R jobs in the heavens

Running R jobs at large scale (remotely)

- A good point: Collaboration and sharing codes and analysis
- Local clusters... you may have access to one in you institution, but
  - high cost for maintenance and administration...
  - it must be accessible to all researchers... independently from subject
  - young guns and basic sciences are not very appealing to deserve access...

Cloud possibilities

Amazon Web Services - AWS

- Getting started with R on Amazon Web Services - https://aws.amazon.com/blogs/opensource/getting-started-with-r-on-amazon-web-services/
- Running an R Code on AWS Batch - https://medium.com/geekculture/running-an-r-container-on-aws-batch-on-production-9be336c34f95
  - Jobs are submitted via Docker files with help of `packrat` `R` package
  - cost? only free trial?

Azure - Microsoft

- R workloads on Azure Batch - https://azure.microsoft.com/pt-br/blog/r-workloads-on-azure-batch/
- Tutorial: Run a parallel workload with Azure Batch using the .NET API - https://docs.microsoft.com/en-us/azure/batch/tutorial-parallel-dotnet
  - paid if you have an institutional account?

Google Cloud Platform

- Running R at Scale on Compute Engine - https://cloud.google.com/architecture/running-r-at-scale
  - If not eligible for a free trial, "... a 6-node cluster composed of n1-standard-4 instances, would cost $1.84/hr ..."

Oracle cloud computing

- Using Oracle R Enterprise Embedded R Execution
  - free credits for new accounts... better performance? lower costs?

R possibilities

R Cloud - AT&T Labs

- Try It, Online or Locally - https://rcloud.social/tryit/index.html
  - support to different languages in the same code
  - everything code abd abakysis is public

Rstudio Server or connect???

- From "What is the difference between rstudio-connect and rstudio-pro" - https://community.rstudio.com/t/what-is-the-difference-between-rstudio-connect-and-rstudio-pro/43949
  - RStudio Server Pro - Very similar to the Desktop IDE, but runs on a server. This is very helpful for a few reasons: it allows you to have more compute resources closer to your data, it provides a uniform environment for teams that makes it easier to collaborate, and it provides controls for administrators to monitor and scale work.
  - RStudio Connect - Makes it easy to share R Markdown reports, deploy shiny web applications, and deploy APIs written in R. Connect can also be used to share Jupyter Notebooks. Whereas GitHub stores your code statically, RStudio Connect knows how to run it, which means your work can be accessible to non R users. For example, a user can visit a shiny application on RStudio Connect and interact with the app without ever knowing R code is involved! Likewise, you can have RStudio Connect run reports on a schedule and send emails with the results.
- R Studio Server on Google Cloud - https://towardsdatascience.com/r-studio-server-on-google-cloud-dd69b8bff80b
- Getting Started with RStudio Connect for GCP - https://support.rstudio.com/hc/en-us/articles/360033988434-Getting-Started-with-RStudio-Connect-for-GCP

Docker

- all starts with Docker???
- primary focused on reproducibility??? ... but also have your code running in any cloud service

Container it! Rstudio with all you need in a docker container and deploy in a virtual lmachine... you dont need to instal everytinhg again and again manually... and have your code "forever"

- You have to setup and install everything for the first time, start and stop the job whenever you want.
  - **Even if you stop the VM if there is files in the storage will be charged by the service**

"the easy way" to get started quickly

Rstudio on google cloud... you can use it all iteratively, but has to install everytime you setup a virtual machine...

https://support.rstudio.com/hc/en-us/articles/115010260627-Getting-Started-with-RStudio-Workbench-RStudio-Server-Pro-Standard-for-GCP