Big Data Engineering In details
From Beginner to Professional

Mostafa Alaa Mohamed
Senior Big Data Engineer
Email: mustafa.alaa.mohamed@gmail.com
Linkedin: Mostafa Alaa [1]

[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

- Graph Applications using Spark

Spark Basics

Spark Basics: Introduction to Spark

- Any Big Data solution working based distributed systems.

Spark Basics: Introduction to Spark

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

Map-Reduce using Spark

- Any Big Data solution working based distributed systems.

Map-Reduce using Spark

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

Spark Programming using RDDs

Spark RDD

- Any Big Data solution working based distributed systems.

Spark RDD

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

Spark Working With Key/Value Pairs

- Any Big Data solution working based distributed systems.

Spark Working With Key/Value Pairs

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

Why we need RDD in Real World Applications

- Any Big Data solution working based distributed systems.

Why we need RDD in Real World Applications

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

Spark Datasets/Dataframe

Introduction to Datasets/Dataframe

- Any Big Data solution working based distributed systems.

Introduction to Datasets/Dataframe

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

- Any Big Data solution working based distributed systems.

Spark SQL

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

Dataframes/Datasets vs. RDDs

- Any Big Data solution working based distributed systems.

Dataframes/Datasets vs. RDDs

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

Spark on Production

Go Production Part 1

- Any Big Data solution working based distributed systems.

Go Production Part 1

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

Go Production Part 2

- Any Big Data solution working based distributed systems.

Go Production Part 2

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

Spark For Data Warehouse

- Any Big Data solution working based distributed systems.

ETL pipeline End-to-End into Spark

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

Spark Streaming

- Any Big Data solution working based distributed systems.

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

- Any Big Data solution working based distributed systems.

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

Spark Structure Streaming

- Any Big Data solution working based distributed systems.

Spark Structure Streaming

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

Spark Streaming Applications

- Any Big Data solution working based distributed systems.

Spark Streaming Applications

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

Spark using other Programming Languages

- Any Big Data solution working based distributed systems.

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

- Any Big Data solution working based distributed systems.

PySpsark Utilizing Python Features

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

RSpark for R Geeks

- Any Big Data solution working based distributed systems.

RSpark for R Geeks

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

Spark For Data Scientist

Data Analysis using Spark

- Any Big Data solution working based distributed systems.

Data Analysis using Spark

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

Machine Leaning using Spark

- Any Big Data solution working based distributed systems.

Machine Leaning using Spark

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

Deep Learning using Spark

- Any Big Data solution working based distributed systems.

Deep Learning using Spark

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

Spark Graph Dataframe/Graphx

Introduction to Spark Graph Dataframe/GraphX

- Any Big Data solution working based distributed systems.

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

- Any Big Data solution working based distributed systems.

Graph Applications using Spark

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

Tuning your Spark Jobs

Tuning Spark Jobs Part 1

- Any Big Data solution working based distributed systems.

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

- Any Big Data solution working based distributed systems.

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

Appendix

Appendix A: Hadoop Map-Reduce

- Any Big Data solution working based distributed systems.

Appendix A: Hadoop Map-Reduce

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

Appendix B: Introduction to Scala

- Any Big Data solution working based distributed systems.

Appendix B: Introduction to Scala

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

- Any Big Data solution working based distributed systems.

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

- Any Big Data solution working based distributed systems.

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

Appendix E: Maven for Automation Cycle

- Any Big Data solution working based distributed systems.

Appendix E: Maven for Automation Cycle

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

- Any Big Data solution working based distributed systems.

Appendix F: How to use Shell Scripts to Automate your tasks

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

Appendix G: Introduction to Oozie for Big Data Development

- Any Big Data solution working based distributed systems.

Appendix G: Introduction to Oozie for Big Data Development

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

Appendix H: Introduction to Deep Learning

- Any Big Data solution working based distributed systems.

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