**Normalization: Transitioning Through 1NF and 2NF to 3NF**

**1NF (First Normal Form)**

In **1NF**, all data must be atomic (no multi-valued attributes), and there must be a unique identifier for each row.

**Customer Management**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CustomerID** | **FirstName** | **LastName** | **Email** | **PhoneNumber** | **AddressLine1** | **AddressLine2** | **City** | **PostalCode** | **RegistrationDate** |
| C123 | John | Doe | [john@email.com](mailto:john@email.com) | 1234567890 | Line1 | Line2 | New York | 10001 | 2024-01-01 |
| C124 | Jane | Smith | [jane@email.com](mailto:jane@email.com) | 0987654321 | Line3 | Line4 | Chicago | 60606 | 2024-02-01 |

**Product Management**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ProductID** | **ProductName** | **Category** | **Brand** | **UnitPrice** | **TaxRate** | **ReorderLevel** | **IsActive** |
| P101 | TV | Electronics | Sony | 1000.00 | 10% | 5 | Yes |
| P102 | Sofa | Furniture | Ikea | 500.00 | 5% | 2 | Yes |

**Stock Management**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **StockID** | **ProductID** | **Warehouse** | **QuantityAvailable** | **LastRestockedDate** |
| S001 | P101 | Warehouse1 | 10 | 2024-01-10 |
| S002 | P102 | Warehouse2 | 5 | 2024-01-15 |

**Discount Management**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DiscountID** | **DiscountName** | **StartDate** | **EndDate** | **DiscountDetails** |
| D001 | Summer Sale | 2024-06-01 | 2024-06-30 | 15%, Percentage |

**Sales Management**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **TransactionID** | **CustomerID** | **TransactionDate** | **PaymentMethod** | **TotalAmount** | **DiscountApplied** | **FinalAmount** | **ProductDetails** |
| T001 | C123 | 2024-01-05 | Credit Card | 2000.00 | 15% | 1700.00 | P101: 2 units, $1000/unit |

**Employee Management**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **EmployeeID** | **FirstName** | **LastName** | **Role** | **ContactInfo** | **HireDate** |
| E001 | Alice | Brown | Manager | [alice@company.com](mailto:alice@company.com), 1234567 | 2020-01-01 |

**2NF (Second Normal Form)**

To achieve **2NF**, all partial dependencies (fields depending only on part of a composite key) are removed. This step involves creating separate tables for attributes that do not directly depend on the primary key.

**Customer Management**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CustomerID** | **FirstName** | **LastName** | **Email** | **RegistrationDate** |
| C123 | John | Doe | [john@email.com](mailto:john@email.com) | 2024-01-01 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CustomerID** | **PhoneNumber** | **AddressLine1** | **AddressLine2** | **City** | **PostalCode** |
| C123 | 1234567890 | Line1 | Line2 | New York | 10001 |

**Product Management**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ProductID** | **ProductName** | **UnitPrice** | **ReorderLevel** | **IsActive** |
| P101 | TV | 1000.00 | 5 | Yes |

|  |  |
| --- | --- |
| **ProductID** | **Category** |
| P101 | Electronics |

|  |  |
| --- | --- |
| **ProductID** | **Brand** |
| P101 | Sony |

|  |  |
| --- | --- |
| **ProductID** | **TaxRate** |
| P101 | 10% |

**Stock Management**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **StockID** | **ProductID** | **Warehouse** | **QuantityAvailable** | **LastRestockedDate** |
| S001 | P101 | Warehouse1 | 10 | 2024-01-10 |

**Discount Management**

|  |  |  |  |
| --- | --- | --- | --- |
| **DiscountID** | **DiscountName** | **StartDate** | **EndDate** |
| D001 | Summer Sale | 2024-06-01 | 2024-06-30 |

|  |  |  |
| --- | --- | --- |
| **DiscountID** | **DiscountType** | **DiscountValue** |
| D001 | Percentage | 15% |

**Sales Management**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **TransactionID** | **CustomerID** | **TransactionDate** | **PaymentMethod** | **TotalAmount** | **DiscountApplied** | **FinalAmount** |
| T001 | C123 | 2024-01-05 | Credit Card | 2000.00 | 15% | 1700.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TransactionID** | **ProductID** | **Quantity** | **UnitPrice** | **LineTotal** |
| T001 | P101 | 2 | 1000.00 | 2000.00 |

**Employee Management**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **EmployeeID** | **FirstName** | **LastName** | **RoleID** | **HireDate** |
| E001 | Alice | Brown | R001 | 2020-01-01 |

|  |  |
| --- | --- |
| **RoleID** | **RoleName** |
| R001 | Manager |

|  |  |
| --- | --- |
| **EmployeeID** | **ContactInfo** |
| E001 | [alice@company.com](mailto:alice@company.com), 1234567 |

To achieve **3NF**, all transitive dependencies (attributes depending on non-primary key attributes) are removed. The following tables demonstrate the fully normalized structure:

**Customer Management**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CustomerID** | **FirstName** | **LastName** | **Email** | **RegistrationDate** |
| C123 | John | Doe | [john@email.com](mailto:john@email.com) | 2024-01-01 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CustomerID** | **PhoneNumber** | **AddressLine1** | **AddressLine2** | **City** | **PostalCode** |
| C123 | 1234567890 | Line1 | Line2 | New York | 10001 |

**Product Management**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ProductID** | **ProductName** | **CategoryID** | **BrandID** | **UnitPrice** | **TaxRateID** | **ReorderLevel** | **IsActive** |
| P101 | TV | CAT01 | BR01 | 1000.00 | TAX01 | 5 | Yes |

|  |  |
| --- | --- |
| **CategoryID** | **CategoryName** |
| CAT01 | Electronics |

|  |  |
| --- | --- |
| **BrandID** | **BrandName** |
| BR01 | Sony |

|  |  |
| --- | --- |
| **TaxRateID** | **TaxRate** |
| TAX01 | 10% |

**Stock Management**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **StockID** | **ProductID** | **WarehouseID** | **QuantityAvailable** | **LastRestockedDate** |
| S001 | P101 | WH01 | 10 | 2024-01-10 |

|  |  |  |  |
| --- | --- | --- | --- |
| **WarehouseID** | **WarehouseName** | **Location** | **ManagerID** |
| WH01 | Warehouse1 | New York City | E001 |

**Discount Management**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DiscountID** | **DiscountName** | **StartDate** | **EndDate** | **DiscountType** | **DiscountValue** |
| D001 | Summer Sale | 2024-06-01 | 2024-06-30 | Percentage | 15% |

**Sales Management**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **TransactionID** | **CustomerID** | **TransactionDate** | **PaymentMethodID** | **TotalAmount** | **DiscountApplied** | **FinalAmount** |
| T001 | C123 | 2024-01-05 | PM01 | 2000.00 | 15% | 1700.00 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TransactionDetailID** | **TransactionID** | **ProductID** | **Quantity** | **UnitPrice** | **LineTotal** |
| TD001 | T001 | P101 | 2 | 1000.00 | 2000.00 |

|  |  |
| --- | --- |
| **PaymentMethodID** | **PaymentType** |
| PM01 | Credit Card |

**Employee Management**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **EmployeeID** | **FirstName** | **LastName** | **RoleID** | **Email** | **PhoneNumber** | **HireDate** |
| E001 | Alice | Brown | R001 | [alice@company.com](mailto:alice@company.com) | 1234567890 | 2020-01-01 |

|  |  |
| --- | --- |
| **RoleID** | **RoleName** |
| R001 | Manager |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **LogID** | **Action** | **PerformedBy** | **ActionTimestamp** | **TableName** |
| LOG001 | Update Stock | E001 | 2024-01-10 10:30:00 | Stock |

**Report Generation**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ReportID** | **ReportName** | **GeneratedBy** | **GeneratedDate** | **ReportType** | **ReportData** |
| REP001 | Daily Sales Report | E001 | 2024-01-11 | Sales | JSON Data |

**Normalization Process: 1NF, 2NF, and 3NF**

Normalization is a systematic approach to organizing data in a database to reduce redundancy and improve data integrity. This document explains normalization through **1NF**, **2NF**, and **3NF**, using examples to demonstrate how redundant data is removed at each stage.

**1NF (First Normal Form)**

**Definition**

1NF ensures that:

* All columns in a table contain atomic (indivisible) values.
* Each record is unique, and there are no duplicate rows.
* There are no repeating groups or arrays within the table.

**Example: Data in Unnormalized Form (0NF)**

Consider a CustomerOrders table that stores customer orders.

|  |  |  |
| --- | --- | --- |
| **CustomerID** | **CustomerName** | **Orders** |
| C001 | John Doe | TV: 2 units, $1000/unit; Sofa: 1 unit, $500 |
| C002 | Jane Smith | TV: 1 unit, $1000/unit |

**Transition to 1NF**

To achieve 1NF:

* Split multi-valued attributes (e.g., Orders) into atomic values.
* Create a separate row for each unique order.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CustomerID** | **CustomerName** | **ProductName** | **Quantity** | **UnitPrice** |
| C001 | John Doe | TV | 2 | 1000 |
| C001 | John Doe | Sofa | 1 | 500 |
| C002 | Jane Smith | TV | 1 | 1000 |

**2NF (Second Normal Form)**

**Definition**

2NF eliminates partial dependencies, where non-primary key attributes depend only on part of a composite primary key. To achieve 2NF:

* Move attributes that are dependent on only part of the primary key to a separate table.
* Ensure all attributes depend on the entire primary key.

**Example: Data in 1NF**

Continuing with the CustomerOrders table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CustomerID** | **CustomerName** | **ProductName** | **Quantity** | **UnitPrice** |
| C001 | John Doe | TV | 2 | 1000 |
| C001 | John Doe | Sofa | 1 | 500 |
| C002 | Jane Smith | TV | 1 | 1000 |

**Identifying Partial Dependencies**

* CustomerName depends only on CustomerID.
* UnitPrice depends only on ProductName.

**Transition to 2NF**

Split the table into two separate tables:

**Customer Table**

|  |  |
| --- | --- |
| **CustomerID** | **CustomerName** |
| C001 | John Doe |
| C002 | Jane Smith |

**Product Table**

|  |  |
| --- | --- |
| **ProductName** | **UnitPrice** |
| TV | 1000 |
| Sofa | 500 |

**CustomerOrders Table**

|  |  |  |
| --- | --- | --- |
| **CustomerID** | **ProductName** | **Quantity** |
| C001 | TV | 2 |
| C001 | Sofa | 1 |
| C002 | TV | 1 |

By eliminating partial dependencies, redundancy is reduced. For example, the price of a product is stored only once in the Product table, rather than being repeated for every order.

**3NF (Third Normal Form)**

**Definition**

3NF eliminates transitive dependencies, where a non-primary key attribute depends on another non-primary key attribute. To achieve 3NF:

* Move attributes that do not directly depend on the primary key to a separate table.

**Example: Data in 2NF**

Consider the Customer table:

|  |  |  |  |
| --- | --- | --- | --- |
| **CustomerID** | **CustomerName** | **City** | **PostalCode** |
| C001 | John Doe | New York | 10001 |
| C002 | Jane Smith | Chicago | 60606 |

**Identifying Transitive Dependencies**

* City depends on PostalCode.
* CustomerID is the primary key.

**Transition to 3NF**

Split the table into two separate tables:

**Customer Table**

|  |  |  |
| --- | --- | --- |
| **CustomerID** | **CustomerName** | **PostalCode** |
| C001 | John Doe | 10001 |
| C002 | Jane Smith | 60606 |

**Location Table**

|  |  |
| --- | --- |
| **PostalCode** | **City** |
| 10001 | New York |
| 60606 | Chicago |

By removing transitive dependencies, redundancy is further reduced. For instance, the city corresponding to a postal code is stored only once in the Location table.

**Summary of Normalization Steps**

|  |  |  |
| --- | --- | --- |
| **Normalization Form** | **Key Action** | **Result** |
| 1NF | Remove multi-valued attributes | Atomic values and unique rows |
| 2NF | Eliminate partial dependencies | Attributes depend on the entire primary key |
| 3NF | Eliminate transitive dependencies | Attributes depend only on the primary key |

By applying normalization, we ensure the database is efficient, consistent, and free of unnecessary redundancy. This improves data integrity and minimizes the risk of anomalies during data manipulation.