

www.sqlbi.com




a brand of


Microsoft Partner
Gold Business Intelligence
Gold Data Platform



Marco Russo
marco@sqlbi.com
twitter @marcorus

Power Query in Modern Corporate BI



Who's Speaking?

- BI Expert and Consultant
- Founder of www.sqlbi.com
 - Problem Solving
 - Complex Project Assistance
 - Data Warehouse Assessments and Development
 - Courses, Trainings and Workshops
- Book Writer
- Microsoft Business Intelligence Partner



a brand of





What is Power Query




What is Power Query

- Excel Add-In for Excel 2010 and Excel 2013
 - Free download
- Features
 - Data Discovery
 - Data Transformation
- Hidden features
 - Language "M"



Power Query Elements

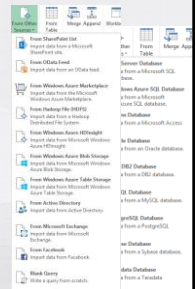
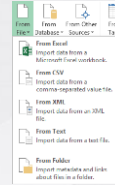
- Connectors
- Transformations (UI)
- The "M" Language
- Sharing Queries and Functions



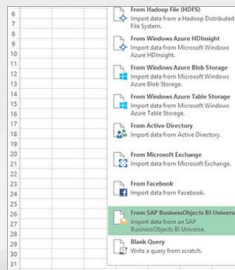
Connectors



Connectors



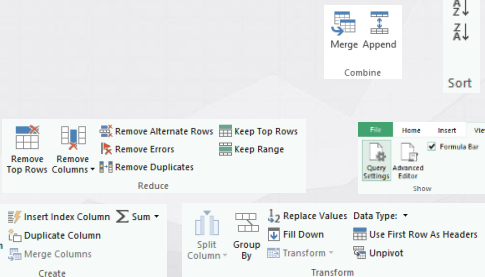
SAP BusinessObjects BI Universe



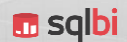
Transformations



Transformations



The "M" Language



M

- Functional language similar to F#
- No “real” iterative statements
 - Lambda functions...
- Flexible, can be invoked dynamically
 - Expression.Evaluate(...)



Query Folding

- Expression evaluation can be pushed to data source
 - Only certain data sources
 - e.g. SQL Server, OData
 - Only certain operations
- You can stop that:
 - Using Table.Buffer()
 - Combining different data sources
 - Writing explicit SQL queries



Sharing Queries & Functions



Sharing Queries and Functions

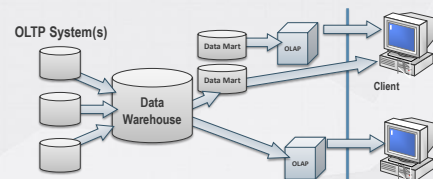
- You can share queries and functions
 - Requires Office 365 account with Power BI
- Published queries can be “certified”
 - Requires Data Steward role
- Accessible through “Online Search”
 - Query is copied and executed locally
 - No execution server-side



Corporate BI Architecture



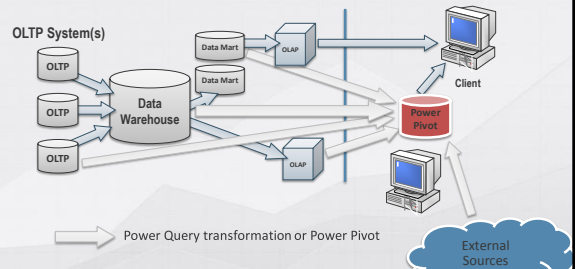
The Classical BI Solution Architecture



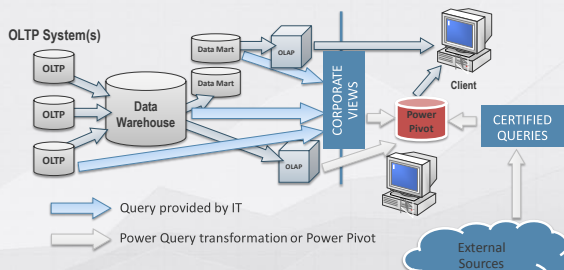
The Self-Service BI Force



What might happen



What should happen



Support for Self-Service BI

Provide Data to Self-Service BI

- Use Views
 - Do not expose table directly
 - Every view define a contract with a consumer
 - Keep views valid over time – do not break existing reports and extraction
- Use Natural (Business) Keys
 - Surrogate Keys are useless to integrate data from different sources
- Expose both Data Warehouse and Data Marts
 - Data Marts are usually report-oriented
 - Data Warehouse might offer better data for Self-Service BI (similar to a private Data Mart)
 - OLTP only for real-time data (limit history available)

Consider Self-Service ETL

- Simple transformations can be “self-service”
 - No longer necessary to apply whatever transformation in DWH/DM
 - Combine columns, concatenate names, denormalization of lookup tables
 - Aesthetic transformations easily done with tools like Power Query
- Complex transformations should be centralized
 - Business logic belongs to the Data Warehouse
 - Week Calendar is ready to use, no ambiguities in usage

Metadata Repository

- Simplify extraction of data from Data Warehouse
 - Navigate in available data
 - Use Views
 - Classify Sources
 - Control who extracts what with automatic tools
- Searching for data
 - Navigate in available data
 - Cloud services available (Azure DataMarket, datamarket.com, ...)
 - Searchable metadata repository required for data warehouse
 - Consider Power Query repository in Office 365



Optimize Data for Columnar Databases

- Control Data Normalization
 - Views should present denormalized entities
 - Columnar databases use high compression for low granularity columns
 - Relationships are much more expensive
- Do not over-denormalize
 - Having multiple entities in the same table makes it hard correlating data from other tables
 - It is important to find the right balance – star schema is still a good guideline



Power Query Usage



Power Query Role

- Self-Service ETL
 - Transform data using a client tool
 - Possible execution server-side (data refresh in Power BI)
 - Share queries and expressions (simple text file, repository available in Office 365)
- Integrate queries in Power Pivot data models
 - Same language regardless of data source
 - Graphical user interface and interactive editor
 - Does not hide the query in the data model
 - Technically it would be possible to analyze dependencies in a "M" expression (not yet implemented)



Surface exposed to Power Query

- Expose Tables
 - **PROS:** automatic detection of relationships
 - **CONS:** lack of decoupling layer in Corporate BI
 - Easier to implement, new table/columns are immediately accessible
- Expose Views
 - **PROS:** better tracking of tables/columns used – enable dependency analysis
 - **CONS:** lack of automatic detection of relationships
 - Act as a decoupling layer between Corporate BI and Self-Service BI, useful for handle data structure change over time



Data Mart vs. Data Warehouse

- Data Warehouse
 - Expose natural (business) keys
 - Raw data, useful for unplanned analysis
 - Should not be the default choice, but it could be wrong hiding it completely
- Data Mart
 - Expose surrogate keys
 - Views of single tables should
 - Expose application keys
 - Hide surrogate keys for SCD Type I
 - Show surrogate keys for SCD Type II
 - Might include report business logic
 - Should be easier to use to obtain same numbers in existing Corporate reports



Share Queries in Power Query

- Publish Queries on Power BI
 - Different visibility levels: private, group, public (within the company)
- Share and Search Easily
 - Features integrated in Power Query ribbon
- Validation Provided by Data Steward
 - Data Stewards can certify a query



Data Stewardship

Demo



Data Steward

- Responsible of queries/issues related to data
 - Identifies and acquires new data sources
 - Creates and maintains master data definitions
 - Monitor published data sources for usage/relevance/quality
 - Maintains business metadata (descriptions) for published data sources
- Connection to the IT department
 - Should be a business user – knows data and business processes



Other Corporate Data Sources



Reporting Services

- OData Feed
 - Use standard protocol
 - Can be imported directly in Power Pivot
 - Power Query **does not support** OData generated by Reporting Services (currently = May 2014)
- Security
 - Data visibility handled by the report (might have complex logic)
 - Granularity exposed according to report design
 - In Power Query it would be possible to call same report with different parameters changing URL



Analysis Services

- One Query, One Table
 - At the moment, no direct support for DAX/MDX
 - Coming soon, similar user interface as SAP BusinessObjects BI connector
 - Can be used with a query in SQL Server and a Linked Server
 - In this way, it is supported also in Power Query scheduled refresh
- Column Names
 - Complete dimension/hierarchy/level name exposed as column name in Power Pivot
 - In Tabular, table name is always present before column name
 - A query can rename all the columns in a proper way



OData Integration in Corporate BI

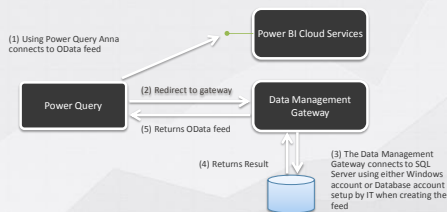


OData feeds in Power BI

- Publish OData through Data Management Gateway
 - Expose data in a controlled way
 - Simplify connection to data source in complex environments
 - Supports only SQL Server and Oracle
- Limitations
 - Currency/Money data type is exposed as decimal
 - Requires physical access to database server
 - Currently it is accessible only by Power Query
 - Excel and Power Pivot do not support Organizational Account
 - Cannot be used to refresh Excel
 - Expose data to organizational accounts
 - Does not expose relationships between tables



Corporate OData Feeds



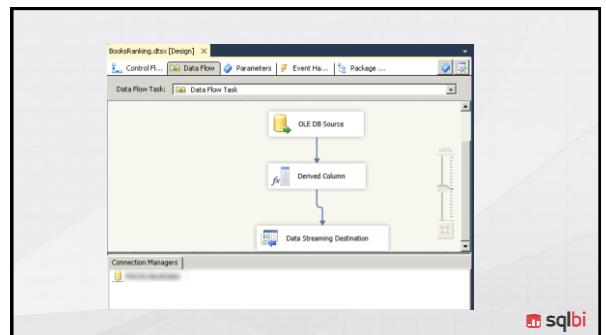
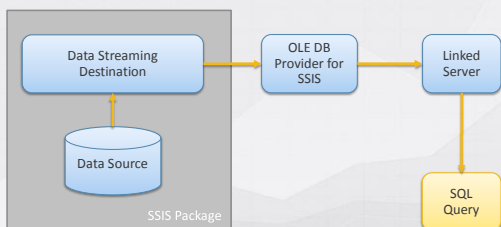
Expose OData in SSIS

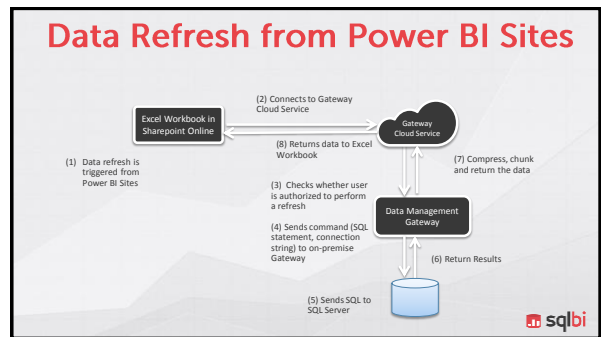
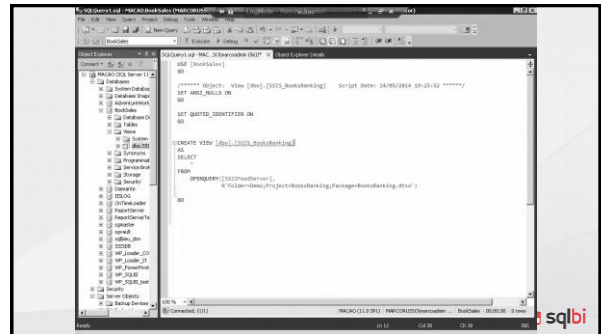
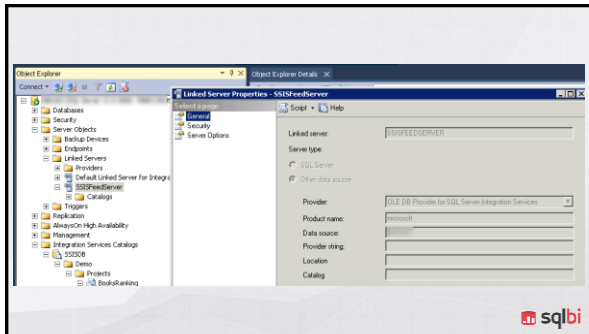
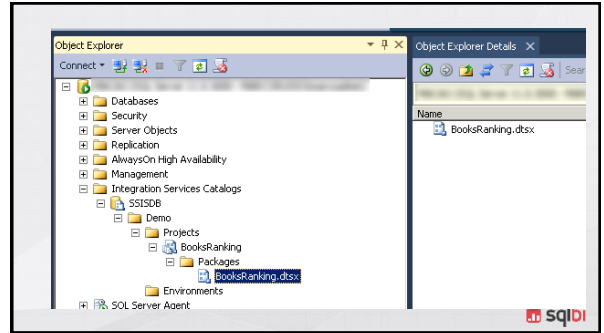
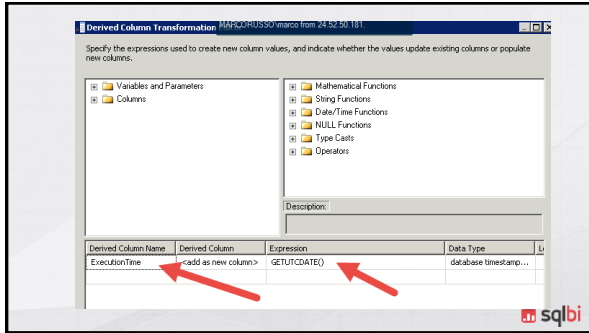
- SSIS Transformations On-Demand
 - Publish SSIS Packages as a view in SQL Server
 - Can be exposed as OData Feed Sources using Data Management Gateway (DMG)
 - Enable Power BI refresh of Data Sources not supported by SQL Server
 - Enable Power Query access to unsupported data sources
- Latency and Arguments
 - Acceptable SSIS execution latency (2-3 seconds on-prem, something more through DMG)
 - SSIS parameter are defined in the view exposed (cannot be passed at execution time)

More info:
<http://office.microsoft.com/en-us/publish-ssis-packages-as-odata-feed-sources-HA104079177.aspx>



Expose SSIS Package as SQL Query/View





Data Refresh with Power Query

- Support in Power BI
 - Scheduled Refresh on Power Query (no direct refresh)
 - Only Oracle and SQL Server data sources
 - Consider moving data to Windows Azure SQL Database and/or SQL Server on IaaS for direct refresh
- Refresh on the Cloud
 - Workload on source databases at refresh time
 - Possible latency for accessing data on-premise
 - Transfer data source on the cloud, transformation on the cloud (after query unfolding)



Conclusion

- Corporate BI supports Self-Service BI
 - Consume Data Warehouse and Data Marts
 - Expose Views
- Power Query can be managed by IT
 - Certify Query
 - Consume Views
- SQL Server as an additional data gateway
 - Power BI refresh only supports SQL Server and Oracle
 - Use linked server to access other providers
 - Use SSIS to extend access to real-time/on-demand transformations



Thank you!

Check our new articles on
www.sqlbi.com