Intro to Databases

Week 02- Day 04


statistics
python
phd
sql
r
hadoop
java
ms
hive
math
matlab
spark

Databases

=

Software to store and access data

Why not CSV or Excel?!

How to make money with a CSV database

09:11:10:07 - Transaction1 (edo, julie, 100)

09:11:10:08 - Transaction2 (julie,edo, 200)

(t1) Read edo account - 1000$

(t1) Read julie account - 2000$

(t2) Read edo account - 1000$

(t2) Read julie account - 2000$

(t1) Write edo-100 = 900, Write julie+100 = 1100

(t2) Write edo+200 = 1200, Write julie-200 = 800

Complex to manage multiple accesses

Slow

Memory limits

Why DBs

Fast / Optimized

Easily manage multiple accesses

What do DBs really do?

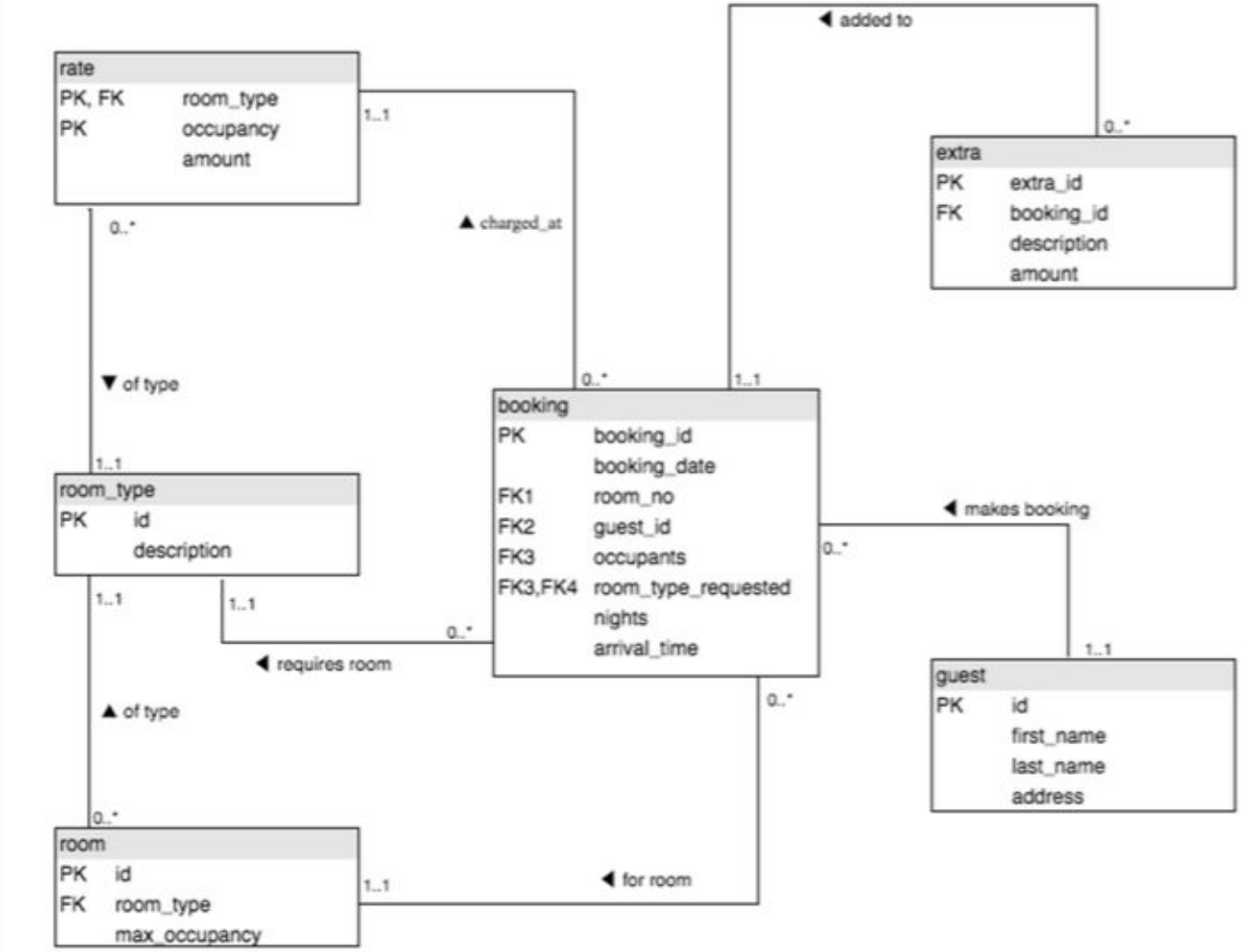
Store data

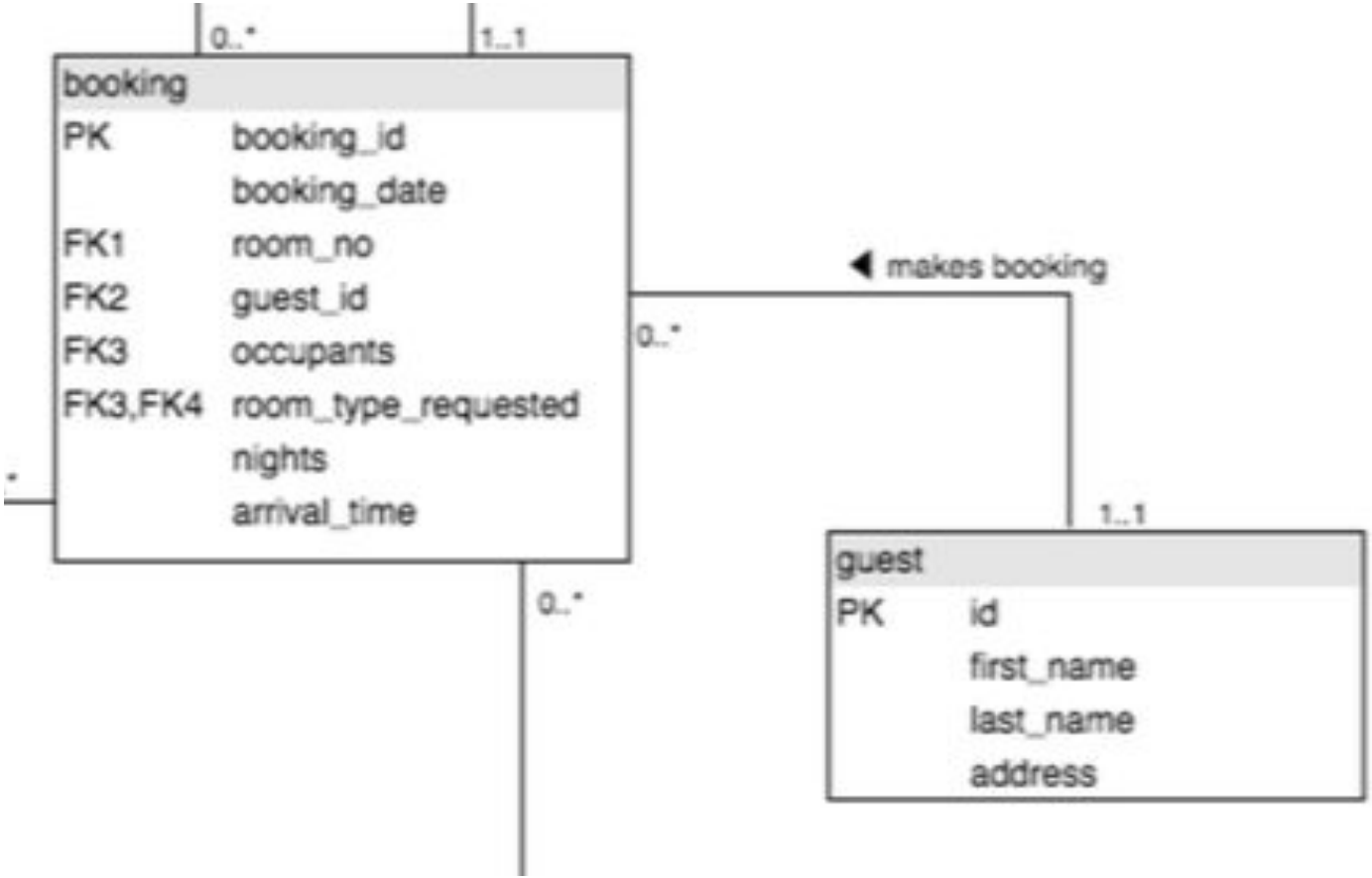
Manage structures to keep the data clean and organized

Keep your data in a clean state

(transactions)

How are DBs organized?


rate
PK, FK room_type
PK occupancy
amount
1..1
charged_at
0..*
added to
0..*
extra
PK extra_id
FK booking_id
description
amount
of type
0..*
1..1
booking
PK booking_id
booking_date
FK1 room_no
FK2 guest_id
FK3 occupants
FK3,FK4 room_type_requested
nights
arrival_time
1..1
room_type
PK id
description
makes booking
0..*
1..1
1..1
0..*
requires room
guest
PK id
first_name
last_name
address
of type
0..*
0..*
room
PK id
FK room_type
max_occupancy
1..1
for room


0..*
1..1
booking
PK booking_id
booking_date
FK1 room_no
FK2 guest_id
FK3 occupants
FK3,FK4 room_type_requested
nights
arrival_time
makes booking
0..*
1..1
0..*
guest
PK id
first_name
last_name
address

Just boring tables!

Voter ID	First Name	Last Name	Turnout Score
1000001	Matt	Brems	0.96
1000002	Sam	Stack	0.43
1000003	Joseph	Nelson	N/A

Primary Key - Unique!

Voter ID	First Name	Last Name	Turnout Score
1000001	Matt	Brems	0.96
1000002	Sam	Stack	0.43
1000003	Joseph	Nelson	N/A

Tables are connected with (external) keys

Exercise: Build the Uber database!

- User ID
- User Name
- Driver ID
- Driver Name
- Ride ID
- Ride Time
- Pickup Longitude
- Pickup Latitude
- Pickup Location
- Drop-Off Location
- Drop-Off Latitude
- Miles
- Travel Time
- Fare
- CC Number

How to build a Database

Step 1 - build an empty table

Step 2 - start populating the table

```
CREATE TABLE Persons (
    PersonID int,
    LastName varchar(255),
    FirstName varchar(255),
    Address varchar(255),
    City varchar(255)
);
```

How to access data in (most of) databases?

```
SELECT name, age
FROM customers
WHERE age>30
```

```
SELECT *
FROM passengers
WHERE destination="sg"
AND departure>"2018-06-12"
```

Different types of databases

RDBMS vs. NoSQL

(key, value) stores

(af98sa: Object1, oieu36: Object2)

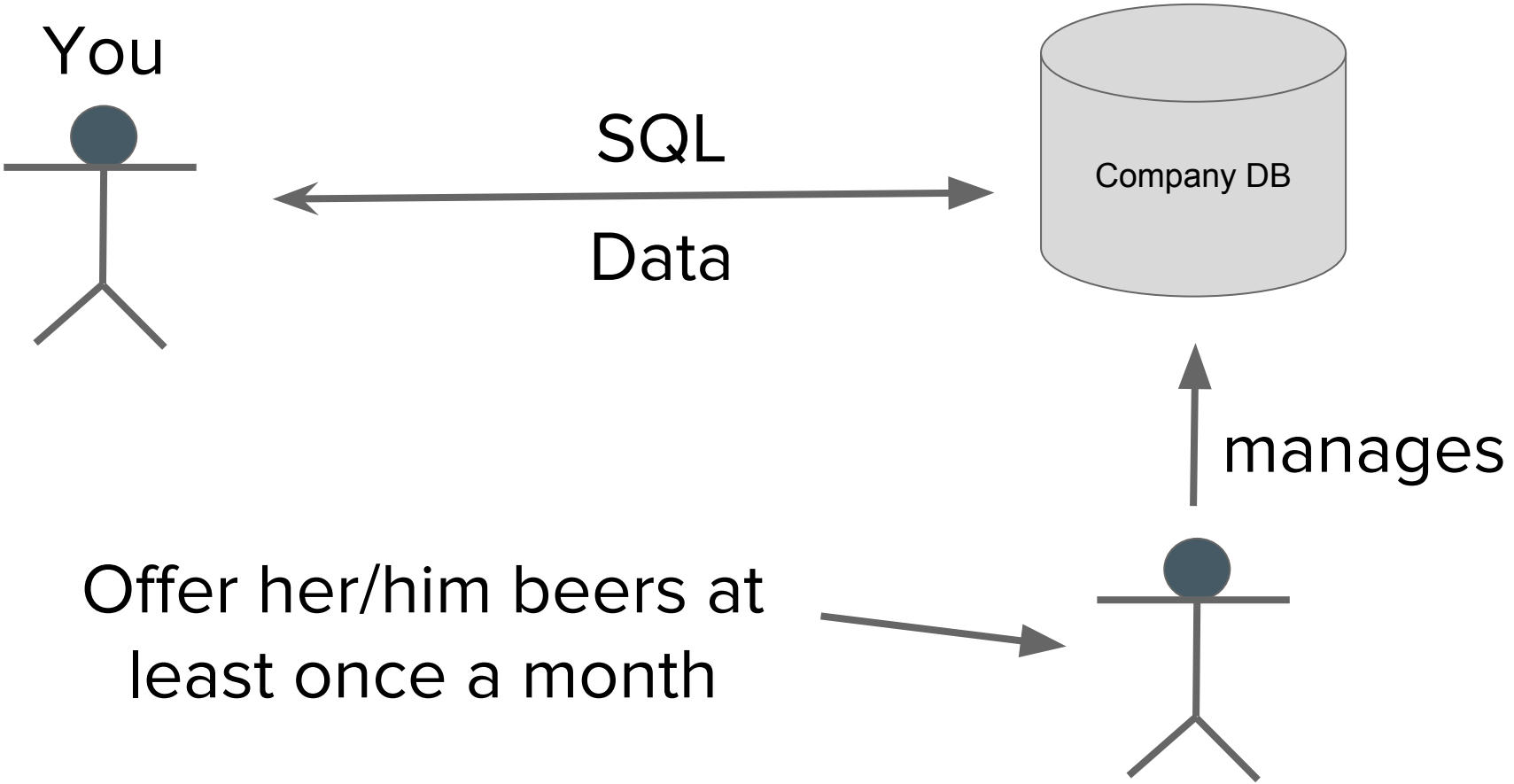
Document databases

{‘name’:’anna’, ‘age’:23, ‘gender’:’f’}

Graph databases

(Twitter, Facebook, etc.)

Real case scenario


You
SQL
Data
Company DB
manages
Offer her/him beers at least once a month