

МАГІСТЕРСЬКА РОБОТА

МР. ШМ - 06.00.00.000 ПЗ

Група ШМ-22-1

Вербовий Вадим

2024

Івано-Франківський національний технічний університет нафти і газу

Інститут інформаційних технологій

Кафедра інженерії програмного забезпечення

Вербовий Вадим Віталійович

(прізвище, ім'я, по батькові)

УДК 004.942
(індекс)

МАГІСТЕРСЬКА РОБОТА

Моделі, методи та алгоритми розробки інформаційно-програмних

засобів інтелектуальної підтримки прийняття рішень

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Вербовий В.В.

(підпис, ініціали та прізвище здобувача освітнього ступеня)

Науковий керівник **Яцишин Микола Миколайович, к.т.н., доцент**

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту

В.о. завідувача кафедри

доц.

Бандура В.В.

(посада) (підпис) (дата) (ініціали та прізвище)

Рецензент

доц.

(посада) (підпис) (дата) (ініціали та прізвище)

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

Івано-Франківськ – 2024

Івано-Франківський національний технічний університет нафти і газу

Інститут інформаційних технологій

Кафедра інженерії програмного забезпечення

Освітній рівень магістр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

В.о. зав. кафедрою ІІЗ

доц. В.В. Бандура

“ 04 ” вересня 2023 р.

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Вербовому Вадиму Віталійовичу

(прізвище, ім'я, по-батькові)

1. Тема магістерської роботи “ Моделі, методи та алгоритми розробки інформаційно-програмних засобів інтелектуальної підтримки прийняття рішень ”

керівник проекту (роботи) Яцишин Микола Миколайович, к.т.н., доцент

затверджені наказом закладу вищої освіти від “ 18 ” грудня 2023 р. № 738/7

2. Строк подання студентом проекту (роботи) 15 січня 2024 р.

3. Вихідні дані до проекту (роботи) Теоретичні концепції та формальні моделі побудови та функціонування інформаційних СППР

4. Зміст розрахунково - пояснювальної записки(перелік питань, які потрібно розробити)

1. Аналіз сучасних систем підтримки прийняття рішень та технологій обробки даних

2. Дослідження моделей та алгоритмів інтелектуального аналізу даних

3. Реалізація системи підтримки прийняття рішень на основі технології OLAP

4. Імітаційне моделювання основних характеристик системи підтримки прийняття рішень

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

1. Структурна схема СППР (рис. 1.1)

2. Типи закономірностей, які дозволяють виявляти методи Data Mining (рис. 1.2)

3. Методи інтелектуального аналізу (рис. 1.3)

4. Ознаки класифікації IC (рис. 1.4)

5. Семантичний шар OLAP клієнтів (рис. 1.5)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата
Нормоконтроль	доц., к.т.н. Вовк Р.Б.	
Перевірка на плагіат	доц., к.т.н. Вовк Р.Б.	

7. Дата видачі завдання 04 вересня 2023 р.

Керівник _____

(підпис)

Завдання прийняв до виконання _____

(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назви етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1	Підбір і вивчення літератури по темі магістерської роботи	01.10.2023	виконано
2	Аналіз концепцій та алгоритмів предметної області	25.10.2023	виконано
3	Аналіз сучасних систем підтримки прийняття рішень та технологій обробки даних	10.11.2023	виконано
4	Дослідження моделей та алгоритмів інтелектуального аналізу даних	22.11.2023	виконано
5	Реалізація системи підтримки прийняття рішень на основі технології OLAP	01.12.2023	виконано
6	Реалізація функціональності запропонованої інформаційної технології	15.12.2023	виконано
7	Затвердження пояснювальної записки роботи завідувачем кафедри	15.01.2024	виконано

Студент – магістр _____

(підпис)

Керівник роботи _____

(підпис)

АНОТАЦІЯ

Магістерська робота: 99 с., 30 рис., 15 табл., 43 джерела.

Тема: Моделі, методи та алгоритми розробки інформаційно-програмних засобів інтелектуальної підтримки прийняття рішень

Об'єкт дослідження: існуючі інформаційні реєстрові системи та системи з багатовимірною структурою представлення даних.

Мета роботи: дослідження процесу підвищення ефективності проведення аналізу накопичених даних та вирішення найбільш актуальних завдань предметних галузей, шляхом застосування моделей та методів аналізу багатовимірних структур даних та інтелектуального аналізу даних.

Предмет дослідження: моделі та методи інтелектуального аналізу багатовимірних даних інформаційних систем.

Результати дослідження:

Виконано вдосконалення багатовимірної моделі та методу агрегації, оперативного аналізу даних, що розповсюджуються на інформаційні системи, які оперують багатовимірними структурами даних та дозволяють оцінити достовірність агрегованих значень при здійсненні аналізу даних.

Висновок:

Запропоновано загальну архітектуру системи підтримки прийняття рішень, що створюється на основі реєстрових систем, визначено основні складові архітектури, використання яких забезпечить можливість функціонування систем у відповідності до їх призначення.

БАГАТОВИМІРНІ ДАНІ, СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ, ІНФОРМАЦІЙНА СИСТЕМА, МЕТОД АГРЕГАЦІЇ, СТРУКТУРА ДАНИХ, ТЕХНОЛОГІЯ OLAP.

ABSTRACT

Master Thesis: 99 pp., 30 fig., 15 tab., 43 sources.

Thesis Subject: Models, methods and algorithms for the development of information and software tools for intelligent decision-making support

Object of research: existing information register systems and systems with multidimensional data representation structure.

Research goal: to study the process of improving the efficiency of the analysis of accumulated data and solving the most pressing problems of subject areas, through the use of models and methods of analysis of multidimensional data structures and data mining.

Subject of research: models and methods of intellectual analysis of multidimensional data of information systems.

The results:

Improvements of the multidimensional model and method of aggregation, operative analysis of data that apply to information systems that operate with multidimensional data structures and allow to assess the reliability of aggregated values in data analysis.

Conclusion:

The general architecture of the decision support system created on the basis of registered systems is offered, the basic components of architecture which use will provide a possibility of functioning of systems according to their purpose are defined.

MULTIDIMENSIONAL DATA, DECISION SUPPORT SYSTEM, INFORMATION SYSTEM, AGGREGATION METHOD, DATA STRUCTURE, TECHNOLOGY.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ.....	9
ВСТУП.....	10
РОЗДІЛ 1.	
АНАЛІЗ СУЧАСНИХ СИСТЕМ, ІНФОРМАЦІЙНО-ПРОГРАМНИХ ЗАСОБІВ ТА ТЕХНОЛОГІЙ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ	14
1.1 Огляд технологій аналізу даних.....	14
1.2 Класифікація та аналіз інформаційних систем.....	24
1.3 Концепція побудови систем підтримки прийняття рішень на основі реєстрових систем.....	34
Висновки до розділу	36
РОЗДІЛ 2.	
ДОСЛІДЖЕННЯ МОДЕЛЕЙ ТА АЛГОРИТМІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ	38
2.1 Технологія проведення інтелектуального аналізу даних.....	38
2.2 Моделі та методи прогнозування виникнення критичних ситуацій	41
2.3 Моделі та методи розрахунку прогнозованих значень критичних показників об’єктів	54
2.4 Взаємодія технологій OLAP та Data Mining.....	57
2.5 Формалізований опис операцій багатовимірної моделі даних	59
Висновки до розділу	66
РОЗДІЛ 3.	
РЕАЛІЗАЦІЯ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ ТЕХНОЛОГІЇ OLAP	67
3.1 Імітаційне моделювання основних характеристик системи підтримки прийняття рішень	67

3.2 Архітектура системи та компонентів.....	84
Висновки до розділу	94
ВИСНОВКИ	95
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	96

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

СППР – систем підтримки прийняття рішень

СКБД – система керування базами даних

ОС – операційна система

ІС – інформаційна система

OLAP – online analytical processing – аналітична обробка у реальному часі

Data Mining – інтелектуальний аналіз

ВСТУП

Актуальність теми.

Ефективне управління організацією, підприємством, галуззю неможливо без прийняття виважених, обґрунтованих рішень. Прийняття даних рішень вимагає від керівників та управлінців проведення аналізу великих об'ємів інформації, як правило, в умовах істотних обмежень у часі. Проведення аналізу великих об'ємів даних є складним та неефективним процесом без застосування відповідних моделей та методів аналізу даних, найбільш ефективними з яких є оперативний аналіз (OLAP) та інтелектуальний аналіз (Data Mining).

Задача розробки моделей та методів аналізу з метою прийняття управлінських рішень, виникла при створенні та експлуатації комплексів технологічно – програмних засобів, що призначені для вирішення завдань автоматизації збору, збереження та обробки даних.

Натомість, експлуатація реєстрових систем та покладена у їх основу реляційна структура даних, показала неможливість ефективного вирішення задач аналізу накопичених об'ємів даних та, на його основі, прийняття управлінських рішень.

Результати проведеного аналізу можливих шляхів вирішення завдань поставлених перед реєстрами, сформувавши необхідність у побудові багатовимірних моделей даних, що доповнять існуючу реляційну структуру даних, підвищать ефективність та розширять границі аналізу, шляхом застосування методів оперативного аналізу даних (OLAP) та формування методів інтелектуального аналізу даних (Data Mining), які нададуть можливість знаходження закономірностей у накопичених великих об'ємах даних, з метою вирішення наявних актуальних задач.

Вибір багатовимірної моделі даних був обумовлений вимогами до можливостей та продуктивності систем підтримки прийняття рішень (СППР), основними з яких є забезпечення оперативності отримання необхідної

інформації при проведенні аналізу великих об'ємів даних. Разом з тим, застосування даного підходу виявило ряд принципових проблем, що виникають у разі недостатньої кількості вихідних даних (первинних даних) та потребують вирішення. Експлуатація означених реєстрових систем показує, що часто, у силу різноманітних обставин, внесена інформація не відповідає у повній мірі заданим вимогам. Задачі підвищення ефективності управління на основі аналізу інформації, що не у повній мірі визначена, виникають у різноманітних сферах діяльності: у економіці, екології, сільському господарстві, тощо. Характерною особливістю таких задач є те, що для їх вирішення та прийняття певного управлінського рішення, вплив наявності повної інформації не є критично важливим. Натомість, при розробці програмних продуктів, призначених для рішення таких задач, факт невідповідності вимогам вихідних даних має принципове значення. Застосування багатовимірних моделей даних у випадку невідповідності інформації призводить до виникнення ряду принципових проблем, що суттєво знижують ефективність використання вихідної інформації. Для вирішення даних проблем необхідна розробка методів аналізу багатовимірної структури даних з розрідженою інформацією. У ряді наукових праць відомих фахівців розглядаються теоретичні основи реляційних структур даних з розрідженою інформацією, однак єдиного підходу до технічної реалізації СППР на основі розріджених, багатовимірних структур даних з неповною інформацією на даний момент не сформовано. Таким чином, розробка ефективних моделей та методів аналізу даних багатовимірних, розріджених структур даних є актуальною задачею.

Використання для проведення аналізу готових засобів інтелектуального аналізу, що представлені на ринку програмного забезпечення, не є ефективним рішенням з огляду на: високу вартість продуктів даного сегменту, кількість рівнів застосування систем; надмірну функціональність, що не буде повністю використана; потребу у доведенні продуктів під потреби конкретних предметних галузей; неможливість розширення або модифікації ринкових

продуктів; відсутність готових рішень; необхідність практично єдиного технологічного рішення для обох СППР.

Мета та задачі досліджень.

Метою роботи є дослідження процесу підвищення ефективності проведення аналізу накопичених даних та вирішення найбільш актуальних завдань предметних галузей, шляхом застосування моделей та методів аналізу багатовимірних структур даних та інтелектуального аналізу даних.

У магістерській роботі вирішуються наступні основні задачі:

- дослідження сучасних систем підтримки прийняття рішень на основі даних, технологій аналізу даних, їх складових та тенденцій їх розвитку;
- опис багатовимірних моделей на основі структур даних інформаційних систем;
- дослідження багатовимірної моделі та методу агрегації, оперативного аналізу даних, що надає можливість оцінити рівень достовірності отриманих значень;
- опис моделей та методів інтелектуального аналізу структур даних.

Об'єктом дослідження є існуючі інформаційні реєстрові системи та системи з багатовимірною структурою представлення даних.

Предметом дослідження є моделі та методи інтелектуального аналізу багатовимірних даних інформаційних систем.

Методи дослідження. В процесі досліджень були використані методи проектування багатовимірних баз даних, теорія множин та графів, методи системного аналізу, методи інтелектуального аналізу, методи статистичного аналізу, теорія ймовірності.

Наукова новизна отриманих результатів полягає у виконанні вдосконалення багатовимірної моделі та методу агрегації, оперативного аналізу даних, що розповсюджуються на інформаційні системи, які оперують багатовимірними структурами даних та дозволяють оцінити достовірність агрегованих значень при здійсненні аналізу даних.

Практичне значення отриманих результатів виконаної роботи полягає розробці загальної архітектури системи підтримки прийняття рішень, що створюються на основі реєстрових систем, визначені основні складові архітектури, використання яких забезпечить можливість функціонування систем у відповідності до їх призначення.

Структура та обсяг роботи. Магістерська робота складається зі вступу, трьох розділів, висновку та списку використаної літератури. Робота викладена на 99 сторінках машинописного тексту, містить 30 ілюстрацій, 15 таблиць, список використаних літературних джерел із 43 найменувань.

РОЗДІЛ 1.

АНАЛІЗ СУЧАСНИХ СИСТЕМ, ІНФОРМАЦІЙНО-ПРОГРАМНИХ ЗАСОБІВ ТА ТЕХНОЛОГІЙ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

1.1 Огляд технологій аналізу даних

Розглянемо основні принципи, властивості та складові сучасних СППР, з метою визначення напрямків та принципів побудови СППР.

1.1.1 Системи підтримки прийняття рішень

Розвинутий механізм керування транзакціями у сучасних СКБД зробив їх основним засобом побудови так званих систем OLTP (оперативна обробка транзакцій). Метою створення OLTP системи - є автоматизація інформаційної підтримки бізнес – процесів, а саме, обробки та збереження інформації, що пов'язана з діяльністю організації, підприємства, галузі, тощо. Прикладами таких систем є “1С-Бухгалтерія”, різноманітні реєстри та реєстрові системи різних предметних галузей, тощо.

Системи OLTP, як правило, будуються на основі реляційних СКБД та включають у себе підсистеми збору, зберігання та інформаційно – пошукового аналізу інформації, а також містять попередньо визначену множину запитів для повсякденної роботи. Кожен новий запит, невизначений при проектуванні такої системи, вимагає втручання спеціаліста (програміста), що є неприпустимим в умовах необхідності виконання нерегламентованих запитів і оперативного прийняття рішення.

Ці та інші обмеження, характерні для систем OLTP, призвели до формування нового напрямку у розвитку інформаційних систем – систем підтримки прийняття рішень (СППР).

Основна мета створення систем підтримки прийняття рішень – надати користувачу інструмент для виконання ефективного аналізу даних. СППР

мають засоби внесення, збереження та аналізу даних, що відносяться до певної предметної галузі з метою пошуку рішення. Слід зазначити, що системи не гарантують правильність рішення, а лише представляють данні у вигляді звітів, таблиць, графіків, тощо – тобто інформацію, що дозволяє провести комплексну оцінку ситуації та забезпечує можливості проведення аналізу великих об'ємів даних в умовах обмеженого часу.

СППР можливо класифікувати як:

- СППР, на основі повідомлень (Communication-Driven DSS);
- СППР на основі даних (Data-Driven DSS);
- СППР на основі документів (Document-Driven DSS);
- СППР на основі знань (Knowledge-Driven DSS);
- СППР на основі моделей (Model-Driven DSS).

При подальшому розгляді у роботі буде розглядатись підтип напрямку СППР - СППР на основі даних, тому далі під терміном СППР мається на увазі саме даний підтип. СППР основані на забезпеченні автоматизованого вирішення наступних задач:

Узагальнення. Деталізовані дані представляються у агрегованому вигляді, дозволяючи провести глобальну оцінку ситуації, виявити загальні закономірності процесу, що досліджується.

Візуалізація. Для підвищення ефективності сприйняття інформації використовуються різноманітні засоби графічної візуалізації, такі як колові та стовпчасті діаграми, графіки, тощо.

Моделювання. Моделювання припускає створення моделі реального процесу та дослідження її властивостей з метою пошуку відповіді на питання: Що у разі? При цьому розглядається поведінка моделі, що досліджується.

Прогнозування. Мабуть найбільш питома і у той же час найбільш трудоемна щодо створення, функція аналітичних систем. Засоби СППР дозволяють здійснити глибинний аналіз даних, з метою автоматизації виявлення скритих закономірностей, тенденцій, тощо.

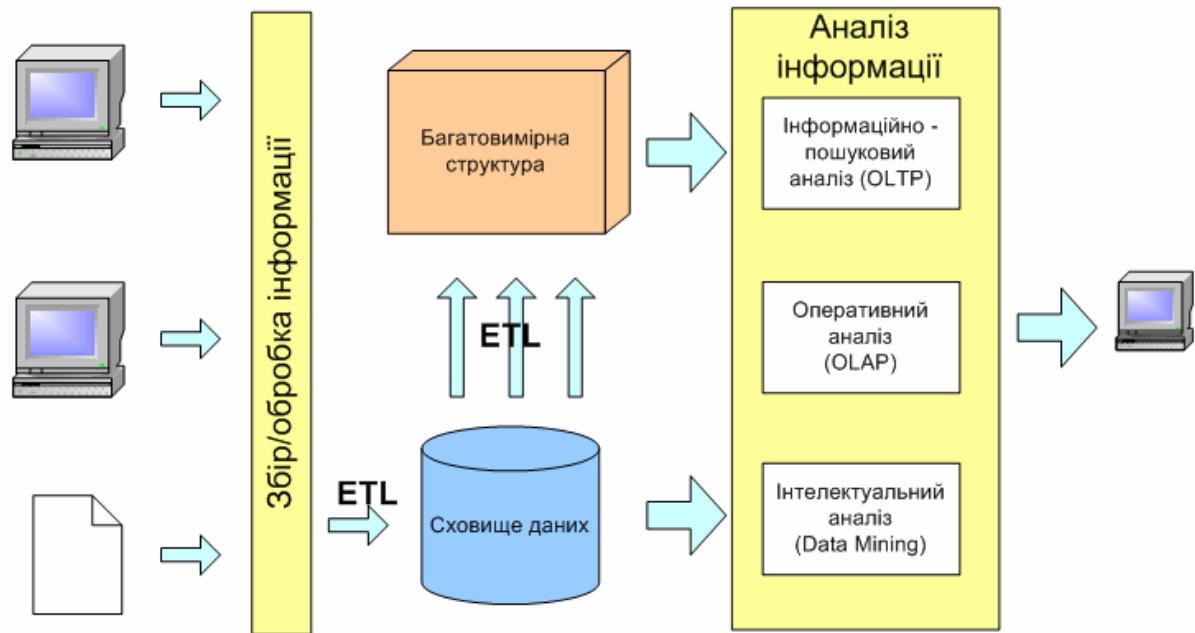


Рис. 1.1. Структурна схема СППР

Структурна схема СППР представлена на рис. 1.1. Основні компоненти СППР:

- *Сховище даних*. Сховища даних будують на основі реляційних структур даних. У процесі імпорту даних з облікових систем, вони проходять певну первинну обробку, що дозволяє забезпечити ефективність наступного аналізу: відбувається очищення даних, впорядкування по часу, агрегування, фільтрація, тощо. Іншим завданням сховищ даних є представлення різноманітних джерел інформації у єдиному інформаційному просторі, з яким зручно та ефективно працювати та маніпулювати.
- *ETL* (Extracting Transforming and Loading) – засоби та методи видобування, обробки та завантаження даних.
- *OLAP* – багатовимірна структура та методи обробки багатовимірної інформації.
- *Data Mining* – методи та засоби інтелектуального аналізу даних.

Таким чином, основними функціональними елементами СППР виступають компоненти OLAP та Data Mining. Розглянемо та проаналізуємо особливості даних технологій та сучасний стан їх розвитку.

Технологія OLAP

У 1993-му році вийшла стаття Е.Кодда, у якій він вперше визначив та використав термін OLAP. У статті було запропоновано 12 правил, яким повинен задовольняти програмний продукт класу OLAP. Данні правила, виступаючи, по суті, визначенням OLAP, виявились доволі складними для сприйняття, суперечливими та були у подальшому перероблені Найгелем Пендсом автором журналу OLAP Report у сукупність п'яти правил, що отримали назву тест FASMI (Fast Analysis of Shared Multidimensional Information – Швидкий Аналіз Розподіленої Багатовимірної Інформації).

Незважаючи на те, що спроби створення моделей представлення даних, відмінних від реляційної, здійснювались з 60-х років, саме стаття Е.Кодда вважається днем народження OLAP - оперативного аналізу даних на основі багатовимірної представлення. Саме багатовимірність представлення даних є ключовою особливістю OLAP та робить механізм отримання необхідних даних доступним та інтуїтивно зрозумілим, а можливість виконання нерегламентованих запитів та швидкість отримання даних дозволяє говорити о можливості проведення оперативного аналізу.

Поняття OLAP тісно пов'язане з концепцією сховищ даних (Data Warehouse). Сховище даних, це інтегрований накопичувач інформації, що зібрана з різноманітних джерел, у тому числі і систем OLTP. Сховища даних орієнтовані на бізнес поняття, а не на бізнес – процеси та містять усю інформацію, що відноситься до даних понять. Дана інформація збирається з оперативних систем за узгоджені періоди часу та не підлягає оперативним змінам. Данні у сховищі суттєво денормалізовані з метою підвищення продуктивності виконання запитів. Як правило, сховище даних організовано на основі реляційної структури по схемі «зірка» або «крижинка». Данні

завантажуються у сховище шляхом так званих ETL алгоритмів (Extraction, Transformation, Loading – видобування, перетворення, завантаження). Сховище даних використовується у якості безпосереднього «поставника» інформації для систем аналізу інформації на основі OLAP.

Сьогодні жодна з серйозних аналітичних систем не обходиться без застосування OLAP та сховищ даних. Сучасні системи підтримки прийняття рішень, засоби BI (Business Intelligence – аналіз та обробка даних масштабу підприємства), CRM (Customer Relationship Management – підвищення ефективності взаємодії з клієнтами), ERP (Enterprise Resource Planning – планування ресурсів підприємства) основані на OLAP чи містять його як компоненту.

Провідні корпорації – виробники ПЗ та СКБД: Oracle, Microsoft, Informix, Sybase, тощо, містять у складі своїх продуктів засоби побудови сховищ даних та аналітичних систем на основі OLAP. Однак при усьому різноманітті програмних продуктів, жоден з виробників не представляє закінченого рішення для конкретного підприємства чи галузі.

Резюмуючи огляд технології, слід відзначити, що OLAP система – це перш за все зручний та потужний інструмент аналізу даних. Зручність полягає у оперативності отримання необхідної інформації. Багатовимірне представлення даних у OLAP – системі є інтуїтивно зрозумілим для користувача та для отримання потрібної інформації, йому не потрібно вивчати структуру бази даних або залучати спеціаліста зі створення звітів. Потужність OLAP полягає у можливості отримання будь яких необхідних даних (як узагальнених так і деталізованих) на усіх етапах аналізу. Технологія OLAP, таким чином, представляє собою сукупність методів та засобів створення OLAP - систем.

Технологія Data Mining

Data Mining – процес виявлення у даних, раніше невідомих, нетривіальних, корисних на практиці та доступних для інтерпретації знань, необхідних для прийняття рішення у різноманітних сферах діяльності.

Виділяють п'ять стандартних типів закономірностей, що виявляються методами Data Mining (рис. 1.2).

Асоціація має місце у випадку, якщо декілька подій пов'язані одна з одною. Асоціація представляється у вигляді правил, що можуть бути використані як для кращого розуміння природи даних, що аналізуються так і для прогнозування подій.

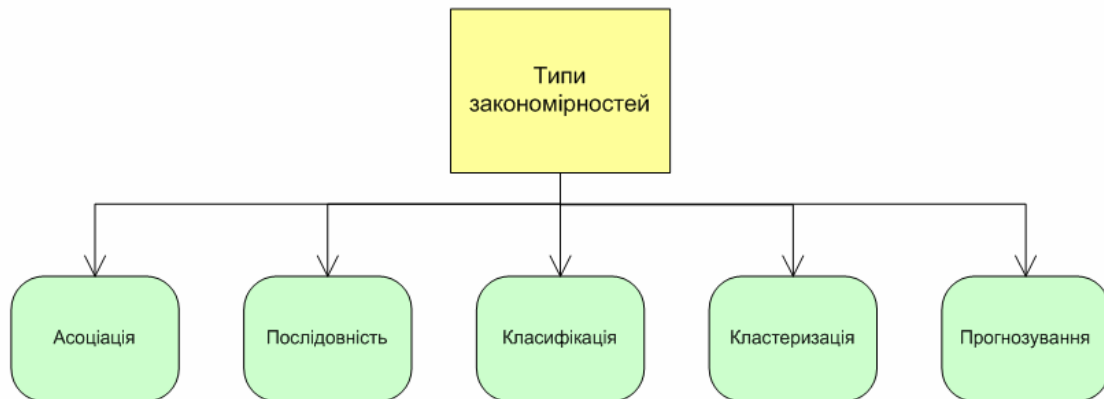


Рис. 1.2. Типи закономірностей, які дозволяють виявляти методи Data Mining

У разі існування ланцюга пов'язаних у часі подій, говорять про *послідовність*. Послідовність визначає закономірності між пов'язаними у часі подіями.

Класифікація – наявні властивості, що характеризують групу до якої належить подія чи об'єкт. Зазвичай при цьому, на основі аналізу вже класифікованих подій (об'єктів) - сформовані певні правила класифікації.

Кластеризація – закономірність, подібна до класифікації та відрізняється тим, що групи не задані заздалегідь, а визначаються у процесі обробки даних. Об'єкти, що належать до групи, повинні бути «схожі» за властивостями та відрізнятись від об'єктів інших груп.

Основою для різноманітних систем *прогнозування* виступає історична інформація. Шляхом побудови шаблонів, що адекватно відображають динаміку

поведінки цільових показників, можливо спрогнозувати поведінку моделі у майбутньому.

На сьогоднішній день існує велика кількість різноманітних методів аналізу даних, основні з яких зображені на рис. 1.3. Резюмуючи огляд технології необхідно відзначити, що сфера застосування Data Mining систем експотенційно збільшується. У даному розвитку приймають участь практично усі великі корпорації. Системи застосовуються по двом основним напрямкам: як масовий продукт для бізнес систем, як інструменти для проведення унікальних досліджень (генетика, хімія, медицина, тощо). Кількість інсталяцій масових продуктів сьогодні досягає десятків тисяч. Лідери пов'язують майбутнє даних систем у використанні їх у якості інтелектуальних додатків вбудованих у корпоративні сховища даних.

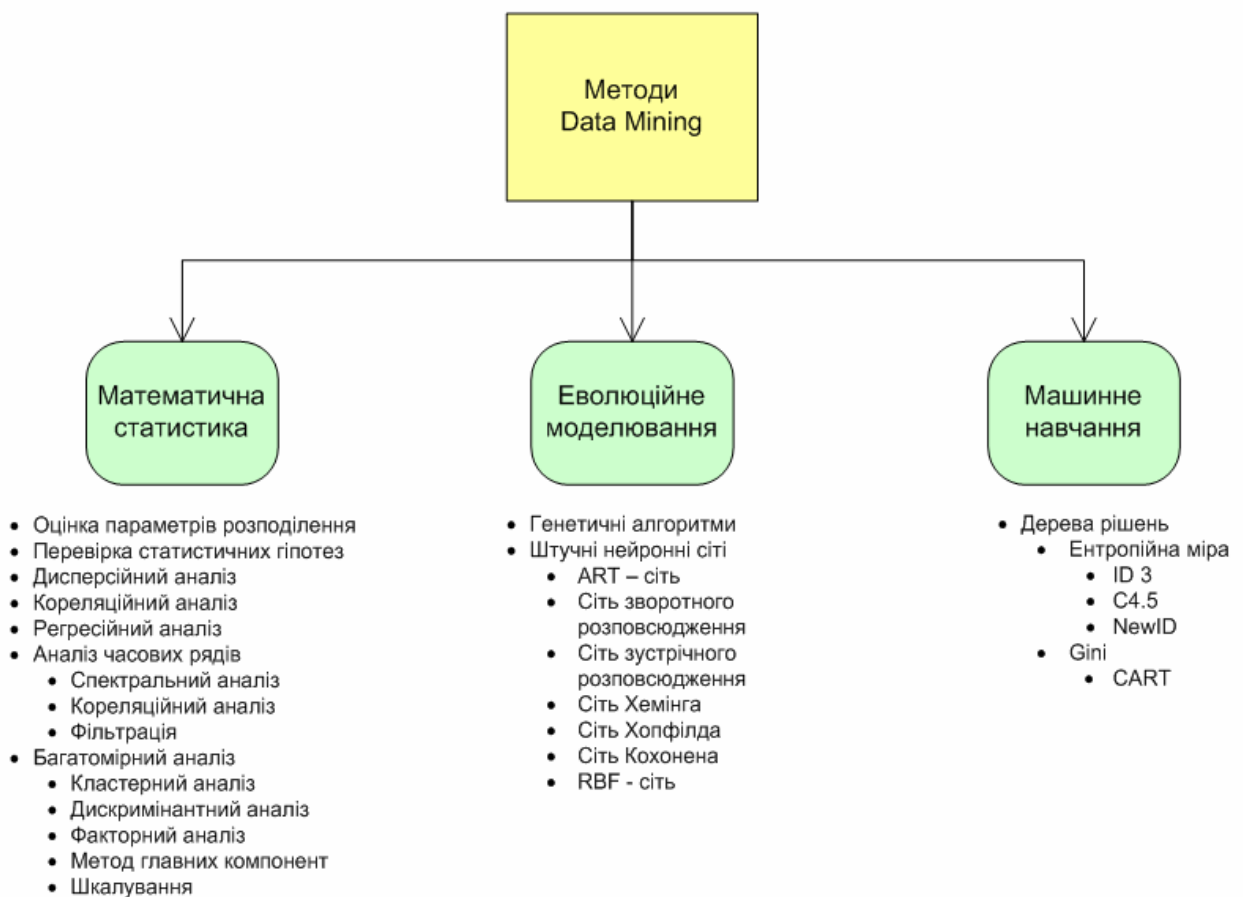


Рис. 1.3. Методи інтелектуального аналізу

Огляд та аналіз сучасних СППР на основі даних, представлених на ринку ПЗ

При створенні інформаційних систем, що орієнтовані на аналіз даних можливі різноманітні підходи. З однієї сторони, існуючі програмні продукти представляють собою налагоджені багатофункціональні системи, що дозволяють задовольнити більшість потреб кінцевого користувача. З іншого боку, за рахунок власної універсальності данні засоби мають надлишкову функціональність (що безумовно відображаються у вартості ПЗ), яка навряд буде використана користувачем. При цьому будь який засіб вимагає додаткових затрат пов'язаних з необхідністю «доведення» системи, з метою використання при вирішенні конкретних задач. Виходячи з вище сказаного, доцільним є підхід, що полягає у комбінованому використанні існуючих програмних продуктів, компонентів та створенні спеціалізованих інтерфейсів, що забезпечать необхідну функціональність.

Більш детально дане питання буде розглянуто нижче, але перед цим доцільно провести огляд сучасних систем, що містять засоби оперативного та інтелектуального аналізу, проаналізувати їх переваги та недоліки з метою можливого подальшого використання у складі систем, що створюються.

Після попереднього огляду існуючих на ринку систем, були виявлені наступні компанії – розробники: Targit, Next Action Technology, Cartesis, Hyperion Solutions, Cristal Decisions, Knosys, Cognos, Brio Technology, Sybase, Arbor Systems, SAS, IQ Software, MicroStrategy, Business Objects, Oracle, Informix, Microsoft, IBM, тощо. Зважаючи на їх кількість, доцільно виділити та розділити коло продуктів, що аналізуються, до продуктів найбільш відомих та лідируючих на ринку компаній, продуктів, що надають можливість створення та маніпулювання багатовимірними структурами, тобто OLAP серверів та продуктів, що мають засоби Data Mining.

До компаній, що відомі своїми досягненнями у галузі СКБД та мають у ряді власних продуктів OLAP сервера було віднесено: Microsoft, IBM, Oracle, Hyperion Solutions, Cognos (табл. 1.1).

Таблиця 1.1.

Поширені OLAP продукти

Компанія розробник	Назва OLAP продукту
IBM	DB2 OLAP Server
Microsoft	MS SQL Server Analysis Services
Oracle	Oracle xg (Oracle OLAP)
Hyperion Solutions	Hyperion Essbase
Cognos	Cognos PowerPlay

Виходячи з обраних для порівняння продуктів та завдань, було визначено критерії за якими можливо оцінити OLAP продукти та здійснено їх оцінку. Результати аналізу обраних OLAP продуктів відображені у табл. 1.2.

Таблиця 1.2.

Порівняння OLAP продуктів

Критерій/Продукт	MS SQL Server Analysis Services	DB2 OLAP Server	Oracle xg (Oracle OLAP)	Hyperion Essbase	Cognos PowerPlay
Підтримка ODBC протоколу	+	+	+	+	+
Підтримка доступу до нереляційних джерел інформації	+	-	+	+	-
Можливість зберігання інформації у режимах MOLAP, HOLAP	+	+	+	HOLAP	+
Тимчасове збереження багатовимірних даних на клієнтських робочих станціях	+	-	-	-	-
Можливість створення міток, що розраховуються	+	+	+	+	+

Сумісна робота з декількома кубами	+	-	+	-	-
Масштабування	+	+	+	+	+
Оптимізація схеми агрегування	+	+	+	-	+
Зручні засоби адміністрування	+	+	+	+	+
Вирішення проблеми «вибуху даних»	+	+	+	+	+
Підтримка доступу множини користувачів до OLAP- серверу	+	+	+	+	+
Можливість попереднього розрахунку агрегатів	+	+	+	+	+
Безпека на рівні комірки даних	+	-	+	-	-
Можливість роботи під керуванням різних операційних систем	-	+	+	-	-

Резюмуючи отримані результати аналізу слід відзначити, що функціональні можливості наведених OLAP - продуктів практично рівні, але продукти компаній Oracle та Microsoft все ж мають деяку перевагу.

Говорячи про Data Mining продукти, слід відзначити, що основне їх призначення це всебічний аналіз даних, моніторинг та звітність. Такі системи найчастіше «прив'язані» до певних бізнес процесів, мають потужні засоби аналізу, та візуалізації інформації. Вартість даних продуктів (одного робочого місця) дуже висока від 10-100 тисяч гривен, не враховуючи адаптацію для вирішення потрібних задач. Серед лідерів слід відзначити наступні компанії у відповідності до класів компонентів Data Mining:

- Статистичні пакети: SAS (SAS Institute), SPSS (SPSS), STATGRAPICS (Manugistics), STATISTICA, STADIA.
- Нейронні мережі: BrainMaker (CSS), NeuroShell (Ward Systems Group), OWL (HyperLogic).

- Дерева рішень: See5/C5.0 (RuleQuest), Clementine (Integral Solutions), SIPINA (University of Lyon), IDIS (Information Discovery), KnowledgeSeeker (ANGOSS).
- Генетичні алгоритми: GeneHunter (Ward Systems Group).
- Компанії, що пропонують комплексні рішення для комплексного аналізу даних: Oracle, Cognos, Business Objects, SAP, Microsoft, Hyperion Solutions, MicroStrategy.

Створення систем підтримки прийняття рішень завжди складається з етапів аналізу даних та процесів Замовника, проектування сховищ та багатовимірних структур даних з урахуванням його потреб та процесів. Кілька десятків фірм пропонують продукти, що здібні вирішувати ті чи інші задачі, що постають перед системою підтримки прийняття рішень. Сюди входять СКБД, засоби ETL, інструменти OLAP та Data Mining аналізу, тощо. На сьогоднішній день не існує визнаного лідера у галузі виробництва програмного забезпечення для побудови систем підтримки прийняття рішень. Жодна з компаній не пропонує готового рішення, що придатне до безпосереднього використання у процесах Замовника, без проведення попередньої адаптації.

1.2 Класифікація та аналіз інформаційних систем

Перед розглядом моделей та методів, що будуть використані для вирішення задач предметних галузей реєстрів певної предметної області, а також побудови на основі даних реєстрових систем - системи підтримки прийняття рішень, доцільно провести скорочену класифікацію та аналіз сучасних напрямків розвитку інформаційних систем (ІС). Класифікація інформаційних систем була проведена за наступними основними ознаками (рис. 1.4).

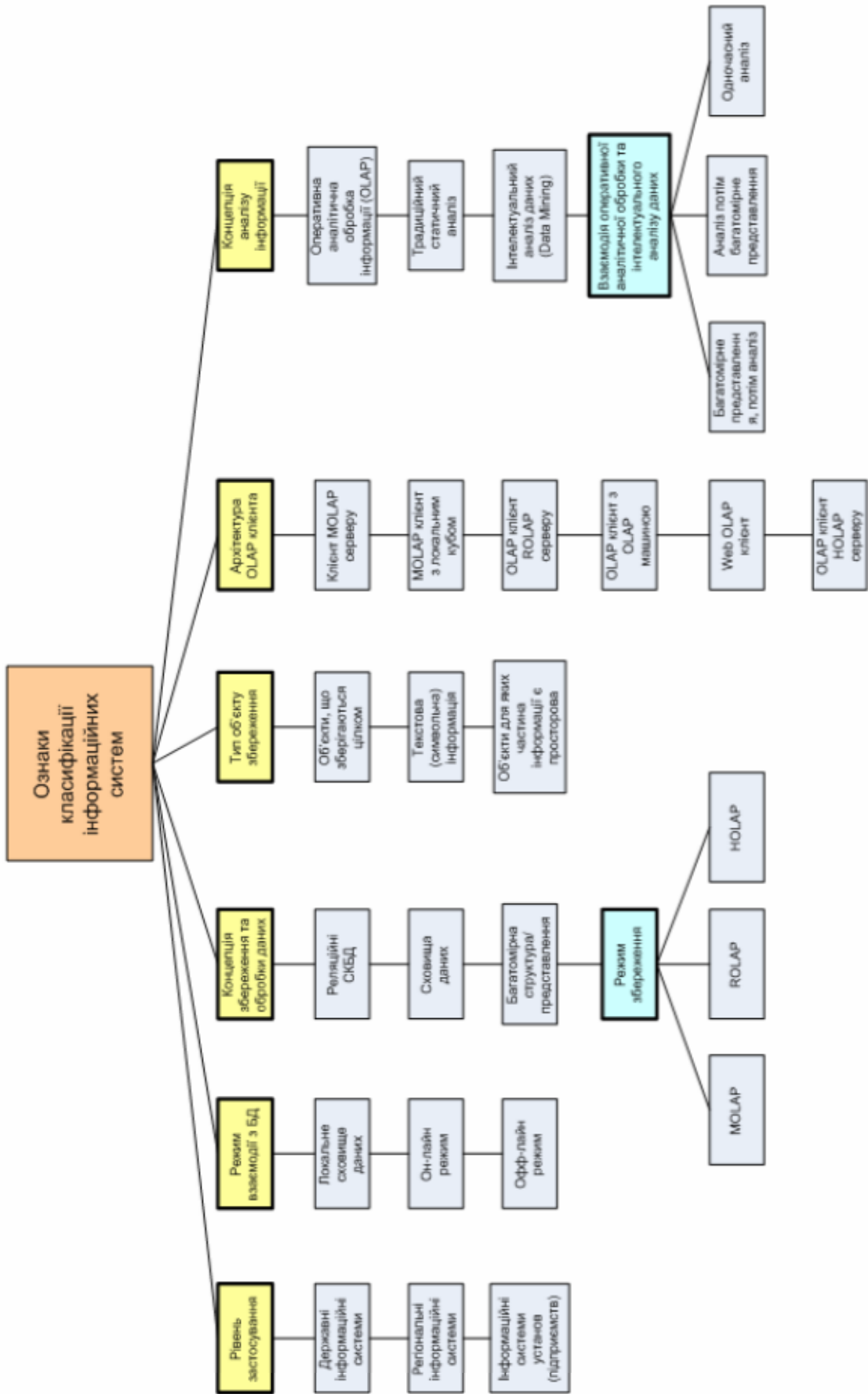


Рис. 1.4. Ознаки класифікації ІС

1.2.1. Рівень застосування

За рівнем застосування ІС можливо класифікувати як:

- *Державні інформаційні системи* - призначені для обліку та аналізу інформації, щодо найважливіших об'єктів країни. На базі використання обчислювальних комплексів ведеться облік та аналіз інформації щодо стратегічно важливих об'єктів та процесів країни, регулюється діяльність окремих ланцюгів практично в усіх галузях застосування. Основним призначенням та особливістю таких систем є те, що інформація у них несе юридичну значимість, тому дані системи повинні задовольняти вимогам щодо надійності збереження, безпеки та конфіденційності інформації у відповідності до вимог чинних стандартів у галузі інформаційних систем України. Архітектурно, данні системі найчастіше мають розподілену, територіально незалежну структуру з центральним сховищем даних, доступ до якого може відбуватись у постійному режимі. Прикладами реалізації та впровадження таких систем можуть виступати комплекси інформаційних систем, що впроваджені під керівництвом міністерства юстиції, міністерства аграрної політики, тощо.
- *Регіональні інформаційні системи* - призначені для обліку та аналізу інформації, щодо об'єктів та процесів обласних та районних адміністративно територіальних одиниць України. Ці системи вирішують задачі, що поставлені перед регіоном, формування звітності й видачі оперативних даних місцевим і керівним державним та господарським органам. Найчастіше бувають двох видів:
 - складові державних інформаційних систем;
 - незалежні регіональні інформаційні систем.
- *Інформаційні системи установ (підприємств)*. Основним завданням даних систем є ведення обліку та аналізу даних необхідних для функціонування відповідних установ та підприємств. Дані системи зберігають інформацію, що не відображає державних інтересів, та не підлягає загальному

державному обліку. Можуть мати різноманітну архітектуру, функціональність, склад інформації та методи доступу до неї. Часто данні системи мають розвинуті засоби аналізу даних та елементи систем підтримки прийняття рішень. Оскільки під цю категорію потрапляють як великі корпорації, так і невеликі установи, тому не можливо однозначно класифікувати дану категорію, щодо можливостей систем.

1.2.2 Режим взаємодії з БД

Основні режими взаємодії зі сховищем даних (БД) можливо класифікувати як:

- *Локальне сховище даних.* Найбільш просте архітектурне рішення, представляє собою від одного до множини робочих місць у рамках однієї локальної мережі. Не вимагає засобів зв'язку, додаткової техніки. Системи з таким режимом можливі для застосування лише на підприємствах або установах, що об'єднані у локальну мережу.

Серед позитивних моментів даного режиму взаємодії слід відзначити: проста, швидкість реалізації; відсутність потреби у каналах зв'язку; простота оновлення версії програмного продукту; швидкість передачі інформації.

До негативних моментів слід віднести: територіальна залежність, що не дозволяє вважати та застосовувати данні системи на державному рівні.

- *Офф-лайн (off-line) режим.* Побудова системи з офф-лайн режимом взаємодії дозволяє говорити про систему як про централізовану. Принцип роботи такої системи полягає у обміні реплікаційними пакетами даних.

Серед позитивних моментів даного режиму взаємодії слід відзначити: можливість реалізації територіально – незалежних інформаційних систем; можливість внесення інформації з багатьох робочих місць; можливість накопичення у центральному сховищі даних, значних об'ємів даних; можливість застосування методів інтелектуального аналізу та багатовимірною аналізу на центральному рівні систем.

До негативних моментів слід віднести: необхідність у каналах зв'язку; складність процесу оновлення клієнтського програмного забезпечення; складність забезпечення вимог безпеки для регіональних БД; низький рівень актуальності інформації; потреба у захисті інформації при передачі.

Облікова система з офф-лайн режимом взаємодії з БД є деяким проміжним варіантом між локальним сховищем даних та он-лайн режимом. Незважаючи на перелічені недоліки, системи основані на даному режимі взаємодії можуть використовуватись у якості державних облікових систем, при цьому містити засоби багатовимірної обробки даних та інтелектуального аналізу.

- *Он-лайн (on-line) режим.* Он-лайн режим взаємодії є найбільш сучасним, та з точки зору виконання основних операції обліку та аналізу інформації - забезпечує найбільші можливості, але є більш складним з точки зору розробки програмного забезпечення. Режим он-лайн гарантує постійний доступ з робочого місця до бази даних системи, при цьому забезпечуючи як внесення інформації так і отримання результатів у будь якому вигляді.

Серед позитивних моментів даного режиму взаємодії слід відзначити: централізоване збереження інформації; висока актуальність інформації; територіальна незалежність; можливість одночасної роботи множин клієнтів; гнучка система безпеки; можливість автоматичного оновлення клієнтського програмного забезпечення; єдине джерело зберігання та походження даних.

До негативних моментів слід віднести: необхідність у каналі зв'язку; складність розробки.

Незважаючи на існуючі недоліки, он-лайн режим є найбільш універсальним для застосування та може використовуватись для вирішення задач усіх типів, дозволяючи створення та впровадження як облікових системи, так і систем інтелектуальної підтримки прийняття управлінських рішень.

Тип об'єкту збереження

Необхідність ведення обліку та проведення аналізу великих об'ємів інформації стосується практично усіх об'єктів оточуючого середовища. Для підприємств це може бути предмет продажу, купівлі або деякі елементи ринку над якими необхідно проводити аналіз, тощо. На рівні держави це об'єкти, інформацією про які оперує держава: об'єкти нерухомого майна, об'єкти правових відносин, об'єкти сільського господарства, тощо.

З огляду на вище сказане - неможливо класифікувати ІС за конкретними об'єктами. Тому у якості ознак класифікації були визначені основні типи об'єктів:

- Інформаційні системи, що оперують об'єктами які **зберігаються цілком**. Інформаційні системи, даного класу оперують інформацією, де об'єкт переставляє собою аудіо, відео формат, сканований документ, тощо. Характеризуючи данні системи з точки зору аналізу інформації необхідно зазначити, що зазвичай ніякої мови про аналіз інформації тут не йдеться.
- Інформаційні системи, що оперують **просторовою інформацією**. Для даного класу інформаційних систем характерна важлива роль, яку відіграє місцезнаходження об'єктів, або їх взаємне розташування. Частина інформації даних систем – просторова інформація. Об'єкти такого типу найчастіше використовуються у державних ІС, що оперують окрім просторовою також великими об'ємами семантичної інформації. Системи, що оперують просторовою інформацією часто мають потужні засоби аналізу даних у яких провідну роль відіграє саме просторова інформація.
- Інформаційні системи, що оперують **текстовою (символьною) інформацією**. Інформаційні системи даного типу містять текстовий опис об'єкту, властивостей його поведінки, або описують процедури над об'єктом. Даний тип є найбільш поширеним серед інформаційних систем.

1.2.3 Концепція збереження та обробки даних

За концепцією збереження даних інформаційні системи можливо класифікувати як:

- Інформаційні системи у яких інформація зберігається у реляційних СКБД. Дані інформаційні системи спрямовані на облік інформації та, як правило, не містять засобів аналізу інформації.
- Інформаційні системи зі сховищем даних. До даного класу інформаційних систем відносяться системи, у яких реалізовані механізми роботи зі сховищем даних. Даний клас інформаційних систем дозволяє проводити інтелектуальний аналіз на основі даних сховища. Можливо стверджувати, що інформаційна система відповідає даній концепції збереження даних у разі відповідності наступним вимогам: предметна орієнтованість, інтегрованість, незмінність, підтримка хронології.
- Інформаційні системи з багатовимірною структурою/представленням інформації. До даного класу інформаційних систем відносяться системи які мають багатовимірну структуру або представлення даних та розроблені механізми роботи з нею. Дані інформаційні системи спрямовані на проведення обліку (при наявності облікової частини) та аналізу, що обумовлюється відповідною багатовимірною структурою/представленням даних, що зберігаються. Можливо стверджувати, що інформаційна система відповідає даній концепції збереження даних у разі відповідності або максимальної наближеності до принципів запропонованих Е.Коддом. За типом організації багатовимірної БД, ІС у свою чергу можливо класифікувати як:

- Системи з багатовимірним OLAP (MOLAP).
- Системи з реляційним OLAP (ROLAP).
- Системи з гібридним OLAP (HOLAP).

Завершуючи аналіз інформаційних систем згідно даної ознаки, слід розглянути відмінності у характеристиках даних систем (табл. 1.3), які визначаються концепцією збереження та обробки даних.

Таблиця 1.3.

Порівняння характеристик ІС з різною концепцією збереження та обробки даних

Характеристики	Облікові системи (Реляційні СКБД)	Сховища даних	Багатовимірні бази даних
<i>Типова операція</i>	Оновлення	Звіт	Аналіз
<i>Частота оновлення</i>	Висока невеликими об'ємами	Мала, великими об'ємами	Мала, дуже великими об'ємами
<i>Джерела інформації</i>	Внутрішні	Зовнішні	Зовнішні
<i>Рівень аналітичних вимог</i>	Низький	Середній	Високий
<i>Об'єм даних транзакції</i>	Невеликий	Від малого до великого	Великі
<i>Рівень інформації</i>	Детальна	Детальна та сумарна	Сумарна
<i>Строки зберігання</i>	Поточні	Історичні та поточні	Історичні, поточні та прогнозовані
<i>Структурні елементи</i>	Записи	Записи	Масиви

Таким чином, як видно з табл. 1.3, характеристики та задачі що вирішують системи з різними концепціями зберігання та обробки інформації доволі різні, але повноцінна СППР повинна об'єднувати усі перелічені концепції.

1.2.4 Концепція аналізу інформації

За концепцією аналізу даних інформаційні системи можливо класифікувати як:

- інформаційні системи, що здійснюють традиційний статичний аналіз. Найпоширеніший вид інформаційних систем, можливості аналізу інформації у яких покладено на традиційні механізми здійснення запитів до реляційних структур даних. Кожен новий запит у системах реалізованих на основі традиційних технологій статичного аналізу повинен бути спочатку

описаний, переданий програмісту, реалізований, і на кінець, виконаний. Таким чином, зовнішня простота систем, що використовують традиційний статичний аналіз обертається втратою гнучкості.

- інформаційні системи, що мають засоби оперативної аналітичної обробки інформації (OLAP). Оперативна аналітична обробка інформації, на відміну від традиційного статичного аналізу інформації, спрямована на обробку нерегламентованих запитів аналітиків до інформації. У основу інформаційних систем даного класу покладено принципи багатовимірного представлення та обробки інформації. В більшості публікацій під аббревіатурою OLAP мається на увазі не лише багатовимірний погляд на данні, але і концепція збереження самих даних, взагалі це не зовсім вірно оскільки реляційні бази даних є найбільш зручною технологією для збереження інформації, і якщо ми говоримо про недоліки статичного аналізу, то насамперед необхідно зазначати, що необхідність існує не у новій концепції збереження інформації, а скоріше у концепції аналізу даних.
- інформаційні системи, що мають засоби інтелектуального аналізу даних (Data Mining). Даний клас інформаційних систем призначений, та має відповідні засоби, для виявлення скритих закономірностей у великих масивах даних. Системи інтелектуального аналізу даних дають змогу особі, що приймає рішення, вирішувати задачі інтелектуального характеру, що вимагають аналітичної обробки великих обсягів, збереженої у базі даних інформації.
- інформаційні системи, що здійснюють оперативну аналітичну обробку інформації та інтелектуальний аналіз. Оперативна аналітична обробка та інтелектуальний аналіз даних – дві складові частини процесу підтримки прийняття рішень. Даний клас інформаційних систем об'єднує ці два напрямки аналізу, принципова схема якого зображена на рис. 1.1. Розрізняють декілька варіантів інтеграції:
 - *Багатовимірне представлення, потім аналіз.* Можливість виконання інтелектуального аналізу інформації повинна бути над кожним

результатом запиту до багатовимірної структури даних, тобто над будь-яким фрагментом будь-якого виміру багатовимірного кубу.

- *Аналіз, потім багатовимірне представлення.* Подібно даним, що витягнуті зі сховища – результати інтелектуального аналізу повинні бути представлені у вигляді, допустимому для наступного багатовимірного аналізу.
- *Одночасний аналіз.* Цей найбільш гнучкий спосіб інтеграції, що дозволяє автоматично активувати однотипні механізми інтелектуальної обробки над результатом кожного кроку багатовимірного аналізу.

1.2.5. Архітектура OLAP клієнта

Дана ознака визначає тип організації доступу та взаємодії клієнтського програмного забезпечення та серверу OLAP.

Загальний принцип роботи OLAP клієнтів спирається на здійсненні попереднього опису семантичного шару, що «ховає» фізичну схему даних від користувача, як це зображено на рис. 1.5 на прикладі структури реєстру в сільськогосподарській галузі. Після створення даного шару, користувач має можливість самостійно маніпулювати зрозумілими для нього об'єктами, використовуючи терміни предметної галузі.

Таким чином за архітектурою OLAP клієнтів інформаційні системи можливо класифікувати як:

- Інформаційні системи з клієнтом MOLAP серверу.
- Інформаційні системи на основі MOLAP клієнта з локальним кубом.
- Інформаційні системи на основі OLAP клієнта та ROLAP сервера.
- Інформаційні системи на основі OLAP клієнта та HOLAP сервера.
- Інформаційні системи на основі OLAP клієнта з OLAP машиною.
- Інформаційні системи з Web OLAP клієнтом.

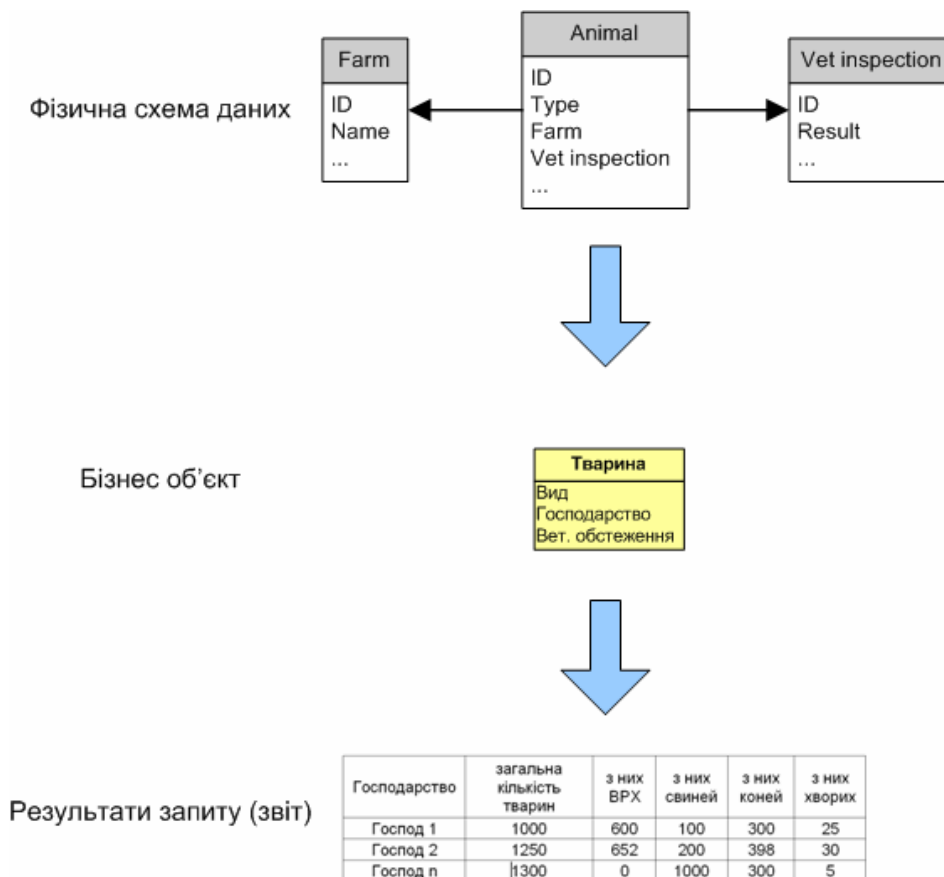


Рис. 1.5. Семантичний шар OLAP клієнтів

1.3 Концепція побудови систем підтримки прийняття рішень на основі реєстрових систем

Сформуємо концепцію (рис. 1.6) розвитку існуючих реєстрових систем, з метою доповнення їх засобами аналізу накопиченої інформації й, як наслідок, задоволення наявних потреб та проведемо аналіз кожного з компонентів концепції.

Облікова частина реалізованих систем залишається без змін, оскільки наявна реляційна структура даних та методи її обробки найбільш доцільні для вирішення задач обліку інформації.

У якості основи для здійснення аналізу інформації, пропонується використання сховища даних, яке буде містити сформовану інформацію у відповідності до вимог, що висуваються до сховищ даних. При цьому

інформація у сховищі буде зберігатись як у традиційній реляційній структурі так і у багатовимірній структурі.

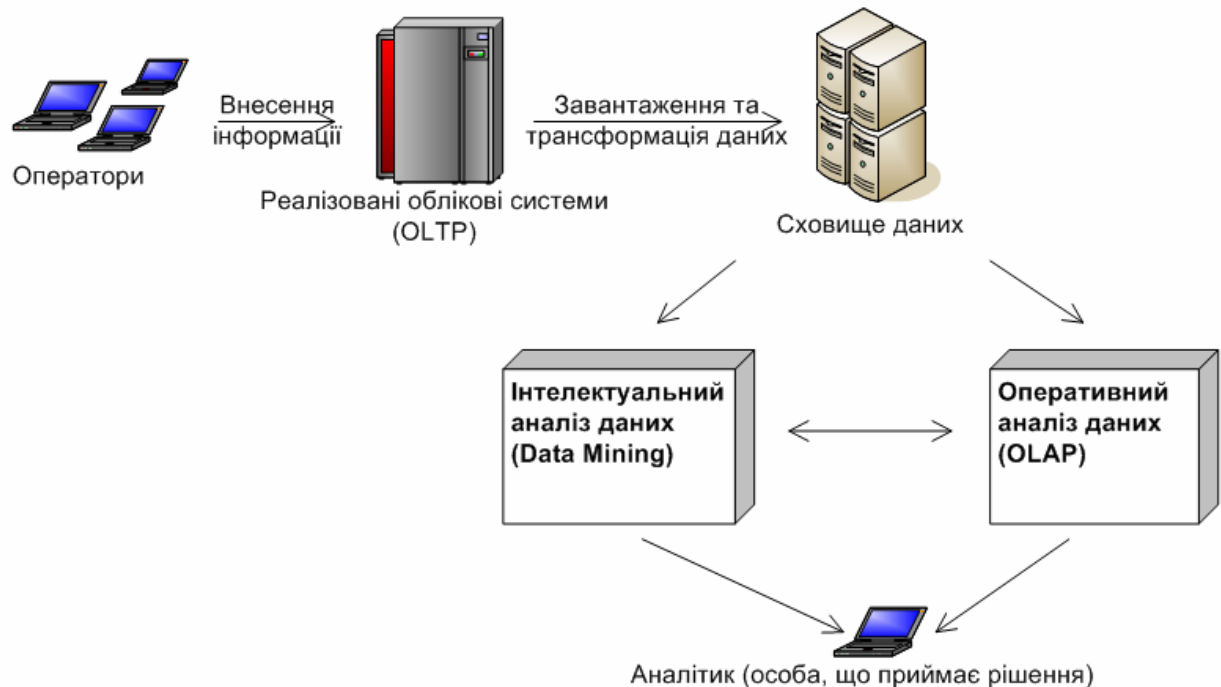


Рис. 1.6. Концепція побудови СППР

Реалізацію сховища даних та у його складі багатовимірну структуру даних планується реалізувати на основі сторонньої СКБД Oracle, оскільки:

- дана СКБД використовуються у реалізованих реєстраційних системах обліку інформації про об'єкти;
- дана СКБД має певні переваги, що були зазначені у попередніх підрозділах;
- реалізація власного сховища даних у тому числі з багатовимірною структурою даних, процес складний та ресурсоємних, виконання якого з точки зору вартості СКБД Oracle та необхідності застосування її лише у якості єдиного центрального сховища системи є недоцільним.

У якості частини систем, що безпосередньо забезпечить підтримку прийняття рішень, доцільно використання комбінованих засобів оперативного

та інтелектуального аналізу, оскільки інтеграція даних технологій обробки інформації забезпечує найбільшу гнучкість та ефективність. На ринку програмного забезпечення представлена певна кількість засобів, що реалізують як оперативний, так і інтелектуальний аналіз, але з огляду на наступні аспекти:

- висока вартість продуктів даного сегменту, враховуючи кількість рівнів застосування системи;
- надмірна функціональність, що не буде використовуватись;
- потреба у доведенні програмних продуктів під потреби відповідних предметних галузей;
- неможливість розширення або модифікації наданих продуктів;
- відсутність готових рішень.

Зважаючи на необхідність практично єдиного технологічного рішення для обох систем, більш доцільним виступає рішення розробки власних засобів інтелектуального та оперативного аналізу. Розробка та впровадження власних засобів має наступні переваги по відношенню до сторонніх продуктів:

- запропоноване рішення буде повністю охоплювати наявні потреби та усі етапи вирішення задач;
- можливість інтеграції усіх компонентів у єдину систему;
- можливість реалізації унікальних методів вирішення наявних задач;
- можливість подальшого розвитку та модифікацій запропонованих методів.

Реалізація засобів інтелектуального та оперативного аналізу вимагає розроблення моделей представлення даних переметної галузі у багатовимірному вигляді, методів їх обробки та інтелектуального аналізу з метою вирішення необхідних задач.

Висновки до розділу

В даному розділі проведено огляд та аналіз сучасних технологій аналізу даних: OLAP та Data Mining, розглянуті підстави їх виникнення, основні

властивості, характеристики та тенденції розвитку. Здійснено оцінку можливостей сучасних СППР на основі даних, що представлені на ринку, з точки зору їх функціональності, готовності до впровадження та вирішення задач поставлених у рамках предметних галузей. На основі проведеного аналізу та здійснених оцінок, прийняте рішення про необхідність розробки власних засобів оперативного та інтелектуального аналізу. Представлені результати проведеного аналізу та класифікації інформаційних систем

РОЗДІЛ 2.

ДОСЛІДЖЕННЯ МОДЕЛЕЙ ТА АЛГОРИТМІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

2.1 Технологія проведення інтелектуального аналізу даних

Для розуміння призначення та місця кожного з механізмів аналізу, що будуть покладені у СППР, що створюються на основі реєстру певної предметної області, розглянемо увесь процес прийняття рішення цілком. Серед множини методик проведення аналізу, найбільш ефективними та поширеними визначено:

- Перший варіант: особа, що приймає рішення використовує інформаційну системи виключно як засіб отримання інформації, а висновки робить сам. Для потреб такого роду, використовуються системи звітності, діаграми та інші засоби візуалізації.
- Другий варіант: система не тільки забезпечує отримання та представлення даних, а й проводить різного роду обробку, а до оброблених даних застосовує методи інтелектуального аналізу – кластеризацію, класифікацію, регресію, тощо. У даному випадку людина, що приймає рішення отримує не сирі данні, а данні, що пройшли обробку та/або представляють певну рекомендацію, щодо прийняття рішення. Тобто людина працює з моделями, які підготовлені системою.

Завдяки тому, що у першому випадку практично все, що пов'язано з механізмами прийняття рішення покладено на людину, проблема з обранням адекватної моделі та вибором методів обробки виносяться за рамки системи та покладені на людину. У деяких випадках цього достатньо, але у разі якщо людину, що приймає рішення, цікавлять знання, що знаходяться доволі «глибоко», прості механізми отримання даних тут недоречні. Необхідна більш серйозна обробка даних, що була описана вище у якості другого варіанту. Оптимальним випадком була б можливість застосування обох підходів до

аналізу. Вони дозволяють забезпечити практично усі потреби аналізу інформації. Розглянемо модель проведення аналізу, яка буде покладена у основу методів та технологій аналізу, що будуть запропоновані у даному розділі для вирішення поставлених задач (рис. 2.1).

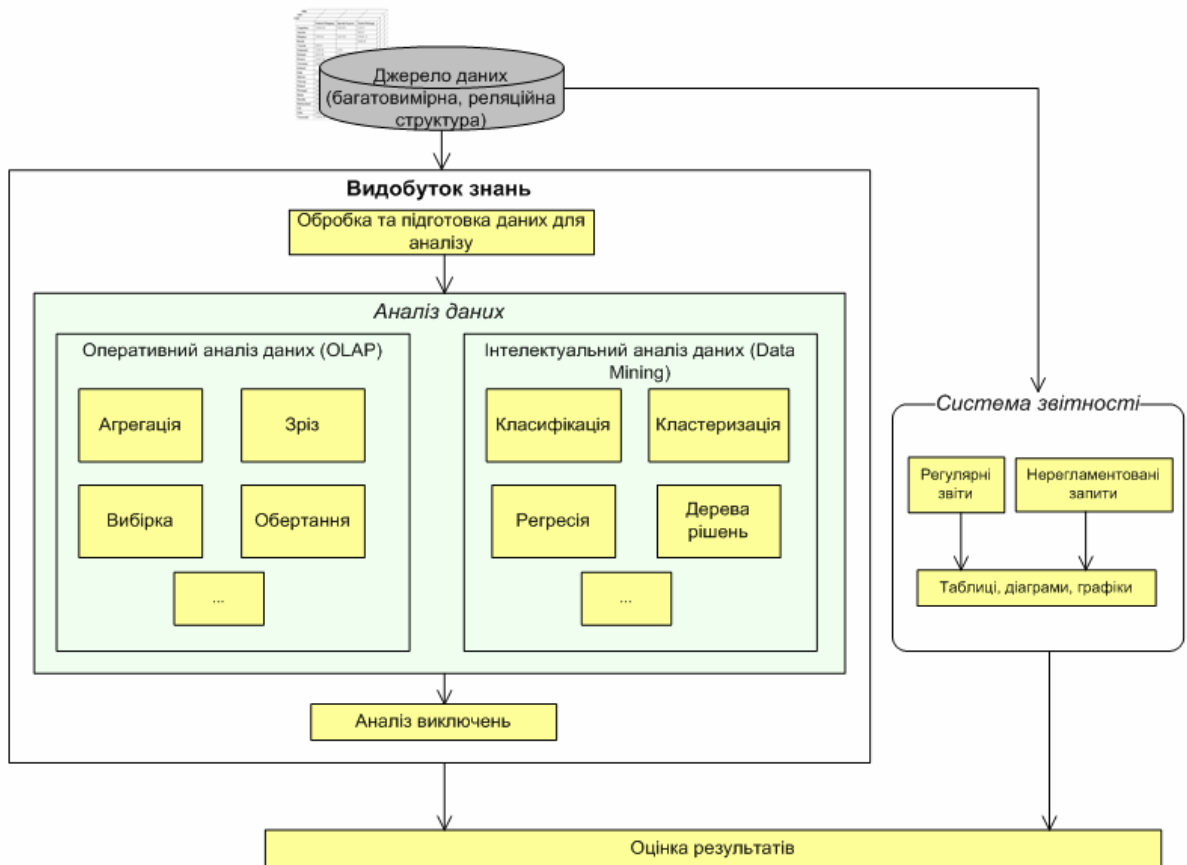


Рис. 2.1. Узагальнена модель проведення аналізу даних

Часто при описі тієї чи іншої задачі застосовуються терміни – ризик менеджмент, прогнозування, тощо. Але в дійсності рішення кожної з цих задач зводиться до застосування одного чи сукупності методів інтелектуального аналізу. Розглянемо кожен з елементів моделі, що зображені на рис. 2.1:

- в якості первинного джерела виступає сховище даних, яке містить реляційну та, при наявності й необхідності, багатовимірну структуру даних на основі яких здійснюється подальший аналіз. Сховище даних не реалізує жодних технологій аналізу даних, натомість воно виступає тією базою, яка

дозволяє використовувати данні у вигляді, що придатний для проведення ефективного інтелектуального аналізу. При відсутності сховища даних на збір та систематизацію даних, необхідних для аналізу, буде витрачатись велика частина часу, що знівелює усі переваги аналізу, оскільки одним з ключових показників будь якої аналітичної системи є можливість швидко отримати результат.

- *системи звітності призначені для відповіді на питання «що відбувається».* Перший варіант використання – регулярні звіти, що використовуються для контролю оперативної ситуації та аналізу відхилень. Другий варіант – обробка нерегламентованих запитів. У разі якщо особа, що приймає рішення, бажає перевірити будь яку гіпотезу, йому необхідно отримати інформацію, що підтверджує чи спростовує гіпотезу. Оскільки данні ситуації виникають спонтанно та відсутнє уявлення про те, якого роду інформація буде потрібна – потрібен механізм, конструктор запитів, що дозволяє швидко та в зручному виді дану інформацію отримати. Отриманні дані як правило представляються у вигляді графіків, діаграм, тощо.
- *видобуток знань* – це процес перетворення даних у знання. Процес включає в себе питання: підготовки даних; вибору інформаційних ознак; очищення даних; застосування методів інтелектуального та оперативного аналізу інформації; постобробки даних; оцінки отриманих результатів.

Привабливість даного процесу полягає у тому, що в незалежності від предметної галузі, ми використовуємо одні і ті операції, що перелічені вище, об'єднані та скомпоновані відповідним чином у конкретну методику проведення аналізу. При необхідності, даний процес може відбуватись ітеративно. Одночасне виконання оперативного аналізу (OLAP) та інтелектуального аналізу (Data Mining) дозволяє значно розширити межі та підвищити ефективність проведення аналізу.

- *оцінка результатів проводиться особою, що приймає рішення.* Повністю виключити участь людини неможливо, оскільки той чи інший результат

повинен бути інтерпретований у відповідності до конкретної предметної галузі.

На рис. 2.1, описана технологія аналізу даних, згідно якої здійснено вирішення поставлених задач, що актуальні для предметних галузей реєстрових систем, шляхом комбінації методів інтелектуального та оперативного аналізу в залежності від конкретної задачі аналізу.

2.2 Моделі та методи прогнозування виникнення критичних ситуацій

Постановка задачі

У загальному випадку вирішення даної задачі можливо шляхом використання накопиченої інформації про об'єкт чи процес та застосування певних методів її обробки. Застосування комплексу даних методів надає можливість спрогнозувати можливість виникнення критичної ситуації та прийняти міри по її недопущенню.

У формалізованому вигляді задачу можливо представити наступним чином: існує певна кількість об'єктів $S_n^{R_i}$, які характеризуються певним набором параметрів $A_m^{S_n^{R_i}}$, що визначають як кількісні показники об'єкту, так і його якісні показники. Оперуючи накопиченою історичною інформацією щодо об'єкту, можливо, оцінюючи поточні значення параметрів $A_m^{S_n^{R_i}}$, насамперед кількісних, визначити можливу майбутню зміну якісних показників вихідних об'єктів, тобто:

$$F(S_n^{R_i}(A_m^{S_n^{R_i}}), S_n^{R_i}(A_m^{S_n^{R_i}})) \rightarrow R(S_n^{R_i}(A_m^{S_n^{R_i}})), \quad (2.1)$$

де: $F()$ – методи обробки даних множин вихідних об'єктів, що застосовуються з метою отримання множини R , у рамках вирішення задачі та згідно запропонованих у даному розділі схем та методів; $R(S_n^{R_i}(A_m^{S_n^{R_i}}))$ – результат

вирішення задачі, що складається з підмножин вихідних об'єктів $S_n^{R_i}$ та прогнозованих значень якісних параметрів $A_m^{S_n^{R_i}}$; $S_n^{R_i}$ – множина об'єктів, для яких здійснюється аналіз; $S_n^{R_i}$ – множина об'єктів, на основі інформації про які здійснюється аналіз; $A_m^{S_n^{R_i}}$ – множина кількісних та якісних атрибутів об'єктів, на основі яких здійснюється аналіз.

Виходячи з вище сказаного, вирішення задачі вимагає визначення та опису методів аналізу $F()$ над множинами інформації щодо об'єктів, які приймають участь у аналізі та об'єктів на основі інформації про які, проводиться аналіз (рис. 2.2).

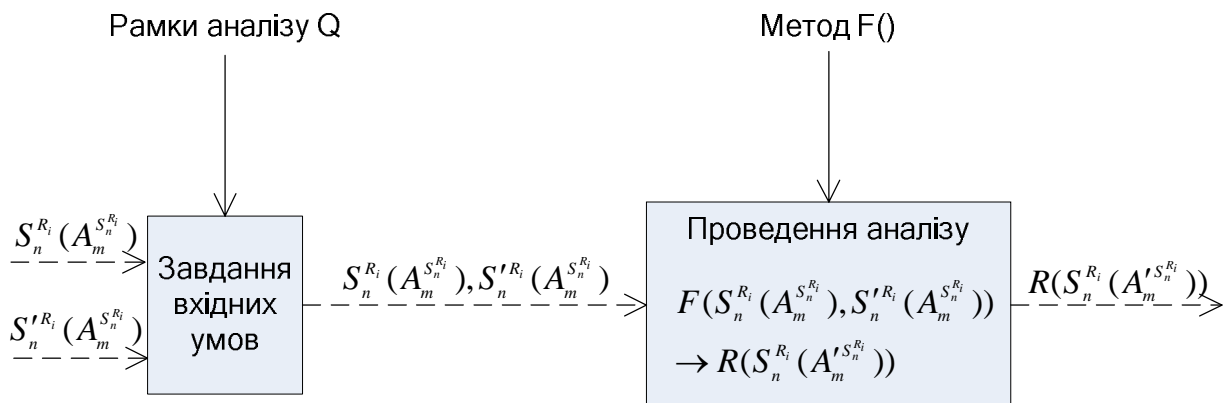


Рис. 2.2. Схема постановки задачі

Множини $S_n^{R_i}(A_m^{S_n^{R_i}})$ та $S_n^{R_i'}(A_m^{S_n^{R_i'}})$ формуються з урахуванням рамок аналізу Q , що визначають набір інформації, який використовується при проведенні аналізу та задаються особою, що приймає рішення. Перелік керуючих параметрів, що утворюють рамки аналізу буде розглянуто далі.

Використання облікової частини системи надає вичерпну інформацію про зв'язок між значенням параметрів об'єкту, його властивостями та подіями, що з ним відбувались, а також можливість проведення аналізу у певних рамках, таких як: окремо обране господарства розведення та утримання тварин, складу

зберігання зерна, регіону, області, тощо, тобто надає можливість абстрагуватись від додаткових, невизначених факторів аналізу.

Опис методу

Для попередження виникнення критичних ситуацій, пропонується за рахунок накопиченої у базі даних інформації $S_n^{R_i}(A_m^{S_i})$ сформулювати метод інтелектуального аналізу, що надасть можливість побудувати кластери, які будуть визначати простір, що сигналізує про високу прогнозовану ймовірність виникнення критичної ситуації (далі – критичні кластери). За допомогою запропонованого методу, встановимо прогнозовану ймовірність виникнення критичної ситуації для об'єктів, що аналізуються $S_n^{R_i}(A_m^{S_i})$ в залежності від їх параметрів та додаткових умов.

У якості сукупність параметрів $A_m^{S_i}$, що впливають на вірогідність виникнення критичної ситуації для об'єкту були визначені певні абстрактні параметри: $(A_2^{S_1^{R_1}})$; $(A_3^{S_1^{R_1}})$; $(A_4^{S_1^{R_1}})$, $(A_1^{S_2^{R_1}})$, $(A_5^{S_1^{R_1}})$, $(A_2^{S_2^{R_1}})$.

На основі даних параметрів побудовані критичні кластери та проведено аналіз потрапляння чи не потрапляння об'єкту до критичного кластеру. Для вирішення поставленої задачі розроблено метод, узагальнена схема якого зображена на рис. 2.3.

Проаналізуємо та дамо опис кожного з етапів вирішення задачі

Завдання вихідних параметрів та умов

Даний етап є підготовчим етапом аналізу та полягає у визначенні рамок аналізу Q та визначенні переліку об'єктів необхідних для проведення аналізу $S_n^{R_i}(A_m^{S_i})$, $S_n^{R_i}(A_m^{S_i})$. У якості вихідних або керуючих даних для проведення аналізу доцільно використовувати наступну інформацію у вигляді параметрів: (Q_1) ; (Q_2) ; (Q_3) ; (Q_4) ; (Q_5) ; (Q_6) .

Визначення даних параметрів регламентує перелік інформації з бази даних, що буде використана при проведенні аналізу. Завдання рамок аналізу

надає можливість отримання більш гнучких результатів аналізу та можливість врахування або не врахування додаткових факторів що впливають на об'єкт. Після визначення рамок аналізу перейдемо до розгляду наступного етапу – визначення критичних кластерів.



Рис. 2.3. Узагальнена схема методу вирішення поставленої задачі

Визначення критичних кластерів

На даному етапі, вирішення поставленої задачі, буде визначено кластери, потрапляння об'єкту до яких свідчить про високу прогнозовану ймовірність

виникнення критичної ситуації для даного об'єкту. Виконання даного етапу вимагає використання певного методу та алгоритму проведення кластеризації. Зважаючи на їх велику кількість та різноманітність, було проведено аналіз найбільш поширених методів та алгоритмів кластеризації, результати якого представлені нижче, та обрано найбільш ефективний алгоритм, виходячи з потреб завдання, що вирішується. У табл. 2.1 розглянемо переваги та недоліки кожної з груп методів кластерного аналізу.

Таблиця 2.1.

Переваги та недоліки методів кластерного аналізу

Метод	Переваги	Недоліки
Ієрархічні методи	<ul style="list-style-type: none"> - Не існує необхідності у визначенні числа кластерів. - Наочність та можливість отримати детальне представлення о структурі даних. - Існує можливість достатньо легко ідентифікувати відхилення у наборі даних та у результаті покращити якість даних. 	<ul style="list-style-type: none"> - Недоцільність використання при великих об'ємах даних. - Складності методів кластеризації: обмеження об'єму даних; вибір міри близькості; негнучкість отриманих результатів.
Неієрархічні методи	<ul style="list-style-type: none"> - Висока стійкість по відношенню до шумів та відхилень, некоректного вибору розмірностей, включення незначних змінних у набір даних, що приймають участь у кластеризації. 	<ul style="list-style-type: none"> - Необхідність попереднього визначення кількості кластерів, кількості ітерацій або правил зупинки.

Проаналізувавши методи кластеризації - визначимо групу методів, що найкраще підходить для вирішення поставлених завдань, враховуючи перелічені недоліки та переваги методів кластерного аналізу та специфіку задачі, що вирішується. Визначення методу проведемо на основі аналізу

відповідності ієрархічних та неієрархічних методів кластерного аналізу, вимогам поставленої задачі, що відображено у табл. 2.2.

Таблиця 2.2.

Аналіз відповідності методів кластеризації та поставленої задачі

Характеристика	Специфіка задачі	Аналіз	Перевага
Об'єм даних	Об'єм даних, що аналізується, буде варіюватись в залежності від рамок аналізу заданих аналітиком. Але, оскільки запропонована технологія проведення аналізу найбільш ефективна при аналізі великих об'ємів даних, відповідно методи кластеризації повинні ефективно працювати з ними.	Ієрархічні методи кластеризації не дозволяють ефективно працювати з великими об'ємами даних, натомість неієрархічні методи кластеризації це дозволяють.	Неієрархічні методи.
Ціль аналізу	Прогнозування виникнення критичної ситуації для об'єкту, узагальнено, можливо виразити через його відстань до центру кластеру.	Розрахунок центру кластеру виконується у групі неієрархічних методів кластеризації та не виконується у групі ієрархічних методів.	Неієрархічні методи.
Кількість кластерів	Поставлена задача не дозволяє однозначно встановити кількість кластерів.	Оскільки технологія проведення аналізу не дозволяє однозначно встановити кількість кластерів необхідних для проведення аналізу, тому згідно даної характеристики, доцільне використання ієрархічних методів кластеризації.	Ієрархічні методи.

Потреба у представленні інформації	Виходячи з кількості даних, що будуть використані для аналізу представлення структури інформації у будь якому наочному вигляді є неможливою та не є критичною для поставленої задачі.	Достатнім результатом буде прогнозування рівня ймовірності виникнення критичної ситуації для об'єкту чи групи об'єктів, звідси даний критерій не надає перевагу жодній з груп методів кластерного аналізу.	-
Стійкість до шумів та відхилень	Дана проблема актуальна для поставленої задачі.	Аналіз методів кластеризації за даною властивістю надає перевагу неієрархічним методам кластеризації, як більш стійким до впливу некоректних наборів даних.	Неієрархічні методи.

Підсумовуючи поставлену задачу та можливості, що надає кожна з груп методів кластерного аналізу, відмічаємо істотну перевагу неієрархічних методів кластерного аналізу. Тому наступним кроком буде визначення алгоритму з групи неієрархічних методів (k – середніх, нечітких k – середніх, Густафсона – Кеселя), що буде використаний для проведення кластерного аналізу у рамках вирішення задачі визначення прогнозованої ймовірності виникнення критичної ситуації для об'єкту або групи об'єктів (табл. 2.3).

Таблиця 2.3.

Аналіз та вибір алгоритму

Характеристика	Специфіка предметної галузі	Аналіз	Перевага
Необхідність оперування нечіткими множинами	У розрізі поставленої задачі немає необхідності у розрахунку вірогідності потрапляння об'єкту до кластеру, оскільки природа кластерів	Розглянуті алгоритми, окрім методу k – середніх, надають можливість розрахунку ймовірності	Жоден алгоритм не має переваги.

	однакова.	потрапляння об'єкту до кластеру	
Точність результатів аналізу	Результатом поставленої задачі є прогнозування ймовірності виникнення критичної ситуації для кожного з об'єктів, інформація про який аналізується, що не допускає спотворення результатів на етапах проведення аналізу.	Алгоритм k - середніх досить чуттєвий до відхилень, алгоритми нечітких k – середніх та Густафсона – Кеселя можуть спотворювати результати за рахунок заданої форми кластеру.	Жоден алгоритм не має переваги.
Швидкість роботи	Враховуючи необхідність роботи з великими об'ємами даних, швидкість роботи є важливим фактором при виборі алгоритму.	Перевагу у швидкості проведення аналізу має алгоритм k – середніх.	Алгоритм k - середніх.

Використання алгоритму k - середніх є найбільш доцільним серед розглянутих алгоритмів за умови вирішення проблеми чуттєвості до відхилень. Тому, пропонується використання модифікованого алгоритму k – середніх. Модифікація алгоритму полягає у визначенні середніх значень за методом медіан, що є менш чуттєвий до випадкових відхилень та забезпечить більшу точність у результатах аналізу (більш детально модифікацію алгоритму буде розглянуто нижче).

Модифікація критичних кластерів

На даному етапі проводиться модифікація сформованих критичних кластерів, шляхом побудови кластерів (стабільних кластерів) які формують об'єкти, для яких не виникала критична ситуація (у заданих рамках аналізу) та оцінка перетину цих кластерів й критичних кластерів. Схема даного етапу приведена на рис. 2.4.



Рис. 2.4. Схема модифікації критичних кластерів

1. *Побудова стабільних кластерів.* На даному етапі визначаються центри та границі кластерів на основі об'єктів для яких не виникала критична ситуація. Кластерний аналіз виконується аналогічно до описаного раніше, з тими ж рамками та умовами, єдиним виключенням є те, що у якості вхідної інформації потрапляють виключно об'єкти для яких не виникала критична ситуація, інформація про які відбирається відповідно до Q_{1-6} .

2. *Аналіз перетину кластерів.* На даному етапі виконується аналіз перетину критичних кластерів та стабільних кластерів, з метою модифікації критичних кластерів. Аналіз проводиться наступним чином:

- для кожного критичного кластеру визначаються його перетини зі стабільними кластерами;
- фіксується простір перетину кластерів (K_i);

- розраховується кількість об'єктів для яких виникла та невиникла критична ситуація $(\sum b_i, \sum z_i)$, що потрапили у простір перетину кластерів.
Схематично, принципи проведення аналізу зображені на рис. 2.5.

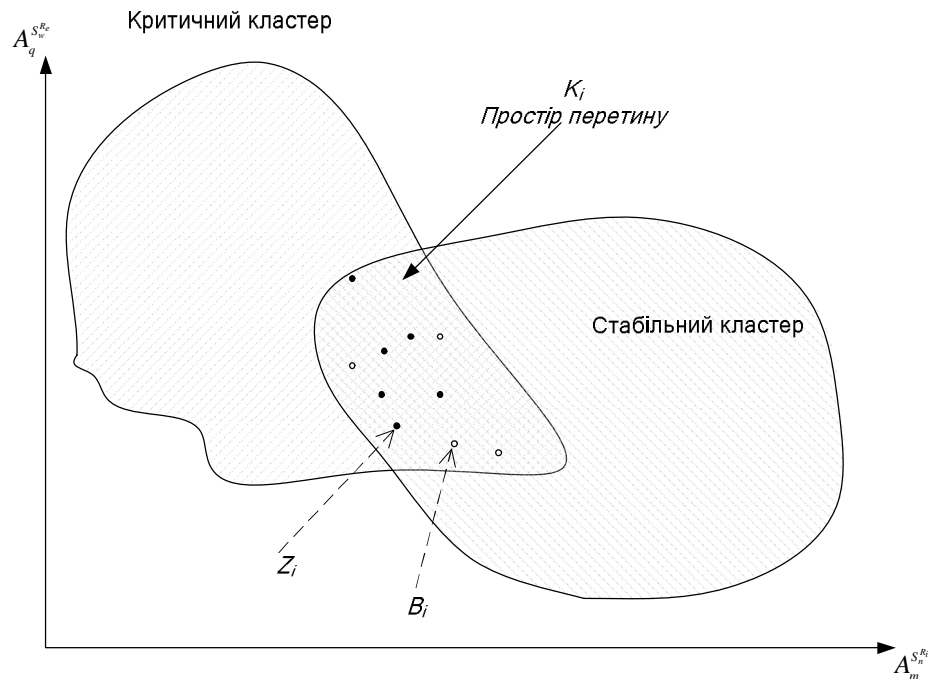


Рис. 2.5. Схема проведення аналізу перетину кластерів

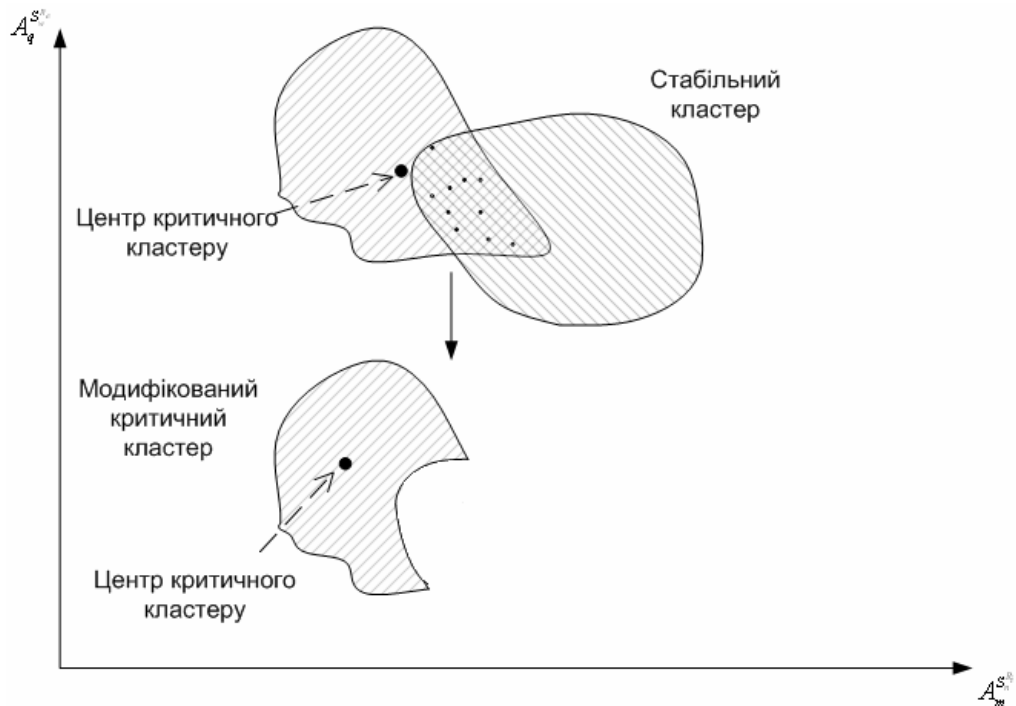


Рис. 2.6. Схема модифікації критичних кластерів

Результатом виконання даного етапу є модифіковані (при необхідності) критичні кластери зі зміненими границями та центрами кластерів. Модифікація критичних кластерів надає можливість зменшити вплив на результати аналізу додаткових та випадкових факторів, що не можливо (складно) врахувати, шляхом використання накопиченої у сховищі даних інформації.

Прогнозування ймовірності для об'єкту / групи об'єктів

Ціллю даного етапу є визначення рівня прогнозованої ймовірності виникнення критичної ситуації для об'єкта, групи об'єктів $S_n^{R_i}$, для яких проводиться аналіз. Узагальнена схема даного етапу зображена на рис. 2.7.



Рис. 2.7. Схема визначення рівня ймовірності

1. *Внесення/завантаження інформації про об'єкт.* На даному етапі, абстрагуючись від реалізації, у систему вноситься (у системі обирається) інформація про об'єкти $S_n^{R_u}$ (тварини), для яких необхідно провести аналіз. Це

може бути інформація як про існуючі об'єкти, так і набір параметрів неіснуючого об'єкту для якого необхідно провести аналіз.

2. *Визначення рівня ймовірності.* На даному етапі кожен з критичних кластерів G_i , умовно ділиться на шість відрізків по кожному виміру (змінній), як це зображено на рис. 2.8.

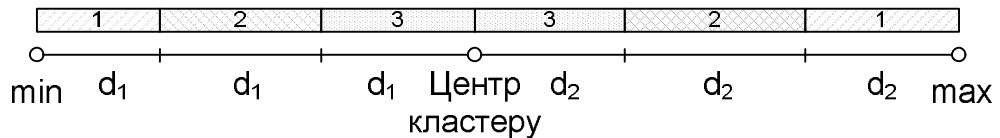


Рис. 2.8. Схема розділення кластеру по виміру

Кластер розбивається на відрізки від центру до мінімальної границі кластеру з кроком:

$$d_1 = \frac{c_x - \min_x}{3}, \quad (2.1)$$

відповідно від центру до максимальної границі:

$$d_2 = \frac{\max_x - c_x}{3}, \quad (2.2)$$

де: d_1, d_2 – крок ділення кластера; c_x – координати центру кластера за виміром (змінній) x ; \min_x, \max_x – відповідно координати мінімальної та максимальної точки кластеру за виміром x .

Кожен з відрізків, в залежності від ступеню віддаленості від центру кластеру, несе відповідно різний рівень прогнозованої ймовірності виникнення критичної ситуації для об'єкту, що у нього потрапив:

- 1 відрізок – низький рівень;
- 2 відрізок – середній рівень;
- 3 відрізок – високий рівень.

Для кожного об'єкту $S_n^{R_i}$, що приймає участь у аналізі визначається:

- приналежність до критичного кластеру $S_n^{R_i} \subseteq G_i$.
- по кожному виміру фіксується відрізок (\min_x ; \max_x) до якого потрапив об'єкт згідно його характеристик (виключно для об'єктів, що потрапили до критичного кластеру).
- визначається загальний, прогнозований рівень ймовірності виникнення критичної ситуації для об'єктів, що належать до критичного кластеру, згідно:

$$Y = \frac{\sum_{n=1}^n R_n}{n}, \quad (2.3)$$

де: Y – загальний рівень прогнозованої ймовірності: [1..1,5) – низький, [1,5..2,5) - середній, [2,5..3] - максимальний; n – кількість змінних (вимірів) за яким проводиться аналіз; R_n – значення рівня прогнозованої ймовірності виникнення критичної ситуації для об'єкту, згідно виміру n :

- 1 – низький рівень;
- 2 – середній рівень;
- 3 – максимальний рівень.

3. *Інтерпретація результатів аналізу.* а даному етапі відображаються результати проведеного аналізу. Для кожного з об'єктів, що приймав участь у аналізі, результати містять наступну інформацію (табл. 2.4).

Таблиця 2.4.

Представлення результатів аналізу

Унікальний ідентифікатор об'єкту	Приналежність до критичної групи	Прогнозований рівень ймовірності
Об'єкт 1	Так/Ні	-/Низький/Середній/Максимальний
...
Об'єкт N	Так/Ні	-/Низький/Середній/Максимальний

Таким чином, описані моделі та методи, надають можливість визначити рівень прогнозованої ймовірності поданих даних на основі накопичених емпіричних даних у реєстрових системах, з метою попередження та недопущення відповідних критичних ситуацій. Дані моделі та методи, можуть бути застосовані для вирішення типових задач у різних сферах людської життєдіяльності.

2.3 Моделі та методи розрахунку прогнозованих значень критичних показників об'єктів

Постановка задачі

Аналіз факторів, що впливають на визначені результативні ознаки та безпосередньо значення самих результативних ознак, свідчить про існування між ними кореляційного зв'язку. Наявність кореляційного зв'язку між параметрами у моделях R_1 та R_2 (моделі реєстру двох абстрактних об'єктів) надає можливість вирішення поставлених задач шляхом розробки відповідних методів проведення аналізу на основі кореляційно – регресійного аналізу. Застосування у основі розроблених методів елементів кореляційно - регресійного аналізу надасть можливість:

- визначити форми та параметри рівнянь зв'язку на основі яких можливо прогнозування відповідних показників.
- виміряти тісноту зв'язку та, як наслідок, відповідним чином інтерпретувати результати аналізу.

У загальному випадку задачу можливо представити у відповідності до рис. 2.9, де на основі параметрів об'єкту, а також додаткових факторів, що впливають на життєвий цикл об'єкту, необхідно здійснити прогнозує модулювання та розрахувати значення певних характеристик об'єкту при поточних чи довільних значеннях вихідних параметрів та факторів.

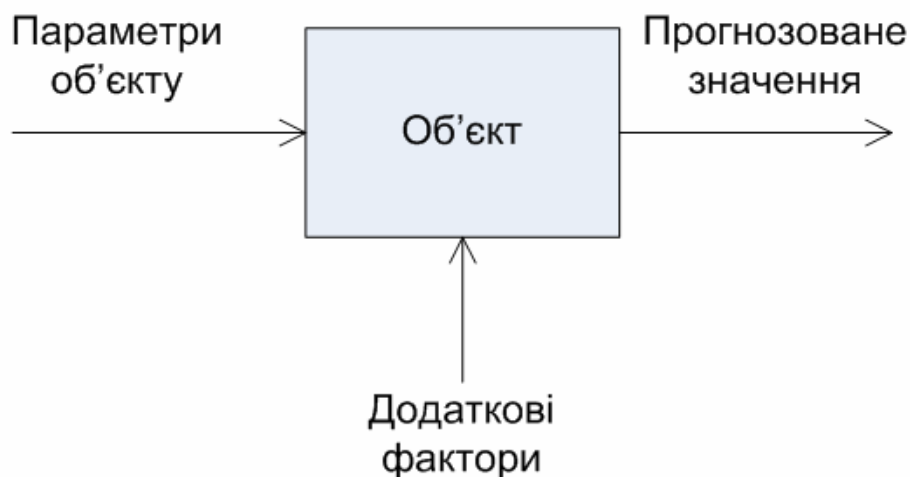


Рис. 2.9. Схема постановки задачі

Проведення аналізу пропонується здійснити на основі схеми, що зображена на рис. 2.10.

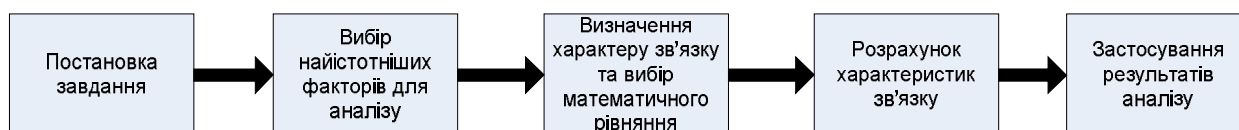


Рис. 2.10. Схема проведення аналізу

Перед розглядом етапів аналізу, розглянемо особливості деяких етапів схеми, з точки зору їх реалізації у системах:

- *вибір найістотніших факторів для аналізу, визначення характеру зв'язку та вибір математичного рівняння.* Данні етапи було здійснено шляхом аналізу факторних ознак, що можуть вчинити вплив на результативну ознаку, емпіричних даних, розрахунку характеристик регресійної моделі та кореляційних зв'язків для переліку усіх можливих факторних ознак та видів регресійної моделі - з метою отримання математичних рівнянь та переліку факторних ознак, застосування яких найкраще відображає наявні зв'язки у предметних галузях. Даний етап не виконується при здійсненні

аналізу, та може бути повторений у разі незадовільних показників експлуатації розробленої моделі та у випадку виникнення додаткових обставин та ознак.

- *розрахунок характеристик зв'язку.* На даному етапі, згідно певної математичної моделі, система розраховує якісні показники регресійної моделі та кореляційного зв'язку. Розрахунок даних параметрів відбувається на основі інформації, що наявна у БД (факторні та результативні ознаки), при чому фільтрація (відбір) даних, згідно яких проводиться аналіз, надає можливість врахувати фактори, інтерпретувати які у математичну модель неможливо. Розрахунок та відображення характеристик зв'язку при відображенні результатів аналізу надає можливість особі, що приймає рішення по - різному оцінювати результати проведеного аналізу.

- *Застосування результатів аналізу.*

Заключним етапом вирішення задачі є застосування результатів аналізу. Розглянемо даний етап та загальну схему вирішення задачі з точки зору взаємодії особи, що приймає рішення (аналітиком) з системами (рис. 2.11).

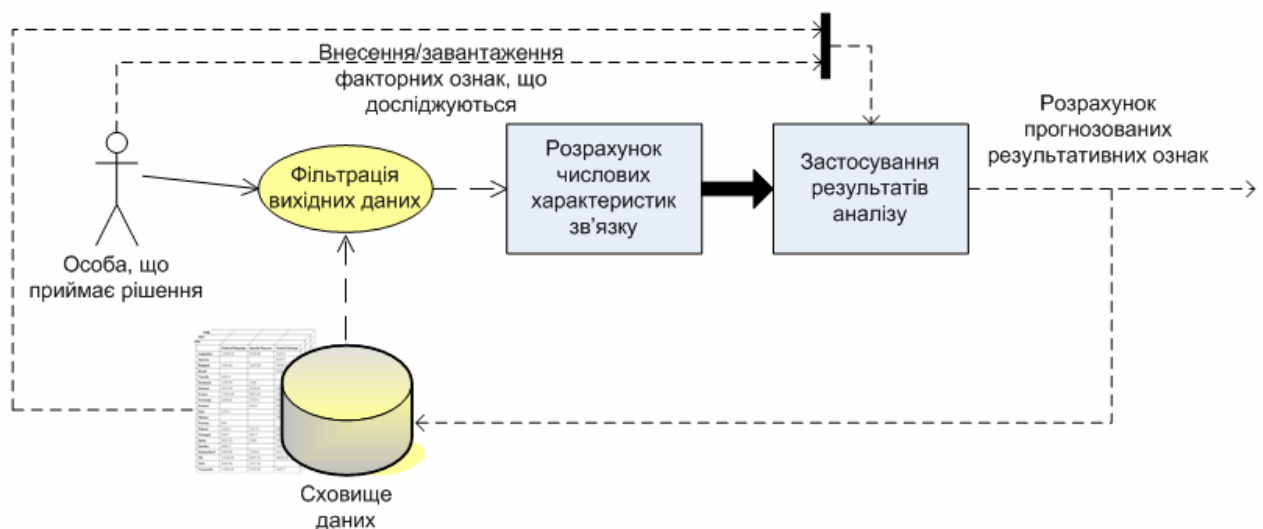


Рис. 2.11. Застосування результатів аналізу

На початку здійснення аналізу Аналітик відбирає (фільтрує) з реляційної або багатовимірної структури БД інформацію, на основі якої буде здійснено аналіз (область, регіон, вид об'єкту, інші фактори та параметри). Інформацію про об'єкти, що досліджуються, Аналітик також завантажує із БД, зовнішніх джерел або вносить вручну.

Після розрахунку характеристик зв'язку, система відображає користувачу результати аналізу, які при необхідності, він також може зберегти у сховище даних та використати для подальшого вивчення та аналізу. На основі отриманих результатів аналізу Аналітик має можливість: прогнозувати значення абстрактних параметрів $A_1^{S_1^{R_1}}$ та $A_1^{S_1^{R_2}}$, проводити оцінку відхилень від прогнозованих результатів, приймати рішення щодо збільшення / зменшення значення того, чи іншого фактору, що впливає на результативну ознаку, тощо.

2.4 Взаємодія технологій OLAP та Data Mining

Як було зазначено раніше, найбільш ефективним при здійсненні аналізу є використання засобів інтелектуального (Data Mining) та оперативного (OLAP) аналізу. У кожному з методів, що були описані у рамках даного розділу, з метою підвищення швидкості проведення аналізу, поряд з методами інтелектуального аналізу були використанні методи оперативного аналізу. Загальна схема даної взаємодії представлена на рис. 2.12, де представлені етапи сформованих методів інтелектуального аналізу, у яких відбувається взаємодія з методами оперативного аналізу на основі побудованих багатовимірних моделей. Основний зміст даної взаємодії полягає у:

- Прискоренні розрахунків при підготовці та обробці даних, за рахунок попереднього обчислення відповідних значень параметрів у побудованих багатовимірних моделях.

- Відбору для проведення аналізу даних, що є коректними, та показовими, виключивши з аналізу – помилкову, некоректну, незначущу інформацію, тощо.

Таким чином, підсумовуючи запропоновані методи аналізу, необхідно відзначити, що при їх застосуванні взаємодія з методами оперативного аналізу не є обов'язковою, але наявність такої взаємодії, надає практично необмежену гнучкість при підготовці, відборі даних для проведення аналізу, при цьому не вимагаючи написання непомірних по розміру та складності запитів, прискорює етапи розрахунку та обробки даних, надає розширені можливості представлення результатів аналізу

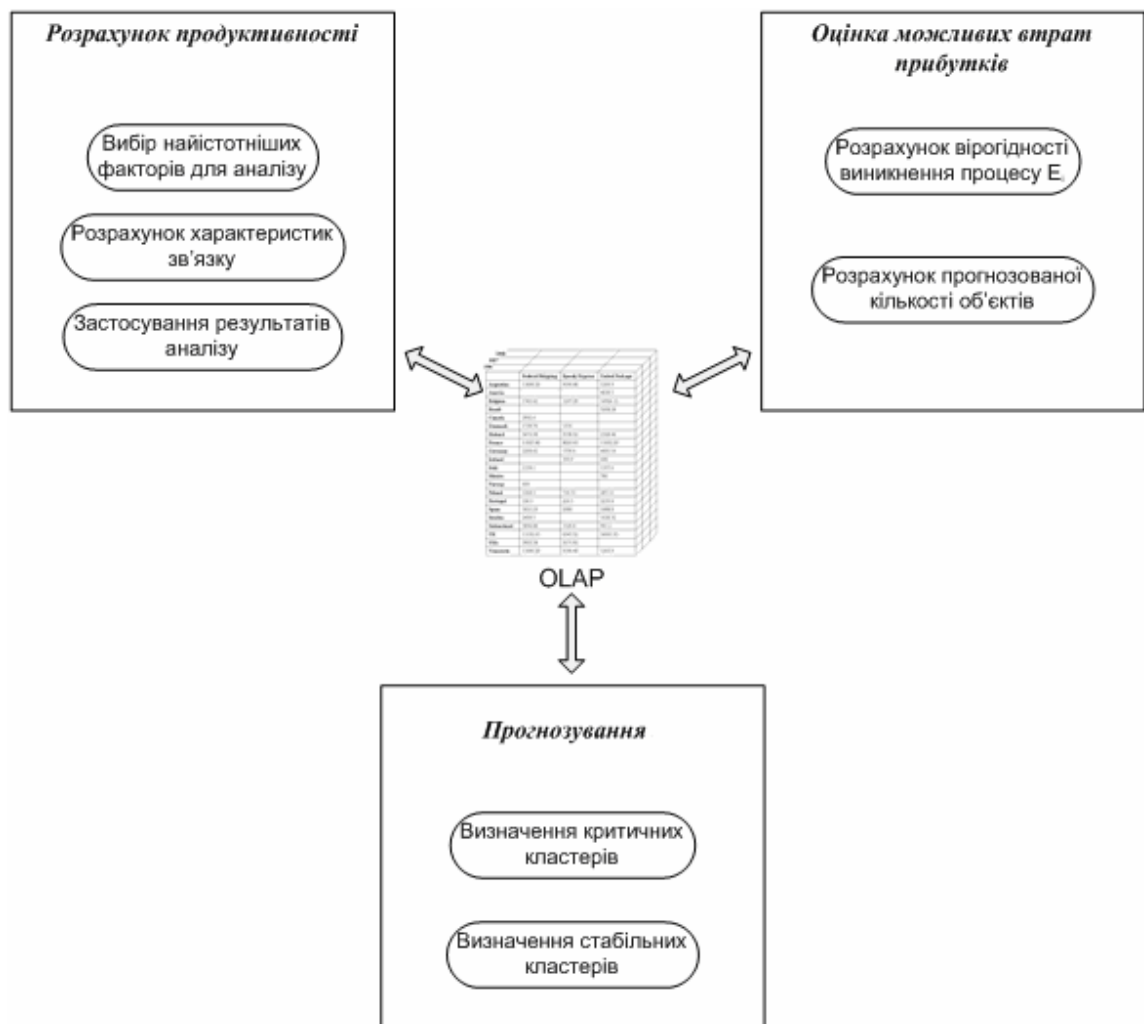


Рис. 2.12. Схема взаємодії методів інтелектуального та оперативного аналізу

2.5 Формалізований опис операцій багатовимірної моделі даних

Оперативний аналіз даних є засобом швидкого доступу, перегляду та аналізу інформації, який маніпулює моделями даних, що представляють собою багатовимірні куби. Осями багатовимірної системи координат виступають основні атрибути об'єкту предметної галузі, що аналізується (рис. 2.13). На перетині осей - вимірів (dimensions) – знаходяться данні, що якісно характеризують процес – міри (measures).

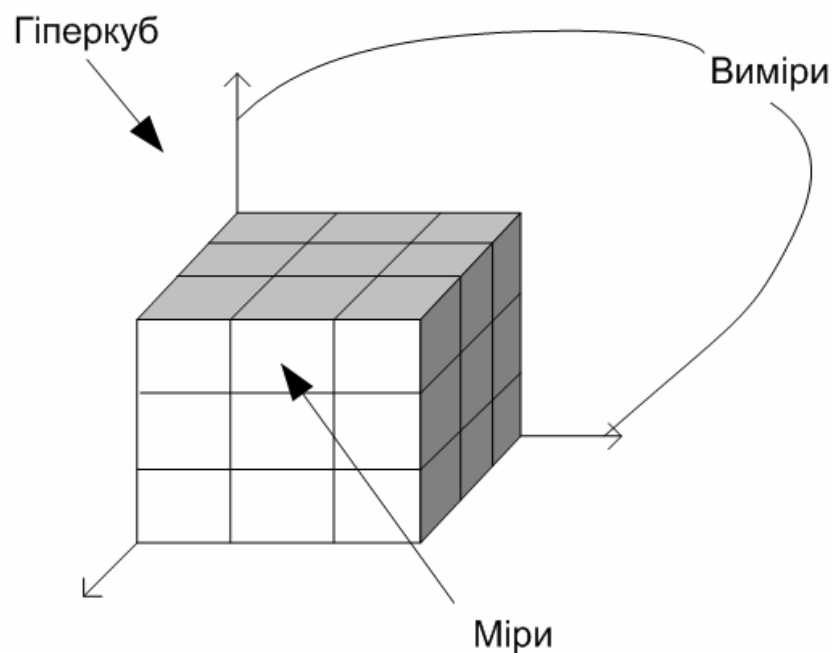


Рис. 2.13. Узагальнена схема багатовимірної моделі

Основними поняттями багатовимірної моделі є:

- багатовимірний куб.
- вимір.
- мітки.
- комірки.
- міри.

Багатовимірний куб даних складається з одного чи більше вимірів та представляє собою впорядкований набір комірок. Кожна з комірок

визначається одним і тільки одним набором значень вимірів – міток. Комірка може містити данні – міру, чи бути порожньою.

Реалізація засобів оперативного аналізу даних у СППР вимагає проектування моделей багатовимірних структур даних та методів їх обробки. Оскільки, принципи формування моделей наведені у попередньому розділі, розглянемо методи обробки багатовимірних структур даних, на основі яких реалізуються засоби маніпулювання даними у багатовимірних моделях.

Такими методами є операції, що здійснюються над багатовимірною інформацією:

- зріз;
- обертання;
- агрегація;
- запит.

Розробка засобів оперативного аналізу даних вимагає формалізованого опису методів обробки багатовимірної інформації з метою їх подальшої реалізації у компонентах програмного продукту. Таким чином у даному розділі, формалізуємо методи обробки багатовимірної інформації, виходячи із сформованих багатовимірних моделей та потреб аналізу.

Багатовимірний куб даних позначимо як множину комірок $H(D, M)$, що відповідають множинам D, M . Підмножину багатовимірного кубу даних, що відповідає множинам фіксованих значень D', M' позначимо як $H'(D', M')$.

Кожній з комірок багатовимірного кубу $h \in H$ відповідає єдино можливий набір міток вимірів $M_h \subset M$. Комірка може бути порожньою (не містити даних) або містити значення показника – міру.

2.5.1. Операція «Зрізу»

Підмножина багатовимірного кубу $H'(D', M')$ отримана внаслідок фіксації міток одного чи більше вимірів називається зрізом. Операція побудови зрізу виконується з ціллю отримання необхідної підмножини комірок $H' \subset H$

та відсічення «непотрібних» значень шляхом послідовної фіксації міток. Зріз, як правило, представляє двомірний масив (таблицю).

Мітка $m_{j_i} \in M$ задає площину зрізу багатовимірного кубу даних, що відповідає виміру $d_i \in D$. Множина фіксованих міток $M' \subseteq M$ задає множину площин зрізів кубу, що відповідає множині фіксованих вимірів $D' \subseteq D$. Перетин даних площин визначає множину комірок (зріз) багатовимірного кубу даних $H'(D', M')$, що і є результатом виконання операції. Таким чином суть виконання операції полягає у побудові зрізу багатовимірного кубу даних $H'(D', M')$ шляхом визначення множин D' та M' .

Фіксація мітки $m_{j_i} \in M$, що відповідає виміру $d_i \in D$, визначає область кубу, яку необхідно отримати. На кожному наступному етапі доступні мітки, що відповідають множині незафіксованих вимірів $D \setminus D'$. Таким чином, фіксація мітки у вимірах $D' \parallel |D \setminus D'| = 2$ надає можливість отримання зрізу багатовимірного кубу даних у вигляді таблиці.

2.5.2. Операція «Обертання»

Зміна порядку представлення (візуалізації) вимірів називається обертанням. Обертання забезпечує можливість візуалізації даних у формі найбільш комфортної для їх сприйняття. У термінах моделі, що розглядається, означає зміну послідовності фіксації міток при побудові зрізу. Результатом обертання для двомірного зрізу (таблиці) буде заміна стовпців на строки та навпаки.

2.2.3. Агрегація даних

Агрегація даних – отримання значень, що відповідають міткам деякого рівня $l, l \geq 1$ ієрархічного виміру D на основі значень рівня $l-1$ [53]. Розглянемо ієрархічний вимір D з L рівнями (рис. 2.14).

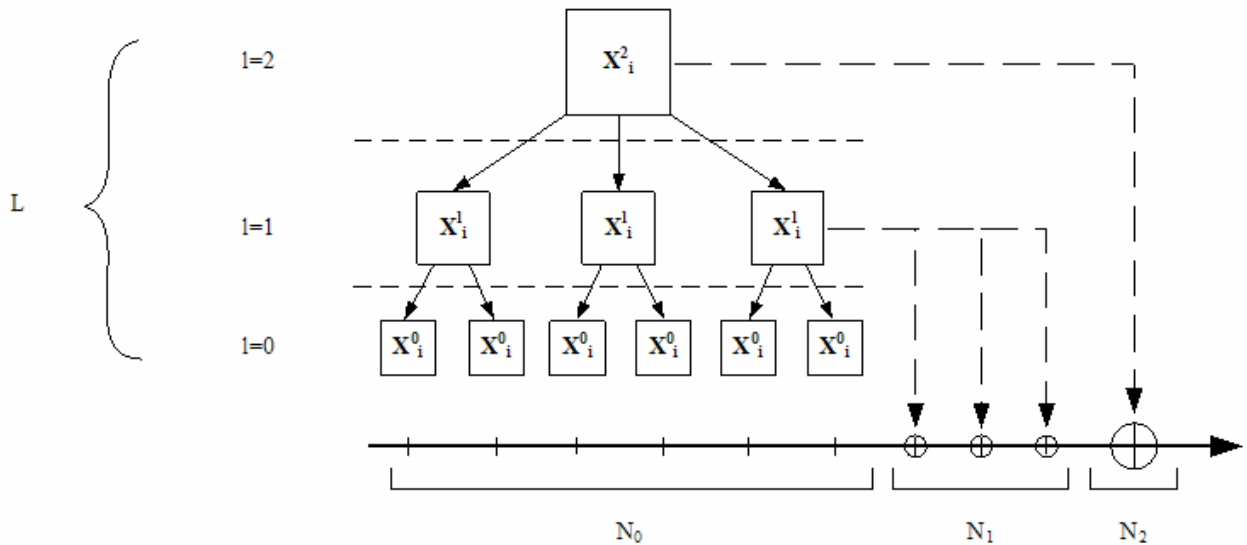


Рис. 2.14. Агрегація багатовимірного кубу даних. Одномірне представлення

Первинні данні (факти) відповідають нижньому рівню ієрархії ($l = 0$). Розрахунок агрегатів виконується у відповідності до методу агрегації. Наприклад, у випадку сумування значень агрегату X_j^l на рівні ієрархії $l = 1$, може бути розраховано за формулою:

$$X_j^l = \sum_{i=1}^{M_j} x_i^0, \quad (2.4)$$

де M_j – кількість фактів, відповідних міток, що є дочірніми по відношенню до мітки j . Узагальнюючи, отримаємо формули розрахунку агрегатів згідно методу сумування на інших рівнях ієрархії:

$$X_j^l = \sum_{i=1}^{M_j} x_i^{l-1}, l = 1, \dots, L; j = 1, \dots, N_l. \quad (2.5)$$

Ось виміру D , що спочатку містить мітки, що відповідають нижньому рівню ієрархії ($l = 0$), може бути доповнена мітками, що відповідають рівням

ієрархії, починаючи з $l = 1$. Таким чином, різниця між мітками, що відповідають первинним даним та мітками, що відповідають агрегатам є – умовною.

Операція згортки даних у даному випадку представляє собою побудову зрізу багатовимірного кубу, що відповідає зміні мітки рівня агрегації $l_1 \mid 0 \leq l_1 < L$ на рівень $l_2 \mid l_1 \leq l_2 < L$. Операція деталізації відповідає зміні мітки рівня $l_1 \mid 1 \leq l_1 \leq L$ на рівень $l_2 \mid 0 \leq l_2 < l_1$.

Кількість агрегатів для одного виміру: $N_A = \sum_{i=1}^L N_i$. Розглянемо випадок двох вимірів (рис. 2.15).

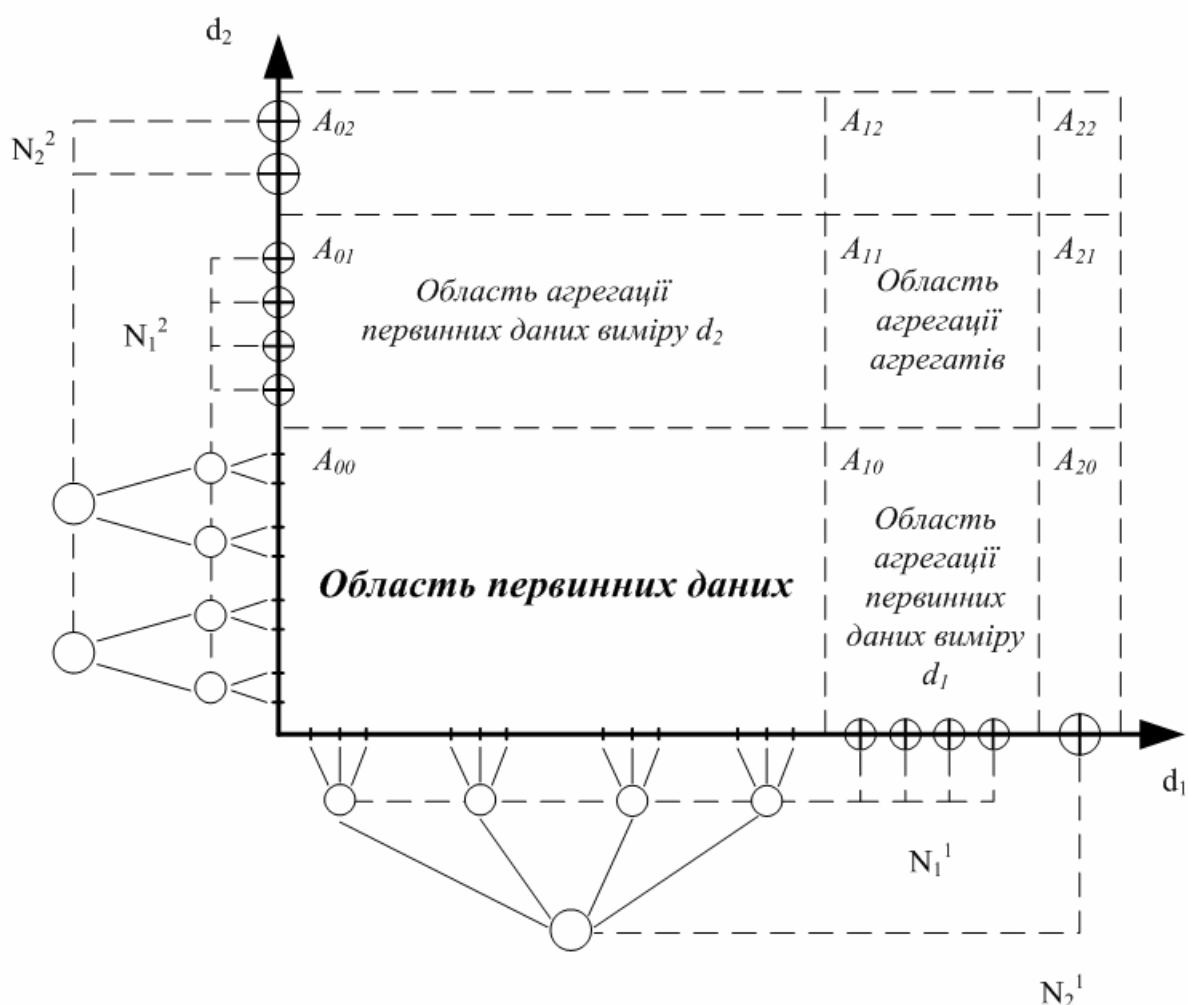


Рис. 2.15. Агрегація багатовимірного кубу даних. Двовірне представлення.

Кількість агрегатів, що зберігається у багатовимірному кубі, наряду з первинними даними, залежить від кількості міток, що відповідають рівням

ієрархії вимірів кубу, починаючи з $l=1$, та може істотно перевищувати кількість первинних даних. У випадку двох вимірів кількість агрегатів буде складати сума значень областей: $A_{01}, A_{02}, A_{10}, A_{11}, A_{12}, A_{20}, A_{21}, A_{22}$.

З іншого боку кількість агрегатів може бути розрахована як різниця кількості усіх значень багатовимірного кубу і кількості значень, що відповідають області первинних даних A_{00} . Таким чином кількість агрегатів N_A багатовимірного кубу у двовірному випадку дорівнює:

$$N_A = (N_0^1 + N_1^1 + \dots + N_{L_1}^1) \times (N_0^2 + N_1^2 + \dots + N_{L_2}^2) - N_0^1 \times N_0^2 = \sum_{i=0}^{L_1} N_i^1 \times \sum_{i=0}^{L_2} N_i^2 - N_0^1 \times N_0^2. \quad (2.6)$$

Узагальнюючи для випадку довільної кількості вимірів D , отримуємо:

$$N_A = \prod_{j=1}^D \sum_{i=0}^{L_j} N_i^j - \prod_{j=1}^D N_0^j, \quad (2.7)$$

де N_i^j - кількість міток i -го рівня ієрархії виміру $j, j=1..D$, а L_j - кількість рівнів ієрархії виміру j .

2.5.4. Вибірка даних

Процес побудови запиту на отримання даних з багатовимірного кубу даних може бути представлено у вигляді сітьового графу (рис. 2.16).

Розглянемо сітьовий граф $G(S, P)$, вершини якого відповідають міткам кубу даних $H(D, M)$. Множина вершин графу S включає n підмножин – шарів $S_i \subset S$, що відповідають крокам користувача при покроковій фіксації міток. Шар $S_i, i=1, \dots, n$ складається з вершин $s_{ji}, j=1, \dots, |M|$, що відповідають міткам, які можуть бути зафіксовані на i -му кроці.

Множина ребер графу P представляє набір вершин $p_{jk}^i = \{s_{j_i}, s_{k_{i+1}}\}$. Ребро графу $p_{jk}^i \in P$, таким чином з'єднує j -у вершину шару S_i , з k -ю вершиною шару S_{i+1} .

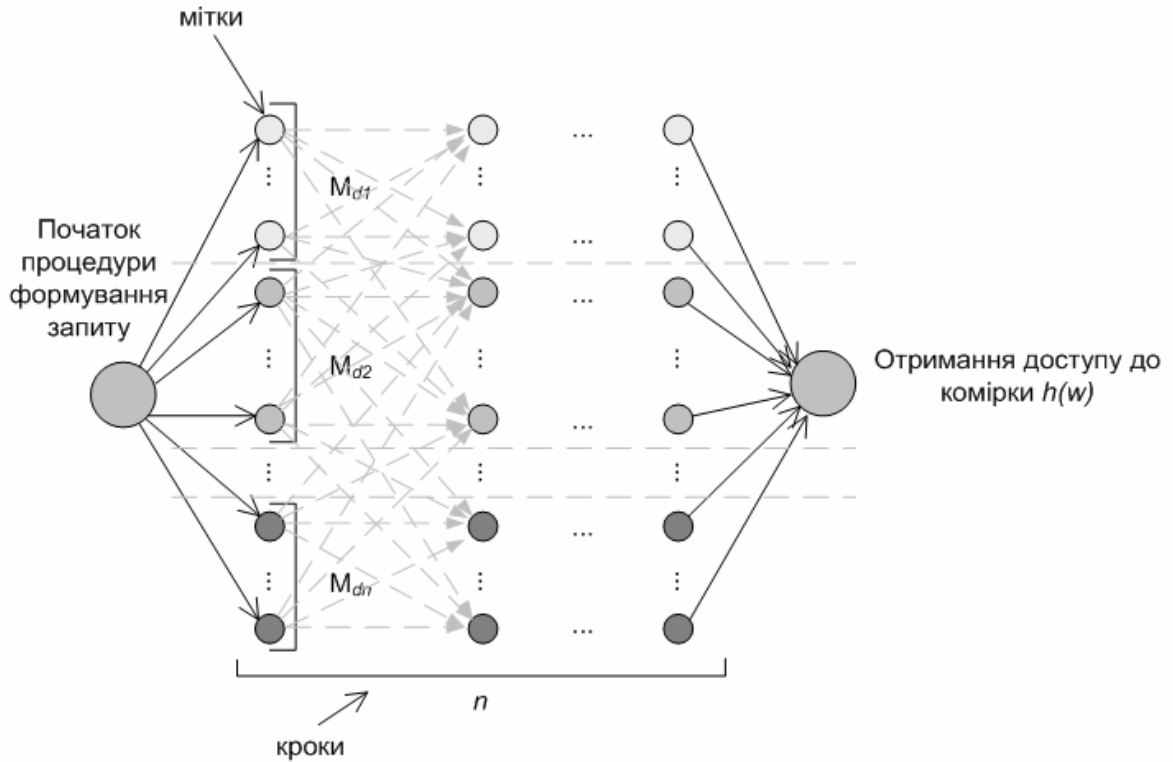


Рис. 2.16. Побудова запиту

Процедура побудови запиту користувача на отримання даних з багатовимірного кубу $H(D, M)$ може бути представлена у виді послідовності кроків, на кожному з яких користувач фіксує мітку $m_{d_i} \in M$ виміру $d_i \in D$. Зафіксувавши мітки в усіх вимірах, отримуємо доступ до комірки, що відповідає зафіксованим міткам.

Послідовність фіксації міток може бути представлена на сітьовому графі у вигляді послідовності ребер шляху $w = \{p_{jk}^0, p_{jk}^1, \dots, p_{jk}^n\}$, що відображає процес формування запиту.

Початкова вершина графу S_0 відповідає початку процедури формування запиту та відсутності фіксованих міток. На першому кроці ($n=1$) фіксується

мітка $m_{d_i} \in M$ виміру $d_i \in D$ багатовимірного кубу даних $H(D, M)$, що відповідає вершині s_{j_i} шару S_1 графу $G(S, P)$. На другому кроці фіксується мітка $m_{d_j} \in M \setminus M_{d_i}$ виміру $d_j \in D \setminus d_i$, так як вимір d_i вже був зафіксований. Процедура повторюється n разів, де n – кількість вимірів, які необхідно зафіксувати для отримання необхідного зрізу даних (у загальному випадку $n = |D|$).

Кінцева вершина $S_{0_{n+1}}$ відповідає закінченню процедури формування запиту та отримання доступу до комірки $h(w)$ багатовимірного кубу даних $H(D, M)$, де w - шлях запиту на графі $G(S, P)$.

Подані багатовимірні моделі даних є неповністю заповненими даними, що призводить до того, що отримана комірка $h(w)$ може виявитись порожньою. Таким чином, час затрачений на формування запиту буде затрачений марно, оскільки результатом запиту буде порожнє значення, що призводить до низької ефективності роботи з багатовимірним кубом даних.

Висновки до розділу

В розділі подана узагальнена технологія проведення аналізу даних, яка покладена у основу вирішення поставлених перед предметними галузями, шляхом комбінації методів інтелектуального та оперативного аналізу. Дослідженні моделі і методи які дозволяють спрогнозувати виникнення критичних ситуацій, що здійснюють негативний вплив на об'єкти предметних галузей. Описані моделі та методи вирішують одну з задач предметних галузей реєстрових систем та дозволяють попередити критичні ситуації, наслідки яких мають небажаний ефект.

РОЗДІЛ 3. РЕАЛІЗАЦІЯ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ ТЕХНОЛОГІЇ OLAP

3.1 Імітаційне моделювання основних характеристик системи підтримки прийняття рішень

Моделювання основних показників багатовимірного сховища даних, при різних технологіях збереження даних

Одним з найважливіших етапів при проектуванні СППР є визначення технології збереження багатовимірних даних з метою досягнення максимальної продуктивності. У даному розділі проведено моделювання роботи багатовимірних сховищ даних на основі СКБД Oracle g та здійснено аналіз отриманих результатів. На основі результатів моделювання визначено режим зберігання багатовимірної інформації.

У якості найбільш важливих критеріїв, при побудові багатовимірних структур даних, відібрані режим збереження та рівень агрегації даних. Режими зберігання Oracle g наведено у табл. 3.1.

Таблиця 3.1.

Режими збереження даних

Режим збереження	Опис
Реляційний OLAP (ROLAP)	Детальні факти та агрегати зберігаються на сервері реляційної БД.
Багатовимірний OLAP (MOLAP)	Дані – факти і агрегати зберігаються на OLAP - сервері у відповідному багатовимірному форматі.
Гібридний OLAP (HOLAP)	Дані – факти зберігаються на сервері реляційної БД, а агрегати – на OLAP сервері у відповідному багатовимірному форматі.

Агрегати представляють попередньо розраховані данні таблиці фактів для певних комбінацій кожного з вимірів. Агрегати використовуються для обробки запитів і створення додаткових агрегатів. Попередній розрахунок усіх можливих агрегатів призводить до значного збільшення об'єму інформації, що зберігається, але з іншого боку – розрахунок агрегатів у момент обробки запиту збільшує необхідний для цього час. Виконаємо порівняльний аналіз різних режимів збереження інформації та рівнів агрегації. Основні характеристики багатовимірних структур даних, які визначають результати моделювання, відображені на рис. 3.1.

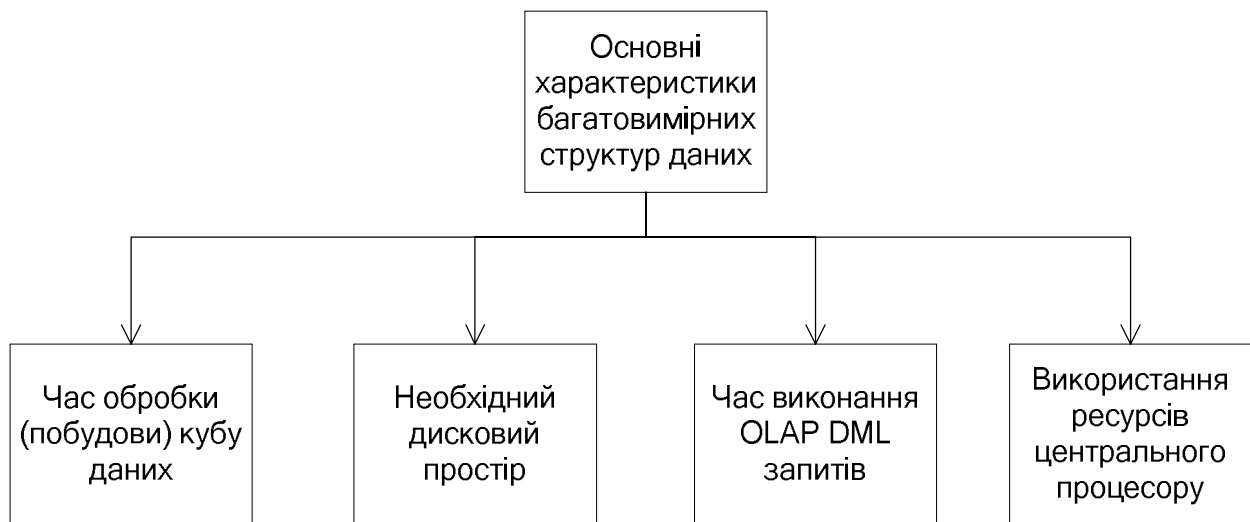


Рис. 3.1. Основні характеристики багатовимірних структур даних

Конфігурація тестової системи

Для проведення імітаційного моделювання було використано наступні апаратно - програмні ресурси:

- Процесор 2 x Intel Xeon 3.0 GHz.
- ОЗУ 4096 Мб.
- Жорсткий диск: 400 Gb SATA.
- ОС Windows.
- Oracle Database Enterprise Edition.

Запити для тестування

З метою проведення моделювання, до даних будуть поставлені наступні бізнес запити (табл. 3.2), що були сформовані за принципами максимального урізноманітнення процесів виконання запиту (предметна область – реєстр сільськогосподарських тварин).

Таблиця 3.2.

Запити для тестування

№	Запит
1	Яка кількість тварин здійснила переміщення з одного господарства у інше у віці від 3 до 4 років за поточний рік?
2	Яка середня кількість «забитої» ВРХ по господарствам за поточний рік?
3	Яка кількість народжених кіз по окремому господарству за поточний та попередній рік?
4	Яка максимальна кількість тварин потрапила до карантину за увесь час, по кожній області України та за кожний рік?
5	Яка кількість зареєстрованих тварин у розрізі областей, видів тварин та агентів - за поточний рік?
6	Яка кількість «жовтих» ветеринарних карток, була видана тваринам у розрізі областей та господарств за поточний рік?
7	Яка кількість державних ветеринарних установ по кожній області?

Побудова схеми «зірка»

Для отримання відповідей про ефективність того чи іншого методу збереження даних було побудовано схему «зірка» (рис. 3.2).

Структура складається з наступних таблиць:

- таблиця «**Тварина**» – містить відомості про тварин;
- таблиця «**Обстеження**» – містить відомості про обстеження тварини, захворювання;
- таблиця «**Господарство**» – містить основні відомості про господарство розведення та утримання тварин;

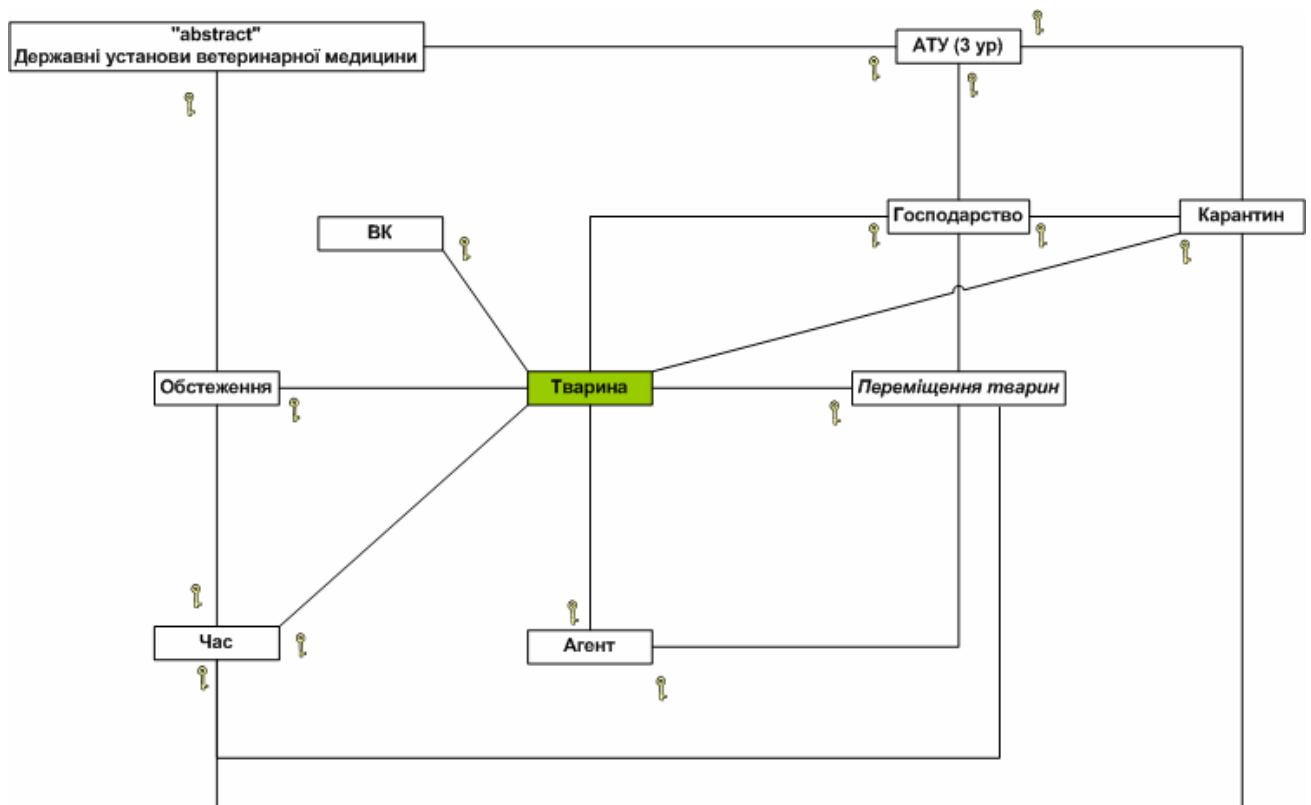


Рис. 3.2. Структура схеми «зірка»

- таблиця «*Переміщення тварин*» – містить відомості про переміщення тварин, посилання на господарства з яких тварина переміщується, період, тощо;
- таблиця «*Агент*» – містить відомості про агента, що провів реєстрацію тварини;
- таблиця «*Карантин*» – містить відомості про карантини, що виникли, сферу їх розповсюдження, вид тварин на які вони поширюються, тощо;
- таблиця «*Державні ветеринарні установи*» – містить відомості про ветеринарні установи;
- таблиця «АТУ» – містить відомості про адміністративно – територіальну одиницю до якої відноситься – господарство, ветеринарна установа, або поширюється зона карантину;
- таблиця «ВК» – містить відомості про ветеринарну картку, що була видана на тварину;

- службова таблиця «Час» – містить часові рамки необхідні для виконання аналізу у розрізі років, місяців, днів.

Кількість записів та розмір таблиць представлено у табл. 3.3.

Таблиця 3.3.

Вихідні характеристики моделювання

Таблиця	Кількість рядків	Розмір	Кількість рівнів у вимірі
Тварина	2 056 135	846.3 Мб	-
Обстеження	2 235 689	212.8 Мб	1
Переміщення	451 321	42.1 Мб	1
Господарство	687 214	268.2 Мб	1
Агент	2 568	3.2 Мб	1
Карантин	10 098	17,6 Мб	1
Державні ветеринарні установи	866	1.1 Мб	1
АТУ	40 843	4 Мб	3
Час	2880	2.1 Мб	3

Побудова OLAP кубів

З метою проведення імітаційного моделювання, було створено сховище даних, що містить 12 кубів однакового змісту, але з різними режимами збереження даних та рівнями агрегації. У побудованих кубах таблиця «тварина» представляє собою – міри (факти) кубу, а інші таблиці, що зображені на рис. 3.2 є вимірами (гранями) кубу. У табл. 3.4 зображено скорочену характеристику кожного з 12 кубів.

Таблиця 3.4.

Багатовимірні куби для моделювання

<i>Назва</i>	<i>Режим зберігання</i>	<i>Рівень агрегації (%)</i>
CubeM0	MOLAP	0
CubeM30	MOLAP	30
CubeM60	MOLAP	60
CubeM90	MOLAP	90
CubeR0	ROLAP	0
CubeR30	ROLAP	30
CubeR60	ROLAP	60
CubeR90	ROLAP	90
CubeH0	HOLAP	0
CubeH30	HOLAP	30
CubeH60	HOLAP	60
CubeH90	HOLAP	90

Модель імітаційного моделювання

Імітаційне моделювання здійснюється у відповідності до узагальненої моделі, принципова схема якої зображена на рис. 3.3.

Розглянемо основні складові та етапи проведення моделювання на основі описаної моделі. У циклі по усім режимам зберігання інформації (табл. 3.4) та, для кожного режиму, по усім напрямкам моделювання здійснюється:

- Виконання підготовчих операцій. Даний етап, в залежності від виду моделювання, може включати операції створення багатовимірної структури, з відповідними до режиму зберігання характеристиками, завантаження даних, тощо.
- Виконання тестового завдання. У відповідності до напрямку моделювання у рамках даного етапу може здійснюватись: побудова багатовимірної структури, завантаження даних, виконання запитів.

- Фіксація результатів моделювання. Збереження показників системи, що відображають результати відповідного напрямку імітаційного моделювання.
- Очищення СКБД. Повне фізичне знищення побудованих структур та завантажених даних, з метою дотримання однакових вихідних вимог при здійсненні моделювання.

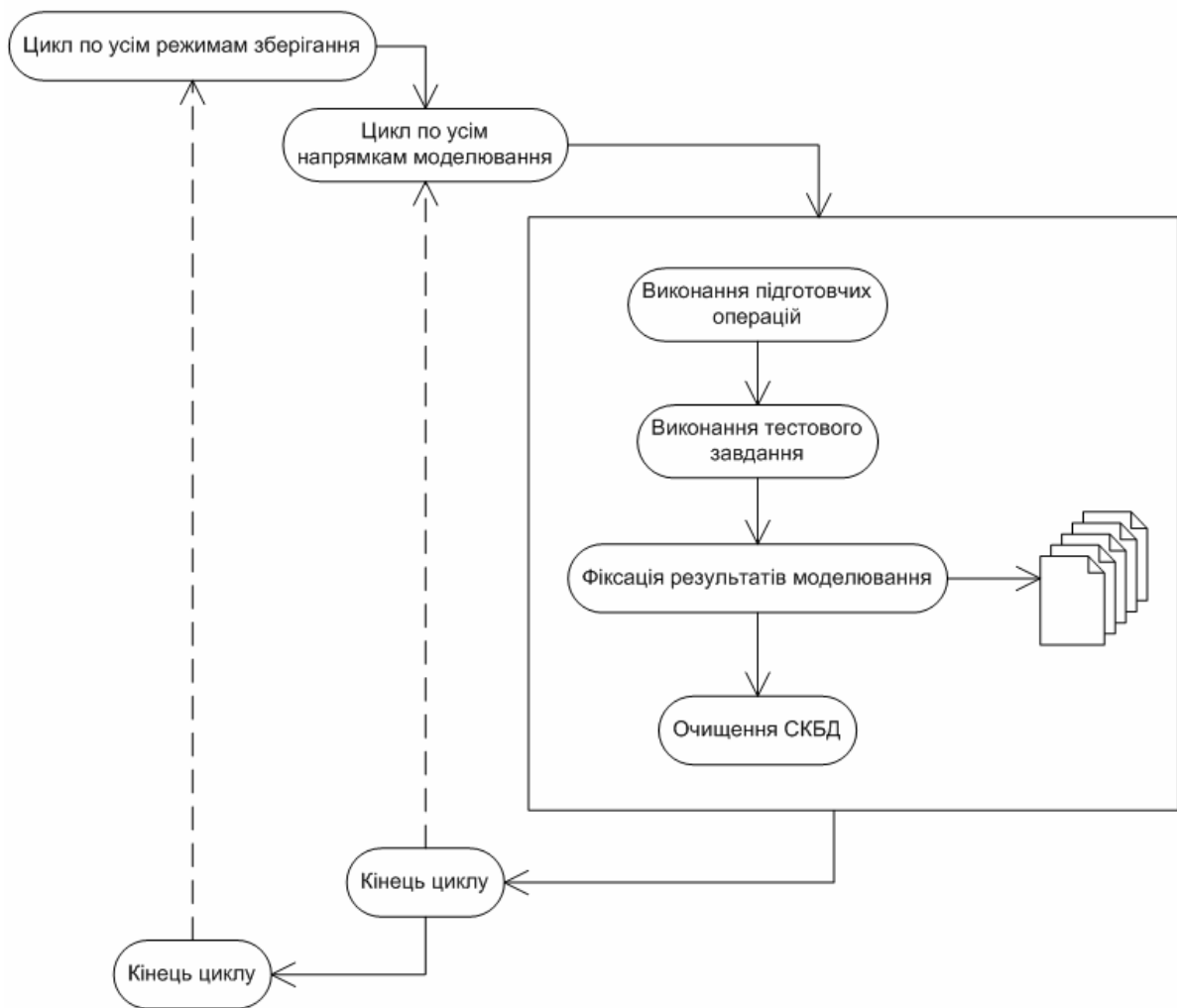


Рис. 3.3. Схема імітаційного моделювання

Слід зазначити, що для уникнення похибок та «шумів» при проведенні імітаційного моделювання, етапи моделі, що зображена на рис. 3.3, повторюються та розраховуються середні показники результатів моделювання.

Результати імітаційного моделювання

Час побудови (обробки) багатовимірного кубу. Результати зображені на рис. 3.4, були отриманні у процесі обробки ідентичних по структурі кубів з різними режимами зберігання інформації та рівнями агрегації.

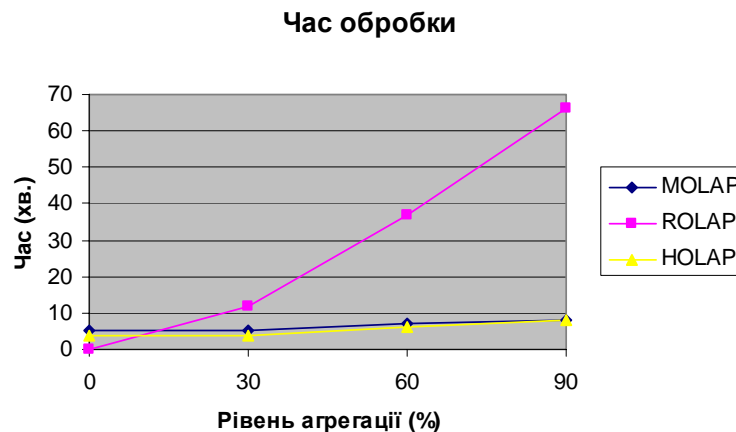


Рис. 3.4. Час затрачений на побудову багатовимірного кубу

З результатів моделювання, зображених на рис. 3.4, можливо зробити наступні висновки:

- при рівні агрегації 0%, ROLAP знадобилося найменша кількість часу для обробки кубу. При цьому данні таблиці фактів і вимірів в куб не додаються та агрегати не розраховуються;
- по мірі збільшення рівня агрегації, ROLAP по відношенню до MOLAP та HOLAP – витрачає усе більший час на обробку багатовимірного кубу;
- різниця між MOLAP та HOLAP у проміжку 30-60% незначна;
- час обробки MOLAP та HOLAP збільшується в проміжку 60-90%, але неістотно;
- час обробки ROLAP збільшується експоненційно у проміжку 60-90%.

Необхідний об'єм дискового простору для кожного режиму зберігання.

На рис. 3.5 зображено зміну необхідного дискового простору в залежності від рівня агрегації для кожного режиму зберігання.

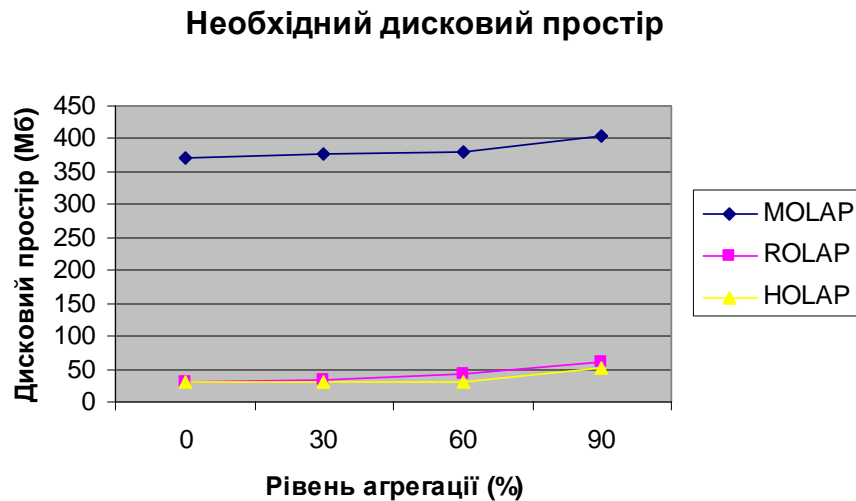


Рис. 3.5. Необхідний дисковий простір для збереження багатовимірному кубу

Виходячи з результатів моделювання, що зображені на рис. 3.5, можливо зробити наступні висновки:

- режим зберігання MOLAP вимагає більше дискового простору, ніж HOLAP чи ROLAP (куби MOLAP містять копії вихідних фактів та вимірів);
- різниця у кількості необхідного дискового простору при режимах ROLAP та HOLAP незначна у інтервалі 0-60% та збільшується по мірі наближення до рівня 90%;
- режим зберігання HOLAP використовує найменшу кількість дискового простору, що обумовлюється тим, що копії вихідних фактів та вимірів відсутні у базі даних OLAP, а агрегати зберігаються у оптимізованому багатовимірному форматі;
- режим зберігання ROLAP вимагає додаткового місця на диску, коли рівень агрегації переважає 30% та наближається до 90% (графік враховує об'єм дискового простору, необхідний для збереження агрегатів даних в реляційній структурі даних).

У табл. 3.5 зображено об'єм необхідного дискового простору для MOLAP кубу у порівнянні з вихідною схемою «зірка» (таблиці фактів та вимірів) у реляційній СКБД.

Таблиця 3.5

Порівняння необхідних об'ємів дискового простору

Рівень агрегації (%)	Дисковий простір для MOLAP кубу	Розмір схеми «зірка» (таблиці фактів та вимірів з індексами)	Ступінь стиснення даних при побудові кубів MOLAP
60	379	1397.4	72,8
90	405	1397.4	71

З табл. 3.5 очевидно, що дисковий простір необхідний для MOLAP кубу складає приблизно 18% від простору необхідного для схеми «зірка». Навіть при 90% рівні агрегації вдалося досягнути високого ступеню стиснення. Додатковий простір, необхідний для побудови MOLAP кубу, залежить від: кількості рівнів в вимірі, кількості мір (фактів) та типів даних.

Моделювання ефективності SQL та OLAP DML запитів. У табл. 3.6 зображено результати моделювання ефективності OLAP DML та SQL запитів, відповідно до багатовимірного MOLAP кубу та реляційної структури (кожен запит виконувався після перевантаження серверу, для гарантії того, що у кеші відсутні результати запитів).

Таблиця 3.6.

Результати моделювання ефективності запитів

Номер запиту (згідно таблиці)	Час, що затрачено на обробку OLAP DML запиту на OLAP сервері (використовуючи режим зберігання MOLAP та рівень агрегації 60%)	Час, що затрачено на обробку SQL запиту до реляційної СКБД.
1	9 сек.	84 сек.
3	5 сек.	62 сек.
6	7 сек.	65 сек.

З табл. 3.6 очевидно, що OLAP DML запити виконуються набагато швидше. Незважаючи на те, що подібне моделювання може здатись не дуже коректним та таким, що безпосередньо не відноситься до цілей розділу, його результати говорять про те, що використання OLAP DML запитів до Oracle OLAP істотно збільшує продуктивність роботи систем, оскільки OLAP куби зберігають попередньо розраховані агрегати, ефективність у OLAP середовищі значно вище, ніж при використанні реляційних структур.

Моделювання ефективності OLAP DML запитів при різноманітних режимах зберігання даних.

Розглянемо час обробки OLAP DML запитів, використовуючи режими зберігання MOLAP, ROLAP та HOLAP, різні рівні агрегації та «теплий» або «холодний» кеш. «Холодний» кеш свідчить про те, що перед виконанням запиту сервер було перевантажено і кеш не містить результатів попередніх запитів, у випадку «теплого» кешу, результати запитів збереглись у кеш пам'яті. Час було зафіксовано у секундах.

На рис. 3.6 зображено залежність середнього часу обробки запитів для MOLAP, ROLAP та HOLAP від рівня агрегації при «холодному» кешу.

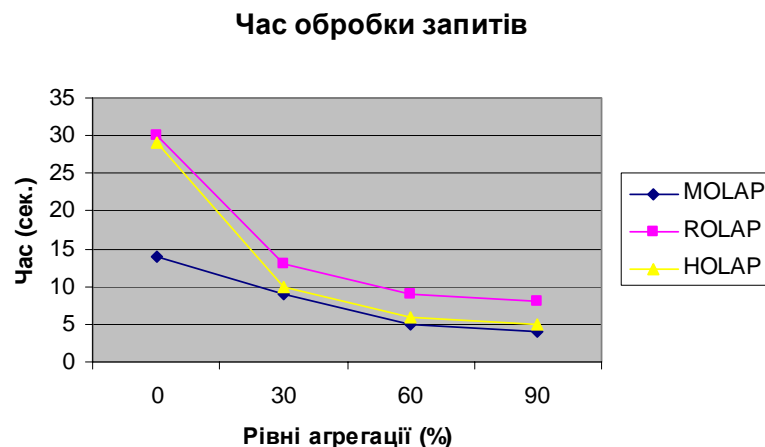


Рис. 3.6. Час обробки OLAP DML запитів при «холодному» кешу

Виходячи з результатів зображених на рис. 3.6, можливо зробити наступні висновки:

- MOLAP дозволяє досягнути найбільшої швидкості при обробці запитів, при чому продуктивність зростає при збільшенні рівня агрегації.
- Для усіх режимів збереження даних, при збільшенні рівня агрегації – збільшується продуктивність, але при збільшенні рівня агрегації з 60 до 90% приріст продуктивності є незначним для усіх режимів.

На рис. 3.7 зображена залежність середнього часу обробки запитів для MOLAP, ROLAP та HOLAP від рівня агрегації у випадку з «теплим» кешем.

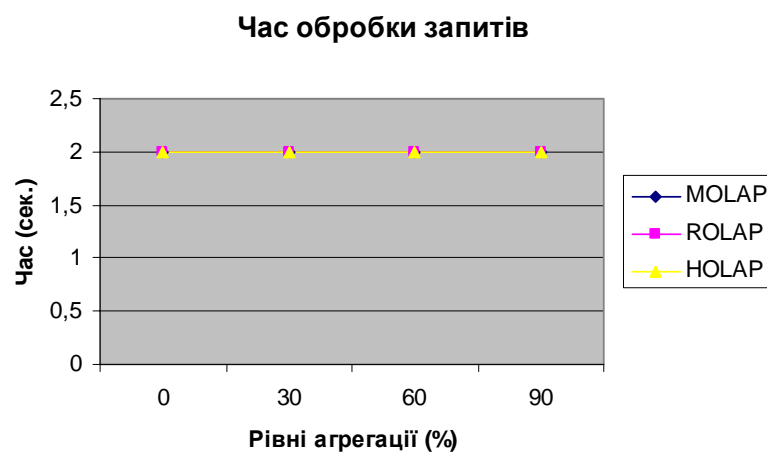


Рис. 3.7. Час обробки OLAP DML запитів при «теплому» кешу

З рис. 3.7 слідує: якщо результати запитів містяться у кеш пам'яті – його обробка виконується практично вмить (біля 2 секунд) для усіх запитів, незалежно від режиму зберігання та рівня агрегації.

Використання центрального процесора (CPU) при обробці запитів. На рис. 3.8 відображено результати моделювання часу роботи процесору при обробці запитів до MOLAP, ROLAP та HOLAP кубів.

З рис. 3.8 можливо зробити наступні висновки:

- MOLAP потребує значно меншого навантаження процесору.

- ROLAP та HOLAP використовують процесор у більшій мірі, це обумовлюється необхідністю проведення великої кількості розрахунків у реляційній частині СКБД. Відносна рівність HOLAP та ROLAP обумовлюється тим, що обрані для проведення моделювання запити не використовують велику кількість розрахованих агрегатів.

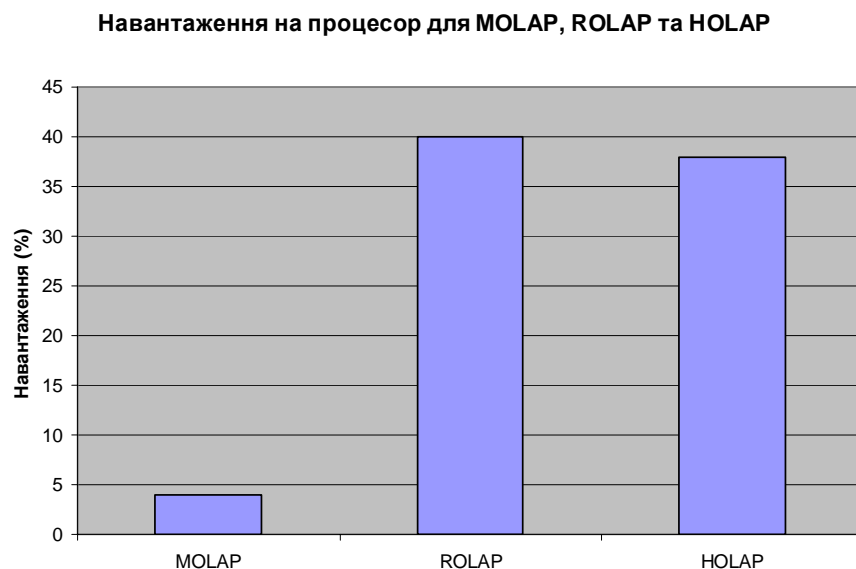


Рис. 3.8. Навантаження на центральний процесор

Таким чином, провівши моделювання роботи OLAP складової Oracle серверу при різних режимах зберігання (ROLAP, MOLAP, HOLAP) та різних рівнях агрегації, зобразимо результати моделювання на зведеному графіку (рис. 3.9), та зробимо узагальнені висновки:

- По мірі збільшення рівня агрегації, більше часу потребує процес побудови/перебудови ROLAP кубів по відношенню до MOLAP та HOLAP кубів.
- MOLAP вимагає більше дискового простору по відношенню до ROLAP чи HOLAP.
- У випадку режиму зберігання MOLAP, дисковий простір, що займає куб, складає приблизно 18 % від вихідних даних.

- Використання багатовимірних запитів (Oracle DML) надає можливість значно прискорити роботу системи.
- Максимальна швидкість обробки запитів досягається при використанні режиму зберігання даних - MOLAP.
- Режим зберігання MOLAP, менше навантажує процесор, оскільки джерелом навантаження є OLAP компонента реляційної структури.

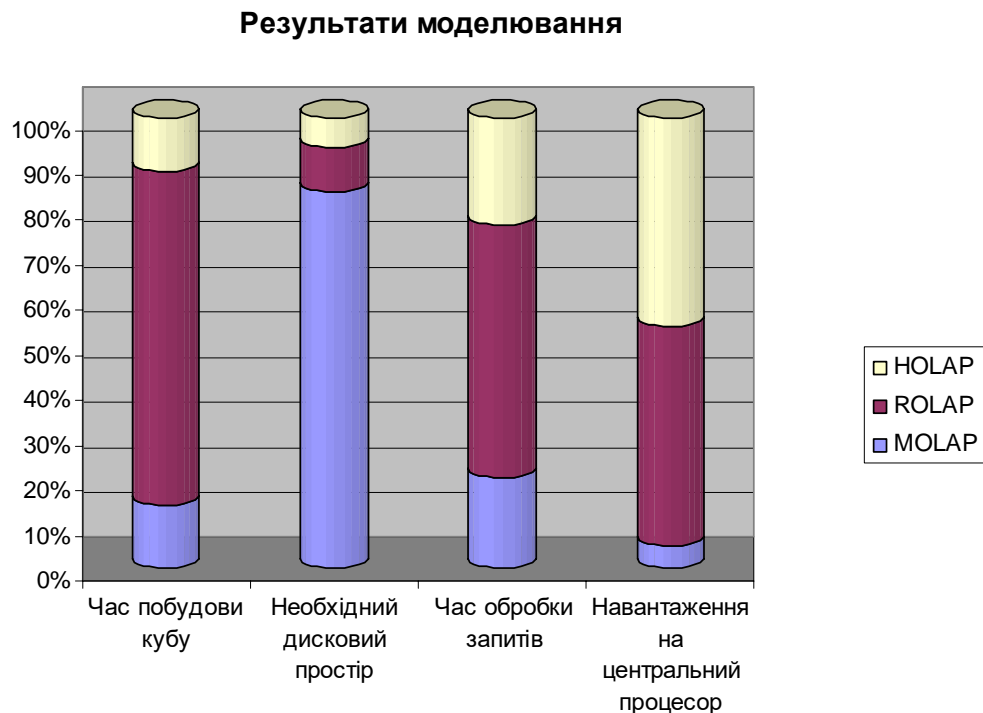


Рис. 3.9. Зведені результати моделювання (рівень агрегації – 60%)

Застосуємо отриманні результати моделювання при виборі режиму зберігання багатовимірних даних для проектування систем. Оскільки кожен з режимів має переваги по різних напрямкам моделювання, визначимо напрямки, що є пріоритетними (критичними) для систем та ті, що є менш важливими:

- Час побудови (обробки) кубу – даний параметр не є визначальним, оскільки побудова/перобудова кубу відбувається рідко, у регламентований (неробочий) час – тому не приймаємо даний параметр у якості визначального.

- Необхідний дисковий простір – даний параметр, при нинішньому рівні розвитку апаратних засобів зберігання інформації та показаних результатах для усіх режимів зберігання, також не є визначальним, оскільки забезпечення необхідного дискового простору для найбільш вимогливого режиму, враховуючи прогнозований ріст об’ємів даних, не ставить надзадач до програмно – апаратних засобів.
- Час обробки запитів – найбільш важливий параметр, оскільки характеризує швидкість роботи системи з особою, що приймає рішення, а враховуючи можливу одночасну роботу множин користувачів, дана залежність збільшується у рази. Тому, приймаємо даний параметр у якості найбільш критичного для функціонування систем.
- Навантаження на центральний процесор – даний параметр також є важливим з точки зору функціонування систем, оскільки одночасно робота множини користувачів, чинитиме доволі значне навантаження на процесори серверу СКБД. Тому, приймаємо даний параметр важливим для функціонування систем.

Виходячи з наведених характеристик напрямків моделювання, режим зберігання MOLAP, є найбільш ефективним для реалізації СППР.

Моделювання границь застосування методу підвищення ефективності конструктора запитів

Реалізація методу, який дозволяє підвищити ефективність побудови запиту користувача до багатовимірного сховища, що є компонентою конструктора запитів, вимагає додаткових розрахункових витрат, пов’язаних з визначенням значення оператора кількості непорожніх комірок для різних співвідношень обраних користувачем вимірів. Час розрахунку значення оператора, у свою чергу, залежить від кількості записів у БД (ступеня наповненості) та потужності апаратно – програмних засобів, що використовуються.

Зміст методу, що дозволяє підвищити ефективність запиту, полягає у скороченні часу, що витрачає користувач на отримання результатів запиту. Застосування даного методу виправдовує себе виключно у випадку, коли час, що витрачається системою на виконання розрахунку усіх необхідних значень оператора кількості непорожніх комірок - не переважає час обрання користувачем значення виміру (інакше користувач буде витрачати час на очікування відгуку системи, що веде до зниження ефективності роботи). Назвемо цей час критичним значенням та умовно визначимо його як t сек. (згідно проведених тестів, 7 секунд необхідно користувачу для вибору значення). Моделювання пропонується здійснити у відповідності до моделі, схема якої зображена на рис. 3.10.

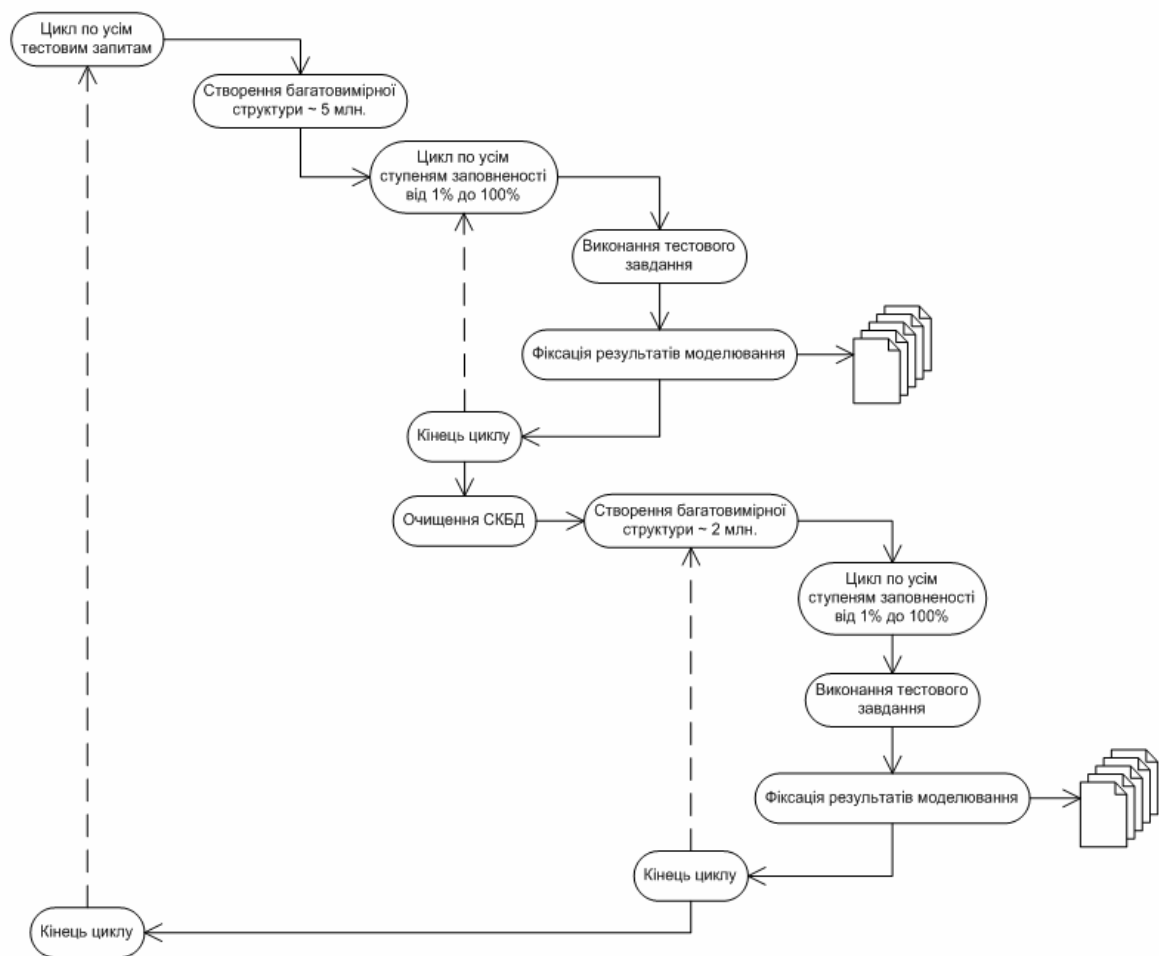


Рис. 3.10. Модель імітаційного моделювання

На рис. 3.11, 3.12 представлені результати проведеного моделювання, графіки, що характеризують залежність часу розрахунку значення оператора кількості непорожніх комірок від ступеню заповненості багатовимірного кубу даних для різних потужностей апаратних засобів. Використане програмно – апаратне забезпечення відображено у табл. 3.7, інформація про тестові набори даних у табл. 3.8.

Таблиця 3.7.

Програмно-апаратне забезпечення моделювання

	Програмно - апаратне забезпечення 1	Програмно - апаратне забезпечення 2
БД	Oracle	Oracle
ОС	ОС Windows 7	ОС Windows 7
Процесор	Pentium 4 2.0 ГГц	2 x Intel Xeon 3.0 GHz;
Оперативна пам'ять	1 Gb	4 Gb
Сумарна потужність	~1,5 GFPLS	~5 GFPLS

Таблиця 3.8.

Тестові набори даних

	Тест 1	Тест 2
Кількість записів (фактів)	~ 1 млн.	~ 2 млн.
Кількість вимірів	6	6

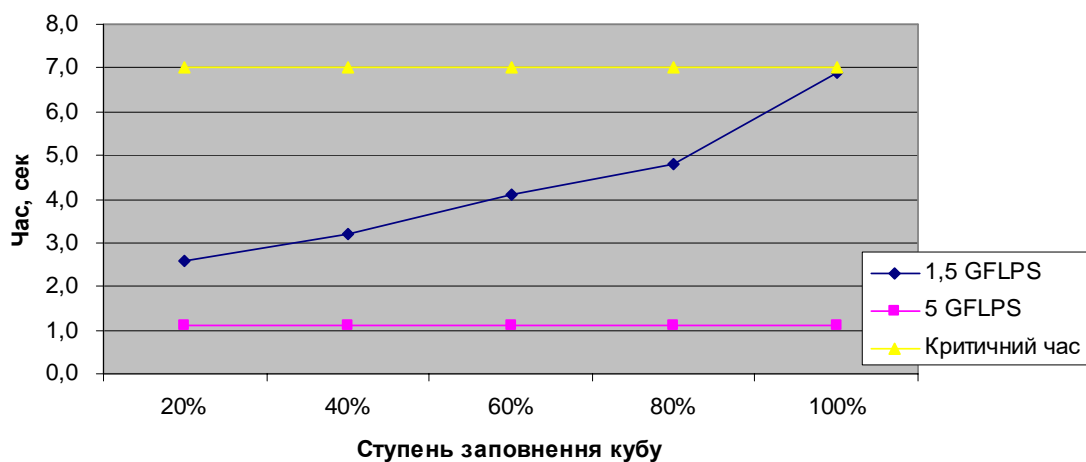


Рис. 3.11. Границі застосування методу, який підвищує ефективність запитів (~1 млн.)

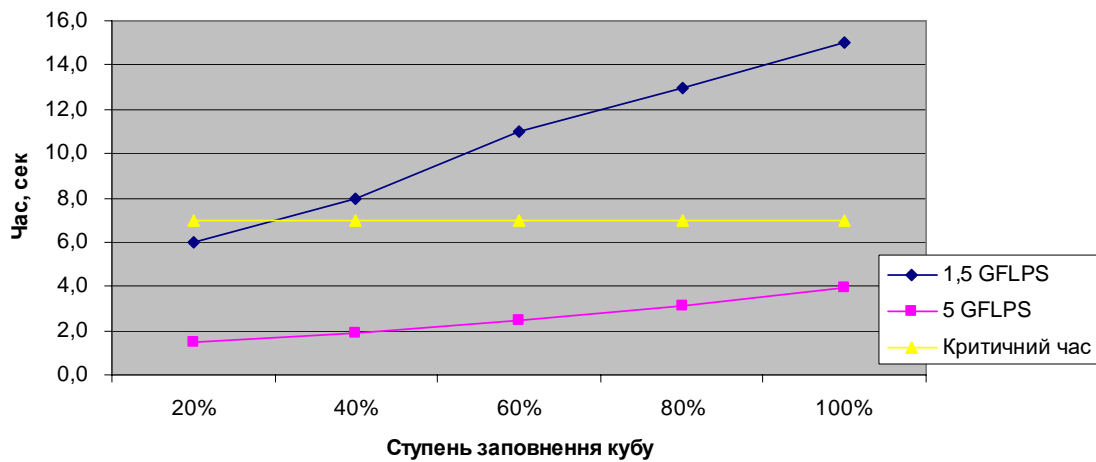


Рис. 3.12. Границі застосування методу, який підвищує ефективність запитів (~2 млн.)

За результатами моделювання можливо зробити наступні висновки:

- При максимальній кількості записів ~ 1 млн., час розрахунку оператора не перевищує критичного значення часу для обох конфігурацій апаратного забезпечення.
- У випадку максимальної кількості запитів ~ 2 млн., для високого рівня заповненості кубу (більше 40%) рекомендується використання апаратних засобів, що перевищують потужність у 1,5 GFLPS.

3.2 Архітектура систем та компонентів

Загальний дизайн архітектури

Призначення систем та функціональні вимоги до них вимагають територіально - розподіленого процесу внесення інформації, в основу якого доцільно покласти архітектурну модель з трьома рівнями: сервер бази даних – сервер додатків – клієнт.

Дане рішення має наступні переваги:

- високий рівень безпеки системи: СКБД повністю схована від сторонніх користувачів;
- гнучкість розподілення навантаження;

- централізоване оновлення клієнтського програмного забезпечення;
- моніторинг та логування.

В загальному вигляді схеми архітектури та технологічного дизайну систем представлено на рис. 3.13 та 3.14.

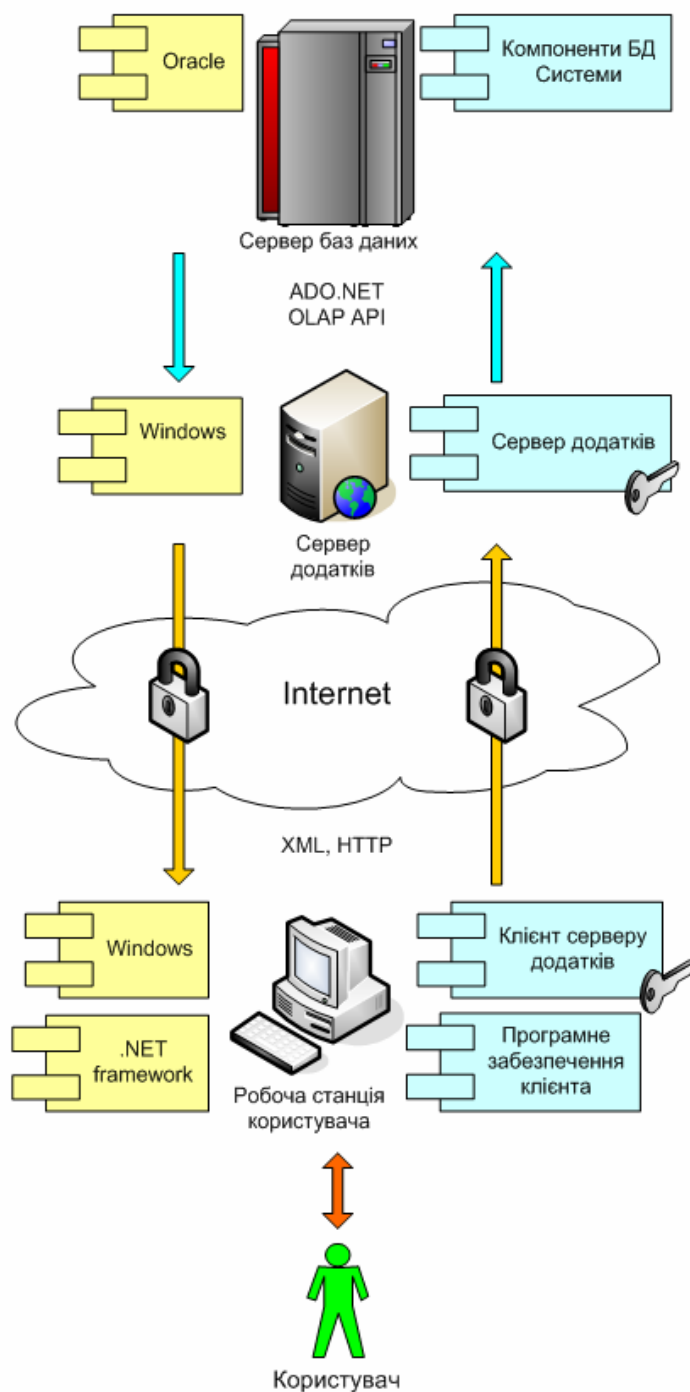


Рис. 3.13. Технологічний дизайн систем

В основу технологічного дизайну (рис. 3.13) програмного продукту покладено платформу Microsoft .NET та технології, що входять до її складу. Microsoft .NET Framework є платформою для побудови програмного забезпечення, яка містить великий набір базових бібліотек та підтримує декілька мов програмування високого рівня.

Платформа містить технології для створення програмного забезпечення різноманітного призначення та рівня складності, таких як програми із графічним інтерфейсом Windows, програмне забезпечення для використання у мережі Інтернет, підтримка взаємодії із різними джерелами даних, віддаленого доступу, тощо.

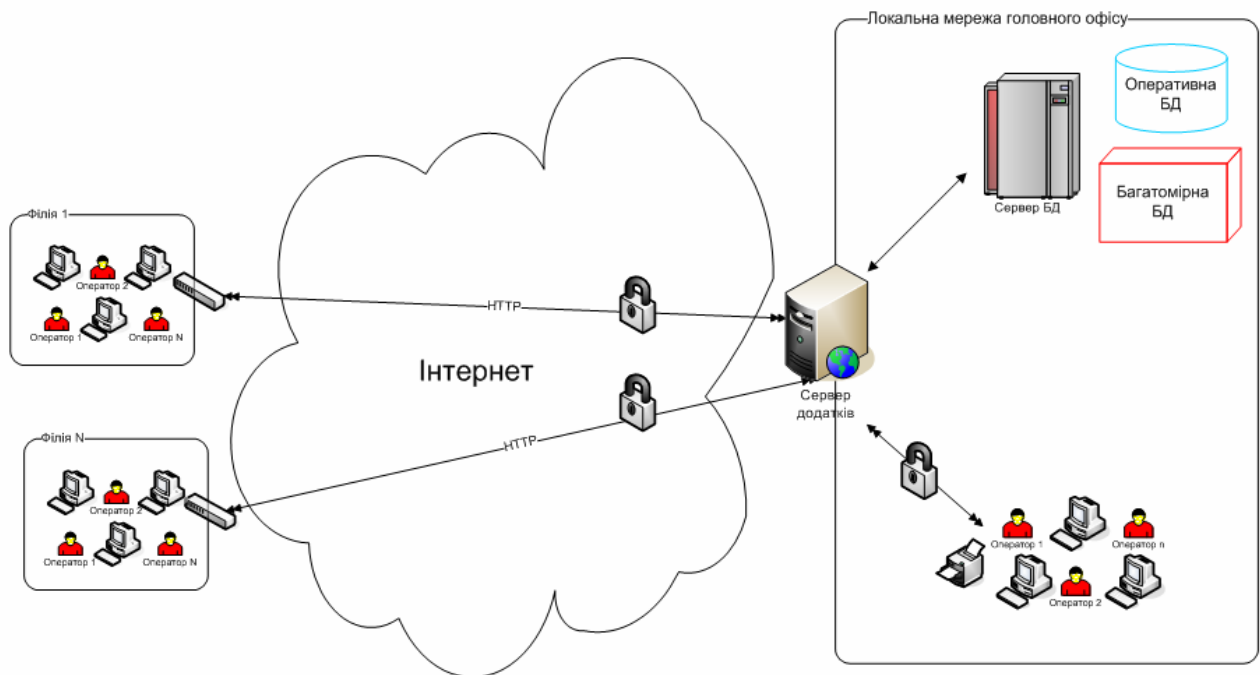


Рис. 3.14. Узагальнена схема архітектури систем

Використання даної платформи дозволить значно скоротити час розробки програмного забезпечення за рахунок широкого використання базових компонентів платформи. Завдяки використанню гнучкого середовища виконання програм, обрана платформа дозволяє розробляти безпечні та добре нарощувані програмні продукти.

Сервер додатків

Сервер додатків в запропонованій архітектурі є єдиним клієнтом серверу бази даних, доступ до якої з мережі Інтернет буде закрито.

Сервер додатків складається з наступних складових (рис. 3.15):

- Загальні компоненти – частина серверу додатків, що забезпечує стиснення та передачу даних, шифрування, контроль сесій, тощо.
- Додаток – складова серверу додатків, що виконує: реєстрацію методів бізнес логіки та делегує їм звернення, виконує перевірку прав користувачів, виконує серіалізацію та десеріалізацію контейнерів даних.
- Бізнес логіка – складова серверу додатків, що містить набір бізнес правил та методів, які відповідають за перевірку коректності даних, підтримку та перевірку цілісності даних, виконання специфічних перевірок, тощо.

Порядок взаємодії компонентів архітектури, в узагальненому вигляді можливо представити наступним чином:

- Програмне забезпечення клієнта запускається на робочій станції, з'єднується із сервером додатків та проводить авторизацію.
- Сервер додатків перевіряє привілеї користувача і в разі надання доступу до системи встановлює канал обміну даними із робочою станцією клієнта через мережу Інтернет. По цьому каналу, за протоколом НТТР, сервер додатків отримує шифровані запити від програмного забезпечення клієнта на доступ чи виконання певних операцій над даними.
- Отримавши запит, сервер додатків виконує його дешифрування, перевіряє достатність привілеїв клієнта і у разі успіху, викликає відповідні методи бізнес логіки.
- Методи бізнес логіки звертаються безпосередньо до бази даних, отримуючи чи модифікуючи інформацію, відповідно до запиту клієнта. Результати виконаних операцій передаються серверу додатків, який проводить шифрування та стиснення даних і відправляє їх до клієнта через мережу Інтернет.

- Отримавши відповідь від сервера, програмне забезпечення клієнта виконує розпакову та дешифрування даних і надає їх для обробки користувачеві. Сформувавши новий запит до Системи, програмне забезпечення клієнта виконує шифрування та стиснення даних і відправляє їх на сервер по встановленому каналу.

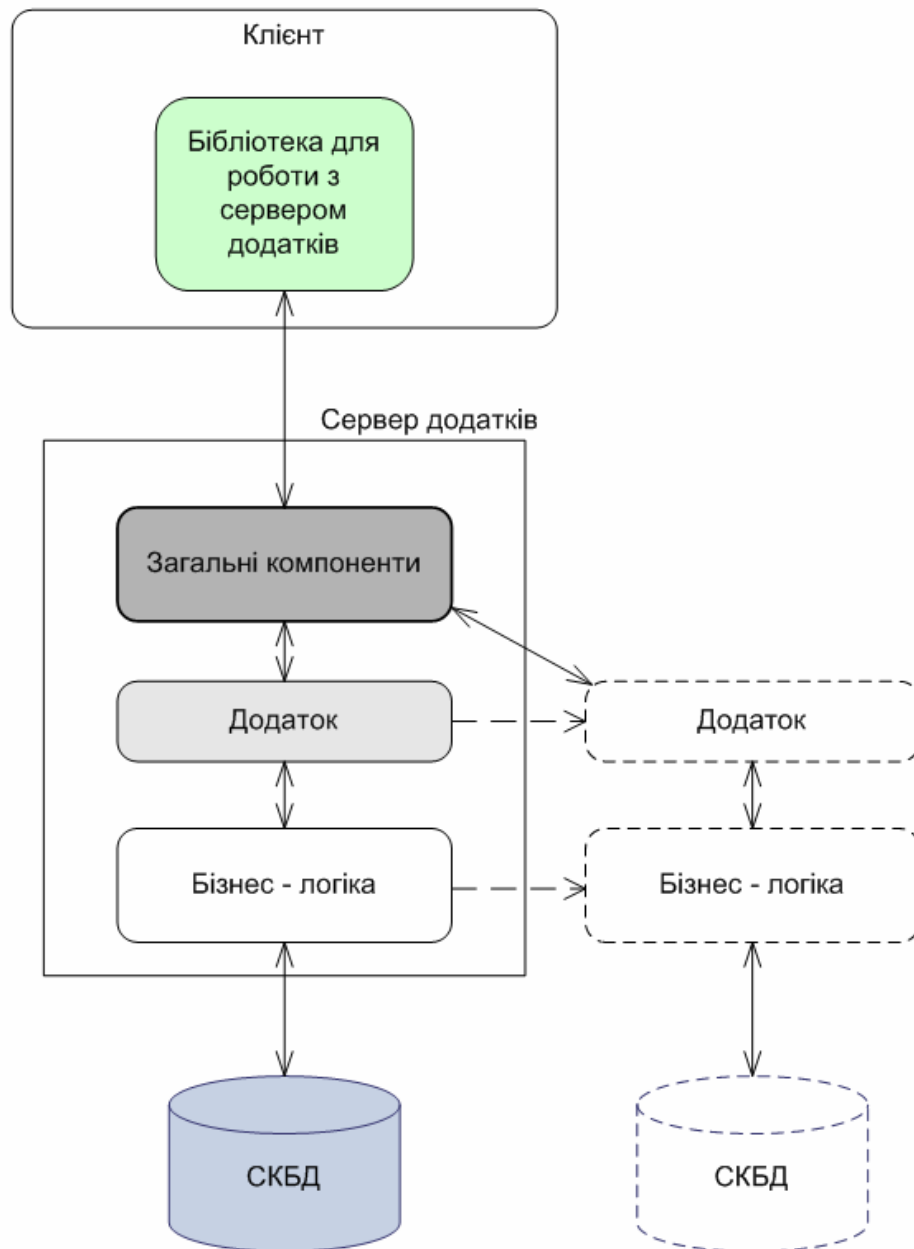


Рис. 3.15. Структура компонентів серверу додатків

Запропонована структура серверу додатків, дозволяє використання єдиного серверу додатків для різних систем, шляхом доповнення його

відповідними компонентами – «додаток» та «бізнес логіка» (як це зображено на рис. 3.15). Використання такого «уніфікованого» підходу, значно зменшує затрати часу на розробку та модифікацію інформаційних систем, а також збільшує гнучкість при їх розгортанні та впровадженні.

Сервер додатків призначений для виконання наступних основних функцій:

Передача даних між сервером бази даних та клієнтами. Бібліотека клієнтського програмного забезпечення оперує спеціальною ієрархічною структурою контейнер. Будь яка взаємодія з сервером додатків повинна відбуватись шляхом створення вхідного клієнтського контейнера та передачі його об'єкту сесії, де відбувається підготовка до передачі, передача до серверу та отримання відповіді. У загальному випадку схема взаємодії відображена на рис. 3.16.

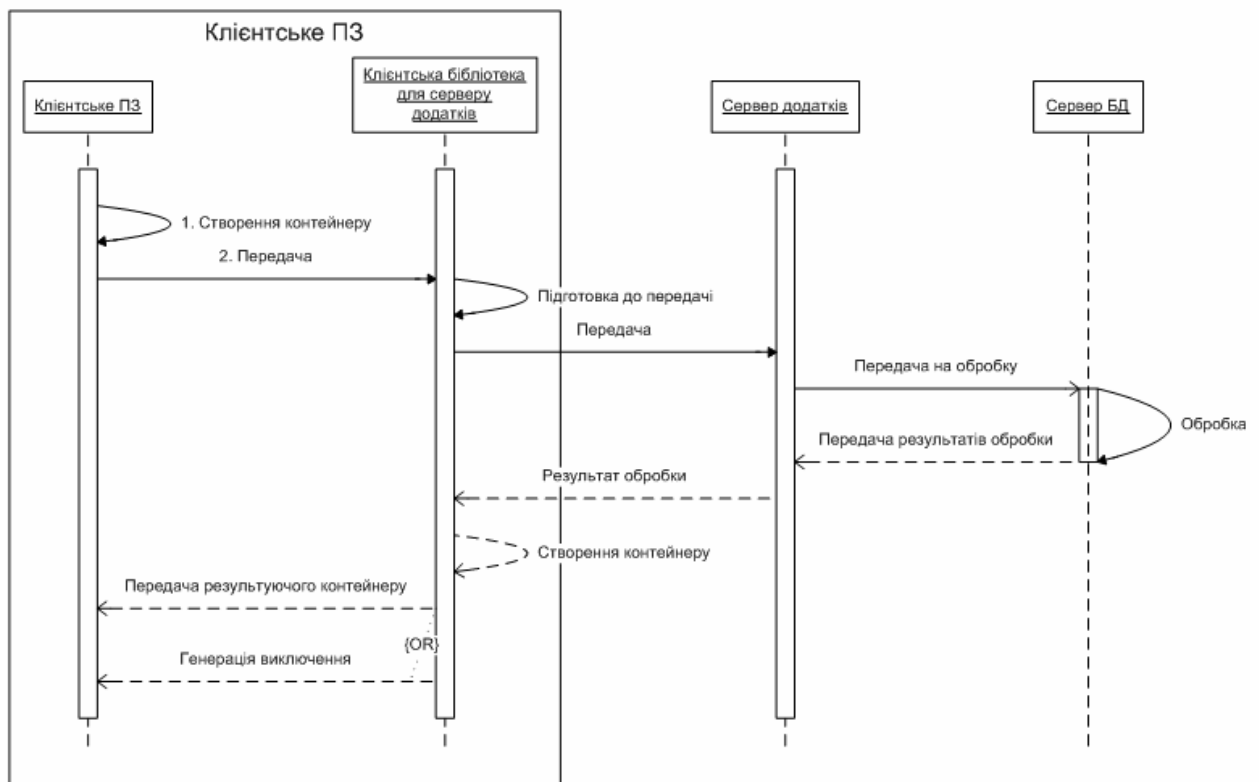


Рис. 3.16. Схема взаємодії складових системи

Підтримка сесій клієнта, з'єднаного з сервером. Клієнтське програмне забезпечення зберігає номер власної сесії, що отримує після виконання аутентифікації. Кожен раз, при надсиланні запитів до серверу додатків (окрім запитів аутентифікації) клієнтське ПЗ записує номер сесії у протокол обміну. Ніяких інших даних про користувача не передається, тобто пароль передається виключно один раз - при початку роботи.

Сесія достатньо стабільний показник, сервер додатків не втрачає відповідність номеру сесії та даних про підключення, навіть після перевантаження, таким чином перевантаження серверу не призводить до знищення сесій клієнтів. Підробити номер сесії або передати інший не є можливим, оскільки номер є доволі великий текстовий рядок, що генерується практично випадково.

Життєвий цикл сесії починається з першого входу до системи та закінчується при виконання операції виходу з системи, або сервер додатків має можливість її примусового завершення, у разі відсутності відповіді від клієнтського ПЗ, впродовж заданого проміжку часу.

Контроль версій клієнтського програмного забезпечення. При виконанні операції аутентифікації - клієнтська бібліотека серверу додатків, отримує від клієнтського програмного забезпечення його скорочену назву та номер версії у текстовому вигляді, після чого йде передачі цієї інформації до серверу додатків у відповідності до рис. 3.16, та визначення можливості роботи з даною версією клієнтського ПЗ. Якщо робота з даною версією клієнтського програмного забезпечення не підтримується сервером – клієнт не отримує можливість подальшої роботи з сервером.

Пакування даних. Модуль пакування підключається як бінарний декодер до протоколу передачі даних. Використовується метод пакування – zip. Засоби для пакування наявні у вигляді вільної бібліотеки сторонніх розробників.

Клієнтське програмне забезпечення.

Архітектура клієнтського програмного забезпечення побудована на основі компонентів, що представляють ядро системи, яке пов'язує інші частини програмного забезпечення (рис. 3.17). Ядро є самостійним компонентом та можливе функціонування програмного забезпечення без єдиної підсистеми.

Підсистеми реалізовані як доповнення, що використовують основні компоненти для інтеграції у програмне забезпечення та взаємодію з іншими підсистемами програмного забезпечення.

Однією з підсистем клієнтського програмного забезпечення, є підсистема виконання нерегламентованих запитів до багатовимірного сховища. Дана підсистема забезпечує користувачу інтерфейс доступу до багатовимірної моделі даних, забезпечуючи його можливостями зручного маніпулювання даними для виконання задач аналізу у різноманітних розрізах багатовимірної інформації.

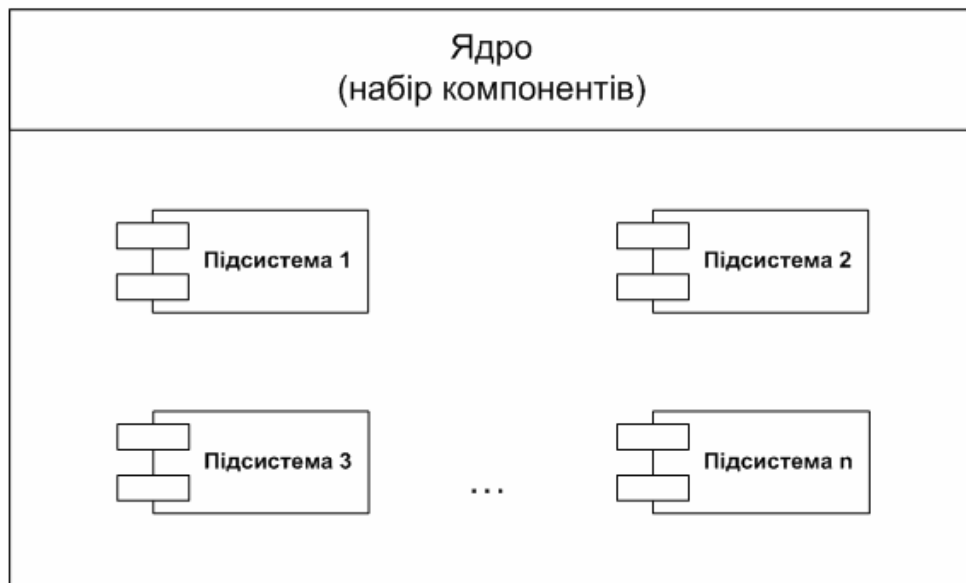


Рис. 3.17. Схема архітектури клієнтського програмного забезпечення

Сервер СКБД

СКБД є розташованим на спеціально виділеному сервері, сховищем даних та методів їх обробки. Управління базою даних Системи покладається на СКБД Oracle, що являє собою потужну систему управління різноманітними видами

даних, пов'язаними із будь-якими галузями людської діяльності. В ядро СКБД Oracle вбудовано новітні методи обробки і аналізу інформації, розвинена система захисту та підтримки цілісності даних, а також підтримка механізмів збереження та маніпулювання багатовимірними даними.

База даних складаються з двох складових:

- оперативної БД - в цій реляційній базі даних містяться усі дані системи, що накопичуються під час її функціонування. База даних знаходиться під керуванням системи управління базами даних Oracle, яка дозволяє поряд із даними зберігати також частину бізнес логіки – програмних методів обробки даних;
- багатовимірної БД (MOLAP) – багатовимірне сховище даних, що містить як деталізовані данні, для генерації кубів, так і агрегати у формі багатовимірних структур. Багатовимірна БД використовується для виконання оперативного та інтелектуального аналізу та виконання нерегламентованих запитів до даних.

Для реалізації бізнес логіки, на рівні бази даних, використовується мова PL/SQL, у якості мови маніпулювання багатовимірними даними – мова OLAP DML. Доступ до даних забезпечується технологіями ADO.NET та OLAP API, відповідно до реляційної та багатовимірної частини даних.

На рис. 3.18 розглянемо більш детально та дамо характеристику потоків інформації між серверами Системи.

- **Робота облікової частини системи.** Сервер додатків, отримуючи запит (виклик процедури) від клієнтського програмного забезпечення, ретранслює запит через ADO.Net до реляційної бази даних, виконує його (ініціює виконання процедури) та повертає результати запиту (при необхідності) до серверу додатків, а той, у свою чергу, до клієнтського програмного забезпечення.
- **Робота аналітичної частини системи.** Сервер додатків, отримуючи запит від клієнтського програмного забезпечення, ретранслює запит до багатовимірного сховища даних. При обробці запиту використовується

данні логічної моделі багатовимірного сховища (метаданні), що міститься у представленні активного каталогу («Active Catalog Views»). Результати запиту до багатовимірного сховища даних, через сервер додатків, повертаються для відображення на клієнтському програмному забезпеченні. Запити до багатовимірного сховища формуються на мові OLAP DML.

- **Завантаження/оновлення даних з реляційної у багатовимірну БД.** При створенні багатовимірної БД відбувається первинне завантаження даних з реляційної БД до багатовимірної. У регламентовані проміжки часу чи по запиту користувача, відбувається оновлення даних багатовимірної бази даних, оновленими даними з реляційної бази даних.

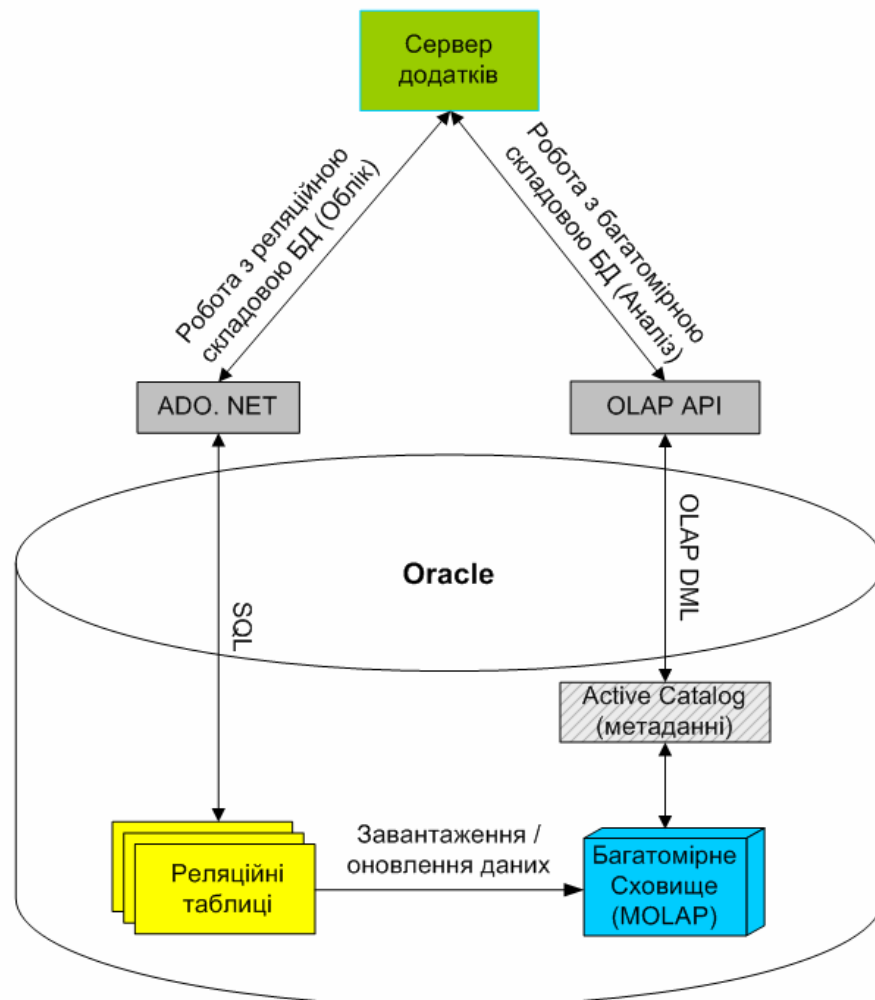


Рис. 3.18. Схема взаємодії серверу додатків та складовими СКБД

Для проведення аналізу даних та вирішення поставлених задач, системи повинні оперувати даними у багатовимірному представленні. Багатовимірні структури побудовані на основі багатовимірних моделей. Багатовимірні структури даних створені на основі облікових (реляційних) баз даних, шляхом представлення вихідної, реляційної структури у багатовимірному представленні.

Висновки до розділу

В розділі здійснено опис імітаційного моделювання основних характеристик багатовимірного сховища даних: час побудови багатовимірної структури; необхідний об'єм дискового простору; ефективність виконання DML запитів; навантаження на центральний процесор; в залежності від режимів збереження даних (MOLAP, ROLAP, HOLAP) та визначено оптимальний режим зберігання даних (MOLAP), для розглянутого кола задач СППР.

ВИСНОВКИ

В магістерській роботі проведений аналіз та здійснено класифікацію інформаційних систем орієнтованих на аналіз даних, здійснено дослідження сучасних технологій аналізу даних OLAP та Data Mining, оцінені можливості основних, найбільш поширених систем підтримки прийняття рішень, які ґрунтуються на використанні даних технологій, з огляду на їх функціональні можливості та готовність до вирішення поставлених задач. Результати проведених досліджень були використані: при здійсненні оцінки перспектив використання існуючих продуктів та компонентів інтелектуального та оперативного аналізу даних; при оцінці перспектив розробки власних компонентів та виборі ефективних шляхів і варіантів розвитку реєстрових систем.

Описана багатовимірна модель вихідних даних, здійснено формальний опис багатовимірної структури даних та методів обробки багатовимірної інформації, які покладені у основу реалізації компонентів СППР.

Представлена узагальнена технологія проведення аналізу даних, яка за рахунок взаємодії елементів оперативного та інтелектуального аналізу, дозволяє підвищити ефективність та гнучкість проведення аналізу великих об'ємів даних. Проведено дослідження границь застосування методу, який підвищує ефективність виконання запиту до неповністю заповненої багатовимірної структури даних, що надало можливість визначення вимог до співвідношення об'ємів даних, ступеню розрідженості та потужності апаратних засобів СППР. Здійснено імітаційне моделювання основних характеристик багатовимірної структури даних, за результатами якого визначено оптимальний режим збереження даних у багатовимірному представленні (MOLAP), виходячи з потреб та специфіки задач, що вирішуються.

Наведена концепція побудови архітектури СППР, досліджені складові архітектури, принципи серверної взаємодії реляційної та багатовимірної структур.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бек В.Л. Практикум з теорії статистики / В.Л. Бек, Г.В. Капленко. - К.: Новий Світ -2000, 2006. - 288 с.
2. Дюк В.А. Data Mining / В.А. Дюк, А.П. Самойленко. - СПб. 2001. - 487 с.
3. Лепа Є.К. Системи підтримки прийняття рішень / Є.К. Лепа, Є.К Міхеєв, В.В. Крініцин. - Х.: ВАТ «ХМД», 2007. - 236 с.
4. Мармоза А.Т. Теорія статистики / А.Т. Мармоза. - К.: Ельга, 2003. - 392 с.
5. http://www.businessobjects.com/pdf/company/products_services.pdf.
6. http://www.cognos.com/products/business_intelligence/analysis/features.html.
7. <http://www.redbooks.ibm.com/redbooks/pdfs/sg246599.pdf>.
8. Keen P. Decision support systems: a research perspective / P. Keen. - New York: Pergamon Press, 1980. - 734 с.
9. Keen P. Decision Support Systems: The next decades / P. Keen // Decision Support Systems. - New York: Pergamon Press, 1987. - № 3. - с. 253 - 265.
10. MicroStrategy OLAP Services, 2008. - режим доступу: http://www.microstrategy.com/Software/Products/Service_Modules/OLAP_Services/index.asp.
11. OLAP Council APB-1 OLAP Benchmark. Release II. November 1998., 2008. - режим доступу: <http://www.olapcouncil.org/>
12. Oracle OLAP., 2008. - режим доступу: <http://www.oracle.com/technology/products/bi/olap/index.html>.
13. Power D.J. A Brief History of Decision Support Systems / D.J. Power., 2003. - режим доступу: <http://DSSResources.COM/history/dsshhistory.html.research/bmarkco.htm>.
14. SQL Server Analysis Services Overview., 2007. - режим доступу: <http://www.microsoft.com/sql/technologies/analysis/overview.mspх>.
15. The OLAP Report., 2008. - режим доступу: <http://www.olapreport.com>.

16. TPC Benchmark R. Standard Specification. Revision 2.1.0. Transaction Processing Performance Council (TPC)., 2008. - режим доступу: <http://www.tpc.org/tpcr/default.asp>.
17. Уманець Т.В. Загальна теорія статистики /Т.В. Уманець. - К.: Знання, 2006. - 239 с.
18. Лепа Є.К. Системи підтримки прийняття рішень / Є.К. Лепа, Є.К Міхеєв, В.В. Крініцин. - Х.: ВАТ «ХМД», 2007. - 236 с.
19. Мармоза А.Т. Теорія статистики / А.Т. Мармоза. - К.: Ельга, 2003. - 392 с.
20. Scott Morton M. S. Management Decision Systems: Computer-based Support for Decision Making / M.S. Scott Morton. - Boston: Harvard University, 1971. - 372 с.
21. Sprague R. H. Building Effective Decision Support Systems / R.H. Sprague, E.D. Carlson. - NJ: Prentice-Hall, 1982. - 658 с.
22. Sprague R.H. A Framework for the Development of Decision Support Systems / R.H. Sprague // MIS Quarterly. - NJ: Prentice-Hall, 1980. - № 4. - с. 1 - 26.
23. 77. Alter S. L. Decision support systems: current practice and continuing challenges / S.L. Alter. - New York: Addison-Wesley Pub., 1980. - 458 с.
24. Bonczek R.H. Foundations of Decision Support Systems / R.H. Bonczek, C. Holsapple, A.B. Whinston. - New York: Academic Press, 1981. - 320 с.
25. BusinessObjects. Overview of products and services. [Електронний ресурс], 2008. - режим доступу: http://www.businessobjects.com/pdf/company/products_services.pdf.
26. Codd E.F. A Relational Model of Data for Large Shared Data Banks / E.F. Codd // Comm. of the ACM. - New York: Academic Press, 1970. - №6. - с. 377-387.
27. Codd E.F. Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate / E.F. Codd, S.B. Codd, C.T. Salley. - New York: E.F.Codd & Associates, 1993. - 540 с.
28. Cognos PowerPlay: Features and Benefits. [Електронний ресурс], 2008. - режим доступу: http://www.cognos.com/products/business_intelligence/analysis/features.html.

29. Davis G. Management Information Systems: Conceptual Foundations, Structure, and Development / G. Davis. - New York: McGraw-Hill, 1974. - 378 с.
30. DB2 OLAP Server V8.1: Using Advanced Functions. [Электронный ресурс], - режим доступа: <http://www.redbooks.ibm.com/redbooks/pdfs/sg246599.pdf>.
31. Druzdzel M.J. Decision Support Systems. Encyclopedia of Library and Information Science / M.J. Druzdzel, R.R. Flynn. - Washington: A. Kent, Marcel Dekker, Inc., 1999. - 546 с.
32. Edwards J.S. Expert Systems in Management and Administration - Are they really different from Decision Support Systems? / J.S. Edwards // European Journal of Operational Research, 1992. - №. 61. - с. 114 - 121.
33. Eom H. Decision Support Systems Applications Research: A Bibliography (1971-1988) / H. Eom, S. Lee // European Journal of Operational Research, 1990. - № 46. - с. 333 - 342.
34. Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation, October 2000. [Электронный ресурс], 2000. - режим доступа: <http://www.w3.org/TR/2000/REC-xml-20001006>.
35. Finlay P. Introducing decision support systems / P. Finlay. - Mass.: Blackwell Publishers, 1994. - 490 с.
36. Ginzberg M.I. Decision Support Systems: Issues and Perspectives // Processes and Tools for Decision Support / M.I. Ginzberg, E.A. Stohr. - Amsterdam: North-Holland Pub.Co, 1983. - с. 129 - 143.
37. Golden B. Decision Insight Systems: A Critical Evaluation / B. Golden, A. Hevner, D.J. Power // Computers and Operations Research. - Mass.: Addison-Wesley Pub. Co, 1986. - №2. - с. 287 - 300.
38. Holsapple C. Decision Support Systems: A Knowledge-based Approach / C. Holsapple, A. Whinston. - Minneapolis: West Publishing Co., 1996. - 586 с.
39. Hyperion Essbase Analytics Features. [Электронный ресурс], - режим доступа: http://dev.hyperion.com/products/essbase_analytic_services/index.cfm?id=hdnnav&link=other_prod_hyp_essbase.

40. IBM DB2 OLAP Server V8.1 -New Enhancements Available. [Електронний ресурс]. - режим доступу: http://www.ibm.com/common/ssi/rep_ca/2/897/ENUS202-172/ENUS202-172.PDF.
41. Дарвен Х. Основы майбутніх систем баз даних. Третій маніфест / Х. Дарвен, К. Дейт., 2004. - 656 с.
42. Parsaye K. Surveying Decision Support: New Realms of Analysis // Database Programming and Design. – 1996. – № 4.
43. Інформаційні технологія збереження та обробки даних [Електронний ресурс]. - режим доступу: https://pns.hneu.edu.ua/pluginfile.php/211290/mod_resource/content/1/%D0%A2%D0%B5%D0%BC%D0%B0%203.pdf