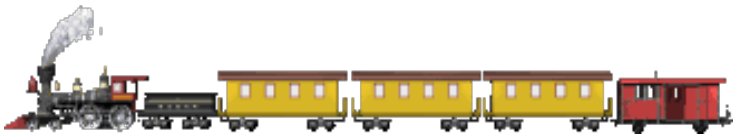




Conception Avancée de Base de Données

Principe des TP



Traduction en cours

Simplification



- Opération sur les données.
- Opération la plus utilisée : "select"
- Recherche d'éléments sur un critère de choix.
- Techniques utilisées dans toutes les applications de traitement de données.
- Opération réalisées à base d'index.
- Select = jointure, une ou deux tables.

Analyse et compréhension du problème



- 1) cas minimum : jointure en mémoire
 - Les relations tiennent en mémoire
 - Pas de segmentation
 - Pas de gestion des partitions sur disque
- 2) cas complexe : jointure sur disque
 - Les relations ne tiennent pas en mémoire
 - Les relations sont stockées dans des blocks disques
 - Gestion des partitions sur disque
 - Gestion des buffers mémoire

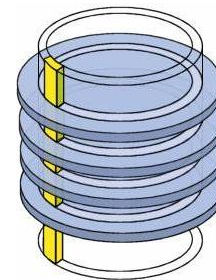
Jointures Physiques

■ Jointure en mémoire

- Nested Loop
- Merge join
- Hash join

■ Jointure sur disque

- Nested Loop
- Sort Merge Join
- Hash join

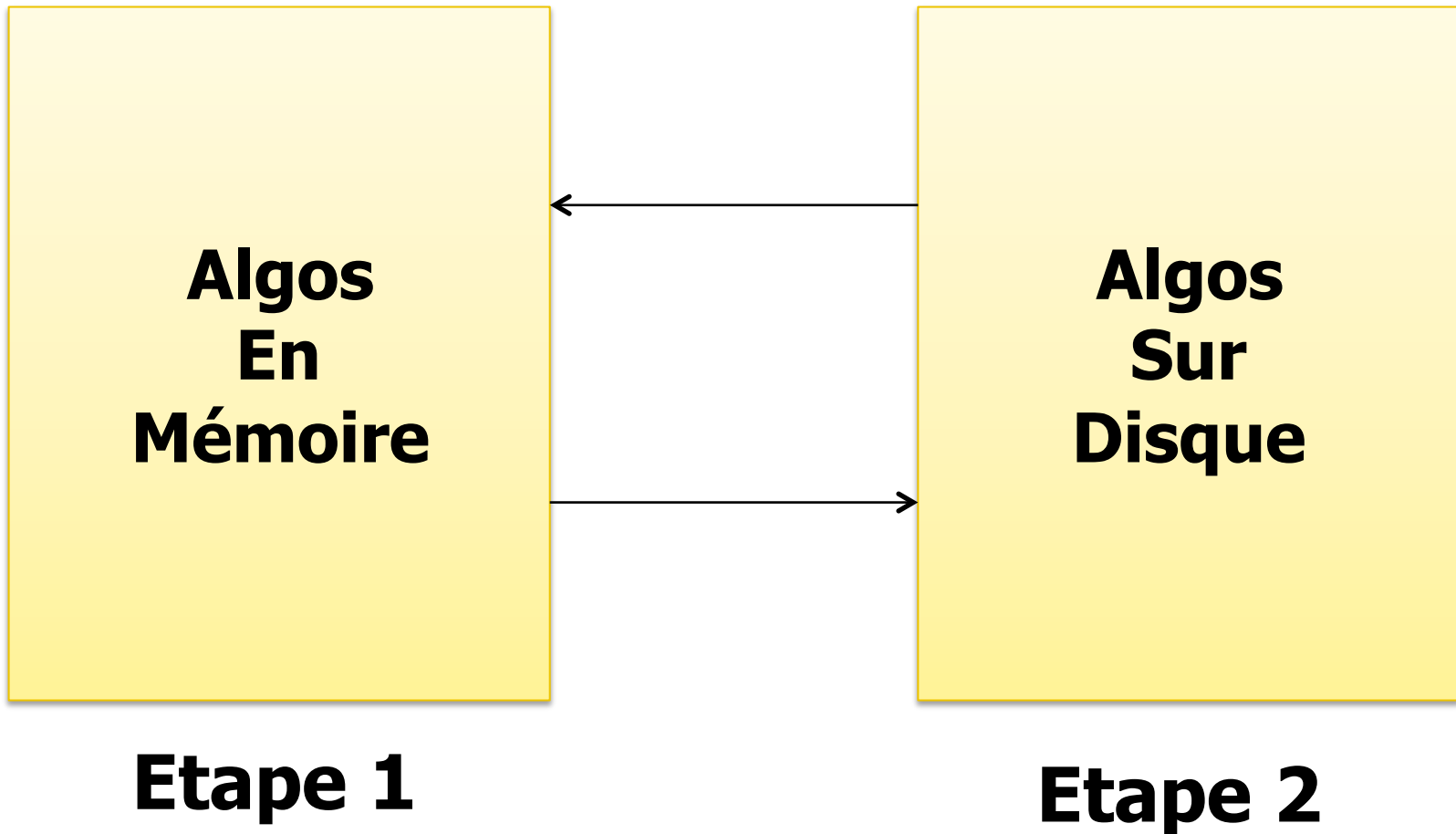


Principes de conception



Jointures en mémoire

Jointures sur disque



Principes de conception



- Objectifs : Les algos sur disque utilisent les algo en mémoire.
- Étape 1 :
 - Simplification des algos en mémoire
 - Ils seront adaptés plus tard pour être utilisés par les algos sur disque.
- Etape 2 :
 - Les algos sur disque rajoutent la gestion des blocs disque et pages mémoire (buffer).

Principes de simplification des algos en mémoire



- Chaque relation tient dans un seul fichier lu en une seule fois.
- Une seule colonne par relation (index)
- Pas de doublons, sauf cas particulier.
- Le résultat de la jointure tient dans un seul fichier écrit en une fois.
- Taille des relations fixée à 10 lignes pour simplifier les tests.
- Seule contrainte : utilisation de tableaux de taille 10.

Principes de simplification des algos en mémoire en Java



- Pas de collections Java
- Que des tableaux de char
 - `char[] r =`
 - `char[] s =`
- Pas de String
 - `String[] r =`
 - `String[] s =`
- Pas de Char, rien qui commence avec une majuscule.

Principes de simplification des algos en mémoire en C



- Pas de lib spécifiques non contrôlées
- Que des tableaux de char
 - `char r [10];`
 - `char s [10];`
- Ce sont des tables pas des string

A
Z
G
J
U
K
E
B
V
D

A	Z	G	J	U	K	E	B	V	D	null
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-------------

Principes de simplification des algos en mémoire

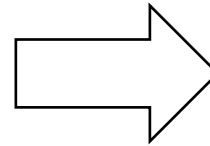


- Tables => Mono Column => Index
- Index :
 - Files
 - Database
 - Big Data

Principes de simplification des algos en mémoire



A	B	C
a	g	v
v	e	g
b	y	e
y	r	h
i	d	j
k	s	l
l	f	m



A
a
v
b
y
i
k
l

Une table

On s'implifie

Tables => Mono Column => Index



Relation R

Attribut A

**Index on
Attribut A**

A
Z
G
J
U
K
E
B
V
D

Tableaux de char (pas Char)



R

char

A
Z
G
J
U
K
E
B
V
D

Table R

S

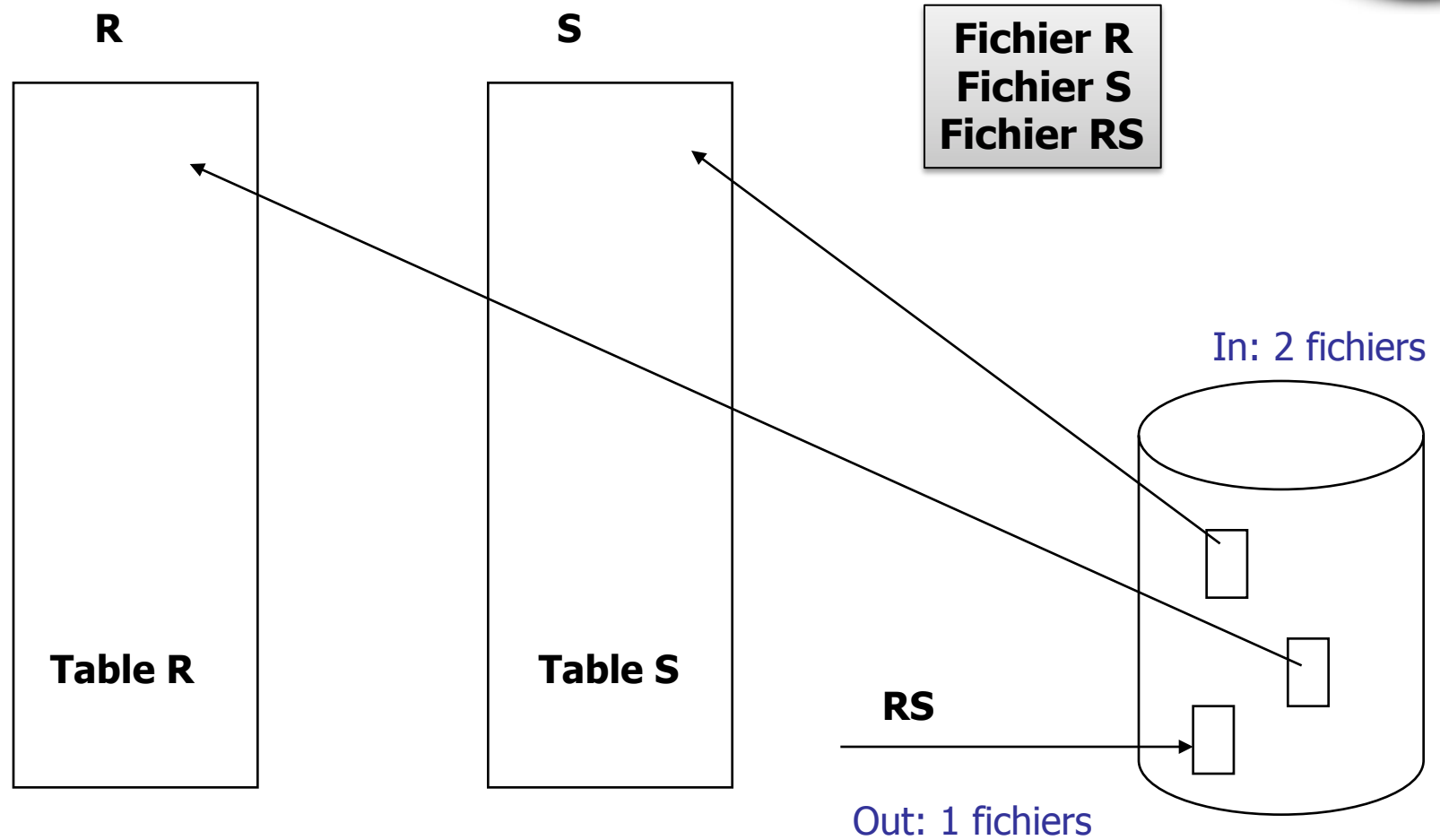
char

B
U
Z
K
X
V
N
L
M
E

Table S

$$RS = \text{join } (R,S)$$

Une table égale un fichier sur disque



Nested Loop Join

