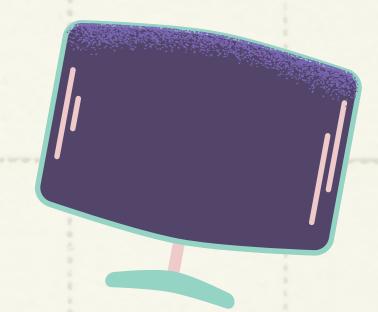


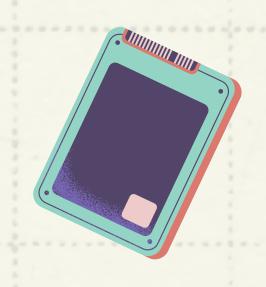
CONTENTS

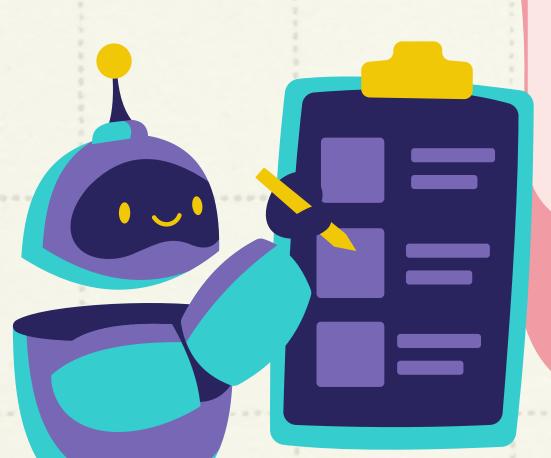


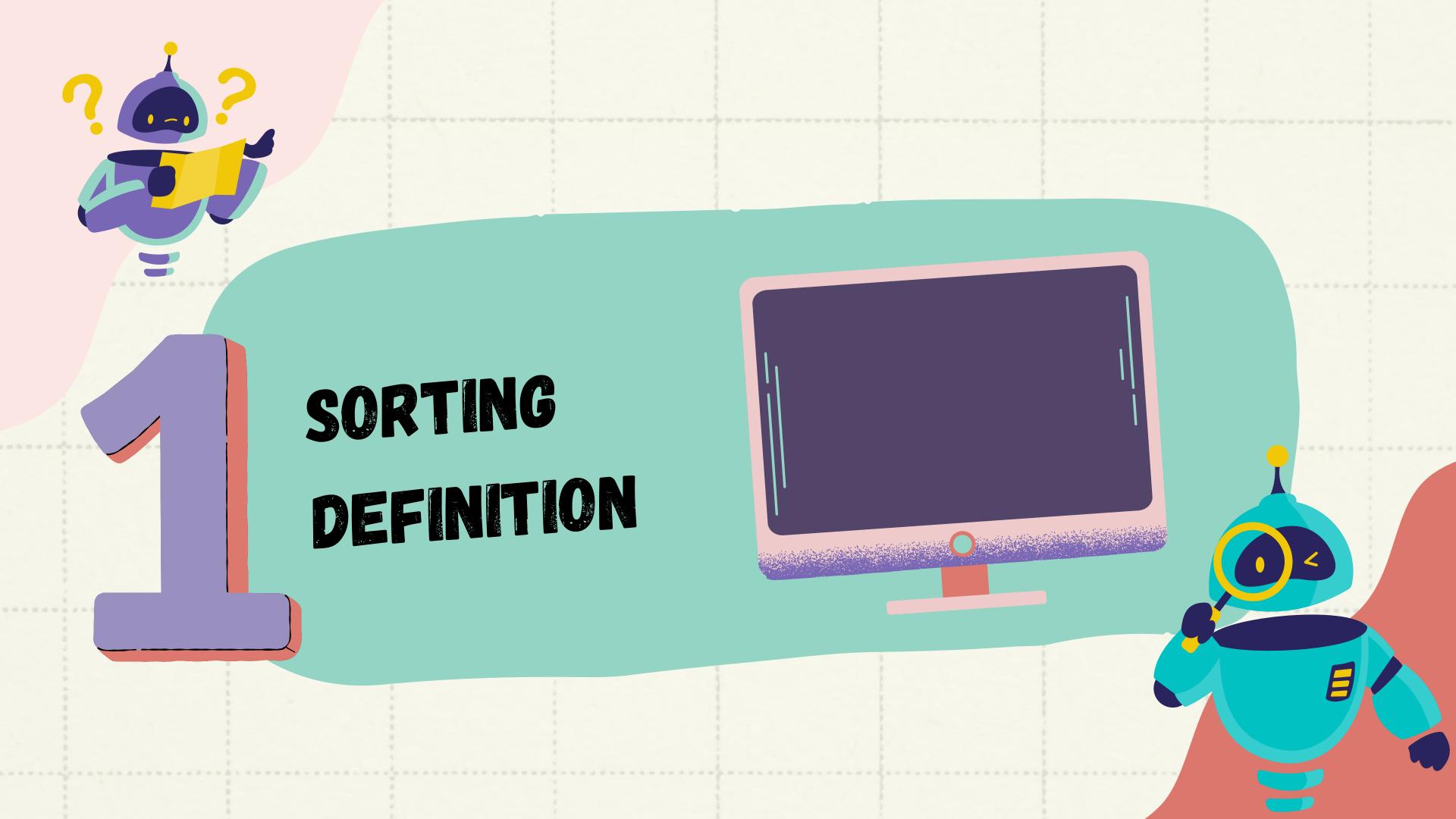
SORTING DEFINITION SORTING CATEGORIES









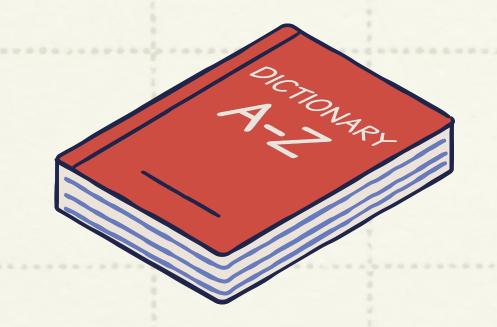


Bài toán sắp xếp

- •Sự sắp xếp dữ liệu theo một thứ tự nhất định
- •Dữ liệu cần sắp xếp có thể là các số nguyên, xâu kí tự hay các đối tượng...

Bài toán sắp xếp

- •Sự sắp xếp dữ liệu theo một thứ tự nhất định
- Dữ liệu cần sắp xếp có thể là các số nguyên, xâu kí tự hay các đối tượng...









SORTING CATERGORIES



ONLINE - OFFLINE

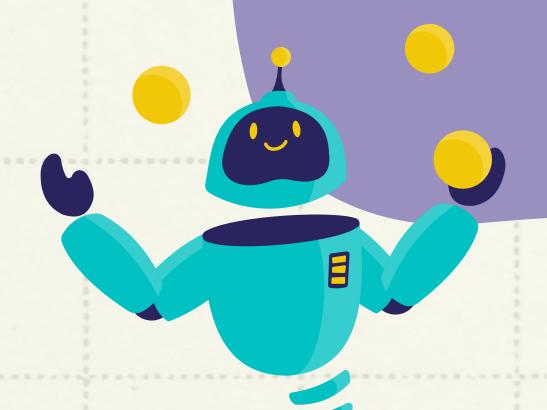
OFFLINE SORTING ALGORITHM

MỘT THUẬT TOÁN OFFLINE (OFFLINE ALGORITHM) LÀ CHƯƠNG TRÌNH ĐƯỢC CUNG CẤP TOÀN BỘ DỮ LIỆU ĐẦU VÀO CẦN THIẾT ĐỂ GIẢI QUYẾT VẤN ĐỀ ĐƯỢC ĐẶT RA

ONLINE SORTING ALGORITHM

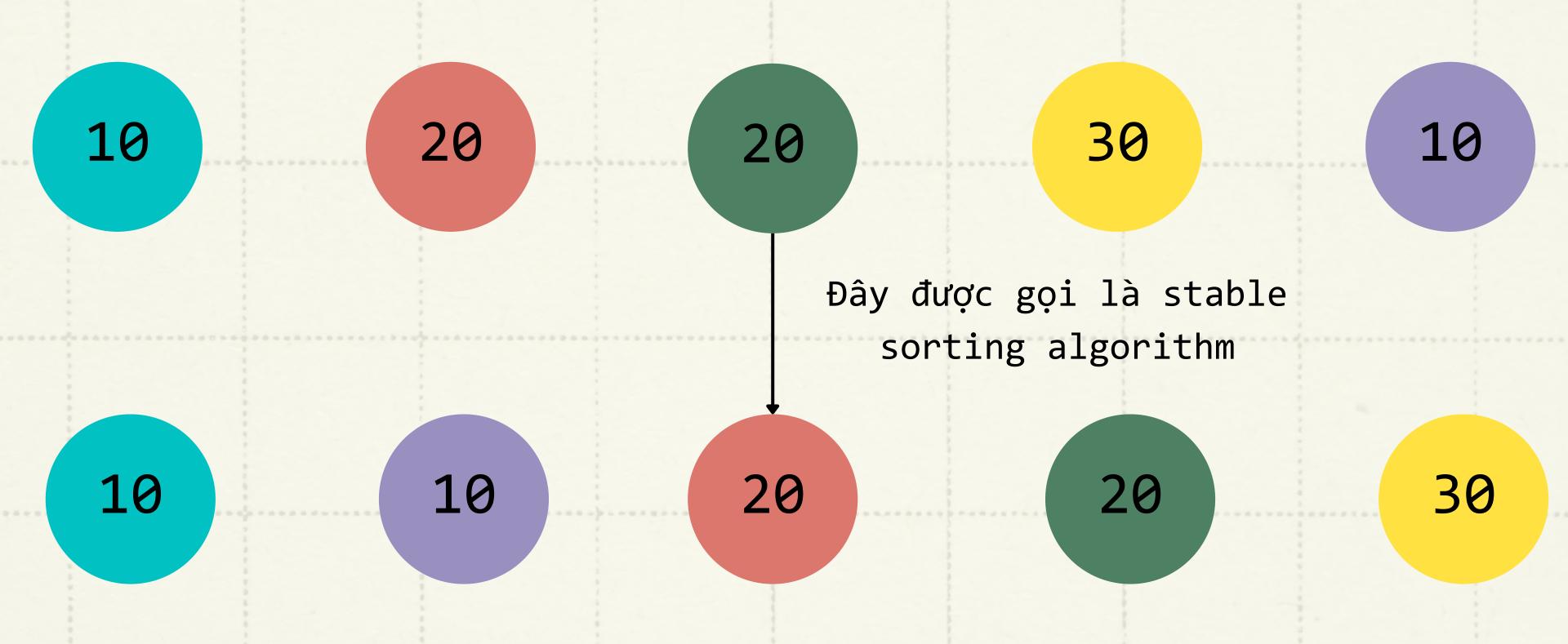
MỘT THUẬT TOÁN ONLINE ALGORITHM SẼ XỬ LÍ DỮ LIỆU ĐẦU VÀO 1 CÁCH TUẦN TỰ

STABLE - UNSTABLE



1 THUẬT TOÁN ĐƯỢC GỌI LÀ ỐN ĐỊNH (STABLE) NẾU 2 PHẦN TỬ CÓ GIÁ TRỊ BẰNG NHAU GIỮ NGUYÊN THỨ TỰ TRƯỚC VÀ SAU KHI SẮP XẾP

Ví dụ



STABLE SORTING ALGORITHM KHÔNG THỰC SỰ THỂ HIỆN ĐƯỢC SỰ HỮU DỤNG NẾU CÁC PHẦN TỬ LÀ SỐ NGUYÊN (INTEGERS)

STABLE SORTING ALGORITHM KHÔNG THỰC SỰ THỂ HIỆN ĐƯỢC SỰ HỮU DỤNG NẾU CÁC PHẦN TỬ LÀ SỐ NGUYÊN (INTEGERS)

CHÚNG CHỈ THỰC SỰ CHỨNG MINH ĐƯỢC HIỆU QUẢ KHI HOẠT ĐỘNG VỚI CÁC KIỂU DỮ LIỆU KHÁC.

VÍ DŲ:

INPUT:

HOW MUCH WOOD WOULD WOODCHUCK CHUCK IF WOODCHUCK COULD CHUCK WOOD

Output:

how 1
much 1
wood 2
would 1
woodchuck 2
chuck 2
if 1
could 1

VÍ DŲ:

First pass, sorted lexicographically:

```
(chuck, 2)
  (could, 1)
   (how, 1)
   (if,1)
  (much, 1)
  (wood, 2)
(woodchuck, 2)
  (would, 1)
```

VÍ DŲ:

SECOND PASS, SORTED BY COUNT USING A UNSTABLE SORT:

```
(wood, 2)
 (chuck, 2)
(woodchuck, 2)
  (could, 1)
  (how, 1)
   (if,1)
 (would, 1)
   (much, 1)
```

KHI DÙNG UNSTABLE SORT THÌ LẦN SORT THỨ 2 ĐÃ KHÔNG ĐẢM BẢO CHO CHÚNG TA VỀ THỨ TỰ ABC.

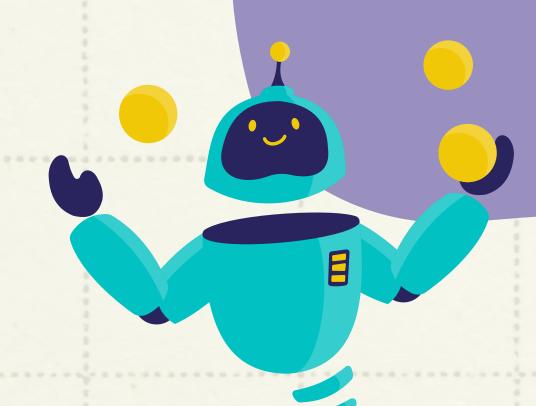
SECOND PASS, SORTED BY COUNT USING A STABLE SORT:

```
(chuck, 2)
  (wood, 2)
(woodchuck, 2)
  (could, 1)
  (how, 1)
   (if, 1)
  (much, 1)
  (would, 1)
```

- VỚI DỮ LIỆU LỚN, UNSTABLE SORT HOẠT ĐỘNG TỐT HƠN VÌ KHÔNG CẦN PHẢI GIỮ NGUYÊN CÁC THỨ TỰ CỦA CÁC PHẦN TỬ.
- UNSTABLE SORT ĐƠN GIẢN HƠN VÀ NHANH HƠN ĐỂ XÂY DỰNG NÊN.

• TUY NHIÊN, VIỆC KHÔNG GIỮ NGUYÊN CÁC THỨ TỰ CỦA CÁC PHẦN TỬ BẰNG NHAU CÓ THỂ DẪN ĐẾN MỘT SỐ VẤN ĐỀ TRONG CÁC TÌNH HUỐNG CỤ THỂ.

INTERNAL - EXTERNAL



EXTERNAL SORTING

LÀ KỸ THUẬT SẮP XẾP DỮ LIỆU **DUNG**LƯỢNG LỚN, VƯỢT QUÁ KHẢ NĂNG LƯU
TRỮ CỦA BỘ NHỚ CHÍNH (RAM) MÁY TÍNH

EXTERNAL SORTING

ĐƯỢC THIẾT KẾ ĐỂ XỬ LÝ CÁC TẬP DỮ LIỆU LỚN KHÔNG THỂ HOÀN TOÀN CHỨA TRONG RAM MÁY TÍNH, ĐÒI HỎI SỬ DỤNG BỘ NHỚ NGOẠI VI NHƯ Ổ ĐĨA CỨNG.

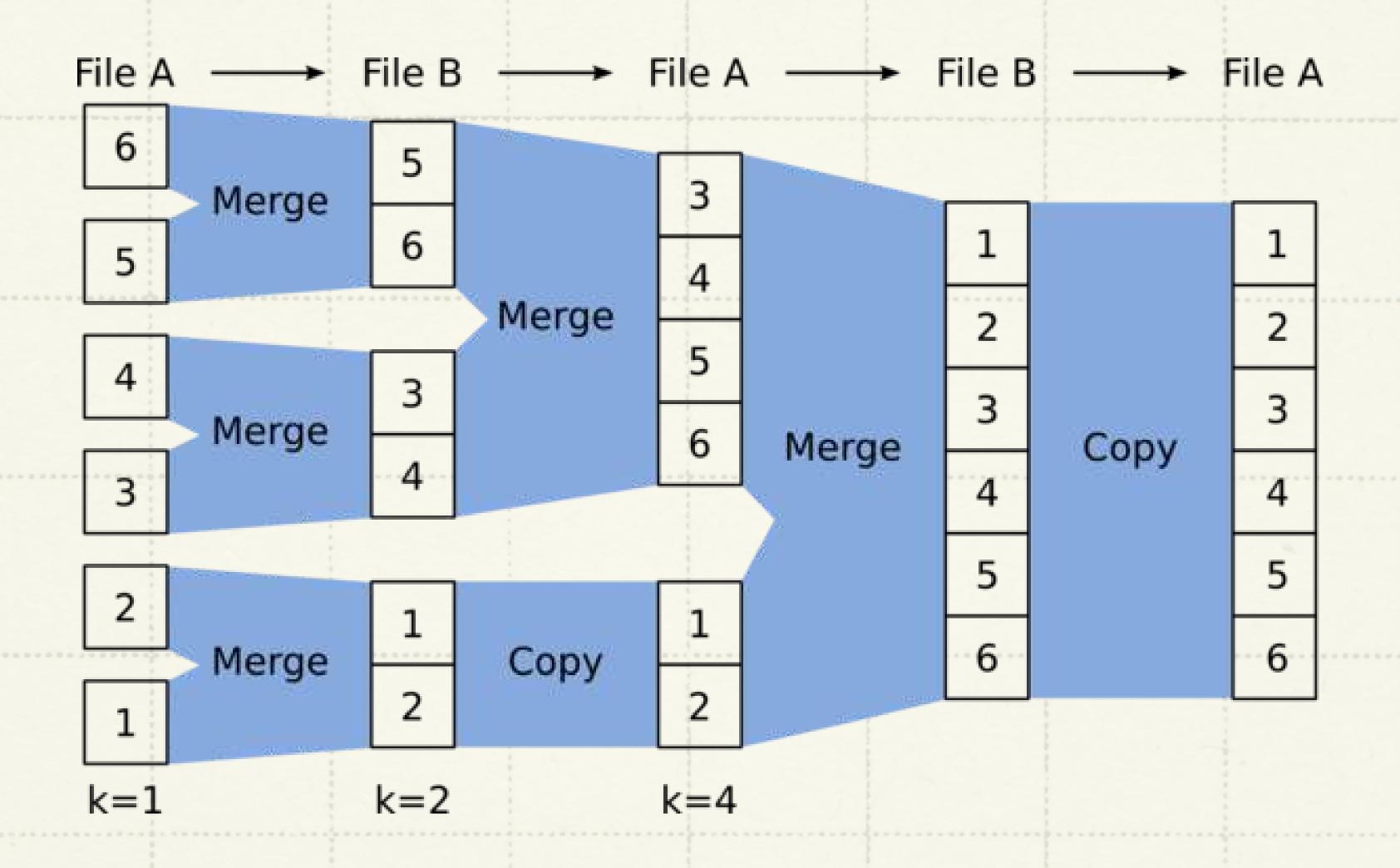
CHATERISTICS

DATASET SIZE: KHI TẬP DỮ LIỆU CẦN ĐƯỢC SẮP XẾP **QUÁ LỚN** ĐỂ CHỨA TRONG **RAM**

EXTERNAL STORAGE: SỬ DỤNG THIẾT BỊ LƯU
TRỮ NGOẠI VI, THƯỜNG LÀ Ổ ĐĨA CỬNG, ĐỂ
LƯU TRỮ VÀ TRUY XUẤT DỮ LIỆU TRONG QUÁ
TRÌNH SẮP XẾP

CHATERISTICS

MERGE-BASED APPROACH: HÂU HẾT CÁC THUẬT TOÁN SẮP XẾP NGOẠI VI TUÂN THEO PHƯƠNG PHÁP GỘP, TRONG ĐÓ TẬP DỮ LIỆU ĐƯỢC CHIA THÀNH CÁC PHẦN NHỎ CÓ THỂ CHỨA TRONG BỘ NHỚ, ĐƯỢC SẮP XẾP RIÊNG LÈ VÀ SAU ĐÓ ĐƯỢC GỘP ĐỂ CÓ KẾT QUẢ CUỐI CÙNG ĐÃ ĐƯỢC SẮP XẾP.



CHATERISTICS

EFFICIENCY: CÁC THUẬT TOÁN NHẰM GIẢM THIỀU SỐ LẦN THỰC HIỆN NHẬP/XUẤT, VÌ TRUY CẬP DỮ LIỆU TỪ LƯU TRỮ NGOẠI VI CHẬM HƠN NHIỀU SO VỚI RAM

OPTIMIZATION TECHNIQUES: NHIỀU KỸ THUẬT TỐI ƯU HÓA ĐƯỢC SỬ DỤNG, NHƯ GIẨM THIỀU TÌM KIẾM ĐĨA, GIẨM SỐ LẦN QUÉT QUA DỮ LIỆU VÀ SỬ DỤNG CẦU TRÚC DỮ LIỆU HIỆU QUẢ CHO VIỆC SẮP XẾP.

THƯỜNG ĐƯỢC SỬ DỤNG TRONG CÁC TÌNH HUỐNG ĐÒI HỎI VIỆC SẮP XẾP QUY MÔ LỚN, NHƯ TRONG CƠ SỞ DỮ LIỆU, HỆ THỐNG TỆP, VÀ KHO DỮ LIỆU. NÓ ĐẶC BIỆT HỮU ÍCH KHI XỬ LÝ LƯỢNG DỮ LIỆU LỚN KHÔNG THỂ XỬ LÝ HOÀN TOÀN TRONG RAM CỦA MÁY TÍNH.

VIỆC SỬ DỤNG SẮP XẾP NGOẠI VI CŨNG ĐẨM BẢO RẰNG QUÁ TRÌNH SẮP XẾP DUY TRÌ HIỆU SUẤT TỐI ƯU NGAY CẢ KHI XỬ LÝ CÁC KÍCH THƯỚC TỆP LỚN. NÓ GIẢM THIỀU CÁC TÁC VỤ ĐỌC/GHI ĐĨA VÀ TỐI ĐA HÓA HIỆU SUẤT TỔNG THỂ CỦA THUẬT TOÁN SẮP XẾP.

INTERNAL SORTING

LÀ QUÁ TRÌNH SẮP XẾP DỮ LIỆU TRONG RAM HOẶC BỘ NHỚ NỘI VI CỦA MÁY TÍNH. NÓ BAO GỒM VIỆC SẮP XẾP LẠI CÁC PHẦN TỬ CỦA TẬP DỮ LIỆU THÀNH MỘT THỨ TỰ CỤ THỂ, CHẮNG HẠN NHƯ TĂNG DẦN HOẶC GIẢM DẦN, DỰA TRÊN MỘT SỐ TIÊU CHÍ NHẤT DINH.

INTERNAL SORTING

CÓ NHIỀU THUẬT TOÁN ĐƯỢC SỬ DỤNG CHO SẮP XẾP NỘI VI, BAO GỒM INSERTION SORT, SELECTION SORT, BUBLE SORT, MERGE SORT VÀ QUICK SORT. MÕI THUẬT TOÁN CÓ ƯU ĐIỂM VÀ HẠN CHẾ RIỆNG VỀ ĐỘ PHỨC TẠP THỜI GIAN, SỬ DỤNG BỘ NHỚ VÀ TÍNH ỔN ĐỊNH.

DATA SIZE: THƯỜNG ĐƯỢC SỬ DỤNG CHO CÁC TẬP DỮ LIỆU NHỎ CÓ THỂ CHỨA HOÀN TOÀN TRONG RAM , GIÚP TĂNG TỐC ĐỘ TRUY CẬP VÀ XỬ LÝ DỮ LIỆU TRONG QUÁ TRÌNH SẮP XẾP.

MEMORY EFFICIENCY: CÁC THUẬT TOÁN NHẰM GIẨM THIỀU VIỆC SỬ DỤNG BỘ NHỚ PHỤ TRONG QUÁ TRÌNH SẮP XẾP, TẬP TRUNG VÀO TỐI ƯU HÓA SỬ DỤNG RAM ĐỂ THỰC HIỆN CÁC HOẠT ĐỘNG SẮP XẾP HIỆU QUẢ.

SPEED: VIỆC SẮP XẾP DỮ LIỆU NỘI VI CÓ THỂ NHANH ĐÁNG KỂ SO VỚI PHƯƠNG PHÁP SẮP XẾP NGOẠI VI, VÌ NÓ LOẠI BỎ CÁC NHU CẦU HOẠT ĐỘNG ĐỌC/GHI ĐĨA TỐN THỜI GIAN. VỚI DỮ LIỆU NẰM TRONG RAM, QUÁ TRÌNH SẮP XẾP HƯỞNG LỢI TỪ THỜI GIAN TRUY CÂP NHANH.

ALGORITHM SELECTION: CÓ NHIỀU THUẬT TOÁN, SỰ CHỌN LỰA THUẬT TOÁN PHỤ THUỘC VÀO CÁC YẾU TỐ NHƯ KÍCH THƯỚC DỮ LIỆU, TÍNH ỔN ĐỊNH, ĐỘ PHỨC TẠP KHÔNG GIAN VÀ ĐỘ PHỨC TẠP THỜI GIAN MONG MUỐN.

STABLE: CÁC THUẬT TOÁN CÓ THỂ ĐƯỢC THIẾT KẾ ĐỂ BẢO TỒN THỨ TỰ TƯƠNG ĐỐI CỦA CÁC PHẦN TỬ BẰNG NHAU. ĐIỀU NÀY CÓ NGHĨA LÀ NẾU HAI PHẦN TỬ CÓ CÙNG GIÁ TRỊ, THỨ TỰ BAN ĐẦU CỦA CHÚNG TRONG TẬP DỮ LIỆU SẼ ĐƯỢC **DUY TRÌ** SAU KHI SĂP XÊP.

APPLICATION: THƯỜNG ĐƯỢC SỬ DỤNG TRONG CÁC TÌNH HUỐNG MÀ TOÀN BỘ TẬP DỮ LIỆU CÓ THỂ ĐƯỢC TẢI VÀO BỘ NHỚ, NHƯ SẮP XẾP CÁC MẢNG HOẶC DANH SÁCH NHỎ HOẶC TỔ CHỨC CẤU TRÚC DỮ LIỆU TRONG BỘ NHỚ ĐỂ TRUY CẬP VÀ TRUY XUẤT HIỆU QUẢ.

SẮP XẾP NỘI VI THƯỜNG ĐƯỢC SỬ DỤNG TRONG NHIỀU TÌNH HUỐNG KHI CẦN TỔ CHỨC MỘT LƯỢNG LỚN DỮ LIỆU MỘT CÁCH HIỆU QUẢ. SAU ĐÂY LÀ MỘT SỐ TRƯỜNG HỢP SẮP XẾP NỘI VI ĐƯỢC SỬ DỤNG:

DATABASE MANAGEMENT SYSTEMS: SÜ DUNG TRONG CÁC HỆ THỐNG QUẢN LÝ CƠ SỞ DỮ LIỆU ĐỂ SẮP XẾP DỮ LIỆU THEO MỘT THỨ TỰ CỤ THỂ, CẢI THIỆN HOẠT ĐỘNG TÌM KIẾM VÀ TRUY XUẤT, ĐẨM BẢO DỮ LIỆU ĐƯỢC LƯU TRỮ CHO PHÉP TRUY CẬP NHANH VÀ GIẨM THỜI GIAN CẦN THIẾT CHO CÁC TRUY VÂN.

DATA ANALYSIS: SỬ DỤNG TRONG CÁC TÁC VỤ PHÂN TÍCH DỮ LIỆU KHI CẦN SẮP XẾP CÁC TẬP DỮ LIỆU LỚN THEO CÁC TIÊU CHÍ CỤ THỂ. BẰNG CÁCH SẮP XẾP DỮ LIỆU, CÁC NHÀ PHÂN TÍCH CÓ THỂ DỄ DÀNG NHẬN BIẾT MÔ HÌNH, XU HƯỚNG, GIÚP HỌ ĐƯA RA QUYẾT ĐINH THÔNG TIN VÀ RÚT RA PHÂN TÍCH SÂU SĂC.

FILE SYSTEMS: QUAN TRỌNG TRONG HỆ THỐNG TỆP KHI TỔ CHỨC CÁC TỆP, THƯ MỤC. SẮP XẾP GIÚP NHANH CHÓNG ĐỊNH VỊ VÀ TRUY CẬP CÁC TỆP MONG MUỐN.

PROGRAMMING AND ALGORITHM DESIGN: LÀ MỘT PHẦN CƠ BẢN, SỬ DỤNG ĐỂ SẮP XẾP MẢNG HOẶC DANH SÁCH CÁC PHẦN TỬ DỮ LIỆU THEO THỨ TỰ, GIÚP TÌM KIẾM VÀ XỬ LÝ MỘT CÁCH HIỆU QUẢ.

DATA MINING AND MACHINE LEARNING: ĐƯỢC SỬ DỤNG TRONG CÁC TÁC VỤ KHAI THÁC DỮ LIỆU VÀ MÁY HỌC ĐỂ TIỀN XỬ LÝ VÀ CHUẨN BỊ DỮ LIỆU CHO PHÂN TÍCH VÀ MÔ HÌNH HÓA. SẮP XẾP DỮ LIỆU HỖ TRỢ LỰA CHỌN ĐẶC TRƯNG, XẾP HẠNG THUỘC TÍNH VÀ CÁC BƯỚC TIỀN XỬ LÝ DỮ LIỆU CẦN THIẾT ĐỂ XÂY DỰNG MÔ HÌNH CHÍNH XÁC VÀ ĐÁNG TIN CÂY.

	Internal Sorting	External Sorting
	In Internal sorting, all the data to sort is stored in main memory at all times while sorting is in progress.	In External sorting, data is stored outside memory (like on disk) and only loaded into memory in small chunks.
	It is performed when the data to be sorted is small enough to fit in main memory.	External sorting is usually applied in cases when data can't fit into main memory entirely.
0	In other words, It is used when size of input is small.	In other words, It is used when size of input is large.
	In this method, the storage device used is only Main memory(RAM).	In this method, the storage devices are both Secondary memory(Hard disk) and Main memory(RAM).
	Examples: Bubble sort, Insertion sort, Quick sort, Heap sort etc.	Example: External Merge Sort.

.



