

Big data  
day 1 : 3

# Cloudera Installation

- 1) oracle virtual box ] To setup Hadoop  
Hive  
Pyspark
- <https://www.virtualbox.org/wiki/Downloads>
- 2) Image for Hadoop
- we need linux environment

<https://www.cloudera.com/downloads/hortonworks-sandbox.html>

Download Hortonworks HDP → HDP Sandbox

HDP is too big file

↳ go with cloudera : find link in Big data resources in google Keep.

→ Download Putty

[ Putty will help to access VM Environment into my local environment ]