Analytics without precomputation

Consider a database that captures customers with a name and state; states with a name; products that have a list price, a name and belong to a category; categories have names and descriptions; sales of a product by a customer capturing quantity and price paid (may be discounted).

Due Friday 4/15 midnight: Produce an SQL schema that captures the above information.

Write the following queries:

1. Show the total sales (quantity sold and dollar value) for each customer.
2. Show the total sales for each state.
3. Show the total sales for each product, for a given customer. Only products that were actually bought by the given customer. Order by dollar value.
4. Show the total sales for each product and customer. Order by dollar value.
5. Show the total sales for each product category and state.
6. For each one of the top 20 product categories and top 20 customers, it returns a tuple (top product, top customer, quantity sold, dollar value)

Due Friday 4/29 midnight: Write and submit the above SQL queries.

Your next goal is to deliver top performance for the queries. Expect a significant amount of experimentation until you tune the performance.

In particular, create any indices that will be beneficially used by the queries. Do not create useless indices. Measure the performance of the Exercise 2 queries on cold runs. Utilize the data generator that Junbo will provide in order to test performance. You may have to reformulate some queries to maximize performance.

Due Friday 5/13 midnight: Submit your index choices (in the form of the script with the CREATE INDEX statements) and the measured query performance on the Amazon instance. The submission will be followed by a demo and discussion with Junbo or Yannis, where you will justify your choices.

Your next goal will be to improve the performance of the Query 6 by appropriate precomputations. It is understood that the maintenance of the precomputed table(s) will lead to some slowdown while viewers interact. Your precomputation choices must be such that the precomputations that you introduce collectively save more time to this option than they cost to maintain. Precomputed tables will also benefit from creating certain indices on them. Build any beneficial indices. Calibrate your solution against cold runs of the insert-and-querying mix created by the activity script that Junbo will provide.

Due Friday 5/27 midnight: Submit the following. The submission will be followed by a demo and discussion with Junbo or Yannis, where you will justify your choices.

- Which precomputed tables you created: CREATE TABLE statement and query that initially loaded it.
- Your new Query 6, which makes the best use of the precomputed tables you chose.
- Which indices you created (CREATE INDEX scripts)