**Project and Labs Tools and Set Up**

# **Cloud Computing**

**Parallel Computing with Cloud -- Potential Platforms:**

Cloud computing engines such as Spark and Hadoop and many other related engines and libraries facilitate parallel computing on Cloud.

For your projects you can use:

1) the Docker Containers distributed to students for Spark and Hadoop.

2) You may open accounts to our own Amarel Lab, log in remotely, and be able to use Spark and many other tools for your project.

<https://oarc.rutgers.edu/resources/amarel/#access>

3) You can use Elastic Map Reduce (<https://aws.amazon.com/elasticmapreduce/>) from amazon to start experimenting with Hadoop and Spark. Both of these tools are installed on EMR and by reading through amazon tutorials you will learn how to launch a cluster on EMR and start your project with Hadoop or Spark. It is preferable that you first install these engines on your local machine, you can follow the below links to do so:

For **hadoop**, <http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/>

For **spark**, <http://spark.apache.org/docs/latest/building-spark.html>

Be careful, to build spark, you should have Maven 3.3.3 or newer and Java 7+."

**You may also visit the following links**:

<https://aws.amazon.com/articles/Elastic-MapReduce/4926593393724923>

<http://ampcamp.berkeley.edu/big-data-mini-course/>

<http://people.csail.mit.edu/matei/papers/2010/hotcloud_spark.pdf>

<https://www.engpaper.com/2019-papers.htm>

4) You may install Spark and Hadoop on your own machine from scratch. This can work better if your OS in Linux. If you would like to get instructions on how to do this please check out the corresponding documents posted on CANVAS under the Project-Lab-Tools Module.