**TEAM ACTIVITY**

**🍮TEAM 2 🍮**

1. Jelaskan dan gambarkan tahapan membuat *Git add, Git status, Git commit, Git log* dan *Git push*!
   1. **Git add**

Add files or directory to staging area

git add [filename.file]

* 1. **Git Status**

Display the state of the repository and staging area

git status

* 1. **Git Commit**

Add files or directory to commit

git commit -m “[add comment]”

* 1. **Git Log**

git log

* 1. **Git Push**

git push [branch name]

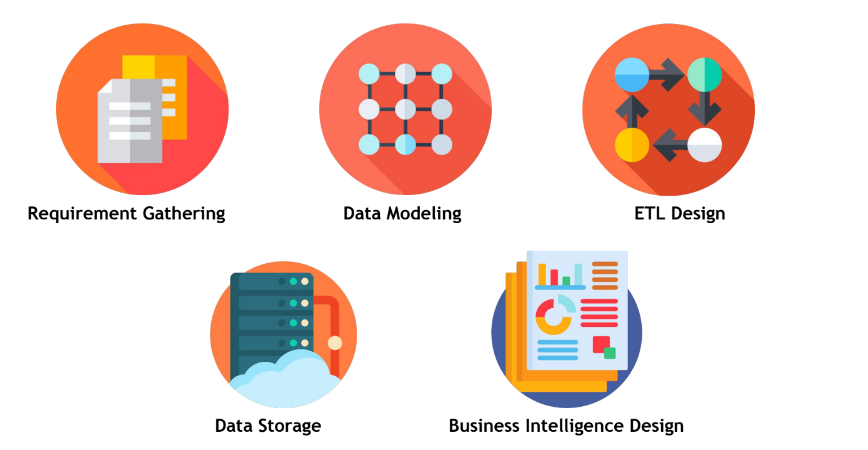
git push origin [branch name]

git push origin master

**TEAM ACTIVITY**

**🍮TEAM 2 🍮**

1. Jelaskan dan gambarkan tahapan mendesain dan membangun *Data Warehouse*!



* Requirement Gathering

Proses untuk mengumpulkan, menganalisis, dan mendokumentasikan kebutuhan bisnis, teknis, dan fungsional yang diperlukan untuk merancang dan mengembangkan data warehouse yang efektif dan sesuai dengan tujuan organisasi.

Beberapa langkah penting dalam Requirement Gathering pada data warehouse:

* Identifikasi semua pihak terkait yang akan menggunakan data warehouse, memahami kebutuhan mereka. Tanyakan pertanyaan terkait tujuan bisnis, jenis data yang diperlukan, tingkat detail yang diharapkan, serta kebutuhan pelaporan dan analisis.
* Mengidentifikasi Sumber Data.
* Menentukan Kualitas Data.
* Merancang Arsitektur Data.
* Dokumentasi semua informasi yang dikumpulkan, termasuk kebutuhan fungsional, teknis, dan bisnis. Dokumentasi ini akan menjadi panduan untuk pengembangan data warehouse.
* Validasi Kebutuhan.
* Perencanaan Proyek.
* Data Modeling

Proses merancang struktur data yang akan digunakan dalam data warehouse untuk menyimpan, mengelola, dan mengakses data dengan efisien. Model data ini memberikan panduan tentang bagaimana data akan diorganisasi, direpresentasikan, dan dihubungkan dalam data warehouse. Terdapat dua jenis utama dari data modeling dalam konteks data warehouse: Dimensional Modeling dan Normalized Modeling.

Fakta (Fact): Ini adalah tabel yang berisi data yang dapat diukur dan berhubungan dengan kejadian bisnis tertentu. Contohnya adalah tabel penjualan yang berisi data seperti jumlah penjualan, tanggal penjualan, dan lokasi penjualan.

Dimensi (Dimension): Ini adalah tabel yang berisi atribut yang digunakan untuk menggambarkan data fakta. Misalnya, dimensi waktu dapat mencakup atribut seperti tahun, bulan, dan hari. Dimensi pelanggan dapat mencakup atribut seperti nama pelanggan dan alamat.

* ETL Design
* Data Storage
* Business Intelligence Design

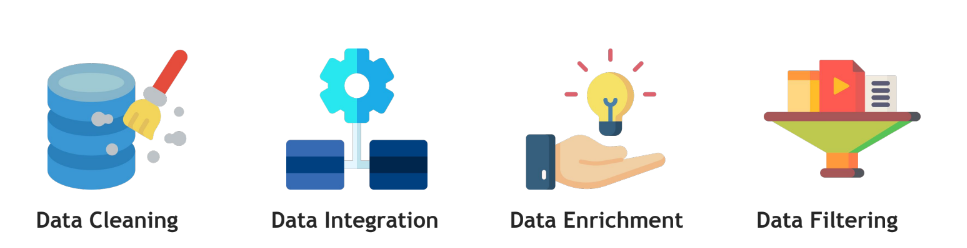
**TEAM ACTIVITY**

**🍮TEAM 2 🍮**

1. Sebutkan 3 teknik extract data, lalu jelaskan masing-masing teknik tersebut biasa digunakan untuk mengambil tipe data apa dari sumber seperti apa!

* **Web Scraping**, dapat digunakan untuk mengambil berita dari situs berita, data penjualan dari situs e-commerce, atau informasi lainnya yang tersedia secara publik di internet. Proses ini melibatkan pengacakan elemen HTML, CSS, atau XPath pada halaman web target, dan dalam beberapa kasus, menggunakan API yang disediakan oleh situs web tersebut. Sumber data yang biasa diperoleh menggunakan teknik ini berupa **html, xml, dan json**. Biasanya, untuk melakukan proses web scraping ini menggunakan **API** yang disediakan oleh situs web yang ingin kita peroleh datanya dan menggunakan tools atau library seperti **beautifulsoap, scrapy, parsehub, dkk**.
* **API**, teknik ekstraksi data menggunakan API adalah pendekatan yang umumnya digunakan untuk mengambil data terstruktur dari berbagai sumber data yang menyediakan antarmuka pemrograman aplikasi. Jenis data yang dapat diekstraksi melalui API sangat bervariasi tergantung pada sumber data yang digunakan dan endpoint API yang tersedia. Data yang diambil melalui API umumnya berupa data terstruktur, seperti **JSON atau XML**, yang dapat dengan mudah diproses dan dianalisis. Sumber data yang cocok untuk teknik ini mencakup platform media sosial seperti **Twitter** atau **Facebook**, layanan cuaca seperti **OpenWeatherMap**, layanan geolokasi seperti **Google Maps**, informasi keuangan dari layanan seperti Alpha **Vantage** atau **Yahoo Finance**, serta data publik dari berbagai **organisasi** atau **pemerintah**. Langkah-langkah umum dalam menggunakan teknik ini termasuk memahami dokumentasi API, mengautentikasi untuk mendapatkan key API jika diperlukan, mengirim permintaan API dengan parameter yang sesuai, mengolah response API, menangani potensi kesalahan, dan memantau serta memelihara integrasi API seiring waktu. Dengan memanfaatkan API, kita dapat mengintegrasikan data dari berbagai sumber ke dalam aplikasi, situs web sehingga membuka peluang untuk meningkatkan fungsionalitas dan informasi yang tersedia dalam proyek.
* **Query**, teknik ekstraksi data menggunakan query adalah pendekatan umum untuk mengambil data terstruktur dari basis data, baik basis data relasional dengan **SQL** atau basis data **NoSQL** dengan bahasa query yang sesuai. Tipe data yang dapat diambil bervariasi sesuai dengan struktur basis data yang sedang diakses, dan sumber data umumnya berupa basis data perusahaan, sistem manajemen basis data, atau penyimpanan data online yang mendukung penulisan query. Dengan menulis query yang tepat, data dapat diambil dalam format tabel atau sesuai dengan struktur yang ada dalam basis data tersebut. Teknik ini sangat berguna untuk analisis bisnis, pembuatan laporan, dan pengambilan keputusan, serta memungkinkan integrasi data internal perusahaan secara efisien.

1. Sebutkan 3 tipe transformation data, lalu jelaskan tujuan dari masing-masing tipe transformation data tersebut!



* **Data cleaning** adalah salah satu jenis transformasi data yang penting dalam pengolahan data. Tujuan dari data cleaning adalah untuk **membersihkan dan mempersiapkan data** agar dapat digunakan dengan baik dalam analisis atau aplikasi lainnya. Proses ini melibatkan **identifikasi** dan **koreksi kesalahan** atau **ketidaksesuaian dalam data**, seperti data yang **hilang**, **duplikat**, **salah ketik**, atau **nilai yang tidak masuk akal**. Data yang bersih dan terstruktur membantu meminimalkan kesalahan dalam analisis, meningkatkan akurasi model prediksi, dan memungkinkan pengambilan **keputusan yang lebih baik**. Selain itu, data yang telah dibersihkan juga **meningkatkan efisiensi dan keandalan** dalam proses pengolahan data selanjutnya, seperti pemodelan statistik atau pembuatan laporan. Dengan demikian, tujuan utama dari data cleaning adalah untuk **menghasilkan data yang lebih berkualitas dan dapat diandalkan** untuk analisis dan pengambilan keputusan yang lebih baik.

1. Sebutkan 3 metode loading data, jelaskan karakteristik dari masing-masing metode tersebut dan kapan metode tersebut perlu digunakan!
   1. Bulk Loading
   2. Full Refresh
   3. Incremental Loading
   4. Streaming
   5. Change Data Capture (CDC)