禪

R، زبانی با پتانسیل بالا در تحلیل

آیا R به زبان رسمی برنامهنویسی بزرگدادهها تبدیل خواهد شد؟

Programming
HTML
</>
.js
PHP
Is R the Next Generation Programming Language for Big Data?

بهرغم مزایای بسیار زیادی که بزرگدادهها برای ما به همراه داشتهاند و باعث پیشرفت بسیاری از کسب و کارها شدهاند، اما واقعیت این است که دردسرهای بسیاری برای طراحان و توسعهدهندگان به همراه داشتهاند . بسیاری از توسعهدهندگان بر این باورند که زبانهای برنامهنویسی سنتی این پتانسیل را ندارند تا برای غلبه بر پیچیدگیها و حجم بالای دادهها مورد استفاده قرار گیرند و بر همین اساس ما به زبانهای ویژهای نیاز داریم که برای این منظور طراحی شدهاند . دانشمندان علم داده و طراحان برای غلبه بر این چالش گزینههای متعددی پیش رو دارند که از جمله این گزینهها میتوان به سکوهای توسعه گرافیکمحور، زبانهای مشتق شده از خانواده سی همچون سی پلاس پلاس، جاوا و درنهایت زبان R اشاره کرد .

در 25 سال گذشته زبان برنامهنویسی R موفق شده است محبوبیت خود را نزد توسعهدهندگان حفظ کند. اما هیچگاه به اندازه زبانهای دیگری همچون زبان سی یا اسلاف آن موفق نشد بهشکل گسترده به نیازهای توسعهدهندگان پاسخ دهد. اما به نظر میرسد اوضاع بهآرامی در حال تغییر است . این زبان در بعضی از زمینهها بهبود پیدا کرده است و همین موضوع زبان R را به ایدهآلترین گزینه برای کار با بزرگ دادهها تبدیل کرده است. به طوری که اکنون زمان آن فرا رسیده است تا بهشکل جدیتری از زبان R استفاده کنیم . اولیور براچت یکی از وبلاگنویسان R در پستی که مخاطب آن متخصصان و توسعهدهندگان هستند، به توصیف مزایا و قابلیتهای زبان R پرداخته است. او در پست خود با اشاره به نقطه نظرهای توسعهدهندگان بزرگ نوشته است که زبان R بهمنظور مدیریت و غلبه بر چالش بزرگ دادهها ایدهآلترین زبان است، به طوری که بدون هیچگونه مشکلی قادر است با این حجم عظیم از دادهها کار کند . همچنین، به این نکته توجه داشته باشید که تسلط بر زبان R و بهویژه بزرگ دادهها مزیتهای بسیاری به همراه دارد . گزارشهایی که

Top 20 Industries Hiring Big Data Expertise
1.05%
0.76%
1.71%
1.22%
0.62%
3.04%
5.70%
27.14%
8.13%
8.20%
9.62%
18.89%
12.35%
Professional, Scientific and Technical Services
Information
Manufacturing
Retail Trade
Sustainabilty, Waste Management & Remediation Svcs
Finance and Insurance
Wholesale Trade
Educational Services
Other Services (except Public Administration)
Accommodation and Food Services
Health Care and Social Assistance
Real Estate and Rental and Leasing
Construction
- 1

Заголовок

Текст текст текст текст текст текст текст текст текст текст текст текст

Текст текст текст текст текст текст текст текст текст (текст текст) 21

Текст текст текст текст текст текст

Текст текст текст текст текст текст текст текст текст текст текст текст

Language Rank
Types
Spectrum Ranking
1. C
100.0
2. Java
98.1
3. Python
98.0
4. C++
95.9
5. R
87.9
6. C#
86.7
7. PHP
82.8
8. JavaScript
82.2
9. Ruby
74.5
10. Go
71.9
Your primary programming language for Analytics, Data Mining, Data Science tasks: [512 voters]
2015 primary programming language:
R (and its packages) (263)
51% (of 2015 votes)
Python (including scikit-learn and other libraries) (151)
29%
Other (Java, MATLAB, SAS, Scala, etc) (89)
17%
none (9)
1.8%
2014 primary programming language:
R (and its packages) (237)
46% (of 2014 votes)
Python (including scikit-learn and other libraries) (117)
23%
Other (Java, MATLAB, SAS, Scala, etc) (118)
23%
none (40)
7.8%

http://spectrum.ieee.org/computing/software/the-2016-top-programming-languages
http://www.kdnuggets.com/polls/2015/r-vs-python.html

شکل **2** - همان
گونه که در تصویر مشاهده میکنید، زبانهای مبتنی بر سی جزء پنج زبان برتر به شمار میروند . از طرف زبان **R** و بستههای
نرمافزاری متعلق به آن و در ادامه پیتون و کتابخانههای مرتبط به آن اصلیترین گزینهها در زمینه دادهکاوی و تحلیل به
شمار میروند .

اگر به دنبال پردازش بزرگ دادهها هستید، وقت خود را روی یادگیری **R** سرمایهگذاری کنید

زبان برنامهنویسی R اولین مرتبه در سال 1993 به دنیای برنامهنویسی وارد شد . این زبان برنامهنویسی در 20 سال گذشته بهشکل محدودی از سوی توسعهدهندگان مورد استفاده قرار میگرفت، اما در چند سال اخیر بهواسطه تواناییش در پردازش بزرگ دادهها یک بار دیگر مورد توجه توسعهدهندگان قرار گرفته است . شرکتهای نرمافزاری و توسعهدهندگان از چند سال پیش بهشکل گستردهای از این زبان بهمنظور جمعآوری بزرگ دادهها و رایانش توزیع شده استفاده کردهاند . همین موضوع باعث شده است تا امروزه شاهد عرضه کتابخانههای گستردهای از این زبان باشیم که در زیرساختهای بزرگ رایانش توزیع شده مورد استفاده قرار میگیرند . جالب آنکه تعدادی از این کتابخانهها را میتوان روی سامانههای کوچکتر و حتی کامپیوترهای شخصی مورد استفاده قرار داد .

مارتین هلر ویراستار سایت اینفوورلد دلایل متعددی را برای انتخاب زبان R بهعنوان ایدهآلترین راهکار مدیریت بزرگ دادهها برشمرده است . او در بخشی از مقاله خود آورده است : «بستهها و توابعی از زبان R وجود دارند که قادرند دادهها را از هر منبع معقول و مناسبی بارگذاری کنند و تنها محدود به فایلهای CSV نیستند . فراتر از کاراکترهای متمایزکنندهای همچون کاما که از طریق تابع ()read.tabe مورد استفاده قرار میگیرند، توسعهدهندگان این توانایی را دارند تا جداول دادهای را کپی کرده و انتقال داده، فایلهای اکسل را خوانده، اکسل را به R متصل کرده، دادههای متعلق به SAS و SPSS را فراخوانی کرده، به بانکهای اطلاعاتی دسترسی داشته، به سرویسهای ارائه شده از سوی Saleforce متصل شده و به رابطهای RESTful دسترسی داشته باشند. در واقع، شما برای وارد کردن دادهها نیاز ندارید ترکیب نحوی هریک از سرویسها و نرمافزارهایی که به آنها اشاره شد را یاد بگیرید . بهواسطه آنکه RStudio����� منوی Tools ویژهای بهنام Import Database برای وارد کردن دادهها در اختیارتان قرار میدهد. این منو به توسعهدهندگان کمک میکند با نگاه کردن به دادهها از درون یک فایل متنی یا نشانی اینترنتی و تنظیم گزینه مناسب تبدیل از طریق منوی کشویی که این ابزار در اختیارشان قرار میگیرد، دستورات تعاملی خود را بهشکل درستی ایجا کنند .»

مطلب پیشنهادی

BATch

این زبانها به شما در درک بهتر برنامهنویسی کمک میکنند

سادهترین زبانهای برنامهنویسی ویژه افراد تازهکار

بارگذاری دادهها از منابع چندگانه

پیش از آنکه بزرگ دادهها به یکی از جریانهای بزرگ دنیای ما تبدیل شوند، اکثر برنامههای معروف دادههای مورد نیاز خود را تنها از یک منبع دریافت میکردند، اما امروزه دیگر این گونه نیست . بزرگ دادهها همراه با ظهور فناوری ناآشنایی بهنام اینترنت اشیا به دنیای ما وارد شدند . امروزه اکثر پروژههای کوچک و بزرگ برای انجام کارهای خود به دادههایی که از منابع چندگانه به دست میآیند وابسته است. برنامههای بازاریابی یک نمونه کلاسیک در این زمینه به شمار میروند . این گونه برنامهها اطلاعات مرتبط با مشتریان را از بانک اطلاعاتی خود، شبکههای اجتماعی و از دستگاههایی که مشتریان مورد استفاده قرار میدهند به دست میآورند . به همین دلیل است که شما به زبان برنامهنویسی ویژهای نیاز دارید که بتواند محاورههای مربوط را بهمنظور جمعآوری دادهها از منابع مختلف به دست آورد و آنها را پردازش کند . (شکل 3)

>install.packages("package name")
1
Pre Modeling Stage
Data Analysis
Useful Libraries in
R
2
Modeling Stage
1. Data Visualization
ggplot2, googleVis
2. Data Transformation
plyr, data.table
3. Missing Value Imputations
MissForest, MissMDA
4. Outlier Detection
Outliers, EVIR
5. Feature Selection
Features, RRF
6. Dimension Reduction
FactoMineR, CCP
1. Continuous Regression
car, randomforest
2. Ordinal Regression
RMiner, CoreLearn
3. Classification
Caret, BigRF
4. Clustering
CBA, RankCluster
5. Time Series
forecast, LTSA
6. Survival
survival, Basta
Post Modeling Stage
3
1. General Model Validation
LSMeans, Comparison
2. Regression Validation
RegTest, ACD
3. Classification Validation
BinomTools, DAIM
4. Clustering Validation
ClustEval, SigClust
5. ROC Analysis
PROC, TimeROC
Other Libraries
A. Improve performance
Rcpp,parallel
B. Work with web
XML, jsonlite, httr
C. Report results
shiny, RMarkdown
D. Text Mining
tm, twitteR
E. Database
sqldf, RODBC, RMongo
F. Miscellaneous
swirl, reshape2, qcc
Analytics Vidhya
Learn Everything About Analytics

شکل **3** -

فهرستی از بستههای نرمافزاری مفیدی که برای تحلیل دادهها در زبان **R** در اختیار توسعهدهندگان قرار دارد . برای دانلود این کتابخانهها به نشانی **https://discuss.analyticsvidhya.com/t/download-the-complete-list-of-powerful-r-libraries-for-data-analysis/2624** مراجعه کنید .

توان وفقپذیری و سازگاری برنامهنویسی

یادگیری هر ترکیب نحوی جدید زمانبر است . متأسفانه زبانهای برنامهنویسی همهمنظوره بهگونهای طراحی شدهاند که باید نکات مربوط به آنها با دقت بیشتری یاد گرفته شوند . تمرکز زیاد روی جزئیات بهویژه در زمان کار با بزرگ دادهها اهمیت پیدا میکند . اما زبان برنامهنویسی R در این زمینه یک استثنا به شمار میرود . بهواسطه آنکه اگر بر دستورات پایه تسلط پیدا کنید، قادر خواهید بود از کتابخانههای مرتبط با این زبان بهمنظور پردازش بزرگ دادهها بدون هیچ مشکل خاصی استفاده کنید .

سازگاری با زبانهای دیگر

یکی از نکات مثبتی که در ارتباط با زبان برنامهنویسی R وجود دارد این است که به شما اجازه میدهد زبان R را همراه با زبانهای مرتبط دیگری با این زبان همچون سی پلاس پلاس مورد استفاده قرار دهید .

توانایی استخراج از زیرساختهای ابری

اگر توسعهدهندگان ترکیب نحوی dplyr در زبان R را فراگیرند، قادر خواهند بود از این ترکیب نحوی بهمنظور اجرای محاورههای بزرگ دادهها روی زیرساختهای مختلف ابری همچون Google BigQuery و Amazon Redshift استفاده کنند .

آموزش نصب R در هاست

نصب R در هاست اشتراکی کار بسیار دشواری است و به همین دلیل بسیاری از کاربران از نصب R در هاست اشتراکی منصرف می‌شوند. اما اگر شما هم جزو این دسته از کاربران هستید که می‌خواهید R را در هاست اشتراکی نصب کنید، نگران نباشید. در این مقاله به شما آموزش می‌دهیم که چگونه R را در هاست اشتراکی نصب کنید.

قبل از شروع آموزش، باید بگوییم که برای نصب R در هاست اشتراکی، شما نیاز به دسترسی به SSH دارید. اگر هاست شما دسترسی به SSH ندارد، متاسفانه نمی‌توانید R را در آن نصب کنید. در این صورت، پیشنهاد می‌کنیم که یک هاست اشتراکی با دسترسی به SSH تهیه کنید.

برای نصب R در هاست اشتراکی، ابتدا باید به SSH هاست خود متصل شوید. برای این کار، می‌توانید از نرم‌افزارهای مختلفی مانند PuTTY استفاده کنید. پس از اتصال به SSH، دستور زیر را وارد کنید:

```
sudo apt-get install r-base
```

این دستور، R را در هاست شما نصب می‌کند. پس از نصب، می‌توانید با دستور زیر R را اجرا کنید:

```
R
```

اکنون R در هاست شما نصب شده است و می‌توانید از آن استفاده کنید.

نصب پکیج‌های R

پکیج‌های R را می‌توانید از CRAN نصب کنید. برای این کار، دستور زیر را وارد کنید:

```
install.packages("package_name")
```

به جای package_name، نام پکیج مورد نظر خود را وارد کنید. برای مثال، اگر می‌خواهید پکیج ggplot2 را نصب کنید، دستور زیر را وارد کنید:

```
install.packages("ggplot2")
```

پس از نصب پکیج، می‌توانید با دستور زیر آن را بارگذاری کنید:

```
library(package_name)
```

به جای package_name، نام پکیج مورد نظر خود را وارد کنید. برای مثال، اگر می‌خواهید پکیج ggplot2 را بارگذاری کنید، دستور زیر را وارد کنید:

```
library(ggplot2)
```

اکنون پکیج ggplot2 بارگذاری شده است و می‌توانید از آن استفاده کنید.

نصب RStudio

RStudio یک محیط توسعه یکپارچه برای R است که به شما امکان می‌دهد کدهای R را به راحتی بنویسید و اجرا کنید. برای نصب RStudio در هاست اشتراکی، ابتدا باید به SSH هاست خود متصل شوید. سپس دستور زیر را وارد کنید:

```
sudo apt-get install rstudio-server
```

این دستور، RStudio Server را در هاست شما نصب می‌کند. پس از نصب، می‌توانید با دستور زیر RStudio Server را اجرا کنید:

```
sudo rstudio-server start
```

اکنون RStudio Server در هاست شما اجرا شده است و می‌توانید از آن استفاده کنید. برای دسترسی به RStudio Server، باید به آدرس زیر بروید:

```
http://your_server_ip:8787
```

به جای your_server_ip، آدرس IP هاست خود را وارد کنید. سپس نام کاربری و رمز عبور خود را وارد کنید و وارد RStudio Server شوید.

نتیجه‌گیری

در این مقاله به شما آموزش دادیم که چگونه R را در هاست اشتراکی نصب کنید. همچنین به شما آموزش دادیم که چگونه پکیج‌های R را نصب کنید و RStudio را در هاست اشتراکی نصب کنید. امیدواریم که این آموزش برای شما مفید بوده باشد.

:برچسب‌ها
آموزش نصب R در هاست
:دسته‌بندی
آموزش R
نصب R در هاست
نصب R در هاست اشتراکی
:تاریخ انتشار
11:55 - 28/06/1396
:مطالب
R آموزش R - نصب R در هاست اشتراکی - آموزش R در هاست اشتراکی

منبع:
https://www.shabakeh-mag.com/networking-technology/9505/%D8%A2%DB%8C%D8%A7-r-%D8%A8%D9%87-%D8%B2%D8%A8%D8%A7%D9%86-%D8%B1%D8%B3%D9%85%DB%8C-%D8%A8%D8%B1%D9%86%D8%A7%D9%85%D9%87%E2%80%8C%D9%86%D9%88%DB%8C%D8%B3%DB%8C-

%D8%A8%D8%B2%D8%B1%DA%AF%E2%80%8C%D8%AF%D8%A7%D8%AF%D9%87%E2%80%8C%D9%87%D8%A7-%D8%AA%D8%A8%D8%AF%DB%8C%D9%84-%D8%AE%D9%88%D8%A7%D9%87%D8%AF-%D8%B4%D8%AF%D8%9F