

DB2 Web Query DataMigrator ETL Extension

Doug Mack
mackd@us.ibm.com

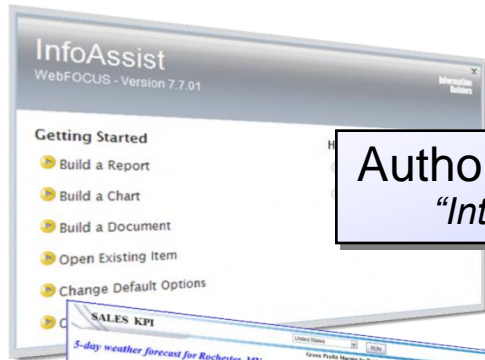
Rob Bestgen
bestgen@ibm.com

DB2 Web Query Team
QU2@us.ibm.com

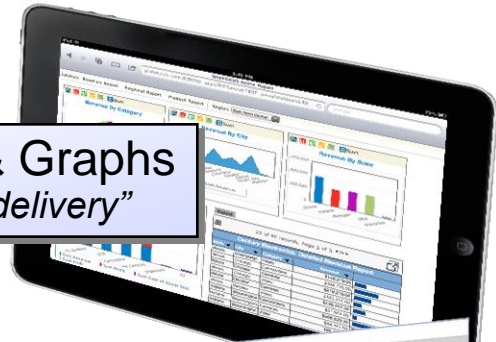


DB2 Web Query for i – Intuitive, Insightful, Extensible

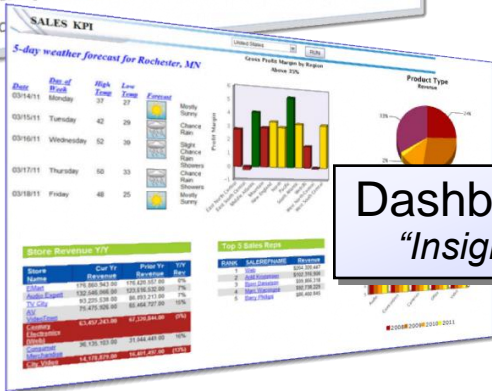
Fast, easy access to business information assets for query, reporting and analysis



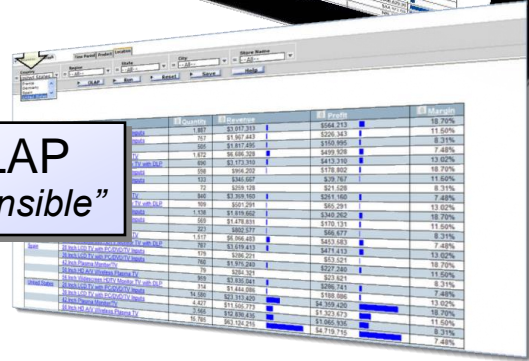
Authoring Tool
"Intuitive"



Reports & Graphs
"Flexible delivery"



Dashboards
"Insightful"



OLAP
"Extensible"

Providing Analytics and Query/400 Modernization for the IBM i Client
 Low Cost Solution Offering Value Upgrade from Query/400
 Leveraging Best Practice Meta Data Layer Approach
 Leveraging DB2 for i Advanced Query Optimization Technology

DB2 Web Query for i v2.1

Providing the Reporting side of DB2 Web Query

▪ Latest Version 2.1

- New Product IDs (5733-WQe/WQs)
- Core based pricing
- Available for IBM i v6.1 and v7.x

▪ Simpler Packaging Options

- Express Edition
 - ✓ Entry level edition
 - ✓ Report development
 - ✓ Great for small, getting started environments
- Standard Edition
 - ✓ Adds report distribution, integration APIs
 - ✓ Adds Microsoft SQLServer Access
 - ✓ Virtually unlimited runtime users support
 - ✓ Optionally add JDEdwards Adapter
- ✓ Add User Licenses to either Edition



DB2 Web Query Data Sources Support

- DB2 Web Query reports can contain data from these relational databases:
 - Any DB2 for i database in your enterprise
 - Any DB2 family member database in your enterprise
 - MS SQL Server (requires Standard Edition)
 - An imported (uploaded) Excel Spreadsheet
- DB2 Web Query fundamentally does not discern if those databases are any of the following designs as meta data maps to the data structure:
 - Production databases designed for transaction processing
 - Operational Data Stores (copies of production databases)
 - Data Marts or Data Warehouses (transformed, consolidated data bases)



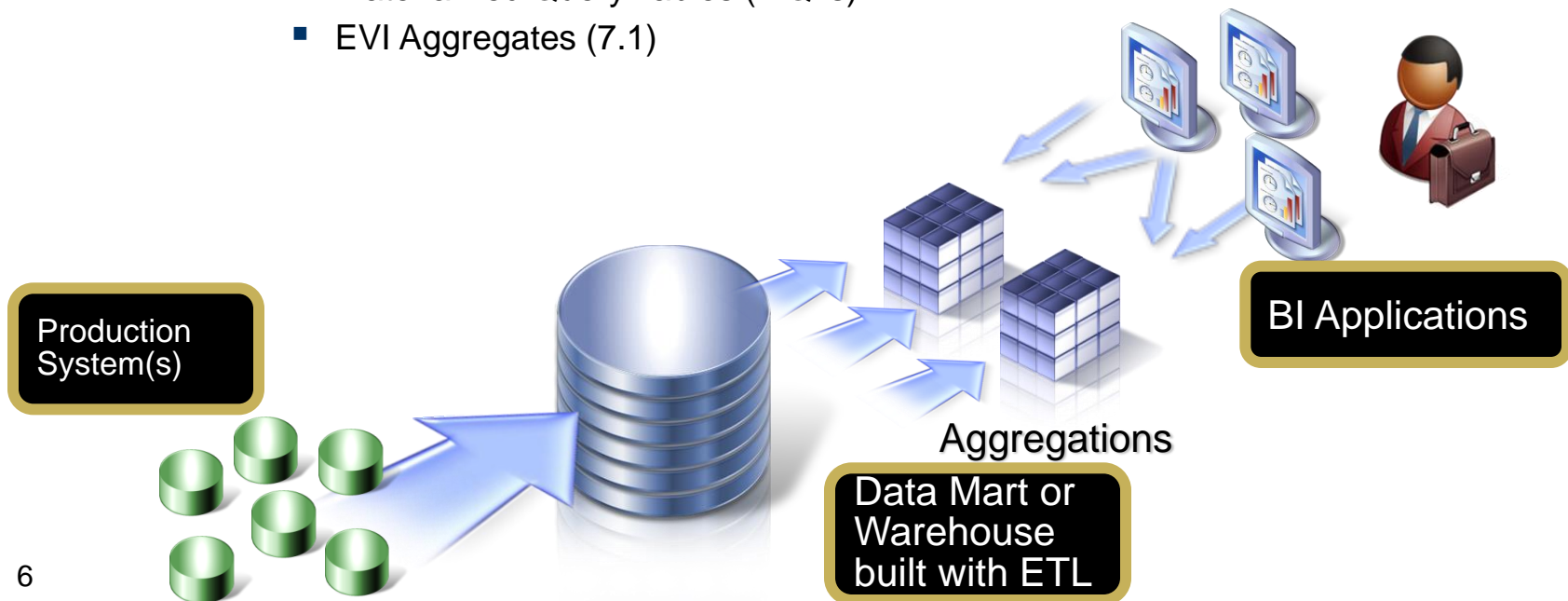
DB2 Web Query Data Sources Support

- For many, using production databases will be a simple solution
 - ❑ Run DB2 Web Query in the same LPAR/Server as your production applications
 - ❑ Existing resources available in a single production LPAR/Server
 - ❑ Smaller data volumes, and all historical data is still present
 - ❑ Tuning and keeping an eye on performance is a critical success factor
 - Options for optimization may be constricted by balanced workload desires
 - ❑ Creating a controlled ad-hoc requirement is a very good best practice to optimize the reports and avoid negative impacts on production systems



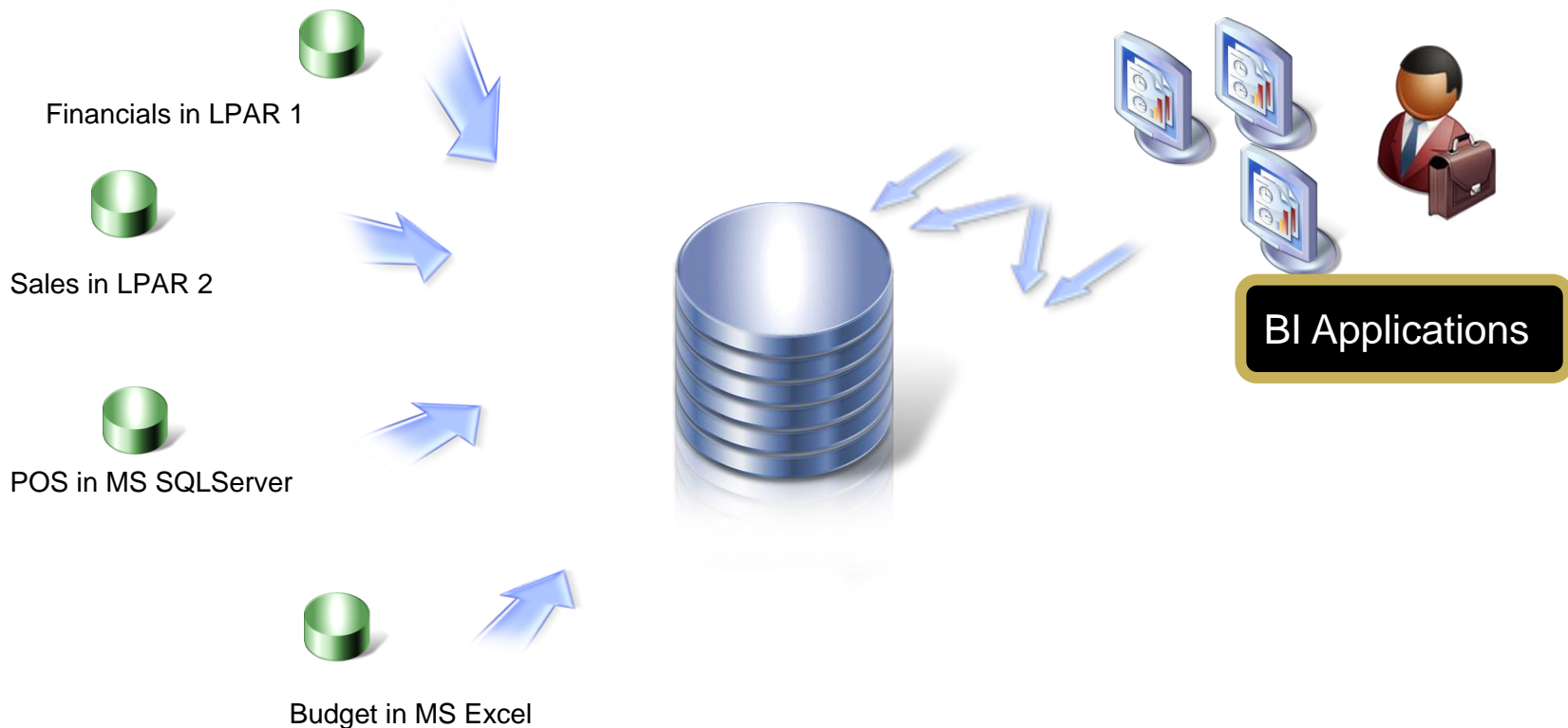
So Why an ETL (Extract, Transformation, Load) tool?

- Many customers will benefit from isolating query workloads from production workloads
 - Different workloads require different tuning
 - Leverage advanced DB2 for i query optimization and tuning techniques to boost performance without impacting transaction processing
 - Symmetric Multi Processing, Indexing Strategies
 - Dedicated subsystems
 - Create Aggregations to support BI type queries
 - Materialized Query Tables (MQTs)
 - EVI Aggregates (7.1)



So Why an ETL tool?

- Consolidate data from multiple data sources
 - ❑ 70% to 80% of a Business Intelligence (Analytics) project is in gathering data
 - ❑ DataMigrator for i automates the gathering and consolidation process from multiple data sources (including Excel spreadsheets)
 - ❑ Cross database/system joins from the reporting server can be painful performance wise



So Why an ETL tool?

■ Prepare the data for analytics

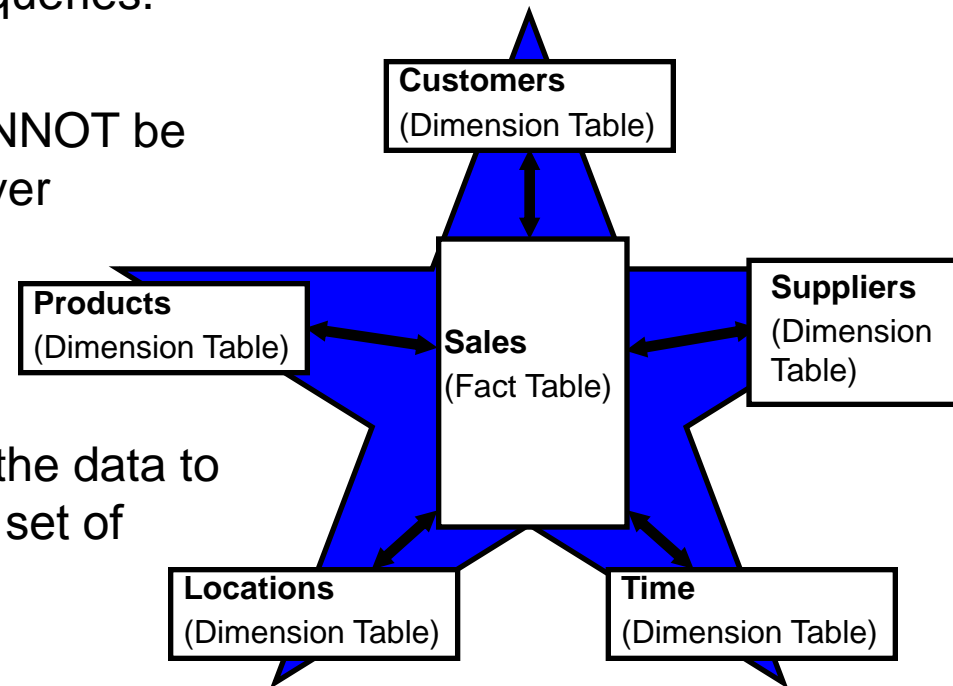
❑ Transaction oriented databases are designed for simple database operations (create, update, or delete a record). They are not designed for analytical queries and are less than optimal when used as the source for such queries.

❑ There are data challenges that CANNOT be solved easily with the meta data layer

- Surrogate Key Generation
- Slowly Changing Dimensions

❑ Transform. Cleanse and SECURE the data to provide “single version of the truth” set of trusted, secured data

- Star Schema model
- More historical data perspective

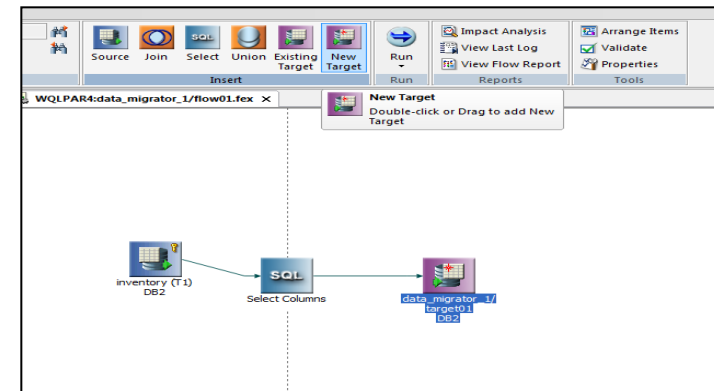
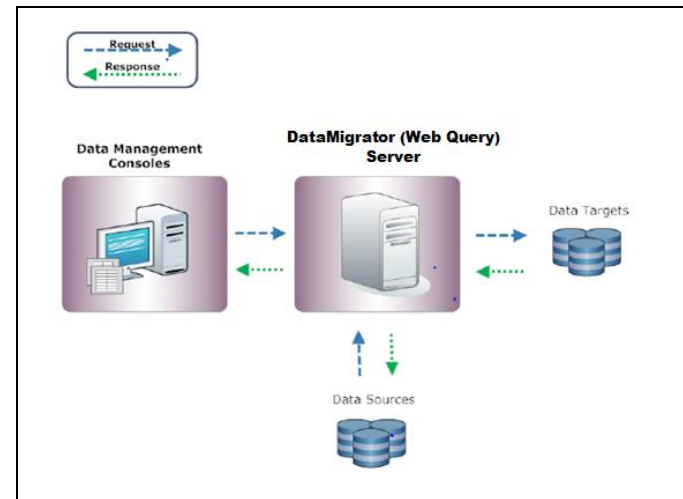


DataMigrator ETL Extension

- New Product ID 5733-WQM
 - Requires either DB2 Web Query Express or Standard Edition V2.1 with Group PTF Level 11
 - Core based. Must have same number of cores as DB2 Web Query core licenses
 - Standard Edition required for MS SQLServer as a data source
 - One license of DB2 Web Query Developer Workbench is required to set up data and process flows

- ETL for IBM i
 - Like DB2 Web Query, all server code and ETL objects run in and/or are stored in IBM i

- INTEGRATED with DB2 Web Query
 - Can leverage meta data created with DB2 Web Query or vice versa
 - Same look and feel
 - Shared services and administration

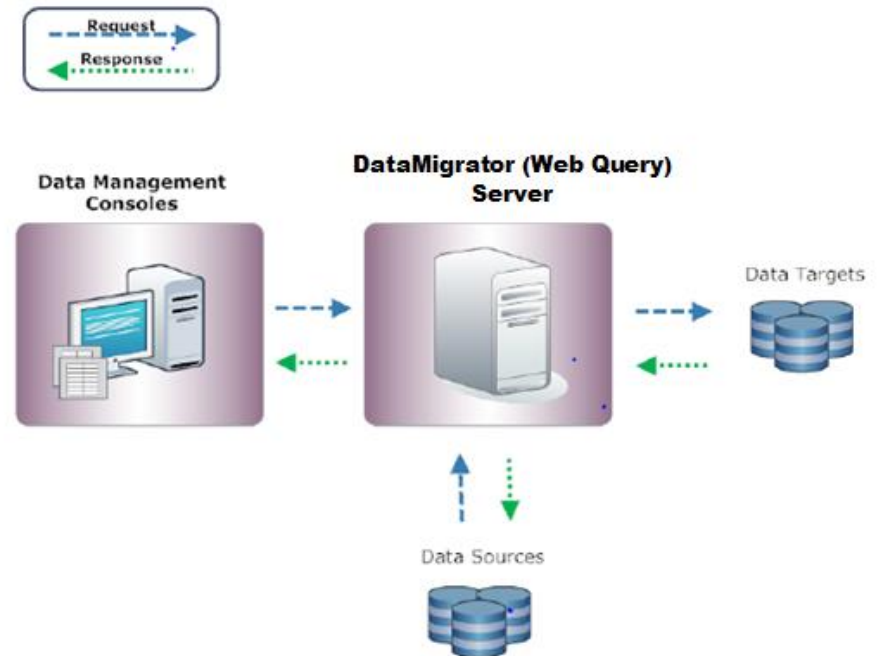


DataMigrator Architecture

DataMigrator for i has two main components:

- DataMigrator Server
 - Engine that stores and runs ETL ‘Flows’
 - Same running server as Web Query

- Data Management Console (DMC)
 - Graphical user interface
 - Build and manage ETL ‘Flows’
 - Included in Developer Workbench

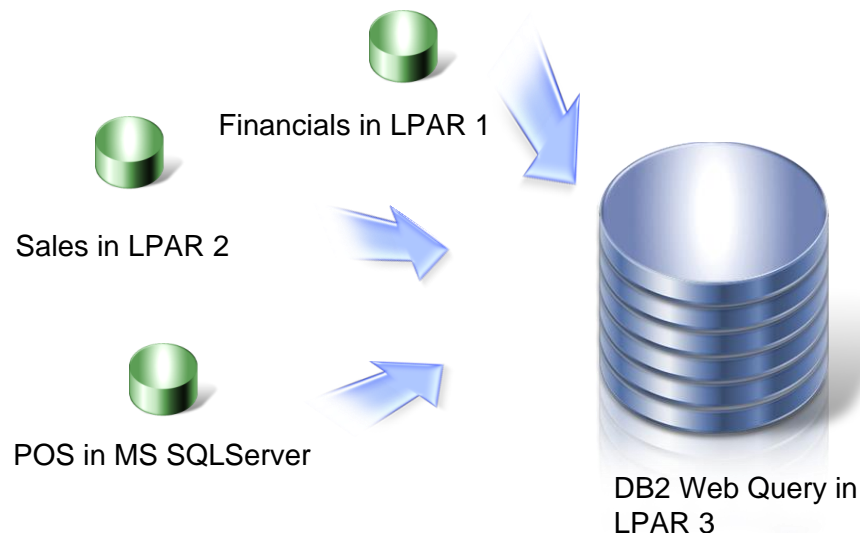


DataMigrator ETL Capabilities

- **Extract** data from
 - Any data source you can create a synonym for today in Web Query
 - Database files, procedures, etc...
 - SQL Server (with Standard Edition)
 - DB2 family (LUW and z)
 - Flat (IFS) files
 - Database file journals
- Apply **Transformation** logic to ‘clean up’, join and convert data into a desirable form
 - Including aggregating data to simplify reporting and decision support
- **Load** data into a set of target tables such as a data warehouse
- Integrate multiple data sources from multiple systems onto a single system
- Schedule updates to synchronize data sources
 - With change data capture (CDC) feature

DataMigrator Example

- Dedicated LPAR or Server for DB2 Web Query and Data Repository
 - DB2 Web Query and DataMigrator ONLY installed in LPAR3
 - DataMigrator automates process of pulling data from sources into LPAR3
 - Pull directly from operational databases (full refresh or based on some selection criteria, example where date=current_date)
 - Pull from a Journal Receiver (Change Data Capture)
 - All reports run against data in LPAR 3 database

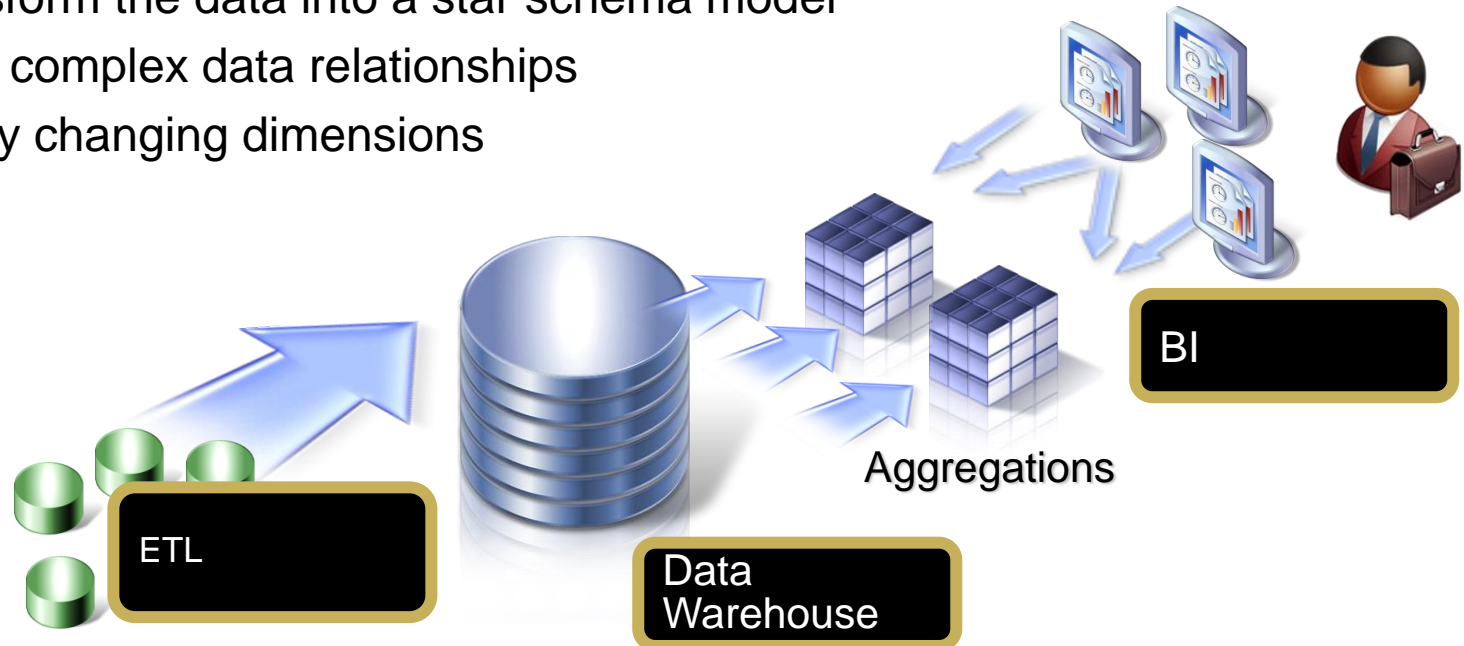


What is an Operational Data Store

- A COPY of the operational (transaction) data base most often used for reporting purposes
- Isolated from production workloads (usually)
- Kept up to date based on requirements
 - Near real time? Nightly?
- The data model stays pretty much the same as production, with possibly some minor changes
 - Could have subset of fields/columns
 - Could contain more historical data than production systems
 - Could contain minimal data transformations
- Ability to enhance for performance
 - Aggregations of the data
 - Indexing
 - Work Management
- Security model might be different than production data
 - Eliminate sensitive data not required for analytics

Why Consider a Data Mart or Data Warehouse?

- When additional Data Sources are required for analytics applications
 - Pulls together and relates data coming from different systems or applications
- When cleansing and/or enhancing of data required
 - Operational data contains anomalies that can be addressed with transformations (e.g., validate data)
 - Use data profiling to help identify those anomalies
- Data model restructuring required or desired
 - Transform the data into a star schema model
- Managing complex data relationships
 - Slowly changing dimensions



Example: Non unique keys across databases



U.S. System

CUST_NO	CUST_NAME
1122	Acme Corporation
1123	IBM
1124	Olympia Brewpub
1125	XYZ Inc



Canada System

CUST_NO	CUST_NAME
1122	Victoria Analytics
1124	Elizabethan Enterpri:
1126	Maple Leaf Brew
1128	Ice R Us

SOURCE Systems contain similar customer numbers for different customers

CUST_NO	CUSTOMER_NAME	REGION	OLD_CUST_NO
0001	Acme	United States	1122
0002	IBM	United States	1123
0003	Olympia Brewpub	United States	1124
0004	XYZ Inc	United States	1125
0005	Victoria Analytics	Canada	1122
0006	Elizabethan Enterprises	Canada	1124
0007	Maple Leaf Brews	Canada	1126
0008	Ice R Us	Canada	1128

TARGET System contains consolidated table with surrogate UNIQUE key

How Does it Work ?

Major concepts of DataMigrator

- The process for defining the source and target of an ETL process and the transformations between them is encapsulated in a **Flow**
- DataMigrator has a concept of initial (bulk load) and incremental updates
 - For small source data sets, repeated (often nightly) bulk loads may be sufficient
 - For larger data sources, normally an initial load is done one time, then incremental updates occur on a scheduled basis (nightly/weekly/hourly)
 - Incremental updates can be based on log scraping (journal receivers)
- Scheduling flows is a critical concept (who would want to manually run loads and updates?)
- Dealing with and transforming data is a messy business!
 - Error handling is common, not rare
 - E-mail notification capabilities

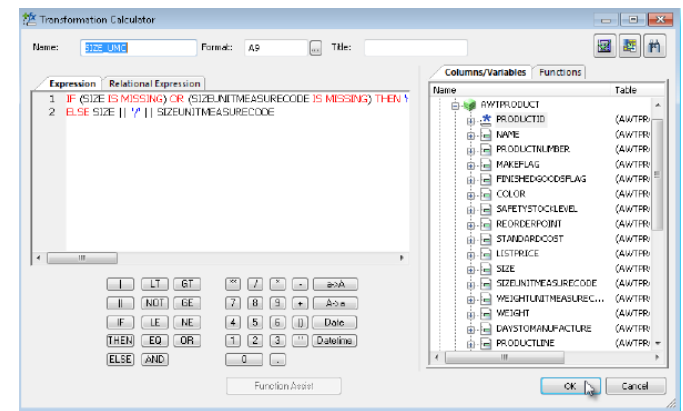
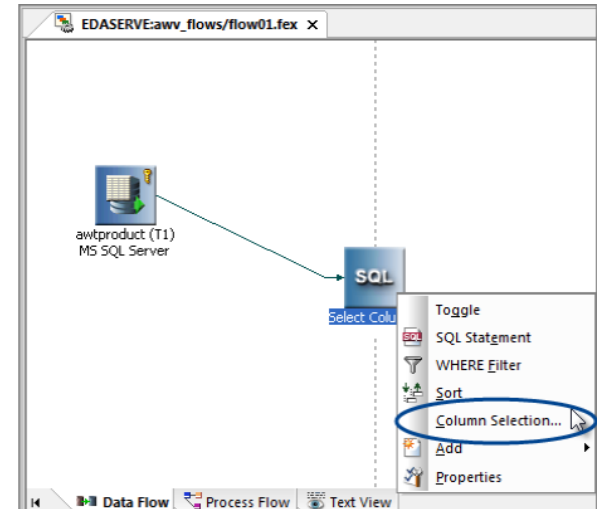
How Does it Work?

- DataMigrator has a **Server** component and a **Client** component
- **Server** - the Server component IS the Web Query reporting server
 - The DataMigrator portion of the server, the part that understands and processes DataMigrator flows and schedules, is enabled via option 8 - DataMigrator (5108)
 - The server is the main piece that runs Flows and handles schedules
 - Can and does run flows and schedules without requiring a client to be connected
 - Can use the web console interface of Web Query to **view** aspects of DataMigrator
 - The number of cores enabled for DataMigrator always matches Web Query
- **Client** - main tool for working with DataMigrator is the **DMC - Data Management Console**
 - Used to create, modify, schedule, and check on Flows
 - Windows PC based tool, part of Developer Workbench
 - Installing Workbench automatically installs DMC
 - Licensed the same way as Workbench, NO new user types
 - Requires option 5 - Workbench Users (5105)
 - Requires a licensed user - Licensed Users (5104), just like WorkBench
 - Enabled when the target (IBM i) server is licensed/enabled for DataMigrator

What is a Data Flow?

A Data Flow is where you define the data sources, transformation, and map to the target tables.

- Leverage existing DB2 Web Query meta data or create as part of the process
- Select columns, add where clause, edit SQL, or join files as part of the source definition
- Create derived columns
- Choose or create new target table(s)
- Determine load type (default is insert/update)
- Test your transformations, edit if necessary



What is a Process Flow?

A Process Flow defines a set of steps that execute a series of components or objects, including the data flows that define your extractions, transformation, and load logic. You can combine:

- Use properties to schedule a process flow.
- Data Flow objects that indicate what data to extract and move from data sources to data targets.
- Email objects that notify users about the status of the process at specified points.
- Stored Procedure objects that perform a variety of supplementary tasks.
- Connector objects, represented as arrows, that specify execution logic for the other objects included in the process flow.
- Group objects that specify the flow of a subset of objects within the total flow.

What makes DataMigrator for i different?

■ LOW PRICE

- Most ETL tools will run from \$50,000 to \$250,000
- ✓ DataMigrator for i is a fraction of that price

ANNOUNCED: 2/23
GA: 3/27

■ OPTIMIZED FOR IBM i

- Almost all other ETL tools will not run in IBM i nor leverage DB2 for i capabilities, will require x86 servers, and add complexity and potential performance and availability challenges

■ INTEGRATED with DB2 Web Query

- While DB2 Web Query does not HAVE to be the BI/Reporting tool of choice to leverage DataMigrator, there are significant integration points that allow you to leverage commonality in usage and meta data synchronization

■ SQL Based

- Set at a time operations on data – versus record at a time – implies more powerful processing of the data

DataMigrator for i FAQs

- How do I order DB2 Web Query for i DataMigrator ETL Extension
 - Support for DataMigrator is provided in IBM's e-configurator tool at General Availability (GA) date (March 27).

- Is it available on a trial basis ?
 - Yes, like most IBM i Licensed Program Products, DataMigrator for i is available under a 70 day trial

- How does one acquire/download the product for trial or with license keys?
 - DataMigrator is included in the Group PTF for DB2 Web Query as of GA date. Installation instructions for DataMigrator are posted on the DB2 Web Query wiki at ibm.co/db2wqwiki

- This all sounds very exciting. What if we need assistance in designing the data mart database and implementing the ETL data and process flows?
 - The DB2 for i Center of Excellence, part of Systems Lab Services, can assist in consulting and training around all the DB2 Web Query products, as well as in the area of data modeling and SQL performance optimization. Reach us at QU2@us.ibm.com

DataMigrator for i FAQs

- Can I use other query, BI, or Analytics tools with a datamart built by DataMigrator, or MUST I only use DB2 Web Query for reporting/analytics?
 - While there are integration points between DB2 Web Query and DataMigrator, the ETL process builds a DB2 for i database which supports open standards that would allow many other tools (Cognos, Business Objects, etc.) to be used with it. So....yes, you can use other tools and are not locked into DB2 Web Query for reporting. But, DB2 Web Query Express or Standard Edition are a pre-req for DataMigrator from a licensing perspective

- Can DataMigrator pull data from an Oracle database?
 - DataMigrator data source and target support are consistent with DB2 Web Query's data source support, which is DB2 for i, or any DB2 family member, and MS SQLServer. It also supports Oracle JDEdwards applications where DB2 for i is the database in question. So the answer is NO, if you are referring to an Oracle database (and not JDE running in IBM i). However, you could supplement DataMigrator with some other tool (see the question about CDC later) to move data from Oracle into a DB2 Web Query supported data source

DataMigrator for i FAQs

- Could I use DataMigrator in a conversion/migration scenario from a Windows platform to IBM i?
 - In theory, yes, as it could provide data movement/replication with mappings. In terms of a dual maintenance strategy (the idea that you want to keep two databases on different platforms synchronized in near real time), it is recommended that you have a design review to ensure it could meet the requirements. Keep in mind the previous comment about data source and target support as well
- IBM already offers a couple of ETL solutions. What makes this different?
 - IBM does have some other ETL solutions, including InfoSphere DataStage and Cognos DataManager. Both of these solutions are robust ETL tools that could meet your customer's needs. Neither runs in IBM i but can support DB2 for i as a source or target. If your customer is looking for a low cost, IBM i centric solution, then DataMigrator is likely to fill that need
- What about IBM's InfoSphere CDC ?
 - While there is some overlap in functionality, CDC is designed as a record oriented REPLICATION tool. However, it falls down in a true ETL environment as it has very minimal data transformations available to it. There may actually be scenarios where these products compliment each other – most likely if you need to access a data source supported by CDC and not by DB2 Web Query (e.g. Oracle)

Additional Collateral and Resources

Product Demos and Getting Started Videos


- Product Demos
 - Auto Drill down (OLAP) reporting
 - Dashboard Options
 - Working with Excel
 - Upload Wizard
 - Meta Data Wizard


- Getting Started
 - Acquiring and Installing
 - Setting up Security
 - Creating Meta Data over Sample database

- Access via DB2 Web Query website or wiki
 - ibm.com/systems/i/db2/webquery
 - ibm.co/db2wqwiki

Additional information

 [DB2 Web Query V2.1 Enhancements \(151KB\)](#)

 [JDEdwards Adapter for DB2 Web Query \(1.72MB\)](#)

 [DB2 Web Query for i quick-look demo \(00:05:57\)](#)

[→ View DB2 Web Query End User Demos](#)

[→ View DB2 Web Query Wizard Analytics in action](#)

[⇌ Get Adobe® Flash® Player](#)

Version 2.1 is here! Find out the [details](#) now!

Check out these wiki pages for more detailed information. For older releases, [click here](#).

- [DB2 Web Query Home Page](#)
- [DB2 Web Query End User Reporting Demonstrations](#)
- Documentation
 - Accessibility Guide
 - Active Technologies Adobe Licensing
 - Active Technologies for Mobile Web Apps
 - Application Integration
 - Articles
 - Backup, High-Availability, and Disaster Recovery
 - Business Partner Support Documents
 - CRTWQSYN CL Command
 - Dynamic Runtime Environments
 - Excel Plug In
 - Functions
 - Installation
 - Interfaces
 - JD Edwards Application Adapter
 - Known Issues
 - Microsoft SQL Server Adapter
 - New Features
 - PHP Integration
 - Query/400 Discovery Tool
 - Redbook
 - Report Broker
 - Release 2.1 Product Manual
 - Run Time User Enablement
 - SSL Enablement
 - SSO Kerberos
 - Support KnowledgeBase
 - System Variables
 - [Wizard Analytics](#)
 - Workload Capping
- Education
- Forum
- [Getting Started Videos](#)
- DB2 for i Center of Excellence Web Query Training and Consulting Services
- PTFs and On-going Service

New Education Offerings

- **OD04 DB2 Web Query – Introduction to Reporting**
 - Updated to the latest level in Q4 2014
 - 1 Day INTRO course focused in report writing
 - Heavily hands-on lab oriented
 - Taught by IBM Global Training Partners
- **OD05 DB2 Web Query for Advanced Users, Developers, and Administrators**
 - Updated to the latest level in Q4, 2014
 - 3 Day Advanced Class covers much deeper content
 - Bring to an individual client or a group of clients
 - IBM Instructor led course
- **SQL Performance Monitoring, Tuning and Analysis Workshop**
 - Get a Database Engineer trained up on DB2 for i
 - EVERY large client should attend this class to keep workloads on i
 - IBM Instructor led – bring to your client or attend public class
 - Upcoming class in MAY in Rochester !



DB2 for i Center of Excellence Offerings

- DB2 Web Query Getting Started Service
 - 3 Day skills transfer and workshops to build prototype
 - Goes beyond tutorials with Best Practices and hand-holding
 - Best way to ensure successful implementation
 - Should include with EVERY DB2 Web Query proposal (proposal insert available)
- Query/400 Modernization
 - Discovery, Analysis, and Conversion
 - Addresses the question of how a customer can move from 100's or 1000's of Query/400 definitions to modernized environment that leverages DB2 query optimization!
 - Out with the (very) OLD, in with the NEW – improve performance, productivity, perception of IBM i, and move into “analytics”
- DB2 Web Query V1 to V2 Migration Service
 - Remote assistance to help customers move to Version 2
 - While there exists migration tooling, migrations can become more complex depending on the customer's DB2 Web Query Environment - inexpensive service to hand hold the migration
- DB2 and SQL Performance Assessments
 - Really critical assessment that should be conducted regularly by the customer
 - But lack of DBAs in our i community means no one is paying attention to the database
 - Assessment collects database (and system) performance data
 - Recommendations made in report

Links

- DB2 Web Query for i Marketing Website
 - www.ibm.com/systems/i/db2/webquery
- DB2 Web Query for i Wiki
 - lbn.co/db2wqwiki
- DB2 Web Query Getting Started Enablement:
 - <https://www.ibm.com/developerworks/community/wikis/form/anonymous/api/wiki/96f36e44-b59d-40c8-8efb-e3267e86b310/page/ae2311b7-c193-4647-b428-cf4aa871825a/attachment/4967ccde-083f-4ff7-b951-c5e7a6e19011/media/db2%20web%20query%20getting%20start%20services.pdf>
- DB2 Web Query Classroom Training
 - https://www.ibm.com/developerworks/community/wikis/home?lang=en_US#!/wiki/W516d8b60d32c4fc5a8115f3d840bf524/page/Education
- Demonstrations:
 - <https://www.ibm.com/developerworks/community/wikis/home?lang=en#/wiki/W516d8b60d32c4fc5a8115f3d840bf524/page/DB2%20Web%20Query%20Demonstrations>
- FAQs
 - http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=ST&infotype=SA&appname=STGE_PO_PO_USEN&htmlfid=POQ12347USEN&attachment=POQ12347USEN.PDF#loaded

