Data Warehouse: The Choice of Inmon versus Kimball

Ian Abramson
IAS Inc.

Agenda

- The 2 Approaches
  - Bill Inmon – Enterprise Warehouse (CIF)
  - Ralph Kimball – Dimensional Design
- Similarities
- Differences
- Choices

DW History

- 1990
  - Inmon publishes “Building the Data Warehouse”
- 1996
  - Kimball publishes “The Data Warehouse Toolkit”
- 2002
  - Inmon updates book and defines architecture for collection of disparate sources into detailed, time variant data store.
    - The top down approach
  - Kimball updates book and defines multiple databases called data marts that are organized by business processes, but use enterprise standard data bus
    - The bottom-up approach

The Data Warehouse Is:

- **Bill Inmon,** an early and influential practitioner, has formally defined a data warehouse in the following terms;
  - Subject-oriented
    - The data in the database is organized so that all the data elements relating to the same real-world event or object are linked together;
  - Time-variant
    - The changes to the data in the database are tracked and recorded so that reports can be produced showing changes over time;
  - Non-volatile
    - Data in the database is never over-written or deleted - once committed, the data is static, read-only, but retained for future reporting; and
  - Integrated
    - The database contains data from most or all of an organization's operational applications, and that this data is made consistent

- **Ralph Kimball,** a leading proponent of the dimensional approach to building data warehouses, provides a succinct definition for a data warehouse:
  - “A copy of transaction data specifically structured for query and analysis.“

Ref: wikipedia

What are they saying?

- These two influential data warehousing experts represent the current prevailing views on data warehousing.

- Kimball, in 1997, stated that
  - "...the data warehouse is nothing more than the union of all the data marts",
  - Kimball indicates a bottom-up data warehousing methodology in which individual data marts providing thin views into the organizational data could be created and later combined into a larger all-encompassing data warehouse.
- Inmon responded in 1998 by saying,
  - "You can catch all the minnows in the ocean and stack them together and they still do not make a whale,"
  - This indicates the opposing view that the data warehouse should be designed from the top-down to include all corporate data. In this methodology, data marts are created only after the complete data warehouse has been created.

What is a Data Warehouse:

- The single organizational repository of enterprise wide data across many or all lines of business and subject areas.
  - Contains massive and integrated data
  - Represents the complete organizational view of information needed to run and understand the business

What is a Data Mart?

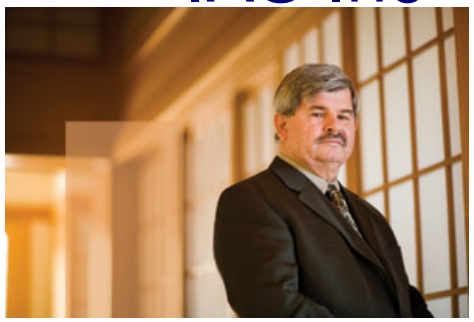
- The specific, subject oriented, or departmental view of information from the organization. Generally these are built to satisfy user requirements for information
  - Multiple data marts for one organization
  - A data mart is built using dimensional modeling
  - More focused
  - Generally smaller, selected facts and dimensions
  - Integrated

Data Warehouses vs. Data Marts

- **Data Warehouses**:
  - Scope
    - Application independent
    - Centralized or Enterprise
    - Planned
  - Data
    - Historical, detailed, summary
    - Some denormalization
  - Subjects
    - Multiple subjects
  - Source
    - Many internal and external sources
  - Other
    - Flexible
    - Data oriented
    - Long life
    - Single complex structure

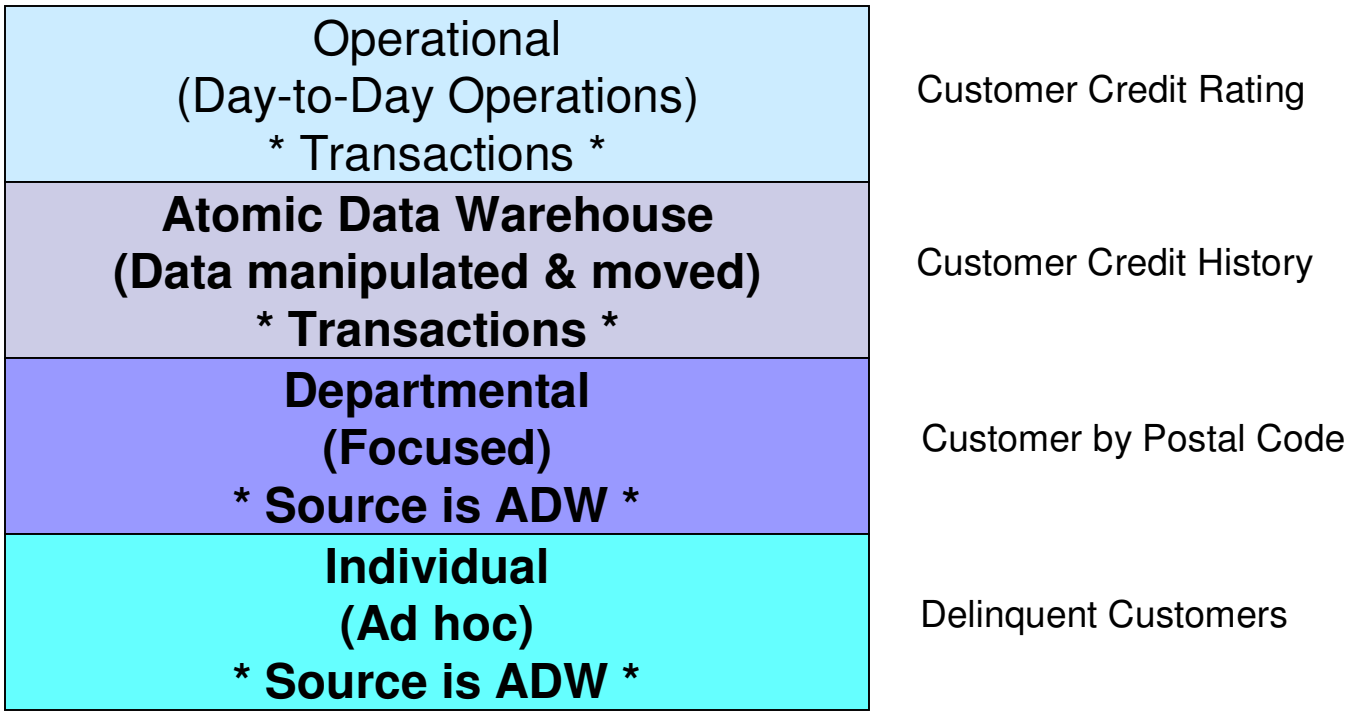
- **Data Marts**:
  - Scope
    - Specific application
    - Decentralized by group
    - Organic but may be planned
  - Data
    - Some history, detailed, summary
    - High denormalization
  - Subjects
    - Single central subject area
  - Source
    - Few internal and external sources
  - Other
    - Restrictive
    - Project oriented
    - Short life
    - Multiple simple structures that may form a complex structure

The Inmon Model



- Consists of all databases and information systems in an organization…..
  - The CIF (Corporate Information Factory)
- Defines overall database environment as:
  - Operational
  - **Atomic data warehouse**
  - **Departmental**
  - **Individual**
- The Warehouse is part of the bigger whole (CIF)

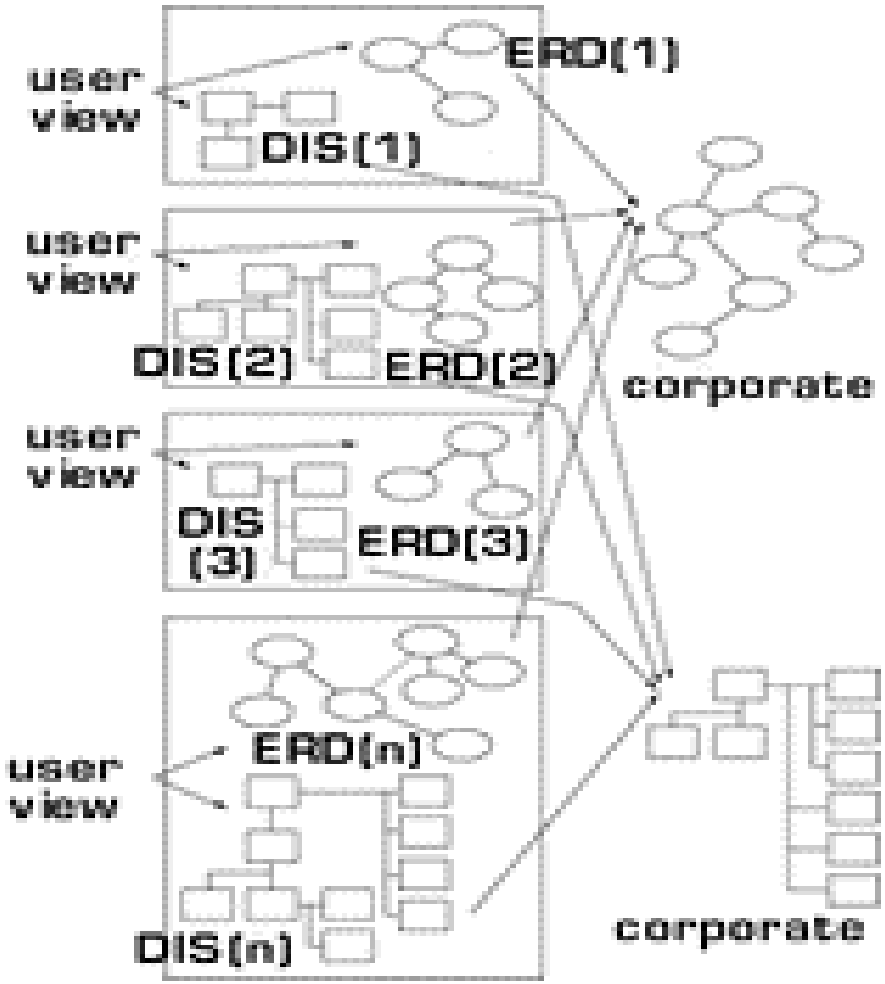
The Data Warehouse



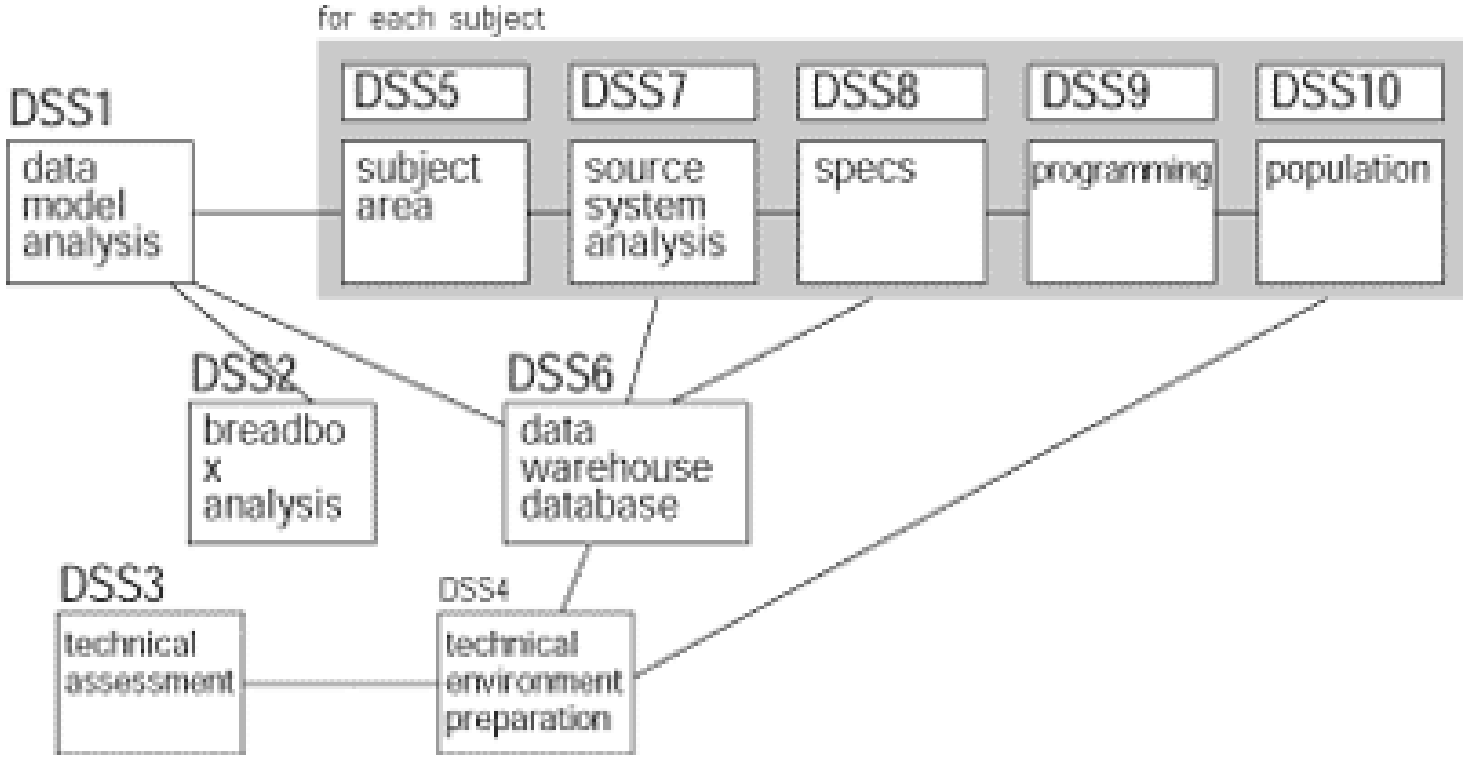
Inmon Modeling

- Three levels of data modeling
  - ERD (Entity Relationship Diagram)
    - Refines entities, attributes and relationships
  - Mid-Level model (*DIS*)
    - Data Item Sets
    - Data sets by department
    - Four constructs:
      - Primary data groupings
      - Secondary data groupings
      - Connectors
      - “Type of” data
  - Physical data model
    - Optimize for performance (de-normalize)

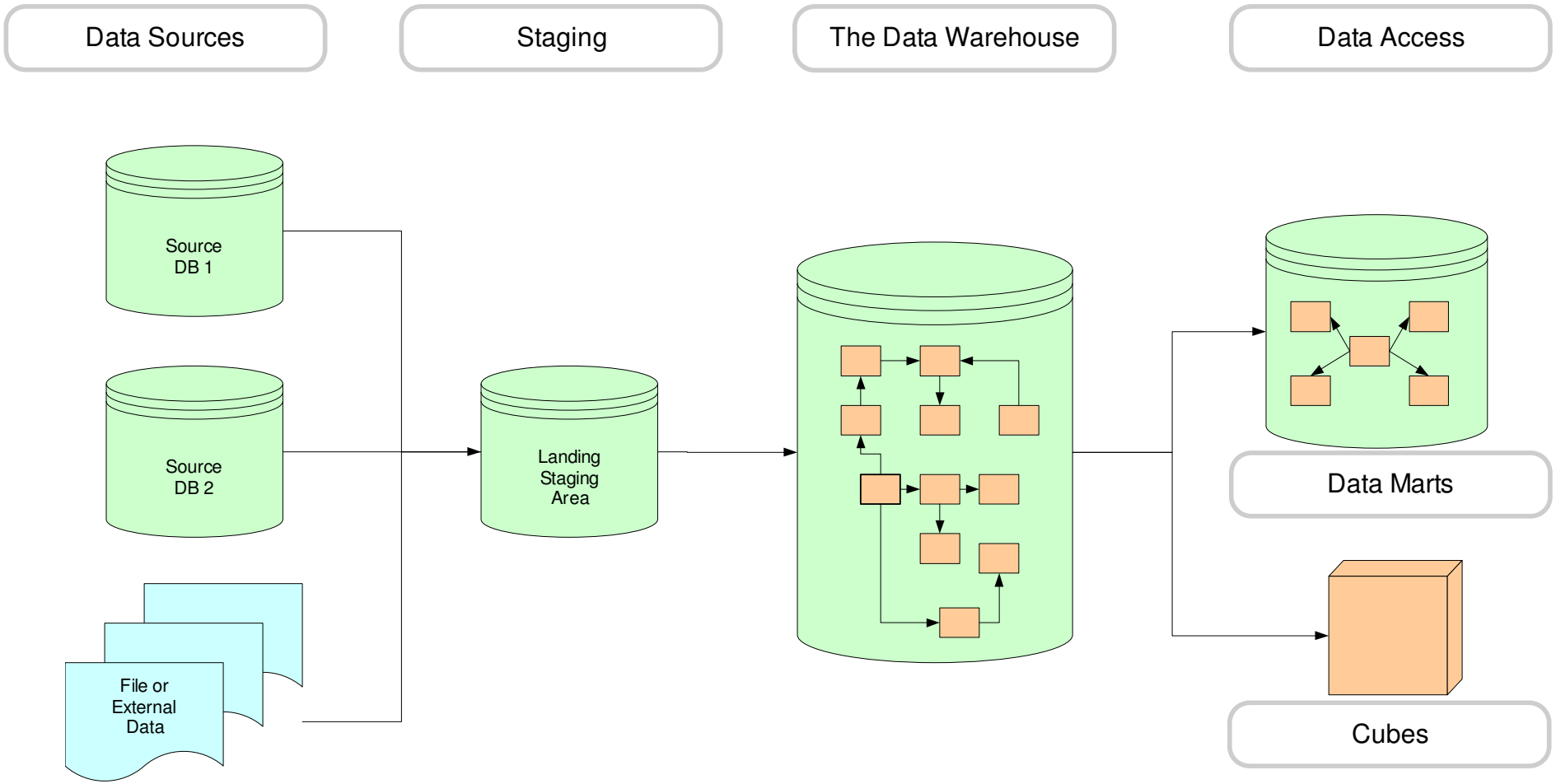
Relationship between Levels One and Two of Inmon's Data model (Inmon,2002)



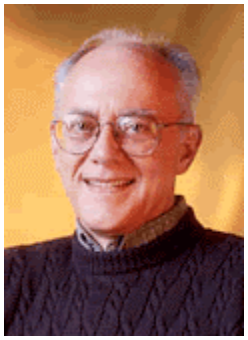
The Warehouse Architecture



The Inmon Warehouse



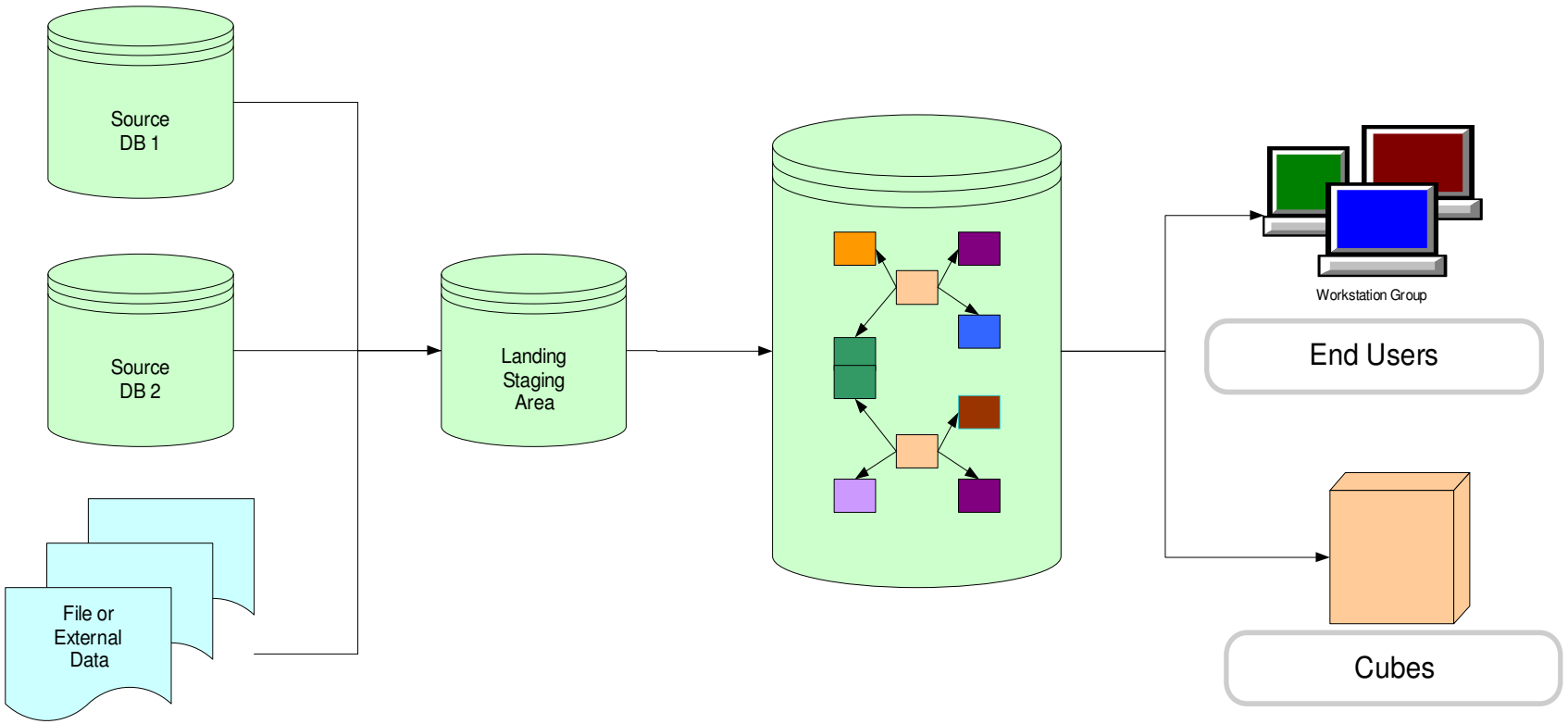
The Kimball Approach



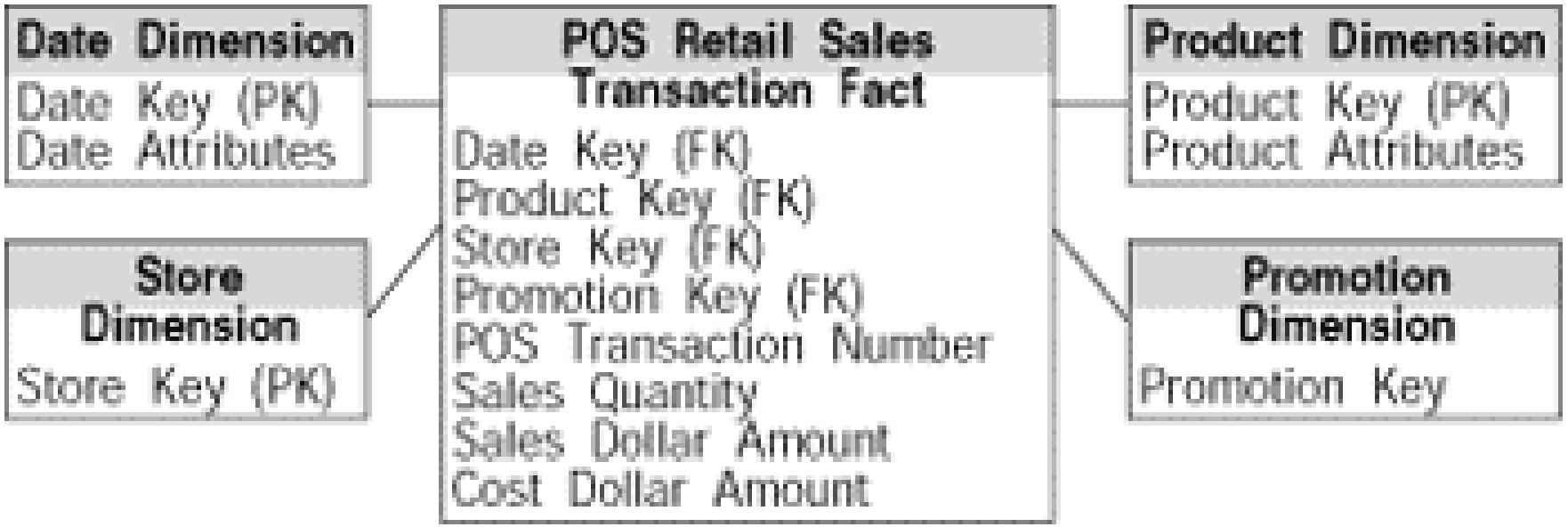
- The Dimensional Data Model
  - Starts with tables
    - Facts
    - Dimensions
  - Facts contain metrics
  - Dimensions contain attributes
    - May contain repeating groups
  - Does not adhere to normalization theory
  - User accessible

The Kimball Data Lifecycle

Data Sources | Staging | The Data Warehouse | Data Access



The Dimensional Model



The Kimball Data Bus

- Data is moved to staging area
  - Data is scrubbed and made consistent
- From Staging Data Marts are created
- Data Marts are based on a single process
- Sum of the data marts can constitute an Enterprise Data Warehouse
- Conformed dimensions are the key to success

The Kimball Design Approach

- Select business process
- Declare the grain
- Choose dimensions
- Identify facts (metrics)

Kimball's Philosophy

- Make data easily accessible
- Present the organization's information consistently
- Be adaptive and resilient to change
- Protect information
- Service as the foundation for improved decision making.

Getting Started with Choices

- Kimball
  - Will start with data marts
  - Focused on quick delivery to users
- Inmon
  - Will focus on the enterprise
  - Organizational focus

Kimball vs. Inmon

- Inmon:
  - Subject-Oriented
  - Integrated
  - Non-Volatile
  - Time-Variant
  - Top-Down
  - Integration Achieved via an Assumed Enterprise Data Model
  - Characterizes Data marts as Aggregates
- Kimball
  - Business-Process-Oriented
  - Bottom-Up and Evolutionary
  - Stresses Dimensional Model, Not E-R
  - Integration Achieved via Conformed Dimensions
  - Star Schemas Enforce Query Semantics

The Comparison (Methodology and Architecture)

	Inmon	Kimball
Overall approach	Top-down	Bottom-up
Architectural structure	Enterprise-wide DW feeds departmental DBs	Data marts model a business process; enterprise is achieved with conformed dims
Complexity of method	Quite complex	Fairly simple

Reference: http://www.bi-bestpractices.com/view-articles/4768

The Comparison (Data Modeling)

	Inmon	Kimball
Data orientation	Subject or data driven	Process oriented
Tools	Traditional (ERDs and DIS)	Dimensional modeling; departs from traditional relational modeling
End user accessibility	Low	High

Reference: http://www.bi-bestpractices.com/view-articles/4768

The Comparison (Dimensions)

	Inmon	Kimball
Timeframe	Continuous & Discrete	Slowly Changing
Methods	Timestamps	Dimension keys

Reference: http://www.bi-bestpractices.com/view-articles/4768

Inmon Continuous & Discrete Dimension Management

- Define data management via dates in your data
  - Continuous time
    - When is a record active
    - Start and end dates
  - Discrete time
    - A point in time
    - Snapshot

Kimball Slowly Changing Dimension Management

- Define data management via versioning
  - Type I
    - Change record as required
    - No History
  - Type II
    - Manage all changes
    - History is recorded
  - Type III
    - Some history is parallel
    - Limit to defined history

The Comparison (Philosophy)

	Inmon	Kimball
Primary Audience	IT	End Users
Place in the Organization	Integral part of the Corporate Information Factory (CIF)	Transformer and retainer of operational data
Objective	Deliver a sound technical solution based on proven methods	Deliver a solution that makes it easy for end users to directly query data and still have reasonable response rate

Reference: http://www.bi-bestpractices.com/view-articles/4768

How to Choose?

Characteristic	Favors Kimball	Favors Inmon
Nature of the organization's decision support requirements	Tactical	Strategic
Data integration requirements	Individual business areas	Enterprise-wide integration
Structure of data	Business metrics, performance measures, and scorecards	Non-metric data and for data that will be applied to meet multiple and varied information needs
Scalability	Need to adapt to highly volatile needs within a limited scope	Growing scope and changing requirements are critical

How to Choose?

Characteristic	Favors Kimball	Favors Inmon
Persistency of data	Source systems are relatively stable	High rate of change from source systems
Staffing and skills requirements	Small teams of generalists	Larger team(s) of specialists
Time to delivery	Need for the first data warehouse application is urgent	Organization's requirements allow for longer start-up time
Cost to deploy	Lower start-up costs, with each subsequent project costing about the same	Higher start-up costs, with lower subsequent project development costs

References

- Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models: by Mary Breslin http://www.bi-bestpractices.com/view-articles/4768
- Inmon CIF glossary: http://www.inmoncif.com/library/glossary/#D
- The Data Warehouse Toolkit, Kimball, 2002
- Inmon, W.H. Building the Data Warehouse (Third Edition), New York: John Wiley & Sons, (2002).
- Kimball, R. and M. Ross. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition), New York: John Wiley & Sons, 2000.

Thanks and Questions?

- Ian Abramson

  IAS Inc

  416-407-2448

  ian@abramson.ca