

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ДИЗАЙНУ
КАФЕДРА ІНФОРМАТИКИ І КОМП'ЮТЕРНИХ НАУК

Пояснювальна записка

до дипломної бакалаврської роботи

на тему:

**Методологія створення бізнес-аналітичних структур даних засобами
Microsoft SQL Server**

Виконала: студентка 4 курсу, групи 4КН
спеціальності 122 «Комп'ютерні науки»

Журба Н.О

Керівник: к.т.н., доцент Вишемирська С.В.

Рецензент: к.т.н., доцент Огнева О.Є.

Факультет	<u>Інформаційних технологій та дизайну</u>
Кафедра	<u>Інформатики і комп'ютерних наук</u>
Рівень вищої освіти	<u>перший (бакалаврський) рівень</u>
Галузь підготовки	<u>12 «Інформаційні технології»</u> (шифр і назва)
Освітньо-професійна програма	<u>Комп'ютерні науки</u> (назва)
Спеціальність	<u>122 «Комп'ютерні науки»</u> (шифр і назва)

ЗАТВЕРДЖУЮ

**Завідувач кафедри ІКН,
професор**

_____ **В.І. Литвиненко**

« ____ » _____ **2021 року**

**ЗАВДАННЯ
НА ДИПЛОМНУ РОБОТУ СТУДЕНТА**

Журба Наталія Олегівна

(прізвище, ім'я, по батькові)

1. Тема роботи: **Методологія створення бізнес-аналітичних структур даних засобами Microsoft SQL Server**

керівник роботи: **Вишемирська Світлана Вікторівна, кандидат технічних наук, доцент кафедри інформатики і комп'ютерних наук**

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

затверджені наказом ХНТУ від 26.11. 2020 р. № 644-с

2. Строк подання студентом роботи: 04.06.2021

3. Вихідні дані до роботи

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити): Вступ. Розділ 1. Аналітичний огляд літературних та інших джерел. Розділ 2. Теоретична частина. Аналіз предметної області. Розділ 3. Практична частина. Розділ 4. Охорона праці.

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

Таблиць – 2,

Формул – 0,

Рисунків – 8.

6. Консультанти розділів роботи:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
4	к.с.-г.н., доцент Малєєв В.А.		

7. Дата видачі завдання: 08.02.2021 р.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів дипломної роботи	Строк виконання етапів роботи	Примітка
1	Аналітичний огляд літературних джерел за тематикою бакалаврської роботи	10.02.2021- 01.03.2021	
2	Обґрунтування проблеми та аналіз методів і засобів її вирішення	02.03.2021- 15.03.2021	
3	Системний аналіз об'єкта дослідження та предметної області	16.03.2021- 01.04.2021	
4	Використання бізнес-аналітичних структур. Інтелектуальний аналіз даних	02.04.2021- 10.04.2021	
5	Методи та засоби розв'язання поставленої задачі	11.04.2021- 21.04.2021	
6	Розробка програми	01.05.2021- 10.05.2021	
7	Аналіз можливих небезпечних і шкідливих факторів, охорона праці	11.05.2021- 21.05.2021	
8	Написання та оформлення пояснювальної записки до дипломної роботи	22.05.2021- 04.06.2021	

Студентка

(підпис)

Н.О. Журба

(прізвище та ініціали)

Керівник роботи

(підпис)

С.В. Вишемирська

(прізвище та ініціали)

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	7
ВСТУП.....	8
РОЗДІЛ 1	11
АНАЛІТИЧНИЙ ОГЛЯД ЛІТЕРАТУРНИХ ТА ІНШИХ ДЖЕРЕЛ	11
1.1 Інтелектуальний аналіз даних	11
1.2 Системи підтримки прийняття рішень.....	13
1.2.1 Сховища даних	17
1.2.2 Моделі сховищ даних	20
1.2.3 Способи структурування сховища даних	20
1.2.4 Інструменти та методи збору, обробки та завантаження даних (ETL)	21
1.2.5 Багатовимірна база даних та засоби аналізу OLAP	21
1.2.6 Джерело даних.....	24
1.2.7 Подання джерела даних.....	25
1.3 Міри	26
1.3.1 Аналіз даних (Data Mining)	27
1.3.2 Класи систем Data Mining	27
1.3.3 Методи.....	30
1.4 Кластерний аналіз.....	31
1.4.1 Алгоритм k-середніх (k-means).....	32
1.4.2 Перевірка якості кластеризації	33
1.4.3 Байєсовські мережі.....	34
1.4.4 Штучні нейронні мережі	35
1.4.5 Навчання нейронних мереж	36
1.4.6 Завдання інтелектуального аналізу даних	38
1.4.7 Інструменти DataMining	40
1.5 Мова структурованих запитів SQL.....	40
1.6 Безпека SQL для бізнесу	45
1.7 Висновок до розділу.....	46

ТЕОРЕТИЧНА ЧАСТИНА. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	47
2.1 Опис предметної області	47
2.2 Реляційна СКБД.....	47
2.2.1 Структурована таблиця	49
2.2.2 Основні ключі	50
2.3 Цілісність даних.....	50
2.3.1 Цілісність на рівні домену.....	50
2.3.2 Цілісність на рівні відносин	51
2.3.3 Цілісність зовнішніх ключів (цілісність на рівні БД)	51
2.4 Microsoft SQL Server	52
2.4.1 Використані технології.....	53
2.5 Створення бази даних	55
2.5.1 Робота з даними [18]	56
2.5.2 Коли використовувати власний клієнт SQL Server [18]	56
2.6 Висновок до розділу.....	57
РОЗДІЛ 3	58
ПРАКТИЧНА ЧАСТИНА.....	58
3.1 Постановка задачі. Побудова моделі “сутність - зв’язок”	58
3.3 Створення діаграми в Microsoft SQL Server	61
3.4 Створення SQL розширених запитів	62
3.5 Створення і використання представлень (Views)	64
3.5.1 Оператор CREATE VIEW.....	64
3.6 Створення БД в MS SQL SERVER	65
3.7 Компоненти Business Intelligence Development Studio	65
3.7.1 Служби SQL Server Integration Services (SSIS)	66
3.7.2 Заповнення сховищ даних та баз даних.....	67
3.7.3 Служби Analysis Services.....	68
3.7.4 Масштабованість і продуктивність	69

3.8 Висновок до розділу.....	71
РОЗДІЛ 4	72
ОХОРОНА ПРАЦІ.....	72
4.1 Характеристика робочого місця	72
4.2 Санітарно-гігієнічні умови праці.....	72
4.3 Висновок до розділу.....	76
ВИСНОВОК	77
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	79
ДОДАТКИ	83

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

Скорочення, термін, позначення	Пояснення
SQL (Structured Query Language)	Мова структурованих запитів
СППР	Система підтримки прийняття рішень
БД	База даних
СКБД	Система керування базами даних
СД	Сховище даних
DSS (Decision Support Systems)	Система підтримки прийняття рішень
BI (Business Intelligence)	Бізнес аналітика
DM (Data Mining)	Інтелектуальний аналіз даних
OLTP (Online Transaction Processing)	Обробка транзакцій системи в режимі реального часу
OLAP (Online Analytical Processing)	Аналітична обробка даних в режимі реального часу
DW (Data Warehouse)	Сховища даних
ETL (Extract, Transform, Load)	Інструменти та методи збору, обробки та завантаження даних
SSIS (SQL Server Integration Services)	Платформа для створення рішень з інтеграції даних та робочих процесів
SSAS (SQL Server Analysis Services)	Служба , що допомагає аналізувати великий об'єм даних
UDM (Unified Dimensional Model)	Уніфікована багатовимірна модель

ВСТУП

Підприємницькі інформаційні системи зазвичай включають програми, призначені для всебічного багатовимірного аналізу даних, їх динаміки, тенденцій тощо. Зрештою, такий аналіз призначений для полегшення прийняття рішень. Часто ці системи називають системами підтримки прийняття рішень.

Неможливо прийняти будь-яке управлінське рішення, не маючи необхідної інформації, як правило, кількісної. Це вимагає створення сховища даних (Data warehouses), тобто процесу збору, перегляду та попередньої обробки даних з метою надання користувачам отриманої інформації для статистичного аналізу (а часто також створення аналітичних звітів).

Бізнес-аналітика - це область на стику інформаційних технологій, баз даних, алгоритмів видобутку даних, математичної статистики та методів візуалізації.

Менеджери приймають рішення, а завдання бізнес-аналітиків - зробити все можливе, щоб ці рішення були оптимальними та вчасно.

Метою роботи є дослідження математичних моделей, методів та інструментів бізнес-аналітики в СКБД SQL Server.

Актуальність теми. Тема створення аналітичної системи обробки даних важлива, оскільки сьогодні інструменти бізнес-аналітики використовуються майже у всіх сферах людського життя:

- в Інтернет-технологіях інструменти Business Intelligence використовуються для побудови систем рекомендацій для інтернет-магазинів та для вирішення проблеми персоналізації відвідувачів веб-сайту;
- у торгівлі технології Business Intelligence дозволяють ефективно просувати товари та безпосередньо управляти попитом на різні товари за допомогою аналізу ринкових кошиків;

- у телекомунікаціях засоби Business Intelligence допомагають вирішити проблему прибутковості та ризику клієнтів, а також проблему захисту від шахрайства;

- у промисловому виробництві прикладом використання технології Business Intelligence може бути прогнозування якості продукції залежно від вимірюваних параметрів технологічного процесу;

- прикладом використання засобів бізнес-аналітики в медицині є побудова діагностичної системи або перевірка ефективності хірургічного втручання;

- класичним прикладом застосування методів аналізу даних на практиці є вирішення проблеми можливої неплатоспроможності клієнтів у банківському секторі.

Мета і задачі дослідження. Метою даної роботи є створення аналітичної системи обробки даних, що охоплює всі основні компоненти систем бізнес-аналітики в СКБД SQL Server.

Для досягнення цієї мети в роботі вирішуються такі **завдання**:

- 1) огляд архітектури СКБД SQL Server та підсистеми Business Intelligence;

- 2) опис та аналіз сфер застосування інструментів Business Intelligence, таких як багатовимірний аналіз даних та видобуток даних;

- 3) огляд мовних інструментів Business Intelligence;

- 4) експериментальне застосування отриманої інформації.

Об'єктом нашого дослідження є масив бізнес-даних.

Предмет дослідження – інструменти бізнес-аналітики системи управління базами даних SQL Server.

Апробація результатів бакалаврської дипломної роботи була на міжнародній науковій конференції ISDMCI'2021 з публікацією тез: Zhurba N., Vyshemyrska S. Methodology for the creation of business intelligence structures

using Microsoft SQL Server / Intellectual Systems for Decision Making and Problems of Computational Intelligence: Conference Proceedings. – Kherson: PP Vyshemirsky V. S., 2021. P. 24.