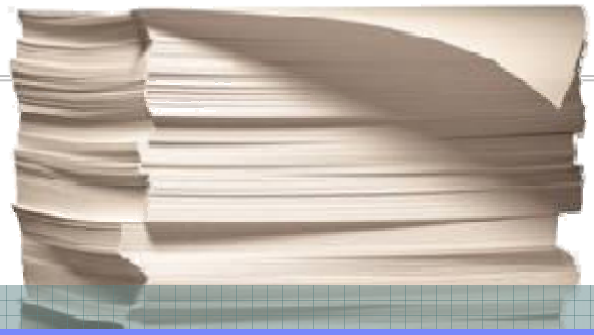
IBM Information Management

DB2 Warehouse Tooling

Eliminating the Barriers



Phil Downey
DB2 Warehouse Technical Marketing

December 2007

© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- *Warehouse Tooling Components*
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*

© 2007 IBM Corporation

What is a Data Warehouse Package ?

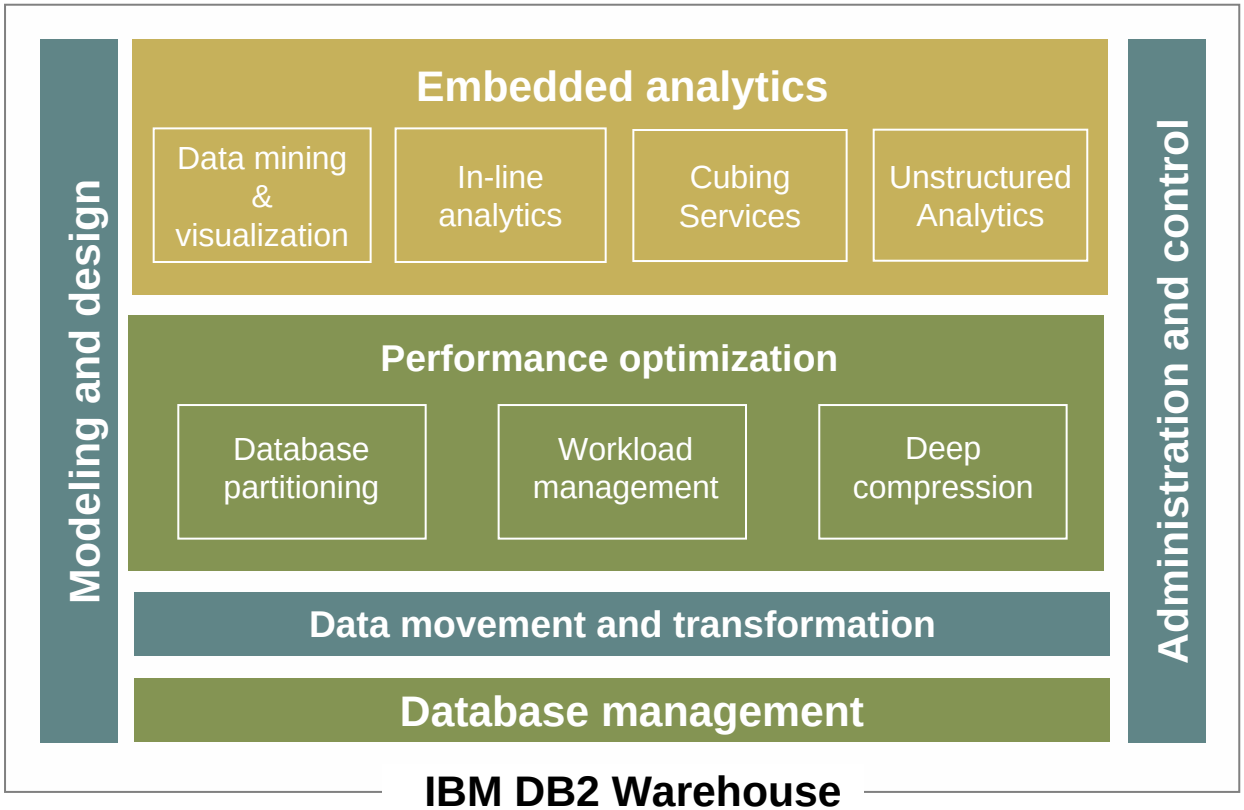
Today there is much Focus on the Platform Components of warehouse environments

Analytics is becoming more of a focal point however there is not as much of a focus for Scalability and Integration of these tools as there should be

© 2007 IBM Corporation

IBM DB2 Warehouse 9.5

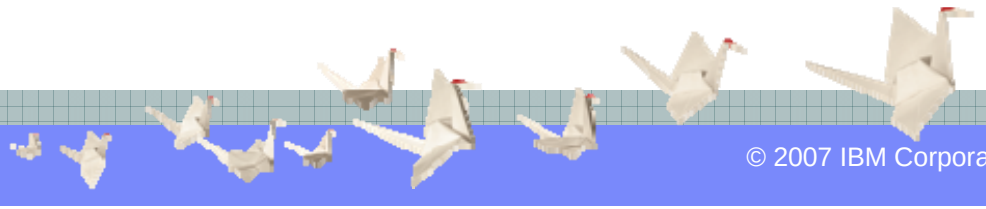
There is more to a Warehouse Environment, right ?



© 2007 IBM Corporation

Agenda

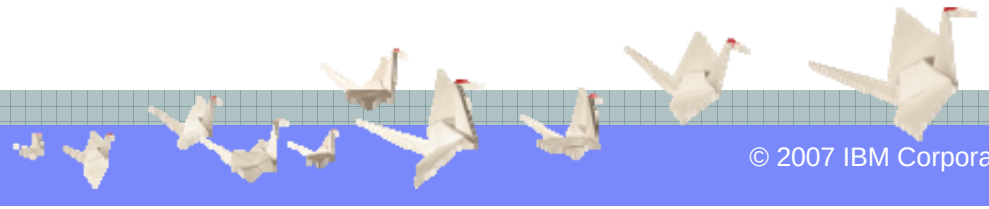
- *Warehouse Tooling*
  - ***What Is it ?***
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- *Warehouse Tooling Components*
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*



© 2007 IBM Corporation

What is DB2 Warehouse Tooling ?

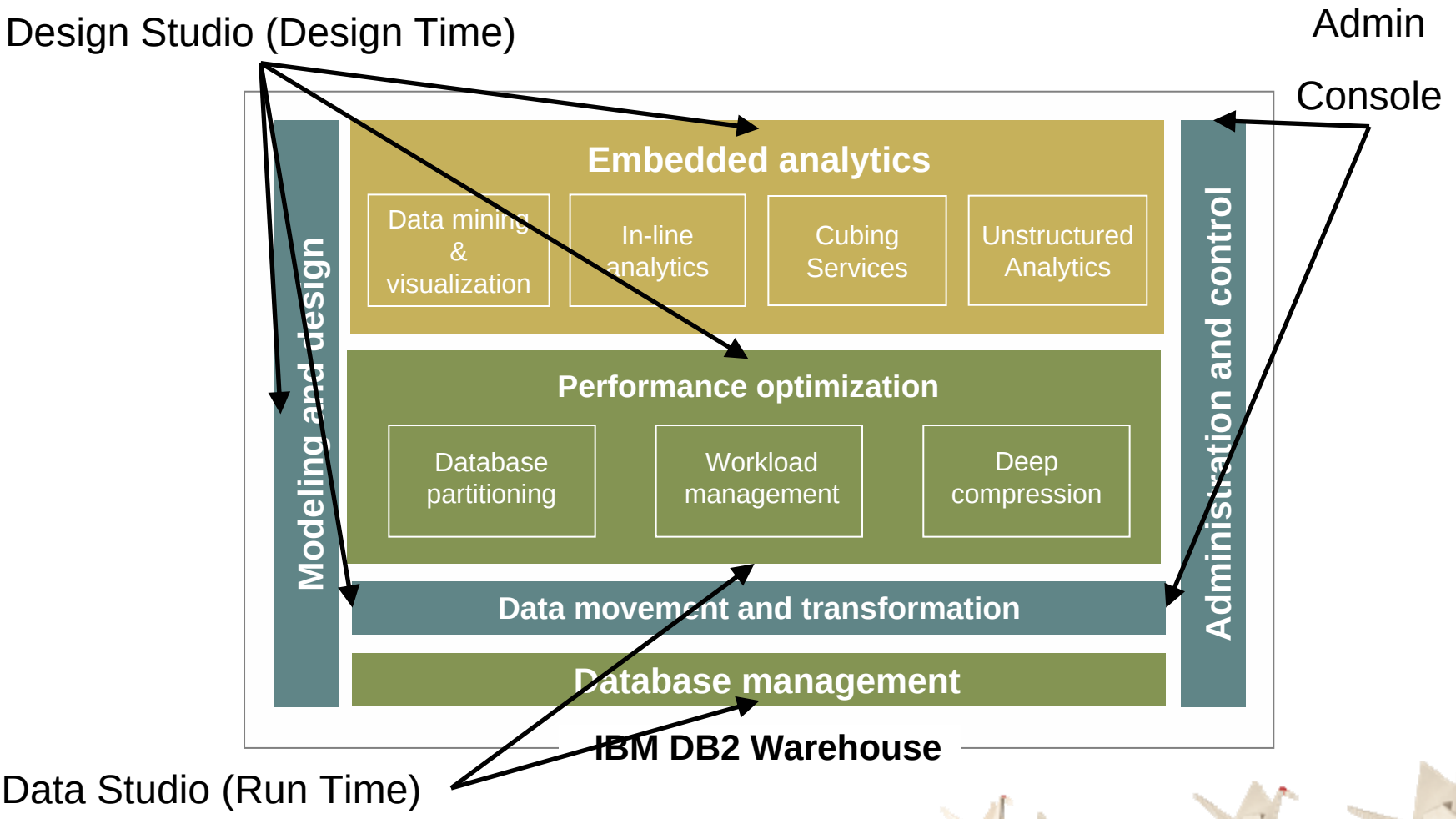
- DB2 Warehouse tooling is made up of three key components
  - Database management & Application Development
    - Data Studio
  - Data Warehouse Development
    - Design Studio
  - Data Warehouse Management
    - Administration Console



© 2007 IBM Corporation

IBM DB2 Warehouse 9.5

What do these Tools help you manage ?



© 2007 IBM Corporation

IBM Data Studio

Data Life Cycle

Design	Develop	Deploy	Manage	Govern
Logical Modeling Physical Modeling Integration Modeling	Java .NET Web Services PHP RUBY COBOL	Stored Procedures SQL XQuery User Defined Functions	Administration Configuration Performance Management Backup & Recovery	Data Auditing Data Archiving Data Masking Data Encryption Security Access Security Analysis
Data Modeling	***Application Development***	***Database Development***	***Database Management***	***Data Governance***
Data Architect	*Application Developer*	*Database Developer*	*Database Administrator*	*Security Administrator*

pureQuery

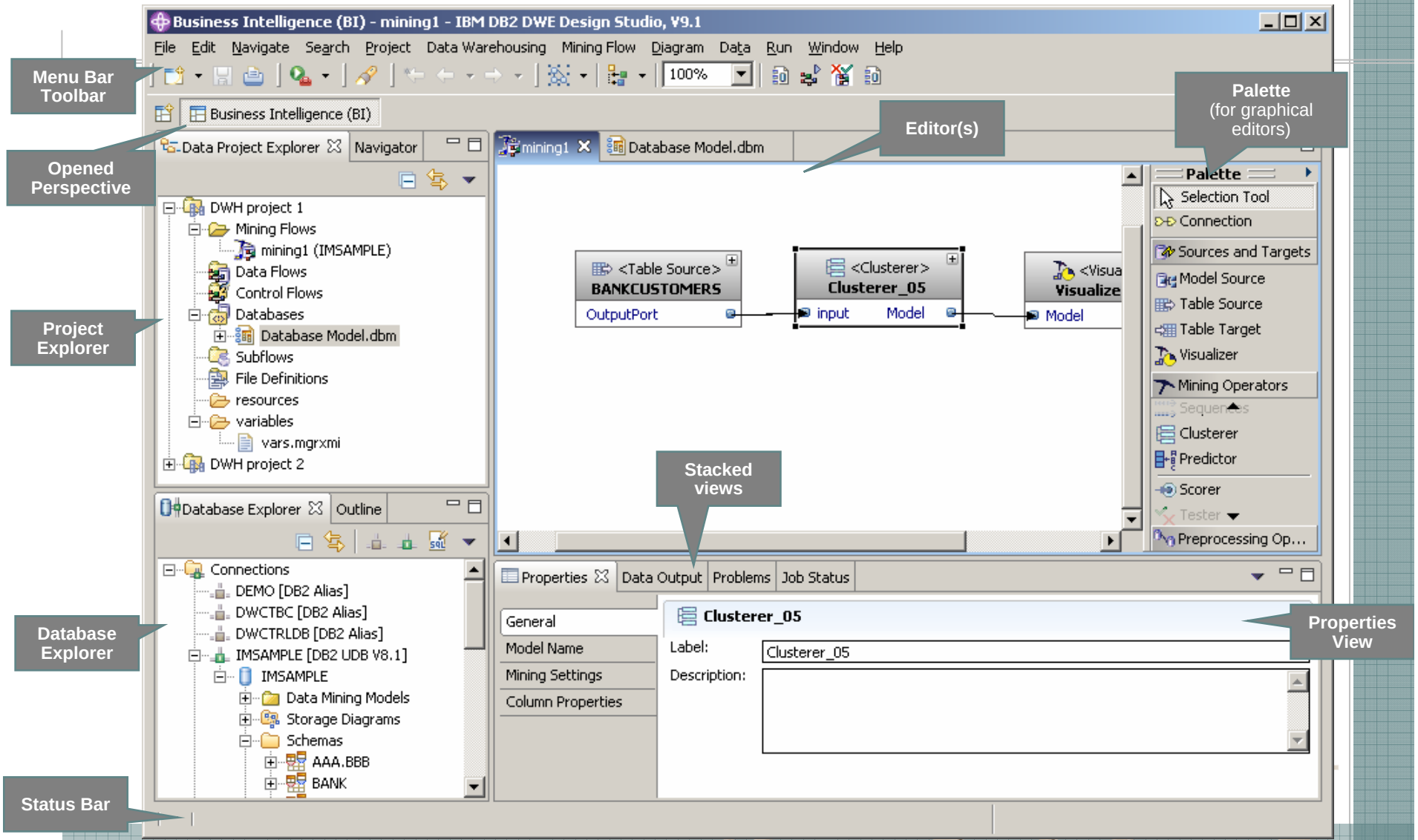
© 2007 IBM Corporation

DB2 Warehouse Design Studio

- Based on the same platform as Data Studio
- Provides all the capabilities required to Build and Deploy Warehouse Applications
  - **Data Modeling & Development**
    - ER Modeling
    - Database Development (stored Procedures, Functions)
  - **Data Transformation**
    - Embedded Transformations
    - External Transformation Interface
  - **OLAP**
    - Supporting ROLAP and OLAP Interfaces
  - **Unstructured Analytics**
    - Intra Database Text Analysis and Classification
  - **Mining**
    - In-Line Data Mining
  - **Workload Management**
    - Dynamically Enabled Workload Management
  - **Embedded Analytics Development**

© 2007 IBM Corporation

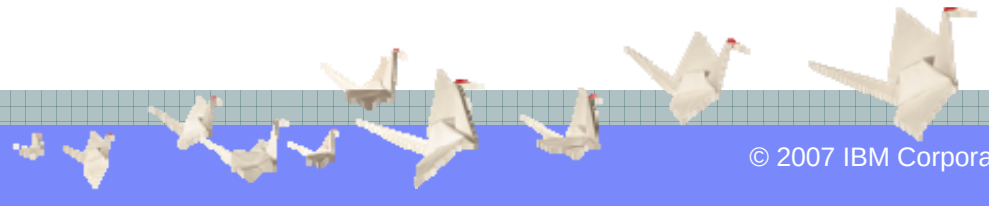
Design Studio



© 2007 IBM Corporation

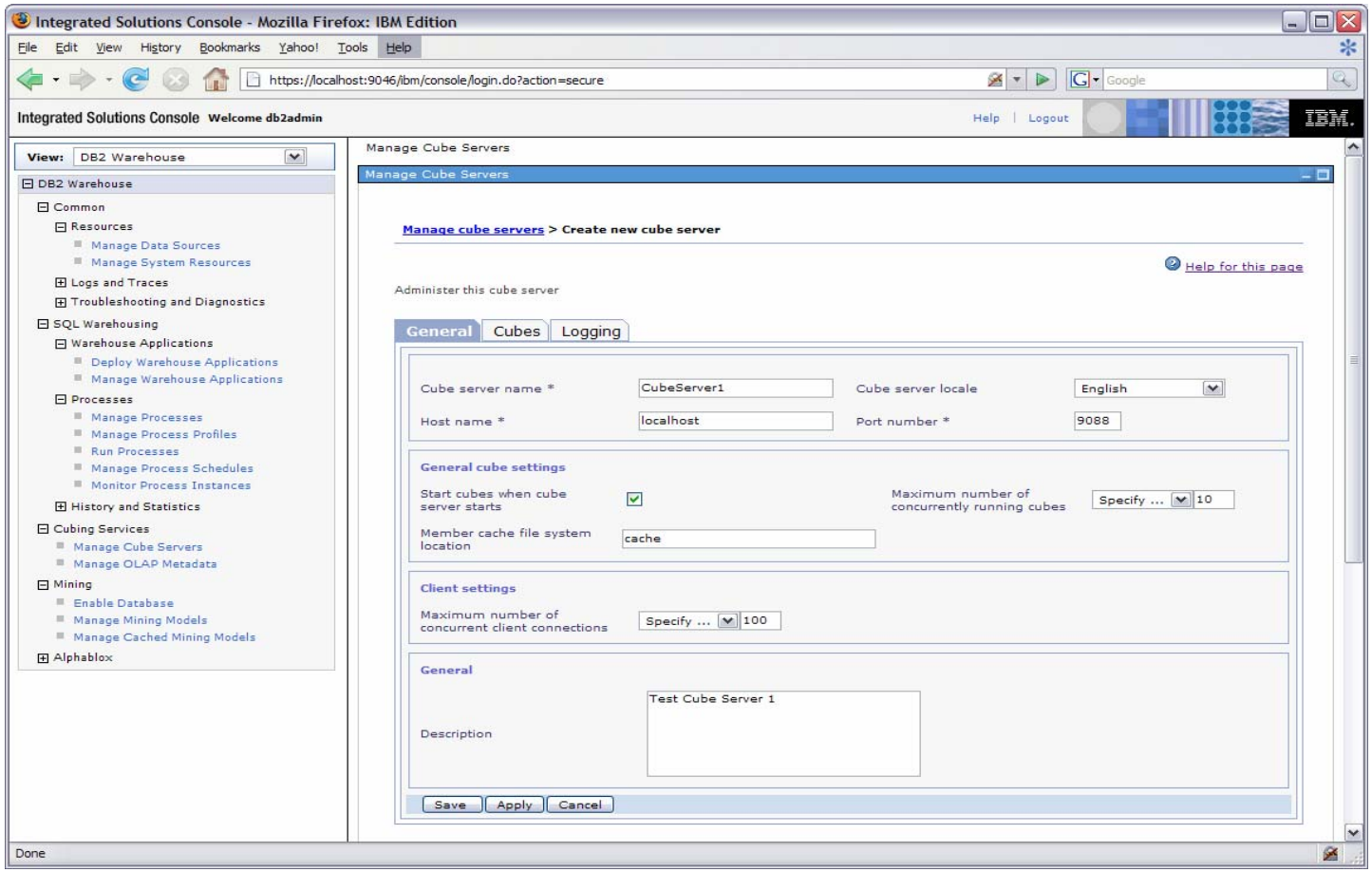
Administration Console

- Built Using IBM Systems Console
  - Use DB2W Admin and Websphere Console with one login
- Warehouse Administration
  - Warehouse Function Deployment e.g
    - OLAP
    - Mining
    - WLM
  - Warehouse Transformation Deployment
  - Warehouse Monitoring e.g.
    - SQW performance
  - Warehouse Deployed object Catalog
  - Cubing Services Engine management



© 2007 IBM Corporation

Administration Console



© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - ***What can it do for me ?***
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- *Warehouse Tooling Components*
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*

© 2007 IBM Corporation

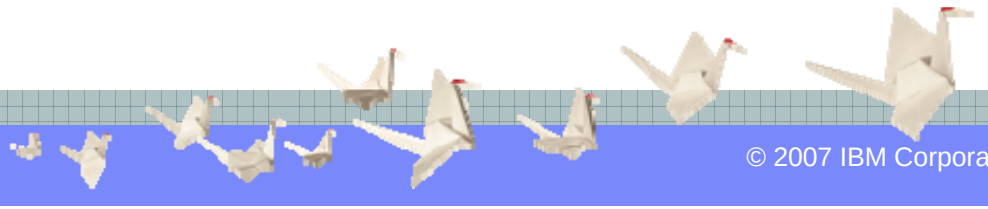
What Can It do for me ?

Provide one cohesive integrated Tool for Developing your Warehouse

- Design, Develop and Deploy
- Full Life cycle approach
- Easy Packaging of BI Applications
  - Great for BI applications that are part of a software package or Consulting tool set

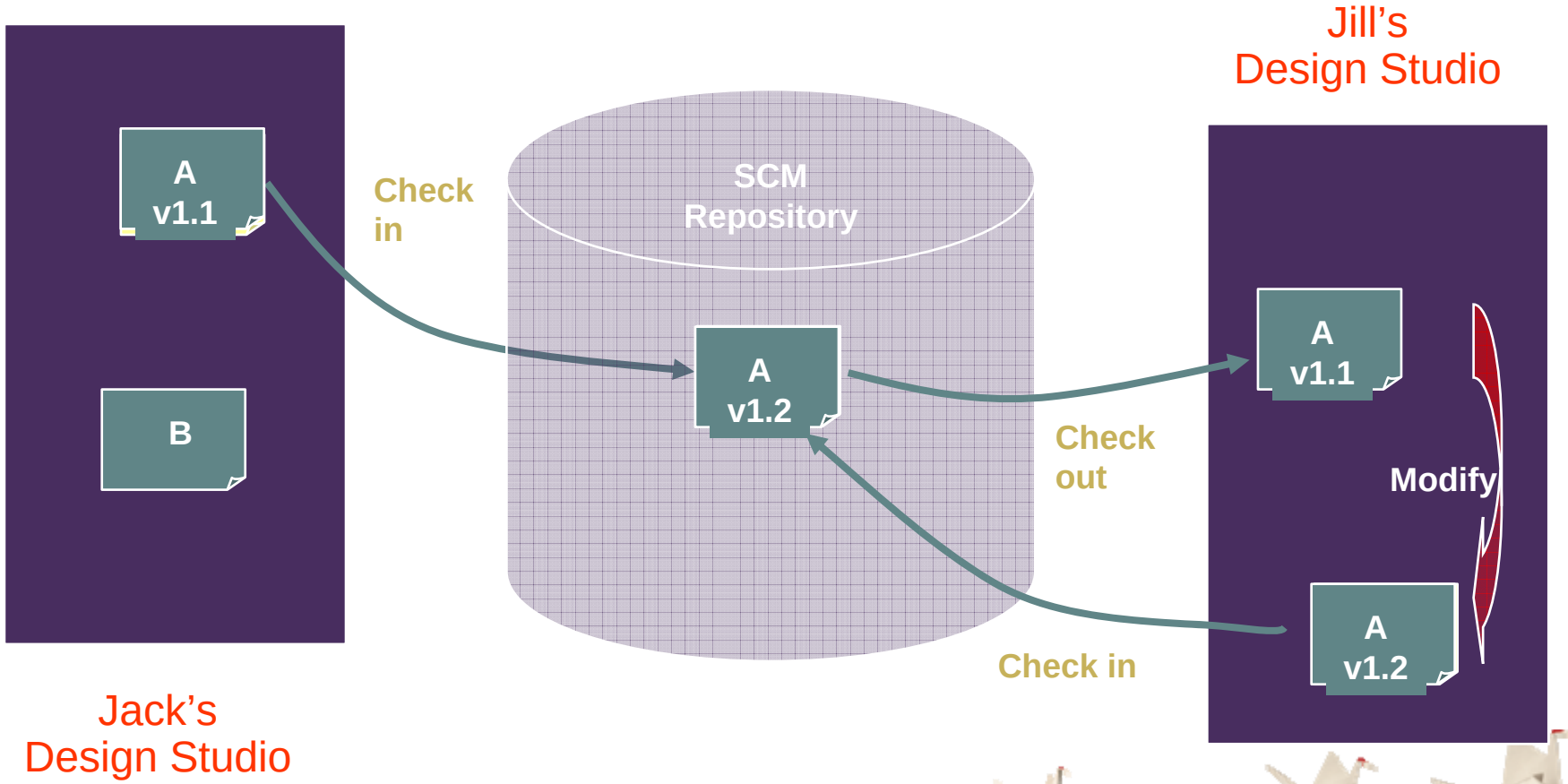
- Delivers Teaming and Versioning Capabilities
  - Using CVS compatible repositories
  - Enables Teams to share project work
  - Enables Teams to work on multiple version
    - Excellent for Distributed warehouse deployments
      (e.g. Head office and Branch environments)
    - Excellent for ensuring consistency between New and in place Warehouse Deployments
      (e.g. changes in existing Dimension or Data transformation jobs can be also integrated easily into new development chains)
    - Enables Greater audit ability
      Compliance

- Provides single point of Warehouse Administration for deployed functionality



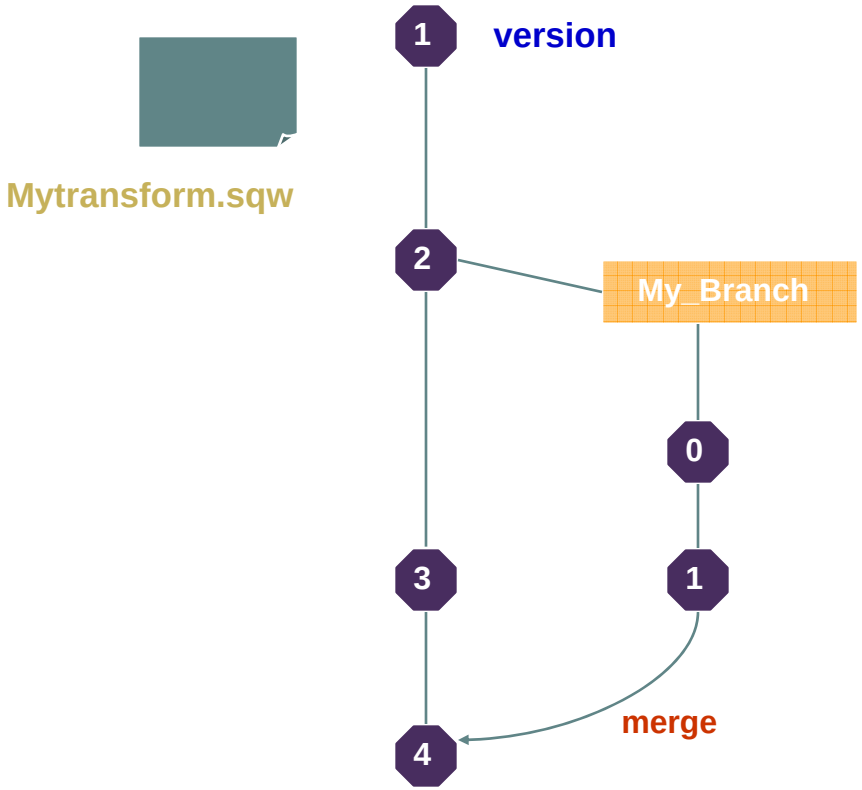
© 2007 IBM Corporation

Teaming and Versioning



© 2007 IBM Corporation

Branching and Merging



© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - ***Can I use it with externally Based Tooling ?***
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- *Warehouse Tooling Components*
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*

© 2007 IBM Corporation

Can I use it with Externally Based Tooling ?

- **Absolutely !**
- **Transformations**
  - Design Studio can schedule external Jobs
  - Design Studio can View Data Stage Servers and schedule Data Stage jobs within a Warehouse Transformation
  - Design Studio can export workloads to DataStage
- **Modeling**
  - Design Studio can import Models from
    - Rational Data Architect
    - ErWin
- **OLAP**
  - Meta Data exchange with BI Tools
  - Coming Q1 2008 ODBO Driver for OLAP
- **Mining**
  - Ability to integrate with SAS and SPSS Data Mining Models
- V9.1.X Integrates with current releases of RDA and Other Rational Tooling
  - 9.5 moving forward will be compatible with New Versions of other tooling

© 2007 IBM Corporation

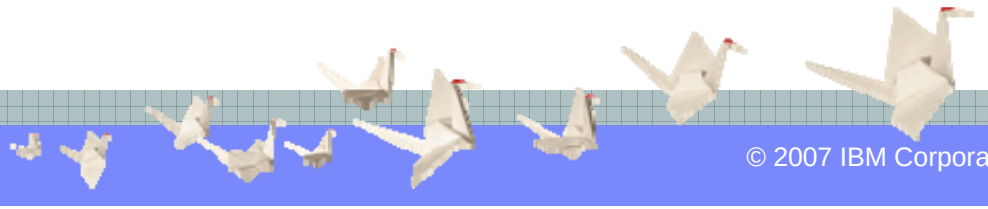
Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- ***Operational Advantages***
  - *Design Studio*
  - *Administration Console*
- *Warehouse Tooling Components*
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*

© 2007 IBM Corporation

Operational Advantages: Design Studio

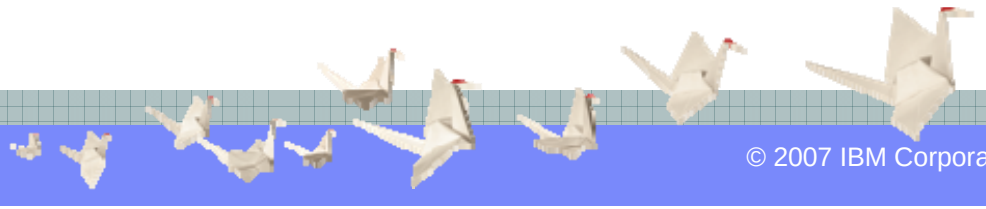
- No requirement to jump between tooling when delivering Warehouse Capabilities
- Integrates with other BI Tooling
  - Mining Workbenches (SPSS,SAS)
  - ER tools
  - DataStage
  - Cognos/BO
- Drives Team Based Development
  - Most Warehouses are developed by teams
  - Enhanced communication
  - Not just a first in first served model
- Enables Warehouse Applications to be packaged and deployed as one project
  - Enables Project based versioning
  - Enables merging and concurrent streams of project development
    - Important for Software Package providers
    - Important for those developing consolidated Warehouse environments
      "Program" based deployment
- Better Compliance & Auditability
  - Versioning
  - Change management
  - Impact analysis



© 2007 IBM Corporation

Operational Advantages: Administration Console

- Single Catalog of deployed components
- Single point of Deployment
  - Enables BI developers to play a greater role in Deployment
  - Better control and change management of Warehouse Components
- Single point of monitoring of Warehouse Functions
  - SQW activity
  - Cubing Services



© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- ***Warehouse Tooling Components***
  - ***Modeling (RDA)***
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*

© 2007 IBM Corporation

Rational Data Architect Plug-in functions in DB2 Warehouse

Discover, model, visualize, relate, and develop diverse and distributed data assets with one tool

Eclipse based framework, common models for integration across the lifecycle

Reverse engineer
Navigation
Discovery
Visualization

Rational. Data Architect

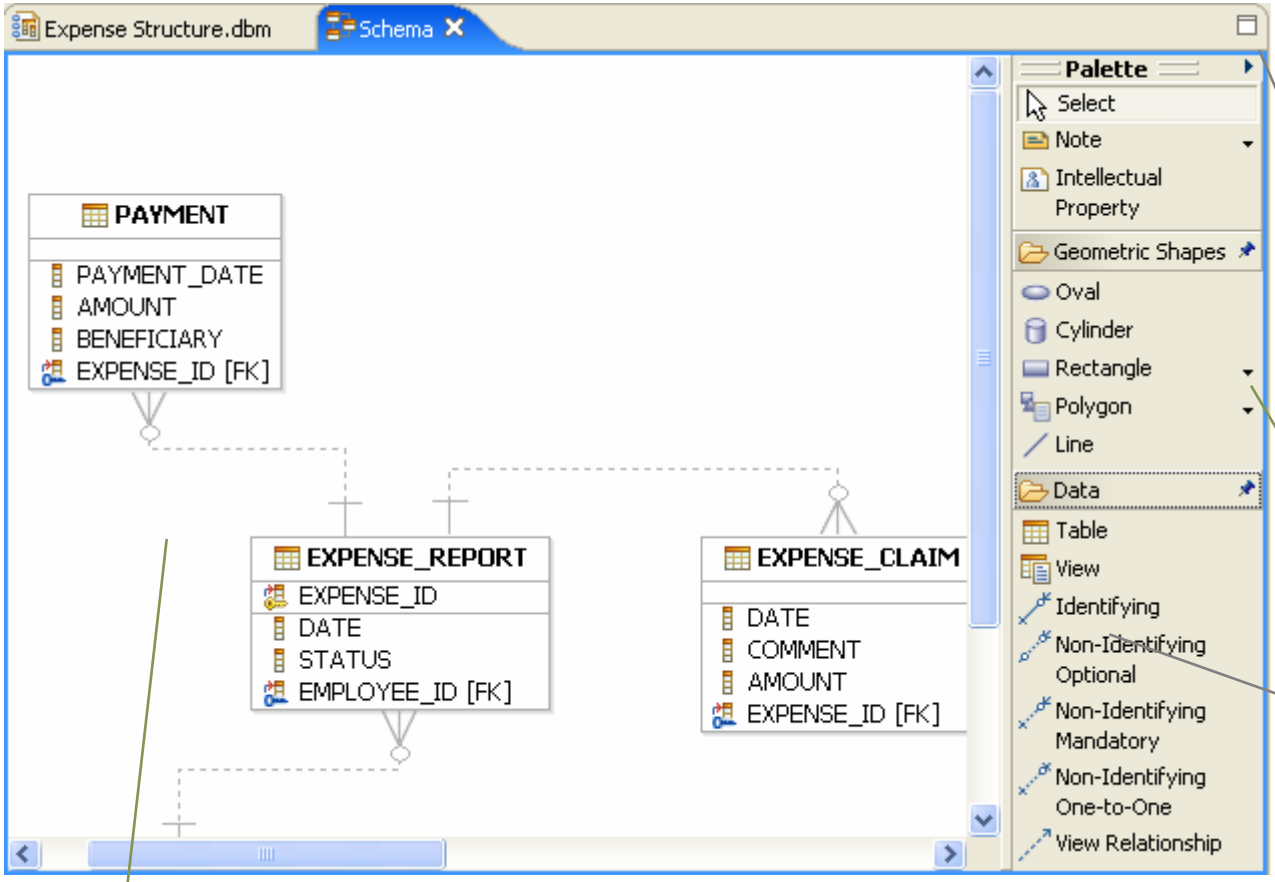
Physical Design
Rules and Model Validation

Lifecycle management
Compare and Sync
Impact Analysis

Edit
Execute
Debug
Deploy
Team Integration

© 2007 IBM Corporation

Physical data modeling – visual representation



Hide/show palette

Geometric shapes

Palette – select the element to create on diagram

Drawing area

© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- ***Warehouse Tooling Components***
  - *Modeling (RDA)*
  - ***Transformation (SQW)***
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*

© 2007 IBM Corporation

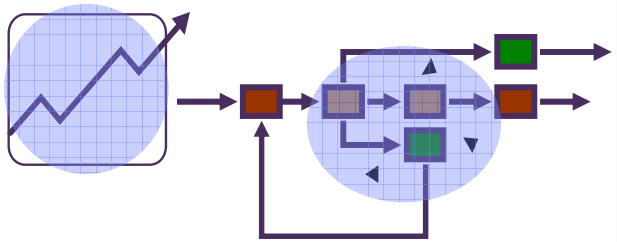
Embedded Data Movement and Transformation
Enhanced SQL Warehousing Tool (SQW)

DB2 SQL Warehousing Tool provides Data Movement and Transformation capabilities based on the DB2 9.5 engine. Transformations and Data Movement jobs are developed in a fully integrated graphical interface within the DB2 Design studio.

Benefits

Easy to Use

- *Graphically build complex Transformations within DB2*
- *Ability to Natively Connect to and source data from Non DB2 sources*
- *Advanced Workflow Control and Scheduling*



Integration

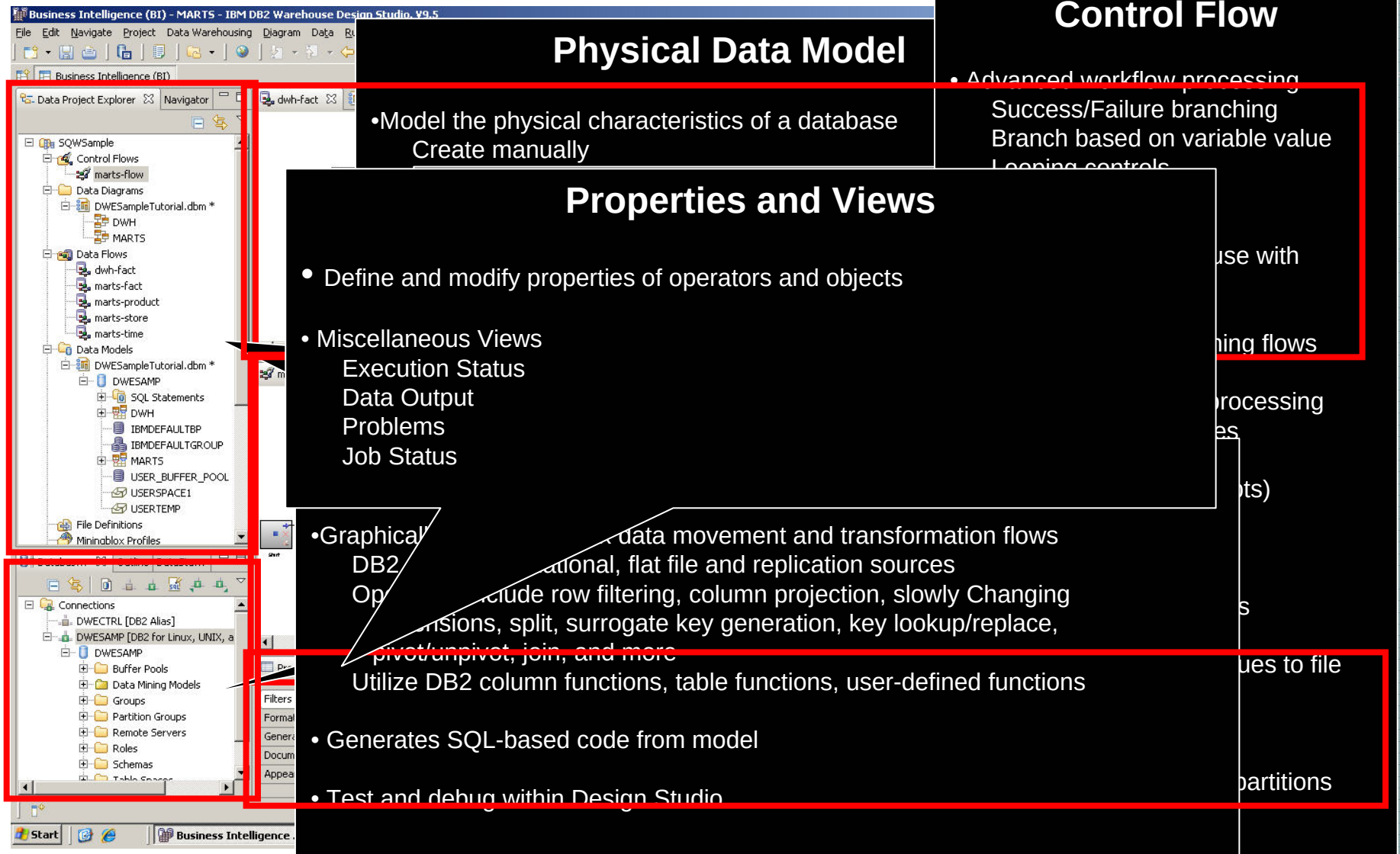
- *Automate Data Mining workloads*
- *Integration with Information Server*
  - *schedule, execute, monitor and control Datastage Jobs*
  - *Ability to integrate Datastage Work streams*

Compliance

- Ability to add Version Management
- Job Monitoring

© 2007 IBM Corporation

SQL Warehousing Tool (SQW)



Physical Data Model

•Model the physical characteristics of a database
Create manually

Control Flow

• Advanced workflow processing
Success/Failure branching
Branch based on variable value
Looping controls

Properties and Views

- Define and modify properties of operators and objects

- Miscellaneous Views
  - Execution Status
  - Data Output
  - Problems
  - Job Status

•Graphical data movement and transformation flows
DB2 ational, flat file and replication sources
Op clude row filtering, column projection, slowly Changing
nsions, split, surrogate key generation, key lookup/replace,
pivot/unpivot, join, and more
Utilize DB2 column functions, table functions, user-defined functions

• Generates SQL-based code from model

• Test and debug within Design Studio

use with

ning flows

rocessing
es

ts)

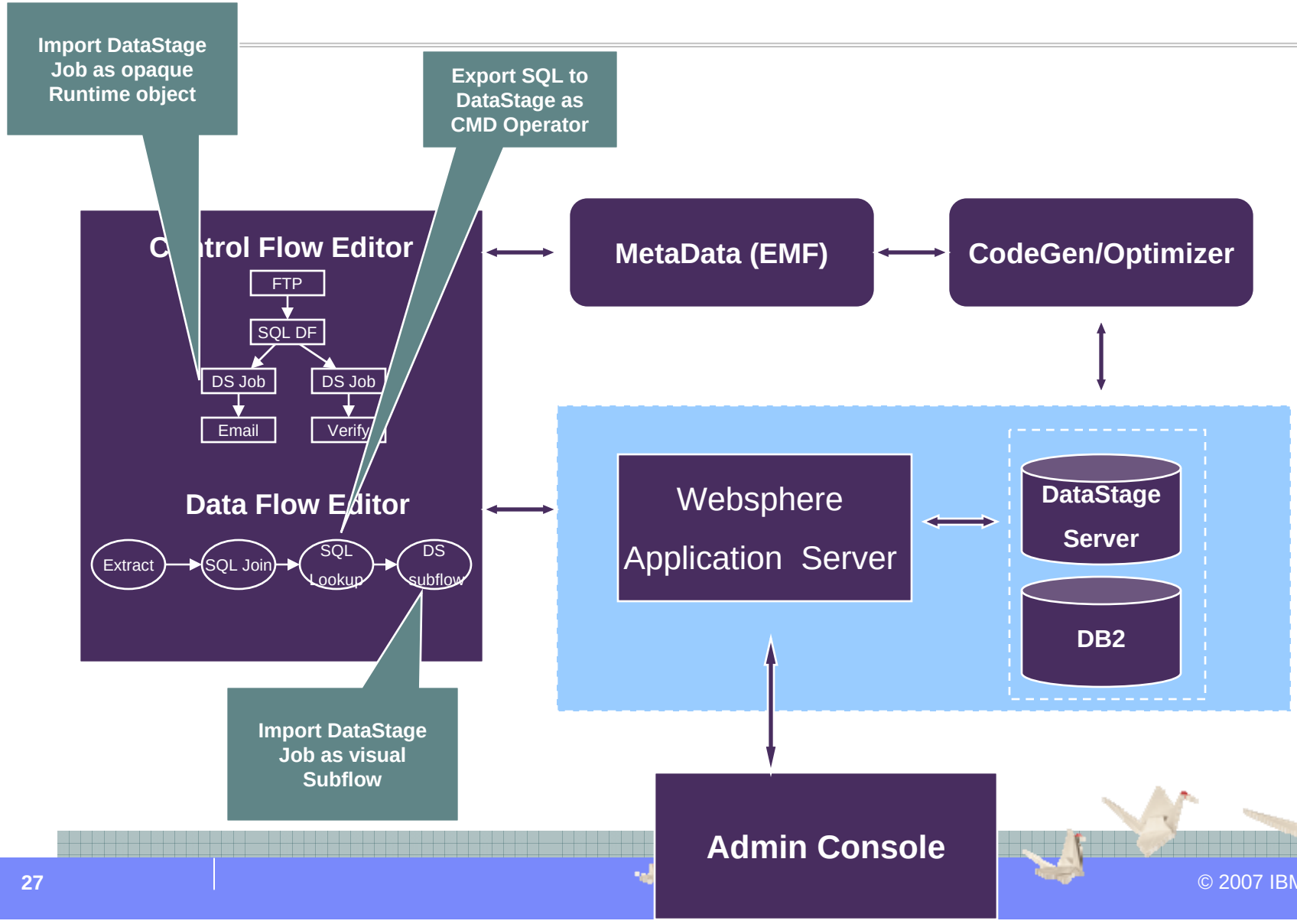
s

ues to file

partitions

© 2007 IBM Corporation

SQW with DataStage: Integration points



© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- ***Warehouse Tooling Components***
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - ***Data Mining***
  - *Integration…..*

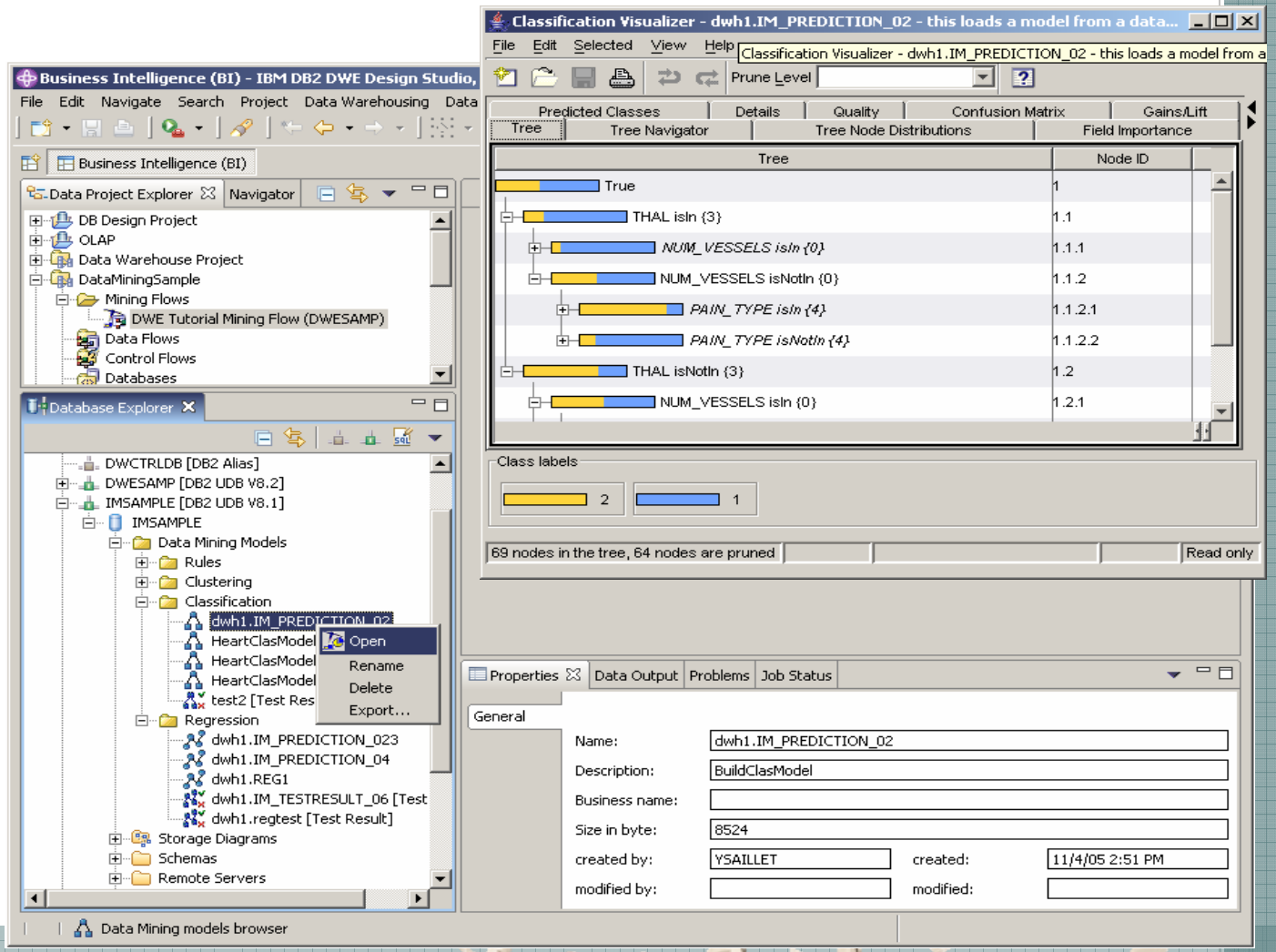
© 2007 IBM Corporation

Data Mining

- Manage your data mining models in the database from the **Database Explorer**

Key Features:

- View the list of the loaded models
- See the information about the models in the Properties view
- Import/Export models from/to the file system
- Open the model with IM-Visualization



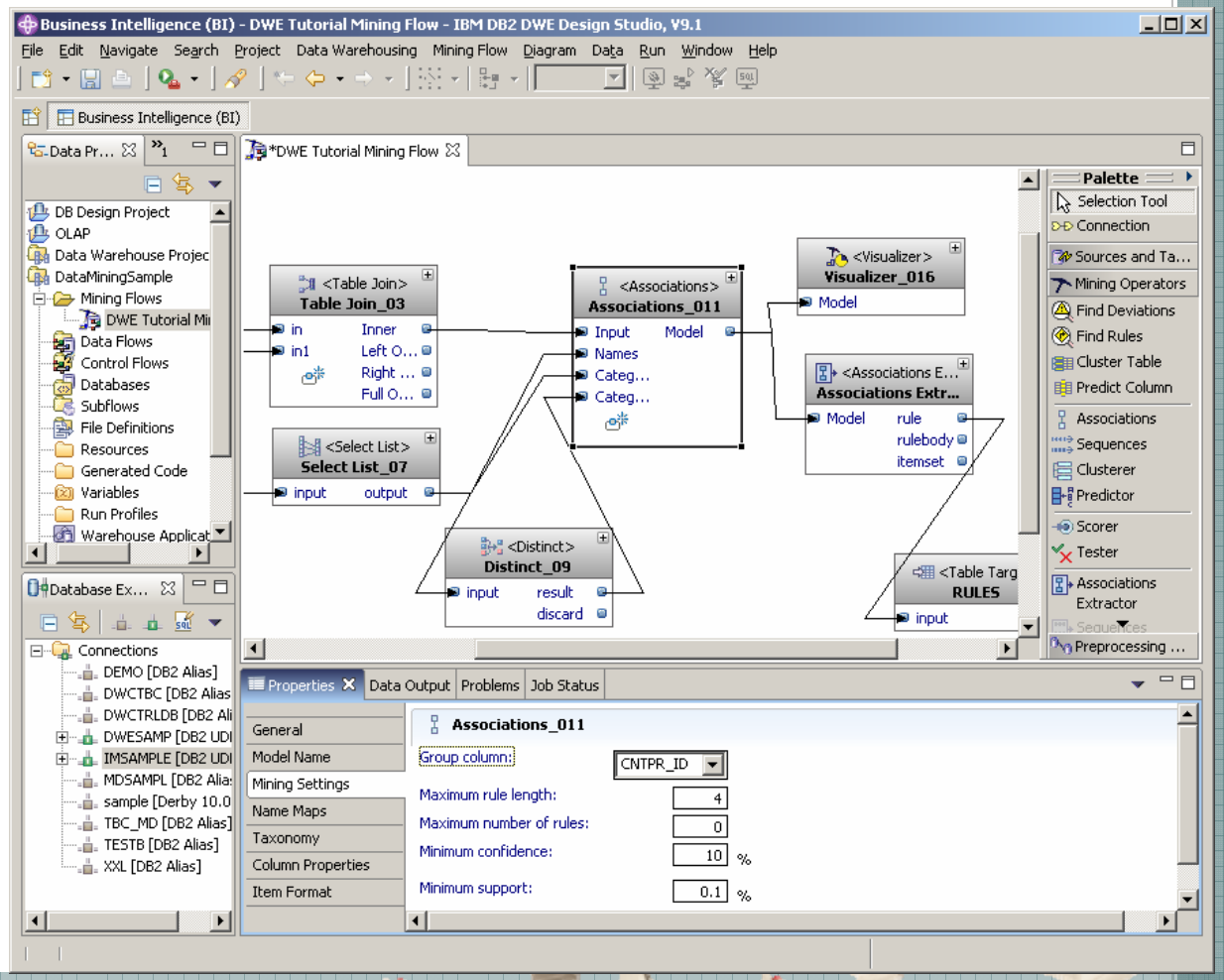
© 2007 IBM Corporation

Data Mining

- Interactive creation of **Mining Flows** to build or apply data mining models
  - Create a "**Data Warehouse Project**" and then a "**Mining Flow**"

Key Features:

- Same editor as for SQLW data flows.
- When editing mining flows, mining-specific operators are available in the palette
- The SQLW data transformation operators are also available in mining flow
- Enhanced interactivity to facilitate the iterative creation of mining models
- Mining flows can be embedded in DWH control flow and be deployed



© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- ***Warehouse Tooling Components***
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - ***Unstructured Analytics***
  - *Data Mining*
  - *Integration…..*

© 2007 IBM Corporation

Unstructured Analytics

Extracts structured information out of previously untapped business text through the ability to apply linguistic technology that extracts targeted information from within unstructured text.

Benefits

- *Provides Additional Data Source for Analysis*
  - *Unstructured data represents up to 80% of data within an organisation*
- *Enables you to target specific information hidden within text*
  - *Complaints/problems*
  - *Claimants*
  - *Involved Parties*
  - *Products*

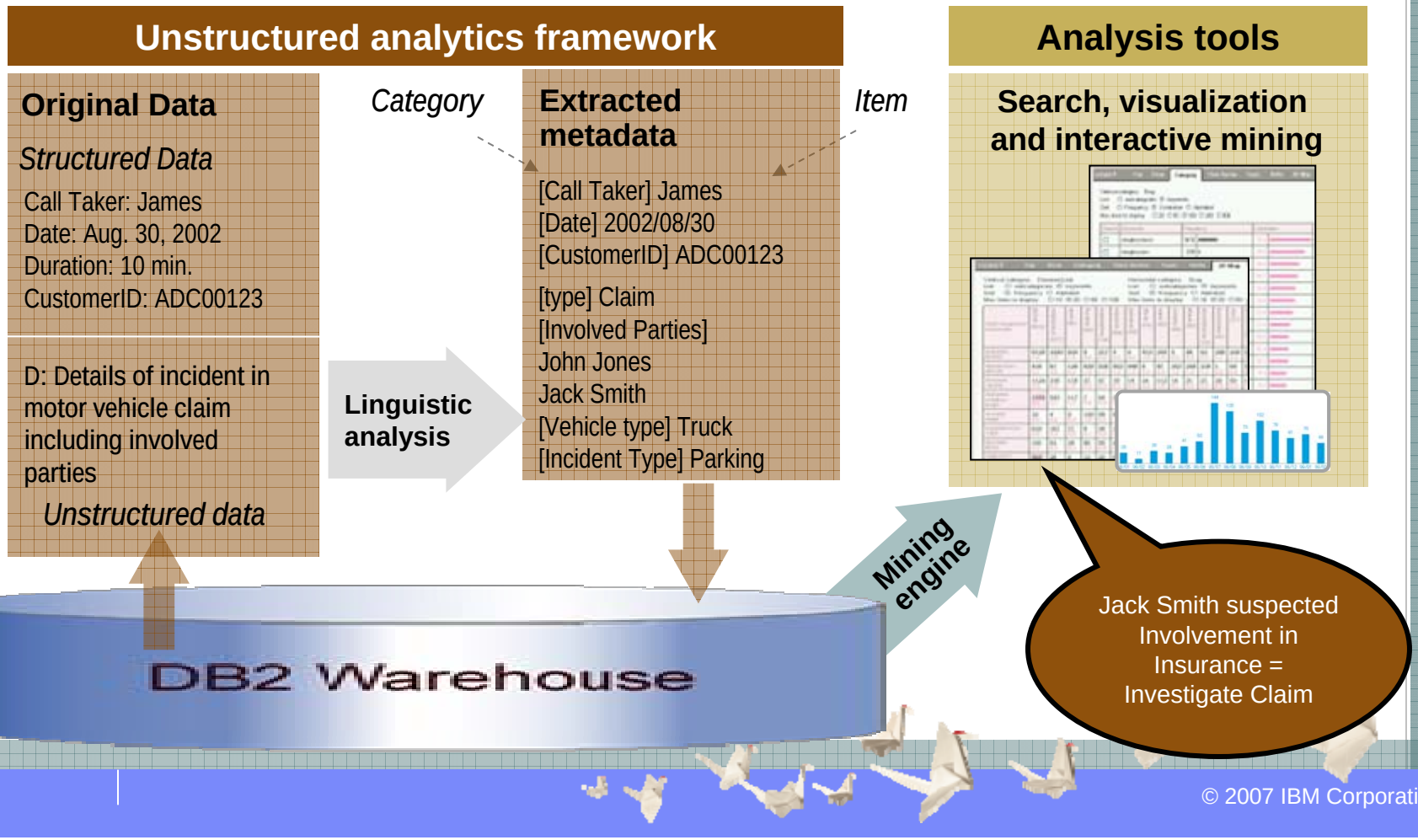
Business Value

- *Early warning on Customer attrition*
- *Fraud Detection*
- *Product Defect Analysis*
- *Better customer profiling*

© 2007 IBM Corporation

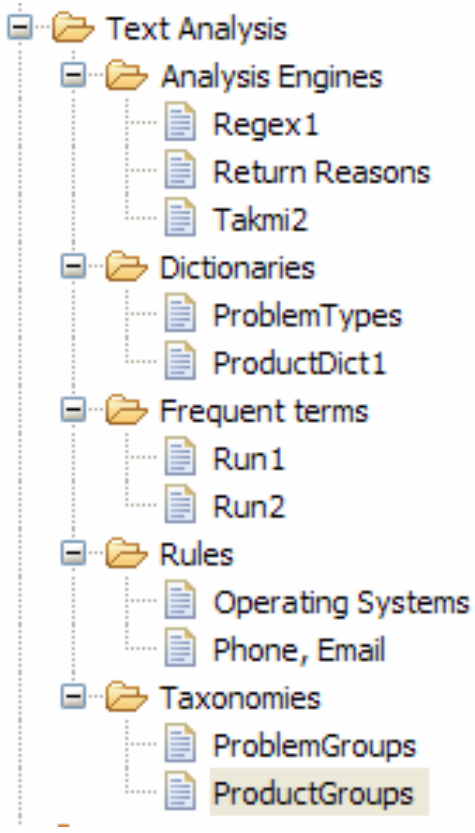
Unstructured Analytics in DB2 Warehouse Enterprise Edition

- Rich analysis interface for combining structured and unstructured data
- Combines search, text analytics and data visualization



© 2007 IBM Corporation

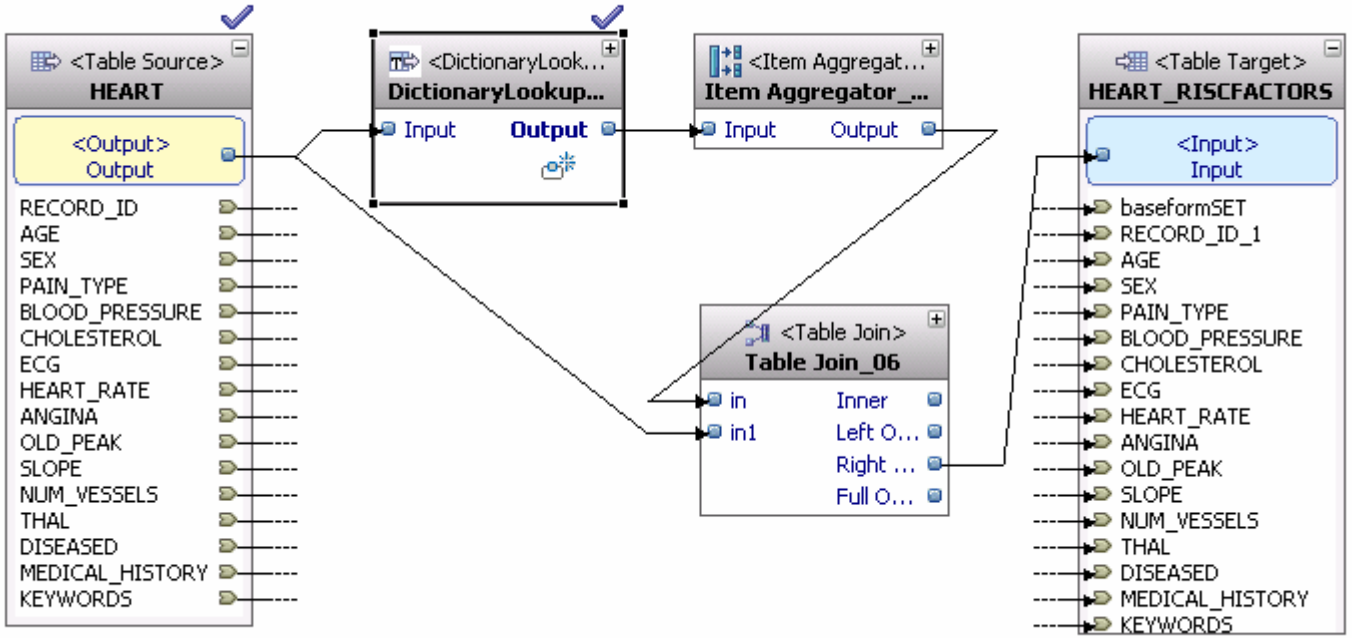
Data Warehouse Design Studio task overview



- Data understanding
  - –View text columns in database, text statistics
  - –frequent terms analysis
- Dictionary editor
  - –Create named entity dictionary from frequent terms or import from table
- Rule editor
  - –Create regular expression rule set
- Taxonomy editor
  - –Create hierarchical classification for concepts extracted from text
  - –Create dimension table from taxonomy for multidimensional analysis
- Analysis Engines
  - –Import pre-configured analysis engine (from IBM research, business partners, Omnifind solutions)
- Flow editor with text specific operators
  - –Create transformation flow that reads text columns in source tables, applies a dictionary lookup or rule lookup and writes the results to target table
  - –TextAnalyzer, Dictionary Lookup, ItemAggregator and Regular Expression Lookup operators

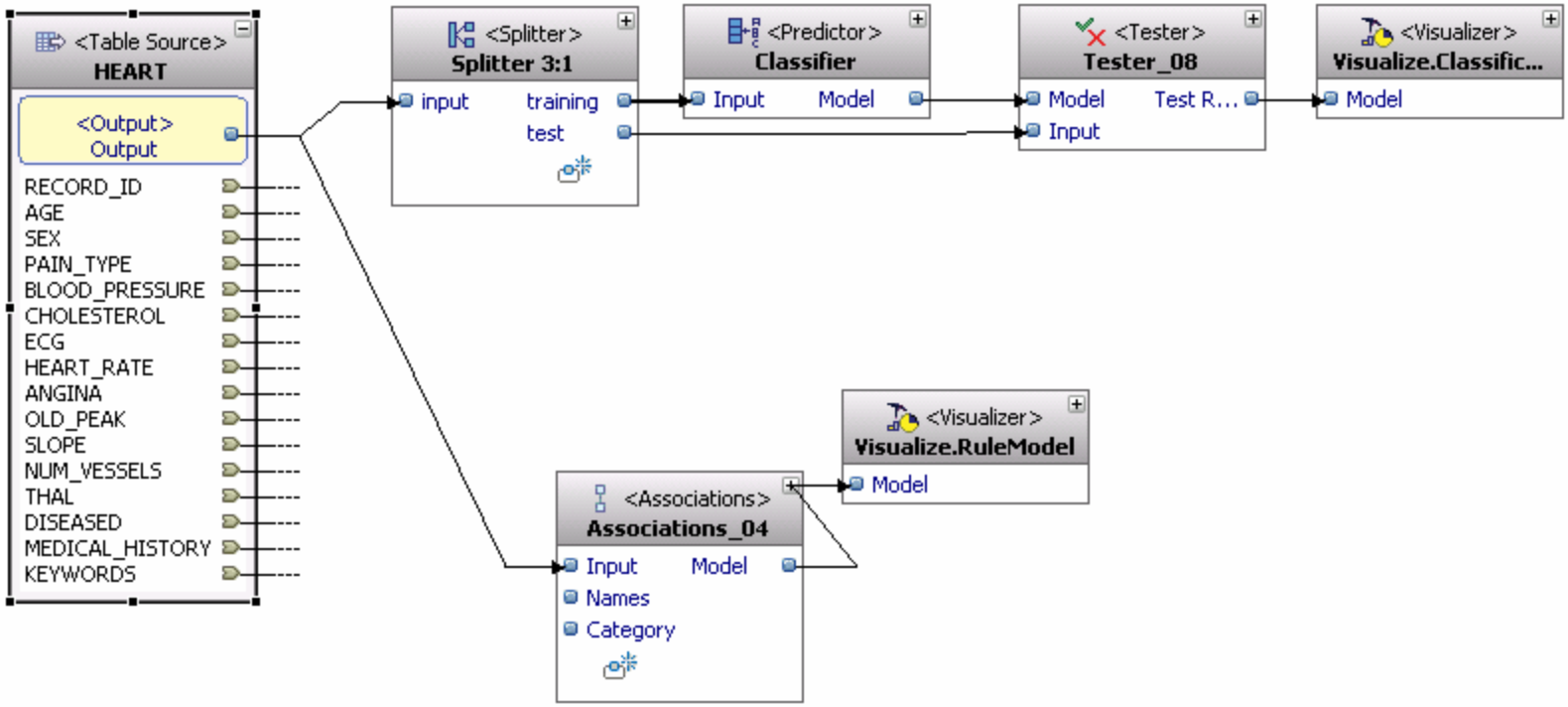
© 2007 IBM Corporation

Extracting symptoms using Dictionary Lookup



© 2007 IBM Corporation

Building the associations and prediction model



© 2007 IBM Corporation

The Association Rules model applied to Extracted Text Data

Visible rules:

Body	Support	Confidence	Lift	▲ Items...
[NUM_VESSELS=3]	6.2500%	83.3333%	1.94	1
[physical inactivity]	23.7500%	78.0822%	1.82	1
[NUM_VESSELS=2]	9.5833%	76.6667%	1.79	1
[alcoholic]	10.0000%	75.0000%	1.75	1
[obesity]	23.7500%	74.0260%	1.72	1
[THAL=7]	28.3333%	73.9130%	1.72	1
[ANGINA=y]	24.5833%	73.7500%	1.72	1
[HEART_RATE < 130 {=1/4}]	14.1667%	72.3404%	1.69	1
[smoker]	17.0833%	68.3333%	1.59	1
[PAIN_TYPE=4]	30.8333%	67.2727%	1.57	1
[SLOPE=2]	28.7500%	63.8889%	1.49	1
[NUM_VESSELS=1]	13.3333%	61.5385%	1.43	1
[AGE >= 55 AND < 65 {=3/4}]	21.6667%	59.0909%	1.38	1
[HEART_RATE >= 130 AND < 150 {=2/4}]	13.3333%	58.1818%	1.36	1
[short of breath]	17.0833%	57.7465%	1.35	1
[CHOLESTEROL >= 250 AND < 300 {=3/4}]	15.8333%	53.5211%	1.25	1
[SEX=m]	36.6667%	53.0120%	1.24	1
[physical inactivity]+[HEART_RATE < 130 {=1/4}]	8.7500%	100.0000%	2.33	2
[NUM_VESSELS=2]+[THAL=7]	6.2500%	100.0000%	2.33	2
[NUM_VESSELS=2]+[ANGINA=y]	5.8333%	100.0000%	2.33	2
[physical inactivity]+[ANGINA=y]	12.9167%	96.8750%	2.26	2
[HEART_RATE >= 130 AND < 150 {=2/4}]+[obesity]	8.7500%	95.4545%	2.22	2
[physical inactivity]+[BLOOD_PRESSURE >= 110 AND < 130 {=2/4}]	8.7500%	95.4545%	2.22	2
[alcoholic]+[SLOPE=2]	7.9167%	95.0000%	2.21	2
[physical inactivity]+[THAL=7]	14.5833%	94.5946%	2.20	2
[physical inactivity]+[SLOPE=2]	14.1667%	94.4444%	2.20	2
[NUM_VESSELS=2]+[SLOPE=2]	6.6667%	94.1176%	2.19	2
[obesity]+[NUM_VESSELS=2]	6.2500%	93.7500%	2.18	2
[physical inactivity]+[NUM_VESSELS=1]	6.2500%	93.7500%	2.18	2
[NUM_VESSELS=2]+[ECG=2]	6.2500%	93.7500%	2.18	2
[alcoholic]+[THAL=7]	5.4167%	92.8571%	2.16	2
[alcoholic]+[ANGINA=y]	5.0000%	92.3077%	2.15	2

© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- ***Warehouse Tooling Components***
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - ***OLAP (Cubing Services)***
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*

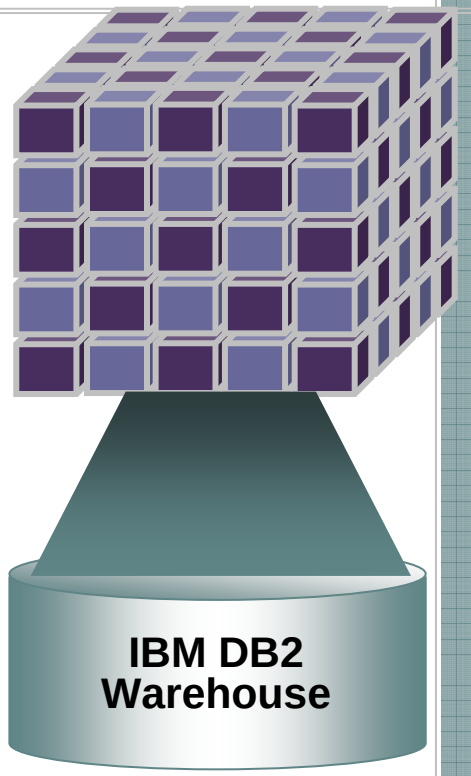
© 2007 IBM Corporation

Cubing Services (OLAP)

DB2 Cubing Services is a Multidimensional Analysis Server that enables OLAP applications access to Terabyte data volumes via industry standard OLAP connectivity

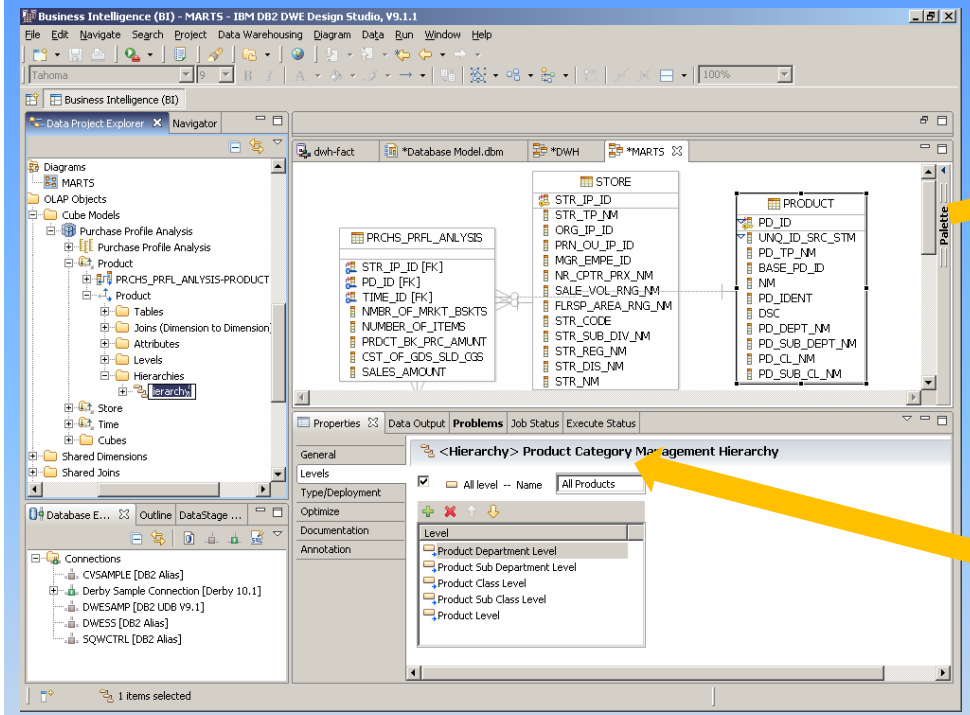
Benefits

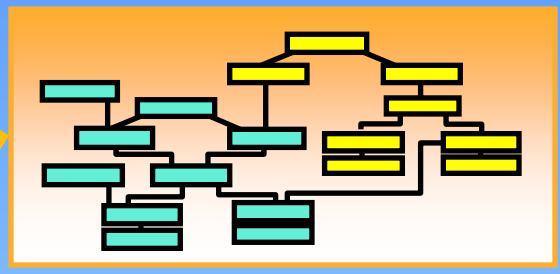
- Uses industry standards OLAP Interfaces for access by a wide variety of tools for presentation and reporting
- Empowers users with ad hoc access to business information.
  - –What is the profitability for Product A across the Branches X,Y,Z?
- Speed of thought access to OLAP data managed in a DB2 Data Warehouse. Thus for your DB2 Warehouse OLAP data there is:
  - –OLAP and SQL shared access to the same information
  - –Single point of Management
  - –Single point of Maintenance
  - –Single point of Performance Tuning
- Enables Access to up to 1 TB of base OLAP Data

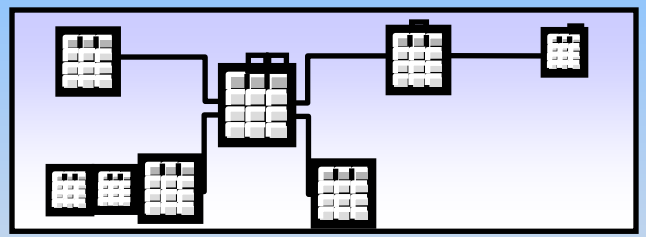


© 2007 IBM Corporation

Cube Modeling in Design Studio





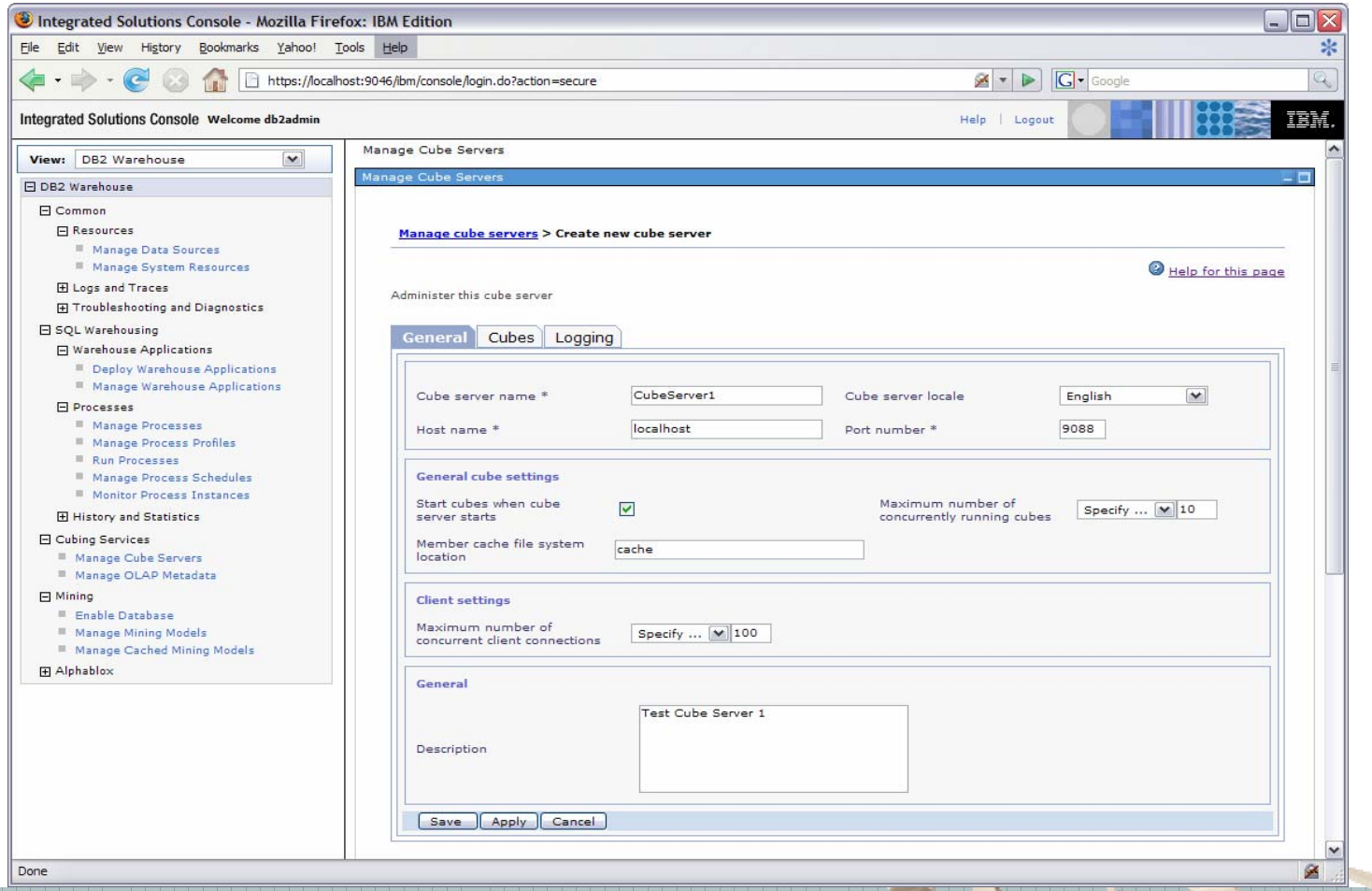


- Multidimensional Business Model
  - Maps logical business concepts to physical relational tables

- Deploy to the Cube Server
  - Available to Applications
  - Database can be optimized

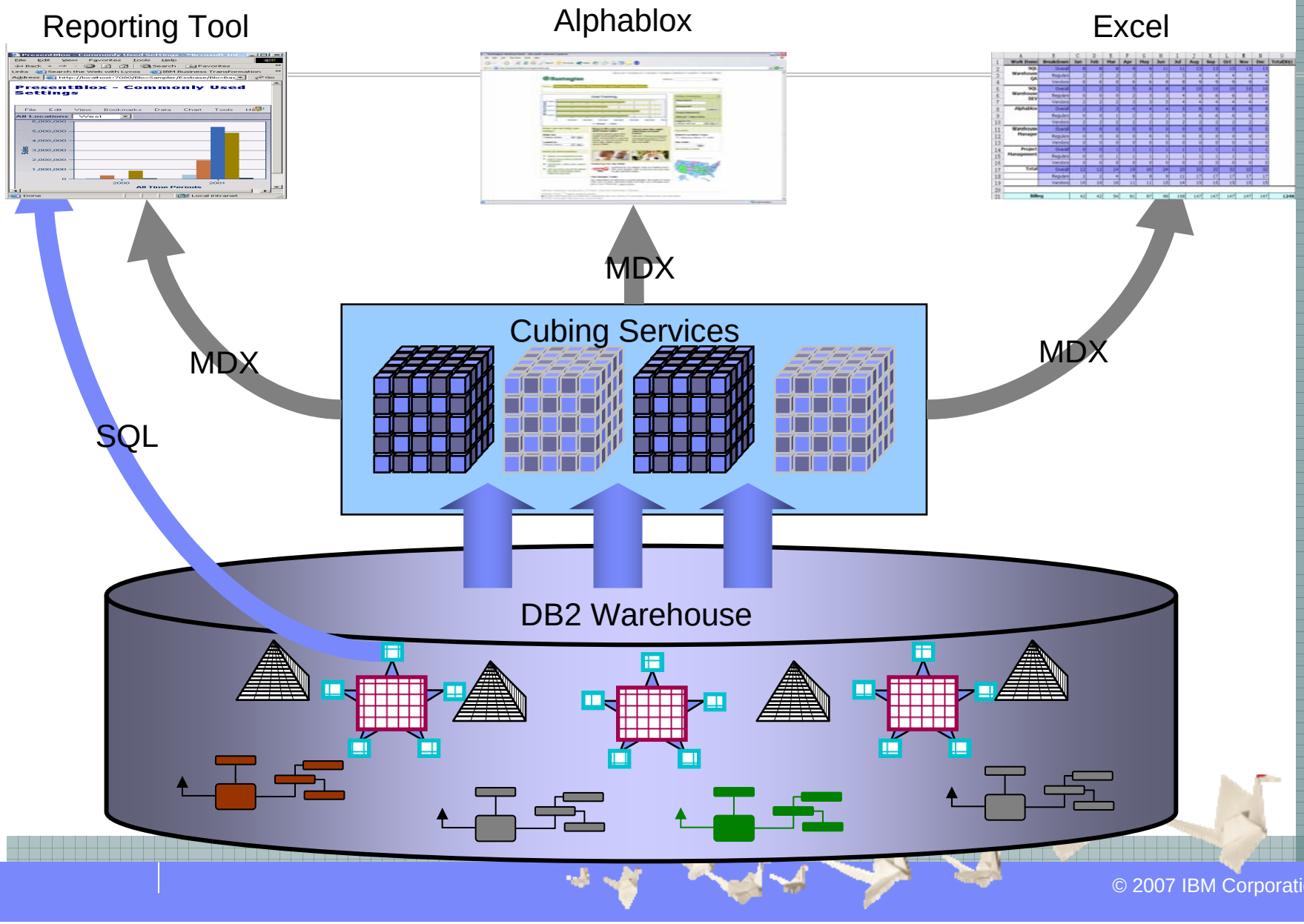
© 2007 IBM Corporation

Cubing Services Management: Create Cube Server



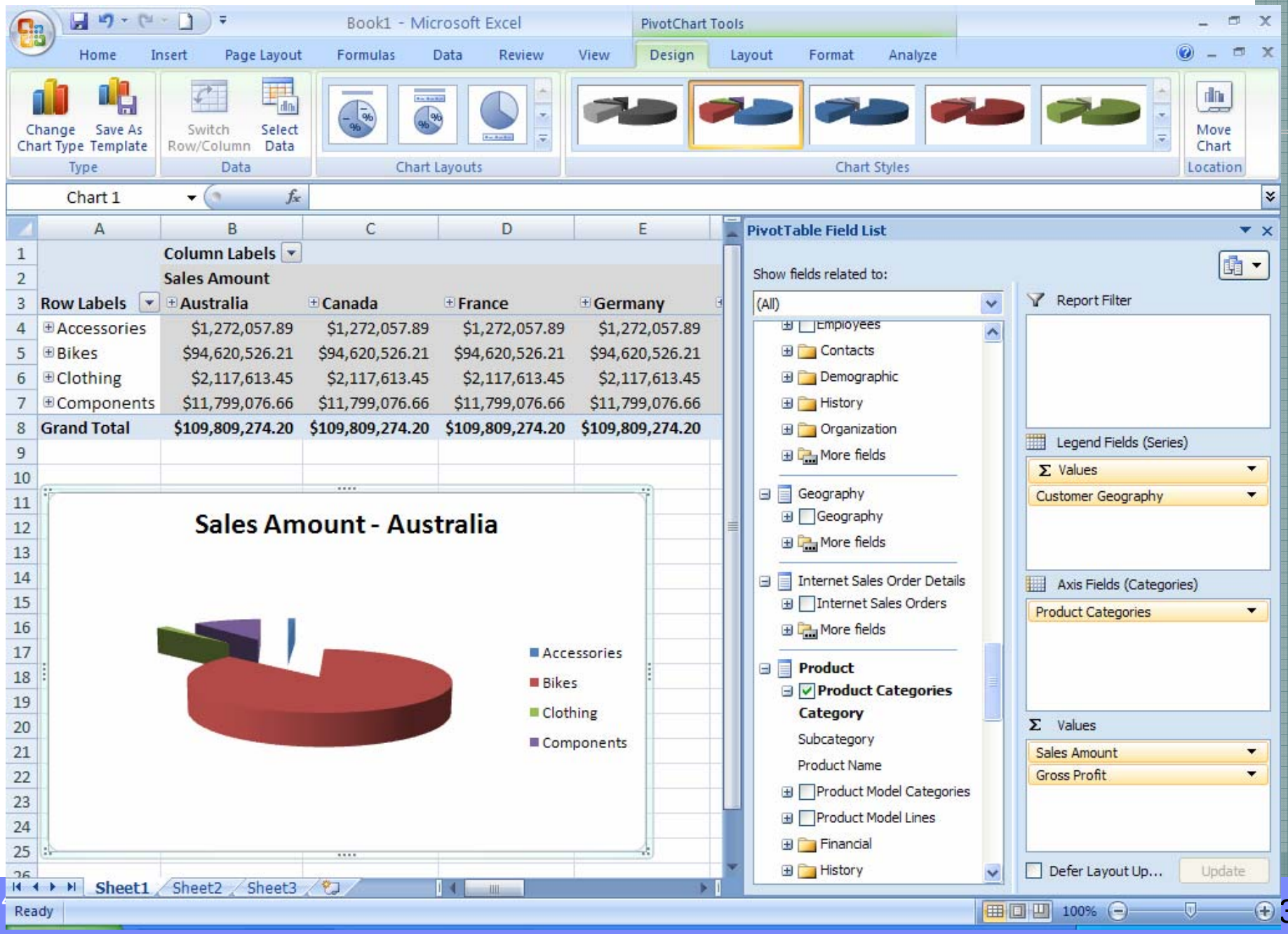
© 2007 IBM Corporation

Cubing Services



© 2007 IBM Corporation

Excel 2007 – Common BI Desktop for DB2W 9.5.1 (Q108)



Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- ***Warehouse Tooling Components***
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - ***Workload Management (WLM)***
  - *Unstructured Analytics*
  - *Data Mining*
  - *Integration…..*

© 2007 IBM Corporation

Workload Management

DB2 Workload Management enables you to enforce and monitor the business service levels for users across your warehouse.

Benefits

- *Enables Consolidation of Users into one Warehouse*
  - *Manages workload based on user priorities*
  - *Prevents Proliferation of Data Marts*
  - *Prevents Runaway workloads*

- *Empowers the Convergence and Management of OLTP and BI workloads onto one Platform*
  - *Real-Time Warehousing*
  - *Operational Data Stores*
  - *Batch Workloads*

- *Improves Monitoring and Performance management*
  - *Real time monitoring*
  - *Real time resource management*

© 2007 IBM Corporation

Workload Management

How Does it Work?

- DB2 Workload
  - Routes workloads to a DB2 Service Class
  - Defined by combinations of
    - System and Session Users and Groups
    - Client Userid. Application Name, Workstation Name or Client Account attributes

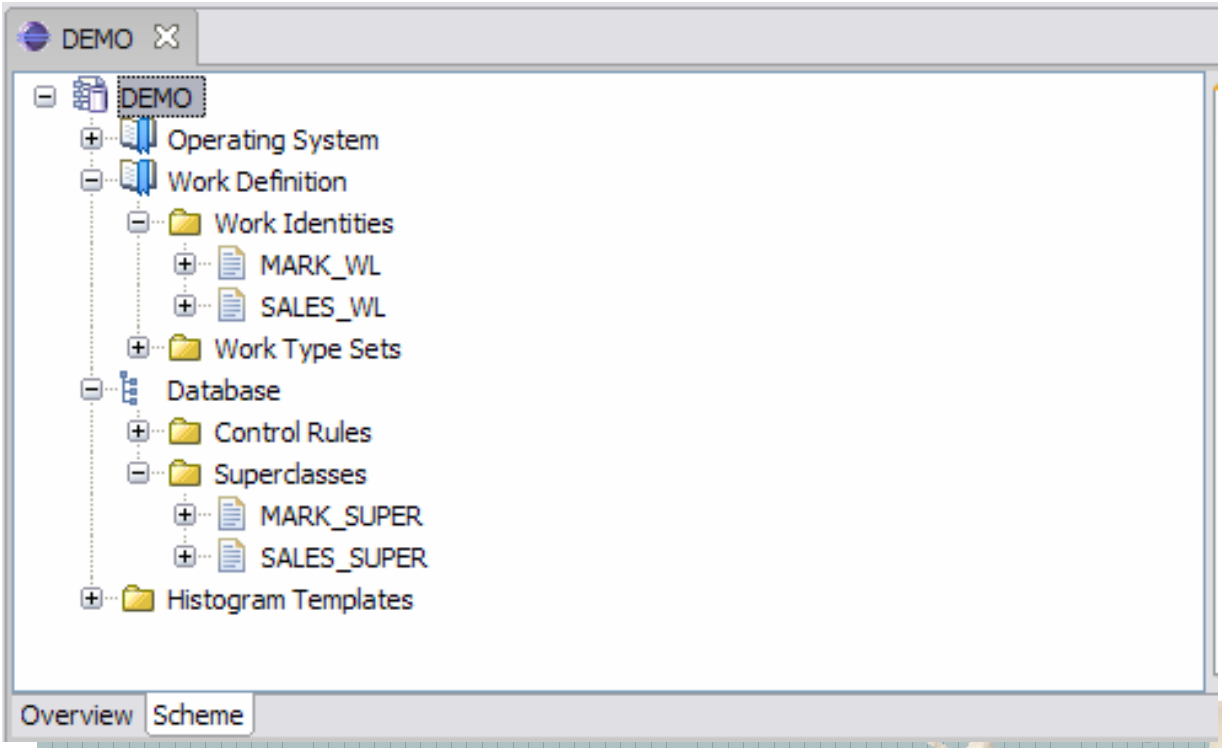
- DB2 Service Class
  - This is where all work for a database is executed
  - Service Classes are broken up into Super Classes and Sub Classes
  - Service Classes dictate
    - CPU Priority
    - Prefetch I/O Priority

- Work Action Set
  - Ability to treat activities differently based on either type or a individual characteristic e.g
    - "Put DML in a different service class than DDL"
    - "Put all Read queries of less than 100 timerons in a different service class than all the other Read queries"
    - "Put a restriction on the number of concurrent Loads allowed to execute on the database at any one time"

© 2007 IBM Corporation

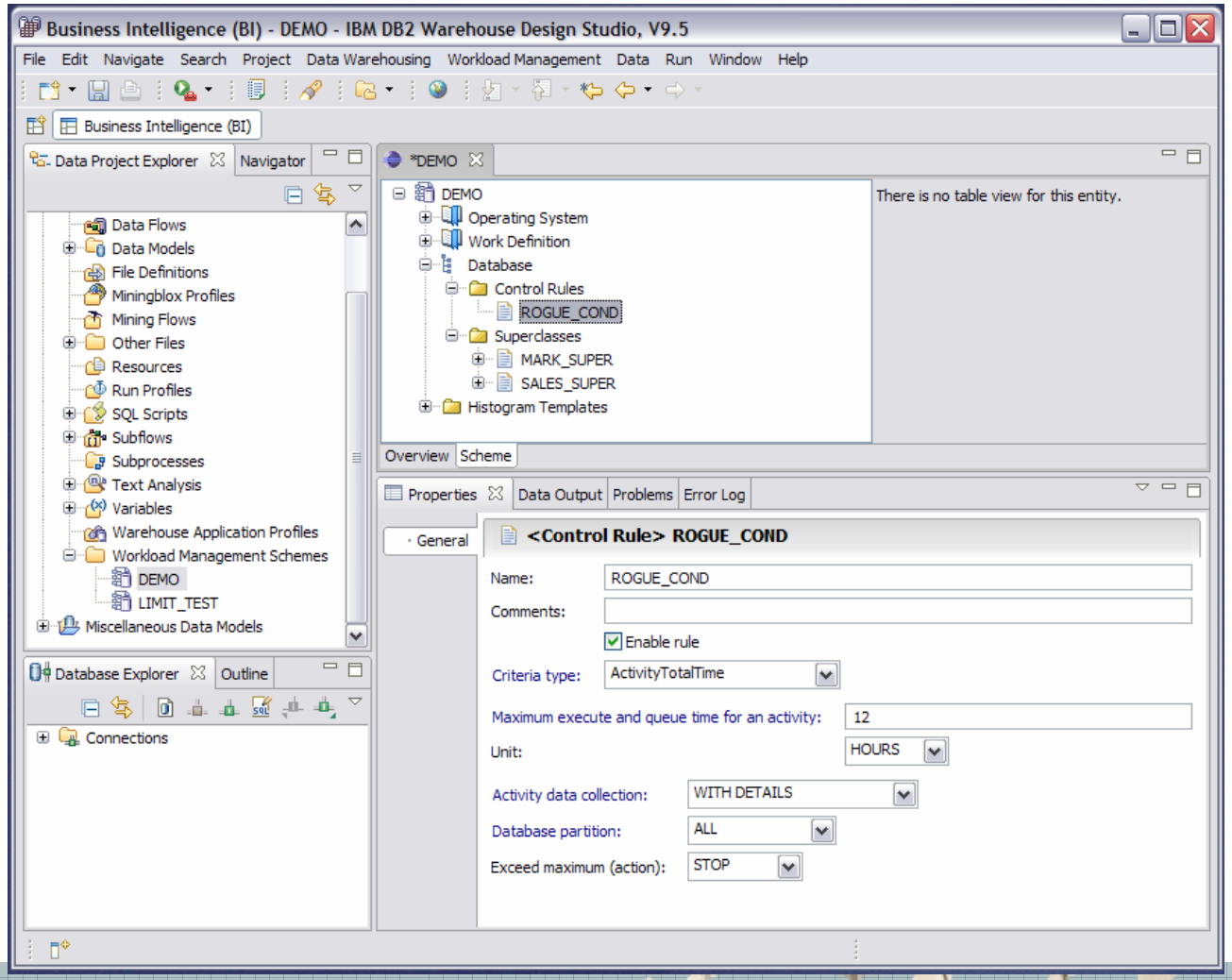
WLM Graphical Interface- Tree View

- Hierarchy of folders for WLM entities
- Provides a global view of WLM solution
- WLM entities can be manipulated directly in tree view



© 2007 IBM Corporation

Configure a Threshold in Tree View



© 2007 IBM Corporation

Agenda

- *Warehouse Tooling*
  - *What Is it ?*
  - *What can it do for me ?*
  - *Can I use it with externally Based Tooling ?*
- *Operational Advantages*
  - *Design Studio*
  - *Administration Console*
- ***Warehouse Tooling Components***
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - ***Integration…..***

© 2007 IBM Corporation

Alphablox BloxBuilder

Alphablox BloxBuilder provides an easy to use interface that accelerates the delivery of user friendly analytical applications. Alphablox provides a framework for delivering custom built analytical applications that can be deployed on their own or integrated into business Applications.

Benefits

Easy to Use

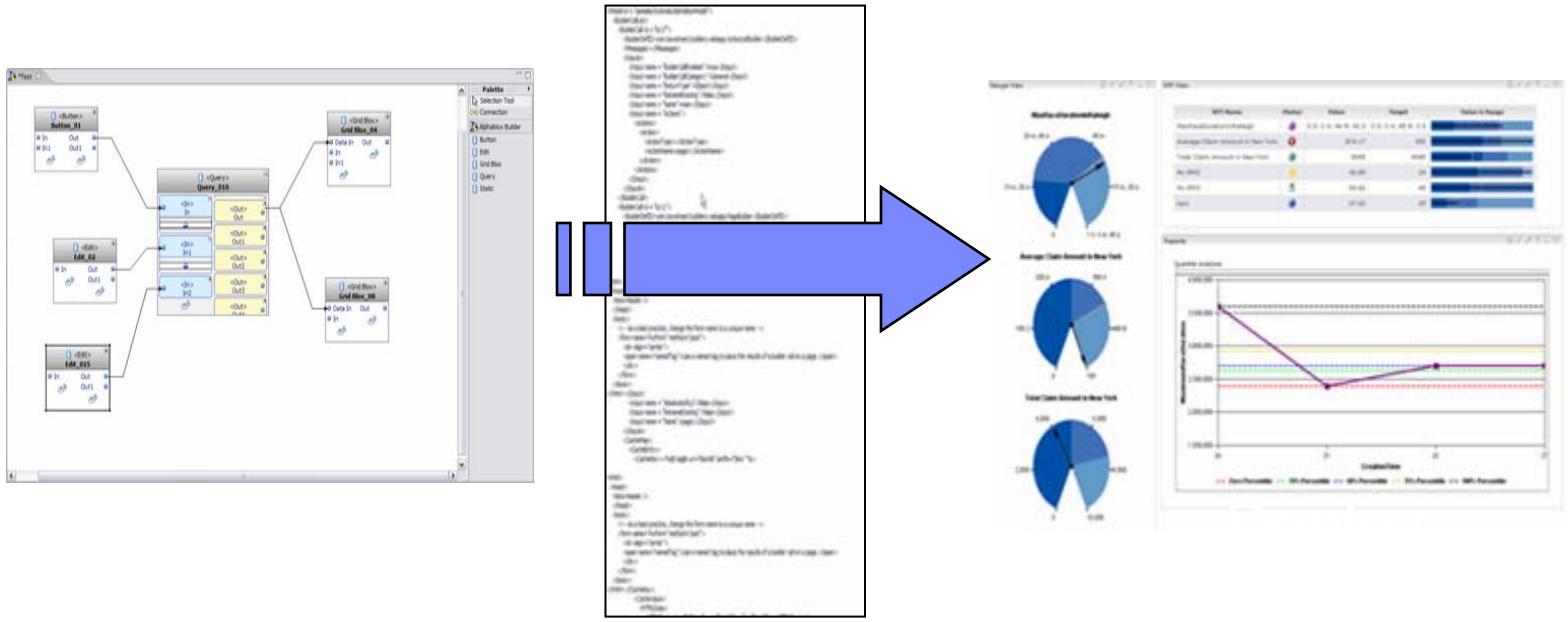
- *Enables the quick assembly of analytical applications through an easy to use graphical interface*
- *Builds fully integratable applications that can be embedded into business applications, such as J2EE applications and Enterprise Portals*

Business Value

- *Applications can be developed and deployed by staff with BI only experience*
- *Provides a low cost reporting and analysis deployment platform*

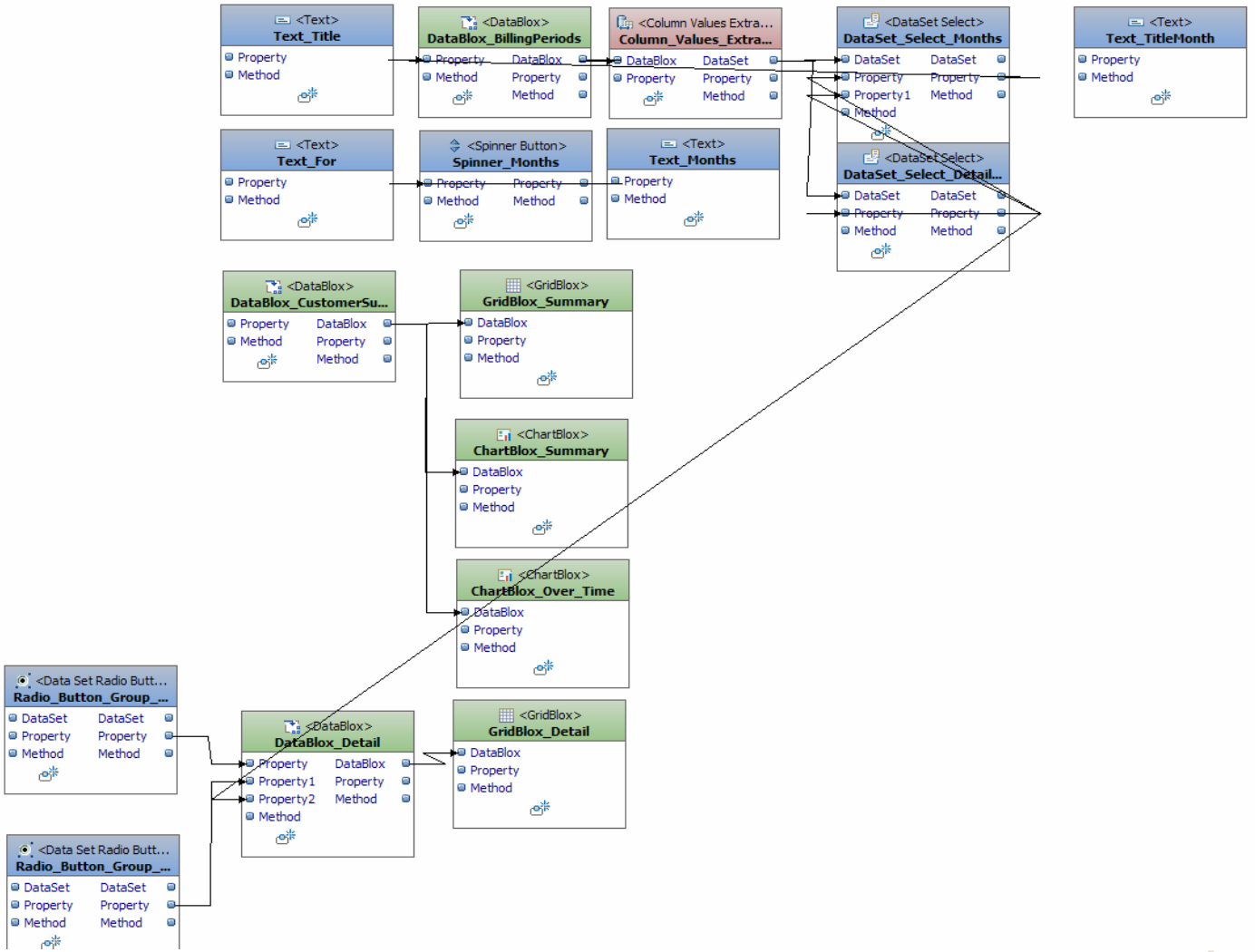
© 2007 IBM Corporation

Alphablox Easy to Build - Generate Reports Without Code



© 2007 IBM Corporation

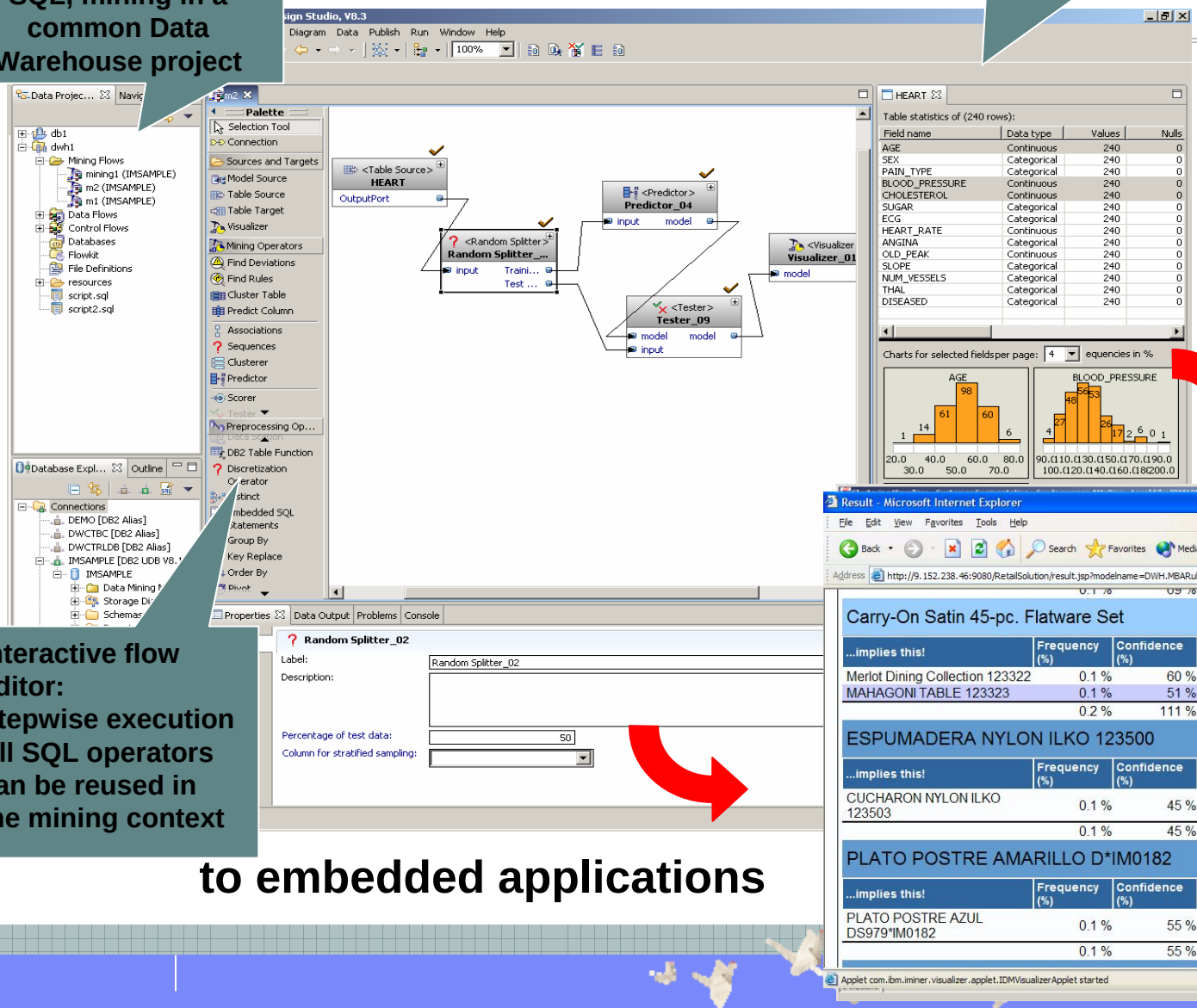
Complex model



© 2007 IBM Corporation

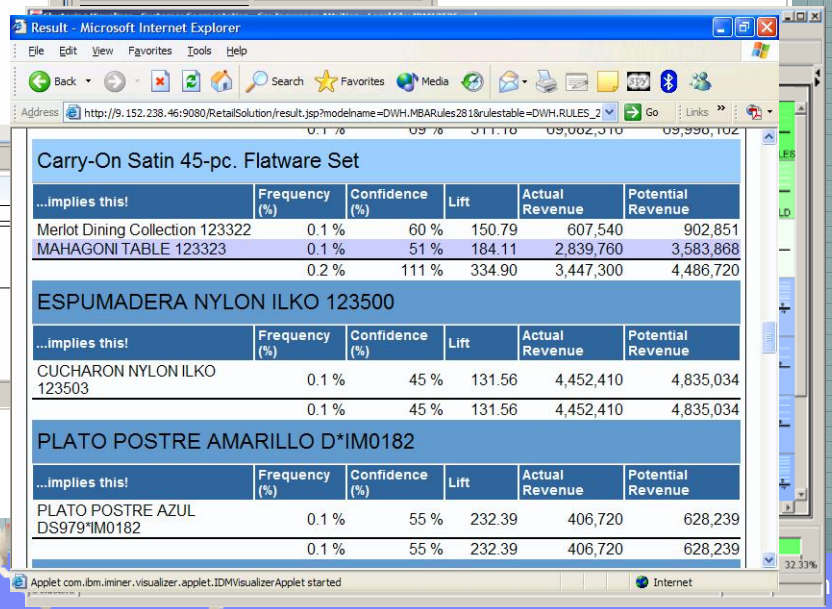
Seamless integration of data modeling, SQL, mining in a common Data Warehouse project

Data exploration components give you an overview on the content and quality of your data.



Design Studio Mining Editor from Design

Interactive flow editor:
Stepwise execution
All SQL operators can be reused in the mining context



to embedded applications

Summary

- *DB2 Warehouse tooling provides significant Operational Advantages*

- *Warehouse Tooling Provides Ease of development through access to the following Components*
  - *Modeling (RDA)*
  - *Transformation (SQW)*
  - *OLAP (Cubing Services)*
  - *Workload Management (WLM)*
  - *Unstructured Analytics*
  - *Data Mining*
  - *and more*

© 2007 IBM Corporation

Copyright information

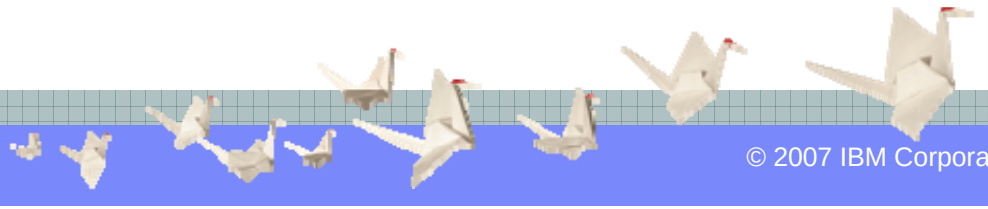
© Copyright IBM Corporation 2007

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
03-07
All Rights Reserved.

DB2, IBM, the IBM logo, OmniFind, Rational and System z are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Other company, product and service names may be trademarks or registered trademarks or service marks of others. The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this documentation, it is provided “as is” without warranty of any kind, express or implied. In addition, this information is based on IBM’s current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation or any other documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.



© 2007 IBM Corporation