

1. (The following four questions are based on *restaurants.json* data in the hands-on session.) Write a MongoDB query to display the fields `restaurant_id`, `name` and `cuisine`, but exclude the field `_id` for all the documents in the collection `restaurant`. Paste the running results for your query. (0.5 point)
2. Write a MongoDB query to find the restaurants who achieved a score more than 85. Paste the running results for your query. (0.5 point)
3. Write a MongoDB query to find the restaurants which do not prepare any cuisine of 'American ' and achieved a grade point 'A' not belongs to the borough Brooklyn. The document must be displayed according to the cuisine in descending order. Paste the running results for your query. (0.5 point)
4. Write a MongoDB query to find the restaurant `Id`, `name`, `address` and geographical location for those restaurants where 2nd element of `coord` array contains a value which is more than 40 and up to 50. Paste the running results for your query. (0.5 point)
5. What is the main problem with big data information integration? (0.5 point)
 - A. Many sources
 - B. Probabilistic Schema Mapping
 - C. Pay-as-you-go model
 - D. Mediated Schema
6. What would be the two possible solutions associated with "big data" information integration as mentioned in lecture? (Choose 2) (0.5 point)
 - A. Attribute Grouping
 - B. Mediated Schema
 - C. Pay-as-you-go Model
 - D. Customer Transactions
 - E. Probabilistic Schema Mapping
7. What are mediated schemas? (0.5 point)
 - A. Schema created from integrating two or more schemas.
 - B. Schemas created entirely from attribute grouping.
 - C. A type of probabilistic schema mapping.
 - D. Schemas created from customer info.

8. In attribute grouping, how would one evaluate if two attributes should go together? (Choose 2) (0.5 point)

- A. Probability of Two Attributes Co-occurring
- B. Integrated Views
- C. Candidate Designs
- D. Customer Interaction
- E. Similarity of Attributes

9. What is data fusion? (0.5 point)

- A. Another term for customer analytics.
- B. Extracting a global value from a data source.
- C. Extracting true sources from a data source.
- D. Extracting the true value of a data item.

10. What do we mean when we say "the true value of a data item"? (0.5 point)

- A. Data created from statistical estimations.
- B. Another term for data fusion.
- C. Extrapolated data from a data item that represents the worth of that item.

11. In a mediator (data integration) system, the mediated schema is defined as below:

CUSTOMER (ID, name, state_name)

There are two data resources:

S1: Customer (ID, full_name, region)

S2: Customer (ID, first_name, last_name, region_code), Region(region_code, region_name)

Assume "name" in the mediated schema means "full_name" in S1. "Region" in data source S1 means "state_name" in the mediated schema. (2 points)

- (a) Can you do schema mediation in Global-as-view? Write down your solutions in SQL views.
- (b) Write a query "List all customer names in Helsinki" in SQL based on the mediated schema. Can the query be answered with your view?