RFID Intelligence

Tiago Peralta

50267

tiago.peralta@tagus.ist.utl.pt

Abstract. The constant search for a greater visibility of product movement in the organizations supply chains as turned Radio Frequency Identification technology more accessible (1), allowing implementations on the own item. However these implementations face the challenge of transforming the RFID data in relevant information for decision support, not so trivial (2) due to the large volume of simple, inaccurate and strongly spatial and temporal data. This data explosion, together with the spatial and temporal component, creates difficulties to the actual decision support systems based in the multidimensional data model cube (3), namely in the search of interesting information about object movement. This is precisely where this work as is relevance, accessing the difficulties that the data model cube (3) has when dealing with this kind of data and proposing and studying a new data model (4) to treat, organize and analyze the RFID data, checking the potential advantages and disadvantages when compared to the traditional data model in a luggage location and monitoring project.

1 Introduction

The constant search for a greater control and visibility of product movement in the organizations supply chains as turned RFID technology in a potential source for information. Although this technology is not new, the size reduction of the devices, standardization and the adoption by the industry “big players” like Wallmart and the Department of Defense of the United States, with the consequent device price reduction has allowed the implementation by a number of industry sectors (1). However, more detailed data is not necessarily a added value. To achieve that added value we need to analyze and relate that data to achieve information and knowledge that we can use to support the decision making. That´s the point where the RFID and BI technologies interact, providing RFID the data and BI the analysis of that data in order to provide relevant information for decision support. However this liaison is not trivial, due to the special characteristics of the RFID data (2). This is precisely where this work as is relevance, accessing the difficulties that the data model cube (3) has when dealing with this kind of data and proposing and studying a new data model (4) to treat, organize and analyze the RFID data, checking the potential advantages and disadvantages when compared to the traditional data model in a luggage location e monitoring project.

2 Problem Definition

RFID readings generate large volumes of simple temporal and spatial data. A Data Warehouse uses the cube data model (3) to aggregate and relate this data in order to provide relevant information for decision support. Each reader generates readings in the form (EPC, location, time), meaning EPC the tag identification, location the physical location of the reading and time the temporal moment of the reading. If the item stays in a location for a period of time, duplicate readings are generated. To remove this duplicates we can group this reading in the form (EPC, location, time_in, time_out), meaning EPC the tag identification, location the physical location where the reading occurred, time_in the time of the first reading of the object in the location and time_out the last reading of the object in the location. If we consider the table that contains these records as the Data Warehouse fact table, the cube data model will calculate all possible group by for this table, aggregating the records that share the same values in all possible dimensions. With this data organization in the Data Warehouse, we can see for example the number of objects of a type that stayed in a location. However we can't obtain a response if we need to know how many of those items traveled to another location (4) (6) because the cube data model doesn't consider the relations between the different records. This dissertation contributes in the resolution of this problem, accessing the difficulties that the cube data model has when dealing with RFID Data, proposing and studying a new data model (4) to treat, organize and analyze the RFID data, checking the potential advantages and disadvantages when compared to the traditional data model in a luggage location e monitoring project.

3 Related Work

The considerations about the special characteristics of RFID data and his challenges have been studied in some works (2) (14). We can face the RFID data management problem using two approaches. The first approach is to consider the on-line treatment of RFID data, by processing the data directly in the data flow (15) (16) (17). The second approach is the off-line treatment of the RFID data, by storing the data and analyzing it (4) (6) (18) (19) (20) (21). The last approach is followed in this work.

In the off-line approach, the EPC model has some limitations when dealing with complex queries that require state information (18). The DRER model expands the relational model to deal with the spatial and temporal characteristics of the RFID data, however this model is only valid when we have reading only tags and fixed readers. The DRMA (20) model expands the DRER model to include a wider range of applications, however this two models are based in traditional OLTP models, more adequate to deal with operational databases and not with historical data used in decision support. The bitmap data type (21) is a more adequate response in organizing and analyzing historical data, reducing the space that this data occupies in the Data Warehouse by compressing the tag identification as a collection of tag identifications. However this type of compression may not work adequately when we can't make groups hover an item property or proximity. Besides, in the cases where the epc_suffixes are not contiguous, the bitmap will have a large number of zeros, reducing the efficiency of the bitmap operations. Plus the incremental actualization of data is extremely inefficient.

An interesting proposal is suggested by Gonzales et al (6), presenting a model based in some RFID data considerations, like the removal of duplicate readings of an item in a same location, the grouping of items that share the same abstraction level and the consideration of paths based on his relevancy for the analysis. In the same study it is suggested that the cube data model presents some limitations when dealing with the RFID data, proposing a new model, the RFID Cuboid, to deal with analysis of the relations between the records. However this model is only relevant when we can group objects that travel together during some points of the paths. In the case study, based in a project of luggage tracking and monitoring, each luggage generates different temporal readings, making it impossible to group luggages. Besides, the GIDS prefix schema uses string comparison to make joins between tables, more inefficient than using numeric comparison.

4 Solution Definition

In order to analyze the historical object movement, in a project of luggage tracking and monitoring we will study the model proposed by Chung et al (4). In a first step, similar to the Gonzales et al (6) approach, we gropu the records that share the same EPC, location butt differs in the temporal moment in the form (EPC, location, time_in, time_out), meaning EPC the tag identification, location the physical location of the reading, time_in the time of the first reading of the object in the location and time_out the last reading of the object in the location. In a second step we transform this records in the form **EPC: Li [Si; Ei] --> … --> Lk [Sk; Ek]** where Li; …; Lk are the locations where the tag is detected, Si the time of the first reading of the object in the location Li and Ei the last reading of the object in the location Li. We will deal with these records instead of the original records. After this initial treatment we apply a data compression based in three steps:

The first step codifies the locations that form a path, using a fundamental theorem of arithmetic, the unique factorization theorem (22) saying:

- Any natural number greater than 1 can is exclusively expressed by the product of prime numbers.
- Supposing we have different locations, and for each location we have attributed a prime number in the form Prime (Li).
- We can define then a Element Number Encoding for the path L1 -> … -> Lk has Prime (L1) x … x Prime (Lk).
- With the Element Number Encoding we know the locations that compose a path, has the Element number Encoding is factorized exclusively by the

product of the prime numbers associated to the locations that compose the path.

However, although we know the locations that compose the path, we need to know also the order of the locations in the path. That can be achieved using another theorem, the Chinese remainder theorem (22):

- If n1; …; nk are pair wise relative prime numbers, then there's X between 0 e N (= n1 … nk) that resolves the system of simultaneous congruence.

- To code the order of the locations in the path we use the system of simultaneous congruence:

    - X mod Prime (L1) = 1

    - …

    - X mod Prime (Lk) = k

- We call Order Encoding to X, being 1; …; e k the levels of the nodes associated to L1; …; Lk.

- Last, to store time information of the objects a time tree is constructed, each node representing a reading of an object in the same location, on the same time. To efficiently search in this tree, the nodes are stored containing a Region Number Scheme, Start and End, obtained during a Depth Search. In this manner, this Regions Number Scheme has the property of expressing the hierarchy relations between locations. A node A is an ancestor of B only if A.Start <*B.*Start e B.End <*A.*End.

5 Case Study

This work accesses the difficulties that the data model cube (3) has when dealing with RFID data, proposing and studying a new data model (4) to treat, organize and analyze the RFID data, checking the potential advantages and disadvantages when compared to the traditional data model in a luggage location e monitoring project. In this chapter we introduce the project. The DUMBO project was developed by Link consulting and has born from the identification of opportunities in the Aviation Industry, namely due to the weak luggage traceability in the airport. According to the Air Transport Users Council, based in data of the Association of European Airlines, the number of lost luggage as increased, in 2007 as represented about 6.2 million of lost luggage. To resolve this problem, several companies have developed experiments to track the luggage movement in the airport, mainly using barcode systems. However the barcode systems have a low reading efficiency, never going beyond the 90%. In this level the RFID is presented and supported by the International Air Transport Association has a possible solution (14). The DUMBO project implements a solution based on RFID, providing the middleware for the decision support system. This case study shares the same assumptions and scope of the DUMBO project, diverting only in the chosen airport in order to simulate large luggage traffic. The chosen airport was the Zurich Airport (24). The readers were positioned in the entries and exits of the major components of the luggage system, namely the Conveyors**:** used to connect the different locations that compose the luggage system, Check-Ins: the entry point of the luggage that are destined to the departure flights, Distribution Center South: luggage distribution center, combining the luggage that comes from the Check-Ins 2 and 3, Primary Sorters: responsible for the baggage identification, security verifications, luggage storing and non read luggage treatment, Secondary Sorters: directs each luggage to its destination and Transfer Offload: entry point for the luggage coming from the arrival flights.

6 Experiments

In order to access the cube data model response to tracking queries over RFID data and to check the potential advantages and disadvantages of the model proposed by Chung et al (4) we developed three modules: The first module was a RFID data generator to simulate the luggage readings and support the two prototypes. The second module was a representation of a traditional system that uses the cube data model. We developed the ETL process and the analysis application, using SQL Server 2008 following the process described in (12). The chosen dimensions were Flight, Luggage, Geography, Date and Time. Finally the third system represented the approach of Chung et al (4) and was developed in C# and SQL Server 2008. After the development of the prototypes we proceeded to the experiments. In this chapter we use the prototypes applied to the case study to: Access the difficulties presented to the cube data model when analyzing tracking data and the need for a new model; Check the response of the new model proposed by Chung et al (4) when dealing with the same situations of the cube data model, comparing the two model response in the context of the case study, a project of luggage tracking. To simulate the data we used the data generator and generated data to simulate reading for 2500 flights, 375334 luggage and 3418340 luggage readings. We first access the loading time for the traditional model:

- ETL: 00:06:32
- Cube Processing: 00:01:37

Then we access the loading time for the model proposed by Chung et al (4):

- Building of Trace Records: 00:12:18
- Encoding Scheme transformations: 02:19:30 (Mostly building the time tree)
- File Loading: 00:04:45

After the ETL process we start to define the queries used in the experiments:

1. What is the path made by the luggage with LPN 10012?
    - Traditional Model Response: 00:00:06
    - New Model Response: 00:00:00.5

2. Witch luggage traveled through the path **Check-in 1 to Plane by Secondary Sorter Dock A?**
    - Traditional Model Response: (Not possible)
    - New Model Response: 00:00:00.5 (Without time information)

- New Model Response: over 00:15:00 (With time information)

3. Witch luggage traveled in the Check-in 1 to Plane by Secondary Sorter Dock E Weast path with a treatment time greater than 20 minutes in the Check-In 1?”
   - Traditional Model Response: (Not Possible)
   - New Model Response: 00:00:02

4. What is the average luggage treatment time for the path Check-in 1 to Plane by Secondary Sorter Dock E Weast?
   - Traditional Model Response: 00:00:00.4
   - New Model Response: 00:01:05

5. What is the total number of luggage that traveled in the path Check-in 1 to Plane by Secondary Sorter Dock E Weast?
   - Traditional Model Response:00:00:00.1
   - New Model Response: 00:00:00.2

7 Conclusion

This work contributes to resolve the problem presented to the cube data model when analyzing the movement of object by accessing the problem, studying a possible implementation in a luggage tracking project. To achieve that we started analyzing the Gonzalez et al (6) model, concluding after a theoric validation that this model is not adequate to a luggae tracking project because we cant group luggages without loss of information. The next potential model, proposed by Chung et al (4), was considered valid to this case after the same validation, so the prototype phase was initiated. With the prototypes developed we accessed that the cube data model can’t respond efficiently to queries like “Witch luggage traveled in the Check-in 1 to Plane by Secondary Sorter Dock E Weast path with a treatment time greater than 20 minutes in the Check-In 1?”. The model proposed by Chung et al (4) responds efficiently to the queries and is a good alternative when we have projects that need to analyse object movements like luggage tracking. However, this implememtation as some potential problems that require attention. In the time tree construction, as luggages have different read times in the entry and exit points of the path segments, the number or nodes in the second level of the tree is very large, making the construction of the time tree a time consuming task and generating a large number of records in the time table, making time related queries inneficient. Besides, this model is based in a relational approach, making the queries based in aggregates less efficient than using the cube data model. Another time consuming task is the conversion of the trace records, especially when treating very large datasets. To supress this problem we suggest loading the operational data frequently. This problem is visible also when constructing the time tree. Another problem that we encountered was the possibility of an overflow of the element encoding number and the order encoding number when dealing with very large paths. This problem was detected by Chung et al in a posterior work, however our work didn’t reflect that approach because when the work was published we were already in experiments. This work has succeeded in accessing the cube data model problems when dealing with tracking data and studied the advantages and problems of a proposed alternative in a real luggage tracking project.

8 Future Work

An interesting step forward would be a more balanced solution, improving the response of this model when dealing with time queries and the large dataset loading of data. An alternative could be a distributed model in order to extend this concept to a grid of airports, allowing a more rich and complete tracking analysis. Another interesting step could be a study of the integration of this model with a BAM, generating SLA values that could be used by the Bam to generate real time alerts.

9 Bibliography

1. **Smith, R.** *RFID: A brief technology analysis.* s.l. : CTOnet.org, 2004.

2. **Derakhshan, R., Orlowska, M.E. and Li, Xue.** *RFID Data Management: Challenges and Opportunities.* s.l. : IEEE International Conference on RFID pp.175-182, 26-28 March 2007.

3. **Gray, J., et al.** *Data cube: a relational aggregation operator generalizing GROUP-BY, CROSS-TAB, and SUB-TOTALS.* s.l. : Proceedings of the Twelfth International Conference on pp.152-159, 26 F, 1996.

4. **Lee, C. and Chung, C.** *RFID Data Processing in Supply Chain Management using a Path Encoding Scheme.* Vancouver, BC, Canada : SIGMOD'08, June 9–12, 2008.

5. **Segura, Consórcio Mala.** *Mala Segura.* s.l. : MultiSector Norte - Consultores em Estratégia, Tecnologia e Gestão Industrial, Lda, Fevereiro 2008.

6. **Stockman, H.** *Communication by Means of Reflected Power.* s.l. : Proceedings of the IRE, vol.36, no.10, pp. 1196-1204, Oct. 1948.

7. **Gonzalez, H., et al.** *Warehousing and Analyzing Massive RFID Data Sets.* s.l. : ICDE '06. Proceedings of the 22nd International Conference on pp. 83-83, 03-07 April 2006.

8. **Korkmaz, E. and Ustundag, A.** *Standards, Security & Privacy Issues about Radio Frequency Identification (RFID).* s.l. : RFID Eurasia 2007 1st Annual pp.1-10, 5-6 Sept. 2007.

9. **D., Brock.** *The Electronic Product Code (EPC) A Naming Scheme for Physical Objects.* s.l. : MIT Auto-Id Center, January 1, 2001.

10. **Kin Seong Leong, Mun Leng Ng, Engels D.** *EPC Network Architecture.* s.l. : AUTOIDLABS, September 2005.

11. **F., Armenio, et al.** *The EPCglobal Architecture Framework.* September 2007.

12. **Kimball, R.** *The Data Warehouse Toolkit Second Edition The complete guide to dimensional modeling.* s.l. : John Wiley and Sons , 2002. , “”, .

13. **Chaudhuri, S. and Dayal, U.** *An overview of data warehousing and OLAP technology.* s.l. : SIGMOD Rec. 26, 65-74., Mar. 1997.

14. **IATA.** *RFID BUSINESS CASE FOR BAGGAGE TAGGING.* 2007.

15. **S. S. Chawathe, V. Krishnamurthy, S. Ramachandrany, and S. Sarma.** *Managing RFID Data.* s.l. : VLDB, pages 1189–1195, 2004.

16. **Y. Bai, F. Wang, P. Liu, C. Zaniolo, and S. Liu.** *Rfid data processing with a data stream query language.* s.l. : ICDE, 2007.

17. **Chui, Chun Kit.** *The Design and Implementation of an OLAP System for Sequence Data Analysis.* s.l. : Proceedings of the Second SIGMOD PhD Workshop on Innovative Database Research (IDAR 2008), June 13, 2008.

18. **F. Wang, S. Liu, P. Liu, and Y. Bai.** *Bridging physical and virtual worlds: Complex event processing for rfid data streams.* s.l. : EDBT, 2006.

19. **Harrison, M.** *EPC Information Service - Data Model and Queries.* s.l. : Auto-ID Center, 2003.

20. **Liu, F. Wang and P.** *Temporal Management of RFID Data.* s.l. : VLDB, 2005.

21. **Shaorong Liu, Fusheng Wang and Peiya Liu.** *A Temporal RFID Data Model for Querying Physical Objects.* s.l. : A TIMECENTER Technical Report, February 8, 2007.

22. **Hu, Y., Sundara, S., Chorma, T., Srinivasan, J.** *Supporting rfid-based item tracking applications in oracle dbms using a bitmap datatype.* s.l. : VLDB pp. 1140–1151 , 2005.

23. **Rosen, K. H.** *Discrete Mathematics and Its Applications.* s.l. : McGraw-Hill, 2003.

24. **T., Vogel.** *Technical System Specification to RFID-enable the Baggage Sorting System at Zurich Airport.* s.l. : Master Thesis, ETH Zurich, 2007.