

# Join

In this tutorial you will learn how to combine information from different sources.

We will use the same setting as in [Simple Query Processing](#). So you should follow steps 1-4.

## Join

One of the most important operators to combine events from different streams is the join operator.

To combine two elements from different streams, two conditions must hold:

1. The time intervals must overlap
2. The join predicate must evaluate to true

Create a new PQL based script (join1.qry) with the following input:

```
#PARSER PQL
#TRANSCFG Standard
#ADDQUERY
out = JOIN({PREDICATE='nexmark:person.id = nexmark:bid.bidder'},nexmark:person, nexmark:bid)
```

This query add to each bid the person that is bidding.

After starting you should see something like in the following.

| timestamp | id | name              | email              | creditcard         | city       | state   | timestamp | auction | bidder | datetime | price | Metadata |
|-----------|----|-------------------|--------------------|--------------------|------------|---------|-----------|---------|--------|----------|-------|----------|
| 15000     | 2  | Pius Shmuely      | Shmuely@earthli... | 9106 9151 8141 ... | Gothenburg | Vermont | 15000     | 3       | 2      | 15001    | 219.0 | 15000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...  | 8116 3675 7280 ... | Abidjan    | Schrodl | 14500     | 2       | 1      | 14500    | 211.0 | 14500 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...  | 8116 3675 7280 ... | Abidjan    | Schrodl | 14000     | 1       | 1      | 14000    | 262.0 | 14000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...  | 8116 3675 7280 ... | Abidjan    | Schrodl | 13500     | 1       | 1      | 13501    | 248.0 | 13500 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 13000     | 1       | 0      | 13012    | 231.0 | 13000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...  | 8116 3675 7280 ... | Abidjan    | Schrodl | 12500     | 1       | 1      | 12500    | 221.0 | 12500 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 12000     | 0       | 0      | 12001    | 150.0 | 12000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...  | 8116 3675 7280 ... | Abidjan    | Schrodl | 11500     | 0       | 1      | 11500    | 141.0 | 11500 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 11000     | 1       | 0      | 11000    | 199.0 | 11000 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 10500     | 1       | 0      | 10501    | 192.0 | 10500 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...  | 8116 3675 7280 ... | Abidjan    | Schrodl | 10000     | 0       | 1      | 10001    | 126.0 | 10000 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 9500      | 1       | 0      | 9500     | 182.0 | 9500 oo  |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 9000      | 1       | 0      | 9000     | 161.0 | 9000 oo  |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 8500      | 0       | 0      | 8500     | 103.0 | 8500 oo  |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 8000      | 0       | 0      | 8000     | 78.0  | 8000 oo  |
| 5000      | 0  | Luitpold Martucci | Martucci@toront... | 6573 1248 4234 ... | Pensacola  | Georgia | 7500      | 0       | 0      | 7521     | 53.0  | 7500 oo  |

WARNING: In this example we use no windows. For Joins this is typically not recommended, as it can lead to memory overflow!

In most cases it is easy to define a window for each input stream.

There are some case, where you can speed up processing or cope with the case where no windows can be defined by using the parameter CARD:

- ONE\_ONE: In each input stream there is exactly one corresponding object. With this setting, windows can potentially be avoided.
- ONE\_MANY: Each element in the right input stream has exactly one corresponding element in the left input stream
- MANY\_ONE: Each element in the left input stream has exactly one corresponding element in the right input stream
- MANY\_MANY: For each element in both input streams there may be multiple corresponding elements. This is the default.

For the example above, each bid has exactly one person. So we can write:

```
#PARSER PQL
#TRANSCFG Standard
#ADDQUERY
out = JOIN({PREDICATE='nexmark:person.id = nexmark:bid.bidder', CARD='ONE_MANY'},nexmark:person, nexmark:bid)
```

| timestamp | id | name              | email               | creditcard         | city       | state    | timestamp | auction | bidder | datetime | price | Metadata |
|-----------|----|-------------------|---------------------|--------------------|------------|----------|-----------|---------|--------|----------|-------|----------|
| 20000     | 3  | Wing Kusalik      | Kusalik@bellatla... | 2536 1450 4675 ... | Idaho      | Illinois | 20000     | 5       | 3      | 20000    | 76.0  | 20000 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 19500     | 1       | 0      | 19502    | 296.0 | 19500 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 19000     | 3       | 1      | 19001    | 257.0 | 19000 oo |
| 15000     | 2  | Pius Shmuely      | Shmuely@earthli...  | 9106 9151 8141 ... | Gothenburg | Vermont  | 18500     | 1       | 2      | 18500    | 273.0 | 18500 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 18000     | 0       | 1      | 18001    | 162.0 | 18000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 17500     | 2       | 1      | 17500    | 236.0 | 17500 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 17000     | 0       | 1      | 17001    | 151.0 | 17000 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 16500     | 1       | 0      | 16501    | 267.0 | 16500 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 16000     | 3       | 0      | 16001    | 241.0 | 16000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 15500     | 3       | 1      | 15512    | 237.0 | 15500 oo |
| 15000     | 2  | Pius Shmuely      | Shmuely@earthli...  | 9106 9151 8141 ... | Gothenburg | Vermont  | 15000     | 3       | 2      | 15000    | 219.0 | 15000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 14500     | 2       | 1      | 14500    | 211.0 | 14500 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 14000     | 1       | 1      | 14016    | 262.0 | 14000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 13500     | 1       | 1      | 13500    | 248.0 | 13500 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 13000     | 1       | 0      | 13000    | 231.0 | 13000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 12500     | 1       | 1      | 12500    | 221.0 | 12500 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 12000     | 0       | 0      | 12001    | 150.0 | 12000 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 11500     | 0       | 1      | 11502    | 141.0 | 11500 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 11000     | 1       | 0      | 11001    | 199.0 | 11000 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 10500     | 1       | 0      | 10501    | 192.0 | 10500 oo |
| 10000     | 1  | Rattasit Schroer  | Schroer@ibm.co...   | 8116 3675 7280 ... | Abidjan    | Schrodl  | 10000     | 0       | 1      | 10001    | 126.0 | 10000 oo |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 9500      | 1       | 0      | 9501     | 182.0 | 9500 oo  |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 9000      | 1       | 0      | 9001     | 161.0 | 9000 oo  |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 8500      | 0       | 0      | 8501     | 103.0 | 8500 oo  |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 8000      | 0       | 0      | 8001     | 78.0  | 8000 oo  |
| 5000      | 0  | Luitpold Martucci | Martucci@toront...  | 6573 1248 4234 ... | Pensacola  | Georgia  | 7500      | 0       | 0      | 7500     | 53.0  | 7500 oo  |

And this lead to the same output.