(Big) Data Engineering In Depth

From Beginner to Professional

Mostafa Alaa Mohamed
Senior Big Data Engineer
MoustafaAlaa Moustafa Alaa @Moustafa_alaa22
mustafa.alaa.mohamed@gmail.com

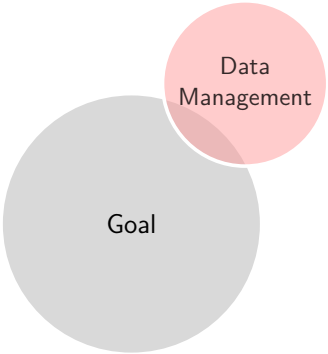
[1]Big Data & Analytics Department, Epam Systems

The Definitive Guide to Big Data Engineering Tasks

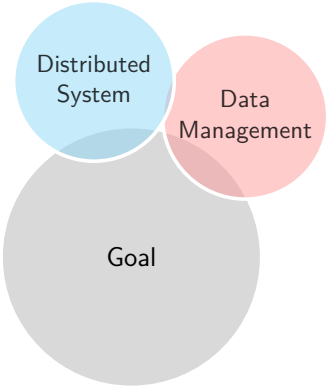
Course Introduction

Course Target

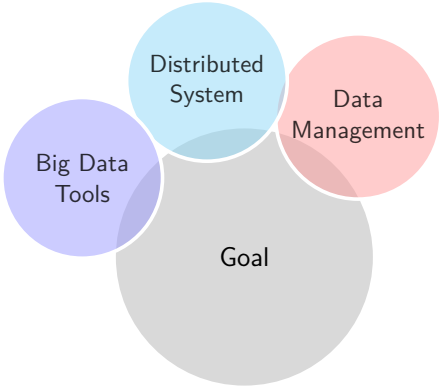
Course Target



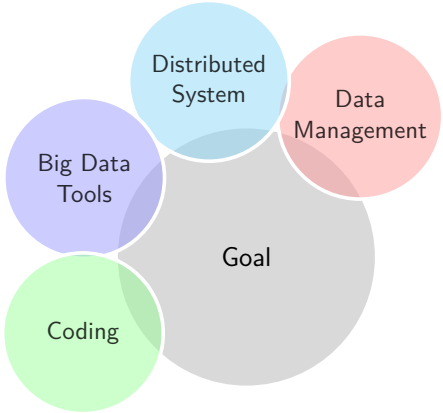
Course Target



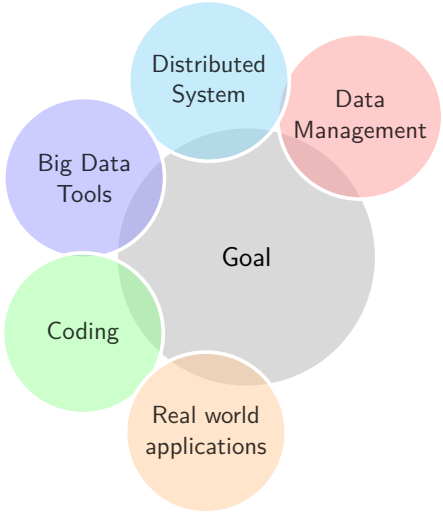
Course Target



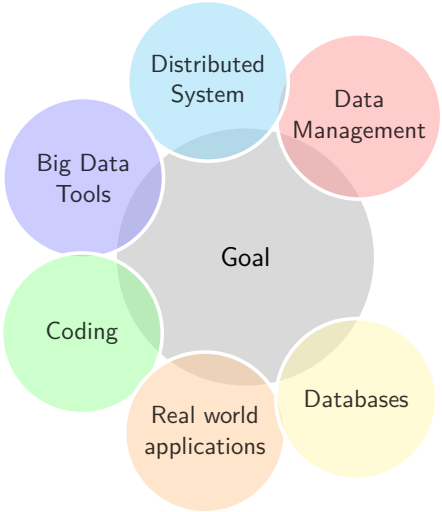
Course Target



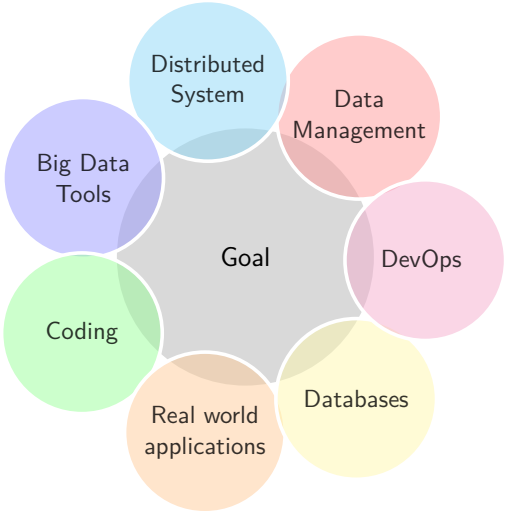
Course Target



Course Target



Course Target



Learning Objectives and Audience

Learning Objectives

- Simplify the concepts in data management.

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.
- Building real-life examples.

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.
- Building real-life examples.
- Applying machine learning over big data.

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.
- Building real-life examples.
- Applying machine learning over big data.
- Automate the data life-cycle process end-to-end (e2e).

Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.
- Building real-life examples.
- Applying machine learning over big data.
- Automate the data life-cycle process end-to-end (e2e).
- Understanding of the DevOps tools and functions in data life-cycle.

Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.

Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.
- Data Warehouse Engineer who needs to know more about big data.

Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.
- Data Warehouse Engineer who needs to know more about big data.
- Software developer who needs to change to data engineering track.

Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.
- Data Warehouse Engineer who needs to know more about big data.
- Software developer who needs to change to data engineering track.
- DevOps engineers who needs to understand the concepts of big data.

Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.
- Data Warehouse Engineer who needs to know more about big data.
- Software developer who needs to change to data engineering track.
- DevOps engineers who needs to understand the concepts of big data.
- Business or entrepreneur who needs to get more information about how to build or manage a data product.

Getting max benefit from this course

Getting max benefit from this course

Take the course advantage

- Follow the videos order as described.

Getting max benefit from this course

Take the course advantage

- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).

Getting max benefit from this course

Take the course advantage

- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.

Getting max benefit from this course

Take the course advantage

- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.
- Do the assignments.

Getting max benefit from this course

Take the course advantage

- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.
- Do the assignments.
- Ask your questions.

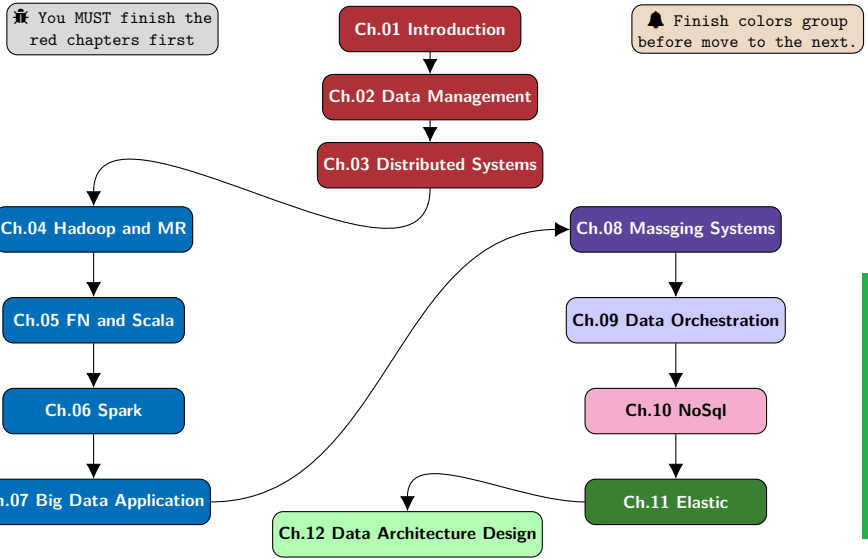
Getting max benefit from this course

Take the course advantage

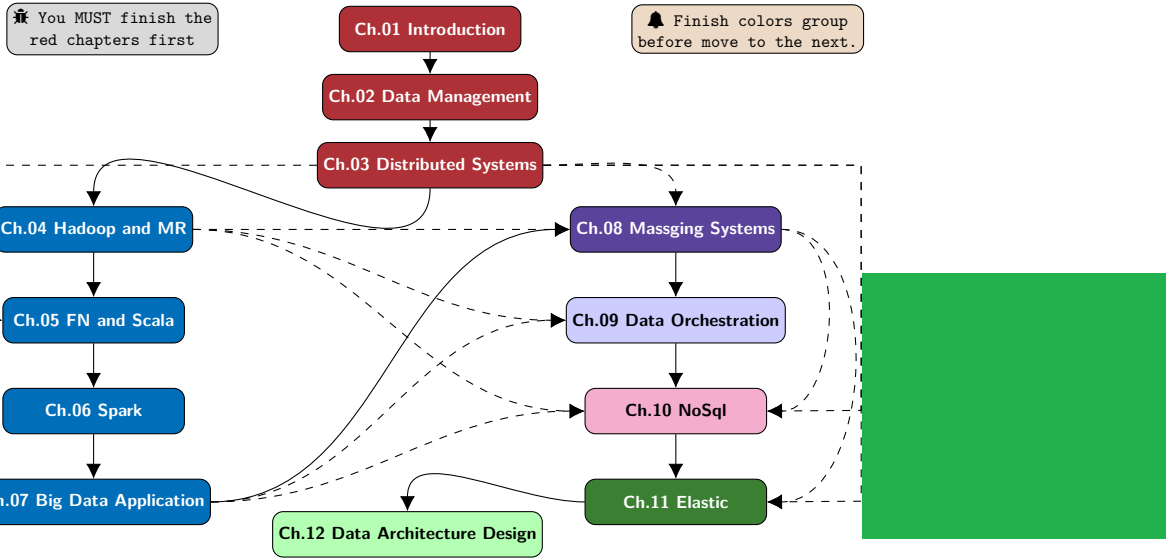
- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.
- Do the assignments.
- Ask your questions.
- Join the online meeting or discussions.

Chapter Dependencies

Chapters Dependencies



Chapter Dependencies (Jump Out Path)



Assignments, Labs, and Text Books

Assignments and Labs

Remark

- Full project code.

Assignments and Labs

Remark

- Full project code.
- Notebooks (Jupyter or Zeppelin).

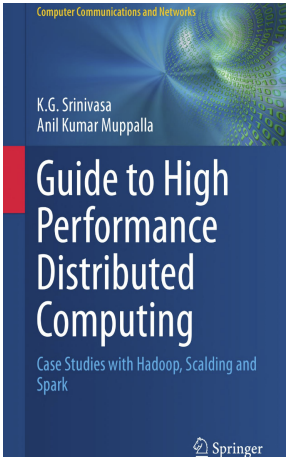
Assignments and Labs

Remark

- Full project code.
- Notebooks (Jupyter or Zeppelin).
- Read the reference.

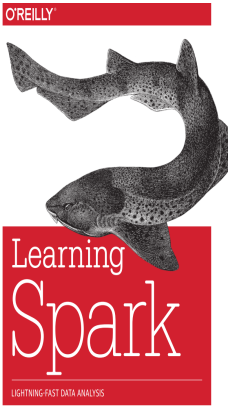
Textbooks-1


O'REILLY
4th Edition
Revised & Updated
Hadoop
The Definitive Guide
STORAGE AND ANALYSIS AT INTERNET SCALE
Tom White


Computer Communications and Networks
K.G. Srinivasa
Anil Kumar Muppalla
Guide to High Performance Distributed Computing
Case Studies with Hadoop, Scalding and Spark
Springer

Textbooks-2

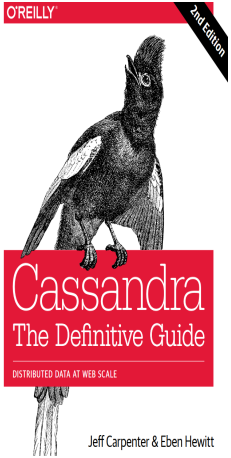
Holden Karau &
Rachel Warren

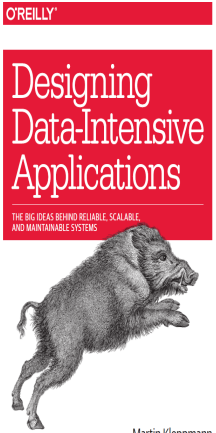


Holden Karau, Andy Konwinski,
Patrick Wendell & Matei Zaharia

Textbooks-3


O'REILLY
Compliments of
Confluent
Kafka
The Definitive Guide
REAL-TIME DATA AND STREAM PROCESSING AT SCALE
Neha Narkhede,
Gwen Shapira & Todd Palino


O'REILLY
2nd Edition
Cassandra
The Definitive Guide
DISTRIBUTED DATA AT WEB SCALE
Jeff Carpenter & Eben Hewitt


O'REILLY
Designing
Data-Intensive
Applications
THE BIG IDEAS BEHIND RELIABLE, SCALABLE,
AND MAINTAINABLE SYSTEMS
Martin Kleppmann

Ugly but important

- User stories or technical discussions are not related to any of my current work or my previous companies.

Ugly but important

- User stories or technical discussions are not related to any of my current work or my previous companies.
- I am working at EPAM Systems. My company approved me for doing this online course public but the materials are not reviewed or assessed by my company. It is on my own responsibilities.