Harshitha Onkar
Sac Id: 220262706

Sort-Join Algorithm

Step 1: Initial Disk Representation of R and S

Disk Structure for R (r_disk): ArrayList<List<List<R>>>, Total no. of Blocks: 8

Disk Structure for S (s_disk): ArrayList<List<List<S>>>, Total no. of Blocks: 7

Y is the join key.

Following is the structural information of R and S.

	R(W, X, Y, M)	S(Y, Z, Q, T)
No. of Elements per tuple	4	4
No. of tuples per block	3	3
Total no. of blocks	8	7

Step 2: Load R into Memory and perform main memory sort on each sublist

Memory Structure for R (memory): ArrayList<List<R>> with memory block size of 6.

Disk Structure of R sublist (sortedSublistR): ArrayList<List<List<R>>>, 2 sublists

R is loaded into memory block by block, and main memory sorting is done. After sorting the sublist it is stored on disk.

Main memory sort is performed using: Collections.sort(memlist, Comparator.comparingInt(R :: getY))

Step 3: Load S into Memory and perform main memory sort on each sublist

Memory Structure for S (memory): ArrayList<List<S>> with memory block size of 6.

Disk Structure of S sublist (sortedSublistS): ArrayList<List<List<S>>>, 2 sublists

R is loaded into memory block by block, and main memory sorting is done. After sorting the sublist it is stored on disk as sortedSublistS.

Main memory sort is performed using: Collections.sort(memlist, Comparator.comparingInt(S :: getY))

Step 4: Load one block of each sorted sublist to memory

One block of each sublist is loaded into memory and the smallest y-value (join-key) among all the memory buffers is selected. After a block is loaded into memory buffer, it is deleted from the sublist.

For each memory buffer the following is done until all sublists are processed:

- if smallest == current_y and it is the last element of the block, then next block is loaded
- else if smallest == current y but not last element in block then index is incremented
- else if current y > smallest stop, break out of loop

Step 5: Join R and S on Y and put the resulting tuple to disk via output buffer

Joined Tuple (result): Result, a class is used to represent the joined tuple of R & S

Output buffer (output_buffer): ArrayList<Result>, used to put the results to the disk block by block

Final Disk Structure (disk): ArrayList<List<Result>>, final output the sort-join algorithm as stored on disk

After all the memory buffers are processed for the current smallest y, the matching tuples of R and S are joined on attribute Y. The joined tuples are added to output buffer. When the output buffer is full, the contents of the buffer are dumped to the disk.

The next smallest y is identified based on the current index position of all the memory buffers and Step 4 is repeated until all the sublists of any one relation is completely processed.

Input:

Relation R: Consists of 8 blocks, each block contains 3 tuples of R.

[
[R [w=1, x=1, y=3, m=1], R [w=1, x=1, y=0, m=1], R [w=1, x=1, y=3, m=1]],
[R [w=2, x=2, y=1, m=2], R [w=2, x=2, y=1, m=2], R [w=2, x=2, y=1, m=2]],
[R [w=3, x=3, y=4, m=3], R [w=3, x=3, y=1, m=3], R [w=3, x=3, y=4, m=3]],
[R [w=4, x=4, y=2, m=4], R [w=4, x=4, y=1, m=4], R [w=4, x=4, y=2, m=4]],
[R [w=5, x=5, y=3, m=5], R [w=5, x=5, y=3, m=5], R [w=5, x=5, y=666, m=5]],
[R [w=6, x=6, y=666, m=6], R [w=6, x=6, y=777, m=6], R [w=6, x=6, y=777, m=6]],
[R [w=7, x=7, y=555, m=7], R [w=7, x=7, y=4, m=7], R [w=7, x=7, y=5, m=7]],
[R [w=8, x=8, y=2, m=8], R [w=8, x=8, y=0, m=8], R [w=8, x=8, y=0, m=8]]
]

Relation S: Consists of 7 blocks, each block contains 3 tuples of S.

[
[S [y=2, z=1, q=1, t=1], S [y=1, z=1, q=1, t=1], S [y=3, z=1, q=1, t=1]],
[S [y=1, z=2, q=2, t=2], S [y=3, z=2, q=2, t=2], S [y=2, z=2, q=2, t=2]],
[S [y=3, z=3, q=3, t=3], S [y=2, z=3, q=3, t=3], S [y=1, z=3, q=3, t=3]],
[S [y=4, z=4, q=4, t=4], S [y=2, z=4, q=4, t=4], S [y=3, z=4, q=4, t=4]],
[S [y=888, z=5, q=5, t=5], S [y=1, z=5, q=5, t=5], S [y=888, z=5, q=5, t=5]],
[S [y=99, z=6, q=6, t=6], S [y=1, z=6, q=6, t=6], S [y=0, z=6, q=6, t=6]],
[S [y=5, z=7, q=7, t=7], S [y=0, z=7, q=7, t=7], S [y=888, z=7, q=7, t=7]]
]

Output:

[
[Result [w=1, x=1, y=0, m=1, z=6, q=6, t=6], Result [w=1, x=1, y=0, m=1, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6]],
[Result [w=8, x=8, y=0, m=8, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6], Result [w=8, x=8, y=0, m=8, z=7, q=7, t=7]],
[Result [w=2, x=2, y=1, m=2, z=1, q=1, t=1], Result [w=2, x=2, y=1, m=2, z=2, q=2, t=2], Result [w=2, x=2, y=1, m=2, z=3, q=3, t=3]],

[Result [w=2, x=2, y=1, m=2, z=5, q=5, t=5], Result [w=2, x=2, y=1, m=2, z=6, q=6, t=6], Result [w=2, x=2, y=1, m=2, z=1, q=1, t=1]],

[Result [w=2, x=2, y=1, m=2, z=2, q=2, t=2], Result [w=2, x=2, y=1, m=2, z=3, q=3, t=3], Result [w=2, x=2, y=1, m=2, z=5, q=5, t=5]],

[Result [w=2, x=2, y=1, m=2, z=6, q=6, t=6], Result [w=2, x=2, y=1, m=2, z=1, q=1, t=1], Result [w=2, x=2, y=1, m=2, z=2, q=2, t=2]],

[Result [w=2, x=2, y=1, m=2, z=3, q=3, t=3], Result [w=2, x=2, y=1, m=2, z=5, q=5, t=5], Result [w=2, x=2, y=1, m=2, z=6, q=6, t=6]],

[Result [w=3, x=3, y=1, m=3, z=1, q=1, t=1], Result [w=3, x=3, y=1, m=3, z=2, q=2, t=2], Result [w=3, x=3, y=1, m=3, z=3, q=3, t=3]],

[Result [w=3, x=3, y=1, m=3, z=5, q=5, t=5], Result [w=3, x=3, y=1, m=3, z=6, q=6, t=6], Result [w=4, x=4, y=1, m=4, z=1, q=1, t=1]],

[Result [w=4, x=4, y=1, m=4, z=2, q=2, t=2], Result [w=4, x=4, y=1, m=4, z=3, q=3, t=3], Result [w=4, x=4, y=1, m=4, z=5, q=5, t=5]],

[Result [w=4, x=4, y=1, m=4, z=6, q=6, t=6], Result [w=4, x=4, y=2, m=4, z=1, q=1, t=1], Result [w=4, x=4, y=2, m=4, z=2, q=2, t=2]],

[Result [w=4, x=4, y=2, m=4, z=3, q=3, t=3], Result [w=4, x=4, y=2, m=4, z=4, q=4, t=4], Result [w=4, x=4, y=2, m=4, z=1, q=1, t=1]],

[Result [w=4, x=4, y=2, m=4, z=2, q=2, t=2], Result [w=4, x=4, y=2, m=4, z=3, q=3, t=3], Result [w=4, x=4, y=2, m=4, z=4, q=4, t=4]],

[Result [w=8, x=8, y=2, m=8, z=1, q=1, t=1], Result [w=8, x=8, y=2, m=8, z=2, q=2, t=2], Result [w=8, x=8, y=2, m=8, z=3, q=3, t=3]],

[Result [w=8, x=8, y=2, m=8, z=4, q=4, t=4], Result [w=1, x=1, y=3, m=1, z=1, q=1, t=1], Result [w=1, x=1, y=3, m=1, z=2, q=2, t=2]],

[Result [w=1, x=1, y=3, m=1, z=3, q=3, t=3], Result [w=1, x=1, y=3, m=1, z=4, q=4, t=4], Result [w=1, x=1, y=3, m=1, z=1, q=1, t=1]],

[Result [w=1, x=1, y=3, m=1, z=2, q=2, t=2], Result [w=1, x=1, y=3, m=1, z=3, q=3, t=3], Result [w=1, x=1, y=3, m=1, z=4, q=4, t=4]],

[Result [w=5, x=5, y=3, m=5, z=1, q=1, t=1], Result [w=5, x=5, y=3, m=5, z=2, q=2, t=2], Result [w=5, x=5, y=3, m=5, z=3, q=3, t=3]],

[Result [w=5, x=5, y=3, m=5, z=4, q=4, t=4], Result [w=5, x=5, y=3, m=5, z=1, q=1, t=1], Result [w=5, x=5, y=3, m=5, z=2, q=2, t=2]],

[Result [w=5, x=5, y=3, m=5, z=3, q=3, t=3], Result [w=5, x=5, y=3, m=5, z=4, q=4, t=4], Result [w=3, x=3, y=4, m=3, z=4, q=4, t=4]],

[Result [w=3, x=3, y=4, m=3, z=4, q=4, t=4], Result [w=7, x=7, y=4, m=7, z=4, q=4, t=4], Result [w=7, x=7, y=5, m=7, z=7, q=7, t=7]]

]

Execution Results:

```
Relation R on Disk
[[R [w=1, x=1, y=3, m=1], R [w=1, x=1, y=0, m=1], R [w=1, x=1, y=3, m=1]], [R [w=2, x=2, y=1, m=2], R [w=2, x=2, y=1, m=2], R [w=2, x=2, y=1,


Relation S on Disk
[[S [y=2, z=1, q=1, t=1], S [y=1, z=1, q=1, t=1], S [y=3, z=1, q=1, t=1]], [S [y=1, z=2, q=2, t=2], S [y=3, z=2, q=2, t=2], S [y=2, z=2, q=2,

-------------------------------------------------------------------------------------------
Sorted sublists of Relation R on Disk

Sublist 1 of R
[[R [w=1, x=1, y=0, m=1], R [w=2, x=2, y=1, m=2], R [w=2, x=2, y=1, m=2]], [R [w=2, x=2, y=1, m=2], R [w=3, x=3, y=1, m=3], R [w=4, x=4, y=1,

Sublist 2 of R
[[R [w=8, x=8, y=0, m=8], R [w=8, x=8, y=0, m=8], R [w=8, x=8, y=2, m=8]], [R [w=7, x=7, y=4, m=7], R [w=7, x=7, y=5, m=7], R [w=7, x=7, y=55

-------------------------------------------------------------------------------------------
Sorted sublists of Relation S on Disk

Sublist 1 of S
[[S [y=0, z=6, q=6, t=6], S [y=1, z=1, q=1, t=1], S [y=1, z=2, q=2, t=2]], [S [y=1, z=3, q=3, t=3], S [y=1, z=5, q=5, t=5], S [y=1, z=6, q=6,

Sublist 2 of S
[[S [y=0, z=7, q=7, t=7], S [y=5, z=7, q=7, t=7], S [y=888, z=7, q=7, t=7]]]

*******************************************************************************************************************************************

*******************************************************************************************************************************************
Smallest Y-value:0

Matching tuples of R
[R [w=1, x=1, y=0, m=1], R [w=8, x=8, y=0, m=8], R [w=8, x=8, y=0, m=8]]
Matching tuples of S
[S [y=0, z=6, q=6, t=6], S [y=0, z=7, q=7, t=7]]
-------------------------------------------------------------------------------------------
Output buffer contents:
[Result [w=1, x=1, y=0, m=1, z=6, q=6, t=6], Result [w=1, x=1, y=0, m=1, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6]]
-------------------------------------------------------------------------------------------
Disk contents:
[[Result [w=1, x=1, y=0, m=1, z=6, q=6, t=6], Result [w=1, x=1, y=0, m=1, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6]]]

*******************************************************************************************************************************************
Smallest Y-value:1

Matching tuples of R
[R [w=2, x=2, y=1, m=2], R [w=2, x=2, y=1, m=2], R [w=2, x=2, y=1, m=2], R [w=3, x=3, y=1, m=3], R [w=4, x=4, y=1, m=4]]
Matching tuples of S
[S [y=1, z=1, q=1, t=1], S [y=1, z=2, q=2, t=2], S [y=1, z=3, q=3, t=3], S [y=1, z=5, q=5, t=5], S [y=1, z=6, q=6, t=6]]
-------------------------------------------------------------------------------------------
Output buffer contents:
[Result [w=8, x=8, y=0, m=8, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6], Result [w=8, x=8, y=0, m=8, z=7, q=7, t=7]]
[Result [w=2, x=2, y=1, m=2, z=1, q=1, t=1], Result [w=2, x=2, y=1, m=2, z=2, q=2, t=2], Result [w=2, x=2, y=1, m=2, z=3, q=3, t=3]]
[Result [w=2, x=2, y=1, m=2, z=5, q=5, t=5], Result [w=2, x=2, y=1, m=2, z=6, q=6, t=6], Result [w=2, x=2, y=1, m=2, z=1, q=1, t=1]]
[Result [w=2, x=2, y=1, m=2, z=2, q=2, t=2], Result [w=2, x=2, y=1, m=2, z=3, q=3, t=3], Result [w=2, x=2, y=1, m=2, z=5, q=5, t=5]]
[Result [w=2, x=2, y=1, m=2, z=6, q=6, t=6], Result [w=2, x=2, y=1, m=2, z=1, q=1, t=1], Result [w=2, x=2, y=1, m=2, z=2, q=2, t=2]]
[Result [w=2, x=2, y=1, m=2, z=3, q=3, t=3], Result [w=2, x=2, y=1, m=2, z=5, q=5, t=5], Result [w=2, x=2, y=1, m=2, z=6, q=6, t=6]]
[Result [w=3, x=3, y=1, m=3, z=1, q=1, t=1], Result [w=3, x=3, y=1, m=3, z=2, q=2, t=2], Result [w=3, x=3, y=1, m=3, z=3, q=3, t=3]]
[Result [w=3, x=3, y=1, m=3, z=5, q=5, t=5], Result [w=3, x=3, y=1, m=3, z=6, q=6, t=6], Result [w=4, x=4, y=1, m=4, z=1, q=1, t=1]]
[Result [w=4, x=4, y=1, m=4, z=2, q=2, t=2], Result [w=4, x=4, y=1, m=4, z=3, q=3, t=3], Result [w=4, x=4, y=1, m=4, z=5, q=5, t=5]]
-------------------------------------------------------------------------------------------
Disk contents:
[[Result [w=1, x=1, y=0, m=1, z=6, q=6, t=6], Result [w=1, x=1, y=0, m=1, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6]], [Result
```

```
**********************************************************************************************************************************************
Smallest Y-value:2

Matching tuples of R
[R [w=4, x=4, y=2, m=4], R [w=4, x=4, y=2, m=4], R [w=8, x=8, y=2, m=8]]
Matching tuples of S
[S [y=2, z=1, q=1, t=1], S [y=2, z=2, q=2, t=2], S [y=2, z=3, q=3, t=3], S [y=2, z=4, q=4, t=4]]
----------------------------------------------------------------------------------------------
Output buffer contents:
[Result [w=4, x=4, y=1, m=4, z=6, q=6, t=6], Result [w=4, x=4, y=2, m=4, z=1, q=1, t=1], Result [w=4, x=4, y=2, m=4, z=2, q=2, t=2]]
[Result [w=4, x=4, y=2, m=4, z=3, q=3, t=3], Result [w=4, x=4, y=2, m=4, z=4, q=4, t=4], Result [w=4, x=4, y=2, m=4, z=1, q=1, t=1]]
[Result [w=4, x=4, y=2, m=4, z=2, q=2, t=2], Result [w=4, x=4, y=2, m=4, z=3, q=3, t=3], Result [w=4, x=4, y=2, m=4, z=4, q=4, t=4]]
[Result [w=8, x=8, y=2, m=8, z=1, q=1, t=1], Result [w=8, x=8, y=2, m=8, z=2, q=2, t=2], Result [w=8, x=8, y=2, m=8, z=3, q=3, t=3]]
----------------------------------------------------------------------------------------------
Disk contents:
[[Result [w=1, x=1, y=0, m=1, z=6, q=6, t=6], Result [w=1, x=1, y=0, m=1, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6]], [Result

**********************************************************************************************************************************************
Smallest Y-value:3

Matching tuples of R
[R [w=1, x=1, y=3, m=1], R [w=1, x=1, y=3, m=1], R [w=5, x=5, y=3, m=5], R [w=5, x=5, y=3, m=5]]
Matching tuples of S
[S [y=3, z=1, q=1, t=1], S [y=3, z=2, q=2, t=2], S [y=3, z=3, q=3, t=3], S [y=3, z=4, q=4, t=4]]
----------------------------------------------------------------------------------------------
Output buffer contents:
[Result [w=8, x=8, y=2, m=8, z=4, q=4, t=4], Result [w=1, x=1, y=3, m=1, z=1, q=1, t=1], Result [w=1, x=1, y=3, m=1, z=2, q=2, t=2]]
[Result [w=1, x=1, y=3, m=1, z=3, q=3, t=3], Result [w=1, x=1, y=3, m=1, z=4, q=4, t=4], Result [w=1, x=1, y=3, m=1, z=1, q=1, t=1]]
[Result [w=1, x=1, y=3, m=1, z=2, q=2, t=2], Result [w=1, x=1, y=3, m=1, z=3, q=3, t=3], Result [w=1, x=1, y=3, m=1, z=4, q=4, t=4]]
[Result [w=5, x=5, y=3, m=5, z=1, q=1, t=1], Result [w=5, x=5, y=3, m=5, z=2, q=2, t=2], Result [w=5, x=5, y=3, m=5, z=3, q=3, t=3]]
[Result [w=5, x=5, y=3, m=5, z=4, q=4, t=4], Result [w=5, x=5, y=3, m=5, z=1, q=1, t=1], Result [w=5, x=5, y=3, m=5, z=2, q=2, t=2]]
----------------------------------------------------------------------------------------------
Disk contents:
[[Result [w=1, x=1, y=0, m=1, z=6, q=6, t=6], Result [w=1, x=1, y=0, m=1, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6]], [Result
```

```
**********************************************************************************************************************************************
Final Disk Result

[[Result [w=1, x=1, y=0, m=1, z=6, q=6, t=6], Result [w=1, x=1, y=0, m=1, z=7, q=7, t=7], Result [w=8, x=8, y=0, m=8, z=6, q=6, t=6]], [Result
```