Problems in the work process 2

- ➢ **Cluster:**
  We needed to re-create and rearrange the cluster, change the master and create the workers. In the process as we ran our code in python via cluster the workers could not find the path for the file so we used hdfs and transferred our code there so that it is in a common for the whole cluster space.

- ➢ **Code, syntax-imports:**
  We changed many codes in an attempt to solve the task. We all had problems with imports in the spark so we needed to install them there as well. We encountered problems with the lamda function process in python and how it reads and processes higgs. Also when we imported HIGGS.csv to read it in the code and there we had problems in how to find the right path.

- ➢ **RDD-Hdfs issues:**
  Here our problems were that we had to empty space so that HIGGS.csv could accept and read it. Also even though HIGGS.csv was in hdfs again the code could not find and read them until we could find the right code.

- ➢ **Caching-persistance-parsing:**
  Our biggest problem was this space so we had to empty the master engine and hdfs, but we had to do cache () and parsing while reading HIGGS.csv and editing it.