IBM Data Studio for Mainframe Developers

David Simpson, Senior Technical Advisor
Themis, Inc.
dsimpson@themisinc.com
www.themisinc.com
Themis and Themis, Inc. are trademarks of Themis, Inc.

DB2, IBM Data Studio, Visual Explain, Stored Procedure Builder, Rational and Control Center are trademarks of the IBM Corporation.

Eclipse is a trademark of the Eclipse Foundation.

Other products and company names mentioned herein may be trademarks of their respective companies. Mention of third party products or software is for reference only and constitutes neither a recommendation nor an endorsement.

© Copyright Themis, Inc. March 2012

David Simpson is currently a Senior Technical Advisor at Themis Inc. He teaches courses on SQL, Application Programming, Database Administration as well as optimization, performance and tuning. He also installs and maintains the database systems used for training at Themis. Since 1993 David has worked as a developer and DBA in support of very large transactional and business intelligence systems. David is a certified DB2 DBA on both z/OS and LUW. David was voted Best User Speaker and Best Overall Speaker at IDUG North America 2006. He was also voted Best User Speaker at IDUG Europe 2006 and is a member of the IDUG Speakers Hall of Fame.
# Table of Contents

IBM DATA STUDIO FOR MAINFRAME DEVELOPERS .........................................................1

IBM Data Studio ..................................................................................................................4
  Eclipse Terminology .........................................................................................................6
  Workspace ..........................................................................................................................7
  The Workbench .................................................................................................................8
  The Welcome View .........................................................................................................9
  Eclipse Tips & Tricks .......................................................................................................10
  The Data Perspective ......................................................................................................11

Configuring a Database Connection ..................................................................................12
  Database Connection Dialog .........................................................................................13
  Database Connection Details .........................................................................................14

Creating a Project ................................................................................................................15

IBM Data Studio Functions ...............................................................................................18

Catalog Interface ..............................................................................................................19

Database Administration .................................................................................................20
  Generating DDL ...............................................................................................................21
  DDL Options ...................................................................................................................22
  DDL Output .....................................................................................................................23

Creating a SQL Script .........................................................................................................24
  Editing the Script .............................................................................................................25
  Running the Script ..........................................................................................................26
  Script Output ..................................................................................................................27

Using Visual Explain .........................................................................................................28
  DB2 10 Explain Tables ....................................................................................................29
  Visual Explain Options .................................................................................................30
  IBM Data Studio Access Path Graph ..............................................................................31
  Reading Access Path Graphs ..........................................................................................32
  Node Detail – Sources ....................................................................................................33
  Node Detail – Operations ............................................................................................35
  Optimizer Predicate Categories .....................................................................................36
  Table Scans and Stage 2 Predicates ..............................................................................37
  Single Query Tuning .......................................................................................................38

Stored Procedure Development .......................................................................................43
  Types of Stored Procedures ..........................................................................................44
  External Stored Procedures .........................................................................................45
  Native Stored Procedures .............................................................................................46
  Creating a New Stored Procedure ................................................................................47
  Copying an Existing Procedure ....................................................................................48
  Drag and Drop ...............................................................................................................49
  Deploying a Stored Procedure ......................................................................................50
  Running a Stored Procedure .........................................................................................51
IBM Data Studio

IBM Data Studio is a free product provided by IBM to allow users, developers and administrators to interact with databases from a client machine. IBM Data Studio may be used on a variety of platforms, but is most commonly used with the DB2 family of products. Most of the functionality from previous client tools such as Visual Explain, Control Center and Stored Procedure Builder (aka Development Center & DB2 Developer Workbench) has migrated or is in the process of being migrated to this new platform. The Eclipse IDE was selected as the environment to be used. Java developers are generally familiar with this interface since it has been used as a Java development environment for about ten years. Many products in IBM's Rational line are built on the Eclipse framework.
Mainframe developers are sometimes less familiar with this environment. This presentation is meant to help those unfamiliar with Eclipse and IBM Data Studio get started with these products.

This link references a video provided by Themis that covers the Eclipse Environment and the IBM Data Studio Environment. Gregg Lippa, a Java and Eclipse expert joins David Simpson, a DB2 DBA to discuss the use of Eclipse by mainframe professionals.

http://www.youtube.com/watch?v=tHqf6yUkQas
Eclipse Terminology

- Workspace
- View
- Perspective
- Project

Eclipse Terminology

Some basic terminology is necessary when discussing the Eclipse Environment.
Workspace

In Eclipse terminology a *workspace* is a folder on your client machine where your work will be stored. This work will include meta data (such as how to connect to various databases), projects (stay tuned for more) and settings about how Eclipse will behave. When you first enter IBM Data Studio, you will be using the default workspace. You can change your workspace at any time by clicking on File -> Switch Workspace. You will then see the folder that Eclipse is using.
The Workbench

When IBM Data Studio is launched the first thing you will see is the workbench. The workbench is made up of one or usually more views where work will be accomplished. Each window inside the workbench is referred to as a view that may be closed, minimized or resized as needed. The default views and positions will depend on which perspective you are using (more on that later).
The Welcome View

Any view may be closed by clicking on the “X” on the tab for the view. Whenever you see the “Welcome” view…. Close it. 😊
Eclipse Tips & Tricks

• Expand and Contract Views by double-clicking the tab

• Use “Window -> Reset Perspective” when bad things happen

• <CTRL> – Spacebar gives context sensitive help

Eclipse Tips & Tricks

Views may also be expanded to consume the entire workbench by double clicking on the tab for the view. Double clicking on the tab again will place it back where it belongs on the workbench. This will become especially important when you are working on something exclusively (such as a query script or stored procedure) and the other views on the screen become “noise”.

Whenever you want to return everything to its normal place in your perspective, you can click “Window” then “Reset Perspective”. This will retrieve any “lost” views.

When using an editor you can get context sensitive help by pressing CTRL-space.
The Data Perspective

In Eclipse, a perspective is a default collection of views and actions along with the size and position of the views on the screen. When using IBM Data Studio it is often helpful to use the “Data” perspective shown here. You can change your perspective using the button on the top right. You can also customize or create your own perspectives. You can add views by clicking Window -> Show View.
Configuring a Database Connection

IBM Data Studio does not require a DB2 client to make database connections. It uses the Type 4 JDBC driver supplied by IBM to make connections directly to DB2 on any platform with no middleware. This driver is provided with a limited use license for use with IBM Data Studio even when connection to DB2 for z/OS databases (this is normally a licensed product that costs money). If you do have a DB2 product installed on your machine with nodes and databases configured, you will see those connections already configured in the Data Source Explorer view. If you do not have a DB2 product installed this view will be empty. It is easy to configure a connection to a new database provided you know the IP address or host name of the server on which it resides and the port number.

Clicking on the button shown here will begin the process of configuring a new database connection.
Configuring a Database Connection

![Database Connection Dialog]

- **Type 4 JDBC Driver**
- **Select Platform**

Database Connection Dialog
Database Connection Details

When you have filled in all the appropriate information it is a good idea to use the “Test Connection” button to avoid too many “under the covers” attempts at logging in with the wrong password later.
Creating a Project

All work in Eclipse is created and stored in a project. You can think of a project as a subfolder in your workspace stored on the local drive of your workstation. A project is a place to put the things you create such as SQL scripts, stored procedures, or the results of explains and query tuning. There are different kinds of projects that can hold different kinds of work. In IBM Data Studio most kinds of work will require a Data Development Project. You can create as many projects as are necessary. Each project will have a default database connection that it uses.
Creating a Project

Give the project a name.
Creating a Project

Every data development project is associated with a database connection where the work will be done.
IBM Data Studio Functions

- Catalog Interface
- Physical Data Models
- Database Administration
- SQL Scripts
- Developing and debugging SQL PL Stored Procedures
- Testing any kind of Stored Procedures
- Query Tuning

IBM Data Studio Functions

IBM Data Studio is a very versatile tool. As mention previously, many of the functions from other client tools have been combined into this product. This presentation will deal with some of these features.
The Data Source Explorer view provides an interface to the DB2 catalog and allows a user to browse the environment. The folder labeled “Schemas” under each database connection contains lists of tables, views, stored procedures and other objects. This example shows navigation to a table and the columns and indexes on that table.
Database Administration

IBM has included some basic Database Administration features in the free version of this product. You can generate DDL for an object by right-clicking on the object and selecting “Generate DDL”.

© 2012 Themis, Inc. All rights reserved.
Generating DDL

DDL options will be presented on the next two screens. You may select what to generate and omit using the checkboxes.
DDL Options

```
Objects
Select the model objects that you want to include in the DDL script.

Generate:
- Check constraint
- Foreign key constraint
- History table
- Indexes
- Primary key constraint

- Privileges
- Tables
- Tablespaces
- Triggers

Select All  Deselect All  Restore Defaults

< Back  Next >  Finish  Cancel
```
DDL Output

The DDL output will be displayed. It may then be imported into a script in your project or saved as a text file.
Creating a SQL Script

The Data Development project contains subfolders that can be used to create and store database objects that you develop.

Creating SQL Statements: Right click on the new project, and select New -> SQL or XQuery Script.

Provide a name. No spaces can be used in the name.

A wizard may be used to help generate a SQL statement or you can code it yourself.
Editing the Script

After clicking Finish, data studio brings up an empty editor box. Like many other Data Studio features, the editor provides user assistance to create SQL statements. The user assistance can be triggered by pressing the key combination *Ctrl+Space*.

Configuration tabs at the top of the query screen will allow you to set the database connection and control the behavior of the script.

Development of XQuery scripts is supported by Data Studio when connecting to a server with XQuery support, such as DB2. XQuery supports path expressions to enable developers to navigate through XML’s hierarchical structures.
Running the Script

There are multiple ways to execute the query:

- Right click in the editor that shows your script
- Right click your script object in the Data Project Explorer
- Hit the F5 key
- Click on the Green Arrow icon
Script Output

The output of the query will show up in the properties view at the bottom of the screen. A history of activity is also kept on the left side. You may navigate to a previous result. If your script has multiple statements, it may be necessary to open the folder on the left side to see the detailed results.
Using Visual Explain

Click Visual Explain icon OR Right-click on your SQL Script & choose Visual Explain

SELECT *
FROM EMP
WHERE LASTNAME = 'Wynton'
AND FIRSTNME = 'John'
AND DEPTNO = 'A00'
DB2 10 Explain Tables

To use Visual Explain or the Query Tuner in IBM Data Studio, the full suite of explain tables must be available. While a PLAN_TABLE is all that is needed to do a regular explain, the extra tables enable Visual Explain to render additional information about the explain. Because there are so many explain tables, DBAs usually ask groups of developers to share a set. This may be done either by creating aliases from each user id to the explain tables to be shared or by creating the tables under the owner of a security group to which the users can set their SQLID.
Visual Explain Options

If you specify different values for CURRENT SCHEMA and CURRENT SQLID, Visual Explain searches for explain tables that are qualified by the value of CURRENT SQLID. If Visual Explain does not find explain tables that are qualified by the value of CURRENT SQLID, Visual Explain attempts to create the explain tables under that value. These values can be set either dynamically:

```
SET CURRENT SCHEMA = 'XXXXX';
SET CURRENT SQLID = 'XXXXXX';
```

Or thru Window-> Preferences -> Data Management > Visual Explain -> Visual Explain Settings.
SELECT *
FROM EMP
WHERE LASTNAME = 'Wynton'
AND FIRSTNAME = 'John'
AND DEPTNO = 'A00'

Index XEMP03
Index XEMP02
Reading Access Path Graphs

In this graph, there are two nodes that indicate they are a source of data. Data is accessed by doing an index scan of the XEMP03 index. Once appropriate rows are identified then the data in the EMP table is retrieved using the identifiers from the index.

If one of these nodes is highlighted by clicking on it the catalog data about the object is displayed on the left side of the screen. Statistical information is displayed as well as the timestamp when statistics were last gathered.

By navigating the tree at the top left of the screen, information may be viewed about the table, tablespace and any other indexes that exist on the referenced table.
Node Detail – Sources

If one of these nodes is highlighted by clicking on it the catalog data about the object is displayed on the left side of the screen. Statistical information is displayed as well as the timestamp when statistics were last gathered.

By navigating the tree at the top left of the screen, information may be viewed about the table, tablespace and any other indexes that exist on the referenced table.
Reading Access Path Graphs

The data retrieval nodes are the areas that show the predicate information and where they are applied. Predicates in an SQL statement are executed at the index level (IXSCAN) or the data file level (FETCH).
If one of these nodes is selected the left side of the screen will show the metrics that DB2 used in determining that this was the appropriate access method. Predicate level data is shown as well as row estimates for how many rows will be passed to the next operation. These estimates may then be compared to reality to determine if the optimizer made a good choice.

In this example, DB2 is showing 2 predicates that will be applied at the index level. Each predicate is shown with is ‘guesstimated’ filter factor. Filter Factor: Is the percentage of rows that DB2 thinks will be affected by that predicate alone. In this example:

LASTNAME = ‘Wynton’ DB2 thinks .1 percent of the data matches

FIRSTNME = ‘John’ DB2 thinks .38 percent of the data matches
Optimizer Predicate Categories

- Stage 1 or Stage 2
- Indexable or Non Indexable
- Indexable matching, Indexable Screening
- Filter Factor
- Predicate Type (Equal, Between, Range, Like, etc..)
Table Scans and Stage 2 Predicates

Each predicate is classified as either Stage 1 or Stage 2. Stage 2 predicates generally require more resources and are not indexable. Visual Explain shows you which predicates are Stage 1 and which are Stage 2. Attention should be given to any Stage 2 predicates to see if they may be re-written.
Data Studio Single Query Tuning

Basic single query tuning features: These can also be done using the Optimization Service Center Tool, which is being deprecated as of V9.

- Capture queries from many data sources (Files, Cobol programs, XML files, etc.)
- View formatted queries.
- View access plan graphs.
- Capture information about the data server that queries run against, a feature which corresponds to Service SQL in Optimization Service Center for DB2 for z/OS
- Generate reports on the performance of queries.
- Run the Query Statistics Advisor to analyze the statistics that are available for the data that a query accesses, check for inaccurate, outdated, or conflicting statistics, and look for additional statistics that you might capture to improve how the data server processes the query.
- See any reformatting of an SQL statement the Optimizer modifies

Single Query Tuning

In addition to running Visual Explain directly from the SQL script, you may also invoke the query tuner. This feature will present the Visual Explain graph in addition to several other features.
Using the Query Tuner

The button shown above will invoke the query tuner.
If you ‘Select what to Run’, you will notice the different options available and not available when it comes to getting extra information in order to ‘Tune a query’.
At this point all the available ‘free’ option are marked. By selecting the ‘Run default Advisors and Tools’ option, all the information will be generated (Formatted query, Access Plan Graph, Statistics Advisor, and report). All of these options are available after the generate on the left side of the following screen:
Selecting the options out of the Review tab on the left will display the extra tuning information available through Data Studio. Note when the ‘Open formatted Query’ is selected, the original and transformed queries are presented. Often times the optimizer will rewrite queries a different way than presented. This shows any rewrite (transformation).
Here’s another video. This one features Themis instructors Linda Claussen and David Simpson discussing Native SQL Stored Procedures on both DB2 9 and DB2 10.

http://www.youtube.com/watch?v=pnwxEdmeJDk
Types of Stored Procedures

- **External**
  - COBOL
  - C, C++
  - PL/I
  - Assembler
  - REXX
  - Java
  - SQL PL

- **Native**
  - SQL PL
  - New in DB2 9 for z/OS

Prior to version 9, DB2 on z/OS supported only external stored procedures. An external stored procedure runs in an address space outside of DB2. An external stored procedure can be written in many different languages, such as COBOL, C, C++, PL/I, JAVA, Assembler, and REXX. The SQL Procedure Language (SQL PL) was introduced to support scripted procedures compatible with stored procedures deployed on other platforms. External stored procedures also need a language environment in which to run, provided by a work load manager. When an external stores procedure is called, the caller's DB2 thread gets suspended while it is switched from the caller's task (SRB or TCB) to the TCB of the external program.

Beginning on DB2 9 for z/OS, SQL PL procedures may be deployed as native stored procedures. This means that they run as interpreted code inside a DB2 address space. When a native stored procedure is called, there is no thread suspension and task switching. DB2 just switches from the callers package to the stored procedure package.
External Stored Procedures

This diagram depicts processing with external stored procedures in a distributed environment. In this example, a group of SQL statements are processed with a single send/receive operation.

Distributed applications consist of at least 2 parts: client programs (which may be calling programs) that reside on the client system or application server, and server programs (which may be stored procedures) that reside on the database server.

On an IBM mainframe server environment, there must be an address space running TCP/IP to interface to the network and Unix System services to provide security functions in the TCP/IP environment. The DB2 Distributed address space must be active to process remote requests and the Database Services address space must be active to process SQL.
Native SQL PL stored procedures were introduced in DB2 9 for z/OS. These stored procedures are syntactically identical to the external SQL PL procedures that were possible in Version 8 and prior. The benefit of native procedures is that they run inside the DBM1 address space that processes all SQL requests. No calls to a WLM managed address space are needed. Currently only procedures coded in SQL PL are eligible to run natively.

Native SQL PL procedures are also eligible to run on zIIP specialty processors when called from a distributed environment which may decrease the cost of these procedures. zIIPs are specialty engines that can handle a percentage of eligible workload's CPU which will not factor into mainframe software prices.

Native stored procedures are zIIP eligible if it is invoked via a remote call through DDF of DB2.

The storage used in a native stored procedure is above the bar.
Creating a New Stored Procedure

You can use the wizard to help create a basic ‘Shell’ program. By selecting the kind of stored procedure you are interested in creating will determine the ‘Shell’ code setup for the new procedure.
Another way to create a new stored procedure is to copy from an already existing procedure from the Data Source Explorer area. Copy / Paste, or Drag / Drop into the Stored Procedure folder in the Data Project Explorer area. After copying into the folder, double click on the stored procedure to get the edit window up, then rename the procedure in the ‘Create’ portion of the code and save. Data Studio will rename the copied stored procedure. See the following pages.
Drag and Drop

Highlight the procedure within the database and schema in the Data Source Explorer. Drag and Drop it into your Data Project, double click on it to edit, rename and save.
Deploying a Stored Procedure

If it is a new stored procedure, the deploy function will create it the first time. If it is an updated procedure, the deploy executes an alter statement.
Running a Stored Procedure

You can run any stored procedure by navigating to it in the “Schema” folder in the Data Source Explorer. The stored procedure you run does not need to be one that was developed using IBM Data Studio. It can even be a COBOL (or any other kind of external) stored procedure. This allows you to unit test stored procedures you have written in any language.
Running a Stored Procedure

When you run a procedure you will be prompted to enter any input parameters in a dialog box.
Running a Stored Procedure

The status box will be populated with messages about the execution of the procedure. If errors are returned to the caller they will be shown here.
Running a Stored Procedure

The input and output values of each parameter are shown in the “Parameters” tab.
If the procedure returns any result sets there will be a “Result” tab for each. IBM Data Studio detects any result sets that are returned by the stored procedure and fetches the rows for display.