MINISTRY OF SCIENCE AND EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

Non-Profit Joint Stock Company
ALMATY UNIVERSITY OF POWER ENGINEERING AND TELECOMMUNICATIONS

Department Telecommunication systems and networks

«Admitted»
Head of the Department Baykenov A.S.
d.t.s., professor
(Surname and initials, degree, rank)
(sign) «_____» _____ **20__ y.**

DIPLOMA PROJECT

Theme: Applying the Big Data technologies in contact center

Specialty: 5B071900 – Radio engineering electronics and telecommunications

Implemented by: Masmakynova S.N. ICTe 14-9
(Student's surname and initials) group

Scientific Supervisor: Tumanbayeva K.Kh
(Surname and initials, degree, rank)
(sign) «_____» _____ 20___ y.

Advisors:
of Economy section:
Tuzelbayev B.I. docent
(Surname and initials, degree, rank)
(sign) «31» may 2018 y.

of Life activity safety section:
Begimbetova A.S senior lecturer
(Surname and initials, degree, rank)
(sign) «23» may 20__ y.

of Computer Science section:
Tumanbayeva K.Kh
(Surname and initials, degree, rank)
(sign) «_____» _____ 20___ y.

Standards compliance controller: Tumanbayeva K.Kh
(Surname and initials, degree, rank)
(sign) «_____» _____ 20___ y.

Reviewer: Sahariev B.B.
(Surname and initials, degree, rank)
(sign) «_____» _____ 20___ y.

Almaty 2018 y.

MINISTRY OF SCIENCE AND EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

Non-Profit Joint Stock Company
ALMATY UNIVERSITY OF POWER ENGINEERING AND TELECOMMUNICATIONS

Institute of Space Engineering and Telecommunications (ISET)
Specialty: 5B071900 – Radio engineering electronics and telecommunications
Department: Telecommunication systems and networks

ASSIGNMENT

For diploma project implementation

Student: Masmakynova Saltanat Nurpeiskyzy
(name, patronymic and surname)

Theme: Applying the Big Data technologies in contact center

Approved by Rector order № 155 of « 23 » 10 2017 y.
Deadline of completed project: « 25 » 05 2018 y.

Initial data for project, required parameters of designing result, object initial data:

Incoming traffic for April: 101842 calls
Incoming traffic for May: 113761 calls
Service time: 2 minutes
Daily traffic: 4144 calls

List of questions for development in diploma project or brief content:

Functions of modern contact centers
Shared systems of contact center
Contact center functionality
Areas of contact center use and their advantages
Architecture of the contact center
Technology of Big data
Definition of Big data
Data volume. Why did Big data become large?
Tasks of Big Data
Application of Big Data in Telecommunication
Mathematical modeling of call services
Analyzing and processing contact center's statistical data
Forecasting by method of trend extrapolation

List of illustrations (with exact specifying of mandatory drawing):

1. Голденштейн В.С., Фрейкман В.А. Call-ue
The general structure of the Contact center
Customer service diagram of contact center
Logical structure of the contact center
Typical architecture of the contact center
Steps for data processing
Classification of Big data technologies
Linear traffic trend for April
Plan for placing equipment and workplace in the control room
The linear trend of traffic for May
Smoothing the series using the moving average method (April 2018)

Recommended main references:

1 Голденштейн В.С., Фрейкман В.А. Call-центры и компьютерная телефония, СПб: БХВ, 2002
2. Бураван Д., Шикунов А., Будникаева С. Аналитический обзор рынка больших данных - М.: Борд, 2015- 51с.
3. A Gilmore, Call Center Management: Is Service Quality a Priority, Managing Service Quality, vol. 11. no 3pp. 133-159.
4. Золотарев С. Три кита больших данных // ИКС-2014-№1-2-С.34
5. Тумайдаева К.Х., Лемешская Э.М Анализ и прогнозирование входящего трафика call-центра // Качество-инновации образование-2009
6. Киселев М.Н IP контакт центр - объединенная система сервиса потребителей. - Москва, 2008г. - 122с

Project adviser with corresponding sections specifying:

Section	Advisor	Dates	Sign
Main part	Тумайбаева К.Кh		
Life safety	Ведяшева А.С.	-25.05.18	
Economic part	Тихвинев В.І	31.05.18	
Application of comp tech	Тумайбаева К.Кh		

SCHEDULE
of diploma project implementation

№	Sections, list of developing questions	Dates of bringing to Scientific Supervisor	Notes
1	Functions of modern contact centers	15.01.18 - 23.01.18	Done
2	Shared systems of call center	23.01.18 - 31.01.18	Done
3	Architecture of contact center	31.01.18 - 04.02.18	Done
4	Definition of Big Data	04.02.18 - 09.02.18	Done
5	Application of Big Data in Telecommunication industry	09.02.18 - 14.02.18	Done
6	Key technologies for Big Data	15.02.18 - 20.02.18	Done
7	Analyzing and processing contact centers statistical data	20.02.18 - 01.03.18	Done
8	Forecasting by the method of trend extrapolation	01.03.18 - 20.03.18	Done
9.	Life safety part - Review of working conditions - Calculation of natural light - Fire safety organization in telecommunication industry	15.01.18 - 26.04.18	Done
10	Economic part - Market analysis. Investigation of market services - Calculation of investment costs - Calculation of income - Calculation of operating costs	15.01.18 - 25.05.18	Done
11	Conclusion	25.05.18 - 30.05.18	Done

Assignment issue date «12» 10 2018 y.

Head of Department: _____ Baykenov A.S
(sign) (Surname and initials)

Scientific Superviser: _____ Tumanbayeva K.Kh
(sign) (Surname and initials)

Assignment submitted for implementation:
_____ Mashmakinova S.N
(sign) (Surname and initials)

Аннотация

Данный дипломный проект посвящен определению особенностей и путей применения технологий Big Data для повышения качества обслуживания современных контакт центров.

В проекте были рассмотрены методы обработки и хранения больших данных, а также влияние конченых результатов на повышения качества работы контакт центров. Для достижения поставленной задачи были обработаны данные о функционировании контакт центра и выполнен статистический анализ.

К тому же, был проведен анализ рабочего места оператора относительно безопасности жизнедеятельности. В экономической части были затронуты вопросы касательно эффективности и срока окупаемости проекта и составлен бизнес-план.

Аңдатпа

Аталмыш дипломдық жоба Big Data технологиясын қолдану арқылы заманауи байланыс орталықтарының қызмет көрсету сапсын арттыру ерекшеліктері мен жолдарын анықтауға бағытталған.

Бұл жобада ауқымды деректерді талдау және сақтау тәсілдерімен қатар олардың соңғы нәтижесіне қарай байланыс орталықтарының сапасын арттыру жолдары қарастырылған. Қойылған мақсаттарды орындау үшін байланыс орталықтарының қызметі туралы мәліметтер өңделді және статистикалық талдаулар жасалды.

Сонымен қатар, өмір тіршілік қауіпсіздігіне қатысты оператордың жұмыс орнының талдамасы қарастырылды. Экономикалық тұрғыдан, жобаның тиімділігі мен өзін өзі ақтау мерзімі анықталып, бизнес-жоспар құрылды.

Abstract

This thesis is dedicated to identify the features and applications of Big Data technology to improve the quality of services in contact centers.

In the project will be considered methods of processing and storing the Big Data, as well as the impact of its final results on improving the quality of services in contact centers. In order to achieve this goal there will performed data about functioning of contact center and statistical analysis.

In addition, there were analyzed the situations of safety of vital, furthermore presented safety of employers in working place. In the economic part of the project were raised questions regarding the effectiveness and payback period of the project and a business plan.

Content

Introduction

In rapidly developing time the most effective tool in communication with the customers nowadays become the contact center. Majority of organization, working with clients, sometimes it can be private company or government and emergency services, which includes in their infrastructure one or more distributed contact centers.

In the recent years, many companies in Kazakhstan felt the need of contact center in connection with the straightening of competition and increasing of requirements for the service provided by them.

Data from the statistical survey presented on the Kazakhstan Contact Centers Association (KCCA), indicate insufficient customer service quality level. On the increase of load and the expansion of the types of incoming traffic to contact center the quality of service will be low.

Investigation of processes in functioning of contact center by the applying the Big Data technologies for achieving the statistical results and development of calculation methods of contact center is became as urgent task.

Purpose of the project is improvement of the quality of services of Contact center based on the Big Data technologies research by using of its functioning and development of calculation methods of characteristics, influencing on the quality of services.

For achieving the purpose its necessary to implement these tasks:

1) Making analysis of current state of contact centers;
2) Statistical analysis and data processing of functioning of contact center;
3) Analytical modeling of functioning of contact center as the system of mass service;
4) Development of a statistical analysis for contact center for predicting quality of service on STATISTICA program;

On the workflow there will be given definition for the Big Data technologies, key performance indicators, differences between traditional data base and Big Data. Also there will be considered questions about how large the data is and who can take benefit from it?

1.Functions of modern contact centers

1.1 Shared systems of contact center

Modern call center, in difference with traditional system of distribution of calls (DoC), has variety functions. Directed incoming calls enter to the first freed operator. As the part of contact center there is installed original hardware and software complex, components for each workplace [1,2,3].

Usually in the components of group devices of contact center, made by conditions of traditional telephony, includes:

- the switching subsystem (depending on the architecture of the contact center or the voice communication server CTI, or a special switch);
- data base of contact center;
- contact center administrator workstation.

Also for regulation and current application processing there might be appointed programmers and provided separate workplaces[1].

As much as the blocks presented in the figure are functional, their arrangement with separation into separate physical devices is performed in an arbitrary form.

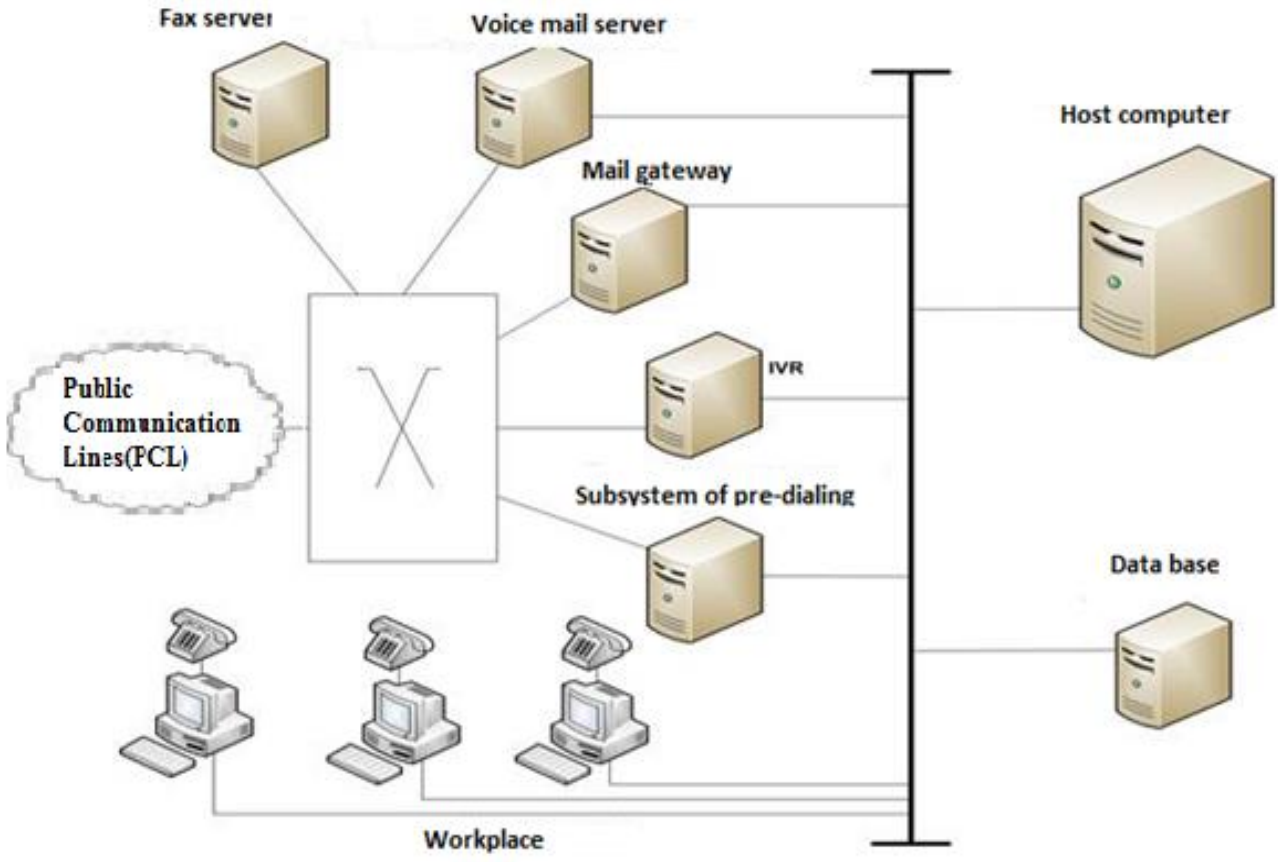

Figure 1.1 - The general structure of the Contact center

Let’s consider a generalized implementation of the main functions of contact center. Typically, contact center supplies support of several groups of operators (functional columns) by a separate line in each group or common wait buffer.

While waiting for the services, the Contact center not only provides the subscriber with the necessary information, supporting the menu of an arbitrary structure through multifrequency navigation of the additional set, but also provides an interactive subsystem of voice responses and IVR recommendations that allow creating multi-level actions [2,3,6].

During such interaction, the subscriber can easily receive voluminous information of interest to himself and for this the operators of the contact center will not waste their time. During communication with the IVR, the subscriber can enter some information on the display in front of the operator. And the operator, when making a call, prepares the data necessary for the subscriber.

Also at certain times of the day, calls can be sent to the IVR. Due to this, subscribers can get basic information even when the service of operators is not functioning.

The Contact center provides a guidance system (guidance) and it provides the ability to respond to incoming calls and provide a service before issuing a response.

At the request of the system administrator, calls are sent to other groups (in the case of the absence of free operators, it is queued), depending on its various parameters (assigned number, caller's number, information entered during IVR communication, day of the week and specific time of day, duration queue, assigned to groups of operators, the number of calls queued in this group of operators, the waiting time for the provision of services by a group of operators, etc.). By mixing such parameters, you can create an optimal algorithm for answering calls (for example, you can service calls received from clients with high status out of the queue).

During their waiting for the provision of services, according to tradition, it is provided to the subscribers of the answering machine about the waiting time in the queue, the state in the queue for a certain group of operators, and the service being called by the subscriber. The subscriber can be informed about the serial number of his call and how much time will have to wait.

Accordingly, after the number dialing of service by client, the call will be sent to the server of their distribution. And server operates according to one of these following scenarios:

- If there is a free operator in the group, using a specially assigned dispatch algorithm for the service, the call is sent directly to its workplace;
- If free operators are not available, the call is queued;
- the call can be directed to the IVR system, after this dialogue it (if necessary) is sent to the workstation of the released operator;
- the call is sent to the IVR system and after the dialogue, if there is no free operator in the corresponding group (service), it is put in the necessary queue.

In order to evenly distribute the load between operators, one of the following three algorithms is used:

- Circular distribution of calls, i.e. The call is sent to any of the free operators;

- after servicing the last call, the choice of the remaining one for the longest time free operator;

- selection of operators who have received the least number of calls compared to other operators since the beginning of the shift.

In the Call-center, as a rule, the subsystem of preliminary dialing of number is provided. The function of this subsystem is necessary for the development of outgoing connections in the process of work. Depending on the task of the center, either the system itself or the operators themselves make up a list of notified customers. According to the compiled list, the system automatically makes calls and checks the status of the called subscriber network. If there is a person waiting for connection in the network, one of the operators (after preliminary delivery of the combined sentence by the auto-sensing subscriber) starts to service.

The Contact center also provides a call forwarding function, which is triggered when the operators do not answer the call. This function provides the possibility of customer servicing in the event that the operator obliged to receive a call for unknown reasons could not answer the call received on its console equipment (multifunction telephone device) (for example, if he left the workplace without warning the system and the system considers it ready to receive calls). Since each of these cases worsens the quality of service provision to subscribers, it is necessary to reduce its effect. If the operator assigned by the system administrator does not answer the call within a certain period of time, the call is redirected to another operator of the group. In order for this situation not to happen again, the workstation of the absent operator is automatically transferred to an unworkable state and is maintained until the given operator returns and redirects its console to the standby mode. Provision is also made for the senior operator to receive reports of each such cases of non-response by operators. Along with messages about the specific time for leaving unanswered queries, a chronological report is also compiled with the name of such operators and the total time for leaving the requests unanswered.

The call-center provides for the formation and preservation of voluminous statistical and user information. The reports compiled on the basis of this information go to the workplace of the system administrator. The call system captures several important parameters of each processed call: a certain type of calls (incoming / outgoing / internal), call arrival time, waiting time, call service duration, number of the operator who serviced the call, during which time the call was lost (if a similar situation took place).

It provides a report on the status of the queue at a specific time to each group of workplaces and operators in the Contact center. In addition, the administrator can get an answer to questions about the length of waiting in the queue and the average duration of calls, etc.

Contact center provides the ability to generate historical reports on the quality of call service and all important parameters of the effectiveness of the activity of the given center (on serviced / lost calls and on calls that did not wait for queue until the

set waiting time or after it, about the total time spent on calls by an operator or group of operators).

1.2 Contact center functionality

The configuration of the system for transporting (transferring) words and data via IP networks provides huge opportunities for access to contact center services from anywhere on the network [1,2,3].

In addition, due to the use of effective packet switching technology and as a result of reducing the bandwidth of talk traffic, it is possible to reduce production costs and optimize the primary and hardware architecture of the communication system.

Actually, the call service centers themselves are equipped with universal traffic processing mechanisms. Thanks to this, the reception, distribution and processing of messages and calls received from various networks is carried out by means of universal algorithms [1]. A new level of integration has led to a change in the term Call-center in the name of the contact center, much better reflects the functionality of the Contact center.

Through the use of VoIP (Voice over Internet Protocol) technology, the contact center provides reception of requests that allow processing (requests that in the future go into the concept of unicasting), delaying the reception of traditional telephone calls coming from the Internet.

Unified Messaging is provided by the following functions:

- multi-network access (through a fixed telephone exchange, via the Internet, through mobile networks of various standards, through a paging network);
- multi-terminal access (via fixed / mobile phone, fax, web browser, etc.);
- Unified mailbox for all types of messages (conversations, fax and e-mail, SMS messages (Short Message Service), messages to the pager).

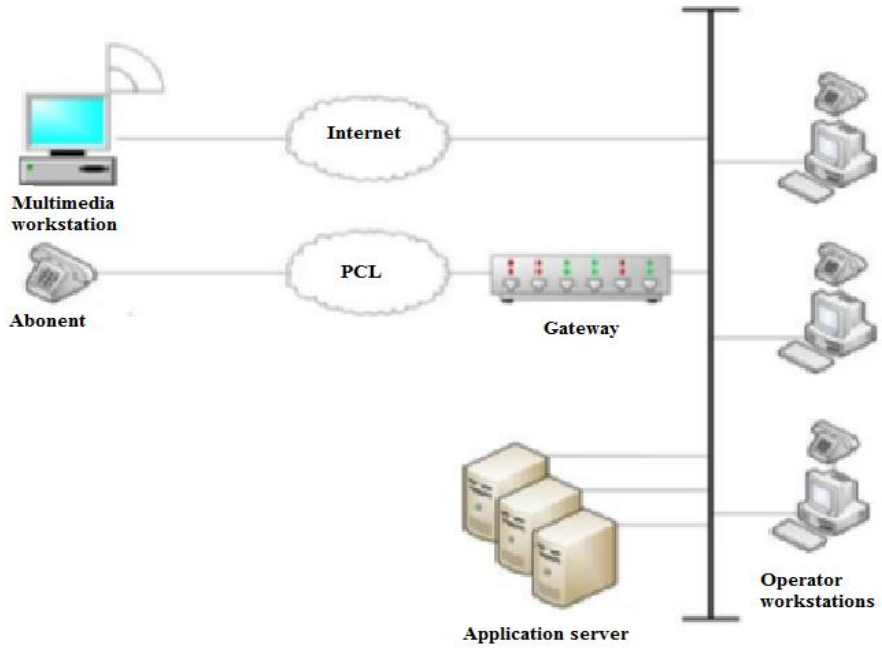

Figure 1.2 - Simplified contact center structure

The Contact Center, therefore, must provide requests for multimedia flat-line communication, support the text chat mode (real-time text information exchange between the subscriber and the operator of the center), ensure the support function of the web pages of the dynamic help system.

In other words, the tasks of the contact center are reduced to the following points:

- providing a wide range of opportunities both in terms of access and services with the use of human resources (operators) and automated systems;
- Guaranteed processing of the transaction, regardless of where the call came in and how to access the resources of the contact center;
- Providing opportunities to integrate with existing contact centers and sufficient equipment with the necessary functions with the use of third-party devices.

In assessing the architecture of the contact center, two ways of its organization can be distinguished (Table 1.1):

- the existing contact centers make necessary software and hardware additions that are not directly related to the system of call distribution (SoCD); On the beginning of integration, many companies which had had contact centers, installed software and hardware additions;
- Companies that implemented the system on the principles of packet switching, can most fully use their own, resulting from this advantage; this is the way of the company, starting all over again or decisively in technical terms, for example, widely applying IP technologies of the company.

Table 1.1 - Characteristics of the traditional call-center and IP-contact-center

Call center	IP contact center	Benefits for the company that uses it
Circuit switching	Packet switching	Providing traffic is beneficial, for supporting the channel switching does not require expensive devices
To access services, you need the support of one environment	To access services, you need the support of the entire environment	Great opportunities regarding accessibility
A large number of additional servers - for each application a separate server is provided	The number of servers depends on their performance and functionality	Ease of distributed allocation and management, low cost
The algorithms for servicing different types of calls are different	The algorithm for serving all types of calls is the same	High quality of services provided by operators and their performance

Table 1.1 Continuation

Call center	IP contact center	Benefits for the company that uses it
Computer telephony integration (integrated functionality)	Reciprocity of computer-to-computer interaction (unified functionality)	The cheap integration of services, the shorter terms of setting complex tasks, the connection of the processing function of talk traffic and information technology ensure fast and efficient economic progress
Traffic processing is centralized	Distributed traffic management	Flexibility, reliability of configuration
Hard attachment of the installation site of the operator consoles to the installation site	Independence of the location of the operator consoles from the installation site	Opportunity to support distributed operator workplaces, opportunity to offer outsourcing services

Packet switching technologies provide the ability to use the IP protocol as a universal transport layer protocol and to abandon large-scale switches by shifting the switching functions to the system itself. In this case, the switching functions of the talk channels are limited to managing media flows between the contact center elements (the flow of conversational information in the form of a packet). All the functionality is implemented by the application servers, working with management information, media streams and in the process of maintenance with information and technological databases.

Thus, each of these servers is responsible for its own set of services. To connect new features, additional servers and applications are provided. The contact center based on IP-technology (Figure 1.3) consists of several types of functional parts, one of which is a hardware-software unit, and the rest - only software products. The software product that manages the direction and queue of calls is the main core of such a system.

The functions of the contact center modules are as follows [1].

IP-telephony gateway. The IP gateway provides interoperability between the packet switching system and the channel switching telephone system. The main functional goal of the gateway is to perform the function of the hardware and software complex that converts the colloquial information coming from the public communication line (PCL) to the appropriate form for sending through networks directed by IP packets. Another function of the gateway is converting the PCL number to an IP address. In addition, the gateway supports interchange of signaling messages and with switching / terminal networks, and with devices working with IP telephony standards.

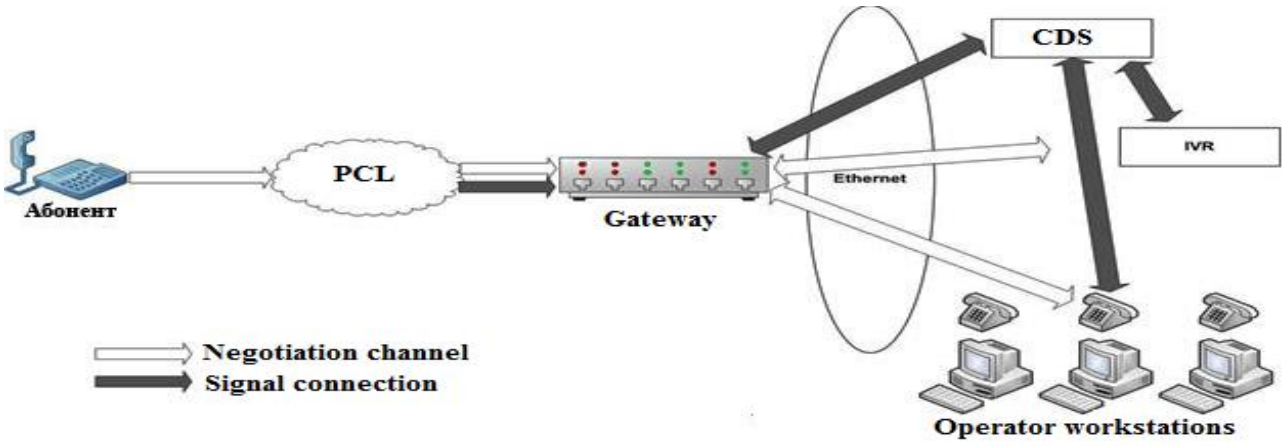

Figure 1.3 - General structure of the contact center based on IP-technology

Servers application. Application servers provide full implementation of the services provided: logics such as interactive interaction, distribution of calls, recording of conversations.

Call servers are the crucial element of any contact center. When serving calls, interacting with the database, they provide support for the interlacing system, routing and distribution of incoming calls to the contact center

Interaction Servers

Interactive chat interaction servers with IVR. IVR performs all functions regarding the organization of computer communication with subscribers who have contacted the contact center. This is the presentation of auxiliary calls to the subscriber by means of a call, receiving additional information from him in the direction of the mono-frequency additional dialing, providing him with various help and service information in automatic mode, supporting the conversation synthesis function and many other functions related to IVR. The actual technological specifics of this subsystem of the contact center consists in the sending and receiving of information in the form of a packet through an IP network that is in communication with other subsystems of the contact center.

Database. The database of the contact center is divided into two categories: information and technology. The technological base stores information about the system configuration, statistical data, some operational information, etc. The information base stores the data of the company using the contact center, about the customers regarding their service environment and other business information.

The maintenance (administration and management) server, in essence, is a simple computer (a workplace of the system administrator), provided with a special program. It performs the functions of configuration and diagnostics of the system, the state of interfaces and talk channels is under control, the collection of operational and statistical information about the operation of the system and about the maintenance of calls, and the functions of generating reports and archiving.

Operator workstations. Operator consoles are organized on the basis of the usual simple standard working computer , provided with a special program (or web browser). The additional equipment of the workplace, the provision of talk traffic

within the contact center directly depends on the technology used. If two networks (computer and telephone) operate independently of each other and the role of the conversation switch performs call distribution servers(CDS), then there is no need for additional equipment. Telephone calls are traditionally received at the operator's console connected to the CDS over the subscriber network, and call information on the personal computer. If the contact center uses an integrated solution based on the IP protocol (there is only one - a computer network), then the operator will not need any telephone console. However, his workplace will need to be fully equipped with multimedia, i.e. voice card and headset.

Processing of voice calls. The call is sent to the IP-telephony gateway. On the basis of the dialed call service number, the IP address on the CDS server (which should serve this call) is determined and sends the call to this server.

Further, the possibility or impossibility of servicing the particular call at a given moment is determined in turn. After that, media channels are opened between the CDS and the telephone call (or when called over the Internet, between the CDS and the subscriber's computer) and the caller is given corresponding greetings and system administrator messages, for example, "Sorry! Unfortunately, at the moment all operators are busy. Your request will be answered without fail. We ask you to stay on the line and wait!). If the logic of the service implies an answer in the form of interactive communication (IVR), simultaneously the receivers of multifrequency signals (DTMF) or the word recognition system are connected and the possibility of interactive communication of the subscriber with the system is provided. Search for a free operator is carried out simultaneously. When there is a free operator, the CDS redirects this call to it. If the operator answers the call, the channel "Gateway-server CDS" is closed and the media channel "gateway-operator" is opened. The caller's console immediately receives information about the caller and external applications connected with the provision of services to these typical calls can be connected (for example, an additional window opens indirectly in the database).

During the service of the call, the operator can give various commands regarding the recording of the call, its waiting, redirection, etc. Both sides have equal opportunities to interrupt a call.

1.3 Areas of contact center use and their advantages

There is no sphere of activity where the use of contact centers would not be an effective solution. But still, if the company performs the following types of actions when setting up a system of rendering services to subscribers, then the use of such contact centers will be much more expedient:

- if in its activity it aims to obtain profit from the provision of reference and dispatching services;
- if the range of services provided by the company is extensive, the types and forms of service provision are diverse and multifaceted, if the advertising and marketing policy of the company increases the flow of clients' flows and if operational and vocational training is required to provide them with services;

- if the company seeks to reduce the cost of a functioning dispatch service and increase the efficiency of the use of existing telecommunications and information resources;

- if the company aims to expand the range of services provided, the development of new market segments.

Contact centers can work, focusing on the following types of services:

- a variety of activities;

- reference and information services, providing clients with a variety of information (services of the MTSP, banks, companies, medical registration services, railway, etc.);

- operational services that receive information from clients about emergency (emergency) situations (police, medical emergency, service to combat the elements, etc.);

- Tax policy service, public services regulating the activities of individuals through the issuance of regulations, etc.

Commercial directions of contact centers used for business:

- shopping centers and reception centers (trade conducted through catalogs, ordering seats, tickets, etc., or booking them, determining the effectiveness of advertising, supporting distributors, conducting customer work, marketing research, transportation);

- Banking centers (payment or verification of accounts, warning of payments);

- centers of technical support services, whose specialists give pre-sales consultations to customers, etc.

The contact center solution provides a unified, fully integrated architecture for the management of users and their data transmitted over the Internet or the public telephone network (PSTN). The advantages of using multimedia contact centers are the following [2]:

- Customer service. This opportunity of contact centers is of the greatest interest to both companies and people. The client can find the answer to most of the questions he is interested in on his own, at any time convenient for him. If he has questions to the operator, he can ask them at any time convenient for him. Since expensive human resources are saved, the company primarily saves its finances;

-the cost of telephone network services is reduced.

- thanks to the use of cheap communication, the owner of the contact center allocates fewer costs to companies via the Internet;

-loads are distributed evenly. Through fax or e-mail, operators can respond to requests while reducing the load. Thus, the efficiency of the operators and the entire contact center as a whole is increased;

-Geographic independence. As a result of the use of IP technology, the contact center supports territorial independence regarding both client resources and server applications and gives administrations the ability to repeatedly define a contact center within the entire corporate system;

-high reliability;

As for a business operating on a real-time scale, i.e. connected with real business times, call management is an urgent matter, the contact center meets the highest quality standards thanks to the use at the data transmission level of means, devices and devices that reliably exceed the traditional reliability requirements of all fail-safe system components;

-Constantly high quality of providing services to customers.

All the company's employees and contact center operators, regardless of how they contacted the customer - via telephone, via e-mail, or through the WEB-site - have access to a single array of information about the history of the relationship with this client. This means that the client can always be confident in providing services in a unified style in every case of contacting the company;

-increase of labor productivity of operators. The organization of a single queue using a variety of communication technologies for contacts with customers increases the efficiency of operators' work. They work in real time with interactive messages (speech, phone, chat), and in their spare time - with messages that can be sent slowly (e-mail, fax).

If we talk about the functions of contact centers, they are technical complex designed for reception and processing by means of operators of a large flow of telephone calls to subscribers. They can have the appearance of a large building where the operators are located, which receive and service incoming and outgoing calls or consist of a small number of jobs distributed among several departments. The main functions of such contact centers are to implement interaction with subscribers in the company's required volume, ineffective work and making profit.

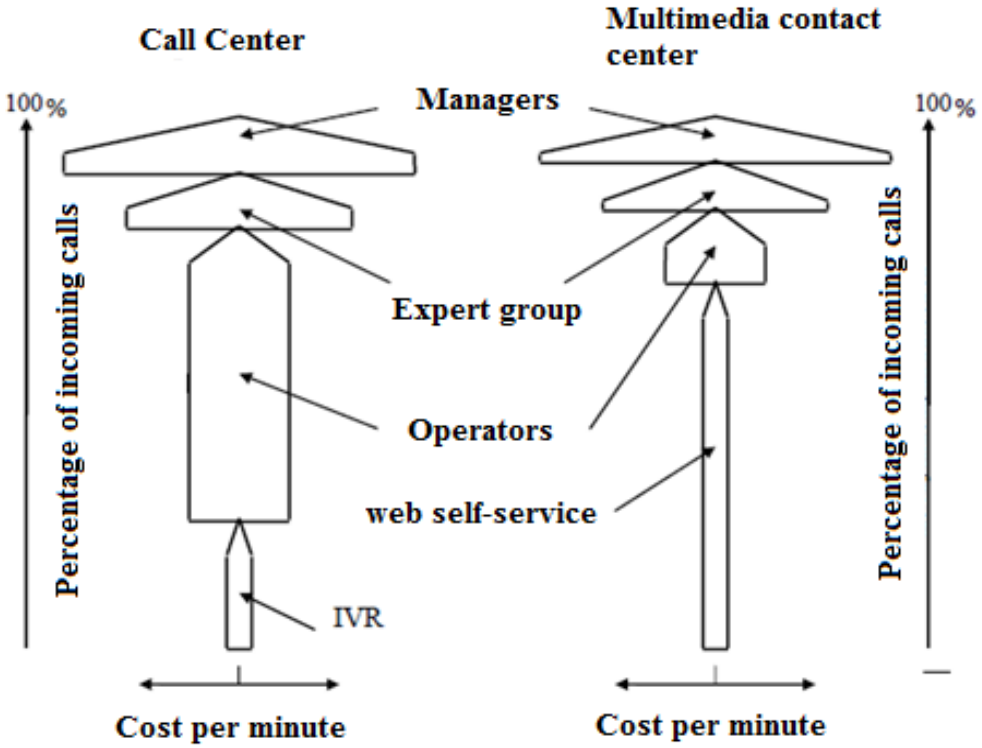

Figure 1.4 - Customer service diagram of the contact center

The main reason for the intensive growth of the service industry Contact center calls and reasons for the popularity of this method of processing requests are as follows:

- reduction in the number of cases of loss of subscribers due to increased load on telecommunication resources;
- increase of efficiency and expansion of opportunities for rendering services to subscribers;
- possibility of conversation with the subscriber separately, in comfortable conditions;
- decrease in the cost price of service of calls of subscribers;
- increasing the efficiency of the use of existing human, telecommunication and information resources;
- use of contact center resources to improve the universality of the company's services, expand their nomenclature and improve the mechanisms for providing services to customers.

The above factors both collectively and separately provide opportunities to increase the overall profitability of all company operations and reduce the costs spent on processing requests received from customers.

The presence in the contact center of technical facilities that allow individual conversation with the subscriber, the exact distribution of incoming requests flows, as well as interactive conversation has a positive impact on the company's image and directly affect the company's revenues. Thus, even enclosed volumetric investments pay off very quickly.

Although in the beginning the contact center was considered only as an infrastructural unit that did not generate direct revenues, then the attitude towards it changed - it turned into one of the profitable production units. Not all branches of the contact center yield the same income. However, even the indirect revenues of such centers in most cases pay back the costs of their development and maintenance. Therefore, many companies that have contact centers are trying to expand and modernize them.

There are two directions for using traditional contact centers [2, 3]:

- telemarketing (outgoing communication);
- provision of services to customers (incoming communication).

Telemarketing gives the company the opportunity to attract new customers and offer them their products. The provision of services increases the company's competitiveness and opens up new sources of profit. And to the client, along with this, information services are provided and support is provided in specific cases (questions).

In both cases, the ways of using contact centers do not change in any way. In telemarketing, the dialer calls the clients, and the speech recognition device filters incoming calls to the answering machine. After finding the customer in the database, he connects with the service provider as a free representative (operator) of the center. That is, as in the case of providing services to customers, this representative only works with certain connections. Although the telemarketing and

customer service functions are similar from a technical point of view, modern contact centers are focused only on providing services to customers.

In accordance with modern requirements, the equipment of contact centers can be in the form of a prefabricated constructor. Of these, you can construct the configurations needed for the desired application.

Such hardware flexibility and openness creates the software needed to develop the application. For today the list of spheres of possible application of contact centers, specific solutions counts tens, hundreds.

The communication company, which provides a variety of services through contact center operators, improves the quality of the services provided to clients and increases its capabilities. For the communication company, the optimized contact center achieves such opportunities [3]:

-call reception and offer him an automated menu;

- switching, connecting the subscriber to the required service or providing him with the necessary information in the operator's or automated form;

-identification of the subscriber or the owner of the service telephone card;

- receipt from the subscriber of the application for connection and quick implementation of the application (in case of impossibility of connection, a message about this to the subscriber);

- calculation of the cost of the service rendered to the subscriber in accordance with the tariffs defined in the system;

-provision of the multilateral conference regime;

-Representation of statistics in real (real-time) and archive (off-time) time.

Automatic caller identifiers in the contact center can provide operators with information about the called subscriber: from the most simple (where the call came from) and ending with the most difficult type of searching for customer information in the database.

1.4 Architecture of the contact center

A classic example of a contact center is a structure designed to receive and process incoming telephone calls from subscribers. However, as a result of the widespread adoption of Internet technology in all spheres of human life, new opportunities have emerged for subscribers to satisfy their requests, such as e-mail, Web-chat or VoIP, etc. This influenced the formation in the contact center of a new multimedia component responsible for receiving and processing requests coming through the Internet. As a result, a method of "duplicated organized" work appeared in multimedia contact centers. Thus, the classical part of the ACD is responsible for the telephone traffic, and the newly developed multimedia subsystem is responsible for receiving, processing requests.

However, the method of organizing such centers of processing requests coming through the Internet has significant drawbacks. The main ones are given below [4, 5]:

- duplicated performance of the function - the growth of the multimedia subsystem as a separate component results in the duplicated performance of some functions performed in the contact center;

- Distributed management − a "duplicate organized" method of providing services to customers leads to distributed management by presenting a given level. As a result, in the actions of the two groups there is inconsistency leading to poor−quality provision of services to clients;

- complicating the archive of contacts − when working on a "double-organized" method, it is extremely difficult to save the archive of contacts. And this directly affects the time and quality of providing services to customers.

As is clear from the above, work on the "duplicated organization" method leads to an increase in unnecessary costs and a deterioration in the quality of services provided to customers. Therefore, it is desirable to implement Internet capabilities taking into account the capabilities of existing software and contact center devices. Such a complex solution can be performed on the basis of modern equipment of a contact center and a software system.

Contact centers differ from the telephone center by the number of channels provided to connect customers with operators. Due to this, the service is expanded and the number of clients served is increasing. Figure 1.5 shows the logical structure of the contact center consisting of functional blocks.

The logical structure of the contact center consists of the following functional blocks [3]:

- Customer Interface - provides the sending of requests using the Internet or PSTN;

- Real Interaction Framework - activated when requests from users arrive, coordinates the operation of all devices and all contact center software, and controls the provision of all services for all incoming requests at a single level;

- Business Management - uses the indicators of managing the work of the contact center by the administration of the center and the efficiency of its performance;

- Enterprise Interface - used to respond operators to customer requests, made by any means;

- Infrastructure Management - provides high availability reality for users of contact center services and provides a guaranteed service of all incoming calls.

One of the main principles in the organization of the contact center integration of the direction and management of web-communications and voice messages. Let's consider the functions performed by the contact center when servicing requests coming through telephone networks or via the Internet (Figure 1.6).

A typical scheme for organizing such a system should include the following main constituent elements:

- Call Distribution System (ACD) or a telephone switching system, they provide the connection of the contact center to the public telephone network and the distribution of incoming calls to the operators' workplaces, depending on the

various circumstances that have developed (subscriber data, which operator is free, the characteristics of the requested information and etc.), connection of the center devices (IP-telephony gateway, STI server, IVR interactive interaction device);

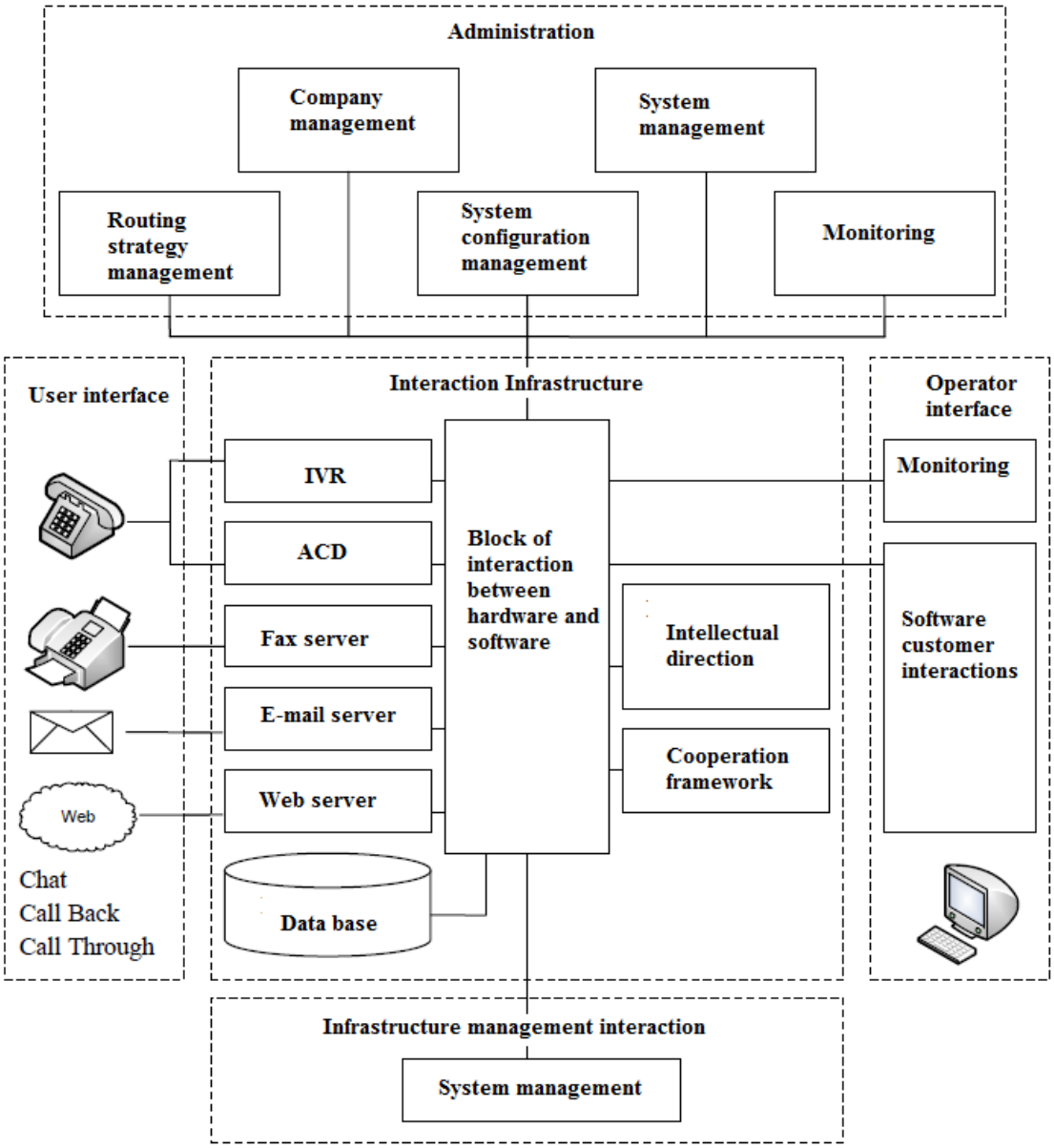

Figure 1.5 - Logical structure of the contact center

- IVR interactive interaction device, which provides automation of the direction of calls to operators or the ability to automatically respond to an incoming request;

- a management and statistics system designed to manage the direction of the Internet requests of the contact center, to collect statistics on the work of the center;

- IP-telephony gateway designed as a separate device or integrated into the switching system to ensure the operation of the contact center with the IP-telephony network;

- CTI server, providing contact of telephone and computer systems of the contact center;

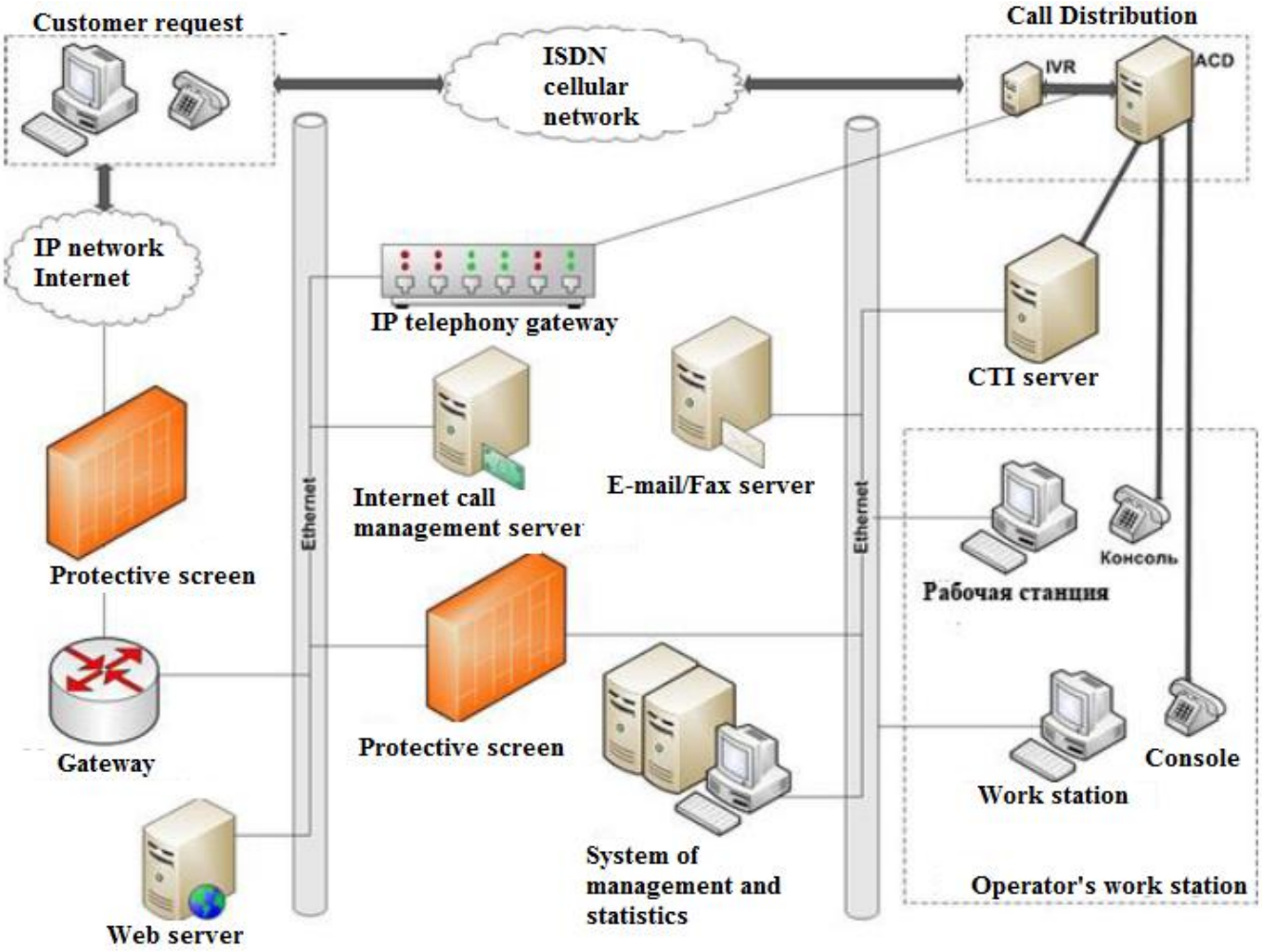

Figure 1.6 - Typical architecture of the contact center

- WEB-server that allows access to the Internet, routers, switches, etc.;
- a computer system consisting of various servers needed to process Web requests and save the database, the Firewall protection screen, the local computer system, etc.;
- workplaces of operators and contact center administrators, equipped with personal computers and telephone headsets, connected to switching or computer systems.

The Contact Center can be supplemented by advanced and service-providing applications:

- voice mail systems integrated into the local network, which allow to create a "universal mail box". This mailbox consists of emails, facsimile and voice messages. Within the contact center, such a system performs the functions of processing "lost", i.e. untimely in time for some reason calls;
- systems providing voice intelligent services, they allow you to organize a complex structure of intelligent autoresponders. Their main function is the recognition of voice simple commands, recognition of pulse sets, the function of text-to-speech transformation, integration with databases, etc.;

- "Home Office" system, which allows to create a distributed contact center, where some operators are sitting at home. The main function of this system is to send the information necessary for the normal operation of the operator outside the office of the company;

- system for creating a voice warning for mass call of subscribers together with a contact center. Provides an assessment of the subscriber's response, direction in accordance with this response and reporting. It is used to inform clients about the debts for various payments, for operative communication of the operator with the client, telemarketing, etc.

2. Technology of Big data

2.1 Definition of Big data

"Big data" is a fashionable term that now appears in almost all professional conferences devoted to data analysis, predictive analytics, data mining, and CRM. The very concept of "big data" is not new, it arose in the time of mainframes and it is related to scientific computer calculations. As you know, knowledge-intensive computing has always been difficult and usually inextricably linked with the need to process large amounts of information.

However, directly the term "large data" appeared in use relatively recently. It is one of the few names that have a very reliable date of birth - September 3, 2008, when a special issue of the oldest British scientific journal Nature was published, dedicated to finding the answer to the question "How can technologies that open up the possibility of working with large amounts of data? ». The special number summed up the previous discussions about the role of data in science in general and in electronic science (e-science) in particular.

You can identify several reasons that caused a new wave of interest in large data. The volumes of information grow exponentially, and its lion's share is related to unstructured data. In other words, the questions of correct interpretation of information flows are becoming more relevant and at the same time more complex. The reaction from the IT market was immediate - large players acquired the most successful highly specialized companies and began developing tools for working with large data, the number of corresponding startups and completely exceeds all the expectations.

Along with the rapid accumulation of information, data analysis technologies are also developing at a rapid pace. If several years ago it was possible, say, to segment customers into groups with similar preferences, now it is possible to build models for each client in real time, analyzing, for example, its movement through the Internet to search for a particular product. Consumer interests can be analyzed, and in accordance with the constructed model, suitable advertising or specific proposals are displayed. The model can also be tuned and rebuilt in real time, which was unthinkable a few years ago.

In the field of telecommunications, for example, technologies have been developed to determine the physical location of cell phones and their owners, and it

seems that the idea described in the sci-fi movie "Special Opinion", 2002, where display of advertising information in shopping centers took into account the interests of specific individuals passing by.

At the same time, there are situations when the enthusiasm for new technologies can lead to disappointment. For example, sometimes authorized data (Sparse data), giving an important understanding of reality, are much more valuable than Big Data, describing mountains, often, not essential information. First of all, under the terms "Big Data", "Big Data" or simply "big date" is hiding a huge set of information. And its volume is so large that processing large amounts of data with standard software and hardware is extremely difficult. In other words, Big Data is a problem. The problem of storing and processing huge amounts of data.

On the other hand, processing large amounts of information is only part of the "iceberg". Typically, when they talk about the term "Big Date", they use the most popular definition of the three "V", which means Volume - the amount of data, Velocity - the need to process information at a high speed and Variety - the variety and often the lack of structured data. For example, the time of the operation to check the balance on the card with the withdrawal of cash is calculated in milliseconds. It is these requirements dictate the market.

The third side of the issue is diversity and unstructured information. More and more often we have to operate with media content, blog posts, poorly structured documents, etc.

Thus, when we talk about big data, we understand that this is related to three aspects: the large amount of information, its diversity, or the need to process data very quickly.

On the other hand, this term is often understood as a very specific set of approaches and technologies designed to solve these problems. One of these approaches is based on a distributed computing system, where the processing of large amounts of data requires for itself not one high-performance machine, but a whole group of such machines that are clustered together.

2.2 Data volume. Why did the data become large?

There are a lot of sources of large data in the modern world. In their capacity, continuously incoming data from measuring devices, events from radio frequency identifiers, messages from social networks, meteorological data, remote sensing data of the earth, streams of data on the location of subscribers of cellular communication networks, audio and video recording devices can act as their quality. Actually, the mass distribution of the above technologies and fundamentally new models of use of various kinds of devices and Internet services served as a starting point for the penetration of large data into almost all spheres of human activity. First of all, scientific research, commercial sector and public administration.

As a rule, the discussion of Big Data is centered around data warehouses (and analysis based on such repositories), the volume is much larger than just a few terabytes.

In particular, some data stores can grow to thousands of terabytes, that is, to petabytes (1,000 terabytes = 1 petabytes).

Outside of petabytes, the accumulation of data can be measured in exabytes, for example, in the manufacturing sector around the world in 2010, an estimated 2 exabytes of new information is accumulated (Manyika et al., 2011).

There are industries where data is collected and accumulated very intensively.

For example, in the production area, such as power plants, a continuous data stream is sometimes generated for tens of thousands of parameters every minute or even every second.

In addition, over the past few years, so-called smart grid technologies have been introduced that allow utilities to measure electricity consumption by individual families every minute or every second.

For applications of this kind in which data must be stored for years, the accumulated data are classified as Extremely Big Data.

The number of Big Data applications is growing among commercial and government sectors, where the amount of data in the storage can be hundreds of terabytes or petabytes.

Modern technologies allow you to "track" people and their behavior in various ways. For example, when we use the Internet, we shop at online stores or large chain stores such as Walmart (according to Wikipedia, the Walmart data store is estimated to be more than 2 petabytes), or we move with the mobile phones on - we leave a trace of our actions that leads to the accumulation of new information.

Different ways of communication, from simple phone calls to downloading information through social networking sites such as Facebook (according to Wikipedia, the information exchange every month is 30 billion units), or sharing videos on sites such as YouTube (Youtube claims that it loads 24 hours of video every minute, see Wikipedia), daily generate a huge amount of new data.

Similarly, modern medical technologies generate large amounts of data related to the provision of medical care (images, video, real-time monitoring).

So, the classification of data volumes can be represented as follows:

Large data sets: from 1000 megabytes (1 gigabyte) to hundreds of gigabytes

Huge data sets: from 1000 gigabytes (1 terabyte) to several terabytes

Big Data: from a few terabytes to hundreds of terabytes

Extremely Big Data: from 1000 to 10,000 terabytes = from 1 to 10 petabytes

Where does the data come from?

Companies collect and use data of the most diverse types, both structured and unstructured. Here are the sources from which the Cisco Connected World Technology Report data:

- 74 percent collect current data;
- 55 percent collect historical data;
- 48 percent take data from monitors and sensors;
- 40 percent use the data in real time, and then erase them. Most often, real-time data are used in India (62 percent), USA (60 percent) and Argentina (58 percent);

- 32 percent of respondents collect unstructured data - for example, video. In this area, China is leading: 56% of those polled gather unstructured data there.

Who benefits from large data

It is difficult to find an industry for which the problems of large data would be irrelevant. Ability to operate large amounts of information, analyze the interrelations between them and make informed decisions, on the one hand, has the potential for companies from different verticals to increase profitability and profitability, increase efficiency. On the other hand, it is an excellent opportunity for additional earnings to partners of vendors - integrators and consultants.

To emphasize the benefits of developing and implementing tools for working with large data, McKinsey offers the statistics below. It is tied primarily to the US market, but it is not difficult to extrapolate it to other economically developed regions.

The potential volume of the healthcare market in the US is $ 300 billion a year - more than twice as much as in Spain. Part of this huge amount is spent on the introduction of modern IT, and obviously, large data will not be left out.

The total annual budget of public authorities in Europe reaches 250 billion euros - this is more than the GDP of Greece.

Monetization of geolocation services can bring $ 600 billion in additional revenues to market participants.The use of "large data" analysis tools in retail chains can potentially increase profitability by 60%.

Only in the USA, 140-190 thousand analysts and more than 1.5 million managers will be needed to manage the "large data" effectively to manage the information arrays.

American companies in 15 of the 17 economic sectors have large amounts of data than the library of the US Congress.

Methods for analyzing large data. There are many different methods for analyzing data sets, based on tools derived from statistics and informatics (for example, machine learning). The list does not pretend to be complete, but it reflects the most popular approaches in various industries. Researchers continue to work on the creation of new techniques and improvement of existing ones. In addition, some of the techniques listed above do not necessarily apply exclusively to large data and can be successfully used for smaller arrays (for example, A / B testing, regression analysis). Undoubtedly, the more voluminous and diversifiable the array is subjected to analysis, the more accurate and relevant data can be obtained at the output.

A / B testing. The technique in which the control sample is compared in turn with others. Thus, it is possible to identify the optimal combination of indicators to achieve, for example, the best response of consumers to the marketing proposal. Large data allows you to conduct a huge number of iterations and thus get a statistically reliable result.

Association rule learning. A set of techniques for identifying relationships, i.e. associative rules, between variables in large data sets. Used in data mining.

Classification. A set of techniques that allows you to predict the behavior of consumers in a certain segment of the market (making purchasing decisions, outflow, volume of consumption, etc.). Used in data mining.

Cluster analysis. Statistical method for classifying objects into groups by identifying previously unknown common features. Used in data mining.

Crowdsourcing. Methods for collecting data from a large number of sources.

Data fusion and data integration. A set of techniques that allows you to analyze the comments of users of social networks and compare with the results of sales in real time.

Data mining. A set of techniques that allows you to identify the most susceptible to the product or service of a consumer category, identify the characteristics of the most successful employees, predict the behavioral model of consumers.

Ensemble learning. In this method, many predicative models are involved, thereby improving the quality of the predictions made.

Genetic algorithms. In this technique, possible solutions are represented in the form of `chromosomes`, which can be combined and mutated. As in the process of natural evolution, the fittest survives.

Machine learning. The direction in informatics (historically, after him, was called the artificial intelligence), which aims to create self-learning algorithms based on the analysis of empirical data.

Natural language processing (NLP). A set of techniques borrowed from informatics and linguistics for recognizing the natural language of a person.

Network analysis. A set of techniques for analyzing connections between nodes in networks. With regard to social networks, it allows analyzing the relationships between individual users, companies, communities, etc.

Optimization. A set of numerical methods for the redesign of complex systems and processes to improve one or more indicators. Helps in making strategic decisions, for example, the composition of the product line being introduced to the market, conducting investment analysis, and so on.

Pattern recognition. A set of techniques with elements of self-learning for predicting the behavioral model of consumers.

Predictive modeling. A set of techniques that allow you to create a mathematical model of a pre-defined probable scenario of events. For example, analysis of the database CRM-system for possible conditions that will prompt subscribers to change the provider.

Regression. A set of statistical methods for revealing the pattern between the variation of the dependent variable and one or more independent ones. It is often used for forecasting and predictions. Used in data mining.

Sentiment analysis. At the heart of methods for assessing the mood of consumers lie the technology of recognizing the natural language of a person. They allow you to isolate from the general information flow messages related to the subject of interest (for example, a consumer product). Further evaluate the polarity of the judgment (positive or negative), the degree of emotionality, and so on.

Signal processing. Borrowed from the radio technology set of techniques, which aims to recognize the signal against the background of noise and its further analysis.

Spatial analysis. A set of methods of spatial data analysis, borrowed from statistics - topology of terrain, geographical coordinates, geometry of objects. The source of large data in this case is often geoinformation systems (GIS).

Statistics. The science of collecting, organizing and interpreting data, including developing questionnaires and conducting experiments. Statistical methods are often used for evaluative judgments about the relationships between events.

Supervised learning. A set of technology-based machine learning techniques that allow us to identify functional relationships in the analyzed data sets.

Simulation. Modeling the behavior of complex systems is often used to predict, predict, and study various scenarios in planning.

Time series analysis. A set of methods of analyzing data sequences that have been repeated over time from statistics and digital signal processing. One of the obvious uses is tracking the securities market or the incidence of patients.

Unsupervised learning. A set of technology-based machine learning techniques that allow you to identify hidden functional relationships in the analyzed data sets. Has common features with Cluster Analysis.

Visualization. Methods for graphically representing the results of analysis of large data in the form of diagrams or animated images to simplify the interpretation of the ease of understanding the results.

2.3 Tasks related to Big Data

There are three types of tasks related to Big Data:

1. Storage and management

The amount of data in hundreds of terabytes or petabytes does not allow you to easily store and manage them using traditional relational databases.

2. Unstructured information

Most of the Big Data data is unstructured. Those. How can you organize text, video, images, etc.?

3. Analysis of Big Data

How to analyze unstructured information? How to compile simple reports based on Big Data, build and implement advanced forecast models?

Storage and management of Big Data

Big Data is usually stored and organized in distributed file systems.

In general, information is stored on several (sometimes thousands) hard disks on standard computers.

The so-called "map" (map) keeps track of where (on which computer and / or disk) a particular piece of information is stored.

To ensure fault tolerance and reliability, each piece of information is usually stored several times, for example - three times.

So, for example, suppose that you collected individual transactions in a large retail chain of stores. Detailed information about each transaction will be stored on different servers and hard disks, and the "map" (index) indexes exactly where the information about the relevant transaction is stored.

With standard hardware and open source software to manage this distributed file system (for example, Hadoop), it's relatively easy to implement reliable petabytes of data storage.

Unstructured information

Most of the information collected in a distributed file system consists of unstructured data, such as text, images, photographs or video.

This has its advantages and disadvantages.

The advantage is that the ability to store large data allows you to store "all data" without worrying about which part of the data is relevant for later analysis and decision making. The disadvantage is that in such cases, subsequent processing of these huge data sets is required to extract useful information.

Although some of these operations can be simple (for example, simple calculations, etc.), others require more complex algorithms that must be specifically designed to work effectively on a distributed file system.

One top manager once told StatSoft that he "spent a fortune on IT and data storage, but has not yet begun to receive money" because he did not think about how best to use this data to improve core business. So, while the amount of data can grow exponentially, the ability to extract information and act on this information is limited and will asymptotically reach the limit.

It is important that methods and procedures for building, updating models, and automating decision-making are developed along with data storage systems to ensure that such systems are useful and beneficial to the enterprise.

Big Data Analysis.This is really a big problem with the analysis of unstructured Big Data data: how to analyze them with benefit. About this issue is written much less than the data storage and management technologies Big Data.

The impact of Big Data on IT departments. Big Data opens new opportunities for IT departments to build value and build close relationships with business units, allowing you to increase revenues and strengthen the company's financial position. Big Data projects make IT departments a strategic partner for business units.

In the opinion of 73 percent of respondents, it is the IT department that will become the main locomotive of the Big Data strategy implementation. At the same time, the respondents believe, other departments will also be involved in the implementation of this strategy. First of all, this concerns the finance departments (it was named 24 percent of respondents), research (20 percent), operational (20 percent), engineering (19 percent), and marketing (15 percent) and sales (14 percent).

Big Data will help increase IT budgets. In the course of the study (spring 2013), the Cisco Connected World Technology Report, conducted in 18 countries by the independent analytical company InsightExpress, interviewed 1,800 college students and the same number of young professionals aged 18 to 30 years. The

survey was conducted to find out the level of the IT departments' readiness to implement the Big Data projects and get an idea of the related problems, technological flaws and strategic value of such projects.

More than half of the IT managers surveyed believe that the Big Data projects will help increase IT budgets in their organizations, as they will face increased demands on technology, staff and professional skills. At the same time, more than half of the respondents expect that such projects will increase IT budgets in their companies this year. 57 percent are confident that Big Data will increase their budgets over the next three years.

81 percent of respondents said that all (or at least some) Big Data projects will require the use of cloud computing. Thus, the spread of cloud technologies can affect the speed of distribution of Big Data solutions and the value of these solutions for business.

Almost half (48 percent) of IT executives forecast a doubling of the load on their networks in the next two years. (This is especially true in China, where 68 percent of respondents hold such a point of view, and Germany - 60 percent). 23 percent of respondents expect a triple network load over the next two years. At the same time, only 40 percent of respondents declared their readiness for explosive growth of network traffic volumes.

27 percent of respondents admitted that they need better IT policies and information security measures. 21 percent need to expand the bandwidth.

Big data and mobile operator. The project implemented in China Unicom, according to Huawei, can be considered as typical for many mobile operators, including Russian ones. Although the standard version of the Telco Big Data solution is focused on the use of its own basic facilities (hardware, SIS), its modular structure allows replacing them with components of other vendors.

If we talk about the methods of large data aimed at obtaining an effect for business, then in the general case, four main areas are considered: the first three are aimed at improving the internal work of the company itself, and the latter is an additional market product for external customers:

- Precise marketing - targeted offering of products and services to those consumers who are most ready to purchase (new tariff plans, additional services, payment terminals, etc.);
- Customer Experience Management (Customer Experience Management) to improve its customer satisfaction in order to prevent user outflow;
- optimization of internal work of the operator and development planning (ROI-based Network Optimization and Planning) on the basis of taking into account all objective factors and opinions of consumers with the aim of maximum guarantees of return of investments in the shortest possible time;
- Monetization of information assets (Data Asset Monetization) - sale in one form or another (including in the form of equity participation in projects) of the operator's data to his partners so that they can solve their problems with their help.

By deploying a large data solution, the mobile operator was able to begin collecting and analyzing substantially more information about the behavior and

interests of its customers, including the intensity of communication usage and geographic location. And all this information could be correlated with the data on the operation of the cellular network itself, including about its download, about the arising failures, etc.

The possibilities of using such methods are evident from the results obtained. So, in early 2013, the effectiveness of marketing proposals (for customers who accepted them) with a total mass mailing was 0.7%. By the end of the year, due to the simple segmentation of subscribers (by age, sex, subscription term), this value was brought to 4%, and during 2014, first increased to 11% (accounting for intensity of use of services and location of customers) and then to 24% (accounting for preferred options for receiving the offer - voice calls, SMS, e-mail, social networks, etc.). During the year, it was possible to reduce the number of ineffective appeals to clients by 11 million, significantly reducing the cost of advertising campaigns.

Based on the analysis of 85 parameters of subscriber behavior, a "risk group" was identified, potentially ready to leave the operator's services. Inside it, there was also a certain segmentation, and for each category of clients a set of measures was developed to increase the level of their loyalty (discounts, other tariff plans, gifts, etc.). The customer conducted the research, dividing the "risk group" into two subgroups: from the first, special actions were taken to retain, on the other, nothing was done. An analysis of this work for the year showed that the company was able to significantly reduce the outflow of its existing consumers, retaining more than 200 thousand subscribers; at the same time it should be borne in mind that the cost of retaining a customer is always significantly lower than attracting a new user.

Before the use of large data, the expansion of the operator's geographic network was actually carried out only on the basis of information on the density of buildings and population, but having implemented this decision, China Unicom proceeded to develop its activity on the basis of multifactor analysis, which took into account such indicators as real traffic congestion and the demand for services for example, taking into account the place of work of people), the "value" of customers (in terms of living standards), the requirements for the quality of communication (the distance between receiving stations), the demand for different categories of services it depends on the use of different equipment), etc.

In terms of monetizing customer data for external partners, two examples were given: first, the optimization of outdoor advertising, both geographically (the place of residence, work or transport communications of the right customers), and taking into account the time for dynamic advertising (depending on from the time of day, the days of the week and the seasons of the year, the composition of the public can vary), and secondly, similar proposals for the development of retail chains (taking into account the location and range). In addition, it is very beneficial to target the distribution of mobile advertising in real time in accordance with the schedule of the person's employment, his interests and physical stay (for example, sending out information about the action films that the client is interested in, especially in his spare time and taking into account nearby cinemas). General

industry experience shows that such targeted methods allow you to increase revenues from the distribution of advertising at times.

How to cope with large data? Unprecedented diversity of data resulting from a huge number of all kinds of transactions and interactions provides an excellent fundamental basis for business to refine forecasts, assess the prospects for product development and whole directions, better control costs, evaluate efficiency - the list is easy to continue for as long as you like. On the other hand, large data sets difficult tasks for any IT department. Not only are they of a fundamentally new nature, in their solution it is important to take into account the budget constraints on capital and current costs.

An IT director who intends to benefit from large structured and unstructured data should be guided by the following technical considerations:

- Divide and rule.

The migration and integration of data is necessary, but both approaches increase capital and operating costs for information retrieval, conversion and download tools. Therefore, do not neglect standard relational environments, such as Oracle, and analytical data stores, such as Teradata.

- Compression and deduplication.

Both technologies have significantly advanced, for example, multi-level compression allows you to reduce the amount of `raw` data by dozens of times. However, it is always worth remembering which part of the compressed data may require restoration, and starting from each specific situation make a decision about using the same compression.

- Not all data is the same.

Depending on the specific situation, the range of requests for business intelligence varies widely. Often, to obtain the necessary information, it is sufficient to receive an answer to the SQL-query, but there are also deep analytical queries that require the use of tools that are equipped with business intelligence and possess the full range of capabilities of the instrument board and visualization. To prevent a sharp increase in operating costs, you need to carefully approach the compilation of a balanced list of necessary proprietary technologies in combination with open source software Apache Hadoop.

- Scaling and manageability.

Organizations are forced to solve the problem of the heterogeneity of databases and analytical environments, and in this regard, the possibility of scaling horizontally and vertically is of fundamental importance. Actually, just the ease of horizontal scaling has become one of the main reasons for the rapid spread of Hadoop. Especially in the light of the possibility of parallel processing of information on clusters from ordinary servers (does not require highly specialized skills from employees) and thus saving investments in IT resources.

- Technological backwardness.

Big data can turn into a big headache or open up great opportunities for government agencies, if only they can use them. Such conclusions were reached in the second quarter of 2012 by the authors of the study with the disappointing title

The Big Data Gap (from the English gap - "discrepancy", in this context between theoretical benefits and the real state of affairs). According to the results of the survey 151 CIOs in the next two years, the amount of data stored in government agencies will be reduced by 1 Petabyte (1024 terabytes). At the same time, it is becoming increasingly difficult to take advantage of the constantly growing information flows, the lack of available space in storage systems, the access to the necessary data, the lack of computing power and qualified personnel is hampered.

Technologies and applications at the disposal of IT managers demonstrate a significant lag behind the requirements of real-world tasks, the solution of which can bring additional data to great data. 60% of representatives of civilian and 42% of defense departments are still engaged in studying the phenomenon of large data and are searching for possible points of its application in their activities. The main, in the opinion of of federal authorities, is to increase the efficiency of work - this is the opinion of 59% of respondents. In second place is the increase in the speed and accuracy of decisions (51%), the third - the ability to build forecasts (30%).

Be that as it may, the data flows continue to grow. 87% of the surveyed indicated the increase in the volume of information stored during the last two years; 96% of the respondents (with an average increase of 64%) expect 96% of the respondents to maintain this trend in the next two years. To be able to take advantage of all the benefits that promise large data, the institutions that participated in the survey will need an average of three years. So far, only 40% of the authorities make strategic decisions based on the accumulated data, and only 28% interact with other organizations to analyze the distributed data.

- Poor data quality.

In a big house it's always more difficult to put things in order than in a tiny apartment. Here you can make a complete analogy with the large data, when working with which it is very important to adhere to the formula `garbage at the entrance - gold at the output`. Unfortunately, modern master data management tools are not effective enough and often lead to reverse situations (`gold at the entrance - garbage at the output`).

- Metadata: knowledgeable - means armed.

A query that does a good job of finding hundreds of lines out of a million may not be able to cope with a table of one hundred billion lines. If data changes frequently, it is extremely important to keep a journal and conduct an audit. The implementation of these simple rules will allow to have an important method for storing and working with data on information about the amount of data, the speed and the frequency of its change.

- Tell me who is your friend and I'll tell you who you are.

Correct interpretation of the tendencies and interconnections hidden in arrays of large data can literally be literally trained specialists. To some extent, they are able to replace the filters and recognizers of structures, but the quality of the results obtained at the output so far leaves much to be desired.

- Visualization.The same section of the article clearly illustrates the complexity and ambiguity of the approaches used to visualize large data. At the

same time, the presentation of results in a form that is accessible to perception is sometimes of critical importance.

-Time is money.

Viewing real-time data means the need for constant recalculation, which is far from always acceptable. We have to compromise and resort to a retrospective method of analytics, for example, on the basis of cubes, and to put up with partly obsolete results.

- Paly from the cannon on sparrows.

It is never possible to know in advance at what time interval large data are of particular value and most relevant. But collecting, storing, analyzing, creating backup copies requires considerable resources. It remains to perfect the storage policy and, of course, do not forget to apply it in practice.

2.4 Application of Big Data in Telecommunication

We are all living in the rapidly developing era of Big data, because every day 2.5 quintillion bytes of data is generated. Today, in every second the huge amount of data is created and transported through connected telecommunication world, so big data has become an integral part of the communication industry. Telecommunication operators have access to own customers information about their behavior, preferences, movement and etc.

Nowadays in telecommunication industry there are not only human-to-human communication but also human-to-machine and machine to-machine (M2M) communication makes huge amount of data which could be helpful for other industries [1]. According to the Industry analysts, projections are of 32 billion devices performing 44 trillion GB of data by 2020.

The telecommunication and technology companies have been applying legacy analytics for a long time, but the full potential is only in the future going to be realized by leveraging on the huge amount of data which is generated every time from social networks, search engines government portals, online businesses and other applications through real time predictive analytics.

The improved analytics solution can provide information which can help in creating new business models and launch innovative products and services. By using of its own data and fusing the data of other different sources, telecommunication operators can gain deeper understanding of their customer interaction and churn, product performance, and this can improve upon the customer experience and value addition. These type insights can be used by telecommunication companies to help other industries such as agriculture, healthcare, education, etc.

Nowadays big data analytics is becoming an integral part of every business, because with the quickly growing technology, big data will become an essential part of every business.

Slow growth of revenue, increasing network costs and rivalry between companies will enforce companies to set up advanced analytic solutions, which will

allow them to improve data traffic management and optimization of the utilization and performance of networks.

While data analytic companies are going through 40% annual growth, there is a huge shortage of specialists. Big data need specialists in data science who can understand the data analytic technology as well as the business objective of a telecom operator [1].

As telecommunication operators encounter spectrum costs and high network and intense competition, advanced analytics will help them to improve their profitability and gain competitive advantage by enhancing customer experience and optimizing the network usage. Creating a clear value proposition using analytics will help to justify and attract investment.

Possessing large data transmission technologies, telecommunication companies can find patterns and causes of the call drop in real and send a message with the apology of the customers, as well as return the money for which leads to increased customer satisfaction and brand value.

We need to build a large data platform, where all operators need to download sample CDR data, network and billing data after masking the details of subscribers and this platform will provide a dropped-call rate (DCR); Call Drop templates; causes drop call and billing template to subscribers and Call Drop, etc. This will result in some transparency regarding network performance and a rejection pattern.

Big data analytics makes considerable value in order to provide more actionable and accurate information in decision making which will support to build competitive advantages and more efficient cost structure. In difference from traditional data warehousing technologies, big data presents following opportunities and advantages for telecommunication operators:

1. Assembling company's networks for the future demands: in order to make networks optimized, robust and scalable, its necessary to take advantage of available information, which we can achieve the results with the help of Big data technology. By analyzing network traffic in real time it can help to optimize quality of service and routing. For example, by revision of 3G-capable cellphone users running their devices on 2G might identify ways of making significant enhancement based on intelligence analysis. To better deliver media content depended on behavior of user, which also impacts on customer experience, it is needed to collect some information with the aid of Big data technology.
2. Understanding of user's experience: Big data can understand right from network data and social media information about customer by its capabilities, which in further establishes customer-centric KPIs, enables to understand user experience.

Insights can be deployed in client contact centers to better respond and solve problems for the client on the phone and allow them to flexibly and profitably modify call plans for subscribers immediately. They can also be used in network operations centers (NOCs) to detect and resolve any problems for larger groups of users. Where necessary, you can collect information about the experience of users in real time. For example, employees of the contact center could make sure that the

client on the phone had problems in a certain place or when using a certain service. The result is improved customer service, a higher level of customer satisfaction and a lower outflow rate. Large data can also help analyze call records in real time to detect fraud.

For the further development of telecommunication value chain, there is listed some Big data application purposed areas [1]:

1. Network Infrastructure Management:

a) Opportunity to inspect packet in real-time to optimize traffic routing and manage network quality of service.

b) Performance measurements of cellular network

c) Measurement of data traffic for supplying

2. Service access integration:

a) Analyzing of real time call data records for identifying the fraud

b) Planning of innovations

c) Measurement and modeling of total customer application

3. Marketing and Sales:

a) Differentiation of responses based on event-based marketing campaigns that use geo-location and social media.

b) Cross- and up-sell targeting

c) Sale of (anonymous) customer insights based on usage data to shops, media agencies, etc.

d) New product/service innovation based on real-time usage patterns

2.5 Challenges faced by telecommunication operators

To create the new revenue streams and to become the winner of the battle for customers big data has the opportunity to place communication services providers (CSPs) in a prime position.

It supplies them with an information about customer’s behavior, preferences, and movements. However, many communication service providers are trying to fully benefit from the large data.

Depending on the variety, velocity and complexity of many communication service providers at macro level big data analytics go through many challenges.

Variety: As shown in the figure 2.1 the social media networks, billing information, government portals, connected devices, call data records, etc., produce huge amount of data. Therefore most of the data coming from sources which written above are different and unstructured.

The telecommunication entrepreneurs need to upgrade their Call Data Records (CDR) with other information such as location-based service, financial information, etc., to institutionalize the data for business intelligence platforms before the analysis can be set on it.

Velocity: Every year the population of Kazakhstan spend more and more on online shopping, for example by the end of 2016 amounted to 157 billion tenge, which is 2.6% of the retail trade, or 0.32% of Kazakhstan's GDP. Every minute 100

hours of video is shared on YouTube[2].The average time spent by a social media user in Kazakhstan 23-25 hours a week.

All this points to the fact that the data generation speed is tremendously high and to gain value from this data, it needs to be processed in proper timeframe[1]. This volume of data requires new real-time operational capabilities for various functions and that in turn demands increased data storage for compliance and potential future uses as well as new tools for mediating, managing, and archiving data within available time frames [1].

Complexity: The user generated data is mostly unstructured and complex because of the lack of standard format to store data. The legacy network and storage devices do not have any specific format to store data which can be relevant for advanced analytics. The data varies with demographics, geography, life style, etc. Analytics may provide unwanted results if the data is not filtered properly [1].

2.6 Key technologies

Architecture of data acquiesce to telecommunication companies to stock new types of data, possess that data longer, and join diverse data sets to derive new insights. This type of reference diagram architecture shows an admixture of those access that we observe between our telecom clients.

In big data technology for efficient processing to perform an effective analysis is convenient to use traditional relational approaches with horizontal scaling where already exists data volume, which assembled by acquisition technology.

Also, in big data science its not remarkable to find that there some lots of different data, which for each of them is needed to find their own approaches and instruments.

Now then we may list some of these categories [2]:

1. Natural language
2. Audio, video and images
3. Graph based
4. Structured data
5. Unstructured data
6. Machine-gene rated
7. Graph based
8. Streaming

Unfortunately, only 20% of data from all internet resources is structures and rest 80% of data, which made by users in social media, physical sensing devices and internet of thighs, is unstructured and dynamic.

Processing of data usually have 6 steps to process the data. The main structure is shown in Figure 1.

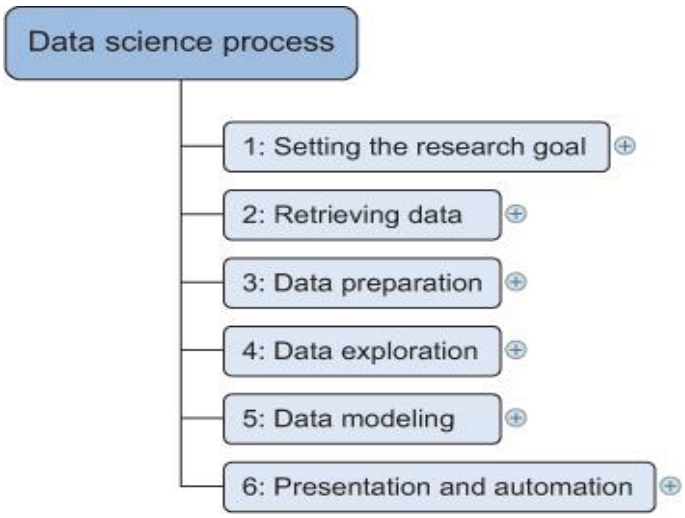

Figure .1 Steps for data processing

1. Setting the goal of research. Typically, data science discipline applied in context of organizations. For example, in business sphere at first is accepted to make project charter. In project charter is written about what you are going to research, what kind of benefit company will get, what kind of researches you will need.'

2. Retrieving data. At the second stage there will be the assembling of data. In project chapter is already written about what type of data is needed and where you can find them. In this stage you are making decision for using data in your program, which means verification the existence, quality and availability of data.

3. Preparing of data. At the data retrieving stage there might be mistakes, therefore in this stage scientist improves the quality of data and prepares them to next employment. This stage consists of three phases: cleaning the data from incorrect values and eliminates discrepancies between sources; integration of data expands the information by combining the data from several sources; transformation of data guarantees that data is in appropriate format for appliance in our model.

4. Exploration of data. Investigation of data is directed to have deep understanding of data. Consequently, you will try to understand how variables interact with each other, assess the distribution of data and identify the presence of emissions. For this, descriptive statistics, visual methods and simple modeling are mainly used.

5. Data modeling. At this stage of modeling will be useful the knowledge from statistics, machine learning, investigation of modeling operations. Model building is an iterative process, during which the researcher selects variables for the model, applies the model, and conducts model diagnostics.

6. Presentation and automation. After all, given 5 stages before, results of investigation will be given to business-side. Such results can exist in different forms, about presentation to reports.

2.7 Instruments and infrastructure for big data

Nowadays, exists several types of instruments and infrastructure for big data. It easy to get confused with the growth of new technologies. Situation might become easy with understanding that ecosystem of big data might can be divided into some small groups by technologies with similar goals and functionality. At the Figure2 shown the components of big data eco system and area of different technologies in this structure. There will be considered different group of instruments and will be ascertained what kind of role plays each of them.

Distributed filesystem**.** Distributed file system is similar to simple file system, but in difference with last one it works on several servers simultaneously. In distributed file system you may do anything, which is possible in simple file system. At the basis of any file system exists such actions as file saving, reading and deleting of data, also realization of data security tools. Distributed file systems have several advantages:

1. They are able to save files which are larger than the size of single computer disk;
2. Files automatically are replicated to multiple servers for creating redundancy and execution of parallel operations, while all the complexities of technical implementation are hidden from the user;
3. System is easily scalable: user is not limited with the volume of memory and disk space of server.

Before scaling was executed by transformation of all system in server with big volume of memory and disk space with faster processor. Nowadays, we have opportunity to add in our system one small server (horizontal scaling). Due to this principle potential of scaling becomes more practically limitless.

Today most popular distributed file system is Hadoop File System (HDFS). It represents realization with Google Hadoop File System with open code. In this graduate work all attention will be on Hadoop File System because it is more often used in daily practice. Also, there is a lot of other distributed file systems such as: Cepth File System, Red Hat Cluster File System, Tachyon File System and so on.

Distributed programming. After memorizing of all data in distributed file system, its is necessary to implement it. One of the most important aspects of working with distributed hard disk is that there is no need to shift data to program, but rather to shift program to data.

Data integration infrastructure. After creating distributed file system its necessary to add data. You should translocate data from one source to other; exactly here all advantages appears of such data integration infrastructures as the Apache Sqoop and Apache Flume. Process similar to the process of extracting, converting, and loading into traditional folding data.

Machine learning infrastructure. When the data is in its place, the time comes for extracting the coveted hidden information. At this stage it will be necessary to use methods from machine learning, statistics and applied mathematics. Most popular machine learning library for Python is Scikitlearn. It is best instrument of

machine learning, and we will use it in this book. Of course, exists other types of libraries for Python[2]:

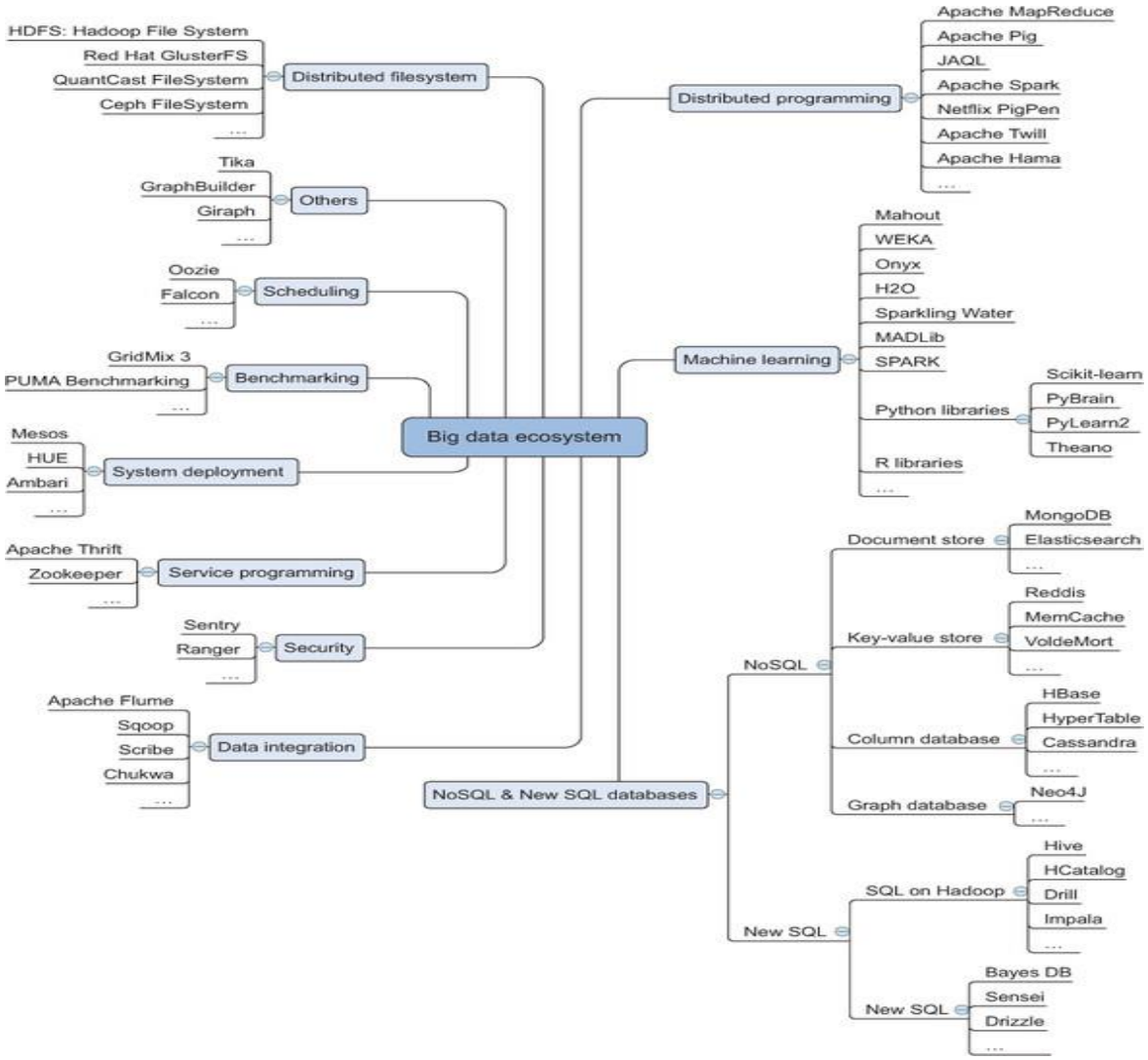

Figure 2. Classification of Big data technologies

1. PyBrain for neural networks– the term “neural networks” denote learning algorithms that model the human brain in relation to the mechanics of learning and complexity.
2. TensorFlow– library for Python for deep learning, provided by Google.
3. NLTK(Natural Language Toolkit) – library for data with natural language. This extensive library comes with a few text boxes that simplify the modeling of your data.

NoSQL databases. For storing huge amounts of data required software that specializes in these data and generate queries to them. Traditionally in this area were the main relational databases– such as Oracle SQL, MySQL, Sybase IQ and others. And although in many situations they still consider the preferred solution, new types of databases have appeared, combined databases NoSQL.

Exits sever types of databases, but we can divide them into:

1. Column Databases– data organized into columns, which allows algorithms allows you to increase the speed of processing requests. In more modern technologies there is used principle of storing into cells. Table structures continue to play an important role in data processing.

2. Document Repository – document repository do not use tables, but stores full information about document. Their distinguishing feature is an extremely flexible data scheme.

3. Streaming data –collection, transformation and aggregation of data is carried out not in accordance with the package principle, but in real time. Although we have allocated streaming data to the database category in order to simplify the choice of the tool, rather they form a special kind of tasks that has generated technologies such as Storm.

4. Storage key-value –data is stored in tables, with each value associated key. This value scales well, but the implementation is almost entirely the responsibility of the developer.

5. SQL in Hadoop – Packet queries in Hadoop are written in a SQL-based language, the internal rendering of which uses the display-convolution infrastructure.

6. Updated SQL-this type combines the scalability of NoSQL databases with the advantages of relational databases. All these databases use the SQL interface and the relational data model.

7. Graph database - not for all tasks the tabular format is optimal. Some tasks are more suitable for presentation in the form of graphs and storage in graph databases. A classic example of this kind is the social network.

Planning Tools. Planning tools make it easy to automate recurring operations and launch tasks by events (for example, when a new file appears in a folder). They are similar to traditional programs like CRON in Linux, but designed specifically for large data. For example, such tools can run the MapReduce task when a new data set appears in the directory.

Benchmarking tools. This class was developed for optimizing the big data installing for providing standardized profiles. Profiles are made on basis of a representative set of operation with big data . Using an optimized infrastructure leads to significant savings. For example, saving 10% in a cluster of 100 servers is equal to the cost of 10 servers [2].

Deploying the system. Preparing an infrastructure for large data is not an easy task. The deployment tools of the system manifest themselves when deploying new applications in large data clusters. They largely automate the installation of large data components.

Programming of Services. Suppose you have created an application for predicting the results of football matches and now want to give all those who wish access to the projections of your application. However, you have no idea what architecture or technology will be used in their systems. Service programming tools provide access to large data applications as to a server. The most famous example is REST-service;

REST means “Representational State Transfer”. These services are often used to transfer data on the web sites.

Security. It is important that everyone can access the necessary data. To do this, you need to organize a mechanism for precise control over access to data, but you do not want to do this at the level of each individual application. The security of large data allows you to create a centralized and highly accurate system for managing access to data.

3 Mathematical modeling of call services

3.1 Analyzing and processing contact center’s statistical data

The statistical data were collected and analyzes from the contact center of one of the telecommunications company in Almaty.

From every contact center, it is possible to obtain the following information:

− the cause of the communication failure when receiving the service of the telecommunications company, information about the time of communication restoration;

− rules for filing an application for the elimination of violations.

In a month, the contact center receives about 100−120 thousand calls. Calls come from public telecommunications network, from the mobile network and via the Internet. Statistics are collected not only about incoming calls, but also about outgoing calls.

The contact center regularly collects statistics on incoming calls, namely, the number of serviced and lost calls, the average service time.

Before evaluating the average service, time and analyzing the factors affecting this value, it was necessary to collect statistical data on the following values:

− number of incoming calls;

− the number of serviced calls;

− number of lost calls;

− average call processing time.

Statistical data on the calls arriving at the contact center were collected within a month.

Call−center regularly collects statistics about calls. Information about the number of incoming (incoming traffic), serviced and rejected (denied service) calls, the average service time of the application is recorded.

The object of the study was the Contact center of one of the communication providers in Almaty. The center, designed for 20 workplaces, operates daytime from 8.00 to 18.00 hours seven days a week. The total staff of the operator is 35 people. Calls come both from mobile phones and from the public telephone network. The number of calls within an hour reaches 250 calls, during the day − 3000 and within a month − 101000 calls.

Operators of the Contact center consult on various issues related to telecommunications services. It accepts applications for installation, transfer,

re−registration, unlocking a city phone, as well as for connection to services provided by a telecommunications operator.

The analysis was carried out on the basis of statistical data on the incoming traffic of contact center for April and May 2018.

For the expediency of applying the method of trend extrapolation, it is necessary to identify the presence of a trend in a number of dynamics, i.e. perform series check for stationarity.

A simple method of analyzing stationarity is to compare its parameters, grouped for the beginning, middle and end of the series. Having divided the series into two groups, for each of them calculate the average value by the formula (3.1) and the variance value by the formula (3.2). Then these parameters are compared for individual segments of the source series.

$$\hat{Y} = \frac{\sum Y_i}{n} \tag{3.1}$$

where n−number of levels in a row, $\sum Y_i$−the sum of all observations;

$$\delta^2 = \frac{\sum (Y_i - \hat{Y}_1)^2}{n_1 - 1} \tag{3.2}$$

where Y_i−individual observations,

$\hat{Y}_1$−arithmetic mean,

n−number of observation.

Thus, in order to establish the stationarity of the series relative to its central tendency (the absence of a non−cyclic trend), comparison of the arithmetic means is made according to the t criterion (Student's test):

$$t = \frac{\hat{Y}_2 - \hat{Y}_1}{\sqrt{\frac{\delta_1^2 (n_1 - 1) + \delta_2^2 (n_2 - 1)}{n_1 + n_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}} \tag{3.3}$$

Let's consider the method of analysis for stationarity of series for January. We divide the series into two parts: the first part, from the first to the 15th of April, and the second from the 16th to the 30st of April.

We find the arithmetic mean for both parts of the series.

$$\hat{Y}_1 = \frac{\sum Y_I}{n} = \frac{"914 + 5675 + 5035 + 4486 + 4010 + 4367 + 2019 + 983 + (\ldots) + 4174 + 3863 + 1740 + 938}{15} = 2490$$

$$\hat{Y}_2 = \frac{\sum Y_I}{n} = \frac{3940 + 4090 + 3677 + 3693 + 3209 + 1515 + 5034 + (\ldots) + 4552 + 4042 + 1350 + 1706}{15} = 3298$$

We find the variances for both parts of the series.

$$\delta_1^2 = \frac{\sum (Y_i - \hat{Y}_1)^2}{n_1 - 1} = \frac{(914 - 3490)^2 + (5675 - 3490)^2 + (5035 - 3490)^2 + (,,,)^2 + (938 - 3490)^2}{14} = 2788967$$

$$\delta_{21}^2 = \frac{\sum (Y_i - \hat{Y}_1)^2}{n_2 - 1} = \frac{(3490 - 3298)^2 + (4090 - 3928)^2 + (3677 - 3298)^2 + (,,,)^2 + (1706 - 3298)^2}{14} = 1659932$$

Now we find the criterion

$$t = \frac{\hat{Y}_2 - \hat{Y}_1}{\sqrt{\frac{\delta_1^2 (n_1 - 1) + \delta_2^2 (n_2 - 1)}{n_1 + n_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}} = \frac{3490 - 3298}{\sqrt{\frac{2788967 \cdot (15 - 1)}{30} \sqrt{\frac{1}{14} + \frac{1}{14}}}} = \frac{192}{883} = 0{,}21$$

The degree of freedom in calculating the t−test is defined as

$$d_f = n_1 + n_2 - 2$$

$$d_f = 14 + 14 - 2 = 26$$

The obtained value is compared with the above estimates of the criterion for P=0,05.

t_{table}=2,04

$t_{calc.}$=0,21

If the estimated $t_{calc.} < t_{table}$ less than the tabulated one in the graph corresponding to the significance level P = 0.05, then the differences between the averages are considered to be insignificant.

3.2 Forecasting by the method of trend extrapolation

The linear trend is a function y = ax + b, where

The value of x is the number of the period in the time series, the month number;

y− is a sequence of values that we analyze (for example, traffic by day of the week.)

b −point of intersection with the y−axis in the graph (minimum level);

a− is the value at which the next value of the time series increases;

Moreover, if a> 0, then the dynamics of growth is positive,

If a <0, the dynamics of the trend is negative, i.e., for example, the traffic forecast will give a big error. A linear trend should be used to predict time series whose data is increasing or decreasing at a constant rate.

To calculate the forecast, we confine ourselves to the period of prehistory from April 1 to April 28, we will find projections for the 29.30 and 31 days of the month and compare them with the actual traffic data for these days the same days.

Table 3.1−Forecast by the method of trend extrapolation (April 2018)

Date	Incoming traffic	y=−29,1t+3845,4
01.04.2018	914	3816
02.04.2018	5675	3786
03.04.2018	5035	3758,1
04.04.2018	4486	3729
05.04.2018	4010	3699,9
06.04.2018	4367	3670,8
07.04.2018	2019	3641,7
08.04.2018	983	3612,6
09.04.2018	4653	3583,5
10.04.2018	4907	3554,4
11.04.2018	4597	3525,3
12.04.2018	4174	3496,2
13.04.2018	3863	3467,1
14.04.2018	1740	3438
15.04.2018	938	3408,9
16.04.2018	3940	3379,8
17.04.2018	4090	3350,7
18.04.2018	3677	3321,6
19.04.2018	3693	3292,5
20.04.2018	3209	3263,4
21.04.2018	1515	3234,3
22.04.2018	854	3205,2
23.04.2018	5034	3176,1
24.04.2018	4552	3147
25.04.2018	4080	3117,9
26.04.2018	3814	3088,8
27.04.2018	3925	3059,7

The linear trend calculated for 28 days of the month (period of prehistory) is described by a function of the form y=−29,1t+3845,4. For forecasting, we need to extend the trend line and determine its values.

Now calculate the linear trend for the month of February. For the calculation of the forecast period in the history confine ourselves to the 1 st to May 28, we find projections for the 29, 30 and 31 of the month and compare them with the actual data traffic these days.

Calculated for 28 days of the month (period of prehistory) linear trend is described by a function of the form y= 34,0x+3124,4. The linear trend of traffic for May is shown in Figure 3.2

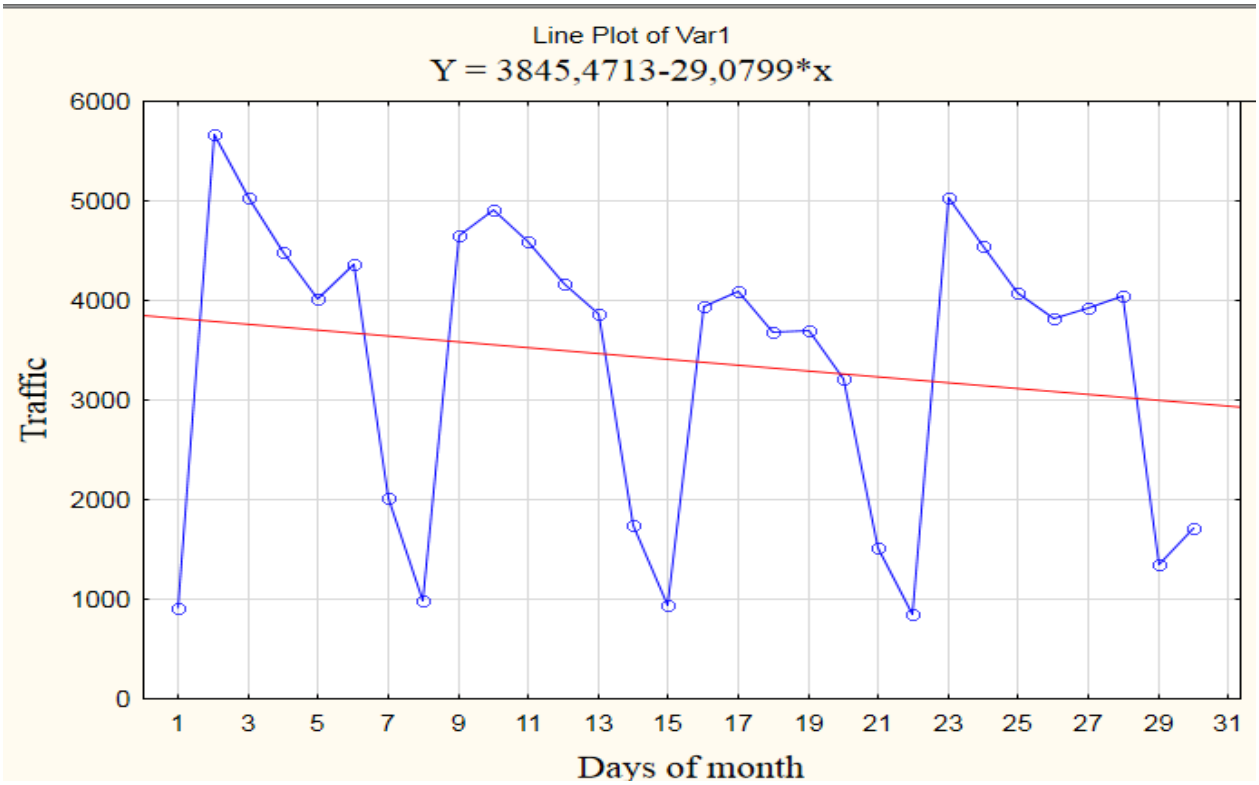

Figure 3.1 − Linear traffic trend for April

Table 3.2 Forecast by the method of trend extrapolation (February 2013)

Date	Incoming traffic	y=34,0x+3124,4
01.05.2018	1148	3158,4
02.05.2018	4518	3192,4
03.05.2018	4683	3226,4
04.05.2018	4947	3260,4
05.05.2018	4862	3294,4
06.05.2018	1468	3328,4
07.05.2018	1401	3362,4
08.05.2018	1420	3396,4
09.05.2018	864	3430,4
10.05.2018	5775	3464,4
11.05.2018	5630	3498,4
12.05.2018	2477	3532,4
13.05.2018	1034	3566,4
14.05.2018	5663	3600,4
15.05.2018	5150	3634,4
16.05.2018	4983	3668,4
17.05.2018	4499	3702,4
18.05.2018	4632	3736,4
19.05.2018	2093	3770,4

Table 3.2 Continuation

Date	Incoming traffic	y=34,0x+3124,4
20.05.2018	1059	3804,4
21.05.2018	4798	3838,4
22.05.2018	5221	3872,4
23.05.2018	4640	3906,4
24.05.2018	4286	3940,4
25.05.2018	4256	3974,4
26.05.2018	1966	4008,4
27.05.2018	1124	4042,4
28.05.2018	4799	4076,4

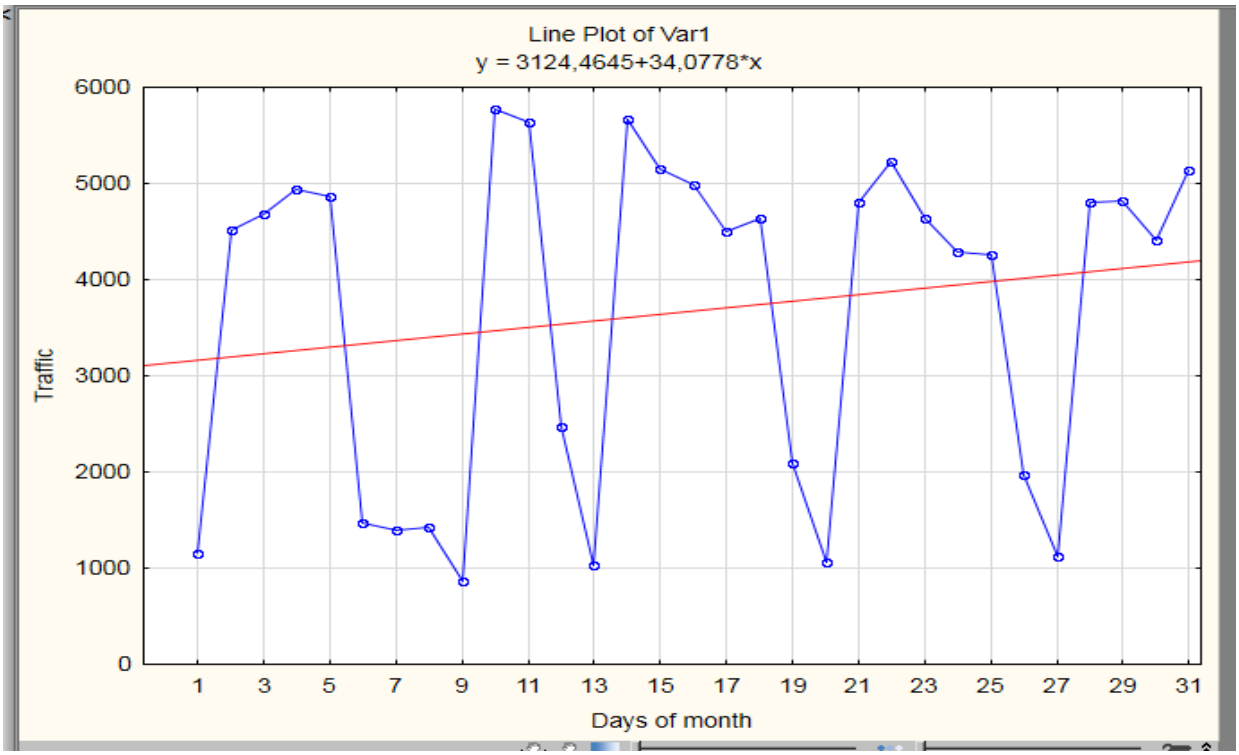

Figure 3.2− The linear trend of traffic for May is shown in Figure 3.2

Comparison of the forecast estimates with the actual data (Tables 3.1 and 3.2) made it possible to determine that the absolute forecast error for one day ahead was 263 calls, respectively the relative error was 6%.

3.3 Forecasting using the weighted moving average method

The second method to which the forecast was calculated was a moving average. Due to the fact that the considered time series has a trend and a seasonal component with a seasonal period of n = 7 days, the length of the smoothing interval is assumed to be n = 2p + 1. Since the value of n is odd, the moving average will be given by formula 3.4

$$y_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \ldots + y_{t+p+1} + y_{t+p}}{2p+1} \qquad (3.4)$$

where y_i−the actual value of the i−th level;

y_t−the moving average at time t;

2p+1 − length of the smoothing interval.

Table 3.2 Forecasting by weighted moving average method (April 2018)

Date	Actual traffic	Forecasting by weighted moving average method
4	4486	3786
5	4010	3796
6	4367	3650
7	2019	3632
8	983	3648
9	4653	3671
10	4907	3599
11	4597	3559
12	4174	3553
13	3863	3451
14	1740	3334
15	938	3203
16	3940	3134
17	4090	3041
18	3677	3008
19	3693	2996
20	3209	3153
21	1515	3219
22	854	3276
23	5034	3294
24	4552	3396
25	4080	3757
26	3814	3828
27	3925	3352

Accept smoothing period of one week (n = 7).

The incoming traffic and its alignment with the moving average on the days of the week are graphically presented in Figure 3.3. Forecast values of incoming traffic for 28,29 and 30 January are assumed equal to the smoothed value on April 27.

Now we calculate the forecast using the moving average method for May. The incoming traffic and its alignment with the moving average on the days of the week are graphically presented in Figure 3.4.The predicted values of incoming traffic for May 30 and May 31 are equal to the smoothed value for May 29.

Figure 3.3 reflects the series, smoothed by the moving average method for May, that is, the series 1 − incoming traffic, the series − 2 alignment using the moving average.

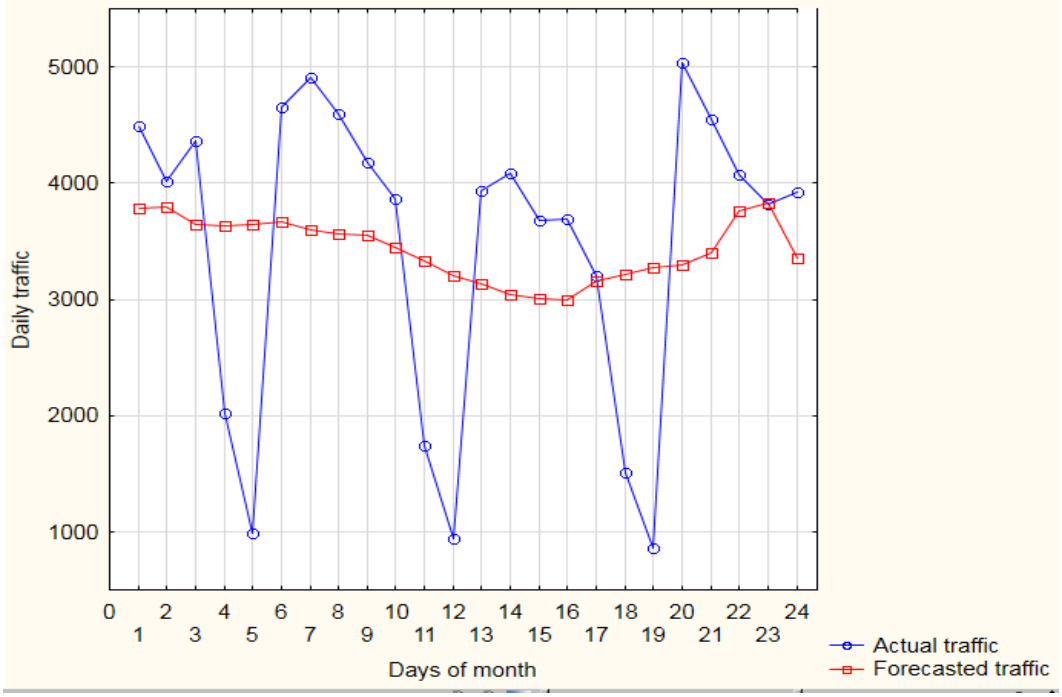

Figure 3.3− Smoothing the series using the moving average method (April 2018)

Table 3.4 Forecasting by weighted moving average method (May 2018)

Date	Actual traffic	Forecasting by weighted moving average method
4	4947	3289
5	4862	3328
6	1468	2806
7	1401	2962
8	1420	3060
9	864	2719
10	5775	2657
11	5630	3266
12	2477	3799
13	1034	4387
14	5663	4205
15	5150	4062
16	4983	4007

Table 3.4 Continuation

Date	Actual traffic	Forecasting by weighted moving average method
17	4499	4011
18	4632	3887
19	2093	3897
20	1059	3848
21	4798	3818
22	5221	3764
23	4640	3746
24	4286	3755
25	4256	3756
26	1966	3699
27	1124	3666
28	4799	3787
29	4826	3179

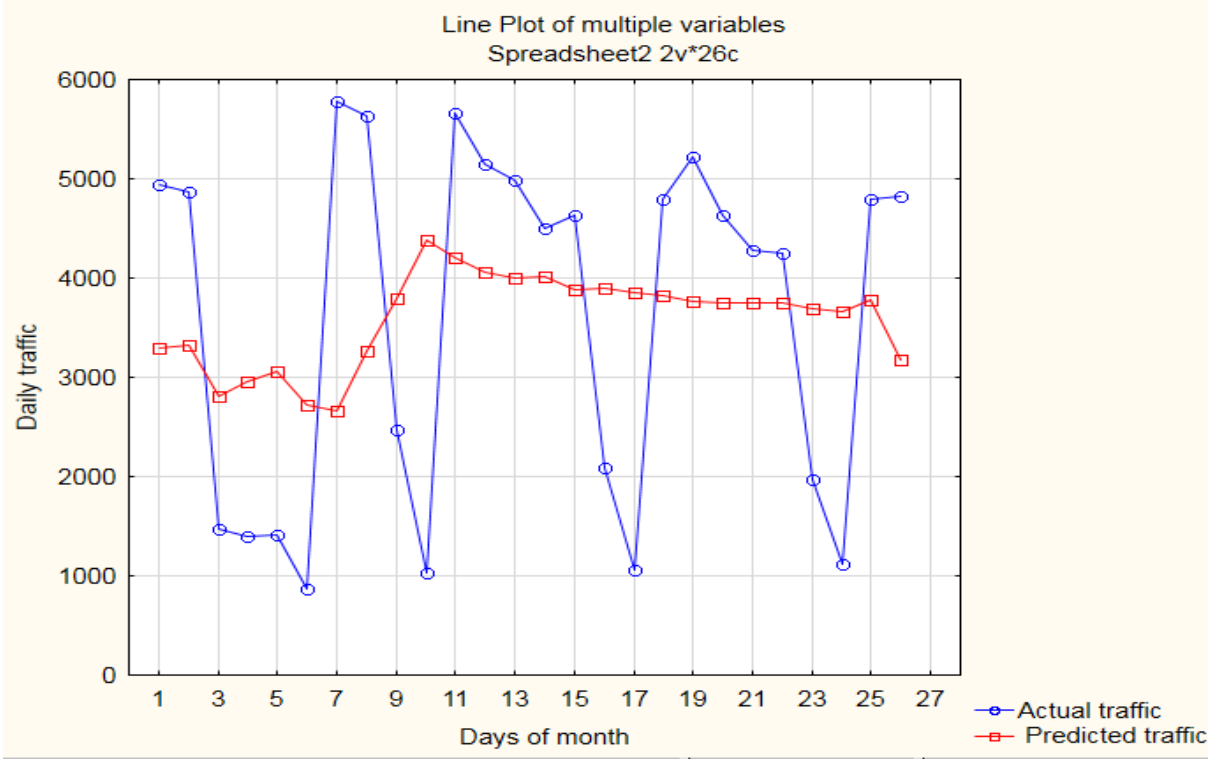

Figure 3.3− Smoothing the series using the moving average method (May 2018)

3.5 Prediction of traffic on a particular day of the week

As can be seen above, we predicted traffic for the month by two methods. Since the highest load of incoming traffic falls on Thursday, we calculate the forecast load for this day of the week, using the incoming traffic data for Thursday for two months (April and May).

To calculate the forecast by the method of trend extrapolation, we limit ourselves to the prehistory period, which includes 9 Thursdays (4 in April and 5 in May), we will find the forecasted values for May 31 (the ninth Monday). The predicted value will be compared with the actual traffic for this number. Plotted in Figure 3.4 graph and given on the x−axis − time ,on the y−axis − traffic.

Calculated linear trend is described by a function of the form Y = 139x + 3752.

Table 3.7. − Calculation of the forecast by the method of trend extrapolation (Thursday)

Date	Incoming traffic	y=139x+ 3752
05.04.2018	4010	3891
12.04.2018	4174	4030
19.04.2018	3693	4169
26.04.2018	3814	4308
03.05.2018	4683	4447
10.05.2018	5775	4586
17.05.2018	4499	4725
24.05.2018	4286	4864

The forecast by the trend extrapolation method on May 31 (the ninth Thursday) was 5003 calls.

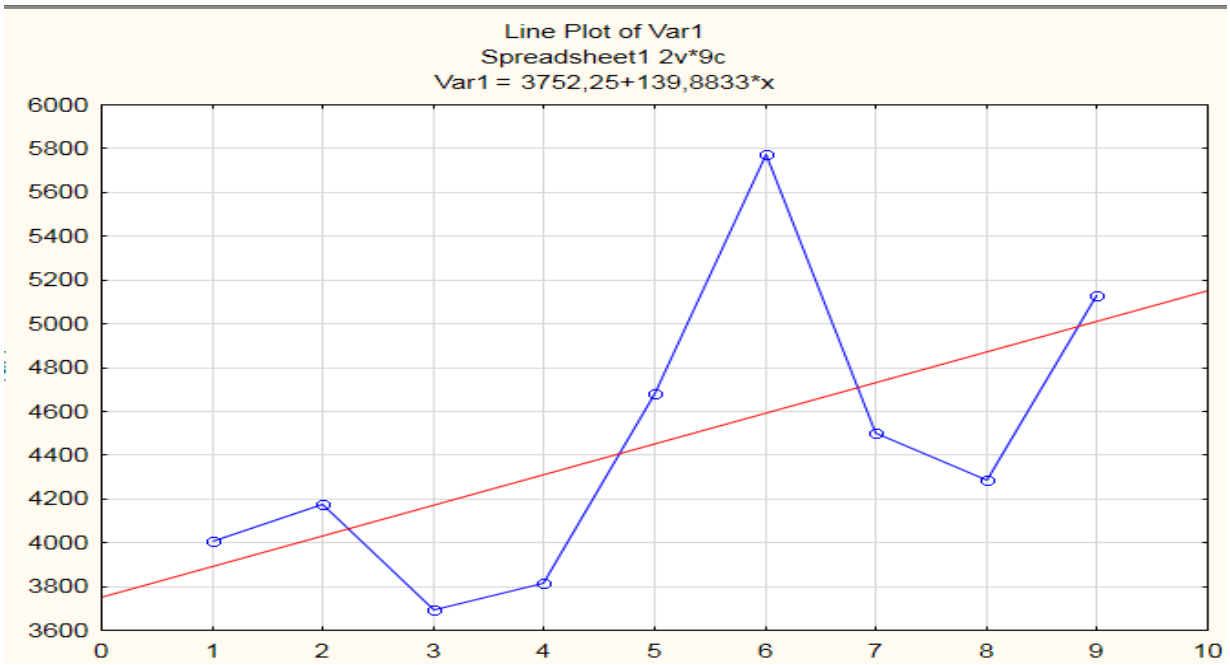

Figure 3.4− Linear trend of traffic for Thursdays

The second method by which the forecast for the eighth Thursday was calculated was a moving average. Since the value of n is odd, the moving average will be given by formula (3.4)

$$y_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \ldots + y_{t+p+1} + y_{t+p}}{2p+1}$$

where y_i–the actual value of the
i–th level;
y_t–the moving average at time t;
2p+1 – length of the smoothing interval.

Table 3.8 – Calculation using the moving average method for Thursday

Date	Incoming traffic	Forecasting by weighted moving average method
05.04.2018	4010	3833
12.04.2018	4174	3959
19.04.2018	3693	3893
26.04.2018	3814	4063
03.05.2018	4683	4757
10.05.2018	5775	4985
17.05.2018	4499	4853
24.05.2018	4286	4638

Incoming traffic and its alignment by means of a moving average for a specific day of the week are graphically presented in Figure 3.4.

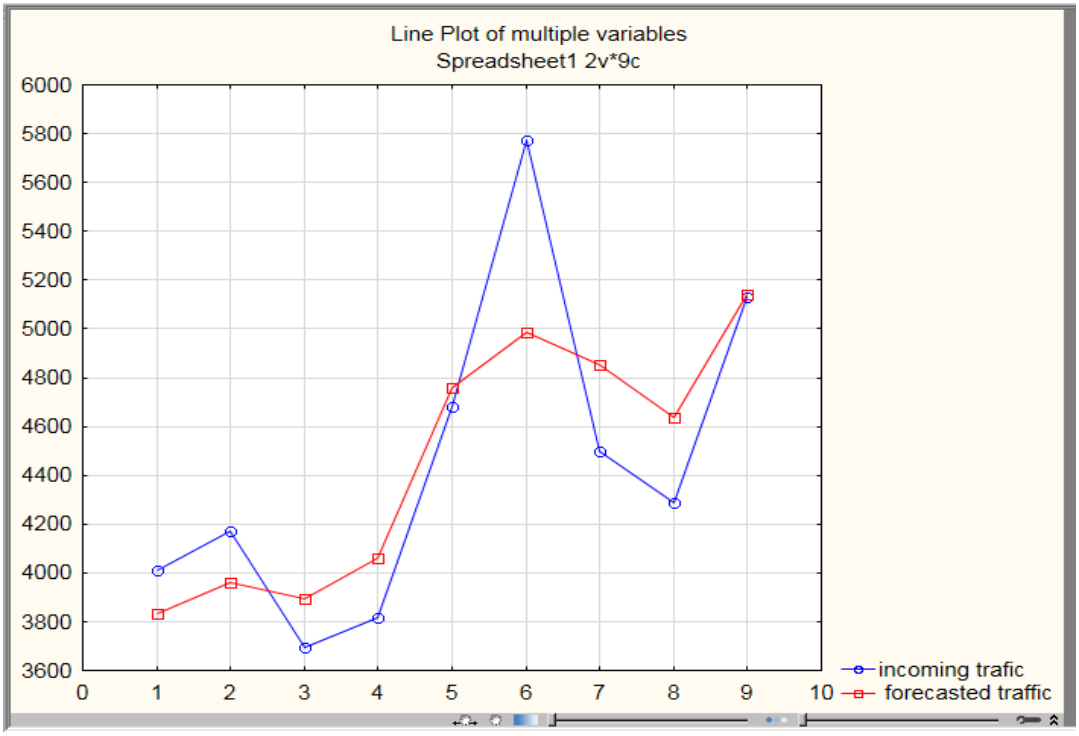

Figure 3.5– Smoothing series by the moving average method

Forecast different estimates of incoming traffic as of May 31 are equal to the smoothed value on the ninth Thursday 4989 calls.

Let's compare the predictions calculated by the two methods in Table 3.9

Table 3.9 Comparison of forecast estimates on May 31

Date	Actual traffic	Weighted moving average method	Method of trend extrapolation
31.05.2018	5131	4989	5003

Table 3.10−Estimate the accuracy of the forecast for Thursday

Date	Actual traffic y_i	Absolute error, $\Delta_1 = y_i - y_x$		Relative error, $\beta = \frac{\Delta_1}{y_i} \cdot 100\%$	
		Weighted moving average method	Method of trend extrapolation	Weighted moving average method	Method of trend extrapolation
31.05.2018	5131	142	128	2,77%	2,49%

The approximation error, calculated for forecasting on each Thursday by formula (3.6), when calculating by the trend extrapolation method taking into account the unevenness index is 2,49% and the moving average method gives an error of approximation of 2,77%.

The results of the short−term forecasting of incoming traffic of the Contact center on Thursday made it possible to reveal that the greatest accuracy is provided by the forecast calculated by the trend extrapolation method taking into account the unevenness index.

3.6 Estimation of forecast accuracy

Obviously, the accuracy of the forecast is higher, the smaller the error value, which is the difference between the predicted and the actual value of the investigated quantity.

Table 3.11 − Projected Traffic Estimates (April 2018)

Date	Actual traffic	Weighted moving average method	Method of trend extrapolation
28.04.2018	4042	3803	3030,6
29.04.2018	2840	1586	3001,5
30.04.2018	2806	1655	2972,4

The approximation error calculated for the 1−3 days forecast by formula (3.6) was 12 % for the linear extrapolation calculation, when calculating by the trend the moving average method gives the approximation error 30%.

$$\varepsilon = \frac{1}{n} \cdot \sum \frac{|y - y_x| \cdot 100}{y} \quad (3.6)$$

Table 3.12−Estimate of forecast accuracy (April 2018)

Date	Actual traffic y_i	Absolute error $\Delta_1 = y_i - y_x$		Relative error $\beta = \frac{\Delta_1}{y_i} \cdot 100\%$	
		Weighted moving average method	Method of trend extrapolation	Weighted moving average method	Method of trend extrapolation
28.04.2018	4042	239	1011,4	6%	25%
29.04.2018	2840	1254	161,5	44%	6%
30.04.2018	2806	1151	166,4	41%	6%

Table 3.13 − The forecasted traffic estimates (May 2018)

Date	Actual traffic	Weighted moving average method	Method of trend extrapolation
29.05.2018	4826	3179	4110,4
30.05.2018	4408	3134	4144,4
31.05.2018	5131	3289	4178,4

Table 3.12−Estimate of forecast accuracy (May 2018)

Date	Actual traffic y_i	Absolute error $\Delta_1 = y_i - y_x$		Relative error $\beta = \frac{\Delta_1}{y_i} \cdot 100\%$	
		Weighted moving average method	Method of trend extrapolation	Weighted moving average method	Method of trend extrapolation
28.04.2018	4826	1647	715,6	34%	15%
29.04.2018	4408	1274	263,6	29%	6%
30.04.2018	5131	1842	952,6	36%	19%

The approximation error calculated for the May forecast for 1−3 days by formula (3.6) was 13% for the linear extrapolation calculation, and the moving average method gives an error of 33 % approximation.

Calculations showed that the method of linear extrapolation of the trend provides the greatest forecast accuracy in April and May of the month. The error does not exceed 13%.

The results of traffic forecasting for a particular contact center were presented. The estimation of stationary traffic, will reveal the trend in the statistical traffic data for the month. The forecasted traffic estimates are calculated using different methods with a forecast horizon of 1−3 days. The forecast for a particular day of the week is determined, which, as a rule, has the greatest load. The method of short−term forecasting is chosen, which ensures the greatest accuracy of the forecast.

3.7 Determination number of operators

As a mathematical model of the contact center, the following queuing systems are used (model designations are given according to the Kendall symbols [5,14]):

− a model with expectation and an infinite queue M / M / N / ∞;
− a model with expectation and an arbitrary distribution of the service time M/G /N;
− model with expectation and the final queue M / M / N / B;

It should be noted that almost all considered models are described by the simplest incoming call flow and exponential distribution of service time, with the exception of the M / G / N model having a common service time distribution.

In practice, systems with the ability to delay incoming calls in the queue are much more common, if their immediate service is impossible. The model with the expectation M/M/N, described by the second Erlang formula, assumes the presence of an infinite queue, and the main area of its application is large contact centers with a large number of operators and the number of incoming lines.

Using statistical data, we determine which model can be described by the investigated call−center.

The M/M/N model is convenient to use and is recommended for calculating the staff of operators in those contact centers where it is important to accurately predict the average queue length. The discrepancy with the empirical value of this parameter for the data averaged over the entire study interval did not exceed 3.3%. Similar results are given also by the model with a finite queue M / M / N / B, but analytically it is more complicated than the M / M / N model. Therefore, it is advisable to apply this model to contact centers where it is important to take into account the possibility of locks, and the number of incoming lines exceeds the number of operators by no more than 10%.

As expected, the closest to the empirical data is the M / G / N model (the uncertainty was less than 0.1% for the whole simulation interval). Therefore, if the prediction of the waiting time is critical for a DSC, then it is advisable to use the

M/G/N model, which takes into account the fact that the time for servicing calls is not distributed exponentially.

The correct choice of a model for planning the characteristics of contact centers is an extremely important task, but it is not easy to solve it. When planning and developing new contact centers, it is difficult to accurately predict the incoming load, maintenance time and other important parameters in advance. After all, the successful choice of the model depends first of all on the volume and variety of the data collected during the operation, their analysis and processing. At the same time, it is very important that the statistics showing the entire sequence of contact centers functioning are complete and detailed.

This allows you to efficiently and accurately calculate the functional parameters of contact centers, as well as make timely changes to its operation. The lack of such data makes it possible to make only approximate estimates of many important characteristics of the quality of service, which, incidentally, is characteristic for the design of most contact centers, when the calculation data for simple models such as M/M/N, M/M/N/N, M/M/N/B.

Mathematical modeling has shown that none of the models considered is absolutely accurate, nor is it absolutely unsuitable. One can judge only the degree of accuracy of the prediction of certain parameters of real systems. The models considered can serve only as a guide for choosing the right strategy for developing new contact centers.

They are designed to solve problems at the lowest level of the management hierarchy contact center. The solution of this task, of course, will not be the key to success the functioning of the whole system as a whole, because at the higher levels of management no less important long−term tactical and strategic decisions must be made. However, the results could be the basis for selecting the appropriate model at the stage of conceptual design contact center.

Consider the case of s identical service channels (M/M/S/8) with the maintenance intensity μ of each device. In our case, the service channels are operators, but in the future we will follow the terms of the queuing theory.

With the simplest input stream and exponential service time, the channel selection rule does not affect the functioning of the system, i.e. on the number of applications that are in the system, for the duration of waiting.

In Figure 3.5 is presented transition graph.

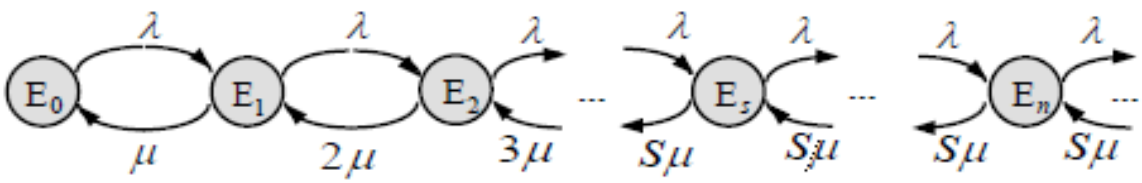

Figure 3.6 − Transition graph

The intensity of service increases with the number of occupied channels. Once all channels are busy with applications, the intensity of service in the system

ceases stops to grow, newly arrived applications become queued, waiting for the release of any of the channels. From the graph we construct the equations of state probabilities:

$$P_1 = \frac{\lambda}{\mu} \cdot P_0,$$

$$P_2 = \frac{\lambda^2}{\mu^2 \cdot 2} \cdot P_0,$$

....

$$P_n = \frac{\psi^N}{n!} \cdot P_0 \text{ where } 0 \le n \le s,$$

$$P_S = \frac{\psi^S}{s!} \cdot P_0,$$

....

$$Pn = \frac{\psi^n}{s! s^{n-s}} \cdot P_0 \text{ ,where n}\geq\text{s.}$$

Considering that $\sum P_n = 1$,

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{\psi^n}{n!} + \sum_{n=s}^{\infty} \frac{\psi^n}{n! s^{n-s}}} = \frac{1}{\sum_{n=0}^{\infty} \frac{\psi^n}{n!} + \frac{\psi^s}{s!(1-\frac{\psi}{s})}}$$

We obtain the basic characteristics of the system:

$$P_{ref} = 0,$$

$$q = 1 - P_{ref} = 1,$$

Absolute bandwidth:

$$A = q \cdot \lambda = \mu \cdot z,$$

hence the average number of occupied channels:

$$z = \psi$$

Average number of requests in the queue:

$$\bar{v} = \sum_{n=s+1}^{\infty} (n-s) \cdot P_n = \sum_{n=s+1}^{\infty} (n-s) \frac{\psi^n}{s! s^{n-s}} \cdot P_0,$$

after the transformations:

$$\bar{v} = \frac{\psi^{s+1}}{(s-1)!(s-\psi)^2} \cdot P_0.$$

Total number of applications in the system: $\bar{n} = \bar{v} + \bar{z}$

Average length of stay in the system:

$$\bar{u} = \frac{\bar{n}}{\lambda} = \frac{1}{\mu}\left[\frac{\psi^s}{(s-1)!(s-\psi)^2} P_0 + 1\right]$$

All formulas presented above are valid only for a stationary regime. In to have a stationary mode, order for a multichannel system with an infinite queue should satisfy condition $\frac{\lambda}{\mu \cdot s} < 1$.

At definition of number of operators of the call−center now use the C−formula of Erlang [10]. Let's consider it in more detail.

Let a simple flow of calls with parameters λ arrive to a fully accessible system with expectation. The length of service is distributed by exponential law:

$$F(t) = 1 - e^{-\beta \cdot t},$$

where $t_3 = \frac{1}{\beta}$ – average service time.

If all lines are busy at the time of the call, the call is queued and waiting for service. Calls from the queue are selected in the order they are received. In Kendall's symbols we have a system M/M/v/r=∞/FF.

It is required to determine the probability of waiting for an incoming call in the queue, the mathematical expectation of the waiting time and the length of the queue.

The problem in such a way was first formulated and solved Erlang. In the problem posed by Erlang, the number of waiting places r =∞.

The solution of this task allows to evaluate the quality of service characteristics of a fully accessible system with the expectation:

a) the probability of waiting in the queue for an incoming call

$$P = P_t = P(\gamma > 0) = \frac{E_v(A)}{1 - \left(\frac{A}{v}\right) \cdot \left[1 - E_v(A)\right]} = D_v(A) \qquad (3.7)$$

Where A− incoming traffic

v − number of lines or number of operators;

$$P_0 = P(v) = \frac{A^v / v!}{\sum_{x=0}^{v} \frac{A^x}{x!}} = E_v(A)$$

Erlang's first formula, in some sources it is called Erlang's B formula;
$D_v(A)$ is called the second Erlang formula or the Erlang's C−formula.

b) The expectation of the length of stay in the queue

$$t = \frac{L}{A \cdot P} \qquad t = \frac{i}{v - A} \tag{3.2}$$

Thus, in the Erlang formula (3.1) we know the values of the incoming load or the number of calls arriving for a certain period, for example during the hour of maximum load, the probability of waiting P, which we set ourselves.

Let us find the number of operators that will provide us with a given probability

expectations using the formula (3.3):

$$V = A \cdot (1 + \frac{1}{L \cdot P}) \tag{3.4}$$

where L− is the average queue length.

And now let's look at the real data that shows how the value of P changes over the course of a day − the probability of waiting for a call in the queue.

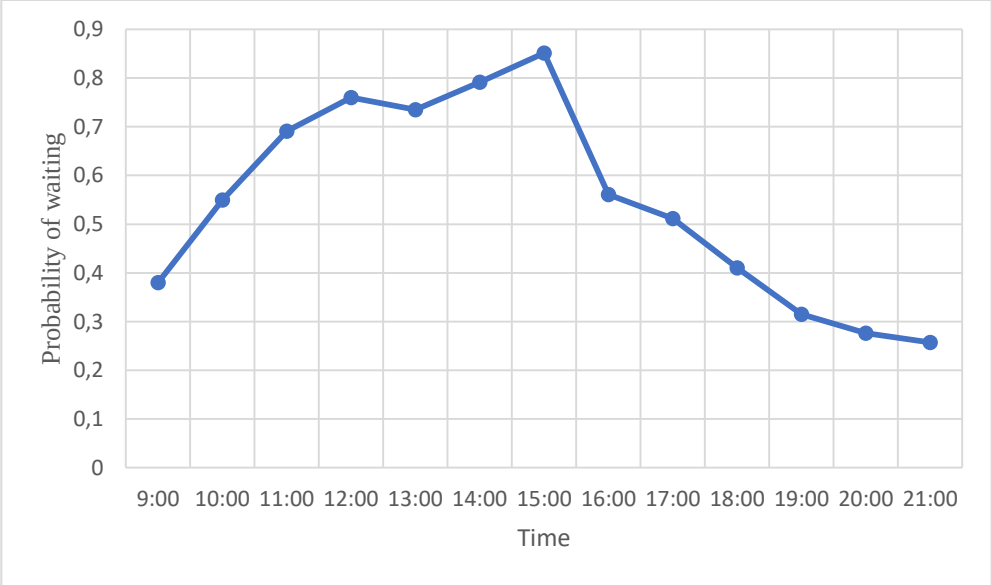

Figure 3.6 − Changes of the probability P during working hours

The graph in Figure 3.6 shows that the highest probability of waiting for a call in the queue was observed between 15.00 and 16.00, and is P = 0.86. The smallest probability P is for the period 20:00−21:00 and is 0.26. Figure 3.7 shows the schedule of load changes during this working day. The greatest number of calls was received during the period 15.00 − 16.00 hours, and made 616 calls, and the smallest number of calls arrived at the end of the day and made 186 calls.

As we can see from the analysis of the received statistical data, the time of the greatest and the least load does not coincide with the time of the greatest and the least probability of waiting in line. This suggests that the work of the contact center was not provided by a sufficient number of operators.

It can be seen from formula (3.4) that the value of v is directly proportional to the value of A, that is, if we take as the constant value the probability of waiting P equal to 0.15 and the length of the queue L = 10, then we determine the form of the dependence (3.4).

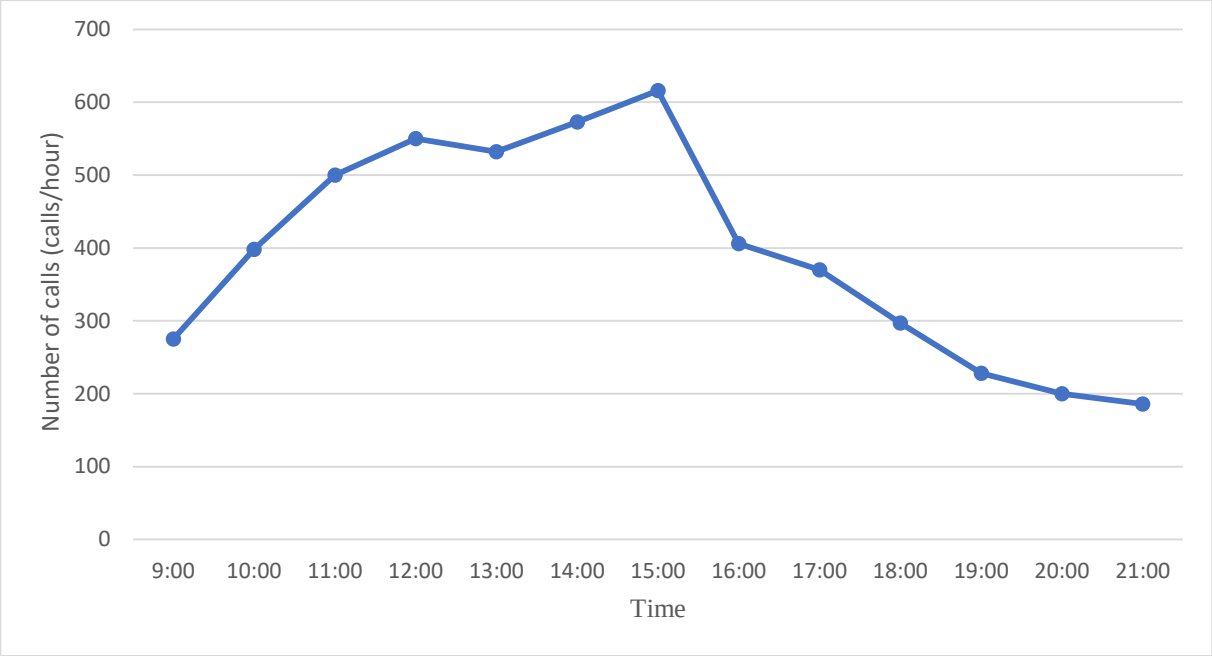

Figure 3.7 - Receipt of calls during the workday

The quantity A is defined as follows. As it is already known from the theory of teletraffic [1], the value of A is calculated by the formula:

$$A = \lambda \cdot t ,$$

where λ is the intensity of the incoming stream, in our case the number of calls coming during in 1 hour;

t – call service time.

Let the number of calls vary from 200 to 500 in steps of 50. The average time for the call is 2 minutes. Then the value of A will vary from 3.3 to 21.5 Erlang.

Table 3.3 - Load value A

Parameters	Values									
Number of calls ,n	200	250	300	350	400	450	500	550	600	650
Load value A(Erlang)	6,7	,3	0,0	1,6	3,3	4,9	6,6	8,2	9,9	1,5

Dependence of 3.4 for this example would look like

$$V = 2.15 \cdot A$$

The result obtained is represented in the form of Table 3.4.

Table 3.4 - Number of operators

Parameters	Values									
Number of calls ,n	200	250	300	350	400	450	500	550	600	650
Load value A(Erlang)	6,7	8,3	10,0	11,6	13,3	14,9	16,6	18,2	19,9	21,5
Number of operators	14	18	22	25	29	32	36	39	43	46

Thus, to ensure the probability of waiting for a call in the queue is not higher than 10% and the queue length is not more than 10, the number of operators working in one shift at the hour of the greatest load should be increased to 46. In addition, forecasting is more efficient to produce by the hour of the day.

Figure 3.8 shows the dependence of the number of operators on the number of incoming calls. Thus, having established by forecasting the number of operators for the previous days, we can determine the number of operators on the schedule, so that the probability of waiting in the queue does not exceed 10%.

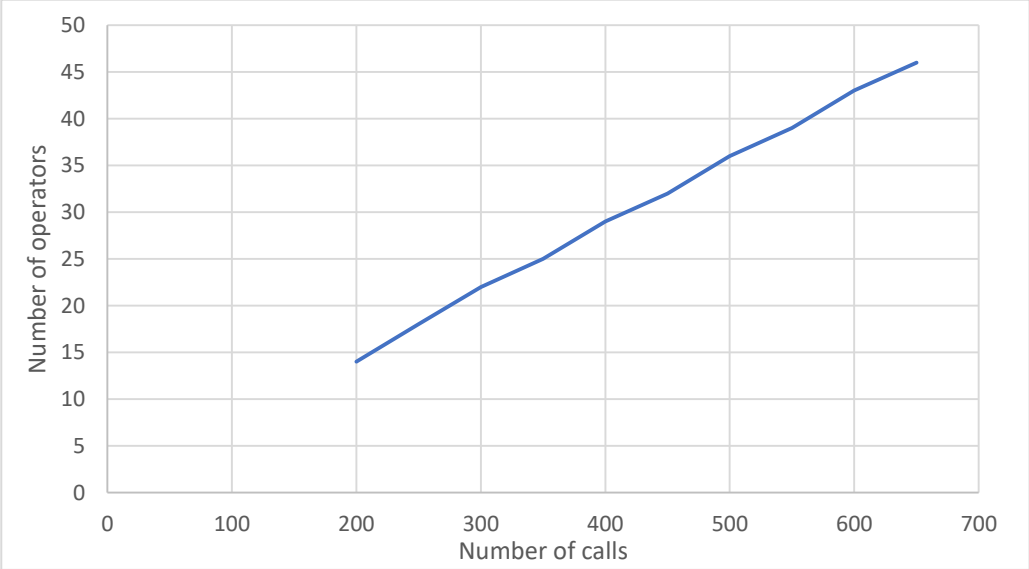

Figure 3.8 - Dependence of the number of operators on the number of calls

4 Safety of vital functions

4.1 Review of working conditions

In this graduation project, in this chapter was analyzed the effectiveness of the organization of the work of contact center operators, I conduct the calculation of the premises for the safety of life activities.

According to the nature of the environment, the room belongs to a classroom with a normal environment, since the air humidity does not exceed 60% and the temperature does not exceed 30℃ , there is no technological dust, an active chemical environment.

The equipment operates optimally under the following conditions:

- temperature from 10℃ to 40℃

- humidity from 5% to 75%

- Power supply: alternating current-voltage 220 V, frequency 50/60 Hz, current 2 - 5 A.

Since all equipment is certified, the class of professional risk is defined as the minimum [20].

Electrical devices for safety measures refer to devices with an operating voltage of up to 1 kV.

By the degree of danger of electric shock, the room belongs to a class without increased danger, since it meets the requirements:

- dry (relative humidity of air does not exceed 60%);

- with a normal temperature (in the cold period, the maximum temperature does not exceed 28 and not fall below 23 degrees, the temperature in the summer is 20-24 degrees, the air speed is 0.1 m / s);

- with insulated floors (non-conductive surfaces, with antistatic impregnation);

- dust-free (no technological dust is released);

- does not have ungrounded objects.

The specifics of the process performed, the properties of the substances and materials use, and the availability of electronic equipment, the premises are classified as D fire hazard category [40].

The workplace for the performance of the work corresponds to the requirements of state standard (GOST 12.2.032-78, " Workplace in the performance of sitting jobs." General ergonomic requirements ", GOST 12.2.033-78".), Electrically adjustable office tables with two operating modes are installed, sitting and standing, which allows to reduce the load by alternating the position during operation, as well as the break during the last 4 hours of operation every hour for 15 minutes [41].

The operator room has the following dimensions: length A = 22 m, hall width B = 16 m, height H = 3,5 m; the height of the work surface h_1=0,8m. The windows start at a height of h=150 cm, the width of windows b=310 cm. Standing next to the building at a distance of 10 m and a height of 7 m, shading no buildings on the other three sides.

The following telecommunications equipment is located in the building:

- personal computers;

- telephone sets;

- office equipment (printer, copier.).

The layout of the equipment is shown in Figure 4.1.

For other equipment, a separate room is allocated for better electrical safety.

In the personnel room there are 35 employees, according to the state standard requirements per person there should be at least 6,2m^2 of the room area. The room has an area of 352,0 m^2, it means that per employee 10,0 m2 of floor space, which meets the requirements of state standard.

The rated supply voltage of the CC is from a reference voltage source minus 60V or minus 48V with a grounded positive pole

The supply voltage is in the range:

- minus (40 .... 57) V at the rated voltage minus 48 V;
- minus (48 ... 72) V at the rated voltage minus 60 V.

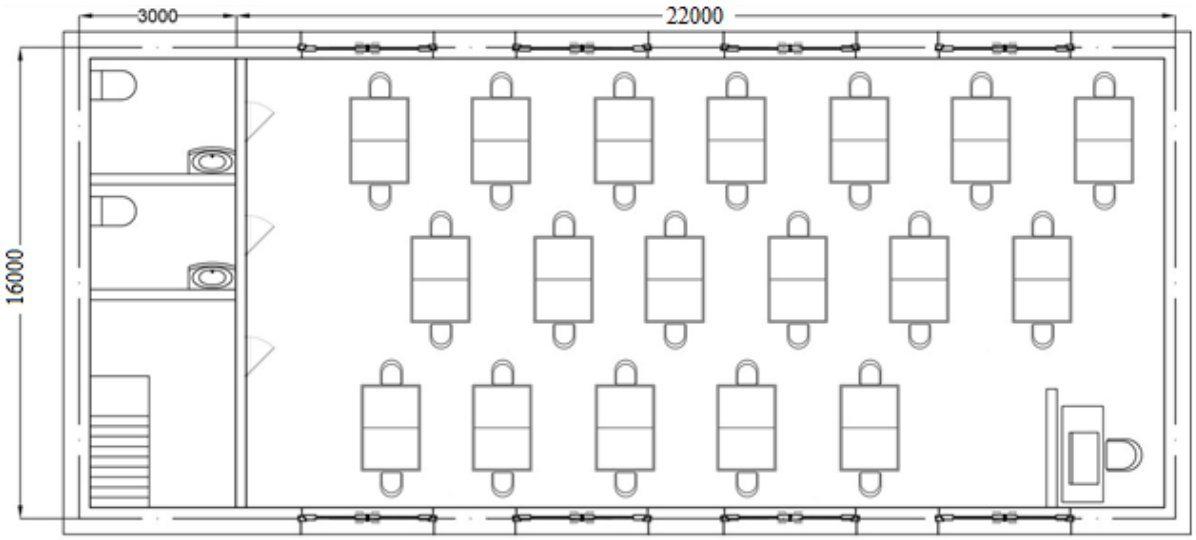

Figure 4.1 - Plan for placing equipment and workplaces in the control room

The supply voltage is in the range:

- minus (40 .... 57) V at the rated voltage minus 48 V;
- minus (48 ... 72) V at the rated voltage minus 60 V.

All components of the CC and power supplies, which are energized above 42 V, are provided with protection against accidental touch of maintenance personnel. In the CC, the ability of electric voltage to fall on the outer metal parts, control buttons, adjustments, adjustments is excluded.

To improve reliability, equipment has enforceability connecting a backup power source, providing the work with the loss of external power.

The uninterruptible power supply is used to protect the equipment from failures in the external network, such as:

- increased or decreased voltage;
- dips or jumps of voltage;
- loss of voltage;
- high noise, distortion, power surges.

If the power supply is uninterrupted, a dead time can be allowed without causing disturbances in the load. For modern technical devices with pulse power supplies this time does not exceed 10-20ms.

In order for the uninterruptible power supply able to supply power to the critical load from the moment of emergency shutdown of the main power source until the backup power source is switched on, the accumulated energy must be ready for immediate use. This energy is in most cases provided by a group of

rechargeable batteries[22].Battery groups require regular maintenance and inspection for corrosion, leakage, and temperature differences between the individual components. Each battery is connected in series to each other by the connecting cable, and each cable is necessary to check the quality of connectivity and the absence of corrosion.

In the battery cabinet, the voltage on the bus voltage is 480V for forty-four-volt batteries and, consequently, 80 terminals that require inspection. This is in addition to the periodic testing of the electrical voltage and internal resistance of storage batteries, and also tests under load.

For uninterrupted power supply, a separate battery room is allocated in the contact center and lead-acid batteries are installed.

Risk factors when working with batteries:

- explosion hazard of storage batteries;
- Acid sulfuric acid, which is a part of the electrolyte of lead-acid batteries;
- High toxicity of lead and sulfuric acid, which are part of most industrial and traction batteries.

According to the degree of exposure to human body, sulfuric acid, according to the sanitary norms, refers to the 2nd class. The maximum permissible concentration (MPC) of sulfuric acid in the air of the working zone according to state standard CH 245-71 is 1 mg / m3, and this concentration is also the limiting one-time. State standard 12.1.005-88 approved the maximum allowable concentration limit for 800 toxic substances [43].

The maximum permissible concentrations of harmful substances in the air of the working area are those concentrations that, within a daily period of 8 hours during the whole working period, can not cause working illnesses or deviations in the state of health detected by modern methods of investigation, directly in the course of work or in the long term.

In accordance with state standard GOST 12.1.007 - 76, hazardous substances are divided into four classes according to the degree of danger: I - extremely dangerous; II - very dangerous; III - dangerous; IV - low-risk. Class I includes substances with a maximum permissible concentration of up to 0.1 mg / m3.

Batteries in power plants and substations, as a rule, should be operated according to the rules of technical operation by the method of constant charging and at the same time the concentration of sulfuric acid in the working area is much lower than the MPC.

Table 4.1 - Maximum allowable concentrations of certain harmful gases, vapors, dust in the air of the work area

Harmful substances	Concentration, $^{mg}/_{m^3}$
Tetraethyl lead	0,005
Vapors or dust of lead, mercury, their inorganic compounds	0,01
Hexachloran, metaphos, ozone	0,1

Table 4.1 continue

Harmful substances	Concentration, $^{mg}/_{m^3}$
Chlorine, sulfuric acid, vapor or copper dust	1
Dust of aluminum, flour dust containing more than 10% of the quartz impurity	2
Tobacco or tea dust	3
Methyl (wood) alcohol, methanol, benzene	5
Dichloroethane, hydrogen sulphide	10
Ammonia, carbon monoxide or sulfur dioxide, naphthalene	20
Xylene, toluene	50
Fuel gasoline	100
Acetone	200
Kerosene, white spirit, transformer oil	300
Ethyl (wine) alcohol	1000

4.2 Calculation of natural light

The premises with the constant presence of people must have, as a rule, natural lighting[1].

It should be noted that the maximum permissible concentration of sulfuric acid in the air of the working area of battery rooms is reached only during the charging of batteries.

When designing new premises, when reconstructing old premises, when projecting the natural lighting of the ship's premises and other objects, it is necessary to determine the area of the light apertures that provide the standardized value of coefficient of natural lightening in accordance with the requirements of Building Norms and Rules of RK 2.04.-05-2002 "Natural lighting and artificial lighting. Design standards "[1] .

Calculation consists in the preliminary determination of the area of light apertures for side and top illumination by the following formulas:

1. At side lighting

$$100\frac{S_0}{S_n} = \frac{e_H \cdot K_{sf} \cdot \eta_0}{\tau_0 \cdot r_2 \cdot k_l} \cdot K_b \qquad (4.1)$$

2. At the top light

$$100\frac{S_0}{S_n} = \frac{e_H \cdot K_{sf} \cdot \eta_l}{\tau_0 \cdot r_2 \cdot k_l} \cdot K_b \qquad (4.2)$$

Where: S_0- the area of light apertures in side lighting, m^2;

S_n-the area of flor, m^2;

e_H-normal value of natural lightening coefficient;

K_{sf}-safety factor;

η_0-Light characteristics of windows;

τ_0-the total light transmittance is determined by the formula:

$$\tau_0 = \tau_1 + \tau_2 + \tau_3 + \tau_4 \qquad (4.3)$$

where τ_1-light transmission coefficient of the material;

τ_2-the coefficient of the light loss in the bindings of the light mist;

τ_3- the coefficient of loss in supporting structures;

τ_4- coefficient for considering the loss coefficient in the sunshield;

τ_5-coefficient of light loss in the protective grid, installed under the lanterns;

r_1-coefficient taking into account the increase in the coefficient of natural illumination at side lightening;

r_1-coefficient taking into account the increase in the coefficient of natural illumination at top lightening

K_b- coefficient that takes into account the eclipse of windows by opposing buildings;

The normalized values of natural light coefficients for a building located in different areas should be determined by the formula:

$$e_N = e_H \cdot m_N \qquad (4.4)$$

Where N-is the natural light group number;

e_H- coefficient of natural lightening;

m_n- coefficient of light climate.

Since the windows are located on the north and south side at the same time, the gap of visual work is V, the coefficient of light climate will be equal to m_n =0,9.

$$e_N = 1 \cdot 0{,}9 = 0{,}9 \qquad (4.5)$$

The main goal of natural lighting is to find the total area of windows for the working space of the operators:

$$S_0 = \frac{S_n \cdot e_H \cdot \eta_0}{100 \cdot \tau_0 \cdot r_1} \cdot K_{sf} \cdot K_b \qquad (4.6)$$

S_0- the area of light apertures in side lighting, m^2;
S_n-the area of flor, m^2;
e_H-normal value of natural lightening coefficient;
K_{sf}-safety factor;
η_0-Light characteristics of wind;
τ_0-the total light transmittance.

Now we calculate the value of the luminous characteristics accordance with the following data:

$$\frac{L}{B} = \frac{22}{16} = 1{,}4$$

$$\frac{H}{h_1} = \frac{3{,}5}{2{,}2} = 1{,}6$$

where h1 is the height above the working surface:

$$h_1 = 0{,}8 + 1{,}4 = 2{,}2$$

Calculated relationships for the window's description of the window we get η_0 = 8.46 and get 8.5. As a light-emitting diode, the project has a two-layer glass, frames made of wood, double-sided window was taken. The sun protecting curtain was taken as a means of protecting against radiation. Corresponding to the table values:

$$\tau_1 = 0{,}8,$$
$$\tau_2 = 0{,}6,$$
$$\tau_3 = 0{,}9,$$
$$\tau_4 = 0{,}1.$$

For the calculation of the light diffraction we will use following formula:

$$\tau_0 = \tau_1 + \tau_2 + \tau_3 + \tau_4 \qquad (4.3)$$
$$\tau_0 = 0{,}8 + 0{,}6 + 0{,}9 + 1$$

To calculate the coefficient, you need to find by the following formula: P/H_b

Here,P is the distance from the point for the one-side lighting. In this case, the height of nearest building is H_b=7m so the relation was taken as 1. If P_ρ=0,5 we will find r_1 coefficient which equals to r_1=2,1. I will find the coefficient K_b of buldings located close to each other. For this condition, the coating layer is 4,5 m and the roof is 2,5 m. In this regard, the height of the building located near:

$$H_b = 4{,}5 + 2{,}5 = 5(m)$$

From the above equation, we find the shadow coefficient of the building facing one another: $F_b = 1$.

Total area of windows:

$$S_0 = \frac{22 \cdot 16 \cdot 0{,}9 \cdot 8{,}5}{100 \cdot 3{,}3 \cdot 2{,}1} \cdot 1{,}2 \cdot 1 = 4{,}66$$

Width of one window:

$$b = \frac{S_0}{h_t} = \frac{4{,}66m^2}{1{,}5m} = 3{,}1m$$

As a result of these calculations, windows with dimensions of 150 × 310 cm^2 will be established.

4.3 Fire safety organization in the communications

The Main Fire Department of the Republic of Kazakhstan through the appropriate departments (departments) of the regions, regions and district inspectorates of fire supervision, as well as fire units and detachments, supervise and monitor compliance with fire safety rules. In violation of fire safety rules, the bodies of the State Fire Safety Administration of the Republic of Kazakhstan have the right to impose fines on the perpetrators persons, to suspend work on the individual sections or the communications enterprise as a whole.

The fire safety rules at the facilities of the Ministry of Communications of the Republic of Kazakhstan were developed with the Ministry of Communications of the Republic of Kazakhstan and coordinated with the Main Fire Department of the Ministry of Internal Affairs of the Republic of Kazakhstan. Responsible for the compliance with fire safety rules at the telecommunications enterprise is its head, for purposes, units, services - their managers appointed by the order for the enterprise.

At all communication enterprises, voluntary fire brigades are set up, headed by the person of administrative personnel, who is appointed by the head of the enterprise. Fire and technical work is organized and conducted by the fire and technical commission, which is the chief engineer of the enterprise. The engineering and technical commission, develops measures for their elimination. At least once a quarter a fire and technical commission inspects communication enterprises.

Measures that eliminate the causes of fires and explosions are subdivided into technical, operational, organizational and regime ones. Technical measures include compliance with fire regulations for buildings, heating and ventilation, selection and installation of electrical equipment, lightning protection. Operational activities underpin the proper operation of production machines, compressor, boiler houses and other power plants and electrical equipment, the correct maintenance of

buildings and theories of the enterprise. Organization activities include the training of production personnel with the opposite rules and the issuance of the necessary instructions and posters. The regime measures include restricting or prohibiting fire, smoking, electric and gas welding works in fire hazardous areas.

In the event of a fire, it is necessary to provide for the evacuation of people. Evacuation routes should cushion the evacuation of all people staying in the premises of the company for the necessary time. Exits are considered evacuation if they lead:

a) from the first floor directly to the outside or through the lobby, corridor, staircase;

b) from the premises of any floor, except the first, into the corridor leading to an internal staircase or flatter cell, which has an outlet directly to the outside or through the vestibule, separated from adjacent corridors by partitions with doors;

c) from a premise in the next premise on the same floor, provided with the outputs specified in subparagraphs a) and b).

Doors on evacuation routes must be opened in the direction of exit from the building. Doors to balconies and platforms intended for evacuation from rooms with a simultaneous stay of no more than 15 people, as well as from clod areas of not more than 200 m^2 and sanitary units may be designed to open inward rooms. Arrangement of a doorway or access door is not allowed. The minimum width of the doors, on the evacuation routes should be 0,8m. The width of the outer doors of staircases should be no less than the width of the ladder march. The height of the crossing on evacuation routes should be at least 2 m. The installation of spiral staircases and runway steps on evacuation routes is not allowed. Between the flights of stairs, it is necessary to provide a horizontal gap of at least 50 mm.

To ensure the safe evacuation of people from premises and buildings, the estimated evacuation time t_{ee} should be less than the required time for evacuation t_{ev}, that is,$t_{ee} \leq t_{ev}$. The necessary time for evacuation of people from rooms up to 6 m in height of industrial buildings I,II, and III degrees of fire resistance is taken in accordance with Table 4.2. At a room height of more than 6 m, the necessary evacuation time is increased: up to 12 m-by 20, up to 18 m-by 30 and up to 24 m or more by 40%.

Table 4.2 Necessary time for evacuation of people

Production category	Room volume, thousand. m^3				
	up to 15	30	40	50	60 and more
A,B	0,5	0,75	1	1,5	1,75
C	1,25	2	2	2,5	3

Estimated time for evacuation of people from the premises and back is established by calculating the time of the diving of one or several human flows through evacuation exits from the most important places of people's

accommodation. When calculating the path of the flow of human flow is divided into sections (a passage, a corridor, a doorway, a flight of stairs, a vestibule) of length L_i and width δ_i. The starting areas are the passages between workplaces, rows of chairs and the like. When determining the estimated time, the length and width of each section of the evacuation path are also measured along the ramps along the ramps. The length of the path in the doorway is assumed to be zero. For example, located in a wall with a thickness of more than 0.7 m, and also the vestibule should be considered independent sections of the horizontal path, having a finite length.

Table 4.3- The speed and intensity of the flow of human flow

The density of people flow in one minute	Horizontal path		Doorway
	Intensity q, m/min.	Speed v, m/min.	Intensity q, m/min.
0,5	5	100	5
9	8	80	8,7
1	12	60	13,4
2	14,1	47	16,5
3	16	40	18,4
4	16,5	33	19,6
5	16,2	27	19
6	16,1	23	18,5
7	15,2	19	17,3
8	13,5	15	8,5
9 and more			

Estimated time of evacuation of people should be defined as the sum of the time of movement of the human flow along separate sections of the path, i.e.

$$t_{ee} = t_1 + t_2 + ... + t_i + ... + t_k = \sum_{i=1}^{i=k} t_i \qquad (4.6)$$

The time of movement of the human flow along the first section of the path

$$t_1 = \frac{L_1}{v_1} \qquad (4.7)$$

where A and B are the length of the path and the slowness of the flow of the human flow at the site is determined from Table 4.3, depending on the human flow.

$$D_1 = \frac{N_1}{L_1 \cdot \delta 1} \qquad (4.8)$$

Where N_1-number of people in the first section of the evacuation route.

Our room specifics of the process performed, the properties of the substances and materials use, and the availability of electronic equipment, the premises are classified as D fire hazard category. To determine the smallest permissible width of the doorway to the staircase and the largest number of workers evacuated from the production premises, up to 1232 m^3 in one minute.

The greatest distance to the exit L_1=26m, transition width δ_1=1,5m. The necessary speed of movement of people $v_1=L_1/t$=26m/1minute=16m/minute. Which equals corresponding to the table 4.3 D_1=10pepole/minute.

$$N_1 = D_1 \cdot L_1 \cdot \delta_1 = 10 \cdot 26 \cdot 1{,}5 = 390$$

Consequently, 390 people can be evacuated throughout the passage

Conclusion:In this section, an analysis of working conditions in the working room for operator staff. Working conditions are revealed by normal, and the results of calculations completely meet the requirements of standards safety and vital activity.

The premises with the constant presence of people must have, as a rule, natural lighting. In calculation of full rooms natural lightening we determined the permissible size of windows for working room of operator staff.

Electrical equipment necessary for the work staff is a possible source of ignition which leads to fire. It was necessary to determine how many people can be evacuated? So, we find 390 people can be evacuated throughout the passage.

5 Business plan

5.1 Objective

The purpose of the study of Big data based on data collected in contact centers, using known methods for data collection, such as the method of reference vectors, linear neural networks, classification trees and regression, to predict the quality of service in contact centers.

Now, when many companies have accumulated enough large data sets, the question arises about their processing. The term Big Data has been around for a long time, but at the same time the Kazakh business is only coming to understand the usefulness of this analytics. It can be applied in various business areas, and, of course, financial organizations - the most likely consumer of such information. Moreover, the pool of tasks that can be solved on the basis of Big Data is almost unlimited.

The growing volume of information and the simultaneous reduction in the cost of computing and data storage generated the notion of large data. It's not just about large data sets, but also about data that is processed at high speed or has a

high variability. Your business processes are served by an information system, for example CRM.

Or you have a website, an application for smartphones, or via e-mail, instant messengers, a phone to your support service, there are appeals from customers. This means that you already have data, which, having accumulated in sufficient volume, can be analyzed.

Methods for analyzing large data are based on the construction of automated analytical models - machine learning. With its help, algorithms are formed that find hidden relationships, train, help to make forecasts, optimize various business processes. Thus, it is possible to work out a much larger number of hypotheses than a person can do.

In this regard, it is necessary to study economically the work of the effective contact-centers, where a wide range of data sources are collected and processed.

5.2 Company and industry

As competition increases, companies have less room to maneuver in the area of lowering prices for products and services. For this reason, companies pay more attention to quality of service to their customers, both existing and potential. An important part of the service process is the responses to requests by mail or via the Internet. Service calls are the «face of the company», because with the quality of service calls customer can draw conclusions about how correctly the company will serve his requests in the future of joint work.

Thus, companies are forced to look for solutions that allow them to efficiently organize customer service. The creation of the contact center in such cases is one of the most popular and productive ways, improving the quality of service.

Contact Center is the next level of technology for high-quality and effective customer service. The contact center makes the best use of the company's available resources (company employees, communication lines, equipment and software) in order to service customers' calls. Unlike the Contact center of the previous generation of contact centers, the modern contact center, in addition to serving traditional telephone calls, provides the same quality of service and Internet calls, such as e-mails, sms messages, web calls and more.

Evaluation of the effectiveness of the investment in the center is made due to the presence of a large number of statistical information about the operator's work, the group of directors, the load of the connecting lines, the contact center as a whole. Such statistics allow you to draw conclusions about how much customer service improves when new contact center implementations are introduced, an increase in the number of operators, etc. It should be noted that the contact center is a living organism, changing and evolving along with the development of business company. At the same time, assessing the effectiveness of the contact center and the return on investment in it is the most important tool.

5.3 Description of services and products

Big Data technology in the telecommunications industry can solve several types of tasks with that is the following positions:

- storage and management;
- analysis of unstructured information;
- effective processing of a wide range of sources.

5.4 Market analysis. Investigation of market services

The field of use of Big Data technologies is extensive. So, with the help of Big Data, you can learn about the preferences of customers, the effectiveness of marketing campaigns or conduct a risk analysis. Below are the results of the survey of the IBM Institute, on the directions of using Big Data in companies [8].

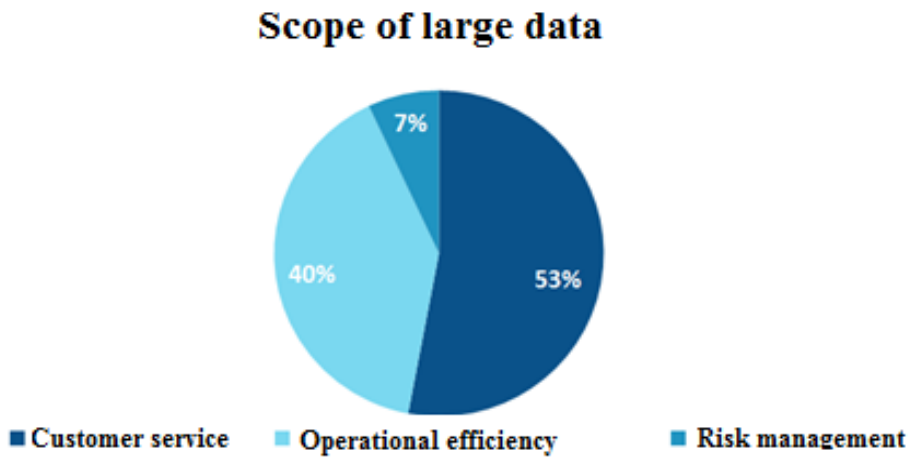

Figure 5.1 – Scope of large data

As can be seen from the diagram, most companies use the Big Data in the customer service area, the second most popular direction is operational efficiency, in the field of risk management, Big Data is less common at the moment.

It should also be noted that Big Data is one of the fastest growing areas of information technology, according to statistics, the total amount of data received and stored doubles every 1.2 years.

Big Data has become widespread in many business sectors. They are used in healthcare, telecommunications, trade, logistics, financial companies, as well as in public administration.

Retail.In the databases of retail stores, a lot of information about customers, the inventory management system, and the supply of marketable products can be accumulated. This information can be useful in all areas of the shops. So, with the help of the accumulated information it is possible to manage the supply of goods, its storage and sale. Based on the accumulated information, you can predict the demand and supply of goods. Also, the data processing and analysis system can solve other problems of the retailer, for example, optimize costs or prepare reports.

Financial services.Large data gives an opportunity to analyze the borrower's creditworthiness, they are also useful for credit scoring and underwriting. The introduction of Big Data technologies will reduce the time for consideration of loan

applications. With the help of Big Data, you can analyze the operations of a specific customer and offer suitable banking services.

Telecommunication industry.In the telecommunications industry, Big Data is widely used by cellular operators.Cellular operators on a par with financial institutions have one of the most extensive databases, which allows them to conduct the most profound analysis of the accumulated information.

The main purpose of data analysis is to retain existing customers and attract new ones. To do this, companies conduct customer segmentation, analyze their traffic, determine the subscriber's social identity.

In addition to using Big Data for marketing purposes, technologies are used to prevent fraudulent financial transactions.

5.5 Financial plan

5.3.1 Calculation of investment costs. Investment costs include the cost of equipment, installation work and transportation services. The amount of capital expenditure is determined. For this purpose, estimates are made for the preparation of equipment. Also there are considered the costs of building buildings, structures and etc.

Table 5.1– Equipment costs

Nomination	The amount of consumed material	Price per unit(tenge)	Total amount(tenge)
Computers for operators	30	180000	5400000
PBX computers	3	150000	450000
Modem	5	8000	40000
Switch	4	10000	40000
Total:			5930000

Total capital investments:

$$\sum K = K_0 + K_c + K_T + K_D \quad (1)$$

Where: K_0 –capital expenditure for the purchase of equipment;
K_C – capital investments for construction;
K_I – capital investments for installation work;
K_D –capital investments for design
K_T– capital expenditure on transport costs.

$K_I=K_0\cdot 0,08=5930000\cdot 0,08=474400$ (tenge),
$K_T=K_0\cdot 0,02=5930000\cdot 0,02=118600$ (tenge),
$K_D=K_0\cdot 0,05=5930000\cdot 0,05=296500$ (tenge),

So overall capital embedding:

$$\sum K=5930000+474400+118600+296500=6819500(\text{tenge})$$

5.3.2 Calculation of income. Revenues from the main activity - income received by businesspeople for the entire volume of communication services sold by the consumer at current tariffs.

$$D_0 = \sum_{i=1}^{n} q_i \cdot U_i \quad (2)$$

Where q_i−The volume of the i-th type of services in natural expressions;

U_i −tariff for the i-th type of services, in tenge;

n−nomenclature of services.

The first main source of income is quality provision of services. On average, each subscriber (private and corporate) will spend about 1200 tenge per month, the control and analysis center receives 5% of the total amount. Then a month to the control and analysis center will receive 60 tenge from each subscriber. Consequently, average income per day from each subscriber will be:

$$D_0 = \sum_{i=1}^{1} 60 \cdot 6578 = 394680\ (tenge)$$

The income of contact centers per month for 8 hours of work is equal to:

$$D = D_0 \cdot 30 = 394680 \cdot 30 = 11840400 \quad (3)$$

5.3.2 Operating costs. In the process of servicing and provision of services for communication, activities are carried out that require the costs of the enterprise. The amount of costs for the year will be the actual production cost or the amount of annual operating expenses.

$$C=FoP+S_t+M+E+D+C+O, \quad (4)$$

Where FoP−fund of payment (main and additional wages);

S_t−social tax (13% from FoP);

M−material costs and spare parts (expenses for spare parts and current repairs are 0.5% of capital investments);

E−electricity for production needs;

D−depreciation deductions;

C−credits (bank charges);

O−overhead costs (indirect costs here can be attributed all the unaccounted expenses, administrative, economic, cost of staff training, transportation costs) Conventionally, it is 75% of the cost.

We calculate the payment for each day's work of each employee. That is, the size of the monthly fee for the number of working days per month ,we define the relation (8 hours to six days of the working week).

For a manager:

$$D = \frac{250000}{24} = 10417 \text{ (tenge/day)}, \quad (5)$$

For an engineer:

$$D = \frac{210000}{24} = 8750 \text{ (tenge/day)},$$

For operators:

$$D = \frac{85000}{24} = 3452 \text{ (tenge/day)}.$$

Table 5.2– Employees' salary

Work executors	Quantity	Salary per hour of work, tenge	Salary per day of work, tenge	Salary per month of work, tenge
Manager	2	1302	10417	250000
Engeneer	2	1094	8750	210000
Operators	30	443	3542	85000
Total:	3470000			

An hourly payroll is calculated by dividing the employee's daily fee by the number of working hours (during 8 hours):

For a manager:

$$D = \frac{250000}{24 \cdot 8} = 1302 \text{ (tenge/hour)} \quad (6)$$

For an engineer:

$$D = \frac{210000}{24 \cdot 8} = 1094 \text{ (tenge/hour)},$$

For operators:

$$D = \frac{85000}{24 \cdot 8} = 443 \text{ (tenge/hour)}$$

Remuneration fund is collected from employees' salaries:

$$T_{main} = 3470000$$

Additional employees' remuneration is 10% of the principal payment:

$$T_{add.} = T_{main} \cdot 10\% \quad (7)$$

$$T_{add.} = 3470000 \cdot 10\% = 347000 \text{(tenge)},$$

The fund of payment(FoP) consists of the sum of additional and main wages:

$$FoP = T_{main} + T_{add} = 3470000 + 347000 = 3817000 \text{(tenge)} \quad (8)$$

Social tax is calculated based on the employee's annual income statement. We get 11% of the revenue to calculate it. For the transfer of social tax pension O_{st} (11%), the budget consists of a pension $O_{pension}$(10%):

$$O_{pension} = FoP \cdot 10\% \quad (10)$$

$$O_{pension}=3817000 \cdot 10\%=381700.$$

Social tax:

$$O_{st}= (FoP - MinWage - O_{pension}) \cdot 11\% \quad (11)$$

The Minimum Wage=28 284 tenge according to the data for Kazakhstan 2018

$$O_{st}=(3817000\text{-}28\ 284\text{-}381700) \cdot 11\%=374771(\text{tenge}).$$

Depreciation deductions:

$$A = \frac{N_{Am} \cdot C_i \cdot N}{100 \cdot 12 \cdot n} \quad (12)$$

Where N_{Am}−depreciation rate,
C_i−initial cost of equipment,
N−number of days in which the work is to be performed,
n−The number of days in a month's work.

$$A_1(computer) = \frac{23 \cdot 180000 \cdot 30}{100 \cdot 12 \cdot 30} = 3750 \text{ (tenge)},$$

$$A_1(\text{mod } em) = \frac{10 \cdot 8000 \cdot 30}{100 \cdot 12 \cdot 30} = 70 \text{ (tenge)},$$

$$A_1(switch) = \frac{10 \cdot 10000 \cdot 30}{100 \cdot 12 \cdot 30} = 85 \text{ (tenge)},$$

$$A_1(PBXcomputers) = \frac{23 \cdot 150000 \cdot 30}{100 \cdot 12 \cdot 30} = 3125 \text{ (tenge)},$$

Table 5.3– Depreciation of fixed assets

Name of equipment	The amount of consumed material	Annual depreciation rate	Price per unit(tenge)	Total :
Computers for operators	30	25%	3750	112500

Table 5.3 Continuation

Name of equipment	The amount of consumed material	Annual depreciation rate	Price per unit(tenge)	Total :
PBXcomputers	3	25%	3125	9375
Modem	5	10%	70	350
Switch	4	10%	85	340
Total depreciation of fixed assets				122565

It is necessary to find out the cost of electric energy, because electric equipment is used in operators' workplaces. Electric power consumption:

$$E = W \cdot T \cdot S \cdot K_i \quad (13)$$

Where W–power consumption, kW;
T–number of operating hours;
S–cost of kilowatt per hour of electricity, (25tg/ kW·h);
K_i–power factor of the nth electrical equipment (adopted that K_i=0.7,0.9).
Electricity consumption is shown in Table 5.4.
There are given necessary and used equipment for the project implementation.

Table 5.4. Electricity consumption

Nomination	Quantity	Passport power, kW	Number of working days	Power factor, Ki	Operating time of equipment	∑ W, kW/h
computers	33	0,5	30	0,9	720	10692
modem	5	0,01	30	0,9	720	32,4
swich	4	0,01	30	0,9	720	25,92
Total :						10750,32

Electricity costs:

$$E = \sum W \cdot S = 10750 \cdot 25 = 268750 \text{ (tenge)} \quad (14)$$

Overhead costs account for 25% of total expenditure:

$$\sum O = (FoP + O_{st} + D + E + C) \cdot 25\% \quad (15)$$

∑ O=(3817000+374771+122565+268750+10000)·25%=1148271(tenge),

C=3817000+374771+122565+268750+10000+1148271=5472607(tenge).

The cost estimates for all listed items are given in Table 5.5. Below is a table of cost estimates for common project costs.

Table 5.4. Estimates for all listed costs

Item of expenditure	Total cost, tenge	Cost structure, %
FoP(Fund of Payment)	3817000	66,6
Social tax	374771	6,5
Depreciation	122565	2,2
Item of expenditure	Total cost, tenge	Cost structure, %
Electricity for production	268750	4,7
Overhead costs	1148271	20,0
Total:	5731357	100

5.3.2 Calculating of economic efficiency. In the final works on the development, expansion and reconstruction of communication enterprises the following indicators of economic efficiency are calculated.

The coefficient of total (absolute) economic efficiency of capital investments, when constructing a new facility or enterprise.

$$E_a=(D-C-H)/K=P_n/K \quad (16)$$

Where D–earnings from key business activities,
C–operating costs,
P –total profit,
P_n–net profit,
H – tax rate (20% of profit).

$$P=D-C=11840400-5731357=6109043(\text{tenge}),$$

$$P_n=P-H=6109043-1221808=4887234,$$

$$E_a=4887234/6819500=0,71.$$

The return on capital investment is reflected in the overall effectiveness ratio:

$$T=1/E_a \quad (17)$$

$$T=\frac{1}{0,71}=1,4 \text{ (year)}$$

Conclusion: The main advantage of big data for statistical research is timely receipt of bulk amounts of information with minimal financial and time costs. Big data is able to provide a wide range of information on various aspects, is not

affected by traditional sources. In addition, the high frequency of obtaining information compared to a routine survey provides a detailed consideration of the processes and the solution of problems at the stage of their origin.

Big Data opens new opportunities for programming to build value and build close relationships with business units, allowing you to increase revenues and strengthen the company's financial position. Big Data projects make IT departments a strategic partner for business units.

Summarizing all the results: it will be profitable to introduce in the field of small and medium-sized businesses the contact center in which Big data technologies are used, from the point of view of economically newcomers to the market. Necessary capital investment amounted to 6819500 tenge. It costs 3-4 times cheaper than other similar contact centers. Economic efficiency ratio is 0.71 , which means that project repayment term and other turnovers is about 16 months. This is a very high index. Thus, large amounts of data processing and storage in contact centers are economically and socially profitable.

Conclusion

In the first chapter, the current state of contact centers was investigated. Purposes and application of modern contact center were considered, also evolution of its technology was reviewed The first part shows the contact center architecture and methods of call distribution.

In the second chapter the definition of the term "Big Data" was given, the relevance of this technology in various professional spheres was examined, and also their structure was studied. Therefore, we found the answer for question on “how large the data is and who can take benefit from it?”.

In the third chapter, the possibility of applying the existing prediction of one-dimensional time series for investigating incoming contact center traffic is considered. The methods of trend extrapolation and moving average are analyzed. The results of traffic forecasting for a particular contact center are presented. The estimation of stationarity of the traffic was made, allowing to reveal the trend in traffic statistics during the month. The forecasted traffic estimates are calculated using different methods with a forecast horizon of 1-3 days. The forecast for a particular day of the week is determined, which, as a rule, has the greatest load

The results of short-term forecasting of the incoming traffic of the contact center made it possible to reveal that the greatest accuracy is provided by the forecast calculated by the method of extraneous linear trend. The proposed forecasting technique can be used in similar contact center operators to provide quality customer service.

In the fourth chapter examination of working conditions in an individual room made it possible to identify and eliminate all the dangerous and harmful factors that operators are subjected to on a daily basis in the course of their labor activity.

The economic effectiveness of the organization of the contact center confirmed by the calculations were carried out in fifth chapter. The payback period of the project does not exceed year and half.

Abbreviated terms

IT- Information Technology
IP – Internet Protocol
FoP-Fund of Payment
KCCA- Kazakhstan Contact Centers Association
DoC- Distribution of Calls
CTI- Computer Telephony Integration
IVR- Interactive Voice Response
VoIP- Voice over Internet Protocol
CRM- *Customer Relationship Management*
MPC- Maximum Permissible Concentration
SoCD- system of call distribution
PCL -public communication line

List of references

1. Гольдштейн В.С., Фрейнкман В.А. Call-центры и компьютерная телефония, СПб.:БХВ, 2002.
2. Гольдштейн Б.С., Исаев В.И., Мамонтова Н.П., Фрейнкман В.А. Анализ, синтез и управление качеством функционирования центров обслуживания вызовов: учебное пособие/ГОУВПОСПбГУТ.-СПб, 2006.
3. Туманбаева К.Х., Лещинская Э.М. , Демесенова К. Повышение эффективности функционирования Call-центра//Труды международной научно - образование, наука практика»», Алматы, КазНТУ, 2012
4. Булавин Д., Чикунов А., Будушкаева С. Аналитический обзор рынка Больших данных. – М.: Борд, 2015. – 51 с.
5. Kuan-Ching Li, Hai Jiang, Laurence T.Yang, Cuzzocrea A. Big Data: algorithms, analytics and applications. – CRC Press. – 2015. – 439 с
6. A. Gilmore, Call Center Management: Is Service Quality a Priority, Managing Service Quality, vol. 11, no. 3 pp. 153-159.
7. A. Parasuraman, Service Quality and Productivity: a Synergistic Perspective, Managing Service Quality, vol. 12, no. 1, pp. 6-9.
8.Davy Cielen, Arno D.B.Meysman, Mohamed Ali. Introducing Data Science. Big data, machine learning, and more, using Python tools–Copyright 2016 Manning Publications, 2016.
9. Золотарев С. Три кита больших данных //ИКС. – 2014. – №1-2. – С. 34.
10. Бабенко Е.И., Кудрявцева Е.П. Система показателей оценки качества телекоммуникационных услуг //Качество, инновации, образование. – 2009. №12. – С. 27.
11. Туманбаева К.Х., Лещинская Э.М. Анализ и прогнозирование входящего трафика call-центра //«Қоғамды ақпараттандыру» III Халықаралық ғылыми-практикалық конференция. – 2012.
12. Туманбаева К.Х., Лещинская Э.М. Моделирование процесса функционирования call-центра //Тезисы докладов 8-ой Международной научн-техн. конференции «Энергетика, телекоммуникации и высшее образование», Алматы. – 2012.
13. Киселев М.Н. IP Контакт Центр – объединённая система сервиса потребительов. – Москва, 2008 г. -128 с.
14. Алексеев М. Big Data – революция в области хранения и обработки данных. [Электронный ресурс]/ М., 2014. URL: http://www.slideshare.net (қолдану уақыты: 21.02.2016).
15. Технологии «больших данных» в важнейших отраслях. CITO Research
16. Барсков А.Г. Операторские центры: классика плюс Интернет-новации // Сети и системы связи. 2001. - №6. - 273 с.
17. Тарасов В.Ю. Корпоративный контакт-центр // Сети и системы связи. - 2003. №10. – 134 с.
18. Учебное пособие по GPSS World. – Казань: Изд-во «Мастер Лайн», 2002

19. Туманбаева К.Х. Моделирование систем телекоммуникаций. Учебное пособие. – Алматы: АИЭС, 2007
20. Т.Л.Саати. Элементы теории массового обслуживания и ее приложения: Пер. с англ. /Под. ред. И.Н. Коваленко, изд-ие 2. М., 1971.
21. Шелухин О.И., Тенякшев А.М., Осин А.В. Моделирование информационных систем.-М.: Радиотехника, 2005.
22. Лещинская Э.М., Ишанова С.Б. Краткосрочное прогнозирование входящего трафика информационно-технического центра //Известия научно-технического общества «КАХАК», Алматы. – 2013. – №2(41).
23. Самгин Э.Б. Освещение рабочих мест. – М.: МИРЭА, 1989. – 186 с.
24. Юдин Е.Я., Борисов Л.А. Борьба с шумом на производстве: Справочник. Под общ. Ред. Юдина Е.Я.
25 Феоктистова Т.Г. Безопасность жизнедеятельности. Производственная санитария и гигиена труда. – М., 2013. – 62 с
26 Голубицкая Е.А. Экономика связи. – М.: ИРМАС, 2006
27 Базылов Қ.Б., Алибаева С.А., Нурмагамбетова С.С. Бітіруші жұмысының экономикалық бөлімі үшін әдістемелік нұсқаулар. 050719 – Радиотехника, электроника және телекоммуникация мамандығының барлық оқе түрінің студенттеріне арналған. – Алматы: АИЭС, 2009. – 9 б.
28 Буров В.П., Новиков О.К. Бизнес-план: методика составления. – М.: ЦИПКК, 1995. – 54 с.