

ІНФОРМАЦІЙНІ Й КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ РАЦІОНАЛЬНОЇ ЕКСПЛУАТАЦІЇ І УПРАВЛІННЯ МІСЬКИМ ГОСПОДАРСТВОМ

АНАЛІЗ МАТЕМАТИЧНИХ МЕТОДІВ ПОДАННЯ BIG DATA

Беленькова К.О.

Науковий керівник – Буласько М.В., канд. техн. наук, доцент

Big Data (Великі дані) за визначенням К. Лінч, Д. Ленеї – набір методів та засобів опрацювання структурованих і неструктурованих різнотипних динамічних даних великих обсягів з метою їх аналізу та використання для підтримки прийняття рішень. Big Data є альтернативою традиційним системам управління базами даних і рішенням класу Business Intelligence. До цього класу відносять також засоби паралельного опрацювання даних (NoSQL, алгоритми MapReduce, Hadoop). На думку компанії DCA (Data-Centric Alliance) під Big Data розуміють не якийсь конкретний об'єм даних і навіть не дані, а методи їх обробки, які дозволяють розподілено обробляти інформацію. Ці методи можна застосовувати як до великих масивів даних (таких як дані всіх сторінок в мережі Інтернет), так і до малих масивів (інформація про денні поступлення товару в магазин).

Метою роботи є формальний опис різних моделей подання даних, таких як: багатовимірної модель, об'єктна модель та графова модель. Виділення операцій та носіїв, а також способів їх сумісного використання.

Основними поняттями багатовимірної моделі даних є:

- гіперкуб даних (G),
- вимір (V),
- атрибут (A),
- комірка (X),
- значення (G (V, A)).

Гіперкуб даних містить один або більше вимірів і є впорядкованим набором комірок. Кожна комірка визначається одним набором значень вимірів – атрибутів. Комірка може містити дані або бути порожньою.

Під виміром розуміють множину атрибутів, що утворюють одну з граней гіперкуба. Прикладом часового виміру є список днів, місяців, кварталів. Прикладом географічного виміру може бути перелік територіальних об'єктів: населених пунктів, районів, регіонів, країн та ін.

Отже, гіперкуб даних можна позначити як множину комірок, що відповідає множинам V, A: G(V, A). Гіперкуби підтримують ієрархію вимірів і формул без дублювання їх визначень. Набір відповідних гі-

перкубів складає багатовимірну базу даних. Багатовимірне представлення даних добре використовувати для задач візуалізації даних та їх аналізу, але у зв'язку з розрідженістю гіперкуба обсяг даних у такому випадку є більший порівняно з реляційним представленням.

ВИБІР ТИПІВ МОДЕЛЕЙ ДАНИХ ДЛЯ ПРЕДСТАВЛЕННЯ BIG DATA

Беседіна А.О.

Науковий керівник – Булаєнко М.В., канд. техн. наук, доцент

Процес побудови інформаційної моделі регіону ускладнюється різноманітністю моделей даних, а також наявністю різних рівнів агрегації даних. Однією з популярних технологій для розроблення систем територіального управління є Big Data. Методи машинного навчання та візуалізації даних дають змогу опрацювати та графічно подати результати аналізу даних великих обсягів (мільйони користувачів).

Модель даних – це сукупність засобів опису структур даних для додатка або класу додатків. Модель даних містить у собі типи даних, структури даних, систему операцій, засоби опису обмежень. Технології Big Data пов'язані з необхідністю обробки інформації з різних типів джерел інформації: структурованих, слабоструктурованих, неструктурованих.

При використанні моделі структурованих даних, на дані накладаються заздалегідь відомі обмеження за типом й довжиною кожного атрибута; структура даних відома й визначена за допомогою схеми даних, її автоматична зміна в процесі роботи моделі ускладнена. Інтерпретувати дані без знання схеми не представляється можливим. Прикладом реалізації моделі структурованих даних може виступати будь-яка реляційна система керування базою даних (СУБД).

Розробка моделі для неструктурованих даних є вкрай складним завданням з наступних причин: дані, як правило, представлені природньою мовою, що ускладнює роботу з ними; повна відсутність визначеної структури накладає серйозні обмеження на можливі операції з даними. Автоматичне виділення структури в таких даних, як правило, не може бути виконане однозначним чином.

Слабоструктурованими даними є будь-які проміжні дані між структурованими й неструктурованими. Структура в таких даних може бути неповною, недовизначеною, а також допускати виключення. Таким чином, у цій моделі повинен існувати інструмент обробки виключень, що дозволяє формувати спосіб запиту до цих даних, ґрунтуючись на заздалегідь заданих критеріях.