Hello Julia: New Data Science Programming Language Shows Promise Despite Flaws

Written by David Telford | Published: 18 June 2016

If you're a programmer, then you may have run into a problem much like Stefan Karpinski did. I'll tell you the story, and you can see if it sounds familiar.

Stefan was an experienced grad student working on a software tool that could emulate wireless networks, and he was getting frustrated. He was using Matlab to code the linear algebra functionality, but Matlab isn't very good for statistics and visualization. For those parts, he was using R.

Neither language is very fast, so to speed things up he was using C for looping and other procedural optimizations. Finally, Ruby was his “glue,” a general purpose tool to keep the other three languages working together. One project, four languages. Later, he would describe it as a “Rube Goldberg” sort of program — complicated, elaborate, but not terribly elegant.

Understandably, Stefan began to wish for a language that could do everything he needed, an all-inclusive language to replace his Frankenstein-ian creation. He mentioned his frustration to fellow grad student Viral Shah, who in turn mentioned it to Jeff Bezanson, a computer scientist specializing in programming languages. Suddenly, the ball was rolling.

That's the origin story of the Julia programming language in a nutshell. Designed by Bezanson, Karpinski, Shah, and Bezanson's graduate advisor at MIT, Alan Edelman, Julia was intended to fix the so-called “two-language” problem in technical computing.

Ideally, a data scientist or mathematician can start and end with Julia, from prototyping to the finished project, without ever having to call on another language. And to some degree, it's working.

This new language's popularity is largely due to three things: support, community, and speed. Though the project is completely open-source, the developers have created a start-up support company for the language. The new business, appropriately called Julia Computing, not only allows the developers to work on Julia full-time, but also helps businesses feel more comfortable adopting the new language because they know they can find assistance.

Those wanting to code in Julia can also seek help from the thriving community, which includes quite a number of MIT students, faculty and alumni (including Bezanson and Edelman themselves), many of whom also contributed to the development of the language. And finally, Julia is primarily used in technical computing, where large data sets mean speed is key. Because Julia is almost as fast as C/++, while having greater functionality, it has seen most of its use in the data science community.

FEATURES
Being open-source is the first of a few powerful features in the Julia arsenal. Getting to the source is quick and straightforward, meaning that users who need grassroots flexibility can get in there and “hack” the language, providing themselves with their own features and tweaks. For advanced programmers, this will be a big advantage over proprietary languages like Matlab, where the source code is unavailable and the feature set unchangeable.