МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ
Факультет кібербезпеки, комп'ютерної та програмної інженерії
Кафедра комп'ютерних систем та мереж

"ДОПУСТИТИ ДО ЗАХИСТУ"
Завідувач кафедри

_____ Жуков І.А.

"_____"_____2020 р.

ДИПЛОМНА РОБОТА

(ПОЯСНЮВАЛЬНА ЗАПИСКА)

випускника освітнього ступеня "МАГІСТР"

спеціальності 123 «Комп'ютерна інженерія»

освітньо-професійної програми «Комп'ютерні системи та мережі»

на тему: **"Інформаційна система аналітичного відділу з використанням хмарних технологій"**

Виконав: _____ Швець А.В.

Керівник: _____ Кудренко С.О.

Нормоконтролер: _____ Надточій В.І.

Засвідчую, що у дипломній роботі немає запозичень з праць інших авторів без відповідних посилань

Швець А.В.

(підпис студента)

Київ 2020

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
NATIONAL AVIATION UNIVERSITY
Faculty of Cybersecurity, Computer and Software Engineering
Computer Systems and Networks Department

"PERMISSION TO DEFEND GRANTED"

The Head of the Department

_____ Zhukov I.A.

"_____"_____2020

MASTER'S DEGREE THESIS

(EXPLANATORY NOTE)

Specialty: 123 Computer Engineering

Educational-Professional Program: Computer Systems and Networks

Topic: **"Information system of the analytical department using cloud technologies"**

Completed by: _____ Shvets A.V.

Supervisor: _____ Kudrenko S.O.

Standard's Inspector: _____ Nadtochii V.I.

Kyiv 2020

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет кібербезпеки, комп'ютерної та програмної інженерії

Кафедра комп'ютерних систем та мереж

Освітній ступінь: «Магістр»

Спеціальність 123 «Комп'ютерна інженерія»

Освітньо-професійна програма: «Комп'ютерні системи та мережі»

"ЗАТВЕРДЖУЮ"
Завідувач кафедри

_____Жуков І.А.

« ____ » _____ 2020 р.

З А В Д А Н Н Я
на виконання дипломної роботи

Швеця Артура Валерійовича
(прізвище, ім'я, по батькові)

1. Тема проекту (роботи): "Інформаційна система аналітичного відділу з використанням хмарних технологій"

затверджена наказом ректора від 25.09.2020 р. № 1793/ст.

2. Термін виконання роботи (проекту): з 1 жовтня 2020 р. до 25 грудня 2020 р.

3. Вхідні дані до роботи (проекту*): Схематичне графічне зображення базових сервісів хмарних технологій, функції працівників аналітичного відділу.*

4. Зміст пояснювальної записки: *Вступ, огляд теми, огляд існуючих технологій розробки інформаційних систем з використанням хмарних технологій, розробка локальної комп'ютерної мережі, висновки по роботі .*

5. Перелік обов'язкового графічного (ілюстративного) матеріалу: *Графічні матеріали результатів дослідження надати у вигляді презентації у форматах .ppt..*

6. Календарний план-графік

№ п/п	Етапи виконання дипломного проекту	Термін виконання етапів	Підпис керівника
1	Узгодження технічного завдання з керівником дипломної роботи	01.10.20 – 05.10.20	
2	Підбір та вивчення науково-технічної літератури за темою дипломної роботи	06.10.20 – 07.10.20	
3	Опрацювати теоретичний матеріал	08.10.20 – 15.10.20	
4	Виконати аналіз впровадження перспективних хмарних технологій в роботу аналітичного віддлу	16.10.20 – 25.10.20	
5	Розробити інформаційну систему та тестування хмарних рішень	26.10.20 – 25.11.20	
6	Оформити пояснювальну записку	26.11.20 – 12.12.20	
7	Оформити графічні матеріали проекту та представлення роботи на антиплагіат	13.12.20 – 14.12.20	
8	Отримати рецензію та відгук керівника. Надати матеріали роботи на кафедру	15.12.20 – 18.12.20	

7. Дата видачі завдання «01» жовтня 2020 р.

Керівник дипломної роботи _____ Кудренко С.О.
(підпис)

Завдання прийняв до виконання _____ Швець А.В.
(підпис студента)

NATIONAL AVIATION UNIVERSITY

Faculty of Cybersecurity, Computer and Software engineering

Department: Computer Systems and Networks

Educational Degree: "Master"

Specialty: 123 "Computer Engineering"

Educational-Professional Program: "Computer Systems and Networks"

"APPROVED BY"

The Head of the Department

_____ Zhukov I.A.

"_____"_____2020

Graduate Student's Degree Thesis Assignment

Shvets Artur Valeriyovych

1. Thesis topic: **"Information system of the analytical department using cloud technologies"** approved by the Rector's order of 25.09.2020 № 1793/st.

2. Thesis to be completed between 1.10.2020 and 25.12.2020.

3. Initial data for the project (thesis): *Schematic graphic representation of basic services of cloud technologies, functions of employees of the analytical department*

4. The content of the explanatory note (the list of problems to be considered):

Introduction, review of the topic, review of existing technologies for the development of information systems using cloud technologies, development of a local computer network, conclusions on the work .

5. The list of mandatory graphic materials: *Graphic materials are given in MS Power Point presentation.*

6. Timetable

#	Completion Stages of Degree Project	Stage Completion Dates	Signature of the supervisor
1	Technical task coordination with the supervisor	01.10.20 – 05.10.20	
2	Selection and study of scientific and technical literature on the topic of the thesis	06.10.20 – 07.10.20	
3	Elaboration of theoretical material	08.10.20 – 15.10.20	
4	Perform analysis of the introduction of promising cloud technologies in the work of the analytical department	16.10.20 – 25.10.20	
5	Design an information system and cloud solution testing	26.10.20 – 25.11.20	
6	Making an explanatory note	26.11.20 – 12.12.20	
7	Design of graphic materials of the project and work presentation on anti-plagiarism checking	13.12.20 – 14.12.20	
8	Receiving review from the supervisor. Deliver materials to the department.	15.12.20 – 18.12.20	

7. Assignment issue date: 01.10.20

Diploma Thesis Supervisor _____ Kudrenko S.O.
(Signature)

Assignment accepted for completion _____ Shvets A.V.
(Signature)

ABSTRACT

The Explanatory Note to the Master's Degree Thesis "Information system of the analytical department using cloud technologies": 80 pages, 27 figures, 5 tables, 25 references.

Object of research – automatization of processes of the analytical department and creating of the information system.

Purpose – to investigate the process of modeling the information system of the analytical department using cloud technologies.

Research methods: exploring of existing cloud technologies services and analytical department needs, comparative analysis, processing of literature sources.

The results of the master's work are recommended to be used during comparison, creating and introducing processes of information system via cloud technologies.

ANALYTICAL DEPARTMENT, CLOUD TECHNOLOGY, INFORMATION SYSTEM, LOCAL NETWORK, NETWORK ARCHITECTURE, SOFTWARE.

CONTENT

LIST OF SYMBOLS, ABBREVEATIONS, TERMS

ASP	–	Active Server Pages
BABOK	–	Business Analysis Body of Knowledge
DSS	–	Decision Support Systems
GNU	–	Gnu's Not Unix
ICMP	–	Internet Control Message Protocol
IEEE	–	Institute of Electrical and Electronics Engineers
IIBA	–	International Institute of Business Analysis
ISO	–	International Standardization Organization
LAN	–	Local Area Network
OSI	–	Open System Interconnection
SLA	–	Service Level Agreement
SSL	–	Secure Socket Layer
WLAN	–	Wireless Local Area Network
HPF	–	High Pass Filter
HOG	–	Histograms of oriented gradients
LPF	–	Low Pass Filter
ML	–	Machine Learning
NIST	–	National Institute of Standards and Technology
OpenCV	–	Open Source Computer Vision Library
PRP	–	Pattern Recognition Problem
RNN	–	Recurrent Neural Network
RGB	–	“Red, Green, Blue” color model
API	–	Application Programming Interface

INTRODUCTION

Actuality of theme. Today, no industry can do without the implementation of automation systems, decision support. When developing software for a department, Business Intelligence (BI) is a collection of software applications, techniques and business systems that play a key role in the business processes of any corporation.

Most companies generate huge amounts of data in the course of their business. To provide access to this data to all departments of the company, a wide arsenal of applications and DBMS is often used.

But the abundance of various programs and systems makes it difficult to extract useful information in a timely manner and draw conclusions based on it. BI software is designed to rectify the situation by replacing disparate data collection and analysis tools.

The history of BI began much earlier than is commonly believed. The earliest mention of BI dates back to 1865, when Richard Devens used the term Business Intelligence to describe the actions of Henry Farness, who studied data, analyzed market conditions, and thereby outstripped his competitors. Later in 1958, Hans Peter Lun of IBM wrote about the potential of BI when you add the power of information technology to this approach. We are all, of course, interested in the technological side of BI, which allows us to find new solutions for business. There are more than 50 different products on the market today, offering BI functionality to one degree or another. But in order to better understand how they differ from each other, and why the functionality of BI solutions began to change dramatically after 2007, it will be useful to look back and trace the evolution of BI as a class of computer systems.

The first decision support systems (DSS - Decision Support Systems) appeared in the 1970s after Edgar Codd proposed a new method of organizing data to the world community. Relational databases made it possible to make a leap forward in the field of BI, including, and decision support tools were actively developing until the mid-1980s.

In the 1980s, active data consolidation began, information began to accumulate in single centers, creating conditions for in-depth and comprehensive analysis. BI tools gained access to large amounts of data, but they remained complex technical tools, so it was impossible to build another report without a smart IT specialist, and therefore the business did not use this category of solutions very often. At least, not in all industries.

In the 1990s, BI technology began to be seen as a business tool, and it has found applications in many industries, from mining to manufacturing and finance. At that time, BI solutions were very expensive and the task could take several days to formulate. It was from this era of the so-called "BI 1.0" that a stereotype arose about the inaccessibility of BI class solutions for "mere mortals"[4].

In the 2000s, a technological base for processing data in real time appeared, which means that companies were able to make decisions instantly, based on current, constantly changing information. Various dashboards displaying a set of indicators began to appear on the market, as well as BI systems with an intuitive interface (such as Tableau or QlikView), in which business users could already make queries. IT professionals were only required to prepare, cleanse, and aggregate data.

The purpose of the thesis is to investigate the process of modeling the information system of the analytical department using cloud technologies.

Business intelligence software, or BI means, is understood as software applications that perform all the tasks of monitoring, collecting and analyzing data, and also allowing you to track patterns in this data and draw conclusions from them the business needs. These applications are also capable of retrieving and using information stored in corporate databases.

The most common enterprise-class BI systems include tools for working with tables and dashboards of operations, data extraction and reporting, data processing and analysis, content viewing, and other tools for managing enterprise resources. Some BI packages also integrate industry-specific tools such as retail, healthcare, or education.

The point is that the tools used by analysts are outdated. Today's business intelligence tools are 10-15 years old, and instead of moving to the cloud, the industry continues to stagnate. Unwillingness or inability to transform leads to a slowdown in

market growth. And if you put it in numbers, then you can mathematically prove a simple truth: cloud technologies help businesses grow faster.

The business intelligence market itself is estimated at $ 150 billion in net profit. And the average annual growth rate will be at least 10%. This is, of course, powerful and rapid growth. However, this market has huge potential, which will allow it to reach $ 300 billion by 2030. In order to achieve these targets, the market needs to do two things.

Research methods – exploring of existing cloud technologies services and analytical department needs, comparative analysis, processing of literature sources.

First, adapt to cloud reality. To simply survive in the current economic situation, a business needs to process a huge amount of data and integrate the results of analysis into business processes on the fly. For this to happen really fast, the BI tools and platform need to be ready for the cloud economy. A living example of such a platform is the Netflix service, the user part of which works in the public cloud, and the analytical platform collects data on user preferences and immediately gives them the most suitable series and films. Cloud analytics that work for users and add value to the service.

The designing object of project – automatization of processes of the analytical department and creating of the information system with introducing cloud technologies. And even though clouds can create tangible value and thus help grow a business, not many people understand the difference between cloud-ready and cloud adoption. Once the clouds are in place, you can easily extract insight and factual value from tons of disparate data like Dropbox files, tweets, RSS feeds, documents, Facebook comments. You are able to process this data in real time and receive useful information for your business. This is just one example out of dozens.

Dominant, however, is a different approach. Business analysts prefer to say that they are ready for the transition to the cloud, but we are not seeing a massive adoption of clouds due to this "readiness". First of all, because this approach does not mean implementation, but a banal transfer of existing, outdated systems to cloud hosting. Thus, without changing practically anything, the analytical department can pretend that

all the analytics of the company are already in the clouds. But virtually no additional value is produced.

Running your infrastructure on a cloud platform is much easier and cheaper than renting a convenient cloud business intelligence tool. As cloud specialist David Linthicum says, “Cloud technology is widely and misunderstood as the ability to push anything into the clouds. Because of this, the business loses added value, and customers receive less value. "

But if we have decided on the value and acceleration of business as the main consequences of the introduction of clouds, then what exactly are cloud technologies for in business intelligence? Mainly in order to strengthen all business departments with analytical results. According to statistics, only 24% of employees of companies that have implemented business intelligence use the insights acquired during data analysis in their operations. What if with the help of the clouds we could increase this figure to 90 or even 100 percent? To do this, you need to understand how business intelligence works in most companies. Usually the data is first received by the financial planning and analytics department. The department is directly dependent on the IT department that supplies the data. The information is processed by analysts and is used within the framework of the same financial planning department[5].

And it turns out that the company's employees who work directly with the client receive less analytical results – which means they cannot draw the correct conclusions about how exactly they need to adjust their activities. Hence the sad figure of 24%, which needs to be changed. Which is more than possible thanks to cloud technology.

The practical significance of the thesis results are recommended to be used during comparison, creating and introducing processes of information system via cloud technologies. In recent years, cloud technologies topic has become one of the most popular in the IT field, many articles have been written about it, an even larger number of conferences have been held, and how many solutions already exist on the market (and are used by us in all our daily life, sometimes even unconsciously) , and never count.

However, as always, there is one "but", namely, most of the users, as before, do not even know what kind of "know-how" cloud technologies are and why they have given up. Well, we will correct the current situation and we will start, as expected, with a theory.

Scientific novelty of the obtained results. Cloud computing is a distributed data processing technology in which computer resources and power are provided to the user as an Internet service. If we explain it in an accessible language, then this is yours, in a sense, a working platform on the Internet, or rather on a remote server.

Cloud providers and hosters use server virtualization technologies. But there are other virtualization technologies in nature as well.

The first is storage virtualization. Its use allows you to combine a large number of physical storage systems into a single logical pool. In other words, a zoo from a storage system can be combined into a single whole, visible from the outside, for example, as 300TB of disk capacity. With this approach, disk fragments of one virtual server can be stored on different storage systems of the provider.

Second is network virtualization. Network virtualization technology allows you to move away from the use of the standard functionality of network equipment, namely, from the use of VLANs. After all, the maximum possible number of VLANs on physical equipment is limited to 4096. And if you want to become a major cloud provider, then you will definitely use this limit. Network virtualization software, in turn, allows you to run an unlimited number of isolated networks and allows you to scale the cloud almost indefinitely.

In addition, the use of network virtualization software allows you to give less importance to the composition of physical network equipment and its manufacturers. Allows you to abstract from hardware, which makes the cloud provider more flexible and not tied to specific vendor technologies.

Using storage and networking virtualization software is not a requirement for a cloud provider, but a testament to its maturity and commitment to becoming a technology leader.

The implementation of a cloud platform for business intelligence will make the results of data processing and analysis available. First of all, employees around the company, those who work with clients. If, say, a call center can receive data on customer reaction to the next sales script in real time, this can bring the company tangible profits due to the constant correction of speech modules.

Simply put, cloud computing is the provision of computing services (including servers, storage, databases, networks, software, analytics, and mining) over the Internet ("the cloud"). These services accelerate innovation, increase resource flexibility, and provide cost savings through high scalability. You usually pay only for cloud services, which can help you reduce operating costs and improve infrastructure management and scale as your business needs change.

PART 1

PRINCIPLES FOR THE DEVELOPMENT OF INFORMATION SYSTEMS TO SUPPORT THE WORK OF THE ANALYTICAL DEPARTMENT

1.1. Analysis of the Target Audience of Users to Support the Work of the Analytical Department

To develop an information system for the analytical department, it is necessary first of all to determine who will use this system and what data will be required for this.

At the moment, there are two main concepts of analysis – "system analysis" and "business analysis". If we look at the origin of the words “system” and “business”, we will see that “system” comes from the ancient Greek word meaning “whole, made up of parts; compound". And the word "business" from the English "business", "enterprise". It turns out that analysts do something that develops and changes an object consisting of parts (in the case of a system) and something on which the whole business and the enterprise (in the case of a business) depend.

Now let's look at what a business analyst does. There is a fundamental work in the world describing the professional standard / set of rules for business analysis, which is being developed by the International Institute of Business Analysis (IIBA) under the name BABOK (Business Analysis Body of Knowledge). By his definition, a business analyst is an employee of a company or a project participant who identifies the needs for changes in the organization (organizational development), summarizes them, records, classifies and recommends decisions on changing the structure of the company's activities (processes), justifies these changes for management and participates in the detailed development and implementation of changes. Most organizations also struggle to get the expected return on their investment in acquiring and maintaining BI. Another common problem is an abundance of complaints from frustrated users who were unable to access data or received erroneous data.

Of course, in general, these changes are needed by the company to generate more profit, i.e. fulfillment of business goals.

The business goals themselves can be born both within the company and under the influence of the external environment and changes in the market in which the company trades. At the same time, the business goal is the benefit for the company that it will receive after implementing a new solution. To put it simply, the business analyst has a finger on the pulse of the company's internal business development, and responds to changes in the pulse. External factors that affect the company's business are recorded or studied by the marketing direction. At the same time, marketing can act as sources for the formation of business needs. In this way, the business analyst identifies the needs of the business and can always explain the reason for a particular change. which she will receive after implementing the new solution. To put it simply, the business analyst has a finger on the pulse of the company's internal business development, and responds to changes in the pulse [7].

External factors that affect the company's business are recorded or studied by the marketing direction. At the same time, marketing can act as sources for the formation of business needs. In this way, the business analyst identifies the needs of the business and can always explain the reason for a particular change. which she will receive after implementing the new solution. To put it simply, the business analyst has a finger on the pulse of the company's internal business development, and responds to changes in the pulse. External factors that affect the company's business are recorded or studied by the marketing direction. At the same time, marketing can act as sources for the formation of business needs. In this way, the business analyst identifies the needs of the business and can always explain the reason for a particular change. At the same time, marketing can act as sources for the formation of business needs. Thus, the business analyst identifies the needs of the business and can always explain the reason for a particular change. At the same time, marketing can act as sources for the formation of business needs. Thus, the business analyst identifies the needs of the business and can always explain the reason for a particular change.

The line between a business analyst and a systems analyst is often so thin that the functions of a systems analyst are attributed to business analyst and vice versa. For example, if we look at the definition of “business analyst” in the 2014 book by Carl Wigers and Joy Beatty (Designing Software Requirements), we see that in their terminology “business analyst” can have a lot of names, including the name "system

analyst". But, despite the "floating" name of the role of a business analyst, his task is also understood – he must find out why users need a new system.

A business analyst draws on his knowledge, experience and the requirements of his work environment. He chooses a method of recording the results of his activities, namely, recording business requirements and their solutions for implementation in the form of business processes. Thus, an intermediate result of a business analyst's work is a set of documentation describing the target device of an organization in its variable part.

The "systems analyst" is often described as the "task manager". Just like a business analyst, a systems analyst does not have a reference, clear definition, and the boundaries of his activities from project to project, from organization to organization, from environment to environment, can be very different. It so happens that a business analyst acts as a customer for a systems analyst, who in turn tries to understand the “desires” of the business, assesses the “adequacy” of requirements and the “feasibility” of their implementation (often this assessment takes place together with the development team and the architect of the software product), after which the systems analyst translates the business requirements into the development language and sets the development task. Thus, the "systems analyst" is a filter and transformer between business and development. The result of the work of a system analyst is the formulation of a problem to create a system, software, application. Its design already depends on the methodology of software product development or project management, but most often, it is a document called "Terms of Reference". The system analyst collects all the requirements and integrates them, systematizes and describes the system setting of the problem[11].

There are also several other types of analysts:

– analyst-designer UX (User eXperience) – often such specialists are called designers or designers of user interfaces, although they also perform analytical work. The word "designer" in the Russian-speaking environment has long acquired the meaning of "draftsman", but the primary translation is design. So is our specialist, he is an interface designer. The most important task of the "UX Designer" is to develop a user interface in such a way that it is intuitive and convenient for the end user and solves his problems. Interface designers analyze the information that the system analyst and business analyst collects and records. Oddly enough, the designer devotes most of the working time to the process of parsing and understanding the requirements for the user interface that displays

the functionality of the system. Therefore, we boldly call him "UX Analyst". Such an analyst puts himself in the user's shoes and figures out exactly how the interface should work. “UX analyst”, possessing knowledge, tools and sound logic, analyzes human behavior and actions (business processes) performed by him. He designs the user interface in such a way that it is comfortable to work in it and he meets the requirements of working with it. In other words, the main task of the UX analyst is to improve the usability of the interface, i.e. make it convenient and understandable based on the behavior of the target audience of product users. The UX design direction is gaining momentum as a distinct class of the profession in the IT environment. Such specialists can be found on projects of multi-user systems and / or portals. Often this class of analysts is absent from projects and its responsibilities lie at the interface between the systems analyst and the designer. On small projects, the system analyst deals with the interface, and there is a problem: the systems analyst often looks at the interface from the point of view of the system, and not the user, and when designing the interface, he does not think about the convenience for the user. In turn, there is also a problem with the UX designer, since he does not see the system as a whole and can be involved in the project as an expert who identifies interface usability problems, but does not solve them himself. Therefore, the ideal way of working on a project is achieved by the joint work of a systems analyst and a "UX analyst". Of course, do not forget that 2 more specialists are involved in the process of designing the user interface:

– Data analyst (Big Data analyst) – an analyst who analyzes the so-called Big Data (big data, although in fact it is any data) and has the knowledge, skills, methods of how to process, classify, structure, store, transform data and to give them in a convenient form to all the “needy”. Such analysts are called mathematicians, mathematicians-programmers, information analysts, and sometimes system analysts too, but with knowledge of working with Big Data. There can be a great variety of names. Of course, the work of these specialists is based on their mathematical education and knowledge in the field of mathematical statistics, data analysis algorithms and mathematical modeling. Such specialists are needed by companies that need to manage customer demand, and these are usually Telecom and the Banking sector, as well as often Retail;

– Analyst / Data Scientist – the activities of a Data Scientist specialist are aimed at processing information to form conclusions, search for hidden patterns and extract

knowledge from large amounts of information. Such conclusions are needed to make decisions in business. The simplest example and familiar to everyone is the question: "To lend money or not?" If a friend comes to you and says: "Give me a loan until next month?" You begin to collect and analyze information: do I have a lot of money, and how good is this friend, and how often does he ask for a loan, and why does he need money, but last time he did not return the money on time, but do I need this friend, eventually! And then you make a decision. Likewise, companies, in order to offer a service and get feedback from a client, conduct research on all information on the client. Many go further introducing data analysis and decision-making systems, they begin to collect detailed information about their customers, right up to the analysis of travel routes and make a conclusion where to put the store (obviously, near the metro station, and which station?), and maybe where to find your partner on business, what to offer to the client, etc. Data Scientist is at the intersection of business analysis (well versed in business) and big data analysis (you can often see that specialists with knowledge of Big Data are looking for a position in this area). Data Scientists are divided into analysts and programmers. Thus, this new direction stands at the intersection of analysis and programming. The specialist must be able to program his mathematical hypotheses and conclusions for analytical purposes. The result of Data Scientist's work is often not immediately noticeable [16].

– integration analyst – again, let's return to large projects with a large number of different specialists. For example, participating in a project for the implementation of a large software product, it must be built into the existing IT environment, or a new one must be built around it. It is for such purposes that a separate direction of integration and analytics is allocated, which work on such projects. Looking ahead, I can say that for several years I worked as an integration analyst, but I can calmly call myself a systems analyst. What is the peculiarity? In certain knowledge and skills. An integration analyst is responsible for connecting systems into a single whole and, ideally, receiving a business process from a business analyst, he can understand the joints of the information systems involved in it, and work out the connection process together with the architect. In the simplest sense, integration is mapping, i.e. comparison of information flows from one system to another. An integration analyst most often draws mapping tables of one information object, converting it to the format of another. The simplest example is the format for presenting a phone number, somewhere it can start with 8, somewhere with +7,

and often systems allow users to enter a phone number in any format, and the integration aligns the data and "releases" them in the correct, clean view of the company's IT landscape. Thus, an integration analyst is a special case of a systems analyst who has knowledge of working with technologies and solutions for building intersystem integration interaction. An integration analyst most often draws mapping tables of one information object, converting it to the format of another. The simplest example is the format for presenting a phone number, somewhere it can start with 8, somewhere with +7, and often systems allow users to enter a phone number in any format, and the integration aligns the data and "releases" them in the correct, clean view of the company's IT landscape. Thus, an integration analyst is a special case of a systems analyst who has knowledge of working with technologies and solutions for building intersystem integration interaction. An integration analyst most often draws mapping tables of one information object, converting it to the format of another. The simplest example is the format for presenting a phone number, somewhere it can start with 8, somewhere with +7, and often systems allow users to enter a phone number in any format, and the integration aligns the data and "releases" them in the correct, clean view of the company's IT landscape. Thus, an integration analyst is a special case of a systems analyst who has knowledge of working with technologies and solutions for building intersystem integration interaction. and often systems allow users to enter a phone number in any format, and integration aligns the data and releases it in the correct, clean form across the company's IT landscape. Thus, an integration analyst is a special case of a systems analyst who has knowledge of working with technologies and solutions for building intersystem integration interaction. and often systems allow users to enter a phone number in any format, and integration flattens the data and releases it in the correct, clean form across the company's IT landscape. Thus, an integration analyst is a special case of a systems analyst who has knowledge of working with technologies and solutions for building intersystem integration interaction [5].

– virus analyst – at present, with the development of the Internet, everyone understands what it means to catch a "virus", and all users install antivirus software on their personal computers. Such an analyst "keeps his finger on the pulse" of information security development and helps to protect users or information systems. He puts himself in the shoes of the creator of the virus and tries to understand the goals and reasons that such a creator wanted in the end and what methods of protection he provided. The analyst,

understanding the work of the virus, offers options for protection against it, as well as, based on his experience and knowledge, tries to get ahead of the creators of new viruses, namely to predict the behavior of new viruses and have time to protect users in advance.

Table presents a comparison of the direction in analytics and shows the main qualities and differences of analysts (Tab. 1.1).

Table 1.1

Summary analysis of the classes of employees of the analytical department

Analyst	**The main purpose of professional activity**	**Typical labor functions**	**Necessary skills, education.**
1	2	3	4
Business analyst	Meeting business development needs for profit and increasing other business metrics	Identifying business needs, assessing them, making decisions to change existing processes or introduce new ones. Description of possible options for business development	Higher education (technical, economic or specialized in relation to the business sphere). Creativity, good understanding of the business development area. It is necessary to be well oriented in business, to understand and apply marketing research, to be in a business trend.
Systems Analyst	Setting tasks for the development of a software product	Filtering business requirements, understanding the feasibility of development, and then fixing functional requirements in an accessible way for development (writing technical specifications, statements, their approval)	Higher technical, engineering education. Coordination of the development process, the ability to quickly provide information, making informed decisions on development problems. Product acceptance and presentation to the customer.
UX Analyst	Designing user interfaces	Revealing the needs of the user, the ability to transfer these needs to the user friendly user interface. Designing web systems and applications	Higher technical education. Possession of web interface design tools. Knowledge of the principles of usability, knowledge of design techniques. Ability to see the big picture from the user's point of view.

Finish of table 1.1.

1	2	3	4
Big Data Analyst	Creation of data warehouses. Analysis of available data	Structuring data for storage, transformation, search, quick and easy access to them, for subsequent processing and analytical reporting or to meet other business needs	Higher mathematical education, good knowledge of mathematical statistics. Knowledge of algorithms for data analysis, mathematical modeling. Knowledge of building databases.
Data Scientist Analyst	Data research to identify metrics for business decision making	Formulates, fixes, programs put forward mathematical hypotheses for decision-making regarding the company's business	Higher mathematical education, good knowledge of mathematical statistics. He is well versed in the business of the company, has knowledge of big data analysis. Possesses the tools and knowledge of working with text for the subsequent classification of key concepts and building relationships between them
Integration analyst	Setting tasks for the development of an integration software product	Intersect with the functions of a system analyst, except for the specifics of integration, namely the construction of mapping tables of various data formats on top of each other	Higher technical, engineering education. Knowledge of integration products, understanding of the main architectural solutions for building integration processes. Coordination of the development process
Virus analyst	Predicting the appearance of new viruses, protecting information from known and potential viruses	Designing options for protecting information from viruses	Education in the field of information security, computer security automation. Understanding of information technology, good programming skills, a desire to delve into hacked software, identifying the causes and methods of infection

1.2. Analysis of the structure of software tools for business intelligence

Not knowing or being able to apply best practices for deploying BI tools tends to lead to ineffective decisions or even approval of a development plan against the wrong priorities. To make the best use of business intelligence, you must understand the

importance of data and its analysis, as well as understand the processes and tools needed to effectively process information [24].

Most organizations also struggle to get the expected return on their investment in acquiring and maintaining BI. Another common problem is an abundance of complaints from frustrated users who were unable to access data or received erroneous data.

There are various reasons for this, but in reality, these unfortunate situations often arise because companies, even with the best software in the industry, do not know how to deploy and use it properly. While it's important to choose easy-to-use BI tools with a user-friendly interface, there are a number of other factors that can hinder you from using the tool effectively and making sure it’s profitable.

For those looking to get the most out of their BI tools, we offer a range of proven guidelines.

1. Enlist the support and participation in the project of all employees of the enterprise. Before implementing BI tools, you need to create a unified business intelligence network that includes all departments of the company – from sales and marketing to logistics and senior management. You need to enlist the support of all departments and collect feedback.

2. Develop a development plan and implementation strategy. By trying to implement business intelligence without a well-thought-out strategy, you can make many mistakes, wasting time and money. Before deciding between BI solutions available on the market, it is very important to identify the core needs of your company and set goals that will meet those needs.

At this stage, you need to involve the entire organization in creating a strategy that will determine what you want and draw up a development plan that will help you achieve this. Your BI tool selection and implementation should be based on this plan.

3. Find Areas Where BI Tools Will Benefit and Deploy There. Do not turn your BI solution deployment into a complex multi-step process — the more difficult it is, the more difficult it will be to master. It is recommended to integrate BI into existing systems that employees are already familiar with.

4. Find out the needs of each department. Most companies already have reporting systems. They may not provide the information you want, or they may be error prone – for example, if Excel documents are created by hand or using self-made scripts. It is

recommended that you review the available reporting options to know what they are doing well and what they are not doing at all, and highlight areas for improvement.

5. Be clear about your goals. When starting a BI implementation project, it is very important to set goals or formulate key expected results. This will help you determine later if you have succeeded.

Knowing what goals you are facing, it will be easier to understand which measures are more likely to contribute to their achievement. In addition, setting targets will make it easier to highlight the most important data.

6. Highlight key KPIs or other statistics to monitor BI uptake. One of the most helpful guidelines for implementing BI is deciding which key performance indicators (KPIs) will most accurately tell you how quickly and successfully your BI implementation is going. But only a few companies succeed in following this advice: the majority is lost in the abundance of various statistics, unable to choose the most significant KPIs.

However, by highlighting only a few of the most important indicators from the mass of indicators, it will be easier for you to determine whether you have achieved a result.

7. Conduct regular surveys of key stakeholders. Meetings with key stakeholders need to be held regularly, be it routine calls, video chats, or armchair meetings. Such meetings allow to establish communication between company management and heads of departments, on the one hand, and data analysts who are involved in maintaining BI systems, and also work with the received data, on the other.

8. Start small. Always start with small tasks, and with practice, you will understand how best to implement BI tools. Most likely, any data obtained in the first stages will entail additional questions. But it's even good: by finding ways to solve problems, you will improve your implementation plan.

9. Use Agile for Deployment. Most agile methodologies are focused on the capabilities of people and are well suited for deploying software, in particular business intelligence tools. It is recommended to build the project on the basis of adaptive planning, self-organization and implementation of short-term deadlines. It is also important that agile methodologies are aimed at continuous improvement in order to improve quality.

10. Determine data sources and required integration toolsю Every company, regardless of its size, constantly collects data about its activities.

Therefore, before implementing a BI system, it is important to determine from which sources this data should come. It is equally important to clearly formulate your expectations and assumptions about such data, and then test your hypothesis in practice.

It is also convenient to use cloud storage to improve access to data, improve usability and productivity.

11. Build your organization's data cultureю You need to constantly pay attention to whether everyone in your organization understands how important data is to achieving organizational goals. It is highly desirable that you maintain a data culture across all departments by encouraging them to focus on a common goal.

12. Lay a solid foundation for data managementю It is very important to provide a solid foundation for data management from the outset. First of all, you need to understand the value of data in the organization (both from internal and external sources) – especially if this data is related to new products, decision making and improvement of business processes.

Rules should be established that govern how data is collected, stored and processed. Reports on the creation, distribution and use of data should be generated taking into account these rules. In addition, they will help determine how to prioritize.

13. Implement a security management planю To protect the BI environment, companies need to plan a lot in advance. It is highly recommended that you implement the necessary security measures prior to implementing BI. A clear cybersecurity roadmap will indicate which actions require authentication or authorization and how to effectively take those measures.

14. Create a learning plan. It is very important that every user working with BI systems knows how to use this software correctly. Relevant information should be available to everyone – for example, in the form of downloadable content such as videos, pdf files, etc. At a minimum, a recording of a lesson conducted by a BI expert should be offered.

15. Select the person responsible for the BI implementationю Your company can assign a dedicated employee to help both executives and frontline executives implement and use BI in the best possible way. He must have technical knowledge of the platform and all business intelligence software. Also, this person will be responsible for training employees and the selection of educational resources for all departments. In addition, the

employee must have a good understanding of how and why people with different job responsibilities use BI to help create templates for analytical reports that meet the needs of different users [16].

1.3. Market analysis of software systems for business intelligence

1.3.1. System analysis Microsoft Power BI

The business intelligence tool in Office 365. Transforms your organization's data into compelling, powerful visualizations to collect and organize, so you can focus on what matters most (Fig. 1.1) [15].

Fig. 1.1. System window example Microsoft Power BI

1.3.2. System analysis Qlik Sense

A visual analytics platform that supports self-service visualization and data exploration. Allows you to quickly create combinations of visual representations, deeply explore data, instantly identify relationships and view opportunities from any perspective (Fig. 1.2).

Fig. 1.2. System window example Qlik Sense

1.3.3. System analysis Roistat

End-to-end business intelligence system. Collects data from the CRM system, advertising platforms and your website. From this data, it generates reports on key business indicators (Fig. 1.3).

Fig. 1.3. System window example Roistat

1.3.4. System analysis GetReport

Service for collecting corporate reports. Collecting data through web forms. Monitoring data collection in real time. Report and Chart Designer (BI). Possibility of integration with internal programs (API) (Fig. 1.4).

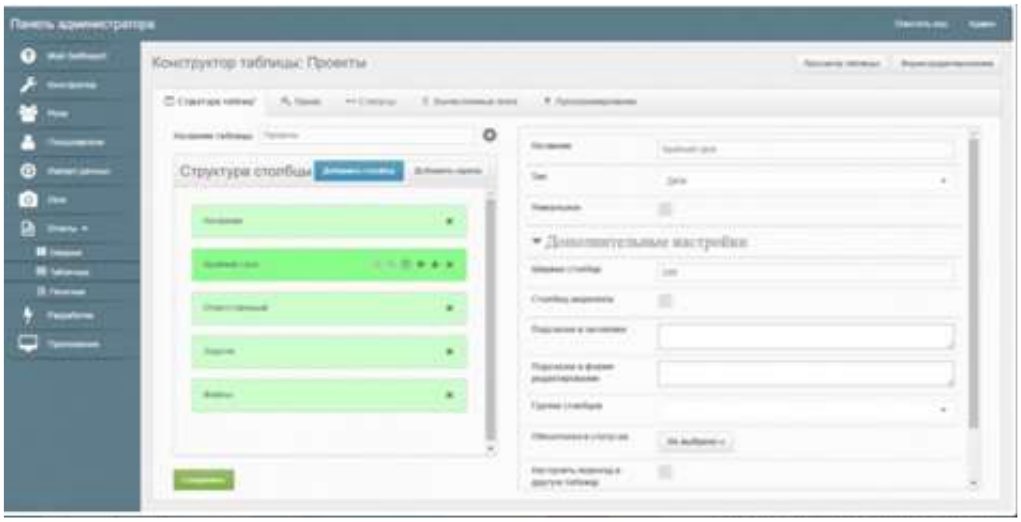


Fig. 1.4. System window example GetReport

1.3.5. System analysis Seeneco

Cloud service for business finance management, including management accounting, financial analytics and cash planning (Fig. 1.5) [24].

Fig. 1.5. System window example Seeneco

1.3.6. System analysis Business Scanner

Analytics service for business. Helps managers make management decisions, assess employee performance, and identify risks based on data, not intuition. Instant creation of reports (sales, finance, marketing, logistics, production, etc.) (Fig. 1.6).

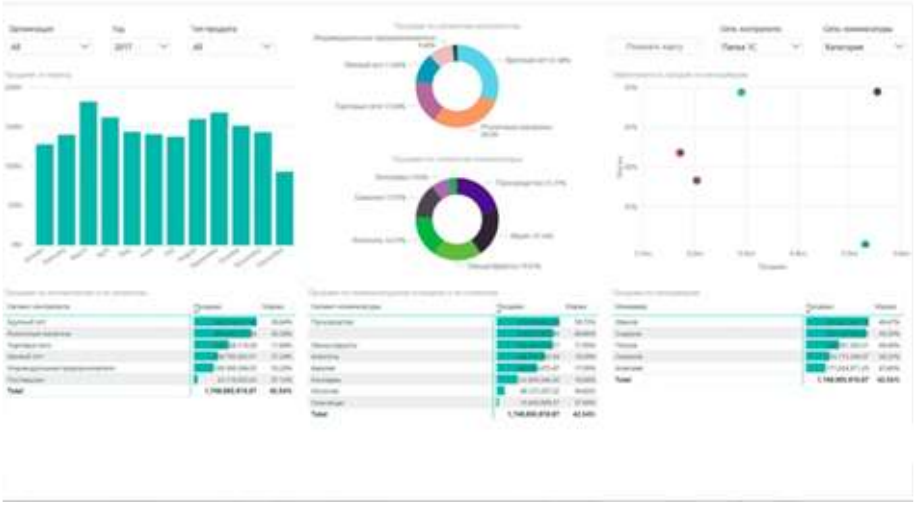


Fig. 1.6. System window example Business Scanner

1.3.7. System analysis Tibco Spotfire

Analytical platform with powerful analysis and visualization tools. Fully customizable for the business processes of a particular organization. Special analysis methods based on spreadsheets are available. There is a Russian interface and local implementers (Fig. 1.7).

Fig. 1.7. System window example Tibco Spotfire

1.3.8. System analysis SAP BusinessObjects

Flexible, scalable business intelligence (BI) system that allows you to find and share data for effective decision making. The product offers a wide range of tools on a single platform and allows IT departments to implement BI into any application or process in different environments (Fig. 1.8).

Fig. 1.8. System window example SAP BusinessObjects

1.3.9. System analysis IBM Cognos

A BI solution that provides a full range of BI capabilities in one product using a single architecture, allows you to define a generalized system of views on the decision-making process for the entire enterprise, reveals the relationship of all previous activities and the current state of the enterprise, and allows you to form the best solution that transforms strategy into action. (Fig. 1.9).

Fig. 1.9. System window example IBM Cognos

1.3.10. System analysisOracle Analytics

An integrated set of analytical tools designed to provide a better vision and understanding of the business to a wide range of users and allows any user of the organization to get fast web-based access to relevant information (Fig. 1.10).

Fig. 1.10. System window example Oracle Analytics

Conclusions on the First Part

With the speed of technology development that we are seeing today, it will become very soon difficult or even impossible for many companies to operate without business intelligence tools.

Companies that implement the appropriate toolkit by following the guidelines above will be able to design systems that will benefit all staff and drive business productivity and profitability.

An analysis of the principles of building information systems to support the analytical department and a review of the market for implemented systems for business intelligence made it possible to determine the signs of a list of the main functions of the system:

– the ability to analyze your mistakes and maximize retention;

– the ability to analyze success and maximize income;

– system of automated notifications;

– a system of recommendations;

– the ability to compare with the market;

– the ability to predict user actions and predict the values of indicators;

– the ability to communicate with end users;

– access to the original data;

– analysis of both traffic and user behavior;

– simplicity and accessibility of the interface;

– technical support;

– data accuracy.

PART 2

CHOOSING A DATA WAREHOUSE ARCHITECTURE

Cloud computing is a distributed data processing technology in which computer resources and power are provided to the user as an Internet service. If we explain it in an accessible language, then this is yours, in a sense, a working platform on the Internet, or rather on a remote server.

According to Drenser research, in 2019, users prefer to work with BI on a subscription basis. This scheme is chosen by 90% of the respondents who participated in the survey. The cloud-based approach allows you to eliminate the risks of payback problems, as well as try a new product using test licenses [18].

BI systems developers also support Trial schemes, as they increase the potential for subsequent purchase of licenses even for small companies that would not have dared to use BI in their practice earlier. According to analysts, the leaders in cloud BI are Amazon AWS, Microsoft Azure, Google Cloud and IBM Bluemix.

A data warehouse is a system that collects data from various sources within a company and uses this data to support management decisions.

Companies are increasingly moving to cloud storage instead of traditional on-premises systems. Cloud data storage has a number of differences from traditional storage:

– there is no need to buy physical equipment;

– cloud storage is faster and cheaper to set up and scale;

– cloud data warehouses can usually run complex analytic queries much faster because they use massively parallel processing.

In traditional architecture, there are three general data warehouse models: virtual warehouse, data mart, and enterprise data warehouse

A distributed system consists of a set of processors connected by a network. The communication network enables information exchange between processors. Communication delays are finite but unpredictable.

Processors do not share shared memory and communicate through messages over the network. There is no global physical clock to which all processors have constant access. The communication environment can deliver messages out of order, messages can be lost, distorted or duplicated due to retransmission, processors can fail, and communications can be broken [25].

2.1. Analyzing Traditional Data Warehouse Architecture

Quite often, the traditional data warehouse architecture has a three-tier structure, consisting of the following layers (Fig. 2.1):

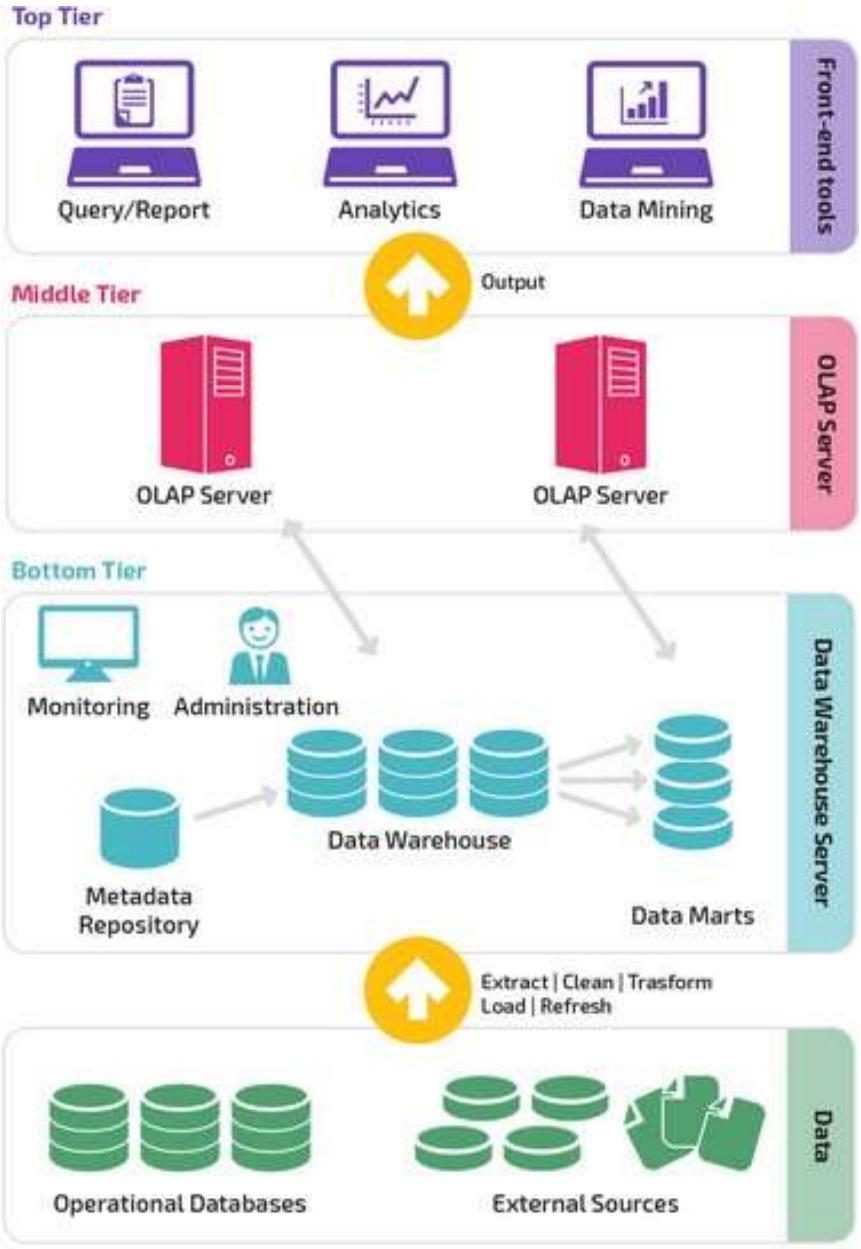


Fig. 2.1. Three-tier storage structure

– lower level: this level contains the database server used to retrieve data from many different sources, such as transactional databases used for front-end applications;

– middle tier: the middle tier contains an OLAP server that transforms data into a structure that is better suited for analysis and complex queries. OLAP Server can operate in two ways: either as an advanced relational database management system that maps operations on multidimensional data into standard relational operations (Relational OLAP), or using a multidimensional OLAP model that directly implements multidimensional data and operations;

– top level: the top level is the client level. This level contains tools used for high-level data analysis, reporting, and data analysis.

There are 2 main approaches to building data warehouses:

– Ralph Kimball's approach, which builds on the importance of data marts, which are data stores belonging to specific lines of business. A data warehouse is simply a combination of different data marts that facilitate reporting and analysis. The Kimball Data Warehouse Project takes a bottom-up approach;

– Bill Inmon's approach, which is based on the fact that the data warehouse is a centralized repository of all corporate data. With this approach, the organization first creates a normalized data warehouse model. Dimension data marts are then created based on the warehouse model. This is known as a top-down data warehouse approach[24].

2.2. Choosing a model and scheme for organizing data warehouses

In traditional architecture, there are three general data warehouse models: virtual warehouse, data mart, and enterprise data warehouse:

A virtual datastore is a collection of separate databases that can be shared so that a user can efficiently access all of the data as if it were stored in a single datastore;

The data mart model is used for reporting and analyzing specific business lines. In this warehouse model, aggregated data from a number of source systems related to a specific business area, such as sales or finance [22].

The enterprise data warehouse model assumes storage of aggregated data that covers the entire organization. This model views the data warehouse as the heart of the enterprise information system with integrated data from all business units.

Star and snowflake schemas are two ways to structure your data warehouse.

A star schema has a centralized data store that is stored in a fact table. The schema splits the fact table into a series of denormalized dimension tables. The fact table contains the aggregated data that will be used for reporting, and the dimension table describes the stored data.

Denormalized projects are less complex because the data is grouped. The fact table uses only one link to attach to each dimension table. The simpler star schema design makes it much easier to write complex queries (Fig. 2.2).

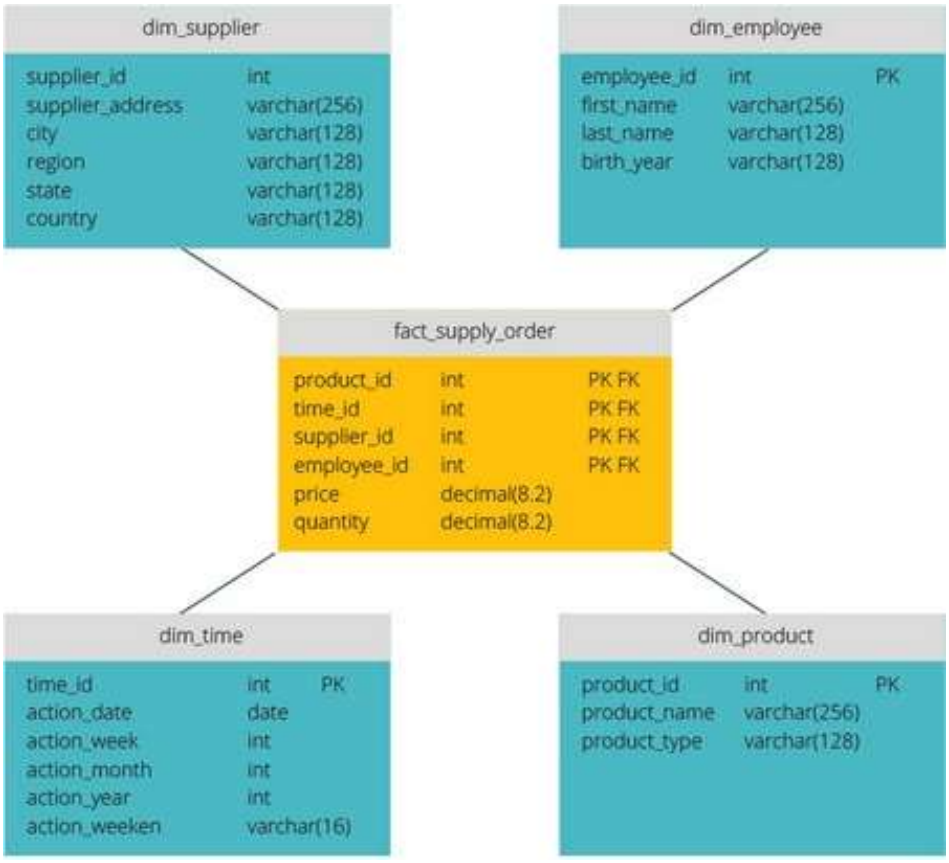


Fig. 2.2. Star Data Warehouse Schema Example

A snowflake schema is different in that it uses normalized data. Normalization means organizing data efficiently so that all data dependencies are defined and each table contains a minimum of redundancy. In this way, the individual dimension tables are forked into separate dimension tables.

The snowflake scheme uses less disk space and better preserves data integrity. The main drawback is the complexity of the queries required to access the data – each query must go through multiple table joins to get the corresponding data.

2.3. Choosing a method for loading data into storage

There are two different ways to load data into the warehouse: ETL and ELT.

ETL (Extract, Transform, Load) first retrieves data from a pool of data sources. The data is stored in a temporary staging database [21]. Transformation operations are then performed to structure and transform the data into an appropriate form for the target data warehouse system. The structured data is then loaded into the warehouse and ready for analysis (Fig. 2.3).

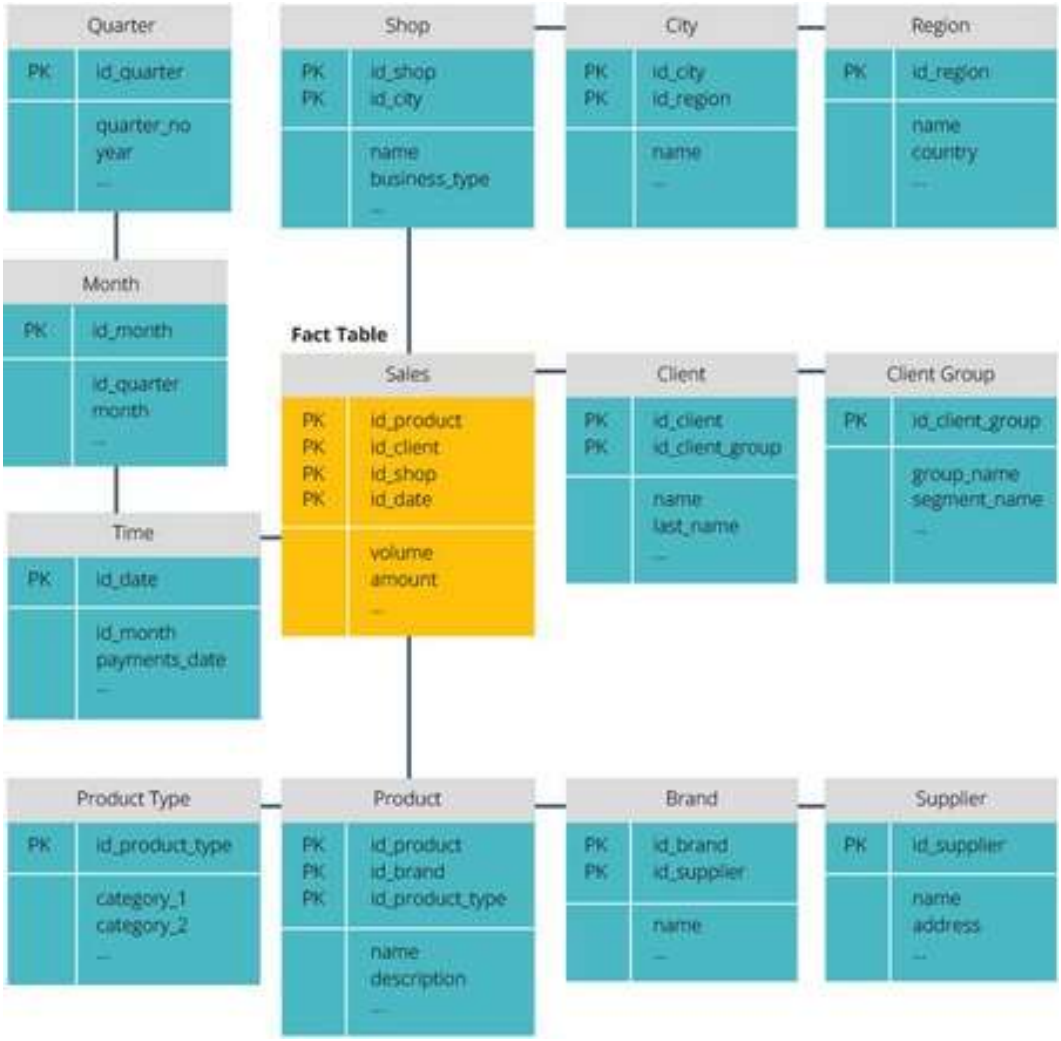


Fig. 2.3. Sample Snowflake Data Warehouse Schema

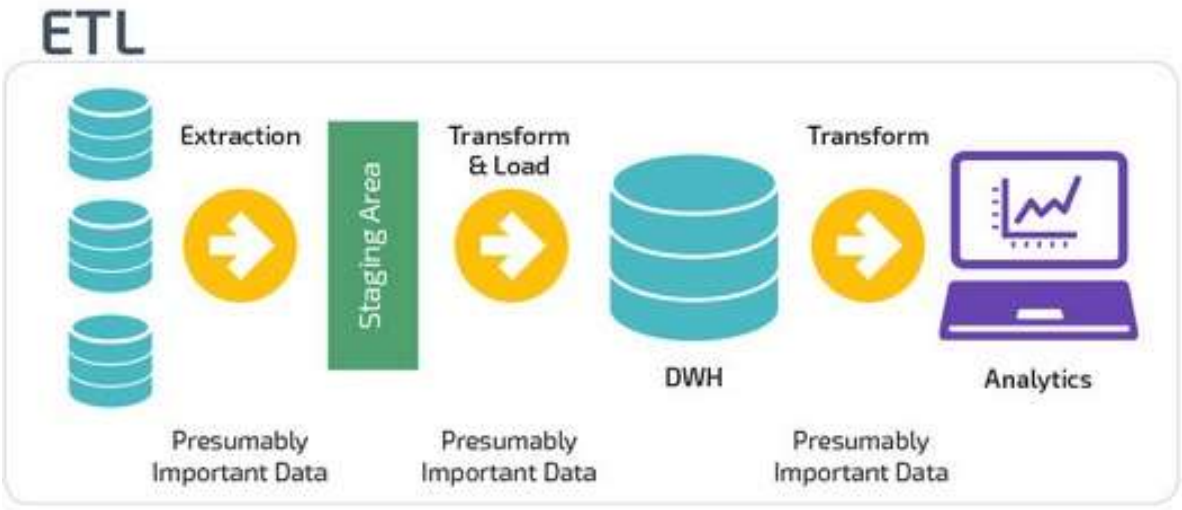


Fig. 2.4. ETL data loading schema

In the case of ELT (Extract, Load, Transform), data is loaded immediately after being extracted from the original data pools. There is no staging database, which means that the data is immediately loaded into a single centralized repository.

The data is transformed in a data warehouse system for use with business intelligence and analytics tools.

Fig. 2.5. ELT data loading scheme

The structure of an organization's data warehouse also depends on its current situation and needs.

The basic structure (Fig. 2.6) allows end users of the warehouse to directly access, report, and analyze summary data from the source systems. This structure is useful for cases where data sources come from the same types of database systems.

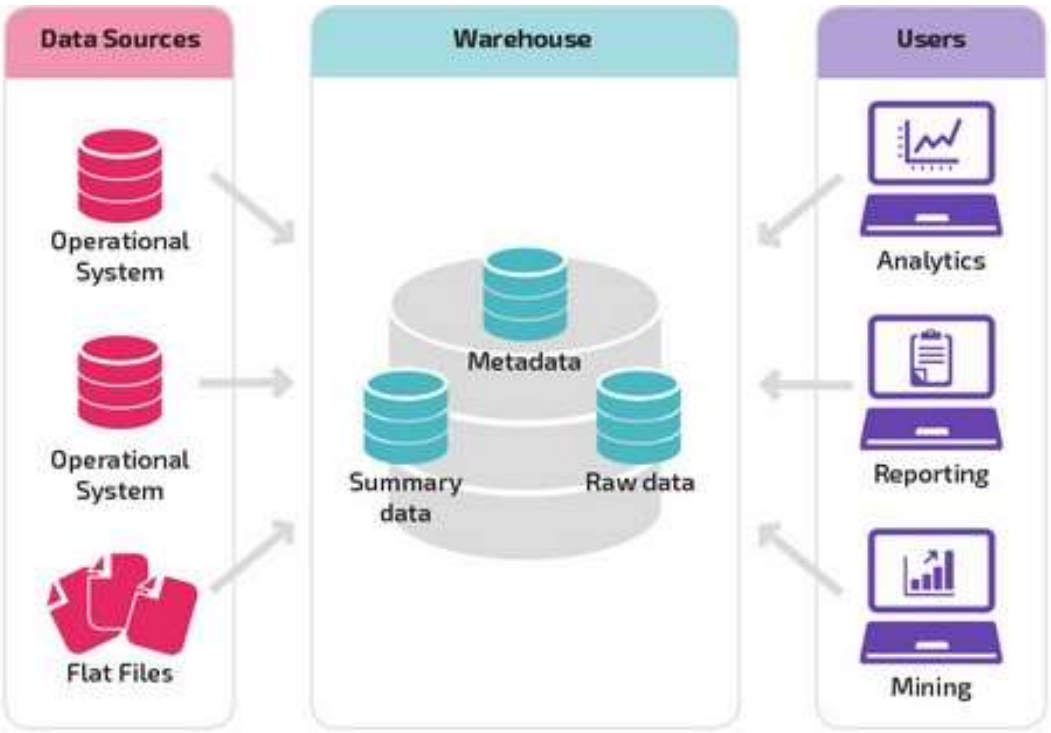


Fig. 2.6. Basic data warehouse structure

Staging area storage (Fig. 2.7) is the next logical step in an organization with heterogeneous data sources with many different types and formats of data. The staging area converts the data into a generalized, structured format that is easier to query using analysis and reporting tools.

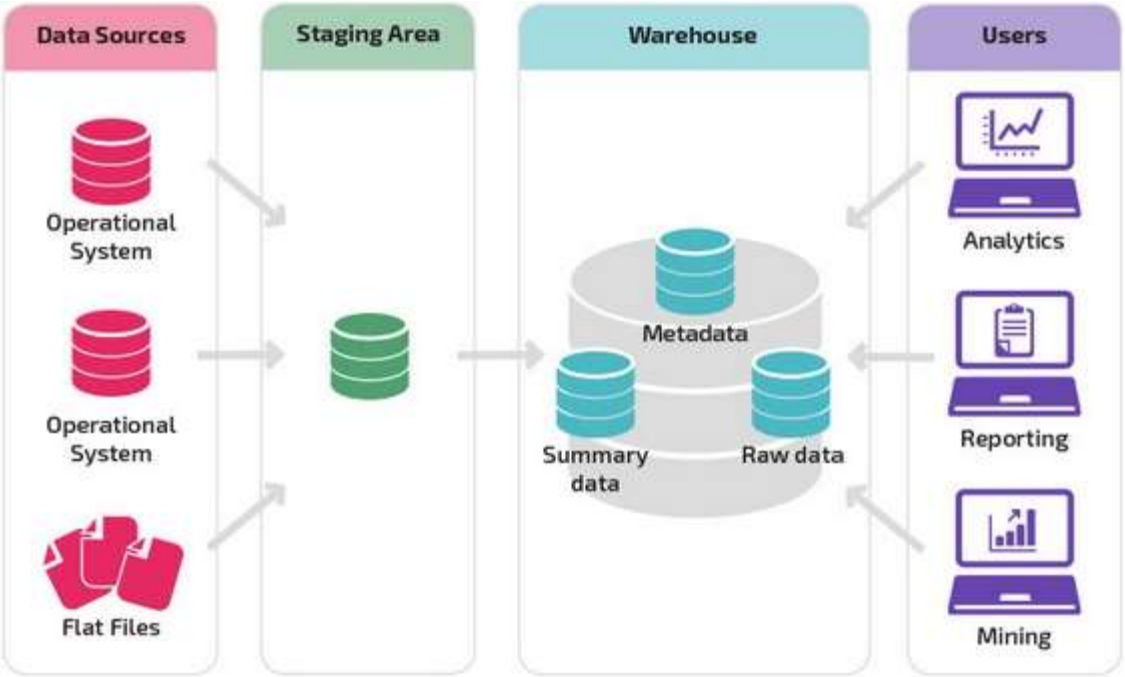


Fig. 2.7. Data warehouse structure with staging area

One type of middleware is adding data marts to a data warehouse. Data marts store summary data for a specific industry, making this data readily available for specific forms of analysis.

For example, adding data marts can enable financial analysts to more easily query detailed sales data and predict customer behavior. Data marts facilitate analysis by tailoring data specifically to meet the needs of the end user.

2.4. Analyzing Amazon Redshift and Google BigQuery data warehouse architectures.

2.4.1. Architecture Amazon Redshift

Amazon Redshift is a cloud-based representation of a traditional data warehouse. Redshift requires compute resources to be provisioned and configured as clusters that contain a collection of one or more nodes [12]. Each node has its own processor, memory, and RAM. The Leader Node compiles requests and passes them on to compute nodes that execute the requests (Fig. 2.8).

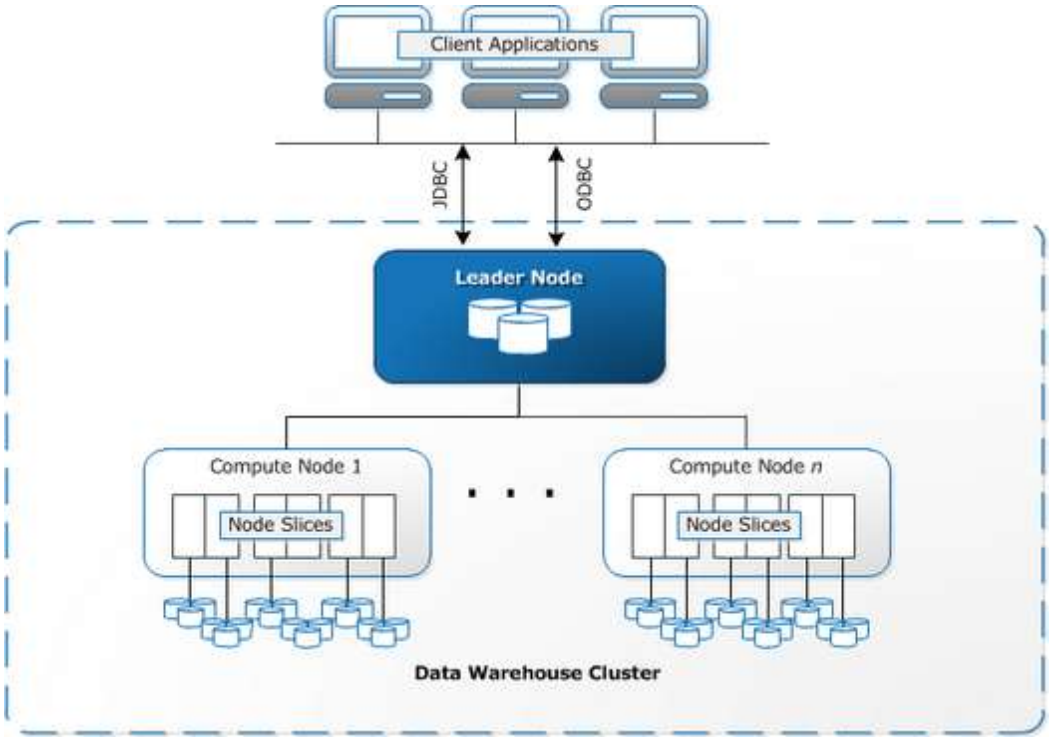


Fig. 2.8. Redshift data warehouse architecture

At each node, data is stored in blocks called slices. Redshift uses columnar storage, which means that each block of data contains values from one column across multiple rows, rather than from a single row with values from multiple columns.

Redshift uses the MPP (Massively Parallel Processing) architecture, breaking up large datasets into chunks that are assigned to slices at each node. Queries are faster because the compute nodes process queries in each slice at the same time. The Leader Node combines the results and returns them to the client application.

Client applications like BI and analytic tools can connect directly to Redshift using open source PostgreSQL JDBC and ODBC drivers. Thus, analysts can perform their tasks directly on the Redshift data.

Redshift can only load structured data. You can load data into Redshift using pre-integrated systems, including Amazon S3 and DynamoDB, by transferring data from any localhost with an SSH connection, or by integrating other data sources using the Redshift API.

2.4.2. Architecture Google BigQuery

BigQuery's architecture doesn't require a server, which means Google dynamically manages the allocation of computer resources. Therefore, all resource management decisions are hidden from the user (Fig. 2.9).

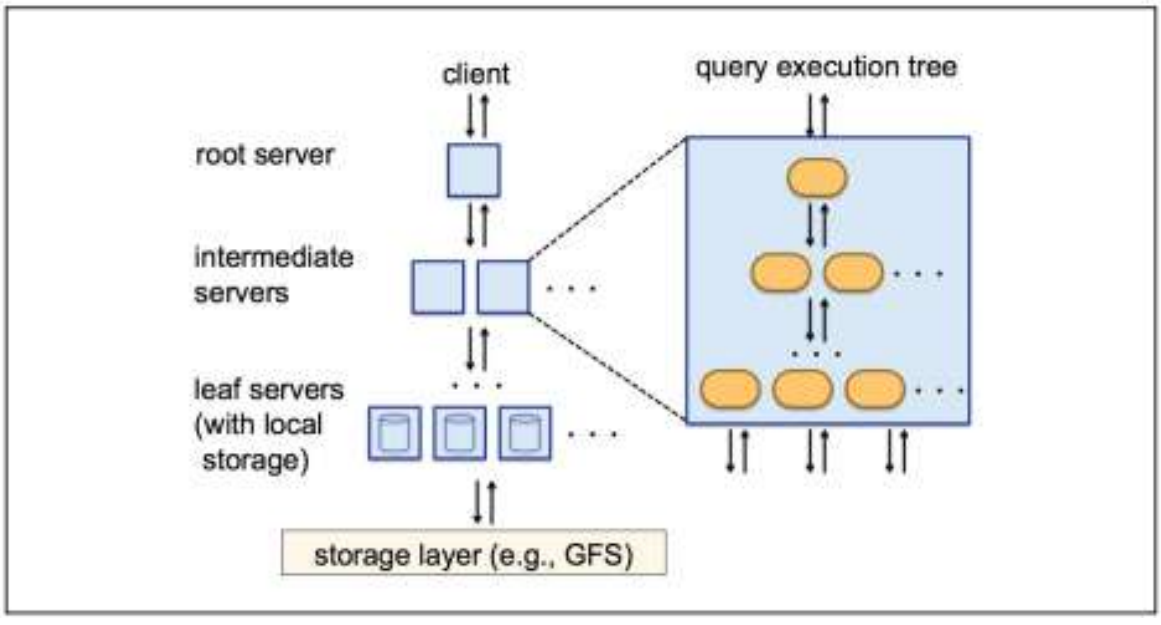


Fig. 2.9. Data warehouse architecture BigQuery

BigQuery allows customers to download data from Google Cloud Storage and other readable data sources. The alternative is data streaming, which allows developers to add data to the data store in real time, line by line, as it becomes available.

BigQuery uses a query engine called Dremel that can scan billions of rows of data in just a few seconds. Dremel uses massively parallel queries to scan data in the underlying Colossus file management system. Colossus splits files into 64 megabyte chunks among a variety of computing resources called nodes, which are grouped into clusters.

Dremel uses a columnar data structure similar to Redshift. The tree-like architecture sends requests to thousands of machines in seconds.

Simple SQL commands are used to query the data.

The architecture of modern distributed and cloud storage of information generates special security requirements for the data storage model. The public cloud infrastructure, on the one hand, allows achieving high scalability and storage efficiency, on the other hand, it makes it necessary to protect user data from unauthorized access by other users or information leaks due to an attack on the data center. A secure storage model must be able to use the cloud without trusting it. Any computers connected to a distributed system, as well as any servers in the system, are considered potentially insecure.

One of the main concerns of using cloud storage is the potential for potential data leaks in multi-user environments. Sharing hardware and transferring information to a potentially untrusted provider, as well as the possibility of an attack on an inadequately protected data center, are also big risks. This problem is solved by encrypting user data on the client side before sending it to the cloud storage. There are many open source and closed commercial solutions for the implementation of basic cryptographic operations (open source - OpenSSL, Libgrypt, Crypto ++, GnuTLS, commercial - pgp sdk, cryptlib).

The key point of protecting cloud systems is a mechanism for identifying and delineating user rights. Even encrypted data can be vulnerable if stored with other users' data. When sharing the same account or gaining access to an account, the data cannot be decrypted but can be deleted or changed.

2.5. Comparison of the cloud and own servers

Based on analytical studies, information was generalized to compare cloud solutions with their own servers. A number of parameters have been introduced, which are considered in detail [23]:

1) Cost - It is often said that clouds can save money on IT. This does happen, but not always. To benefit from cloud hosting, you need to understand the mechanism of its formation: after all, there are no miracles, especially in the economy. If you not only reduce your costs, but also let the intermediary between you and the iron (the cloud operator) earn bread and butter, then this can be explained as follows.

- Reducing the share of conditionally fixed costs. There are components in the costs of the computing process that are weakly dependent on the number of working servers - conditionally fixed costs (object and information security, rental of premises, the number of administrators, etc.). And those that strongly depend on the amount of computation - variable costs (disks, processors, electricity, software subscription, etc.);

- Resubscription of resources - a wonderful property of virtualization systems - the redistribution of idle resources. For example, if there are two cores, then it is possible to give them to two virtual machines at once, but only on the condition that they will not use them at the same time. The virtual machines will be sure they have 4 cores for two. Overwriting RAM and disk is more dangerous and difficult, but it is also quite possible. Many cloud operators only allow you to pay for the time your machine has been running, down to the minute or hour. Pay can also increase with intensive machine operation and decrease during rest moments. When one user's computations are idle, the cloud operator loads the hardware with tasks from another client.

- Discounts on hardware and software from manufacturers - all cloud keepers act as wholesale buyers. Hardware and software manufacturers are ready to give cloud operators, who take their products in the hundreds and thousands, a discount much higher than the market.

Total: savings when using the cloud depends on the volume. The less your need for computing resources, the more profitable it is to rent them. It is not for nothing that large players in the IT market are building their own data centers.

2) Flexibility. The cost considerations are valid for long term server hosting. But it also happens that the power is needed once, for a short time: for some event, scientific work, etc. The Olympiad, World Championship, Universiade or mass remote work can be cited as major examples. Therefore, even if a large number of servers and infrastructure are required, but for a short period after which they will become useless, then purchasing your own resources and building engineering systems will hardly pay off (and it is not a fact that you will have time to do this).

Total: for urgent and one-time events, the rental of computing power is well suited. Clouds are more flexible to your needs. And this is not only a question of price, but also of timing: the operator already has qualified personnel, scaling plans and experience that cannot be bought quickly even for a lot of money.

3) Capital CapEx and operating OpEx. Some companies very often convert capital costs (CapEx) to operating costs (OpEx) when estimating costs. It is preferable for them to pay regularly in small portions, even if there is an overpayment, than to buy some resource at once and then amortize it for a long time, i.e. renting a server is more profitable than buying it.

But everything suddenly changes when it comes to state or semi-state companies. In most cases, they prefer CapEx over operating costs. At least this is the case in education, healthcare, science, etc. The fact that the purchased expensive equipment has a finite term of the recommended work, as well as the fact that it does not need free maintenance, repair (with a very rough estimate, they put 10-15% of the cost of equipment for each year of operation), they are usually not taken into account.

Total: the cloud is better for those who prefer OpEx, and its own infrastructure is better for CapEx lovers. But this is only in terms of financial strategy, and not in terms of "more expensive - cheaper", which we have already discussed above.

3) Availability - For some reason, there is a common perception among executives that the clouds are extremely accessible. But the problem is that anything and anyone can fail (even Amazon had idle zones for several days).

Due to the multitude of risks, some companies maintain replicas of their resources from different operators in different countries (mainly to the markets of which their activities are oriented).

The opinion about the reliability of clouds is greatly exaggerated. They suffer not only from technical failures, but also due to the actions of owners, authorities and even other tenants (there are clouds with a toxic reputation that are blocked by firewalls and antiviruses). If the availability of services is critical for you, then when using the cloud you should have a plan for an emergency transfer of resources to your own capacities or to another operator.

4) Network load - This factor is sometimes underestimated when planning a move to the cloud. We must not forget about these processes:

- taking a backup out of the cloud (all clouds offer backup at their capacities, but if you value the data);

- replication between sites, if one day you decide to deploy a mirror in another cloud or locally;

- if you transfer workplaces of employees of an organization or other capacious processes to the cloud, then you need to take care of better and more reliable channels no longer in the cloud, but in the office.

When assessing the cost of clouds, you need to consider the cost of a productive and reliable network. This can significantly affect the final cost of the solution.

5) Control over data -transferring data from one cloud to another is difficult. Firstly, due to the limited network bandwidth (imagine that you needed to pick up a few TB in the current state, and do it without stopping the virtual machine). Secondly, due to the lack of universal tools. Thirdly, if the cloud has "fallen", then you already have no tools to extract data.

If there is a fear of physical seizure or blocking of data, then a remote cloud will help you. But it can also become a problem if there are problems with its availability.

6) Financial security - for large companies, changing the cloud is a whole story that can drag on for years. The cloud operator understands this, so he can dictate any prices.

You need to be careful about cloud discounts and promotions. They come and go, and the infrastructure is long term.

7) Requirements for personnel - when you maintain a small fleet of your own servers (for example, a couple of productive servers and storage systems) you may face

a personnel problem. Simply will have nowhere to take personnel of the appropriate qualifications for competent maintenance of hardware, virtualization and backup environments. And the problem isn't just money. The server administration industry is changing rapidly. Teed at least 2-3 people who can insure each other on vacation and sick leave.

For small projects, cloud deployment of machines, taking backups, etc. will take a couple of clicks and allow you to save on staff. But for complex projects, the interaction of several clouds with replication and monitoring may, on the contrary, require a larger number of people with completely different qualifications.

Clouds are relevant if the volume of maintenance does not allow maintaining at least a small department of hardware administrators with a more or less uniform load. In this case, you can save on staff. But moving to the cloud for large projects will not necessarily lead to a reduction in staff, and on the contrary, it may increase it.

8) Compliance with SLA - Service Level Agreement, it is also SLA, it is also an agreement on the quality of services is not always possible to check.Cannot afford to constantly run benchmarks, and simply observing the performance of a virtual machine may not always indicate problems. The operator may deliberately or inadvertently limit the read speed, CPU quota provided, network bandwidth, and so on.

Cloud SLA may fall short of expectations due to hidden savings from the operator. Proving this can be difficult or impossible.

9) Legal requirements - many states have requirements for mandatory protection of information (personal data, bank secrets, etc.). Compliance with national requirements is best (and sometimes only possible) to be done in a national cloud or in your own data center. For example, it is problematic to process data of Ukrainian citizens in the Amazon cloud. At least a copy of them must be located on domestic servers.

Choosing a cloud to comply with the requirements of the law in the field of information security is a very delicate matter. But at least the choice of the cloud, which takes over the protection of information, removes from you, if not technical, then at least legal risks.

10) Use of available resources - Some engineering and scientific systems are very, very poorly surviving moving to the cloud. Either for security reasons (burglar alarm, access control system - ACS, automated process control systems - APCS) or for cost reasons (object video surveillance, if it has hundreds and thousands of cameras, special computers, supercomputers, multi-terabyte results computed tomography, etc.). The organization operating them is simply forced to maintain the engineering infrastructure of a server room or a small data center. These costs refer to conditionally fixed costs, therefore, if they already exist, then the organization can use them to organize other computing processes.

If you already have an engineering infrastructure, then the transition to the cloud becomes less profitable. The savings on infrastructure maintenance simply will not come or will be small. Perhaps, in this case, it is worth keeping only some backup servers in the cloud if the reliability of the engineering infrastructure does not fully correspond to the required availability.

The system should provide means of integrity control to identify the fact that no unauthorized modification of the data has occurred, as well as the ability to control the authenticity of the data, that is, to verify that the data has been modified only by those users who have access to record this data.

The dissemination of information throughout the system and the sharing of resources by users should be simple and transparent to the user. This is solved by building an encryption key exchange mechanism. Secure key exchange between users can be implemented using a PKI server, for example, OpenSSL PKI or GnuPG key ring.

Cloud storage should send any unencrypted information over a secure connection, for example, using one of the SSL implementations (OpenSSL, GnuTLS, cryptlib). Cloud storage performance is also an important aspect.

Conclusions on the Second Part

Analysis of cloud data warehouses allowed us to identify a number of problems when setting them up:

– loading data into cloud data warehouses is not trivial, and large-scale data pipelines require configuration, testing and support of the ETL process. This part of the process is usually done by third-party tools;

– updates, inserts and deletes can be complex and must be done carefully to avoid degrading query performance.

– difficulties with processing semi-structured data – they need to be normalized in a relational database format, which requires the automation of large data streams;

– nested structures are usually not supported in cloud storage. You need to convert the nested tables to formats that the data warehouse understands;

– cluster optimization. There are various options for configuring a Redshift cluster to run your workloads. Different workloads, datasets, or even different types of queries may require different customization. To achieve optimal performance, it is necessary to constantly review and, if necessary, further adjust the configuration;

– query optimization – custom queries may not be in line with best practices and will therefore take much longer to complete.

– backup and recovery – although storage vendors provide many options for backing up your data, they are non-trivial to configure and require monitoring and close attention.

PART 3

THE IMPLEMENTATION OF THE INFORMATION SYSTEM OF THE ANALYTICAL DEPARTMENT USING CLOUD TECHNOLOGIES

3.1. Stages of database design for use in the system

3.1.1. Basics of database design

The process of building database applications should be formally divided into several stages [13]:

1) Defining the purpose and objectives of the application.

2) Designing the database structure and application processes required to implement these tasks.

3) Implementation of the project in the application by building the necessary database objects and program objects.

4) Testing applications for compliance with the objectives.

5) Installation of applications for user operation.

The first step in designing any application requires a clear definition of the purpose of the application and the specific tasks it will solve. The next step is to divide the tasks into operational processes.

The application design process is closely intertwined with the database design process.

When building a database, first determine the basic operating processes and tables to service these processes and what fields they should contain. This process is called data modeling. There are several approaches to data modeling. Consider two approaches – modeling of the central application and types of modeling. It also makes sense to build a schematic representation of the complete database project and build data flow diagrams for the processes that make up the application.

The simulation of the central application meets the needs of the application in data through a single database. Each application has its own database or set of tables.

Species modeling organizes tables in databases according to the types of data they represent. For example, a table that has a list of products manufactured by a company may be in a database related to trade activities, while a description of components for the manufacture of products, employees, is in the production database. The report, which shows the total profit for each unit of production, should be based on data from both tables, despite the fact that they are placed in completely different databases.

A key element in designing databases is a table. Each table can be represented as a plane, drawn in rows and columns. As in a spreadsheet, rows represent records in the table, and columns represent fields.

The record is the main content of the table, it is associated with a real data object. This can be a bank account number or the customer's last name.

A field is an element inside a record, it represents the characteristics of the object represented by the record.

A field or set of fields that distinguishes one record in a table from another is called an element identifier. It uniquely identifies each record in the table. An example of an item ID is the account number field in the table of accounts.

The primary key is the field or fields of the table that are used as the element identifier. Like an identifier, the value of the table's primary key is always unique for each record. The fields that make up the primary key are used to build an index designed to quickly access the rows of the table.

A foreign key is a field or fields in a table that, when not used as an identifier, is often used when combined with other tables.

3.1.2. General description of the MySQL database management system

MySQL- very fast, reliable relational database management system. The database allows you to efficiently store, search, sort and retrieve data. The MySQL server controls access to data by allowing multiple users to work with it at the same time,

provides fast access to data, and guarantees access only to those users who have the right to do so. Accordingly, MySQL is a multi-user, multi-threaded server. It uses SQL, the standard database query language used around the world. MySQL appeared on the market in 1996, but its development began in 1979.

Currently, the MySQL package is available as open source software, but if necessary, you can obtain commercial licenses [6].

3.1.3. Building a database structure

When designing a database that is designed to serve the planned operating processes, you need to think about the elements that will have to remember these tables. Using the list of previously developed tables, you need to make a list of all information elements (fields) with which the application will work, determine the optimal data type and size for each field, making sure that there is no duplication of fields.

Distribute the fields from this list in the structure of the tables that have been outlined in advance. Divide all tables into base and transaction tables (changes). Base tables contain elements that are unique by definition. An example of a base table is a table listing the district capitals. Transaction tables receive data collected from hardware or entered by users. Their content is often subject to change.

3.1.4. Master type tables

Relational databases do not allow duplication of data within a specific table. Data should be duplicated at the tabular level. This follows from the fact that the primary key of one table must be duplicated by the primary or foreign key of another table in order to establish a relationship between them. These relationships are expressed in two varieties: one-to-one and one-to-many. One-to-many relationships form the basis for most database management applications. Tables focused on one-to-many relationships are called master and tables. Within a group of transactional tables, identify the subordinate tables, ie those that depend on others when presenting relevant information, and decide which tables are the main ones.

Some tables in relation to one table play the role of subordinates, and in relation to others are the main ones. Master tables are linked by a combination of primary and

foreign keys. An example of a relationship between a master table is a pair of tables, one of which functions as the master and stores the header information for the account, and the other is used as the stores the full characteristics for the account. It is very useful to display the relationships between tables graphically. The purpose of such a relationship diagram is to visually represent the relationships between the tables.

It also makes sense to build a schematic representation of the complete database project and build data flow diagrams for the processes that make up the application.

3.2. Information model and database structure

Any application designed to work with databases will be easier to use, easier to understand and faster to work with if its database has a clear, well-organized structure. The decision on the structure of the database was made at the design stage of the system.

Due to the need for connected work with an existing system, a large flow of information at the entrance and the direction to use this system in the network, the decision was made to use the database MySQL, and this was justified by a number of reasons:

1) ease of building databases;

2) the ability to create more powerful and secure database structures;

3) compatibility with the PHP programming language.

The database is located on the server's hard disk, which has constant access from the Internet. The choice of media type was justified by the following reasons:

1) high access speed;

2) high reliability of information integrity due to duplication;

3) constant creation of backups in automatic mode (2 times a day).

To place the database for the first year of operation of the system on the hard disk requires approximately 1 GB of free space for processed databases, although this figure

may vary depending on the amount of information processed. Each subsequent year will require 200 to 500 MB of hard disk space.

The process of creating an information model begins with defining the conceptual requirements of a number of users. Conceptual requirements can be determined for some tasks (applications), which are not planned to be implemented in the near future (perspective). This may slightly increase the complexity of the work, but does not prevent the most complete consideration of all the nuances of the functionality required for the development system, which will reduce the likelihood of further processing. The requirements of individual users are integrated into a single "generalized view". The latter is called a conceptual model.

The conceptual model represents objects and their relationships without describing ways to physically preserve them. Thus, the conceptual model is, in fact, a model of the subject area. When designing a conceptual model, all the efforts of the developer should be focused mainly on structuring the data and identifying the relationships between them without considering the specifics of implementation and issues of processing efficiency. The design of the conceptual model is based on the analysis of the data processing tasks solved at this enterprise. The conceptual model includes descriptions of objects and their relationships of interest in the subject area and the data identified as a result of the analysis. The conceptual model is then transformed into a data model compatible with the selected database. Perhaps, that reflected in the conceptual model of the relationship between objects will be subsequently unrealized means of the selected database. This will require a change in the conceptual model, a version of which, provided by a specific database, is called a logical model.

The logical (external) model reflects the logical relationships between data elements, regardless of their content and storage environment. Users are given subsets of this logical model, called external models, that reflect their perception of the subject area. The external model corresponds to the representations that users receive based on the logical model, while the conceptual requirements reflect the initial representations of the users that formed the basis for the development of the conceptual model. The logical model is mapped to the physical model.

The physical model that determines data placement, access methods, and indexing techniques is called the internal system model.

Let's carry out the conceptual analysis of necessary fields of a database. The operation of the contact accounting subsystem requires a set of information in the form of a database, which is accumulated by monitoring new orders and contacts. Information enters the subsystem in two ways:

1) in the form of documents drawn up in the prescribed form (for the convenience of storage and processing);

2) a separate entry from the registration form, which is adjusted after the successful conclusion of the agreement on the basis of the documents described in paragraph 1.

The basic information received about the actions of customers is a record of the operator, which is compiled in a certain form:

1) the client;

2) the contact person;

3) type of contact;

4) the result of contact;

5) date and time of contact;

6) a note on the purchase of the product or the conclusion of the agreement.

We will reveal the conceptual content of each of the points:

– Client = Company name, Country, City, Full address, Phone;

– Contact person = Company name, name, contact phone, position, e-mail address, city of residence;

– Contact type = Type, Contact description, Contact result type;

– Contact Result = Contact Result Type, Contact Result Description, Transaction Type;

– Date and time of contact = Date of first contact, Date of successful contact, Date of conclusion of the agreement.

Note on the purchase of a product or the conclusion of a transaction – Description of the transaction, Attached file of the transaction, Volume of the order, Amount of the transaction, Amount of discounts, Amount of deductions to intermediaries.

The database used is relational. This means that it has several tables that are linked by the values of certain fields. One field in the table is the primary key by which the information in the file is indexed.

Based on the conceptual model, we will build a database, which is a set of data tables that are interdependent on certain principles.

The main data table is CONTACTS, which is associated with the ACCOUNTS table with customer information, ACTIVITES is a table that expands the contact information (communication with CONTACTS 1 to 1), it is associated with the table OPPORTUNITES, which contains information about the content of the agreement or other productive results of the contact.

The formation of source documents is as follows. Based on the query to the system, it is determined which fields need to be displayed on the screen, based on this choice, a query is automatically generated, which contains certain fields. If you use fields from different tables, these tables are automatically merged in one of the merge scenarios.

3.3. Development of directories

The specifics of the subsystem requires that each table be associated with directories of two types – system and user. The difference is in the editing access rights: system edits are edited only by the database administrator, custom edits are edited automatically by the user when adding new parameters that were not previously used in the subsystem.

Presents the main system and user directories used in the system, indicating the type and description of the fields (Fig. 3.1).

OPP_TYPE : таблица

Имя поля	Тип данных	Описание
Opportunity Type	Счетчик	Код Типу контакту
Nazv Type	Текстовый	Назва типу контакту
Prim_OP_t	Текстовый	Примітка

JOB_ROL : таблица

Имя поля	Тип данных	Описание
JOB ROL	Счетчик	Код посади
Nazva JR	Текстовый	Назва посади
Otetstv	Текстовый	Описання обов'язків
Primitka	Текстовый	Примітка

Fig. 3.1. System directories

3.4. Relationships between database tables

This diagram shows the main types of connections that were used in the development of the subsystem.

But links between tables are used not only when creating a database, but also with each query.

The main methods that were implemented in the thesis were aimed at working with queries to the tables CONTACTS, ACTIVITES and OPPORTUNITES.

It is the relationships between these tables that have been used in queries to obtain reference information.

Thus, in order to implement the request for output of data on agreements, it is necessary to combine the tables CONTACTS, ACTIVITES and OPPORTUNITES, and if necessary, to help the company name, add the table ACCOUNTS. The connection diagram for this query is shown (Fig. 3.2).

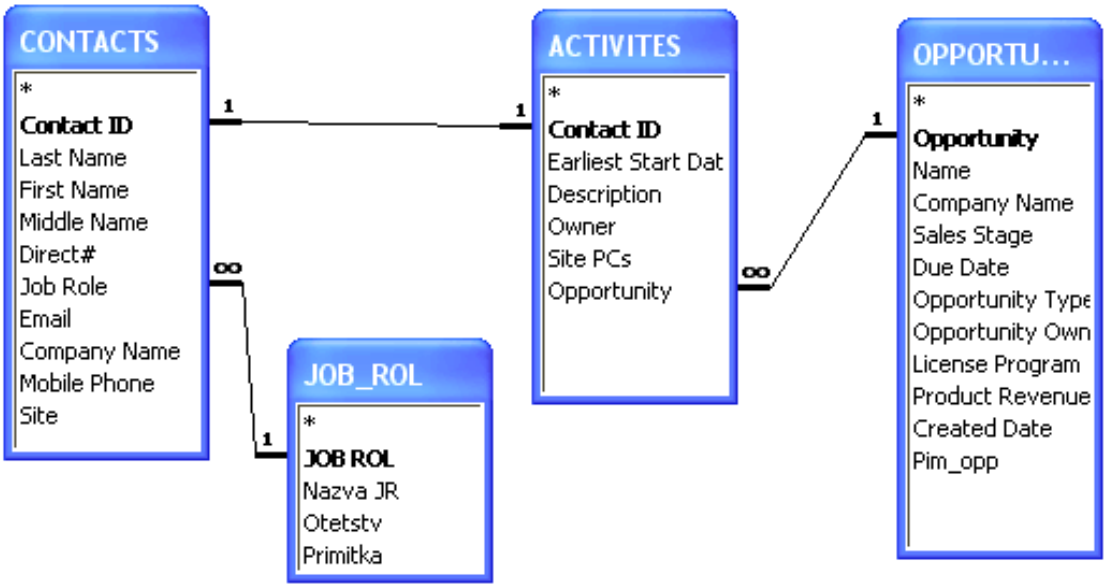


Fig. 3.2. Graphical representation of the connection scheme of the request for input and output of data on transactions

If you need information about the positions held by contact persons, it is enough to link the tables CONTACTS and JOB_ROL. If it is necessary to obtain only a list of positions, grouping by the field of the job title is used (Fig. 3.3).

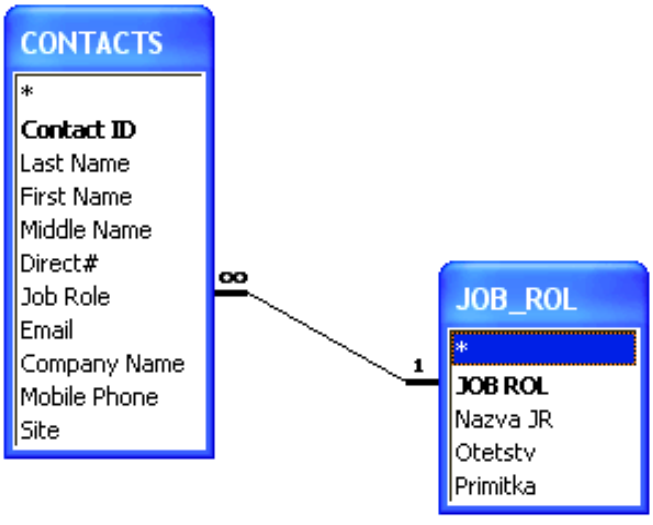


Fig. 3.3. Graphical representation of the connection scheme of the request to obtain information on the positions of contact persons

The general scheme of connections between the database tables used in the program is presented in fig. (Fig. 3.4).

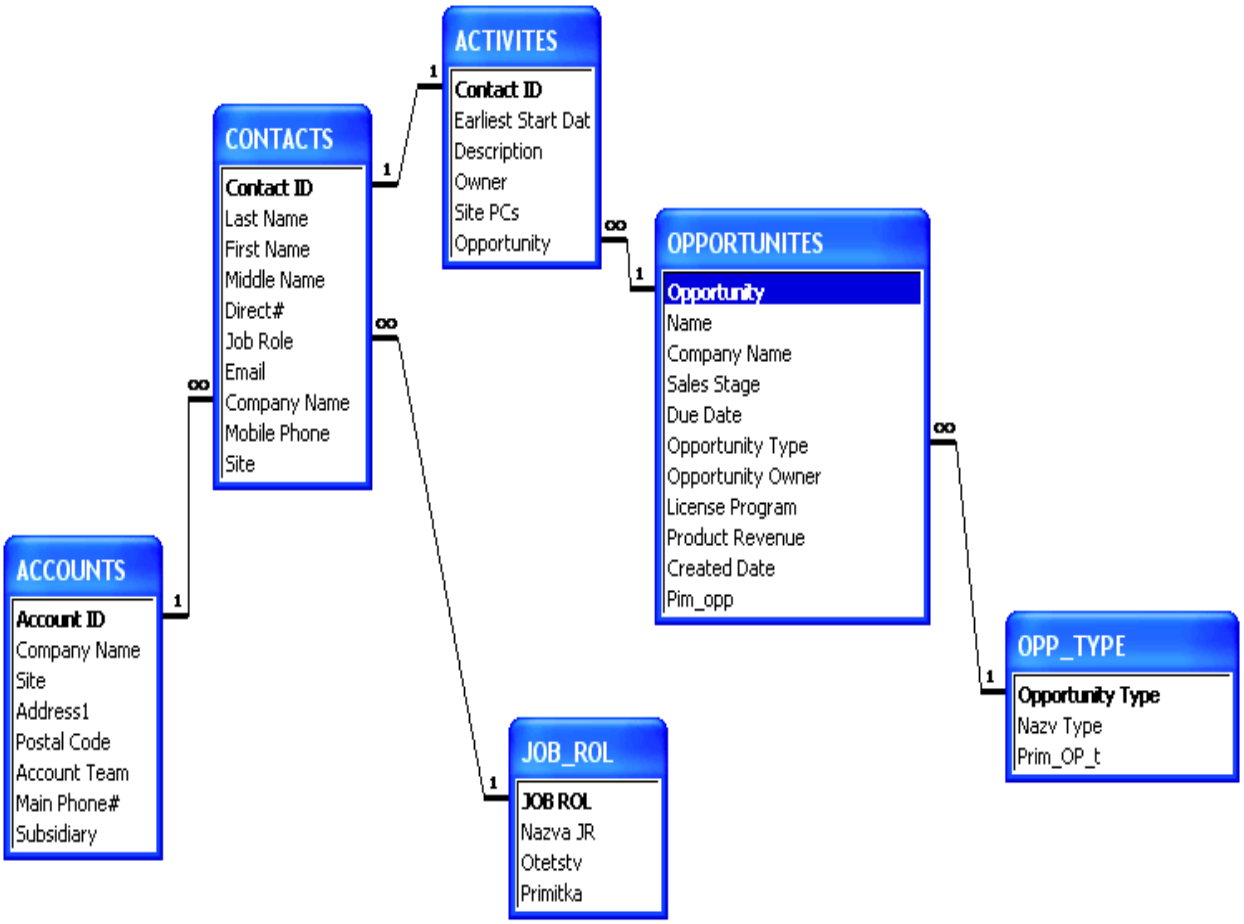


Fig. 3.4. General scheme of relationships between database tables used in the program

As you can see from the figure, the program uses only part of a large database.

To create any standard report, queries are automatically generated that use parameters from the fields of web pages.

3.5. Software development and software description

3.5.1. Description of the programming environment

PHP language was used when writing the program. PHP is a server-side scripting language designed specifically for the Web. You can include PHP code in the HTML

page, which will be executed each time you visit it. PHP code is interpreted by the Web server and generates HTML or other output that the page visitor sees.

PHP development began in 1994 and was first performed by one person, Rasmus Lerdorf. This language was adopted by a wide range of people and underwent three major editions until it became a mature product that is widely used and that we are dealing with today. As of January 2001, it was used in nearly five million domains worldwide and continues to grow. You can find out the number of domains that use PHP on the page www.php.net / usage.php.

PHP Is an open source product. Each user has access to the source code. It can be used, modified and freely distributed to other users or organizations.

Initially, the abbreviation PHP was an abbreviation of Personal Home Page, but later this name was changed in accordance with the recursive agreement to the name GNU (Gnu's Not Unix) and now means PHP Hypertext Preprocessor.

PHP's competitors include Perl, Active Server Pages (ASP), Java Server Pages, and Allaire Cold Fusion.

PHP has many advantages over these products, namely:

– high productivity;

– PHP is very effective. Using a single low-cost server, you can service millions of requests a day;

– availability of interfaces to many different database systems;

– PHP has a built-in connection to many databases (MySQL, PostgreSQL, mSQL, Oracle, dbm, Hyperware, Informix, InterBase, Sybase). You can also use the Open Database Connectivity Standard (ODBC) to connect to any database for which there is an ODBC driver. This applies to products from Microsoft and many other companies;

– built-in libraries to perform many common tasks related to the Web;

– with the help of PHP you can "on the fly" generate GIF-images, connect to other network services, send e-mails, work with cookies and generate PDF-documents – and all this with a few lines of code;

– cost;

– PHP package is free. The latest version can be downloaded for free at any time;

– ease of study and use;

– PHP syntax is based on other programming languages, primarily C and Perl. If you are already familiar with C, Perl or C-like language, such as C ++ or Java, you can use PHP almost immediately;

– mobility;

– PHP can be used under different operating systems. PHP code can be created in free Unix-like systems such as Linux and FreeBSD, commercial versions of Unix such as Solaris and IRIX, or various versions of Microsoft Windows. Typically, programs will run without any changes in different environments with PHP installed;

– availability of source code;

– the user has access to the PHP source code. Unlike commercial closed products, if you need to change or add something in this language, you can always do it. You can not wait until the manufacturer releases edits (patches). There is no need to worry that the manufacturer is going to leave the market or stop supporting the product.

3.5.2. Description of standard window forms

After logging in, one of the main windows is the Clients window (Fig. 3.5)

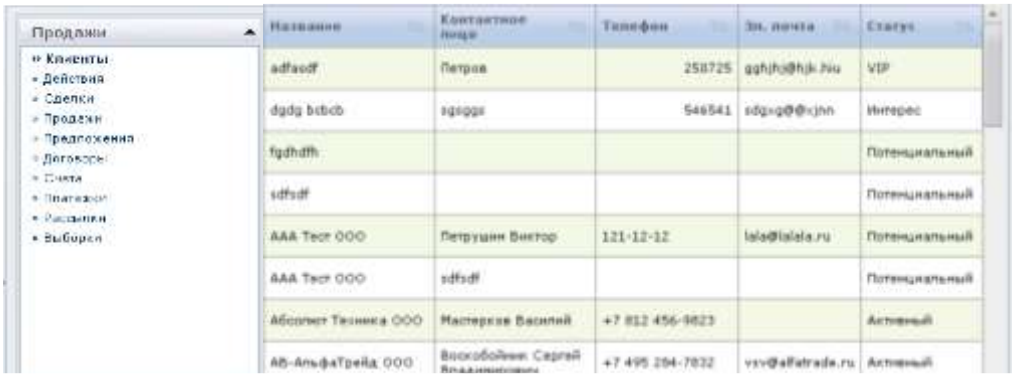


Fig. 3.5. Document List Window: Clients

This window contains basic information about customers and in addition a phone number, e-mail address and a subjective parameter as a status in which the manager conditionally indicates the degree of interest of the company in the customer.

Using the create command, you can enter a new client into the database. To do this, fill in the main fields in the form of Fig. 3.6 (those marked with a red asterisk): name, type of counterparty, status, date and time of registration, as well as enter all

known additional information (if no other data, they will be entered when working with the client).

The following mechanism was used to create a page with many nested windows:

my_c_menu.make_menu (name_u, parent_name_u, text_u, link_u, target_u, width_u, height_u, regImage_u, overImage_u, regClass_u, overClass_u, align_u, rows_u, nolink_u, onclick_u, onmouseover_u, onmouseout_u).

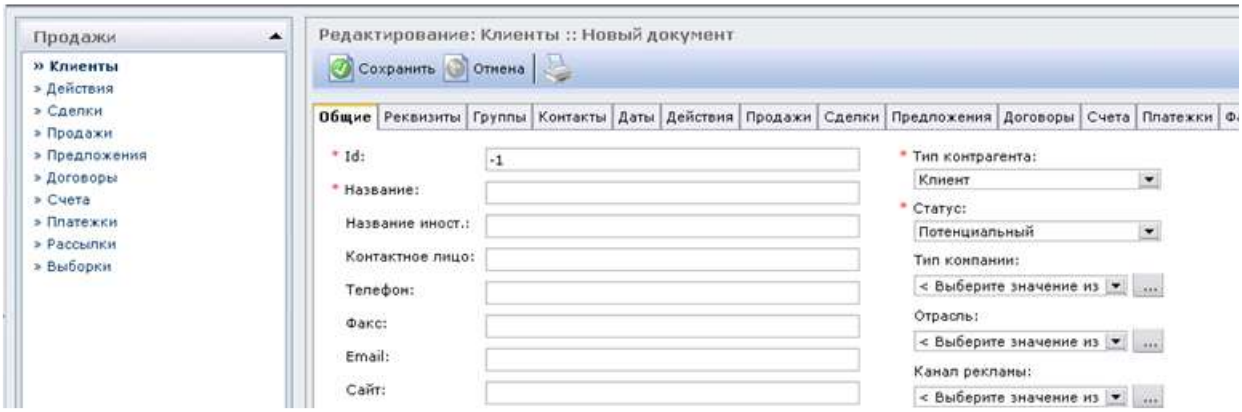


Fig. 3.6. Form of adding a new client

This mechanism is based on the use of JavaScript elements, which allow you to place several elements in one zone, and replace them when external calls to certain procedures or functions. This creates a page layout effect in which only one page is displayed at a time.

In the case of concluding a contract or at least an oral agreement, these data must also be entered as information about the client (Fig. 3.7).

The action section (Fig. 3.8) shows all the events that have taken place in the company since the beginning of the registration of information in the system.

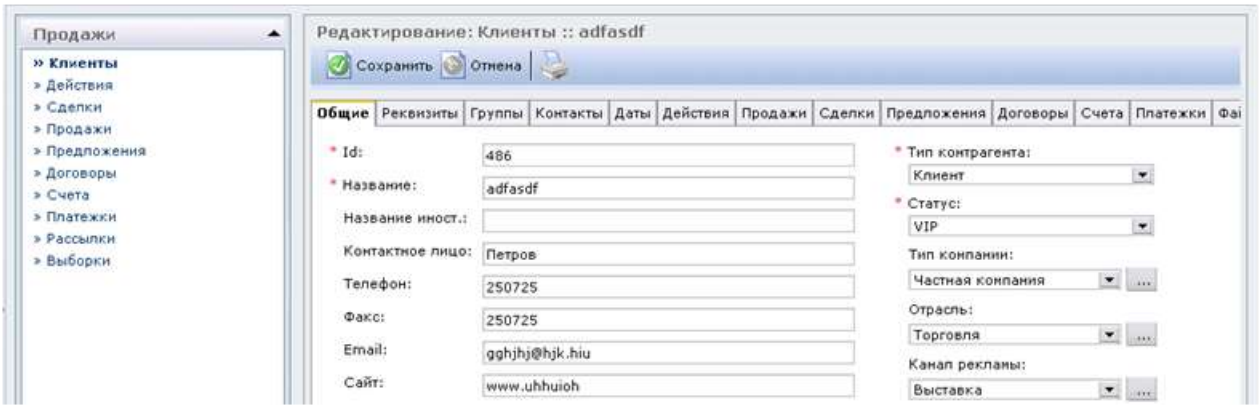


Fig. 3.7. Example of editing customer information

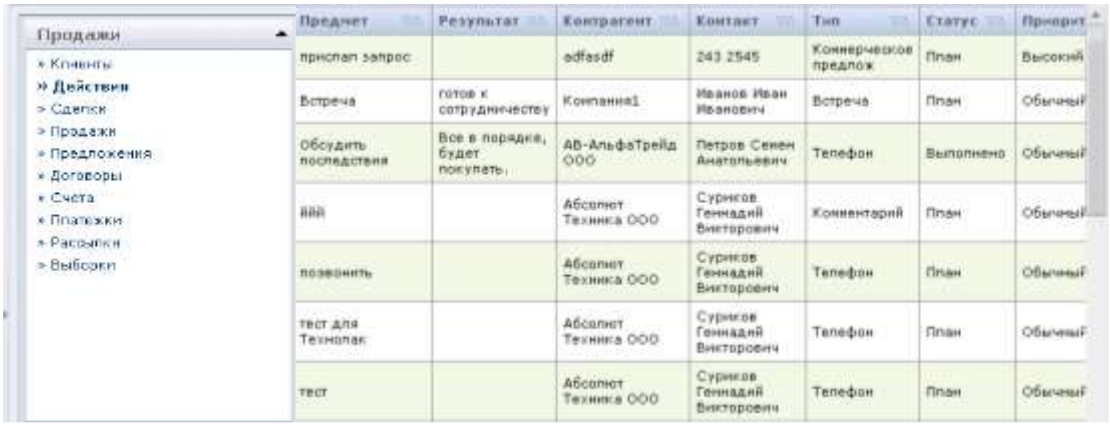


Fig. 3.8. Document List Window: Actions

Thus each event is described not only by standard elements (subject, type, status), and also the subjective estimation of the given action (Priority) is exposed and the free description of result and if necessary additional information on action is added (Fig. 3.9).

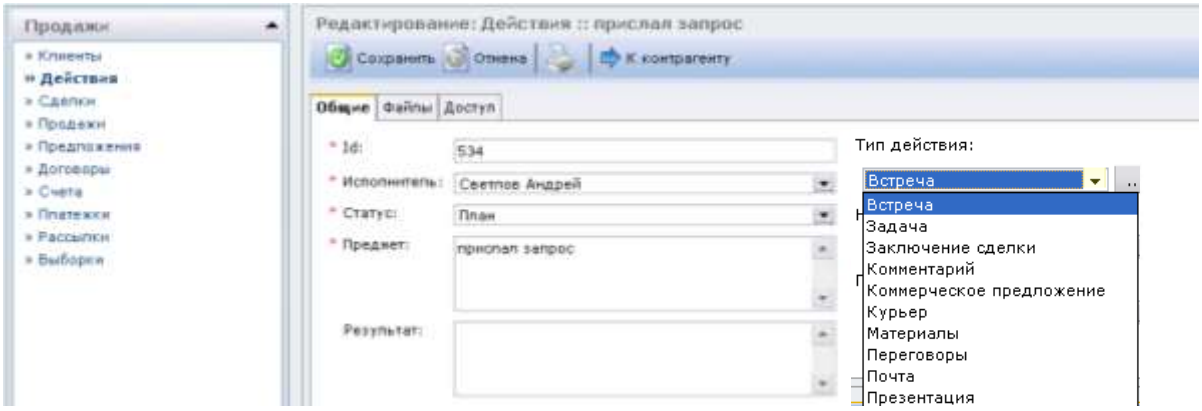


Fig. 3.9. Example of editing action information

A separate section highlights all the agreements that have been concluded with customers (Fig. 3.10). They must indicate the contractors and the status of the agreement (contact, order, etc.). Additional information on the cost and deadlines is also possible, but in the case of a final agreement, this information is required.

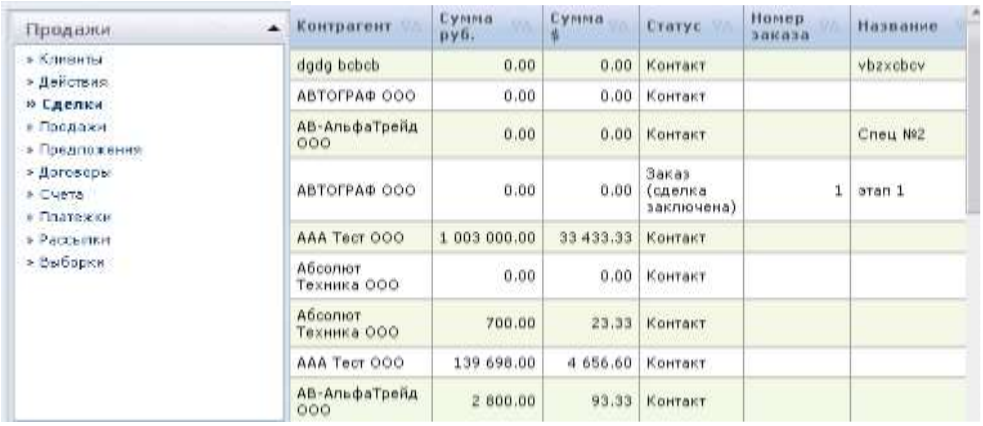


Fig. 3.10. Document List: Agreements window

3.5.3. The latest innovations in the system

The architecture used in the main system allows you to embed individual elements on existing windows or add individual modules that connect to the system.

Some elements that were introduced during the writing of the thesis are the module for generating a report in the format of MS Excel spreadsheets and an advanced module for searching information on the database for multi-parameter analysis of the entered information.

To simplify the work with the database or the ability to provide partial information during the test, the function of exporting data to external programs was added to the system (at the initial stage of modernization it is possible to export only to Excel). Each form has a button marked with the logo of MS Excel – when you click on this button you will be prompted to export data from the worksheet (Fig. 3.11). When generating a report export, you can select all the fields from the query, and only those selected by the user.

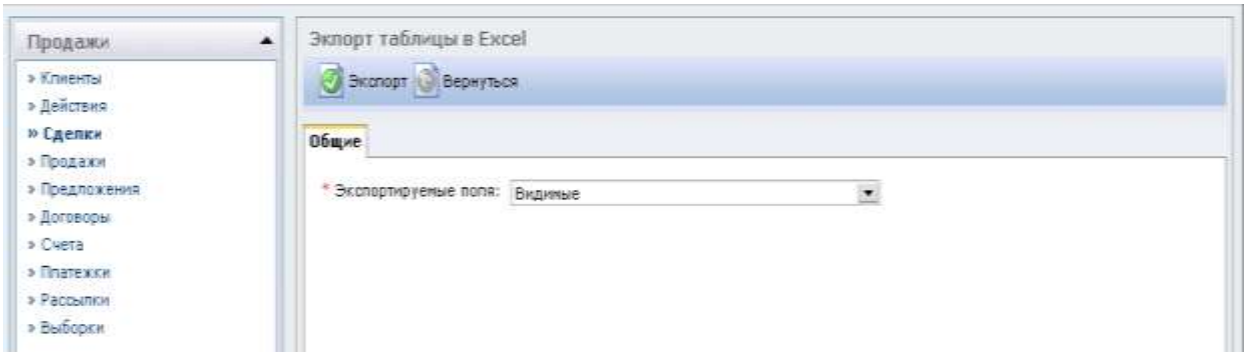


Fig. 3.11. Export table to MS Excel table

To do this, you need to determine which fields will be exported and click "Export". After this action all defined fields will be transferred to the export.xls file (Fig. 3.12).

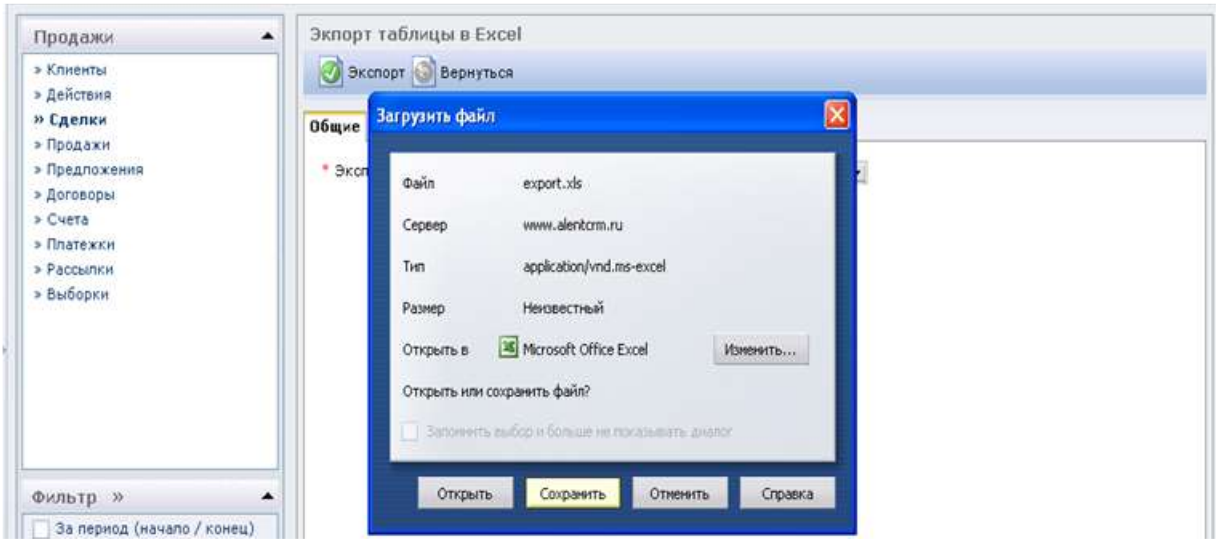
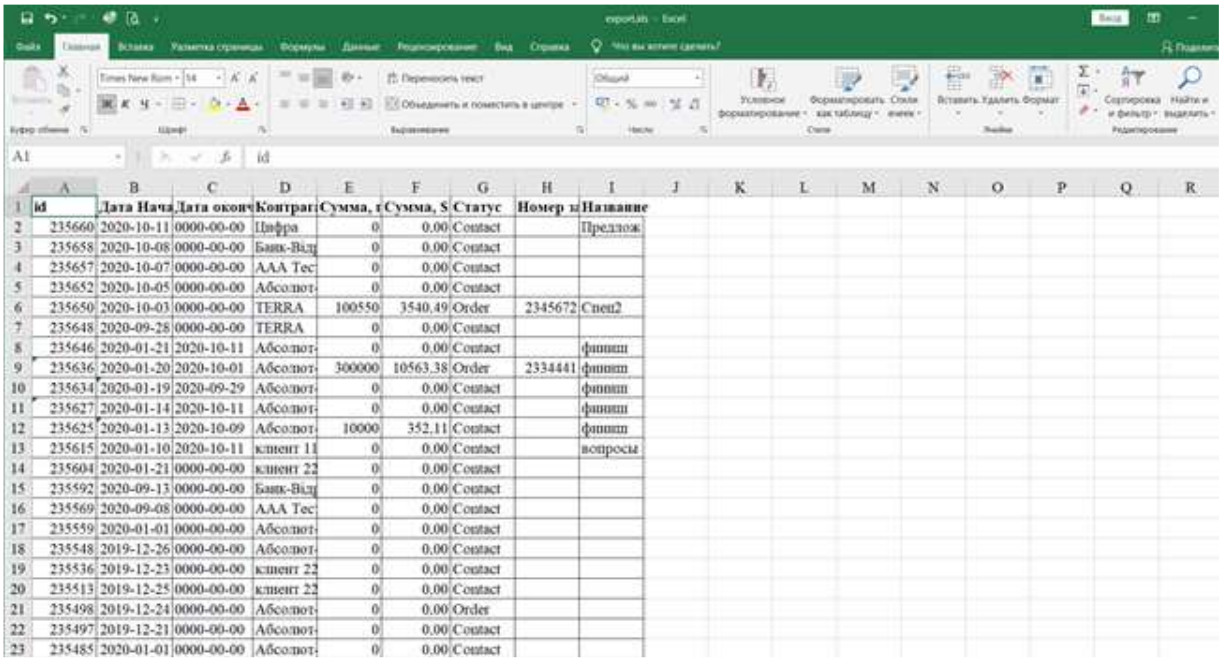


Fig. 3.12. Save data in the export.xls file

In MS Excel, this data will be presented in the form of a table, each column of which is responsible for the user-selected field (Fig. 3.13).



	A	B	C	D	E	F	G	H	I
1	id	Дата Нача	Дата окон	Контраг	Сумма, г	Сумма, $	Статус	Номер з	Название
2	235660	2020-10-11	0000-00-00	Цифра	0	0.00	Contact		Предлож
3	235658	2020-10-08	0000-00-00	Банк-Відр	0	0.00	Contact		
4	235657	2020-10-07	0000-00-00	ААА Тес	0	0.00	Contact		
5	235652	2020-10-05	0000-00-00	Абсолют-	0	0.00	Contact		
6	235650	2020-10-03	0000-00-00	TERRA	100550	3540.49	Order	2345672	Спец2
7	235648	2020-09-28	0000-00-00	TERRA	0	0.00	Contact		
8	235646	2020-01-21	2020-10-11	Абсолют-	0	0.00	Contact		финиш
9	235636	2020-01-20	2020-10-01	Абсолют-	300000	10563.38	Order	2334441	финиш
10	235634	2020-01-19	2020-09-29	Абсолют-	0	0.00	Contact		финиш
11	235627	2020-01-14	2020-10-11	Абсолют-	0	0.00	Contact		финиш
12	235625	2020-01-13	2020-10-09	Абсолют-	10000	352.11	Contact		финиш
13	235615	2020-01-10	2020-10-11	клиент 11	0	0.00	Contact		вопросы
14	235604	2020-01-21	0000-00-00	клиент 22	0	0.00	Contact		
15	235592	2020-09-13	0000-00-00	Банк-Відр	0	0.00	Contact		
16	235569	2020-09-08	0000-00-00	ААА Тес	0	0.00	Contact		
17	235559	2020-01-01	0000-00-00	Абсолют-	0	0.00	Contact		
18	235548	2019-12-26	0000-00-00	Абсолют-	0	0.00	Contact		
19	235536	2019-12-23	0000-00-00	клиент 22	0	0.00	Contact		
20	235513	2019-12-25	0000-00-00	клиент 22	0	0.00	Contact		
21	235498	2019-12-24	0000-00-00	Абсолют-	0	0.00	Order		
22	235497	2019-12-21	0000-00-00	Абсолют-	0	0.00	Contact		
23	235485	2020-01-01	0000-00-00	Абсолют-	0	0.00	Contact		

Fig. 3.13. Example of data exported to MS Excel

3.5.4. Module for creating new reports based on selected scenarios

But a separate problem of updating any automated system is the flexibility of requests for data stored in it. One of the tasks that was implemented in the thesis was to create an interface for building queries to the system database. But the main difficulty was that most employees do not have the necessary skills to create SQL queries to the database.

For this purpose on the basis of information on inquiries and specially developed algorithm of automatic construction of inquiry which was presented in section 3, the software module which allows to form new inquiries on the basis of the fields chosen by the client was created.

To process the parameters specified by the client on this form, we had to develop a separate algorithm for element-by-element survey of form fields with the ability to check the content of these fields and take data to generate a query. And in general, the block diagram of the form polling algorithm is presented in (Fig. 3.14).

This algorithm involves reading all the fields that are represented on it when calling the form. For this form, information is provided in advance about the correspondence of the ordinal number of the field with the name of the table field to which the filtering will be applied. Additionally, the type of this field is read.

The next step is to read the field values directly. If they are full, the value of the field is read, then its name is selected from the previously read list of fields. If the parameter type corresponding to this field matches, the read value is assigned.

If the field is not filled or the data type in it does not match the defined one, the data placed in this field are ignored, and a text string is formed for the user about the error sign in the query parameters, which is added to the existing one with the command ‘+=’.

Based on the read data, a query string is formed. Then go to the next field on the form.

As a result of the query, only those records will be displayed on the screen in which the field parameters are within the entered values (Fig. 3.15).

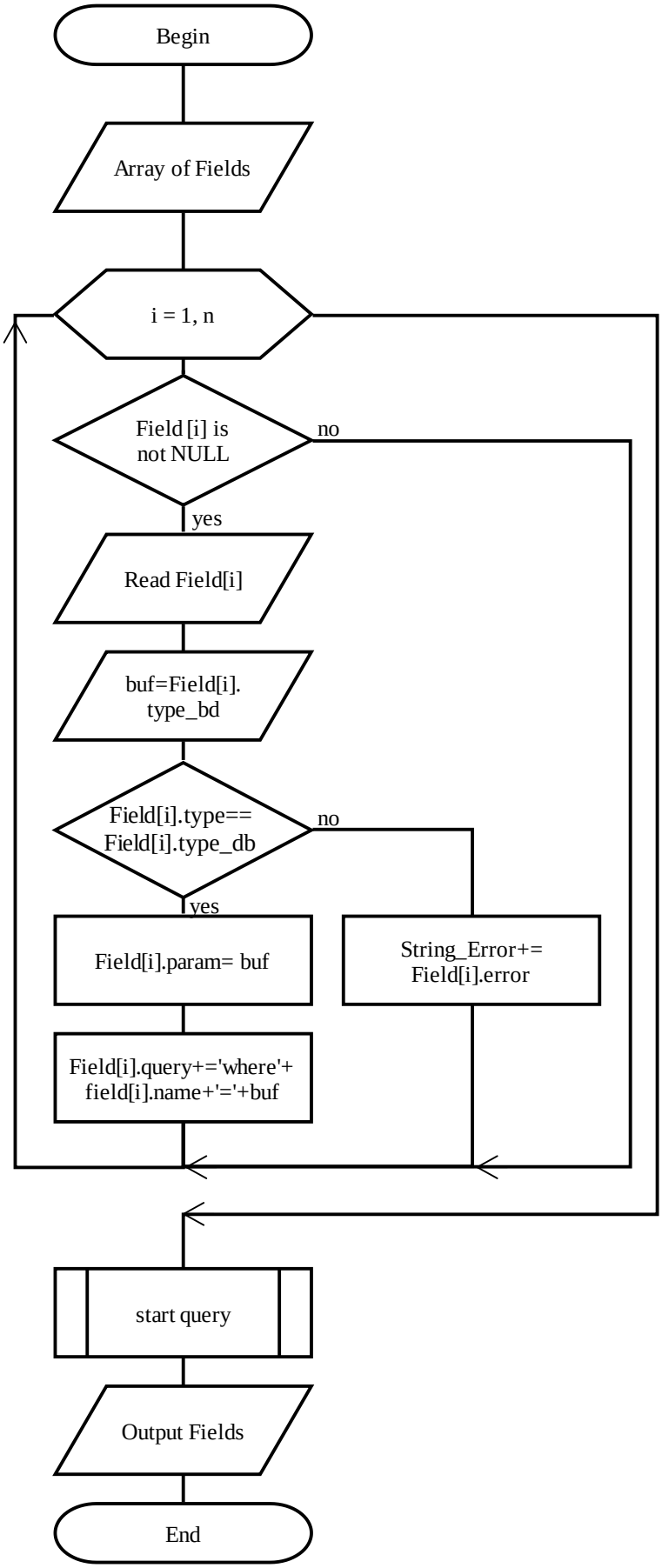


Fig. 3.14. Scheme of the algorithm for reading data from the form for automatic query generation

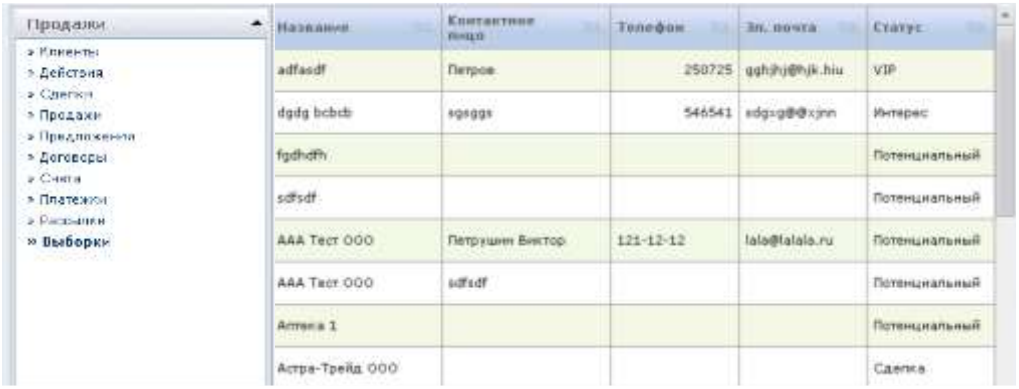


Fig. 3.15. Window of the result of sampling on the basis

At the end of the algorithm, all generated lines are passed to the procedure of forming and executing the query, which is described in section 3, and a line of errors is displayed separately. If there are no errors, a space will be displayed on the screen, which is placed in the default error bar.

This algorithm significantly reduced the program code, but the processing time remained the same. Prior to that, to generate a query, each form field was assigned a separate line with a separate variable to read its value. Then for each form field the procedure of formation of a liquid request with two parameters was called: the name of the field and value of the parameter. It was impossible to automate the reading of field names with such an organization of the program. Now on the basis of the entered marks and parameters the request to a database is automatically generated, in the absence of any marks this field does not take part in request formation.

3.6. Testing the performance of cloud services

Actually, the whole difference lies solely in the method of storing and processing data. If all operations take place on your computer (using its power), then this is not a "cloud", but if the process takes place on a server in the network, then this is exactly the trendy thing that is commonly called "cloud technology".

In other words, cloud technologies are various hardware, software, methodologies and tools that are provided to the user as Internet services to achieve their goals, objectives, projects.

The base of the infrastructure pyramid is a collection of physical devices (servers, hard drives, etc.), a "platform" is built above it - a set of services and the top - software available at the request of users.

Everything related to cloud computing (hereinafter CC) is usually called aaS. It stands for it simply - "as a Service", that is, "as a service", or "as a service".

We cannot trust the estimates of cloud services for two main reasons:

1) most often, companies conducting research do not indicate the parameters for launching tests, the number of iterations, the principles of choosing servers and their configuration;

2) there is no trust in service sellers (sometimes more expensive packages can be inferior in some parameters to cheaper ones).

We will conduct a study of three platforms Google Cloud, AWS, Azure.

3.6.1. Research methodology

All selected providers perform the following set of actions:

1) three virtual machines with 4 CPUs, 8 GB of RAM and a 50 GB system disk are launched sequentially in different zones of availability (if there are two zones, then 1 machine in the first zone and 2 in the second). Processor / Instance Type is the newest available, if you have a choice. The virtual machine type is shared with full core allocation. Disk type - network SSD with the ability to remount to another VM. IOPS guaranteed allocation options or machines optimized for this were not used, unless it is provided for by the standard conditions of use and cannot be canceled. The default file system is ext4. Operating system - Ubuntu 16.04 latest patch level available.

2) a series of 3 tests is launched on each of the machines, the total values for each machine are averaged. According to the following script:

```
TIME 60
cat> fio-rand.fio << EOL
[global]
name_user = fio-rand-RW
filename_user = fio-rand-RW
rw_user = randrw
```

```
rwmixread_user = 70
rwmixwrite_user = 30
bs_user = 4K
direct_user = 1
numjobs_user = 1
time_based_user = 1
runtime_user = $ {TIME}

[file1]
size_user = 2G
iodepth_user = 16
EOL

echo "Run FIO"
for i in {1..3}; do
 echo "$ i iter:"
 fio fio-rand.fio | grep -E "(read | write | bw | iops | READ | WRITE)" | grep -v "Disk"
done
echo "Run stress-ng."
for i in {1,2,4}; do
for z in {1..3}; do
 echo -n "$ z iter. Stress-NG for $ i CPU:"
stress-ng --cpu $ i --cpu-method matrixprod --metrics-brief -t $ TIME 2> & 1 | sed -n '6p' | awk '{print $ 5}'
 done
done
for i in {1,2,4}; do
for z in {1..3}; do
 echo -n "$ z iter. Sysbench CPU for $ i thread (s):"
```

sysbench --num-threads_user = $ i --max-time_user = $ TIME --test_user = cpu run 2> & 1 | grep "total time:" | awk '{print $ 3}'

done

done

for i in {1,2,4}; do

for z in {1..3}; do

echo -n "$ z iter. Sysbench Memory for $ i thread (s):"

sysbench --num-threads_user = $ i --max-time_user = $ TIME --test_user = memory run 2> & 1 | grep "Operations performed:"

done

done

3) the final platform performance is expressed as the arithmetic mean of the averaged test values for each of the virtual machines.

4) the cost parameter is involved in the assessment of the service. The cost is calculated without taking into account any bonus charges from the provider, without taking into account the cost of traffic, based on the operation of a virtual machine for a full calendar month without stopping. Also, some platforms allow, by means of certain restrictions (which it is quite possible to survive skillfully), to significantly reduce the cost of resources. For AWS, these are Spot instances, for GCE, these are Preemptible instances.

5) for comparison with cloud architectures, a regular server with the following characteristics was introduced into testing: E5-4650L @ 2.60GHz, 8 GB DDR3 EEC, SSD SmartBuy;

6) For all tests except Sysbench CPU, a higher value determines a higher performance rating.

3.6.2. Research findings

Table 3.1-3.4 summarized test results are presented

Table 3.1

Real server test results

Parameter	**Average**

Parameter	Average
FIO READ IOPS	3761.33
FIO WRITE IOPS	1611.00
STRESS-NG 1 CPU	12253.33
STRESS-NG 2 CPU	22874.67
STRESS-NG 4 CPU	40580.33
Sysbench CPU for 1	11.05
Sysbench CPU for 2	5.57
Sysbench CPU for 4	2.85
Sysbench Mem t1	2515283.51
Sysbench Mem t2	3289569.89
Sysbench Mem t4	4209258.70

Google Cloud test results (table. 3.2): performance in 4 threads hardly differs from two, but does not degrade, in general, each core is very productive and half as powerful as the core of the test VM.

The only thing worth mentioning separately is the excellent homogeneity. Each of the machines differs in performance by no more than a measurement error, which gives excellent predictability and ease of planning.

Table 3.2.

Google Cloud test results

Parameter	Average
FIO READ IOPS	931.67
FIO WRITE IOPS	399.67
STRESS-NG 1 CPU	14209.89
STRESS-NG 2 CPU	28642.56
STRESS-NG 4 CPU	29738.56

Parameter	Average
Sysbench CPU for 1	12.61
Sysbench CPU for 2	6.36
Sysbench CPU for 4	3.50
Sysbench Mem t1	2064102.94
Sysbench Mem t2	1426786.53
Sysbench Mem t4	2335620.59

AWS test results (table. 3.3): the problem manifests itself clearly only under certain types of load (it is not visible in Sysbench), but considering the results of other platforms, this is clearly not a problem with the test, but a performance limitation. The homogeneity of the results is just excellent, everything is predictable and without surprises.

Table 3.3.

AWS Test Results

Parameter	Average
FIO READ IOPS	1957.33
FIO WRITE IOPS	841.00
STRESS-NG 1 CPU	21651.67
STRESS-NG 2 CPU	43255.33
STRESS-NG 4 CPU	40041.67
Sysbench CPU for 1	8.77
Sysbench CPU for 2	4.40
Sysbench CPU for 4	2.52
Sysbench Mem t1	3042532.67
Sysbench Mem t2	2163753.34

Parameter	Average
Sysbench Mem t4	2818418.98

Azure test results (tabl. 3.4): Decent performance, one of the best platforms presented. True, the price spoils everything.

Conclusions on the Third Part

The choice of DBMS has recently become more commercial than functional, because almost all DBMS developers have reached a sufficiently high level of reliability, speed and optimality of systems. The main difference is the price and some special teams.

More important is the structure of the database itself, rather than the database in which it is developed and used. Recently, almost any database supports the use of third-party databases with the ability to export and import not only information but also the structures themselves, which are usually implemented in SQL.

Table 3.4.

Azure test results

Parameter	Average
FIO READ IOPS	1094.56
FIO WRITE IOPS	469.11
STRESS-NG 1 CPU	10353.56
STRESS-NG 2 CPU	20819.33
STRESS-NG 4 CPU	40362.44
Sysbench CPU for 1	10.10
Sysbench CPU for 2	5.19
Sysbench CPU for 4	2.71

Parameter	Average
Sysbench Mem t1	2536842.28
Sysbench Mem t2	2409424.00
Sysbench Mem t4	2752661.55

The role assigned to the design of the database structure of the system is one of the most important, because it is during the design that critical errors can be made, which will be detected only at the stage of using the system, and which can lead to loss of necessary information or misidentification. bindings. Also, special attention should be paid to the optimal relationship between tables, which affect the speed of data processing in the database, especially for databases containing more than 100,000 records.

Properly designed database structure not only ensures reliability, but also simplifies the work of programmers who use already implemented links and intermediate tables, rather than creating new ones.

This section presents the main points of development of window forms, which are presented in defense of the thesis. The main advantages of the PHP language for the implementation of these forms are considered separately.

Each developed screen form is accompanied by program code for its formation. As separate innovations, which are a significant contribution to the modification of the existing system, the modules of obtaining data export in the format of Excel spreadsheets and automatic generation of multi-parameter query to the database are considered.

An algorithm for reading the parameters entered on the form and automatically generating a query to the database is described. The block diagram of this algorithm is presented.

Based on the study of the performance of cloud platforms, we can formulate the following conclusions:

– AWS holds the lead in average measured performance for single- and dual-core loads. In second place is Google Cloud;

– in terms of the speed of working with memory, the best AWS for single-threaded mode, for two-thread – Azure;

– in terms of disk speed, we have a clear winner – AWS.

CONCLUSIONS

With the speed of technology development that we are seeing today, it will very soon become difficult or even impossible for many companies to operate without business intelligence tools.

Companies that implement the appropriate toolkit by following the guidelines above will be able to design systems that will benefit all staff and drive business productivity and profitability.

An analysis of the principles of building information systems to support the analytical department and a review of the market for implemented systems for business intelligence made it possible to determine the signs of a list of the main functions of the system:

– the ability to analyze your mistakes and maximize retention;

– the ability to analyze success and maximize income;

– system of automated notifications;

– a system of recommendations;

– the ability to compare with the market;

– the ability to predict user actions and predict the values of indicators;

– the ability to communicate with end users;

– access to the original data;

– analysis of both traffic and user behavior;

– simplicity and accessibility of the interface;

– technical support;

– data accuracy.

In the process of writing diploma work the existing and technologies of information systems development in the Internet environment on the example of information system of the analytical department. When creating this system and writingthesis the following tasks were solved:

– the main activities of contact centers are analyzed;

– analyzed the types of Web-server interfaces that allow you to generate Web-documents according to customer requests;

– outlines the main aspects of modern Internet technologies used in the development of network systems on the Internet;

– conceptual and physical data model developed;

– additional modules for export and analysis of information are built.

– describes the algorithm for reading the parameters that were entered on the form, and the automatic generation of the query to the database, as well as presents a block diagram of this algorithm.

As a result of the analysis, it was found that the company that thinks about the customer not on the scale of one operation, but will be interested in the prosperity and success of its consumer, will win in the market. A company that will achieve consumer loyalty. A company that will have deep knowledge about its client and his needs. A company that will not only own but also actively use CRM technology.

But it is necessary to be clearly aware of the problem that the company will face in the implementation of this system and know the full range of opportunities offered by the systems offered in the software market.

It is also necessary to pay attention to the complexity of providing WWW-access to databases, which consists of the complexity of work in the implementation of one of the three main scenarios for updating data in a remote database. When evaluating the scope of work, it was found that the development of means of outputting the contents of the table in HTML format with the necessary formatting and text support takes about 1-3 days for one developer, and the development of tools to build an index structure to data can take 1-3 weeks for one developer.

At the stage of DBMS selection, it was found that recently this choice is more commercial than functional, because almost all DBMS developers have reached a fairly high level of reliability, speed and optimality of systems. The main difference is the price and some special teams.

The role assigned to the design of the database structure of the system is one of the most important, because it is during the design that critical errors can be made, which will be detected only at the stage of system use, and which may lead to loss of necessary information or its erroneous identification when binding. Also, special attention should be paid to the optimal relationship between tables, which affect the

speed of data processing in the database, especially for databases containing more than 100,000 records.

With the help of the algorithm used in the system for automatic generation of requests to the system, it is possible to significantly simplify the program code. This algorithm is fully used in the construction of an extended query to the database.

Today, virtually no system is developed without the ability to export data to third-party systems, which allows the use of existing third-party analyzers, which can be either more successful or more visual.

The software module, which was developed during the thesis, has a practical use, and the general concepts and principles of automatic query generation can be used to present information both on the Internet and on local systems.

Analysis of cloud data warehouses allowed us to identify a number of problems when setting them up:

– loading data into cloud data warehouses is not trivial, and large-scale data pipelines require configuration, testing and support of the ETL process. This part of the process is usually done by third-party tools;

– updates, inserts and deletes can be complex and must be done carefully to avoid degrading query performance.

– difficulties with processing semi-structured data – they need to be normalized in a relational database format, which requires the automation of large data streams;

– nested structures are usually not supported in cloud storage. You need to convert the nested tables to formats that the data warehouse understands;

– cluster optimization. There are various options for configuring a Redshift cluster to run your workloads. Different workloads, datasets, or even different types of queries may require different customization. To achieve optimal performance, it is necessary to constantly review and, if necessary, further adjust the configuration;

– query optimization – custom queries may not be in line with best practices and will therefore take much longer to complete.

– backup and recovery – although storage vendors provide many options for backing up your data, they are non-trivial to configure and require monitoring and close attention.

Based on the study of the performance of cloud platforms, we can formulate the following conclusions:

– AWS holds the lead in average measured performance for single- and dual-core loads. In second place is Google Cloud;

– in terms of the speed working with memory, the best is AWS for single-threaded mode, for two-thread – Azure;

– in terms of disk speed, we have a clear winner – AWS.

REFERENCES

1. ДСТУ 3008-95. Документація. Звіти у сфері науки і техніки. Структура і правила оформлення. – Чинний від 23.02.1995.

2. Computer Engineering: Graduation Project Guidelines Compilers: I.A.Zhukov, S.O. Kudrenko, N.B. Fomina. – K. NAU, 2019. – 46 p.

3. All PTC Mathcad Tutorials [Електроний ресурс]. - http://learningexchange.ptc.com/tutorials/by_product/ptcmathcad (звертання 21 листопада 2020).

4. Buyya R., Broberg J., Goscinski A., Cloud Computing. Principles and Paradigms, John Wiley & Sons, Inc., New Jersey, pp. 637.

5. Focus Group on Cloud Computing Technical Report, 2012, Part 1: Introduction to the cloud ecosystem: definitions, taxonomies, use cases and high-level requirements, ver.1.0, c.62.

6. Khaled S., Boutaba R., Estimating service response time for elastic cloud applications, Proc. Of the International Conference on Cloud Networking CLOUDNET, IEEE, 2012, c.12-16.

7. Recommendation ITU-T, 2016, Cloud computing framework and high-level requirements, Y.3501, c.27

8. Recommendation ITU-T, 2013, Cloud computing infrastructure requirements, Y.3510, 22.

9. Recommendation ITU-T, 2013, Cloud computing framework for end-to-end resource management, Y.3520, c.16

10. Tucci-Piergiovanni S..B.R. Eventual Leader Election in Infinite Arrival Message-Passing System Model with Bounded Concurrency, IEEE Xplore Digital Library Published in: Dependable Computing Conference (EDCC), 2010 European Date of Conference: 28-30 April 2010. 127 - 134 pp.

11. Батура Т.В., Мурзин Ф.А., Семич Д.Ф. Облачные технологии: основные понятия, задачи и тенденции развития, 2014. [Електронний ресурс]. -

http://swsys-web.ru/cloud-computing-basicconcepts-problems.html (звертання 21 листопада 2020).

12. Гореткина Е. Amazon уходит в отрыв на рынке IaaS/PaaS / под. ред. Е. Гореткина. PC Week. - М., 2013. - № 15. - С. 46-48.

13. Гудкова И.А., Масловская Н.Д., Вероятностная модель для анализа задержки доступа к инфраструктуре облачных вычислений с системой мониторинга, 2014.

14. Ермышкин, В.О. Использование облачного хранилища при программировании под ОС Android / В.О. Ермышкин, О.Ю. Чередникова // 69-я Международная студенческая научно-техническая конференция, Астрахань, 15-19 апреля 2019 года [Электронный ресурс]: материалы / Астрахан. гос. техн. ун-т. - Астрахань: Изд-во АГТУ, 2019.

15. Кузнецов А. Тест: какое облачное хранилище самой быстрое. [Электронный ресурс] // iguides.ru: [сайт]. [2018] - Режим доступа: https://www.iguides.ru/main/other/test_kakoe_oblachnoe_khranilishche_samoe_bystroe/.

16. Мамзелев А.И. Основные принципы концепции облачных вычислений, 2011 [Електронний ресурс]. http://cyberleninka.ru/article/n/osnovnyeprintsipy-kontseptsii-oblachnyh-vychisleniy (звертання 21 листопада 2020).

17. Облачные вычисления (Cloud computing), 2015 [Електронний ресурс]. – http://www.tadviser.ru/index.php/Cloud_computing (звертання 21 листопада 2020).

18. Облачные технологии [Електронний ресурс]. - http://www.ixbt.com/cm/cloud-computing.shtml (звертання 22 жовтня 2020).

19. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы. Учебник. – М.:, 2014. – 1026 с.

20. Описание Google Диск. [Электронный ресурс] // Startpack: [сайт]. [2019] - Режим доступа: https://startpack.ru/application/google-drive.

21. Орлов П. А. Сравнительный анализ эффективности использования современных облачных хранилищ // Молодой ученый. - 2017. - №20. - С. 154-156. [Электронный ресурс] // Молодой учёный: [сайт]. [2017] - Режим доступа: https://moluch.ru/archive/154/43468/.

22. Сафронова А. Какой облачный сервис подойдет вашей компании: на что нужно обратить внимание. [Электронный ресурс] // RusBase: [сайт]. [2018] - Режим доступа: https://rb.ru/opinion/oblachnyj-servis/.

23. Топ 10: бесплатные облачные хранилища 2018. [Электронный ресурс] // ТехноПросто: [сайт]. [2018] - Режим доступа: https://www.tehnoprosto.ru/top-10-besplatnye-oblachnye-xranilishha-2018/.

24. Худобина К.Н. Облачные хранилища данных: актуальность и тенденции развития - No. 03 (027). [Электронный ресурс] // Nauka-rastudent.ru: [сайт]. [2016] - Режим доступа: http://nauka-rastudent.ru/27/3300/.

25. Щербо В. К., Киреичев В. М., Самойленко С. И. Стандарты по локальным вычислительным сетям: Справочник./ Под ред. С. И. Самойленко. – М.: Радио и связь, 2010.