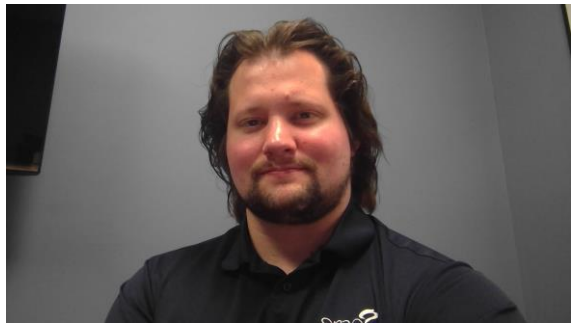
Exploring and Using the Python Ecosystem

Adam J. Cook, Chair of SME Chapter 112

FOR PUBLIC RELEASE

About the Presenter

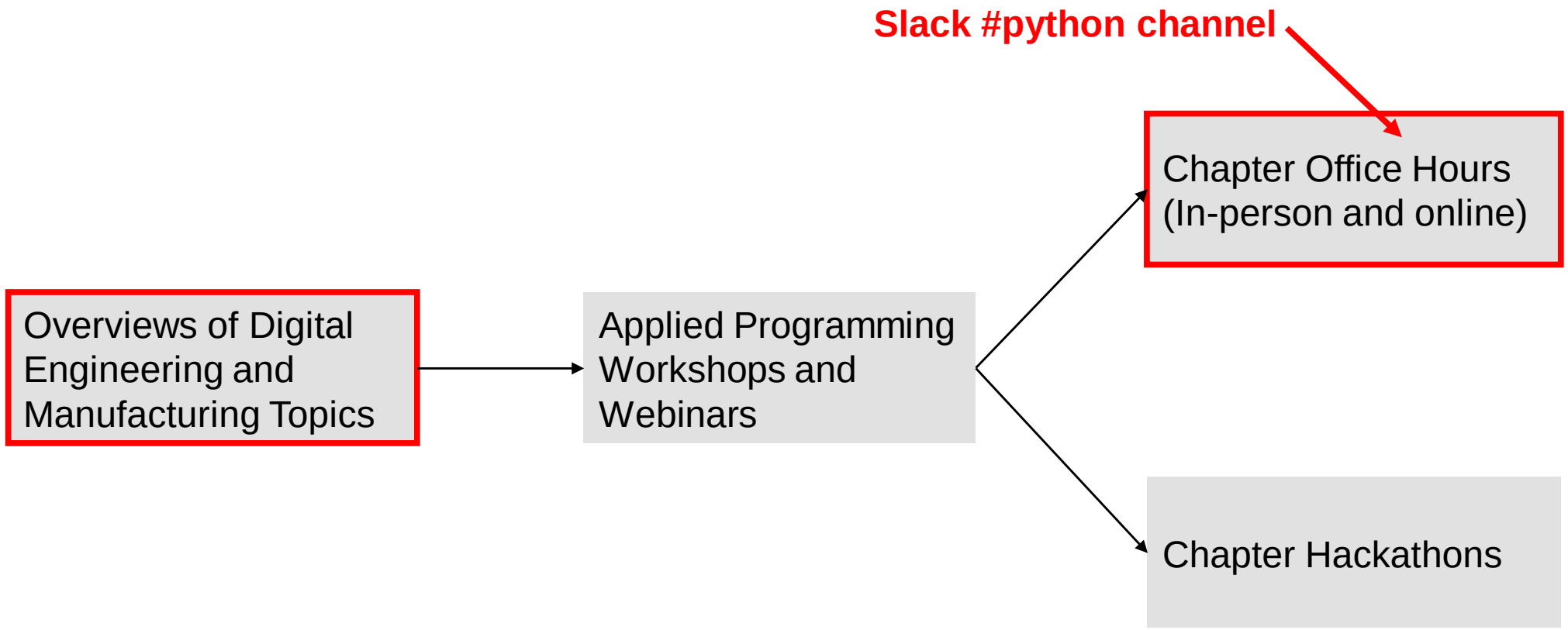


- Adam Cook
- B.S. in Mechanical Engineering from Purdue University West Lafayette.
- Chief Technical Officer of Alliedstrand in Hammond, Indiana.
- Chair of SME Chapter 112 (Northwest Indiana and South Chicago).
- Embedded systems engineering, custom automation systems, industrial software.
- Lives in Chicago.
- Contact me at adam.j.cook@alliedstrand.com.

FOR PUBLIC RELEASE

Chapter “Digital Initiative”

Slides and code:
http://bit.ly/2uzCQqR



FOR PUBLIC RELEASE

What is Python?

Slides and code:
http://bit.ly/2uzCQqR

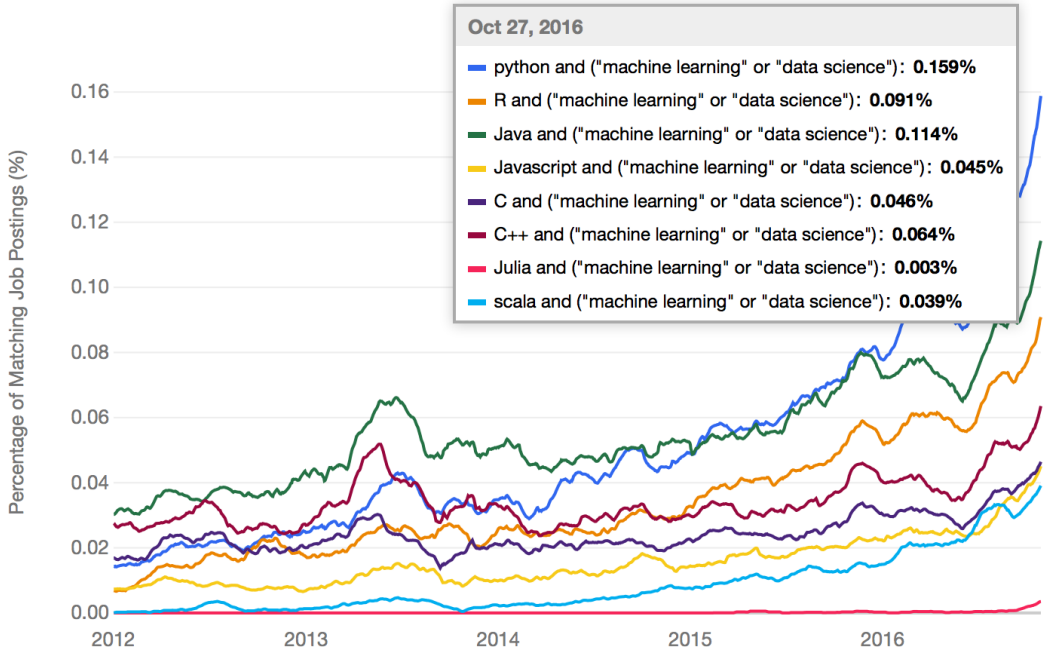
- High-level programming language.
- Free and open-source.
- Cross-platform.
- **Extensive standard library.**
- Designed to be highly readable, explicit and **productive**.
- Proven to be quite versatile (and popular).

FOR PUBLIC RELEASE

Why use Python in Manufacturing?

Slides and code:
http://bit.ly/2uzCQqR

- Python is fast becoming one of the most popular languages in data analytics and machine learning. Coincidentally, manufacturing processes are producing more **valuable** data than ever!
  Source:
  https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Language_Is_Best_For_Machine_Learning_And_Data_Science?lang=en



FOR PUBLIC RELEASE

Today's Agenda

Slides and code:
http://bit.ly/2uzCQqR

- Look at PyPI (also called pip) and how it can help you.
- Brief overview of the Anaconda Python distribution and why you might want to use it (for data science, you should really just use it).
- **Super** high-level overview of data science/analytics. This is important. Data can be **tricky** and **deceptive**.
- Small recap of where we are in these Python webinars.

FOR PUBLIC RELEASE

Caveats and Warnings

Slides and code:
http://bit.ly/2uzCQqR

- This event assumes you are a novice. If you have keep in mind that we will be watering down a bunch.
- Programming and data analytics is challenging – the following presentation will not make you into an expert. **Practice and read code.**
- For data analytics and machine learning applications, in particular, knowing Python is **not enough**.
- We are starting to get **advanced** now. Application architecture patterns are difficult. Data problems are very deep and a very active area of research. The industry is **extremely** fluid. **Do not try to memorize everything!**
- We are going to talk today at a **high-level**. Let us know if you want to break down things into separate webinars.
- **Think about what kind of actual applications you want to build and let us know.** After a couple of projects, things will start clicking together.

FOR PUBLIC RELEASE

Demonstration

Slides and code:
http://bit.ly/2uzCQqR

Let's take a look at pip!
(we will use the code from http://bit.ly/2w62Sk4)

```
> pip install <package name>
```

FOR PUBLIC RELEASE

Other Resources

Slides and code:
http://bit.ly/2uzCQqR

There is another great resource called [Awesome Python](#).

FOR PUBLIC RELEASE

Python vs. Anaconda

Slides and code:
http://bit.ly/2uzCQqR

Anaconda
(think of it as “Python Plus”)

Contains the following out-of-the-box:

- SciPy
- Jupyter
- Other Continuum tools

Python

FOR PUBLIC RELEASE

SciPy, NumPy, Matplotlib, Jupyter…and bears, oh my!

- SciPy – umbrella package containing NumPy, Matplotlib and SymPy.
  - NumPy – provides sophisticated N-dimensional array handling
  - Matplotlib – provides powerful 2D plotting functionality for data and result visualizations
  - SymPy – provides symbolic mathematics functionality (computer algebra system)
- Jupyter – interactive, web browser-based “notebook” which allows you to share Python code, run experiments and capture results.

FOR PUBLIC RELEASE

Demonstration

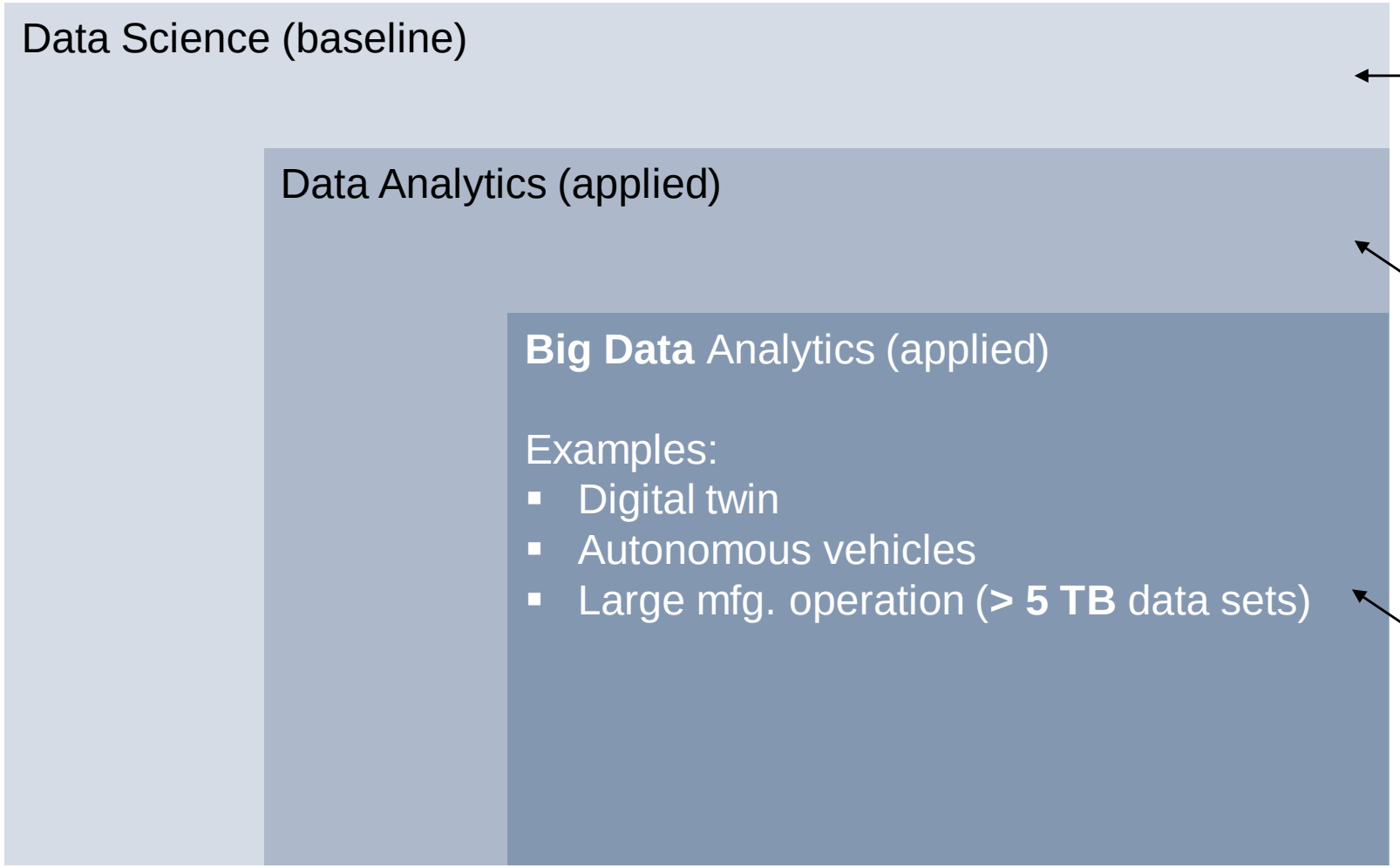
Slides and code:
http://bit.ly/2uzCQqR

Let’s take a look at Jupyter!
(we will use the code from http://bit.ly/2wlz56p)

FOR PUBLIC RELEASE

Word Soup

Slides and code:
http://bit.ly/2uzCQqR



- Calculus
- Statistics
- SQL
- Unstructured data
- Machine learning

- Python
- PostgreSQL
- Algorithm design
- Data visualization
- Data wrangling

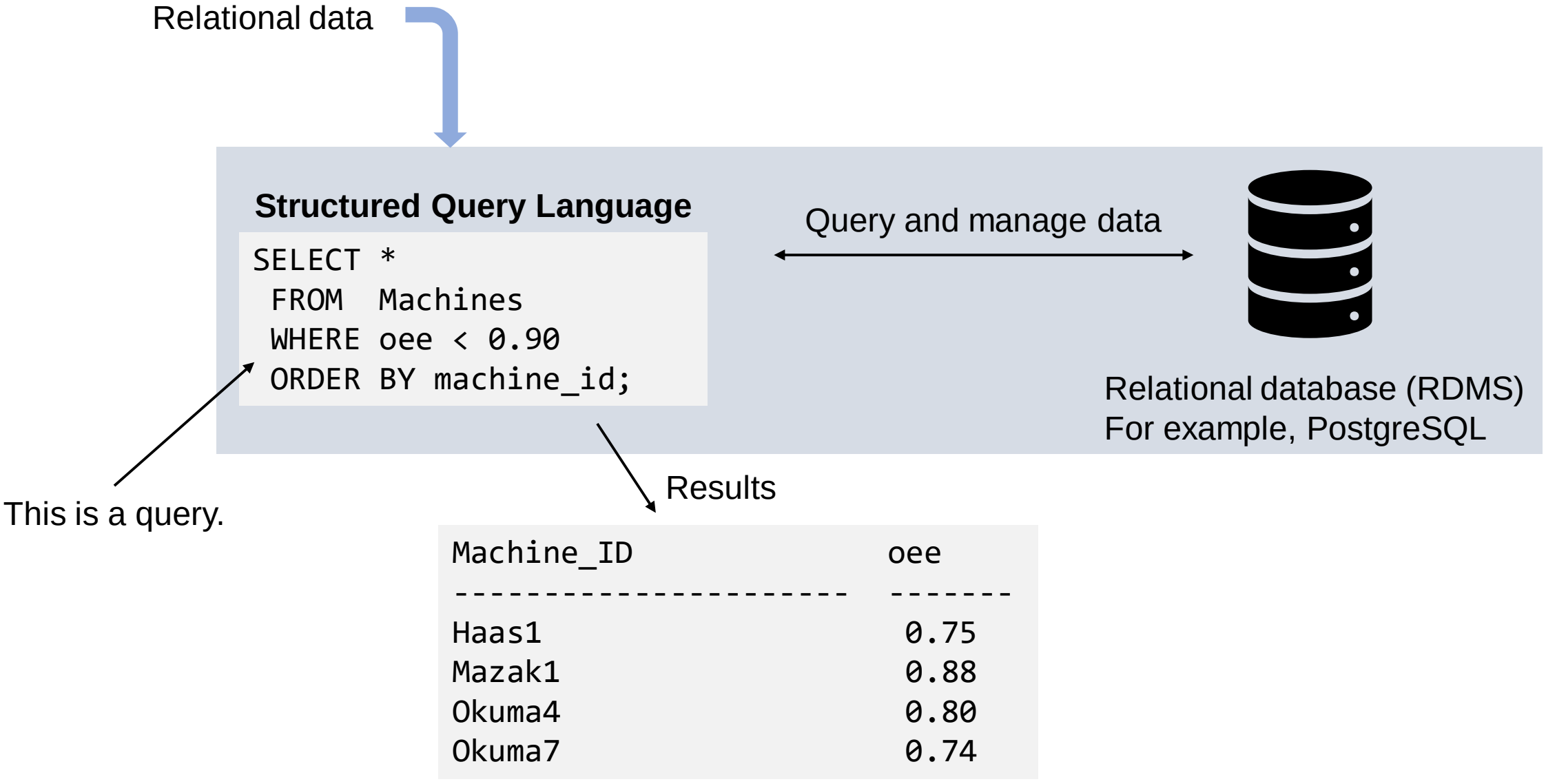
- Java/C#/C++/JavaScript
- Hadoop
- Computational parallelism (Python is **not** good here)
- MapReduce
- Distributed systems

Big Data is **hard** (really hard)! **Make sure you need it!**

FOR PUBLIC RELEASE

SQL

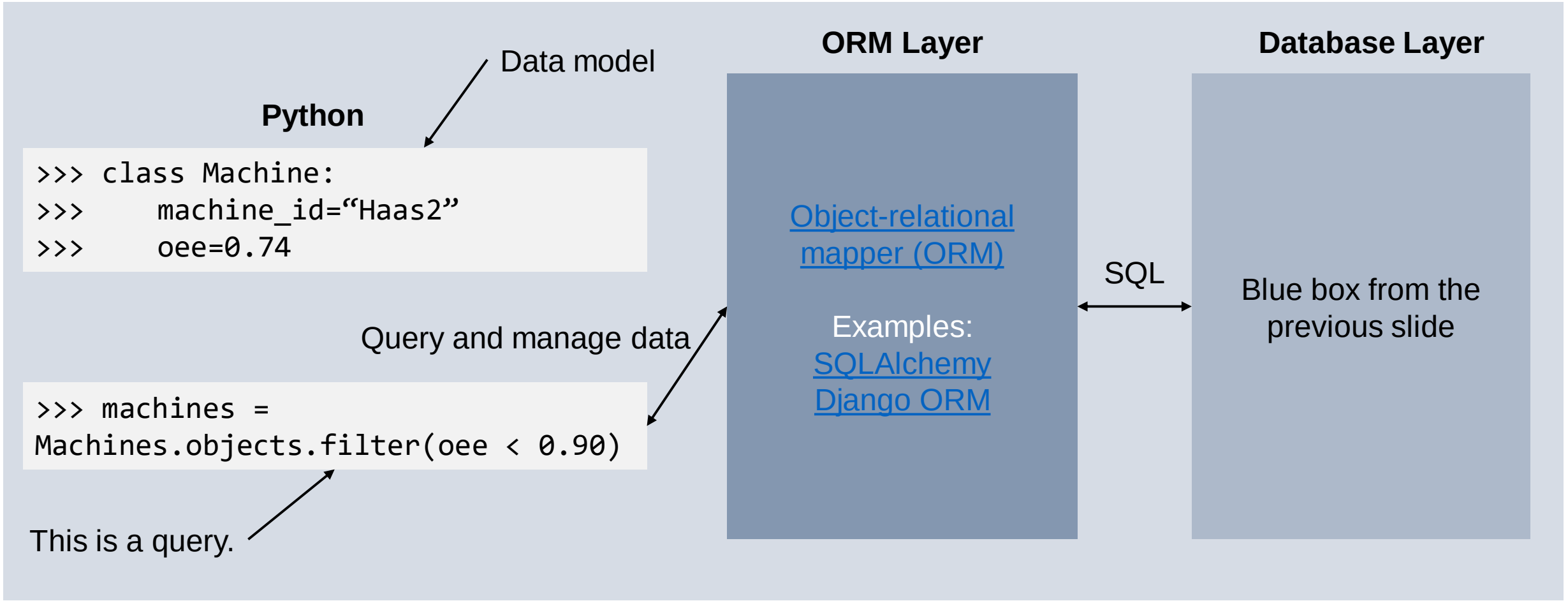
Slides and code:
http://bit.ly/2uzCQqR



FOR PUBLIC RELEASE

Python “hiding” SQL

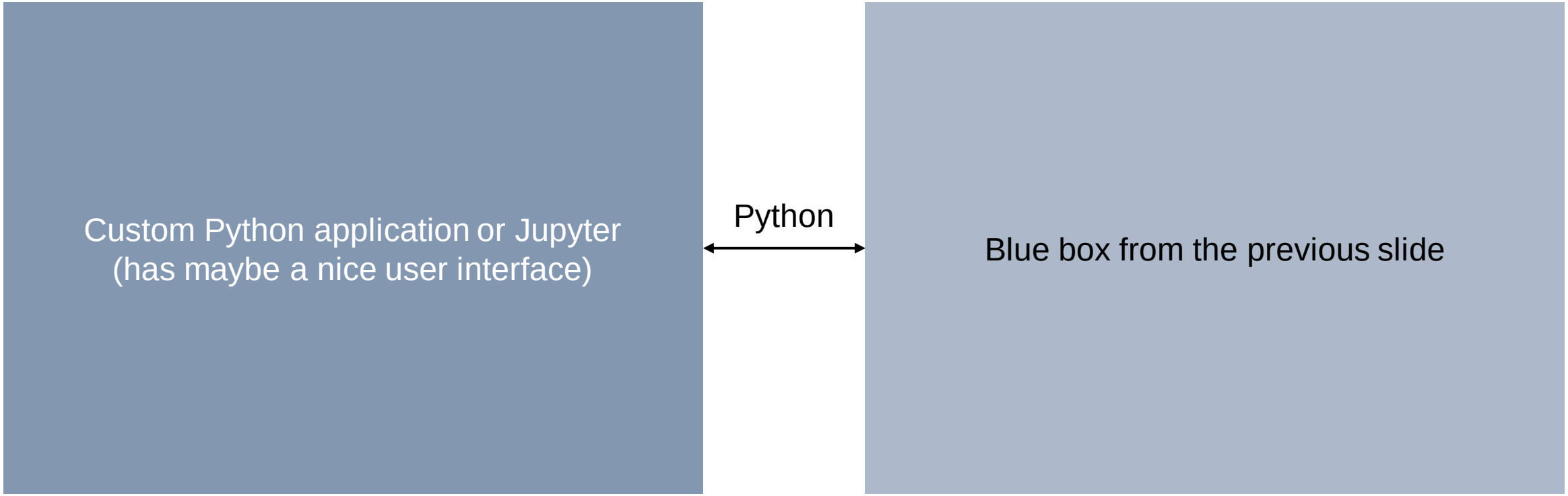
Slides and code:
http://bit.ly/2uzCQqR



FOR PUBLIC RELEASE

Python “hiding” everything

Slides and code:
http://bit.ly/2uzCQqR

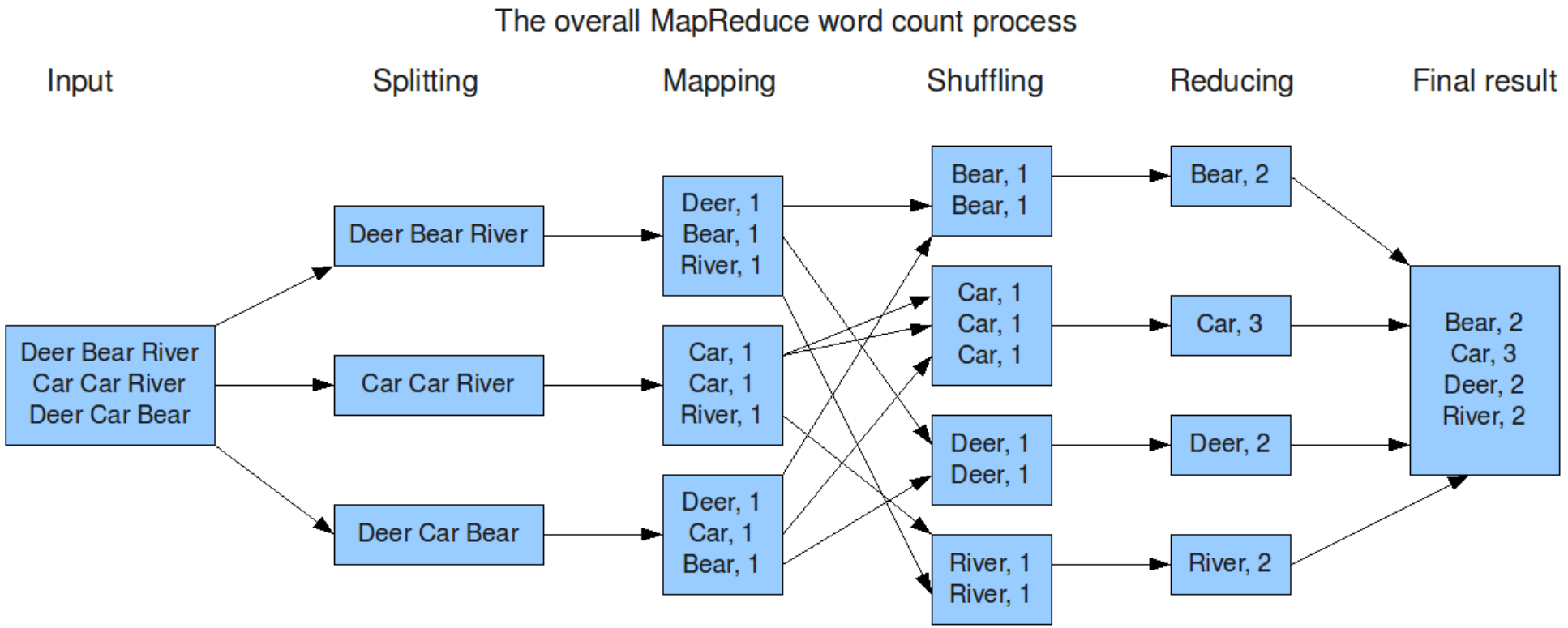


Key word: **Abstractions!**

FOR PUBLIC RELEASE

MapReduce

Slides and code:
http://bit.ly/2uzCQqR



Source: http://datascienceguide.github.io/

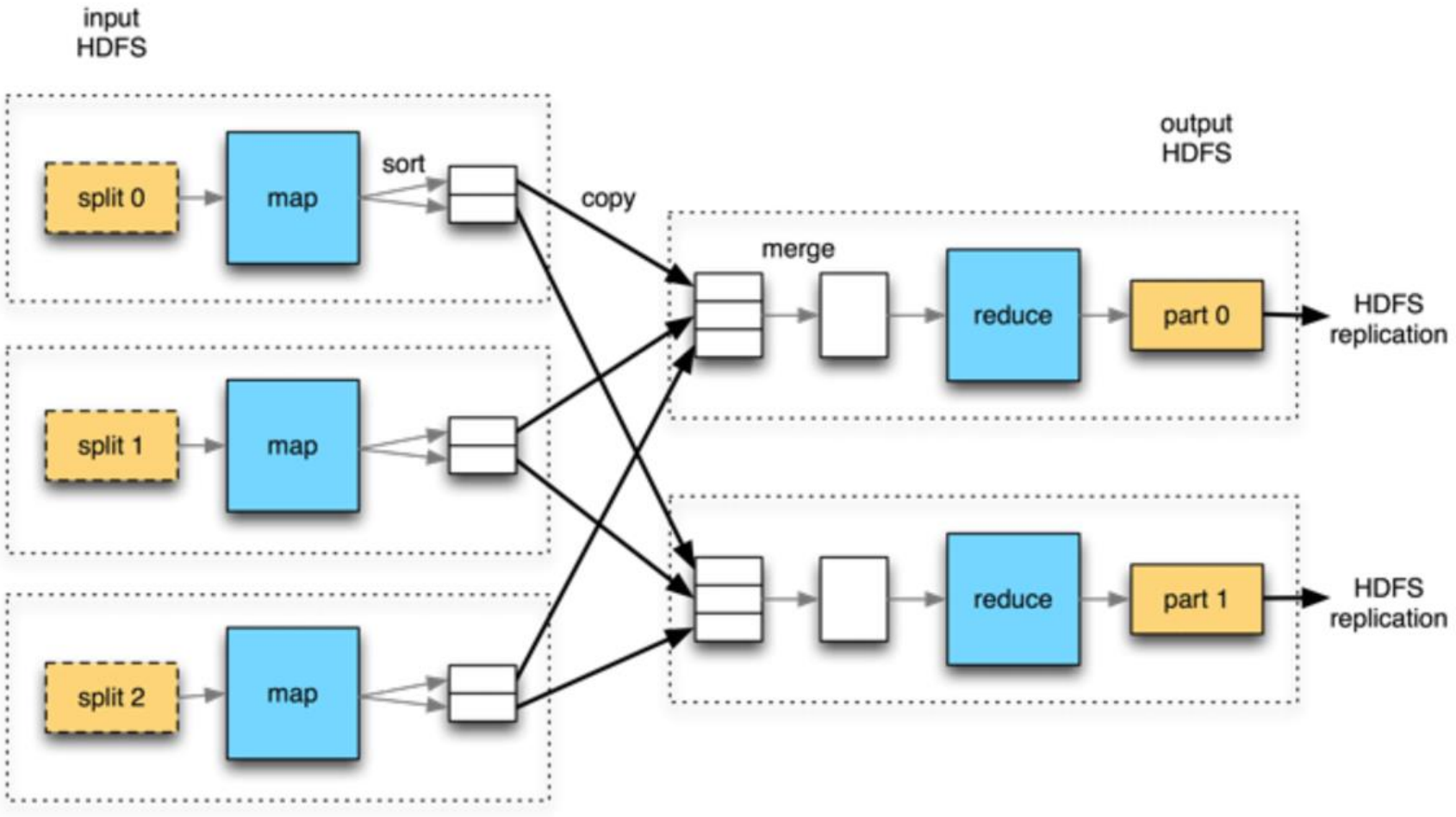
FOR PUBLIC RELEASE

Hadoop

Slides and code:
http://bit.ly/2uzCQqR

Hadoop consists of two (2) parts:

1. Hadoop Distributed File System (HDFS)
2. Processing Part (MapReduce)

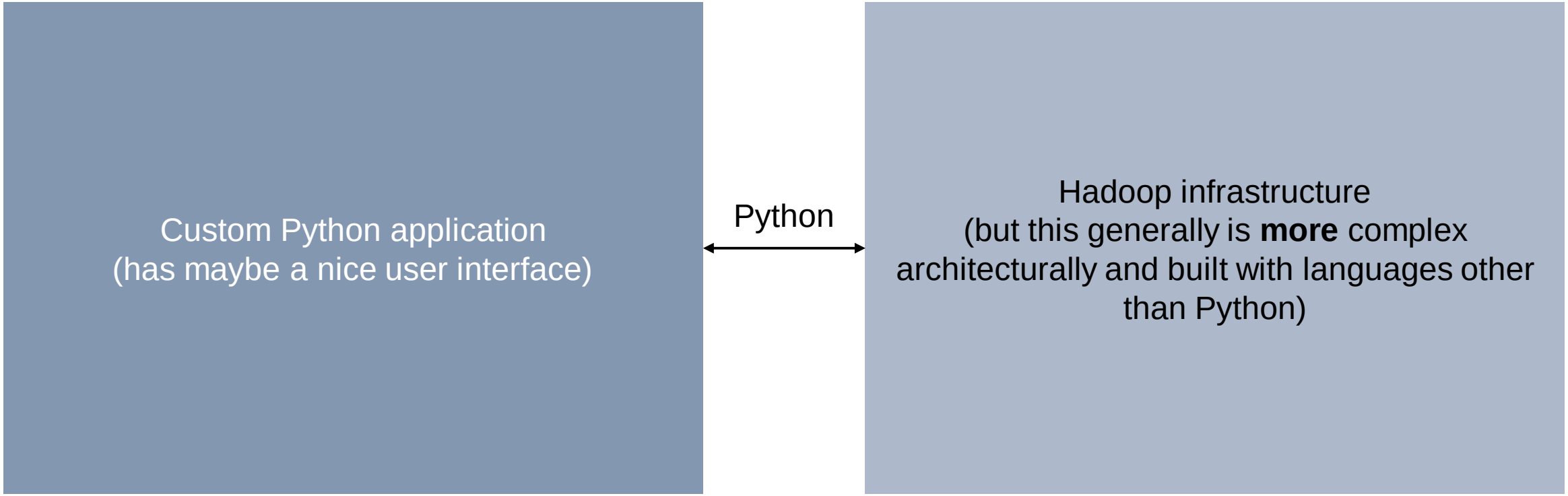


Source: http://ubm.io/2vipYqj

FOR PUBLIC RELEASE

Big Data and Python

Slides and code:
http://bit.ly/2uzCQqR



FOR PUBLIC RELEASE

Slides and code:
http://bit.ly/2uzCQqR

If you are not sure, then you do not need Big Data.
(just use PostgreSQL)

FOR PUBLIC RELEASE

Slides and code:
http://bit.ly/2uzCQqR

What does this all have to do with Python?

FOR PUBLIC RELEASE

Data Sanity

Slides and code:
http://bit.ly/2uzCQqR

- Data can (and it will, at times) lie to you.
- Think about data delivery – particularly if it is arriving from human sources.
- Data anomalies will occur. How do you address them?
- Are you collecting the right data and, more importantly, enough **relevant** data?
- Careful of biases (i.e. confirmation bias). **Be scientific!**

FOR PUBLIC RELEASE

Resources

Slides and code:
http://bit.ly/2uzCQqR

Books

- Raschka, S. (2015). *Python machine learning: unlock deeper insights into machine learning with this vital guide to cutting-edge predictive analytics.* Birmingham (U.K.): Packt Publishing.
- VanderPlas, J. (2017). *Python data science handbook: Essential tools for working with data.* Sebastopol, CA: O'Reilly.
- Klein, P. N. (2013). *Coding the matrix: linear algebra through applications to computer science.* Newton, MA: Newtonian Press.

Videos

- Sarah Guido - Hands-on Data Analysis with Python - PyCon 2015
- Jake VanderPlas - Machine Learning with Scikit-Learn (I) - PyCon 2015
- Olivier Grisel - Machine Learning with Scikit-Learn (II) - PyCon 2015

FOR PUBLIC RELEASE

Online Course

Slides and code:
http://bit.ly/2uzCQqR

http://bit.ly/2danP4n
(Applied Data Science with Python Specialization – University of Michigan)

FOR PUBLIC RELEASE

Next Webinar

Slides and code:
http://bit.ly/2uzCQqR

Machine Learning with scikit-learn (mostly)

FOR PUBLIC RELEASE

Where can I get this slide deck and code?

Slides and code:
http://bit.ly/2uzCQqR

http://bit.ly/2uzCQqR
(actually, go ahead and bookmark this link – this web page will be updated constantly with new content)

FOR PUBLIC RELEASE

Python-ish Feedback Received So Far

Slides and code:
http://bit.ly/2uzCQqR

Cloud Computing

M2M

Big Data

Regular Expressions

Statistics

Data Analytics

IIoT

Machine Vision

Data Visualization

Siemens NX Python API

Machine Learning

CAE

Deep Learning

Realtime (Streaming) Data

Computational Geometry

Embedded Systems

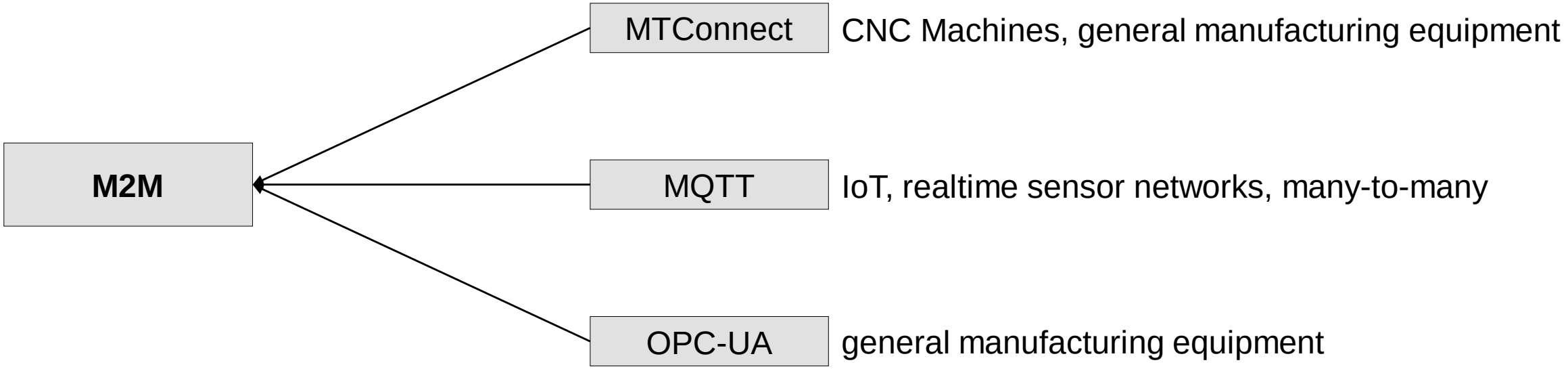
Robotics

Linear Algebra

FOR PUBLIC RELEASE

Deeper Look at Machine-to-Machine (M2M)

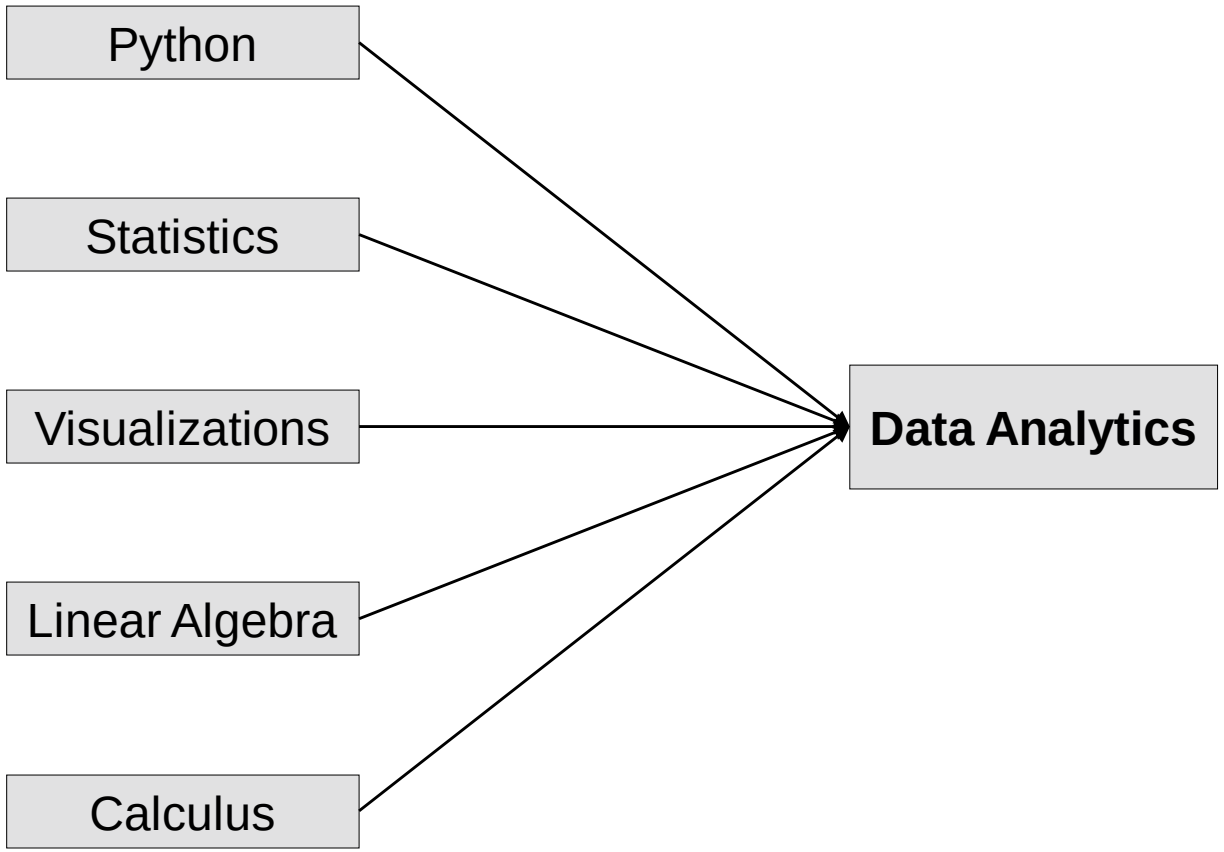
Slides and code:
http://bit.ly/2uzCQqR



FOR PUBLIC RELEASE

Another Example of a Basic Learning Path

Slides and code:
http://bit.ly/2uzCQqR



FOR PUBLIC RELEASE

Call for Feedback

Slides and code:
http://bit.ly/2uzCQqR

Please provide us with feedback!

Want to Keep the Conversation Going?

Slides and code:
http://bit.ly/2uzCQqR

We have a Slack channel! Send me an invite request at my e-mail address. (we are working on an automatic invite link)

FOR PUBLIC RELEASE

Thank you!

Slides and code:
http://bit.ly/2uzCQqR

Thanks for attending!

Special thanks to our hosting partner – GreenCow Coworking. Check them out at greencow.space!

Suggestions? Feedback? Comments? Complaints? Contact us below!

Adam J. Cook

Chief Technical Officer of Alliedstrand

Chair of SME Chapter 112

adam.j.cook@alliedstrand.com

https://linkedin.com/in/adam-j-cook

https://github.com/adamjcook

SME

www.sme.org

https://facebook.com/SMEmfg

https://twitter.com/SME_MFG

https://linkedin.com/company/sme

SME Chapter 112

Serving Northwest Indiana and Chicagoland

https://facebook.com/sme112

https://linkedin.com/company/sme112

https://github.com/sme112

FOR PUBLIC RELEASE