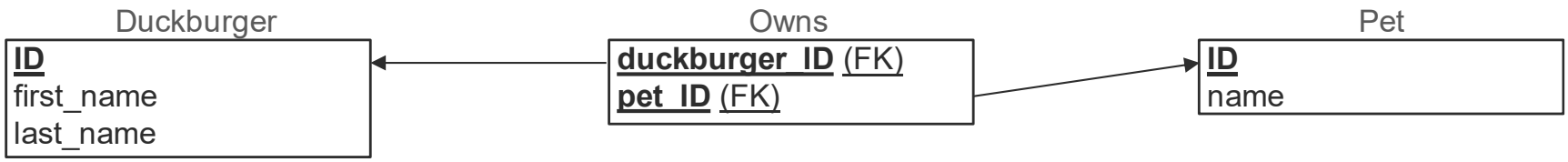
Aggregate Queries

- distinct
- group by
- having
- count, min, max..

Relational Model



Aggregate Queries

- There are two ways of removing duplicates from a result set
  - distinct
  - group by (many uses)

Aggregate Queries

- Let's do a query that prints out the last names of all duckburgers:

```
MariaDB [duckburg]> select last_name from
duckburger;
+-----------+
| last_name |
+-----------+
| Duck      |
| McDuck    |
| Duck      |
| De Spell  |
| Mouse     |
+-----------+
```

- We notice that the name "Duck" appears on the list more than once.

distinct

- Let's create a query that again lists the last names of all duckburgers, but each unique name is only printed out once.

```
MariaDB [duckburg]> select distinct last_name from duckburger;
+-----------+
| last_name |
+-----------+
| Duck      |
| McDuck    |
| De Spell  |
| Mouse     |
+-----------+
```

group by

- Let's write a query that lists each unique last_name once, but this time using grouping.

```
MariaDB [duckburg]> select last_name
    -> from duckburger
    -> group by last_name;
+-----------+
| last_name |
+-----------+
| De Spell  |
| Duck      |
| McDuck    |
| Mouse     |
+-----------+
```

- Here the result is the same as in the previous distinct example.
- More versatile uses for the group by statement are introduced later in these slides.

count

- Let's write a query that prints out the pet_id and counts how many owners each pet has.

```
MariaDB [duckburg]> select pet_id, count(*)
    -> from owns
    -> group by pet_id;
+--------+----------+
| pet_id | count(*) |
+--------+----------+
|      1 |        2 |
|      2 |        1 |
|      3 |        1 |
+--------+----------+
```

- For this we need group by from the last slide
- New skill: count(*)

having count(*)

- Let's make a query that lists the id of all pets with at least two owners.

```
MariaDB [duckburg]> select pet_id, count(*)
    -> from owns
    -> group by pet_id
    -> having count(*) >= 2;
+--------+----------+
| pet_id | count(*) |
+--------+----------+
|      1 |        2 |
+--------+----------+
```

- We need the group by statement again.
- New skill:
  - having count(*)
  - Usually needed when the condition is expressed with "at least" or "not more than"

min, max, avg, sum

- Let's query for the name of the pet with the largest pet id.

```
MariaDB [duckburg]> select name
    -> from pet
    -> where id in (
    -> select max(id)
    -> from pet
    -> );
+---------+
| name    |
+---------+
| Ratface |
+---------+
```

- **Notice that we must do this using a subquery**. Otherwise we would be able to find the largest id, but it would have no correlation with the pet name. (A common mistake).