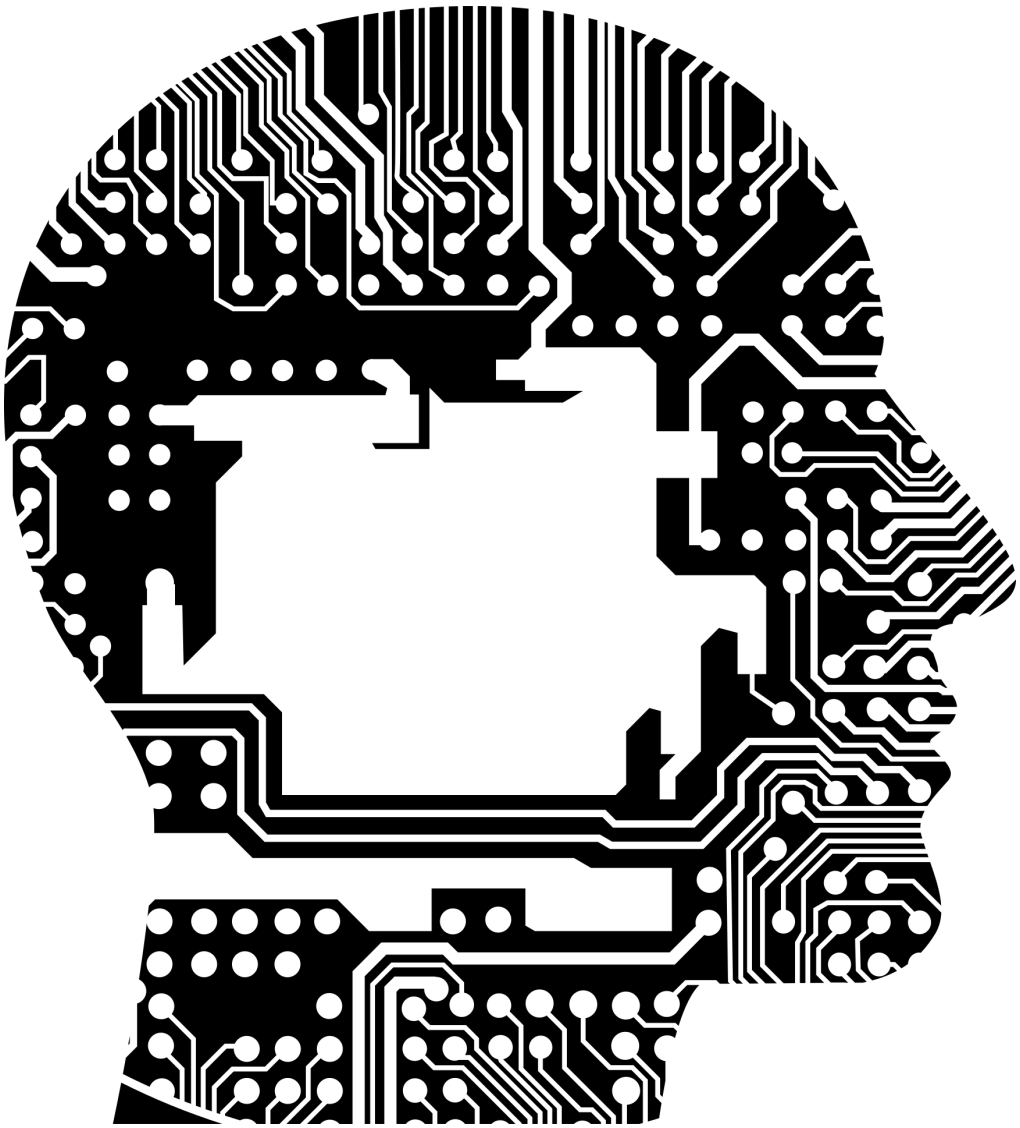


TOP DATA SCIENTIST

PROGRAMMING LANGUAGE

Table of Contents

Introduction	3
The R Programming Language	3
Python Programming Language	5
MATLAB Programming Language	6
Hadoop Programming Language	8
The SQL Programming Language	9
The SAS Programming Language	10
Java Programming Language	10
The C++ Programming Language	12

Everywhere we look, Big Data — all the digital bits of information that help businesses flourish – is present. Big Data has become more reliant on programming languages. Programming languages make the programming process faster and easier, and as a result, they are becoming increasingly important in the field of Big Data and Analytics.

This eBook ranks the top ten programming languages for Data Scientists in terms of importance and application.

1 R Programming Language

R is one of the most widely used programming languages and software environments for graphics and statistical computing among data scientists. It's a critical tool for firms that rely on analytics and finance, such as Facebook, Google, and LinkedIn.

The R system is gaining popularity thanks to its dialect of the S language and lexical scoping semantics. Every few months, the language gains new capabilities and features, cementing its position as the most significant tool for data visualization, computational statistics, and data science.

There are no fees associated with any of the programming language's versions. Microsoft Windows 32-bit versions are available for Linux, as well as OS X for UNIX and Macintosh. It's also accessible through the R Archive Network's Comprehensive R Archive Network (CRAN).

Rick Becker, John Chambers, and Allan Wilks created the R Language at AT&T Bell Laboratories, and it was launched in 1995.

R's users usually utilize a command-line interpreter to access it because it is an interpreted language. The computer responds with 4 when a user puts 2+2 at the R command prompt and enters data.

R uses vectors, matrices, arrays, data frames (akin to tables in a relational database), and lists as data structures.

FEATURES

- Matrix arithmetic is supported.
- The GNU General Public License (GPL) is used to make this software available for free.
- There are pre-compiled binary versions for every operating system included.
- The language employs a command line interface.
- Implements a wide range of statistical and graphical techniques, including as linear and non-linear modelling, time-series analysis, traditional statistical tests, classification, clustering, and more.
- It is easy expandable due to functions and extensions.
- The majority of R's standard functions are defined in the language itself, making it easy for users to track algorithm modifications.
- R has better object-oriented programming capabilities than other statistical computing languages.
- R's lexical scoping rules make it easier to extend it.
- It includes static Graphics, which generates graphs of publication quality, complete with mathematical symbols.
- Interactive and dynamic visuals are provided as part of supplementary packages.
- R has its own LaTeX-like documentation format, unlike other computer languages. This format is excellent for providing extensive documentation in a variety of formats, both online and in print copy.
- R provides procedural programming for some functions while object-oriented programming is supported for others.

This programming language is used by data scientists and statisticians all over the world to address some of their most difficult challenges in domains ranging from computational biology to quantitative marketing.

Because charts and graphs are used to portray complicated data, language has become an important aspect of the data analysis process.

Bill Cleveland and Edward Tufte, thought leaders in data visualization, have influenced the language, making it easier to derive meaning from multidimensional data with multi-panel charts, 3D surfaces, and more.

Almost every data manipulation, graphic, and statistical model required by a modern data scientist have ID built-in.

This cutting-edge community-reviewed statistics and predictive modelling method from prominent data scientists are easy to access, download, and use.

Download for free.

2 Python Programming Language

Python is a high-level, general-purpose programming language. Python's design philosophy emphasizes code readability. Because of the syntax in Python, programmers may convey their ideas in fewer lines of code than they would in languages like Java or C++.

The idea for a programming language resembling Python was conceived in the late 1980s, and Guido van Rossum of CWI in the Netherlands implemented it in December 1989. In 1991, it was released.

It was created as a successor to the ABC language, with the goal of being able to handle and interface with the Amoeba Operating System with ease. C and Modula 3 were also inspirations for the language.

FEATURES

- The language provides constructs with the goal of allowing clear programmes to be written on a large and small scale.
- Various programming paradigms, including imperative, object-oriented, and functional or procedural programming, are supported by the language.
- It includes a dynamic system, a huge standard library, and a memory management mechanism that is automated.
- The program also functions as a glue language for C, C++, and Forton-based programs.

FEATURES

- It has a significant advantage in terms of scope. As an illustration, consider the following: R can run Machine Learning Algorithms on a preprocessed dataset. Python, on the other hand, excels with data processing. Python makes use of Panda, an extremely helpful package that can perform everything SQL and R do, plus a lot more.
- Python is well-known in the programming community for its ease of use.
- NumPy, SciPy, Pandas, Matplotlib, and IPython are among its key libraries.
- In terms of prototyping and creating reusable and tiny systems, the language is extremely productive.
- It's a programming language that can be used for anything.

Data scientists are frequently involved in network application wiring, data processing job scripting and automation, web programming, and other procedures such as data munging. They believe it is preferable to accomplish all of this in one language, in addition to the actual analysis and modelling.

Professionals begin learning Python to become proficient in an all-in-one language. Python programmers are in great demand since it is employed by larger firms for crucial tasks such as evaluating large data sets.

Click here to Download Python.

3 Matlab

Programming Language

Matrix Laboratory, also known as MATLAB, is a multi-paradigm numerical computing environment and a fourth-generation programming language. It was first launched in 1984, and it was developed by MathWorks. In 2004, Matlab reached a million users in both industry and academia. Matlab users come from a variety of backgrounds, including engineering, economics, and science. A number of research institutions and industrial enterprises use the programming language.

FEATURES

- The programming language provides for the plating of functions and data, matrix manipulations, algorithm implementation, connecting with other languages such as C, C++, Forton, Java, and Python, and the building of user interfaces.
- Despite the fact that the language was designed for numerical computing, it now has an optional toolbox that uses the MuPAD symbolic engine to provide access to symbolic computing features.
- Simulink, a separate software, adds multi-domain graphical simulation and model-based design for dynamic and embedded systems to the mix.
- For design, iterative exploration, and problem resolution, the language provides an interactive environment.
- Statistics, linear algebra, filtering, Fourier analysis, optimizations, solving ordinary differential equations, and numerical integrations are among the mathematical functions available.
- It features graphics built-in for visualizing tools and data in order to construct bespoke charts.
- To increase code quality and maintainability, as well as to maximize performance, development tools are available.

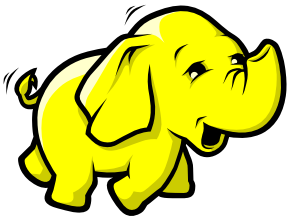
Despite its apparent limitations, this language is used by a large segment of the data analysis and scientific communities to address problems that are represented as matrix problems. Data scientists can execute analysis and acquire insight into data faster with this programming language than with C, C++, or Visual Basic. Data analysts use MATLAB to retrieve information from spreadsheets, files, databases, data gathering hardware, and other software and examine data for trends, estimate uncertainty, and test hypotheses. It aids data analysts in creating customized algorithms, models, and visualizations, as well as the publication of customized reports.

Download MATLAB.

4

Hadoop

Programming Language



Hadoop is a free and open-source platform for distributed processing and storage of massive data volumes. In simple terms, it is a framework that enables professionals to handle and store large amounts of data using simple programming models across clusters of computers in a distributed environment. The language is built to scale from a single server to millions of machines, each with its own storage and computing.

Hadoop is written in Java, and all of its modules are built on the premise that hardware problems are inevitable and should be handled automatically by the software.

In 2005, two Yahoo workers, Doug Cutting and Mike Cafarella, built Hadoop. Hadoop is a cross-platform operating system that was released in 2011.

The following modules comprise the Apache Hadoop framework's foundation:

- Hadoop Common — is a module that contains utilities and libraries that are required by other Hadoop modules.
- Hadoop Distributed File System (HDFS) — The HDFS module is a distributed file system for storing data on commodity devices with high aggregate bandwidth across a cluster.
- Hadoop YARN — YARN is a resource management framework that manages the computation of resources in clusters and uses them to schedule applications for users.
- Hadoop MapReduce — is a programming technique for processing massive amounts of data.

Hadoop gave data scientists additional options for storing and processing data. Hadoop offers concurrent distributed processing of enormous volumes of data across industry-standard servers, rather than relying on proprietary hardware and other systems for processing and storing data. There is no such thing as too much data with Hadoop.

Download Hadoop.

5

SQL

Programming Language

In the 1970s, Donald D Chamberlin and Raymond F Boyce at IBM created the Structured Query Language (SQL). The language was originally built on relational algebra and Tuple relational calculus.

It's a special-purpose programming language designed for managing data in a Relational Database Management System (RDMS) or stream processing in a Relational Data Stream Management System (RDMS).

SQL is one of the most widely used languages today, and it is recommended for all new data kinds and data cases. SQL is not simply a default choice in Big Data; it's also a recognition that SQL is one of the best-suited languages for basic analysis.

The foundation of this language is the concept of relational algebra, which provides a framework for the structure and manipulation of data sets. This mathematical system is briefly used in SQL syntax.

FEATURES

- The language consists of three parts: data definition, data control, and data manipulation.
- The language is a database creation and manipulation computer language that is ISO and ANSI standard.
- Users can create, update, delete, and retrieve data from a database using this programme.
- The language itself is simple and easy to pick up.
- Oracle, DB2, Sybase, MS Access, MS SQL Server, and more products are compatible with the language.
- In the data cleansing stage, SQL is crucial.
- It enables meaningful data to be queried and extracted from vast and complicated databases.

6 SAS Programming Language

SAS is a statistical analysis computer programming language. It began as a university initiative at North Carolina State University. Data from standard databases and spreadsheets are read into the language. The analysis' results are subsequently sent out in the form of tables and graphs.

The language is supported by compilers for Linux, Microsoft Windows, and a variety of other UNIX and mainframe systems.

In the commercial analytics arena, the language has been the unchallenged market leader.

SAS was created at North Carolina State University from 1966 until 1976.

FEATURES

- A wide range of statistical functions are available in the language.
- It offers an easy-to-use graphical user interface and even superior technical support.
- In the Private Organizations, it has the largest market share.
- The language is simple to learn and provides a convenient solution for professionals with prior SQL experience.

Download SAS.

7 Java Programming Language

The language was created in 1991 by the Green Team, a group of technologists. Java was originally called Oak, and it was created for handheld devices and set-up boxes. After failing miserably as OAK, the name was changed to JAVA in 1995 and the software was rewritten to make use of the World Wide Web.

Java is a general-purpose, object-oriented programming language that is regarded as one of the best for programmers and developers. It currently holds the top spot for the best programming language and is the most popular operating system with Android.

The language is utilized across a variety of platforms, including mobile-based applications, desktop application development, enterprise-level applications, and the construction of Android apps for smartphones and tablets.

FEATURES

- The application's goal is to 'Write Once, Run Anywhere' (WORA), which means writing code once and having it run on all platforms that accept the language.
- The language's syntax is adapted from C++.
- The language has been streamlined to remove language features that are known to cause programming problems.
- Java's importance stems from its extensive library offerings, which provide solutions to most of the major issues that developers face while creating enterprise applications.
- Simple: Most concepts are derived from C++, making it simple to read and write.
- Secure: The program cannot affect other system programs, and it offers a safe way to create internet apps.
- Java programs are portable, meaning they can execute in any environment that has a Java run-time system.
- Multithreaded: It allows you to program in multiple threads.
- Java programs are dynamic because they contain a large quantity of run-time information that is needed to verify and resolve object access at run time.

Download Java.

⑧ C++ Programming Language

C++ is a general-purpose programming language that is widely used in high-volume, high-frequency trading.

C++ has objective-oriented and generic programming tools that allow for memory manipulation at a low level.

Bjarne Stroustrup created the language, which was launched in 1983.

FEATURES

- It is based on the previous C programming language.
- Simula characteristics are used in the language.
- Unlike other management systems, C++ is the only language that supports RAII - it allows you to regulate the lifetime of objects.
- Its flexible programming and proximity to hardware add to its performance.
- Because of the language's simplicity, top tech companies such as Facebook and Apple employ it.

⑨ Julia Programming Language

Je Bezanson created the language, which first emerged in 2012.

Julia was established as a high-level dynamic programming language to suit the needs of high-performance scientific and numerical computation, as well as general-purpose programming.

FEATURES

- It includes multiple dispatches, which allows you to design functional behaviour for a variety of argument types.
- For documentation, dispatch, and optimization, a dynamic type system is used.
- It comes with a package manager.
- It has metaprogramming features, such as lisp-like macros.
- The PyCall package is used.
- It offers a sophisticated shell-like capacity for managing processes.
- It was created with parallelism and distributed computing in mind.
- Built-in types are faster and more compact than user-defined types.
- For various argument types, automatic development of efficient, customized code is performed.
- Conversions and promotions for numeric and other kinds that are elegant and extendable.

Download Julia.

Check out the Python training that we are offering here

For other programming language, you may check out here