Announcements

- Homework 1 due Oct. 11th (this Friday!)
- Project Milestone 2 due Oct. 18th
- Homework 2 open Oct. 11th, due Oct. 25th

Announcements

- Exam on Oct 16th

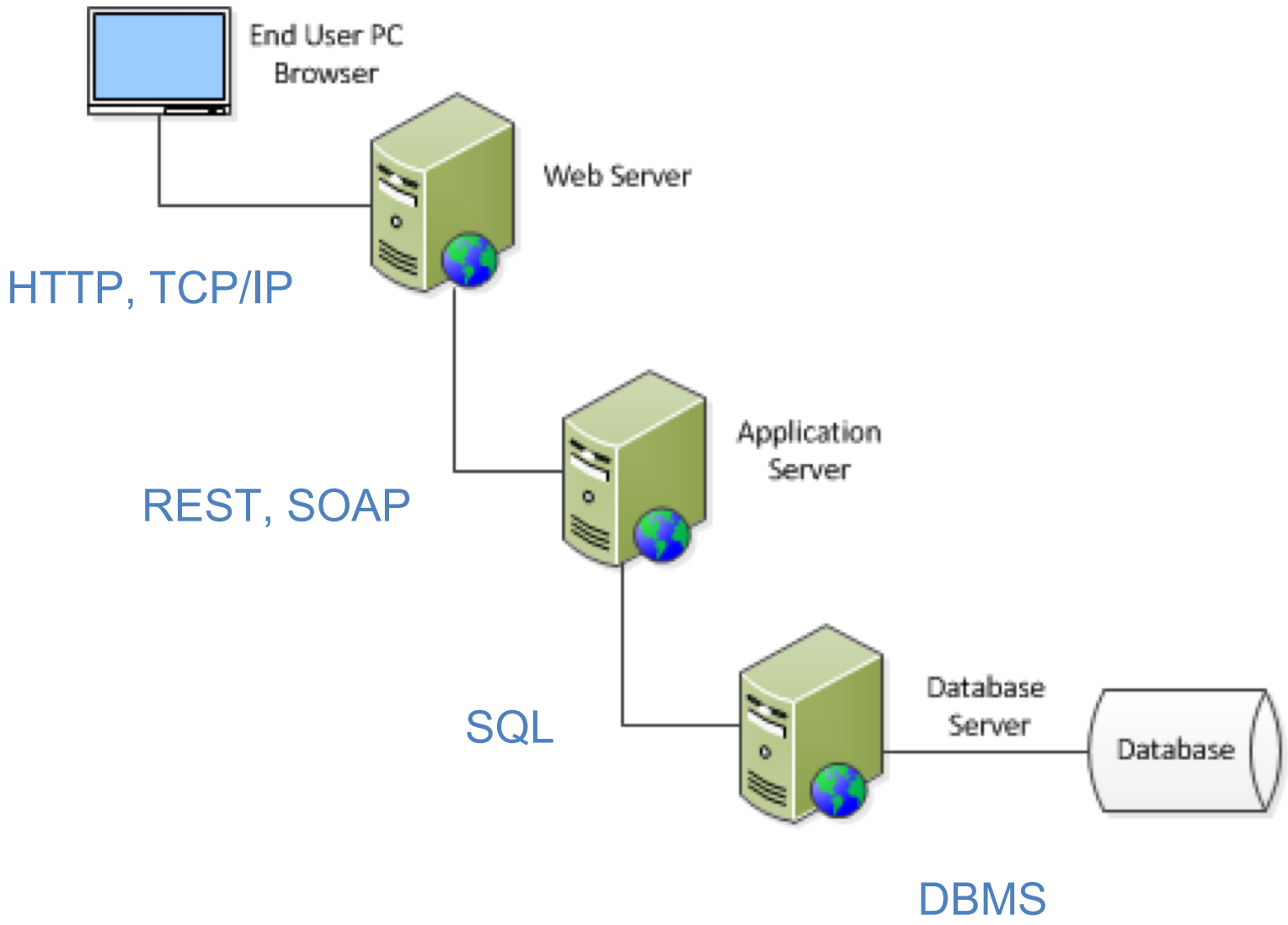
Applications, DBMS, and SQL

Applications are computer programs that users interact with.

Database Management System (DBMS) refers to software that allows us to create databases, load data into databases, retrieve data from databases.

Structured Query Language (SQL) is an internationally recognized standard database language that is used by commercial DBMSs.

Application Architecture



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

Football Example

Let’s track our football team:

- Players
- Games
- …

Who played in which game?

Football Spreadsheet

Name	Year	Major	Games	Score	Date
Bob Smith	Senior	Math	Arizona State	28-21	08-13-2018
			UCLA	34-42	10-17-2018
John Billings	Junior	Journalism	Arizona State	28-21	08-13-2018
			Utah	7-21	09-23-2019

Games	Score	Date	Player	Year	Major
Arizona State	28-21	08-13-2018	Bob Smith	Senior	Math
			John Billings	Junior	Journalism
UCLA	34-42	10-17-2018	Bob Smith	Senior	Math
			John Billings	Junior	Journalism
			Bob Smith	Senior	Math
Utah	7-21	09-23-2019	John Billings	Junior	Journalism

Football Spreadsheet

Name	Year	Major	Games	Score	Date
Bob Smith	Senior	Math	Arizona State	28-21	08-13-2018
			UCLA	34-42	10-17-2018
John Billings	Junior	Journalism	Arizona State	28-21	08-13-2018
			Utah	7-21	09-23-2019

Games	Score	Date	Player	Year	Major
Arizona State	28-21	08-13-2018	Bob Smith	Senior	Math
			John Billings	Junior	Journalism
UCLA	34-42	10-17-2018	Bob Smith	Senior	Math
			John Billings	Junior	Journalism
			Bob Smith	Senior	Math
Utah	7-21	09-23-2019	John Billings	Junior	Journalism

Football Spreadsheet

Name	Year	Major	Games	Score	Date
Bob Smith	Senior	Math	Arizona State	28-21	08-13-2018
			UCLA	34-42	10-17-2018
John Billings	Junior	Journalism	Arizona State	28-21	08-13-2018
			Utah	7-21	09-23-2019

Games	Score	Date	Player	Year	Major
Arizona State	28-21	08-13-2018	Bob Smith	Senior	Math
			John Billings	Junior	Journalism
UCLA	34-42	10-17-2018	Bob Smith	Senior	Math
			John Billings	Junior	Journalism
			Bob Smith	Senior	Math
Utah	7-21	09-23-2019	John Billings	Junior	Journalism

Relational Databases

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

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

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)
  - Players
  - Games

Attribute

- An attribute is a FACT or CHARACTERISTIC describing the occurrences of an entity
  - Player first name, last name, year, major, etc.
  - Game date, score, etc.

Relation

- 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

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.

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

Relation?

Name	Year	Major	Phone
Bob Smith	Senior	Math	303-717-8888, 720-840-9000, 970-666-5555
John Billings	Junior	Journalism	303-455-6666, 1-800-800-8000
Edward James	Junior	Physics	303-123-4567
Bob Smith	Junior	Math	303-303-3003

Relation?

Name	Year	Major	Phone
Bob Smith	Senior	Math	303-717-8888, 720-840-9000, 970-666-5555
John Billings	Junior	Journalism	303-455-6666, 1-800-800-8000
Edward James	Junior	Physics	303-123-4567
Bob Smith	Junior	Math	303-303-3003

Relation?

Player Id	Name	Year	Major	Email	Phone
1	Bob Smith	Senior	Math	bob@colorado.edu	
				Home phone:	720-840-9000
				Mobile phone:	970-666-5555
4	John Billings	Junior	Journalism	john@colorado.edu	303-455-6666

Relation?

Player Id	Name	Year	Major	Email	Phone
1	Bob Smith	Senior	Math	bob@colorado.edu	
				Home phone:	720-840-9000
				Mobile phone:	970-666-5555
4	John Billings	Junior	Journalism	john@colorado.edu	303-455-6666

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

- 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

Football Player Relation

Player Id	Name	Year	Major	Home phone	Cell phone
1	Bob Smith	Senior	Math	123-1234-12345	970-721-9000
2	John Billings	Junior	Journalism	555-5555-5555	303-444-7777
3	Edward James	Junior	Physics	450-888-9999	720-123-4500
4	Will Miller	Junior	Math	303-310-8756	720-909-9000

Football Relations

Player Id	Name	Year	Major	Home phone	Cell phone
1	Bob Smith	Senior	Math	123-1234-12345	970-721-9000
2	John Billings	Junior	Journalism	555-5555-5555	303-444-7777
3	Edward James	Junior	Physics	450-888-9999	720-123-4500
4	Will Miller	Junior	Math	303-310-8756	720-909-9000

Visitor	Home Score	Visitor Score	Date
Arizona State	42	34	08-21-2018
Utah	7	21	09-12-2019

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.

Football Relations

Player Id	Name	Year	Major	Home phone	Cell phone
1	Bob Smith	Senior	Math	123-1234-12345	970-721-9000
2	John Billings	Junior	Journalism	555-5555-5555	303-444-7777
3	Edward James	Junior	Physics	450-888-9999	720-123-4500
4	Will Miller	Junior	Math	303-310-8756	720-909-9000

Visitor	Home Score	Visitor Score	Date	Players
Arizona State	42	34	08-21-2018	[1, 2, 3, 4]
Utah	7	21	09-12-2019	[2, 3]

The 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.

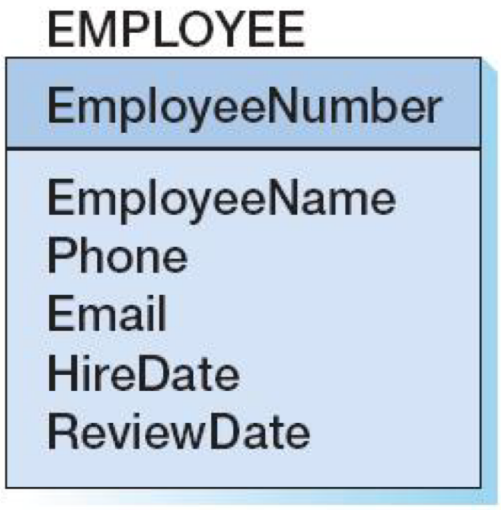
- In other words, the constraint keeps me from adding a row to a table if the value in a foreign key column is “not on file”

Football Relations

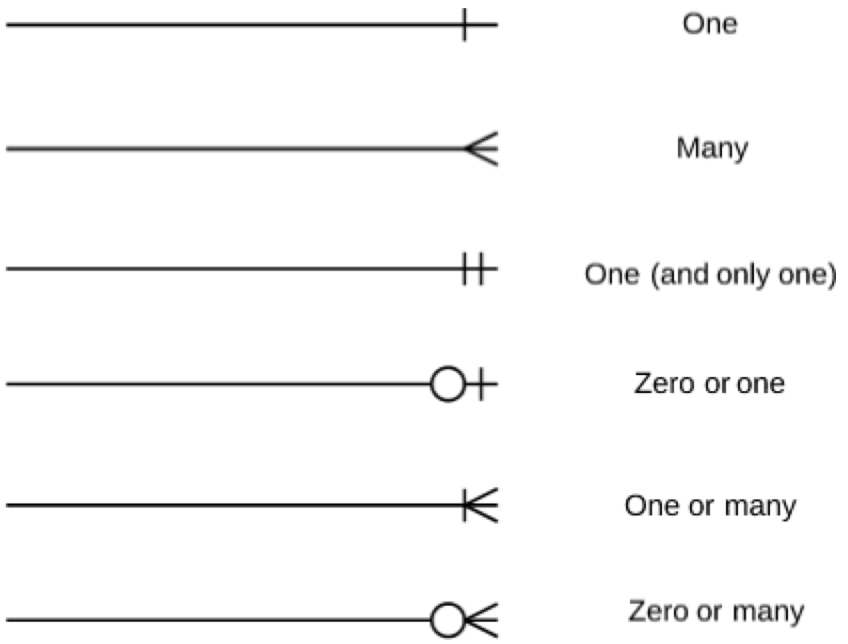
Player Id	Name	Year	Major	Home phone	Cell phone
1	Bob Smith	Senior	Math	123-1234-12345	970-721-9000
2	John Billings	Junior	Journalism	555-5555-5555	303-444-7777
3	Edward James	Junior	Physics	450-888-9999	720-123-4500
4	Will Miller	Junior	Math	303-310-8756	720-909-9000

Visitor	Home Score	Visitor Score	Date	Players
Arizona State	42	34	08-21-2018	[1, 2, 3, 4]
Utah	7	21	09-12-2019	[2, 3]
UCLA	0	21	09-13-2019	[2, 3, 4, 5]

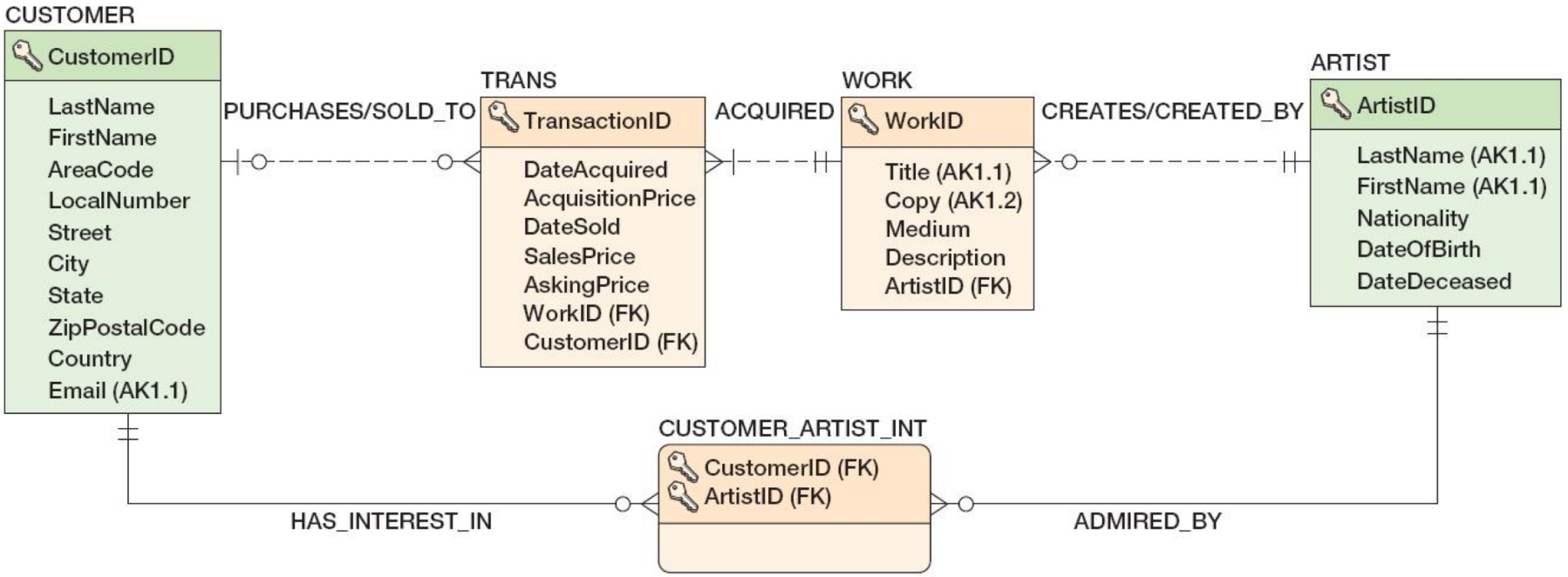
Entity (rectangle) with an entity name, Primary Key, and Attributes listed



Cardinality & Optionality Symbols



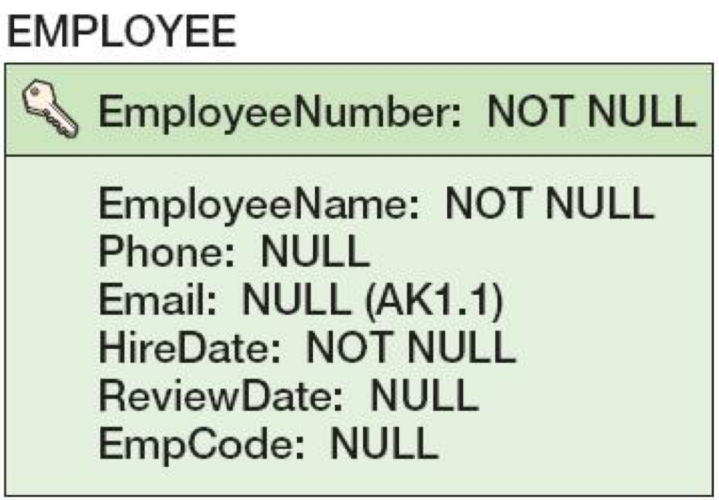
Data Modeling



Data Modeling

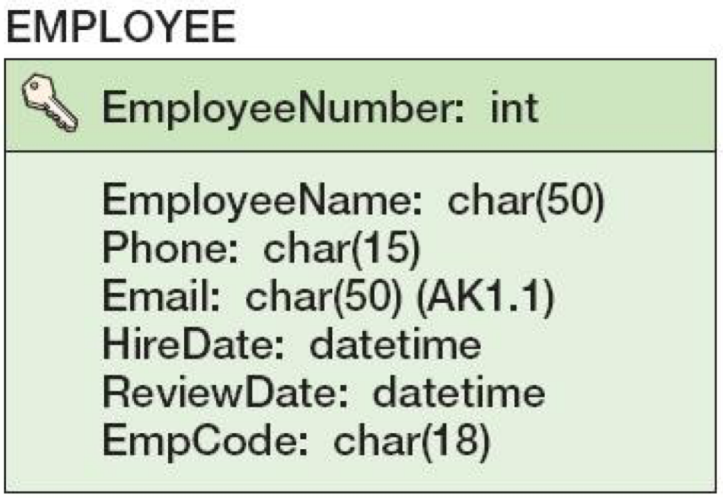
- What are the names of the five entities?
- What is the primary key of each entity?
- Why are some of the relationship lines dashed, and some are solid?
- Why do 4 of the entities have square corners and one has rounded corners?
- Which entity has a composite (or “concatenated”) key?
- Relationship 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?

- **Null status** indicates whether or not the value of the column can be NULL.

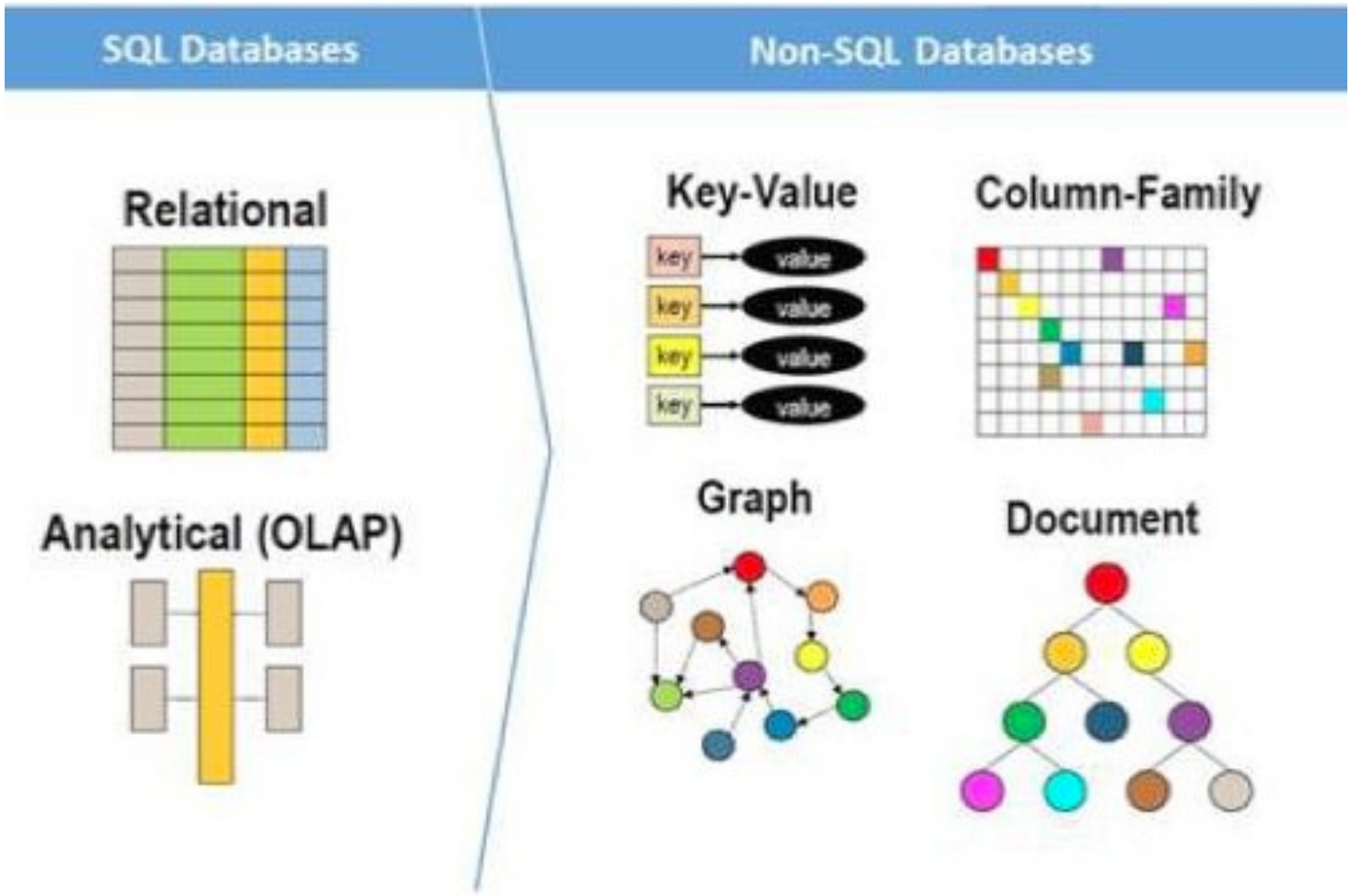


Specify Column Properties: Data Type

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



SQL and No SQL



ToDo for Wednesday

1. Setup PostgreSQL
   - https://postgresapp.com/
   - https://sreeshanath.github.io/lab-instructions/lab1_initial_setup/index.html

2. PostgreSQL GUI Tool
   - https://www.jetbrains.com/datagrip/ (free student license available)