



Hadoop & Spark - References

Big Data Analytics Techniques and Applications, 2022

Hadoop

- Hadoop Tutorial Videos
 - <https://www.youtube.com/watch?v=iANBytZ26MI>
- This Hadoop tutorial introduce what is Big Data, what is Hadoop, how Hadoop came into existence, what are the various components of Hadoop and an explanation on Hadoop use case.

Hadoop

- Hadoop Tutorial
 - <https://www.tutorialspoint.com/hadoop/index.htm>
- This tutorial provides a quick introduction to Big Data, MapReduce algorithm, and Hadoop Distributed File System.

Spark

- Spark Tutorial Videos
 - https://www.youtube.com/watch?v=_C8kWso4ne4
- There are complete introduction for Spark in these videos, including comparison with Hadoop, Spark's components(core, SQL, MLlib..... etc), RDD, and application examples.

Spark

- Spark Documentation
 - <https://spark.apache.org/docs/latest/>
- Official Spark documentation. It contains more useful examples, and will be more convenient when we need to check some usages.

PySpark

- PySpark Tutorial Videos
 - <https://www.youtube.com/watch?v=znBaI3Earms>
- The videos are pyspark tutorials. The author will use codes to introduce complete data analysis workflow, from data preprocessing to modeling.

PySpark

- PySpark Tutorial For Beginners
 - <https://sparkbyexamples.com/pyspark-tutorial/>
- There are complete introduction for pyspark in the tutorial, and it provides sample examples in Python code to introduce pyspark functions.

PySpark

- PySpark Tutorial
 - <https://www.tutorialspoint.com/pyspark/index.htm>
- There are some simple introductions for pyspark in the tutorial, and codes examples in each section.