



Limiting Data Load

Introduction

Methodologies exist within the QlikView script to control the values being loaded into a specific table – and furthermore, into the data model. You can limit the data to only load in data that matches specific values within a field(s) or based on values which do/do not already exist in the same data model. This white paper details different criteria that can be used with this function. A second methodology of achieving this is also detailed below which takes into account the Keep() function. The examples used throughout this white paper are constructed using QlikView 12.0 SR5.

Where Clause

Using Where enables only load in data where a specific string is matched, or where a numeric field values lies above a specific threshold or data which is prior to a given date. Many more scenarios can be used however these are a selected few. An example of these three are given with syntax.

```
Facts:
LOAD      Name,
          Item,
          Quantity,
          Value,
          Date
FROM
[..\DataSource\Data.xlsx]
(ooxml, embedded labels, table is Data)
Where
```

The space after the Where clause is left blank – examples of different criteria is stated below:

```
Where Item = 'Jacket';
Where Quantity <> 10;
Where DateofSale > Date('01/01/2017', 'DD/MM/YYYY') *1;
```

The above examples will only pull in data values which meet the specified criteria – for example the if the first line is used, only rows where the Item field is listed as ‘Jacket’ will be loaded. Furthermore, more than one criteria can be used as shown below. Here we are only including the Item ‘Jacket’ where the Quantity field that is not equal to 10.

```
Where Item = 'Jacket' and Quantity <> 10;
```



Where Exists()

Where Exists() can be used in similar fashion to the above and in this case is to limit data based on data which already exists within the data model. An example of this is you may wish to pull in data from an external product details table that can be tied to a Product ID already loaded into the facts table.



Figure 1 – Data model of Facts (10 unique records) and Product Details tables

In this scenario we only want to load in records from Product Details which relate to any of the ten unique Product IDs in Facts. To accomplish this, one method is to use Where Exists (X,Y) in the load script.

X = Field that exists within the current data model (Product ID in Facts)

Y = Field within the current table being loaded (productid in Product Details)

Keep()

Another method is to use the Keep function – similar to a Join, however here the two tables will be kept separate (whereas with a join you are left with one table). Using Keep() will load the combination of values that belong to any matching fields between the two tables - the table which you are limiting the values to be loaded into the data model depends on the direction of the keep function (more of which will be detailed in a separate white paper). The syntax for the above scenario is shown below.

```
Facts:
LOAD [Order ID],
     [Product ID]
FROM
[..\DataSource\Orders.xlsx]
(ooxml, embedded labels, table is Orders);

Left Keep(Facts)
LOAD productid,
     [Product Name]
FROM
[..\DataSource\ProductDetails.xlsx]
(ooxml, embedded labels, table is ProductDetails);
```



The above syntax will only load in records where the productid value exists in the [Product ID] field (Facts).

TIP: Both the above method works and for where you would like to load values which do not match any values that current exist in a previously loaded table, the following two options are available.

- Where not Exists(x,y) – the same criteria applies here to Where Exists()
- Keep() – in this scenario we would use Right Keep() – the records in the second table would be loaded which have no matching values with a previously loaded table.

Using Where Exists()/Where not Exists() breaks optimization

Where Exists() cannot be used with direct query to a source database – a QlikView function may not be interpreted as expected in direct database query. Alternatively you can load all the data use Keep() to load data based on what does/does not exist in the data model.

If you have any queries, please contact Ricky Tanna.

