Lesson 5: Data Modeling Basics

1. What is a primary key in a table?

A primary key is a unique identifier for each row in a table. Example: CustomerID in Customers.csv or ProductID in Products.csv.

---

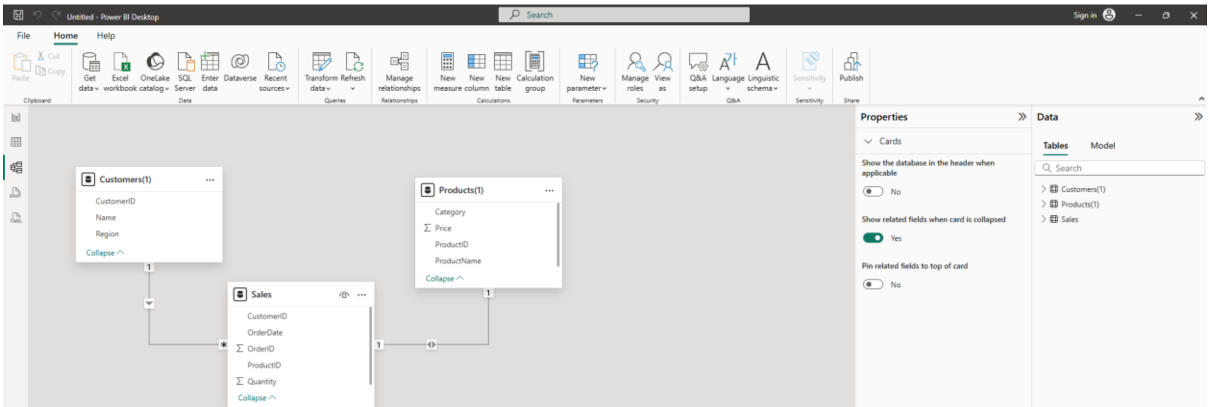
2. Name the two types of table relationships in Power BI.

- **One-to-Many (1:*):** One record in a table relates to many in another (e.g., one customer → many sales).
- **Many-to-Many (:):** Both tables have overlapping keys, usually requiring a bridge table.

---

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

- Go to **Model view** → drag the key column from one table (e.g., CustomerID in Customers.csv) onto the corresponding column in the other (e.g., CustomerID in Sales.csv).



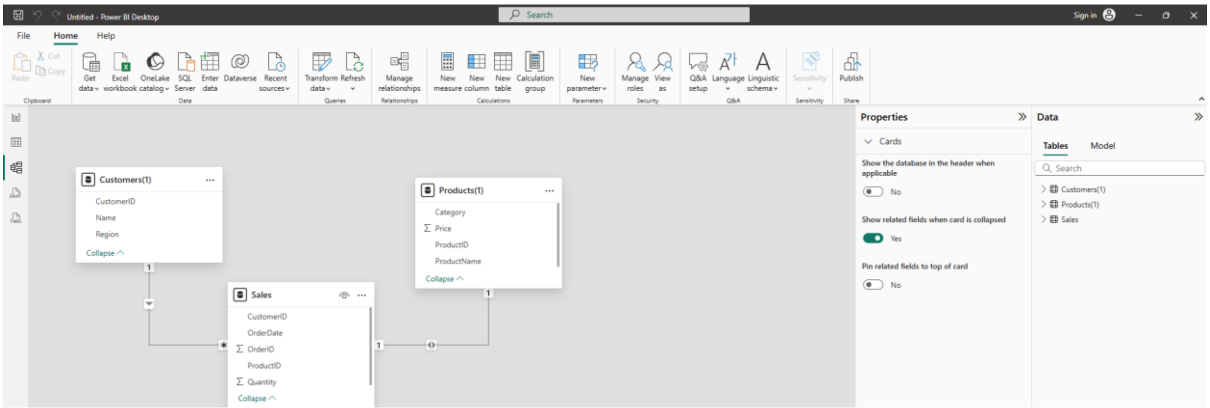
Properties
Cards
Show the database in the header when applicable: No
Show related fields when card is collapsed: Yes
Pin related fields to top of card: No

---

4. What is a "star schema"?

A star schema is a data model with:

- **Fact tables** in the center (e.g., Sales).
- **Dimension tables** surrounding it (e.g., Customers, Products).
  This design simplifies queries and improves performance.



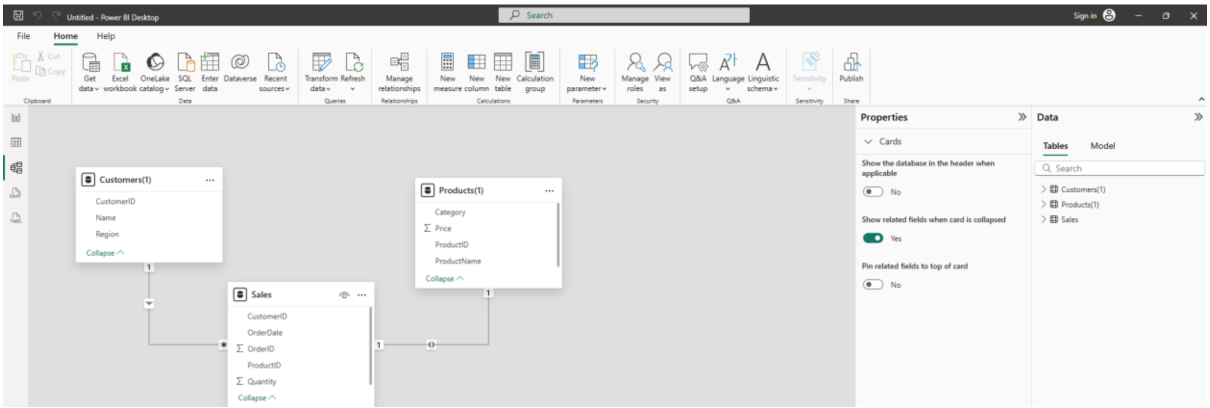
Properties
Cards
Show the database in the header when applicable: No
Show related fields when card is collapsed: Yes
Pin related fields to top of card: No

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

☞ Sales.csv is the fact table since it contains **transactions** (Quantity, OrderDate, CustomerID, ProductID).

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

- Customers.CustomerID → Sales.CustomerID.
- Relationship: **1 (Customers) → Many (Sales)**.



Properties

Cards

Show the database in the header when applicable: No

Show related fields when card is collapsed: Yes

Pin related fields to top of card: No

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

Because it references the unique ProductID in the **Products.csv** table. It ensures sales transactions are tied to valid products.

8. Fix a relationship error where ProductID has mismatched data types.

- Open Power Query → select both columns.
- Convert ProductID in Sales and Products tables to the same type (e.g., **Whole Number**).
- Reapply the relationship.

9. Explain why a star schema improves performance.

- Reduces complexity compared to snowflake models.
- Fewer joins → faster queries.
- Easy to maintain and extend (add more dimensions like Region or Time).

10. Add a new column TotalSales in Sales (Quantity * Price from Products).

DAX formula in Sales table:

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

---

11. Optimize a model with circular relationships—how would you resolve it?

- Avoid bi-directional filters on all relationships.
- Introduce a **bridge table** if two dimension tables connect indirectly through fact tables.
- Remove unnecessary relationships.

---

12. Create a role-playing dimension for OrderDate and ShipDate.

- Duplicate the **Date table** into two tables: OrderDateTable and ShipDateTable.
- Link Sales[OrderDate] to OrderDateTable[Date].
- Link Sales[ShipDate] to ShipDateTable[Date].

---

13. Handle a many-to-many relationship between Customers and Products.

- Create a **bridge table** (e.g., CustomerProduct mapping).
- Use DAX measures for calculations across the relationship.

---

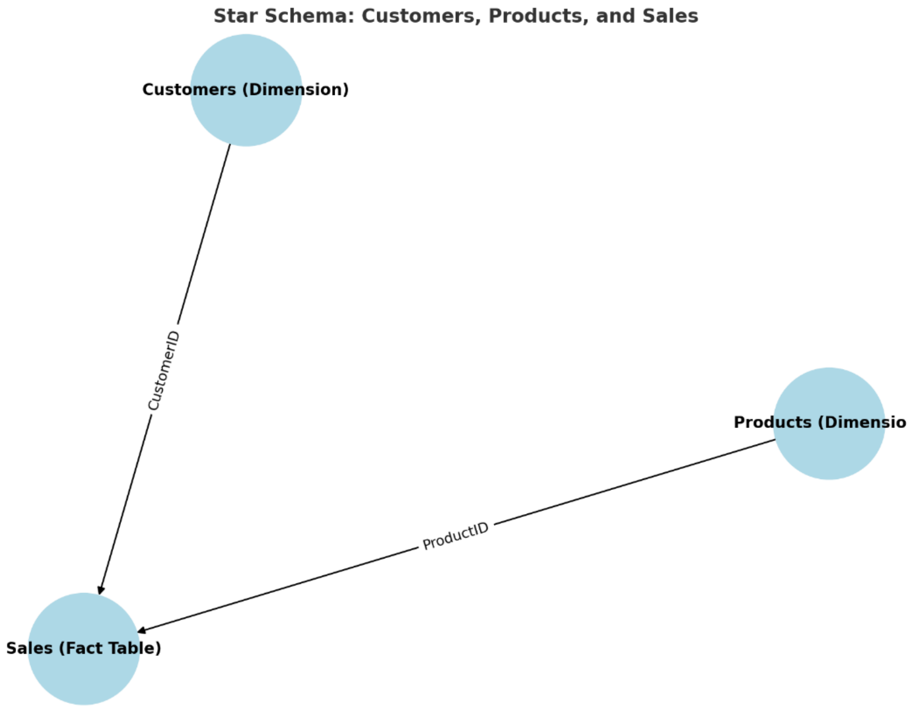
14. Use bidirectional filtering sparingly—when is it appropriate?
When:

- You need filtering to flow both ways (e.g., filtering Customers by Product preferences).
- Data exploration requires dynamic filtering across dimensions.
  But overuse can cause performance issues and ambiguous results.

---

15. Write DAX to enforce referential integrity if a CustomerID is deleted.
Example measure to count only valid customers:

```
ValidSales =
CALCULATE(
   SUM(Sales[Quantity]),
   NOT(ISBLANK(RELATED(Customers[CustomerID])))
)
```


Star Schema: Customers, Products, and Sales
Customers (Dimension)
CustomerID
Products (Dimensio
ProductID
Sales (Fact Table)