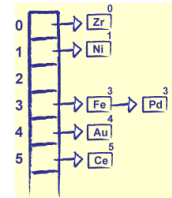
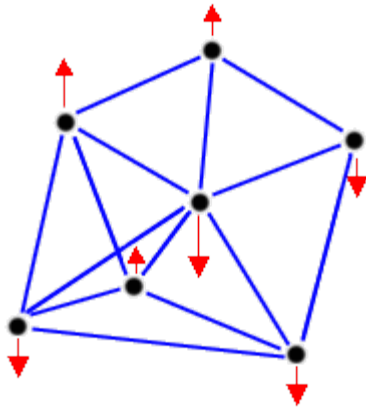


Conception Avancée de Bases de Données

Hash Join



Traduction en cours

Table R and S



$$RS = R \bowtie S$$

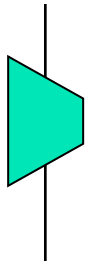
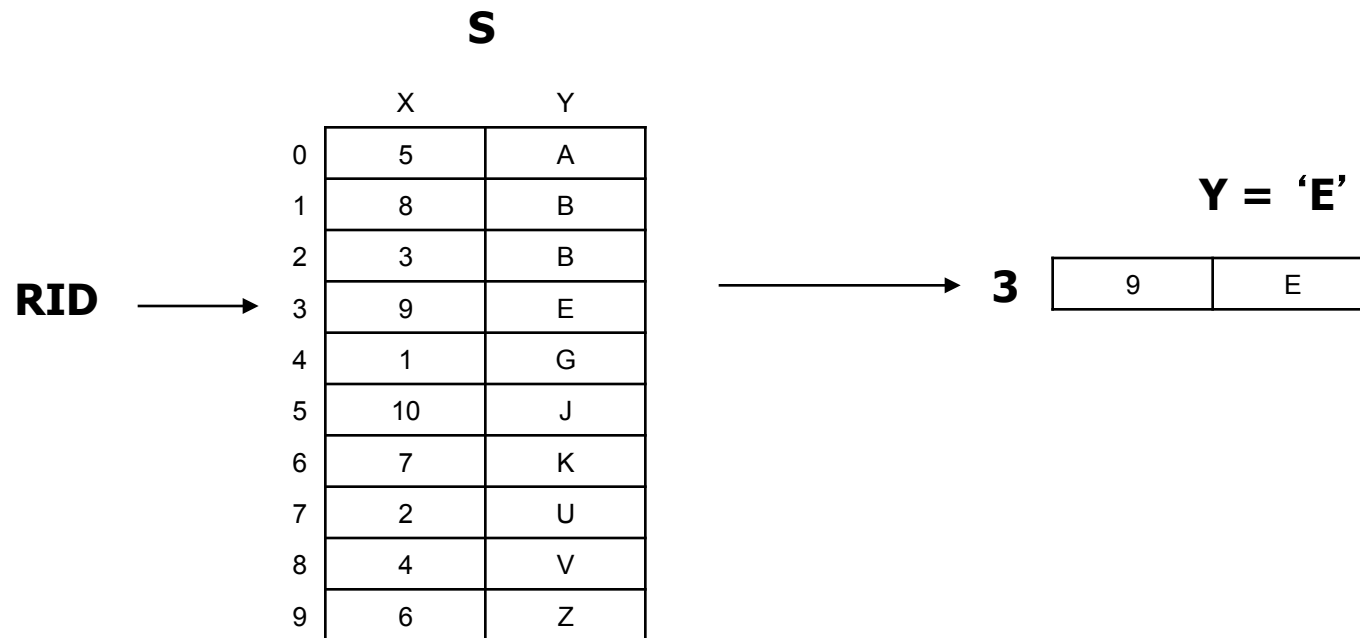
R

Y	Z
Z	20
B	22
E	28
K	24
M	25
N	30
U	27
L	23
V	21
X	26

S

X	Y
5	A
8	B
3	B
9	E
1	G
10	J
7	K
2	U
4	V
6	Z

Raw Identifier (RID)

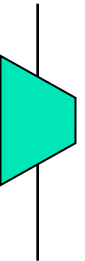
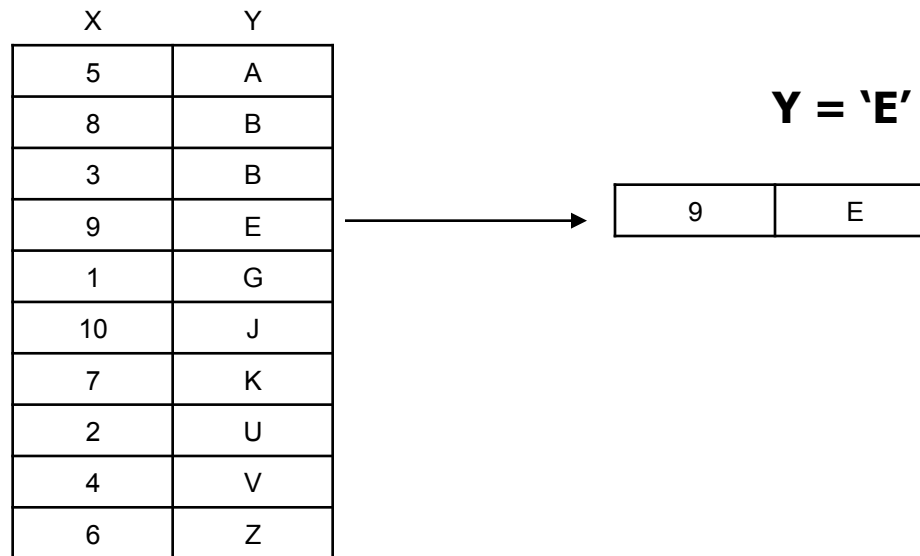


Sequential Scan select : Y = 'E'

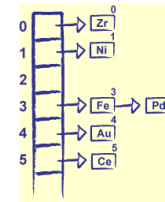


- Seek for Y attribute value "E"

S



Seq Scan select : $Y = 'E'$



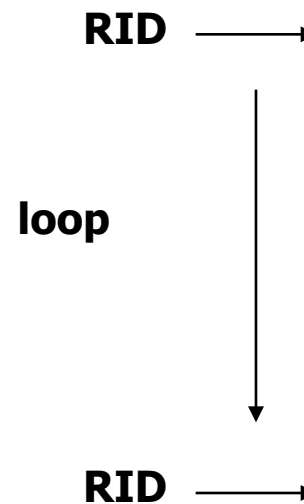
Seq Scan :

While i in index range

if $Y = 'E'$ then OutPut = $S(i)$

End while

S



	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z

Nested Loop with Seq Scan



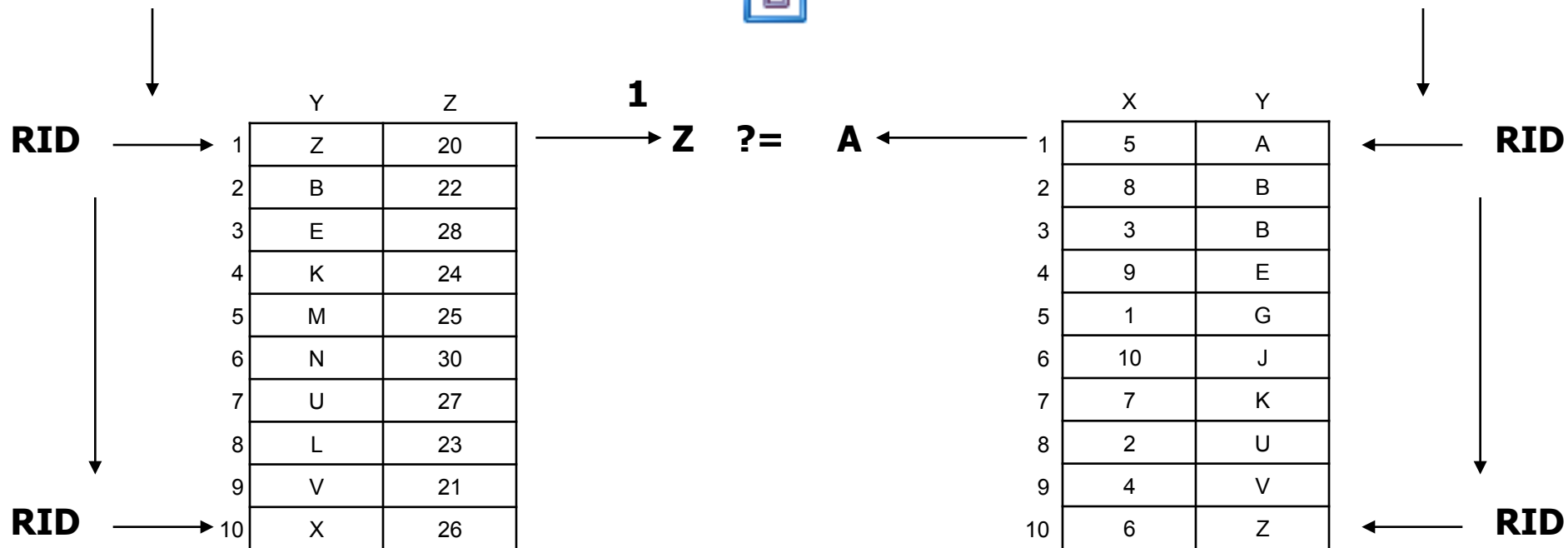
Read()

R



S

Seq Scan



Nested Loop with Seq Scan



Read()

R



S

Seq Scan

RID

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

1

Z ?= B

B

RID

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

RID

Nested Loop with Seq Scan



Read()

R



S

Seq Scan



RID

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

1

Z ?= B

RID

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

RID

Nested Loop with Seq Scan



Read()

R



S

Seq Scan

RID →

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

→ **1**

RID →

Z **?=** **E**

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

← **RID**

← **RID**

Nested Loop with Seq Scan



Read()

R



S

Seq Scan

RID →

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

→ **1**

Z **?=** **G**

← **RID**

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

← **RID**

Nested Loop with Seq Scan



Read()

R



S

Seq Scan



RID

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

1

Z ?= J

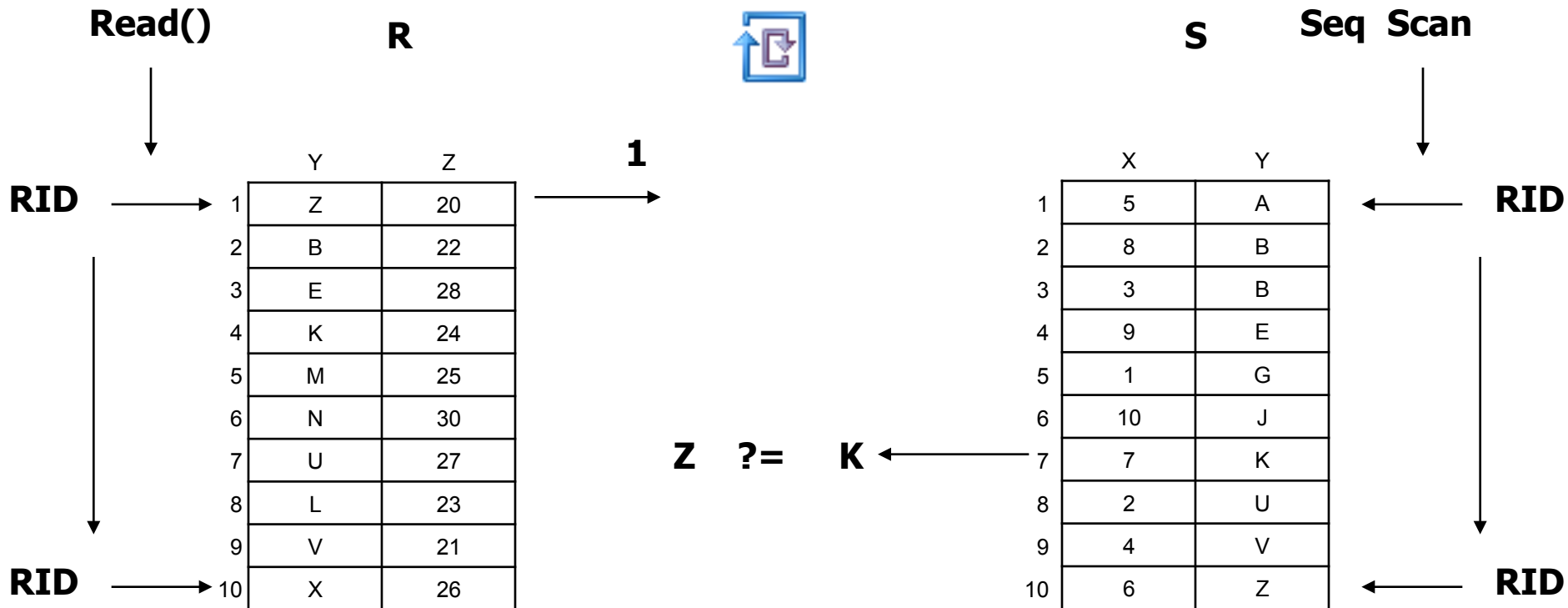
J

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

RID

RID

Nested Loop with Seq Scan



Nested Loop with Seq Scan



Read()

R



S

Seq Scan

RID →

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

→ **1**

RID →

Z **?=** **U**

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

← **RID**

← **RID**

Nested Loop with Seq Scan



Read()

R



S

Seq Scan



RID

RID

RID

RID

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

1

Z ?= V

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

Nested Loop with Seq Scan



Read(i)

R



S

Seq Scan



RID

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

1

Z ?= Z

Z

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

RID

RID

Nested Loop with Seq Scan



Read(i)

R



S

Seq Scan



RID

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

1

RID

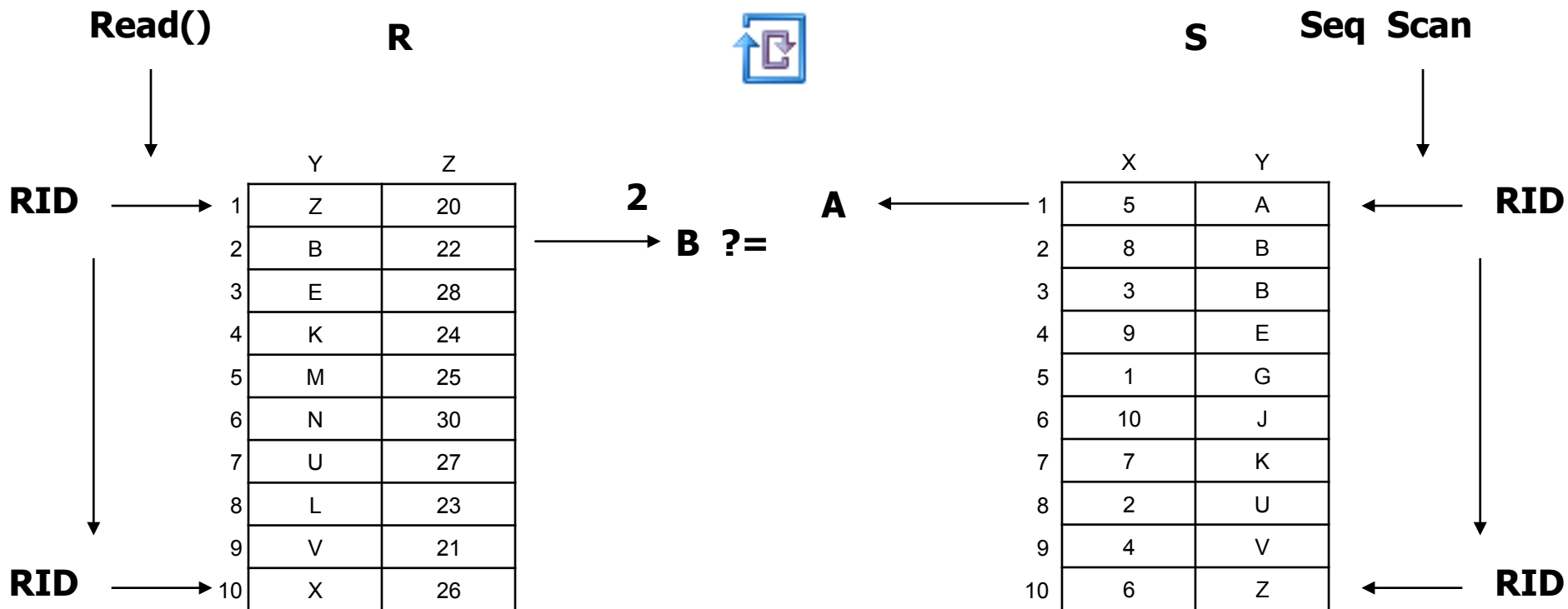
	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

10

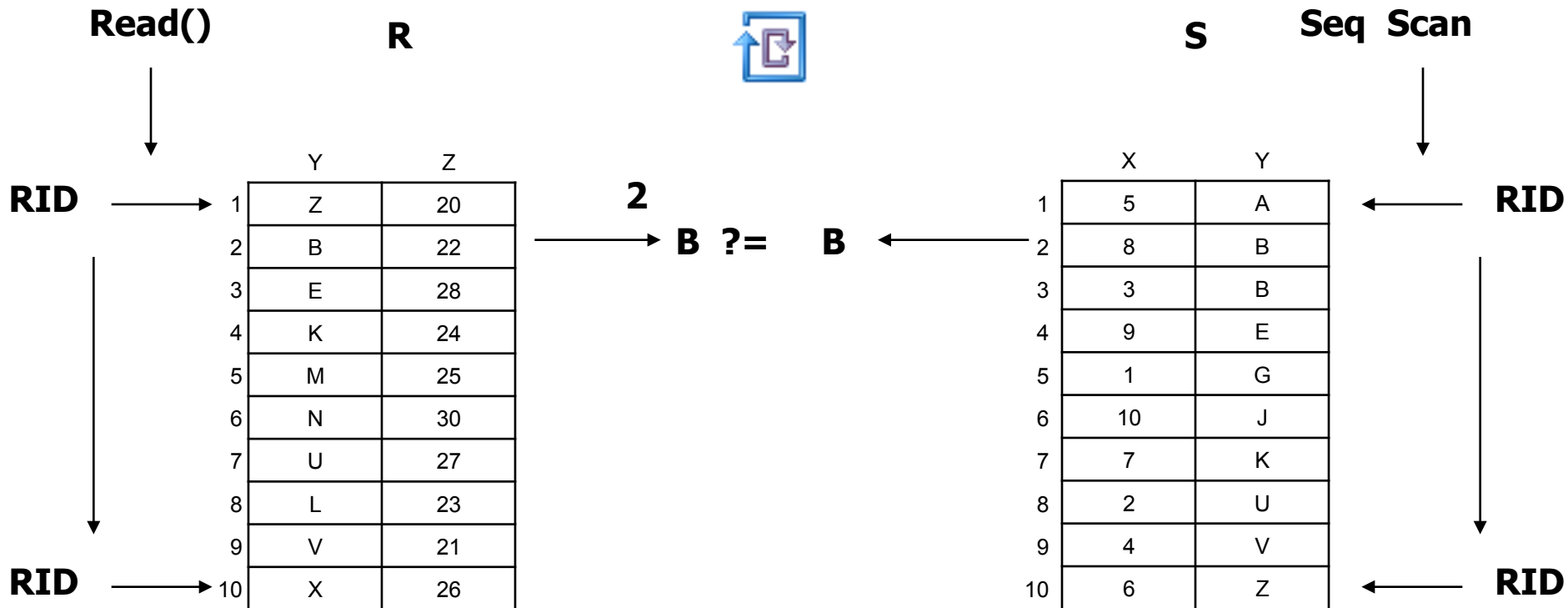
RID

OutPut (S.Read(10) + S.Read(10))

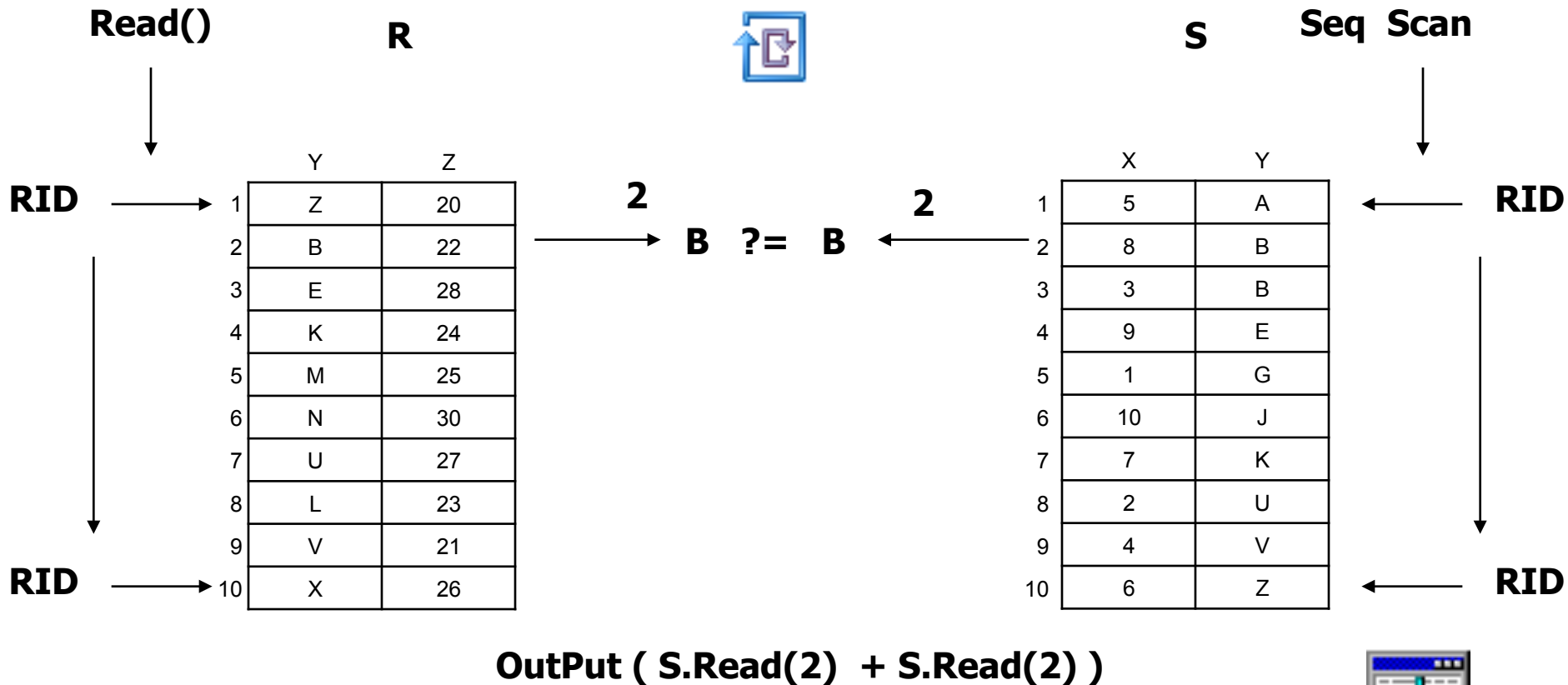
Nested Loop with Seq Scan



Nested Loop with Seq Scan



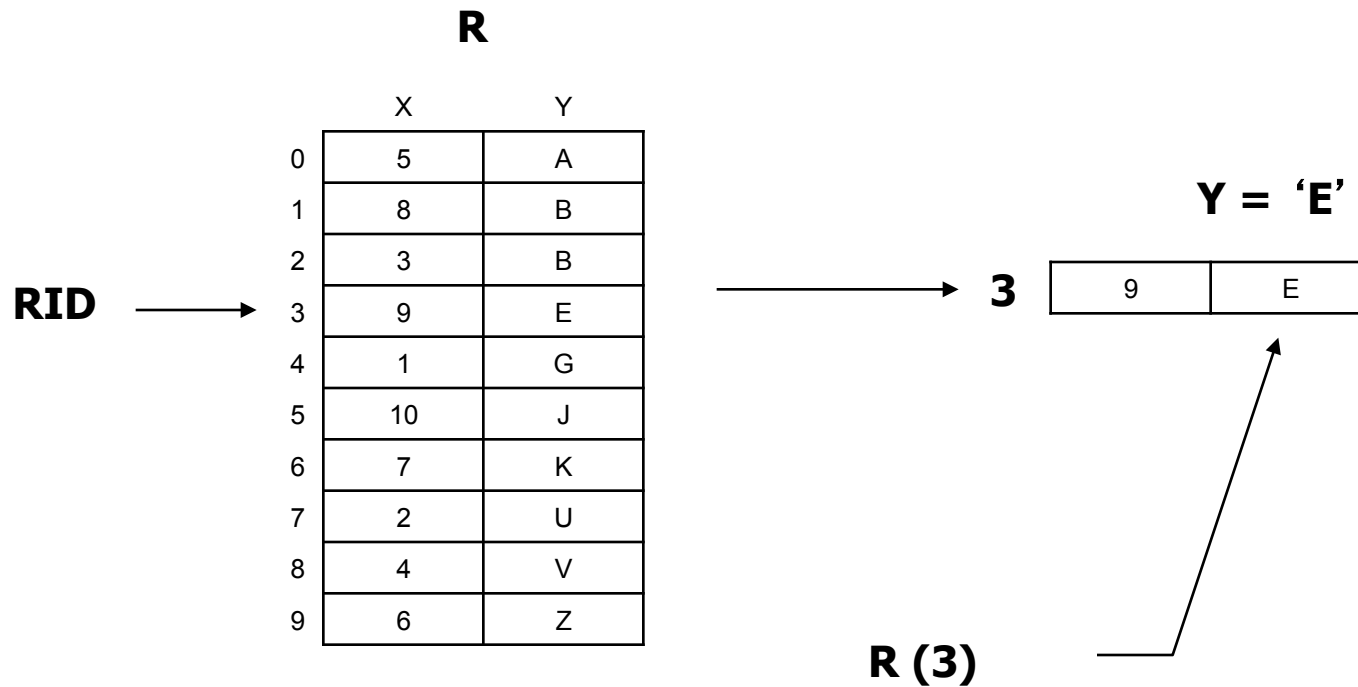
Nested Loop with Seq Scan



Tuple physical access : RID



Raw Identifier (RID)



Hash table access



Hash

R

	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z

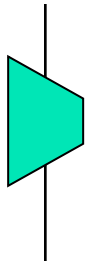
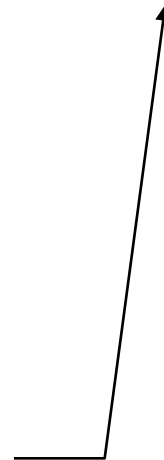
Y = 'E'



3

9	E
---	---

Hash(E)



Hash Join Principle : Hash Key, Hash Value



Relation R

	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z

hash keys = Attribut Y

hash values = RID

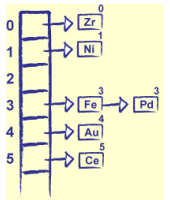
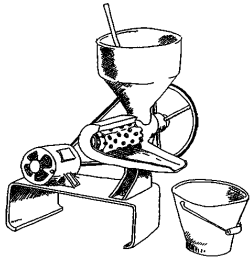
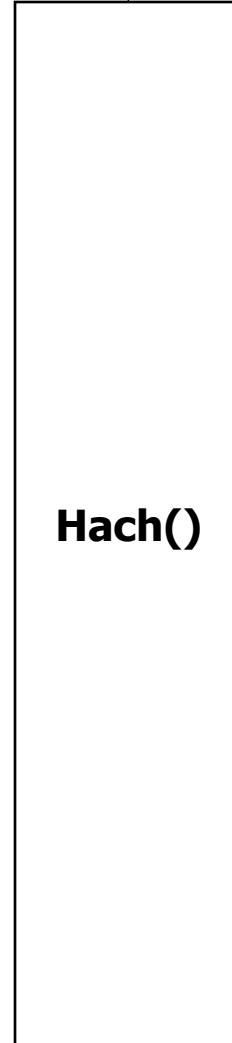


Tableau associatif



« c »



Hach()

hashtable

Relation

RID

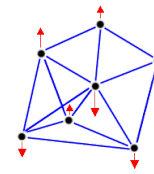
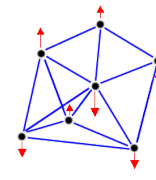


Tableau associatif

« c »

key



Relation

RID

Hach()

hashtable



Tableau associatif

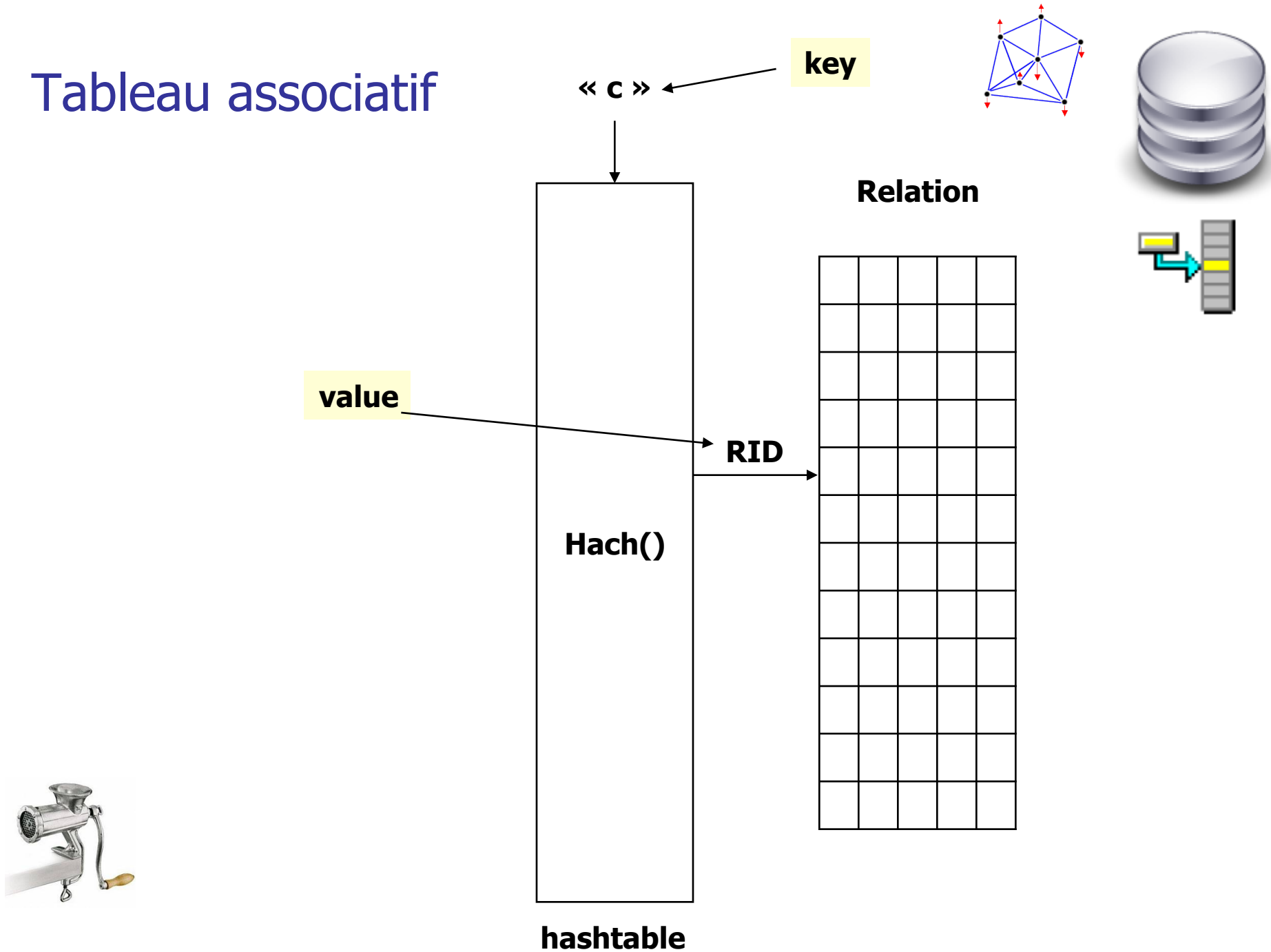
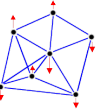
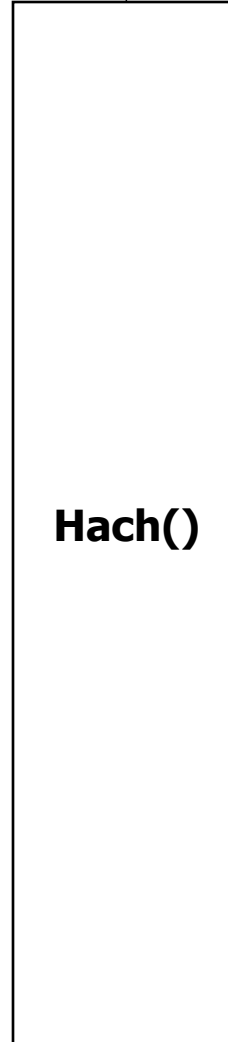


Tableau associatif



« E »



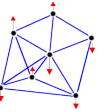
3

	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z

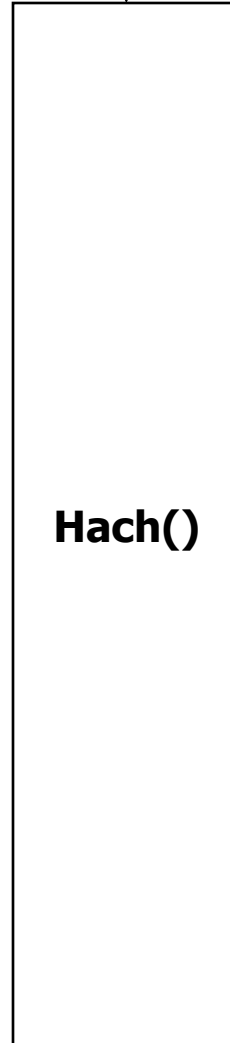


hashtable

Tableau associatif



« U »



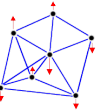
7

	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z

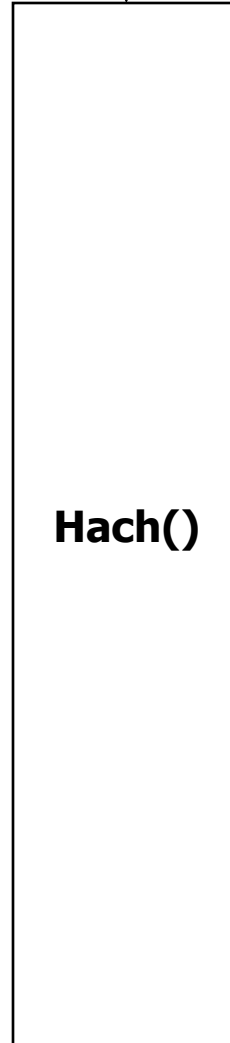


hashtable

Tableau associatif



« G »



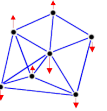
4

	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z

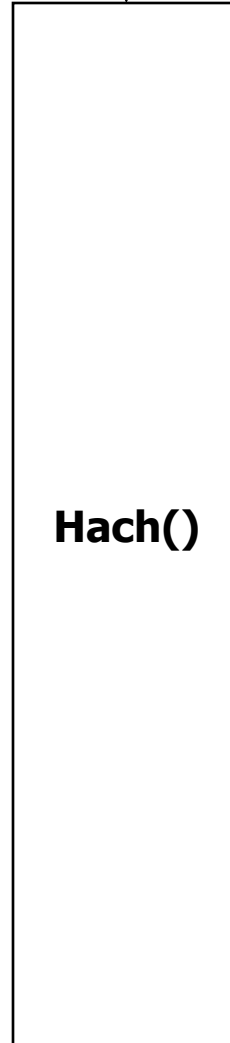


hashtable

Tableau associatif



« B »



1

2

	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z

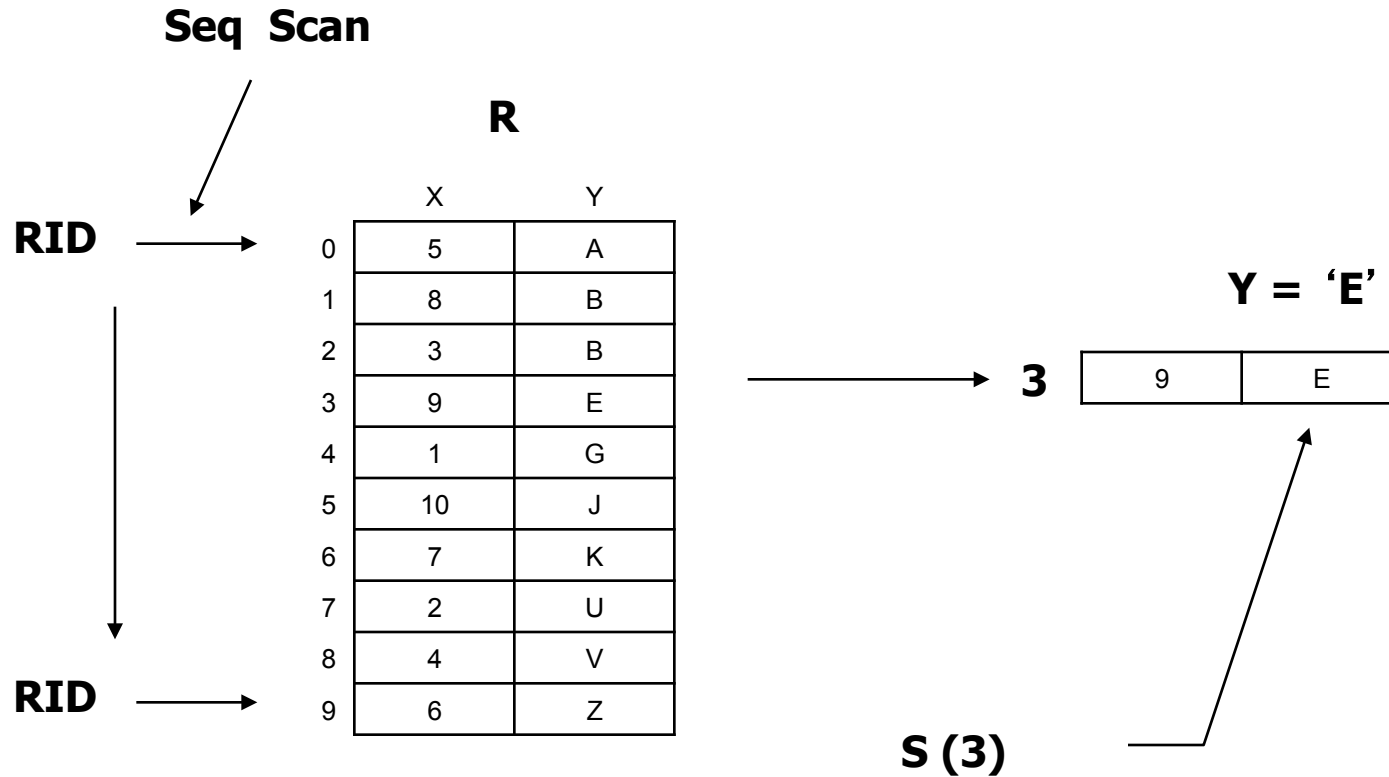


hashtable

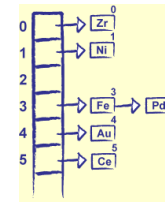
Reminder : Seq Scan



Raw Identifier (RID)



Reminder : Seq Scan select Y = 'E'



Seq Scan :

While i in index range

if Y = 'E' then OutPut = S(i)

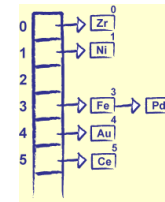
End while

S

	X	Y
RID →	5	A
	8	B
	3	B
loop ↓	9	E
	1	G
	10	J
	7	K
	2	U
	4	V
RID →	6	Z

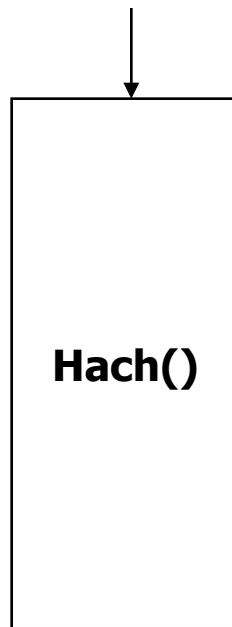


Equivalent Hash select : $Y = 'E'$



Output = R (Hash('E')°;

<< E >>



Hash Table

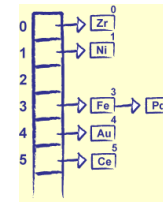
3

R

	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z



Seq scan VS Hash scan



Seq Scan :

While i in index range

if Y = 'E' then OutPut = S(i)

End while

loop

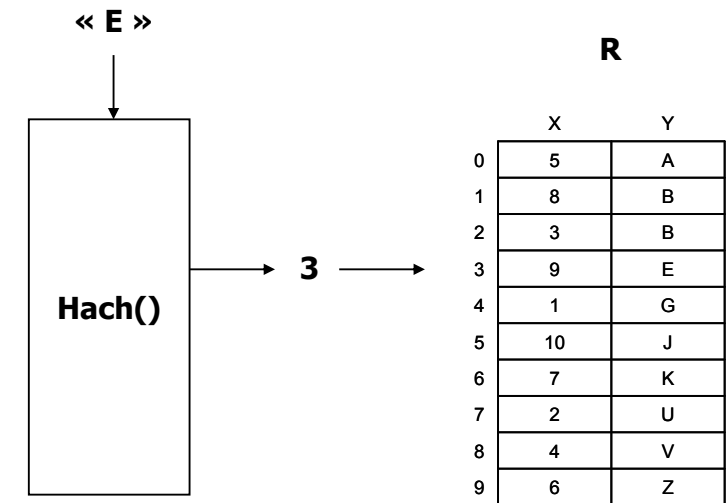
S

	X	Y
0	5	A
1	8	B
2	3	B
3	9	E
4	1	G
5	10	J
6	7	K
7	2	U
8	4	V
9	6	Z

RID → 0
RID → 9



OutPut = R(Hash('E'));



Hash Table



Hash Join



R

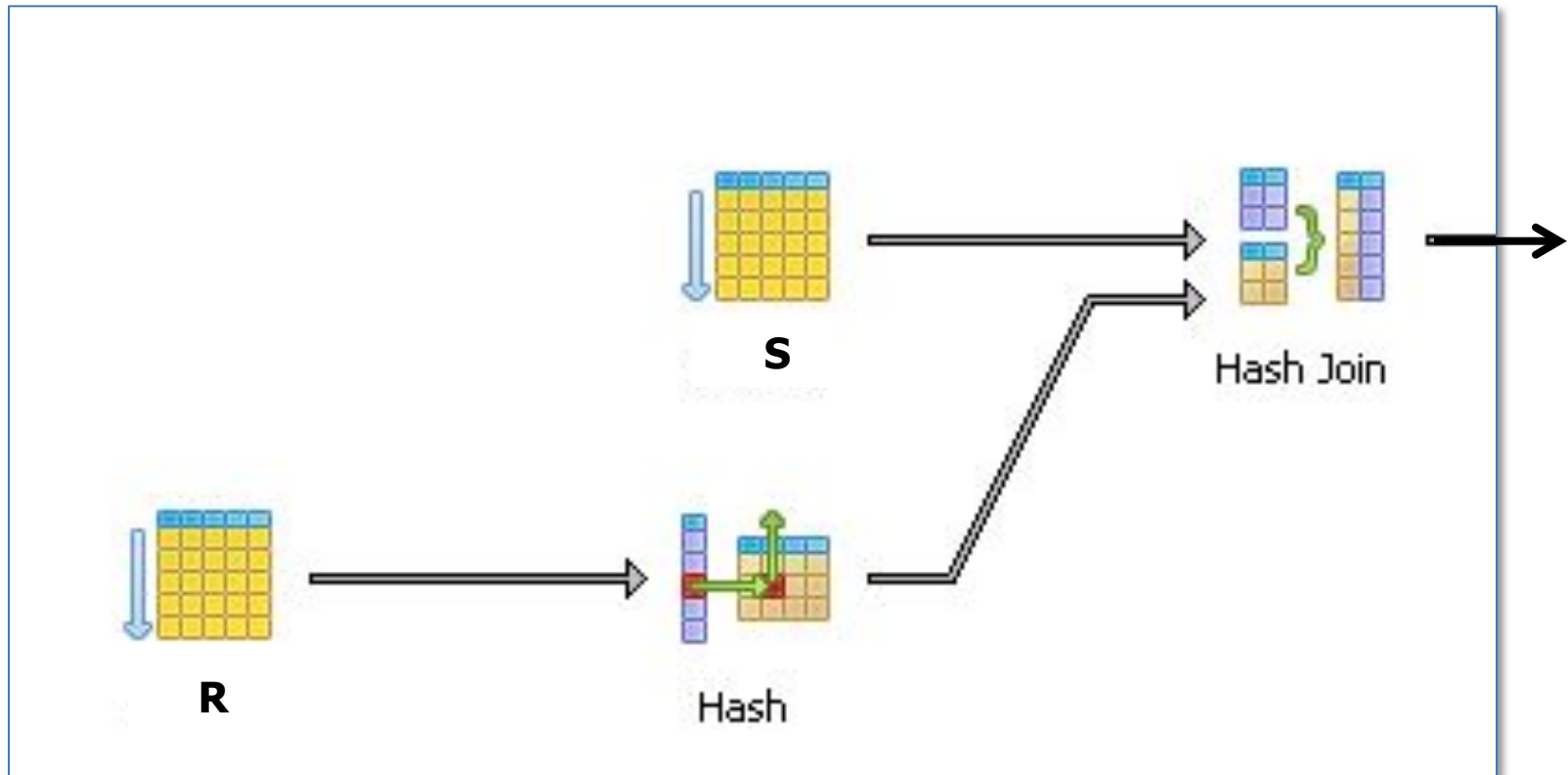


S

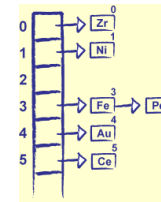
	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

Hash Join



R hash table Build phase



$\text{Hash}(i, R_y(i))$

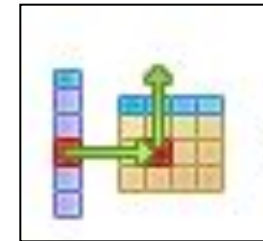
R

	Y	Z
0	Z	20
1	B	22
2	E	28
3	K	24
4	M	25
5	N	30
6	U	27
7	L	23
8	V	21
9	X	26

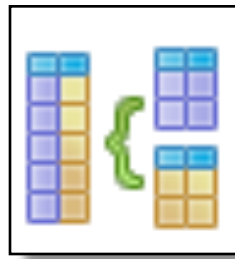
Y

Hach()

Hash Table



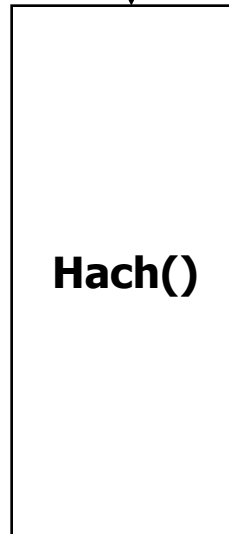
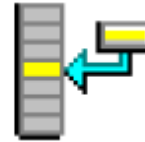
Hash Join



R

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

« A »



Hash Table

S

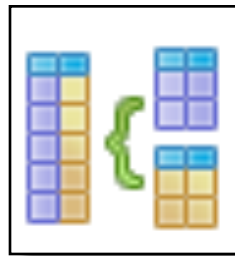
	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

Read S(i)

RID

RID

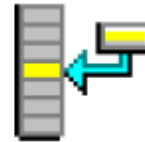
Hash Join



R

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

« B »



S

Read S(i)

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

RID

RID

2

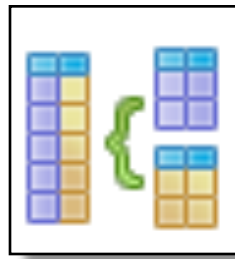
Hach()

Hash Table

OutPut (S.Read(2) + S.Read(2))



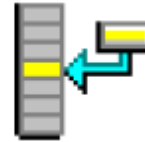
Hash Join



R

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

« B »



2

Hach()

Hash Table

S

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

Read S(i)

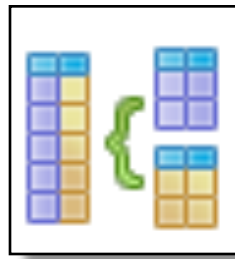
RID

RID

OutPut (S.Read(2) + S.Read(3))



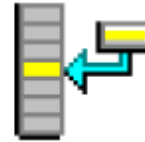
Hash Join



R

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

« E »



S

Read S(i)

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

RID

RID

3

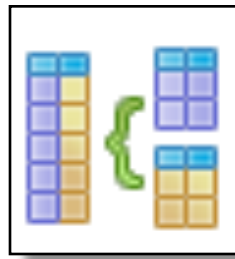
Hach()

Hash Table

OutPut (S.Read(3) + S.Read(3))



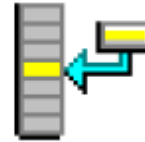
Hash Join



R

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

« G »



Hach()

Hash Table

S

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

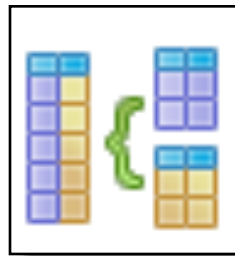
Read S(i)

RID

RID



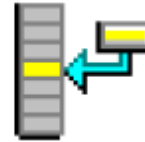
Hash Join



R

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

« J »



Hach()

Hash Table

S

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

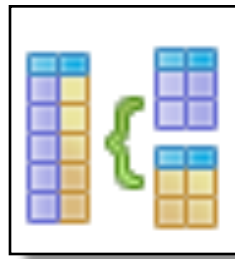
Read S(i)

RID

RID



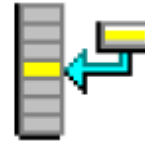
Hash Join



R

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

« K »



4

Hach()

Hash Table

S

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

Read S(i)

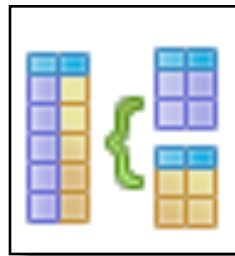
RID

RID

OutPut (S.Read(4) + S.Read(7))



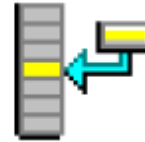
Hash Join



R

	Y	Z
1	Z	20
2	B	22
3	E	28
4	K	24
5	M	25
6	N	30
7	U	27
8	L	23
9	V	21
10	X	26

« U »



Hach()

7

Hash Table

S

	X	Y
1	5	A
2	8	B
3	3	B
4	9	E
5	1	G
6	10	J
7	7	K
8	2	U
9	4	V
10	6	Z

Read S(i)

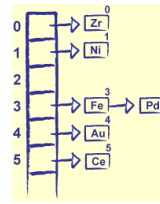
RID

RID

OutPut (S.Read(7) + S.Read(8))

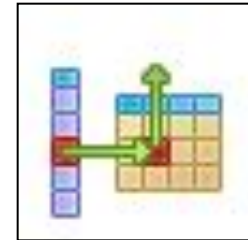


Hash join R S



- R is stored in the R hash table :

- Join Attribute is the Hash Key,
 - Y= Key,
- The RID is the value
 - RID = Value



- Seq scan on S to probe R hash table

- For each record of S we use the Join Attribute as the probe key in R hash table
 - If S Join Attribute probe is positive then the value is used as index to get R Join Attribute value.

