Muni University

BACHELOR OF INFORMATION SYSTEMS DEGREE

Year 3, Semester 2

Course Name: **Data Warehousing and Business Intelligence**

Course Code: **ISM 3203**

NKAMWESIGA Nicholas

Introducing Dimensional Modeling

- Dimensional modeling overview
- Dimensional modeling
  - Fact tables
  - Dimension tables
  - Four-step dimensional design process
- Enterprise data warehouse bus architecture
- Fables and falsehood about dimensional modeling

Dimensional modeling overview

- ***Dimensional modelling*** is a logical design technique for structuring data so that it's intuitive to business users and delivers fast query performance.
- Dimensional models stored in a relational database platform are known as **star schemas**.
- Dimensional models stored in multidimensional **online analytical processing (OLAP)**

Cont…

- ***Normalized Modeling:***
  - The design technique that seeks to eliminate data redundancies.
  - Data is divided into many discrete entities, each of which becomes a table in the relational database.
  - Benefit:
    - Simple and fast transaction loading and updating
- **Dimensional** and **3NF** modeling both result in physical tables substantiated in a relational database system.

Cont…

- ***Steps for converting normalized structures to dimensional structure:***
  - ***Step 1:*** Designate the many to many relationships in the normalized model containing numeric and additive non-key metrics as fact tables.
    - Fact tables are typically normalized to 3NF
  - ***Step 2:*** Denormalize the remaining tables into flat dimension tables with single-part keys that connect directly to the fact table.
    - Dimension tables most often resemble second normal form tables with many low cardinality descriptors.

Cont…

- ***Benefits of dimensional modeling***
  - **Understandability**; easy of use
  - **Query performance**
  - **Symmetry** between dimension and its equivalent entry point into the fact table
    - This allows dimensional model to withstand unexpected changes in query patterns.
    - The query optimization for query optimization for star join databases is simple, predictable, and controllable
  - Best-of-breed ETL and BI tool vendors have **incorporated dimensional intelligence** into their products, like pre-built wizards to facilitate standard dimensional capabilities e.g. handling of slowly changing dimension.
  - Dimensional models are gracefully **extensible** to accommodate unexpected new data.

Dimensional modeling- **Fact tables**

- Fact tables store the performance measurements generated by the organization's business activities or events.
  - Fact is a single performance measure
    - A fact is a variable;
    - It's evaluation happens at the time of measurement events
    - It's numeric and additive
    - Facts conform if their definition are identical
      - Conformed facts are allowed to have the same standardized name in separate tables
    - Facts are huge, with over millions of rows

Dimensional modeling- **Fact tables**

- **Fact Tables**
  - Are characterized by a multipart key made up of foreign keys from the intersecting dimension tables involved in a business process.
  - The multipart key means that fact tables always express a many-to-many relationship.
  - Shouldn’t violate the referential integrity; every foreign key in the fact table must match to a unique primary key in the corresponding dimension table.
    - The fact table’s foreign keys should never be null.
  - The primary key of the fact table is typically a subset of the dimension foreign keys.

Dimensional modeling- **Fact tables**

- **Fact table granularity**
  - The table's ***grain*** is the business definition of the measurement event that produces the fact row.
  - All rows in the fact table should have a single grain
  - The grain is the anchor which needs enough dimensional context in its design to implementation stages.
  - Fact tables should contain the lowest and most detailed grain captured by a business process.
  - Atomic data withstand assaults from unpredictable ad hoc queries because the details can be captured.
  - The granular atomic fact tables are more impervious to changes
    - They can be extended by adding newly sourced facts, newly sourced dimension attributes, and by entirely addition new dimension.

Dimensional modeling- **Dimension Table**

- Dimension Tables are full of stark and sleek; they are filled with big and bulky description fields.
  - The power of the warehouse is correlates to its quality and depth of the dimension attribute.
    - Robust dimensions result in robust querying and analysis capabilities.
- Dimension attributes are either textual or discrete numbers that behave like texts.
- Characteristics of dimensional attributes; they should be;
  - Verbose (labels consisting full words)
  - Descriptive
  - Complete (no missing values)
  - Discretely valued (take only one value for each row in the dimension)
  - Quality assured (no misspellings, impossible values, obsolete or orphaned values, or cosmetically different versions of the same attribute)

Dimensional modeling- **Dimension Table**

- Dimension tables represent hierarchical relationships.
- Dimension tables consists of highly correlated clumps of attributes grouped to represent the key objects of a business, such as products, customers, employees, or facilities
- **Dimension Table Keys**
  - Whereas fact tables have a multipart key, dimension rows are uniquely identified by single key field.
  - Dimension table's primary keys are simple integers assigned in sequence starting with 1; when a key for a new dimension record is created, 1 is simply added to the previously calculated value.
  - The surrogate keys are meaningless, i.e. they merely serve as join fields between the fact and dimension tables

Dimensional modeling- **Dimension Table**

Advantages of surrogate dimension keys instead of referencing the operational natural keys:

- Performance
- Buffer from operational key management practices
- Mapping to integrate disperse sources
- Handle unknown or not applicable conditions
- Track changes in dimension attribute values

Dimensional modeling- **Dimension Table**

Conformed Dimensions

- Are share across the enterprise's data warehouse environment, joining to multiple fact tables representing various business processes.
- Dimensions conform when one dimension is a perfect subset of a more detailed, granular dimension table.
- Shrunken conformed dimension tables are created to describe fact tables that either naturally capture measurements at a higher level of detail, or facts that have been aggregated to a less granular, rolled-up level for performance reasons.

Dimensional modeling- **Dimension Table**

Benefits of conformed dimensions

- Consistency
  - They ensure every fact table is filtered consistently and the resulting query answer sets are labelled consistently
- Integration
  - Strict adherence to conformed dimensions allows the DW/BI environment to operate as an integrated whole.
- Reduced development time to market
  - While there's an organizational investment required to define and manage conformed dimensions, once built, they can dramatically reduce the development time for a project because the common dimensions are available without recreating them.

Dimensional modeling- **Four-step dimensional design process**

- Step 1: Choose the business process
  - Done during the prioritization activity with senior business management.
- Step 2: Declare the grain
  - Done by the design team
  - Fact tables designed with most granular data produce the most robust design
- Step 3: Identify the dimension
  - Begin to think about the foreign keys
  - The grain determines the a primary or nominal dimensions
  - The design is established with additional dimensions that take on a unique value at the declared grain of the fact table
- Step 4: Identify the facts
  - Critically select the facts or metrics that are applicable to the business process

Enterprise data warehouse bus architecture

- Planning crisis
- Bus architecture
- Value chain implication
- Common matrix mishaps
- Taking the pledge
- More dimensions
  - Date and Time
  - Degenerate Dimension
  - Slowly changing dimension
  - Role-playing dimension
  - Junk dimension
  - Snow flaking and Outriggers
  - Handling hierarchies

Fables and falsehood about dimensional modeling

- Fables caused by focusing on departmental reports
- Fables caused by premature summarization
- Fables caused by overvaluing normalization

End

Thank you!