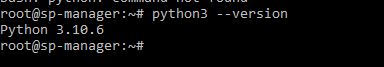
1. Get inside manager container
2. Check if python is installed. Pyspark requires at least python 3.7 or above.

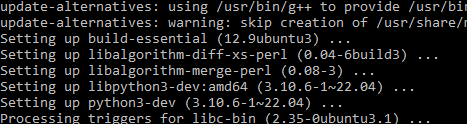
$ python –version



1. As we have python 3.10 , we are good to go with pyspark
2. Install pip

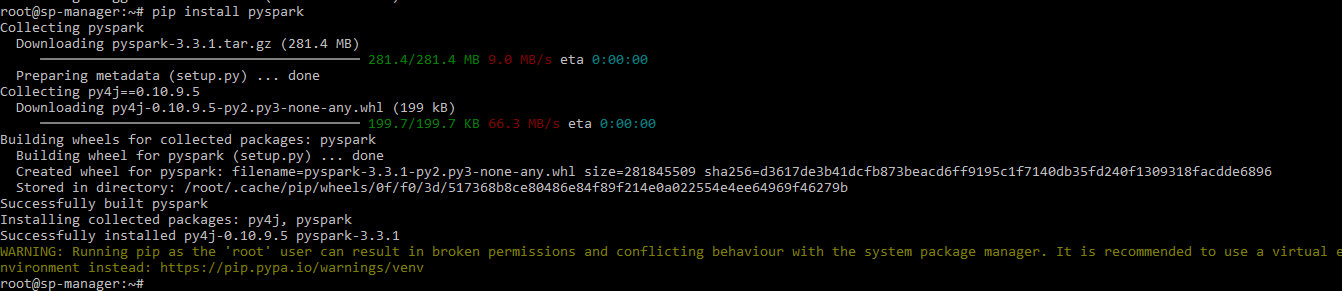
$ apt-get update

$ apt-get install pip



1. Install pyspark and other dependencies

$ pip install pyspark



*# Spark SQL*

pip install pyspark[sql]

*# pandas API on Spark*

pip install pyspark[pandas\_on\_spark] plotly *# to plot your data, you can install plotly together.*

References

<https://spark.apache.org/docs/latest/api/python/getting_started/install.html>