DS-2002: Data Systems

An Overview of SQL Databases

Prof. Jon Tupitza – Fall 2022

UNIVERSITY of VIRGINIA

SQL Database Design

Understanding the Principles that Govern Database Structure

Fundamental Structures: Enforcing Data Integrity

Essential Design Concepts & Database Objects Required for Enforcing Data Integrity

Entity Integrity

- Enforced by the **Table**
  - Entities (nouns):
    - People, Places and Things
  - Concrete: Employees, Customers, Products
  - Conceptual: Sales, Scenarios, etc.

Domain Integrity

- Enforced by the **Column**
  - Data Type definition:
    - Int, Decimal, Float, Char, Nchar, Varchar, Nvarchar, DateTime
  - Constraints:
    - Primary Key, Check, Unique, & Default

Relational Integrity

- Enforced by the **Foreign Key Relationship**
  - One-to-Many: Foreign key relates to Primary key
  - Many-to-Many: Primary keys relate to Foreign keys via a *Juncture table*

Database Normalization: The Normal Forms

There are other Normal Forms, but Resolving to 3rd NF is Considered Appropriate

First Normal Form (1NF)

- A tables columns must contain only atomic values; none may not contain multiple values
- **Ex:** a column named **telephone_number** may contain only one phone number.

Second Normal Form (2NF)

- The table must first satisfy the first normal form.
- The table must be free of partial dependencies; i.e., all columns that are not the Primary Key must depend on the Primary Key

Third Normal Form (3NF)

- The table must first satisfy both the first and second normal forms
- The table must be free of transitive dependencies; i.e., no column may depend on any column that is not the Primary Key.

Workload Characteristics: Form Follows Function

Two Essentially Incompatible Workloads… They Have a Contentious Relationship

Online Transaction Processing (OLTP)

- Characterized by a large volume of transactions each of which affect a small number of rows
- Online Sales, Bank Deposits & Transfers
- Highly Normalized Database Schema

Online Analytical Processing (OLAP)

- Characterized by a small volume of read transactions each of which affect a large number of rows
- Periodic Post-hoc Analysis *(What Happened?)*
- De-Normalized Multi-Dimensional Schema

! These two **don't** play well together: They contend for the same hardware resources!

Database Paradigms: Design Approaches

The Design Approach Accommodates the Workload Characteristic

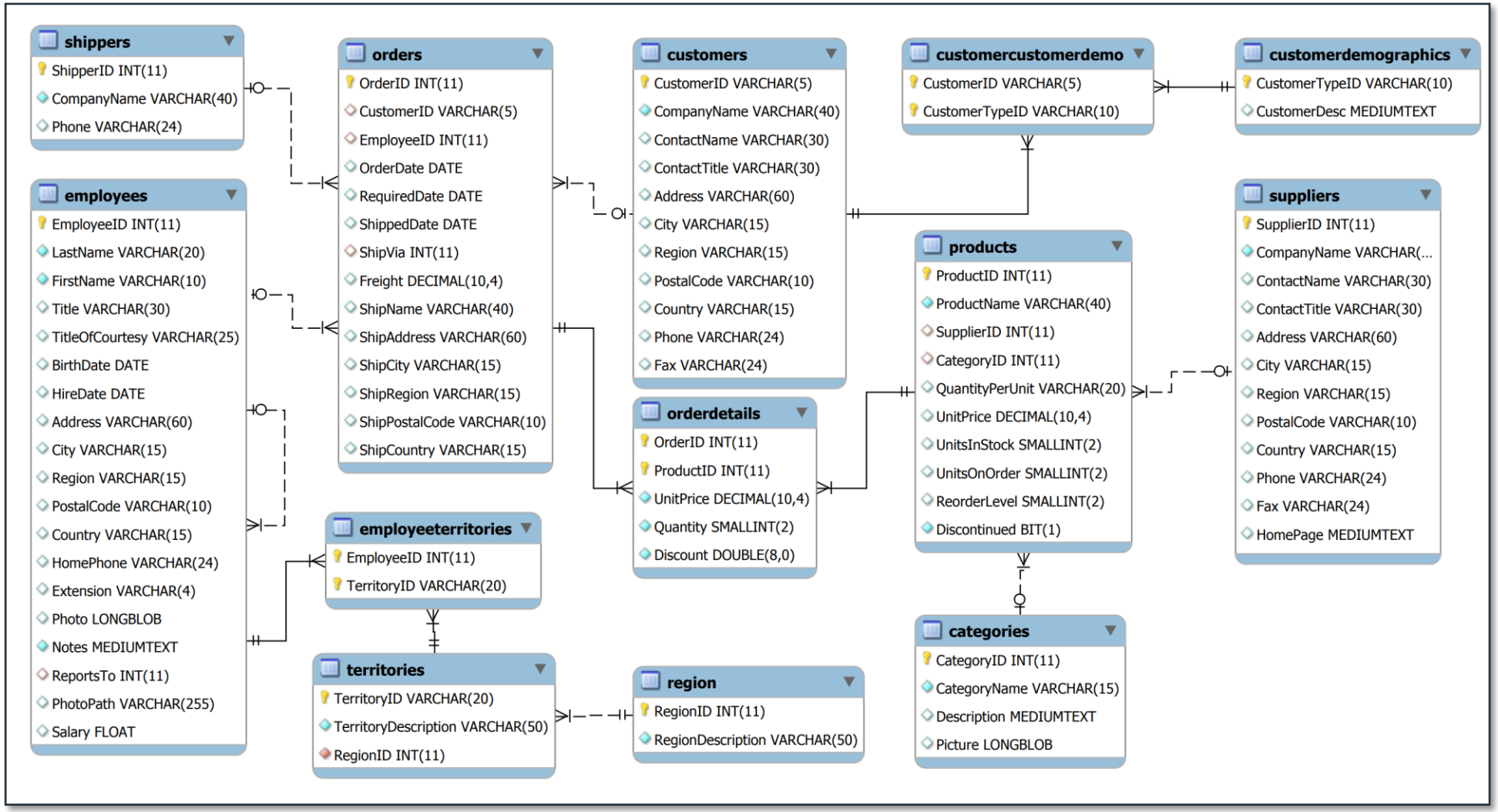
Normalized Relational Database:

- Optimized for Online Transaction Processing (OLTP) workloads
- Aims to Eliminate Data Redundancy and Minimize Storage Requirements
- **Complex:** Sacrifices User-Friendliness in Favor of Transactional Performance

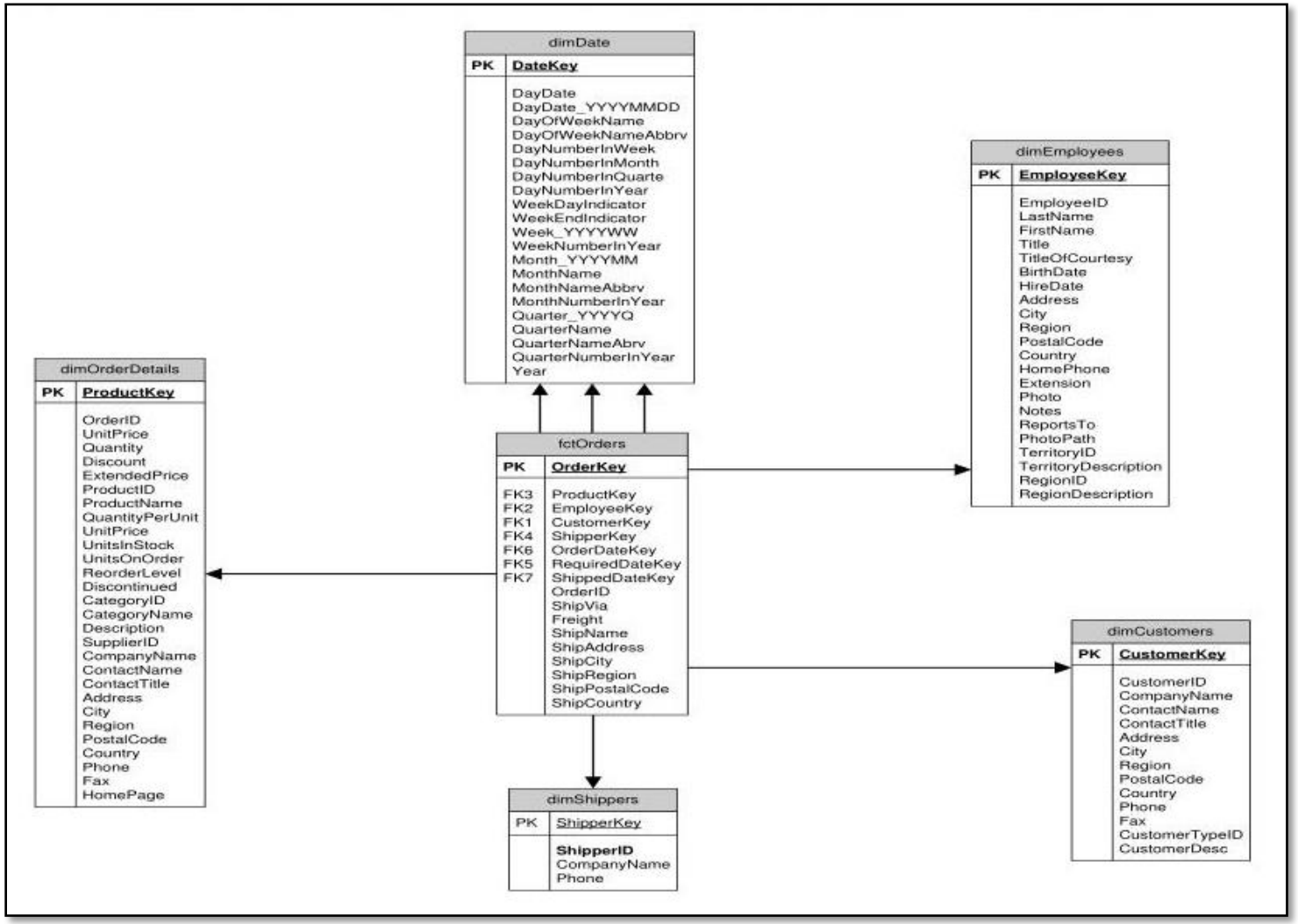
Multi-Dimensional Relational Database:

- Optimized for Online Analytical Processing (OLAP) workloads
- Aims to Optimize Query Performance and Provide an Intuitive User Experience
- **Simple:** Accepts Data Repetition in Favor of User-Friendliness and Improved Query Performance

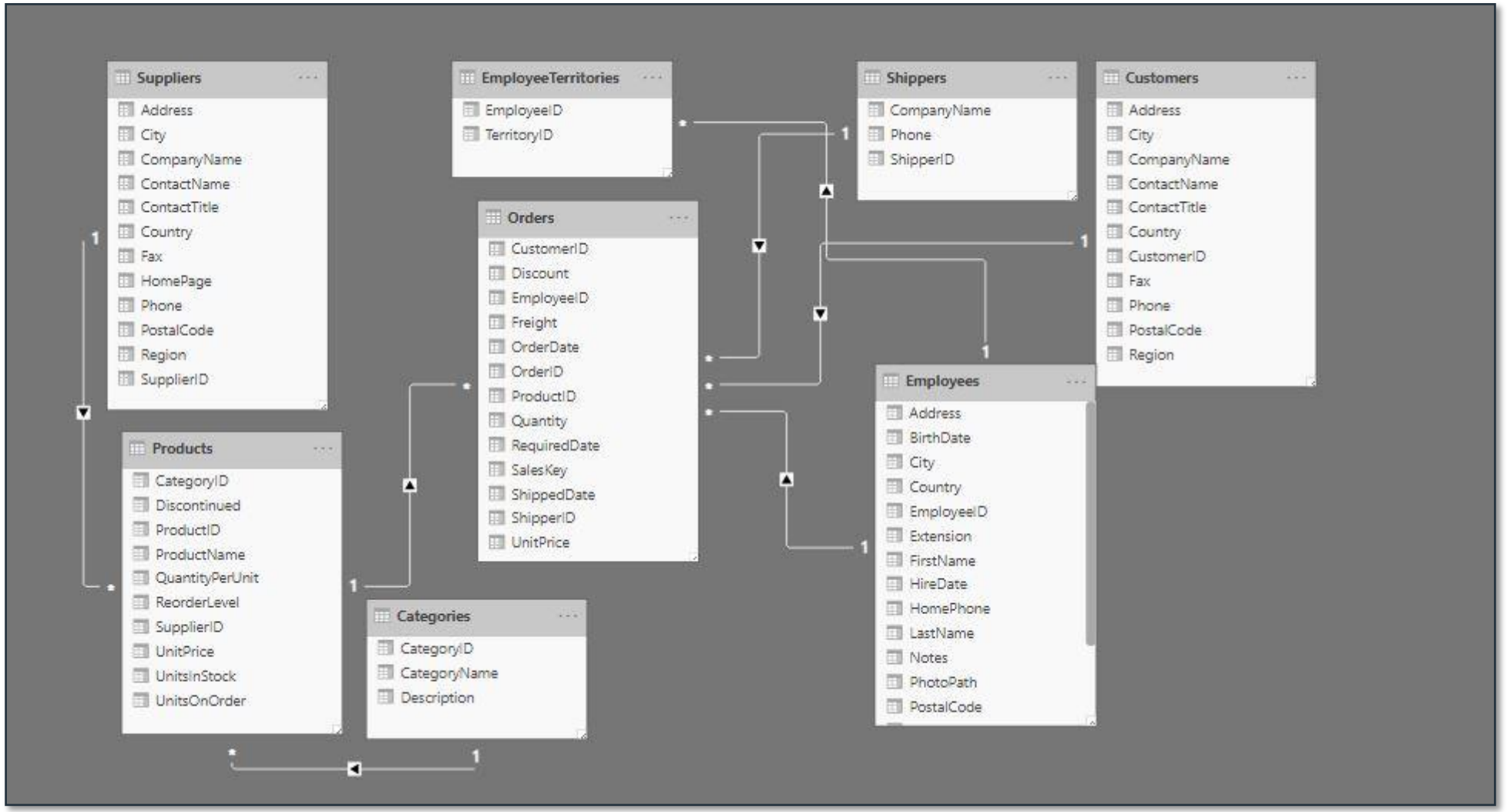
OLTP Database: Normalized Schema



OLAP Database: Multi-Dimensional (Star) Schema



OLAP Database: Snowflake Schema



The SQL Language

Understanding the Structured Query Language

UNIVERSITY of VIRGINIA

The SQL Language: Principal Components

Three Primary Aspects of the ANSI-Compliant SQL Language

Data Definition Language (DDL)

- CREATE, ALTER, DROP, TRUNCATE TABLE, ENABLE & DISABLE TRIGGER
- *Used to manage database structures*

Data Control Language (DCL)

- GRANT, REVOKE, DENY, EXECUTE AS
- *Used to control access to server & database objects (permissions)*

Data Manipulation Language (DML)

- SELECT, INSERT, UPDATE, DELETE, MERGE, and BULK INSERT
- *Used to manipulate database content (data)*

Microsoft Docs | Transact-SQL Reference (Database Engine)

Query a SQL Database: The SELECT Statement

Essential Components of Data Retrieval

SELECT	for specifying the required columns
FROM	for specifying the Table(s) being targeted
JOIN	for specifying additional Table(s)
GROUP BY	for specifying points of aggregation
ORDER BY	for sorting rows of the result set

Filtering Statements:

ON	WHERE	HAVING
specifies the column(s) that enable the joining of two Tables	specifies conditions by which to reduce the rows returned	specifies conditions by which Groups or Aggregates may be reduced

Q & A

A Survey of Data Management Systems