CSC 370 — Database Systems
July 7, 2015
Assignment No. 7

Note 1 **This assignment is to be done individually.**

- Due date: July 14, 2015, at the beginning of the class.
- This assignment is worth 1% of your total course mark.

Objectives

We made it! This is the last assignment of the term. After completing this assignment, you will have experience:

- Learn how the DBMS estimates the number of tuples of a result.

Your task, should you choose to accept it

For this assignment use the database **imdb**.

Part A

1. How many blocks does the relation *productions* use? Postgresql uses the term pages.
2. On the average how many records per block does this relation use?
3. How many blocks does the index on *productions(id)* use?

Part B

Assume the following queries.

```
a. select * from  productions where year = 2012;
b. select * from  productions where year IS NULL;
c. select * from  productions where year IS NULL and year = 2012;
c. select * from  productions where year IS NULL or year = 2012;
d. select * from  productions where year >= 1974 and year < 1977;
```

For each of these queries:

1. Using EXPLAIN ANALYZE, record the number of tuples that Postgresql estimates each query will return and the actual number of tuples returned.
2. Using the information stored in the system tables (pg_stats, pg_class, etc) calculate from scratch the estimated number of tuples in the result in the same way Postgresql does it. Explain any discrepancy between your result and the one calculated by Postgresql.

Hints:

- Use relations *pg_class* and *pg_stats*.
- Use section 52 of Postgresql 9.3 manual (`http://www.postgresql.org/docs/9.3/static/indexam.html`).
- Assume that each bucket in the histogram has a uniform distribution (inside a given bucket, any year is equally likely to appear).
- The counts in the buckets do not include nulls nor the most common values.

What to submit

Submit a paper copy of your solution before the beginning of class.