My design is pretty simple :

First I create a hashtable associating customer keys to the columns that are selected from the customer relation (in this case, custkey).
I then filter the order relation to keep only the one for which the join attribute is in the hashtable. I take the columns that are in the select statement and concatenate them with the one from the customer relation (by looking in the hashtable).

Once this is done I save the returned rdd in a file