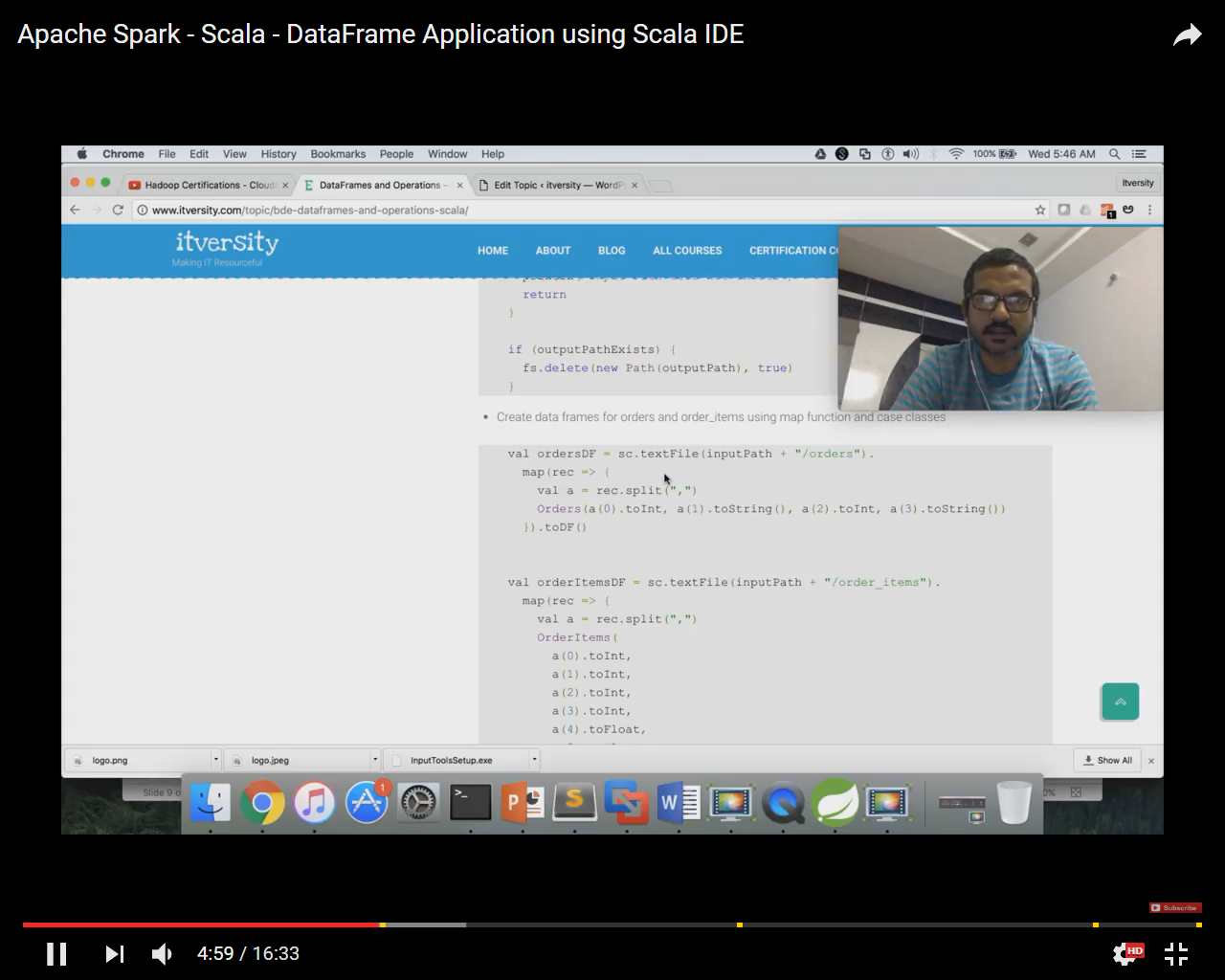
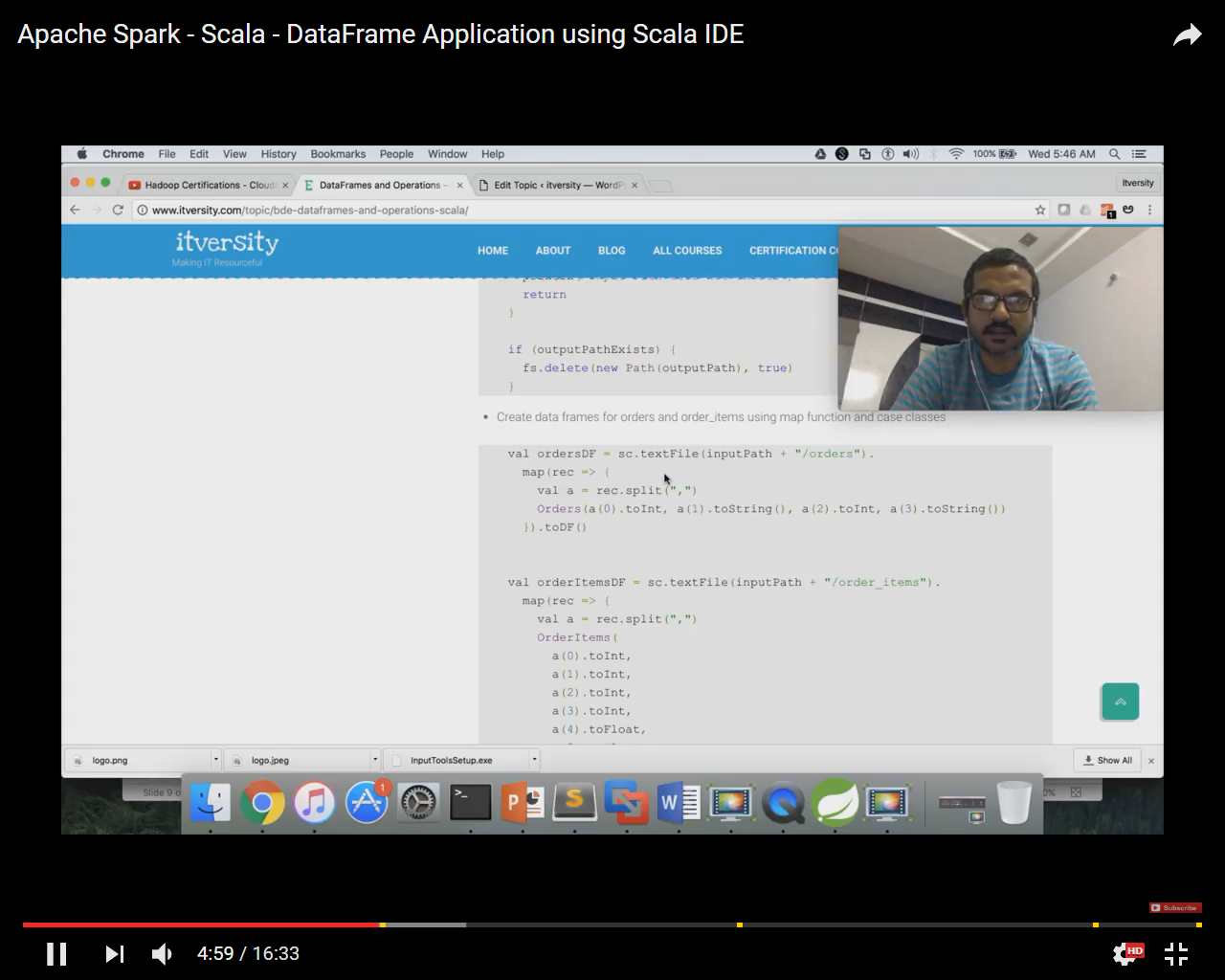
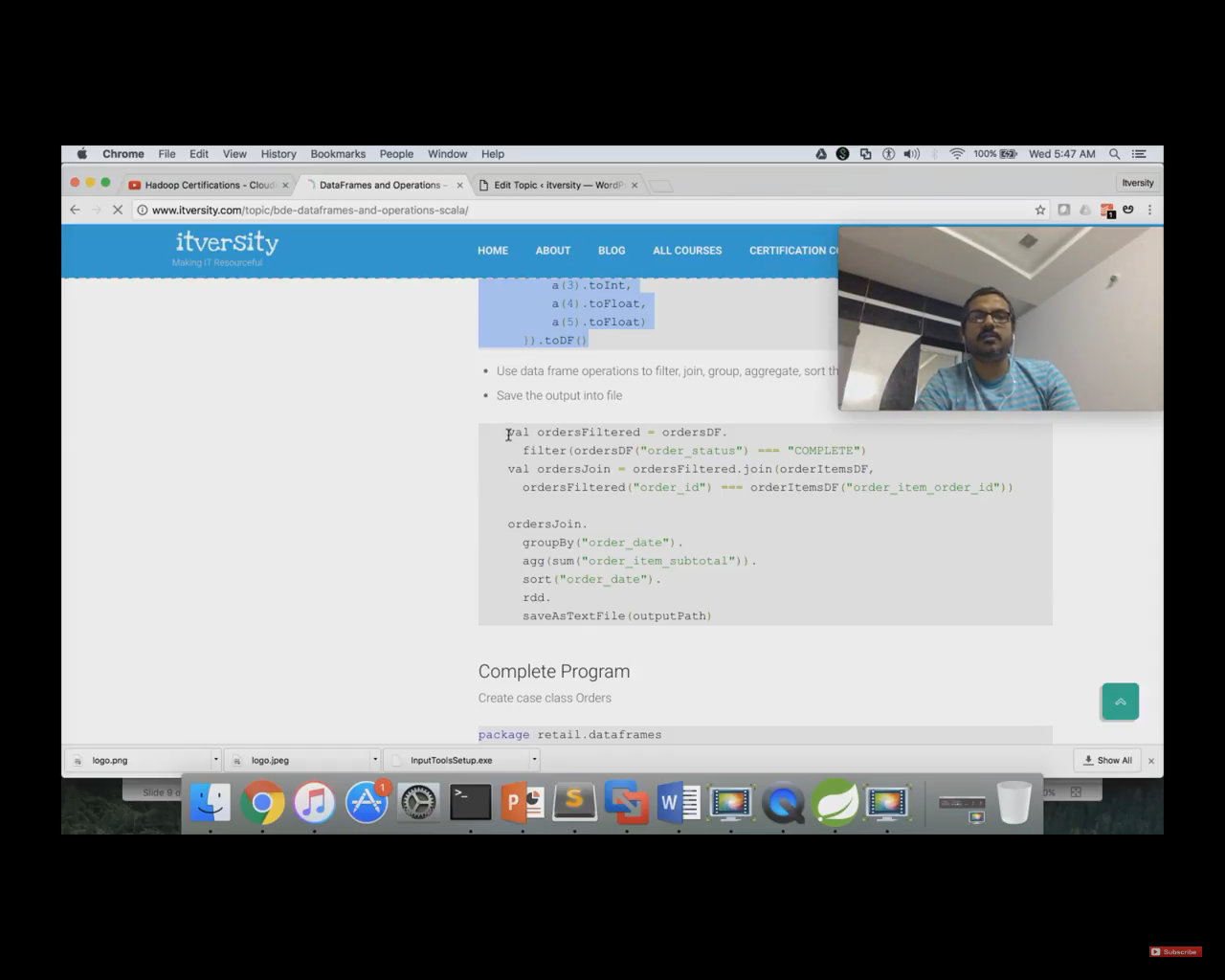


We cant use return statement in spark-shell so remove it if you are trying to run this on scala console

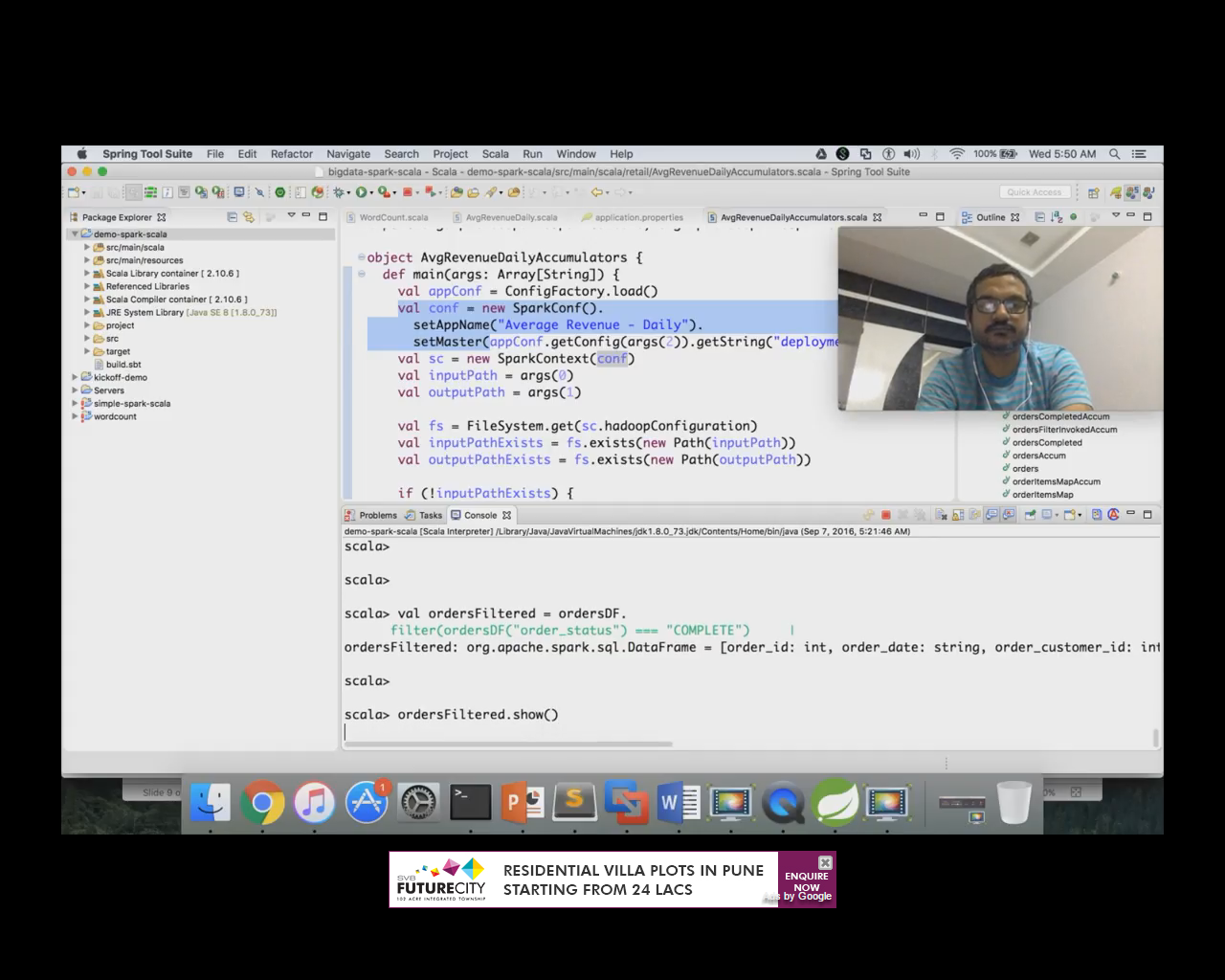


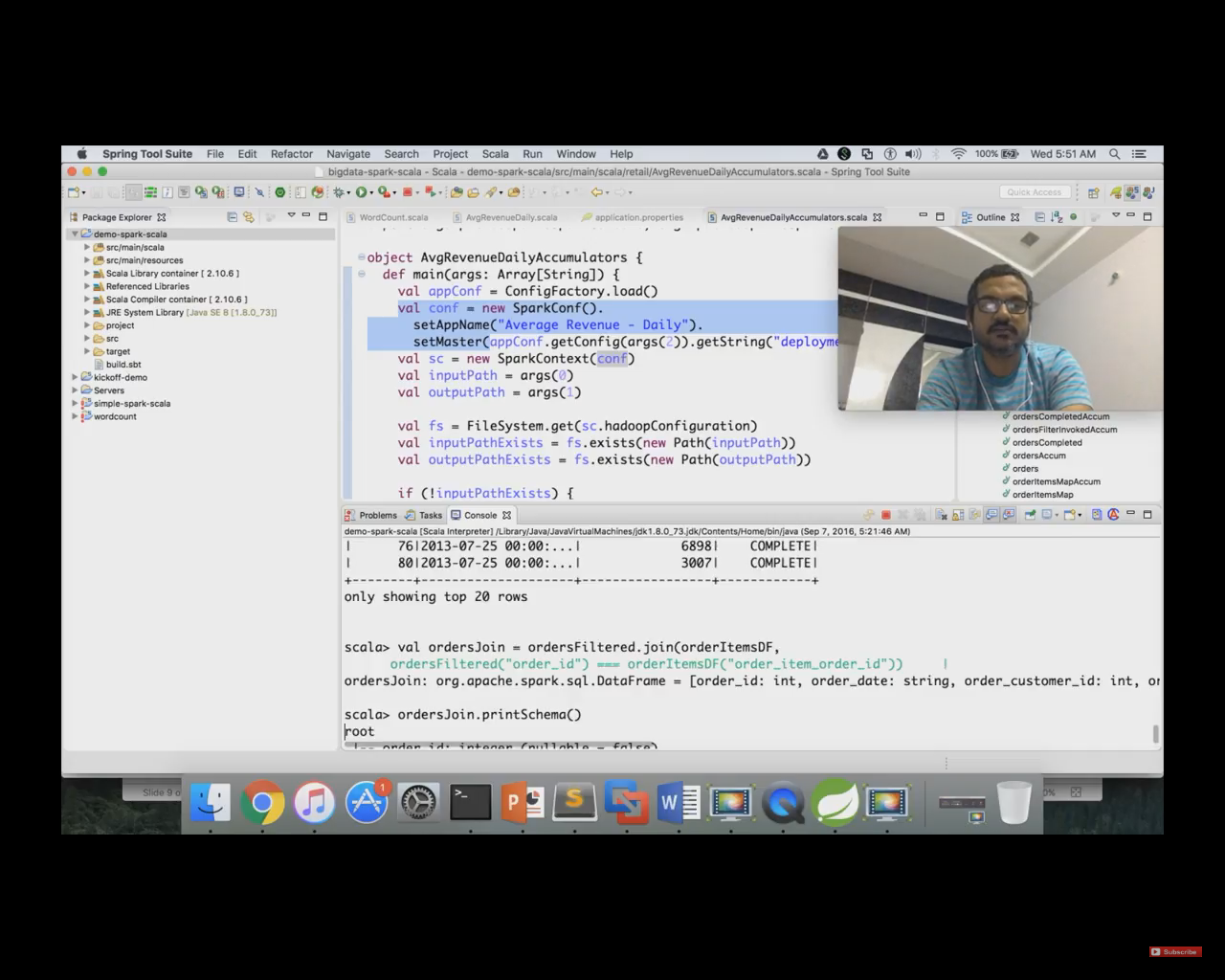


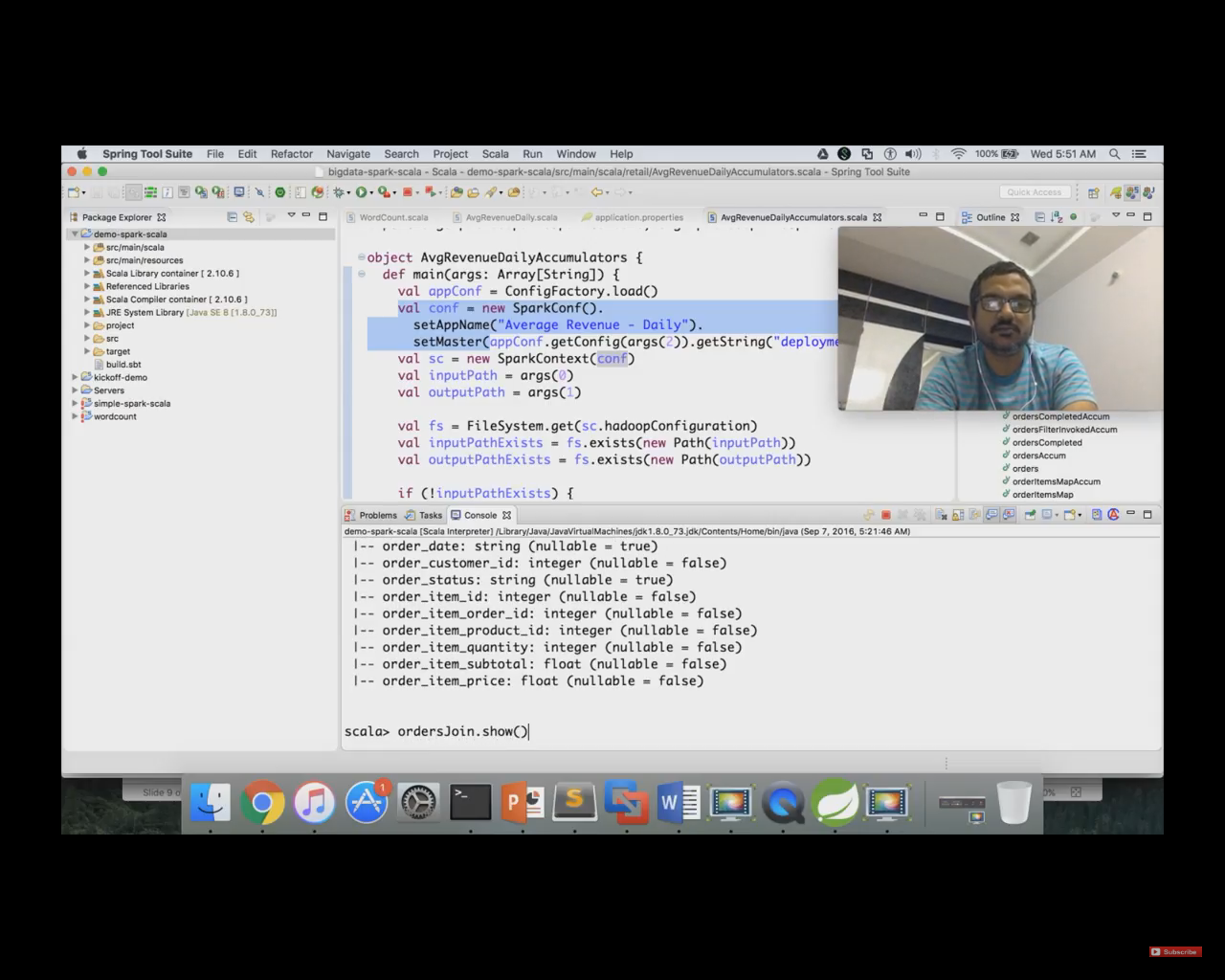


All the operation performed on dataframe creates other dataframes

After aggregation convert datframe to rdd, without converting to rdd we need to use dataframe based operation to save as text file, this is the simplest way to save as text file







https://github.com/dgadiraju/code/blob/master/hadoop/edw/cloudera/spark/spark\_demo\_scala.txt