

Data Cube Technology

Anggota Kelompok

1. Ibrahim Al Anshor
2. Rabal
3. Zidan

Daftar Pembahasan

1. Latar Belakang
2. Definisi Data Cube
3. Manfaat Data Cube
4. Arsitektur Data Cube : Multidimensional Data Model
5. OLAP (Online Analytical Processing)
6. Skema Untuk Multidimensional Data Model
7. Kelebihan Data Cube
8. Kekurangan Data Cube
9. Implementasi Data Cube Di Dunia Nyata
10. Kesimpulan
11. Referensi

1. Latar Belakang

Data Cube penting dalam dunia teknologi dan bisnis karena memungkinkan analisis data multidimensional yang mendalam, pengambilan keputusan cepat, optimasi operasional, personalisasi produk, penemuan wawasan baru, penanganan Big Data, dan menjaga keamanan data. Data Cube membantu organisasi menjadi lebih efisien, inovatif, dan kompetitif.

2. Definisi Data Cube

1. Apa itu Data Cube

Data Cube adalah representasi multidimensional dari data yang memungkinkan penyimpanan, analisis, dan pemahaman yang lebih baik tentang data. Dalam Data Cube, data diorganisir dalam bentuk tabel tiga dimensi atau lebih, di mana masing-masing dimensi mewakili atribut atau variabel tertentu. Ini memungkinkan pemodelan dan analisis data yang lebih kompleks daripada yang dapat dicapai dengan pendekatan tabel datar.

Data Cube adalah alat yang sering digunakan dalam sistem OLAP untuk analisis data interaktif yang efisien. Ini memungkinkan pengguna untuk melakukan operasi seperti slicing, dicing, dan pivoting untuk mendapatkan wawasan dari data yang kompleks.

2. Bagaimana Cara Kerja Data Cube

Cara kerja Data Cube melibatkan proses pengorganisasian data dalam bentuk multidimensional untuk memungkinkan analisis yang lebih baik. Berikut adalah langkah-langkah umum cara kerja Data Cube:

1. Pengumpulan Data: Data pertama dikumpulkan dari berbagai sumber, seperti database, file Excel, atau sumber data lainnya.
2. Pemilihan Dimensi: Anda harus memilih atribut atau dimensi apa yang akan digunakan untuk menganalisis data. Misalnya, dalam analisis data penjualan, dimensi mungkin mencakup waktu (bulan/tahun), produk (jenis produk), dan wilayah geografis.
3. Mengisi Data Cube: Data kemudian dimasukkan ke dalam Data Cube, dengan setiap nilai data sesuai dengan koordinat di dimensi yang sesuai.

Misalnya, angka penjualan produk tertentu pada bulan tertentu di wilayah tertentu akan menjadi nilai dalam Data Cube.

4. Pengolahan dan Agregasi: Data dalam Data Cube kemudian dapat diolah dan disagregasi, seperti menjumlahkan penjualan produk dalam semua wilayah untuk satu bulan atau mencari rata-rata penjualan produk selama beberapa tahun.
5. Analisis Data: Pengguna dapat melakukan analisis data dengan menggunakan operasi OLAP (Online Analytical Processing) seperti slicing (memilih satu layer dari Cube), dicing (mengiris Cube dalam dua dimensi), dan pivoting (mengubah orientasi Cube) untuk mendapatkan wawasan dari data multidimensional.
6. Visualisasi: Hasil analisis dapat divisualisasikan dalam bentuk grafik, grafik batang, atau tabel, yang membuatnya lebih mudah dipahami.
7. Pengambilan Keputusan: Hasil analisis membantu pengambilan keputusan yang lebih baik dalam berbagai bidang, seperti bisnis, manajemen, dan ilmu pengetahuan.
8. Penyimpanan dan Pemeliharaan: Data Cube perlu disimpan dan diperbarui secara berkala sesuai dengan perubahan data sumber.

Penting untuk dicatat bahwa Data Cube dapat diimplementasikan dalam berbagai platform dan alat perangkat lunak, dan pendekatan kerjanya dapat bervariasi tergantung pada alat yang digunakan. Namun, prinsip dasarnya adalah untuk mengorganisir data dalam tampilan multidimensional untuk analisis yang lebih baik.

3. Konsep dasar dari data multidimensional.

Konsep dasar dari data multidimensional adalah pengorganisasian data dalam bentuk yang mencakup beberapa dimensi atau atribut. Ini memungkinkan pemahaman data yang lebih komprehensif dan hubungan antara variabel yang

berbeda. Beberapa konsep dasar yang terkait dengan data multidimensional meliputi:

1. Dimensi: Dimensi adalah atribut atau kategori tertentu yang digunakan untuk mengklasifikasikan atau mengelompokkan data. Misalnya, dalam analisis penjualan, dimensi bisa mencakup waktu (bulan, tahun), produk (jenis produk), dan wilayah geografis.
2. Hierarki: Hierarki adalah hubungan hierarkis antara elemen dalam dimensi. Sebagai contoh, dimensi waktu dapat memiliki hierarki yang melibatkan tahun, kuartal, bulan, dan hari.
3. Fakta: Fakta adalah data aktual yang ingin dianalisis. Ini adalah nilai atau matrix yang diukur atau diamati dalam konteks data multidimensional. Misalnya, fakta dalam analisis penjualan bisa berupa jumlah penjualan atau pendapatan.
4. Data Cube: Data Cube adalah representasi multidimensional dari data, di mana data diorganisir dalam bentuk kubus tiga dimensi atau lebih. Setiap koordinat dalam kubus mewakili kombinasi dari elemen-elemen dalam dimensi yang berbeda.
5. Operasi OLAP: OLAP (Online Analytical Processing) adalah kumpulan operasi yang digunakan untuk menganalisis data multidimensional. Ini termasuk operasi slicing (memilih satu layer dari Cube), dicing (mengiris Cube dalam dua dimensi), dan pivoting (mengubah orientasi Cube).
6. Aggregasi: Aggregasi melibatkan penggabungan atau penjumlahan data dalam Cube untuk menciptakan ringkasan data yang lebih tinggi atau lebih rendah. Ini memungkinkan analisis dari tingkat yang berbeda, misalnya, dari detail hingga ringkasan.

7. Visualisasi: Data multidimensional sering divisualisasikan dalam bentuk grafik, grafik batang, tabel silang, atau laporan multidimensional untuk membantu pemahaman dan analisis.

Konsep dasar data multidimensional sangat penting dalam memahami cara data diorganisir dan dianalisis dalam konteks Data Cube dan OLAP. Ini membantu dalam mengungkapkan wawasan yang mendalam dari data dalam beberapa dimensi sekaligus.

3. Manfaat Data Cube

Penggunaan Data Cube memiliki manfaat yang signifikan dalam berbagai bidang, termasuk bisnis, analisis data, dan pengambilan keputusan. Berikut adalah beberapa manfaat utama penggunaan Data Cube:

1. Bisnis:

- a. Analisis Penjualan: Data Cube memungkinkan perusahaan untuk menganalisis penjualan berdasarkan berbagai dimensi seperti waktu, produk, dan wilayah, sehingga membantu dalam mengidentifikasi tren penjualan dan peluang pasar.
- b. Manajemen Stok: Data Cube dapat digunakan untuk mengelola stok dengan lebih efisien, mengidentifikasi permintaan produk yang tinggi dan rendah, serta meminimalkan biaya penyimpanan.
- c. Perencanaan Sumber Daya Manusia (SDM): Dalam manajemen SDM, Data Cube membantu dalam perencanaan tenaga kerja, pemantauan kinerja karyawan, dan pengambilan keputusan rekrutmen yang lebih baik.

2. Analisis Data:

- a. Eksplorasi Data: Data Cube memungkinkan analis data untuk mengeksplorasi data dengan lebih mendalam, menemukan hubungan yang tersembunyi, dan mengidentifikasi pola yang mungkin tidak terdeteksi dengan data datar.
- b. Penggalian Data (Data Mining): Dengan Data Cube, teknik penggalian data seperti clustering, klasifikasi, dan asosiasi dapat diterapkan untuk mengungkapkan wawasan dari data multidimensional.

3. Pengambilan Keputusan:

- a. Keputusan yang Cepat: Data Cube memungkinkan pengambilan keputusan yang cepat karena memungkinkan akses instan ke data

yang relevan dan kemampuan untuk melihat data dari berbagai sudut pandang.

- b. Perencanaan Strategis: Dalam bisnis, Data Cube membantu dalam perencanaan strategis jangka panjang dan pengembangan produk dengan wawasan yang lebih dalam.

Penggunaan Data Cube membantu organisasi untuk mendapatkan wawasan yang lebih dalam dari data mereka, meningkatkan efisiensi operasional, dan membuat keputusan yang lebih baik dalam berbagai bidang. Hal ini membantu dalam mencapai keunggulan kompetitif dan merespons perubahan pasar dengan lebih baik.

4. Arsitektur Data Cube : Multidimensional Data Model

Kita biasanya menganggap kubus sebagai struktur geometri 3 Dimensi, dalam pengolahan data, data cube adalah n-dimensi.

Untuk memahami data cube dan model data multidimensional dengan lebih baik, mari kita mulai dengan melihat data cube 2 Dimensi sederhana yang sebenarnya adalah tabel atau spreadsheet untuk data penjualan dari AllElectronics.

Secara khusus, kita akan melihat data penjualan AllElectronics untuk barang yang terjual per kuartal di kota Vancouver.

Data ini ditampilkan dalam Tabel 1. Dalam representasi 2-D ini, penjualan untuk Vancouver ditampilkan sehubungan dengan dimensi waktu (diorganisir dalam kuartal) dan dimensi item (diorganisir berdasarkan jenis item yang dijual).

Fakta atau ukuran yang ditampilkan adalah jumlah penjualan dalam dolar (dalam ribuan).

2-D view of Sales Data

location = "Vancouver"					
time (quarter)	item (type)				
	home	entertainment	computer	phone	security
Q1	605	825	14	400	
Q2	680	952	31	512	
Q3	812	1023	30	501	
Q3	927	1038	38	580	

Tabel 1. Data penjualan AllElectronics berdasarkan time dan item

Sekarang, anggaplah kita ingin melihat data penjualan dengan dimensi ketiga.

Misalnya, kita ingin melihat data berdasarkan waktu dan item, serta lokasi, untuk kota-kota Chicago, New York, Toronto, dan Vancouver.

Data 3-D ini ditampilkan dalam Tabel 2. Data 3-D dalam tabel ini direpresentasikan sebagai serangkaian tabel 2-D.

Secara konseptual, kita juga dapat merepresentasikan data yang sama dalam bentuk data cube 3-D, seperti yang ditunjukkan dalam Tabel 2.

3-D view of Sales Data

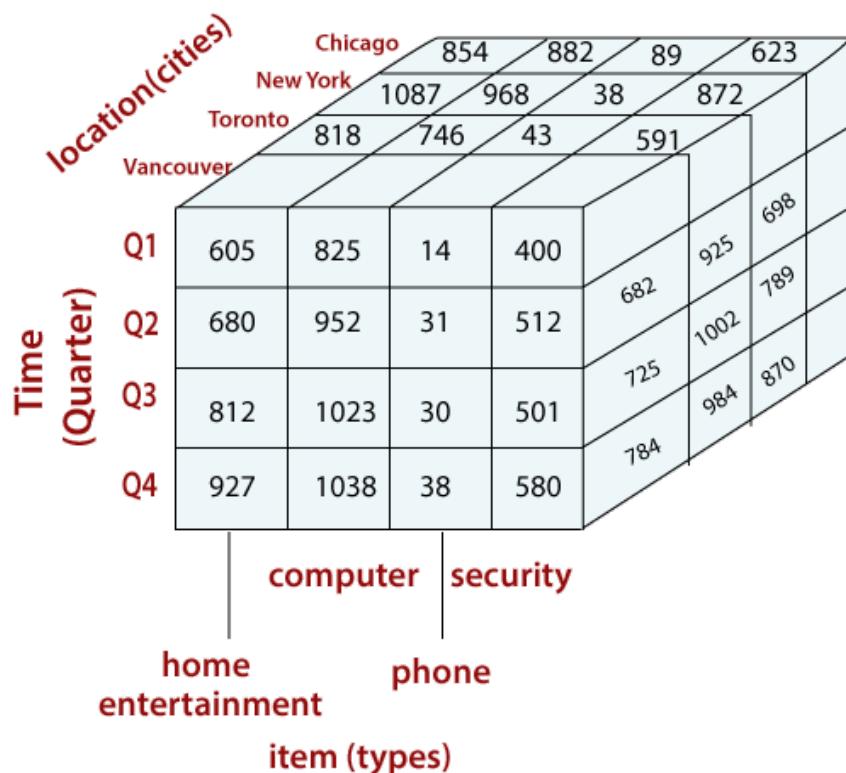
location ="Chicago"					location ="New York"					location ="Toronto"				
item					item					item				
home					home					home				
time	ent.	comp.	phone	sec.	time	ent.	comp.	phone	sec.	time	ent.	comp.	phone	sec.
Q1	854	882	89	623	1087	968	38	872		818	746	43	591	
Q2	943	890	64	698	1130	1024	41	925		894	769	52	682	
Q3	1032	924	59	789	1034	1048	45	1002		940	795	58	728	
Q4	1129	992	63	870	1142	1091	54	984		978	864	59	784	

Tabel 2. Data penjualan AllElectronics berdasarkan time ,item, dan lokasi

Dengan cara ini, kita dapat menampilkan data n-dimensi apa pun sebagai serangkaian "kubus" (n-1)-dimensi. Data cube adalah metafora untuk penyimpanan data multidimensional. Penyimpanan fisik aktual dari data tersebut mungkin berbeda dari representasi logisnya. Hal penting yang perlu diingat adalah bahwa data cube adalah n-dimensi dan tidak membatasi data pada 3-D.

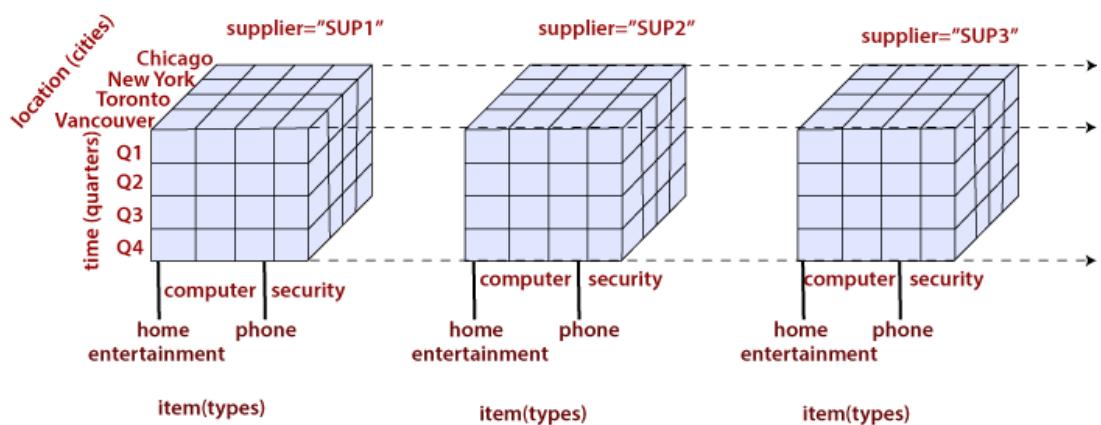
Representasi 3D Data Cube

3-D Data Cube



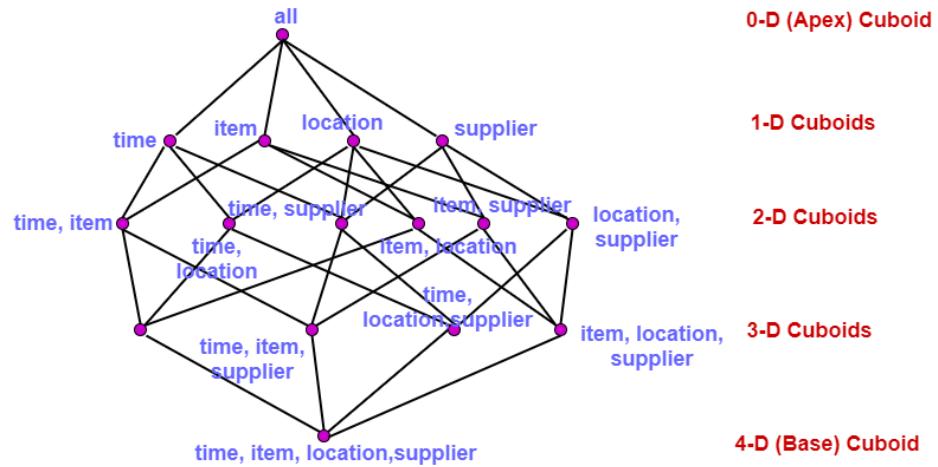
Gambar 1. Data penjualan AllElectronics berdasarkan time ,item, dan lokasi

Representasi 4D Data Cube, sering disebut sebagai cuboid.



Gambar 2. Data penjualan AllElectronics berdasarkan time ,item, lokasi, dan supplier

Lattice cuboid membentuk data cube. Gambar 3 menunjukkan lattice cuboid yang menciptakan data cube 4-D untuk dimensi waktu, item, lokasi, dan pemasok. Setiap cuboid mewakili tingkat ringkasan yang berbeda.



Gambar 4. Lattice Of Cuboid

5. Skema Untuk Multidimensional Data Model

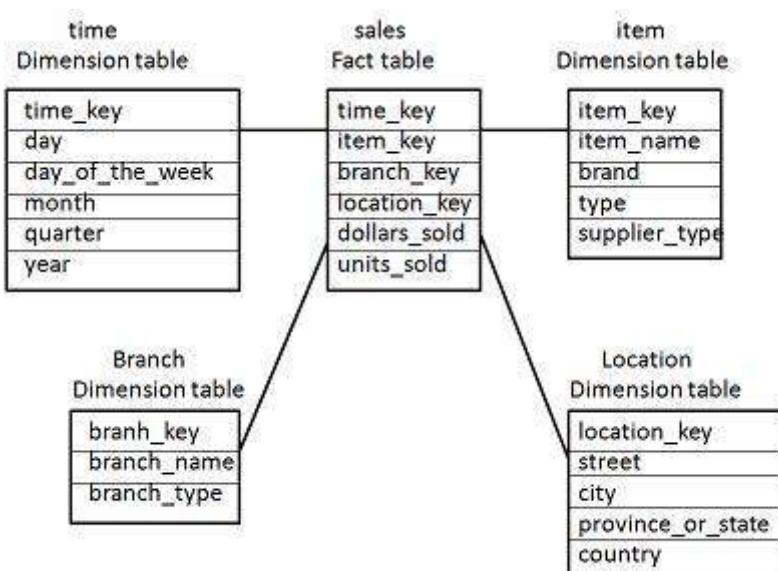
Skema pada data warehouse adalah struktur organisasi data yang menentukan bagaimana data disimpan, diatur, dan dihubungkan. Ini mencakup bagaimana tabel-tabel data dan relasinya dirancang untuk mendukung analisis dan pelaporan bisnis. Skema ini mencakup tabel fakta yang berisi data bisnis inti dan tabel dimensi yang menyediakan konteks terhadap data fakta. Skema dapat berupa skema bintang, skema salju, atau variasi lainnya tergantung pada kebutuhan dan kompleksitas data bisnis yang dikelola. Tujuan utama dari skema dalam data warehouse adalah untuk menyediakan struktur yang memudahkan akses, analisis, dan pelaporan data untuk mendukung pengambilan keputusan bisnis.

1. Star Schema

Setiap dimensi dalam skema star direpresentasikan dengan hanya satu tabel dimensi.

Tabel dimensi ini berisi sekumpulan atribut.

Diagram berikut menunjukkan data penjualan sebuah perusahaan dengan empat dimensi, yaitu waktu, item, cabang, dan lokasi.



Gambar 5. Skema Star

Ada sebuah tabel fakta di tengah. Tabel ini berisi kunci untuk masing-masing dari empat dimensi.

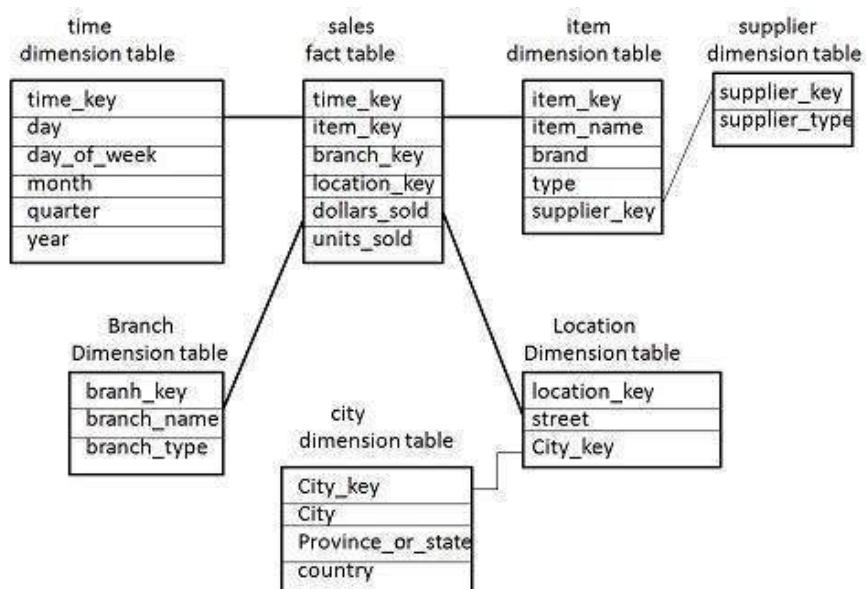
Tabel fakta juga berisi atribut-atribut, yaitu jumlah penjualan dalam dolar dan unit yang terjual.

2. Snowflake Schema

Beberapa tabel dimensi dalam Snowflake schema dinormalisasi.

Normalisasi memecah data ke dalam tabel tambahan.

Berbeda dengan Star schema, tabel dimensi dalam Snowflake schema adalah dalam bentuk yang dinormalisasi. Misalnya, tabel dimensi item dalam skema bintang adalah dalam bentuk yang dinormalisasi dan dipisah menjadi dua tabel dimensi, yaitu tabel item dan tabel pemasok.



Gambar 6. Snowflake Star

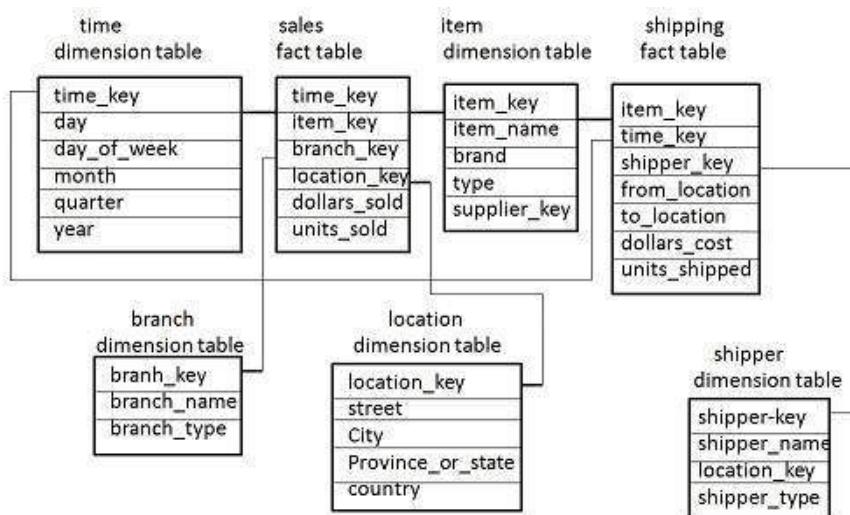
Sekarang tabel dimensi item berisi atribut-atribut item_key, item_name, tipe, merek, dan supplier-key.

Supplier key terhubung ke tabel dimensi supplier. Tabel dimensi supplier berisi atribut-atribut supplier_key dan supplier_type.

3. Fact Constellation Schema

Fact constellation memiliki beberapa tabel fakta. Ini juga dikenal sebagai skema galaksi (galaxy schema).

Diagram berikut menunjukkan dua tabel fakta, yaitu penjualan (sales) dan pengiriman (shipping).



Gambar 7. Fact Constellation Schema

Tabel fakta penjualan sama dengan yang ada dalam skema bintang (star schema).

Tabel fakta pengiriman memiliki lima dimensi, yaitu item_key, time_key, shipper_key, dari_lokasi, ke_lokasi.

Tabel fakta pengiriman juga berisi dua ukuran, yaitu jumlah penjualan dalam dolar dan unit yang terjual.

Mungkin juga berbagi tabel dimensi antara tabel fakta. Sebagai contoh, tabel dimensi waktu, item, dan lokasi dibagi antara tabel fakta penjualan dan tabel fakta pengiriman.

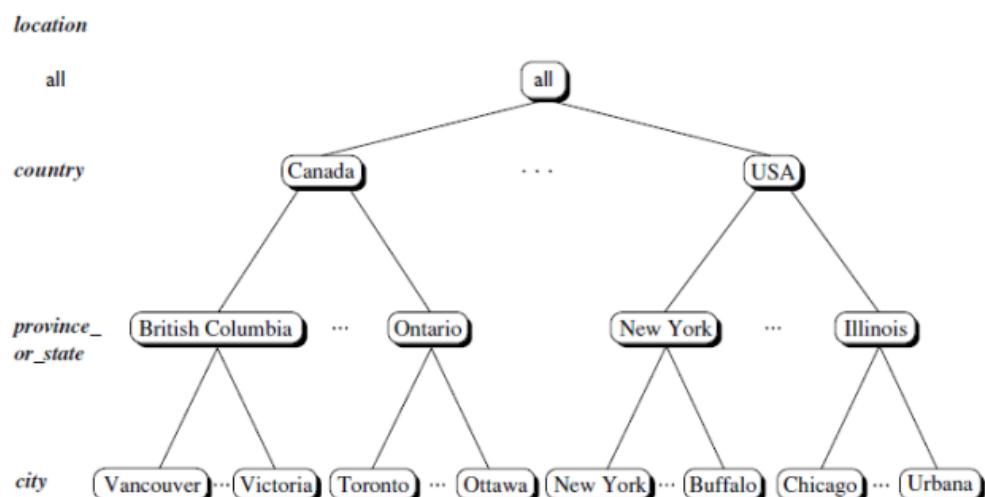
6. Hirarki Konsep

Hirarki Konsep mendefinisikan urutan pemetaan dari sekumpulan konsep tingkat rendah ke konsep yang lebih tinggi dan lebih umum.

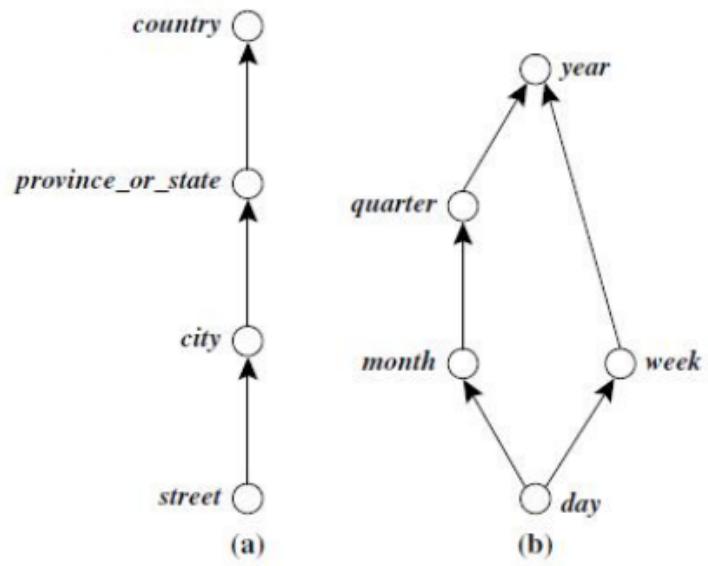
Pertimbangkan sebuah hirarki konsep untuk dimensi lokasi. Nilai kota untuk lokasi meliputi Vancouver, Toronto, New York, dan Chicago.

Namun, setiap kota dapat dipetakan ke provinsi atau negara bagian tempatnya berada.

Sebagai contoh, Vancouver dapat dipetakan ke British Columbia, dan Chicago ke Illinois.



Gambar 8. Hirarki Konsep untuk location



Gambar 9. Struktur Hierarki dan Struktur Jaringan.

Struktur Hierarki dan Struktur Jaringan Atribut dalam Dimensi Gudang: (a) hierarki untuk lokasi dan (b) jaringan untuk waktu.

7. OLAP (Online Analytical Processing)

Dalam model multidimensional, data diorganisir ke dalam beberapa dimensi, dan setiap dimensi berisi beberapa tingkat abstraksi yang didefinisikan oleh hierarki konsep. Organisasi ini memberikan fleksibilitas kepada pengguna untuk melihat data dari berbagai sudut pandang. Ada sejumlah operasi kubus data OLAP yang digunakan untuk menyajikan pandangan-pandangan yang berbeda ini, memungkinkan kueri interaktif dan analisis data yang ada. Oleh karena itu, OLAP memberikan lingkungan yang ramah pengguna untuk analisis data interaktif.

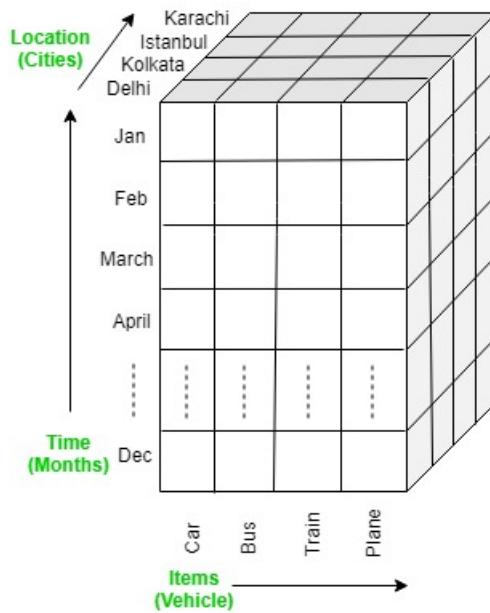
OLAP operations:

1. Drill down

Dalam operasi drill-down, data yang kurang terperinci dikonversi menjadi data yang sangat terperinci. Ini dapat dilakukan dengan:

- a. Pergi ke bawah dalam hierarki konsep
- b. Menambah dimensi baru

Pada kubus yang diberikan dalam bagian gambaran, operasi drill-down dilakukan dengan memindah ke bawah dalam hierarki konsep dimensi Waktu (Kuartal -> Bulan).



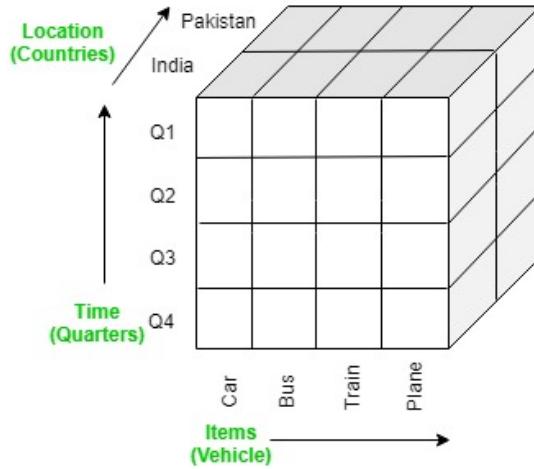
Gambar 10. Drill Down

2. Roll up

Ini adalah kebalikan dari operasi drill-down. Ini melakukan agregasi pada kubus OLAP. Ini dapat dilakukan dengan:

- Naik ke atas dalam hierarki konsep
- Mengurangi dimensi

Pada kubus yang diberikan dalam bagian gambaran, operasi roll-up dilakukan dengan naik ke atas dalam hierarki konsep dimensi Lokasi (Kota -> Negara).

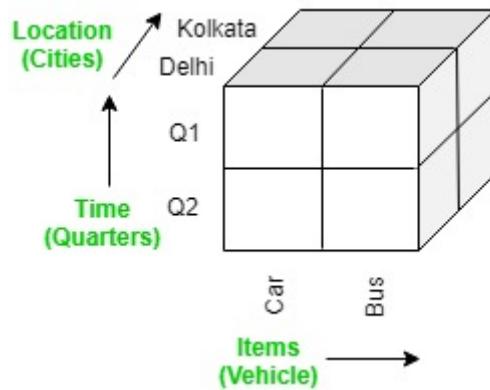


Gambar 11. Drill Down

3. Dice

Ini memilih sub-kubus dari kubus OLAP dengan memilih dua atau lebih dimensi. Pada kubus yang diberikan dalam bagian gambaran, sebuah sub-kubus dipilih dengan memilih dimensi-dimensi berikut dengan kriteria:

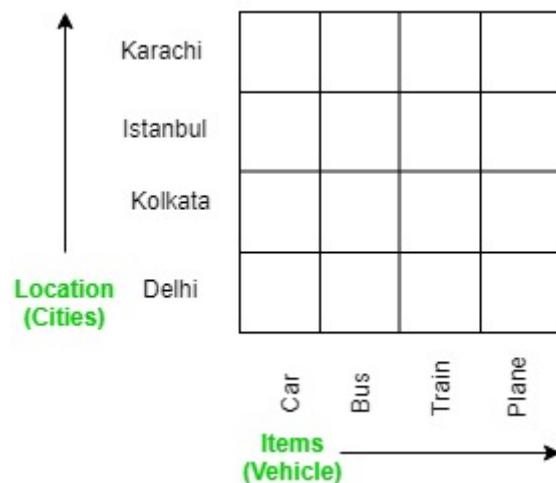
- a. Lokasi = "Delhi" atau "Kolkata"
- b. Waktu = "Q1" atau "Q2"
- c. Item = "Mobil" atau "Bus"



Gambar 12. Dice

4. Slice

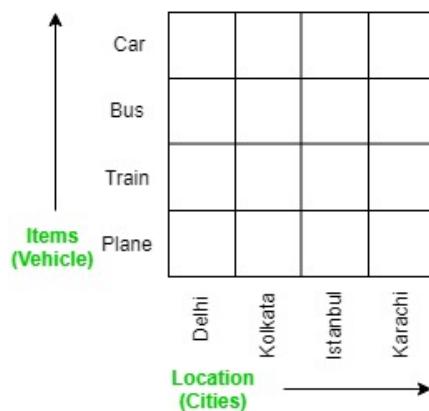
Ini memilih satu dimensi dari kubus OLAP yang menghasilkan pembuatan sub-kubus baru. Pada kubus yang diberikan dalam bagian gambaran, Slice dilakukan pada dimensi Waktu = "Q1".



Gambar 13. Slice

5. Pivot

Ini juga dikenal sebagai operasi rotasi karena memutar tampilan saat ini untuk mendapatkan tampilan representasi yang baru. Dalam sub-kubus yang diperoleh setelah operasi Slice, melakukan operasi pivot memberikan tampilan yang baru darinya.



Gambar 13. Pivot

8. Kelebihan Data Cube

1. Analisis Multi-Dimensi: Data cube memungkinkan analisis multi-dimensi data bisnis, memungkinkan pengguna melihat data dari berbagai perspektif dan tingkat detail.
2. Interaktivitas: Data cube memberikan akses interaktif ke sejumlah besar data, memungkinkan pengguna dengan mudah menavigasi dan memanipulasi data untuk mendukung analisis mereka.
3. Kecepatan dan Efisiensi: Data cube dioptimalkan untuk analisis OLAP, memungkinkan kueri dan agregasi data yang cepat dan efisien.
4. Agregasi Data: Data cube mendukung perhitungan kompleks dan agregasi data, memungkinkan pengguna dengan cepat dan mudah merangkum sejumlah besar data.
5. Peningkatan Pengambilan Keputusan: Data cube memberikan pandangan yang jelas dan komprehensif tentang data bisnis, memungkinkan pengambilan keputusan yang lebih baik dan intelijen bisnis.
6. Aksesibilitas: Data cube dapat diakses dari berbagai perangkat dan platform, memudahkan pengguna untuk mengakses dan menganalisis data bisnis dari mana saja.
7. Membantu memberikan pandangan yang ringkas tentang data.
8. Data cube menyimpan data besar dengan cara yang sederhana.
9. Operasi data cube memberikan analisis yang cepat dan lebih baik.
10. Meningkatkan kinerja data.

9. Kekurangan Data Cube

1. Kompleksitas: Sistem OLAP dapat kompleks dalam pengaturan dan pemeliharaannya, memerlukan keahlian teknis khusus.
2. Batasan Ukuran Data: Sistem OLAP dapat mengalami kesulitan dengan kumpulan data yang sangat besar dan mungkin memerlukan agregasi atau ringkasan data yang ekstensif.
3. Masalah Kinerja: Sistem OLAP dapat lambat ketika berurusan dengan sejumlah besar data, terutama ketika menjalankan kueri atau perhitungan yang kompleks.
4. Integritas Data: Definisi data yang inkonsisten dan masalah kualitas data dapat memengaruhi akurasi analisis OLAP.
5. Biaya: Teknologi OLAP dapat mahal, terutama untuk solusi tingkat perusahaan, karena memerlukan perangkat keras dan perangkat lunak khusus.
6. Ketidakluwesan: Sistem OLAP mungkin tidak dengan mudah menyesuaikan dengan perubahan kebutuhan bisnis dan mungkin memerlukan upaya yang signifikan untuk memodifikasi atau memperluasnya.

10. Implementasi Data Cube Di Dunia Nyata

Dilansir dari Dr. Omid Shabestari, Director of Health Analytics dan Shea Jessee, Team Lead, dari Carilion Clinic.

Implementasi "Health Data Cube" di Carilion Clinic, dengan fokus pada penggunaan data cube dalam analitik kesehatan. Data cube digunakan untuk menjawab pertanyaan penting, seperti rata-rata jumlah pesanan obat per kunjungan pasien dan peringkat kepuasan pasien. Implementasi data cube ini memungkinkan akses lebih cepat dan efisien ke data yang sebelumnya hanya tersedia melalui alat pelaporan.

Data cube ini dibangun di atas data dari berbagai sumber, termasuk data rekam medis elektronik (EMR), survei kepuasan pasien, dan data lainnya. Cube ini memungkinkan perhitungan agregat dan perbandingan data dengan cepat. Data cube ini diproses setiap malam sehingga data yang tersedia selalu up-to-date dalam waktu 24 jam. Cube ini dapat digunakan dengan berbagai alat analisis, seperti lembar kerja dan alat visualisasi data.

Pengembangan data warehouse dan data cube memerlukan perubahan organisasi, termasuk pendirian dewan pengaturan data dan kelompok pengurus data. Cube ini bertujuan untuk menjadi sumber data tunggal dalam organisasi dan mendukung pengambilan keputusan yang lebih baik.

11. Kesimpulan

Data Cube adalah konsep yang mewakili representasi multidimensional dari data, memungkinkan penyimpanan dan analisis data dari berbagai dimensi seperti waktu, lokasi, dan atribut lainnya. Data Cube memungkinkan pengguna untuk melihat data dari berbagai sudut pandang dan menggabungkan konsep hierarki dalam analisis. Ini juga berhubungan erat dengan OLAP (Online Analytical Processing), yang memungkinkan pengguna menjelajahi dan menganalisis data dengan lebih mendalam, membuat keputusan yang lebih baik, dan merespons perubahan dengan lebih cepat dalam berbagai konteks bisnis dan ilmiah.

12. Referensi

- What is Data Cube?

<https://www.javatpoint.com/data-warehouse-what-is-data-cube>

- Data Warehousing - Schemas

https://www.tutorialspoint.com/dwh/dwh_schemas.htm

- OLAP Operations in the Multidimensional Data Model

<https://www.javatpoint.com/olap-operations>

- OLAP Operations in DBMS

<https://www.geeksforgeeks.org/olap-operations-in-dbms/>

- Health Data Cube – Implementation of a data pre-aggregation platform for improving information availability

[https://www.healthtechmagazines.com/health-data-cube-implementation-of-a-data-pr
e-aggregation-platform-for-improving-information-availability/](https://www.healthtechmagazines.com/health-data-cube-implementation-of-a-data-pre-aggregation-platform-for-improving-information-availability/)

- Data Cube or OLAP approach in Data Mining

<https://www.geeksforgeeks.org/data-cube-or-olap-approach-in-data-mining/>