Connecting Spotfire® to Data Sources

Harald Dauwa-Stummer, Manager Analytics Consulting EMEAI
November 3, 2016

HUMAN HEALTH • ENVIRONMENTAL HEALTH
© 2014 PerkinElmer
Content

• How does Spotfire consume data / Enterprise data access

• Data access using Spotfire Information Designer
How does Spotfire consume data?

In-Memory
Load data from source in to memory

In-Database
Leave data in DB Dynamically load and discard data to visualize

On-Demand
Dynamically swap data in and out of memory.

DATA SOURCES
- RDBMS
- Spreadsheets
- Flat Files
- XML
- Hadoop & Big Data stores
- Analytical DWs e.g. Exadata
- Cubes
- Event Data Streams
- Active Spaces
Enterprise Data Access

Databases
- SQL Server
- Oracle
- MySQL
- Teradata
- Netezza
- JDBC/ODBC
- Hadoop
- SFDC
- PostgreSQL
- Etc.

Local data sources
- Excel
- Access
- STDF

Drag-and-drop

Information Services
(join, transform, reusable, parameterized, dynamic query for in-memory use)

Direct Query
(dynamically query and retrieve data for visualization and analysis)

Databases
- Teradata
- Oracle
- MySQL
- Hadoop
- MS SSAS
- TeradataAster
- Netezza
- Etc.

Custom GUI-driven data access via SDK

Direct connection
- ODBC
- OLE DB
- SqlClient

RDBMS
- SAP R/3
- SAP BW
- Oracle E-Business
- Siebel eBusiness
- OBIEE

Web Services
- Salesforce

XML
- Web Services

Flat Files
- XML

Spread-sheets
- Flat Files

JDBC/ODBC
- Databases

Hadoop
- Local data sources

OBIEE
- Information Services

OBIEE
- Information Services

Databases
- Etc.

SAP E-Business
- Information Services

Excel
- Local data sources

Access
- Local data sources

STDF
- Local data sources
Spotfire® Information Designer

Information Links are created within the Information Designer.

Information Designer is the tool used to manage access and connections to JDBC-accessible data sources.

- Information Links are methods to communicate with a database and retrieve a specific set of data, which is then loaded directly into Spotfire.

- Information Links can include limiting options, aggregation methods, data transformations and filtering options to help control the size and specificity of the data set.

- Filters can be static or prompted
Requirements

- License options must be granted
- To set up a data source, the database must accept JDBC connections.
- A good knowledge of the underlying database structure is required
Opening Information Designer
The Interface
Setting up a data source

Data sources are the connections to the database. They are used to create the elements later used as columns, filters, procedures and joins.
Creating Data Elements

- Columns allow you to apply aggregations, filtering and grouping
- Multiple columns can be mapped quickly but do not allow the Column options

<table>
<thead>
<tr>
<th>Aggregation Method</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIN(%1)</td>
<td>Smallest figure in %1</td>
</tr>
<tr>
<td>COUNT(%1)</td>
<td>Number of rows in %1</td>
</tr>
<tr>
<td>CORR(%1, %2)</td>
<td>Coefficient of Correlation of %1 and %2</td>
</tr>
<tr>
<td>COVAR_POP(%1)</td>
<td>Sample Covariance of %1 and %2</td>
</tr>
<tr>
<td>STDDEV(%1)</td>
<td>Sample Standard Deviation of %1</td>
</tr>
<tr>
<td>VAR.Pop(%1)</td>
<td>Population Variance in %1</td>
</tr>
<tr>
<td>VAR_SAMP(%1)</td>
<td>Sample Variance in %1</td>
</tr>
<tr>
<td>VARIANCE(%1)</td>
<td>Variance of %1</td>
</tr>
</tbody>
</table>
Filters allow removal of data prior to loading based on the expression set. Remember to insert the condition in the expression field using the given column alias.
Procedures

• Using Information Designer you select pre-made database procedures and configure these to be accessible in Information Links. These Information Links are available to the Spotfire users to retrieve or manipulate data.

• Parameters are variable values that must be set or passed to the Information Link when loaded. These values can change but are required when loading data. These can be defined in the Expression’s fields.

• Parameters are defined following the next syntax: “?ParameterName”. Later, they will appear in the Parameters section after clicking Refresh.
Procedures can accept input parameters and used to create Information Links

- Select Procedure of Type Query
- Provide the default value for the input parameter, either hardcoded or as a parameter indicated by “?”
Tip: Using the Procedure of Type “Query”

• There is a big advantage in using the ‘Procedure’ from Spotfire Information Designer for creating Information Links to load data into Spotfire Data Table. The Procedure of Type Query returns the data set with the columns specified in Stored Procedure (in SQL server database) / Stored Function (in Oracle database) code into the data table in a Spotfire analysis.

• If there are any changes needed to the table columns, joins..etc, the procedure / function code changes can be modified in the database. With the usage of Procedure element, the new columns and modified old columns in the dataset will be automatically loaded into the Spotfire data table via the Information Links. This approach will eliminate explicit changes to be made to Spotfire Information Model elements and provide easy maintenance of Spotfire IM elements.
Create Information Links from Elements

- Choose the desired Information Elements, click on “Add” and order the columns.
- Click on “Open Data” to review the data in Spotfire
Filter Types

- **Value Filters**

- **Range Filters**

- This filtering option can be set from the Information Link Tab. The difference between this Filter Option and the Filter Element is that the latter can be used regardless of the columns mapped in the Information Link, whereas the Filter Option must be set using an available column from the Elements box of the Information Link Tab.
Prompts

A prompt is a dialog window where the user can choose to limit the amount of data to be loaded from the Information Link and can be changed each time the data loads.

- Values
- Range
- Single Selection
- Multi Selection
SQL Button – allows you to manually edit the SQL

When using the Information Designer to create an Information Link, a SQL is being written according to the settings selected.
Sometimes when creating Information Links, columns from more than one table must be used. In those cases to create a useful Information Link, a join between the tables is needed. The join relates the tables by matching data in columns that identify rows uniquely, e.g. primary key columns.
Open and load Information from a Spotfire Session
Data loaded
Common issues and troubleshooting on Information Designer

- Use right mouse “Validate” on Information Elements. The most likely cause of the “Validation unsuccessful” Information Link message is that the Data Source is not accessible (wrong data source, wrong credentials, database down, ...)
- Invalid SQL: Test your SQL in the source database first to make sure it is valid (e.g. with database vendor tools like SQL+)
- Join Columns issue: Source columns used have been modified or dropped in the source database

For more information about executed SQL queries performed when an Information Link is executed check the SQL Log file on the Spotfire server
Questions? Thank you for your time!

Harald Dauwa-Stummer
Manager Analytics Consulting EMEAI
harald.dauwa-stummer@perkinelmer.com;
InformaticsAnalyticsServices@perkinelmer.com

+43 664 8322305