

Using Data Warehouse for ABA Basketball League Statistics

Z. Kazi, M. Stasevic, B. Radulovic, B. Markoski and Lj. Kazi

University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
zoltan.kazi@tfzr.rs, stasevicmiljan23@gmail.com, biljana.radulovic@tfzr.rs, markonins@yahoo.com, ljubica.kazi@tfzr.rs

Abstract - This paper shows how data warehouse can be used for creating statistics for regional basketball league. Data for analyses were collected from Internet site of ABA league and transformed into a relational database, in a form suitable for processing in data warehouse tool. Two OLAP cubes were projected and created in order to generate analytical data for reporting.

I. INTRODUCTION

Basketball is one of the most popular sports in the world [1]. The success of basketball team in competition during seasons depends on the individual ability of each player, on a good basketball coach, but also on a good statistics and data for training process and for preparing tactics for every match in season. Also, there are many interested potential users of the sport data sets like various academia, governments and authorities, sport governing bodies [2].

Therefore, it is very important the use the data analysis for basketball coach's activities and job. So, the coach must use the data processing model and data analysis technology, which provide the coach with scientific basis and help him make the specific decisions to win the final competition [1].

Data warehouse, as a well-known concept for creating various reports and extracting analytical data from database, could be used as a support system for making decisions in basketball team management.

II. DATA WAREHOUSE CONCEPT

Decision makers in organizations make their decisions on the basis of analysis of facts obtained from various processes and data sources. These facts are stored mostly in OLTP (on-line transaction processing) systems as well as some other data sources. Purpose of data warehouse system is to integrate the data obtained from existing OLTP and other data sources, conduct computation that enable performing analysis for decision making support [3].

Data warehouse systems are based on-online analytical processing (OLAP), which is intended for interactive analysis and reporting. OLAP systems are based on multidimensionality of denormalized data [4]. OLAP architecture requires data to be transferred to

multidimensional structure called a cube, by extracting and transferring data, creating aggregated data, dimensions and measures [5]. Processing OLAP data cubes is a time-consuming process depending on OLAP server and data warehouse tools performances [3].

The main goal of data warehousing is data collecting and distribution of information throughout the organisation and use of any information, anywhere, at any time due to the realization of the principle of "always be at the service for information users". One of the main data warehouse building goals is not only to store data, but also to enable managers to conduct data analysis. Decision makers in organizations are often under pressure, because they must make their decisions on the basis of analysis of current facts obtained from various business situations, processes and data sources. These facts are stored mostly in OLTP systems as well as some the data sources. Integrated obtaining data for analysis is not very easy to perform.

Purpose of data warehouse system is to transform the data obtained from existing OLTP system into a form suitable for processing which enables performing analysis with tools for business decision making support. End user of data warehouse system needs are met by enabling functionality such as using large amount of data from the company business processes for analysis, setting questions and getting answers about different business issues and diverse reporting. Data warehouse systems are based on OLAP, which is intended for interactive analysis and reporting, as opposed to the production system designed for data update and transaction processing – OLTP systems. OLAP systems like data warehouses are using multidimensionality principle of de-normalized data. Basic elements of OLAP systems are: Database – provides data for analysis, OLAP server – process and manipulate data, Interface system – for interaction with a user and other applications and administrative tools [6].

III. CREATING OLAP CUBES

Data used for analyses were collected from Internet site of ABA basketball regional league and transformed into a relational database [7]. In our case, we will conduct an analysis of the success achieved by players and teams in the ABA league in season 2013-2014. The second transformation of this data from existing OLTP DBMS

system was into a form suitable for processing which enables performing analysis with data warehouse tool.

This database with basketball data was made up of eight tables: city, player, team, league, season, team statistics, and individual player statistics per season and association that connects teams and league (figure 1). The database is designed so that the team achieves and the results of playing games are in the form of statistics within one season in the league. The player plays for the proper club (team) placed in a city and achieved results during the season are the data that was entered in the statistics table for each individual. This data are: total number of points, assists, offensive rebounds, defensive rebounds, blocks, errors, number of made and attempts for one, two and three points, and minutes made in the game.

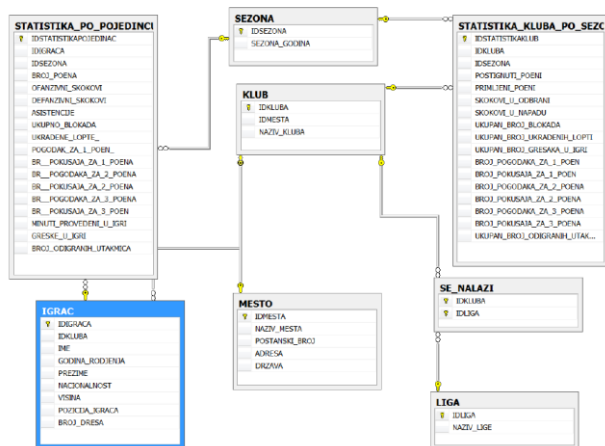


Figure 1. Relational database schema

Olap cubes have been created in the tool for creating data warehouse Microsoft Business Intelligence Development Studio (BIDS) which is based on transactional and relational Microsoft SQL Server database management studio (DBMS). A new project created in BIDS is analysis services type of project. This project has a data source that connects BIDS with relational database with basketball data. After creating a data source view, user must define which part of database will be analyzed and has to define data source view. Our database is a small database with only eight tables, so the created data source views include all tables. In a greater database with hundreds or thousands of tables this would not be practical.

The next step in designing the olap cube is to define measures and dimensions.

Physical architecture of dimensional model in OLAP can be described by a scheme that is defined by the stars or snowflakes from two types of tables - dimensional tables and fact tables that contain measures. Fact tables contain quantitative data that users want to analyze. These data are usually numeric type and can consist of up to several million rows and columns. Dimensional tables are much smaller and contain information describing data on which the analysis is performed. These data are called attributes.

The basic feature of the star scheme is that the dimensional tables are de-normalized. This approach uses

data in DBMS to simplify the repetitive design and performance. De-normalization is the process of combining tables to improve performance. This procedure reduces the number of connections that must be processed by entering queries. This has a direct impact on improving system performance, because the lower the number of connections, the system is faster information requested.

Standard dimensions of the cube projected with star schema are:

- team - team name,
- season - season year,
- player – first name,
- player – last name.

Measures for this cube are all defined for team:

- one point attempts,
- one point made,
- two points attempts,
- two points made,
- three points attempts,
- three points made,
- total number of scored points,
- total number of accepted points,
- played games, offensive rebounds,
- defensive rebounds,
- number of errors,
- number of steals.

Standard dimensions of the cube projected with snowflake schema are:

- team - team name,
- player – first name,
- player – last name,
- player – team name,
- player – height,
- player – nationality,
- player – birth year,
- city – name,
- city – country,
- season – season year.

Measures for this cube are defined for single player during the whole season:

- two points attempts,
- two points made,
- three points attempts,

- three points made,
- number of assists,
- played games,
- minutes played per game,
- total number of points.

For the cube based on snowflake schema (figure 2) we had three computable fields: player percentage for one point (1%), player percentage for two points (2%), player percentage for three points (3%) and they are obtained by dividing the number of hits with the number of attempts for one, two and three points.

For the cube based on star schema we had two computable fields: player average minutes per game, average points per game and they are obtained by dividing the points scored by the number of games played and total minutes played with the number of matches.

As we mentioned earlier two OLAP cubes are projected, the first cube presents statistics clubs achieved in a particular season (star schema). The second cube presents an analysis of players with most scored points in a particular season and the league (snowflake schema).

There are two ways for creating cubes. The first method is to create an empty cube, and then subsequently add the dimensions and measures. Another way is

through the wizard within the steps to choose the dimensions and measures.

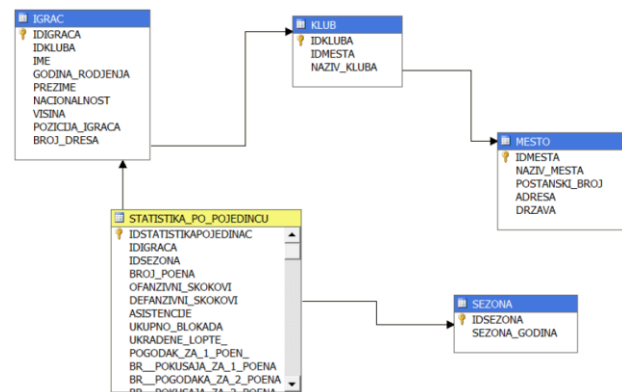


Figure 2. OLAP cube snowflake schema

The next step in using data warehouse tool is a cube creation. User can choose to create an empty cube, to start the wizard or to activate a template. We used a first solution to create an empty cube. The next step is to choose the appropriate data source view. In cube editor were added previously created dimensions, measures with aggregate functions (sum, count, count of rows, minimum, maximum, etc.) are defined as it is described in paper earlier, and finally the calculated fields were defined in the last step.

| IMSEZONA | | Drop Column Fields Here | | | | | | | | | | | | | |
|---------------|------------|-------------------------|-----------|------------|-------------------|-------|------------------|------------------|------------------|------------------|------------------|------------------|--------|-------------------|-------------------|
| IME | PREZIME | POZICIJA_IG | NAZIV_KLU | BROJ_POENA | Prosečno poena po | ASIST | BR_POGODAKA_ZA_1 | BR_POKUSAJA_ZA_1 | BR_POGODAKA_ZA_2 | BR_POKUSAJA_ZA_2 | BR_POGODAKA_ZA_3 | BR_POKUSAJA_ZA_3 | MINUTI | BROJ_ODIGRANIH_UT | Prosečno minuta p |
| Bogdan | Bogdanovic | SG | Partizan | 374 | 13.85 | 94 | 76 | 168 | 55 | 162 | 776 | 27 | 29 | | |
| | | | | 374 | 13.85 | 94 | 76 | 168 | 55 | 162 | 776 | 27 | 29 | | |
| | | Total | Total | 374 | 13.85 | 94 | 76 | 168 | 55 | 162 | 776 | 27 | 29 | | |
| | | | | 374 | 13.85 | 94 | 76 | 168 | 55 | 162 | 776 | 27 | 29 | | |
| Dario | Sarik | PF | Cibona | 467 | 16.68 | 90 | 134 | 243 | 30 | 87 | 921 | 28 | 33 | | |
| | | | | 467 | 16.68 | 90 | 134 | 243 | 30 | 87 | 921 | 28 | 33 | | |
| | | Total | Total | 467 | 16.68 | 90 | 134 | 243 | 30 | 87 | 921 | 28 | 33 | | |
| | | | | 467 | 16.68 | 90 | 134 | 243 | 30 | 87 | 921 | 28 | 33 | | |
| Edo | Muric | SF | | 281 | 12.22 | 33 | 68 | 147 | 39 | 108 | 664 | 23 | 29 | | |
| | | | | 281 | 12.22 | 33 | 68 | 147 | 39 | 108 | 664 | 23 | 29 | | |
| | | Total | Total | 281 | 12.22 | 33 | 68 | 147 | 39 | 108 | 664 | 23 | 29 | | |
| | | | | 281 | 12.22 | 33 | 68 | 147 | 39 | 108 | 664 | 23 | 29 | | |
| Jaka | Blazic | SG | | 290 | 10.74 | 41 | 77 | 142 | 25 | 84 | 640 | 27 | 24 | | |
| | | | | 290 | 10.74 | 41 | 77 | 142 | 25 | 84 | 640 | 27 | 24 | | |
| | | Total | Total | 290 | 10.74 | 41 | 77 | 142 | 25 | 84 | 640 | 27 | 24 | | |
| | | | | 290 | 10.74 | 41 | 77 | 142 | 25 | 84 | 640 | 27 | 24 | | |
| Marko | Matinovic | PG | | 352 | 13.54 | 190 | 53 | 110 | 50 | 184 | 797 | 26 | 31 | | |
| | | | | 352 | 13.54 | 190 | 53 | 110 | 50 | 184 | 797 | 26 | 31 | | |
| | | Total | Total | 352 | 13.54 | 190 | 53 | 110 | 50 | 184 | 797 | 26 | 31 | | |
| | | | | 352 | 13.54 | 190 | 53 | 110 | 50 | 184 | 797 | 26 | 31 | | |
| Nemanja | Gordic | PG | | 374 | 14.38 | 121 | 98 | 202 | 33 | 88 | 735 | 26 | 28 | | |
| | | | | 374 | 14.38 | 121 | 98 | 202 | 33 | 88 | 735 | 26 | 28 | | |
| | | Total | Total | 374 | 14.38 | 121 | 98 | 202 | 33 | 88 | 735 | 26 | 28 | | |
| | | | | 374 | 14.38 | 121 | 98 | 202 | 33 | 88 | 735 | 26 | 28 | | |
| Nolan Derec | Smith | PG | | 372 | 13.29 | 93 | 131 | 283 | 21 | 66 | 766 | 28 | 27 | | |
| | | | | 372 | 13.29 | 93 | 131 | 283 | 21 | 66 | 766 | 28 | 27 | | |
| | | Total | Total | 372 | 13.29 | 93 | 131 | 283 | 21 | 66 | 766 | 28 | 27 | | |
| | | | | 372 | 13.29 | 93 | 131 | 283 | 21 | 66 | 766 | 28 | 27 | | |
| Robko | Varda | C | | 356 | 13.69 | 30 | 94 | 188 | 19 | 58 | 568 | 26 | 22 | | |
| | | | | 356 | 13.69 | 30 | 94 | 188 | 19 | 58 | 568 | 26 | 22 | | |
| | | Total | Total | 356 | 13.69 | 30 | 94 | 188 | 19 | 58 | 568 | 26 | 22 | | |
| | | | | 356 | 13.69 | 30 | 94 | 188 | 19 | 58 | 568 | 26 | 22 | | |
| Savo | Lesic | PF | | 411 | 15.81 | 22 | 123 | 208 | 20 | 60 | 769 | 26 | 30 | | |
| | | | | 411 | 15.81 | 22 | 123 | 208 | 20 | 60 | 769 | 26 | 30 | | |
| | | Total | Total | 411 | 15.81 | 22 | 123 | 208 | 20 | 60 | 769 | 26 | 30 | | |
| | | | | 411 | 15.81 | 22 | 123 | 208 | 20 | 60 | 769 | 26 | 30 | | |
| Tarenece Arth | Kinsey | SG | | 274 | 11.91 | 43 | 85 | 183 | 21 | 57 | 629 | 23 | 27 | | |
| | | | | 274 | 11.91 | 43 | 85 | 183 | 21 | 57 | 629 | 23 | 27 | | |
| | | Total | Total | 274 | 11.91 | 43 | 85 | 183 | 21 | 57 | 629 | 23 | 27 | | |
| | | | | 274 | 11.91 | 43 | 85 | 183 | 21 | 57 | 629 | 23 | 27 | | |
| Grand Total | | | | 3551 | 13.66 | 757 | 939 | 1874 | 313 | 954 | 7265 | 260 | 28 | | |

Figure 3. Data warehouse cube reader

When user choose to create a cube by wizard, through the set of windows there are following actions: choosing tables that will allow us to represent the facts and then program provides some measures that can be used or not. After that the list of dimensions is displayed with options for user to choose all of them or some, depending on the analysis type.

When the cube structure is finally projected user has to deploy the cube because the whole project is placed in various files in the project folder and the data is placed in one or more DBMS software files. Deploying may take a while, depending on the complexity of cube structure and the amount of data from database that have to be de-normalized and placed into the cube.

After deployment BIDS software add a new tab in cube editor called Browser where users have to place dimensions, measures and calculated fields in a form that looks like pivot table. There are three regions for placing dimensions: on rows, on columns or on page as a filter for all aggregated data that have to be displayed. Measures and calculated fields must be dragged and dropped in central part of table.

IV. REPORTING AND RESULTS

The reports were created in the Microsoft Business Intelligence Management Studio for table type reports

| FIRST NAME | LAST NAME | TEAM | NUMBER OF POINTS | 2 POINTS MADE | 2 POINTS ATTEMPTS | 3 POINTS MADE | 3 POINTS ATTEMPTS | PLAYED GAMES | MINUTES PER GAME |
|-------------|------------|---------------|------------------|---------------|-------------------|---------------|-------------------|--------------|------------------|
| Nolan Derec | Smith | Cedevita | 372 | 131 | 283 | 21 | 66 | 28 | 27.3571428571 |
| Dario | Saric | Cibona | 467 | 134 | 243 | 30 | 87 | 28 | 32.8928571428 |
| Jaka | Blazic | Crvena zvezda | 290 | 77 | 142 | 25 | 84 | 27 | 23.7037037037 |
| Nemanja | Gordic | Igokea | 374 | 98 | 202 | 33 | 88 | 26 | 28.2692307692 |
| Edo | Muric | Krka | 281 | 68 | 147 | 39 | 108 | 23 | 28.8695652173 |
| Ratko | Varda | Mega Vizura | 356 | 94 | 188 | 19 | 58 | 26 | 21.8461538461 |
| Bogan | Bogdanovic | Partizan | 374 | 76 | 168 | 55 | 162 | 27 | 28.7407407407 |
| Tarenece | Kinsey | Partizan | 274 | 85 | 183 | 21 | 57 | 23 | 27.3478260869 |
| Marko | Matinovic | Radnicki | 352 | 53 | 110 | 50 | 184 | 26 | 30.6538461538 |
| Savo | Lesic | Radnicki | 411 | 123 | 208 | 20 | 60 | 26 | 29.5769230769 |

Figure 4. Data warehouse reporting services table

In BIDS tool a new project was created from the list of templates as a report server type of project. The wizard guides user to create a report with a new and separate data source to OLAP cube. After that in a special tool called Query Designer, user has to create a Multidimensional (MDX) query for slicing data in cube that will appear in the report. MDX query for reading data that is necessary to design a layout table with similar syntax to Structured Query Language (SQL), but the difference is in working with cube dimensions and measures in MDX while SQL handles tables and columns of relational database. One page of report is shown on figure 4, while the whole report has a dozen of pages.

(figure 4) and Microsoft Excel for graphs and charts (figures 5 and 6). Excel is a very suitable tool for processing tabular data orderly. It is very useful in modern business where quality information is required. The advantage of Excel is that it provides an interactive and direct work with data, so that it can be used by people with different levels and types of education for automation primarily his personal business, without any programming knowledge.

For creating charts in Excel that are connected to data from OLAP cube, the pivot table must be put on worksheet. Pivot table is a dynamic table with pooled data from an external data source. This kind of table was a source for creating a special type of chart (Pivot Chart), with possibility for dynamic changes of graph series when the data in pivot is filtered for example.

Figure 5 is one of the created graphs with clubs percentages in shots for one, two and three points. First column filled with blue color presents percentage for one point, second marked as red is percentage for two points, and the green is percentage for two points.

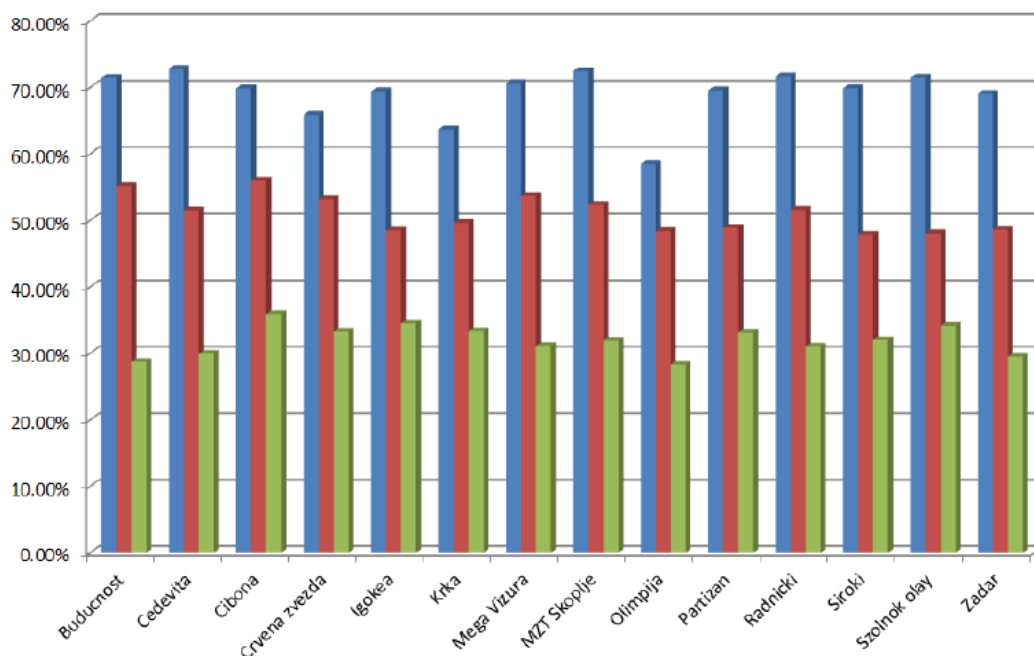


Figure 5. Percentages in shots for one, two and three points

Figure 6 shows in six series total shots for one, two, and three points. First series (column filled with light blue color) on the left shows total one point made. Second column filled with dark blue color shows one point attempts. Third column filled with light green

shows two points made; while fourth column is for two points attempts. Fifth column filled with yellow color shows three points made; while sixth column is for three points attempts during the whole season.

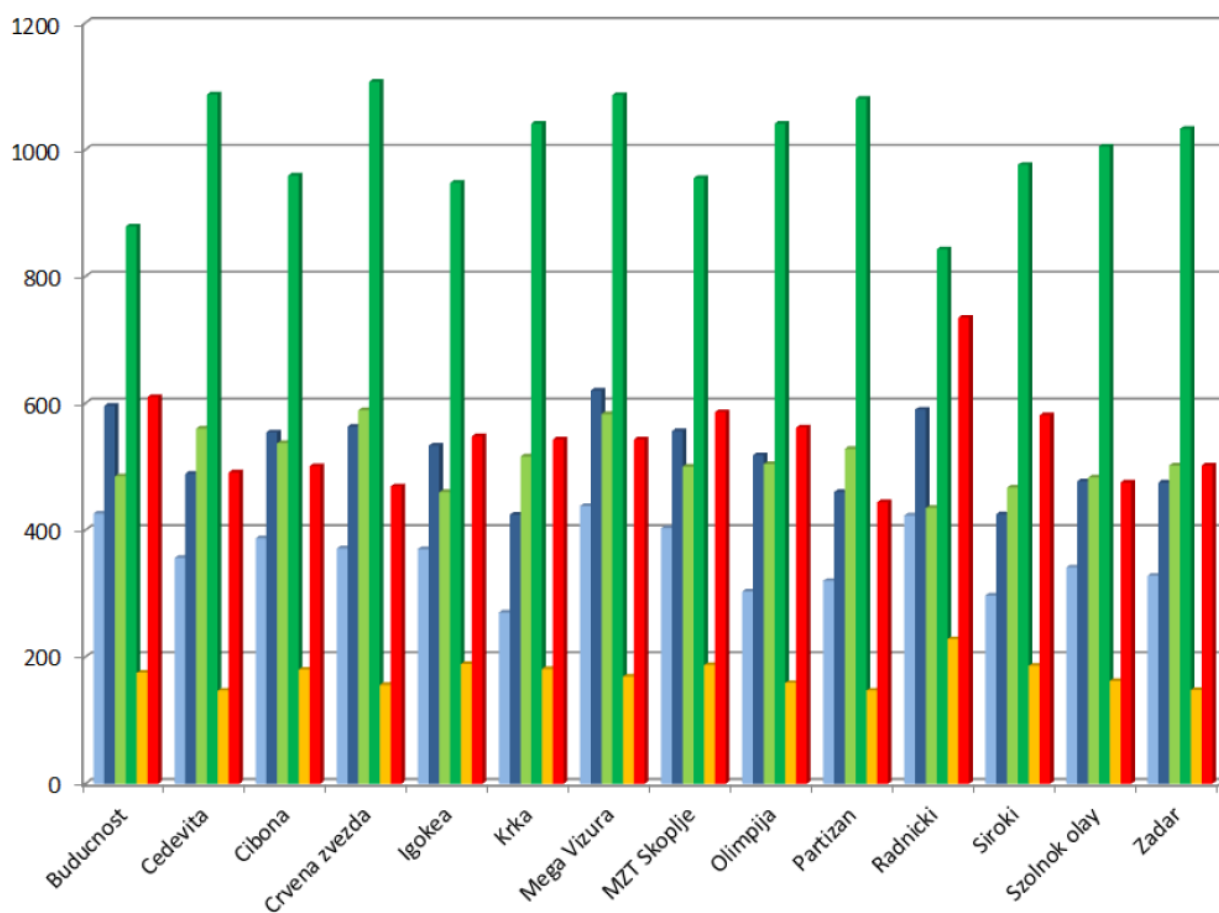


Figure 6. Statistics in shots for one, two and three points

V. CONCLUSION

Aim of this paper is to show how OLAP cubes and data warehouse could support statistical computation of data for basketball team management. Results of projecting and creating data warehouse are cubes with browsed analytical data, reports designed in reporting services and charts created in spreadsheet program.

From this kind of charts and statistics a basketball coach has decision making support and could prepare and modify training process or prepare tactics for different matches.

Further work could be creating mining structures in BIDS for more complex analysis for finding significance in the relationships between a specific data elements and creating data patterns.

REFERENCES

- [1] L. Wu, The Application of Basketball Coach's Assistant Decision Support System, *Journal of Theoretical and Applied Information Technology*, Vol. 49 No. 3, 2013.
- [2] C. Breuer, Data warehouse concept of the German Sport Development Report, http://ec.europa.eu/sport/library/documents/08_breuer_data_warehouse.pdf
- [3] M. I. Hwang, H. Xu: The Effect of Implementation Factors on Data Warehousing Success: An Exploratory Study, *Journal of Information, Information Technology and Organizations* Volume 2., 2007.
- [4] P. Mogin, I. Luković I., M. Govedarica: *Principles of Database Projecting*, University of Novi Sad, Stylos publisher, Serbia, 2000.
- [5] Z. Kazi, B. Radulovic, M. Radovanovic, Lj. Kazi: MOLAP Data Warehouse of a Software Products Servicing Call Center, XXXIII International Symposium Computers in Education, IEEE Region, Opatija, Croatia, Proceedings MiproBIS: pp. 246-250, 2010.
- [6] Z. Kazi, S. Filip, Lj. Kazi, and B. Radulović, "Modeling data warehouse for analysis of air pollution data in urban areas", In proceedings of 1st International conference Ecology of Urban Areas, Ecka-Zrenjanin, Serbia, pp. 552-559.
- [7] ABA league data, <http://www.abaliga.com/n123/Statistics>