Data Warehousing

Keith Henry
Data Warehouse Architect

UK
UNIVERSITY OF KENTUCKY
Center for Clinical and Translational Science

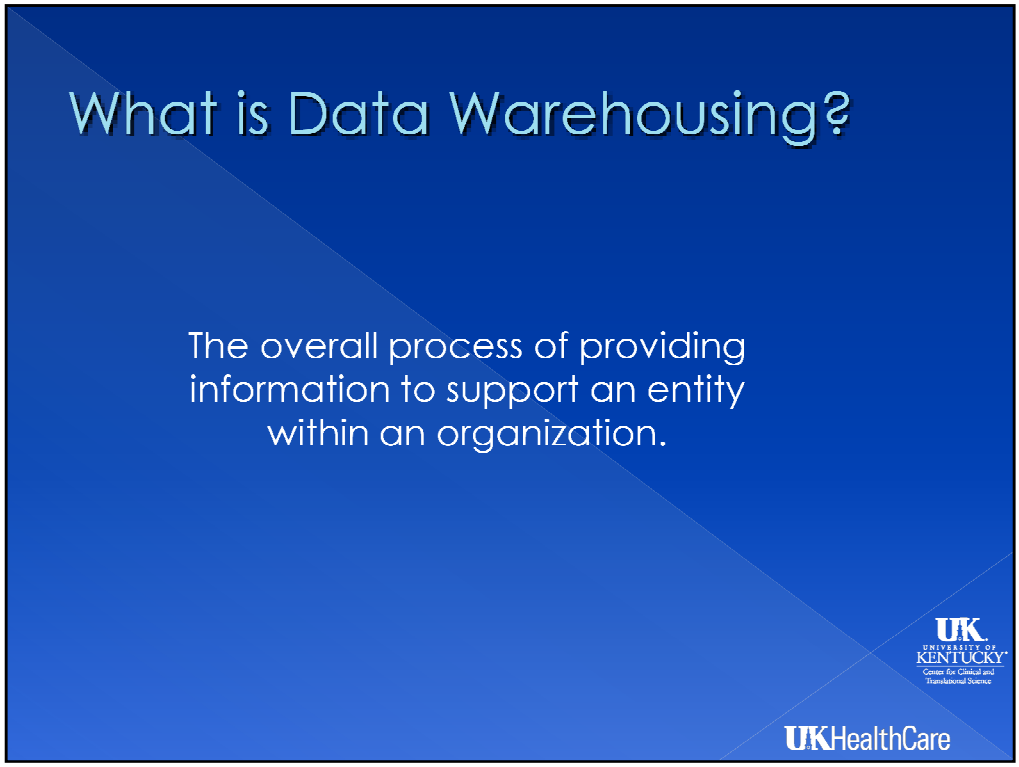
UKHealthCare

Data Warehousing

- What is Data Warehousing?
- Why do we need a DW?
- Goals of a Data Warehouse
- Transactional Model
- Dimensional Model

UK University of Kentucky Center for Clinical and Translational Science

UKHealthCare



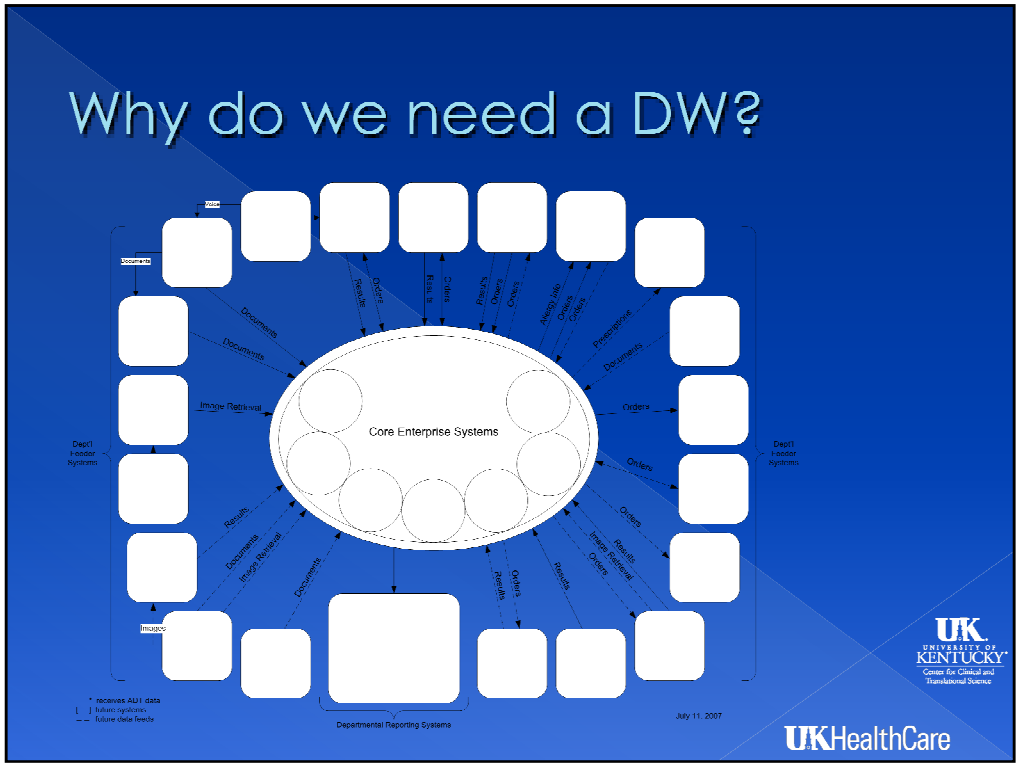
So, It IS NOT just a database. Process, process…..

It is called Data Warehousing, NOT “Database” Warehousing.

It’s a simple idea

Consists of 2 components:

Back room & Front room

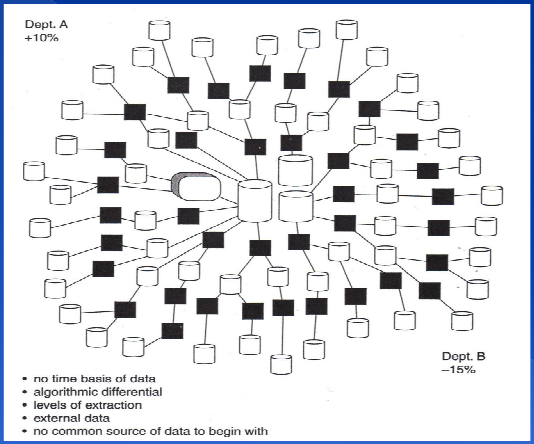


What is the problem?

This is a distilled version.

There is a complex process to gather the information from multiple systems. Sometimes even finding the data you need can be daunting.

Why do we need a DW?



What is the problem?

The Spiderweb problem

The core systems from the prior slide are in the middle here.

Why do we need a DW?

- Need a centralized easy way for people to retrieve and analyze information
- Facilitate the ability to merge data from different source systems.
  - i.e. SCM Clinical with Perinatal
- Relieve the online systems from the burden of reporting.

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

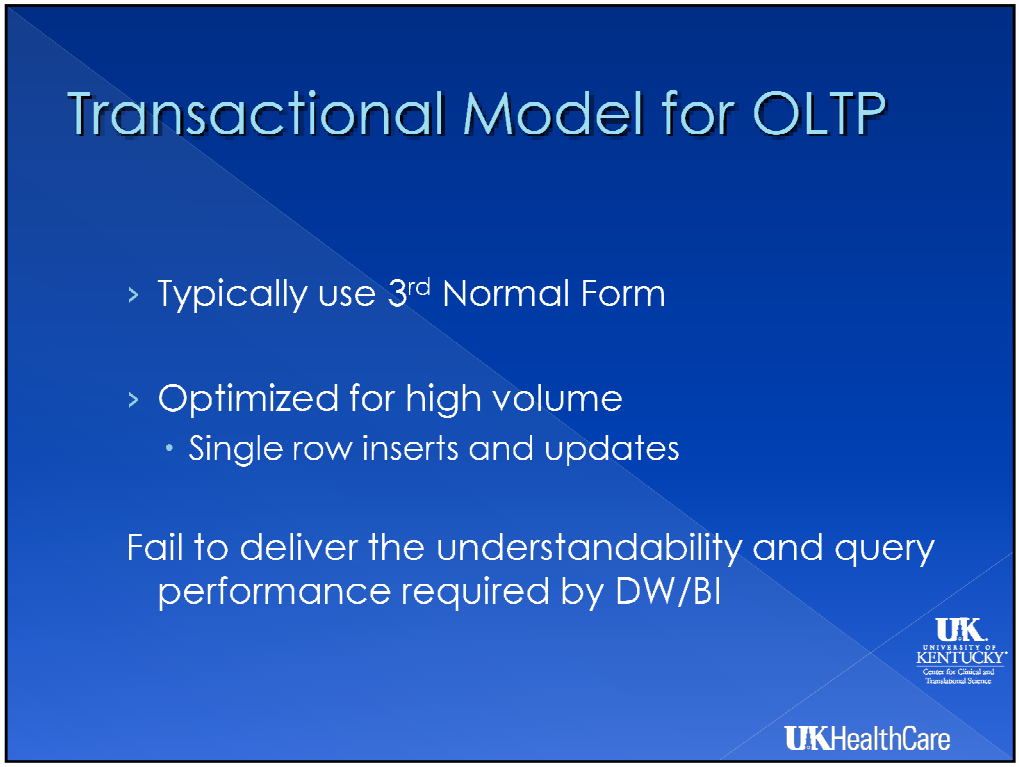
UKHealthCare

Consistent view of data example:

Physician code different at Good Sams vs. UK Healthcare

So, if you want to measure something about the physician you need to know this.

Tools need to be easy to use and more importantly, easy to learn to use.



Normalized model. Eliminating redundant data. Thus joins are used to put data back together.

Dimensional Model for DW

- What is dimensional modeling?

A logical design technique for structuring data so that it's intuitive to business users and delivers fast query performance

Dimensional modeling divides the world into *measurements* and *context*.

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

Dimensional Model for DW

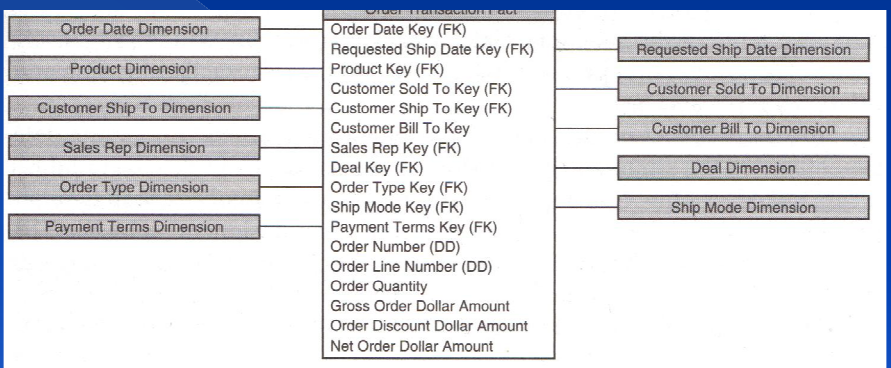


Figure 6-1 Dimensional model of the orders business process for a manufacturer.

Dimensional Model for DW

- Benefits of Dimensional Modeling
  - Understandability
  - Query Performance
  - Symmetry

"Make everything as simple as possible, but not simpler"
Albert Einstein

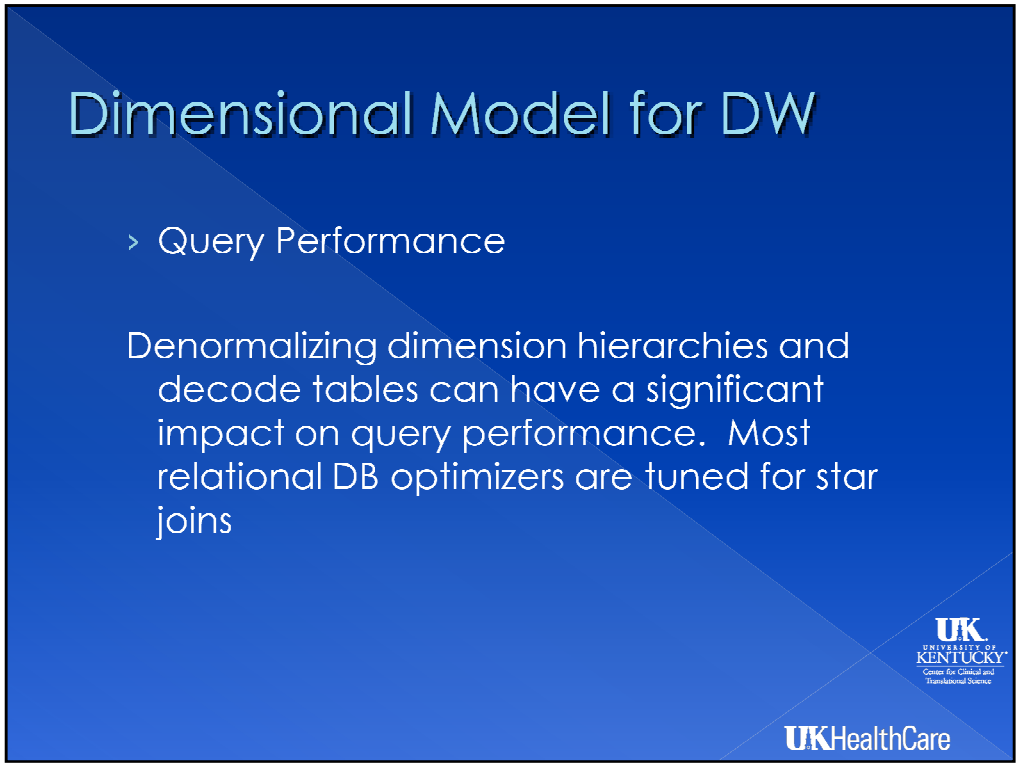
UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

Dimensional Model for DW

- Understandability

The dimensional model is easier for a business user to understand than the typical normalized model because information is grouped into coherent business categories (dimensions)



The database engine leverages the star join by first constraining the dim tables and gathering keys satisfying the query filters, and then querying the fact table with the cartesian product of the relevant dimension keys.

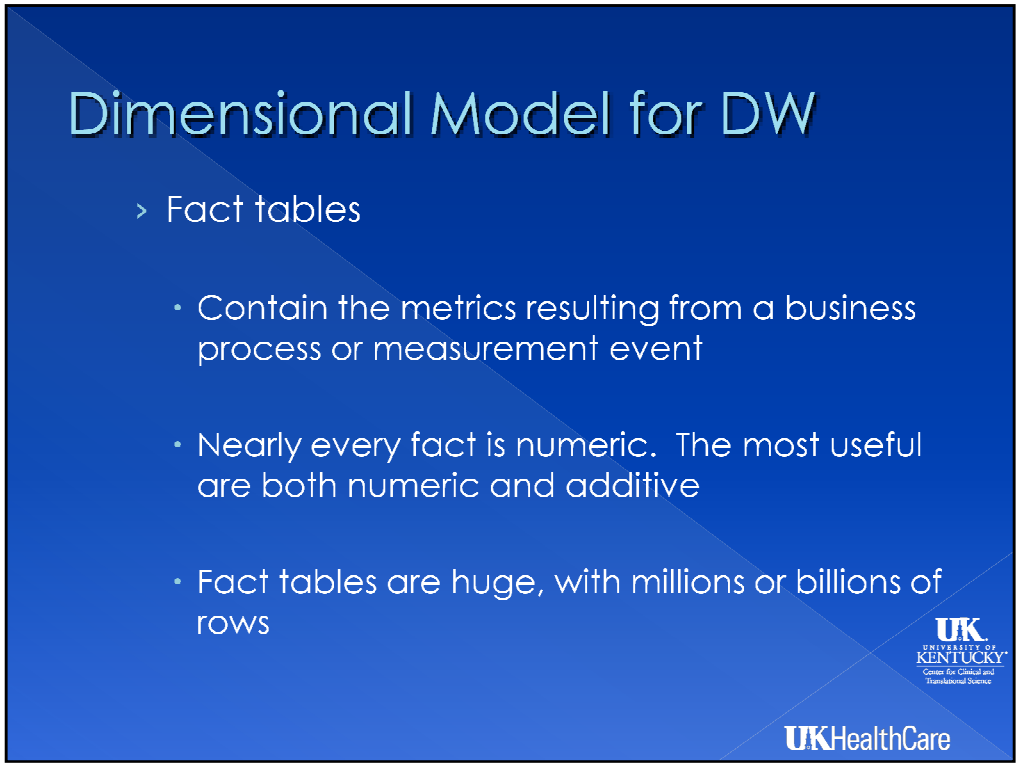
Dimensional Model for DW

- Symmetry

Each dimension is an equivalent entry point into the fact table.

Dimensional Model for DW

- The two primary constructs of dimensional modeling are:
  - Fact Tables
  - Dimension Tables.



Fact tables ex: such as sales ordering process or service call event.

Dimensional Model for DW

- Fact table keys
  - Fact table keys are characterized by a multipart key made up of foreign keys from the intersecting dimension tables
  - The multipart key means that fact tables always express a many-to-many relationship

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

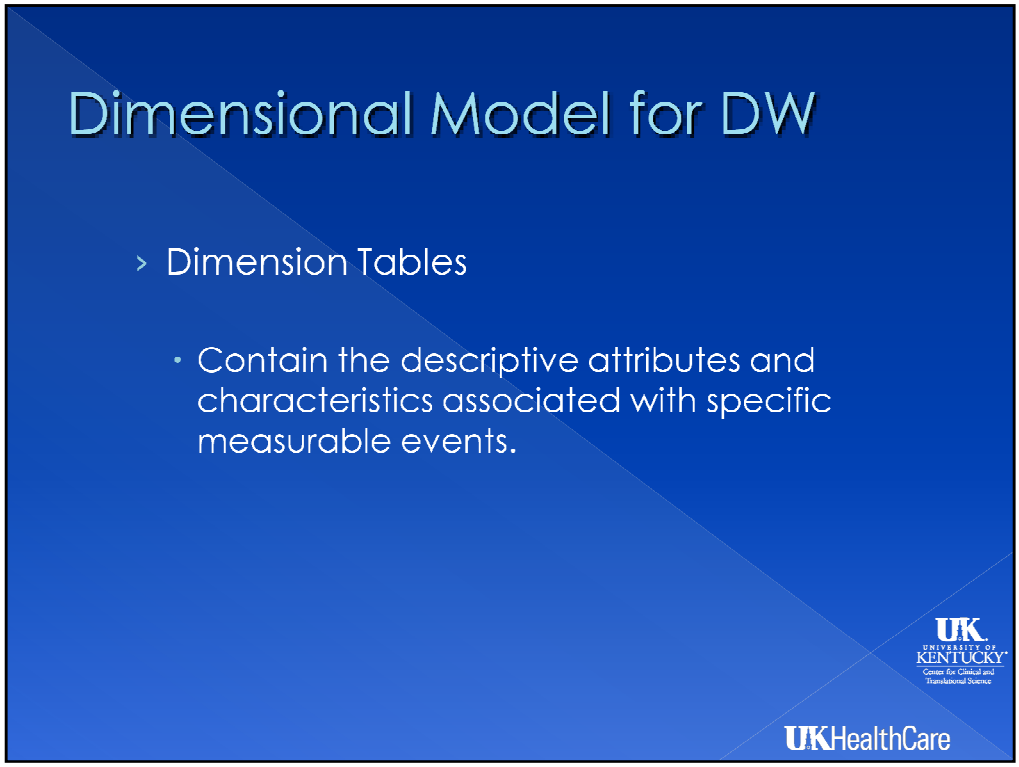
See diagram a few slides back.

Dimensional Model for DW

- Fact table granularity
  - The fact table's grain is the business definition of the measurement event that produces the fact row.
  - A fact row is created when _____ occurs.
  - Fact tables should contain the lowest, most detailed atomic grain captured by a business process.

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare



Dimension table ex: such as customer, product, or sales rep associated with an order

Dimensional Model for DW

- Driven by the following requirement:
  - I want to measure ? By ?

I want to see PROFIT by CUSTOMER
I want to see PROFIT by BUSINESS AREA

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

Dimensional Model for DW

- Dimension table keys
  - Dimension rows are uniquely identified by a single key field.
  - These primary keys should be simple integers.
  - These surrogate keys are meaningless. They merely serve as join fields between the fact and dimension tables.

UK University of Kentucky Center for Clinical and Translational Science

UK HealthCare

Performance

Insulation from source system key changes. Re-using obsolete products, etc.

Handle unknown or NA conditions. Avoid the NULL

Track changes in dimensions, re type 2 SCD

Dimensional Model for DW

- Conformed dimensions
  - Standardized conformed dimensions are the goal for any well-architected data warehouse.
  - Shared across the DW environment, joining to multiple fact tables representing various business processes.

Dimensional Model for DW

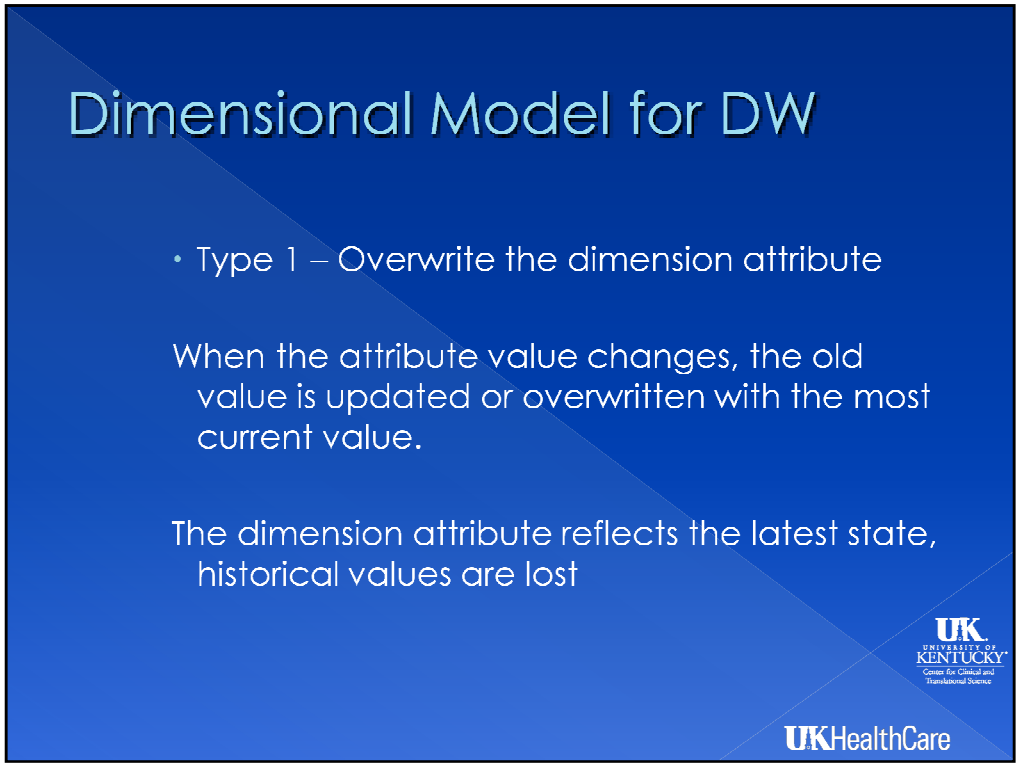
- Conformed dimensions deliver the following:
  - Consistency. Ensure that every fact table is filtered consistently and the query answer sets are labeled consistently.
  - Integration. Allow queries to drill across fact tables.
  - Reduced development time to market

UK University of Kentucky Center for Clinical and Translational Science

UKHealthCare

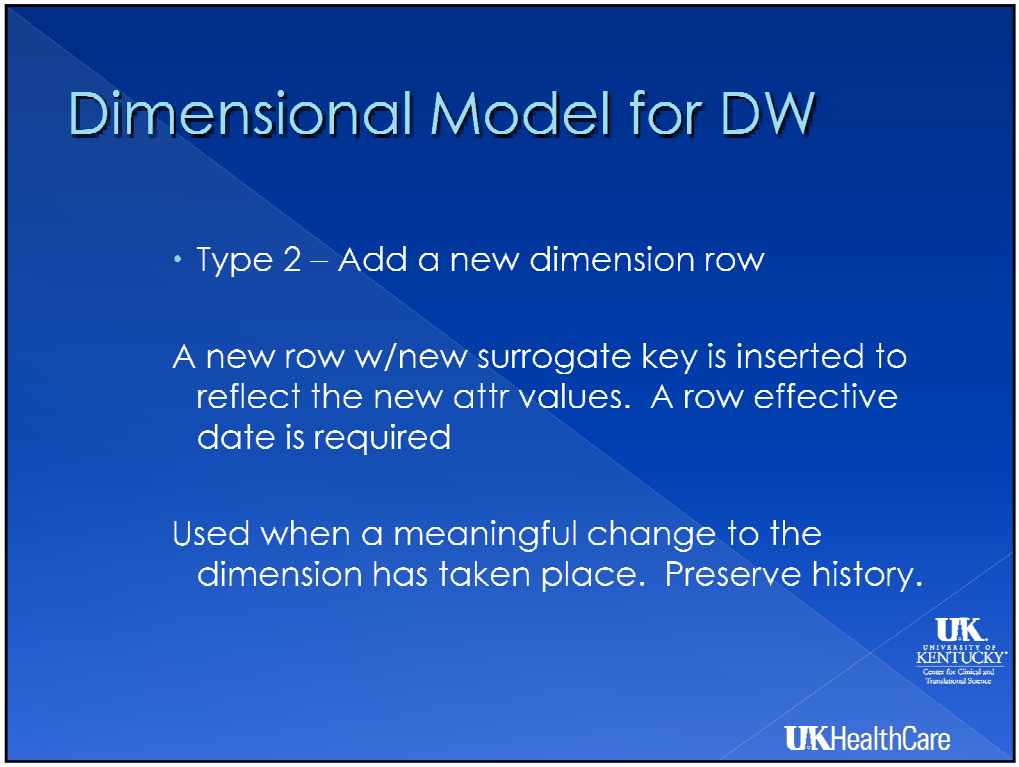
Dimensional Model for DW

- Slowly changing dimensions:
  - Type 1 – Overwrite the dimension attribute
  - Type 2 – Add a new dimension row
  - Type 3 – Add a new dimension attribute

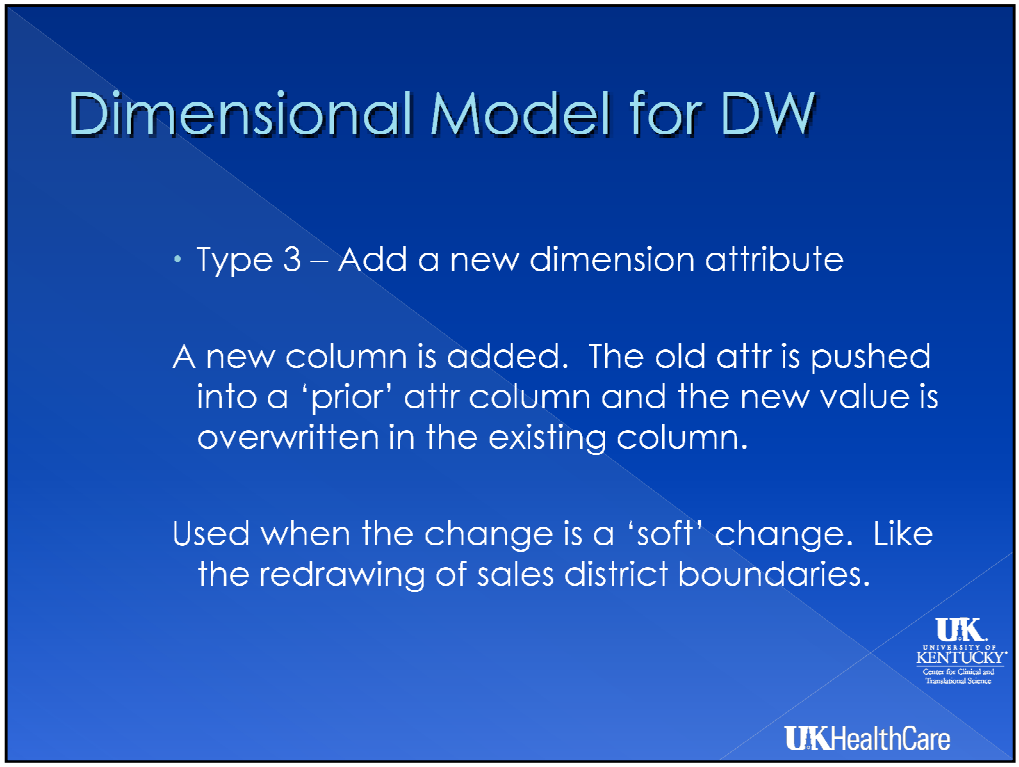


Easiest to implement, but caveat….

If fact data has been previously aggregated based on the type 1 attr, when the dimension value is overwritten, then the aggregation is broken.



Each surrogate key is used in the corresponding fact rows during the time when the particular instance was valid.



In other words although the change has occurred it is still possible to act as if the change had not occurred

Dimensional Model for DW

- The special Date dimension
  - Virtually every fact table is a time series of observations; fact tables always have one or more date dimensions.
  - Calendar navigation should be driven through the date dimension table, not through hard coded application logic.

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

Surrogate keys are used.

Dimensional Model for DW

Date Dimension
Date Key (PK)
Date
Full Date Description
Day of Week
Day Number in Calendar Month
Day Number in Calendar Year
Day Number in Fiscal Month
Day Number in Fiscal Year
Last Day in Week Indicator
Last Day in Month Indicator
Calendar Week Ending Date
Calendar Week Number in Year
Calendar Month
Calendar Month Number in Year
Calendar YYYY-MM
Calendar Quarter
Calendar Year-Quarter
Calendar Year
Fiscal Week
Fiscal Week Number in Year
Fiscal Month
Fiscal Month Number in Year
Fiscal Year-Month
Fiscal Quarter
Fiscal Year
Holiday Indicator
Weekday Indicator
Workday Indicator
SQL Date Stamp
... and more

Figure 6-6 Example date dimension table.

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

Dimensional Model for DW

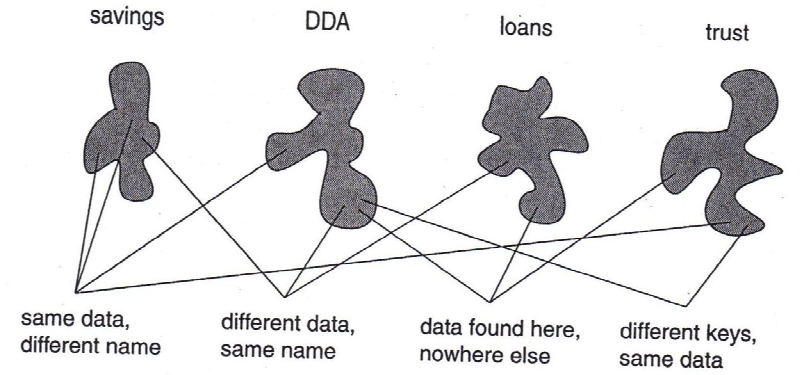


Figure 3.2 Data across the different applications are terribly unintegrated.

UK HealthCare

Extra slides for discussion, if necessary

Dimensional Model for DW

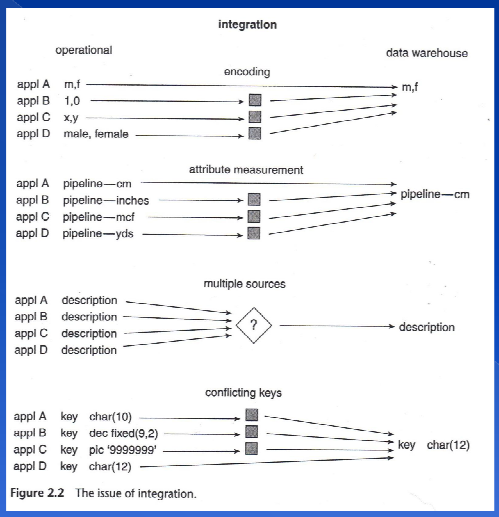


Figure 2.2 The issue of integration.

Extra slides for discussion, if necessary

Dimensional Model for DW

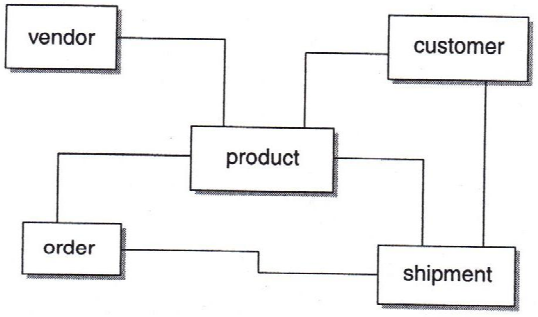


Figure 3.56 A simple two-dimensional data model gives the impression that all entities are equal.

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

Extra slides for discussion, if necessary

Dimensional Model for DW

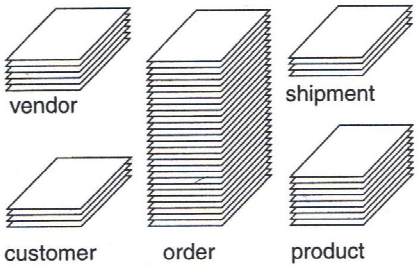


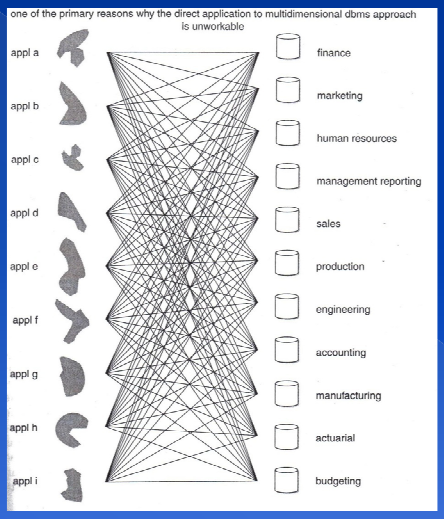
Figure 3.57 A three-dimensional perspective of the entities shows that the entities are anything but equals. Some entities contain far more occurrences of the data than other entities.

UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

Extra slides for discussion, if necessary

Dimensional Model for DW



UK UNIVERSITY OF KENTUCKY Center for Clinical and Translational Science

UKHealthCare

Extra slides for discussion, if necessary

References:

Building the Data Warehouse, 2nd edition, W.H. Inmon, 1996

The Data Warehouse Toolkit (Practical Techniques for Building Dimensional Data Warehouses), Ralph Kimball, 1996

Data Warehouse Design Solutions, Christopher Adamson, Michael Venerable, 1998

UK University of Kentucky Center for Clinical and Translational Science

UKHealthCare

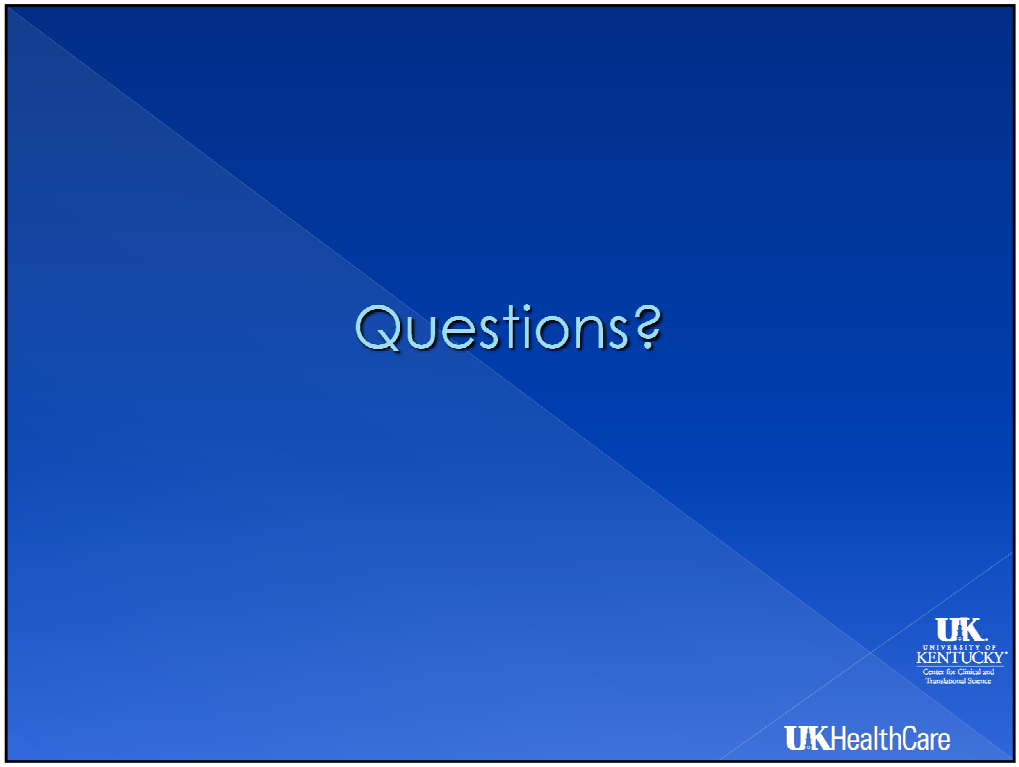
References:

The Data Warehouse ETL Toolkit (Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data), Ralph Kimball, Joe Caserta, 2004

The Data Warehouse Lifecycle Toolkit 2nd ed (Practical Techniques for Building Data Warehouse and Business Intelligence Systems), Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker, 2008

UK University of Kentucky Center for Clinical and Translational Science

UKHealthCare


Questions?
UK
UNIVERSITY OF KENTUCKY
Center for Clinical and Translational Science
UKHealthCare