



Whitemarsh
Information Systems Corporation

Entity Relationship
versus
Star Schema Modeling

December 15, 2011

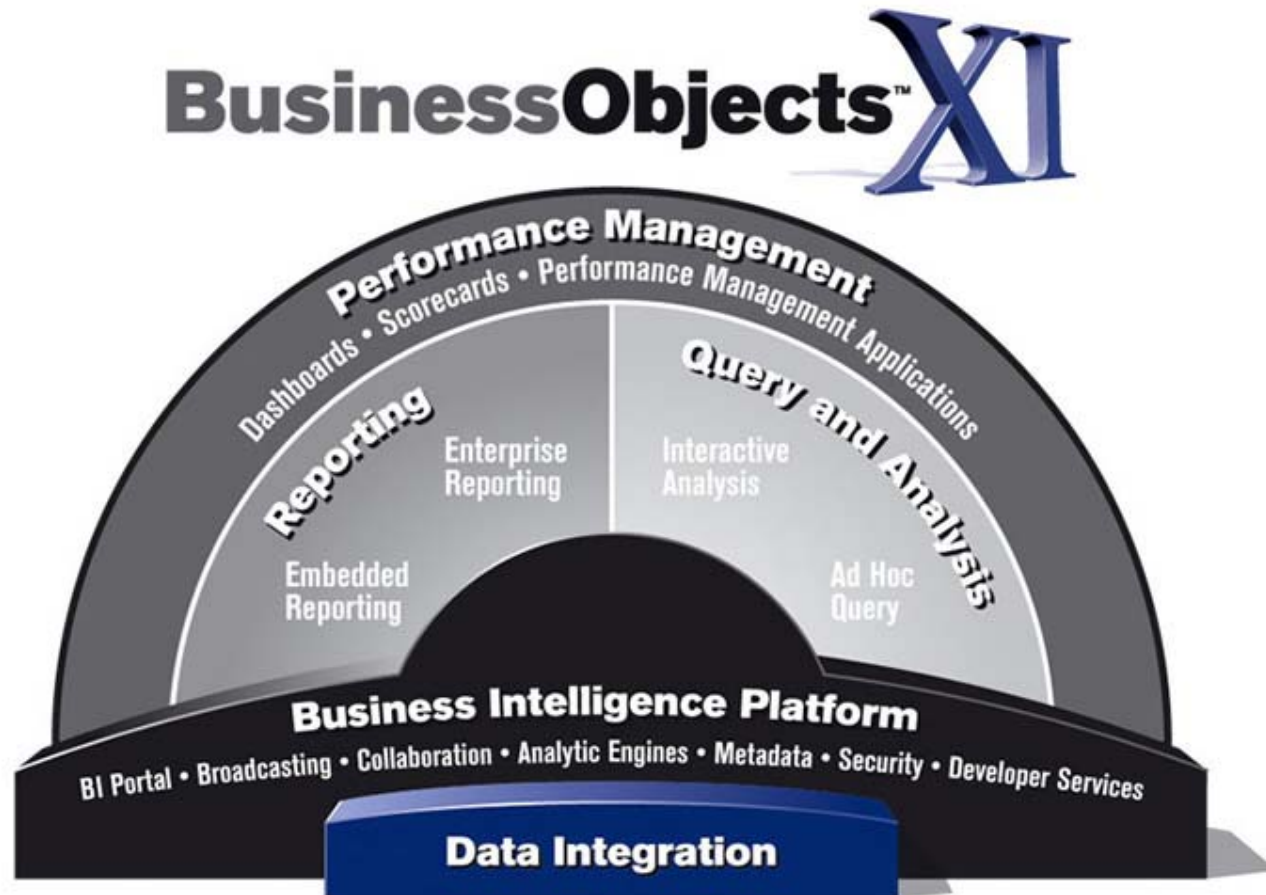
Whitemarsh Information Systems Corporation
2008 Althea Lane
Bowie, Maryland 20716
Tele: 301-249-1142
Email: Whitemarsh@wiscorp.com
Web: www.wiscorp.com

Topics

1. What is Business Intelligence ?	1
2. Fundamental Data Management Terms	3
3. Data is Executed Policy (core principles)	6
4. Two Data Warehouse Strategies	8
4.1 ER-Model Based (Inmon)	9
4.2 Star Schema Based (Kimball)	10
4.3 Key “Data Warehouse Architecture” Differences	12
4.4 Key Data Modeling Differences	15
4.5 Key Build Strategy Differences	18
5. Conclusions and Recommendations	22



1. What is Business Intelligence ?



2. Fundamental Data Management Terms

Term	Definition
Data	The persistent evidences of policy executions, that is, the execution of the procedures of the enterprise.
Database	A highly organized repository of enterprise-data that contains its own internal structure ensuring that data can be entered, updated, reported, and held secure.
Database Management	The strategies and procedures for managing enterprise data within databases in an orderly and well-reasoned manner.
Database Management System (DBMS)	An IT software system that accomplishes the definition, instantiation, evolution, maintenance, evolution, and reporting of databases. Example: Oracle, DB2.
Data Element	A Data Element is a context independent business fact semantic template that can be employed as the meaning and rules basis for an attribute, column, field, screen element, etc.



Term	Definition
Data Model	A model of the data that is to be contained in a database. A complete data model consists of: 1) data record structure (a.k.a., tables), 2) relationships among the tables, and 3) formally defined operations on the rows of data and the relationships.
Concepts Data Model	A type of data model that represents the data models (see above) of individual concepts. E.g., Person, Address, Invoice, and Customer. Commonly this type of data model is high-level and likely only contains entities, attributes, and relationships. Possible 100s of data models of concepts. Data structure pattern of a concept. “Ether-land”
Logical Data Model	A database data model that consists of a schema, tables, columns, and relationships. These are typically constructed through the employment of data models of concepts within the database’s data model. These models are DBMS independent. Commonly these models are in third normal form.
Physical Data Model	A database models tuned to the needs of a specific application and DBMS. Commonly these models are derived from one or more logical data models and may not be in third normal form.



Term	Definition
Metadata	An imprecise and vague term to mean the specifications of real objects. Example: if a database's data is "real," the database's schema is metadata. Simply put, metadata is the abstraction-level about the considered object. Book example: while the chapters are "real," the table of contents and index are metadata.
Metadata Management System	A software application, commonly with a database that is managed by a DBMS that stores, manages, and reports metadata.
Business Intelligence	The sets of data that a business declares is needed to understand its past, and undertake tactical and strategic operations about its current, and future activities.



3. Data is Executed Policy (core principles)

- The persistent evidences of policy executions, that is, the execution of the procedures of the enterprise.
- Processes are the procedures through which policy is executed and data is created.
- Policy and procedures run the business.
- Policy executions produce data.
- Data stored in databases are the persistent memory of the organization.



Data is Executed Policy (cont.)

- "Data" specifications are **policy definitions**.
- "Process" specifications are **procedure definitions**.
- All data (i.e., policy) specifications are metadata.
- All process (i.e., procedure) specifications are metadata.
- A metadata database, i.e. a metabase, is a database for all Policy and Procedure Specifications.
- Data Management manages both metadata and data.



4. Two Data Warehouse Strategies



Bill Inmon – ER Model -based

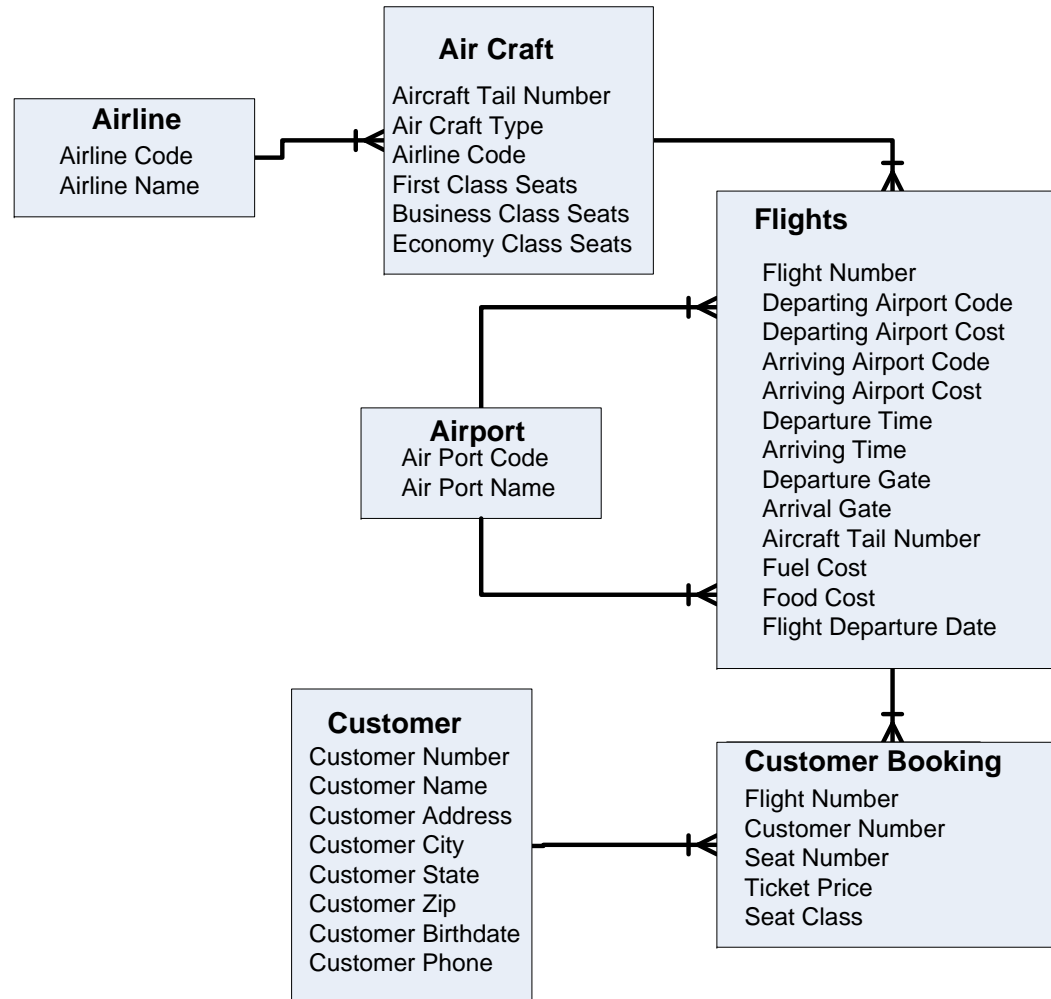


(c) Fenryll FA058

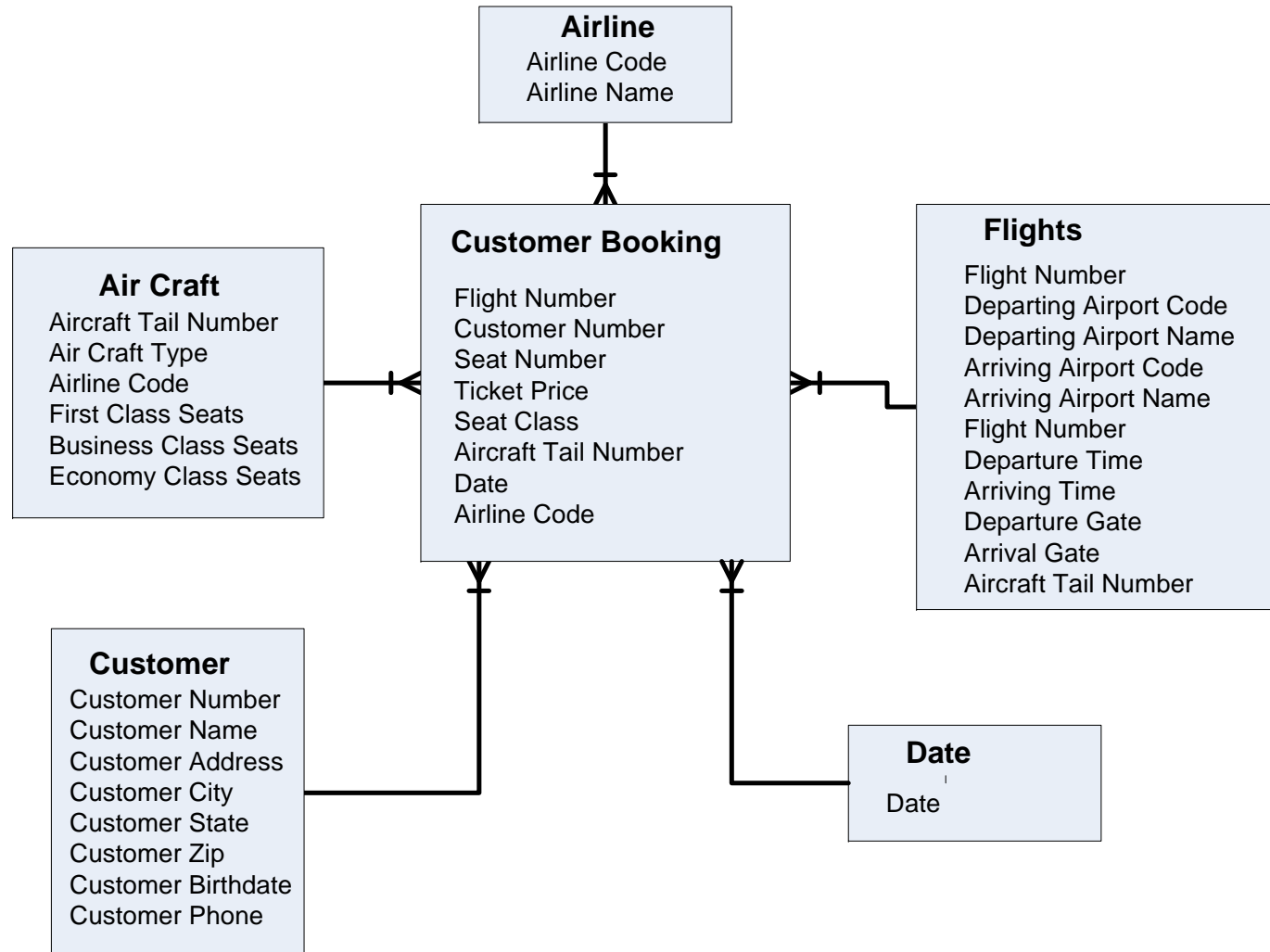
Ralph Kimball – Star Schema-based

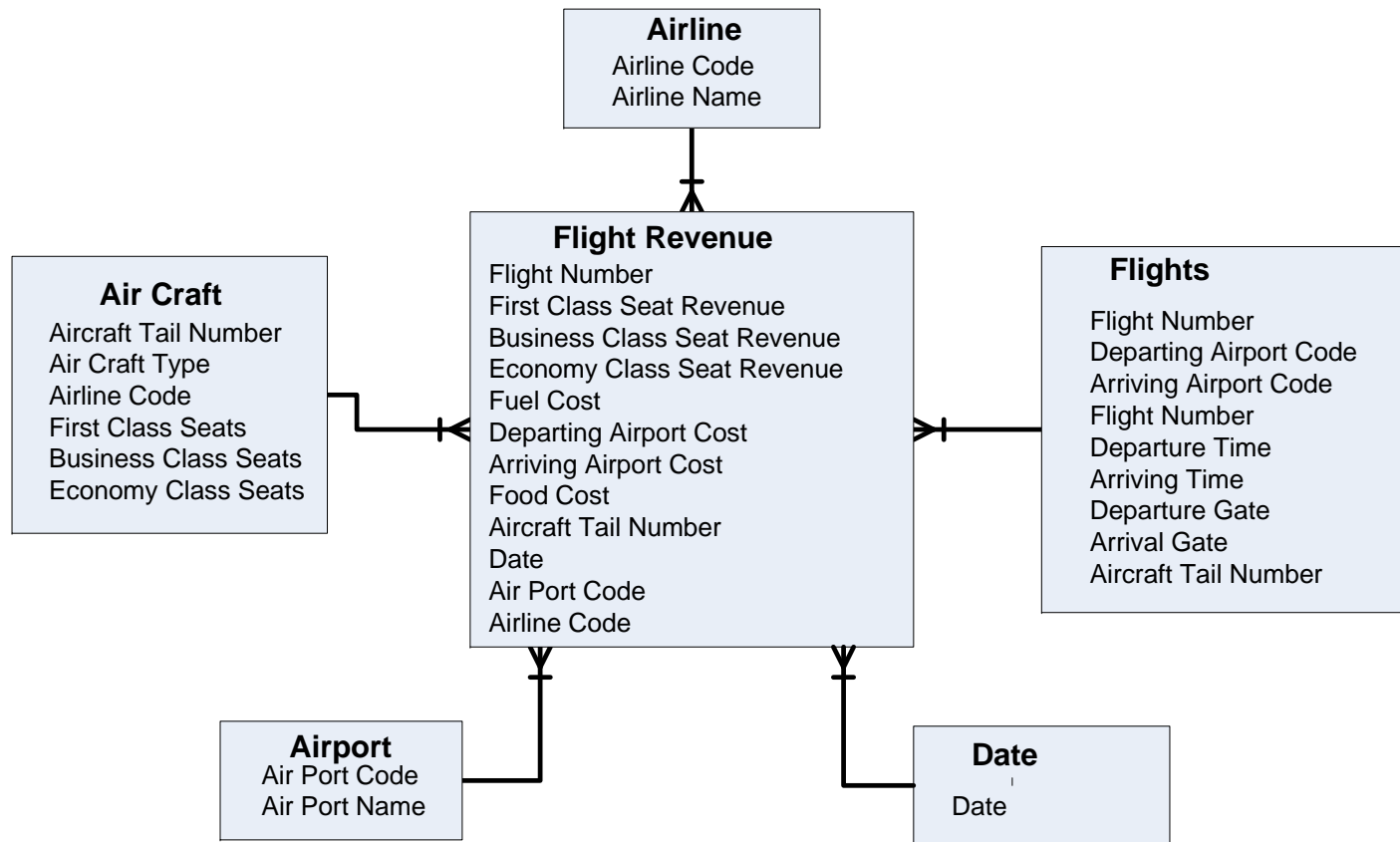


4.1 ER-Model Based (Inmon)



4.2 Star Schema Based (Kimball)





4.3 Key “Data Warehouse Architecture” Differences

1. **Kimball** approach has many Star Schemas.... Sort of a “have a need? Make a Start-Schema Data Mart!” Multiple Star Schemas share “dimensions.”
2. **Inmon** is a top-down approach. Data Marts “pop” out of the bottom of a cohesive Enterprise-wide (or at least subject-wide) database architecture.
3. **Kimball would say:** The Corporate Data Warehouse is the collection of all Data Marts in which each corresponds to the past and current information about a discrete decision domain.
4. **Inmon would say:** The Corporation Data Warehouse is a unified, integrated, and non-redundant enterprise-level operational, tactical, and strategic database. A Data Mart is a materialized discrete data-based view of a small decision domain.



Quotes from Kimball and Inmon

“... The data warehouse is nothing more than the union of all the data marts ...”

Ralph Kimball Dec. 29, 1997.

“You can catch all the minnows in the ocean and stack them together and they still do not make a whale.” **Bill Inmon** Jan. 8, 1998.



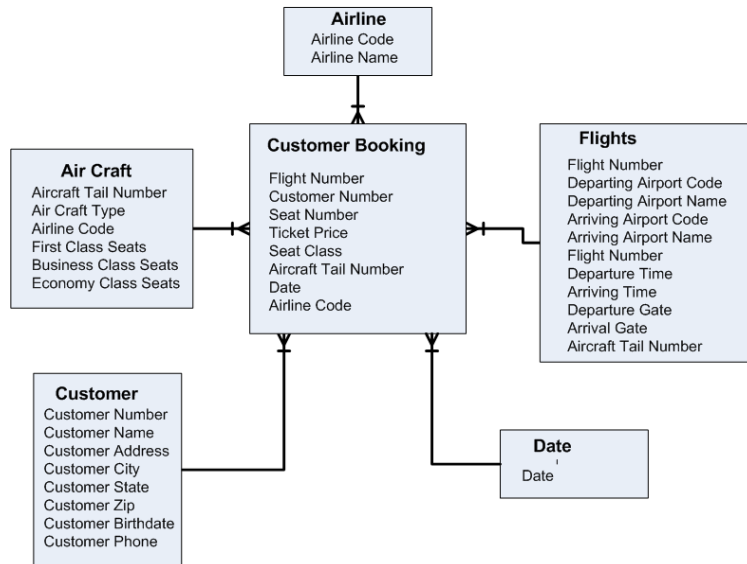
In short –

Kimball--First Data Marts--The “Sum” of which *is* the Enterprise Data Warehouse.

Inmon---First the Enterprise Data Warehouse--Later the Data Marts.



4.4 Key Data Modeling Differences



1. Each dimension is a “flat” table of collapsed hierarchies. Hence very un-normalized.

2. Each dimension with history has a “Current” flag for the most recent row.

3. Time-sequence is managed by a special Time-Dimension

4. Fact tables record state changes or fact measurements. Facts are selectable or

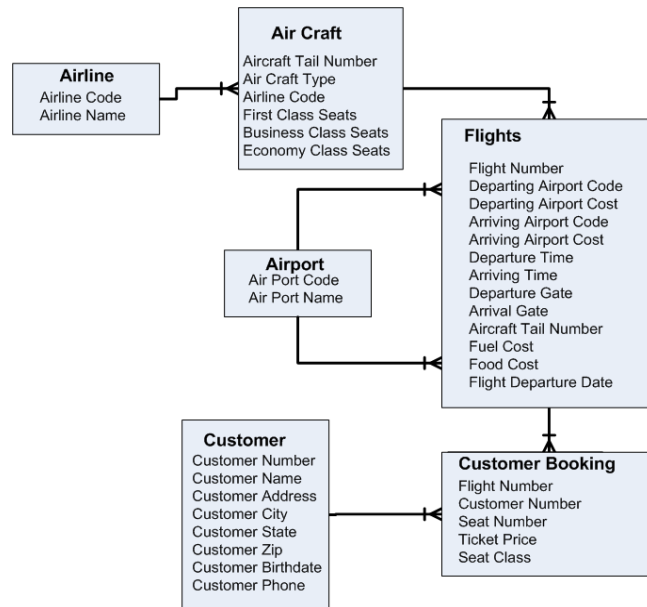
distinguishable by related dimension information.

5. Granularity, precision, and timeliness is set to the need of the decision domain of the specific star schema.

6. Dimensions can be shared between and among fact tables thus giving the ability to “sort of” report different measurements based on common “dimensions.”



ER-Schema Data Models



1. Every table is normalized to the maximum degree possible.

2. No special treatment of history other than what is “naturally” engineered into database tables.

3. Time- sequence is represented by time-stamped columns

3. No “manufactured” dimensions. No “manufactured” fact tables.

4. Granularity, precision, and timeliness is set to that of the enterprise, not of the data warehouse tables.

5. Star Schemas Data Marts are built from corporate data warehouse.



Key Data Modeling Differences Summary

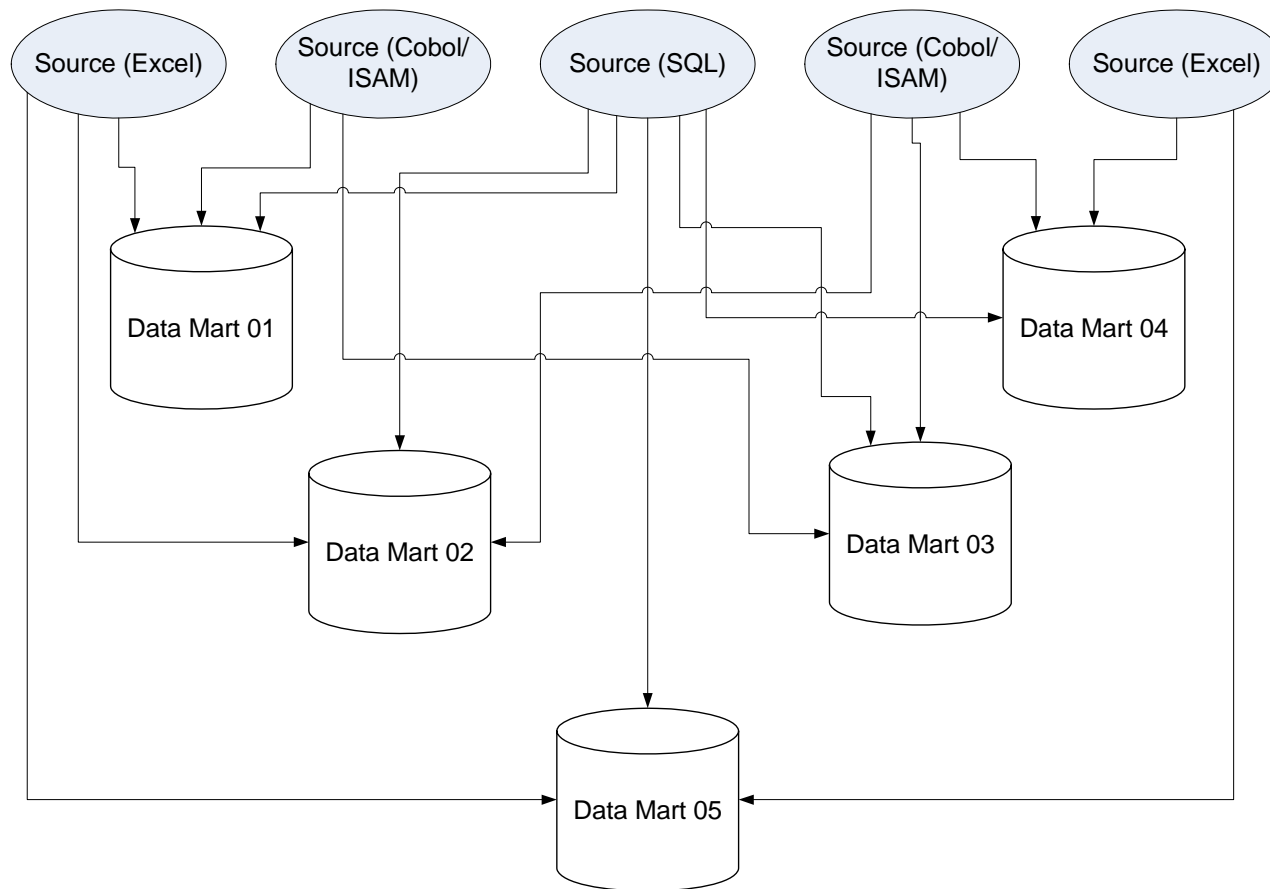
1. Inmon's approach is NOT for directly building data marts. Rather it's to build Enterprise Data Warehouses from which data marts are generated.
2. Kimball's approach is to build collections of Star Schema data marts with shared dimensions. The Kimball EDW is THIS collection.
3. So really, arguing for a Kimball or Inmon approach is almost like arguing which is better, a car's engine or its transmission. An argument based on a false premise.

But if you are going to build Data Marts, the Build Strategy Differences are worth noting.

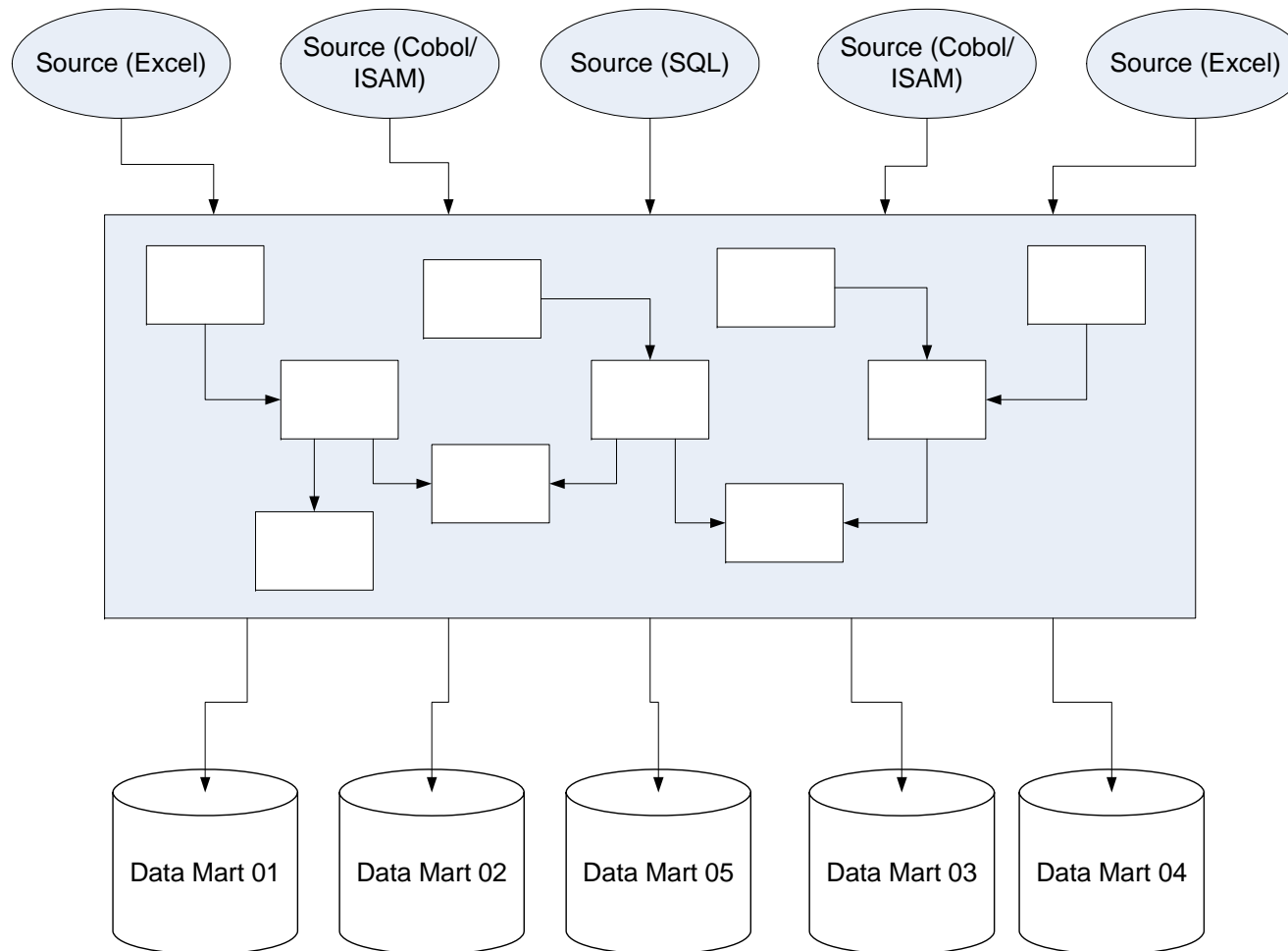


4.5 Key Build Strategy Differences

Star Schema Extract Transform Load (ETL) Strategy (worst case?)



ER-Model Extract Transform Load (ETL) Strategy



Critical Statistics	
Kimball	Inmon
15 sets of business rules for granularity, precision, and timeliness.	10 sets of business rules for granularity, precision, and timeliness.
15 sets of value domain engineering.	5 sets of value domain engineering.
15 sets of ETL software processes.	10 sets of ETL software processes.
Hi probability of conflicting Business rules for the 15 ETL Scripts.	Low probability of conflicting Business rules for the 10 ETL Scripts.

Counts drawn from strategies contained on Slides 25 and 26.



Issue	Critical Comparison	
	Kimball	Inmon
Consistent Semantics	Risk if each ETL and DM has its own.	Resolved and able to be controlled.
Synchronized Granularity and Precision	Risk if each ETL and DM has its own.	Resolved and able to be controlled.
Agreed-upon Timeliness	Risk if each ETL and DM has its own.	Resolved and able to be controlled.
Quality engineered two-way relationships versus collapsed hierarchies	Risk because there's not highly engineered pair-related sets to draw from.	Resolved and able to be controlled.

Comparisons drawn from strategies contained on Slides 25 and 26.



5. Conclusions and Recommendations

- Kimball Data Marts have value and are needed within narrow, well-defined decision domains.
- Kimball Data Marts have an intuitive way of perceiving and processing data by end users.
- Kimball's Data Marts built directly from sources are accomplished faster (i.e., Source to Data Mart) than Source-to-Inmon-to-Kimball.
- Direct population of Kimball Data Marts from Data Sources creates risks in the areas of semantics, granularity, precision, timeliness, and uncontrolled value domains.
- Proliferation of Data Marts can cause more data and semantics stove-pipes if there's no enterprise-wide data-centric engineering and architecture.



Conclusions and Recommendations (cont.)

- While a Inmon-Kimball strategy will take longer than one-off Kimball data marts, it has positive effects on synchronization of semantics, granularity, precision, timeliness, and controlled value domains across Kimball-based data marts.
- Direct querying an Inmon Data warehouse requires more specialized and advanced database knowledge than from Kimball Data Marts.
- Inmon data warehouse design is optimized for data loading and update.
- An Inmon-Kimball architecture has greater stability, flexibility, and evolution capability over either architecture individually.



Bottom Line Recommendation

- Build Inmon data warehouses at the Subject Area database scope.
- Build Kimball Star-Schema data marts from Inmon data warehouses.
- Inmon-Kimball is preferred over either Inmon or Kimball alone.

