

Making Data visualization More Efficient and Effective

Visualisasi data merupakan proses mengubah data abstrak menjadi bentuk visual, seperti grafik, diagram, dan peta. Tujuan utama dari visualisasi data adalah untuk mempermudah pemahaman dan interpretasi informasi kompleks dengan cara yang intuitif dan mudah dipahami. Dengan meningkatnya volume data yang dihasilkan oleh berbagai organisasi, visualisasi data menjadi alat yang penting untuk menyajikan informasi yang besar dan kompleks dalam format yang lebih mudah diakses dan dianalisis. Sehingga memungkinkan pengguna untuk dengan cepat mengenali pola, tren, dan hubungan dalam data serta mendukung pengambilan keputusan yang lebih baik berdasarkan wawasan yang diperoleh dari visualisasi tersebut.

Pipeline Visualisasi Data:

Pipeline visualisasi data terdiri dari lima langkah utama yang saling terhubung:

1. **Pengambilan Data:** Mengumpulkan data dari berbagai sumber yang relevan, seperti basis data, file CSV, API, atau sensor. Data yang dikumpulkan harus sesuai dengan kebutuhan analisis dan visualisasi yang diinginkan.
2. **Persiapan Data:** Setelah data dikumpulkan, langkah selanjutnya adalah menyiapkan data untuk visualisasi. Hal ini melibatkan normalisasi data untuk memastikan konsistensi, perbaikan kesalahan atau anomali, dan pengisian nilai yang hilang melalui interpolasi. Data yang bersih dan terstruktur merupakan kunci untuk visualisasi yang efektif.
3. **Manipulasi Data:** Pada langkah ini, data yang telah disiapkan diolah lebih lanjut untuk visualisasi. Hal ini mencakup penyaringan data untuk memilih subset yang relevan, penggabungan data dari berbagai sumber, dan pengelompokan data untuk analisis yang lebih tepat.
4. **Pemetaan:** Data yang telah diproses dipetakan ke dalam bentuk geometris seperti titik, garis, atau area. Atribut visual seperti warna, posisi, dan ukuran ditentukan untuk memberikan representasi visual yang sesuai dengan tujuan analisis.

5. **Rendering:** Proses rendering menghasilkan representasi visual dari data yang telah dipetakan. Tahap ini adalah langkah akhir di mana data yang telah diproses dan dipetakan diubah menjadi grafik, diagram, atau peta yang dapat dilihat dan dipahami oleh pengguna.

Spesifikasi Visualisasi:

Spesifikasi visualisasi menentukan cara data akan ditampilkan dalam visualisasi. Beberapa pendekatan spesifikasi meliputi:

1. **Bahasa Tingkat Rendah:** Menggunakan bahasa pemrograman seperti D3.js dan Vega, yang memungkinkan spesifikasi mendetail untuk setiap elemen visualisasi. Pendekatan ini memberikan kontrol penuh terhadap elemen visual tetapi memerlukan pemahaman teknis yang tepat.
2. **Bahasa Tingkat Tinggi:** Bahasa seperti ggplot2 dan Vega-Lite menawarkan default yang masuk akal dan menyederhanakan pembuatan visualisasi tanpa memerlukan pengetahuan teknis yang mendalam. Sehingga memudahkan pengguna untuk membuat visualisasi yang efektif dengan lebih sedikit konfigurasi.
3. **Alat Berbasis GUI:** Alat seperti Tableau dan Microsoft Power BI memungkinkan pengguna untuk membuat visualisasi melalui antarmuka grafis tanpa menulis kode. Sehingga membuat visualisasi data lebih mudah diakses bagi pengguna yang tidak memiliki keterampilan pemrograman.
4. **Spesifikasi yang Kurang Ditetapkan:** Pendekatan ini mengizinkan pengguna memberikan "petunjuk" atau panduan kepada sistem visualisasi, yang kemudian mencoba menginterpretasikan dan melengkapi spesifikasi tersebut. Sehingga memungkinkan fleksibilitas dalam visualisasi tanpa harus menentukan semua detail secara eksplisit.

Pendekatan Efisien untuk Visualisasi Data:

Berikut ini merupakan teknik digunakan untuk meningkatkan efisiensi visualisasi data:

1. **Visualisasi Data yang Tepat:**

- **Penerjemahan Kueri:** Mengubah kueri visualisasi menjadi kueri SQL yang diproses oleh sistem manajemen basis data (DBMS) untuk memanfaatkan mesin pemrosesan data yang ada, mempercepat proses pengambilan data.
- **Integrasi Sistem Visualisasi dengan DBMS:** Mengintegrasikan proses pengambilan data dan rendering untuk mengurangi waktu yang dibutuhkan untuk menghasilkan visualisasi.
- **Penggunaan Column Stores dan Indeks:** Meningkatkan kinerja pencarian dan agregasi data dengan menggunakan teknik penyimpanan kolom dan pembuatan indeks.
- **Komputasi Paralel:** Memproses beberapa kueri secara bersamaan untuk mempercepat visualisasi dan mengurangi waktu pemrosesan.

2. Visualisasi Data Perkiraan:

- **AQP Berbasis Sampling:** Menggunakan subset data yang representatif untuk menghasilkan visualisasi perkiraan dengan cepat, mengurangi kebutuhan untuk memproses seluruh dataset.
- **Sampling Inkremental:** Menghasilkan visualisasi awal dengan sampel data kecil dan kemudian meningkatkan kualitas visualisasi dengan menambah ukuran sampel secara bertahap.
- **Pendekatan Berbasis Persepsi Manusia:** Menghentikan sampling ketika penambahan data tidak secara signifikan meningkatkan kualitas visualisasi menurut persepsi manusia, mengoptimalkan proses visualisasi.

3. Visualisasi Data Progresif:

- **Agregasi Hirarkis:** Menggunakan struktur hirarkis untuk mengagregasi data pada berbagai tingkat, memungkinkan eksplorasi data secara bertahap dan memberikan pandangan yang lebih tepat.

Rekomendasi Visualisasi:

1. **Rekomendasi Berbasis Spesifikasi:** Menerima spesifikasi visualisasi yang tidak lengkap dan menyarankan visualisasi yang sesuai berdasarkan aturan atau algoritma machine learning, membantu pengguna menemukan opsi yang tepat.
2. **Rekomendasi Berbasis Perilaku:** Mempelajari perilaku pengguna sebelumnya untuk memprediksi visualisasi yang mungkin diinginkan di masa depan, mempercepat proses eksplorasi data dan memberikan saran yang lebih relevan.
3. **Rekomendasi yang Dipersonalisasi:** Menyesuaikan rekomendasi visualisasi berdasarkan preferensi dan kebutuhan spesifik pengguna, menggunakan algoritma machine learning untuk mempelajari preferensi pengguna dari waktu ke waktu dan meningkatkan pengalaman pengguna.