

Running an R Script on a Schedule: Docker Containers on gitlab

Posted on September 24, 2020 by Roel M. Hogervorst in R bloggers | 0 Comments

[This article was first published on [Category R on Roel's R-tefacts](#), and kindly contributed to [R-bloggers](#). (You can report issue about the content on this page [here](#))

Want to share your content on R-bloggers? [click here](#) if you have a blog, or [here](#) if you don't.

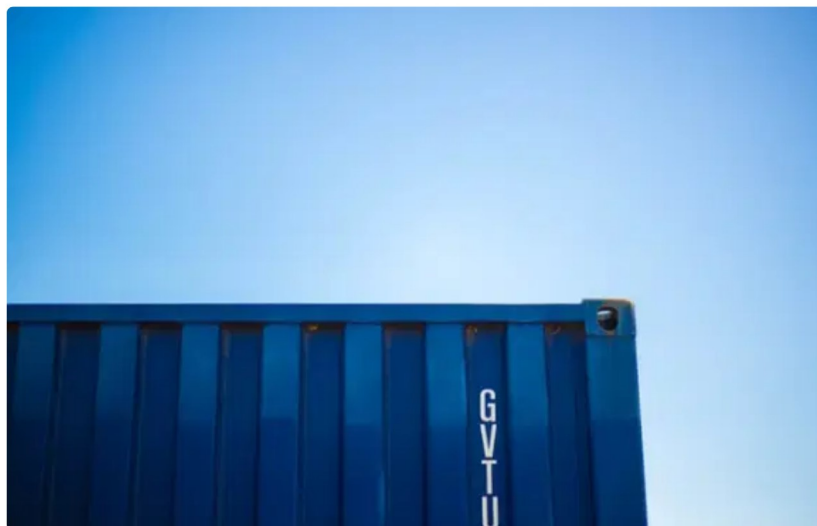
 Share

 Tweet

In this tutorial/howto I show you how to run a docker container on a schedule on gitlab.

Docker containers are awesome because, once made, they run everywhere! It does not matter what type of computer^ [Though I believe there is a problem with ARM based vs other CPU's]. you have. Once I build a container you can run my container on a linux box, windows machine or mac. This is also why people love containers for production, you can finally truly pick up a container from development and hand it over to production.

Thanks to the massive work by [the rocker team](#) we have containers ready that 'just work' with R, there are even containers with Rstudio installed!



My docker file

I'm using the versioned container `rocker/r-ver` set to the latest R version as of now `4.0.2`. I could set it to `latest` but that would mean the container could just break when those containers are refreshed. Keeping this fixed version saves us a lot of headache later on.

For building of packages though, it is super useful to always test against the latest versions!

For how to set up a docker file I refer you to the excellent tutorials '[An introduction to Docker for R users](#)' by Colin Fay, and the [larger tutorial](#) by ROpensci.

My [docker file](#) has 3 steps

Search R-bloggers..

Go

Your e-mail here

Subscribe

52793 readers

 Follow @rbloggers

92.5K followers



R bloggers

 Like Page

Most viewed posts (weekly)

Complexity is a source of income in open source ecosystems
Little useless-useful R functions – Looping through variable names and generating plots
April 2021: "Top 40" New CRAN Packages
Build and improve a Machine Learning Classification model with TidyModels and R
How to find dataset differences in R Quickly
Compare Datasets
Correlation in R (NA friendliness, accepting matrix as input data, returning p values, visualization and Pearson vs Spearman)
Stringr in r 10 data manipulation Tips and Tricks

Sponsors

EARL Conference online
tickets now on sale

6-10 September, 2021



Learn R by doing

 datacamp

Start For Free

- 1. take the basic container and update it

```
FROM rocker/r-ver:4.0.2
# before step (in gitlab)
# - update, and set for maximum cpu use
# - make a renv folder and install renv
RUN apt-get update
RUN echo "options(Ncpus = $(nproc --all))" >> /usr/local/lib/R/etc/Rprofile.site
```

- 2. setup renv

```
RUN mkdir -p ~/.local/share/renv
RUN R -e 'install.packages("renv")'
```

- 3. user settings: install the systems libraries, copy files to the container and install all the necessary packages from the lockfile

```
# user settings
# - install the systems libraries
# - copy script and lock file
RUN apt-get install -y --no-install-recommends libcurl4-openssl-dev libssl-dev
COPY run_job.R run_job.R
COPY renv.lock renv.lock
# I found that renv::restore does not use the super fast
# rstudio package manager, and so by pre installing rtweet and ggplot2
# and all their dependencies we get way faster building speed
RUN R -e 'install.packages(c("ggplot2", "rtweet"))'
RUN R -e 'renv::restore()'
# on running of the container this is called:
CMD Rscript run_job.R
```

Build the container with `docker build -t <name_for_container> .`

Test it out using `docker run --env-file .Renviron <name_for_container>`

Using gitlab

gitlab has its own docker registry where you can push to. see the [gitlab docs](#) for more info.

You have to set up some authentication with docker first, but after that you can do a docker build and push it to a registry on gitlab. This is super useful for many things. We could build a container with everything we want already installed and use that to test our new code!

But we can take it even one step further. We can make gitlab create the container, save it and run it!

The `.gitlab-ci.yml` file is quite clean and was super easy to modify from the [examples given here](#). Gitlab has amazing documentation!

The configuration has 2 stages: build and test.

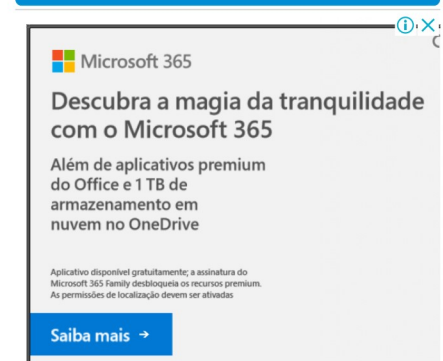
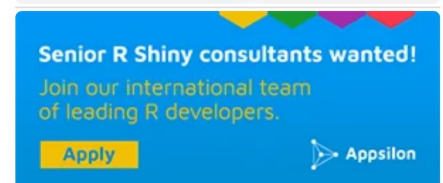
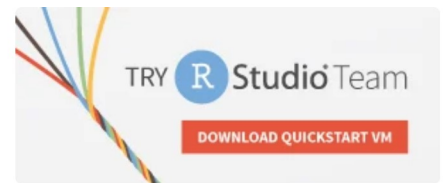
- Build** makes the container and pushes it to the local registry
- Test** pulls the container from the registry and runs the container

It is waaay faster than my previous approach. A nice effect of having two stages in this file is that if the later stage fails you can rerun that stage without rerunning the first part.

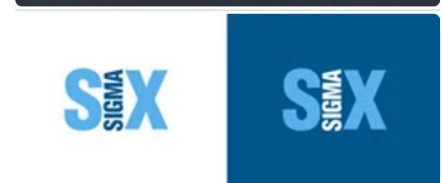
Roel Hogervorst > dockerized_script > Jobs

Status	Job	Pipeline	Stage	Name	Timing
failed	#755536629 Y main -> 433cf958	#194039179 by [user]	test	test	00:00:55 36 minutes ago
passed	#755536599 Y main -> 433cf958	#194039179 by [user]	build	build	00:03:30 37 minutes ago
failed	#755516652 Y main -> de5568de	#194034546 by [user]	build	build	00:00:06 49 minutes ago

You can schedule it again like [in this gitlab post](#) go to CI/CD – schedules and make a



Managed RStudio Infrastructure



schedule

It is also possible to push docker containers to github, GCP or other cloud providers. Some of which I will explore in the future.

That is all for today!

References

- Find more tutorials by me in [this tutorial overview page](#)
- Find all posts about scheduling an R script [here](#)
- gitlab [container registry docs](#)
- Photo by [Victoire Joncheray](#) on [Unsplash](#)

Reproducibility

- At the moment of creation (when I knitted this document) this was the state of my machine: ****click here to expand****

Related

[Running an R Script on a Schedule: Gitlab](#)

In this tutorial I have an R script that creates a plot and tweets it, it runs every day on gitlab runners. The use case is this: You have a September 23, 2020 In "R bloggers"

[Running an R Script on a Schedule: Azure Functions \(Serverless\)](#)

In this post I will show how I run an R script on a schedule, by making use of 'serverless' computing service on the Microsoft January 21, 2021 In "R bloggers"

[Investigating Docker and R](#)

This post is regularly updated (cf. GH issue) and available under the URL [http](#)

December 14, 2016 In "R bloggers"

f Share

🐦 Tweet

To **leave a comment** for the author, please follow the link and comment on their blog: [Category R on Roel's R-tefacts.](#)

[R-bloggers.com](#) offers **daily e-mail updates** about R news and tutorials about [learning R](#) and many other topics. [Click here if you're looking to post or find an R/data-science job.](#)

Want to share your content on R-bloggers? [click here](#) if you have a blog, or [here](#) if you don't.

← Previous post

Next post →

EXPLORE R IN THE OC

Learn more about our M.S. and Ph.D. programs in Computational and Data Sciences



CHAPMAN UNIVERSITY

esR!
5-9 JULY 2021



Our ads respect your privacy. Read our [Privacy Policy](#) page to learn more.

Contact us if you wish to help support R-bloggers, and place **your banner here**.

Recent Posts

more air for MCMC
Tips And Tricks For Data Scientists Vol.8
GooglyPlusPlus2021 is now fully interactive!!!
Mario Kart 64 World Records
Bioconductor Asia Membership Increasing Due to Going Virtual
Working with files and folders in R-Ultimate Guide
Partial dependence plots with tidymodels and DALEX for #TidyTuesday Mario Kart world records
A comparison of terra and raster packages Using {pagedown} in Docker
A forecasting tool (API) with examples in curl, R, Python
R-Ladies Philly – Building our Online Community During the Pandemic
Correlation in R (NA friendliness, accepting matrix as input data, returning p values, visualization and Pearson vs Spearman)

Best Practices for R with Docker
Learning R: Creating Truth Tables
datatable editor-DT package in R

Jobs for R-users


Junior Data Scientist / Quantitative economist
Senior Quantitative Analyst
R programmer
Data Scientist – CGIAR Excellence in
Agronomy (Ref No: DDG-R4D/DS/1/CG/EA
/06/20)
Data Analytics Auditor, Future of Audit Lead
@ London or Newcastle

python-bloggers.com (python/data-science news)

Interview Questions Vol.1
How To Get A Sentiment Score For Words
with Naive Bayes
A forecasting tool (API) with examples in curl,
R, Python
Naive Bayes Classification in NLP tasks from
Scratch
Market Basket Analysis and Association Rules
from Scratch
Cryptocurrency Trading Strategy by Detecting
the Leaders and the Followers
How to Find Similar Documents using
N-grams and Word Embeddings

Full list of contributing R-bloggers

Archives

Select Month 

Other sites

SAS blogs
Jobs for R-users