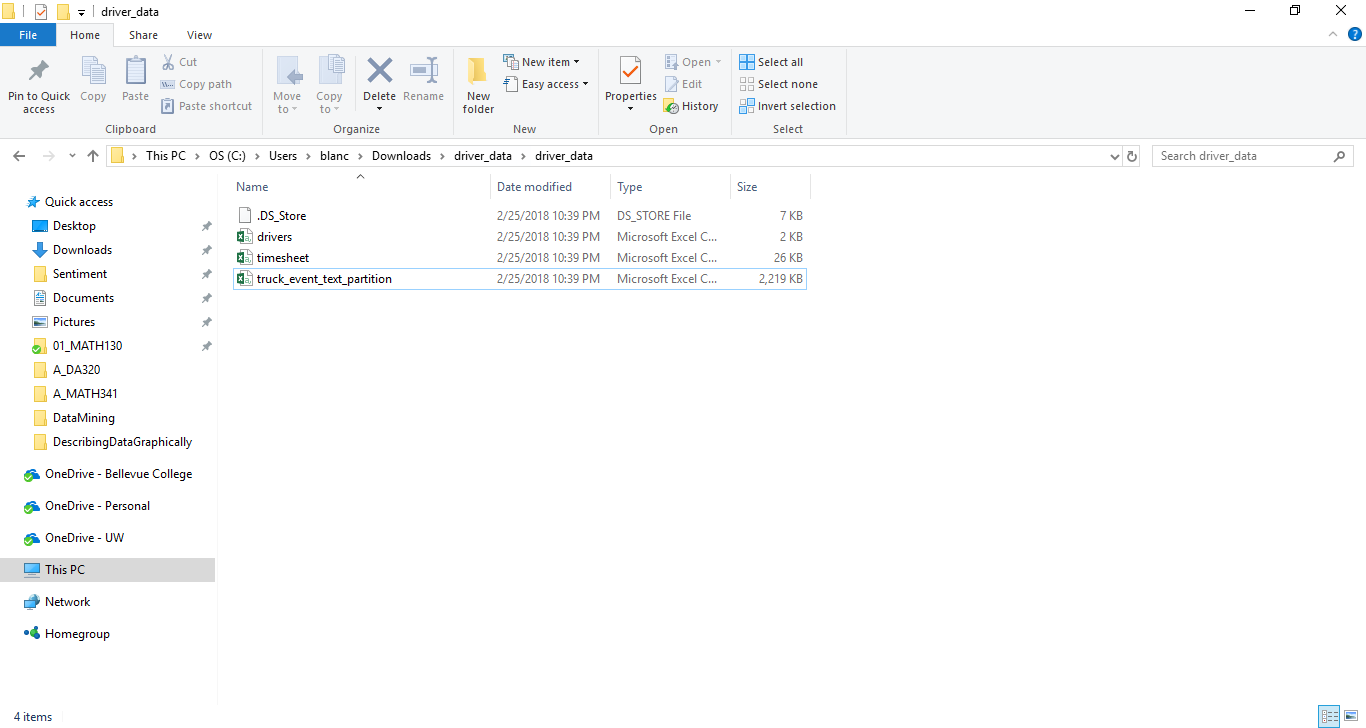
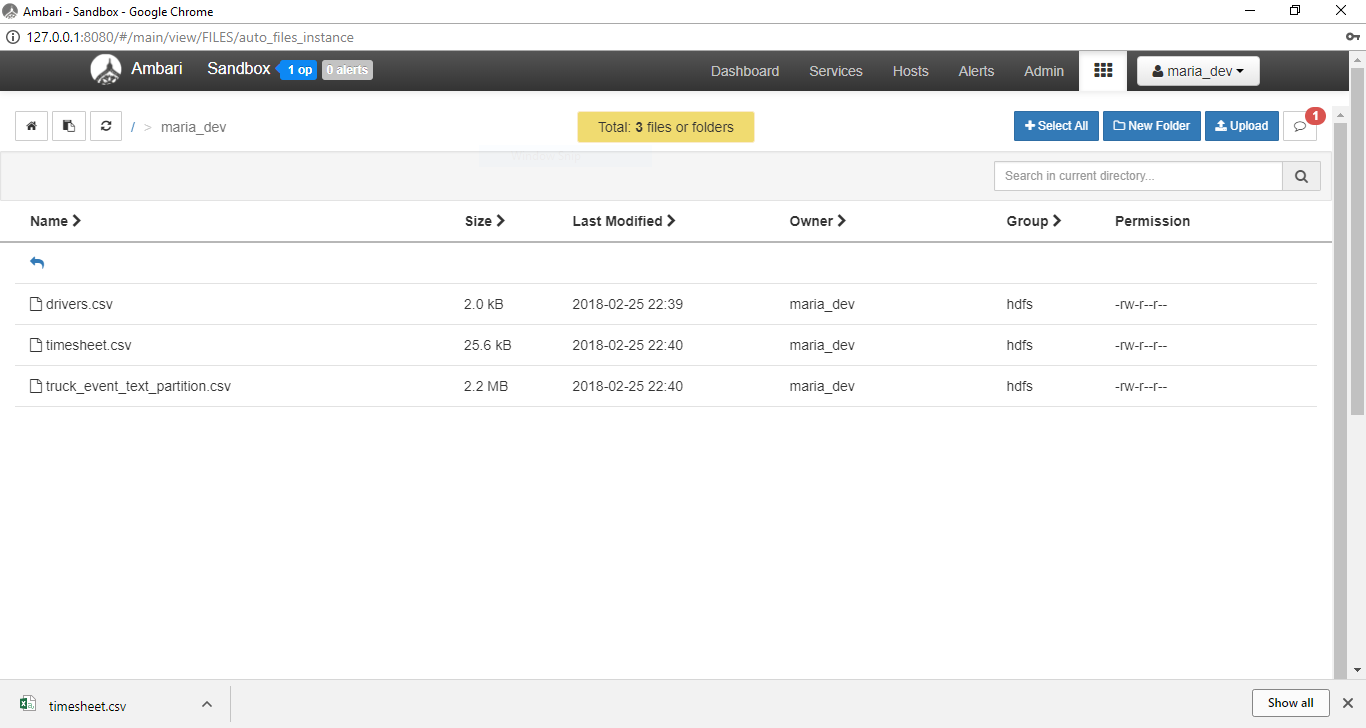
DA 320 Project 8 W. Li

**Part 1**: Complete the basic Pig commands tutorial found at [http://hortonworks.com/hadooptutorial/how-to-use-basic-pig-commands/.](http://hortonworks.com/hadoop-tutorial/how-to-use-basic-pig-commands/) In order to complete the tutorial, you need to download Hortonworks Sandbox found at [http://hortonworks.com/products/hortonworks-sandbox/.](http://hortonworks.com/products/hortonworks-sandbox/) Make sure you include a print screen after completing each step, also submit the two results in text files, in **step 10** (part-r-00000.txt) and **step 13**.

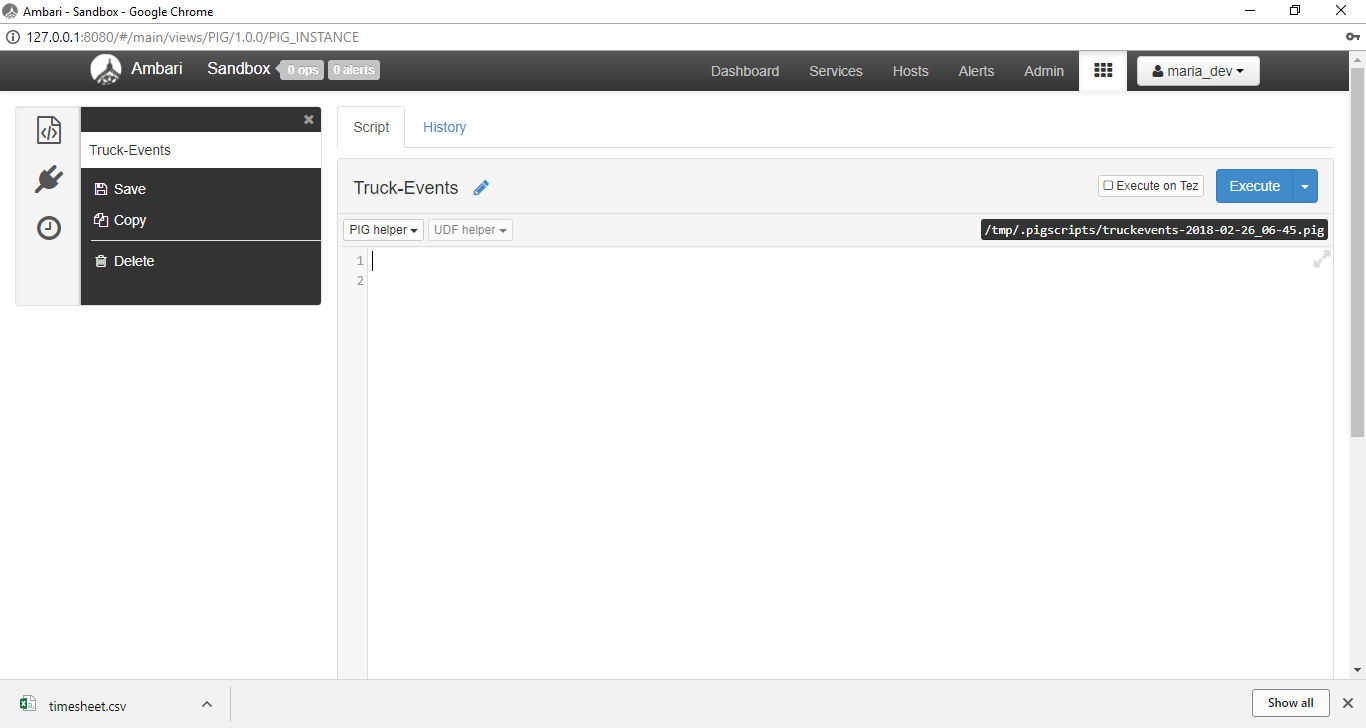
STEP 1: DOWNLOAD THE DATA



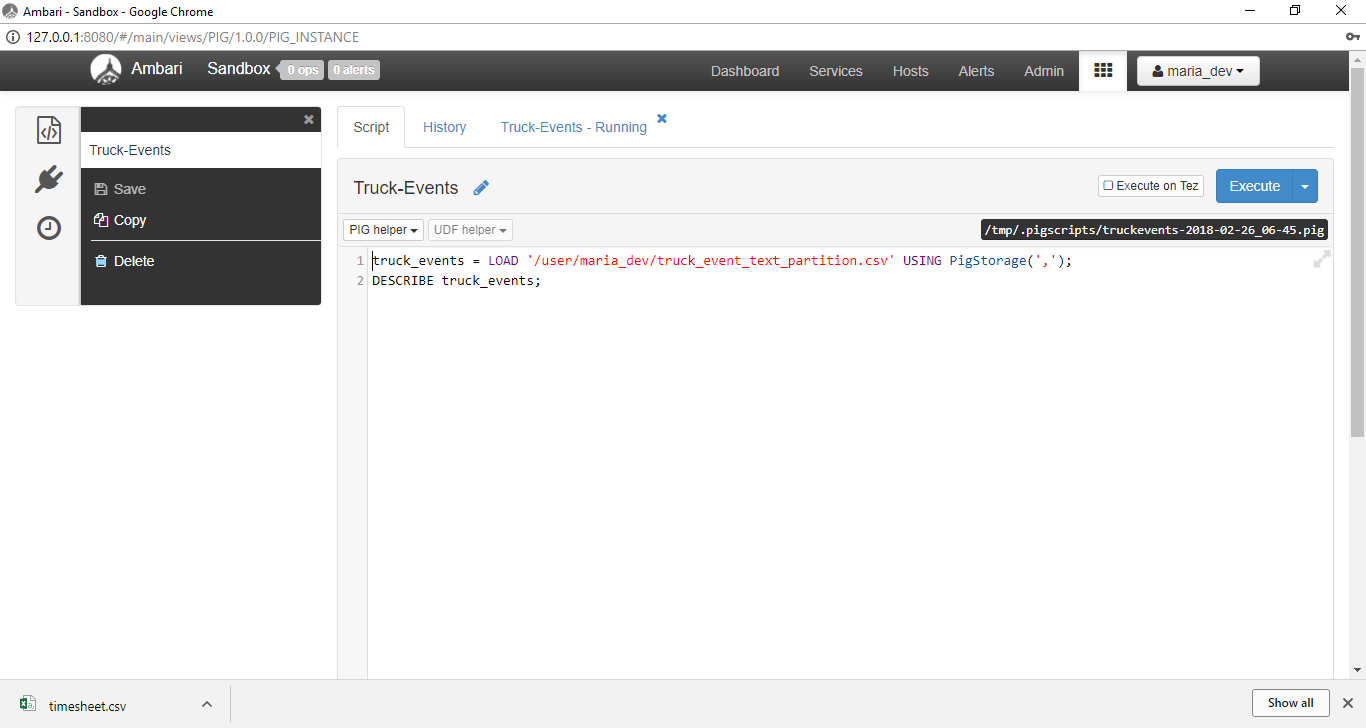
STEP 2: UPLOAD THE DATA FILES



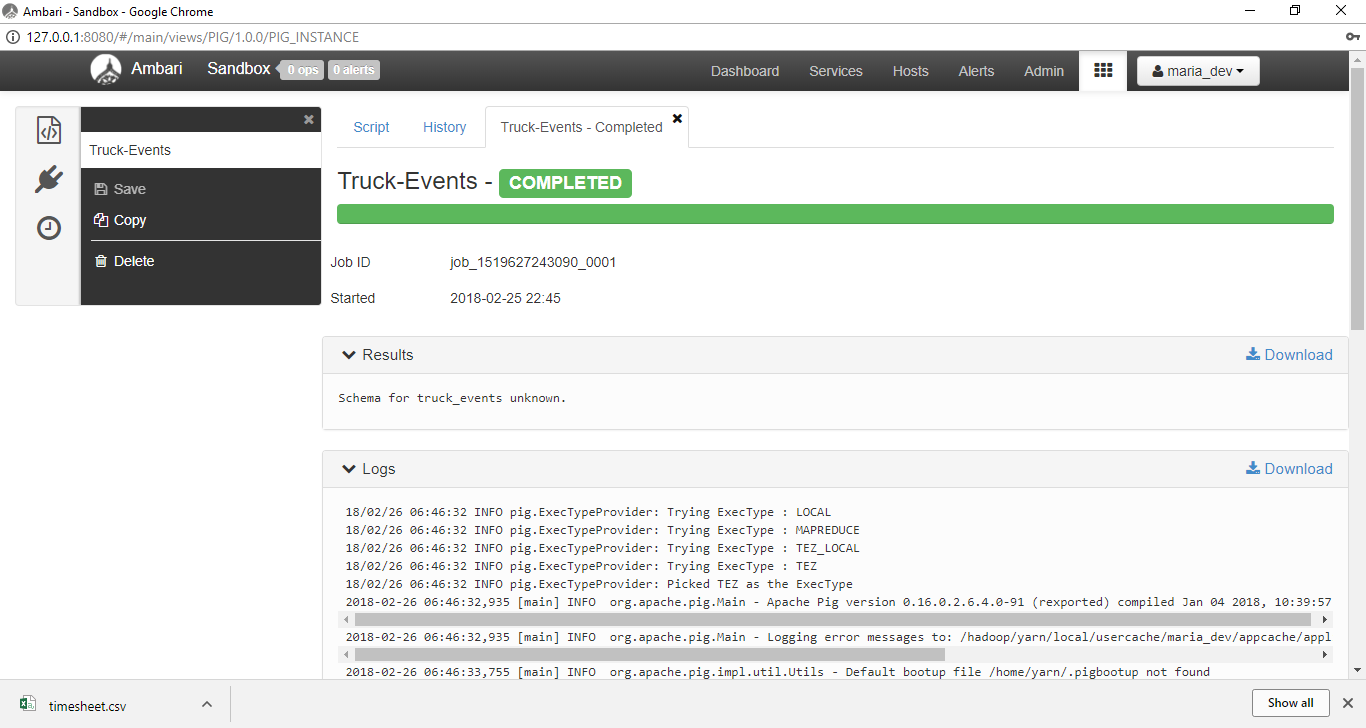
STEP 3: CREATE YOUR SCRIPT



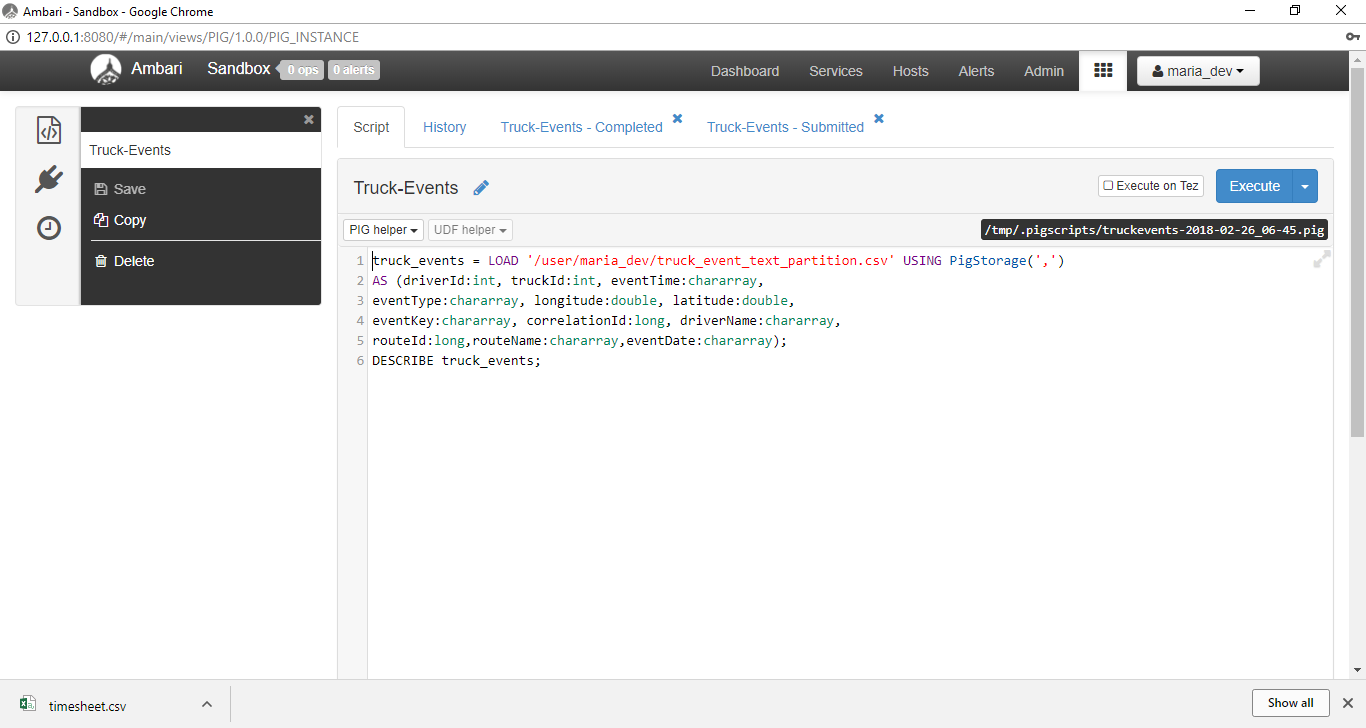
STEP 4: DEFINE A RELATION

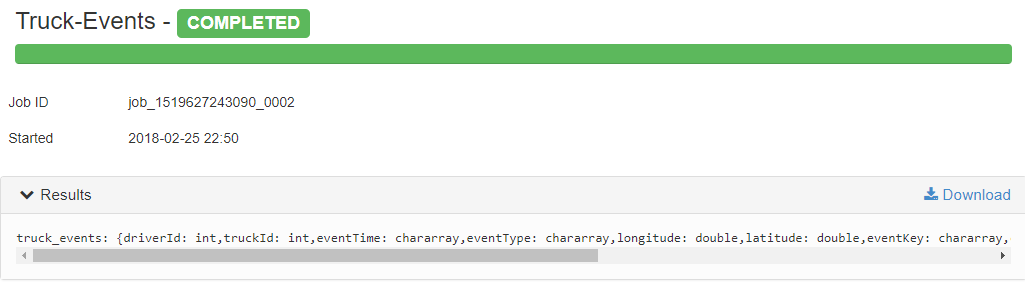


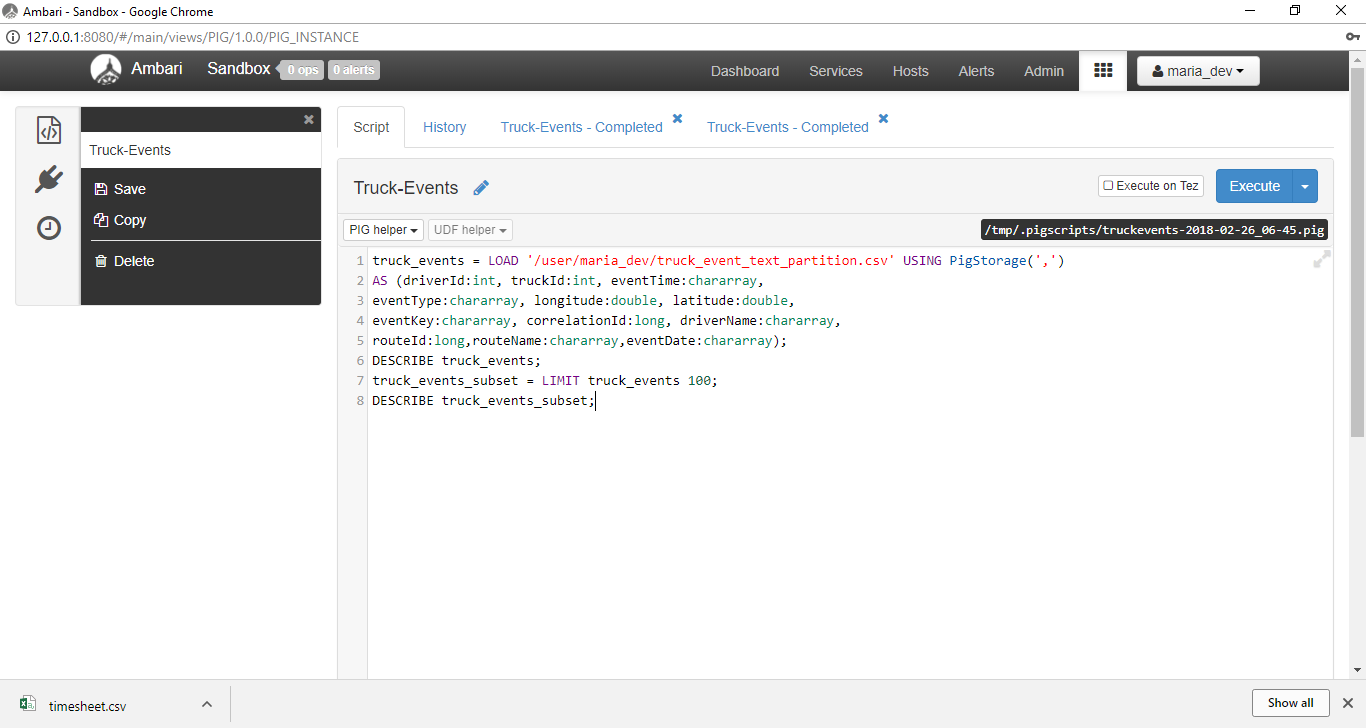
STEP 5: SAVE AND EXECUTE THE SCRIPT

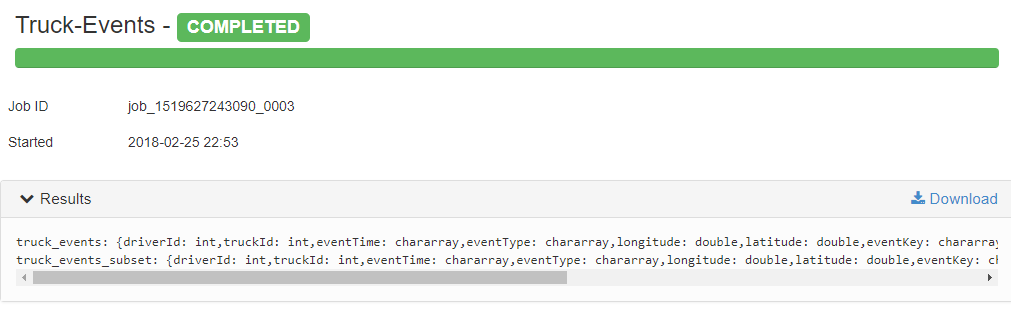


STEP 6: DEFINE A RELATION WITH A SCHEMA

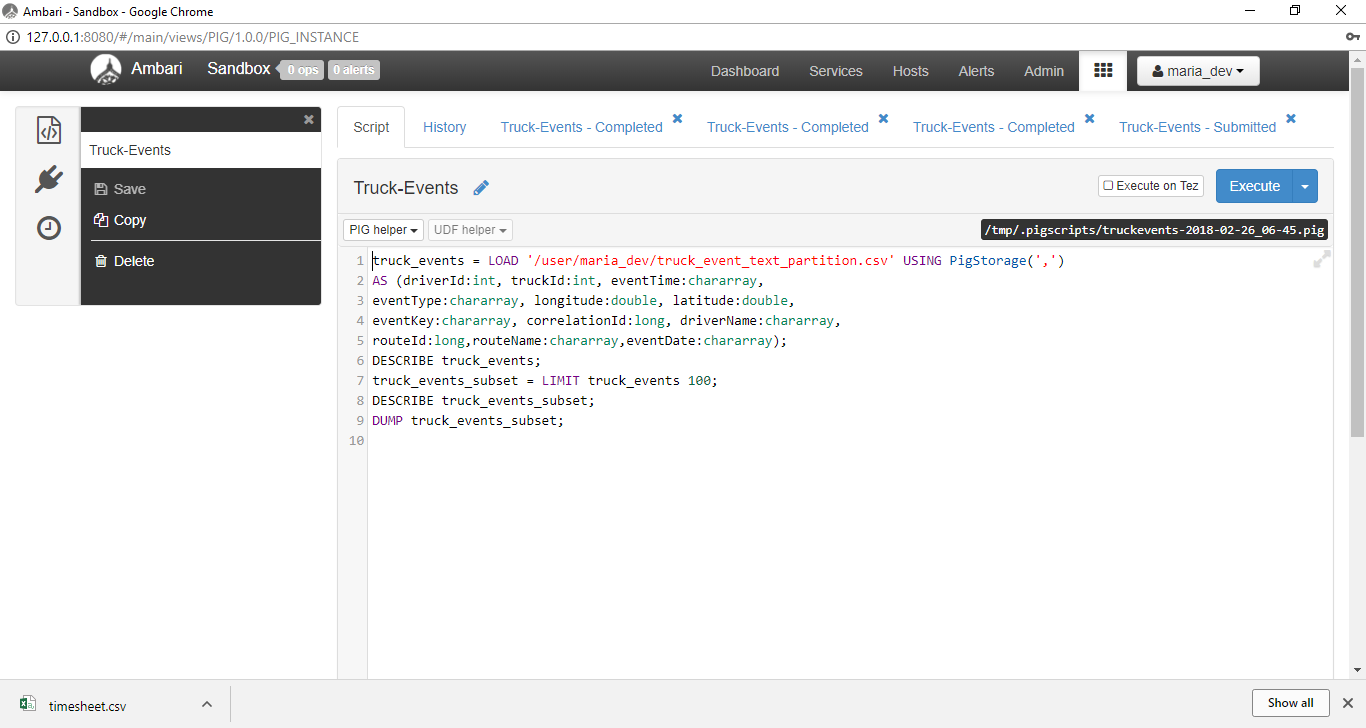


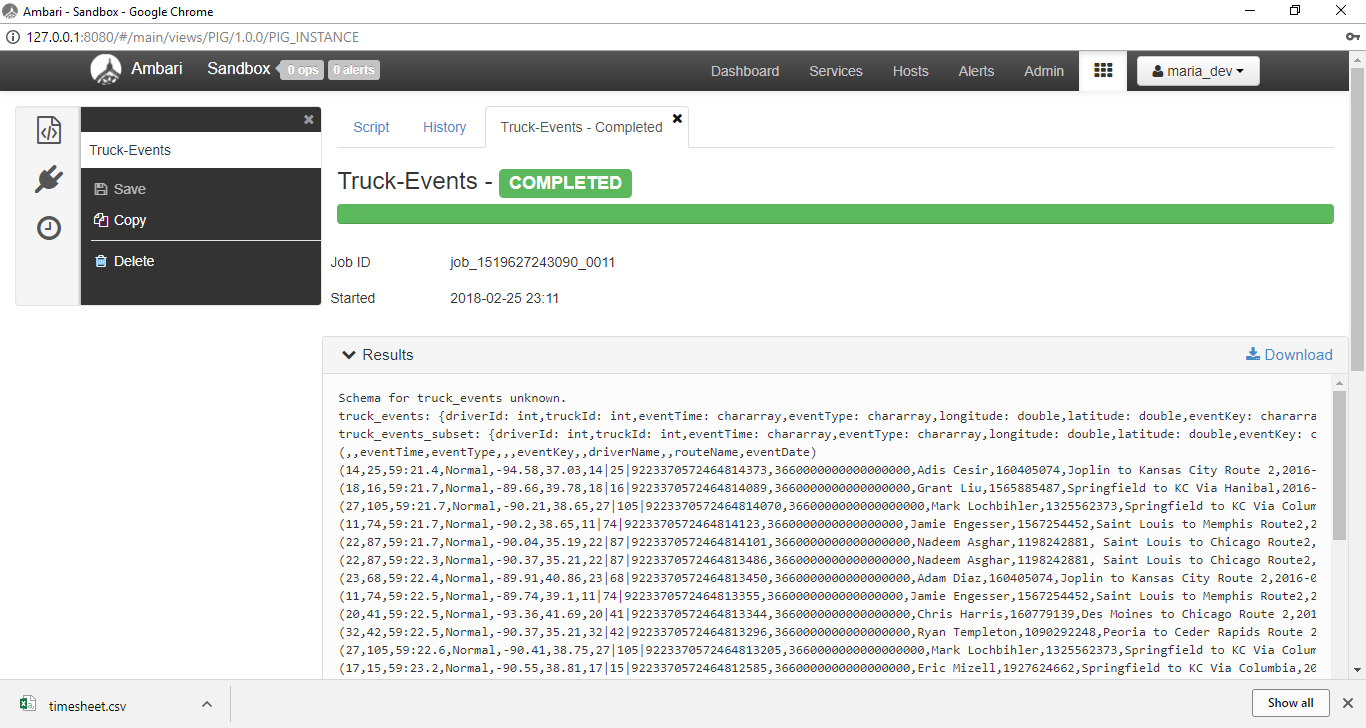




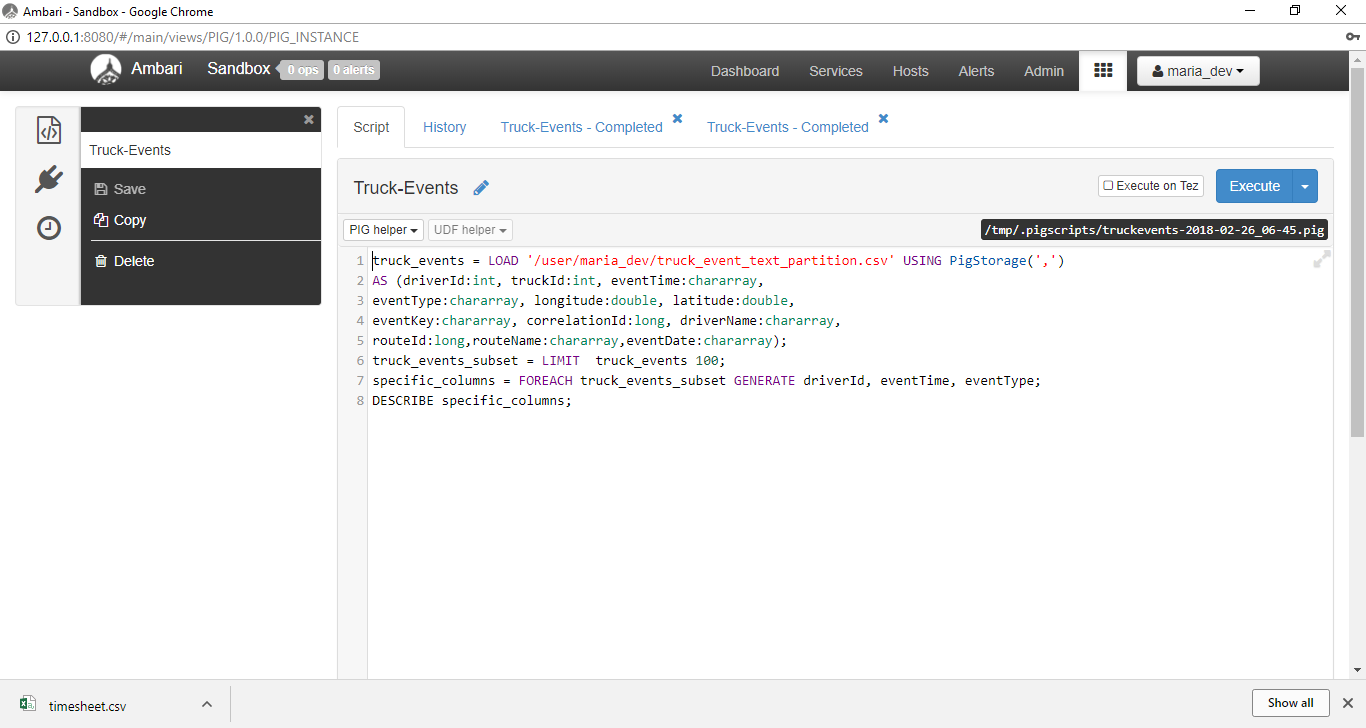


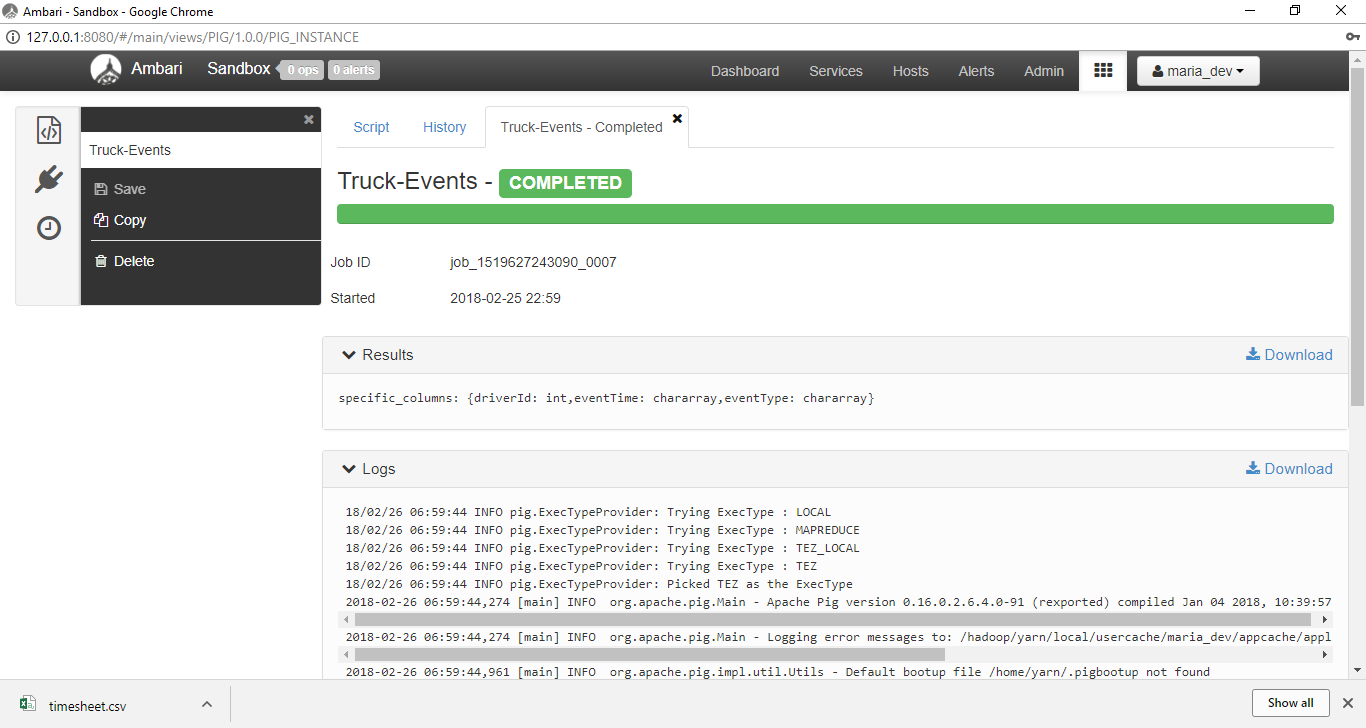
STEP 8: VIEW THE DATA



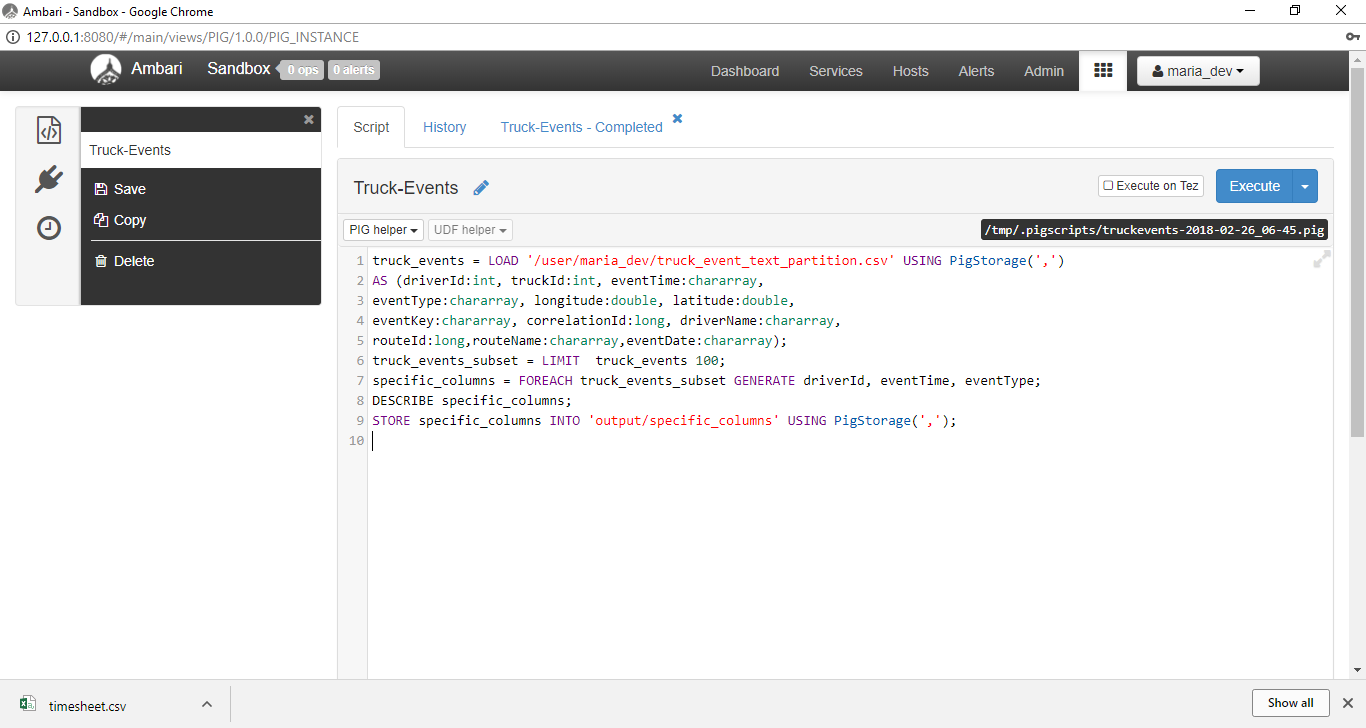


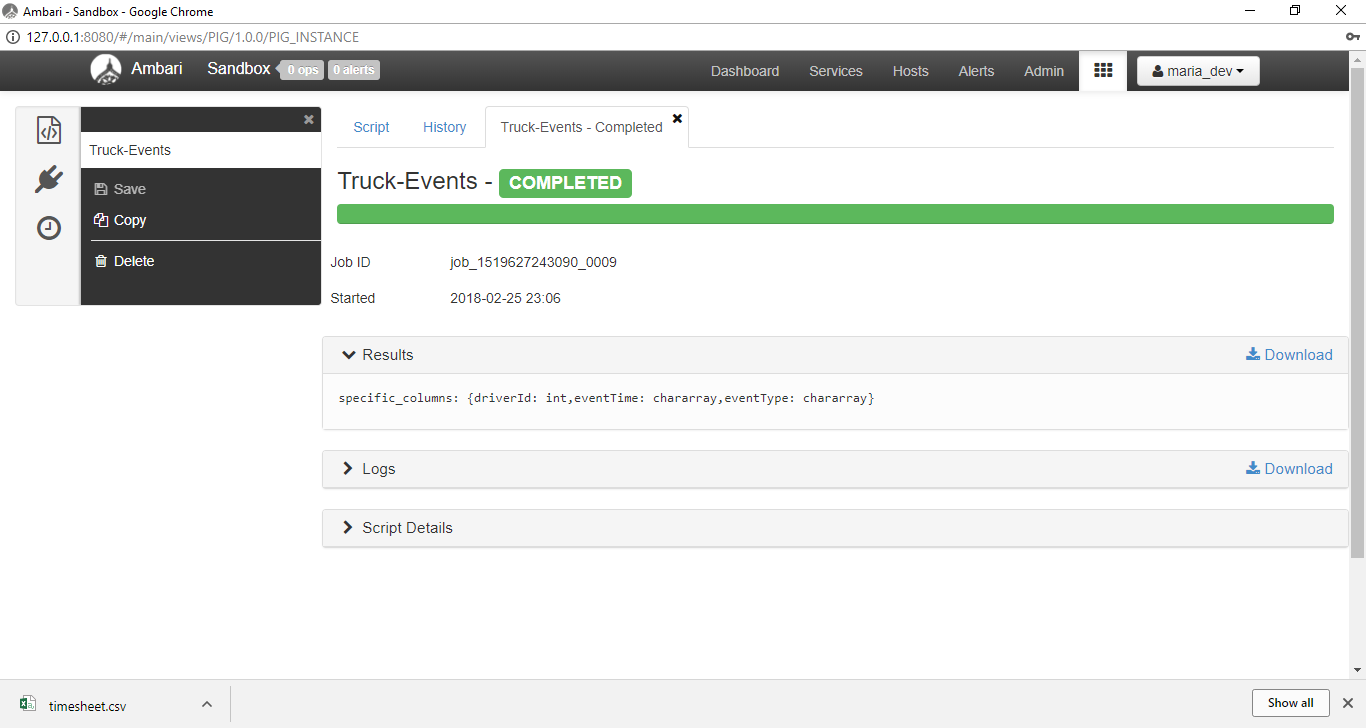
STEP 9: SELECT SPECIFIC COLUMNS FROM A RELATION

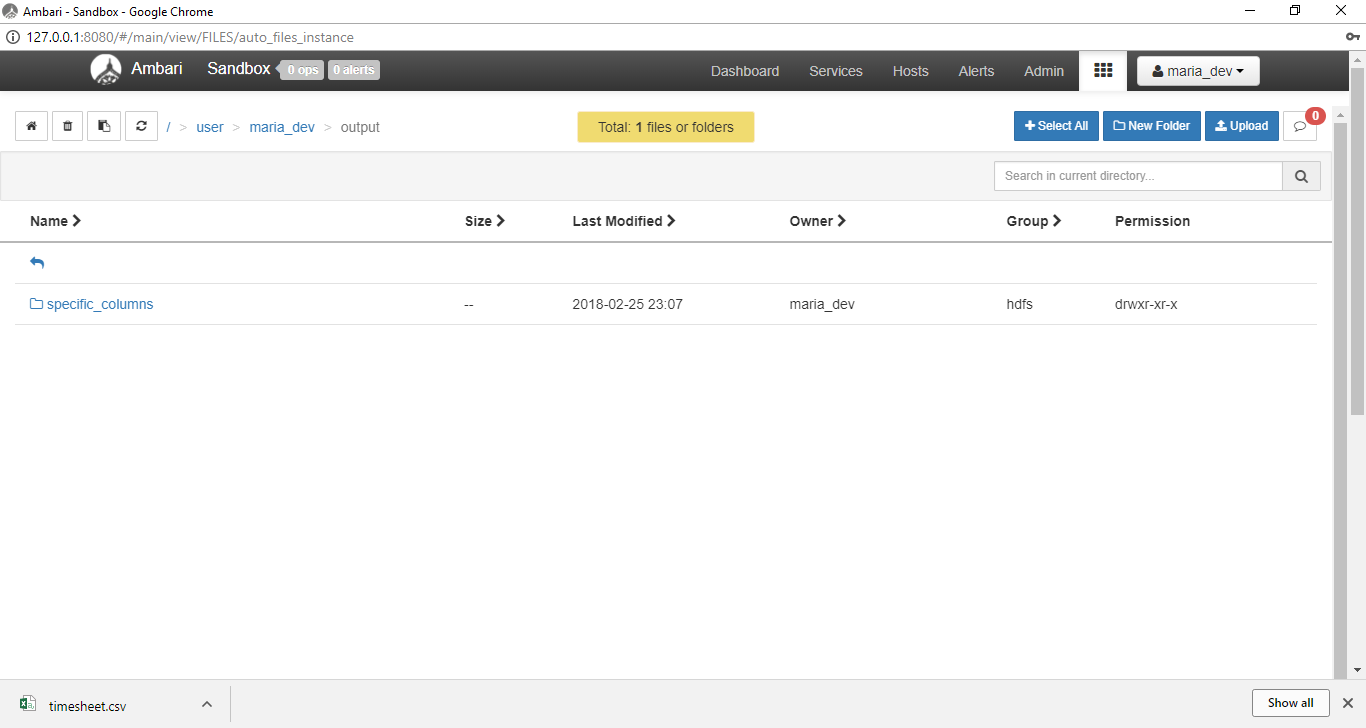


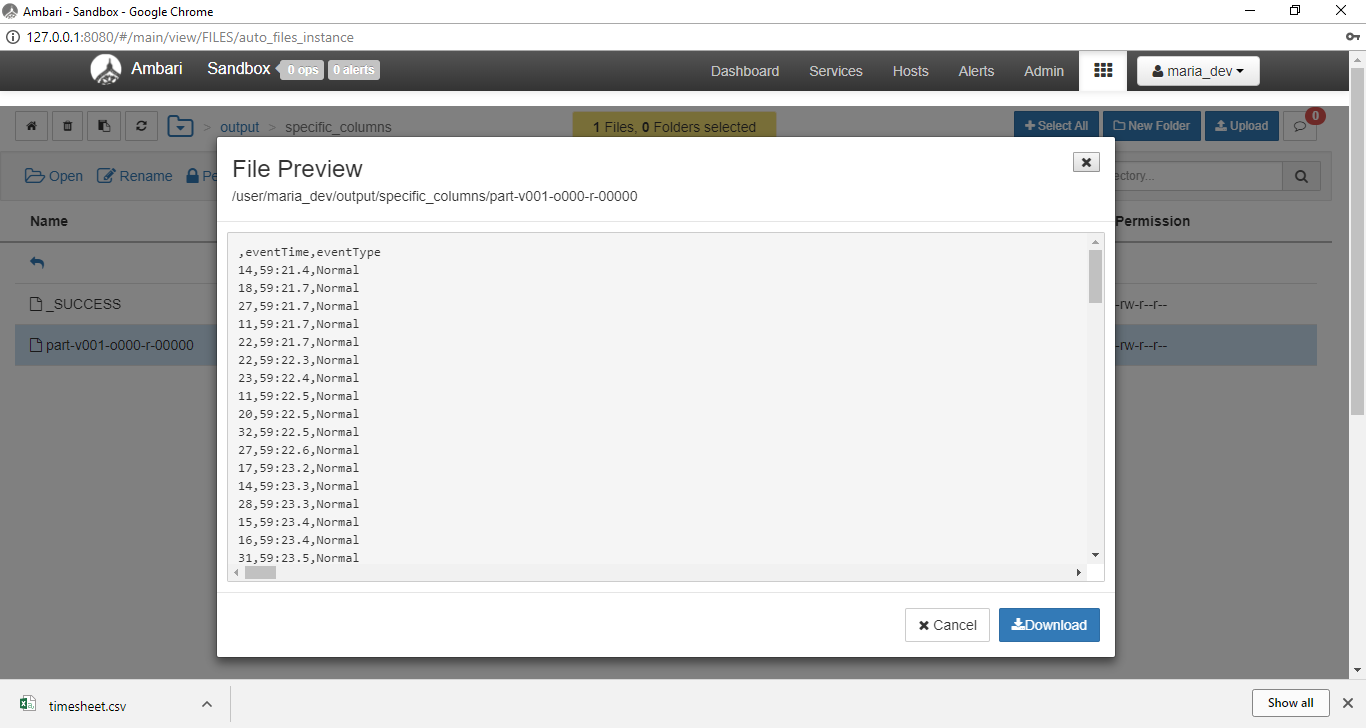


STEP 10: STORE RELATIONSHIP DATA INTO A HDFS FILE

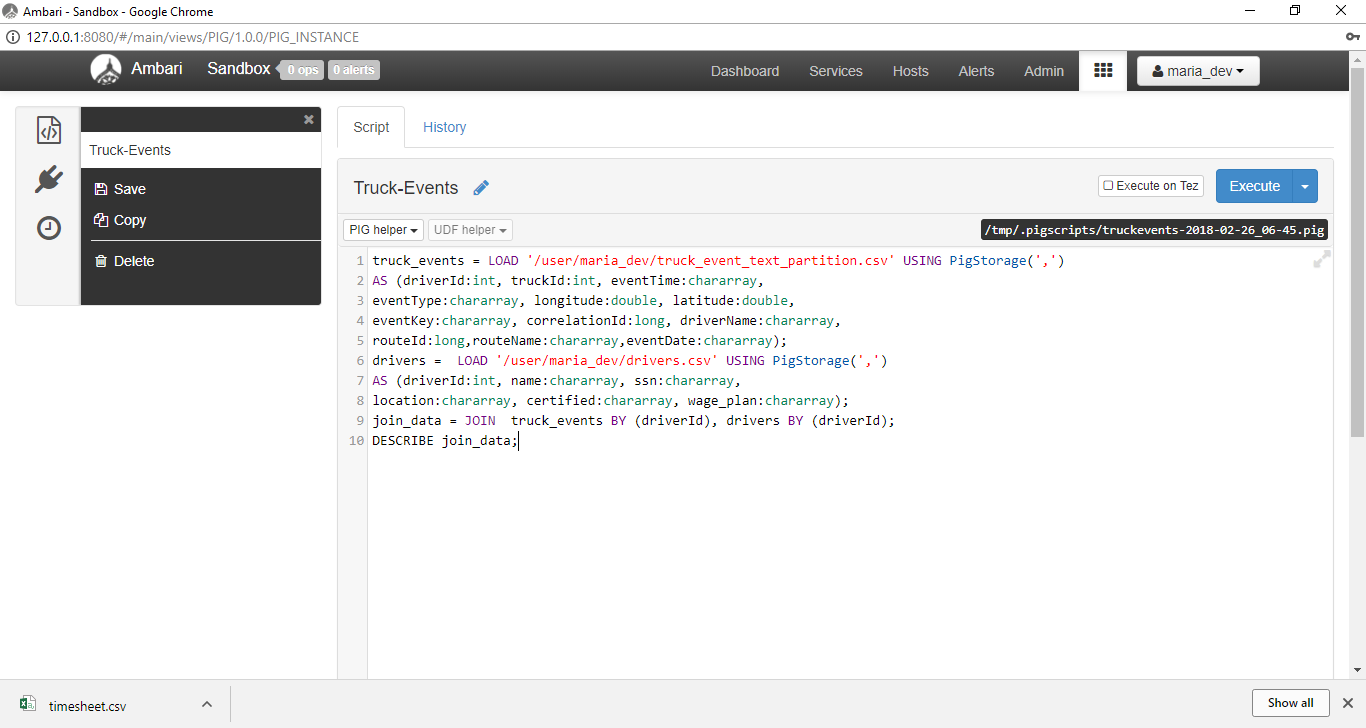


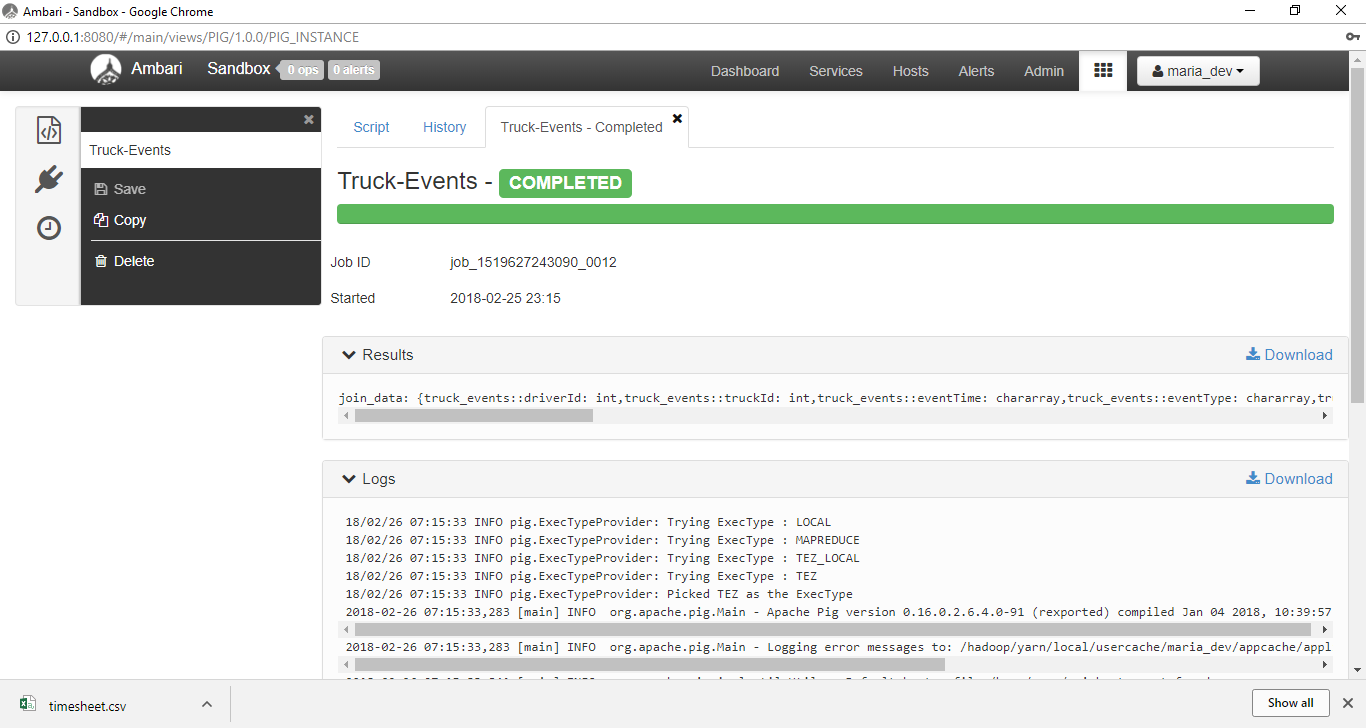


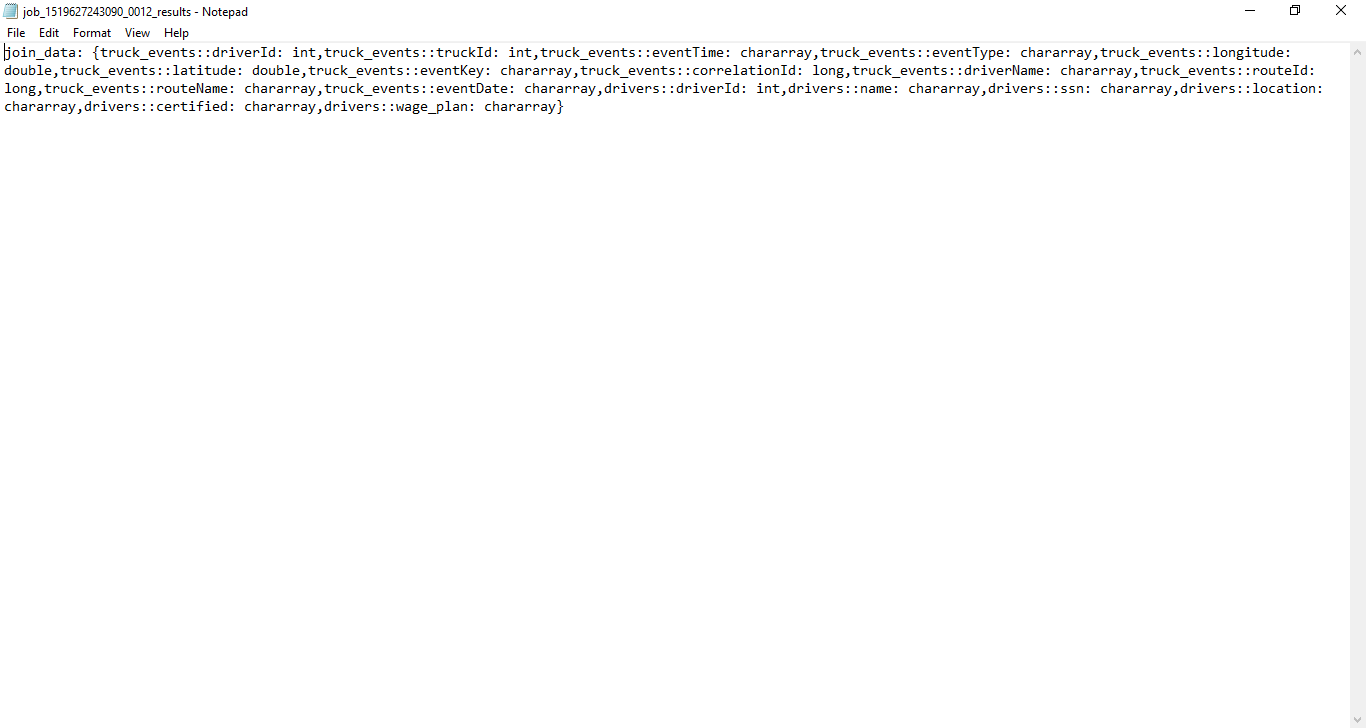




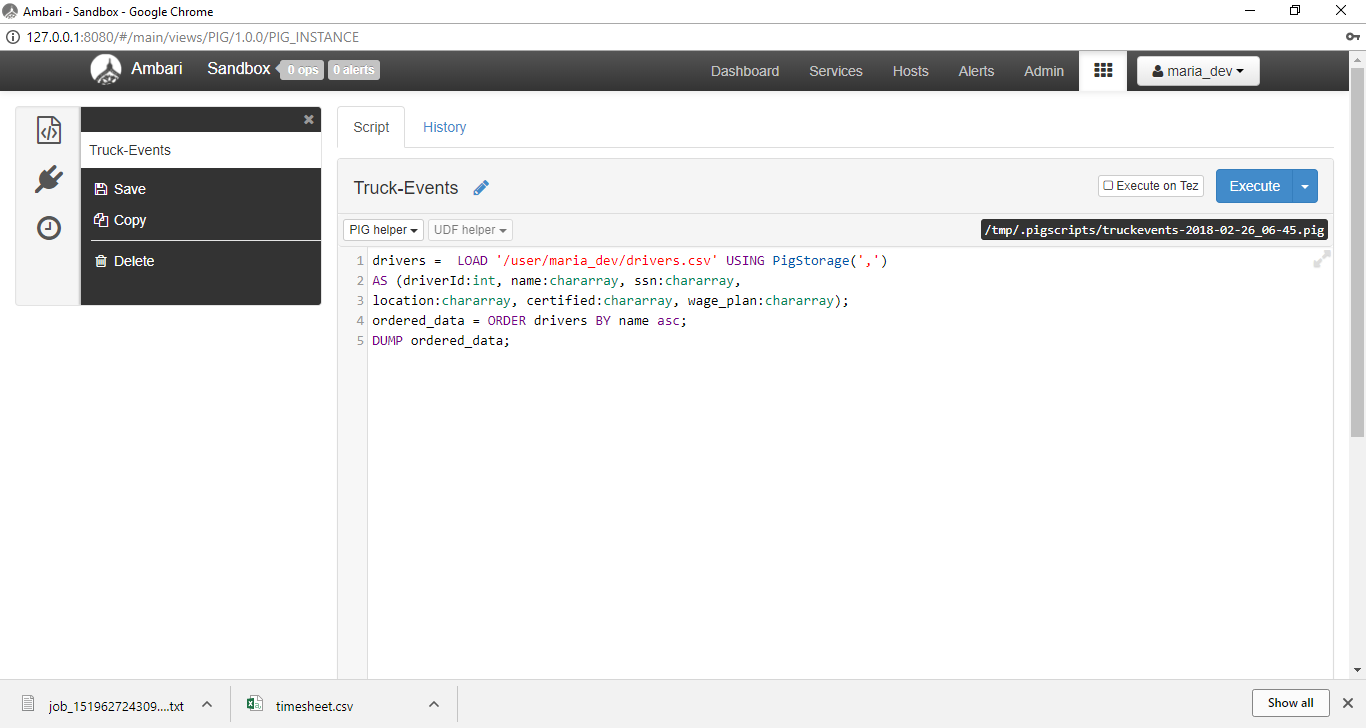
STEP 11: PERFORM A JOIN BETWEEN 2 RELATIONS

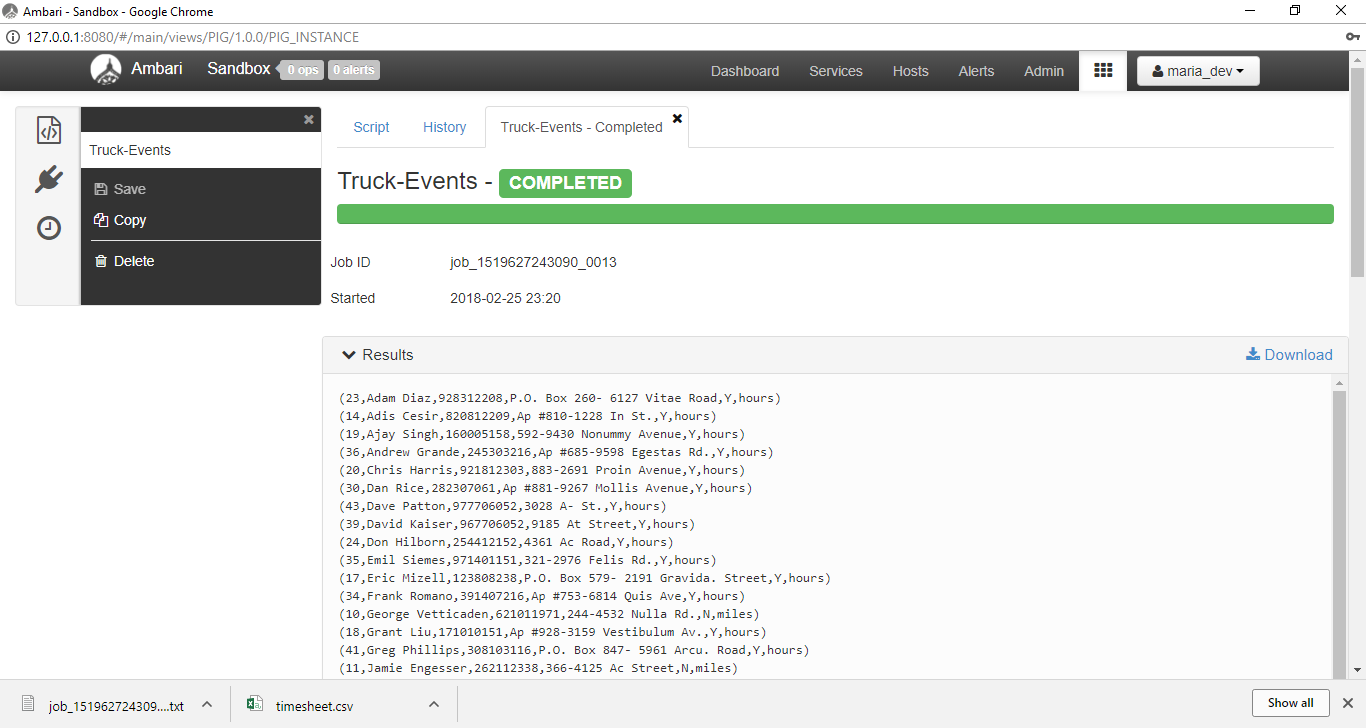




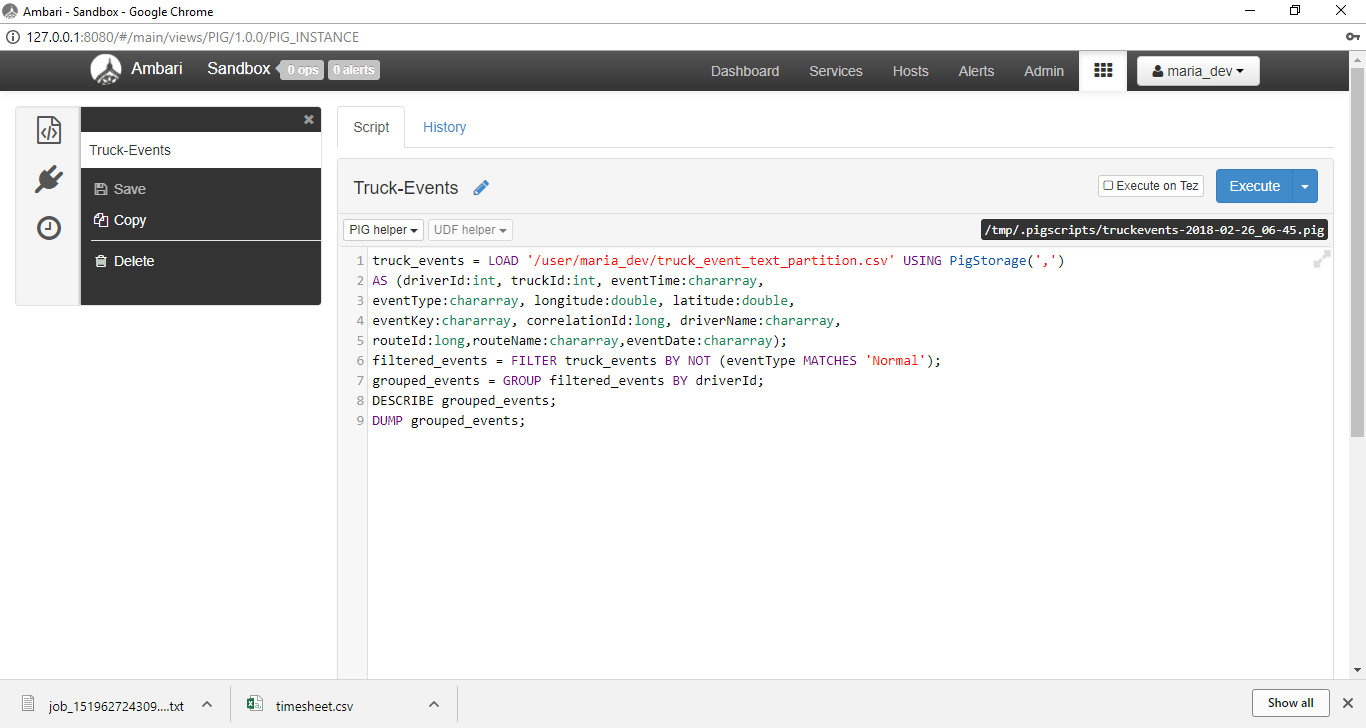


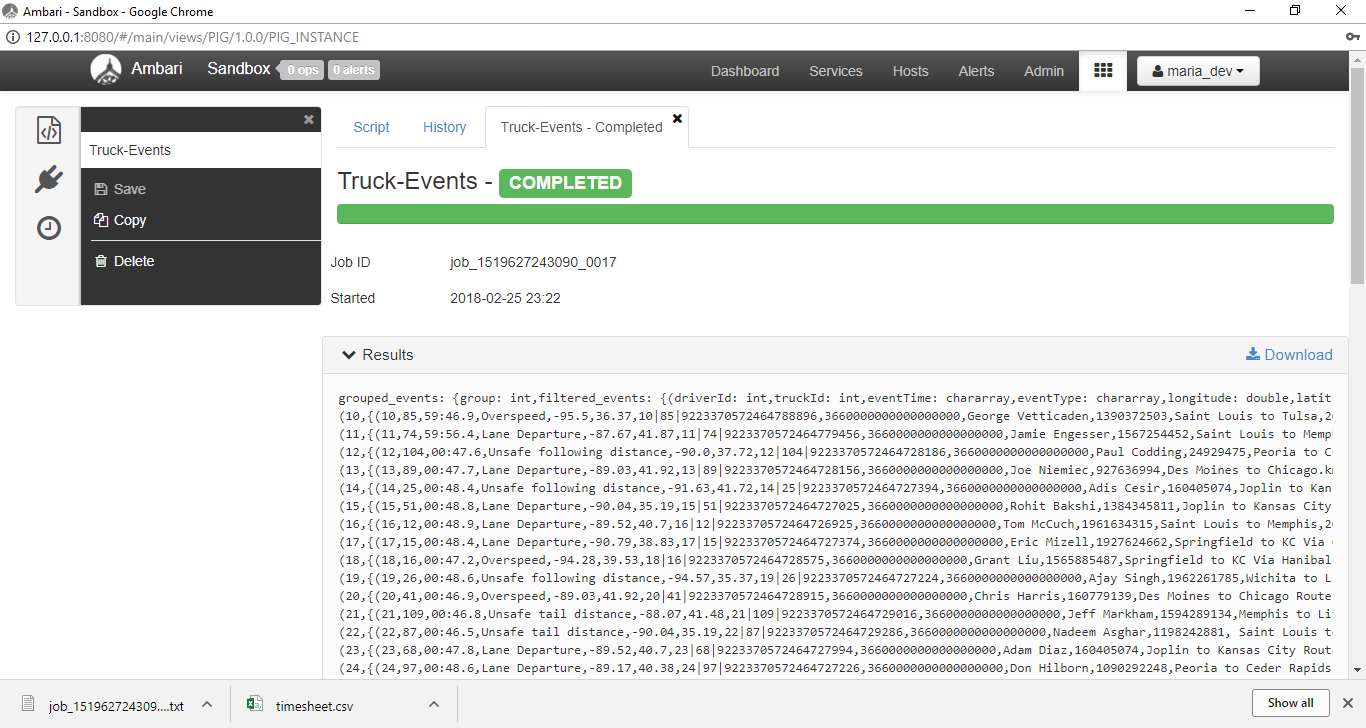
STEP 12: SORT THE DATA USING “ORDER BY”





STEP 13: FILTER AND GROUP THE DATA USING “GROUP BY”







The following files

File: part-v001-o000-r-00000 (step 10)

File: job\_1519627243090\_0017\_results (step 13)

**Part 2**: Complete the basic Cascading tutorial found at http://docs.cascading.org/tutorials/etllog/ Make sure you include a print screen after completing each key step.