



Assignment 1

CMPUT 391 section A1

Fall 2015

Assignment due in class on **Monday October 19, 2015**.
Assignment counts for 5% overall

Question 1 [9 points]

Consider the following set of functional dependencies:

$$F = \{C \rightarrow A, GB \rightarrow CF, F \rightarrow ED, A \rightarrow G, E \rightarrow B\}$$

Prove that the following functional dependencies are entailed by F , using **only** Armstrong's axioms **but not decomposition or union rule**.

- A) [3 points] $FA \rightarrow GED$
- B) [6 points] $EC \rightarrow D$

Question 2 [14 points]

Consider the following relational schema:

$$R = (ABCDE; \{DE \rightarrow A, C \rightarrow E, A \rightarrow B, B \rightarrow A\}).$$

Give two possible decompositions of this schema into BCNF, using the decomposition algorithm explained in class. Make sure that one of these two decompositions is dependency preserving [8 points]; the other one [6 points] does not have to be dependency preserving. Show in detail the steps that you apply to obtain the decompositions, and prove whether your decomposition is dependency preserving or not.

Question 3 [12 points]

Consider the following relational schema:

$$R = (ABCDEFGHI; \{FIH \rightarrow ADG, I \rightarrow A, D \rightarrow EB, FH \rightarrow D, FI \rightarrow AG\}).$$

Synthesize a 3NF schema based on a minimal cover for the functional dependencies. Explain all steps.

Question 4 [6 points]

Consider a relation R with attributes A, B, C, D which have all the same data type. Assume that you have 1000 buffer pages in main memory. The size of a buffer page is 4 Kilobyte (corresponding to the disk page size). Assume a perfect hash function with respect to the values of attribute B , i.e., a hash function that distributes the values of B evenly over its bucket. Compute the maximum size of the input file (in an ideal case) for which a hash-based projection on B , Π_B , can be computed without having to externally sort the buckets.

Question 5 [7 points]

Consider the relation schema $R(A, B, C, D)$ with the following characteristics:

- Total number of tuples: 2,000,000
- 20 tuples per page
- Attribute A is the primary key, its values are numeric ranging from 1 to 2,000,000
- There is a clustered but not integrated B+ tree index of depth 4 on A
- Attribute B has 100,000 distinct values
- There is an unclustered B+-tree index on B

Estimate the number of page transfers needed to evaluate each of the following queries for each of the given methods (assume uniform and independent distribution of values if necessary).

- $\sigma_{A < 2050}$
 1. using sequential scan [1 point]
 2. using the index on A [2 points]
- $\sigma_{A > 2050 \wedge A < 30000 \wedge B = 5}$
 1. using the index on A [2 points]
 2. using the index on B [2 points]

Question 6

Consider the relation schema $R(A, B, C, D)$ and $S(B, E)$ with the following characteristics:

- Total number of tuples in R: 2,000,000
- Total number of tuples in S: 1,000
- Attribute A is the primary key of R with values ranging from 1 to 2,000,000
- Attribute B is the primary key of S, with values ranging from 1 to 1,000
- Attribute C has values “low”, “med. low”, “medium”, “med. high, and “high”
- Attribute D and E have each 100 distinct values
- Assume uniform and independent distribution of values, and assume that all values occurring in S.B also occur in R.B

Estimate the result size of the following relational algebra expressions. Give the result size in number of tuples (assume uniform and independent distribution of values if necessary).

$$\sigma_{R.C = \text{medium and } R.B = 5} ((R \bowtie S))$$