IEEE

mipro 2016

ISSN 1847-3938

organizer upro

39th

international convention

May 30 - June 03, 2016, Opatija – Adriatic Coast, Croatia

Lampadem tradere

mipro - path to knowledge and innovation

mipro proceedings

MIPRO 2016

39th International Convention

May 30 – June 03, 2016
Opatija, Croatia

Proceedings

Conferences:

Microelectronics, Electronics and Electronic Technology /MEET

Distributed Computing, Visualization and Biomedical Engineering /DC VIS

Telecommunications & Information /CTI

Special Session on Future Networks and Services /FNS

Computers in Education /CE

Computers in Technical Systems /CTS

Intelligent Systems /CIS

Special Session on Biometrics & Forensics & De-Identification and Privacy Protection /BiForD

Information Systems Security /ISS

Business Intelligence Systems /miproBIS

Digital Economy and Government, Local Government, Public Services / DE-GLGPS

MIPRO Junior - Student Papers /SP

Edited by:
Petar Biljanović

International Program Committee

Petar Biljanović, General Chair, Croatia
S. Amon, Slovenia
V. Anđelić, Croatia
M.E. Auer, Austria
M. Baranović, Croatia
A. Badnjević, Bosnia and Herzegovina
B. Bebel, Poland
L. Bellatreche, France
E. Brenner, Austria
A. Budin, Croatia
Ž. Butković, Croatia
Ž. Car, Croatia
M. Colnarič, Slovenia
A. Cuzzocrea, Italy
M. Čičin-Šain, Croatia
M. Delimar, Croatia
T. Eavis, Canada
M. Ferrari, Italy
B. Fetaji, Macedonia
T. Galinac Grbac, Croatia
P. Garza, Italy
L. Gavrilovska, Macedonia
M. Golfarelli, Italy
S. Golubić, Croatia
F. Gregoretti, Italy
S. Groš, Croatia
N. Guid, Slovenia
Y. Guo, United Kingdom
J. Henno, Estonia
L. Hluchy, Slovakia
V. Hudek, Croatia
Ž. Hutinski, Croatia
M. Ivanda, Croatia
H. Jaakkola, Finland
L. Jelenković, Croatia
D. Jevtić, Croatia
R. Jones, Switzerland
P. Kacsuk, Hungary
A. Karaivanova, Bulgaria
M. Mauher, Croatia
I. Mekjavić, Slovenia
B. Mikac, Croatia
V. Milutinović, Serbia
V. Mrvoš, Croatia
J.F. Novak, Croatia
J. Pardillo, Spain
N. Pavešić, Slovenia
V. Peršić, Croatia
T. Pokrajčić, Croatia
S. Ribarić, Croatia
J. Rozman, Slovenia
K. Skala, Croatia
I. Sluganović, Croatia
V. Sruk, Croatia
U. Stanič, Slovenia
N. Stojadinović, Serbia
J. Sunde, Australia
A. Szabo, IEEE Croatia Section
L. Szirmay-Kalos, Hungary
D. Šarić, Croatia
D. Šimunić, Croatia
Z. Šimunić, Croatia
D. Škvorc, Croatia
A. Teixeira, Portugal
E. Tijan, Croatia
A.M. Tjoa, Austria
R. Trobec, Slovenia
S. Uran, Croatia
T. Vámos, Hungary
M. Varga, Croatia
M. Vidas-Bubanja, Serbia
B. Vrdoljak, Croatia
D. Zazula, Slovenia

Organization of tax data warehouse for legal entities

Miodrag Sretenović* Božidar Kovačić** Vladan Jovanović***
* MI&DA d.o.o., Karlovac, Hrvatska, m.sretenovic@mida.hr
** Odjel za informatiku, Sveučilište u Rijeci, Hrvatska, bkovacic@inf.uniri.hr
*** Georgia Southern University, Statesboro, USA, vladan@georgiasouthern.edu

Abstract - Legal entities information systems for tracking tax changes belong to the group of business office information systems, which provide answers to the user requirements on the monitoring of tax changes. By its structure, such information system represents only one segment of the overall legal entities information system, but despite this it has a central role in reporting to the Tax authorities and it represents the link between legal entities and the Tax authorities. This paper describes a model for storing tax data that has been implemented in practical business applications. Furthermore it describes the process of building and organization of the data warehouse and implemented ETL process for the extraction, cleansing and transformation of data in a multidimensional model of tax data warehouses, which is based on the metadata model and model of the set of legal rules. In addition it states a concept and basic methods used in the development of business intelligence systems for the construction and use of a warehouse and it gives guidelines for future development. Model warehouse tax information forms the basis for the execution of various analysis for decision making. Formalizing the process performance analysis directly affects the quality of business intelligence.

I. INTRODUCTION

Laws in the field of tax policy regulate the relationship between taxpayers and tax authorities implementing tax regulations. Due to increasing needs of public finance, tax laws are subject to frequent change. But despite these changes, such laws also require storing of historical tax data for several years.

A. Business Intelligence

Business intelligence is the process of collecting data and information from internal and external business environments and their transformation into business knowledge used for making business decisions[1].

B. Data Warehouse

Data warehouse is a segment of business intelligence, provided that the data within the warehouse are filtered. There are several ways of using filtered data. For example, by making inquiries that cannot be made in transactional databases because data are unstructured or semi-structured, or because inquiries could slow down the usual operation of databases. Data warehouse represents a central source of previously cleaned and transformed data, which are used for providing support in making business decisions[2].

Data warehouse is a set of procedures, technologies and tools that enable end users to perform data analyses, which help in the decision-making process and improve information resources[3].

Tax data warehouse represents a source of tax data that are filtered from the database by applying rules and processes based on tax legislation in a form suitable for analysis and creation of tax reports.

C. Data Mart

Data mart is a data warehouse that covers one segment of business, and includes data related to a single topic of business[4]. One of the leading experts in the field of application and implementation of data marts is Ralph Kimball, who founded Red Brick Systems in 1986[5].

Kimball's model [6] was used as a model for designing the model of tax data warehouse. In the dimensional approach of warehouse architecture, adopted by Kimball's model [6], data are divided into fact table, which is the largest one with respect to the amount of data; and this table is connected to one or more dimension tables, which contain attributes, and with the application of adequate tools, enables multidimensional view to the user, so we can call such structures multidimensional or "hypercube". Within the tax data warehouse, dimensions are independent by their structure, so they can be shown as an n-dimensional coordinate system.

By applying the relational structure of databases, access to data warehouse has been optimized, which results in a significantly improved performance with respect to the former technological level of databases.

D. Data Warehouse Architecture

Data warehouse architecture includes the process of defining data warehouse elements, methods for creating and connecting warehouse elements in accordance with set requirements. Data warehouse architecture represents a formal description of a data warehouse system[7].

The reasons that contributed to the architecture and development of data warehouses was the inability of databases to support complex requests for processing in a reliable and safe manner, and an increasing need for analysing a large amount of data and information.

This paper describes the development of tax data warehouses in chapter two, and presents the concept of tax data warehouses in chapter three. Chapter four contains description of a data warehouse model, the model implementation is presented in chapter five, and conclusion in chapter six.

II. DEVELOPMENT OF TAX DATA WAREHOUSE

Tax data warehouse must be able to quickly adapt to potential changes in legislation, so agile methodology is imperative when it comes to the development of data warehouse. Agile methodology applied in software development enables fast adjustment to changes in the environment[8].

Therefore, modelling requires:

- Conceptual modelling of tax data,
- Business process modelling according to ETL (Extraction, Transformation and Loading) process elements:
  - data extraction process,
  - data cleansing process,
  - data transformation process,
- Modelling of data marts within tax data warehouses,
- Enabling the presentation of data on prescribed tax forms, and enabling flexible access to all relevant data, in order to allow system users to further analyse tax changes.

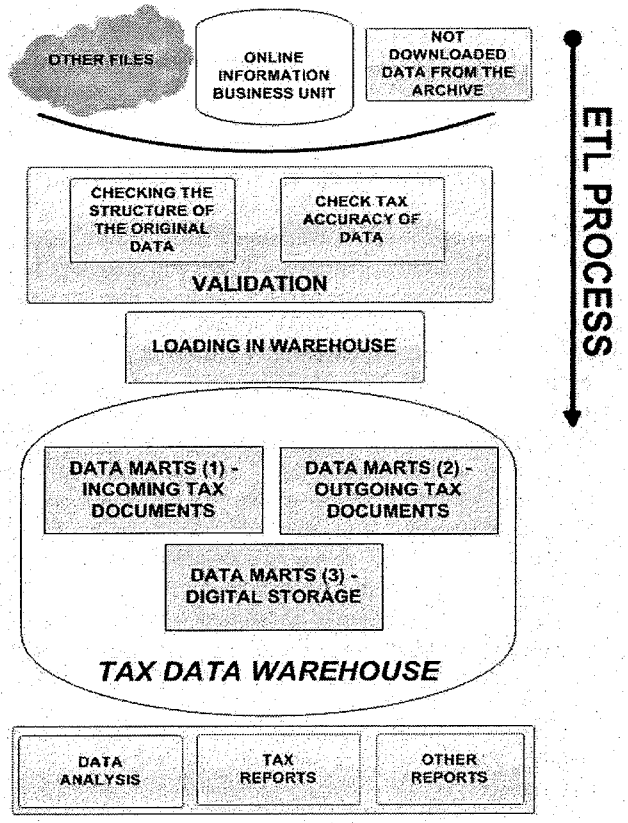

Figure 1. Shows the process of building a data warehouse

A. Conceptual Modelling of Tax Data

Conceptual modelling of data prescribed by tax legislation is a highly abstracted process of data modelling. The structure of data within the information system for tracking tax changes must reflect the structure of prescribed abstract reports.

The conceptual data model is a complete, consistent and concise description of data within an information system[9].

B. ETL Process Modelling

The process of building a tax data warehouse is shown in Figure 1 and starts by defining ETS processes (Extract, Transform, Load).

- The first step includes extracting the data from data sources. The most common data sources are relational databases, with the exception of tax data stored in various file formats generated from incompatible applications. In the process of data extraction, it is crucial to trace the source of data, date and time of their entry and data on entry operators, in order to enable complete insight to auditors. Data Vault concept emphasises the need to record where the data came from, as well as the time they were stored[10]. At the moment ETL tools pull data from relational databases, they pull metadata about their structure. Data structure in relational databases must reflect the final, abstract and prescribed form of report. Therefore, we may say that the tax data warehouse is based on the model of metadata and the model of prescribed legal rules.
- Transform means the transformation of data from the source to the warehouse. The preparations for loading data include an abstracted process of preparation and transformation of format and structure, validating the accuracy and quality of source data. The abstracted process within the tax data warehouse is conditioned primarily by regulatory nomenclature. Therefore, it requires an automatic process of cleansing and removing useless, incomplete or inaccurate data and returning them to the source for reprocessing.
- Load is a process of loading transformed data into the warehouse. These data will be suitable for further processing and enabling access and creation of prescribed tax documents, which is the ultimate aim of the user.

Legal regulations regarding the manner and structure of reporting to tax authorities directly prescribe the obligation to periodically load the data into the tax data warehouse, in intervals of at least 30 days.

C. Modelling of Data Marts within Tax Data Warehouse

Data marts within the tax data warehouse have specific structure, and they are optimized for the field of regulatory requirements for users. They are defined as abstracted reports and do not allow users much creativity or design.

Each tax data warehouse contains multiple data marts, which are coordinated and synchronised, and can be classified as supermarts according to their structure.

Supermarts consist of multiple mutually coordinated and synchronised tables. Table synchronisation is performed using foreign keys coming from conformed dimension. The structure of reporting the course of tax changes to tax authorities requires designing of at least three data marts, which are mutually connected by common dimensions and dimension attributes. Based on this fact, we have modelled the following:

1. data mart containing outgoing tax data or tax liabilities. Outgoing tax documents are invoices issued to customers for performed services or delivered goods.
2. data mart containing incoming tax data or tax receivables. Incoming tax documents are invoices received by suppliers for received services or goods.
3. data mart containing digital tax documents.Data mart containing digital tax documents includes: digital incoming invoices, digital outgoing invoices, digital documents by tax authorities, digital documents submitted to tax authorities, as well as any other tax-related digital documents.

Data marts have shared dimensions (Taxable period, Business units, Type of tax document, Entry operators, Invoice items, Tax partners, Date) and their own dimension attributes. Each dimension table has its surrogate key as a unique identifier of history of dimension data.

D. Modelling of Reports Within Tax Data Warehouse

Tax reports are abstracted reports that must contain elements prescribed by law. Since this is the ultimate aim of the tax data warehouse, it is necessary to perform optimization of all processes and thoroughly examine sources of data that will be used for modelling reports.

III. CONCEPT OF TAX DATA WAREHOUSE

Each change in tax regulations results in a change in the structure of databases at the source. Such changes increase requirements towards the data warehouse with respect to storing data and all previous changes, including the structure of data and metadata, covering a longer period of time. Figures 2, 3 and 4 show Entity-Relationship Diagram (ERD) of data marts, according to Martin Entity-Relationship Diagram[11].

Data Marts are used for the following purposes:

1. Data mart input tax data for the storage of incoming tax documents.
2. Data mart output tax data for the storage of output tax documents.
3. Data mart digital archive for the storage of digitized tax documents.

Entity is anything we wish to collect and store information about, entity-relationship method is a diagram of mutually connected groups of data within a system in question [12].

A. Defining the Structure of Data Marts Containing Incoming Tax Data

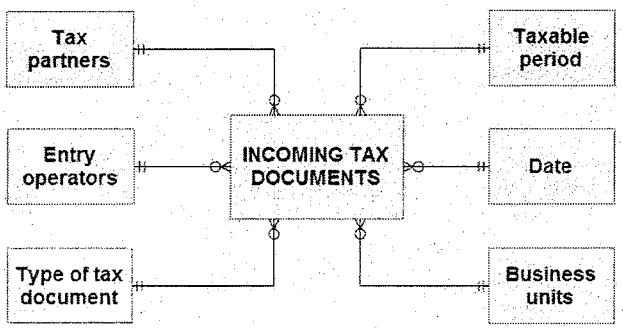

Figure 2. Data mart containing incoming tax data

Figure 2 shows the structure of a data mart containing incoming tax data. The structure of tax reports for the purpose of reporting the amounts and structure of tax receivables to tax authorities is defined by the data mart containing incoming tax data. Data mart structure diagram is shown in figure 2, where the entity INCOMING TAX DOCUMENTS is connected to other entities, which are partly shared with other data marts.

B. Defining the Structure of Data Marts Containing Outgoing Tax Data

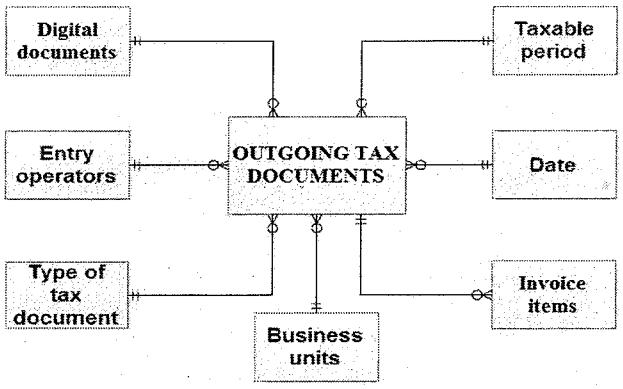

Figure 3. Data mart containing outgoing tax data

Figure 3 shows the structure of a data mart containing outgoing tax data. The structure of tax reports for the purpose of reporting the amounts and structure of tax liabilities to tax authorities is defined by the data mart containing outgoing tax data. Data mart structure diagram is shown in figure 3, where the entity OUTGOING TAX DOCUMENTS is connected to other entities, which are partly shared with other data marts.

C. Defining the Structure of Digital Storage Data Mart

Figure 4 shows the structure of digital storage data mart. Designing a digital storage data mart for storing tax documents in digital form enables fast and accurate access to digital tax documents. Figure 4 shows data mart structure diagram, where the entity DIGITAL STORAGE is connected to other entities, which are partly shared with other data marts.

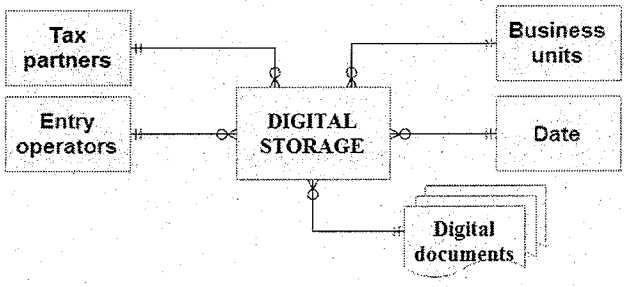

Figure 4. Digital storage data mart

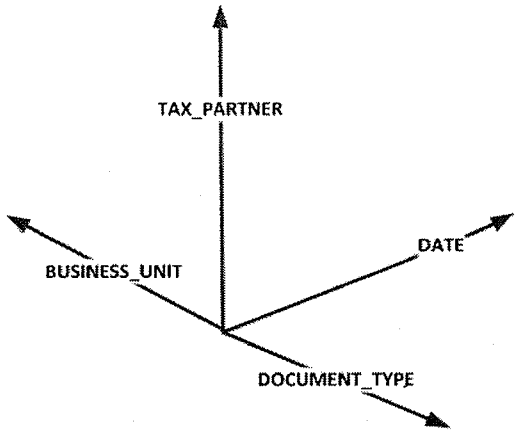

Figure 5. Four dimensional digital archives

The model allows you to perform analysis based on various criteria, for example in Figure 5.

The purpose of verification of tax data. The analysis presents a logical sequence of tax digitized documents in the preparation of binding documents. An analysis enables quality decision-making in the domain of tax activities, and also decided upon in terms of objection to the issued tax solutions.

IV. MODEL OF TAX DATA WAREHOUSE

Before building data marts implemented the ETL process modeling. The process of extraction is carried out from relational databases, with the exception of tax data stored in various file formats generated from incompatible applications (xml, txt, xlsx). Validation is performed by checking the data structures and check the correctness of the data. After validation was performed import data into the warehouse. Data warehouse consists of three data marts, and data marts are mutually connected by using shared dimension tables. Figure 6 shows the organisational scheme of tax data warehouse, with fact tables and dimension tables visible.

A. Defining Dimension Tables

Dimension tables used in the tax data warehouse, visible in figures 2, 3, 4 and 6 are as follows:

- *Taxable period* - data on taxable periods established by legislation, taxable period is usually one calendar month, three calendar months or a calendar year, depending on the status of tax payer and in accordance with the decision of tax authorities.
- *Type of tax documents* – data on defined types of tax records and reports in line with regulations,
- *Tax partners* – data on tax partners and their tax statuses, these are usually customers, suppliers and tax authorities,
- *Business units* – data on business units that created or received a document,
- *Date* – data about dates,
- *Entry operator* –data on operators authorized to enter, update or view data,
- *Invoice items* - data about items on issued invoices with the following elements-data: products or services, unit price, unit price in currency, official currency code, approved discount (rebate) and tax rate,

By using more dimensions with fact tables, we can model subspecialist views and have the opportunity to analyse business processes in detail.

B. Fact Tables OUTGOING TAX DOCUMENTS

This is a fact table with data about invoices issued to customers for performed services or delivered goods, as well as data on tax liabilities, which contains the following attributes:

- *Invoice_number* – Number of issued invoice,
- *Taxable_period_code* – Code of taxable period,
- *Tax_document_type_code* – Code determining the type of tax records,
- *Operator_code* – Code of the operator generating the data,
- *Business_unit_code* – Business unit,
- *Business_partner_tax_number* – Code of business partner,
- *Date_code* – Date,
- *Customer_invoice_amount* – Amount of receivables from a customer,
- *VAT_amount_5%_rate* – Amount of tax liability at a 5% rate,
- *VAT_amount_13%_rate* – Amount of tax liability at a 13% rate,
- *VAT_amount_25%_rate* – Amount of tax liability at a 25% rate,
- *Local_reverse_charge* – Amount of local reverse charge,
- *Delivery_of_goods_in_other_EU_members* – Amount of delivery of goods in other EU member states,
- *Delivery_of_goods_within_EU*– Amount of delivery of goods within EU,
- *Delivery_of_services_within_EU*– Amount of delivery of services within EU,
- *Services_provided_to_persons_without_residence_in_RoC* – Amount of services provided to persons without residence in the Republic of Croatia,
- *Goods_installed_or_assembled_in_other_EU_members* – Amount of goods installed and assembled in another EU member state,
- *Not_subject_to_taxation_within_EU* - Amount that is not subject to taxation in EU,
- *Not_subject_to_domestic_taxation* - Amount that is not subject to taxation in the Republic of Croatia,
- *Export_deliveries* - Amount of deliveries outside EU,
- *Other_exemptions* – Amount of other exemptions,

C. Fact Tables INCOMING TAX DOCUMENTS

This is a fact table with data on invoices received from suppliers for received goods or services, as well as data on tax receivables, which contains the following attributes:

- *Invoice_number* – Internal number of received invoice,
- *Taxable_period_code* – Code of taxable period,

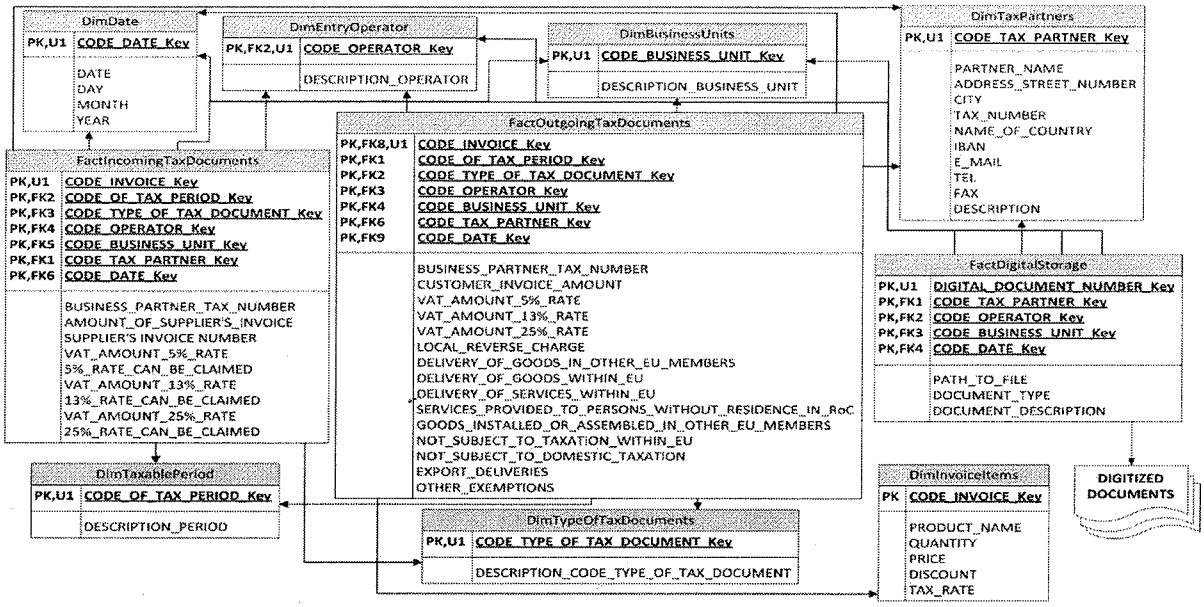

Figure 6. The organization of the warehouse tax data

- *Tax_document_type_code* – Code determining the type of tax records,
- *Operator_code* – code of the operator generating the data,
- *Business_unit_code* – Business unit,
- *Supplier's_invoice_number* – Invoice number of the supplier,
- *VAT_amount_5%_rate* – Amount of total tax shown on the invoice at a 5% rate,
- *5%_rate_can_be_claimed* - Amount of tax that is deductible and can be claimed at a 5% rate,
- *VAT_amount_13%_rate* – Amount of total tax shown on the invoice at a 13% rate,
- *13%_rate_can_be_claimed* - Amount of tax that is deductible and can be claimed at a 13% rate,
- *VAT_amount_25%_rate* – Amount of total tax shown on the invoice at a 25% rate,
- *25%_rate_can_be_claimed* - Amount of tax that is deductible and can be claimed at a 25% rate,

D. Fact Tables DIGITAL STORAGE

This is a fact table with data about digital tax documents, containing the following attributes:

- *Digital_document_number* – Represents the internal number of digital document,
- *Business_partner_tax_number* – Code of business partner
- *Operator_code* – Code of the operator generating the data,
- *Business_unit_code* – Business unit,
- *Date_code* – Date,
- *Path_to_file* – Represents the path to the location in the warehouse where the file with corresponding name of digital document is located,
- *Document_type* – Type of digital document (for example, incoming invoices, outgoing invoices, decision, appeal, tax return),
- *Document_description* – Short description of digital document.

- *Business_partner_tax_number* – Code of business partner,
- *Date_code* – Date,
- *Amount_of_supplier's_invoice* – Amount due to the supplier,

V. IMPLEMENTATION

The application for legal entity tax data warehousing has been created in Embarcadero studio architect technology using programming language Delphi XE5, and Embarcadero InterBase XE5 database technology.

Figure 7 shows one of the fundamental bookkeeping reports, outgoing invoices register which records all the invoices issued to customers, from which the following can be read:

- date of bookkeeping change that has occurred,
- data on customers receiving goods or services,
- data on the amount of receivables from customers,
- data on tax liabilities analytically shown according to tax rates, including tax bases shown according to the analytics of tax rates,
- data on tax exemptions analytically shown according to the type of tax exemption,

The report is created within a taxable period and is a result of data search from the fact table OUTGOING TAX DOCUMENTS, as well as corresponding dimension tables, which is shown in figures 3 and 6.

The incoming invoices register in figure 8 is a result of data search within the taxable period of the fact table INCOMING TAX DOCUMENTS, as well as corresponding dimension tables, which is shown in figures 2 and 6. Input accounts records all invoices from suppliers for goods or services, consisting of:

- date of bookkeeping change that has occurred,
- data on suppliers of goods or services,
- data on invoice number of the supplier,
- data on the amount due to the supplier,
- data on total taxes analytically shown according to tax rates, including tax bases shown according to the analytics of tax rates,
- data on deductible amounts of tax, and on tax that is not deductible, shown analytically according to tax rates,