Open Source @ IBM
Python Community Update
Jerome Nilmeier
@JNilmeier, nilmeier@us.ibm.com
Spark Technology Center

Susan Malaika, Open Technologies
@sumalaika, malaika@us.ibm.com
March, 2018
Today's Agenda

• What is Python?
• How was Python created?
• Where is the Python community?
• Technical overview
• What's the current status of Python?
• What is next for the Python community?
• Python at IBM
• Let's Get Started
• Call to Action
What is Python?
Python is a widely-used high-level dynamic programming language popular in many domains including web-site creation, devops, and increasingly, in data science, scientific computing, and machine learning. The Python Software Foundation, founded in 2001, promotes and advances the language.

- Easy to learn, runs on most systems, extensive libraries for specific applications, such as data science and artificial intelligence (AI).
  - Website developers use Python e.g., in conjunction with JavaScript
  - Data scientists use Python, e.g. via the PyData platform
  - AI developers use Python. Software written in Python includes Keras, TensorFlow, PyTorch
  - Almost all GPU accelerated code will have a Python API
- Estimated 8 Million Python Developers worldwide
- PyPi (the Python Package Index) is where Python developers can get and share over 116,694 code packages
  https://pypi.python.org/pypi using “pip install”
- Ten of thousands of libraries of packages
  - Popular package examples in data science libraries: NumPy, SciPy, Pandas, Scikit-Learn, Pyspark

Python has a solid claim to being the fastest-growing major programming language.

September 6, 2017
https://stackoverflow.blog/2017/09/06/incredible-growth-python/
How was Python created?

For readability and easy to learn (interpreted language)

- Created by Guido van Rossum and first released in 1991. An interpreted language
- Displacing C and Java at universities - Python is growing as the 1st programming language taught in universities - and for data science, artificial intelligence (AI/ML/DL), and beginner IoT.
- Python is now taught to more primary school students than French. http://www.information-age.com/python-overtakes-french-most-popular-language-taught-primary-schools-123460073/
- June 2017 was the first month that Python was the most visited tag on Stack Overflow within high-income nations… It has grown by 2.5-fold since 2012. https://stackoverflow.blog/2017/09/06/incredible-growth-python/

CPython, the reference implementation of Python, is open source software and has a community-based development model.
CPython is managed by the non-profit Python Software Foundation https://www.python.org/psf-landing/
Guido van Rossum is the President.

https://cacm.acm.org/blogs/blog-cacm/176450-python-is-now-the-most-popular-introductory-teaching-language-at-top-u-s-universities/fulltext
Where is the Python community?
Developers, Data Scientists, Academia

- Enterprise Developers
  - Python has earned its way into the enterprise for all kinds of scripting and website building
- Data Scientists/AI Developers
  - Python is fast becoming the tool of choice to data scientists and AI developers
    - R is another popular language for Data Scientists
- Academia
  - Python is increasingly taught at universities
  - Academic Research code is largely in python

“In a nutshell, Python is better for data manipulation and repeated tasks, while R is good for ad hoc analysis and exploring datasets.”
From “If you want to upgrade your data analysis skills, which programming language should you learn?”
Where is the Python Community?

The Python Community can be found at meetups and conferences

The Python Software Foundation (PSF) promotes, protects, and advances the Python programming language, and to support and facilitate the growth of a diverse and international community of Python programmers. PSF runs PyCon. https://www.python.org/psf/

PyData is an educational program of NumFOCUS, that provides a forum for the international community of users and developers of data analysis tools. PyData runs PyData Conferences. It also manages the open data science stack https://pydata.org/downloads.html
Python: Technical Overview

Python Packages around Data Science: PyData

Sample PyData packages:

- **NumPy**: (Commits: 15980, Contributors: 522) fundamental package for scientific computing with Python
- **SciPy**: (Commits: 17213, Contributors: 489) SciPy is a library of software for engineering and science.
- **Pandas**: (Commits: 15089, Contributors: 762) Data wrangling tools
- **Matplotlib**: Matplotlib (Commits: 21754, Contributors: 588) Python 2D plotting library
- **Scikit-learn**: (Commits: 21793, Contributors: 842) Machine learning and image processing in Python
- **Theano**: (Commits: 25870, Contributors: 300) Machine learning, integrates with NumPy

Most of these packages can be downloaded from https://pydata.org/downloads.html

NumFOCUS (a non-profit organization) sponsors PyData activities https://www.numfocus.org/
Python: Technical Overview

Cconda can be used to manage Python packages & more

Conda [https://conda.io/docs/](https://conda.io/docs/) can be used to manage package dependencies & more

In addition to pip install (python index package)
Anaconda provides conda install for packages in the conda index… the dependencies are sometimes better managed in conda
What is the current status of Python?

Many Corporate SDKs & Curated Packages

- IBM [https://console.bluemix.net/docs/runtimes/python/getting-started.html](https://console.bluemix.net/docs/runtimes/python/getting-started.html)
- IBM Watson SDK
  - [https://github.com/watson-developer-cloud/python-sdk](https://github.com/watson-developer-cloud/python-sdk) (32 contributors, 710 commits)
- Anaconda curated data science packages for Python
  - [https://docs.anaconda.com/anaconda/packages/pkg-docs](https://docs.anaconda.com/anaconda/packages/pkg-docs)
- More curated Python packages here:
  - [https://pydata.org/downloads.html](https://pydata.org/downloads.html)

Latest Python Language Release 3.6.3 [https://www.python.org/downloads/](https://www.python.org/downloads/)
What's next for the Python community?
More growth! Both Enterprise, Data/AI Developers, and Academia

“With the raise of machine learning, AI, data science.. it's no surprise to me. There was none Python programmer in my company before except for me. Now for our new R&D, BI, the main/only language we are using is Python.” – A developer

https://stackoverflow.blog/2017/09/06/incredible-growth-python/
Python at IBM

Tools, Interconnects and Courses

- IBM Technology Contributions
  - Pixiedust Magic for your Python Notebook
  - Apache Toree (incubating): the middleman between an application and a Spark cluster
    https://toree.apache.org/
  - Brunel Visualizations
    https://github.com/Brunel-Visualization/Brunel
  - Python Connectors for Loading and Saving Data
    https://datascience.ibm.com/docs/content/analyze-data/python_load.html
  - Streaming Analytics & Python
  - Deploy Python Functions as REST API in Bluemix
    https://datascience.ibm.com/blog/deploy-your-python-functions-as-a-rest-api/
  - IBM Python Journeys
- IBM Education and Getting Started
  - Python courses at Cognitive Class
    https://cognitiveclass.ai/courses/
  - Learn about Python on DeveloperWorks
    https://www.ibm.com/developerworks/topics/python
  - Python QuickStart on Bluemix
    https://console.bluemix.net/docs/runtimes/python/getting-started.html#gettingStarted

Two IBM Distinguished Engineers expert in Python

Jean-Francois Puget
@jfpuget

David Taieb
@DTAIEB55
Python: Let's Get Started
Learn, Attend, Participate

- Learn Python and its packages
  - IBM step by step Education for Python
  - IBM Python journeys
  - Read the many Python books – e.g., from O'Reilly
    - https://wiki.python.org/moin/PythonBooks
  - All Time PyPi Rankings
    - http://pypi-ranking.info/alltime

- Attend conferences and meetups
  - PyCON
    - https://us.pycon.org/2018/ organized by Python Software Foundation
  - PyData
    - https://pydata.org/ organized by NumFOCUS - More than 7 conferences a year in various countries
  - Meetups
    - https://www.meetup.com/topics/python/

- Participate in Community Organizations
  - Software Foundation
  - NumFOCUS
    - https://www.numfocus.org/ (IBM is a partner)
Call to Action

• Get involved – Learn about Python at IBM:
  • see https://www.ibm.com/developerworks/topics/python

• Meet up communities – Most cities have Python Meetup groups:
  • see https://www.meetup.com/topics/python/
  • see PyData meetups https://www.meetup.com/pro/pydata/ (NUMFocus Sponsored)

• Conferences – Look out for events in these series:
  • see PyData https://pydata.org/
  • see PyCon https://us.pycon.org/

• Membership links
  • Join the NUMFocus Community https://www.numfocus.org/community/

• Contributing links
  • Join PyCon 2018 Development Sprints
    • Participate in sprints May 14–17, 2018 - Monday through Thursday https://us.pycon.org/2018/
    • View the intro to Python 2017 sprinting link
      • see https://github.com/chalmerlowe/intro_to_sprinting
Open Source @ IBM
Python Community Update

Jerome Nilmeier
@JNilmeier, nilmeier@us.ibm.com
Spark Technology Center

Susan Malaika, Open Technologies
@sumalaika, malaika@us.ibm.com
March, 2018