Information Package

Dr. Shalini Gambhir

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

- Information Packages – novel idea for determining and recording information requirements for a data warehouse.
- Determining requirements for a data warehouse is based on business dimensions
- The relevant dimension and measurements in that dimension are captured and kept in a data warehouse
- This creates an information package for a specific subject

- Information Packages are a concept used to determine and record the information requirements for a data warehouse. Essentially, they are a way to organize and structure the data stored in the warehouse based on business dimensions.
- For example, let's say a retail company wants to create an Information Package for their sales data. The relevant dimensions could include product, time, location, customer, and sales channel. Within each dimension, specific measurements or attributes are captured. For the product dimension, this could include attributes like product category, brand, and SKU. In the time dimension, it could include

- By organizing the data in this way, the data warehouse creates a cohesive Information Package for sales, allowing users to easily access and analyze sales data based on different dimensions and measurements. This structured approach helps in efficiently managing and querying the data warehouse to derive meaningful insights for decision-making.

- 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

- **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

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		

Hierarchies ↓

Measured Facts: Forecast Sales, Budget Sales, Actual Sales

✔*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

Information Packages : Auto Sales Analysis

Dimensions

Hierarchies/Categories

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		House-hold Size	Date First Operation	
Day of Month	Exterior Color		Vehicles Owned		
Season	Interior Color		Home Value		
Holiday Flag	First Year		Own or Rent		
Facts: Actual Sale Price, MSRP, Options Price, Full Price, Dealer Add-ons, Dealer Credits, Dealer Invoice, Down Payment, Proceeds, Finance					

Information Packages : Hotel Occupancy

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			
Facts: Occupied Rooms, Vacant Rooms, Unavailable Rooms, Number of Occupants, Revenue					

Hierarchies/Categories

Business dimensions for automobile manufacturer

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		House-hold Size	Date First Operation	
Day of Month	Exterior Color		Vehicles Owned		
Season	Interior Color		Home Value		
Holiday Flag	First Year		Own or Rent		
Facts: Actual Sale Price, MSRP Sale Price, Options Price, Full Price, Dealer Add-ons, Dealer Credits, Dealer Invoice, Down Payment, Proceeds, Finance					

Hierarchies / Categories

Hierarchies/Categories

- Hierarchies are paths for drilling down or rolling up in our analysis
- Non hierarchical data which are very important in analyzing are called categories
  - Holiday flag is a category which helps in evaluating the sales on a holiday

Methods

- **Requirement Gathering Methods**
- Requirement gathering is a crucial step in the development of data mining and warehousing systems. Here are some commonly used methods for gathering requirements in this context:
- Interviews: Conducting interviews with key stakeholders, including business users, data analysts, and IT personnel, to understand their needs and expectations. These interviews help identify the desired functionalities, data sources, data types, and analytical goals of the data mining and warehousing system.

- Surveys and Questionnaires: Distributing surveys or questionnaires to a wider audience to gather information about their requirements and preferences. This method allows for collecting input from a larger sample size and can help uncover a broader range of requirements and perspectives.
- Workshops and Focus Groups: Organizing collaborative workshops or focus groups with relevant stakeholders to discuss and brainstorm requirements. These sessions facilitate group discussions, idea sharing, and clarification of expectations, enabling a deeper understanding of the needs and challenges involved.

- Observation and Job Shadowing: Observing users and analysts in their work environment to gain insights into their daily tasks, processes, and pain points. This method provides a firsthand understanding of how data mining and warehousing tools are used, and helps identify specific requirements based on actual workflows.
- Document Analysis: Reviewing existing documentation, such as business reports, data models, system documentation, and user manuals. This analysis helps identify gaps, inconsistencies, and areas for improvement, providing valuable insights into the requirements for the new system.

- Prototyping and Proof of Concept: Building prototypes or proof-of-concept models to demonstrate potential functionalities and gather feedback from stakeholders. This iterative approach allows stakeholders to visualize and experience the system, providing valuable insights and requirements for further refinement.
- Benchmarking and Market Research: Conducting research on existing data mining and warehousing solutions in the market to identify best practices, industry standards, and emerging trends. This research helps gather requirements that align with the latest technologies and

- Collaboration and Feedback: Encouraging ongoing collaboration and feedback from stakeholders throughout the requirement gathering process. This can involve regular meetings, demos, and presentations to validate and refine the gathered requirements based on stakeholder input.
- Existing System Analysis: Analyzing the strengths and weaknesses of the existing data mining and warehousing systems, if any, to identify areas that need improvement or new requirements that arise from current limitations.

- Regulatory and Compliance Requirements: Identifying any specific regulatory or compliance requirements that impact data mining and warehousing, such as data privacy regulations (e.g., GDPR) or industry-specific guidelines. Ensuring compliance with these requirements is essential and should be factored into the overall system design.
- By employing a combination of these requirement gathering methods, organizations can effectively capture the necessary inputs and expectations to develop a successful data mining and warehousing system that meets their business needs.