D

Data Warehousing and Mining in Supply Chains

Richard Mathieu
Saint Louis University, USA

Reuven R. Levary
Saint Louis University, USA

INTRODUCTION

Every finished product has gone through a series of transformations. The process begins when manufacturers purchase the raw materials that will be transformed into the components of the product. The parts are then supplied to a manufacturer, who assembles them into the finished product and ships the completed item to the consumer. The transformation process includes numerous activities (Levary, 2000). Among them are

- Designing the product
- Designing the manufacturing process
- Determining which component parts should be produced in house and which should be purchased from suppliers
- Forecasting customer demand
- Contracting with external suppliers for raw materials or component parts
- Purchasing raw materials or component parts from suppliers
- Establishing distribution channels for raw materials and component parts from suppliers to manufacturer
- Establishing of distribution channels to the suppliers of raw materials and component parts
- Establishing distribution channels from the manufacturer to the wholesalers and from wholesalers to the final customers
- Manufacturing the component parts
- Transporting the component parts to the manufacturer of the final product
- Manufacturing and assembling the final product
- Transporting the final product to the wholesalers, retailers, and final customer

Each individual activity generates various data items that must be stored, analyzed, protected, and transmitted to various units along a supply chain.

A *supply chain* can be defined as a series of activities that are involved in the transformation of raw materials into a final product, which a customer then purchases (Levary, 2000). The flow of materials, component parts, and products is moving downstream (i.e., from the initial supply sources to the end customers). The flow of information regarding the demand for the product and orders to suppliers is moving upstream, while the flow of information regarding product availability, shipment schedules, and invoices is moving downstream. For each organization in the supply chain, its customer is the subsequent organization in the supply chain, and its subcontractor is the prior organization in the chain.

BACKGROUND

Supply chain data can be characterized as either transactional or analytical (Shapiro, 2001). All new data that are acquired, processed, and compiled into reports that are transmitted to various organizations along a supply chain are deemed transactional data (Davis & Spekman, 2004). Increasingly, transactional supply chain data is processed and stored in enterprise resource planning systems, and complementary data warehouses are developed to support decision-making processes (Chen R., Chen, C., & Chang, 2003; Zeng, Chiang, & Yen, 2003). Organizations such as Home Depot, Lowe's, and Volkswagen have developed data warehouses and integrated data-mining methods that complement their supply chain management operations (Dignan, 2003a; Dignan, 2003b; Hofmann, 2004). Data that are used in descriptive and optimization models are considered analytical data (Shapiro, 2001). Descriptive models include various forecasting models, which are used to forecast demands along supply chains, and managerial accounting models, which are used to manage activities and costs. Optimization models are used to plan resources, capacities, inventories, and product flows along supply chains.

Data collected from consumers are the core data that affect all other data items along supply chains. Information collected from the consumers at the point of sale include data items regarding the sold product (e.g., type, quantity, sale price, and time of sale) as well as information about the consumer (e.g., consumer address and method of payment). These data items are analyzed often. Data-mining techniques are employed to determine the types of items that are appealing to consumers. The items are classified according to consumers' socioeconomic backgrounds and interests, the sale price that consumers are willing to pay, and the location of the point of sale.

Data regarding the return of sold products are used to identify potential problems with the products and their uses. These data include information about product quality, consumer disappointment with the product, and legal consequences. Data-mining techniques can be used to identify patterns in returns so that retailers can better determine which type of product to order in the future and from which supplier it should be purchased. Retailers are also interested in collecting data regarding competitors' sales so that they can better promote their own product and establish a competitive advantage.

Data related to political and economic conditions in supplier countries are of interest to retailers. Data-mining techniques can be used to identify political and economic patterns in countries. Information can help retailers choose suppliers who are situated in countries where the flow of products and funds is expected to be stable for a reasonably long period of time.

Manufacturers collect data regarding a) particular products and their manufacturing process, b) suppliers, and c) the business environment. Data regarding the product and the manufacturing process include the characteristics of products and their component parts obtained from CAD/CAM systems, the quality of products and their components, and trends in theresearch and development (R & D) of relevant technologies. Data-mining techniques can be applied to identify patterns in the defects of products, their components, or the manufacturing process. Data regarding suppliers include availability of raw materials, labor costs, labor skills, technological capability, manufacturing capacity, and lead time of suppliers. Data related to qualified teleimmigrants (e.g., engineers and computer software developers) is valuable to many manufacturers. Data-mining techniques can be used to identify those teleimmigrants having unique knowledge and experience. Data regarding the business environment of manufacturers include information about competitors, potential legal consequences regarding a product or service, and both political and economic conditions in countries where the manufacturer has either facilities or business partners. Data-mining techniques can be used to identify possible liability concerning a product or service as well as trends in political and economic conditions in countries where the manufacturer has business interests.

Retailers, manufacturers, and suppliers are all interested in data regarding transportation companies. These data include transportation capacity, prices, lead time, and reliability for each mode of transportation.

MAIN THRUST

Data Aggregation in Supply Chains

Large amounts of data are being accumulated and stored by companies belonging to supply chains. Data aggregation can improve the effectiveness of using the data for operational, tactical, and strategic planning models. The concept of data aggregation in manufacturing firms is called *group technology (GT).* Nonmanufacturing firms are also aggregating data regarding products, suppliers, customers, and markets.

Group Technology

Group technology is a concept of grouping parts, resources, or data according to similar characteristics. By grouping parts according to similarities in geometry, design features, manufacturing features, materials used, and/or tooling requirements, manufacturing efficiency can be enhanced, and productivity increased. Manufacturing efficiency is enhanced by

- Performing similar activities at the same work center so that setup time can be reduced
- Avoiding duplication of effort both in the design and manufacture of parts
- Avoiding duplication of tools
- Automating information storage and retrieval (Levary, 1993)

Effective implementation of the GT concept necessitates the use of a classification and coding system. Such a system codes the various attributes that identify similarities among parts. Each part is assigned a number or alphanumeric code that uniquely identifies the part's attributes or characteristics. A part's code must include both design and manufacturing attributes.

A classification and coding system must provide an effective way of grouping parts into part families. All parts in a given part family are similar in some aspect of design or manufacture. A part may belong to more than one family.

A part code is typically composed of a large number of characters that allow for identification of all part attributes. The larger the number of attributes included in a part code, the more difficult the establishment of standard procedures for classifying and coding. Although numerous methods of classification and coding have been developed, none has emerged as the standard method. Because different manufacturers have different requirements regarding the type and composition of parts' codes,