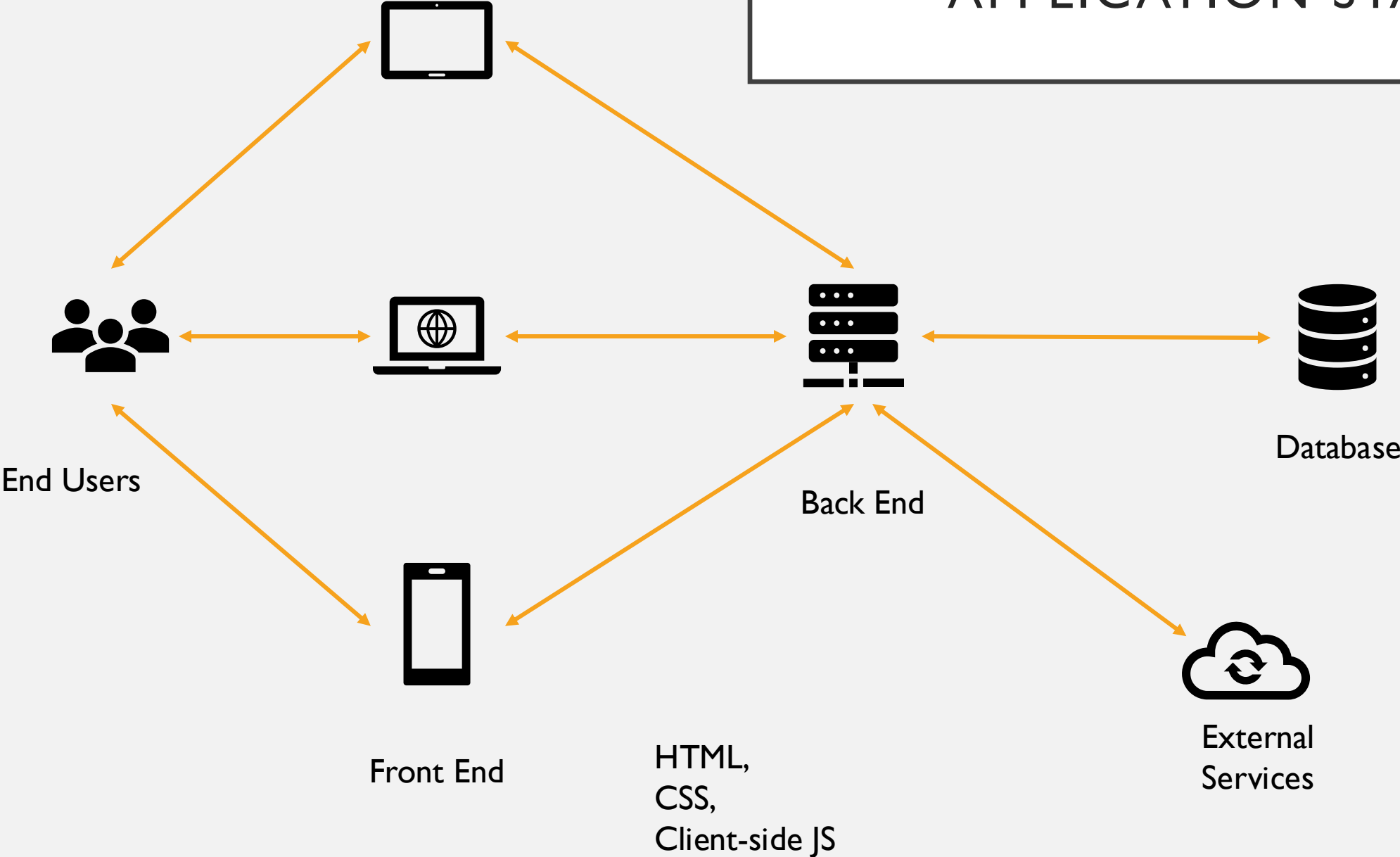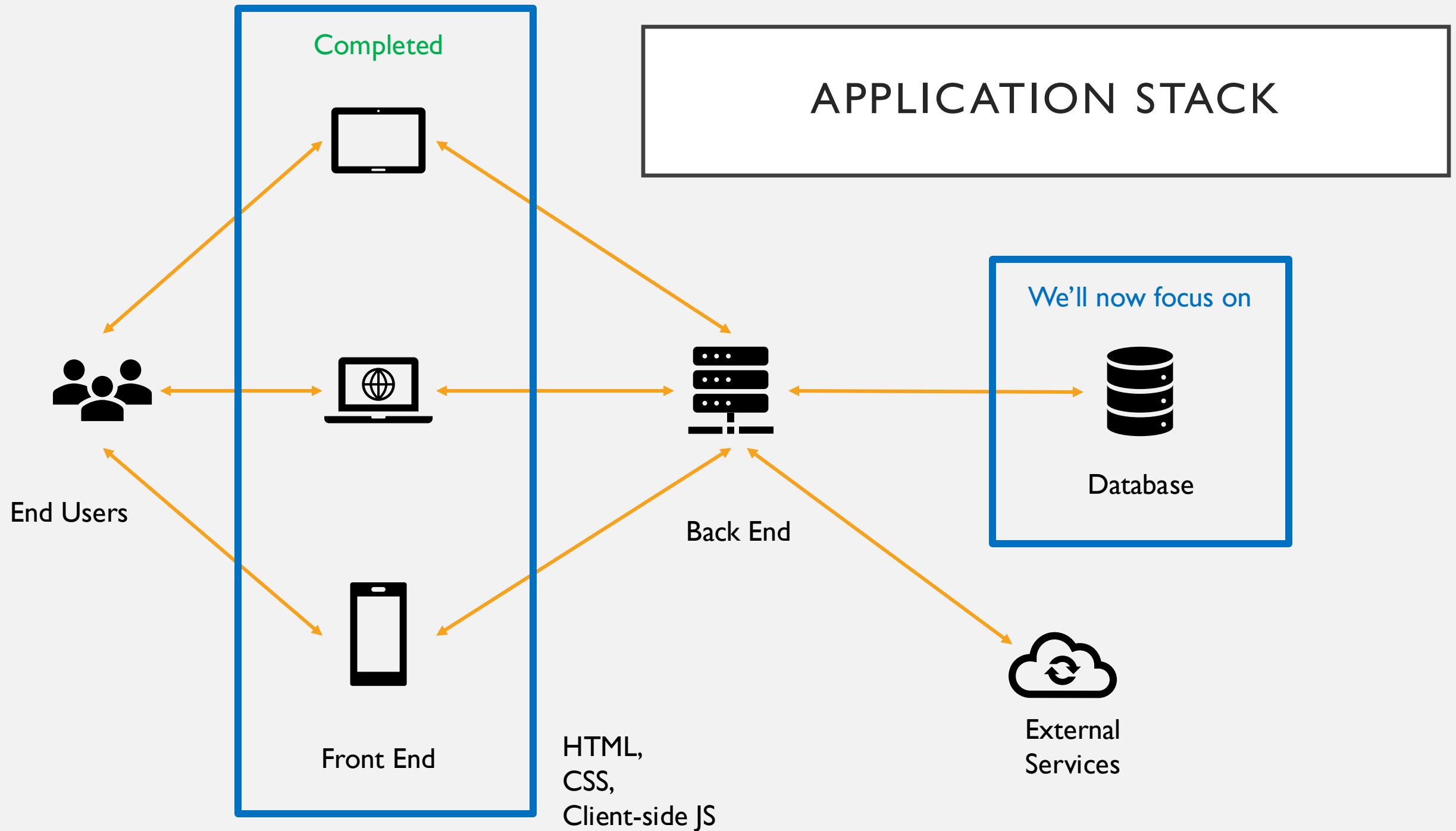# RDBMS

Feb 10, 2025

AGENDA

- Where are we in the tech stack?
- Relational Databases
- ACID Properties
- Designing databases
- Primary Key and Surrogate Keys
- Normalization
- Referential Integrity
- Data Modeling
- Data types in SQL

APPLICATION STACK

End Users

Front End

HTML,
CSS,
Client-side JS

Back End

Database

External
Services

Completed

APPLICATION STACK

We'll now focus on

Database

End Users

Back End

Front End

HTML,
CSS,
Client-side JS

External
Services

RELATIONAL DATABASES

- The purpose of a database is to help people track things of interest to them

- Data is stored in tables, which have rows and columns

- A database may have multiple tables, where each table stores data about a different thing
  - Example: a STUDENT table, a CLASS table

- Each row in a table stores data about one occurrence of the thing of interest
  - Example: one student's data, one class's data

Students

Student ID	Student First Name	Student Last Name	Student Phone	<< other fields >>
60001	Zachary	Erlich	553-3992	....
60002	Susan	McLain	790-3992	....
60003	Joe	Rosales	551-4993	....

Student Schedule (*Linking Table*)

Student ID	Class ID
60003	900001
60001	900003
60003	900003
60002	900002
60001	900001

Classes

Class ID	Class Name	Instructor ID	<< other fields >>
900001	Intro. to Political Science	220087	....
900002	Adv. Music Theory	220039	....
900003	American History	220148	....

ACID

A transaction is a single logical unit of work which accesses and possibly modifies the contents of a database.

- **Atomicity**
  - All or nothing
- **Consistency**
  - Correctness of data
- **Isolation**
  - Transaction occur independently
- **Durability**
  - Transactions are never lost

Account Holder	Balance
Derek	$ 2800 - $300 = $2500
Courtney	$ 2500
Ali	$ 2000 – Didn't receive
Belinda	$ 3000

In this case we should revert the transaction because it is was not completed correctly.

ACID

A transaction is a single logical unit of work which accesses and possibly modifies the contents of a database.

- **Atomicity**
  - All or nothing
- **Consistency**
  - Correctness of data
- **Isolation**
  - Transaction occur independently
- **Durability**
  - Transactions are never lost

Account Holder	Balance
Derek	$ 2800 - $300 = $2500
Courtney	$ 2500
Ali	$ 2000 – Didn't receive
Belinda	$ 3000

Total = $4800

Total = $4500. Where did the remaining money go?

This is data inconsistency introduced by a transaction on the table

ACID

A transaction is a single logical unit of work which accesses and possibly modifies the contents of a database.

- **Atomicity**
  - All or nothing
- **Consistency**
  - Correctness of data
- **Isolation**
  - Transaction occur independently
- **Durability**
  - Transactions are never lost

Account Holder	Balance
Derek	$ 2800 - $300 = $2500
Courtney	$ 2500
Ali	$ 2000 + $300 = $2300
Belinda	$ 3000

If I try to read the content while I'm in the middle of a transaction, the data will be incorrect. Hence, only after all changes have been made to main memory should you read the data.

ACID

A transaction is a single logical unit of work which accesses and possibly modifies the contents of a database.

- **Atomicity**
  - All or nothing
- **Consistency**
  - Correctness of data
- **Isolation**
  - Transaction occur independently
- **Durability**
  - Transactions are never lost

Account Holder	Balance
Derek	$ 2500
Courtney	$ 2500
Ali	$ 2300
Belinda	$ 3000

Later, unless more transfers have taken place on these accounts, this is the data that should be visible to other read/write transactions.

USE CASE DIAGRAM FOR A RECIPE WEBSITE
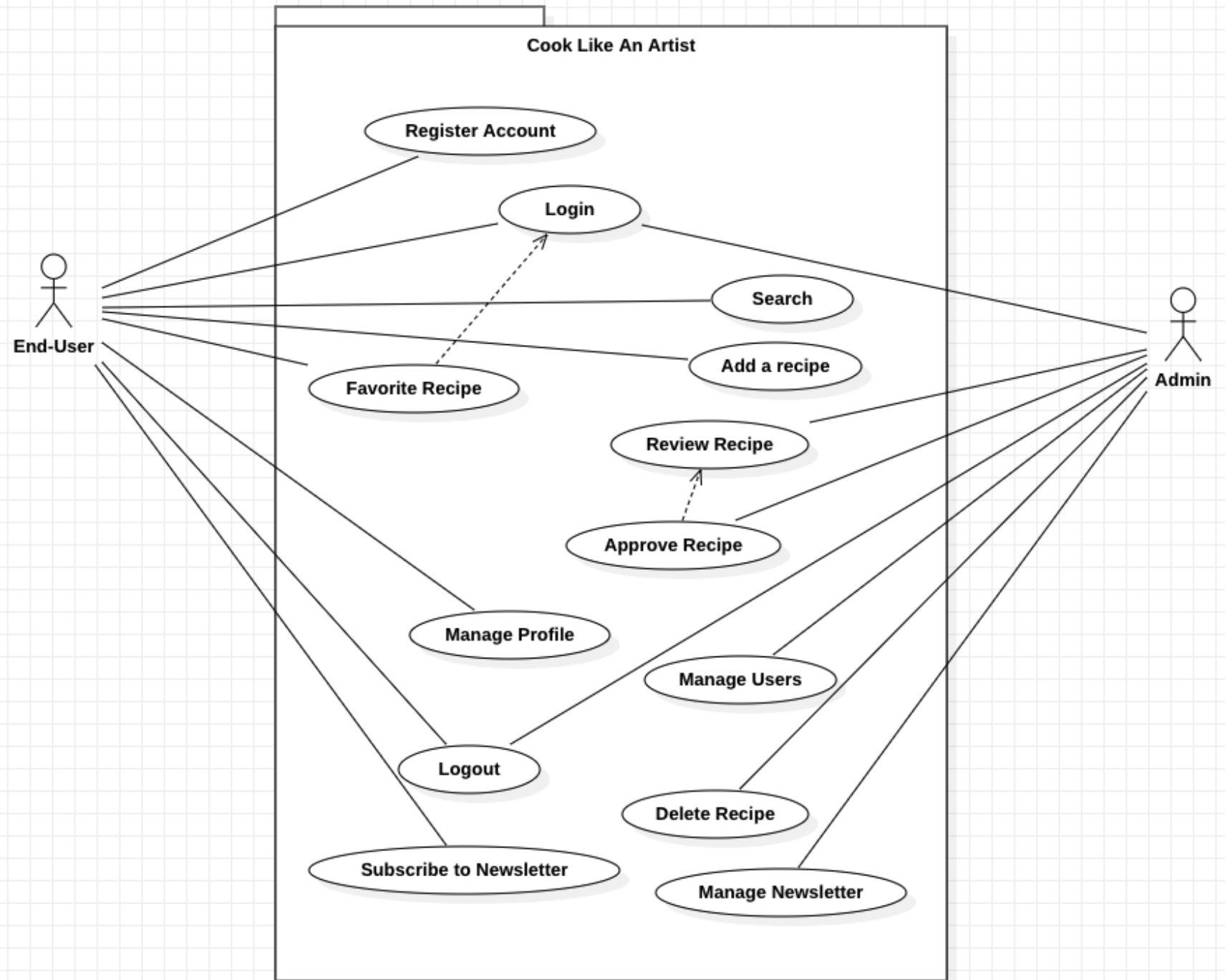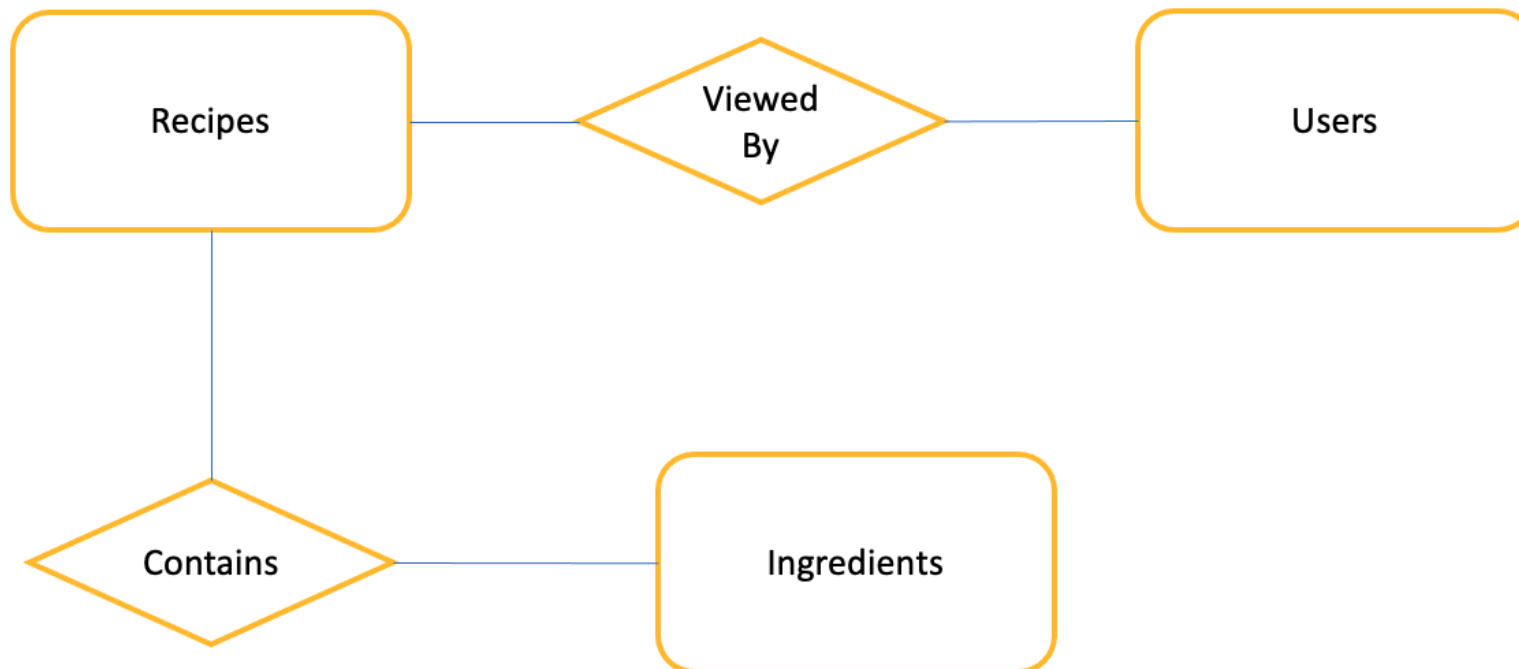
EXAMPLE

- Let's track our recipes:
    - Recipes
    - Users
    - Ingredients

RELATIONAL DATABASES

- A relational database stores BOTH data (in tables) AND relationships (between tables)

DESIGNING A DATABASE

- Data must be normalized

- Construct a Data Model that defines
  - Entities
  - Attributes
  - Relations

ENTITY

- An entity is some identifiable person, place, thing or event that users want to keep track of (that is, store data about)

  - Recipes

  - Users

  - Ingredients

ATTRIBUTE

- An attribute is a FACT or CHARACTERISTIC describing the occurrences of an entity

- Recipes – Recipe Name, Recipe Text/URL, Image, Duration, Author, Likes

- Users – User Id, UserName, Email, Password, isAdmin, favorites

- Ingredients - Ingredient Name, Price

RELATIONS

- A relation is a two-dimensional table that has the following characteristics:
  - Rows contain data about an entity.
  - Columns contain data about attributes of the entity.
  - All entries in a column are of the same kind.
  - Each column has a unique name.
  - One cell of the table holds a single value.
  - The order of the columns is unimportant.
  - The order of the rows is unimportant.
  - No two rows may be identical.
  - Every row has a column that uniquely identifies the row

RELATION?

Recipe Name	Duration	Author	Image	Recipe Link	Ingredients
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	Eggs, cream, butter, salt, pepper
Aglio e Olio	30 min	CookingIsLife	https://abc.com	https://abc_recipe.com	Spaghetti, garlic, olive oil, salt, pepper, parsley, cheese
Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	https://pqr_recipe.com	Eggs, peanut butter, sugar, baking soda

RELATION?

Recipe Name	Duration	Author	Image	Recipe Link	Ingredients	Quantity
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	Eggs	3
					Cream	1 tbsp
					Butter	2 tbsp
					Salt	To taste
					Pepper	A pinch
Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	Mix the 4 ingredients well together and bake for 10-12 mins at 350F	Eggs, peanut butter, sugar, baking soda	

RELATION?

Recipe Name	Duration	Author	Image	Recipe Link
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com
Aglio e Olio	30 min	CookingIsLife	https://abc.com	https://abc_recipe.com
Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	https://pqr_recipe.com

KEYS

- A key is a combination of one or more columns that is used to identify rows in a relation.

- A composite key is a key that consists of two or more columns (also referred to as a concatenated key)

- A primary key is a candidate key selected as the primary means of identifying rows in a relation.

  - There is only one primary key per relation.

  - The primary key may be a composite key.

  - The ideal primary key is short, numeric, and never changes.

PRIMARY KEY?

UserName	Email	Password	isAdmin
CookingIsLife	abc@abc.com	<some encrypted text>	0
BakingGoddess	xyz@abc.com	<some encrypted text>	1

SURROGATE KEYS

- A surrogate key is an artificial column added to a relation to serve as a primary key.

  - DBMS supplied (a sequence number + 1)

  - Short, numeric, and never changes—an ideal primary key

  - Has artificial values that are meaningless to users

SURROGATE KEYS

User Id	UserName	Email	Password	isAdmin
1	CookingIsLife	abc@abc.com	<some encrypted text>	0
2	BakingGoddess	xyz@abc.com	<some encrypted text>	1

WHAT COULD BE A SURROGATE KEY FOR THIS RELATION?

Recipe Id	Recipe Name	Duration	Author	Image	Recipe Link
	Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com
	Aglio e Olio	30 min	CookingIsLife	https://abc.com	https://abc_recipe.com
	Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	https://pqr_recipe.com

FOREIGN KEYS

- A foreign key is the primary key of one relation that is placed in another relation to form a link between the relations.
  - A foreign key can be a single column or a composite key.
  - The term refers to the fact that key values are not primary to the relation in which they appear as foreign key values.

UserName	Email	Password	isAdmin
CookingIsLife	abc@abc.com	<some encrypted text>	0
BakingGoddess	xyz@abc.com	<some encrypted text>	1

Foreign Key

Recipe Name	Duration	Author	Image	Recipe Link
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com
Aglio e Olio	30 min	CookingIsLife	https://abc.com	https://abc_recipe.com
Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	https://pqr_recipe.com

REFERENTIAL INTEGRITY CONSTRAINT

- A referential integrity constraint is a rule that limits the values of the foreign key to those already existing as primary key values in the corresponding relation.

- In other words, the constraint keeps me from adding a row to a table if the value in a foreign key column is "not on file"

UserName	Email	Password	isAdmin
CookingIsLife	abc@abc.com	<some encrypted text>	0
BakingGoddess	xyz@abc.com	<some encrypted text>	1

Recipe Name	Duration	Author	Image	Recipe Link
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com
Aglio e Olio	30 min	CookingIsLife	https://abc.com	https://abc_recipe.com
Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	https://pqr_recipe.com
Shepherd's pie	70 min	WholesomeDish	https://pie.com	https://pie_recipe.com

NORMALIZATION

- A structured, defined, detailed process to arrange the data into a series of clearly defined relations:
  - Each with a primary key
  - All attributes are fully dependent on the primary key

"The data depends on the key (1NF), the whole key (2NF) and nothing but the key (3NF), so help me Codd."

NORMALIZATION STEP-BY-STEP

- **First Normal Form**
  - Remove any multi-valued cells and/or any rows requiring a specific sequence
- **Second Normal Form**
  - For entities with composite keys, make sure that all attributes are dependent on the full key
- **Third Normal Form**
  - Make sure that no attributes are dependent on any other non-key attributes

1NF?

Recipe Name	Duration	Author	Image	Recipe Link	Ingredients
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	Eggs, cream, butter, salt, pepper
Aglio e Olio	30 min	CookingIsLife	https://abc.com	https://abc_recipe.com	Spaghetti, garlic, olive oil, salt, pepper, parsley, cheese
Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	https://pqr_recipe.com	Eggs, peanut butter, sugar, baking soda

NOW THE TABLE IS IN 1NF

Split the multi-valued cells into multiple rows. Here we split the ingredients for one recipe as an example here.

Recipe Name	Duration	Author	Image	Recipe Link	Ingredients
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	Eggs
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	cream
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	butter
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	salt
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	pepper

Primary Key(Recipe Name, Author)

2NF?

However, Email Id depends only on author. Not on Recipe Name. Hence it needs to stored in a different table.

Recipe Name	Duration	Author	Image	Recipe Link	Email Id
Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	abc@abc.com
Aglio e Olio	30 min	CookingIsLife	https://abc.com	https://abc_recipe.com	abc@abc.com
Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	https://pqr_recipe.com	xyz@abc.com
Scrambled Eggs	10 min	BakingGoddess	https://lmn.com	https://lmn_recipe.com	xyz@abc.com

Primary Key(Recipe Id)

3NF?

Email Id depends on author. Author is a non-key attribute in this table

| Recipe Id	Recipe Name	Duration	Author	Image	Recipe Link	Email Id
1	Scrambled Eggs	10 min	CookingIsLife	https://xyz.com	https://xyz_recipe.com	abc@abc.com
2	Aglio e Olio	30 min	CookingIsLife	https://abc.com	https://abc_recipe.com	abc@abc.com
3	Peanut butter cookies	22 min	BakingGoddess	https://pqr.com	https://pqr_recipe.com	xyz@abc.com

DATA MODELING

Entity (rectangle) with an entity name, Primary Key, and Attributes listed

Cardinality & Optionality Symbols
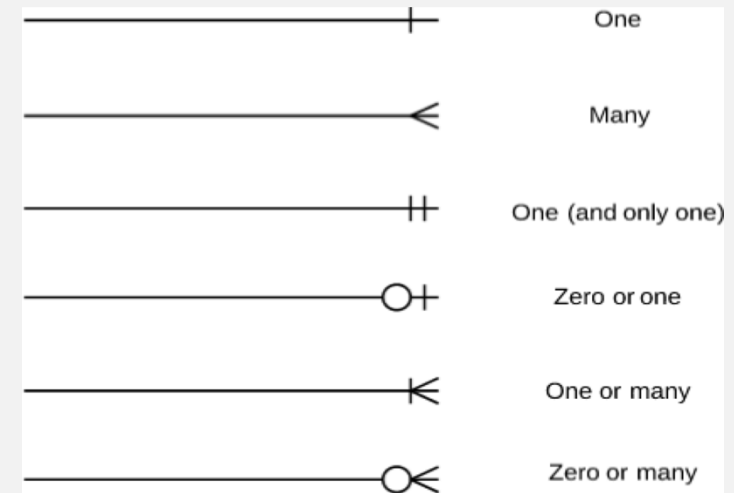
DATA MODELING
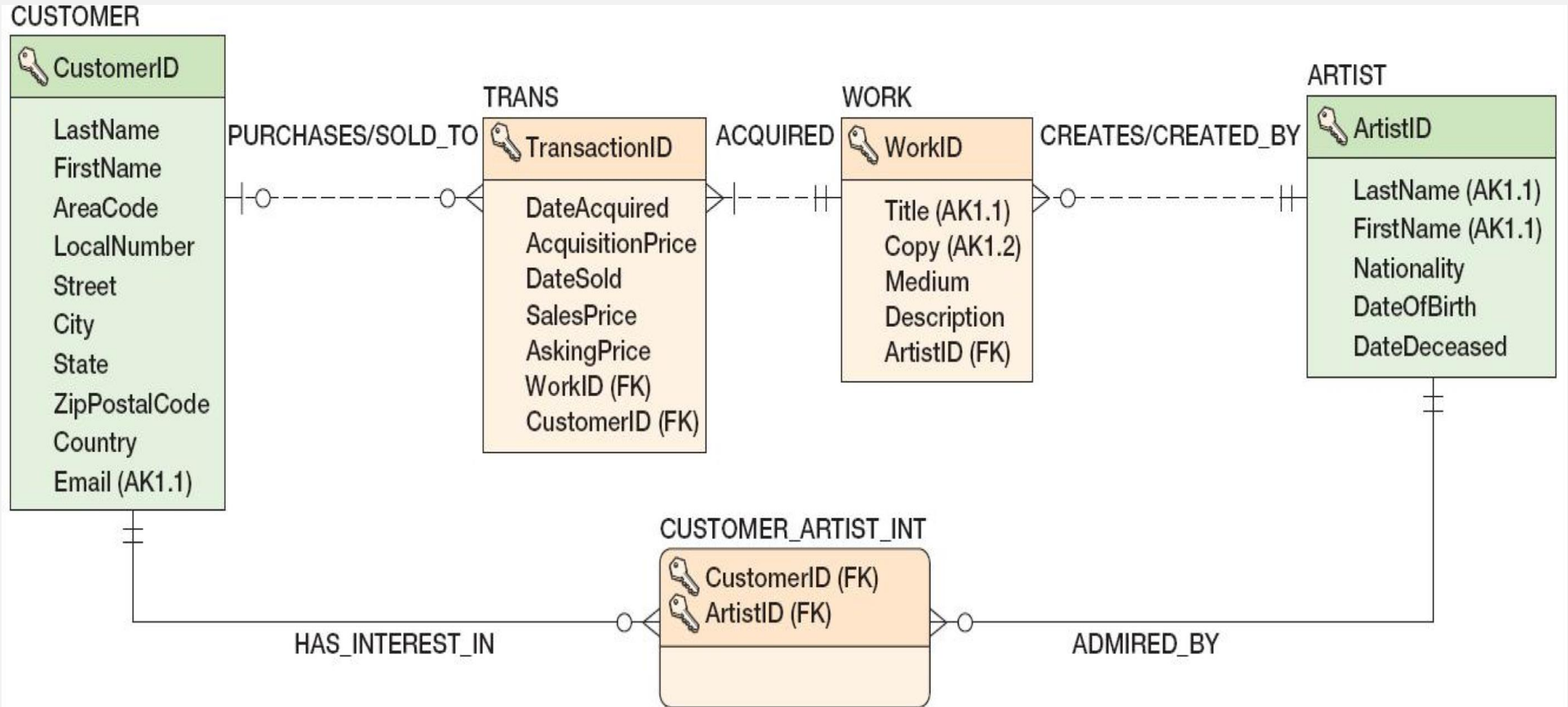
DATA MODELING

- What are the names of the five entities?

- What is the primary key of each entity?

- Why are some of the relationship lines dashed, and some are solid?

- Why do 4 of the entities have square corners and one has rounded corners?

- Which entity has a composite (or "concatenated") key?

- Relationship descriptions are read clockwise: a customer purchases a work; a work is sold to a customer.

  - Which entity represents the fact that a customer purchased a work?

DEFINING DATA

- Null status indicates whether the value of the column can be NULL.

EMPLOYEE

🔑 EmployeeNumber: NOT NULL

EmployeeName: NOT NULL
Phone: NULL
Email: NULL (AK1.1)
HireDate: NOT NULL
ReviewDate: NULL
EmpCode: NULL

SPECIFY COLUMN PROPERTIES: DATA TYPE

- Generic data types:
  - CHAR(n)
  - VARCHAR(n)
  - DATE
  - TIME
  - MONEY
  - INTEGER
  - DECIMAL

EMPLOYEE

🔑 EmployeeNumber: int

EmployeeName: char(50)
Phone: char(15)
Email: char(50) (AK1.1)
HireDate: datetime
ReviewDate: datetime
EmpCode: char(18)

SQL AND NOSQL DATABASES