

Лекція 02: Зберігання даних та
файлова структура

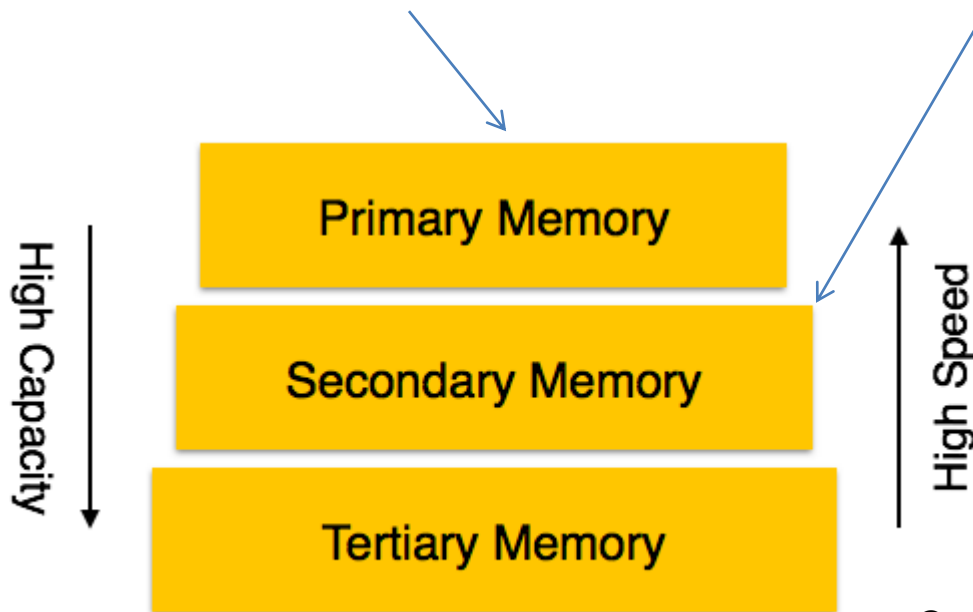
Lecture 02: Storage and file
structure

Система зберігання даних / Storage system

- Бази даних зберігаються у форматах файлів, що містять записи. На фізичному рівні фактичні дані зберігаються на деякому пристрої. Ці пристрої зберігання даних можна розділити на три типи.
- Databases are stored in file formats, which contain records. At physical level, the actual data is stored in electromagnetic format on some device. These storage devices can be broadly categorized into three types.

Первинна пам'ять – безпосередньо доступна для CPU
The memory storage that is directly accessible to the CPU.

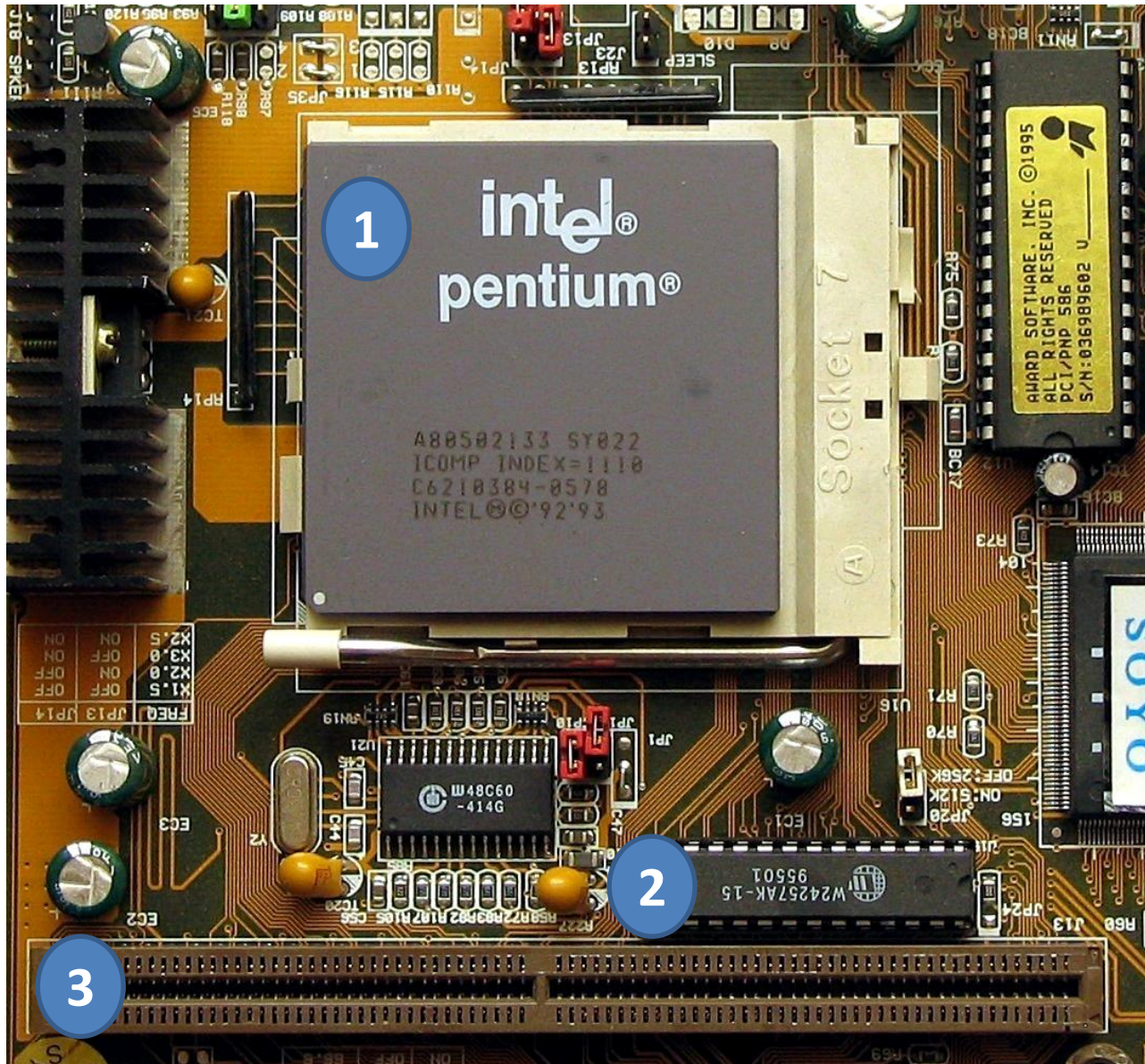
Вторинна пам'ять – використовується для збереження даних з метою їх подальшого використання, або в якості бекапу.
Secondary storage devices are used to store data for future use or as backup.



Стороння пам'ять – використовується для зберігання великих обсягів даних, наприклад, для резервного копіювання всієї системи.
Tertiary storage is used to store huge volumes of data. These storage devices are mostly used to take the back up of an entire system.

Первинна пам'ять / Primary storage

- Пам'ять, що безпосередньо доступна для CPU, потрапляє до цієї категорії. Внутрішня пам'ять процесора (registers), швидка пам'ять (cache) та основна пам'ять (RAM) безпосередньо доступні для CPU, оскільки всі вони розміщені на материнській платі чи на чіпсеті CPU. Цей накопичувач, як правило, дуже малий, надшвидкий та нестабільний. Основне сховище вимагає постійного джерела живлення, щоб підтримувати його стан. У випадку відключення живлення всі його дані втрачаються.
- The memory storage that is directly accessible to the CPU comes under this category. CPU's internal memory (registers), fast memory (cache), and main memory (RAM) are directly accessible to the CPU, as they are all placed on the motherboard or CPU chipset. This storage is typically very small, ultra-fast, and volatile. Primary storage requires continuous power supply in order to maintain its state. In case of a power failure, all its data is lost.



Вторинна пам'ять / Secondary storage

- Вторинні накопичувачі використовуються для зберігання даних для подальшого використання або резервного копіювання. Вторинна пам'ять включає в себе пристрої пам'яті, які не є частиною чіпсета CPU або материнської плати, наприклад, магнітні диски та стрічки, оптичні диски (DVD, CD та ін.), HDD, USB-накопичувачі та SSD.
- Secondary storage devices are used to store data for future use or as backup. Secondary storage includes memory devices that are not a part of the CPU chipset or motherboard, for example, magnetic disks and tapes, optical disks (DVD, CD, etc.), HDD, flash drives, and SSD.



<https://www.pcworld.com/article/2864385/are-pc-hard-drives-destined-to-die-at-the-hand-of-the-cloud-maybe-analysts-say.html>

Стороння пам'ять / Tertiary storage

- Сторонній накопичувач використовується для зберігання величезних обсягів даних. Оскільки такі пристрої зберігання знаходяться поза комп'ютером, вони є найбільш повільними. Ці пристрої зберігання даних найчастіше використовуються для резервного копіювання всієї системи.
- Tertiary storage is used to store huge volumes of data. Since such storage devices are external to the computer system, they are the slowest in speed. These storage devices are mostly used to take the back up of an entire system.



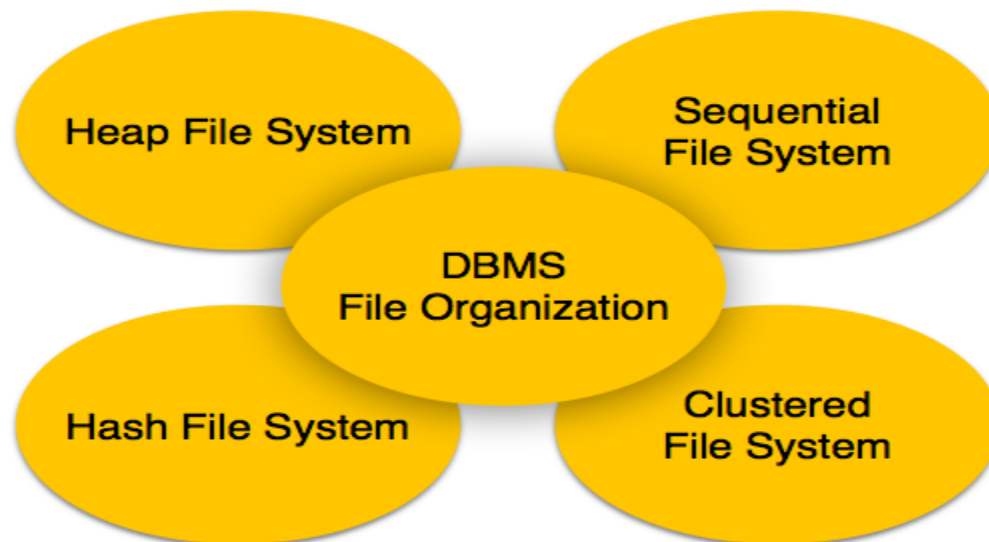
<https://www.bhphotovideo.com/explora/computers/buying-guide/recommended-external-hard-drives-photo-video-and-audio-production>

Файлова структура / File structure

- Пов'язані дані та інформація зберігаються разом у форматах файлів. Файл – це послідовність записів, що зберігаються у двійковому форматі. Дисковий привід відформатовано в кілька блоків, які можуть зберігати записи. Файлові записи відображаються на ці дискові блоки.
- Relative data and information is stored collectively in file formats. A file is a sequence of records stored in binary format. A disk drive is formatted into several blocks that can store records. File records are mapped onto those disk blocks.

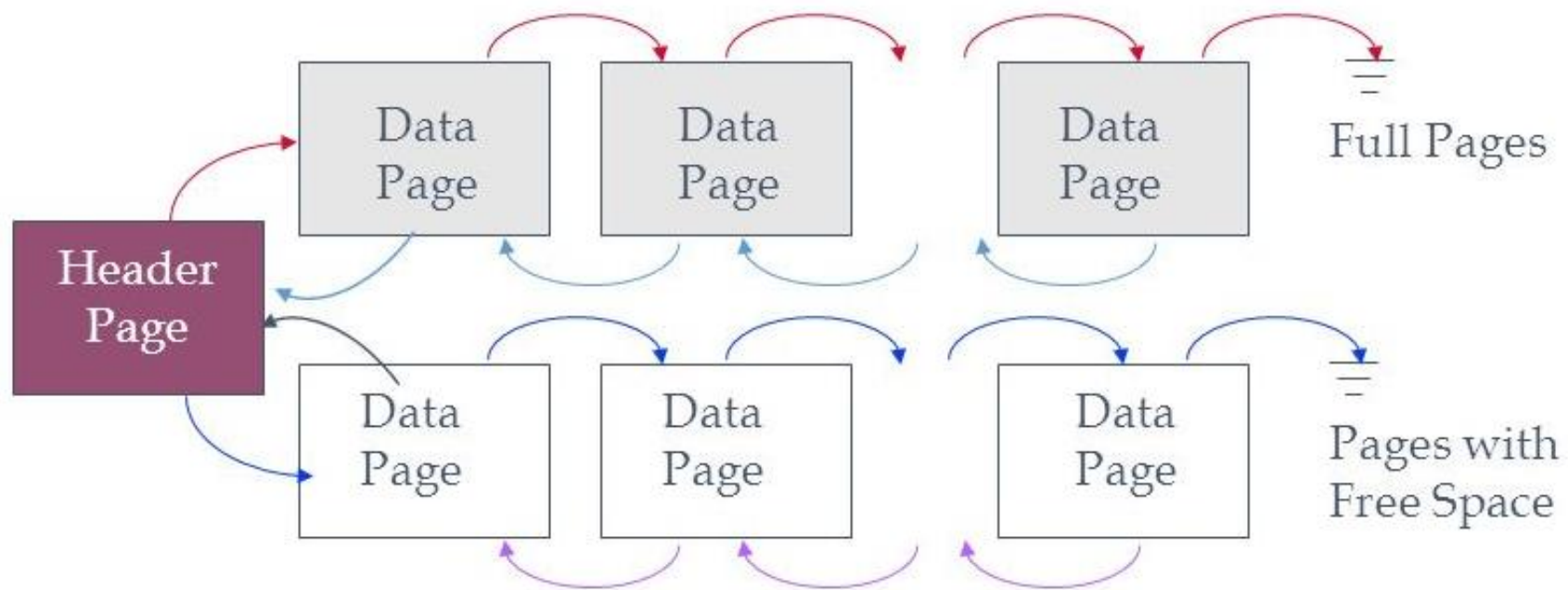
Організація файлів / File organization

- Організація файлів визначає спосіб записування файлів на блоки дисків. Існує чотири типи організації файлів для організації файлових записів.
- File Organization defines how file records are mapped onto disk blocks. We have four types of File Organization to organize file records.



Організація файлів у купі / Heap file organization

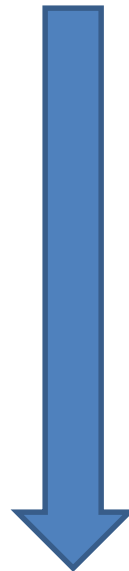
- Коли файл створюється за допомогою купи, операційна система виділяє область пам'яті для цього файлу без будь-яких додаткових облікових даних. Файлові записи можна розміщувати в будь-якій частині виділеної області пам'яті. Відповідальність за керування записами несе програмне забезпечення. Файлова купа не підтримує сортування, упорядкованість чи індексування самотійно.
- When a file is created using Heap File Organization, the Operating System allocates memory area to that file without any further accounting details. File records can be placed anywhere in that memory area. It is the responsibility of the software to manage the records. Heap File does not support any ordering, sequencing, or indexing on its own.



Послідовна організація файлів / Sequential file organization

- Кожний запис файлу містить поле даних (атрибут), щоб однозначно ідентифікувати цей запис. У послідовній організації файлів записи розміщуються у файлі в деякому послідовному порядку на основі унікального ключового поля або ключа пошуку. Практично неможливо зберігати всі записи послідовно у фізичній формі.
- Every file record contains a data field (attribute) to uniquely identify that record. In sequential file organization, records are placed in the file in some sequential order based on the unique key field or search key. Practically, it is not possible to store all the records sequentially in physical form.

A-217	Brighton	750	
A-101	Downtown	500	
A-110	Downtown	600	
A-215	Mianus	700	
A-102	Perryridge	400	
A-201	Perryridge	900	
A-218	Perryridge	700	
A-222	Redwood	700	
A-305	Round Hill	350	



Записи в файлі
впорядковані
за допомогою
ключа пошуку

The records in the file
are ordered by
a search-key

Хеш організація файлів / Hash file organization

- Хеш організація файлів використовує обчислення хеш-функцій на деяких полях записів. Вивід хеш-функції визначає розташування блоку на диску, де потрібно розмістити записи.
- Hash File Organization uses Hash function computation on some fields of the records. The output of the hash function determines the location of disk block where the records are to be placed.

bucket 0

bucket 1

15151	Mozart	Music	40000

bucket 2

32343	El Said	History	80000
58583	Califieri	History	60000

bucket 3

22222	Einstein	Physics	95000
33456	Gold	Physics	87000
98345	Kim	Elec. Eng.	80000

bucket 4

12121	Wu	Finance	90000
76543	Singh	Finance	80000

bucket 5

76766	Crick	Biology	72000

bucket 6

10101	Srinivasan	Comp. Sci.	65000
45565	Katz	Comp. Sci.	75000
83821	Brandt	Comp. Sci.	92000

bucket 7

Hash file organization of *instructor* file, using *dept_name* as key (see previous slide for details).

Кластерна організація файлів / Clustered file organization

- Кластерна організація файлів не вважається корисною для великих баз даних. У цьому механізмі пов'язані записи з одного або декількох відношень зберігаються в тому ж дисковому блоці, тобто упорядковування записів не ґрунтується на первинному ключі або ключі пошуку.
- Clustered file organization is not considered good for large databases. In this mechanism, related records from one or more relations are kept in the same disk block, that is, the ordering of records is not based on primary key or search key.

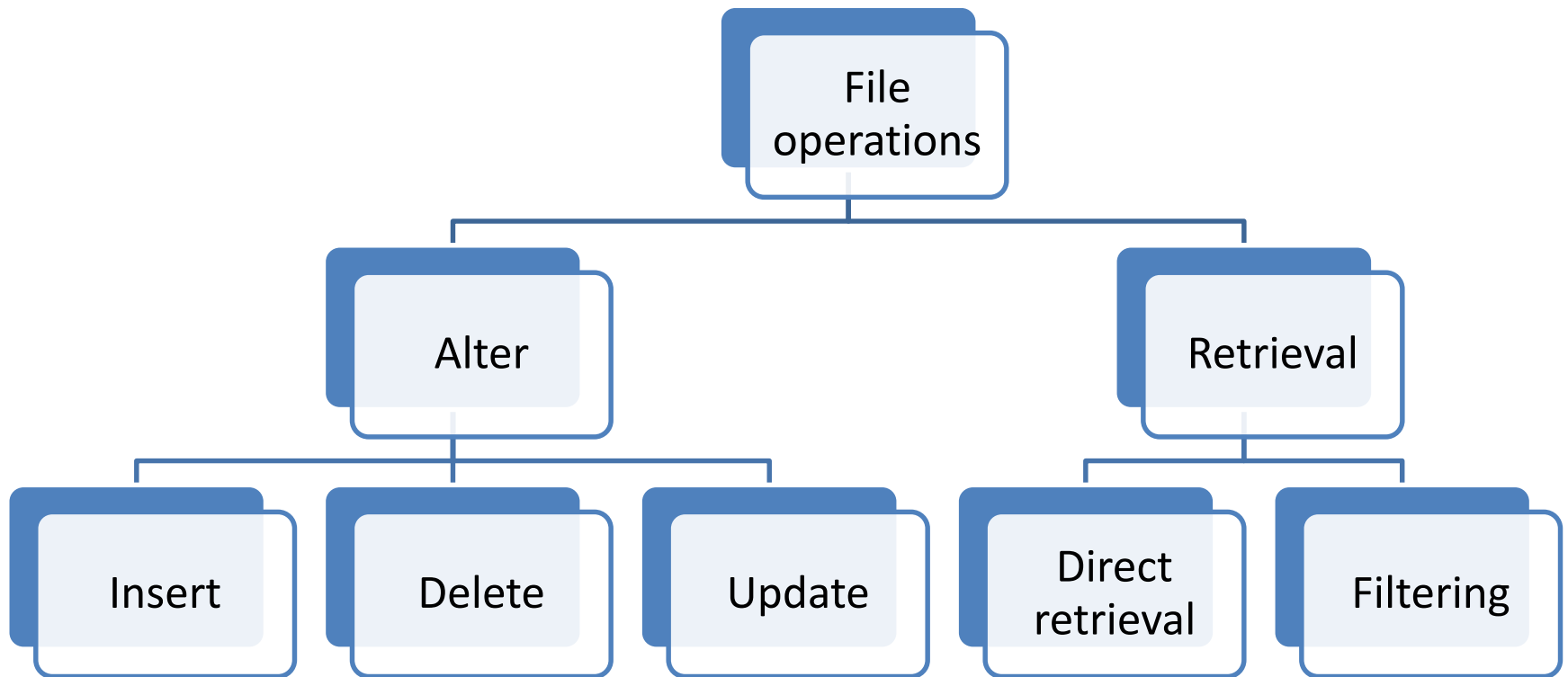
CLUSTER KEY



DEP_ID	DEP_NAME	EMP ID	EMP_NAME	EMP_ADD
D_101	ECO	01	JOE	CAPE TOWN
		02	PETER	CROY CITY
		03	MARY	NOVI
D_102	CS	04	JOHN	FRANSISCO
D_103	JAVA	05	ANNIE	FRANSISCO
D_104	MATHS	06	SAKACHI	TOKYO
D_105	BIO	07	SONI	W.LAND
D_106	CIVIL	08	LUNA	TOKYO

DEPARTMENT + EMPLOYEE

Файлові операції / File operations



Відкриття – файл може бути відкритий в одному з двох режимів: режим **читання** або режим **запису**. У режимі читання операційна система не дозволяє нікому змінювати дані. Інакше кажучи, дані лише для читання. Файли, відкриті в режимі читання, можуть бути розподілені між декількома сутностями. Режим запису дозволяє змінювати дані. Файли, відкриті в режимі запису, можна читати, але їх не можна поділити.

Open – A file can be opened in one of the two modes, **read** mode or **write** mode. In read mode, the operating system does not allow anyone to alter data. In other words, data is read only. Files opened in read mode can be shared among several entities. Write mode allows data modification. Files opened in write mode can be read but cannot be shared.

Переміщення – кожен файл має покажчик файлу, який показує поточну позицію, де дані потрібно читати чи записувати. Цей покажчик можна відповідно скорегувати. Використовуючи операцію пошуку, покажчик можна пересувати вперед або назад.

Locate – Every file has a file pointer, which tells the current position where the data is to be read or written. This pointer can be adjusted accordingly. Using find (seek) operation, it can be moved forward or backward.

Читання – за замовчуванням, коли файли відкриваються в режимі читання, покажчик файлу вказує на початок файлу. Існують варіанти, за яких користувач може повідомити операційній системі, де знаходити покажчик файлу під час відкриття файлу. Зчитуються наступні до покажчика файлу дані.

Read – By default, when files are opened in read mode, the file pointer points to the beginning of the file. There are options where the user can tell the operating system where to locate the file pointer at the time of opening a file. The very next data to the file pointer is read.

Запис – користувач може вибрати, щоб відкрити файл у режимі запису, який дає їм змогу редагувати його вміст. Це може бути видалення, вставка або модифікація. Показчик файлів може знаходитися під час відкриття або може бути динамічно змінений, якщо операційна система це дозволяє.

Write – User can select to open a file in write mode, which enables them to edit its contents. It can be deletion, insertion, or modification. The file pointer can be located at the time of opening or can be dynamically changed if the operating system allows to do so.

Закриття – це найважливіша операція з точки зору операційної системи. Коли створюється запит на закриття файлу, операційна система:

- видаляє всі замки (якщо в режимі спільного використання);
- зберігає дані (якщо вони змінені) на вторинні носії інформації;
- випускає всі буфери та обробники файлів, пов'язані з файлом.

Close – This is the most important operation from the operating system's point of view. When a request to close a file is generated, the operating system

- removes all the locks (if in shared mode);
- saves the data (if altered) to the secondary storage media;
- releases all the buffers and file handlers associated with the file.