

# DEFINING the BUSINESS REQUIREMENTS

Dr. Shalini Gambhir

# Introduction

- OLTP and DW planning is different in term of requirements clarity
- Planning DW is about solving users' problems and providing strategic information to the user.
- OLTP systems are primarily data capture systems.
- Data warehouse systems are information delivery systems.
- OLTP systems are needed to run the day-to-day business and returns are seen immediately.
- No immediate payout is seen in a decision support system.

- OLTP (Online Transaction Processing) and DW (Data Warehousing) serve different purposes in terms of requirements and functionality.
- OLTP systems focus on capturing and processing day-to-day transactions for immediate business operations. For example, an online shopping website uses OLTP to process customer orders, update inventory, and handle transactions in real-time. The requirements for OLTP systems are often centered around ensuring quick and accurate data transactions to support operational tasks.
- On the other hand, DW systems are designed for strategic decision-making and information delivery. Planning for a Data Warehouse involves understanding users' analytical needs and providing a platform to extract insights from historical and consolidated data. For instance, a company may use a data warehouse to analyze sales trends over several years. The requirements for DW systems are centered around providing a comprehensive and integrated view of data for decision support.

- In an OLTP scenario, consider an airline reservation system. The OLTP system handles real-time transactions like booking seats, updating passenger details, and processing payments. The immediate goal is to ensure a smooth and efficient process for customers during ticket reservations, reflecting the day-to-day operational needs of the airline.
- Now, for a Data Warehouse in the same context, the focus shifts to strategic decision-making. The airline might use a data warehouse to analyze historical booking patterns, customer preferences, and revenue trends. Planning for the Data Warehouse involves understanding these analytical needs, providing a platform to extract insights, and delivering reports that assist in long-term planning, marketing strategies, and optimizing routes.
- In essence, while OLTP is about quick and precise data capture for immediate business operations, Data Warehousing is about aggregating and analyzing data over time to provide valuable insights for strategic decision support. The planning considerations for each align with these distinct objectives and timelines.

# Dimensional Analysis : Usage of Information is Unpredictable

- The users are familiar with operational systems because they use these in their daily work
  - Easy to visualize the requirements for new operational systems
  - A data warehouse system cannot be related to anything used before.
  - Process of defining requirements for a data warehouse is vague.

- Dimensional Analysis in the context of data warehousing often involves dealing with unpredictable information usage. Here's an explanation with an example:
- Consider a retail business implementing a data warehouse. In this scenario, users, such as sales managers or marketing analysts, may find it challenging to precisely define their requirements for analyzing data. They might be unfamiliar with the capabilities of a data warehouse and the types of insights it can provide. Additionally, for many users, it could be their first experience with a data warehouse.
- Given the unpredictable nature of information usage, building a data warehouse requires a different approach to requirements gathering. Instead of relying on users to provide specific and clear requirements upfront, dimensional analysis involves working closely with stakeholders to understand the nature of their business, potential analytical needs, and the dimensions along which they want to analyze data (e.g., time, geography, product categories).
- The iterative and collaborative nature of dimensional analysis allows for the discovery and refinement of requirements over time. Through prototypes and feedback loops, users gradually gain a better understanding of what insights they can derive from the data warehouse. This approach accommodates the fact that users might not have a precise vision initially, and the data warehouse evolves based on continuous interaction and exploration of data possibilities.

# Dimensional Analysis : Usage of Information is Unpredictable

- The users are generally unable to define their requirements clearly.
- For most of the users, this could be the very first data warehouse.
- How can you build something the users are unable to define clearly and precisely?
- Need different approach of requirements gathering.

# Dimensional Analysis : Usage of Information is Unpredictable

- What to do?
- Include every piece of data you think users will be able to use in the data warehouse?
- How can you build something the users are unable to define clearly and precisely?
  1. Collect data on the overall business of the organization.
  2. Check on the industry's best practices.
  3. Gather some business rules guiding the day-to-day decision making.
- But these are generalities and are not sufficient to determine detailed requirements
  - managers think of the business in terms of business dimensions.



- When faced with the challenge of unpredictable information usage in Dimensional Analysis, including every piece of data upfront may not be practical. Instead, a strategic approach is needed. Here's how you can tackle it:
- **Collect data on the overall business of the organization:**
  - Example: In a healthcare organization, gather data on patient demographics (age, geographic location, race), medical services provided, and operational aspects such as resource utilization. This broad dataset forms the foundation for understanding the organization's overall business.
- **Check on the industry's best practices:**
  - Example: Explore how successful healthcare organizations utilize data. Identify common metrics (mortality rate) and key performance indicators (KPIs) (eg: Readmission rate) in the industry, such as patient outcomes, resource efficiency (eg: LOS, Bed Turnover Rate), and compliance measures. This provides a benchmark for potential data points.
- **Gather some business rules guiding day-to-day decision making:**
  - Example: Understand the rules governing patient admission (Emergency Room Triage Protocol), discharge processes (a comprehensive assessment to determine their post-discharge care needs, including medication reconciliation, home health services, and follow-up appointments), and billing. These rules influence daily decisions and can contribute to defining data dimensions and measures that are crucial for decision support.

- While these steps provide generalities, the key lies in engaging with managers and stakeholders to understand the business in terms of specific business dimensions.
- For instance, in healthcare, managers may think in terms of dimensions like patient demographics, medical specialties, and geographical locations. By collaboratively identifying and refining these dimensions through ongoing discussions and prototypes, you can tailor the data warehouse to meet the evolving and often unpredictable needs of the users.
- This iterative approach allows for flexibility and ensures the data warehouse aligns closely with the actual business requirements.

# Dimensional Analysis: Dimensional Nature of Business Data

- Users can tell you how they think about the business
  - what measurement units are important for them,
  - How they measure success
  - how they combine the various pieces of information for strategic decision making.
- The actual proposed usage of a data warehouse could be unclear, the business dimensions used by the managers for decision making are quite clear.
  - Managers think of business in terms of business dimensions

- Here's an explanation with an example:
- **Dimensional Nature of Business Data:**
- **Users can tell you how they think about the business:**
  - *Example:* In a retail business, managers may think in terms of dimensions such as product categories, geographical regions, and time periods. They could express that the performance of the business is measured by sales revenue, customer satisfaction, and inventory turnover.
- **What measurement units are important for them:**
  - *Example:* In the retail context, the measurement units could include metrics like total sales, customer satisfaction scores, and inventory levels in units. These are the quantifiable aspects that managers consider crucial for assessing business performance.
- **How they measure success and combine information for strategic decision-making:**
  - *Example:* Success for a retail manager may be defined by the growth in sales revenue, the ability to manage inventory efficiently, and customer loyalty. Combining information across dimensions like product categories and geographical regions helps in strategic decision-making, such as optimizing product placement or marketing strategies.
- While the proposed usage of a data warehouse may initially be unclear, managers' clarity about the business dimensions they use for decision-making provides a foundation. By capturing and modeling data in alignment with these dimensions, a data warehouse can effectively support the strategic goals of the business. In essence, managers naturally think of the business in terms of these dimensions, and Dimensional Analysis leverages this understanding to structure and organize data for meaningful insights.

- THINK OF BUSINESS DIMENSIONS WHILE COLLECTING REQUIREMENTS

- Business dimensions of marketing manager

- ✓ Product
- ✓ Product category
- ✓ Time(day,week,month)
- ✓ District
- ✓ Division

- Business dimensions of financial controller

- ✓ Budget line
- ✓ Time(month,quarter,year)
- ✓ District
- ✓ Division

- Not interested in a single number:
- Break the info down by
  - ✓ Month
  - ✓ Division
  - ✓ Customer demographics
  - ✓ Sales office
  - ✓ Product
- These are the business dimensions for vice president

# Dimensional Analysis: Business Dimensions

## Marketing Vice President

How much did my new product generate  
month by month, in the southern division, by customer demographic, by  
sales office, relative to the previous version, and compared to plan?

## Marketing Manager

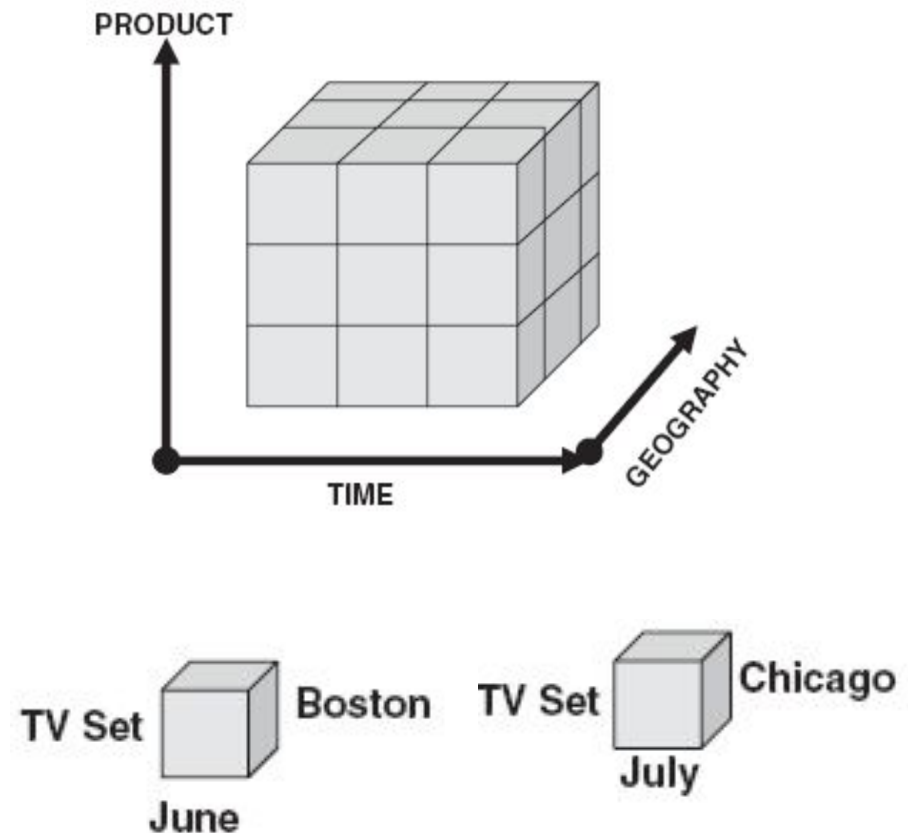
Give me sales statistics  
by products, summarized by product categories, daily, weekly, and  
monthly, by sale districts, by distribution channels.

## Financial Controller

Show me expenses  
listing actual vs budget, by months, quarters, and annual, by budget line  
items, by district, division, summarized for the whole company.

# Dimensional Analysis: Business Dimensions

- Get a good grasp of the dimensional nature of the business data
  - analysis of sales units along the three business dimensions of product, time, and geography.
  - Three dimensions are plotted against three axes of coordinates.
  - three dimensions form a collection of cubes.
  - In each of the small dimensional cubes, you will find the sales units for that particular slice of time, product, and geographical division.
- If there are more than three dimensions visualize multidimensional cubes, also called hypercubes.

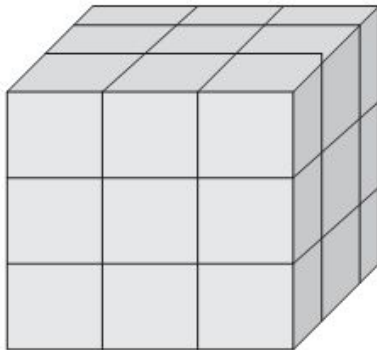


**Slices of product sales information (Units sold)**

# Dimensional Analysis

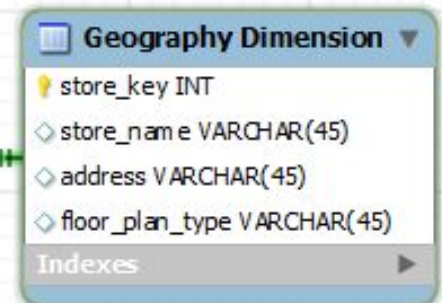
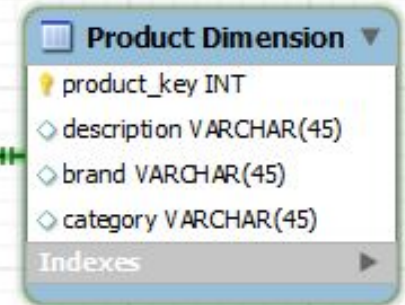
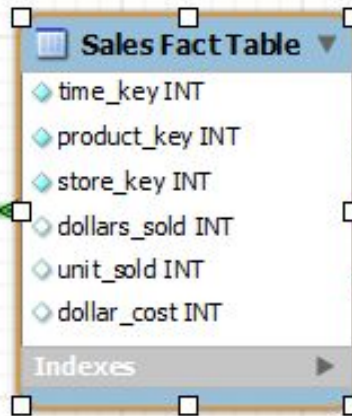
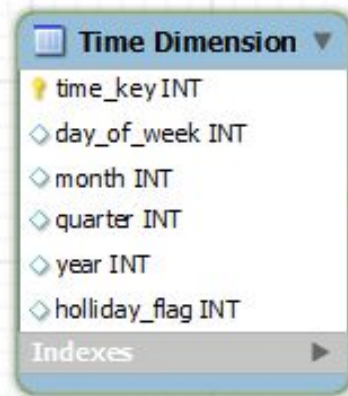
Possible to  
analyze at  
different  
time  
hierarchies

PRODUCT



TIME

GEOGRAPHY





- Here's an explanation with an example:
- **Business Dimensions and Multidimensional Cubes:**
- **Get a good grasp of the dimensional nature of the business data:**
  - *Example:* Imagine a retail business analyzing sales units. The three critical business dimensions are product, time, and geography. These dimensions provide a comprehensive view of sales data.
- **Analysis of sales units along the three business dimensions:**
  - *Example:* Sales units can be analyzed by looking at how they vary across different products, over time periods, and in various geographical regions.
- **Three dimensions are plotted against three axes of coordinates:**
  - *Example:* Picture a 3D graph where the x-axis represents time, the y-axis represents product categories, and the z-axis represents geographical regions. Each point in this space represents a unique combination of time, product, and geography.
- **Three dimensions form a collection of cubes:**
  - *Example:* These cubes represent specific combinations of time, product, and geography. Each cube contains the sales units for that particular slice of the business.

- **In each small dimensional cube, you find sales units for that particular slice:**
  - *Example:* If you focus on a specific cube in the space, it might represent sales of a particular product in a specific region during a certain time period. The value inside the cube would be the sales units for that combination.
- **Visualize multidimensional cubes (hypercubes) for more than three dimensions:**
  - *Example:* If additional dimensions like customer segments or marketing channels are considered, you move into the realm of hypercubes. This involves visualizing a space with more than three axes to accommodate the complexity of multidimensional data.
- In essence, dimensional analysis visualizes business data in a way that makes it easier for stakeholders to understand and derive insights. Each dimension provides a unique perspective, and the combination of these dimensions in cubes allows for a comprehensive exploration of the data landscape.

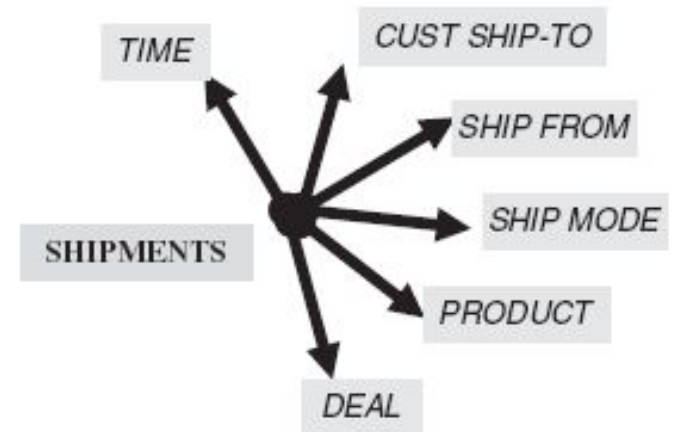
# More Complex Dimensional Model

- Business dimensions are relevant to
  - the business
  - to subject of analysis
- Almost all business analyses are performed over time
  - Time dimension is common

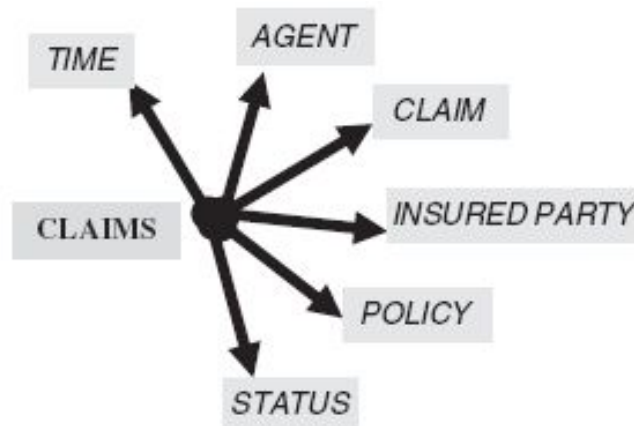
Supermarket Chain



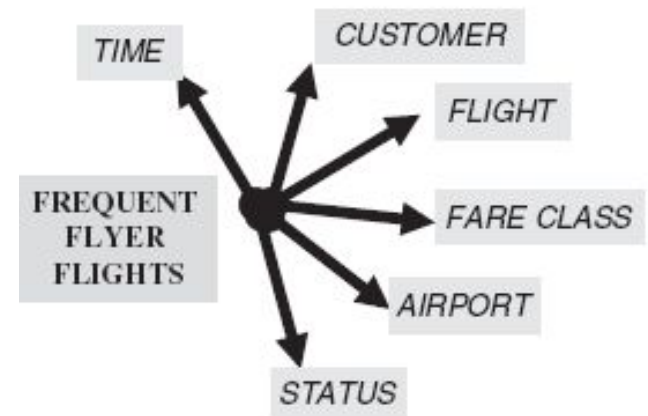
Manufacturing Company



Insurance Business



Airlines Company



# Information Packages

- A new idea for determining and recording information requirements for a data warehouse.
- This concept helps us to give a concrete form to
  - various insights,
  - unclear thoughts,
  - opinions expressed during the process of collecting requirements.
- Useful for taking data warehouse development to next phases

- Example scenario to illustrate how an information package could be used in the context of a data warehouse project.
- **Example Scenario: Building a Retail Data Warehouse**
- Imagine a retail corporation planning to develop a data warehouse to centralize and analyze data from its various sales channels, inventory systems, customer interactions, and marketing campaigns. The corporation aims to leverage this data to improve decision-making, optimize operations, and enhance customer experiences.

- **Information Package Components:**
- **Business Requirements:**
  - **Objective:** Improve sales performance and customer satisfaction through data-driven insights.
  - **Goals:** Increase revenue, reduce inventory costs, enhance customer segmentation, and personalize marketing efforts.
  - **KPIs:** Sales revenue, profit margins, inventory turnover rate, customer retention rate, and customer lifetime value (CLV).
- **Data Sources:**
  - **Point-of-Sale (POS) Systems:** Transactional data capturing sales transactions, including product details, prices, and quantities sold.
  - **Inventory Management Systems:** Data on stock levels, replenishment orders, and warehouse inventory.
  - **Customer Relationship Management (CRM) Systems:** Customer data, including demographics, purchase history, and interactions.
  - **Marketing Platforms:** Data on marketing campaigns, promotions, and customer responses.

- **Data Transformations:**

- Data cleansing: Remove duplicates, correct errors, and standardize formats to ensure data consistency.
- Data integration: Combine data from disparate sources using ETL processes to create a unified view.
- Data enrichment: Enhance customer data with external sources (e.g., demographic data from third-party providers) to enrich customer profiles.

- **Data Quality Requirements:**

- Accuracy: Ensure that data is reliable and free from errors to support accurate analysis and decision-making.
- Completeness: Ensure that all relevant data fields are populated and available for analysis.
- Consistency: Maintain consistency across data sources and ensure alignment with business rules and definitions.
- Timeliness: Ensure that data is updated in a timely manner to support real-time or near-real-time analytics.

- **Reporting and Analysis Needs:**
- Operational Reports: Daily sales reports, inventory status reports, and order fulfillment dashboards for operational insights.
- Strategic Reports: Quarterly sales performance analysis, customer segmentation analysis, and marketing campaign ROI reports for strategic decision-making.
- Ad Hoc Analysis: Ability to perform ad hoc queries and exploratory analysis to uncover actionable insights based on evolving business needs.



# Information Packages : Requirements not fully Determinate

- Why Information Packages?:
  - when requirements cannot be fully determined, we need a new and innovative concept to gather and record the requirements.
- The new methodology for determining requirements for a data warehouse system is based on business dimensions.
  - the need of the users to base their analysis on business dimensions.
  - incorporates the basic measurements and the business dimensions along which the users analyze these basic measurements.
- Discover the measurements and the relevant dimensions that must be captured and kept in the data warehouse.
  - known as an information package for the specific subject.

# Information Packages : Structure

- Subject : Sales
- Measurements shown at the bottom
- Business dimensions along which measurements will be taken are shown at the top as column headings
- Rows represent the hierarchy
  - Year down to individual day

Information Subject: Sales Analysis

Dimensions					
Time Periods	Locations	Products	Age Groups		
Year	Country	Class	Group 1		
Measured Facts: Forecast Sales, Budget Sales, Actual Sales					

Hierarchies

- ✓ *primary goal of requirements definition phase :*
  - ✓ *compile information packages for all the subjects for the data warehouse.*
- ✓ *After the information packages are defined clearly, proceed to the other phases.*

# Benefits of Information Packages

Information Packages enable designers to

- Define the common subject areas
- Design key business metrics
- Decide how data must be presented
- Determine how users will aggregate or roll up
- Decide the data quantity for user analysis or query
- Decide how data will be accessed
- Establish data granularity
- Estimate data warehouse size
- Determine the frequency for data refreshing
- Ascertain how information must be packaged

# Information Packages : Business Dimensions

- Business dimensions form the underlying basis of the new methodology for requirements definition.
- Data must be stored to provide for the business dimensions.
- The business dimensions and their hierarchical levels form the basis for all further phases.
  - Identify business dimensions and their hierarchical levels.
  - Choose the proper and optimal set of dimensions related to the measurements.

# Information Packages : Dimension Hierarchies or Categories

- The measurements along a business dimension are analyzed
  - First as summaries
  - Then at various levels of detail.
- Traverse the hierarchical levels of a business dimension for getting the details at various levels.
- The dimension hierarchies are the paths for drilling down or rolling up in analysis.
- Within each major business dimension there are categories of data elements that can be useful for analysis.
  - Example: holidays in the year dimension
  - Such data elements within the business dimension are called categories.

# Information Packages : Key Business Metrics or Facts

- The data warehouses users think of their business subjects in terms of business dimensions for obtaining information and for doing analysis.
  - What exactly are the users analyzing?
  - What numbers are they analyzing?
- The numbers the users analyze are the measurements or metrics that measure the success of their departments.
- These are the facts that indicate to the users how their departments are doing in fulfilling their departmental objectives.
- The metrics or facts go into the bottom section of the information package.
- The business dimensions will be the column headings.
  - In each column, you will include the hierarchies and categories for the business dimensions.

# Information Packages : Auto Sales Analysis

Hierarchies/Categories	Dimensions				
	Time	Product	Payment Method	Customer Demo-graphics	Dealer
	Year	Model Name	Finance Type	Age	Dealer Name
	Quarter	Model Year	Term (Months)	Gender	City
	Month	Package Styling	Interest Rate	Income Range	State
	Date	Product Line	Agent	Marital Status	Single Brand Flag
	Day of Week	Product Category		Household Size	Date First Operation
	Day of Month	Exterior Color		Vehicles Owned	
	Season	Interior Color		Home Value	
	Holiday Flag	First Year		Own or Rent	
<b>Facts:</b> Actual Sale Price, MSRP, Options Price, Full Price, Dealer Add-ons, Dealer Credits, Dealer Invoice, Down Payment, Proceeds, Finance					

# Information Packages : Hotel Occupancy

Hierarchies/Categories	Dimensions				
	Time	Hotel	Room Type		
	Year	Hotel Line	Room Type		
	Quarter	Branch Name	Room Size		
	Month	Branch Code	Number of Beds		
	Date	Region	Type of Bed		
	Day of Week	Address	Max. Occupants		
	Day of Month	City/State /Zip	Suite		
	Holiday Flag	Construction Year	Refrigerator		
		Renovation Year	Kichennette		
<b>Facts:</b> Occupied Rooms, Vacant Rooms, Unavailable Rooms, Number of Occupants, Revenue					