



UNIT V – Physical Storage and MongoDB

Data Storage and Indexes – RAID- File organization-Indexing and Hashing –Ordered Indices – B+ tree Index Files – B tree Index Files – Static Hashing – Dynamic Hashing. Query Processing Overview-Algorithms for Selection and Sorting Basics of MongoDB, Procedural Language

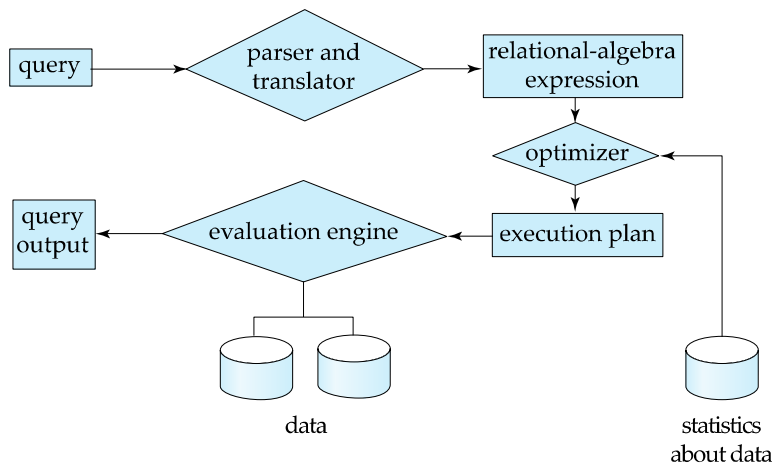
Query Processing Overview

Query Processing includes translations of high-level Queries into low-level expressions that can be used at the physical level of the file system, query optimization, and actual execution of the query to get the actual result.

High-level queries are converted into low-level expressions during query processing. It is a methodical procedure that can be applied at the physical level of the file system, during query optimization, and when the query is actually executed to obtain the result. It needs a basic understanding of relational algebra and file organization. It includes the variety of tasks involved in getting data out of the database. It consists of converting high-level database language queries into expressions that can be used at the file system's physical level.

The process of extracting data from a database is called query processing. It requires several steps to retrieve the data from the database during query processing. The actions involved actions are:

1. Parsing and translation
2. Optimization
3. Evaluation



Parsing and translation

- translate the query into its internal form. This is then translated into relational algebra.
- Parser checks syntax, verifies relations

Evaluation

- The query-execution engine takes a query-evaluation plan, executes that plan, and returns the answers to the query.

Optimization

- A relational algebra expression may have many equivalent expressions

Example: $\sigma_{salary < 75000}(\Pi_{salary}(instructor))$ is equivalent to

$$\Pi_{salary}(\sigma_{salary < 75000}(instructor))$$

- Each relational algebra operation can be evaluated using one of several different algorithms

Correspondingly, a relational-algebra expression can be evaluated in many ways.

- Annotated expression specifying detailed evaluation strategy is called an evaluation-plan.

E.g., can use an index on *salary* to find instructors with salary < 75000, or can perform complete relation scan and discard instructors with salary $\geq$ 75000

- Query Optimization: Amongst all equivalent evaluation plans choose the one with lowest cost.

Cost is estimated using statistical information from the database catalog. e.g. number of tuples in each relation, size of tuples, etc.