BINUS UNIVERSITY GRADUATE PROGRAM

People
Innovation
Excellence

LECTURE NOTES

7023T Advanced Database System

Session 05

Dimensional Modelling 1

People
Innovation
Excellence

LEARNING OUTCOMES

- Peserta diharapkan mampu memahami tahapan proses *dimensional modelling*
- Peserta diharapkan dapat menjelaskan hal-hal yang perlu dipersiapkan dalam tahap *preparation* pada proses *dimensional modelling.*
- Peserta diharapkan mampu memahami tahap *iterative model development* pada proses *dimensional modelling.*
- Peserta diharapkan dapat menjelaskan bagaimana proses *review* dan *validation* pada proses *dimensional modelling.*

OUTLINE MATERI (Sub-Topic):

1. *Process Overview*
2. *Preparation*
3. *Iterative Model Development*
4. *Review and Validation*

Process Overview

Proses desain *model dimensional* dilakukan dalam tiga tahap yaitu *preparation, iterative model development,* dan *review* and *validation.* Pada tahap *preparation* dibuat diagram *star schema* berdasarkan *enterprise bus matrix*. Diagram ini akan menjelaskan grain dari tabel fakta, atribut dan metrik yang akan digunakan sebagai *measure*. Diagram ini juga akan menjelaskan tabel dimensi beserta atributnya. Selanjutnya pada tahap *iterative model development* dilakukan pemetaan yang menunjukkan tabel-tabel pada sumber data yang akan digunakan untuk membentuk tabel fakta maupun tabel dimensi pada sistem DW/BI. Proses pengembangan model dilakukan secara iteratif tabel demi tabel, dengan analisis yang semakin detail pada setiap iterasinya. Model yang dihasilkan selanjutnya akan direview dan divalidasi oleh perwakilan dari pengguna bisnis maupun IT untuk meyakinkan desain yang dihasilkan memenuhi kebutuhan.

Preparation

Tahap *preparation* atau persiapan untuk *dimensional modelling* terdiri dari beberapa fase, yaitu:

- Mengidentifikasi tim yang akan dibentuk serta siapa partisipannya, seperti dijelaskan pada tabel 1. Mayoritas pekerjaan akan dilakukan oleh tim inti, yang terdiri dari *data modeller, business analyst* dan anggota tim ETL.
- Melakukan review terhadap dokumen kebutuhan bisnis, mulai dari *dimensional modelling*.
- Melakukan review terhadap data sumber, dapat dilakukan dengan cara mengeksekusi *query* sederhana atau memanfaatkan perangkat lunak *data profiling*. Fase ini bertujuan untuk meningkatkan pemahaman terhadap karakteristik data sumber.
- Mempersiapkan lingkungan yaang akan digunakan untuk memodelkan dimensi. Sketsa awal dari model dapat disajikan dalam bentuk *spreadsheet* seperti diilustrasikan pada tabel 2. Pendekatan ini akan menghasilkan model ini dapat dikembangkan secara iteratif secara mudah.
- Menentukan aturan penamaan tabel.

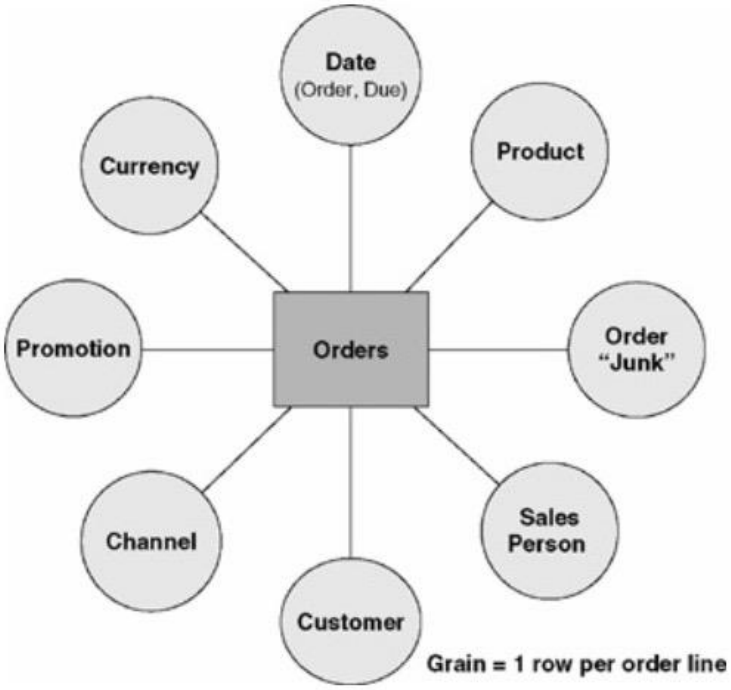
Tabel 1. Partisipan dan peranannya dalam *dimensional modelling*

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

Tabel 2. Contoh sketsa awal pemodelan dimensi

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	V olume 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

Setelah semua aktifitas diatas dilakukan, tahap persiapan dilanjutkan dengan pengembangan model awal yang dimulai dengan diagram model *dimensional*. Diagram ini merepresentasikan tabel fakta dan dimensi untuk proses bisnis yang dipilih. Diagram ini kadang-kadang disebut sebagai *bubble-chart.* Setiap tabel fakta dari proses bisnis tertentu perlu digambarkan dalam diagram terpisah. Peletakan tabel-tabel dimensi disarankan sesuai dengan tingkat kepentingannya, dimana dimensi yang penting diletakkan pada bagian atas seperti diilustrasikan pada gambar 1.



Gambar 1. Contoh diagram *high-level*

Iterative Model Development

Proses *iterative model development* terdiri dari empat langkah, yakni:

1. **Memilih proses bisnis** – menggunakan *enterprise data warehouse bus matrix*, dimana setiap baris dari matriks merepresentasikan proses bisnisnya. Pemilihan proses bisnis dilakuan secara iteratif, satu proses dalam satu iterasi.
2. **Mententukan grain dari tabel fakta** – grain dapat digunakan untuk menjawab pertanyaan "apa yang direpresentasikan oleh tiap baris dari tabel fakta?". Grain menyatakan tingkat detail dari data yang disimpan pada tabel fakta. Semakin tinggi

tingkat detail akan semakin baik. Melewatkan langkah ini merupakan kesalahan desain yang paling sering terjadi.

3. **Mengidentifikasi tabel-tabel dimensi** – jumlah minimum dari tabel dimensi mengikuti grain yang dipilih pada tabel fakta. Dimensi lain dapat ditambahkan kemudian, namun perlu memiliki tingkat grain yang sama dengan tabel fakta.
4. **Mengidentifikasi tabel fakta** – biasanya diturunkan dari hasil pengukuran. Semua fakta harus memiliki tingkat grain yang sama.

Memulai dengan tabel dimensi adalah pendekatan yang paling mudah, contohnya dimensi *date*. Lakukan desain satu atau dua tabel untuk setiap sesinya, hal ini dilakukan untuk menghindarkan kelihangan detail pada model *high-level*.

Identifikasi sumber data terbaik yang akan digunakan untuk mengisi tabel target, baik tabel fakta maupun tabel-tabel dimensi. Mulai dengan menyusun list dari semua kandidat sumber data; sebagian kandidat mungkin saja berasal dari dokumen *business requirement*, dan sisanya mungkin berasal dari personil IT. Selanjutnya lakukan analisis dan *profiling* terhadap setiap kandidat. Kemudian pilih sumber data terbaik yang akan digunakan dalam model dan dokumentasikan alasan pemilihan tersebut.

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.

Gambar 2. Contoh desain detail dimensional model

Identifikasi dimensi-dimensi yang dapat digunakan bersama (*conformed dimensions*) serta tabel fakta dasar dan turunannya. Tabel fakta aditif dapat dihasilkan dari perhitungan pada tabel fakta lain untuk baris yang sama. Apabila kita memiliki banyak tabel fakta turunan, disarankan untuk mendokumentasikannya pada sebuah worksheet. Rancangan detil

dari setiap tabel fakta dan dimensi perlu didokumentasikan dalam *detailed design worksheet* seperti yang diperlihatkan pada Gambar 2. Proses desain secara detail meningkatkan pemahaman kita terhadap bisnis proses dan dapat menjadi panduan saat diperlukan perubahan terkait proses bisnis, tabel fakta, maupun tabel dimensi.

Review and Validation

Model dimensional perlu direview oleh beberapa pihak dengan sudut pandang dan keahliannya masing-masing. Pihak yang pertama kali disarankan untuk melakukan review adalah yang berasal dari DBA (database administrator) atau *system developer* dari departemen IT. Mereka memiliki pemahaman yang baik terhadap sistem transaksional namun biasanya tidak memiliki pengetahuan yang mencukupi mengenai *dimensional model*. Oleh karena itu perlu diberikan tutorial singkat kepada mereka, agar proses review dapat berjalan sesuai dengan harapan. Review berikutnya disarankan dilakukan oleh pengguna dari kalangan bisnis yang tidak terlibat langsung dengan proses desain. Review ini juga merupakan sesi edukatif bagi mereka, dengan cara menunjukkan beberapa contoh sederhana yang dapat mengilustrasikan kemampuan analitik dari model *dimensional*. Tujuannya adalah untuk melakukan verifikasi bahwa model dapat menjawab pertanyaan-pertanyaan terkait proses bisnis.

Setelah desain direview dan divalidasi, langkah berikutnya adalah mendokumentasikan desain tersebut. Dokumen detail desain final dapat disajikan dalam bentuk *worksheet* (seperti yang sudah diilustrasikan pada gambar 2) yang dapat menjadi dasar untuk dokumen *source-to-target mapping,* yang menjelaskan dari mana setiap kolom pada tabel target memperoleh sumber datanya. *Source-to-target mapping* merupakan salah dokumen yang sangat penting bagi tim ETL. Beberapa aspek penting dari proses *dimensional modelling* dirangkum pada gambar 3.

Terdapat beberapa teknik selain e*ntity relationship modelling* dan *dimensional modelling* untuk memodelkan data. Alternatifnya seperti *subject modeling, domain modeling, fact qualifier matrix/modeling, state transition modeling,* dan *information flow modeling. Fact qualifier modelling* merupakan teknik yang berguna untu mendapatkan kebutuhan *granularity/reporting.* Masing-masing model memfokuskan pada jenis informasi yang berbeda. Terdapat lebih dari satu pendekatan untuk mendapatkan kebutuhan bisnis, dan setiap

proyek memiliki berbagai jenis kebutuhan. Merupakan hal yang baik untuk melihat peluang untuk menerapkan model lain yang cocok untuk proyek yang akan dijalankan.

Managing the Effort and Reducing Risk

Follow these best practices to keep the dimensional modeling effort and risk in check:

- Assure the data modeler leading the design process is an expert dimensional modeler. If not, consider supplementing with an outside resource.
- Every member of the design team must thoroughly and completely understand the business requirements.
- Be sure to include power users as part of the design team.
- Continuously probe the proposed source data with your data profiling tools to assure the data required to support the proposed data model is available.

Assuring Quality

Key steps to assure quality include:

- Adhere to dimensional modeling best practices.
- Insist on active participation of power users in the design process.
- Conduct extensive data profiling.
- Do not skip the IT and user review sessions.

Key Roles

Key roles for designing the dimensional model include:

- The data modeler leads the dimensional modeling efforts.
- The business analyst, power users, and BI application developers represent the business users' analytic needs.
- Data stewards help drive to organizational agreement on the dimensional model's names, definitions, and business rules.
- Source system experts bring knowledge of the operational systems.
- The ETL team learns about the sources, targets, and gets a sense of the heavy lifting they'll need to do to convert from one to the other.
- Interested parties in the IT and broader business communities will review and provide feedback on the design.
- The database and data access tool vendors provide principles for database design to optimize their products.

Key Deliverables

Key deliverables for designing the dimensional model include:

- High level model diagram
- Attribute and metrics list
- Detailed dimensional design worksheet
- Issues list

Gambar 3. Blueprint for action untuk proses *dimensional processing*

BINUS UNIVERSITY GRADUATE PROGRAM

People
Innovation
Excellence

BINUS UNIVERSITY GRADUATE PROGRAM

People
Innovation
Excellence

SIMPULAN

- Penentuan *grain* dari tabel fakta merupakan salah satu kesalahan yang paling sering dilakukan saat proses *dimensional modelling.*
- Salah satu dokumen penting yang dihasilkan dari proses *dimensional modelling* adalah *source-to-target mapping* yang akan menjadi panduan bagi tim ETL dalam melaksanakan tugasnya.
- Perlu diadakan tutorial singkat mengenai *dimensional model* kepada para pengguna dari kalangan bisnis maupun personil IT agar proses *review* dan *validation* dapat berjalan sesuai dengan tujuan.

BINUS UNIVERSITY GRADUATE PROGRAM

People
Innovation
Excellence

DAFTAR PUSTAKA

1. Kimball, R. (2008). *The Data Warehouse Lifecycle Toolkit*. John Wiley & Sons.
2. Kimball, R., & Ross, M. (2011). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. John Wiley & Sons.
3. Inmon, W. H. (2005). *Building the Data Warehouse*. John wiley & sons.