

2.0 Data Warehousing Lifecycle and Project Management

Eugene Rex L. Jalao, Ph.D.

Associate Professor

Department Industrial Engineering and Operations Research
University of the Philippines Diliman

Module 2 of the Business Intelligence and Analytics Track of UP NEC and the UP Center of Business Intelligence

Outline for This Training

- 1. Introduction to Data Warehousing
- 2. DW Lifecycle and Project Management
 - Case Study on DW PM
- 3. Dimensional Modeling
- 4. Designing Fact Tables
- 5. Designing Dimension Tables
 - Case Study on Dimension Modeling
- 6. Extraction Transformation and Loading
 - Case Study on ETL Planning
- 7. Transformation and Loading Methodologies
 - Case Study on ETL



Outline for This Session

- Kimball Lifecycle Overview
- Launching and Managing the BI/DW Project
- Business Requirements Definition
- Technical Architecture Design



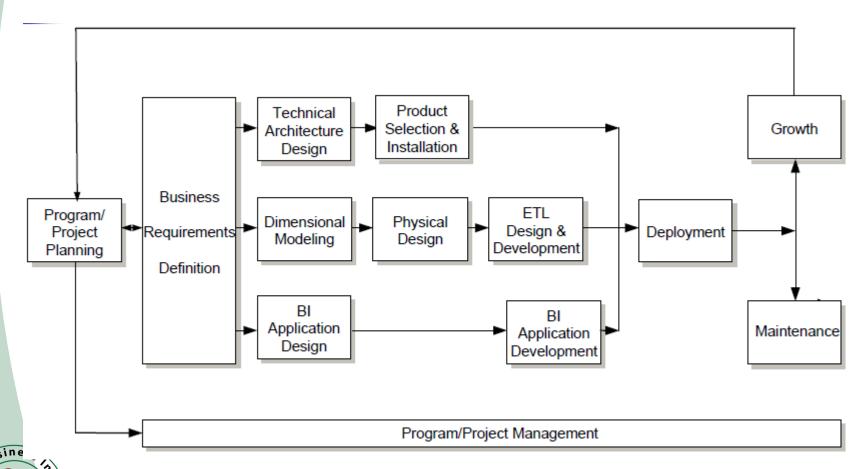


Figure 2.1: Kimball Lifecycle Framework

- Program/Project Planning
 - Define and scope the DW
 - Readiness assessment
 - Resource planning including hardware, software and staffing requirements
 - Define and sequence tasks for entire DW lifecycle
 - Estimate tasks, durations
 - Assign staff to tasks, balance resources
 - Communicate the Project Plan



- Program/Project Management
 - Keep project on track; avoid scope creep
 - Track and resolve issues and bugs
 - Maintain continuous communications
 - Manage expectations
 - Enable creeping commitment
 - Establish and maintain a DW Executive Steering Committee



- Business Requirements Definition
 - Understand the business process
 - Understand business user requirements
 - Business requirements establish foundation for three parallel tracks
 - Data track
 - Technology track
 - Application track
 - Business case and justification



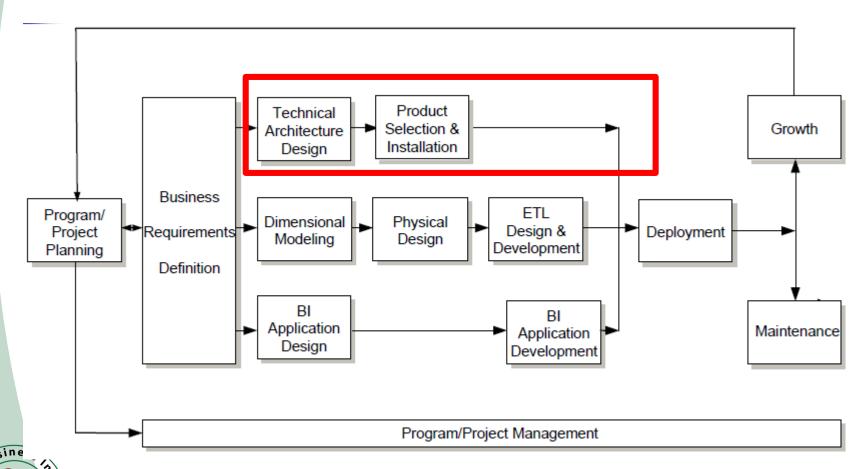


Figure 2.1: Kimball Lifecycle Framework

- Technology Track: Technical Architectural Design
 - Consider three factors simultaneously:
 - Business requirements
 - Current technical environment
 - Planned strategic technical directions
 - Design back room architecture
 - Design ETL (data staging) environment
 - Identify DBMS operating system and hardware environment
 - Design front room architecture
 - Design BI environment
 - Infrastructure and metadata
 - Security requirements



- Technology Track: Product Selection and Installation
 - Evaluate and select:
 - Hardware platform
 - DBMS
 - ETL tool (data staging tool)
 - BI tool (end user data access tool)
 - Install and test to assure end-to-end integration
 - Train team



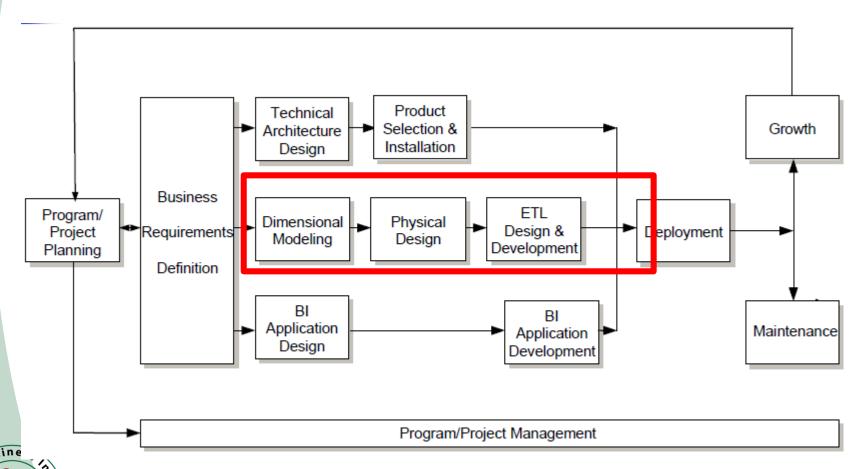


Figure 2.1: Kimball Lifecycle Framework

- Data Track: Dimensional Modeling
 - Identify business processes/events and the associated fact tables and dimensions
 - Construct Business Process/Event v. Common Dimensions Matrix
 - Analyze relevant operational source systems
 - Develop dimensional model:
 - 1. Choose the business process
 - 2. Declare fact table grain
 - 3. Identify the dimensions
 - 4. Identify the facts
 - Develop preliminary aggregation plan

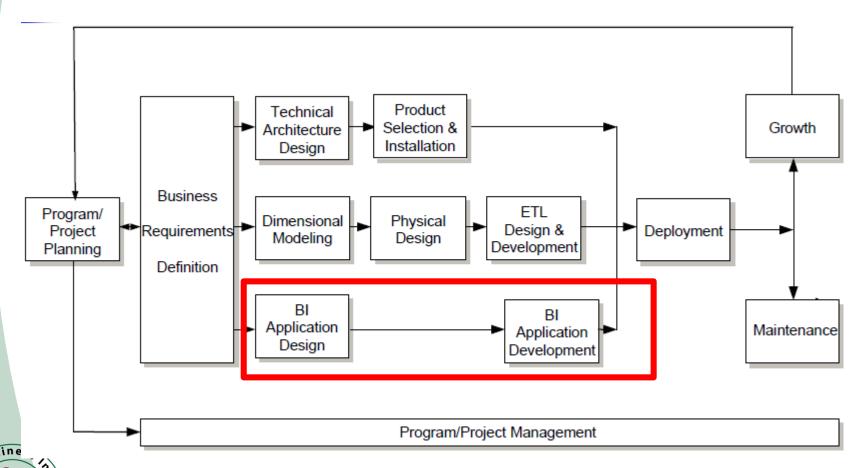


- Data Track: Physical Design
 - Define data naming standards
 - Set up database environment
 - Determine indexing and partitioning strategies



- Data Track: ETL Design and Development
 - Three major steps: Extract, Transform, Load (ETL)
 - Develop source-to-target data mappings
 - Extract data from source operational systems
 - Expose data quality issues buried in source systems
 - Transform to move and clean/correct data
 - Load two staging processes
 - Initial load, including available historical data
 - Incremental loads, often daily
 - Typically underestimated





- Application Track: BI Application Design
 - Identify standard analytic and report requirements to meet 80% –
 90% of user needs
 - Plan and assure ad hoc query and reporting capability
 - Develop report templates for report families
 - Get user signoff on report templates and commit to them
 - Identify metrics and metric calculations, Key Performance Indicators (KPIs)



- Application Track: BI Application Development
 - Ideally, use a single advanced BI tool that meets all user needs
 - Advanced tools provide significant productivity gains for the application development team
 - Good BI design enables end users to modify existing reports and develop ad hoc reports quickly
 - The best tools provide powerful Web-enabled capability



Deployment

- Develop and implement user testing plan
- Develop test protocols to provide thorough, explicit, reusable documents for testing and training
- Obtain user signoff via User Acceptance Test (UAT)
- Develop and implement user training plan
 - Classes
 - Online manual
- Develop and implement user support plan
 - Help desk
 - Problem reporting, tracking, resolution



Maintenance

- Adapt to business changes
- Ongoing user training and support
- Maintain and monitor DW usage statistics
- Purge and archive data



Growth

- Add new business dimensional projects (formerly called data marts)
- Leverage existing dimensions
- Repeat the Lifecycle iteratively for each project



Outline for This Session

- Kimball Lifecycle Overview
- Launching and Managing the BI/DW Project
- Business Requirements Definition
- Technical Architecture Design



Launching and Managing the BI/DW Project

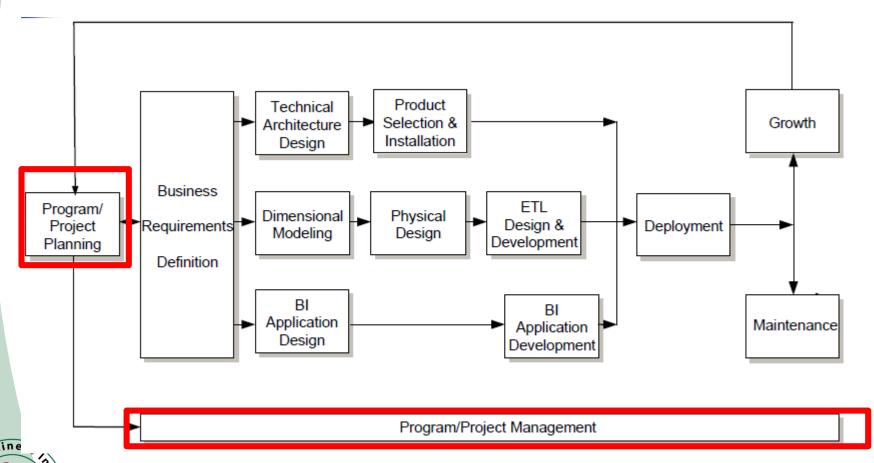


Figure 2.1: Kimball Lifecycle Framework

Launching and Managing the BI/DW Project

Steps:

- Define the Project
- Build the Business Case and Justification
- Plan the Project
- Manage the Project
- Manage the Program



- Assess Your Readiness for a DW project
- Address Shortfalls Risks and Risk Mitigation
- Develop the Preliminary Project Scope and Charter Document



- Assess Your Readiness for DW: Critical Readiness Factors
 - Strong senior business management sponsor(s)
 - Visionary, strategic thinker
 - Willing to be accountable
 - Influential leader, respected by others
 - Politically astute, well connected, "maze bright"
 - Demanding, realistic, supportive
 - Makes things happen
 - Compelling business motivation
 - Feasibility



- Assess Your Readiness for DW: Compelling business motivation
 - Strategic view of the business
 - Driven by competition seeking competitive edge
 - Motivated by industry trends
 - Go beyond today's hot button initiatives CRM, SCM, WA, MDM
 - Strong desire to meet business objectives
 - Eye on the growth and the bottom line



- Assess Your Readiness for DW: Feasibility Factors
 - Primary factor: Quality of the source data (data profiling)
 - Acceptance of the fact that you need to walk before you run agree on a doable project; defer profitability analysis and dashboards until later
 - Willingness to provide the time and money to do it (build the DW)
 right the first time



- Addressing Shortfalls
 - Poor data quality
 - Use a data profiling tool to assess
 - Use the proposed DW project to escalate action
 - Weak business or IT-only sponsor
 - Conduct high-level strategic IT plan study
 - Understand strategic business initiatives
 - Look higher or wider for your champion
 - Go for real thing, not Proof of Concept (POC)
 - POC useful in tool selection, but avoid using it to sell the project



- Addressing Shortfalls
 - Legacy of underperforming, isolated "data warehouse" or "data mart" silos
 - Educate executives and others on need for conformed dimensions and other best practices a la Kimball
 - Consider the past a learning experience
 - Do it right this time



- Addressing Shortfalls
 - Too much demand from business execs
 - Requires prioritization
 - Manage execs who have the lower priority projects –insist rigorously on the "single version of the truth" requirement
 - In search of demand
 - Typical of IT-led initiative
 - Requires user champion
 - Demand from business analysts
 - Requires "selling up"



- Develop Preliminary Scope and Charter: Select Initial DW Project
 - Identify potential business processes
 - Ideally: Select single business process supported by single source system
 - Initial project should be meaningful but doable
 - Select based on business requirements, not deadline



- Develop Project Scope/Charter Document
 - See example in Exhibit 1.0: Project Charter
 - Document Outline:
 - Background/Statement of Need
 - Scope/Project Description
 - Risks and Risk Reduction Plan
 - Scope Exclusions
 - Project Objectives/Success Criteria
 - Milestones and Timelines
 - Project Team Roles and Responsibilities
 - Communication Strategy



Launching and Managing the Project

- Define the Project
- Build the Business Case and Justification
- Plan the Project
- Manage the Project
- Manage the Program



Build the Business Case and Justification

- A: Determine the Financial Investments and Costs
- B: Determine the Returns and Benefits
 - Hard/Tangible Benefits (Financial)
 - Soft/Intangible Benefits
- C: Perform the Cost/Benefit Analysis (ROI)



Build the Business Case and Justification

- Initial Financial Investments and Costs
 - Hardware
 - Software Licenses DBMSs, DW extensions
 - Tools ETL, BI, Data Profiling
 - Internal Staff
 - Consultants
 - Training and Travel



Build the Business Case and Justification

- Ongoing Investments and Costs
 - Ongoing maintenance expenses
 - Ongoing support expenses
 - Expenses to support growth



5-Year Enterprise DW Deployment Map

BASE ATTRIBUTES	Year 1	Year 2	Year 3	Year 4	Year 5
Number of Data Marts	1	2	4	6	6
Number of Sources	3	6	9	12	15
Type of Sources	DDA	Loans	Mortgages	Credit Cards	Trust
Physical Location	CORP	CORP	EUROPE	ASIA	AUSTRALIA
LOB Organization	Retail	Retail	Comm	Comm	Retail
Subject Area	MKTG	SERVICE	FINANCE	SALES	SALES
Size of Database (GB)	100	200	500	800	1,000
USERS					
Corporate	50	100	200	300	400
Europe	10	20	30	30	30
Asia/Pacific	10	20	40	80	160
Total Users	70	140	270	410	590

37

- Determine the Returns and Benefits
 - Reasons for Building a DW
 - Hard/Tangible Benefits (Financial)
 - Soft/Intangible Benefits



- Five Types of Justification
- 1. Cost savings
- 2. Cost avoidance
- 3. Revenue increase
- 4. Cost of doing business
- 5. Basis of doing business



Financial Benefits

- Cost savings and Cost avoidance
 - materials
 - personnel/labor/time
 - inventory
 - facilities
- Revenue increase
 - increase sales
 - increase funding
 - accounts receivable collection



Launching and Managing the Project

- Define the Project
- Build the Business Case and Justification
- Plan the Project
- Manage the Project
- Manage the Program



- Discussion Topics:
 - Project Roles and Staffing
 - Develop the Project Plan
 - Develop the Communication Plan



- Project Roles and Staffing
 - Fans / Front Office / Coaches
 - Business Users
 - Front Office
 - Business Sponsor / Business Driver
 - DW/BI Director / Program Manager
 - Coaches
 - Project Manager
 - Business Project Lead



- Project Roles and Staffing: Regular Line-Up
 - Business Analyst
 - Data Steward / QA Analyst
 - Data Architect / Modeler
 - Data Administrator / DW DBA
 - Metadata Manager
 - ETL Architect / ETL Developer
 - BI Architect / Application Designer / Portal Developer



- Project Roles and Staffing: Special Teams
 - Technical Architect / Technical Support Specialists
 - Security Manager
 - Lead Tester
 - Data Mining / Stats Specialist
 - Educator



- Developing the Project Plan
 - See Exhibit 2.0
 - What to Define
 - Project tasks and subtasks
 - Work Breakdown Structure (WBS) identification
 - Estimated task durations and work
 - Task sequence (predecessors)
 - Available resources project roles and names
 - Task responsibilities by role/name
 - Task start/end dates
 - Gantt chart



- Develop the Communication Plan
 - Project manager develops and maintains the project plan
 - Team members participate in developing the plan, validating their roles, responsibilities, and task time estimates
 - Share a summary of the plan with non-team members business sponsors, business community, and IT colleagues
 - Schedule periodic briefings with your sponsors



Launching and Managing the Project

- Define the Project
- Build the Business Case and Justification
- Plan the Project
- Manage the Project
- Manage the Program



Unique BI/DW Characteristics

- Large, cross-functional teams
- Iterative development lifecycle
- Incredible number of data quality issues exposed
- Elevated visibility

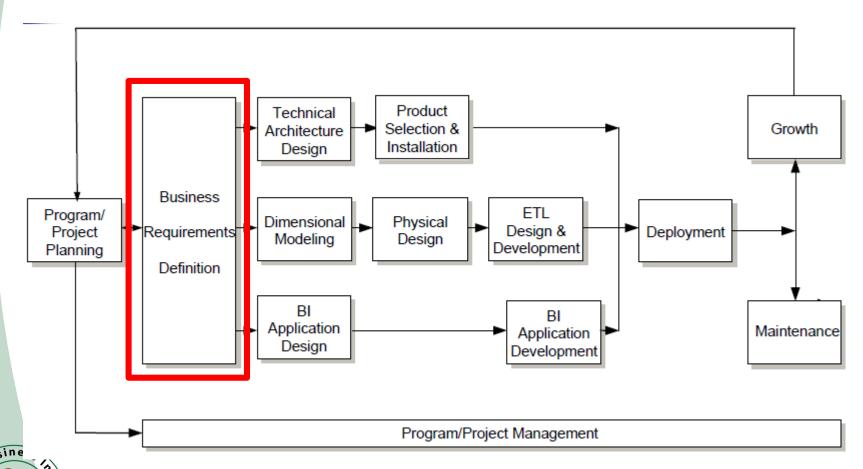


Key Project Management Activities

- Project Kickoff Meeting
- Regular Project Status Meetings
- Regular Project Status Report
- Project Issues / Action Items log, assign, track
- Project Changes log, decide, track



The Kimball Lifecycle



Outline for This Session

- Kimball Lifecycle Overview
- Launching and Managing the BI/DW Project
- Business Requirements Definition
- Technical Architecture Design



Business Requirements Definition: Topics

- Defining Business Processes
 - selected subsets, intuitive
 - comprehensive enterprise-wide, rigorous
- Defining Business Entities
 - selected subsets, intuitive
 - comprehensive enterprise-wide, rigorous
- Enterprise Data Warehouse Bus Architecture Matrix
 - Process to Organization Matrix



DW Bottom Line Objective

Continuously improve business processes for better decisions at all levels of the enterprise!



What is a Business Process?

- Business Process: a logical group of activities and decisions performed by an enterprise
 - Example: The Customer Order Process consists of all the activities and decisions performed by the enterprise to receive an order for products from a customer.
 - Example: The Shipping Process consists of all the activities and decisions performed by the enterprise to fill an order for products from a customer.



About Business Processes

- Processes stretch across functional/departmental lines
- Named in verb object form
- Mutually exclusive one from another
- Defined at a high level using value chain for enterprisewide completeness
- Hierarchical in nature, broken down into subprocesses and sub-subprocesses
- Determined by business knowledge and judgment, not an exact science



Define Business Processes

- Business process definition format
 - <Process Name>
 - The <Process Name> process consists of all the activities and decisions performed by the enterprise to
 - Subprocess A
 - Subprocess B
 - Subprocess C
 - Subprocess D
 - Subprocess E



Sample Business Process Definition

Process Customer Order

- The Process Customer Order process consists of all the activities and decisions performed by the enterprise to receive an order for products from a customer and fill the order. This includes:
 - Receive customer order via Web
 - Verify order info is complete and valid
 - Verify product availability
 - Fill customer order with available products
 - Schedule customer orders for delivery
 - Process back orders
 - Record sales revenue
 - Produce Process Customer Order (sales) reports



Value Chain Template (Exhibit 3)

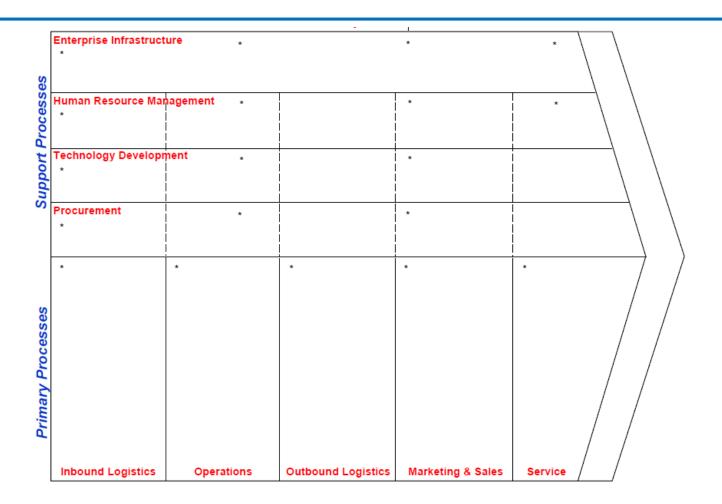




Figure 2.2: Value Chain Template

What Is a Business Entity?

- A Business Entity is something of lasting interest to the enterprise, about which data can be stored, and which is uniquely identifiable.
- Business Entities may be tangible (persons, places, or things) or intangible (concepts or events).
- Examples:
 - Customer, Employee, Location, Product/Service Project, Customer
 Order



Why Business Entities?

- Business entities serve three main purposes
 - First, to manage an entity requires data about the entity
 - Second, entities are a non-redundant way to group data
 - Third, entities are natural foundation for database design –
 operational and data warehouse
- Business entities become the most important dimensions in our data warehouse
 - Recommendation: Identify primary enterprise dimensions during Business Requirements Definition milestone task



Define Business Entities

- The definition of an entity must include a description of the business entity (in sentence form), a unique identifier, and sample data.
- Examples
 - Customer
 - Employee



Business Entity: Customer

- A Customer is any organization or individual that purchases products and/or services from the enterprise.
 Includes prospects and former customers.
- Unique identifier: Customer Number
- Sample data: name, address, phone, customer locations, customer status, credit information,



Ways To Identify Business Entities

- Use source metadata
 - Example: select entities from an entity relationship (ER) diagram
- Use application metadata
 - Example: identify entities on a customer order
- Use the generic entities tool
 - Example: brainstorm entities from a generic list of entities



Some Generic Entities

- Persons:
- Customer
- Employee
- Supplier
- Competitor
- Applicant
- Alliance
- Places:
- Facility/Building/
- Real Estate
- Demographic Area
- Things:
- Product/Service
- Part/Component
 - Material/Supply
 - Physical Asset/
 - Equipment

- Events:
- Customer Order
- Shipping Order
- Customer Invoice
- Requisition
- Purchase Order
- Receipt
- Supplier Invoice
- Disbursement
- Campaign/Promoti on
- Audit
- Project
- Incident
- Work Order
- Test/Inspection

- Concepts:
- Organizational Unit
- Job
- G/L Account
- Financial
 Statement
- Tax Rate Table
- Policy/Procedure
- Law/Regulation
- Contract
- Strategy
- Course
- Community
 Service
- Information Object
- Program



Generic Dimension and Fact Table



Dimension Tables

Persons:

Customer Employee Supplier Competitor Applicant Alliance

Places:

Facility/Building Demographic Area Location

Things:

Product/Service Part/Component Material/Supply Physical Asset/ Equipment

Concepts:

Organizational Unit

Job

G/L Account

Financial Statement

Tax Rate Table

Policy/Procedure

Law/Regulation

Contract

Strategy

Course

Campaign/Promotion

Audit

Project

Fact Tables

Events:

Customer Order

Shipping Order

Customer Invoice

Requisition

Purchase Order

Receipt

Supplier Invoice

Disbursement

Incident

Work Order

Test/Inspection



— h

Enterprise Data Warehouse (EDW) Bus Architecture Matrix

- Picture the flow of processes through a value chain
- Each process can be supported by a DW/BI project, represented by a dimensional model with its fact tables and dimensions
- This intersection of processes and dimensions can be diagrammed in a "Bus" Matrix.
- Other names for the Bus Matrix are:
 - Enterprise Data Warehouse (EDW) Bus Architecture Matrix
 - Process to Dimension Matrix
 - Data Mart to Dimension Matrix
 - Business Process v. Business Entity Matrix



Use of Enterprise DW Bus Architecture Matrix

- Used to dramatize the enterprise-wide need to share data across business processes
 - Note the "density" of the matrix.
- Used to identify and prioritize DW projects
 - Projects are based on business processes
- Used to identify the major dimensions needed for each DW/BI project
- See exhibit 4



Process to Entity Matrix

Process	Entity	Customer	Product	Warehouse	Supplier	Shipper	Customer Order	Invoice	Shipping Order	Purchase Order	Supplier Contract	
Process Customer Order		X	X	Χ		Χ	X	X				
Manage Warehouse Inventory			X	X	X	X			X	X	X	
Ship Customer Order		X	X	X		X	X		X			
Process Purchase Order			Χ	X	X					X	Χ	
Contract Supplier			Χ		Χ						Χ	

Figure 2.3: Process to Entity Matrix

Sample Enterprise Process to Entity Matrix

Process / Data Mart	TIME	CUSTOMER	SERVICE	RATE CATEGORY	LOCAL SERVICE PROVIDER	CALLING PARTY	CALLED PARTY	LONG DISTANCE PROYIDER	INTERNAL ORGANIZATION	EMPLOYEE	LOCATION	EQUIPMENT TYPE	SUPPLIER	ITEM SUPPLIED	VEATHER
Customer Billing	×	х	х	х	Х			Х			х				
Service Orders	×	х	х		Х			Х	х	Х	х	Х			x
Trouble Reports	×	х	х		х	х		Х	х	Х	х	Х	х	х	x
Yellow Page Ads	×	X		х		х			х	Х	х				
Customer Inquiries	×	Х	X	X	×	X		×	X	Х	Х				x
Promotions & Communication	×	Х	X	X	×	X		X	Х	X	Х	×	Х	×	
Billing Call Detail	×	х	X	X	×	X	X	X	Х		Х	×	X	X	x
Network Call Detail	×	х	X	X	×	X	X	X	Х		X	×	X	X	x
Customer Inventory	×	х	X	X	×			X	х		х	×	X	X	
Network Inventory	×		х						х	X	х	X	х	X	
Real Estate	×								х	X	х	X			
Labor & Pagroll	×								х	X	х				
Computer Charges	X	х	х		X			Х	х	X	х	X	х	X	
Purchase Orders	X								х	Х	х	X	х	х	
Supplier Deliveries	X								х	Х	х	X	х	х	
Combined Field Operations	X	х	x	X	х	X		х	х	х	х	X	х	X	x
Customer Relation Management	X	х	x	X	X	X	х	х	х	х	х	X	х	X	X
Customer Profit	X	х	х	х	х	х	х	Х	х	Х	х	Х	х	х	х



Outline for This Session

- Kimball Lifecycle Overview
- Launching and Managing the BI/DW Project
- Business Requirements Definition
- Technical Architecture Design



The Kimball Lifecycle

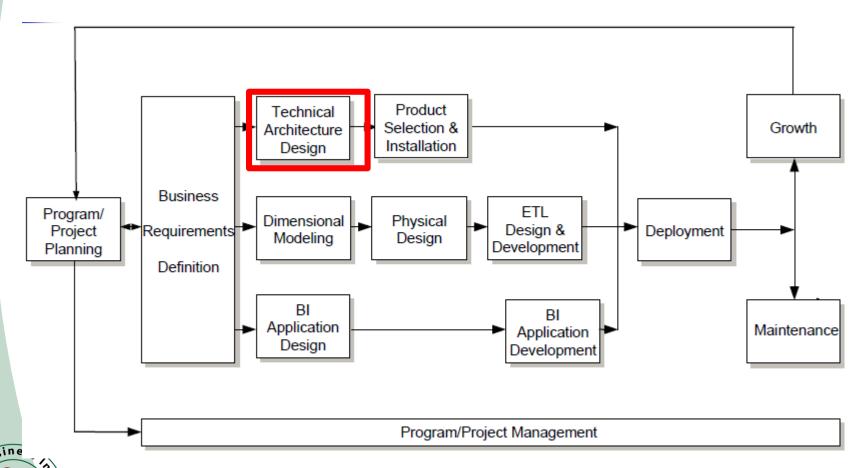


Figure 2.1: Kimball Lifecycle Framework

The Value of Architecture

Communication

- "Show me a picture" diagrams, slides
- Management, business users, IT, DW/BI team

Planning

 Drill down to greater levels of detail

Flexibility and maintenance

- A baseline
- Analyze changes, extensions
- Learning
 - New DW/BI team members
 - User training
- Productivity
 - Saves DW team time



The Value of Architecture

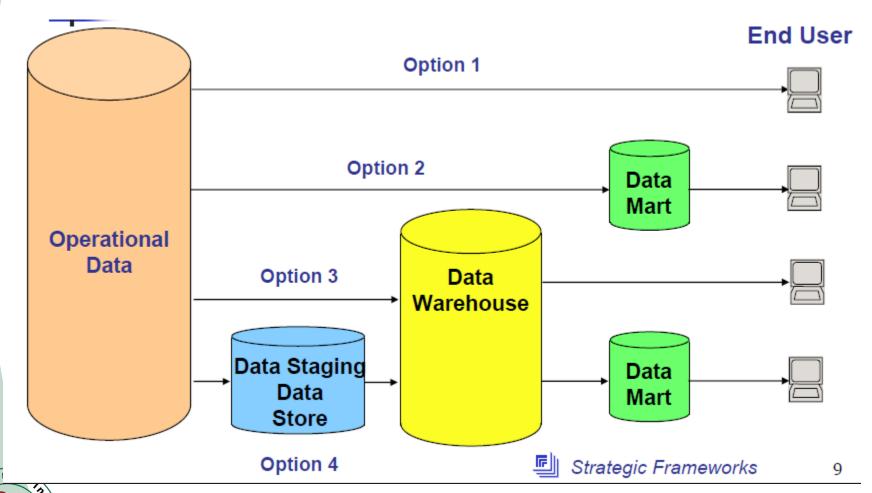
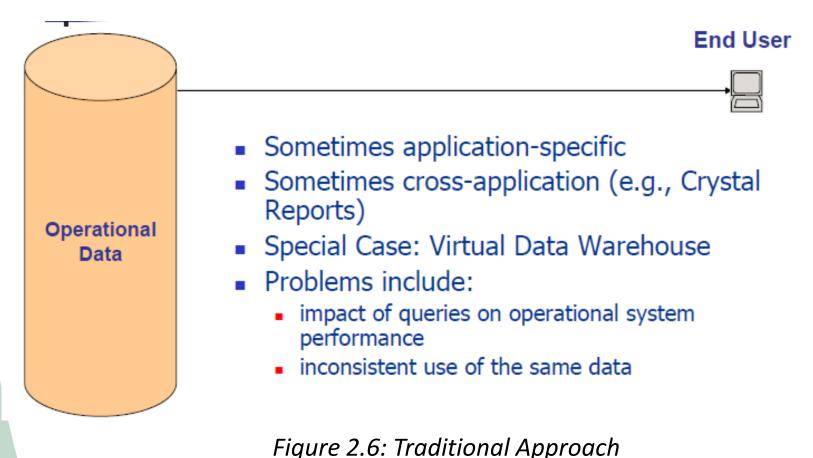


Figure 2.5: Implementation Architectures

Option 1: Traditional Application Reporting Tools





The Virtual Data Warehouse

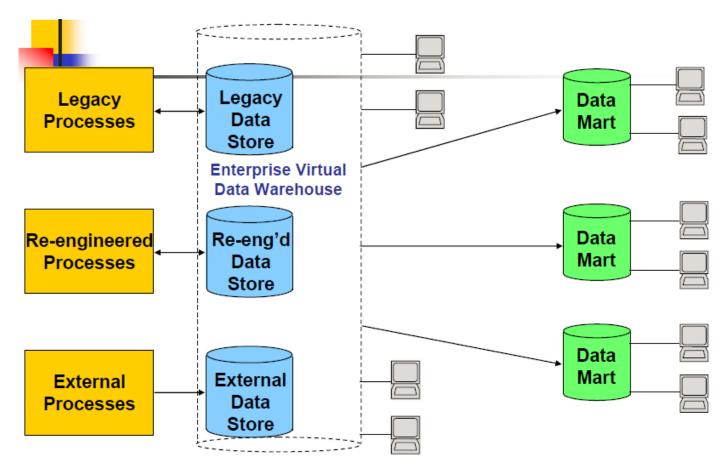




Figure 2.7: Virtual Data Warehouse

Option 2: Data Mart Approach

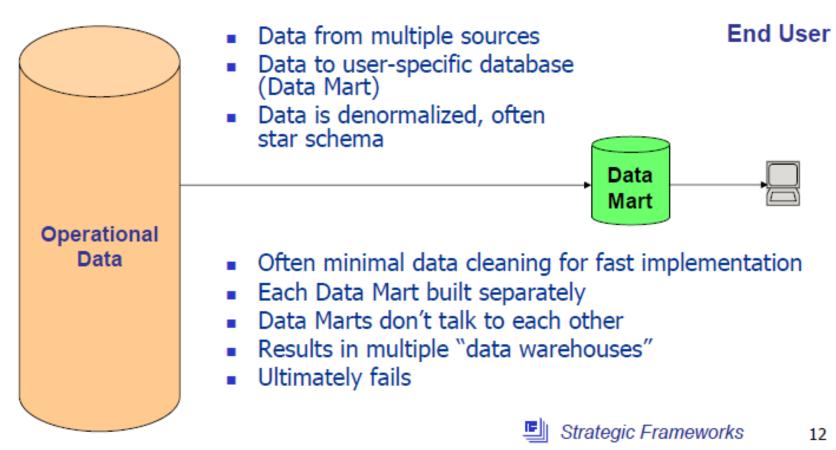




Figure 2.8: The Data Mart Approach

Option 3: ODS Data Warehouse

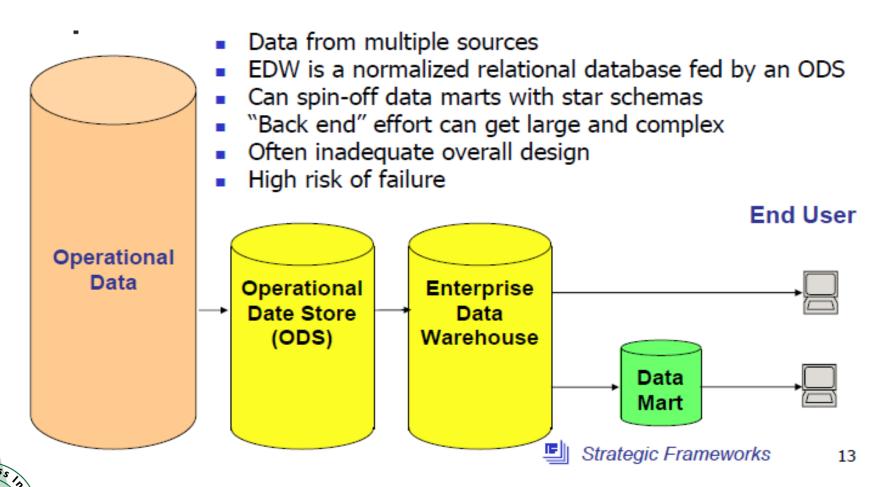


Figure 2.9: The ODS Approach

Option 4: Enterprise DW with Data Staging and Data Marts

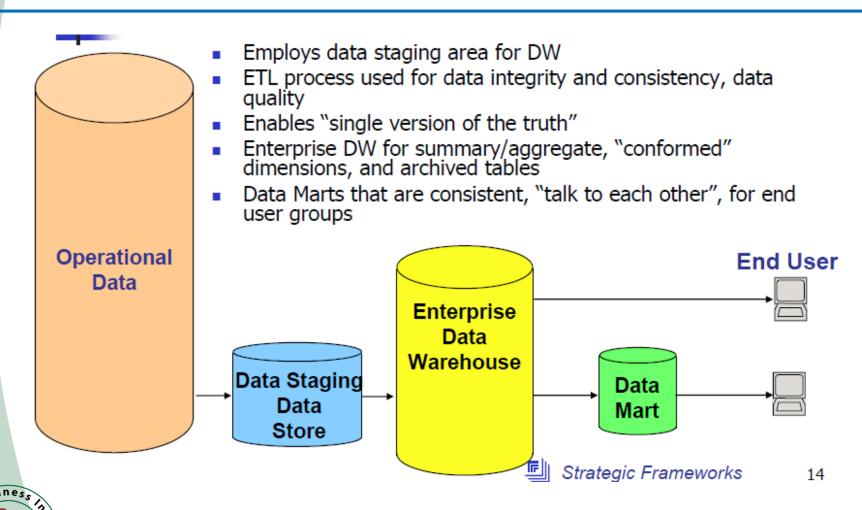


Figure 2.10: Enterprise DW with Data Staging and Data Marts

Technical Architecture Overview

- Back room and front room, which share the presentation server
- Back room ("kitchen")
 - Extracts data from a variety of sources
 - Internal workings of the ETL system not seen by users
 - Cannot be queried
- Front room ("dining room")
 - Contains the query-able data warehouse database
 - Users see the data by using BI applications
- Meta data driven
- Stores and services
 - Stores are the temporary or permanent data storage locations
 - Services are the DW functions, e.g., copying data from store to store



High-Level DW/BI System Architecture Model

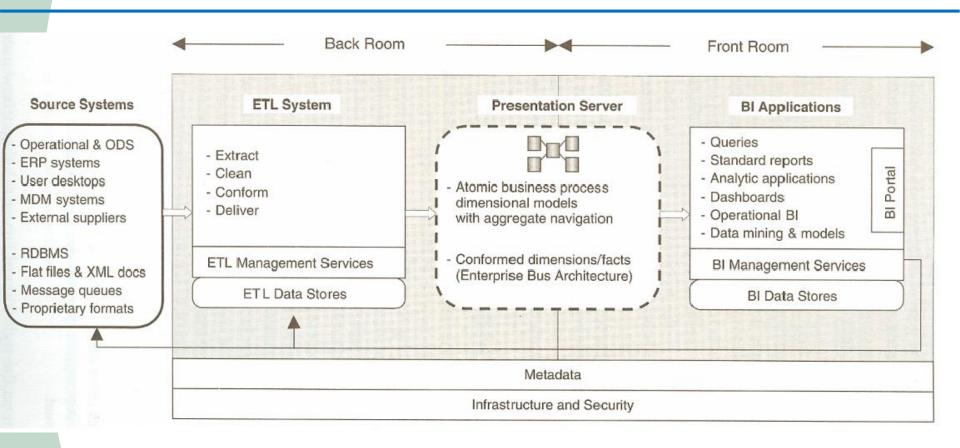




Figure 2.11: High-Level DW/BI System Architecture Model

The Great Religious Modeling War: Inmon vs. Kimball

- Bill Inmon (EDW/DM)
 - The EDW should be in at least 3rd normal form.
 - But the data marts should be in dimensional form.
- Ralph Kimball (Architected EDW)
 - The EDW is based on dimensional models integrated by the conformed dimension bus
 - There is no need for an ODS as a prerequisite to a data warehouse. But if you have one already, use it as another source for your data warehouse.
- Mike Schmitz (Dimensional Normal Form EDW)
 - The EDW should be in dimensional normal form (DNF), that is, E/R analyzed and implemented with formal dimension families.
 - It should be accessible to the business person. The presentation layer can be co-located with the data staging area.
 - Agrees with Kimball on role of an ODS



Source Systems

- Definition: An operational system of record whose function it is to capture the transactions of the business/enterprise.
 - Often a mission-critical legacy mainframe system
 - Main priorities: uptime and availability
 - Maintains little historical data
 - Management reporting is a burden



Source Systems

Operational systems

 Examples: order entry, production, shipping, customer service, accounting, ERP systems

Desktop data

 Examples: budget plans and forecasts, semi-static files (product hierarchies, customer types), contracts/agreements

External systems

 Examples: customer demographic data (U.S. Census, etc.), mailing lists, zip codes, customer business segments (SIC codes), competitive sales and financial data, supplier catalogs and order data



Source System Formats

Source Systems

- Operational & ODS
- ERP systems
- User desktops
- MDM systems
- External suppliers
- RDBMS
- Flat files & XML docs
- Message queues
- Proprietary formats

- RDBMS
- Flat files
- XML documents
- Message queues, log files, redo files
- Proprietary formats



Backroom Architecture

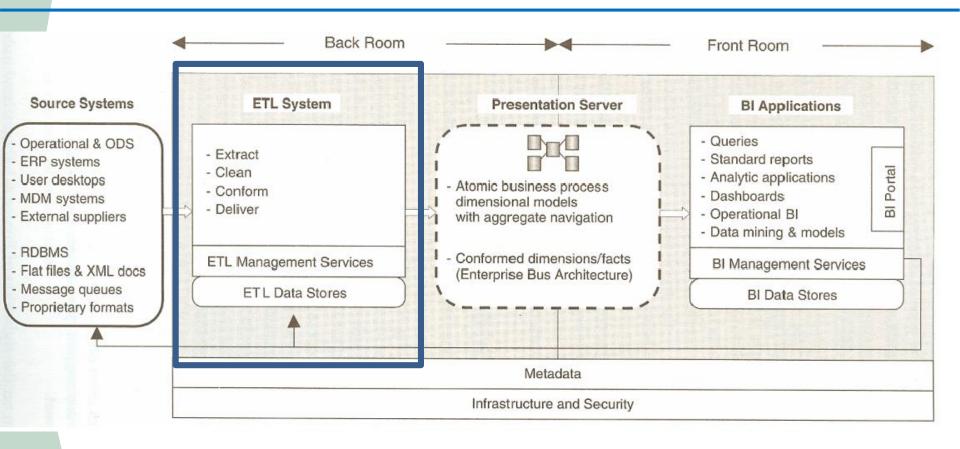




Figure 2.11: High-Level DW/BI System Architecture Model

Back Room Overview

- Also Known As: data staging area
- ETL involves four major operations:
 - Extract data from source systems
 - Clean and Conform data as it is moved from source to target data stores. Commonly called Transform .
 - Deliver transformed data to the presentation server. Commonly called Load.
- ETL Management Services which handle all the services that support the ETL process



ETL Data Stores

- ETL System Data Stores
 - Master copies of source data
 - Intermediate ETL tables
- Lookup and Decode Tables
 - Content enhancement lookup tables
 - Hosted hierarchies for organizations, products
- Data Quality Data Stores
 - Audit tables
 - Error tables



Presentation Server

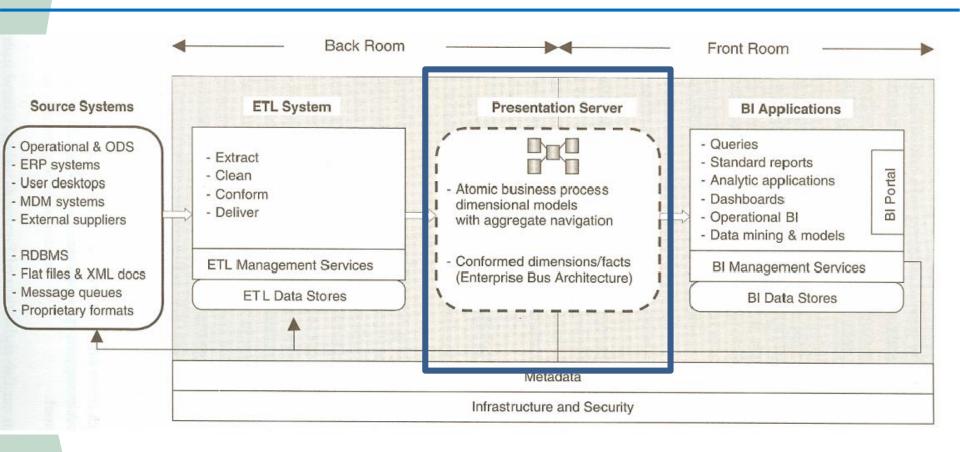




Figure 2.11: High-Level DW/BI System Architecture Model

Presentation Server

- Definition: The target physical machine[s] on which the data warehouse data is organized and stored for direct querying by end users, report writers, and other applications
 - Must present data in a dimensional framework
 - If based on a relational database (RDBMS), then tables organized as star
 schemas
 - If based on multidimensional database (MDDB) (nonrelational, on-line analytic processing OLAP), then will have cubes based on dimensions and facts



The Kimball Lifecycle

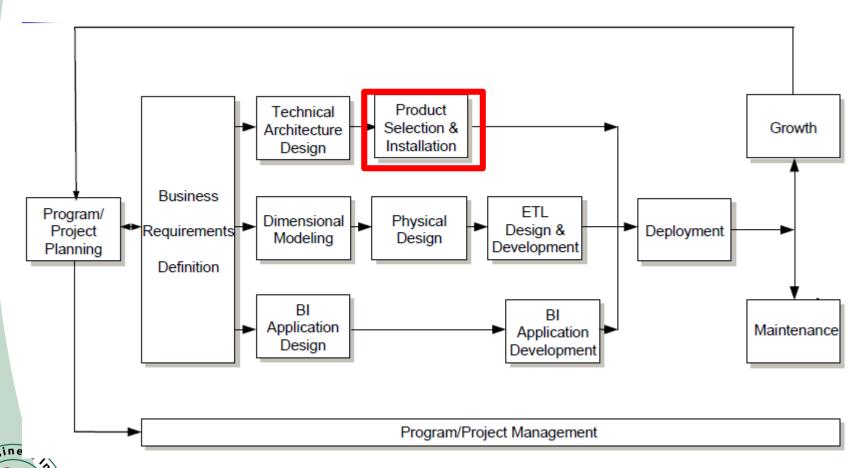


Figure 2.1: Kimball Lifecycle Framework

Presentation Server

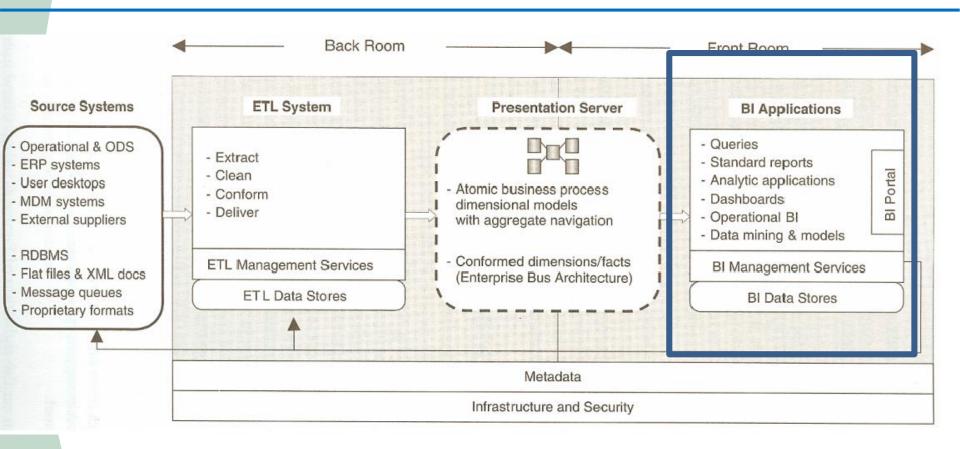




Figure 2.11: High-Level DW/BI System Architecture Model

Tool Selection Methodology

- Define Selection Criteria
- Research Tools
- Select the Top 3 Tools in Each Category
- Hold Product Presentations and Demonstrations
- Prototype/Proof of Concept (POC)
- Choose the Top Tool
- Post Implementation Review



Product Selection Criteria

Vendor Support

- Documentation
- Training
- Technical Support
- Reputation
- Consulting
- Technical Criteria
 - Total Cost of Ownership (TCO)
 - End User Evaluation (data access products)



Selecting the DW Products

- The four major product decisions
 - 1. Hardware/OS platform
 - 2. DBMS
 - 3. ETL tool
 - 4. BI tool(s)



Platform Requirements

- Data Warehouse Databases
 - Support both relational and multidimensional databases
 - Support distribution of data across both types
- OLAP
 - Must support cube functionality
- Data Mining
 - Now becoming mainstream
- Interfaces
 - To relational, multidimensional, and data mining databases
- Build and Manage Capabilities
 - Able to extract from variety of source systems
 - Built-in ETL tools



Complete BI Platforms

- Three BI platforms meet the criteria:
 - IBM DB2/UDB
 - Oracle
 - Microsoft SQL Server



Packaging and Pricing

Microsoft

- \$20K per processor
- OLAP, data mining, build/manage bundled

Oracle

- \$40K per processor
- OLAP, data mining, build/manage priced separately

IBM

- \$25K per processor (UDB), DB2 slightly higher
- Build/manage (Data Warehouse Center) included with UDB
- OLAP, data mining priced separately



Applicability to DW Size and Function

MS SQL Server

- DW size low-end, middle, touches high-end
- For Windows-based servers

Oracle

- Best for high-end
- Rich features bitmapped indexes, partitioning, join capabilities
- UNIX and Windows

IBM

- Low to high end
- Versions for Windows, Unix, and IBM's proprietary S/390
- mainframe and AS/400 midrange



References

- Simon, Alan. CIS 391 PPT Slides
- Tan et al. Intro to Data Mining Notes
- UCI Irvine Data Warehousing Notes
- Analytic Platforms (2010) http://vertica.com/wpcontent/uploads/2010/12/beyond-traditional-datawarehouse.pdf



Case Study 1

Project Management for DW

