Relational Database Design

CS 327E
September 25, 2017

Announcements:

- Next week: Lab 2
- Next class: Postgres RDS instance
- **Before next class**: Install psql locally (see snippets wiki for instructions)

1)Which of the following help avoid most of the problems that arise from poor database design?

A) Serialization
B) Realization
C) Normalization
D) Data Transformation

2) An **Unnormalized** model means that the database schema is in First Normal Form.

A) True
B) False

3) A **functional dependency** is a relationship among attributes.

A) True
B) False

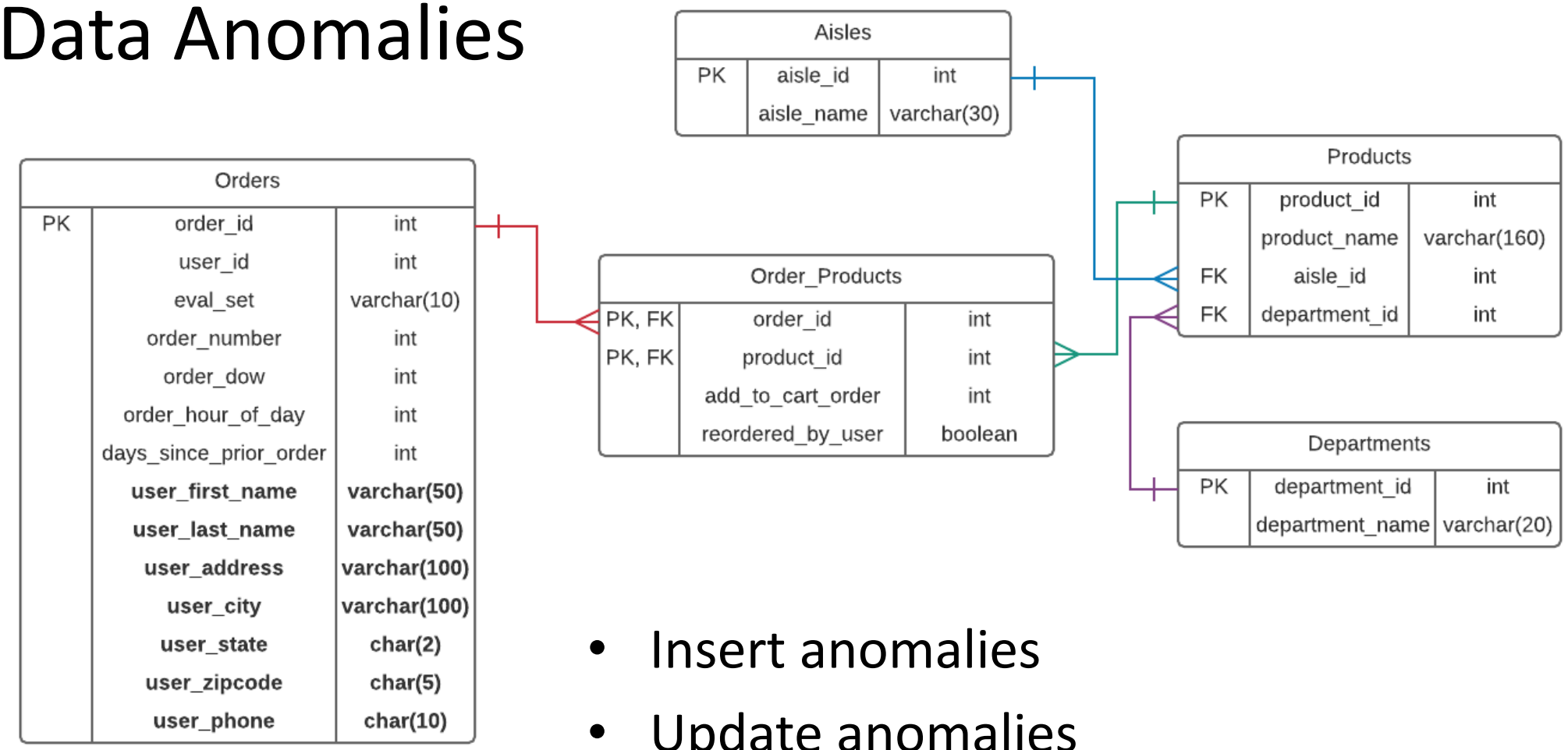
4) Keys made up of multiple fields are known as **composite keys**.

A) True
B) False

5) The higher the normal form, the fewer the number of tables in a database.

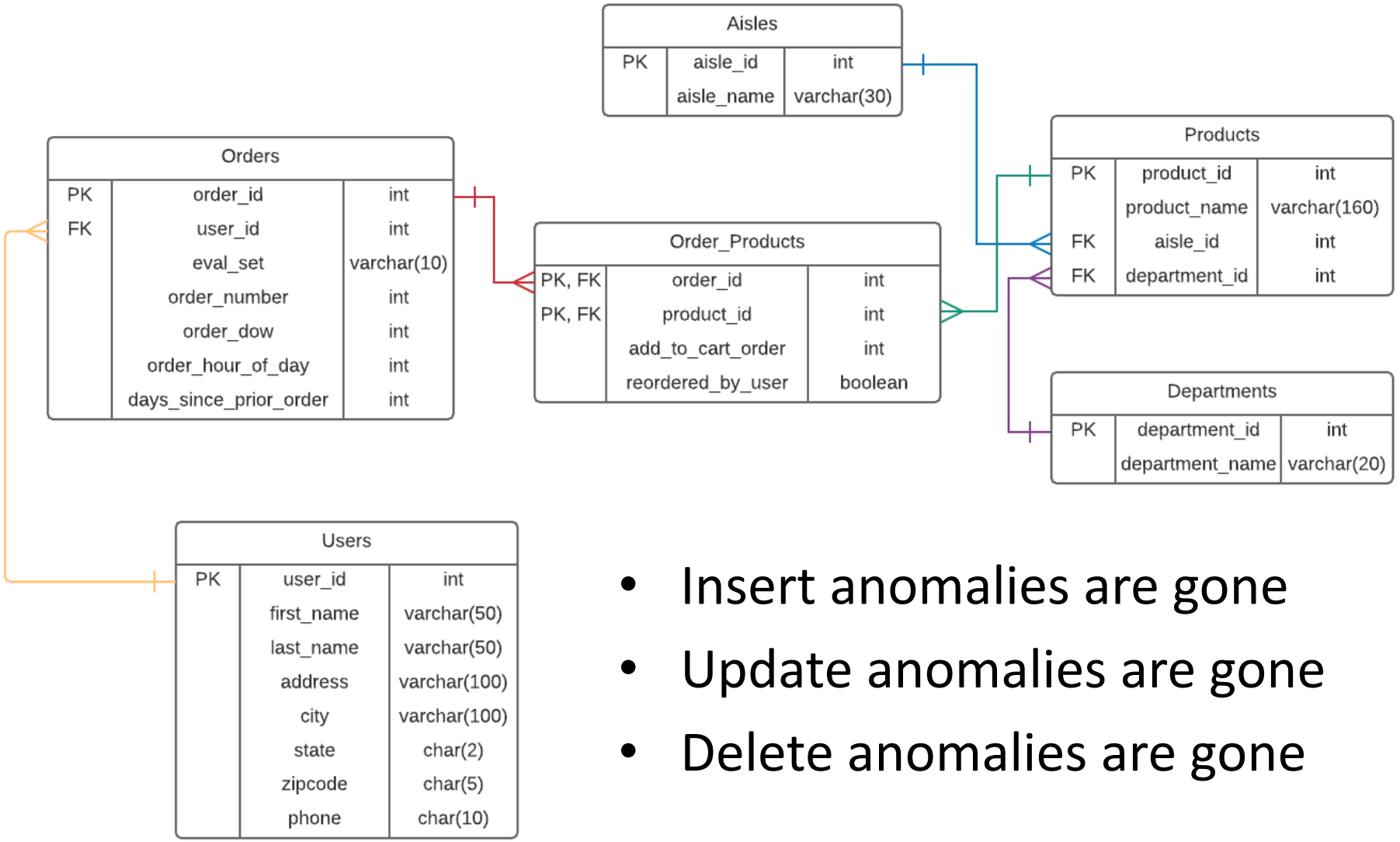
A) True
B) False

Data Anomalies



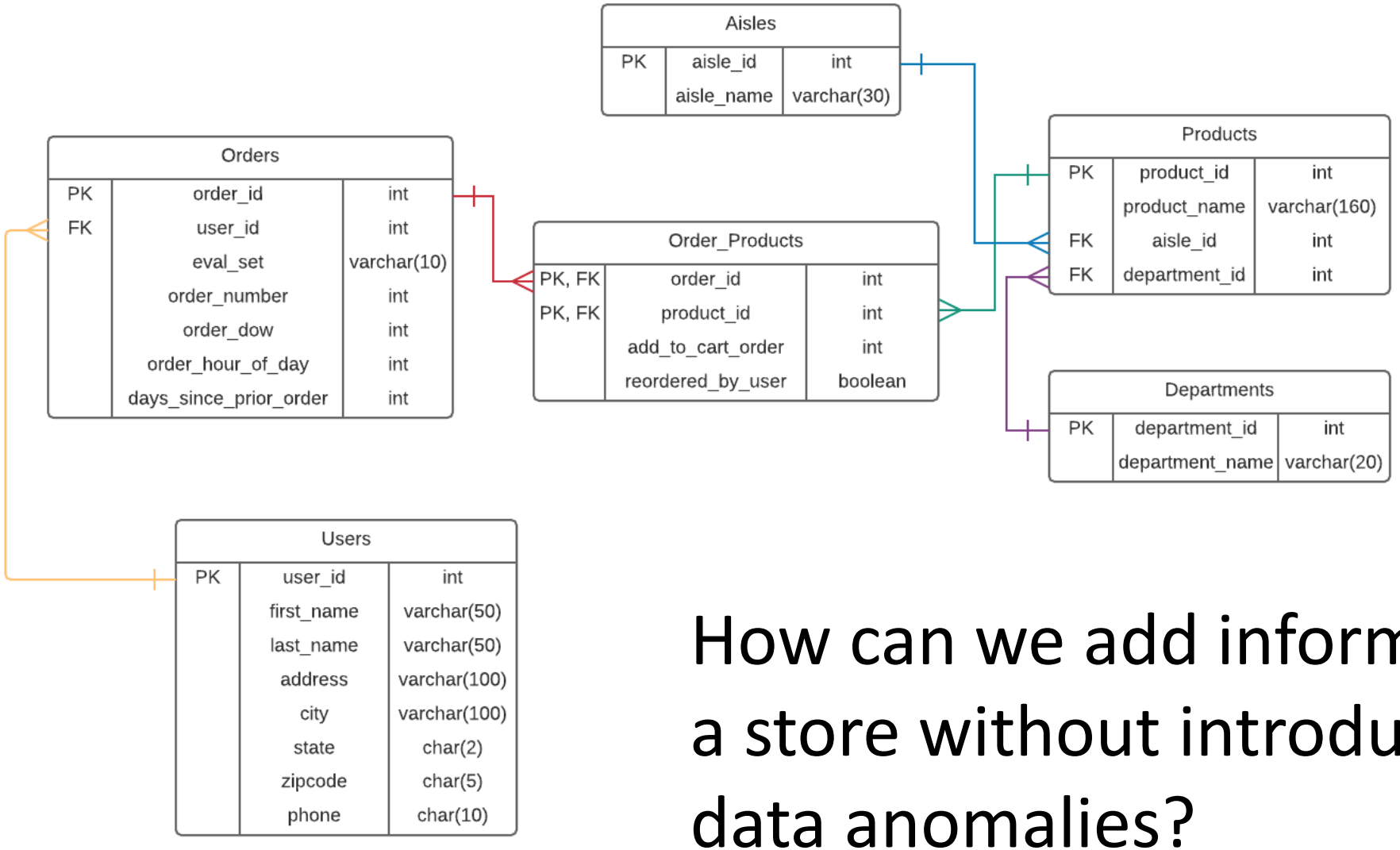
- Insert anomalies
- Update anomalies
- Delete anomalies

Relation Decomposition



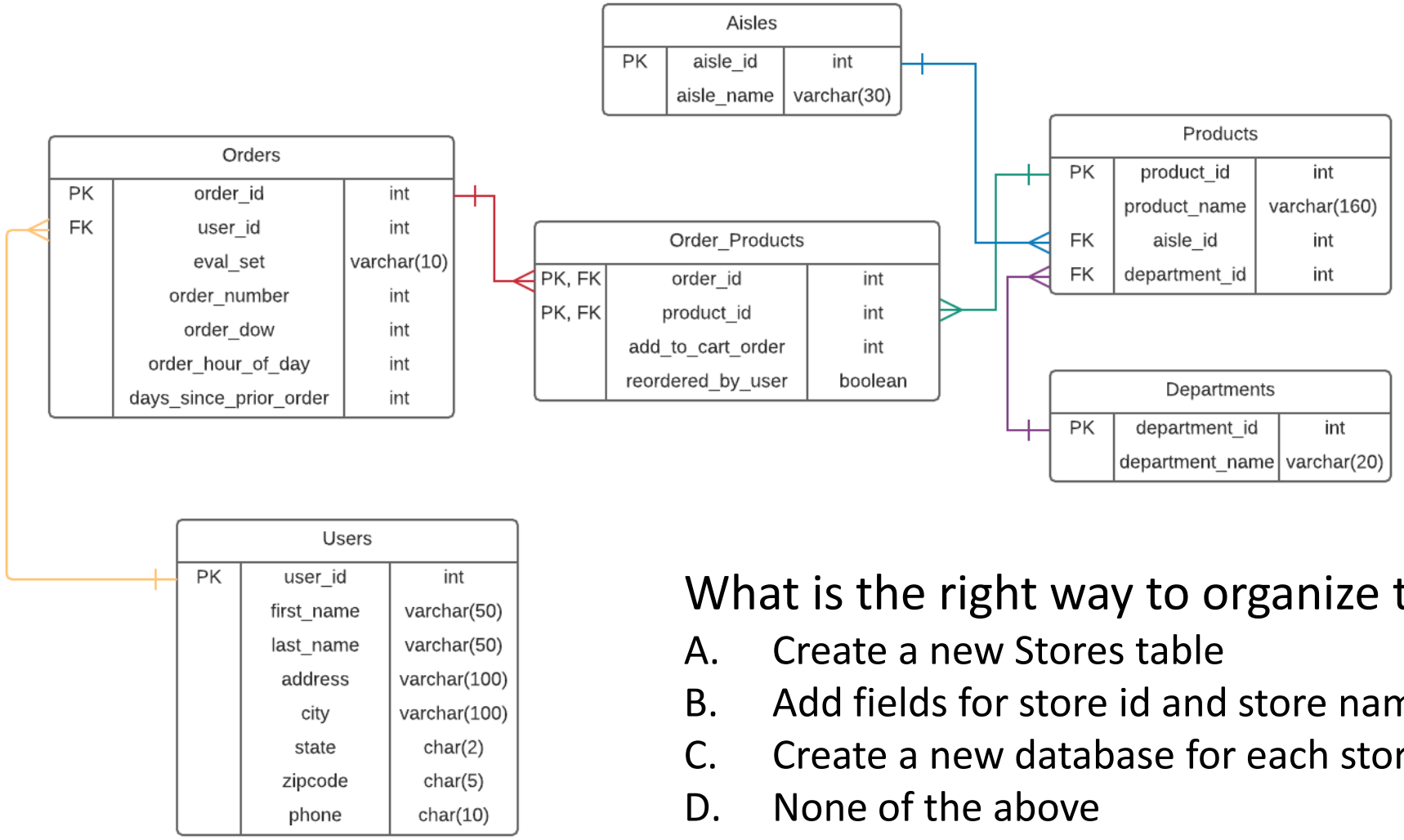
- Insert anomalies are gone
- Update anomalies are gone
- Delete anomalies are gone

Practice Problem 1: Instacart users can order at a number of stores (e.g. Whole Foods, HEB, etc.).



How can we add information about a store without introducing new data anomalies?

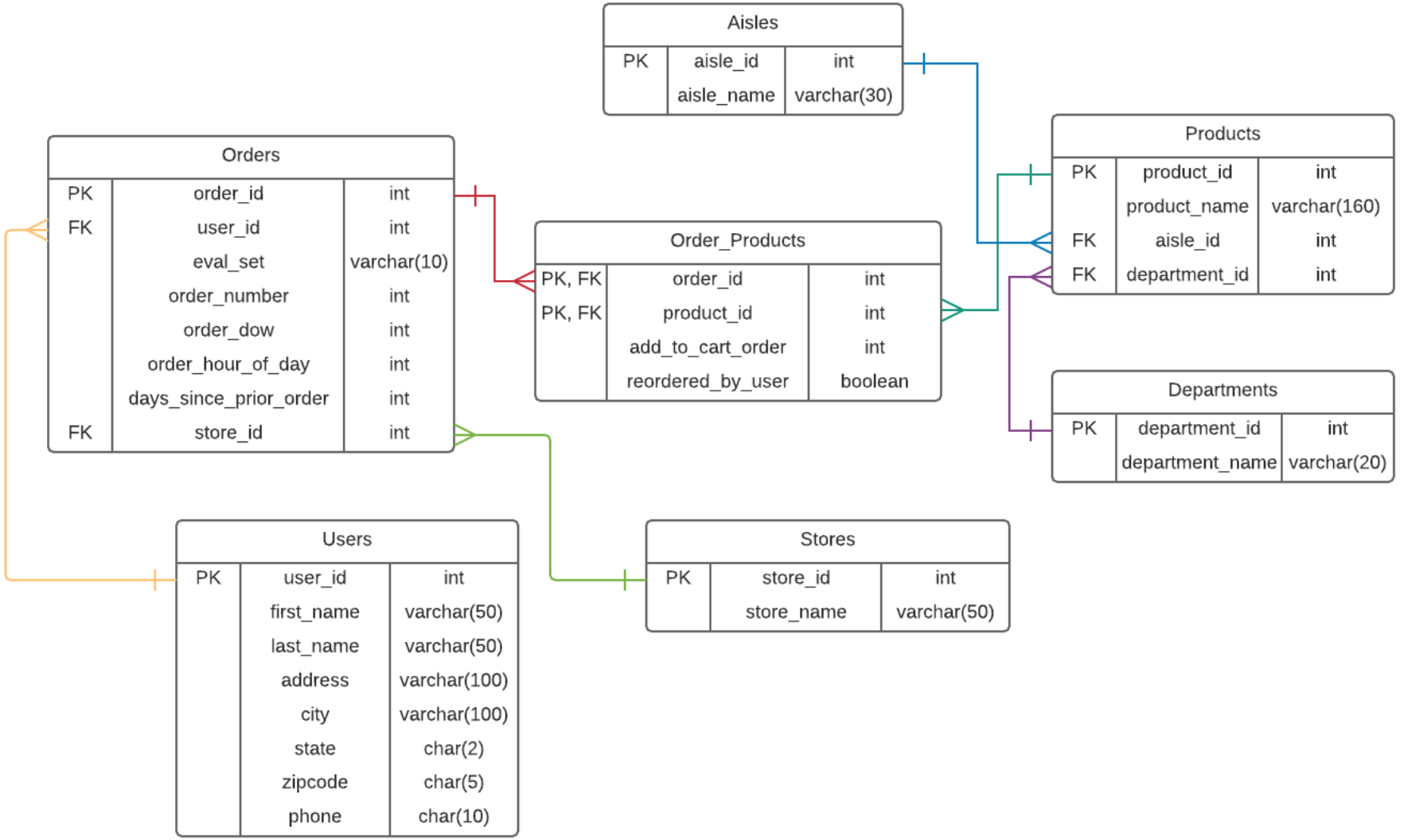
Practice Problem 1: How can we add information about a store without introducing new data anomalies?



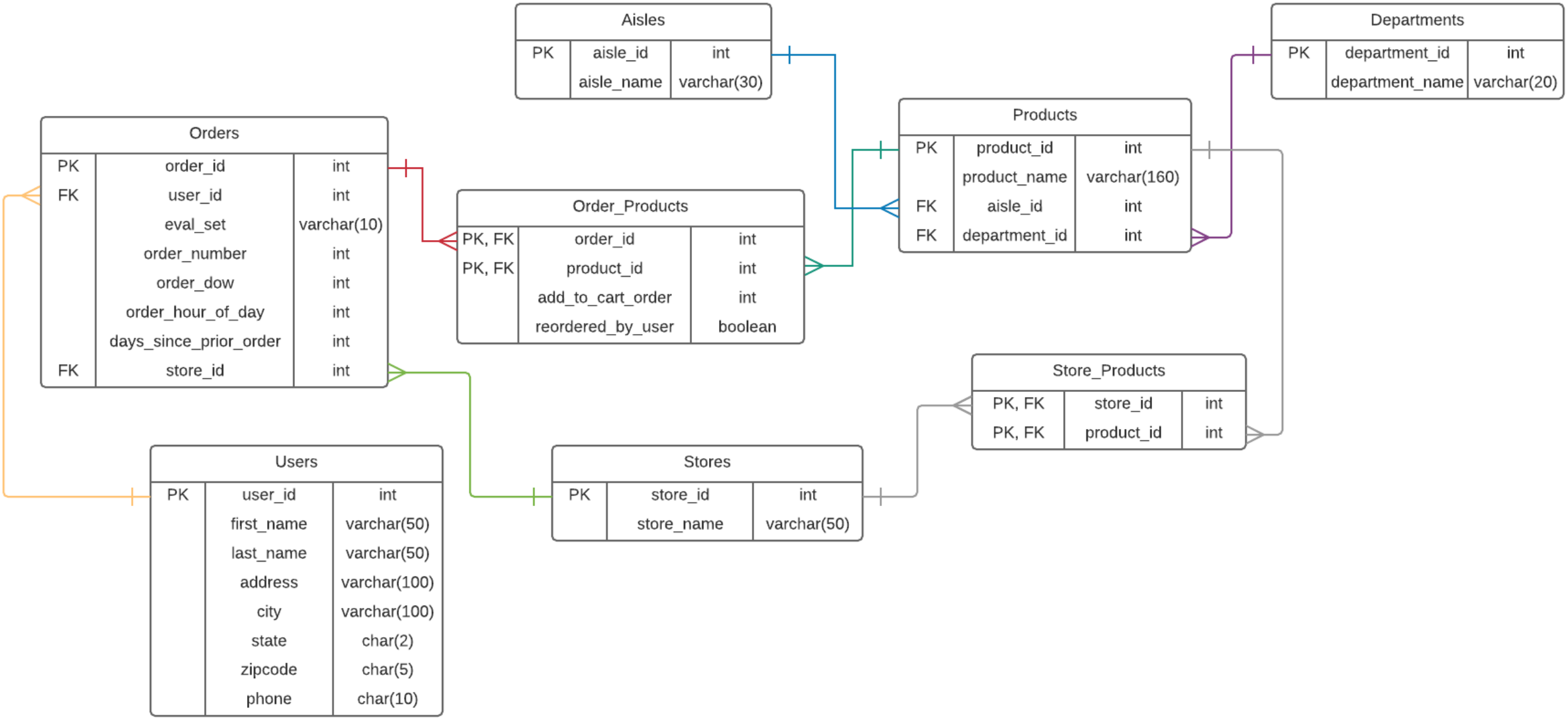
What is the right way to organize the data?

A. Create a new Stores table
B. Add fields for store id and store name to existing tables
C. Create a new database for each store
D. None of the above

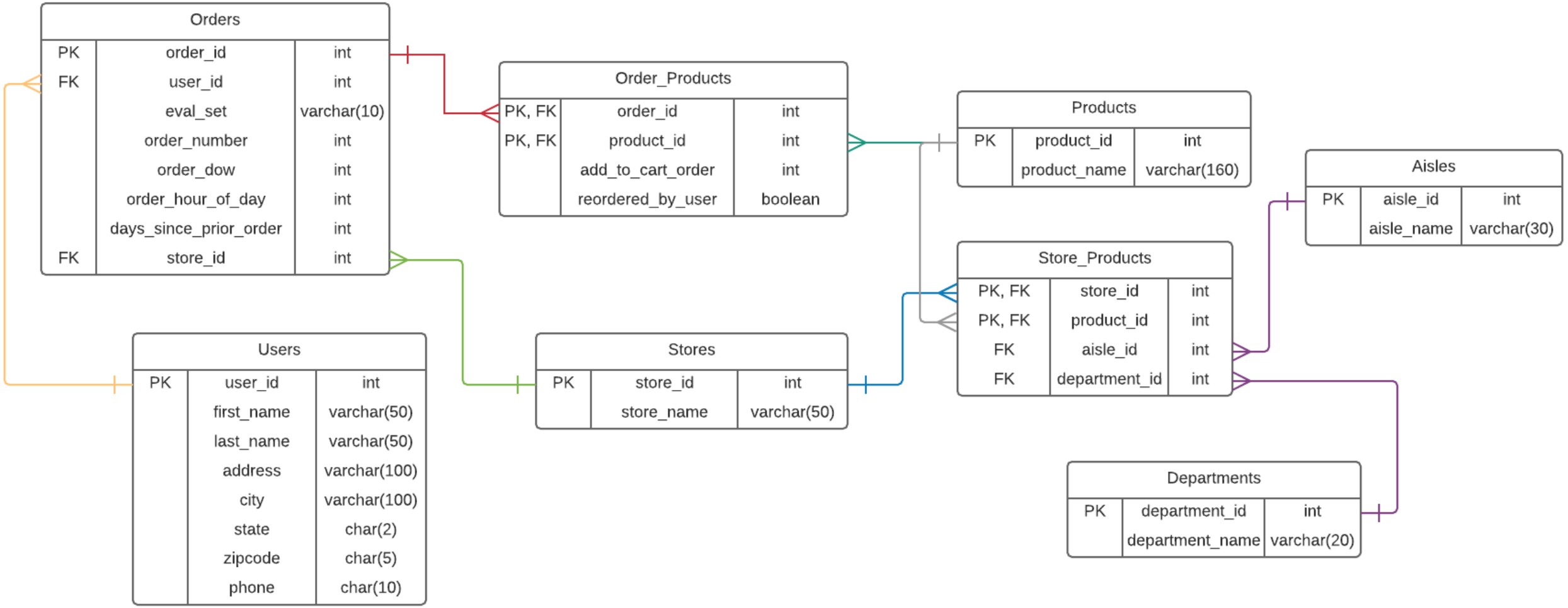
Practice Problem 1: How can we add information about a store without introducing new data anomalies?



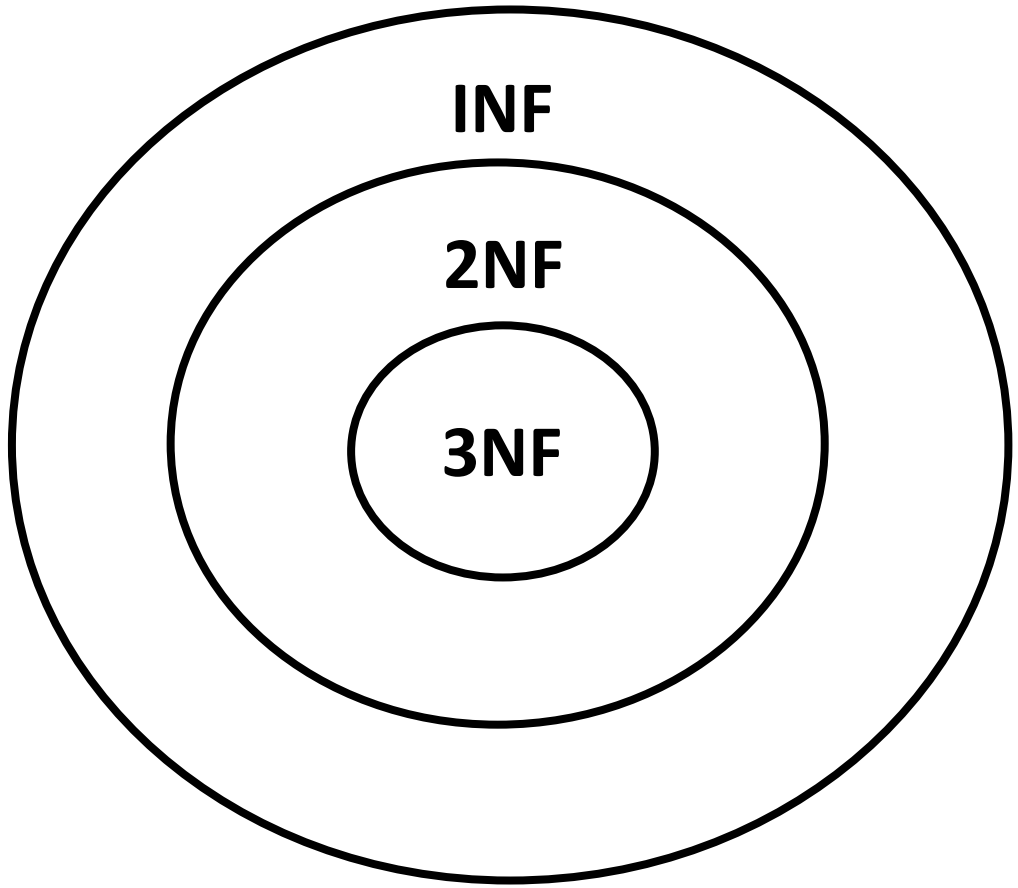
Practice Problem 1: How can we add information about a store without introducing new data anomalies?



Practice Problem 1: How can we add information about a store without introducing new data anomalies?



Normalization Theory



Unnormalized

Unnormalized to 1NF

Rule: A database schema is in 1NF *iff* all attributes have scalar values.

Student_Semester

EID	Semester	GPA	Classes
alice1	Fall17	3.9	Stats: A DB: A Alg: A-
bob20	Fall17	3.7	DB: A Alg: B+
carol30	Fall17	3.5	Stats: A- Alg: B+

Unnormalized

Student_Semester'

EID	Semester	Class	Grade	GPA
alice1	Fall17	Stats	A	3.9
alice1	Fall17	DB	A	3.9
alice1	Fall17	Alg	A-	3.9
bob20	Fall17	DB	A	3.7
bob20	Fall17	Alg	B	3.7
carol30	Fall17	Stats	3.5	3.5
carol30	Fall17	Alg	3.5	3.5

1NF

Functional Dependencies

Definition:

If two records agree on the attributes

$A_1, A_2, \ldots, A_n$

then they must also agree on the attributes

$B_1, B_2, \ldots, B_n$

Formally:

$A_1, A_2, \ldots, A_n \rightarrow B_1, B_2, \ldots, B_n$

FD Example

Which FDs **hold** and **do not hold** on this table?

ID	**Name**	**Phone**	**City**
C0012	Smith	5555	Austin
C3412	Wallace	9876	Houston
C1111	Smith	9876	Dallas
C2323	Johnston	5555	Austin

ID → Name, Phone, City
City → Phone

Not Phone → City
Not Name → Phone

1NF to 2NF

Rule: A database schema is in 2NF *iff* it is in 1NF and there exists no partial FDs on the primary key (i.e. all non-key attributes must be dependent on the entire PK)

Student_Semester

EID	Semester	Class	Grade	Sem_GPA
alice1	Fall17	Stats	A	3.9
alice1	Fall17	DB	A	3.9
alice1	Fall17	Alg	A-	3.9
bob20	Fall17	DB	A	3.7
bob20	Fall17	Alg	B+	3.7
carol30	Fall17	Stats	A-	3.5
carol30	Fall17	Alg	B+	3.5

1NF

FDs:

1. EID, Semester, Class → Grade
2. EID, Semester → Sem_GPA

Student_Semester_Grade

EID	Semester	Class	Grade
alice1	Fall17	Stats	A
alice1	Fall17	DB	A
alice1	Fall17	Alg	A-
bob20	Fall17	DB	A
bob20	Fall17	Alg	B+
carol30	Fall17	Stats	A-
carol30	Fall17	Alg	B+

2NF

Student_Semester_GPA

EID	Semester	GPA
alice1	Fall17	3.9
bob20	Fall17	3.7
Carol30	Fall17	3.5

2NF

2NF to 3NF

Rule: A database schema is in 3NF *iff* it is in 2NF and there exists no non-key attributes that are functionally determined by other non-key attributes.

Student_Major

2NF

EID	**Name**	**Major**	**College**
alice1	Alice	Math	Natural Sciences
bob20	Bob	CS	Natural Sciences
carol30	Carol	Math	Natural Sciences

FDs:

1. EID → Name, Major
2. Major → College

Student_Major'

3NF

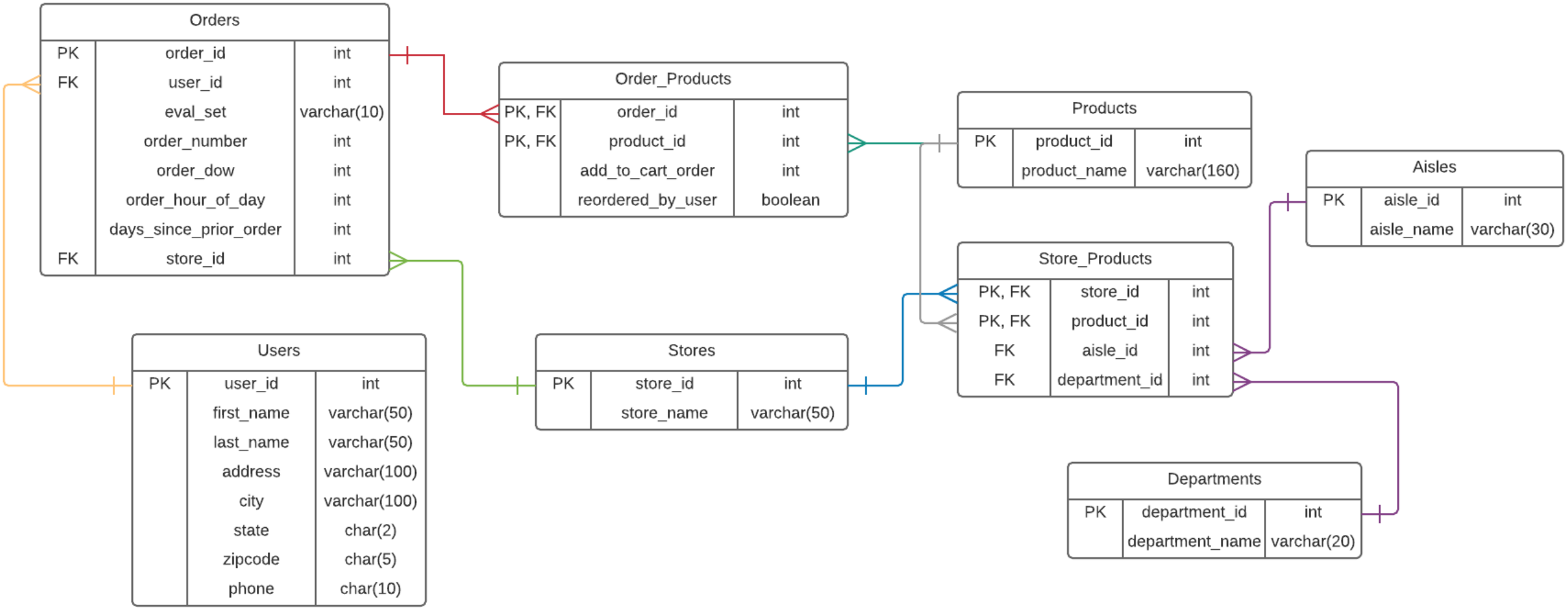
EID	**Name**	**Major**
alice1	Alice	Math
bob20	Bob	CS
carol30	Carol	Math

Major_College

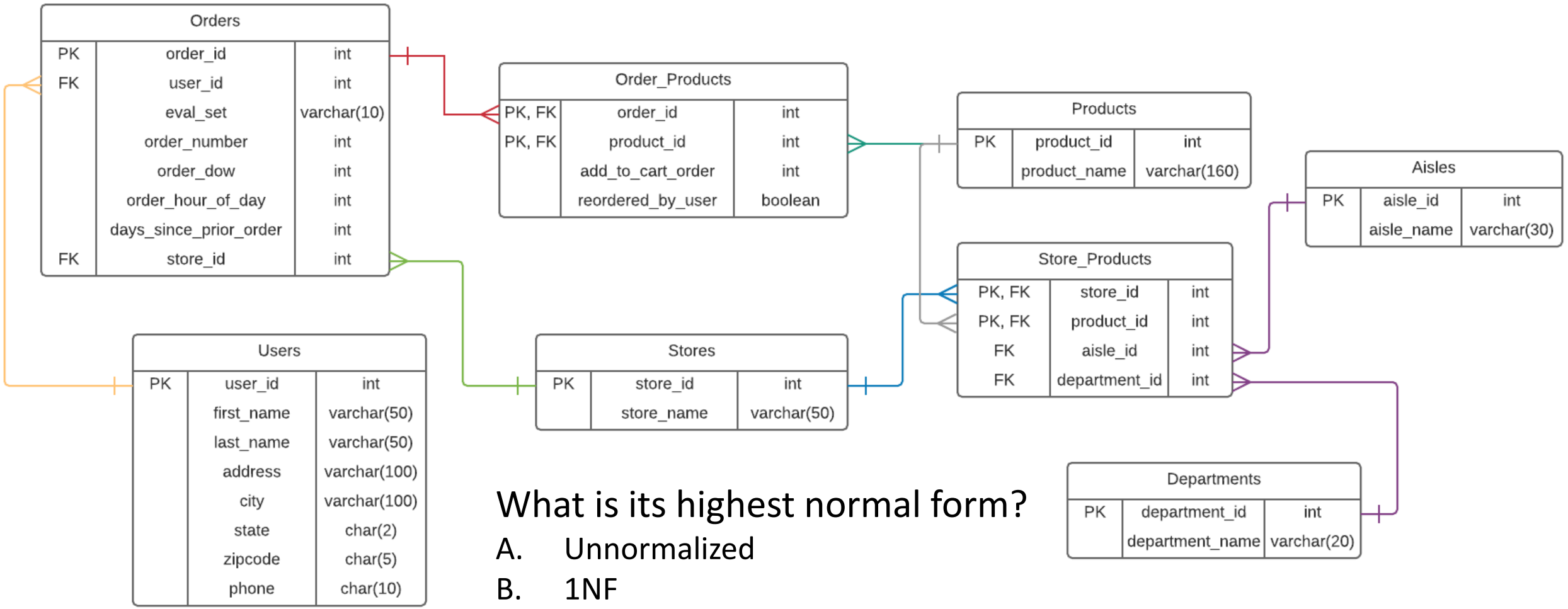
Major	**College**
Math	Natural Sciences
CS	Natural Sciences

3NF

Practice Problem 2: Is the Instacart schema normalized?



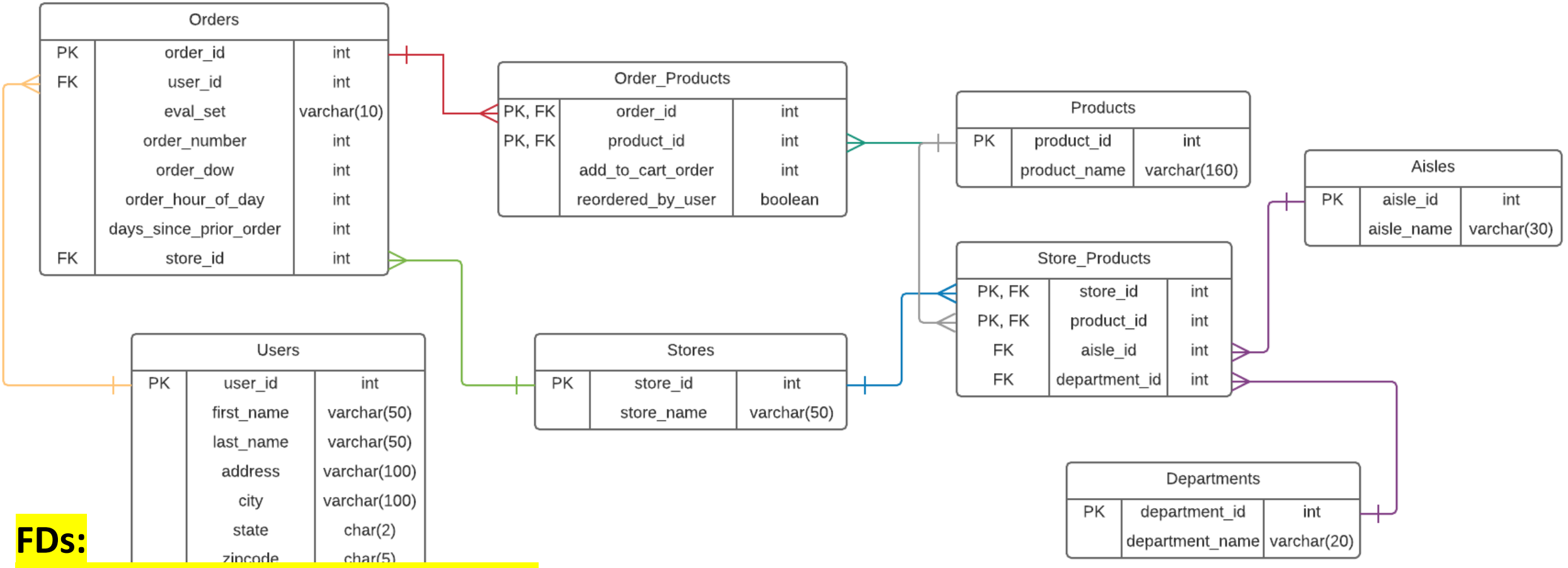
Practice Problem 2: Is the Instacart schema normalized?



What is its highest normal form?

A. Unnormalized
B. 1NF
C. 2NF
D. 3NF

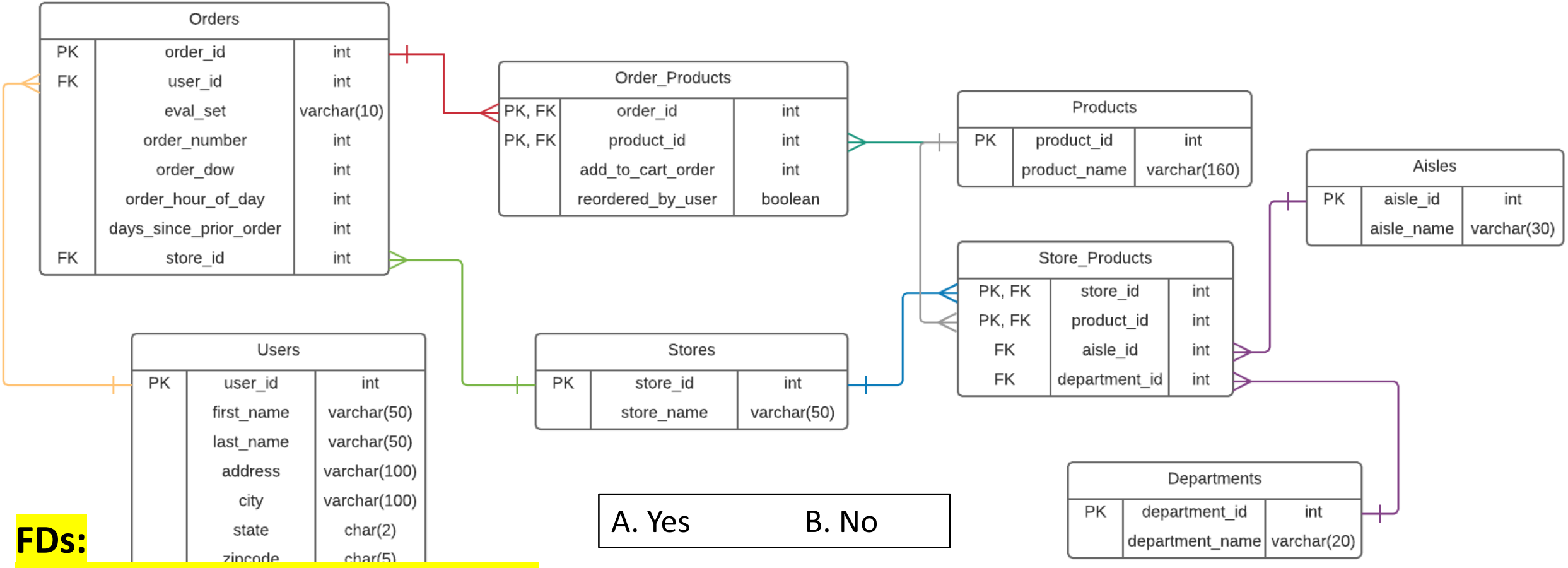
Practice Problem 3: Suppose we discover some new functional dependencies in the data. Do these new FDs cause any violations of 3NF or even 2NF?



FDs:

1. product_id → department_id
2. order_id, prev_order_id → order_number, days_since_prior_order

Practice Problem 3: Suppose we discover some new functional dependencies in the data. Do these new FDs cause any violations of 3NF or even 2NF?

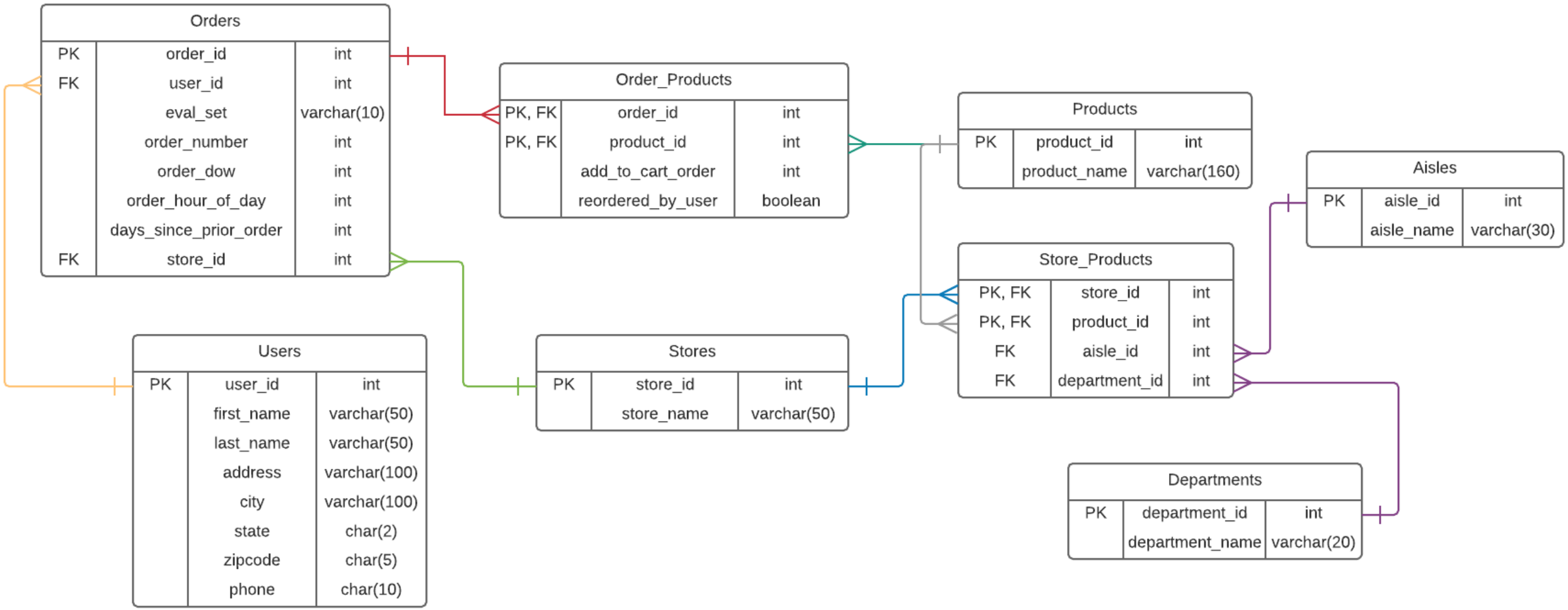


A. Yes B. No

FDs:

1. product_id → department_id
2. order_id, prev_order_id → order_number, days_since_prior_order

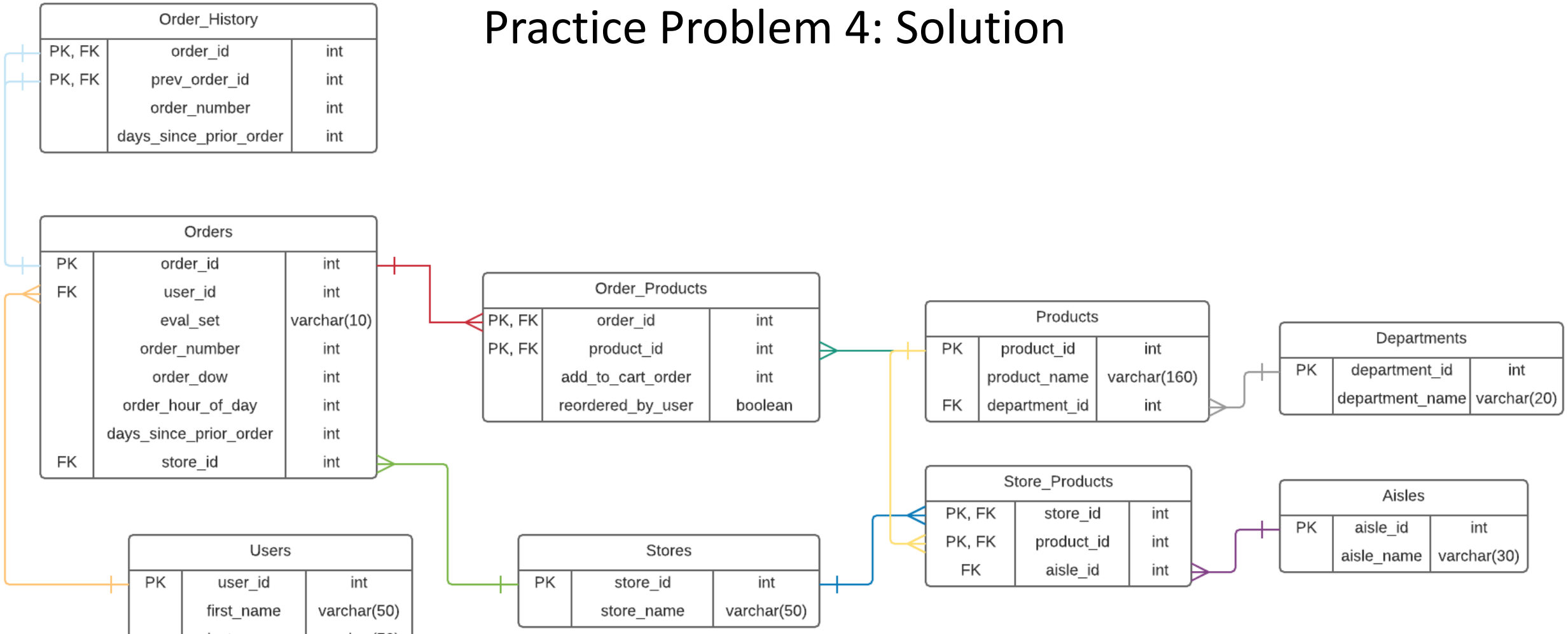
Practice Problem 4: Fix the schema to address the 2NF and 3NF violations.



FDs:

1. product_id → department_id
2. order_id, prev_order_id → order_number, days_since_prior_order

Practice Problem 4: Solution



FDs:

1. product_id → department_id
2. order_id, prev_order_id → order_number, days_since_prior_order