Database System Architectures (INFO-H-417)

Mahmoud SAKR

Written Exam Topics

S2 – Refreshing SQL
• Express SQL queries including selection, projection, join, aggregation, CTE, and subqueries.
S3 - Translating SQL into Relational Algebra
• Describe the concepts of declarative and procedural query languages • Describe the different relational algebra (and extended RA) operators • Translate a given SQL statement into an equivalent RA expression. • Differentiate the set and bag semantics in RA, and describe how the query results may differ applying one semantic or the other. • Illustrate the transformation of a sub-query into a join then into an RA expression.
S4,5: System R & Query Optimizations
• Describe the architecture components of system R. • In light of the System R paper, describe the concepts of: cataloge, tuple identifier, image, clustering image, view, cost-based query optimization, and access path • Discuss how far the concepts in the previous point are implemented in PostgreSQL • Given an RA expression, apply equivalence rules to transform it into other equivalent RA expressions *Note: you don't need to memorize equivalence rules. They will be given if needed*. • Assess whether or not a given equivalence rules is valid, under set or bag semantics, and whether there are constraints for its validity • Illustrate the computational challenge of cost based query optimization, and ways to reduce the cost (e.g., the join ordering problem, heuristic optimization) • Illustrate the use of statistics in estimating the cost of a query plan • Discuss the benefits and costs of using materialized views in query optimization
S6 - Indexing
• Explain the use of indexes in query processing • Explain the concepts of sequential file, dense index, sparse index, 1st level index, and 2nd level index, secondary index • Illustrate the insertion/deletion stategies in convential indexes, also for the case of duplicate keys • Illustrate the benfits of buckets in secondary indexes • Illustrate the Btree (also called B+tree and B-tree) index, its parameters, and how insertions and deletions are performed • Illustrate the properties of the Btree that allows us to answer inequality (<,

<=, >, >=) or range searches (between) efficiently
- Illustrate the use of index only scans
- Discuss how nulls are handled in Btree

S7 – Extending database systems

- Explain the architectural components that make PostgreSQL extensible:
  - What is the role of the catalog ?
  - How is PostgreSQL able to process user types (storage, input, output, statistics, etc)
  - How is PostgreSQL able to compute functions over user types
  - How is PostgreSQL able to use its generalized index structures over user types
  - What is the role of extensions in PostgreSQL
- Describe (in English not in coding) the steps/tasks that one would need in order to create a PostgreSQL extension similar to the complex numbers extension that you created in the exercise session

S9 - MobilityDB

- Discuss what are the types of moving object data, and applications
- Describe what is a moving object database
- Describe at an abstract level the components of MobilityDB as a PostgreSQL extension: types, functions, index support, optimizer statistics

S10 – Distributed databases

- Discuss what are the benfits of distributing a database
- Describe the concepts of: distributed table, replicated table, distribution key, range distribution, hash distribution, spatial distribution, rebalancing, reference tables.
- Illustrate the importance of co-location
- Discuss the approach of the multi-relational algebra as a model for distributed RA
- Illustrate methods for computing non co-located joins in distributed databases, and reflect on their cost

With my best wishes.