# EFK를 활용한 로그 수집과 분석 (3일 과정)

(Elasticsearch + Fluentd + Kibana)

2022 - 09

# 1 일차: EFK, Elasticsearch

- MSA와 Log 수집·분석의 필요성
- 검색엔진과 Elasticsearch의 차이점
- Logstash vs Fluentd 비교 분석
- EFK 환경 구축
  - 1) EFK 설치
  - 2) 간단한 Log 분석 (CRUD Document 검색)
  - 3) EFK Log 분석 및 시각화

[별첨] Window 실습(Local) 환경구축: WSL2, Python, Docker & Kubernetes 설치

## 교육 내용 및 일정 소개

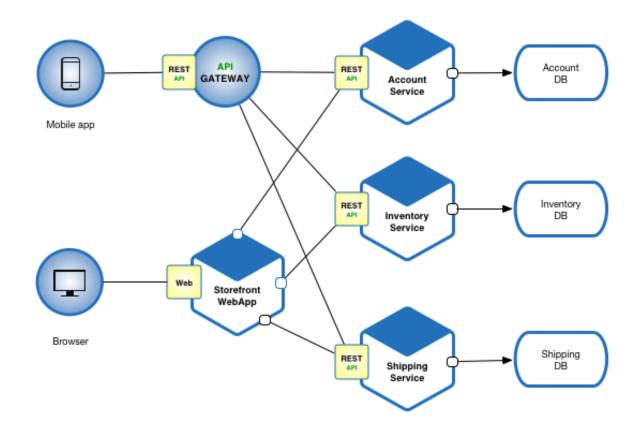
	주제	시간	학습 내용	
교육 내용	Elasticsearch 개요	1H	<ul> <li>MSA(Microservice Architecture)와 Log 분석의 필요성</li> <li>검색 엔진 DBMS, Elasticsearch 차이점 비교</li> <li>Elasticsearch architecture (Cluster, Node, Shard) 이해</li> </ul>	
	Logstash vs Fluentd 비교분석	1H	■ Fluentd와 Logstash 비교 분석 ■ Fluentd와 Logstash의 Hybrid 구성 사례	
	EFK 구축과 실습	1H	■ Elasticsearch와 Kibana 설치 (WSL2 설치, Docker 설치, Cloud 설치)	
		1H	■ elasticsearch.yml 파일 설정	
		1H	■ Elasticsearch와 Kibana CRUD(Create, Read, Update, Delete) Document 검색	
		1H	■ Fluentd 설치 및 /etc/td-agent/td-agent.conf 파일 설정	
		1H	■ [실습] EFK를 이용한 간단한 Log 수집 및 시각화	

#### 마이크로서비스는 무엇인가요?

- 함께 작동하는, 독립적으로 배포 가능한 서비스는 비즈니스 도메인 중심의 모델링
- 네트워크를 통해 서로 통신
- 데이터베이스는 서비스 경계 내부에 숨겨져 있음
- 정보 은폐의 원칙

#### 마이크로서비스(MSA, Microservice Architecture)와 Log 분석의 필요성

MSA의 정의: An architectural style that structures an application as a collection of services



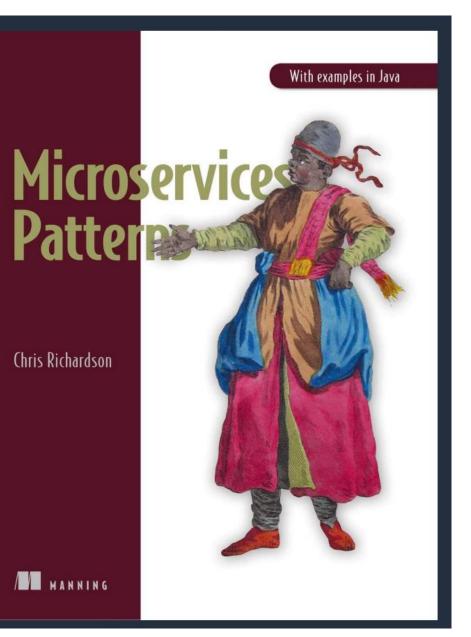
O'REILLY"

# Monolith to Microservices

Evolutionary Patterns to Transform Your Monolith



Sam Newman



#### 독립적인 배포와 데이터 소유권 문제

- 변경 사항을 독립적으로 변경하고 배포합니다.
- 느슨한 결합:
  - 서비스 간의 잘 정의되고 안정적인 계약.
  - 일부 구현 선택은 이를 어렵게 만듭니다. 예: 데이터베이스 공유.
- 자신의 데이터를 소유하십시오 <----- 오늘의 주제



#### **Elasticsearch**

- 오픈소스 (open source)
- 실시간 분석 (real-time)
- 전문(full text) 검색 엔진
- RESTFul API

#### Elasticsearch Stack의 구성

Kibana
Explore, Visualize, Engage

Elasticsearch
Store, Search, Analyze

Integrations
Connect, Collect, Alert

#### EFK(Elasticsearch + Fluentd + Kibana) 란?

- 컨테이너를 이용해 편리하게 배포를 하게 되면서 컨테이너를 관리해 주는 오케스트레이션 툴인 "Kubernetes"가 등장
- 쿠버네티스 Node에 POD가 배포되는데, 이 POD에 여러 컨테이너가 존재한다.
- 내가 관리하지 않아도 자동으로 관리된다. 얼마나 편리한가.
- 그런데.. 쿠버네티스는 POD가 실행이 안 되면 그 POD를 죽이고 새로 생성한다.
- 그러면.. 죽은 POD에 있던 컨테이너가 남긴 로그는 어떻게 될까.
- 로그가 다 사라지게 됩니다.
- POD내의 컨테이너의 로그를 남기는 방법 -> 그게 EFK 또는 ELK(Elasticsearch+Logstash+Kibana)

#### Elasticsearch와 검색엔진의 차이?

- SON 문서 기반 Elasticsearch는 내부적으로는 역파일 색인 구조로 데이터를 저장
- Elasticsearch는 Rest API를 기본으로 지원하며 모든 데이터 조회, 입력, 삭제를 http 프로토콜을 통해 Rest API로 처리
- Elasticsearch의 데이터들은 인덱스(Index) 라는 논리적인 집합 단위로 구성되며 서로 다른 저장소에 분산되어 저장하고 서로 다른 인덱스들을 별도의 커넥션 없이 하나의 질의로 묶어서 검색하고, 검색 결과들을 하나의 출력으로 도출할 수 있는데, Elasticsearch의 이러한 특징을 멀티테넌시 라함

Relational Database	ElasticSearch
Database	Index
Table	Туре
Row	Document
Column	Field
Index	Analyze
Primary key	_id
Schema	Mapping
Physical partition	Shard
Logical partition	Route
Relational	Parent/Child, Nested
SQL	Query DSL

## Logstash Vs Fluentd — 차이 비교

	Logstash	Fluent Bit
공식 플러그인 개수	162개	85개
플러그인	공식 깃허브에서 관리	Golang으로 원하는 플러그인 개발 가능
파일을 읽어올 때	기본적으로 sincedb를 사용하여 파일을 얼마나 읽었는지 추적	db를 사용하여 파일을 모니터링 할 수 있음
로그 파싱	정규식	기반
민감한 필드 익명화, 제외	0	익명화 X, 제외 가능
배치 처리	0	데이터를 Chunk 단위로 처리
데이터 처리	병렬 처리 유동적인 데이터 처리로 부하 상황에서도 안정적	코루틴을 사용한 동시성 프로그래밍, 스트림 프로세싱으로 데이터가 이동중일 때에도 쿼리 가능
데이터 라우팅	if 조건문 기반	Tag 기반
이벤트 전송 보장	최소 1회 보장	filesystem 버퍼링을 사용한 full data safety 보장
커뮤니티	discuss.elastic.co	fluent/fluent-bit 깃허브
메모리 사용량	1G	35MB
운영사/재단	elastic.co	CNCF
언어	jruby 호스트에 자바 런타임 필요	C + Golang 자바 런타임 필요 X

#### **Logstash Vs Fluentd — Basic Comparison**

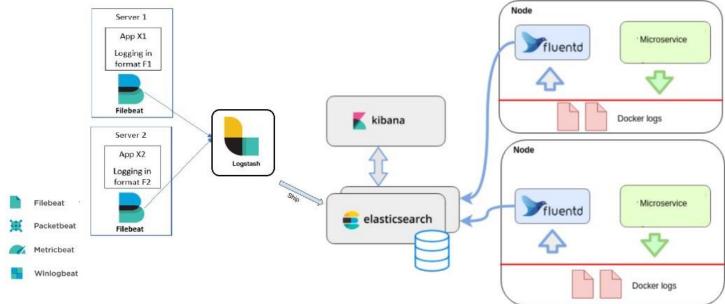
	Logstash	Fluentd	Comment
Endorsing Company	LogStash is part of the popular ELK stack.	Fluentd is built by Treasure Data and is part of the CNCF.	Fluentd also has excellent support for Elastic. For CNCF hosted project (e.g. Kubernetes, OpenTracing or Prometheus), Fluentd could be a better choice.
Code Language	JRuby – so it requires java runtime on the host machine.	CRuby – No java runtime required	Fluentd has advantage here as no java runtime required for Fluentd
Event Routing	Events are routed based on if-else conditions.  Eg- output { if [loglevel] ==      "ERROR" and      [deployment] ==      "production" { { }	Events are routed based on tags.  Eg- <match logtype.error=""> type </match>	Fluentd has better routing approach as it is easier to tag events then use if-else for each event type. If there are too many different types of logs then logstash filters can be difficult to manage.

#### **ELK-EFK Hybrid Architecture**

**For monolithic applications** on traditional VMs, Logstash looks like a clear choice and way to proceed as it supports multiple agents for collection of logs, metrics, health etc.

For microservices hosted on Docker/Kubernetes, Fluentd looks like a great choice considering built in logging driver and seamless integration. It supports all commonly used parsers like json, nginx, grok etc.

#### **ELK-EFK Hybrid Architecture** Kubernetes cluster Server 1 App X1 fluentd



[출처: https://www.techmanyu.com/logstash-vs-fluentd-which-one-is-better-adaaba45021b]

#### **Logstash Vs Fluentd — Basic Comparison**

Plugins	About 200 plugins	About 500 plugins	Logstash has all plugins in the official git repo while Fluentd still does not have a centralized repo for plugins.
Performance	Consumes more memory	Consumes less memory	Fluentd has a slightly better reputation wrt performance. But both these have light weight log shippers so performance wise both are very efficient.
Transport	Logstash is limited to an on-memory queue that holds 20 events (fixed size) and relies on an external queue like redis or kafka for persistence across restarts.	Fluentd has a highly configurable buffering system that can be inmemory or on-disk with a seemingly endless array of	As discussed in this talk at OpenStack Summit 2015, both perform well in most use cases and consistently grok through 10,000+ events
Enterprise Support	Yes – Link	Yes – Link	Logstash has better enterprise support as it is part of official Elastic

#### **CLOUD NATIVE Computing Foundation**

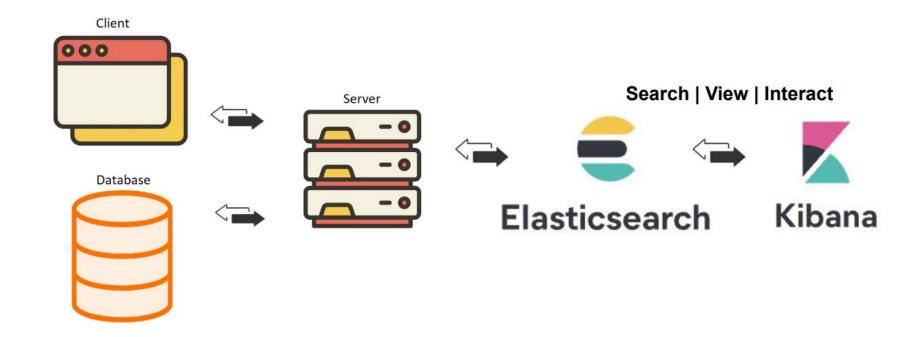
- Fluent Bit은 CNCF의 Graduated Project인 Fluentd의 에코시스템 중 하나입니다. CNCF의 다른 Graduated Project로는 Kubernetes, containerd, etcd, helm, Prometheus 등이 있습니다.
- 따라서 Fluent Bit은 Kubernetes, Prometheus와 같은 다른 CNCF 프로젝트와 함께 사용하기 용이하며 Elastic.co의 시스템과도 통합을 지원합니다.
- Fluentd는 Kubernetes와 같은 CNCF 스택이기 때문에 마이크로 서비스의 로깅에 더 적합하고, Ruby언어로 작성되어 있어 Logstash에 비해 메모리를 적게 사용하고 메시 지 손실이 없다고 알려져 있습니다.

https://www.cncf.io/

#### Elasticsearch Stack의 용도

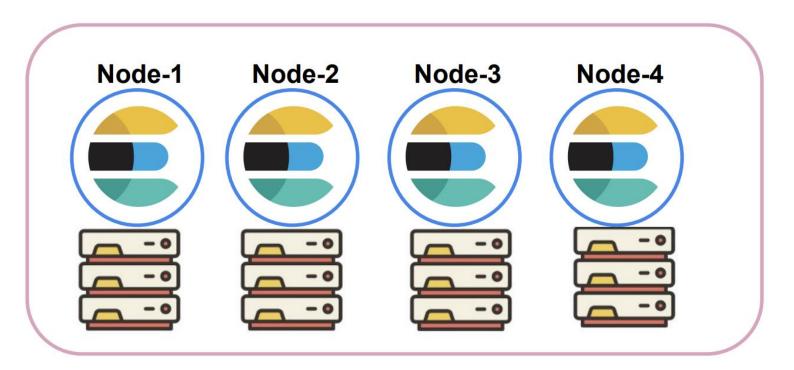
- Logging
- Metrics
- Security Analytics
- Business Analytics

#### Elasticsearch + Kibana의 구성

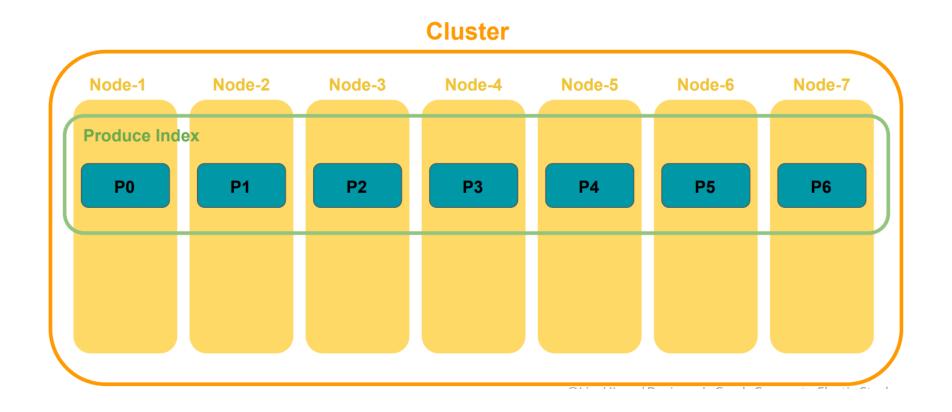


#### Elasticsearch의 Architecture: Cluster 와 Node

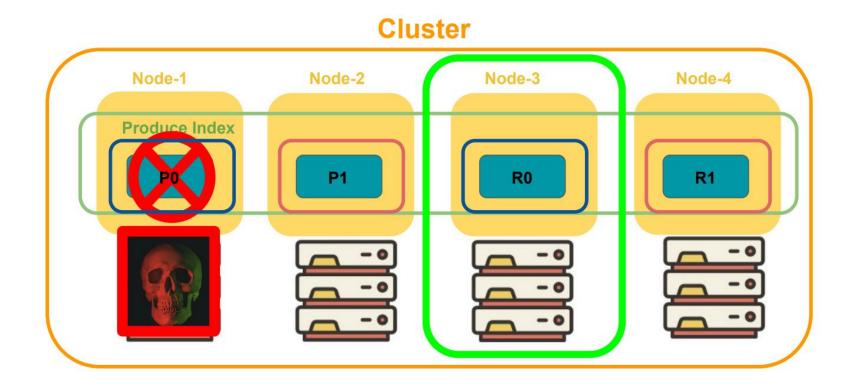
#### Cluster



#### Elasticsearch Architecture: Shard



### Elasticsearch의 Architecture: Replica shards



#### Elasticsearch + Kibana 설치

https://github.com/JSJeong-me/EFK/blob/main/00\_EFK\_Installation.md

● /etc/security/limits.conf 파일에 (max file descriptord 4096 에서 65536 으로 증가) {리눅스계정} - nofile 65535

● /etc/sysctl.conf 파일에 아래 내용 추가 (virtual memory 영역을 65530 에서 262144 로 증가) vm.max\_map\_count=262144

참조: https://www.elastic.co/guide/en/elasticsearch/reference/current/setting-system-settings.html#limits.conf

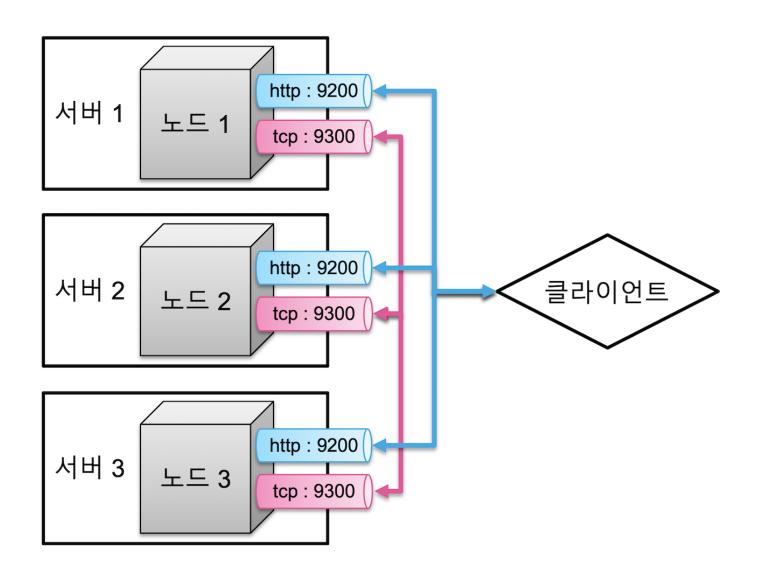
#### Elasticsearch 노드설정 및 클러스터 구성

● config/elasticsearch.yml 파일 설정

```
cluster.name: "es-cluster"
node.name: "node-1"
network.host: ["_local_", "_site_"]
discovery.seed_hosts: ["elastic-1"]
cluster.initial_master_nodes: ["node-1"]
```

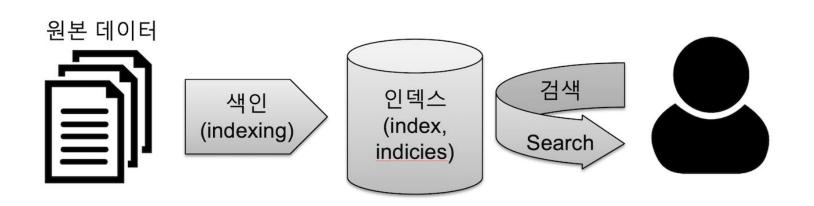
- network.host 값을 실제 ip 주소 또는 hostname으로 변경하면 외부에서도 elasticsearch 접속가능
- 단, 이때부터는 elasticsearch 실행 시 실제 운영 환경 클러스터 구동에 적합한지 bootstrap 체크

#### Elasticsearch 노드설정 및 클러스터 구성



#### 데이터 색인

- [동사] 색인 (indexing) : 데이터가 검색될 수 있는 구조로 변경하기 위해 원본 문서를 검색어 토큰들으로 변환하여 저장하는 일련의 과정입니다. 이 책에서는 색인 또는 색인 과정이라고 표기합니다.
- [명사] 인덱스 (index, indices): 색인 과정을 거친 결과물, 또는 색인된 데이터가 저장되는 저장소입니다. 또한 Elasticsearch에서 도큐먼트들의 논리적인 집합을 표현하는 단위이기도 합니다. 이 책에서는 인덱스라고 표기합니다.
- 검색 (search) : 인덱스에 들어있는 검색어 토큰들을 포함하고 있는 문서를 찾아가는 과정입니다.
- 질의 (query) : 사용자가 원하는 문서를 찾거나 집계 결과를 출력하기 위해 검색 시 입력하는 검색어 또는 검색 조건입니다. 이 책에서는 질의 또는 쿼리라고 표현합니다.



#### **REST APIS**

Elasticsearch는 http 프로토콜로 접근이 가능한 REST API를 지원합니다. 자원별로 고유 URL로 접근이 가능하며 http 메서드 **PUT, POST, GET, DELETE** 를 이용해서 자원을 처리합니다. 이런 특성을 가진 시스템을 보통 **RESTFul** 한 시스템이라고 말합니다.

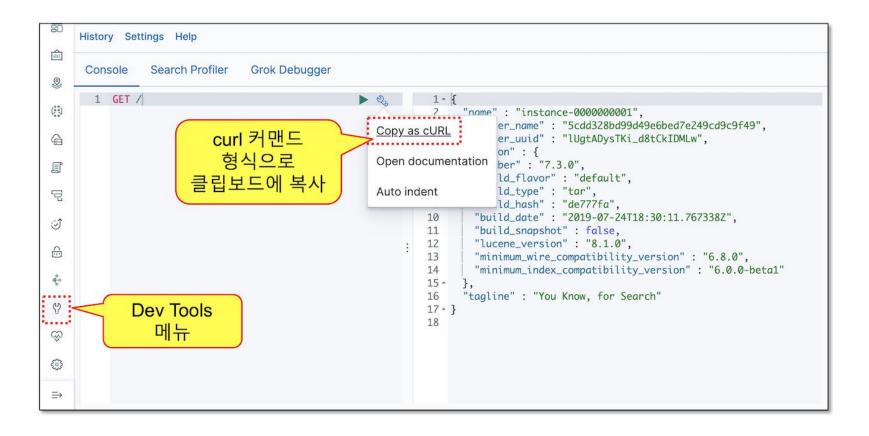
REST API를 지원하는 시스템은 kim 이라는 사용자에 대해 항상 단일 URL로 접근을 하고 PUT, GET, DELETE 같은 http 메서드로 데이터를 처리합니다.

#### RESTFul 한 시스템에서의 데이터 처리

- 입력: PUT http://user.com/kim -d {"name":"kim", "age":38, "gender":"m"}
- 조회 : GET <u>http://user.com/kim</u>
- 삭제 : DELETE http://user.com/kim

#### Kibana Dev Tools을 이용한 CRUD

Kibana에는 elasticsearch 에서 REST API를 간편하게 실행할 수 있는 Dev Tools 라는 도구를 제공 Kibana를 실행한 뒤 웹 브라우저를 열고 http://localhost:5601 로 접속하면 Kibana를 바로 사용



#### CRUD - 입력, 조회, 수정, 삭제

curl 도구를 이용해서 my\_index 인덱스에 도큐먼트 id가 1인 데이터를 입력하는 예제

```
$ curl -XPUT "http://localhost:9200/my_index/_doc/1" -H 'Content-Type: application/json' -d' {
    "name": "Korea",
    "message": "안녕하세요 Elasticsearch"
}'
{"_index":"my_index","_type":"_doc","_id":"1","_version":1,"result":"created","_shards":{"total":2,"suc cessful":1,"failed":0},"_seq_no":0,"_primary_term":1}
```

#### 입력 (PUT)

데이터 입력을 할 때는 PUT 메서드를 이용합니다. 다음은 Kibana 에서 my\_index 인덱스에 도큐 먼트 id가 1인 데이터를 입력하는 예제입니다.

```
PUT my_index/_doc/1
{
   "name":"Korea",
   "message":"안녕하세요 Elasticsearch"
}
```

#### 조회 (GET)

GET 메서드로 가져올 도큐먼트의 URL을 입력하면 도큐먼트의 내용을 가져옵니다. 다양한 정보가 함께 표시되며 문서의 내용은 \_source 항목에 나타납니다.

GET my\_index/\_doc/1

#### 삭제 (DELETE)

DELETE 메서드를 이용해서 도큐먼트 또는 인덱스 단위의 삭제가 가능합니다. 두 경우에 차이가 있는데 먼저 DELETE my\_index/\_doc/1 명령으로 하나의 도큐먼트를 삭제하면 다음과 같이 도큐먼트가 삭제되었다는 "result": "deleted" 결과가 리턴됩니다

DELETE my\_index/\_doc/1

#### 수정 (POST)

POST 메서드는 PUT 메서드와 유사하게 데이터 입력에 사용이 가능합니다. 도큐먼트를 입력할때 POST 메서드로 <인덱스>/\_doc 까지만 입력하게 되면 자동으로 임의의 도큐먼트id 가 생성됩니다. 도큐먼트id의 자동 생성은 PUT 메서드로는 동작하지 않습니다.

```
POST my_index/_doc
{
  "name":"Korea",
  "message":"안녕하세요 Elasticsearch"
}
```

#### \_update

입력된 도큐먼트를 수정하기 위해서는 기존 도큐먼트의 URL에 변경될 내용의 도큐먼트 내용을 다시 PUT 하는 것으로 대치가 가능합니다. 하지만 필드가 여럿 있는 도큐먼트에서 필드 하나만 바꾸기 위해 전체 도큐먼트 내용을 매번 다시 입력하는 것은 번거로운 작업일 것입니다. 이 때는 POST <인덱스>/\_update/<도큐먼트 id> 명령을 이용해 원하는 필드의 내용만 업데이트가 가능합니다.

```
POST my_index/_update/1
{
  "doc": {
   "message":"안녕하세요 Kibana"
  }
}
```

#### Elasticsearch 8.x 에서 https - Disable 방법

```
# Enable security features
xpack.security.enabled: false
xpack.security.enrollment.enabled: false
# Enable encryption for HTTP API client connections, such as Kibana, Logstash,
and Agents
xpack.security.http.ssl:
 enabled: false
 keystore.path: certs/http.p12
# Enable encryption and mutual authentication between cluster nodes
xpack.security.transport.ssl:
 enabled: false
 verification_mode: certificate
 keystore.path: certs/transport.p12
 truststore.path: certs/transport.p12
```

#### EFK 실습 예제 : EFK를 이용한 간단한 Log 수집 및 시각화

- td-agent 설치 (https://docs.fluentd.org/quickstart/td-agent-v2-vs-v3-vs-v4)
   sudo apt update
- curl -L https://toolbelt.treasuredata.com/sh/install-ubuntu-focal-td-agnet4.sh | sh
- 데몬 실행

\$ sudo systemctl start td-agent.service (서비스 시작)

\$ sudo systemctl status td-agent.service (상태 확인)

[참조: https://docs.fluentd.org/installation/install-by-deb]

#### EFK 실습 예제 : EFK를 이용한 간단한 Log 수집 및 시각화

- 간단한 Log 생성 Python 코드: main.py
  - \$ tail –f log.json
  - \$ sudo service td-agent status
  - \$ sudo Is /etc/td-agent/.
  - \$ sudo vi /etc/td-agent/td-agent.conf
  - \$ sudo service td-agent stop
  - \$ sudo service td-agent start
  - \$ python main.py

### [별첨] WSL2, Python, Docker & Kubernetes 설치 및 환경설정

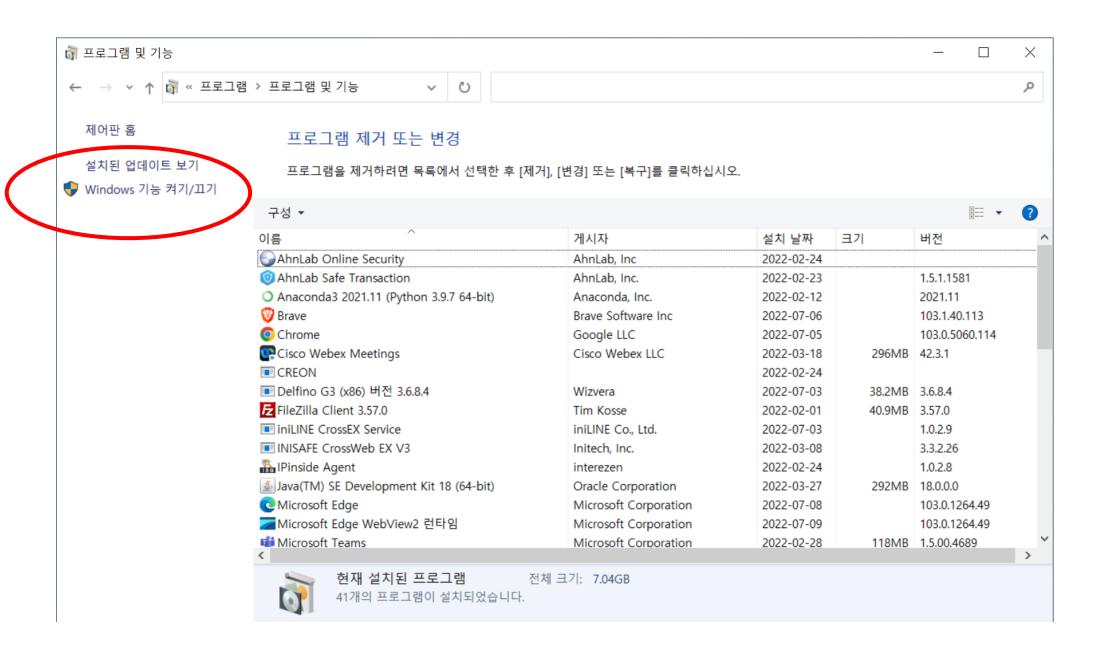
# PC(Local) 환경구축: WSL2 & Docker Install

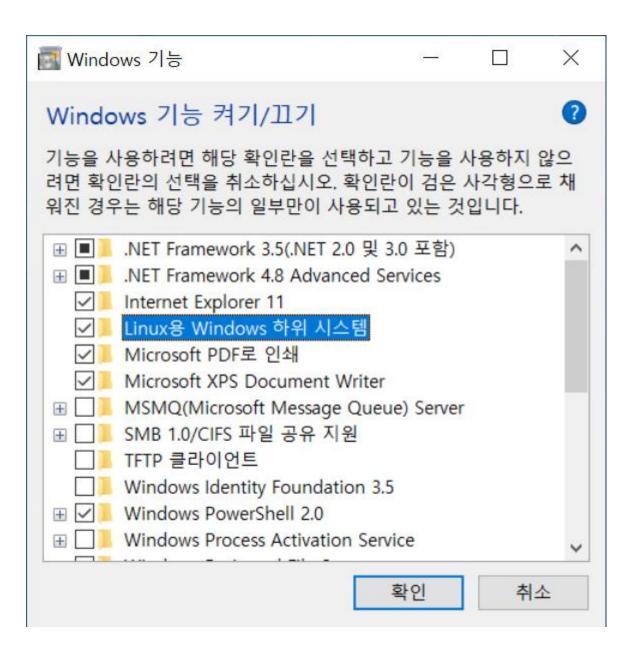
#### 윈도우 환경에서 원하는 Linux를 실행

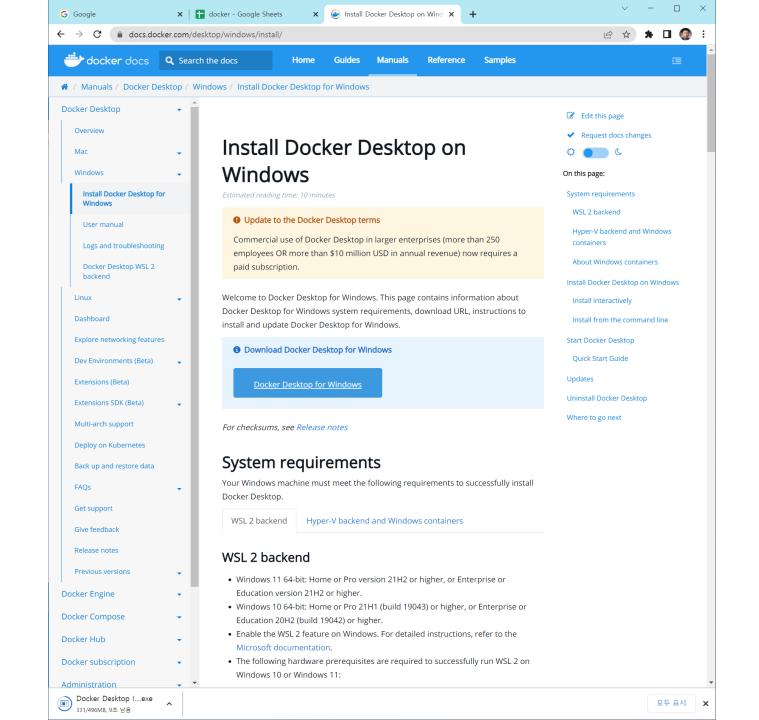
- 1. 관리자 권한으로 명령 프롬프트(CMD) 실행
- 2. <a href="https://docs.microsoft.com/ko-kr/windows/wsl/tutorials/gui-apps">https://docs.microsoft.com/ko-kr/windows/wsl/tutorials/gui-apps</a>
- 3. C> Enable-WindowsOptionalFeature -Online -FeatureName Microsoft-windows-Subsystem-Linux
- 4. 또는 <a href="https://ivyit.tistory.com/264">https://ivyit.tistory.com/264</a>
- 5. Microsoft Store 열기
- 6. Ubuntu 또는 Debian 선택

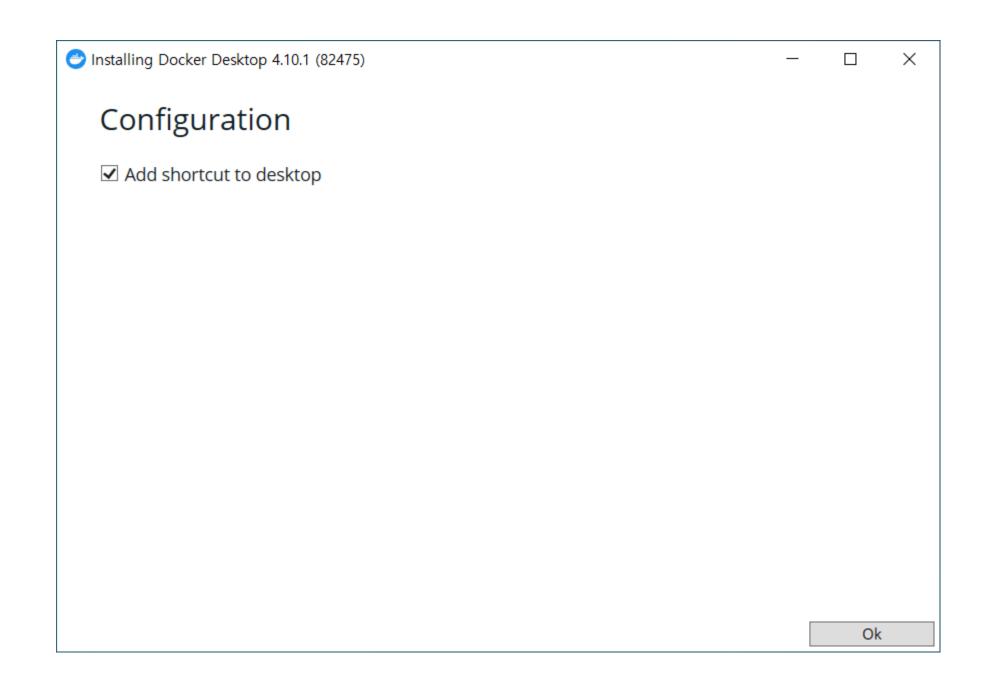
#### https://docs.microsoft.com/ko-kr/windows/wsl/install

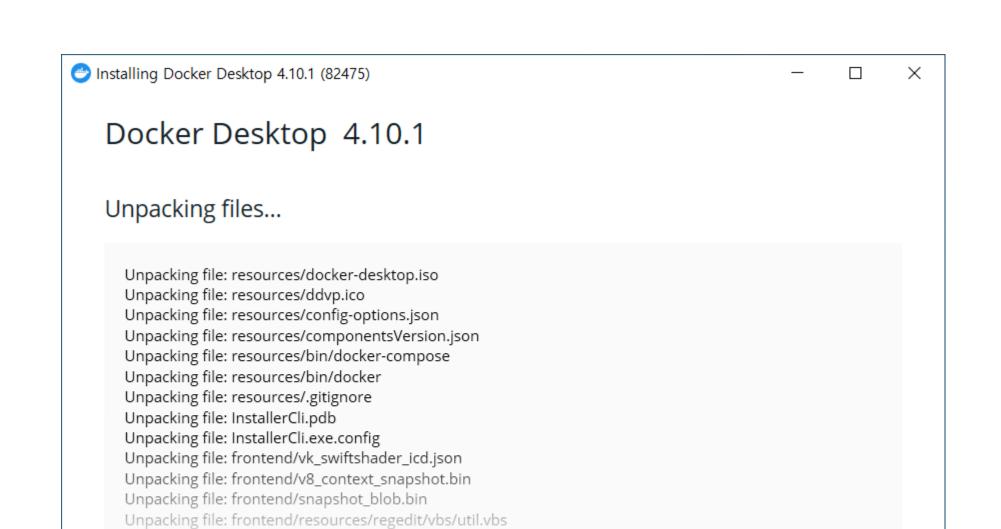
https://docs.microsoft.com/ko-kr/windows/wsl/install-manual#step-4---download-the-linux-kernel-update-package

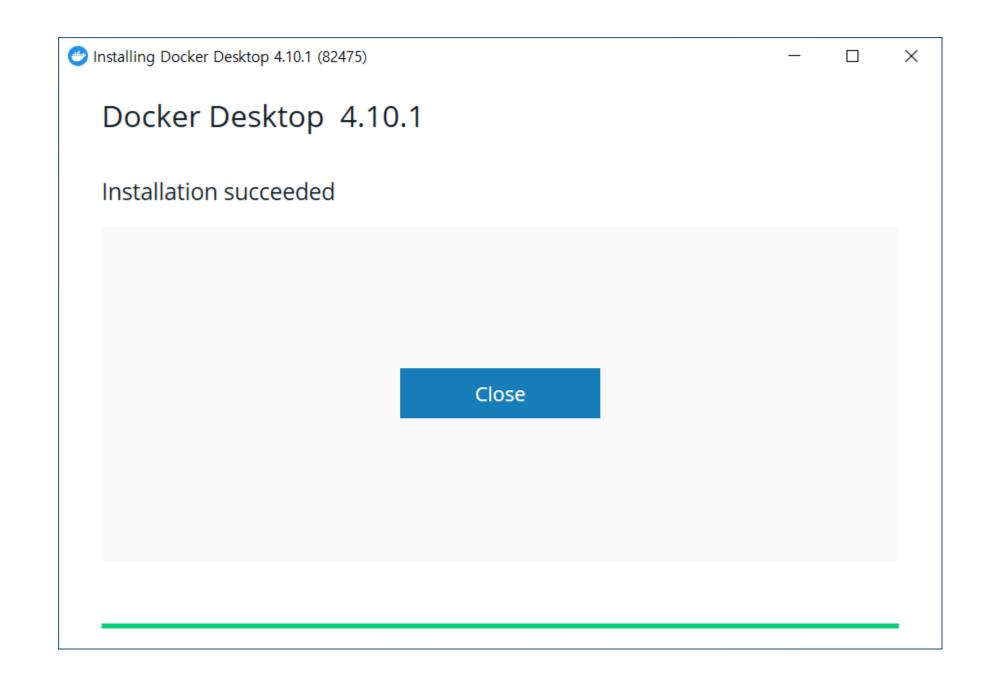














Containers

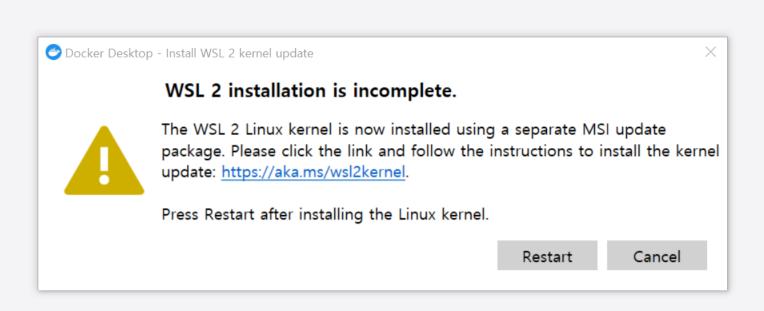
Images

Volumes

Dev Environments BETA

Extensions BETA

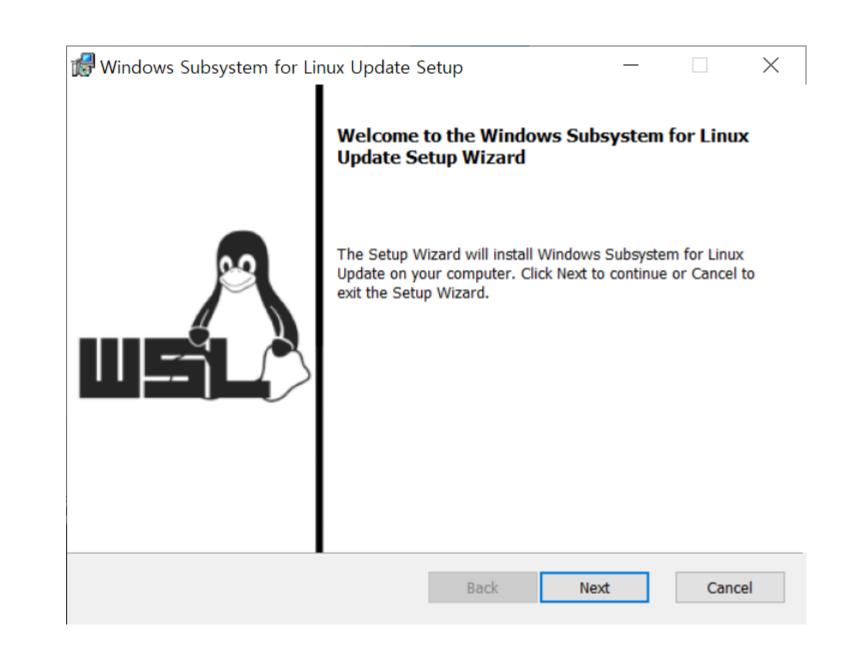
Add Extensions



drjsjeong 😩

# WSL 2 installation incomplete error 발생시 설치 Package

https://docs.microsoft.com/ko-kr/windows/wsl/install-manual#step-4--download-the-linux-kernel-update-package



Microsoft Windows [Version 10.0.19044.1766] (c) Microsoft Corporation. All rights reserved.

C:₩windows\system32>wsl --set-default-version 2 WSL 2와의 주요 차이점에 대한 자세한 내용은 https://aka.ms/wsl2를 참조하세요 작업을 완료했습니다.

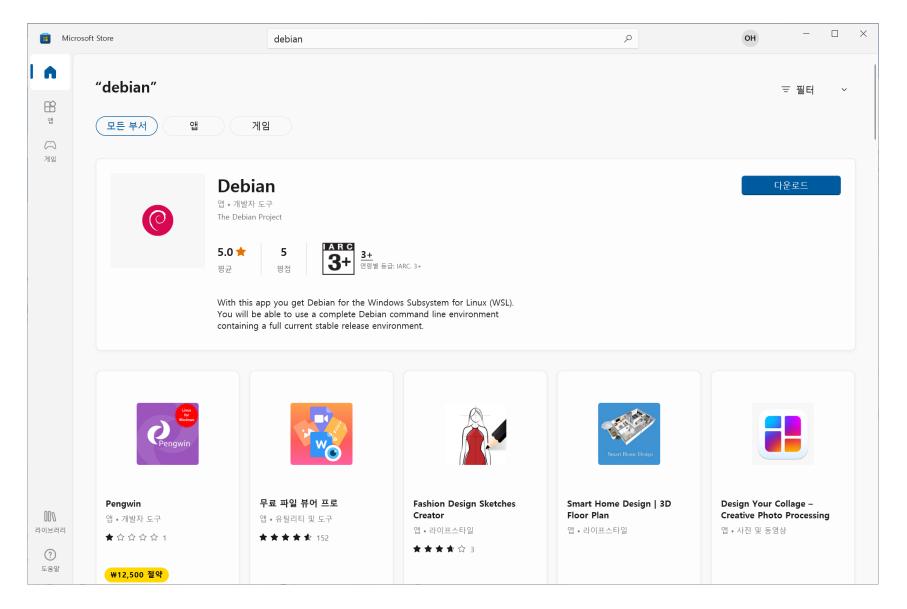
C:₩Windows₩system32>\_

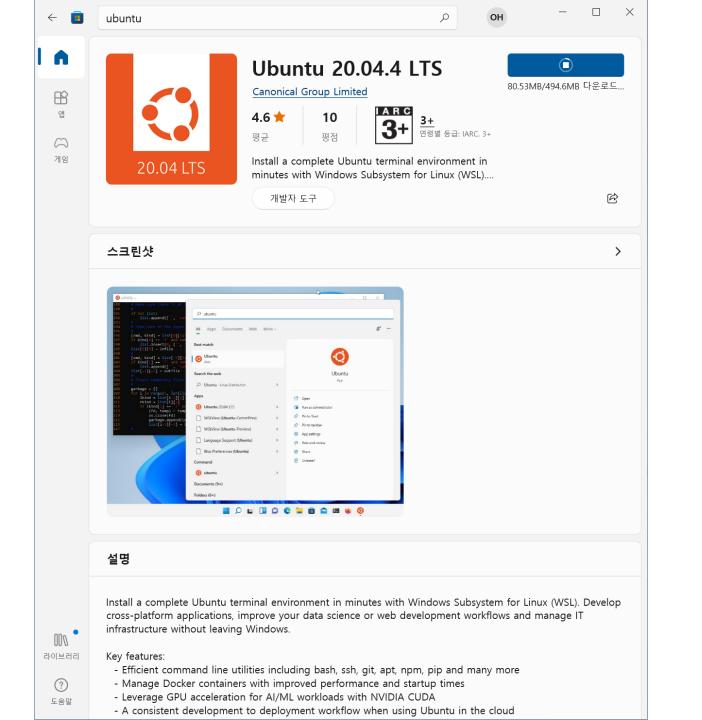
V

### 설치 가능한 리눅스 배포본은 "-I -o" 옵션으로 확인



# Microsoft Store에서: Linux 선택





Installing, this may take a few minutes... Please create a default UNIX user account. The username does not need to match your Windows username. For more information visit: https://aka.ms/wslusers Enter new UNIX username: admin

### Linux 계정 설정

**UNIX** username: me

passwd: diana

\$ sudo passwd root

[sudo] password for me: diana

# https://positivemh.tistory.com/583

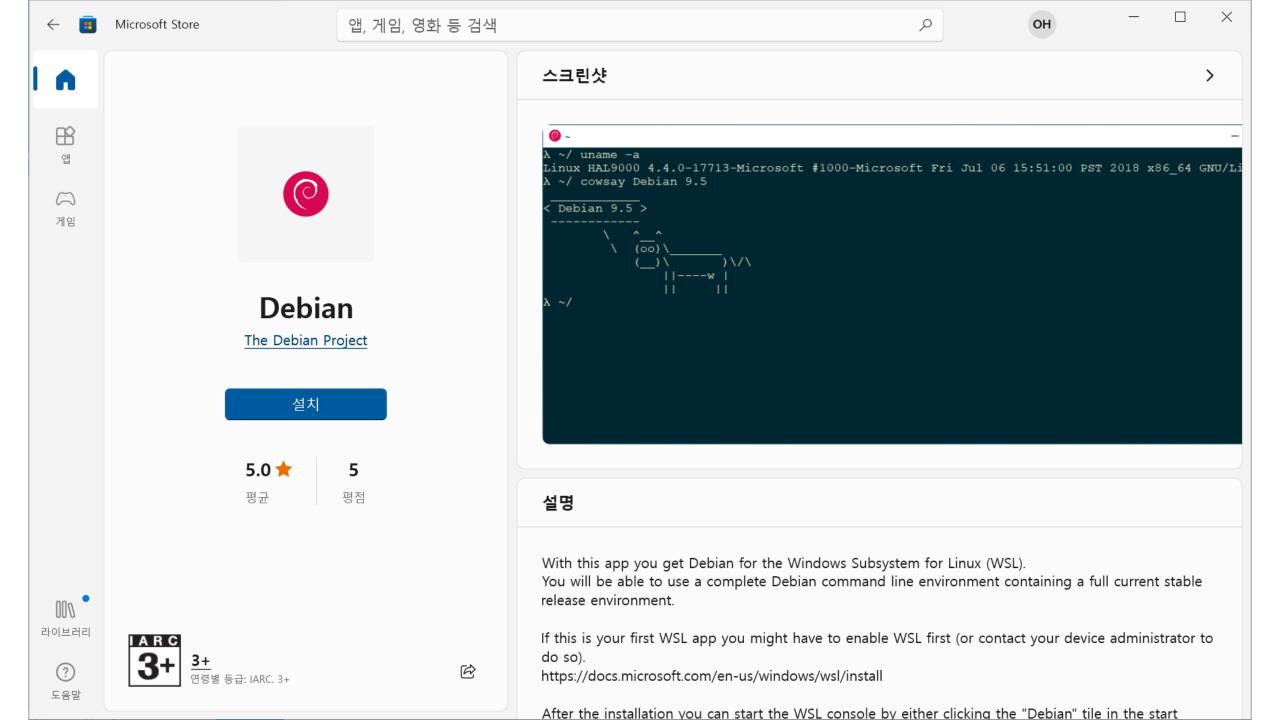
```
me@DESKTOP-HAAI0JO: ~
                                                                                                                Retype new password:
passwd: password updated successfully
Installation successful!
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo root" for details.
Welcome to Ubuntu 20.04.4 LTS (GNU/Linux 5.10.16.3-microsoft-standard-WSL2 x86 64)
 * Documentation: https://help.ubuntu.com
                  https://landscape.canonical.com
 * Management:
* Support:
                  https://ubuntu.com/advantage
 System information as of Sat Jul 9 14:52:19 KST 2022
 System load: 0.06
                                  Processes:
 Usage of /: 0.5% of 250.98GB Users logged in:
                                  IPv4 address for eth0: 172.30.199.88
 Memory usage: 0%
 Swap usage:
 update can be applied immediately.
To see these additional updates run: apt list --upgradable
The list of available updates is more than a week old.
To check for new updates run: sudo apt update
This message is shown once a day. To disable it please create the
/home/me/.hushlogin file.
 e@DESKTOP-HAAIOJO:~$
```

Microsoft Windows [Version 10.0.19044.1766] (c) Microsoft Corporation. All rights reserved.

C:₩windows\system32>wsl --set-default-version 2 WSL 2와의 주요 차이점에 대한 자세한 내용은 https://aka.ms/wsl2를 참조하세요 작업을 완료했습니다.

C:₩Windows₩system32>\_

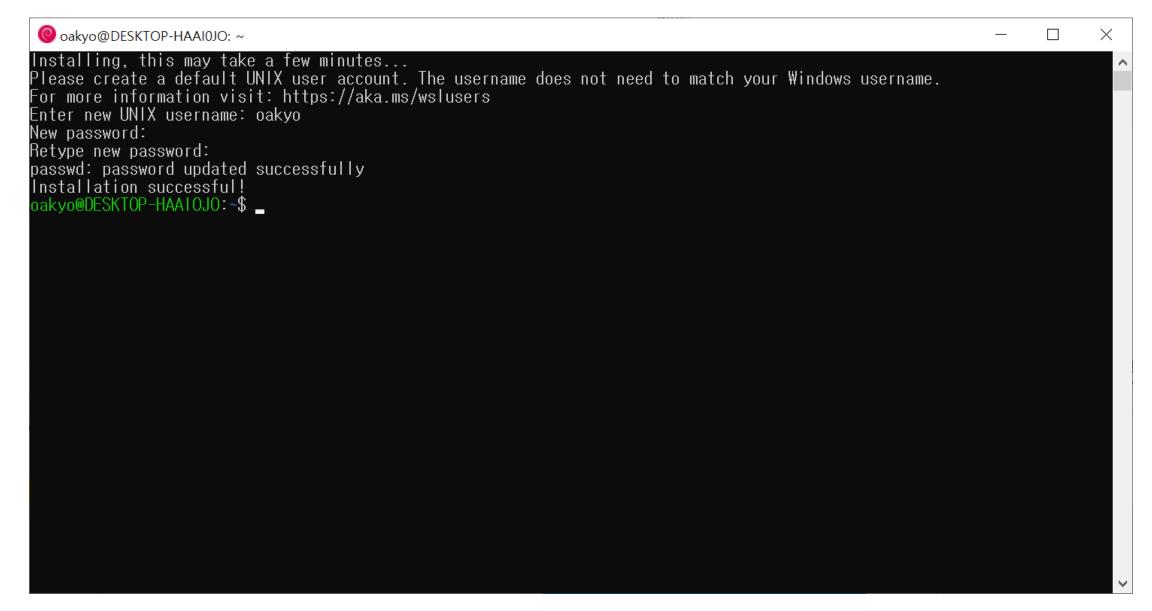
V



# Debian 설치 과정



# Debian 설치 과정



# Python 설치 과정

- \$ sudo apt update
- \$ sudo apt upgrade
- \$ sudo apt-get install wget

#### Anaconda 설치: <a href="https://repo.anaconda.com/archive/">https://repo.anaconda.com/archive/</a>

- \$ wget <a href="https://repo.anaconda.com/archive/Anaconda3-2021.11-Linux-x86">https://repo.anaconda.com/archive/Anaconda3-2021.11-Linux-x86</a> 64.sh
- \$ bash Anaconda3-2021.11-Linux-x86\_64.sh

#### .profile 에 PATH 추가

\$ export PATH=/home/{login ID}/anaconda3/bin:\$PATH \$ . .profile

### Python version 확인

\$ python –version

### PC(Local) 파일 확인

\$ explorer.exe .

### Java 설치 과정

\$ sudo apt install default-jre

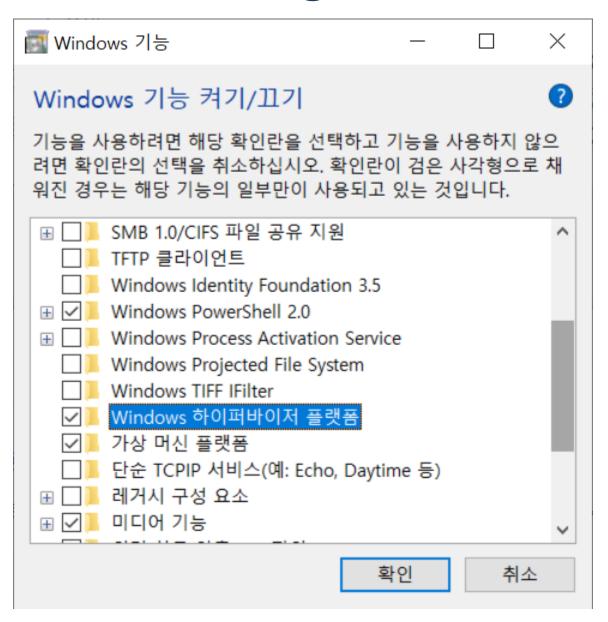
\$ sudo apt install default-jdk

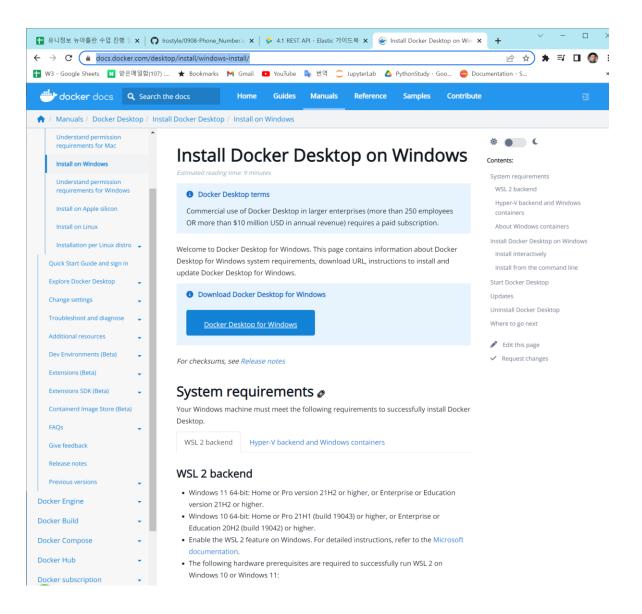
\$ export JAVA\_HOME=/usr/lib/jvm/java-1.11.0-openjdk-amd64

\$ java –version

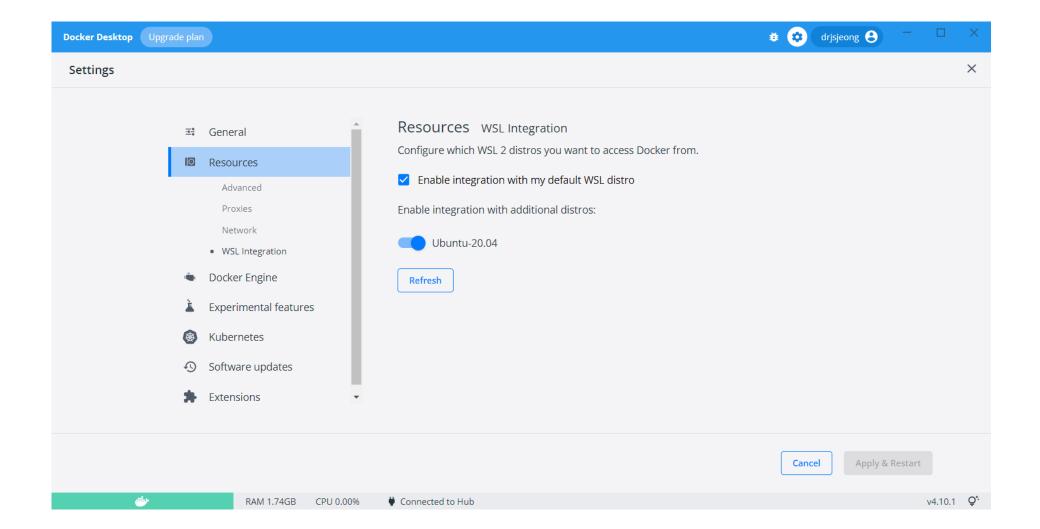
https://serverspace.io/support/help/how-to-install-java-with-apt-on-ubuntu-18-04/

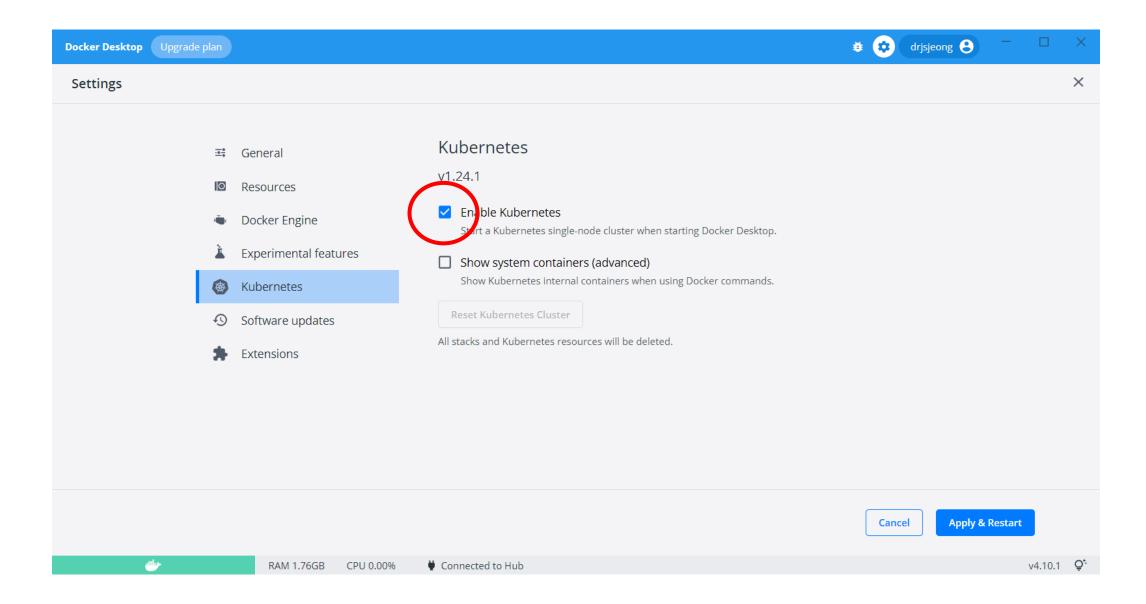
# **Kubernetes Getting Started on Windows**

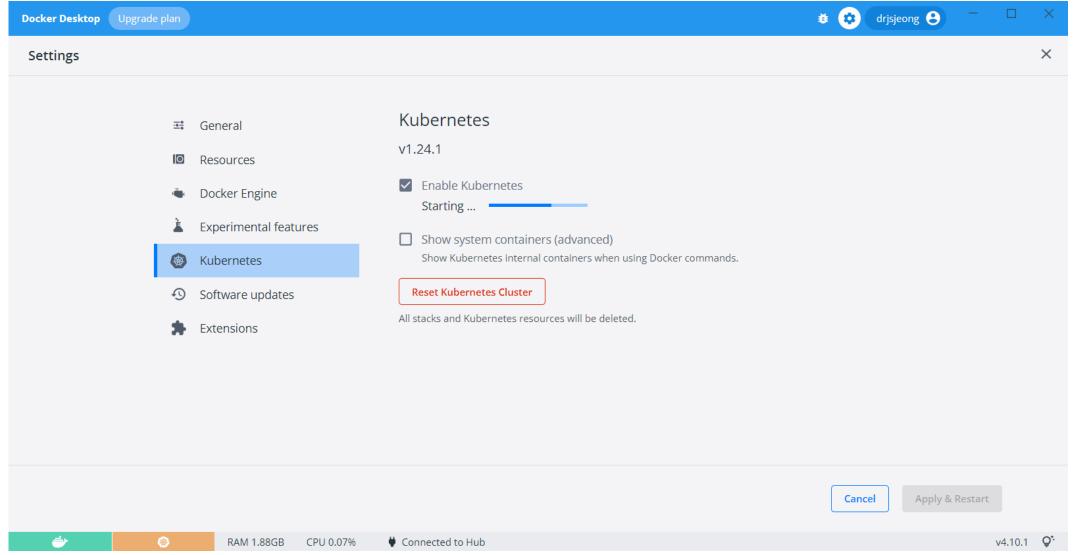




https://docs.docker.com/desktop/install/windows-install/







# Kubernetes 설치 확인

```
📆 명령 프롬프트
                                                                                                             "kind": "ReplicaSetList",
   "apiVersion": "apps/v1",
   "metadata": {
       "resourceVersion": "716"
   "kind": "PodList",
   "apiVersion": "v1",
    "metadata": {
        "resourceVersion": "716"
   "items": []
 :#Users#oakyo>kubectl config current-context
docker-desktop
 :#Users#oakyo>kubect| get node
                STATUS ROLES
                                        AGE VERSION
docker-desktop Ready control-plane 8mls v1.24.1
 ::#Users#oakyo>kubect| get pods
No resources found in default namespace.
 ∷#Users#oakyo>
 ::#Users#oakyo>_
```

- C> kubectl config current-context
- C> kubectl get node
- C> kubectl get pods