

LECTURE NOTES

COMP8014

Knowledge Data Discovery

**Data Warehousing and Online
Analytical Processing**

LEARNING OUTCOMES

L0 1: Mahasiswa diharapkan mampu menjelaskan konsep dasar Knowledge Data Discovery

LO3: Mahasiswa diharapkan mampu mengidentifikasi infrastruktur sistem data mining

OUTLINE MATERI :

1. Apa itu gudang data?
2. Model data multi dimensi
3. Arsitektur Data Warehouse
4. Dari Data Warehouse hingga Data Mining

ISI MATERI

Apakah Data Warehouse?

- Database pendukung keputusan yang dipelihara secara terpisah dari database operasional organisasi
- Dukung pengolahan informasi dengan menyediakan platform data historis dan konsolidasi yang terkonsolidasi.
- "Data warehouse adalah koleksi data berorientasi objek, terpadu, waktu-varian, dan nonvolatile untuk mendukung proses pengambilan keputusan manajemen." - W. H. Inmon
- Proses membangun dan menggunakan gudang data

Apa perbedaan Heterogeneous DBMS dan Data Warehouse?

Integrasi DB heterogen tradisional: Pendekatan berbasis query Data Warehouse: update-driven, high performance

- | | |
|--|--|
| <ul style="list-style-type: none">• Build Wrappers / mediator di atas database heterogen• Meta-dictionary digunakan untuk menerjemahkan kueri ke kueri dan hasilnya diintegrasikan ke dalam kumpulan jawaban global• Penyaringan informasi yang kompleks, bersaing untuk mendapatkan sumber daya | <ul style="list-style-type: none">• Informasi dari sumber heterogen terintegrasi terlebih dahulu dan disimpan di gudang untuk permintaan dan analisis langsung |
|--|--|

Apa perbedaan Heterogeneous DBMS dan Data Warehouse?

- OLTP (on-line transaction processing)
 - Major task of traditional relational DBMS
 - Day-to-day operations: purchasing, inventory, banking, manufacturing, payroll, registration, accounting, etc.
- OLAP (on-line analytical processing)
 - Major task of data warehouse system
 - Data analysis and decision making
- Distinct features (OLTP vs. OLAP):
 - User and system orientation: customer vs. market
 - Data contents: current, detailed vs. historical, consolidated
 - Database design: ER + application vs. star + subject
 - View: current, local vs. evolutionary, integrated
 - Access patterns: update vs. read-only but complex queries

Apakah saja operasi pada OLAP?

- Roll up (drill-up): summarize data
 - *by climbing up hierarchy or by dimension reduction*
- Drill down (roll down): reverse of roll-up
 - *from higher level summary to lower level summary or detailed data, or introducing new dimensions*
- Slice and dice: *project and select*
- Pivot (rotate):
 - *reorient the cube, visualization, 3D to series of 2D planes*
- Other operations
 - *drill across: involving (across) more than one fact table*
 - *drill through: through the bottom level of the cube to its back-end relational tables (using SQL)*

SIMPULAN

1. Gudang data adalah koleksi data berorientasi objek, terpadu, waktu-varian, dan nonvolatile untuk mendukung proses pengambilan keputusan manajemen.
2. Data warehouse didasarkan pada model data multidimensional yang melihat data berupa data kubus
3. Empat tampilan mengenai desain gudang data: Tampilan top-down, tampilan sumber data, tampilan gudang data dan tampilan query Bisnis.
4. Arsitektur OLAP: OLAP relasional (ROLAP), OLAP multidimensional (MOLAP), OLAP Hybrid (HOLAP) (mis., Microsoft SQLServer) dan server SQL khusus (misalnya, Redbricks).

DAFTAR PUSTAKA

1. Han, J., Kamber, M., & Pei, Y. (2006). "Data Mining: Concepts and Technique". Edisi 3. Morgan Kaufman. San Francisco
2. Tan, P.N., Steinbach, M., & Kumar, V. (2006). "Introduction to Data Mining". Addison-Wesley. Michigan
3. Witten, I. H., & Frank, E. (2005). "Data Mining : Practical Machine Learning Tools and Techniques". Second edition. Morgan Kaufmann. San Francisco