



# (Big) Data Engineering In Depth

Moustafa Mahmoud  
Data Solution Architect



# Outline

1. Hadoop and Map-Reduce
  - 1.1 Hadoop Architecture
  - 1.2 Map-Reduce
  - 1.3 Pig
  - 1.4 Hive
  - 1.5 ZooKeeper
  - 1.6 Further Readings and Assignment

# Chapter: 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?



## Section: Hadoop Architecture



# Hadoop Architecture

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



## Sub-Section: Storage



# Storage

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



## Sub-Section: YARN



# YARN

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



## Sub-Section: Hadoop I/O





# Hadoop I/O

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



## Sub-Section: Processing



# Processing

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



## Section: Map-Reduce



# Map-Reduce

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



## Sub-Section: Map-Reduce Components



# Map-Reduce Components

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



## Sub-Section: Word-Count Example





# Word-Count Example

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



## Section: Pig



# Pig

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



## Section: Hive



# Hive

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



## Section: ZooKeeper



# ZooKeeper

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



## Section: Further Readings and Assignment





Thank you for watching!



See you in the next video 😊

