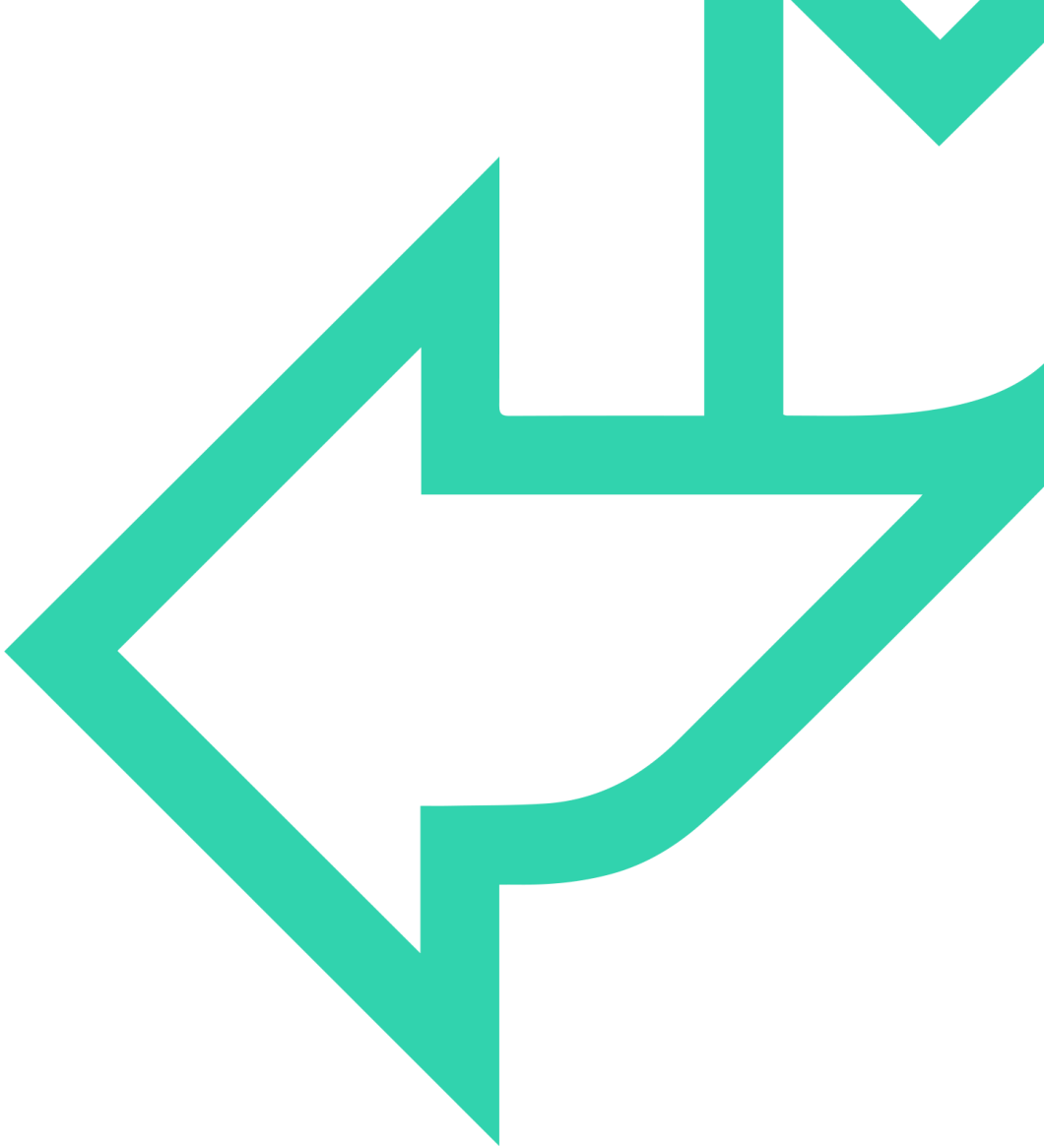
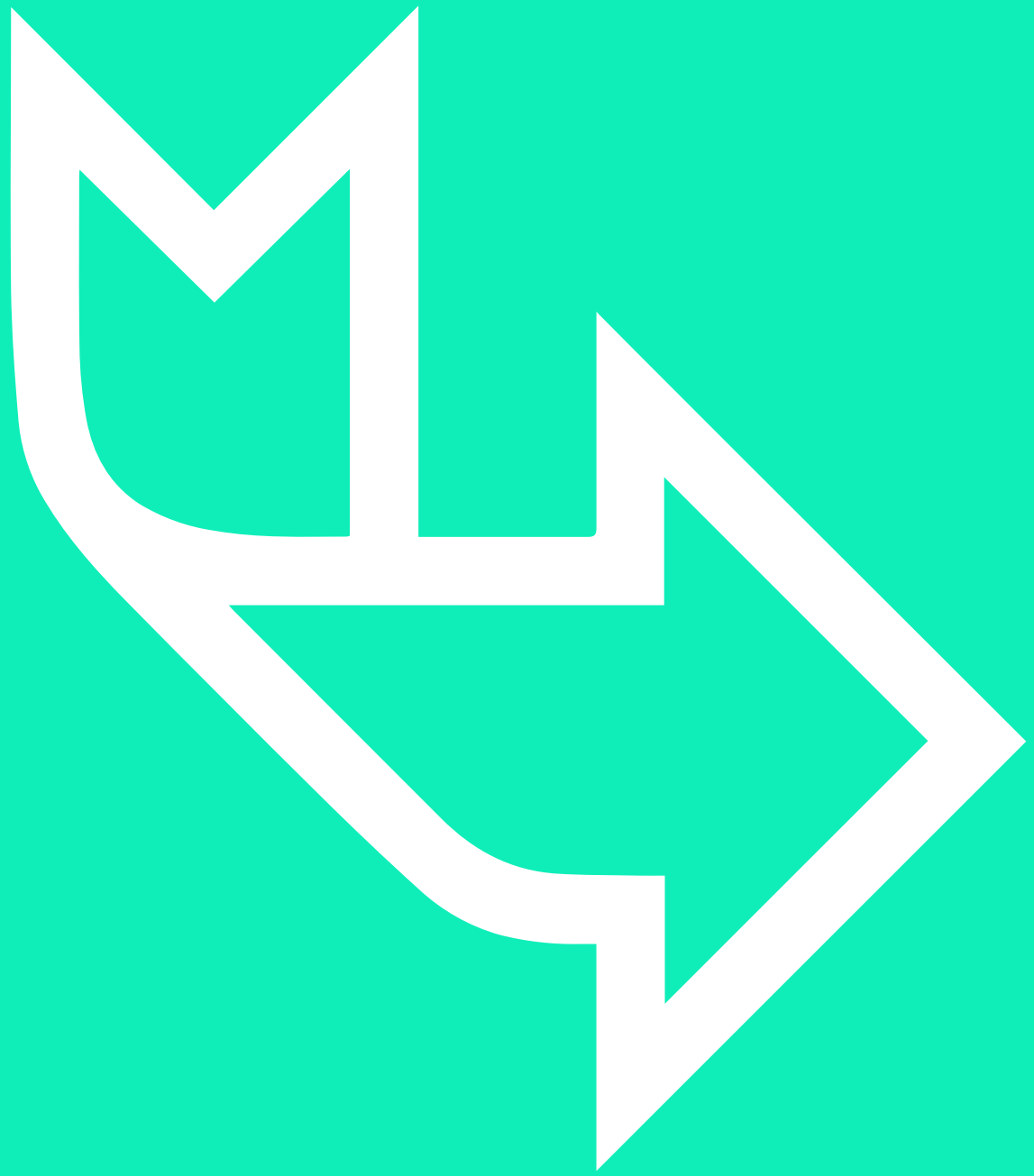
POWERING POTENTIAL



Normalisation

Peter Fellows



Data Normalisation

Objectives

→ Gain an understanding of how to normalise data

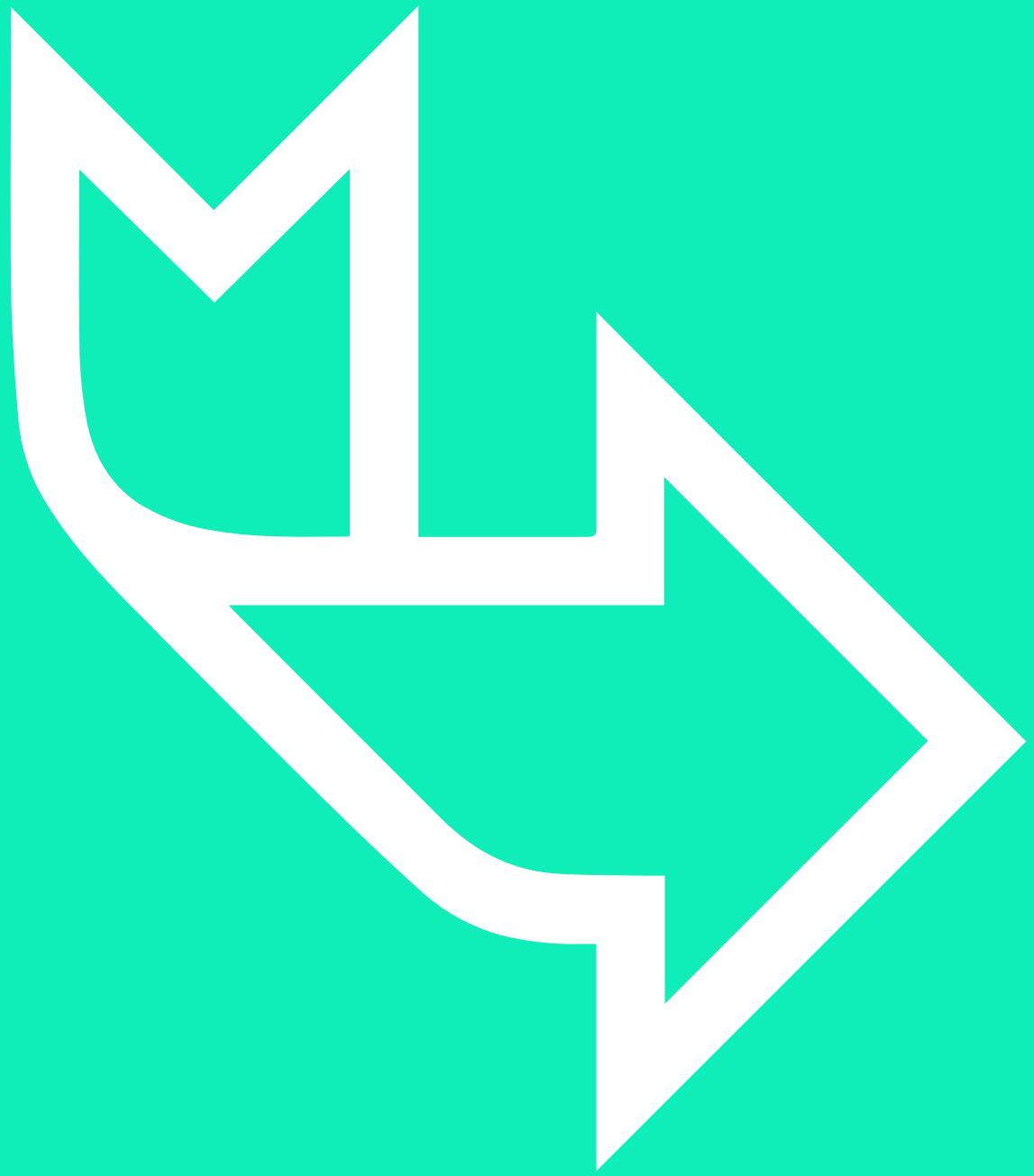
Contents

→ Overview

→ First Normal Form

→ Second Normal Form

→ Third Normal Form



Data Normalisation

The Need for Normalisation

Some Data has a Simple Structure:

→ Customer

→ Name, Address, Telephone, Balance

Some Data has a Complex Structure:

→ But, what if Customer has more than one address?

→ Home phone, Mobile, business phone...

Customers and Orders

QA Games 22 Gamesville Road, Gamesville, GM1 8GG **To**: Games-R-Us, 45 Shoppington Ave. Shoppington Bucks, SH8 2AB	**ORDER FORM** **Order No**:100001 **Date**: 31/07/2011 **Customer No**: 8573

Quantity	Product ID	Description	Unit Price	Amount
2	849	Alpine Ski Instructor	29.99	£59.98
28	773	SIM Accountant	23.46	£656.88
6	382	Recorder Hero – Legends of Folk	19.99	£119.94
			Total Due	£836.80

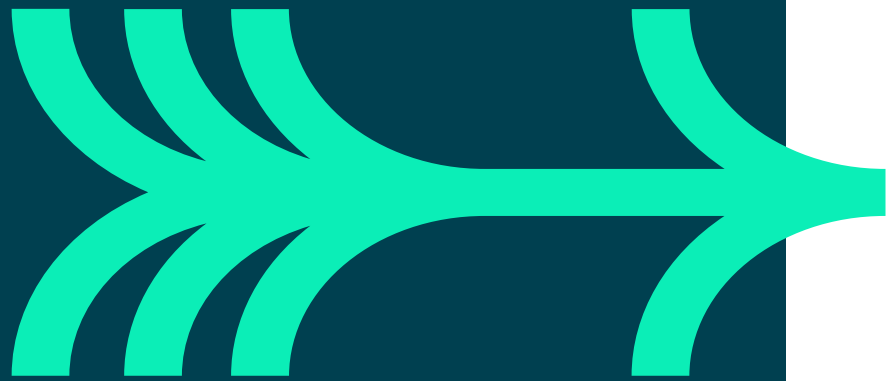
Customers and Orders – Excel Style

How many copies of Alpine Ski Instructor did *Games-R-Us* order in 2011?

	A	B	C	D	E	F	G		J	K	L	M	N
1	Ord. No.	Date	Cust. No.	Cust. Name.	Cust. Address	Cust. Town	Cust. County	Cust. PostCo	Prod. Desc.	Prod. Qty.	Prod. Price	Prod. Amount	Ord. Total
2	100001	31/07/2011	8573	Games-R-Us	45 Shoppingt	Shoppingtc	Bucks.	SH8 2AB	Alpine Ski Instr	2	£29.99	£59.98	£836.80
3									M Accountan	28	£23.46	£656.88	£836.80
4									order Hero	6	£19.99	£119.94	£836.80
5	100002	31/07/2011	5644	Gamerz	304 The Mall	Mallville	Herts.	MV9 1AQ	Alpine Ski Instr	200	£29.99	£5,998.00	£6,772.50
6									Deck Chair Att	50	£15.49	£774.50	£6,772.50
7													

What are total sales of Deck Chair Attendant Deluxe in the county of Buckinghamshire?

DATA NORMALISATION



Normalisation method:

- Approach to database design from ‘bottom up’.
- Formal and mechanical approach.
- Rigorous, unlike Entity-Relationship method.
- Useful for transforming existing files.
- Check existing databases, validate extensions.
- Lossless Decomposition.

DATA NORMALISATION

Benefits:

- Strategy for key and table selection.
- Improved ad-hoc queries and interfaces.
- Eliminates insertion anomalies.
- Eliminates deletion anomalies.
- Reduction in DB modification time.
- Correct representation of physical database.
- Problems show up early in the design.

Moving on from the Spreadsheet

This spreadsheet…

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Ord. No.	Date	Cust. No.	Cust. Name.	Cust. Address	Cust. Town	Cust. County	Cust. PostCode	Prod. ID	Prod. Desc.	Prod. Qty.	Prod. Price	Prod. Amount	Ord. Total
2	100001	31/07/2011	8573	Games-R-Us	45 Shoppingt	Shoppingtc	Bucks.	SH8 2AB	849	Alpine Ski Instr	2	£29.99	£59.98	£836.80
3									773	SIM Accountan	28	£23.46	£656.88	£836.80
4									382	Recorder Hero	6	£19.99	£119.94	£836.80
5	100002	31/07/2011	5644	Gamerz	304 The Mall	Mallville	Herts.	MV9 1AQ	849	Alpine Ski Instr	200	£29.99	£5,998.00	£6,772.50
6									562	Deck Chair Att	50	£15.49	£774.50	£6,772.50
7														

Could be drawn as this table:

- If only it wasn't for the irritating repeating items.

Repeating items

Orders

Order_Number
Order_Date
Customer_Number
Customer_Name
Customer_Address
Customer_Town
Customer_County
Customer_PostCode
Product_ID
Product_Description
Product_Quantity
Product_Price
Product_Total_Amount
Order_Total

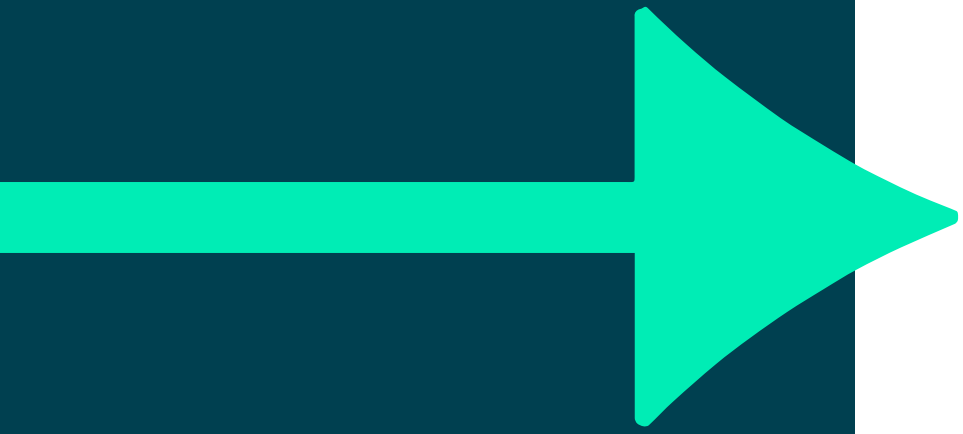
A POSSIBLE SOLUTION?

What if a customer wants to order more than 3 items?

What if a customer wants only 1 item?

Orders	
Order_Number	Product_ID_2
Order_Date	Product_Desc_2
Customer_Number	Product_Quantity_2
Customer_Name	Product_Price_2
Customer_Address	Product_Tot_Amt_2
Customer_Town	Product_ID_3
Customer_County	Product_Desc_3
Customer_PostCode	Product_Quantity_3
Product_ID_1	Product_Price_3
Product_Desc_1	Product_Tot_Amt_3
Product_Quantity_1	Order_Total
Product_Price_1	
Product_Tot_Amt_1	

STEPS TO ACHIEVE 3NF – SUMMARY



1NF - No repeating columns

1. Add extra rows.

2NF - 1NF + all nonkey attributes must be dependent on the whole key

1. Identify the key.
2. Identify functional dependencies (FDs).
3. Identify columns that depend on only part of the key.
4. Create new tables for these columns and partial key.

3NF - 2NF + all nonkey attributes must only be dependent on the key

1. Identify dependencies on nonkey attributes.
2. Create new tables for those columns.

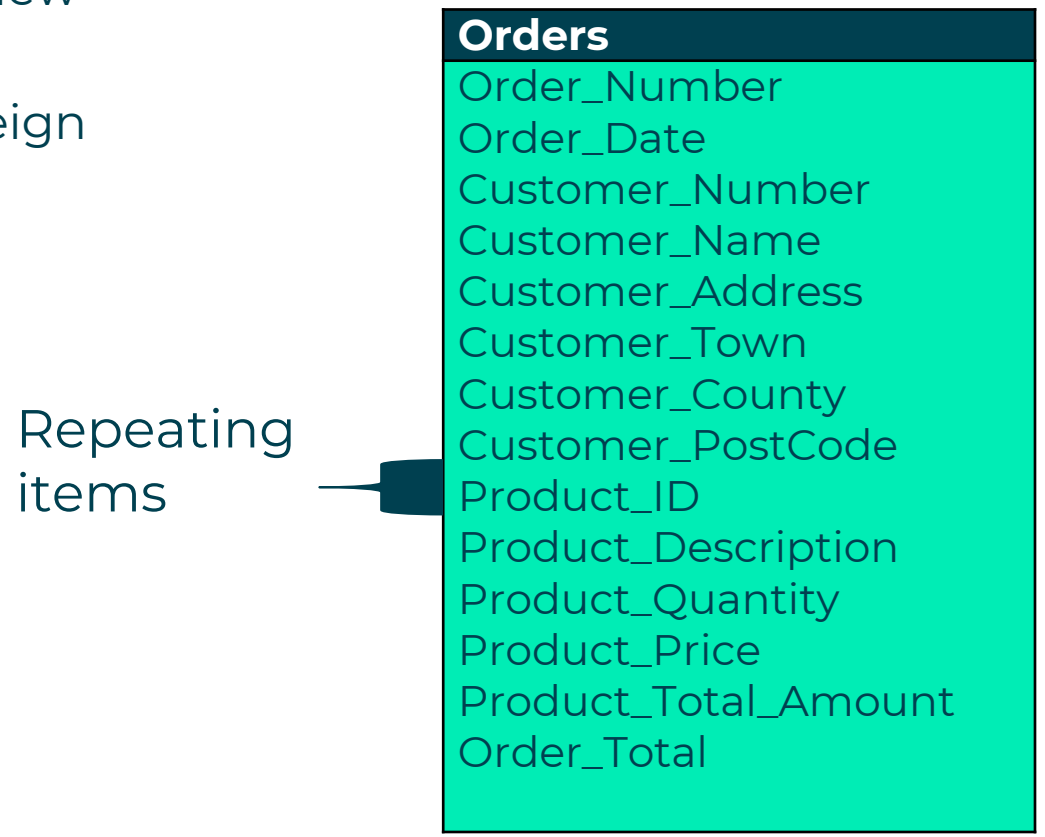
Advanced normal forms

- 4th Normal Form, 5th Normal Form, Boyce/Codd Normal Form.
- Almost never necessary in business systems.

Converting to First Normal Form (1NF)

To convert the entity to First Normal Form (1NF), we must:

- Remove the repeating group of attributes to form a new entity.
- Add the original key to it (this value is known as a foreign key, because it is a primary key in another table).



Converting to First Normal Form (1NF)

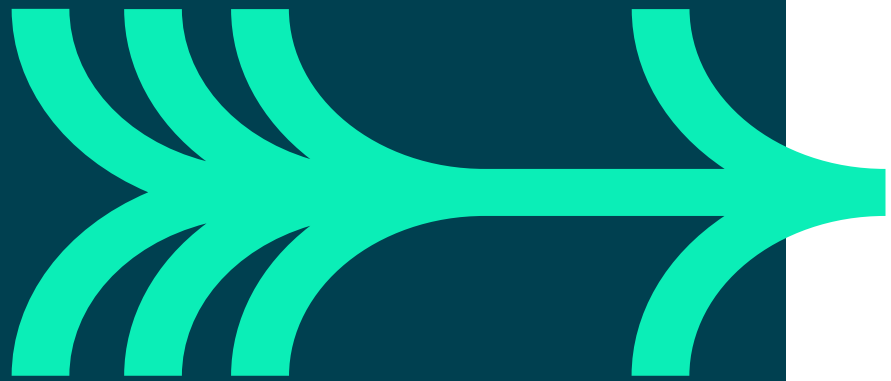
1NF – The Second Part of the Story

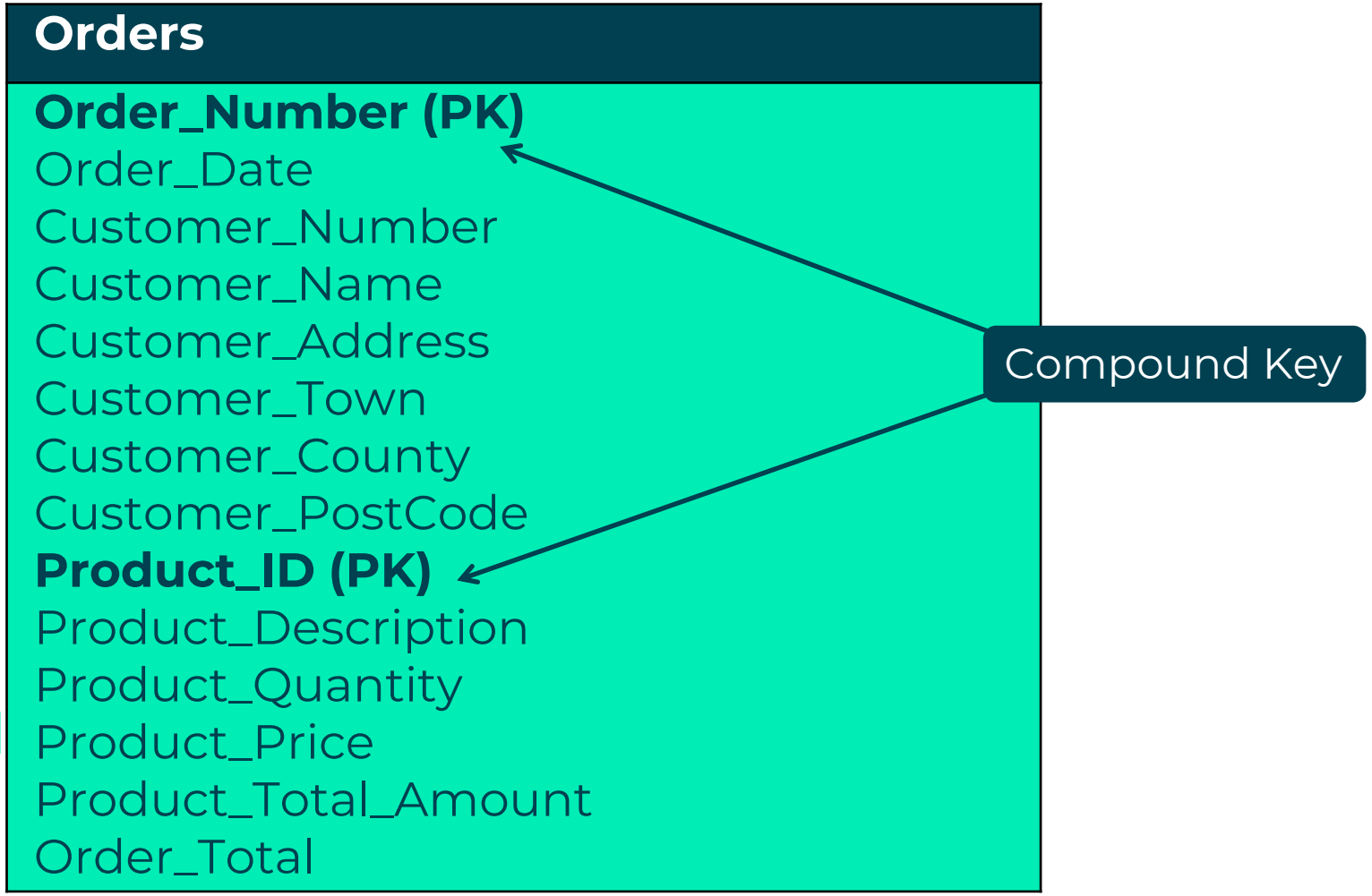
- A row of data cannot contain repeating groups of similar data (atomicity)
- Each row of data must have a unique identifier (or Primary Key)

Orders
Order_Number (PK)
Order_Date
Customer_Number
Customer_Name
Customer_Address
Customer_Town
Customer_County
Customer_PostCode
Order_Total

Order Items
Order_Number (PK)
Product_ID (PK)
Product_Description
Product_Quantity
Product_Price
Product_Total_Amount

PRIMARY KEYS





The Database View

Orders Table:

Results | Messages

	Order_Number	Order_Date	Customer_Number	Customer_Name	Customer_Address	Customer_Town	Customer_County	Custo
1	100001	2011-07-31 00:00:00.000	8573	Games_R_Us	45 Shoppington Ave.	Shoppington	Bucks.	SH8
2	100002	2011-07-31 00:00:00.000	5644	Gamerz	304 The Mall	Mallville	Herts.	MV9

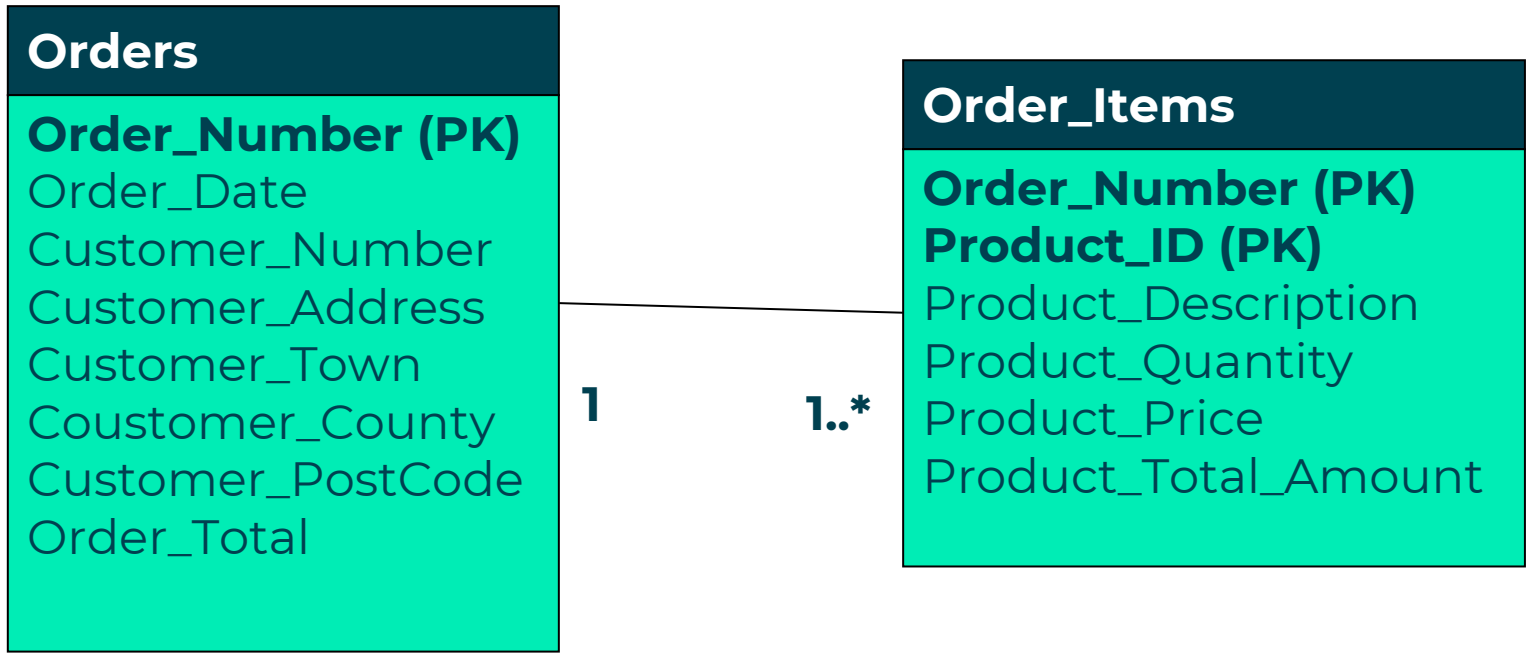
Each Order_Item knows which Order it is a part of.

Order_Items Table:

esults | Messages

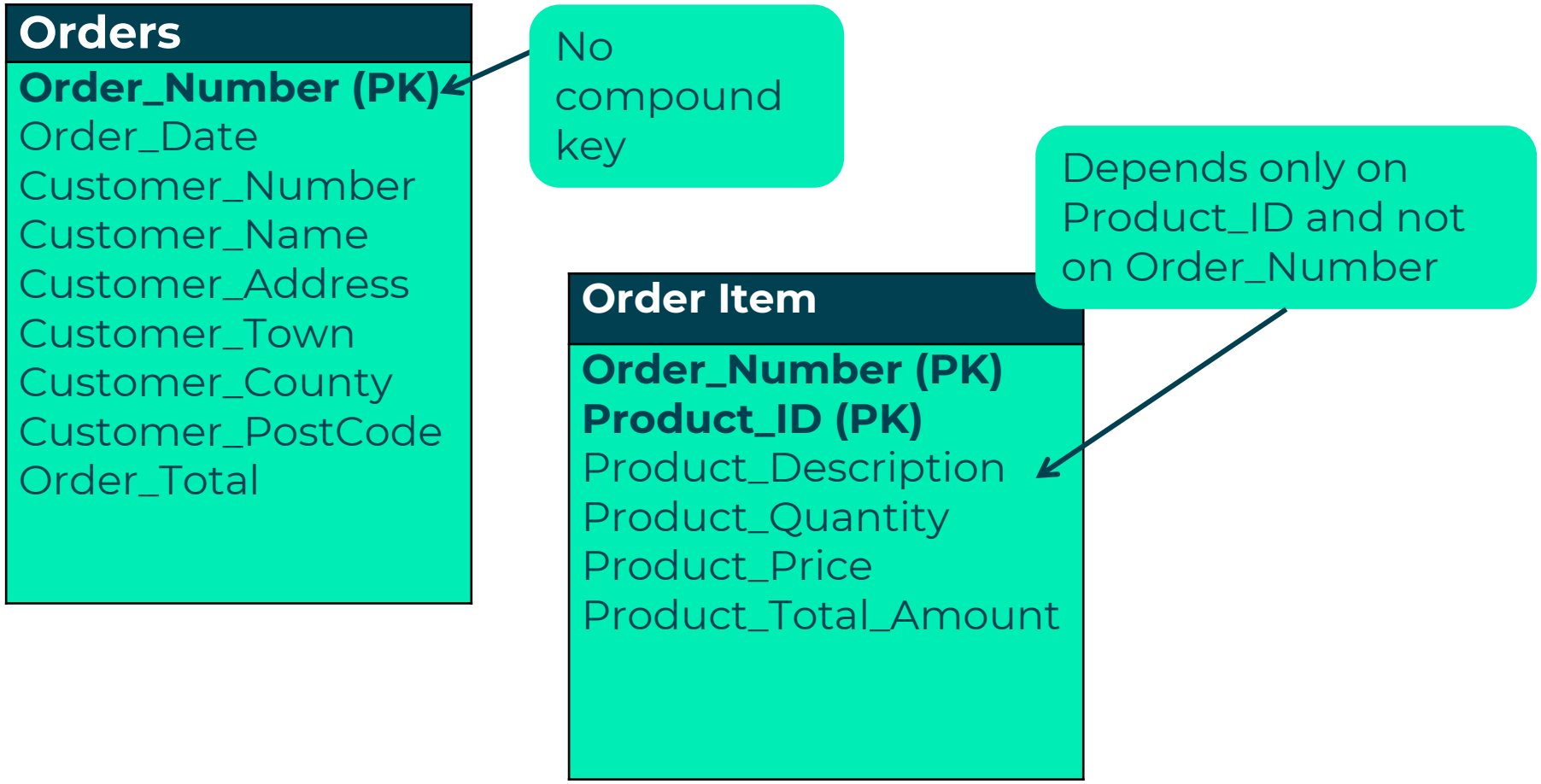
	Order_Number	Product_ID	Product_Description	Product_Quantity	Product_Price
1	100001	849	Alpine Ski Instructor	2	29.99
2	100001	773	SIM Accountant	28	23.46
3	100001	382	Recorder Hero - Legends of Folk	6	19.99
4	100002	849	Alpine Ski Instructor	200	29.99
5	100002	562	Deck Chair Attendant Deluxe	50	15.49

The Entity Relationship Diagram

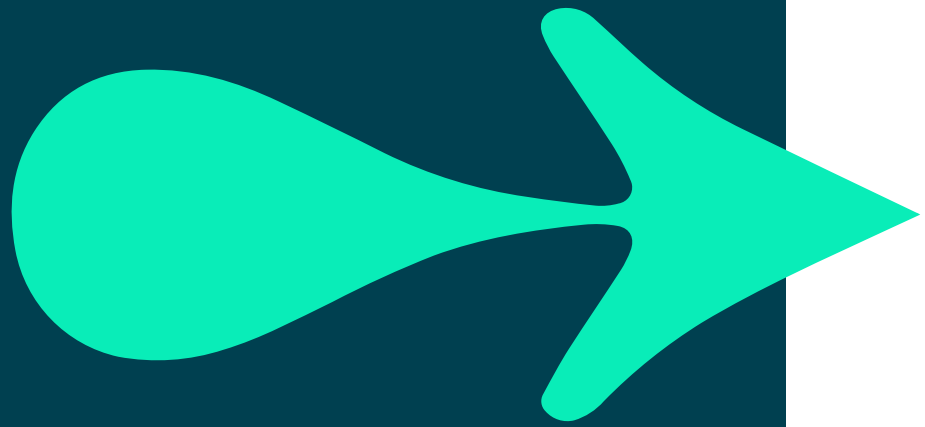


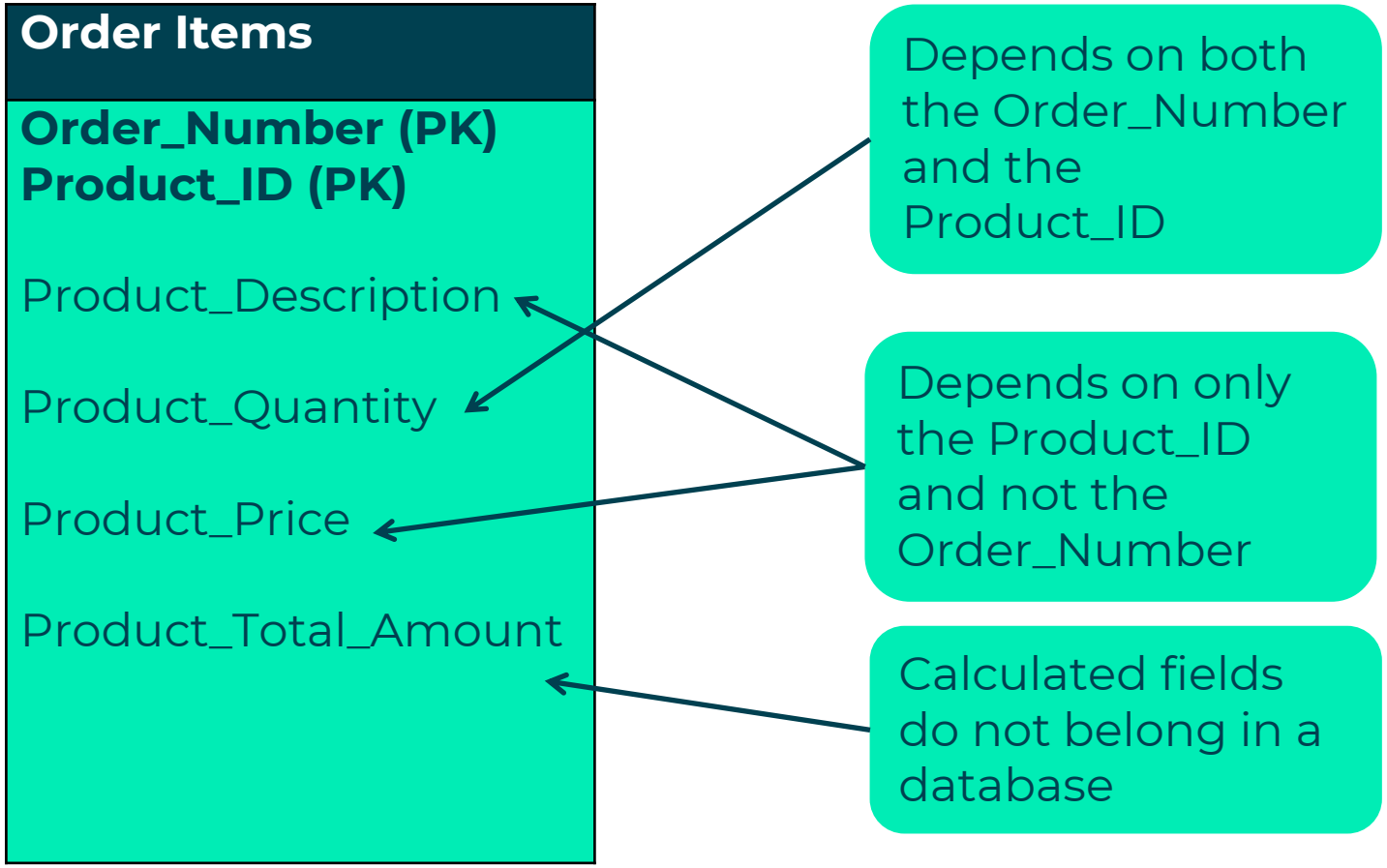
2nd Normal Form

- Converting to Second Normal Form (2NF)
- No Partial Dependencies on a Compound Key



CONTINUING THE HUNT





CONTINUING THE HUNT

The story so far...

Order Items
Product_ID (PK) **Product_Description X (Depends only on Product_ID)** **Product_Quantity ✓** **Product_Price X (Depends only on Product_ID)** **~~Product_Total_Amount~~**

Continuing the Hunt

- Passing the 2NF Test

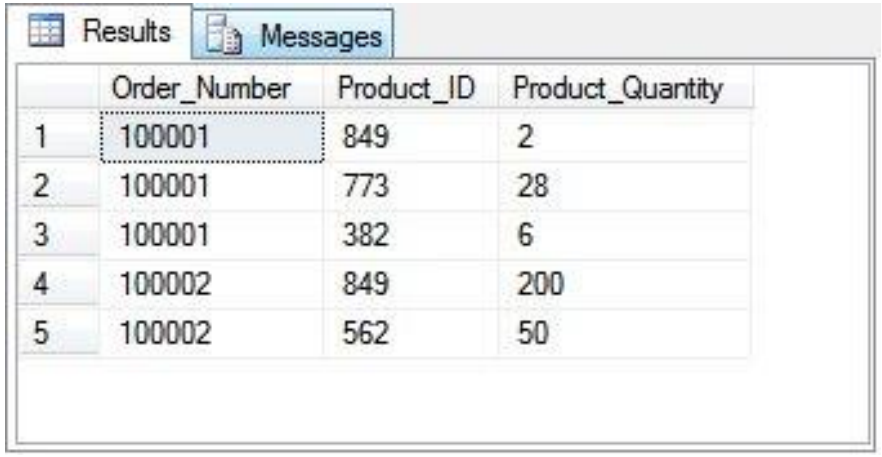
Orders
Order_Number (PK)
Order_Date
Customer_Number
Customer_Name
Customer_Address
Customer_Town
Customer_County
Customer_PostCode
Order_Total

Order_Items
Order_Number (PK)
Product_ID (PK)
Product_Quantity

Products
Product_ID (PK)
Product_Description
Product_Price

DATABASE TABLES

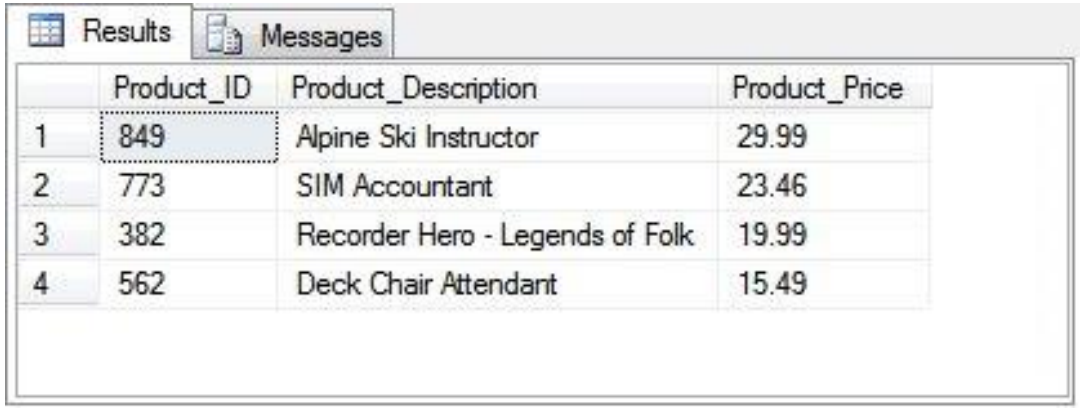
Order_Items:



Results | Messages

	Order_Number	Product_ID	Product_Quantity
1	100001	849	2
2	100001	773	28
3	100001	382	6
4	100002	849	200
5	100002	562	50

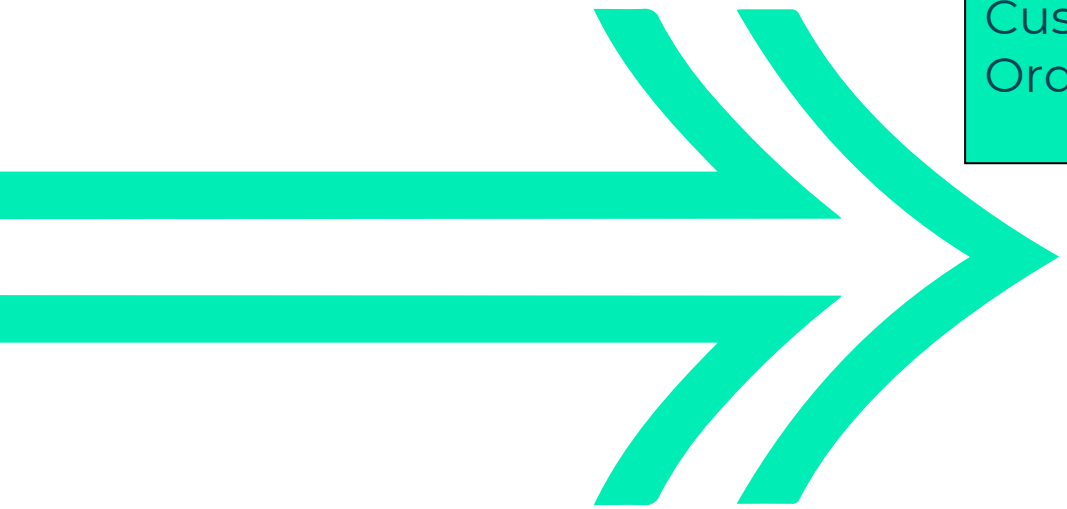
Products:

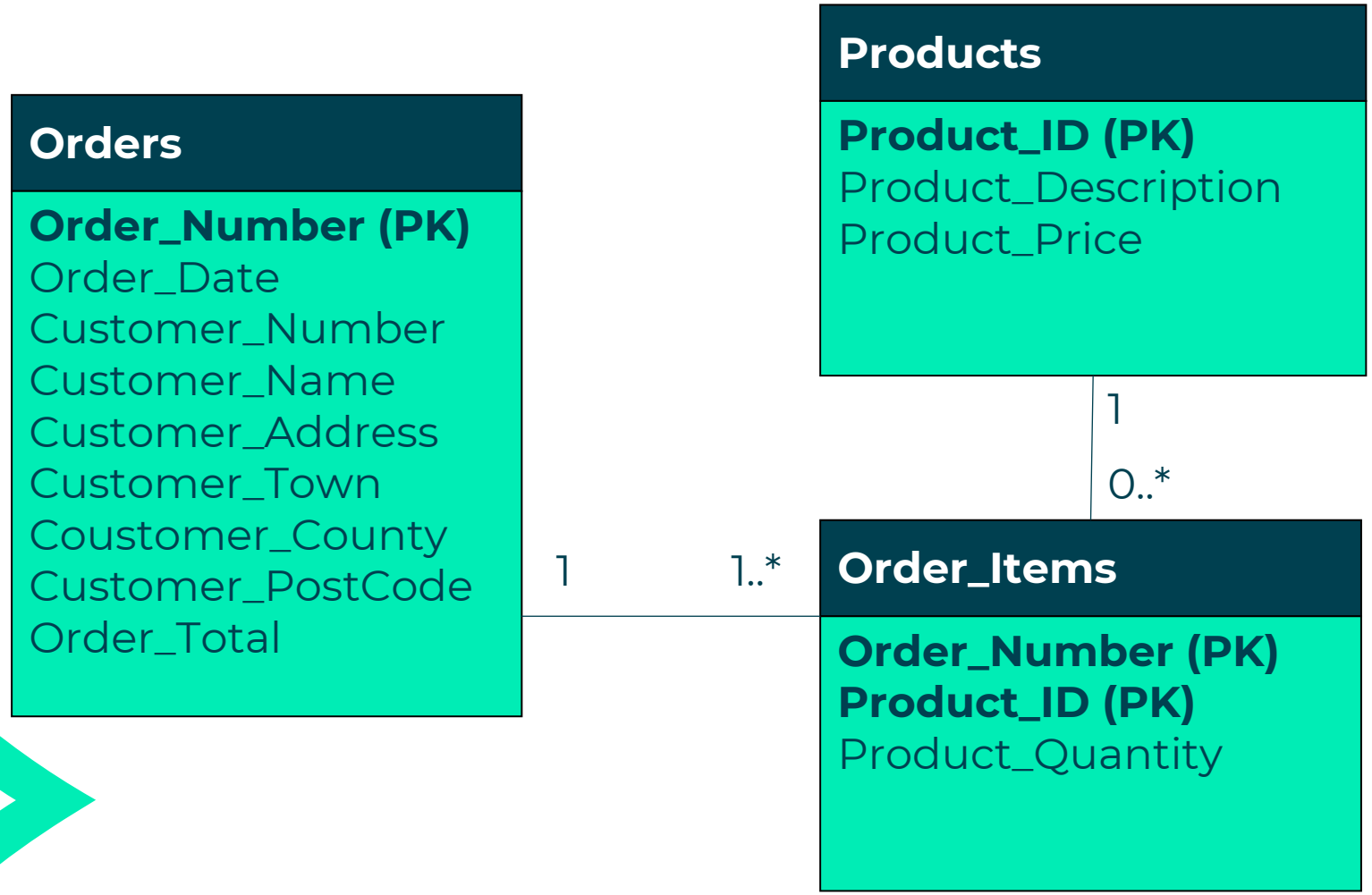


Results | Messages

	Product_ID	Product_Description	Product_Price
1	849	Alpine Ski Instructor	29.99
2	773	SIM Accountant	23.46
3	382	Recorder Hero - Legends of Folk	19.99
4	562	Deck Chair Attendant	15.49

Entity Relationship Diagram





A Problem

What happens if the price of a product changes after an order has been placed?

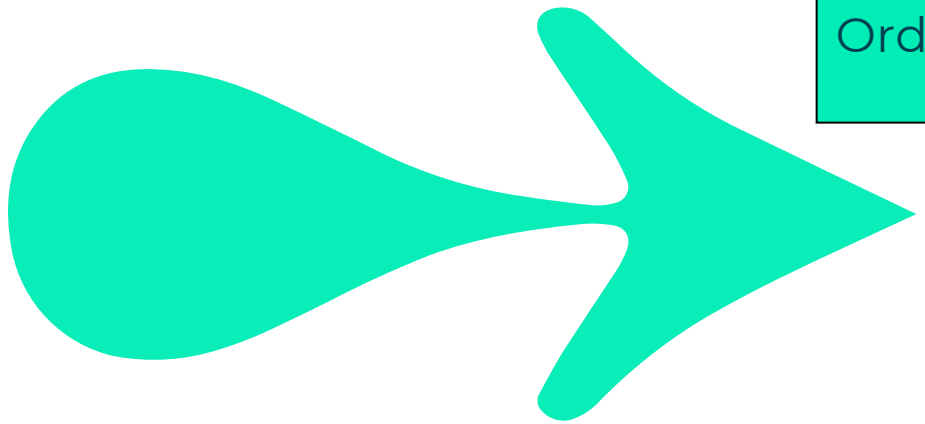
- What should the customer pay?
- The old price or the new?

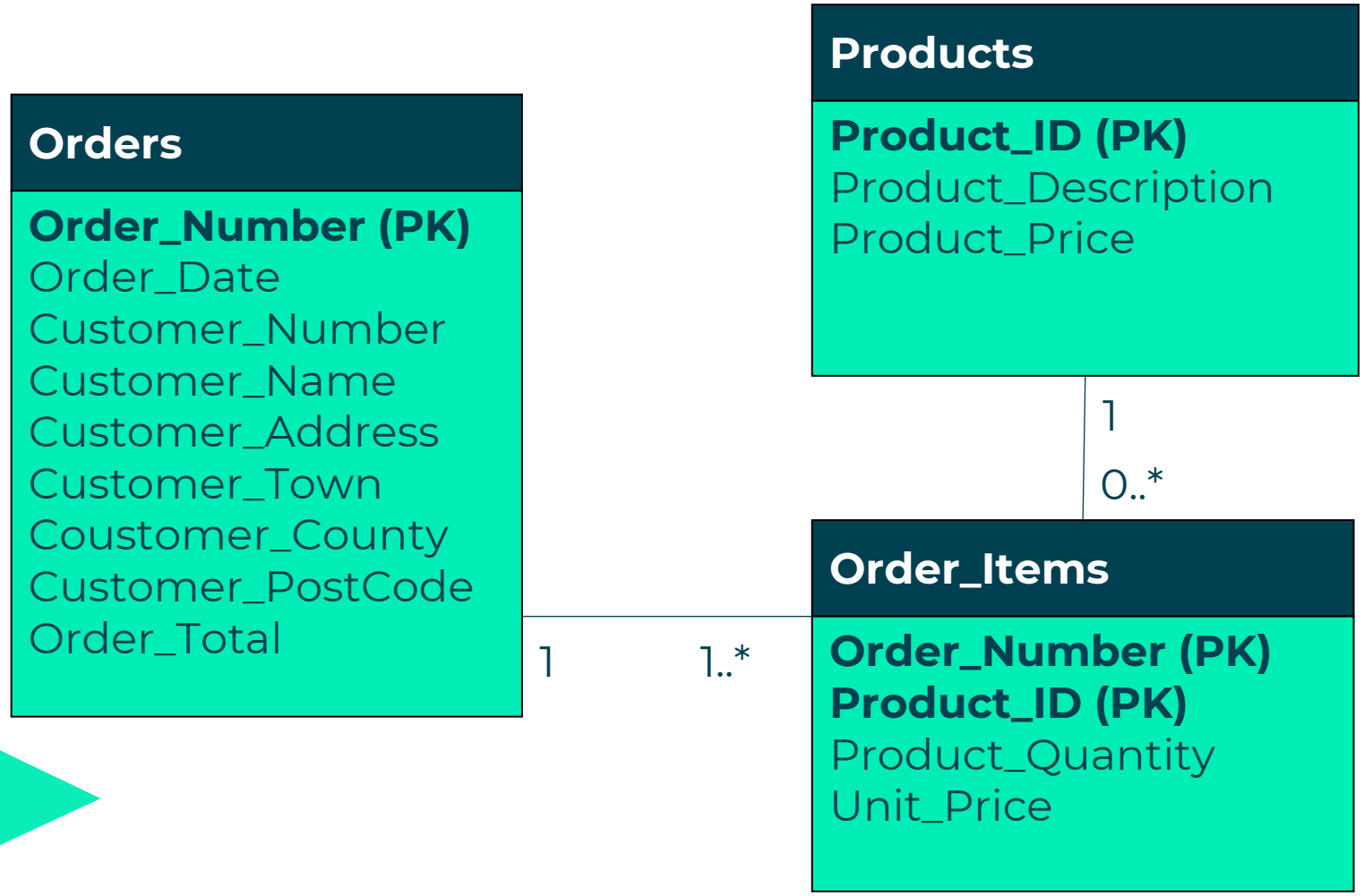
Customer purchases game for £50.

Product Price increases to £55

Customer returns game and demands a refund.

A solution




Orders
Order_Number (PK)
Order_Date
Customer_Number
Customer_Name
Customer_Address
Customer_Town
Coustomer_County
Customer_PostCode
Order_Total
Products
Product_ID (PK)
Product_Description
Product_Price
1
0..*
1
1..*
Order_Items
Order_Number (PK)
Product_ID (PK)
Product_Quantity
Unit_Price

CONVERTING TO THIRD NORMAL FORM (3NF)

To Convert to Third Normal Form:

- Identify any dependencies between non-key attributes within each table.
- Remove them to form a new table.
- Promote one of the attributes to be the key of the new table.
- This becomes a Foreign Key link in the original table (shown by an *).

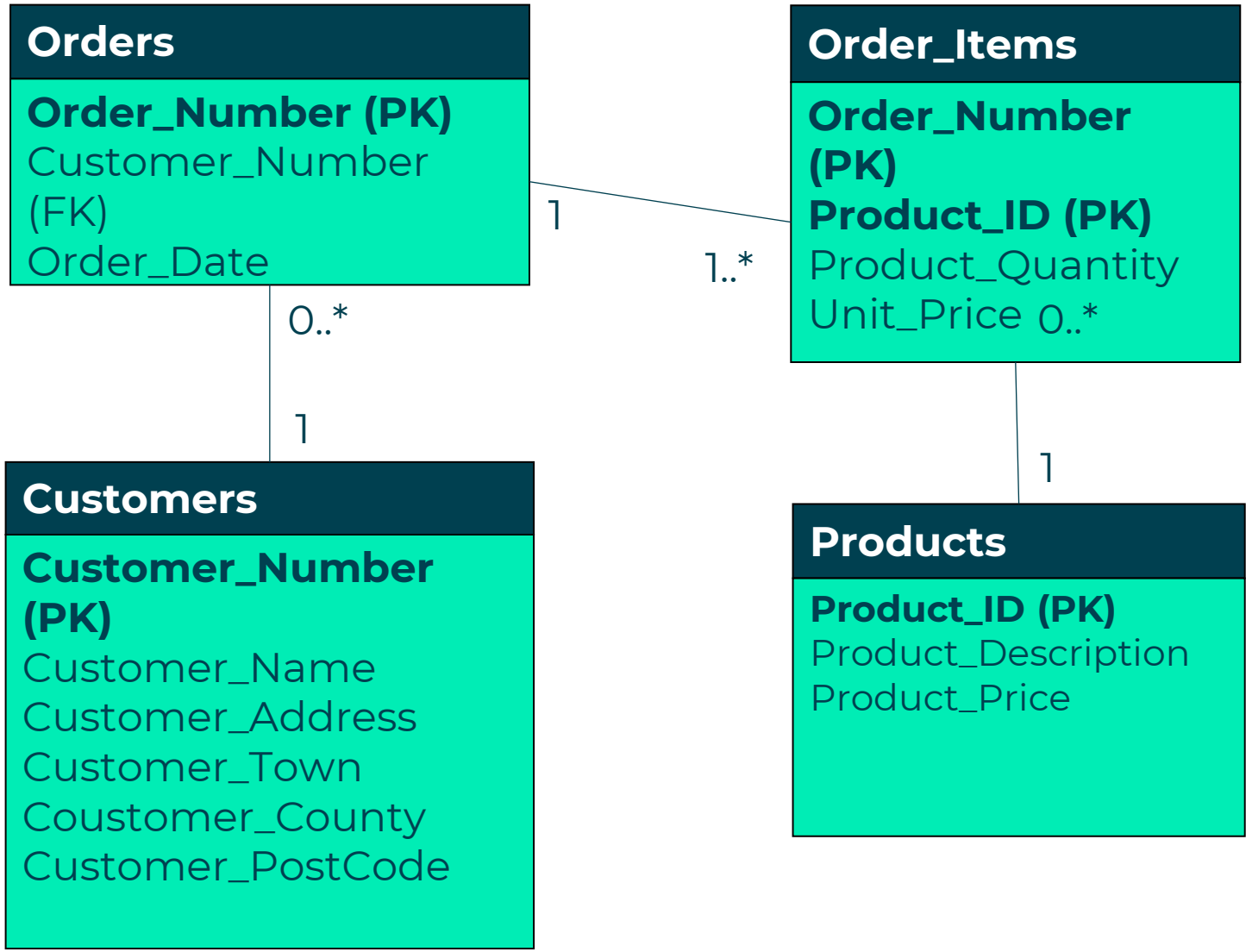
THE CUSTOMER PROBLEM

Orders
Order_Number (PK)
Order_Date
Customer_Number
Customer_Name
Customer_Addres
Customer_Town
Coustomer_County
Customer_PostCode
Order_Total

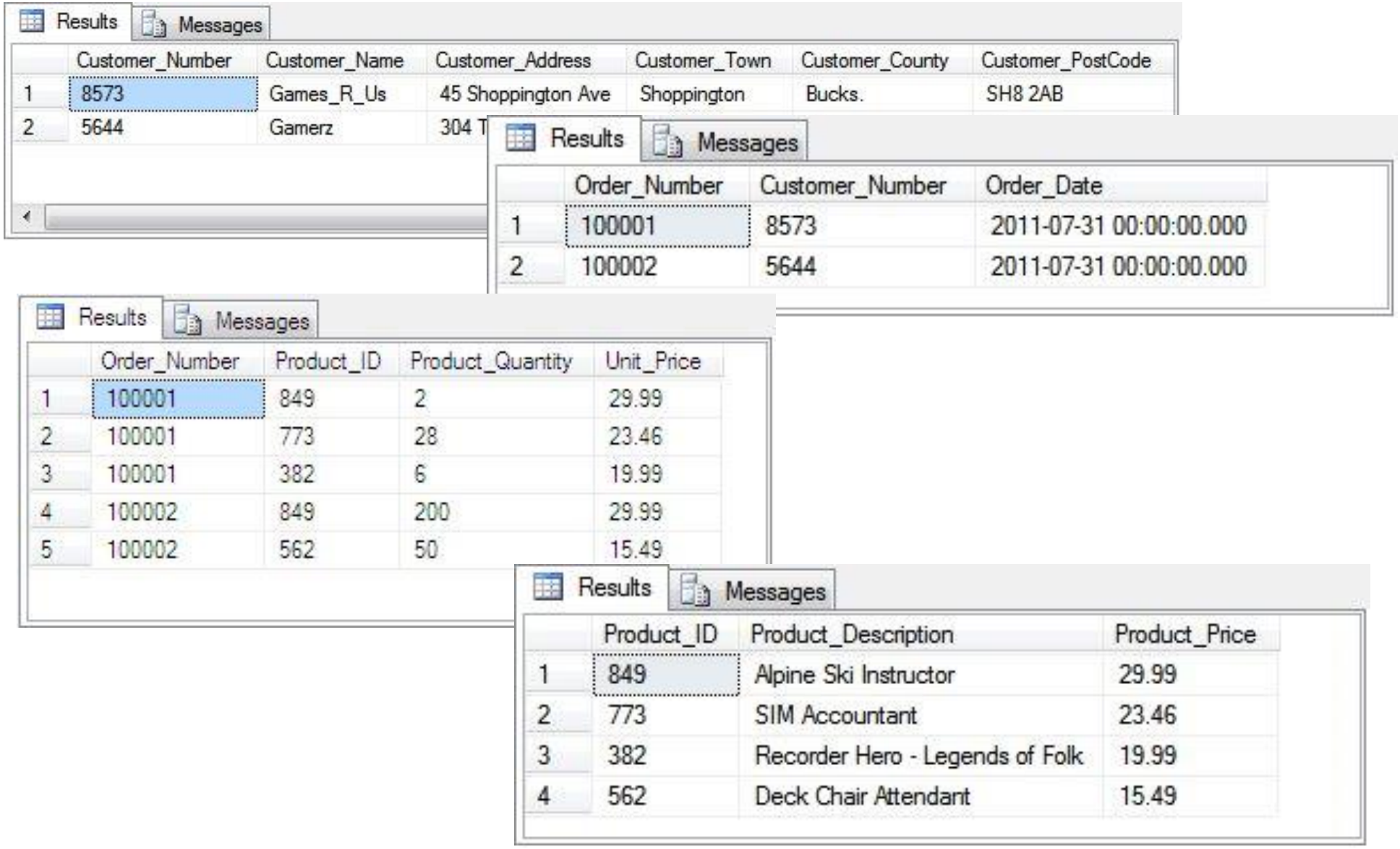
Depends on the Customer_Number and not the Order_Number

Calculated fields do not belong in a database

Entity Relationship Diagram



The Database View



Results | Messages

	Customer_Number	Customer_Name	Customer_Address	Customer_Town	Customer_County	Customer_PostCode
1	8573	Games_R_Us	45 Shoppington Ave	Shoppington	Bucks.	SH8 2AB
2	5644	Gamerz	304 T			

Results | Messages

	Order_Number	Customer_Number	Order_Date
1	100001	8573	2011-07-31 00:00:00.000
2	100002	5644	2011-07-31 00:00:00.000

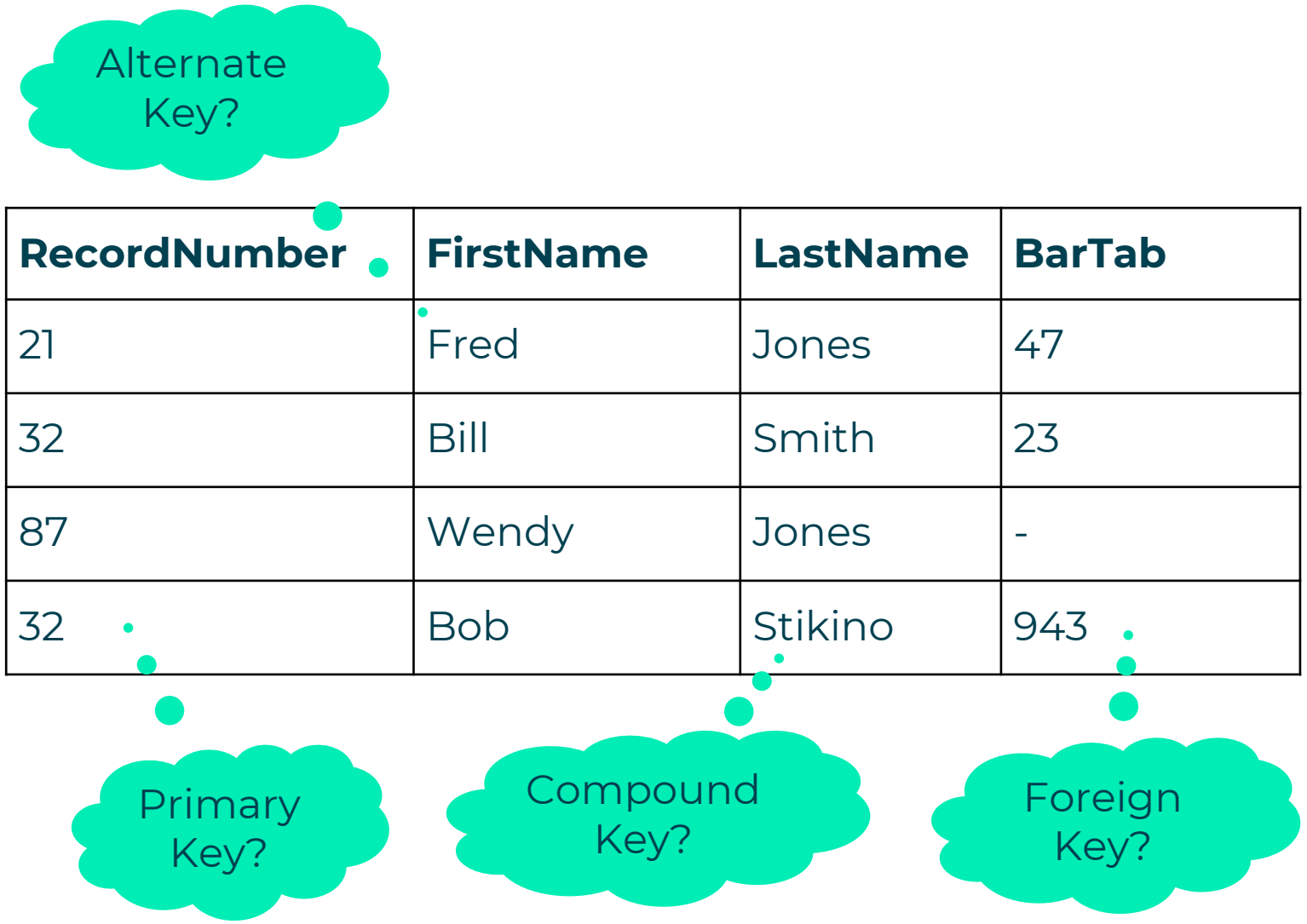
Results | Messages

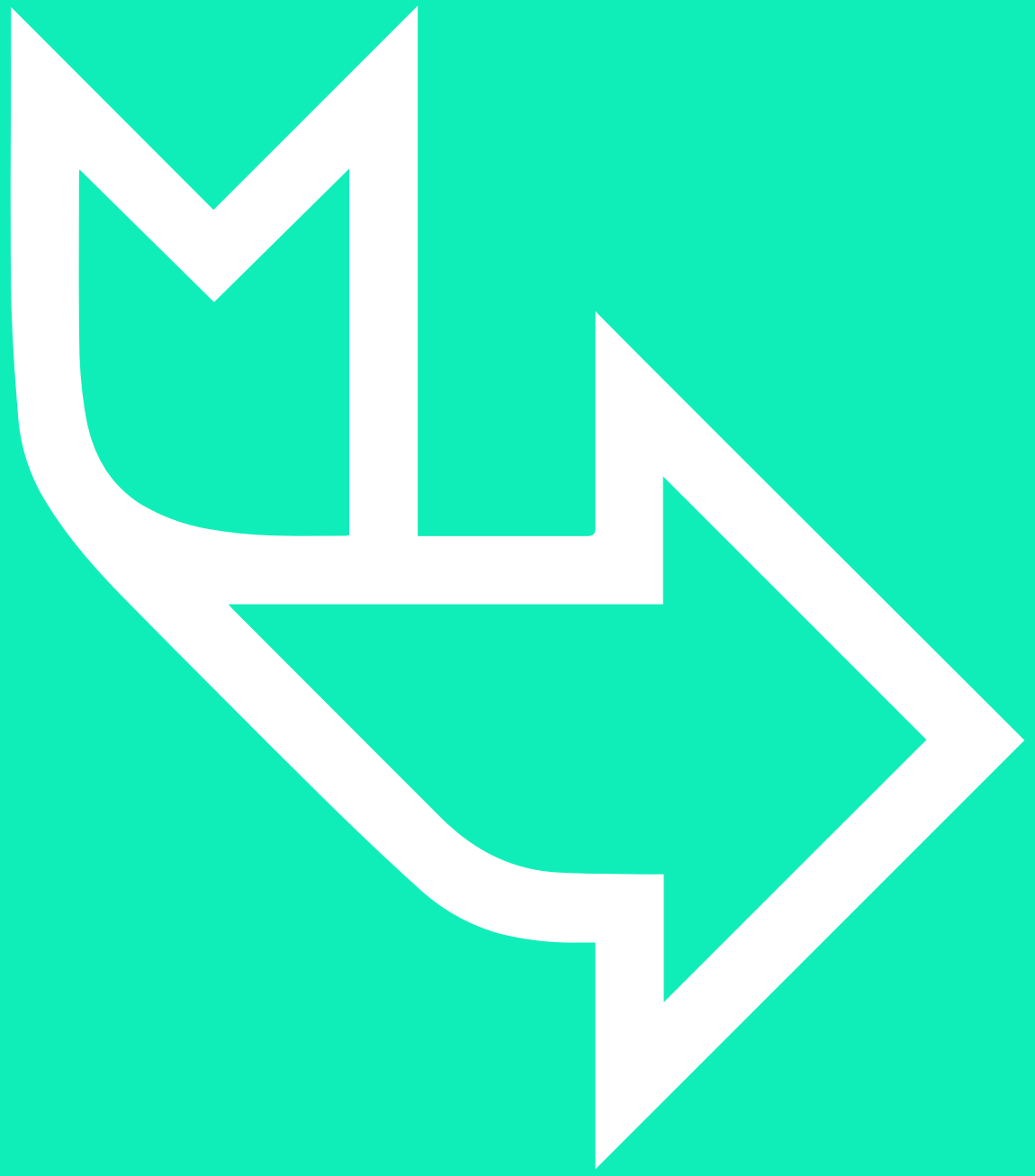
	Order_Number	Product_ID	Product_Quantity	Unit_Price
1	100001	849	2	29.99
2	100001	773	28	23.46
3	100001	382	6	19.99
4	100002	849	200	29.99
5	100002	562	50	15.49

Results | Messages

	Product_ID	Product_Description	Product_Price
1	849	Alpine Ski Instructor	29.99
2	773	SIM Accountant	23.46
3	382	Recorder Hero - Legends of Folk	19.99
4	562	Deck Chair Attendant	15.49

Keys - Revisited





Exercises

Games R Us

The Acme Practice Doctor's Surgery

Exercise

The Acme Practice Doctor's Surgery

The Acme Practice Prescription Record			
Patient Number:	A2059	**Patient Name**:	I.M. Sick
Address:	7 NotFeelingTooWell Drive, Coughsville, Beds, CO2 1AF	**Date of Birth**:	02/06/1992

Date	**Time**	**Prescription Item Code**	**Prescription Item Name**
23/10/2010	9:00am	FJ01	Flu Jab
02/01/2011	9:15am	PK03	Pain Killers
13/03/2011	4:30pm	CM02	Cough Mixture
15/03/2011	4:45pm	M09	Morphine