Generic Data Warehousing for Consumer Electronics Retail Industry

S. Habte, K. Ouazzane, P. Patel, S. Patel

***Abstract*—**The dynamic and highly competitive nature of the consumer electronics retail industry means that businesses in this industry are experiencing different decision making challenges in relation to pricing, inventory control, consumer satisfaction and product offerings. To overcome the challenges facing retailers and create opportunities, we propose a generic data warehousing solution which can be applied to a wide range of consumer electronics retailers with a minimum configuration. The solution includes a dimensional data model, a template SQL script, a high level architectural descriptions, ETL tool developed using C#, a set of APIs, and data access tools. It has been successfully applied by ASK Outlets Ltd UK resulting in improved productivity and enhanced sales growth.

***Keywords*—**Consumer electronics retail, dimensional data model, data analysis, generic data warehousing, reporting.

I. Introduction

As multi-channel consumer electronics retailers operate in dynamic and highly competitive market conditions, they are under constant pressure to invest in new communication and information technologies that automate or support retail activities to increase productivity and gain competitive advantage. Various software applications are in current use by retailers to assist with customer sale transactions, online business, inventory management, purchasing process, finance and human resource management. These software applications generate a huge volume of transactional and other types of data such as web logs, customer feedback, and foot fall counters. Retailers are facing challenges in utilizing the generated datasets to support managers in deciding optimal inventory, choosing a profitable price, introducing new product lines, and satisfying customers. To address these challenges, data warehousing is one of the most popular and effective solutions since it integrates all available data sources into a central corporate data repository and organizes data in such a way that it answers business questions easily and quickly [1], [2].

Designing, implementing, and deploying a data warehousing solution requires a significant commitment in terms of time, money, and expertise. Therefore, Small to Medium Enterprises (SMEs) are not willing to invest on a data warehousing project as financial resources are scarce. However, under the current economic climate, it is difficult for any company to remain competitive in the market without using data warehousing platform.

In this paper, we propose a generic data warehousing solution which meets the needs of retailers who favor a hybrid multi-channel business model in which agility of decision-making is crucial for survival and growth. The generic nature of our solution means that it can be readily applied to any consumer electronics retailer with minimum configuration and will be capable of significantly reducing the time, money, and expertise required to design, implement, deploy and maintain a data warehousing provision.

The remainder of this report is organised as follows. In Section II, we present our case study based methodology. In each of the subsequent three sections, analysis of business processes, dimensional data modelling, and implementation of the proposed data warehousing solution are described. Finally, the conclusions and aspects of further work are presented.

II. Methodology

A deductive approach is applied to design and develop a generic data warehousing solution. First, a specific consumer electronics retailer was considered as a case study and then a data warehousing solution was designed and developed based on their reporting and data analysis requirements. Based on Kimball's data warehousing life cycle approach [1], we applied a method that includes three stages to design and develop a generic data warehousing solution that is applicable to a wide range of consumer electronics retailers. The three stages are: analysis of business processes, dimensional data modelling, and implementation of data warehousing [1]. The details of these three stages are described below.

III. Analysis of Business Processes

A business process is a set of activities performed to accomplish an organizational goal [3], [4]. Based on the analysis of the case study, we identified five common business processes: serving customers, marketing, purchasing, handling inventory, and managing customer relationship. Once the business processes are identified, each business process is described, and its reporting and analysis requirements are gathered. We applied a standard technique that includes interview, questionnaire, and written documentation to gather requirements.

A. Serving Customers

This business process involves recording transactions, receiving payments from customers, invoicing, dispatching

K. Ouazzane and P. Patel are with the London Metropolitan University, London, N7 8DB UK (e-mail: k.ouazzane@staff.londonmet.ac.uk, p.patel@staff.londonmet.ac.uk).

S. Habte (KTP Research Associate) and S. Patel (managing director) are with the ASK Outlets Ltd. (e-mail: s.habte@staff.londonmet.ac.uk, s.patel@askdirect.co.uk).

ordered products to customers, handling returns, and answering customer enquiries. The reporting and data analysis requirements are to analyse sales by outlets, products, and customers.

B. Marketing

Marketing enables retailers to understand customer buying behaviours. It involves promoting products, setting or updating prices, understanding the prevalent market conditions, and identifying new selling channels and new stores. Evaluation of the impact of price changes or promotional trade offerings on sales volumes, and analysis of sales are the data analysis requirements for this business process.

C. Managing Inventory

The main tasks for managing inventory are reviewing inventory below minimum levels, receiving inventory and storing it for sale, and maintaining inventory accuracy. The reporting and analysis requirements are to monitor inventory performance, identify overstocks and under-stocks, and measure inventory accuracy.

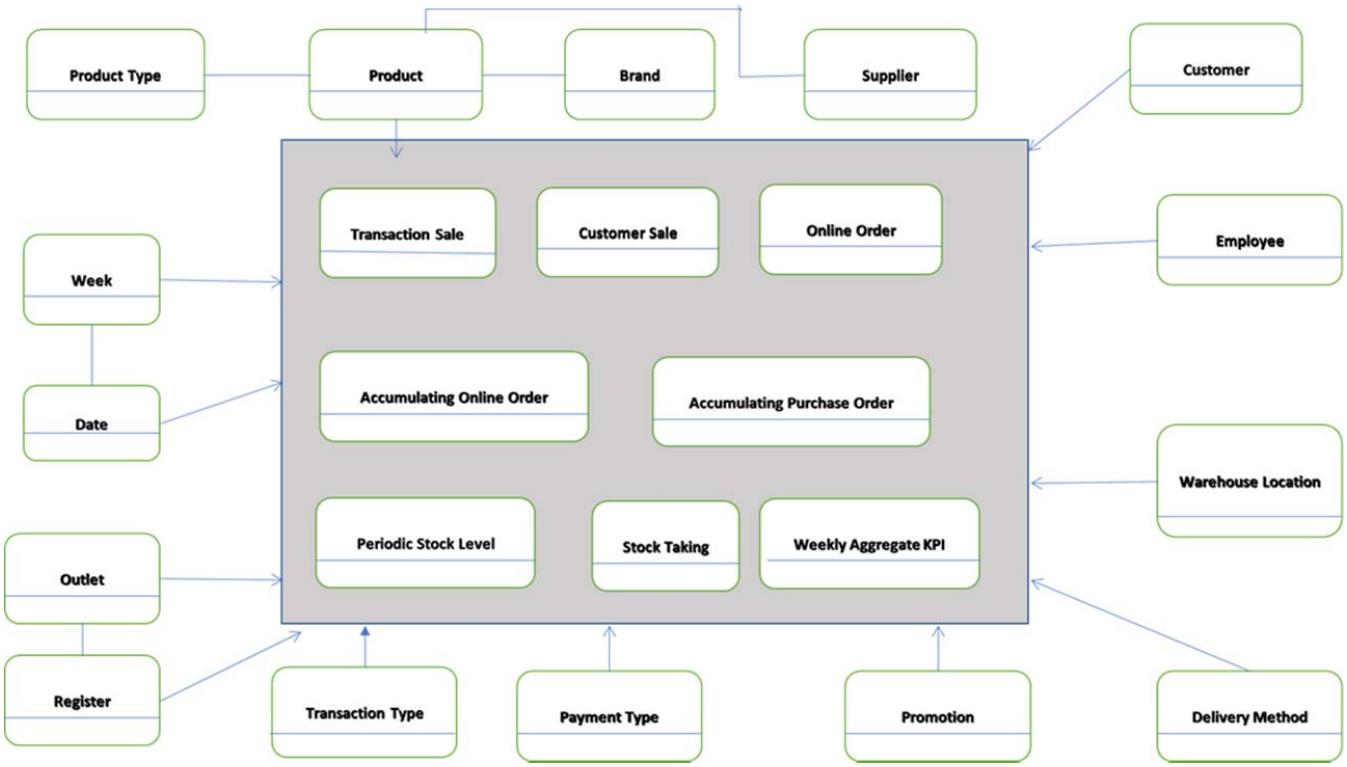

Fig. 1 Dimensional data model for a typical consumer electronics retailer

D. Purchasing

The goal is to develop profitable product ranges that reflect current trends and offer customers product choice. The main activities involved in purchasing business process lie in choosing product lines or models to stock, determining each product quantity to order, introducing new product ranges, deleting product ranges, dealing with suppliers, and handling delivery schedules. The reporting and analysis requirements for this business process are to: identify popular and profitable products, forecast consumer demands, monitor performance of suppliers, and track delivery schedules.

E. Managing Customer Relationship

This business process consists of a set of strategies, programs and systems that focuses on identifying and building loyalty with the retailer's most valued customers. The reporting and analysis requirements are to:

1. Maintain customer information (i.e. transaction history, customer contacts, customer preferences, demographic data, and responses to marketing activities).
2. Analyze transaction and contact data.
3. Identify target customers, the best and most profitable customers, patterns in the customer data that can improve the effectiveness of retailing decisions.

IV. Dimensional Data Modelling

Dimensional data modelling is a logical design technique for structuring data which enables easy and quick answers to business questions [5]. It uses fact and dimension tables to represent and structure data. Fact tables store performance measurements generated by business activity or events whereas dimension tables are a set of descriptive fields that describe the context of facts [1].

Once we have defined the business processes and gathered the reporting and analysis requirements, we adopt an iterative process to build dimensional data models that represent the identified business processes and meet the reporting and

analysis requirements of a typical consumer electronics retailer. The process of building dimensional data model includes three steps, namely building high level dimensional model, identifying attributes of dimensions and measures of fact tables, and defining and representing the structure of the dimension data model using schemas such as star, snowflake and hybrid [1].

To model the above business processes eight fact tables and 15 dimension tables have been identified. The identified fact tables are transaction sale, customer sale, online order, accumulating online order, accumulating purchased order, periodic stock level, stock taking, and weekly aggregate KPI. The dimensions are product, product type, brand, supplier, customer, week, date, outlet, register, transaction type, payment type, promotion, employee, warehouse location, and delivery method.

An overview of the dimensional model is depicted in Fig. 1. As shown in the figure, a hybrid schema is chosen to structure the relationship between fact and dimension tables. The hybrid schema allows a combination of the advantages of both simple star and snowflake schemas. The dimension tables are normalized in the same way as in the snowflake schema for future flexibility and at the same time they contain also de-normalized attributes as in the simple star schema for improved query performance [2].

V. Data Warehousing Solution

This section describes how the data warehousing solution is implemented based on the dimensional data model developed in the previous section. The proposed data warehousing solution is based on the use of four software tools. The high-level architecture of the proposed data warehousing solution is shown in Fig. 2.

The first tool is a relational database engine which stores and manages fact, dimension, staging, and lookup tables. It includes two relational databases. The first database stores the fact and dimension tables whereas the second database stores staging and lookup tables. Any relational database management system such as Microsoft SQL Server, Oracle, IBM DB2, and MySQL can be used to create and manage the two databases [6]. A template SQL script that creates the two databases is part of the proposed solution.

The second software tool is responsible for populating the dimension and fact tables. It is known as Extract, Load, and Transform (ETL) which are the main tasks required to populate a dimension or fact table [7]. There are different ways of developing and deploying an ETL tool. Some of the popular ETL tools are SAP Data Services, SQL Server Integration Services (SSIS), Oracle Data Integrator (ODI), and Talend Studio for Data Integration [8]. Most of them are commercial and require IT expertise for use and maintenance. Scheduling ETL tools to populate the data warehouse regularly is also another issue. For example, ETL tools developed using Microsoft SQL Server Integration Services (SSIS) cannot be scheduled or executed without SQL Server. The developed ETL tool based on the use C# object oriented programming language can be scheduled without any relational database server and does not need IT expertise to be maintained. It can be easily developed using Java object oriented programming language so that it can run also on operating systems other than windows.

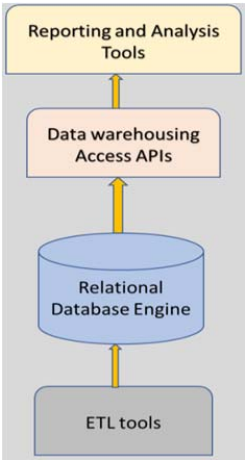

Fig. 2 High level architecture of the proposed data warehousing solution

The ETL tool has two components; the first component extracts raw data from different data sources (flat file, excel or csv files, web services, relational tables) and loads it into the staging database, while the second component cleans and transforms the extracted raw data and loads it into fact and dimension tables. Only the first component needs to be customized depending on which data sources are available. The second component can be easily applicable to a wide range of consumer electronics retailers as it extracts data from the staging database.

The third software tool is a set of data warehousing access Application Programming Interfaces (APIs) for exposing the data warehouse using HTTP protocol and JSON data format. They are developed using ASP.NET Web API framework.

For each dimension, either all records or some records filtered by its attributes can be queried using the APIs. The APIs can be used to retrieve records of fact tables by any possible combination of their dimensions. For example, a transaction sale fact table can be queried by date, product, transaction type or outlet. It can also be queried by date and product to compare the sale of each product for different months, weeks or years.

The fourth software tool is a reporting and data analysis which depends on the services provided by the data warehousing access APIs. Its purpose is to implement all the reporting and data analysis requirements which are useful for a consumer electronics retailers. It provides an easily accessible user interface for managers to generate reports and perform analysis.

The graphical user interface is shown in Fig. 3. It displays a list of dimensions and measures. The user can select either a dimension or a measure. When a user selects a dimension, the relevant attributes and possible measures are displayed to enable users to analyze a measure or the dimension by its

attributes. When a user selects a measure, the relevant dimensions are displayed to enable users to analyze the measure by each dimension or by a combination of the dimensions.

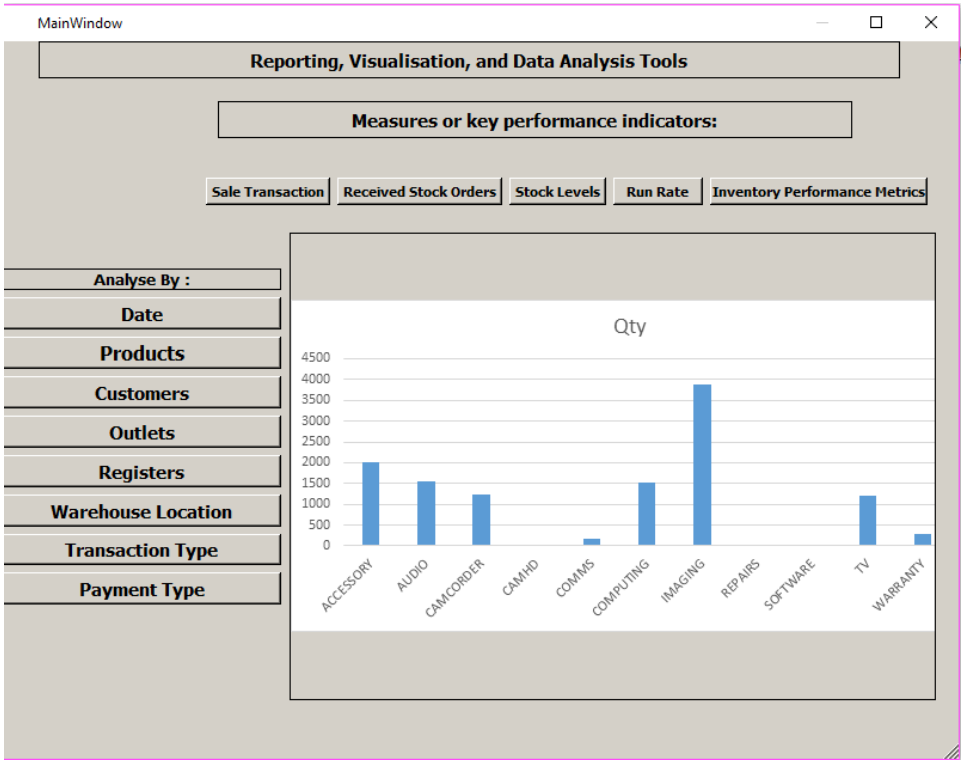

Fig. 3 User interface of the reporting, visualization, and data analysis tools

VI. Conclusion

The generic and platform-independent nature of the proposed data warehousing solution makes it a viable option for immediate adoption by a wide range of consumer electronics retailers. The proposed solution has been successfully applied by ASK Outlets Ltd, a multi-channel consumer electronics retailer based in London, UK, who report on early, high-value impact to their business in terms of enhancing revenue, increasing market share, reducing inventory handling cost, and increasing customer satisfaction.

The proposed solution is built mainly based on five common business processes: serving customers, marketing, managing inventory, purchasing, and managing customer relationships. As a result, it can address the reporting and data analysis requirements relevant to these five business processes which constitute the core elements of the electronics retail industry.

Acknowledgment

The research undertaken in this research paper was funded by Innovate UK and ASK Outlets Ltd and carried out in partnership with London Metropolitan University.

References

[1] Ralph Kimball and Margy Ross. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (2nd ed.).* John Wiley & Sons, Inc., New York, NY, USA. 2002.

[2] Acuma Training Institute: Data Warehouse Design.

[3] Ruth Sara Aguilar-Savén, Business process modelling: Review and framework, International Journal of Production Economics, Volume 90, Issue 2, 28 July 2004, Pages 129-149, ISSN 0925-5273, http://doi.org/10.1016/S0925-5273(03)00102-6.

[4] http://www.appian.com/about-bpm/definition-of-a-business-process (accessed on 18/04/2017).

[5] John Paredes. The Multidimensional Data Modeling Toolkit: Making Your Business Intelligence Application Smart with Oracle OLAP. 2009.

[6] Connolly, T. & Begg, C. Database Systems - A Practical Approach to Design, Implementation and Management (6th ed.), Pearson Education. 2014.

[7] Michelle C. Dy, Michael W. Hawkins, Mark Humphries ,Data Warehousing: Architecture and Implementation' Prentice Hall, *Release Date: December 1998.*

[8] https://www.etltool.com/list-of-etl-tools (accessed on 3/4/2017).