



Join Stage



What is a Join Stage?



Join Stage:



- The Join stage is a processing stage.
- Join stage is one of three stages that join tables based on the values of key columns. The other two are Lookup stage and Merge stage
- It can have a left link, a right link and multiple intermediate links.
- It has any number of input links and a single output link.
- The data sets input to the Join stage must be key partitioned and sorted.

Inner join outputs records when the input datasets have identical key values.

- **Left outer join:**

outputs all left link records and transfers attributes from matching(on key value) right link records

- **Right outer join:**

outputs all right link records and transfers attributes from matching(on key value) left link records

- **Full outer join:**

transfers records from both left and right links.

- The values for attributes for non-matching records are populated as Nulls.

Join stage: Partitioning



- **Partitioning in Join stage:**
- The data sets input to the Join stage must be key partitioned and sorted.
- This ensures that rows with the same key column values are located in the same partition and will be processed by the same node.
- It also minimizes memory requirements because fewer rows need to be in memory at any one time.

Thank You

Email us – support@intellipaat.com

Visit us - <https://intellipaat.com>