# Spark setup

**The best materials:**

* [www.cloudduggu.com](https://www.cloudduggu.com/spark/installation-multi-node/) - spark multi node set up with hadoop.
* [www.confessionsofadataguy.com](https://www.confessionsofadataguy.com/create-your-very-own-apache-spark-hadoop-cluster-then-do-something-with-it/) - Setup of both Hadoop and Spark. there is no info about on which nodes (servers) execute which commands in terminal for setting up hadoop. But the part for Spark seems fine.

**Other materials:**

* [youtube](https://www.youtube.com/watch?v=f_XsaYcETnI) – spark multi node set up on hadoop
* [youtube](https://www.youtube.com/watch?v=-5TSKMXAygc) – spark multi node set up. I am not sure if it is using hadoop.
* [data-flair.training/blogs](https://data-flair.training/blogs/install-apache-spark-multi-node-cluster/) - spark multi node set up on hadoop

# Spark theory

* [youtube - Data Engineering](https://www.youtube.com/watch?v=Tyg1FVNq40g&list=PLGhXxbu7qYooyn_aWk1DqpIF1CjBzaSUn&index=3) – Hadoop and Spark (9h video)
* [youtube - Data Engineering](https://www.youtube.com/watch?v=rsOSrEbK7sU&list=PLLa_h7BriLH0FzTY5aBFpH-vciOiEf4Br) - Spark

# Submitting Spark jobs

* <https://www.youtube.com/watch?v=o_pne3aLW2w>
* <https://www.youtube.com/watch?v=L3VuPnBQBCM>
* <https://learn.microsoft.com/en-us/azure/hdinsight/spark/spark-best-practices>
* <https://stackoverflow.com/questions/53344285/is-there-a-way-to-submit-spark-job-on-different-server-running-master>