

**University "Ss. Cyril and Methodius"
Faculty of Computer Science and Engineering**

**Proceedings of the Ninth International
Conference on Informatics and Information Technology**

CIIT 2012

April 19-22, 2012, Molika, Bitola, Macedonia

**Editors:
Verica Bakeva,
Dejan Gjorgjevikj**

Skopje, 2012

Preface

The Conference of Informatics and Information Technology was held for the ninth time, traditionally in Bitola, Macedonia during April 19-22, 2012. This year, for the first time it was organized by the Faculty of Computer Science and Engineering (FCSE). FCSE is the result of the unification of the two largest institutions in the area of informatics and computer technologies in Macedonia – the Institute of Informatics at Faculty of Natural Sciences and Mathematics and the Institute of Computer Techniques and Informatics at Faculty of Electrical Engineering and Information Technologies. The previous eight conferences were organized by the Institute of Informatics at the Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius in Skopje. Now, FCSE continues the tradition of giving the opportunity to researchers to present their latest results in the field of Informatics and Information Technologies.

During the three days of the conference 81 presentations were given in 13 regular sessions. One project meeting and special student session were also held. On the student session 15 student projects were presented and the best one (chosen from the student participants) was awarded.

Professor Vedran Mornar from Faculty of Electrical Engineering and Computing in Zagreb gave the invited lecture “State Matura and National Information System for Application to Higher Education Institutions”.

The rich variety of topics covered by the presentations provided a setting for numerous fruitful discussions on different concepts, methods and technologies for the benefit of advancing these research areas.

As editors we hope that the CiiT conference will continue its growth toward becoming an influential international conference with great impact to ICT research and development.

The Editors,
Verica Bakeva, PhD
Dejan Gjorgjevikj, PhD

Editors

Verica Bakeva, PhD

Dejan Gjorgjevikj, PhD

Organizing Committee

Nevena Ackovska, PhD

Vesna Dimitrova, PhD

Dejan Gjorgjevikj, PhD

Vladimir Trajkovikj, PhD

Verica Bakeva, PhD

Mile Jovanov, MSc

Magdalena Kostoska, MSc

Program committee

Smile Markovski, PhD
Dragan Mihajlov, PhD
Margita Kon-Popovska, PhD
Vancho Kusakatov, PhD
Lupcho Kocarev, PhD
Marjan Gushev, PhD
Katerna Zdravkova, PhD
Suzana Loshkovska, PhD
Zhaneta Popeska, PhD
Kosta Mitreski, PhD
Verica Bakeva, PhD
Vladimir Trajkovikj, PhD
Ana Madevska Bogdanova, PhD
Dejan Gjorgjevikj, PhD
Dimitar Trajanov, PhD
Sonja Gievska, PhD
Andrea Kulakov, PhD
Ljupcho Antovski, PhD
Slobodan Kalajdzhiski, PhD
Nevena Ackovska, PhD
Marija Mihova, PhD
Sonja Filiposka, PhD
Anastas Mishev, PhD
Ivan Chorbev, PhD
Goce Armenski, PhD
Dejan Spasov, PhD
Goran Velinov, PhD
Vesna Dimitrova, PhD
Boro Jakimovski, PhD
Lasko Basnarkov, PhD
Igor Trajkovski, PhD
Ivica Dimitrovski, PhD
Aleksandra Mileva, PhD
Zoran Ivanovski, PhD
Pece Mitrevski, PhD
Azir Aliu, PhD
Sime Arsenovski, PhD
Adrijan Bozinovski, PhD

List of Participants

Adam Markovski
Ahmad Zakeri
Aleksandar Bahovski
Aleksandar Tenev
Aleksandra Kanevce
Aleksandra Mileva
Aleksandra Popovska-Mitrovikj
Ana Madevska Bogdanova
Andrea Kulakov
Andreja Naumoski
Biljana Spireva
Blagojce Jankulovski
Bojan Kostadinov
Bojana Koteska
Boro Jakimovski
Damjan Temelkovski
Dejan Gjogjevikj
Dejan Spasov
Dimitar Trajanov
Dragan Jakimovski
Dragan Sahnaski
Dragi Zlatkovski
Dushan Novak
Eftim Zdravevski
Elena Janevska
Elena Vlahu-Gjorgievska
Emil Stankov
Georgina Mirceva
Gjorgji Madjarov
Gjorgji Mikarovski
Goce Armenski
Goce Gavrilov
Goran Bakraceski
Goran Dimitrovski
Goran Velkoski
Hristina Mihajloska
Igor Kulev
Igor Mishkovski
Ilija Hristoski
Ilinka Ivanoska
Ivan Chorbev
Ivan Kitanovski
Ivica Dimitrovski
Jane Jovanovski
Katarina Trojancanec
Kire Trivodaliev
Kiril Kirovski
Kosta Budimovski
Kosta Mitreski
Koste Budinoski
Kristina Spirovska
Ljupco Jovanoski
Ljupco Kocarev
Magdalena Kostoska
Maja Ivanova
Maja Siljanoska
Marija Alagjozovska
Marija Mihova
Marija Petkovska
Marija Stankova
Marina Ivanova
Marina Vasileva
Marjan Efremov
Marjan Gushev
Marko Vučković
Martin Mitrevski
Mende Sugarevski
Metodija Angjelkoski
Mile Jovanov
Milos Jovanovik
Nataša Šuteva
Nebojša Škrbina
Nenad Bojkovski
Nevena Ackovska
Nikola Koteli
Nino Karas
Ognjen Ognenoski
Oliver L. Iliev
Pance Ribarski
Pavle Sazdov
Petar Ristoski
Petre Lameski
Renata Petrevska–Neckoska
Riste Stojanov
Sanja Stefanova
Sashko Ristov
Saso Koceski
Silvana Krsteva
Sime Arsenovski
Slavcho Chungurski
Slobodan Kalajdziski
Smile Markovski
Stanislava Stoilova
Stefan Mitev
Stefan Spasovski
Stojance Spasov

Suzana Loshkovska
Tomche Delev
Tome Dimovski
Vancho Kusakatov
Vasil Grozdanoski
Vassil Grozdanov
Velimir Graorkoski
Verche Cvetanoska
Verica Bakeva
Vesna Dimitrievska Ristovska

Vesna Dimitrova
Vesna Kirandziska
Vladimir Apostolski
Vladimir Trajkovik
Vladimir Zdraveski
Vladislav Bidikov
Zlatka Trajcheska
Zoran Kotevski
Zoran Zdravev
Zorica Dimishkova

PROCESS OF CREATING AND USING DATA WAREHOUSE IN A WHOLESALE

Zoran Zdravev, PhD
 University “Goce Delcev”
 Faculty of computer science
 Stip, Macedonia

Cveta Martinovska, PhD
 University “Goce Delcev”
 Faculty of computer science
 Stip, Macedonia

Goran Vitanov
 University “Goce Delcev”
 Faculty of computer science
 Stip, Macedonia

ABSTRACT

This paper presents the process of creating data, information, and knowledge through a real live database example. It describes the On Line Transaction Processing and its purpose. In addition, it clarifies the lifecycle of a data warehouse: how to create it from heterogeneous data sources and synchronize it. After this phase, also known as collecting experience from the past, the next step is to transform this experience into knowledge. The goal is to identify and explain problems step by step.

I. INTRODUCTION

According to the practice, business applications in today’s corporate world are created to gather data in many forms. Data is closely associated with functional process of the organization and can originate from multiple sources, inputs, and systems. Customers, invoices, orders, employees, and manufacturing data are some examples of collecting, processing, storing, and accumulating an extensive amount of data. Storing data in databases is used in day to day operations with no additional value. Those systems are designed for OLTP (On Line Transaction Processing), to handle a large number of on-line transactions (INSERT, UPDATE and DELETE). OLTP systems perform processing records and maintaining data integrity in multi-access environments. Their performance is measured in a number of transactions per second. In our case we have OLTP database, which is supporting 100 concurrent users and its structure and function will be further explained. This step completes the data collection.

The performance of a business application is of high importance. Its performance is impacted when the system is overflowed with reports. This is the first signal to separate the reporting from the transaction workload. In addition to keeping transactional data into different database the application usually enables the organization to consolidate data from several sources. Data warehouse is a relational database, but it is designed for analysis and query rather than transactional processing. Its central integrated data storage is designed for reporting and maintaining history. According to the situations a data warehouse is created to keep the data from the last four years, and it is updated with new data once a day at 01:00AM. The data warehouse has ETL (extraction, transportation, transformation and loading) solution, OLAP (On Line Analytical Processing) and data mining capabilities. This step concludes the memory creation and denormalization

of the database structure, which is a model for a business language.

Table 1: OLTP vs. OLAP database system

	OLTP	OLAP
	(Operational System)	(Data Warehouse)
Data source	Operational data	Data comes from various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and decision support
Data Contents	Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries Return relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved; batch data refreshes and complex queries may take many hours; Query speed can be improved by creating indexes
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; It requires more indexes than OLTP
Database Design	Highly normalized with many tables	Typically denormalized with fewer tables

Separating and regrouping the information allows seeing a trend, exceptions, patterns, and relationships. This enables to analyze multidimensional data from multiple perspectives using consolidation, drill-down, slicing and dicing. With OLAP cubes data is published in a user friendly form, already aggregated and computed in multiple dimensions. For business users an OLAP cube presents an analytical tool, but not knowledge. This step is usually called “gathering experience from the past”. This is enabled through Microsoft Analysis Services (being the engine) as the back end and Microsoft Excel (being the interface) as the front end.

The next question after data collection and using OLAP cubes as analytical tool is if one can further analyze the data and what techniques can be used. Mathematical statistics methods can be used to identify hidden trends and unusual patterns into data in order to predict the future. This technique is called data mining and its main purpose is to extract previously unknown patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining), through automatic or semi-automatic analysis of large datasets. For this purpose it’s used in Microsoft Data Mining tools.

II. CREATING DATA AND INFORMATION

The first goal is to create OLTP database in order to support collecting, fast processing and storing of the data. For that purpose is created a transactional database in SQL Server 2008 with 167 tables. To simplify the work, we are going to analyze the tables which are related to sales. In figure 1 is presented database diagram, and in figure 2 are described tables and its purpose.

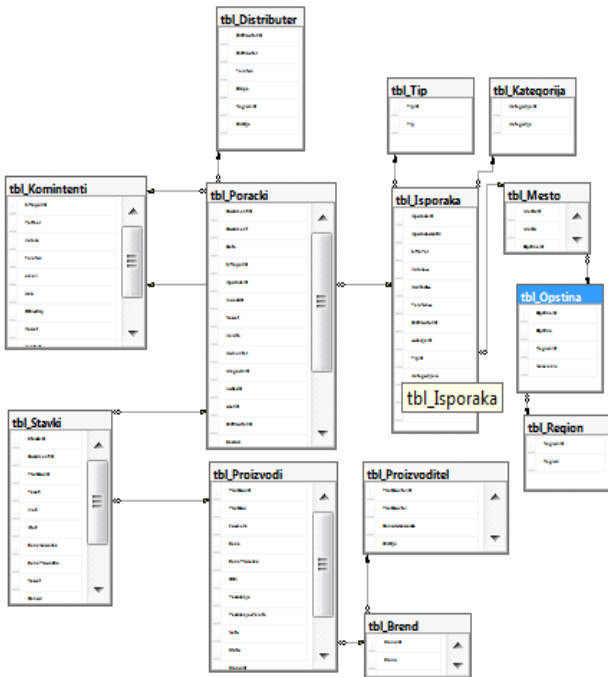


Figure 1: OLTP Database diagram.

Tables	Data stored
tbl_Poracki tbl_Stavki	Master and detail data for sales
tbl_Komintenti tbl_Isporaka	Customers and place of delivery
tbl_Tip, tbl_Kategorija	Customers categorized
tbl_Meto, tbl_Opstina tbl_Regioni	Geographical data
tbl_Proizvodi tbl_Brend tbl_Proizvoditel	Products and product classification
Tbl_Distributer	Details related to sales people

Figure 2: Tables and data description.

The entire data is associated with the functional processes of the company. In this case the extensive amount of sales, customers, categorization, and products related data is collected, stored and processed. The process is going through the following steps: first the operator is creating invoice, adding customer code, date of payment, searching products by code and adding it, second the system is adding the price of the product automatically, adding quantity and discount and finally the system is calculating the amount. With this step the system has collected the following data: DocumentID, Document, Data, SifraparID, IsporakaId, Rabat, Valuta etc. Figure 4 presents sample data from the system.

StavkaID	DokumentID	ProizvodID	Paketi	Vlez	Izlez	CenaNabavna	CenaProdazna	Rabat	Danok	Iznos	UserID
1	1	1	10	0	200	18	20	0	5	4000	1
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

DokumentID	Dokument	Data	SifraparID	IsporakaID	IndeksiID	Rabat	Valuta	Komentar	MaqacinID	VozaciID
1	1	2012-03-25 ...	1	1	1	0	15		1	15
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Figure 3: Data sample.

As the number of transactions is increased, the “wealth of data” is growing. Individually all this data is stored in different tables. The meaning of data SifraparID or IsporakaId doesn’t mean anything unless they are presented in conjunction with other data from related tables. According to this conclusion the next step is to create a query, extract data from the database and provide the users with reporting capabilities, or simply turn the data into information. For example, the following queries are created:

- Sales by customers (top customers)
- Warehouse stock
- Best-selling products
- Top selling brand
- Top selling region etc.

```
SELECT TOP (60) PERCENT dbo.dim_ProizvodKategorija.Proizvo
SUM(ROUND(dbo.dim_Prodazba.Amount, 0)) AS Износ
FROM dbo.dim_Prodazba INNER JOIN
dbo.dim_ProizvodKategorija ON dbo.dim_Prodazba.P
GROUP BY dbo.dim_ProizvodKategorija.Proizvoditel
ORDER BY Износ DESC
```

	Категорија на производи	Количина	Износ	Нето
▶	ПИЈАЛОЦИ	15346979	967759465	11917
	ЛЕБ	1508843	570928895	84416
	ГРИЦКИ	6575967	393634456	75666
	АЛКОХОЛНИ ПИЈАЛОЦИ	366273	304521826	21162
	ДЕТЕРГЕНТИ	902114	125780433	49840
	СРЕДСТВА ЗА САДОВИ	891116	105228105	12893
	КОНЗЕРВИРАН ЗЕЛЕНЧУК	1376452	95892462	67836
	КЕЧАП	207558	91420986	38681
	МЛЕЧНИ ПРОИЗВОДИ	415551	67390594	75994
	ДЕЗОДЕРАНСИ	310155	62268099	82094

Figure 4: Turning the data into information

By collecting the data and placing it in a context that produces meaning, the business application now has an ad hoc query and reporting capabilities and provide the users with the ability to rise up from the data layer and create information.

III. CREATING A DATA WAREHOUSE

Increased number of transactions, included ad hock query and reporting, new ideas to analyze information and determine relationships, patterns, trends and exceptions, resulted in a decreased performance of the database, the next step is to separate the OLTP database and create a data warehouse. In order to create data warehouse the following questions should be answered:

- What tables need to be created
- How to update the tables with the new data
- When to make updates

A. New tables and data

The point of interest is to analyze sales, so in data warehouse the following tables are created:

Table 2: Data Warehouse tables

Data Warehouse Tables	Data stored / (Tables used)
dim_Prodazba	Aggregated sales data
	(tbl_Poracki, tbl_Stavki)
dim_Kupuvaci	Customer dimension
	(tbl_Komintenti, tbl_Isporaka)

dim_KupuvaciKategorija	Customer categorized dimension (tbl_Komintenti, tbl_Isporaka, tbl_Tip, tbl_Kategorija)
dim_ProdaznaTeritorija	Geographical dimension (tbl_Mesto, tbl_Opstina, tbl_Regioni)
dim_ProizvodKategorija	Product dimension (tbl_Proizvodi, tbl_Brend, tbl_Proizvoditel)
dim_Vreme	Date, month, year dimension (tbl_Poracka)

Why the table dim_Prodazba is created? As is shown Figure 6, in the transactional database sales, the data is stored in two tables. Since fields Zakluci, Rabat, Userid, Danok, Vlez etc. are not related to sales, the first step is to select related fields. Second step is to create calculated fields and make aggregations. Those fields are: Data, IsporakaID, DistributerID, VozacId, ProizvodID, Sum(Izlez) as Quantity, Sum(Izlez) as Amount.

According to the selected fields from transactional database, the new table dim_Prodazba has the following fields: Data, IsporakaId, DistributerID, VozacID, ProizvodID, Quantity and Amount. Following this logic, the rest of the data warehouse tables are created.

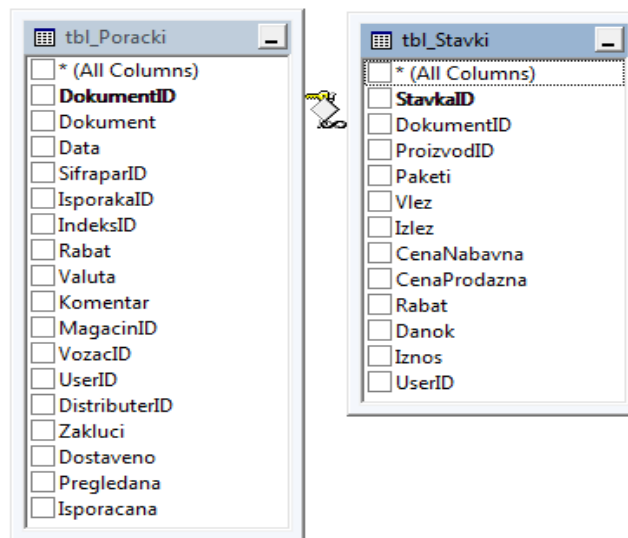


Figure 6: Transactional database tables

B. How and when to update the tables

Once the tables are created, it is necessary to load them with data, process known ETL. It can be completed using Microsoft SSIS (solution for automating data movements).

Here the job is divided in two parts. The first part is to insert data from the past years. The second part is to update

the tables with data from the current year. The process for both is same with one difference. The data from the current year needs to be updated on a daily base because of ongoing updates, deletes and inserts.

For example the process of ETL on dim_Prodazba is the following:

- Using SQL code to build query and to extract data from the source tables:

```
SELECT TOP (100) PERCENT dbo.tbl_Poracki.Data,
dbo.tbl_Poracki.IsporakaID, dbo.tbl_Poracki.DistributerID,
dbo.tbl_Poracki.VozacID, dbo.tbl_Stavki.ProizvodID,
SUM(dbo.tbl_Stavki.Izlez) AS Kolicina,
SUM(dbo.tbl_Stavki.Iznos) AS Iznos FROM
dbo.tbl_Poracki INNER JOIN dbo.tbl_Stavki ON
dbo.tbl_Poracki.DokumentID = dbo.tbl_Stavki.DokumentID
WHERE (dbo.tbl_Poracki.Zakluci IS NULL)
GROUP BY dbo.tbl_Poracki.Data,
dbo.tbl_Poracki.IsporakaID, dbo.tbl_Poracki.DistributerID,
dbo.tbl_Poracki.VozacID, dbo.tbl_Stavki.ProizvodID
ORDER BY dbo.tbl_Poracki.Data
```

- Define the destination table (dim_Prodazba) and fields where to insert the data:

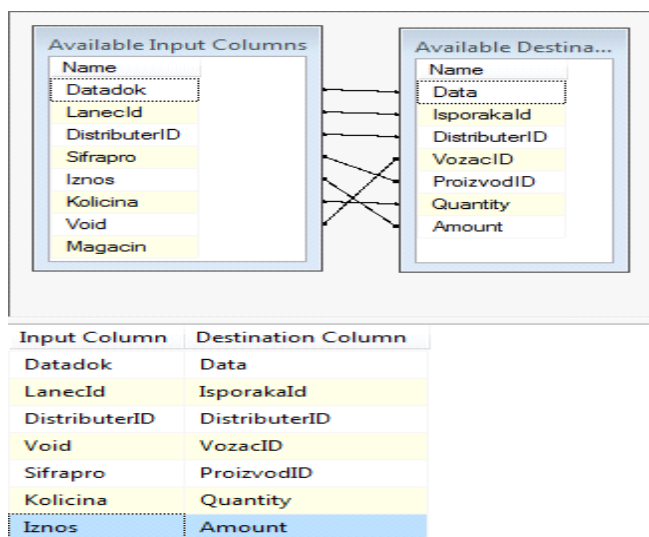


Figure 7: Source and destination tables and fields

Since the source and the destination tables and fields are defined, the system knows how to load the Data Warehouse. This process is known as Data Flow Task (DFT), and group of DFL's creates the ETL process.

- The last step is scheduling the package to run in the exact period and time. Scheduling is once a day at 1:00AM.

Following the logic of source>destination>schedule operation, the process of extraction, transportation, transformation and loading of the data in the data warehouse is fully automated.

IV. CREATING AND USING OLAP

OLAP is a technology used to organize large database into cubes. Cubes are organized to fit the way of retrieving and analyzing data so it's easy to use it in reports. Cubes summarize numerical data of measures Kolicina, Iznos into hieratical dimensions as Proizvodi, Prodazna_Teritorija, Vreme, KupuvaciKategorija etc. Since the Data Warehouse is loaded with data, and cube is created we can start analyzing and answering business related questions. First it is necessary to have the following information. For example:

- Are there differences between products and category sales in the regions?
- Is there any potential to improve sales in different category of customers?
- When, where, and on what brands to lunch an advertising campaign?

For this purpose, the created cube is imported into Microsoft Excel so the business user can start drill and slice the data in user friendly environment. To answer the first question the user has to select dimensions Prodazna_Teritorija and Vreme in Column Labels and dimension Proizvodi in Row Labels. The measure Amount is in Values. So we have the following situation:

	A	B	C	D
1	Amount	Column Labels		
2		2011		
3		Запад		
4	Row Labels	Гостивар	Скопје	Тетово
5	АЛКОХОЛНИ ПИЈАЛОЦИ	4,047,136	97,753,272	16,185,650
6	ВИНО	276,906	4,364,095	585,913
7	ВИСКИ		1,304	
8	ЛИКЕР	51,706	1,977,330	
9	МАСТИКА		1,349,104	430,063
10	ПИВО	3,140,613	82,549,232	13,866,549
11	АМСТЕЛ 0.5Л		80,880	18,390
12	БИТОЛСК 1.5Л	128,322	8,845,466	1,459,889
13	ГОРСКО 1.5Л	2,119	64,973	33,020
14	ЗЛАТЕН ДАБ 0.5Л		42,639	27,604
15	КАМЕНИЦА 0.5Л	617,378	28,913,350	3,639,105
16	КЕНБАХ 0.5Л		270,578	6,090
17	КРАЛИ МАРКО 1.5Л	41,627	6,223,827	663,190
18	СКОПСКО 0.5Л	2,347,813	37,759,476	7,977,434
19	ТУБОРГ 0.33Л	3,355	103,352	32,253
20	ХАЈНЕКЕН 0.33Л		244,706	9,574
21	УЗО	577,912	7,508,826	1,303,125
22	ШТОК		3,385	
23	БОЈА ЗА КОСА	381,701	24,781,464	1,623,365
24	БОНБОНИ			
25	ГРИЦКИ	18,785,114	174,225,168	17,599,244

Figure 8: Drill and slice data through Pivot table.

From here the user can gather experience from the past in order to recognize patterns and make decisions. The navigation is very simple through expand (+) and collapse (-). There is no need for additional calculations or any additional programming knowledge.

V. KNOWLEDGE DISCOVERY

The definition of data mining or Knowledge Discovery refers to the process of analyzing a given data set from different precepts and scenarios in order to discover patterns in a given data set. It is a process of estimation in unknown situations. It can help answer questions such as: "what will be our sales in different sales channels?"

To answer this question is used data modeling – estimate technique in Microsoft Excel. This tool builds estimation model to predict continues values on one column based on the values from other column.

The process is following:

- Create a connection to the data source. It is a connection to Data Warehouse and a query which retrieve the fields: Region, Opstina, Kategorija, Tip, Sum(Amount), Sum(Quantity).
- Use the Estimate technique by selecting the fields which are of interest.
- Analyze the results.

Analyzing the data in order to make business decision is not a subject in this paper.

VI. CONCLUSION

This paper discussed the process of creating the data, information and knowledge. One practical example was used to explain a real life situation in order to implement our knowledge step by step. The point of this work was not only to show the technique of creating the system, but also to put focus on the great importance of using advanced techniques such as OLAP and prediction analysis.

The system that was created is implemented in business environment and it's subject to further research and improvements. Since the Data Warehouse exists our further work will be focused on prediction analysis techniques. Using prediction analysis, through real life business situation we plan to predict the future and make strategic decision. The goal is to describe the entire process of prediction analysis, and how it helps the company to make new sales strategy.

VII. REFERENCES

- [1] S. Q. M. Mark Whitehorn and M. Keith Burns, "Microsoft Corporation," 07 2008. [Online]. Available: [http://technet.microsoft.com/en-us/library/cc719165\(v=sql.100\).aspx](http://technet.microsoft.com/en-us/library/cc719165(v=sql.100).aspx). [Accessed 27 03 2012].
- [2] G. S. L. Michael J.A. Berry, Data Mining Techniques For Marketing, Sales, and Customer Relationship Management, Wiley Publishing, Inc., Indianapolis, Indiana, 2004.
- [3] "The Transition of Data into Wisdom," 20 03 2012. [Online]. Available: <http://www.information-management.com/news/2784-1.html>.
- [4] D. C. A. M. Gene Bellinger, "Data, Information, Knowledge, and Wisdom," [Online]. Available: <http://www.systems-thinking.org/dikw/dikw.htm>. [Accessed 16 03 2012].
- [5] Dijcks, Jean Pierre, "Oracle: Big Data for the Enterprise," Oracle, 2012.
- [6]"OLTP vs. OLAP," [Online]. Available: <http://datawarehouse4u.info/OLTP-vs-OLAP.html>. [Accessed 12 03 2012].