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

¹Big Data & Analytics Department, Epam Systems

The Definitive Guide to Big Data Engineering Tasks

Table of Contents I

- Spark Basics
 - Introduction to Spark
 - Map-Reduce using Spark
- Spark Programming using RDDs
 - Spark RDD
 - Spark Working With Key/Value Pairs
 - Why we need RDD in Real World Applications
- Spark Datasets/Dataframe
 - Introduction to Datasets/Dataframe
 - Spark SQL
 - Dataframes/Datasets vs. RDDs
- Spark on Production
 - Go Production Part 1
 - Go Production Part 2
- Spark For Data Warehouse
 - ETL pipeline End-to-End into Spark

Table of Contents II

- Spark Streaming
 - Introduction to Spark Streaming
 - Spark Streaming with Kafka
 - Spark Structure Streaming
 - Spark Streaming Applications
- Spark using other Programming Languages
 - PySpsark for Python Geeks
 - PySpsark Utilizing Python Features
 - RSpark for R Geeks
- Spark For Data Scientist
 - Data Analysis using Spark
 - Machine Leaning using Spark
 - Deep Learning using Spark
- Spark Graph Dataframe/Graphx
 - Introduction to Spark Graph Dataframe/GraphX

Table of Contents III

Graph Applications using Spark

- 10 Tuning your Spark Jobs
 - Tuning Spark Jobs Part 1
 - Tuning Spark Jobs Part 2

Appendix

Spark Basics

Spark Basics: Introduction to Spark

Spark Basics: Introduction to Spark

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

Map-Reduce using Spark

Map-Reduce using Spark

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

Spark Programming using RDDs

Spark RDD

Spark RDD

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

Spark Working With Key/Value Pairs

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

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

Introduction to Datasets/Dataframe

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

Spark SQL

Spark SQL

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

Dataframes/Datasets vs. RDDs

Dataframes/Datasets vs. RDDs

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



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

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

Spark For Data Warehouse

ETL pipeline End-to-End into Spark

ETL pipeline End-to-End into Spark

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



Introduction to Spark Streaming

Introduction to Spark Streaming

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

Spark Streaming with Kafka

Spark Streaming with Kafka

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

Spark Structure Streaming

Spark Structure Streaming

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

Spark Streaming Applications

Spark Streaming Applications

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

Spark using other Programming Languages

PySpsark for Python Geeks

PySpsark for Python Geeks

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

PySpsark Utilizing Python Features

PySpsark Utilizing Python Features

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

RSpark for R Geeks

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

Data Analysis using Spark

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

Machine Leaning using Spark

Machine Leaning using Spark

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

Deep Learning using Spark

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

Introduction to Spark Graph Dataframe/GraphX

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

Graph Applications using Spark

Graph Applications using Spark

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



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

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



Appendix A: Hadoop Map-Reduce

Appendix A: Hadoop Map-Reduce

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

Appendix B: Introduction to Scala

Appendix B: Introduction to Scala

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

Appendix C: SQL Syntax and Tips for Data Science

Appendix C: SQL Syntax and Tips for Data Science

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

Appendix D: DWH Concepts and Modeling

Appendix D: DWH Concepts and Modeling

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

Appendix E: Maven for Automation Cycle

Appendix E: Maven for Automation Cycle

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

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

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

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

Appendix H: Introduction to Deep Learning

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