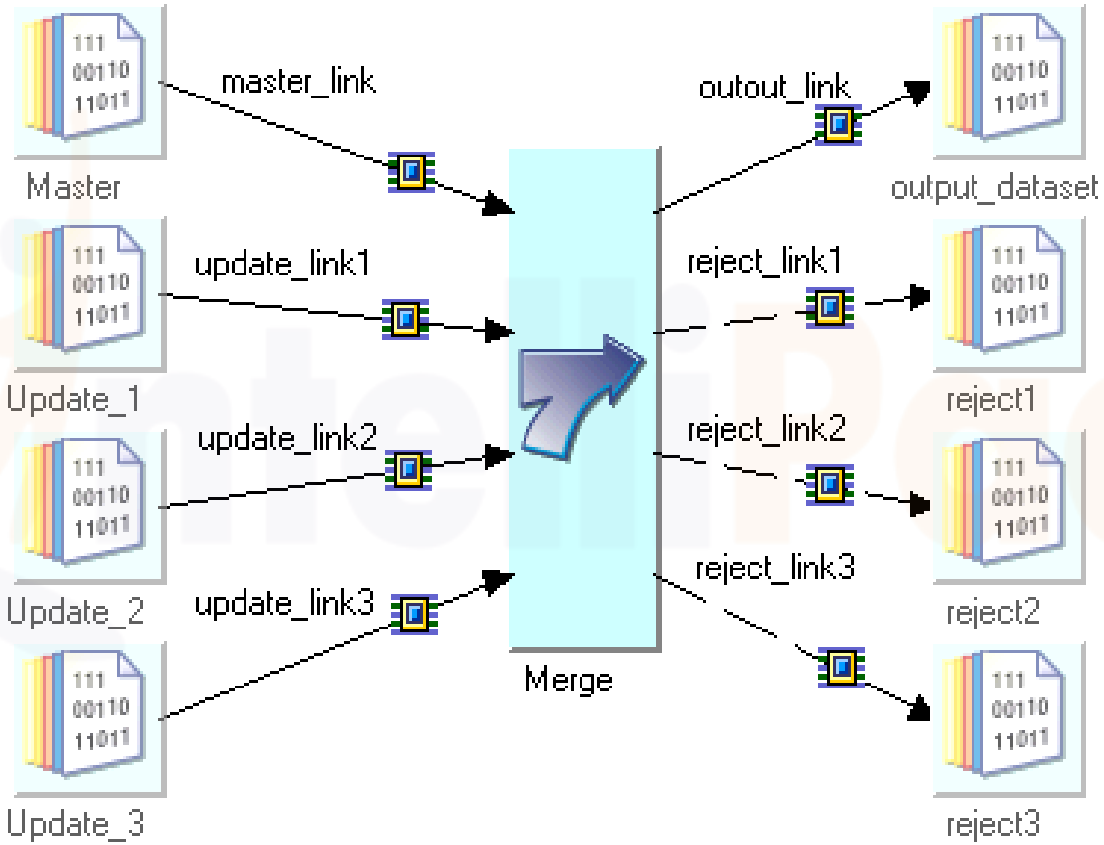
Merge Stage

What do you mean by **Merge Stage?**

Merge Stage:

- The Merge stage is a processing stage.
- It can have any number of input links, a single output link, and the same number of reject links as there are update input links.
- The Merge stage is one of three stages that join tables based on the values of key columns. The other two are: Join stage and Lookup stage.
- The Merge stage combines a master data set with one or more update datasets.
- A master record and an update record are merged only if both of them have the same values for the merge key column(s) that you specify.

Merge Stage:



Partitioning in Merge stage:

The data sets input to the Merge 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.

-As part of preprocessing ,duplicate records should be removed from the master data set.

-In case of more than one update data sets, duplicate records from the update data sets must be removed as well.

Join Vs Lookup Vs Merge:

- Lookup is faster when reference dataset is small and can fit into RAM.
- Join gives better performance with larger reference dataset by sorting data on input links.
- Merge stage is used when multiple update and reject links are needed.

Thank You

Email us – support@intellipaat.com

Visit us - https://intellipaat.com