Extracting Value from Data: Building Data Warehouses with IBM Informix Dynamic Server

Lester Knutsen

President

Advanced DataTools Corporation

Advanced DataTools

Building Data Warehouses with Informix

Agenda

- Success Stories – We have been developing Data Warehouses with Informix for over 12 years!

- What's New – New IBM Informix Data Warehouse Features

- Keys to Building a Successful Data Warehouse

Advanced DataTools

Success Stories

Bureau of Land Management

- Financial Management Information System

- Performance Measures

- Activity Based Cost Accounting

Farm Service Agency

- CORE Financial Data Warehouse

- Commodity Credit Corporation – Financial Data Warehouse

- Financial Inquires – Financial Statements for Every Farmer

- Transparency and Recovery Act Reporting Data Mart

Advanced DataTools

3

BLM – Using the Data Warehouse to Improve Business Practices

- Provide *all BLM managers and employees* with on-line access to BLM business and management data.

- Eliminate redundant automated systems and data entry in procurement, finance, budget, and other business systems.

- Automate systems to record and track accomplishments and produce reports that provide a verifiable record.

Advanced DataTools

BLM Timeline: 9 Months to Implement First Data Mart

- August/September 1997 - Project Prototype
  - Built Prototype in Denver
- November 1997 - Start Development and Data Conversion
  - Loaded all 1997 and 1998 Data from FFS
  - Software Development and Conversion
  - Dec/March - BLM Road Show to State Offices
- March 1998 - Installation at Business Center in Denver
- March 31, 1998 – First Data Mart - Went Live
- October 1, 1998 – Official Production System

Advanced DataTools

BLM – Additional Data Marts

- Performance and Workload Measures
- Activity Based Cost Accounting
- Collections and Billing System
- Budget Planning System

Advanced DataTools

BLM – Years Later

- MIS Data Warehouse is the official reporting system for financial information.

- Provides analysis and reporting for Cost Management:
  - How did BLM spend the money?
  - What performance measure goals were accomplished with this spending?
  - What workload measures results were accomplished with this spending?
  - What did it cost to achieve these results?

- The Bureau of Land Management, Management Information System was one of seven finalists out of over 100 Federal Systems nominated in 2002.

- Usage
  - 6,722 unique users
  - average of 442 concurrent users

Advanced DataTools

FSA Business Problem

- Implementation of new accounting system.
  - How to replace existing reports?
  - Significant cost in developing new reports
- Need to easily view data detail
- Reconciliation difficult without knowing the activity behind a balance
- Old standard reports were not dynamic

Advanced DataTools

Objectives of the FSA Data Warehouse

- Efficient distribution of information via the Web

- Minimize technical involvement by enabling users to generate and maintain their own reports

- Create a user-friendly reporting environment

- Provide easy access to data from different sources

- Lay the foundation and develop plans for full data warehouse development and implementation

Advanced DataTools

FSA Timeline: 5 Months to Implement First Data Mart

- January - Data Warehouse Pilot
- March - Benchmarks
- May - Production Loads
- May - Web Server & Go Live
- June - Operations & User Training

Advanced DataTools

FSA – Additional Data Marts

- Debt Management Data Mart
- Export Tracking System Data Mart
- Commodity Credit Corporation Financial Data Mart
- Global Supply Management Data Mart
- Customer Payments Data Mart
- Public Financial Inquiry Data Mart

Advanced DataTools

FSA – Additional Data Marts

- 1614 Farm Bill Data Mart

- Web Financial Services Data Mart

- FFATA/FAADS – Recovery Act, Transparency Act Reporting Data Mart

- Financial Web Applications Data Mart – Release I

- Financial Web Applications Data Mart – Release II – Released September 14, 2009

FSA Accomplishments

- The USDA Farm Service Agency Financial Data Warehouse was awarded the 2002 Brio Achievement Award (Co-sponsored by Forbes, Computerworld, and Intelligent Enterprise magazines) for the Best Government Solution.
- Usage
    - Over 13,000 unique users
    - Planned for 1,000+ concurrent users
    - Adding 1 TB of disk per year
- Meet the needs of Congress – 1614 Data Mart
- Meet the needs of FSA employees in 2,000+ offices
- Meet the needs of Farmers/Producers who use FSA Financial Statements

Advanced DataTools

New Informix Data Warehouse Features - A Technical Overview

Release in April 2009

Advanced DataTools

Informix Warehouse 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

  – Has a client/server architecture

  – Supplies state of the art ETL / ELT 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

Advanced DataTools

Components

- Informix IDS 11

- Design Studio

- SQL Warehouse Tool (SQLW)

- Warehouse Admin Console

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

Advanced DataTools

IDS 11 – Warehouse 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
- Data Compression

- 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

Advanced DataTools

17

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

Advanced DataTools

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 procedures
- 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

Advanced DataTools

Warehousing Operations

Visualize your current customer script flows for easier troubleshooting and deployment

SQL Transform Operators

Load into target table using target operator

Advanced DataTools
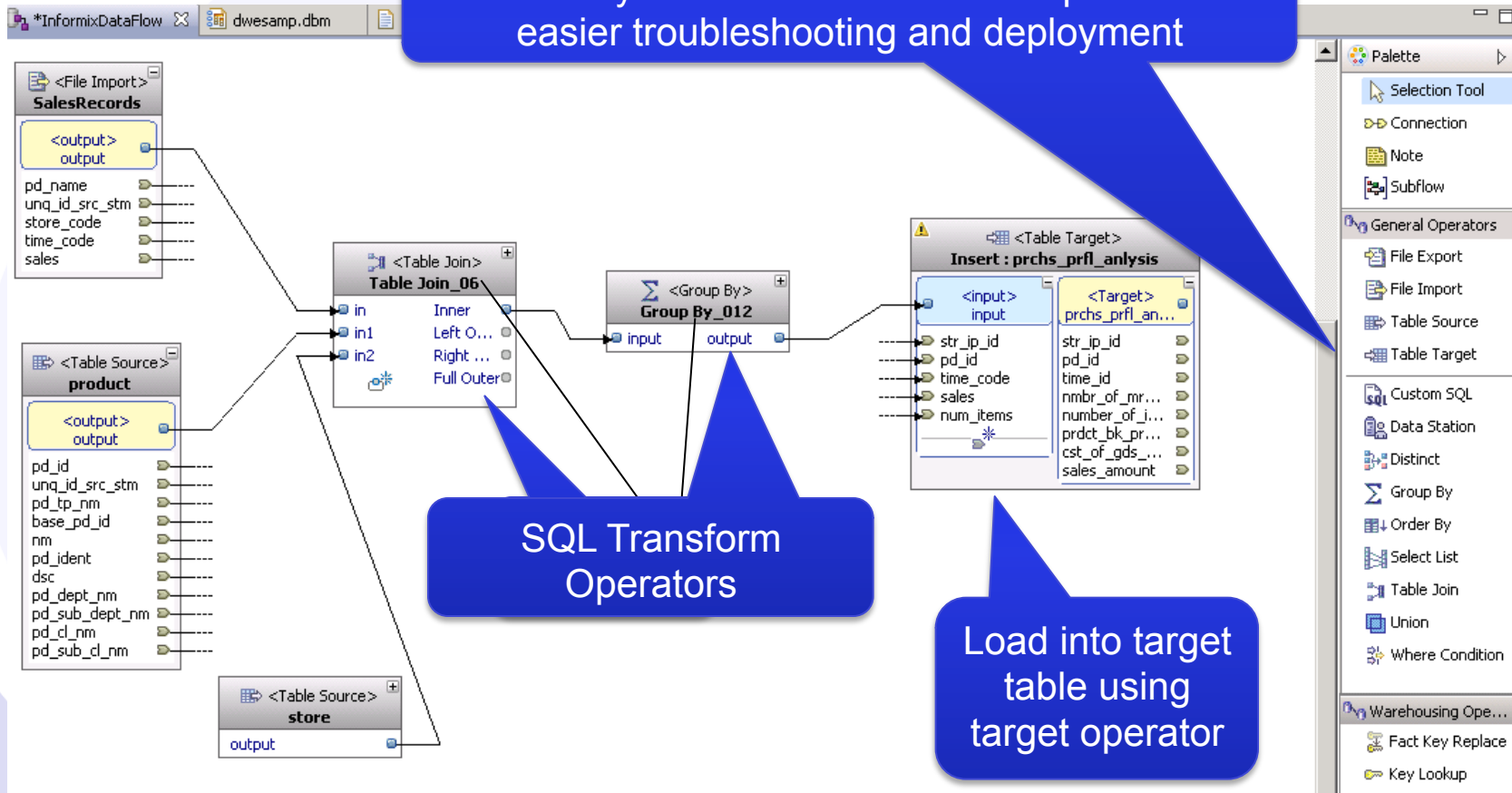
Projects – Extract, Load, Transform Process

- 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"

Advanced DataTools

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
  - From 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

Advanced DataTools

Generate a Warehouse Data Model
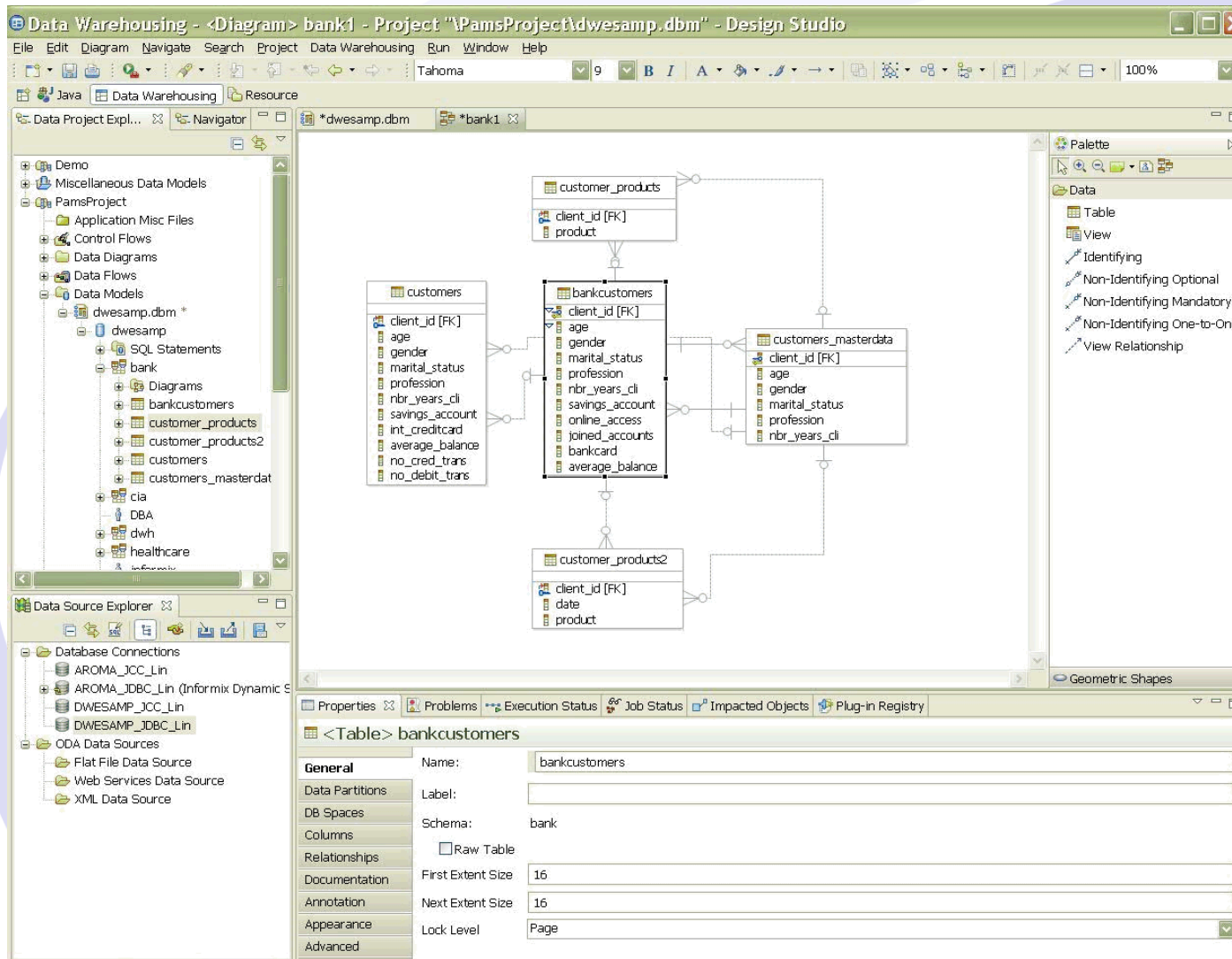
- Visualized data modeling
- Impact analysis
- Reverse engineering or new from scratch
- Compare & sync
- Generated DDL
- Overview diagram

Advanced DataTools

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

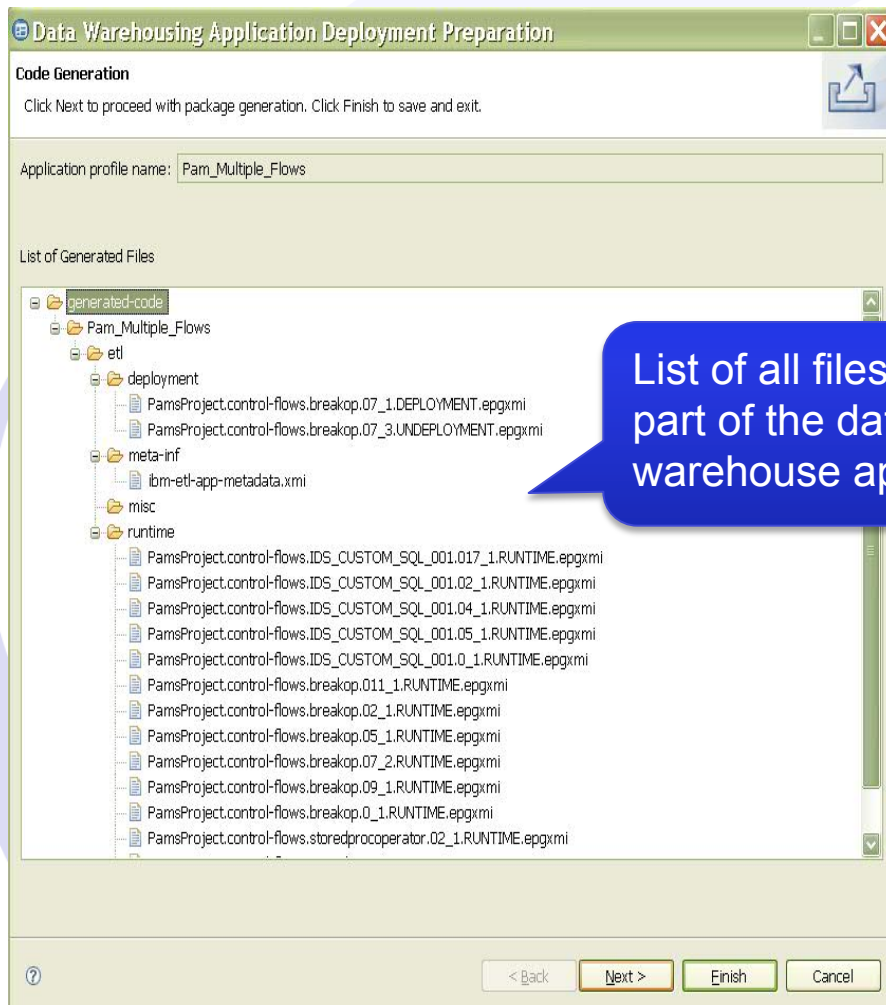**Advanced DataTools**
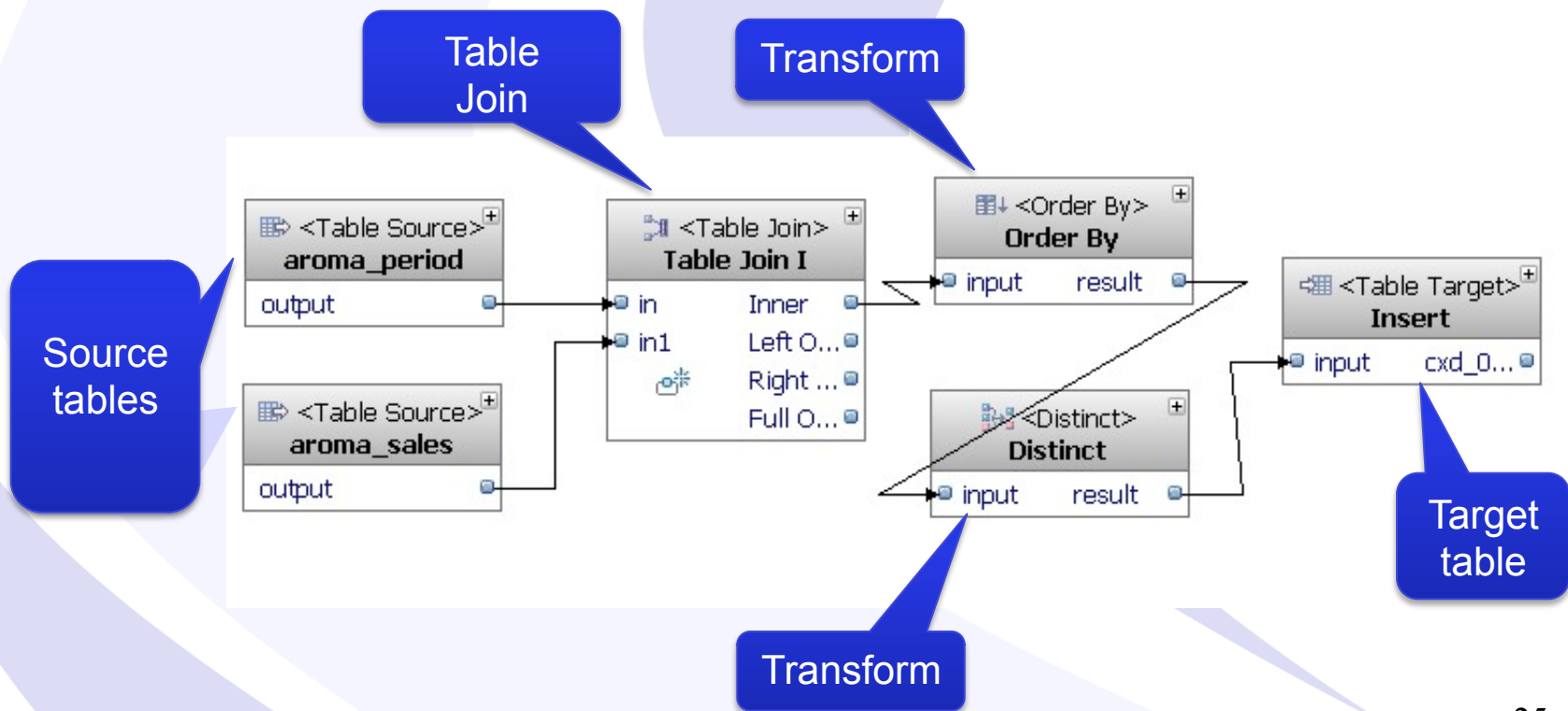
Control Flows

- 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

Advanced DataTools

26

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

Advanced DataTools

Building the Data Warehouse
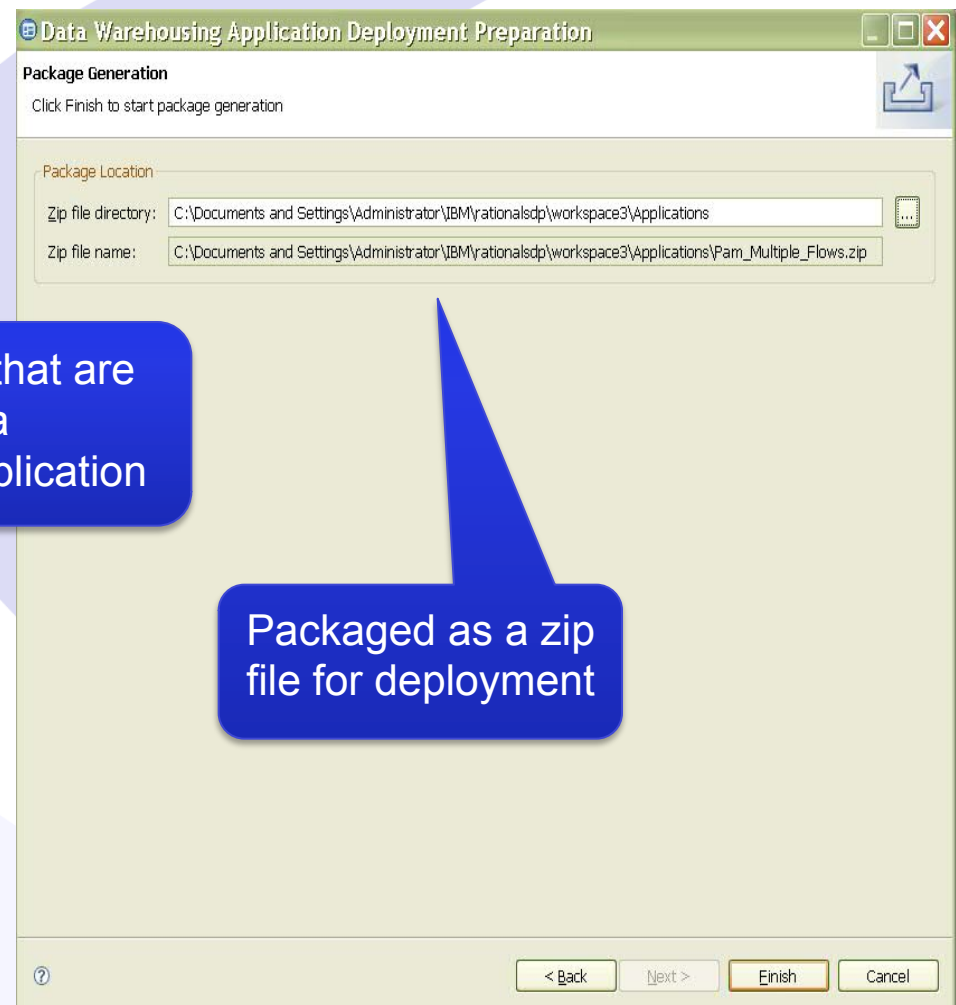
Management and Scheduling

- 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

Advanced DataTools

Informix Warehouse Summary

- 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

Keys to Success

Advanced DataTools

Keys to Success

- Focus scope on one business question
- Requirements Workshop to get consensus on needs and scope
- Build a Prototype – a Model that users can use and try out
- Design an easy to Query schema – Star Schema or Dimensional Model
- Create Summary tables – Pyramid of Summary tables for drill into and roll-up
- Automate data QA in the ETL process – users need to trust the data
- Incorrect data – Fix it in the source system and then let the ETL bring it to the Data Mart
- Disk I/O – must be fast
- Refine, Refine and Refine

Advanced DataTools

Implementation Phases of a Data Warehouse Project

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
Creating a Data Warehouse Prototype	Procuring the Data Warehouse Equipment and Consulting Services	Developing the Data Warehouse Software and Converting the Initial Data	Installing the Data Warehouse Hardware, Software and Converted Data	Training the Data Warehouse Users and Operational Staff	Refining the Data Warehouse Data, Queries and Reports
Data Warehouse Prototype	Contracts and Implementation Plan	Operational Software, Initial Queries, Reports and Data	Data Warehouse Goes Live	Software Documentation and User Manuals	Revised Queries and Reports
6 Weeks	1 Months	3 Months	1 Months	1 Month	3 Months

Advanced DataTools

Data Warehouse
Design Considerations

- What is the "Key Question" you want to ask of your data?

- Organize and summarize your data to answer this question quickly

- Allow for unexpected ad-hoc queries of the data to provide new insight into your business

Advanced DataTools

Build a Prototype in 6 Weeks!

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Creating the Vision and Objectives of the Data Warehouse	Developing the Data Warehouse Design	Converting the Data for the Data Warehouse Prototype	Converting the Data for the Data Warehouse Prototype	Configuring and Loading the Data Warehouse Prototype	Using the Data Warehouse Tools and Presenting the Results
Data Warehouse Launch Scope Document - User Requirements	Data Warehouse Launch Plan Document - Design and Specifications	Converted Data	Converted Data	Prototype Database	Data Warehouse Launch Summary Document - Project Plan

Prototype is a Team Effort

- Our approach requires the full-time involvement of a Project Team

- Customer staff is crucial to the success of the project:
    - Executive Sponsors
    - Data Warehouse Project Manager
   - Data Warehouse Technical Staff
    - End Users

Advanced DataTools

Results of a Prototype

- Converted Actual Data
- Sample Queries and Reports
- Implementation Plans
- A System You can Evaluate

- Excited users who want the system!

Build a Schema - Easy to Query

- Use a Star Schema or Dimensional Model
  - Fact Tables
  - Dimension Tables
  - Summary Tables
- Plan to reuse Dimension Tables

BLM Data Model
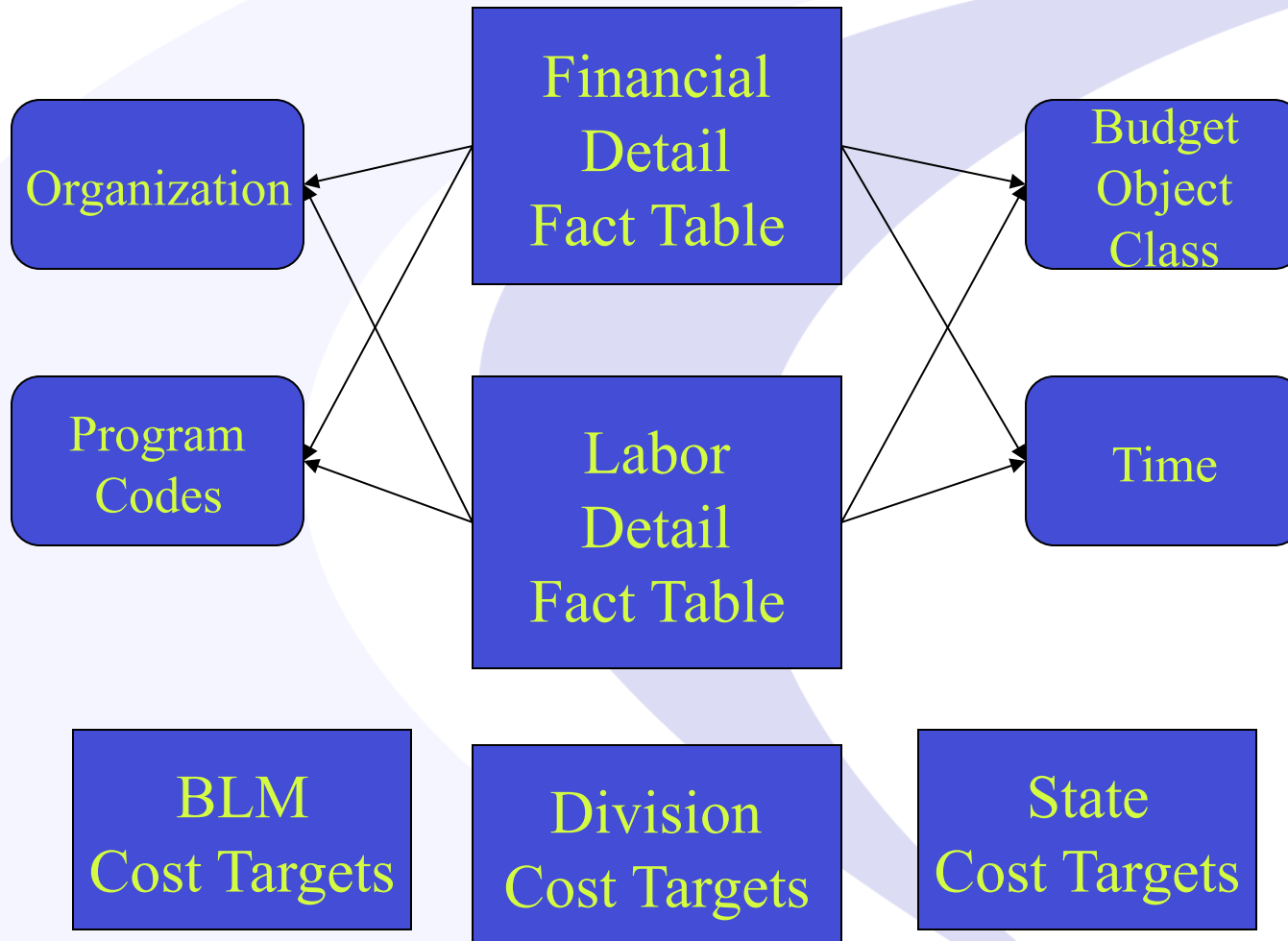
Create Summary Tables

- Pyramid of Summary tables for:
  - Drill down into detail
  - Roll-up to totals

Summary level 2

Summary level 1

Fact Table – billions of rows

Advanced DataTools

BLM Summary Tables

Financial Summary

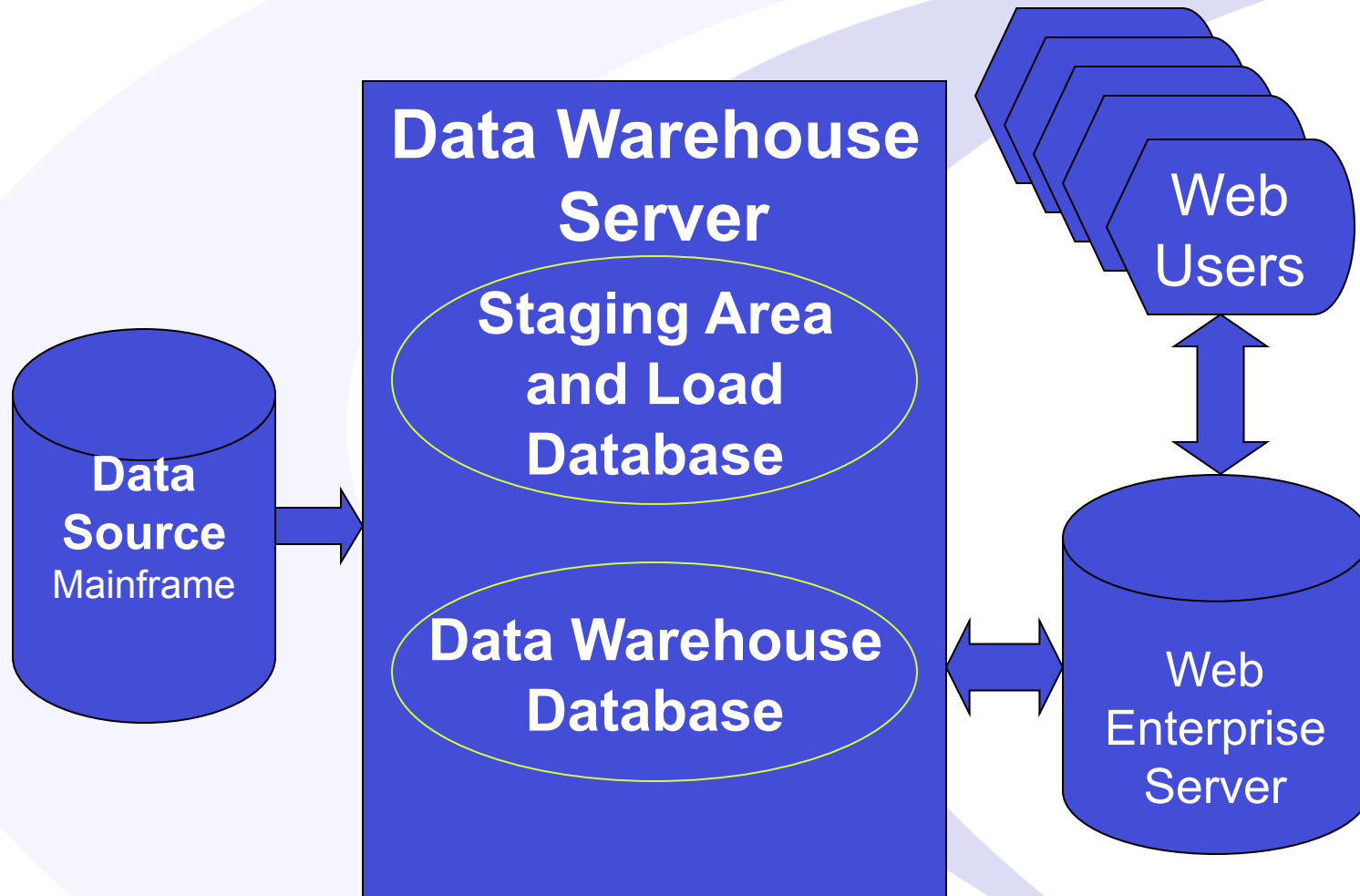BLM Cost Target Summary

Division Cost Target Summary

State Cost Target Summary

- *The goal of the summary tables is to answer 90% of all queries*
- *Saves system resources for handling complex ad-hoc queries*
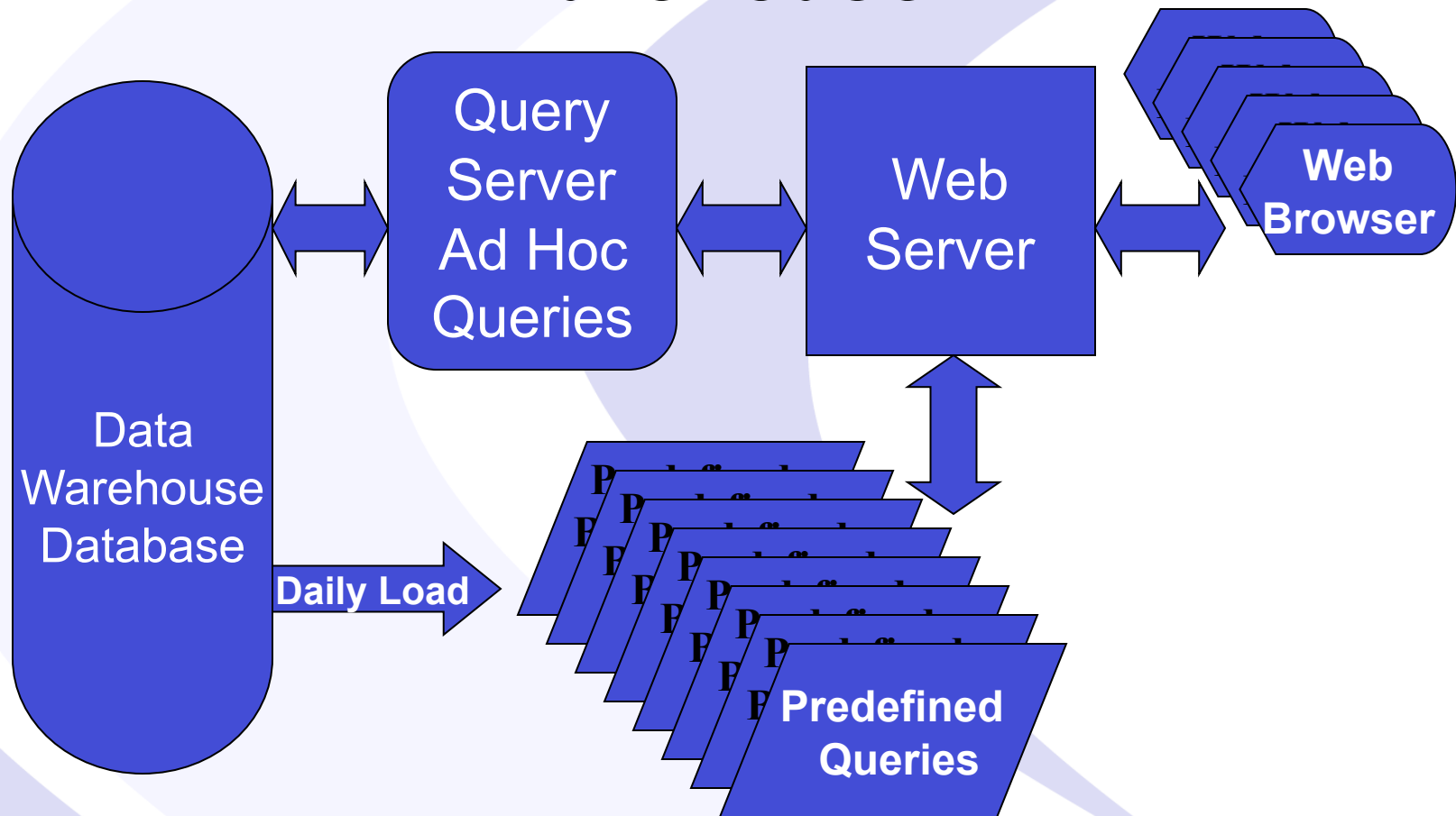
Advanced DataTools

Automate QA in the ETL Process

- Part of the ETL process is to check the data

- Do not load if Source Data is incorrect

- Have a procedure to quickly roll-back bad loads

- Incorrect data – Fix it in the source system and then load it to Data Mart

- *Users must be able to trust the data!*

Advanced DataTools

Data Warehouse Architecture

Data Source Mainframe

Data Warehouse Server
- Staging Area and Load Database
- Data Warehouse Database

Web Users

Web Enterprise Server

Advanced DataTools

43

Web Access to the Data Warehouse

Advanced DataTools

Queries must be *FAST*

- Performance bottleneck will be disk I/O
- Need to Partition the data on disk
- Need to map/layout the data on disk
  - Cannot depend on the SAN to automatically do it
  - Layout based on "How users will query it?"
- Informix IDS 11.5 – Parallel Data Query
- Informix IDS 11.5 – Compression

Advanced DataTools

Steps to Building a Data Warehouse with Informix

- Form a data warehouse team
- Build a working prototype
- Refine the prototype and put it into production
- Load and clean the source data
- Train the users
- Refine, refine, refine

Advanced DataTools

Extracting Value from Data: Building Data Warehouses with IBM Informix Dynamic Server

- Involve users from all levels of the organization. This promotes ownership and successful adoption of the system.
- Most dramatic and successful ideas come from the end users
- Maintain scope and break projects up into small, achievable parts
- Prototypes are no longer "throw-away" in rapid application development
- Refine, Refine, Refine
- ***Insure the Data Warehouse is Fast and Flexible***

Advanced DataTools

Thank You

Lester Knutsen

Advanced DataTools Corporation

lester@advancedatatools.com

Links for more information:

http://www.advancedatatools.com

http://www.informix.com

http://www.ibm.com/informix/warehouse

International Informix User Group http://www.iiug.org

Advanced DataTools