних з використанням віртуалізації».

14