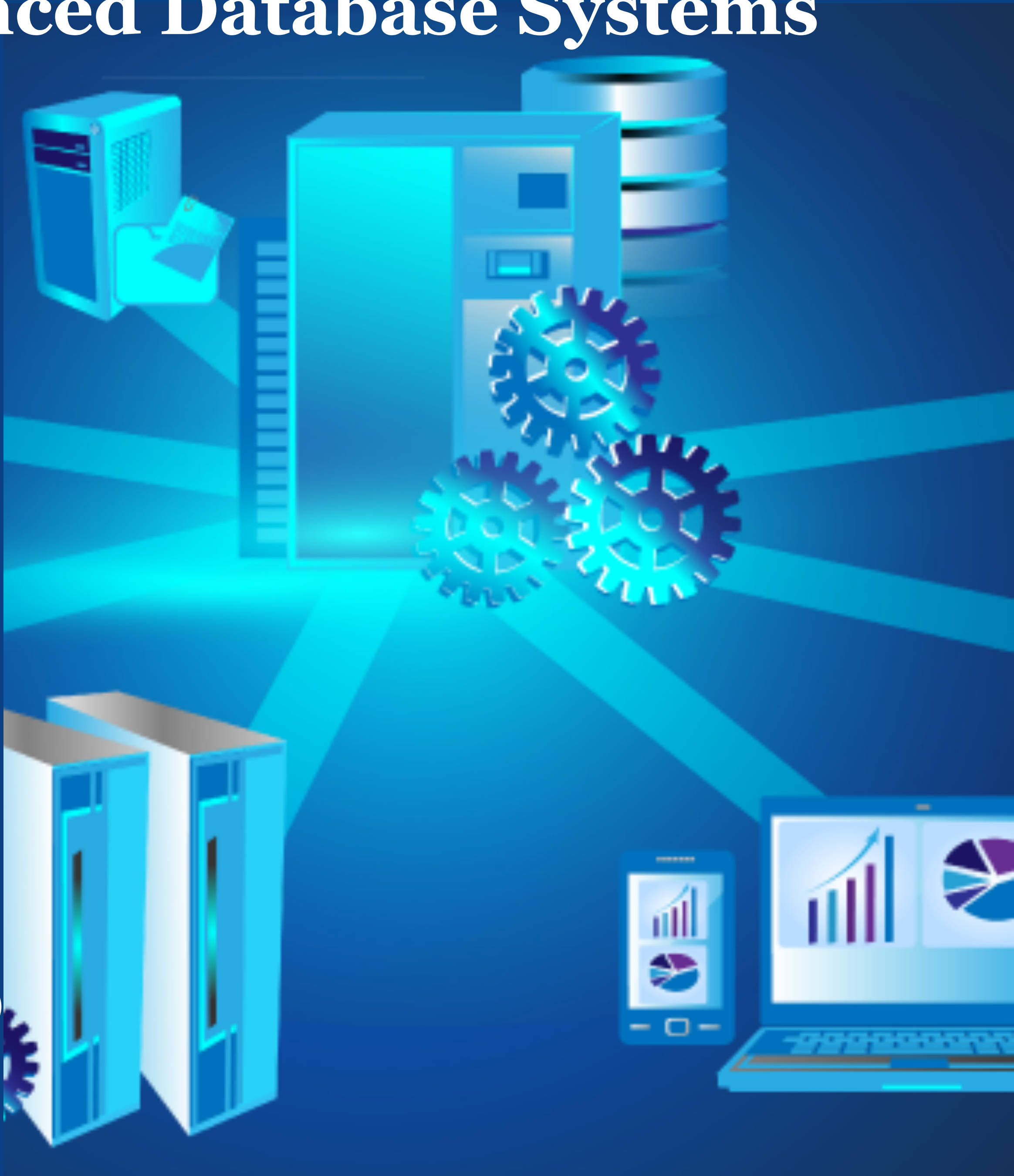
THE UNIVERSITY OF
MELBOURNE

COMP90050
Advanced Database Systems

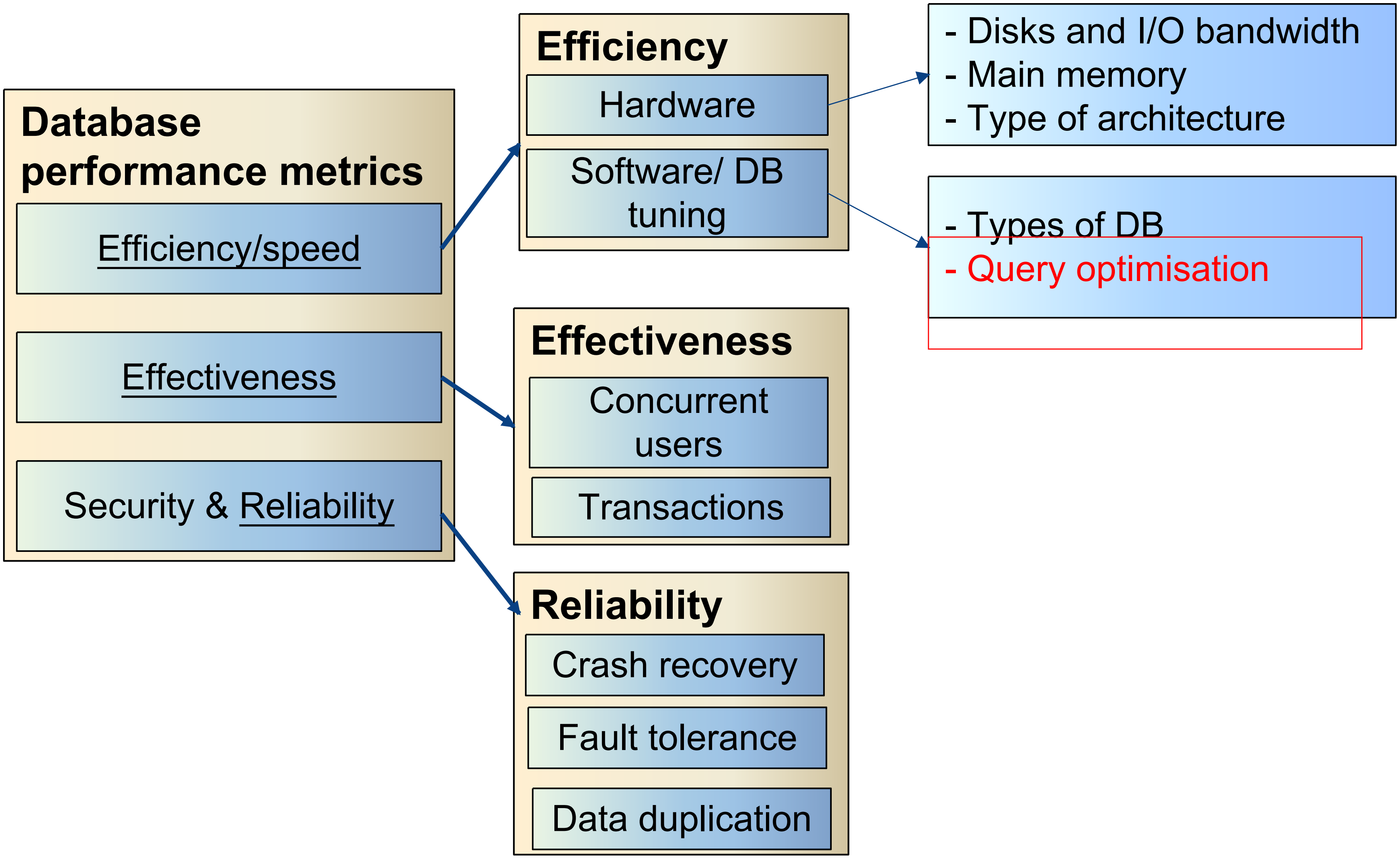


Semester 2, 2024

Lecturer: Farhana Choudhury (PhD)

Live lecture – Week 2

Core Concepts of Database management system



Database performance metrics

Efficiency/speed

Effectiveness

Security & Reliability

Efficiency

Hardware

Software/ DB tuning

- Disks and I/O bandwidth
- Main memory
- Type of architecture

- Types of DB
- Query optimisation

Effectiveness

Concurrent users

Transactions

Reliability

Crash recovery

Fault tolerance

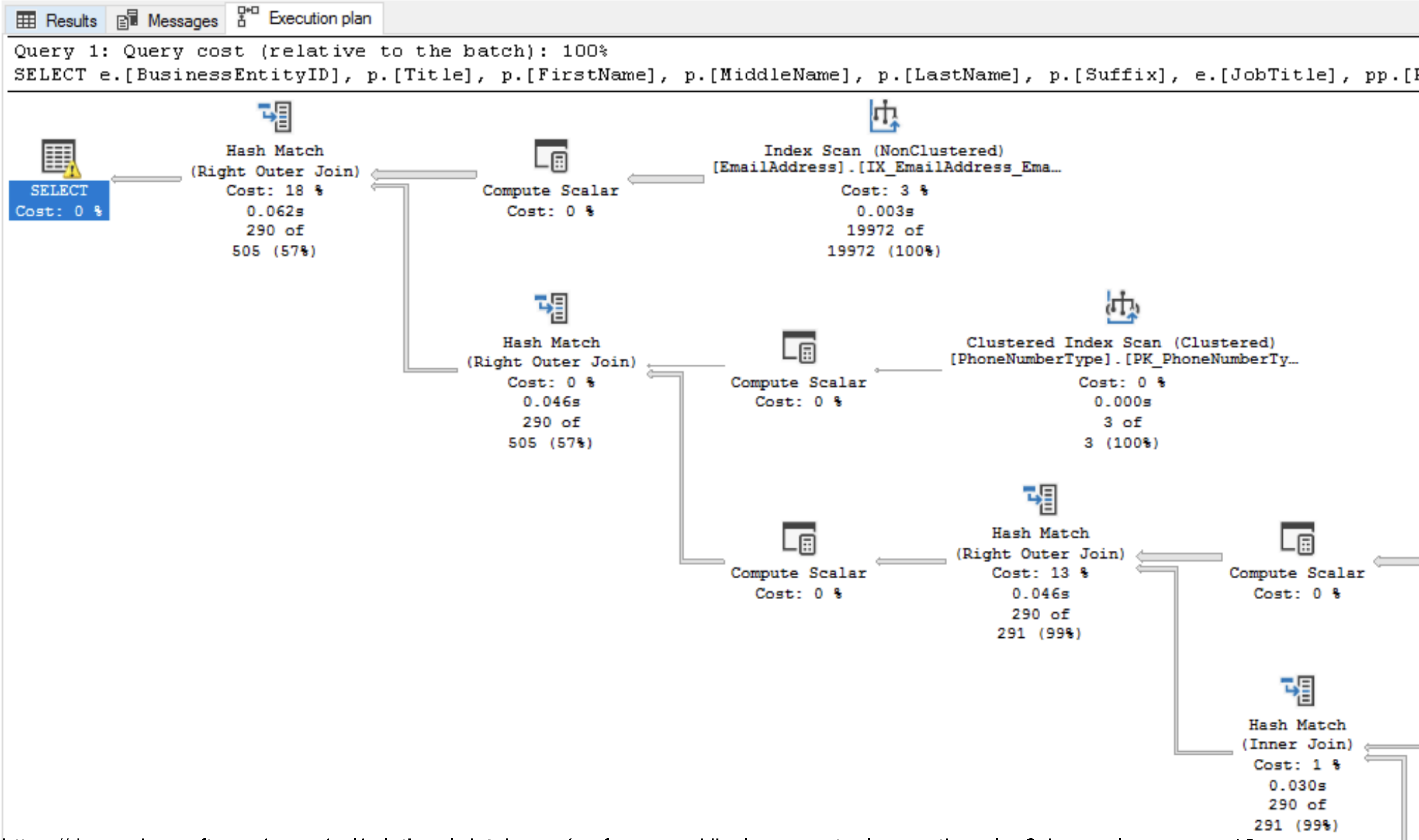
Data duplication

How do query optimiser make the choices?

Steps in **cost-based query optimisation**

1. Generate logically equivalent expressions of the SQL statement
2. Annotate resultant expressions to get alternative query plans
3. Choose the cheapest plan based on the estimated cost

View an actual execution plan



```
Query 1: Query cost (relative to the batch): 100%
SELECT e.[BusinessEntityID], p.[Title], p.[FirstName], p.[MiddleName], p.[LastName], p.[Suffix], e.[JobTitle], pp.[F
```

https://docs.microsoft.com/en-us/sql/relational-databases/performance/display-an-actual-execution-plan?view=sql-server-ver16

How do query optimizer make the choices?

Steps in **cost-based query optimisation**

1. Generate logically equivalent expressions of the SQL statement
2. Annotate resultant expressions to get alternative query plans
3. Choose the cheapest plan based on the estimated cost

disk I/Os mainly

What is the cost of a query?

The number of pages/ disk blocks that are accessed from disk to answer the query

⇒ 若能排好了的table ⇒ 只要retrieve part of them即可而不用全部
这些都会影响到你的选择

Most dominant cost

(Recall from memory hierarchy!

(read data from disk will spend 100 times longer than read data on memory)

minimize the most dominant cost with reduce entire largely.

How do query optimiser make the choices?

Steps in **cost-based query optimisation**

1. Generate logically equivalent expressions of the SQL statement

2. Annotate resultant expressions to get alternative query plans

3. Choose the cheapest plan based on the estimated cost

Estimation of plan cost based on:

- Statistical information about tables. Example:

  number of distinct values for an attribute

- Statistics estimation for intermediate results to compute cost of complex expressions

- Cost formulae for algorithms, computed using statistics again

Query optimisation

For a cost-based query optimization, is the query optimiser optimistic or pessimistic about the query cost estimation?

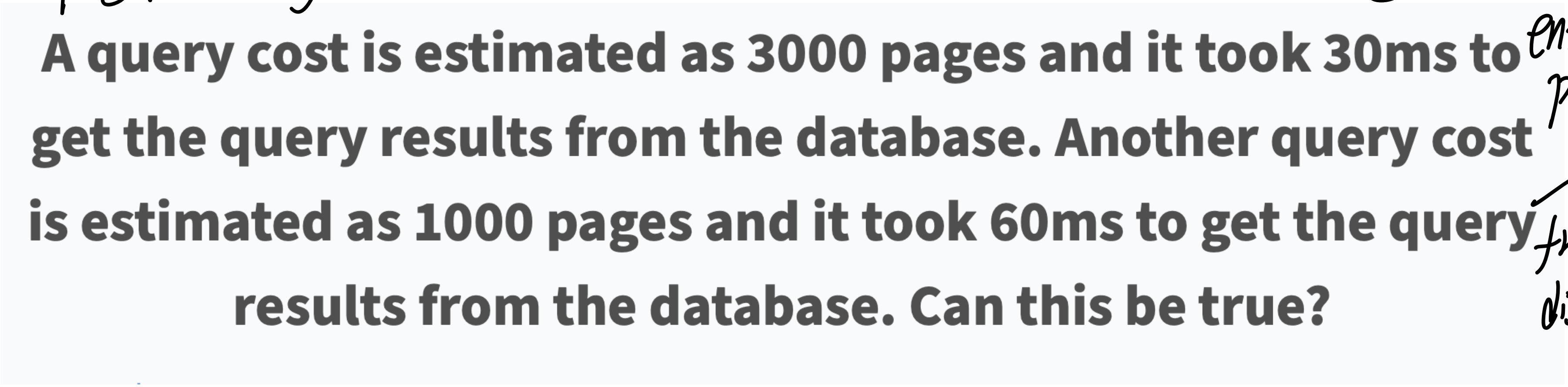
Time for a poll - Pollev.com/farhanachoud585

Query optimisation

For a cost-based query optimization, is the query optimiser optimistic or pessimistic about the query cost estimation?

ans: based on estimation, query plan will choose the one ~~10 the~~ with 1000 pages estimation.

query plan always consider the worst case scenario, it will think that the query will retrieve entire 3000 pages from disk.



A query cost is estimated as 3000 pages and it took 30ms to get the query results from the database. Another query cost is estimated as 1000 pages and it took 60ms to get the query results from the database. Can this be true?

[illegible]

What if some pages are already in main memory? – Difference between logical reads and physical reads

even if some pages are already in main memory (or all of them) so it chooses the plan that gives you smaller page #.

我们能告诉 optimizer data 在main memory上了吗?

When a query is actually being executed, it will check data which part of are already on cache or memory, then it doesn't need to retrieve that. However In planny step; It doesn't do that.

reason: The main memory always change, when the query being executed, main memory could be different compare with the moment when it check. (because many users, apps ~~not~~ using the DB on the same disk and they continously load data into main memory. So the data on main memory ~~Ean~~ at this moment might be changed or remove instantly ~~a~~ next moment, leading to worse estimation of query cost

Solution: memory optimized table (create this in main memory)

if ~~some~~ the table(s) can be frequently used retrieved => can tell DB to keep (it) the table in me as much Long as possible

当告诉 optimizer ~~因~~ 在main memory上有这个表时，然后呢？

Optimizer will be informed some tables very likely be found in main memory, when it do cost calculation, these tables will ~~be~~ become free of cost. => query plan will change.

Memory optimised table

Can you make a table stay in main memory?

```
CREATE TABLE dbo.Customer (
    CustomerID char (5) NOT NULL PRIMARY KEY,
    ContactName varchar (30) NOT NULL
) WITH (MEMORY_OPTIMIZED=ON)
```

Query processing and query optimisation for memory optimised tables?

Performance tuning

When you identify a query with suboptimal performance, what can you do?

Time for another poll - Pollev.com/farhanachoud585

Source: https://docs.microsoft.com/en-us/sql/relational-databases/performance/best-practice-with-the-query-store?view=sql-server-ver16#Insight

Performance tuning

When you identify a query with suboptimal performance, what can you do?

Time for another poll - Pollev.com/farhanachoud585

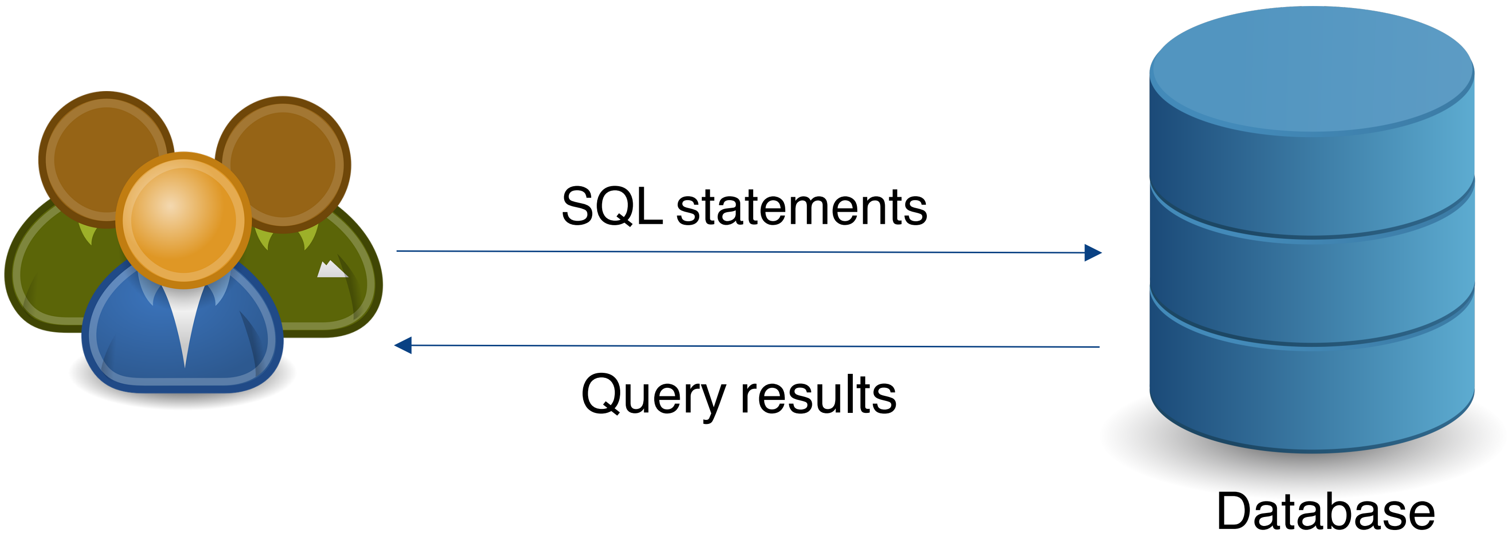
- Force a query plan instead of the plan chosen by the optimizer
- Do we need an index?
- Enforce statistic recompilation
- Rewrite query? ✓
- Memory optimised table? ☆ ✓: if there are some tables that are frequently used, we can creat Mot. (we can).

Limite the scope of selection before join.

要是在整 memory optimise whole database, we cannot see any improvement ⇒ no enough memory.

Source: https://docs.microsoft.com/en-us/sql/relational-databases/performance/best-practice-with-the-query-store?view=sql-server-ver16#Insight

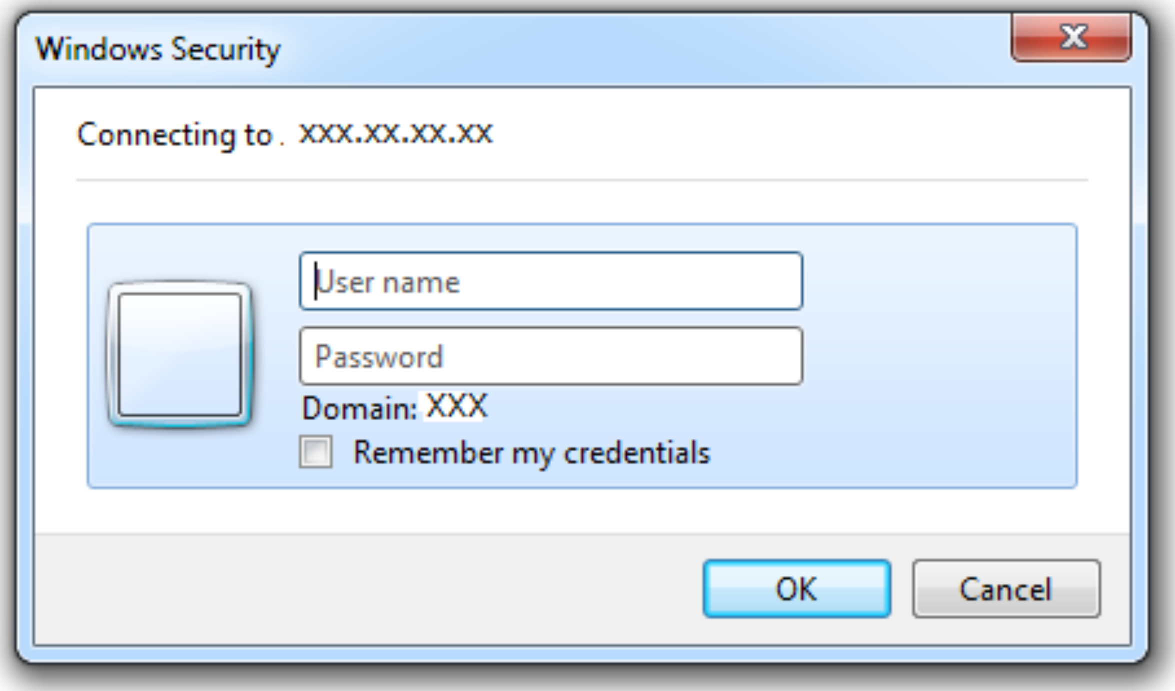
Discussion topic – SQL injection



Can happen when an application executes database query using user-input data, and the user input or part of the user input is treated as SQL statement

SQL syntax

SELECT * FROM `login` WHERE `user`=‘farhana’ AND `pass`=‘comp90050’



Windows Security

Connecting to . XXX.XX.XX.XX

User name

Password

Domain: XXX

□ Remember my credentials

LOGIC: ‘a’=‘a’

Example: SELECT * FROM `table` WHERE ‘a’=‘a’

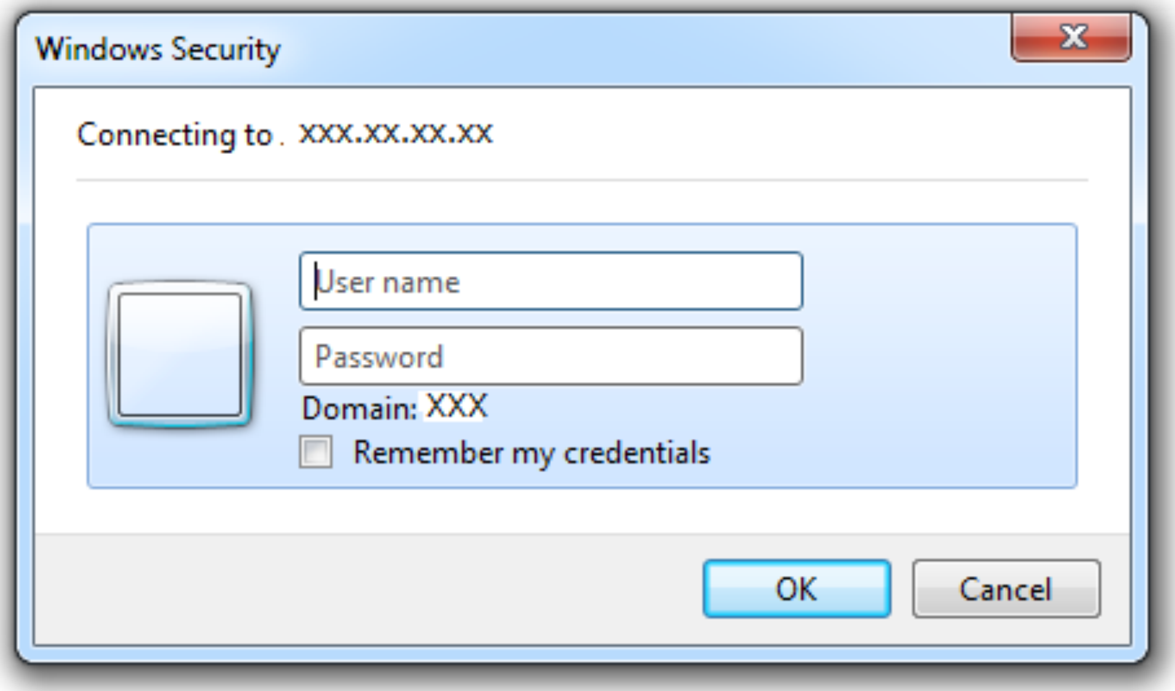
SELECT * FROM `login` WHERE `user`=‘’ OR ‘a’=‘a’ AND `pass`=‘’ OR ‘a’=‘a’

SQL syntax

MULTI STATEMENTS: S1; S2

Example: SELECT * FROM `table`; DROP TABLE `table`;

SELECT * FROM `login` WHERE `user`=‘farhana’ AND `pass`=‘comp90050’



Windows Security

Connecting to . XXX.XX.XX.XX

User name

Password

Domain: XXX

□ Remember my credentials

Any statement(s) can go here

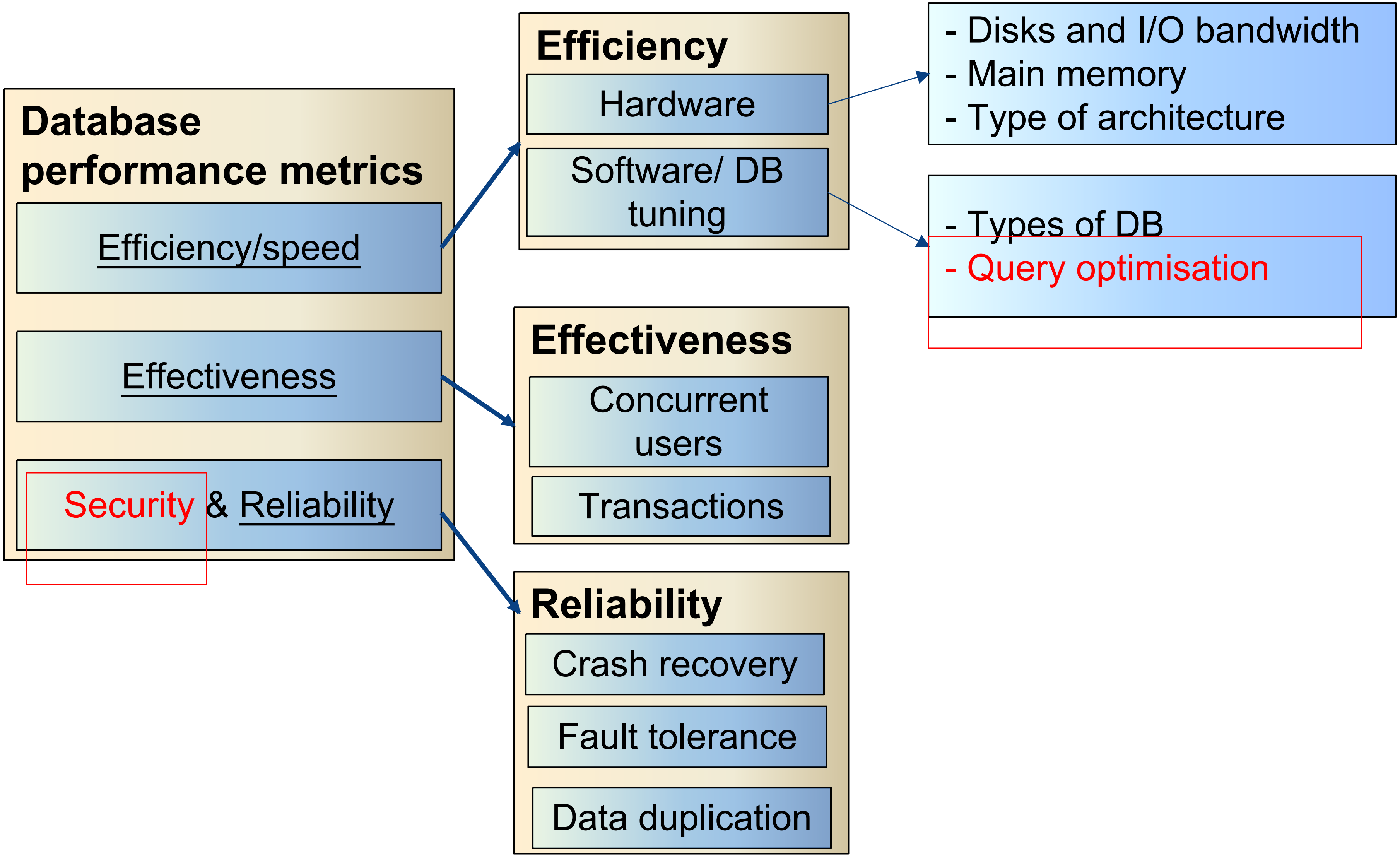
SELECT * FROM `login` WHERE `user`=‘’; DROP TABLE `login`; --’ AND `pass`=‘’

Prevention

User parameterized query/prepared statement - allows the database to distinguish between code and data

```
String query = "SELECT * from login where user = "   + request
.getParameter(”userName");
```

Summary



Database performance metrics

- Efficiency/speed
- Effectiveness
- Security & Reliability

Efficiency

- Hardware
- Software/ DB tuning

- Disks and I/O bandwidth
- Main memory
- Type of architecture

- Types of DB
- Query optimisation

Effectiveness

- Concurrent users
- Transactions

Reliability

- Crash recovery
- Fault tolerance
- Data duplication