Understanding Join and Concatenate

The Qlikview script functions JOIN and CONCATENATE can sometimes be used to tackle the same problem, but there are important differences that should be understood.

Examine the sample tables below. Note that they share one common field name, "Key". Also note that Table1 has a Key value "3" that is not present in Table2.

Table1

Key	A	B
1	A1	B1
2	A2	B2
3	A3	B3

Table2

Key	C	D
1	C1	D1
2	C2	D2

JOIN will combine rows where the Key value matches. The keyword OUTER will also retain rows that do not match rows in the other table. Here's what the merged table will look like after an outer join.

```
OUTER JOIN (Table1) LOAD * RESIDENT Table2;
```

JOIN Result

Row	Key	A	B	C	D
1	1	A1	B1	C1	D1
2	2	A2	B2	C2	D2
3	3	A3	B3	-	-

Values A1 and C1, which were in different tables, now occupy the same row in the result table. The row with Key 3 has missing values for C & D, because there was no matching Key in Table2.

Creating a chart that uses "Key" for dimension will produce results similar to the Table Box above.

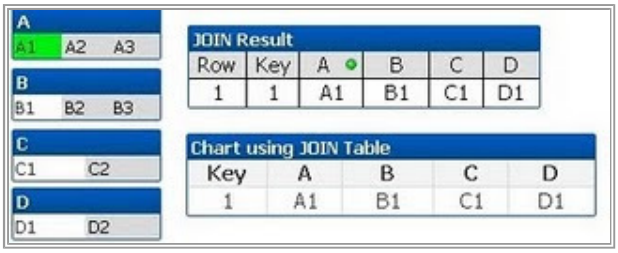
The important

Chart using JOIN Table

Key	A	B	C	D
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	-	-

point is that values with the same Key value have been merged together into a single row. If value A1 is selected, note that values C1 & D1 remain associated (white). The set A1,B1,C1,D1 is indivisible.

Now let's look at Concatenate. Concatenate appends the rows of one table to another.



Concatenate never merges any rows. The number of rows in a concatenated table is always the sum of the rows from the two input tables. Here's what our sample data will look like after Concatenate.

```
CONCATENATE (Table1) LOAD * RESIDENT Table2;
```

CONCATENATE Result

Row	Key	A	B	C	D
1	1	A1	B1	-	-
2	2	A2	B2	-	-
3	3	A3	B3	-	-
4	1	-	-	C1	D1
5	2	-	-	C2	D2

Rows with like Key values are **not** merged together. The rows from Table2 are simply appended to Table1. Because the tables have different fields, rows will have null values for the fields from the "other" table.

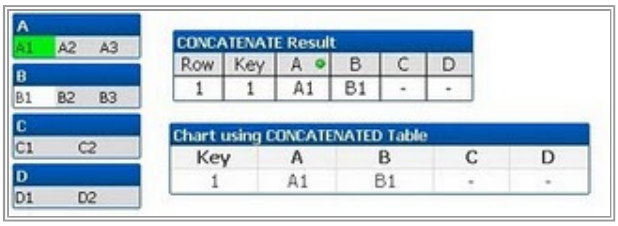
If the data is used to build a chart that utilizes the common field "Key" as dimension, the chart looks just like the JOINed table.

Let's make the selection "A1" in

Chart using CONCATENATED Table

Key	A	B	C	D
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	-	-

Field A and see it's impact on our visible charts and tables.



When A1 is selected, the association to C1 & D1 is lost and C/D values become null in both the Chart and Tablebox. We cannot select both A1 and C1. This is a different result than the JOINed example.

Let's consider a more realistic example where we may choose between JOIN and CONCATENATE. Consider the two tables below. Note that only one BudgetAmount row is present for each Region-Year combination. In the Sales table, the SalesAmount is broken down by Department within Region.

Budget

Region	Year	BudgetAmount
West	2008	10,000
West	2009	6,000
East	2008	1,500
East	2009	1,000

Sales

Region	Year	Department	SalesAmount
West	2008	Clothes	150
West	2008	Sports	3,000
West	2008	Toys	4,000
West	2009	Sports	200
West	2009	Toys	300
East	2008	Sports	1,490
East	2008	Toys	1,200
East	2009	Sports	500
East	2009	Toys	480

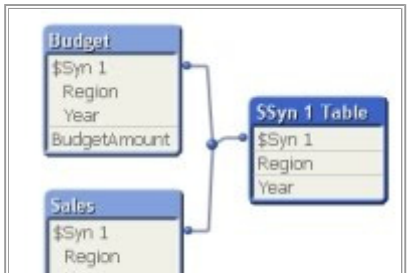
If we load both tables we can produce a chart using expressions like =Sum (BudgetAmount).

Budget vs Sales

Year	Region	Budget	Sales	Variance
2008	East	1,500	2,690	1,190
2008	West	10,000	7,150	-2,850
2009	East	1,000	980	-20
2009	West	6,000	500	-5,500
Total		**18,500**	**11,320**	**-7,180**

The Budget and Sales values have been summed correctly.

We then notice that we have an undesirable synthetic key, created by the Budget and Sales tables sharing the Year



and Region fields. One approach to eliminate the synthetic key would be JOIN or CONCATENATE. But which one in this case?

Let's try JOIN and see what the Chart looks like.

```
OUTER JOIN (Budget) LOAD * RESIDENT SALES
```

Budget

Year	Region	Budget	Sales	Variance
2008	East	3,000	2,690	-310
2008	West	30,000	7,150	-22,850
2009	East	2,000	980	-1,020
2009	West	12,000	500	-11,500
Total		47,000	11,320	-35,680

The summed Budget numbers are incorrect!

A look at raw data of the joined table will identify the problem. The JOIN repeated the BudgetAmount value on each Department row.

Let's try with

JOINed Data

Year	Region	Department	BudgetAmount	SalesAmount
2008	East	Sports	1,500	1,490
2008	East	Toys	1,500	1,200
2008	West	Clothes	10,000	150
2008	West	Sports	10,000	3,000
2008	West	Toys	10,000	4,000
2009	East	Sports	1,000	500
2009	East	Toys	1,000	480
2009	West	Sports	6,000	200
2009	West	Toys	6,000	300

CONCATENATE.

```
CONCATENATE (Budget) LOAD * RESIDENT Sales;
```

Budget

Year	Region	Budget	Sales	Variance
2008	West	10,000	7,150	-2,850
2008	East	1,500	2,690	1,190
2009	West	6,000	500	-5,500
2009	East	1,000	980	-20
Total		18,500	11,320	-7,180

The numbers are now correct and we've accomplished the goal of eliminating the synthetic key.

A peek at the data in the Concatenated table will make it clear why the chart is now correct. There is only BudgetAmount value or each Year-Region.

JOIN and

CONCATENATED Data

Year	Region	Department	BudgetAmount	SalesAmount
2008	East	Sports	-	1,490
2008	East	Toys	-	1,200
2008	East	-	1,500	-
2008	West	Clothes	-	150
2008	West	Sports	-	3,000
2008	West	Toys	-	4,000
2008	West	-	10,000	-
2009	East	Sports	-	500
2009	East	Toys	-	480
2009	East	-	1,000	-
2009	West	Sports	-	200
2009	West	Toys	-	300
2009	West	-	6,000	-

CONCATENATE are both very useful and frequently used functions in Qlikview. It's important to understand the differences between them.

-Rob

Read the complete post at http://qlikviewnotes.blogspot.com/2009/11/understanding-join-and-concatenate.html

Posted 11-20-2009 12:39 by Qlikview Notes | Add post to favorites | Add blog to favorites