



QlikView 11.2 DIRECT DISCOVERY

QlikView Technical Addendum

Published: November, 2012

www.qlikview.com

A decorative graphic in the top left corner of the page, consisting of several overlapping, flowing green lines that create a sense of movement and depth. The lines are thin and have a slight transparency, allowing them to overlap and create darker shades of green.

Overview

This document provides a technical overview of the QlikView 11.2 Direct Discovery feature.

QlikView Direct Discovery capability expands the potential use cases for Business Discovery, enabling business users to conduct associative analysis on big data sources. It provides QlikView's complete associative experience on top of data coming directly from external big data sources, and enables users to combine that big data with data stored in memory. With QlikView Direct Discovery, business users can leverage any data useful for analysis without scalability limitations.

The following part of the paper provides a technical overview of implementing and using QlikView Direct Discovery. It also provides information on the best practices and limitations of this feature with this release.

What is QlikView Direct Discovery

QlikView Direct Discovery capability combines the associative capabilities of the QlikView in memory dataset with a query model where the source data is not directly loaded into the QlikView data model, instead the aggregated query result is passed back to QlikView user interface. The direct discovery data set is still part of the associative experience where the user can navigate both on the in-memory data and the direct discovery data associatively.

The business users can associatively make selections on either of the data sets, and see what is associated and not associated with the same QlikView association colours; green, gray, and white. They can create charts that would help them analyze data from both data sets together.

A decorative graphic in the top left corner of the page, consisting of several overlapping, flowing green lines that create a sense of movement and depth.

Technical Details

With QlikView's unique associative experience in combination with Direct Discovery users are able to navigate and interact with data by a combination of methods.

I. HOW DOES DIRECT DISCOVERY WORK

QlikView decides which data resides in-memory and which data is direct discovery data by using the special script syntax, "DIRECT SELECT". This syntax allows certain data elements not to be loaded into the QlikView data model during the script reload process, but still available for query purposes from the QlikView User Interface and to be combined for analysis with the QlikView in memory dataset.

Once the direct discovery structure is established, the direct discovery fields can be used with certain QlikView objects. When a direct discovery field is used on the QlikView object, QlikView will automatically create the appropriate SQL query to run on the external data source. The result of the query will be displayed on the QlikView object. When selections are made on the QlikView application, the associated data values of the direct discovery fields will be used in the WHERE conditions of the queries. With each selection, the direct discovery charts will be calculated, where the calculations and aggregations will be done on the source table by executing the SQL query created by QlikView. It is possible to use calculation condition feature of the QlikView charts to set a condition indicating when the chart should be calculated. Until that condition is met, QlikView will not run queries and the chart will not be calculated. Please note that QlikView will execute SQL queries on the data source for some of the list boxes that use direct discovery fields. This is required to achieve the associative navigation capability.

II. DATA LOADING

Within the script editor a new syntax is introduced to connect to data in direct discovery form. Traditionally the following syntax is required to load data from a database table:

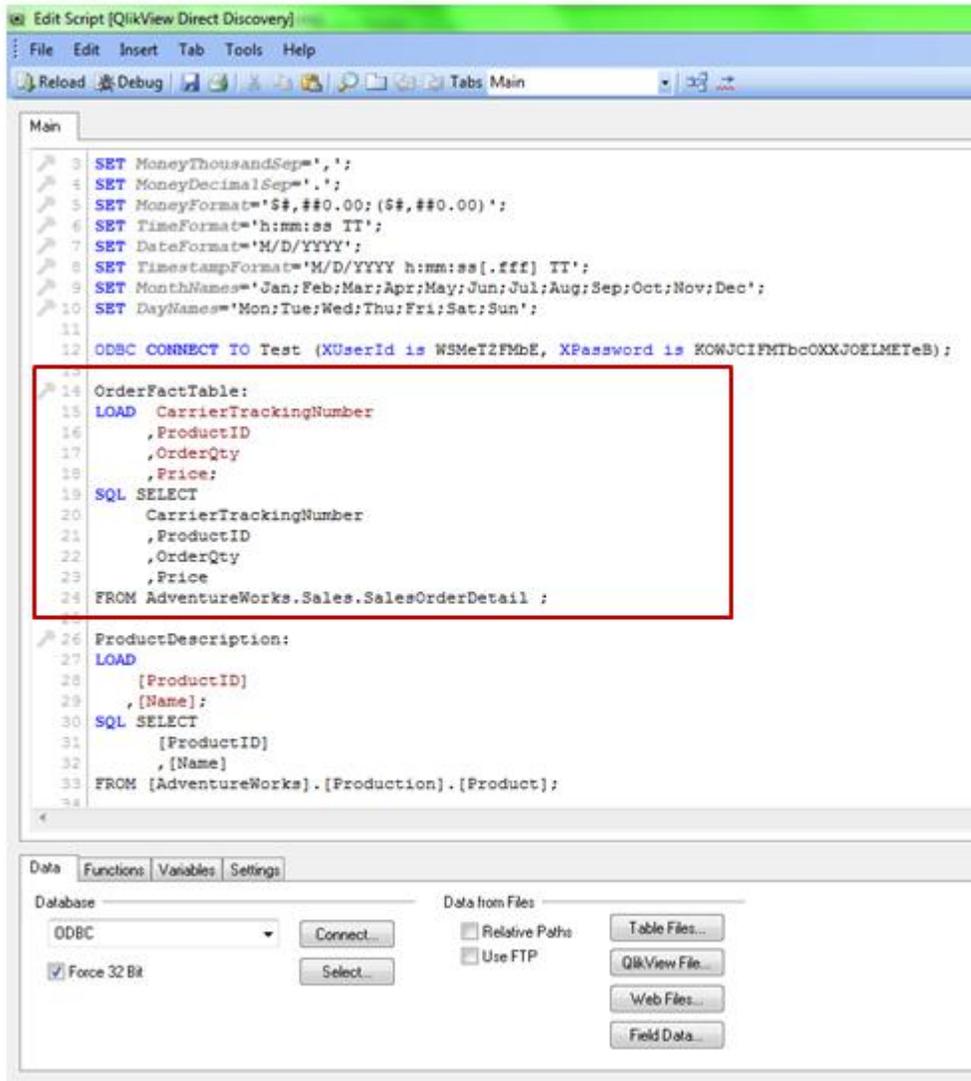


Figure 1. Traditional QlikView Load Script Syntax

To invoke the direct discovery method, the keyword “SQL SELECT” is replaced with “DIRECT SELECT”:

```

2 SET DecimalSep='.';
3 SET MoneyThousandSep=',';
4 SET MoneyDecimalSep='.';
5 SET MoneyFormat='$#,##0.00;($#,##0.00)';
6 SET TimeFormat='h:mm:ss TT';
7 SET DateFormat='M/D/YYYY';
8 SET TimestampFormat='M/D/YYYY h:mm:ss[.fff] TT';
9 SET MonthNames='Jan;Feb;Mar;Apr;May;Jun;Jul;Aug;Sep;Oct;Nov;Dec';
10 SET DayNames='Mon;Tue;Wed;Thu;Fri;Sat;Sun';
11
12 ODBC CONNECT TO Test (XUserId is WSMetZFMbE, XPassword is KOWJCIFMTbcOXXJOELMETeB);
13
14 OrderFactTable:
15 DIRECT SELECT
16     CarrierTrackingNumber
17     , ProductID
18 FROM AdventureWorks.Sales.SalesOrderDetail ;
19
20 ProductDescription:
21 LOAD
22     [ProductID]
23     , [Name];
24 SQL SELECT
25     [ProductID]
26     , [Name]
27 FROM [AdventureWorks].[Production].[Product];
28

```

Figure 2. QlikView Load Script Syntax for Direct Discovery

In the example above, the source data table “SalesOrderDetail” has 4 fields; CarrierTrackingNumber, ProductID, OrderQty and Price. With the use of “DIRECT SELECT” keyword, only column CarrierTrackingNumber and ProductID are loaded in memory as symbol tables (they are referred as explicit direct discovery fields). These two columns are loaded in memory to create the associations between the in-memory data and the direct discovery data. These fields can also be used for selection purposes on the QlikView user interface. Other columns, OrderQty and Price, exist in the source data table within the database and they are not part of the in memory data model. OrderQty and Price fields are referred as “IMPLICIT” fields. An implicit field is a field that QlikView is aware of on a “meta level”. The actual data of an implicit field resides only in the database but the field may be used in QlikView expressions. More information on the implicit fields is provided in the next section. Please note that preceding load cannot be used with Direct Discovery as it only relates to the data that is loaded in memory.

Looking at the table view and data model, the implicit direct discovery fields are not within the model (DirectTable is the table created on the data model with the explicit direct discovery fields. This table allows QlikView to query the source database table as users make selections on the user interface):

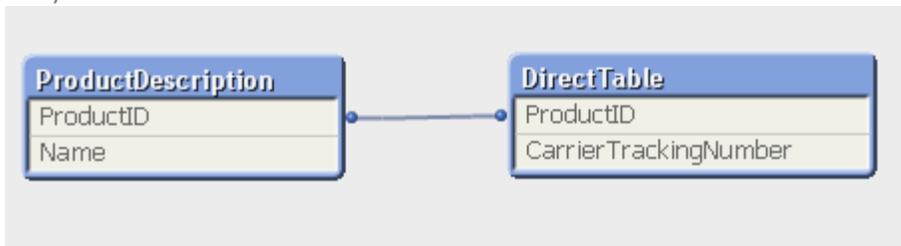


Figure 3. Associative Data Model for Direct Discovery and In-memory data sets

Once the direct discovery structure is established, the direct discovery data can be joined with the in-memory data with the common field names (Figure 3). In this example, “ProductDescription” table is loaded in memory and joined to the direct discovery data with the ProductID field. This allows the user to associatively navigate both on the direct discovery and in memory data sets. As figure 4 demonstrates, the business users can associatively make selections on either of the data sets (direct discovery or in-memory), and see what is associated and not associated with the same QlikView association colors; green, gray, and white.

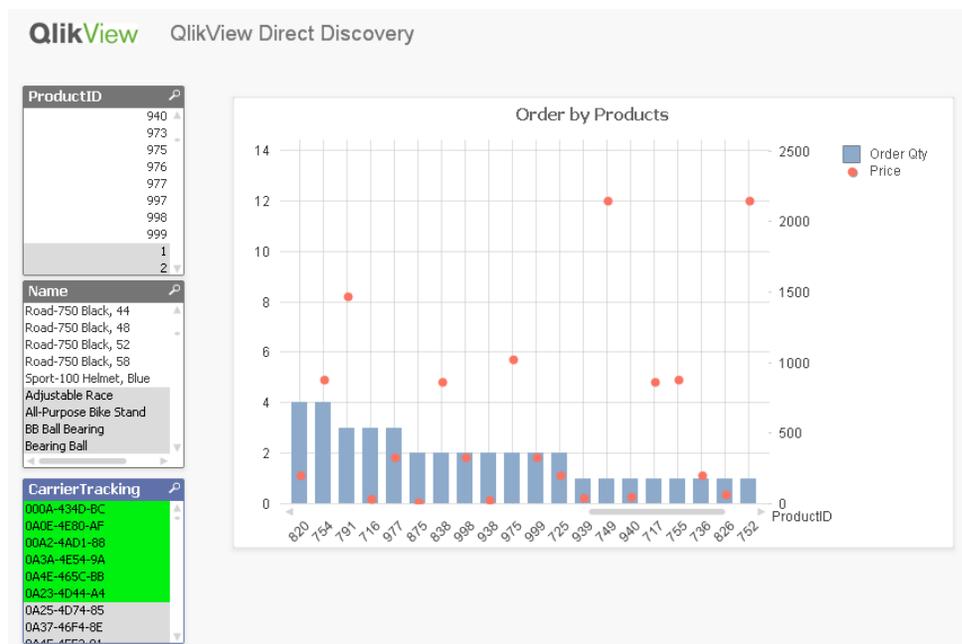


Figure 4. Associative Business Discovery on direct discovery and in-memory data sets

A decorative graphic in the top left corner of the page, consisting of several overlapping, wavy, light green lines that create a sense of motion and depth.

Background and definition of implicit fields

An implicit field is a field that QlikView is aware of on a “meta level”. When QlikView recognizes the keyword “DIRECT SELECT” on the load script, it only automatically loads the metadata of the fields from the source table (even though the fields are not listed on the load script), and makes them “IMPLICIT” fields. The actual data of an implicit field resides only in the database but the field may be used in QlikView expressions. The idea is that an implicit field will be treated as any other field when the user works with expressions in QlikView.

The reason to introduce the concept of implicit fields is for QlikView to decide if an aggregation should be done on the database instead of being processed by QlikView itself.

For example, if the expression Count(Address) is used in a QlikView chart and Address is an implicit field, QlikView will let the database do the count (obviously since QlikView doesn’t have any data for the Address field). To put it in another way; implicit fields will add support for aggregating in the database while using the same expression syntax QlikView already has. The result is that the usage of the direct discovery fields from the database will be transparent for the user.

For more advanced usages or database specific tasks, the user may utilize the new SQL(“) function (Figure 5) in the QlikView chart expressions. This new function would allow the user to run SQL syntax directly from QlikView expression definition. Please note that the SQL statements can only be created against the database table that is defined as Direct Discovery table on the load script. When SQL(“) is used in the QlikView chart expression, the expression totals will not be displayed and the use of snowflake dimensions on the chart is not supported.

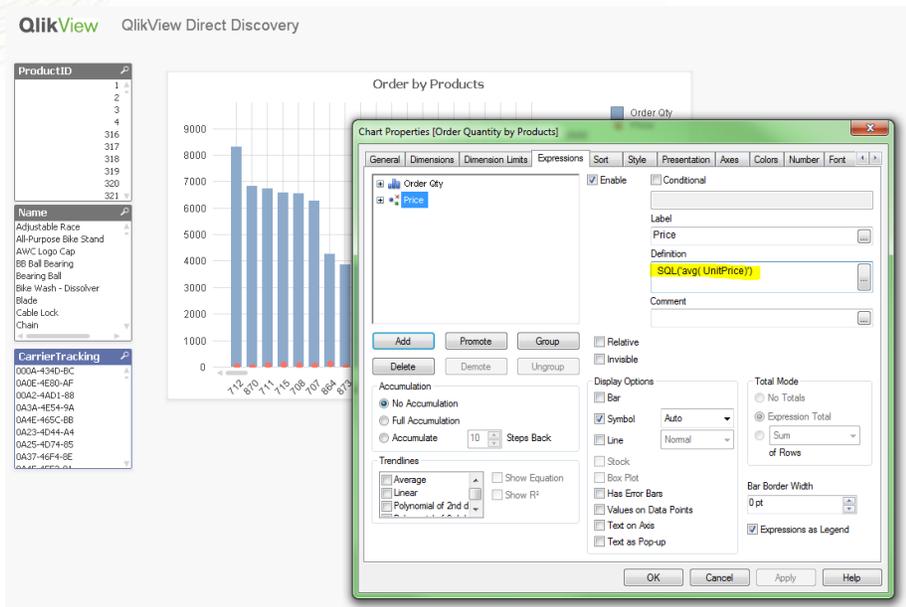


Figure 5. The use of SQL() function in the chart expression – SQL('avg(UnitPrice)')

How to define implicit fields

There are two ways of defining implicit fields. The most straightforward way is to only load the fields which will be used for selections purposes or to join the direct discovery data to other in memory data. All other fields of the database table will be traded as IMPLICIT fields. On direct discovery table script load, QlikView will fetch all of the other fields from the database table and store the knowledge of them as implicit fields.

The second way is to use the keyword "IMPLICIT" on the load script. Once this keyword is used, QlikView will only load the fields listed as IMPLICIT in addition to the fields listed after the keyword "DIRECT". Only these fields from the database will be available on the QlikView application user interface for Business Discovery. This scenario can be used by the developer in the cases where there are fields on the database table that should be hidden from the user.

Please see examples below for these two scenarios.

Examples

Database fields: Field1, Field2, Field3, Field4, Field5

- **DIRECT SELECT Field1, Field3 FROM TableName**

A decorative graphic in the top left corner of the page, consisting of several overlapping, wavy, light green lines that create a sense of motion and depth.

Field2, Field4 and Field5 will be known as implicit fields to QlikView

- **DIRECT SELECT Field1, Field3 IMPLICIT Field2, Field4 FROM TableName**
Only Field2 and Field4 will be known as implicit fields to QlikView

It is possible to rename the IMPLICIT fields on the load script. This would require using “IMPLICIT” keyword.

DIRECT SELECT Field1, Field3 IMPLICIT Field2 AS FieldB FROM TableName

Field2 is renamed to FieldB and only FieldB will be known as implicit fields to QlikView

III. CREATING QLIKVIEW OBJECTS WITH DIRECT DISCOVERY FIELDS

Once the direct discovery structure is setup, the direct discovery fields from the source table will be available on the user interface. Please note that only fields specified after the DIRECT SELECT statement (explicit fields) will be loaded into QV in the traditional manner. These fields can be used for selections or to enable associations between in memory data fields and direct discovery fields. All of the other fields from the data source will be available as implicit fields.

Due to the interactive and SQL specific nature of QlikView Direct Discovery, only certain QlikView objects can use implicit direct discovery fields. Table box, statistics box, pivot and mini charts are not supported with implicit direct discovery fields.

On the field list, additional information is provided to notify the user that the field is an implicit field (figure 5). This is useful information as some of the implicit fields from the big data sources may have billion values and using these fields on the user interface may slow down the user experience. Detailed attention should be paid with the use of implicit fields. Although the data values for the implicit fields are not loaded in memory, they still consume memory and CPU once they are used on the user interface of a QlikView application.

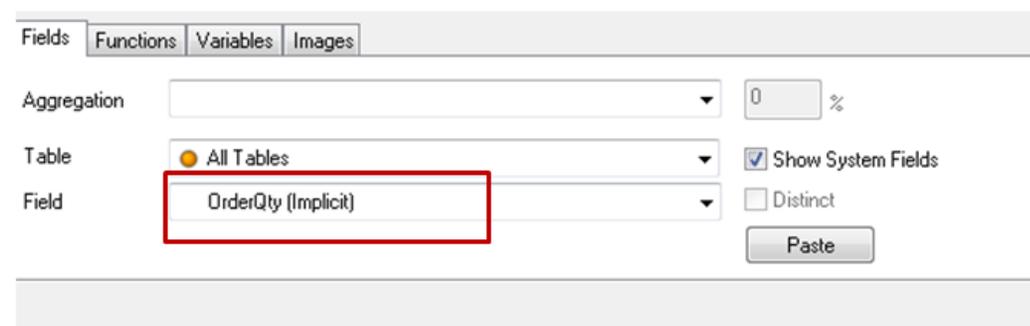


Figure 5. The direct discovery fields are marked with the keyword “Implicit” on the field list.

How to create list boxes with direct discovery fields

The traditional way of creating list box is used to create list boxes with the explicit direct discovery fields (fields that are listed after “DIRECT SELECT” keyword on the load script).

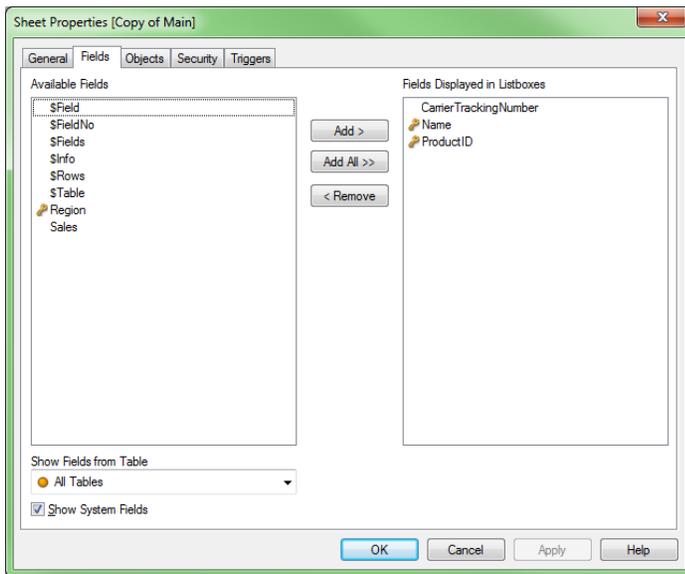


Figure 6. Traditional way of creating list boxes

In the previous example, CarrierTrackingNumber and ProductID fields can be used in list boxes as these are the explicit direct discovery fields.

Please note that to achieve the associative navigation capability; there will be cases where QlikView will execute SQL queries on the direct discovery data source when there is a selection made on the list box using a direct discovery field.

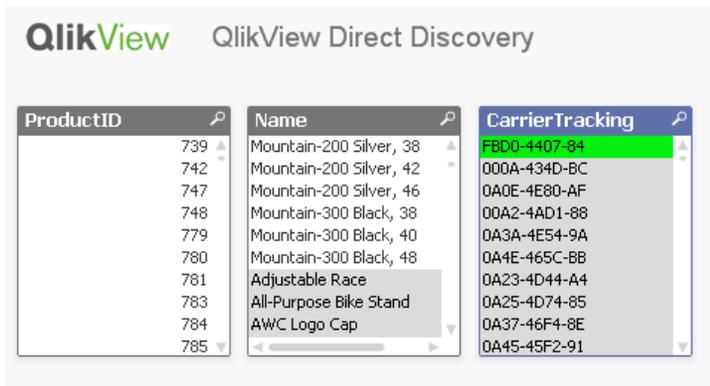


Figure 7. Associative data discovery with list boxes using direct discovery and in-memory data fields

It is possible to use the expression option of the list box with an implicit direct discovery field. The `aggr()` function should be used to show the aggregated value of the implicit field for any dimension field. For example, the list box expression “`aggr(sum(OrderQty),Name)`” will list the aggregated OrderQty values for product names in a list box (OrderQty is an implicit direct discovery field).

How to create QlikView charts with direct discovery fields

The traditional way of creating charts is used to create a chart using direct discovery fields. It is possible to use the direct discovery fields as dimensions and/or expressions on the charts. However, please note that only implicit direct discovery fields can be used in expressions and explicit direct discovery fields can be used as dimensions. In the previous example, only CarrierTrackingNumber, ProductID fields can be used as chart dimensions and OrderQty, Price fields can be used as chart expressions.

For the QlikView charts that use only direct discovery fields, all of the aggregations are done on the database. If the QlikView chart has fields both from the in-memory and direct discovery tables, a second level aggregation is done on the chart level for the in memory fields once the database level aggregations are done.

As QlikView Direct Discovery generates SQL as its base code, not all QlikView expression functionality will be compatible with the feature. The expression functions that are supported

with this initial release are; Sum, Avg, Count, Min, Max. Other functionality may be available pending time frames.

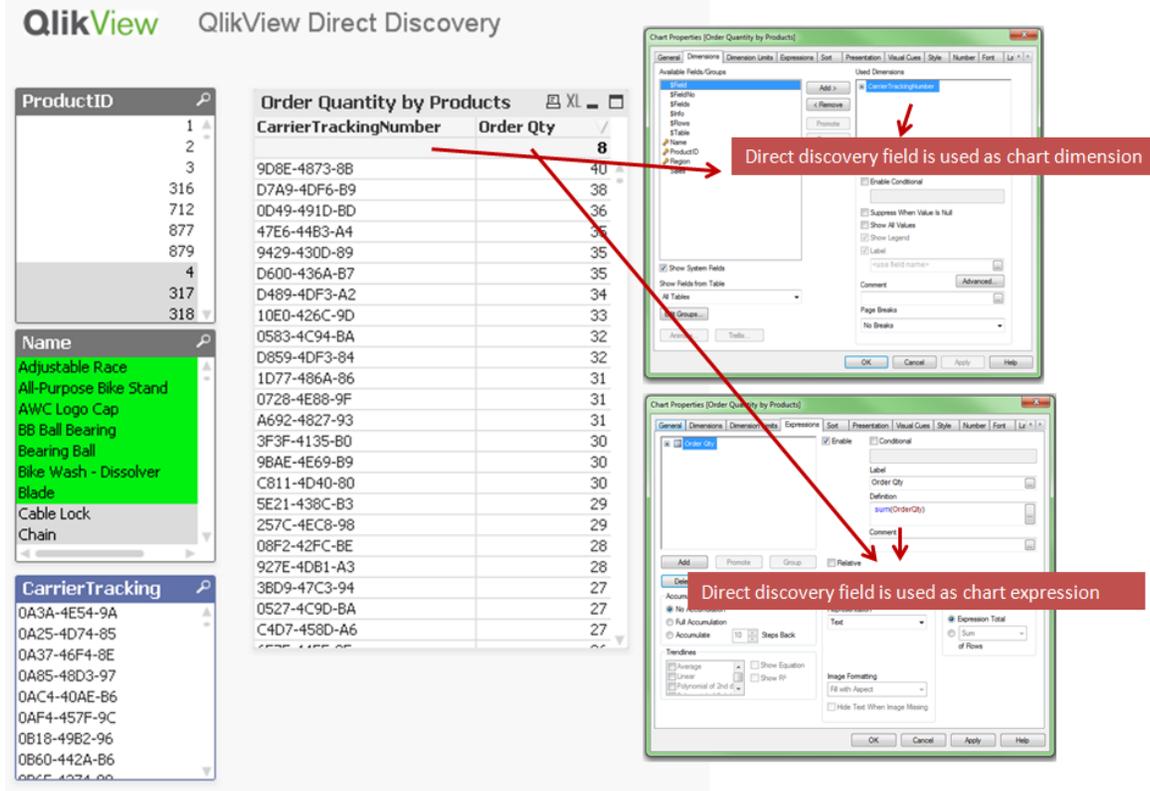


Figure 8. QlikView chart using direct discovery fields as dimension and expression

Most of the QlikView chart functionalities (interactive sorting, formatting, visual clues, dimension limits etc...) are still available to be used on the charts that use direct discovery fields. This provides the capability to leverage the same ease of use and rapid app development QlikView experience on the big data analysis.

Due to the SQL syntax specific nature of the direct discovery, pivot tables and mini charts are not supported with the use of direct discovery fields.

It is also possible to use the in memory fields and direct discovery fields on the same chart. In these cases, QlikView associative technology automatically handles the associations between the in memory fields and direct discovery fields. The business user still has the same easy to develop QlikView experience and does not need to worry about how the data sets are joined together.

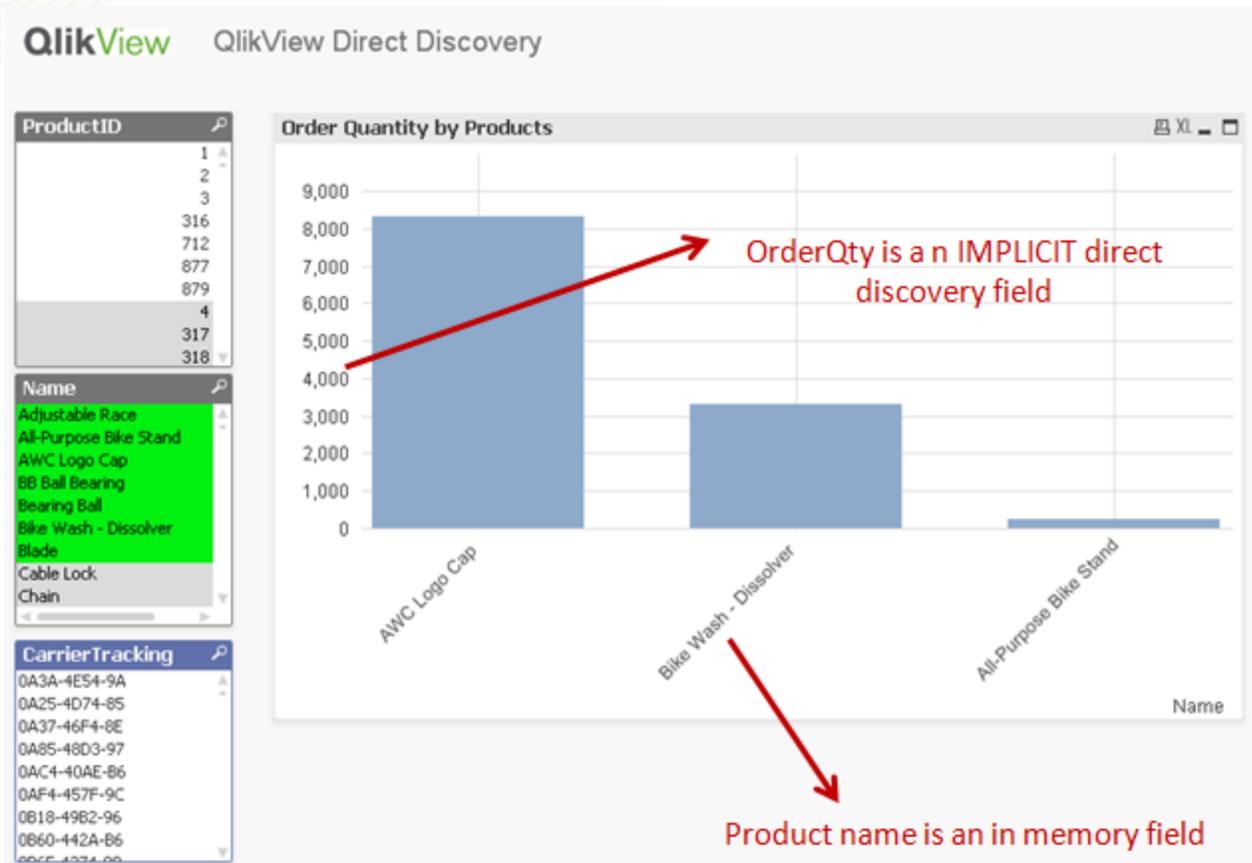


Figure 9. QlikView chart using in-memory and direct discovery fields

Please note that because of the SQL nature of direct discovery “Suppress missing” option of QlikView charts does not work and should be disabled. Once it is disabled, “Suppress when value is null” option will work as expected (Figure 10).

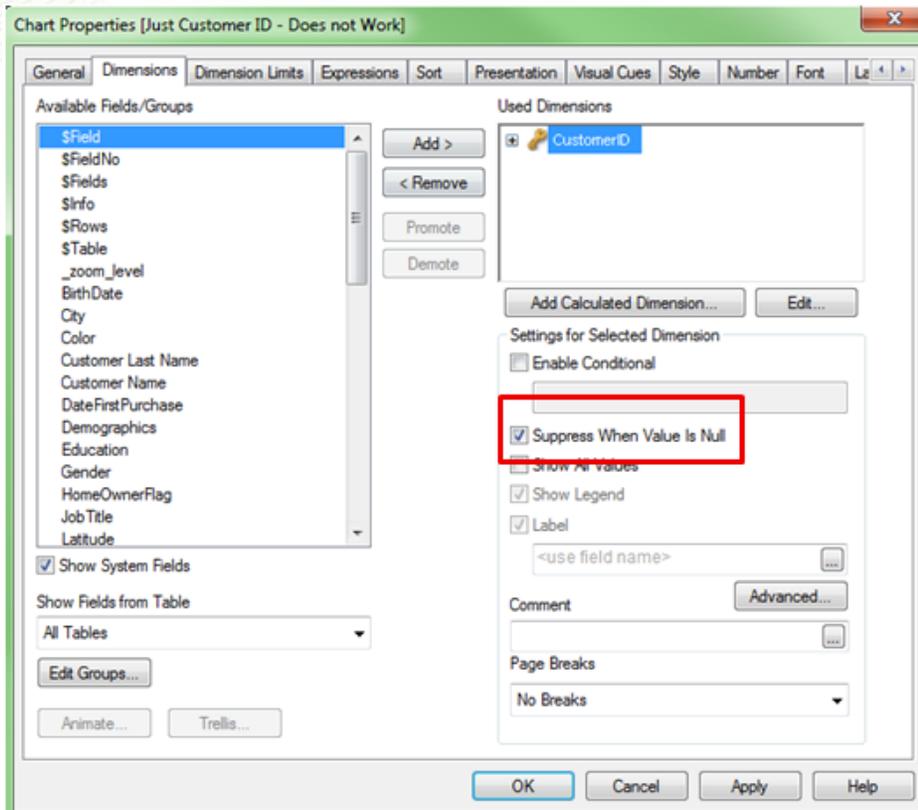


Figure 10. Suppress when value is null option

Supported data types

All data types are supported, however there may be cases where specific source data formats need to be defined to QlikView. This can be done on the load script by using the “SET Direct...Format” syntax. Below example demonstrates how to define the date format of the source database that is used as the source for Direct Discovery.

```
SET DirectDateFormat='YYYY-MM-DD';
```

Figure 11. Load Script Syntax to define database specific data formats

IV. DATA SOURCES

QlikView Direct Discovery can only be used against SQL compliant data sources. The following data sources are supported;

- ODBC/OLEDB data sources - All ODBC/OLEDB sources are supported, including SQL Server, Teradata and Oracle.
- Custom connectors which support SQL – SAP SQL Connector, Custom QVX connectors for SQL compliant data stores.

Both the 32-bit and 64-bit connections are supported with this feature.

V. CANCELLING DIRECT DISCOVERY QUERY

The cancel icon on QlikView charts can be used to abort the direct discovery query. The direct discovery query running for a chart will be automatically aborted when the QlikView application is closed or a new tab is selected.

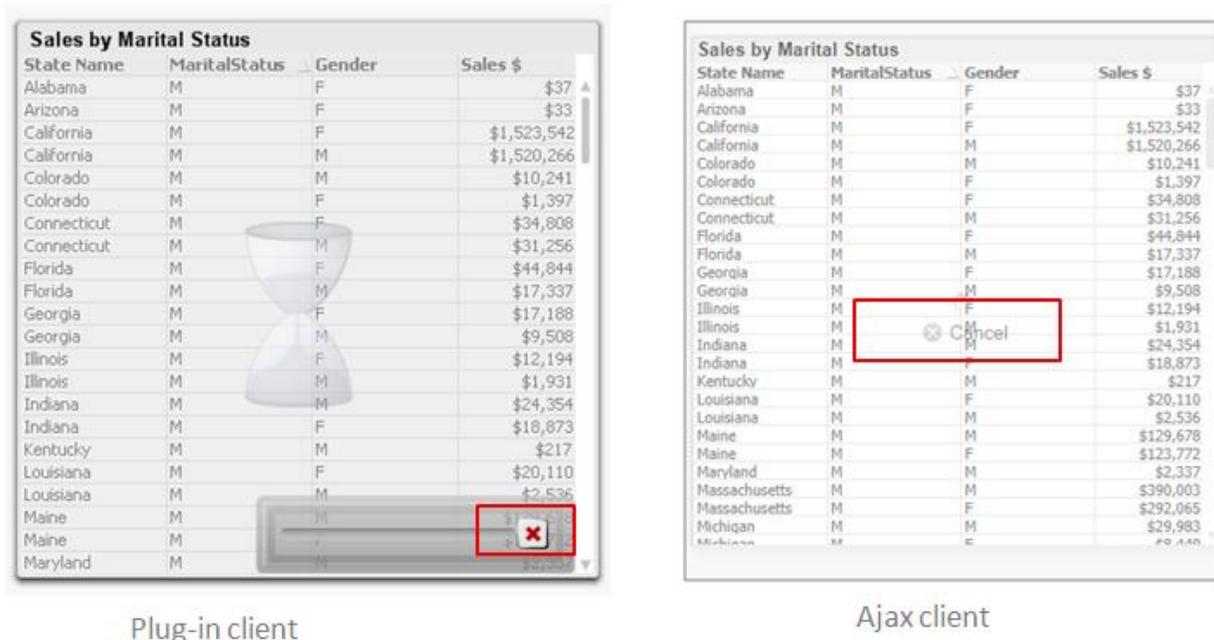


Figure 12. Cancelling direct discovery query

VI. SUPPORTED QLIKVIEW FUNCTIONALITY

Due to the interactive and SQL syntax specific nature of the Direct Discovery approaches, the following QlikView features are not supported;

- Advanced calculations
- Calculated dimensions
- Comparative Analysis (Alternate State) on the QlikView objects that use Direct Discovery fields
- Direct Discovery fields are not supported on Global Search
- Binary load from a QlikView application with Direct Discovery table
- Section access and data reduction
- Loop and Reduce
- Synthetic keys on the Direct Discovery table
- Table naming in script does not apply to the Direct table
- Document chaining
- The use of "*" after DIRECT SELECT keyword on the load script (e.g. DIRECT SELECT *)

VII. PERFORMANCE CONSIDERATIONS

The performance of the Direct Discovery feature fundamentally reflects the performance of the underlying system as the feature queries an external system from QlikView. Please note that, in-memory processing will always be faster than in-database processing. It is possible to use standard database and query tuning best practices for this feature. All of the performance tuning should be done on the source database. Direct Discovery feature does not provide any support for query performance tuning from the QlikView application. However, it is possible to do asynchronous, parallel calls to the database by using the connection pooling capability. The load script syntax to setup the pooling capability is as follows:

```
SET DirectConnectionMax=10;
```

Figure 13. Load Script Syntax to setup the pooling capability

To improve the overall user experience, QlikView's caching is used. Please see the section on caching for more information.

VIII. SECURITY

Some security best practices should be taken into considerations when using Direct Discovery feature.

- All of the users using the same QlikView application with the Direct Discovery capability will be using the same connection. With this initial release, authentication pass through or credentials per user are not supported.
- Section Access is not supported.
- With the new SQL() expression function, it would be possible to execute custom SQL statements in the database. It is advised that the database connection set up in the load script should use an account with only read access to the database.
- Please note that with the use of the new SQL() expression function, it would be possible for the users to create SQL statements that could affect the performance of the database system. If this is the case, the cancel functionality could be used to cancel the queries from the QlikView user interface.
- With this release of QlikView Direct Discovery, there is no logging capability. However it is possible to use the ODBC tracing capability. Please see the logging section for more details on this.
- It is possible to flood the database with requests from the client.
- It is possible to get detailed error messages from the QlikView Server log files.

IX. QLIKVIEW SERVER SETTINGS

Some settings on the QlikView Server and on the config.xml file should be reviewed if the Direct Discovery capability is used on a QlikView application. Please note that once these settings are changed, it will affect all of the QlikView applications that are on the same QlikView Server.

Object Calculation Time Limit

As Direct Discovery feature queries an external system from QlikView, the chart calculation time is dependent on the performance of the underlying system. It is advised to set the object calculation time limit setting on the QlikView Management Console high enough to allow enough time for the QlikView chart to get the direct discovery query results back from the data source. This setting is located under Performance tab of QlikView Server listed on the QlikView Management Console (Figure 14).

Max Symbols in Charts

Max symbols in charts setting is used to set the number of data points to be displayed on QlikView charts. Please note that as Direct Discovery query can return many distinct values, it is advised to review this setting to allow QlikView to display the desired number of data points on charts (Figure 14).

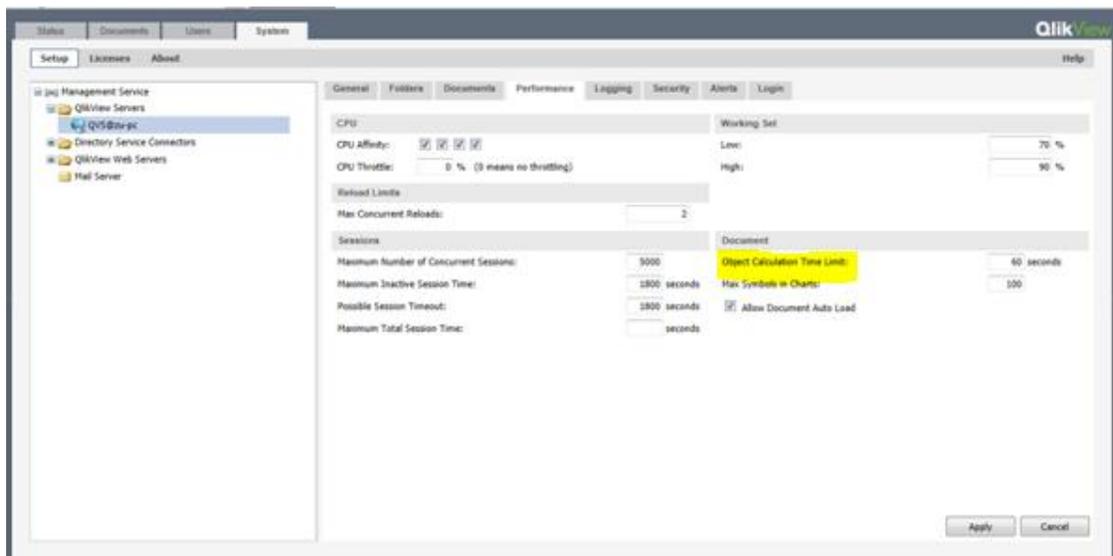


Figure 14. Object calculation time limit and Max symbols in charts settings on QlikView Management Console

QVS Time Out Setting on the Config.xml File

As Direct Discovery feature queries an external system from QlikView, the QlikView Server time out setting on the config.xml file could be adjusted to allow enough time to QlikView to get the query results back. This change should be made if “Lost connection to server” error is seen when using the Ajax client.

The default setting on this option is 60 seconds. It is advised to increment this value to the possible maximum query time. Config.xml file is located under C:\ProgramData\QlikTech\WebServer folder. Please note that during upgrades this file will be overwritten with the default value.

```
<QvsTimeout>60</QvsTimeout>
```

Figure 15. QlikView Server time out setting on config.xml file

X. CACHING

The performance of querying an external system from QlikView fundamentally reflects the performance of the underlying system. To improve the overall user experience, QlikView’s caching is used and stores selection states of queries in memory. As the same types of selections are made, QlikView will leverage the query from the cache (Figure 17). These cached result sets are shared across users. Please note that, when the same types of selections are done, QlikView will not query the source data and will leverage the cached result set. Also, when back and forward buttons are used, QlikView will leverage the query results from the cache.

It is possible to set a time limit on caching. A special syntax, “STALE” can be used on the load script to set a caching limit on the Direct Discovery query results (Figure 16). Once this time limit hits, QlikView Server will clear the cache for the Direct Discovery query results that were generated for the previous selections. When this happens, QlikView will query the source data for the selections and will create the cache again for the designated time limit.

```

Edit Script [QlikView Direct Discovery]
File Edit Insert Tab Tools Help
Reload Debug
Main
2 SET DecimalSep='.';
3 SET MoneyThousandSep=',';
4 SET MoneyDecimalSep='.';
5 SET MoneyFormat='$#,##0.00;($#,##0.00)';
6 SET TimeFormat='h:mm:ss TT';
7 SET DateFormat='M/D/YYYY';
8 SET TimestampFormat='M/D/YYYY h:mm:ss[.fff] TT';
9 SET MonthNames='Jan;Feb;Mar;Apr;May;Jun;Jul;Aug;Sep;Oct;Nov;Dec';
10 SET DayNames='Mon;Tue;Wed;Thu;Fri;Sat;Sun';
11
12 ODBC CONNECT TO Test (XUserId is WSMtZFMbE, XPassword is KOWJCIbTbcOXXJOELMETeB);
13
14 OrderFactTable:
15 DIRECT (Stale after 5 seconds) SELECT
16     CarrierTrackingNumber
17     ,ProductID
18 FROM AdventureWorks.Sales.SalesOrderDetail ;
19
20 ProductDescription:
21 LOAD
22     [ProductID]
23     ,[Name];
24 SQL SELECT
25     [ProductID]
26     ,[Name]
27 FROM [AdventureWorks].[Production].[Product];
28
  
```

Figure 16. “Stale” option on the load script to clear the direct discovery table cache

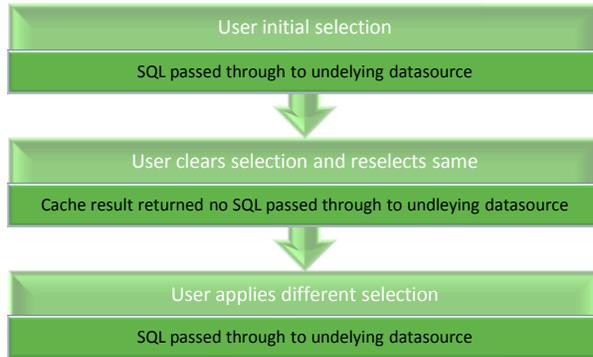


Figure 17. QlikView caching process for Direct Discovery data

XI. LOGGING

Behind the scenes a direct SQL statement is passed to the underlying data source as such the statement passed can be viewed through the trace files of the underlying connection. An example of this is shown below for a standard ODBC connection:

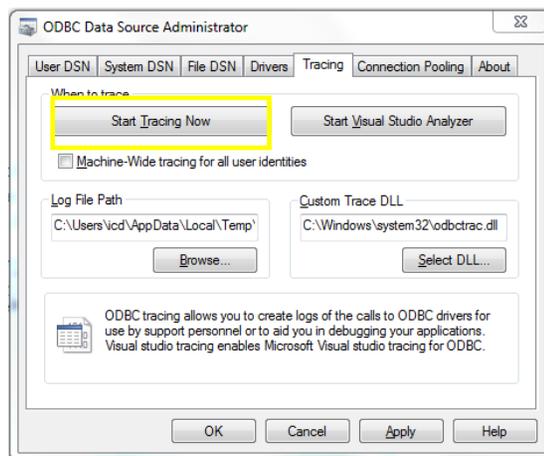


Figure 18. Standard ODBC connection tracing tab

The subsequent trace file details SQL statements generated through the user selections and interactions.

XII. ERROR MESSAGES

Because of the SQL syntax nature of Direct Discovery some errors may happen on the QlikView charts when direct discovery fields are used. Below is a brief description of these errors.

- Direct query attempted against missing or non-direct table
This error message is displayed on the chart when the load script has been changed so that a previously defined direct query is now against a table that is no longer there or against an in memory table.
- Mixed DBMS and in-memory aggregation not supported
This error message is displayed when the new SQL(“”) expression function is used with snowflake dimension fields.
- Direct query failed
This error message is displayed when the query that is executed for direct discovery is failed.

XIII. KNOWN ISSUES

Please refer to QlikView 11.2 release notes document.

XIV. ADDITIONAL RESOURCES

QlikView Direct Discovery Data Sheet

<http://www.qlikview.com/us/explore/resources/brochures-datasheets?language=english>

QlikView Direct Discovery FAQ

<http://www.qlikview.com/us/explore/resources/brochures-datasheets?language=english>

QlikView and Big Data

<http://www.qlikview.com/us/explore/products/big-data>