

УДК 654.01

FEATURES OF DATA ANALYSIS TOOLS IN TRANSPORT SYSTEMS
ОСОБЛИВОСТІ ЗАСОБІВ АНАЛІЗУ ДАНИХ В ТРАНСПОРТНИХ
СИСТЕМАХ

Lyamzin A.A. / Лямзін А.О.

s.t.s., docent, / к.т.н., доц.

ORCID: 0000-0002-6964-845X

Приазовський державний технічний університет,

Маріуполь, вул. Університетська 7, 87500

Pryazovskyi State Technical University,

Mariupol, Universytetska st., 7, 87500

Polyshin D/L. / Полушин Д.Л.

student / студент

Приазовський державний технічний університет,

Маріуполь, вул. Університетська 7, 87500

Pryazovskyi State Technical University,

Mariupol, Universytetska st., 7, 87500

Анотація. В роботі розглядаються методи комп'ютеризації та інформаційні технології дозволяють удосконалити і полегшити виробничий процес, що дозволяє частково або повністю полегшити працю, пов'язану з виконанням небезпечних і складних для життя трудових дій. Сьогодні інформаційні технології дозволяють вирішувати багато проблем і надають можливість зробити світ зручніше і комфортніше, сучасніше, краще.

Ключові слова: інформаційні технології, життєвий цикл, трансформація інформаційних потоків.

Abstract. In the robot, I look at the methods of computerization and information technology that allow you to refine and take care of the viral process, who often allow it, and if you need to take care of the public, you need to take care of it. Thanks to the information technology, you can let us have a wide range of problems and give you the opportunity to feel better and more comfortable, comfortable and beautiful.

Key words: information technology, life cycle, transformation of information flows.

Вступ.

Досягнення сучасних інформаційно-комунікаційних та дорожньо-транспортних технологій дозволяють забезпечити якісно новий рівень організації перевезень і управління транспортними потоками та транспортно-дорожньою інфраструктурою. На жаль, в даний час результати їх використання незначні в порівнянні з можливостями, через відсутність міждисциплінарного

системного підходу до вирішення поставлених завдань. Основна увага звернена на створення безпілотних транспортних засобів, що виглядає дуже привабливим. І це правда. Але для їх успішного впровадження потрібно створювати інфраструктурні рішення, в яких транспортні засоби, дороги і, головне, користувач, стають єдиною інформаційною системою, що працює на сучасних цифрових інтелектуальних рішеннях.

Одним із засобів аналізу даних є технологія OLAP (Online Analytical Processing - оперативна аналітична обробка даних). Це клас додатків і технологій, призначених для збору, зберігання і аналізу багатовимірних даних з метою підтримки прийняття управлінських рішень. Технологія OLAP дозволяє аналітикам, менеджерам і керуючим сформуванню своє власне бачення даних, використовуючи швидкий, однаковий, оперативний доступ до різноманітних форм подання інформації. Джерело: [1,2].

Основний текст.

Життєвий цикл аналітики даних може включати в себе: виявлення проблем аналізу даних, збір набору даних, проектування, аналіз даних, візуалізацію даних.

Для початку необхідно зрозуміти, якою є реальна кінцева мета проекту, яку користь він може принести, оцінити його критерії успішності або провалу, вибрати технології для збору, трансформації та аналізу даних. Потрібно постаратися відповісти на такі додаткові питання: чи достатньо у вас ресурсів на реалізацію проекту? з ким ви будете контактувати по ходу проекту? чи достатньо у вас вхідних даних? чи були вже в компанії спроби аналізу? чи достатньо у вас часу на реалізацію проекту? з якими проблемами ви можете зіткнутися?

До етапу виявлення проблеми (вивчення) входять також такі пункти: визначення набору необхідних знань, які потрібні для орієнтації в предметній області; наявність доступних ресурсів (люди, інструменти, дані); структурування даних з точки зору аналітики; вивчення історії бізнесу.

Після вивчення слід перейти до підготовки даних. Визначте використовувані програмні засоби, тобто виберіть ПО, БД, інструменти аналізу та візуалізації. Отримайте, очистіть і завантажте дані в систему, після чого оцініть кількість і якість даних.

На етапі проектування моделі потрібно визначити методи, технології, робочі процеси, необхідні для розрахунку моделі і обсяг даних. Також важливо визначити кореляцію між змінними: стовпцями таблиць і полями даних.

Побудова моделі полягає в тому, щоб розробити набори даних для тестування, навчання і виробництва. Потім оцінити життєздатність і надійність даних для використання в моделі.

І в кінцевому підсумку вибрати робоче оточення, тобто апаратні і програмні засоби, на яких можна налаштувати процес.

В результаті попереднього кроку вийшли якісь результати. Тепер необхідно визначити, чи вдалося вам досягти результату на основі критеріїв проекту, а також визначити ключові результати дослідження. Залежно від цільової аудиторії можна розробити діаграми і графіки і сформулювати підсумки і рекомендації.

Останній етап - практична реалізація. В процесі цього етапу відбувається доставка фінальних рекомендацій, звітів, коду і технічних документів, запуск пілотного проекту, реалізація моделі у виробничому середовищі і інтеграція аналітичних оцінок в панель управління або в операційній системі.

Починаючи працювати з великими даними, багато хто стикається з труднощами, які можна уникнути, виконуючи найпростіші поради, такі як: завжди проводити глибоке вивчення предметної області, розбивати завдання на більш дрібні, робити аналітику гнучкою і масштабованою, передбачити можливість повторення кожного з етапів з можливістю внесення змін в попередній етап, бути готовим до негативного результату, уважно оцінювати витрачений час на кожен етап.

Одночасний аналіз по декількох вимірах визначається як багатовимірний аналіз. Кожен вимір включає напрямки консолідації даних, що складаються з

послідовних рівнів узагальнення, де кожен рівень відповідає більшою мірою агрегації даних по відповідному вимірюванню.

OLAP є інструментом для аналізу великих обсягів даних в режимі реального часу і забезпечує наступні можливості роботи з багатовимірними даними: гнучкий перегляд інформації, довільні зрізи даних, деталізація, згортка або консолідація, обертання, порівняння в часі (рис. 1).

Візьмемо для прикладу таблицю Invoices, яка містить замовлення фірми. Поля в даній таблиці будуть наступні: дата замовлення, країна, місто, назва замовника, компанія-постачальник, назва товару, кількість товару, сума замовлення.

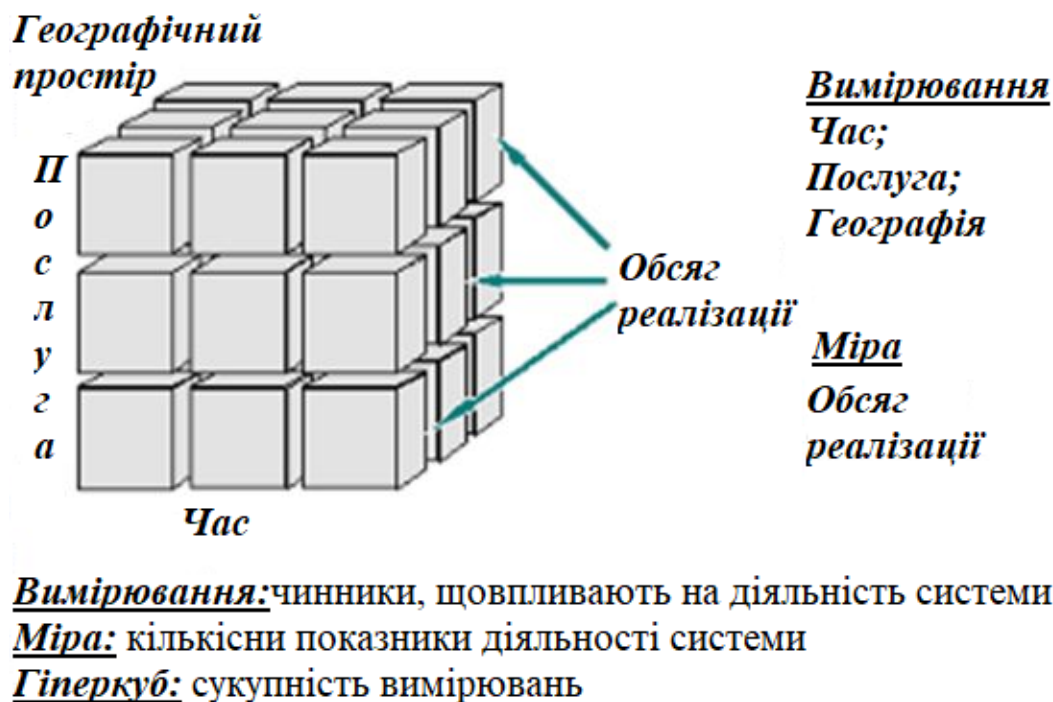


Рисунок 4.1 – Схема функціоналів складових OLAP

Які агрегатні дані ми можемо отримати на основі цього подання? Зазвичай це відповіді на наступні питання. Яка сумарна вартість замовлень, зроблених клієнтами з певної країни? Яка сумарна вартість замовлень, зроблених клієнтами з певної країни і доставлених певною компанією? Яка сумарна вартість замовлень, зроблених клієнтами з певної країни в заданому році і доставлених певною компанією?

Всі ці дані можна отримати з цієї таблиці цілком очевидними SQL-запитами з угрупованням. Результатом цього запиту завжди буде стовпець чисел і список атрибутів, його описують (наприклад, країна) - це одновимірний набір даних або, кажучи математичною мовою, вектор.

Уявімо собі, що нам треба отримати інформацію за сумарною вартістю замовлень з усіх країн і їх розподіл по компаніям доставщиків - ми отримаємо вже таблицю (матрицю) з чисел, де в заголовках колонок будуть перераховані постачальникам, в заголовках рядків - країни, а в осередках буде сума замовлень. Це - двовимірний масив даних. Такий набір даних називається зведеною таблицею (pivot table) або крос-таблицею.

Якщо ж нам захочеться отримати ті ж дані, але ще в розрізі років, тоді з'явиться ще одна зміна, тобто набір даних стане тривимірним (умовним тензором 3-го порядку або 3-х мірним «кубом»).

Очевидно, що максимальна кількість вимірювань - це кількість всіх атрибутів (Дата, Країна, Замовник і т.д.), що описують наші дані (суму замовлень, кількість товарів і т.п.)

Також варто додати, що для зручності роботи з OLAP існують мови запитів до багатовимірним кубів, наприклад, MDX.

В осередках багатовимірного куба поміщаються числові параметри, призначені для аналізу, наприклад, обсягів продажів. Вимірами OLAP-куба можуть служити такі параметри, як час, продукти, регіони, продавці. Продажі за часом в консолідованому вигляді можуть представлятися по роках, при деталізації - по кварталах, місяцях і днях.

Заключення та висновки.

Було розглянуто існуючі шкали та методи оцінки ефективності транспортних систем.

Було доведено необхідність покрокового підходу до визначення ступеня ефективності транспортних систем в умовах кризи.

Література

1. Майер-Шенбергер В., Кукъер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. Language Arts & Disciplines – 2013. – 599с.

1. Силен Д., Мейсман А., Али М. Основы DataScience и BigData. Python и наука о данных. СПб: Питер. – 2017. – 336с.

© Лямзін А.О.

© Полушин Д.Л.