R Advantages and Disadvantages

R is the most popular programming language for statistical modeling and analysis. Like other programming languages, R also has some advantages and disadvantages. It is a continuously evolving language which means that many cons will slowly fade away with future updates to R.

There are the following pros and cons of R



Pros

1) Open Source

An open-source language is a language on which we can work without any need for a license or a fee. R is an open-source language. We can contribute to the development of R by optimizing our packages, developing new ones, and resolving issues.

2) Platform Independent

R is a platform-independent language or cross-platform programming language which means its code can run on all operating systems. R enables programmers to develop software for several competing platforms by writing a program only once. R can run quite easily on Windows, Linux, and Mac.

3) Machine Learning Operations

R allows us to do various machine learning operations such as classification and regression. For this purpose, R provides various packages and features for developing the artificial neural network. R is used by the best data scientists in the world.

4) Exemplary support for data wrangling

R allows us to perform data wrangling. R provides packages such as dplyr, readr which are capable of transforming messy data into a structured form.

5) Quality plotting and graphing

R simplifies quality plotting and graphing. R libraries such as ggplot2 and plotly advocates for visually appealing and aesthetic graphs which set R apart from other programming languages.

6) The array of packages

R has a rich set of packages. R has over 10,000 packages in the CRAN repository which are constantly growing. R provides packages for data science and machine learning operations.

7) Statistics

R is mainly known as the language of statistics. It is the main reason why R is predominant than other programming languages for the development of statistical tools.

8) Continuously Growing

R is a constantly evolving programming language. Constantly evolving means when something evolves, it changes or develops over time, like our taste in music and clothes, which evolve as we get older. R is a state of the art which provides updates whenever any new feature is added.

Cons

1) Data Handling

In R, objects are stored in physical memory. It is in contrast with other programming languages like Python. R utilizes more memory as compared to Python. It requires the entire data in one single place which is in the memory. It is not an ideal option when we deal with Big Data.

2) Basic Security

R lacks basic security. It is an essential part of most programming languages such as Python. Because of this, there are many restrictions with R as it cannot be embedded in a web-application.

3) Complicated Language

R is a very complicated language, and it has a steep learning curve. The people who don't have prior knowledge or programming experience may find it difficult to learn R.

4) Weak Origin

The main disadvantage of R is, it does not have support for dynamic or 3D graphics. The reason behind this is its origin. It shares its origin with a much older programming language "S."

5) Lesser Speed

R programming language is much slower than other programming languages such as MATLAB and Python. In comparison to other programming language, R packages are much slower.

In R, algorithms are spread across different packages. The programmers who have no prior knowledge of packages may find it difficult to implement algorithms.

R integration with Hadoop

What is Hadoop?

**Hadoop** is an open-source framework which was founded by the **ASF - Apache Software Foundation**. It is used to store process and analyze data that are huge in volume. Hadoop is written in Java, and it is not OLAP (Online Analytical Processing). It is used for batch/offline processing. It is being used by Facebook, Google, Twitter, Yahoo, LinkedIn, and many more. Moreover, it can be scaled up just by adding nodes in the cluster.

Why integrate R with Hadoop?

R is an open-source programming language. It is best suited for statistical and graphical analysis. Also, if we need strong data analytics and visualization features, we have to combine R with Hadoop.

The purpose behind R and Hadoop integration:

1. To use Hadoop to execute R code.
2. To use R to access the data stored in Hadoop.

R Hadoop Integration Method

Hadoop and R complement each other very well in terms of big data visualization and analytics. There are four ways of using Hadoop and R together, which are as follows:



R Hadoop

The R Hadoop methods are the collection of packages. It contains three packages i.e., rmr, rhbase, and rhdfs.

**The rmr package**

For the Hadoop framework, the rmr package provides MapReduce functionality by executing the Mapping and Reducing codes in R.

**The rhbase package**

This package provides R database management capability with integration with HBASE.

**The rhdfs package**

This package provides file management capabilities by integrating with HDFS.

Hadoop Streaming

Hadoop Streaming is a utility that allows users to create and run jobs with any executable as the mapper and/or the reducer. Using the streaming system, we can develop working Hadoop jobs with just enough knowledge of Java to write two shell scripts which work in tandem.

The combination of R and Hadoop appears as a must-have toolkit for people working with large data sets and statistics. However, some Hadoop enthusiasts have raised a red flag when dealing with very large Big Data excerpts. They claim that the benefit of R is not its syntax, but the entire library of primitives for visualization and data. These libraries are fundamentally non-distributed, making data retrieval a time-consuming affair. This is an inherent flaw with R, and if you choose to ignore it, both R and Hadoop can work together.

RHIPE

RHIPE stands for **R and Hadoop Integrated Programming Environment**. Divide and Recombine developed RHIPE for carrying out efficient analysis of a large amount of data.

RHIPE involves working with R and Hadoop integrated programming environment. We can use Python, Perl, or Java to read data sets in RHIPE. There are various functions in RHIPE which lets HDFS interact with HDFS. Hence, this way we can read, save the complete data which is created using RHIPE MapReduce.

ORCH

ORCH is known as Oracle R Connector. This method is used to work with Big Data in Oracle appliance particularly. It is also used on a non- Oracle framework like Hadoop.

This method helps in accessing the Hadoop cluster with the help of R and also helps to write the mapping and reducing functions. It allows us to manipulate the data residing in the Hadoop Distributed File System.