# Indexing and Query optimization

Hogeschool Rotterdam
Rotterdam, Netherlands

# Introduction



## Lecture topics

- Query optimization.
- Examples of slow query operations.
- Hashing.
- Trees.

# Query optimization



## Reasons

- Query needs to be fast.
- Sometimes they are not.
- You do not want to see your nephew born before retrieving the book you are looking for from Amazon.

# Query optimization



## Causes

- Too much data (Big data analysis).
  - Data clustering.
  - Better hardware. (arrays of disks, caching, ...)
- Too complex queries (DBMS optimization)
  - Refactor query. (Access planner)
  - Refactor data. (Indexing)

# Query optimization



## Indexes

- Query refactoring not always possible.
- Build additional data to speed up the data retrieval.

# Query optimization



## Indexes

- Take your text book and look for the paragraph titled "Key constraints" without using the index. How many pages have you looked?

# Query optimization



## Indexes

- Take your text book and look for the paragraph titled "Key constraints" without using the index. How many pages have you looked?
- **Answer:** 29 (from page 3 to 32).

# Query optimization



## Indexes

- Take your text book and look for the paragraph titled "Key constraints" without using the index. How many pages have you looked?
- **Answer:** 29 (from page 3 to 32).
- Do the same using the index. How many pages have you looked?

# Query optimization



## Indexes

- Take your text book and look for the paragraph titled "Key constraints" without using the index. How many pages have you looked?
- **Answer:** 29 (from page 3 to 32).
- Do the same using the index. How many pages have you looked? **Answer:** 2 (1 in the index, 1 in the text).

# Where



```
SELECT name
FROM ships
WHERE firepower >= 500
```

| ships | | | | |
|---|---|---|---|---|
| name | type | firepower | speed | position |
| Red 1 | X-Wing | 10 | 300 | (1,3,1) |
| Red 2 | X-Wing | 10 | 300 | (1,2,1) |
| Red 3 | X-Wing | 10 | 300 | (0,2.5,1) |
| Red 4 | X-Wing | 10 | 300 | (2,2.5,1) |
| Red 5 | X-Wing | 10 | 300 | (2,2.5,0) |
| Red 6 | X-Wing | 10 | 300 | (2,2.5,0) |
| Tantine IV | Corellian Corvette | 60 | 300 | (4,2.5,0) |
| Tyranny | Imperial Star Destroyer | 1500 | 100 | (12,0,0) |
| Accuser | Imperial Star Destroyer | 1500 | 100 | (-12,0,0) |
| Bombard | Victory Star Destroyer | 500 | 175 | (-6,1,0) |

# Where



```
SELECT name
FROM ships
WHERE firepower >= 500
```

| ships | | | | |
|---|---|---|---|---|
| name | type | firepower | speed | position |
| Red 1 | X-Wing | 10 | 300 | (1,3,1) |
| Red 2 | X-Wing | 10 | 300 | (1,2,1) |
| Red 3 | X-Wing | 10 | 300 | (0,2.5,1) |
| Red 4 | X-Wing | 10 | 300 | (2,2.5,1) |
| Red 5 | X-Wing | 10 | 300 | (2,2.5,0) |
| Red 6 | X-Wing | 10 | 300 | (2,2.5,0) |
| Tantine IV | Corellian Corvette | 60 | 300 | (4,2.5,0) |
| Tyranny | Imperial Star Destroyer | 1500 | 100 | (12,0,0) |
| Accuser | Imperial Star Destroyer | 1500 | 100 | (-12,0,0) |
| Bombard | Victory Star Destroyer | 500 | 175 | (-6,1,0) |

Number of comparisons: 10

# WHERE



## Indexes

- How many comparisons we do at most in a table with R records?

# WHERE



## Indexes

- How many comparisons we do at most in a table with R records?
- **R comparisons**

# WHERE



## Indexes

- How many comparisons we do at most in a table with R records?
- **R comparisons**
- How many comparisons we do at least in a table with R records?

# WHERE



## Indexes

- How many comparisons we do at most in a table with R records?
- **R comparisons**
- How many comparisons we do at least in a table with R records?
- **R comparisons**

# WHERE



- Sorting and grouping requires to sort the column values.
- The best sorting algorithm requires about $R \log R$ operations, where R is the number of records.
- Running the query below requires about $10 * \log 10 \simeq 23$ operations.

```
SELECT type
FROM ships
WHERE firepower >= 500
ORDER BY firepower DESC
```

# JOIN



- Generate pairs with one element from the first table and the second from the other followed by a selection.
- Same problem of the selection.
- Consider the following query applied to ship and the table below.

```
SELECT s.name,p.damage
FROM ships s,projectiles p
WHERE s.position = p.position AND
      s.name = p.target
```

| Projectiles | | |
|---|---|---|
| target | position | damage |
| Tantine IV | (0,1,0) | 30 |
| Tantine IV | (3,1,-2) | 50 |
| Tantine IV | (4,2.5,0) | 100 |