

Обами активно використовували big data для аналізу настроїв виборців та коригування програми кандидата);

- процес планування міст та координація процесів на визначеній конфігурації місцевості (соціальні мережі та геолокаційні сервіси представляють величезні обсяги інформації, аналіз якої використовується при вирішенні прикладних задач містобудування, проектування транспорту, виявлення та координації надзвичайних ситуацій тощо);

- обробка процедур в медицині (big data допомагають запобігти розвитку хвороб на ранній стадії завдяки аналізу серцево-судинного тиску, пульсу, дихання та рівня цукру в крові);

- аналіз політичної ситуації (технології Big Data в політичних відносинах використовують у виборчих кампаніях); [3]

- формування інформаційної бази для процесів прийняття управлінських рішень в промисловості (згідно звіту компанії McKinsey «Global Institute, Big data: The next frontier for innovation, competition, and productivity», дані стали таким само важливим фактором виробництва, як трудові ресурси чи виробничі активи). За рахунок використання великих даних, компанії можуть отримувати відчутні конкурентні переваги.

Технології Big Data можуть бути корисними при вирішенні наступних задач: прогнозування ринкової ситуації; маркетинг і оптимізація продажів; вдосконалення продукції; ухвалення управлінських рішень; підвищення продуктивності праці; ефективна логістика; моніторинг стану основних фондів.[4]

Список літератури

1. Офіційний сайт caseware. URL: <https://caseware.com.ua/використання-big-data-у-сучасному-світі/>.
2. Офіційний сайт вікіпедія. URL: https://uk.wikipedia.org/wiki/Великі_дані
3. Сайт новин lb. URL: https://lb.ua/blog/katerina_odarchenko/395903_big_data_yak_tse_pratsyuie_ukraini.html.
4. Офіційний сайт компанії It-enterprise. URL: <https://www.it.ua/knowledge-base/technology-innovation/big-data-bolshie-dannye>.

УДК 004.43

*Курдупов О. Л., студент
Зелінська О.В., к.т.н., доцент, доцент
кафедри інформаційних технологій*

ІНТЕЛЕКТУАЛЬНА РОЗРОБКА ДАНИХ ДЛЯ МІГРАЦІЇ В БЕЗПЕЧНІ СЕРЕДОВИЩА НА ОСНОВІ NoSQL

Донецький національний університет імені Василя Стуса, м.Вінниця

Постановка проблеми

Бази даних NoSQL мають низку переваг перед RDBs (Реляційними базами даних), включаючи можливість обробляти різні типи даних, будь то структуровані, неструктуровані або напівструктуровані, дуже гнучкі, розподільні та масштабовані. Вони можуть значно знизити вартість товарного обладнання, оскільки воно може легко масштабуватися (горизонтальна масштабованість) на товарних кластерах. Бази даних NoSQL є економічно ефективними, оскільки вони дозволяють зберігати величезну кількість даних за набагато нижчих витрат. Бази даних NoSQL мають більш прості моделі даних, розподіл даних та автоматичний ремонт, що призводить до менших вимог до технічного обслуговування та налаштування, тому вони вимагають меншого управління на відміну від RDB. Враховуючи проблеми великих даних для RDB, сучасні організації, які мають системи великих даних, швидко переходять від існуючих RDB до баз даних NoSQL.

Мета дослідження

Аналіз даних - це ще одна проблема для організацій, що працюють з великими даними. Наше дослідження спрямоване на те, щоб уявити автоматизований підхід до перетворення існуючої бази даних RDB в базу даних NoSQL і застосувати метод очищення даних для підвищення якості даних. Очищення даних необхідне для досягнення точності аналізу даних.

Результати дослідження

У запропонованому підході використовується метод перетворення моделі бази даних SQL Server до бази даних Oracle NoSQL. Ключовими особливостями баз даних NoSQL є реплікація, розподілені індекси та використання оперативної пам'яті для зберігання даних, горизонтальної масштабованості, секціонування та гнучкої схеми.

Системи управління (СУБД) не здатні керувати величезним зростанням обсягу даних на одному сервері (вертикальна масштабованість). Для цього перетворення необхідно, щоб дизайнери-архітектори мали глибоке розуміння RDB, а також бази даних NoSQL.

Один з методів призначений для перетворення даних із бази даних RDB (MySQL) до NoSQL (MongoDB). Міграція з реляційного на NoSQL спочатку складається з кількох кроків; створюється з'єднання з базою даних

MySQL, а потім здійснюється доступ до деталей бази даних через прототип програмного забезпечення. Надалі зіставлення виконується між RDB (MySQL) та NoSQL (MongoDB). Використовується підхід, який має два модуля: модуль перетворення та модуль очищення даних. Перший з них полягає в перетворенні RDB на базу даних Oracle NoSQL за допомогою перетворення моделі БД. Останній забезпечує методику очищення даних для покращення якості даних.

Трансформація моделі - це автоматизований спосіб створення та використання концептуальних моделей предметної галузі. Для цієї мети доступна відносна метамодель, але відсутня метамодель для Oracle NoSQL. Необхідними елементами для перетворення моделі є вхідна модель, опис перетворення, механізм перетворення та правила перетворення. В описуваному

підході використовується так званий механізм SiTra як механізм перетворення вхідної моделі в вихідну модель та зіставляються поняття вхідної моделі (SQL Server, RDBs model) з вихідною моделлю (Oracle NoSQL model) для створення правил перетворення, що використовуються SiTra для перетворення [1]. MDA (Model Driven Architecture) - це заснований на моделі підхід, який має різні типи перетворення моделі, наприклад модель в модель, модель в текст і текст у моделі. Перетворення "модель - модель" - головне завдання, яка перетворює одну модель на іншу. Підхід MDA зіставляє поняття вихідної метамоделі із відповідними поняттями цільової метамоделі. Це зіставлення використовується для створення правил перетворення, які використовуються механізмом даного перетворення для автоматичного створення цільової метамоделі із вихідної метамоделі.

База даних Oracle NoSQL містить таблиці, еквівалентні таблицям бази даних SQL Server. Кожна таблиця Oracle NoSQL зберігає дані у вигляді рядків та полів Oracle NoSQL, які еквівалентні рядкам та стовпцям бази даних SQL Server. Рядок-це набір полів, і кожен рядок – це пара ключа та значення, де ключ унікальний і має значення довільної довжини. Кожна таблиця Oracle NoSQL складається з безлічі рядків, еквівалентних рядкам у базі даних SQL Server [2]. Кожен рядок Oracle NoSQL виймається за допомогою унікального ключа цього конкретного рядка.

Методи, що використовуються для підтримки якості даних – це очищення даних, моніторинг якості даних та інтеграція даних. Очищення даних виявляє та видаляє невідповідності та помилки даних. Очищення даних стає необхідною при перетворенні застарілої бази даних на нову систему баз даних, наприклад, бази даних SQL-server до бази даних NoSQL [2]. Невірні, неповні та неточні дані виявляються за допомогою різних методів. Ми використовували синтаксичний аналізатор для виявлення синтаксичних помилок, метод кластеризації для ідентифікації нерелевантних записів та метод контрольної суми файлів для виявлення дублювання даних.

Синтаксичний аналізатор дозволяє виявити синтаксичні помилки. Робота цього парсера аналогічна роботі парсера, що використовується для синтаксичного аналізу мови та граматики. Синтаксичний аналізатор буде приймати рішення про прийняття чи відхилення рядка даних. Він приймає дані, якщо вони відповідають специфікації (наприклад, JSON). Алгоритм кластеризації забезпечує підтримку для ідентифікації нерелевантних даних. Використовується комбінований алгоритм кластеризації, що враховує Евклідову відстань між записами. Алгоритм намагається скоригувати розмір кластерів, роблячи автоматичне регулювання розмірів кластера швидко[3]. Виявлені кластери поєднують записи, що мають певну подібність. Запис, що має аномальну відстань, буде ідентифікований як нерелевантні дані.

Дублювання даних означає, що одна сутність має більше одного уявлення. Кращий спосіб виявити повторювані записи-це провести порівняння між записами; порівняти кожен запис з усіма іншими записами. Ми використовували метод контрольної суми для виявлення дублювання даних. Цей метод дозволяє точно та швидко ідентифікувати дублювання шляхом порівняння контрольної суми записів.

Висновки

У цій роботі було розглянуто підхід перетворення RDB на базу даних NoSQL, що складається з двох модулів: модуля перетворення та модуля очищення даних.

Запропонований підхід придатний для перетворення баз даних та очищення даних. Як майбутній напрямок цей підхід може бути розширений, щоб забезпечити можливість багаторазового введення даних у разі використання великого обсягу даних. Час трансформації може бути додатково скорочено за допомогою найсучасніших технологій. Модуль перетворення даних може бути вдосконалений в якості майбутньої роботи для підтримки інших RDB та NoSQL. Модуль очищення даних може бути додатково вдосконалено з використанням статистичних методи для ідентифікації дублікатів.

Список літератури

1. S. Ramzan, I. Bajwa, and R. Kazmi, "An intelligent approach for handling complexity by migrating from conventional databases to big data," *Symmetry*, vol. 10, no. 12, p. 698, Nov. 2018.
2. MongoDB Production Deployments [Електроний ресурс].
[URL: http://www.mongodb.org/about](http://www.mongodb.org/about)
3. A. Davoudian, L. Chen, and M. Liu, "A survey on NoSQL stores," *ACM Comput. Surv.*, vol. 51, no. 2, p. 40, Apr. 2018.

УДК 004.855:519.216

Дужак А. О., студент
Зелінська О.В., к.т.н., доцент, доцент
кафедри інформаційних технологій

ПРИКЛАДНІ АСПЕКТИ ВИКОРИСТАННЯ ВЕЛИКИХ ДАНИХ ДЛЯ РОЗВ'ЯЗАННЯ ОКРЕМИХ ПРОБЛЕМ СТАТИСТИКИ

Донецький національний університет імені Василя Стуса, м. Вінниця

«Великі дані (Big Data) – позначення структурованих и неструктурованих даних величезних обсягів і значного розмаїття, що піддаються ефективній обробці програмних інструментів, які горизонтально масштабуються та з'явилися у кінці 2000-х років, і альтернативних традиційних систем управління базами даних і рішенням класу рішень Business Intelligence».

Основні принципи роботи з великими даними такі:

- **Горизонтальна масштабованість.** Це — базовий принцип обробки великих даних.
- **Відмовостійкість.** Цей принцип витікає з попереднього.