Australian National University

Assignment Project Exam Help

Query Processing

https://tutorcs.com

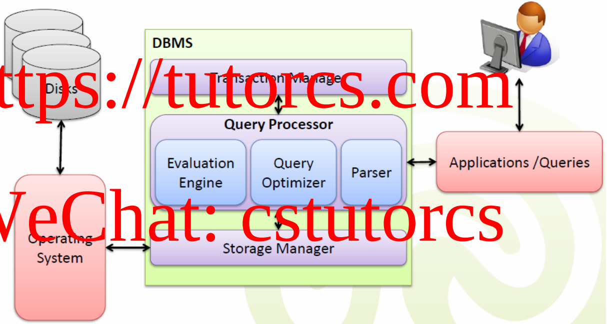
WeChat: cstutorcs

Query Processing – Overview

Assignment Project Exam Help

1. Users **submit** SQL queries to a DBMS.
2. The DBMS **processes and executes** them in a database.

https://tutorcs.com



WeChat: cstutorcs

- **Note:** SQL is a declarative language, so it is the task of DBMSs to decide how SQL queries should be executed.

Query Processing – Example

Assignment Project Exam Help

- **From:**

```
SELECT name FROM Person WHERE age<21;
```

https://tutorcs.com

- **To:**

name
Rickon
Bran

WeChat: cstutorcs

- **Questions:**
  - **How does a relational DBMS process this?**
  - **How can a relational DBMS process this efficiently?**

Query Processing – Example

Assignment Project Exam Help

```
SELECT name FROM Person WHERE age<21
```

High-level language (SQL)

$\Downarrow$

$$\pi_{name}(\sigma_{age<21}(\texttt{Person}))$$

Low-level language (Relational Algebra)

$\Downarrow$

https://tutorcs.com

$$\begin{array}{c} \pi_{name} \\ | \\ \sigma_{age<21} \\ | \\ \texttt{Person} \end{array}$$

Execution plan (Query tree)

WeChat: cstutorcs

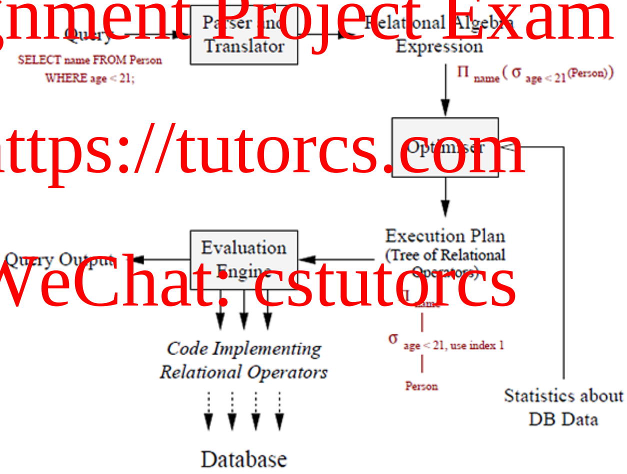
$\Downarrow$

name
Rickon
Bran

Query result

Query Processing – Example

Assignment Project Exam Help
https://tutorcs.com
WeChat: cstutorcs



Query Processing Steps

Assignment Project Exam Help

- **Query parser and translator**
  1. Check the syntax of SQL queries
  2. Verify that the relations do exist
  3. Transform into relational algebra expressions

https://tutorcs.com

- **Query optimiser**
  1. Transform into the best possible execution plan
  2. Specify the implementation of each operator in the execution plan

WeChat: cstutorcs

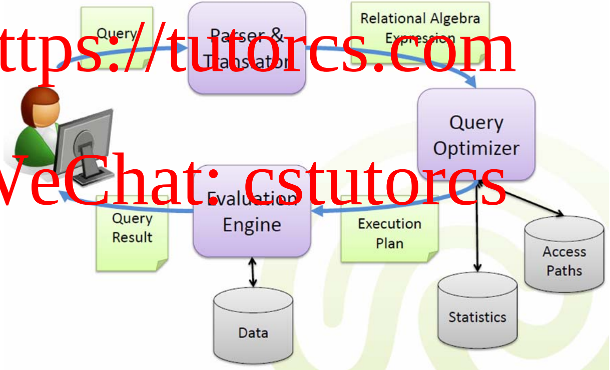
- **Evaluation engine**
  1. Evaluate the query execution plan
  2. Return the result to the user

Australian National University

Query Processing – Parser

- The **parser** checks the syntax of the query:
  - Validation of table names, attributes, data types, access permission ...;
  - Either the query is executable or an error message is generated.

Assignment Project Exam Help
https://tutorcs.com
WeChat: cstutorcs



Query Processing – Parser

Assignment Project Exam Help

- Consider the relation schema:

  `Person(id:integer, name:string, age:integer, address:string)`

- **Note:** **System catalog** (also called **data dictionary**) is used at this stage, which contains the information about data managed by the DBMS.

  Example:

https://tutorcs.com

attr_name	rel_name	type	position
id	Person	integer	1
name	Person	string	2
age	Person	integer	3
address	Person	string	4
...	...	...	...

WeChat: cstutorcs

- **Question:** Can the following query be accepted by the parser?

```
SELECT fname, lname FROM Person WHERE address<21;
```

Query Processing – Parser

Assignment Project Exam Help

- Consider the relation schema:

  `Person(id:integer, name:string, age:integer, address:string)`

https://tutorcs.com

- **Question:** Can the following query be accepted by the parser?

```
SELECT fname, lname FROM Person WHERE address<21;
```

WeChat: cstutorcs

- **Answer:** The query **would be rejected** because
  1. The attributes `fname` and `lname` are not defined;
  2. The attribute `address` is not comparable with 21.

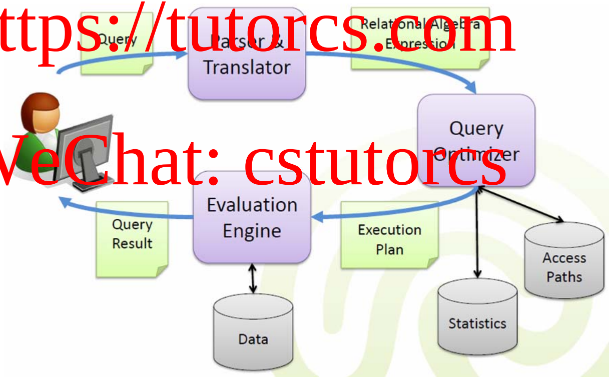
Query Processing – Translator

Assignment Project Exam Help

- The **translator** translates queries into RA expressions (not necessarily equivalent due to duplicates):
  - A query is first decomposed into **query blocks**.
  - Each query block is translated into an RA expression.

https://tutorcs.com

WeChat: cstutorcs



Recall: RA and SQL Queries

Assignment Project Exam Help

- **RA operators**
  - **selection** σ_{φ}
  - **projection** $\pi_{A_1,\dots,A_n}$
  - **Cartesian product** $R_1 \times R_2$
  - **join** $R_1 \bowtie_{\varphi} R_2$ and $R_1 \bowtie R_2$
  - **renaming** $\rho_{R(A_1,\dots,A_n)}$
  - **union** $R_1 \cup R_2$
  - **intersection** $R_1 \cap R_2$
  - **difference** $R_1 - R_2$

https://tutorcs.com

- **SQL statement**

```
SELECT attribute_list
  FROM table_list
 [WHERE condition]
[GROUP BY attribute_list
[HAVING group_condition]]
[ORDER BY attribute_list];
```

WeChat: cstutorcs

$$\sigma_{\varphi}(R) \Leftrightarrow \texttt{SELECT * FROM } R \texttt{ WHERE } \varphi;$$

$$\pi_{A_1,\dots,A_n}(R) \Leftrightarrow \texttt{SELECT DISTINCT } A_1, \dots, A_n \texttt{ FROM } R;$$

$$R_1 \times R_2 \Leftrightarrow \texttt{SELECT DISTINCT * FROM } R_1, R_2;$$

$$\dots$$

- Aggregate operations in SQL require extended RA expressions.

Australian National University

Recall: RA and SQL Queries

Assignment Project Exam Help

- Nested subqueries are decomposed into separate query blocks.
- **Example:**

https://tutorcs.com

```
SELECT Lname, Fname
  FROM EMPLOYEE
 WHERE Salary > (SELECT Salary
                   FROM EMPLOYEE
                  WHERE ssn=5);
```

WeChat: cstutorcs

Outer query block

`SELECT Lname, Fname FROM EMPLOYEE WHERE Salary > c`

$\Downarrow$ translated

$\pi_{Lname, Fname}(\sigma_{Salary > c}(\text{EMPLOYEE}))$

Inner query block

`(SELECT Salary FROM EMPLOYEE WHERE ssn=5)`

$\Downarrow$ translated

$\pi_{Salary}(\sigma_{ssn=5}(\text{EMPLOYEE}))$

Query Processing – Query Optimiser

Assignment Project Exam Help

1. Transform into the best possible execution plan

   There are different possible relational algebra expressions for a single query!

   https://tutorcs.com

   (will be covered in this course)

2. Specify the implementation of each operator in the execution plan

   WeChat: cstutorcs

   There are different possible implementations for a relational algebra operator!

   (will not be covered in this course)

Query Processing – Query Optimiser

Assignment Project Exam Help

- SQL queries only specify **what data to be retrieved** and **not how to retrieve data**.

https://tutorcs.com

- There are **many possible execution plans** for a SQL query.

- Query optimiser is responsible for identifying **an efficient execution plan**:

  1. enumerating alternative plans (typically, a subset of all possible plans);
  2. choosing the one with the least estimated cost.

WeChat: cstutorcs

- Query optimisation is one of the most important tasks of a relational DBMS. **A good DBMS must have a good query optimiser!**

Equivalent RA Expressions

Assignment Project Exam Help

- Suppose that we have:

```
Students(matNr, firstName, lastName, email)
Exams(matNr, crsNr, result, semester)
Courses(crsNr, title, unit)
```

https://tutorcs.com

```
SELECT lastName, result, title
  FROM STUDENTS, EXAMS, COURSES
 WHERE STUDENTS.matNr=EXAMS.matNr AND
          EXAMS.crsNr=COURSES.crsNr AND result≤1.3;
```

WeChat: cstutorcs

- **Question:**

How many equivalent RA expressions for this SQL query can you find?

Equivalent RA Expressions

Assignment Project Exam Help

```
Students(matNr, firstName, lastName, email)
Exams(matNr, crsNr, result, semester)
Courses(crsNr, title, unit)
```

```
SELECT lastName, result, title
FROM STUDENTS, EXAMS, COURSES
WHERE STUDENTS.matNr=EXAMS.matNr AND
      EXAMS.crsNr=COURSES.crsNr AND result≤1.3;
```

https://tutorcs.com

- **Answer:**

WeChat: cstutorcs

1. $\pi_{lastName,result,title}(\sigma_{result \leq 1.3}((\texttt{Students} \bowtie_{\texttt{Students}.matNr=\texttt{Exams}.matNr} \texttt{Exams}) \bowtie_{\sigma_{\texttt{Exams}.crsNr=\texttt{Courses}.crsNr}} \texttt{Courses}))$
2. $\pi_{lastName,result,title}(\sigma_{result \leq 1.3}(\sigma_{\text{EXAMS}.crsNr=\texttt{Courses}.crsNr}(\sigma_{\texttt{Students.matNr=Exams.matNr}}(\texttt{Students} \times \texttt{Exams} \times \texttt{Courses}))))$
3. $\pi_{lastName,result,title}((\texttt{Students} \bowtie_{\texttt{Students}.matNr=\texttt{Exams}.matNr} (\sigma_{result \leq 1.3}(\texttt{Exams}))) \bowtie_{\sigma_{\texttt{Exams}.crsNr=\texttt{Courses}.crsNr}} \texttt{Courses})$

Query Trees

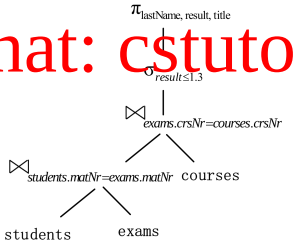
Assignment Project Exam Help

- Each RA expression can be represented as a **query tree**:
  - **leaf nodes** represent the input relations;
  - **internal nodes** represent the intermediate result;
  - **the root node** represents the resulting relation.
- **Example:**

https://tutorcs.com

$\pi_{lastName, result, title}(\sigma_{result \leq 1.3}((\texttt{Students} \bowtie_{\texttt{Students}.matNr = \texttt{Exams}.matNr} \texttt{Exams})$
$\bowtie \sigma_{\texttt{Exams}.crsNr = \texttt{Courses}.crsNr} \texttt{Courses}))$

WeChat: cstutorcs



Query Trees

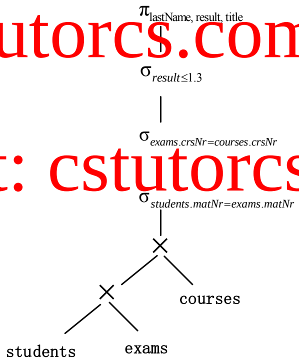
Assignment Project Exam Help

- **Exercise:** Can you draw the query tree for the following RA expression?

$$\pi_{lastName, result, title}(\sigma_{result \leq 1.3}(\sigma_{\texttt{Exams.crsNr=Courses.crsNr}}(\sigma_{\texttt{Students.matNr=Exams.matNr}}(\texttt{Students} \times \texttt{Exams} \times \texttt{Courses}))))$$

https://tutorcs.com

WeChat: cstutorcs



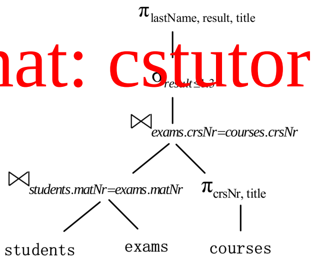
Query Trees

Assignment Project Exam Help

- For each query tree, computation proceeds **bottom-up**:
  - child nodes must be executed before their parent nodes;
  - but there can exist multiple methods of executing sibling nodes, e.g.,
    - process sequentially,
    - process in parallel.

https://tutorcs.com

WeChat: cstutorcs



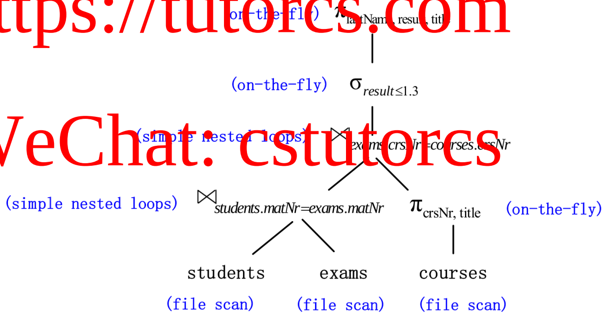
Execution Plan

Assignment Project Exam Help

- A **query execution plan** consists of an (extended) query tree with additional annotation at each node indicating:
  - (1) the *access method* to use for each table, and
  - (2) the *implementation method* for each RA operator.

https://tutorcs.com

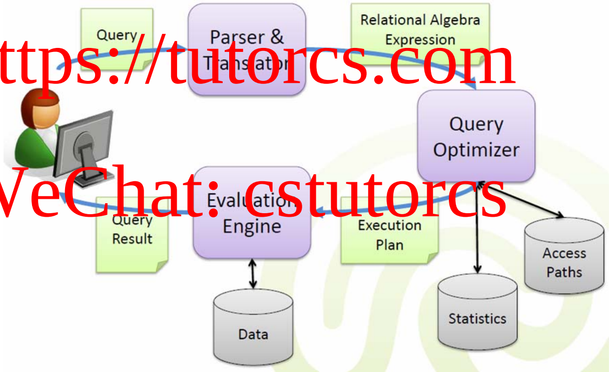
WeChat: cstutorcs



Query Processing – Evaluation Engine

Assignment Project Exam Help

- The **evaluation engine** executes an execution plan, and returns the query answer to the user.



https://tutorcs.com
WeChat: cstutorcs