



Informix Warehouse

InfoBahn 2010

Agenda

- **Introduction**
- **Overview of Informix Warehouse Components**
 - IDS 11
 - Design Studio
 - SQL Warehouse Tool
 - Warehouse Admin Console
- **Overview of Features**
 - Physical Data Models
 - Data Transformation and movement
 - Data Flows
 - Control Flows
 - Data warehouse applications
 - Application Management and Scheduling
- **Summary**

Introduction

- **Informix Warehouse is an offering**
 - For Informix customers with data warehouse requirements
 - Aimed at reducing operation complexity and cost
 - Using Informix for transactional and warehouse data management
 - Informix Warehouse Feature
 - Informix Warehouse Workgroup
 - Informix Warehouse Enterprise

- **Informix Warehouse**
 - Has a client/server architecture
 - Supplies state of the art ETL(Extract,Transform,Load) /
ELT(Extract,Load,Transform) tools
 - In an intuitive graphical environment
 - Enables BI applications and tools to leverage data better
 - Provides the foundation to cost effectively
 - Build and deploy next generation analytic solutions with Informix

Overview of Components

- **IDS 11**
 - **Design Studio**
 - **SQL Warehouse Tool (SQLW)**
 - **Warehouse Admin Console**
 - WebSphere Application Server 7.0
 - SQL Warehouse Tool Services
- } Eclipse environment

Design Studio, SQL Warehouse Tool

Platform	Operating System
Intel x86-32	Microsoft Windows® XP SP2 Professional, Windows Vista Business/Enterprise/Ultimate Microsoft Windows® XP and Vista with FDCC support Linux® SLES 10 SP2, RHEL 5.2

Warehouse Administration Console

Platform	Operating System
Intel x86-32	Microsoft Windows® 2003 SP2 (32-bit)
Intel/AMD x86-64	Microsoft Windows® 2003 SP2 (64-bit) Linux® SLES 10 SP2, RHEL 5.2
Sun SPARC	Sun Solaris® 9, 10 (64-bits)
IBM Power PC	IBM AIX® 5.3 TL8, 6.1 SP4



Informix Warehouse Components

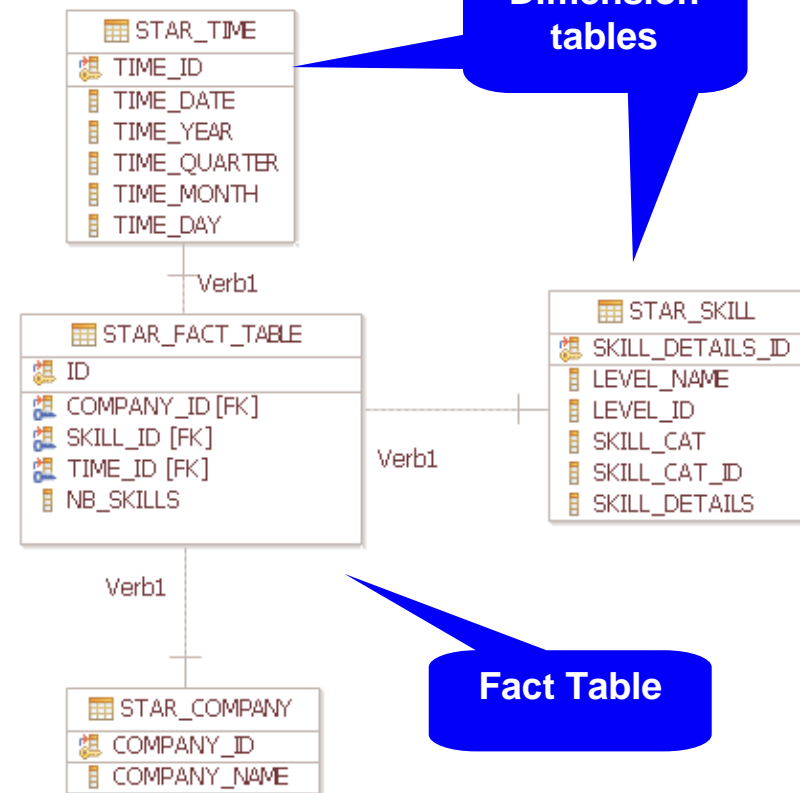
IDS 11 – Warehouse Enabling Capabilities

- **Multi-threaded Dynamic Scalable Architecture (DSA)**
 - Scalability and Performance
 - Optimal usage of hardware and OS resources
- **DSS Parameters to optimize memory**
 - DSS queries
 - Efficient hash joins
- **Parallel Data Query for parallel operations**
 - Light scans, extensive calculations, sorts, multiple joins
 - Ideal for DSS queries and batch operations
- **Time cyclic data management**
 - Fragment elimination, fragment attach and detach
 - Data/index distribution schemas
 - Improve large data volume manageability
 - Increase performance by maximizing I/O throughput
- **Configurable Page Size**
 - On disk and in memory
 - Additional performance gains
- **Large Chunks support**
 - Allows IDS instances to handle large volumes
- **Quick Sequential Scans**
 - Essential for table scans common to DSS environments

IDS – OLTP and Star Schemas

- **OLTP**
 - Highly normalized database structure
 - Designed for:
 - Fast inserts, updates, deletes
- **Star Schema**
 - Designed for fast query processing
 - Redundancy part of design
 - Consists of:
 - Central fact and dimension tables
- **IDS Warehouse features enable**
 - Quick processing of decision support queries
 - Based on star schemas

A SAMPLE STAR SCHEMA



Data Warehouse building steps

- **Define a data warehouse project**
- **Design data warehouse model**
- **Build processing jobs**
 - Processing data
 - Control jobs
- **Deploy**

Design Studio

- **Eclipse based design environment**
 - Workspace
 - Central repository for data files
 - Perspectives
 - Contains views, editors
 - Controls display of menu and toolbars
 - Projects
 - Create objects as part of data transformation process

- **Graphical Capabilities**
 - Connect source and target databases
 - Create, modify, generate DDL for physical data models
 - Build SQL-based data transformations
 - Build Control Flows
 - Application Packaging

Projects

- **Represented as icons in Data Project Explorer**
- **Organize resources for data warehouse**
- **Associated with a sub-directory on disk**
 - Within a workspace directory
 - Metadata file within the directory “.project”
- **Build and test validity**
 - Without impacting database
- **Work with physical models**
 - Project type has to be “data design”

The screenshot displays the IBM Informix Data Warehousing Design Studio interface. The main workspace shows a data flow diagram with several operators: two Table Source operators (aroma_period and aroma_sales), a Table Join operator (Table Join 1), an Order By operator, a Distinct operator, and a Table Target operator (Insert). The interface includes a Navigator on the left, a Properties pane at the bottom, and an Operator Palette on the right. Blue callout boxes highlight specific features: 'Data Warehousing Perspective' points to the top menu bar; 'Data Flow and Control Flow Editor View' points to the central diagram; 'Operator Palette' points to the right-hand palette; and 'Database Connections' points to the Data Source Explorer on the bottom left.

Physical Data Models

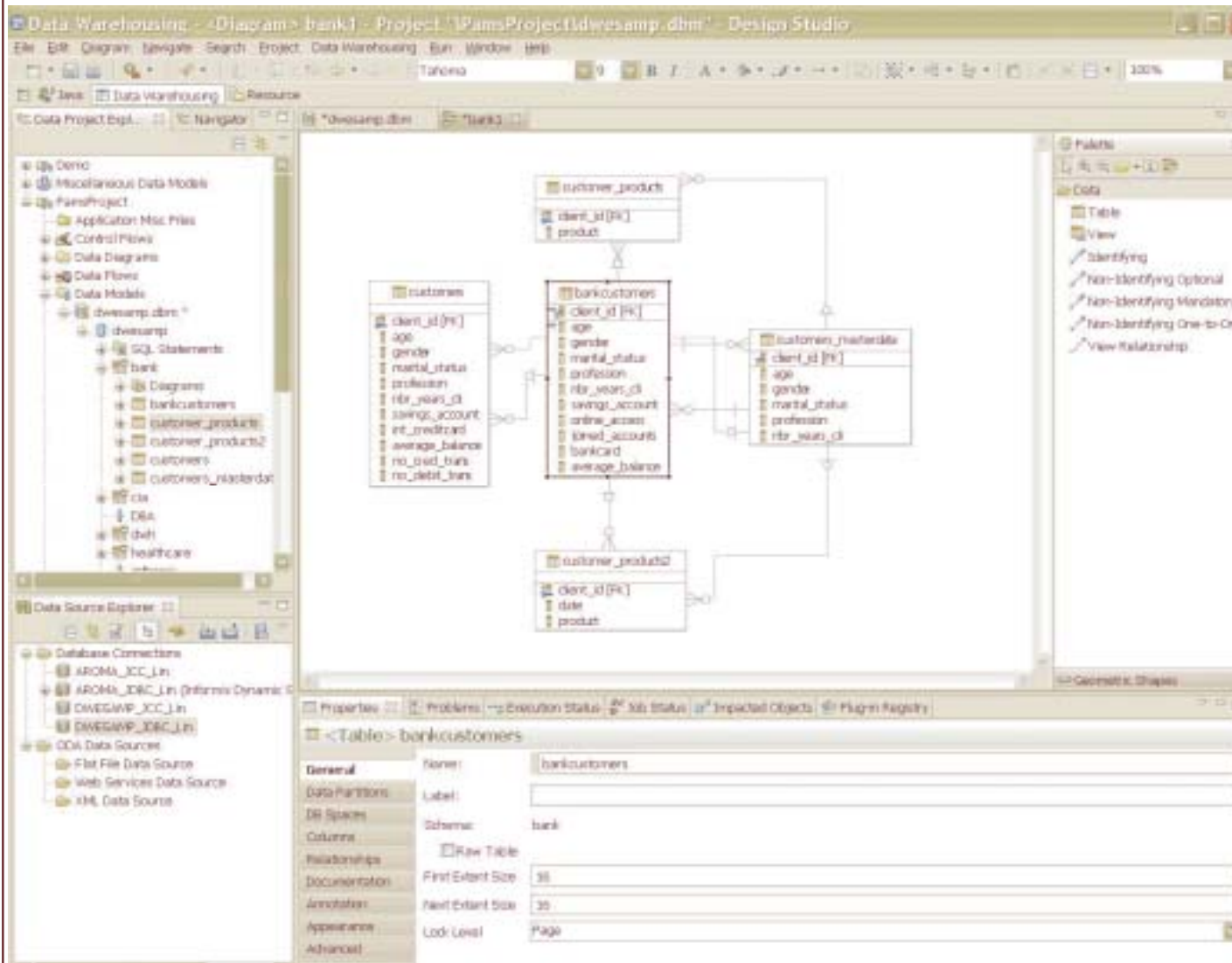
- **Database specific model**
- **Represents relational data objects and relationships**
 - Example: Tables, columns, primary keys, foreign keys
- **Create Data Models**
 - From scratch or reverse engineer from live database
 - For Source, target databases, staging tables
- **Deploy generated DDL directly to database server**
- **Compare data objects**
 - Analyze impact and dependencies
 - Copy changes or merge properties between objects
 - Export structural differences to XML file on disk

Generate an Informix Warehouse Data Model

The screenshot shows the IBM Data Warehousing - Design Studio interface. The 'Data Project Explorer' on the left displays a project structure with 'Data Model' selected. A context menu is open over 'Data Model', showing 'New' with a sub-menu containing 'Physical Data Model'. The 'New Physical Data Model' dialog box is open, showing the following fields and options:

- Model File**: Specify the database, version, and location of the new model file.
- Destination folder**: /PamsProject
- File name**: AROMA
- Database**: Informix
- Version**: 11.1
- Options**:
 - Create from template
 - Create from reverse engineering
- Navigation**: < Back, Next >, Finish, Cancel

Work with your Data Model



- Visualized data modeling
- Impact analysis
- Reverse engineering or new from scratch
- Compare & sync
- Generated DDL
- Overview diagram

SQL Warehouse Tool

- **Works in conjunction with Design Studio**
- **Solves data movement and integration problems**
- **Data Flow/Transform, Control operators**
 - File export and import
 - Join, group by, order by, distinct
 - Variable assignment/comparison, stored procedure
- **Warehouse operators**
 - Fact key replace, key lookup
- **Informix specific operators**
 - Attach partition, detach partition, update statistics
- **Code generation system**
 - Translates flow models into optimized SQL code

Warehousing Operations

CONTROL FLOW OPERATORS

A screenshot of a software palette titled 'Palette'. It contains several categories of operators:

- Selection Tool** (highlighted)
- Connection
- Note
- Common Operators** (expanded):
 - Command (highlighted)
 - Data Flow
 - DataStage Job Sequence
 - DataStage Parallel Job
 - Email
 - File Wait
 - File Write
 - Iterator
 - Parallel Container
 - Secure Command
 - Secure FTP
 - Stored Procedure
 - Subprocess
 - Variable Assignment
 - Variable Comparison
- Break
- Continue
- End
- Fail
- Informix Operators

INFORMIX SPECIFIC OPERATORS

A screenshot of a software palette titled 'Informix Operators'. It contains the following operators:

- Attach partition
- Detach partition
- IDS Custom SQL
- IDS SQL Script
- Update Statistics

DATA FLOW and TRANSFORM OPERATORS

A screenshot of a software palette titled 'Palette'. It contains several categories of operators:

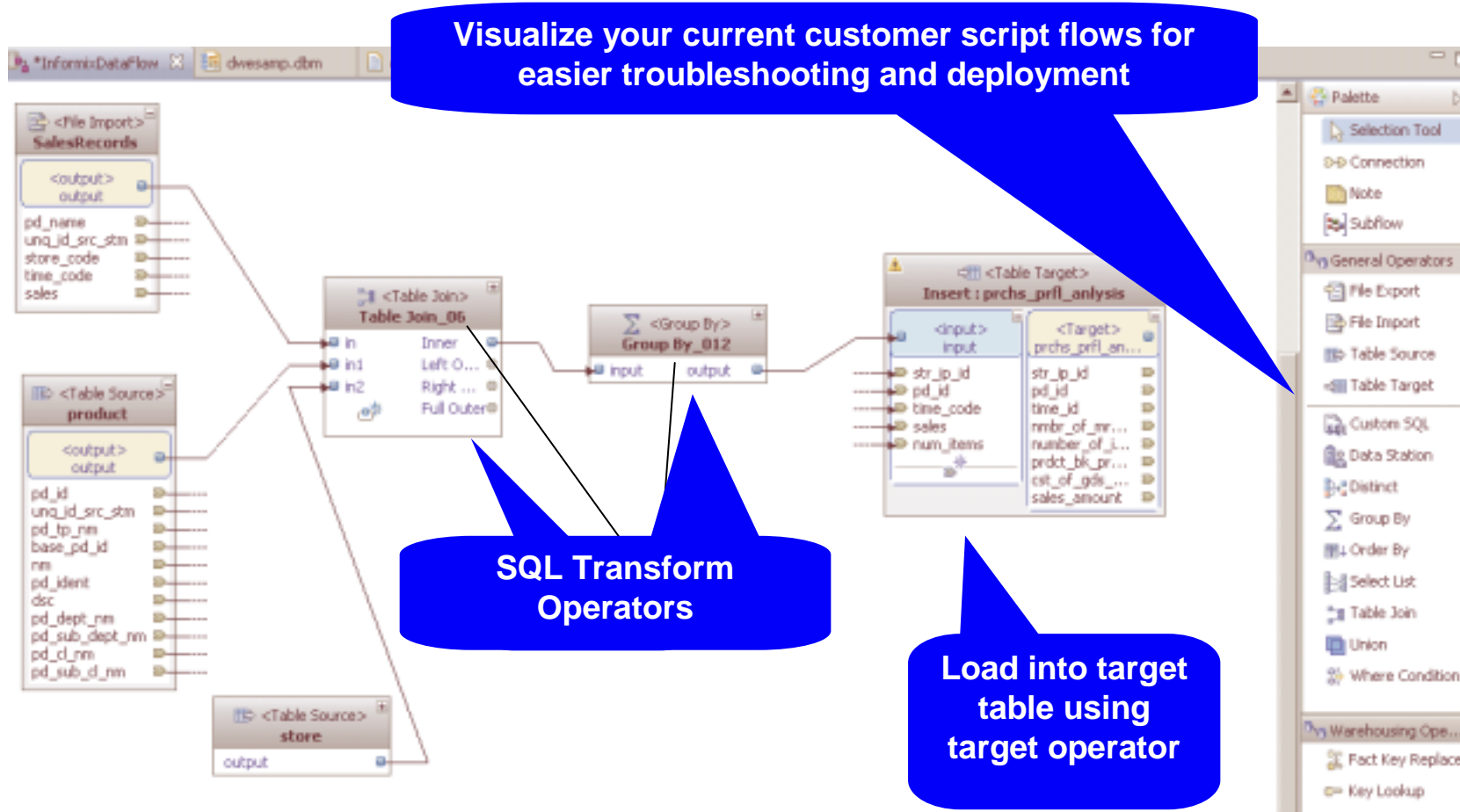
- Selection Tool** (highlighted)
- Connection
- Note
- Subflow
- General Operators** (expanded):
 - File Export
 - File Import
 - Table Source
 - Table Target
- Custom SQL
- Data Station
- Distinct
- Group By
- Order By
- Select List
- Table Join
- Union
- Where Condition
- Warehousing Operators

WAREHOUSE SPECIFIC OPERATORS

A screenshot of a software palette titled 'Warehousing Operators'. It contains the following operators:

- Fact Key Replace
- Key Lookup
- Pivot
- Splitter
- Unpivot

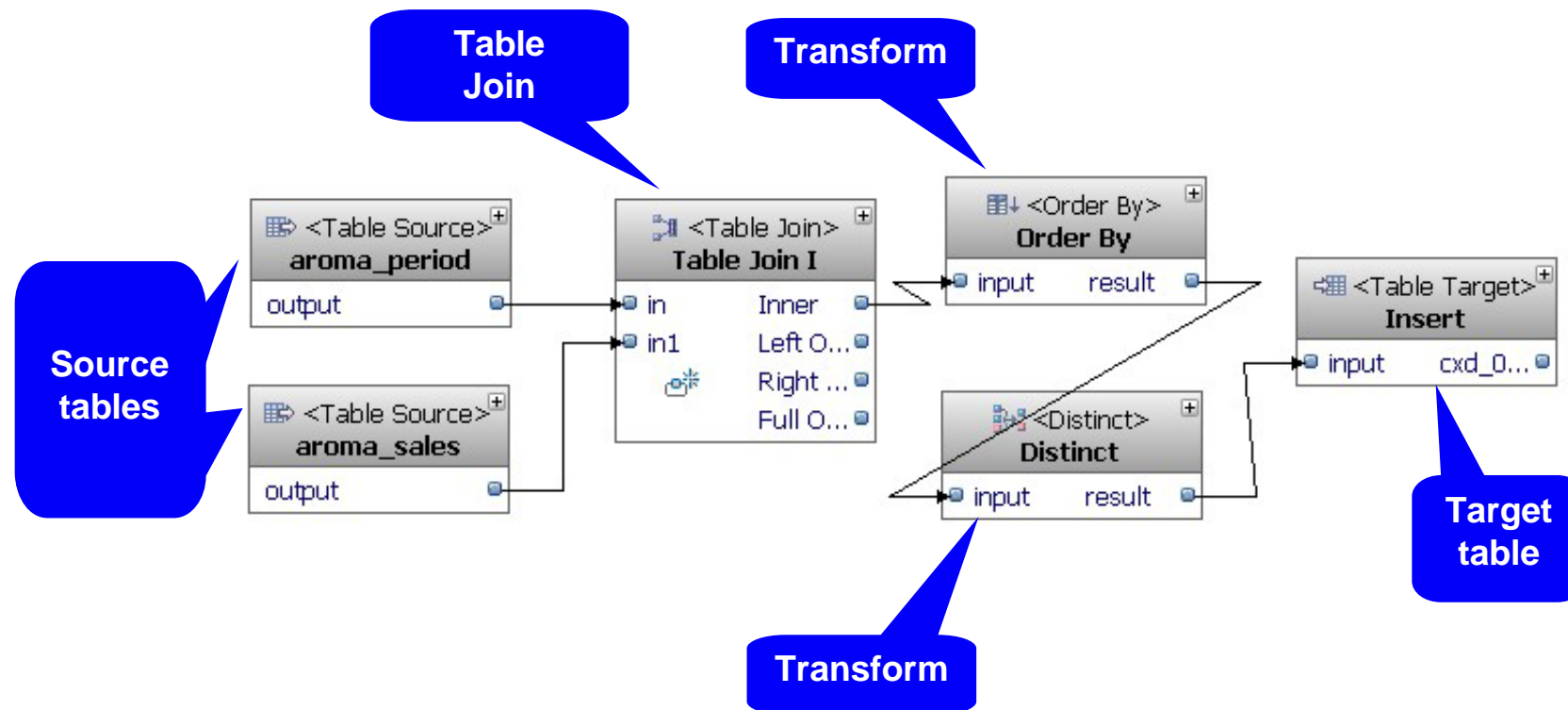
Warehousing Operations



A Simple Data Flow

■ Data Flows

- Define data transformation activities
 - Extract data from flat files or relational table sources
- Visualize and design using Design Studio
- Model flow activity steps with SQL Warehouse Tool



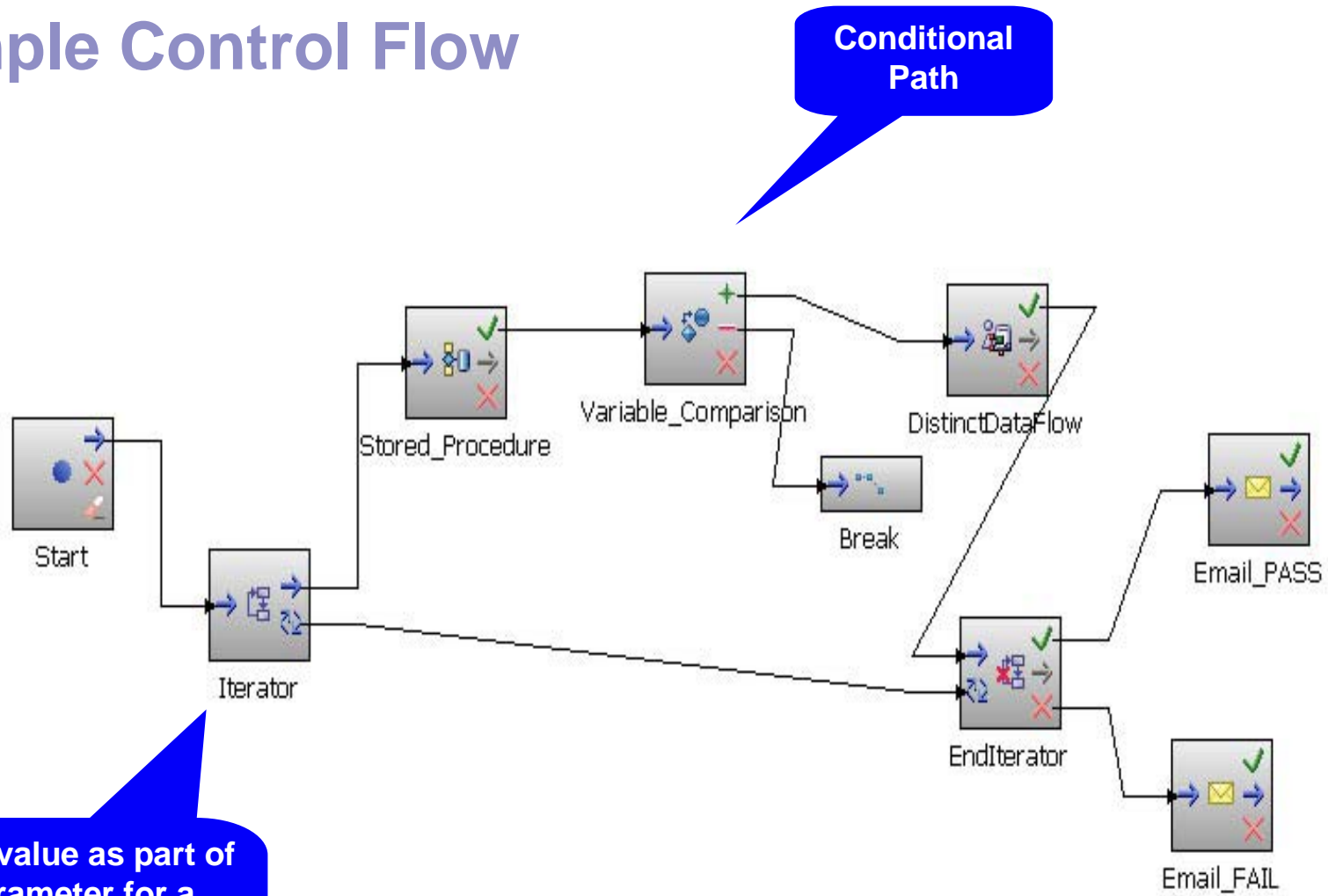
Control Flows

- **Processes to coordinate activities**
 - Data flows, other control flows
 - Define parallel processing, conditional paths
 - Error handling

- **Reusable within other control flows**
 - Sub-processes embedded within main control flows

- **Operators available via SQL Warehouse Tool**
 - Iterators, variable comparisons
 - E-mail, FTP

A Simple Control Flow



Conditional Path

Iterator value as part of IN parameter for a stored procedure

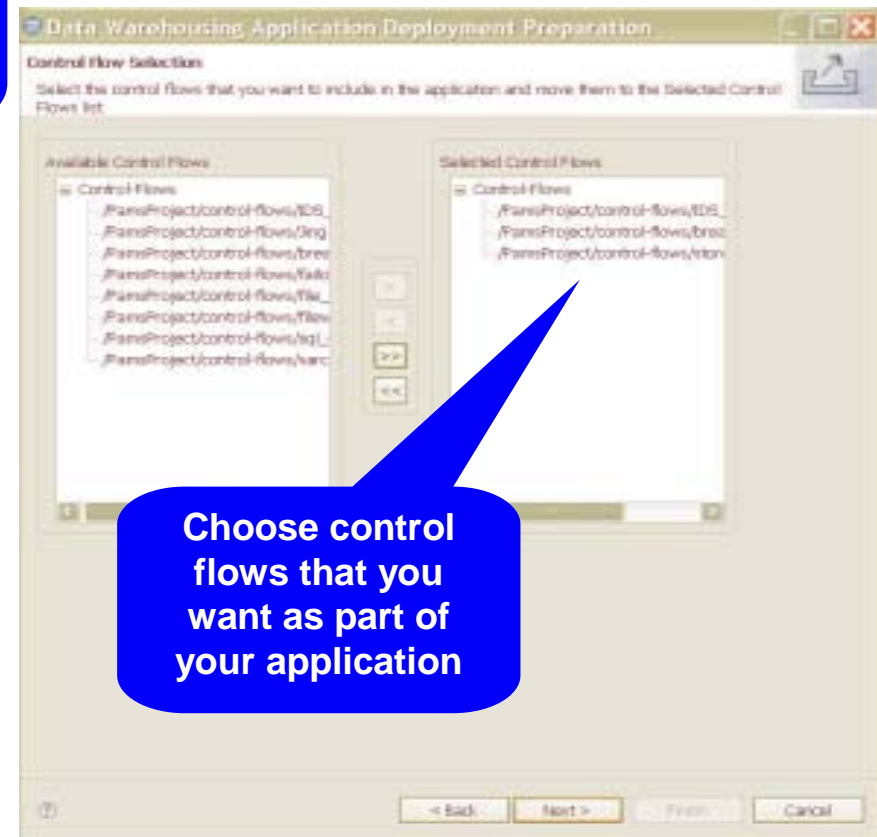
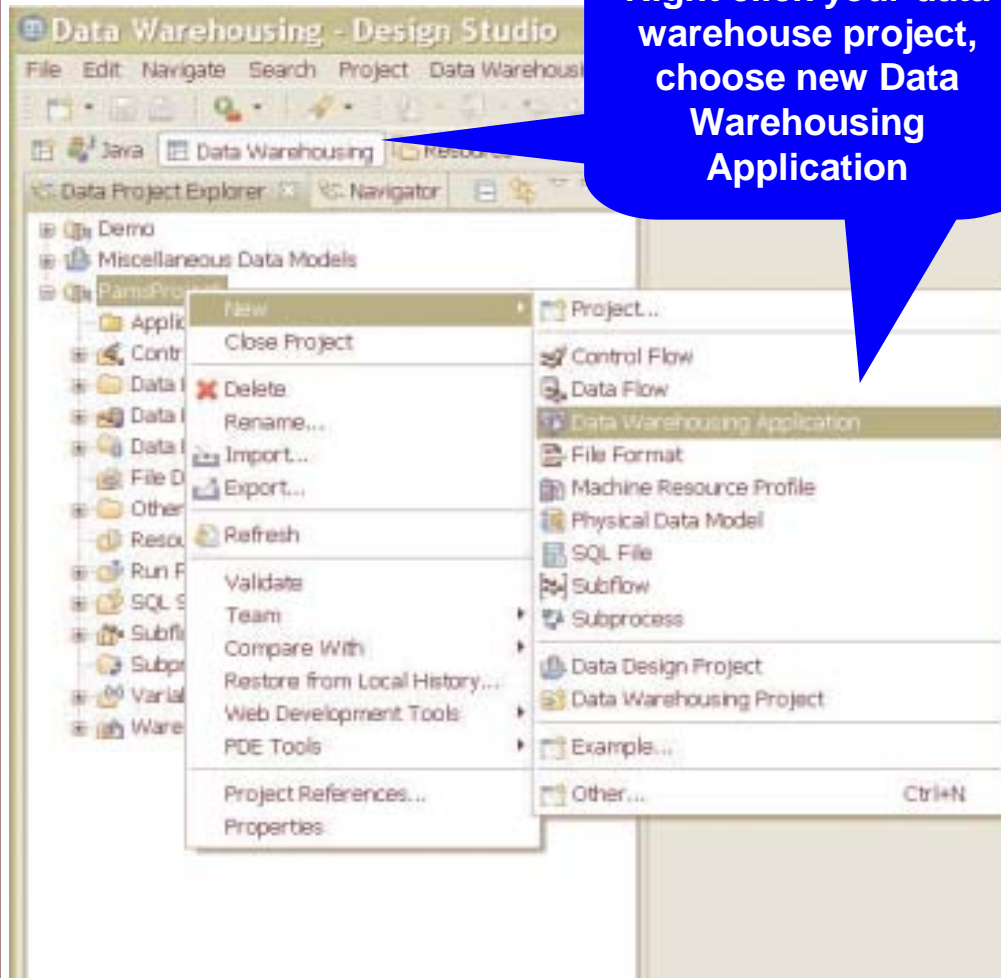
Data Warehouse Application

- **Package control flows into deployable applications**
 - Select the control flows from your project
 - Edit any variables application may require
 - Select any other files you may need for the application
 - Generate application “.zip” file

- **Use Administration Console to deploy application**

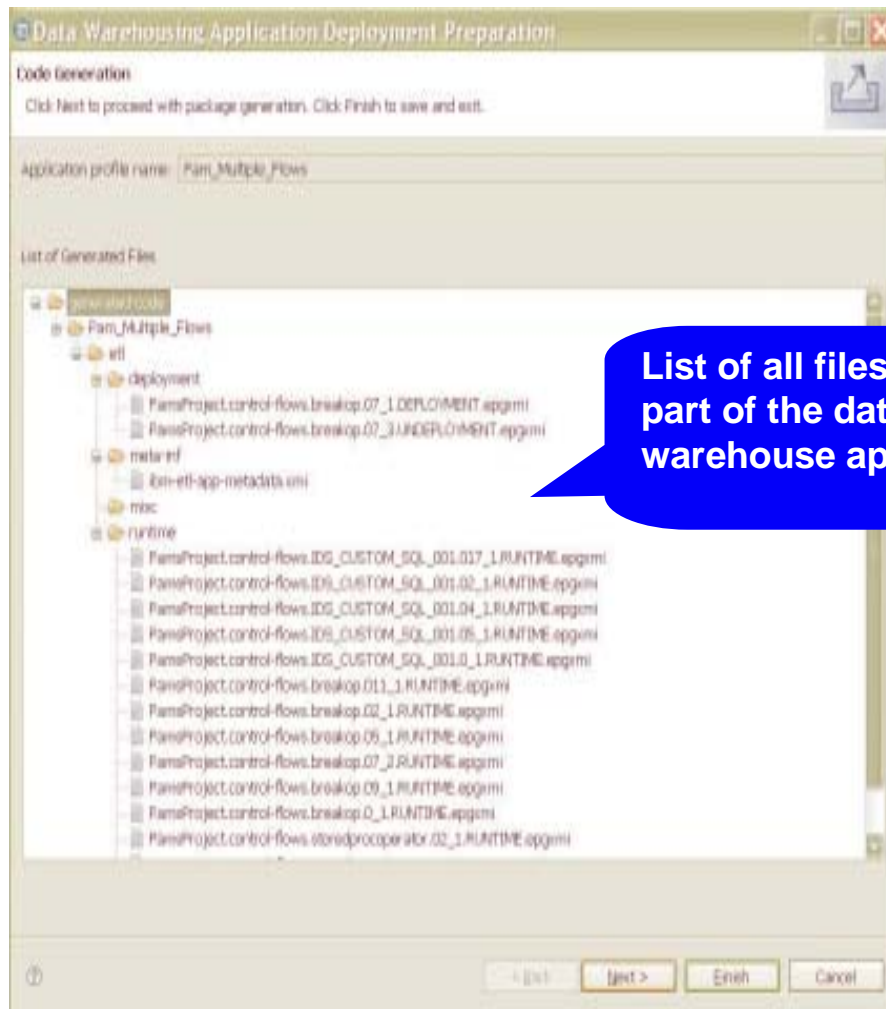
Building the Data Warehouse

Right click your data warehouse project, choose new Data Warehousing Application

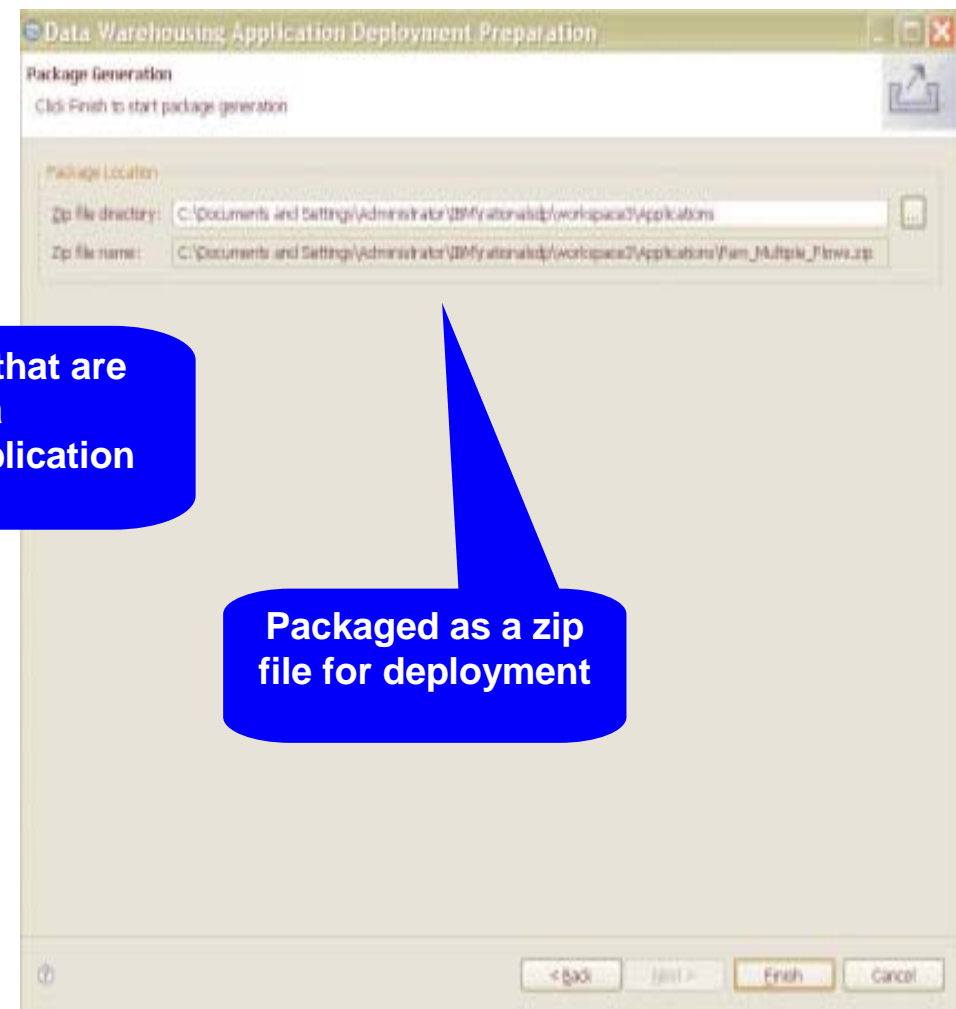


Choose control flows that you want as part of your application

Building the Data Warehouse



List of all files that are part of the data warehouse application



Packaged as a zip file for deployment

Warehouse Administration Console

- **Deploy applications**
- **Manage resources**
 - Database connections, machine resources
 - Schedule the execution of control flows
 - Monitor execution status

The screenshot displays the Warehouse Administration Console interface. At the top, there is a navigation bar with tabs for 'Welcome', 'Configuration', 'Manage Logs', 'Manage Connections', 'Manage System Resources', and 'SQL Warehousing'. Below this, there are sub-tabs for 'Manage Applications', 'Manage Control Flows', and 'Manage Instances'. A message states: 'The deployed applications are listed below. You can deploy a new application or deploy changes to an existing application. Learn more.' Below the message are buttons for 'Deploy...', 'Refresh', 'Add', 'Remove', and 'OK'. The main content area features a table with the following data:

Application Name	Status	Description
Stored Procedure Case 1 SP Business	◆	
Exp1_Script	◆	
Variable Alphabet	◆	
Variable Variable_Scale	◆	

At the bottom right of the table area, there are navigation icons and the text 'Page 1 of 1'.

Application Management and Scheduling

- Manage warehouse applications, control flows
- Create, modify, delete schedules for control flow runs
 - Example: Scheduling a Control Flow to repeat every day for 9 days

The screenshot displays the Informix Application Management and Scheduling interface. The main window shows a table of control flows with the 'Manage Schedules' button highlighted. A modal dialog titled 'Create Schedule for Control Flow: EmailFlow' is open, showing the 'Schedule Setting' step. The dialog includes fields for 'Start Date and Time' (03/22/2010 03:00 PM), 'Frequency' (Every 1 Days), 'Start Date' (03/22/2010), and 'End' (3 total runs). The 'Back', 'Finish', and 'Cancel' buttons are visible at the bottom of the dialog.

Control Flow Name	Application Name
Variable	Variable_Alphabet
EmailFlow	EmailFlowTab_Data Flow and Email Test

Monitoring Application Instances

The screenshot shows a web-based monitoring interface. A central dialog box titled "Monitor Instance" is open, displaying a table of activity logs for a specific instance named "EmailTest_J". The table has the following columns: Activity Name, Status, Activity Type, Start Time, and Elapsed Time. Two rows are visible in the table, both with a green diamond status icon.

Activity Name	Status	Activity Type	Start Time	Elapsed Time
Data_Flow_04	Success	DataFlowActivity	Wed February 04, 2009 11:48:20 AM	1.020s
Email_05	Success	Email	Wed February 04, 2009 11:48:21 AM	4.900s

The dialog box also includes a "Refresh" button at the top and a "Close" button at the bottom. The background interface shows a navigation menu on the left and a main data table on the right, partially obscured by the dialog box.

Informix Warehouse Feature

- **Offering for Informix customers with warehouse requirements**
- **Helps extend legendary IDS performance and scalability**
 - To data warehouse environments
 - Reduce operation complexity and cost
- **Client Server Architecture**
- **Intuitive Graphical Interface**
 - Extract data from various data sources
 - Create physical data models using Design Studio
 - Build SQL-based data transformations with SQL Warehouse Tool
 - Create control flows that can be packaged as applications
 - Deploy, run and manage application using Warehouse Administrator

Informix Warehouse – more information

- **Redbook – “Data Warehousing with the Informix Dynamic Server” -**
<http://www.redbooks.ibm.com/abstracts/sg247788.html?Open>

- **IBM developersWorks**
 - Get started with Informix Warehouse Feature, Part 1: Model your data warehouse using Design Studio -
<http://www.ibm.com/developerworks/data/tutorials/dm-0904warehouse1/index.html>
 - Get started with Informix Warehouse Feature, Part 2: Extract, load, and transform your data in Design Studio -
<http://www.ibm.com/developerworks/data/tutorials/dm-0911warehouse2/index.html>



Informix Warehouse

InfoBahn 2010