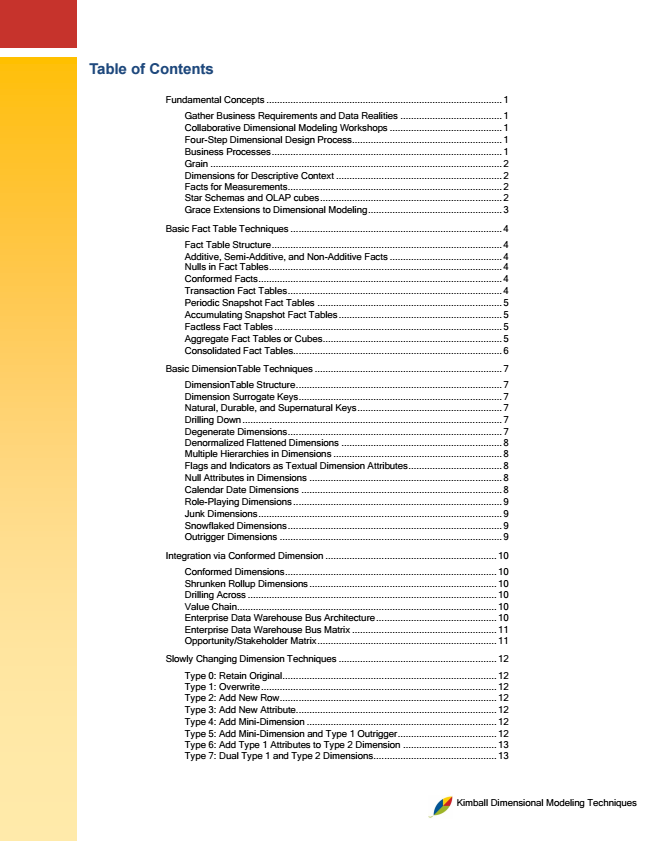


Kimball Dimensional Modeling Techniques



**Table of Contents**

Fundamental Concepts ........................................................................................ 1

Gather Business Requirements and Data Realities ...................................... 1 Collaborative Dimensional Modeling Workshops .......................................... 1 Four-Step Dimensional Design Process ........................................................ 1 Business Processes ...................................................................................... 1 Grain ............................................................................................................. 2 Dimensions for Descriptive Context .............................................................. 2 Facts for Measurements ................................................................................ 2 Star Schemas and OLAP cubes .................................................................... 2 Grace Extensions to Dimensional Modeling .................................................. 3

Basic Fact Table Techniques ............................................................................... 4

Fact Table Structure ...................................................................................... 4 Additive, Semi-Additive, and Non-Additive Facts .......................................... 4 Nulls in Fact Tables ....................................................................................... 4 Conformed Facts ........................................................................................... 4 Transaction Fact Tables ................................................................................ 4 Periodic Snapshot Fact Tables ..................................................................... 5 Accumulating Snapshot Fact Tables ............................................................. 5 Factless Fact Tables ..................................................................................... 5 Aggregate Fact Tables or Cubes ................................................................... 5 Consolidated Fact Tables .............................................................................. 6

Basic DimensionTable Techniques ...................................................................... 7

DimensionTable Structure ............................................................................. 7 Dimension Surrogate Keys ............................................................................ 7 Natural, Durable, and Supernatural Keys ...................................................... 7 Drilling Down ................................................................................................. 7 Degenerate Dimensions ................................................................................ 7 Denormalized Flattened Dimensions ............................................................ 8 Multiple Hierarchies in Dimensions ............................................................... 8 Flags and Indicators as Textual Dimension Attributes ................................... 8 Null Attributes in Dimensions ........................................................................ 8 Calendar Date Dimensions ........................................................................... 8 Role-Playing Dimensions .............................................................................. 9 Junk Dimensions ........................................................................................... 9 Snowflaked Dimensions ................................................................................ 9 Outrigger Dimensions ................................................................................... 9

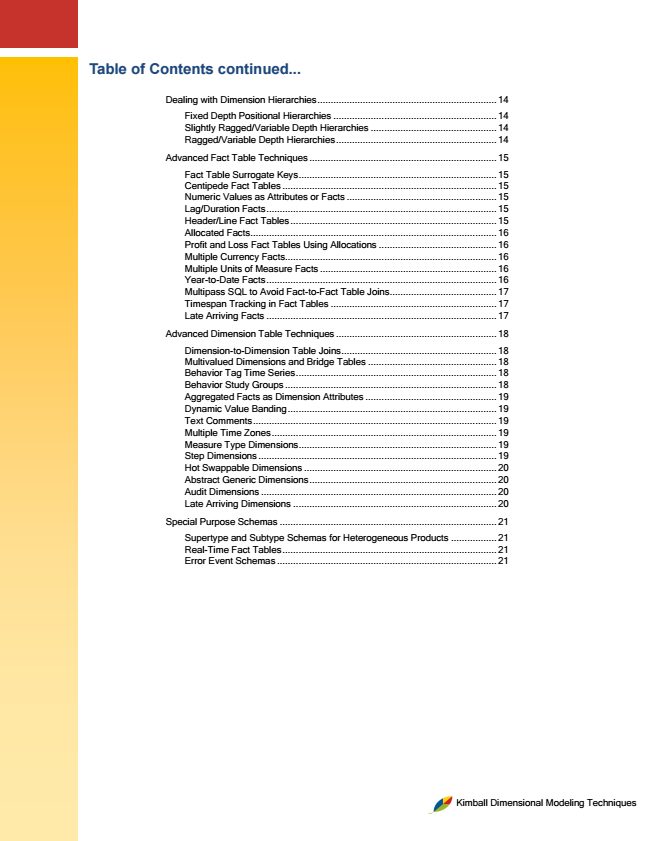
Integration via Conformed Dimension ................................................................ 10

Conformed Dimensions ............................................................................... 10 Shrunken Rollup Dimensions ...................................................................... 10 Drilling Across ............................................................................................. 10 Value Chain ................................................................................................. 10 Enterprise Data Warehouse Bus Architecture ............................................. 10 Enterprise Data Warehouse Bus Matrix ...................................................... 11 Opportunity/Stakeholder Matrix ................................................................... 11

Slowly Changing Dimension Techniques ........................................................... 12

Type 0: Retain Original ................................................................................ 12 Type 1: Overwrite ........................................................................................ 12 Type 2: Add New Row ................................................................................. 12 Type 3: Add New Attribute........................................................................... 12 Type 4: Add Mini-Dimension ....................................................................... 12 Type 5: Add Mini-Dimension and Type 1 Outrigger ..................................... 12 Type 6: Add Type 1 Attributes to Type 2 Dimension ................................... 13 Type 7: Dual Type 1 and Type 2 Dimensions .............................................. 13

Kimball Dimensional Modeling Techniques



**Table of Contents continued...**

Dealing with Dimension Hierarchies ................................................................... 14

Fixed Depth Positional Hierarchies ............................................................. 14 Slightly Ragged/Variable Depth Hierarchies ............................................... 14 Ragged/Variable Depth Hierarchies ............................................................ 14

Advanced Fact Table Techniques ...................................................................... 15

Fact Table Surrogate Keys .......................................................................... 15 Centipede Fact Tables ................................................................................ 15 Numeric Values as Attributes or Facts ........................................................ 15 Lag/Duration Facts ...................................................................................... 15 Header/Line Fact Tables ............................................................................. 15 Allocated Facts ............................................................................................ 16 Profit and Loss Fact Tables Using Allocations ............................................ 16 Multiple Currency Facts ............................................................................... 16 Multiple Units of Measure Facts .................................................................. 16 Year-to-Date Facts ...................................................................................... 16 Multipass SQL to Avoid Fact-to-Fact Table Joins ........................................ 17 Timespan Tracking in Fact Tables .............................................................. 17 Late Arriving Facts ...................................................................................... 17

Advanced Dimension Table Techniques ............................................................ 18

Dimension-to-Dimension Table Joins .......................................................... 18 Multivalued Dimensions and Bridge Tables ................................................ 18 Behavior Tag Time Series ........................................................................... 18 Behavior Study Groups ............................................................................... 18 Aggregated Facts as Dimension Attributes ................................................. 19 Dynamic Value Banding .............................................................................. 19 Text Comments ........................................................................................... 19 Multiple Time Zones .................................................................................... 19 Measure Type Dimensions .......................................................................... 19 Step Dimensions ......................................................................................... 19 Hot Swappable Dimensions ........................................................................ 20 Abstract Generic Dimensions ...................................................................... 20 Audit Dimensions ........................................................................................ 20 Late Arriving Dimensions ............................................................................ 20

Special Purpose Schemas ................................................................................. 21

Supertype and Subtype Schemas for Heterogeneous Products ................. 21 Real-Time Fact Tables ................................................................................ 21 Error Event Schemas .................................................................................. 21

Kimball Dimensional Modeling Techniques



© Kimball Group. All rights reserved. Kimball Dimensional Modeling Techniques 1

Ralph Kimball introduced the data warehouse/business intelligence industry to dimensional modeling in 1996 with his seminal book, The Data Warehouse Toolkit. Since then, the Kimball Group has extended the portfolio of best practices. Drawn from The Data Warehouse Toolkit, Third Edition (coauthored by Ralph Kimball and Margy Ross, 2013), here are the “official” Kimball dimensional modeling techniques.

**Fundamental Concepts**

Gather Business Requirements and Data Realities Before launching a dimensional modeling effort, the team needs to understand the needs of the business, as well as the realities of the underlying source data. You uncover the requirements via sessions with business representatives to understand their objectives based on key performance indicators, compelling business issues, decision-making processes, and supporting analytic needs. At the same time, data realities are uncovered by meeting with source system experts and doing high-level data profiling to assess data feasibilities.

Collaborative Dimensional Modeling Workshops Dimensional models should be designed in collaboration with subject matter experts and data governance representatives from the business. The data modeler is in charge, but the model should unfold via a series of highly interactive workshops with business representatives. These workshops provide another opportunity to flesh out the requirements with the business. Dimensional models should not be designed in isolation by folks who don’t fully understand the business and their needs; collaboration is critical!

Four-Step Dimensional Design Process The four key decisions made during the design of a dimensional model include:

1. Select the business process. 2. Declare the grain. 3. Identify the dimensions. 4. Identify the facts The answers to these questions are determined by considering the needs of the business along with the realities of the underlying source data during the collaborative modeling sessions. Following the business process, grain, dimension, and fact declarations, the design team determines the table and column names, sample domain values, and business rules. Business data governance representatives must participate in this detailed design activity to ensure business buy-in.

Business Processes Business processes are the operational activities performed by your organization, such as taking an order, processing an insurance claim, registering students for a class, or snapshotting every account each month. Business process events generate or capture performance metrics that translate into facts in a fact table. Most fact tables focus on the results of a single business process. Choosing the process is important because it defines a specific design target and allows the grain, dimensions, and facts to be declared. Each business process corresponds to a row in the enterprise data warehouse bus matrix.



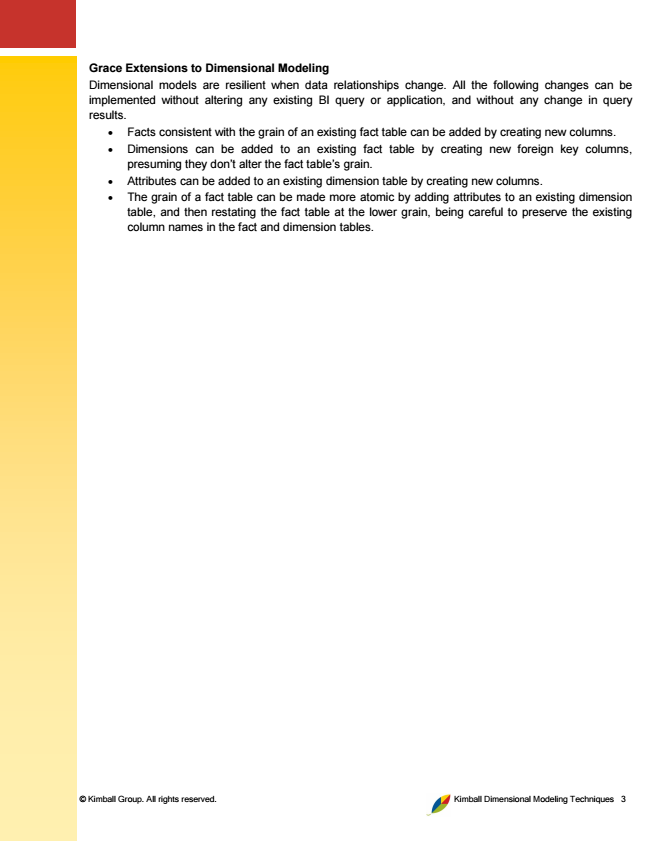
© Kimball Group. All rights reserved. Kimball Dimensional Modeling Techniques 2

Grain Declaring the grain is the pivotal step in a dimensional design. The grain establishes exactly what a single fact table row represents. The grain declaration becomes a binding contract on the design. The grain must be declared before choosing dimensions or facts because every candidate dimension or fact must be consistent with the grain. This consistency enforces a uniformity on all dimensional designs that is critical to BI application performance and ease of use. Atomic grain refers to the lowest level at which data is captured by a given business process. We strongly encourage you to start by focusing on atomic- grained data because it withstands the assault of unpredictable user queries; rolled-up summary grains are important for performance tuning, but they pre-suppose the business’s common questions. Each proposed fact table grain results in a separate physical table; different grains must not be mixed in the same fact table.

Dimensions for Descriptive Context Dimensions provide the “who, what, where, when, why, and how” context surrounding a business process event. Dimension tables contain the descriptive attributes used by BI applications for filtering and grouping the facts. With the grain of a fact table firmly in mind, all the possible dimensions can be identified. Whenever possible, a dimension should be single valued when associated with a given fact row. Dimension tables are sometimes called the “soul” of the data warehouse because they contain the entry points and descriptive labels that enable the DW/BI system to be leveraged for business analysis. A disproportionate amount of effort is put into the data governance and development of dimension tables because they are the drivers of the user’s BI experience.

Facts for Measurements Facts are the measurements that result from a business process event and are almost always numeric. A single fact table row has a one-to-one relationship to a measurement event as described by the fact table’s grain. Thus a fact table corresponds to a physical observable event, and not to the demands of a particular report. Within a fact table, only facts consistent with the declared grain are allowed. For example, in a retail sales transaction, the quantity of a product sold and its extended price are good facts, whereas the store manager’s salary is disallowed.

Star Schemas and OLAP cubes Star schemas are dimensional structures deployed in a relational database management system (RDBMS). They characteristically consist of fact tables linked to associated dimension tables via primary/ foreign key relationships. An online analytical processing (OLAP) cube is a dimensional structure implemented in a multidimensional database; it can be equivalent in content to, or more often derived from, a relational star schema. An OLAP cube contains dimensional attributes and facts, but it is accessed through languages with more analytic capabilities than SQL, such as XMLA. OLAP cubes are included in this list of basic techniques because an OLAP cube is often the final step in the deployment of a dimensional DW/BI system, or may exist as an aggregate structure based on a more atomic relational star schema.



© Kimball Group. All rights reserved. Kimball Dimensional Modeling Techniques 3

Grace Extensions to Dimensional Modeling Dimensional models are resilient when data relationships change. All the following changes can be implemented without altering any existing BI query or application, and without any change in query results.

• Facts consistent with the grain of an existing fact table can be added by creating new columns.

• Dimensions can be added to an existing fact table by creating new foreign key columns, presuming they don’t alter the fact table’s grain.

• Attributes can be added to an existing dimension table by creating new columns.

• The grain of a fact table can be made more atomic by adding attributes to an existing dimension table, and then restating the fact table at the lower grain, being careful to preserve the existing column names in the fact and dimension tables.



© Kimball Group. All rights reserved. Kimball Dimensional Modeling Techniques 4

**Basic Fact Table Techniques**

Fact Table Structure A fact table contains the numeric measures produced by an operational measurement event in the real world. At the lowest grain, a fact table row corresponds to a measurement event and vice versa. Thus the fundamental design of a fact table is entirely based on a physical activity and is not influenced by the eventual reports that may be produced. In addition to numeric measures, a fact table always contains foreign keys for each of its associated dimensions, as well as optional degenerate dimension keys and date/time stamps. Fact tables are the primary target of computations and dynamic aggregations arising from queries.

Additive, Semi-Additive, and Non-Additive Facts The numeric measures in a fact table fall into three categories. The most flexible and useful facts are fully additive; additive measures can be summed across any of the dimensions associated with the fact table. Semi-additive measures can be summed across some dimensions, but not all; balance amounts are common semi-additive facts because they are additive across all dimensions except time. Finally, some measures are completely non-additive, such as ratios. A good approach for non- additive facts is, where possible, to store the fully additive components of the non-additive measure and sum these components into the final answer set before calculating the final non-additive fact. This final calculation is often done in the BI layer or OLAP cube.

Nulls in Fact Tables Null-valued measurements behave gracefully in fact tables. The aggregate functions (SUM, COUNT, MIN, MAX, and AVG) all do the “right thing” with null facts. However, nulls must be avoided in the fact table’s foreign keys because these nulls would automatically cause a referential integrity violation. Rather than a null foreign key, the associated dimension table must have a default row (and surrogate key) representing the unknown or not applicable condition.

Conformed Facts If the same measurement appears in separate fact tables, care must be taken to make sure the technical definitions of the facts are identical if they are to be compared or computed together. If the separate fact definitions are consistent, the conformed facts should be identically named; but if they are incompatible, they should be differently named to alert the business users and BI applications.

Transaction Fact Tables A row in a transaction fact table corresponds to a measurement event at a point in space and time. Atomic transaction grain fact tables are the most dimensional and expressive fact tables; this robust dimensionality enables the maximum slicing and dicing of transaction data. Transaction fact tables may be dense or sparse because rows exist only if measurements take place. These fact tables always contain a foreign key for each associated dimension, and optionally contain precise time stamps and degenerate dimension keys. The measured numeric facts must be consistent with the transaction grain.



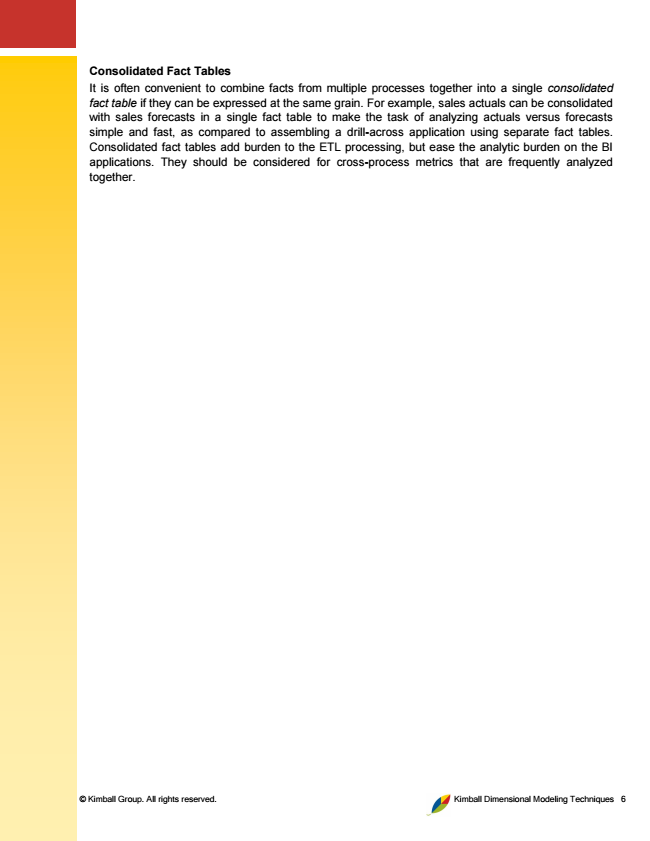
© Kimball Group. All rights reserved. Kimball Dimensional Modeling Techniques 5

Periodic Snapshot Fact Tables A row in a periodic snapshot fact table summarizes many measurement events occurring over a standard period, such as a day, a week, or a month. The grain is the period, not the individual transaction. Periodic snapshot fact tables often contain many facts because any measurement event consistent with the fact table grain is permissible. These fact tables are uniformly dense in their foreign keys because even if no activity takes place during the period, a row is typically inserted in the fact table containing a zero or null for each fact.

Accumulating Snapshot Fact Tables A row in an accumulating snapshot fact table summarizes the measurement events occurring at predictable steps between the beginning and the end of a process. Pipeline or workflow processes, such as order fulfillment or claim processing, that have a defined start point, standard intermediate steps, and defined end point can be modeled with this type of fact table. There is a date foreign key in the fact table for each critical milestone in the process. An individual row in an accumulating snapshot fact table, corresponding for instance to a line on an order, is initially inserted when the order line is created. As pipeline progress occurs, the accumulating fact table row is revisited and updated. This consistent updating of accumulating snapshot fact rows is unique among the three types of fact tables. In addition to the date foreign keys associated with each critical process step, accumulating snapshot fact tables contain foreign keys for other dimensions and optionally contain degenerate dimensions. They often include numeric lag measurements consistent with the grain, along with milestone completion counters.

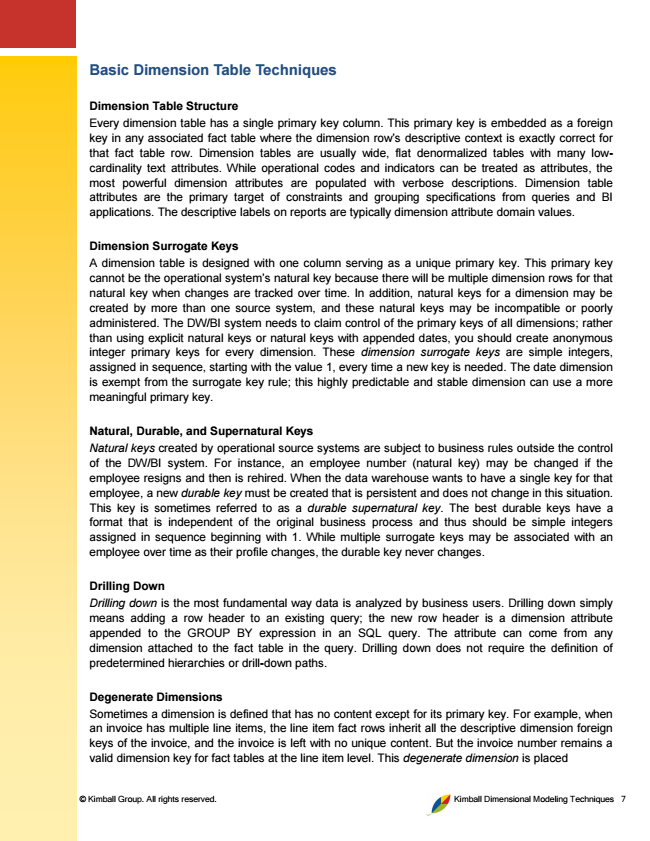
Factless Fact Tables Although most measurement events capture numerical results, it is possible that the event merely records a set of dimensional entities coming together at a moment in time. For example, an event of a student attending a class on a given day may not have a recorded numeric fact, but a fact row with foreign keys for calendar day, student, teacher, location, and class is well-defined. Likewise, customer communications are events, but there may be no associated metrics. Factless fact tables can also be used to analyze what didn’t happen. These queries always have two parts: a factless coverage table that contains all the possibilities of events that might happen and an activity table that contains the events that did happen. When the activity is subtracted from the coverage, the result is the set of events that did not happen.

Aggregate Fact Tables or Cubes Aggregate fact tables are simple numeric rollups of atomic fact table data built solely to accelerate query performance. These aggregate fact tables should be available to the BI layer at the same time as the atomic fact tables so that BI tools smoothly choose the appropriate aggregate level at query time. This process, known as aggregate navigation, must be open so that every report writer, query tool, and BI application harvests the same performance benefits. A properly designed set of aggregates should behave like database indexes, which accelerate query performance but are not encountered directly by the BI applications or business users. Aggregate fact tables contain foreign keys to shrunken conformed dimensions, as well as aggregated facts created by summing measures from more atomic fact tables. Finally, aggregate OLAP cubes with summarized measures are frequently built in the same way as relational aggregates, but the OLAP cubes are meant to be accessed directly by the business users.



© Kimball Group. All rights reserved. Kimball Dimensional Modeling Techniques 6

Consolidated Fact Tables It is often convenient to combine facts from multiple processes together into a single consolidated fact table if they can be expressed at the same grain. For example, sales actuals can be consolidated with sales forecasts in a single fact table to make the task of analyzing actuals versus forecasts simple and fast, as compared to assembling a drill-across application using separate fact tables. Consolidated fact tables add burden to the ETL processing, but ease the analytic burden on the BI applications. They should be considered for cross-process metrics that are frequently analyzed together.



© Kimball Group. All rights reserved. Kimball Dimensional Modeling Techniques 7

**Basic Dimension Table Techniques**

Dimension Table Structure Every dimension table has a single primary key column. This primary key is embedded as a foreign key in any associated fact table where the dimension row’s descriptive context is exactly correct for that fact table row. Dimension tables are usually wide, flat denormalized tables with many low- cardinality text attributes. While operational codes and indicators can be treated as attributes, the most powerful dimension attributes are populated with verbose descriptions. Dimension table attributes are the primary target of constraints and grouping specifications from queries and BI applications. The descriptive labels on reports are typically dimension attribute domain values.

Dimension Surrogate Keys A dimension table is designed with one column serving as a unique primary key. This primary key cannot be the operational system’s natural key because there will be multiple dimension rows for that natural key when changes are tracked over time. In addition, natural keys for a dimension may be created by more than one source system, and these natural keys may be incompatible or poorly administered. The DW/BI system needs to claim control of the primary keys of all dimensions; rather than using explicit natural keys or natural keys with appended dates, you should create anonymous integer primary keys for every dimension. These dimension surrogate keys are simple integers, assigned in sequence, starting with the value 1, every time a new key is needed. The date dimension is exempt from the surrogate key rule; this highly predictable and stable dimension can use a more meaningful primary key.

Natural, Durable, and Supernatural Keys Natural keys created by operational source systems are subject to business rules outside the control of the DW/BI system. For instance, an employee number (natural key) may be changed if the employee resigns and then is rehired. When the data warehouse wants to have a single key for that employee, a new durable key must be created that is persistent and does not change in this situation. This key is sometimes referred to as a durable supernatural key. The best durable keys have a format that is independent of the original business process and thus should be simple integers assigned in sequence beginning with 1. While multiple surrogate keys may be associated with an employee over time as their profile changes, the durable key never changes.

Drilling Down Drilling down is the most fundamental way data is analyzed by business users. Drilling down simply means adding a row header to an existing query; the new row header is a dimension attribute appended to the GROUP BY expression in an SQL query. The attribute can come from any dimension attached to the fact table in the query. Drilling down does not require the definition of predetermined hierarchies or drill-down paths.

Degenerate Dimensions Sometimes a dimension is defined that has no content except for its primary key. For example, when an invoice has multiple line items, the line item fact rows inherit all the descriptive dimension foreign keys of the invoice, and the invoice is left with no unique content. But the invoice number remains a valid dimension key for fact tables at the line item level. This degenerate dimension is placed