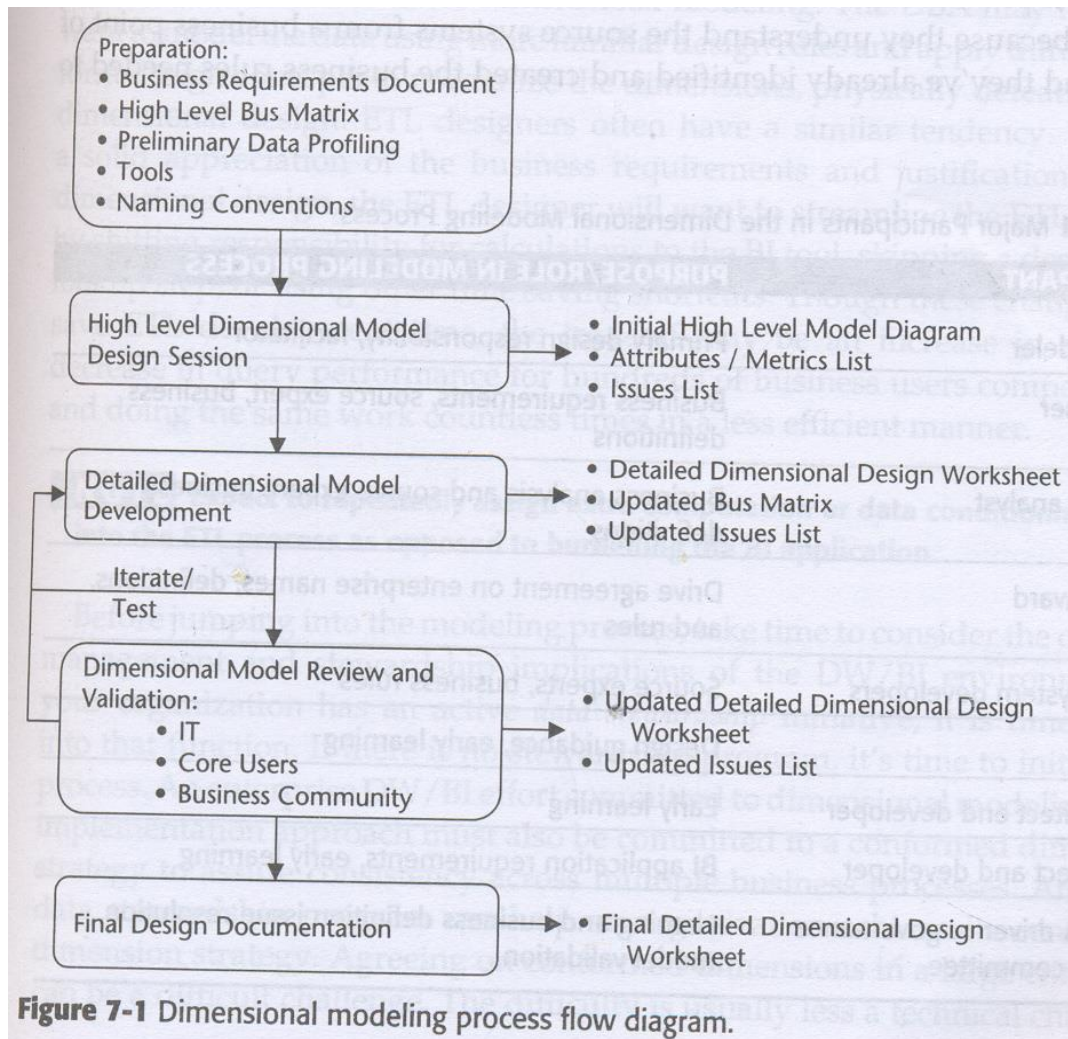# Designing the Dimensional Model

Modeling Process Overview

- Graphical design from bus matrix

- Scope of design

- Identify grain of facts

- Identify dimensions

- Reviewing and validation of models

- Models evolve through stages

- Tested against business requirements

- Three to four weeks but vary with complexity

Preparation:
- Business Requirements Document
- High Level Bus Matrix
- Preliminary Data Profiling
- Tools
- Naming Conventions

High Level Dimensional Model Design Session
- Initial High Level Model Diagram
- Attributes / Metrics List
- Issues List

Detailed Dimensional Model Development
- Detailed Dimensional Design Worksheet
- Updated Bus Matrix
- Updated Issues List

Iterate/ Test

Dimensional Model Review and Validation:
- IT
- Core Users
- Business Community
- Updated Detailed Dimensional Design Worksheet
- Updated Issues List

Final Design Documentation
- Final Detailed Dimensional Design Worksheet

Figure 7-1 Dimensional modeling process flow diagram.

Identify Design Participants

Table 7-1 Major Participants in the Dimensional Modeling Process

PARTICIPANT	PURPOSE/ROLE IN MODELING PROCESS
Data modeler	Primary design responsibility, facilitator
Power user	Business requirements, source expert, business definitions
Business analyst	Business analysis and source expert, business definitions
Data steward	Drive agreement on enterprise names, definitions, and rules
Source system developers	Source experts, business rules
DBA	Design guidance, early learning
ETL architect and developer	Early learning
BI architect and developer	BI application requirements, early learning
Business driver or governance steering committee	Naming and business definition issue resolution, model validation

- Dimension manager
- -defining, building and publishing one or more conformed dimensions
- -can have multiple Dimension manager
- -work with data stewards

- Fact Provider
- -designing ,building and publishing a specific fact table on consulting dimension manager

Revisit the Requirements

- Review requirement s documentation
- Modeling team must understand challenges and opportunities the business users face
- Modeling team to translate business requirements into flexible model that can support range of analysis
- Data modeler should understand business and review business requirements before conforming data model
- Requirement document includes list of proposed data elements, sample questions and desired reports.
- Don't move to designing skipping business requirements review

Use Modeling Tools

- Use spreadsheets as initial modeling tool so as to change it quickly.

- Use specialized modeling tools at later stages.

- Advanced tools features are capture information in their metadata stores and allows it to export to other formats

- Helps DBA to forward engineer the model into database like creating tables ,indexes,views etc

Promotion

Attribute Name	Description	Alternate Names	Sample Values
Special Offer ID	Source system key		
Special Offer Name	Name / description of the Special Offer	Promotion name, Special offer description	Volume Discount 11 to 14; Fall Discount 2006
Discount Percent	Percent item is discounted		
Special Offer Type	Description of the type of promotion, special offer or discount.	Promotion Type	Volume Discount; Discontinued Product
Special Offer Category	Channel to which the Promotion applies	Promotion Category	Reseller; Customer
Start Date	First day the promotion is available		6/15/2008
End Date	Last day the promotion is available		12/31/2008
Minimum Quantity	Minimum quantity required to qualify for the promotion		0
Maximum Quantity	Maximum quantity allowed under the promotion		NULL

Attributes Not Elsewhere Classified

Attribute Name	Description	Alternate Names	Sample Values
SalesReasonID	Sales reason ID from source system		
Sales Reason	Reason the customer bought the product, as reported by the customer (Internet only)		Demo Event; On Promotion; Price; Review; Sponsorship
Sales Reason Type	Grouping for Sales Reason		Marketing; Promotion; Other
Channel	Channel through which the item was sold		Customer; Reseller; Field Sales

Figure 7-2 Example attribute and metrics list.

Establish Naming Conventions

- Names should be descriptive and consistent
- Agreeing on common definitions and common labels
- Document naming conventions
- Column names should be like

  PrimeWord_zeroOrMoreQualifies_ClassWord

-fields in sales fact table representing amount sold

  sales_dollar_amount

Seek agreement on naming conventions

Provide For Source Data research and Profiling

- To enhance source data understanding

- Resources include source systems, data expert, query and reporting systems, data profiling tools

- Data profiling uses query and reporting tools to explore content and relationships in system

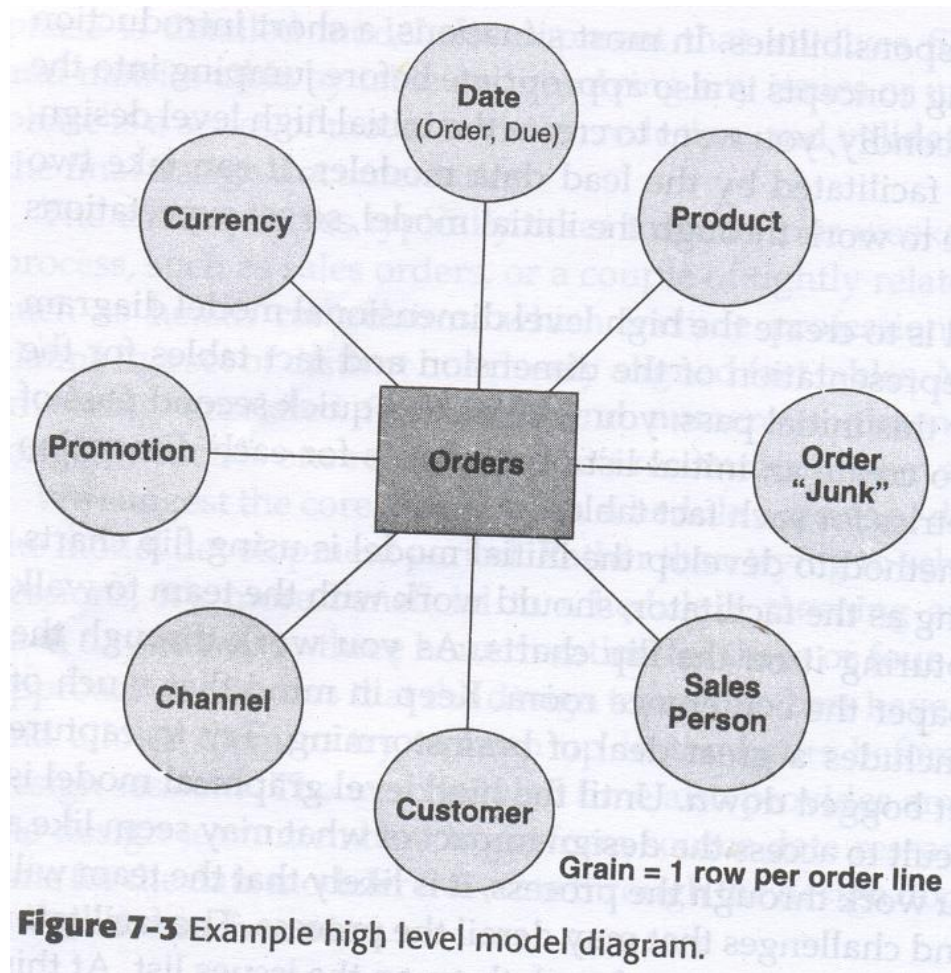# Four Step Modeling Process

- Choose business process

- Declare grain

- Identify the dimensions

- Identify the facts

Design Dimensional Model

- Iterative process
- Based on Business user and source files
- Change model as you learn
- Identify another model and grain adjustments
- Three phases
- First is High level design session
- -defines scope of business process dimensional model
- Second is Detailed model development
- -filling attributes and metrics
- Third is model reviews, redesign and validation step

Build High level Dimensional Model

- Conduct initial Design session
- -bring together core modeling team with interested participants from source and ETL team
- -discuss project objectives, scope and roles and responsibilities.
- -to create high level dimensional model diagram
- -after that create list of attributes for each dimension table and metrics for each fact table

Figure 7-3 Example high level model diagram.

The diagram shows a central box labeled "Orders" connected to surrounding circles:
- Date (Order, Due)
- Product
- Order "Junk"
- Sales Person
- Customer
- Channel
- Promotion
- Currency

Grain = 1 row per order line

- Document high level model Diagram
- -handy introduction to design while communicating about project, its scope and contents
- -facilitates discussions between design team and business partners
- -useful training tool

- Identify the attributes and Metrics
- -initial attribute and metric list
- -utilize spread sheet
- -issues list for any controversies

Develop detailed Dimensional Model

- -about enriching high level model with missing info, design issues, and constantly testing the business requirements
- -determine source locations, definitions, and business rules that specify how these attributes and metrics are populated
- Identify the data sources
- Understand candidate data sources
- -through business requirements, dimensions in bus matrix
- Profile and select data sources
- -evaluate and select using documentations, data models, record layouts etc
- -data profiling

- Outcomes of data profiling

- Select reject decision

- Source system data quality issues

- Data quality at ETL stage

- Business rules foreign and primary key relationship

- Establish Confirmed dimension
- -data stewards and business analyst should decide table and attribute naming, descriptions and definitions
- -multiple sources for some dimensions like customer
- -customer for sales, marketing
- Identify base and derived facts
- -two types of derived facts as well as base facts
- -additive base and derived fact

Document Detailed Table designs

- Use spreadsheet to store attribute and metrics list
- Purpose is to capture details design for communication to other interested stakeholders including business users, BI application developers and ETL developers.
- Each dimension and fact table should have separate detailds design
- It includes attribute/fact name, description, sample values, slowly changing dimension type indicators for every dimension attribute
- Also includes rules for each fact to indicate its additive, semiadditive or non additive
- Should capture attributes and facts desired by user but not available

- Update Bus Matrix

- -Detailed modeling leads to introduction of new fact table, new dimensional table, splitting or combining of dimensions

- -Acts as communication tool for design teams

- Identify and resolve modeling issues

Business Process	Fact Tables	Granularity	Facts	Date	Policyholder	Coverage	Covered Item	Employee	Policy	Claim	Claimant	3rd Party
Policy Transactions	Corporate Policy Transactions	1 row for every policy transaction	Policy Transaction Amount	X Trxn Eff	X	X	X	X	X			
	Auto Policy Transactions	1 row per auto policy transaction	Policy Transaction Amount	X Trxn Eff	X	X Auto	X Auto	X	X			
	Home Policy Transactions	1 row per home policy transaction	Policy Transaction Amount	X Trxn Eff	X	X Home	X Home	X	X			
Policy Premium Snapshot	Corporate Policy Premiums	1 row for every policy, covered item, and coverage each month	Written Premium Revenue Amount, Earned Premium Revenue Amount	X	X	X	X	X Agent	X			
	Auto Policy Premiums	1 row per auto policy, covered item, and coverage each month	Written Premium Revenue Amount, Earned Premium Revenue Amount	X	X	X Auto	X Auto	X Agent	X			
	Home Policy Premiums	1 row per home policy, covered item, and coverage each month	Written Premium Revenue Amount, Earned Premium Revenue Amount	X	X	X Home	X Home	X Agent	X			
Claim Transactions	Claim Transactions	1 row for every claim transaction	Claim Transaction Amount	X Trxn Eff	X	X	X	X	X	X	X	X
	Claim Accumulating Snapshot	1 row per covered item and coverage on a claim	Original Reserve Amount, Assessed Damage Amount, Reserve Adjustment Amount, Current Reserve Amount, Open Reserve Amount, Claim Amount Paid, Payments Received, Salvage Received, Number of Transactions	X	X	X	X	X Agent	X	X	X	
	Accident Event	1 row per loss party and affiliation in an auto claim	Implied Accident Count	X	X	X Auto	X Auto		X	X Auto	X	

Figure 7-7 Example detailed bus matrix.

Chng Flag	Issue #	Task / Topic	Issue	ID Date	Rptd By	Resp	Date Closed	Status	Priority
	27	Employee	Research availability of historical data.	2/18/2008	Team	BH	-	Open	High
	28	Sales Territory	Research relationships and history among Sales Territory, Sales Rep, and Customer tables in the source system.	2/18/2008	Team	RB	-	Open	High
	29	Customer	Assess projected impact of combined Field Sales, Internet and Reseller on data size and growth of Customer table.	2/18/2008	Team	BB	-	Open	High
	30	Customer	Verify understanding of and support for combined Customer table with core users.	2/18/2008	Team	BB	-	Open	High
	31	Promotions	Discuss special offer / promotions with marketing to understand how these might change.	2/18/2008	Team	LB	-	Open	Med
	32	Order_Info	Verify concept of Channel with Marketing and Sales.	2/18/2008	Team	CK	-	Open	High
	33	Order_Info	Verify usability of combined Channel and Sales Reason fields in the same table.	2/18/2008	Team	CK	-	Open	High
	34	Order_Info	Determine list of possible combinations between Channel and Sales Reason.	2/18/2008	Team	CK	-	Open	High

Figure 7-8 Example data model issues list.

Review and validate model

- Source system developers and DBA
- Core business or power user
- Perform IT data model review
- Review with core users
- Present to the business users
- Finalize design documentation