

Folie 5 (Basics: Getting to know the MovieDB):

How many tables do we have? (19)

How large are our tables?

actors2awards (107741) movies2grossopeningweekend (11309)

actors2movies (1048575) movies2grossworldwide (19031)

awards (268559) persons (1391353)

awards2movies (268559) producers2awards (54645)

directors2awards (83248) producers2movies (184692)

directors2movies (33606) ratings (30625)

genres (82073) runtimes (40747)

locations (133400) writers2awards (85928)

movies (30625) writers2movies (71148)

movies2budget (13675)

Which is the largest?

(1. persons 1.4 Mio., 2. actors2movies 1 Mio., 3. awards2movies 105k)

Which schema does it follow?

.schema persons

Output:

```
CREATE TABLE persons (personid text PRIMARY KEY, name text NOT NULL);
```

Folie 6 (Finding S. Kubrick, a step by step guide):

--- Where should he be/Which table(s) do we need? (persons)

--- What is his personid? (nm0000040)

select personid from persons where name like '%Kubrick%';

More than one?:

select name, personid from persons where name like '%Kubrick%';

BONUS: Which of these directors produced the most movies?

select name, personid, count(*) as cnt from persons natural join directors2movies where name like '%Kubrick%' order by cnt;

Why is that wrong? (group by personid MISSING)

Example:

select name, personid, count(*) as cnt from persons natural join directors2movies where name like '%Ku%' order by name;

Akira Kurosawa, 501 wirkt hier enorm hoch!

Correct would be:

select name, personid, count(*) as cnt from persons natural join directors2movies where name like '%Ku%' group by personid order by cnt desc limit 10;

select name, personid, count(*) as cnt from persons natural join directors2movies where name like '%Kubrick%' group by personid order by cnt;

To really be sure:

select name, personid, count(*) as cnt from persons natural join directors2movies where name like 'Stanley Kubrick' order by cnt;

(Stanley Kubrick is the only director and he directed 13 movies)

--- Are there any other Kubricks? (Yes)

--- Which movies did he direct? (... , The Shining, Full Metal Jacket, ...)

select originaltitle, name, personid from persons natural join directors2movies natural join movies where name like 'Stanley Kubrick';

--- How much profit did they make? (203.406.946 Total)

For each title:

```
select originaltitle, grossworldwide - budget as profit from persons natural join
directors2movies natural join movies natural join movies2grossworldwide natural join
movies2budget where name like 'Stanley Kubrick';
```

For total:

```
select name, sum(grossworldwide - budget) as totalprofit from persons natural join
directors2movies natural join movies natural join movies2grossworldwide natural join
movies2budget where name like 'Stanley Kubrick' group by name;
```

--- How many awards did they/he win? (104 Total, 34 Total)

THEY (104):

way of thought:

.schema awards2movies

```
select distinct awardoutcome from awards;
```

Minimum for getting all awards awarded to Stanley Kubrick:

```
select name, personid, count(*) as cnt from persons natural join directors2movies natural
join awards2movies natural join awards where personid='nm0000040' and
awardoutcome='Winner';
```

Solution with originaltitles:

```
select originaltitle, count(*) as cnt from directors2movies natural join awards2movies
natural join awards natural join movies where personid='nm0000040' and
awardoutcome='Winner' group by movieid;
```

HE (34):

```
select name, count(*) as cnt from persons natural join directors2awards natural join awards
wh ere personid='nm0000040' and awardoutcome='Winner' group by name;
```

--- Bonus: Are there any S. Kubrick movies, that didn't win an award? Which?

Subquerie:

```
select movieid from directors2movies natural join awards2movies natural join awards
natural join movies where personid='nm0000040' and awardoutcome='Winner' group by
movieid;
```

```
select originaltitle, movieid from directors2movies natural join movies where
personid='nm0000040' and movieid not in (select movieid from directors2movies natural
join awards2movies natural join awards natural join movies where personid='nm0000040'
and awardoutcome='Winner' group by movieid);
```

Folie 7 (Queries):

--- Which movie made the most profit and when was it released? (Avatar, 2009, 2.553.439.092)

First way, that comes to mind:

```
select originaltitle, year, grossworldwide - budget as profit from movies natural join movies2grossworldwide natural join movies2budget order by profit desc limit 20;
```

Use max function:

```
select originaltitle, year, max(grossworldwide - budget) as profit from movies natural join movies2grossworldwide natural join movies2budget;
```

--- Who directed it? (James Cameron)

```
select name, originaltitle from persons natural join directors2movies natural join movies where originaltitle='Avatar';
```

You may realize, that it's important, that '=' is case sensitive, while 'like' is NOT

```
select name, originaltitle from persons natural join directors2movies natural join movies where originaltitle like 'avatar';
```

--- Which movie has the most directors and how many? (36)

```
select movieid, count(*) as cnt from directors2movies group by movieid order by cnt desc limit 20;
```

Weird Number, what happened?

```
select originaltitle, title, year, movieid, count(*) as cnt from directors2movies natural join movies group by movieid order by cnt desc limit 20;
```

Let's check that on the IMovieDB:

<https://www.imdb.com/>

https://www.imdb.com/title/tt0973844/?ref_=nv_sr_srg_0_tt_1_nm_0_q_Chacun%2520son%2520cin%25C3%25A9ma%2520ou%2520Ce%2520petit%2520coup

--- How many movies have the same title (1093)?

Subquerie:

```
select originaltitle, count(*) as cnt from movies group by originaltitle having cnt > 1 order by cnt desc;
```

```
select count(*) as duplicatecnt from (select originaltitle, count(*) as cnt from movies group by originaltitle having cnt > 1 order by cnt desc);
```

--- What are the

most popular actor names?

subquerie (List of names (distinct since personid) and in how many movies they played in):

```
select name, count(*) as cnt from persons natural join actors2movies group by personid limit 20;
```

better:

```
select name from persons natural join actors2movies group by personid limit 20;
```

solution REMEMBER TO NOT LIMIT THE SUBQUERIE:

```
select name, count(*) as cnt from (select name, personid from persons natural join actors2movies group by personid) group by name order by cnt desc limit 20;
```

Top 10 Movies with the least/most actors? (1, 1301)

```
select originaltitle, count(*) as cnt from movies natural join actors2movies group by movieid order by cnt limit 10;
```

```
select originaltitle, count(*) as cnt from movies natural join actors2movies group by movieid order by cnt desc limit 10;
```

Top 20 directors who made the most movies (Alfred Hitchcock)

```
select name, count(*) as cnt from persons natural join directors2movies group by personid order by cnt desc limit 20;
```

Top 20 directors with the most awards? (Steven Spielberg)

```
select name, count(*) as awardcnt from persons natural join directors2movies natural join awards2movies natural join awards where awardoutcome='Winner' group by personid order by awardcnt desc limit 20;
```

Top 10 Actors that played in the most movies?

What could be a problem while finding this solution? (An actor could be listed more than once for a given movie)

Thats why we group by personid:

```
select name, count(*) as cnt from persons natural join actors2movies group by personid order by cnt desc limit 10;
```

Folie 8 (Bonus: Errors)

--- Duplicates

Find our duplicate in directors2movies:

```
select personid, awardid, count(*) cnt from directors2awards group by personid, awardid  
order by cnt desc limit 20;
```

Get the name (Hong-Seung Yoon):

```
select * from directors2awards natural join persons where personid='nm3460905';
```

Which award:

```
select * from directors2awards natural join persons natural join awards where  
personid='nm3460905';
```