

Інструкція до самостійної роботи №1

Тема: Інформаційні технології обробки збережених даних.

Мета: Вивчити технології обробки збережених даних.

Основні питання теми:

1. Технологія оперативної обробки даних (OLTP).
2. Технологія оперативного аналізу даних (OLAP).
3. Технологія інтелектуального аналізу даних (DMg).

Рекомендована література:

1. Інформаційні системи і технології в обліку й аудиті: Навчальний посібник. [Електронний ресурс]. – Режим доступу: https://www.ltklntu.org.ua/wp-content/uploads/2022/09/ISTOA_Adamyk.pdf

Завдання до самостійної роботи №1

1. Записати:
 - технологію оперативної обробки даних (OLTP): визначення, ознаки, сутність;
 - технологію оперативного аналізу даних (OLAP): сутність, базові операції, класифікаційні класи;
 - технологію інтелектуального аналізу даних (DMg): визначення, опис даних, стадії, задачі, переваги.

Питання для самоконтролю:

1. Охарактеризувати технологію оперативної обробки даних (OLTP): визначення, ознаки, сутність.
2. Охарактеризувати технологію оперативного аналізу даних (OLAP): сутність, базові операції, класифікаційні класи.
3. Охарактеризувати технологію інтелектуального аналізу даних (DMg): визначення, опис даних, стадії, задачі, переваги.

Вид контролю:

поточний – семінарське заняття №2.

Опорний конспект

OLTP (англ. Online Transaction Processing) – онлайнова обробка транзакцій. Спосіб організації БД, при якому система працює з невеликими за розмірами транзакціями, що йдуть великим потоком, і при цьому клієнту потрібний від системи максимально швидкий час відповіді.

OLAP (англ. online analytical processing, аналітична обробка у реальному часі) – це інтерактивна система що дозволяє переглядати різні підсумки по багатовимірних даних. Термін «в реальному часі» (англ. online) означає що нові результати отримуються протягом секунд, без довгого очікування на результат запиту.

OLAP-куб містить базові дані та інформацію про вимірювання (агрегати). Куб потенційно містить всю інформацію, потрібну для відповідей на будь-які запити.

Найпоширенішими є такі назви класифікаційних угруповань: клас, підклас, група, підгрупа, вид, підвид, тип.

Технологія інтелектуального аналізу даних (DMg)

Файли DMG в цілому є файл-образ диска, який при відкритті з'являється на робочому столі. Файл містить дані RAW, які зазвичай стиснуті та зашифровані. Пристрої з ОС Mac розглядають файли DMG як диск, вставлений у дисковод.

Файл DMG (з англ. Mac OS X Disk Image) є образом диска, створеним в операційній системі Mac OS. Такий образ диска, з метою читання даних, слід монтувати на віртуальний пристрій.

Для того, щоб відкрити DMG у системі Windows, потрібні утиліти емуляції дисків із серії Daemon Tools. Найчастіше DMG файли використовуються операційною системою Mac, в якій програми-інстальатори, завантажені з мережі, застосовують подібне розширення файлу. Наприклад, у такій операційній системі, як Mac OS Classic, використовувався .IMG формат файлу замість файлу з розширенням .DMG. А в нових версіях типу OSX, був реалізований саме формат DMG. Варто відзначити, що формат файлу DMG не сумісний з OS 9 і більш застарілими модифікаціями платформ в системі Mac.

Власникам комп'ютерів компанії з Купертіно відкрити DMG файл дозволяє штатна програма Apple Disk Utility. Основним методом вирішення проблеми відкриття формату DMG в середовищі операційної системи Windows є його конвертація. DMG-образ даних потрібно конвертувати у файл у форматі, який підтримується тій чи іншій ОС. Відкрити файл DMG можна за допомогою Acute Systems TransMac, існують і більш поширені програми, серед яких можна виділити програму UltraISO, а також конвертери DMG2IMG або DMG2ISO, що дозволяють надалі відкрити файл, який є раніше недоступним.

Задачі DMg класифікують перш за все за типами інформації, що витягується, тобто за видами знайдених закономірностей. Виділяють наступні п'ять видів задач:

1. Класифікація – дозволяє виявити ознаки, що характеризують однотипні групи об'єктів (класи), для того, щоб за відомими значеннями цих характеристик можна було віднести новий об'єкт до того або іншого класу.

Передбачається, що характеристики класів відомі наперед (до аналізу). Як методи рішення задачі класифікації застосовують алгоритм найближчого сусіда (Nearest Neighbor), індукцію дерев рішень, «навчені вчителем» нейронні мережі.

2. Кластеризація – розповсюдження ідеї класифікації на складніший випадок, коли самі класи не призначені. В результаті виконання процедури кластеризації початкові дані розбиваються на однорідні групи (кластери). Це дозволяє установі виробити по відношенню до кожної з груп (наприклад, до груп покупців) певну політику. Задача кластеризації значно складніша за задачу класифікації. У даний час найбільш часто задачі кластеризації вирішуються методом «навчання без вчителя» спеціального виду нейронних мереж.

3. Виявлення асоціацій. Асоціація – закономірність в даних, що фіксує настання двох (або більш) одночасних подій. Типовий приклад асоціації досліджується в задачі визначення пар виписаних одночасно документів (наприклад, накладна і податкова накладна, платіжне доручення про виплату заробітної плати на дебетові картки працівників і платіжне доручення про сплату податків і зборів, пов'язаних з оплатою праці, тощо).

4. Виявлення послідовностей. Послідовність – закономірність у даних, що фіксує настання подій не одночасно, а з деяким певним розривом в часі. Таким чином, асоціація є окремий випадок послідовності з нульовим тимчасовим лагом. Наприклад, виписування відрядного посвідчення працівникові й авансовий звіт про витрачені кошти; податкова декларація і платіжне доручення про сплату податків.

5. Прогнозування – формалізована процедура прогнозу, яка на основі дослідження поточних і минулих даних дозволяє оцінити майбутні значення числових показників.