Quiz 2

- There are 5 problems in this quiz, worth 12 points in total.
- You must not communicate with other students during this test.
- Books or notes are not allowed.
- Electronic devices except calculators are not allowed.
- You cannot use your mobile phone as calculator.
- The duration of this quiz is 30 minutes.
- Wait for instructions before turning this page.

1. Fill in your information:

Full Name: _____

NetID: _____

1/1. (3 points) The rank of a number, given an unsorted list, is its position in the the list, after sorting the list in **descending** order. For example, the rank of 7 in list $[1, 9, 5, 3, 7]$ is 2 (The sorted version of this list is $[9, 7, 5, 3, 1]$, where 7 occupies the second position). Assume, the *rank* function returns the rank of each element in a list. For example, for input $[1, 9, 5, 3, 7]$, the *rank* function will return $[5, 1, 3, 4, 2]$. Is *rank* a distributive, an algebraic, or a holistic measure? Explain why.

Solution. *rank* is a holistic measure. There is no constant bound on the storage size needed to describe a sub-aggregate of *rank* function.

1/2. (3 points) The rank of a number, given an unsorted list, is its position in the the list, after sorting the list in **ascending** order. For example, the rank of 7 in list $[1, 9, 5, 3, 7]$ is 4 (The sorted version of this list is $[1, 3, 5, 7, 9]$, where 7 occupies the fourth position). Assume, the *rank* function returns the rank of each element in a list. For example, for input $[1, 9, 5, 3, 7]$, the *rank* function will return $[1, 5, 3, 2, 4]$. Is *rank* a distributive, an algebraic, or a holistic measure? Explain why.

Solution. *rank* is a holistic measure. There is no constant bound on the storage size needed to describe a sub-aggregate of *rank* function.

2/1. (3 points)

The following table shows the base cells of a datacube with four dimensions. Compute all closed cells for this datacube. Hint: $(a_1, b_1, c_1, d_1 : 1)$ is a closed cell.

Table 1: Datacube

A	B	C	D
a_1	b_1	c_1	d_1
a_1	b_1	c_1	d_3
a_1	b_2	c_2	d_2

Solution. $(a_1, *, *, * : 3)$, $(a_1, b_1, c_1, * : 2)$, $(a_1, b_1, c_1, d_1 : 1)$, $(a_1, b_1, c_1, d_3 : 1)$, $(a_1, b_2, c_2, d_2 : 1)$

2/2. (3 points)

The following table shows the base cells of a datacube with four dimensions. Compute all closed cells for this datacube. Hint: $(d_1, b_1, c_1, f_1 : 1)$ is a closed cell.

Table 2: Datacube

D	B	C	F
d_1	b_1	c_1	f_1
d_1	b_1	c_1	f_3
d_1	b_2	c_2	f_2

Solution. $(d_1, *, *, * : 3)$, $(d_1, b_1, c_1, * : 2)$, $(d_1, b_1, c_1, f_1 : 1)$, $(d_1, b_1, c_1, f_3 : 1)$, $(d_1, b_2, c_2, f_2 : 1)$

3/1. (3 points) A data cube has 5 binary (two possible value) dimensions. Assume that there are no concept hierarchies associated with the dimensions. (i) What is the maximum number of cells possible in the base cuboid? (ii) What is the minimum number of cells possible in the base cuboid? Please provide explanation for each sub-question.

Solution. For computing maximum base cells, we consider the presence of all possible combination of attribute values in the base cells ($a0b0c0d0e0, a0b0c0d0e1, a0b0c0d1e0, \ldots, a1b1c1d1e1$). For minimum base cells, we align the possible values of the attributes in such a way that there's minimum number of base cells ($a0b0c0d0e0, a1b1c1d1e1$).
(i) 2^5, (ii) 2.

3/2. (3 points) A data cube has 4 ternary (three possible value) dimensions. Assume that there are no concept hierarchies associated with the dimensions. (i) What is the maximum number of cells possible in the base cuboid? (ii) What is the minimum number of cells possible in the base cuboid? Please provide explanation for each sub-question.

Solution. Same explanation as 3/1.
(i) 3^4, (ii) 3.

4/1. (1 point) What are the differences between a star and a snowflake schema?

Solution. Star schema consists of a fact table in the middle connected to a set of dimension tables. Snowflake schema is a refinement of star schema where some dimensional hierarchy is normalized into a set of smaller dimension tables, forming a shape similar to snowflake.

4/2. (1 point) What are the differences between a star and a snowflake schema?

Solution. Star schema consists of a fact table in the middle connected to a set of dimension tables. Snowflake schema is a refinement of star schema where some dimensional hierarchy is normalized into a set of smaller dimension tables, forming a shape similar to snowflake.

5/1. (2 points) For the datacube shown in the following table, return the result of the following query: $< a_1, ?, *, ? : count >$.

Table 3: Datacube

A	B	C	D
a_1	b_1	c_1	d_1
a_1	b_1	c_1	d_3
a_1	b_2	c_2	d_2
a_2	b_2	c_2	d_1

Solution. $< a_1, b_1, *, d_1 : 1 >, < a_1, b_1, *, d_3 : 1 >, < a_1, b_2, *, d_2 : 1 >$

5/2. (2 points) For the datacube shown in the following table, return the result of the following query: $< a_1, *, ?, ? : count >$.

Table 4: Datacube

A	B	C	D
a_1	b_1	c_1	d_1
a_1	b_1	c_1	d_3
a_1	b_2	c_2	d_2
a_2	b_2	c_2	d_1

Solution. $< a_1, *, c_1, d_1 : 1 >, < a_1, *, c_1, d_3 : 1 >, < a_1, *, c_2, d_2 : 1 >$
