**National Research University**

**Higher School of Economics**

Faculty of Business Informatics

Department of Business Analytics

DRAFT

of the paper

«Development of Data Warehouse Facilities to Support Strategic Decision-making in Business»

Student Dinara Alieva

Group 471

Argument Consultant Shcherbinin O.P.

Style and Language Consultant Bagramyants N.L.

2013

Contents

[Introduction 4](#_Toc348794105)

[The actuality of the chosen data domain 5](#_Toc348794106)

[Goals and Objectives 6](#_Toc348794107)

[Research methods 7](#_Toc348794108)

[Literature Overview 8](#_Toc348794109)

[Graduation work’s structure 10](#_Toc348794110)

[Conclusion 12](#_Toc348794111)

[References 13](#_Toc348794112)

Abstract

This paper is devoted to the topic of the data warehouse development and the structure of my graduation work which is entitled “Development of a Data Warehouse Facilities to Support Strategic Decision-making in Business”. Data warehouse is a database that is used to store information for decision support systems.

In this paper the key terms and definitions, as well as the goals and objectives of the graduation paper will be given. It also contains the actuality of the chosen topic and the research methods used to write the graduation paper.

# Introduction

Every company makes up to at least 3 big decisions every month – what should be the price of a new product, what new service they should offer or where they should open a new store. These decisions are based on the data, collected from the everyday operations (Hillard, 2010). But how can these companies get all those data in one place? That is what data warehouses are for.

This paper is devoted to the topic of the data warehouse development and to the structure of my graduation work, entitled “Development of Data Warehouse Facilities to Support Strategic Decision-making in Business”.

In the first part of this paper to the actuality of the topic and the main goals and objectives of my graduation work are described. The second part summarizes the research methods. The third part covers the literature review, and is divided into: main definitions, the history of the data warehouses and main approaches to data warehouses. In the next part is my graduation work’s structure. The concluding part contains a review of key points touched upon in this paper.

# The actuality of the chosen data domain

It is wildly recognized that decision-making is vital for every company, big or small. So is the process of collecting operational data to support these decisions. Nowadays there almost every company has some kind of an operational system to track everyday operations. Those systems are designed in the way that allows creating and updating data very quickly and without data loss in a specific area. However, when it is needed to select large amount of data, especially from several areas this operational systems are not always suitable.

For instance, there is a chain of small shops around town owned by the same company. Since these shops probably were not opened all at once, there is every possibility that each of these shops uses its own application that allows monitoring the products in stock. One day the owner of this company wants to know which products are bought the most which time of the year, so he can plan the purchase orders for the whole chain. Creating a single report based on the data from all these shops is an elaborate task that can freeze the operating system for hours.

The solution to this problem is to create a single data warehouse, where the data from all the different sources can be loaded.

The main advantages of the data warehouses are:

* Keeping data history
* Integrating data from multiple sources
* Improving data quality by data cleaning
* Restructuring the data so that it delivers better query performance, without impacting the operational systems
* Providing a single common data model

Provided with this data one can easily deduce why data warehouses are so important.

# Goals and Objectives

The primary goal of my diploma paper is to analyse the realm of internet hosting business, develop data warehouse facilities that will support decision-making process and study the results.

This goal can be achieved through several objectives:

1. Analysing the main business processes of the internet hosting business
2. Finding the areas that need decision support system
3. Designing a typical structure of an operational system in the sphere
4. Designing the structure of the data warehouse
5. Developing the data warehouse in the Microsoft SQL Server
6. Analysing the effectiveness of the developed system

# Research methods

In order to conduct a thorough analysis and development of a data warehouse a lot of literature needs to be studied. There a lot of national and international practises of this topic, so a lot of these practises need to be examined and generalized.

Theoretical analysis and synthesis of the literature on this topic will make it possible to study the chosen sphere and to develop data warehouse facilities, so that the primary goal of the graduation paper is achieved.

# Literature Overview

Key terms

Online Transaction Processing (OLTP) – processing in which the system responds immediately to user requests. A class of systems that manages transact-oriented applications.

Online Analytical Processing (OLAP) - is an approach to answering multi-dimensional analytical queries swiftly.

Operational system – a system that is designed to support day-to-day operations of an organization. These systems are usually designed to support high-end transaction processing.

Decision support system (DSS) – a set of related computer applications required to assist with business or organizational analysis and decision-making.

Data warehouse – a database used in reporting and data analysis. The data stored in a warehouse is usually extracted, transformed and loaded (ETL) from operational systems. This data is then used in DSS.

Data mart – a layer of a data warehouse that is presented to the actual user.

ETL - Extract, Transform and Load, refers to a process in database usage and especially in data warehousing that involves:

* Extracting data from outside sources
* Transforming it to fit operational needs
* Loading it into the end target

The history of the Data Warehouses

The concept of a data warehouse dates back to the late 1980s. The first 'business data warehouse' was developed by IBM researchers Barry Devlin and Paul Murphy. The main idea behind the data warehouse was to provide a model for the data flow from operational systems to decision support environments. The concept of the data warehouses was invented to solve some of the problems of this flow, mainly the high costs.

There usually were a lot of decision support environments that operated independently of each other, but used the same data. Each environment had to gather, clean and integrate the same data which led to partial replications of the same processes. The lack of a unified system also resulted in a complicated process of redevelopment as new requirements emerged.

This all led to the creation of such a database, that can collect and keep all the necessary data – data warehouse.

What is a data warehouse?

There are two main approaches to data warehouses. Both of them are valid but the premises for using a specific approach are different.

The first approach is presented by Ralph Kimball, and is usually referred to as a bottom-up approach. In this approach for every specific area of business is created a data mart. This are can be for example “Sales” or “Marketing”. Then a Bus is created to link all the data marts together. Every data mart does not depend on another date mart, but always depends on the Bus. Thus, the data warehouse is the merge of all the data marts and the Bus. As a result a single data mart can be created very quickly. As soon as the first data mart is ready the organization can benefit from it. However, this approach makes it very hard to add a new data mart when there are already plenty of them.

Bill Inmon is a leading proponent of a top-down approach. In this approach a data warehouse is a centralized repository for the entire enterprise. The development of the data warehouse starts from the beginning and can take up to several years. Data marts can be created from the data warehouse, only when the warehouse is fully developed, so nobody can see the results until the development is finished. But this approach makes it easier to create further data marts, and extract new data.

Both these approaches can be used, but the bottom-up approach is usually better for small and medium companies and the top-down approach is used for big companies.

# Graduation work’s structure

The structure of my graduation work is the following:

Chapter 1. Theoretical material

* Operational systems
* Decision support systems
* Difference between operational systems and decision support systems
* What is a data warehouse?
* The necessity of data warehouses.

Chapter 2. Data warehouse creation – practical part

* Detailed goal statement
* Work plan
* Internet hosting sphere description
* Data warehouse design
* Detailed description of a data warehouse development

Chapter 3. Results analysis – analytical part

* The result of implementing a data warehouse
* Results analysis
* Disadvantages
* Alternatives

Chapter 4. Resume

* What improvements can be made and how the data warehouse can be used in future

Chapter 1 is called “Theoretical material” and it will consist of all of the necessary theory connected to the data warehouses development. Firstly the data warehouses will be discussed in terms of their capabilities. Secondly the necessity of implementing a data warehouse will be shown. Finally the functionality of the data warehouse and its ways of retrieving data from the operating system will be discussed. Concluding the Chapter I will give a brief diploma’s goal statement.

Chapter 2 represents practical part of the diploma. The chapter will begin with a detailed goal statement, followed by a plan of data warehouse development. It will be followed by a description of Internet hosting sphere, it's main business-process and its typical operational system overview.

After there will be given a detailed description of the actual data warehouse will be design and development.

Chapter 3 is entitled “Results analysis” and represents diploma’s analytical part. Here the results of implementing a data warehouse for a specific project will be shown. Results will be analysed from a position of what was planned, what it turned out to be in reality and what should be done in future to eliminate mistakes. Moreover, I will try to highlight some disadvantages and possible alternatives.

Chapter 4 will give the resume to everything written in the paper. In this part some improvements for the system and future use for the data warehouse will be suggested.

# Conclusion

The data warehouses are vital in decision-making process. There are several approaches used to develop a data warehouse and it is important to choose the proper one.

In this paper the graduation work “Development of Data Warehouse Facilities to Support Strategic Decision-making in Business” was reviewed. The actuality, goals, objectives and research methods were examined.

# References

* 1. Becker B., Thornthwaite W., Mundy J., Kimball R., Ross M. (2011). *The data warehouse lifecycle toolkit, 2nd edition.* Indianapolis: Wiley Publishing, Inc.
  2. Hillard R. (2010). *Information-Driven Business*. Indianapolis: Wiley Publishing, Inc.
  3. Inmon W. H. (2005). *Building the data warehouse, 4th edition*. Indianapolis: Wiley Publishing, Inc.
  4. Ross M. (2003). The bottom-up misnomer. *Business Intelligence and Data Warehouse Articles*, Sep 17.
  5. Golfarelli M., Rizzi S. (2009). *Data warehouse design: Modern principles and methodologies.* New York: McGraw-Hill, Inc.