

Egyptian Big Data Geeks

1st Meeting

How to Start in the field?

Mostafa Alaa, Data Engineer at Etisalat Misr, DWH Team.

MustafaAlaa.Mohamed@gmail.com

Start From Cloudera VM

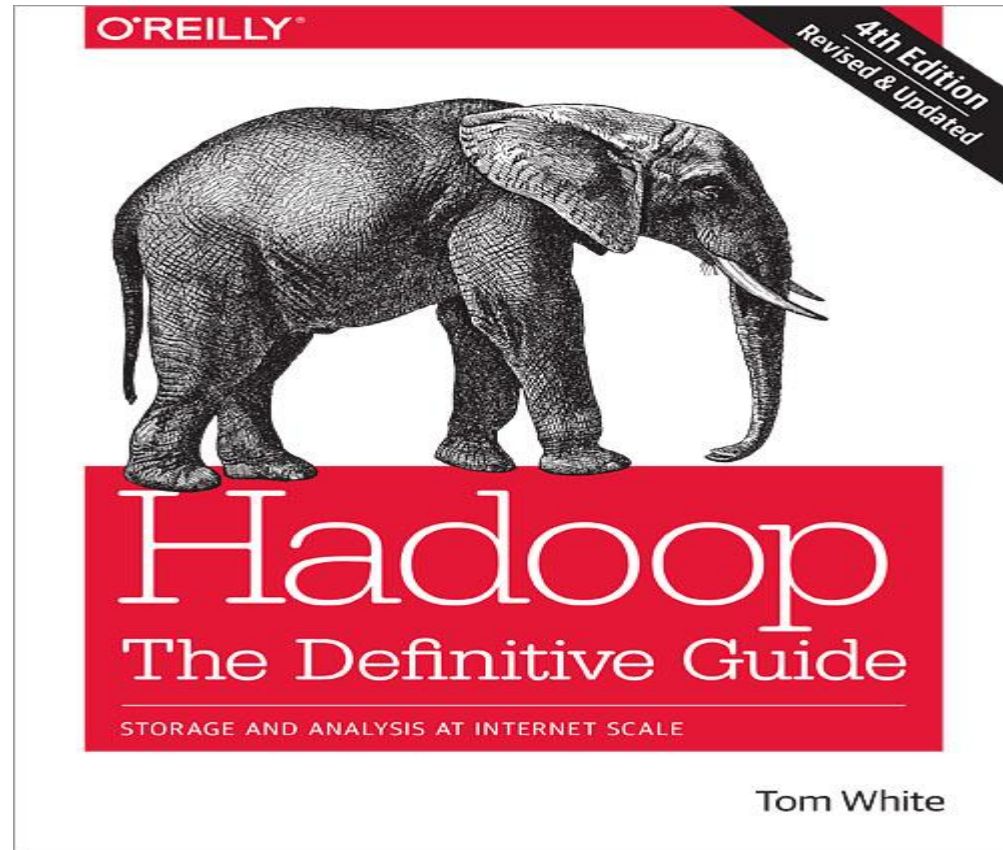
- ▶ Why we need to download cloudera VM?
 - ▶ Easy to start.
 - ▶ Cloudera Stack.
 - ▶ Tools already installed.

http://www.cloudera.com/downloads/quickstart_vms/5-5.html

BIG DATA

► Books

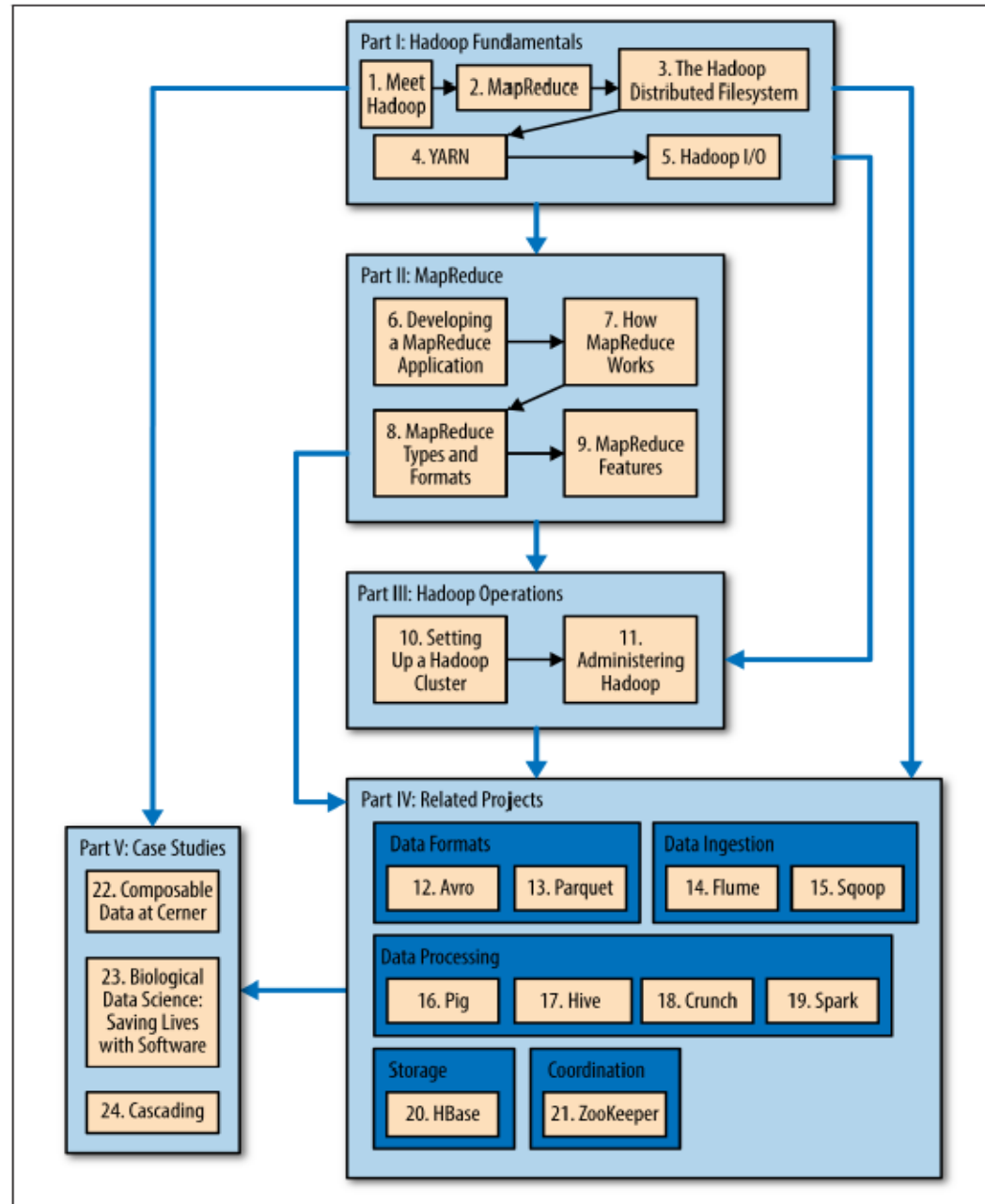
► Hadoop: The Definitive Guide, 4th Edition



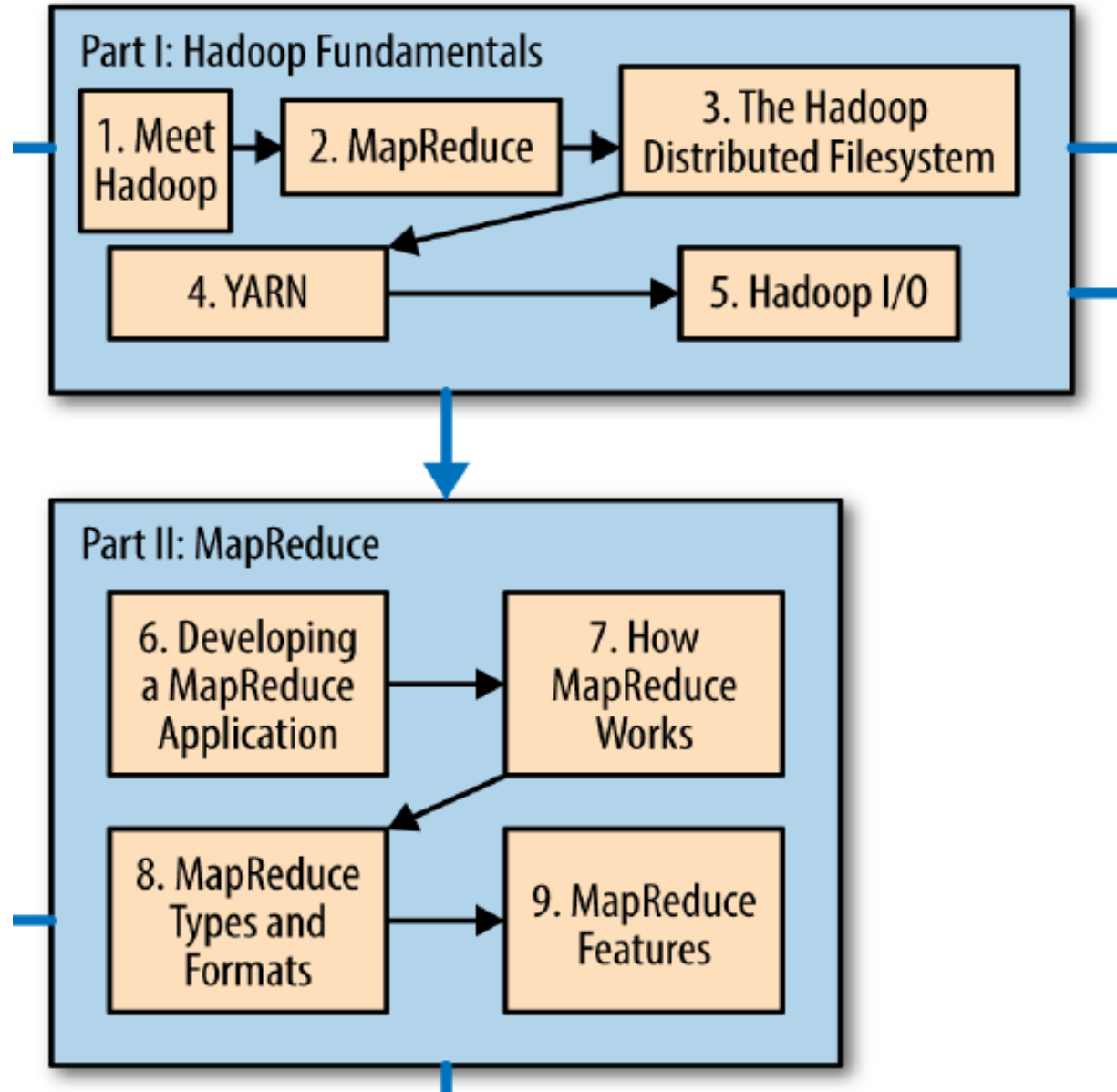
4/25/2016

3

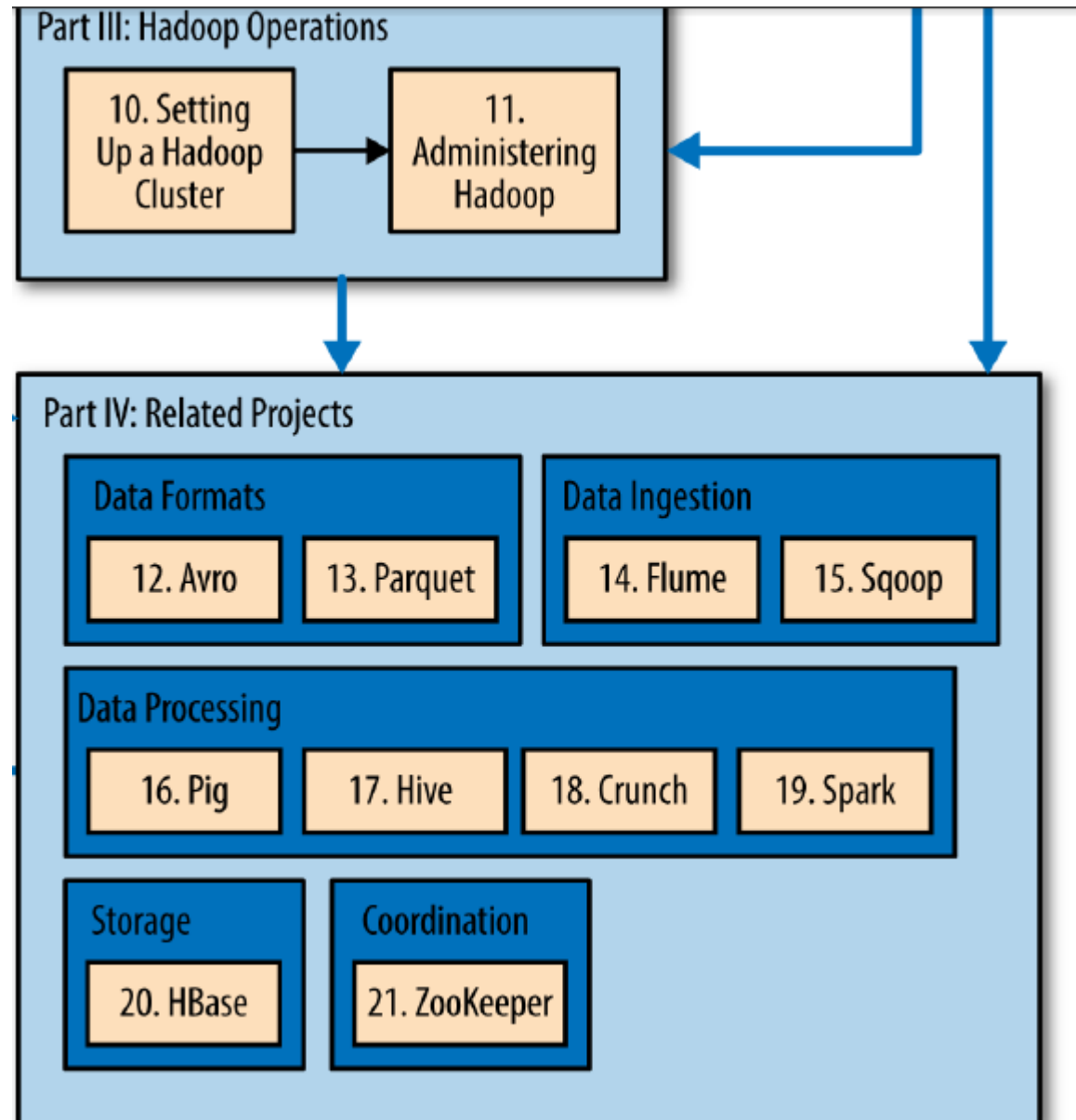
How to read Hadoop The Definitive Guide, 4th Edition



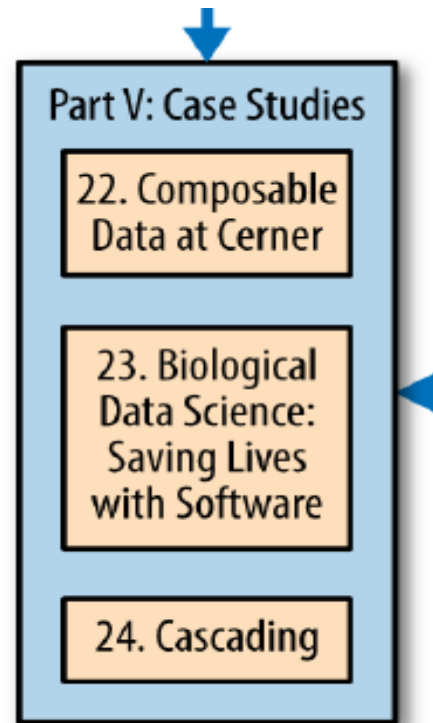
How to read Hadoop The Definitive Guide, 4th Edition



How to read Hadoop The Definitive Guide, 4th Edition



How to read Hadoop The Definitive Guide, 4th Edition



BIG DATA

- ▶ Online Courses.

UC San Diego

- ▶ Coursera Specialization.

- ▶ Introduction to Big Data.
 - ▶ Hadoop Platform and Application Framework.
 - ▶ Introduction to Big Data Analytics.
 - ▶ Machine Learning With Big Data.
 - ▶ Graph Analytics for Big Data.

BIG DATA

- ▶ Another Big Data Books.
 - ▶ Hadoop Operations.
 - ▶ Hadoop Application Architectures.
 - ▶ Learning Spark.
 - ▶ Advanced Analytics with Spark.

BIG DATA

▶ Another Online Courses.

▶ Edx.

▶ Data Science and Engineering with Spark.

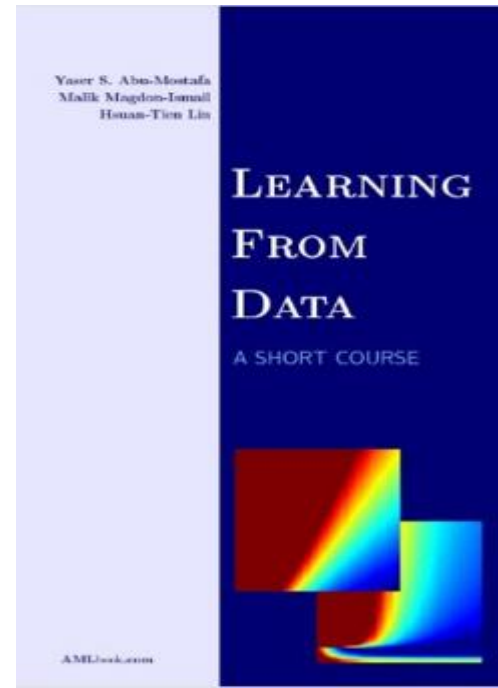
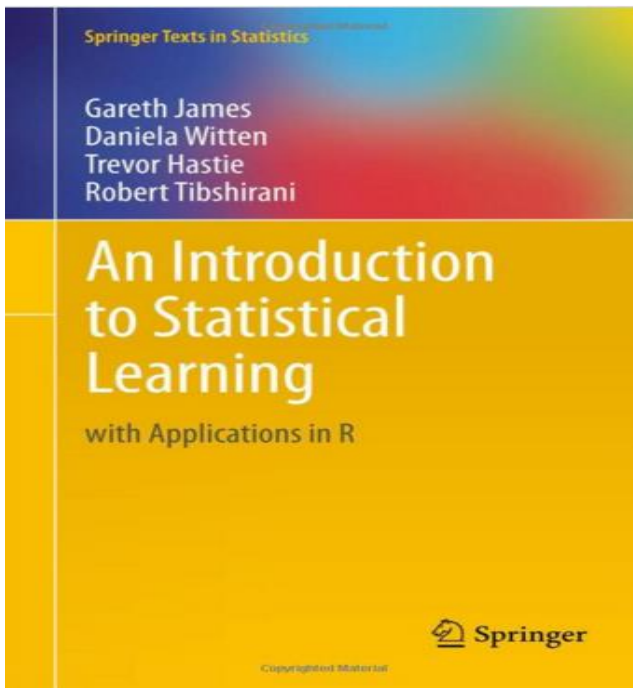
- ▶ Introduction to Spark.
- ▶ Big Data Analysis with Spark.
- ▶ Distributed Machine Learning with Spark.
- ▶ Advanced Distributed Machine Learning with Spark.
- ▶ Advanced Spark for Data Science and Data Engineering.



Data Science

► Books

- Introduction to Statistical Learning With Applications in R by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani.
- Learning From Data by Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin.



Data Science

► Online Courses

► Coursera

- Data Science Specialization by Johns Hopkins University.

- <https://www.coursera.org/specializations/jhu-data-science>

- Machine Learning Specialization University of Washington

- <https://www.coursera.org/specializations/machine-learning>

Data Science

► Data Science Specialization by Johns Hopkins University.

1. The Data Scientist's Toolbox.
2. R Programming.
3. Getting and Cleaning Data.
4. Exploratory Data Analysis.
5. Reproducible Research.
6. Statistical Inference.
7. Regression Models.
8. Practical Machine Learning.
9. Developing Data Products.

Data Science

► Machine Learning Specialization University of Washington

1. Machine Learning Foundations: A Case Study Approach.
2. Machine Learning: Regression.
3. Machine Learning: Classification.
4. Machine Learning: Clustering & Retrieval.
5. Machine Learning: Recommender Systems & Dimensionality Reduction.

Data Science

- ▶ How to deal with Mathematics, Probability and Statistics?

- ▶ Dr. Waleed A. Yousef

<http://www.helwan.edu.eg/university/staff/dr.waleedyousef/HTML/Teaching.html>

- ▶ MA 112 (Discrete Mathematics).
- ▶ ST 121 (Probability and Statistics I).
- ▶ ST 122 (Probability and Statistics II).
- ▶ CS 395 Essential Mathematics for Engineers and Computer Scientists.
- ▶ IT 342 (Pattern Recognition).

