DO NOT OPEN UNTIL INSTRUCTED TO DO SO!!!!

Write clearly, precisely, and briefly!!

<table>
<thead>
<tr>
<th>ID:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grades</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Section</td>
<td>Max</td>
<td>Scored</td>
</tr>
<tr>
<td>A</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>
A. Questions from Chapter 2 (Data Preprocessing)

1. Briefly explain each of the following two measures. Also, explain what the benefit of understanding them is. 

   a. Algebraic measure

   b. Holistic measure

2. Using smoothing by bin boundary, remove the noise from the following data:

   5, 9, 16, 22, 22, 25, 26, 29, 35
3. Briefly explain the two main problems that are resolved during data integration. [3]

4. How is correlation analysis used for data integration? [3]

5. Data transformation involves normalization. Using an example, explain the use of normalization for clustering. [3]
6. Explain the following data reduction techniques:

   a. cluster sample

   b. stratified sample

   c. clustering

   d. Stepwise forward selection algorithm
7. Explain how discretization is done using the following techniques [6]
   
a. ChiMerge

b. Entropy-based discretization

c. 3-4-5 Rule
Questions from Chapter 3 (DW & OLAP)

1. What is a data warehouse

2. Suppose that a data warehouse consists of 3 dimensions, namely, TIME, PRODUCT, and BRANCH, and two measures namely, COUNT, and SALES. Assume the distinct values in TIME, PRODUCT, and BRANCH are 100, 10, and 5 respectively

   a. Draw a star schema diagram for the data warehouse
b. If each dimension consists of 2 levels, what is the maximum number of cuboids that can be created?

c. By how many cells does the base cuboid grow every day.

d. What is the number of cells in the data cube? (Assume we have a full cube.)

e. How would you generate all the cuboids efficiently? (You can a graph to show which cuboid will be generated from which.)
3. Why the star schema gives better performance than the snow flake schema? [3]

4. Using an example, explain the following OLAP operations [4]
   a. Drill down
   b. Dice
5. Briefly explain the architecture of a data warehouse. [3]

6. Compare ROLAP and MOLAP in terms of flexibility and performance. [3]