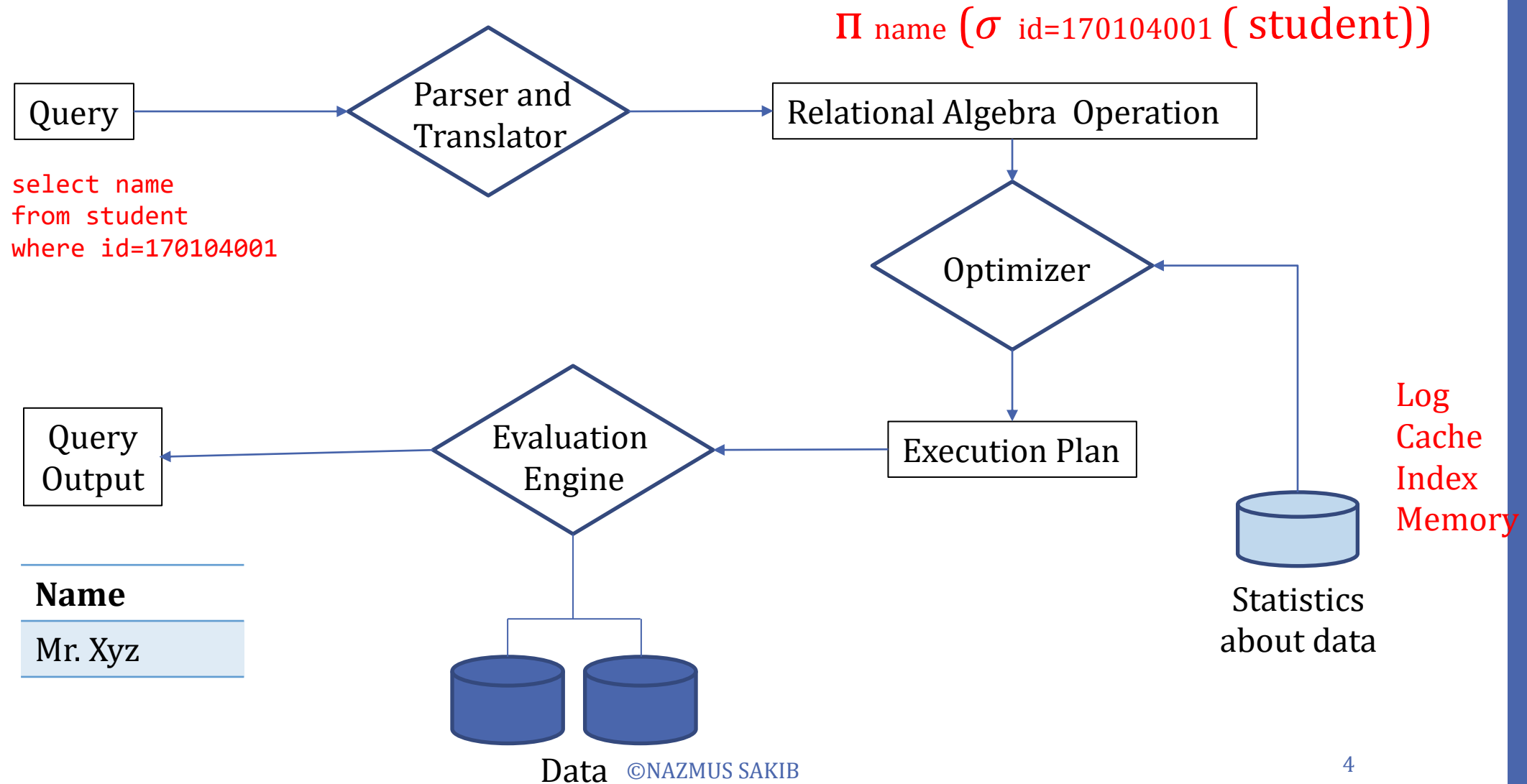DATABASE

CSE3103 : Database
FALL 2020

Nazmus Sakib
Assistant Professor
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Query Processing

Query processing refers to the range of activities involved in extracting data from a database. The basic steps involved in processing a query are:

- Parsing and Translation
- Optimization
- Evaluation

Query Processing

$$\Pi_{name}\left(\sigma_{id=170104001}(student)\right)$$

```
Query
```

select name
from student
where id=170104001

Parser and Translator

Relational Algebra Operation

Optimizer

Execution Plan

Log Cache Index Memory

Statistics about data

Evaluation Engine

Query Output

Name
Mr. Xyz

Data ©NAZMUS SAKIB

4

Basic Steps in Query Processing

- 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.

Basic Steps in Query Processing : Optimization

- A relational algebra expression may have many equivalent expressions
  - **E.g., $\sigma_{salary<75000}(\prod_{salary}(instructor))$ is equivalent to $\prod_{salary}(\sigma_{salary<75000}(instructor))$**

- 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

Measures of Query Cost

- Cost is generally measured as total elapsed time for answering query
  - Many factors contribute to time cost
    - *disk accesses, CPU*, or even network *communication*

- Typically disk access is the predominant cost, and is also relatively easy to estimate. Measured by taking into account
  - **Number of seeks   * average-seek-cost**
  - **Number of blocks read   * average-block-read-cost**
  - **Number of blocks written   * average-block-write-cost**

  - Cost to write a block is greater than cost to read a block
    - data is read back after being written to ensure that the write was successful

Measures of Query Cost

- For simplicity we just use the **number of block transfers** *from disk and the* **number of seeks** as the cost measures

  - t_T – **time to transfer one block**
  - t_S – **time for one seek**

  - Cost for b block transfers plus S seeks
    $$b * t_T + S * t_S$$

- We ignore CPU costs for simplicity
  - Real systems do take CPU cost into account

Let's Have a Math

- In a disk, there are 125 blocks of code and to build the program it requires 21 seeks. Time to transfer a block of code and time for one individual seek are 7ms and 11ms respectively. What is the cost of the query processing?

- Here,

b = 125

t_T = 7 ms

S = 21

t_S = 11ms

So,
$B * t_T + S * t_S$ = 125 * 7 *210 *11
= 202125 ms
= 202.125 s