Chapter 3

Warehousing RFID and Location-Based Sensor Data

Hector Gonzalez
University of Illinois at Urbana-Champaign, USA

Jiawei Han
University of Illinois at Urbana-Champaign, USA

Hong Cheng
University of Illinois at Urbana-Champaign, USA

Tianyi Wu
University of Illinois at Urbana-Champaign, USA

ABSTRACT

Massive Radio Frequency Identification (RFID) datasets are expected to become commonplace in supply-chain management systems. Warehousing and mining this data is an essential problem with great potential benefits for inventory management, object tracking, and product procurement processes. Since RFID tags can be used to identify each individual item, enormous amounts of location-tracking data are generated. Furthermore, RFID tags can record sensor information such as temperature or humidity. With such data, object movements can be modeled by movement graphs, where nodes correspond to locations, and edges record the history of item transitions between locations and sensor readings recorded during the transition. This chapter shows the benefits of the movement graph model in terms of compact representation, complete recording of spatio-temporal and item level information, and its role in facilitating multidimensional analysis. Compression power and efficiency in query processing are gained by organizing the model around the concept of gateway nodes, which serve as bridges connecting different regions of graph, and provide a natural partition of item trajectories. Multi-dimensional analysis is provided by a graph-based object movement data cube that is constructed by merging and collapsing nodes and edges according to an application-oriented topological structure.

DOI: 10.4018/978-1-60566-328-9.ch003
1 INTRODUCTION

The increasingly wide adoption of RFID technology by retailers to track containers, pallets, and even individual items as they move through the global supply chain, from factories in producer countries, through transportation ports, and finally to stores in consumer countries, creates enormous datasets containing rich multi-dimensional information on the movement patterns associated with objects along with massive amounts of important sensor information collected for each object. However, this information is usually hidden in terabytes of low-level RFID readings, making it difficult for data analysts to gain insight into the set of interesting patterns influencing the operation and efficiency of the procurement process. For example, we may discover a pattern that relates humidity and temperature during transportation to return rates for dairy products. In order to realize the full benefits of detailed object tracking and sensing information, we need a compact and efficient RFID cube model that provides OLAP-style operators useful to navigate through the movement data at different levels of abstraction of both spatio-temporal, sensor, and item information dimensions. This is a challenging problem that cannot be efficiently solved by traditional data cube operators, as RFID datasets require the aggregation of high-dimensional graphs representing object movements, not just that of entries in a flat fact table.

The problem of constructing a warehouse for RFID datasets has been studied in (Gonzalez, Han, Li, & Klabjan, 2006; Gonzalez, Han, & Li, 2006; Gonzalez, Han, Li, & Klabjan, 2006) introduced the concept of the RFID-cuboid, which compresses and summarizes an RFID dataset by recording information on items that stay together at a location with stay records, and linking such records through the use of a map table that connects groups of items that move and stay together through several locations. This view carries an implicit notion of the graph structure of RFID datasets but it fails to explicitly recognize the concept of movement graph as a natural model for item movements, and thus neglects the study of the topological characteristics of such a graph and its implications for query processing, cube computation, and data mining. In this chapter, we approach the RFID data warehouse from a movement graph-centric perspective, which makes the warehouse conceptually clear, better organized, and obtaining significantly deeper compression and an order of magnitude performance gain over (Gonzalez, Han, Li, & Klabjan, 2006) in the processing of path queries.

The importance of the movement graph approach to RFID data warehousing can be illustrated with an example.

Example 1. Consider a large retailer with a global supplier and distribution network that spans several countries, and that tracks objects with RFID tags placed at the item level. Such a retailer sells millions of items per day through thousands of stores around the world, and for each such item it records the complete set of movements between locations starting at factories in producing countries, going through the transportation network, and finally arriving at a particular store where it is purchased by a customer. The complete path traversed by each item can be quite long as readers are placed at very specific locations within factories, ships, or stores (e.g., a production lane, a particular truck, or an individual shelf inside a store). Further, for each object movement, we can record sensor readings, such as weight loss, humidity, or temperature. These lead not only to a tremendous amount of “scattered” data but also to a rather complicated picture.

The questions become “how can we present a clean and well organized picture about RFID objects and their movements?” and “whether such a picture may facilitate data compression, data cleaning, query processing, multi-level, multi-dimensional OLAPing, and data mining?”

The movement-graph approach provides a nice and clean picture for modeling RFID tagged