

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ІВАНА ПУЛЮЯ  
ФАКУЛЬТЕТ КОМП'ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ І ПРОГРАМНОЇ  
ІНЖЕНЕРІЇ  
КАФЕДРА КОМП'ЮТЕРНИХ НАУК

**ГНУТЕЛЬ МАКСИМ МИХАЙЛОВИЧ**

УДК 004.9

**МЕТОДИ І ЗАСОБИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ АНАЛІТИЧНОГО  
ОПРАЦЮВАННЯ BIGDATA В ПРОЕКТАХ КЛАСУ «РОЗУМНЕ МІСТО»**

122 «Комп'ютерні науки»

**Автореферат**  
дипломної роботи на здобуття освітнього ступеня «магістр»

Тернопіль  
2018

Роботу виконано на кафедрі комп'ютерних наук Тернопільського національного технічного університету імені Івана Пулюя Міністерства освіти і науки України

**Керівник роботи:** доктор технічних наук,  
професор кафедри комп'ютерних наук  
**Пасічник Володимир Володимирович,**  
Тернопільський національний технічний університет  
імені Івана Пулюя,

**Рецензент:** кандидат технічних наук,  
доцент кафедри комп'ютерно-інтегрованих технологій  
**Золотий Роман Захарійович,**  
Тернопільський національний технічний університет  
імені Івана Пулюя,

Захист відбудеться 24 грудня 2018 р. о 9<sup>00</sup> годині на засіданні  
екзаменаційної комісії № 1 у Тернопільському національному технічному  
університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56,  
навчальний корпус №1, ауд. 702



## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми роботи** зумовлена зростанням вимог до адаптації аналітичної обробки та необхідністю її оптимізації в умовах роботи з розподіленими системами «розумних міст».

**Мета роботи:** підвищення ефективності обробки інформації (BigData) в прогресивних проектах «Розумних міст» за рахунок адаптації методів для аналітичної обробки інформації до роботи в міському гетерогенному середовищі.

**Об'єкт, методи та джерела дослідження:** Процес обробки інформації (BigData) в розподілених інформаційних системах «Розумних міст».

Основні методи дослідження загальної задачі – це методи системного аналізу та моделювання. В якості інформаційних джерел використовуються наукові публікації та інтернет джерела.

### **Наукова новизна отриманих результатів:**

- Досліджено адаптований метод аналітичної обробки муніципальних колекцій великих даних, що дозволяє прискорити швидкість обробки при роботі в розподіленому гетерогенному середовищі проектів класу «розумне місто».
- Запропоновано прототип сервісу для аналітичного опрацювання інформації (BigData), котрий дозволяє видобувати приховані в муніципальній інформації корисні інформаційні набори.

### **Практичне значення отриманих результатів.**

Теоретичні розробки виконані в дипломній роботі готові до програмної реалізації муніципального сервісу для аналітичного опрацювання великих за обсягом інформаційних колекцій «Розумного міста».

**Апробація.** За результатами досліджень проведених в рамках магістерської роботи зроблено дві доповідні на VI науково-технічній конференції «Інформаційні моделі, системи та технології» 12-13 грудня 2018 року з публікацією тез доповідей.

**Структура роботи.** Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 8 частин, висновків, переліку посилань та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 155 арк. формату А4, графічна частина – 20 слайдів презентації.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

**У вступі** проведено огляд сучасного стану інформаційних технологій великих даних в проектах «Розумних міст» та охарактеризовано основні завдання, які необхідно вирішити.

**В першому розділі** розглянуто концепт «Розумне місто» та концепцію «Великих даних». Визначено роль BigData в «Розумному місті». Наведено приклади застосування великих даних в «розумному місті». Проведено аналіз проблем опрацювання різнотипної інформації в проектах «розумних міст». Проаналізовано концепції роботи з Великими даними в проектах класу «Розумне місто». Виконано



аналіз програмного забезпечення для роботи з Великими даними в проектах «розумних міст». Досліджено особливості перетворення NoSQL в інші формати.

**В другому розділі** описано методи аналітичного опрацювання Big Data. Зокрема, описано алгоритми аналітичної обробки даних застосовні для проектів «Розумних міст». Проаналізовано алгоритми кластеризації. Подано формальне визначення кластеризації та її етапи. Розглянуто алгоритми асоціації. Описано процеси пошуку асоціативних правил та їх числові характеристики.

**В третьому розділі** досліджено сучасні технології підвищення продуктивності опрацювання великих за обсягом колекцій даних в проектах класу «Розумне місто». Розглянуто розподілені системи зберігання даних застосовні для проектів класу «Розумне місто», такі як Google File System, Hadoop File System, Google BigTable, HBase. Проаналізовано моделі програмування і технології розподіленої обробки. Зокрема такі як модель програмування MapReduce, платформа Apache Hadoop та Технологія Microsoft Dryad. Досліджено питання оптимізації аналітичної обробки.

**В четвертому розділі** подано хмарне управління, аналіз та представлення BigData як сервісу для проектів класу «розумне місто». Досліджено Hadoop в хмарі. Подано архітектуру Amazon Web Services. Розглянуто архітектуру Amazon Elastic MapReduce.

**В спеціальній частині** виконано формування мети експерименту. Описано ефекти збільшення загального розміру даних та збільшення вузлів, котрі проявилися в процесі виконання експерименту.

**В розділі «Обґрунтування економічної ефективності»** розраховано основні техніко-економічні показники проведених досліджень методів і засобів підвищення ефективності аналітичного опрацювання BigData в проектах класу «Розумне місто».

**В частині «Охорона праці та безпека в надзвичайних ситуаціях»** описано електромагнітні випромінювання та вимоги електробезпеки до приміщень з ЕОМ. Розглянуто забезпечення функціонування Державної системи моніторингу довкілля, як складова частина національної інформаційної інфраструктури, сумісної з аналогічними системами інших країн. Досліджено оцінку дії радіоактивного забруднення місцевості після ядерного вибуху на виробничу діяльність муніципальних підприємств, установ та організацій.

**В частині «Екологія»** досліджено методи узагальнення екологічної інформації, котрі використовуються на третьому етапі екологічних досліджень. Проаналізовано екологічну відповідальність та організацію «Зеленого офісу».

**У загальних висновках щодо дипломної роботи** описано прийняті в проекті технічні рішення і організаційно-технічні заходи, які забезпечують виконання завдання на проектування; оригінальні технічні рішення, прийняті автором в процесі роботи; технічні рішення роботи, які можуть бути впроваджені у виробництво;

В додатках до пояснювальної записки приведено ксерокопії тез доповідей.

В графічній частині подано тему, мету, об'єкт та предмет дослідження. Подано завдання до дипломної роботи. Описана чотирьохрівнева структура системи управління «Розумного міста» на основі великих даних та операційна діаграма «Розумного міста». Наведена таблиця порівняння SQL та NoSQL в контексті BigData. Розглянуті рівні BigData як сервісів хмарних послуг. Розглянута порівняльна характеристика OLAP. Подано архітектури Google BigTable,



MapReduce, модель роботи Apache Hadoop, архітектуру Microsoft Dryad, структури Hadoop та AWS. Описані основні результати та висновки.

## **ВИСНОВКИ**

В процесі виконання дипломної роботи освітнього рівня «магістр», в першому розділі:

- Розглянуто концепт «Розумне місто».
- Висвітлено концепцію «Великих даних».
- Визначено роль BigData в «Розумному місті».
- Наведено приклади застосування великих даних в «розумному місті».
- Проведено аналіз проблем опрацювання різнотипної інформації в проектах «розумних міст».
- Проаналізовано концепції роботи з Великими даними в проектах класу «Розумне місто».
- Виконано аналіз програмного забезпечення для роботи з Великими даними в проектах «розумних міст».
- Досліджено особливості перетворення NoSQL в інші формати.

В другому розділі описано методи аналітичного опрацювання Big Data. Зокрема:

- Описано алгоритми аналітичної обробки даних застосовні для проектів «Розумних міст».
- Проаналізовано алгоритми кластеризації. Подано формальне визначення кластеризації та її етапи.
- Розглянуто алгоритми асоціації. Описано процеси пошуку асоціативних правил та їх числові характеристики.

В третьому розділі досліджено сучасні технології підвищення продуктивності опрацювання великих за обсягом колекцій даних в проектах класу «Розумне місто».

А саме:

- Розглянуто розподілені системи зберігання даних застосовні для проектів класу «Розумне місто», такі як Google File System, Hadoop File System, Google BigTable, HBase.
- Проаналізовано моделі програмування і технології розподіленої обробки. Зокрема такі як модель програмування MapReduce, платформа Apache Hadoop та Технологія Microsoft Dryad.
- Досліджено питання оптимізації аналітичної обробки.

В четвертому розділі подано хмарне управління, аналіз та представлення BigData як сервісу для проектів класу «розумне місто»:

- Досліджено Hadoop в хмарі.
- Подано архітектуру Amazon Web Services.
- Розглянуто архітектуру Amazon Elastic MapReduce.

В спеціальній частині:

- Виконано формування мети експерименту.
- Описано ефекти збільшення загального розміру даних та збільшення вузлів, котрі проявилися в процесі виконання експерименту.



В шостому розділі дипломної роботи розраховано основні техніко-економічні показники проведених досліджень.

В сьомому розділі дипломної роботи описано електромагнітні випромінювання та вимоги електробезпеки до приміщень з ЕОМ. Розглянуто забезпечення функціонування Державної системи моніторингу довкілля, як складова частина національної інформаційної інфраструктури, сумісної з аналогічними системами інших країн. Досліджено оцінку дії радіоактивного забруднення місцевості після ядерного вибуху на виробничу діяльність муніципальних підприємств, установ та організацій.

У восьмому розділі досліджено методи узагальнення екологічної інформації, котрі використовуються на третьому етапі екологічних досліджень. Проаналізовано екологічну відповідальність та організацію «Зеленого офісу».

## **СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ**

1. Семантичні моделі даних в проектах класу «Розумне місто» / [Гнутель М.М. та ін.]. // Матеріали VI Міжнародної науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя, (Тернопіль, 12 – 13 грудня 2018 р.). – Тернопіль: Тернопільський національний технічний університет імені Івана Пулюя – 2018. – С. 15.

2. Ієрархічна масштабована База даних інформаційних ресурсів «Розумні міста» / [Гнутель М.М. та ін.]. // Матеріали VI Міжнародної науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя, (Тернопіль, 12 – 13 грудня 2018 р.). – Тернопіль: Тернопільський національний технічний університет імені Івана Пулюя – 2018. – С. 53.

## **АНОТАЦІЯ**

Дипломна робота присвячена аналітичному опрацюванню методів і засобів підвищення ефективності аналітичного опрацювання BigData в проектах класу «Розумне місто».

В першому розділі дипломної роботи розглянуто два сучасні інформаційно-технологічні концепти «розумне місто» та «великі дані».

В другому розділі дипломної роботи проведено аналітичний огляд алгоритмів аналітичної обробки записів застосовних для проектів класу «розумне місто». Проаналізовано алгоритми аналітичної обробки даних застосовні для проектів «Розумних міст». Досліджено алгоритми кластеризації та асоціації.

В третьому розділі дипломної роботи розглянуто сучасні технології підвищення продуктивності опрацювання великих за обсягом колекцій даних в проектах класу «Розумне місто». Досліджено розподілені системи зберігання даних застосовні для проектів класу «Розумне місто».

В четвертому розділі висвітлено хмарне управління, аналіз та представлення BigData як сервісу для проектів класу «розумне місто».

В розділі «Спеціальна частина» описано формування мети експерименту. Досліджено ефекти збільшення загального розміру даних та збільшення вузлів.



В повному обсязі виконано розділи «Обґрунтування економічної ефективності», «Охорона праці та безпека в надзвичайних ситуаціях» та «Екологія».

Об'єкт дослідження. Процес обробки інформації (BigData) в розподілених інформаційних системах «Розумних міст».

Предмет дослідження. Методи обробки інформації (BigData) та платформи для організації роботи з великими об'ємами даних «Розумного міста».

*Мета і задачі дослідження.* Метою магістерської дипломної роботи є підвищення ефективності обробки інформації (BigData) в прогресивних проектах «Розумних міст» за рахунок адаптації методів для аналітичної обробки інформації до роботи в гетерогенному міському обчислювальному середовищі.

Основні результати:

- Проведено аналіз проблем опрацювання різнотипної інформації в проектах «розумних міст».
- Виконано аналіз програмного забезпечення для роботи з Великими даними в проектах «розумних міст».
- Досліджено методи аналітичного опрацювання Big Data.
- Розглянуто сучасні технології підвищення продуктивності опрацювання великих за обсягом колекцій даних в проектах класу «Розумне місто».
- Висвітлено хмарне управління, аналіз та представлення BigData як сервісу для проектів класу «Розумне місто».

**Ключові слова:** ВЕЛИКІ ДАНІ, БАЗИ ДАНИХ, ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ІНФОРМАЦІЙНІ МОДЕЛІ, ОПРАЦЮВАННЯ, РОЗУМНЕ МІСТО.

## ANNOTATION

The thesis is devoted to analytical work on methods and means of increasing the effectiveness of analytical processing of BigData in the projects of the "Smart City" class.

In the first section of the thesis two modern information technology concepts "smart city" and "large data" are considered.

In the second section of the thesis an analytical review of algorithms for analytical processing of records applicable to "smart city" projects is carried out. Analyzed algorithms of analytical data processing are applicable for "Smart cities" projects. The algorithms of clustering and association are investigated.

The third section of the thesis deals with modern technologies for increasing the productivity of processing large volumes of data collections in the projects of the "Smart City" class. Distributed storage systems are investigated for Smart City projects.

The fourth section covers cloud management, analysis, and presentation of BigData as a service for smart city projects.

In the section "Special part" describes the formation of the purpose of the experiment. The effects of increasing the total size of data and increasing the nodes were studied.

The sections "Justification of economic efficiency", "Labor protection and safety in emergencies" and "Ecology" are executed in full.

Object of study. The process of processing information (BigData) in distributed intelligence systems "Smart cities".



Subject of study. Methods of processing information (BigData) and platforms for the organization to do with large volumes of data "Smart city".

The purpose and tasks of the study. The aim of the master's thesis is to increase the effectiveness of information processing (BigData) in the progressive projects of "Smart cities" by adapting methods for analytical processing of information to do in a heterogeneous urban computing environment.

Main results:

- The analysis of problems of working out of various types of information in projects of "smart cities" is carried out.
- An analysis of the software for working with Great data in the projects of "smart cities".
- The methods of analytical processing of Big Data are investigated.
- Modern technologies of increasing the productivity of processing large volumes of data collections in the projects of the "Smart City" class are considered.
- Illustrated cloud management, analysis and presentation of BigData as a service for Smart Classroom projects.

Keywords: BIGDATA, DATABASES, INFORMATIONAL TECHNOLOGIES, INFORMATION MODELS, STRUCTURE, SMART CITY.