**My Future Profession**

Hello. My name is Nikita. I am a 3rd year student of Software Engineering at Bauman Moscow State Technical University.

I would like to become a big data analyst.

This profession involves analyzing large datasets to uncover patterns, trends, and insights that can help organizations make informed decisions.

Big Data Analysts play a crucial role in interpreting complex data and presenting it in a meaningful way.

In this role, I will be responsible for collecting and processing data from various sources.

I will also clean and organize this data to ensure it is accurate and usable.

By applying statistical and analytical techniques, I will identify significant trends and correlations within the data.

One of the key tasks of a Big Data Analyst is to create reports and visualizations that clearly communicate findings to stakeholders.

This helps businesses understand their data and use it to drive strategy and innovation.

Additionally, I will develop models to predict future trends and outcomes based on historical data.

Working as a Big Data Analyst requires strong problem-solving skills and attention to detail.

I will need to be curious and constantly willing to learn, as the field is always evolving.

Effective communication is also important, as I will often need to explain complex data concepts to non-technical team members.

I am excited about the potential to make a significant impact through data analysis.

By turning raw data into valuable insights, I will help organizations improve their operations and achieve their goals.

The combination of technical challenges and the opportunity to influence business decisions makes this career particularly appealing to me.

In the future, I aim to become proficient in this field and contribute to the success of the organizations I work for.

The role of a Big Data Analyst is both challenging and rewarding, and I am eager to begin my journey in this exciting profession.

**Computer Architecture**

A computer is an electronic machine that accepts data, processes it, and provides results as information in a specified format.

Information can include texts, tables, drawings, photos, sound recordings, and more, stored and processed digitally in bytes, kilobytes (KB), megabytes (MB), and gigabytes (GB).

Data is entered into the computer's memory, processed by a program, and results are displayed on the screen or printed.

New computers require software to function, with the operating system acting as an intermediary between application software and hardware.

The three most common operating systems are Microsoft Windows, macOS, and Linux.

A computer system consists of software and hardware. Hardware includes the monitor, CPU, circuits, displays, power supplies, cables, keyboards, printers, and mice.

Software, which lacks a physical form, includes programs, data, and protocols, such as web browsers, games, and word processors.

The three basic hardware sections are the central processing unit (CPU), main memory, and peripherals.

The CPU, or microprocessor, is the computer's "heart," located on the motherboard, and controls processing speed.

The main memory (RAM) holds instructions and data for the CPU.

Peripherals, attached to the computer, provide input or output, including storage devices (hard drives, DVD drives, flash drives) and input/output devices (keyboard, mouse, display, printer).

Peripherals also include hard disk drives, floppy disk drives, and optical disk drives, serving as both input and output devices.

Various ports on the computer allow connection of devices like modems, cameras, and scanners, facilitating communication.

Modern PCs have USB ports and memory card readers on the front panel.

**Operating system**

When a new computer comes off the factory assembly line, it can do nothing: hardware needs software to make it work.

An operating system is a set of programs that lies between applications software and the computer hardware.

The most important program in the operating system, the program that manages the operating system, is the supervisor program, most of which remains (остается) in memory, this program is named resident.

The supervisor controls the entire operating system and loads into memory other operating system programs (called non-resident) from disk storage only as needed (по мере необходимости).

An operating system has **three main functions**:

(1) manage the computer's resources, such as the central processing unit, memory, disk drives, and printers

(2) establish (создание) a user interface

(3) execute and provide services for applications software.

But much of the work of an operating system is hidden from the user (for example, managing the computer's resources).

Operating systems usually come **pre-loaded** on any computer you buy.

Most people use the operating system that comes with their computer, but it's possible to upgrade or even change operating systems.

The three most common operating systems for personal computers are **Microsoft Windows**,**macOS**, and **Linux**.

Modern operating systems use a **graphical user interface** (**GUI)**.

A GUI lets you use your mouse to click **icons**, **buttons**, and **menus**, and everything is clearly displayed on the screen using a combination of **graphics** and **text**.

Microsoft created the **Windows** operating system in the mid-1980s.

There have been many different versions of Windows, but the most recent one is **Windows 10** (released in 2015).

Windows comes **pre-loaded** on most new PCs, which helps to make it the **most popular operating system** in the world.

**MacOS** (previously called **OS X**) is a line of operating systems created by Apple.

It comes preloaded on all Macintosh computers, or Macs.

Only 10% of people use macOS, because  Apple computers tend to be more expensive.

**Linux** is a family of **open-source** operating systems, which means they can be modified and distributed by anyone around the world.

This is different from **software** like Windows, which can only be modified by the company that owns it.

The advantages of Linux are that it is **free**, and there are many different **distributions** — or versions —you can choose from.

**Mobile devices** such as **phones**, **tablet computers**, and **MP3 players** are different from desktop and laptop computers, so they run operating systems that are designed specifically for mobile devices.

Examples of mobile operating systems include **Apple iOS** and **Google Android.**

**NETWORK**

Network is a system of computers and other devices, such as printers, that are linked together.

A network can consist of two computers connected with cables or millions of computers that are spread over a large geographical area and connected by telephone lines, fiber-optic [файбер] cables (оптоволоконными кабелями), or radio waves (радиоволнами).

The Internet is an example of very large network.

The Internet is the global system of interconnected computer networks that use the Internet protocol suite (набор протоколов) to (для) link devices worldwide.

It is a network of networks that consists of private, public, academic, business (деловых) and government networks of local to global scope, linked by electronic, wireless, and optical networking (сетевых) technologies.

Compared to printed media (сми), books, encyclopedias and traditional libraries, the World Wide Web has enabled the decentralization of information on a large scale.

A website is a collection (набор) of related web pages, including multimedia content, typically (обычно) identified with a common domain name, and published on at least one web server.

Websites can have many functions and can be used in various fashions (по-разному); a website can be a personal website, a corporate website for a company, a government website, an organizational website, etc.

They are typically dedicated to a particular topic or purpose, ranging (начиная) from entertainment and social networking to providing news and education.

A web browser is a software (программное) application.

The primary (основная) purpose of a web browser is to bring information resources to the user, allowing them to view the information and then access other information.

The most popular web browsers are Chrome, Internet Explorer, Edge, Safari, Opera and Firefox.

Modern technologies are not standing still (не стоят на месте) and are constantly evolving (развиваются).

Speaking of networks, we cannot fail to mention some wireless means (средства) of communications using the type of connection of the last generations, such as *3G and 4G*.

Each generation is faster, more secure and more reliable.

3G set the standards for most of the wireless technology we have come to know and love: web browsing (просмотр веб-стр), email, video downloading, picture sharing and other technologies.

With 4G network users get the advantage of better uninterrupted connectivity (соединения), especially for advanced tasks such as video chats and conferences.

They can stream music, video and movies at a much faster rate (скорость) than ever before and can also easily share information online.

Rather than (вместо) installing software on each machine or server within your organization, you rent applications from the ASP (application service provider), which provides remote access to the software.

The 21st century software trend is the following: lease instead of buy (давайте в аренду вместо покупки)!

**PROGRAMMING LANGUAGES**

Computers can deal with different kinds of problems if they are given the right instructions for what to do.

Instructions are first written in one of the high-level languages, depending on the type of problem to be solved.

A program written in one of these languages is often called a source program (исходная программа).

The source code (исходный код) of a program has to go through some kind of translation into machine code and the machine can actually (фактически) run.

An applications program (прикладная программа) is a program written in one of these high-level languages is designed (предназначена) to do a specific type of work.

Software packages are a set of programs designed to perform (выполнения) certain applications which conform (соответствуют) to the particular specifications (требованиям) of the user.

Institutions (учреждения) either purchase these programs as packages (в виде пакетов) or commission (поручают) their own programmers to write them to meet the specifications of the users.

=============

There are some kinds of programming languages:

## 1. Compiled languages

Compiled languages get translated into runnable files of binary machine code by a special program called a *compiler*.

Once the binary has been generated, you can run it directly without looking at the source code again.

Compiled languages tend to give excellent performance (производительность) and have the most complete access to the OS, but also to be difficult to program in.

Compiled languages are: C and C++, the languages in which Unix itself is written; FORTRAN, but it is much more primitive and COBOL, which is widely used for financial and business software.

**2. Interpreted languages**

An *interpreted language* depends on an interpreter program that reads the source code and translates it on the fly into computations (вычисления) and system calls.

The source has to be re-interpreted each time the code is executed (выполняется).

Interpreted languages tend to be slower than compiled languages, but they tend to be easier to program.

Interrupted languages are: Basic and Lisp.

## 3. P-code languages (pseudo code)

P-code languages are like compiled languages in that the source is translated to a compact binary form, but that form is not machine code, it's *pseudo code* (or *p-code*), which is usually a lot simpler but more powerful than a real machine language.

In the P-code, programming speed is still slower than compiled but is faster than interpreted.

Important p-code languages include Python, Perl, and Java.

**My favorite programming language is python**. Here's the reasons why I absolutely love Python:

1. Clean Syntax. There are no semicolons, braces to remember, whitespaces and tabs automatically do it for you,
2. Awesome Libraries. Whatever project you are doing, whatever module you need, Python has all the libraries you need.
3. Works on any computer. If you want to start learning Python, you don’t need to worry about what computer and operating system you use (Mac, Windows, or Linux).

Also, this means that you can write and use Python code across different systems without making any changes to the code.

1. Versatility (универсальность)

You can use Python for a variety of different projects, such as:

* Building web applications
* Conducting data analysis
* Machine learning
* Building games
* Visualizing data

So, one of the most significant advantages of Python is its versatility and flexibility.

All in all, Python is a robust (надежный) programming language that allows you to manage your coding project more efficiently, and it’s easy to debug, too.

**Databases**

A database is a collection of information that is organized so that it can easily be accessed, managed and updated.

In one view, databases can be classified according to types of content: **bibliographic, full-text, numeric** and **images.**

In computing, databases are sometimes classified according to their organizational approach (подход).

* The most pr’evalent (распространенный) approach is the **relational database**, a tabular database in which data is defined so that it can be reorganized and accessed by different ways.
* A **distributed** (распределенная) database is one that can be dispersed (рассредоточена) or replicated (реплицирована/копирована) among different points in a network.
* An **object-oriented** programming database is one that is c’ongruent (соответствует) with the data defined in object classes and subclasses.
* **NoSQL** database is used for large sets of distributed data. This type of computers database is very efficient in analyzing large-size unstructured data.
* In a **document-oriented** database, the data is kept in document collections, usually using the XML, JSON, BSON formats.

One record can store as much data as you want, in any data type (or types) you prefer.

The data in a database should have the following features:

* The data is organized and related.
* The data in database can be shared among different users and applications.
* The data in database is constant.
* The data is secured.
* The data is easily accessible and available to users.

Computer databases typically contain aggregations (совокупность) of data records or files, such as sales transactions (транзакции продаж), product c’atalogs and customer (клиент) prof[i]les.

Also databases are used to hold patients` details so they can be accessed from anywhere within a hospital or hospital network.

With the recent improvements in image compression techniques (методы сжатия изобр-й), X-rays (рентгеновский снимок) and scan output can also be held in databases and accessed in the same way.

Information is entered into the database via fields.

Each field holds a separate piece of information, and the fields are grouped together in records.

Therefore, a record about an employee might consist of several fields which give their name, address, phone number, date of birth, salary and other information.

Multi-user databases are managed by a piece of software called a database management system (DBMS).

DBMS is a software or application that enable the user to interact with the database, thus allowing users to insert, retrieve, update and delete data.

The DBMS provides a view (представление) of the data that is elevated (находятся) above the hardware level (на уровне выше аппаратного), and supports user-requests written in a higher-level language.

The DBMS also determines (определяет) the amount and type of information that each user can access from a database.

When a user wants to access a database, he makes an access request using a particular data-manipulation (язык обработки данных) language understood by the DBMS.

The DBMS receives the request, and checks it for syntax errors.

The DBMS then inspects (проверяет), in turn (в свою очередь), the external (внешнюю) sch[к]ema, the conceptual schema, and the mapping (соответствие) between the conceptual schema and the internal schema.

And then performs the necessary operations on the stored data.

So, all requests for access to data from users are handled by the DBMS.

Users are the persons who use the database for accessing data.

Types of database users include administrators, developers, and other users.

========================================

**Datamining**

Data Mining is a process of finding potentially useful patterns from huge data sets.

It is a skill that uses machine learning, statistics, and AI (artificial intelligence) to extract (извлечь) information to evaluate (оценить) future events probability.

The information derived (полученная) from Data Mining is used for marketing, fraud detection (обнаружение мошенничества), scientific discovery.

The most popular tool used when mining is artificial intelligence (AI).

AI technologies try to work the way the human brain works, by making intelligent guesses (предположения), learning by example, and using deductive reasoning (рассуждения).

Some of the more popular AI methods used in data mining include neural networks, clustering, and decision trees.

* **Neural networks** look at the rules of using data, which are based on the connections found or on a sample set of data.
* **Clustering** divides data into groups based on similar features.

Clusters are used when data isn't labelled in a favourable way to mining.

* **Decision trees**, like clusters, separate the data into subsets and then analyse the subsets to divide them into further subsets, and so on.

The final subsets are small enough that the mining process can find interesting patterns and relationships within the data.

When the process is complete, the mining software generates a report.

An analyst check to see if further work needs to be done, such as refining (уточнение) parameters, using other data analysis algorithm, or even delete data if it's unusable.