

HADOOP

이영환

01. 빅데이터란?

◆ 빅데이터 개념

- 빅데이터는 마케팅용어 외에 단순히 큰 데이터를 말하는 것이 아니고 이런 데이터를 처리하는 방식을 동시에 가리킨다.

◆ 빅데이터는

- RDBMS와의 차이점
: 기존의 RDBMS는 데이터를 빨리 찾는 데 최적화
- NoSQL 기술과의 차이점
: NoSQL은 RDBMS와 달리 빠른 기록에 최적화
- 데이터를 병렬처리할 수 있는 기반구조를 만들었다.
: 시스템(파일시스템)을 계속 추가확장하도록 만든 프레임워크 - 스케일아웃(Scale Out)
작업을 병렬처리할 수 있도록 분해(맵리듀스)

02.HADOOP의 역사

◆ 하둡의 역사

하둡은 더그커팅이 구글에 영감을 얻어 10년 이상 개발된 빅데이터 처리를 위한 자바기반의 분산파일시스템 + 병렬처리 프레임워크이다.

◆ 하둡의 특징

시스템을 지속 추가확장가능한 스케일아웃구조이다.

- 하둡은 자바를 기반으로 만들 프레임워크이다.

- 배포판

- Apach Hadoop(기본), CDH(클라우드라판), HDP(오픈웍스판)

03.하둡의 아키텍처

◆ 분산파일시스템(HDFS)

- 고가용성(HA:HighAvailability)

:동일한 데이터블록을 여러 번 복제해서 장애에도 문제없이 작동 가능하도록 만듦

◆ 맵리듀스(MapReduce)

하나의 작업을 두 가지 단계(맵/리듀스)로 나눠 병렬처리가 가능하도록 만든 분산 처리 프레임워크

◆ HDFS

네임노드 + 데이터노드

◆ MapReduce

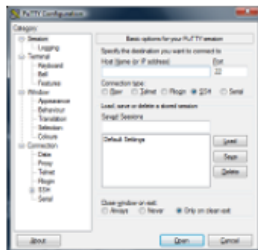
잡트래커 + 태스크트래커

◆ 네임노드가 죽으면 전체 서비스 장애가 생김

보조네임노드사용(Secondary NameNode)

04. 하둡을 설치하기 준비

- 하이퍼 바이저 - Virtual Box 나 도커 (download)
- 우분투(download)
- 자바(우분투에서 다운로드 설치)
- SSH(우분투에서 다운로드 설치)
- Putty (SSH Client : download)
- NANO(기본에디터)



Download PuTTY

PuTTY is an SSH and telnet client, developed originally by Simon Tatham for the Windows platform. PuTTY is open source software that is available with source code and is developed and supported by a group of volunteers.

You can download PuTTY [here](#).

Package files

You probably want one of these. They include all the PuTTY utilities.

(Not sure whether you want the 32-bit or the 64-bit version? Read the [FAQ entry](#).)

MSI ('Windows Installer')

32-bit: [putty-0.70-installer.msi](#) [\(or by FTP\)](#) [\(signature\)](#)

64-bit: [putty-64bit-0.70-installer.msi](#) [\(or by FTP\)](#) [\(signature\)](#)

VIRTUAL BOX DOWNLOAD

<https://www.virtualbox.org/wiki/Downloads>



[About](#)

[Screenshots](#)

[Downloads](#)

[Documentation](#)

[End-user docs](#)

[Technical docs](#)

[Contribute](#)

[Community](#)

VirtualBox

Download VirtualBox

Here you will find links to VirtualBox binaries and its source code.

VirtualBox binaries

By downloading, you agree to the terms and conditions of the respective license.

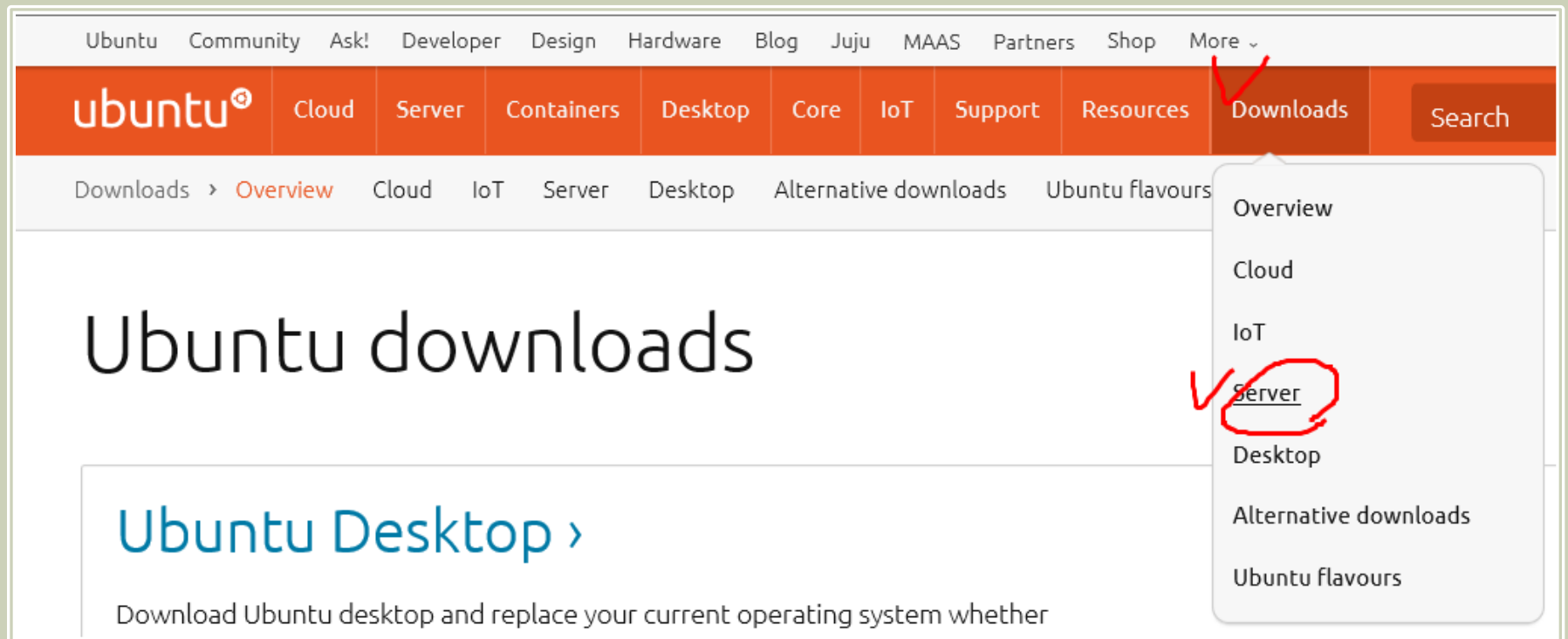
If you're looking for the latest VirtualBox 5.1 packages, see [VirtualBox 5.1 builds](#). Consider upgrading.

VirtualBox 5.2.12 platform packages

- [Windows hosts](#)
- [OS X hosts](#)
- [Linux distributions](#)
- [Solaris hosts](#)

우분투 다운로드1

<https://www.ubuntu.com/download>



The screenshot shows the Ubuntu website's download section. The top navigation bar includes links for Ubuntu, Community, Ask!, Developer, Design, Hardware, Blog, Juju, MAAS, Partners, Shop, and a More dropdown. Below this is a secondary navigation bar with links for Cloud, Server, Containers, Desktop, Core, IoT, Support, Resources, Downloads, and a Search button. The 'Downloads' link is highlighted with a red checkmark. A dropdown menu is open from 'Downloads', listing Overview, Cloud, IoT, Server, Desktop, Alternative downloads, and Ubuntu flavours. The 'Server' option is circled in red with a red checkmark. The main content area features the heading 'Ubuntu downloads' and a card for 'Ubuntu Desktop' with a right-pointing chevron. Below the card, it says 'Download Ubuntu desktop and replace your current operating system whether'.

Ubuntu Community Ask! Developer Design Hardware Blog Juju MAAS Partners Shop More ▾

ubuntu[®] Cloud Server Containers Desktop Core IoT Support Resources Downloads Search

Downloads ▸ Overview Cloud IoT Server Desktop Alternative downloads Ubuntu flavours

Ubuntu downloads

Overview
Cloud
IoT
Server
Desktop
Alternative downloads
Ubuntu flavours

Ubuntu Desktop ▸

Download Ubuntu desktop and replace your current operating system whether

우분투 다운로드2

Download Ubuntu Server

Ubuntu Server 18.04 LTS

The long-term support version of Ubuntu Server, including the Queens release of OpenStack and support guaranteed until April 2023 — 64-bit only.

This release uses our new installer, Subiquity. If you require advanced networking and storage features, such as RAID and LVM, please use the traditional installer found on the [alternative downloads](#) page.

[Ubuntu Server 18.04 LTS release notes](#)

Download

For other versions of Ubuntu including torrents, the network installer, a list of local mirrors, and past releases [see our alternative downloads](#).