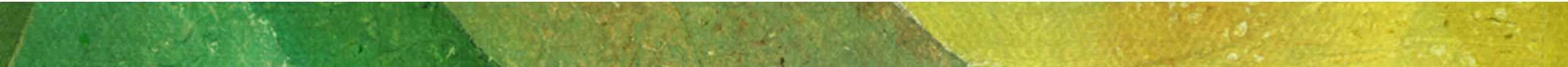
Hard Disk Drive and Static Indexes

NDBI007: Assignment 1



Task 1

- Define a data structure for the *Aircraft* record that has the following attributes:
  - Unique aircraft *identifier* (*primary key*)
  - Aircraft *manufacturer identifier* (suppose there are only 4 aircraft manufacturers in the world)
  - Aircraft *serial number* (not necessary unique value, assigned by the aircraft manufacturer)
  - In addition, at least *6 more attributes* (e.g., binary, numeric or string values)
    - E.g., see https://schema.org/Vehicle

- *Draw any diagram* of the proposed data structure and include the size of each attribute (in bytes)
- Also review the assignments for all tasks to verify that your proposed Airplane data structure meets the requirements from all tasks
  - I.e., appropriate secondary index selection and bitmap utilization

- **Points:** 0 (note that the data structure is only required to solve the remaining tasks)

- For all the remaining tasks, assume we have 20,000,000 Aircraft records stored in the database
  - Block size is 4 kB
  - Pointer size is 4 B

- Before the computation, round the size of Aircraft data structure to the lowest higher power of 2

Task 2

- Propose a static index to search for records about specific aircraft if we have a primary dataset stored sequentially and sorted by the value of the primary key, namely:
  A) Primary key index
  B) Secondary key direct index (i.e., appropriately select a secondary key)
  C) Secondary key indirect index

- For all cases A), B), and C), determine the *height of the index*, the *number of blocks* in which each level is stored, and the *total size* of the index in MB, and *compare* all indexes with each other
- In addition, in case B) also *justify the choice* of the attribute to create the *secondary key*
  - I.e., not all attributes are suitable for use

- **Points:** 2.5

Task 3

- Having an index from the task 2A), *determine the time required to read* information about one aircraft from the sequentially stored file on an IBM Deskstar HDD (75 GB)
  - You may use the following formulas to estimate the time:

$$T = 2 \bullet (s + r + btt) + 2 \bullet r + btt$$

$$btt = \frac{2 \bullet r}{TC} \bullet B$$

- Most of the required values can be found in the IBM Deskstar HDD white paper (see practical class website)
  - As for the track capacity, use $TC = 0.3\ MB$

- Points: 1

Task 4

- Propose two *appropriate ways to use a bitmap* to store the values of a selected attribute from the Aircraft data structure
  - *Justify the choice* of attribute over which to create the bitmap
  - *Compare the two methods* (e.g., estimate theirs size in MB, effectivity of read operations etc.)

- Points: 1.5

Bonus Task 5

- Describe how a *reading and writing data in an SSD* works
  - *Compare* this method with reading and writing data in HDD
  - Indicate *typical use cases* for HDDs and SSDs
    - Compare the typical cases and justify your answer

- When citing any source in your answer, please provide a reference (or a citation) to the literature

- Points: 1

Submission

- The solution can be handed in by email or on paper before the beginning of the next practical class
  - email: pavel.koupil@matfyz.cuni.cz

- Deadline: By the start of the next practical class where you are *enrolled in SIS*