CH1. 빅데이터 처리 개요

전성환

72160261@dankook.ac.kr

• 빅데이터 처리 과정과 기술







• 빅데이터 분석 및 처리 기술

기술 분류	설명	세부기술	
분석기법	데이터 집합을 분석하는데 활용될 수 있는 통계 및 컴퓨터 공학 분야의 다양한 기법	A/B testing Association rule learning Classification Cluster analysis Crowdsourcing Data fusion and integration Data mining Ensemble learning Genetic algorithms Machine learing Natural language processing Network analysis Neural networks	Optimization Pattern recognition Predictive modeling Regression Sentiment analysis Signal processing Simulation Spatial analysis Statistics Supervised learning Time series analysis Unsupervised learning Visualization
처리기술	분석에 필요한 데이터를 수집, 조작, 관리하거나 분석기법을 지원하기 위해 개발된 기술	Big Table Business Intelligence Cassandra Cloud computing Data mart Data warehouse Distributed system Dynamo Extract, transform and load Google File System Hadoop Hbase	MapReduce Mashup Metadata Non-relational R Relational database Semi-structured SQL Stream processing Structured data Unstructured data Visualization

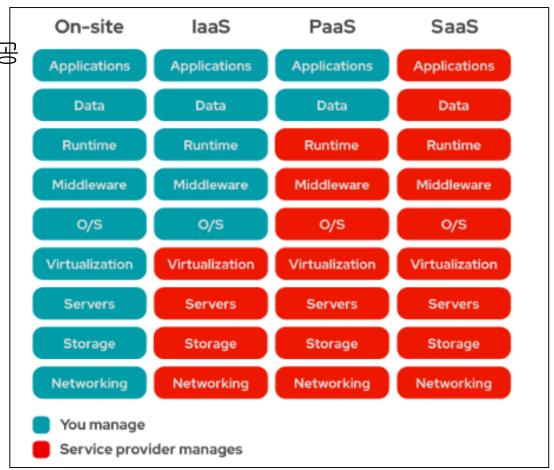






• 클라우드 컴퓨팅

- IaaS(Infrastructure as a Service)
 - 서버, 스토리지, 네트워크를 가상화 환경으로 제공
 - AWS EC2
- PaaS(Platform as a Service)
 - 서비스를 개발할 수 있는 환경 제공
 - Microsoft Azure
- SaaS(Software as a Service)
 - 클라우드 환경에서 동작하는 응용 프로그램
 - 네이버 클라우드







• 클라우드 컴퓨팅의 이점

왜 Cloud Computing 인가? - 6가지 이점

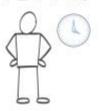
초기 선투자 비용 없음

고정비용을 가변비용으로 대체 미리 서버를 구매할 필요 없음



속도 및 민첩성

수 분 만에 인프라 구축 가능 빠르게 변화에 대응



운영 비용 절감

사용한 만큼만 지불하며 규모의 경제로 인한 지속적인 비용 절감



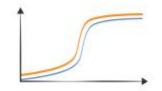
비즈니스에만 집중 가능

혁신을 위한 다양한 실험 가능 불필요한 인프라 관리 업무 제거



탄력적인 운영 및 확장

필요 용량에 대한 예측 불필요 수요에 맞춘 유연한 확장



글로벌 확장

빠른 시간 내 글로벌 서비스 구현 가능









- 아마존 EC2는 클라우드에서 컴퓨팅의 규모를 자유롭게 변경할 수 있는 웹 서비스 이며, 개발자가 웹 스케일 컴퓨팅을 쉽게 사용할 수 있도록 설계함.
- EC2는 웹 서비스 인터페이스를 사용하여 다양한 운영체제로 인스턴스를 시작.
- EC2는 메모리 및 CPU 자원을 시간 단위로 지정하여 컴퓨터를 임대할 수 있는 웹 서비스.
- EC2는 슈퍼유저 권한으로 리눅스나 윈도우 서버 네트워크에 접근할 수 있으며, 소프트웨어 설치 및 시스템 설정이 자유로움.
- Elastic Block Storage 의로 이 보고 보고 되었다. Elastic Block Storage 의로 기존의 파일 시스템과 동시에 활용할 수 있고, 엘라스틱 Elastic 맵리듀스 서비스로 하둡 클러스터를 간단히 구성함.
- 가상화 계층이 실제 하드웨어 구성 설정을 제한하기는 하지만, 네트워크 통신이나 데이터 전송 과정에서 병목 현상이 발생하면 자동으로 많은 가상 머신을 활용하므로 성능이 비교적 낮은 시 스템으로도 문제를 해결할 수 있다는 장점.







• 학습 목표

- Amazon EC2 개요에 대해 알 수 있다.
- Amazon EC2에서 제공하는 기능 및 활용에 대해 알 수 있다.
- Amazon EC2를 실질적으로 사용하기 위한 방법에 대해 알 수 있다.

• 학습 내용

- Amazon EC2 개요
- Amazon EC2에서 제공하는 인스턴스 종류 및 부가기능
- 회원가입/로그인 및 서버 위치 설정
- 인스턴스 네트워크 및 보안 설정
- 인스턴스 생성
- 엘라스틱 IP 생성 및 할당
- Putty를 이용한 인스턴스 접속





1.1 Amazon EC2 개요

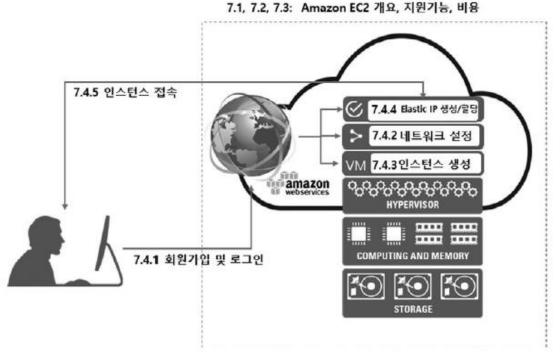
• Amazon에서 제공하는 IaaS서비스 중 하나인 Amazon

EC2에 대하여 알아보고 Amazon EC2에서 제공하는 기능

을 활용하여 사용자가 원하는 서버를 생성하고 이를 관

리하기 위한 실습을 수행

Amazon EC2계정생성 및 사용자가 원하는 서버를 생성하고 이를 관리하기 위한 네트워크 및 보안설정을 위한 실습을 수행 SSH을 이용하여사용자가 생성한 서버에 접속하기 위한 실습을 수행







1.1 Amazon EC2 개요

Amazon Elastic Compute Cloud(Amazon EC2)

- 아마존의 대표적인 Iaas 서비스
- 하드웨어 서버를 가상화하여 하드웨어 자원을 사용자에게 제공
- 사용자는 아마존에서 제공해주는 가상화 서버 위에 운영체제와 소프트웨어를 설치해 서버 인프라를 구축할 수 있음
- 사용자는 물리적인 하드웨어를 구비하지 않고도 더욱 빠르게 원하는 서버 인프라를 구축할 수 있음





1.1 Amazon EC2 개요

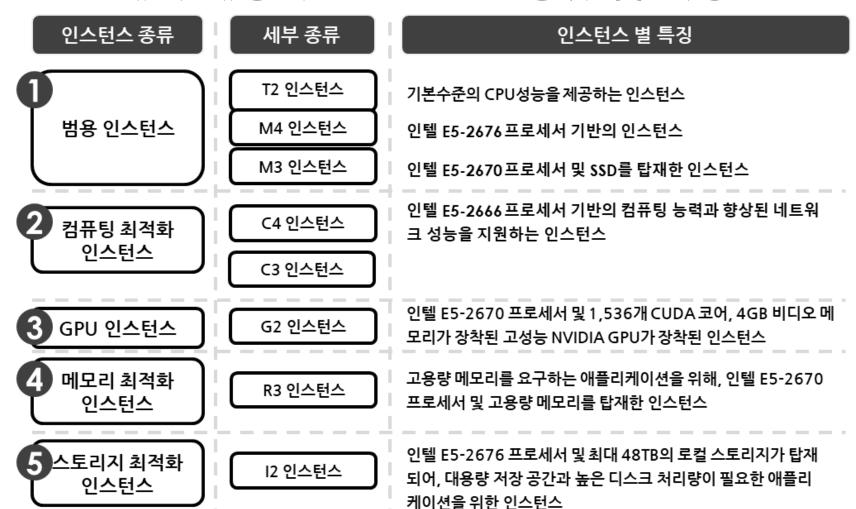
- Amazon EC2 인스턴스 유형
 - 각 사용 사례에 최적화된 다양한 인스턴스(가상 서버) 유형 제공
 - 인스턴스 유형
- CPU, 메모리, 스토리지 및 네트워킹 용량의 다양한 조합으로 구성
- 애플리케이션에 따라 적합한 리소스 조합을 선택할 수 있음
- 각 인스턴스 유형은 하나 이상의 인스턴스 크기를 포함
 - •목표로 하는 워크로드 요구 사항까지 리소스를 확장할 수 있음





1.2 Amazon EC2에서 제공하는 인스턴스 및 부가기능

• Amazon EC2에서 제공하는 인스턴스 종류 및 특징



1.2 Amazon EC2에서 제공하는 인스턴스 및 부가기능

• Amazon EC2에서 제공하는 부가기능

부가기능 종류 부가기능에 대한 설명 인스턴스와 연동이 가능한 영구적인 저장 공간 제공 Elastic Block Store(EBS) 인스턴스추가비용을지불하여BS에대한높은IOPS(IOperSecond)를제공 EBS 최적화 인스턴스 다중위치 인스턴스를여러위치에배치할수있는기능제공 외부에서클라우드컴퓨팅내부인스턴스로접근이가능한고정된P 엘라스틱 IP 주소 클라우드 리소스 및 애플리케이션에 대한 모니터링 기능 지원 Amazon CloudWatch 사용자가정의한조건에따라인스턴스의용량을자동으로확장하거나축소 오토스케일링 엘라스틱 로드밸런싱 인스턴스가 수신하는 애플리케이션 트래픽을 여러 인스턴스에 분신하여 각 인 스턴스에발생하는로드를분산 클러스터 컴퓨팅, 클러스터 GPU, 고용량 메모리 클러스터 기능을 제공하여 고성 HPC 클러스터 능컴퓨팅환경제공





• AWS 프리티너



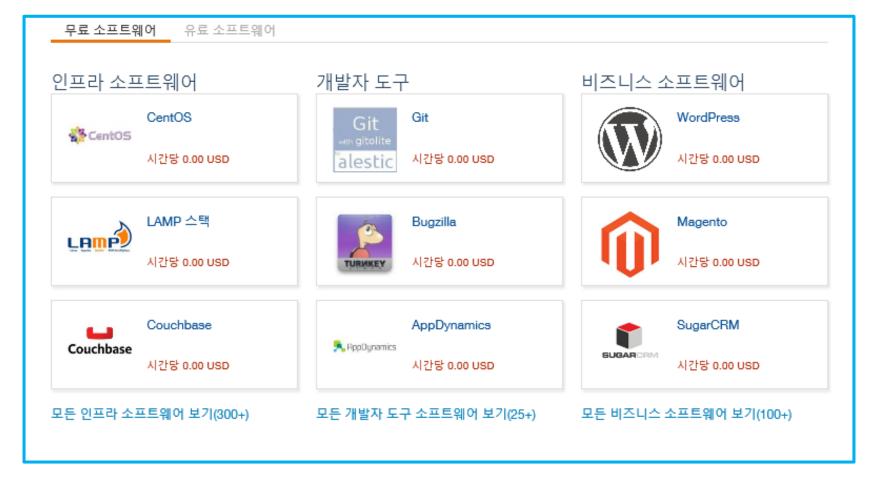


- Linux를 구동하는 EC2 t2.micro 인스턴스 750시간 사용
- Microsoft Windows Server를 구동하는 EC2 t2.micro 인스턴스 750시간 사용
- 엘라스틱 로브 밸런싱 750시간 및 15GB의 데이터 처리 제공
- EBS 30GB 및 스냅샷 스토리지 1GB 제공
- 모든 AWS 서비스를 합산하여 15GB의 데이터 전송
- 1GB의 내부 데이터 전송 제공





• AWS 프리티너







서비스에 요금이 부과되지 않는 경우 사용 범위에서는 한도가 있을 수 있음

많은 서비스가 여러 유형의 한도를 가짐

- 시간, 분 단위 등의 시간 한도
 - Amazon EC2에는 사용 가능한 인스턴스의 유형과 1개월 동안 사용 가능한 시간 모두에 대해 한도가 있음
- 사용자가 서비스로 전송하는 요청, 즉 API작업을 기준으로 하는 요청 한도
 - Amazon S3는 사용 가능한 메모리 사용량에 대한 한도가 있으며 매달 특정 작업을 호출할 수 있는 횟수에 대한 한도도 있음



- 온디멘드 인스턴스 방식의 요금 체계
 - 컴퓨팅 자원의 사용량에 따른 비용을 지불하는 방식
 - 필요에 따라 컴퓨팅 자원을 증감할 수 있음







- 예약 인스턴스 방식의 요금체계
 - 계약기간 동안 예약한 인스턴스를 자유롭게 사용하는 방식
 - 온디멘드 인스턴스 방식의 요금체계와 비해, 비교적 저렴

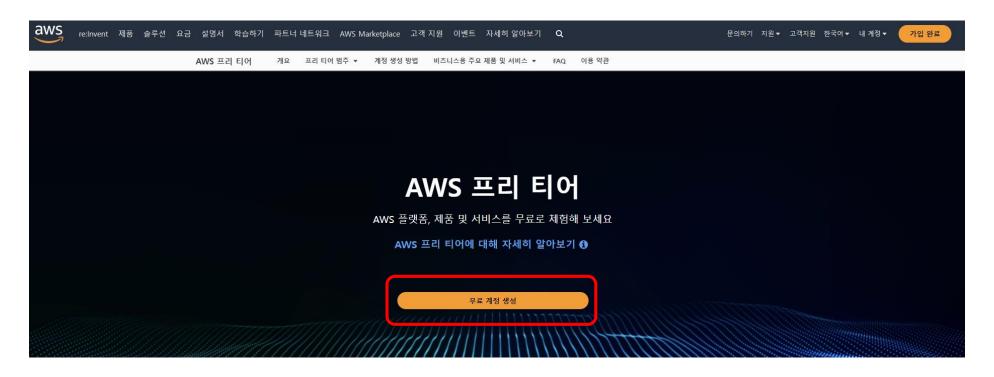


사용이 꾸준하거나 예측 가능한 애플리케이션, 예약 용량이 필요한 애플리케이션, 또는 총 컴퓨팅 비용을 절감하기 위해 선 수금 지급이 가능한 경우 적합





- 회원가입 / 로그인 및 서버 위치 설정
 - Amazon Web Service 초기 화면 (https://aws.amazon.com)
 - 계정이 없는 사용자는 계정을 생성 후 다음 단계를 진행







• 회원가입 / 로그인 및 서버 위치 설정



새로운 AWS 계정으로 프리 티어 제품 을 살펴보세요.

자세히 알아보려면 aws.amazon.com/free를 방문하세요.



AWS에 가입

루트 사용자 이메일 주소 계정 복구 및 일부 관리 기능에 사용

AWS 계정 이름

계정의 이름을 선택합니다. 이름은 가입 후 계정 설정에 서 변경할 수 있습니다.

이메일 주소 확인

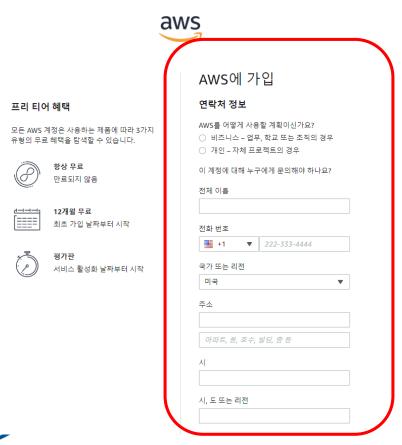
또는

기존 AWS 계정에 로그인





• 회원가입 / 로그인 및 서버 위치 설정





AWS에 가입

Support 플랜 선택

비즈니스 또는 개인 계정에 대한 Support 플랜을 선택합니다. 플랜 및 요금 예시를 비교 🗹해 보세요. nsole에서 플래을 변경할 수 있습니다.

Ο 기본 지원 - 무료

- AWS를 처음 시작하는 신규 사용자에게 권장
- AWS 리소스에 대한 연 중무휴 24시간 셀프 서 비스 액세스
- 계정 및 청구 문제 전
- Personal Health Dashboard 및 Trusted Advisor에 대한 액세스



- 개발자 지원 시작가 ○ 비즈니스 지원 - 시작 는 29 USD/윌
- AWS를 체험해보는 개 발자에게 권장
- 업무 시간 중 AWS Support에 대한 이메 일 액세스
- 12시간(업무 시간 기 준) 이내의 응답 시간



- 가는 100 USD/윌
- AWS 기반 프로덕션 워 크로드 실행에 추천
- 이메일, 전화 및 채팅 을 통한 연중무휴 24시 간 기술 지원
- 1시간 이내의 응답 시
- Trusted Advisor 모범 사례 권장 사항 전체





엔터프라이즈 수준의 지원이 필요하신가요?

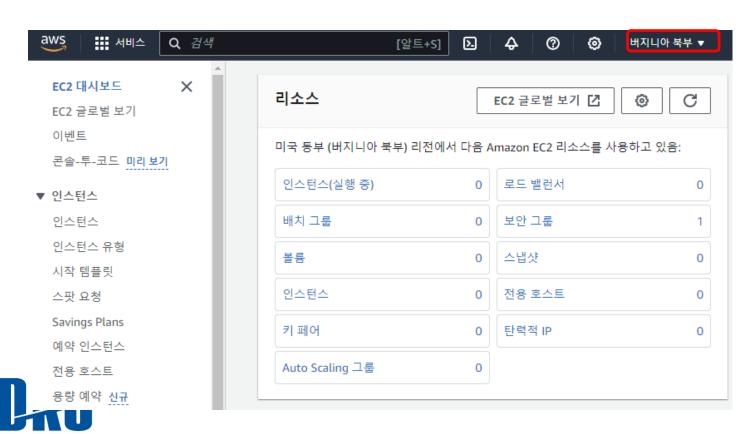
최저 윌 15,000 USD로 15분 이내에 응답을 받을 수 있으며 기술 지원 관리자가 배정된 컨시어지 스타일의 서비스를 이용할 수 있습니다. 자세히 알아보기 ☑





• 회원가입 / 로그인 및 서버 위치 설정

- Amazon EC2 인스턴스 서버 위치 설정
- Amazon AWS는 전 세계 곳곳에 서버를 구축하고 있으므로 인스턴스 사용 목적에 적합한 서버의 위치를 선택

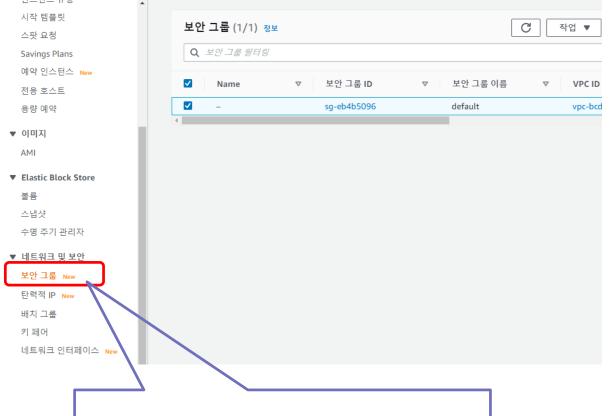


인스턴스 서버의 물리적 위치를 선택



- 인스턴스 네트워크 및 보안 설정
 - 보안 그룹 생성을 위한 웹 인터페이스

인스턴스를 생성하기 전 보안그룹(Security Group)을 설정 보안그룹은 추후 생성할 인스턴스에 적용할 방화벽 규칙

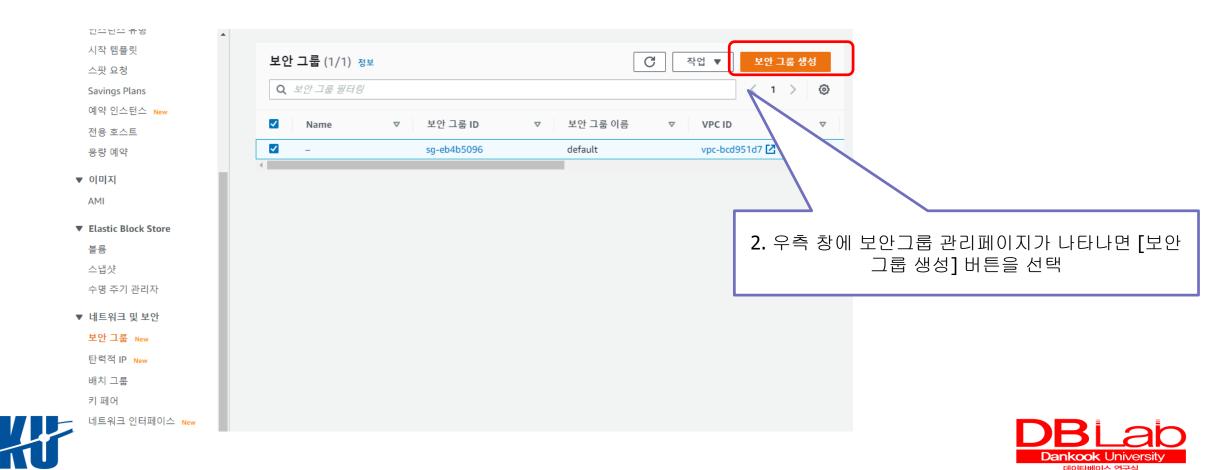




1. 좌측 메뉴 리스트에서 [보안 그룹]을 선택



- 인스턴스 네트워크 및 보안 설정
 - 보안 그룹 생성을 위한 웹 인터페이스

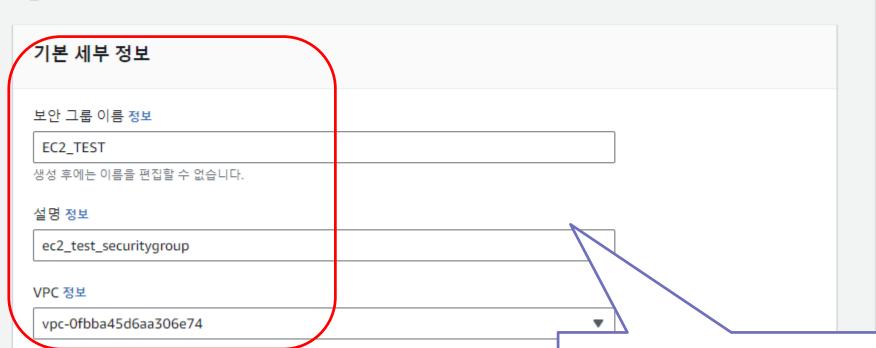


• 인스턴스 네트워크 및 보안 설정

• 보안 그룹 생성을 위한 웹 인터페이스

보안 그룹 생성 정보

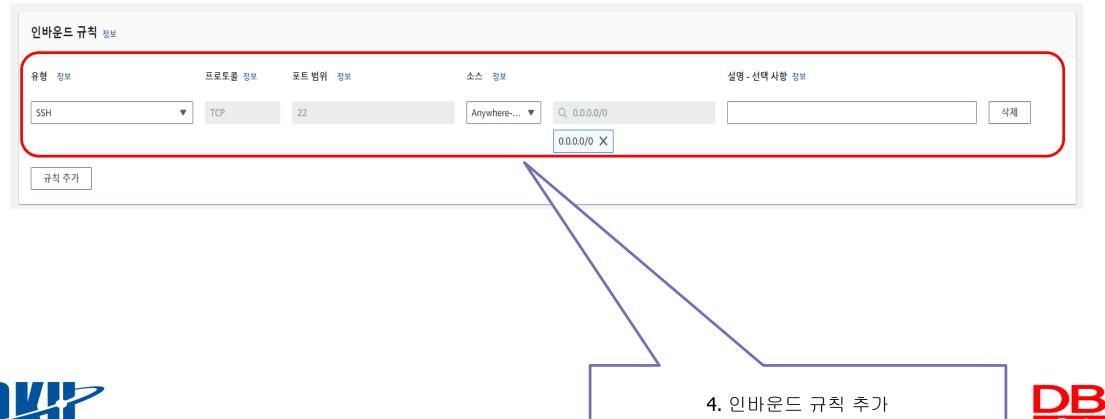
보안 그룹은 인바운드 및 아웃바운드 트래픽을 관리하는 인스턴스의 가상 방화벽 역할을 합니다. 새 보안 그룹을 생성하려면 아래의 필드를 작성하십시오.



3. 보안그룹 이름 및 보안그룹에 대한 설명을 기술하고 VPC는 Default 값을 선택



- 인스턴스 네트워크 및 보안 설정
 - 보안 그룹 생성을 위한 웹 인터페이스

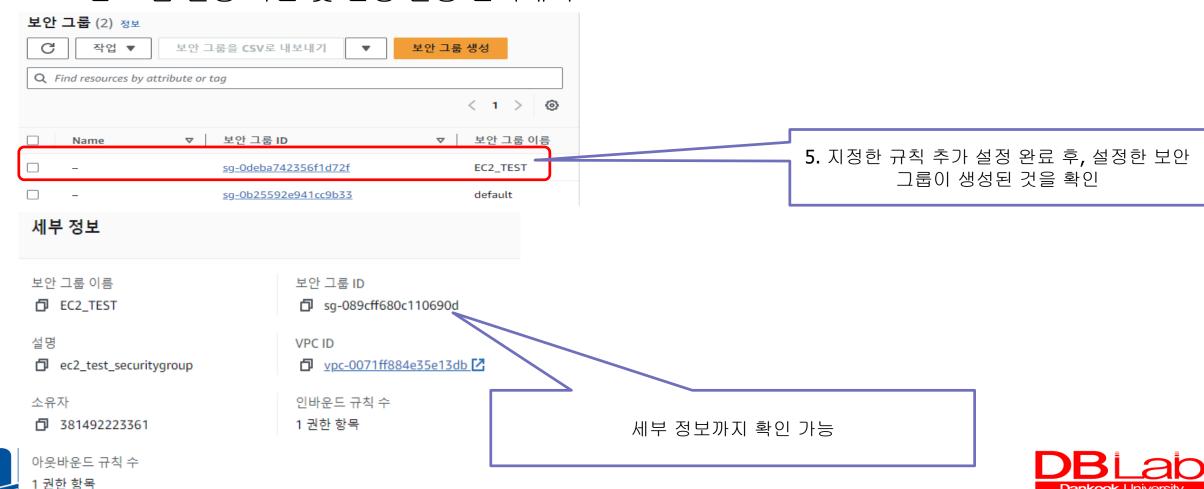






• 인스턴스 네트워크 및 보안 설정

• 보안 그룹 설정 확인 및 환경 설정 인터페이스



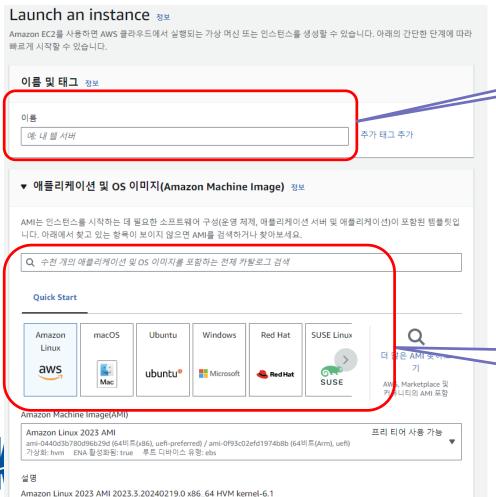
• 인스턴스 생성

• Amazon EC2 인스턴스 생성 웹 인터페이스



• 인스턴스 생성

• Amazon EC2 인스턴스 생성 웹 인터페이스



2. 인스턴스 생성 후 인스턴스 이름 설정

3. 애플리케이션 및 OS 선택
-> 검색으로 원하는 애플리케이션 및 OS를 선택해도 되고, Quick Start에서 선택해도 됨.





• 인스턴스 생성

- Amazon EC2 인스턴스 생성 웹 인터페이스
- ▼ **인스턴스 유형** 정보 | 조언받기

4. 인스턴스 유형을 프리 티어로 선택

인스턴스 유형

t2.micro

프리 티어 사용 가능

₩

패밀리: t2 1 vCPU 1 GiB 메모리 현재 세대: true 온디맨드 Windows 기본 요금: 0.0162 USD 시간당 온디맨드 SUSE 기본 요금: 0.0116 USD 시간당 온디맨드 RHEL 기본 요금: 0.0716 USD 시간당 온디맨드 Linux 기본 요금: 0.0116 USD 시간당

소프트웨어가 사전 설치된 AMI에는 추가 비용이 적용됩니다.

모든 세대

인스턴스 유형 비교

▼ 키 페어(로그인) 정보

키 페어를 사용하여 인스턴스에 안전하게 연결할 수 있습니다. 인스턴스를 시작하기 전에 선택한 키 페어에 대한 액세스 권 한이 있는지 확인하세요. 5. 새 키 페어 생성 선택

키 페어 이름 - 필수

선택

C 새 키 페어 생성



• 인스턴스 생성

• Amazon EC2 인스턴스 생성 웹 인터페이스

키 페어 생성

×

키 페어 이름

키 페어를 사용하면 인스턴스에 안전하게 연결할 수 있습니다.

testkey

이름에는 최대 255개의 ASCII 문자가 포함됩니다. 앞 또는 뒤에 공백을 포함할 수 없습니다.

키 페어 유형

RSA

RSA 암호화된 프라이빗 및 퍼블릭 키 페 어 O ED25519

ED25519 암호화된 프라이빗 및 퍼블릭 키 페어

프라이빗 키 파일 형식

o .pem

OpenSSH와 함께 사용

O .ppk PuTTY와 함께 사용

▲ 메시지가 표시되면 프라이빗 키를 사용자 컴퓨터의 안전하고 액세스 가능한 위치에 저장합니다. 나중에 인스턴스에 연결할 때 필요합니다. 자세히 알아보기

6. 키 페어 이름 작성 -> 키 페어 유형 RSA -> 프라이빗 키 파일 형식 pem -> 키 페어 생성 클릭 -> 생성하면 자동으로 키 페어 다운로드가 됨



• 인스턴스 생성

- Amazon EC2 인스턴스 생성 웹 인터페이스
- ▼ **인스턴스 유형** 정보 | 조언받기

4. 인스턴스 유형을 프리 티어로 선택

인스턴스 유형

t2.micro

프리 티어 사용 가능

₩

패밀리: t2 1 vCPU 1 GiB 메모리 현재 세대: true 온디맨드 Windows 기본 요금: 0.0162 USD 시간당 온디맨드 SUSE 기본 요금: 0.0116 USD 시간당 온디맨드 RHEL 기본 요금: 0.0716 USD 시간당 온디맨드 Linux 기본 요금: 0.0116 USD 시간당

소프트웨어가 사전 설치된 AMI에는 추가 비용이 적용됩니다.

모든 세대

인스턴스 유형 비교

▼ 키 페어(로그인) 정보

키 페어를 사용하여 인스턴스에 안전하게 연결할 수 있습니다. 인스턴스를 시작하기 전에 선택한 키 페어에 대한 액세스 권 한이 있는지 확인하세요. 5. 새 키 페어 생성 선택

키 페어 이름 - 필수

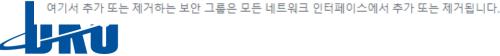
선택

C 새 키 페어 생성



• 인스턴스 생성

• Amazon EC2 인스턴스 생성 웹 인터페이스 6. 생성한 보안그룹 선택 ▼ 네트워크 설정 정보 편집 네트워크 정보 vpc-0fbba45d6aa306e74 서브넷 정보 기본 설정 없음(가용 영역의 기본 서브넷) 퍼블릭 IP 자동 할당 정보 활성화 방화벽(보안 그룹) 정보 ○ 보안 그룹 생성 ○ 기존 보안 그룹 선택 일반 보안 그룹 정보 보안 그룹 선택 보안 그룹 규칙 비교 EC2_TEST sg-0deba742356f1d72f X VPC: vpc-0fbba45d6aa306e74



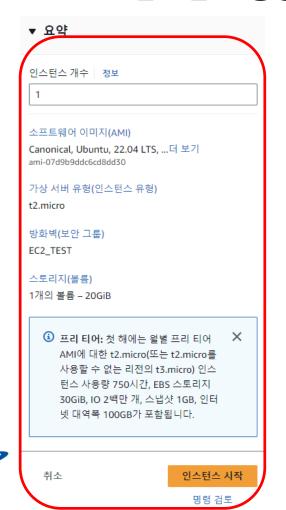
• 인스턴스 생성

• Amazon EC2 인스턴스 생성 웹 인터페이스 7. 프리 티어는 최대 30GB까지 선택가능. ▼ 스토리지 구성 정보 루트 볼륨 (암호화되지 않음) GiB 1x 20 gp2 ① 프리 티어를 사용할 수 있는 고객은 최대 30GB의 EBS 범용(SSD)또는 마그네틱 스토리지를 사용할 수 있습 니다. 새 볼륨 추가 선택한 AMI에 인스턴스가 허용하는 것보다 많은 인스턴스 스토어 볼륨이 포함되어 있습니다. AMI에서 처음 0개의 인스턴스 스토어 볼륨에만 액세스할 수 있습니다. ① 백업 정보를 보려면 새로 고침 클릭 C 할당한 태그에 따라 Data Lifecycle Manager 정책으로 인스턴스를 백업할지 여부가 결정됩 니다. 0 x 파일 시스템 편집



• 인스턴스 생성

• Amazon EC2 인스턴스 생성 웹 인터페이스



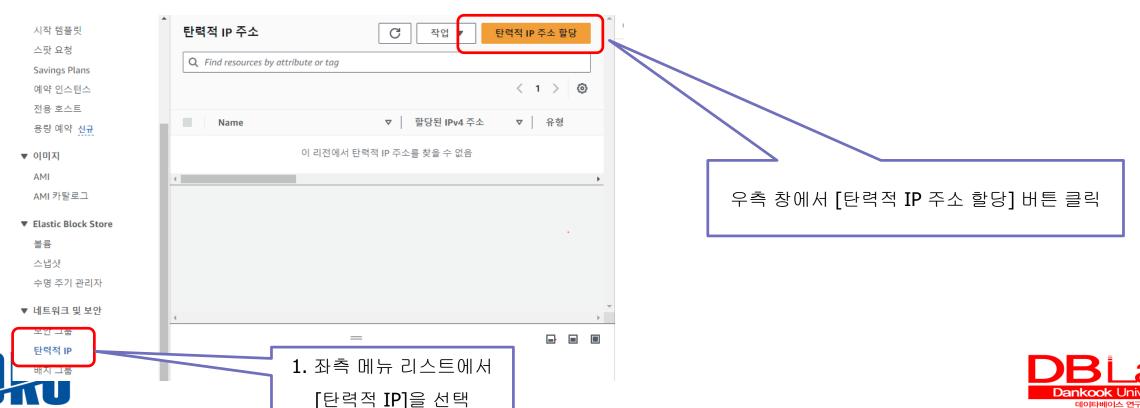
8. 요약 탭에서 설정한 인스턴스 개수, 스토리지 등다양한 정보를 확인할 수 있다. 인스턴스 시작 클릭.





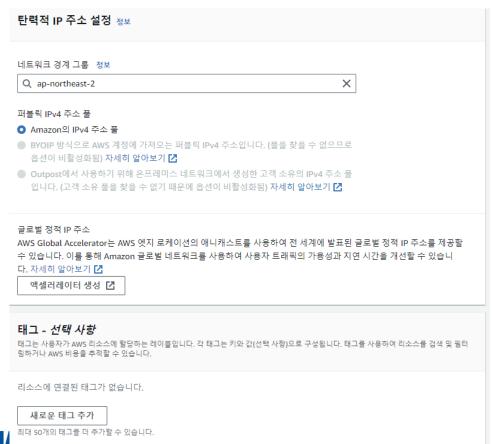
• 엘라스틱 IP (탄력적 IP) 생성 및 할당

- 인스턴스 생성이 끝난 뒤 해당 인스턴스의 Public DNS를 사용하여 생성한 인스턴스에 접속 할 수 있음
- 긴 접속 주소로 인해 불편함을 초래되거나, 클라우드 기반 서비스를 제공하기 위해 고정된 접 속주소가 필요한 경우가 발생함





• 엘라스틱 IP (탄력적 IP) 생성 및 할당

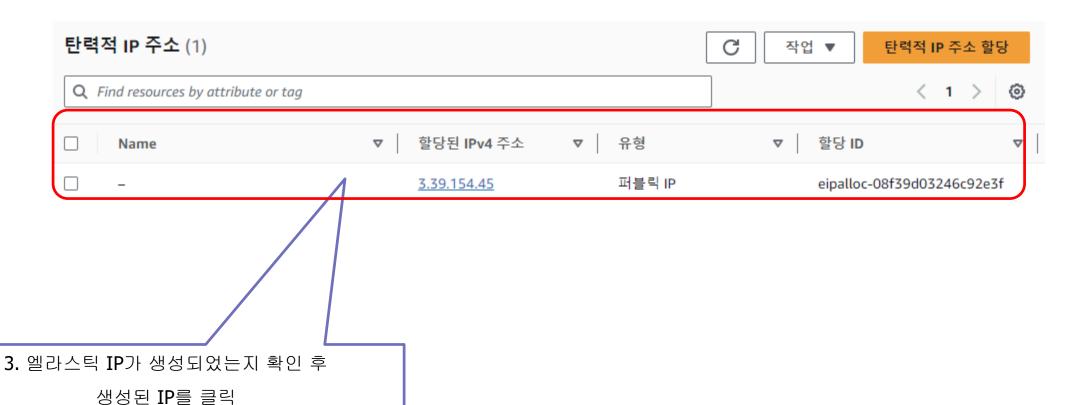


2. Default 값으로 [할당] 버튼 클릭



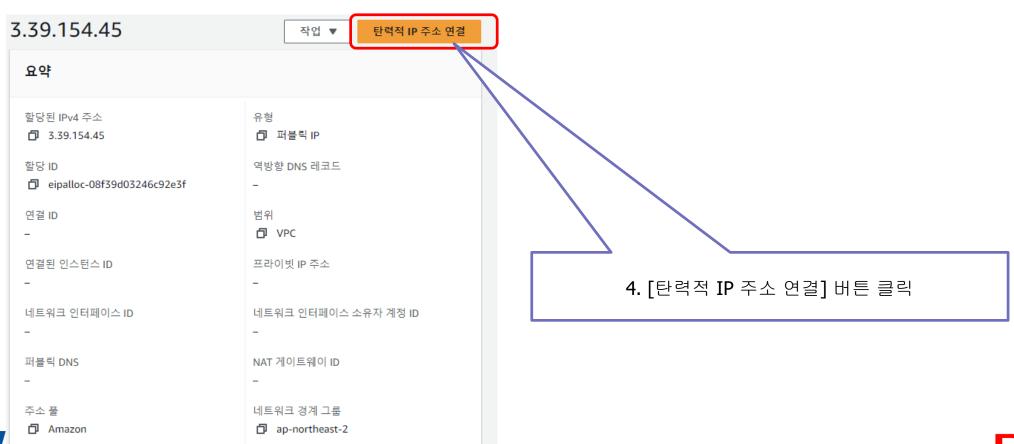


- 엘라스틱 IP (탄력적 IP) 생성 및 할당
 - 엘라스틱 IP와 인스턴스 연결





- 엘라스틱 IP (탄력적 IP) 생성 및 할당
 - 엘라스틱 IP 할당 인터페이스





• 엘라스틱 IP (탄력적 IP) 생성 및 할당

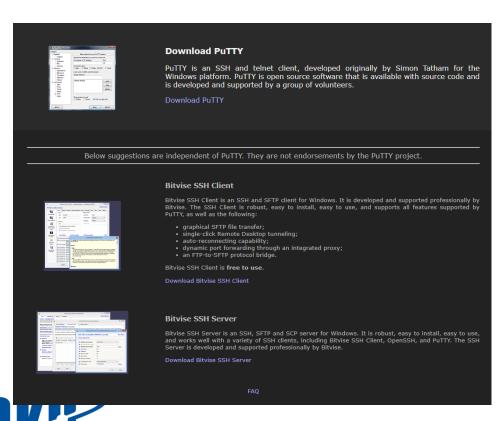


5. 생성한 인스턴스와 프라이빗 IP설정 후 [연결] 버튼 클릭





- Putty를 이용한 인스턴스 접속
 - 인스턴스에 접속하기 위해 가장 많이 사용되는 프로그램인 Putty를 사용
 - 접속에 필요한 키파일을 생성하기 위해 PuTTYgen 프로그램을 이용

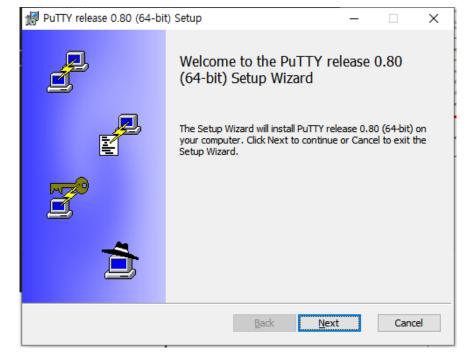


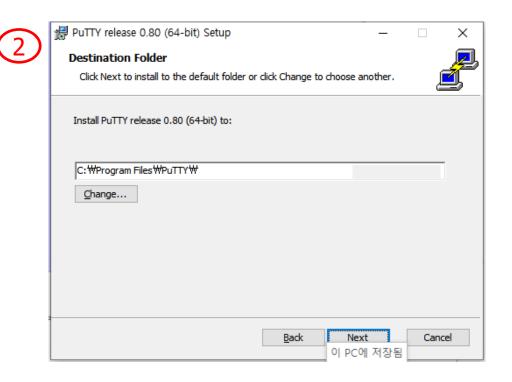




• Putty를 이용한 인스턴스 접속



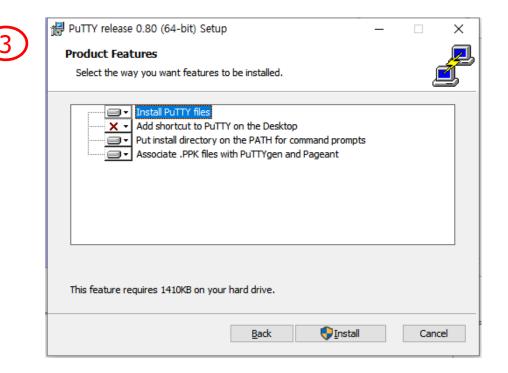


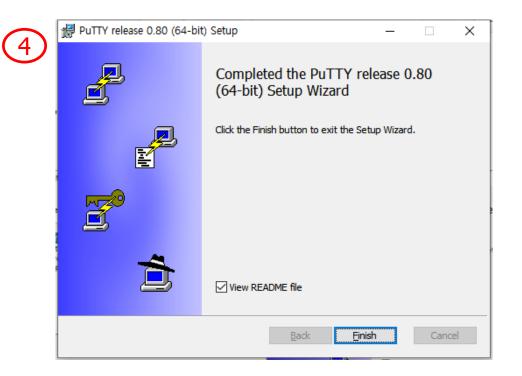






• Putty를 이용한 인스턴스 접속

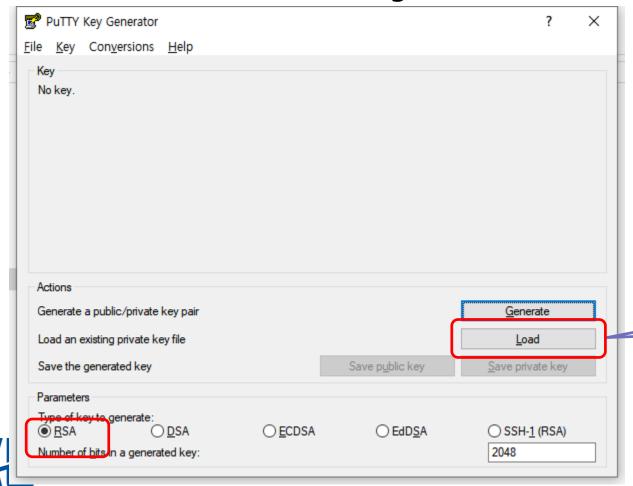








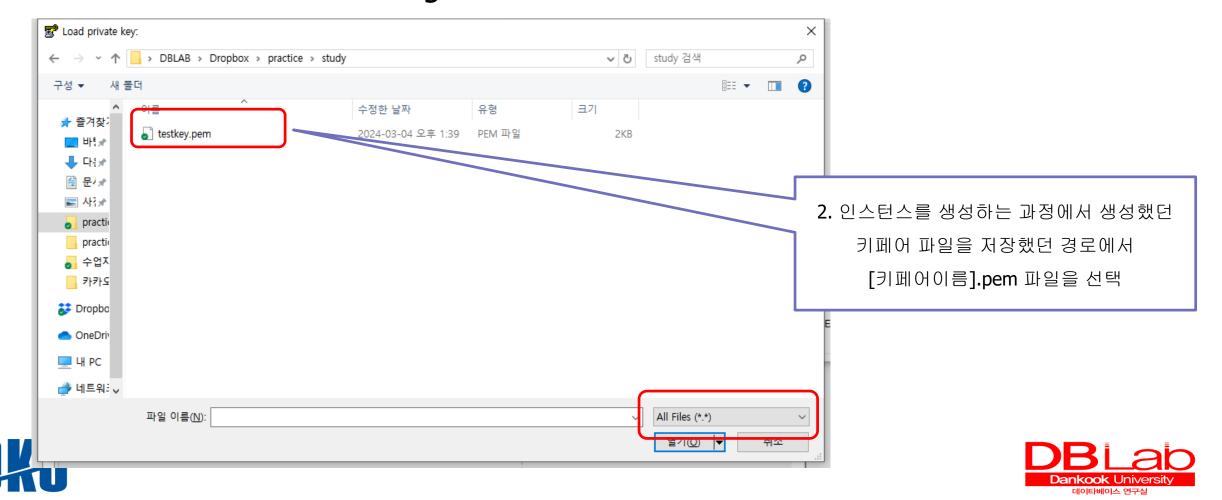
- Putty를 이용한 인스턴스 접속
 - 개인키 생성을 위한 PuTTYgen 인터페이스



1. PuTTYgen을 이용하여 개인키 파일을 생성

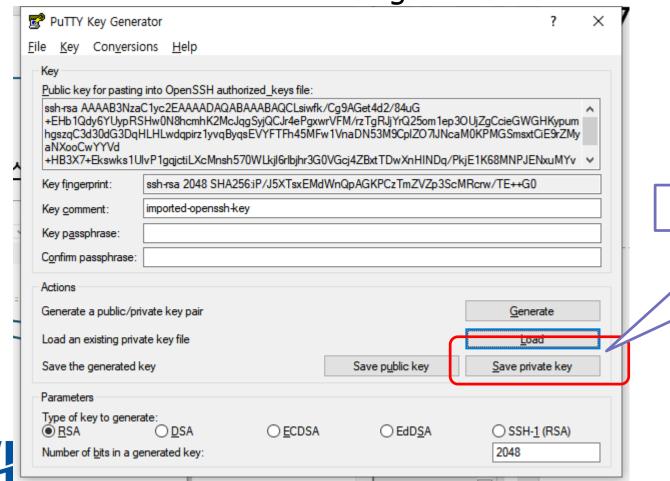


- Putty를 이용한 인스턴스 접속
 - 개인키 생성을 위한 PuTTYgen 인터페이스



• Putty를 이용한 인스턴스 접속

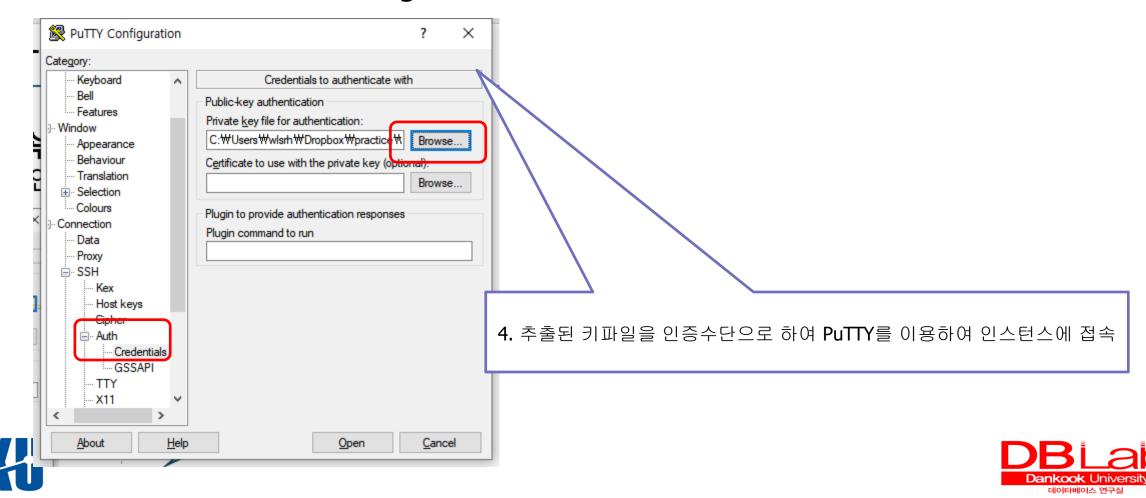
• 개인키 생성을 위한 PuTTYgen 인터페이스



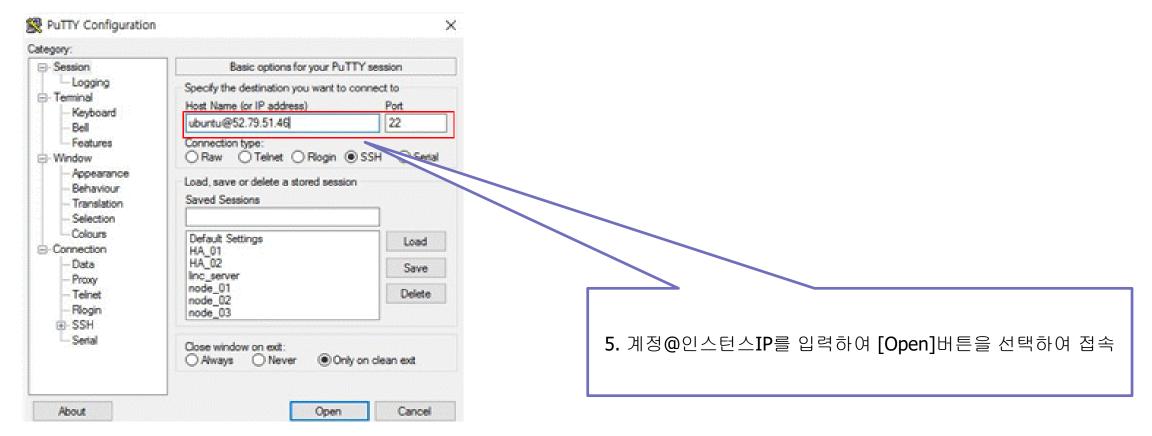
3. 키페어 파일에서 추출한 개인키 파일을 저장



- Putty를 이용한 인스턴스 접속
 - 개인키 생성을 위한 PuTTYgen 인터페이스



- Putty를 이용한 인스턴스 접속
 - PuTTY 접속을 위한 IP 설정







- Putty를 이용한 인스턴스 접속
 - AMI별 기본 계정

Amazon Linux AMI의 경우 사용자 이름은 'ec2-user'

RHEL5 AMI의 경우 사용자 이름은 'root' 또는 'ec2-user'

Ubuntu AMI의 경우 사용자 이름은 'ubuntu'

Fedora AMI의 경우 사용자 이름은 'fedora' 또는 'ec2-user'

SUSE Linux의 경우 사용자 이름은 'root' 또는 'ec2-user'





- Putty를 이용한 인스턴스 접속
 - PuTTY를 이용한 인스턴스 접속

```
@ ubuntu@ip-172-31-9-52: ~
 System load: 0.0
                                  Processes:
                                                       97
 Usage of /: 2.6% of 29.39GB Users logged in:
 Memory usage: 5%
                                 IP address for eth0: 172.31.9.52
 Swap usage: 0%
 Graph this data and manage this system at:
   https://landscape.canonical.com/
 Get cloud support with Ubuntu Advantage Cloud Guest:
   http://www.ubuntu.com/business/services/cloud
 packages can be updated.
 updates are security updates.
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
ubuntu@ip-172-31-9-52:~$
```



