DATA ENGINEERING

140+

DATA ENGINEERING 101

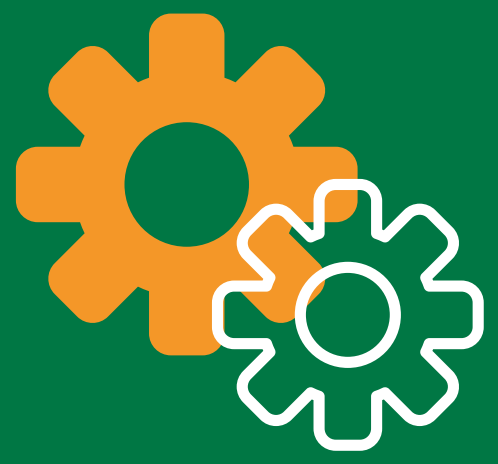
ETL - TERMINOLOGY

Everything you need to begin the journey

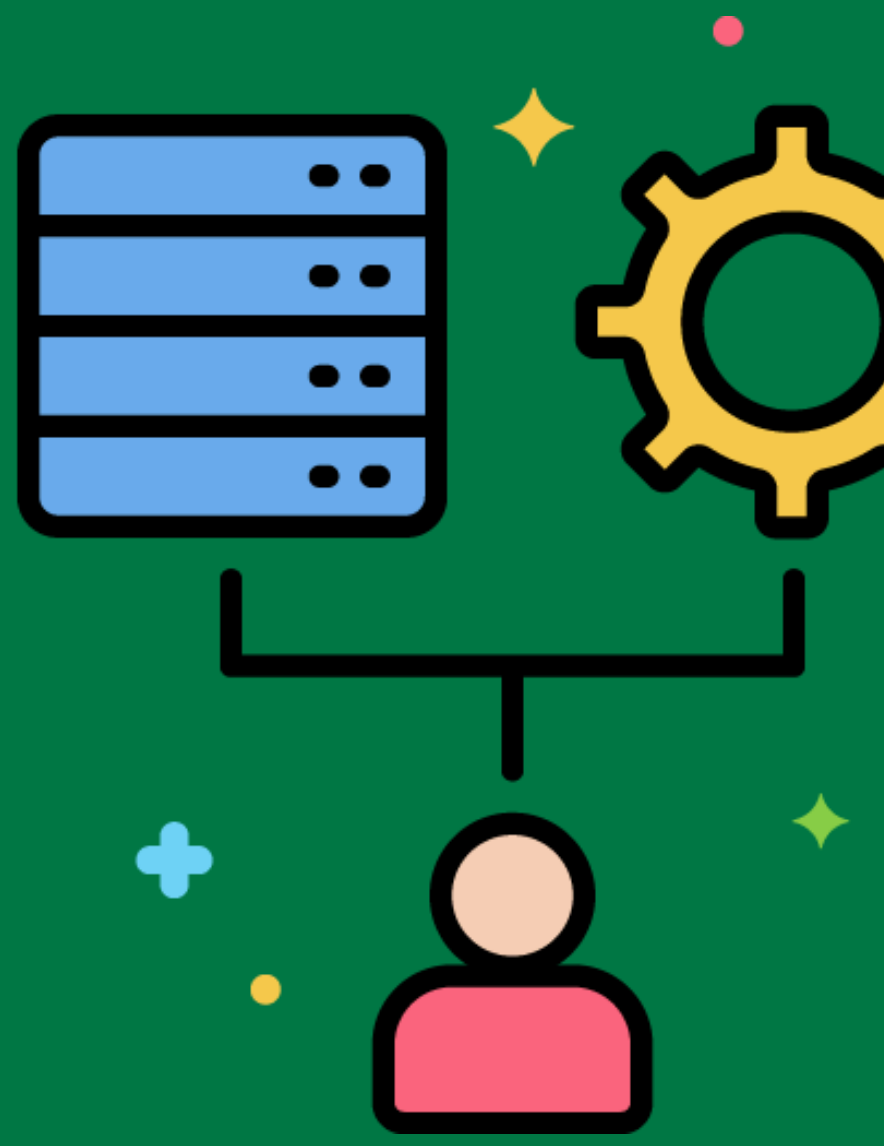
Swipe Left

Data Engineering 101 - ETL Terminology

ETL System



Extract, Transform, Load (ETL) system is responsible for extracting data from source systems, transforming it to fit business needs, and loading it into a data warehouse.

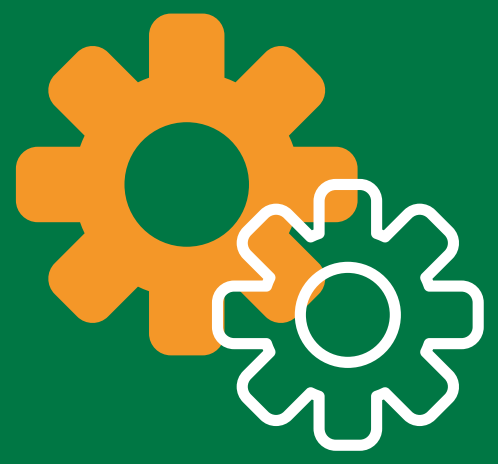


A retail company extracts sales data from its point-of-sale system, transforms it to analyze seasonal trends, and loads it into a data warehouse.

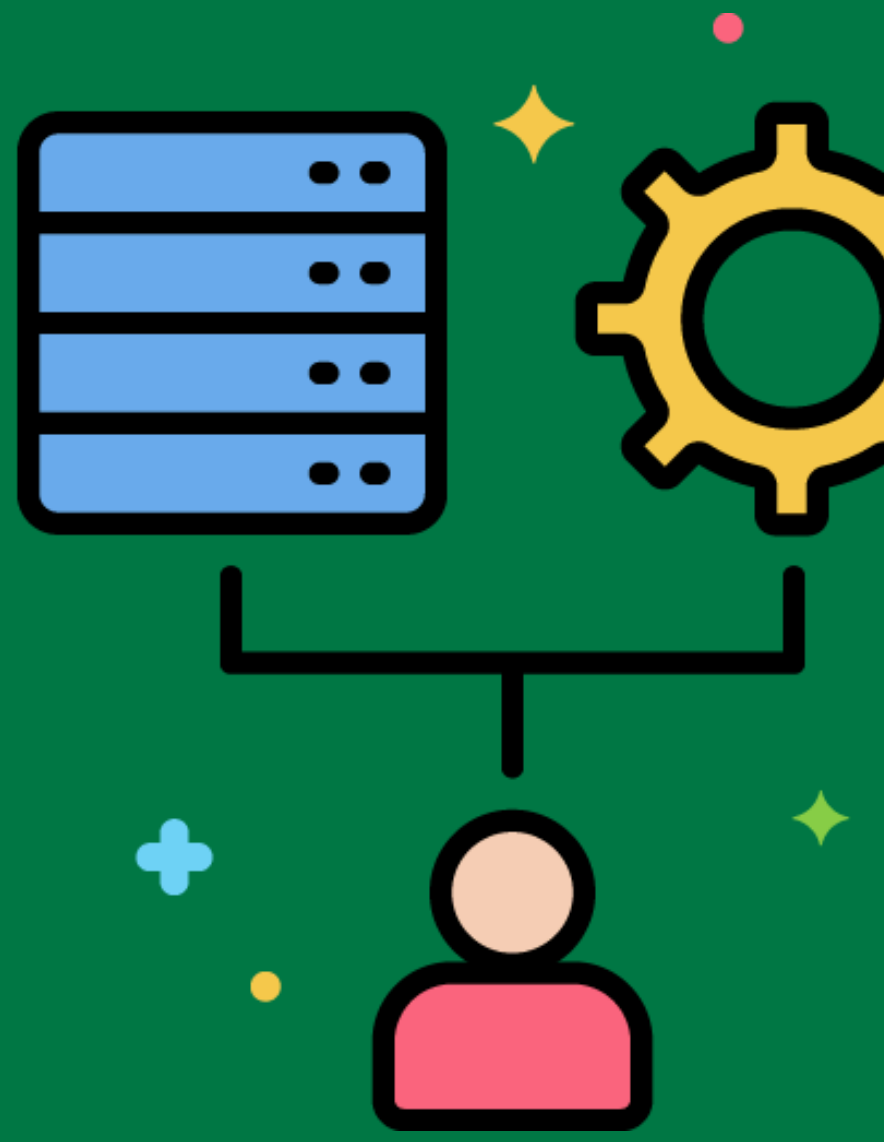
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Warehouse



A central repository of integrated data from multiple sources, designed to support decision making and business intelligence activities.

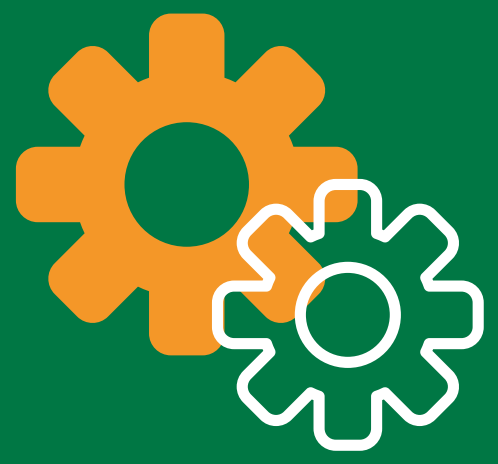


A healthcare organization uses a data warehouse to integrate patient data from various departments for comprehensive reporting and analysis.

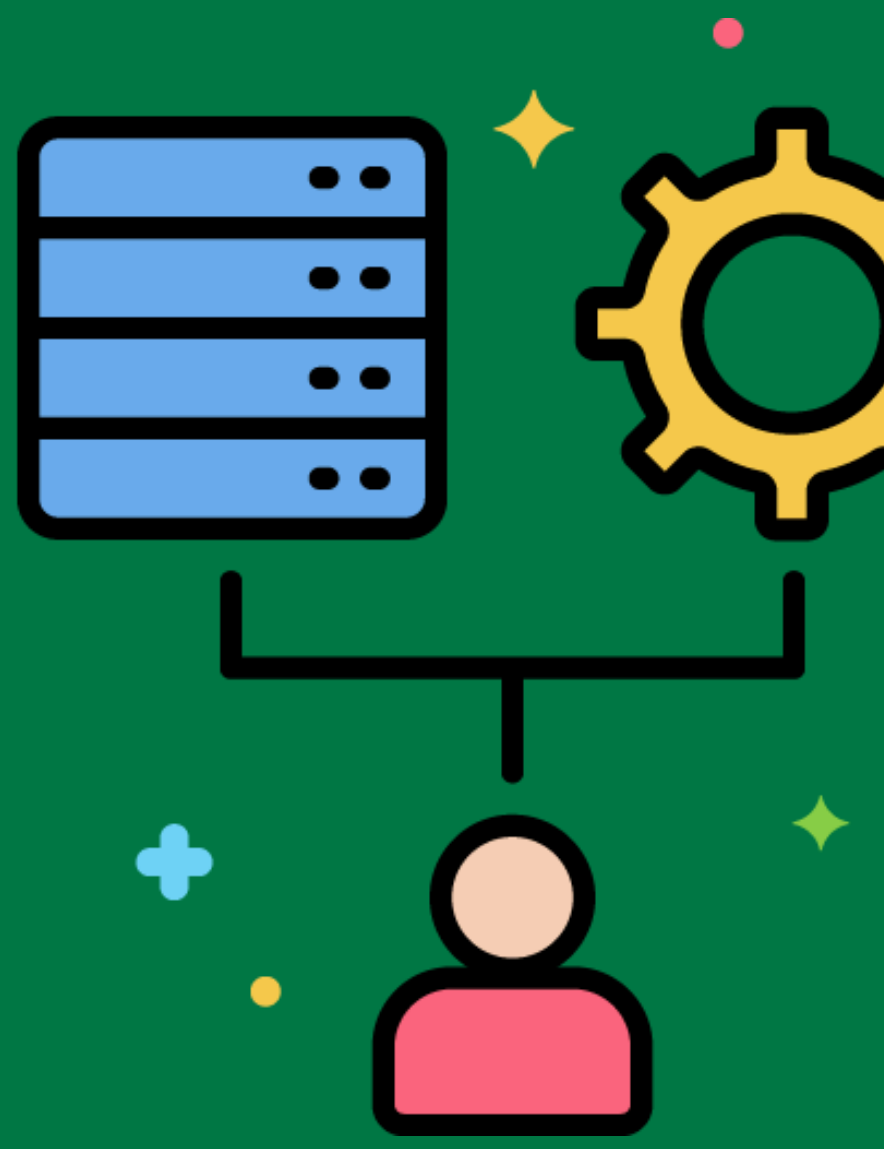
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Extract



The process of retrieving data from different source systems.

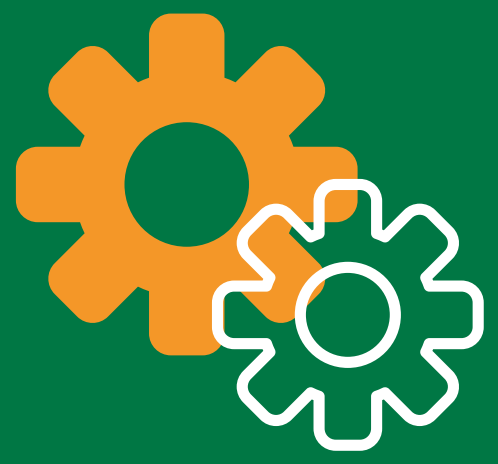


Extracting customer data from a CRM system and sales data from an ERP system.

Shwetank Singh
GritSetGrow - GSGLearn.com

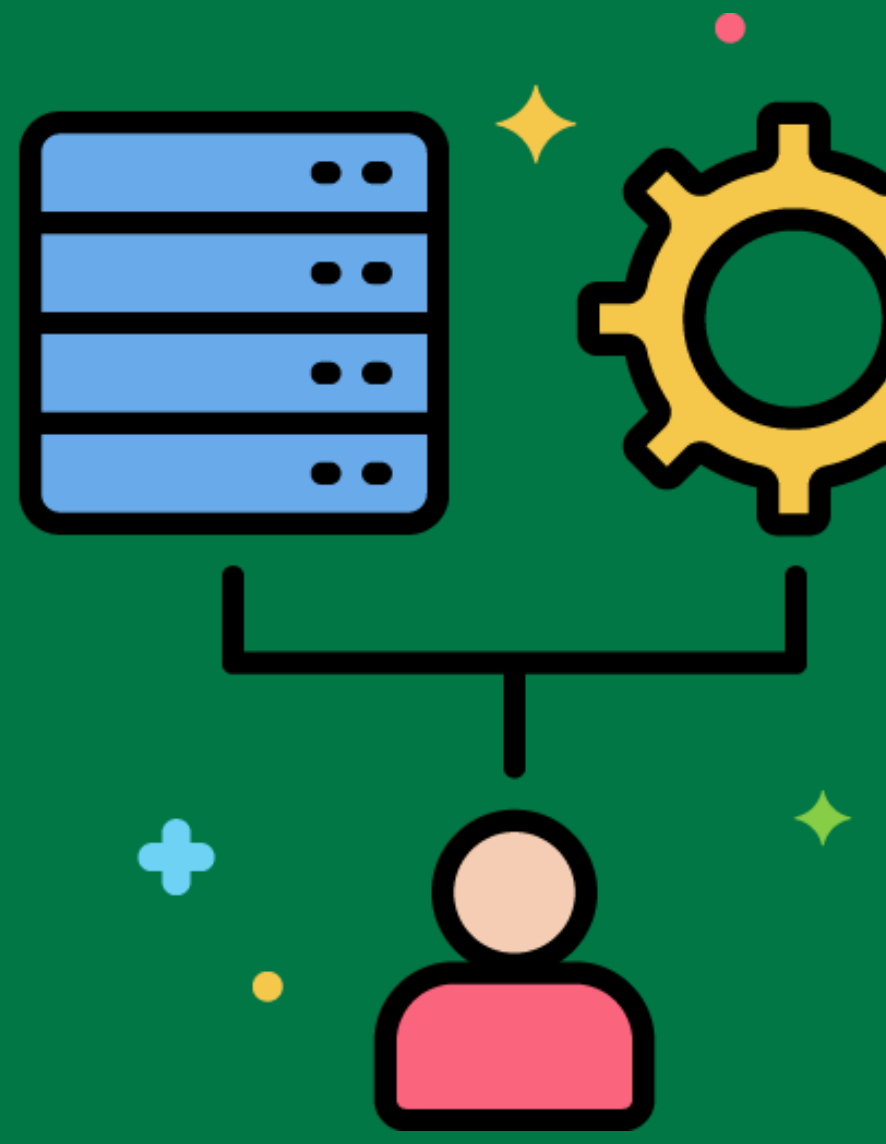
Data Engineering 101 - ETL Terminology

Transform



The process of converting extracted data into a format that can be loaded into the data warehouse, including cleaning, conforming, and integrating data from multiple sources.

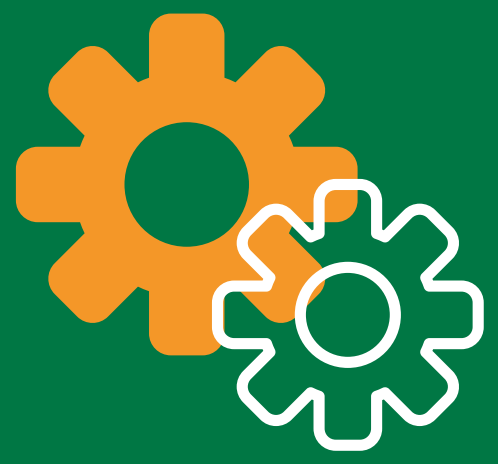
Standardizing date formats and removing duplicates from customer records before loading them into the data warehouse.



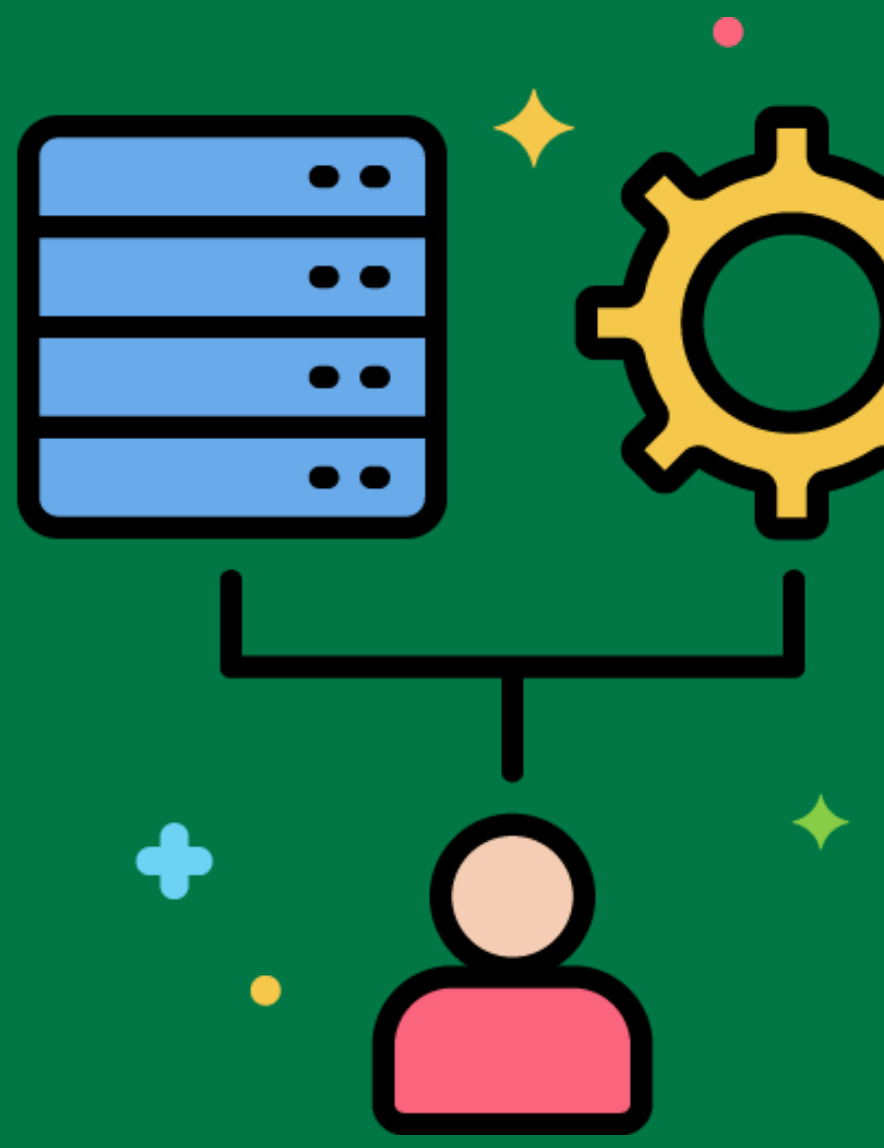
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Load



The process of loading transformed data into the target data warehouse.

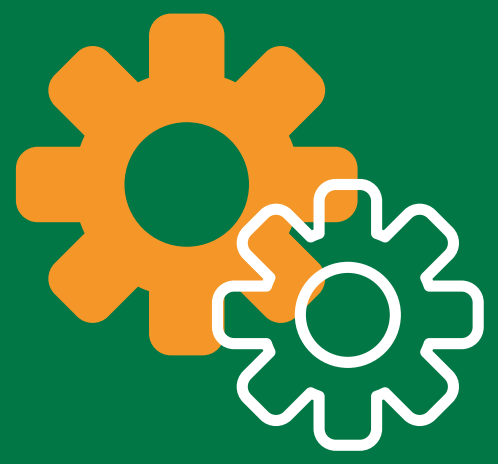


Loading cleaned and standardized customer data into the customer dimension table in the data warehouse.

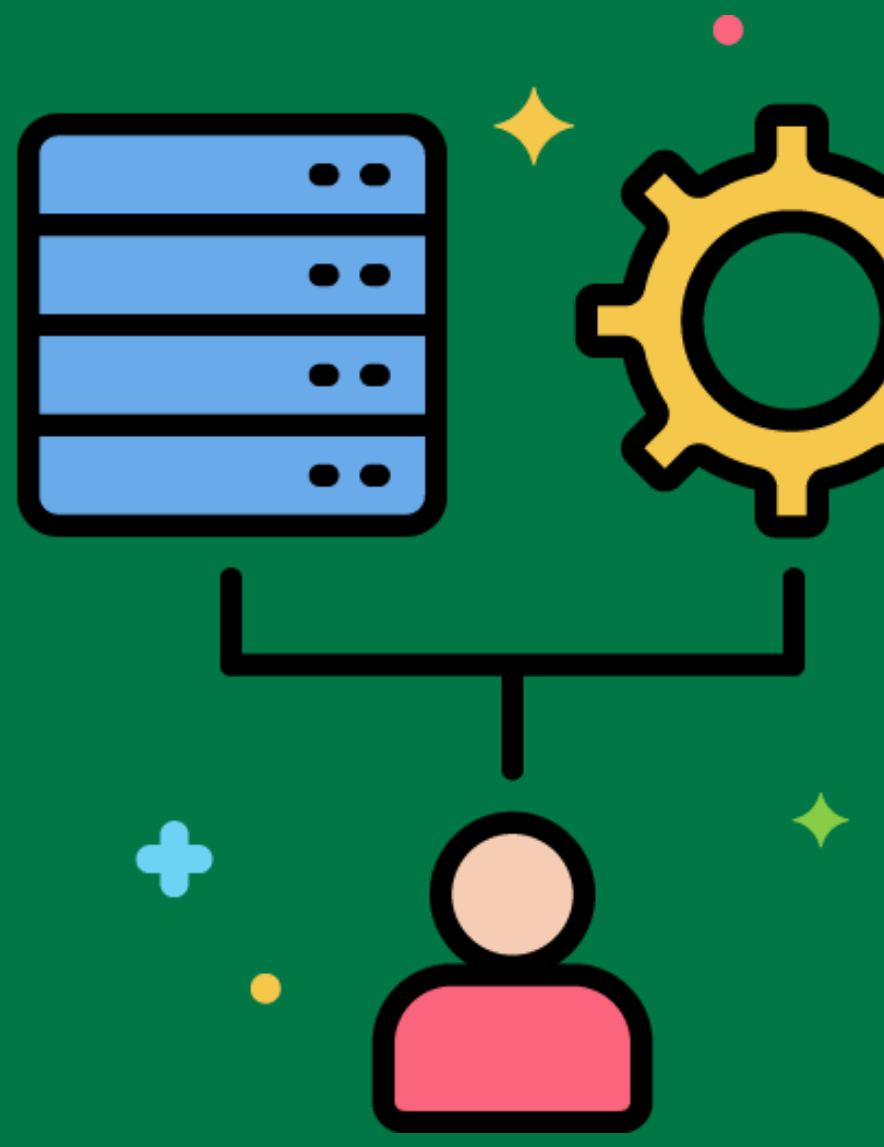
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Staging Area



An intermediate storage area used for data processing during the ETL process.

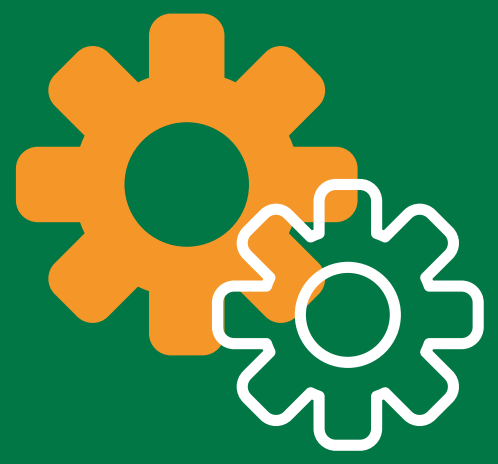


Temporarily storing raw sales data extracted from an ERP system before transformation.

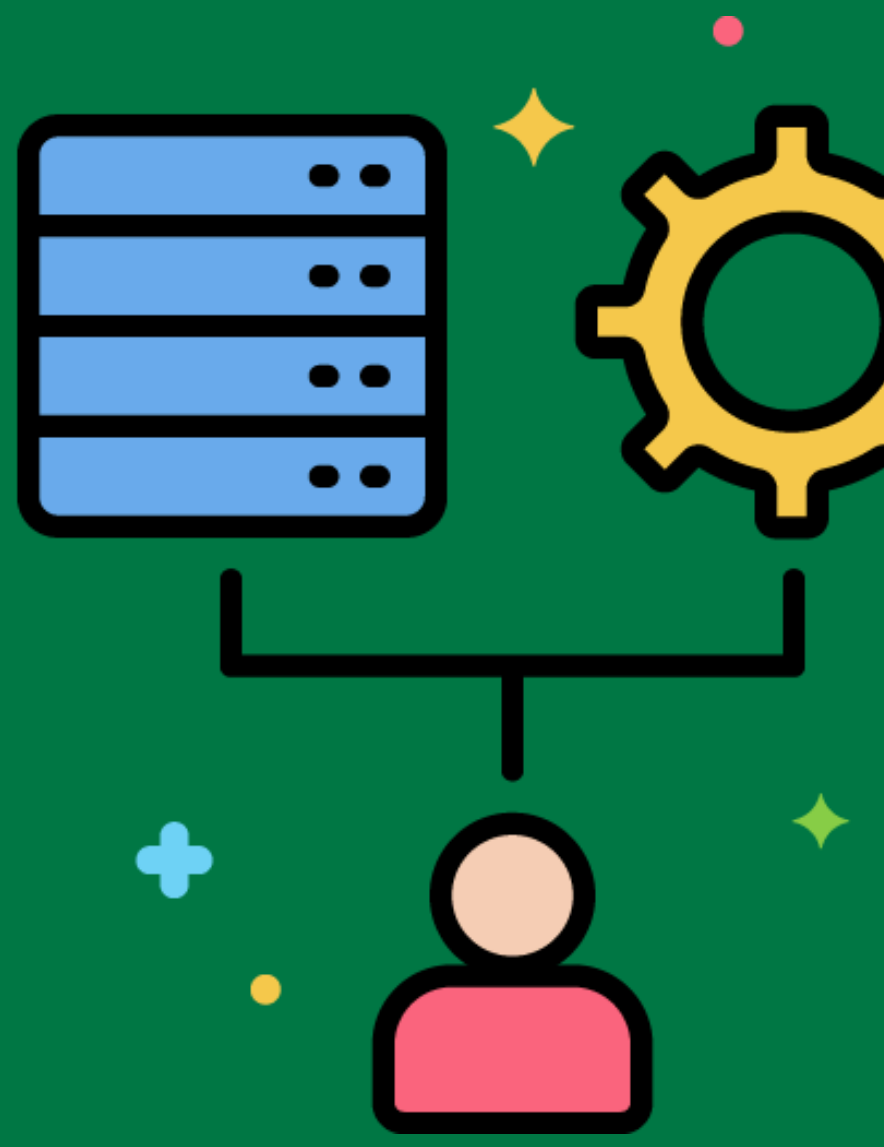
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Profiling



Analyzing source data to assess its quality, completeness, and fitness for purpose.

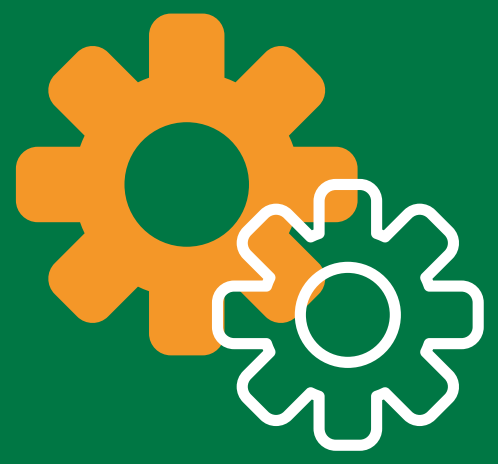


Using a data profiling tool to find missing values, outliers, and inconsistencies in the customer data from the CRM system.

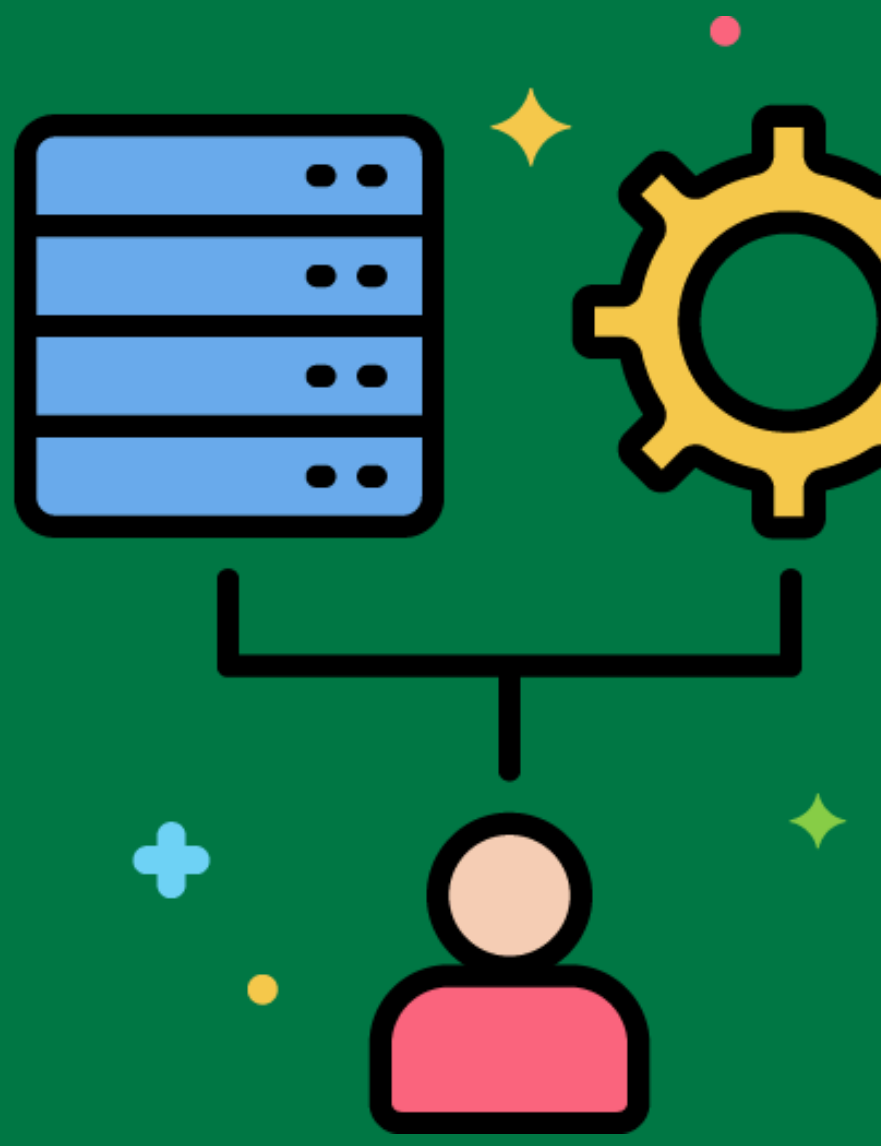
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Cleaning



The process of identifying and fixing errors and omissions in the data.

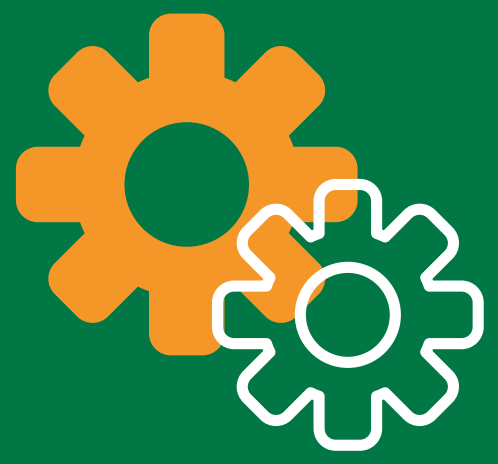


Standardizing addresses in customer data by correcting misspellings and formatting issues.

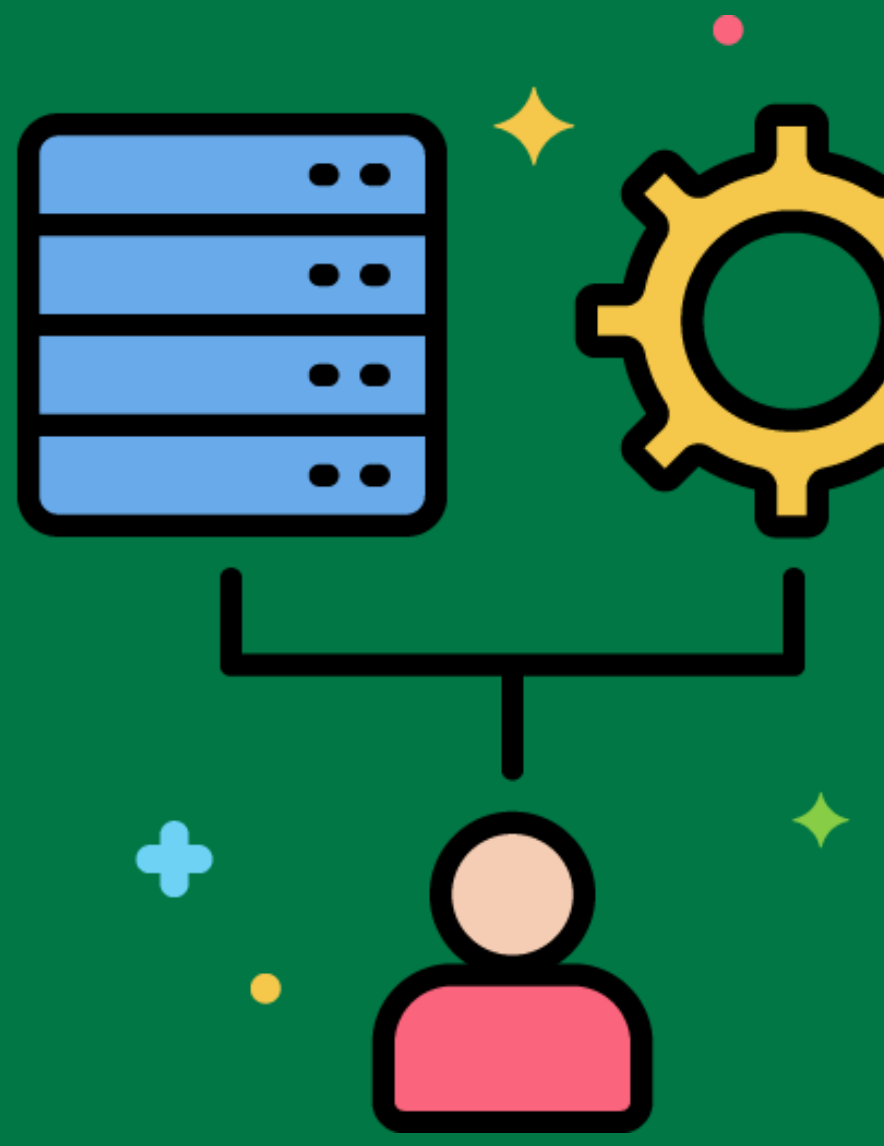
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Conforming



Resolving labeling conflicts between potentially incompatible data sources so that they can be used together in the data warehouse.

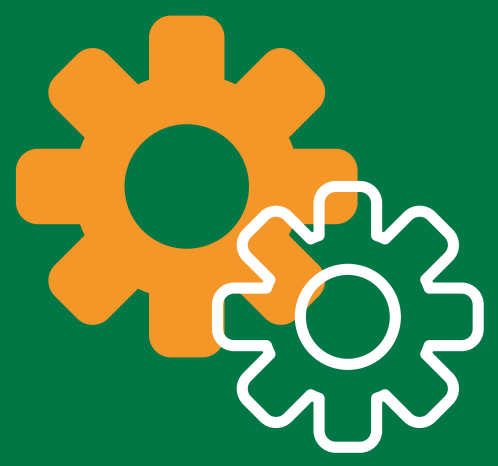


Conforming product category names from different source systems to a standard set of categories used in the data warehouse.

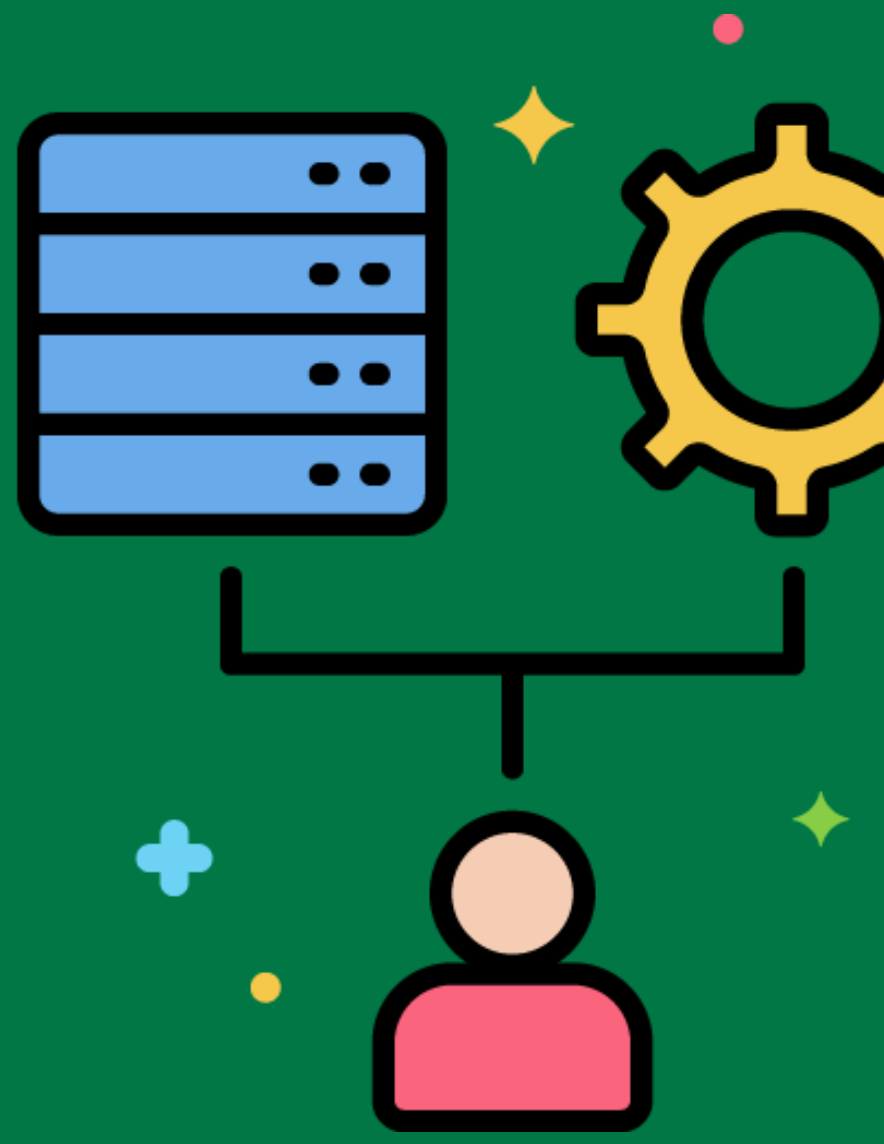
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Change Data Capture



Techniques for capturing changes in the source data to keep the data warehouse up-to-date.

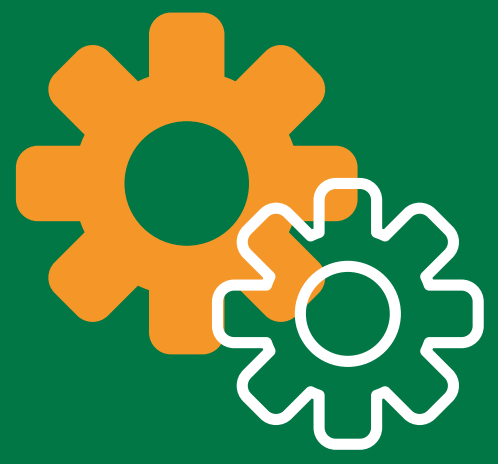


Implementing triggers in the CRM system to capture inserts, updates, and deletes and storing these changes in a staging table.

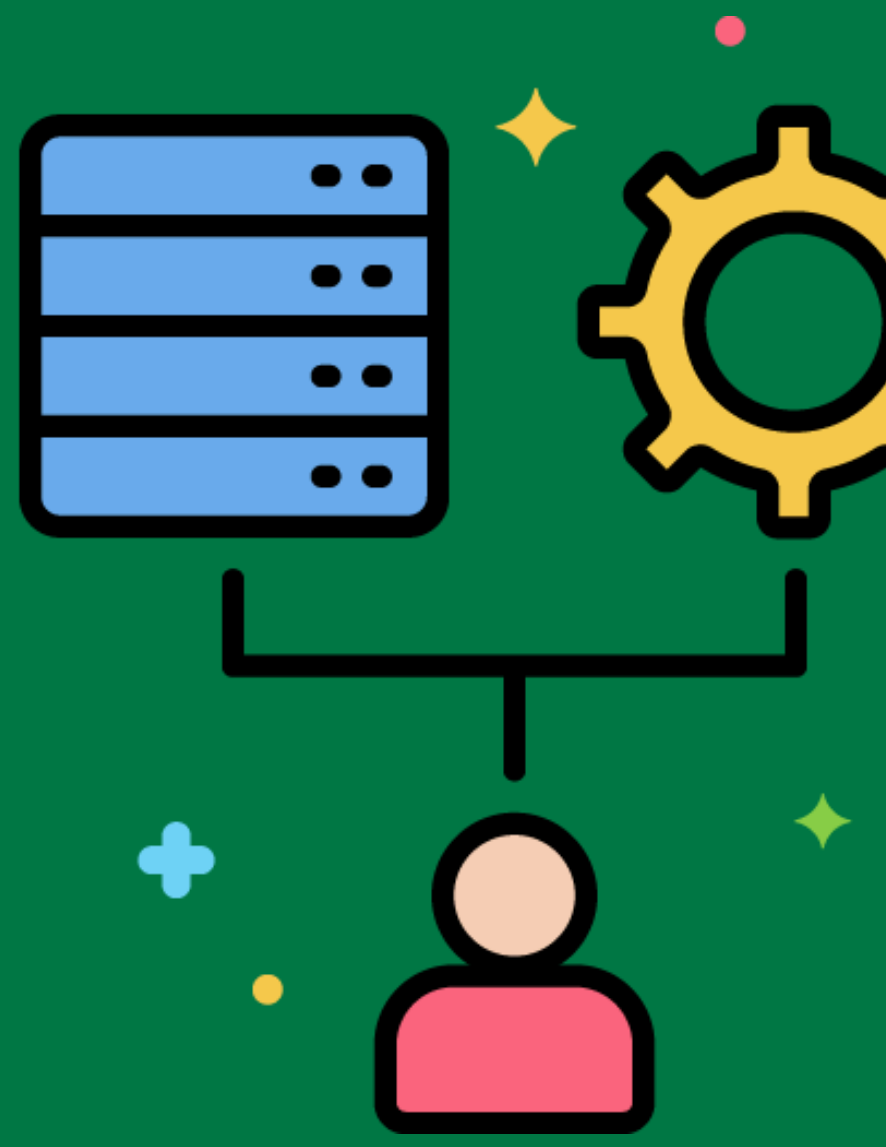
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Slowly Changing Dimensions (SCD)



Handling changes in dimension data in a way that preserves historical data.

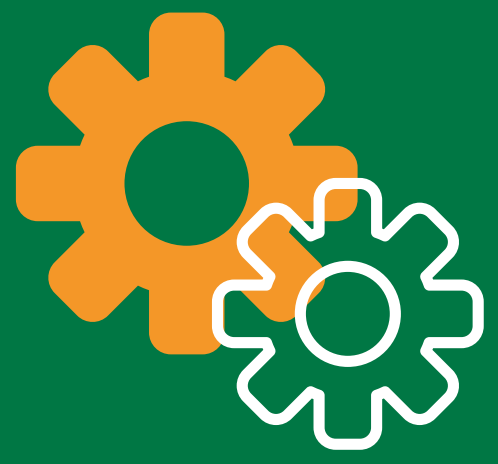


Implementing SCD Type 2 for customer addresses, where a new record is created for each change, preserving the history of address changes.

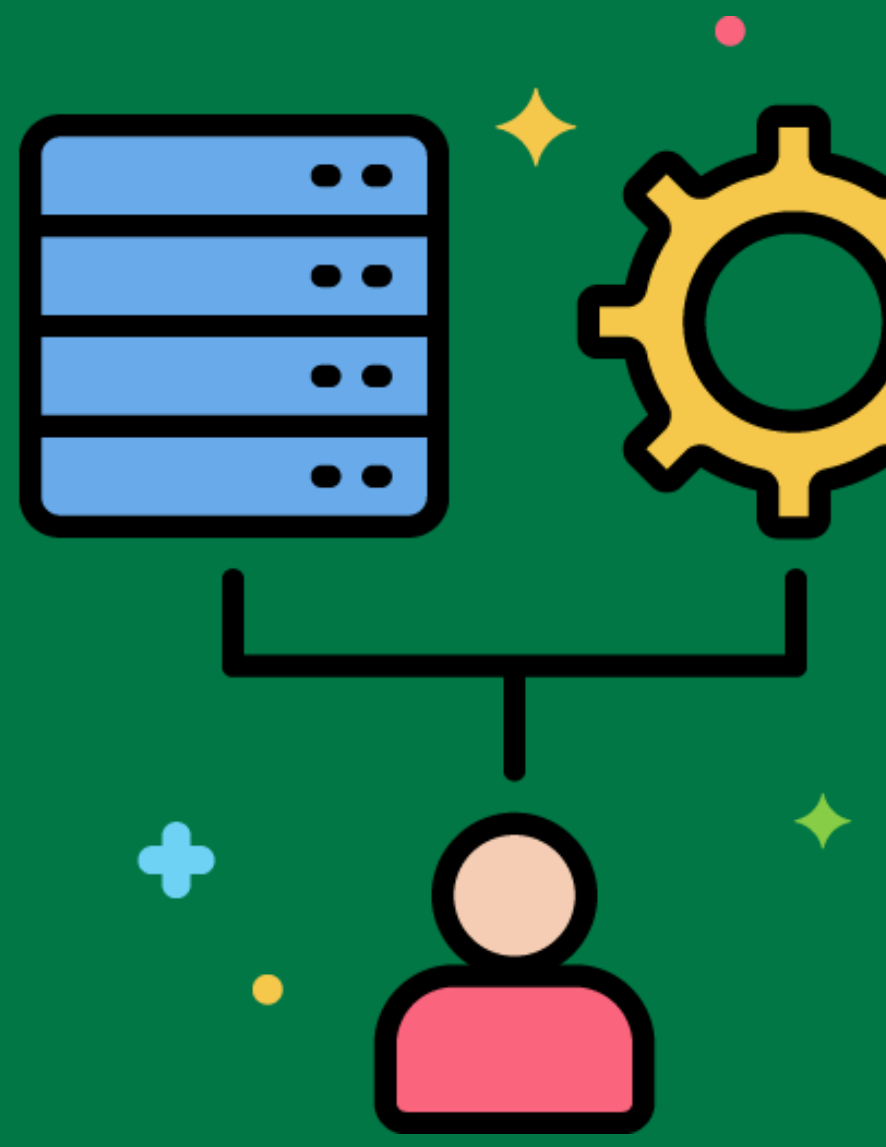
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Surrogate Key Assignment



Assigning unique keys to records in the dimension tables to avoid using business keys from the source systems.

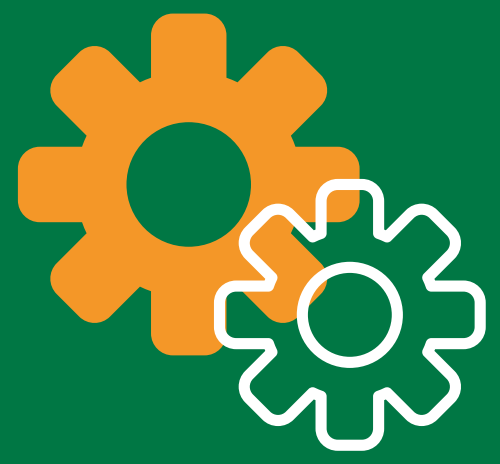


Generating surrogate keys for customer records in the data warehouse to replace the customer IDs from the CRM system.

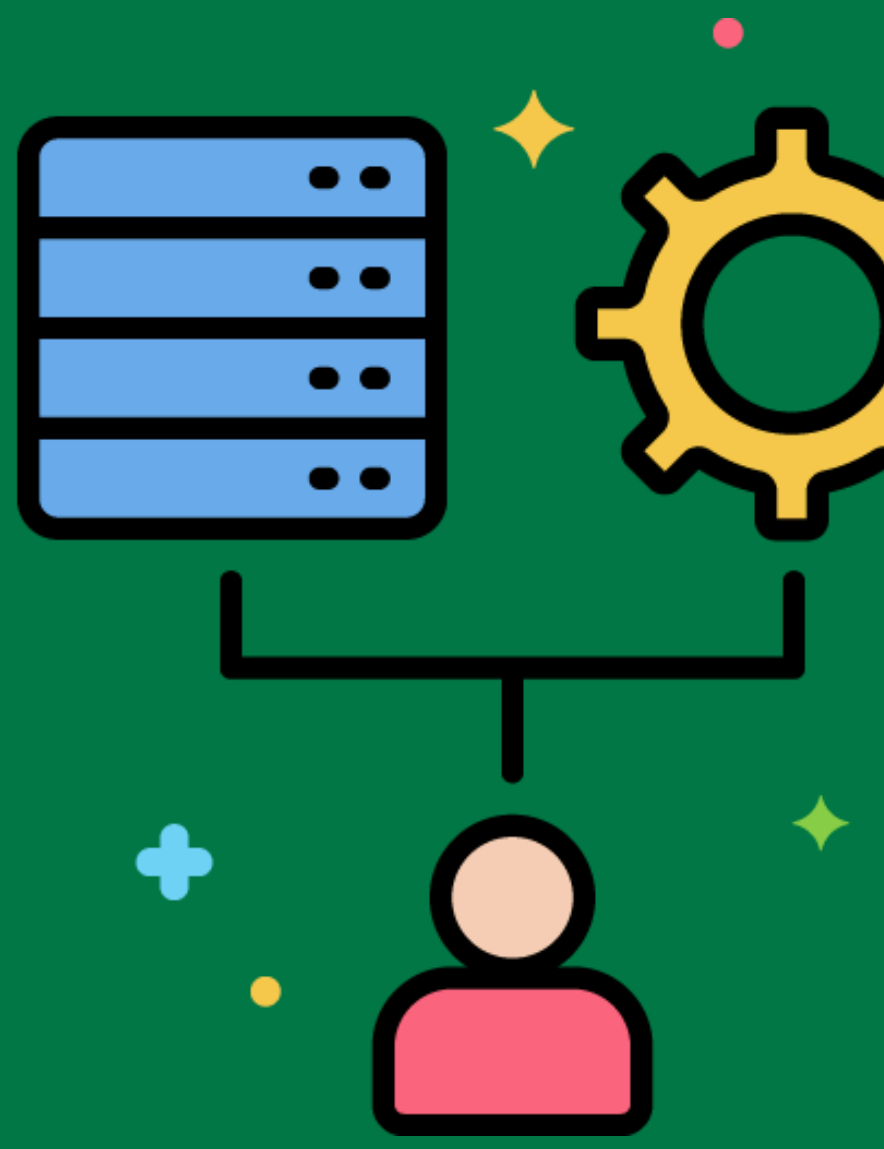
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Fact Table Loading



The process of loading transactional data into fact tables, often involving complex transformations and calculations.

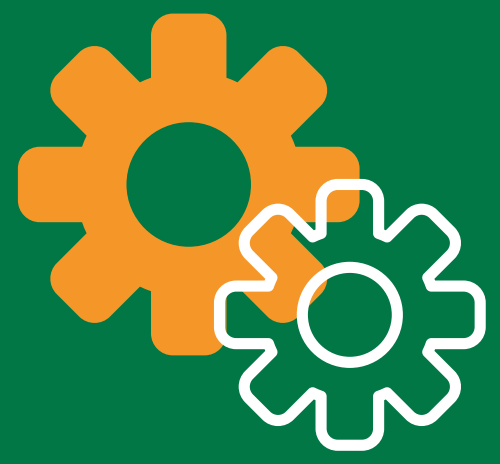


Loading sales transactions into the sales fact table, including calculations for total sales amount and discount applied.

Shwetank Singh
GritSetGrow - GSGLearn.com

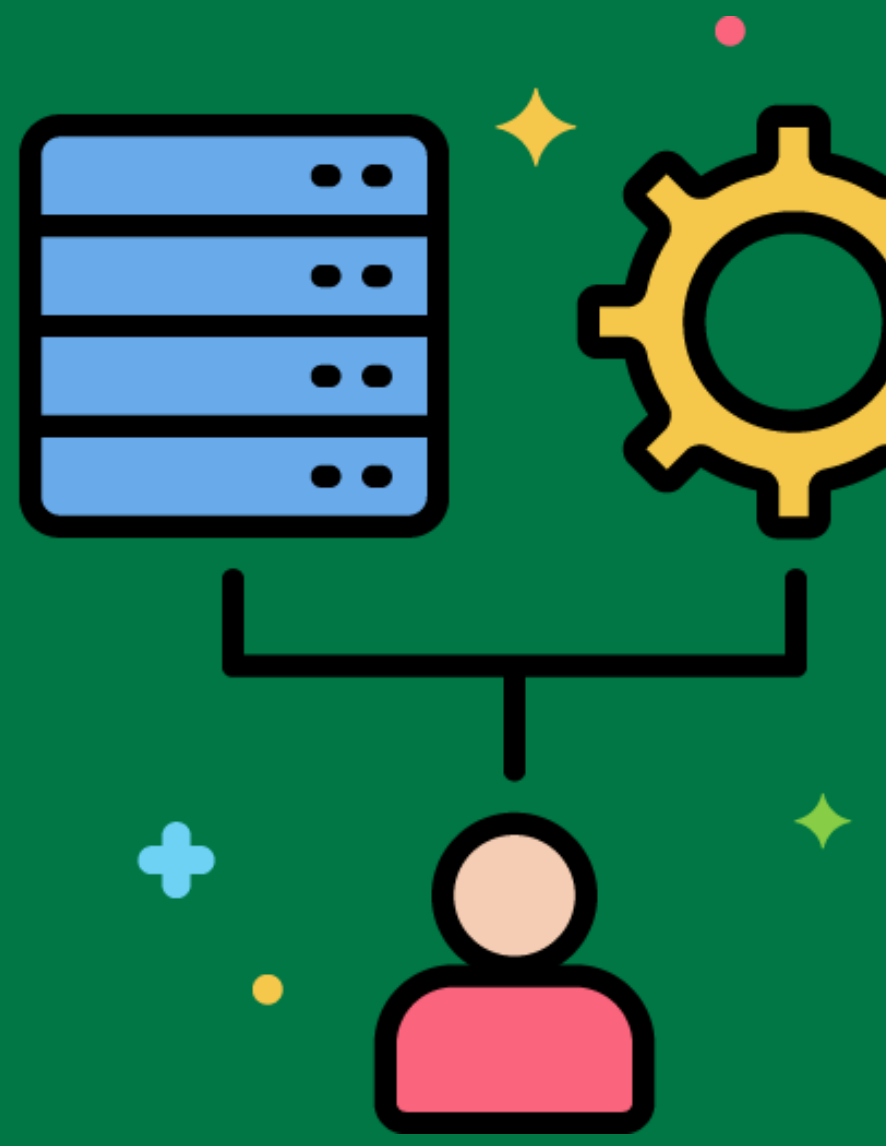
Data Engineering 101 - ETL Terminology

ETL Process Automation



Automating the ETL process to run at scheduled intervals or in response to specific events.

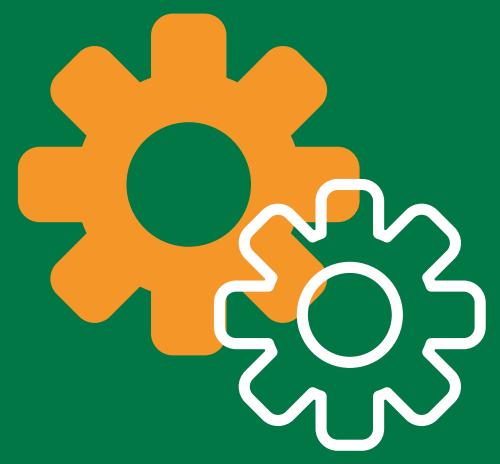
Setting up a workflow in an ETL tool to automatically extract, transform, and load sales data from the source systems to the data warehouse every night.



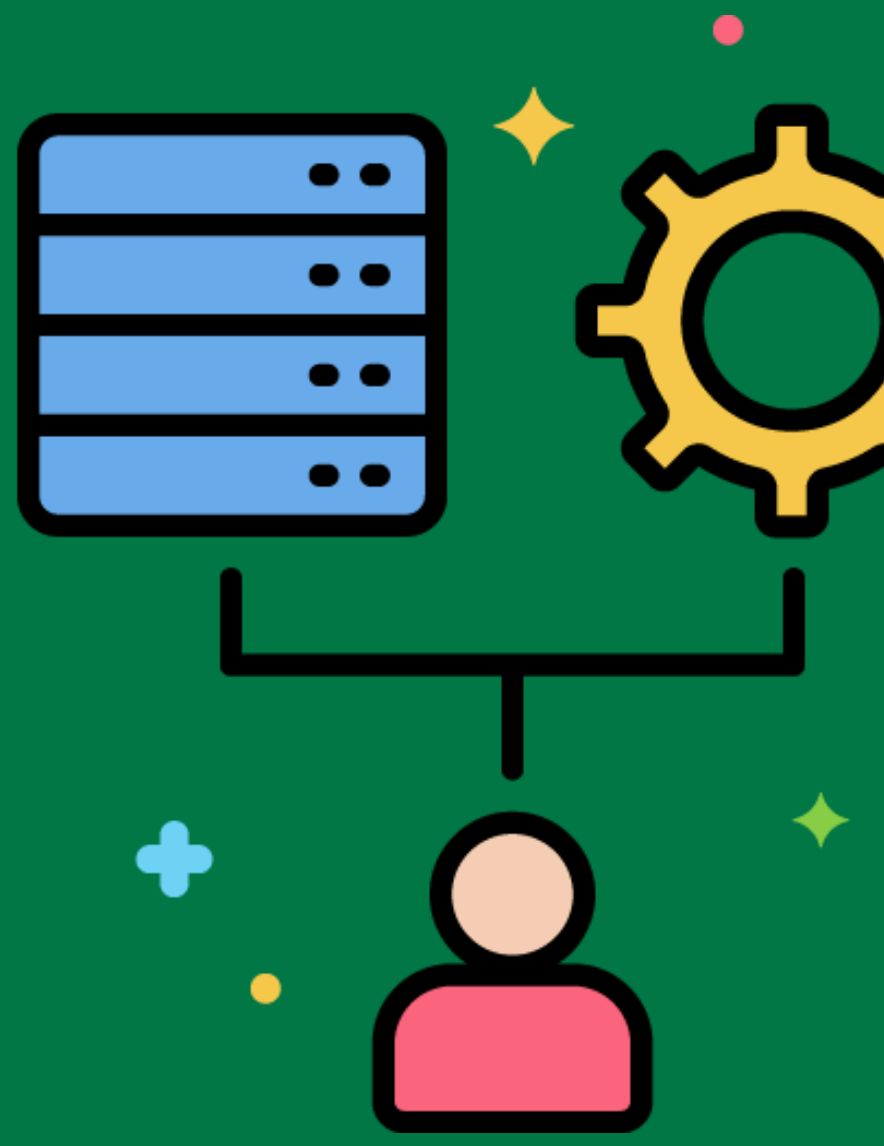
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Logical Data Map



A document that ties the beginning of the ETL system to the end, describing the relationship between source fields and destination fields.

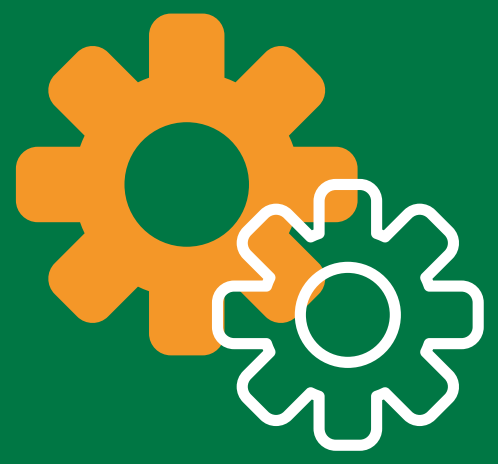


A logical data map for a sales data warehouse showing how sales records from different source systems map to the unified sales table in the data warehouse.

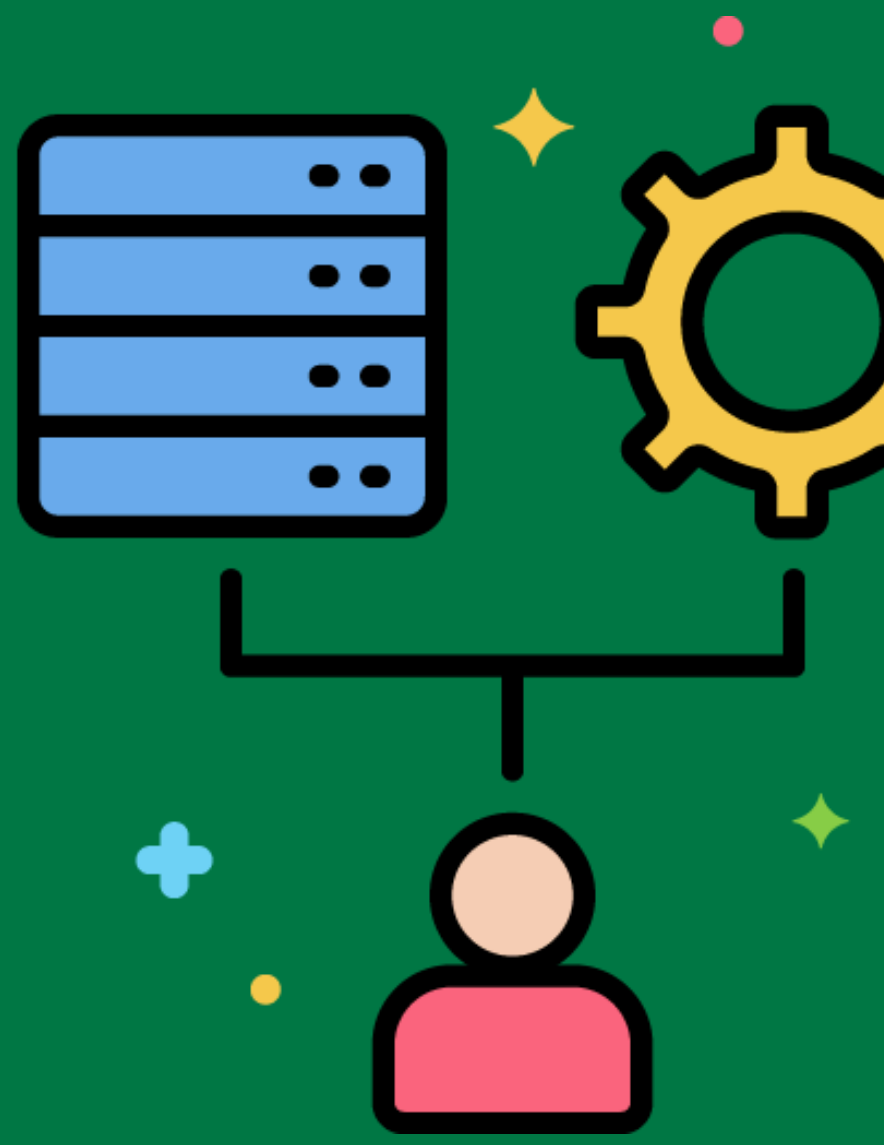
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Discovery



The phase where source systems are identified and analyzed to determine the required source data for the data warehouse.

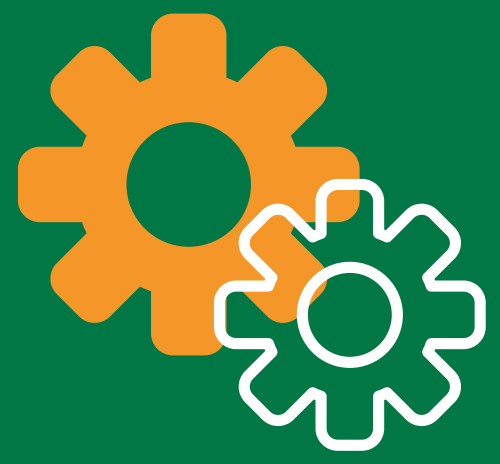


Identifying the CRM system as the primary source for customer data and analyzing its data structure.

Shwetank Singh
GritSetGrow - GSGLearn.com

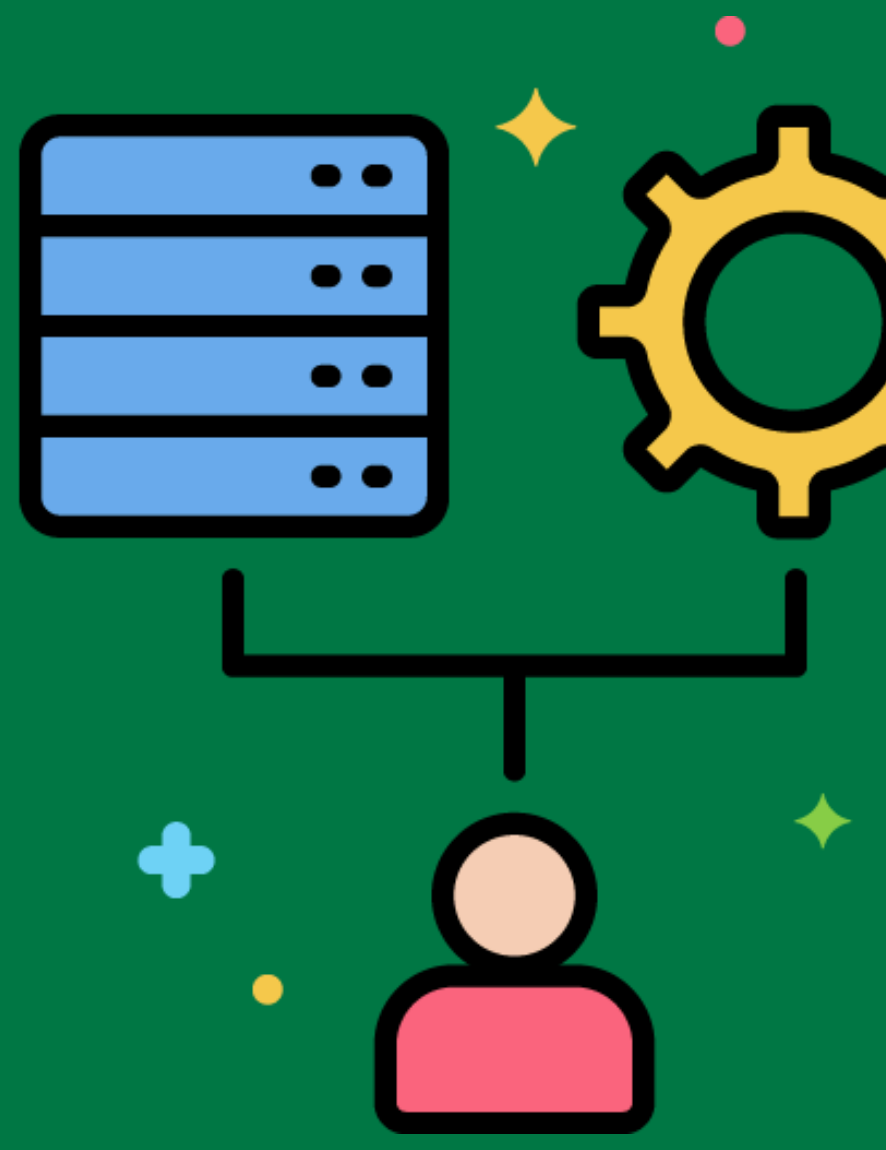
Data Engineering 101 - ETL Terminology

Data Integration



Combining data from different sources to provide a unified view.

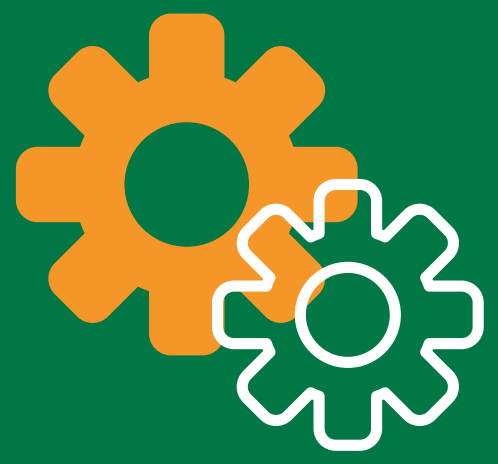
Integrating customer data from the CRM system with sales data from the ERP system to analyze customer buying patterns.



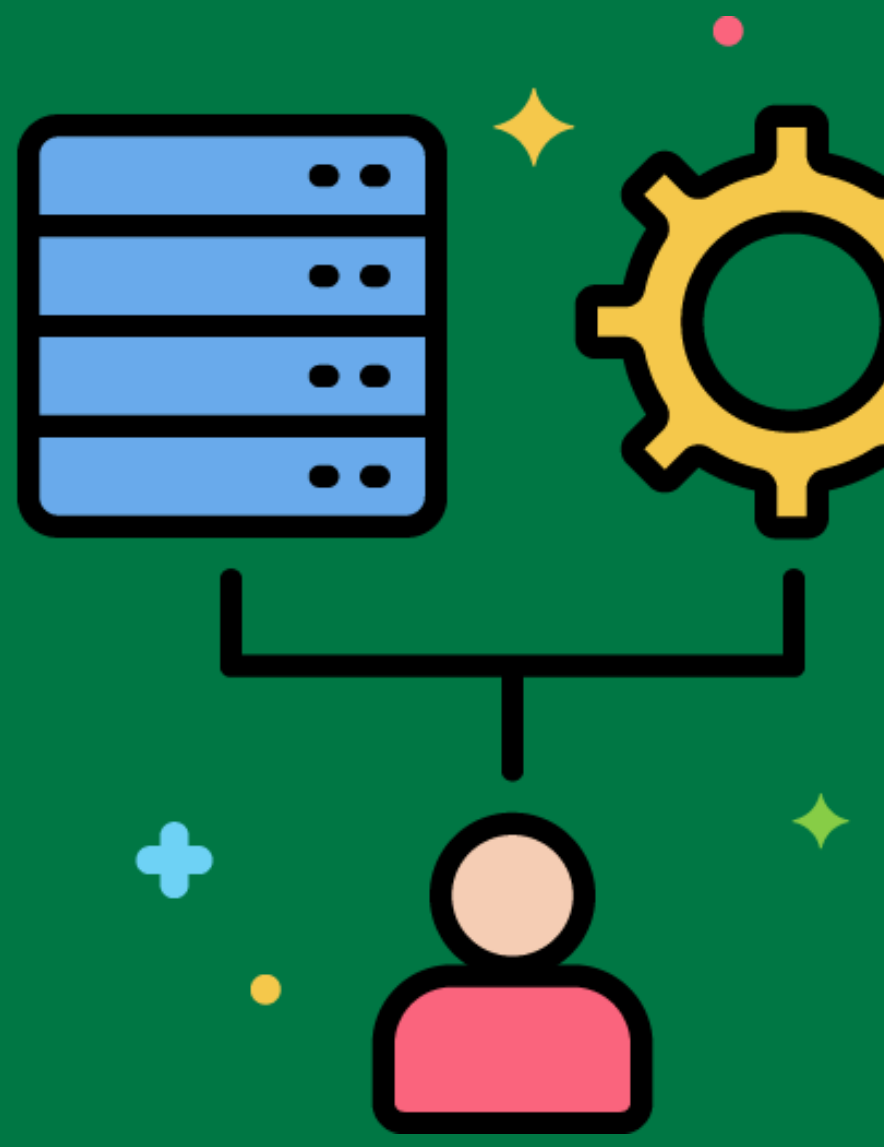
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Metadata Management



Managing data about data, providing information about the source, transformation, and destination of data.

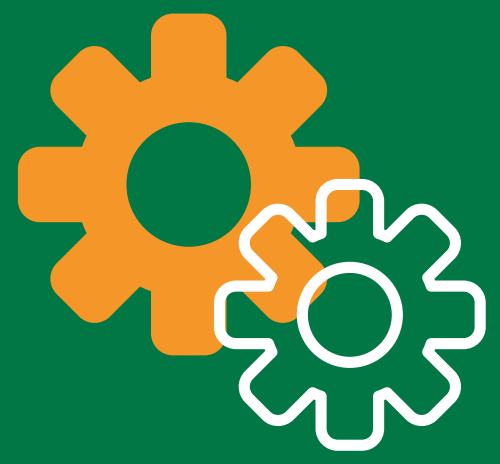


Documenting the source, transformation rules, and target tables for sales data in the data warehouse.

Shwetank Singh
GritSetGrow - GSGLearn.com

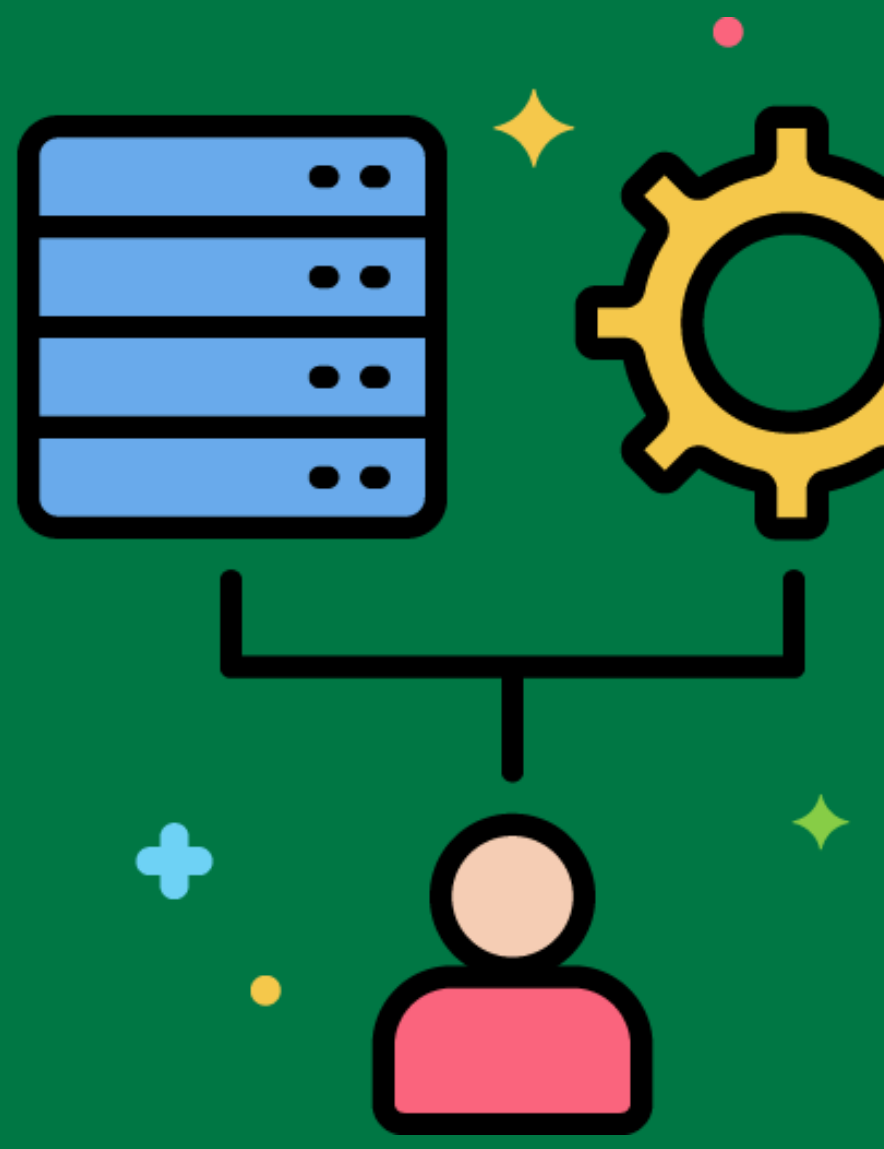
Data Engineering 101 - ETL Terminology

Data Lineage



Tracking the flow of data from source to destination, showing how data has been transformed.

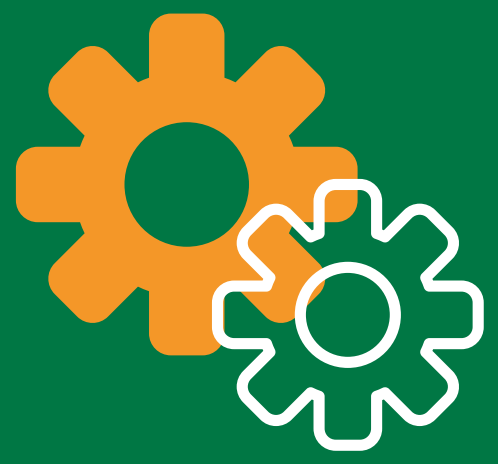
Tracing a sales figure in a report back to the original transaction in the point-of-sale system.



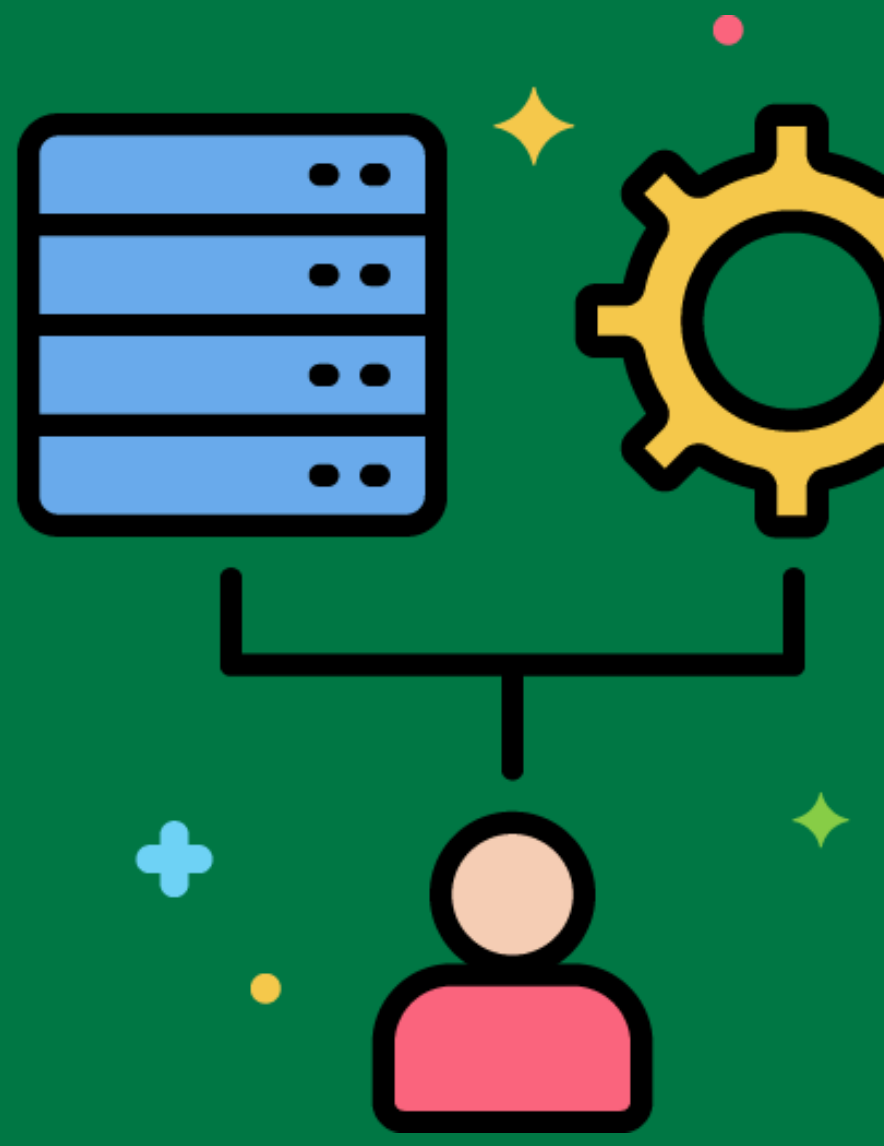
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Quality



Ensuring that the data in the data warehouse is accurate, complete, and reliable.

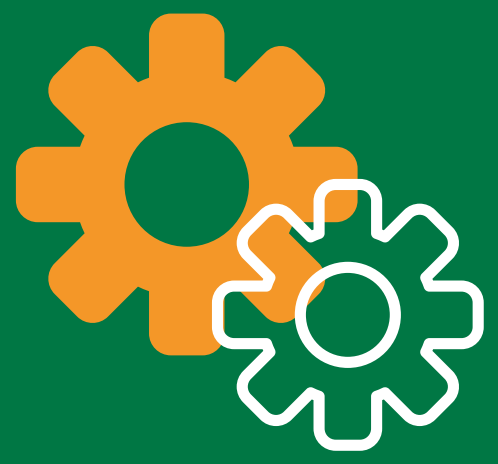


Implementing data validation rules to check for missing values and incorrect data types in customer records.

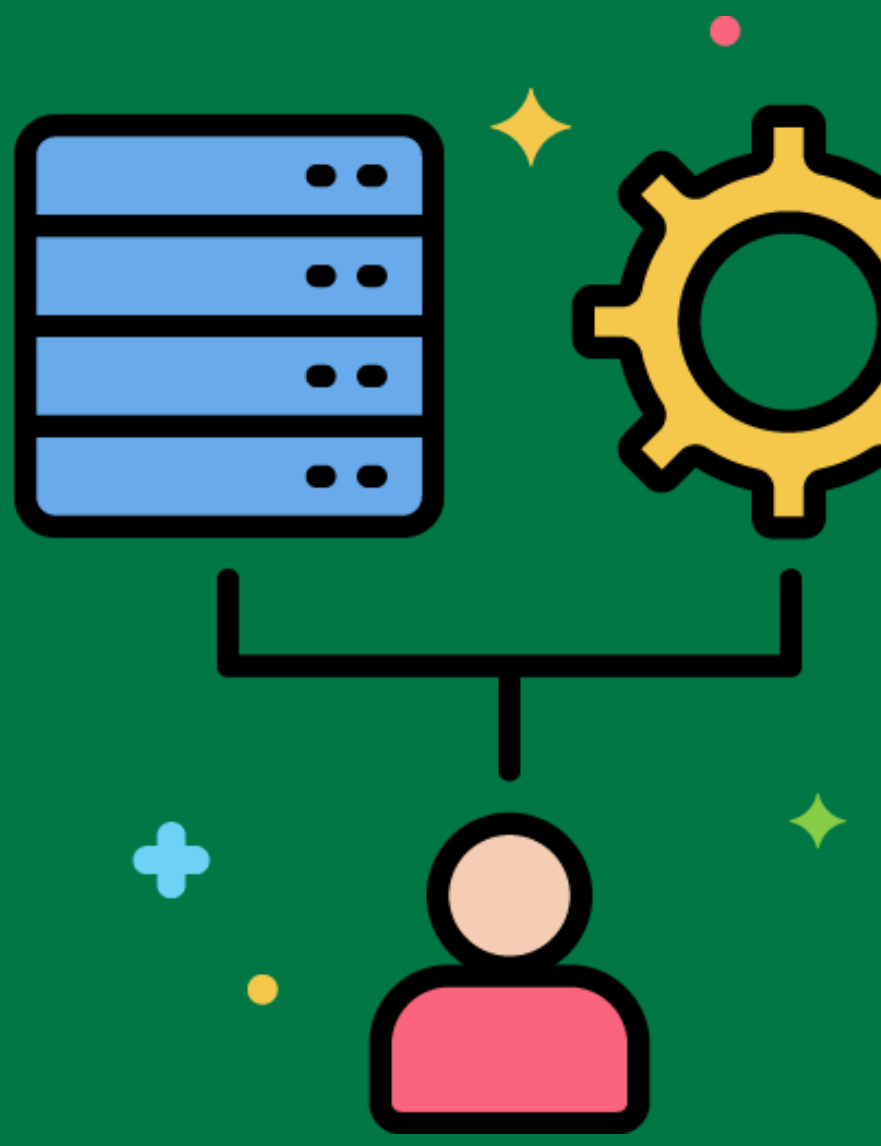
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Business Rules



Specific conditions and policies that guide how data should be processed and transformed.

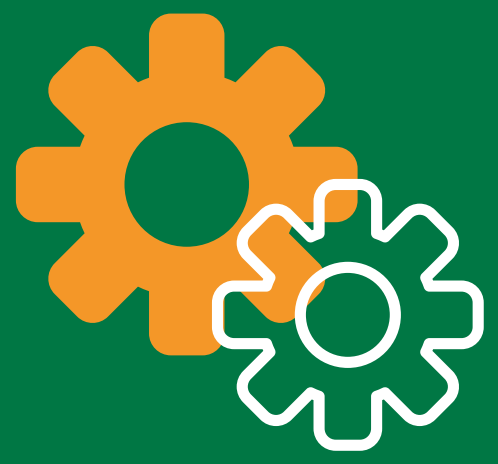


Applying a business rule to calculate discounts based on customer loyalty status during the ETL process.

Shwetank Singh
GritSetGrow - GSGLearn.com

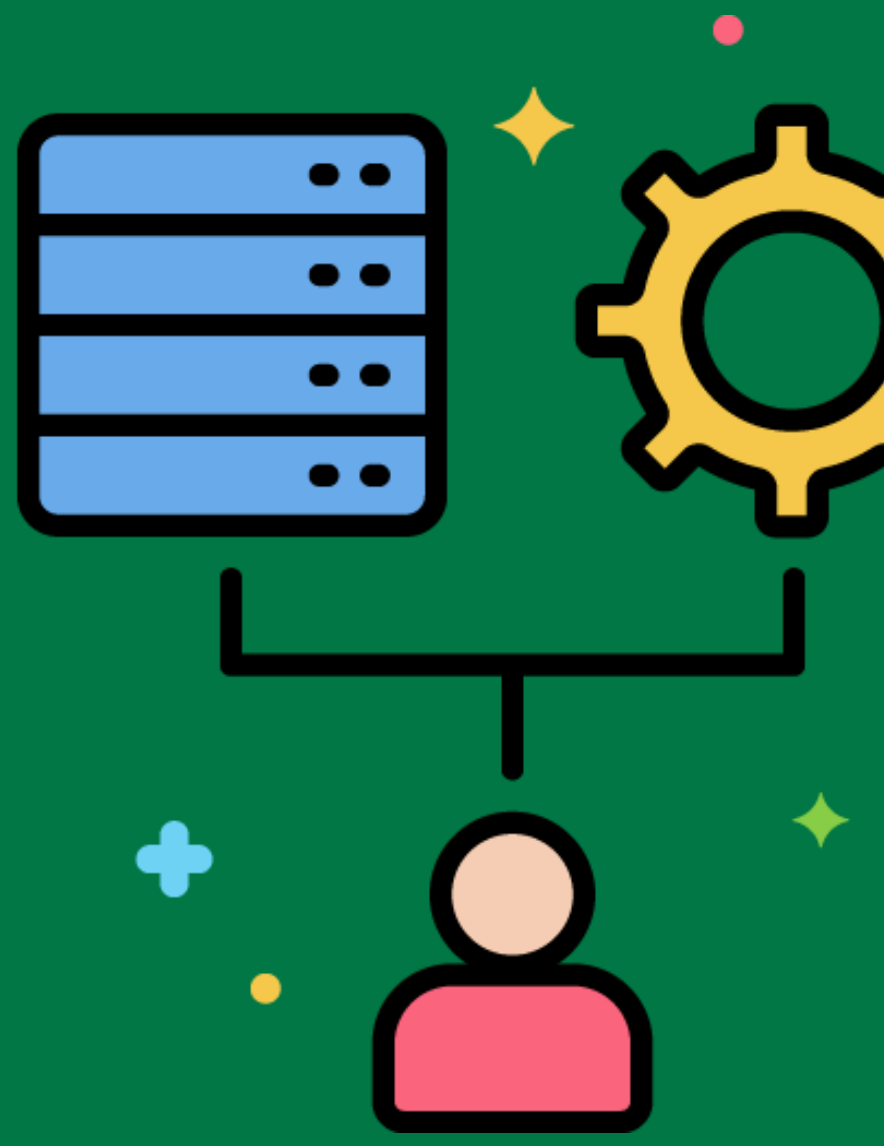
Data Engineering 101 - ETL Terminology

Error Handling



Identifying, logging, and managing errors that occur during the ETL process.

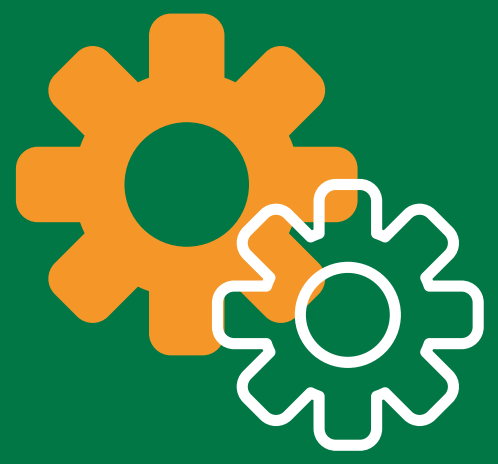
Logging and handling errors when there are missing values in the customer data during the extraction process.



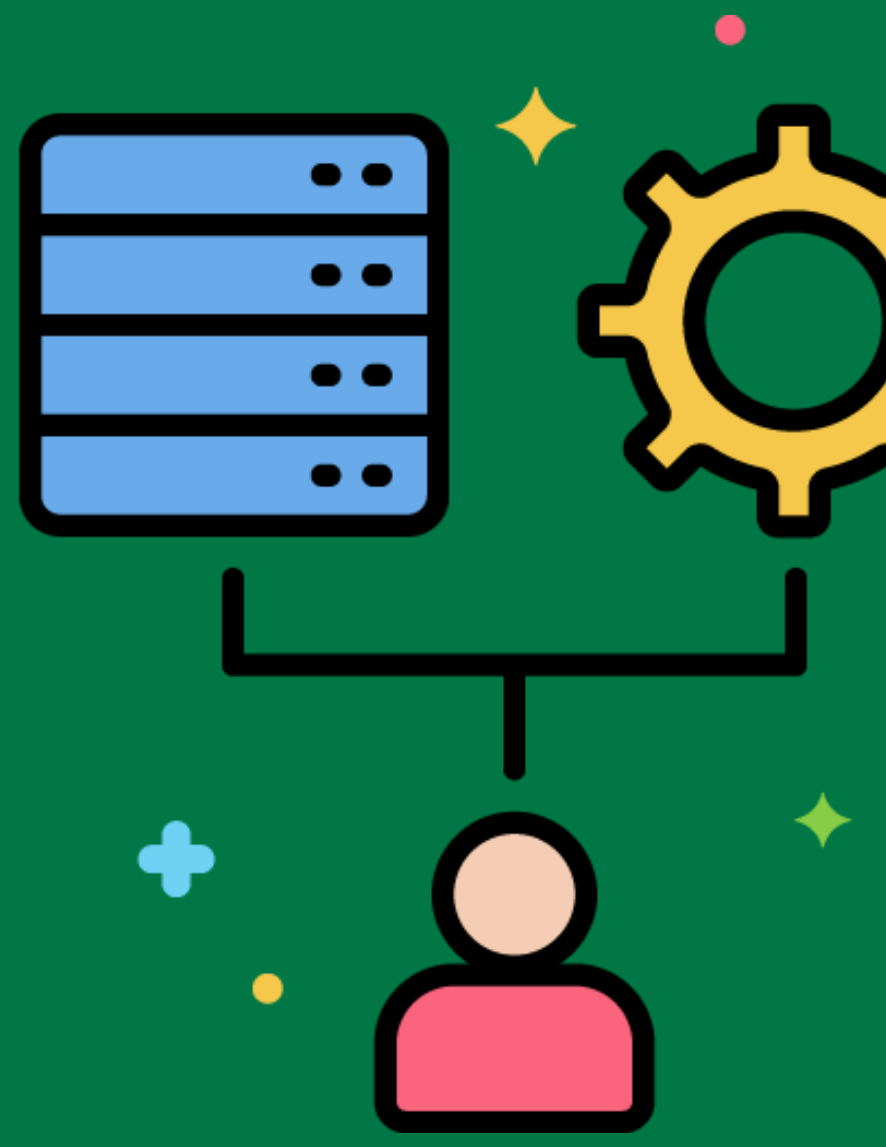
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Incremental Load



Loading only the data that has changed since the last ETL run.

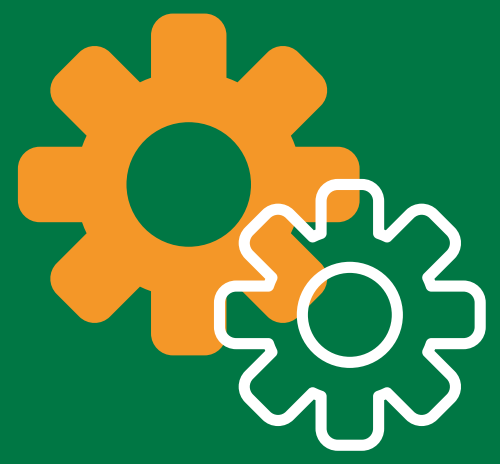


Loading only new and updated sales transactions from the ERP system into the data warehouse.

Shwetank Singh
GritSetGrow - GSGLearn.com

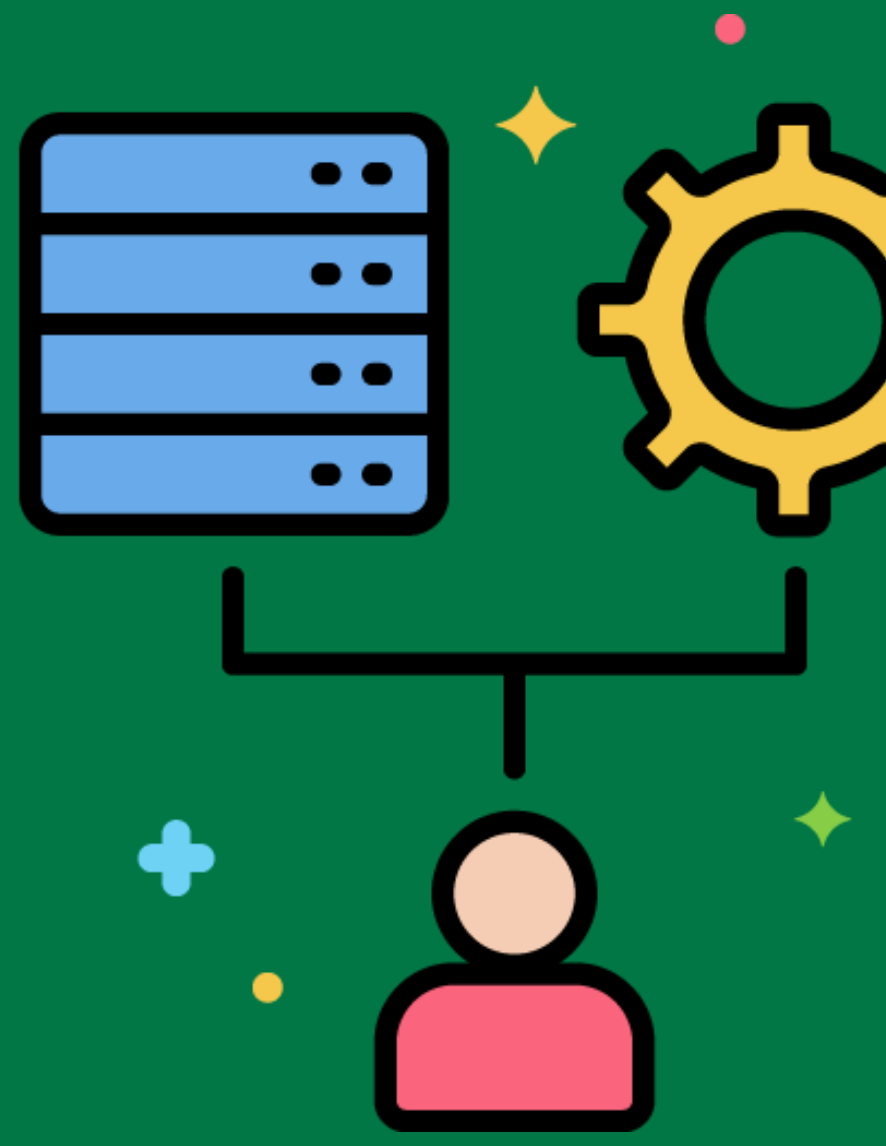
Data Engineering 101 - ETL Terminology

Full Load



Reloading all data from the source system into the data warehouse.

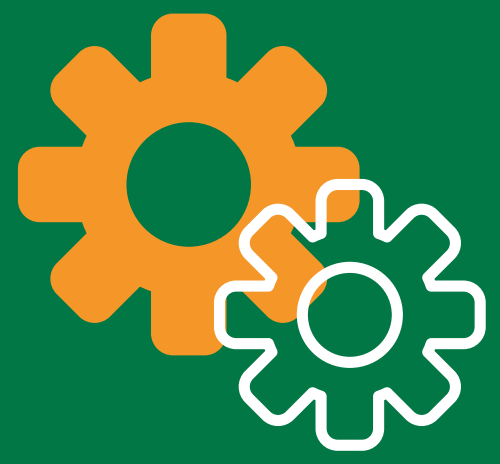
Performing a full load of historical sales data into the data warehouse during the initial ETL process.



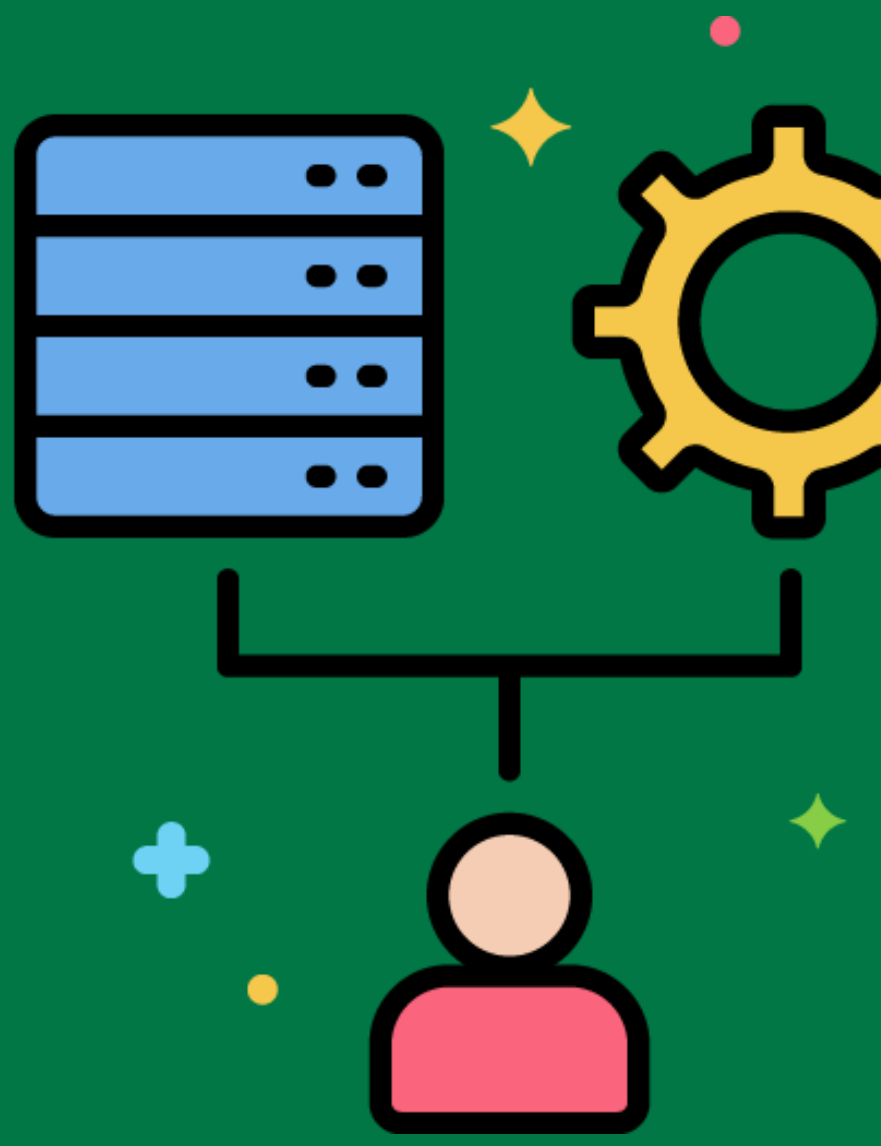
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Archiving



Storing historical data that is no longer actively used but may be needed for future reference.

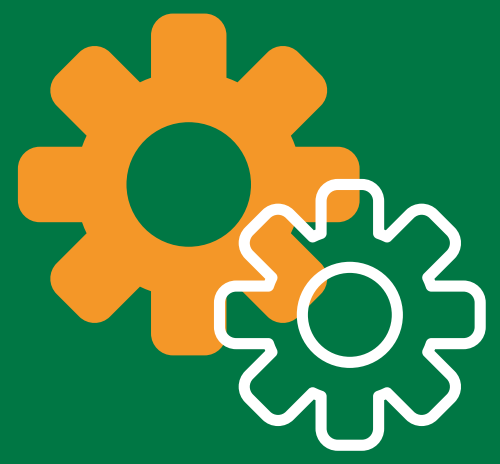


Archiving old sales transaction data to free up space in the active data warehouse tables.

Shwetank Singh
GritSetGrow - GSGLearn.com

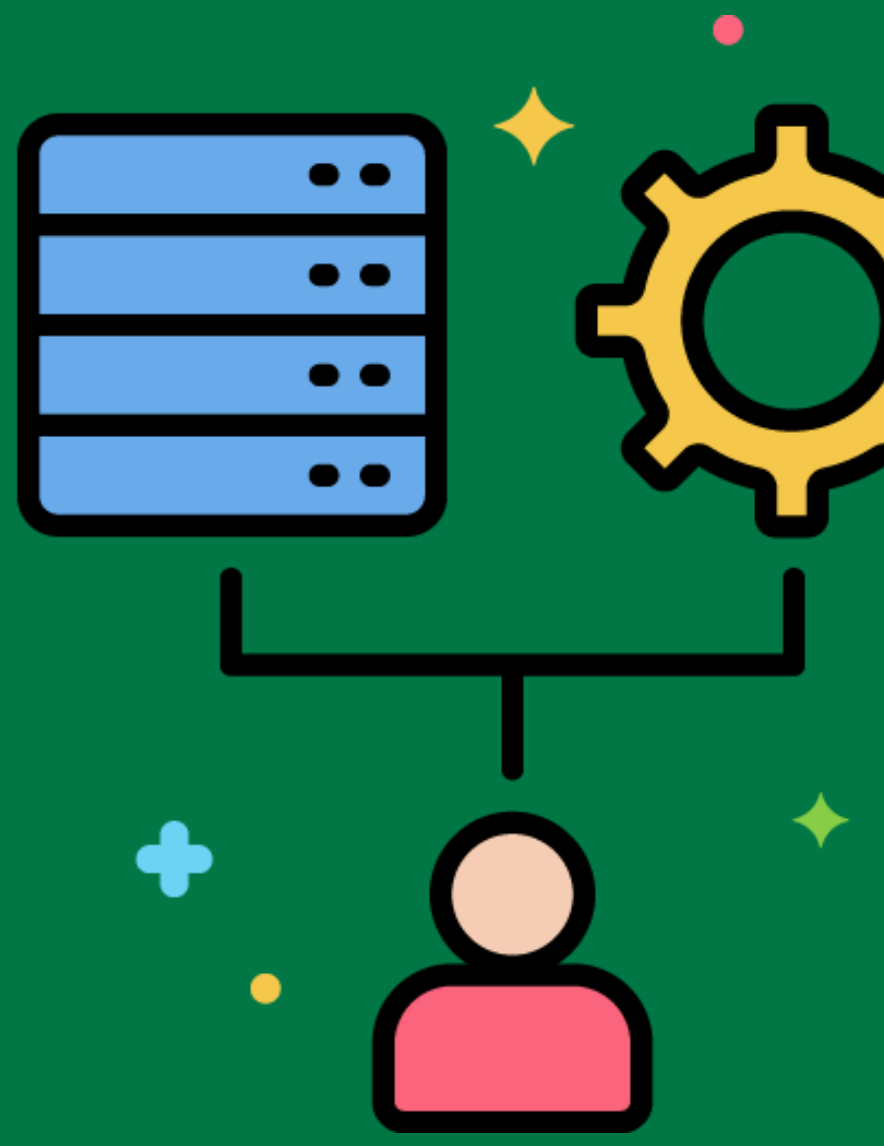
Data Engineering 101 - ETL Terminology

Data Purging



Removing data that is no longer needed from the data warehouse.

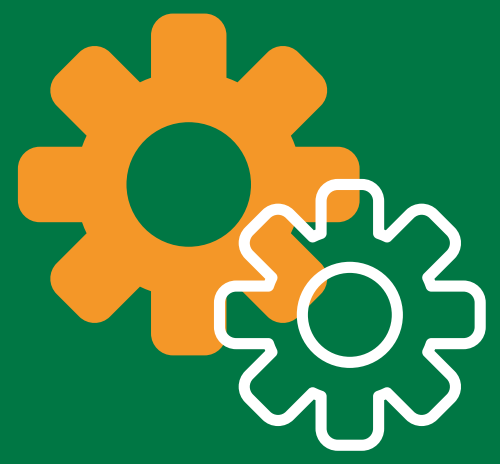
Deleting customer records that have been inactive for more than ten years.



Shwetank Singh
GritSetGrow - GSGLearn.com

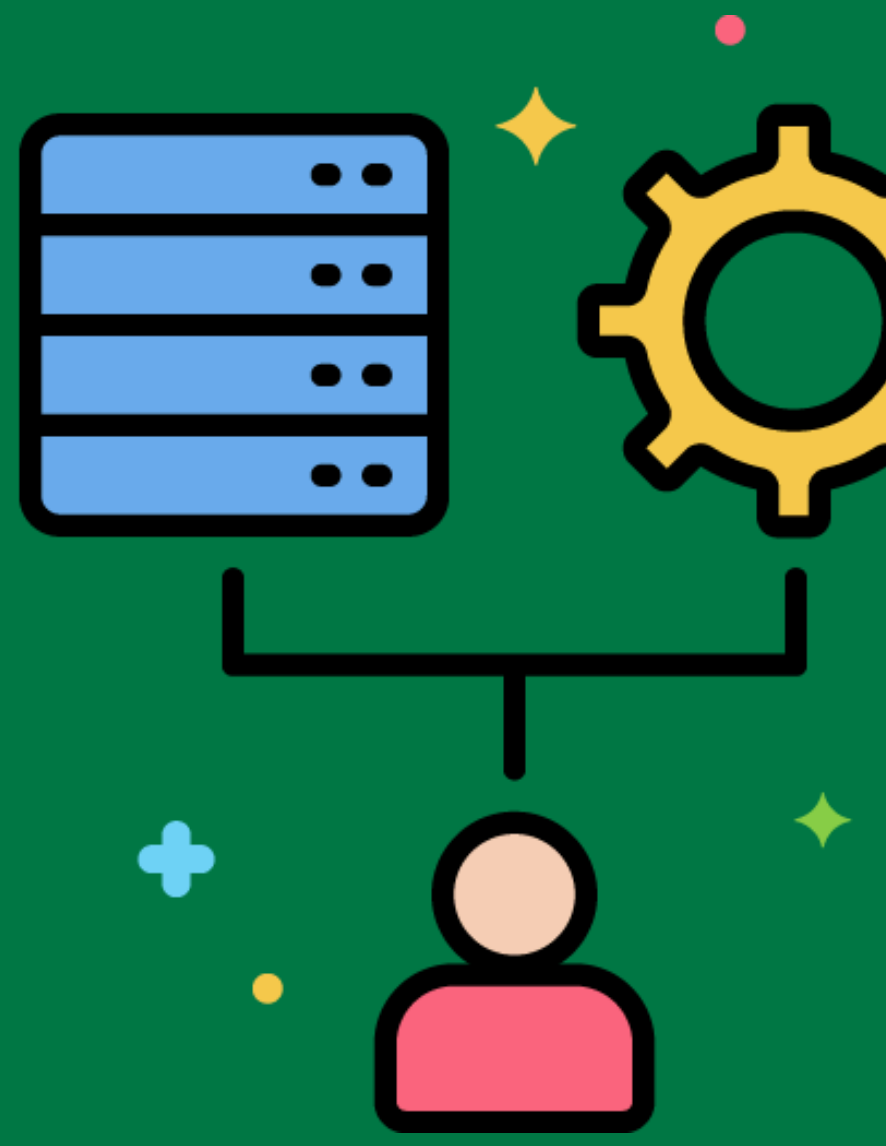
Data Engineering 101 - ETL Terminology

Data Backup



Creating copies of data to protect against data loss.

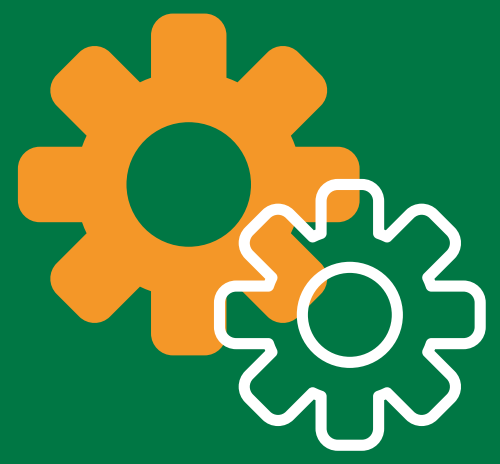
Regularly backing up the data warehouse to an off-site storage location.



Shwetank Singh
GritSetGrow - GSGLearn.com

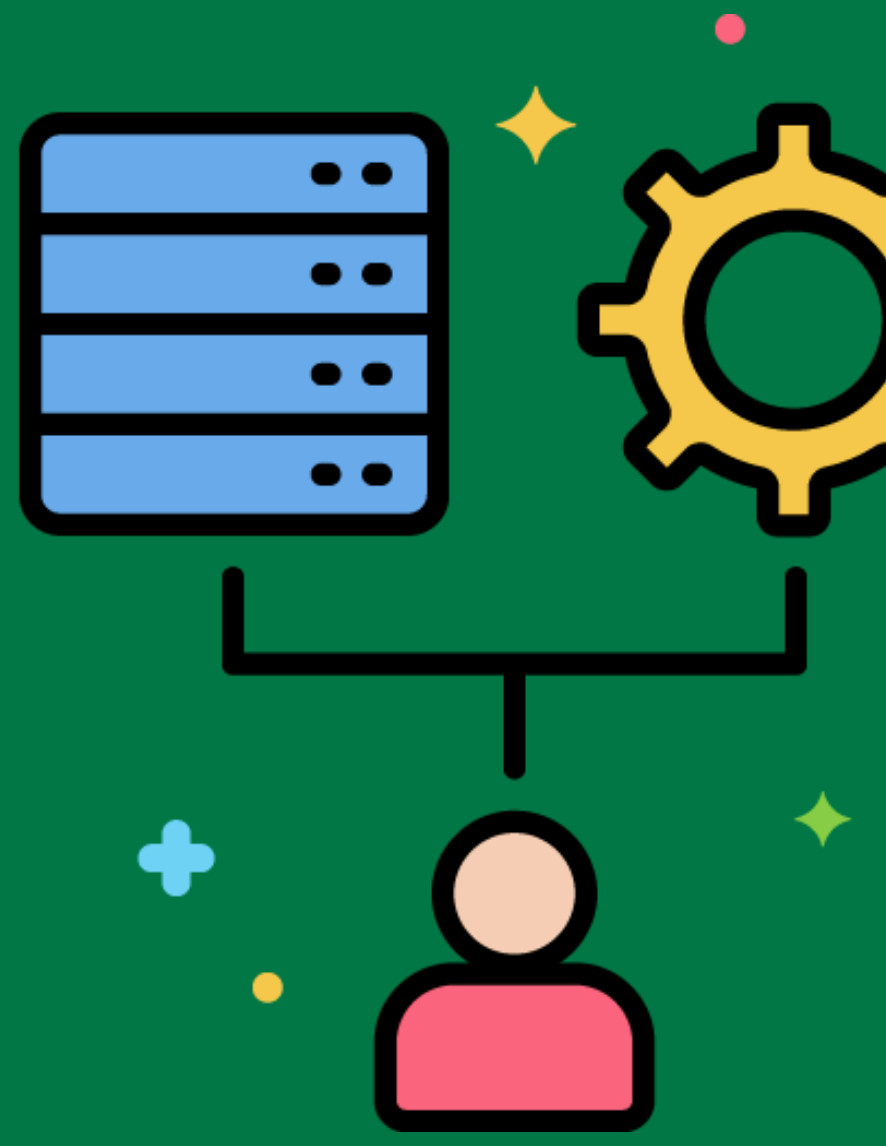
Data Engineering 101 - ETL Terminology

Data Recovery



Restoring data from backups in case of data loss or corruption.

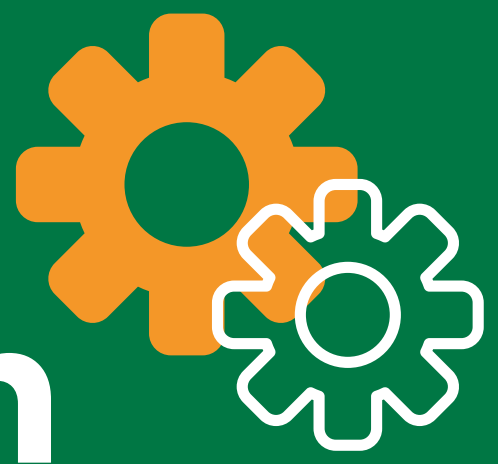
Restoring the data warehouse from a backup after a hardware failure.



Shwetank Singh
GritSetGrow - GSGLearn.com

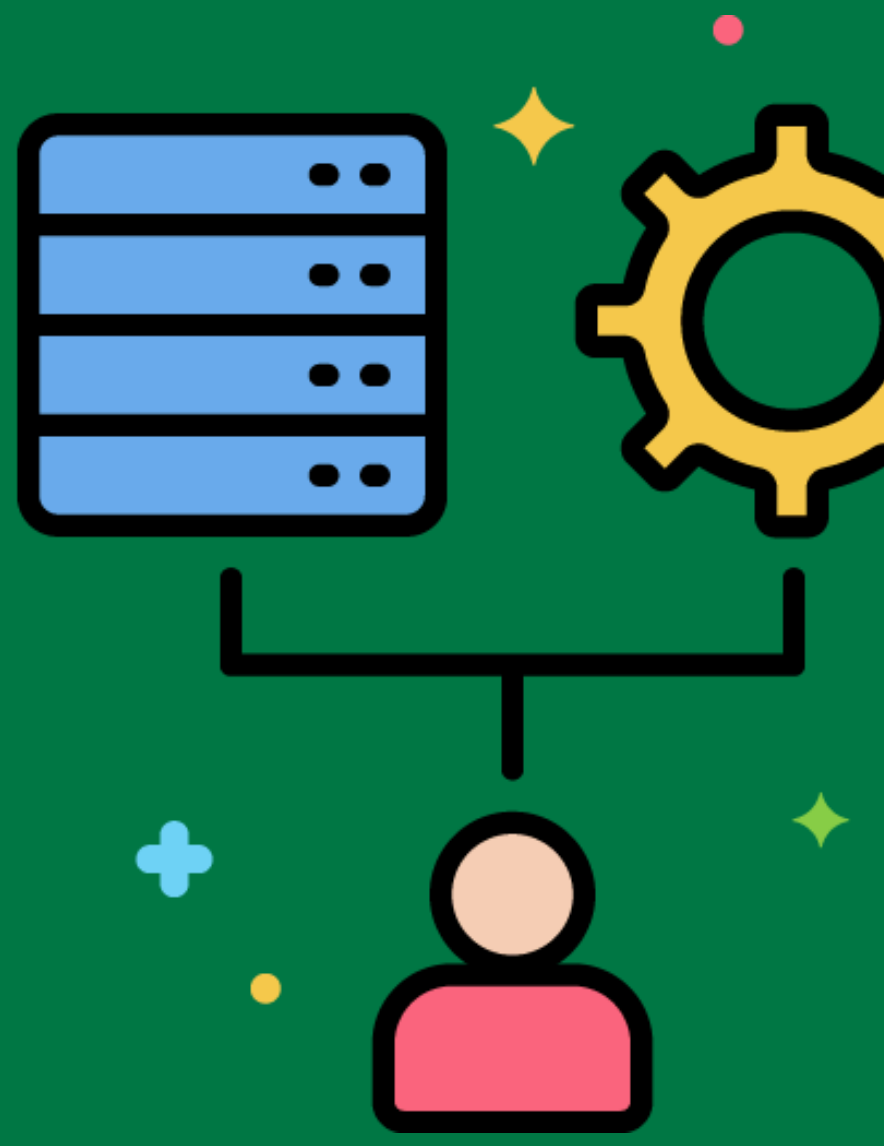
Data Engineering 101 - ETL Terminology

Data Anonymization



Removing or masking personally identifiable information (PII) in the data warehouse.

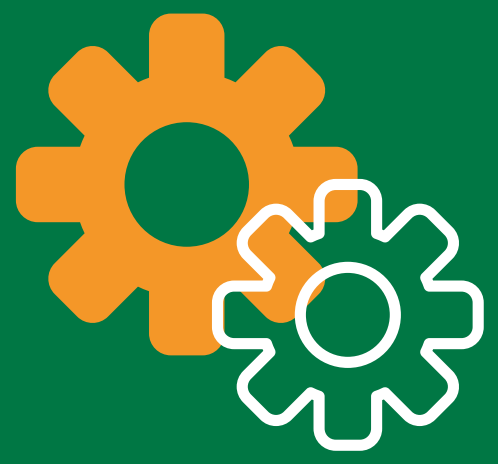
Anonymizing customer names and addresses in a dataset used for public analysis.



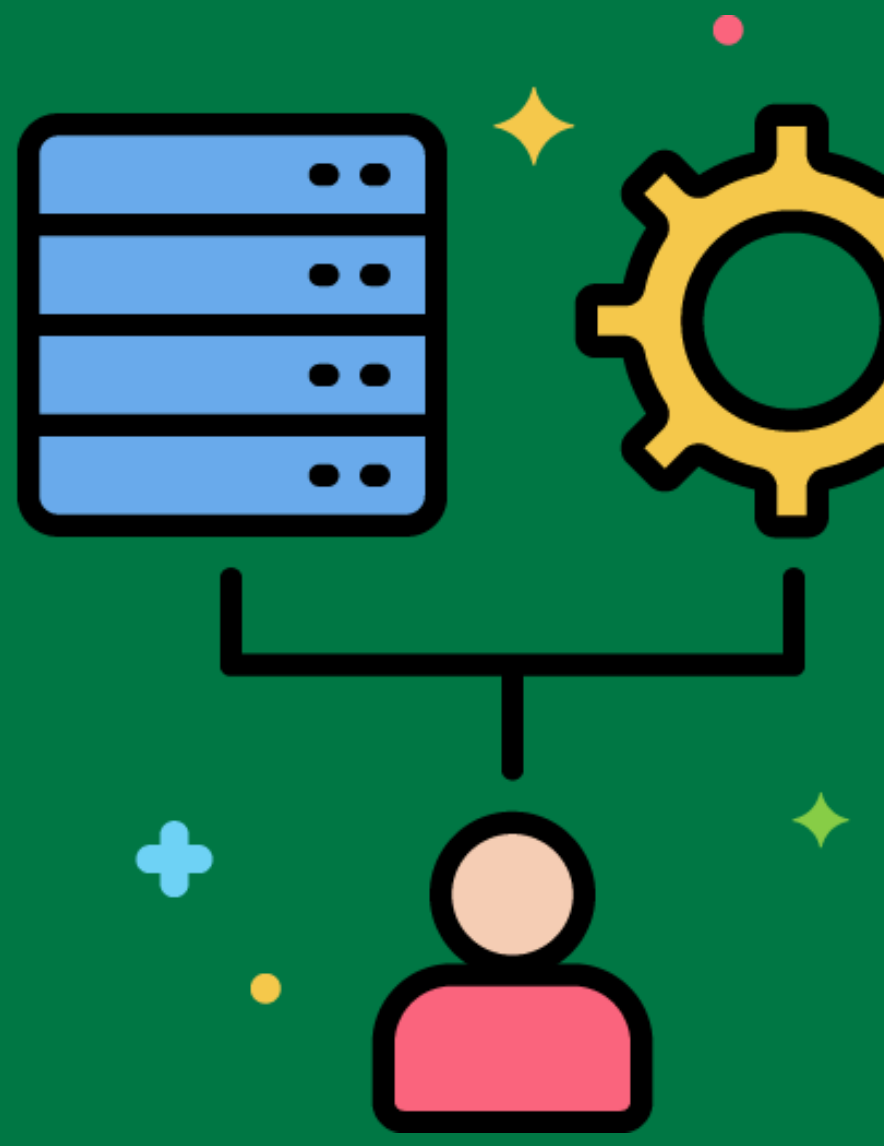
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Encryption



Protecting data by converting it into a secure format that can only be read with the proper decryption key.

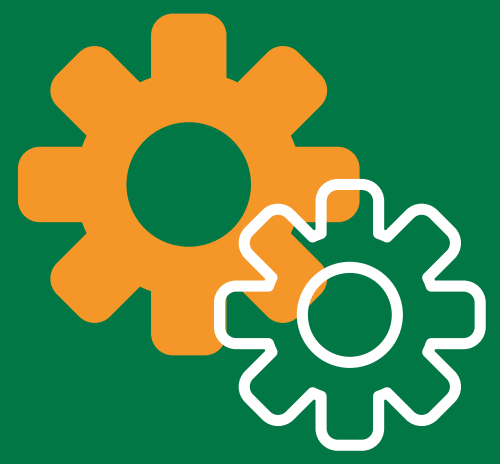


Encrypting sensitive customer data before loading it into the data warehouse.

Shwetank Singh
GritSetGrow - GSGLearn.com

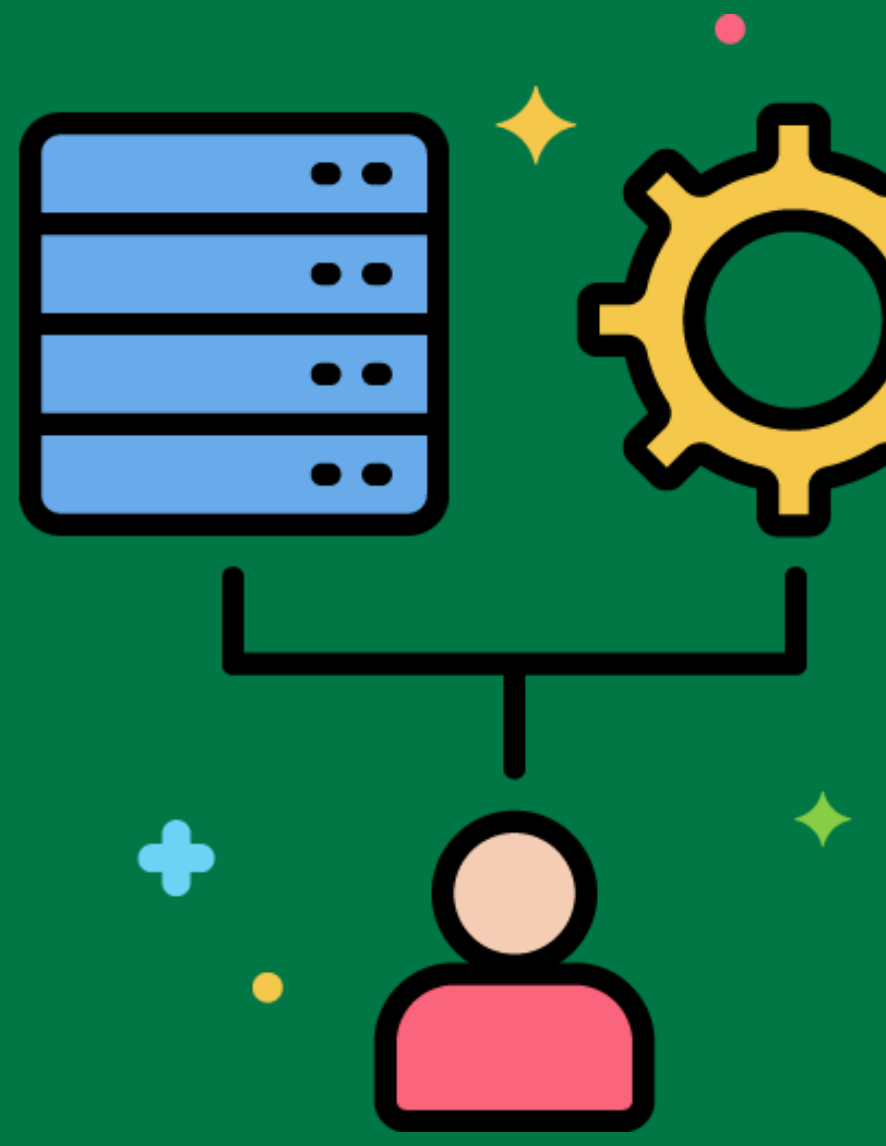
Data Engineering 101 - ETL Terminology

Data Masking



Hiding sensitive data by replacing it with fictional but realistic data.

Masking credit card numbers in the data warehouse to protect against unauthorized access.

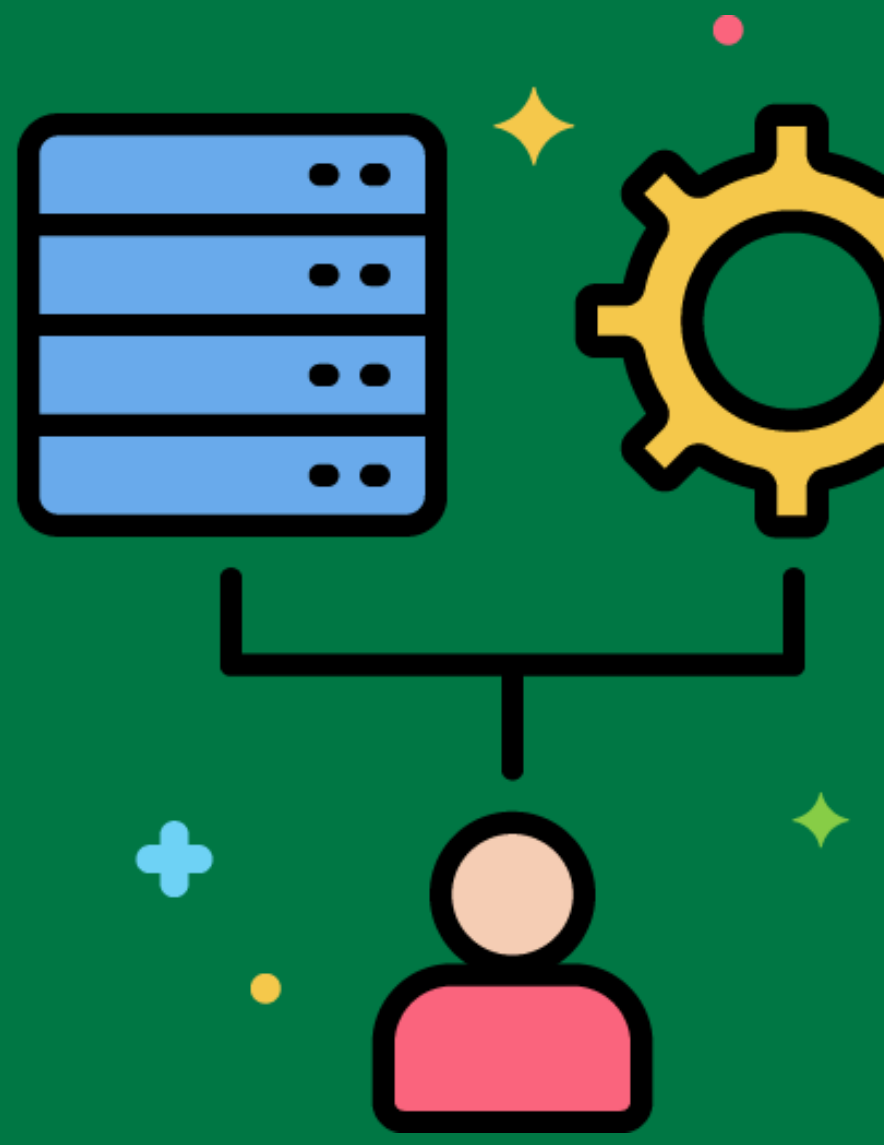


Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation

Converting data from one format or structure to another to meet the requirements of the data warehouse.

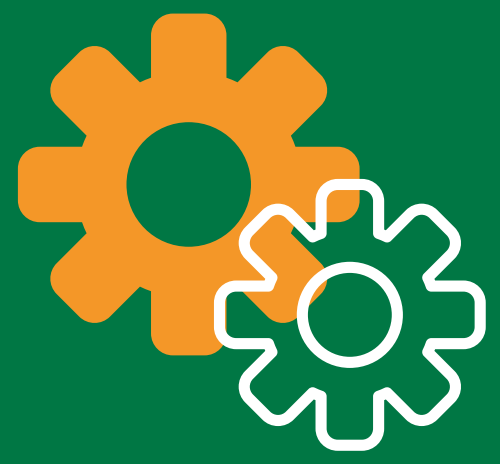


Transforming date formats from MM/DD/YYYY to YYYY-MM-DD during the ETL process.

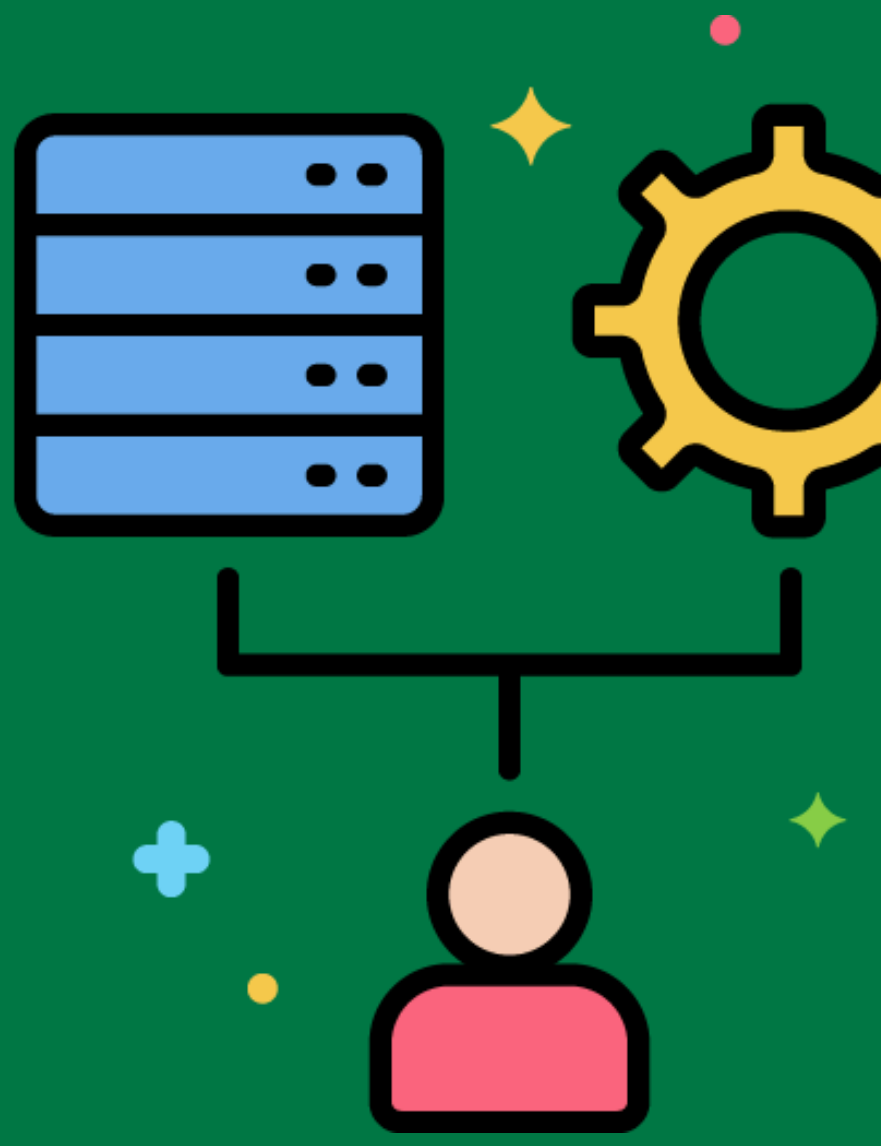
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Validation



Checking data for accuracy and consistency before loading it into the data warehouse.

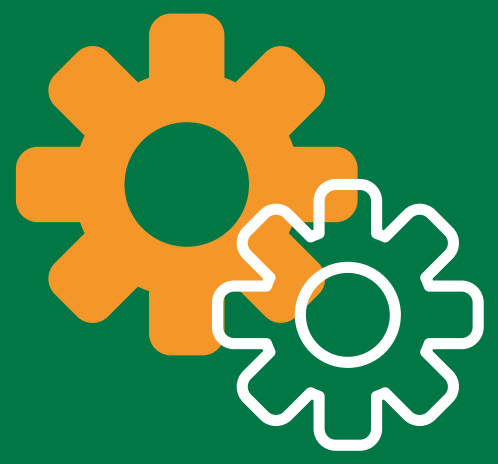


Validating email addresses in customer records to ensure they are in the correct format.

Shwetank Singh
GritSetGrow - GSGLearn.com

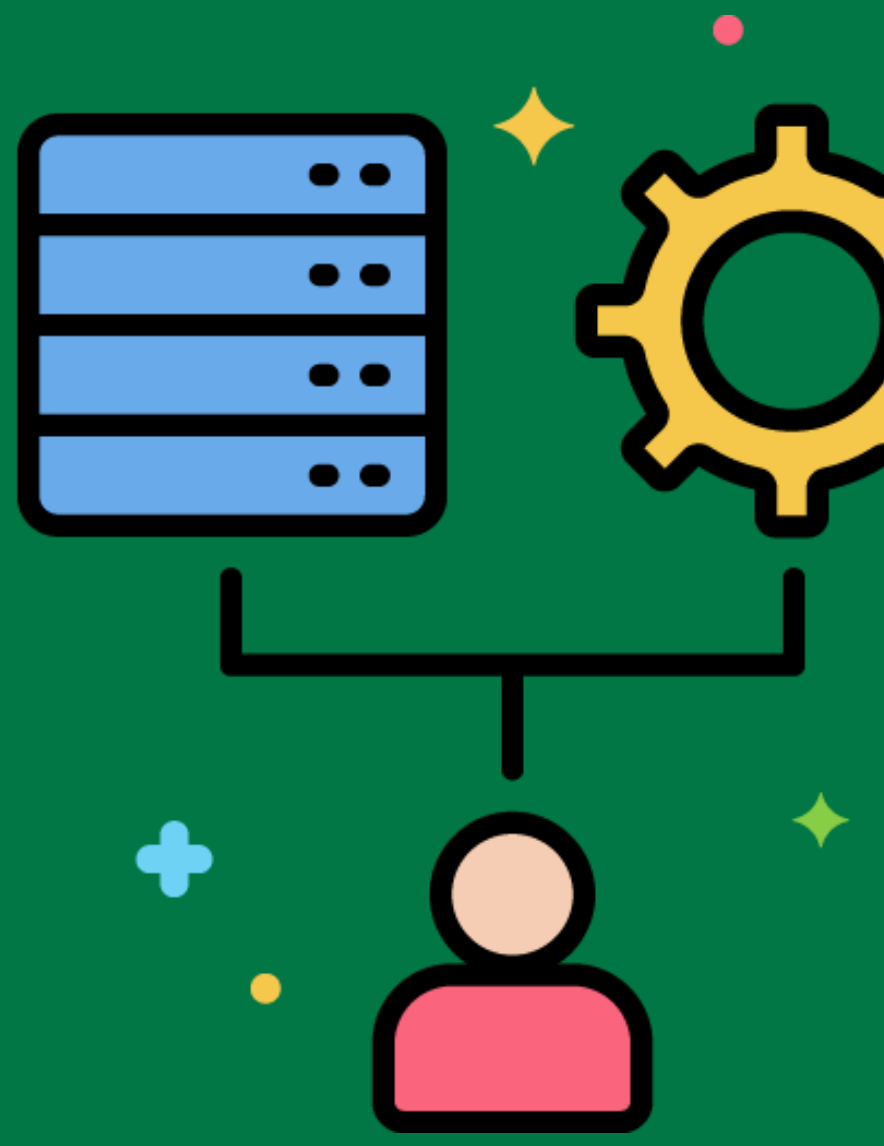
Data Engineering 101 - ETL Terminology

ETL Logging



Recording the steps and events that occur during the ETL process for monitoring and debugging purposes.

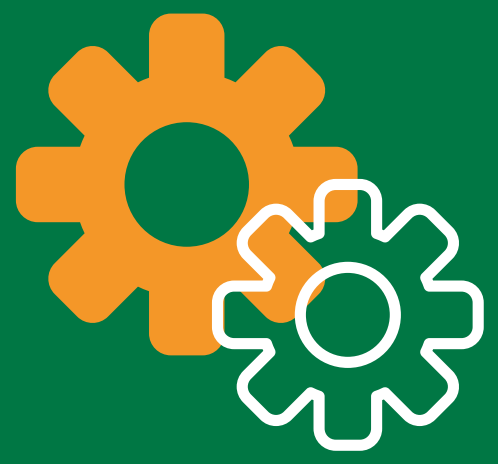
Keeping a log of all extraction, transformation, and load activities, including any errors encountered.



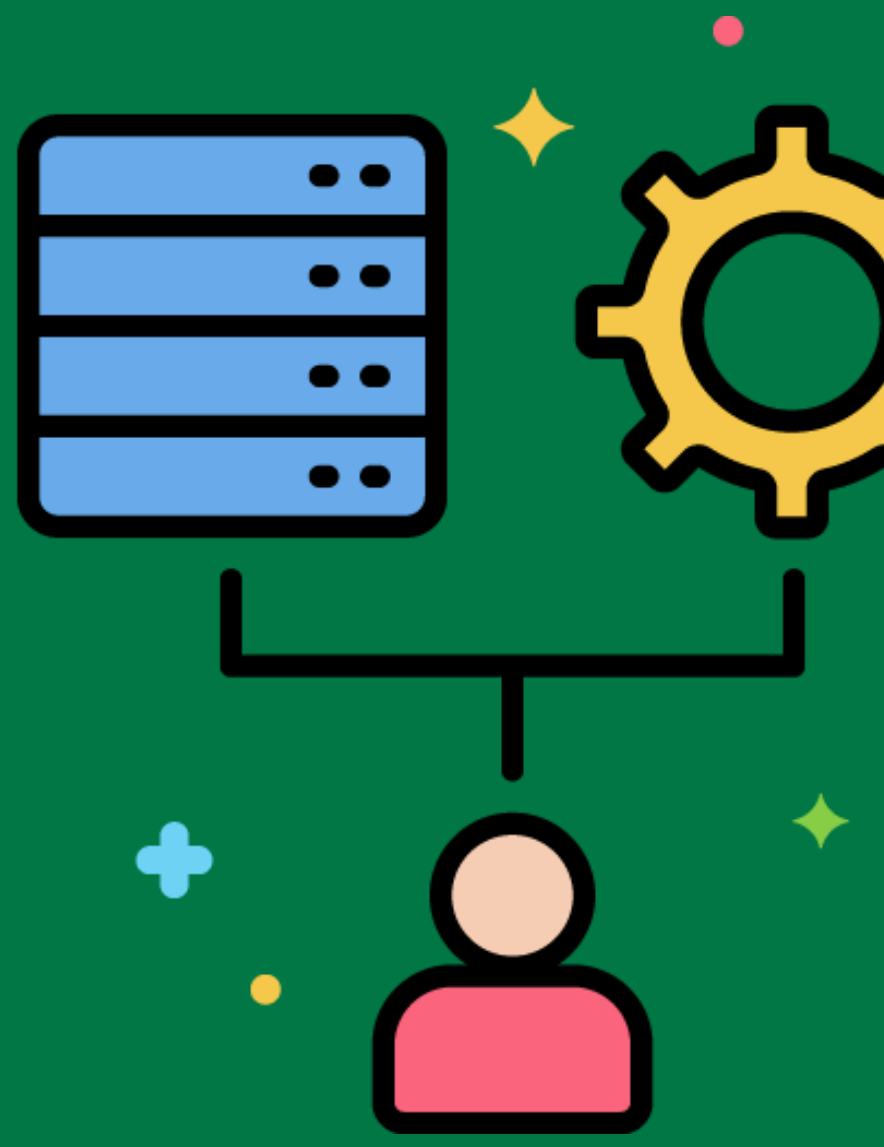
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Monitoring



Continuously tracking the performance and status of the ETL process to ensure it runs smoothly.

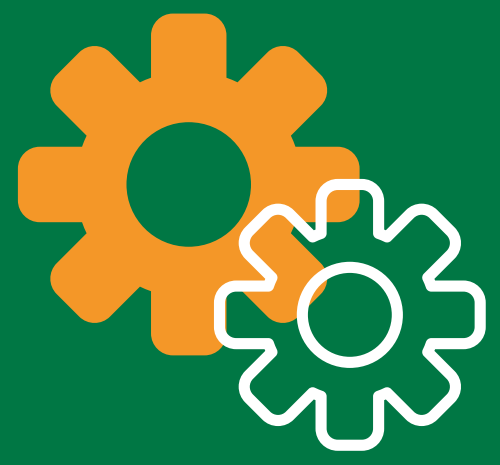


Using an ETL monitoring tool to track the progress and performance of daily data loads.

Shwetank Singh
GritSetGrow - GSGLearn.com

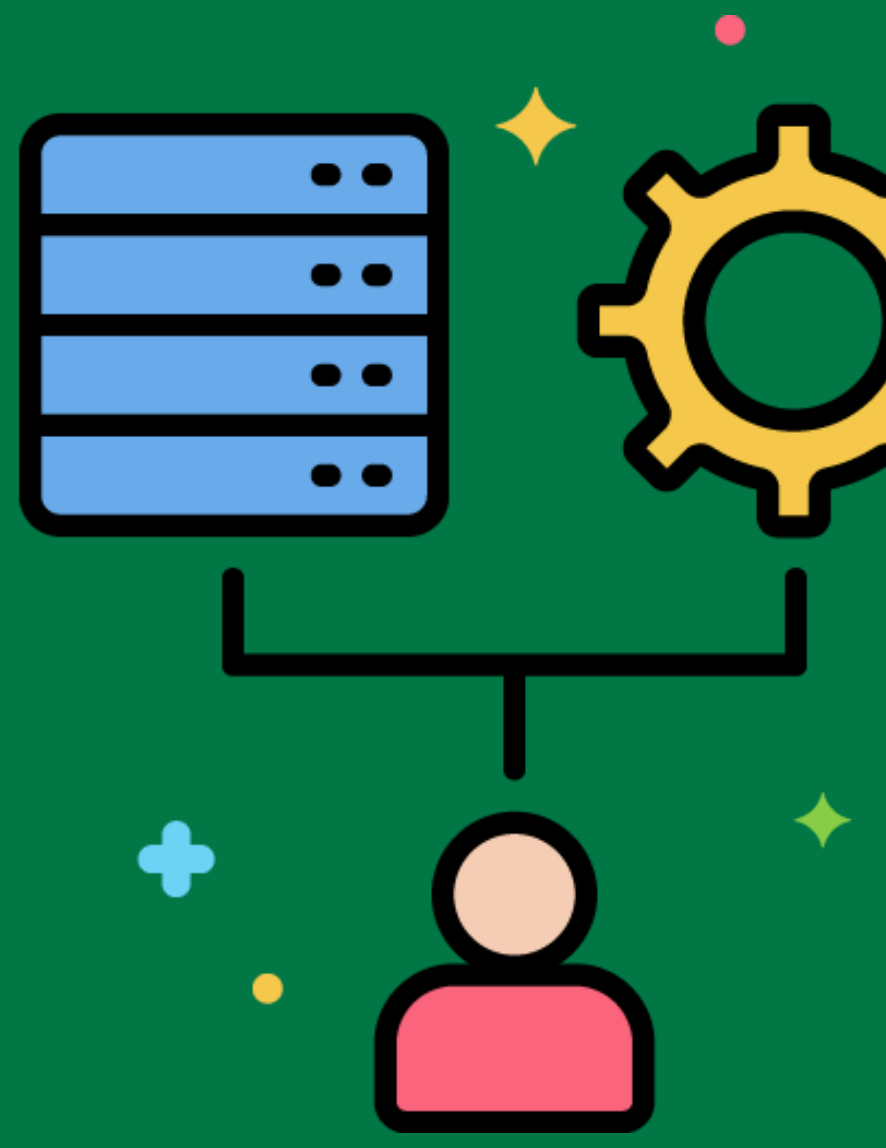
Data Engineering 101 - ETL Terminology

ETL Performance Tuning



Optimizing the ETL process to improve speed and efficiency.

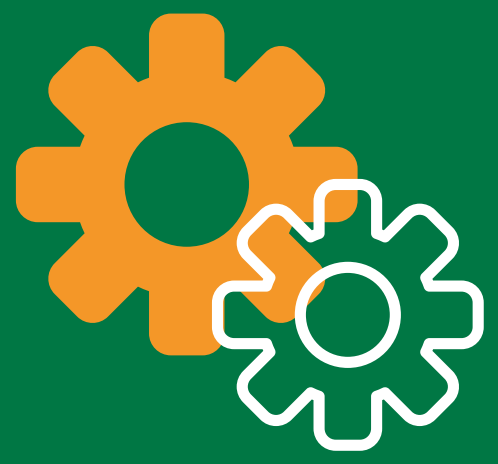
Adjusting the ETL process to reduce the time it takes to load sales data by optimizing SQL queries and parallelizing tasks.



Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Mart



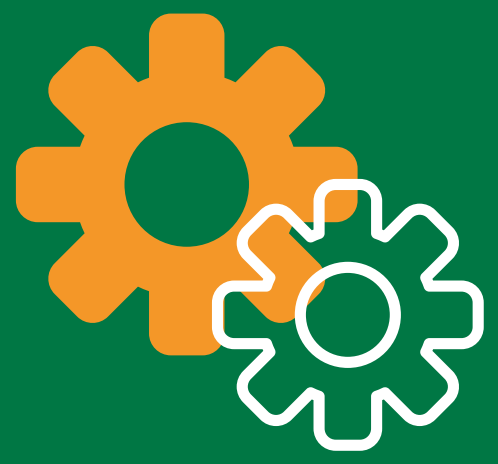
A subset of the data warehouse that focuses on a specific area or department within the organization.

Creating a sales data mart that includes only sales-related data for the sales department.

Shwetank Singh
GritSetGrow - GSGLearn.com

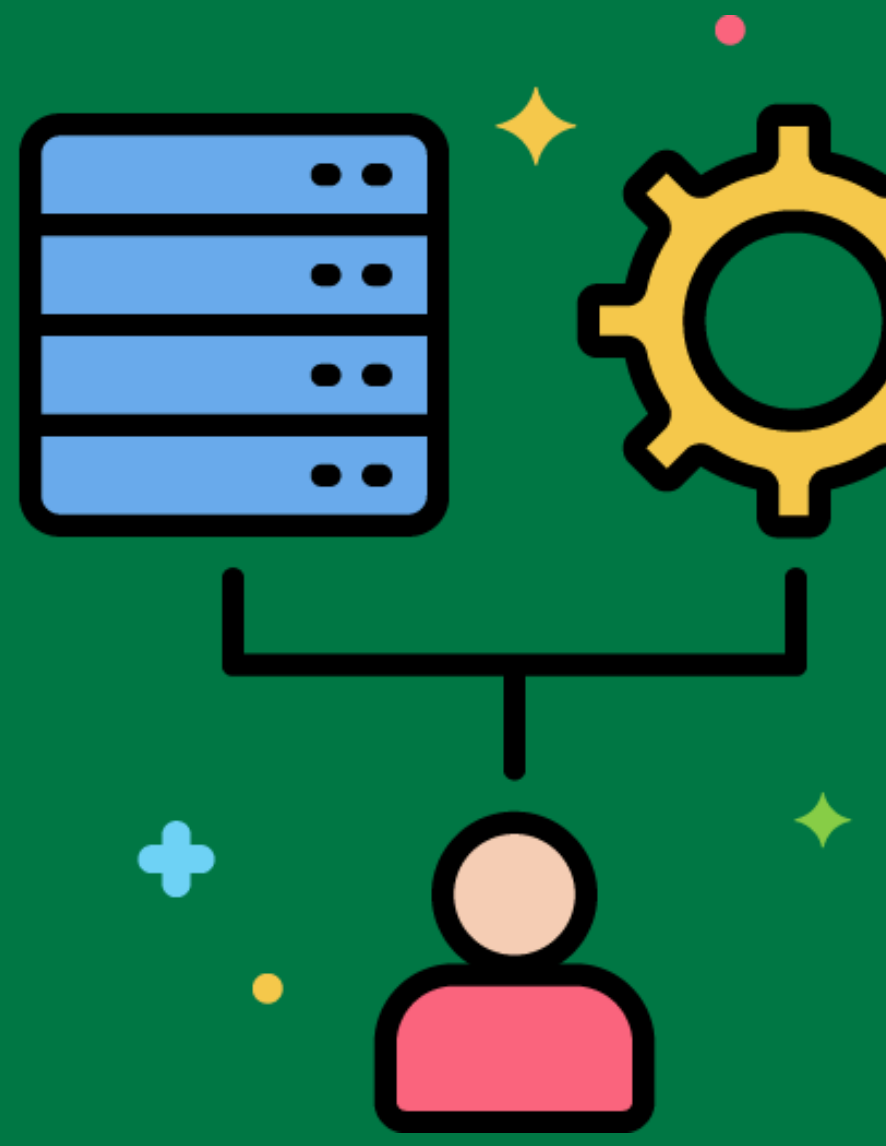
Data Engineering 101 - ETL Terminology

Data Lake



A centralized repository that allows you to store all your structured and unstructured data at any scale.

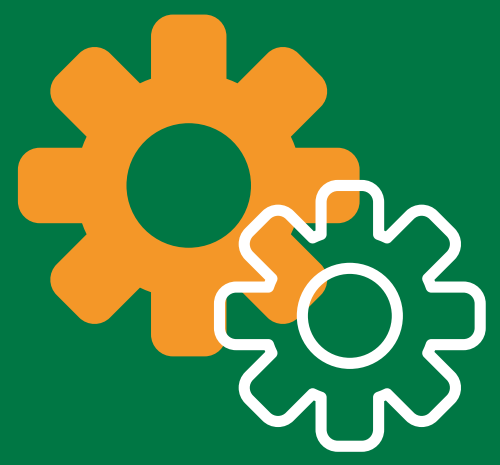
Storing raw and unprocessed data from various sources in a data lake for later analysis.



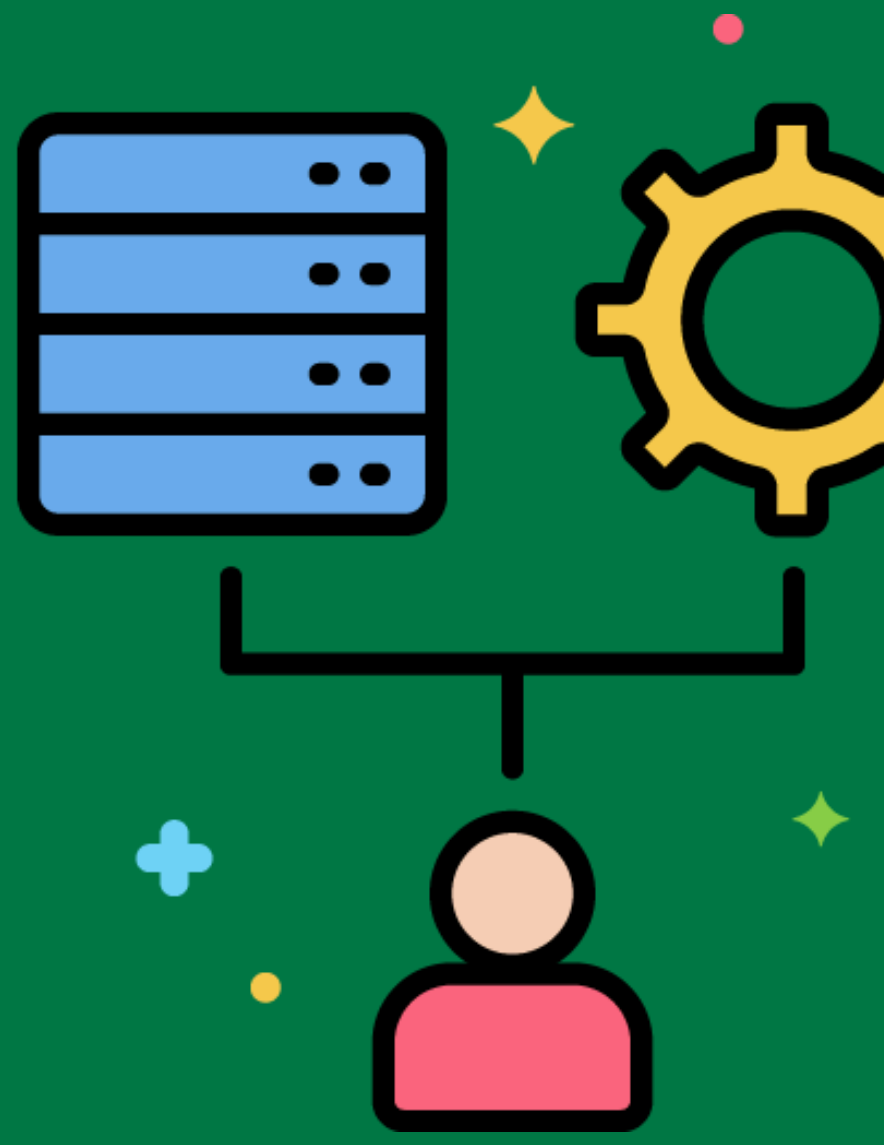
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Integration Tools



Software tools that help to combine data from different sources and provide a unified view.

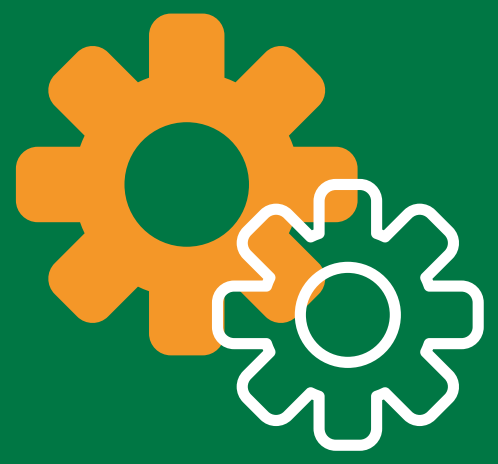


Using Informatica or Talend for integrating data from multiple source systems into the data warehouse.

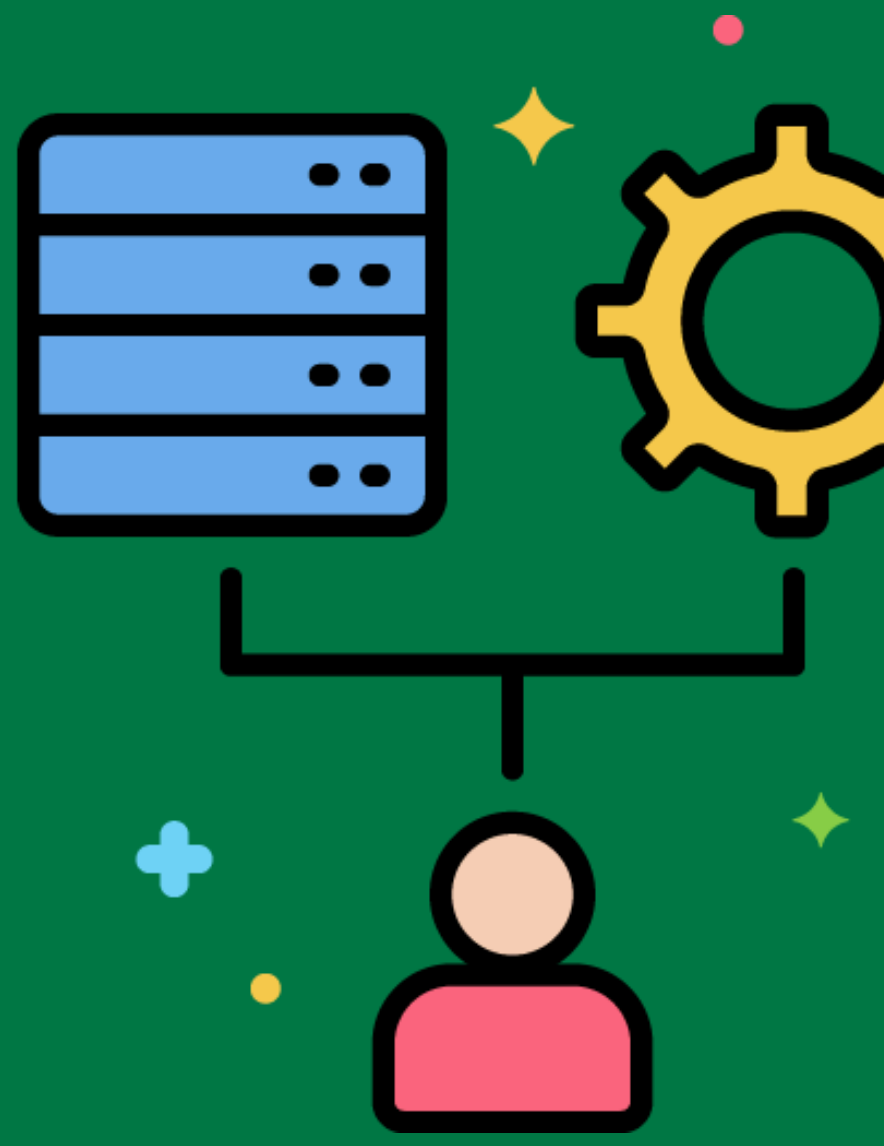
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Tool



Software tools that facilitate the extraction, transformation, and loading of data into the data warehouse.

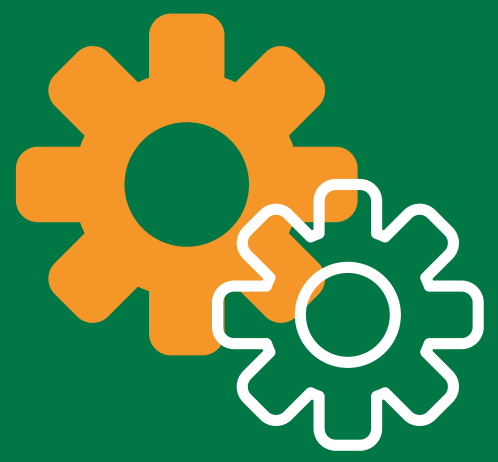


Using tools like Apache Nifi or Microsoft SSIS to automate and manage the ETL process.

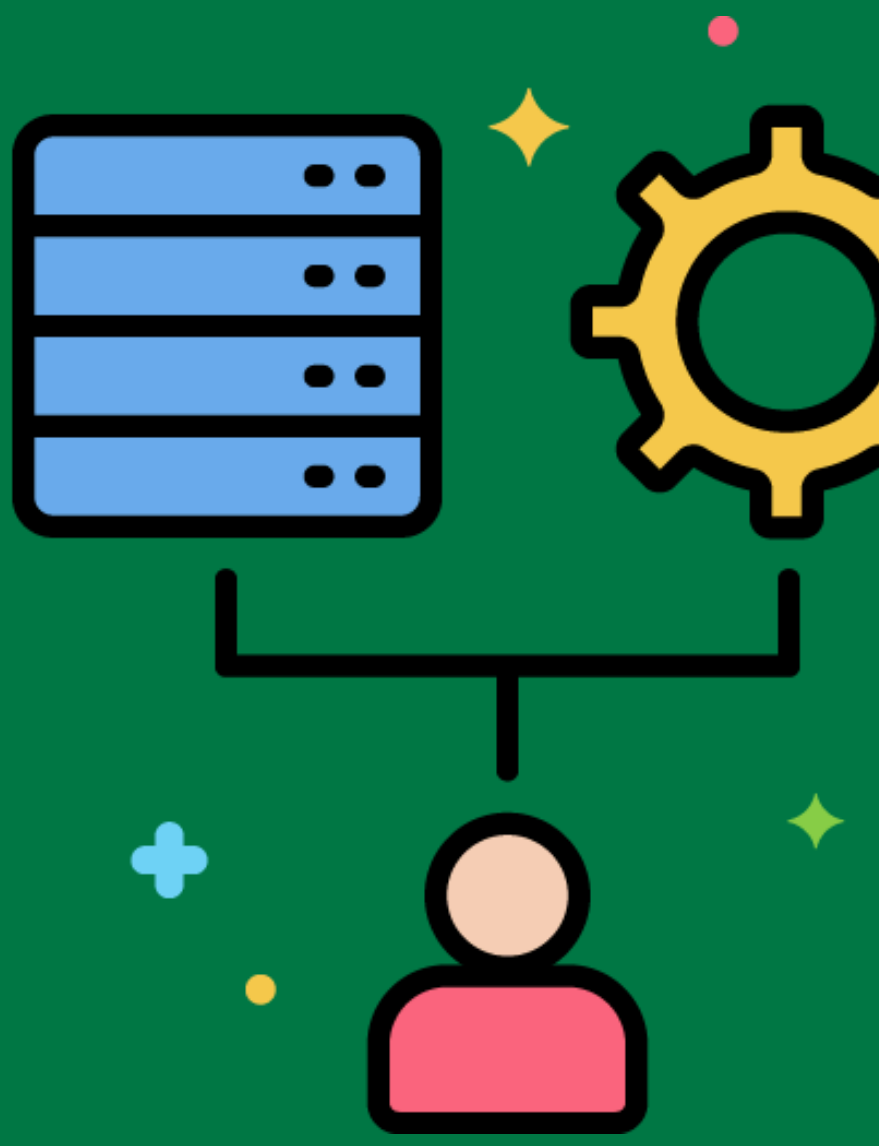
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Governance



The management of data availability, usability, integrity, and security in the data warehouse.

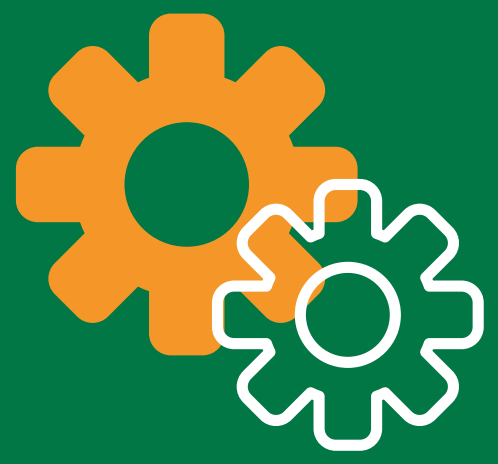


Implementing data governance policies to ensure data quality and compliance in the data warehouse.

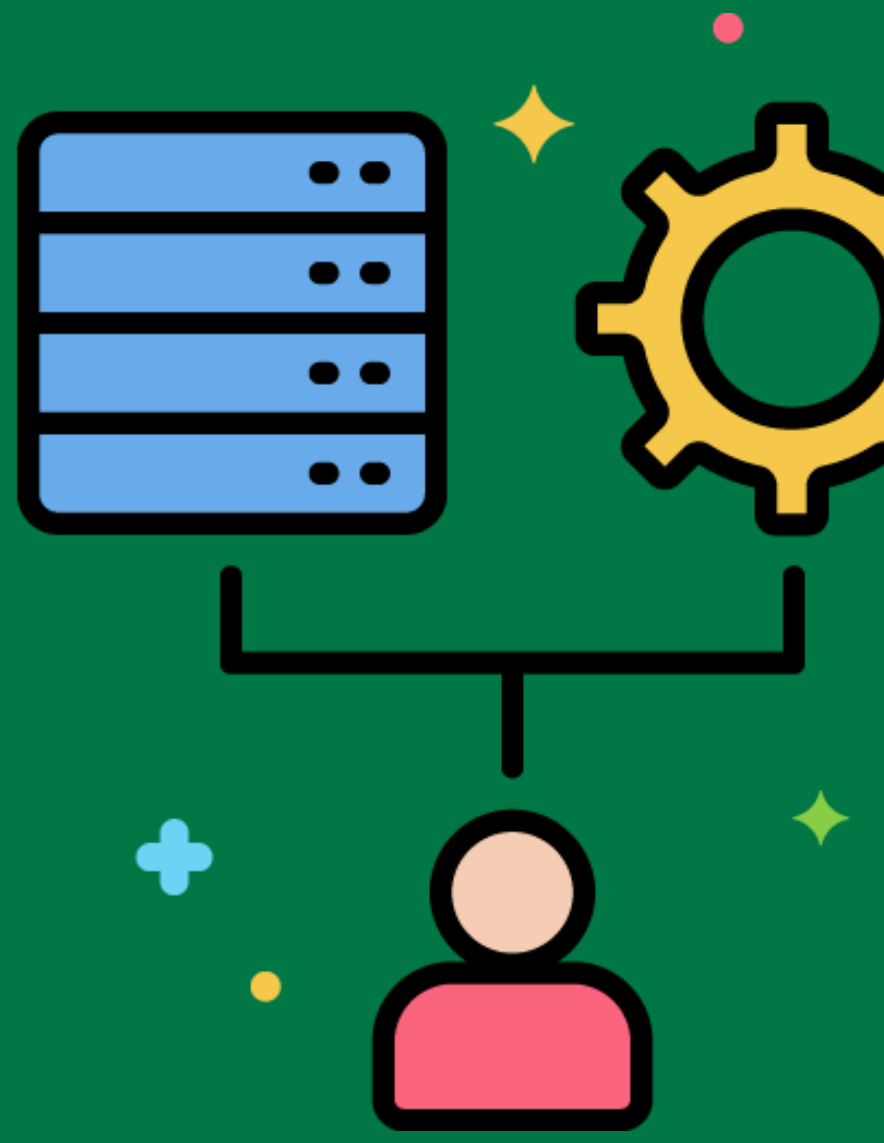
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Stewardship



The management and oversight of an organization's data assets to help provide business users with high-quality data that is easily accessible.

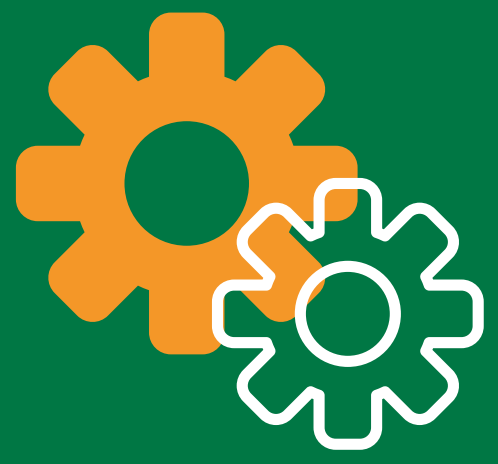


Assigning data stewards to oversee data quality and data management practices in the organization.

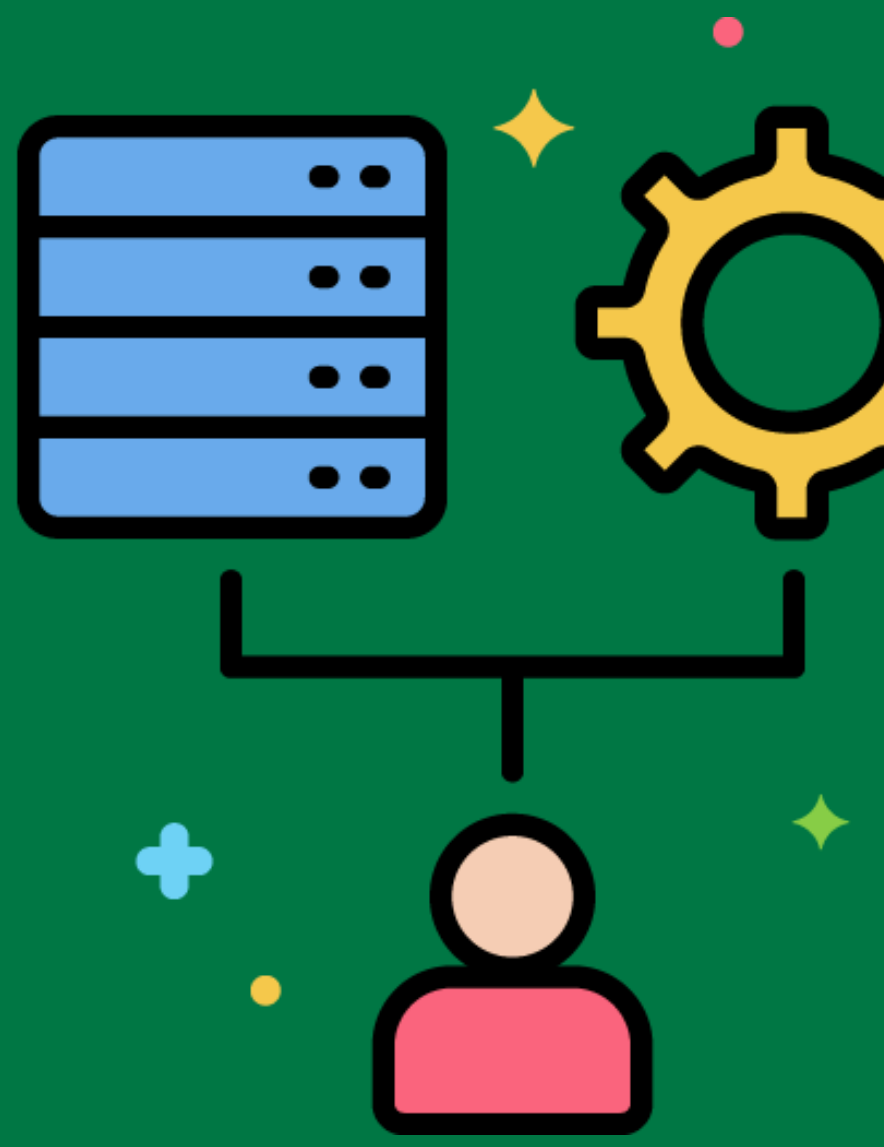
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Warehouse Architecture



The design and structure of a data warehouse, including its components and their relationships.

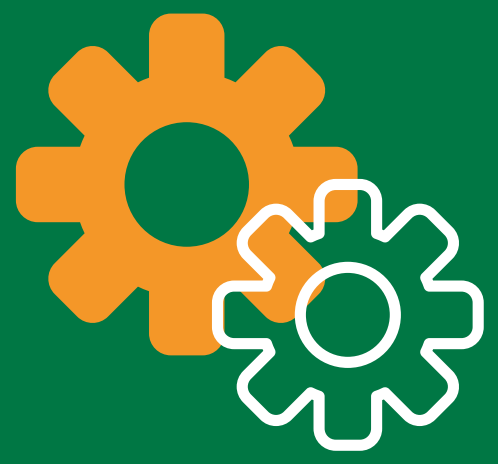


Designing a data warehouse architecture that includes a staging area, ETL process, and presentation layer.

Shwetank Singh
GritSetGrow - GSGLearn.com

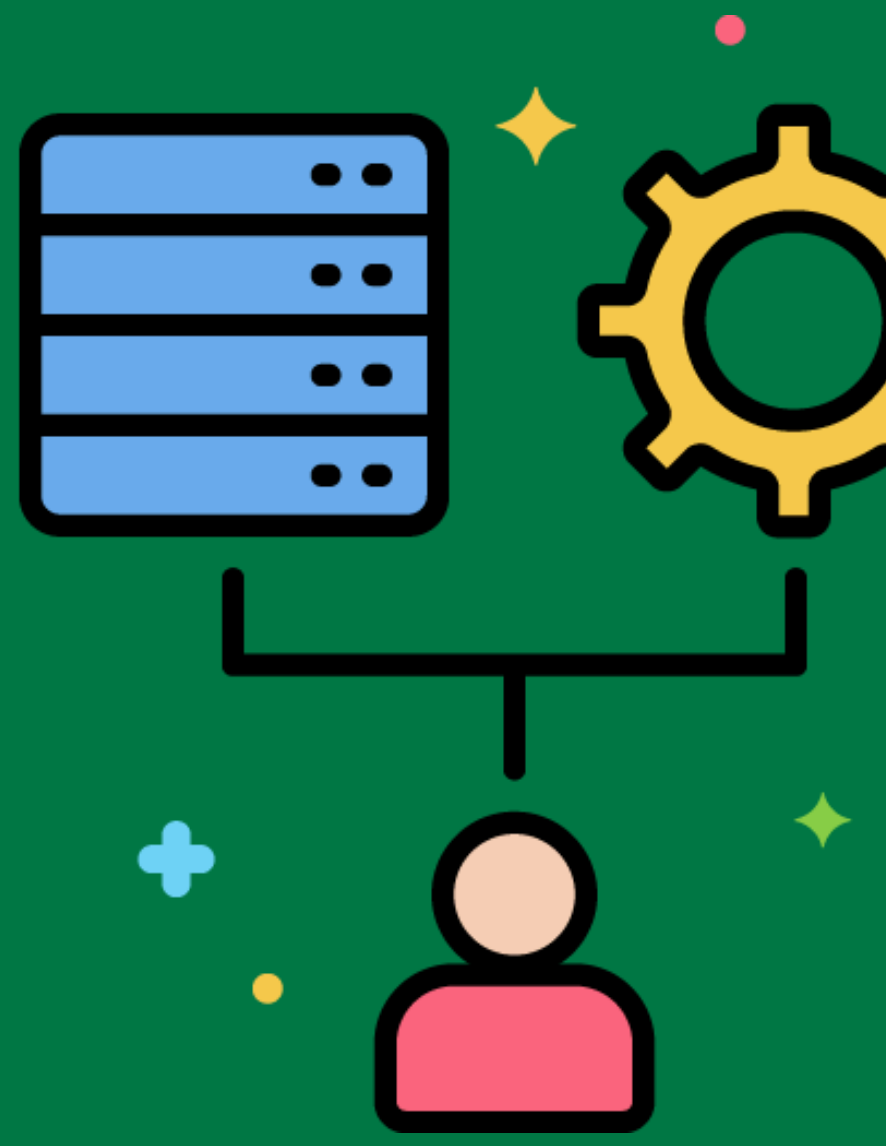
Data Engineering 101 - ETL Terminology

Dimensional Modeling



A data modeling technique optimized for data warehouse and OLAP cube implementations.

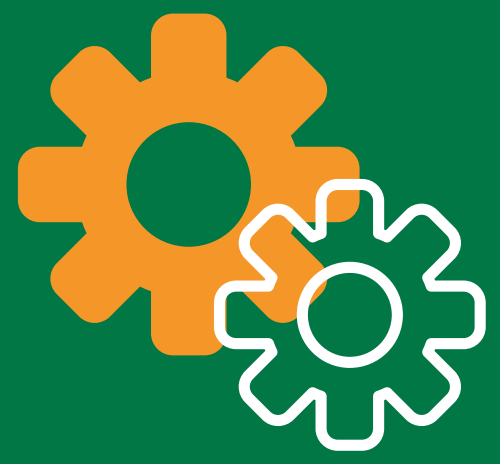
Designing a star schema or snowflake schema for the sales data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

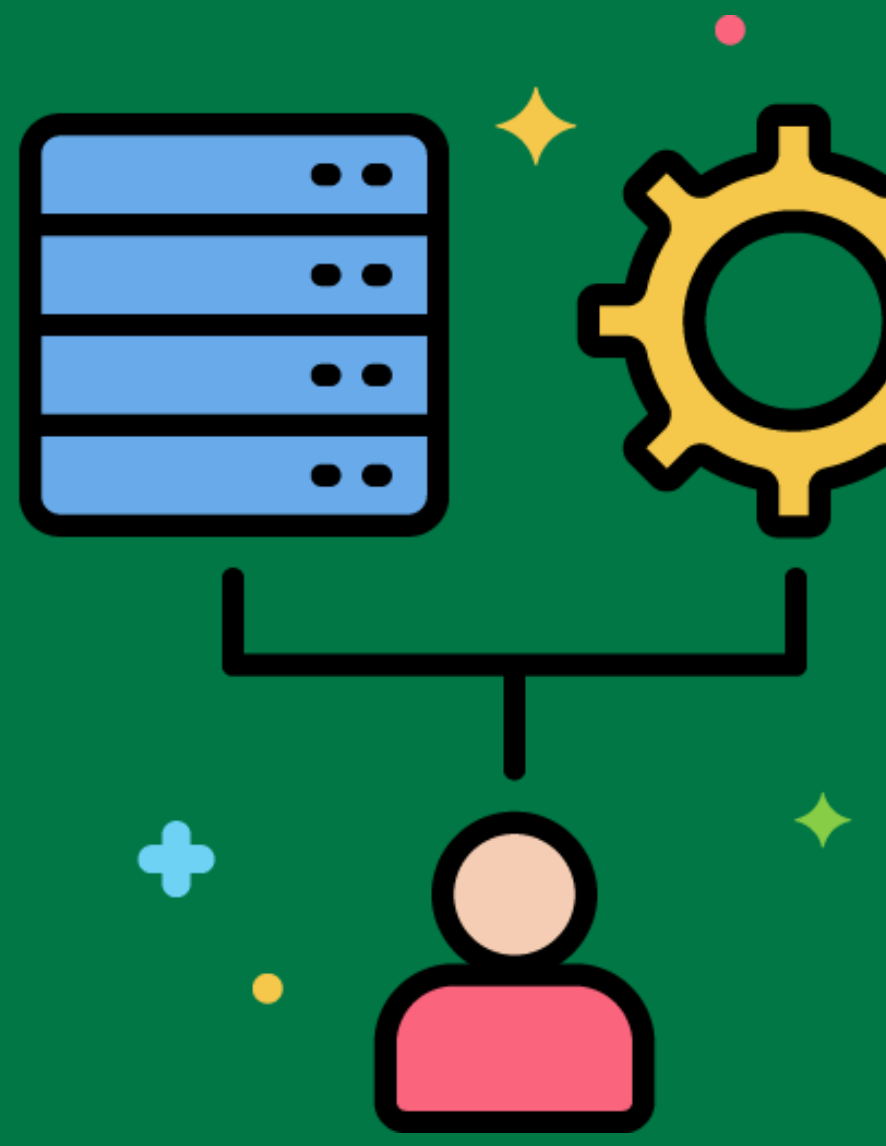
Data Engineering 101 - ETL Terminology

Star Schema



A type of database schema that is composed of a single fact table and multiple dimension tables.

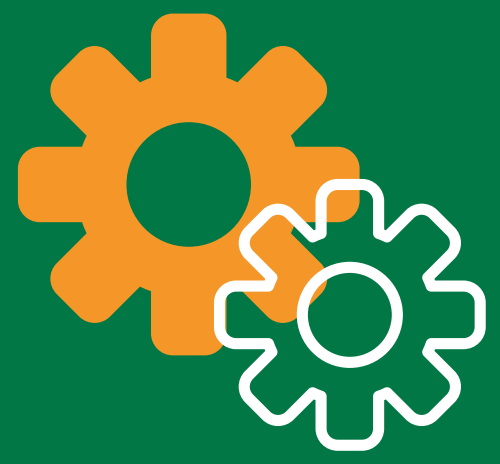
Designing a star schema with a central sales fact table connected to dimension tables for products, customers, and time.



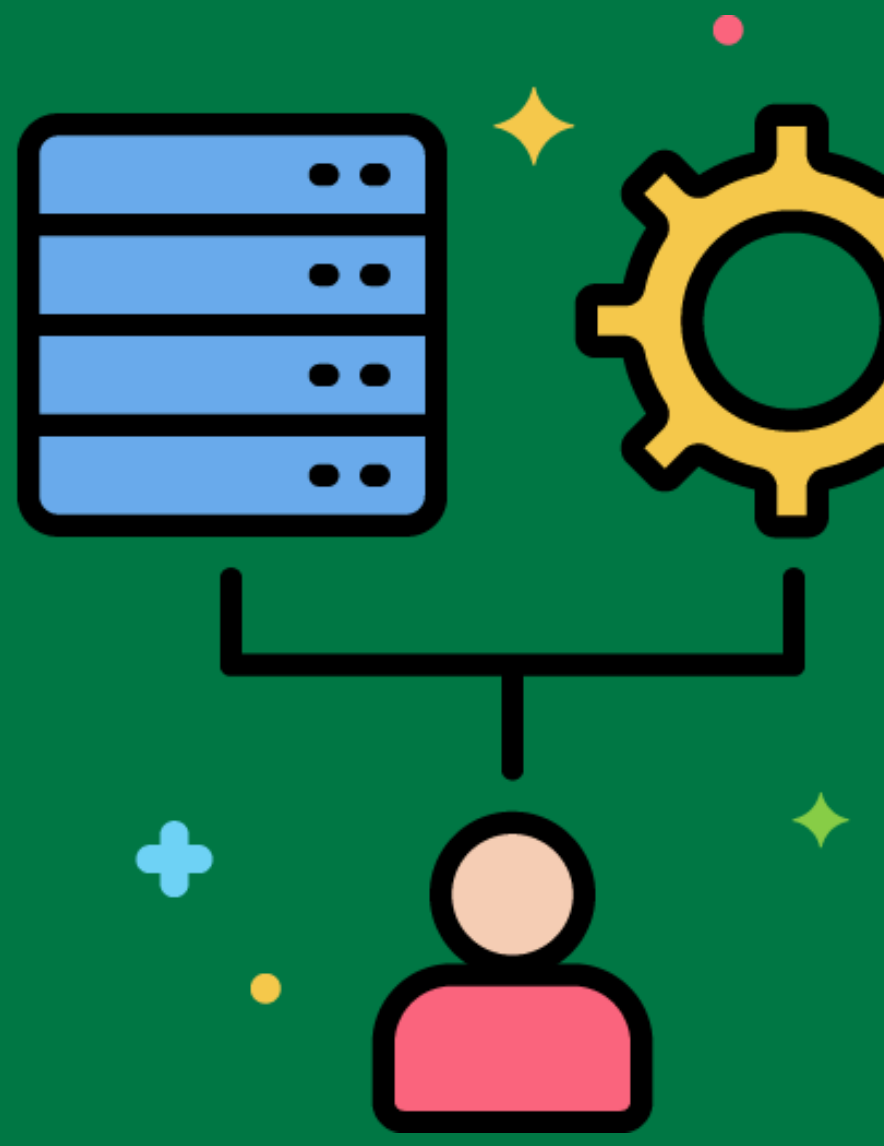
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Snowflake Schema



A type of database schema that is composed of a fact table and multiple dimension tables, where the dimension tables are normalized.

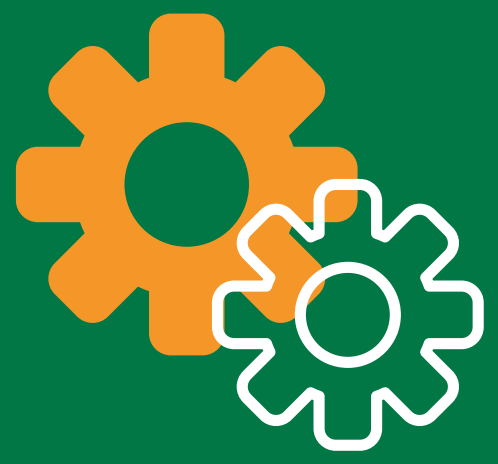


Designing a snowflake schema where the product dimension table is normalized into separate tables for product categories and subcategories.

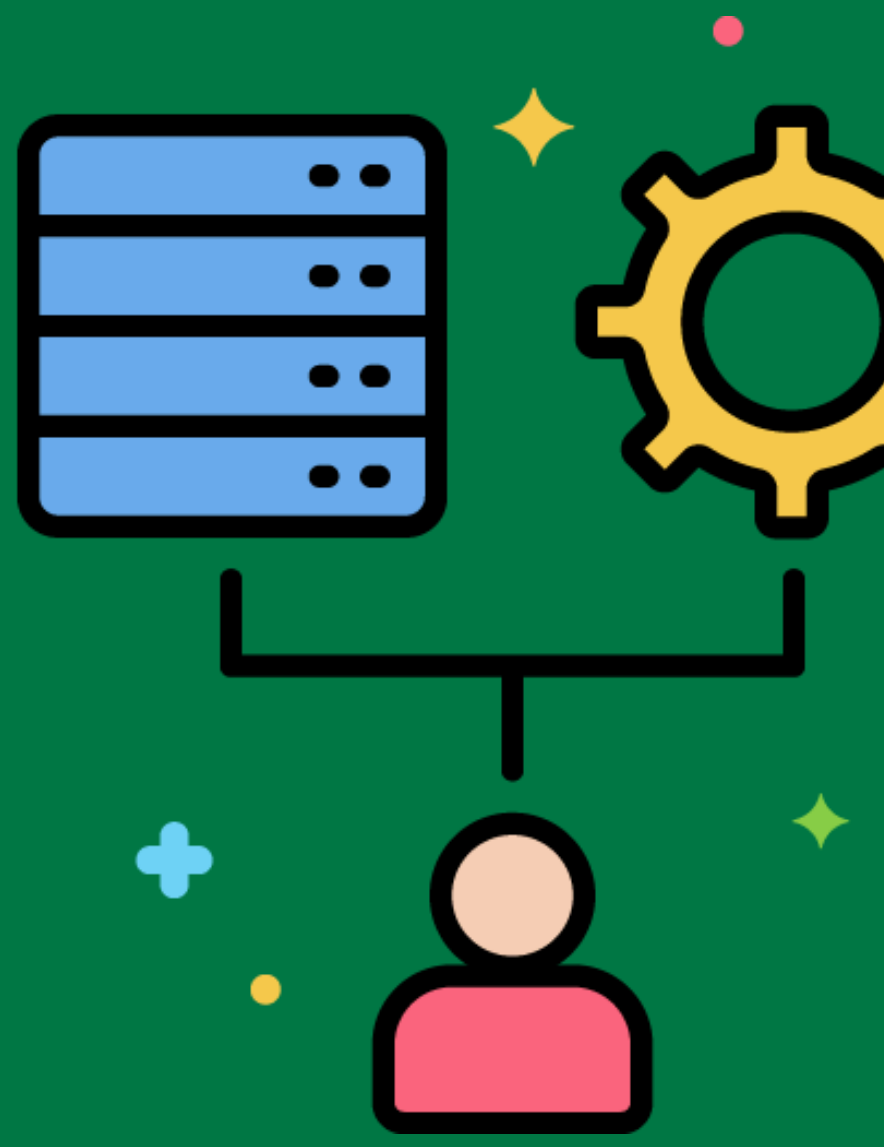
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Fact Table



A central table in a star or snowflake schema that contains the numerical measures of a business process.

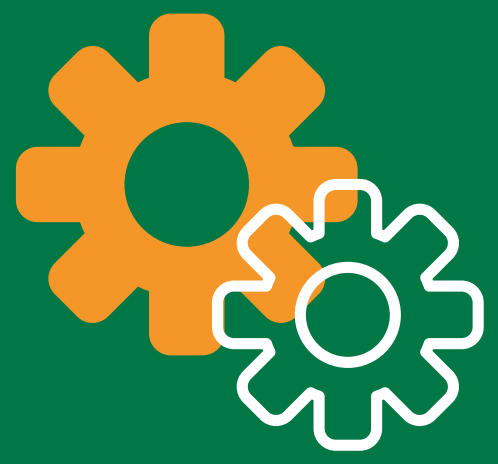


Creating a sales fact table that records sales transactions, including quantities sold and sales amounts.

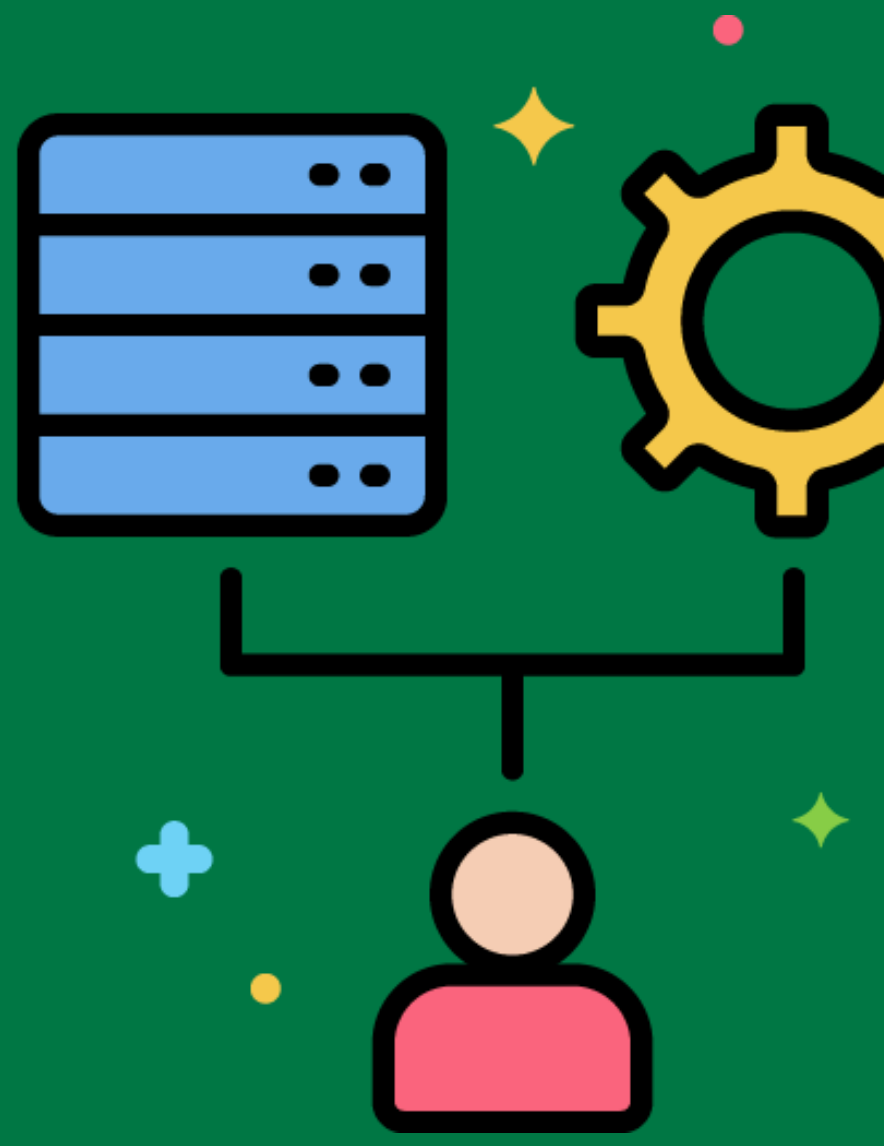
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Dimension Table



A table in a star or snowflake schema that contains descriptive attributes related to the dimensions of the business process.

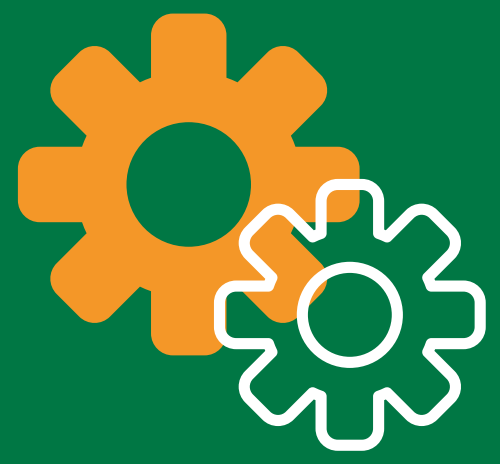


Creating a customer dimension table that includes attributes like customer name, address, and contact information.

Shwetank Singh
GritSetGrow - GSGLearn.com

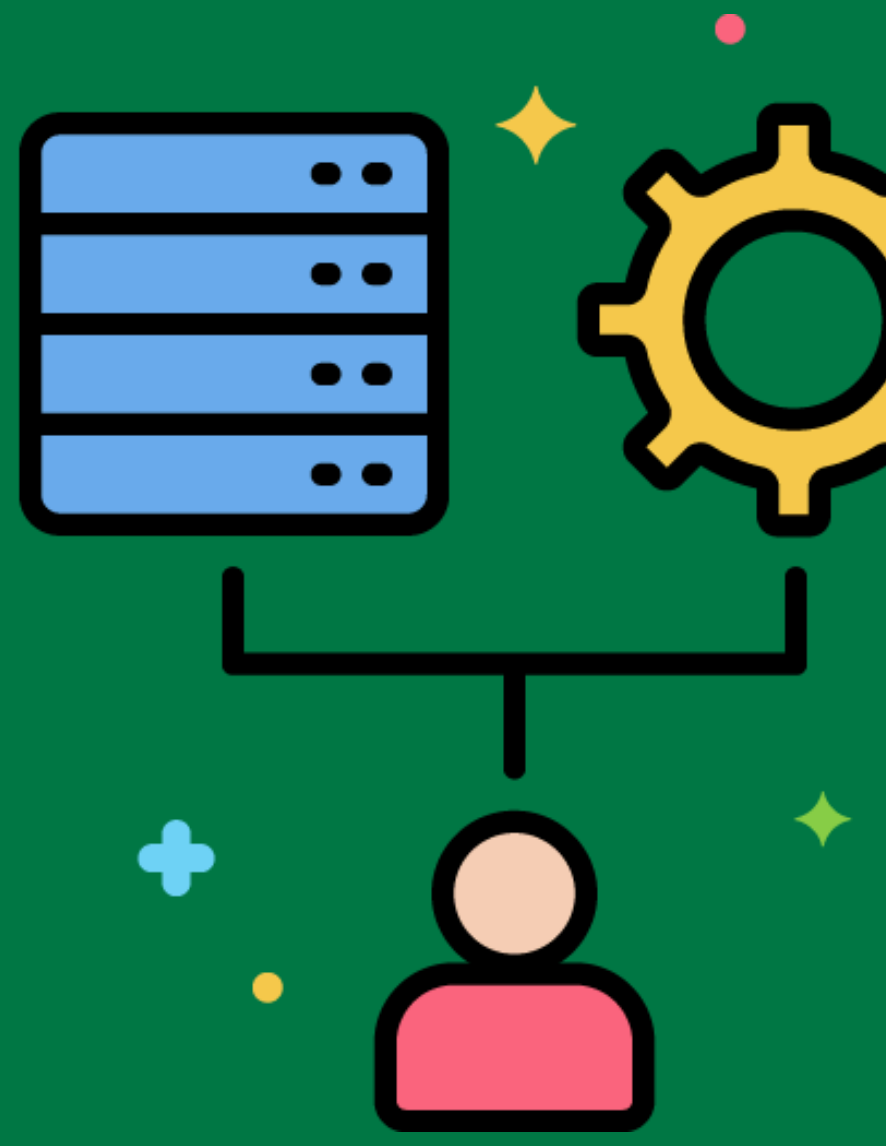
Data Engineering 101 - ETL Terminology

Data Aggregation



The process of summarizing detailed data for analysis and reporting.

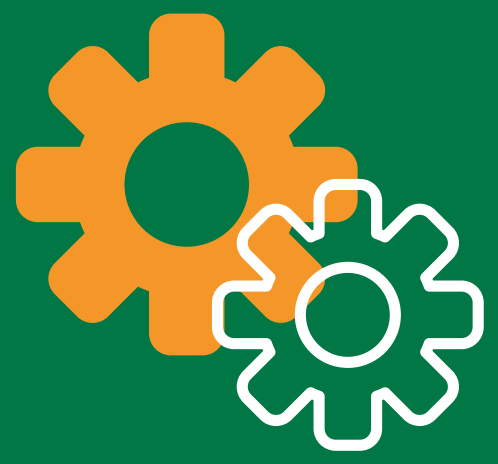
Aggregating daily sales data into monthly sales summaries for trend analysis.



Shwetank Singh
GritSetGrow - GSGLearn.com

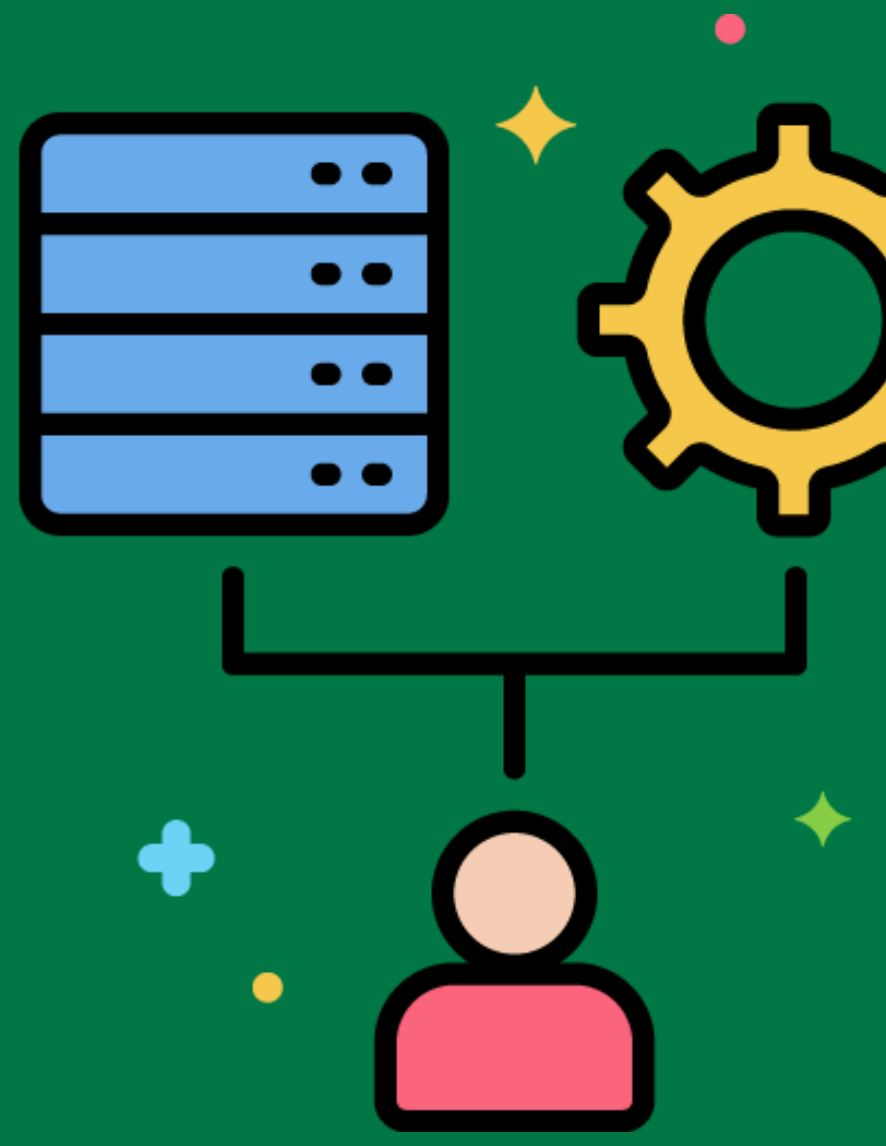
Data Engineering 101 - ETL Terminology

Granularity



The level of detail represented by the data in the data warehouse.

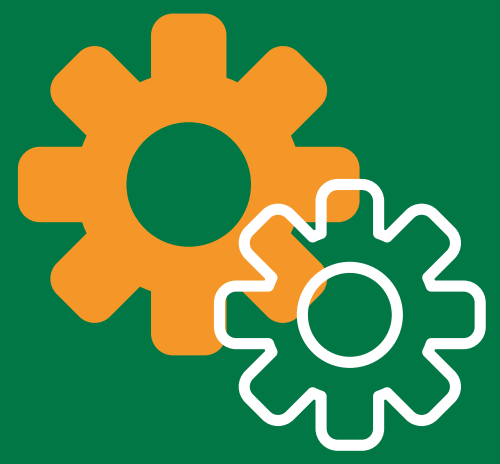
Deciding on the granularity of the sales fact table, such as recording sales at the transaction level versus daily summaries.



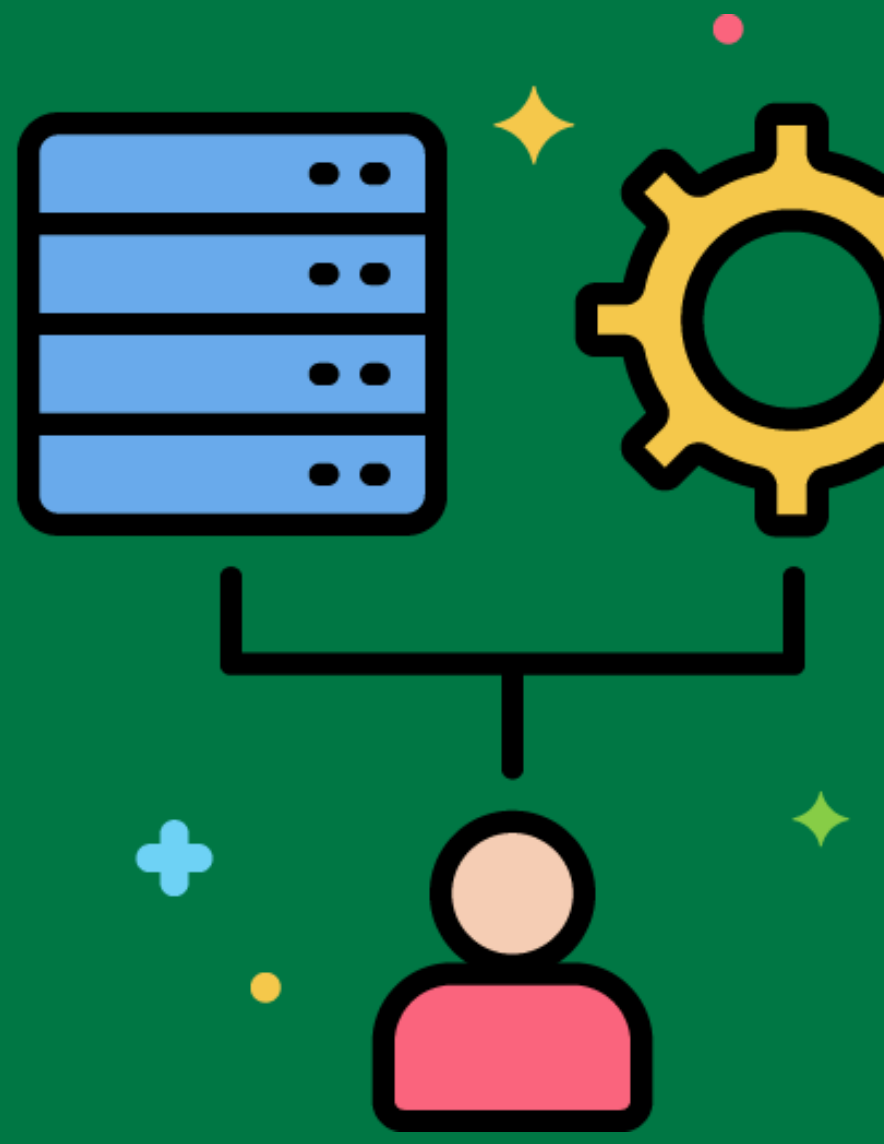
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Historical Data



Data that represents past events and is stored in the data warehouse for analysis.

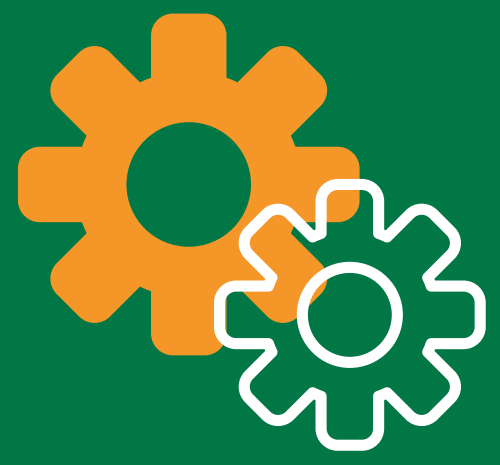


Storing historical sales data for the past ten years to analyze long-term trends.

Shwetank Singh
GritSetGrow - GSGLearn.com

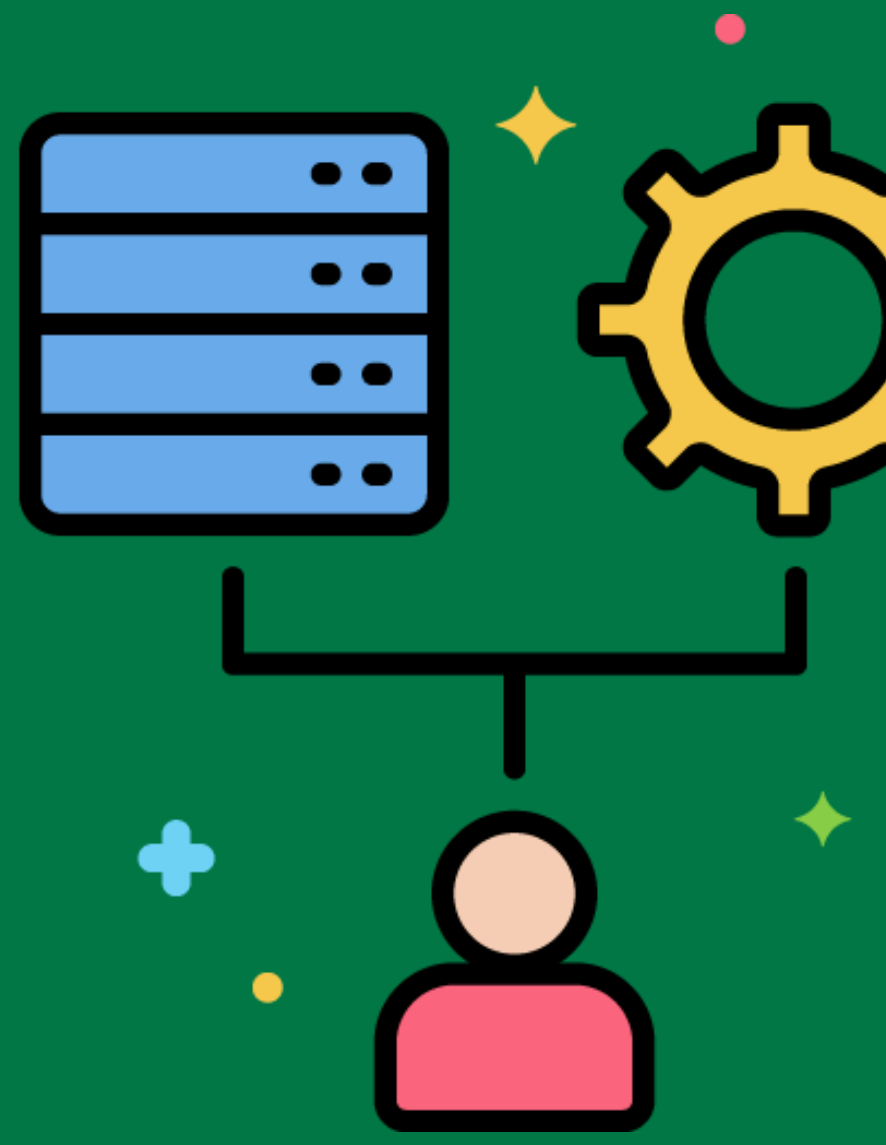
Data Engineering 101 - ETL Terminology

Real-Time Data Warehousing



The process of updating the data warehouse in real-time as new data becomes available.

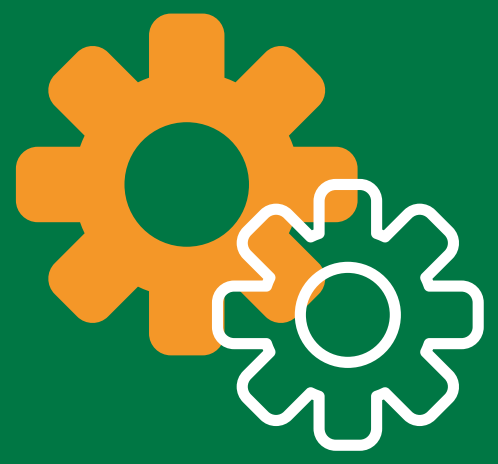
Implementing a real-time ETL process to load streaming data from IoT devices into the data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

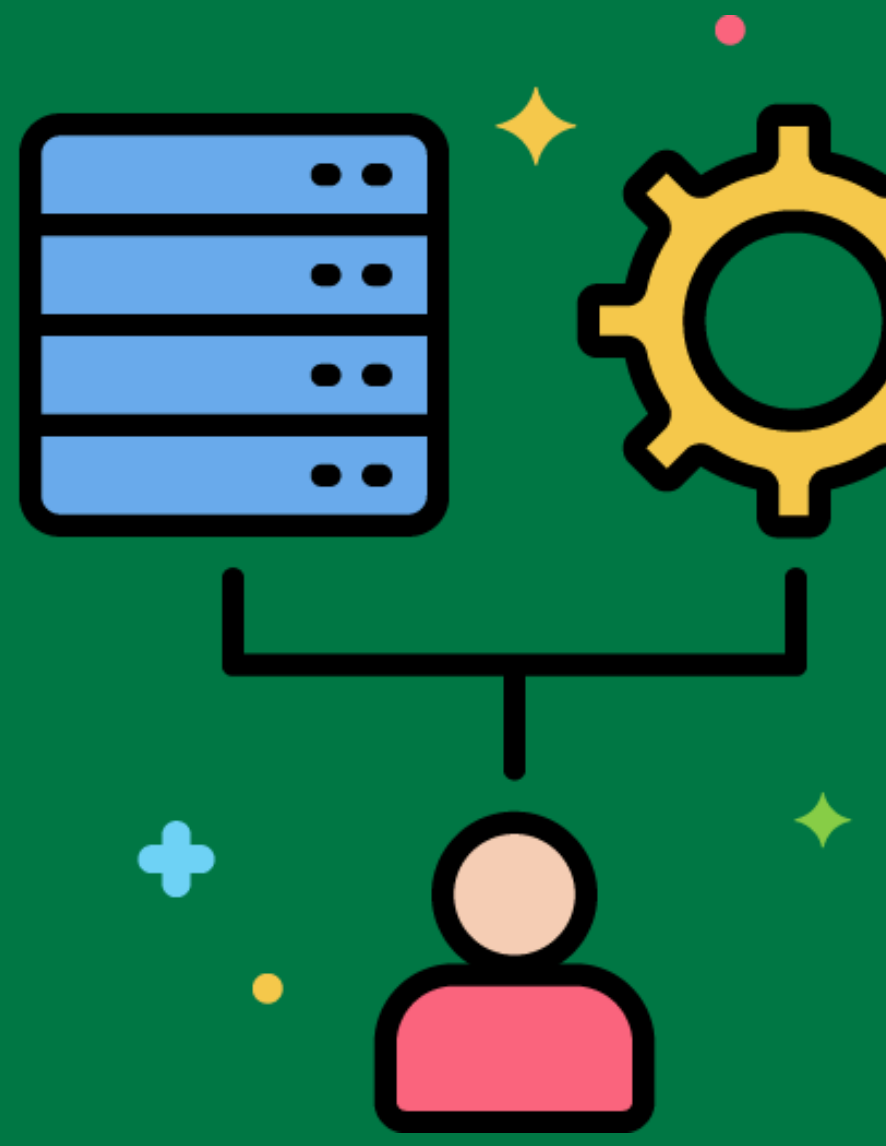
Data Engineering 101 - ETL Terminology

Batch Processing



The processing of data in large groups or batches at scheduled intervals.

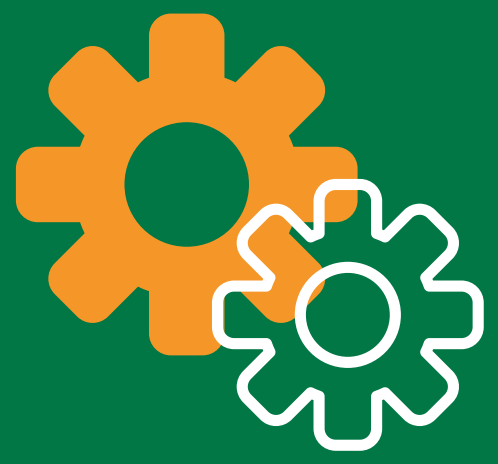
Running a nightly batch process to load the day's sales data into the data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

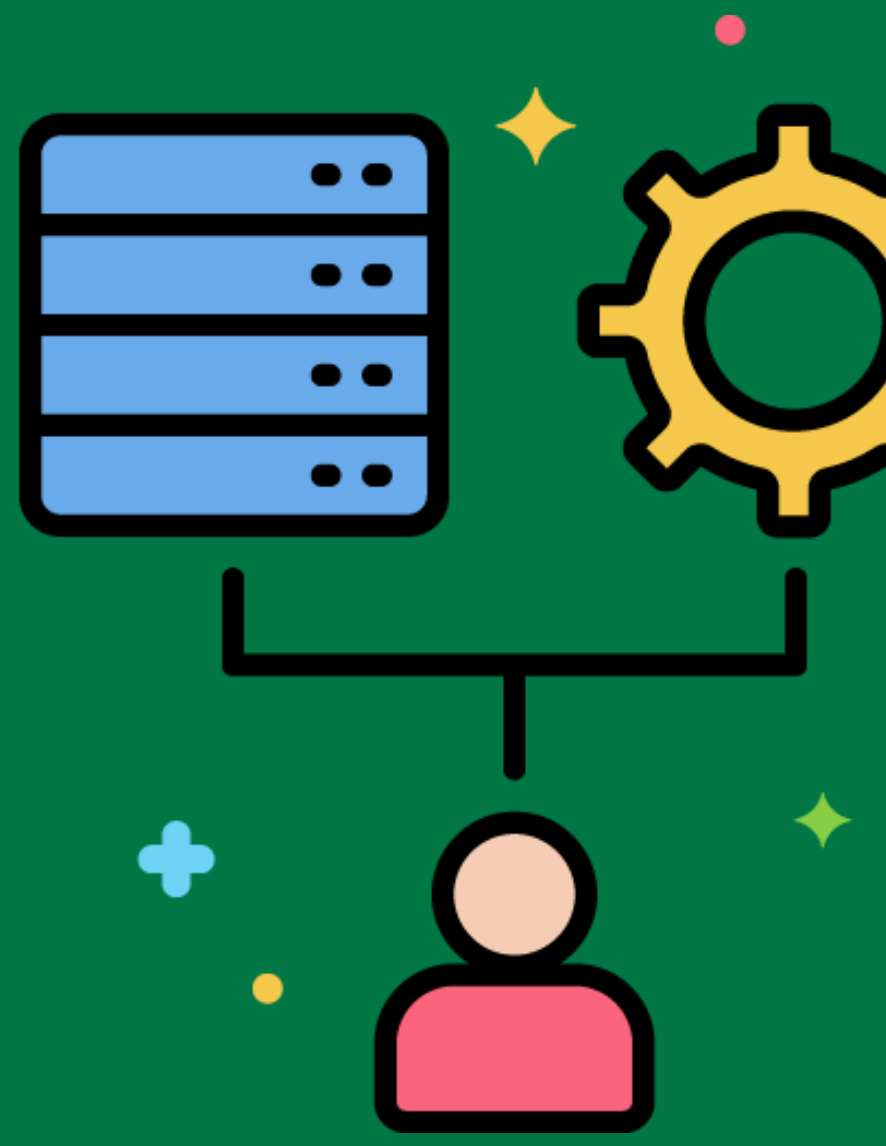
Data Engineering 101 - ETL Terminology

Parallel Processing



The simultaneous processing of multiple tasks to speed up the ETL process.

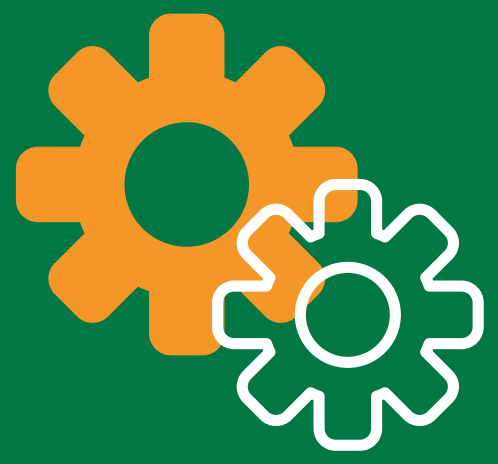
Using parallel processing to transform and load multiple data sources concurrently.



Shwetank Singh
GritSetGrow - GSGLearn.com

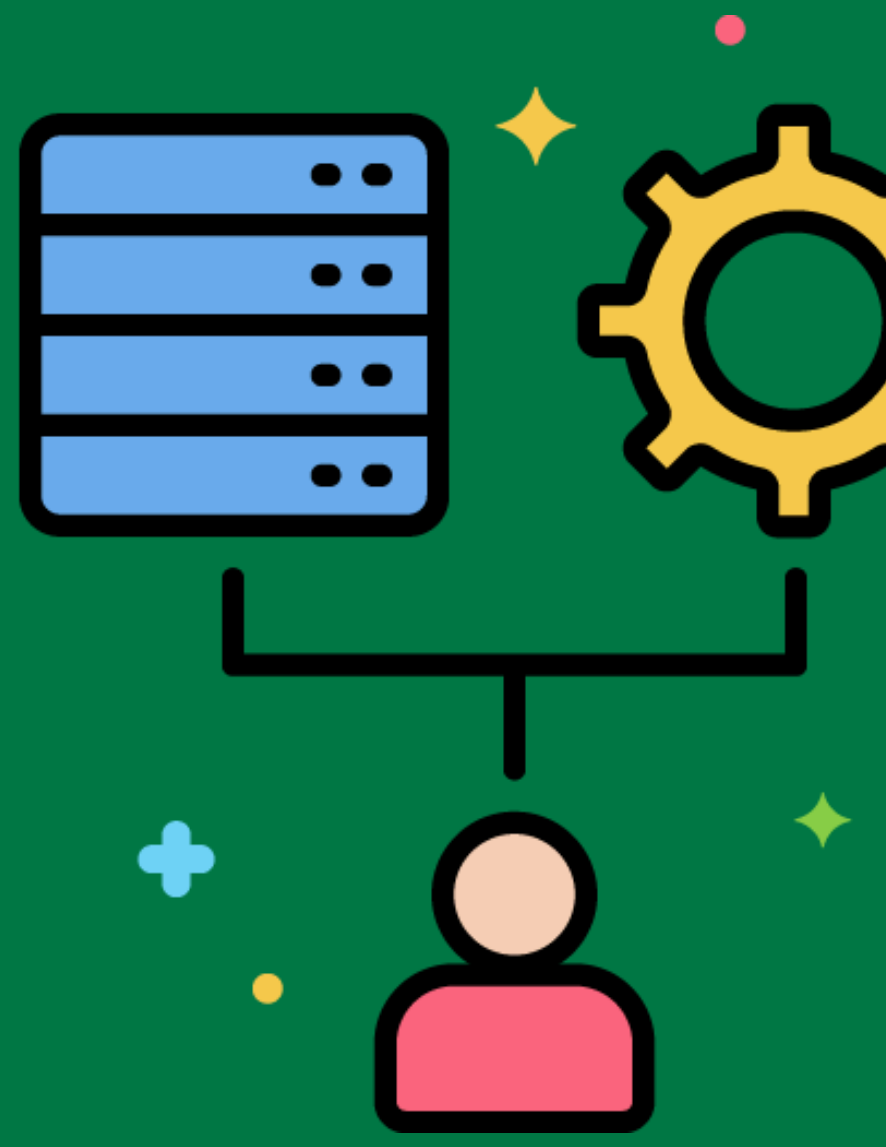
Data Engineering 101 - ETL Terminology

Data Lakehouse



An architecture that combines the benefits of data lakes and data warehouses.

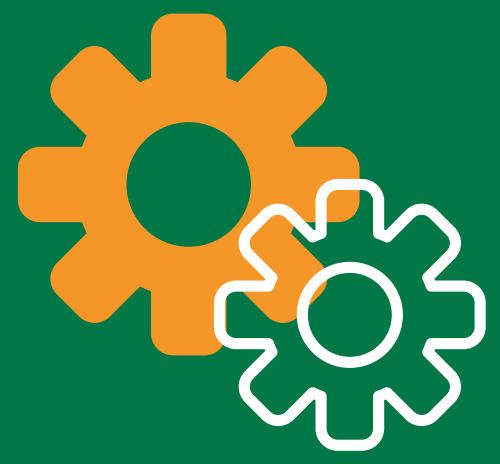
Implementing a data lakehouse to store both structured and unstructured data for comprehensive analysis.



Shwetank Singh
GritSetGrow - GSGLearn.com

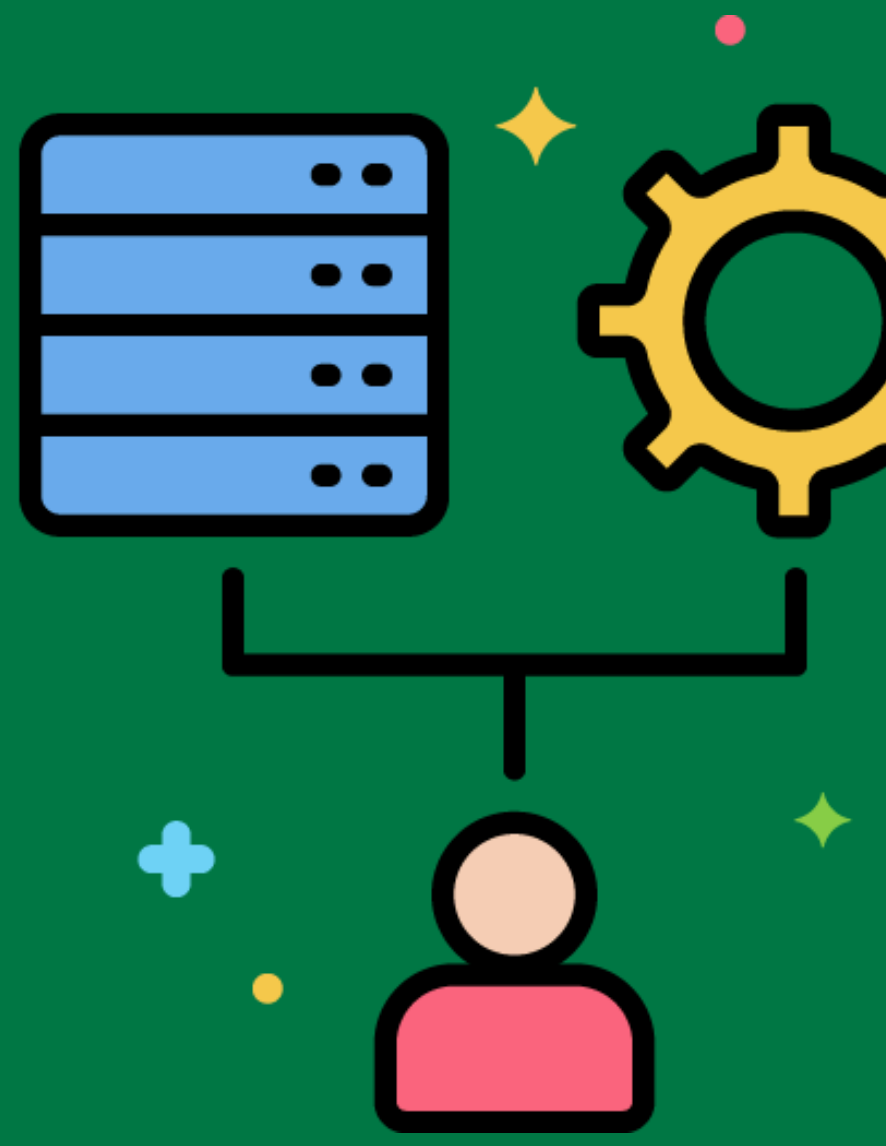
Data Engineering 101 - ETL Terminology

Data Vault Modeling



A database modeling method designed to provide long-term historical storage of data from multiple operational systems.

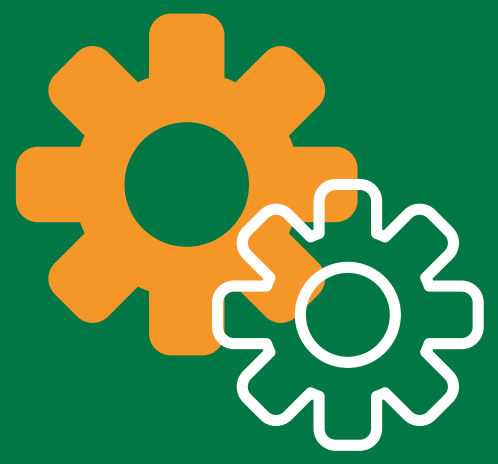
Using data vault modeling to capture and store all historical changes in the data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

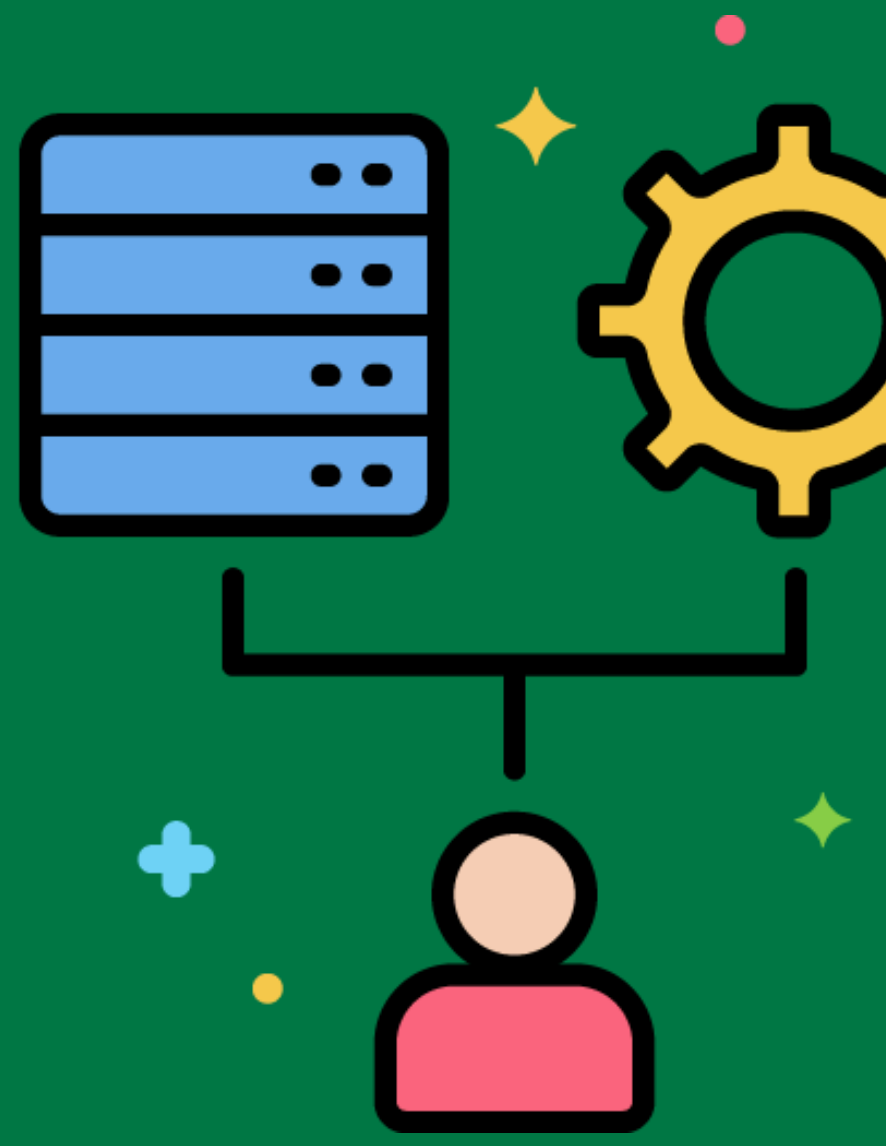
Data Engineering 101 - ETL Terminology

Data Quality Dimensions



Attributes used to measure data quality, including accuracy, completeness, consistency, timeliness, and validity.

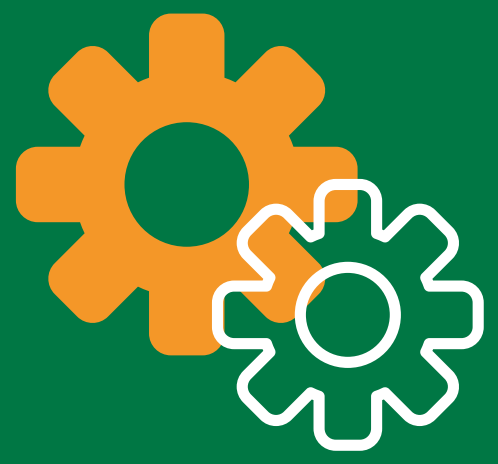
Assessing the accuracy and completeness of customer data in the data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

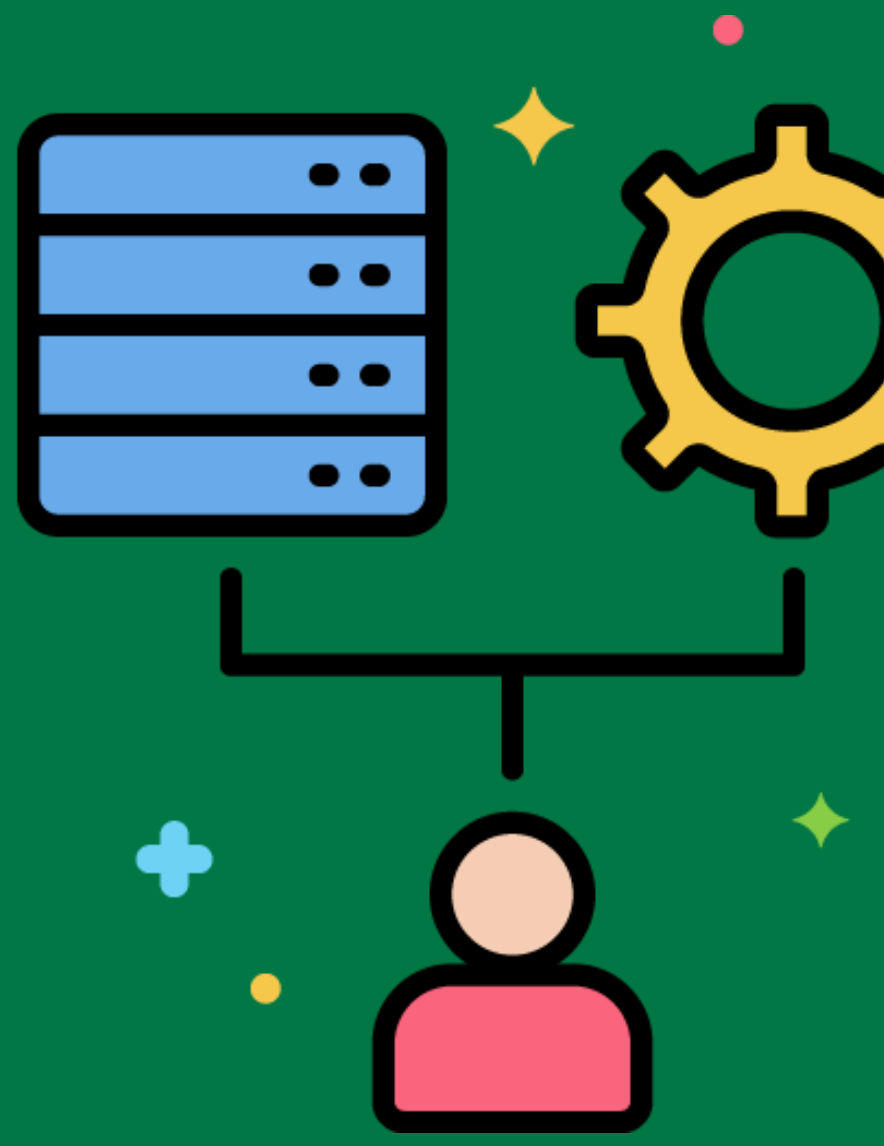
Data Engineering 101 - ETL Terminology

ETL Metadata



Data about the ETL process, including source-to-target mappings, data lineage, and transformation rules.

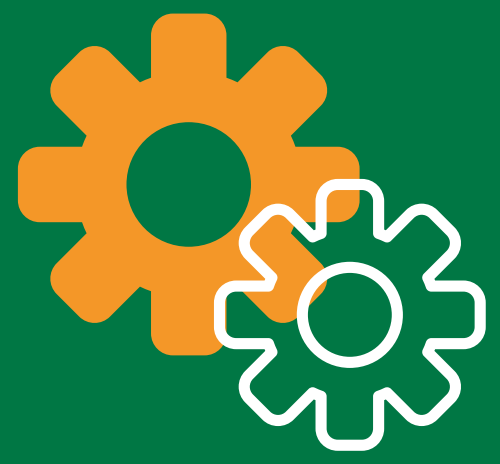
Documenting the transformation rules and source-to-target mappings in the ETL metadata repository.



Shwetank Singh
GritSetGrow - GSGLearn.com

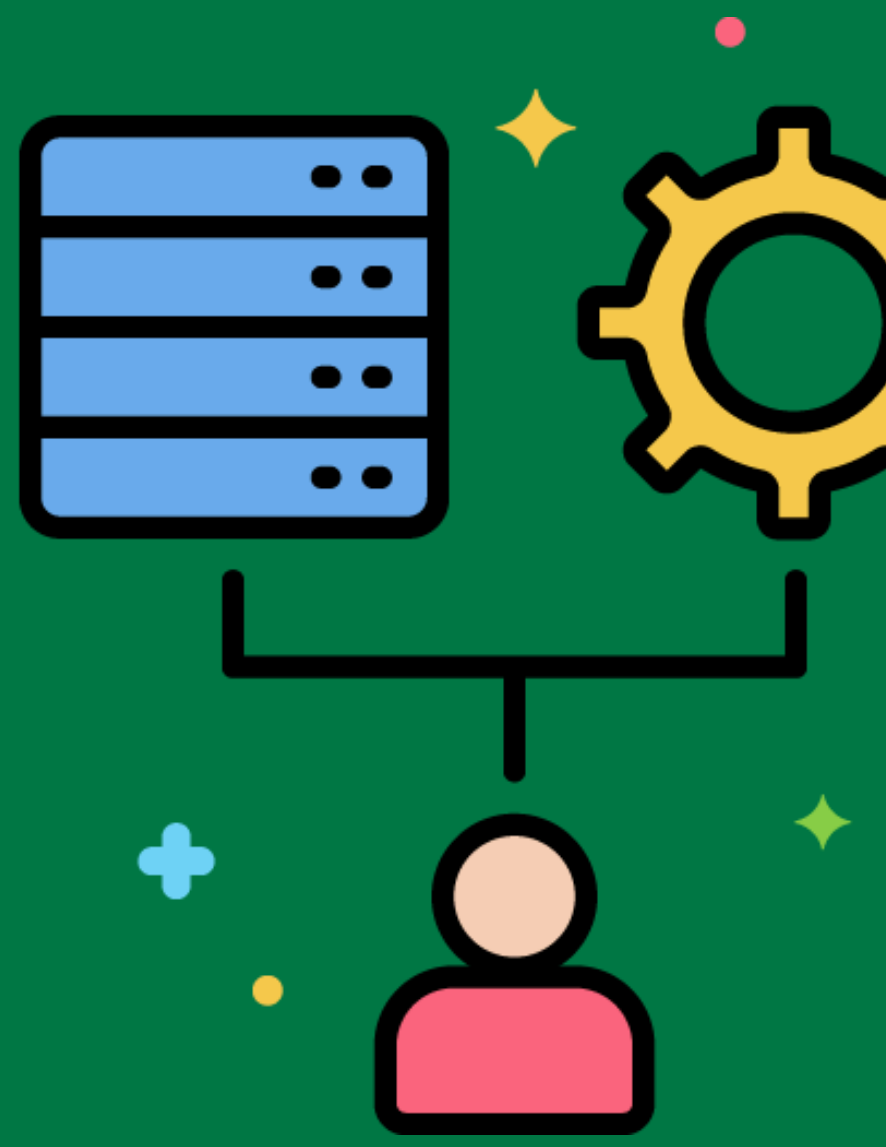
Data Engineering 101 - ETL Terminology

ETL Orchestration



Coordinating and managing the execution of multiple ETL processes.

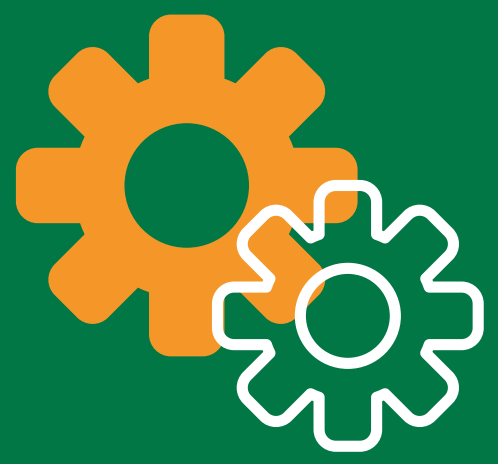
Using an ETL orchestration tool like Apache Airflow to manage the workflow of data extraction, transformation, and loading tasks.



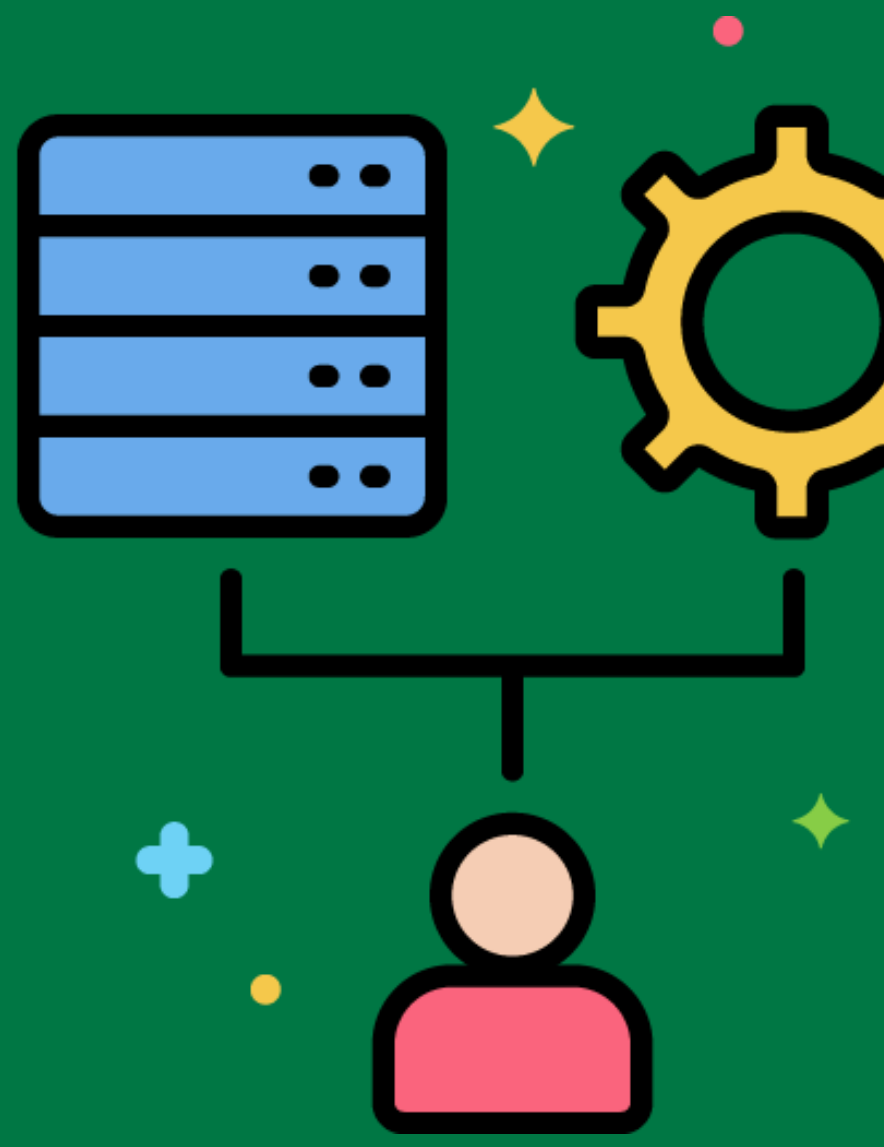
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Integration Patterns



Common design patterns used to integrate data from different sources, such as ETL, ELT, and CDC.

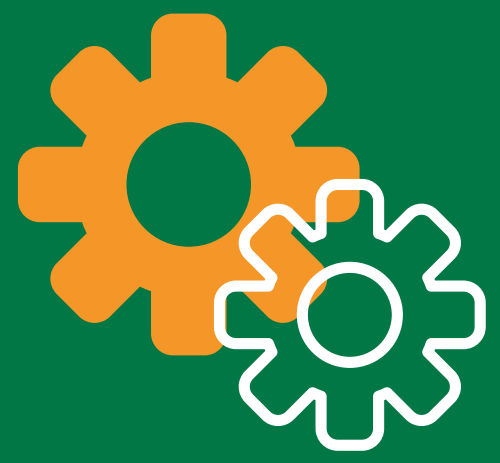


Implementing an ELT pattern where data is first loaded into the data warehouse and then transformed.

Shwetank Singh
GritSetGrow - GSGLearn.com

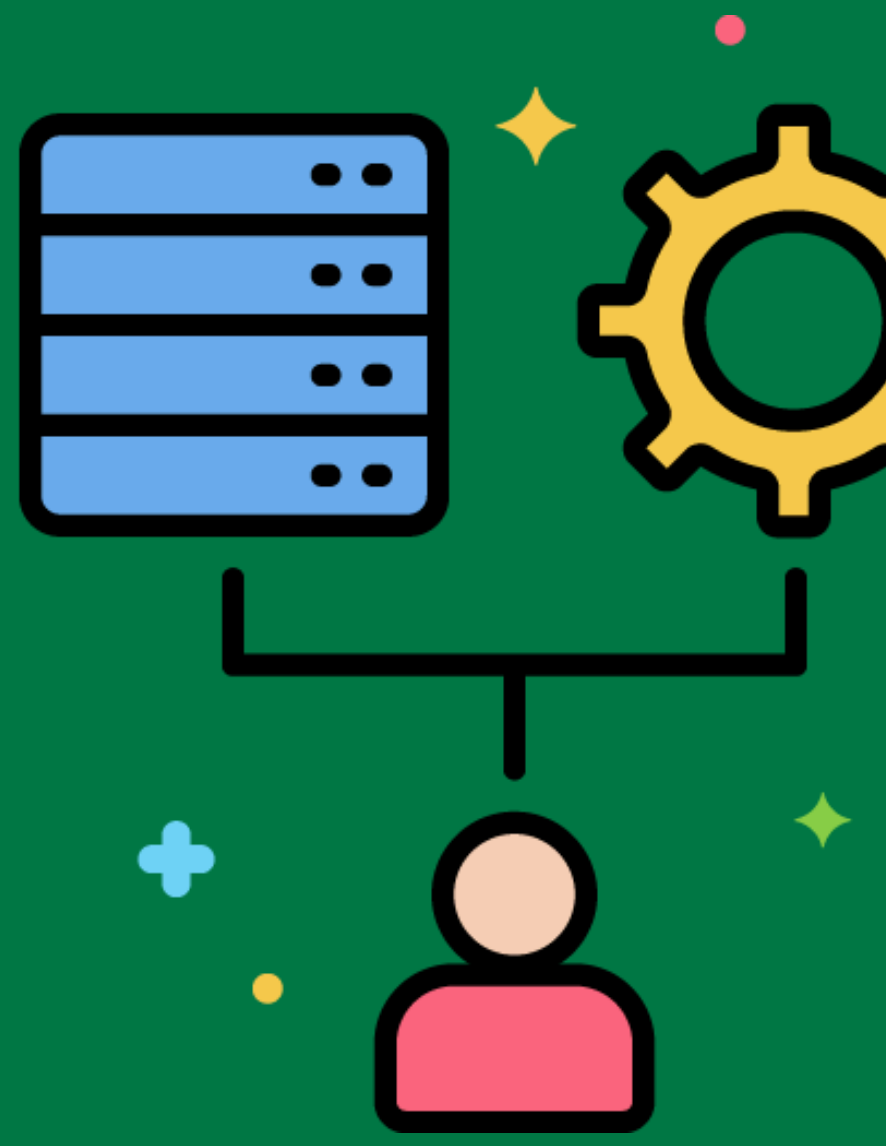
Data Engineering 101 - ETL Terminology

ELT (Extract, Load, Transform)



A data integration pattern where data is first loaded into the data warehouse and then transformed.

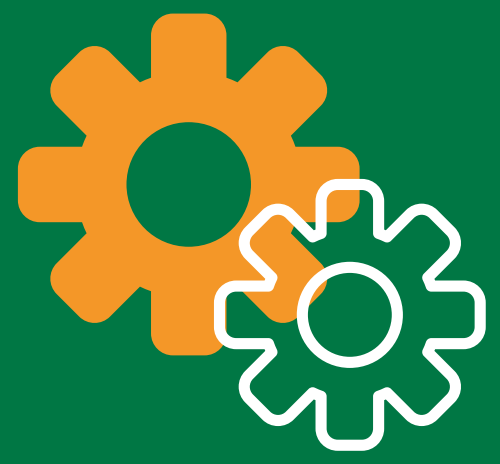
Loading raw sales data into the data warehouse and then performing transformations within the data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

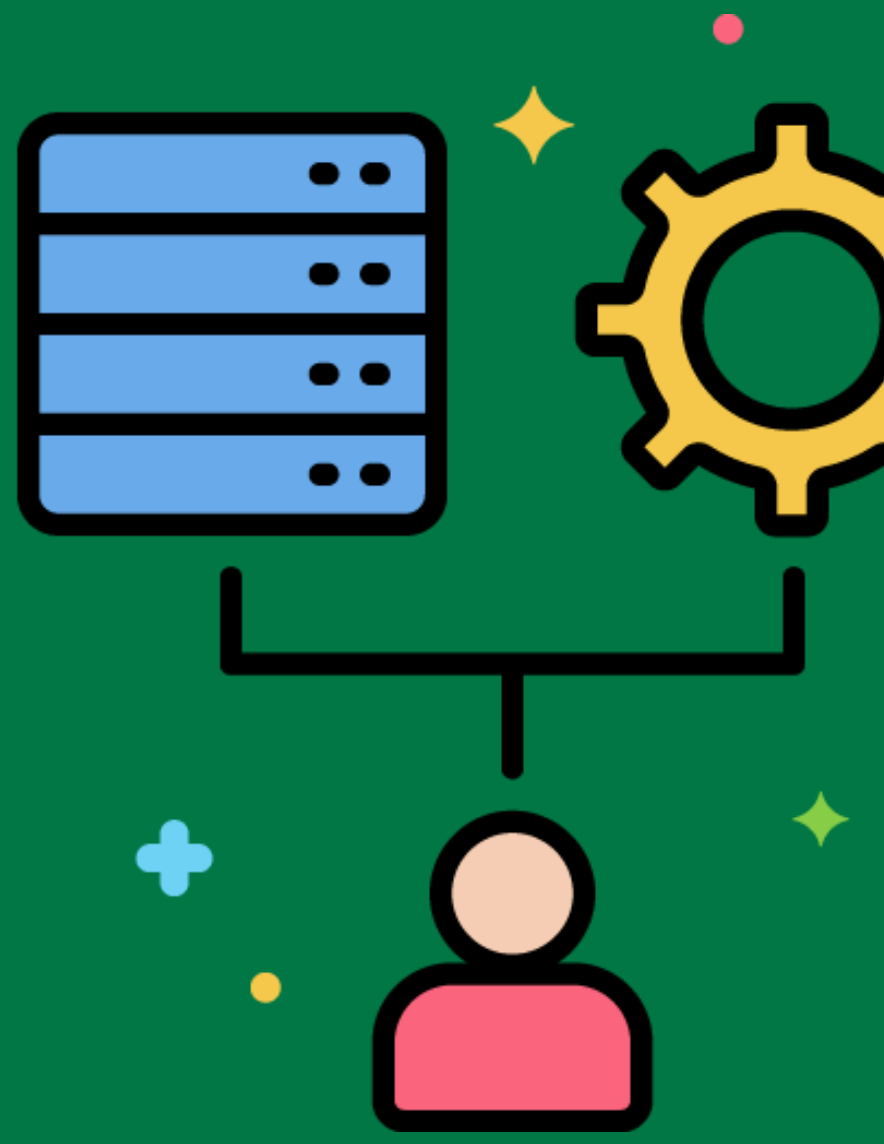
Data Engineering 101 - ETL Terminology

Data Source



The system or location from which data is extracted for the ETL process.

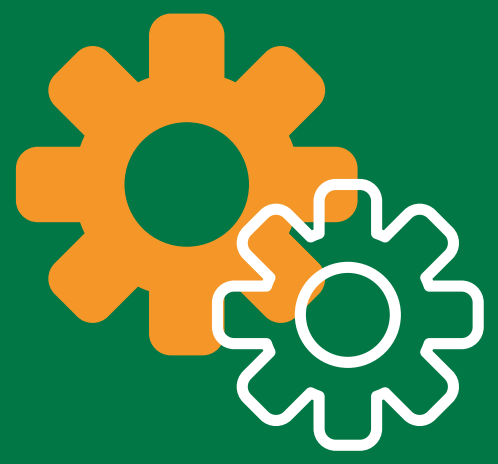
Extracting customer data from a CRM system and sales data from an ERP system.



Shwetank Singh
GritSetGrow - GSGLearn.com

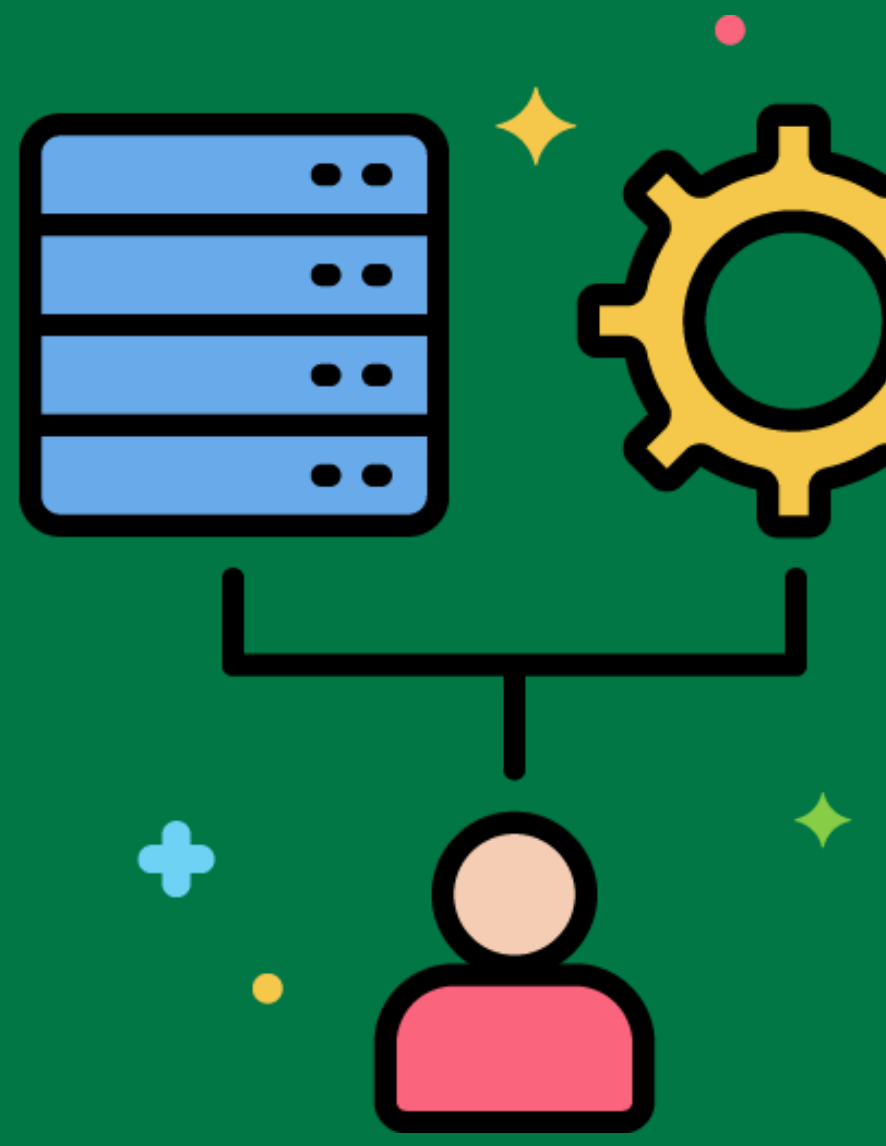
Data Engineering 101 - ETL Terminology

Target System



The system or location where transformed data is loaded during the ETL process.

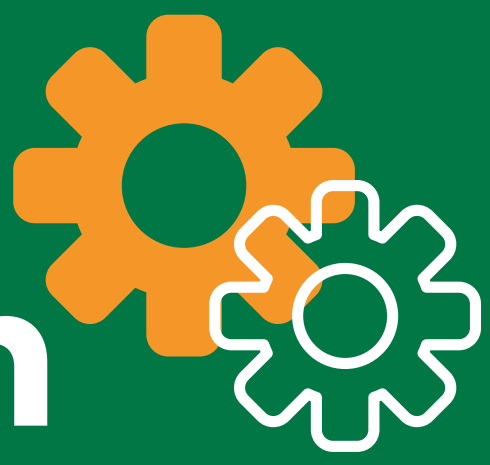
Loading cleaned and transformed data into the data warehouse.



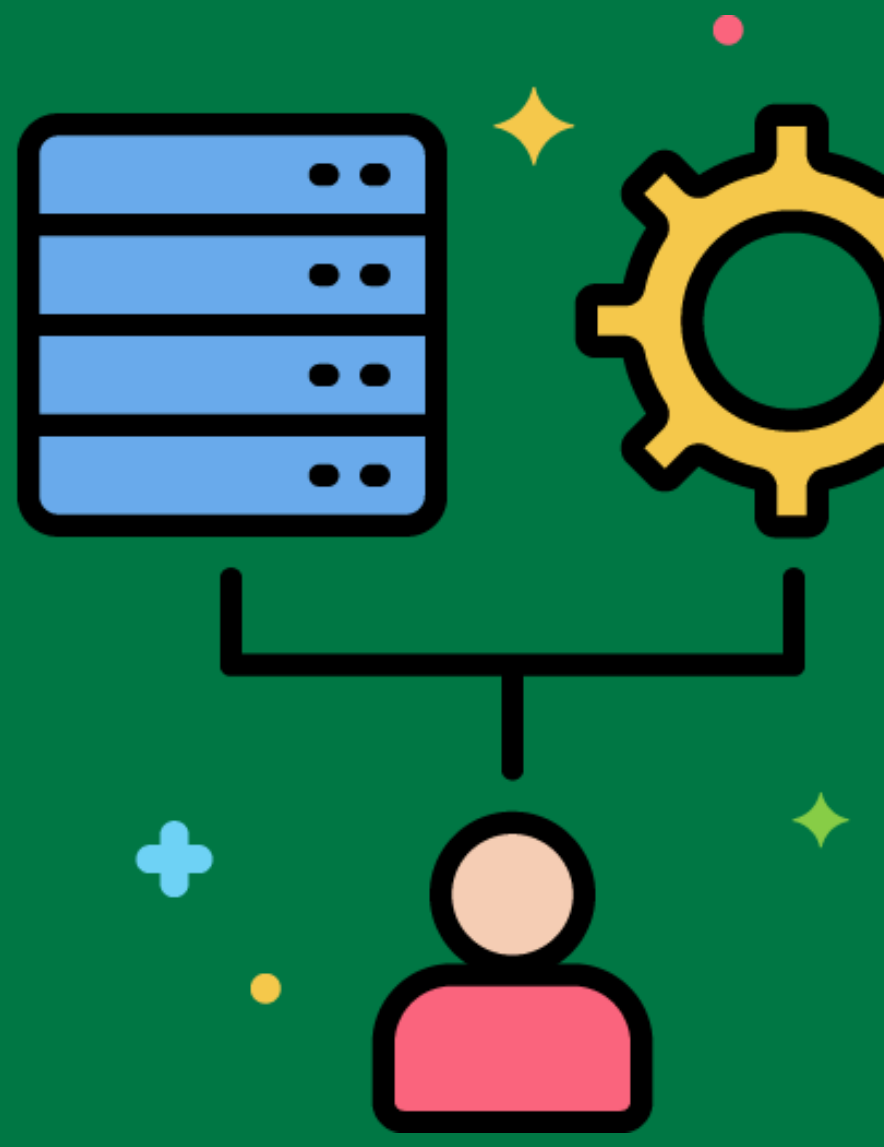
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Synchronization



The process of ensuring that data in different systems is consistent and up-to-date.

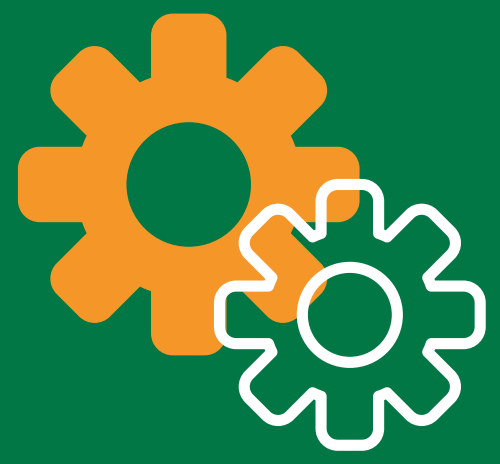


Synchronizing customer data between the CRM system and the data warehouse.

Shwetank Singh
GritSetGrow - GSGLearn.com

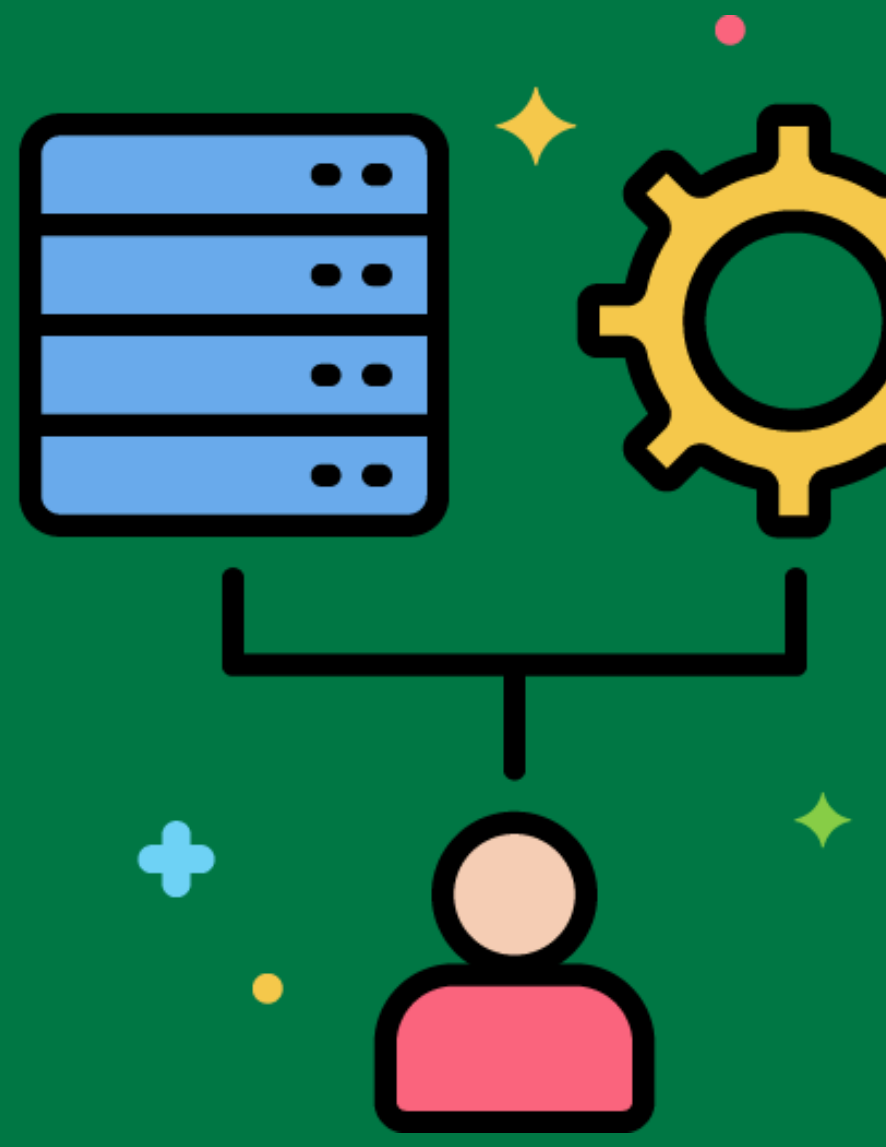
Data Engineering 101 - ETL Terminology

Data Federation



The integration of data from multiple sources into a single, unified view without physically moving the data.

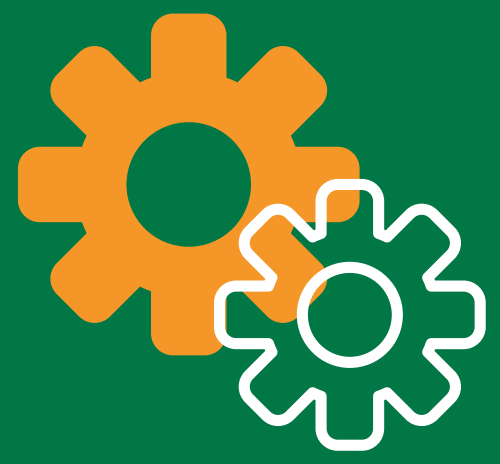
Using a data federation tool to provide a unified view of customer data stored in different systems.



Shwetank Singh
GritSetGrow - GSGLearn.com

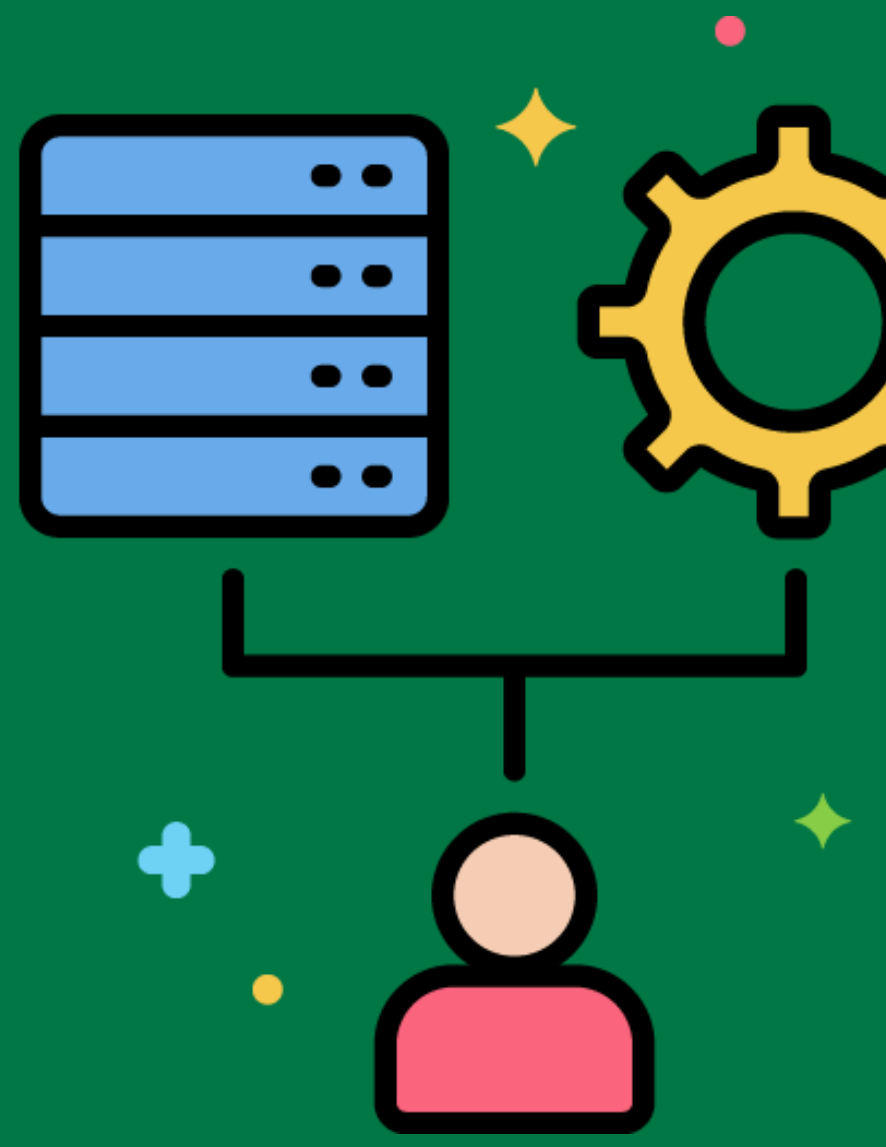
Data Engineering 101 - ETL Terminology

Data Virtualization



The process of creating a virtual layer that provides a unified view of data from multiple sources without moving or replicating the data.

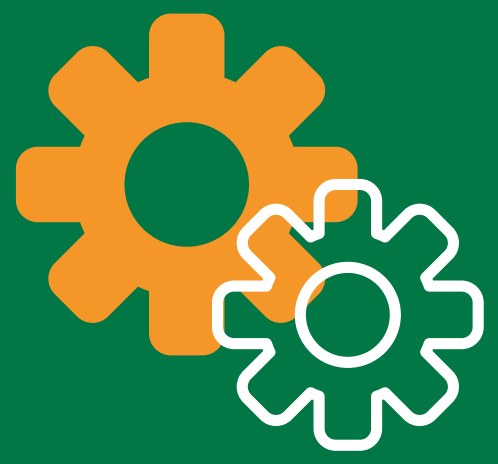
Implementing a data virtualization solution to access and query data from different databases and systems in real-time.



Shwetank Singh
GritSetGrow - GSGLearn.com

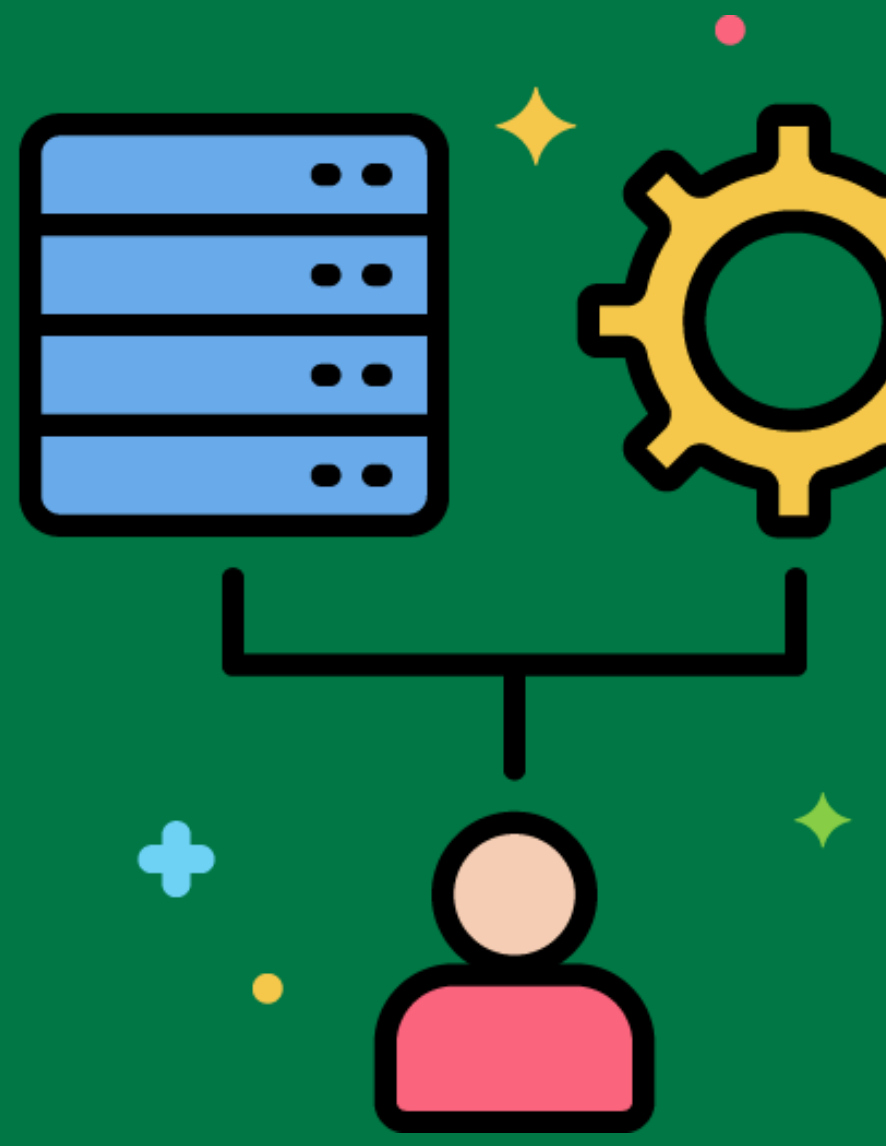
Data Engineering 101 - ETL Terminology

Data Blending



Combining data from different sources to create a new dataset for analysis.

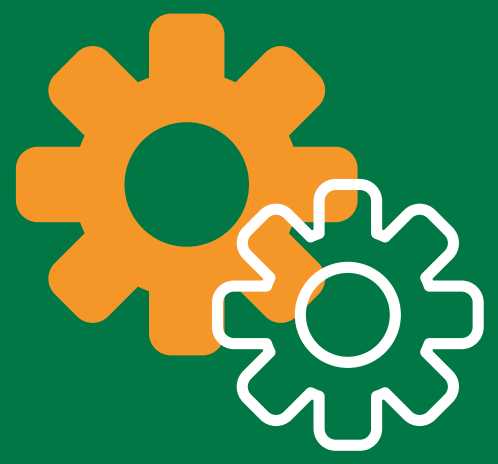
Blending sales data from an ERP system with customer data from a CRM system to analyze customer purchasing behavior.



Shwetank Singh
GritSetGrow - GSGLearn.com

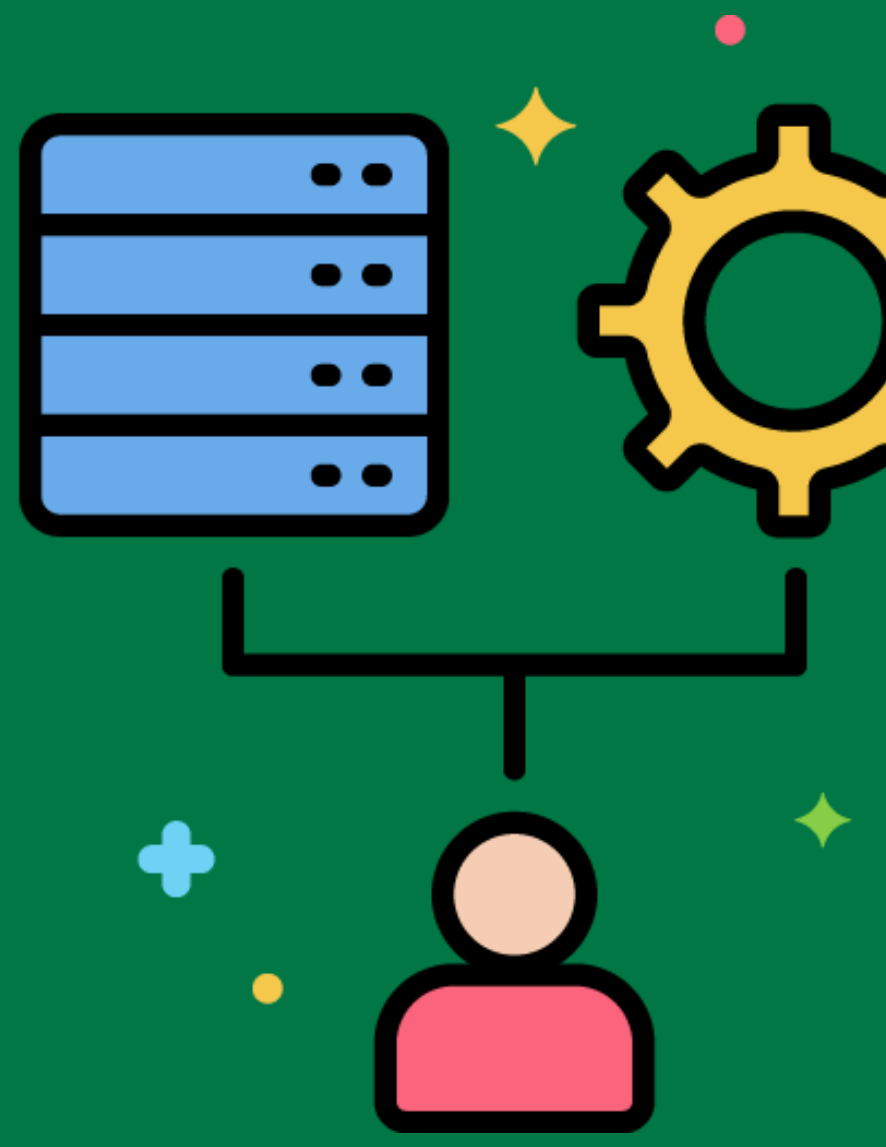
Data Engineering 101 - ETL Terminology

ETL Pipeline



A series of processes that extract, transform, and load data from source systems to the data warehouse.

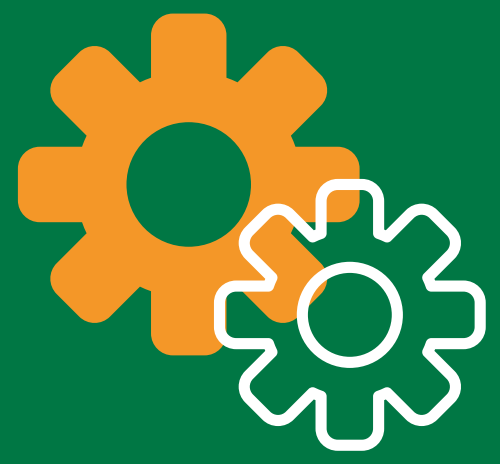
Designing an ETL pipeline that extracts data from various sources, applies business rules, and loads the data into the data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

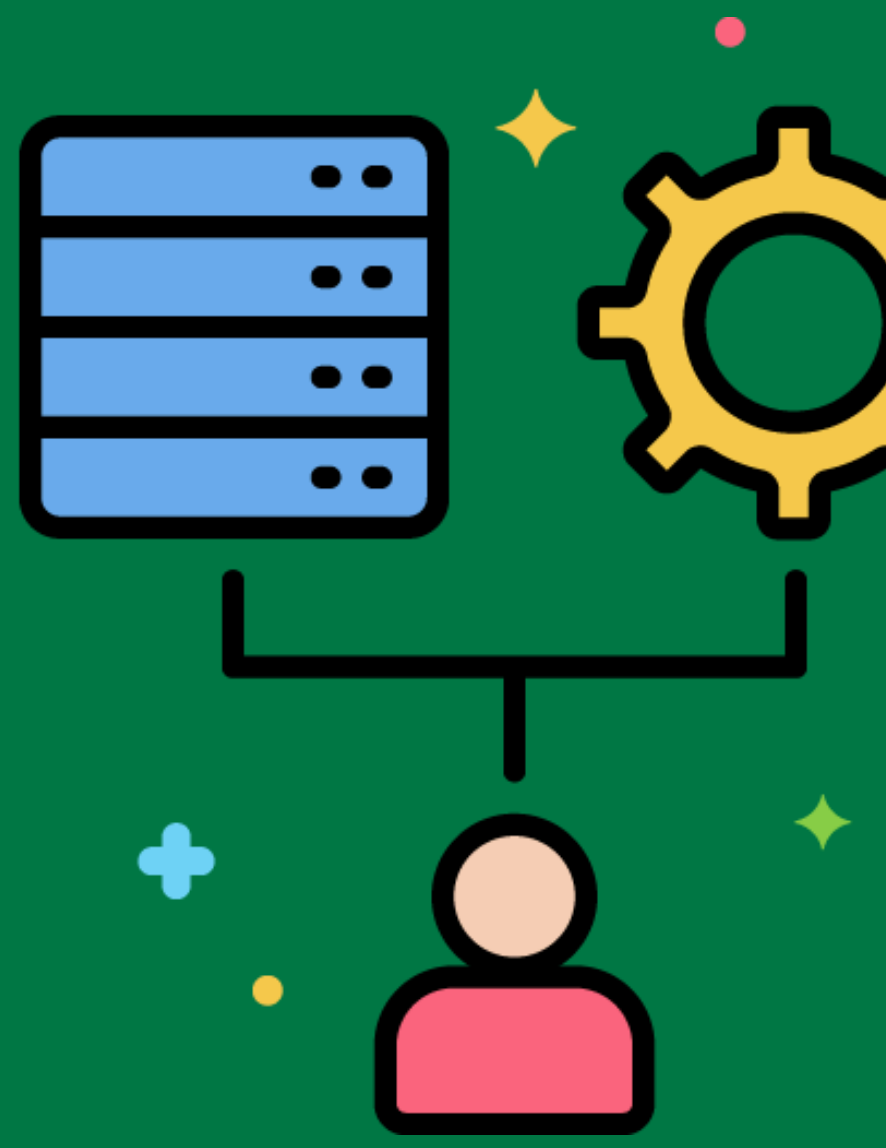
Data Engineering 101 - ETL Terminology

ETL Workflow



The sequence and flow of tasks involved in the ETL process.

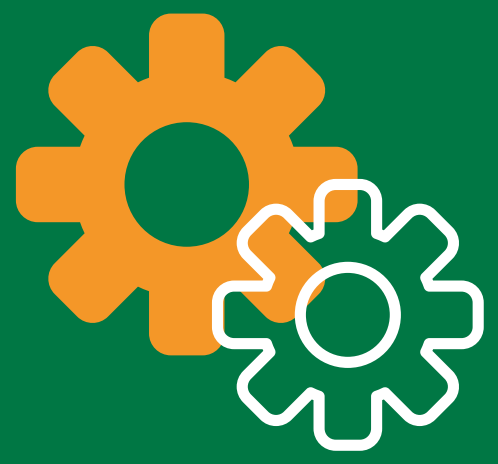
Creating an ETL workflow that includes tasks for data extraction, data transformation, and data loading.



Shwetank Singh
GritSetGrow - GSGLearn.com

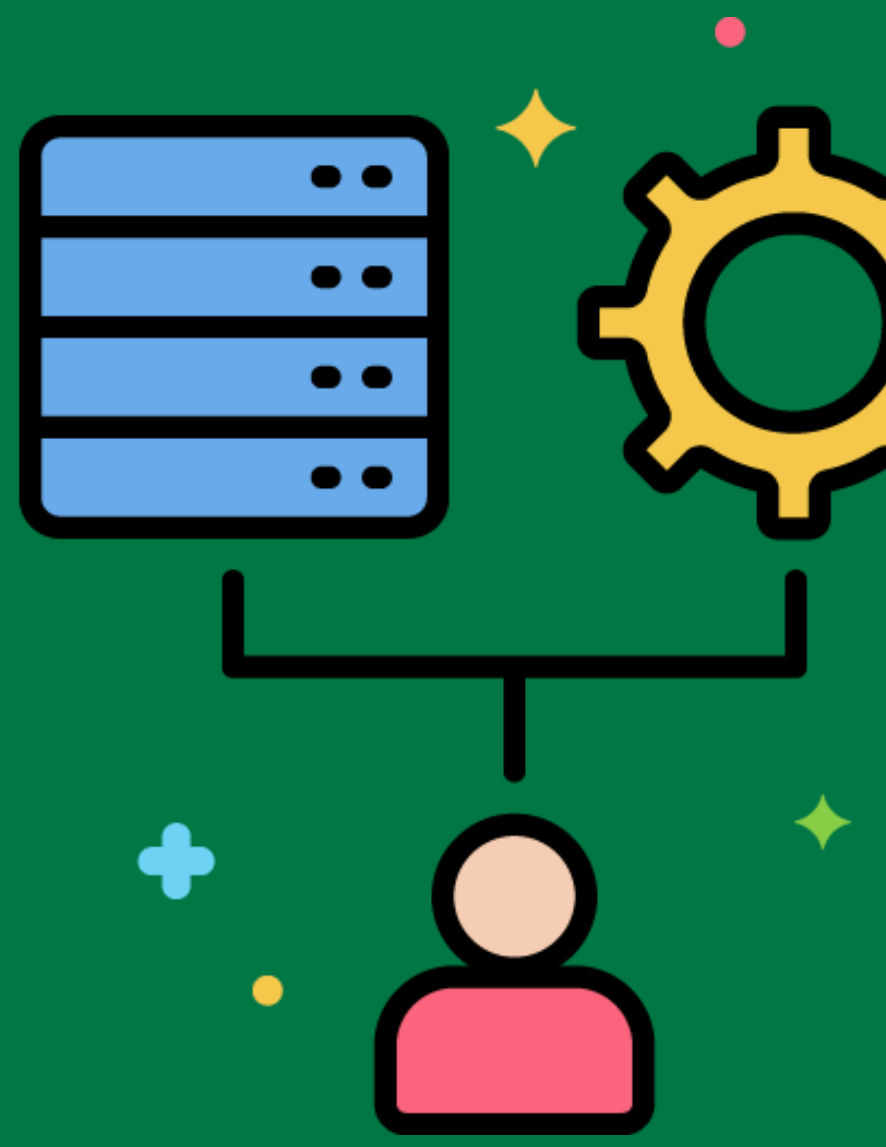
Data Engineering 101 - ETL Terminology

ETL Scheduler



A tool that manages the timing and execution of ETL processes.

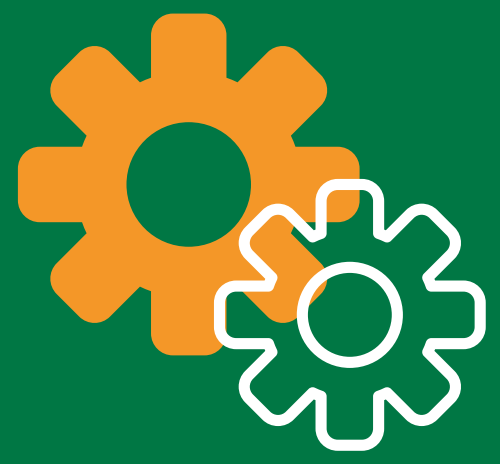
Using an ETL scheduler to run data extraction and loading jobs at specified times.



Shwetank Singh
GritSetGrow - GSGLearn.com

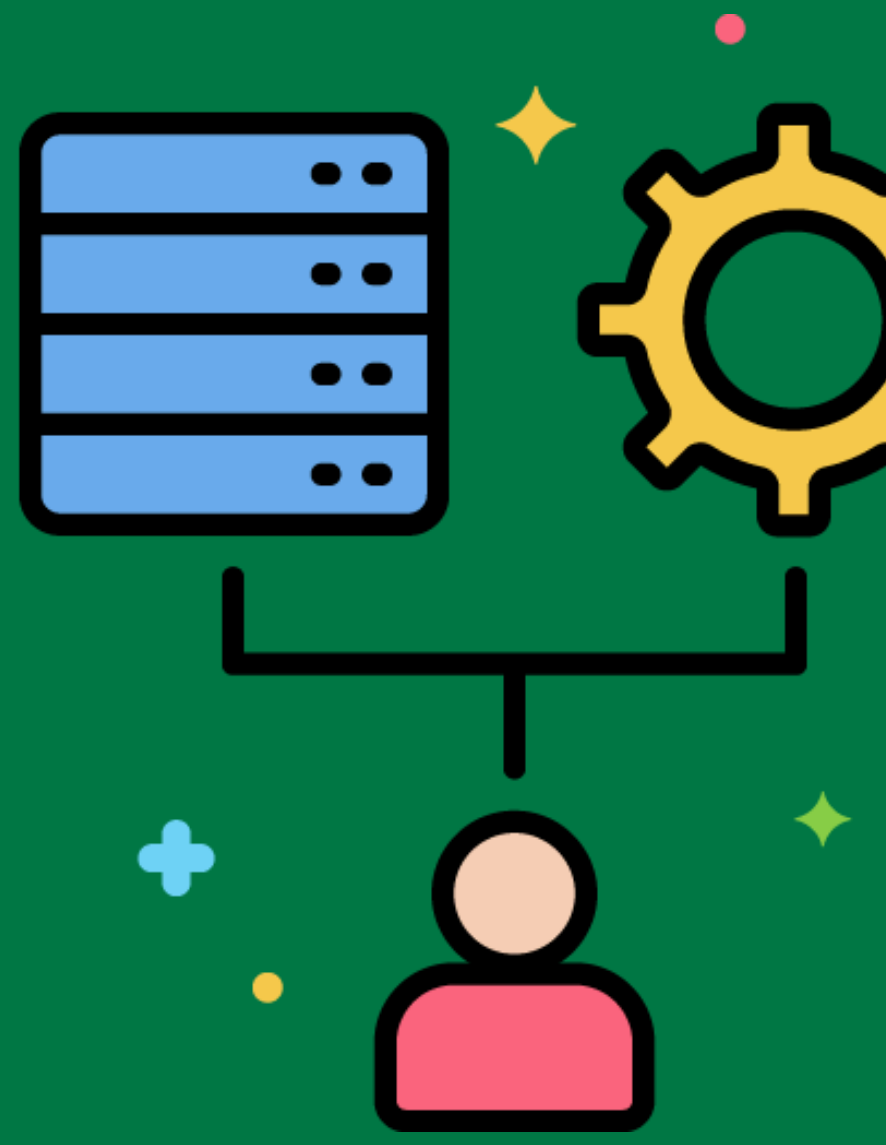
Data Engineering 101 - ETL Terminology

ETL Job



A single task or unit of work in the ETL process.

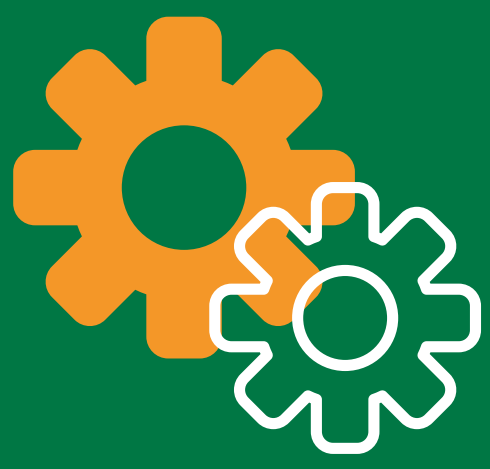
Creating an ETL job to extract customer data from the CRM system.



Shwetank Singh
GritSetGrow - GSGLearn.com

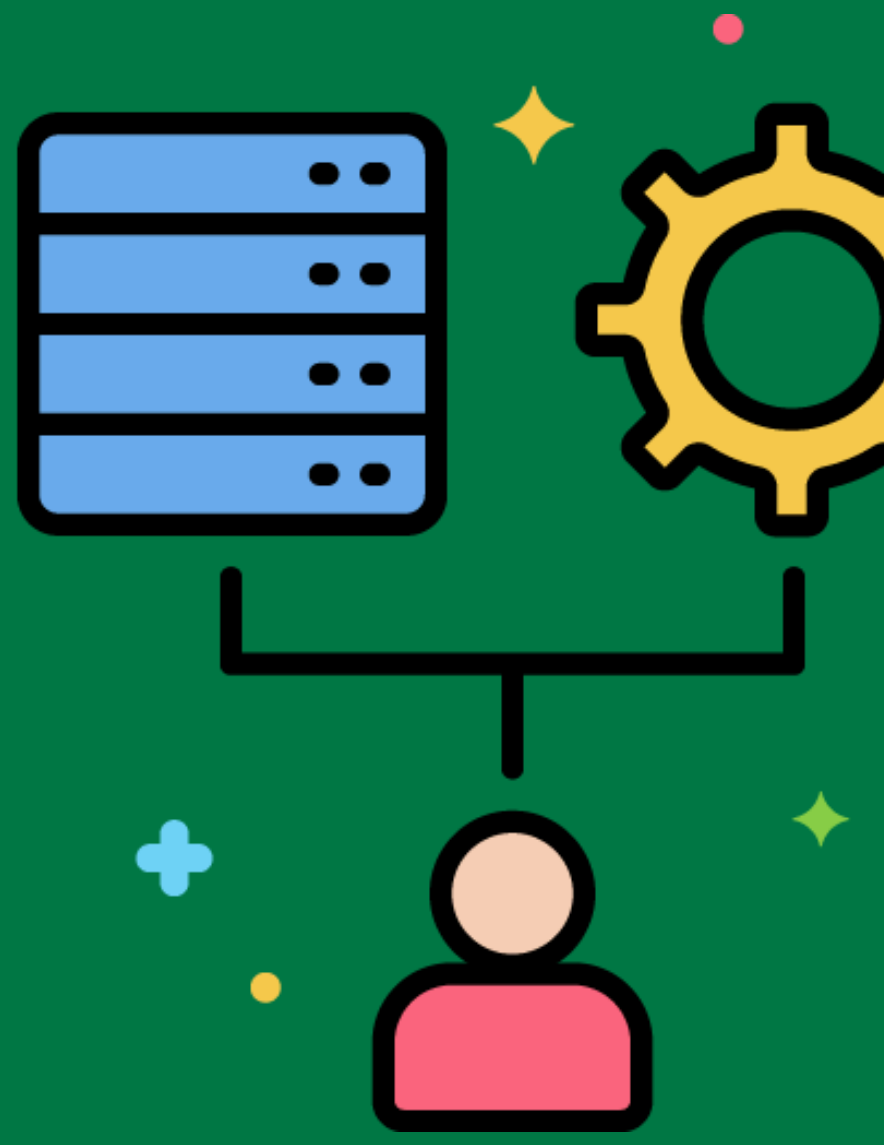
Data Engineering 101 - ETL Terminology

Data Transformation Rules



Specific rules and logic applied to data during the transformation process.

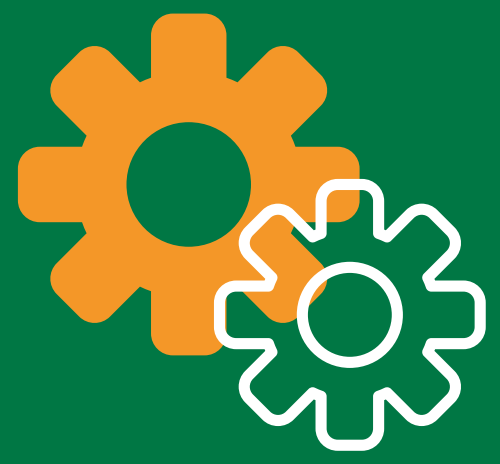
Defining transformation rules to convert all product prices to a common currency.



Shwetank Singh
GritSetGrow - GSGLearn.com

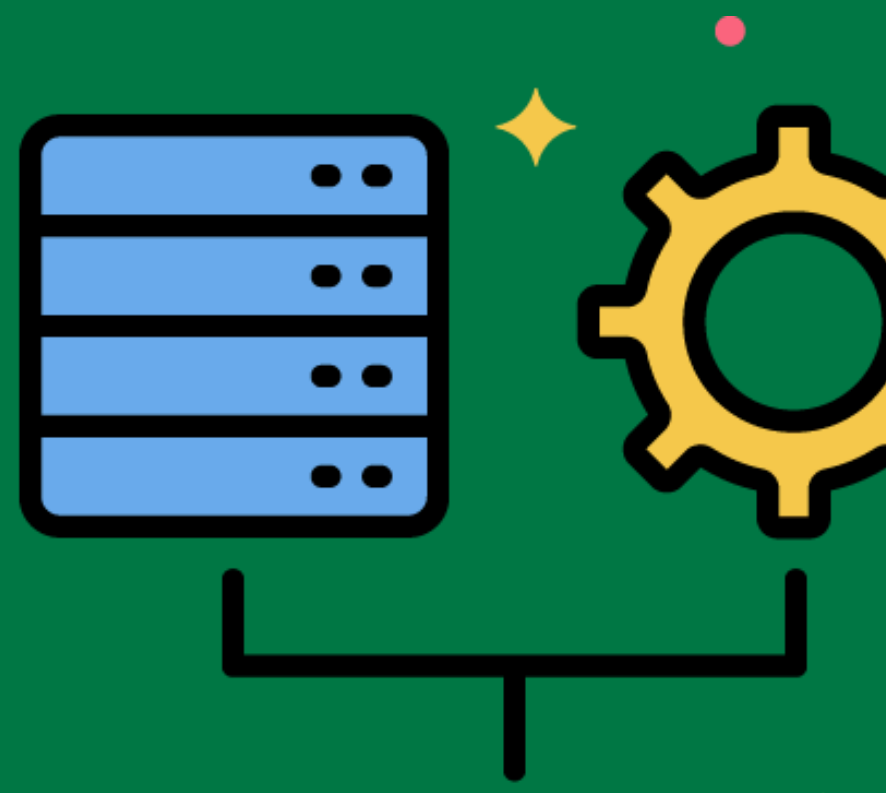
Data Engineering 101 - ETL Terminology

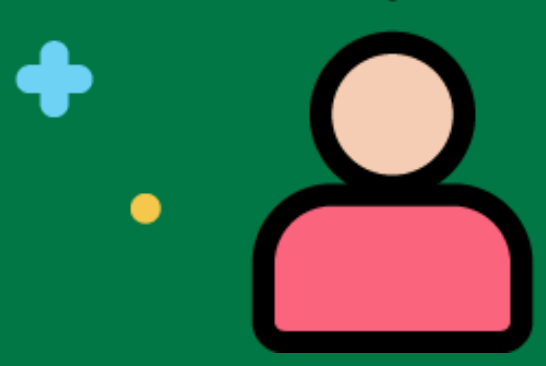
Data Quality Rules



Specific criteria and checks to ensure data quality during the ETL process.

Implementing data quality rules to check for missing values and incorrect data types in customer records.



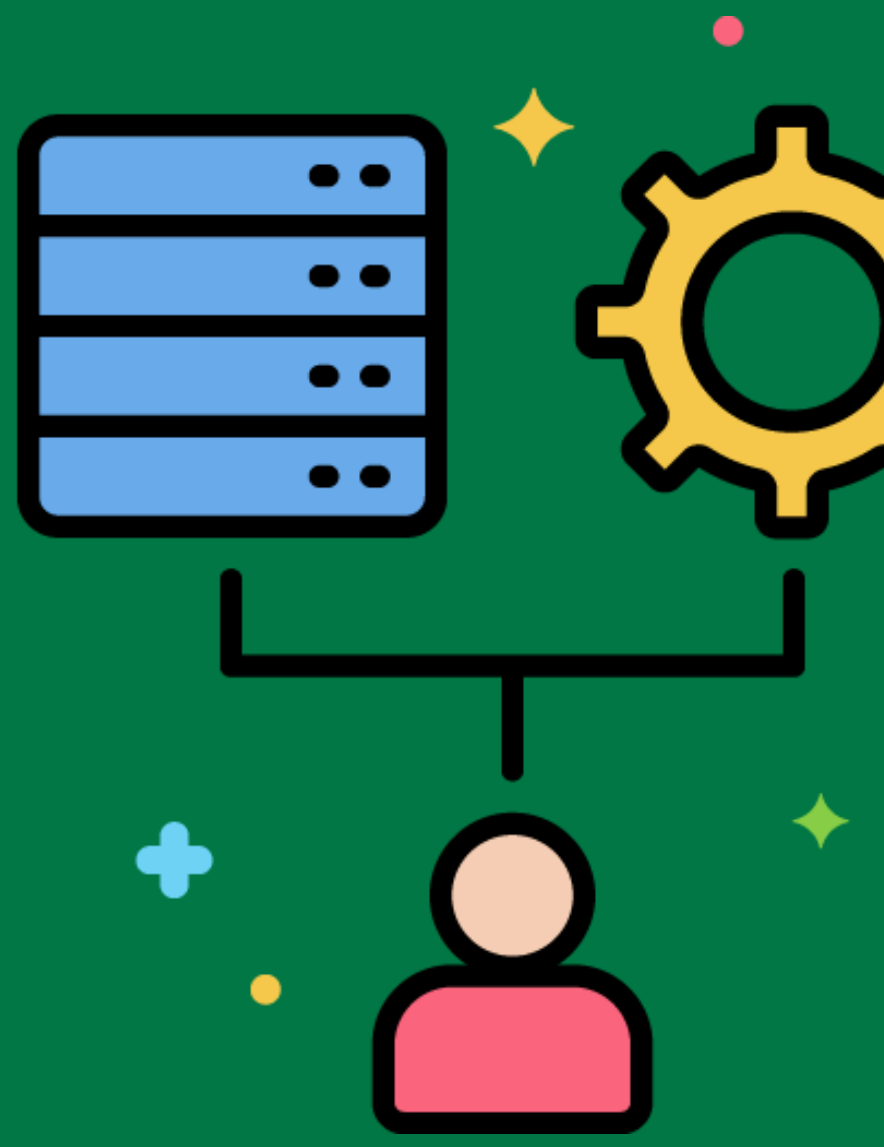


Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Functions

Functions and operations used to transform data during the ETL process.

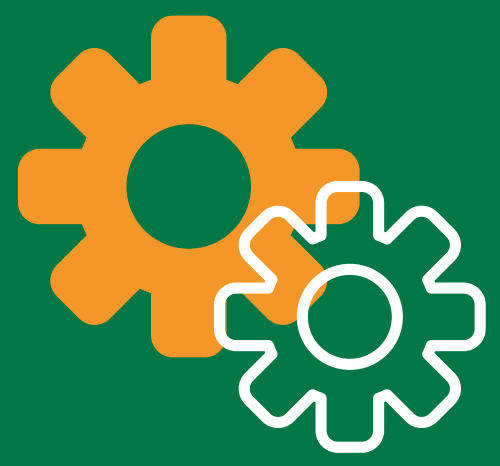


Using a transformation function to concatenate first name and last name into a full name field.

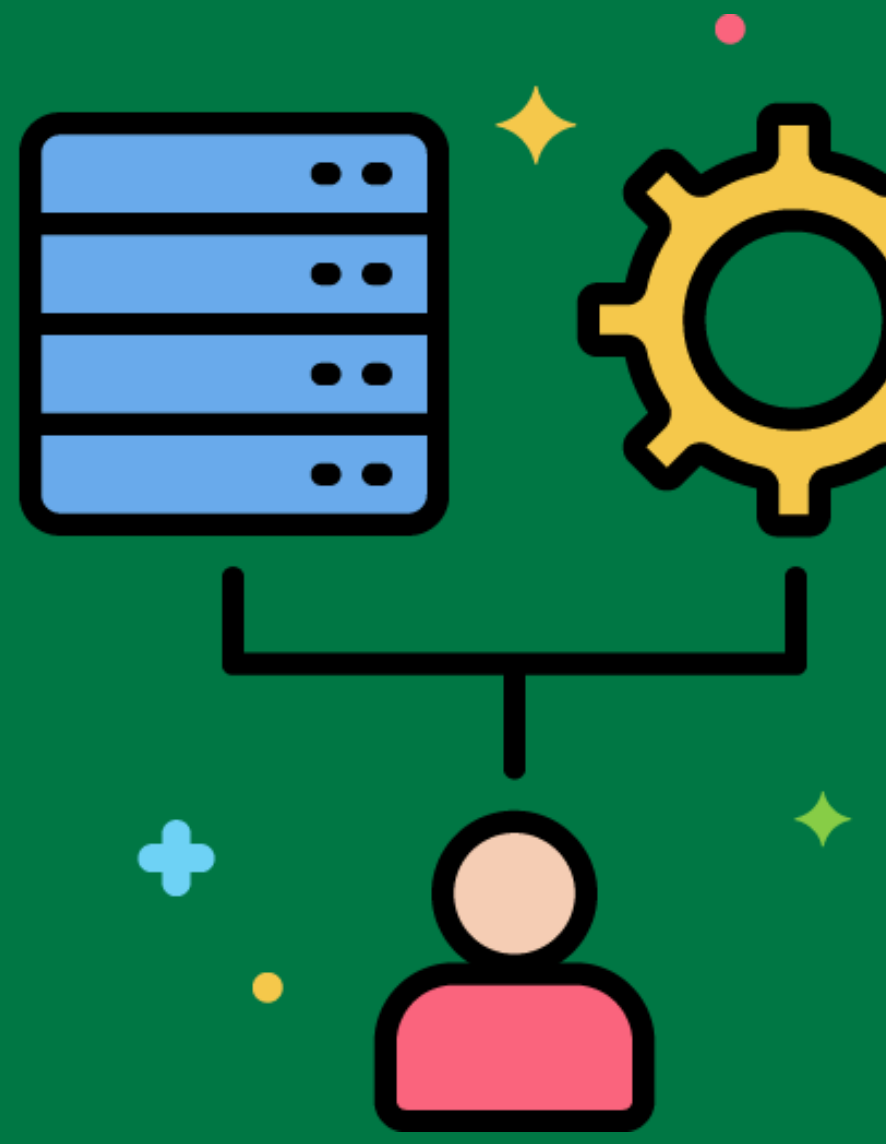
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Integration Strategy



A plan for combining data from different sources to provide a unified view.

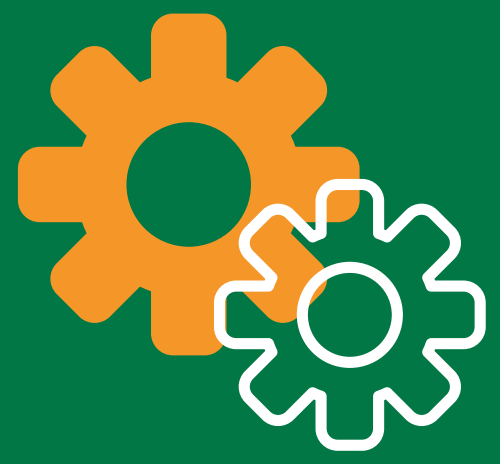


Developing a data integration strategy to consolidate customer data from multiple systems into a single customer profile.

Shwetank Singh
GritSetGrow - GSGLearn.com

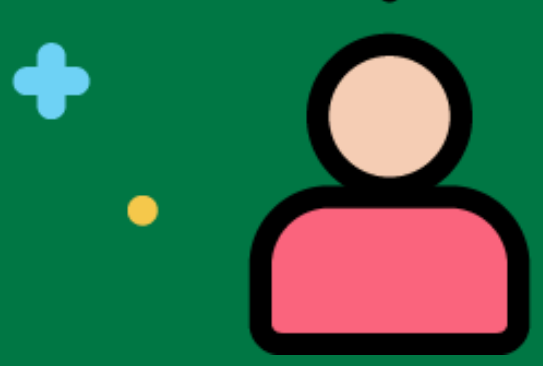
Data Engineering 101 - ETL Terminology

ETL Testing



The process of verifying that the ETL process is working correctly and that the data is accurate and complete.

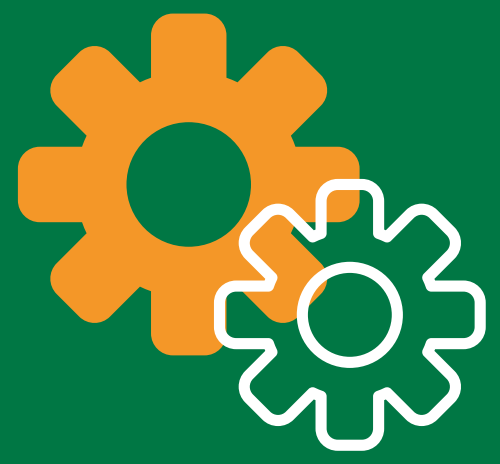
Conducting ETL testing to ensure that all data transformations have been applied correctly and that the loaded data matches the source data.



Shwetank Singh
GritSetGrow - GSGLearn.com

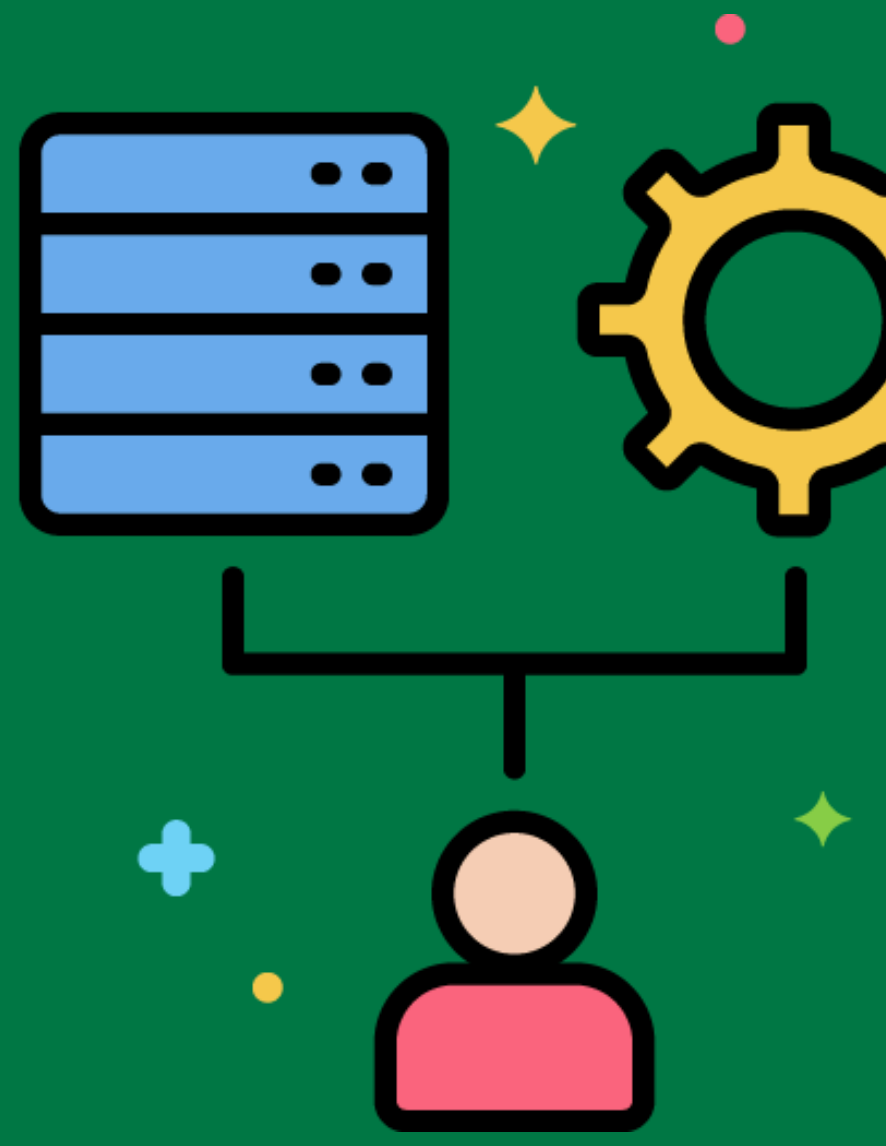
Data Engineering 101 - ETL Terminology

ETL Debugging



Identifying and fixing issues in the ETL process.

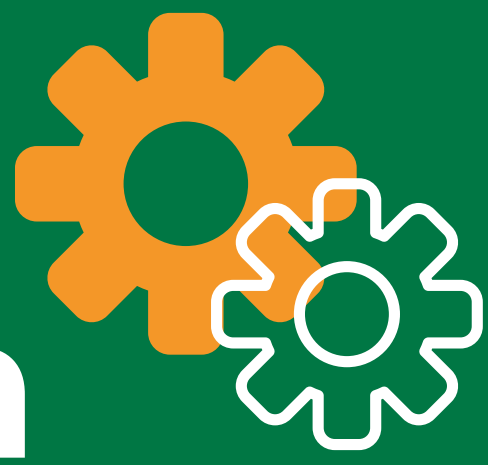
Debugging an ETL process to resolve an error in the data transformation logic.



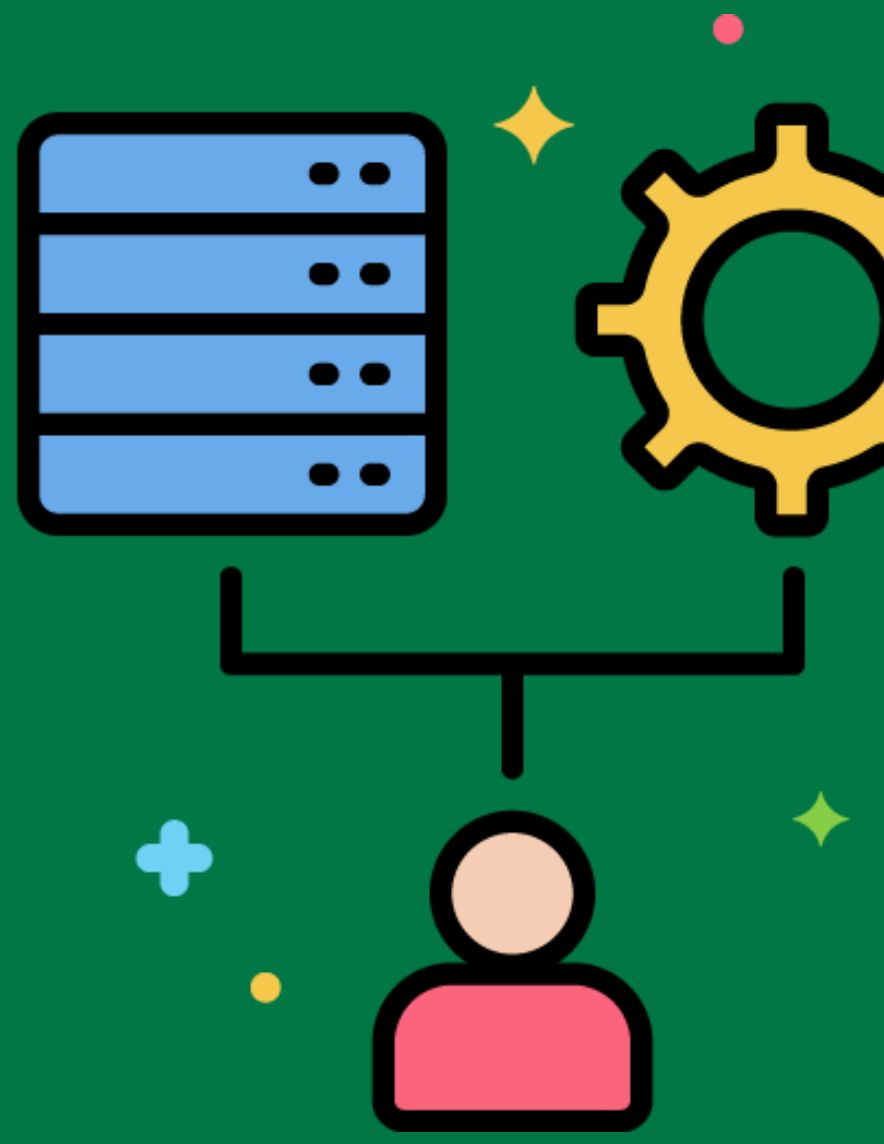
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Logic



The specific logic and rules used to transform data during the ETL process.

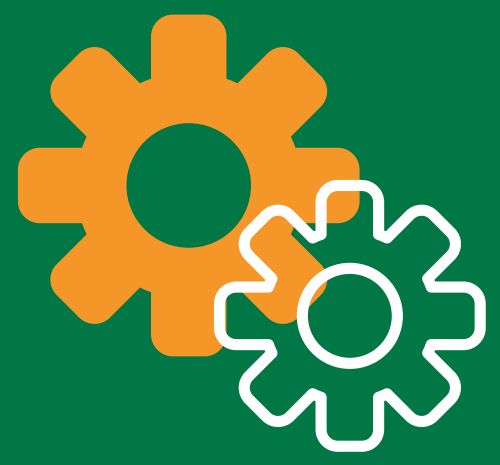


Writing transformation logic to convert product codes into product names.

Shwetank Singh
GritSetGrow - GSGLearn.com

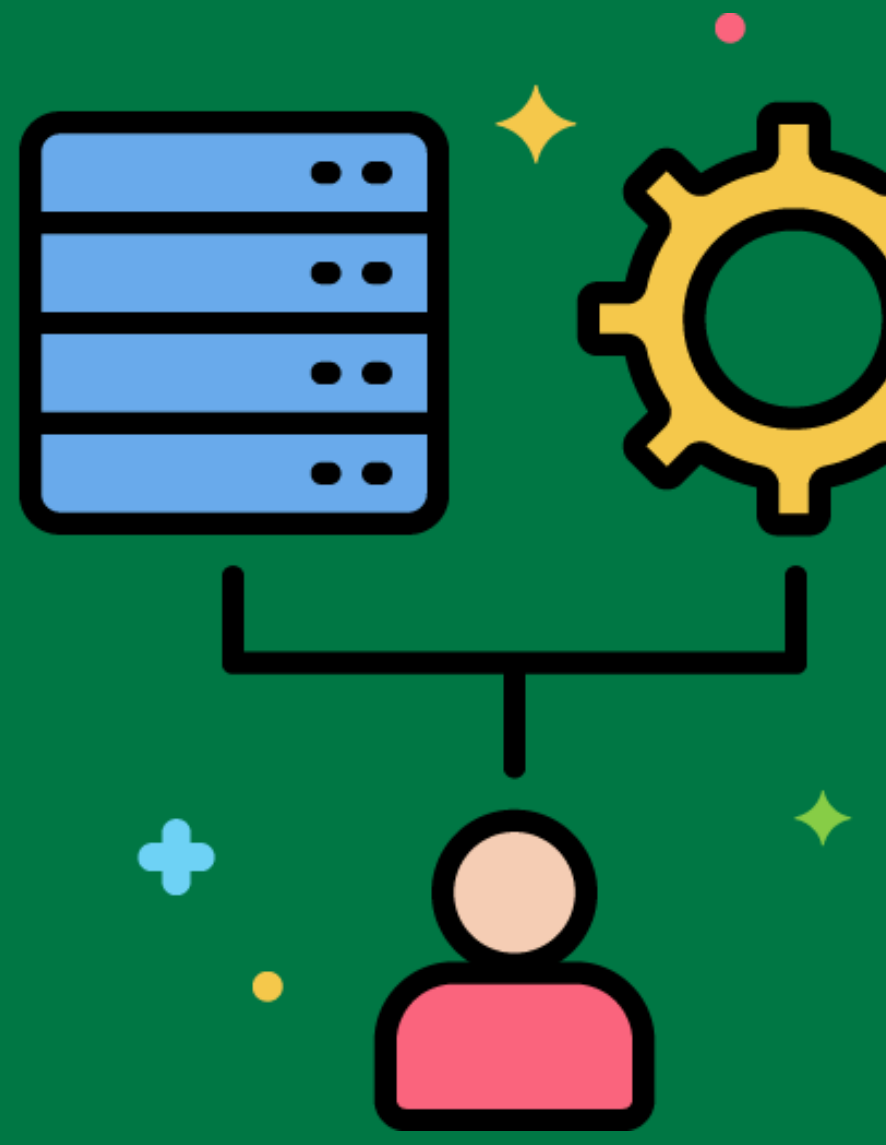
Data Engineering 101 - ETL Terminology

Data Validation Rules



Specific criteria and checks to ensure data validity during the ETL process.

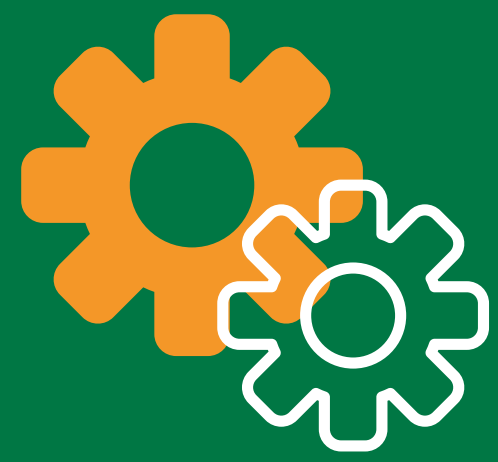
Implementing validation rules to ensure that email addresses are in the correct format and that date fields contain valid dates.



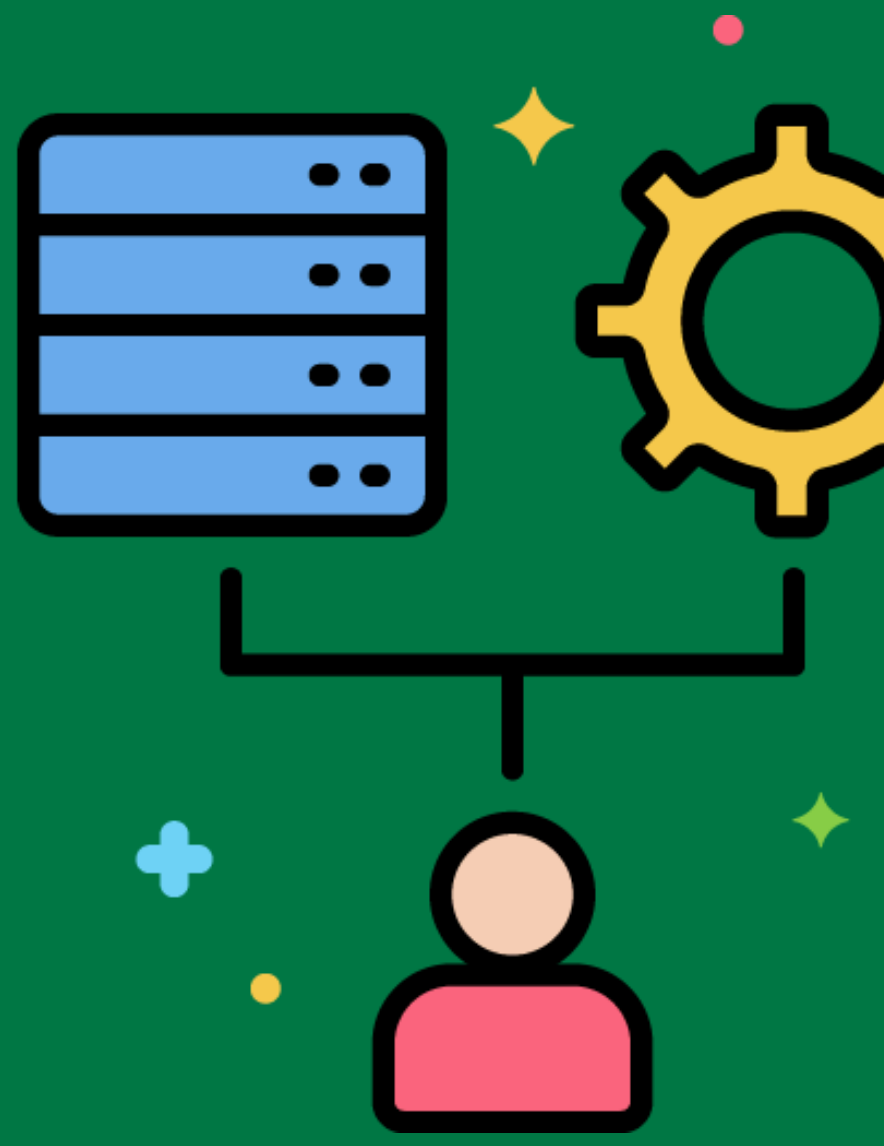
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Scripts



Scripts that automate data transformations during the ETL process.

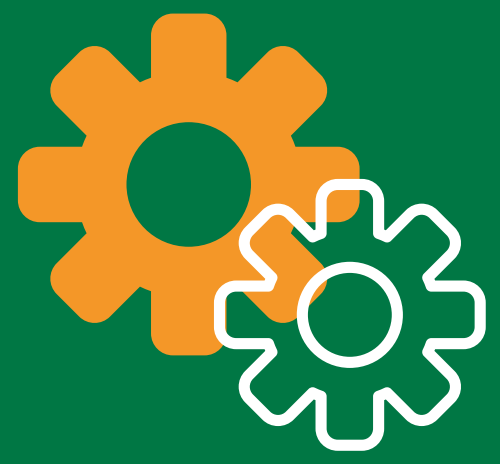


Writing a script to automatically transform raw sales data into a format suitable for the data warehouse.

Shwetank Singh
GritSetGrow - GSGLearn.com

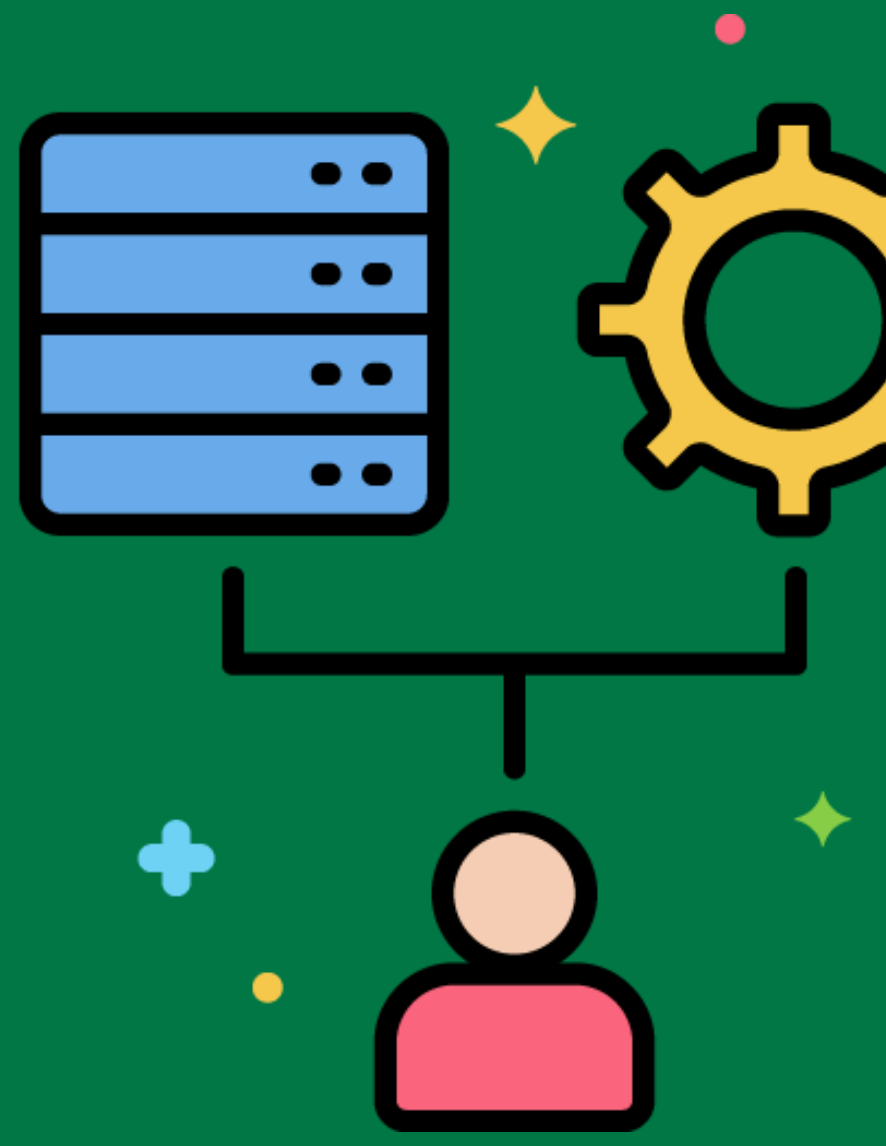
Data Engineering 101 - ETL Terminology

ETL Architecture



The design and structure of the ETL system, including its components and their relationships.

Designing an ETL architecture that includes a staging area, ETL engine, and data warehouse.

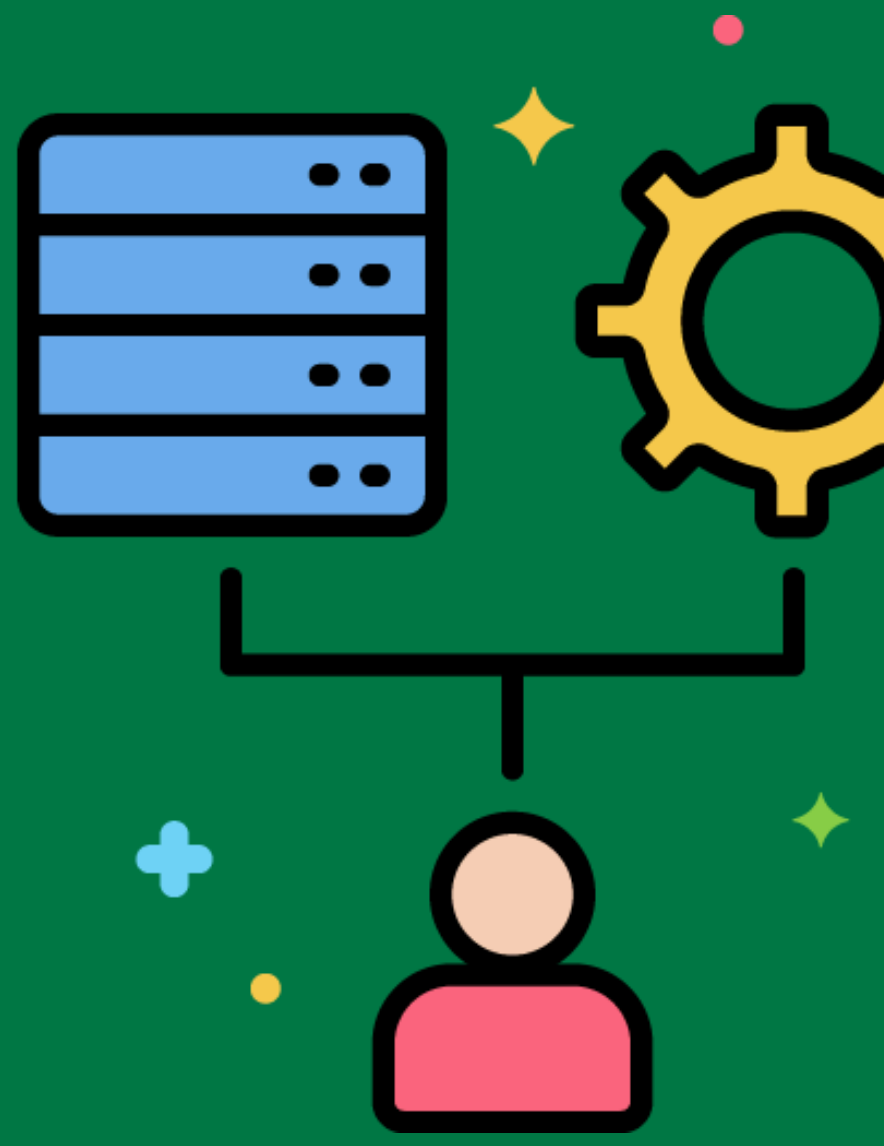


Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Framework

A set of tools and best practices for designing and implementing data transformations.

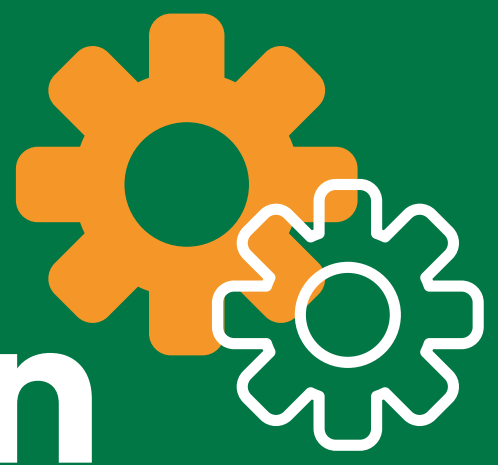


Using a data transformation framework to standardize and simplify the development of ETL processes.

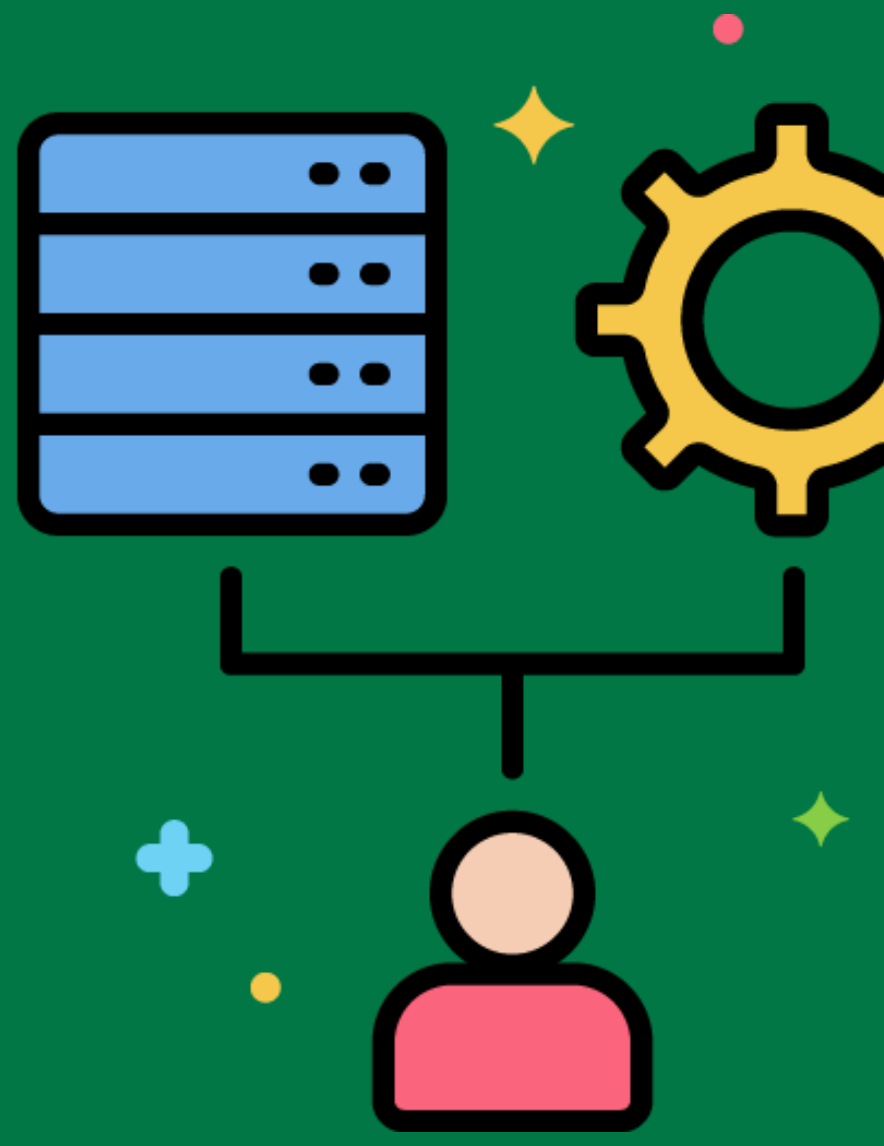
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Workflow



The sequence and flow of tasks involved in the data transformation process.



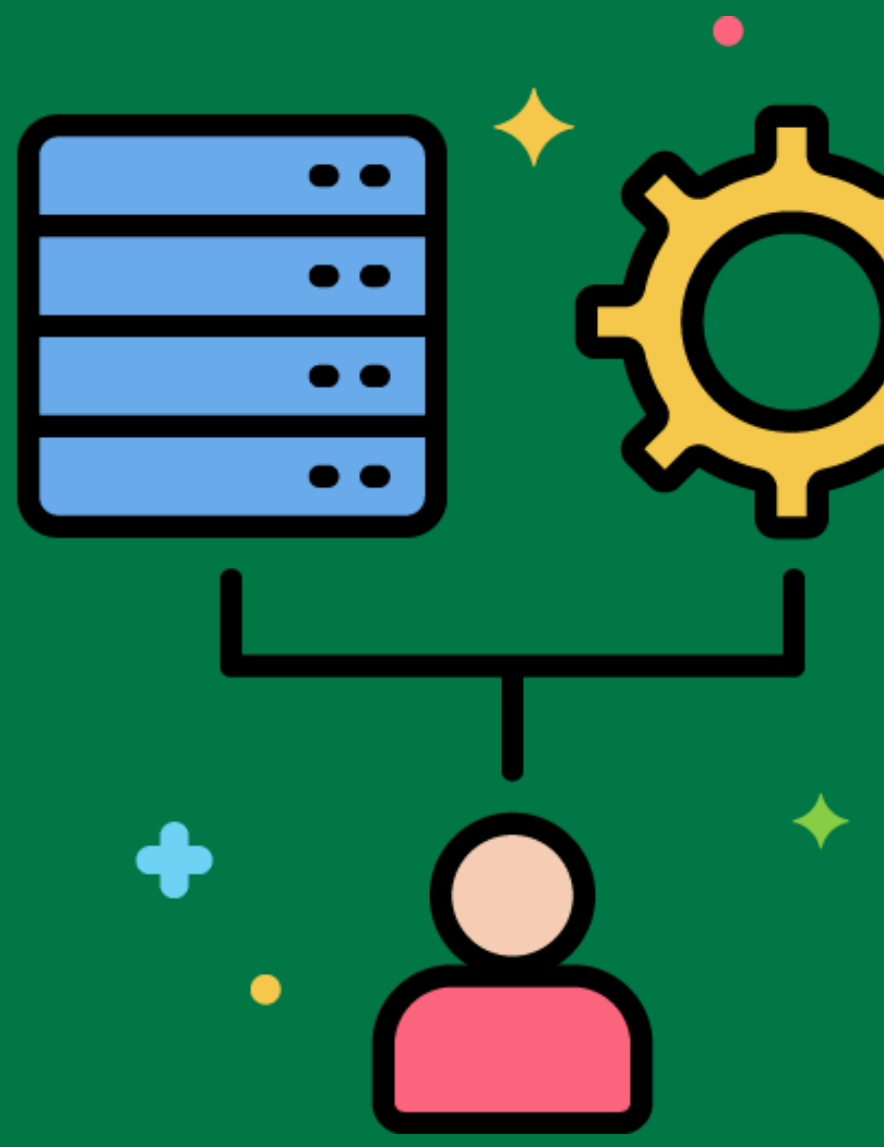
Creating a workflow that includes tasks for data extraction, data transformation, and data loading.

Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Pipeline

A series of processes that transform data from its raw form into a format suitable for the data warehouse.

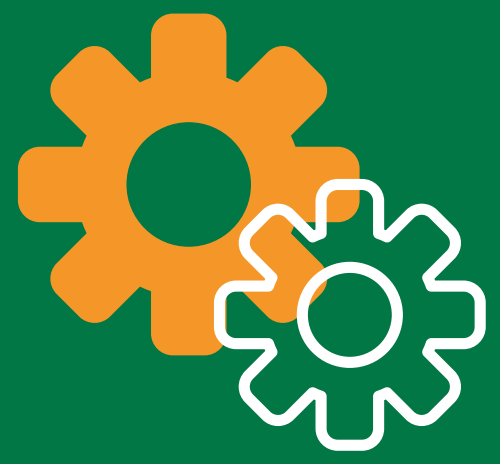


Designing a pipeline that extracts raw data from various sources, applies transformation rules, and loads the transformed data into the data warehouse.

Shwetank Singh
GritSetGrow - GSGLearn.com

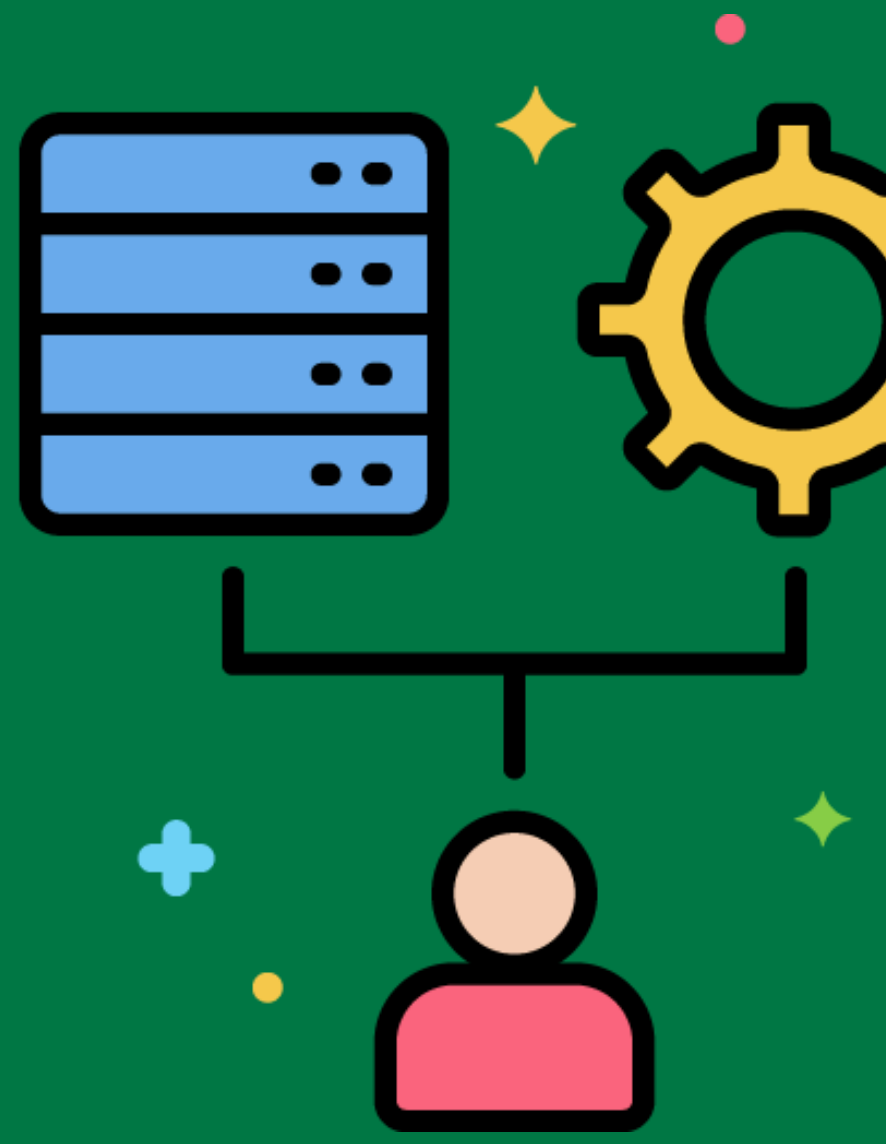
Data Engineering 101 - ETL Terminology

ETL Tools and Technologies



Software tools and technologies used to implement the ETL process.

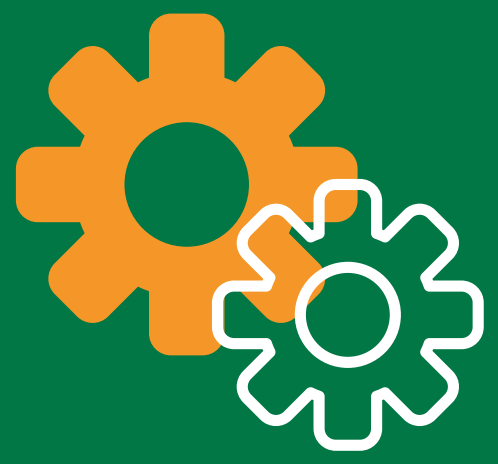
Using ETL tools like Talend, Apache Nifi, or Microsoft SSIS to automate and manage the ETL process.



Shwetank Singh
GritSetGrow - GSGLearn.com

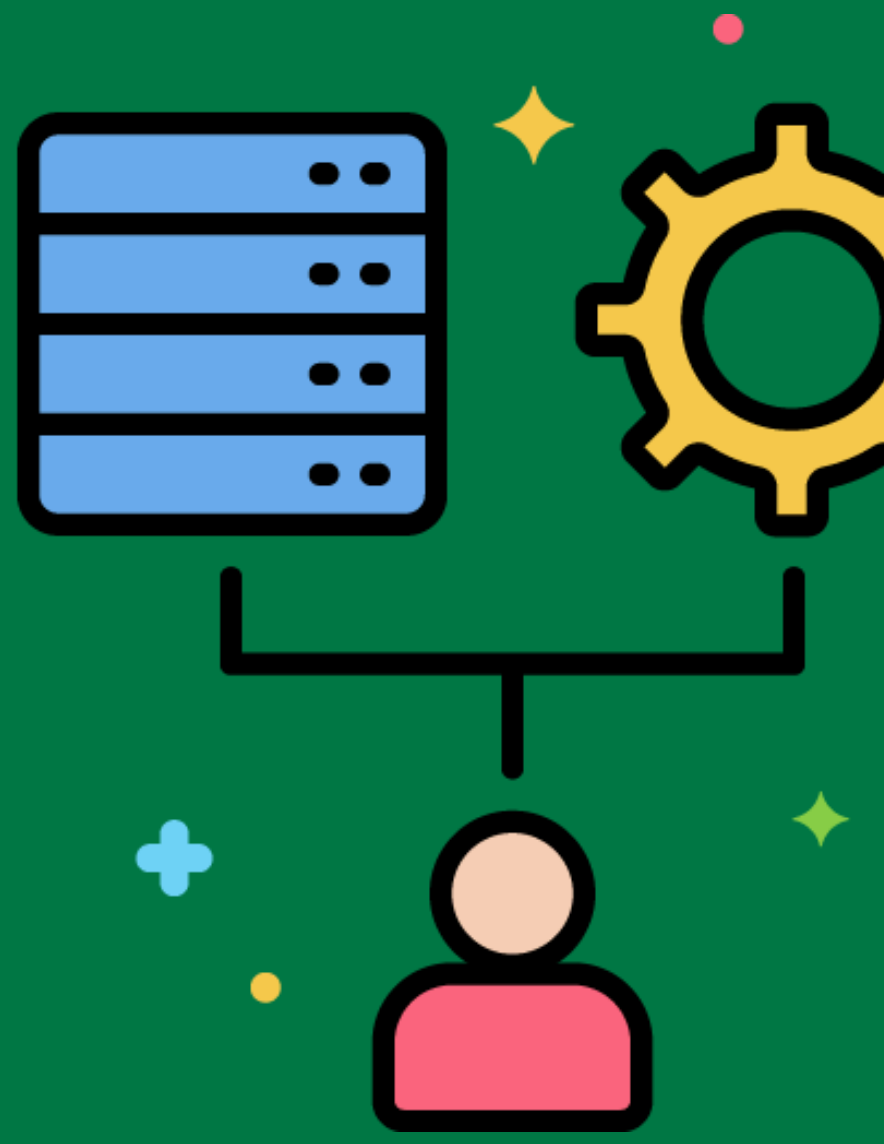
Data Engineering 101 - ETL Terminology

ETL Resiliency



The ability of the ETL process to recover from failures and continue processing.

Implementing checkpointing in ETL processes to allow for restarting from the last successful step after a failure.



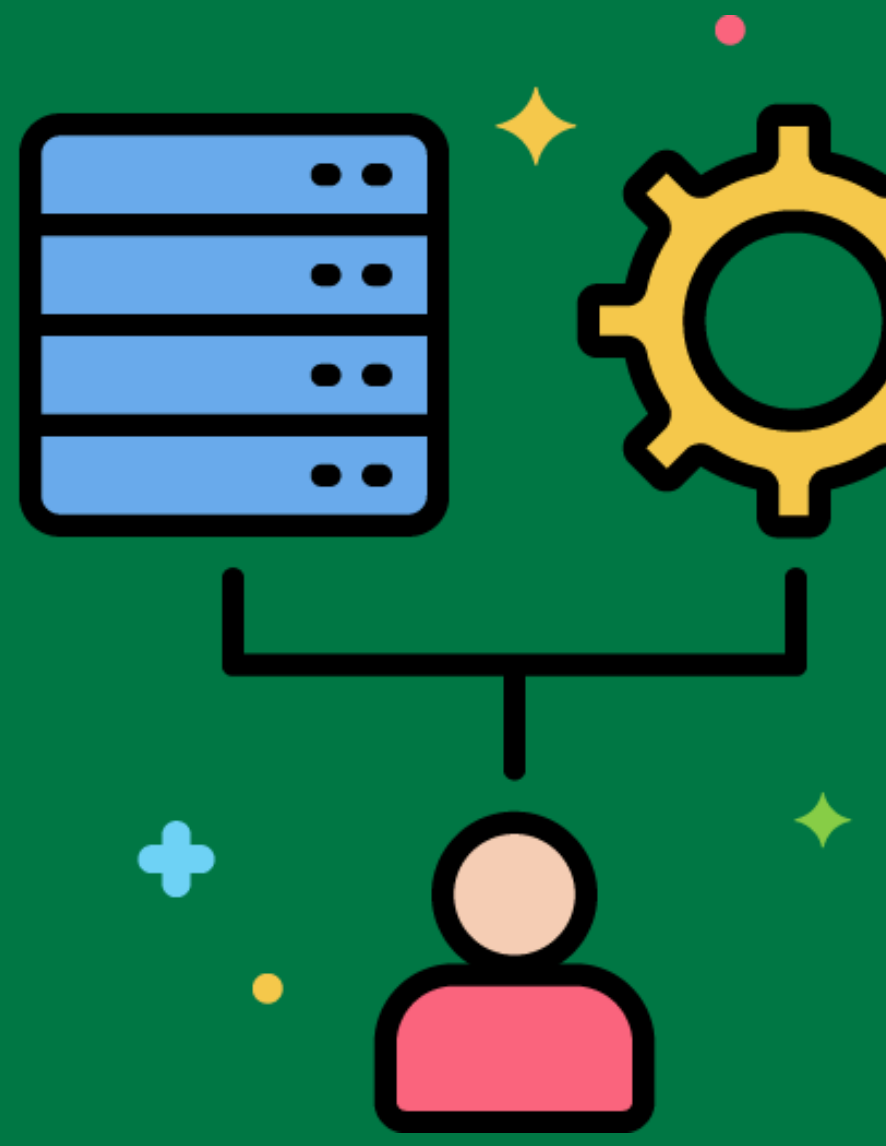
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Patterns

Common methods and best practices for transforming data in ETL processes.

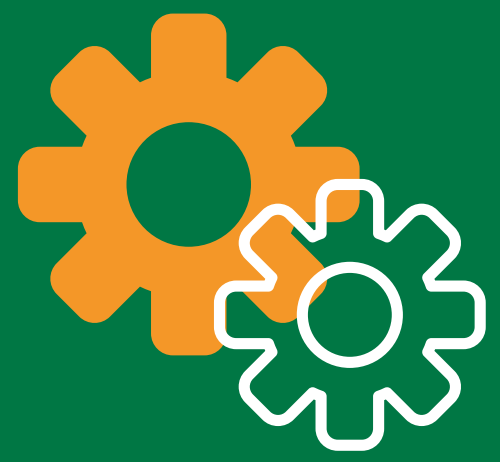
Using map-reduce patterns for processing large datasets in parallel during transformation.



Shwetank Singh
GritSetGrow - GSGLearn.com

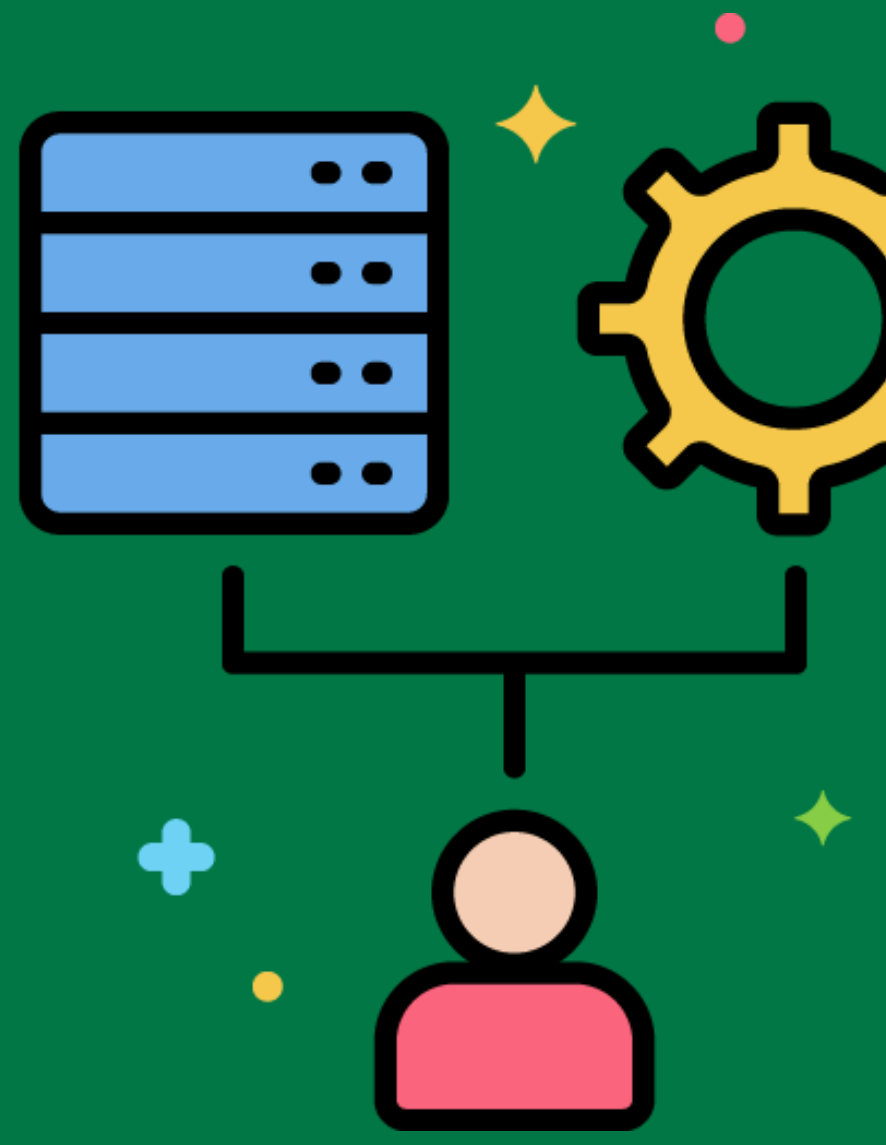
Data Engineering 101 - ETL Terminology

Data Quality Monitoring



Continuously tracking data quality metrics to ensure the accuracy and reliability of data in the data warehouse.

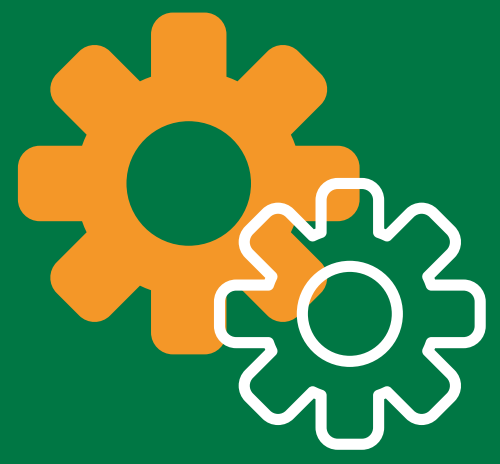
Using dashboards to monitor data quality metrics like completeness, accuracy, and consistency in real-time.



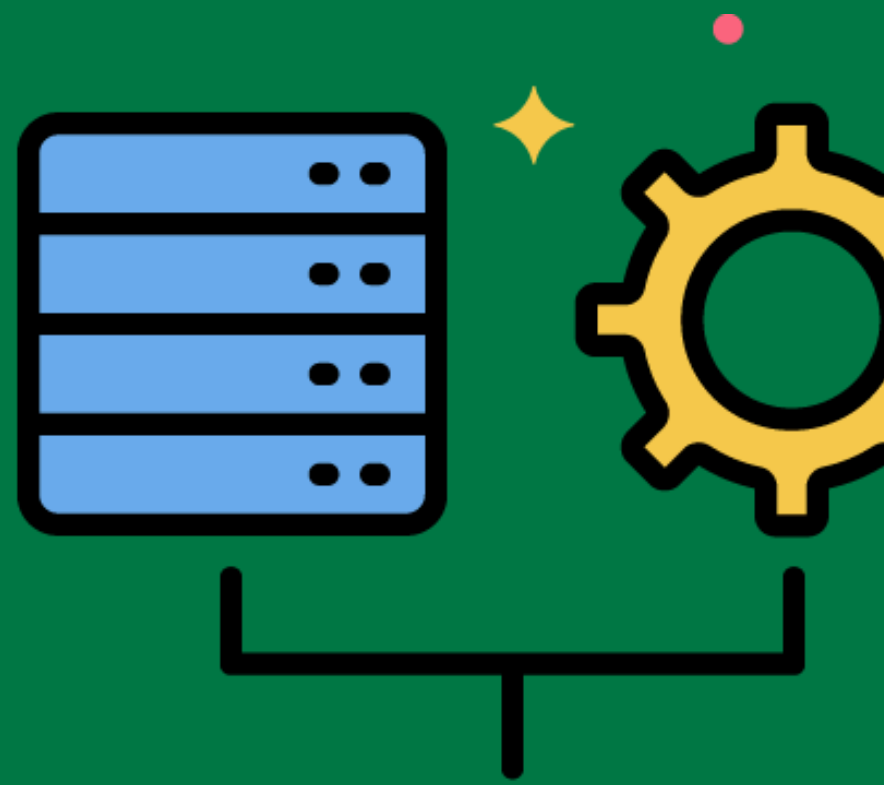
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

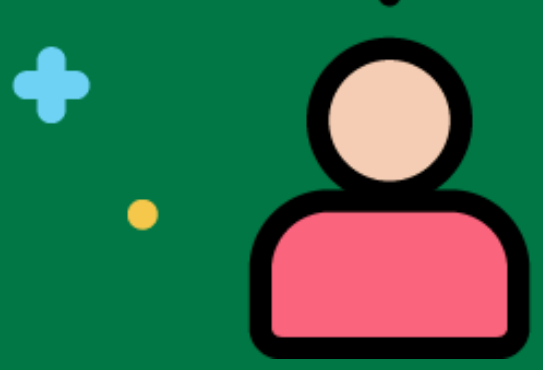
ETL Scheduling Strategies



Approaches for planning and managing the execution of ETL jobs.



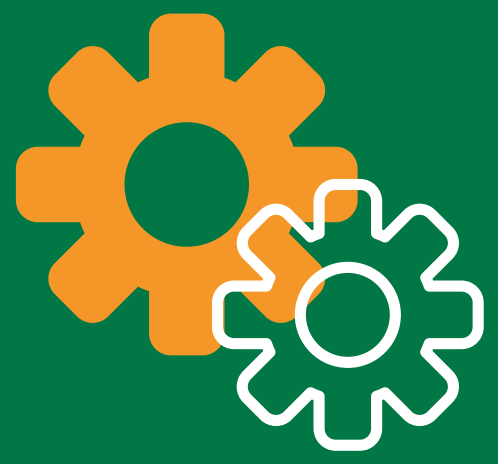
Using a time-based scheduling strategy to run ETL jobs every night at midnight.



Shwetank Singh
GritSetGrow - GSGLearn.com

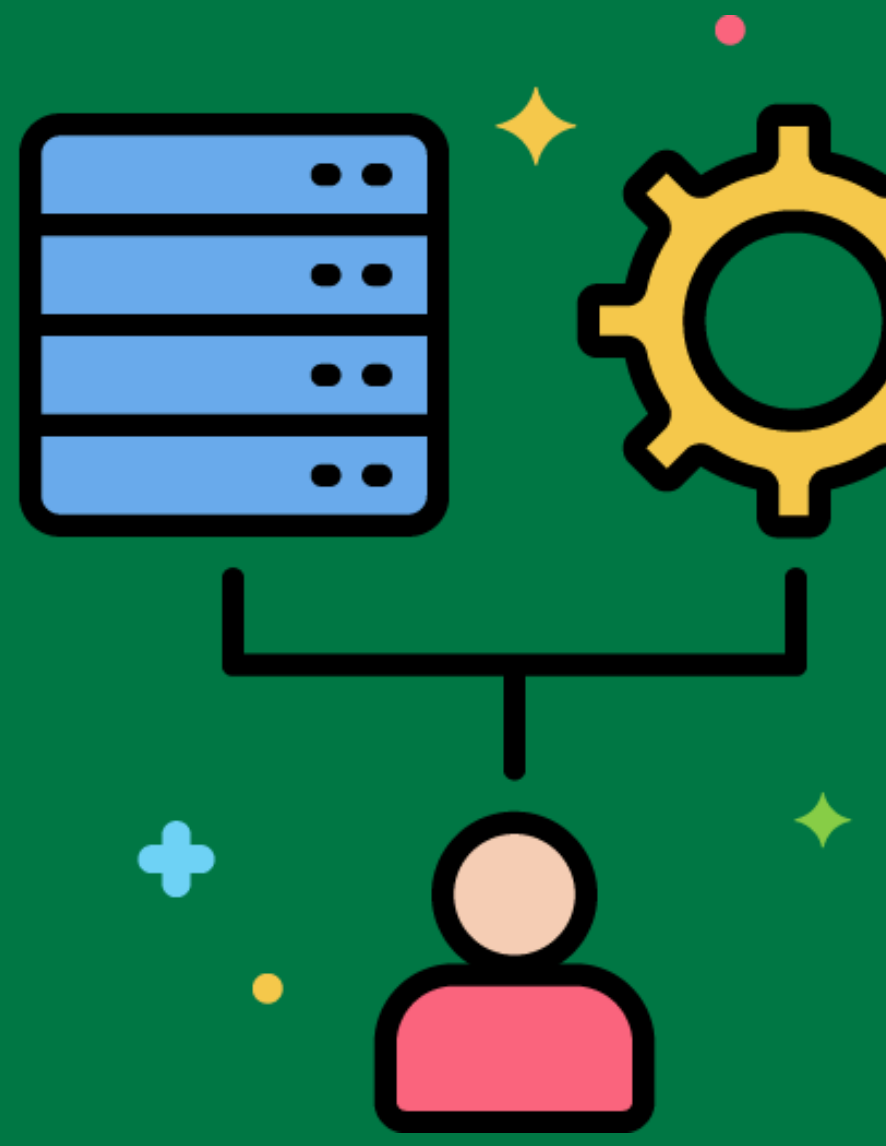
Data Engineering 101 - ETL Terminology

ETL Load Balancing



Distributing ETL workloads across multiple servers or processes to optimize performance.

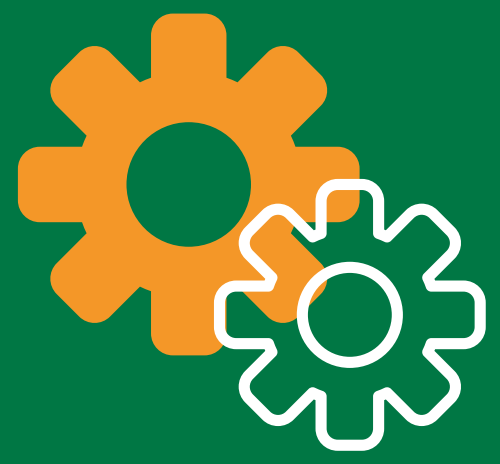
Implementing load balancing to distribute data transformation tasks across a cluster of ETL servers.



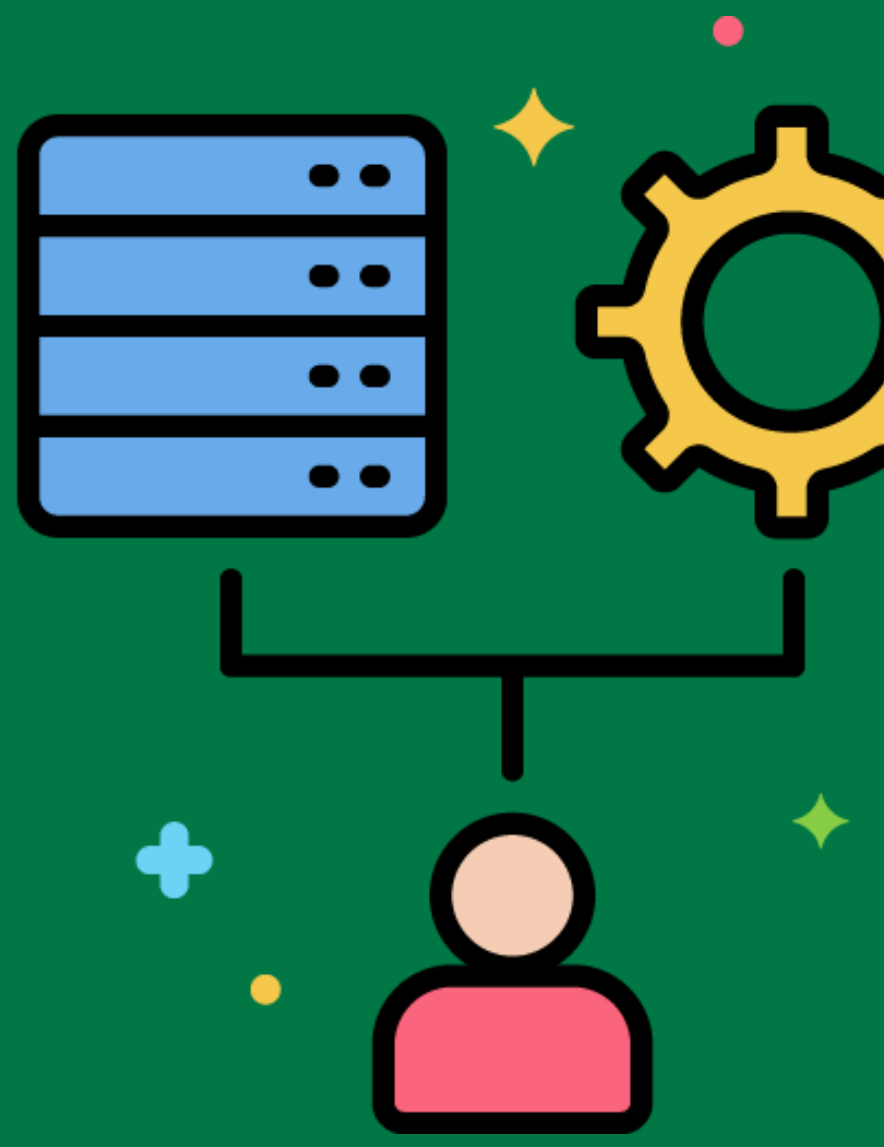
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Load Optimization



Techniques to improve the efficiency and speed of loading data into the data warehouse.



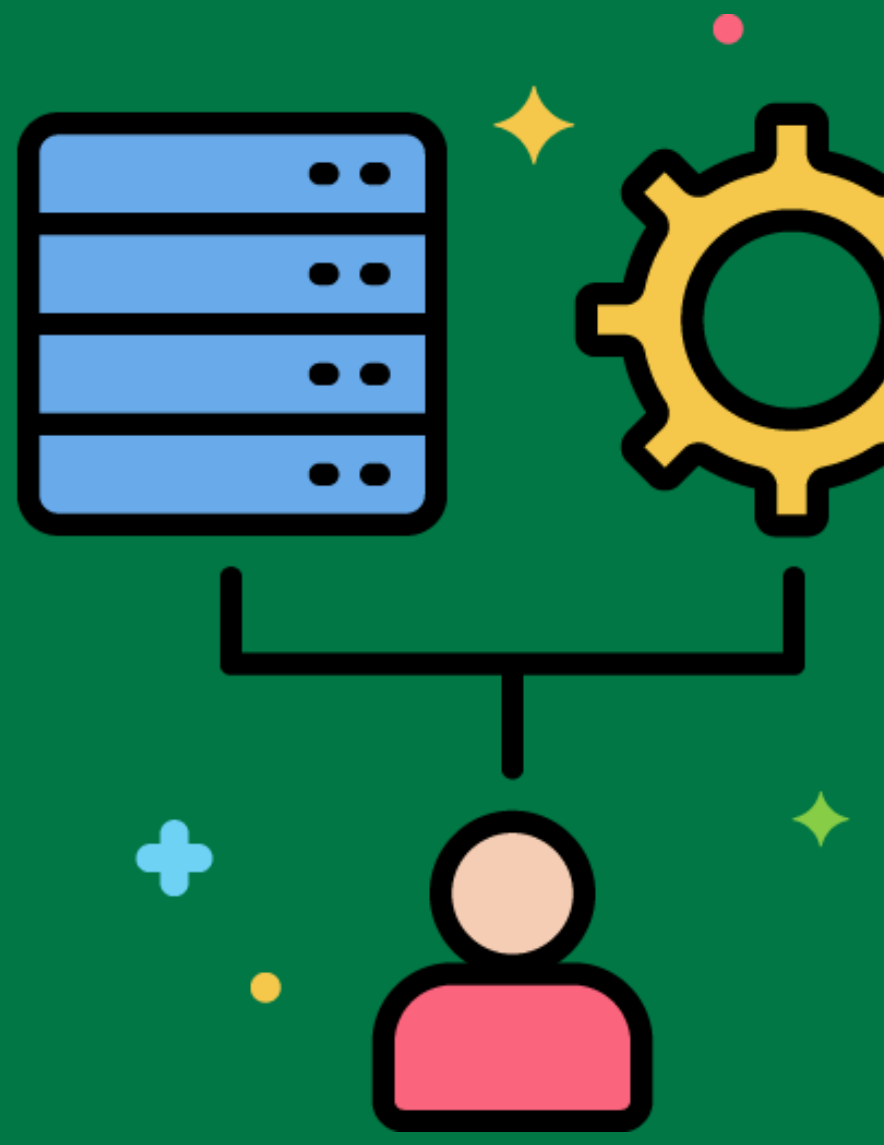
Using bulk load utilities and optimizing database indexes to speed up the loading process.

Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Engine

A component of the ETL system that performs data transformation tasks.

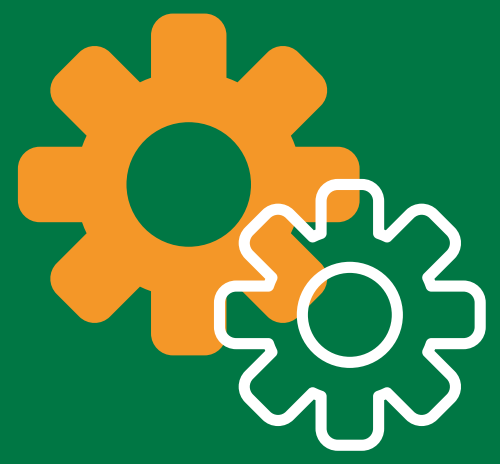


Using a dedicated transformation engine to apply complex business rules and transformations to extracted data.

Shwetank Singh
GritSetGrow - GSGLearn.com

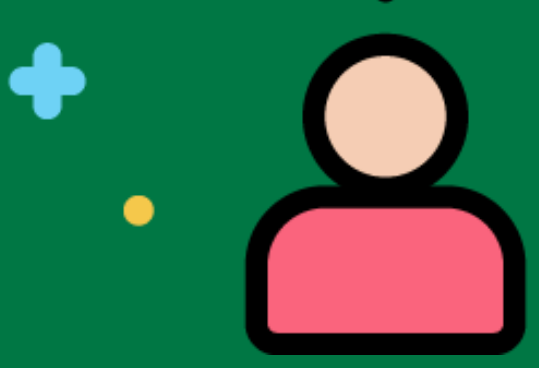
Data Engineering 101 - ETL Terminology

ETL Error Recovery



Strategies and mechanisms to handle and recover from errors during the ETL process.

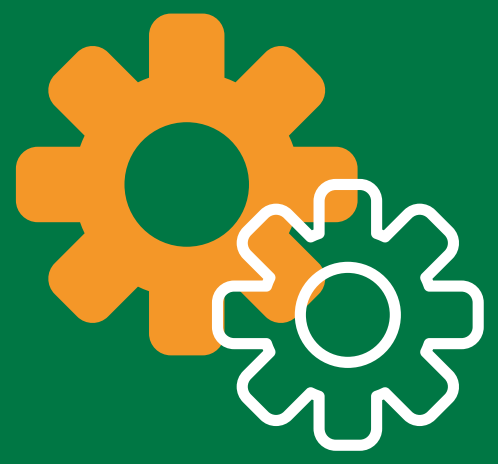
Implementing a retry mechanism to handle temporary network failures during data extraction.



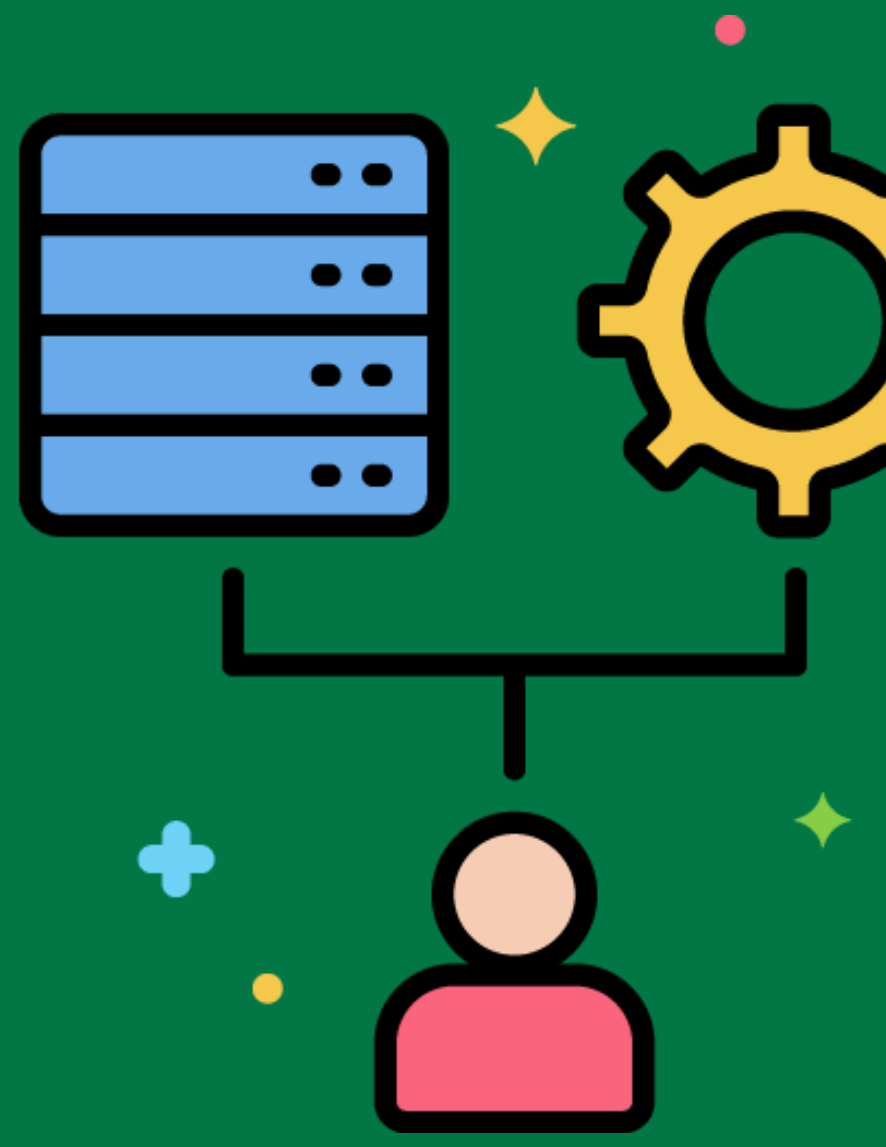
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Rollback



The ability to revert changes made by an ETL process in case of failure.

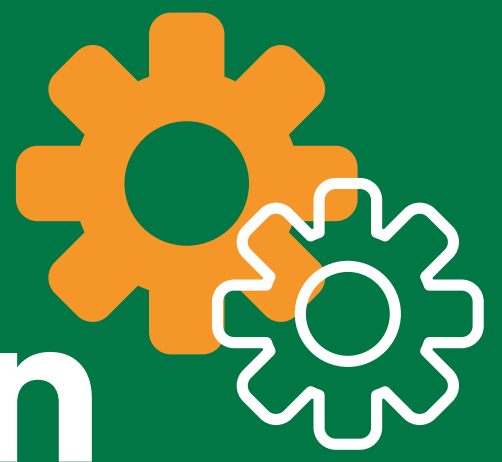


Using database transactions to ensure that changes can be rolled back if an error occurs during data loading.

Shwetank Singh
GritSetGrow - GSGLearn.com

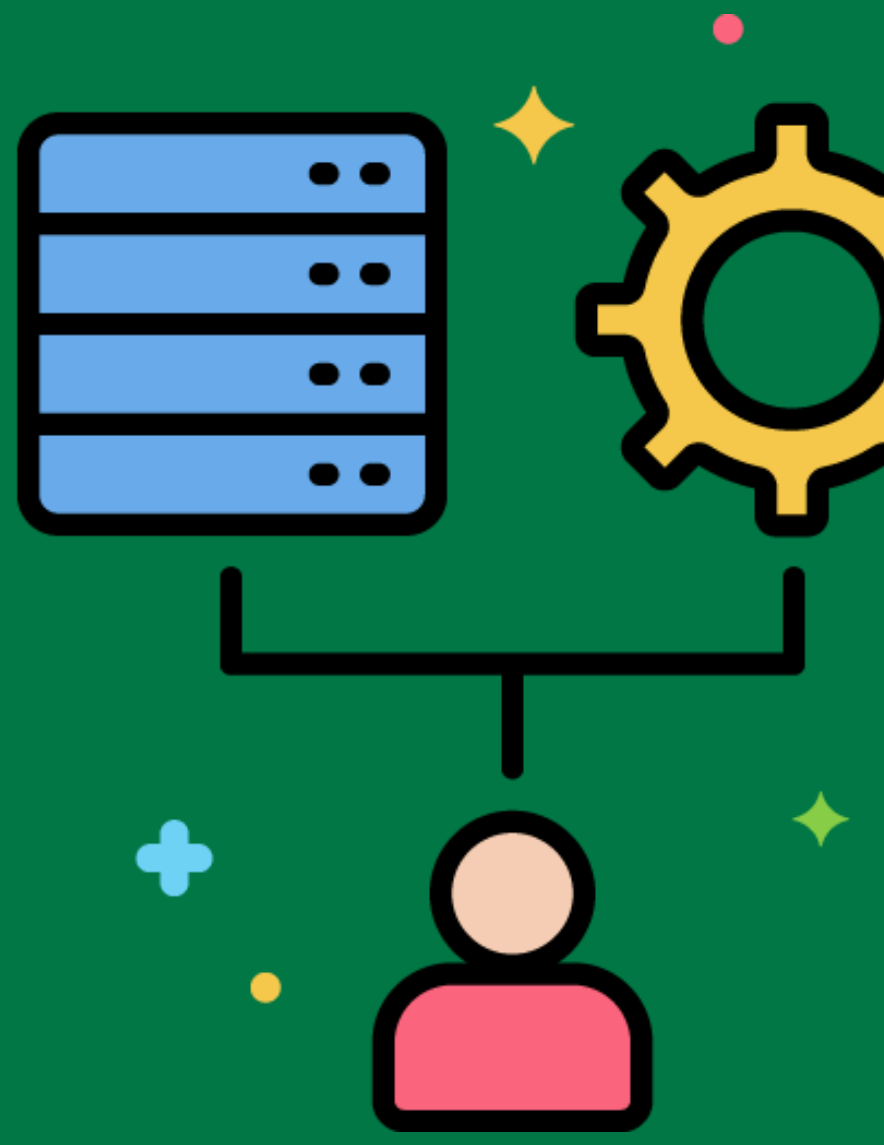
Data Engineering 101 - ETL Terminology

Data Transformation Pipelines



A series of connected data transformation tasks that process data in sequence.

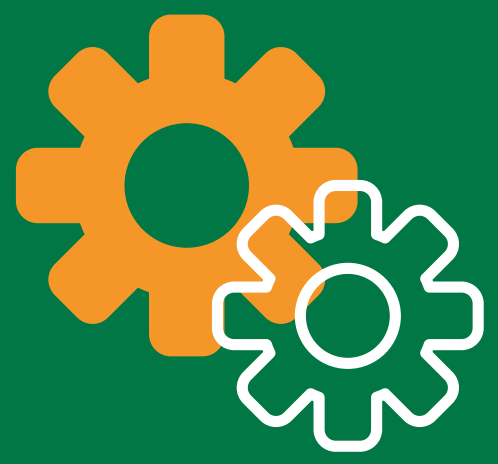
Designing a pipeline that extracts data, applies validation rules, transforms it, and loads it into the data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

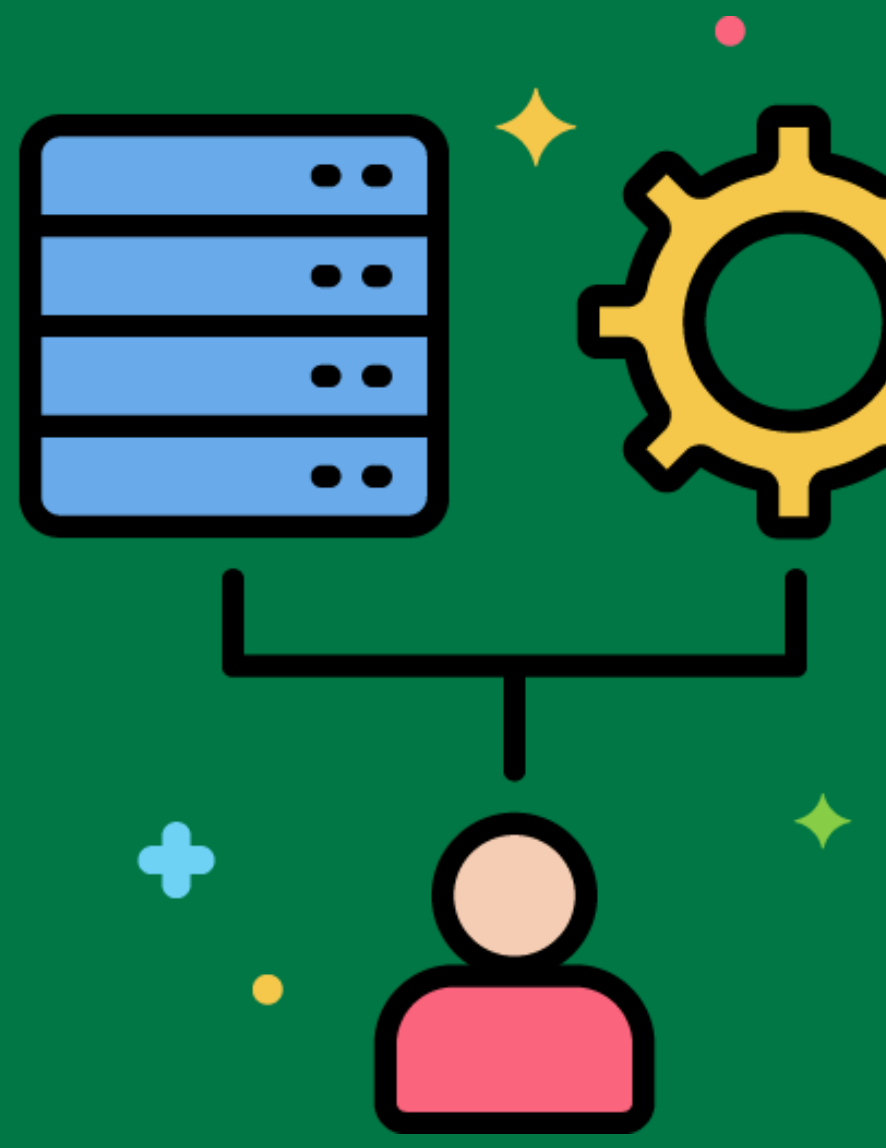
Data Engineering 101 - ETL Terminology

ETL Job Dependency Management



Managing the dependencies between different ETL jobs to ensure they run in the correct order.

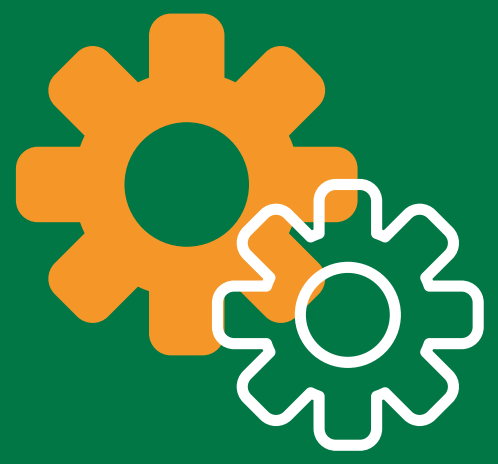
Defining dependencies between ETL jobs so that data extraction is completed before transformation and loading tasks begin.



Shwetank Singh
GritSetGrow - GSGLearn.com

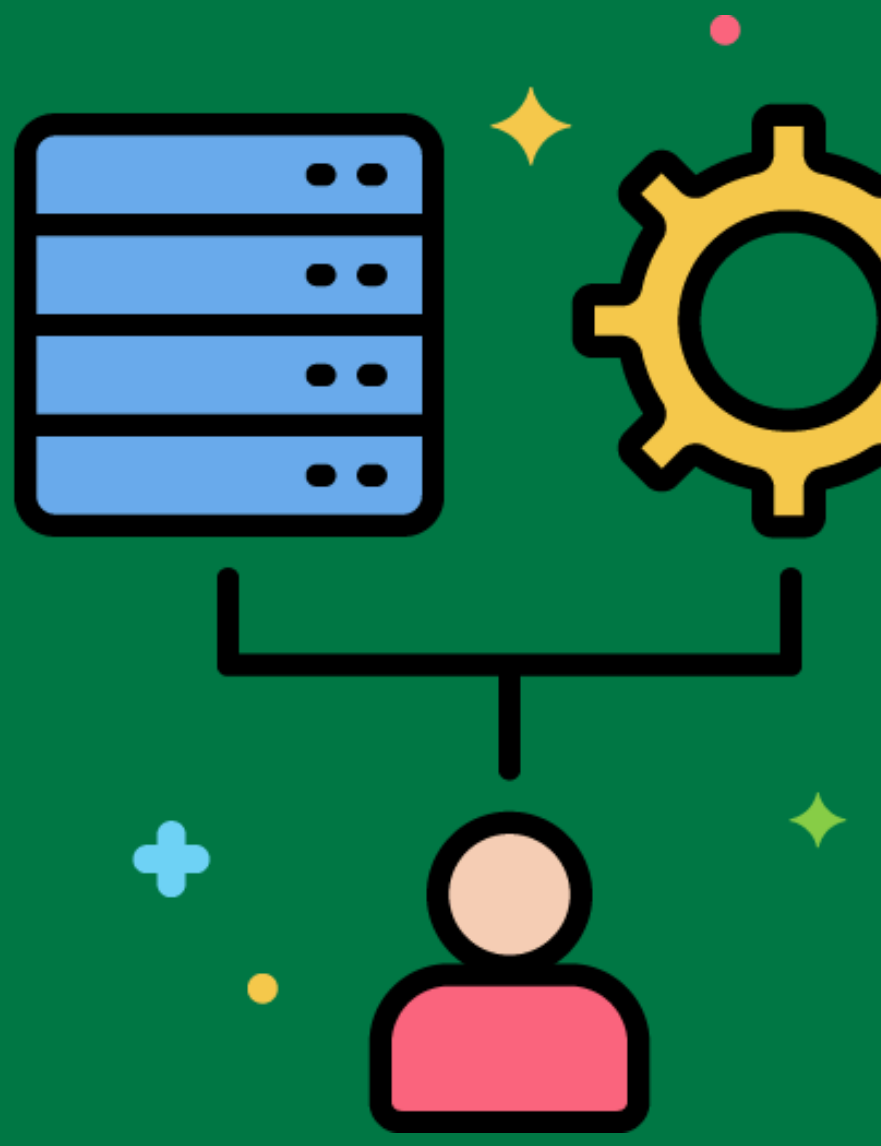
Data Engineering 101 - ETL Terminology

ETL Process Parallelism



Executing multiple ETL tasks concurrently to improve performance and reduce processing time.

Using parallel processing to run multiple data extraction tasks at the same time.



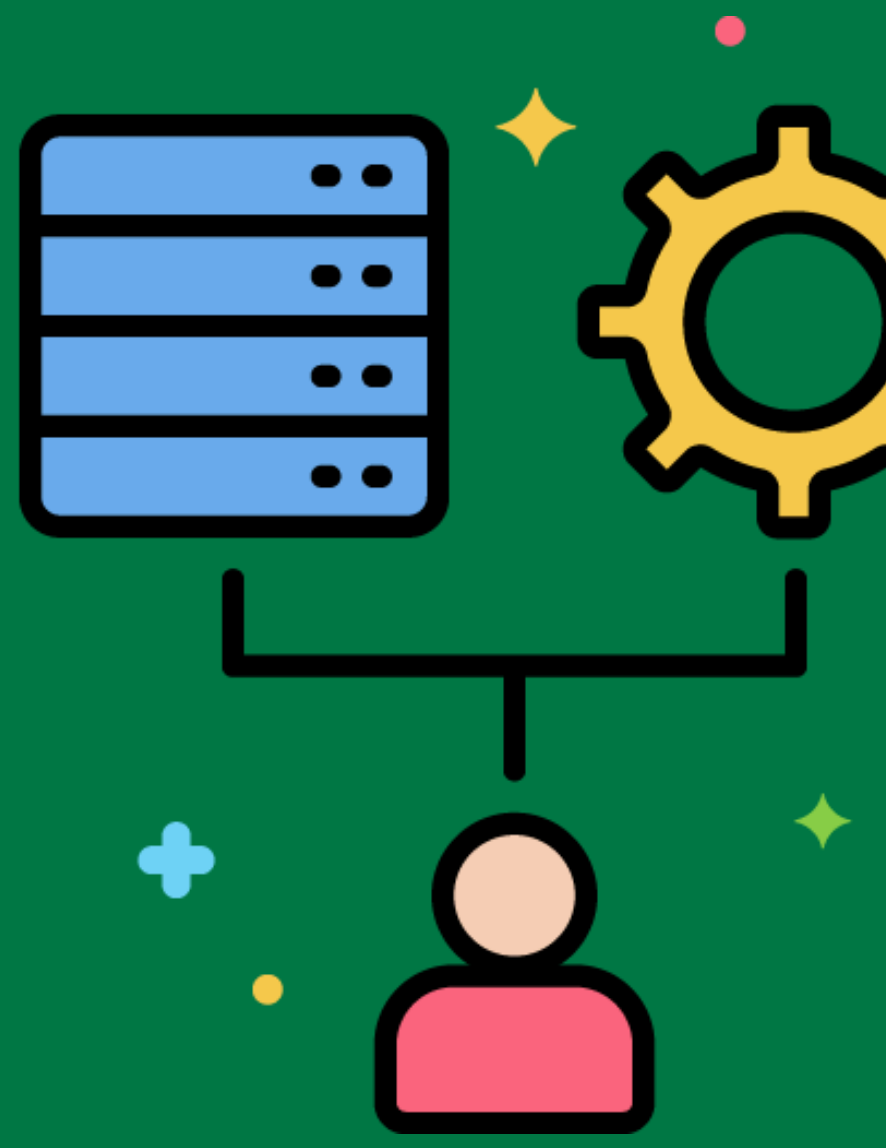
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

Data Transformation Libraries

Reusable libraries of data transformation functions and routines.

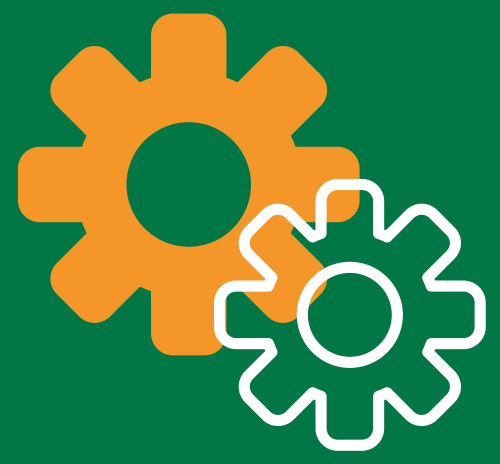
Creating a library of commonly used transformation functions, such as date format conversions and string manipulations.



Shwetank Singh
GritSetGrow - GSGLearn.com

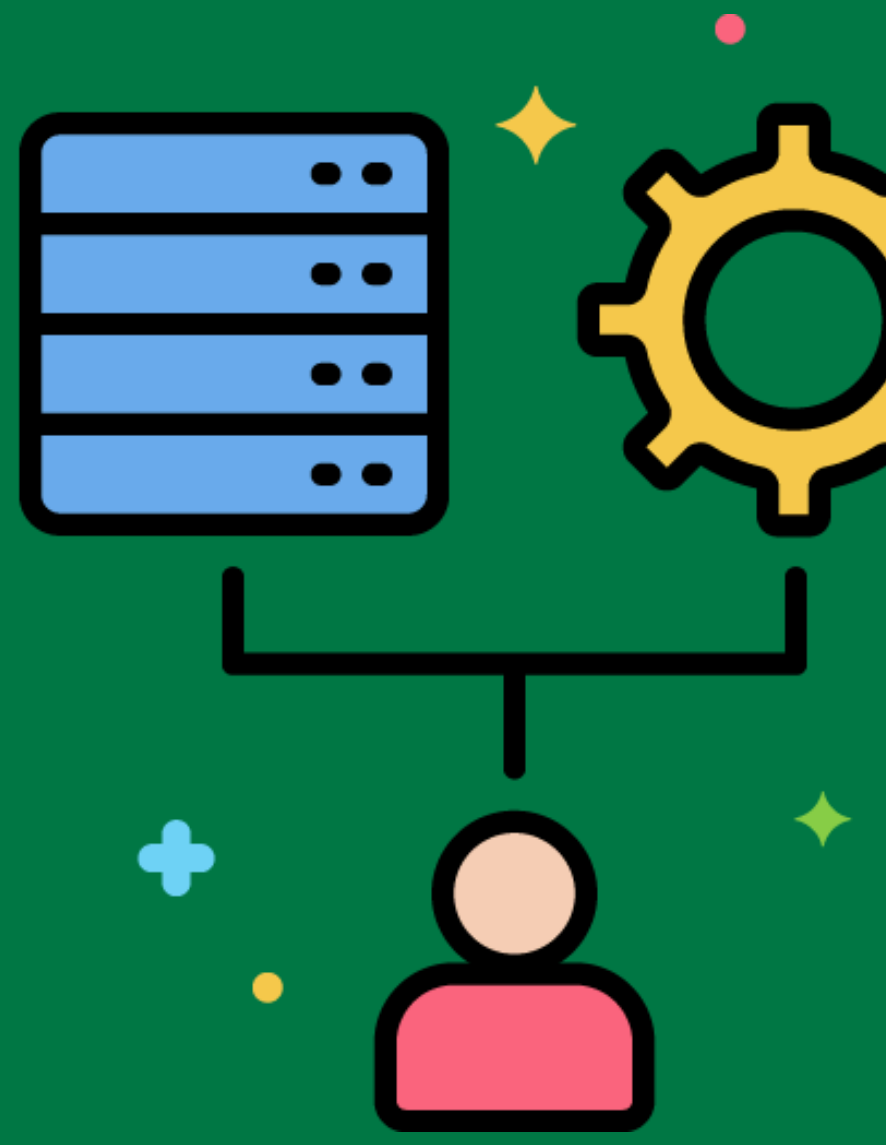
Data Engineering 101 - ETL Terminology

ETL Validation Framework



A structured approach to implementing and managing data validation rules during the ETL process.

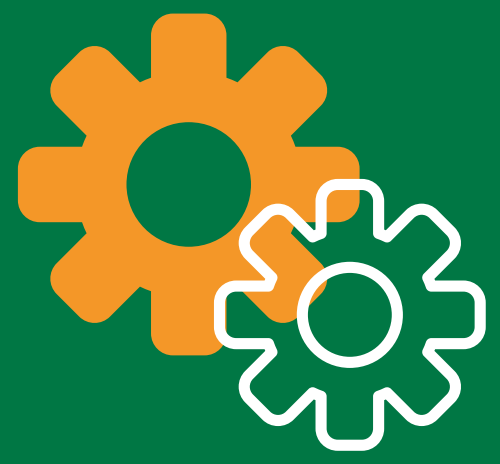
Using a validation framework to apply and manage data validation rules consistently across all ETL processes.



Shwetank Singh
GritSetGrow - GSGLearn.com

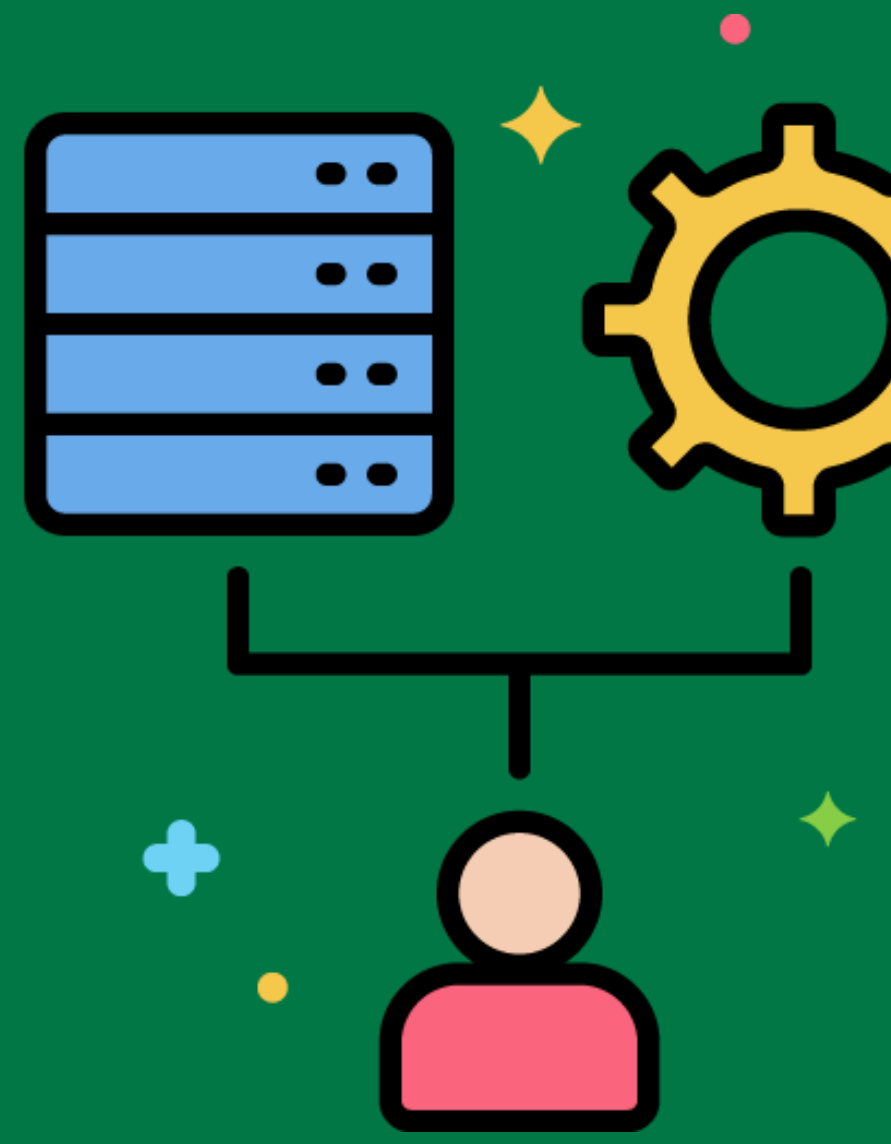
Data Engineering 101 - ETL Terminology

ETL Monitoring Dashboard



A visual interface for tracking the status and performance of ETL processes.

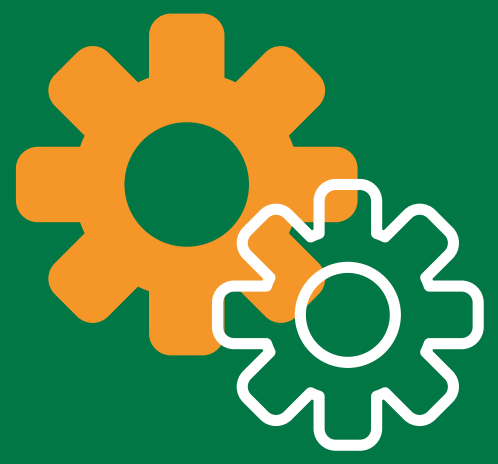
Implementing a dashboard to monitor ETL job execution times, error rates, and data quality metrics.



Shwetank Singh
GritSetGrow - GSGLearn.com

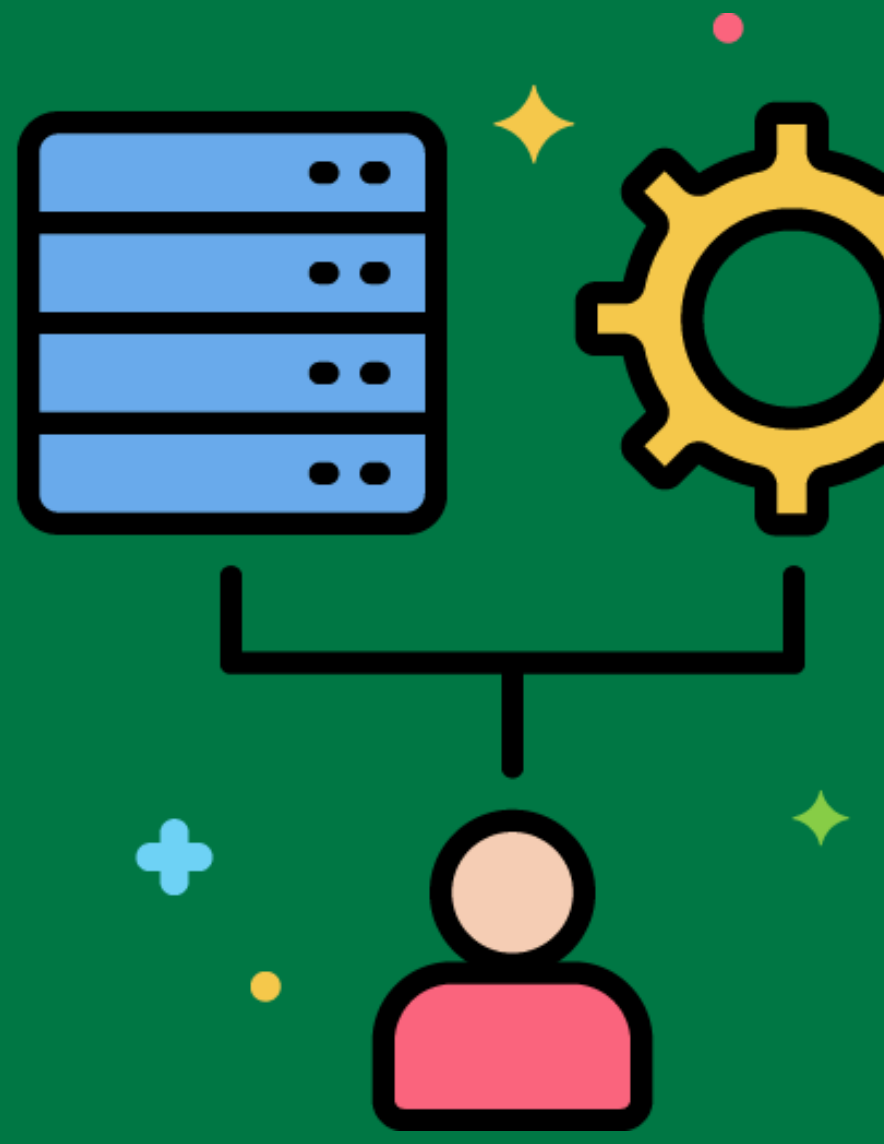
Data Engineering 101 - ETL Terminology

ETL Audit Trail



A record of all changes and transformations applied to data during the ETL process.

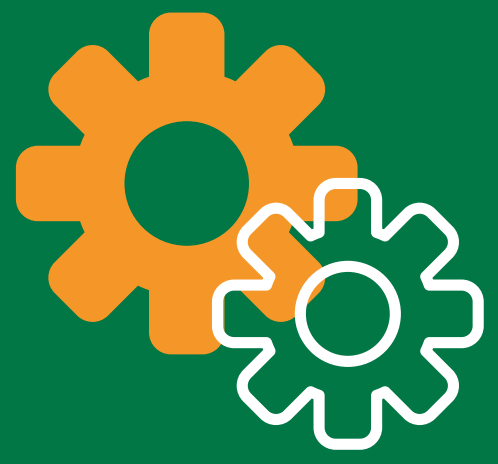
Maintaining an audit trail to track all data transformations and changes for compliance and debugging purposes.



Shwetank Singh
GritSetGrow - GSGLearn.com

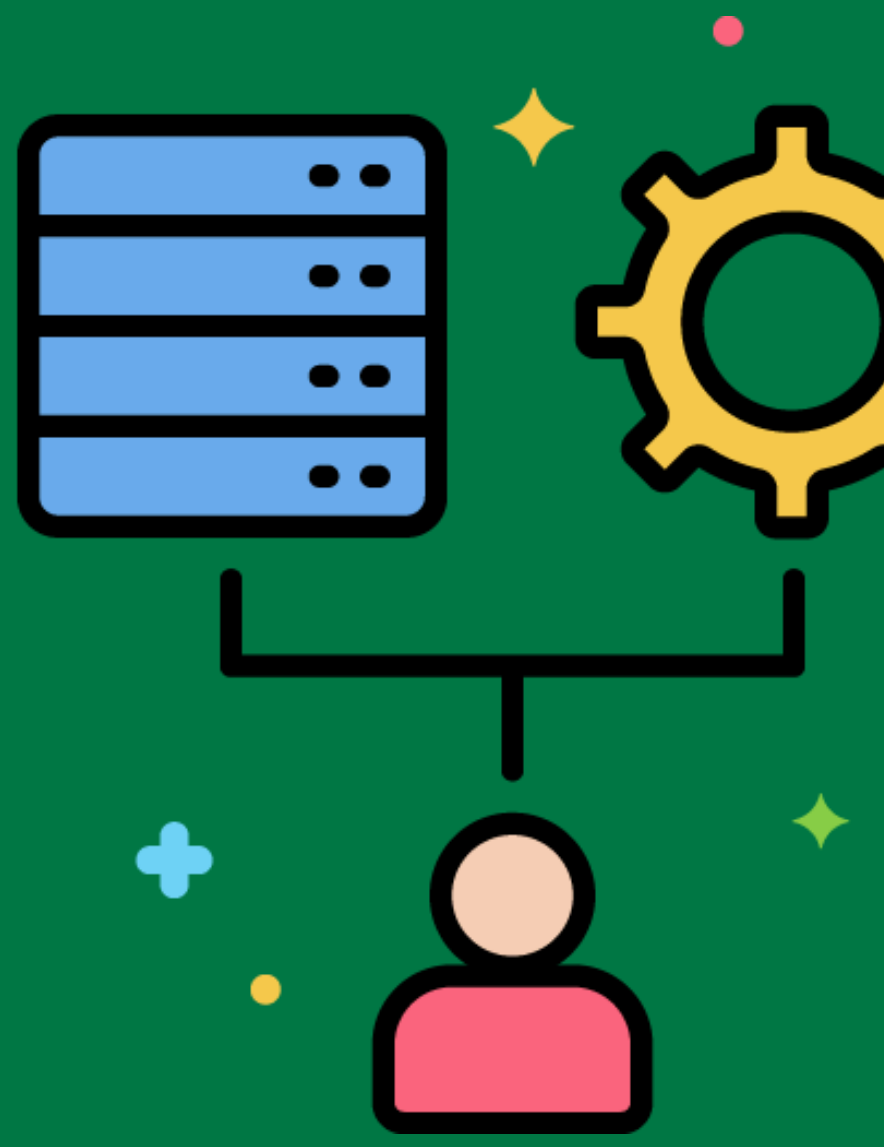
Data Engineering 101 - ETL Terminology

ETL Metadata Repository



A centralized store of metadata related to the ETL process, including source-to-target mappings, transformation rules, and data lineage.

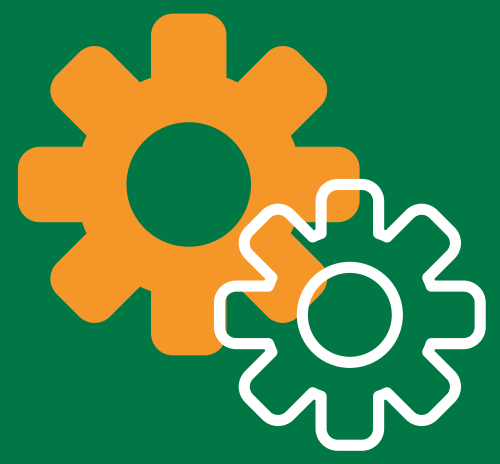
Using a metadata repository to document and manage all ETL processes, including data mappings and transformation logic.



Shwetank Singh
GritSetGrow - GSGLearn.com

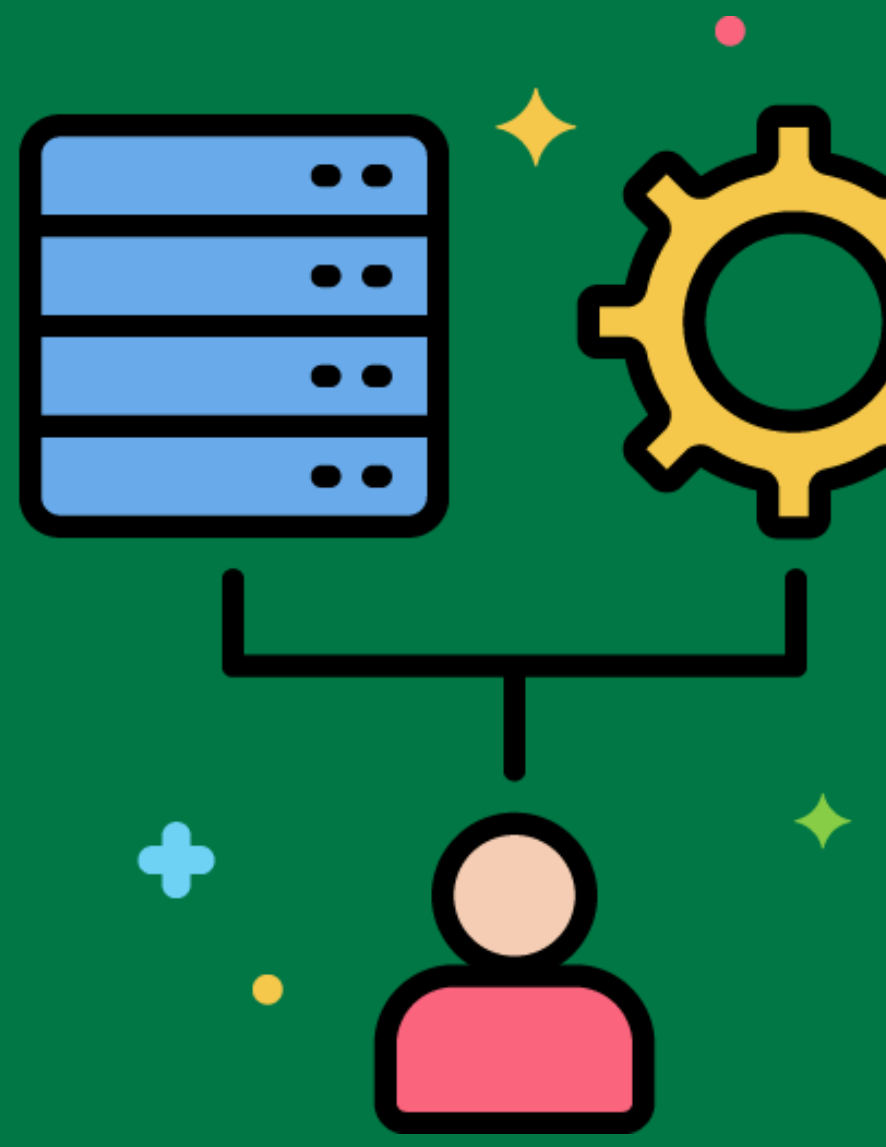
Data Engineering 101 - ETL Terminology

ETL Data Flow



The movement of data from source systems through extraction, transformation, and loading stages to the data warehouse.

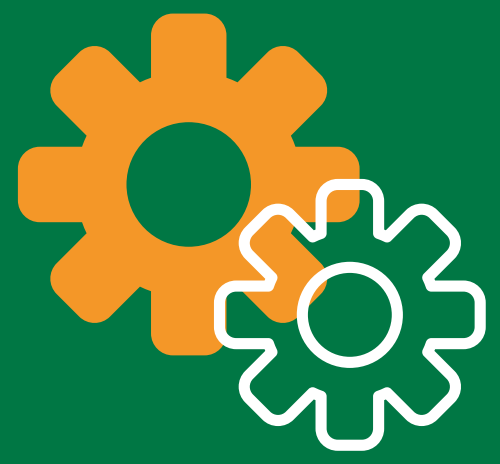
Designing the ETL data flow to move customer data from the CRM system, apply transformations, and load it into the data warehouse.



Shwetank Singh
GritSetGrow - GSGLearn.com

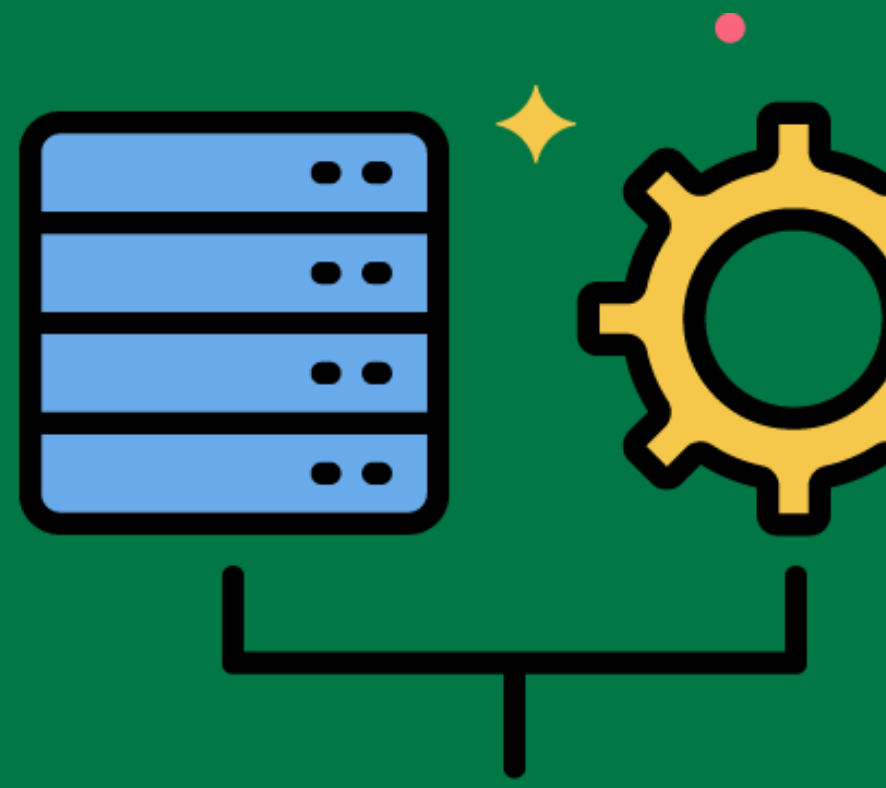
Data Engineering 101 - ETL Terminology

ETL Code Management



Practices for managing and versioning the code used in ETL processes.

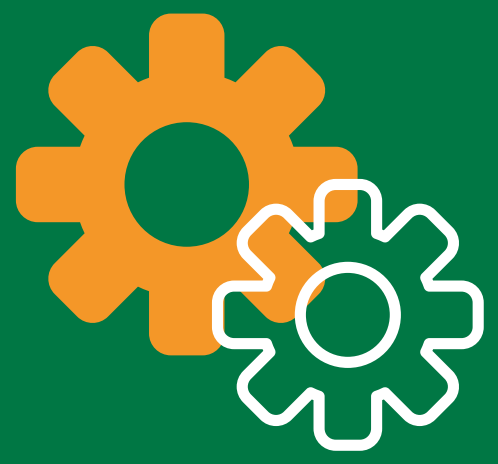
Using version control systems like Git to manage ETL scripts and transformation logic.



Shwetank Singh
GritSetGrow - GSGLearn.com

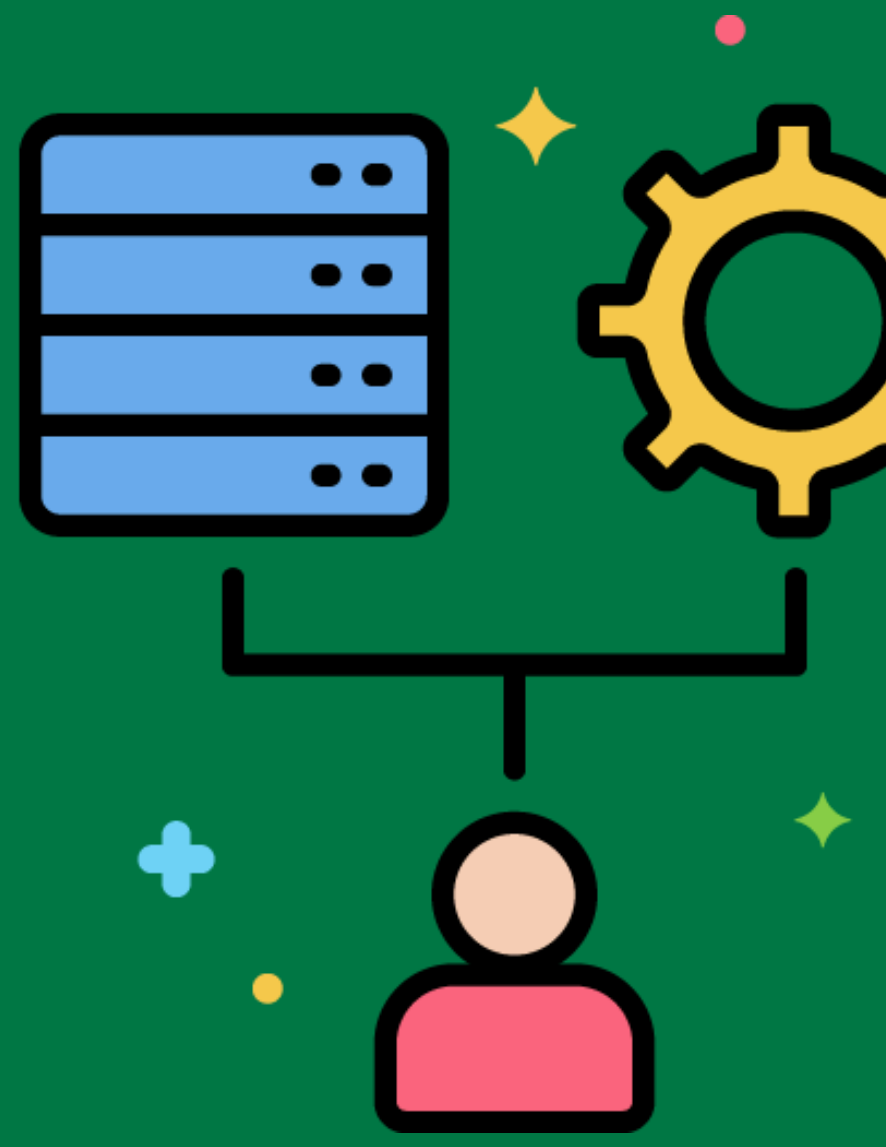
Data Engineering 101 - ETL Terminology

ETL Code Review



The process of reviewing ETL code to ensure it meets quality standards and best practices.

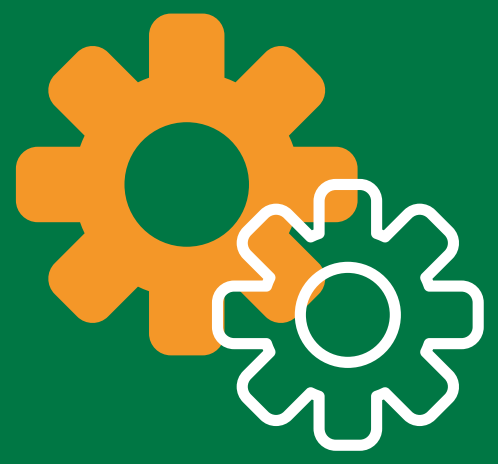
Conducting code reviews to identify and fix issues in ETL scripts before deployment.



Shwetank Singh
GritSetGrow - GSGLearn.com

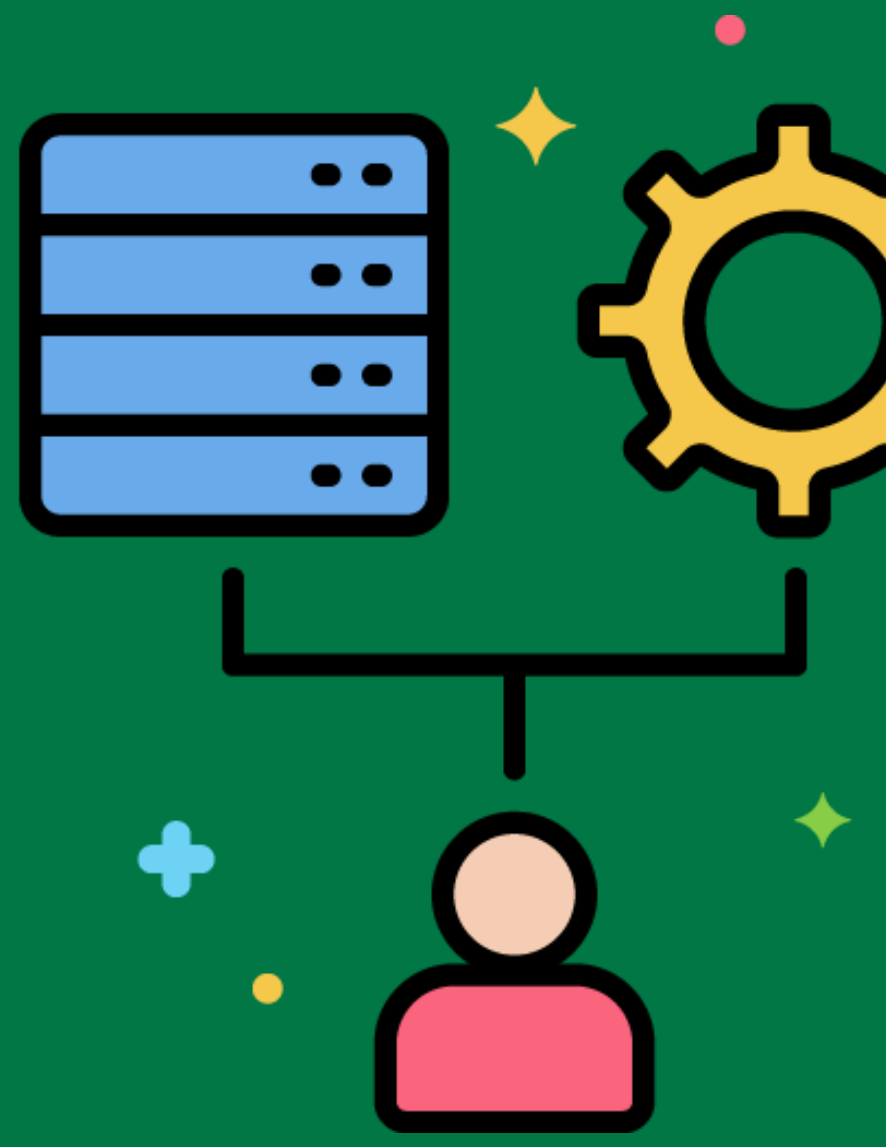
Data Engineering 101 - ETL Terminology

ETL Testing Framework



A set of tools and best practices for testing ETL processes to ensure they work correctly.

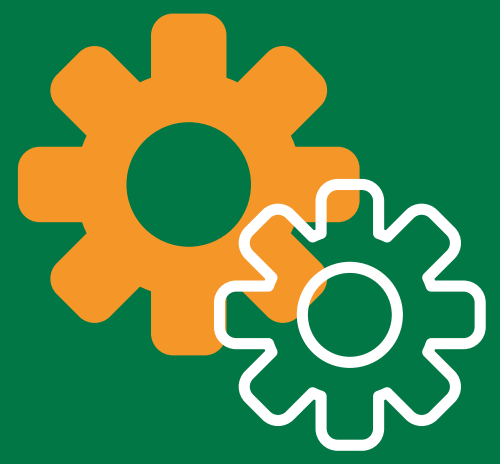
Using an ETL testing framework to automate data validation and transformation tests.



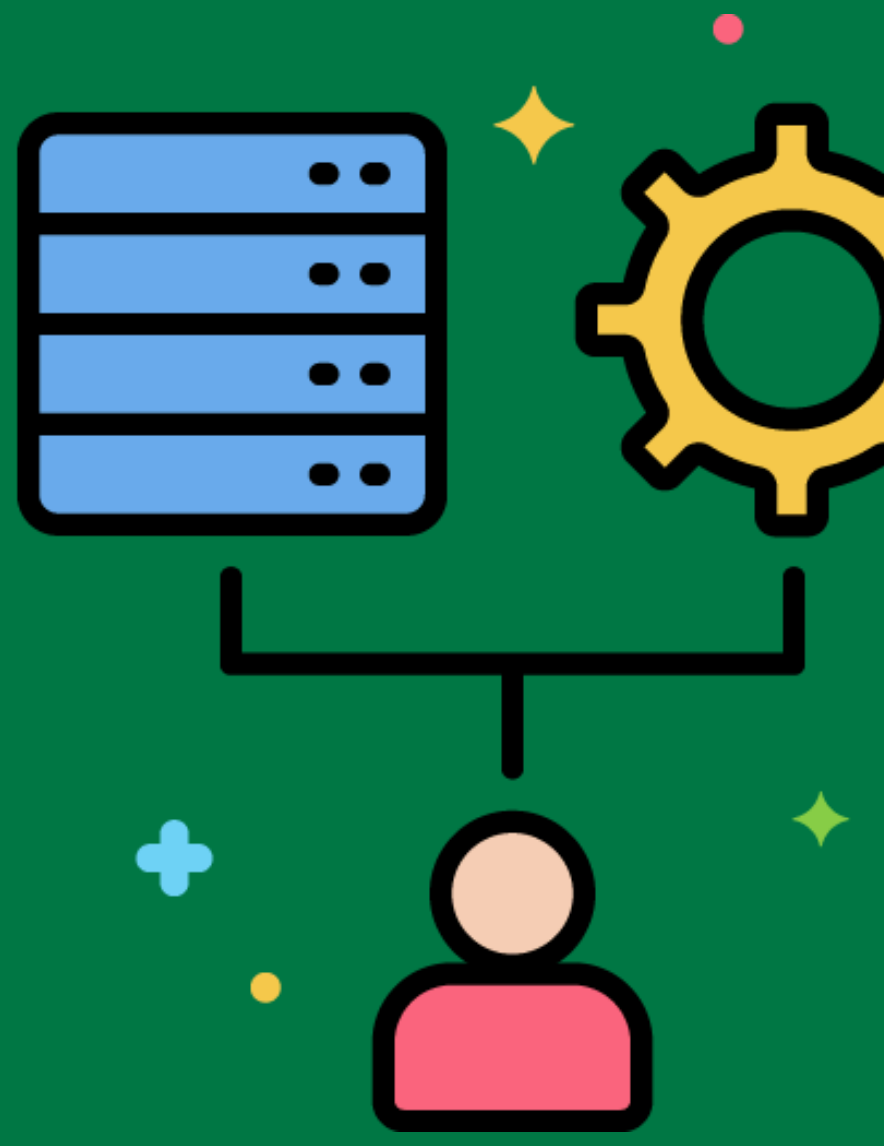
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Best Practices



Recommended approaches and techniques for designing, developing, and maintaining ETL processes.

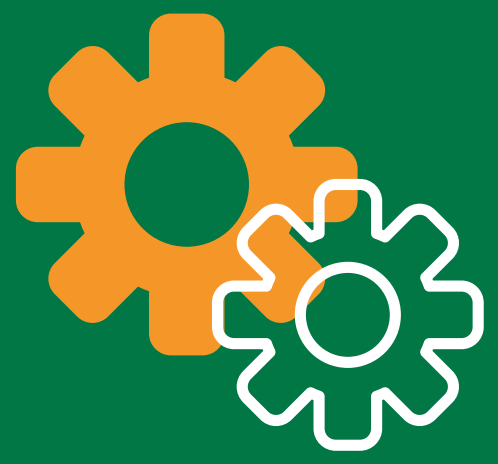


Following best practices like modularizing ETL processes, using descriptive naming conventions, and implementing robust error handling.

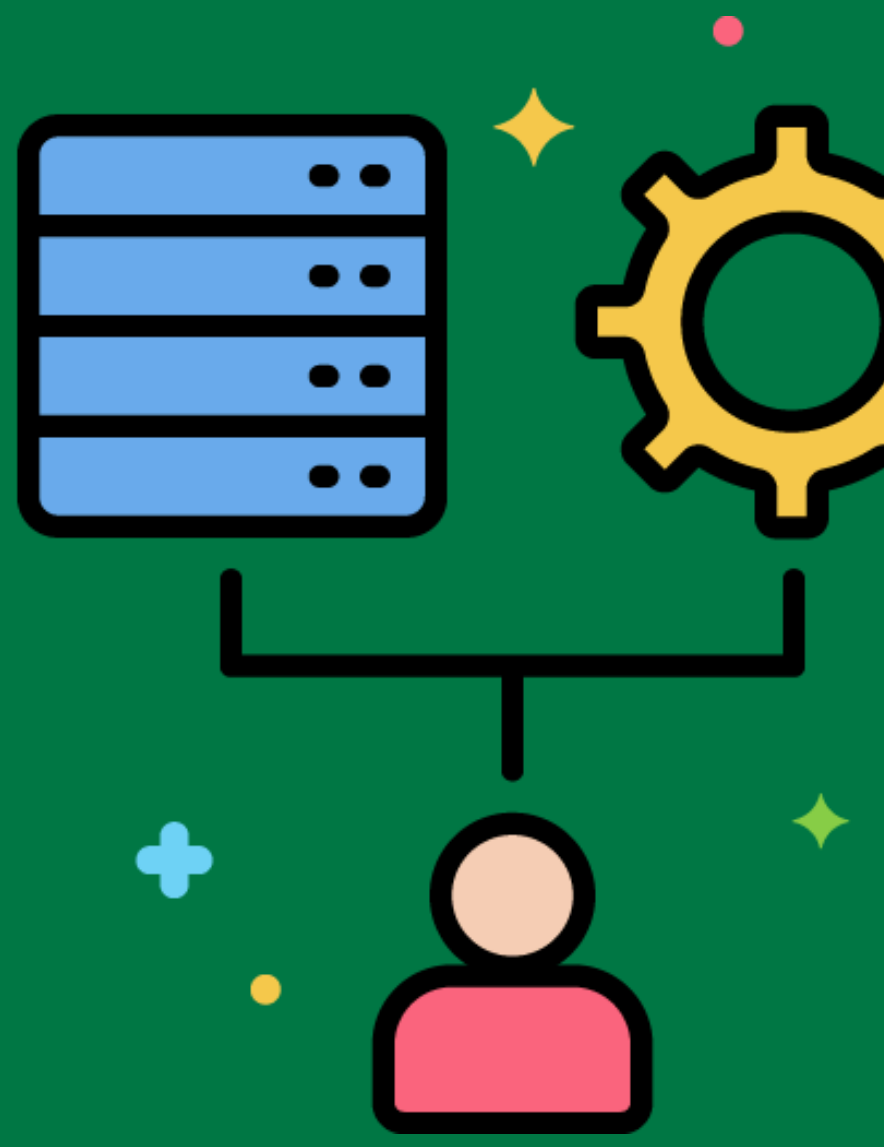
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Lifecycle



The stages involved in the development and maintenance of ETL processes, from planning to deployment and monitoring.

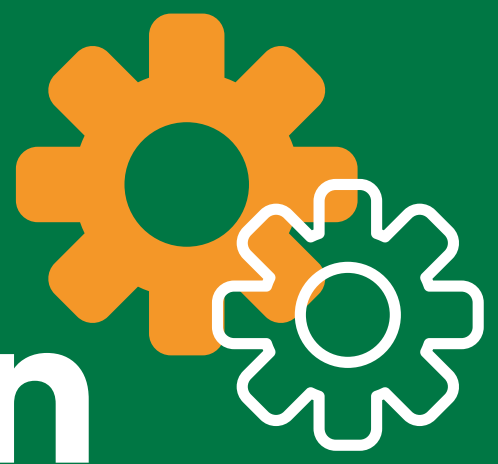


Managing the ETL lifecycle from initial requirements gathering and design to development, testing, deployment, and ongoing maintenance.

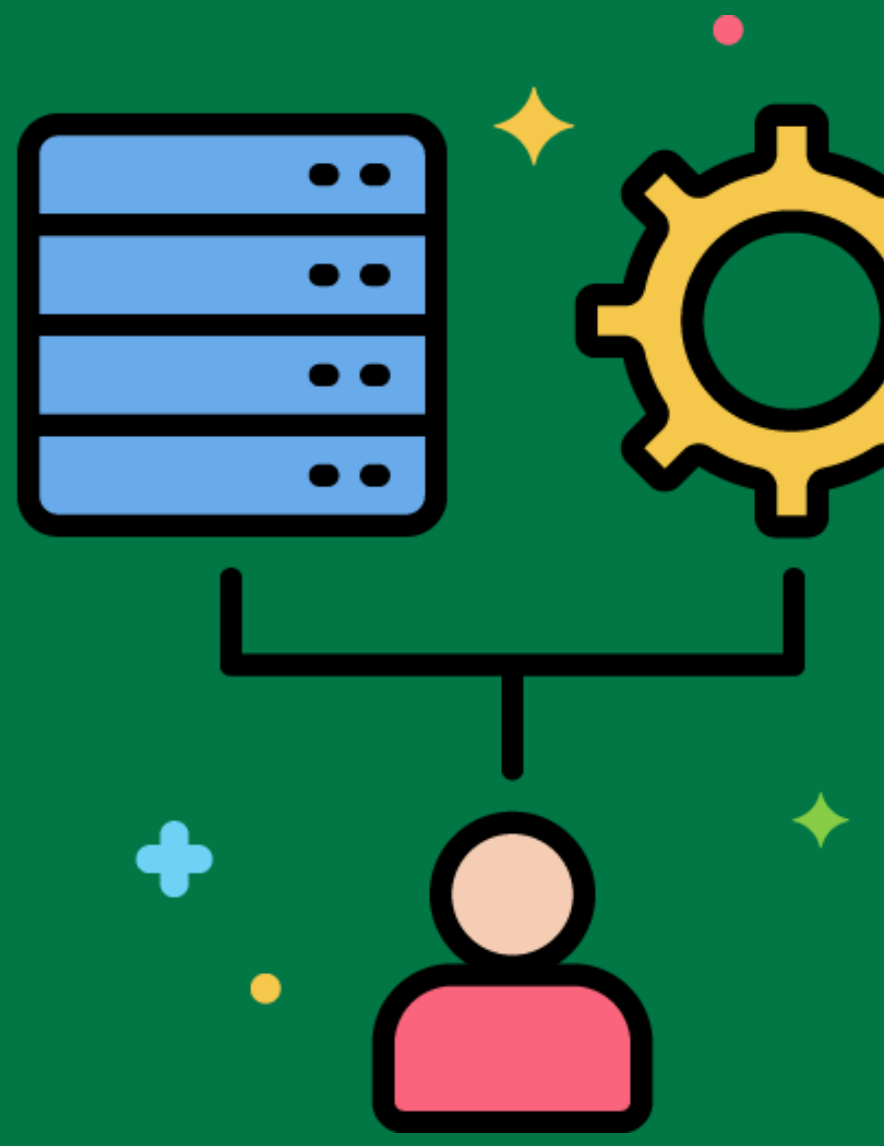
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Documentation Standards



Guidelines for creating and maintaining documentation for ETL processes.

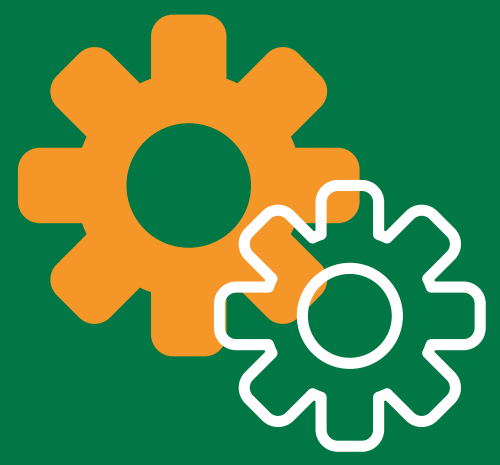


Implementing documentation standards to ensure that all ETL processes are well-documented, including data mappings, transformation rules, and job schedules.

Shwetank Singh
GritSetGrow - GSGLearn.com

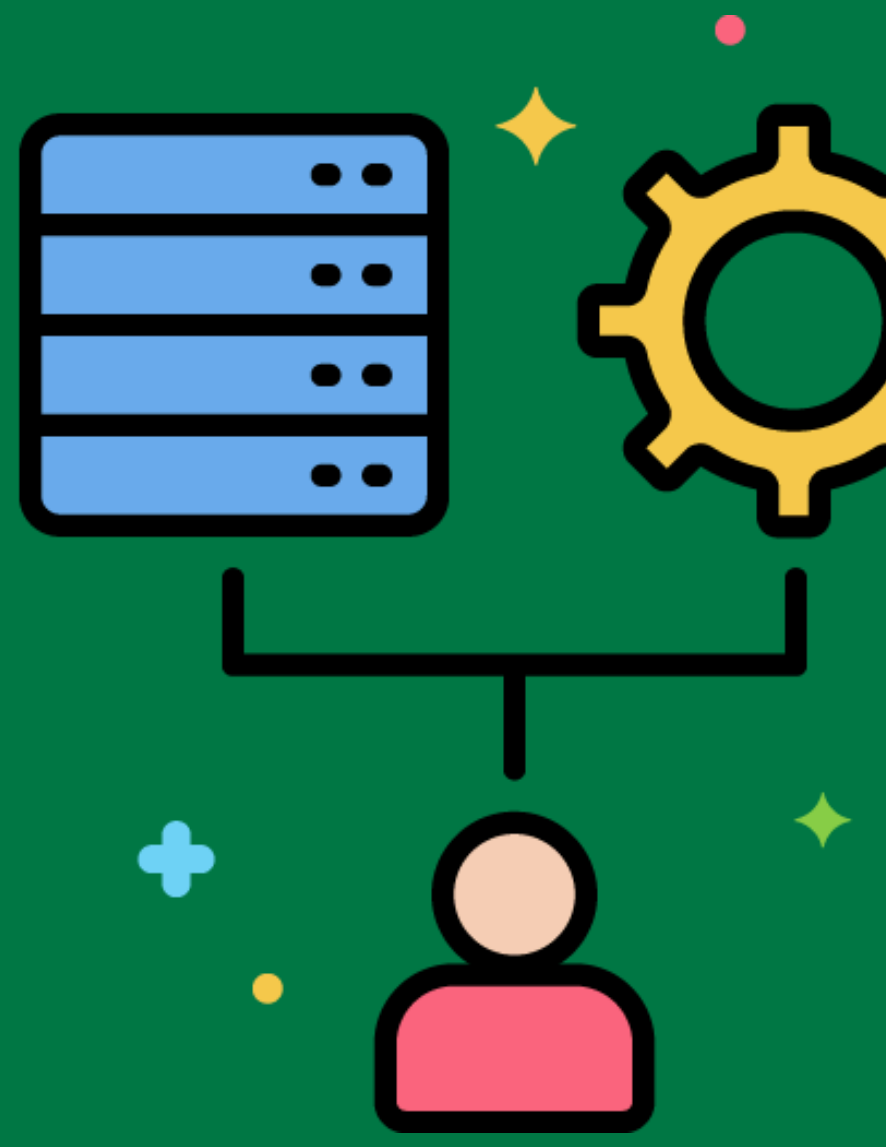
Data Engineering 101 - ETL Terminology

ETL Process Automation Tools



Tools that facilitate the automation of ETL processes, including scheduling, monitoring, and error handling.

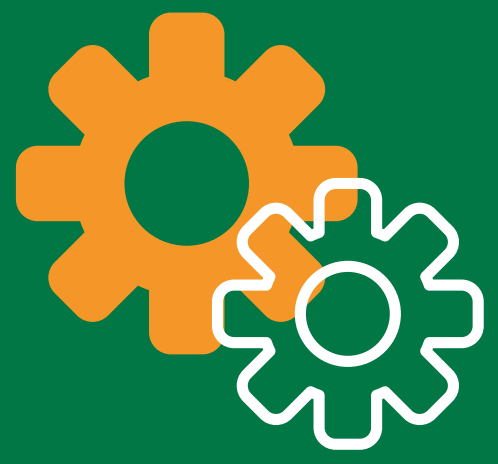
Using tools like Apache Airflow or Informatica PowerCenter to automate and manage ETL workflows.



Shwetank Singh
GritSetGrow - GSGLearn.com

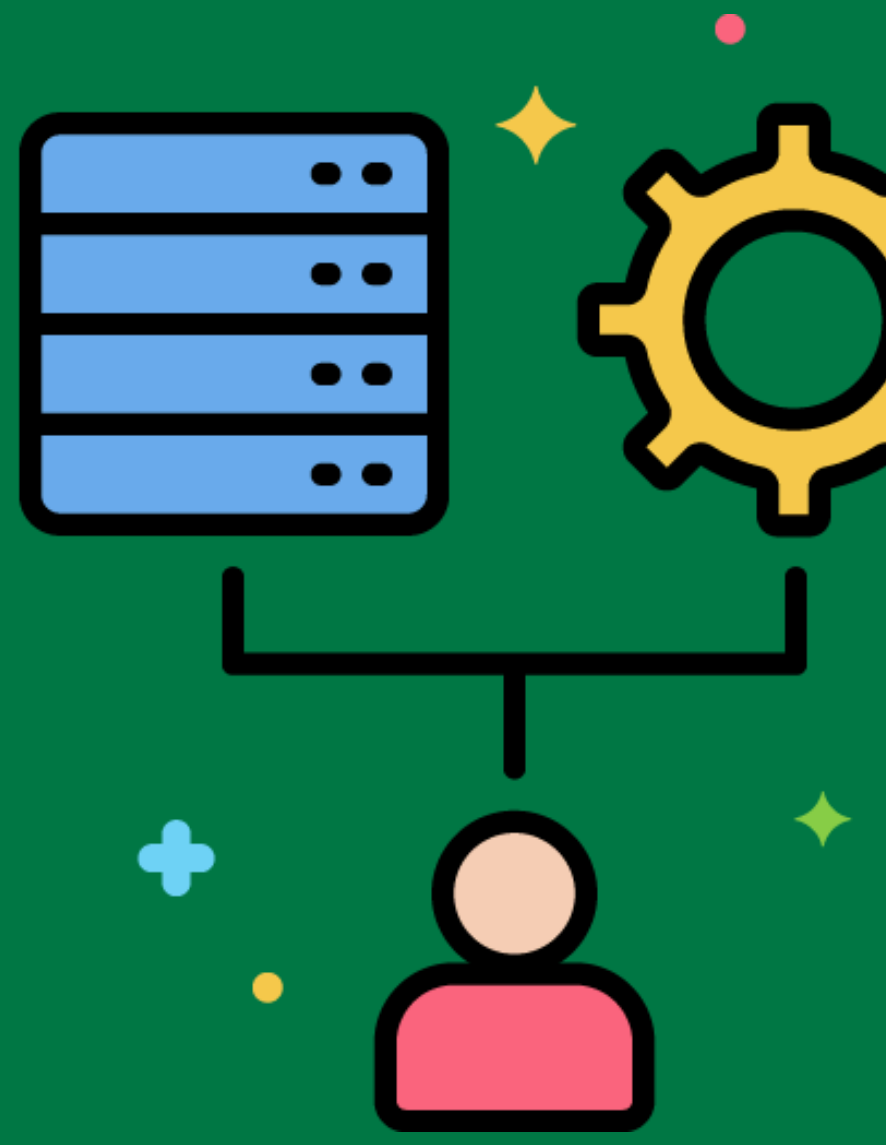
Data Engineering 101 - ETL Terminology

ETL Operational Monitoring



The continuous tracking of ETL process performance and health to ensure smooth operation.

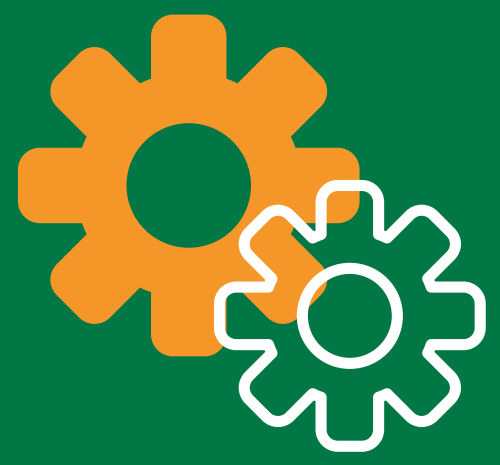
Using monitoring tools to track ETL job execution times, resource usage, and error rates.



Shwetank Singh
GritSetGrow - GSGLearn.com

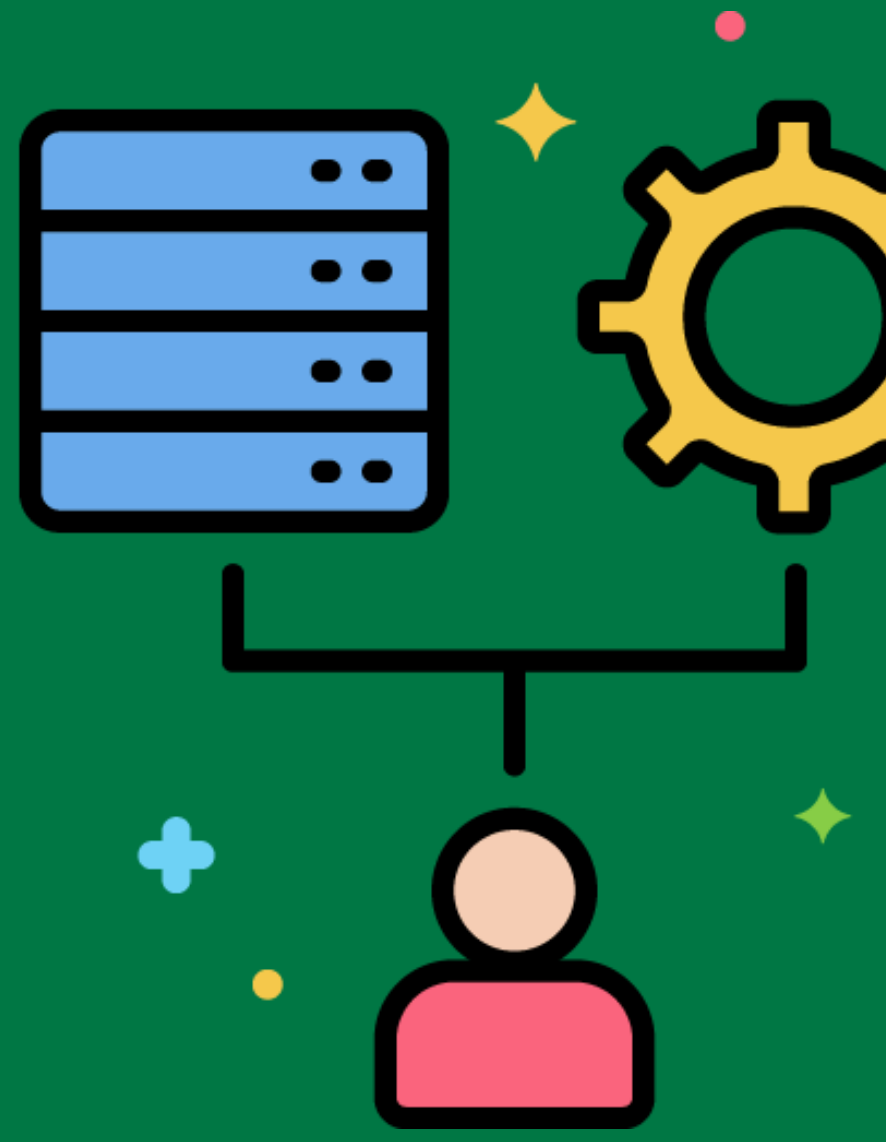
Data Engineering 101 - ETL Terminology

ETL Data Quality Framework



A structured approach to ensuring data quality in the ETL process, including validation, cleansing, and monitoring.

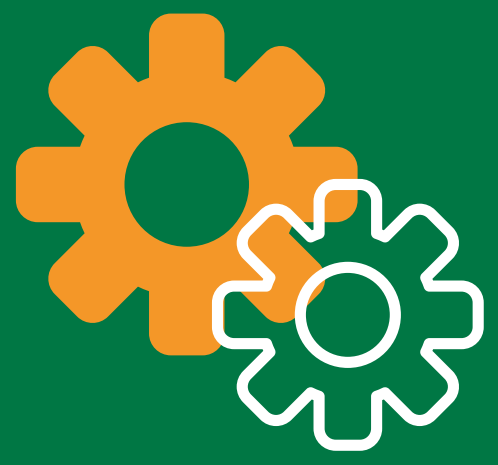
Implementing a data quality framework to apply validation rules, clean data, and monitor data quality metrics during the ETL process.



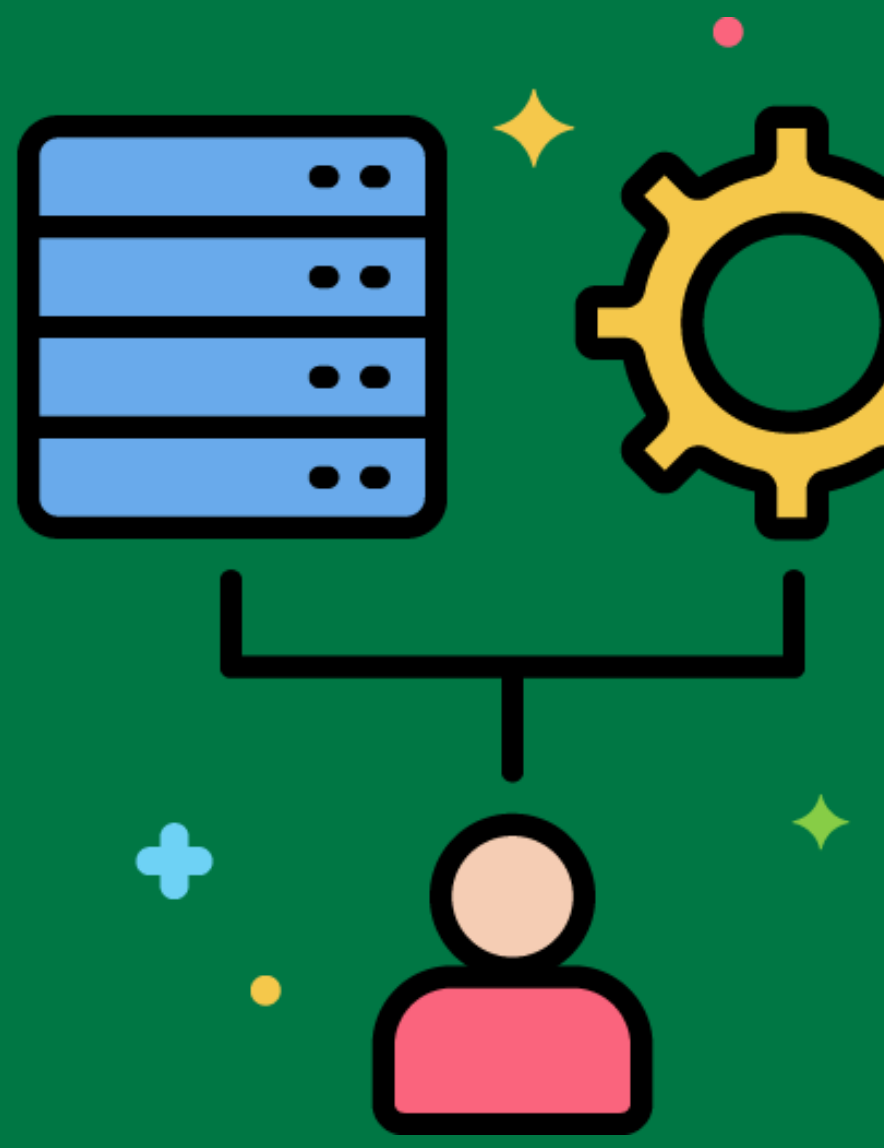
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Security Best Practices



Recommended approaches and techniques for securing ETL processes and the data they handle.

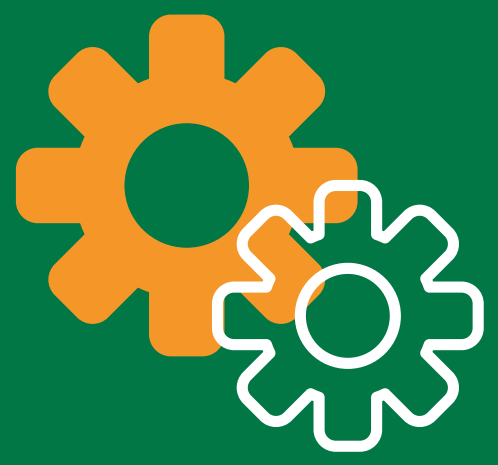


Implementing security best practices like data encryption, access controls, and secure data transfer methods.

Shwetank Singh
GritSetGrow - GSGLearn.com

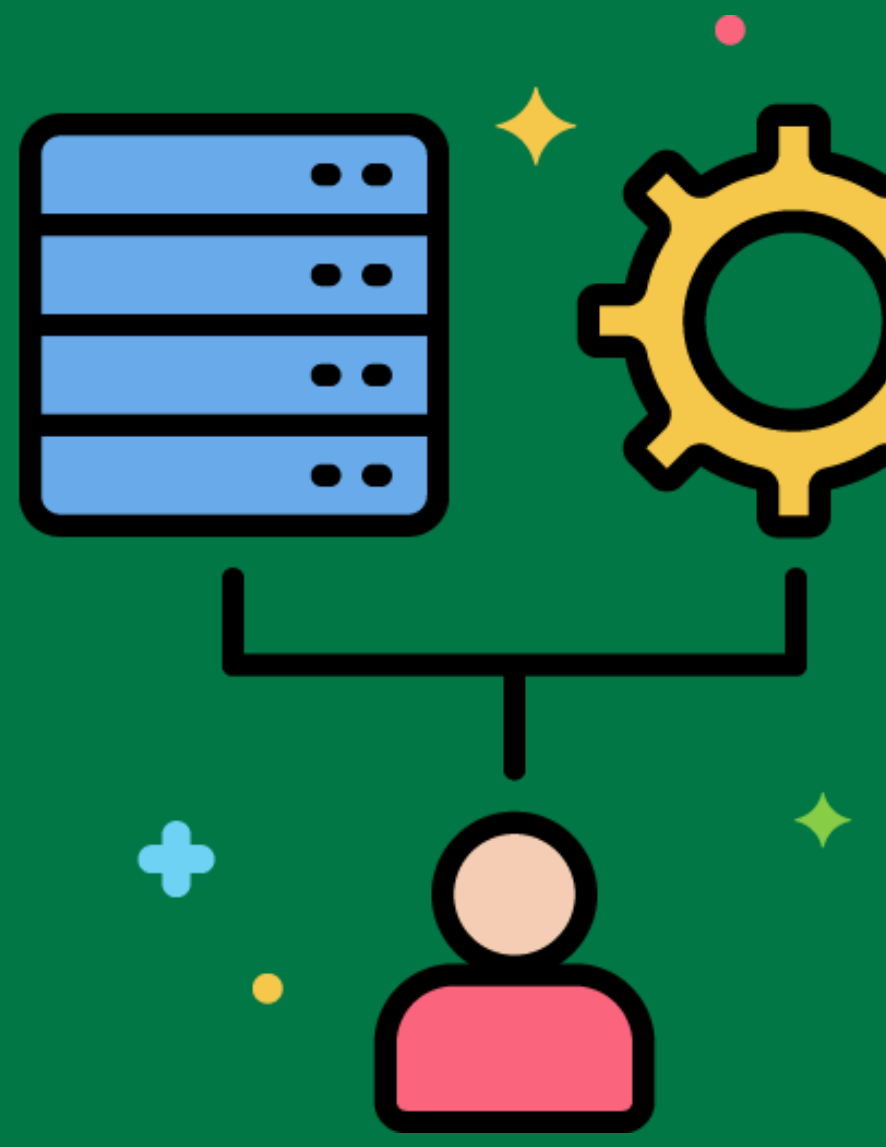
Data Engineering 101 - ETL Terminology

ETL Compliance Framework



A structured approach to ensuring that ETL processes comply with relevant regulations and standards.

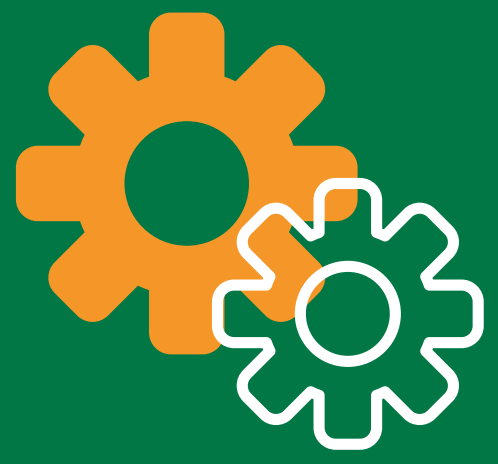
Implementing a compliance framework to ensure that ETL processes adhere to data protection regulations like GDPR and HIPAA.



Shwetank Singh
GritSetGrow - GSGLearn.com

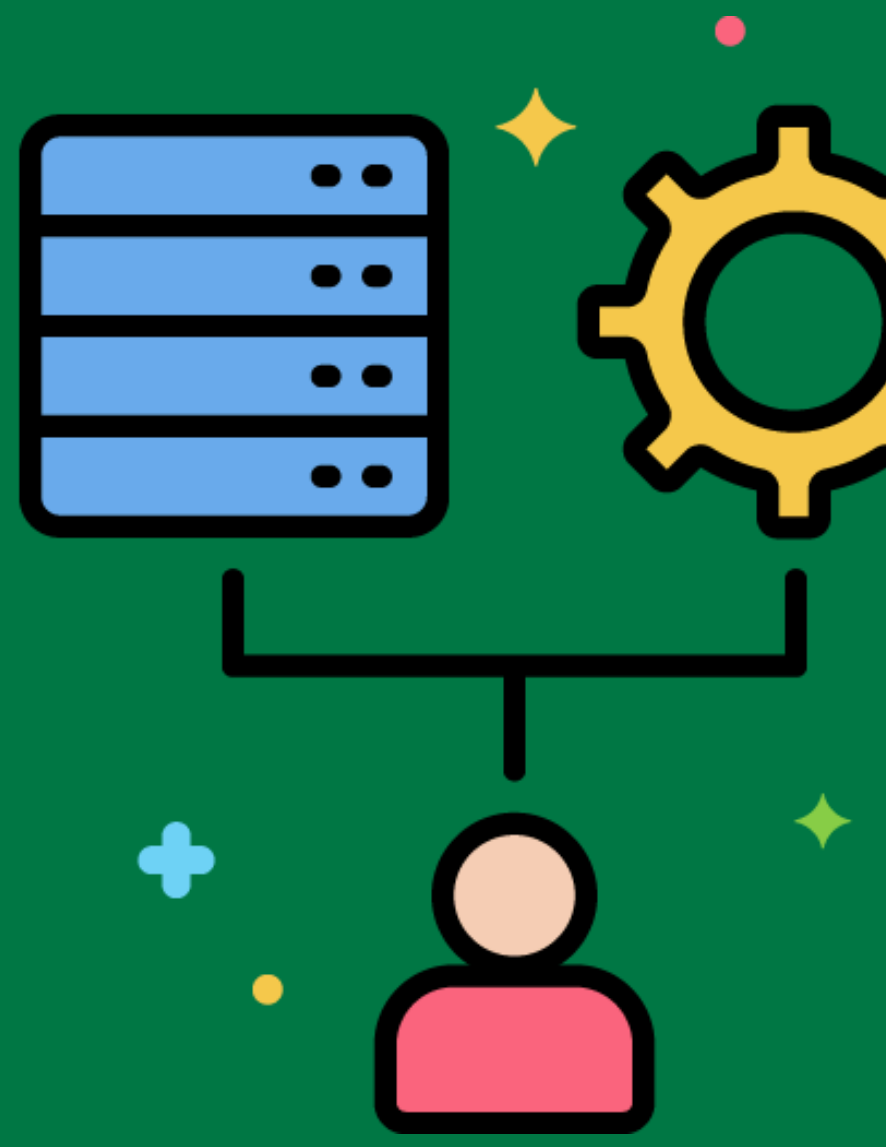
Data Engineering 101 - ETL Terminology

ETL Auditing and Compliance



Practices for tracking and documenting ETL process activities to meet regulatory requirements.

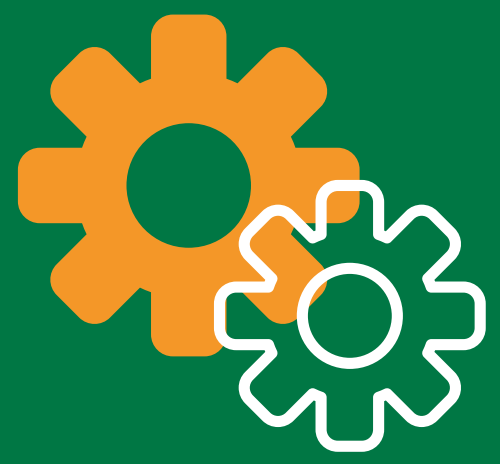
Implementing auditing mechanisms to log all data transformations and ensure compliance with regulations.



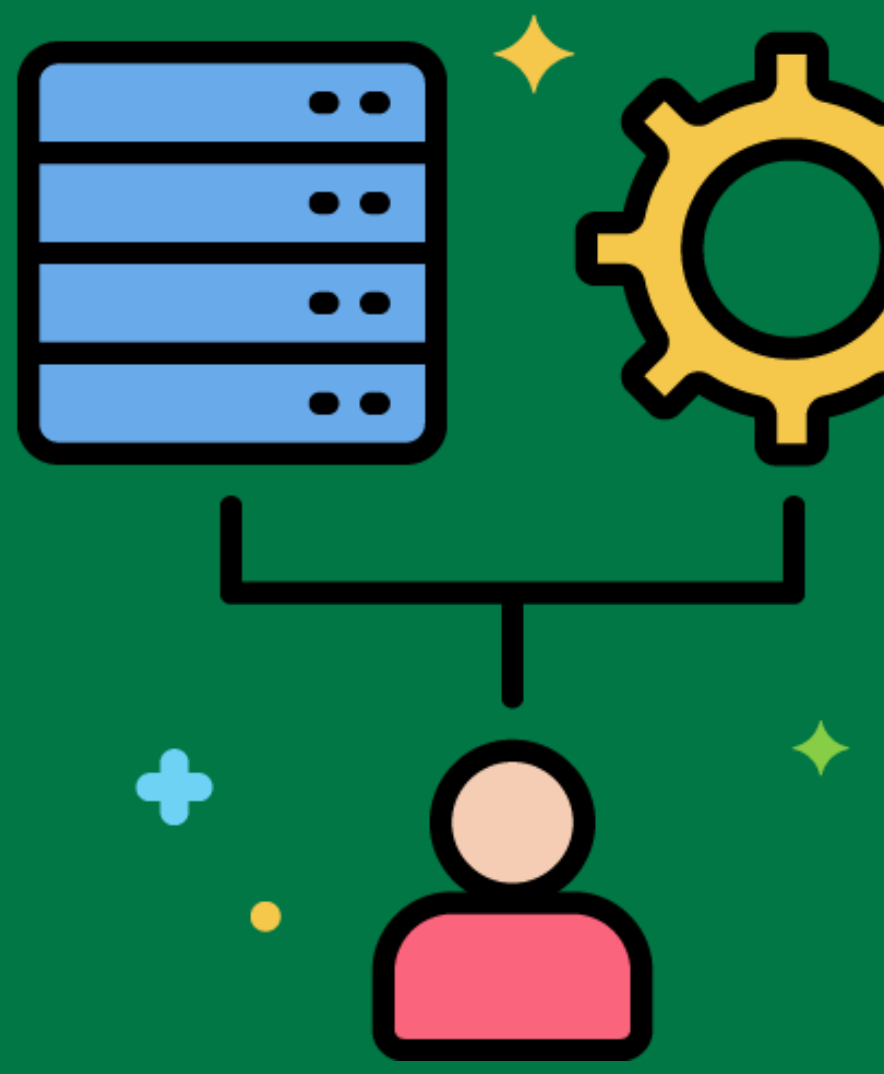
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Integration Best Practices



Recommended approaches and techniques for integrating data from multiple sources into a unified view.

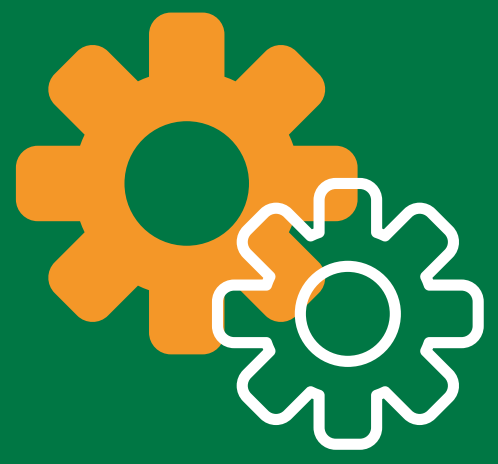


Following best practices for data integration, such as standardizing data formats, resolving data conflicts, and ensuring data consistency.

Shwetank Singh
GritSetGrow - GSGLearn.com

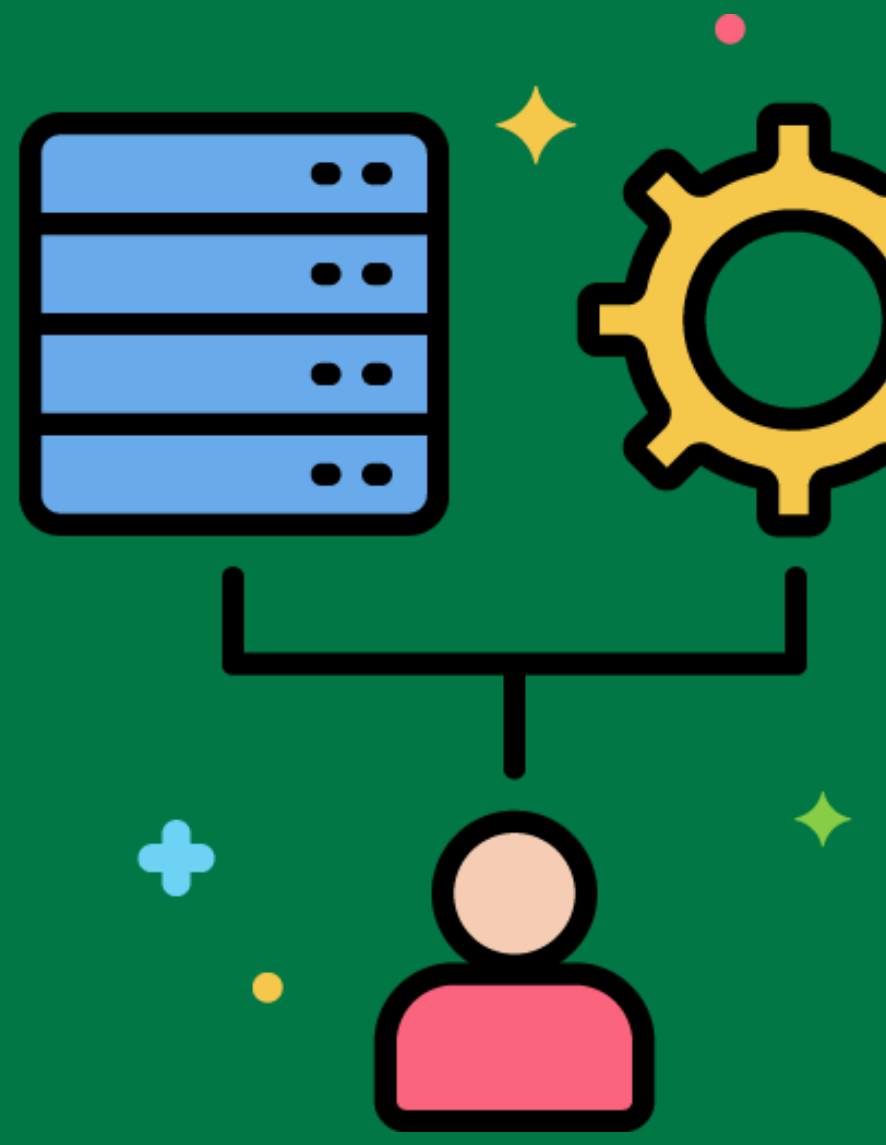
Data Engineering 101 - ETL Terminology

ETL Performance Optimization



Techniques for improving the performance and efficiency of ETL processes.

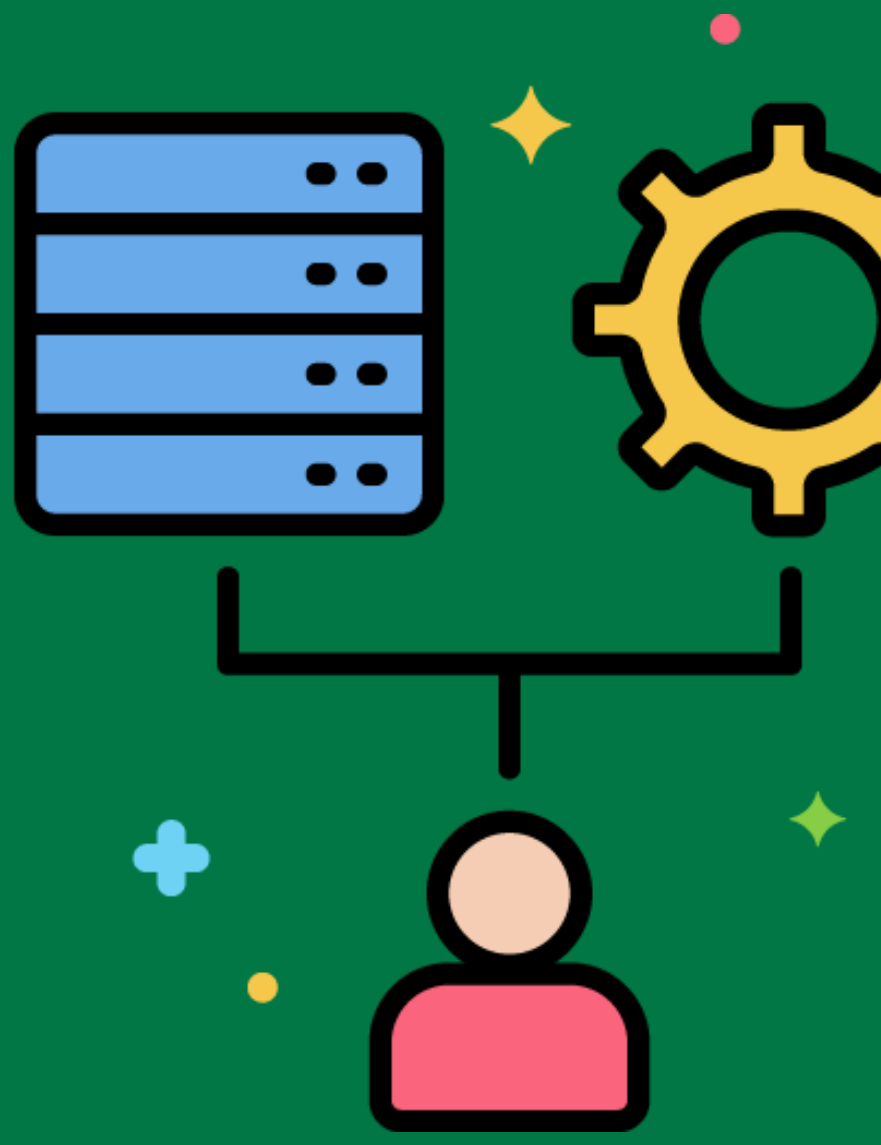
Tuning SQL queries, optimizing data transformations, and using parallel processing to speed up ETL processes.



Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Transformation Best Practices



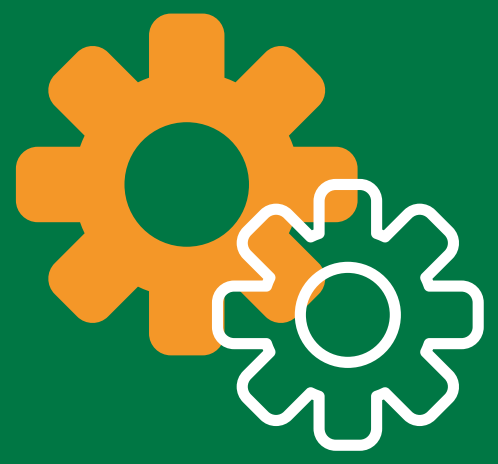
Recommended approaches and techniques for transforming data during the ETL process.

Following best practices for data transformation, such as using reusable transformation functions and ensuring data integrity.

Shwetank Singh
GritSetGrow - GSGLearn.com

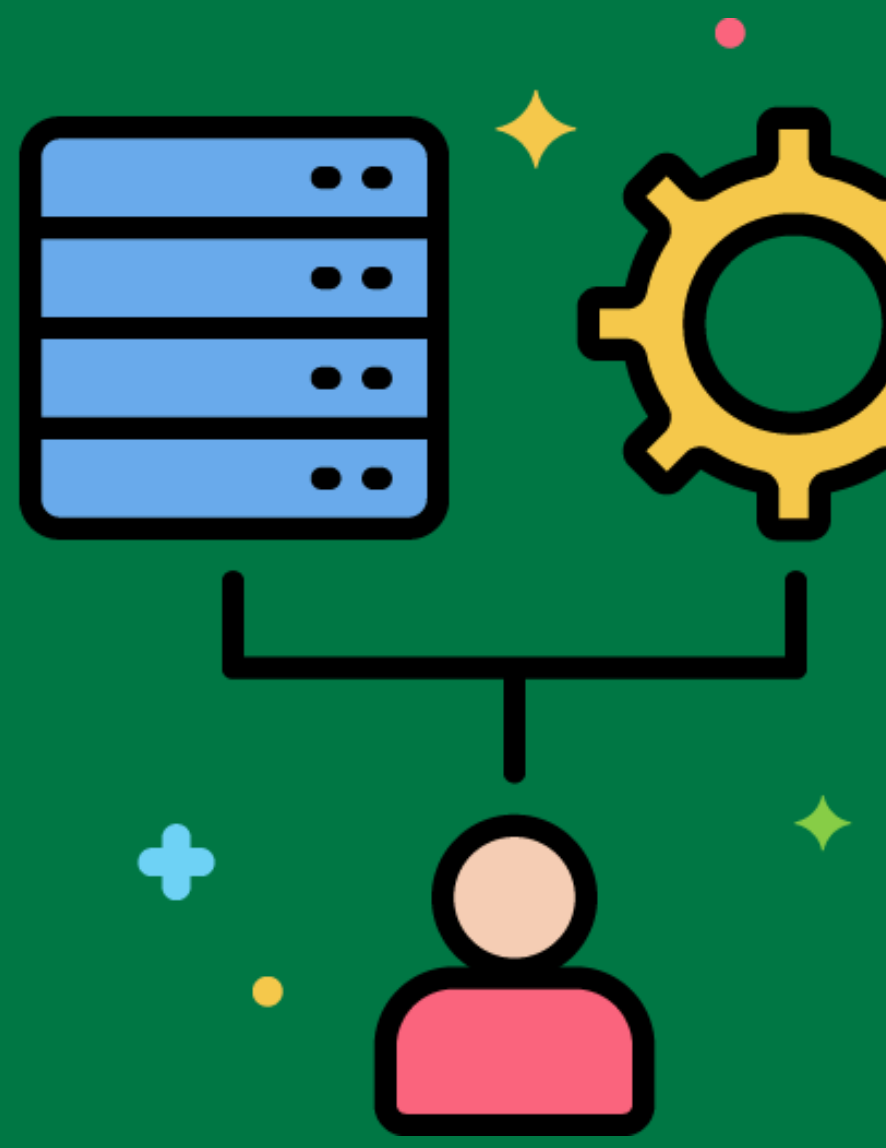
Data Engineering 101 - ETL Terminology

ETL Job Scheduling



Planning and managing the execution of ETL jobs to ensure they run at the right time and in the right order.

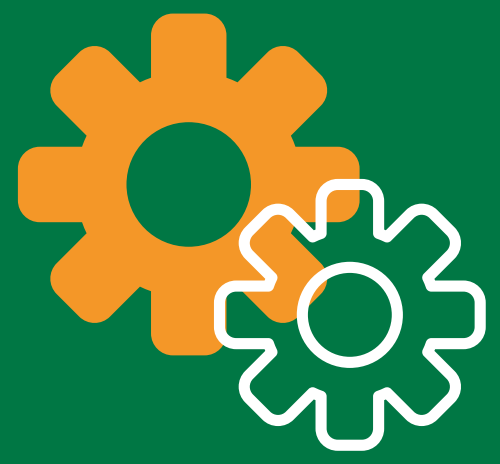
Using a job scheduler to run ETL jobs during off-peak hours to minimize impact on source systems.



Shwetank Singh
GritSetGrow - GSGLearn.com

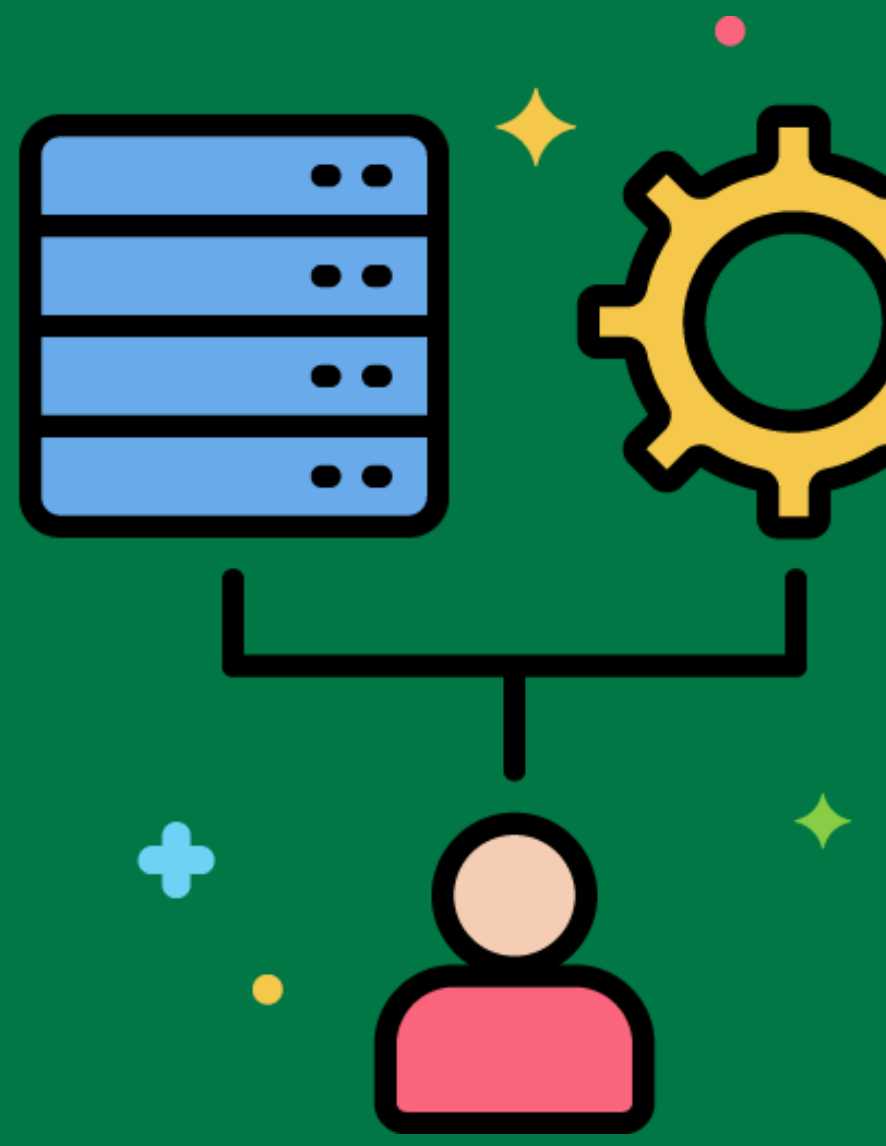
Data Engineering 101 - ETL Terminology

ETL Process Orchestration



Coordinating and managing the execution of multiple ETL processes to ensure they work together seamlessly.

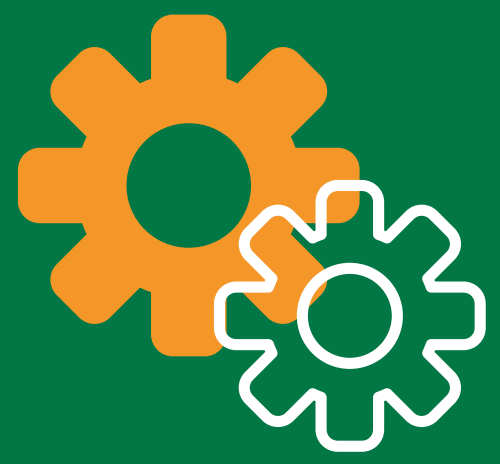
Using an orchestration tool to manage the dependencies between ETL jobs and ensure they run in the correct order.



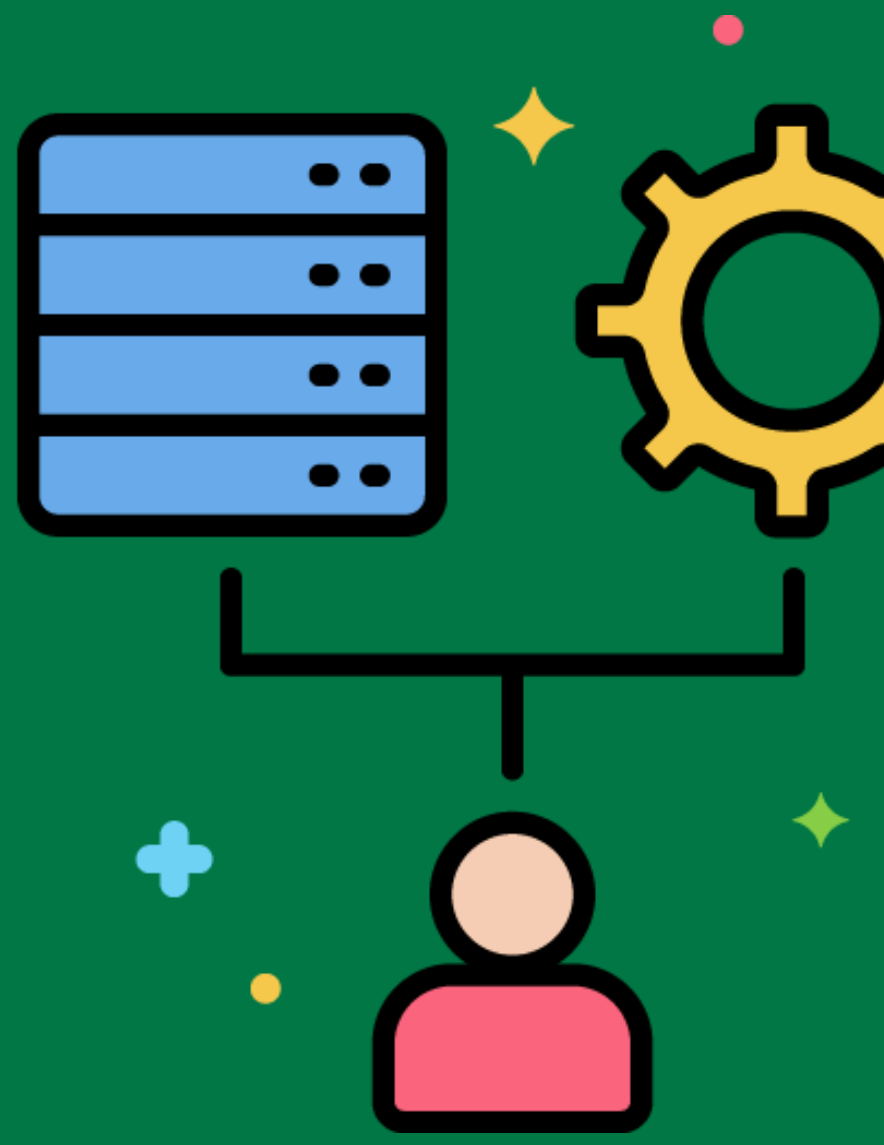
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Validation Best Practices



Recommended approaches and techniques for validating data during the ETL process.



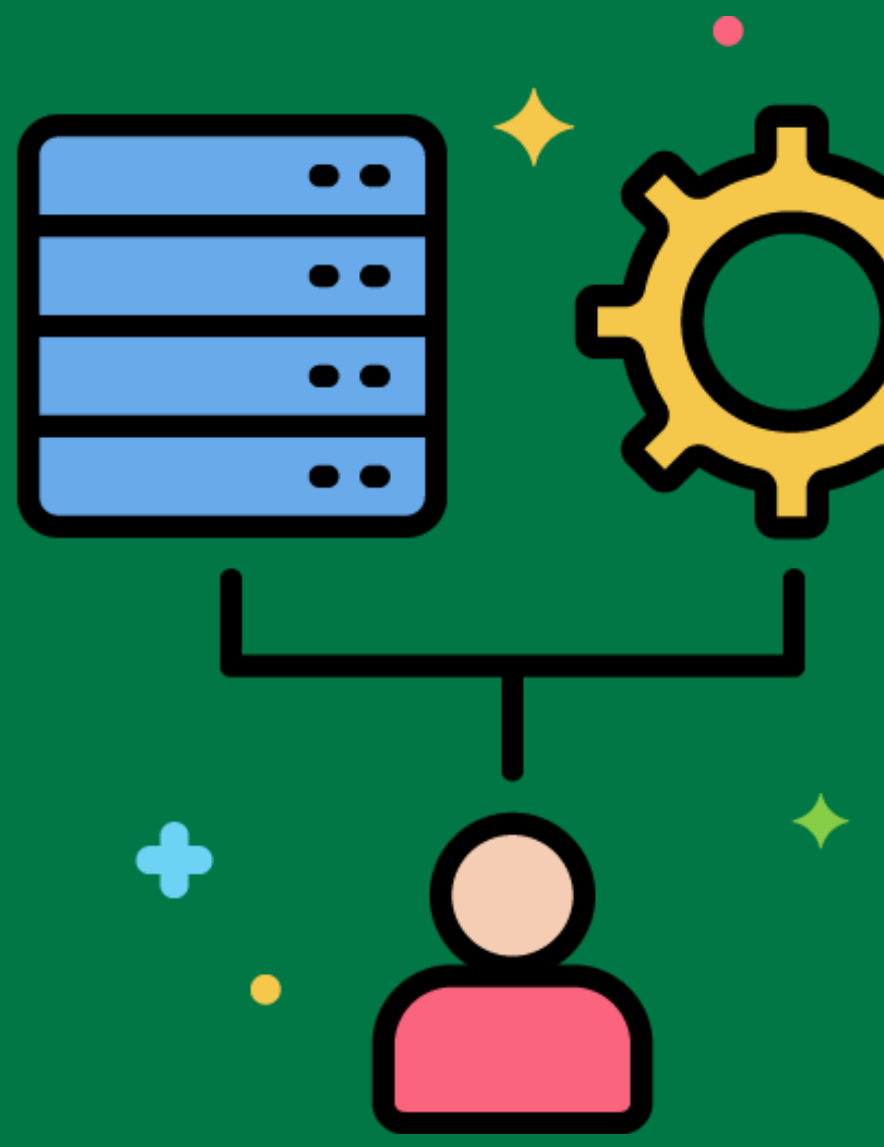
Following best practices for data validation, such as implementing validation rules to check data accuracy and completeness.

Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Transformation Testing

The process of verifying that data transformations are applied correctly and produce the expected results.

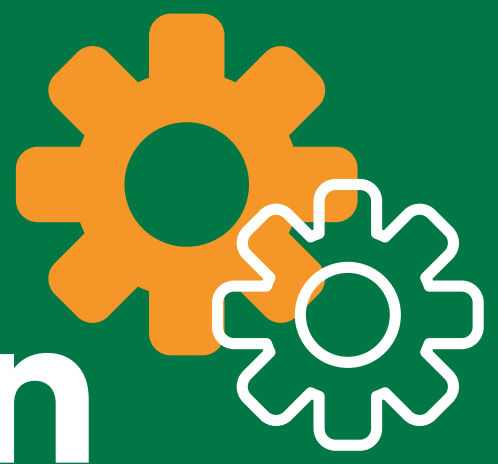


Conducting transformation tests to ensure that data is transformed correctly according to business rules.

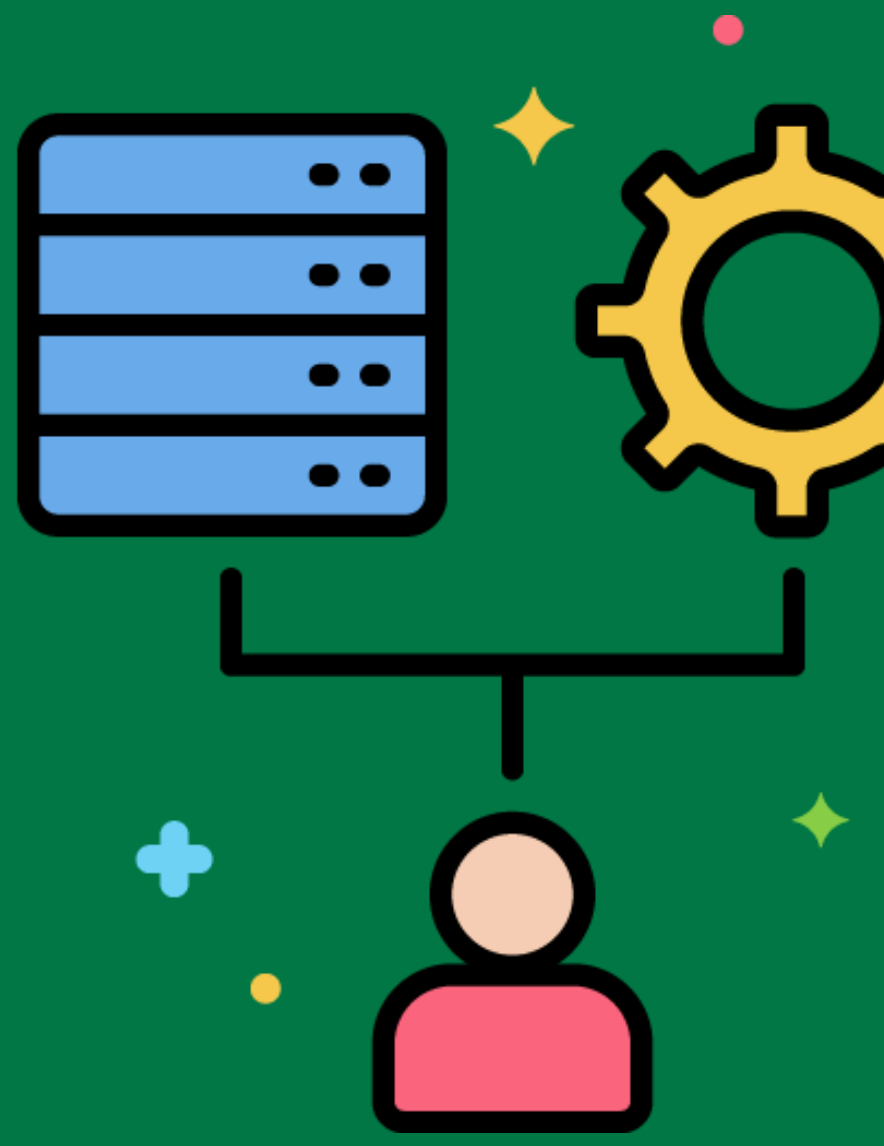
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Process Documentation



Creating and maintaining detailed documentation for ETL processes, including data mappings, transformation rules, and job schedules.

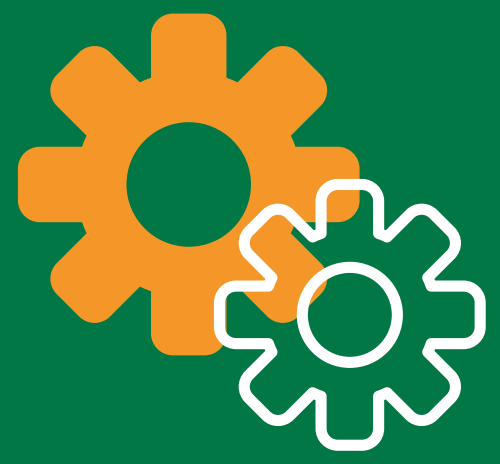


Documenting the entire ETL process, from data extraction to transformation and loading, to ensure clarity and maintainability.

Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Code Quality



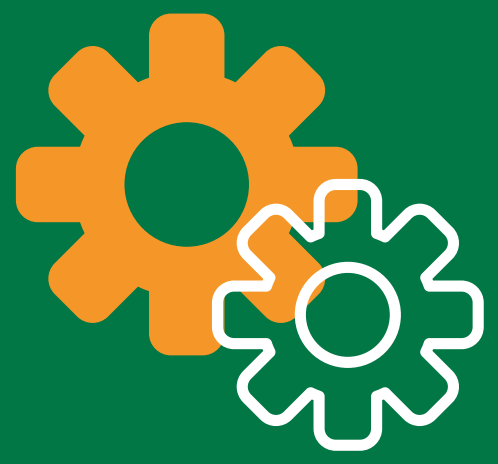
Ensuring that the code used in ETL processes meets quality standards and best practices.

Conducting code reviews and using static code analysis tools to ensure ETL code quality.

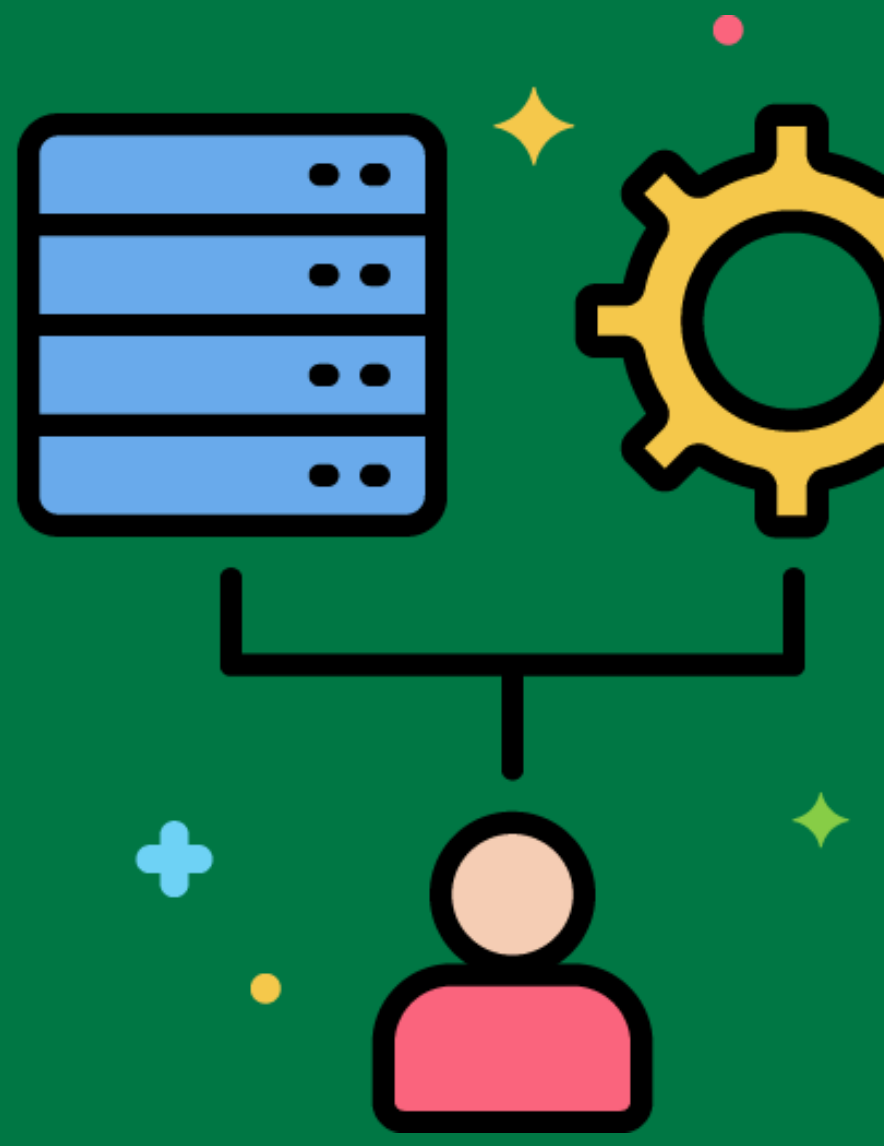
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Process Governance



Managing and overseeing the ETL process to ensure it meets organizational standards and objectives.

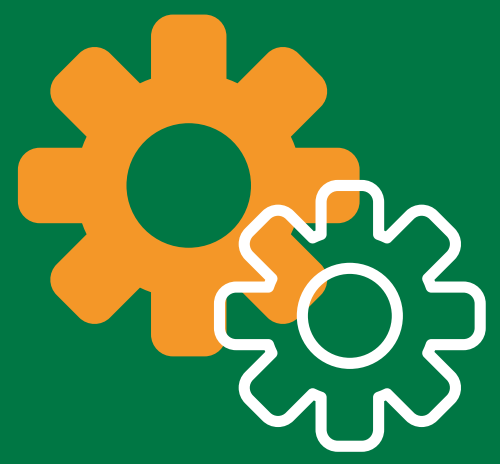


Implementing governance policies to ensure that ETL processes are aligned with business goals and comply with regulatory requirements.

Shwetank Singh
GritSetGrow - GSGLearn.com

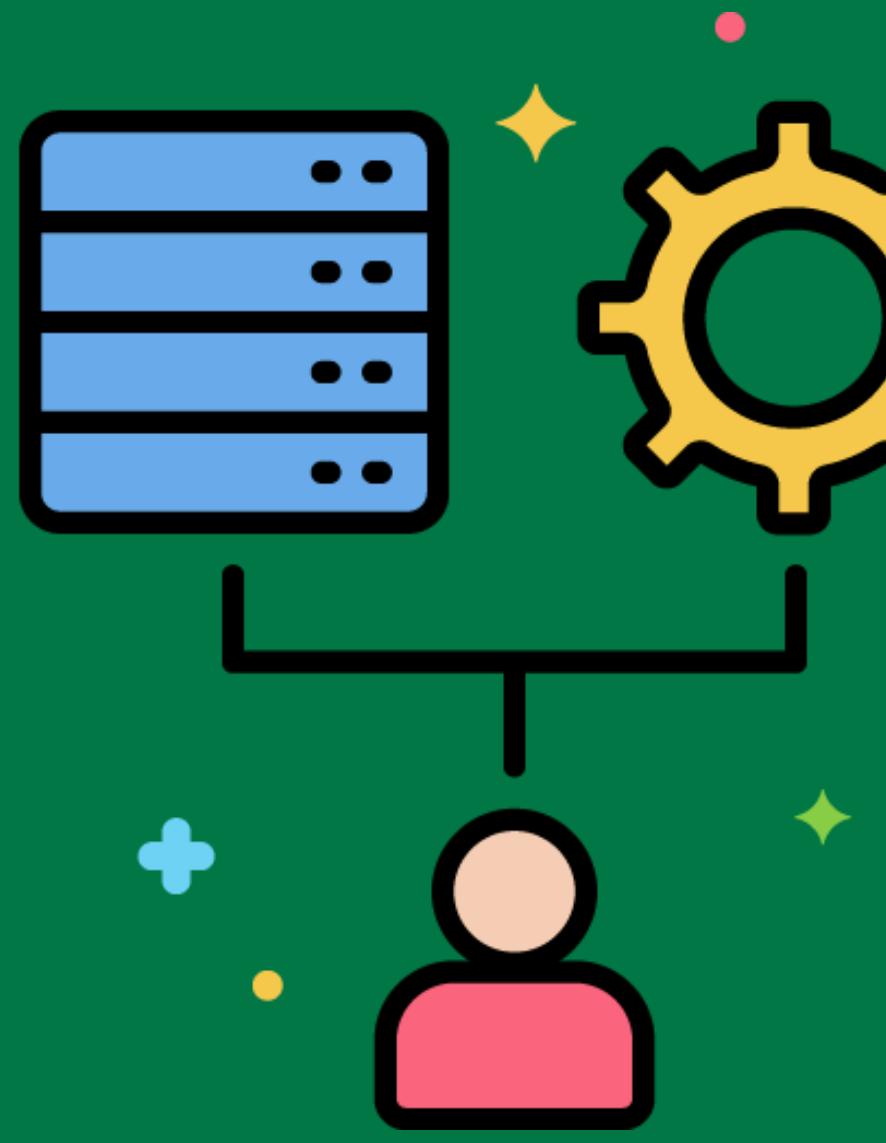
Data Engineering 101 - ETL Terminology

ETL Job Scheduling Best Practices



Recommended approaches and techniques for scheduling ETL jobs to ensure they run efficiently and reliably.

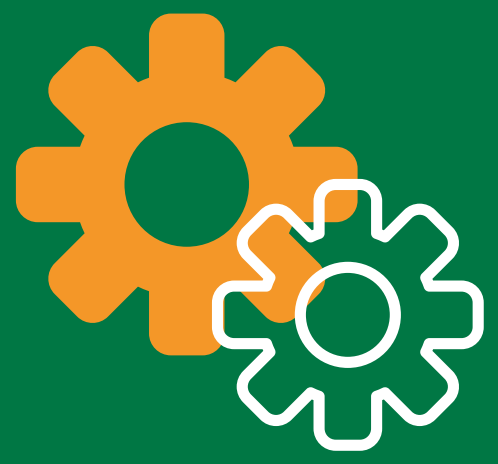
Following best practices for job scheduling, such as defining job dependencies and using time-based scheduling strategies.



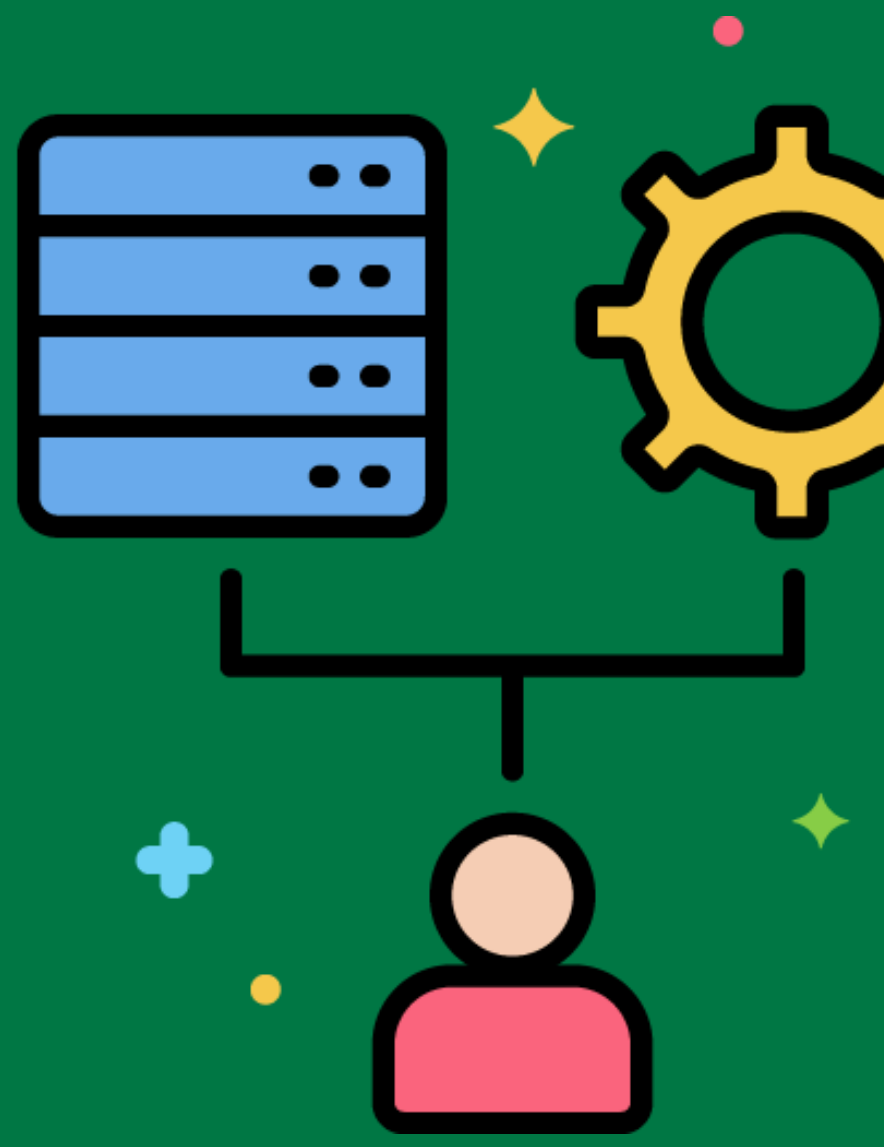
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Process Resilience



Ensuring that ETL processes can withstand and recover from failures and disruptions.

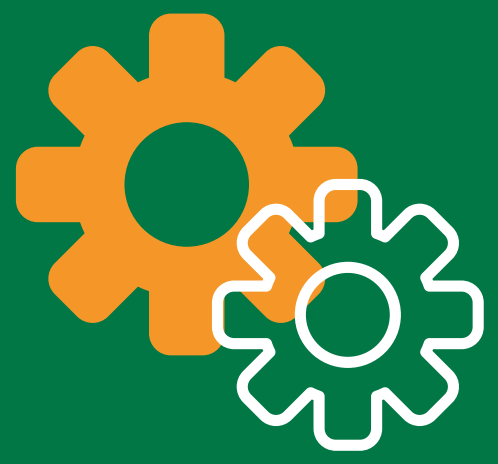


Implementing checkpointing and retry mechanisms to improve ETL process resilience.

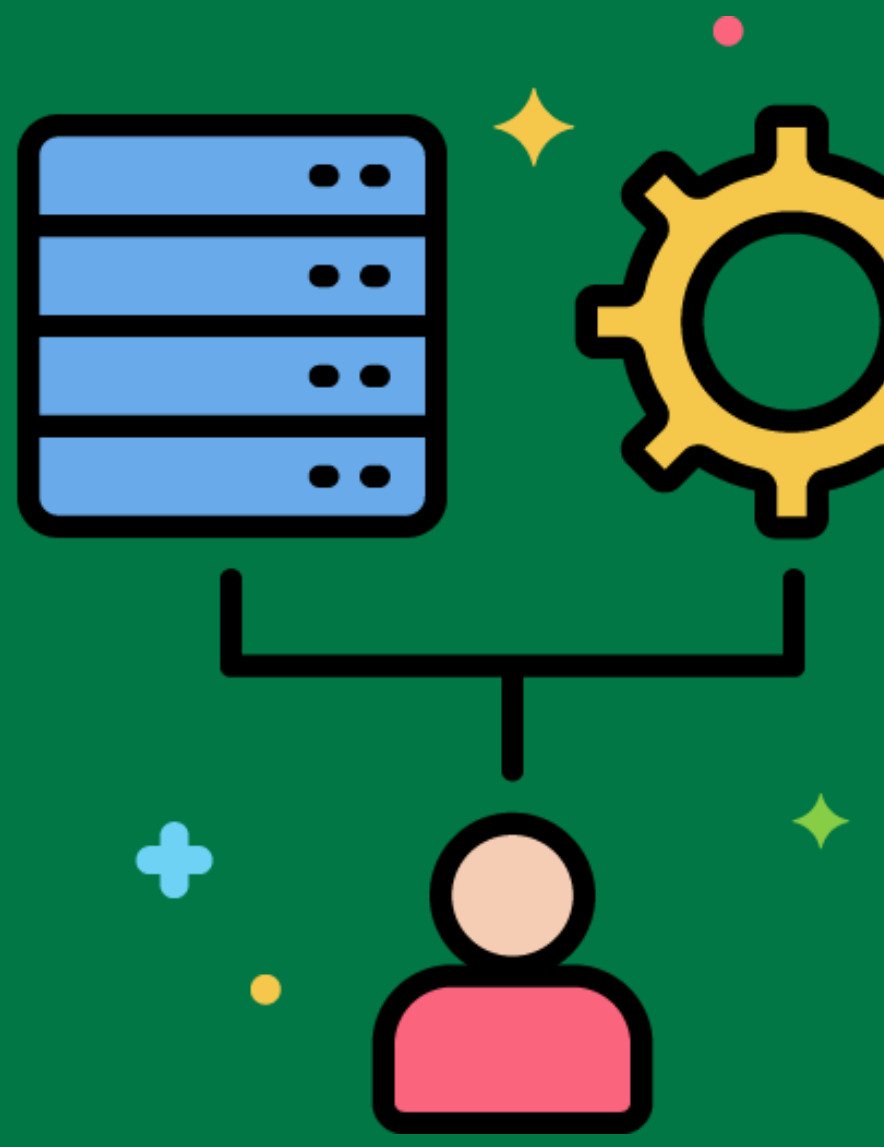
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Process Scalability



Ensuring that ETL processes can handle increasing volumes of data and processing demands.

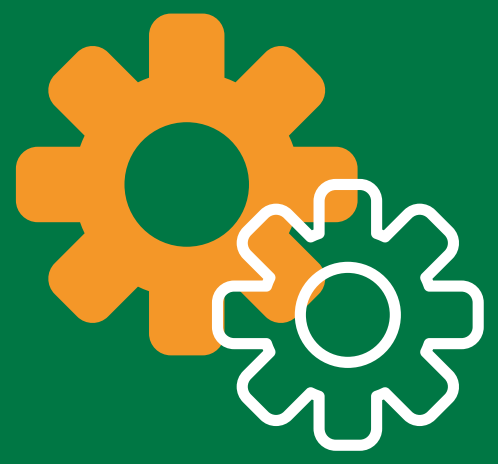


Designing ETL processes to scale horizontally by adding more servers or processes to handle growing data volumes.

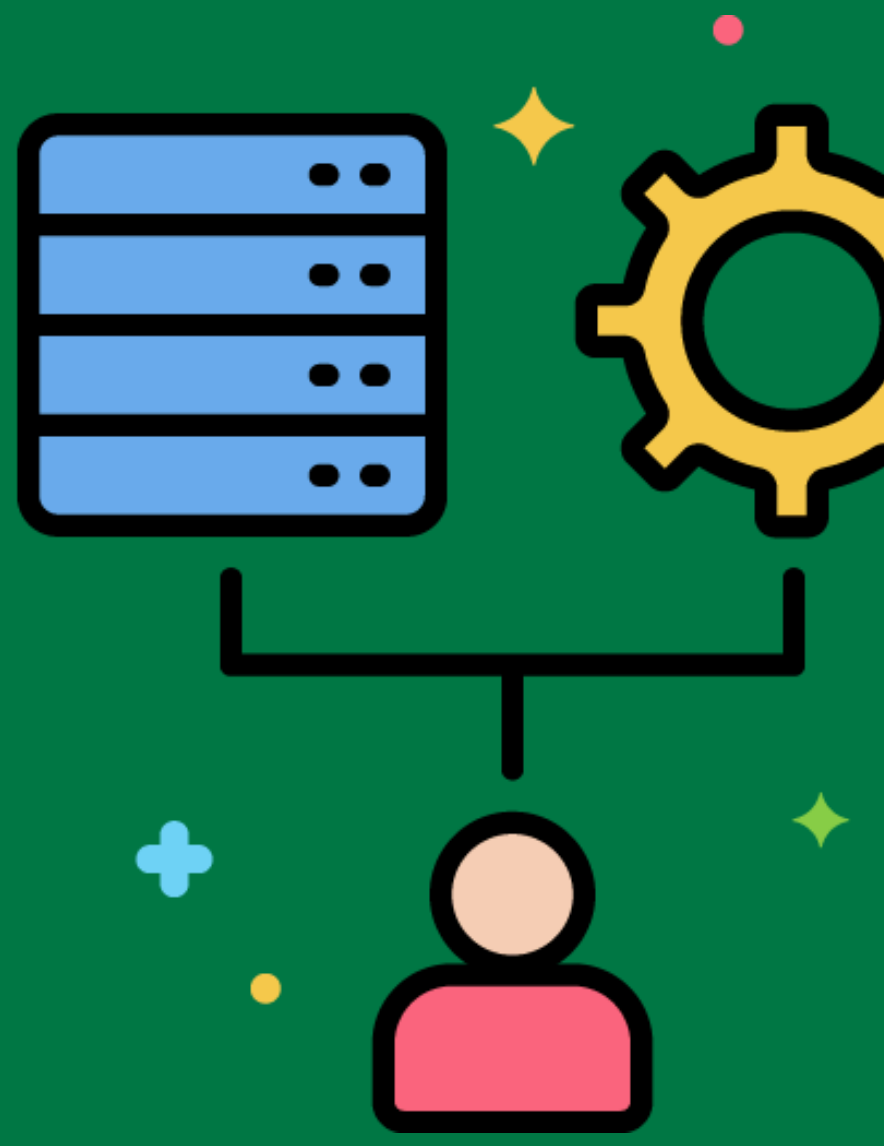
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Process Flexibility



Ensuring that ETL processes can adapt to changing business requirements and data sources.

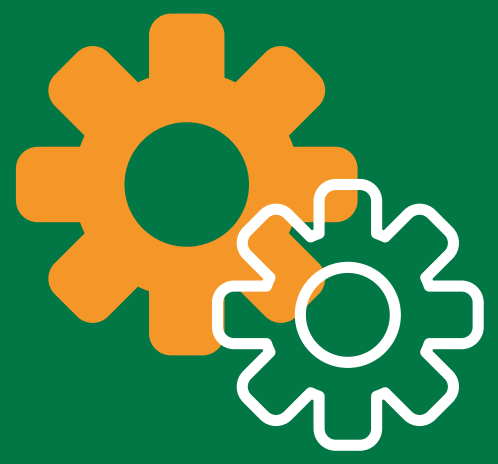


Designing ETL processes with modular components that can be easily modified or replaced to accommodate changes.

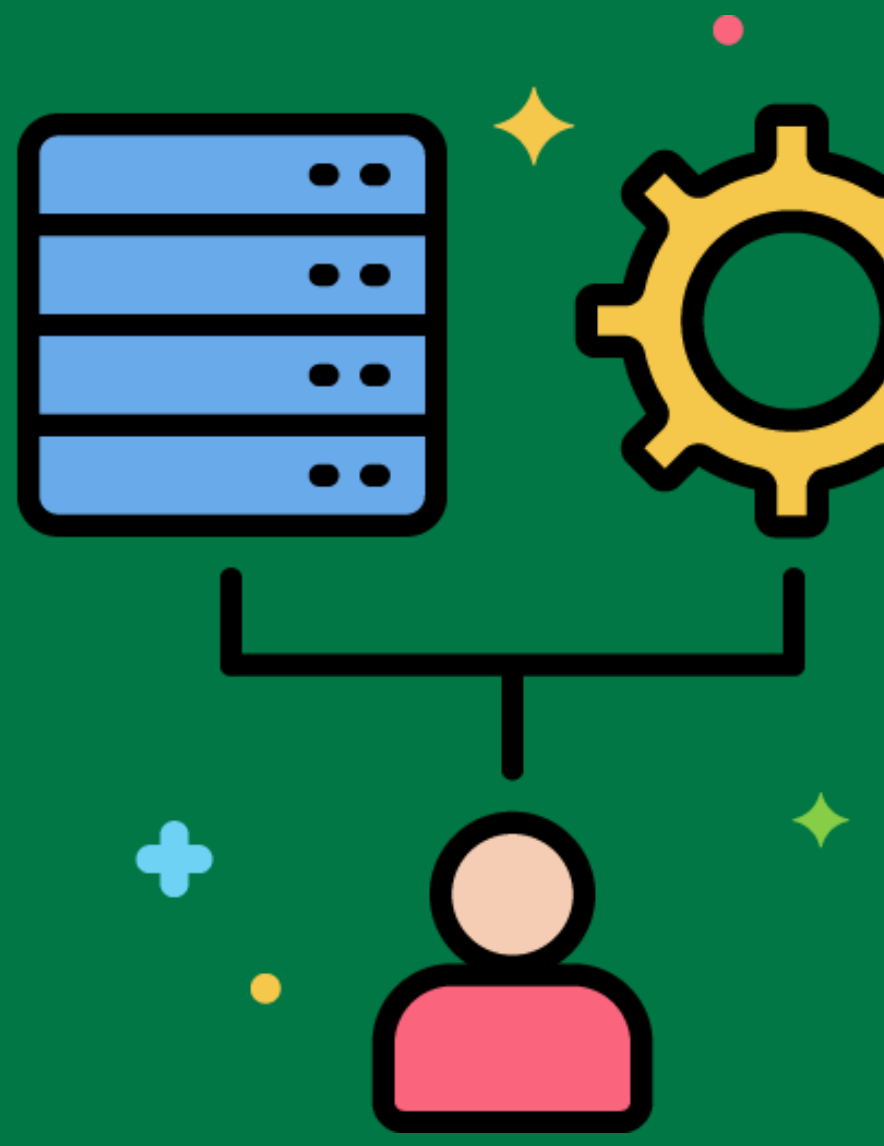
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Security



Protecting data during the ETL process to ensure its confidentiality, integrity, and availability.

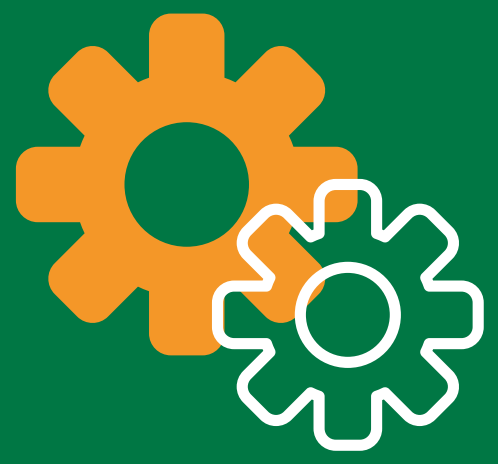


Implementing data encryption, access controls, and secure data transfer methods to protect data during the ETL process.

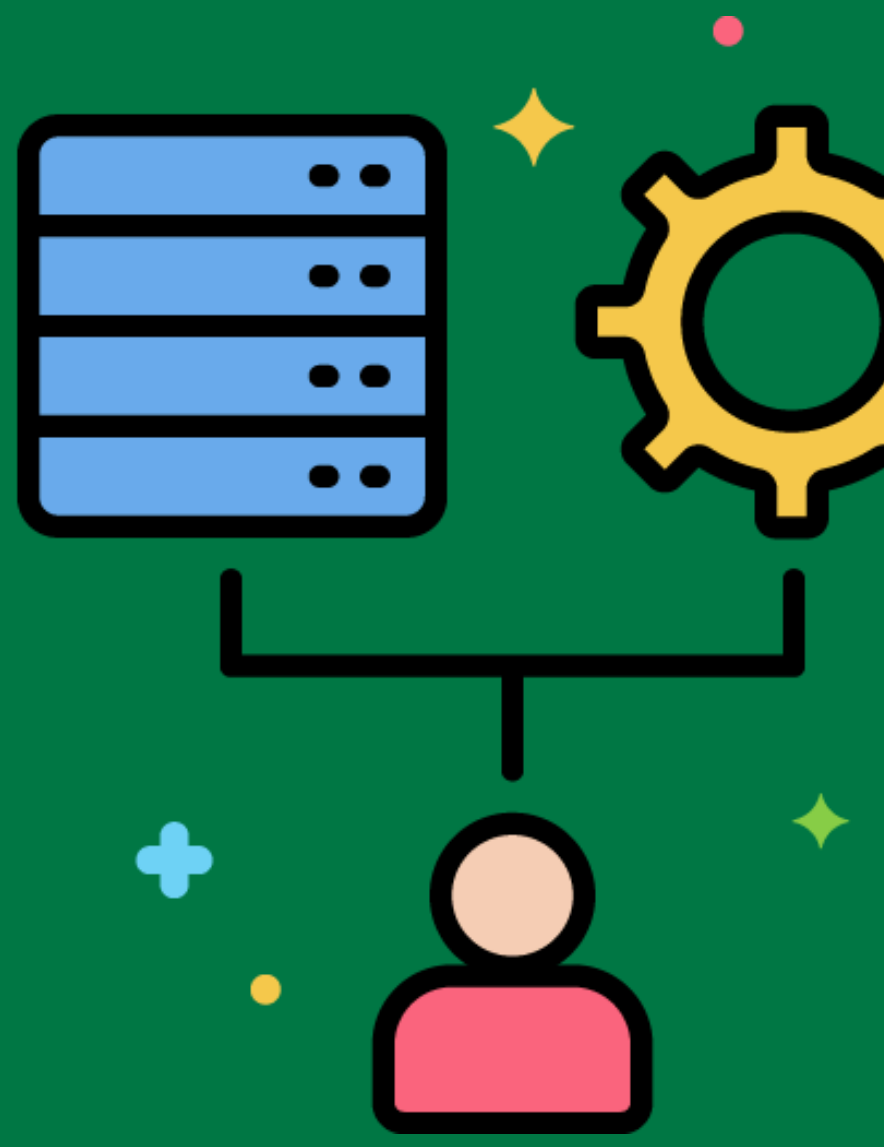
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Privacy



Ensuring that ETL processes comply with data privacy regulations and protect sensitive information.

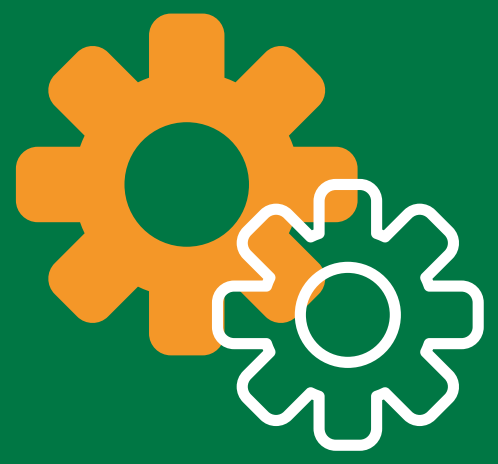


Implementing data anonymization and masking techniques to protect personally identifiable information (PII) during the ETL process.

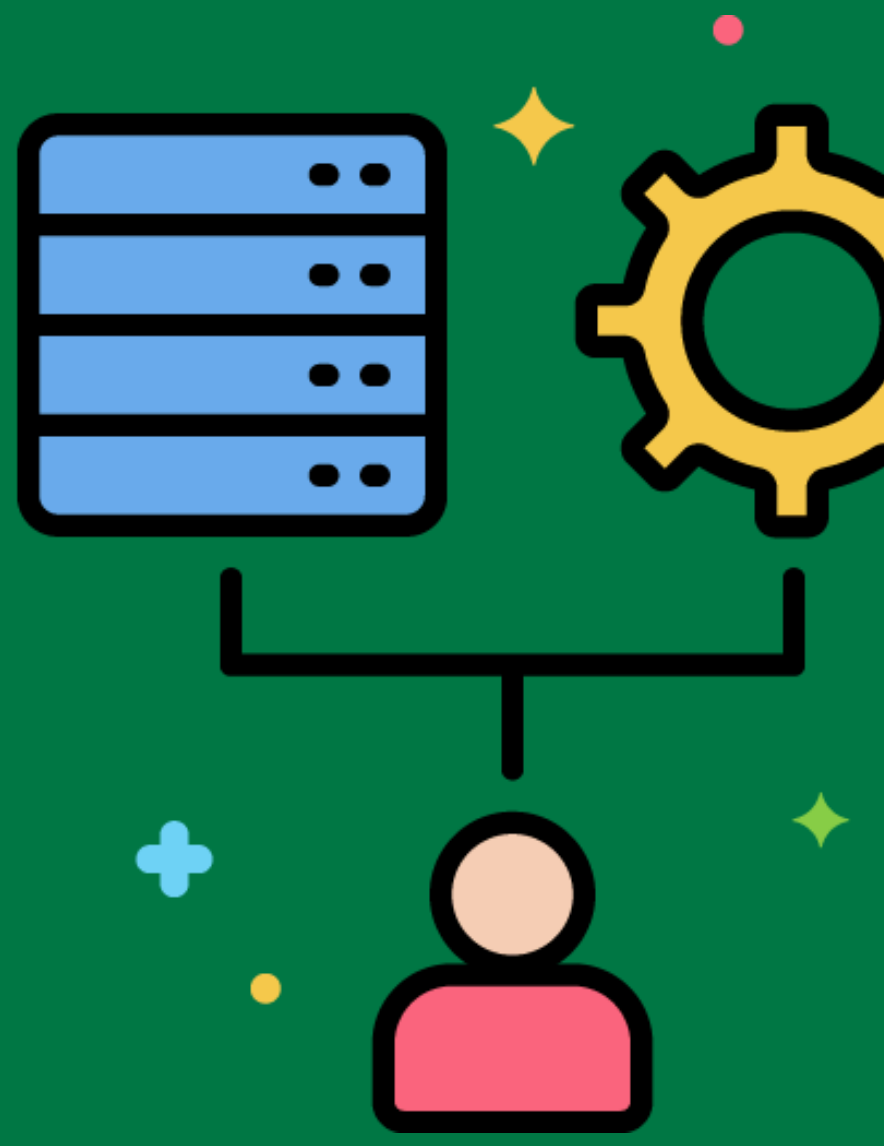
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Integrity



Ensuring that data is accurate, complete, and consistent throughout the ETL process.

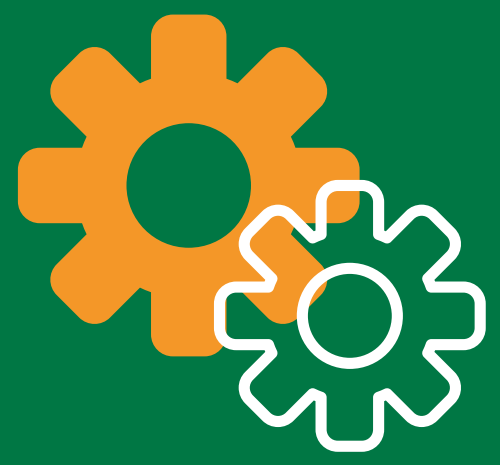


Implementing validation and reconciliation checks to ensure data integrity during extraction, transformation, and loading.

Shwetank Singh
GritSetGrow - GSGLearn.com

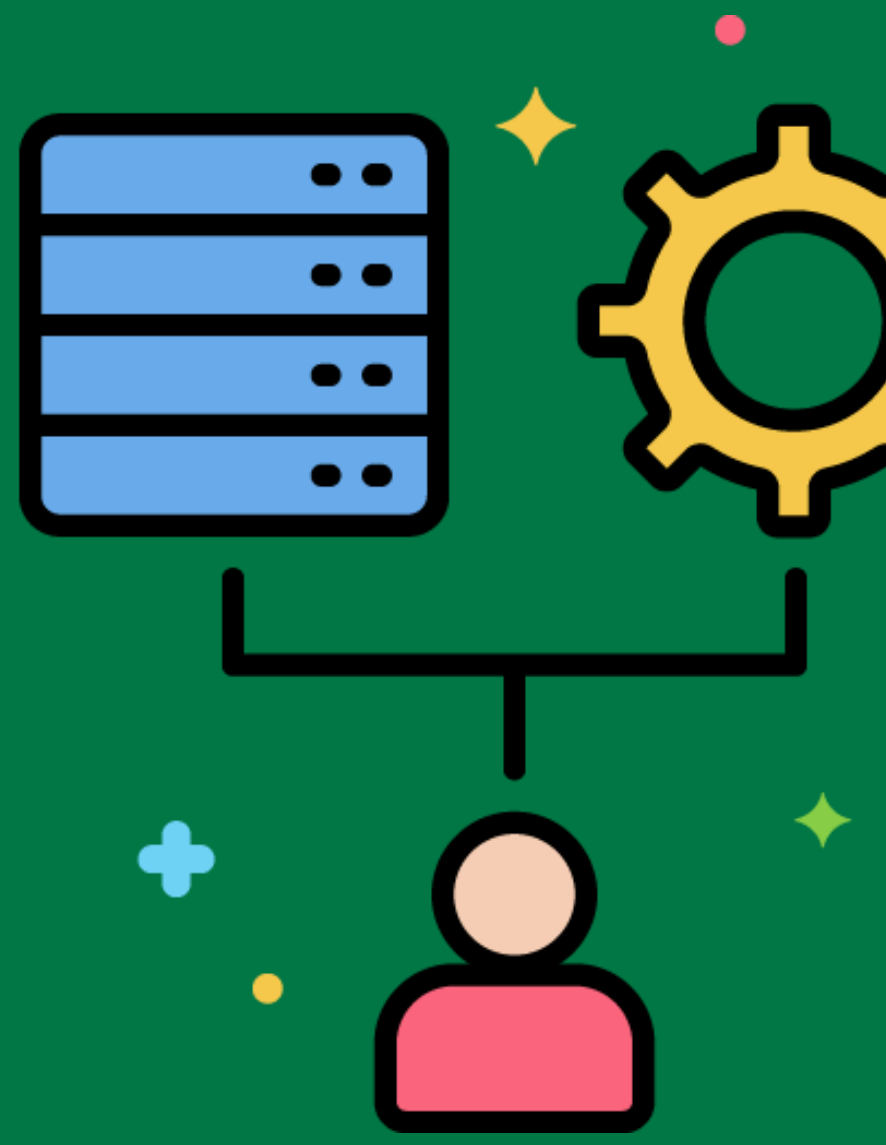
Data Engineering 101 - ETL Terminology

ETL Process Automation Best Practices



Recommended approaches and techniques for automating ETL processes to improve efficiency and reliability.

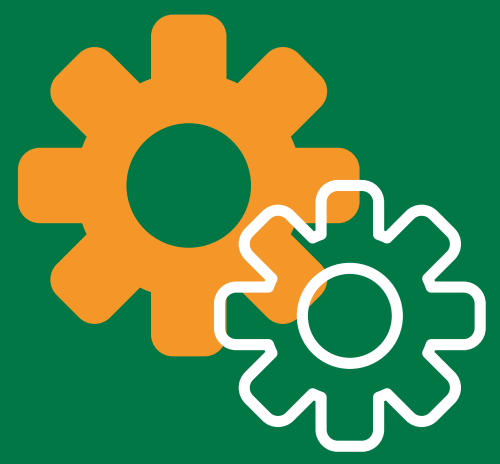
Following best practices for ETL automation, such as using orchestration tools and defining clear job dependencies.



Shwetank Singh
GritSetGrow - GSGLearn.com

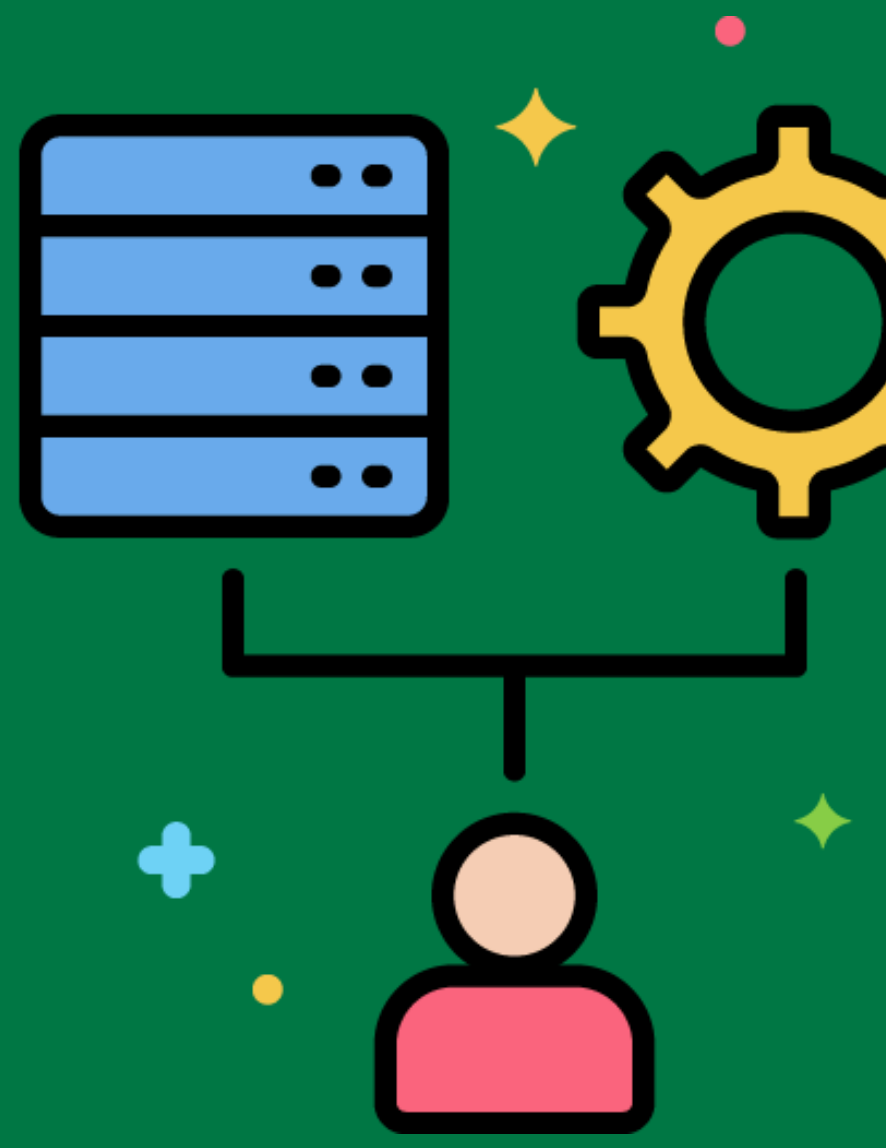
Data Engineering 101 - ETL Terminology

ETL Process Monitoring Best Practices



Recommended approaches and techniques for monitoring ETL processes to ensure they run smoothly and efficiently.

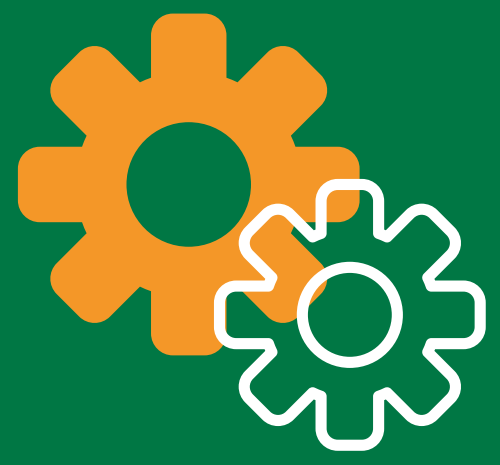
Implementing monitoring tools and dashboards to track ETL process performance and identify issues.



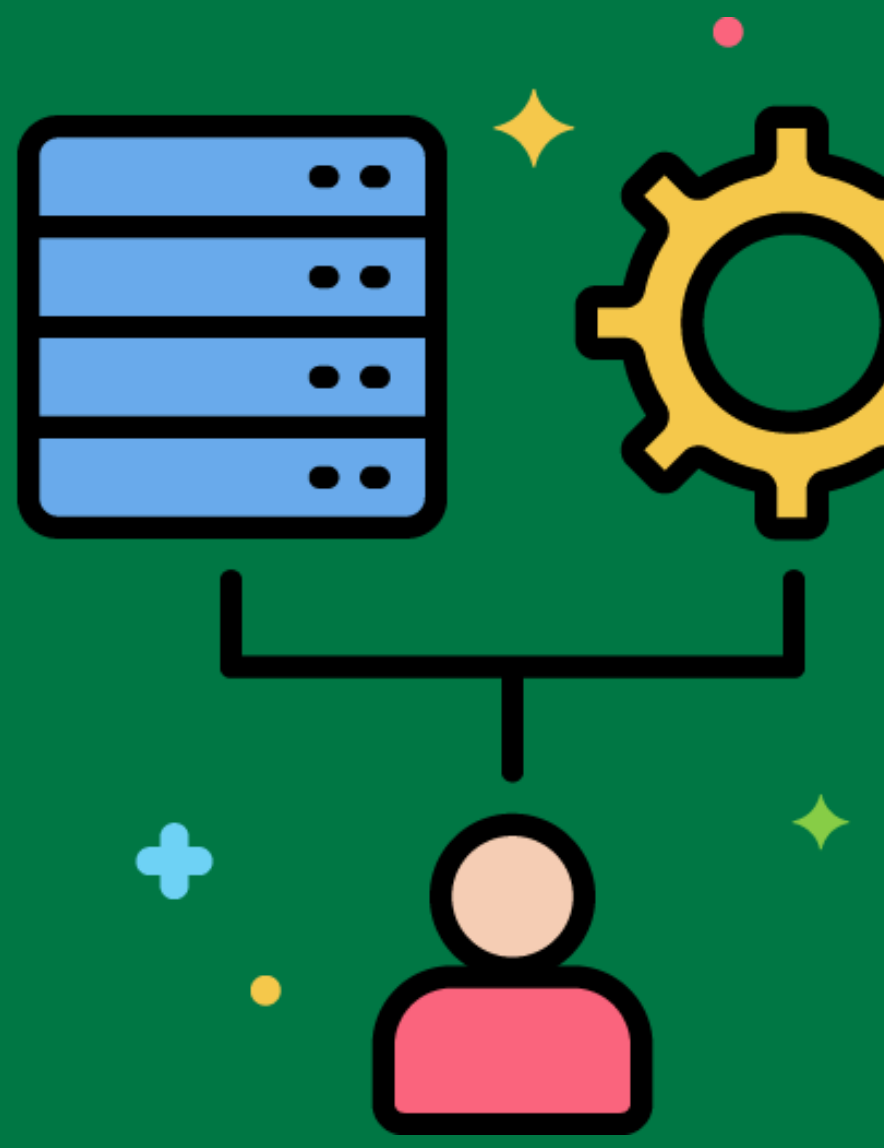
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Performance Tuning Best Practices



Recommended approaches and techniques for optimizing the performance of ETL processes.

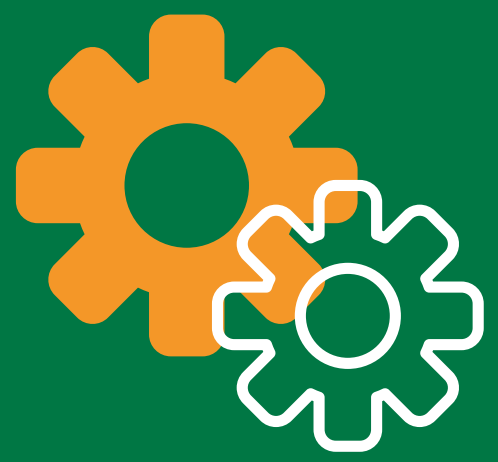


Following best practices for performance tuning, such as optimizing SQL queries, using parallel processing, and minimizing data movement.

Shwetank Singh
GritSetGrow - GSGLearn.com

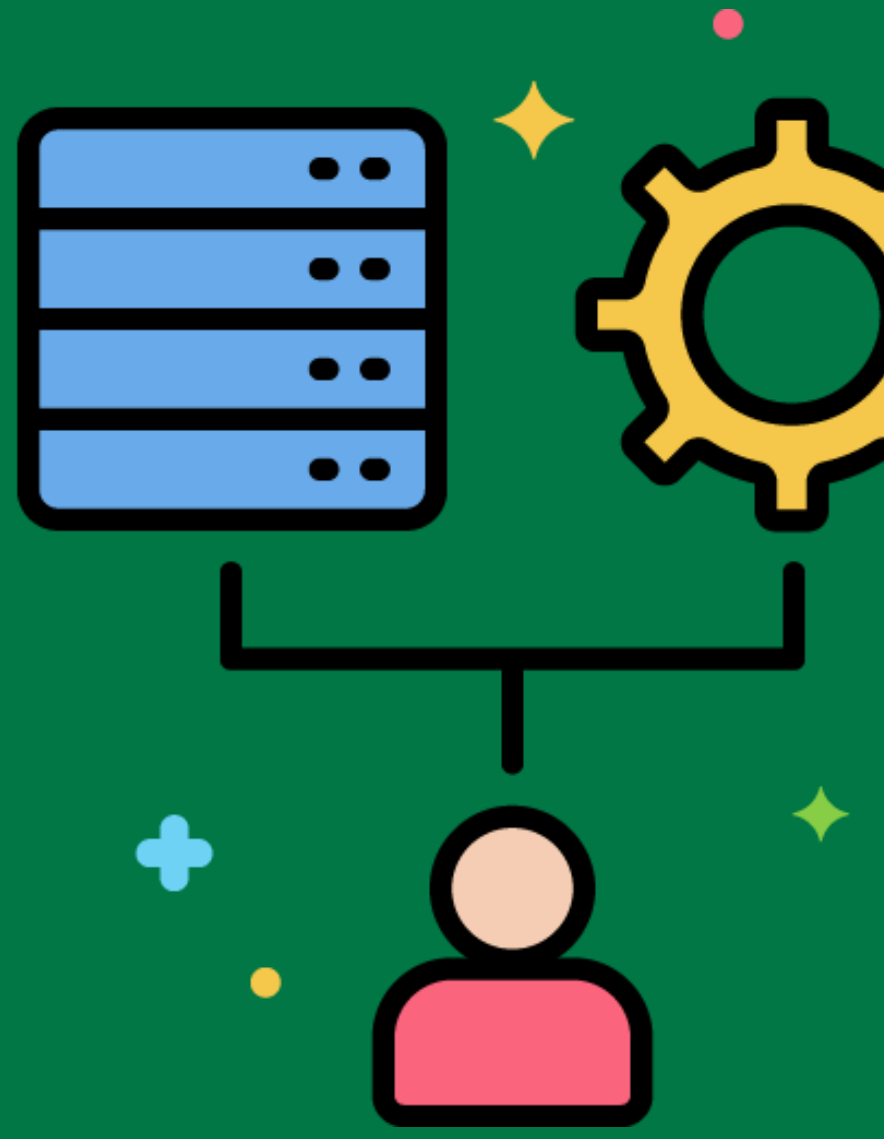
Data Engineering 101 - ETL Terminology

ETL Data Quality Management



Ensuring that data in the data warehouse meets quality standards and is suitable for analysis and reporting.

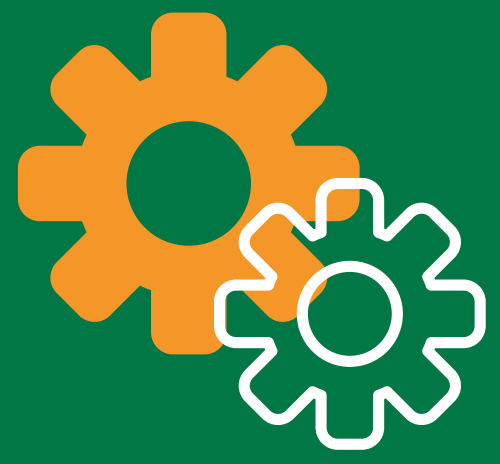
Implementing data quality management practices, such as validation, cleansing, and monitoring, to ensure data quality in the data warehouse.



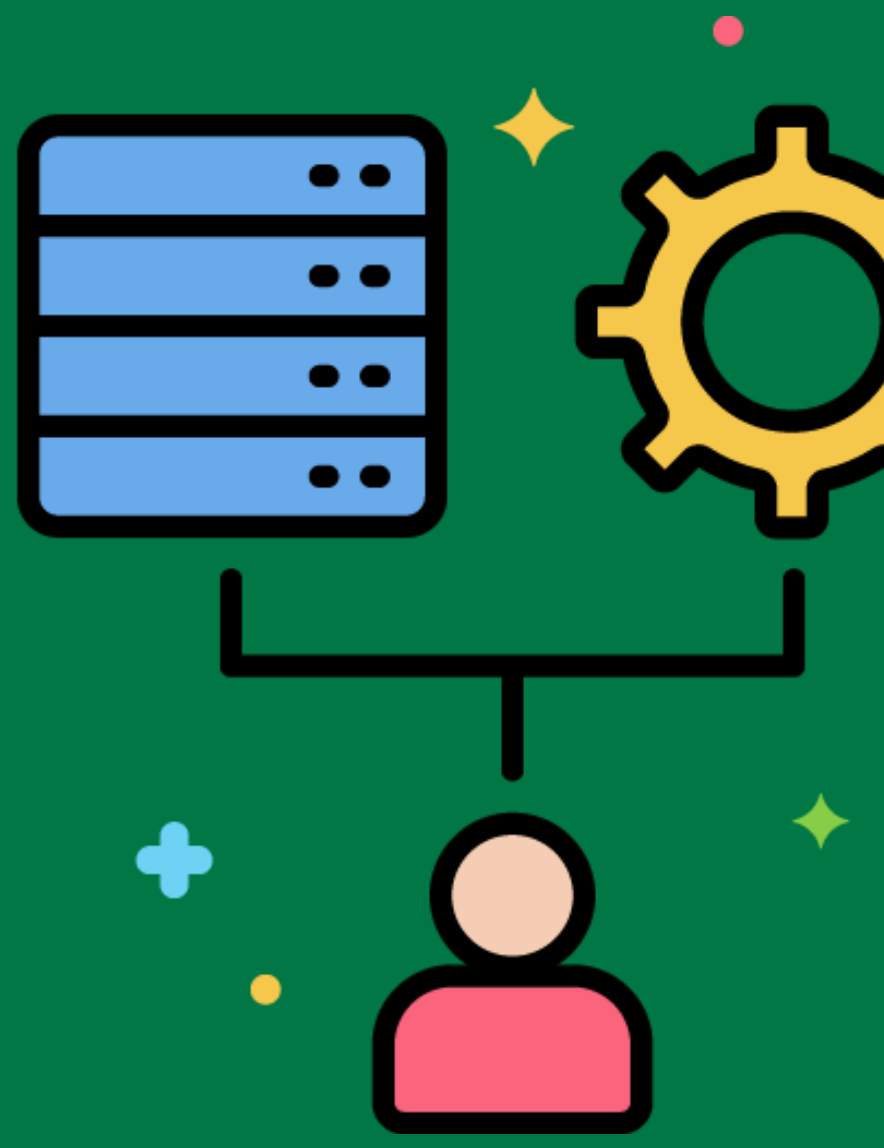
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Metadata Management Best Practices



Recommended approaches and techniques for managing metadata related to the ETL process.

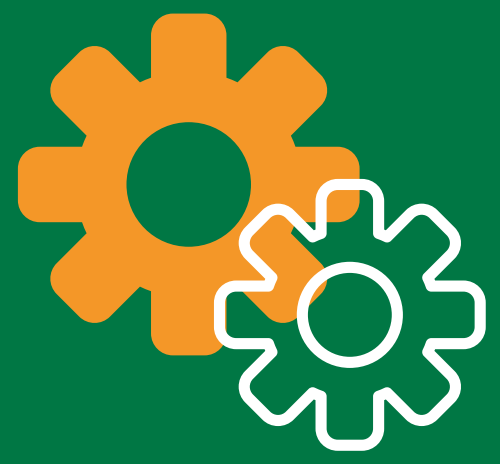


Following best practices for metadata management, such as documenting source-to-target mappings, transformation rules, and data lineage.

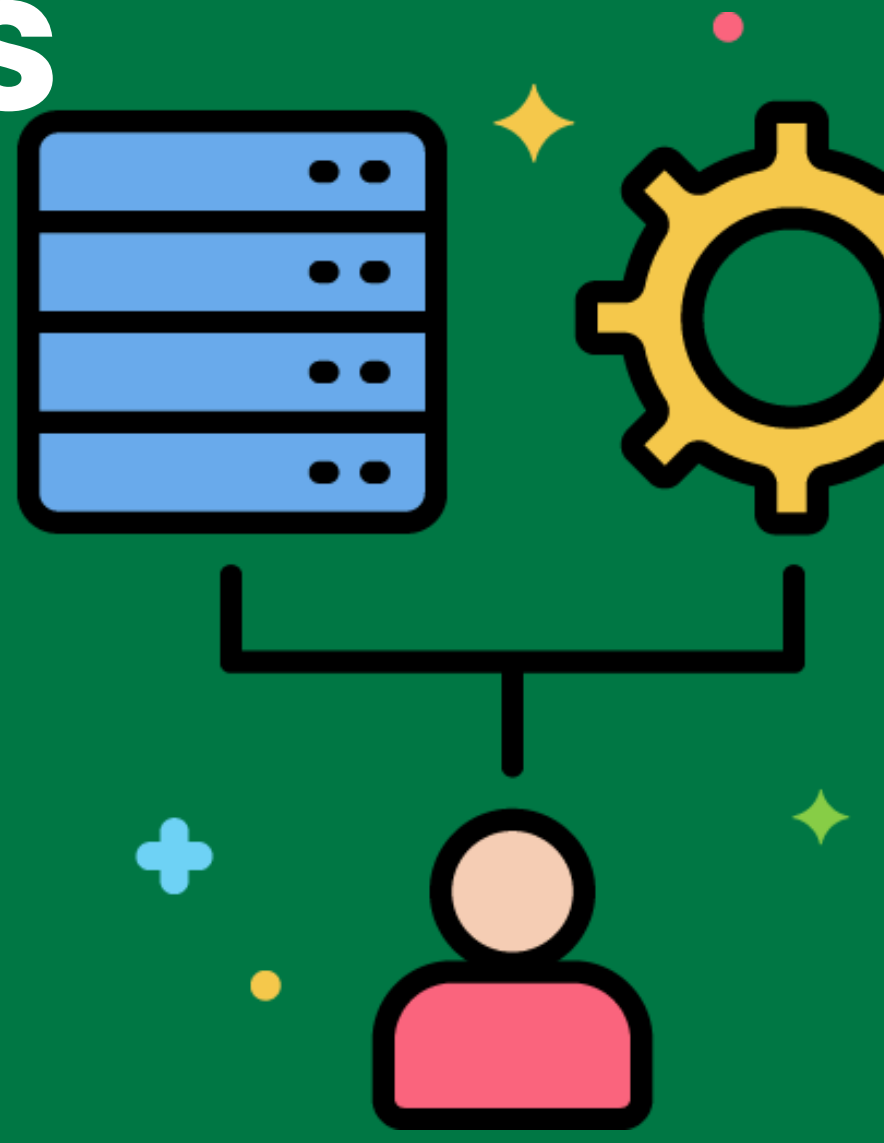
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Integration Tools and Technologies



Tools and technologies used to integrate data from multiple sources and provide a unified view.

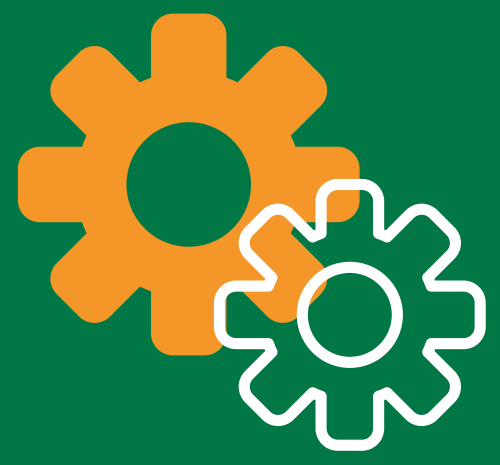


Using data integration tools like Talend, Informatica, or Apache Nifi to combine data from different sources into the data warehouse.

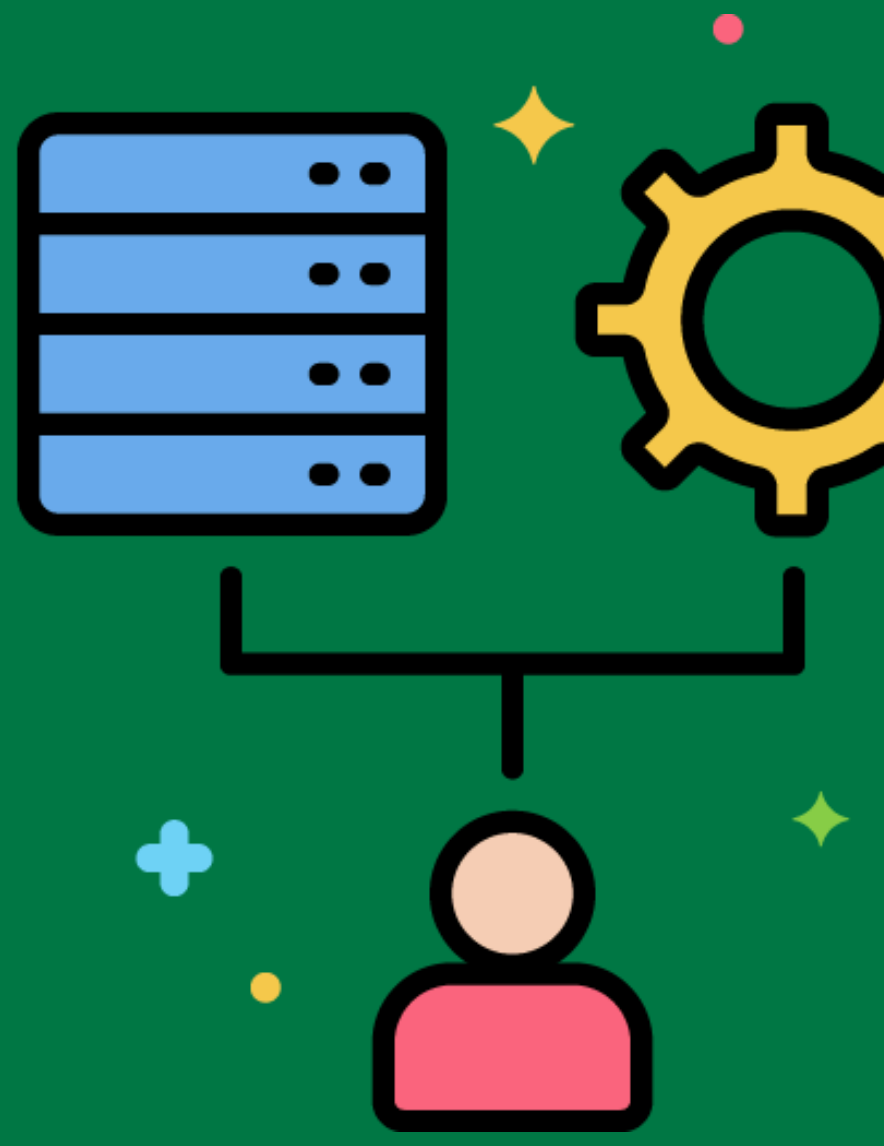
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Process Design Best Practices



Recommended approaches and techniques for designing ETL processes that are efficient, maintainable, and scalable.

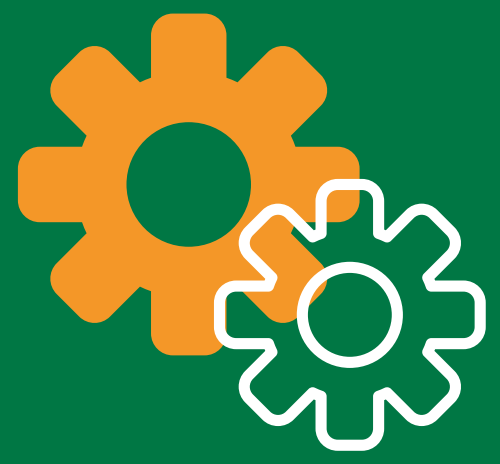


Following best practices for ETL process design, such as modularizing tasks, using descriptive naming conventions, and implementing robust error handling.

Shwetank Singh
GritSetGrow - GSGLearn.com

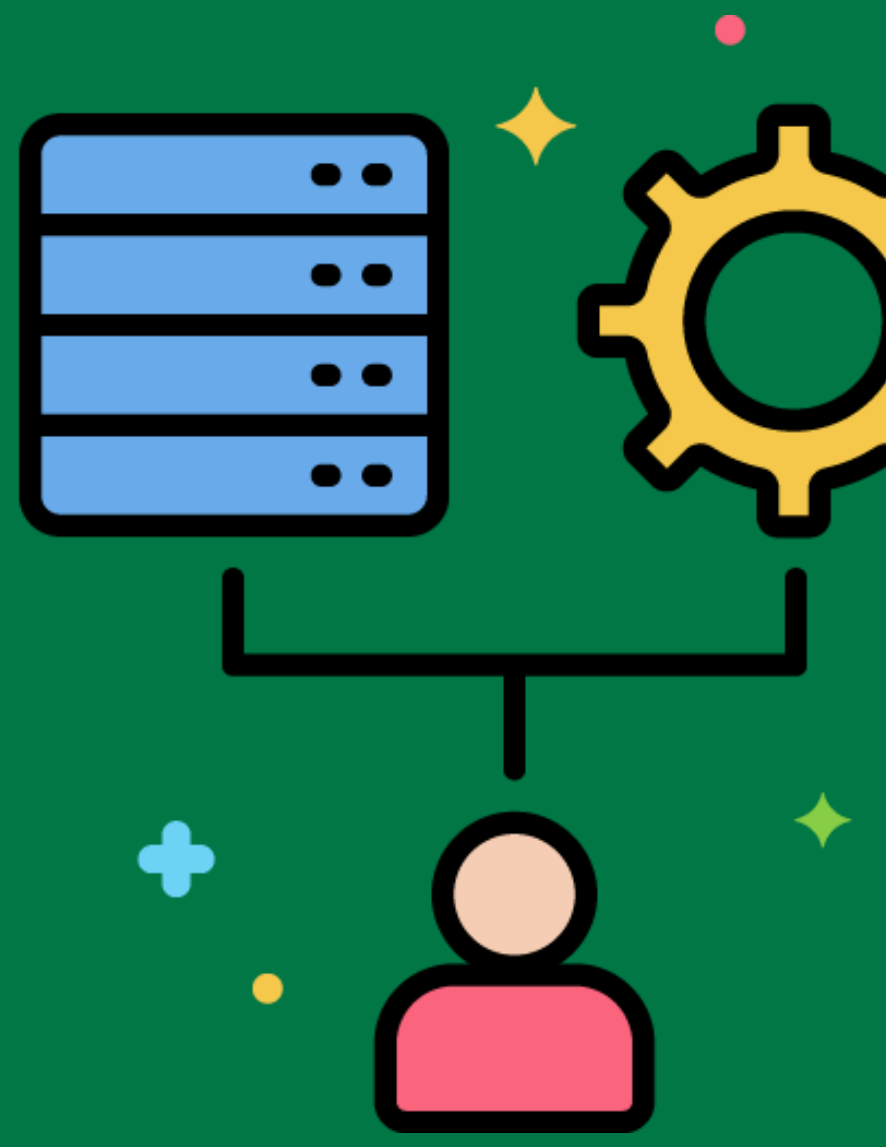
Data Engineering 101 - ETL Terminology

ETL Process Management



Overseeing the development, deployment, and maintenance of ETL processes to ensure they meet organizational objectives.

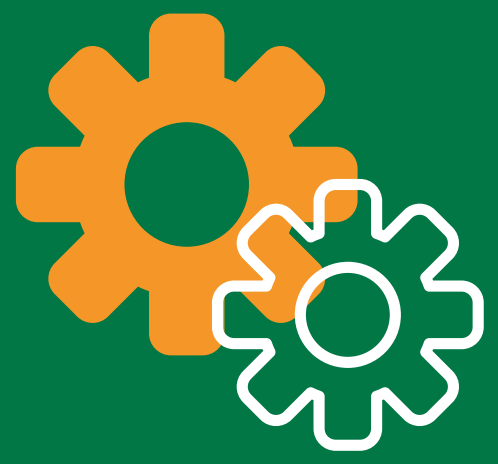
Implementing process management practices to coordinate ETL development, monitor performance, and ensure continuous improvement.



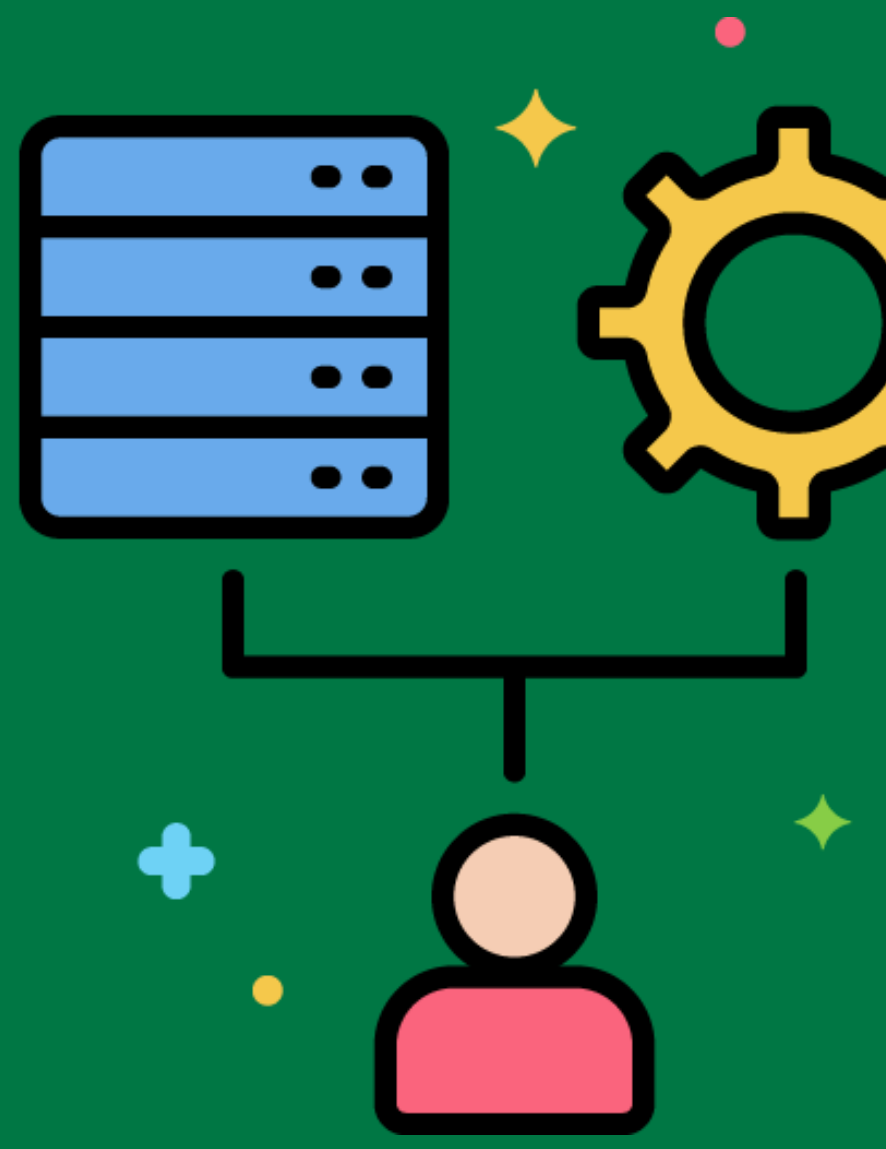
Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Development Tools



Software tools that facilitate the design, development, and testing of ETL processes.

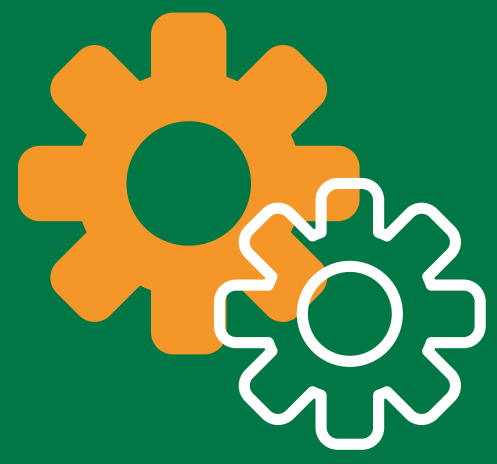


Using development tools like Talend Studio, Informatica PowerCenter, or Microsoft SSIS to create and manage ETL processes.

Shwetank Singh
GritSetGrow - GSGLearn.com

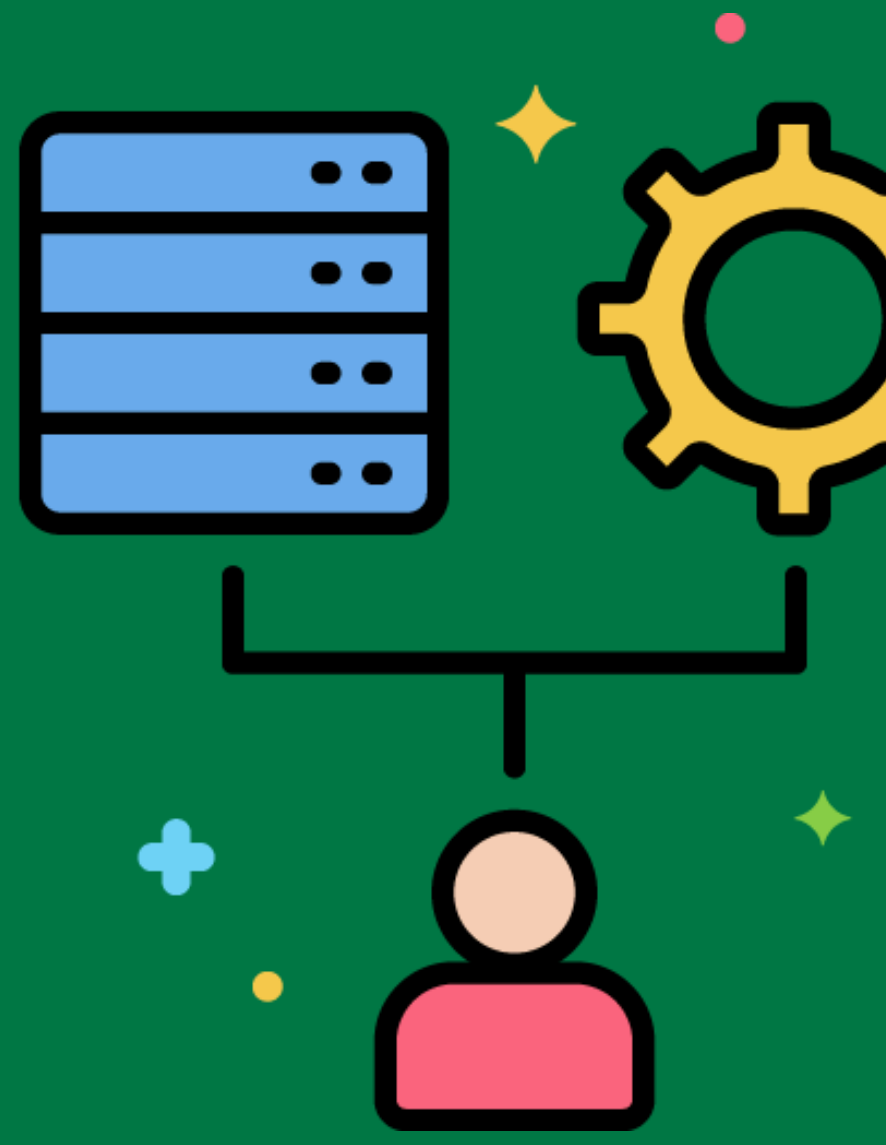
Data Engineering 101 - ETL Terminology

ETL Deployment Strategies



Approaches for moving ETL processes from development to production environments.

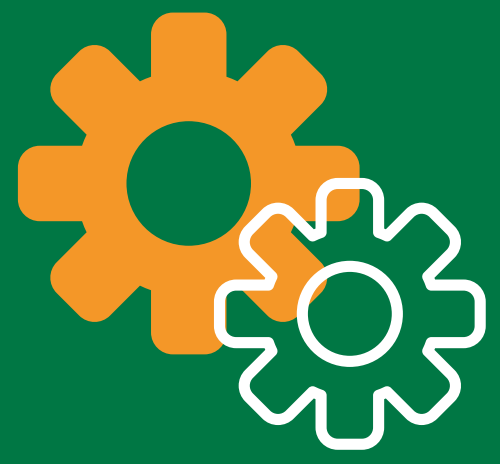
Using a phased deployment strategy to gradually roll out ETL processes and minimize disruptions.

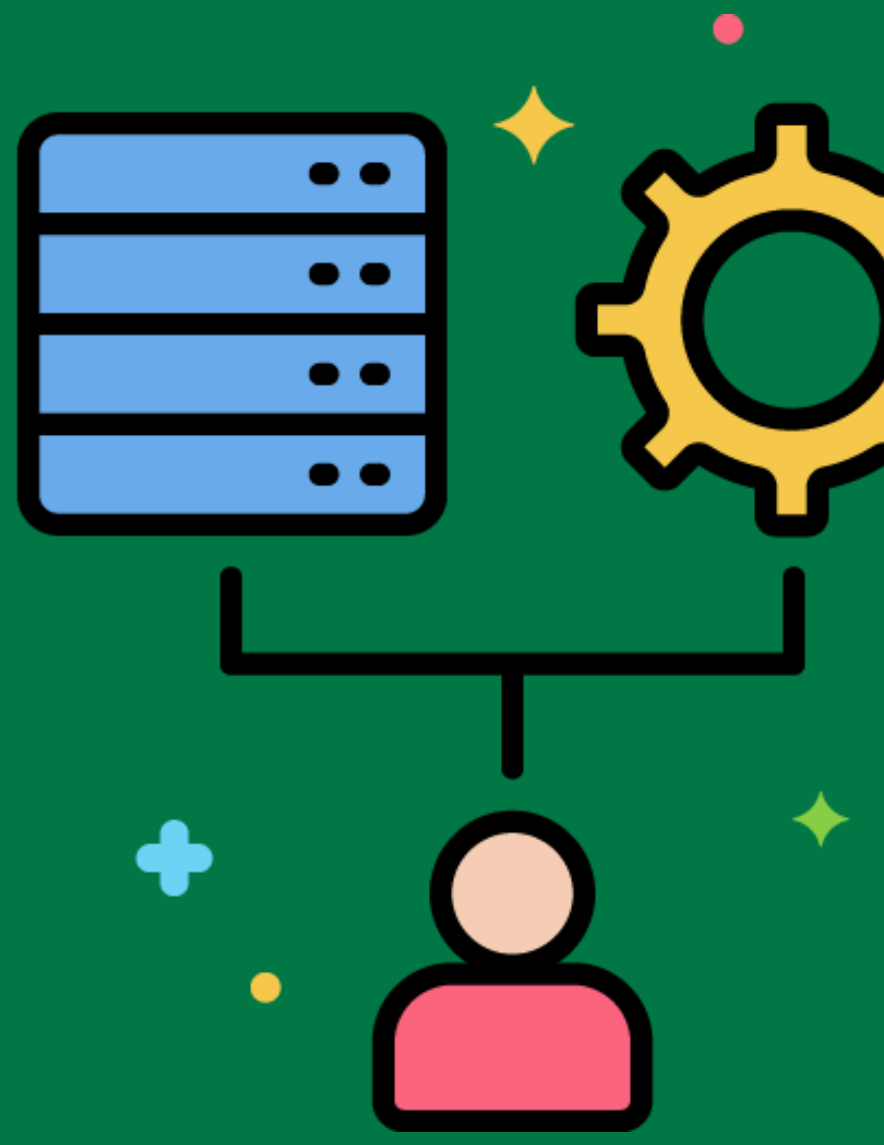


Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Maintenance Strategies





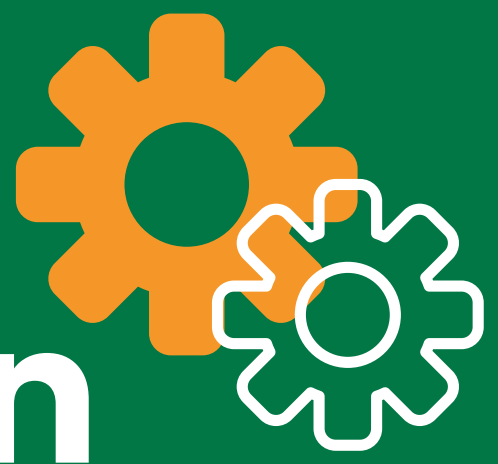
Approaches for maintaining and updating ETL processes to ensure they continue to meet business requirements.

Implementing maintenance strategies like regular process reviews, performance tuning, and error handling improvements.

Shwetank Singh
GritSetGrow - GSGLearn.com

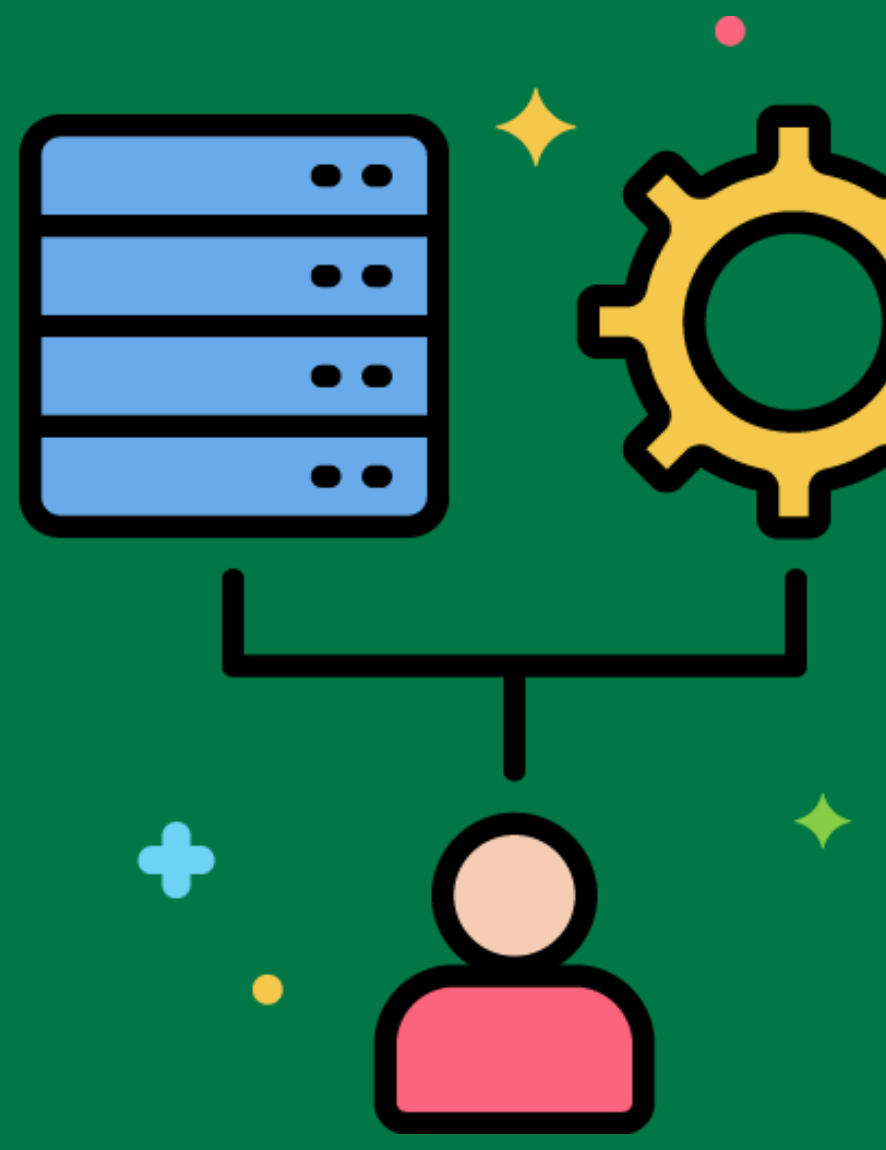
Data Engineering 101 - ETL Terminology

ETL Documentation Tools



Software tools that facilitate the creation and maintenance of documentation for ETL processes.

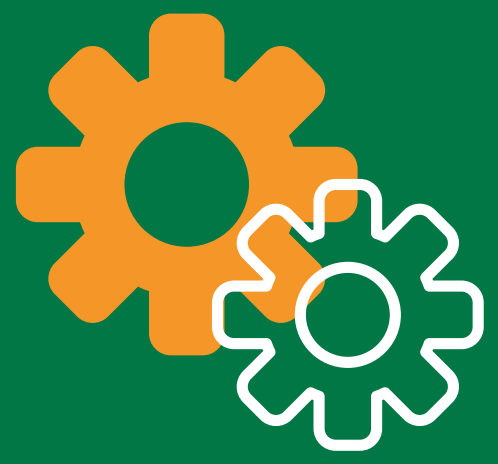
Using documentation tools like Confluence, JIRA, or Microsoft Word to document ETL processes, including data mappings and transformation rules.



Shwetank Singh
GritSetGrow - GSGLearn.com

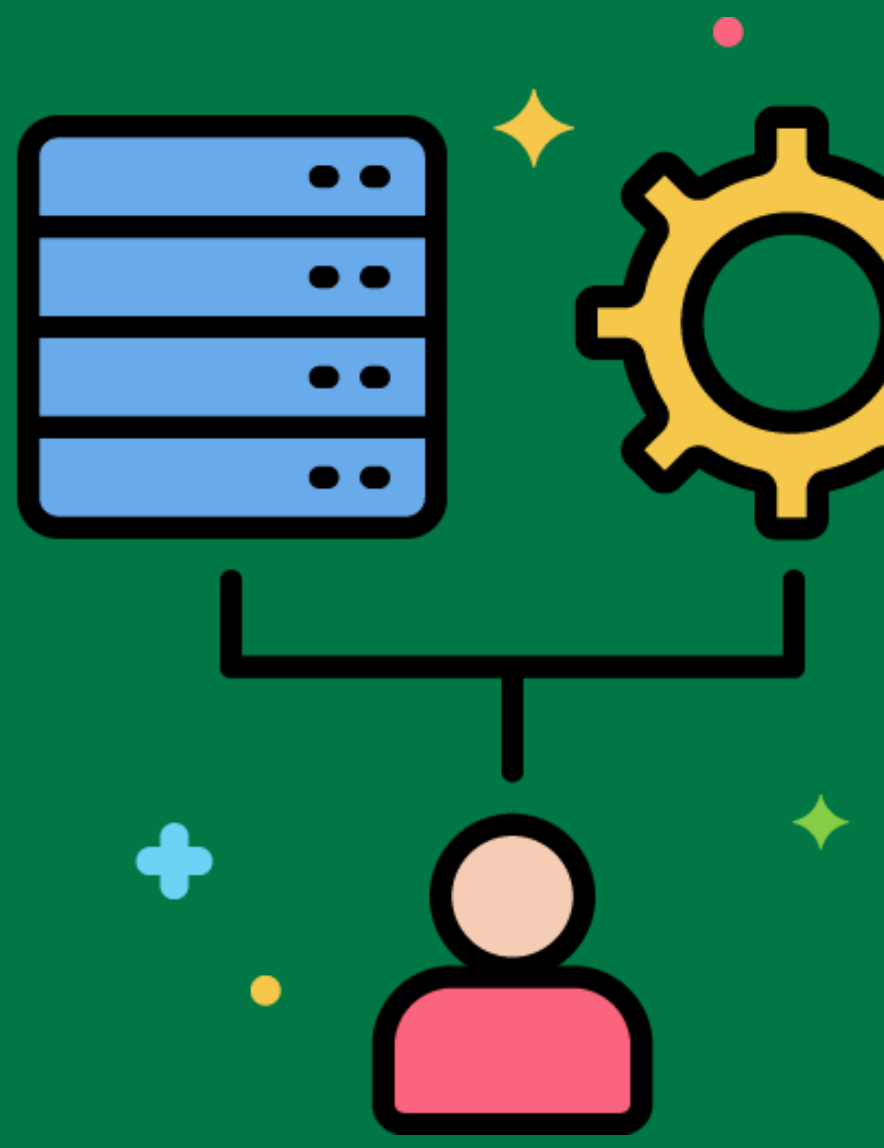
Data Engineering 101 - ETL Terminology

ETL Process Versioning



Managing different versions of ETL processes to track changes and ensure consistency.

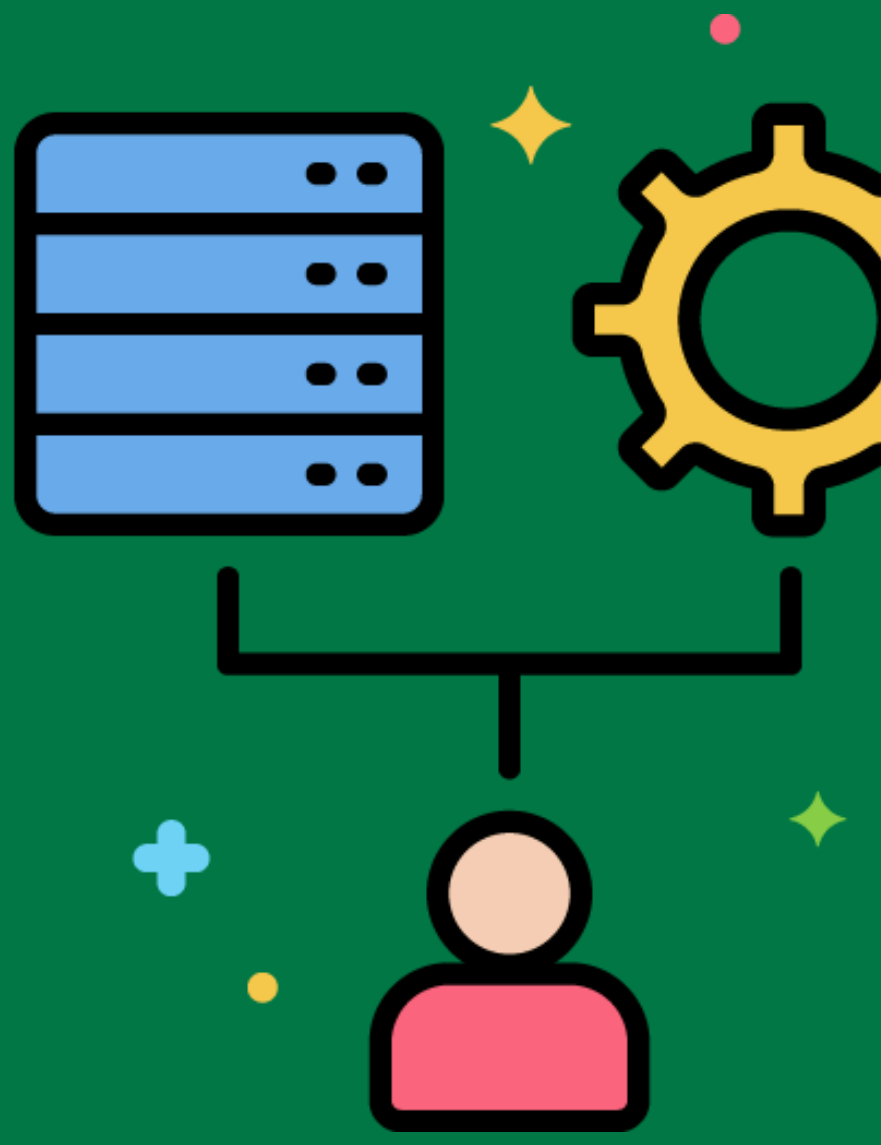
Using version control systems like Git to manage and track changes to ETL scripts and transformation logic.



Shwetank Singh
GritSetGrow - GSGLearn.com

Data Engineering 101 - ETL Terminology

ETL Data Transformation Tools



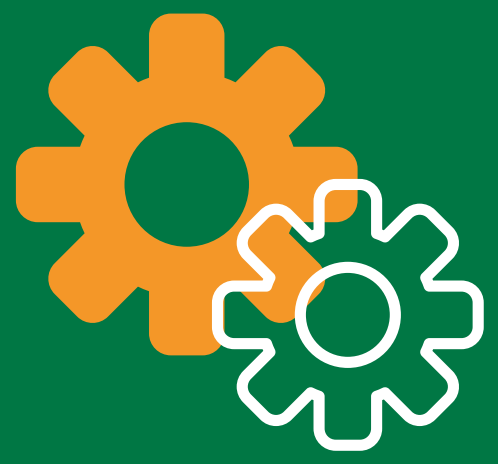
Software tools that facilitate the transformation of data during the ETL process.

Using data transformation tools like Talend, Informatica, or Pentaho to apply complex transformations to extracted data.

Shwetank Singh
GritSetGrow - GSGLearn.com

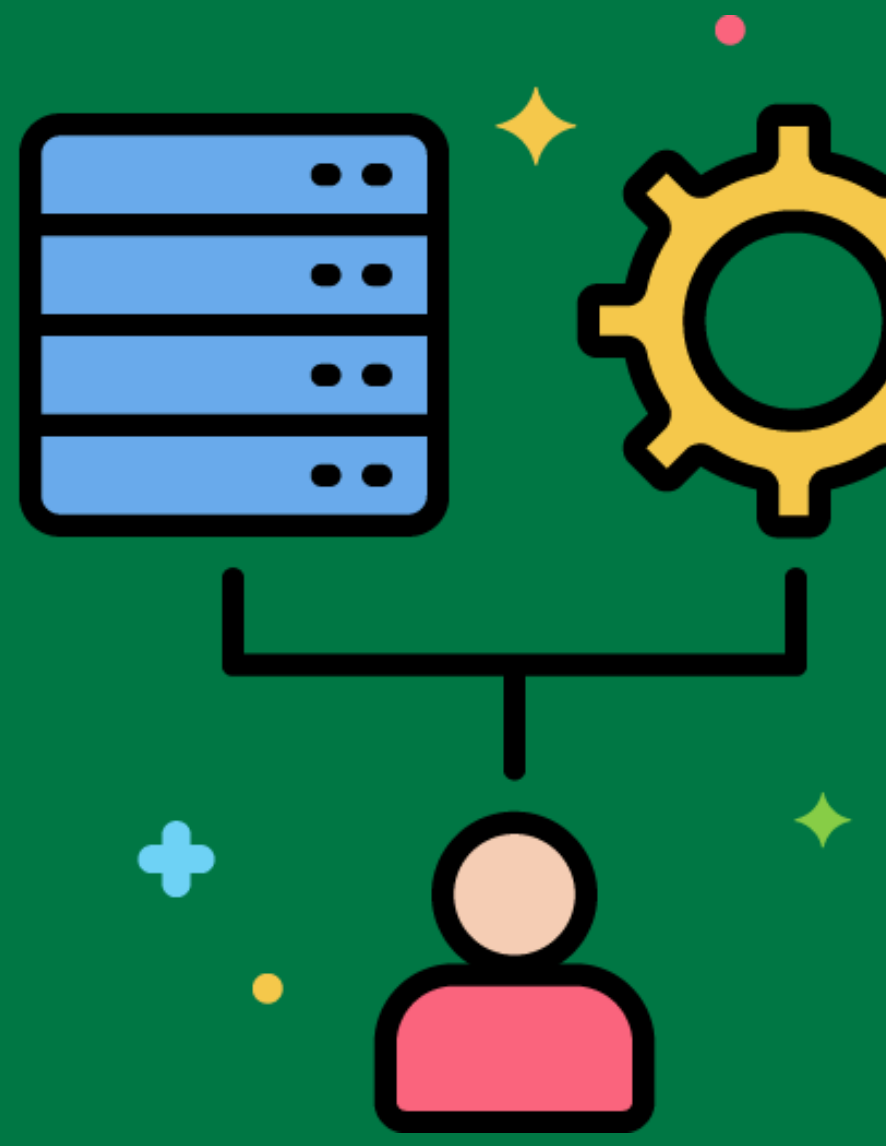
Data Engineering 101 - ETL Terminology

ETL Process Execution



The actual running of ETL jobs to extract, transform, and load data into the data warehouse.

Executing ETL jobs on a scheduled basis to keep the data warehouse up-to-date with the latest data from source systems.



Shwetank Singh
GritSetGrow - GSGLearn.com