Introduction to Spark
Udacity

Pramod Duvvuri

October, 2019

1. Introduction to Big Data

2. Big Data Ecosystem

3. Numbers Everyone Should Know

4. Distributed Computing

5. Spark

6. Hadoop vs Spark

7. Map Reduce

8. **Modes in Spark**

   (a) Local
   (b) Distributed
        i. Standalone
        ii. YARN
        iii. Mesos

9. Functional Programming

10. Directed Acyclic Graphs (DAGs)

11. Lazy Evaluation

12. Master and Worker Nodes

13. Distributed File Storage

    (a) HDFS
    (b) AWS S3

14. Components of a Spark Program

    (a) SparkContext
    (b) SparkConfig