NobleProg

Advaced Analytics with TIBCO Spotfire

Information Links and Data Connectors

Bringing Data in Spotfire

- Direct file access
- Data connections via connectors
- Information Links

NobleProg

In-memory vs in-database analytics

- When data is loaded into the in-memory data engine, an entire copy is loaded in your computer/server.
- Spotfire's in-memory data engine is highly optimized for working with datasets of all sizes, but eventually you do run out of memory.
- This engine compresses data, removes redundant columns and swaps data to disk if needed.

In-memory vs in-database analytics (II)

- Loading all data in-memory might not be feasible, nor needed.
- Spotfire can instead query the database, extract the data it needs and only keep the aggregated/transformed data required for the chart.
- To handle this, Spotfire uses **data-on-demand,** so that data is retrieved only when something changes.

Information Links

- Information Links are Spotfire's database abstraction method. They specify references to:
  - Columns
  - Join elements
  - Filter elements
- Support for joins between multiple tables (across databases), stored procedures, custom SQL and write back to DB.
- They support only **in-memory** data.

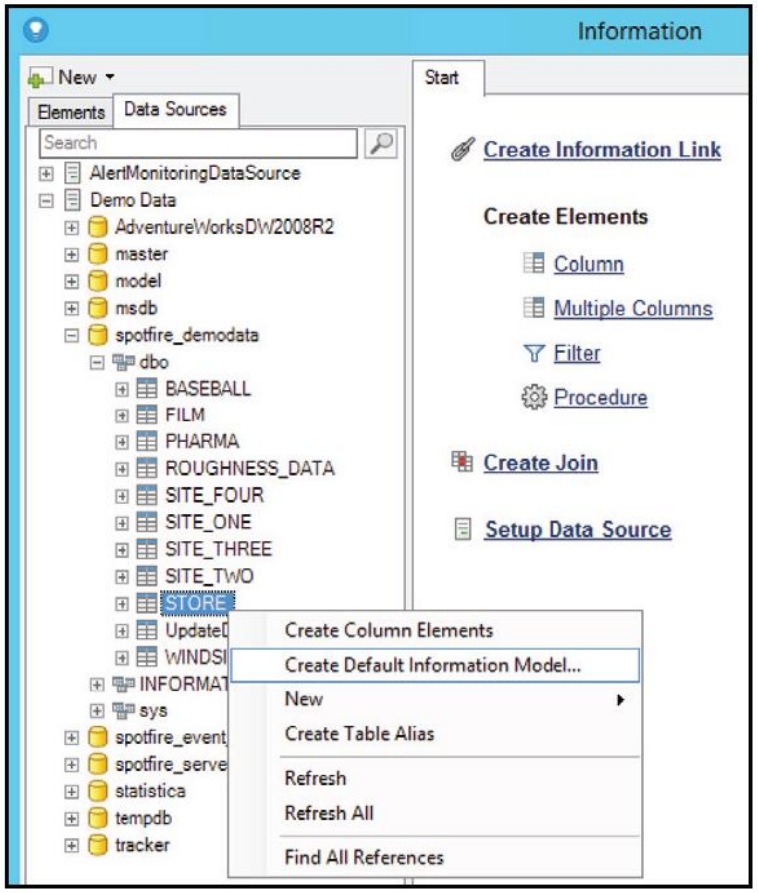
NobleProg

Information Link vs Data Connectors

- With Information Links the data will be in memory, not external.

- Connecting to an external data source (e.g. Teradata) would require that all clients have the driver. With information links, you only need the JDBC.

To create an Information Link



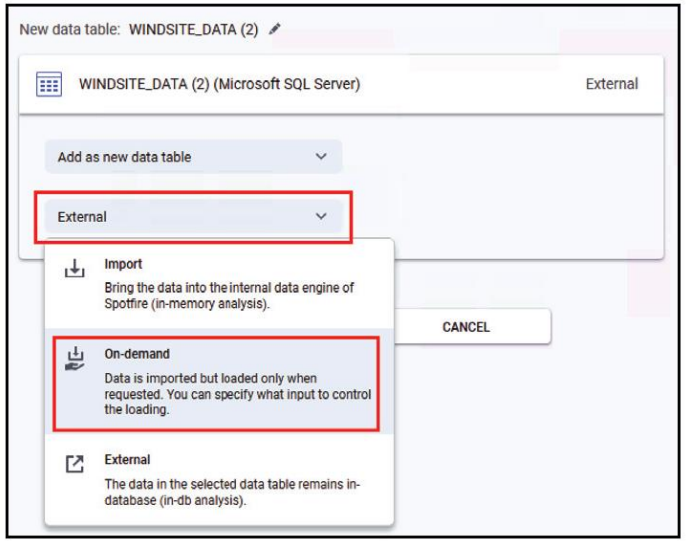
- Connect to a data source.
- Create a default information model (see screenshot)
- To define or manage information links, the Information Designer can be found under the **Data** menu.

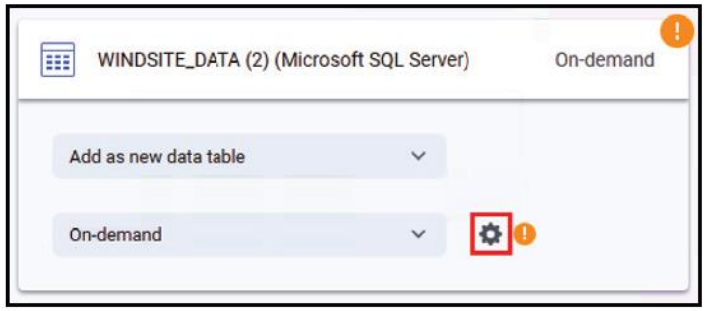
➔ Create Connector

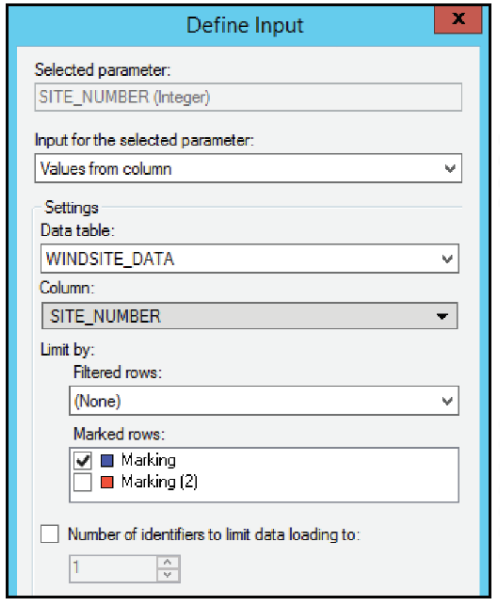
NobleProg

Data On-Demand

- Data snapshots will be reloaded as needed.
- While setting the connection, you can choose data to be on-demand.







NobleProg

Custom Expressions

- Custom Expressions can be used, but there are limitations depending on the connector.

- In general, we can perform custom expressions within visualizations, not on calculated columns.

Important points when working with external data

- Filters are not created by default.
- Box plots cannot be used on external data.
- Spotfire‘s Transformations cannot be used on external data.
- Only the data source‘s functions for calculated column expressions can be used.

NobleProg

Caching in Spotfire

NobleProg

Why caching?

- Caching is storing data in RAM or hard disk for performance.

- It helps to
  - Avoid hitting database every time anyone opens the analysis.
  - Eliminate wait time.

- Common use case:
  - When many dashboards require data from the same Information Link.

How it works?

- Individual Information Links are cached at the Spotfire Server level.
- The cache is a Spotfire-specific Binary file stored on the hard drive of Spotfire Server.
- The cache is self-cleaning depending on timeouts and validation query. A server restart will also clear it up.

Creating cache

1. **First user takes the hit**.
2. **Automation Services and fixed schedules**: schedule a job that opens a file (which will trigger the Information Link).
3. **Automation Services and triggers**: Run a bat file programatically or call Automation Services webservices to wait for instance for an ETL process to finish.
4. **Scheduled Updates.** In this case cache is retained in RAM.

NobleProg

Considerations

- Enough disk space should be available.

- To use the benefits of caching, at least one report should have the Information Link completely loaded. Subsequent reports using the cache should be triggered with enough time in between.

Considerations II

- Caching does not exempt you from good practices like data on-demand and in-database analytics. Spotfire allows for building great hybrid solutions, but does not do miracles.

- Caching 15 queries at the same time may not be as fast as 3x5. This is environment-specific, so one should try different things.

NobleProg

Considerations III

- Scheduled Updates will reduce load time in the Web Player, but not in the Analyst Client.

- Saving an analysis as embedded (dxp) or data (sbdf) may not be an option if this is larger than 2GB and the application database is a Microsoft SQL Server.

NobleProg