Data Warehouse for FER e-Invoice System

Luka Humski*, Igor Lažegić** and Zoran Skočir *
* University of Zagreb Faculty of Electrical Engineering and Computing / Department of Fundamentals of Electrical Engineering and Measurements, Zagreb, Croatia
** Ericsson Nikola Tesla d.d., Zagreb, Croatia
{luka.humski, zoran.skocir}@fer.hr, igor.lazegic@ericsson.com

Abstract - An e-invoice is the most widely used electronic document in the world. Although it covers only one segment of the entire supply chain, the e-invoice has a central role in the development of electronic business. An electronic invoicing system named an *FER e-invoice* is built at the University of Zagreb Faculty of Electrical Engineering and Computing (FER). The *FER e-invoice* system contains business processes that should be monitored and analyzed. Therefore, a data warehouse, containing data extracted from e-invoice XML documents, has been developed. Multidimensional model of the data warehouse consists of two fact tables, one for the invoice heading and one for the invoice items, and of 14 dimensional tables. In this paper the developed data warehouse is presented.

Key words: data warehouse, e-invoice, electronic business, XML, XSLT, ETL, dimension table, fact table

I. INTRODUCTION

Electronic business (e-business) includes business transactions and information interchange using information and communication technology in an enterprise, between enterprises and their customers or between enterprises and public administration [1]. The basic part of electronic business is the supply chain which consists of business processes that include catalog, tendering, contract, order, scheduling, invoice, payment and remittance advice. To be able to implement such a supply chain together with the entire process of e-business it is necessary to establish specific electronic documents and procedures such as e-catalog, e-order, e-invoice, e-invoice user registry, e-payment, e-signature, e-identity altogether with semantic and technical interoperability. The goal of electronic business is to achieve the highest possible rate of automation and digitalization of business processes. For example, in a world where electronic documents do not exist, a customer would have to write his order on a piece of paper and send it to the seller using the mail service, fax or electronic mail service. The seller would have to manually analyze the order and approve it if it is a feasible one. After that, he would have to manually copy the elements from the order to the invoice and send the invoice through the same communication channel he received the order. In a world where electronic business is used the same procedure would be arranged by a simple exchange of a digital order, usually in a XML format. After the seller would approve the order, it would be feasible to automatically generate an invoice from its elements. That way it is possible to save time (faster invoice delivery, processing and charging), paper (environment friendly) and eliminate possible mistakes during manual copy of information which in the end results with financial gain.

E-invoice is the most outspread electronic document in the world. Its development is the key to the success of electronic business although it covers just one segment of the entire supply chain. It contains important data for the calculation of VAT therefore its usage has extra significance for the enforcement of the national fiscal policy. The widespread usage of internet banking in Croatia is the cornerstone to achieving the maximum effects of e-invoice introduction. Electronic document interchange in Croatia is regulated by the Electronic trade law [2], Electronic document law [3] and Electronic signature law [4]. Accounting law [5] regulates book-keeping documents composed in electronic notation while Common tax law [6] prescribes electronic data management. The VAT regulation has obstructed the vast implementation of e-invoice interchange by not allowing enterprises involved in the VAT system to use e-invoices for tax purposes. That obstacle was bypassed finally in July 2011 by the new VAT regulation [7] that states that electronic invoices are legally equal to the paper ones.

E-invoice implementation is a good example of a horizontal integration inside e-business and therefore should attract interest on one side from large and small enterprises and on the other from the public sector [8]. A system, where all entities must receive e-invoices in tax purposes (VAT) in case two conditions have been met, has been established in all European Union countries. The first condition is that the e-invoice receiver must agree to receive the invoice in an electronic form whereas the second one requires that integrity of the contents and the authenticity of the origin must be guaranteed during transport and archiving [9]. The transition from the paper invoice to the e-invoice brings several benefits. E-invoice utilization brings faster and easier management of invoice publishing, receiving and archiving. Moreover, all the sent and received invoices are archived in an electronic form. European analyses show that the costs of sending and receiving invoices should be lower for 60 - 70 percent compared to today's costs. According to [10], the implementation of e-invoices only in Croatia should bring six billion kuna savings annually, whereof 350 million just in public administration.

Electronic invoices in XML format as like as paper invoices contain important data that can be used for

business analysis. Data warehouse can be built based upon the data collected from e-invoices. Data warehouse is a database that contains historical unchangeable data which is collected and processed to support business decision making [11]. In such a data warehouse business processes can be analyzed with regard to the supplier, customer, way of payments, etc. As the source data is in XML format, it has to be transformed by using XSLT transformations in order to eliminate irrelevant data. Finally, the data is inserted into the data warehouse.

The paper is organized as it follows: chapter II briefly describes the *FER e-invoice* system (FER is abbreviation of Faculty of Electrical Engineering and Computing in Croatian); chapter III defines the concept of a data warehouse and specifies the procedure of data warehouse development and data loading; chapter IV depicts the development, data loading and the usage of the data warehouse implementation in the *FER e-invoice* system; the final chapter V contains the conclusion.

II. *FER E-INVOICE* SYSTEM DESCRIPTION

A pilot system for electronic invoice interchange has been developed on the University of Zagreb Faculty of Electrical Engineering and Computing. E-invoice interchange implies processes of e-invoice formatting, sending, receiving, storing and searching. A model for e-invoice users registry has been proposed and developed in order to support the e-invoice interchange system. Users can utilize the system through two basic applications. The first one is the application for e-invoice interchange which consists of the application for sending the e-invoice and the application for accessing the e-invoice archive. The second one is the application for accessing the e-invoice users registry. An additional application (registry form) is used for registering new e-invoice users.

A. *System architecture*

FER e-invoice system architecture is based upon 4 corner model (4C). As the name suggests, it has four basic components: user in role of an e-invoice sender, sender's service provider, receiver's service provider and user in a role of an e-invoice receiver. Fig. 1 shows the design of the 4C model (user 1 and 2 – e-invoice sender/receiver, SP – service provider).

E-invoices are being sent in such a way that the sender forms (manually or through the application) an e-invoice and forwards it to its service provider. The sender's service provider forwards the invoice to the receiver's service provider which passes it over to the e-invoice receiver. The e-invoice is archived on the sender's and receiver's service provider.

E-invoice user's data needs to be stored permanently. That is the reason why every service provider has an e-invoice users registry which stores all important data concerning the users that are the service provider's clients.

The important feature of this system architecture is the fact that the central part of the system is totally decentralized and at the same time built around a set of service providers. However, as the number of information

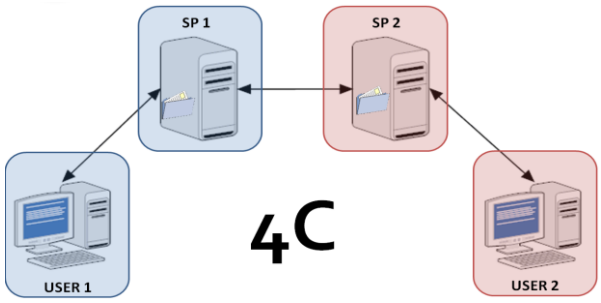

Fig. 1. The design concept of the 4C model

providers increases, it becomes more difficult to know precisely to which service provider the invoice must be sent to reach the designated receiver. That is the reason why the central e-invoice users registry has been developed and introduced into the system. It stores the essential information needed for e-invoice routing. After the e-invoice sender hands over the e-invoice to its service provider, the provider will fetch the receiver's OIB (personal identification number in Croatia) parameter from the invoice. Then it will contact the central users registry to find out the service provider it needs to send the e-invoice to so that the e-invoice reaches its designated receiver. The skeleton of the developed system for e-invoice interchange is shown in the Fig. 2. The image depicts the system's integral components and web services used for their communication.

FER e-invoice system consists of:

1. Applications that facilitate e-invoice formatting and sending, reading the archive of received and sent e-invoices and registering new users together with examining the data about current users
2. Service providers (with local e-invoice users registries) that facilitate e-invoice interchange
3. Central e-invoice users registry
4. Web services that enable the e-invoice interchange from the technical point of view.

III. DATA WAREHOUSES

Data warehouse is an important functional extension of *FER e-invoice* system described in previous section. It allows different business analyses based on e-invoice. This chapter shows basic characteristics of the data warehouse, while design of a *FER e-invoice* data warehouse is described in the next chapter.

The data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process [12]. Subject-oriented means that the data warehouse can be used to analyze a particular subject area. For example, "sales" can be a particular subject. Integrated means that data warehouse integrates data from multiple data sources. For example, source A and source B may have different ways of identifying a product, but in a data warehouse, there will be only a single way of identifying a product. Time-variant means that historical data is kept in a data warehouse. This contrasts with a transactions system, where often only the most recent data is kept. Non-volatile means that once data is in the data warehouse, it will not change. So, historical data in a data warehouse should never be altered.

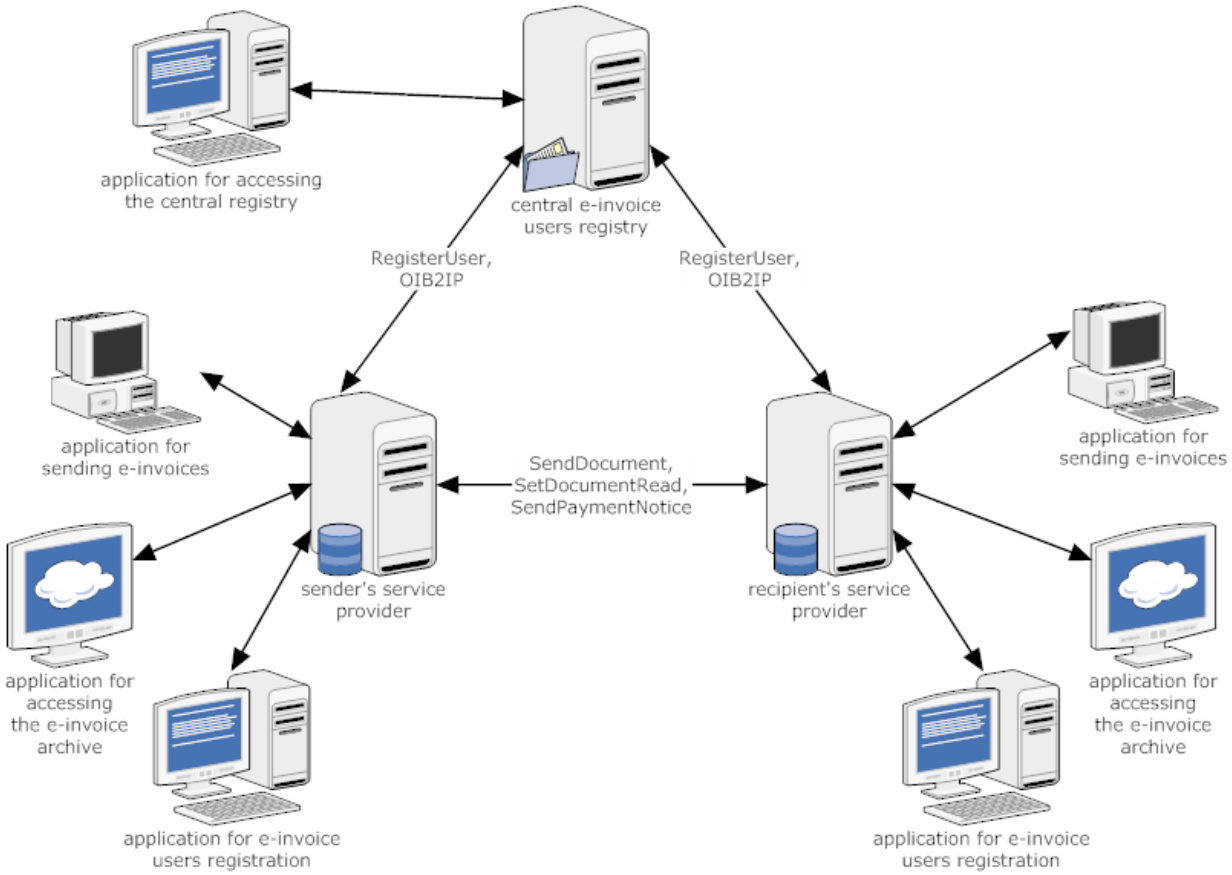

Fig. 2 The concept of *FER e-invoice* system architecture

The usual way of developing transactional databases is not suited for developing data warehouses. The most important difference between transactional databases and data warehouses is that data in transactional databases, unlike data warehouses, can often change. That is the reason why the transactional database must be normalized (in third normal form). In data warehouses the query execution speed is more important and that is why denormalized multidimensional modeling is used for developing data warehouses. Fig. 3 shows the basic activities that can be done on the data warehouse. Those are extraction, transformation and data loading (ETL); data storing and data management; data usage in the process of business decision making. Multidimensional modeling consists of conceptual, logical and physical modeling [13]. The result of conceptual modeling is the data warehouse schema that is DBMS (database management system) independent. Logical modeling results with a logical data warehouse schema which defines the way the data warehouse will be implemented. Physical modeling depends of the used DBMS.

Data warehouse development consists of four phases [14]:

1. Select the business process
2. Declare the grain
3. Choose the dimensions
4. Identify the facts (measures).

Data warehouse once designed and implemented must be loaded with data which can be done with the ETL procedure. ETL can be divided into initial and incremental loading.

IV. *FER E-INVOICE* SYSTEM DATA WAREHOUSE

FER e-invoice system holds important data about the business activities of every participant in the process of e-invoice interchange through its archived e-invoices in

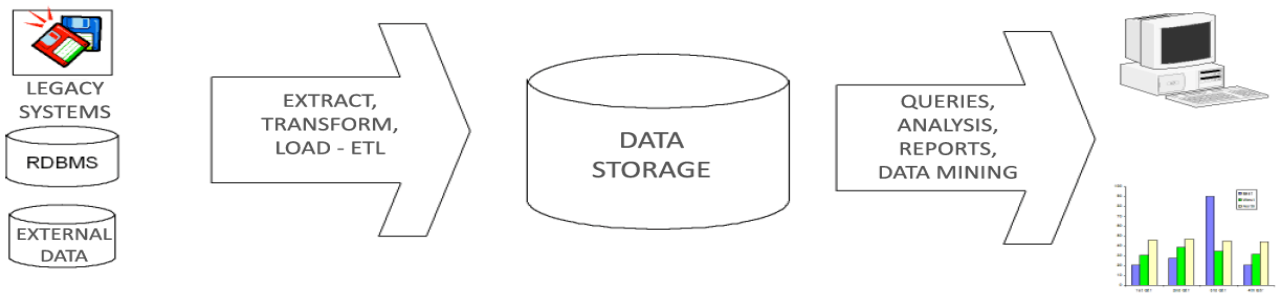

Fig. 3 The essential activites during the process of data warehousing

XML format. The data is related to the business entities (supplier, customer, deliverer and tax agent), business processes (delivery, payment method) and business objects (invoice, invoice item, tax, expenses, discount, article). Based upon such data with the goal of providing a service of quality business analysis to the business entities involved in the system, a data warehouse has been developed and implemented into the *FER e-invoice* system. The important characteristic of the system is that every service provider has its own data warehouse (data warehouse would be hosted on the separate server), but with the same conceptual model that enables later data integration for entire *FER e-invoice* system. Thanks to it, the provider can offer to his clients access to the data that are of their interest. In other words, it can offer business entities a possibility of analyzing their own business activities.

A. *Data warehouse model*

The first step of the data warehouse modeling was the analysis of the input specification. The input specification was the technical specification of e-invoices that are being interchanged in the *FER e-invoice* system. It is specified in [15]. Based upon the analysis, a data warehouse conceptual multidimensional model has been built [16]. Fig. 4 illustrates it. Business processes that have been modeled are the e-invoice and its items. These processes represent the master/detail relationship that is modeled by two fact tables: *fInvoice* and *fInvoiceItem*. Data source system's granularity has been adopted to the data warehouse's granularity, meaning that table *fInvoice* has one entry for each e-invoice while the table *fInvoiceItem* has one record for every invoice item of every e-invoice. Dimension tables relate to entities and their elements present in the e-invoice: *dCost* (contains cost data), *dDiscount* (contains discount data on the invoice level), *dDelivery* (contains delivery data), *dCustomer* (contains customer data), *dTaxAgent* (contains tax agent data), *dDeliverer* (contains deliverer data), *dSupplier* (contains supplier data), *dPaymentMethod* (contains payment

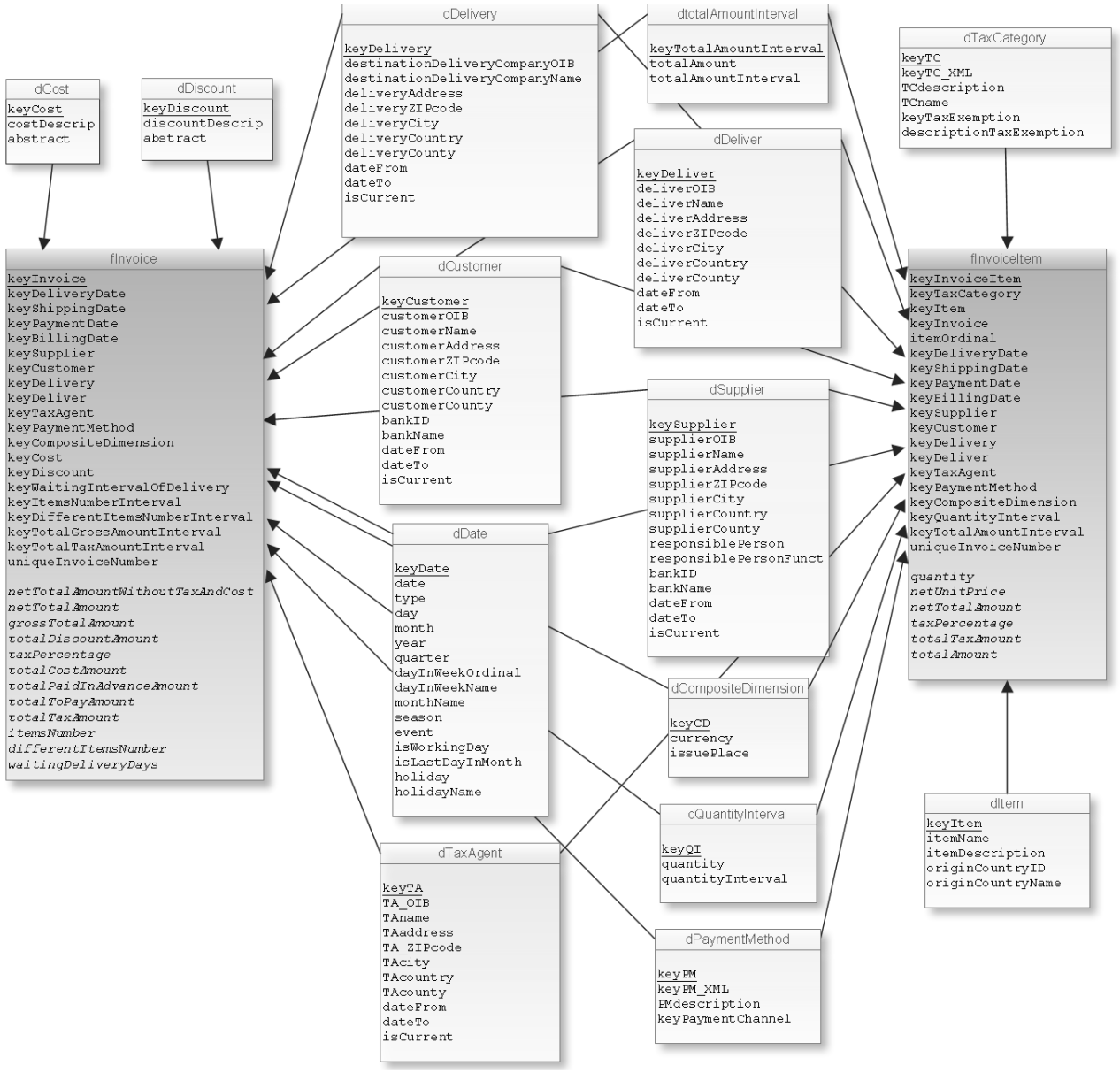

Fig. 4 Conceptual multidimensional model

method data), *dTaxCategory* (contains tax category data that is calculated for every invoice item) i *dArticle* (contains article data). The *dCompositeDimension* table is an example of a composite dimension. It is a dimension table that contains different types of flags and indicators of low cardinality. The role of composite dimension is to store flags and indicators outside of the fact table to achieve significant savings in disk space while suffering an insignificant loss in performances. Dimension table *dDate* is the data warehouse's time dimension. That means it takes over the meaning of all the date entries in the e-invoice (delivery date, dispatch date and similar). Interval dimension tables (*dQuantityInterval* i *dTotalAmountInterval*) have been created in the process of enrichment of the fact tables. Namely, attributes *itemsNumber*, *differentItemsNumber*, *grossTotalAmount*, *waitingDeliveryDays*, *totalDiscountAmount*, *quantity* and *totalAmount* can be inserted in specific intervals. That is the reason why dimension tables that define such intervals have been created. That way through the foreign key in the fact table it is noted which interval a certain value refers to. The foreign key points to the appropriate record in the interval table. As a result the representation of data is enriched and a sum report can be developed in an easier way during the process of conducting analytical processing over data in the data warehouse. Most of the dimension tables refer to the whole e-invoice (*fInvoice*), however the same dimension tables are connected also with the e-invoice items (*fInvoiceItem*) because the data that refers to the whole e-invoice refers to the every e-invoice item as well.

Measures are attributes listed in the bottom part of the fact tables shown in Fig. 4. Most of them are a direct copy of the appropriate elements of the e-invoice, while others (*fInvoice*: *itemsNumber*, *waitingDeliveryDays* i *discountPercentage*; *fInvoiceItem*: *totalAmount*) are the result of additional calculations done on the basic ones.

B. Data warehouse data loading

Data warehouse data loading consists of "manual" data loading of six dimension tables (*dDate*, *dQuantityInterval*, *dtotalAmountInterval*, *dTaxCategory*, *dPaymentMethod* i *dCompositeDimension*) and of an ETL procedure. It is used to insert the data into the rest of the dimension tables and into the fact tables based on the data stored in the e-invoices.

The ETL procedure can be divided into two parts: data warehouse initial data loading and incremental data loading which is used for keeping the data warehouse up-to-date. Fig. 5 shows the flow of the entire ETL process. The first step is to fetch e-invoices (initially all, after that just newer ones that have not been fetched) from the e-invoice archive. The e-invoice files have to be transformed with a series of XSLT transformations in order to eliminate data that is irrelevant for the data warehouse. Finally, data from the transformed files is inserted into the data warehouse with the ETL operations developed in the *SQL Server Business Intelligence Development Studio* tool.

C. Business analyses by using data warehouse

The way the data warehouse has been developed and loaded with data has been described so far. All that remains is to explain how data stored in the data warehouse can be used for business analysis. The data warehouse has been implemented in the *Microsoft SQL Server 2008 R2* DBMS. *SAP Business Objects* tool is used for complex analyses. For simpler analysis a web application has been developed. It can be easily used for running queries in the data warehouse. Fig. 6 depicts the application's user interface. Application has been developed in the Croatian language since the *FER e-invoice* system has been built for use in the Republic of Croatia. In the upper left corner the user can pick the fact table that will be analyzed. Under the list of the fact tables, is the list of measures and dimensions that can be selected. Based upon the selected measures and dimensions an SQL query shall be generated and executed on the selected fact table. The query results are shown in a form of a table in the central part of the web page. The table attributes are the selected measures and dimensions. To familiarize the user with the data warehouse the generated SQL query is shown over the results table.

V. CONCLUSION

The paper describes the architecture of the *FER e-invoice* system, the concept of data warehouses and the procedure of data warehouse development and data loading. The main part of the paper describes the *FER e-invoice* data warehouse system. The data warehouse is implemented around two fact tables: *fInvoice* and *fInvoiceItem*. Table *fInvoice* stores data that refers to the whole invoice, while *fInvoiceItem* contains data referring to every single invoice item. ETL procedure fetches data from the e-invoice archive, cleans them and as such inserts into the data warehouse. It includes the initial data loading and the incremental data warehouse data loading. An additional web application for data warehouse data analysis has been developed and incorporated in the *FER e-invoice* system.

The data warehouse developed out of e-invoices in XML format is an important contribution because it allows efficient realization of the different queries about the business activities of an enterprise.

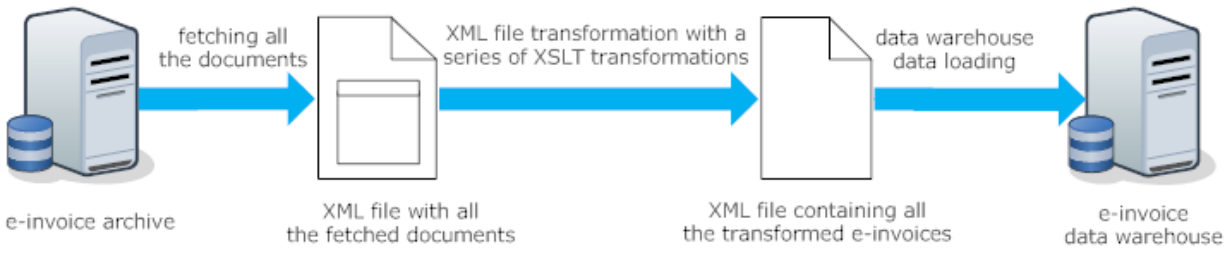

Fig. 5 ETL process

Fig. 6 Application for business activity analysis

REFFERENCES

[1] A. Bažant , Ž. Car, G. Gledec, D. Jevtić, G. Ježić, M. Kunštić, I. Lovrek, M. Matijašević, B. Mikac, Z. Skočir, *Telecommunications - Technology and Market (in Croatian: Telekomunikacije – tehnologija i tržište)*, Element, 1st edition, Zagreb, 2007.

[2] *Electronic trade law (in Croatian: Zakon o elektroničkoj trgovini)*, http://www.zakon.hr/z/199/Zakon-o-elektroni%C4%8Dkoj-trgovini, 4. 2. 2012.

[3] *Electronic document law (in Croatian: Zakon o elektroničkoj ispravi)*, http://www.zakon.hr/z/272/Zakon-o-elektroni%C4%8Dkoj-ispravi, 4. 2. 2012.

[4] *Electronic signature law (in Croatian: Zakon o elektroničkom potpisu)*, http://www.zakon.hr/z/211/Zakon-o-elektroni%C4%8Dkom-potpisu, 4. 2. 2012.

[5] *Accounting law (in Croatian:Zakon o računovodstvu)*, http://www.zakon.hr/z/118/Zakon-o-ra%C4%8Dunovodstvu, 4. 2. 2012.

[6] *Common tax law (in Croatian: Opći porezni zakon)*, http://www.zakon.hr/z/100/Op%C4%87i-porezni-zakon, 4. 2. 2012.

[7] *VAT regulation (in Croatian: Pravilnik o porezu na dodanu vrijednost)*, http://propisi.porezna-uprava.hr/index_open.asp?idPropisa=2201&jid=1&ime=Pravilnik%20o%20porezu%20na%20dodanu%20vrijednost, 4. 2. 2012.

[8] *Development Strategy of e-business in the Republic of Croatia for the perion 2007 – 2010 (in Croatian: Strategija razvitka e-poslovanja u Republici Hrvatskoj za razdoblje 2007. - 2010.)*, Ministry of Economy, Labor and Entrepreneurship, Zagreb, October 2007.

[9] *Council directive 2001/115/EC*, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:015:0024:0028:EN:PDF, 5. 2. 2012.

[10] M. Bačelić, E-invoice: The state is still losing because they do not change VAT regulation (in Croatian: *E-račun: Država i dalje gubi jer ne mijenja Pravilnik o PDV-u)*, magazine Lider, Zagreb, January 2011.

[11] Z. Skočir, I. Matasić, B. Vrdoljak, D*ata processing (in Croatian: Organizacija obrade podataka)*, 1st edition, Zagreb, 2007.

[12] W. H. Inmon, *Building the Data Warehouse, 4th edition, John Wiley & Sons, 2005.*

[13] M. Golfarelli, S. Rizzi, *Designing the data warehouse: key steps and crucial issues*, Journal of Computer Science and Information Management, 1999.

[14] R. Kimbal, T*he Data Warehouse Toolkit, 2nd edition, John Wiley & Sons, 2002.*

[15] Čičak, A., *E-Business interoperability,* diploma thesis, Zagreb, June 2011.

[16] Lažegić, I., *Using Data Warehouse for e-Business,* diploma thesis, Zagreb, June 2011.