

DIMENSIONAL MODEL DATA WAREHOUSE: AN INTRODUCTION (WHAT)

Dan Kirpes, Fireman's Fund Insurance Company, Novato, CA

ABSTRACT

This paper will focus on the Dimensional Model Data Warehouse and is the follow up presentation to Dimensional Model Data Warehouse: An Overview. This paper will first address the model vocabulary. The components of the model: grain, fact table, and dimension table will be covered in more detail. Metadata will be briefly discussed. Graphical examples of the model will be displayed. Additional information sources and resources will be shared.

DIMENSIONAL MODELING VOCABULARY

Data Warehouse

The queryable source of data in the enterprise.

Data Mart

A logical and physical subset of the data warehouse. Data marts represent data from a single business process. The choice of the data mart, and hence the source of data, is the first step in the design of the dimensional model.

Grain

Grain or granularity is the meaning of a single row in a fact table. The declaration of the grain of a fact table is the second of four key steps in the design of a dimensional model.

Fact Table

In a Dimensional model, the central table with numeric performance measurements characterized by a composite key, each of whose elements is a foreign key drawn from a dimension table.

Dimension Table

A dimensional table is a table in a dimensional model with a single-part primary key and descriptive attribute columns.

DESIGNING A DIMENSIONAL MODEL:

Step 1: Select the Business Process

Step 2: Declare the Grain.

Step 3: Identify the Facts

Step 4: Choose the Dimensions

BUSINESS PROCESS

The first step in the design is to decide what business process to model. This requires an understanding of the business requirements and the availability of data.

GRAIN

It is important to declare the grain of the fact table before creating the fact table. There are 3 grain categories: transactions, periodic snapshot, and accumulating snapshot.

Transactions

This is the granularity in one row for the lowest level of detail captured by a transaction.

Periodic Snapshot

Periodic snapshot fact table grain represents business performance at the end of each regular, predictable time period. For example, monthly.

Accumulating Snapshot

This type of fact table contains multiple dates, which represent the major milestones of a relatively short-live process or pipeline. The fact table is revisited and updated as activity occurs. A record is placed in an accumulating snapshot fact table just once, when the item is first created.

FACT TABLE

A fact table is the primary table in a dimensional model where the numerical performance measurements of the business are stored.

Additive

The most useful facts are numeric and additive, such as dollar amounts.

Non-additive

Some numeric measures cannot be added across any dimension. For example, ratios or averages are non-additive numerical measures.

Keys

All fact tables have two or more foreign keys, as designated by the FK notation that connect to the dimension tables' primary keys.

The derived fact worksheet, shown in Figure 1, is helpful when developing the facts. The worksheet contains the following columns:

- Change flag: Indicates a row that has changed since the last release of the worksheet.
- Fact Group (Data Mart): The name of the data mart
- Measurement (Fact) Name: The name of the derived fact

- Measurement (Fact) Description: A brief description of the fact.
- Type: Indicates the kind of fact
 - Column
 - Constraints
 - Transformation
 - Calculation
 - Column with limits
- Aggregation rule: Facts summed or recalculated
- Formula: Specific mathematical formula
- Constraints: Indicates the constraints that must be applied to the data in order to create this fact.
- Transformations: Indicates specific translations that must be calculated when creating the fact.

DIMENSION TABLE

Description

Dimension tables are integral companions to a fact table. The dimension tables contain the textual descriptors of the business. In a well-designed dimensional model, dimension tables have many columns or attributes. These attributes describe the rows in the dimension table. It is not uncommon for a dimension table to have 50 to 100 attributes.

Detail

Full descriptive information must be provided to support the model. Each dimension attribute should have the following:

- Attribute name: The business attribute name.
- Attribute definition: A brief description of the attribute.
- Cardinality: The number of distinct values for the attribute
- Sample data: Sample values that the attribute may contain.
- Slowly changing policy.
 - Type 0: the record is never updated
 - Type 1: the record is overwritten
 - Type 2: new record created when a change is detected
 - Type 3: an old and a new version of this attribute is continuously maintained.

See Figure 2 for an example of the Dimension attribute detail description worksheet.

THE DIMENSIONAL MODEL: PUTTING IT ALL TOGETHER

Now that we have reviewed the individual parts of the dimensional model, it is time to develop the logic and physical model. Figure 3 depicts a typical dimensional model.

METADATA DATABASE

Description

Meta data is any data that is maintained to support the operations or use of a data warehouse, similar to an encyclopedia for the data warehouse.

There are several types of metadata:

- Source System Metadata
- Data Staging Metadata
- DBMS Metadata
- End-User Metadata

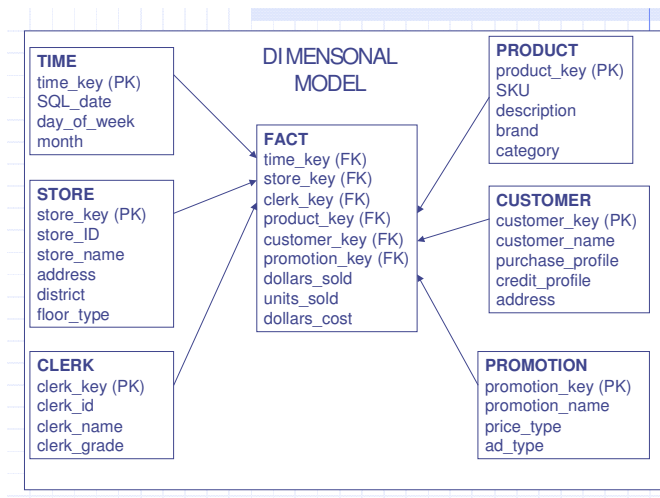
FIGURE 1: DERIVED FACT WORKSHEET

	Example 1	Example 2
Change Flag		
Fact Group	Point of Sale	Point of Sale
Measure Name	\$ Sales	Total US \$ Sales
Measure Description	Dollar amount sold	Dollar amt sold US
Type	col	cnstr
Aggregation Rule	sum	sum
Formula	sum(dollar sales)	sum(dollar sales)
Constraints	none	Geography=US
Transformations	none	none

FIGURE 2: DIMENSION ATTRIBUTE WORKSHEET

Attribute Name	Attribute Description	Card- inality	Sample Values
Day	specific date	365	1/14/98
Holiday	calendar holiday	14	Thanksgiving
Day of Week	name of the day	7	Thursday
Calendar Week	week ending in Saturday	53	WE 01/17/1998
Calendar Month	calendar month	12	1998/01
Calendar Quarter	calendar quarter	4	1998 Q1
Calendar Year	calendar year	1	1998

FIGURE 3: DIMENSIONAL MODEL



CONCLUSION

In this paper, we explored the major components of the dimension data model. We discussed the fact table, dimension tables, and the grain.

REFERENCES

Dimensional Modeling, www.ou.edu/class/aschwarz/DataWarehouse

Kimball, Ralph, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, New York, NY: John Wiley and Sons, Inc., 2002. 436pp.

Kimball, Ralph, *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses*, New York, NY: John Wiley and Sons, Inc., 1998. 771pp.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Dan Kirpes
Fireman's Fund Insurance Company
777 San Marin Drive
Novato, CA 94998
(415) 899-3561:
Email: dkirpes@ffic.com