



Virtual Internship Experience

Data Pipeline

Konsep Data Pipeline

Data Pipeline

Data pipeline adalah sarana untuk memindahkan data dari satu tempat (sumber) ke tujuan (seperti gudang data). Sepanjang jalan, data diubah dan dioptimalkan, tiba dalam keadaan yang dapat dianalisis dan digunakan untuk mengembangkan wawasan bisnis. Sebuah Data pipeline pada dasarnya adalah langkah-langkah yang terlibat dalam menggabungkan, mengatur, dan memindahkan data. Data pipeline modern mengotomatiskan banyak langkah manual, ini termasuk memuat data mentah ke dalam tabel staging untuk penyimpanan sementara dan kemudian mengubahnya sebelum akhirnya memasukkannya ke dalam tabel tujuan.

Suatu organisasi pastinya berurusan dengan sejumlah data yang besar. Untuk menganalisis semua data itu, memerlukan satu tampilan dari seluruh kumpulan data. Ketika data itu berada di beberapa sistem dan layanan, itu perlu digabungkan dengan cara yang masuk akal untuk analisis mendalam. Aliran data itu sendiri bisa tidak dapat diandalkan: ada banyak titik selama pengangkutan dari satu sistem ke sistem lain di mana korupsi atau kemacetan dapat terjadi. Ketika luas dan cakupan dari peran yang dimainkan data meningkat, masalahnya hanya diperbesar dalam skala dan dampak.

Itulah mengapa jalur data pipeline sangat penting. Mereka menghilangkan sebagian besar langkah manual dari proses dan memungkinkan aliran data otomatis yang lancar dari satu tahap ke tahap lainnya. Data pipeline penting untuk analitik real-time untuk membantu membuat keputusan yang lebih cepat dan berdasarkan data. Beberapa keuntungan data pipeline:

1. Analisis data secara real time
2. Menyimpan data di cloud
3. Menyimpan data dari berbagai sumber

Dengan menggabungkan data dari berbagai silo menjadi satu sumber kebenaran, kita dapat memastikan kualitas data yang konsisten dan memungkinkan analisis data yang lebih cepat untuk mendapatkan wawasan bisnis.



Sumber:

1. <https://www.snowflake.com/guides/data-pipeline>