

생물정보학
AWS 클라우드 컴퓨팅

2020.07.25

한주현

kenneth.jh.han@snu.ac.kr

강사소개

강사 소개	
이름	한주현
소속	3billion, Bioinformatics Engineer Seoul National University, Medical Informatics
메일	kenneth.jh.han@snu.ac.kr
주요 업무	Human Genome Analysis (WGS, WES) Rare Disease Analysis Bioinformatics Algorithms Analysis Pipeline / Platform Development Full Stack Development Cloud Computing
주 언어	Python, JAVA, JavaScript, Bash shell
저서	니콜라스 볼커 이야기 (2016.10, 금창원 외 공역) 바이오파이썬으로 만나는 생물정보학 (2019.03, 한주현)
웹 페이지	https://korbillgates.tistory.com (블로그)

강의 내용 또는 생물정보학, 취업 및 진로에 관련하여
궁금한점이 있으시면 언제든 메일로 문의해주세요

목차

- 강의 소개
- AWS 클라우드 컴퓨팅 소개
- 계정 만들기
- EC2 만들기
- EC2 접속하기 (AWS 콘솔)
- EC2 접속하기 (윈도우)
- EC2 접속하기 (맥 or 리눅스)
- 기본 리눅스 명령어
- EC2에서 파이썬 설치하기
- EC2에서 포트 열어 jupyter notebook 구동하기
- DNA 분석 파이프라인 소개
- EC2에서 생물정보학 툴 설치하기
- EC2 삭제하기
- 계정 삭제하기

강의 소개

- 본 강의는 AWS 클라우드 컴퓨팅 서비스를 만들어보고
클라우드 컴퓨팅 환경에서 작업을 진행하며
생물정보학 현업에서 많이 사용하는 리눅스의 기본적 명령어와
DNA 분석 파이프라인에서 나오는 파일들을 리눅스 명령어로
살펴보고 익혀보는 강의입니다.

AWS 클라우드 컴퓨팅 소개

- AWS (Amazon Web Service) 클라우드 컴퓨팅은 클라우드 컴퓨팅의 대명사라고 할 만큼 높은 점유율의 사용자를 가지고 있습니다.
- 클라우드 컴퓨팅은 서버실에 있는 on-premise 서버와 비교하였을 때 높은 가소성을 가지고 있습니다. 즉, 필요할 때 빌려서 쓰고 필요 없을 때는 반납하여 비용 절감을 할 수 있습니다.

AWS 계정 만들기

계정 만들기

Google aws

All Images News Maps Videos More Settings Tools

About 206,000,000 results (0.64 seconds)

Ad · aws.amazon.com/account/sign-up ▾

Amazon Web Services - Sign Up For a Free AWS Account

Build, Deploy, and Manage Websites, Apps or Processes On AWS' Secure, Reliable Network. Sign Up for a Free Account & Experience AWS' Secure, Reliable, Scalable Services. Durable, Safe & Secure. Performance At Scale. Secure Solutions.

Why Amazon Web Services
Focus on What Differentiates Your Business - Not the Infrastructure.

AWS Cloud Compute Service
AWS Lightsail: Easy-To-Use Platform For Simple Computing. 1 Year Free.

Ad · www.datadoghq.com/aws-monitoring ▾

AWS Monitoring Platform - Modern, Full-Stack Monitoring

Get Faster Insights Into All The AWS Services That Power Your Business. Start Monitoring AWS Performance In Minutes With Dashboards, Alerts, And More. Complete Observability. Clear Visualizations. Single Pane View. Smarter Alerting. Metrics, Traces, & Logs.

Logging Without Limits · Application Monitoring · Network Monitoring · Serverless Monitoring

aws.amazon.com ▾

Amazon Web Services (AWS) - Cloud Computing Services

Amazon Web Services offers reliable, scalable, and inexpensive cloud computing services. Free to join, pay only for what you use.

AWS Cloud
Cloud computing with AWS.
Amazon Web Services (AWS ...)

Amazon EC2
Amazon Elastic Compute Cloud
(Amazon EC2) is a web service ...

Pricing
Amazon EC2 Pricing - Amazon S3
Pricing - Amazon RDS Pricing

Products
Compute - Blockchain on AWS -
Networking - Developer Tools

Amazon Web Services  IT service management company

aws.amazon.com

Amazon Web Services is a subsidiary of Amazon that provides on-demand cloud computing platforms and APIs to individuals, companies, and governments, on a metered pay-as-you-go basis.

[Wikipedia](#)

CEO: Andy Jassy (Apr 2016–)
Headquarters: Seattle, Washington, United States
Founder: Amazon.com
Founded: 2006
Subsidiaries: Amazon Web Services Japan K.K., CloudEndure, [MORE](#)

Parent organization: Amazon.com

[Disclaimer](#)

People also search for [View 10+ more](#)

 Microsoft  IBM  Microsoft Azure  accenture  GE
Microsoft Corporation IBM Microsoft Azure Accenture General Electric

구글에서 aws를 검색하고
링크를 클릭합니다

<https://aws.amazon.com/>

The screenshot shows the AWS homepage with a red box highlighting the 'Create an AWS Account' button in the top right corner. The page features a dark blue header with the AWS logo and navigation links like 'Products', 'Solutions', 'Pricing', etc. A banner at the top reads 'See the AWS Initiatives and Response to COVID-19 ». Below it, a large section titled 'Start Building on AWS Today' includes a 'Get Started for Free' button and three call-to-action cards: 'Sign up for an AWS account', 'Enable remote work & learning', and 'Launch your first application in minutes'.

Explore Our Products



Analytics



Application Integration



AR & VR



AWS Cost Management



Blockchain



Business Applications

Compute



Containers



Customer Engagement



Database



Developer Tools



End User Computing



Game Tech



Internet of Things



Machine Learning

Create an AWS Account
를 클릭합니다.

<https://portal.aws.amazon.com/billing/signup#/start>

Create an AWS account

Email address

* Email is a required field.

Password

Confirm password

AWS account name ⓘ

Continue

[Sign in to an existing AWS account](#)

© 2020 Amazon Web Services, Inc. or its affiliates.
All rights reserved.
[Privacy Policy](#) | [Terms of Use](#)

내용을 채워 넣습니다.

Contact Information

All fields are required.

Please select the account type and complete the fields below with your contact details.

Account type ⓘ
 Professional Personal

Full name

* Full Name is a required field.

Company name

Phone number

Country/Region

Address

City

State / Province or region

Postal code

Check here to indicate that you have read and agree to the terms of the [AWS Customer Agreement](#)

Create Account and Continue

한국인 경우,
Country/Region은
Korea, Republic of
를 선택합니다.

Payment Information

All fields are required.

We use your payment information to verify your identity and only for usage in excess of the [AWS Free Tier Limits](#). We will not charge you for usage below the AWS Free Tier Limits. To learn more about payment options, review our [Frequently Asked Questions](#).

 When you submit your payment information, we will charge \$1 USD/EUR to your credit card as a verification charge to ensure your card is valid. The amount may show as pending in your credit card statement for 3-5 days until the verification is completed, at which time the charge will be removed. You may be redirected to your bank website to authorize the verification charge.

Credit/Debit card number



AWS accepts most major credit and debit cards.

Expiration date

07 2020

Cardholder's name

Billing address

Use my contact address

Seokchonhosu-ro 210, 903ho
Seoul Seoul 05611
KR

Use a new address

Verify and Add

© 2020 Amazon Web Services, Inc. or its affiliates. All rights reserved.

[Privacy Policy](#) | [Terms of Use](#) | [Sign Out](#)

결제 정보를 넣습니다.
우리는 무료 계정을 사용할 것이고
강의가 마치고 나면 계정 삭제할 것이므로
결제 청구 되지 않습니다.

1달러는 확인용(verification) 청구입니다.

Confirm your identity

Before you can use your AWS account, you must verify your phone number. When you continue, the AWS automated system will contact you with a verification code.

How should we send you the verification code?

Text message (SMS) Voice call

Country or region code

Korea, Republic of (+82)

Cell Phone Number

Security check



Type the characters as shown above

Send SMS

© 2020 Amazon Web Services, Inc. or its affiliates. All rights reserved.

[Privacy Policy](#) | [Terms of Use](#) | [Sign Out](#)

Enter verification code

Enter the 4-digit verification code that you received on your phone.

Verify Code

Having trouble? Sometimes it takes up to 10 minutes to receive a verification code. If it's been longer than that, [return to the previous page](#) and enter your number again.



Your identity has been verified successfully.

Continue

Select a Support Plan

AWS offers a selection of support plans to meet your needs. Choose the support plan that best aligns with your AWS usage. [Learn more](#)



Basic Plan

Free

Developer Plan

From \$29/month

Business Plan

From \$100/month

- Included with all accounts
 - 24x7 self-service access to AWS resources
 - For account and billing issues only
 - Access to Personal Health Dashboard & Trusted Advisor
- For early adoption, testing and development
 - Email access to AWS Support during business hours
 - 1 primary contact can open an unlimited number of support cases
 - 12-hour response time for nonproduction systems
 - For production workloads & business-critical dependencies
 - 24/7 chat, phone, and email access to AWS Support
 - Unlimited contacts can open an unlimited number of support cases
 - 1-hour response time for production systems

Need Enterprise level support?

Contact your account manager for additional information on running business and mission critical-workloads on AWS (starting at \$15,000/month). [Learn more](#)

© 2020 Amazon Web Services, Inc. or its affiliates. All rights reserved.

[Privacy Policy](#) | [Terms of Use](#) | [Sign Out](#)

Free 를 누릅니다.

Welcome to Amazon Web Services

Thank you for creating an Amazon Web Services Account. We are activating your account, which should only take a few minutes. You will receive an email when this is complete.

Sign In to the Console

Contact Sales

Personalize Your Experience

Fill in the blanks below to receive recommendations catered to your role and interests.

My role is: [select role](#)

I am interested in: [select area](#)

Submit

Try a Tutorial on the Free Tier

COMPUTE



Launch a Linux Virtual Machine



Deploy Docker Containers



Run a Serverless "Hello, World!"

Sign In to the Console
을 누릅니다.



Sign in

Root user

Account owner that performs tasks requiring unrestricted access. [Learn more](#)

IAM user

User within an account that performs daily tasks.
[Learn more](#)

Root user email address

username@example.com

Next

New to AWS?

Create a new AWS account

가입한 메일 주소를 넣습니다.

Password

[Forgot password?](#)

|

Sign in

[Sign in to a different account](#)

[Create a new AWS account](#)

패스워드를 넣습니다.

AWS Management Console

AWS services

Find Services
You can enter names, keywords or acronyms.
 Example: Relational Database Service, database, RDS

▶ All services

Build a solution
Get started with simple wizards and automated workflows.

Launch a virtual machine With EC2 2-3 minutes 	Build a web app With Elastic Beanstalk 6 minutes 	Build using virtual servers With Lightsail 1-2 minutes 	Register a domain With Route 53 3 minutes 
Connect an IoT device With AWS IoT 5 minutes 	Start migrating to AWS With CloudEndure Migration 1-2 minutes 	Start a development project With CodeStar 5 minutes 	Deploy a serverless microservice With Lambda, API Gateway 2 minutes 

▶ See more

Stay connected to your AWS resources on-the-go
Download the AWS Console Mobile App to your iOS or Android mobile device. [Learn more](#)

Explore AWS

Amazon SageMaker Resources
Learn about SageMaker's features, use cases, and available workshops. [Learn more](#)

Free Digital Training
Get access to 350+ self-paced online courses covering AWS products and services. [Learn more](#)

Amazon EFS for Containers
Simplify file storage for modern applications, securely persist and share data with zero management required. [Learn more](#)

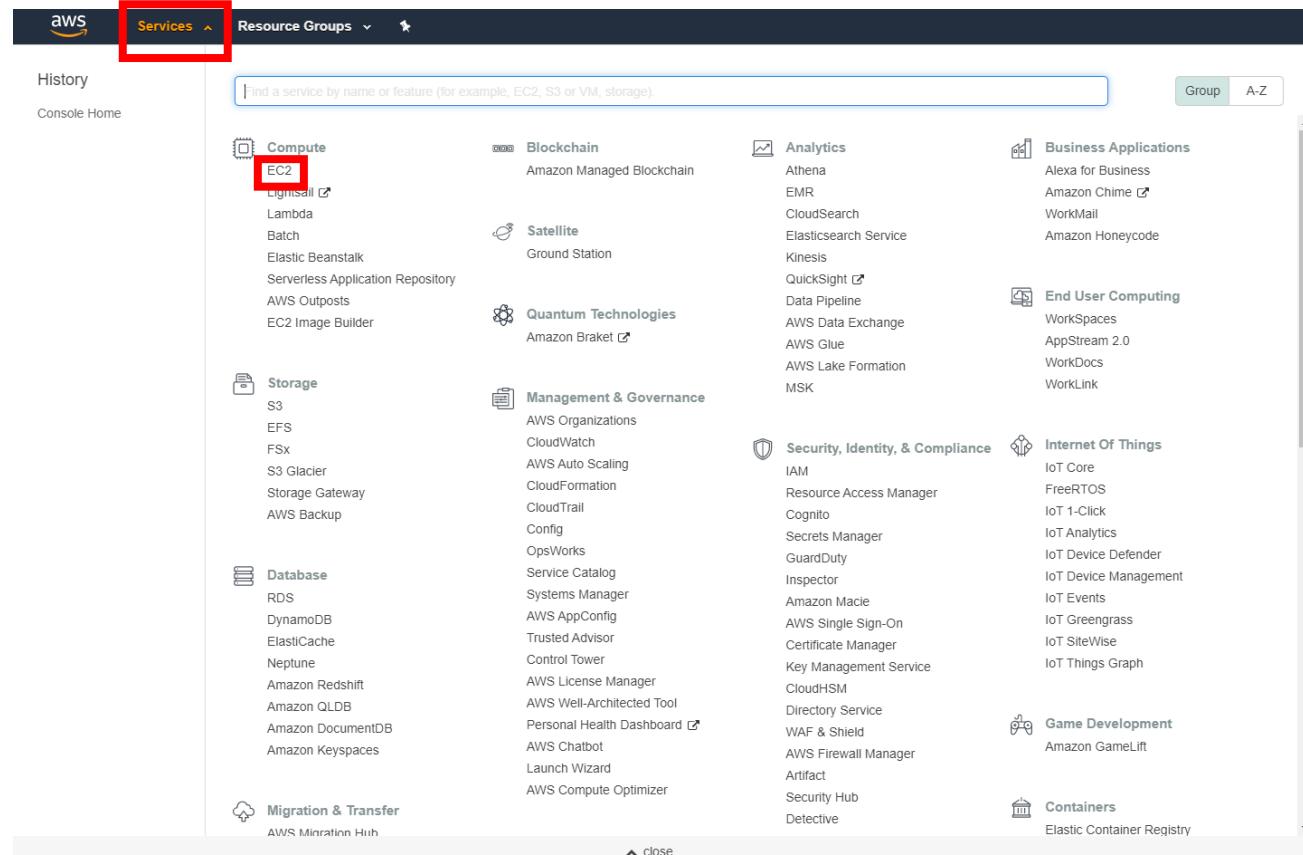
Amazon DocumentDB (with MongoDB compatibility)
New role-based access control support helps you enforce least privilege access and build multi-tenant applications. [Get started](#)

Have feedback?

가입 완료!

AWS EC2 만들기

EC2 만들기



상단의 Services 를 누르면 AWS 가 제공하는 여러 서비스들이 나옵니다.

우리는 여기서 EC2를 살펴보겠습니다.

The screenshot shows the AWS EC2 Instances page. On the left sidebar, under the 'Instances' section, the 'Instances' link is highlighted with a red box. The main content area displays a welcome message about the new EC2 console, resource statistics, and a 'Launch instance' button. The 'Service health' and 'Zone status' sections show that the service is operating normally across three zones.

Welcome to the new EC2 console!
We're redesigning the EC2 console to make it easier to use and improve performance. We'll release new screens periodically. We encourage you to try them and let us know where we can make improvements. To switch between the old console and the new console, use the New EC2 Experience toggle.

Resources

You are using the following Amazon EC2 resources in the US East (Ohio) Region:

Running instances	0	Elastic IPs	0	Dedicated Hosts	0
Snapshots	0	Volumes	0	Load balancers	0
Key pairs	0	Security groups	1	Placement groups	0

Easily size, configure, and deploy Microsoft SQL Server Always On availability groups on AWS using the AWS Launch Wizard for SQL Server. [Learn more](#)

Launch instance

To get started, launch an Amazon EC2 instance, which is a virtual server in the cloud.

[Launch instance](#)

Note: Your instances will launch in the US East (Ohio) Region

Scheduled events

US East (Ohio)
No scheduled events

Service health

Region	Status
US East (Ohio)	This service is operating normally

Zone status

Zone	Status
us-east-2a (use2-az1)	Zone is operating normally
us-east-2b (use2-az2)	Zone is operating normally
us-east-2c (use2-az3)	Zone is operating normally

[Enable additional zones](#)

Account attributes

Supported platforms [\[?\]](#)

- VPC

Default VPC [\[?\]](#)
vpc-297bdd42

Settings

EBS encryption

Zones

Default credit specification

Console experiments

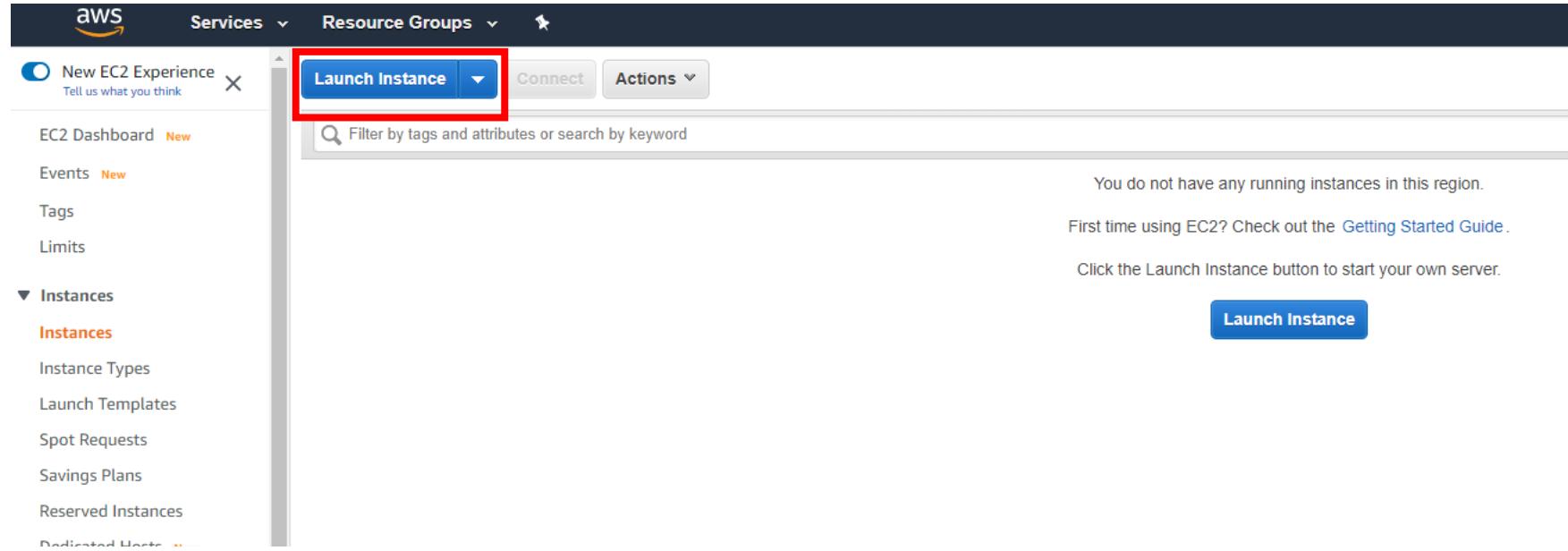
Explore AWS

Easily launch third-party AMI products
AWS Marketplace has thousands of third-party AMI products that you can find, buy, and deploy with 1-click using the Amazon EC2 console. [Learn more](#)

Save up to 90% on EC2 with Spot Instances
Optimize price-performance by combining EC2 purchase options in a single EC2 ASG. [Learn more](#)

Enable Best Price-Performance with AWS Graviton2
AWS Graviton2 powered EC2 instances enable up to 40% better price performance for a broad spectrum of cloud workloads. [Learn more](#)

Instances 를 클릭합니다.



Launch Instance 를 클릭합니다.

Step 1: Choose an Amazon Machine Image (AMI)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

Search for an AMI by entering a search term e.g. "Windows"

Search by Systems Manager parameter

Quick Start

My AMIs

AWS Marketplace

Community AMIs

Free tier only (i)



Amazon Linux 2 AMI (HVM), SSD Volume Type - ami-016b213e65284e9c9 (64-bit x86) / ami-0e067567dbf210b67 (64-bit Arm)

Amazon Linux 2 comes with five years support. It provides Linux kernel 4.14 tuned for optimal performance on Amazon EC2, systemd 219, GCC 7.3, Glibc 2.26, Binutils 2.29.1, and the latest software packages through extras.

Select

64-bit (x86)
 64-bit (Arm)



Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type - ami-0bdcc6c05dec346bf

The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages.

Select

64-bit (x86)



Red Hat Enterprise Linux 8 (HVM), SSD Volume Type - ami-0a54aeef4ef3b5f881 (64-bit x86) / ami-0ffd59b53e6797671 (64-bit Arm)

Red Hat Enterprise Linux version 8 (HVM), EBS General Purpose (SSD) Volume Type

Select

64-bit (x86)
 64-bit (Arm)



SUSE Linux Enterprise Server 15 SP1 (HVM), SSD Volume Type - ami-054462c52d9c225ce (64-bit x86) / ami-0c17b20c5844a1acb (64-bit Arm)

SUSE Linux Enterprise Server 15 Service Pack 1 (HVM), EBS General Purpose (SSD) Volume Type. Public Cloud, Advanced Systems Management, Web and Scripting, and Legacy modules enabled.

Select

64-bit (x86)
 64-bit (Arm)



Ubuntu Server 18.04 LTS (HVM), SSD Volume Type - ami-0a63f96e85105c6d3 (64-bit x86) / ami-0f330b426f35ac7f4 (64-bit Arm)

Ubuntu Server 18.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>).

Select

64-bit (x86)
 64-bit (Arm)

Amazon Linux, Red Hat, Ubuntu 등의 여러 리눅스들이 있습니다.

우리는 무료 버전인 Ubuntu Server 18.04 LTS (HVM), SSD Volume Type 을 선택하겠습니다.

Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. [Learn more](#) about instance types and how they can meet your computing needs.

Filter by: All instance types ▾ Current generation ▾ Show/Hide Columns

Currently selected: t2.micro (Variable ECUs, 1 vCPUs, 2.5 GHz, Intel Xeon Family, 1 GiB memory, EBS only)

	Family	Type	vCPUs <small>i</small>	Memory (GiB)	Instance Storage (GB) <small>i</small>	EBS-Optimized Available <small>i</small>	Network Performance <small>i</small>	IPv6 Support <small>i</small>
<input type="checkbox"/>	General purpose	t2.nano	1	0.5	EBS only	-	Low to Moderate	Yes
<input checked="" type="checkbox"/>	General purpose	t2.micro <small>Free tier eligible</small>	1	1	EBS only	-	Low to Moderate	Yes
<input type="checkbox"/>	General purpose	t2.small	1	2	EBS only	-	Low to Moderate	Yes
<input type="checkbox"/>	General purpose	t2.medium	2	4	EBS only	-	Low to Moderate	Yes
<input type="checkbox"/>	General purpose	t2.large	2	8	EBS only	-	Low to Moderate	Yes
<input type="checkbox"/>	General purpose	t2.xlarge	4	16	EBS only	-	Moderate	Yes
<input type="checkbox"/>	General purpose	t2.2xlarge	8	32	EBS only	-	Moderate	Yes
<input type="checkbox"/>	General purpose	t3a.nano	2	0.5	EBS only	Yes	Up to 5 Gigabit	Yes
<input type="checkbox"/>	General purpose	t3a.micro	2	1	EBS only	Yes	Up to 5 Gigabit	Yes
<input type="checkbox"/>	General purpose	t3a.small	2	2	EBS only	Yes	Up to 5 Gigabit	Yes
<input type="checkbox"/>	General purpose	t3a.medium	2	4	EBS only	Yes	Up to 5 Gigabit	Yes
<input type="checkbox"/>	General purpose	t3a.large	2	8	EBS only	Yes	Up to 5 Gigabit	Yes
<input type="checkbox"/>	General purpose	t3a.xlarge	4	16	EBS only	Yes	Up to 5 Gigabit	Yes

Cancel

Previous

Review and Launch

Next: Configure Instance Details

Instance 유형을 고르는 창입니다.
 무료 버전이 가능한 t2.micro 를 고르고
 Next: Configure Instance Details 를 클릭합니다.

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

Number of instances Launch into Auto Scaling Group [\(i\)](#)

Purchasing option [\(i\)](#) Request Spot instances

Network [\(i\)](#) vpc-297bdd42 (default) [\(C\)](#) Create new VPC

Subnet [\(i\)](#) No preference (default subnet in any Availability Zone) [\(C\)](#) Create new subnet

Auto-assign Public IP [\(i\)](#) Use subnet setting (Enable) [\(C\)](#)

Placement group [\(i\)](#) Add instance to placement group

Capacity Reservation [\(i\)](#) Open [\(C\)](#) Create new Capacity Reservation

IAM role [\(i\)](#) None [\(C\)](#) Create new IAM role

Shutdown behavior [\(i\)](#) Stop [\(C\)](#)

Stop - Hibernate behavior [\(i\)](#) Enable hibernation as an additional stop behavior

Enable termination protection [\(i\)](#) Protect against accidental termination

Monitoring [\(i\)](#) Enable CloudWatch detailed monitoring
Additional charges apply.

Tenancy [\(i\)](#) Shared - Run a shared hardware instance [\(C\)](#)
Additional charges may apply when launching Dedicated instances.

Elastic Inference [\(i\)](#) Add an Elastic Inference accelerator
Additional charges apply.

T2/T3 Unlimited [\(i\)](#) Enable
Additional charges may apply

File systems [\(i\)](#) [Add file system](#) [\(C\)](#) Create new file system

[Cancel](#)

[Previous](#)

[Review and Launch](#)

[Next: Add Storage](#)

건드리는 것 없이 기본값으로 둡니다.
Next: Add Storage 를 클릭합니다.

Step 4: Add Storage

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. [Learn more](#) about storage options in Amazon EC2.

Volume Type <i>i</i>	Device <i>i</i>	Snapshot <i>i</i>	Size (GiB) <i>i</i>	Volume Type <i>i</i>	IOPS <i>i</i>	Throughput (MB/s) <i>i</i>	Delete on Termination <i>i</i>	Encryption <i>i</i>
Root	/dev/sda1	snap-021b833c41d050331	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input checked="" type="checkbox"/>	Not Encrypted

[Add New Volume](#)

Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage. [Learn more](#) about free usage tier eligibility and usage restrictions.

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Add Tags](#)

건드리는 것 없이 기본값으로 둡니다.
Next: Add Tags 를 클릭합니다.

Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver.

A copy of a tag can be applied to volumes, instances or both.

Tags will be applied to all instances and volumes. [Learn more](#) about tagging your Amazon EC2 resources.

Key (128 characters maximum)	Value (256 characters maximum)	Instances <small>i</small>	Volumes <small>i</small>
<p><i>This resource currently has no tags</i></p> <p>Choose the Add tag button or click to add a Name tag. Make sure your IAM policy includes permissions to create tags.</p>			
Add Tag (Up to 50 tags maximum)			

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Configure Security Group](#)

건드리는 것 없이 기본값으로 둡니다.

Next: Configure Security Group 을 클릭합니다.

Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. [Learn more](#) about Amazon EC2 security groups.

Assign a security group: Create a new security group

Select an existing security group

Security group name: launch-wizard-1

Description: launch-wizard-1 created 2020-07-23T21:35:35.706+09:00

Type	Protocol	Port Range	Source	Description
SSH	TCP	22	Custom 0.0.0.0/0	e.g. SSH for Admin Desktop

Add Rule



Warning

Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

[Cancel](#) [Previous](#) [Review and Launch](#)

건드리는 것 없이 기본값으로 둡니다.
Review and Launch 를 클릭합니다.

Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

⚠ Improve your instances' security. Your security group, launch-wizard-1, is open to the world.

Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only.

You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. [Edit security groups](#)

AMI Details

[Edit AMI](#)

 **Ubuntu Server 18.04 LTS (HVM), SSD Volume Type - ami-0a63f96e85105c6d3**

Free tier eligible Ubuntu Server 18.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>).
Root Device Type: ebs Virtualization type: hvm

Instance Type

[Edit instance type](#)

Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
t2.micro	Variable	1	1	EBS only	-	Low to Moderate

Security Groups

[Edit security groups](#)

Security group name launch-wizard-1

Description launch-wizard-1 created 2020-07-23T21:35:35.706+09:00

Type <small>i</small>	Protocol <small>i</small>	Port Range <small>i</small>	Source <small>i</small>	Description <small>i</small>
SSH	TCP	22	0.0.0.0/0	

Instance Details

[Edit instance details](#)

Number of instances 1

Network vpc-297bdd42

Subnet No preference (default subnet in any Availability Zone)

EBS-optimized No

Purchasing option On demand

[Cancel](#)
[Previous](#)
[Launch](#)

건드리는 것 없이 기본값으로 둡니다.
Launch 를 클릭합니다.

Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

⚠ Improve your instances' security. Your security group, launch-wizard-1, is open to the world.

Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only.

You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. [Edit security groups](#)

AMI Details

Ubuntu Server 18.04 LTS (HVM), SSD Volume Type - ami-0a63f96e8510

Free tier eligible

Ubuntu Server 18.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support a

Root Device Type: ebs Virtualization type: hvm

Instance Type

Instance Type	ECUs	vCPUs	Memory (GiB)	Insta
t2.micro	Variable	1	1	EBS

Security Groups

Security group name: launch-wizard-1
Description: launch-wizard-1 created 2020-07-23T21:35:35.706+09:00

Type	Protocol	Port Range
SSH	TCP	22

Instance Details

Number of instances: 1
Network: vpc-297bdd42
Subnet: No subnet (default subnet in your Availability Zone)
EBS-optimized: No

Purchasing option: On demand

빨간 네모쪽을 눌러서 Create a new key pair를 선택하고,
아래 Key pair name에서 자신이 원하는 key 이름을 씁니다.
저는 bioinfo-lecture-key 라고 쓰겠습니다.
그리고 Download Key Pair를 눌러서 key를 받습니다.
key는 한 번만 다운로드 가능하며 분실하면 다시 받을 수 없으니 주의하세요.
key를 받으면 Launch Instances 버튼이 활성화 됩니다. 클릭해주세요.

Select an existing key pair or create a new key pair

A key pair consists of a **public key** that AWS stores, and a **private key file** that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about [removing existing key pairs from a public AMI](#).

Create a new key pair

Key pair name:

bioinfo-lecture-key

Download Key Pair

You have to download the **private key file** (*.pem file) before you can continue. **Store it in a secure and accessible location.** You will not be able to download the file again after it's created.

Cancel

Launch Instances

File name: bioinfo-lecture-key
Save as type: PEM File

Hide Folders

Save Cancel

Edit security groups

Edit instance details

Cancel Previous Launch

Launch Status

Your instances are now launching

The following instance launches have been initiated: i-09b27bf28513f2b10 [View launch log](#)

Get notified of estimated charges

Create [billing alerts](#) to get an email notification when estimated charges on your AWS bill exceed an amount you define (for example, if you exceed the free usage tier).

How to connect to your instances

Your instances are launching, and it may take a few minutes until they are in the **running** state, when they will be ready for you to use. Usage hours on your new instances will start immediately and continue to accrue until you stop or terminate your instances.

Click [View Instances](#) to monitor your instances' status. Once your instances are in the **running** state, you can **connect** to them from the Instances screen. [Find out](#) how to connect to your instances.

▼ Here are some helpful resources to get you started

- [How to connect to your Linux instance](#)
- [Learn about AWS Free Usage Tier](#)
- [Amazon EC2: User Guide](#)
- [Amazon EC2: Discussion Forum](#)

While your instances are launching you can also

- [Create status check alarms](#) to be notified when these instances fail status checks. (Additional charges may apply)
- [Create and attach additional EBS volumes](#) (Additional charges may apply)
- [Manage security groups](#)

[View Instances](#)

View Instances 를 클릭합니다.

aws Services Resource Groups

New EC2 Experience Tell us what you think

EC2 Dashboard New

Events New

Tags

Limits

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Launch Instance Connect Actions

Filter by tags and attributes or search by keyword

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)	IPv4 Public IP	IPv6 IPs	Key Name	Monitoring	Last
i-██████████	t2.micro	us-east-2b	running	Initializing	None	ec2-3-██████████.ec2.us-east-2.amazonaws.com	3.██████████	-	-	bioinfo-lecture-...	disabled	July

K < 1 to 1 of 1 > |

A screenshot of the AWS EC2 Instances page. The top navigation bar includes the AWS logo, services dropdown, resource groups, and Ohio region. A 'New EC2 Experience' feedback button is present. The left sidebar shows navigation links for EC2 Dashboard, Events, Tags, Limits, Instances (selected), and various EC2-specific options like Instance Types, Launch Templates, and Savings Plans. The main content area displays a table of instances. One instance is listed: it has a Name of 'i-██████████', an Instance ID of 'i-██████████', an Instance Type of 't2.micro', an Availability Zone of 'us-east-2b', and an Instance State of 'running'. The Status Checks are 'Initializing' and the Alarm Status is 'None'. The Public DNS (IPv4) is 'ec2-3-██████████.ec2.us-east-2.amazonaws.com' and the IPv4 Public IP is '3.██████████'. The Key Name is 'bioinfo-lecture-' and Monitoring is disabled. The last update was in July. A search bar at the top allows filtering by tags and attributes or searching by keyword.

EC2 생성 완료!

AWS EC2 접속하기

EC2 접속하기 (AWS 콘솔)

The screenshot shows the AWS EC2 Instances page. At the top, there are three buttons: "Launch Instance" (blue), "Connect" (gray), and "Actions" (gray dropdown). Below these are search and filter fields. The main table lists an instance: Name (redacted), Instance ID (i-**REDACTED**), Instance Type (t2.micro), Availability Zone (us-east-2b), Instance State (running), Status Checks (2/2 checks ...), Alarm Status (None), and Public DNS (IPv4) (redacted). To the right of the instance details is a context menu with a "Connect" option highlighted by a red box. A modal window titled "Connect to your instance" is open, showing three connection methods: "A standalone SSH client" (radio button), "Session Manager" (radio button), and "EC2 Instance Connect (browser-based SSH connection)" (radio button, also highlighted by a red box). The "EC2 Instance Connect" option is selected. Below the modal, a note says "Connect using a custom user name, or default to the user name for the AMI used to launch the instance." and a "Learn more" link. At the bottom of the modal are "User name" (ubuntu) and "Connect" (blue) buttons, with the "Connect" button also highlighted by a red box.

Launch Instance ▾ Connect Actions ▾

Filter by tags and attributes or search by keyword

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)
i- REDACTED	t2.micro	us-east-2b	running	2/2 checks ...	None	Connect	Get Windows Password Create Template From Instance Launch More Like This

Connect to your instance

Connection method

- A standalone SSH client ⓘ
- Session Manager ⓘ
- EC2 Instance Connect (browser-based SSH connection) ⓘ

Connect using a custom user name, or default to the user name for the AMI used to launch the instance.
[Learn more](#)

User name: ubuntu ⓘ

Close Connect

EC2 접속하기 (윈도우)

<https://www.putty.org/>



Download PuTTY

PuTTY is an SSH and telnet client, developed from source code and is developed and supported by volunteers.

You can download PuTTY [here](#).

Below suggestions are independent of the aut

Download PuTTY: latest release (0.74)

[Home](#) | [FAQ](#) | [Feedback](#) | [Licence](#) | [Updates](#) | [Mirrors](#) | [Keys](#) | [Links](#) | [Team](#)
Download: [Stable](#) · [Snapshot](#) | [Docs](#) | [Changes](#) | [Wishlist](#)

This page contains download links for the latest released version of PuTTY. Currently this is 0.74, released on 2020-06-27.

When new releases come out, this page will update to contain the latest, so this is a good page to bookmark or link to. Alternatively, here is a [permanent link to the 0.74 release](#).

Release versions of PuTTY are versions we think are reasonably likely to work well. However, they are often not the most up-to-date version of the code available. If you have a problem with one of these, it may already have been fixed in those versions.

Package files

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

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

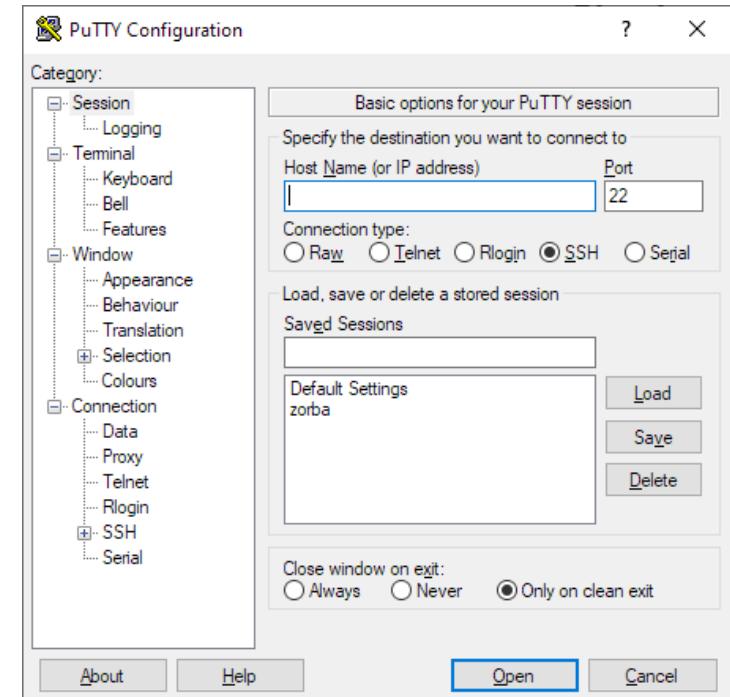
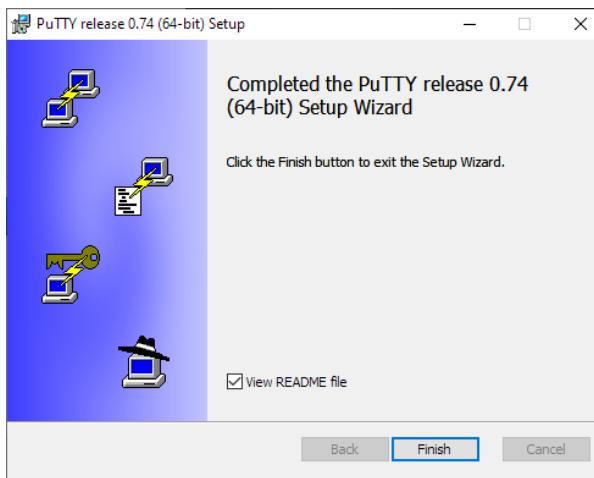
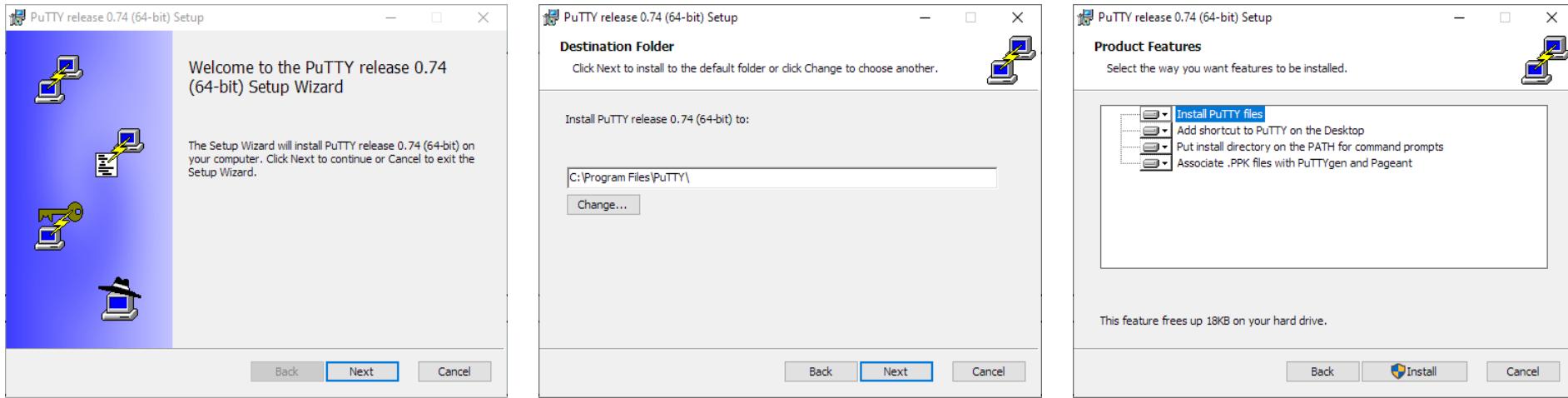
MSI ('Windows Installer')

32-bit:	putty-0.74-installer.msi	(or by FTP)	(signature)
64-bit:	putty-64bit-0.74-installer.msi	(or by FTP)	(signature)

Unix source archive

.tar.gz:	putty-0.74.tar.gz	(or by FTP)	(signature)
----------	-----------------------------------	------------------------------	-------------------------------

ssh접속 도구인
putty를 설치합니다.



ssh접속 도구인
putty를 설치합니다.

pem 키를 ppk 키로 변환하기].

<https://www.chiark.greenend.org.uk/~sgtatham/putty/>



PuTTY: a free SSH and Telnet client

[Home](#) | [FAQ](#) | [Feedback](#) | [Licence](#) | [Updates](#) | [Mirrors](#) | [Keys](#) | [Links](#) | [Team](#)
Download: [Stable](#) · [Snapshot](#) | [Docs](#) | [Changes](#) | [Wishlist](#)

PuTTY is a free implementation of SSH and Telnet for Windows and Unix platforms, along with an xterm terminal emulator. It is written and maintained primarily by [Simon Tatham](#).

The latest version is 0.74. [Download it here](#).

LEGAL WARNING: Use of PuTTY, PSCP, PSFTP and Plink is illegal in countries where encryption is outlawed. We believe it is legal to use PuTTY, PSCP, PSFTP and Plink in England advice before downloading it. You may find useful information at [cryptolaw.org](#), which collects information on cryptography laws in many countries, but we can't vouch for its correctness.

Download PuTTY: latest release (0.74)

[Home](#) | [FAQ](#) | [Feedback](#) | [Licence](#) | [Updates](#) | [Mirrors](#) | [Keys](#) | [Links](#) | [Team](#)
Download: [Stable](#) · [Snapshot](#) | [Docs](#) | [Changes](#) | [Wishlist](#)

This page contains download links for the latest released version of PuTTY. Currently this is 0.74, released on 2020-06-27.

When new releases come out, this page will update to contain the latest, so this is a good page to bookmark or link to. Alternatively, here is a [permanent link to the 0.74 release](#).

Release versions of PuTTY are versions we think are reasonably likely to work well. However, they are often not the most up-to-date version of the code available. If you have a problem with already been fixed in those versions.

Package files

You probably want one of these. They include versions of 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-0.74-installer.msi	(or by FTP)	(signature)
64-bit:	putty-64bit-0.74-installer.msi	(or by FTP)	(signature)

Unix source archive

.tar.gz:	putty-0.74.tar.gz	(or by FTP)	(signature)
----------	-----------------------------------	-------------	-------------

chiark.greenend.org.uk/~sgtatham/putty/latest.html

32-bit: [pageant.exe](#) (or by FTP) (signature)
64-bit: [pageant.exe](#) (or by FTP) (signature)

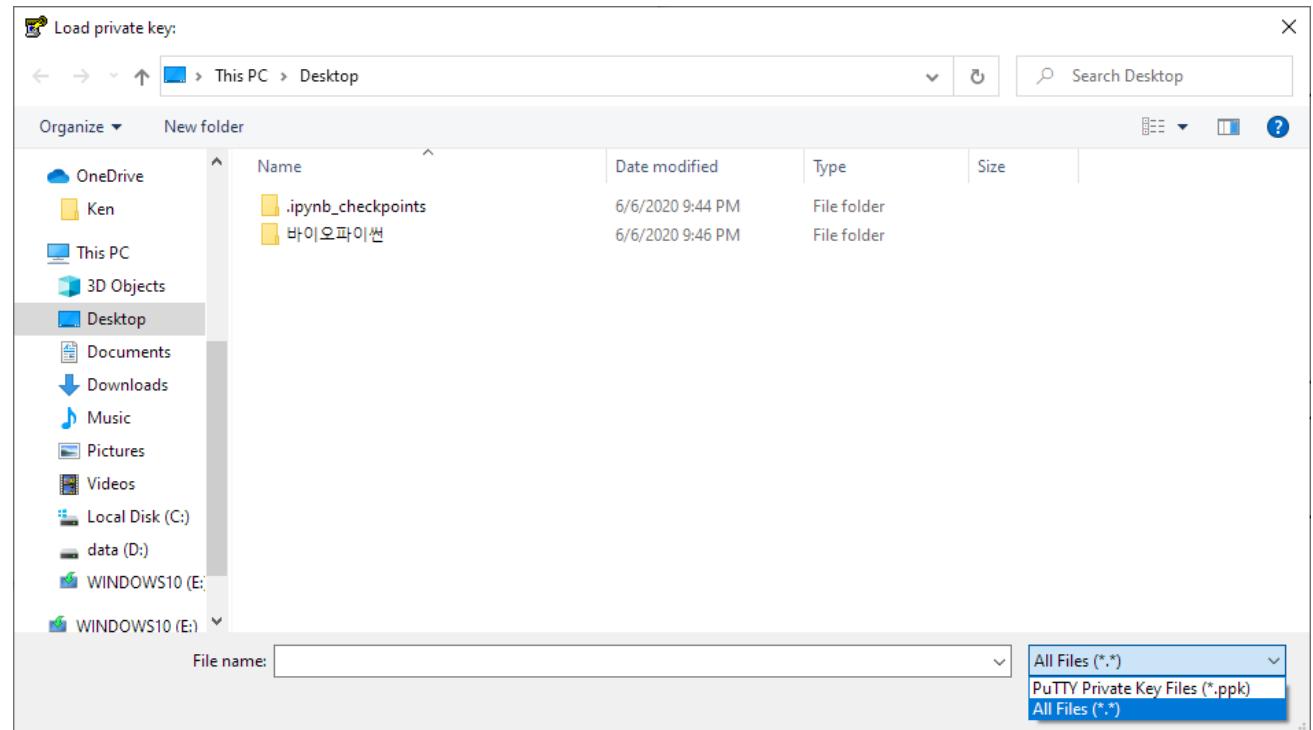
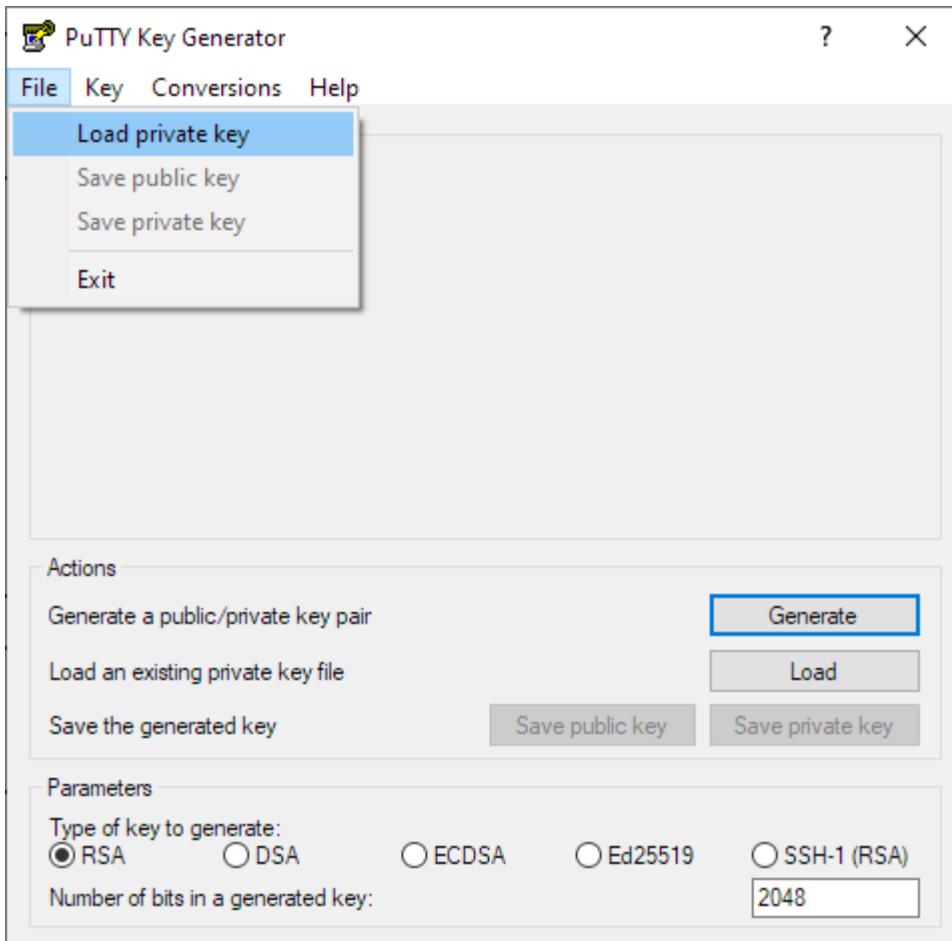
puttygen.exe (a RSA and DSA key generation utility)

32-bit:	puttygen.exe	(or by FTP)	(signature)
64-bit:	puttygen.exe	(or by FTP)	(signature)

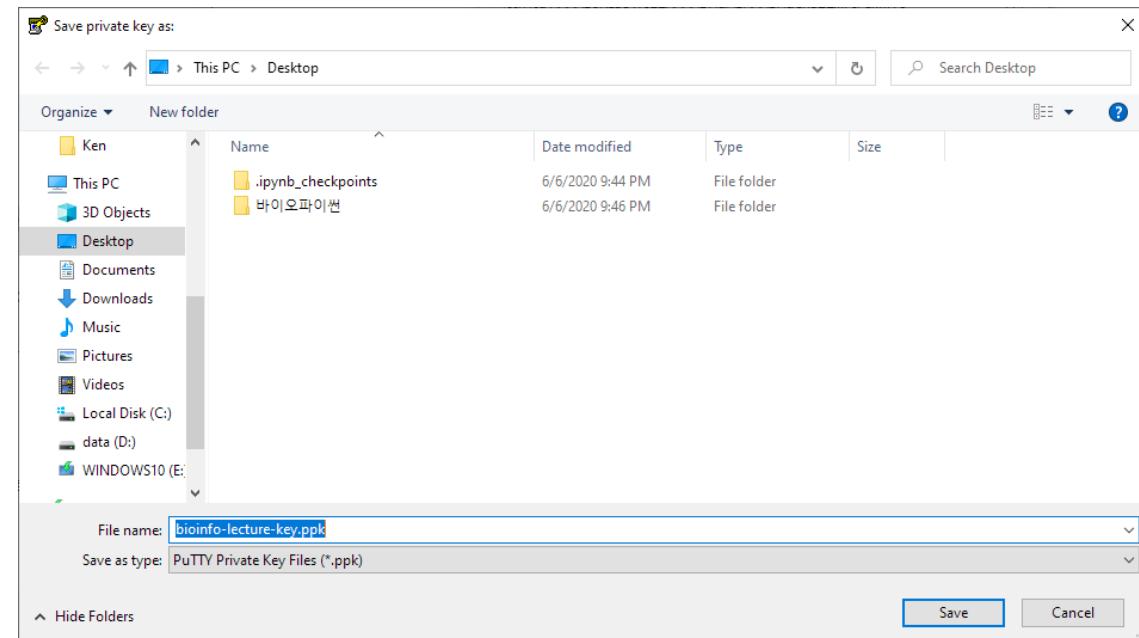
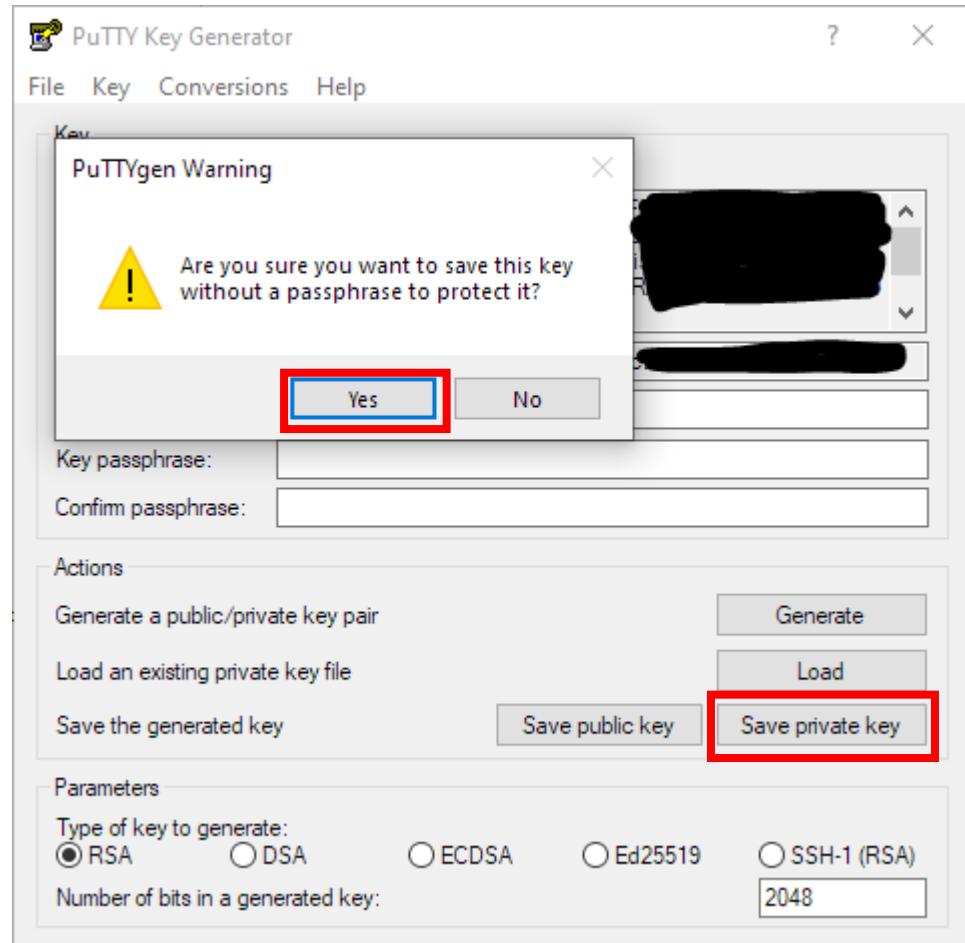
putty.zip (a .ZIP archive of all the above)

32-bit:	putty.zip	(or by FTP)	(signature)
64-bit:	putty.zip	(or by FTP)	(signature)

puttygen.exe 파일을 받습니다.



File - Load private key 를 클릭하여
EC2를 생성하면서 만든 pem key 를 불러옵니다.



Launch Instance		Connect	Actions
<input type="text"/> Filter by tags and attributes or search by keyword			
	Name	Instance ID	Instance Type
<input type="checkbox"/>	i-09b27bf28513f2b10	t2.micro	us-east-2b

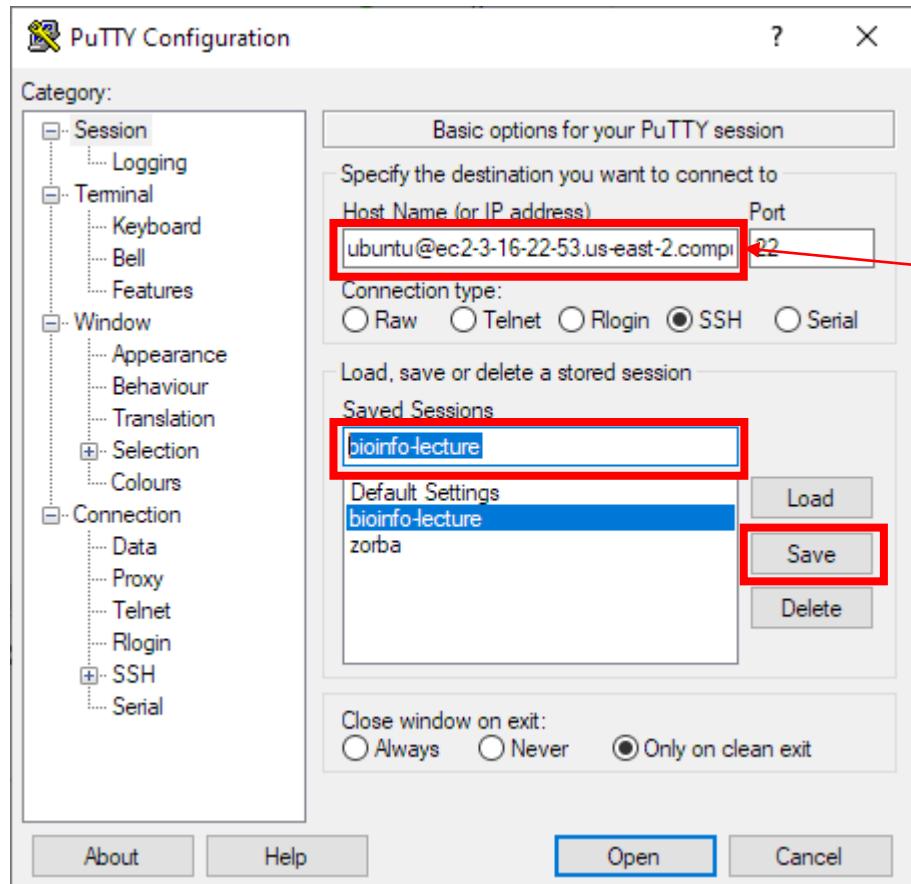
Instance: i-09b27bf28513f2b10 Public DNS: ec2-3-16-22-53.us-east-2.compute.amazonaws.com

Description Status Checks Monitoring Tags

Instance ID: i-09b27bf28513f2b10
Instance state: running
Instance type: t2.micro
Finding: Opt-in to AWS Compute Optimizer for recommendations. [Learn more](#)

Public DNS (IPv4): **ec2-3-16-22-53.us-east-2.compute.amazonaws.com**
IPv4 Public IP: **3.16.22.53**
IPv6 IPs: -
Elastic IPs: -

Public DNS (IPv4) 부분을 복사한다.



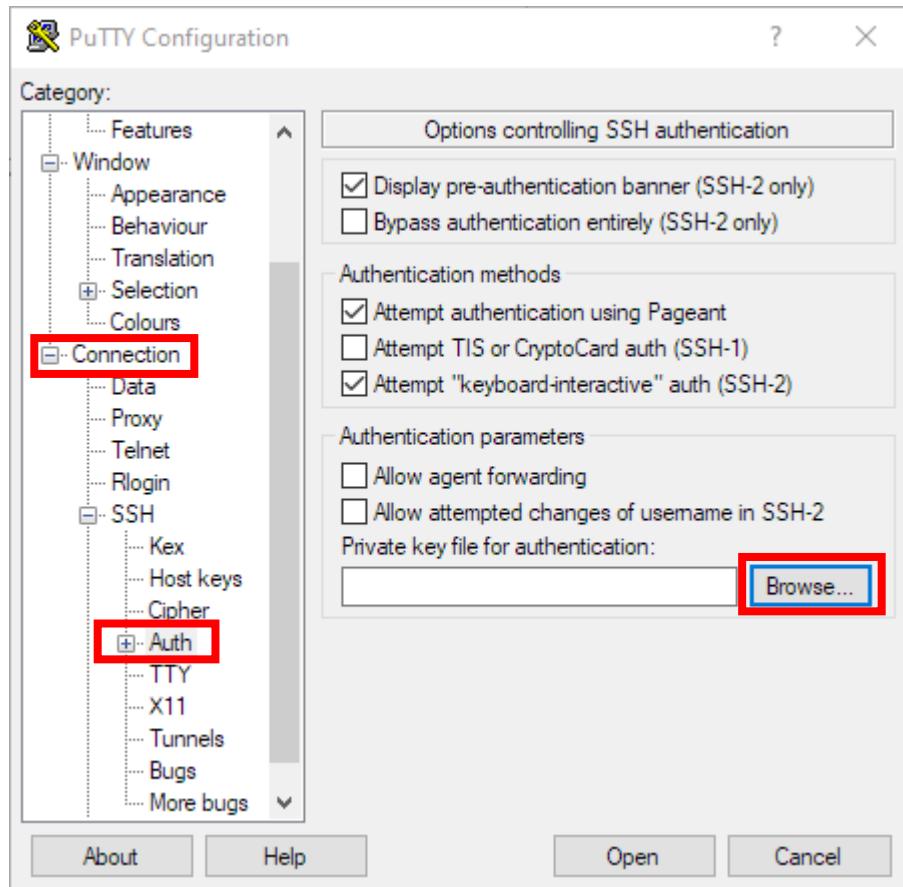
이곳 빨간 네모 부분에

ubuntu@

이라 쓰고

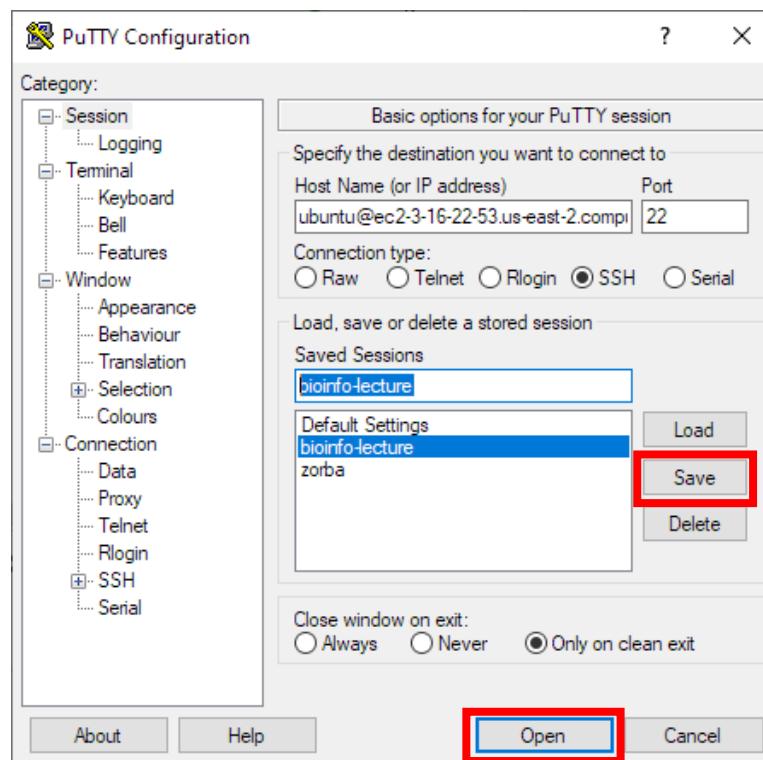
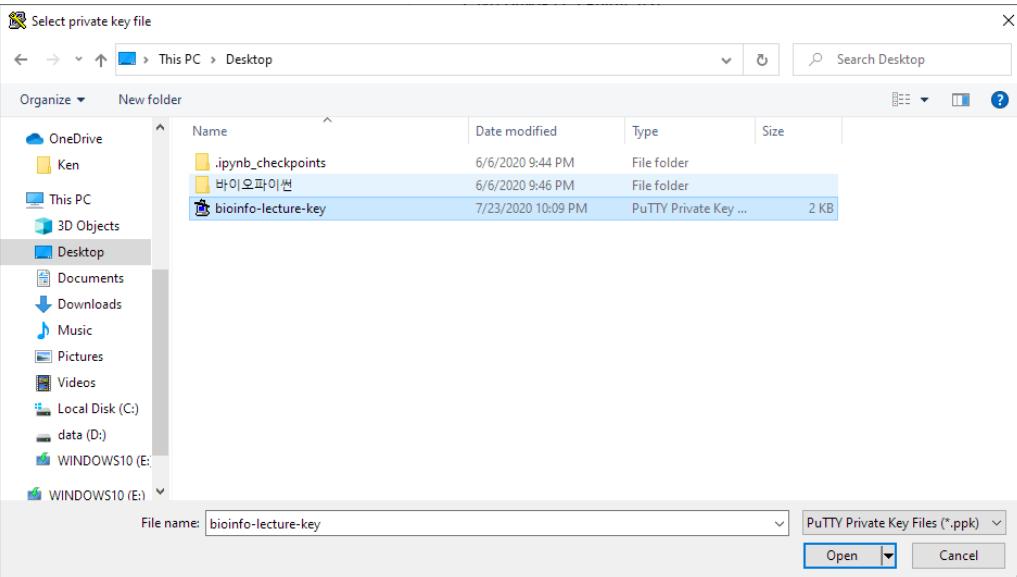
복사한 Public DNS (IPv4) 값을 뒤에 붙여 넣는다.

이후 Saved Sessions 아래에
bioinfo-lecture 라고 쓰고
Save 버튼을 누른다.

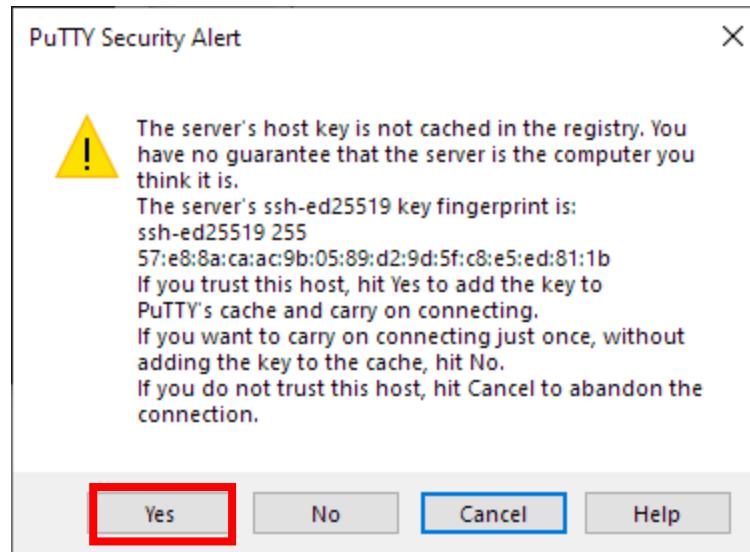


Connection 아래 Auth 를 누르고
Browse 를 누른다.

아까 만든 ppk 파일을 찾아서 넣어준다.



다시 Session 으로
돌아가서 Save 를
누르고
Open 을 누른다.



처음 접속할 때 나오는
모습으로 Yes 를 누릅니다.

AWS EC2 인스턴스에 접속
성공하였습니다!

```
ubuntu@ip-172-31-30-114: ~
[1] Using username "ubuntu".
[2] Authenticating with public key "imported-openssh-key"
Welcome to Ubuntu 18.04.4 LTS (GNU/Linux 5.3.0-1023-aws x86_64)

 * Documentation: https://help.ubuntu.com
 * Management: https://landscape.canonical.com
 * Support: https://ubuntu.com/advantage

System information as of Thu Jul 23 13:12:21 UTC 2020

System load: 0.0 Processes: 90
Usage of /: 13.8% of 7.69GB Users logged in: 0
Memory usage: 16% IP address for eth0: 172.31.30.114
Swap usage: 0%

0 packages can be updated.
0 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.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@ip-172-31-30-114:~$
```

EC2 접속하기 (맥 or 리눅스)

The screenshot shows a terminal window titled "ubuntu@ip-172-31-30-114: ~". The session has been running for 18 seconds at 22:44:03. The terminal displays the following commands and output:

```
ubuntu@ip-172-31-30-114: ~
> chmod 400 bioinfo-lecture-key.pem
> ssh -i "bioinfo-lecture-key.pem" ubuntu@ec2-3-16-22-53.us-east-2.compute.amazonaws.com
Welcome to Ubuntu 18.04.4 LTS (GNU/Linux 5.3.0-1023-aws x86_64)

 * Documentation: https://help.ubuntu.com
 * Management:   https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Thu Jul 23 13:44:08 UTC 2020

System load:  0.0          Processes:      91
Usage of /:   14.0% of 7.69GB  Users logged in:  0
Memory usage: 16%
Swap usage:   0%

0 packages can be updated.
0 updates are security updates.

Last login: Thu Jul 23 13:43:50 2020 from 14.32.217.164
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@ip-172-31-30-114:~$
```

받았던 pem key 를 chmod 400 으로 설정합니다.

Fingerprint 가 나오면 Yes 를 타이핑 합니다.

AWS EC2 인스턴스에 접속 성공하였습니다!

리눅스 기본 명령어

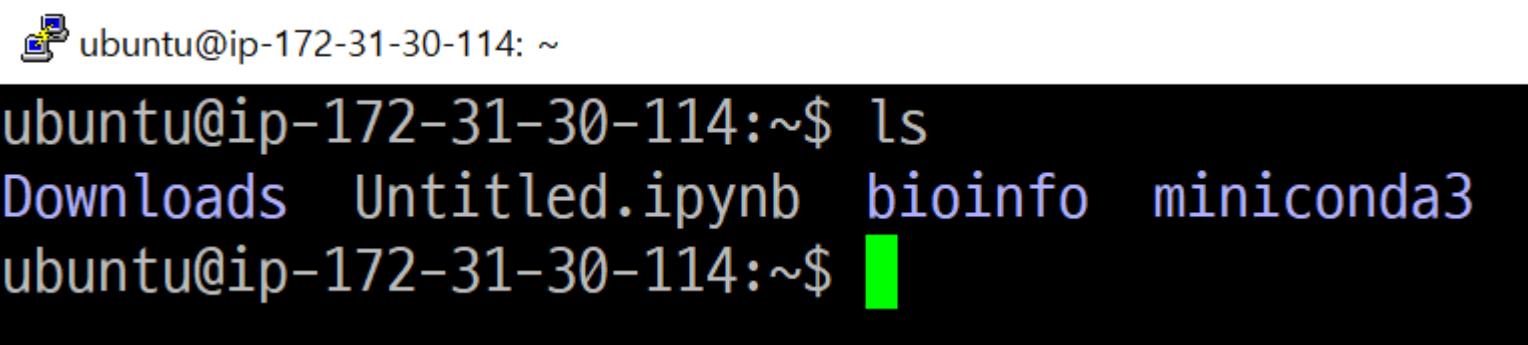
pwd

설명	print name of current/working directory 현재 작업 중인 디렉터리를 출력하는 커맨드입니다.
사용법	pwd [OPTION]…
사용 예	 ubuntu@ip-172-31-30-114: ~ ubuntu@ip-172-31-30-114:~\$ pwd /home/ubuntu ubuntu@ip-172-31-30-114:~\$ █

mkdir

설명	Make directories 디렉터리를 생성하는 커맨드입니다.
사용법	<code>mkdir [OPTION]… DIRECTORY…</code>
사용 예	 ubuntu@ip-172-31-30-114: ~ ubuntu@ip-172-31-30-114:~\$ <code>mkdir bioinfo</code> ubuntu@ip-172-31-30-114:~\$ █

ls

설명	list directory contents 디렉터리의 내용을 보여줍니다.
사용법	ls [OPTION]… [FILE]…
사용 예	 <pre>ubuntu@ip-172-31-30-114: ~ ubuntu@ip-172-31-30-114:~\$ ls Downloads Untitled.ipynb bioinfo miniconda3 ubuntu@ip-172-31-30-114:~\$ █</pre>

ls -l

설명	Use a long listing format 파일과 함께 여러 정보들을 보여줍니다.
사용법	ls -l
사용 예	 ubuntu@ip-172-31-30-114: ~ ubuntu@ip-172-31-30-114:~\$ ls -l total 16 drwxrwxr-x 2 ubuntu ubuntu 4096 Jul 23 13:55 Downloads -rw-rw-r-- 1 ubuntu ubuntu 812 Jul 23 14:29 Untitled.ipynb drwxrwxr-x 2 ubuntu ubuntu 4096 Jul 24 09:03 bioinfo drwxrwxr-x 15 ubuntu ubuntu 4096 Jul 23 13:57 miniconda3 ubuntu@ip-172-31-30-114:~\$

cd

설명	Change the working directory 작업 디렉토리를 이동합니다.
사용법	cd DIRECTORY
사용 예	 ubuntu@ip-172-31-30-114: ~/bioinfo ubuntu@ip-172-31-30-114:~\$ cd bioinfo ubuntu@ip-172-31-30-114:~/bioinfo\$ pwd /home/ubuntu/bioinfo ubuntu@ip-172-31-30-114:~/bioinfo\$ ll total 8 drwxrwxr-x 2 ubuntu ubuntu 4096 Jul 24 09:03 ./ drwxr-xr-x 12 ubuntu ubuntu 4096 Jul 24 09:13 ../ ubuntu@ip-172-31-30-114:~/bioinfo\$ █

touch

설명	Change file timestamps 파일의 타임스탬프를 변경합니다. 새 빈 파일을 만들때도 사용합니다.
사용법	touch [OPTION]... FILE...
사용 예	 ubuntu@ip-172-31-30-114: ~/bioinfo ubuntu@ip-172-31-30-114:~/bioinfo\$ touch 1.txt 2.txt 3.txt ubuntu@ip-172-31-30-114:~/bioinfo\$ ls 1.txt 2.txt 3.txt ubuntu@ip-172-31-30-114:~/bioinfo\$ █

rm

설명	Remove files or directories 파일 또는 디렉터리를 삭제합니다.
사용법	rm [OPTION]... [FILE]...
사용 예	<pre>ubuntu@ip-172-31-30-114:~/bioinfo\$ mkdir test ubuntu@ip-172-31-30-114:~/bioinfo\$ ls 1.txt 2.txt 3.txt test ubuntu@ip-172-31-30-114:~/bioinfo\$ rm 1.txt 2.txt ubuntu@ip-172-31-30-114:~/bioinfo\$ ls 3.txt test ubuntu@ip-172-31-30-114:~/bioinfo\$ rm -r test ubuntu@ip-172-31-30-114:~/bioinfo\$ ls 3.txt ubuntu@ip-172-31-30-114:~/bioinfo\$ █</pre>

wget

설명	The non-interactive network downloader. 네트워크를 통해 파일을 받는 명령어입니다.
사용법	wget [OPTION]... URL...
사용 예	<pre>ubuntu@ip-172-31-30-114:~/bioinfo\$ wget https://raw.githubusercontent.com/KennethJHan/2020-07-bio-lecture/master/src/covid19.fasta --2020-07-24 09:26:52-- https://raw.githubusercontent.com/KennethJHan/2020-07-bio-lecture/master/src/covid19.fasta Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 199.232.64.133 Connecting to raw.githubusercontent.com (raw.githubusercontent.com) 199.232.64.133 :443... connected. HTTP request sent, awaiting response... 200 OK Length: 30429 (30K) [text/plain] Saving to: 'covid19.fasta' covid19.fasta 100%[=====] 29.72K --.-KB/s in 0.01s 2020-07-24 09:26:53 (2.60 MB/s) - 'covid19.fasta' saved [30429/30429] ubuntu@ip-172-31-30-114:~/bioinfo\$ ls 3.txt covid19.fasta ubuntu@ip-172-31-30-114:~/bioinfo\$ </pre>

다음 주소를 사용하세요

wget <https://raw.githubusercontent.com/KennethJHan/2020-07-bio-lecture/master/src/covid19.fasta>

cat

설명	Concatenate files and print on the standard output 파일을 표준 출력으로 출력하는 명령어입니다.
사용법	cat FILE
사용 예	<pre>~/bioinfo\$ cat covid19.fasta >NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome ATTAAAGGTTATACCTTCCCAGGTAAACAAACCAACCAACTTCGATCTCTGTAGATCTGTTCTAA CGAACCTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAAAC TAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTGCAAGGCTGCTTACGGTTCGTCCGTG TTGCAGCCGATCATCAGCACATCTAGGTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGT CCTGGTTTCAACGAGAAAACACACGTCCAACTCAGTTGCCTGTTACAGGTTCGACGTGCTCGTAC GTGGCTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTAAACATCTTAAAGATGGCACTTGTGG</pre>

less

설명	Opposite of more 파일 내부를 확인할 수 있는 명령어입니다.
사용법	less FILE
사용 예	<pre>~/bioinfo\$ less covid19.fasta</pre> >NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome ATTAAGGTTTACCTCCCAGGTAAACAAACCAACCAACTTCGATCTTGTAGATCTGTTCTCTAAA CGAACTTAAAATCTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAAAC TAATTACTGTCGTTGACAGGACACGAGTAACTCGCTATCTTCTGCAGGCTGCTTACGGTTCGTCCGTG TTGCAGCCGATCATCAGCACATCTAGGTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGT CCTGGTTTCAACGAGAAAACACACGTCCAACTCAGTTGCCTGTTTACAGGTTCGCAGGTGCTCGTAC GTGGCTTGGAGACTCCGTGGAGGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGG CTTAGTAGAAGTTGAAAAAGGCCCTTGCCCTCAACTGAACAGGCCATGTGTTCATCAAACGTTCGGAT

head

설명	Output the first part of files 파일의 첫 몇 줄을 출력하는 커맨드입니다.
사용법	head [OPTION]… [FILE]…
사용 예	<pre>~/bioinfo\$ head covid19.fasta >NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome ATTAAAGGTTTATACCTTCCCAGGTAAACAAACCAACCAACTTCGATCTCTGTAGATCTGTTCTCTAAA CGAACTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAAAC TAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTGCAGGCTGCTACGGTTCGTCCGTG TTGCAGCCGATCATCAGCACATCTAGGTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTC CCTGGTTCAACGAGAAAACACACGTCCAACTCAGTTGCCTGTTACAGGTTCGCGACGTGCTCGTAC GTGGCTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGG CTTAGTAGAAGTTGAAAAAGGCCTTGCCTCAACTGAACAGCCCTATGTGTTCATCAAACGTTCGGAT GCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACTCGAAGGCATTCAAGTACGGTC GTAGTGGTGAGACACTTGGTGTCCCTCATGTGGCGAAATACCAGTGGCTACCGCAAGGTTCT ~/bioinfo\$</pre>

tail

설명	Output the first part of files 파일의 마지막 몇 줄을 출력하는 커맨드입니다.
사용법	tail [OPTION]... [FILE]...
사용 예	<pre>~/bioinfo\$ tail covid19.fasta TGAATAAGCATATTGACGCATACAAACATTCCCACCAACAGAGCCTAAAAAGGACAAAAGAAGAAGGC TGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAACAGCAAACGTGACTCTTCTGCTGCAGAT TTGGATGATTCTCAAACAATTGCAACAATCCATGAGCAGTGCTGACTCAACTCAGGCCTAAACTCATG CAGACCACACAAGGCAGATGGGCTATATAACGTTTCGCTTTCCGTTACGATATATAGTCTACTCTT GTGCAGAATGAATTCTCGTAACTACATAGCACAAGTAGATGTAGTTAATCTCACATAGCAATCT TTAACAGTGTGTAACATTAGGGAGGACTTGAAAGAGGCCACCATTTCACCGAGGCCACGCCAGTAC GATCGAGTGTACAGTGAACAATGCTAGGGAGAGCTGCCTATATGGAAGAGCCCTAATGTGTAAAATTAAT TTTAGTAGTGCTATCCCCATGTGATTTAATAGCTTCTAGGAGAATGACAAAAAAAAAAAAAAA AAAAAAA</pre> <pre>~/bioinfo\$</pre>

WC

설명	Print newline, word, and byte counts for each file 파일의 줄, 단어, 문자 수를 출력하는 명령어입니다.
사용법	wc [OPTION]… [FILE]…
사용 예	<pre>~/bioinfo\$ wc covid19.fasta 430 439 30429 covid19.fasta ~/bioinfo\$ █</pre>

WC -l

설명	wc 에서 line (줄) 수를 출력하는 명령어입니다.
사용법	wc -l [FILE]
사용 예	<pre>~/bioinfo\$ wc -l covid19.fasta 430 covid19.fasta ~/bioinfo\$ █</pre>

cp

설명	Copy files and directories 파일이나 디렉터리를 복사하는 명령어입니다.
사용법	cp SOURCE ... DEST
사용 예	<pre>~/bioinfo\$ ls 3.txt covid19.fasta ~/bioinfo\$ cp covid19.fasta covid19.fasta.copy ~/bioinfo\$ ls 3.txt covid19.fasta covid19.fasta.copy ~/bioinfo\$</pre>

df

설명	Report file system disk space usage 디스크 사용량을 알려주는 명령어입니다.
사용법	df
사용 예	<pre>~/bioinfo\$ df -h Filesystem Size Used Avail Use% Mounted on udev 468M 0 468M 0% /dev tmpfs 98M 760K 98M 1% /run /dev/xvda1 7.7G 2.2G 5.6G 28% / tmpfs 490M 0 490M 0% /dev/shm tmpfs 5.0M 0 5.0M 0% /run/lock tmpfs 490M 0 490M 0% /sys/fs/cgroup /dev/loop0 18M 18M 0 100% /snap/amazon-ssm-agent/1566 /dev/loop1 98M 98M 0 100% /snap/core/9289 /dev/loop2 97M 97M 0 100% /snap/core/9665 /dev/loop3 29M 29M 0 100% /snap/amazon-ssm-agent/2012 tmpfs 98M 0 98M 0% /run/user/1000 ~/bioinfo\$</pre>

du

설명	Estimate file space usage 디렉터리의 사용량을 알려주는 명령어입니다.
사용법	du
사용 예	<pre>~/bioinfo\$ du -sh 68K . ~/bioinfo\$ █</pre> <p>du -sh 로 현재 디렉터리의 사용량을 확인할 수 있습니다.</p>

vim

설명	Vi IMproved, a programmer's text editor 텍스트 에디터 입니다.
사용법	vim
사용 예	<pre>~/bioinfo\$ vim hello.sh  ubuntu@ip-172-31-30-114: ~/bioinfo echo "hello world" ~ ~ "hello.sh" 1L, 19C written</pre>

텍스트 에디터 vim 사용 방법

vim 에는 두 가지 모드가 있습니다.

명령모드와 입력모드입니다.

처음 vim에 들어가면 명령모드이며

i 를 눌러 입력모드로 들어갑니다.

타이핑을 진행하고, 마치면 ESC키를 눌러

명령모드로 돌아갑니다.

:w [엔터]를 눌러 편집내용을 저장합니다

:q [엔터]를 눌러 vim을 종료합니다.

텍스트 에디터 vim 사용 방법

```
$ vim hello.sh
```

i 를 눌러 입력 모드로 들어갑니다.

```
echo "hello world"
```

라고 입력합니다.

```
~/bioinfo$ sh hello.sh
hello world
~/bioinfo$
```

ESC 키를 눌러 명령 모드로 돌아갑니다.

:w [엔터] → :q [엔터] 를 눌러 저장&종료 합니다.

```
$ sh hello.sh
```

로 작성한 쉘스크립트를 실행합니다.

chmod

설명	Change file mode bits 파일의 읽기,쓰기,실행 권한을 바꾸는 명령어입니다.
사용법	chmod MODE FILE...
사용 예	<pre>~/bioinfo\$ chmod 775 hello.sh ~/bioinfo\$./hello.sh hello world ~/bioinfo\$ ll total 76 drwxrwxr-x 2 ubuntu ubuntu 4096 Jul 24 14:41 . drwxr-xr-x 12 ubuntu ubuntu 4096 Jul 24 14:41 .. -rw-rw-r-- 1 ubuntu ubuntu 0 Jul 24 09:22 3.txt -rw-rw-r-- 1 ubuntu ubuntu 30429 Jul 24 09:26 covid19.fasta -rw-rw-r-- 1 ubuntu ubuntu 30429 Jul 24 14:29 covid19.fasta.copy -rwxrwxr-x 1 ubuntu ubuntu 19 Jul 24 14:41 hello.sh* ~/bioinfo\$</pre>

whoami

설명	Print effective userid 현재 사용자를 출력해주는 명령어입니다.
사용법	whoami [OPTION]…
사용 예	<pre>~/bioinfo\$ whoami ubuntu ~/bioinfo\$ █</pre>

date

설명	Print or set the system date and time 시스템의 시간을 출력 또는 셋팅해주는 명령어입니다.
사용법	date [OPTION]… [+FORMAT]
사용 예	<pre>~\$ date Sat Jul 25 00:11:12 UTC 2020 ~\$ TZ=Asia/Seoul date Sat Jul 25 09:11:15 KST 2020 ~\$ █</pre>

연습문제

- 1) 홈 디렉터리 (~) 아래에서 bioinfo 디렉터리를 bioinformatics로 복사해보세요
- 2) 복사 후 bioinfo 디렉터리를 bioinfo_bak 으로 이름을 바꿔보세요
- 3) vim으로 hello_time.sh 를 생성하여 hello와 현재 디렉터리, 현재 시간을 출력하는 shell script를 생성해보세요
- 4) covid_19.fasta 의 상위 5줄과 하위 5줄을 covid_19.fasta.new 파일에 저장해보세요

연습문제

wget으로 다음 파일들을 다운로드 해보세요

https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh

https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982_1.filt.fastq.gz

https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982_2.filt.fastq.gz

https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982.mapped.sorted.markdup.bam

https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982.mapped.sorted.markdup.bam.bai

https://raw.githubusercontent.com/KennethJHan/Bioinformatics_Programming_101/master/GATK_BestPractice/SRR000982.filtered.variants.annotated.vcf

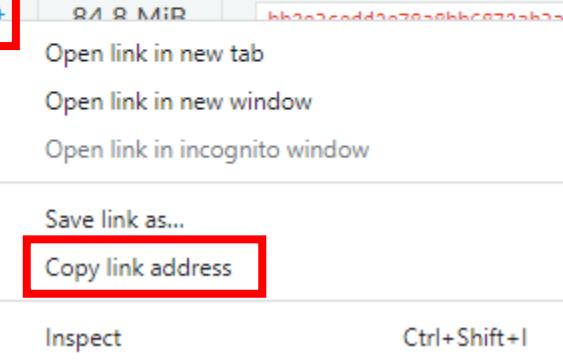
EC2에서 파이썬 설치하기

파이썬 설치하기 (miniconda)

Linux installers

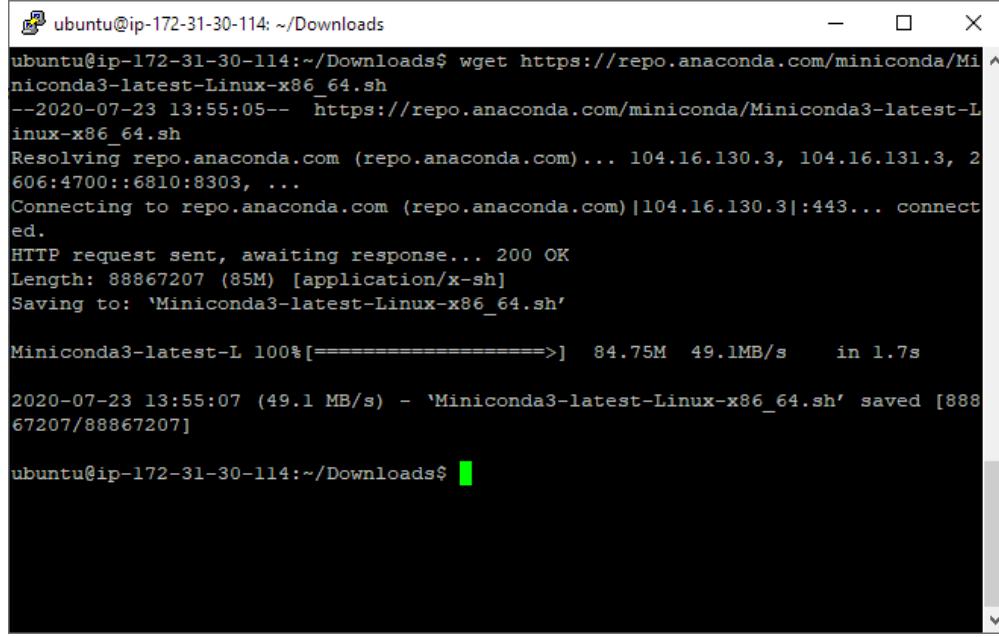
Linux			
Python version	Name	Size	SHA256 hash
Python 3.7	Miniconda3 Linux 64-bit	81.8 MiB	5b1266a90f8c7004a22d8dc2ea5effeb6a439a
	Miniconda3 Linux 32-bit		2d59065b30205c2758a8b86b4c27144adafcc4
Python 2.7	Miniconda2 Linux 64-bit		e7300a252b33b5befd078a15d4a017476b8979
	Miniconda2 Linux 32-bit		26b1a2f2782d0994facdecbb8e0baf51749979

Installing



The screenshot shows a browser context menu for the 'Miniconda3 Linux 64-bit' link. The menu items are: Open link in new tab, Open link in new window, Open link in incognito window, Save link as..., Copy link address (which is highlighted with a red box), and Inspect.

<https://docs.conda.io/en/latest/miniconda.html>

A screenshot of a terminal window titled "ubuntu@ip-172-31-30-114: ~/Downloads". The window shows the execution of a wget command to download the Miniconda3 latest Linux x86_64 installer. The output includes connection details, file download progress, and a success message indicating the file was saved.

```
ubuntu@ip-172-31-30-114:~/Downloads$ wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
--2020-07-23 13:55:05-- https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
Resolving repo.anaconda.com (repo.anaconda.com)... 104.16.130.3, 104.16.131.3, 2606:4700::6810:8303, ...
Connecting to repo.anaconda.com (repo.anaconda.com)|104.16.130.3|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 88867207 (85M) [application/x-sh]
Saving to: 'Miniconda3-latest-Linux-x86_64.sh'

Miniconda3-latest-L 100%[=====] 84.75M 49.1MB/s in 1.7s

2020-07-23 13:55:07 (49.1 MB/s) - 'Miniconda3-latest-Linux-x86_64.sh' saved [88867207/88867207]

ubuntu@ip-172-31-30-114:~/Downloads$
```

wget

https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh

을 입력하고 엔터를 눌러서 다운로드 받습니다.

```
ubuntu@ip-172-31-30-114: ~/Downloads
=====
End User License Agreement - Anaconda Individual Edition
=====

Copyright 2015-2020, Anaconda, Inc.

All rights reserved under the 3-clause BSD License:

This End User License Agreement (the "Agreement") is a legal agreement between you and Anaconda, Inc. ("Anaconda") and governs your use of Anaconda Individual Edition (which was formerly known as Anaconda Distribution).

Subject to the terms of this Agreement, Anaconda hereby grants you a non-exclusive, non-transferable license to:

    * Install and use the Anaconda Individual Edition (which was formerly known as Anaconda Distribution),
    * Modify and create derivative works of sample source code delivered in Anaconda Individual Edition; and
    * Redistribute code files in source (if provided to you by Anaconda as source) and binary forms, with or without modification subject to the requirements set forth below.

--More--
```

```
ubuntu@ip-172-31-30-114: ~/Downloads
cryptography
A Python library which exposes cryptographic recipes and primitives.

pycryptodome
A fork of PyCrypto. It is a self-contained Python package of low-level cryptographic primitives.

pycryptodomex
A stand-alone version of pycryptodome.

libsodium
A software library for encryption, decryption, signatures, password hashing and more.

pynacl
A Python binding to the Networking and Cryptography library, a crypto library with the stated goal of improving usability, security and speed.

Last updated February 25, 2020

Do you accept the license terms? [yes|no]
[no] >>> yes
```

```
ubuntu@ip-172-31-30-114: ~/Downloads
A stand-alone version of pycryptodome.

libsodium
A software library for encryption, decryption, signatures, password hashing and more.

pynacl
A Python binding to the Networking and Cryptography library, a crypto library with the stated goal of improving usability, security and speed.

Last updated February 25, 2020

Do you accept the license terms? [yes|no]
[no] >>> yes

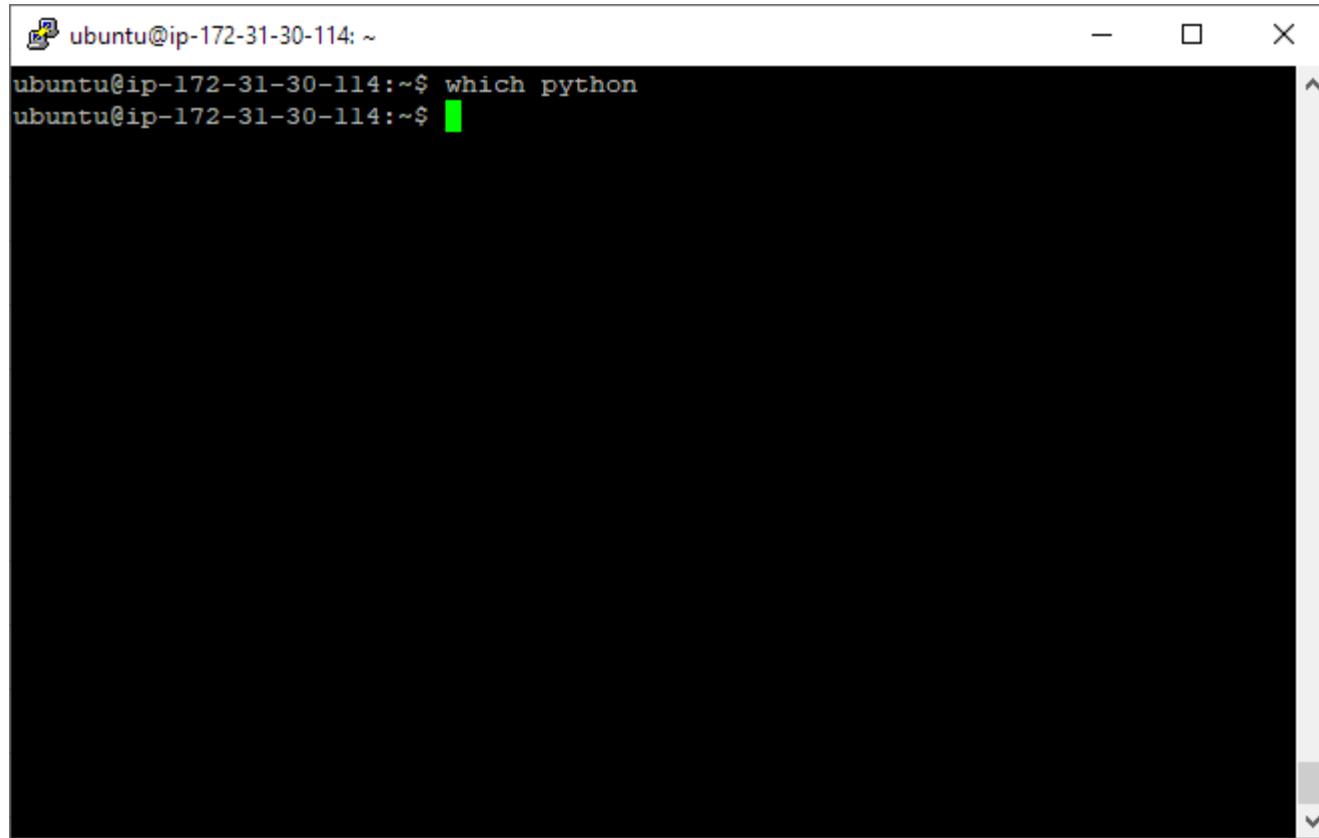
Miniconda3 will now be installed into this location:
/home/ubuntu/miniconda3

- Press ENTER to confirm the location
- Press CTRL-C to abort the installation
- Or specify a different location below

[/home/ubuntu/miniconda3] >>>
```

```
ubuntu@ip-172-31-30-114: ~/Downloads
pkgs/main/linux-64::pyopenssl-19.1.0-py37_0
pkgs/main/linux-64::pysocks-1.7.1-py37_0
pkgs/main/linux-64::python-3.7.7-hcff3b4d_5
pkgs/main/linux-64::readline-8.0-h7b6447c_0
pkgs/main/linux-64::requests-2.23.0-py37_0
pkgs/main/linux-64::ruamel_yaml-0.15.87-py37h7b6447c_0
pkgs/main/linux-64::setuptools-46.4.0-py37_0
pkgs/main/linux-64::six-1.14.0-py37_0
pkgs/main/linux-64::sqlite-3.31.1-h62c20be_1
pkgs/main/linux-64::tk-8.6.8-hbc83047_0
pkgs/main/noarch::tqdm-4.46.0-py_0
pkgs/main/linux-64::urllib3-1.25.8-py37_0
pkgs/main/linux-64::wheel-0.34.2-py37_0
pkgs/main/linux-64::xz-5.2.5-h7b6447c_0
pkgs/main/linux-64::yaml-0.1.7-had09818_2
pkgs/main/linux-64::zlib-1.2.11-h7b6447c_3

Preparing transaction: done
Executing transaction: done
installation finished.
Do you wish the installer to initialize Miniconda3
by running conda init? [yes|no]
[no] >>>
```



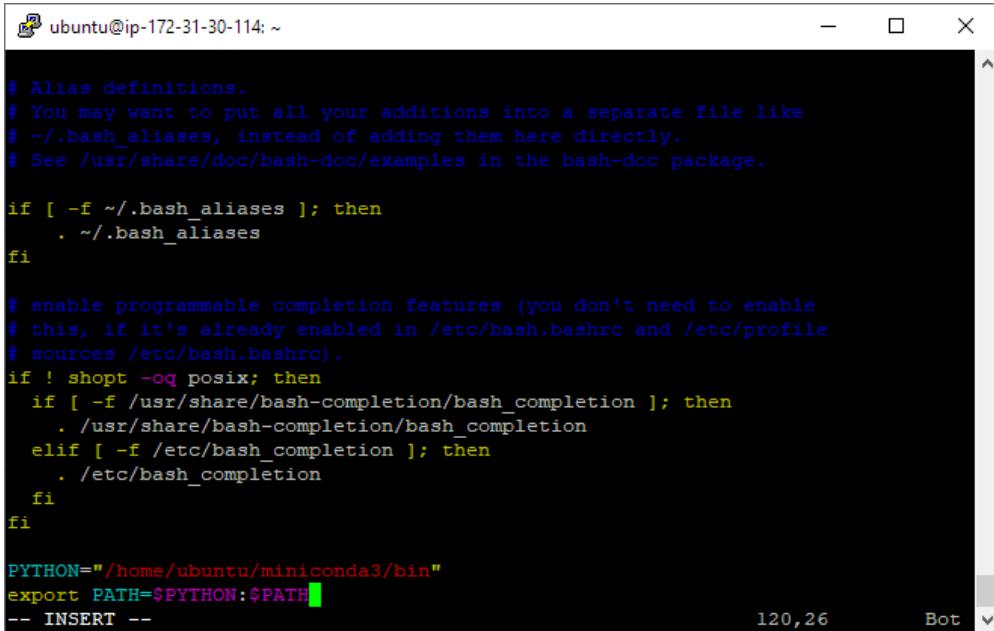
A screenshot of a terminal window titled "ubuntu@ip-172-31-30-114: ~". The window contains the following text:

```
ubuntu@ip-172-31-30-114:~$ which python
```

which python

을 하였을 때 안 나오는 것을 확인

```
ubuntu@ip-172-31-30-114:~$ vi ~/.bashrc
```

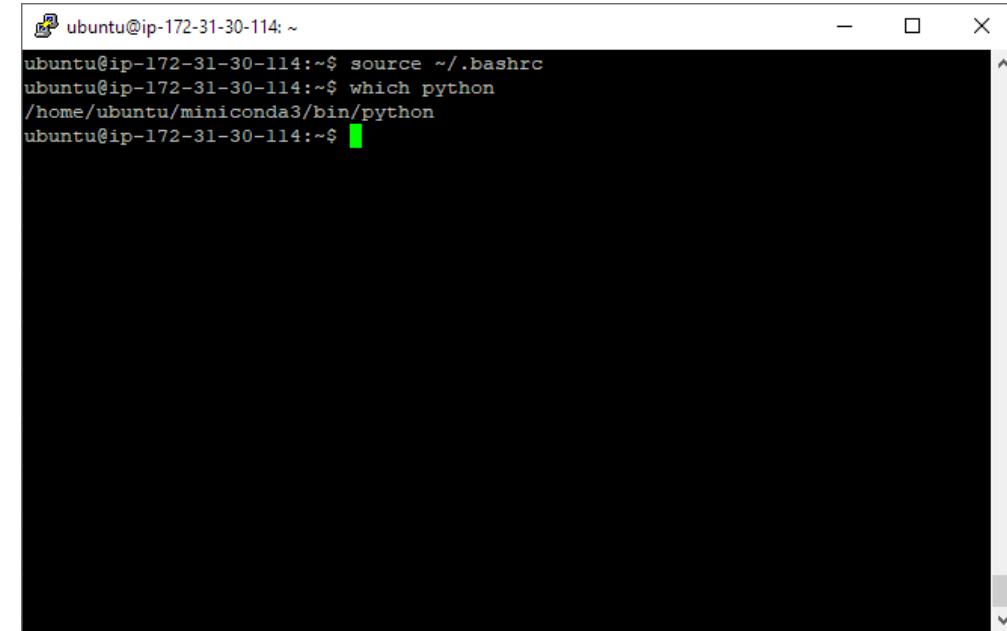


```
# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
  . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -q posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

PYTHON="/home/ubuntu/miniconda3/bin"
export PATH=$PYTHON:$PATH
-- INSERT --
```



```
ubuntu@ip-172-31-30-114:~$ source ~/.bashrc
ubuntu@ip-172-31-30-114:~$ which python
/home/ubuntu/miniconda3/bin/python
ubuntu@ip-172-31-30-114:~$
```

```
ubuntu@ip-172-31-30-114:~$ source ~/.bashrc
```

python path 가 잡혔다.

EC2에서 포트열어
jupyter notebook 구동하기

EC2에서 jupyter notebook 구동하기

EC2에서 jupyter notebook을 구동하기 위해서 세 가지 단계를 실행하면 됩니다.

- 1) 파이썬에서 jupyter 설치
- 2) EC2 리눅스 상에서 포트 (8888) 열기
- 3) AWS console 보안에서 포트 (8888) 열기

```
ubuntu@ip-172-31-30-114: ~/miniconda3/bin  
ubuntu@ip-172-31-30-114:~/miniconda3/bin$ sudo ufw allow 8888  
Rules updated  
Rules updated (v6)  
ubuntu@ip-172-31-30-114:~/miniconda3/bin$
```

sudo ufw allow 8888

을 입력합니다.

New EC2 Experience [Tell us what you think](#)

EC2 Dashboard [New](#)

Events [New](#)

Tags

Limits

Instances

- Instances
- Instance Types
- Launch Templates
- Spot Requests
- Savings Plans
- Reserved Instances
- Dedicated Hosts [New](#)
- Capacity Reservations

Images

- AMIs

Elastic Block Store

- Volumes
- Snapshots
- Lifecycle Manager

Network & Security

- Security Groups [New](#)**
- Elastic IPs [New](#)
- Placement Groups [New](#)
- Key Pairs [New](#)

EC2 > Security Groups

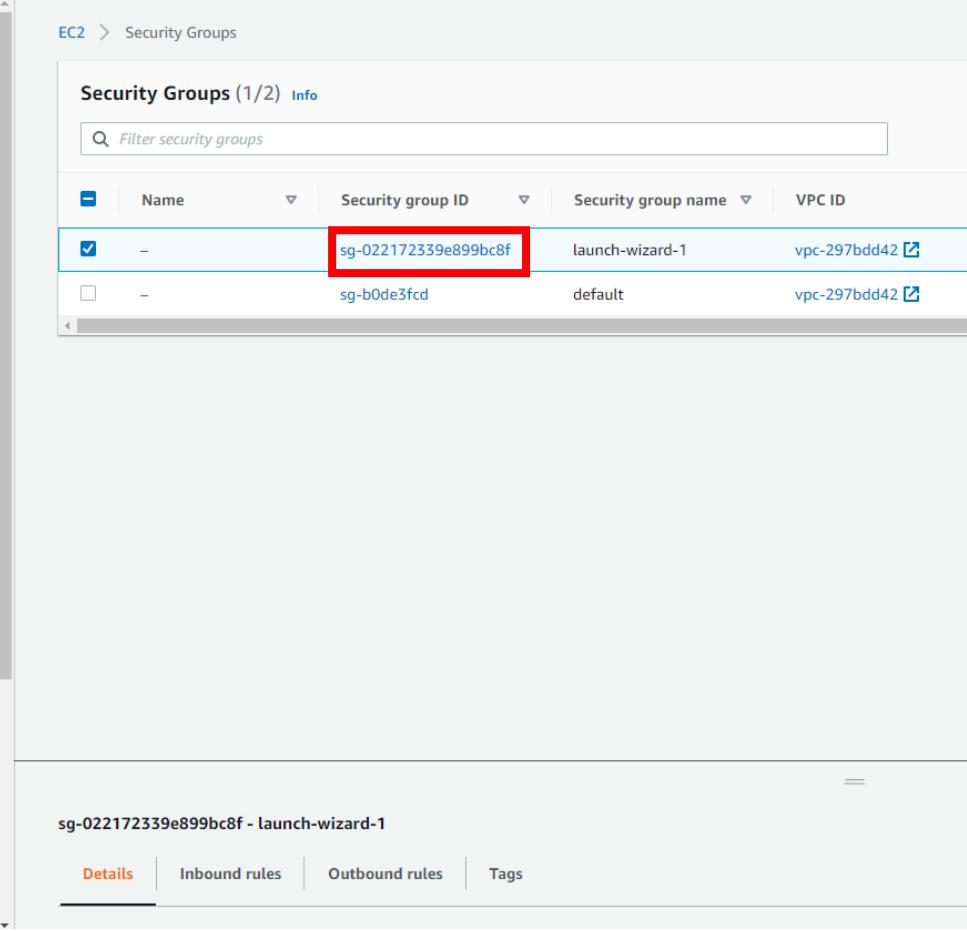
Security Groups (1/2) [Info](#)

Filter security groups

Name	Security group ID	Security group name	VPC ID
<input checked="" type="checkbox"/> -	sg-022172339e899bc8f	launch-wizard-1	vpc-297bdd42
<input type="checkbox"/> -	sg-b0de3fcfcd	default	vpc-297bdd42

sg-022172339e899bc8f - launch-wizard-1

Details Inbound rules Outbound rules Tags



Security Groups 을 클릭하고
launch-wizard-1 옆에 있는
sg- 로 시작하는 Security group ID를
클릭합니다.

New EC2 Experience [Tell us what you think](#)

EC2 Dashboard [New](#)

Events [New](#)

Tags

Limits

Instances

- Instances
- Instance Types
- Launch Templates
- Spot Requests
- Savings Plans
- Reserved Instances
- Dedicated Hosts [New](#)
- Capacity Reservations

Images

- AMIs

Elastic Block Store

- Volumes
- Snapshots
- Lifecycle Manager

Network & Security

- Security Groups [New](#)
- Elastic IPs [New](#)
- Placement Groups [New](#)

Inbound security group rules successfully modified on security group (sg-022172339e899bc8f | launch-wizard-1)

▶ Details

EC2 > Security Groups > sg-022172339e899bc8f - launch-wizard-1

sg-022172339e899bc8f - launch-wizard-1

[Delete security group](#) [Copy to new security group](#)

Details			
Security group name	Security group ID	Description	VPC ID
launch-wizard-1	sg-022172339e899bc8f	launch-wizard-1 created 2020-07-23T21:35:35.706+09:00	vpc-297bdd42
Owner	Inbound rules count	Outbound rules count	
981427888788	2 Permission entries	1 Permission entry	

[Inbound rules](#) [Outbound rules](#) [Tags](#)

Inbound rules					Edit inbound rules
Type	Protocol	Port range	Source	Description - optional	
Custom TCP	TCP	8888	0.0.0.0/0	-	
SSH	TCP	22	0.0.0.0/0	-	

Edit inbound rules 를 클릭합니다.

Edit inbound rules Info

Inbound rules control the incoming traffic that's allowed to reach the instance.

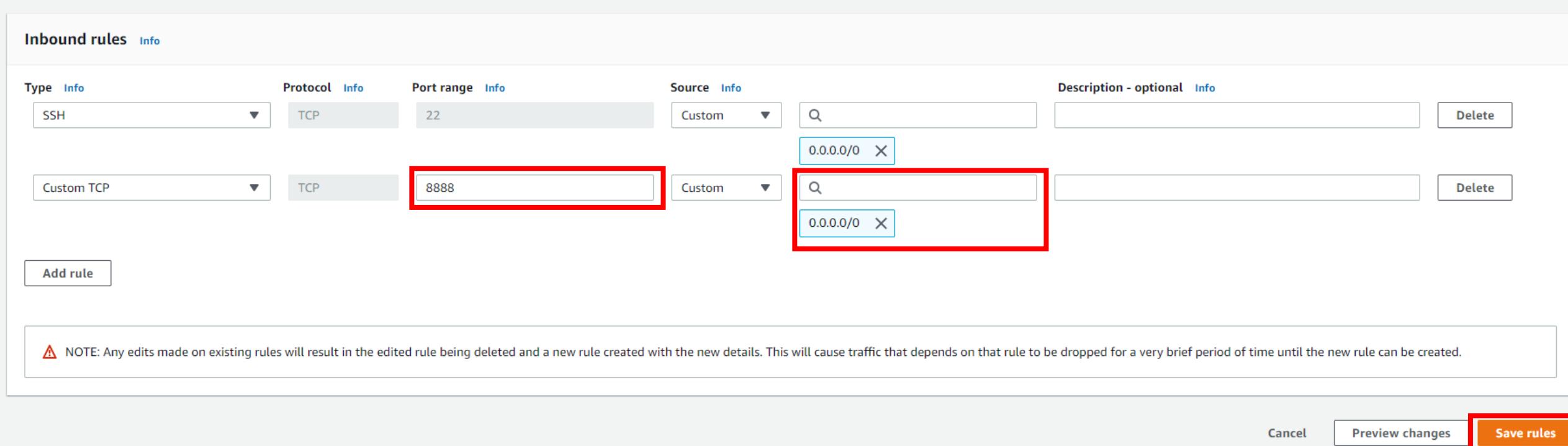
Inbound rules Info

Type <small>Info</small>	Protocol <small>Info</small>	Port range <small>Info</small>	Source <small>Info</small>	Description - optional <small>Info</small>	
SSH	TCP	22	Custom	0.0.0.0/0 <small>X</small>	Delete
Custom TCP	TCP	8888	Custom	0.0.0.0/0 <small>X</small>	Delete

Add rule

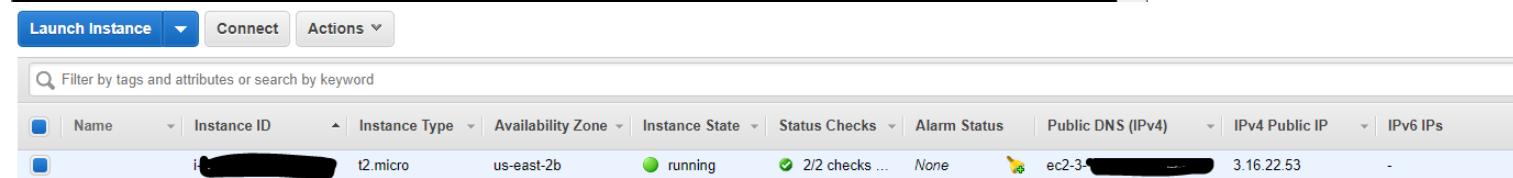
⚠ NOTE: Any edits made on existing rules will result in the edited rule being deleted and a new rule created with the new details. This will cause traffic that depends on that rule to be dropped for a very brief period of time until the new rule can be created.

Cancel Preview changes **Save rules**



8888 과 0.0.0.0/0 을 넣고 Save rules 를 클릭합니다.

```
ubuntu@ip-172-31-30-114: ~
ubuntu@ip-172-31-30-114:~$ jupyter notebook --ip=*
```



```
Instance: i-09b27bf28513f2b10 Public DNS: ec2-3-16-22-53.us-east-2.compute.amazonaws.com

Description Status Checks Monitoring Tags
Instance ID: i-09b27bf28513f2b10
Instance state: running
Instance type: t2.micro
Finding: Opt-in to AWS Compute Optimizer for recommendations. Learn more

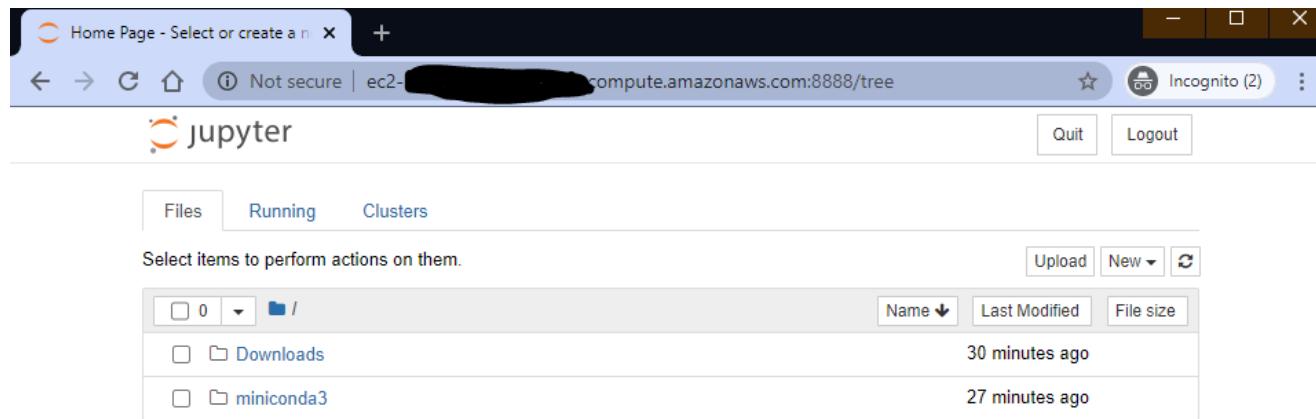
Public DNS (IPv4) ①
IPv4 Public IP: ec2-3-16-22-53.us-east-2.compute.amazonaws.com
IPv6 IPs: -
Elastic IPs: -
```

[I 14:19:35.736 NotebookApp] The Jupyter Notebook is running at:
[I 14:19:35.736 NotebookApp] http://ip-172-31-30-114:8888/?token=84efb799981d48517335ab93b0938606b2250541714b2963
[I 14:19:35.736 NotebookApp] or http://127.0.0.1:8888/?token=84efb799981d4851735ab93b0938606b2250541714b2963
[I 14:19:35.736 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[W 14:19:35.740 NotebookApp] No web browser found: could not locate runnable browser.
[C 14:19:35.741 NotebookApp]

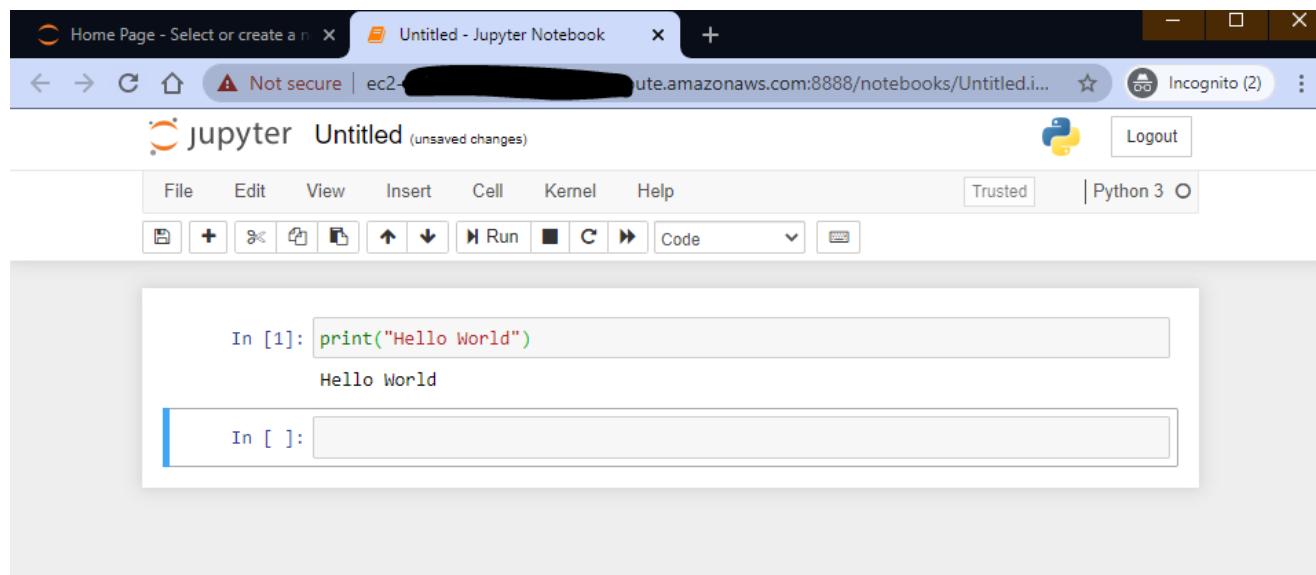
To access the notebook, open this file in a browser:
file:///home/ubuntu/.local/share/jupyter/runtime/nbserver-2677-open.html
Or copy and paste one of these URLs:
http://ip-172-31-30-114:8888/?token=84efb799981d48517335ab93b0938606b2250541714b2963
or http://127.0.0.1:8888/?token=84efb799981d48517335ab93b0938606b2250541714b2963

①과 ②를 합쳐서 하나의 주소로 만들고

사용하고 있는 PC(노트북) 크롬 창에 붙여넣기 합니다.



와! 성공!!!



DNA 분석 파이프라인

실습에 필요한 파일

FASTQ

https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982_1.filt.fastq.gz

https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982_2.filt.fastq.gz

BAM

https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982.mapped.sorted.markdup.bam

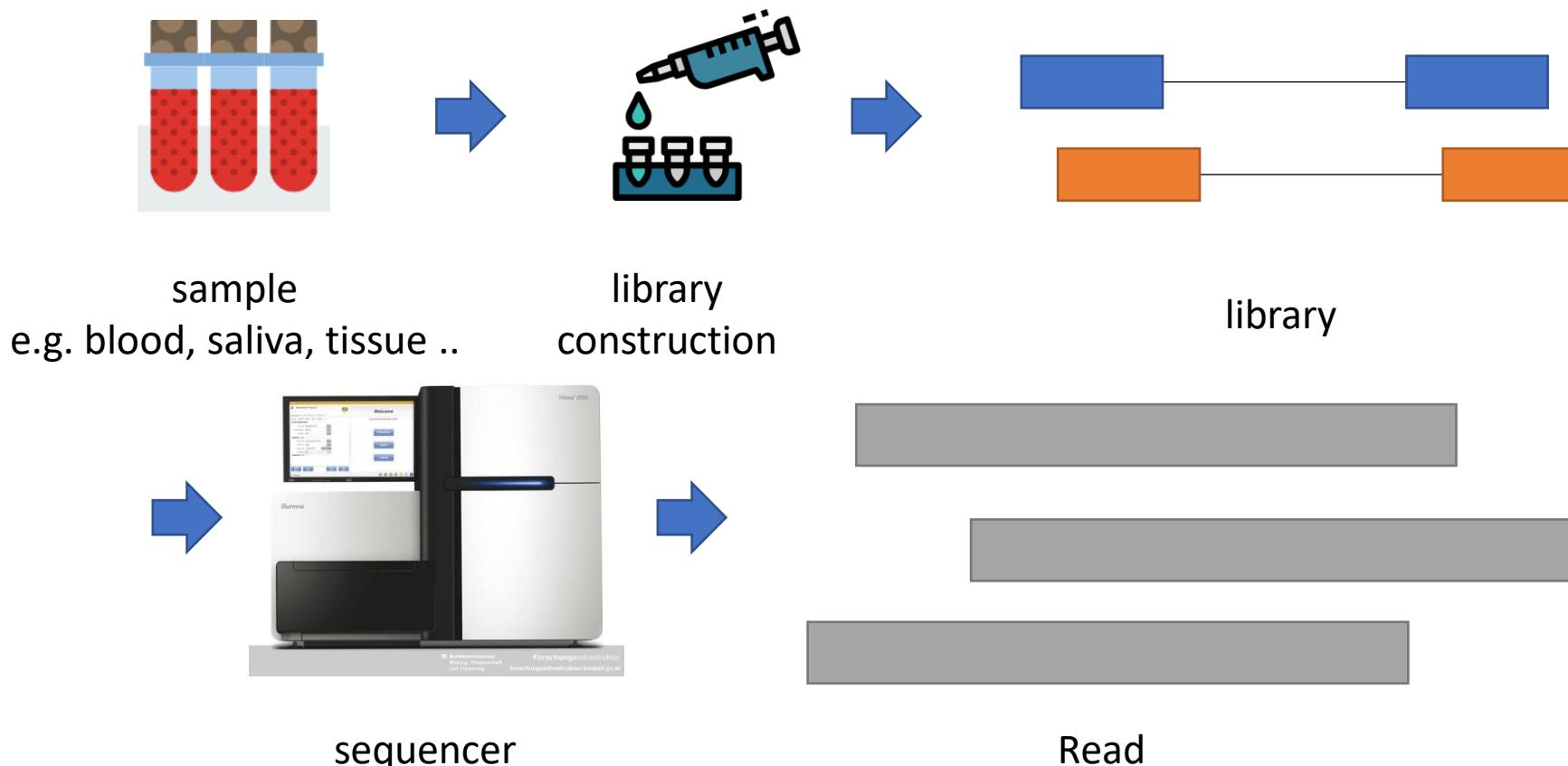
https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982.mapped.sorted.markdup.bam.bai

VCF

https://raw.githubusercontent.com/KennethJHan/Bioinformatics_Programming_101/master/GATK_BestPractice/SRR000982.filtered.variants.annotated.vcf

wget 명령어로 다운로드 받아놓읍시다.

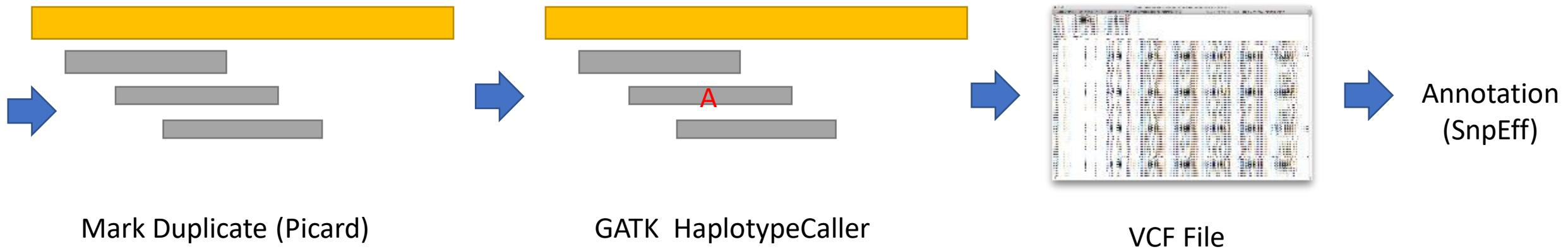
Sequencing의 전반적 개요



혈액, 타액, 조직 등과 같은 샘플에서 DNA를 추출하여 sequencer가 읽을 수 있도록 라이브러리를 제작하고 sequencing을 합니다. Sequencer가 라이브러리를 읽으면 read라고 하는 단위로 염기서열을 읽습니다. Illumina(일루미나) sequencing 장비의 경우 라이브러리를 앞뒤로 읽어서 read1, read2로 두 개의 read 파일을 생성합니다.



읽어낸 read는 기준서열(reference sequence)와 비교하여 조각을 맞추는 align 작업을 하고, pcr 과정에서 생긴 duplicate을 picard와 같은 툴로 제거합니다.



그리고 난 뒤 기준 서열과 다른 염기를 찾아내는 과정인 variant calling을 진행합니다. Variant call을 하게 되면 VCF (Variant Calling Format) 파일이 생성됩니다. VCF 파일은 테이블과 같은 구조로 생긴 텍스트 파일입니다. VCF 파일에는 변이의 위치와 기준서열, 변이서열 정보만 있어서 이 변이가 어떤 의미인지 붙이는 작업이 필요한데, 이를 annotation이라고 합니다.

Tool summary

Tool name	Role	Input	Output
BWA	Align Fastq read to reference sequence	Read1.fastq Read2.fastq	Raw.bam
samtools sort	Sort aligned read in chromosome position order	Raw.bam	Sorted.bam
picard MarkDuplicate	MarkDuplicate to remove artificial effect during sequencing	Sorted.bam	Markdup.bam
GATK BaseRecalibrate	By using known variants(e.g. dbsnp), recalibrate variant quality	Markdup.bam	Recal.bam
GATK HaplotypeCaller	Variant call from bam file	Recal.bam	Raw.vcf
GATK VariantFilter	Apply hard-filter	Raw.vcf	Filtered.Varian ts.vcf
Annotation	Annotate variant information	Filtered.Varia nts.vcf	Annotated.vcf

FASTQ → BAM → VCF

FASTQ 파일 살펴보기

FASTQ 파일은 보통 gzip으로 압축된 형태이기에 눈으로 보려면 zless 명령어를 사용하여 봅니다.

FASTQ 파일은 네 줄이 하나의 read로 구성되어 있습니다.
첫 번째 줄은 @로 시작하는 헤더
두 번째 줄은 염기서열
세 번째 줄은 + 기호로 된 구분 줄
네 번째 줄은 염기서열에 대한 quality score 입니다.

```
kenneth_jh_han@instance-1:~/Downloads$ zless SRR000982_1.filt.fastq.gz
@SRR000982.5 E745RJU01DDHJ6 length=113
AAGGCACCATGCAGAGATGCAAGGCCCTTCTAAGCCTAGACTTCTGGATGACACTTCTAGAAACACCCCTGGCCAGAAGTGAACCTGTCGCCCTGAAGGAAATACTCG
+
DDDDDDDDDDDDDDDDFFDBB:::@DDDDDDDDDDFEDDAAADDxDDDDDDDDDDDA8666@DD@#866AAADDxDDDDDDDDDDDDDDDDCCC@AACDDDDDD
@SRR000982.26 E745RJU01BNUR length=113
AAGGCACCATGCAGAGATGCAAGGCCCTTCTAAGCCTAGACTTCTGGATGACACTTCTAGAAACACCCCTGGCCAGAAGTGAACCTGTCGCCCTGAAGGAAATACTCAG
+
DDDDDDDDDDDDDDDDDD66555566@DDDDDDDGIIIFFEBBDDDDDDDBBBBDDD@#8552@DD@#8669@#DDDDDDDDDDDDDB@000000005566BBDDDDDD
@SRR000982.32 E745RJU01BGMLP length=66
AGGGAAAGGACTCTATAAGATGATATAGTAGACATCTGAAGTCAGCAAGGTATGAGCAAT
+
FFFFFFFFFFFFFFFIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIFFFFFFFFFFFFFFFF
```

Question

FASTA 라고 들어본적이 있는데 FASTQ와 FASTQ의 차이점은 무엇인가요?

FASTA

KT225476.2.fasta

```
>KT225476.2 Middle East respiratory syndrome coronavirus  
isolate MERS-CoV/THA/CU/17_06_2015, complete genome  
AGTGAATAGCTGGCTATCTCACTTCCCCTCGTTCTTGCAGAACTTGATTAAACGAA  
CTTAAATAA  
AAGCCCTGTTAGCGTATTGTTGCACTTGTCTGGTGGATTGTGGCACTAATTGCCT  
GCTCATCTA  
... 중간 생략 ...  
ATCAACAGACCAAATGTCTGAACCTCCAAAGGAGCAGCGTGTGCAAGGTAGCATCACTCAG  
CGCACTCGC  
ACCCGTCCAAGTGTTCAGCCTGGTCCAATGATTGATGTTAACACTGATTAGTGTCACTC
```

>로 시작하는 헤더와
두 번째 줄부터는 서열이 있다.



FASTQ 파일에서 염기 세기

```
kenneth_jh_han@instance-1:~/Downloads$ python3 read_base.py SRR000982_1.filt.fastq.gz  
3743235 2466934 2408976 3770887  
kenneth_jh_han@instance-1:~/Downloads$ python3 read_base.py SRR000982_2.filt.fastq.gz  
3630460 2403252 2436010 3625665  
kenneth_jh_han@instance-1:~/Downloads$
```

read1, read2 FASTQ 파일에서 A, C, G, T 염기수를 세어보세요.
결과는 A, C, G, T 순으로 그림과 같이 출력합니다.
gzip 파일 상태로 읽기 위해서 import gzip 을 사용해보세요.
gzip 사용법을 구글에서 스스로 찾아서 사용방법을 익혀봅시다.

리눅스에서 툴 설치하기

Samtools 설치

1) Samtools 홈페이지

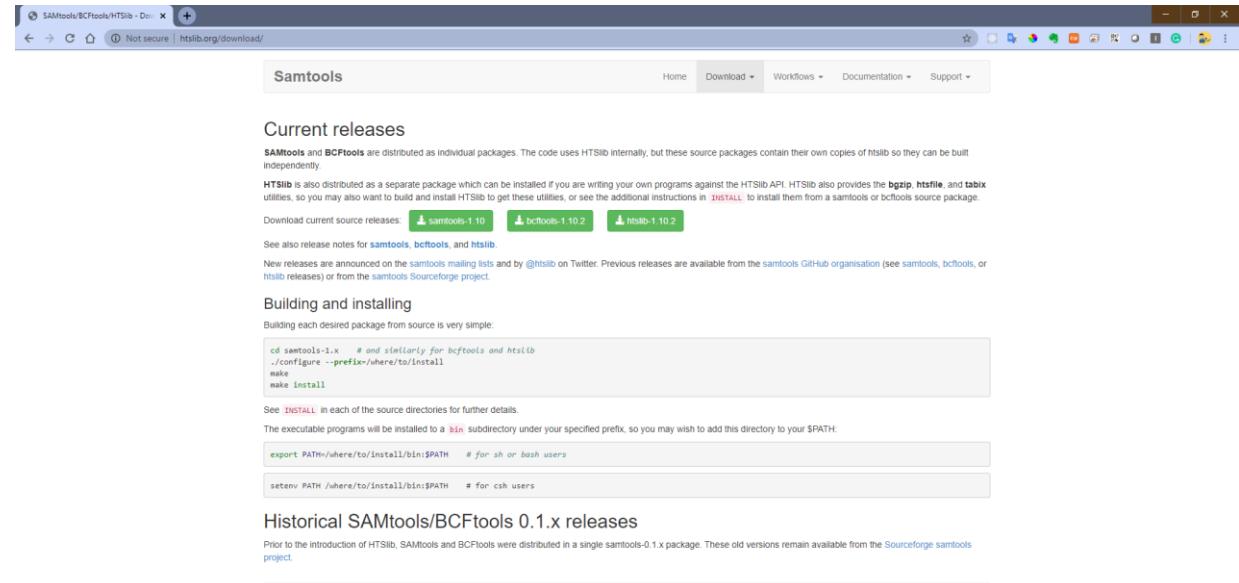
<http://www.htslib.org/download/>

에 들어가서 samtools-1.10 으로 된 버튼에서 우클릭 후 주소복사를 합니다.

2) 복사한 주소

<https://github.com/samtools/samtools/releases/download/1.10/samtools-1.10.tar.bz2>

를 리눅스에서 wget을 사용하여 다운로드 받습니다.



samtools 설치

3) tar xf 명령어로 압축을 해제합니다.

4) 압축을 해제한 디렉터리에 들어가서
./configure 라고 입력합니다.

만약 오른쪽 아래 그림같이
gcc... no 라고 뜬다면
c compiler가 없는 것으로
설치해주어야합니다. 4-1 참조.

4-1) sudo apt-get install gcc g++
을 타이핑하고 엔터를 눌러 c compiler를
설치해줍니다. 중간에 묻는것이 나오면
y를 눌러서 계속 진행해줍니다.

```
kenneth_jh_han@instance-1:~/Downloads$ wget https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982.mapped.sorted.markdup.bam
--2020-05-21 09:32:52-- https://github.com/KennethJHan/Bioinformatics_Programming_101/raw/master/GATK_BestPractice/SRR000982.mapped.sorted.markdup.bam
Resolving github.com (github.com)... 140.82.114.4
Connecting to github.com (github.com) |140.82.114.4|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://raw.githubusercontent.com/KennethJHan/Bioinformatics_Programming_101/master/GATK_BestPractice/SRR000982.mapped.sorted.markdup.bam [following]
--2020-05-21 09:32:52-- https://raw.githubusercontent.com/KennethJHan/Bioinformatics_Programming_101/master/GATK_BestPractice/SRR000982.mapped.sorted.markdup.bam
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 151.101.192.133, 151.101.128.133, 151.101.64.133
...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com) |151.101.192.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 23967674 (23M) [application/octet-stream]
Saving to: 'SRR000982.mapped.sorted.markdup.bam'
kenneth_jh_han@instance-1:~/Downloads$ tar xf samtools-1.10.tar.bz2
kenneth_jh_han@instance-1:~/Downloads$ cd samtools-1.10/
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ ls
AUTHORS      bam2depth.c   bam_mate.c      bam_tview.c      configure.ac    padding.c      stats_isize.c
ChangeLog.old  bam_addrprg.c  bam_md.c       bam_tview.h      coverage.c     phase.c       stats_isize.h
INSTALL       bam_aux.c     bam_plbuf.c    bam_tview_curses.c  cut_target.c  sam.c        test
LICENSE       bam_cat.c     bam_plbuf.h    bam_tview_html.c  dict.c        sam.h       tmp_file.c
Makefile       bam_color.c   bam_plcmd.c   bamshuf.c      doc          sam_opts.c   tmp_file.h
NEWS          bam_endian.h  bam_quickcheck.c  bamtk.c       examples      sam_opts.h   version.sh
README         bam_fastq.c   bam_reheader.c  bedcov.c      faidx.c      sam_utils.c
bam.c          bam_flags.c   bam_rmdup.c    bedidx.c      htslib-1.10  sam_view.c
bam.h          bam_index.c   bam_rmdupse.c  bedidx.h      install-sh   sample.c
bam2bcf.c     bam_lpileup.c  bam_sort.c    config.h.in   lz4         sample.h
bam2bcf.h     bam_lpileup.h  bam_split.c   config.mk.in  m4          samtools.h
bam2bcf_indel.c  bam_markdup.c  bam_stat.c   configure     misc         stats.c
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ ./configure
checking for gcc... no
checking for cc... no
checking for cl.exe... no
configure: error: in `/home/kenneth_jh_han/Downloads/samtools-1.10':
configure: error: no acceptable C compiler found in $PATH
See `config.log' for more details
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ sudo apt-get install gcc g++
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  grub-pc-bin libnuma1
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  binutils binutils-common binutils-x86_64-linux-gnu cpp cpp-7 g++-7 gcc-7 gcc-7-base libasan4 libatomic1
  libbinutils libc-dev-bin libc6-dev libcc1-0 libcilkrt5 libgcc-7-dev libgomp1 libisl19 libltsan0 libmpc3
  libmpx2 libquadmath0 libstdc++-7-dev libtsan0 libubsan0 linux-libc-dev manpages-dev
Suggested packages:
  binutils-doc cpp-doc gcc-7-locales g++-multilib g++-7-multilib gcc-7-doc libstdc++-6-7-dbg gcc-multilib make
  autoconf automake libtinfo flex bison gdb gcc-doc gcc-7-multilib libgcc1-dbg libgomp1-dbg libitm1-dbg
  libatomic1-dbg libasan4-dbg libtsan0-dbg libubsan0-dbg libcilkrt5-dbg libmpx2-dbg
  libquadmath0-dbg glibc-doc libstdc++-7-doc
The following NEW packages will be installed:
  binutils binutils-common binutils-x86_64-linux-gnu cpp cpp-7 g++-7 gcc gcc-7 gcc-7-base libasan4 libatomic1
  libbinutils libc-dev-bin libc6-dev libcc1-0 libcilkrt5 libgcc-7-dev libgomp1 libisl19 libltsan0 libmpc3
  libmpx2 libquadmath0 libstdc++-7-dev libtsan0 libubsan0 linux-libc-dev manpages-dev
0 upgraded, 30 newly installed, 0 to remove and 24 not upgraded.
Need to get 41.8 MB of archives.
After this operation, 160 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
```

samtools 설치

4-1) which gcc 와 which g++ 가

다음과 같이 나온다면 제대로 설치
된 것입니다.

5) ./configure 를 하였을 때 다음과
같이 FAILED 로 나올 수 있습니다.

이건 뭔가 잘못해서 그런건 아니고,
리눅스에서 프로그램을 설치할 때

필요한 dependency (의존)

프로그램이 없어서 그런것입니다.

다들 프로그래밍을 해보셔서 잘 아시겠지만

스스로 디버깅 할 수 있는 수준이 되면

프로그래밍이 쉬워집니다. 디버깅을 하는 방법은
오류 메시지를 잘 읽어 보는 것 입니다.

마찬가지로 다음 그림의 노란색 네모친 부분은
보시면 libncurses5-dev 를 설치하라고 나옵니다.

(ubuntu 기준)

6) sudo apt-get install libncurses5-dev

를 실행합니다.

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ which gcc
/usr/bin/gcc
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ which g++
/usr/bin/g++
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ ./configure
Checking for gcc... gcc
Checking whether the C compiler works... yes
Checking for C compiler default output file name... a.out
Checking for suffix of executables...
Checking whether we are cross compiling... no
Checking for suffix of object files... o
Checking whether we are using the GNU C compiler... yes
Checking whether gcc accepts -g... yes
Checking for gcc option to accept ISO C89... none needed
Checking for grep that handles long lines and -e... /bin/grep
Checking for C compiler warning flags... -Wall
Checking for special C compiler options needed for large files... no
Checking for _FILE_OFFSET_BITS value needed for large files... no
Checking location of HTSlib source tree... htllib-1.10
Checking for NcursesW wide-character library... no
Checking for Ncurses library... no
Checking for Curses library... no
configure: error: curses development files not found
```

The 'samtools tview' command uses the curses text user interface library.
Building samtools with tview requires curses/ncurses/etc development files
to be installed on the build machine; you may need to ensure a package such
as libncurses5-dev (on Debian or Ubuntu Linux) or ncurses-devel (on RPM-based
Linux distributions) is installed.

FAILED. Either configure --without-curses or resolve this error to build
samtools successfully.

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ sudo apt-get install libncurses5-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  grub-pc-bin libnuma1
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  libtinfo-dev
Suggested packages:
  ncurses-doc
The following NEW packages will be installed:
  libncurses5-dev libtinfo-dev
0 upgraded, 2 newly installed, 0 to remove and 24 not upgraded.
Need to get 256 kB of archives.
After this operation, 1422 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://us-west1.gce.archive.ubuntu.com/ubuntu bionic-updates/main amd64 libtinfo-dev amd64 6.1-1ubuntu1.18.04 [81.3 kB]
Get:2 http://us-west1.gce.archive.ubuntu.com/ubuntu bionic-updates/main amd64 libncurses5-dev amd64 6.1-1ubuntu1.18.04 [174 kB]
Fetched 256 kB in 0s (2755 kB/s)
Selecting previously unselected package libtinfo-dev:amd64.
(Reading database ... 70400 files and directories currently installed.)
Preparing to unpack .../libtinfo-dev_6.1-1ubuntu1.18.04_amd64.deb ...
Unpacking libtinfo-dev:amd64 (6.1-1ubuntu1.18.04) ...
Selecting previously unselected package libncurses5-dev:amd64.
Preparing to unpack .../libncurses5-dev_6.1-1ubuntu1.18.04_amd64.deb ...
Unpacking libncurses5-dev:amd64 (6.1-1ubuntu1.18.04) ...
Setting up libtinfo-dev:amd64 (6.1-1ubuntu1.18.04) ...
Setting up libncurses5-dev:amd64 (6.1-1ubuntu1.18.04) ...
Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

Samtools 설치

7) 다시 ./configure 를 입력합니다.
이번에도 FAILED가 나왔는데,
노란색 네모친 부분을 보면
이유가 조금 다릅니다.
sudo apt-get install zlib1g-dev
로 dependency를 설치해줍니다.

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ ./configure
checking for gcc... gcc
checking whether the C compiler works... yes
checking for C compiler default output file name... a.out
checking for suffix of executables...
checking whether we are cross compiling... no
checking for suffix of object files... o
checking whether we are using the GNU C compiler... yes
checking whether gcc accepts -g... yes
checking for gcc option to accept ISO C89... none needed
checking for grep that handles long lines and -e... /bin/grep
checking for C compiler warning flags... -Wall
checking for special C compiler options needed for large files... no
checking for _FILE_OFFSET_BITS value needed for large files... no
checking location of HTSlib source tree... htslib-1.10
checking for NcursesW wide-character library... no
checking for Ncurses library... yes
checking for working ncurses/curses.h... no
checking for working ncurses.h... yes
checking for zlib.h... no
checking for inflate in -lz... no
configure: error: zlib development files not found
```

Samtools uses compression routines from the zlib library <<http://zlib.net>>.
Building samtools requires zlib development files to be installed on the build
machine; you may need to ensure a package such as zlib1g-dev (on Debian or
Ubuntu Linux) or zlib-devel (on RPM-based Linux distributions) is installed.

FAILED. This error must be resolved in order to build samtools successfully.

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ sudo apt-get install zlib1g-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  grub-pc-bin libnuma1
Use 'sudo apt autoremove' to remove them.
The following NEW packages will be installed:
  zlib1g-dev
0 upgraded, 1 newly installed, 0 to remove and 24 not upgraded.
Need to get 176 kB of archives.
After this operation, 457 kB of additional disk space will be used.
Get:1 http://us-west1.gce.archive.ubuntu.com/ubuntu bionic amd64 zlib1g-dev amd64 1:1.2.11.dfsg-0ubuntu2 [176 kB]
Fetched 176 kB in 0s (11.1 MB/s)
Selecting previously unselected package zlib1g-dev:amd64.
(Reading database ... 70450 files and directories currently installed.)
Preparing to unpack .../zlib1g-dev_1%3a1.2.11.dfsg-0ubuntu2_amd64.deb ...
Unpacking zlib1g-dev:amd64 (1:1.2.11.dfsg-0ubuntu2) ...
Setting up zlib1g-dev:amd64 (1:1.2.11.dfsg-0ubuntu2) ...
Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

samtools 설치

8) 다시 ./configure를 입력합니다.

마치 프로그래밍 디버깅 하듯이

오류 메시지를 보고 필요한

dependency 들을 설치해주면 됩니다.

윈도우나 맥과 같은 환경에서는 그냥 클릭하면

알아서 잘 설치해주었는데 리눅스에서는

조금은 불편하겠지만, 이렇게 설치해 주어야

합니다.

sudo apt-get install libbz2-dev

를 입력합니다.

```
checking for library containing log... -lm
checking for zlib.h... yes
checking for inflate in -lz... yes
checking for library containing recv... none required
checking for bzlib.h... no
checking for BZ2_bzBuffToBuffCompress in -lbz2... no
configure: error: libbz2 development files not found
```

```
The CRAM format may use bzip2 compression, which is implemented in HTSlib
by using compression routines from libbz2 <http://www.bzip.org/>.
```

```
Building HTSlib requires libbz2 development files to be installed on the
build machine; you may need to ensure a package such as libbz2-dev (on Debian
or Ubuntu Linux) or bzip2-devel (on RPM-based Linux distributions or Cygwin)
is installed.
```

```
Either configure with --disable-bz2 (which will make some CRAM files
produced elsewhere unreadable) or resolve this error to build HTSlib.
configure: error: ./configure failed for htslib-1.10
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ sudo apt-get install libbz2-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  grub-pc-bin libnuma1
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  bzip2-doc
The following NEW packages will be installed:
  bzip2-doc libbz2-dev
0 upgraded, 2 newly installed, 0 to remove and 24 not upgraded.
Need to get 324 kB of archives.
After this operation, 514 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://us-west1.gce.archive.ubuntu.com/ubuntu bionic-updates/main amd64 bzip2-doc all 1.0.6-8.1ubuntu0.2 [294 kB]
Get:2 http://us-west1.gce.archive.ubuntu.com/ubuntu bionic-updates/main amd64 libbz2-dev amd64 1.0.6-8.1ubuntu0.2 [30.0 kB]
Fetched 324 kB in 0s (13.8 MB/s)
Selecting previously unselected package bzip2-doc.
(Reading database ... 70478 files and directories currently installed.)
Preparing to unpack .../bzip2-doc_1.0.6-8.1ubuntu0.2_all.deb ...
Unpacking bzip2-doc (1.0.6-8.1ubuntu0.2) ...
Selecting previously unselected package libbz2-dev:amd64.
Preparing to unpack .../libbz2-dev_1.0.6-8.1ubuntu0.2_amd64.deb ...
Unpacking libbz2-dev:amd64 (1.0.6-8.1ubuntu0.2) ...
Setting up libbz2-dev:amd64 (1.0.6-8.1ubuntu0.2) ...
Setting up bzip2-doc (1.0.6-8.1ubuntu0.2) ...
Processing triggers for install-info (6.5.0.dfsg.1-2) ...
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

samtools 설치

9) 다시 ./configure를 합니다.

이번에도 에러가 났는데

sudo apt-get install liblzma-dev
로 설치해줍니다.

```
checking for fdatasync... yes
checking for library containing log... -lm
checking for zlib.h... yes
checking for inflate in -lz... yes
checking for library containing recv... none required
checking for bzlib.h... yes
checking for BZ2_bzBuffToBuffCompress in -lbz2... yes
checking for lzma.h... no
checking for lzma_easy_buffer_encode in -llzma... no
configure: error: liblzma development files not found
```

The CRAM format may use LZMA2 compression, which is implemented in HTSlib
by using compression routines from liblzma <<http://tukaani.org/xz/>>.

```
Building HTSlib requires liblzma development files to be installed on the
build machine; you may need to ensure a package such as liblzma-dev (on Debian
or Ubuntu Linux), xz-devel (on RPM-based Linux distributions or Cygwin), or
xz (via Homebrew on macOS) is installed; or build XZ Utils from source.
```

Either configure with --disable-lzma (which will make some CRAM files
produced elsewhere unreadable) or resolve this error to build HTSlib.

```
configure: error: ./configure failed for htslib-1.10
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ sudo apt-get install liblzma-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  grub-pc-bin libnuma1
Use 'sudo apt autoremove' to remove them.
Suggested packages:
  liblzma-doc
The following NEW packages will be installed:
  liblzma-dev
0 upgraded, 1 newly installed, 0 to remove and 24 not upgraded.
Need to get 145 kB of archives.
After this operation, 669 kB of additional disk space will be used.
Get:1 http://us-west1.gce.archive.ubuntu.com/ubuntu bionic/main amd64 liblzma-dev amd64 5.2.2-1.3 [145 kB]
Fetched 145 kB in 0s (8586 kB/s)
Selecting previously unselected package liblzma-dev:amd64.
(Reading database ... 70493 files and directories currently installed.)
Preparing to unpack .../liblzma-dev_5.2.2-1.3_amd64.deb ...
Unpacking liblzma-dev:amd64 (5.2.2-1.3) ...
Setting up liblzma-dev:amd64 (5.2.2-1.3) ...
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

samtools 설치

10) 다시 ./configure를 합니다.

이번엔 제발 되었음 좋겠습니다.

다음 그림 같이 FAIL, error 없이 나온다면

configure가 제대로 된 것 입니다.

만약 다른 dependency가 없어서 오류가 난다면

스스로 필요한 것을 설치하여 해결해봅시다.

11) 지금까지는 설치에 필요한 요소들을

체크하는 과정이었고 이제 정말 설치를 해보겠습니다.

그 전에 which make 를 타이핑 합니다.

오른쪽 그림과 같이 아무 결과가 없다면 make가

없는 것 입니다.

which 명령어는 프로그램이 리눅스 PATH 중 어디에 있는지

찾아서 알려주는 명령어입니다.

sudo apt-get install make 를 입력합니다.

설치 후 which make를 하였을 때 다음 그림처럼

나오면 제대로 make가 설치된 것 입니다.

```
checking how to run the C preprocessor... gcc -E
checking for egrep... /bin/grep -E
checking for ANSI C header files... yes
checking for sys/types.h... yes
checking for sys/stat.h... yes
checking for stdlib.h... yes
checking for string.h... yes
checking for memory.h... yes
checking for strings.h... yes
checking for inttypes.h... yes
checking for stdint.h... yes
checking for unistd.h... yes
checking for stdlib.h... (cached) yes
checking for unistd.h... (cached) yes
checking for sys/param.h... yes
checking for getpagesize... yes
checking for working mmap... yes
checking for gmtime_r... yes
checking for fsync... yes
checking for drand48... yes
checking whether fdatasync is declared... yes
checking for fdatasync... yes
checking for library containing log... -lm
checking for zlib.h... yes
checking for inflate in -lz... yes
checking for library containing recv... none required
checking for bzlib.h... yes
checking for BZ2_bzBuffToBuffCompress in -lbz2... yes
checking for lzma.h... yes
checking for lzma_easy_buffer_encode in -llzma... yes
checking for libdeflate.h... no
checking for libdeflate_deflate_compress in -ldeflate... no
checking for curl_easy_pause in -lcurl... no
checking for curl_easy_init in -curl... no
configure: WARNING: libcurl not enabled: library not found
configure: WARNING: GCS support not enabled: requires libcurl support
configure: WARNING: S3 support not enabled: requires libcurl support
checking whether PTHREAD_MUTEX_RECURSIVE is declared... yes
configure: creating ./config.status
config.status: creating config.mk
config.status: creating htplib.pc.tmp
config.status: creating config.h
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ 
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ which make
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ 
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ sudo apt-get install make
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  grub-pc-bin libnumctl
Use 'sudo apt autoremove' to remove them.
Suggested packages:
  make-doc
The following NEW packages will be installed:
  make
0 upgraded, 1 newly installed, 0 to remove and 24 not upgraded.
Need to get 154 kB of archives.
After this operation, 381 kB of additional disk space will be used.
Get:1 http://us-west1.gce.archive.ubuntu.com/ubuntu bionic/main amd64 make amd64 4.1-9.1ubuntu1 [154 kB]
Fetched 154 kB in 0s (9749 kB/s)
Selecting previously unselected package make.
(Reading database ... 70533 files and directories currently installed.)
Preparing to unpack .../make_4.1-9.1ubuntu1_amd64.deb ...
Unpacking make (4.1-9.1ubuntu1) ...
Setting up make (4.1-9.1ubuntu1) ...
Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ 
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ which make
/usr/bin/make
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ 
```

Samtools 설치

- 11) sudo make 를 타이핑 합니다.
- 12) sudo make install 을 타이핑 합니다.
- 13) which samtools 를 타이핑 합니다.
- 14) samtools 를 실행하였을 때
오른쪽 그림같이 나오면 제대로 된 것
입니다.

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ sudo make install
mkdir -p -m 755 /usr/local/bin /usr/local/bin /usr/local/share/man/man1
install -p samtools /usr/local/bin
install -p misc/ace2sam misc/maq2sam-long misc/maq2sam-short misc/md5fa misc/md5sum-lite misc/wgsim /usr/local/bin
install -p misc/blast2sam.pl misc/bowtie2sam.pl misc/export2sam.pl misc/interpolate.sam.pl misc/novo2sam.pl misc/pl
ot-bamstats misc/psl2sam.pl misc/sam2vcf.pl misc/samtools.pl misc/seq_cache_populate.pl misc/soap2sam.pl misc/wgsim
_eval.pl misc/zoom2sam.pl /usr/local/bin
install -p -m 644 doc/samtools*.1 misc/wgsim.1 /usr/local/share/man/man1
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$
```

```
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ which samtools
/usr/local/bin/samtools
kenneth_jh_han@instance-1:~/Downloads/samtools-1.10$ samtools

Program: samtools (Tools for alignments in the SAM format)
Version: 1.10 (using htllib 1.10)

Usage: samtools <command> [options]

Commands:
-- Indexing
dict          create a sequence dictionary file
faidx         index/extract FASTA
fqidx         index/extract FASTQ
index          index alignment

-- Editing
calmd         recalculate MD/NM tags and '=' bases
fixmate       fix mate information
reheader      replace BAM header
targetcut    cut fosmid regions (for fosmid pool only)
addreplacerg adds or replaces RG tags
markup        mark duplicates

-- File operations
collate       shuffle and group alignments by name
cat           concatenate BAMs
merge         merge sorted alignments
mpileup       multi-way pileup
sort          sort alignment file
split         splits a file by read group
quickcheck   quickly check if SAM/BAM/CRAM file appears intact
fastq         converts a BAM to a FASTQ
fasta         converts a BAM to a FASTA

-- Statistics
bedcov        read depth per BED region
coverage      alignment depth and percent coverage
```

FASTQ → BAM → VCF

BAM 파일 살펴보기

1) bam 파일을 받았던 경로로 가서,
samtools view SRR000982.mapped
를 입력합니다.

2) 오른쪽 아래 그림과 같이
나오게 됩니다.

3) 이번엔
samtools tview SRR000982.mapped.sorted.markdup.bam
를 입력합니다.

```
kenneth_jh_han@instance-1:~/Downloads$ ll
total 49124
drwxrwxr-x 3 kenneth_jh_han kenneth_jh_han 4096 May 21 09:39 .
drwxr-xr-x 6 kenneth_jh_han kenneth_jh_han 4096 May 21 09:34 ..
-rw-rw-r-- 1 kenneth_jh_han kenneth_jh_han 173146 May 21 09:33 SRR000982.filtered.variants.annotated.vcf
-rw-rw-r-- 1 kenneth_jh_han kenneth_jh_han 23967674 May 21 09:32 SRR000982.mapped.sorted.markdup.bam
-rw-rw-r-- 1 kenneth_jh_han kenneth_jh_han 3520184 May 21 09:33 SRR000982.mapped.sorted.markdup.bam.bai
-rw-rw-r-- 1 kenneth_jh_han kenneth_jh_han 8502024 May 21 09:31 SRR000982_1.filt.fastq.gz
-rw-rw-r-- 1 kenneth_jh_han kenneth_jh_han 9392196 May 21 09:32 SRR000982_2.filt.fastq.gz
drwxrwxr-x 9 kenneth_jh_han kenneth_jh_han 4096 May 21 10:13 samtools-1.10/
-rw-rw-r-- 1 kenneth_jh_han kenneth_jh_han 4721173 Dec 6 16:47 samtools-1.10.tar.bz2
kenneth_jh_han@instance-1:~/Downloads$ samtools view SRR000982.mapped.sorted.markdup.bam | less -S
kenneth_jh_han@instance-1:~/Downloads$ 
```

SRR000982.91192 115	chrM	9	60	102M	=	3300	3326	GTCTATCACCTATTAAACCACTCAGGGAGNTCTC
SRR000982.385325	1139	chrM	9	60	102M	=	3327	3326 GTCTATCACCTATTAAACGACTCGGG
SRR000982.271454	65	chrM	25	60	143M	=	14316	14292 ACCACTCACGGGAGCTCTCCATGCATT
SRR000982.125609	117	chrM	29	0	*	=	29	0 ACCCATATAACCCTCCCCCAAATTTC
SRR000982.125609	185	chrM	29	60	5M1D15M1D6M1D162M	=	29	0 CTCACGGAGCT
SRR000982.132204	177	chrM	63	60	4S10M1I129M	=	14215	14101 CGGTTCTGGGGGGTAGTGC
SRR000982.159836	131	chrM	65	60	82M2S	=	3045	2981 TGGGGGGTATGCACGGATAGCATTGCA
SRR000982.237678	65	chrM	67	60	113M	=	13393	13327 GGGGGTATGCACGGATAGCATTGCGA
SRR000982.469601	177	chrM	90	60	62M	=	13979	13962 ACGAGACGCTGGAGCCGGAGCACCTTA
SRR000982.124004	131	chrM	91	60	103M	=	2153	2063 CGAGACGCTGGAGCCGGAGCACCTAT
SRR000982.245083	131	chrM	97	60	33M	=	3131	3035 GCTGGAGCCGGAGCACCTATGTCGCA
SRR000982.11440 177	chrM	122	60	68M	=	14794	14726 CAGTATCTGTCITTGATTCCTGCTCATCCCATT	
SRR000982.192107	65	chrM	124	60	79M1D31M	=	14913	14790 GTATCTGCTTGATTCCT

BAM 파일 살펴보기

samtools tview 의 모습입니다.

c 키를 누르면 염기서열 별로

색을 입혀서 볼 수 있습니다.

방향키로 움직이면 보는 위치를
바꾸면서 볼 수 있습니다.

/ 키를 입력하면 특정 위치로 이동할 수 있는
입력 창이 나옵니다.

chrM:150 을 입력해봅시다.



```
1      11     21     31     41     51     61     71     81     91     101
N*NN
GTCTATCACCTATTAAACACTCACGGGAGCTCCATGCATTTGGTATTCGCTGGGGGT RIGCACCGGATAGCATTGCGAGACGCTGGAGCCGGAGCACCC
gtctatcacctattaaccactcacggag,tctccatgcatttgcattttcgctgggggt*atgcacgcgatagcattgcgagacgctggagccggagc
gtctatcacctattaaccactcacggag,tctccatgcatttgcattttcgctgggggt*atgcacgcgatagcattgcgagacgctggagccggagc
          ACCACTCACGGGAGCTCCATGCATTTGGTATTCGCTGGGGGT*ATGCACCGGATAGCATTGCGAGAC.CTGGAGCCGGAGCACCC
          ctcac*ggaggctccatgcattttcgctgggggt*atgcacgcgatagcattgcgagacgctggagccggagcc
          tctgggggttagtgccacgcgatagcattgcgagacgctggagccggagcc
          TGGGGGT*ATGCACCGGATAGCATTGCGAGACGCTGGAGCCGGAGCACCC
          GGGGT*ATGCACCGGATAGCATTGCGAGACGCTGGAGCCGGAGCACCC
          accagacgcgtggagccggagcc
          CGAGACCGCTGGAGCCGGAGCACCC
          GCTGGAGCCGGAGCACCC
```



```
1      11     21     31     41     51     61     71     81     91     101
N*NN
GTCTATCACCTATTAAACACTCACGGGAGCTCCATGCATTTGGTATTCGCTGGGGGT RIGCACCGGATAGCATTGCGAGACGCTGGAGCCGGAGCACCC
gtctatcacctattaaccactcacggag,tctccatgcatttgcattttcgctgggggt*atgcacgcgatagcattgcgagacgctggagccggagc
gtctatcacctattaaccactcacggag,tctccatgcatttgcattttcgctgggggt*atgcacgcgatagcattgcgagacgctggagccggagc
          ACCACTCACGGGAGCTCCATGCATTTGGTATTCGCTGGGGGT*ATGCACCGGATAGCATTGCGAGAC.CTGGAGCCGGAGCACCC
          ctcac*ggaggctccatgcattttcgctgggggt*atgcacgcgatagcattgcgagacgctggagccggagcc
          tctgggggttagtgccacgcgatagcattgcgagacgctggagccggagcc
          TGGGGGT*ATGCACCGGATAGCATTGCGAGACGCTGGAGCCGGAGCACCC
          GGGGT*ATGCACCGGATAGCATTGCGAGACGCTGGAGCCGGAGCACCC
          accagacgcgtggagccggagcc
          CGAGACCGCTGGAGCCGGAGCACCC
          GCTGGAGCCGGAGCACCC
```



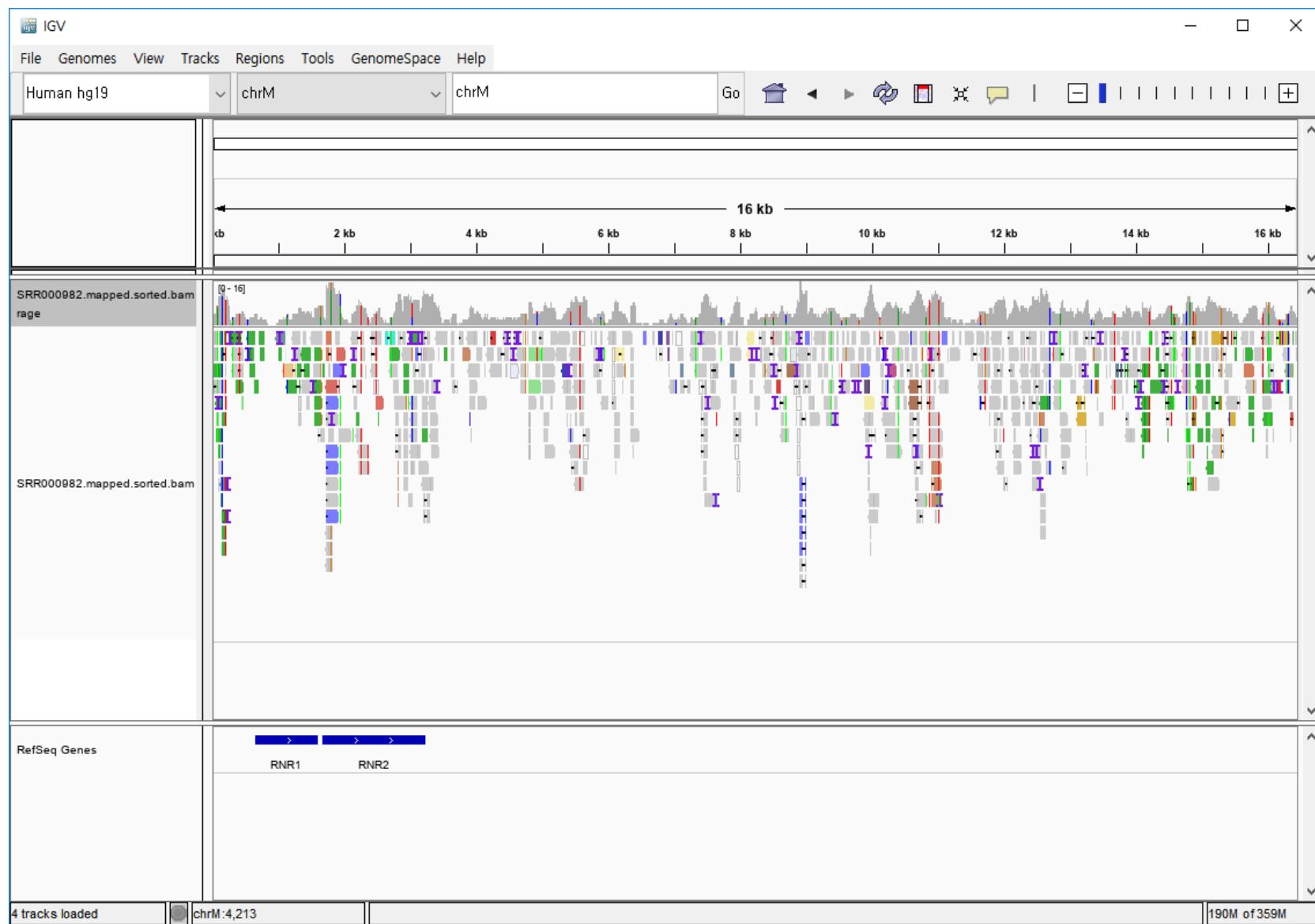
```
151      161      171      181      191      201      211      221      231      241      251
N*N*NNN N*NNN
CCCCATTATTATGCACCTACGTTCAATATTACAGGGAACATACTTACT AAAAGTGTAAATT AAT GCTTGAGGACATAATAAACATTGAATGTCTGCA CAG
CCCCATTATTATCGCACC
          tacttactaaagtgtgttaatttaatttaatttgtgttgcggacata
          gccccac
cccattatttatcgacccatcggtcaatattacaggcgaacatacttactaaagtgtgttgcggacataattttaa
cccattatttatcgacccatcggtcaatattacaggcgaacatacttactaa
          TGTTAATTAAATT AAT GCTTGAGGACATAATAAACATTGAATGTCT
CCCCATTATTATGCACCTACGTTCAATATTACAGGGAACATA
          t'aat'gtttgttgcggacataataaaacaaatttgaaatgtctgca*
ccattatttatcgacccatcggtcaatattacaggcgaacatacttactaaagtgtgttgcggacataataaaacaaatttgaaatgtctgca*
CCCCATTATTATGCACCTACGTTCAATATTACAGGGAACATACTTACT AAAA*TGTAAATT AAT GCTTGAGGACATA
cccattatttatcgacccatcggtcaatattacaggcgaacatacttactaaagtgtgttgcggacataataaaacaaatttgaaatgtctgca*
CCCCATTATTATGCACCTACGTTCAATATTACAGGGAACATACTTACT AAAA*TGTAAATT AAT GCTTGAGGACATA
cccattatttatcgacccatcggtcaatattacaggcgaacatacttactaaagtgtgttgcggacataataaaacaaatttgaaatgtctgca*
CCCCATTATTATGCACCTACGTTCAATATTACAGGGAACATACTTACT AAAA*TGTAAATT AAT GCTTGAGGACATA
cccattatttatcgacccatcggtcaatattacaggcgaacatacttactaaagtgtgttgcggacataataaaacaaatttgaaatgtctgca*
```

BAM 파일 살펴보기

실무에서 bam 파일을 보는 경우가
가끔 있는데, 리드들이 잘 쌓여있는지
기준 서열과 다른 변이들이
잘 있는지 보기 위해 살펴봅니다.

리드라고 하는 것은 시퀀서에서 한 번에 읽은 서열을 의미합니다.
염기서열들이 있으며 기준서열(reference genome)에
붙이면(mapping) 우리가 위에서 본 것과 같이 나옵니다.

1GV 실행



MarkDuplicate

TTTCATACTAACTAGCCTGCGGTCTGTGTTCCGACTTCTGAGTCATGGGGTTCAATGCCTATAGATTCT
.....C.
.....
.....T.....
.....C.....
.....
.....
.....C.....
..C.....

TTTCATACTAACTAGCCTGCGGTCTGTGTTCCGACTTCTGAGTCATGGGGTTCAATGCCTATAGATTCT
.....
.....T.....
.....C.....
.....C.....
.....C.....
.....C.....
.....
.....

MarkDuplicate

The screenshot shows a web browser window titled "Explain SAM Flags" at the URL <https://broadinstitute.github.io/picard/explain-flags.html>. The page is part of the Picard project, specifically the "Decoding SAM flags" section. At the top, there are download links for "Latest for Release", "Source Code ZIP File", "Source Code TAR Ball", and "View On GitHub". Below this, a sub-header "Decoding SAM flags" is displayed. A text block explains that the utility makes it easy to identify properties of a read based on its SAM flag value or vice versa. It instructs users to enter a SAM flag value (e.g., 1139) and click "Explain" to see a summary of properties. A "Switch to mode" link is also present. On the left, a "Find SAM flag by property" section lists various properties with checkboxes: "read paired" (checked), "read mapped in proper pair" (checked), "read unmapped" (unchecked), "mate unmapped" (unchecked), "read reverse strand" (checked), "mate reverse strand" (checked), "first in pair" (checked), "second in pair" (unchecked), "not primary alignment" (unchecked), "read fails platform/vendor quality checks" (unchecked), "read is PCR or optical duplicate" (checked), and "supplementary alignment" (unchecked). To the right, a "Summary" section lists the corresponding SAM flag values: "read paired (0x1)", "read mapped in proper pair (0x2)", "read reverse strand (0x10)", "mate reverse strand (0x20)", "first in pair (0x10)", and "read is PCR or optical duplicate (0x400)".

<https://broadinstitute.github.io/picard/explain-flags.html>

MarkDuplicate

```
$ samtools view
```

...

-f INT only include reads with all of the FLAGs
in INT present [0]

```
$ samtools view -f1024 SRR000982.mapped.sorted.markdup.bam | head -1
```

```
SRR000982.385325 1139 chrM 9 60 102M = 3327 3326
GTCTATCACCTATTAAACCACTCACGGGAGNTCTCCATGCATTGGTATTCGTCTGGGGGTATGCAC
GCGATAGCATTGCGAGACGCTGGAGGCCGGAGC
DDDDDDDDDDDDDDDDDDDDDDDDDDDB>>DB!BDDDDDDDDDDDDDDDDDDDDDDDDDD@ @@
@ @@EFFFFGHIIIIIIIIIDDDDDDDDDDDDDDDDD MD:Z:30C33G37 PG:Z:MarkDuplicates
RG:Z:test NM:i:2 AS:i:95 XS:i:51
```

Question

SAM과 BAM의 차이점은 무엇인가요?

BAM은 SAM의 binary 버전이다.
쉽게 말해 텍스트 파일인 SAM을 압축한 버전
크기가 1/10 정도로 줄어든다.



FASTQ → BAM → VCF



BED

BED

BED (Browser Extensible Data) format

구간을 나타내는 포맷

구간에 annotation 정보를 넣을 수 있다

```
browser position chr1:12081-12251
track name="Covered" description="Agilent SureSelect DNA - SureSelect Human All Exon V6 - Genomic regions covered by probes" color=0,0,128 db=hg19
chr1 12080 12251 ref|DDX11L1,ref|NR_046018,ens|ENST00000518655,ens|ENST00000450305,ens|ENST00000456328,ens|ENST00000515242
chr1 12595 12802 ref|DDX11L1,ref|NR_046018,ens|ENST00000518655,ens|ENST00000450305,ens|ENST00000456328,ens|ENST00000515242
chr1 13163 13658 ref|DDX11L1,ref|NR_046018,ens|ENST00000518655,ens|ENST00000450305,ens|ENST00000456328,ens|ENST00000515242
chr1 14620 15015 ref|WASH7P,ref|NR_024540,ens|ENST00000438504,ens|ENST00000538476,ens|ENST00000488147,ens|ENST00000541675,ens|ENST00000423562
chr1 15795 15914 ref|WASH7P,ref|NR_024540,ens|ENST00000538476,ens|ENST00000438504,ens|ENST00000488147,ens|ENST00000541675,ens|ENST00000423562
chr1 16743 17098 ref|WASH7P,ref|NR_024540,ens|ENST00000538476,ens|ENST00000438504,ens|ENST00000488147,ens|ENST00000541675,ens|ENST00000423562
...
chrY 59358000 59358120 -
chrY 59359070 59359565 ref|DDX11L16,ref|NR_110561
chrY 59359925 59360132 ref|DDX11L16,ref|NR_110561
chrY 59360476 59360647 ref|DDX11L16,ref|NR_110561
```

BED

Name	Length
SureSelect_V5	50,390,601 bp (50Mbp)
SureSelect_V5+UTR	74,569,526 bp (74Mbp)
SureSelect_V6	60,456,963 bp (60Mbp)

FASTQ → BAM → VCF

VCF 파일 살펴보기

Meta-information Line

Header Data Line

FORMAT column

tag	meaning
GT	Position에 대한 샘플의 Genotype 0/0 - homozygous reference를 의미 0/1 - heterozygous REF/ALT를 의미 1/1 - homozygous ALT를 의미
AD	Allele Depth
DP	Depth

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER<ID=q10,Description="Quality below 10">
##FILTER<ID=s50,Description="Less than 50% of samples have data">
##FORMAT<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT<ID=QQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS .
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1|1:43:5:...
20 1110696 rs6040355 G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

VCF 파일은 기준서열과 다른 변이들을 테이블 형식으로 표현한 파일입니다.

첫 번째 (#) 두 개인 ##에는 VCF 파일을 생성할 때 사용한 옵션 등의 설명이 있고 한 개인 #에는 각 컬럼의 타이틀이 붙어있습니다.

9개의 의무적 컬럼이 있고 10번째 부터는 샘플의 FORMAT에 해당 값이 있습니다.

VCF 파일 살펴보기

SNP (Single Nucleotide Polymorphism)
: 기준서열과 비교하여 하나의 서열이 바뀐것을 의미

기준 서열	ACAA <ins>G</ins> GTT
Read	ACAT T GGTT

Insertion
: 기준서열과 비교하여 서열이 추가된 것을 의미

기준 서열	ACAA <ins>A</ins> -GGTT
Read	ACAA A GGTT

Deletion
: 기준서열과 비교하여 서열이 제거된 것을 의미

기준 서열	ACAA <ins>G</ins> GTT
Read	ACA*GGTT

변이라고 하는 것은 기준서열과 일치하지 않는 염기를 의미합니다.

VCF 파일 살펴보기

AD (Allele Depth)는 GT (Genotype)를 기준으로 표현하는데,

CHROM	POS	REF	ALT	FORMAT	SAMPLE
chr20	1234	A	T	GT:AD:DP	0/1:23,11:34

Variant caller 중 하나인 GATK 를 기준으로 설명하자면
AD 에서 첫 번째는 GT의 0, 두 번째는 GT의 1, 세번째는 GT의 2 번째 를
의미합니다.

예를 들어 chr20:1234 A → T 에서, A는 GT의 0, T는 GT의 1입니다.
AD는 23,11 로 써져있는데 순서대로 0인 A가 → 23, 1인 T가 → 11 만큼의
depth를 나타냅니다.

VCF 파일 살펴보기

CHROM	POS	REF	ALT	FORMAT	SAMPLE
chr1	1234	A	T	GT:AD:DP	0/1:23,11:34
chr3	2222	TC	T,TCC	GT:AD:DP	1/2:2,31,21:5 4
chr21	2830	T	G	GT:AD:DP	1/1:0,44:44

Chr1: 1234 위치의 쌍인 ReadDepth는 34다. (O/X)

Chr1:1234에서 A의 개수는 23개 T의 개수는 11개다. (O/X)

Chr3:2222 위치의 변이 종류는 SNP, Insertion, Deletion이다. (O/X)

Chr3:2222 위치의 REF Depth는 31, ALT Depth는 21이다. (O/X)

정답은 다음 슬라이드에

VCF 파일 살펴보기

CHROM	POS	REF	ALT	FORMAT	SAMPLE
chr1	1234	A	T	GT:AD:DP	0/1:23,11:34
chr3	2222	TC	T,TCC	GT:AD:DP	1/2:2,31,21:5 4
chr21	2830	T	G	GT:AD:DP	1/1:0,44:44

Chr1: 1234 위치의 쌍인 ReadDepth는 34다. (O)

Chr1:1234에서 A의 개수는 23개 T의 개수는 11개다. (O)

Chr3:2222 위치의 변이 종류는 SNP, Insertion, Deletion이다. (X)
Insertion (TC → TCC) 과 Deletion (TC → T) 만 있고 SNP는 없다.

Chr3:2222 위치의 REF Depth는 31, ALT Depth는 21이다. (X)
GT에서 TC:0, T:1, TCC:2로 볼 수 있는데,
AD에서 2,31,21에서 TC,T,TCC의 순서로 일치하기에
REF인 TC는 2, ALT1인 T는 31, ALT2인 TCC는 21이다.

VCF 파일 살펴보기

```
kenneth_jh_han@instance-1:~/Downloads$ python3 calc_snp_indel.py SRR000982.filtered.variants.annotated.vcf  
364 10 24  
kenneth_jh_han@instance-1:~/Downloads$
```

VCF 파일에서 SNP, Insertion, Deletion 의 숫자를 세어보세요.

ALT 컬럼에서 쉼표로 구분되어 alt가 여러개 있는 경우도 있으니 이 점 참고하여 진행해주세요.

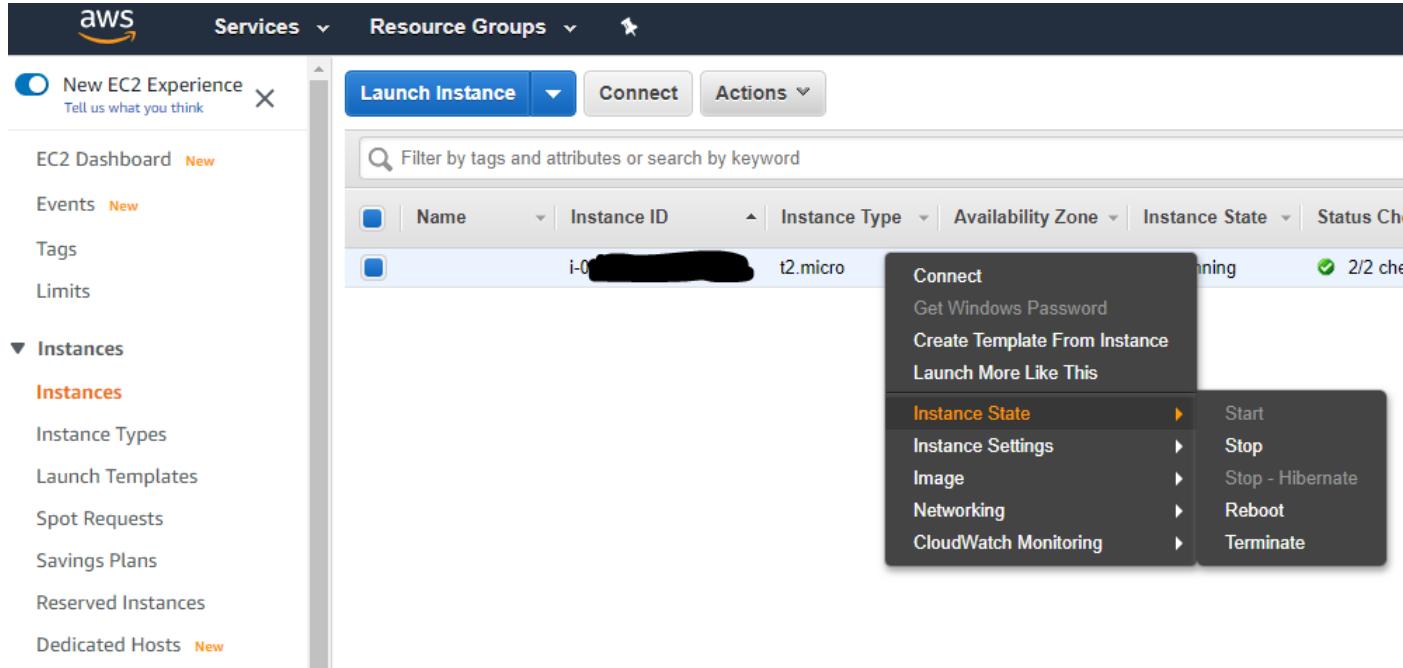
답은 그림과 같이 나옵니다.

실습 시간!

<https://github.com/kennethJHan/genome-analysis-tutorial>

EC2 삭제하기

EC2 삭제하기



EC2 Instances에서, 삭제하고 싶은 인스턴스에서 우 클릭 후

Instance State -> Stop

Instance State 가 Stop 이 되면,

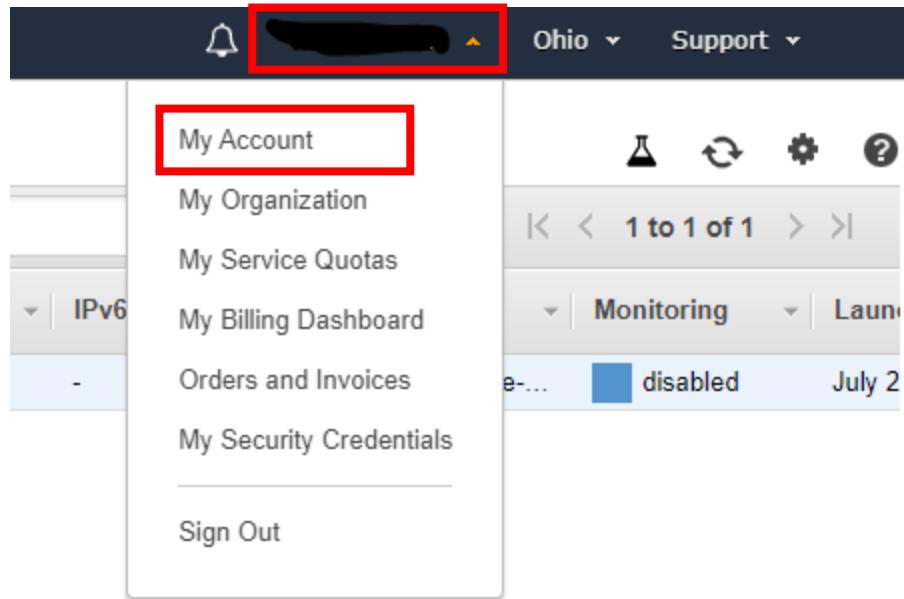
Instance State -> Terminate 를 합니다.

Instance 에 함께 엮인 스토리지 (EBS) 가 함께 삭제되도록 합니다.

(남아있으면 사용하지 않아도 과금이 되므로 주의!)

AWS 계정 삭제하기

계정 삭제하기



우측 상단에 계정 이름을 클릭
My Account 를 클릭 합니다.

This screenshot shows the 'Close Account' section of the AWS GovCloud (US) console. It includes a 'Sign up for AWS GovCloud (US)' button and several checkboxes for account closure terms. At the bottom is a large red 'Close Account' button.

Close Account

I understand that by clicking this checkbox, I am closing my AWS account. The closure of my AWS account serves as notice to AWS that I wish to terminate the AWS Customer Agreement or any other agreement with AWS that governs my AWS account, solely with respect to that AWS account.

Monthly usage of certain AWS services is calculated and billed at the beginning of the following month. If I have used these types of services this month, then at the beginning of next month I will receive a bill for usage that occurred prior to termination of my account. In addition, if I have any active subscriptions (such as a Reserved Instance for which I have elected to pay in monthly installments), then even after my account is closed I may continue to be billed for the subscription until the subscription expires or is sold in accordance with the terms governing the subscription.

I acknowledge that I may reopen my AWS account only within 90 days of my account closure (the "Post-Closure Period"). If I reopen my account during the Post-Closure Period, I may be charged for any AWS services that were not terminated before I closed my account. If I reopen my AWS account, I agree that the same terms will govern my access to and use of AWS services through my reopened AWS account.

If I choose not to reopen my account after the Post-Closure Period, any content remaining in my AWS account will be deleted. For more information, please see the [Amazon Web Services Account Closure page](#).

I understand that after the Post-Closure Period I will no longer be able to reopen my closed account.

I understand that after the Post-Closure Period I will no longer be able to access the Billing Console to download past bills and tax invoices.
If you wish to [download any statements you can do so here](#). Select the month and expand the summary section to download the payment invoices and/or tax documents.

I understand that after the Post-Closure Period I will not be able to create a new AWS account with the email address currently associated with this account.
If you wish to update your e-mail address, [follow the directions here](#).

Close Account

연결되는 페이지에서 스크롤을 가장 아래로 내리면 Close Account 가 있습니다.
모두 체크하고 Close Account 를 클릭하고,
다시 한 번 Close Account 를 클릭하면 계정이 삭제됩니다.
주의할 점은 계정을 닫으면 다시는 사용했던 이메일로
재 가입이 되지 않습니다.

강사소개

강사 소개	
이름	한주현
소속	3billion, Bioinformatics Engineer Seoul National University, Medical Informatics
메일	kenneth.jh.han@snu.ac.kr
주요 업무	Human Genome Analysis (WGS, WES) Rare Disease Analysis Bioinformatics Algorithms Analysis Pipeline / Platform Development Full Stack Development Cloud Computing
주 언어	Python, JAVA, JavaScript, Bash shell
저서	니콜라스 볼커 이야기 (2016.10, 금창원 외 공역) 바이오파이썬으로 만나는 생물정보학 (2019.03, 한주현)
웹 페이지	https://korbillgates.tistory.com (블로그)

강의 내용 또는 생물정보학, 취업 및 진로에 관련하여
궁금한점이 있으시면 언제든 메일로 문의해주세요