

Lesson 5

Topic: Data Modeling Basics

Prerequisites: Download customer.csv, product.csv, sales.csv

1. What is a primary key in a table?
2. Name the two types of table relationships in Power BI.
3. How do you create a relationship between two tables in Power BI?
4. What is a "star schema"?
5. Which table is typically the fact table in a sales dataset?
6. Link Sales.csv to Customers.csv using CustomerID (one-to-many).
7. Why is ProductID in Sales.csv a foreign key?
8. Fix a relationship error where ProductID has mismatched data types.
9. Explain why a star schema improves performance.
10. Add a new column TotalSales in Sales (Quantity * Price from Products).
11. Optimize a model with circular relationships—how would you resolve it?
12. Create a role-playing dimension for OrderDate and ShipDate.
13. Handle a many-to-many relationship between Customers and Products.
14. Use bidirectional filtering sparingly—when is it appropriate?
15. Write DAX to enforce referential integrity if a CustomerID is deleted.

Sales

OrderID	CustomerID	ProductID	Quantity	OrderDate
1001	101	1	2	1/5/2023
1002	102	2	5	1/10/2023
1003	101	3	1	1/15/2023

Products

ProductID	ProductName	Category	Price
1	Laptop	Electronics	1200
2	Mouse	Accessories	25
3	Keyboard	Accessories	80

Customer

CustomerID	Name	Region
101	Alice	North
102	Bob	South
103	Charlie	East

Решения:

Lesson 5: Data Modeling Basics

1. What is a primary key in a table?

Primary key — это столбец (или набор столбцов), который:

- **уникально идентифицирует** каждую строку таблицы
- не содержит `null`
- используется для создания связей между таблицами

Пример: `CustomerID` в таблице `Customers`.

2. Two types of table relationships in Power BI

1. **One-to-Many (1:*)** — самый распространённый тип
 2. **Many-to-Many (:)** — используется реже, требует аккуратного моделирования
-

3. How do you create a relationship in Power BI?

Способы:

- **Model View → Drag & Drop** поля между таблицами
- **Manage Relationships → New**

Указываются:

- столбцы
 - тип связи
 - направление фильтрации
-

4. What is a star schema?

Star schema — модель данных, где:

- в центре находится **fact table**
 - вокруг — **dimension tables**
 - связи только ***1: ****
 - нет цепочек и циклов
-

5. Which table is the fact table in a sales dataset?

→ **Sales**

Потому что содержит:

- транзакции
 - числовые показатели (Quantity)
 - внешние ключи
-

Practical Tasks

6. Link Sales.csv to Customers.csv (one-to-many)

- Customers[CustomerID] → **1**
 - Sales[CustomerID] → **many**
 - Relationship: **One-to-Many**
 - Filter direction: **Single**
-

7. Why is ProductID in Sales.csv a foreign key?

Потому что:

- ссылается на **primary key** Products[ProductID]
 - связывает факт продаж с размерностью продуктов
 - не обязан быть уникальным
-

8. Fix mismatched data types for ProductID

Steps:

1. Проверить тип данных в обеих таблицах
2. Установить **Whole Number** в Sales[ProductID] и Products[ProductID]
3. Обновить связь

9. Why does a star schema improve performance?

- Упрощает фильтрацию
- Снижает количество JOIN'ов
- Улучшает сжатие VertiPaq
- Повышает скорость DAX-вычислений

10. Add TotalSales column in Sales

```
TotalSales =  
Sales[Quantity] *  
RELATED ( Products[Price] )
```

11. Circular relationships — how to resolve?

Решения:

- Удалить лишние связи
- Использовать **bridge table**
- Заменить bidirectional на single
- Применять DAX вместо связей

12. Role-playing dimension (OrderDate & ShipDate)

Steps:

1. Создать таблицу **Date**
2. Создать две связи:
 - Date → Sales[OrderDate]
 - Date → Sales[ShipDate]
3. Одна связь активная, вторая — неактивная
4. Использовать USERELATIONSHIP() в DAX

13. Handle many-to-many (Customers ↔ Products)

Решение:

- Создать **bridge table** (CustomerID, ProductID)
 - Связать:
 - Customers → Bridge (1:*)
 - Products → Bridge (1:*)
-

14. Use bidirectional filtering sparingly — when?

Подходит, если:

- есть many-to-many
- требуется фильтрация в обе стороны
- вы понимаете влияние на производительность

В стандартной звезде — **не рекомендуется**

15. DAX to enforce referential integrity

```
Valid Sales =  
IF (  
    ISBLANK (  
        LOOKUPVALUE (  
            Customers[CustomerID],  
            Customers[CustomerID],  
            Sales[CustomerID]  
        )  
    ),  
    BLANK(),  
    Sales[Quantity]  
)
```

Исключает продажи с удалённым CustomerID.

Sample Data

Sales

OrderID	CustomerID	ProductID	Quantity	OrderDate
---------	------------	-----------	----------	-----------

1001	101	1	2	1/5/2023
1002	102	2	5	1/10/2023
1003	101	3	1	1/15/2023

Products

ProductID	ProductName	Category	Price
-----------	-------------	----------	-------

1	Laptop	Electronics	1200
2	Mouse	Accessories	25
3	Keyboard	Accessories	80

Customers

CustomerID	Name	Region
------------	------	--------

101	Alice	North
102	Bob	South

CustomerID Name Region

103 Charlie East