DATA ENGINEERING 101: RELATIONAL DATABASES

RELATIONAL MODEL

EXPLANATION

A data model based on the concept of relations, which are sets of tuples sharing the same attributes.

EXAMPLE

A table representing a list of customers with columns for ID, Name, and Address.

Shwetank Singh
GritSetGrow - GSGLearn.com

DATABASE DESIGN

EXPLANATION

The process of creating a detailed data model of a database, including all necessary logical and physical design choices.

EXAMPLE

Designing tables and relationships for a customer order management system.

Shwetank Singh
GritSetGrow - GSGLearn.com

NORMALIZATION

EXPLANATION

The process of organizing data to minimize redundancy and improve data integrity.

EXAMPLE

Breaking down a large customer order table into separate tables for customers, orders, and order items.

Shwetank Singh
GritSetGrow - GSGLearn.com

FIRST NORMAL FORM (1NF)

EXPLANATION

A property of a relation where each column contains atomic, indivisible values, and each entry is unique.

EXAMPLE

A table where each cell contains only a single value, like a table of students with unique IDs.

Shwetank Singh
GritSetGrow - GSGLearn.com

FUNCTIONAL DEPENDENCY

EXPLANATION

A relationship where one attribute uniquely determines another attribute.

EXAMPLE

In a table of employees, EmployeeID -> EmployeeName (EmployeeID uniquely determines EmployeeName).

Shwetank Singh
GritSetGrow - GSGLearn.com

CANDIDATE KEY

EXPLANATION

A minimal set of attributes that can uniquely identify a tuple in a relation.

EXAMPLE

In a table of employees, EmployeeID and SocialSecurityNumber might both be candidate keys.

Shwetank Singh
GritSetGrow - GSGLearn.com

PRIMARY KEY

EXPLANATION

A candidate key chosen to uniquely identify tuples in a relation.

EXAMPLE

The "EmployeeID" column in a table of employees.

Shwetank Singh
GritSetGrow - GSGLearn.com

FOREIGN KEY

EXPLANATION

An attribute in a table that links to the primary key of another table, establishing a relationship between the two tables.

EXAMPLE

The "DepartmentID" column in an Employees table linking to the DepartmentID in a Departments table.

Shwetank Singh
GritSetGrow - GSGLearn.com

REFERENTIAL INTEGRITY

EXPLANATION

A property that ensures that foreign key values in a table match primary key values in the referenced table.

EXAMPLE

Ensuring every order in an Orders table references a valid customer in a Customers table.

Shwetank Singh
GritSetGrow - GSGLearn.com

SECOND NORMAL FORM (2NF)

EXPLANATION

A relation is in 2NF if it is in 1NF and all non-key attributes are fully functionally dependent on the entire primary key.

EXAMPLE

Splitting a table with composite keys into two tables where each table's non-key attributes depend on the full primary key.

Shwetank Singh
GritSetGrow - GSGLearn.com

THIRD NORMAL FORM (3NF)

EXPLANATION

A relation is in 3NF if it is in 2NF and all the attributes are functionally dependent only on the primary key.

EXAMPLE

A table where no non-primary key attribute depends on another non-primary key attribute.

Shwetank Singh
GritSetGrow - GSGLearn.com

BOYCE-CODD NORMAL FORM

EXPLANATION

A stricter version of 3NF where every determinant is a candidate key.

EXAMPLE

Ensuring no redundancy by having only one candidate key determining other attributes.

Shwetank Singh
GritSetGrow - GSGLearn.com

DENORMALIZATION

EXPLANATION

The process of deliberately introducing redundancy into a database to improve read performance.

EXAMPLE

Combining tables to reduce the number of joins in queries for better performance.

Shwetank Singh
GritSetGrow - GSGLearn.com

JOIN DEPENDENCY

EXPLANATION

A constraint that specifies a relation should be decomposable into two or more relations without loss of information.

EXAMPLE

A table of orders can be decomposed into separate tables for order headers and order details.

Shwetank Singh
GritSetGrow - GSGLearn.com

MULTIVALUED DEPENDENCY

EXPLANATION

A type of dependency where one attribute in a table uniquely determines another attribute, independent of other attributes.

EXAMPLE

In a table of books, Author and Title are multivalued dependencies if each book can have multiple authors.

Shwetank Singh
GritSetGrow - GSGLearn.com

FOURTH NORMAL FORM (4NF)

EXPLANATION

A relation is in 4NF if it has no multivalued dependencies other than a candidate key.

EXAMPLE

Splitting a table with multivalued dependencies into separate tables to ensure that each dependency is represented by a separate table.

Shwetank Singh
GritSetGrow - GSGLearn.com

FIFTH NORMAL FORM (5NF)

EXPLANATION

A relation is in 5NF if it cannot be decomposed into any number of smaller tables without loss of data.

EXAMPLE

Ensuring complex join dependencies are handled correctly by decomposing tables into the smallest possible tables.

Shwetank Singh
GritSetGrow - GSGLearn.com

DOMAIN-KEY NORMAL FORM

EXPLANATION

A normal form that ensures all constraints on the database are logical consequences of the domain and key constraints.

EXAMPLE

Defining strict domain constraints and primary keys for each table to ensure data integrity.

Shwetank Singh
GritSetGrow - GSGLearn.com

ENTITY-RELATIONSHIP MODEL

EXPLANATION

A data modeling technique that creates a graphical representation of the entities and their relationships in a database.

EXAMPLE

An ER diagram showing tables for Customers, Orders, and Products with lines connecting related entities.

Shwetank Singh
GritSetGrow - GSGLearn.com

TUPLE

EXPLANATION

A single entry in a table, representing a set of related data items.

EXAMPLE

A row in a table representing a single employee with attributes like EmployeeID, Name, and Department.

Shwetank Singh
GritSetGrow - GSGLearn.com

ATTRIBUTE

EXPLANATION

A column in a table, representing a data field.

EXAMPLE

The "Name" column in a table of employees.

Shwetank Singh
GritSetGrow - GSGLearn.com

DATA INTEGRITY

EXPLANATION

The accuracy and consistency of data stored in a database.

EXAMPLE

Implementing constraints and validations to ensure accurate and consistent data entry.

Shwetank Singh
GritSetGrow - GSGLearn.com

RELVAR

EXPLANATION

A variable that represents a relation in a relational database.

EXAMPLE

A table in a database where data can be inserted, updated, or deleted.

Shwetank Singh
GritSetGrow - GSGLearn.com

SCHEMA

EXPLANATION

The structure of a database, defined by its tables, columns, data types, and relationships.

EXAMPLE

A database schema for an e-commerce system including tables for Users, Products, Orders, and OrderItems.

Shwetank Singh
GritSetGrow - GSGLearn.com

VIEW

EXPLANATION

A virtual table in a database created by a query that joins and filters data from one or more tables.

EXAMPLE

A view that shows only active customers from the Customers table.

Shwetank Singh
GritSetGrow - GSGLearn.com

INDEX

EXPLANATION

A database object that improves the speed of data retrieval operations on a table.

EXAMPLE

Creating an index on the CustomerID column in an Orders table to speed up queries on customer orders.

Shwetank Singh
GritSetGrow - GSGLearn.com

COMPOSITE KEY

EXPLANATION

A key that consists of two or more attributes that uniquely identify a tuple in a relation.

EXAMPLE

In a table of student enrollments, a composite key could be {StudentID, CourseID}.

Shwetank Singh
GritSetGrow - GSGLearn.com

SURROGATE KEY

EXPLANATION

An artificial key created for uniquely identifying a tuple, usually a sequential number.

EXAMPLE

An auto-incremented ID column used as the primary key in a table.

Shwetank Singh
GritSetGrow - GSGLearn.com

NATURAL KEY

EXPLANATION

A key that is formed of attributes that already exist in the real world and have business meaning.

EXAMPLE

Using a social security number (SSN) as a primary key in an Employees table.

Shwetank Singh
GritSetGrow - GSGLearn.com

INTEGRITY CONSTRAINT

EXPLANATION

A rule that ensures the correctness and validity of data in a database.

EXAMPLE

A constraint that enforces unique email addresses in a Users table.

Shwetank Singh
GritSetGrow - GSGLearn.com

TRIGGER

EXPLANATION

A procedural code that is automatically executed in response to certain events on a particular table or view.

EXAMPLE

A trigger that updates the inventory count after an order is placed.

Shwetank Singh
GritSetGrow - GSGLearn.com

STORED PROCEDURE

EXPLANATION

A precompiled collection of SQL statements that can be executed as a single unit.

EXAMPLE

A stored procedure that calculates monthly sales totals for a store.

Shwetank Singh
GritSetGrow - GSGLearn.com

ACID PROPERTIES

EXPLANATION

A set of properties that guarantee that database transactions are processed reliably (Atomicity, Consistency, Isolation, Durability).

EXAMPLE

Ensuring a banking transaction either fully completes or fully rolls back to maintain consistent state.

Shwetank Singh
GritSetGrow - GSGLearn.com

ATOMICITY

EXPLANATION

Ensures that each transaction is treated as a single unit, which either succeeds completely or fails completely.

EXAMPLE

Transferring funds between accounts where both debit and credit operations must succeed or both fail.

Shwetank Singh
GritSetGrow - GSGLearn.com

CONSISTENCY

EXPLANATION

Ensures that a database remains in a consistent state before and after a transaction.

EXAMPLE

Ensuring that all foreign key references are valid after a transaction completes.

Shwetank Singh
GritSetGrow - GSGLearn.com

ISOLATION

EXPLANATION

Ensures that transactions are executed in isolation from each other, preventing concurrent transactions from interfering with each other.

EXAMPLE

Using isolation levels to prevent dirty reads in concurrent transactions.

Shwetank Singh
GritSetGrow - GSGLearn.com

DURABILITY

EXPLANATION

Ensures that the results of a transaction are permanently stored in the database and will not be lost, even in the event of a system failure.

EXAMPLE

Committing a transaction to ensure data is saved even if the database crashes afterward.

Shwetank Singh
GritSetGrow - GSGLearn.com

ENTITY

EXPLANATION

An object that exists and is distinguishable from other objects, often representing real-world objects in a database.

EXAMPLE

A customer, product, or order in a database.

Shwetank Singh
GritSetGrow - GSGLearn.com

RELATIONSHIP

EXPLANATION

An association between two or more entities in a database.

EXAMPLE

The relationship between Customers and Orders in a sales database.

Shwetank Singh
GritSetGrow - GSGLearn.com

CARDINALITY

EXPLANATION

The uniqueness of data values contained in a particular column (attribute) of a database table.

EXAMPLE

A one-to-many relationship between customers and orders.

Shwetank Singh
GritSetGrow - GSGLearn.com

DOMAIN

EXPLANATION

The set of permissible values for a given attribute.

EXAMPLE

Defining a domain for the "age" attribute to be between 0 and 120.

Shwetank Singh
GritSetGrow - GSGLearn.com

CONSTRAINT

EXPLANATION

A rule applied to data in a database to enforce data integrity.

EXAMPLE

A UNIQUE constraint on an email column to ensure no duplicate emails are entered.

Shwetank Singh
GritSetGrow - GSGLearn.com

COMPOSITE ATTRIBUTE

EXPLANATION

An attribute that can be subdivided into other attributes.

EXAMPLE

An address attribute that can be subdivided into street, city, state, and zip code.

Shwetank Singh
GritSetGrow - GSGLearn.com

DERIVED ATTRIBUTE

EXPLANATION

An attribute whose value is calculated from other attributes.

EXAMPLE

An "age" attribute derived from the "date of birth" attribute.

Shwetank Singh
GritSetGrow - GSGLearn.com

OPTIONAL ATTRIBUTE

EXPLANATION

An attribute that may not have a value in every instance of an entity.

EXAMPLE

A "middle name" attribute that is optional for a person's record.

Shwetank Singh
GritSetGrow - GSGLearn.com

MANDATORY ATTRIBUTE

EXPLANATION

An attribute that must have a value in every instance of an entity.

EXAMPLE

A "last name" attribute that is mandatory for a person's record.

Shwetank Singh
GritSetGrow - GSGLearn.com

DOMAIN INTEGRITY

EXPLANATION

Ensures that all values in a column fall within a defined domain.

EXAMPLE

Ensuring that all entries in an "age" column are non-negative integers.

Shwetank Singh
GritSetGrow - GSGLearn.com

ENTITY INTEGRITY

EXPLANATION

Ensures that each entity has a unique identifier and that this identifier is not null.

EXAMPLE

Using a primary key to ensure each row in a table is uniquely identifiable.

Shwetank Singh
GritSetGrow - GSGLearn.com

REFERENTIAL INTEGRITY

EXPLANATION

Ensures that foreign key values must match primary key values in the related table or be null.

EXAMPLE

Preventing deletion of a customer record that is referenced by an order record.

Shwetank Singh
GritSetGrow - GSGLearn.com

NULLABILITY

EXPLANATION

Determines whether a column can accept null values.

EXAMPLE

Allowing a "middle name" column to have null values if the person does not have a middle name.

Shwetank Singh
GritSetGrow - GSGLearn.com

CHECK CONSTRAINT

EXPLANATION

A rule that specifies a condition that must be met by the data in a table.

EXAMPLE

A check constraint ensuring that the salary of an employee is within a specified range.

Shwetank Singh
GritSetGrow - GSGLearn.com

UNIQUE CONSTRAINT

EXPLANATION

Ensures that all values in a column or a set of columns are unique across the rows in a table.

EXAMPLE

Ensuring that all entries in an email column are unique within a Users table.

Shwetank Singh
GritSetGrow - GSGLearn.com

PRIMARY KEY CONSTRAINT

EXPLANATION

Ensures that a column or a set of columns uniquely identifies each row in a table.

EXAMPLE

Defining a primary key constraint on the CustomerID column in a Customers table.

Shwetank Singh
GritSetGrow - GSGLearn.com

FOREIGN KEY CONSTRAINT

EXPLANATION

Ensures that a value in one table matches a value in another table, establishing a link between the two tables.

EXAMPLE

Enforcing that each Order's CustomerID matches a valid CustomerID in the Customers table.

Shwetank Singh
GritSetGrow - GSGLearn.com

INDEX

EXPLANATION

A database object that improves the speed of data retrieval operations on a table.

EXAMPLE

Creating an index on the LastName column in an Employees table.

Shwetank Singh
GritSetGrow - GSGLearn.com

TRANSACTION

EXPLANATION

A sequence of operations performed as a single logical unit of work.

EXAMPLE

A banking transaction that includes debit from one account and credit to another.

Shwetank Singh
GritSetGrow - GSGLearn.com

ROLLBACK

EXPLANATION

The process of undoing a transaction, returning the database to its previous state.

EXAMPLE

Reverting a transaction that failed to complete due to a system error.

Shwetank Singh
GritSetGrow - GSGLearn.com

COMMIT

EXPLANATION

The process of making a transaction's changes permanent in the database.

EXAMPLE

Committing a purchase transaction to finalize the order.

Shwetank Singh
GritSetGrow - GSGLearn.com

SAVEPOINT

EXPLANATION

A point within a transaction to which a transaction can be rolled back.

EXAMPLE

Setting a savepoint before making multiple updates in a transaction.

Shwetank Singh
GritSetGrow - GSGLearn.com

DEADLOCK

EXPLANATION

A situation where two or more transactions are waiting for each other to release locks, preventing all transactions from proceeding.

EXAMPLE

Two transactions each holding a lock on a resource the other needs, causing a standstill.

Shwetank Singh
GritSetGrow - GSGLearn.com

LOCK

EXPLANATION

A mechanism to control concurrent access to a database resource.

EXAMPLE

Locking a row in a table to prevent other transactions from modifying it concurrently.

Shwetank Singh
GritSetGrow - GSGLearn.com

ISOLATION LEVEL

EXPLANATION

A setting that controls the visibility of changes made by concurrent transactions.

EXAMPLE

Setting the isolation level to Serializable to prevent dirty reads.

Shwetank Singh
GritSetGrow - GSGLearn.com

DIRTY READ

EXPLANATION

A situation where a transaction reads data that has not yet been committed.

EXAMPLE

Reading an uncommitted change from another transaction, which may later be rolled back.

Shwetank Singh
GritSetGrow - GSGLearn.com

PHANTOM READ

EXPLANATION

A situation where a transaction re-reads data and finds rows that were not visible before.

EXAMPLE

A transaction that finds new rows inserted by another transaction after its initial read.

Shwetank Singh
GritSetGrow - GSGLearn.com

NON-REPEATABLE READ

EXPLANATION

A situation where a transaction reads the same row twice and finds different data each time.

EXAMPLE

Reading a row's data, another transaction updates it, then reading it again and finding the new data.

Shwetank Singh
GritSetGrow - GSGLearn.com

SERIALIZABLE ISOLATION

EXPLANATION

The highest isolation level, ensuring complete isolation from other transactions.

EXAMPLE

Ensuring that transactions are executed in a serial order, one after the other.

Shwetank Singh
GritSetGrow - GSGLearn.com

READ COMMITTED ISOLATION

EXPLANATION

An isolation level that prevents reading uncommitted data but allows non-repeatable reads.

EXAMPLE

Reading only committed changes but may see different data on subsequent reads of the same row.

Shwetank Singh
GritSetGrow - GSGLearn.com

READ UNCOMMITTED ISOLATION

EXPLANATION

The lowest isolation level, allowing transactions to read uncommitted data.

EXAMPLE

Reading changes made by other transactions before they are committed.

Shwetank Singh
GritSetGrow - GSGLearn.com

REPEATABLE READ ISOLATION

EXPLANATION

An isolation level that prevents non-repeatable reads but allows phantom reads.

EXAMPLE

Ensuring the same data is read multiple times within a transaction, but new rows may be added by others.

Shwetank Singh
GritSetGrow - GSGLearn.com

TWO-PHASE COMMIT

EXPLANATION

A protocol to ensure all participants in a distributed transaction agree to commit or roll back the transaction.

EXAMPLE

Coordinating a transaction across multiple databases to ensure consistency.

Shwetank Singh
GritSetGrow - GSGLearn.com

CONCURRENCY CONTROL

EXPLANATION

Mechanisms to manage simultaneous operations without conflicting.

EXAMPLE

Using locking or timestamping to manage concurrent access to data.

Shwetank Singh
GritSetGrow - GSGLearn.com

OPTIMISTIC LOCKING

EXPLANATION

A concurrency control method that checks for conflicts only at the end of a transaction.

EXAMPLE

Allowing multiple users to update data and only checking for conflicts during the commit phase.

Shwetank Singh
GritSetGrow - GSGLearn.com

PESSIMISTIC LOCKING

EXPLANATION

A concurrency control method that locks data during a transaction to prevent conflicts.

EXAMPLE

Locking a record when it is read to prevent other transactions from modifying it until the lock is released.

Shwetank Singh
GritSetGrow - GSGLearn.com

DATA WAREHOUSE

EXPLANATION

A central repository of integrated data from multiple sources, used for reporting and analysis.

EXAMPLE

A system that consolidates sales, inventory, and customer data for business intelligence purposes.

Shwetank Singh
GritSetGrow - GSGLearn.com

ETL PROCESS

EXPLANATION

Extract, Transform, Load - a process of extracting data from source systems, transforming it, and loading it into a data warehouse.

EXAMPLE

Extracting sales data from transactional systems, cleaning and aggregating it, and loading it into a data warehouse.

Shwetank Singh
GritSetGrow - GSGLearn.com

OLAP

EXPLANATION

Online Analytical Processing - tools for analyzing data stored in a data warehouse, supporting complex queries and data analysis.

EXAMPLE

Analyzing sales trends over time using OLAP cubes.

Shwetank Singh
GritSetGrow - GSGLearn.com

OLTP

EXPLANATION

Online Transaction Processing - systems designed to manage transactional data, supporting day-to-day operations.

EXAMPLE

A banking system that processes deposits, withdrawals, and transfers in real-time.

Shwetank Singh
GritSetGrow - GSGLearn.com

SHARDING

EXPLANATION

A database architecture pattern that horizontally partitions data across multiple database instances.

EXAMPLE

Distributing customer data across multiple databases based on geographic regions.

Shwetank Singh
GritSetGrow - GSGLearn.com

REPLICATION

EXPLANATION

The process of copying and maintaining database objects in multiple databases to ensure data redundancy and availability.

EXAMPLE

Replicating a primary database to a secondary database for failover and disaster recovery.

Shwetank Singh
GritSetGrow - GSGLearn.com

MASTER-SLAVE REPLICATION

EXPLANATION

A replication model where one database (master) handles write operations, and replicas (slaves) handle read operations.

EXAMPLE

Using master-slave replication to distribute read requests across multiple replicas for load balancing.

Shwetank Singh
GritSetGrow - GSGLearn.com

EVENTUAL CONSISTENCY

EXPLANATION

A consistency model for distributed databases where updates are propagated gradually and all nodes eventually reach a consistent state.

EXAMPLE

Ensuring that updates to a distributed database are eventually reflected in all replicas.

Shwetank Singh
GritSetGrow - GSGLearn.com

STRONG CONSISTENCY

EXPLANATION

A consistency model where a system guarantees that all reads return the most recent write.

EXAMPLE

Ensuring that all read operations return the latest data in a distributed database.

Shwetank Singh
GritSetGrow - GSGLearn.com

THANK YOU