LECTURER: TAI LE QUY

PROGRAMMING WITH PYTHON

TOPIC OUTLINE

TOPIC OUTLINE

Version Control

6

UNIT 4

PYTHON IMPORTANT LIBRARIES

STUDY GOALS

- Know how to use most popular Python libraries
- Discover how to perform scientific calculations using NumPy, Pandas, and SciPy libraries
- Understand how to use Numba to speed up programs
- Visualize different data types using Matplotlib, Seaborn and Bokeh libraries
- Learn to access MySQL databases with SQLAlchemy and PyMySQL client driver

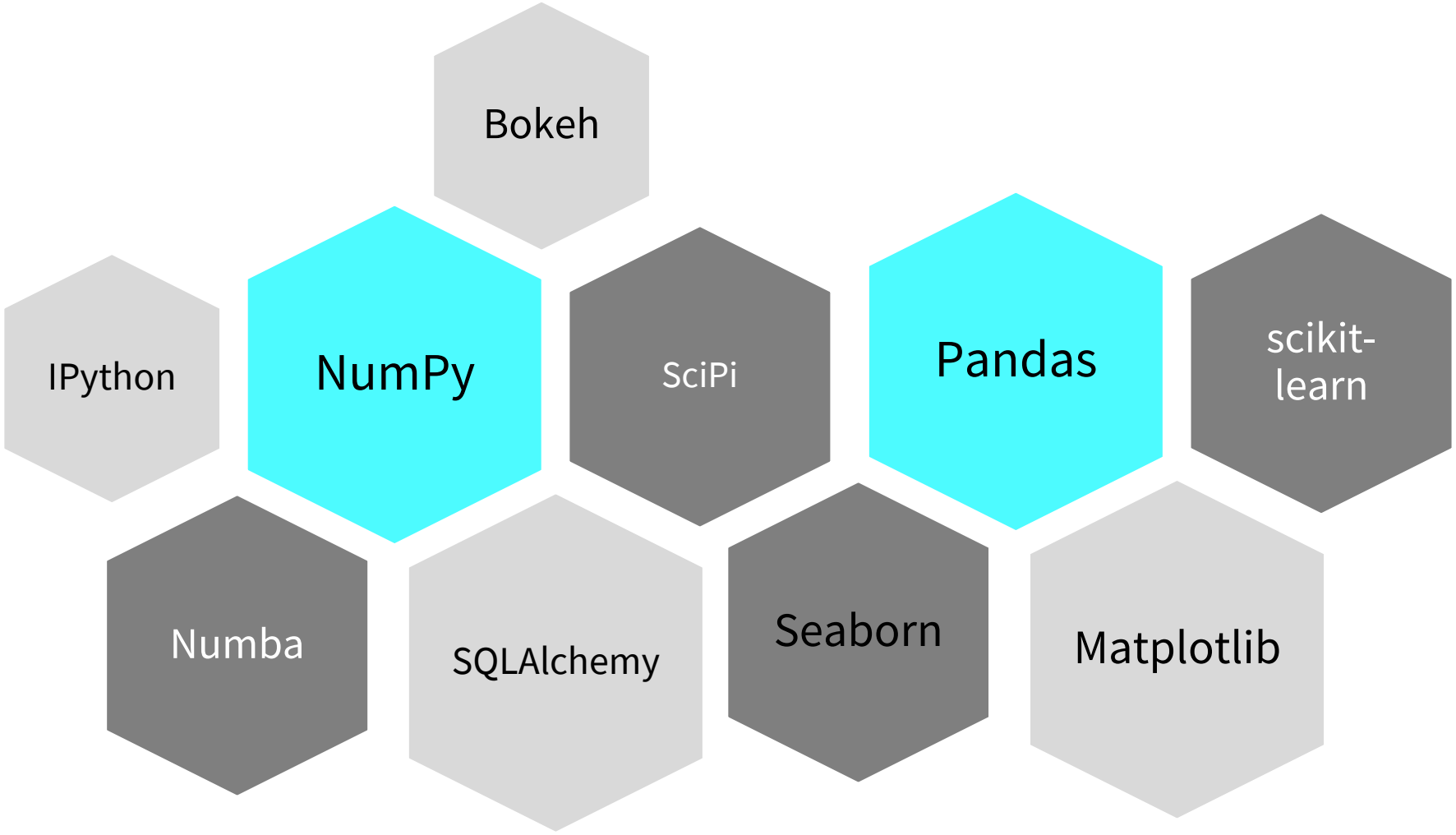
EXPLAIN SIMPLY

1. Why are Python standard libraries important?

2. Which libraries are most used for scientific computing and projects?

3. Which libraries are most used for data visualization?

PYTHON LIBRARIES



Source of the text: McKinney, 2017.

NumPy: most used for numerical computations

Widely used

- Most important library of Python eco-system

High performance

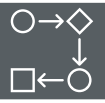
- Efficient even with high dimensional data

NumPy Array

- Wide range computations of arrays and matrices

Pandas: main processing tool for data analysis and visualization

Data processing



- Easy and efficient handling of large data

Data visualization

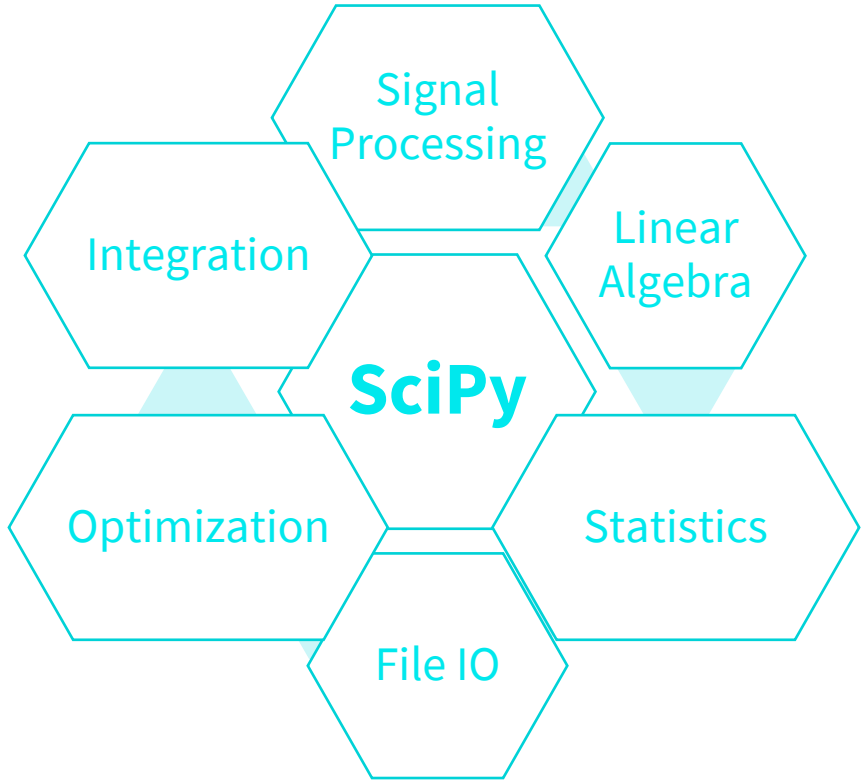
- Tools for visualizing heterogeneous types of data

Series and timeframes

- One or two dimensional data structures used in data analytics machine learning

SciPy: higher level scientific algorithms

- Built on top of NumPy
- Multiple modules for advanced computations
- High speed and widely used



High volume and multi-dimensional data demand speedy programs

CYTHON

- C-like performance
- Extension modules

PYPY

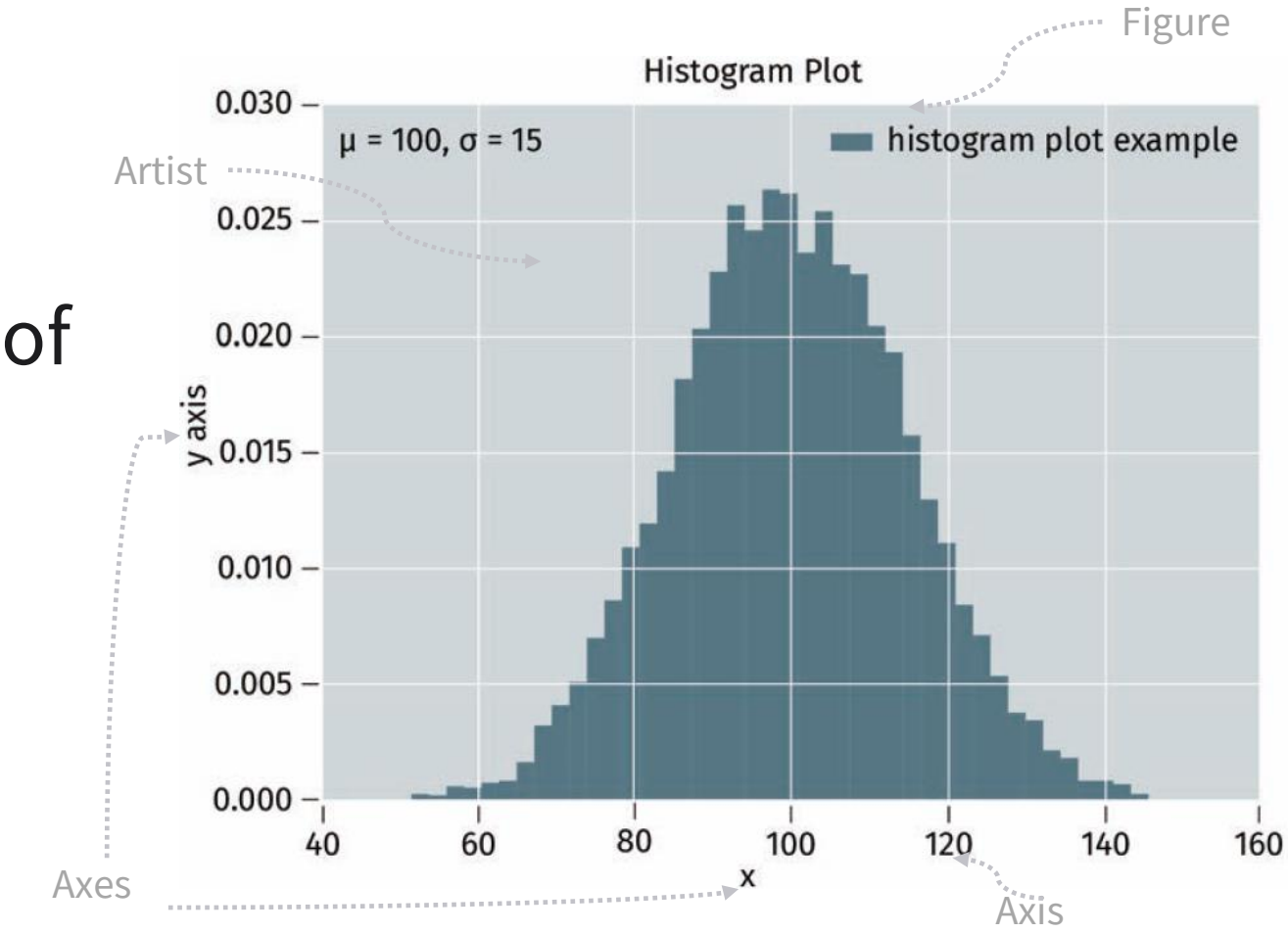
- Built using R
- Compile versions foster execution speed-up

Numba

- Parallel code
- Used most frequently
- Efficient with NymPy and functions

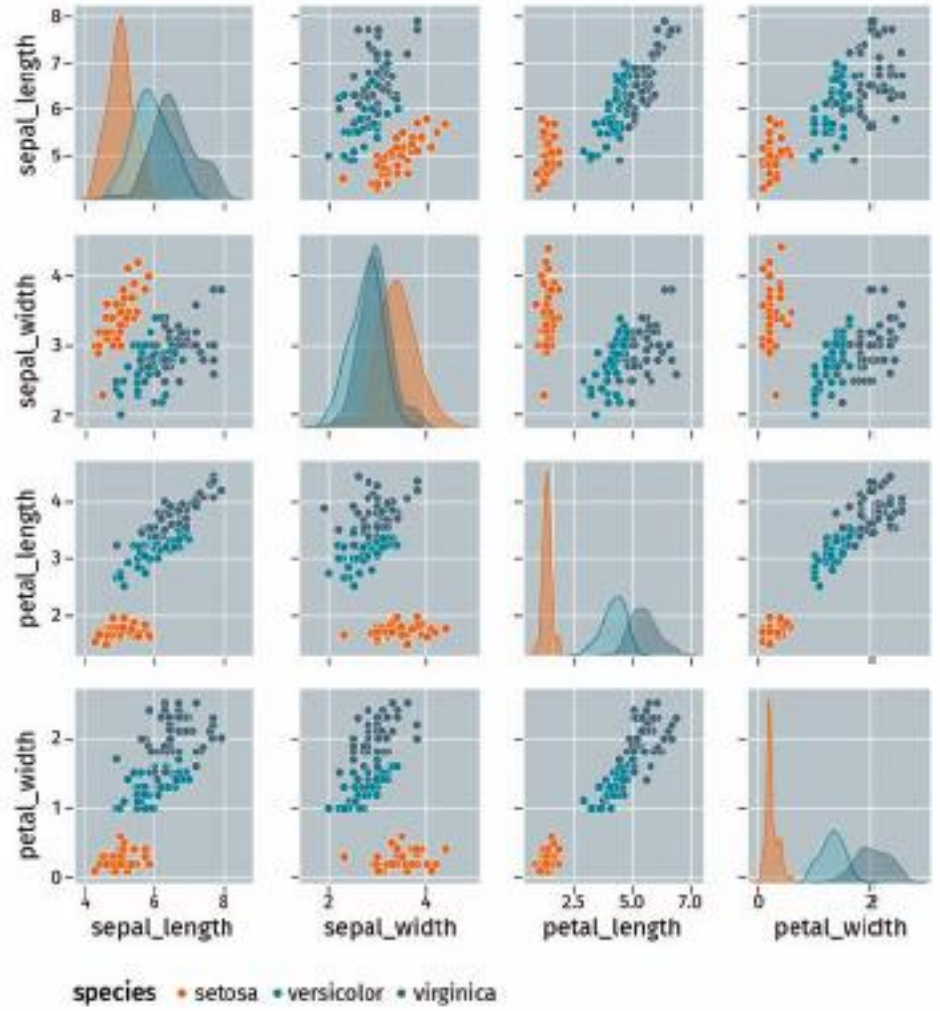
Matplotlib: most used library for plotting

- Wide range graphs
- Matplotlib figure consists of
  - Figure
  - Axes, axis
  - Artist



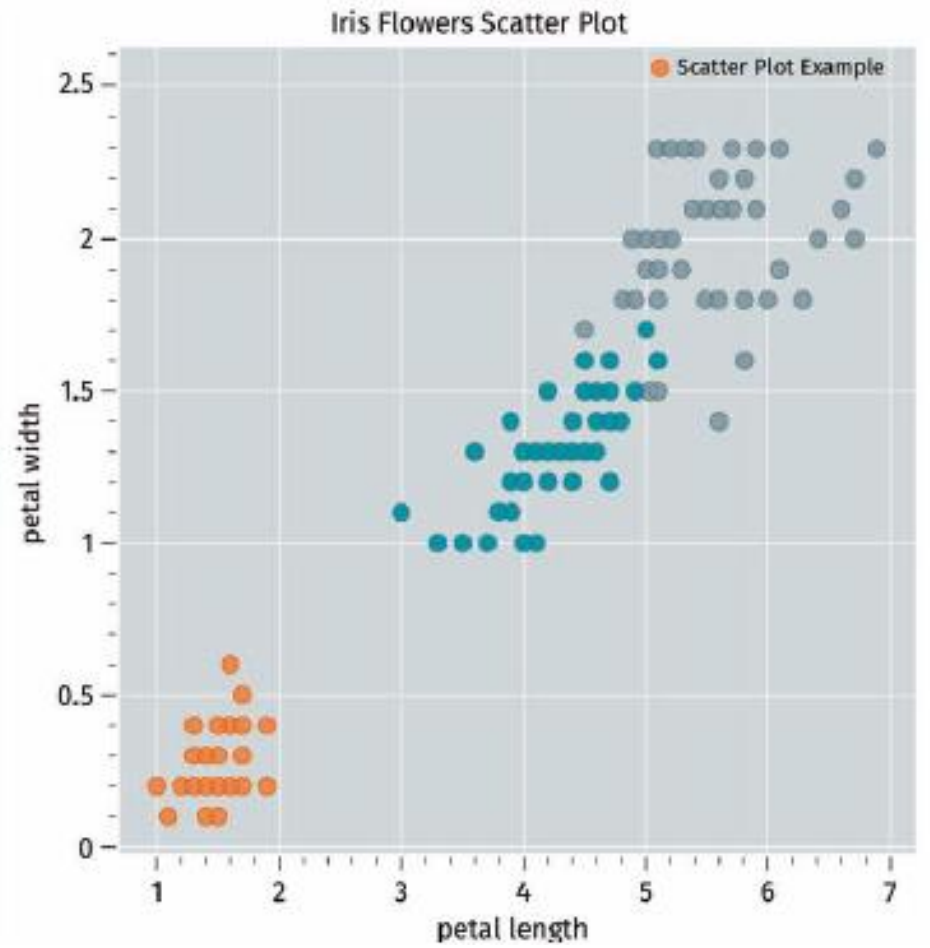
Seaborn: statistical data visualization

- Built on top of matplotlib
- Effective with statistical models, time-series, pandas and more



Bokeh: web browser's interactive data visualization

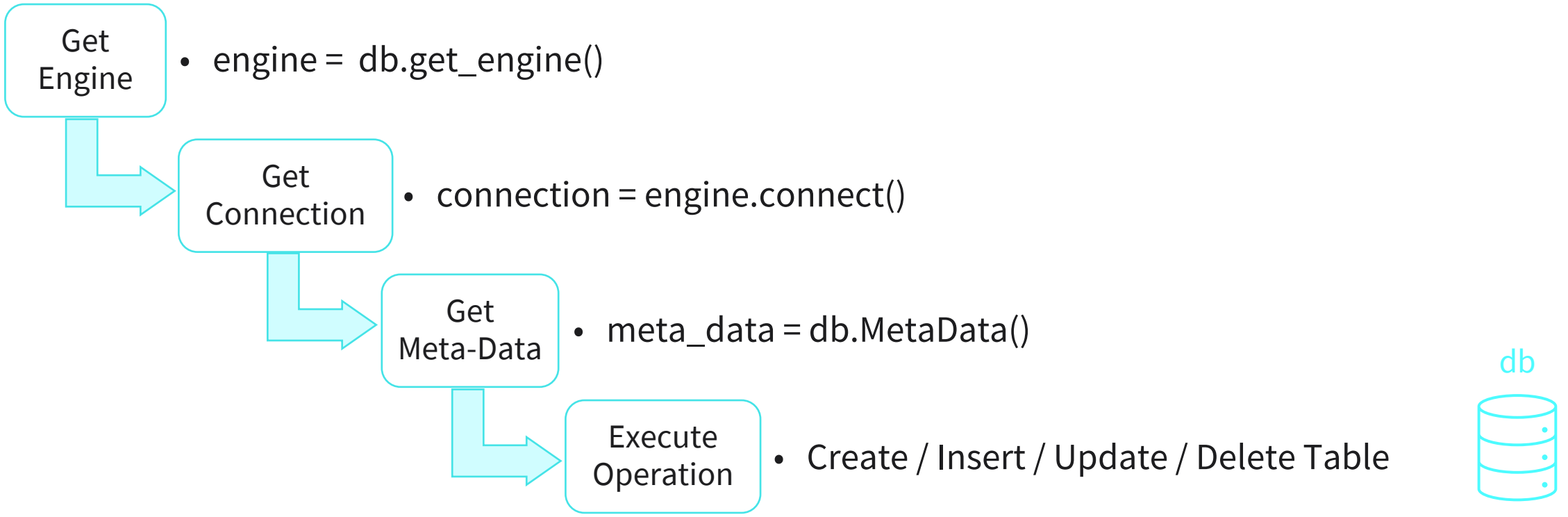
- Produces JSON files
  - Data structures and objects
  - Standard data interchange
  - Data transmission between web application and server



SQLAlchemy: open source SQL toolkit and object-relational mapper

- Object relational model allows use of SQL
- Relational connection reduces amount of code

Engine and **connection** objects: key components of SQLAlchemy



REVIEW STUDY GOALS

- Know how to use most popular Python libraries
- Discover how to perform scientific calculations using NumPy, Pandas, and SciPy libraries
- Understand how to use Numba to speed up programs
- Visualize different data types using Matplotlib, Seaborn and Bokeh libraries
- Learn to access MySQL databases with SQLAlchemy and PyMySQL client driver

UNIT 4

TRANSFER TASK

TRANSFER TASK

Group Work: Machine Learning Algorithms through scikit-learn

READ Research the scikit-learn library

IDENTIFY Identify most fundamental algorithms used for machine learning

EXPLAIN Aim to group together similar (i.e., same family of) algorithms

DISCUSS Discuss your findings and compare them with the other groups

TRANSFER TASK
PRESENTATION OF RESULTS

Please present your results.

The results will be discussed in plenary.

1. NumPy library is defined as …
   a) … a statistical and fundamental package for scientific computing in Python
   b) … a numerical and fundamental package for computing simulation in Python
   c) … a numerical and fundamental package for scientific computing in Python
   d) … a numerical and fundamental package for scientific validation in Python

2. Bokeh library is defined as…
   a) … an interactive visualization library for windows-based application
   b) … an interactive visualization library for web-based application
   c) … an interactive visualization library for mobile-based application
   d) … an interactive visual library for linux-based application

3. Which are the most popular visualization libraries in Python?
   a) Matplotlib, Seaborn, Bokeh
   b) Plotlib, Seaborn, Bokeh
   c) Matplotlib, SeaVisual, Bokeh
   d) Matplotlib2D, Seaborn, Bokeh3D

LIST OF SOURCES

McKinney, W. (2017). *Python for data analysis* (2nd ed.). O'Reilly.