

CMSC 724, Spring 2015: Homework 2

Due Friday February 27, 2015, 11:59pm.

The homework is to be done by yourself. You are free to make reasonable assumptions about the problem if something is not fully specified – make sure to state any assumptions. Try to be concise in your answers – most answers should be 1-2 paragraphs long. Total weight = 5%.

1. Explain two reasons why hash joins are preferred over sort-merge joins in today's database systems.
2. One of the main innovations of *VectorWise* over *MonetDB* was use of a *vectorized execution model*. Explain how this model strikes a balance between MonetDB's approach and tuple-at-a-time processing, and its main pros and cons.
3. Briefly explain two advantages of late materialization and why it may sometimes be slower than the alternative.
4. Consider the following query with an EXCEPT over two relations: $R(A, B)$ and $S(A, C)$.

```
(select A from R where B = 10)
EXCEPT
(select A from S where C = 10)
```

Rewrite this query to a single-block query by first converting it to a two-block query (with a subquery), and then converting it into a single-block query. Suggest a general approach to handling such queries. Keep in mind that EXCEPT eliminates duplicates by default. Suggest reasons why such a rewrite may give large performance improvements.

5. Briefly explain in your words the idea of “using semijoin-like techniques for optimizing multi-block queries” (Section 4.3, Chaudhuri paper).