SS G515 - Data Warehousing

Dr. Yashvardhan Sharma
Assistant Professor, CS & IS Dept.
BITS-Pilani

Data Marts

- What is a data mart?
- Advantages and disadvantages of data marts
- Issues with the development and management of data marts

Data Marts

- A subset of a data warehouse that supports the requirements of a particular department or business process
- Data Mart is a subset of corporate-wide data that is of value to a specific groups of users. Its scope is confined to specific, selected groups, such as marketing data mart.
- Characteristics include:
  - Does not always contain detailed data unlike data warehouses
  - More easily understood and navigated
  - Can be dependent or independent

Data Marts

- Data Mart: A scaled-down version of the data warehouse
- A data mart is a small warehouse designed for the department level.
- It is often a way to gain entry and provide an opportunity to learn
- Major problem: if they differ from department to department, they can be difficult to integrate enterprise-wide

Reasons for Creating Data Marts

- Proof of Concept for the DW
- Can be developed quickly and less resource intensive than DW
- To give users access to data they need to analyze most often
- To improve query response time due to reduction in the volume of data to be accessed

Kimball vs Inmon

- *Bill Inmon's paradigm:* Data warehouse is one part of the overall business intelligence system. An enterprise has one data warehouse, and data marts source their information from the data warehouse. In the data warehouse, information is stored in 3rd normal form.

- *Ralph Kimball's paradigm:* Data warehouse is the conglomerate of all data marts within the enterprise. Information is always stored in the dimensional model.

Kimball vs Inmon

- *Bill Inmon:* Endorses a Top-Down design

  Independent data marts cannot comprise an effective EDW. Organizations must focus on building EDW

- *Ralph Kimball:* Endorses a Bottom-Up design

  EDW effectively grows up around many of the several independent data marts – such as for sales, inventory, or marketing

Kimball vs Inmon: War of Words

"...The data warehouse is nothing more than the union of all the data marts...,"

Ralph Kimball, December 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, January 8, 1998.

Kimball vs. Inmon

There is no right or wrong between these two ideas, as they represent different data warehousing philosophies. In reality, the data warehouse in most enterprises are closer to Ralph Kimball's idea. This is because most data warehouses started out as a departmental effort, and hence they originated as a data mart. Only when more data marts are built later do they evolve into a data warehouse.

Data Warehousing Process

- Enterprise-wide warehouse, top down, the Inmon methodology
- Data mart, bottom up, the Kimball methodology
- When properly executed, both result in an enterprise-wide data warehouse

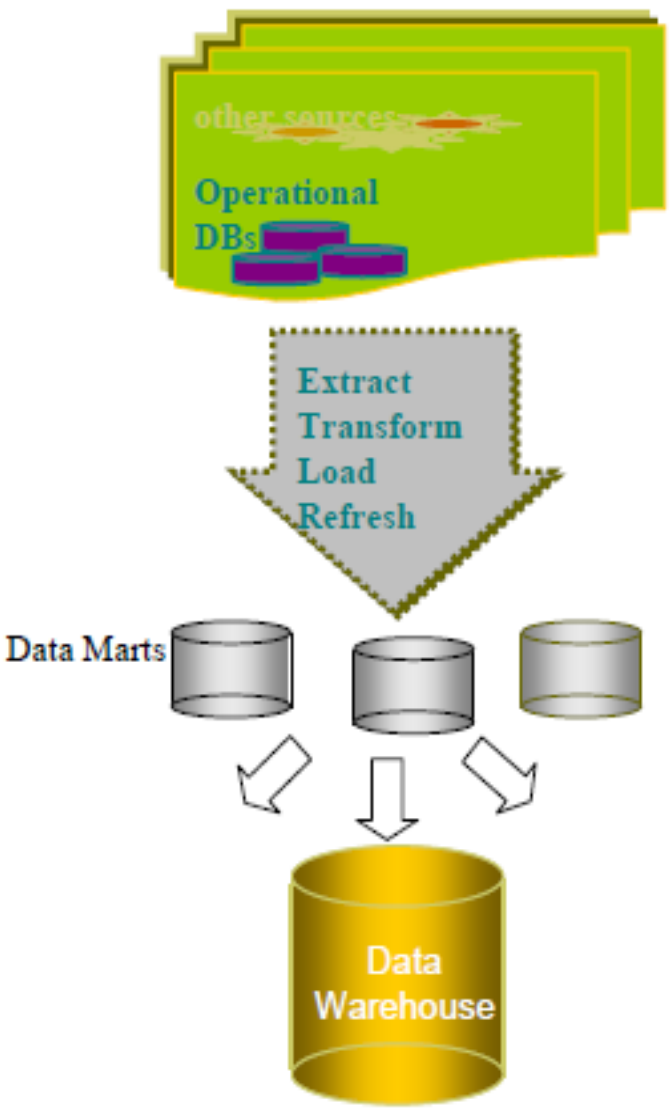
Data warehouse versus data mart.

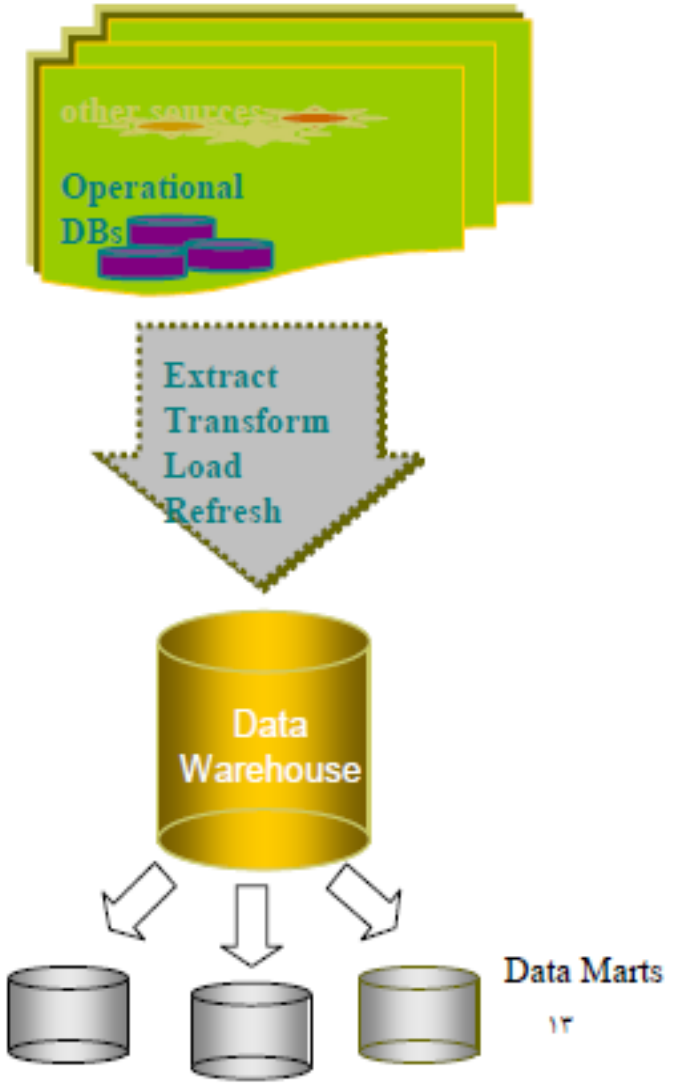
DATA WAREHOUSE	DATA MART
Corporate/Enterprise-wide	Departmental
Union of all data marts	A single business process
Data received from staging area	Star-join (facts & dimensions)
Queries on presentation resource	Technology optimal for data access and analysis
Structure for corporate view of data	Structure to suit the departmental view of data
Organized on E-R model	

Building a Data Mart

- Questions to be asked:
  - Top-down or bottom-up approach?
  - Enterprise-wide or departmental?
  - **Which first—data warehouse or data mart?**
  - Build pilot or go with a full-fledged implementation?
  - Dependent or independent data marts?

Bottom-Up Versus Top-Down Approach





Data Warehouse or Data Mart First?

- Top-Down vs. Bottom-Up Approach
- Advantages of Top-Down
  - A truly corporate effort, an enterprise view of data
  - Inherently architected-not a union of disparate DMs
  - Single, central storage of data about the content
  - Central rules and control
  - May be developed fast using iterative approach

Data Warehouse or Data Mart First?

- Disadvantages of Top-Down
  - Takes longer to build even with iterative method
  - High exposure/risk to failure
  - Needs high level of cross functional skills
  - High outlay without proof of concept
  - Difficult to sell this approach to senior management and sponsors

Data Warehouse or Data Mart First?

- Advantages of Bottom-Up Approach
  - Faster and easier implementation of manageable pieces
  - Favorable ROI and proof of concept
  - Less risk of failure
  - Inherently incremental; can schedule important DMs first
  - Allows project team to learn and grow

Data Warehouse or Data Mart First?

- Disadvantages of Bottom-Up Approach
  - Each DM has its own narrow view of data
  - Permeates redundant data in every DM
  - Difficult to integrate if the overall requirements are not considered in the beginning
- Kimball's approach is considered as a Bottom-Up approach, but he disagrees

Dependent Data Marts

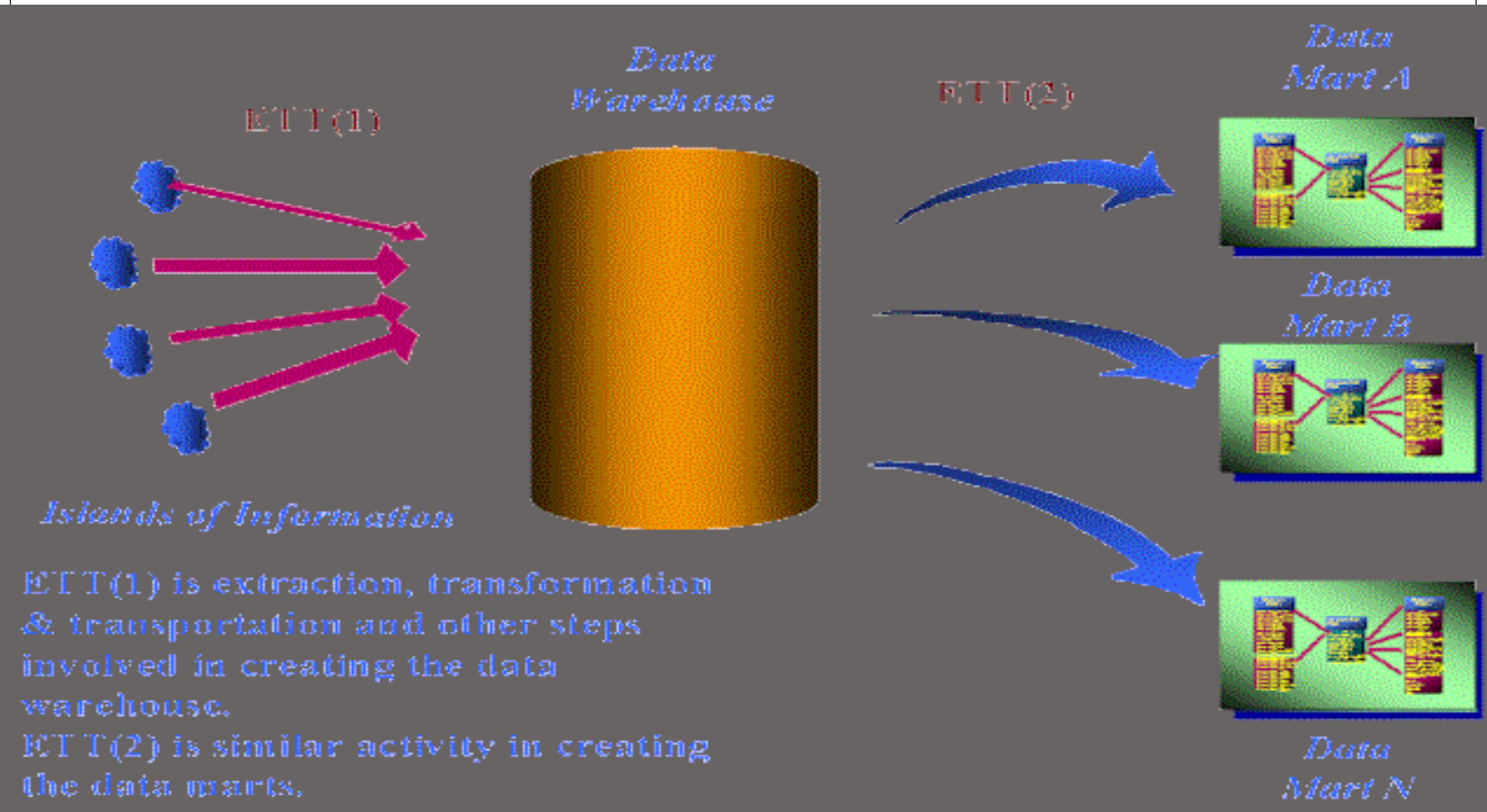


Figure One -- The Top Down Flow from Data Warehouses to Data Marts

Independent Data Marts

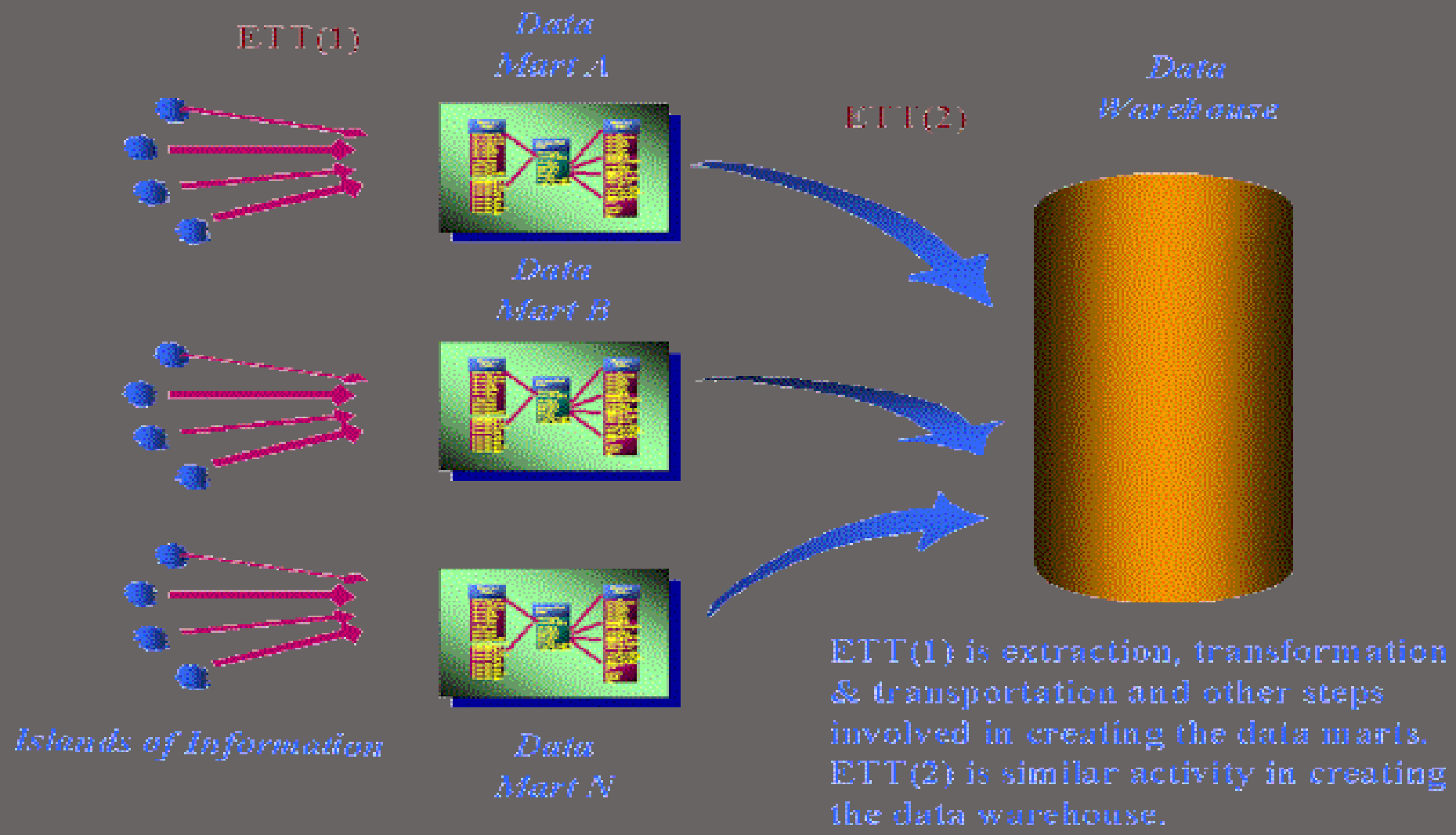


Figure Two -- The Bottom-up Flow from Data Marts to the Data Warehouse

The Bottom-Up Misnomer

Kimball encourages you to broaden your perspective both "vertically" and "horizontally" while gathering business requirements while developing data marts

The Bottom-Up Misnomer

- Vertical
  - Don't just rely on the business data analyst to determine requirements
  - Inputs from senior managers about their vision, objectives, and challenges are critical
  - Ignoring this vertical span might cause failure in understanding the organization's direction and likely future trends

The Bottom-Up Misnomer

- Horizontal
  - Look horizontally across the departments before designing the DW
  - Critical in establishing the enterprise view
  - Challenging to do if one particular department if funding the project
  - Ignoring horizontal span will create isolated, department-centric databases that are inconsistent and can't be integrated
  - Complete coverage in a large organization is difficult
  - One rep. from each dept. interacting with the core development team can be of immense help

Data Warehouse or Data Mart First?

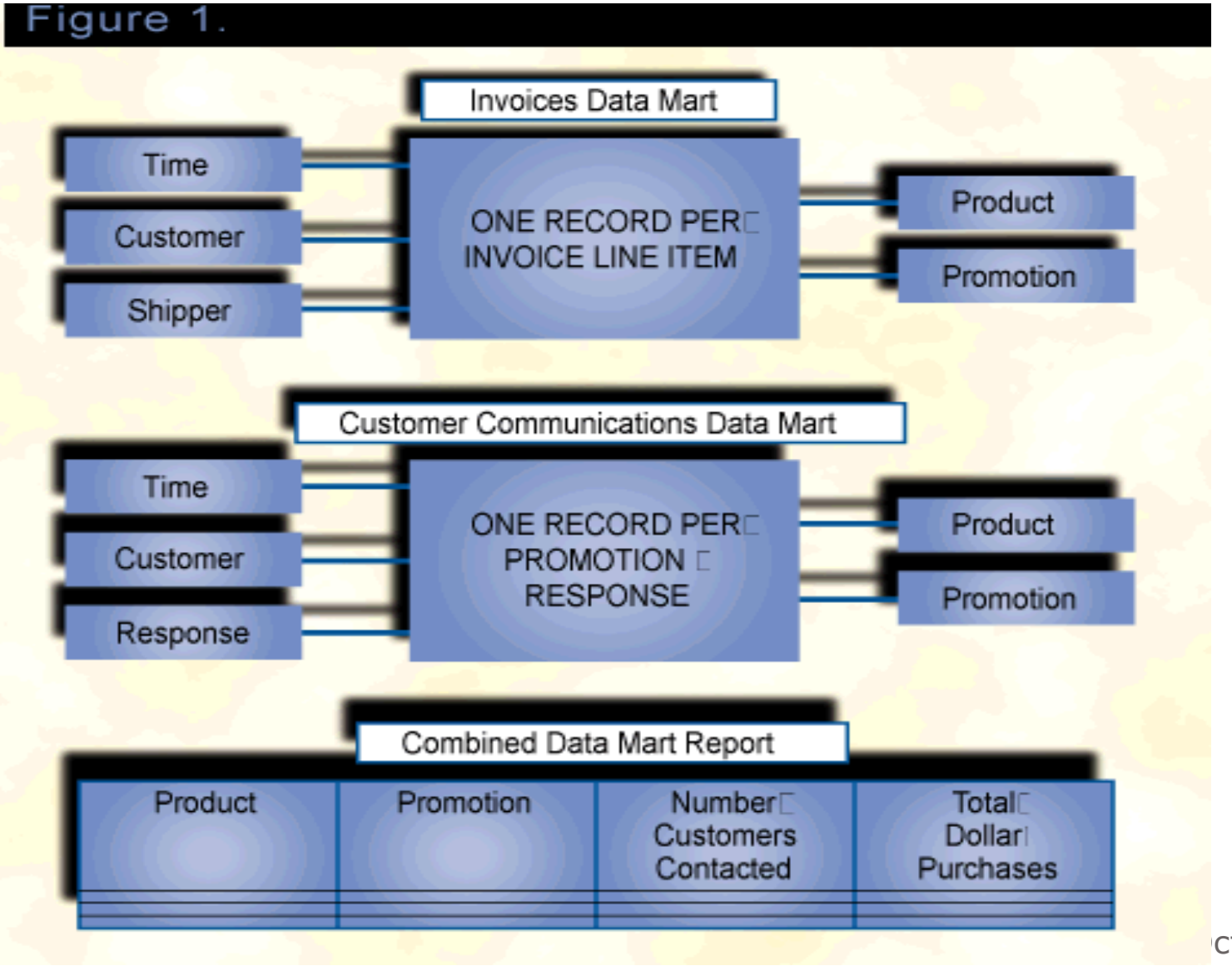
New Practical approach by Kimball

1. Plan and define requirements at the overall corporate level
2. Create a surrounding architecture for a complete warehouse
3. Conform and standardize the data content
4. Implement the Data Warehouse as a series of Supermarts, one at a time

A Word about SUPERMARTS

- Totally monolithic approach vs. totally stovepipe approach
- A step-by-step approach for building an EDW from granular data
- A Supermart s a data mart that has been carefully built with a disciplined architectural framework
- A Supermart is naturally a complete subset of the DW
- A Supermart is based on the most granular data that can possible be collected and stored
- Conformed dimensions and standardized fact definitions

A Word about SUPERMARTS



Pilot Projects: Risk vs. Reward

- Start with a pilot implementation as the first rollout for DW
- Pilot projects have advantage of being small and manageable
- Provide organization with a "proof of concept"

Pilot Projects: Risk vs. Reward

Functional scope of a pilot project should be determined based on:

1. The Degree of risk enterprise is willing to take
2. The potential for leveraging the pilot project
   - Avoid constructing a throwaway prototype
   - Pilot warehouse must have actual value to the enterprise

Pilot Projects: Risk vs. Reward

A Practical Approach

- Most people employ a ***Hybrid approach with*** elements of ***Top-Down and Bottom-Up***
- Again, practitioners don't always concentrate on these issues and use this terminology, and just focus on best-practice
- That would include;
  - Build incrementally according to a business function
  - Employ an enterprise perspective
  - Dimensionally model data
  - Utilize conformed dimensional models
  - Employ a Staging Area or Data Warehouse
  - Store atomic data