

OPEN SOURCE SPECIAL

CIOReview

The Navigator for Enterprise Solutions

NOVEMBER 30, 2016

CIOREVIEW.COM

IN MY OPINION

KEN OWENS,
CTO-CLOUD PLATFORMS & SERVICES GROUP,
CISCO

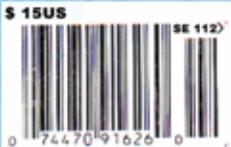
CXO INSIGHTS

JIM ZEMLIN,
EXECUTIVE DIRECTOR,
THE LINUX FOUNDATION

RStudio

The Futu(R)e of Data Science

Taref Kawaf,
President



CIOReview
44700, S Gimmer Blvd.
#202, Fremont, CA 94538

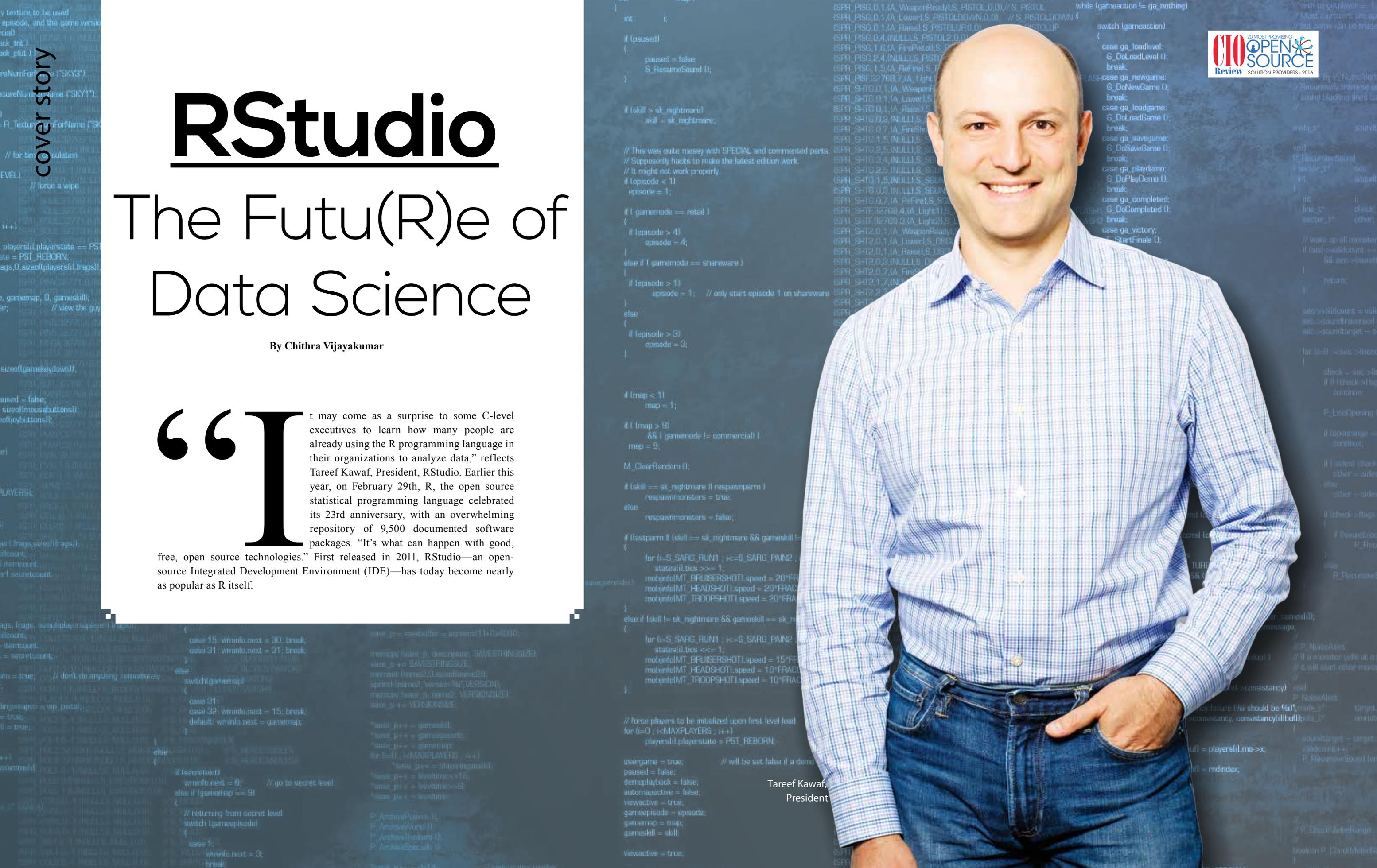
RStudio

The Futu(R)e of Data Science

By Chithra Vijayakumar

“It may come as a surprise to some C-level executives to learn how many people are already using the R programming language in their organizations to analyze data,” reflects Tareef Kawaf, President, RStudio. Earlier this year, on February 29th, R, the open source statistical programming language celebrated its 23rd anniversary, with an overwhelming repository of 9,500 documented software packages. “It’s what can happen with good, free, open source technologies.” First released in 2011, RStudio—an open-source Integrated Development Environment (IDE)—has today become nearly as popular as R itself.

Tareef Kawaf, President



When CEO, founder, and principal software developer, J.J. Allaire, started working on the concept of a web-based IDE for R, he didn't know whether a business would come out of it, but he knew that better open source software for data analysis, freely available to anyone with access to a computer, would be useful to academia, research, and industry. The status quo—spreadsheets, domain-specific tools, web applications, proprietary statistical software, and business intelligence products—was challenged as data became ubiquitous. “We are all working through the implications of living in an age where we can amass data far faster than we can consume and understand it,” observes Kawaf. “It's not just about the way we analyze data but also how we're going to run our businesses in an informed way.” Using an open source ecosystem like R can diminish the dependence on less flexible and often expensive tools while injecting a much-needed emphasis on reproducibility. “At RStudio, our engineering is focused on enabling a flexible and scalable “data science” toolchain for people using the R ecosystem but also for anyone who consumes their work in the nearly infinite forms of applications, reports, dashboards, and plots that R allows you to create,” says Kawaf.

A Substratum for Developers

R recently climbed to the fifth spot of IEEE's annual ranking of most popularly used programming languages. “R is a fantastic language for data analysis and its design makes it particularly amenable for creating domain specific languages to help solve problems in a human readable way,” states Kawaf. It is the only language of IEEE's top five, catering exclusively to the data science needs of today. For a growing number of data scientists, RStudio's open source tools have become the go-to means for accessing data, understanding it, and communicating their findings to others.

The flagship IDE tool, RStudio IDE has all the bread-and-butter features of a GUI development environment. Available from the RStudio website as a free download, licensed under the Affero General Public License (AGPL), the RStudio IDE is widely adopted by academicians, enterprises, and open source enthusiasts. The growing number of downloads which have crossed a hundred thousand per week, stands as a testament to its prominence. RStudio's stratagem so far rebuts those who would say ‘what good can come out of free software as opposed to commercially available proprietary software.’

Designed and developed under the open core model, the RStudio IDE open source version resides on GitHub where the community can submit pull requests. The RStudio IDE can be deployed as a standalone desktop, or as a server instance accessed by multiple users via a web browser, across a variety of Operating Systems. It provides a customizable workbench with all of the tools required to work with R in one place (console, source, plots, workspace, help, history, etc.), a syntax highlighting editor with code completion, and allows users to execute code directly from the source editor or within a notebook.

Another of RStudio's popular products is Shiny, which enables data scientists to create interactive web applications directly from R, without requiring web development skills. RStudio's Shiny is an open source package that data scientists use to communicate the results of data analysis to people who don't need to know R. “Shiny allows data scientists to create interactive web applications by generating them directly from R in a versionable, reproducible manner,” notes Kawaf. In addition, with Shiny Server or RStudio Connect, data scientists can easily share their applications with other teams and

“Data science with R is a liberating technical innovation that has really just started”



Tareef Kawaf,
President

users, giving the users the ability to access sophisticated algorithms and models through the familiar interactivity of the web.

Finding your way with Shiny

Illustrating Shiny's proficiency from the heart of Silicon Valley is Google's Waze, the world's largest community-based traffic and navigation app. Waze pulls a lot of real-time geospatial information from its logged-in users and the data is noisy by nature, involving a labyrinth of complex data structures. The team required a rapid and interactive framework that is efficient enough to process and visualize the complex and dynamically changing data. Above all, they needed the flexibility of a “do-it-yourself programming approach.” “We've been avid R users for a long time now, so in a matter of days from starting to work with RStudio Shiny Server, we deployed our first dashboard with much hype and success,” reported Daniel Marcous, Waze Data Wizard, Google. “As desired, it included both an interactive tool for analyzing geospatial data over a map, and some statistical analysis in the form of anomaly detection for irregular traffic.” The project's ongoing success reverberates every time someone taps on their smartphone to find quickest possible route to a destination.

“At RStudio, our engineering is focused on enabling a flexible and scalable ‘data science’ toolchain for people using the R ecosystem”

RStudio's open source and commercial products have also largely impacted enterprises outside Silicon Valley, such as Samsung, Nestlé, AstraZeneca, and GE due to the various advantages it holds against commercial platforms. Interestingly, RStudio disassociates itself from providing professional services around its solutions. “It is an important part of our strategy to stay focused on software and let third party entities with certified enterprise expertise on specific domains build their businesses on our foundation,” says Kawaf.



When RStudio Goes Big

Ticking one item off their product roadmap, RStudio recently released its newest product to extend the reach of shareable items via R. Dubbed RStudio Connect, the tool allows a horde of content such as plots, notebooks, dashboards, documents, presentations, and custom applications to be published with fewer clicks. “RStudio Connect allows data scientists to schedule and disseminate their work easily, while providing business users with a centralized location to interact and refresh their analyses,” says Kawaf. “It offers the most complete enterprise platform for publishing and managing all the work a team creates in R.”

In the golden age of big data that is burgeoning at an alarming rate, it has been observed that about 2.5 quintillion bytes of data are created every day. In the enterprise universe, the term big data is becoming as ubiquitous as data itself. Stemming from the need to support analytics on distributed data computation engines used in big data like Spark, RStudio recently developed Sparklyr, a Spark package for R. “Sparklyr makes it much easier for data scientists to call the powerful features of Spark and yet leverage the visualization, advanced analytical capabilities, and communication tooling within R,” explains Kawaf.

Data analysis is as universal an activity as one can imagine, spanning geography, company size, and industry. Although the open source R language has been available for more than two decades, its usefulness continues to grow, enhanced by RStudio, package developers, and millions of enthusiasts. As the explosion of data reaches newer levels, the need to turn data into value is of utmost importance. Predicting exciting times for his company, RStudio and the R ecosystem, Kawaf assures, “Data science with R—from data access to understanding, and communication in one coherent workflow—is a liberating technical innovation that has really just started.” 

CIOReview

The Navigator for Enterprise Solutions

OPEN SOURCE SPECIAL

NOVEMBER 30, 2016

CIOREVIEW.COM

20 Most Promising Open Source Solution Providers 2016

Digital transformations drive IT industry to remodel its infrastructure in a more agile and cost-effective manner. By spurring innovation for the past decade, open source technologies have reduced these demands by offering new possibilities, alternative business models and expanding markets in the information and communication technology space.

Open source technology gains in stature not only because of its flexibility and collaborative nature but also because of the potential benefits it caters to all the users. Numerous organizations prefer open source as it reduces lead times, improves quality, and interoperability. With significantly reduced R&D costs, especially important for new entrants, solutions can be brought to market faster.

Furthermore, the flood tide of new open-source technologies namely OpenStack and Linux Containers supports new use cases such as programmable infrastructure

through API, integration capability, and advanced application development. These changeovers are addressing the topical challenges of both enterprise IT and infrastructure software vendors, assisting them in planning and scaling with time. Today, IT infrastructure vendors are looking to seize the opportunities of open-source technology by choosing innovative solutions and integrating it in a comprehensive manner.

To help CIOs navigate through a plethora of open source technology vendors, our distinguished selection panel, comprising CEOs, CIOs, VCs, industry analysts and the editorial board of CIOReview has selected the 20 Most Promising Open Source Solution Providers. The listing exhibits the competence in delivering open source technologies poised for exponential growth in IT.

We present to you CIOReview's 20 Most Promising Open Source Solution Providers 2016.



Company:
RStudio

Description:
Developer of free and open source IDE and other packages for the R statistical computing environment

Key Person:
Tareef Kawaf
President

Website:
rstudio.com