



Networking
Academy

Big Data & Analytics



Content of the course

**Preperation for the
final exam**

Hands On

Competencies

Section 1.0
Introduction

Section 1.1
Value of Data

Section 1.2
Data and Big Data

Section 1.3
Managing Big Data

Section 1.4
Summary

Topic 1.1.1
The Data Aspect of a
Connected World

Topic 1.1.2
Data is Growing
Exponentially

Topic 1.1.3
Data Growth Changes
Our Lives

Section 1.0
Introduction

Section 1.1
Value of Data

Section 1.2
Data and Big Data

Section 1.3
Managing Big Data

Section 1.4
Summary

Topic 1.2.1
Where Does Big Data
Come From

Topic 1.2.2
Open Data and Private
Data

Topic 1.2.3
Structured and
Unstructured Data

Topic 1.2.4
Data at Rest and Data in
Motion

Content of the course

Section 1.0
Introduction

Section 1.1
Value of Data

Section 1.2
Data and Big Data

Section 1.3
Managing Big Data

Section 1.4
Summary

Topic 1.3.1
Evolution to Big Data

Topic 1.3.2
Basic Data Management
Technologies

Section 2.0
Introduction

Topic 2.1.1
Analytics Models

Section 2.1
What is Data Analysis?

Section 2.2
Using Big Data

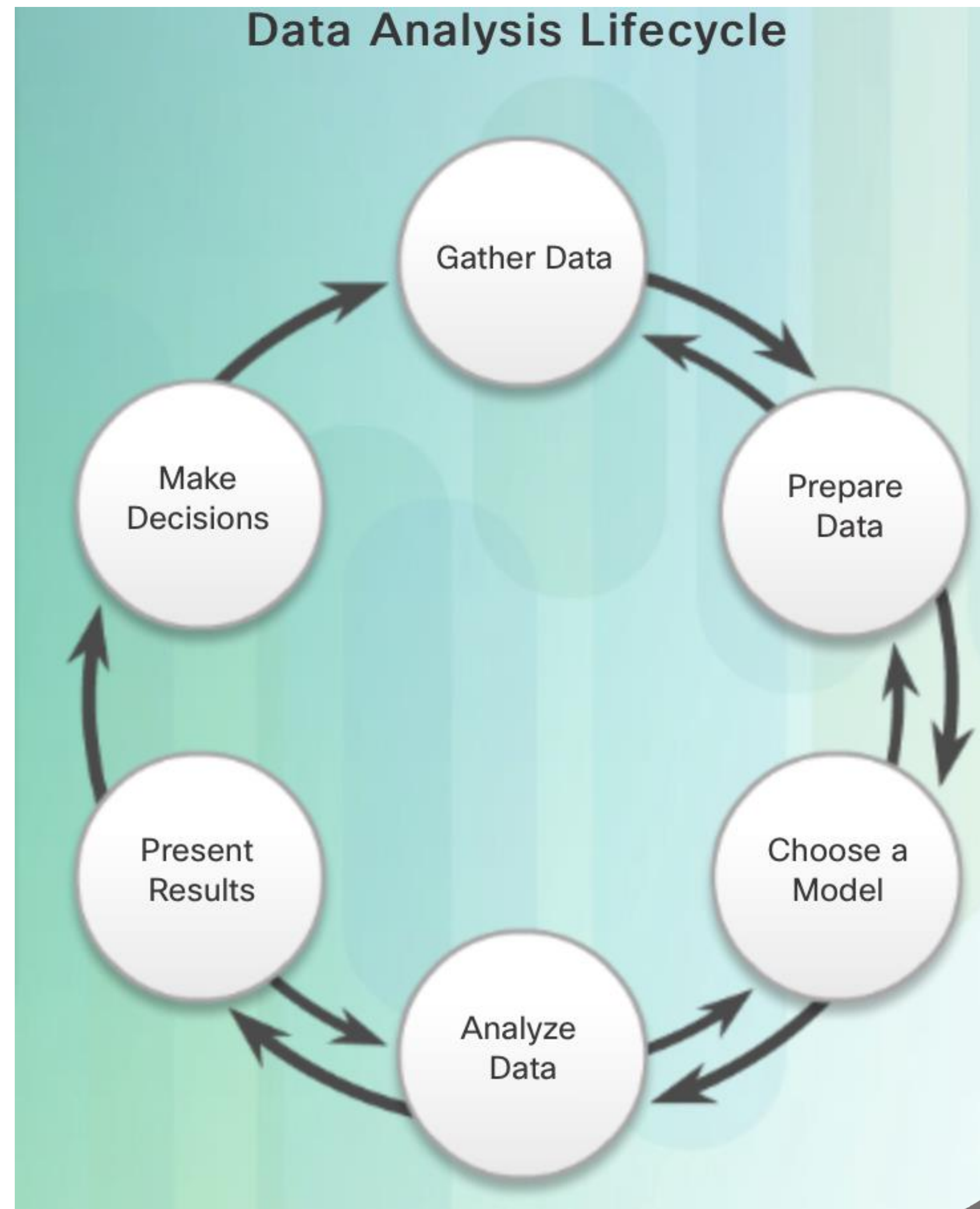
Section 2.3
Data Acquisition and Preparation

Section 2.4
Big Data Ethics

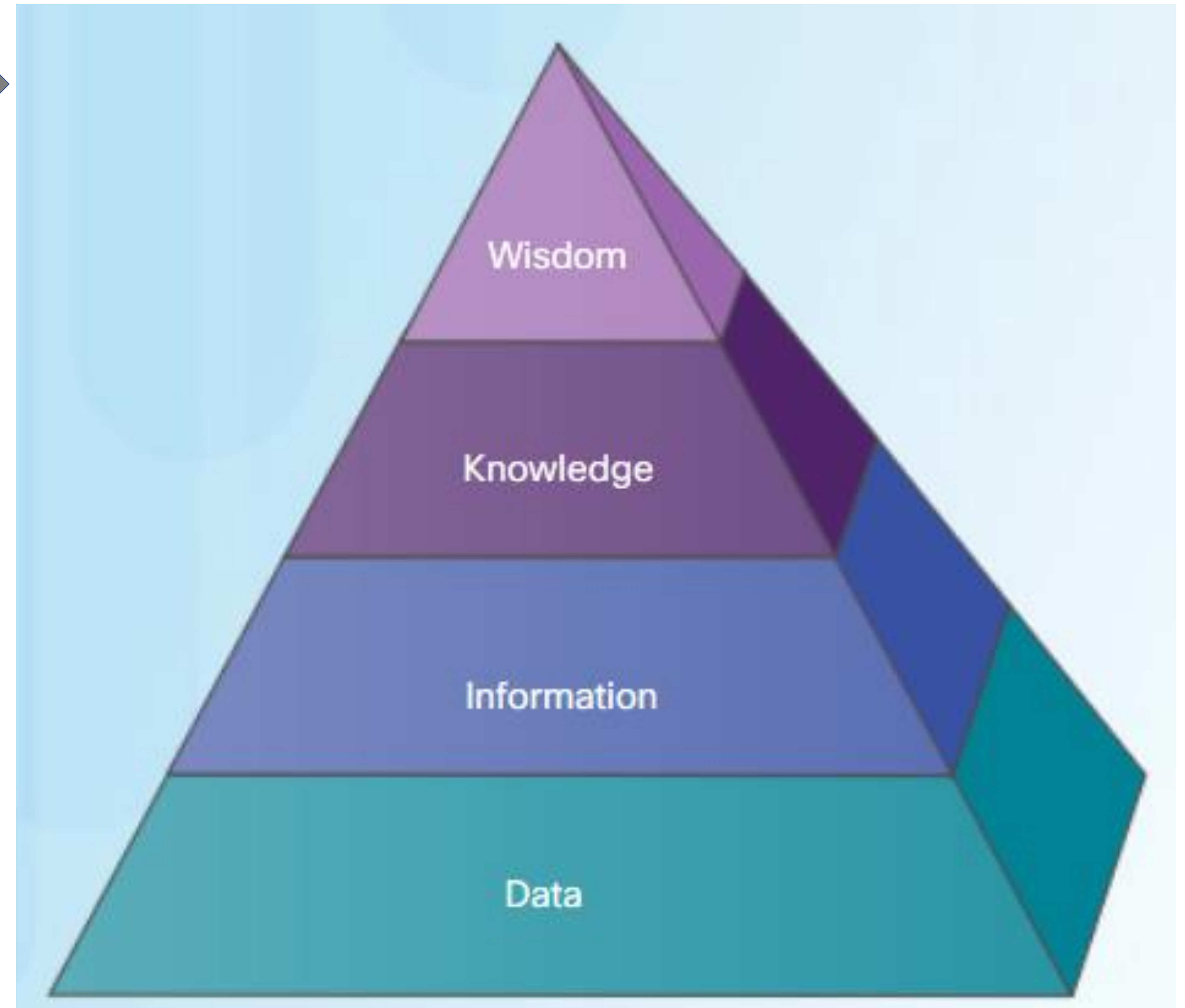
Section 2.5
Preparation for Chapter 2
Internet Meter Labs

Section 2.6
Summary

Data Analysis Lifecycle



Section 2.0 Introduction	Topic 2.2.1 Why Analyze Big Data?
Section 2.1 What is Data Analysis?	Topic 2.2.2 Types of Data Analysis
Section 2.2 Using Big Data	Topic 2.2.3 Timely Analysis of Big Data
Section 2.3 Data Acquisition and Preparation	Topic 2.2.4 Data Analysis Lifecycle
Section 2.4 Big Data Ethics	
Section 2.5 Preparation for Chapter 2 Internet Meter Labs	
Section 2.6 Summary	



Content of the course

Section 2.0 Introduction	Topic 2.2.1 Why Analyze Big Data?
Section 2.1 What is Data Analysis?	Topic 2.2.2 Types of Data Analysis
Section 2.2 Using Big Data	Topic 2.2.3 Timely Analysis of Big Data
Section 2.3 Data Acquisition and Preparation	Topic 2.2.4 Data Analysis Lifecycle
Section 2.4 Big Data Ethics	
Section 2.5 Preparation for Chapter 2 Internet Meter Labs	
Section 2.6 Summary	



Descriptive	Standard Reporting	What happened?
	Ad Hoc Reporting	How many, how often, where?
	Data Queries	What exactly is the problem?
Predictive	Simulation	What could happen?
	Forecasting	What if these trends continue?
	Predictive Modeling	What will happen next?
Prescriptive	Optimization	How can we have the best outcome?
	Optimizations Under Uncertainty	How can we have the best outcome, given variability?

Section 2.0

Introduction

Section 2.1

What is Data Analysis?

Section 2.2

Using Big Data

Section 2.3

Data Acquisition and Preparation

Section 2.4

Big Data Ethics

Section 2.5

Preparation for Chapter 2
Internet Meter Labs

Section 2.6

Summary

Topic 2.2.1

Why Analyze Big Data?

Topic 2.2.2

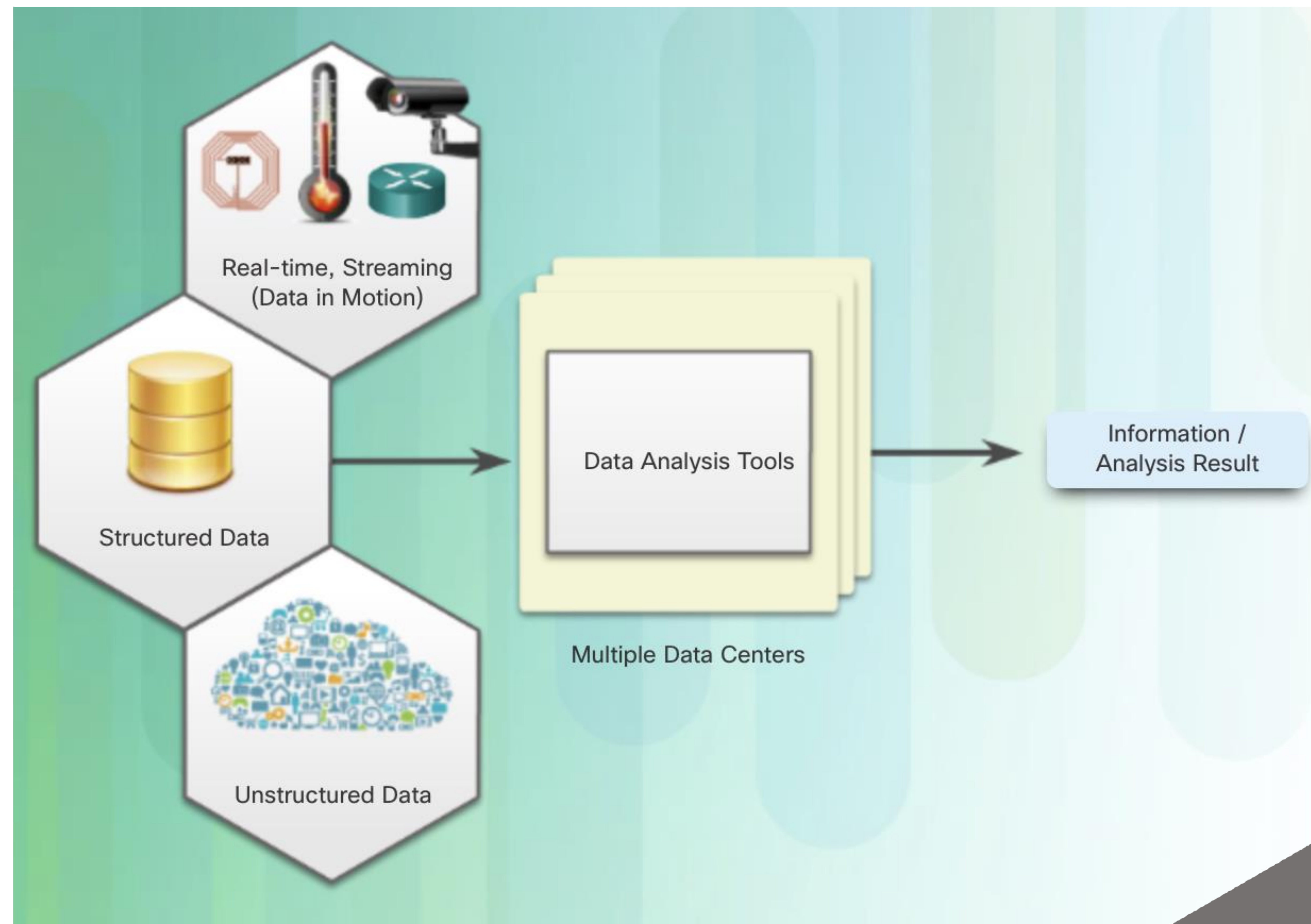
Types of Data Analysis

Topic 2.2.3

Timely Analysis of Big
Data

Topic 2.2.4

Data Analysis Lifecycle



Content of the course

Section 2.0
Introduction

Section 2.1
What is Data Analysis?

Section 2.2
Using Big Data

Section 2.3
Data Acquisition and Preparation

Section 2.4
Big Data Ethics

Section 2.5
Preparation for Chapter 2
Internet Meter Labs

Section 2.6
Summary

Topic 2.3.1
Sources of Data

Topic 2.3.2
Data Preparation

- Files
- Databases
- Sensors
- Internet

- load data
- extract data
- transform data

Section 2.0
Introduction

Section 2.1
What is Data Analysis?

Section 2.2
Using Big Data

Section 2.3
Data Acquisition and Preparation

Section 2.4
Big Data Ethics

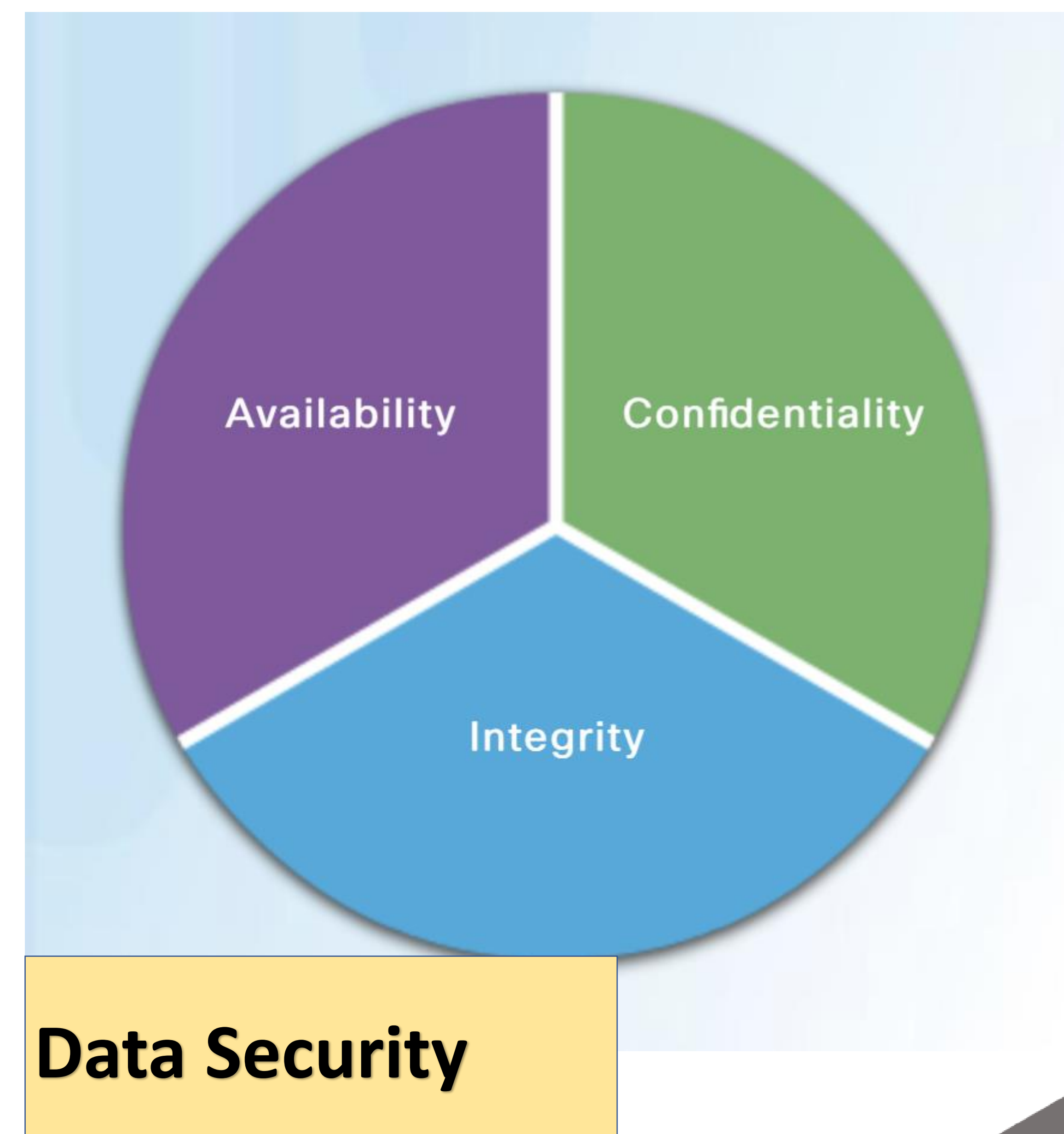
Section 2.5
Preparation for Chapter 2
Internet Meter Labs

Section 2.6
Summary

Topic 2.4.1
What are the Ethical
Concerns?

Data Ethics:

- What?
- How?
- Where?
- Why?
- How Long?



Content of the course

Section 2.0 Introduction

Section 2.1 What is Data Analysis?

Section 2.2 Using Big Data

Section 2.3 Data Acquisition and Preparation

Section 2.4 Big Data Ethics

Section 2.5 Preparation for Chapter 2 Internet Meter Labs

Section 2.6 Summary

Lab - Internet Meter Data Analysis

In this lab, you will acquire Internet speed statistics and store the live data in a comma separated values (csv) file. You will also load the stored data from the csv file to a Python data structure, the Pandas DataFrame, and use its functionalities to explore the data and manipulate it so that it is easily readable.

- Use the PL-App Launcher to connect to your Raspberry Pi.

Select the link for the Course Materials folder and then the link for the Big Data & Analytics folder.

Select lab: 2.5.1.4 - Internet Meter Data Analysis

Lab - Working with Python and SQLite

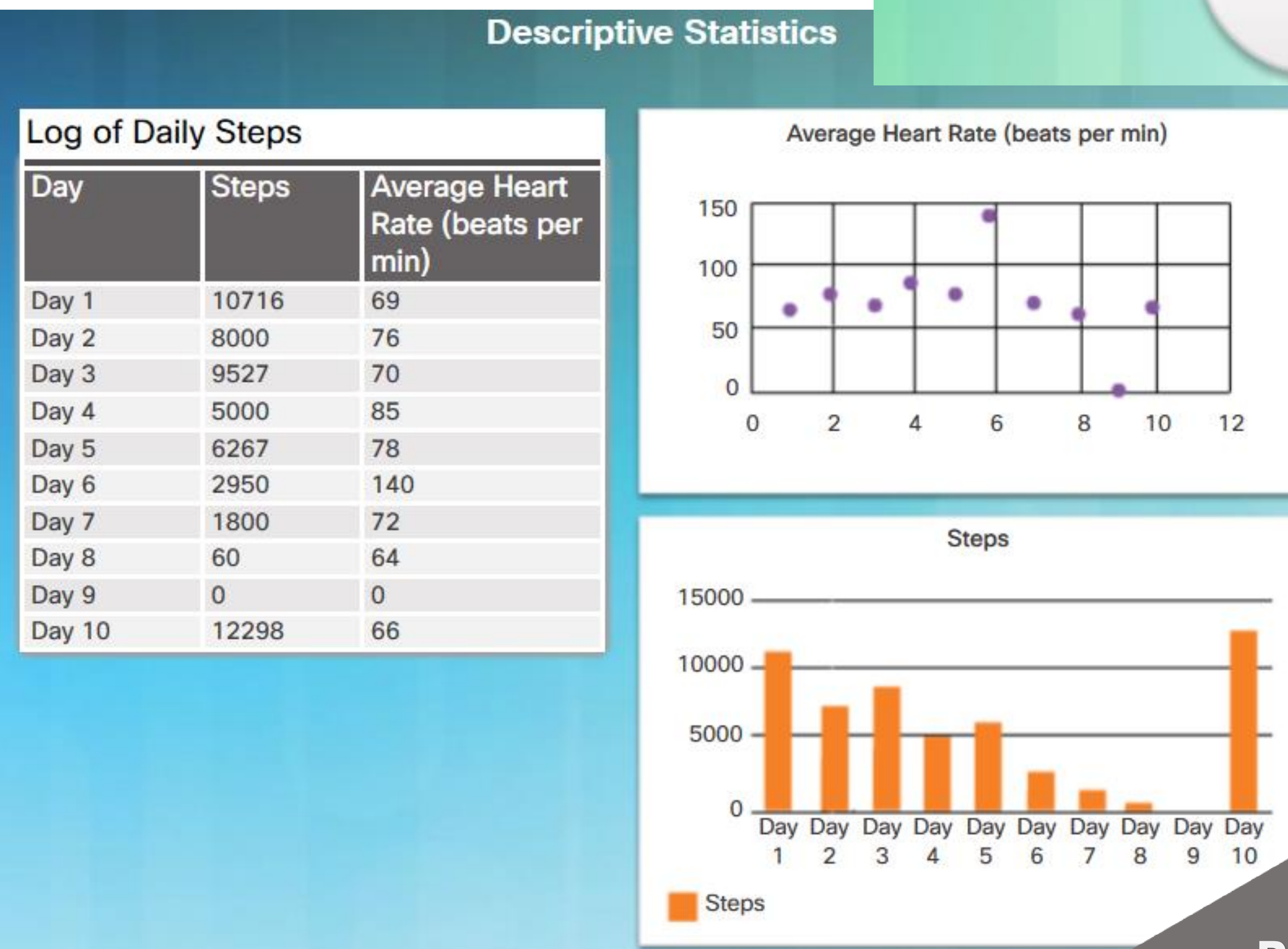
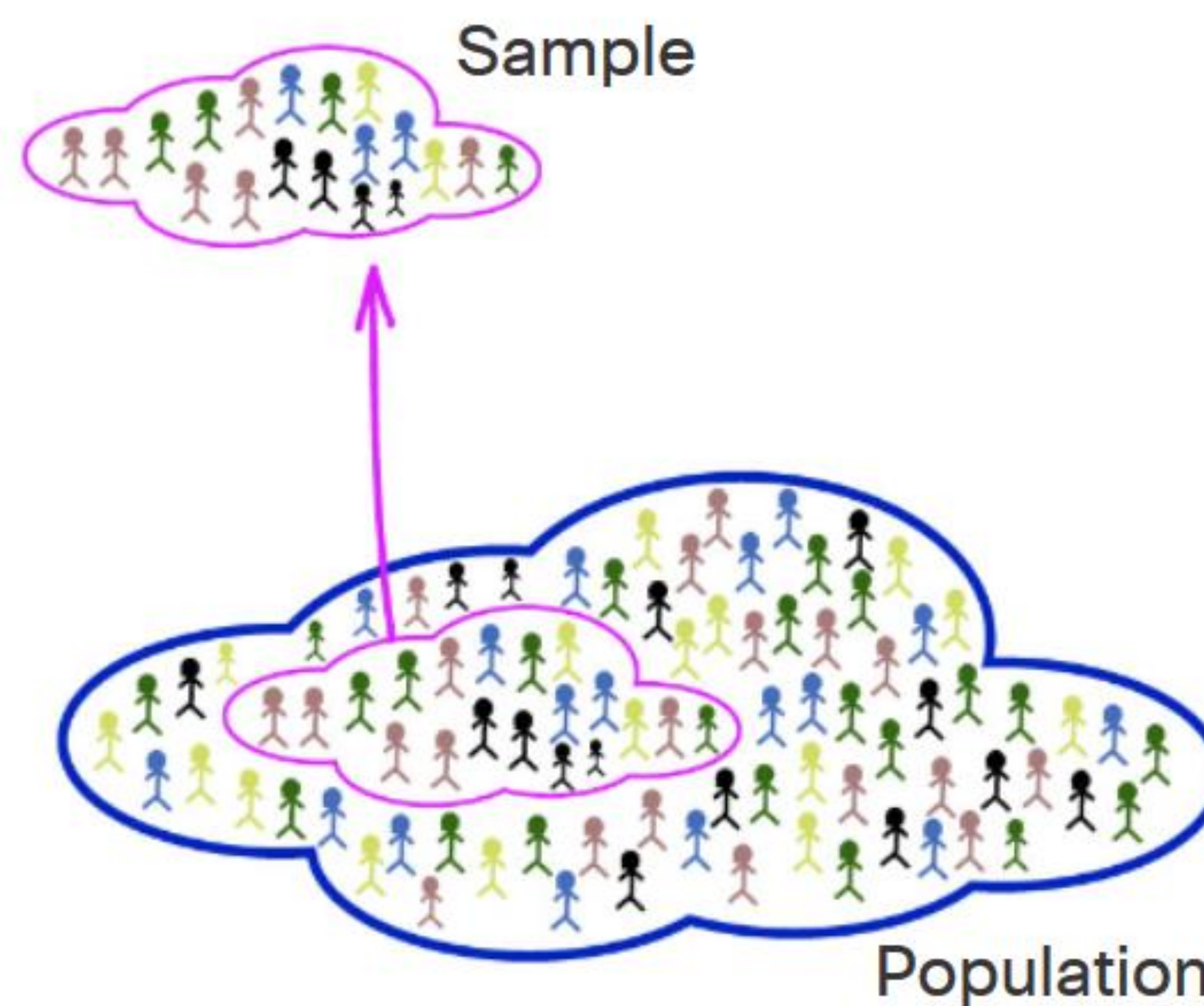
In this lab, you will learn how to connect to a relational database, place queries, and retrieve the data through a Jupyter notebook. You will also learn how to transfer the data from a Microsoft Excel spreadsheet to a SQLite database.

- Use the PL-App Launcher to connect to your Raspberry Pi.

Select the link for the Course Materials folder and then the link for the Big Data & Analytics folder.

Select lab: 2.5.2.4 - Working with Python and SQLite

Section 3.0 Introduction	Topic 3.1.1 Preliminaries
Section 3.1 Analyzing Data	Topic 3.1.2 Statistical Analysis
Section 3.2 Preparation for Chapter 3 Internet Meter Lab	Topic 3.1.3 Characteristics of Samples
Section 3.3 Summary	Topic 3.1.4 Analysis Using Descriptive Statistics
	Topic 3.1.5 Analysis Using Correlation



Pandas 

Columns

rows

Regd. No	Name	Marks%
1000	Steve	86.29
1001	Mathew	91.63
1002	Jose	72.90
1003	Patty	69.23
1004	Vin	88.30

Topic 3.2.1
Basic Analysis with
pandas.

Section 3.0
Introduction

Section 3.1
Analyzing Data

Section 3.2
Preparation for Chapter 3
Internet Meter Lab

Section 3.3
Summary

- Fast and efficient DataFrame object with default and customized indexing.
- Tools for loading data into in-memory data objects from different file formats.
- Data alignment and integrated handling of missing data.
- Reshaping and pivoting of date sets.
- Label-based slicing, indexing and subsetting of large data sets.
- Columns from a data structure can be deleted or inserted.
- Group by data for aggregation and transformations.
- High performance merging and joining of data.
- Time Series functionality.

Section 4.0
Introduction

Section 4.1
Predictive Analytics

Section 4.2
Model Evaluation

Section 4.3
Preparation for Chapter 4 Labs

Section 4.4
Summary

Topic 4.1.1
Machine Learning

Topic 4.1.2
Regression

Topic 4.1.3
Classification

Supervised Machine
Learning

Regression Algorithm

Classification Algorithm

Unsupervised Machine
Learning

Clustering Algorithm

Association Algorithm

Content of the course

Section 4.0
Introduction

Section 4.1
Predictive Analytics

Section 4.2
Model Evaluation

Section 4.3
Preparation for Chapter 4 Labs

Section 4.4
Summary

Topic 4.2.1
Validity and Reliability

Topic 4.2.2
Error in Analyses

Topic 4.2.3
Evaluating Analytic
Models

Section 4.0
Introduction

Section 4.1
Predictive Analytics

Section 4.2
Model Evaluation

Section 4.3
Preparation for Chapter 4 Labs

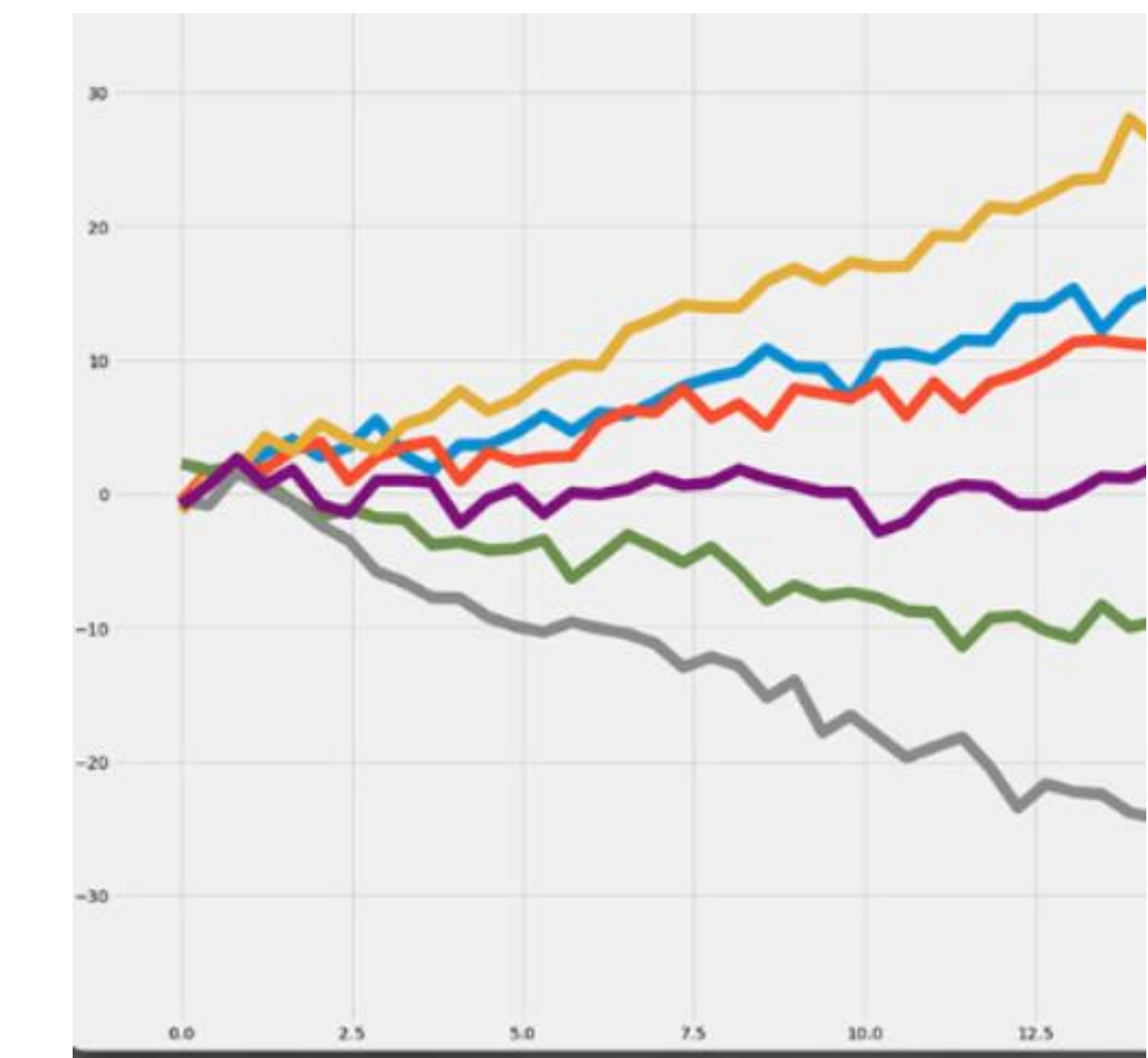
Section 4.4
Summary

Topic 4.3.1
Regression and
Prediction Lab

Topic 4.3.2
Lab - Internet Meter
Anomaly Detection

Using scikit-learn for Regression Analysis

In the first lab, you will use regression analysis to view historical data about the growth of Internet traffic. You will quantify the relationship between the year and the measurement of Internet traffic. In order to do this in Python, you will use scikit-learn, a popular machine learning library. This library contains many useful tools for data analysis and is built on NumPy, SciPy, and matplotlib. Click [here](#) to learn more about the scikit-learn library.



Style Sheets for Plots

You will also install pandas, numpy, and matplotlib. The matplotlib library includes different styles for showing your plots. The figure shows the fivethirtyeight plot style you will use in the lab.

Lab - Internet Meter Linear Regression

In this lab, you will become familiar with the concepts of regression and prediction and the concept of working with data.

Use the PL-App Launcher to connect to your Raspberry Pi.

Section 5.0
Introduction

Section 5.1
Building a Data Story

Section 5.2
The Power of Visualization

Section 5.3
Preparation for Chapter 5 Labs

Section 5.4
Summary

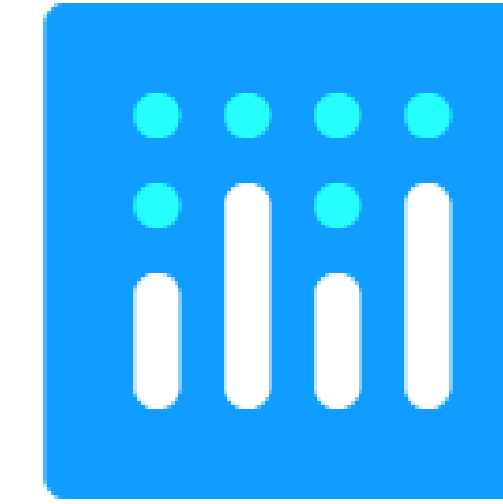
Topic 5.1.1
Know Your Purpose

Topic 5.1.2
Proposition and Evidence

Using Evidence

Deductive Reasoning

Inductive Reasoning



plotly

- makes interactive, publication-quality graphs online,
- Examples of how to make line plots,
 - ✓ scatter plots,
 - ✓ area charts,
 - ✓ bar charts,
 - ✓ error bars,
 - ✓ box plots,
 - ✓ histograms,
 - ✓ heatmaps,
 - ✓ subplots,
 - ✓ multiple-axes,
 - ✓ polar charts, and
 - ✓ bubble charts

Section 5.0 Introduction	Topic 5.2.1 Pyplot
Section 5.1 Building a Data Story	Topic 5.2.2 Plotly
Section 5.2 The Power of Visualization	Topic 5.2.3 Choosing the Right Visualization for the Job
Section 5.3 Preparation for Chapter 5 Labs	
Section 5.4 Summary	

Section 5.0
Introduction

Section 5.1
Building a Data Story

Section 5.2
The Power of Visualization

Section 5.3
Preparation for Chapter 5 Labs

Section 5.4
Summary

Topic 5.3.1
Chapter 5 Labs

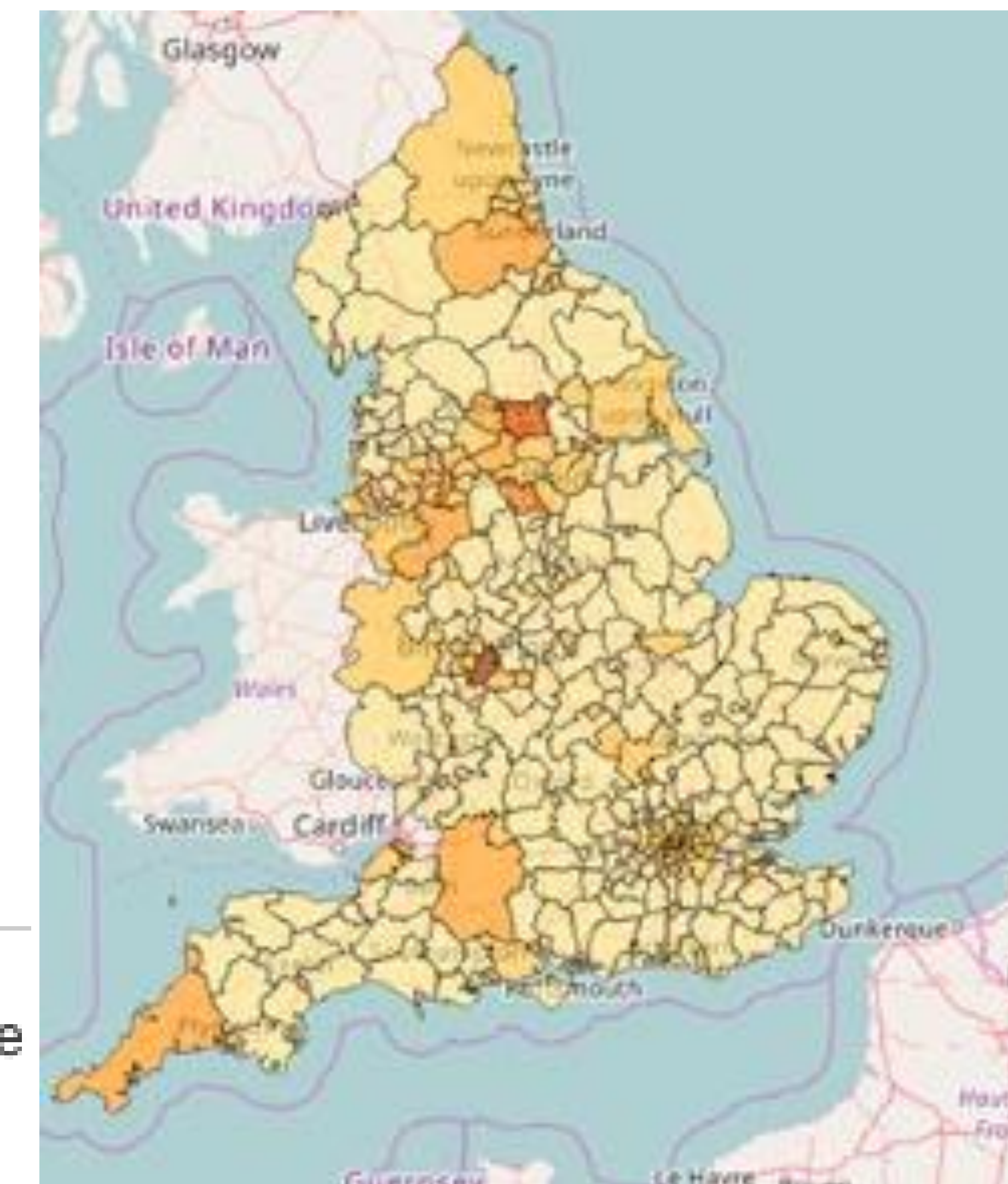
Folium Library

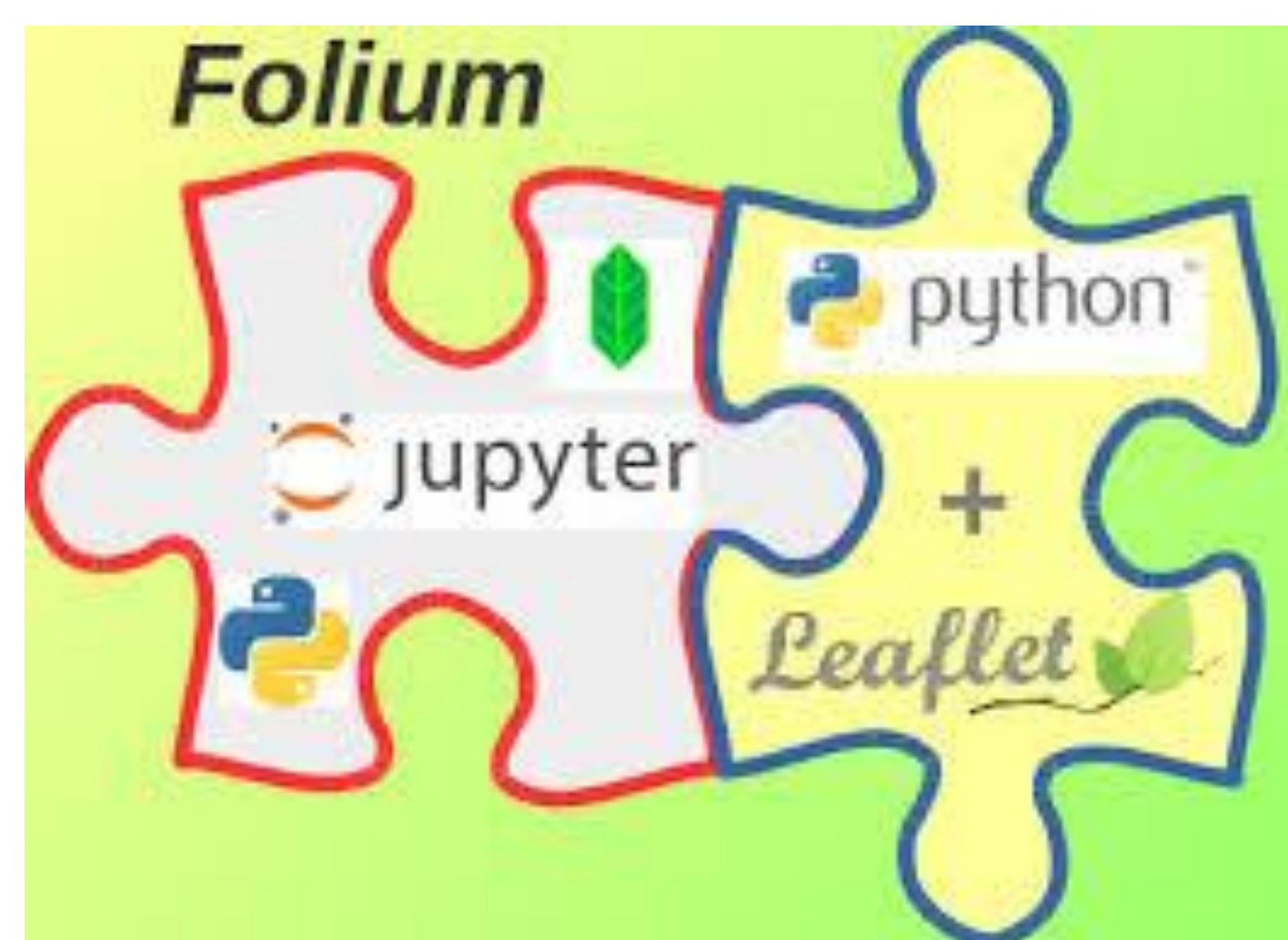
You can also use Python libraries to plot maps. In the first lab, you will import a Python script that calls methods from the Folium library. Folium combines the strength of Python with the mapping abilities of the Leaflet.js library. Folium allows you to take your Python data frames and display them on an interactive Leaflet map.

Lab - Advanced Data Visualization

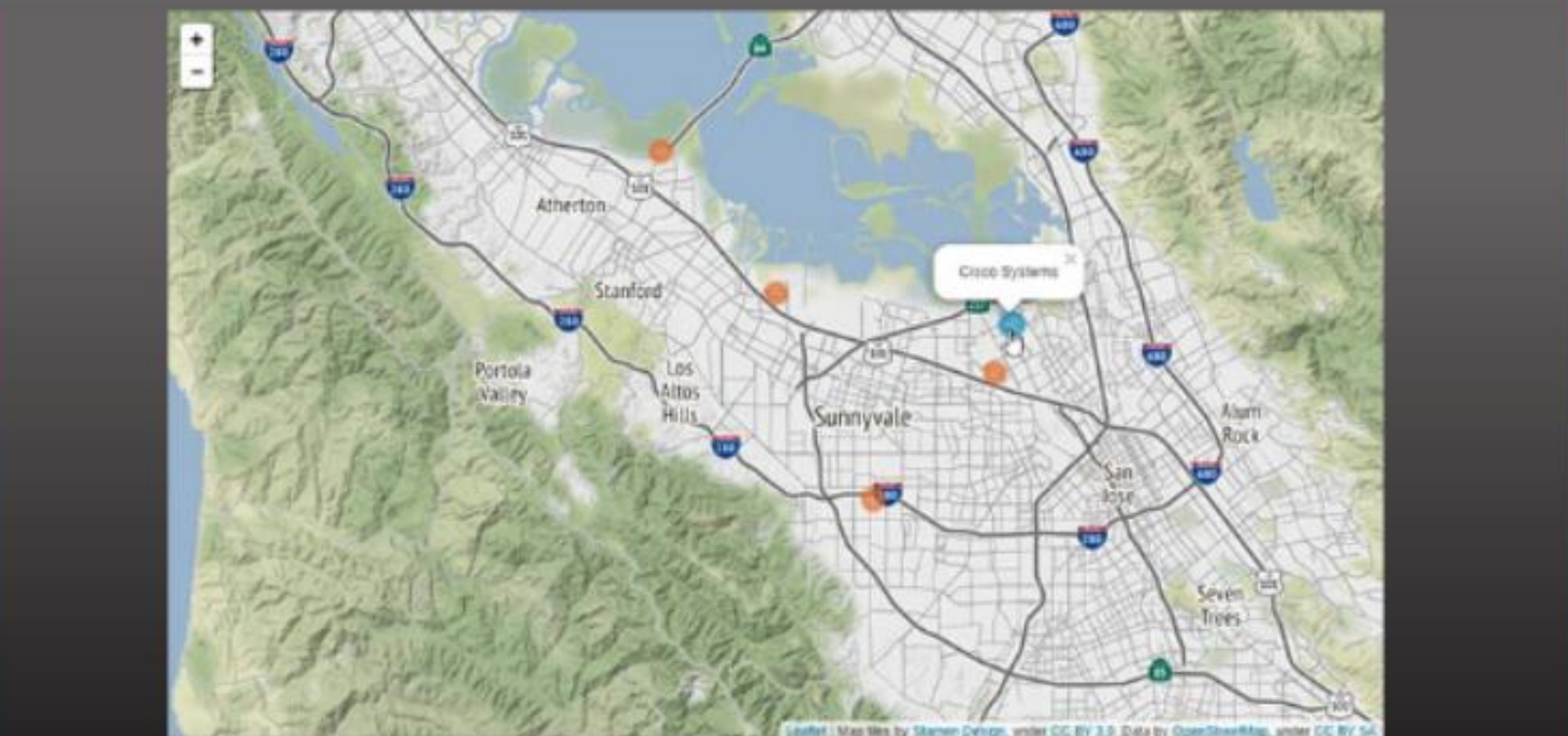
In this lab, you will learn how to combine an SQLite database, JSON files, and pandas DataFrames. You will interact with a wrapper for the library Folium that enables you to plot data on a geographical map. You will produce a map of the United Kingdom that is divided into areas.

Lab - Internet Speed Compliance





```
import folium
svmap = folium.Map(location=[37.3861, -122.0839], tiles="Stamen Terrain", zoom_start=11)
folium.CircleMarker([37.409006, -121.954078], popup='Cisco Systems', color='#0F8ABE', fill
folium.CircleMarker([37.331697, -122.030628], popup='Apple Computer', color='#E46924', fil
folium.CircleMarker([37.422392, -122.084208], popup='Google', color='#E46924', fill_color=
folium.CircleMarker([37.484610, -122.147930], popup='Facebook', color='#E46924', fill_colo
folium.CircleMarker([37.387583, -121.963523], popup='Intel', color='#E46924', fill_color='
svmap
```



Section 5.0 Introduction	Topic 5.3.1 Chapter 5 Labs
Section 5.1 Building a Data Story	
Section 5.2 The Power of Visualization	
Section 5.3 Preparation for Chapter 5 Labs	
Section 5.4 Summary	

Section 6.0
Introduction

Section 6.1
Scaling Data Analytics

Section 6.2
Introduction to Data Engineering

Section 6.3
The Big Data Pipeline

Section 6.4
The Image Processing Labs

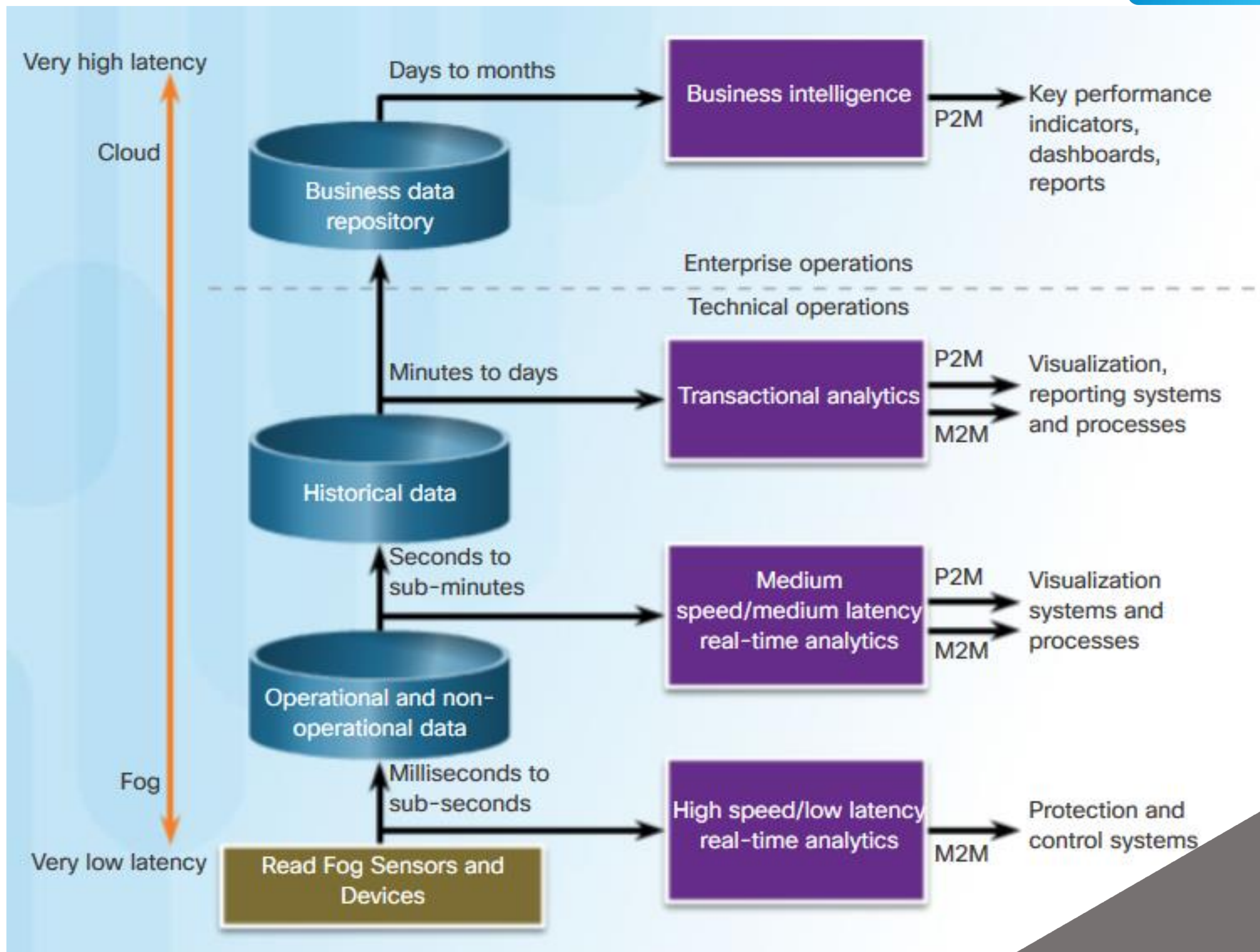
Section 6.5
Summary

Topic 6.1.1
The Cloud, the Fog and
the Edge

Topic 6.1.2
Virtualization

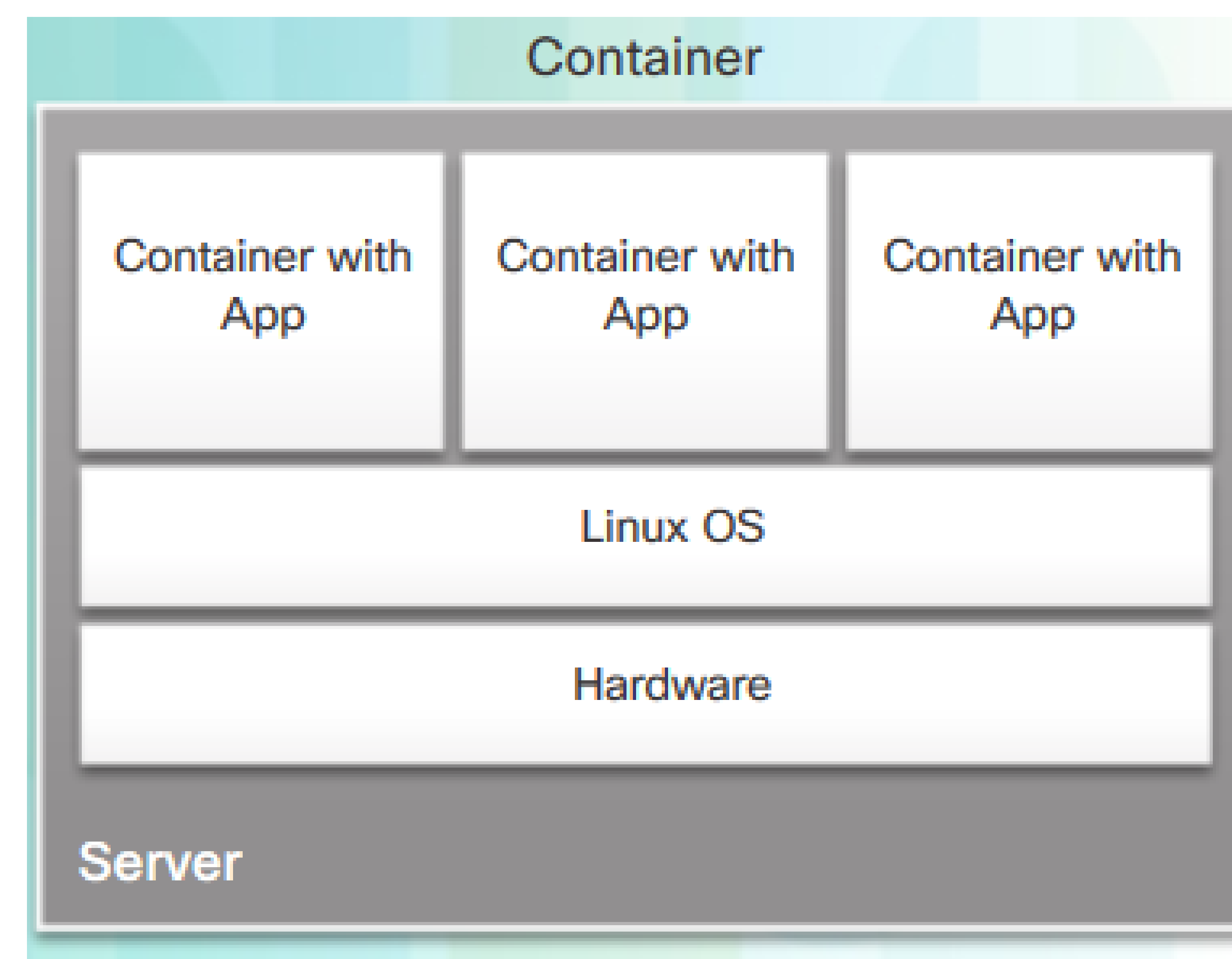
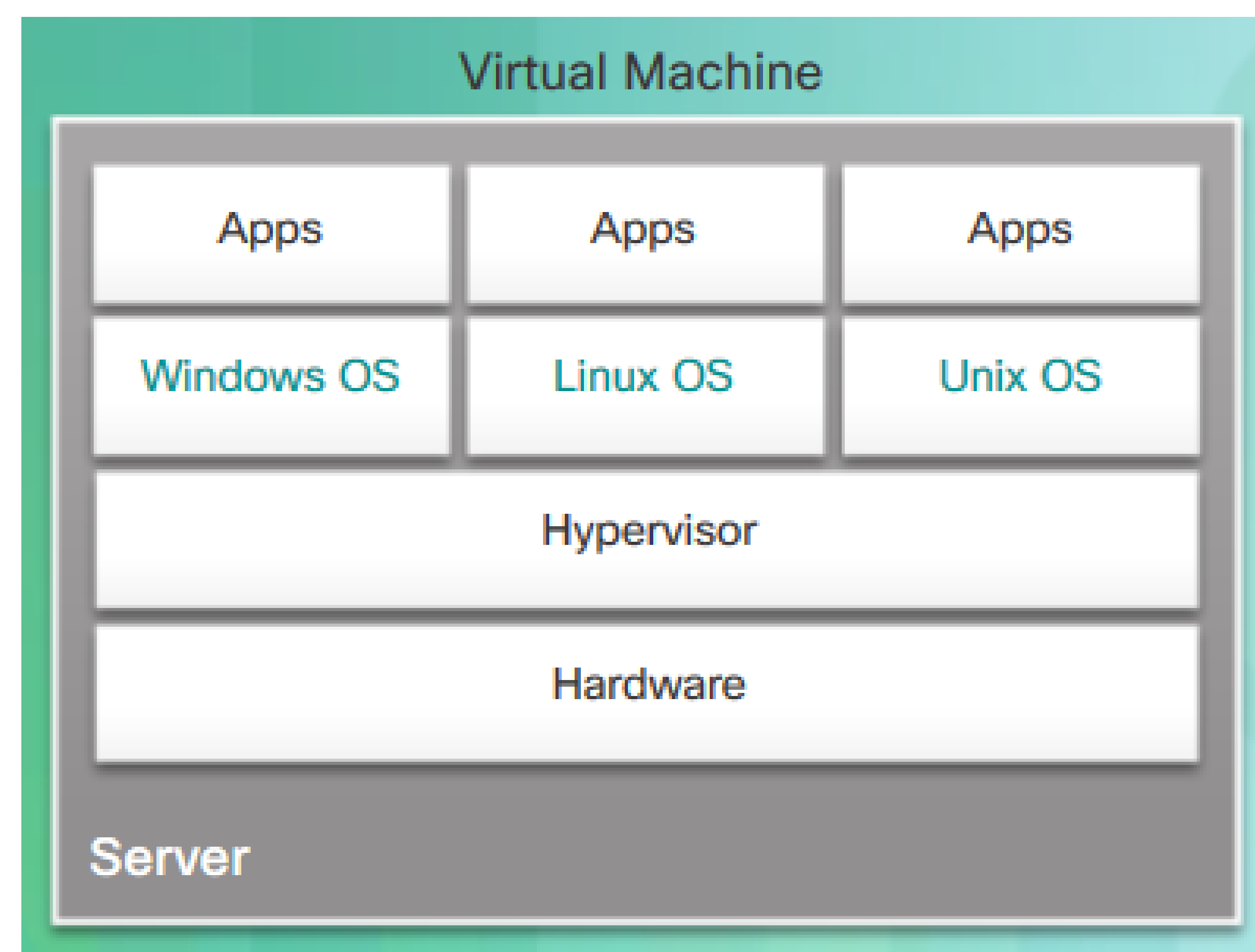
Topic 6.1.3
The Virtualized Data
Center

Content of the course






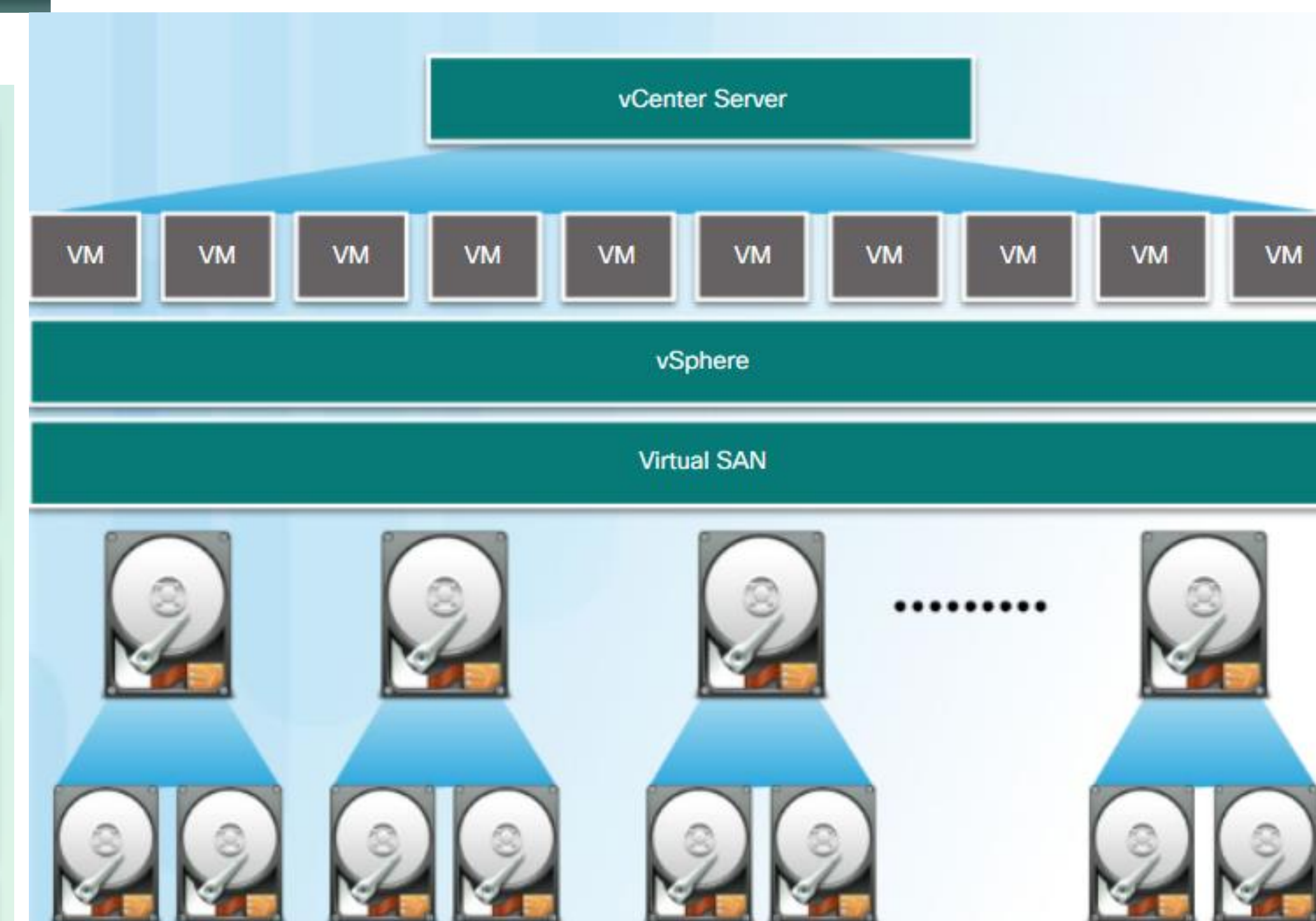
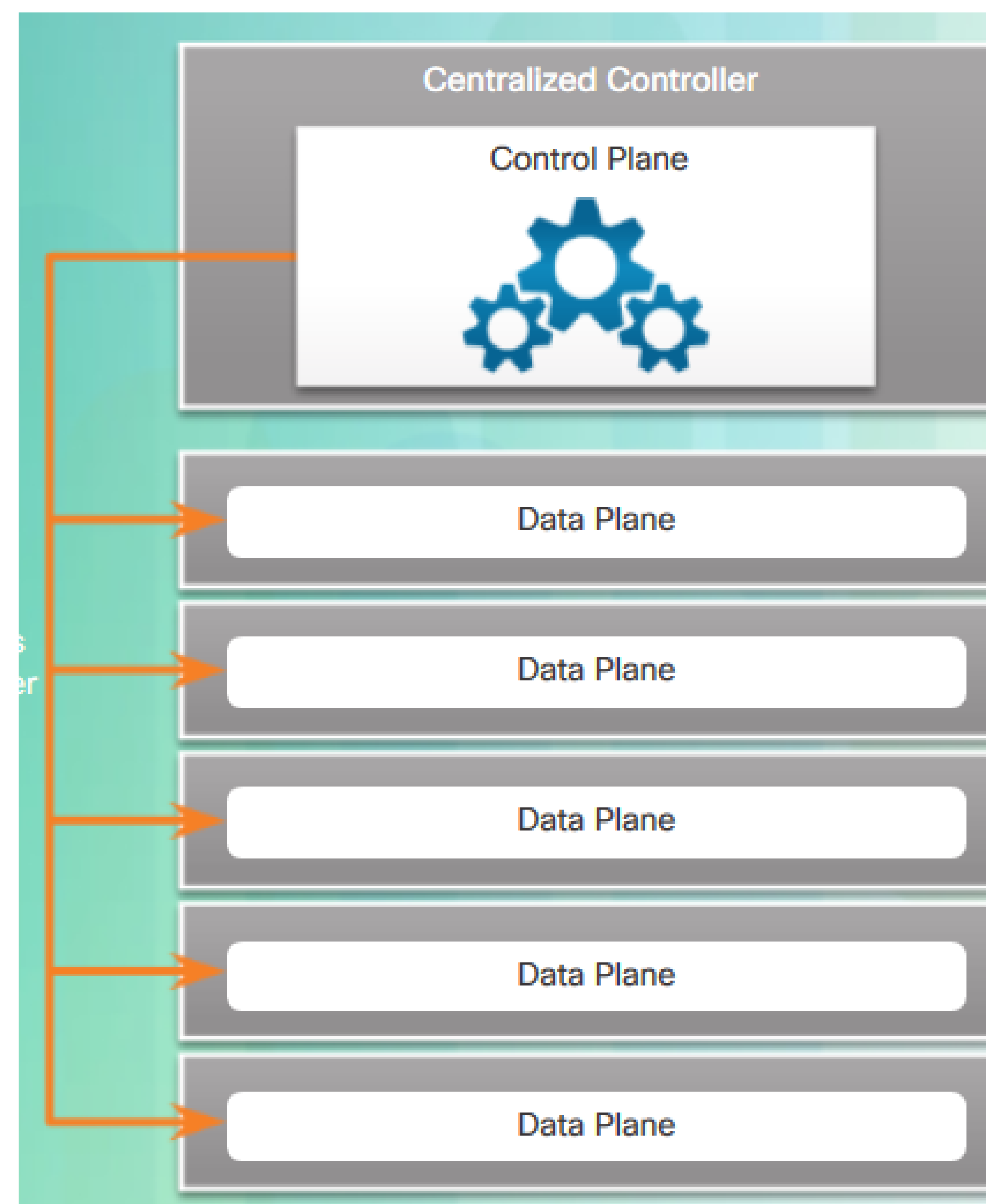
Section 6.0 Introduction	Topic 6.1.1 The Cloud, the Fog and the Edge
Section 6.1 Scaling Data Analytics	Topic 6.1.2 Virtualization
Section 6.2 Introduction to Data Engineering	Topic 6.1.3 The Virtualized Data Center
Section 6.3 The Big Data Pipeline	
Section 6.4 The Image Processing Labs	
Section 6.5 Summary	

Section 6.0 Introduction	Topic 6.1.1 The Cloud, the Fog and the Edge
Section 6.1 Scaling Data Analytics	Topic 6.1.2 Virtualization
Section 6.2 Introduction to Data Engineering	Topic 6.1.3 The Virtualized Data Center
Section 6.3 The Big Data Pipeline	
Section 6.4 The Image Processing Labs	
Section 6.5 Summary	

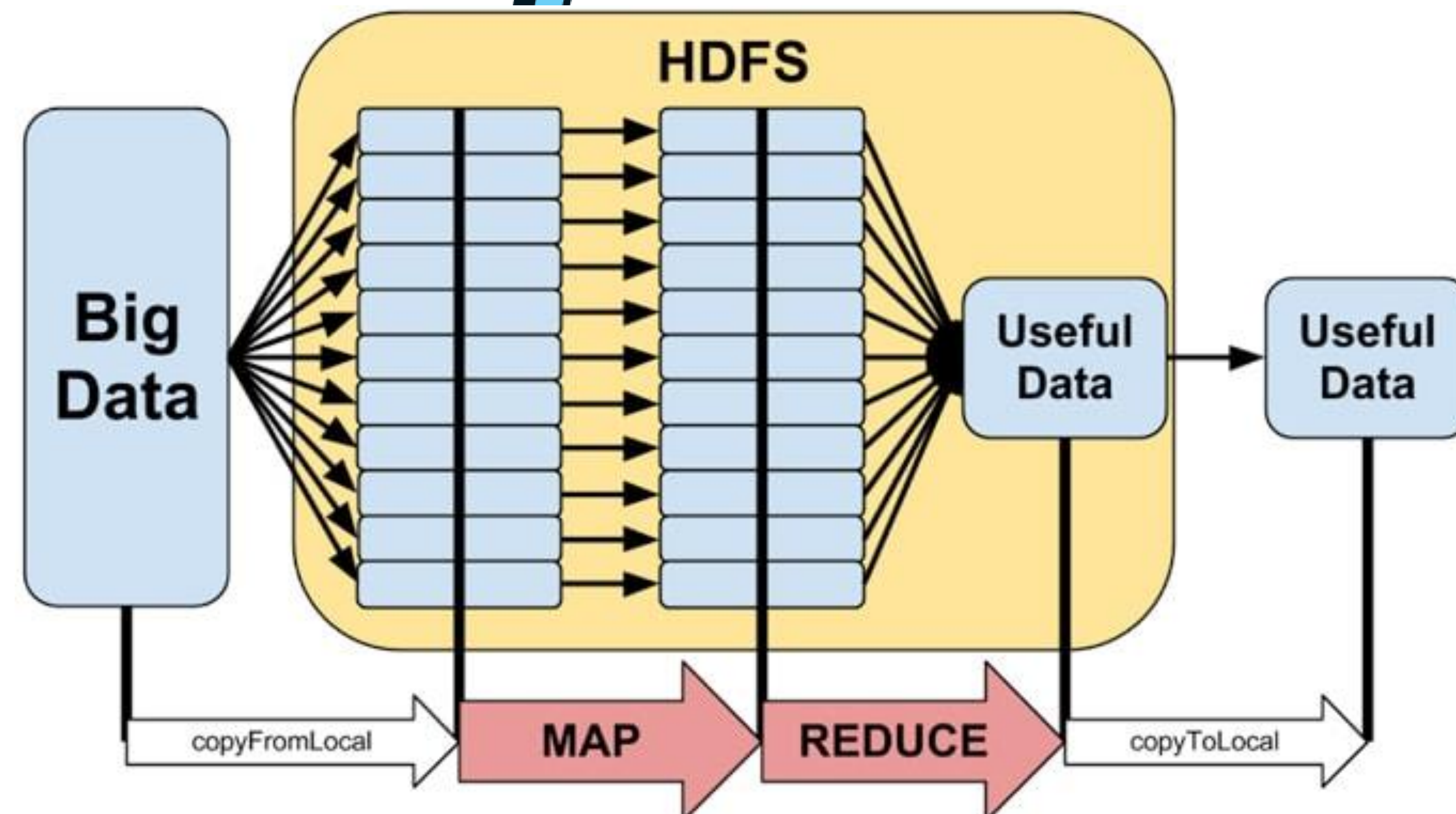
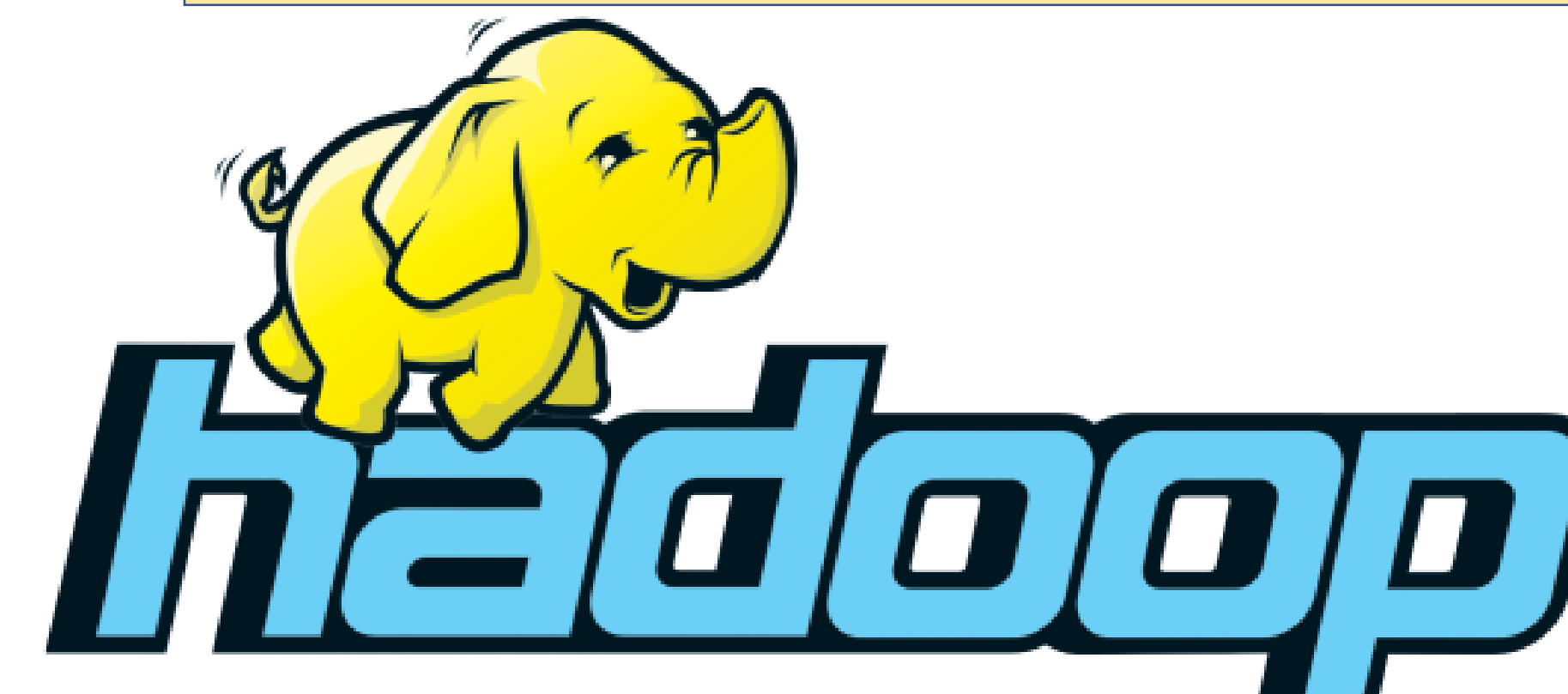


Section 6.0 Introduction	Topic 6.1.1 The Cloud, the Fog and the Edge
Section 6.1 Scaling Data Analytics	Topic 6.1.2 Virtualization
Section 6.2 Introduction to Data Engineering	Topic 6.1.3 The Virtualized Data Center
Section 6.3 The Big Data Pipeline	
Section 6.4 The Image Processing Labs	
Section 6.5 Summary	

-  Software as a Service (SaaS)
-  Platform as a Service (PaaS)
-  Infrastructure as a Service (IaaS)

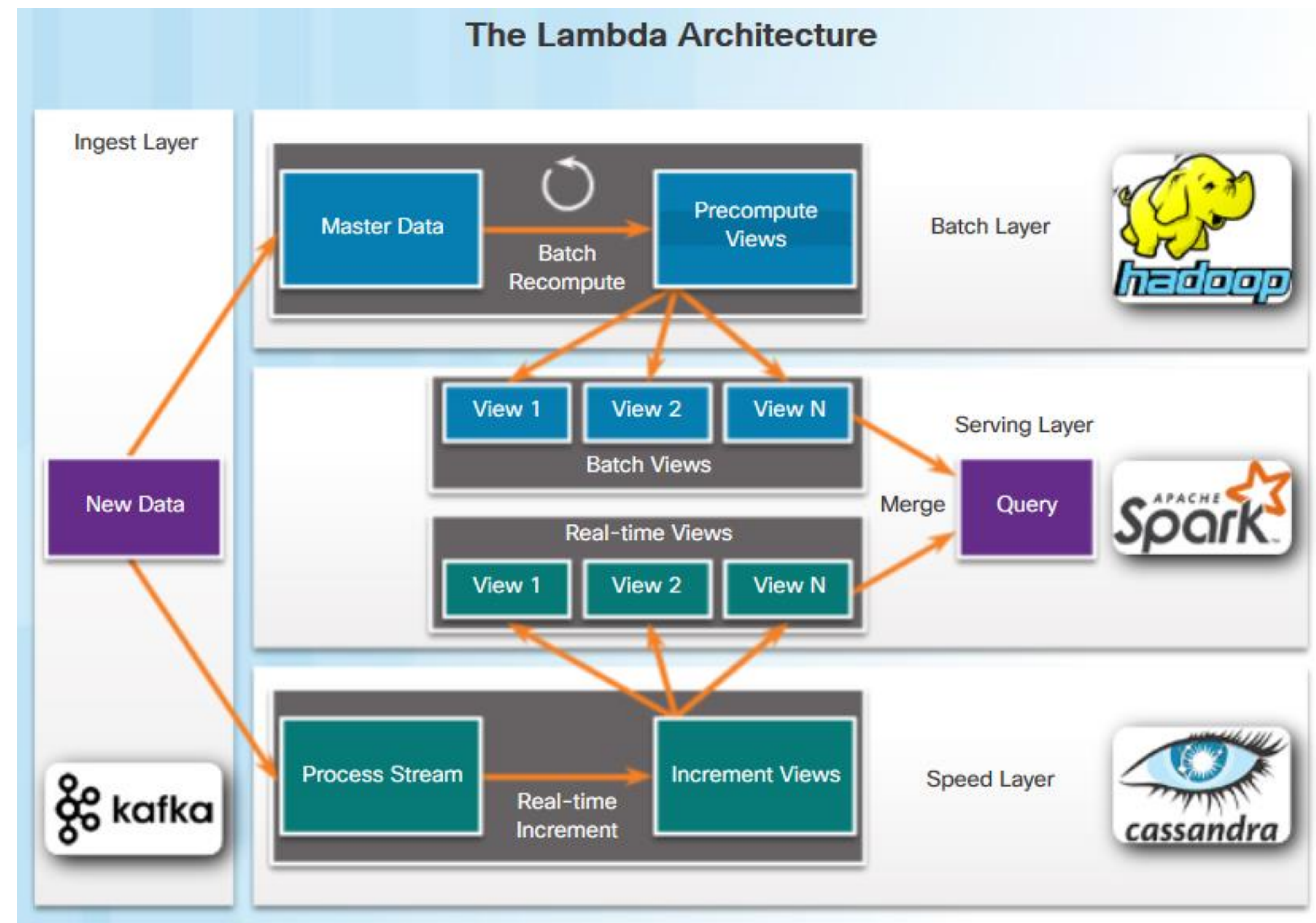


- Scalable Systems
- Distributed filesystem (HDFS)
- Parallel Processing (map reduce)
- Master-Slave-System



Section 6.0 Introduction	Topic 6.2.1 History of Data Engineering
Section 6.1 Scaling Data Analytics	Topic 6.2.2 Big Data Systems
Section 6.2 Introduction to Data Engineering	Topic 6.2.3 What is Hadoop?
Section 6.3 The Big Data Pipeline	
Section 6.4 The Image Processing Labs	
Section 6.5 Summary	

Section 6.0 Introduction	Topic 6.3.1 Data Ingestion
Section 6.1 Scaling Data Analytics	Topic 6.3.2 Data Storage
Section 6.2 Introduction to Data Engineering	Topic 6.3.3 Compute
Section 6.3 The Big Data Pipeline	Topic 6.3.4 The Lambda Architecture
Section 6.4 The Image Processing Labs	
Section 6.5 Summary	



Section 6.0
Introduction

Section 6.1
Scaling Data Analytics

Section 6.2
Introduction to Data Engineering

Section 6.3
The Big Data Pipeline

Section 6.4
The Image Processing Labs

Section 6.5
Summary

Topic 6.4.1
Digital Images as Data

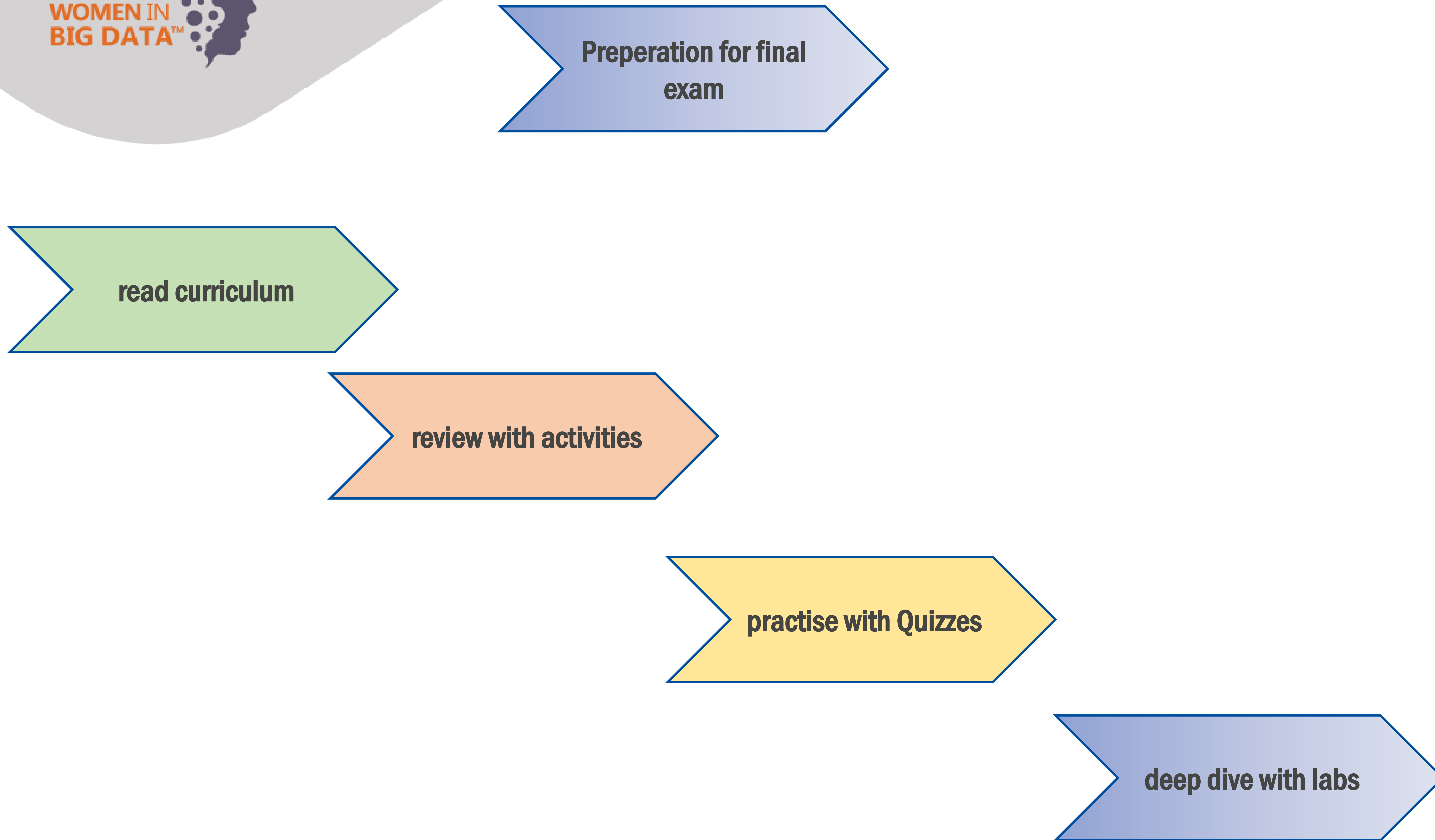


Lab - Smile Detection

Objectives

The objective of this lab is to use machine learning to detect whether a person is smiling when given a picture or a video.

- **Part 1: Detect Faces**
- **Part 2: Data Processing: Face Standardization**
- **Part 3: Load a Pre-trained Model**
- **Part 4: Real-time Smile Detection**



Preperation for final
exam

read curriculum

Chapter 1
Data and the Internet of Things

▶ 1.2
Data and Big Data

▶ 1.2.3
Structured and Unstructured Data

▶ 1.2.3.1
Structured Data



Structured Data



Structured Data

Previously, we have classified data in terms of its accessibility; data is either open or private. Data can also be classified by the way it is arranged, either structured or unstructured.

Structured data refers to data that is entered and maintained in fixed fields within a file or record. Structured data is easily entered, classified, queried, and analyzed by a computer. This includes data found in relational databases and spreadsheets. For example, when you submit your name, address, and billing information to a website, you are creating structured data. The structure will force a certain format for entering the data to minimize errors and make it easier for a computer to interpret it.

If the data set is small enough, structured data is often managed with Structured Query Language (SQL), a programming language created for querying data in relational databases. SQL only



Preperation for final
exam

review with activities

Activity - Compare Structured and Unstructured Data

Instructions

Determine the data type based on each description in the table. Click the appropriate field next to each data type description to indicate your answers.

Data Type Description	Structured	Unstructured
Spreadsheet	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Email in your inbox	<input type="checkbox"/>	<input checked="" type="checkbox"/>
PDF file	<input type="checkbox"/>	<input checked="" type="checkbox"/>
SQL Database	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Word document	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Comma-separated values (CSV) file	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Paperback book	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Electronic book (e-book)	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Check

Reset



Preperation for final
exam

practise with Quizzes

11:48 Sa. 2. März

1403093.netacad.com



My
NetAcad



Account



Dashboard



Courses



Calendar



Inbox



Help



Question 9

2 pts

What are two examples of unstructured data? (Choose two.)

- video content
- user account data
- SQL queries
- blog entry
- customer account spreadsheet



Question 10

2 pts

What is a characteristic of structured data?

- Structured data is subject to intellectual property restrictions.
- It has a predefined organization.
- It is raw data.
- It generates new knowledge.

Preperation for final
exam

deep dive with labs



Cisco Networking Academy®

Mind Wide Open™

Lab - Smile Detection

Objectives

The objective of this lab is to use machine learning to detect whether a person is smiling when given a picture or a video.

- **Part 1: Detect Faces**
- **Part 2: Data Processing: Face Standardization**
- **Part 3: Load a Pre-trained Moel**
- **Part 4: Real-time Smile Detection**

Scenario/Background

In this lab, you will get to work on a complete image processing pipeline. You will use a machine learning model both to detect a face inside of an image and to understand whether the detected face is smiling. You will then build a real-time smile detector, that will be used to take pictures when a person is smiling.

Required Resources

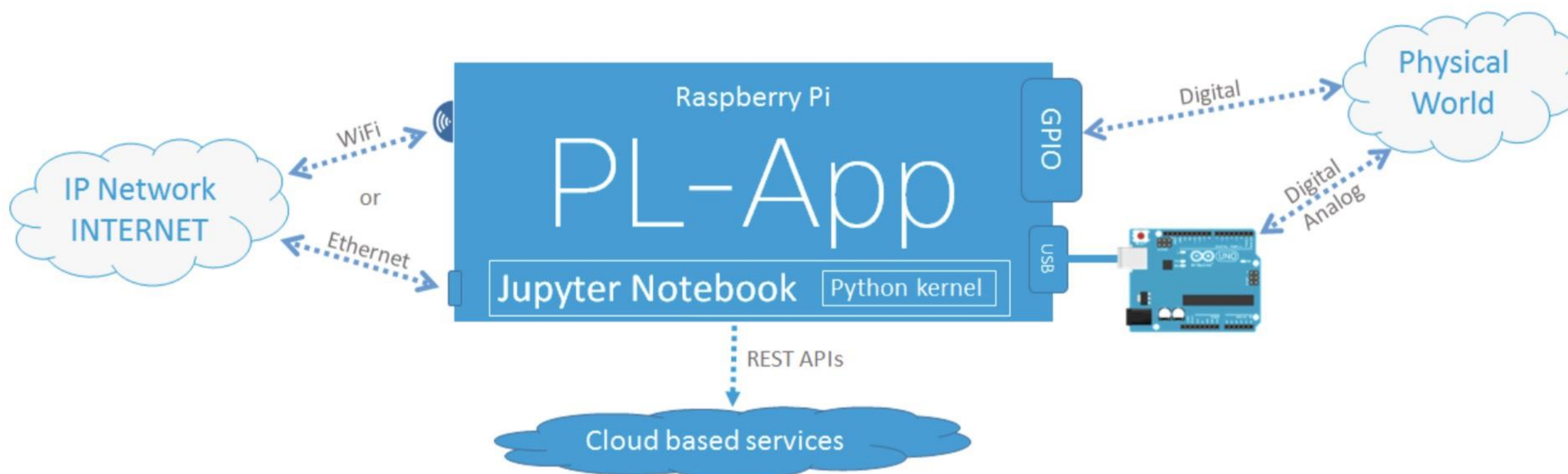
- 1 PC with Internet access
- Raspberry Pi version 2 or higher
- Python libraries: picamera, cv2, time, matplotlib, IPython, numpy, scipy, sklearn
- Datafiles: smile.jpg, nosmile.jpg

Preperation for final
exam

Hands On

Lab: Setting up PL-App with a Raspberry Pi

Lab Topology



Preparation for final
exam

Hands On

LAN

Board Provisioning
List of Devices
Login to Devices

PL-App Launcher

Discovery of devices

PL-App Image

PL-Kit

PL-App

Jupyter Notebooks:
Labs, Code examples,
Dev environment



Competencies

Chapter	Big Data & Analytics	Summary Description
1	Data and the Internet of Things	Understand the concepts of BigData&Analytics
2	Fundamentals of Data Analysis	<p>Knowledge about</p> <ul style="list-style-type: none"> • the basics of descriptive statistic, • the practical aspects in acuring data from a sensor, • how to create visual representations of the data
3	Data Analysis	<p>Ability to</p> <ul style="list-style-type: none"> • explore Data using statistics and visualization to extract information and create hypotheses
4	Advanced Data Analytics and Machine Learning	<p>Ability to</p> <ul style="list-style-type: none"> • predictive analytics, • supervised and unsupervised approaches to machine learning • how to apply models to make predictions from the data
5	Storytelling with Data	<p>Ability</p> <ul style="list-style-type: none"> • to transform analytics results into a clear and convincing narrative and visual communication
6	Architecture for Big Data and Data Engineering	<p>Knowledge about</p> <ul style="list-style-type: none"> • basic principles behind the most important scalable solutions for Big Data • such as Apache Hadoop and the related ecosystem of technologies



Certificate to
demonstrate
competencies



Networking
Academy



Certificate of Completion

Presented to:

Name

For completing the Cisco Networking Academy® Big Data & Analytics course, and demonstrating the ability to perform the following:

- Explain the value of data analytics for an IoT solution.
- Describe and Apply the data analysis process to solve a problem.
- Apply Python scripts to collect, prepare, analyze and visualize or export data.
- Apply basic machine learning algorithms.
- Apply data visualization and storytelling techniques to communicate results of analysis.
- Describe approaches to data management including SQL and No-SQL solutions.
- Explain the fundamental concepts behind the main Big Data platforms.

Date

This is
my
thank you
dance!

