The Scripture of Python: An Inside Look at the Use of Scripting Programming Languages on Wall Street

Author: Anthony Malakian
Source: [Waters Technology](#) | June 2, 2015

In the wake of the financial crisis, firms have invested in research and data analytics platforms. For many of those tools, the coding underpinning them is the Python language. Anthony Malakian looks at how various capital markets firms are using this open-source option to improve risk and data management practices while seeking alpha.

In today's financial markets, the developers and programmers are the soldiers on the front lines executing the orders of traders, portfolio managers, CIOs and CEOs. In the ranks of IT, the language that you're proficient at is a Shibboleth – a sort of linguistic code word for how savvy and advanced you are at the trade.

In the 1980s, a man named Guido van Rossum created the Python programming language. It's not a complex language. It's not a tongue twister. In fact, its elegance resides in its simplicity. But if you want to get a job on Wall Street as a programmer or developer, you had better be able to say the word, "Python".

According to research conducted by media outlet Quartz, the average salary of a programmer proficient in Python is $100,717, third-highest behind only Ruby on Rails and Objective-C, and ahead of stalwarts C++, Java, C, R and SQL (see chart at bottom of the page). Andy Fundinger, a developer at Risk Focus, recently conducted a search of job listings and found that four of the top five banks by size, and seven of the top 10 banks, had open job listings for Python developers.

When it comes to scripting languages, Python has emerged over the past five-plus years as the language of choice. There are many reasons for this, which we will explore in this article, but the underlying trend is that since the global financial crisis, Wall Street firms are investing more in analytics and research tools, while at the same time managing a deluge of datasets flowing into the organization. And as this potion is being mixed, banks, traditional asset managers and hedge funds are becoming more comfortable with using open-source tools.

For this feature, Waters spoke with a cross-section of banks, brokers, prop-trading shops, hedge funds and technology vendors to get a better understanding as to how capital markets firms are using Python. While the open-source scripting language is still deemed undesirable for low-latency trading systems, it has proven a valuable tool when it comes to research, analytics and data delivery.

All In

Like most every other hedge fund, after 2008, AQR had to regroup to recoup lost assets. By 2009, assets under management had dipped below $30 billion, but as of Q1 this year, that number had ballooned to more than $130 billion. One of the reasons for that resurgence was a significant investment into its research team, which included a conversion from its legacy C++ code base to Python, Craig Austin, head of research development at AQR, tells Waters.

Today, Austin says, AQR's front-office research team is "100 percent Python". While C++, Java, C# and Scala are prevalent for other teams inside the organization, when it comes to research, the Greenwich, Connecticut-based manager has bought fully into Python.

There are other research-designed languages that are popular among hedge funds – such as R or Matlab – but Austin says that Python was helpful when it came to attracting new developers and it also served as an area of common ground between researchers and software engineers.

"Python is helpful because we have researchers – who are not computer scientists – working side-by-side with software engineers, and it is the language that makes both of those groups happy," he says. "If we were to use C++, the researchers might be lost a bit, and if we used R or Matlab, the developers might not be happy."

By standardizing its languages, Austin adds that the firm has cut down on coding errors and the need to work with IT to fix mistakes.
"Python reduces the need to rewrite code. A lot of groups like ours, researchers will write in one language – say R or Matlab – and the developers will rewrite in C++. We don't do that because we use Python front-to-back. So it helps us to be more efficient because we don't have to do any rewriting," he says.

Additionally, as the firm has grown, the research department has been able to easily scale along with that influx of assets. If they were using separate languages, then for every new researcher that was hired, the firm would have to hire an equal number of developers to couple with research.

There are drawbacks, however. The major knock on Python is its performance. Compared to compiled languages like C and C++, Python is relatively slow. Austin says that for performance-critical elements of their code, they'll write in C or C++ and then call that code from Python.

"So the researchers might not realize it, but sometimes they're calling C++ behind the scenes. So that helps mitigate some of the performance issues with Python," he says.

In the Mix

Since 2008, there have been several high-profile projects that have involved Python. Bank of America Merrill Lynch's Quartz platform, which prices trades, manages positions and calculates risk across the organization, is open source and was built using Python.

JP Morgan's Athena Platform, its cross-asset market risk and trading platform, uses a blend of Python, C++ and Java. Mark Ashton-Rigby, global co-head of technology JPMorgan's Corporate & Investment Bank, tells Waters Python serves as Athena's front-end language and "glues" different components together, which allows for faster application development. And since Athena is JPMorgan's "strategic front-office platform" for the bank's trading operations, as new apps are being rolled out for the system, those are also being written using Python.

Ashton-Rigby says that they considered using other scripting languages, but Python won out. The only other option was to develop a proprietary language, "but this made little commercial sense," he says.

In October 2014, when in-memory database provider McObject released v6.0 of its eXtremeDB Financial Edition database, it added support for Python. Market data provider Xignite has recently created a new software development kit (SDK) for Python. In November, analytics and data management solutions provider Xenomorph added Python to its TimeScape data management platform.

Brian Sentance, CEO of Xenomorph, says that the vendor turned to Python after a client request. The client was using Remote Python Call (RPyC), which allows for remote procedure calls and distributed computing.

"Unlike regular RPC mechanisms, such as ONC RPC, CORBA or Java RMI, RPyC is transparent, symmetric, and requires no special decoration or definition languages," Sentance explains. "Moreover, it provides programmatic access to any pythonic element, be it functions, classes, instances or modules."

At information solutions provider Integration Systems, the vendor has "significantly turned to Python," says Paul Short, infrastructure services director. The company chose to standardize much of its scripting and found Python to be better than Perl or Java when it came to deployment. Python is also close to C# (not to be confused with C) and C++.

"It's really helped us because it's the glue between all the heavy lifting of C# and C++," Short says. "The other thing that put the nail in the coffin for me was the adoption by the quants. There are a lot of quants coming through academia and education in the sciences and it seems to be the language of choice."

Monster Datasets

Kevin Kernan, director of product development at broker-dealer Wolverine Execution Services, says that WEX uses Python to help load and send "monster" datasets to and from clearing firms, and they use it to normalize data when different languages are used. Wolverine Trading, the proprietary-trading arm of the organization, uses Python for ad hoc analysis of tick data, internal quote data, market data updates and various other reports.

"Scripting languages fell out of favor as people moved toward compiled languages, but I'm not sure why that is," Kernan says. "Everyone started writing in Excel and when they ran into limitations they moved to compiled things, but now people are taking a step back and saying that there's a middle ground with something like Python."

It also helps that Python is more similar to popular compiled languages. WEX's first adoption of Python came when it wanted to modernize several simple Microsoft .NET libraries, so it turned to the IronPython implementation tool.

"In my own use, I find it very easy to go from C to Python," he says. "I would say it's closer to the tools that are already being used. I also think that the existence of IronPython [an open-source implementation of Python for .NET and Mono libraries] is a convenience that a lot of people use. Through IronPython, we can interact with existing compiled .NET applications very easily."

Ann Neidenbach, chief information officer of agency broker and services provider Convergex, says that right now "the ruling king of scripting languages at Convergex is R," although they are "experimenting with using more Python."

Neidenbach acknowledges that R is not as good at text or file handling as Python, but that R's built-in statistical and graphing functions are "without match" for Convergex's model prototyping and research applications.
Community

Neidenbach adds, though, that Python's developer pool is "very strong" and tools like NumPy and SciPy, developed specifically for Python, help to add analytical power and that "quant developers don't have to reinvent the wheel on every application."

Alexei Miller, managing director at consulting firm DataArt, says that Python is "getting a second life as the analytics language of choice on Wall Street" and that it is "slowly becoming the language of choice for quants."

When compared to C and C++, Python is relatively simple, allowing novices to pick it up quickly. "You can read a book on how to code in C++, but it's really easy to write really bad code in C++," Miller says. "In Python, you can write decent code a week into your training."

Even for firms using other scripting languages like Perl, Ruby or JavaScript, Python is still being used because of its simplicity for certain one-off data analysis procedures, says Al Chang, chief technology officer at Xignite.

"The simplicity for the scripting makes it easy to do one-off data analysis," Chang says. "A lot of what I do here is take data from a lot of different exchanges and venues and sift through it and find patterns [within it], but it's often necessary to write one-off scripts that you might later want to be able to codify. Python makes it so easy to make that first call."

Because Python is open source, it is free, unlike Matlab and other research languages. AQR's Austin notes that when his researchers are trying to figure out the answer to a problem – like how to run a new type of linear regression – they simply have to do a Google search and many of the results will be in Python.

"You can find Python solutions to a lot of common problems simply by searching on Google because that's where a lot of the focus is these days," he says.

Not a Panacea

When it comes to programming, there isn't one language to rule them all. It's a sauce made up of different ingredients. Compiled languages like C and C++ seem to be the most popular when it comes to high-performance trading systems. For research and analytics platforms, and to move large datasets around or to quickly deploy new tools, scripting languages are preferable because of their ease of use, and Python, at this time, is the language of choice over R, Perl, Ruby or JavaScript.

Integration Systems' Short notes that his company is working on a trading application where the entire back-end is written in C or C++, depending on the module, but the front-end that the user interacts with is written in Python using Django, an open-source web application framework written in Python that allows for quick front-end development.

Antonio Hallak, CEO of London-based investment bank Sibyl Trading, says that while the firm uses Python "heavily," it's usually for front-end presentations, while anything related to trade execution is done in C and C++. He only has one rule at his firm, he says: "We don't use any Java – we don't allow it."

And JPMorgan's Ashton-Rigby adds that while users can expect some latency issues with Python, it can still be used to compose any type of app, "including trading systems for high-frequency trading," he says, with the low-latency components written in C++.

While performance will always be the greatest hurdle Python has to negotiate, Risk Focus' Andy Fundinger adds that those issues are increasingly being dealt with.

"Performance in Python or anywhere else is a very peculiar subject. If you put a superior developer and tell them to optimize the living daylights out of something, they're probably going to create something that's pretty fast. If you're talking about true, algorithmic institutional trading, pretty fast isn't anywhere near good enough. But if you're concerned about performance and you're also concerned about development times, and you're not turning milliseconds into millions of dollars, then the story is much more interesting because you can get a Python developer to apply some common sense," he says.

It's important to remember that while Wall Street is currently infatuated with open source languages, that can always change with a language like Matlab coming back in favor again. While Wall Street firms are investing in open source in their data centers, whether or not that continues to extend to development tools, workflow engines, VPN engines, or messaging, remains to be seen, DataArt's Miller adds.

For both humans and computers, language evolves. So to think that Python is here to stay would be wishful and probably foolish. But it's also what's hot right now and it's what trading firms are looking for. But the next "hot language" is, in all likelihood, already out there. Objective-C is getting a lot of play because it's the primary programming language used when writing code for Apple OSX and iOS, while AQR's Austin says that Scala is "growing in popularity these days," as it looks to supplant Java.

But for the time being, on Wall Street anyway, Python is the Shibboleth that seems to be unlocking the most doors, and if you can't speak the language, entry will be difficult.

The Most Valuable Programming Skills to Have on a Resume

1. Ruby on Rails - $109,460
2. Objective-C - $108,225
3. Python - $100,717
4. Java - $94,908
5. C++ - $93,502
6. JavaScript - $91,461
7. C - 90,134
8. R - $90,055
9. C# - $89,074
10. Visual Basic - $85,962
11. SQL - $85,511
12. Perl - $82,513

Salient Points

● Firms are increasingly turning to Python to build their analytics and research platforms, to move large datasets and to rapidly deploy new tools.
● Python has recently taken on greater adoption compared to scripting languages like R, Ruby, Perl and JavaScript.
● Python's largest drawback is its performance, especially when compared to compiled languages like C and C++, which are still largely used for low-latency execution platforms.
● When conducting job searches among the largest banks, most programming positions are Python-related.