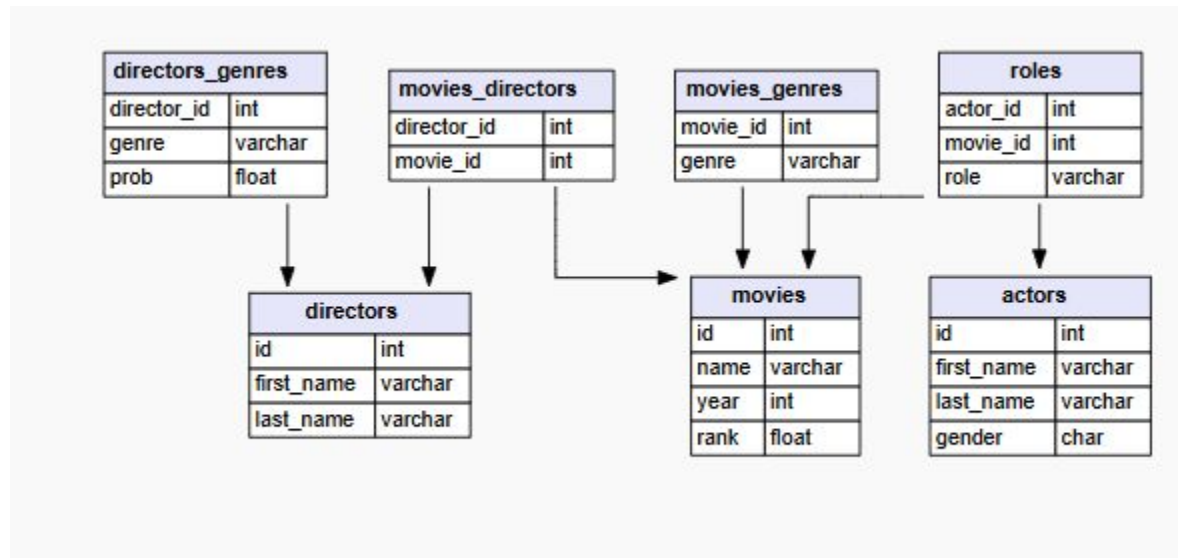


# Databases for data analytics

<https://github.com/evidencebp/databases-course/>

Continue with DML

# When we want more than a single table



# Joining tables - It is set theory down below

- Cartesian product
- Joins help to “walk step by step” on keys
- Joins help to find missing records when used with null

# Data modification - table creation

# Creating a new table based on another

DROP TABLE IF EXISTS imdb\_ijs.tmp\_movies; # Drop is important for reruns

CREATE TABLE imdb\_ijs.tmp\_movies

as

select \*

from

imdb\_ijs.movies

;

# Data modification - update

```
update imdb_ijs.tmp_movies
```

```
set year = year + 1
```

```
where
```

```
id >= 0 # To bypass safe update mode - beware!
```

```
;
```

# In class exercises

- All movies with a Superman role
- Pair of movies directed by the same director (use Hitchcock to reduce volume)

# Exercises

- All Hitchcock's movies with an 'Assassin' role (consider a movie with two of them as a corner case).
- All movies with a 'detective', 'murderer' roles (case insensitive)
- All actors that played with Marilyn Monroe
- Copy directors to directors\_tmp and delete all those that did not direct any movie