**Mempersiapkan Data untuk Eksplorasi**

**Semua Hal Tentang Data Base**

Database merupakan kumpulan data yang tersimpan di sistem komputer, tetapi sistem penyimpanan ini hanyalah awal dari pembahasan. Anda akan mempelajari cara kerja database untuk menemukan informasi yang Anda butuhkan dengan tepat untuk analisis Anda.

Metadata adalah data tentang data. Seperti yang sudah saya katakan sebelumnya: mendalam. Metadata sangat penting ketika bekerja dengan database. Anggap saja seperti panduan dari sebuah referensi. Tanpa panduan, semua yang Anda miliki adalah sekelompok data tanpa penjelasan konteks apa arti dari data-data tersebut Metadata menjelaskan dari mana asal data, kapan, dan bagaimana data itu dibuat, serta semua hal yang menggambarkan data itu.

**Database relasional**

Database relasional adalah database berisi serangkaian tabel yang dapat dihubungkan untuk memperlihatkan relasi di antara mereka. Pada dasarnya, database relasional memungkinkan analis data untuk mengatur dan menautkan data berdasarkan kesamaan data.

Pada tabel non-relasional, Anda akan menemukan semua variabel yang mungkin menarik bagi Anda untuk dianalisis dan dikelompokkan bersama. Hal ini dapat menyulitkan penyortiran. Inilah salah satu alasan mengapa database relasional umum digunakan dalam analisis data: mereka menyederhanakan banyak proses analisis dan membuat data lebih mudah ditemukan dan digunakan di seluruh database.

* **Primary key** adalah pengidentifikasi nilai keunikan tiap data yang merujuk pada suatu kolom database. Anda bisa menganggapnya sebagai penanda unik untuk setiap baris di tabel.
* **Foreign key** adalah bidang di sebuah tabel yang merupakan primary key di tabel lain. Atau, foreign key adalah bagaimana satu tabel dapat dihubungkan ke tabel lainnya.

Primary key digunakan untuk memastikan data di kolom tertentu adalah unik. Key ini secara unik mengidentifikasi suatu record di sebuah tabel database relasional. Hanya satu primary key yang diperbolehkan dalam satu tabel dan nilainya tidak boleh berisi nol atau kosong. Foreign key adalah kolom atau kelompok kolom di tabel database relasional yang membuat ada hubungan antara data dalam dua tabel. Key ini mengacu pada bidang dalam tabel yang merupakan primary key dari tabel lain. Terakhir, penting untuk dicatat bahwa satu tabel diperbolehkan memiliki lebih dari satu foreign key.

Database menggunakan bahasa khusus untuk berkomunikasi yang disebut bahasa kueri (query). Structured Query Language (SQL) adalah jenis bahasa kueri yang memungkinkan analis data berkomunikasi dengan database. Jadi, seorang analis data akan menggunakan SQL untuk membuat kueri guna melihat data spesifik yang mereka inginkan dari kumpulan data yang lebih besar. Dalam database relasional, analis data dapat menulis kueri untuk mendapatkan data dari tabel-tabel terkait. SQL adalah alat yang ampuh untuk bekerja dengan database. Itulah sebabnya Anda akan mempelajari lebih lanjut tentang hal itu.

**Metadata sama pentingnya dengan data itu sendiri**

Metadata adalah data tentang data. Dalam pengelolaan database, itu menyediakan informasi tentang data lain dan membantu analis data menafsirkan isi data dalam sebuah database.Terlepas dari apakah Anda bekerja dengan data dalam jumlah besar atau kecil, metadata adalah tanda yang diberikan oleh tim analitis berpengetahuan luas, yang membantu mengomunikasikan data di seluruh perusahaan dan mempermudah penggunaan data kembali. Pada intinya, metadata memberitahu hal-hal terkait siapa, apa, kapan, di mana, yang mana, bagaimana, dan mengapa tentang data.

**Jenis-jenis Metadata**

* **Metadata deskriptif** adalah metadata yang menggambarkan sekeping data dan dapat digunakan untuk mengidentifikasinya di lain waktu. Misalnya, metadata deskriptif dari sebuah buku di perpustakaan akan menyertakan kode yang Anda lihat di tepi buku, yang dikenal sebagai Internasional Standard Book Number yang unik, disebut juga ISBN.
* **Metadata struktural**, yang merupakan metadata yang menunjukkan bagaimana sekeping data diatur dan apakah itu bagian dari satu atau lebih kumpulan data. Mari kita kembali ke perpustakaan. Contoh data struktural adalah bagaimana halaman-halaman dari sebuah buku disatukan untuk membuat bab yang berbeda. Penting untuk dicatat bahwa metadata struktural juga melacak hubungan antara dua hal. Misalnya, metadata dapat menunjukkan kepada kita bahwa dokumen digital dari suatu manuskrip buku sebenarnya adalah versi asli dari buku yang sekarang sedang dicetak.
* **Metadata administratif,** adalah metadata yang menunjukkan sumber teknis suatu aset digital. Saat kita melihat metadata di dalam foto, itu adalah metadata administratif. Metadata ini menunjukkan kepada Anda jenis file itu, tanggal dan waktu pengambilannya, dan banyak lagi.

**Elemen-elemen metadata**

Sebelum melihat contoh-contoh metadata, penting untuk memahami jenis informasi yang biasanya disediakan oleh metadata.

* **Judul dan deskripsi**, Apa nama file atau situs web yang sedang Anda periksa? Apa jenis konten yang dikandungnya?
* **Tag dan kategori**, Apa gambaran umum dari data yang Anda miliki? Apakah data diindeksasi atau dijelaskan dengan cara tertentu?
* **Siapa yang membuatnya dan kapan**, Dari mana data itu berasal, dan kapan dibuat? Apakah baru-baru ini, atau sudah ada sejak lama?
* **Siapa yang terakhir mengubahnya dan kapan**, Apakah ada perubahan yang dilakukan pada data? Jika ya, apakah perubahannya baru-baru ini?
* **Siapa yang dapat mengakses atau memperbaruinya**, Apakah kumpulan data ini bersifat publik? Apakah ada izin khusus yang diperlukan untuk menyesuaikan atau memodifikasi dataset itu?

**Contoh metadata**

Di dunia digital saat ini, metadata ada di mana-mana, dan menjadi praktik yang lebih umum untuk menyediakan metadata di banyak media dan informasi yang berinteraksi dengan Anda. Berikut adalah beberapa contoh dunia nyata tempat metadata dapat ditemukan:

* **Foto**. Setiap kali foto diambil dengan kamera, metadata seperti nama file kamera, tanggal, waktu, dan geolokasi dikumpulkan dan disimpan dengannya.
* **Email**. Saat email dikirim atau diterima, ada banyak metadata yang terlihat seperti baris subjek, pengirim, penerima, serta tanggal dan waktu pengiriman. Ada juga metadata tersembunyi yang mencakup nama server, alamat IP, format HTML, dan detail perangkat lunak.
* **Spreadsheet dan dokumen**, Spreadsheet dan dokumen sudah diisi dengan data yang cukup banyak sehingga tidak mengherankan jika metadata juga akan menyertainya. Judul, penulis, tanggal pembuatan, jumlah halaman, komentar pengguna, serta nama tab, tabel, dan kolom adalah metadata yang dapat ditemukan di spreadsheet dan dokumen.
* **Situs web**, Setiap halaman web memiliki sejumlah bidang standar tentang metadata, seperti tag dan kategori, nama pembuat situs, judul dan deskripsi halaman web, waktu pembuatan, dan ikonografi apa pun.
* **File digital**, Biasanya, jika Anda mengklik kanan pada file komputer mana pun, Anda akan melihat metadata-nya. Ini bisa terdiri dari nama file, ukuran file, tanggal pembuatan dan modifikasi, serta jenis file.
* **Buku,** Metadata tidak hanya digital. Setiap buku memiliki sejumlah metadata standar di sampul dan di dalamnya yang akan memberi tahu Anda tentang judulnya, nama penulis, daftar isi, informasi penerbit, deskripsi hak cipta, indeks, dan deskripsi singkat tentang isi buku.

Metadata menjadi sumber tunggal yang tepercaya karena data di dalamnya selalu konsisten dan seragam. Metadata juga membuat data lebih handal, akurat, tepat, relevan, dan dapat diakses tepat waktu. Metadata juga memudahkan analis data untuk mengidentifikasi akar penyebab dari tiap masalah yang mungkin akan timbul.

**Respositori Metadata**

Adalah database yang khusus dibuat sebagai penyimpanan metadata. Repositori metadata dapat disimpan dalam bentuk fisik, ataupun virtual, seperti penyimpanan data melalui sistem cloud. Repositori ini bisa menyediakan informasi dari mana metadata berasal, menyimpannya sehingga dapat diakses dengan cepat dan mudah, serta menyimpan metadata dalam struktur yang umum sehingga semua orang dapat menggunakannya. Repositori metadata membuat metadata yang didapat dari banyak sumber menjadi lebih mudah dan lebih cepat disatukan untuk digunakan dalam analisis data. Repositori bekerja dengan cara menggambarkan keadaan dan lokasi asal metadata, struktur tabel di dalamnya, dan bagaimana data mengalir melalui repositori. Repositori bahkan mampu melacak siapa saja yang mengakses metadata dan kapan waktu mengaksesnya.

Metadata disimpan di satu tempat yang terpusat dan memberikan informasi yang didapat melalui data yang sudah distandarkan untuk digunakan perusahaan. Cara ini dilakukan melalui dua cara. Pertama, metadata memiliki informasi mengenai keberadaan tiap sistem dan letak dataset yang ada di dalam sistem-sistem tersebut. Kedua, metadata dapat menjelaskan bagaimana semua data terhubung di antara berbagai sistem.

**Tata kelola data** adalah sebuah proses manajemen formal yang mengelola aset data di perusahaan. Tata kelola ini memberikan suatu organisasi cara mengontrol data yang lebih baik serta membantu perusahaan mengelola masalah yang terkait dengan keamanan dan privasi data, integritas data, penggunaan, serta aliran data internal maupun eksternal.