HW4 - Example Solution

1. **1.A) Hardware and DBMS Design**

   MULTI | 5 points | 0.10 penalty | Multiple

   1.A) Select the true statements:

   (a) SSDs are especially well suited to improve performance of retrieving large result sets using unclustered indexes. (50%)
   (b) Before a transaction can be committed, all the disk pages it has updated must be written to secondary storage to ensure durability. (−50%)
   (c) In a main memory system, it is not necessary to include the "old" values when logging changes made by transactions to secondary storage. (50%)
   (d) As discussed in a lecture, Google claims to have built a distributed system that offers both consistency and availability in the face of network partitions; this proves that the CAP theorem is wrong. (−50%)

2. **1.B) Hardware and DBMS Design**

   ESSAY | 5 points | 0.10 penalty | monospaced

   1B) Write your reflections here:

   Notes for grader:

   - `In the context of ACID transactions, consistency refers primarily to maintenance of primary keys and foreign keys. While globally unique identifiers are feasible, using foreign keys would require either the ability to partition the data such that those constraints can be checked locally, or significant work to check constraints at update time. Since data collections that are large enough to require distributed systems are rarely relational, this effort is not considered worthwhile.`

3. **2.A) Data Systems for Analytics**

   MULTI | 5 points | 0.10 penalty | Multiple

   2.A) Select the true statements:

   (a) Today's key-value stores implement all the functionality required to support *all* needs of analytics applications. (−50%)
   (b) In big data, "velocity" means that it is necessary to react quickly to the large amounts of data being added to the system. (50%)
   (c) Emails are an example of semi-structured data. (50%)
   (d) Spark supports low-latency big data analytics. (−50%)

4. **2.B) Data Systems for Analytics**

   ESSAY | 5 points | 0.10 penalty | monospaced

2.B) Write your arguments here:

Notes for grader:

- `A distributed computing framework, such as Spark or Hadoop, would be a good system in this scenario, as it allows complex computations, distributed well to large clusters, and has good throughput. Spark and Hadoop do not offer low latency, but latency is not needed in this scenario. Spark is more flexible and more efficient than Hadoop, and would be the system of choice today.`

5. **3.1 (Projects) FDs**

ESSAY | 5 points | 0.10 penalty | monospaced

3.1 (Projects) Write down FDs for the Projects relation.

Notes for grader:

- PID → PN, SID → SN, ID → MID, MID → MN, ID → MN

6. **3.2 (Projects) Decomposition**

ESSAY | 5 points | 0.10 penalty | monospaced

3.2 (Projects) Write down the decomposed relations for the Projects relation.

Notes for grader:

- Projects1(MID, MN)
  Projects2(ID, MID)
  Projects3(PID, PN)
  Projects4(SID, SN)
  Projects0(ID, PID, SID)

7. **3.3 (Projects) DDL**

ESSAY | 5 points | 0.10 penalty | monospaced

3.3 (Projects) Write down SQL commands to create and populate the decomposed relations for Projects.

Notes for grader:

- See Homework4-SOL3.sql

8. **3.4 (Projects) Normal Form**

MULTI | 5 points | 0.10 penalty | Single

3.4 (Projects) Select the lowest normal form for any relations of the Projects schema:

(a) All relations are in BCNF (and therefore also 3NF) (100%)
(b) Some relations are only in 3NF, not in BCNF

9. **4.1) ER-diagram**

   ESSAY | 10 points | 0.10 penalty | monospaced

   4.1) Submit ER-diagram.

   Notes for grader:

   - See Homework4-SOL4.PDF

10. **4.2) DDL**

    ESSAY | 10 points | 0.10 penalty | monospaced

    4.2) SQL DDL to implement the diagram.

    Notes for grader:

    - See Homework4-SOL4.sql

11. **5(a) Numerical answer**

    NUMERICAL | 2.5 points | 0.10 penalty

    5(a) Run the query and paste the result here (an integer):

    - 2 ✓

12. **5(b) Numerical answer**

    NUMERICAL | 2.5 points | 0.10 penalty

    5(b) Run the query and paste the result here (an integer):

    - 3883371 ✓

13. **5(c) Numerical answer**

    NUMERICAL | 2.5 points | 0.10 penalty

    5(c) Run the query and paste the result here (an integer):

    - 833 ✓

14. **5(d) Numerical answer**

    NUMERICAL | 2.5 points | 0.10 penalty

    5(d) Run the query and paste the result here (an integer):

    - 24 ✓

15. **5(e) Numerical answer**

    NUMERICAL | 2.5 points | 0.10 penalty

    5(e) Run the query and paste the result here (an integer):

- 421 ✓

16. **5(f) Numerical answer**

    NUMERICAL | 2.5 points | 0.10 penalty

    5(f) Run the query and paste the result here (an integer):

    - 2072 ✓

17. **5(g) Numerical answer**

    NUMERICAL | 2.5 points | 0.10 penalty

    5(g) Run the query and paste the result here (a floating point number):

    - 1.1994.0001 ✓

18. **5(h) Numerical answer**

    NUMERICAL | 2.5 points | 0.10 penalty

    5(h) Run the query and paste the result here (an integer):

    - 1278 ✓

Total of marks: 80