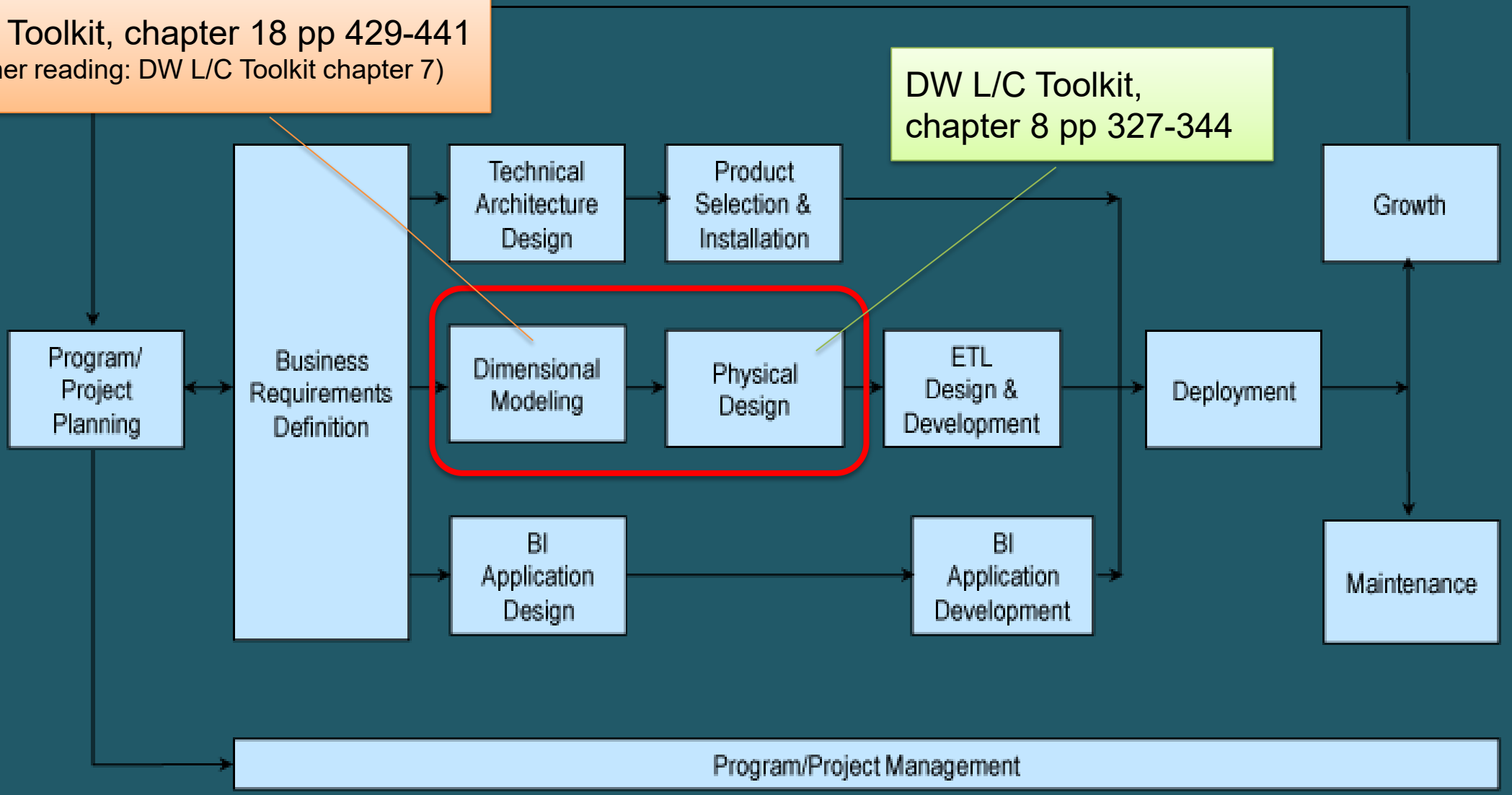
Designing the dimensional model and the physical database

Part 1

Where are we

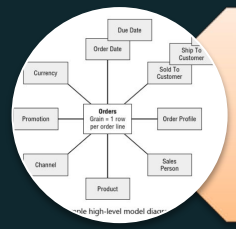
DW Toolkit, chapter 18 pp 429-441
(further reading: DW L/C Toolkit chapter 7)

DW L/C Toolkit,
chapter 8 pp 327-344

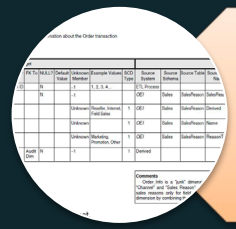


Study unit outcomes

Describe the dimensional modelling process



Develop a high-level model diagram



Develop a detailed dimensional model

Document the dimensional model design

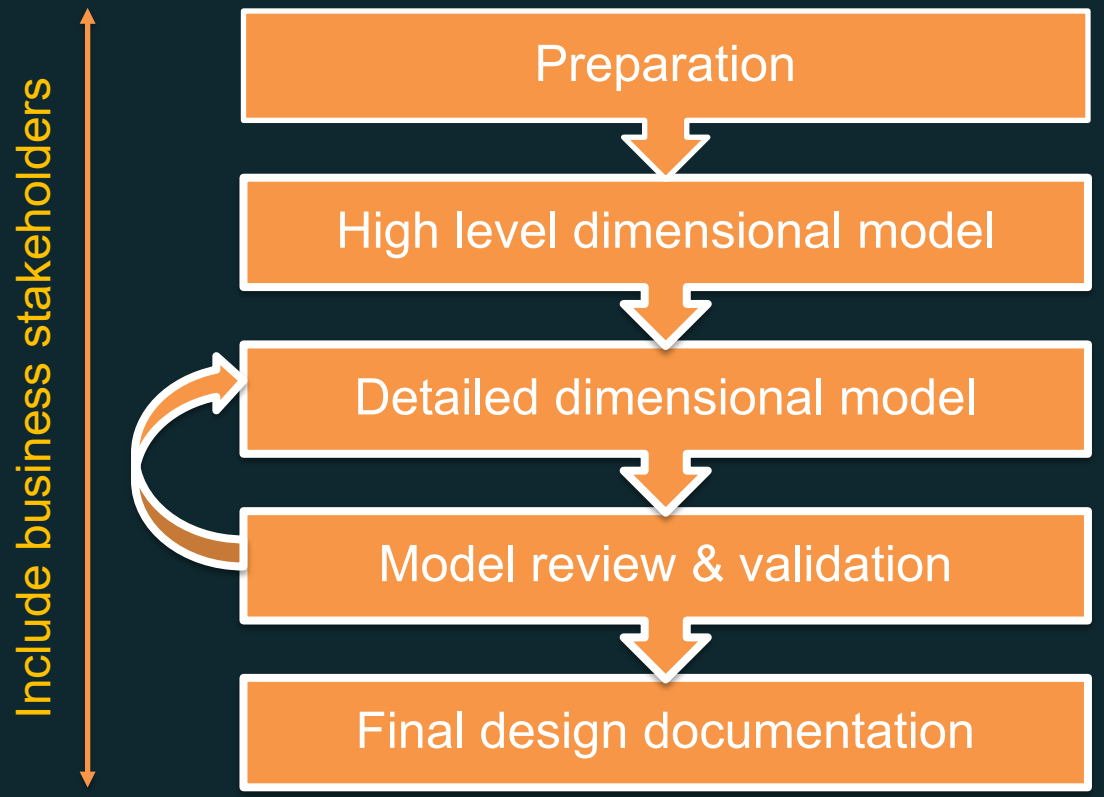
Study unit outcomes (cont.)

Describe the high-level physical design process

Discuss the development of standards for the DW/BI system

Develop a physical data model and database

Dimensional modelling process



Highly iterative & dynamic

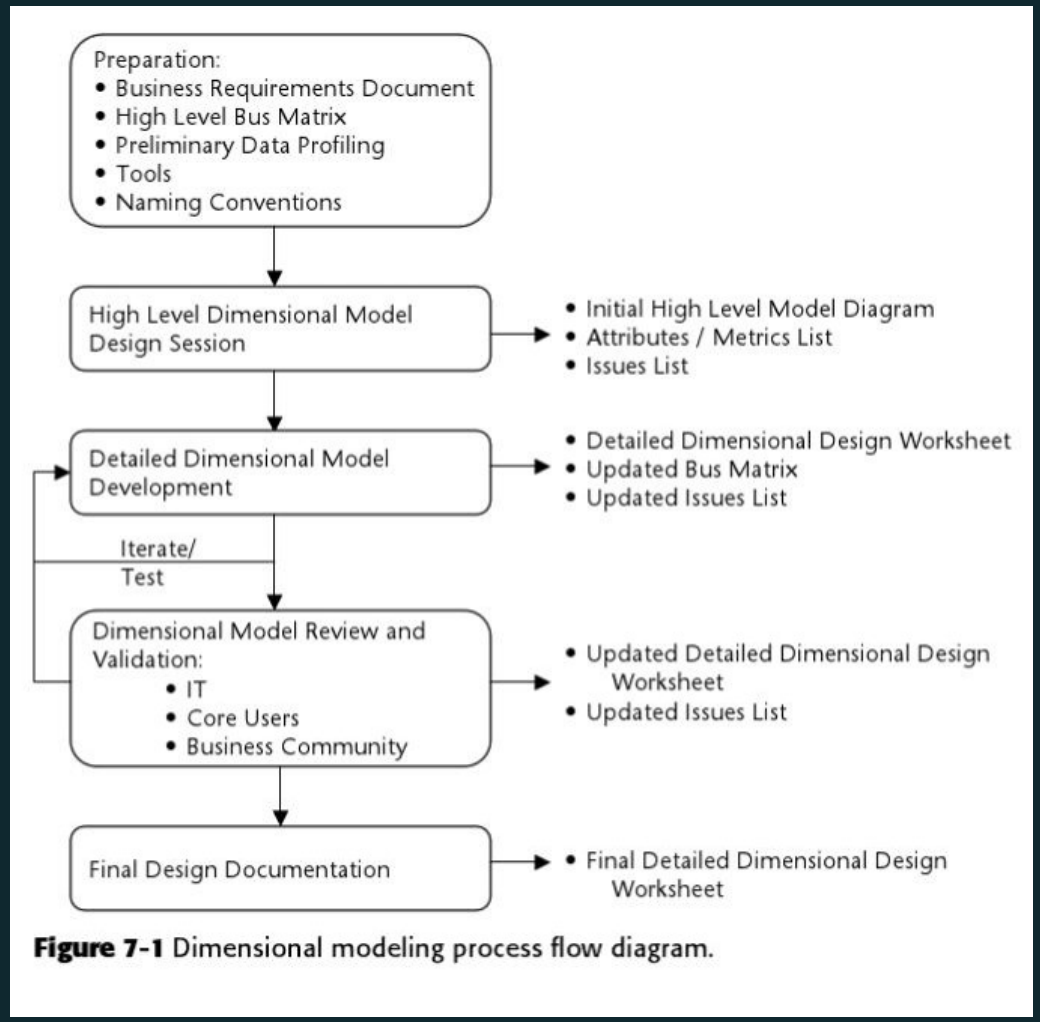
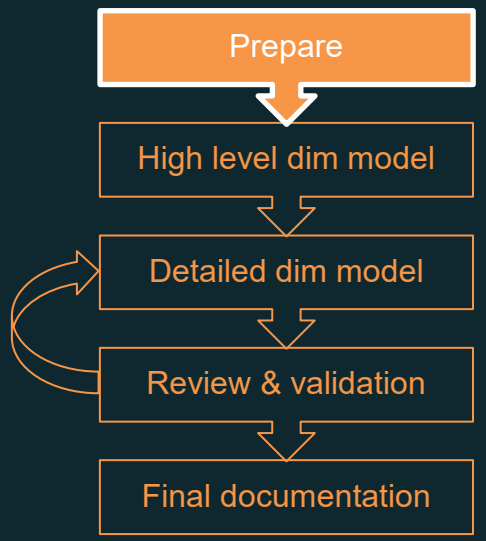


Figure 7-1 Dimensional modeling process flow diagram.

With deliverables



1. Identify Participants, especially business representatives

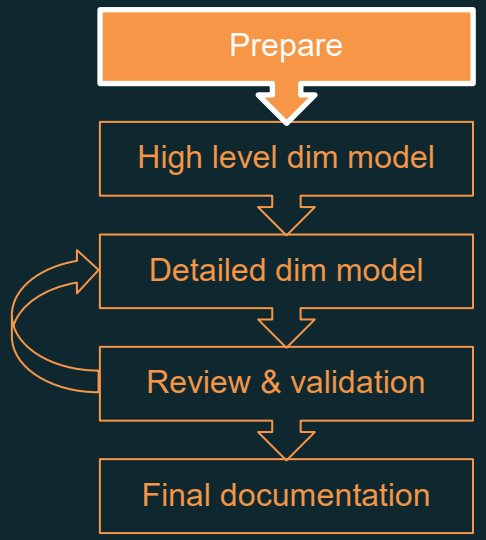
Core modeling team

- **Data modeler**
- Business analyst
- ETL representative
- Power user

Extended team

- **Business driver or governance steering committee**
- Data steward*
- Source system developers
- DBA
- ETL architect & developer
  - Dimension manager
  - Fact provider
- BI architect & developer

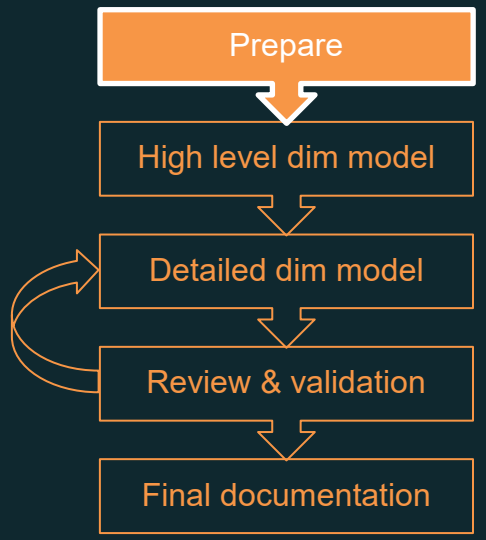
**Data stewardship: An organizational function to address data definitions, consistency, integration, quality, and knowledge in an enterprise.*



1. Identify Participants, especially business representatives

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



Deliverables:

- Business requirements
- High level bus matrix
- Preliminary data profiling
- Tools
- Naming conventions

2. Review the business requirements

- Proposed data elements, sample questions & reports
- Bring modelling team up to date & get their input

3. Select modeling tools

- Spreadsheets are handy to start with – can be easily modified
- list attribute, metrics, business and ETL rules, metadata

4. Preliminary data profiling

- To explore source data content and relationships

5. Establish naming conventions

- Agree on common definitions and labels
- Labels must be descriptive and consistent: think UX

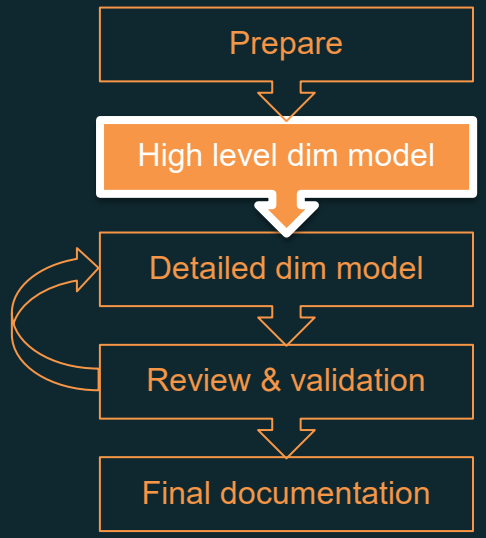
`PrimeWord_Qualifiers_ClassWord`

`Business_Address_PostalCode`

`Employee_Project_Timesheet_Hours`

6. Coordinate calendars and facilities

- Schedule design sessions and book conference rooms



- Follow the four-step process:
  1. Identify the business process
  2. Declare the grain of the business process
  3. Identify the dimensions
  4. Identify the facts

During requirements -> priorities

- Initial high level model
  - Diagram (bubble chart*)
  - From bus matrix
  - Include the grain!

Deliverables:

- Initial high level model diagram
- Initial attributes & metrics list
- Issues list

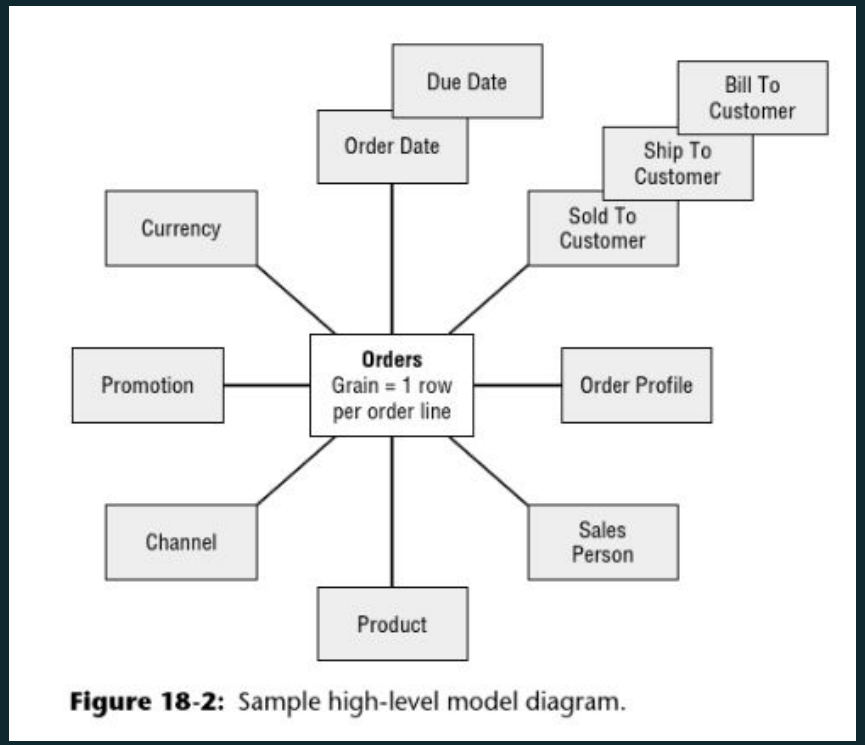
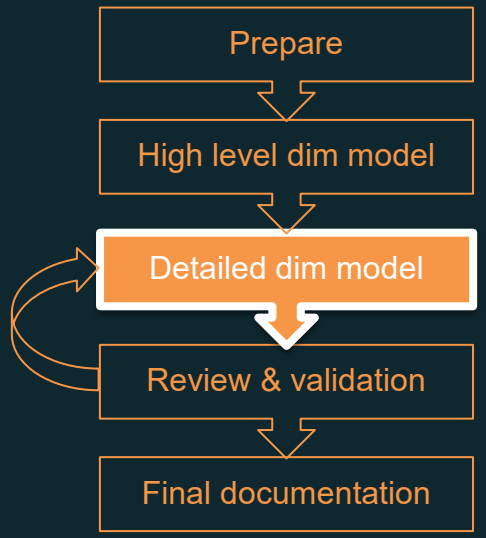


Figure 18-2: Sample high-level model diagram.

**Bubble chart: A high level graphical representation of a business process dimensional data model. Useful for communicating data models to a non-technical audience.*



Deliverables:

- Detailed dimensional design worksheets
- Updated bus matrix
- Updated issues list

- Complete attributes and metrics table by table
- Start with (easier) dimension tables
- Keep examining the source data
- Actions to include:
  a. Identify dimensions and their attributes — Conformed!
  b. Identify the facts – base and derived — True to the grain!
  c. Identify SCD techniques — Include source system experts
  d. Document the **details table designs** — Non-technical communication tool
  e. Track modelling issues
  f. Maintain and update the bus matrix

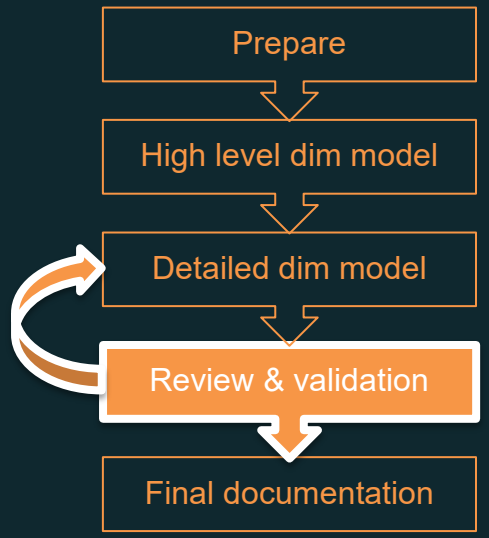
Source to target map

Table Name:	DimOrderInfo
Table Type	Dimension
View Name	OrderInfo
Description	OrderInfo is the "junk" dimension that includes miscellaneous information about the Order transaction
Used in schemas	Orders
Generate script?	Y

	Target										Source						
Column Name	Description	Datatype	Size	Key?	FK To	NULL?	Default Value	Unknown Member	Example Values	SCD Type	Source System	Source Schema	Source Table	Source Field Name	Source Datatype	ETL Rules	Comments
OrderInfoKey	Surrogate primary key	smallint		PK ID		N		–1	1, 2, 3, 4...		ETL Process					Standard surrogate key	
BKSalesReasonID	Sales reason ID from source system	smallint				N		–1			*OEI*	Sales	SalesReason	SalesReasonID	int	Convert to char; left-pad with zero. R for reseller row.	We need to insert a single row for Reseller
Channel	Sales channel	char	8					Unknown	Reseller, Internet, Field Sales	1	*OEI*	Sales	SalesReason	Derived		"Internet" for real sales reasons. "Reseller" for reseller row.	
SalesReason	Reason for the sale, as reported by the customer	varchar	30					Unknown		1	*OEI*	Sales	SalesReason	Name	nvarchar(50)	Convert to varchar; "Reseller" for reseller row.	
SalesReasonType	Type of sales reason	char	10					Unknown	Marketing, Promotion, Other	1	*OEI*	Sales	SalesReason	ReasonType	nvarchar(50)	Convert to varchar; "Reseller" for reseller row.	
AuditKey	What process loaded this row?	int		FK	Audit Dim	N		–1		1	Derived					Populated by ETL system using standard technique	

Comments
Order_Info is a "junk" dimension with only a handful of rows based on "Channel" and "Sales Reason". We currently have only three channels and sales reasons only for field sales and Internet sales. We can eliminate a dimension by combining these two.

Figure 7-6 Example detailed dimensional design worksheet.

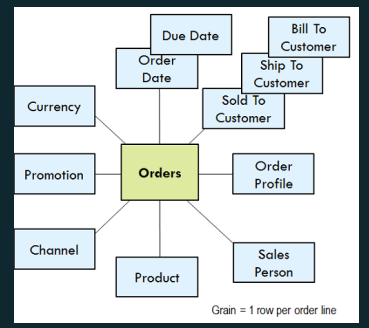


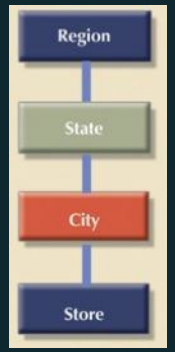
- Review, validate & redesign -> final design:
  - Ensure design meet requirements
  - Verify data are available
  - Provide sound source-to-target mapping
- Get feedback from interested parties:
  - Data model review with IT peers
    - Other DW/BI team members
    - Source system experts
    - DBAs
  - Review with core users
  - Present to broader user community
- Start with the bus matrix

Deliverables:

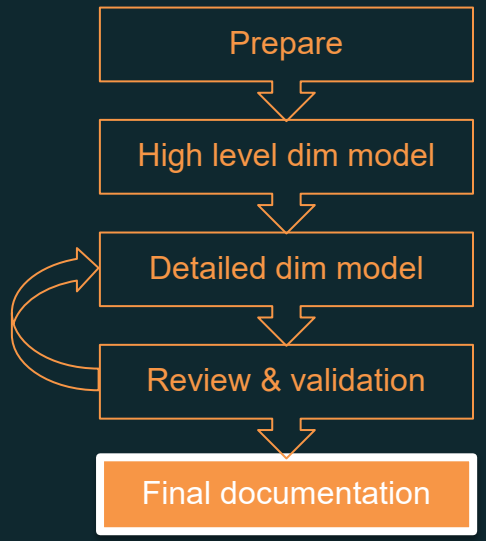
- Updated detailed dimensional design worksheets
- Updated issues list

	time	customers	products	suppliers	warehouses	employees	sales reps	stores
actuals sales	X	X	X	X			X	X
sales budget	X							X
actual orders	X	X	X		X		X	X
actual production	X		X	X	X	X		
production forecast	X		X	X				
inventory	X		X	X	X			X





List of common design flaws to scout for when performing a review

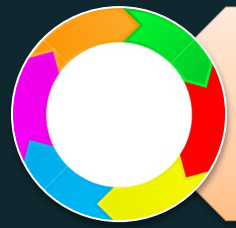


Deliverables:

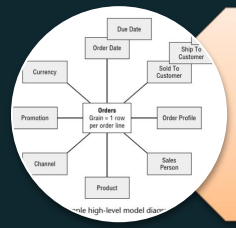
- Final detailed dimensional design worksheet

1. Brief description of the project business processes
2. High-level business requirements
3. High-level data model diagram
4. Detailed dimensional design worksheets — Core deliverable
5. Open issues lists
6. Known limitations of the design
7. Other items: design compromises, source data concerns, etc. …

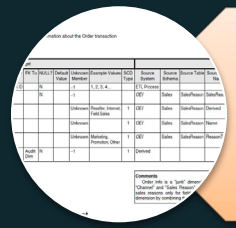
Recap - study unit outcomes



Describe the dimensional modelling process

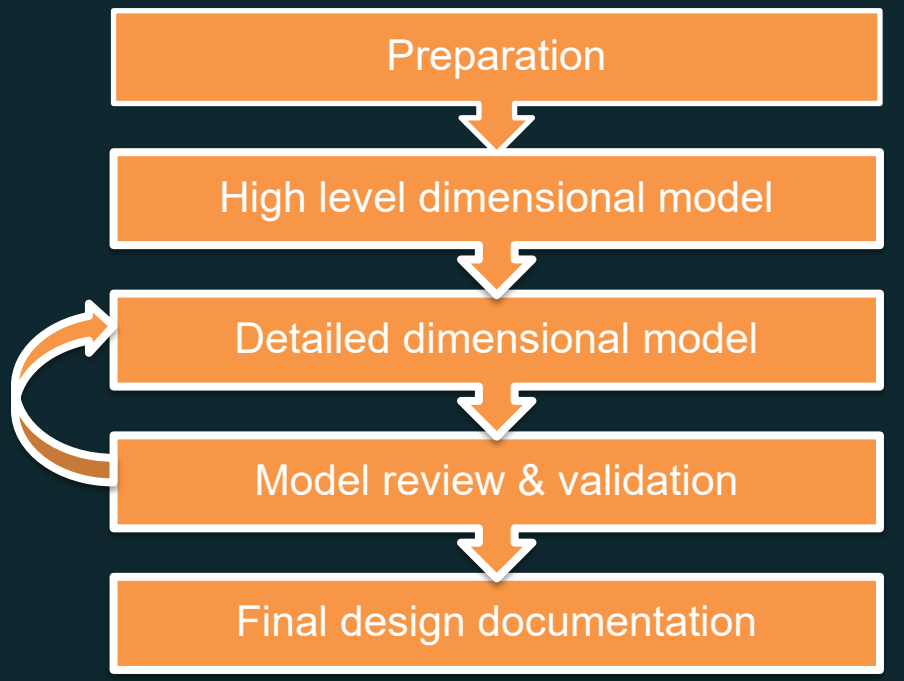


Develop a high-level model diagram



Develop a detailed dimensional model

Document the dimensional model design



Going Agile? Start with the Bus Matrix

End – Part 1