(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

Table of Contents I

Table of Contents II

Table of Contents III

Table of Contents IV

Table of Contents V

Table of Contents VI

Table of Contents VII

Table of Contents VIII

Table of Contents IX

Hadoop and Map-Reduce

Chapter Objectives

- Introduction to Hadoop and its echo-systems.

Chapter Objectives

- Introduction to Hadoop and its echo-systems.
- Why we need Hadoop?

Chapter Objectives

- Introduction to Hadoop and its echo-systems.
- Why we need Hadoop?
- Understand the concept of HDFS and Map-Reduce.

Chapter Objectives

- Introduction to Hadoop and its echo-systems.
- Why we need Hadoop?
- Understand the concept of HDFS and Map-Reduce.
- Developing Map-Reduce applications.

Chapter Objectives

- Introduction to Hadoop and its echo-systems.
- Why we need Hadoop?
- Understand the concept of HDFS and Map-Reduce.
- Developing Map-Reduce applications.
- Using HiveQL over Map-Reduce.

Chapter Objectives

- Introduction to Hadoop and its echo-systems.
- Why we need Hadoop?
- Understand the concept of HDFS and Map-Reduce.
- Developing Map-Reduce applications.
- Using HiveQL over Map-Reduce.
- Hadoop advantages and disadvantages with use cases?

Hadoop Architecture

Hadoop Architecture

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Storage

Storage

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

YARN

YARN

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Hadoop I/O

Hadoop I/O

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Processing

Processing

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Map-Reduce

Map-Reduce

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Map-Reduce Components

Map-Reduce Components

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Word-Count Example

Word-Count Example

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Pig

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Hive

Hive

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

ZooKeeper

ZooKeeper

- Any Big Data solution working based distributed systems.
- What is distributed systems in brief?

Further Readings and Assignment