

Introducción a la Genómica en la Nube con Google Cloud

Dr. Matthieu J. Miossec (@RealMattJM)

Bioinformatics analyst @ Wellcome Centre for human genetics



UNIVERSITY OF
OXFORD

Programa Unidad 9

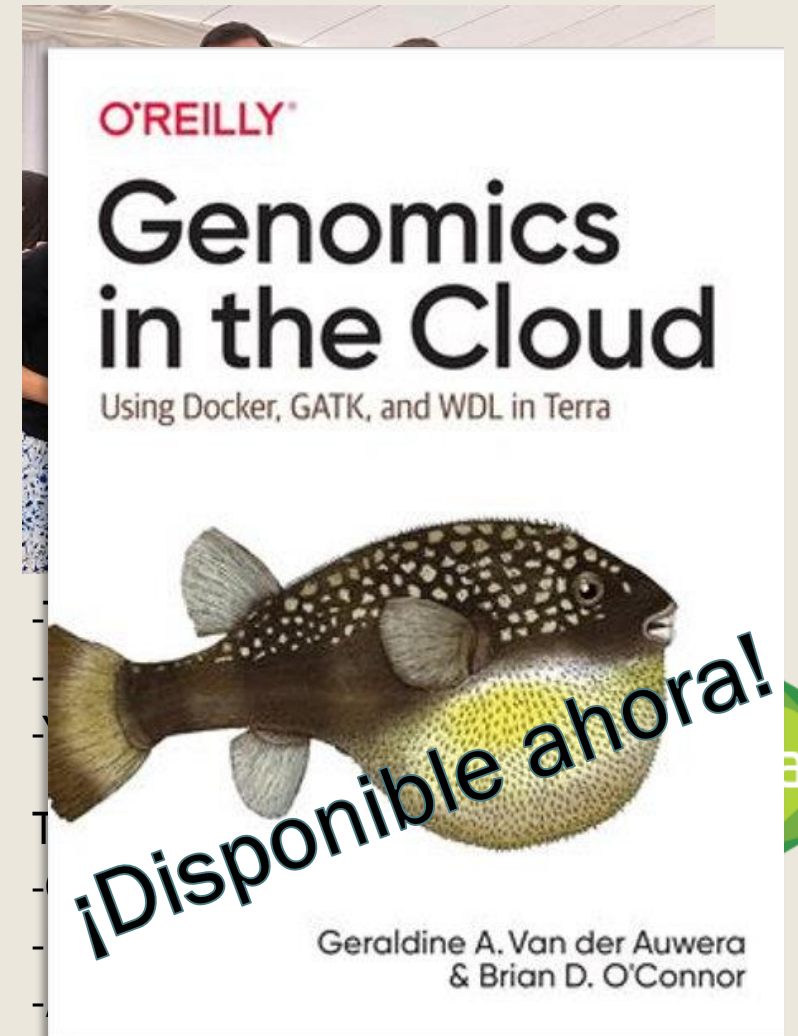
- 17 de mayo (lunes) – Introducción a la genómica en la nube con Terra.
- 19 de mayo (miércoles) – Introducción a la plataforma Terra
- 24 de mayo (lunes) – Otras herramientas en Terra

Agradecimientos

- Data Sciences Platform  **BROAD**
INSTITUTE
Broad Institute of Harvard and MIT
<https://gatk.broadinstitute.org/>

Por los materiales (Terra, workflows, docs...), gráficos y el apoyo otorgado durante la preparación de la clase de 2019.

Agradecimiento especiales,
al equipo de Viña del Mar (Nov 2018)



-Kate Noblett

¿Por qué trabajar en la nube?

- Trabajar y compartir muchos datos (públicos como privados) en línea sin deber almacenarlos localmente.
- Facilita colaboraciones internacionales, se comparte más fácilmente.
- Alternativa a un HPC local difícil de acceso o sobreocupado o inadecuado.
- Permite probar nuevas herramientas sin preocuparse tanto de la instalación (no necesita reinventar la rueda) → Docker
- Hace que todo un estudio sea reproducible (y accesible).

Genómica en la nube



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



<https://azure.microsoft.com/en-us/services/genomics/>



Google Cloud

<https://cloud.google.com/life-sciences>

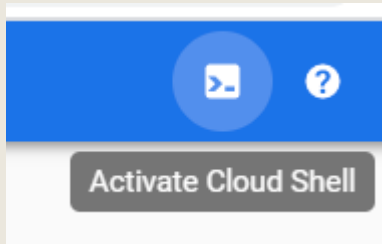




Google Cloud Platform (GCP)

- Cada nuevo usuario puede aprovechar de **\$300 (US)** en créditos nube que pueden ser usados durante **90 días***.
(pero necesita ingresar una tarjeta de crédito)
- Algunos servicios quedan gratuitos bajo una cierta limite, detalles en: <https://cloud.google.com/free>
- Para la duración de esa unidad, van a usar la GCP en un proyecto vinculado al billing 'Bioinfo2021'.

*La oferta puede haber cambiado.



El Cloud Shell

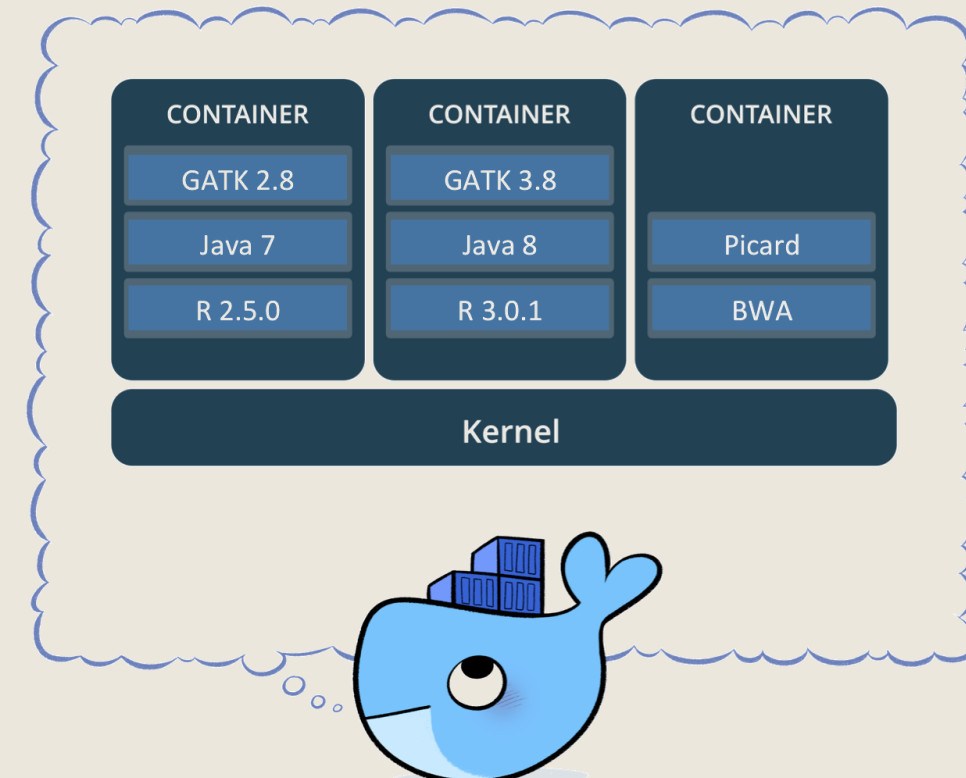
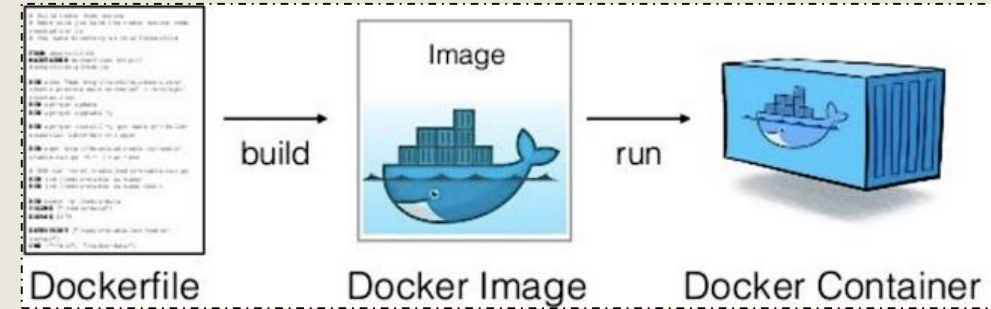
- El Cloud Shell nos da acceso a la primera maquina virtual que vamos a usar en la nube
 - Fácil de acceso, no necesita configuración, siempre gratis.
 - Linux Debian, 5Gb para almacenamiento, paquetes básicos preinstalado; incluso el Google Cloud SDK (gcloud).
- Ideal para pequeñas tareas o pruebas y administración de la cuenta. Limitado para actividades más computacionalmente intensivas como genómica.

Google Storage Utilities → gsutil

- Nos permite interactuar con el Google Cloud Storage (GCS), un espacio de almacenamiento de archivos en la nube.
 - Unidades de almacenamiento se llaman 'buckets' (~cubos).
 - Son accesible a través de una dirección empezando con **gs://**
- Comandos basados en las de UNIX (ej. ls, cat, cp...)
- Existe una versión Desktop para subir/descargar archivos a/desde la nube.

Docker

- Container: Similar a una maquina virtual.
 - Crea una ‘imagen’ de un sistema operativo (versión de Linux) como base.
+ los programas necesarios para ejecutar un conjunto de tareas predefinidas.
- Posible ejecutar varios ‘containers’ en la misma maquina si existen incompatibilidades entre sistemas.



¡Apagar (o borrar) la maquina virtual!

VM instances

CREATE INSTANCE

IMPORT VM

REFRESH

START/RESUME

STOP

OPERATIONS

INSTANCES

INSTANCE SCHEDULE

Filter

Enter property name or value

Name ↑	Zone	Recommendations	In use by	Internal IP	External IP	Connect
instance-1	us-central1-a			10.128.0.3 (nic0)	34.134.182.225	SSH

Start/Resume

Stop

Suspend

Reset

Delete

View network details

Create new machine image

View logs

View monitoring

Please select

