

Keys for Relational Data

INTRODUCTION TO DATA MODELING IN SNOWFLAKE



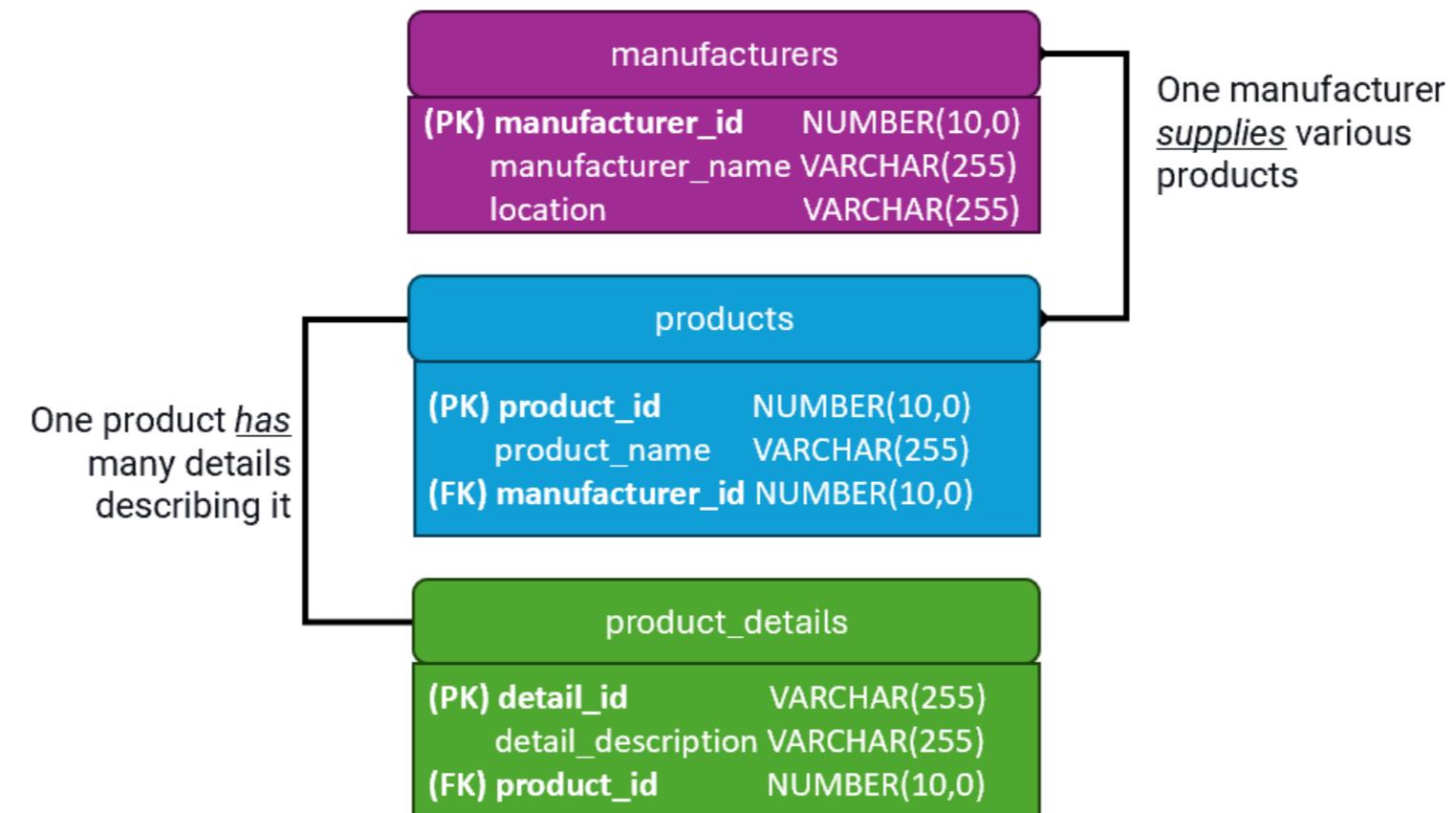
Nuno Rocha
Director of Engineering

Introduction to relational data

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Introduction to relational data (1)

- **Relational data:** Structured data organized into individual entities and keys that establish relationships between them



The power of keys in relational data

- **Data structure & integrity:** Keys organize data into tables and maintain accuracy
- **Real-world connections:** Keys facilitate meaningful links and reflect complex interactions
- **Efficient access:** Keys streamline data retrieval and querying
- **Scalability:** Keys ensure data adapts to growth with integrity

Recap of primary and foreign keys

- Primary Key: Unique identifier of each record in an entity

```
CREATE OR REPLACE TABLE products (
    id NUMBER(10,0) PRIMARY KEY,
    name VARCHAR(255)
);
```

Recap of primary and foreign keys (1)

- **Primary Key:** Unique identifier of each record in an entity
- **Foreign Key:** Links one entity to another

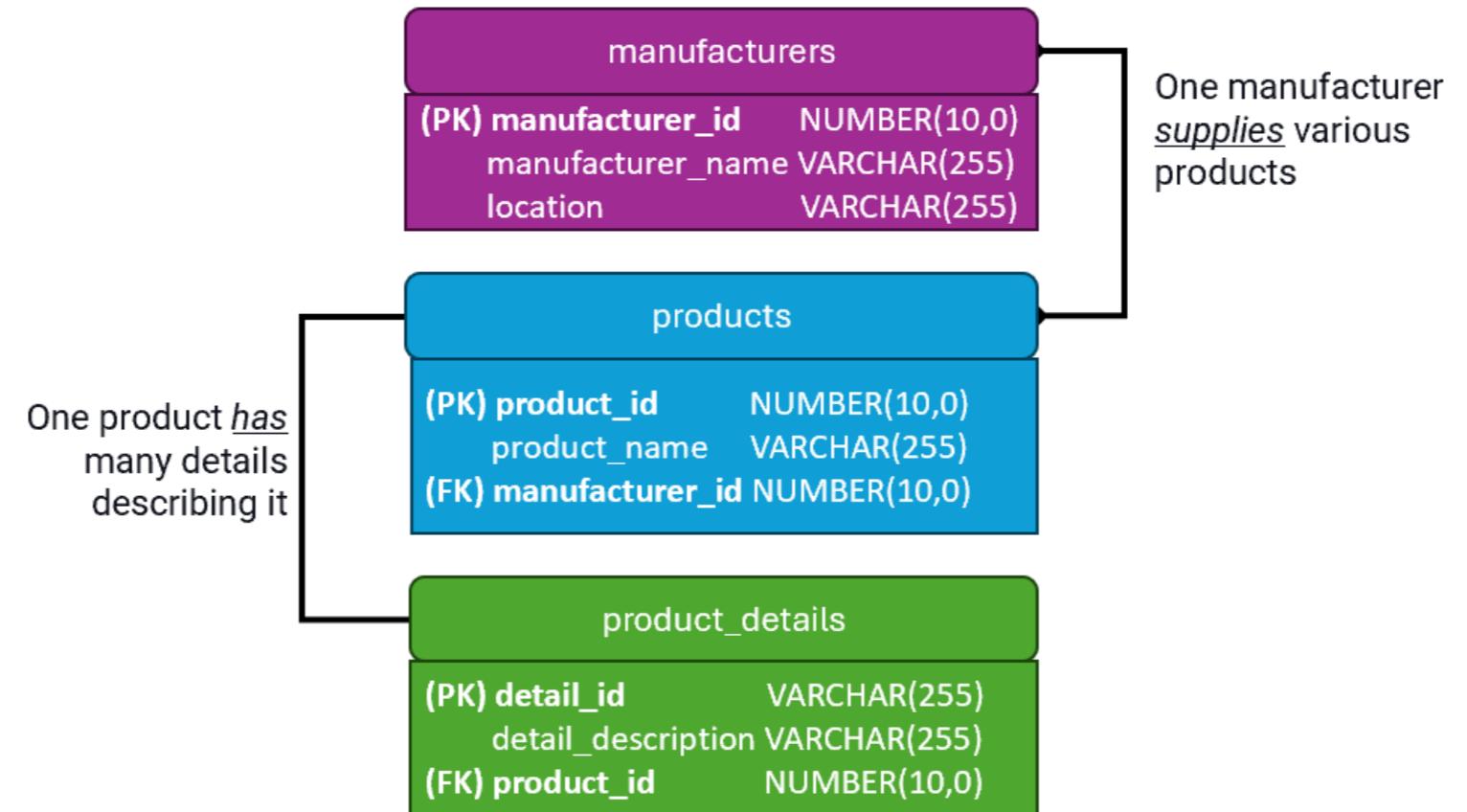
```
CREATE OR REPLACE TABLE products (
    product_id NUMBER(10,0) PRIMARY KEY,
    product_name VARCHAR(255)
);
```

Recap of primary and foreign keys (2)

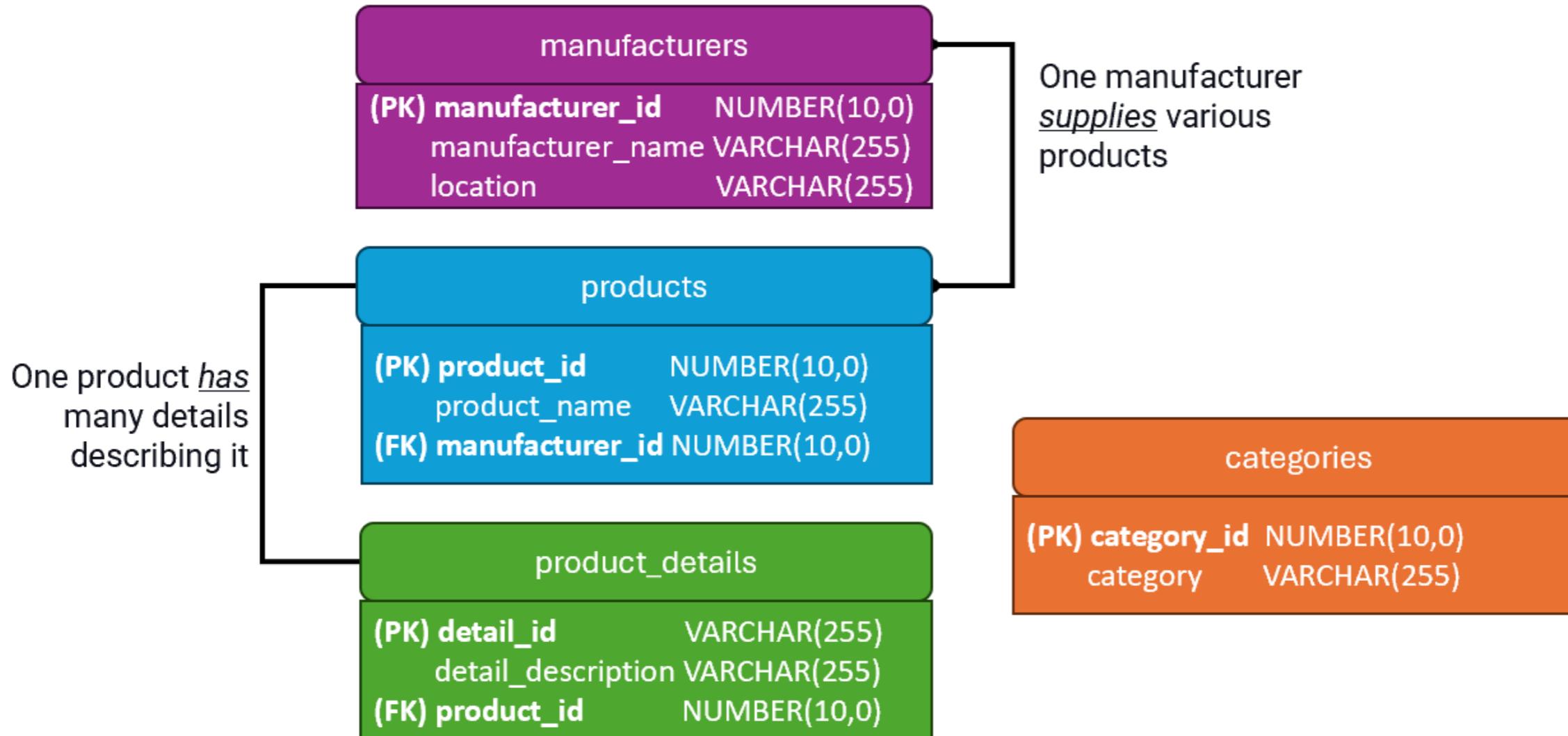
- **Primary Key:** Unique identifier of each record in an entity
- **Foreign Key:** Links one entity to another

```
CREATE OR REPLACE TABLE products (
    product_id NUMBER(10,0) PRIMARY KEY,
    product_name VARCHAR(255),
    manufacturer_id NUMBER(10,0),
    FOREIGN KEY (manufacturer_id) REFERENCES manufacturers(manufacturer_id)
);
```

Relationships in the data model

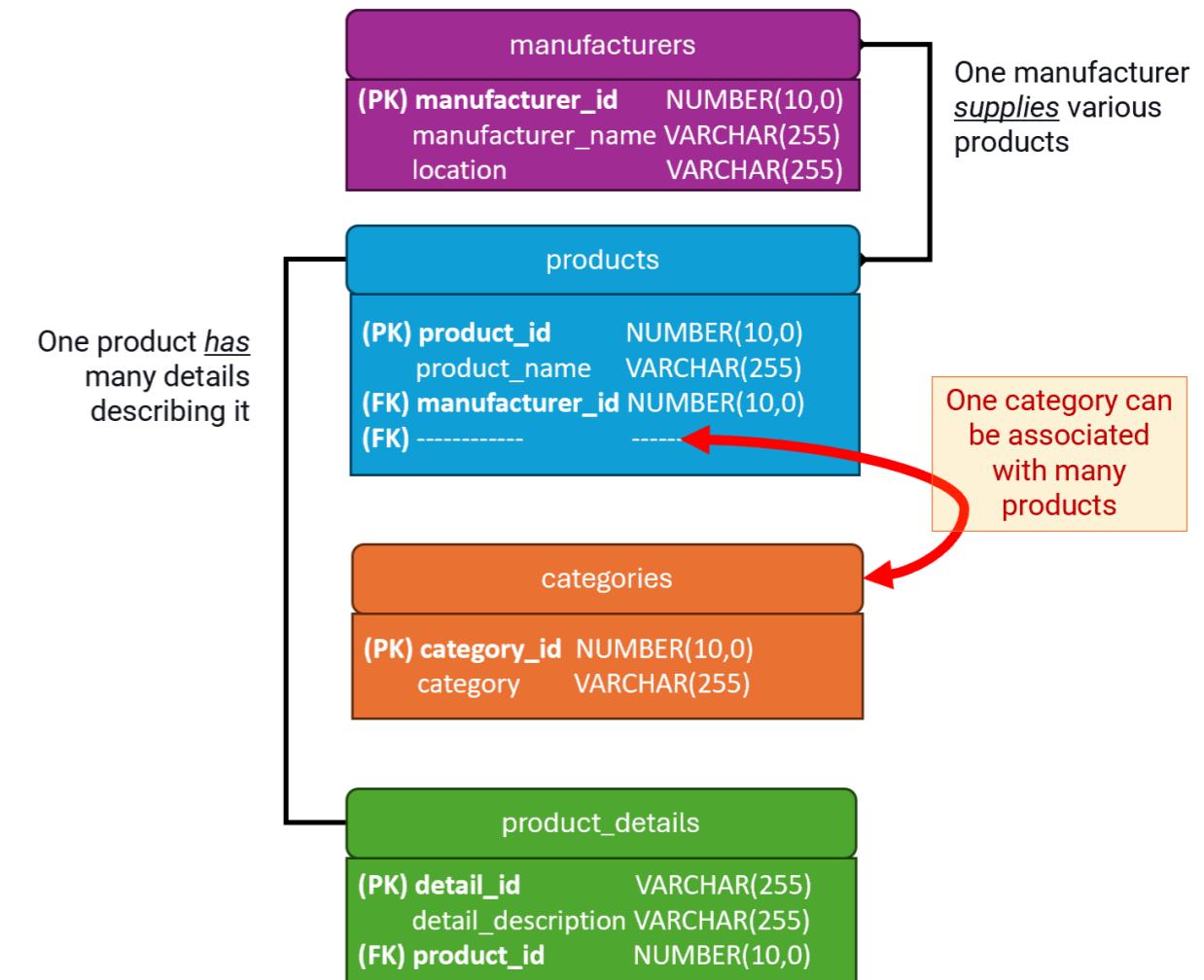


Relationships in the data model (1)



Adjusting the data model

```
CREATE OR REPLACE TABLE categories (
    category_id NUMBER(10,0) PRIMARY KEY,
    category VARCHAR(255)
);
```



Adjusting the data model (1)

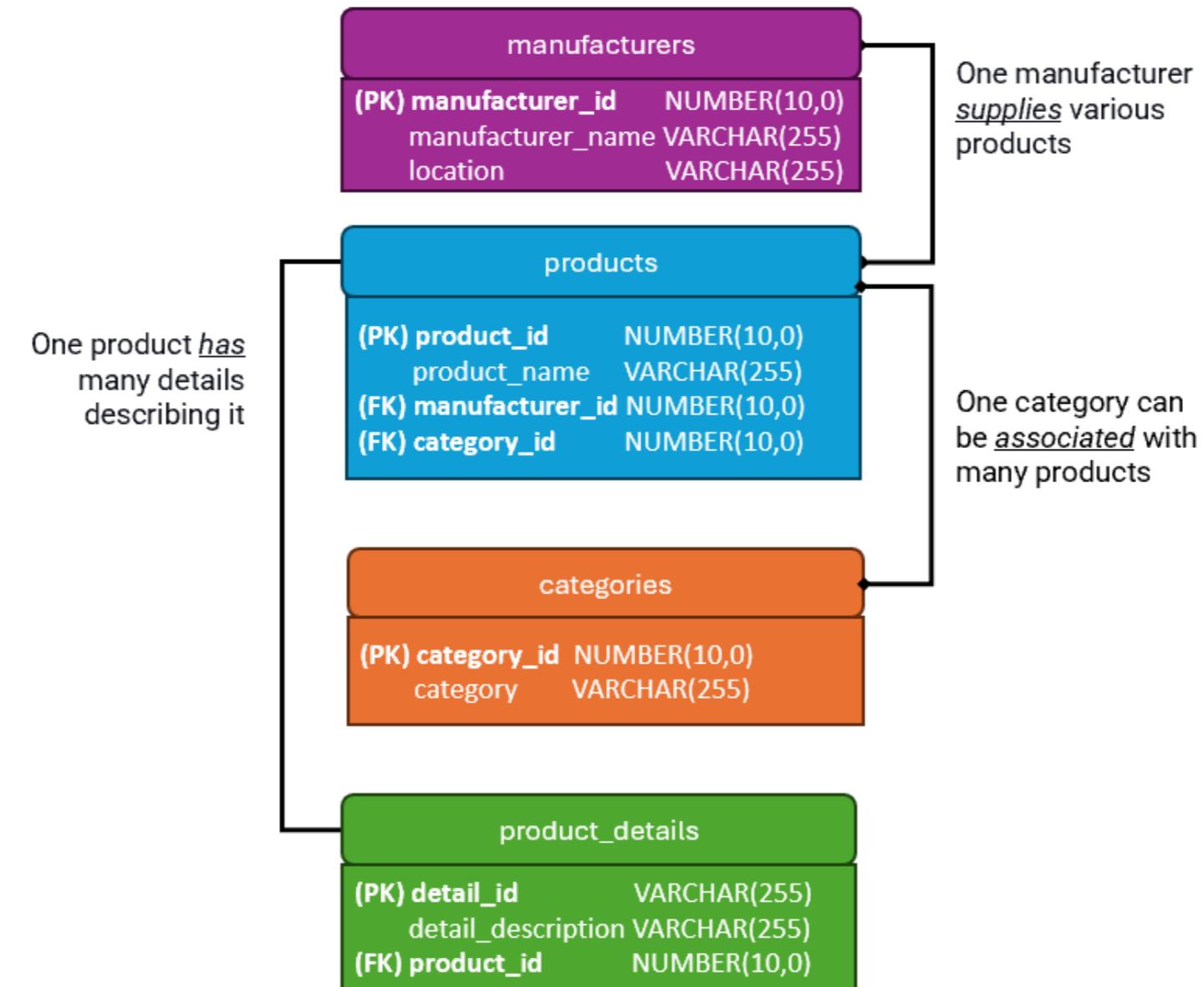
```
ALTER TABLE products
```

```
ADD COLUMN category_id NUMBER(10,0);
```

```
ALTER TABLE products
```

```
ADD FOREIGN KEY (category_id) REFERENCES categories(category_id);
```

Adjusting the data model (2)



Terminology and functions overview

- **Relational data:** Structured data organized into individual entities and keys that establish relationships between them
- **ALTER TABLE :** SQL command used to modify the structure of an existing entity
- **ADD :** SQL command, used with **ALTER TABLE** , to add new elements to the entity

```
ALTER TABLE table_name
```

```
ADD COLUMN column_name column_datatype;
```

```
ALTER TABLE table_name
```

```
ADD PRIMARY KEY (column_name);
```

```
ALTER TABLE table_name
```

```
ADD FOREIGN KEY (column_name) REFERENCES foreign_table(PK_from_foreign_table);
```

Let's practice!

INTRODUCTION TO DATA MODELING IN SNOWFLAKE

Normalizing Relational Data

INTRODUCTION TO DATA MODELING IN SNOWFLAKE



Nuno Rocha
Director of Engineering

Understanding unnormalized data

- **Unnormalized data (UNF):** Data that might lack a structure, be disorganized, contains repetitions and/or anomalies

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Understanding unnormalized data

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Understanding unnormalized data

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Understanding unnormalized data (3)

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Understanding unnormalized data (4)

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Understanding unnormalized data (5)

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Problems with unnormalized data

- **Data redundancy:** Unnecessary repetition of data
- **Data anomalies:** Irregularities or inconsistencies in the data

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Identifying unnormalized data

```
SELECT manufacturer_id,  
       manufacturer_name,  
       location,  
       COUNT(*) AS repetitions  
  FROM allproducts  
 GROUP BY manufacturer_id,  
          manufacturer_name,  
          location  
 HAVING COUNT(*) > 1;
```

manufacturer_id	manufacturer_name	location	product_count
101	Fresh Orchard Inc.	U.S.A.	3
102	Tropical Paradise	Costa Rica	2
103	Vineyard Bounty	France	3
104	Summer Harvest	Nicaragua	3
106	Hilltop Farms	Canada	2
105	Green Gardens	Spain	1

Identifying unnormalized data

```
SELECT DISTINCT category  
FROM allproducts;
```

category
C
A+
A, A+
B
AA
B, X, O
L, A, AA
L
B, L

Identifying unnormalized data

```
SELECT DISTINCT product_name,  
    category  
FROM allproducts  
WHERE category = 'L';
```

product_name	category
Lemon	L

Identifying unnormalized data

```
SELECT DISTINCT product_name,  
    category  
FROM allproducts  
WHERE category = 'L';
```

product_name	category
Lemon	L
Kiwi	L, A, AA
Raspberry	B, L

Normalized data

- **Normalized data:** Organized data into distinct and atomic entities
- **Benefits:**
 - Improves data accuracy and reliability
 - Enhances query performance
 - Optimized resource use
 - Scalability

Terminology and functions overview

- **Unnormalized data (UNF)**: Data that might lack a structure, be disorganized, or contain repetitions or anomalies
- **Normalized data**: Organized data into distinct and atomic entities
- **SELECT FROM** : SQL clause to retrieve data from a specific entity
- **DISTINCT** : SQL clause to return unique (different) values from an attribute
- **COUNT** : SQL clause that counts the number of rows that match the specified criteria
- **GROUP BY** : SQL clause that groups rows with the same values by the specified attributes
- **HAVING** : SQL clause used with GROUP BY to filter groups based on a condition
- **WHERE** : SQL clause to filter records based on a set condition
- **AS** : SQL clause used to rename a column or table with an alias

Functions overview

-- Querying unique values while being filtered by a specific condition

```
SELECT DISTINCT column_name  
FROM table_name  
WHERE column_name condition value;
```

-- Counting the values aggregated by a specific column while filtering the results

```
SELECT column_name,  
       COUNT(*) AS alias_name  
  FROM table_name  
 GROUP BY column_name  
 HAVING COUNT(*) condition value;
```

Let's practice!

INTRODUCTION TO DATA MODELING IN SNOWFLAKE

The First Norm

INTRODUCTION TO DATA MODELING IN SNOWFLAKE

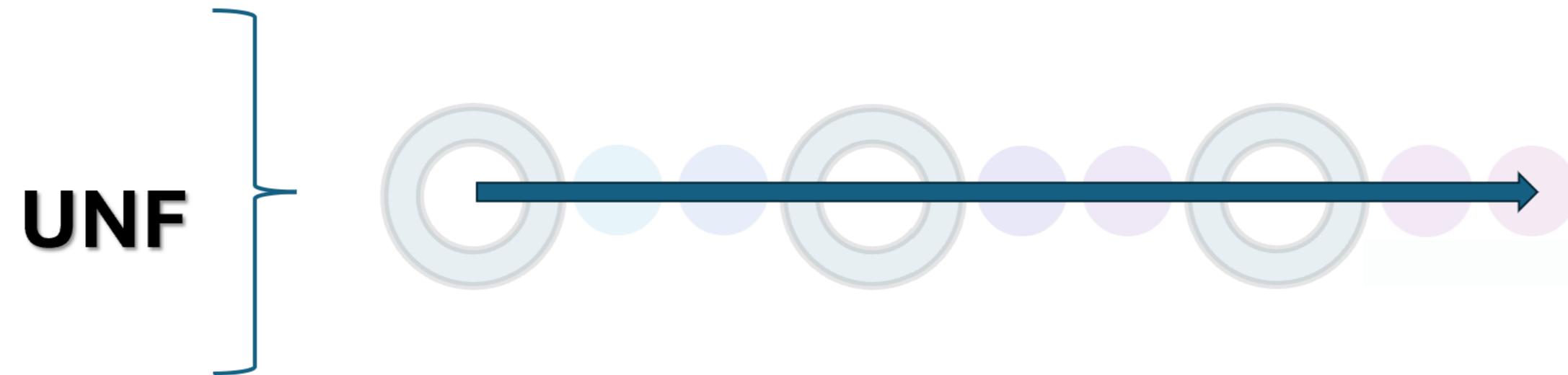


Nuno Rocha

Director of Engineering

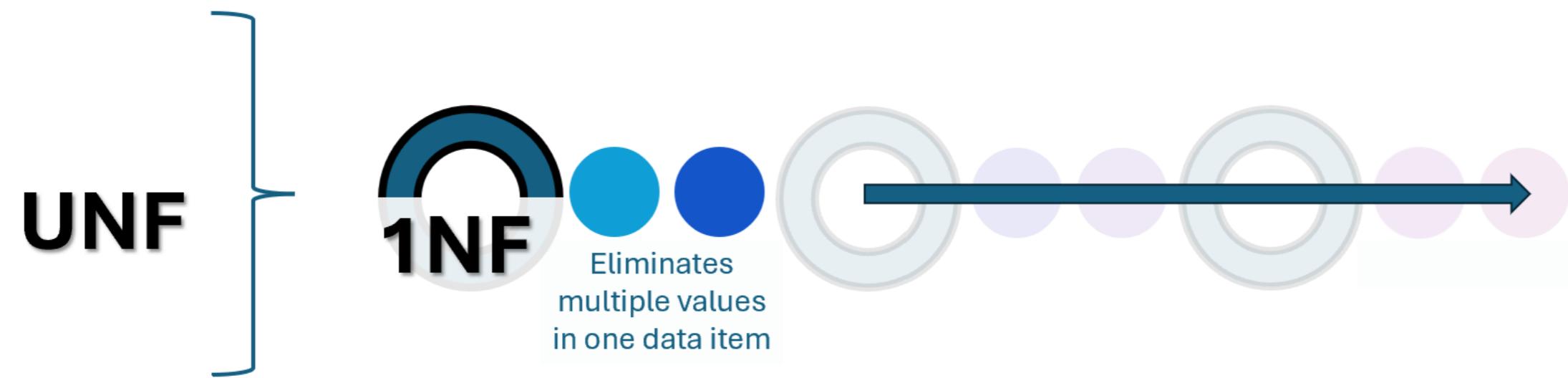
The process of data normalization

- **Data normalization:** Multi-step process of structuring data to minimize duplication and dependencies



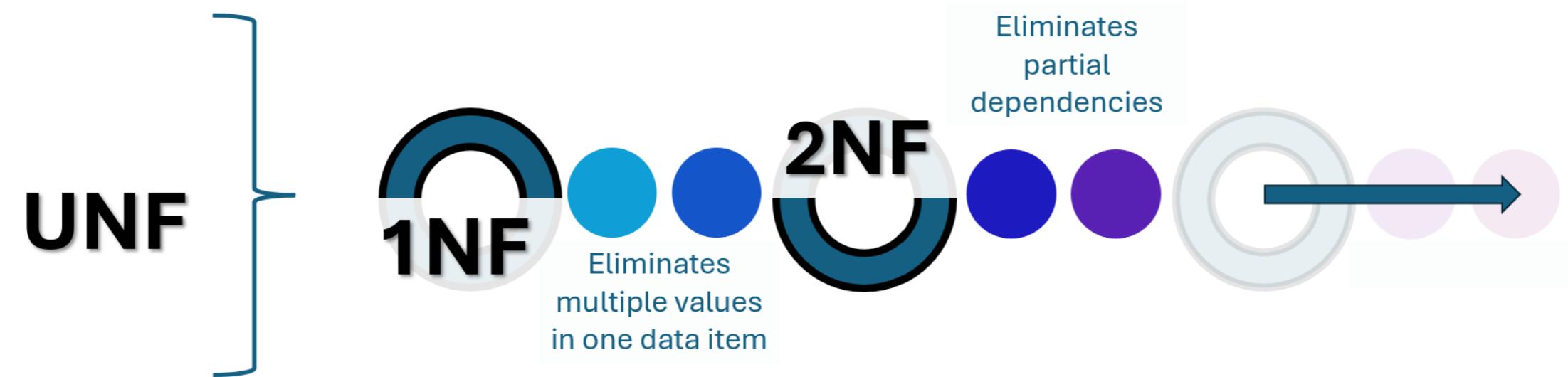
The process of data normalization(1)

- **Data normalization:** Multi-step process of structuring data to minimize duplication and dependencies



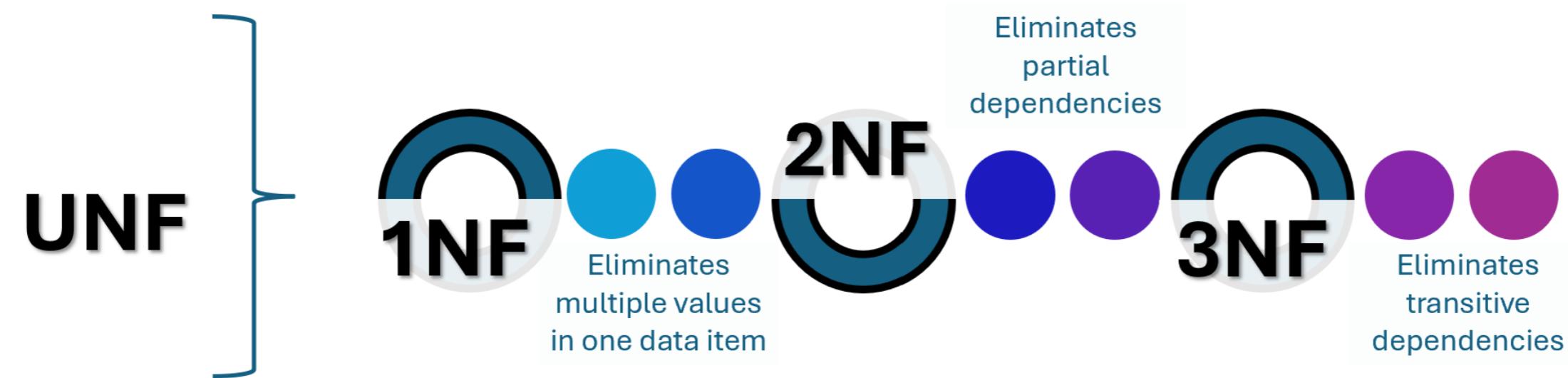
The process of data normalization (2)

- **Data normalization:** Multi-step process of structuring data to minimize duplication and dependencies



The process of data normalization (3)

- **Data normalization:** Multi-step process of structuring data to minimize duplication and dependencies



The first normal form

- **First normal form (1NF):** Ensures each column in an entity holds unique atomic values

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

The first normal form (1)

- UNF category data: L category is not isolated to be able to update the values

product_name	category
Lemon	L
Kiwi	L, A, AA
Raspberry	B, L

The first normal form (2)

- Validation step: Query the unique values to evaluate compliance with 1NF.

```
SELECT DISTINCT category  
FROM allproducts;
```

category
C
A+
A, A+
B
AA
B, X, O
L, A, AA
L
B, L

Snowflake functions for 1NF

- TRIM
- LATERAL & FLATTEN
- SPLIT

Snowflake functions for 1NF

- TRIM: Removes empty spaces from the start and end of values

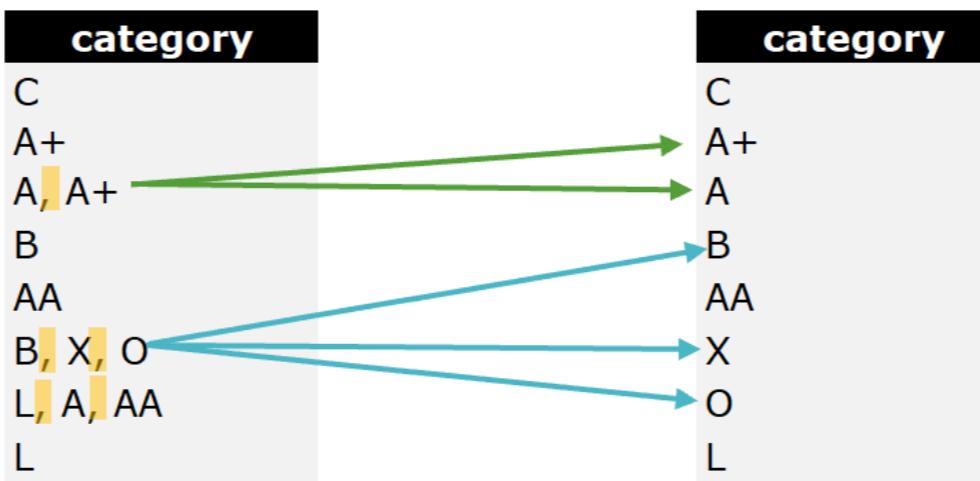
```
SELECT TRIM(category)  
FROM allproducts;
```

category
C
A+
A, A+
B
AA
B, X, O
L, A, AA
L
B, L

Snowflake functions for 1NF

- **TRIM**: Removes empty spaces from the start and end of values.
- **LATERAL & FLATTEN**: Treat list of values like a table with individual items.
- **SPLIT**: Separates values based on a delimiter.

```
SELECT TRIM(f.value) -- Clean empty values  
FROM allproducts,  
LATERAL FLATTEN(INPUT => SPLIT(allproducts.category, ',')) f;
```



Applying 1NF

- Step 1: Create a new entity to transfer the UNF attribute values

```
CREATE OR REPLACE TABLE categories (
    category_id NUMBER(10,0) PRIMARY KEY,
    category VARCHAR(255)
);
```

Applying 1NF

- Step 2.1: Fill new entity with data from the initially unnormalized entity
- **INSERT INTO** : SQL command to insert new rows into a table.

```
INSERT INTO categories (category_id, category)
```

```
---
```

Applying 1NF

- Step 2.2: Select the data from allproducts, the initially unnormalized entity

```
INSERT INTO categories (category_id, category)
```

```
SELECT
```

```
----,
```

```
---
```

```
FROM allproducts;
```

Applying 1NF

- Step 2.3: Use function to split values within a specific attribute

```
INSERT INTO categories (category_id, category)
SELECT
    --,
    TRIM(f.value)
FROM allproducts,
LATERAL FLATTEN(INPUT => SPLIT(allproducts.category, ',')) f;
```

Applying 1NF

- Step 2.4: Select the row number to define a unique identifier

Modify the script to insert this data into the ingredients entity, as ingredient_id and ingredient attribute.

```
INSERT INTO categories (category_id, category)
SELECT
    ROW_NUMBER() OVER (ORDER BY TRIM(f.value)), Create a sequential row number based on each result from TRIM(f.value).
    TRIM(f.value)
FROM allproducts,
LATERAL FLATTEN(INPUT => SPLIT(allproducts.category, ';')) f;
```

ROW_NUMBER() OVER (ORDER BY manufacturer, company_location) AS manufacturer_id,

Applying 1NF

- Step 2.4: Aggregate the data to generate unique category values

```
INSERT INTO categories (category_id, category)
SELECT
    ROW_NUMBER() OVER (ORDER BY TRIM(f.value)),
    TRIM(f.value)
FROM allproducts,
LATERAL FLATTEN(INPUT => SPLIT(allproducts.category, ';')) f
GROUP BY TRIM(f.value);
```

Towards data normalization

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

UNF

1NF

category_id	category
1	C
2	A+
3	B
4	AA
5	A
6	K
7	B
8	O
9	X
10	L

Modify manufacturers to remove the location attribute.

-- Modify entity

ALTER TABLE manufacturers

-- Remove attribute

DROP COLUMN IF EXISTS company_location;

Terminology and functions overview

- **Data normalization:** Process of structuring data to minimize duplication and dependency
- **Normal forms:** Guidelines to apply data normalization
- **First normal form (1NF):** Ensures each column in an entity holds unique atomic values
- **INSERT INTO :** SQL command to insert new rows into a table
- **TRIM :** SQL function to remove spaces at the start and end of values
- **LATERAL FLATTEN(INPUT => SPLIT()) :** Snowflake function to split values into a rows
- **ROW_NUMBER() OVER (ORDER BY) :** SQL function to generate a sequential number
- **GROUP BY :** SQL clause to aggregate data that have the same values

Functions overview

```
-- Fill a entity with data from a query result
INSERT INTO table_name (column_name, other_columns)
SELECT
    -- Generate a unique value using the row number
    ROW_NUMBER() OVER (ORDER BY TRIM(alias.value)),
    TRIM(alias.value)
FROM another_table,
-- Split a text attribute value based on a delimiter
LATERAL FLATTEN(INPUT => SPLIT(another_table.column_name, 'delimiter_value'))alias
-- Aggregate the data to ensure uniqueness of values
GROUP BY TRIM(alias.value);
```

Let's practice!

INTRODUCTION TO DATA MODELING IN SNOWFLAKE

2NF and 3NF

INTRODUCTION TO DATA MODELING IN SNOWFLAKE



Nuno Rocha

Director of Engineering

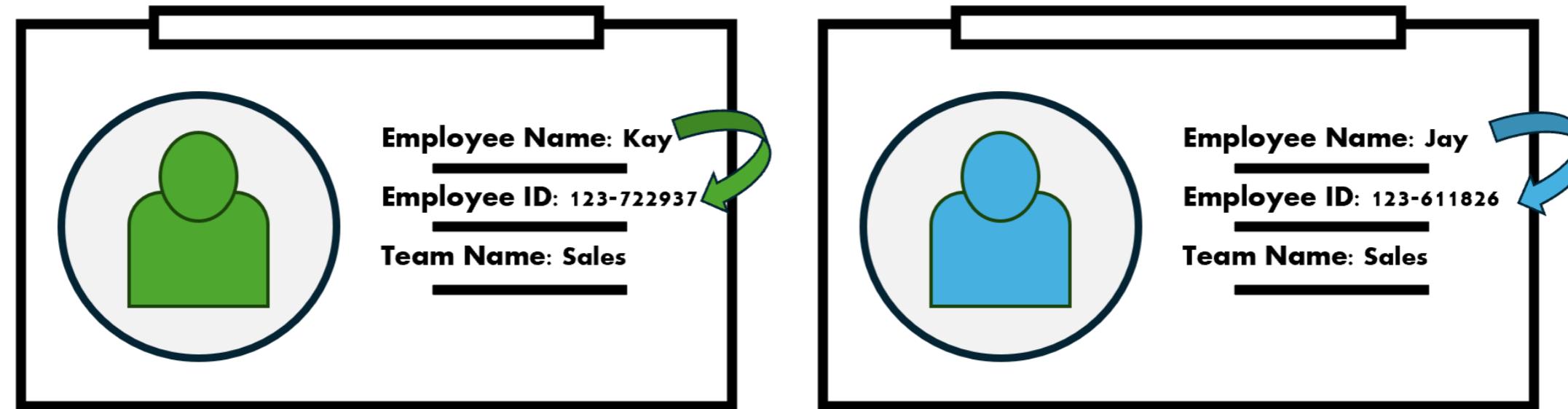
Introduction to 2NF

- Second normal form (2NF): Eliminates partial dependencies; every non-key attribute must functionally depend on the primary key



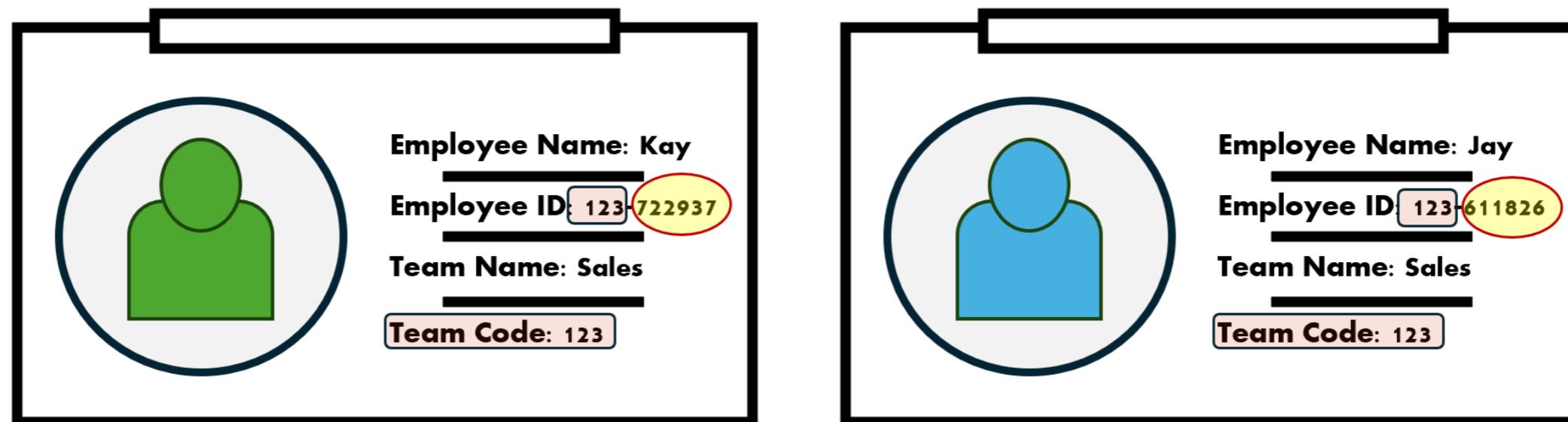
Introduction to 2NF (1)

- **Second normal form (2NF):** Eliminates partial dependencies; every non-key attribute must functionally depend on the primary key
- **Functional Dependency:** The primary key explicitly identifies an attribute



Introduction to 2NF (2)

- **Second normal form (2NF):** Eliminates partial dependencies; every non-key attribute must functionally depend on the primary key.
- **Functional Dependency:** The primary key explicitly identifies an attribute.
- **Partial Dependency:** Only part of the primary key is needed to identify an attribute.



The second normal form



A green checkmark icon is located in the top right corner of the table header.

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

The second normal form

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

The second normal form

product_id	product_name	manufacturer_id	manufacturer_name	location	detail_id	detail_description	category
9274	Apple	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	C
1442	Banana	102	Tropical Paradise	Costa Rica	501	Organically Grown	A+, A
1436	Pineapple	102	Tropical Paradise	Costa Rica	502	Non-GMO	B
3901	Grapes	103	Vineyard Bounty	France	800	Seedless variety	AA
5502	Watermelon	104	Summer Harvest	Nicaragua	504	Pesticide-Free	C
4345	Mango	104	Summer Harvest	Nicaragua	501	Organically Grown	K
4012	Strawberry	104	Summer Harvest	Nicaragua	589	Locally Sourced	B, X, O
3421	Blueberry	103	Vineyard Bounty	France	100	Handpicked	B, X, O
9274	Apple	106	Hilltop Farms	Canada	502	Non-GMO	X
3432	Kiwi	105	Green Gardens	Spain	101	Rich in Vitamins	L, A, AA
5645	Lemon	101	Fresh Orchard Inc.	U.S.A.	1011	Rich in Vitamins	L
7778	Cherry	106	Hilltop Farms	Canada	506	Seasonal Availability	A
9734	Raspberry	103	Vineyard Bounty	France	506	Seasonal Availability	B, L
3901	Grapes	101	Fresh Orchard Inc.	U.S.A.	501	Organically Grown	A+

Transitioning to 2NF

- Step 1: Create new entities to allocate the attributes that had a partial dependency.

```
CREATE OR REPLACE TABLE manufacturers (
    manufacturer_id NUMBER(10,0) PRIMARY KEY,
    manufacturer VARCHAR(255),
    location VARCHAR(255)
);
```

```
CREATE OR REPLACE TABLE details (
    detail_id NUMBER(10,0) PRIMARY KEY,
    detail VARCHAR(255)
);
```

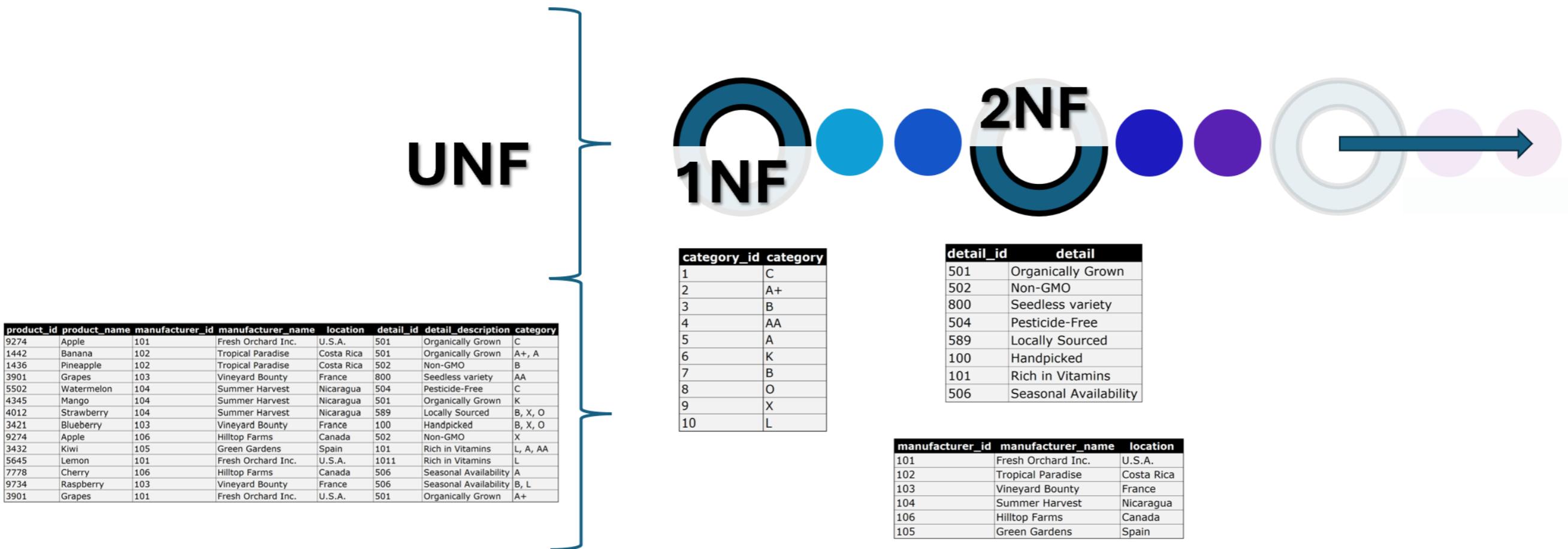
Transitioning to 2NF

- Step 2: Fill entities with data from the initially unnormalized entity.

```
INSERT INTO manufacturers (manufacturer_id, manufacturer, location)
SELECT DISTINCT manufacturer_id,
    manufacturer_name,
    location
FROM allproducts;
```

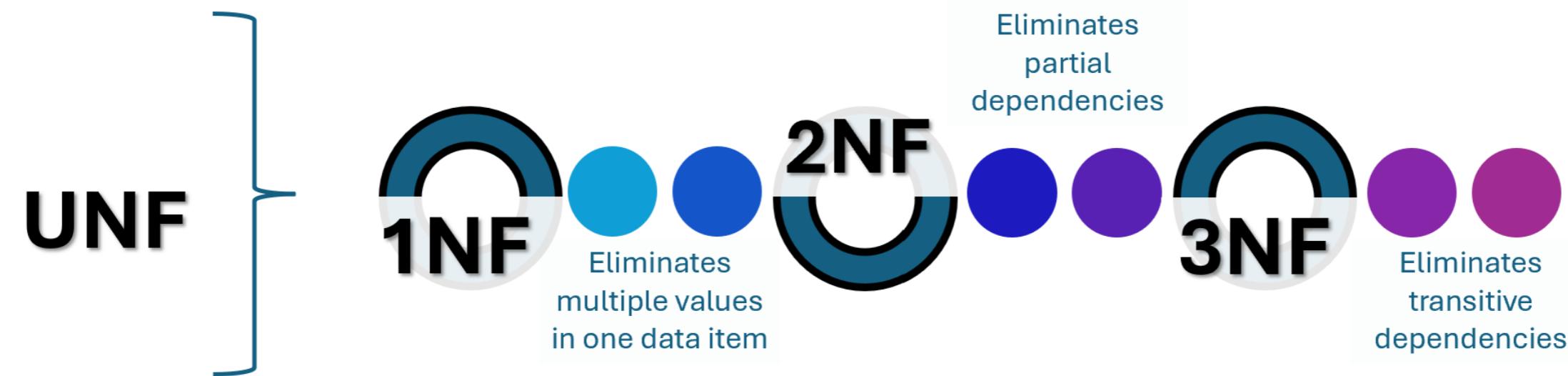
```
INSERT INTO details (detail_id, detail)
SELECT DISTINCT detail_id,
    detail_description
FROM allproducts;
```

Transitioning to 2NF



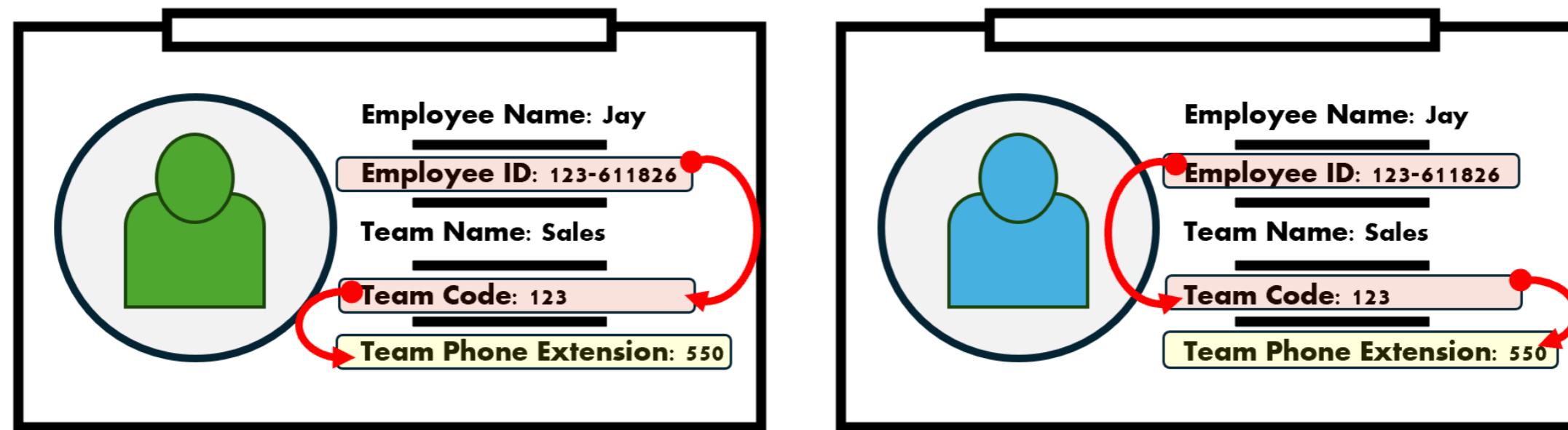
Introduction to 3NF

- Third normal form (3NF): Eliminates transitive dependencies; non-key attributes must directly depend on the primary key.



Introduction to 3NF

- **Third normal form (3NF):** Eliminates transitive dependencies; non-key attributes must directly depend on the primary key.
- **Transitive Dependency:** An attribute depends on another attribute, which is not the primary key.



The third normal form

A diagram illustrating a self-referencing foreign key constraint. A red curved arrow originates from the 'location' column of the first row (manufacturer_id 101) and points back to the 'manufacturer_id' column of the same row. Another red curved arrow originates from the 'location' column of the last row (manufacturer_id 105) and points back to the 'manufacturer_id' column of the same row. Both arrows feature a red 'X' at their midpoints, indicating that this relationship is incorrect or不合規範.

manufacturer_id	manufacturer_name	location
101	Fresh Orchard Inc.	U.S.A.
102	Tropical Paradise	Costa Rica
103	Vineyard Bounty	France
104	Summer Harvest	Nicaragua
106	Hilltop Farms	Canada
105	Green Gardens	Spain

Transitioning to 3NF

- Step 1: Create new entity to allocate the attributes that had a transitive dependency.

```
CREATE TABLE locations (
    location_id NUMBER(10,0) PRIMARY KEY,
    location VARCHAR(255)
);
```

Transitioning to 3NF

- Step 2: Fill entities with data from the initially unnormalized entity.

```
INSERT INTO locations (location_id, location)
SELECT ROW_NUMBER() OVER (ORDER BY location),
       location
FROM manufacturers
GROUP BY location;
```

```
ALTER TABLE manufacturers
DROP COLUMN location;
```

Transitioning to 3NF

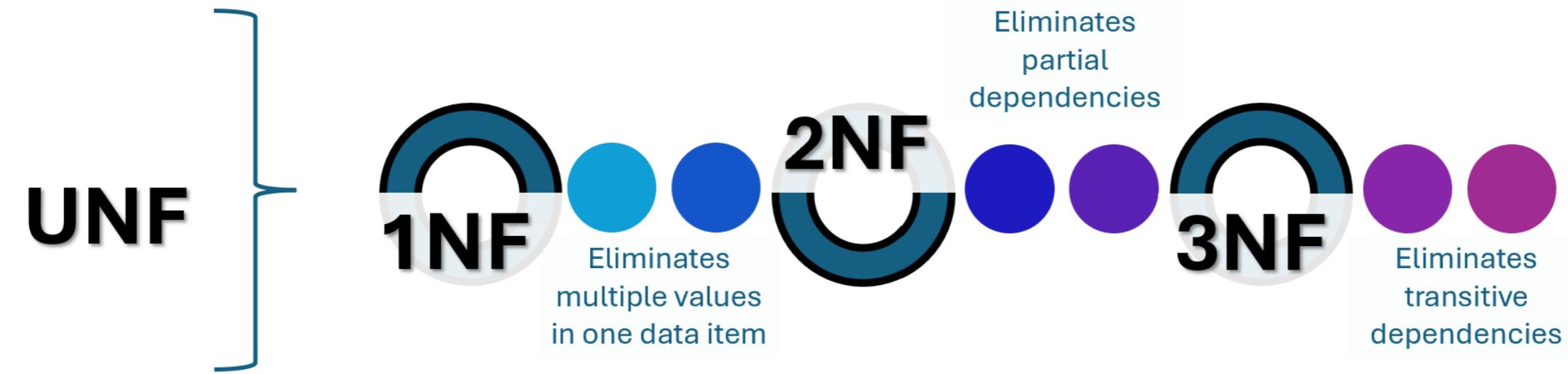
- **Step 3:** Create a new entity to extract from the unnormalized entity the remaining attributes.

```
CREATE OR REPLACE TABLE products (
    product_id NUMBER(10,0) PRIMARY KEY,
    name VARCHAR(255)
);
```

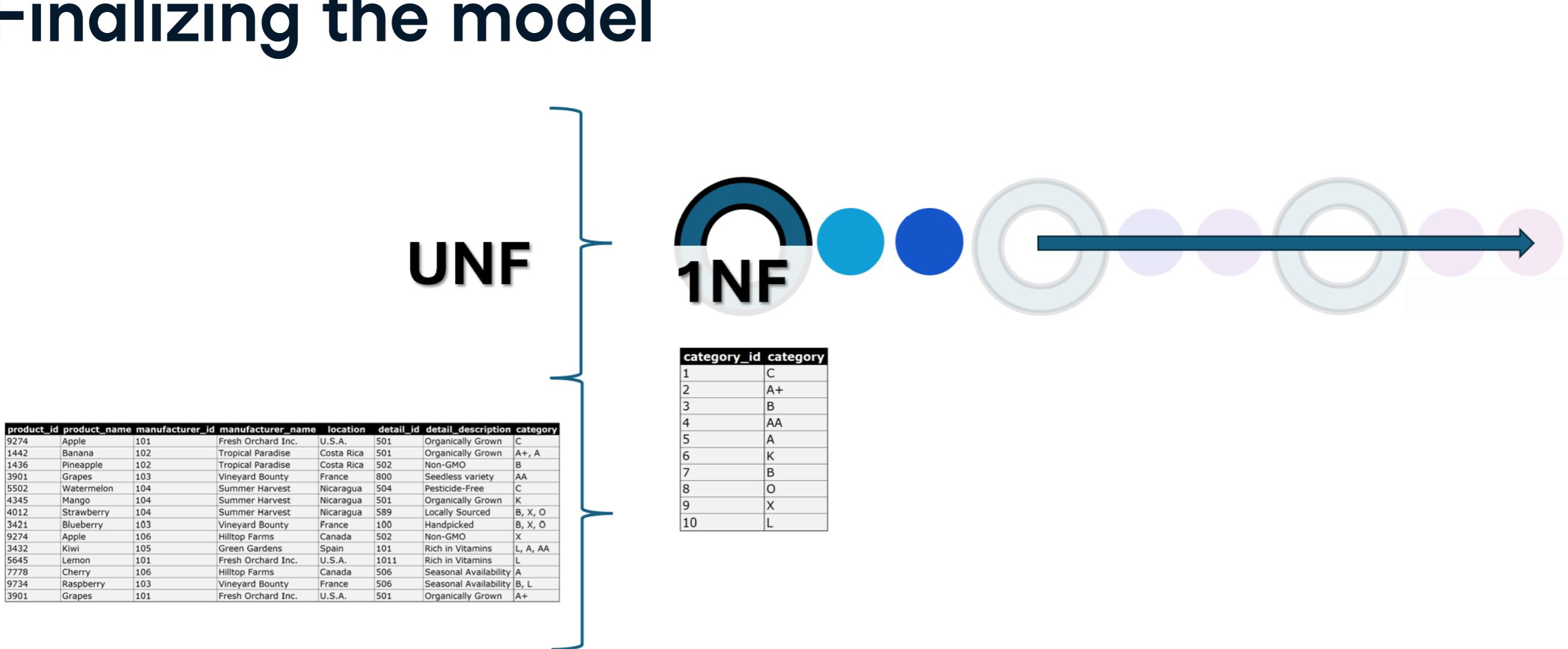
- **Step 4:** Fill the entity with the unique values left in the unnormalized entity.

```
INSERT INTO products (product_id, name)
SELECT DISTINCT product_id,
    product_name
FROM allproducts;
```

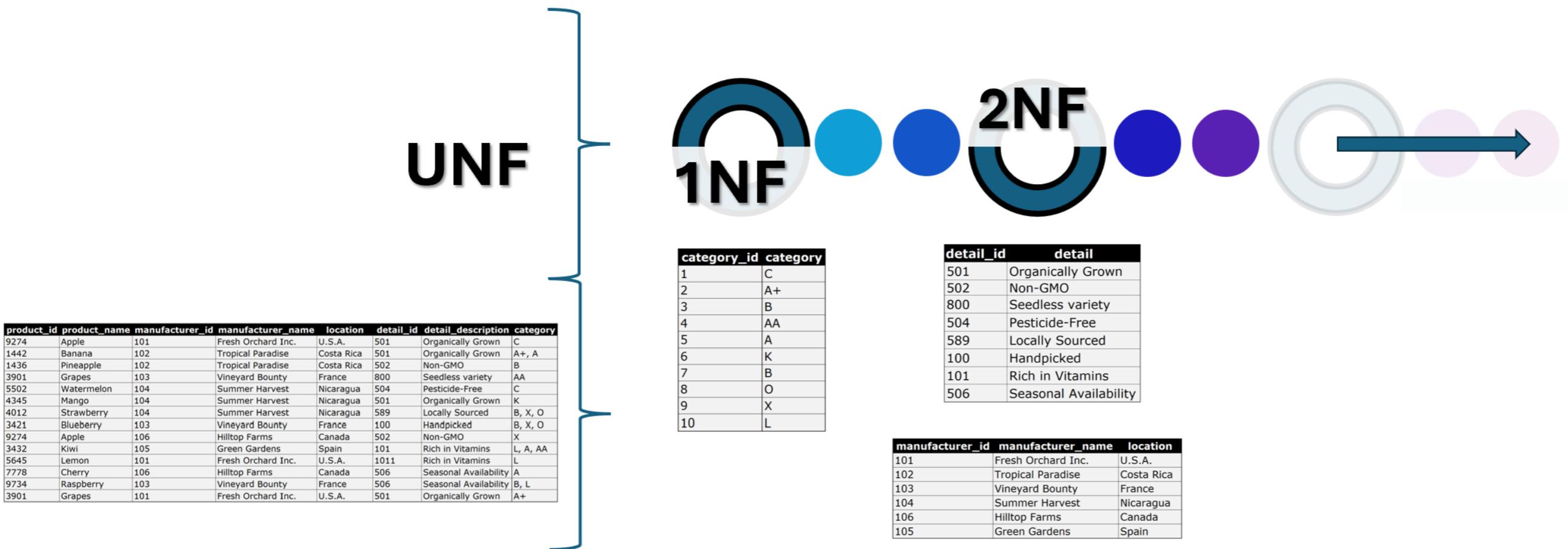
Finalizing the model



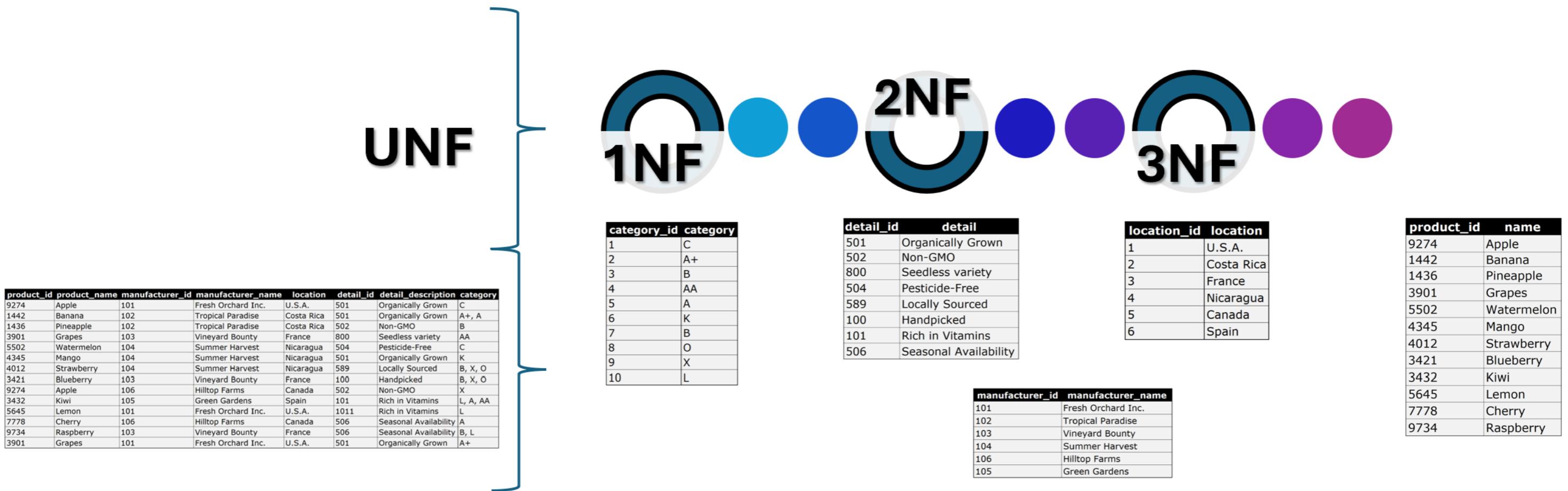
Finalizing the model



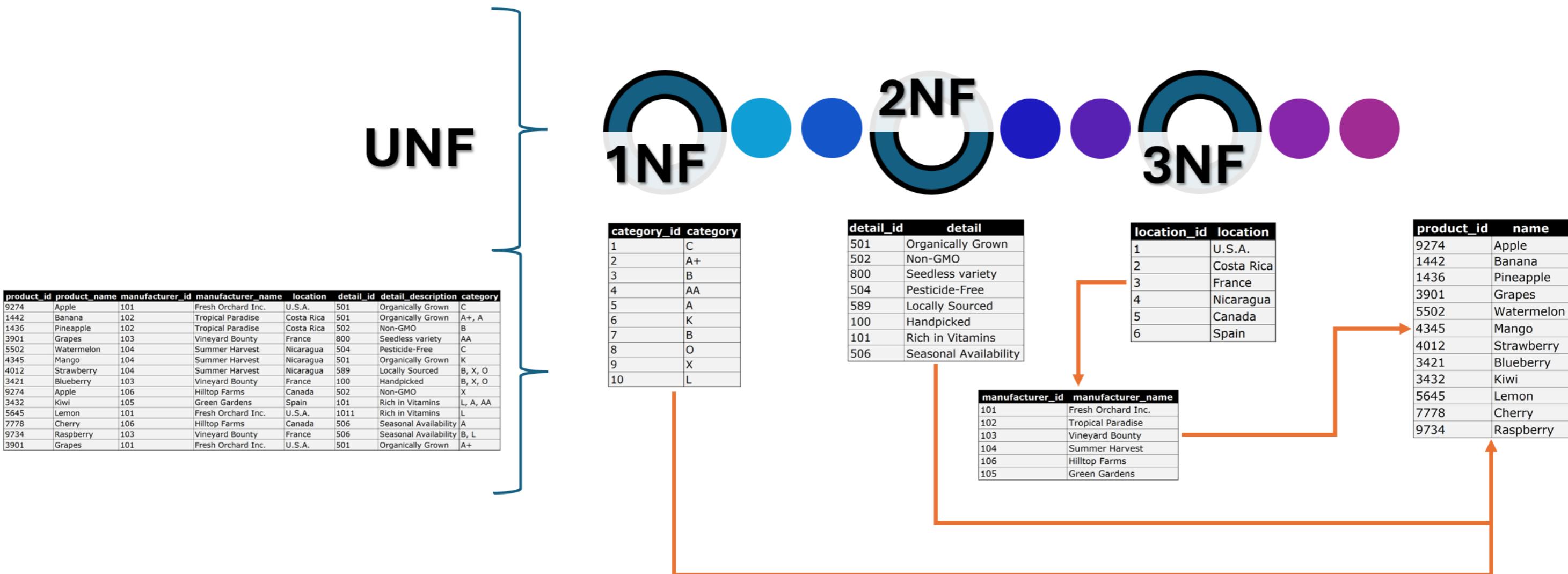
Finalizing the model



Finalizing the model



Finalizing the model



Terminology and functions overview

- **Second normal form (2NF)**: Eliminates partial dependencies; every non-key attribute must functionally depend on the primary key.
- **Third normal form (3NF)**: Eliminates transitive dependencies; non-key attributes must directly depend on the primary key.
- **Functional Dependency**: The primary key explicitly identifies an attribute.
- **Partial Dependency**: Only part of the primary key is needed to identify an attribute.
- **Transitive Dependency**: An attribute depends on another attribute, which is not the primary key.
- **ROW_NUMBER() OVER (ORDER BY)** : SQL function to generate a sequential number.
- **DISTINCT** : SQL clause to return unique (different) values from an attribute.
- **DROP** : SQL command, used with **ALTER TABLE**, to remove elements to the entity.

Functions overview

```
INSERT INTO table_name (column_name, other_columns)  
SELECT DISTINCT column_name,  
    other_columns  
FROM another_table;
```

```
INSERT INTO table_name (column_name)  
SELECT ROW_NUMBER() OVER (ORDER BY column_name)  
FROM another_table  
GROUP BY TRIM(column_name);
```

```
ALTER TABLE table_name  
DROP COLUMN column_name;
```

Let's practice!

INTRODUCTION TO DATA MODELING IN SNOWFLAKE