Physical Design

by Sophia

WHAT'S COVERED

This lesson explores the last step in designing a database. We will explore physical model design, in three parts. Specifically, this lesson will cover:

1. Rationale for Physical Design

Creating a **physical model** is the final step in the database design process. This data model describes how the database will be implemented using a specific database management system. The database administrator and the database developers create this model. Data requirements and relationships are represented in the conceptual model; however, a database management system (DBMS) is used to implement the technical implementation details of the database.

The physical model converts the conceptual model into a format optimized for the data management system. This is the build stage of the design process that relies on the conceptual and logical stages to be complete. Any issues not addressed early will appear at the final physical stage.

In the conceptual model, entities and attributes are represented by physical tables, columns, and data types.

An index optimizes data retrieval and query performance by determining which columns should be indexed. Data integrity constraints, including primary keys, foreign keys, and check constraints, are implemented to ensure data is accurate and consistent. Data is partitioned and distributed across multiple servers or partitions in distributed databases to achieve scalability and load balancing. We implement access controls and security measures to protect sensitive data and ensure data privacy.

Query optimization executes plans based on the physical model's indexing and data distribution strategies. Optimizing database parameters and configuring settings assist in achieving optimal performance under given hardware and software conditions. Some of this might be physical hardware, networking, RAM, and other underlying components, perhaps even including the access time of the storage devices on which the data resides. Optimization takes many forms, including caching frequently requested information and other ways of speeding up data presentation and retrieval. You can do many things to speed up a database's performance.

A well-designed physical model is essential in order to meet the application's performance and scalability requirements. DBAs use this document to create and configure the actual database schema, tables, and indexes within the selected database management system (DBMS). Databases can handle large amounts of data by converting conceptual models into well-optimized, technically sound physical designs that support multiple concurrent users and deliver fast and reliable query responses.

☆ BIG IDEA

It is important to note that the physical data model will be different depending on the relational database management system (e.g., PostgreSQL) that has been selected. Each database will use its own specific data types.

TERM TO KNOW

Physical Model

A technical plan for implementing the details outlined in the conceptual and logical models using a specific database management system (DBMS).

2. Details of the Database

The physical design focuses on data storage, security measures, and performance measures. Before defining the data storage organization, we have to know the volume of the data and usage patterns. With the sheer quantity of data that a company creates, it is very important to understand the volume, variety, and velocity of data for your design to work well.

KEY CONCEPT

The three V's of data are volume, velocity, and variety.

Large-scale data sets present unique challenges, and these characteristics describe what makes big data different from traditional data management:

- **Volume** refers to the enormous amount of data generated and collected. Data is being produced at an unprecedented rate due to the proliferation of digital devices and systems. Big data storage and processing solutions are required to handle massive volumes of data efficiently.
- **Velocity** relates to the speed of data generation and the need to process and analyze data in real time. Applications that use big data often deal with data streams that require continuous analysis so that timely insights can be derived, and dynamic decisions can be made.
- **Variety** signifies the diverse types and formats of data and big data. This includes structured data (e.g., traditional databases), semi-structured data (e.g., JSON, XML), and unstructured data (e.g., text, images, videos). Various data sources and formats necessitate flexible data storage and processing approaches.

Managing, storing, processing, and analyzing high-volume data presents challenges. A combination of advanced data management techniques, distributed computing, cloud technologies, and innovative data analytics methods is required to address these aspects. As a result, big data's vast and diverse universe can provide valuable insights and opportunities. This is a unique but growing subset of database systems. While we allude to them throughout this course, understanding how the “3 V's” apply to most databases will help you design a much more capable and functional database system.

The column data types and sizes are also defined here. We need the indexes for each table beyond the primary keys, depending on the usage patterns or performance requirements. If there are any anticipated views or reports (especially recurring reports or dashboards) to create for the database, they would be useful to define now. If you are revising an existing system, using the older report, view, and dashboards, it is a good idea to get started with optimizing the system.

With respect to security, we take the time to define the group roles and user roles that can access this database and what level of privileges they should have for each object. This goes beyond the physical data model itself, but it is a crucial step to take when planning for the physical database design. Some data is regulated under government programs such as HIPAA, FERPA, or COPA. This is the optimal time to ensure that the data storage, access, and processing meets all legal, regulatory, and industry-level requirements.

All of the constraints we have defined as part of the logical data model, such as requiring that certain attributes contain unique or non-null values, would also be implemented as part of the physical data model. Once this model has been created, it should be ready to be implemented in the database.

TERMS TO KNOW

Variety

The diverse types and formats of data and big data. This includes structured data (e.g., traditional databases), semi-structured data (e.g., JSON, XML), and unstructured data (e.g., text, images, videos). The variety of data sources and formats necessitates flexible data storage and processing approaches.

Velocity

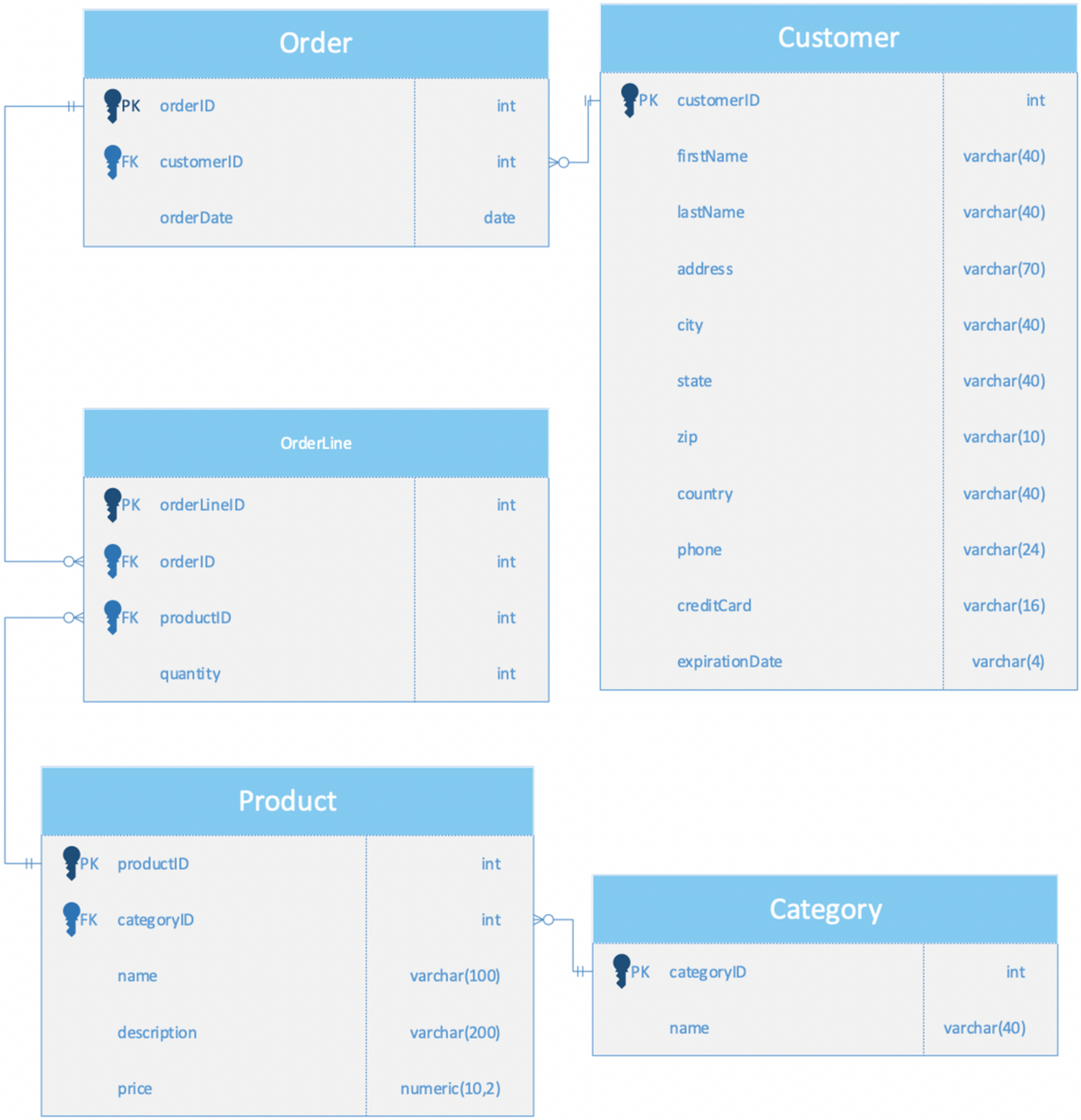
The speed of data generation and the need to process and analyze data in real time.

Volume

The enormous amount of data generated and collected. Big data storage and processing solutions are required to handle massive volumes of data efficiently.

3. E-Commerce Example

Here is a complete physical data model for the e-commerce database. You may notice that many of the core elements are identical to the logical data model, with added data types and sizes. In other databases, changes with added constraints may depend on the business rules or other criteria being incorporated.



Order

PK	orderID	int
FK	customerID	int
	orderDate	date

Customer

PK	customerID	int
	firstName	varchar(40)
	lastName	varchar(40)
	address	varchar(70)
	city	varchar(40)
	state	varchar(40)
	zip	varchar(10)
	country	varchar(40)
	phone	varchar(24)
	creditCard	varchar(16)
	expirationDate	varchar(4)

OrderLine

PK	orderLineID	int
FK	orderID	int
FK	productID	int
	quantity	int

Product

PK	productID	int
FK	categoryID	int
	name	varchar(100)
	description	varchar(200)
	price	numeric(10,2)

Category

PK	categoryID	int
	name	varchar(40)

SUMMARY

In this lesson, you learned about the **physical design** of a database. You learned that in database design, a physical model refers to the actual schema implementation within a database management system (DBMS). Translating the logical model, which represents the data requirements and relationships, into an optimized DBMS format is the focus of the process. You also learned about specific **details related to databases**. Physical models define the data types, column lengths, and constraints for expressing the entities and attributes defined in the logical model. To ensure accurate and consistent data, they implement integrity constraints, such as primary keys and foreign keys, to optimize data retrieval and query performance. The physical model also addresses data partitioning, allocation, and distribution to achieve scalability and load balancing, particularly in distributed databases. The physical model also requires performance tuning and optimization, such as parameter configuration, query optimization, and access controls.

You learned in the **e-commerce example** that physical models provide the blueprint for creating real databases within the selected database management system. To meet the application's performance, scalability, and security requirements, the physical model guides database administrators and developers through implementing the database schema, tables, indexes, and access controls. Its successful implementation ensures that the database operates efficiently, delivering fast and reliable query responses while handling large volumes of data.

Source: THIS TUTORIAL WAS AUTHORED BY DR. VINCENT TRAN, PHD (2020) AND Faithe Wempen (2024) FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

TERMS TO KNOW

Physical Model

A technical plan for implementing the details outlined in the conceptual and logical models using a specific database management system (DBMS).

Variety

The diverse types and formats of data and big data. This includes structured data (e.g., traditional databases), semi-structured data (e.g., JSON, XML), and unstructured data (e.g., text, images, videos). The variety of data sources and formats necessitates flexible data storage and processing approaches.

Velocity

The speed of data generation and the need to process and analyze data in real time.

Volume

The enormous amount of data generated and collected. Big data storage and processing solutions are required to handle massive volumes of data efficiently.