

ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ
Факультет комп'ютерної інженерії та управління
Кафедра електронних обчислювальних машин
Рівень вищої освіти другий (магістерський)

КВАЛІФІКАЦІЙНА РОБОТА

Модель оптимального збереження великих обсягів інформації у гібридних сховищах

Здобувач
Данило АБРАМОВИЧ
гр. СПМ-23-5

Керівник:
доц. Олексій ПІСКАРЬОВ

Харків
2025

1

МЕТА ТА ЗАДАЧІ РОБОТИ

Метою цієї роботи є розробка та дослідження адаптивної моделі оптимального збереження великих обсягів інформації у гібридних сховищах, що забезпечує динамічний розподіл даних між локальними та хмарними середовищами з урахуванням чутливості, частоти доступу та продуктивності системи.

Задачі:

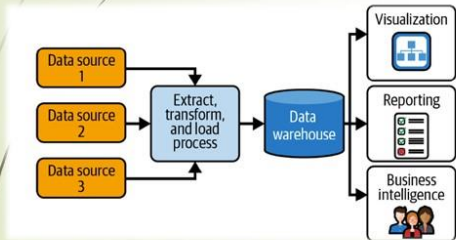
- ▶ провести комплексне дослідження сучасних архітектур та методів зберігання великих обсягів інформації у гібридних сховищах даних;
- ▶ проаналізувати підходи до організації систем збереження: Data Warehouse, Data Lakes, Data Lakehouse, Data Vault 2.0 та гібридних архітектур;
- ▶ запропонувати адаптивну математичну модель на базі ADH-DVL, що дозволить оптимально розподіляти дані між локальними та хмарними середовищами;
- ▶ провести моделювання роботи моделі при різних сценаріях навантаження.

2

АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО СТВОРЕННЯ СХОВИЩ ВЕЛИКИХ ОБСЯГІВ ДАНИХ

Сховище даних (Data Warehouse, DWH) — це програмна система для централізованого збереження та управління великими обсягами структурованих даних, які збираються з різних джерел (реляційних баз даних - RDBMS, сховищ даних - Data Lakes, зовнішніх джерел чи файлів) для їх подальшого використання в аналітиці, звітності, прогнозуванні та ухваленні бізнес-рішень

Структура корпоративного сховища даних

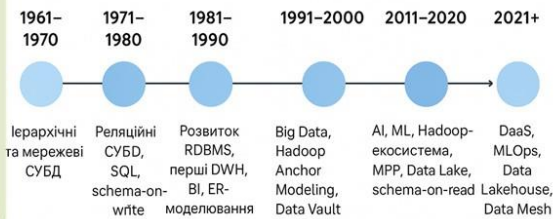


Якщо сховище даних використовується всією компанією, його називають корпоративним сховищем даних (Enterprise Data Warehouse, EDW). EDW є розширеною версією DWH, яка охоплює ширший спектр джерел та типів даних для підтримки всіх бізнес-підрозділів організації. Таким чином, EDW виконує функцію єдиного джерела правдивої інформації (Single Version of Truth, SVOT) для компанії, забезпечуючи узгоджене, централізоване представлення даних

3

АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО СТВОРЕННЯ СХОВИЩ ВЕЛИКИХ ОБСЯГІВ ДАНИХ

Еволюція архітектур даних

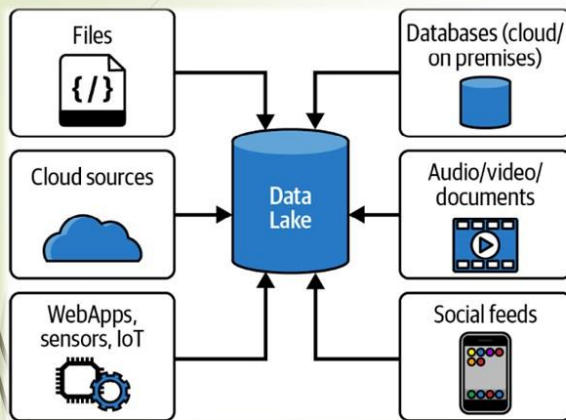


- У 1961–1970 роках з'явилися перші комерційні СУБД (IMS від IBM, IDS від General Electric), що заклали основи аналітики та роботи зі структурованими даними на основі ієрархічних і мережових моделей.
- У 1971–1980 роках відбулося становлення реляційних баз даних і мови SQL, сформувався концепції ETL, OLTP та підхід schema-on-write. У 1981–1990 роках розвинулись RDBMS (Oracle, IBM DB2, Microsoft SQL Server), з'явилися ER-моделювання, DWH та перші BI-інструменти.
- У 1991–2000 роках активно впроваджувалися DWH, OLAP, UML, dimensional та об'єктно-реляційні моделі, зокрема PostgreSQL.
- У 2001–2010 роках, зростання обсягів даних сприяло розвитку Big Data, Hadoop, Anchor Modeling та Data Vault.
- У 2011–2020 роках інтегрувались AI, ML, хмарні технології, розвинулись Hadoop-екосистема, MPP, Data Lake, schema-on-read та системи горизонтального масштабування (Cassandra, MongoDB, Google BigQuery).
- З 2021 року активно розвиваються хмарні сервіси DaaS, MLOps, архітектури Data Lakehouse та Data Mesh.

4

АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО СТВОРЕННЯ СХОВИЩ ВЕЛИКИХ ОБСЯГІВ ДАНИХ

Структура сховища Data Lakes



Data Lake — це сховище, яке дозволяє зберігати всі можливі типи даних, включаючи структуризовані, напівструктуризовані та неструктуризовані, без необхідності попереднього моделювання або перетворення. На відміну від традиційних DWH, Data Lake не має власного нативного обчислювального механізму і виступає лише як гнучке сховище даних у їх початковому форматі. Цей підхід базується на концепції *schema-on-read*, що означає, що структура даних визначається лише під час читання та аналізу, а не при записі.

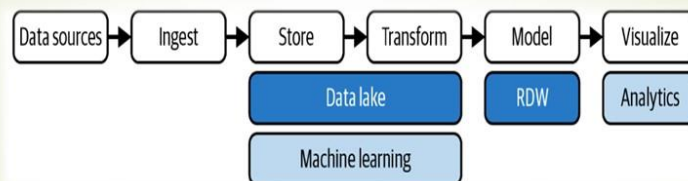
5

АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО СТВОРЕННЯ СХОВИЩ ВЕЛИКИХ ОБСЯГІВ ДАНИХ

Сучасні сховища даних сховища (Modern Data Warehouse)

Попри переваги, Data Lakes не змогли повністю замінити реляційні DWH, оскільки не забезпечували якісного аналітичного опрацювання, залишаючись ефективними для зберігання та підготовки необроблених даних. Це призвело до розвитку гібридних архітектур, які поєднують гнучкість Data Lake з продуктивністю DWH, утворюючи сучасне сховище даних (Modern Data Warehouse, MDW). MDW базується на розподіленому зберіганні, хмарних технологіях і підтримці різних форматів даних, де Data Lake збирає та обробляє інформацію, а DWH забезпечує глибоку аналітику та звітність.

Структура сховища Modern Data Warehouse



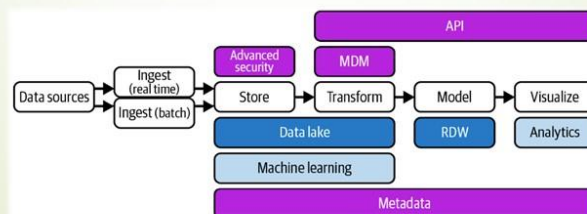
6

АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО СТВОРЕННЯ СХОВИЩ ВЕЛИКИХ ОБСЯГІВ ДАНИХ

Data Fabric

Зі зростанням складності даних і потреби в автоматизації з 2016 року почала розвиватися архітектура Data Fabric як еволюція MDW з розширеними можливостями. Вона поєднує обробку даних у реальному часі, централізований каталог метаданих, управління еталонними даними (MDM), API-інтерфейси та вдосконалені політики доступу, забезпечуючи швидку, автоматизовану та масштабовану роботу з інформацією. Data Fabric дає змогу ефективно працювати з розподіленими джерелами в локальних, хмарних і гібридних середовищах, що підвищує ефективність аналітики та створює основу для подальшої автоматизації управління даними.

Структура сховища Data Fabric



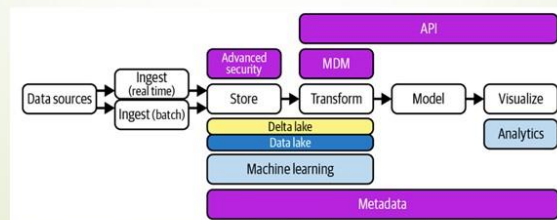
7

АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО СТВОРЕННЯ СХОВИЩ ВЕЛИКИХ ОБСЯГІВ ДАНИХ

Data Lakehouse

З 2020 року, після презентації компанією Databricks концепції Data Lakehouse і медальйонної архітектури, набув поширення підхід, що замінює реляційні DWH єдиним сховищем Data Lake для зберігання всіх типів даних та виконання аналітики безпосередньо в ньому. На відміну від класичних Data Lakes, у Data Lakehouse додано транзакційний програмний шар, який забезпечує продуктивність, подібну до DWH, без дублювання даних. Для його реалізації створено формати Delta Lake, Apache Iceberg і Apache Hudi, що підтримують одночасну обробку читання та запису без втрати продуктивності та цілісності даних. Таким чином, Data Lakehouse поєднує масштабованість і гнучкість Data Lake із аналітичною ефективністю DWH.

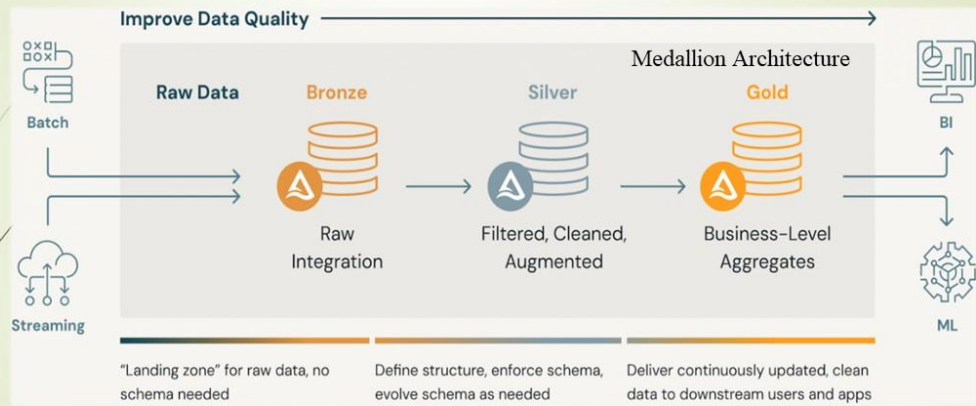
Структура сховища Data Lakehouse



8

СТРУКТУРА DWH ТА МОДЕЛІ ДАНИХ

Шарова структура DWH та архітектура «медальйон»



Архітектура передбачає три рівні: Bronze — зберігає сирі дані у вихідному форматі; Silver — містить очищені, узгоджені та структуровані дані після базових трансформацій; Gold — включає агреговані, оптимізовані дані для аналітики, звітності та машинного навчання. Такий підхід поєднує гнучкість Data Lakehouse з класичними методами моделювання даних.

9

СТРУКТУРА DWH ТА МОДЕЛІ ДАНИХ

Повільно-змінювані вимірювання

Схема «зірка»

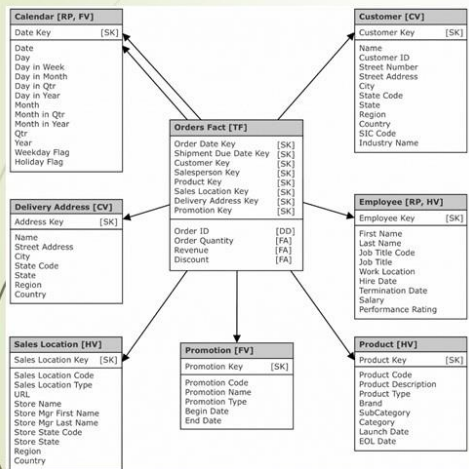
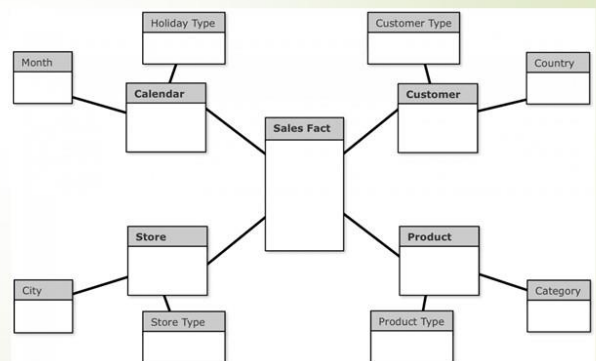


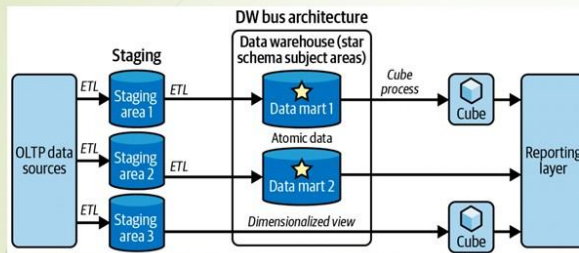
Схема «сніжинка»



10

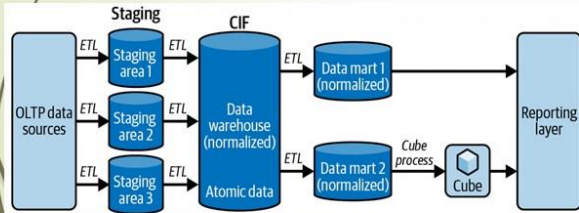
СТРУКТУРА DWH ТА МОДЕЛІ ДАНИХ

Побудова DWH за методом Кімбулла



Кімбалл розробив низку методів для аналітичних сховищ, зокрема погоджені, повільно змінювані, мусорні та міні-вимірювання, місткові таблиці й таблиці фактів періодичних знімків. Його методологія побудови DWH базується на bottom-up підході: дані з OLTP спочатку потрапляють у стейджингові таблиці без попередньої обробки, після чого проходять інтеграцію, очищення й організуються у розмірні моделі з таблицями фактів і вимірювань. Такий підхід дозволяє швидко створювати аналітичні рішення, що поступово інтегруються у єдине сховище, і є ефективним для оперативної аналітики та бізнес-звітності.

Побудова DWH за методом Інмона

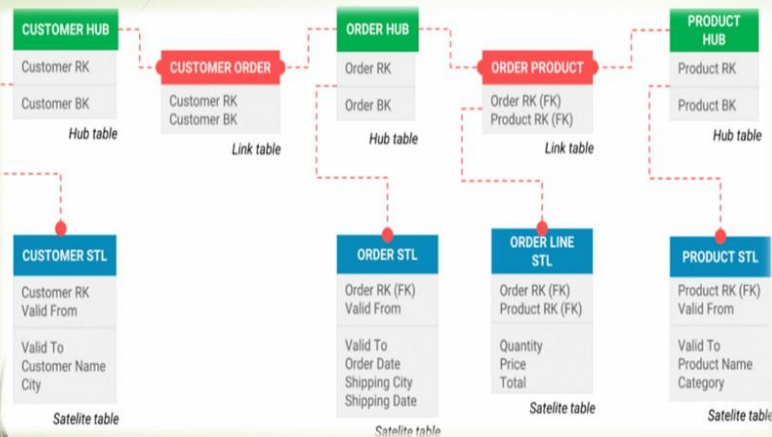


Інмон пропонує створення нормалізованого DWH як єдиного сховища з уніфікованими даними, доповненого аналітичними вітринами (Data Marts). Кімбалл орієнтується на швидке впровадження аналітики через окремі Data Marts для конкретних бізнес-завдань. Підхід Інмона більше підходить великим корпораціям, а Кімбалла — організаціям, що потребують швидких результатів.

11

СТРУКТУРА DWH ТА МОДЕЛІ ДАНИХ

Концепція Data Vault

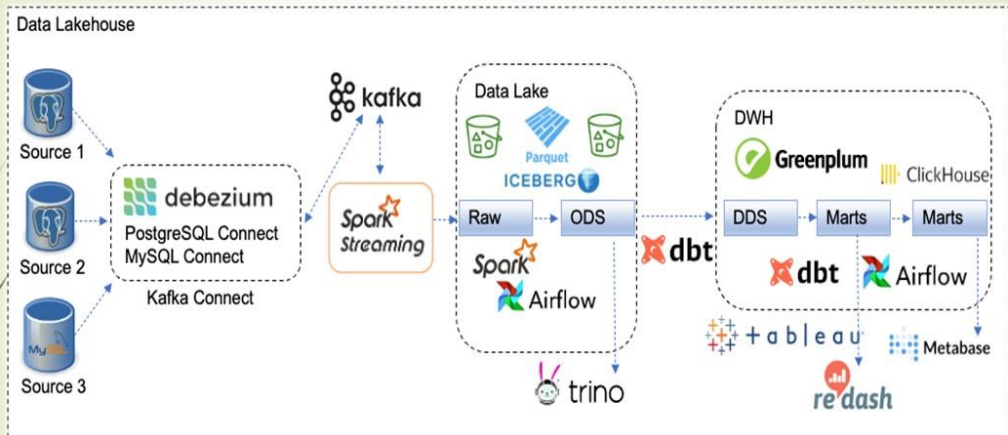


Створена Деном Лінстедтом у 2000 році, концепція Data Vault є методологією проектування DWH, орієнтованою на гнучкість, масштабованість і збереження історії змін, краще пристосованою до Big Data. Модель складається з хабів (бізнес-сутності), послань (зв'язки між сутностями) та сателітів (описові атрибути й історія змін). Така структура дозволяє ефективно масштабувати сховище, інтегрувати нові джерела й відстежувати зміни без втрати продуктивності.

12

СТРУКТУРА DWH ТА МОДЕЛІ ДАНИХ

Приклад архітектури ELT у Data Lakehouse

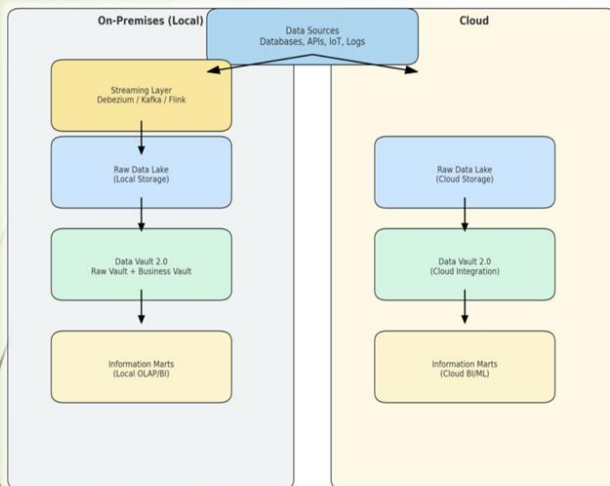


Архітектура ELT у Data Lakehouse використовує сучасні технології для обробки великих даних. Дані з OLTP-баз, логів, API та потоків (Kafka, Flink) витягуються без трансформації та завантажуються в Raw Zone об'єктних сховищ (MinIO, HDFS, ClickHouse). Трансформації здійснюють Spark, dbt і Trino. Аналітичні системи (Superset, Metabase, Redash) працюють із підготовленими даними, інтегруючись із ML-моделлями. Для обробки змін у реальному часі застосовуються Debezium, Kafka, Flink і Spark Streaming, а процесами керують Airflow і dbt.

13

МОДЕЛЬ ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ У ГІБРИДНИХ СХОВИЩАХ

Архітектура запропонованої моделі



Запропонована модель на базі ADH-DVL інтегрує сучасні технології та методології для гібридних систем збереження даних, поєднуючи локальні та хмарні ресурси задля гнучкого розподілу, оперативності та високої доступності. Дані з різних джерел надходять у Raw Data Lake у сирому вигляді, зберігаючись як локально (для безпеки та швидкодії), так і в хмарі (для масштабованості). На інтеграційному рівні застосовується Data Vault 2.0: хаби представляють бізнес-сутності, зв'язки — їх взаємодії, а сателіти — атрибути та історію змін. Далі формується Business Vault, де дані трансформуються за бізнес-правилами, агрегуються й готуються до аналітики. Аналітичний рівень Information Marts забезпечує денормалізовані структури для звітності, BI та OLAP. Окремий Streaming Layer здійснює захоплення змін у режимі реального часу (CDC), використовуючи Kafka Connect, Kafka, Flink та Spark Streaming для обробки потоків і оперативного оновлення аналітичних даних.

14

МОДЕЛЬ ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ У ГІБРИДНИХ СХОВИЩАХ

Математична модель розподілу даних

$$V = V_L + V_C.$$

У центрі моделі лежить розподіл даних між локальною та хмарною частинами системи, що формалізується рівнянням балансу: загальний обсяг даних V складається з локальної частини V_L і хмарної V_C . Це забезпечує збереження повноти інформації та слугує основою для адаптивного розподілу даних у системі.

$$V_L = \alpha V (f_A + S).$$

Розподіл даних між локальною та хмарною частинами залежить від політики безпеки, чутливості даних та частоти їх використання. Формула для обсягу локального зберігання враховує коефіцієнт локалізації $\alpha \in [0, 1]$, нормалізовану частоту доступу f_A та рівень чутливості S . Чим більшим є α , тим більше даних лишається локально. Обидва параметри f_A та S нормалізовані (0–1), що дозволяє гнучко враховувати вимоги при визначенні локального обсягу зберігання.

$$R(t) = \gamma_L (f_A(t) - \bar{f}_A).$$

Модель включає адаптивну функцію корекції локального обсягу зберігання. Змінна $R(t)$ визначається як різниця між поточною частотою доступу $f_A(t)$ та її середнім значенням \bar{f}_A за період спостереження. Коефіцієнт γ регулює швидкість і амплітуду адаптації. За зростання активності система автоматично збільшує локальне зберігання, знижуючи залежність від хмари та затримки доступу.

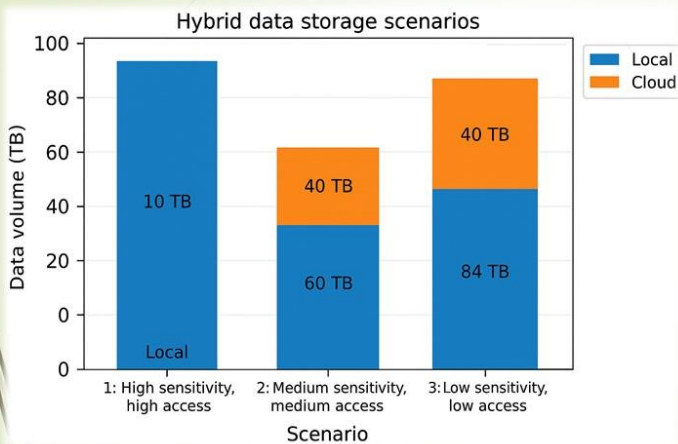
$$L_{total} = L_{ingest} + L_{stream} + L_{ETL} + L_{query}.$$

Оцінка ефективності системи враховує часові витрати на обробку даних у всіх ключових компонентах через рівняння загальної затримки. L_{ingest} — затримка захоплення змін (CDC, Debezium), L_{stream} — затримка стрімінгової обробки (Flink, Spark Streaming), L_{ETL} — затрати часу на трансформацію в Data Vault (нормалізація, очищення, історизація, збагачення), L_{query} — затримки виконання запитів у BD-системах. Сума цих затримок визначає загальний час відповіді та слугує основою для оптимізації продуктивності.

15

ДОСЛІДЖЕННЯ МОДЕЛІ ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ У ГІБРИДНИХ СХОВИЩАХ

Оцінка ефективності моделі



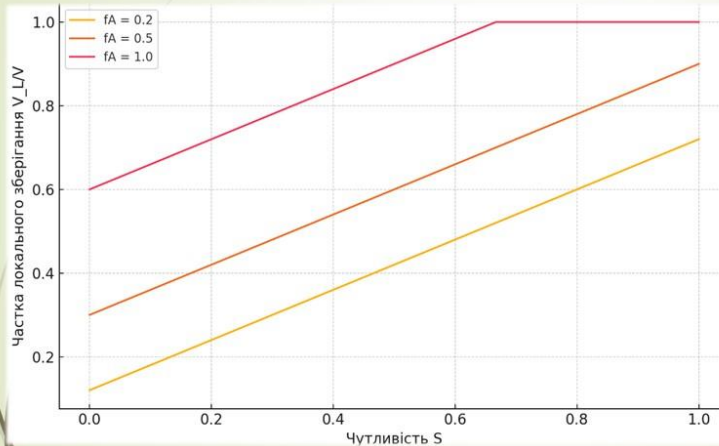
Ефективність адаптивної моделі ADH-DVL оцінювалась шляхом моделювання трьох сценаріїв із різною чутливістю даних, частотою доступу та політикою локалізації. У першому сценарії (10 ТБ, чутливість 0.8, доступ 1.0, $\alpha=0.8$) весь обсяг даних залишено локально. У другому (100 ТБ, чутливість 0.5, доступ 0.5, $\alpha=0.6$) — 60 ТБ зберігаються локально, 40 ТБ — у хмарі. У третьому (100 ТБ, чутливість 0.2, доступ 0.2, $\alpha=0.4$) — локально залишається 16 ТБ, решта 84 ТБ — у хмарній інфраструктурі. Результати підтверджують адаптивність моделі до різних умов, забезпечуючи баланс безпеки, ефективності та економічності при масштабуванні гібридної інфраструктури.

16

ДОСЛІДЖЕННЯ МОДЕЛІ ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ У ГІБРИДНИХ СХОВИЩАХ

Дослідження поведінки гібридної моделі на базі ADH-DVL

Залежність частки локального зберігання від чутливості



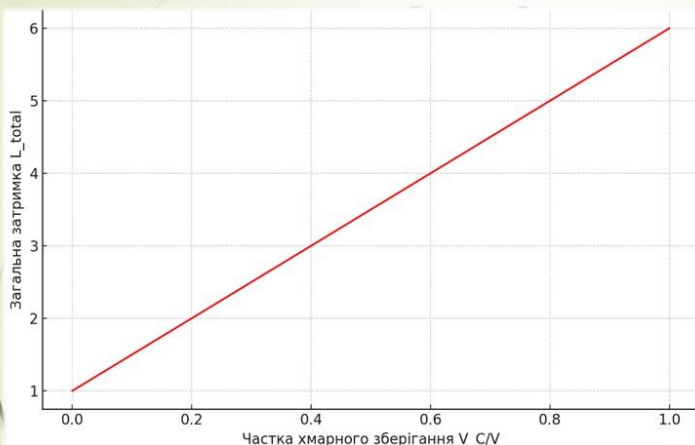
Перша серія моделювання показала, як зі зростанням чутливості даних збільшується частка локального зберігання. Вища конфіденційність і бізнес-критичність вимагають локалізації для мінімізації ризиків витоків і загроз. Особливо це помітно за високої частоти доступу: навіть помірною чутливістю при активному використанні призводить до значної локальної присутності. Модель коректно балансує між безпекою та ефективністю, забезпечуючи оптимізацію ресурсів і дотримання регуляторних вимог.

17

ДОСЛІДЖЕННЯ МОДЕЛІ ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ У ГІБРИДНИХ СХОВИЩАХ

Дослідження поведінки гібридної моделі на базі ADH-DVL

Залежність затримки від частки хмарного зберігання



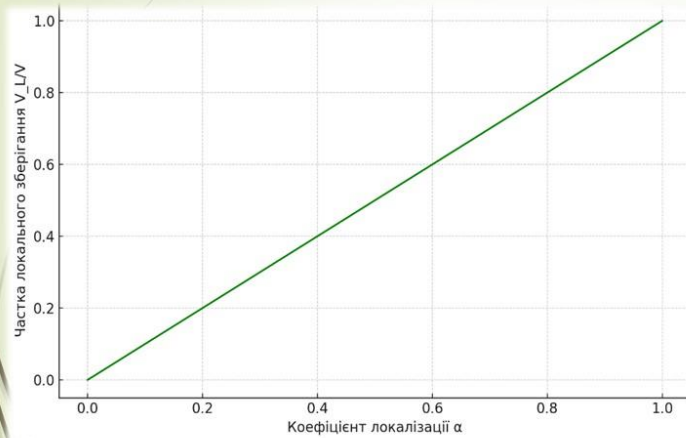
Друге моделювання показало, що зі збільшенням обсягу даних у хмарі зростає загальна затримка системи через мережеві комунікації та розподілену обробку. Компроміс між масштабованістю і швидкістю особливо важливий для систем реального часу — фінансових, телемедицини чи виробничих, де затримки можуть призвести до фінансових втрат або порушень технологічних процесів.

18

ДОСЛІДЖЕННЯ МОДЕЛІ ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ У ГІБРИДНИХ СХОВИЩАХ

Дослідження поведінки гібридної моделі на базі ADH-DVL

Залежність частки хмарного зберігання від коефіцієнту локалізації



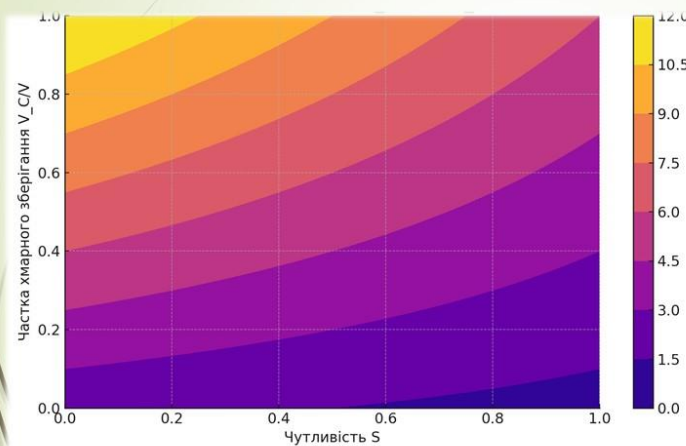
Третє моделювання показало вплив адміністративного параметра політики локалізації, який дозволяє гнучко змінювати частку локального зберігання без зміни властивостей даних. Це дає змогу оперативно реагувати на нові нормативи, ризики чи зміни вартості хмарних послуг, що особливо важливо для банківських, державних, оборонних і медичних систем. Лінійна залежність підтверджує простоту й ефективність впровадження цього механізму у великих датацентрах.

19

ДОСЛІДЖЕННЯ МОДЕЛІ ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ У ГІБРИДНИХ СХОВИЩАХ

Дослідження поведінки гібридної моделі на базі ADH-DVL

Вплив чутливості та хмарного зберігання на затримку



Четверте моделювання досліджує взаємодію чутливості даних і частки хмарної компоненти, формуючи аналітичний простір компромісів між продуктивністю, безпекою та витратами. Найбільші затримки виникають при високій частці хмарних даних з низькою чутливістю. При зростанні чутливості модель автоматично збільшує локальне зберігання, покращуючи затримки. Це дозволяє підприємствам будувати оптимальні сценарії гібридних сховищ, балансуючи витрати та SLA.

20

ДОСЛІДЖЕННЯ МОДЕЛІ ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ У ГІБРИДНИХ СХОВИЩАХ

Програмна реалізація

Фрагмент робочого коду, що демонструє застосування моделі на базі ADH-DVL у реальній системі зберігання даних

```

# Ініціалізація роботи системи збереження
class HybridStorageSystem:
    def __init__(self, total_capacity_tb, local_capacity_tb):
        self.total_capacity = total_capacity_tb
        self.local_capacity = local_capacity_tb
        self.local_used = 0
        self.cloud_used = 0
    def adaptive_store(self, data_size_tb, alpha, fb, S):
        # Розрахунок оптимального розміщення даних
        VL = alpha * data_size_tb * (fb + S)
        if VL > data_size_tb:
            VL = data_size_tb
        # Перевірка наявності вільного місця локально
        if self.local_used + VL <= self.local_capacity:
            self.local_used += VL
            self.cloud_used += (data_size_tb - VL)
            print(f'Дані обсягом {data_size_tb} TB збережено: {VL} TB локально, {data_size_tb - VL} TB у хмарі.')
        else:
            # Якщо локально місця недостатньо – все відправляється в хмару
            self.cloud_used += data_size_tb
            print(f'Недостатньо локального місця. Дані обсягом {data_size_tb} TB збережено повністю у хмарі.')
    def status(self):
        print(f'Використано локально: {self.local_used} TB з {self.local_capacity} TB.')
        print(f'Використано у хмарі: {self.cloud_used} TB.')
        print(f'Загальне збереження: {self.local_used + self.cloud_used} TB з {self.total_capacity} TB в.')

# Ініціалізація гібридної системи
total_storage = 500 # Загальний обсяг системи, наприклад 500 TB
local_storage = 200 # Місце локально: 200 TB
system = HybridStorageSystem(total_storage, local_storage)

# Нові надходження даних:
system.adaptive_store(data_size_tb=50, alpha=0.5, fb=1.0, S=0.5)
system.adaptive_store(data_size_tb=100, alpha=0.5, fb=0.1, S=0.4)
system.adaptive_store(data_size_tb=150, alpha=0.2, fb=0.2, S=0.1)
system.status()

```

Розроблена програмна реалізація демонструє роботу керуючого модуля гібридного сховища з адаптивним розподілом ресурсів. Клас HybridStorageSystem містить параметри ємності системи та поточний стан локального й хмарного зберігання. Метод adaptive_store() приймає рішення про розміщення нових даних залежно від частоти доступу, чутливості та політики розподілу, автоматично балансує між локальними й хмарними ресурсами. Метод status() відображає поточний стан розподілу даних у системі.

Наукові публікації за темою кваліфікаційної роботи



Секція 2. Інтеграція IT у промисловість: сучасні тенденції

ДК 697.9:004.032.26:004.652

МОДЕЛЬ ІНТЕЛЕКТУАЛЬНОЇ ПІДТРИМКИ МІКРОКЛІМАТУ У ВЕЛИКИХ ПРИБЛИЖНИХ ІЗ ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖЕВИХ ПІДХОДІВ ТА ГІБРИДНИХ СХОВИЩ ІНФОРМАЦІЇ

Пісаров О.М., к.т.н., доц.

Абрамчик Д.О., здобувч РВО магістр

Пугун В.В., здобувч РВО магістр

Харківський національний університет радіоелектроніки

м. Харків, Україна, oleksii.pisarov@nure.ua

Анотація. У статті розглянуто модель інтелектуальної підтримки мікроклімату у великих приміщеннях, що ґрунтується на алгоритмі штучного нейронного мережі та гібридній архітектурі зберігання даних. Запропоновано підхід до обробки та накопичення великих обсягів сенсорної інформації у системі моніторингу мікроклімату, що дозволяє забезпечити як високу точність прогнозування, так і надійне збереження даних. Архітектура системи виможує спеціальної динамічної середовища та дозволяє адаптуватися до змінних умов, оптимізуючи керування мікрокліматом. Модель реалізує інтеграцію хмарних та локальних сховищ з нейромережами, що навігаються на основі даних та забезпечує гнучке регулювання параметрів мікроклімату.

Ключові слова: мікроклімат, нейронна мережа, гібридне сховище, інтелектуальна система, прогнозування, оптимізація

Автоматизація керування мікрокліматом у великих приміщеннях – одне з пріоритетних завдань сучасного теплого середовища, оскільки забезпечення стабільних умов сприяє як підвищенню комфорту, так і ефективному енергоспоживанню. Системи моніторингу та регулювання потребують обробки великих масивів даних, зокрема з різноманітних сенсорів, що постійно фіксують зміни параметрів. Водночас ці дані мають бути збережені у надійній формі з можливістю швидкого доступу для навігації моделей прогнозування. Використання гібридних сховищ дозволяє зберігати перебіги локальних і хмарних рішень, а штучні нейромережі – ефективно працювати з інструментованими та нелінійними залежностями, характерними для процесів формування мікроклімату.

Метою роботи є створення адаптивної інтелектуальної середовища, яке забезпечує комфорт, енергоефективність та можливість довготривалого зберігання скроєних даних для аналізу. Загальна архітектура системи представлена на рис. 1. Вона включає модуль збору даних на основі мікроконтролера STM32F407, який зчитує сигнали з датчика температури, вологості CO2 та освітлення. Дані передаються до гібридного сховища [1], що складається з локальної та хмарної частин, завдяки чому відбувається розподіл даних між локальними та хмарними сховищами. На основі прогнозу система адаптивного керування формує керуючі сигнали для виконавчих механізмів, забезпечуючи оптимальні умови в приміщенні відповідно до

Секція 2. Інтеграція IT у промисловість: сучасні тенденції

PMV або вручну заданим), розподілом відключення та коригує режим роботи пристрою у режимі реального часу. Це дозволяє уникати некорисних витрат і забезпечити стабільність мікроклімату навіть при різких зовнішніх змінах [3].

Запропонована модель інтелектуальної підтримки мікроклімату у великих приміщеннях демонструє ефективне поєднання технологій штучного інтелекту, гібридного зберігання даних та мікроконтрольного керування. В основі системи лежить використання нейромережевих алгоритмів (LSTM) для прогнозування зміни параметрів середовища, що дозволяє здійснювати адаптивне регулювання у режимі реального часу. Інтеграція моделі PMV забезпечує можливість адекватного оцінювання теплого комфорту, а застосування гібридного сховища підвищує надійність і масштабованість архітектури, що особливо важливо при обробці великих обсягів сенсорної інформації [4].

Практична значущість запропонованої системи полягає у можливості її впровадження у різноманітні галузі: офісні простори, торгові комплекси, логістичні та виробничі центри, навчальні заклади. Використання адаптивної інтелектуальної керування дає змогу не лише підвищити комфортність середовища, але й суттєво знизити енергоспоживання за рахунок точного прогнозування та своєчасного втручання у керуванні їй.

Перспективні розробки системи включають розширення функціональності шляхом інтеграції з іншими елементами "розумного будинку" або "розумного виробництва", зокрема, з системами безпеки, розпізнавання присутності, адаптивного освітлення тощо. Крім того, передбачається вивчення можливостей використання більш складних нейромережевих архітектур, таких як трансформери, для глибокого аналізу причинно-наслідкових залежностей між параметрами середовища [5].

Загалом, представлена система закладає підґрунтя для створення нових поколінь інтелектуальних керуючих комплексів, що здатні ефективно функціонувати в умовах швидкозмінних середовища та великого обсягу даних, зберігаючи стабільність, адаптивність і масштабованість.

Список літератури

1. Павлов І.І., Сирен С.В. Система збереження великих даних: архітектура та моделі // Вісник НТУУ "КПІ". Серія: Комп'ютерна наука. – 2021. – № 2. – С. 45–53.
2. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on CVPR. – 2016. – P. 770–778.
3. Тараненко В.С., Котенко І.М. Нейронні мережі та їх застосування в автоматизованих системах керування // Автоматизація процесів керування. – 2020. – № 3. – С. 12–18.
4. Zhang Y., Zheng Y., Qi D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction // Proceedings of AAAI. – 2017.
5. Власов М.М., Колосовський В.М. Хмарні та гібридні технології зберігання інформації // Системи обробки інформації. – 2022. – № 1. – С. 89–94.

МНІК «Інформаційні технології в сучасному світі», Харків, ДВУ, 2025

ціліхових стандартів комфорту. Координація між модулями реалізована за допомогою стандартних шифрованих протоколів та API.



Рис. 1. Структурна схема системи підтримки мікроклімату

Середня система є мікроконтролером STM32F407, що виконує регулювання даних з шифрованих сенсорів температури, вологості, рівня CO2, освітленості та ін. Дані передаються через протоколи I2C, SPI та UART. Обробка інформації на нижньому рівні (фільтр, первинна фільтрація, агрегація) відбувається безпосередньо на рівні контролера. У системі реалізовано апаратну підтримку адаптивного керування на основі моделі PMV (Predicted Mean Vote), що дозволяє кількісно оцінити тепловий комфорт користувача на основі фізіологічних параметрів та параметрів середовища [2].

Усі отримані дані передаються на центральний сервер, де реалізовано гібридне сховище даних: частина даних розміщується локально в базі даних (наприклад, SQLite або PostgreSQL), решта синхронізується з хмарними сховищами (наприклад, Amazon S3 або Firebase). Така архітектура забезпечує високу доступність, надійність та масштабованість системи. Дані останніх двох днів зберігаються в RAM-пам'яті для швидкого доступу нейроної мережі, а решта даних – в хмарі з підтримкою резервного копіювання та ротації.

Для прогнозування параметрів мікроклімату у системі застосовано штучну нейронну мережу типу LSTM (Long Short-Term Memory), що враховує тимчасову залежність у даних. Навчання мережі проводиться на історичних даних, отриманих з гібридного сховища. Мережева модель навчає на основі реальних даних з експериментальної теплого середовища та офісного середовища. При цьому були враховані типові сценарії зміни: прохід людей у приміщення, денні коливання температури, відкриття вікон, зміни погоди. Завдяки використанню методу скользящего вікна та адаптивного градієнта, модель здатна оперативно адаптуватися до нових умов і прогнозувати навігання без винесення керування.

Регулювання здійснюється автоматично за принципом замкнутого зв'язку: система порівнює прогнозовані стан з бажаним (визначеним через

Наукові публікації за темою кваліфікаційної роботи



вул. Ільїна, 6/1,
м. Одеса, Україна, 65101
www.glybetska.ua
mailbox@glybetska.ua

Станціонер: 048 709 38 69
Vodafone: 095 934 48 28
Kyivstar: 097 723 06 08

ДОВІДКА

Видавничий дім «Гельветика» за домовленістю з Міжрегіональним Академією управління персоналом є офіційним видавцем журналу «Інформаційні технології та суспільство» та займається усіма видавничо-копірафікованими процесами, до яких належить: вибір статей до чергового випуску; рецензування; перевірка на плагіат; коректорська випитка; верстка; присвоєння кожному матеріалу РOІ; розміщення електронної версії видання на офіційному сайті журналу; надсилання електронної версії видання до Національної бібліотеки України імені В. І. Вернадського на репозитарій України; розсилка обов'язкового безоплатного примірника до державних установ України.

Цією довідкою повідомляємо, що наукова стаття Пискарьова Олександра, Абрамовича Данила, Плутина Владислава «Застосування нейронерг та гібридних сховищ при створенні комп'ютерних систем» прийнята редакцією журналу «Інформаційні технології та суспільство» для розміщення в Випуску 2 (17) за 2025 рік.

Директор

Видавничого дому «Гельветика»



Олег ГОЛОВКО

УДК 004.77

Олександр ПИСКАРЬОВ, кандидат технічних наук, доцент, доцент кафедри електронних обчислювальних машин Харківської національної технічної університет радіоелектроніки, oleksii.piskarov@nure.ua ORCID: 0000-0002-6980-984X

Данило Абрамович, Здобувач РВО магістр-науковець Харківської національної технічної університет радіоелектроніки, danylo.abramovych@nure.ua

Владислав Плутич, Здобувач РВО магістр-науковець Харківської національної технічної університет радіоелектроніки, vladislav.plutin@nure.ua

Застосування нейронерг та гібридних сховищ при створенні комп'ютерних систем

Анотація. У сучасних комп'ютерних системах зростає потреба у швидкій обробці надійному зберіганні та точному аналізі великих обсягів даних у реальному часі. Це особливо актуально для автоматизованих технологічних процесів, інтелектуальних середовищ та керування розподіленою об'єктами. У таких умовах традиційні архітектури зберігання та аналітики даних стають недостатньо ефективними та масштабованими. Метою статті є огляд сучасних підходів до побудови комп'ютерних систем прогнозування стану дозволити об'єктам шляхом поєднання методів машинного навчання з гібридною архітектурою зберігання інформації. У рамках роботи здійснюється оцінка можливості різних моделей штучного інтелекту для коректного прогнозування параметрів, а також аналізується ефективність використання гібридних сховищ даних для забезпечення надійного, масштабованого та продуктивного функціонування таких систем у реальному часі. **Методи дослідження:** використовуються методи машинного навчання, зокрема рекурсії та аналітичні моделі, а також експериментальне моделювання на основі часових рядів. Для оцінки ефективності запропонована модель впроваджена та аналіз впливу різних параметрів. Обробка зберігання даних реалізована з використанням гібридної інфраструктури. **Наукова новизна дослідження** полягає в тому, що проведений аналіз впливу різних параметрів інтеграції високошвидкісних методів машинного навчання з гібридними сховищами інформації

для створення адаптивної комп'ютерної системи прогнозування, здатної до оперативної реакції на динамічні зміни в умовах обробки великих обсягів даних у реальному часі. **Висновки.** Проведене дослідження дозволяє визначити використання машинного навчання у поєднанні з гібридною архітектурою зберігання даних для створення інтелектуальних комп'ютерних систем прогнозування. На основі експериментального моделювання встановлено, що методи ХІВові демонструють надійну точність, стабільність результатів і здатність до інтеграції впливу різних параметрів на вхідні дані, значно перевершуючи інші пропозитивні алгоритми. Важливо складовою запропоновано підходу стала побудова гібридного сховища інформації, яке поєднує переваги локального зберігання, NoSQL-систем та хмарної інфраструктури. Така структура дозволяє забезпечити як оперативне реагування на зміни вхідних параметрів, так і проведення глибокого історичного аналізу для довготривалого планування та оптимізація алгоритмів. Запропонована система може бути адаптована для широкого кола завдань у сфері автоматизованого керування, цифровою моніторингу, управління виробничими процесами та в інших промислових напрямках. Вона дозволяє не лише підвищити ефективність роботи комп'ютерних систем, а й закласти основу для створення самонавчальних, адаптивних рішень нового покоління, здатних працювати в умовах реального часу з великими обсягами даних.

Ключові слова: комп'ютерні системи, машинне навчання, гібридні сховища, інтелектуальне керування, автоматизація, хмарні технології.

Oleksii Piskarov, Danilo Abramovich, Vladislav Plutin. Application of neural networks and hybrid storage in the creation of computer systems

Abstract. In modern computer systems, there is a growing need for fast processing, reliable storage, and accurate analysis of large amounts of data in real time. This is especially true for automated technological processes, intelligent environments, and distributed object management. In such circumstances, traditional data storage and analytics architectures become insufficiently flexible and scalable. The purpose of the article is to review modern approaches to building computer systems for predicting the state of dynamic objects by combining machine learning methods with a hybrid information storage architecture. The paper evaluates the accuracy of various artificial intelligence models for short-term parameter forecasting and analyzes the effectiveness of using hybrid data warehouses to ensure reliable, scalable, and productive operation of such systems in real time. **Research methods:** The study uses machine learning methods, including regression and ensemble models, as well as experimental modeling based on time series. Accuracy metrics and analysis of the influence of input parameters are used to evaluate the efficiency. Data processing and storage are implemented using a hybrid infrastructure. **The scientific novelty of the study** is that the analysis revealed significant prospects for integrating high-precision machine learning models with hybrid information warehouses to create

ВИСНОВКИ

У ході кваліфікаційної роботи проведено комплексне дослідження сучасних архітектур та моделей збереження великих обсягів даних у гібридних сховищах, які об'єднують локальні та хмарні компоненти.

Проаналізовано еволюцію підходів — від класичних централізованих DWH до гібридних моделей, що включають Data Lakehouse, Data Mesh і Data Fabric. Розглянуто інтеграцію структурованих, напівструктурованих та неструктурованих даних в умовах стрімкого зростання інформаційних обсягів.

Запропоновано адаптивну модель ADH-DVL, що поєднує концепції Lakehouse, Data Vault 2.0, стрімінгову обробку та розподілене зберігання. Математична модель враховує чутливість даних, частоту доступу та політики локалізації, оптимізуючи баланс між безпекою, продуктивністю й витратами. Числові моделювання і графічні дослідження підтвердили ефективність моделі в управлінні гібридними ресурсами. Розроблена концепція може бути використана для створення промислових рішень у фінансових, медичних, промислових та державних інформаційних системах, де критично важлива ефективність гібридної архітектури.

За темою роботи опубліковано тези доповіді в рамках міжнародної науково-практичної конференції здобувачів вищої освіти і молодих вчених "Інформаційні технології у сучасному світі" та підготовлено наукову статтю до наукового журналу «Інформаційні технології та суспільство».

ДОДАТОК Б

Наукові публікації за темою кваліфікаційної роботи



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ БІОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ КІБЕРПОРТ
КАФЕДРА АВТОМАТИЗАЦІЇ ТА КОМП'ЮТЕРНО-ІНТЕГРОВАНИХ ТЕХНОЛОГІЙ



ЗБІРНИК МАТЕРІАЛІВ

МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
ЗДОБУВАЧІВ ВИЩОЇ ОСВІТИ І МОЛОДИХ УЧЕНИХ

**«ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ
В СУЧАСНОМУ СВІТІ»**

29 квітня 2025 р.
м. Харків

УДК 697.9:004.032.26:004.652

**МОДЕЛЬ ІНТЕЛЕКТУАЛЬНОЇ ПІДТРИМКИ МІКРОКЛІМАТУ
У ВЕЛИКИХ ПРИМІЩЕННЯХ ІЗ ВИКОРИСТАННЯМ
НЕЙРОМЕРЕЖЕВИХ ПІДХОДІВ ТА ГІБРИДНИХ СХОВИЩ
ІНФОРМАЦІЇ**

Піскаръов О.М., к.т.н., доц.
Абрамович Д.О., здобувач РВО магістр
Плугін В.В. здобувач РВО магістр
Харківський національний університет радіоелектроніки
м. Харків, Україна, oleksii.piskarov@pure.ua

Анотація. У статті розглянуто модель інтелектуальної системи підтримки мікроклімату у великих приміщеннях, що поєднує алгоритми штучних нейронних мереж та гібридну архітектуру збереження даних. Запропоновано підхід до обробки та накопичення великих обсягів сенсорної інформації у системах моніторингу мікроклімату, що дозволяє забезпечити як високу точність прогнозування, так і надійне збереження даних. Архітектура системи враховує специфіку динамічного середовища та дозволяє адаптуватися до змінних умов, оптимізуючи керування мікрокліматом. Модель реалізує інтеграцію хмарних та локальних сховищ з нейромережею, що навчається на історичних даних та забезпечує гнучке регулювання мікрокліматичних параметрів.

Ключові слова: мікроклімат, нейронна мережа, гібридне сховище, інтелектуальна система, прогнозування, оптимізація

Автоматизація керування мікрокліматом у великих приміщеннях – одне з пріоритетних завдань сучасного техногенного середовища, оскільки забезпечення стабільних умов сприяє як підвищенню комфорту, так і ефективному енергоспоживанню. Системи моніторингу та регулювання потребують обробки значних масивів даних, зокрема з різноманітних сенсорів, що постійно фіксують змінні параметри. Водночас ці дані мають бути збережені у надійний спосіб із можливістю швидкого доступу для навчання моделей прогнозування. Використання гібридного сховища дозволяє поєднувати переваги локальних і хмарних рішень, а штучні нейромережі — ефективно працювати з неструктурованими та нелінійними залежностями, характерними для процесів формування мікроклімату.

Метою роботи є створення адаптивного інтелектуального середовища, яке забезпечує комфорт, енергоефективність та можливість довготривалого зберігання історичних даних для аналітики. Загальна архітектура системи представлена на рис. 1. Вона включає модуль збору даних на основі мікроконтролера STM32F407, який зчитує сигнали з датчиків температури, вологості, CO₂ та освітлення. Дані передаються до гібридного сховища [1], що складається з локальної та хмарної частин, звідки надходять у нейромережевий модуль для прогнозування параметрів мікроклімату. На основі прогнозу система адаптивного керування формує керуючі сигнали для виконавчих механізмів, забезпечуючи оптимальні умови в приміщенні відповідно до

цільових показників комфортності. Комунікація між модулями реалізована за допомогою стандартних цифрових протоколів та API.



Рис. 1. Структурна схема системи підтримки мікроклімату

Серцем системи є мікроконтролер STM32F407, що виконує зчитування даних з цифрових сенсорів температури, вологості, рівня CO₂, освітленості та ін. Дані передаються через протоколи I²C, SPI та UART. Обробка інформації на нижчому рівні (відбір, первинна фільтрація, агрегація) відбувається безпосередньо на рівні контролера. У системі реалізовано апаратну підтримку адаптивного керування на основі моделі PMV (Predicted Mean Vote), що дозволяє кількісно оцінити тепловий комфорт користувача на основі фізіологічних параметрів та параметрів середовища [2].

Уся отримана інформація далі передається на центральний сервер, де реалізовано гібридне сховище даних: частина даних розміщується локально в базі даних (наприклад, SQLite або PostgreSQL), решта синхронізується з хмарними сервісами (наприклад, Amazon S3 або Firebase). Така архітектура забезпечує високу доступність, надійність та масштабованість системи. Дані останніх двох діб зберігаються в RAM-диску для швидкого доступу нейронної мережі, а архівні дані – в хмарі з підтримкою резервного копіювання та реплікації.

Для прогнозування параметрів мікроклімату у системі застосовано штучну нейронну мережу типу LSTM (Long Short-Term Memory), що враховує тимчасову залежність у даних. Навчання мережі проводиться на історичних даних, отриманих із гібридного сховища. Мережева модель навчена на основі реальних даних з експериментального тепличного комплексу та офісного середовища. При цьому були враховані типові сценарії змін: прихід людей у приміщення, денні коливання температури, відкриття вікон, зміни погоди. Завдяки використанню методу скользячого вікна та адаптивного градієнта, модель здатна оперативно адаптуватися до нових умов і продовжити навчання без повного перезапуску.

Регулювання здійснюється автоматично за принципом зворотного зв'язку: система порівнює прогнозований стан з бажаним (визначеним через

Секція 2. Інтеграція ІТ у промисловості: сучасні тенденції

PMV або вручну заданим), розраховує відхилення та коригує режими роботи пристроїв у режимі реального часу. Це дозволяє уникати перевитрати енергії та забезпечити стабільність мікроклімату навіть при різких зовнішніх змінах [3].

Запропонована модель інтелектуальної системи підтримки мікроклімату у великих приміщеннях демонструє ефективне поєднання технологій штучного інтелекту, гібридного збереження даних та мікроконтролерного керування. В основі системи лежить використання нейромережових алгоритмів (LSTM) для прогнозування зміни параметрів середовища, що дозволяє здійснювати адаптивне регулювання у режимі реального часу. Інтеграція моделі PMV забезпечує можливість кількісного оцінювання теплового комфорту, а застосування гібридного сховища підвищує надійність і масштабованість архітектури, що особливо важливо при обробці великих обсягів сенсорної інформації [4].

Практична значущість запропонованої системи полягає у можливості її впровадження у різноманітні галузі: офісні простори, тепличні комплекси, логістичні та виробничі центри, навчальні заклади. Використання адаптивного інтелектуального керування дає змогу не лише підвищити комфортність середовища, але й суттєво знизити енергоспоживання за рахунок точного прогнозування та своєчасного втручання у керуючі дії.

Перспективи розвитку системи включають розширення функціональності шляхом інтеграції з іншими елементами "розумного будинку" або "розумного виробництва", зокрема, з системами безпеки, розпізнавання присутності, адаптивного освітлення тощо. Крім того, передбачається вивчення можливостей використання більш складних нейромережових архітектур, таких як трансформери, для глибшого аналізу причинно-наслідкових залежностей між параметрами середовища [5].

Загалом, представлена система закладає підґрунтя для створення нових поколінь інтелектуальних керуючих комплексів, що здатні ефективно функціонувати в умовах швидких змін середовища та великого обсягу даних, зберігаючи стабільність, адаптивність і масштабованість.

Список літератури

1. Пахомов І.І., Єршов С.В. Системи збереження великих даних: архітектура та моделі // Вісник НТУУ "КПІ". Серія: Комп'ютерні науки. – 2021. – № 2. – С. 45–53.
2. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on CVPR. – 2016. – P. 770–778.
3. Тараненко В.С., Костенко І.М. Нейронні мережі та їх застосування в автоматизованих системах керування // Автоматизація процесів керування. – 2020. – № 3. – С. 12–18.
4. Zhang Y., Zheng Y., Qi D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction // Proceedings of AAAI. – 2017.
5. Власов М.М., Козловський В.М. Хмарні та гібридні технології збереження інформації // Системи обробки інформації. – 2022. – № 1. – С. 89–94.



вул. Інглезі, 6/1,
м. Одеса, Україна, 65101
www.helvetica.ua
mailbox@helvetica.ua

Стационар: 048 709 38 69
Vodafone: 095 934 48 28
Kyivstar: 097 723 06 08

ДОВІДКА

Видавничий дім «Гельветика» за домовленістю з Міжрегіональною Академією управління персоналом є офіційним видавцем журналу «Інформаційні технології та суспільство» та займається усіма видавничо-поліграфічними процесами, до яких належить: набір статей до чергового випуску; рецензування; перевірка на плагіат; коректорська вичитка; верстка; присвоєння кожному матеріалу РОЇ; розміщення електронної версії видання на офіційному сайті журналу; надсилання електронної версії видання до Національної бібліотеки України імені В. І. Вернадського на репозитарне зберігання та представлення на порталі в інформаційному ресурсі «Наукова періодика України»; розсилка обов'язкового безоплатного примірника до наукових установ України.

Цією довідкою повідомляємо, що наукова стаття Піскарьова Олексія, Абрамовича Данило, Пругіна Владіслава «Застосування нейромереж та гібридних сховищ при створенні комп'ютерних систем» прийнята редакцією журналу «Інформаційні технології та суспільство» для розміщення у Випуску 2 (17) за 2025 рік.

Директор
Видавничого дому «Гельветика»



Олег ГОЛОВКО

ДОДАТОК В

Фрагмент робочого коду, що демонструє застосування моделі на базі
ADH-DVL у реальній системі зберігання даних

```

# Імітація роботи системи збереження
class HybridStorageSystem:
    def __init__(self, total_capacity_tb, local_capacity_tb):
        self.total_capacity = total_capacity_tb
        self.local_capacity = local_capacity_tb
        self.local_used = 0
        self.cloud_used = 0
    def adaptive_store(self, data_size_tb, alpha, fA, S):
        # Розрахунок рекомендованого розміщення даних
        VL = alpha * data_size_tb * (fA + S)
        if VL > data_size_tb:
            VL = data_size_tb

        # Перевірка наявності вільного місця локально
        if self.local_used + VL <= self.local_capacity:
            self.local_used += VL
            self.cloud_used += (data_size_tb - VL)
            print(f"Дані обсягом {data_size_tb} ТБ збережено:
{VL} ТБ локально, {data_size_tb - VL} ТБ у хмарі.")
        else:
            # Якщо локального місця недостатньо – все
            # відправляється в хмару
            self.cloud_used += data_size_tb
            print(f"Недостатньо локального місця. Дані обсягом
{data_size_tb} ТБ збережено повністю у хмарі.")

    def status(self):
        print(f"Використано локально: {self.local_used} ТБ з
{self.local_capacity} ТБ.")
        print(f"Використано у хмарі: {self.cloud_used} ТБ.")
        print(f"Загальне збережене: {self.local_used +
self.cloud_used} ТБ з {self.total_capacity} ТБ.\n")

# Ініціалізація гібридної системи:
total_storage = 500 # Загальний обсяг системи, наприклад 500 ТБ
local_storage = 200 # Місце локально: 200 ТБ
system = HybridStorageSystem(total_storage, local_storage)

# Нові надходження даних:
system.adaptive_store(data_size_tb=50, alpha=0.8, fA=1.0, S=0.9)
system.adaptive_store(data_size_tb=100, alpha=0.5, fA=0.3,
S=0.4)
system.adaptive_store(data_size_tb=150, alpha=0.3, fA=0.2,
S=0.1)
system.status()

```