

## 빅데이터 분산 플랫폼 아키텍처와 머신러닝 기반 데이터 분석 알고리즘

박경규

### 강사소개



- 이름:박경규
- 강의경력: 딥러닝, 자연어처리, 빅데이터, 파이썬, 에너지솔루션
- 개발경력

ADD플랫폼(Al Data Discovery Platform) 개발

BEMS (Building Energy Management System) 개발

PMS (Project Management System) 개발

- 관심분야: Self Development, Reading , Overseas Travel, Investment
- 깃허브 주소 : https://github.com/kgpark88
- 실습자료: https://github.com/kgpark88/bigdata

### **Contents**

1. 4차 산업혁명과 최신 기술 트렌드

5. 데이터 검색엔진 Elasticsearch

2. 빅데이터 개요 및 비정형 데이터 분석

6. 데이터분석 및 시각화

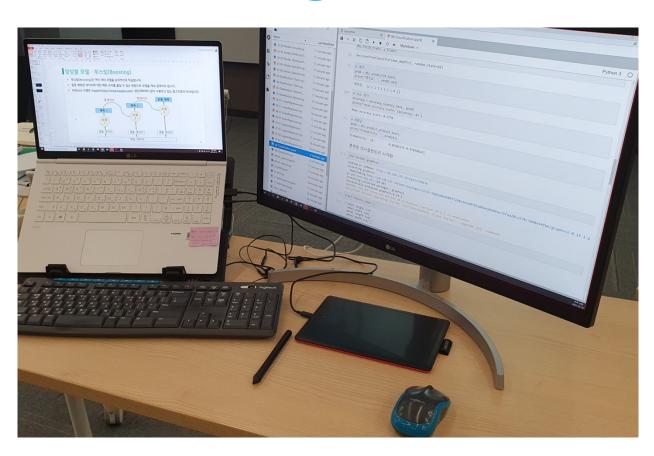
3. 빅데이터 플랫폼 Hadoop

7. 머신러닝 핵심 알고리즘

4. 빅데이터 프로세싱 엔진 Spark

8. 딥러닝 원리 및 주요모델

# 실습 환경 구축



### VirtualBox 설치

### https://www.virtualbox.org/



#### VirtualBox 7.0.10 platform packages

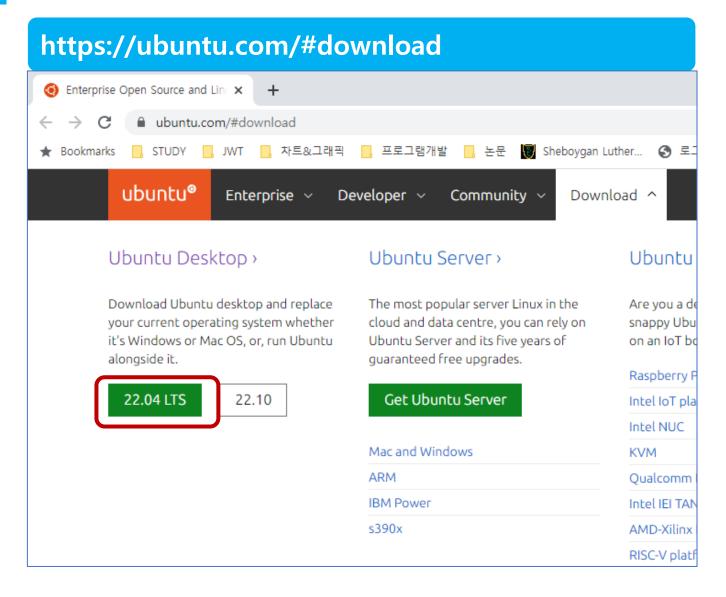
- ➡Windows hosts
- ⇒macOS / Intel hosts
- Linux distributions
- Solaris hosts
- Solaris 11 IPS hosts

#### <u>VirtualBox 7.0.10 Oracle VM</u> VirtualBox Extension Pack

→ All supported platforms

Support VirtualBox RDP, disk encryption, NVMe and PXE boot for Intel cards. See this chapter from the User Manual for an introduction to this Extension Pack. The Extension Pack binaries are released under the VirtualBox Personal Use and Evaluation License (PUEL). Please install the same version extension pack as your installed version of VirtualBox.

### Ubuntu 설치



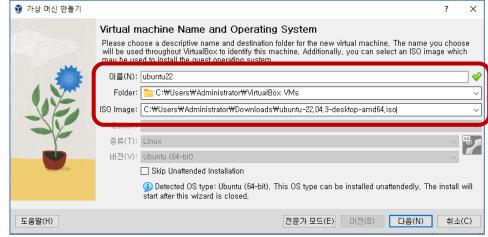
#### VirtualBox 관리자에서 [새로만들기] 클릭





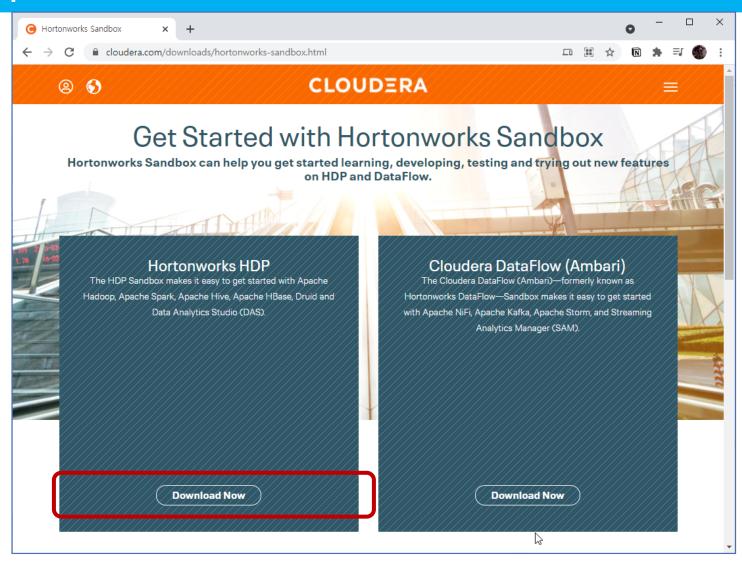
#### 서버 이름 입력 설치 위치 선택

ubuntu-22.04.3-desktop-amd64.iso 선택

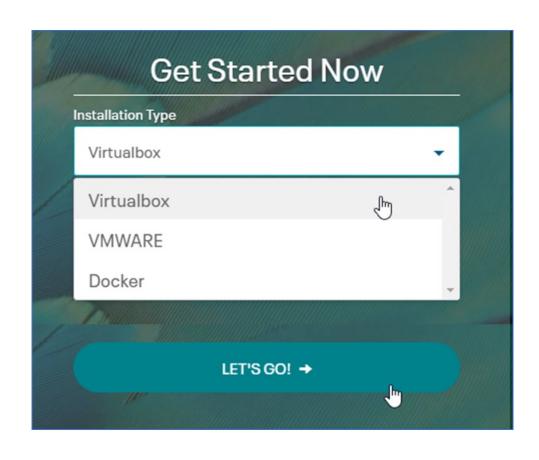


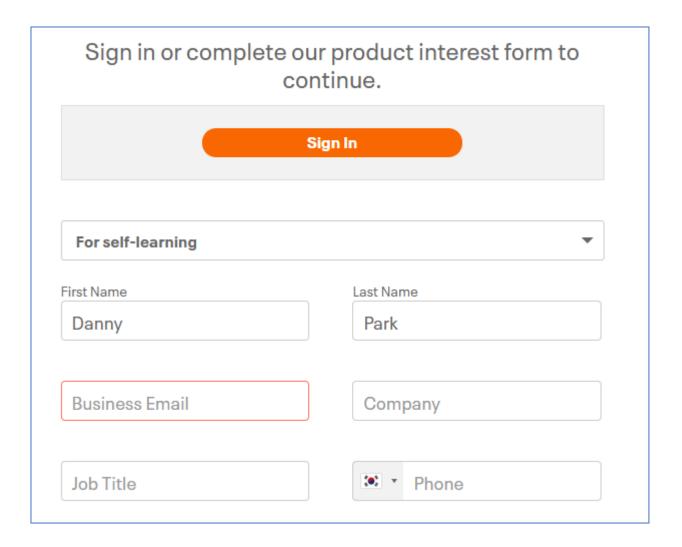
### HDP(Hortonworks Data Platform ) 다운로드

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



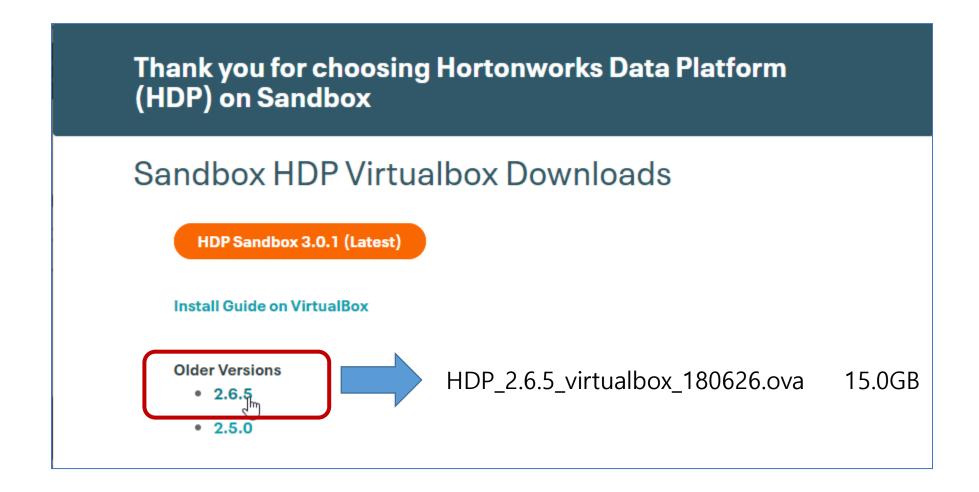
## HDP 다운로드





■ 참고 <a href="https://sungwoonblog.tistory.com/1">https://sungwoonblog.tistory.com/1</a>

### HDP 다운로드



Direct Download Link: <a href="https://archive.cloudera.com/hwx-sandbox/hdp/hdp-2.6.5/HDP\_2.6.5\_virtualbox\_180626.ova">https://archive.cloudera.com/hwx-sandbox/hdp/hdp-2.6.5/HDP\_2.6.5\_virtualbox\_180626.ova</a>