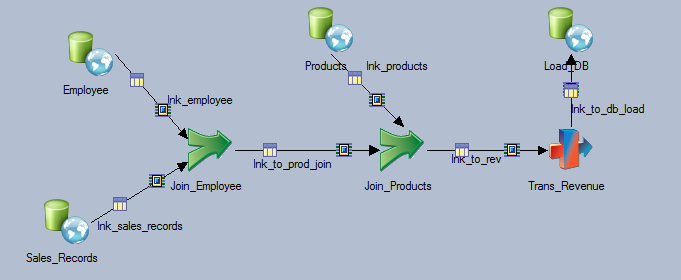
**Join With DataStage**

The Join stage is one of three stages that join tables based on the values of key columns. The others are Lookup stage and Merge stage. Join is usually used when you have large reference dataset (i.e. does not fit in RAM). If the reference dataset is small enough to fit in RAM, it is faster to use Lookup. The performance of Join can be improved by sorting data on input links.

**Key points for the Join Stage**

* The key columns must be the same name between tables.
* It can have multiple input links (as long as table has the same key columns between them) and a single output link.
* The performance of Join can be improved by key-sorting data on input links (‘Auto’ partitioning mode is usually fine).
* If the reference dataset is small enough to fit in RAM, it is faster to use Lookup.
* There are four join options: inner join, left outer join, right outer join and full outer join.
* We need to make sure input links are in the right order. This can be set from Stage 🡪 Link Ordering.

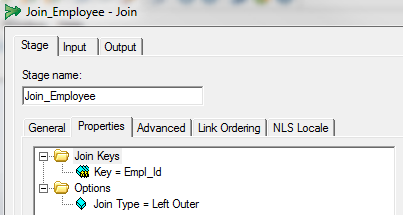
**Example**



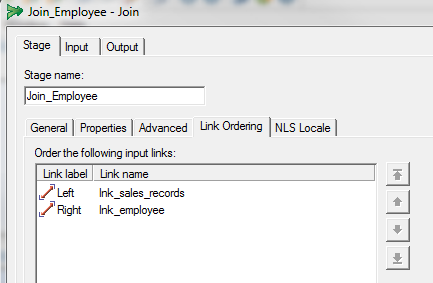
In this example, we join Employee and Products tables to Sales\_Records based on Emp\_Id and Product\_Id. Then, calculate the revenue by multiplying the price column from Products by the number of units sold.

**Key Steps**

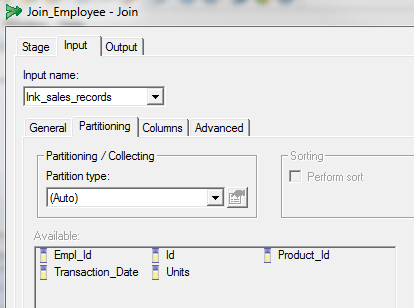
1. **In each Join stage, choose join key and type.**



**In the Link Ordering Tab, make sure the link order is correct.**



**Partitioning can be auto.**



**In the Transformer stage, you can calculate Revenue by multiplying Unit\_Price by Units. The data type for Units is interger and Unit\_Price is double. Therefore, set the Revenue’s data type as double.**

