



## Why of cloud computing

### Overview

**Teaching:** 5 min

**Exercises:** 0 min

#### Questions

- What is cloud computing?
- What are the tradeoffs of cloud computing?

#### Objectives

- Understand benefits of working on a remote computer system

There are a number of reasons why accessing a remote machine is invaluable to any scientists working with large datasets. In the early history of computing, working on a remote machine was standard practice - computers were bulky and expensive. Today we work on laptops or desktops that are more powerful than the sum of the world's computing capacity 20 years ago, but many analyses (especially in genomics) are too large to run on these laptops/desktops. These analyses require larger machines, often several of them linked together, where remote access is the only practical solution.

In recent years, computational power has become a commodity and entire companies have been built around a business model that allows you to “rent” one or more linked computers for as long as you require, at relatively low cost. This is the basic principle behind the **cloud**. You define your computational requirements and off you go.

You'll know you need to start working on the cloud when:

- Your computer does not have enough resources to run the desired analysis (memory, processors, disk space, network bandwidth).
- Your computer is taking hours or days to get through an analysis.
- You cannot install software on your computer (application does not have support for your operating system, conflicts with other existing applications, etc.)
- You cannot afford the infrastructure (hardware, bandwidth, power supply, administrator) required to acquire and maintain sufficiently large computers for your analysis.
- You need to continue working on your analyses while you travel.

The cloud is a part of our everyday life (e.g. using Amazon, Google, Netflix, or an ATM involves remote computing). The topic is fascinating, but this lesson says ‘5 minutes or less’ so let's get connected.

## Choosing a cloud platform

The most important thing about the **cloud** is choice - instead of purchasing a physical computer, you can obtain on-demand computing at almost any scale. This power comes with advantages and disadvantages:

## Advantages of Cloud Computing

- Access large amounts of computing power on demand
- Full administrative rights - install anything
- Use pre-configured “images” (machine snapshot where operating system and software are already installed)
- Your local operating system doesn't matter - once you connect to the cloud you can run any UNIX software

## Disadvantages of Cloud Computing

- It takes time to upload data and download results
- Cloud computing costs money (you must keep track of your costs)
- If you need help, you may not have a local system administrator
- Images may be poorly documented (you may not be clear on what is installed, or how to use it)
- Form of payment (credit card)\*
- Understanding of Amazon's billing and payment (See: Getting started with AWS Billing and Cost Management (<https://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/billing-getting-started.html>))

\* You can use some of Amazon Web Services (AWS) for free, or see if you qualify for AWS Cloud Credits for Research (See: <https://aws.amazon.com/research-credits/> (<https://aws.amazon.com/research-credits/>)). The free level of service *will not* be sufficient for working with the amount of data we are using for our lessons.

## Cloud platform choices

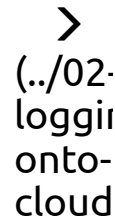
There are several cloud providers to choose from. Some scientific clouds may either be free or allocate resources competitively. Commercial clouds can be very powerful, but choice can be overwhelming. We will come back to this at the end of the lesson, but for right now, we're going to continue using the Amazon Web Services as our resource.

Learn more about cloud computing in bioinformatics

Fusaro VA, Patil P, Gafni E, Wall DP, Tonellato PJ (2011) **Biomedical Cloud Computing With Amazon Web Services**. PLoS Comput Biol 7(8): e1002147. doi: 10.1371/journal.pcbi.1002147

### Key Points

- Cloud computing increases processing speed and efficiency



Licensed under CC-BY 4.0 () 2018–2019 by The Carpentries (<https://carpentries.org/>)  
 Licensed under CC-BY 4.0 () 2016–2018 by Data Carpentry (<http://datacarpentry.org>)

Edit on GitHub ([https://github.com/datacarpentry/cloud-genomics/edit/gh-pages/\\_episodes/01-why-cloud-computing.md](https://github.com/datacarpentry/cloud-genomics/edit/gh-pages/_episodes/01-why-cloud-computing.md)) /  
 Contributing (<https://github.com/datacarpentry/cloud-genomics/blob/gh-pages/CONTRIBUTING.md>) / Source  
 (<https://github.com/datacarpentry/cloud-genomics/>) / Cite (<https://github.com/datacarpentry/cloud-genomics/blob/gh-pages/CITATION>) / Contact (<mailto:team@carpentries.org>)

Using The Carpentries theme (<https://github.com/carpentries/carpentries-theme/>) — Site last built on: 2019-08-27 17:14:54 +0000.