
Analytic
Health.

R in Production?

What R we going to talk about?

1. Introduction

2. What is ‘production’?

3. A look into our production process
    1. Data gathering and data processing
    2. Logging
    3. Storage
    4. Building products
    5. Deployment
    6. Monitoring

4. Our tips

5. Resources

2011: Bsc Business Economics
2015: Msc Data Science

Tilburg University

Eneco
2016: Data Scientist

Dr.Reddy's
2018: Data Scientist

HYPEBRIGHT
2019: Data Science consultant

2021: Managing Director & Head of Data Science

Analytic Health.

Shiny JS VUE R

Analytic Health – who are we?

We develop intelligent and accessible technology which gives organisations the tools they need to accelerate innovation in healthcare.

Pharmly Analytics

Pharmly Cloud Data

https://analytichealth.co.uk

Veerle van Leemput
Managing Director &
Head of Data Science

Greg Mills
Managing Director &
Founder

Jana Kopecna
Head of Operations

What is Production?

Production

Your code is being **used** in a real-world situation, where people **rely upon** your code to function properly. If something goes wrong, this has **real life impact**.

Those people can be **users** and **developers**

+

Organizing your environment: how we make sure people can rely upon on our code

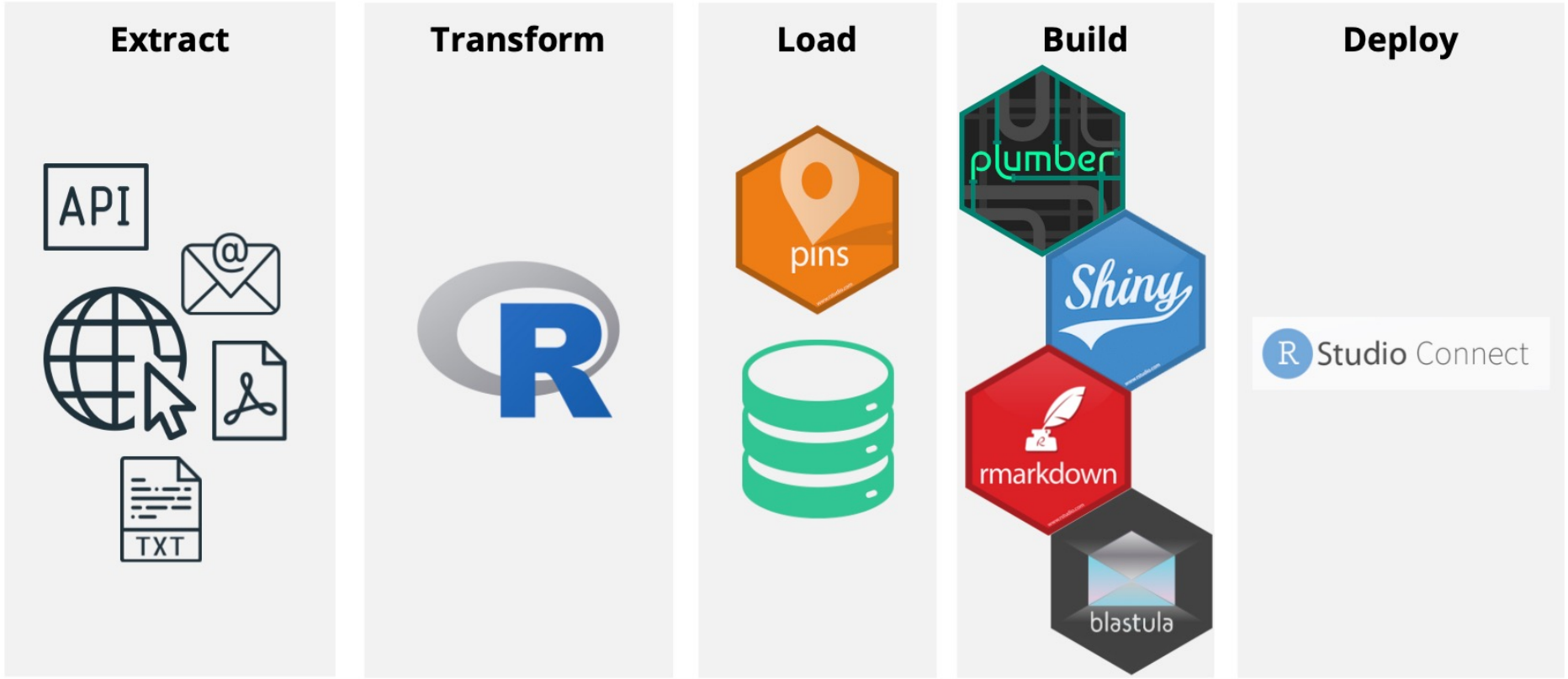
Our code should always be...

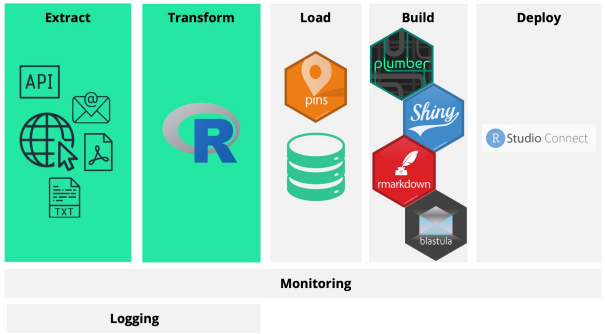
1. **Available**: our code (aka apps, ETL pipelines, reports) needs to be accessible at all times

2. **Validated**: mistakes in our ETL pipelines have big consequences, we validate the outcomes of our processes at all times

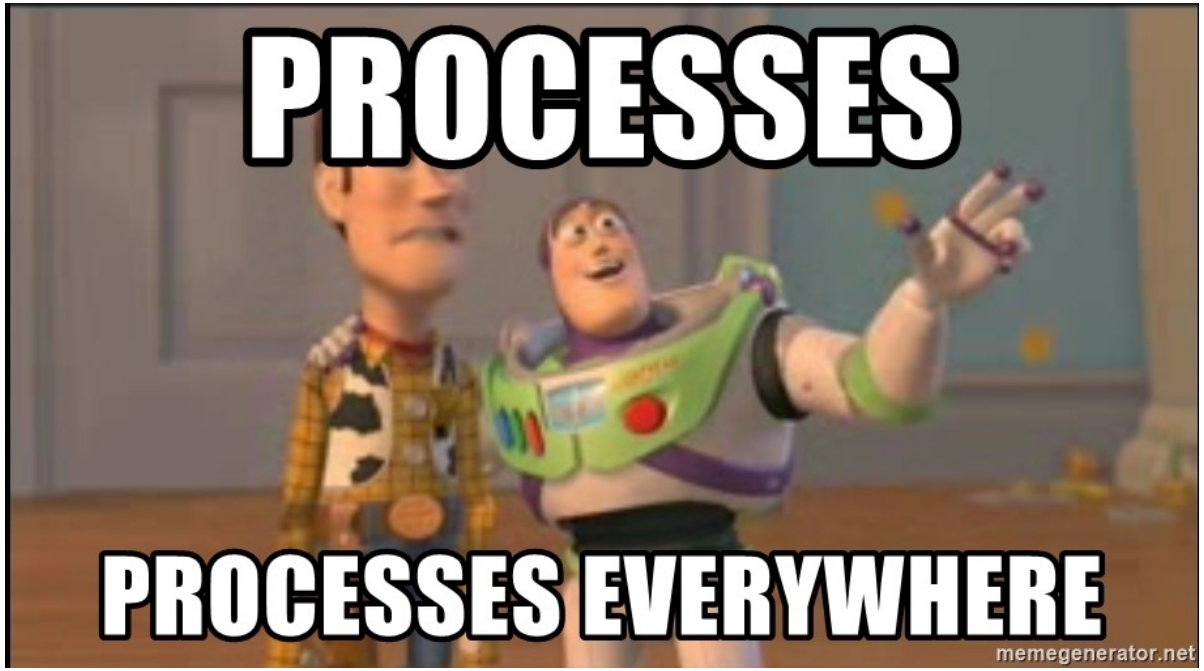
3. **Version controlled & monitored**: we know what's going on

A look into our process

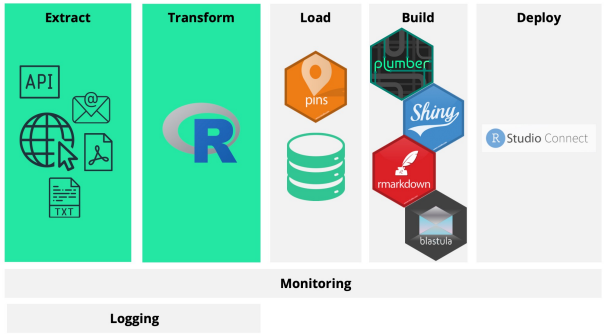




Data Gathering and Processing



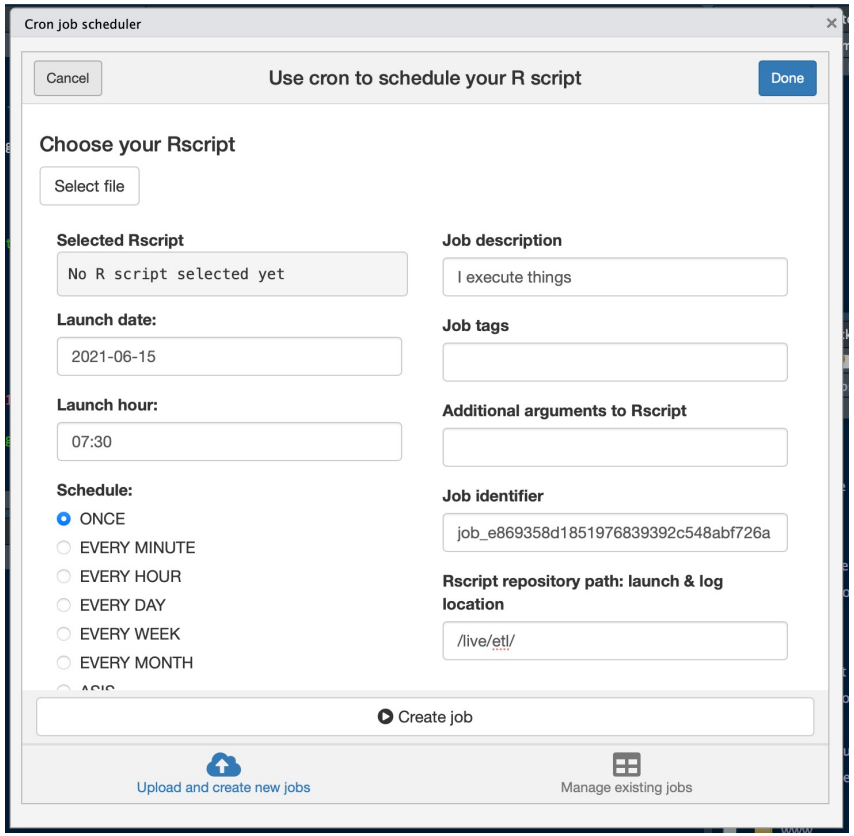
We have **32** ETL processes running in R

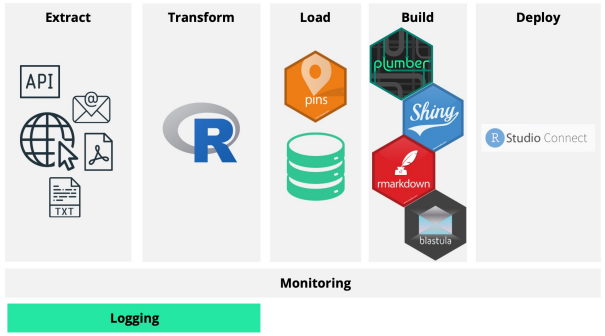


Data Gathering and Processing

cronR is your friend

```
# * * * * *  command to execute
# ┬ ┬ ┬ ┬ ┬
# │ │ │ │ │
# │ │ │ │ └──── day of week (0 - 7)
# │ │ │ └─────── month (1 - 12)
# │ │ └──────── day of month (1 - 31)
# │ └───────── hour (0 - 23)
# └──────── min (0 - 59)
```

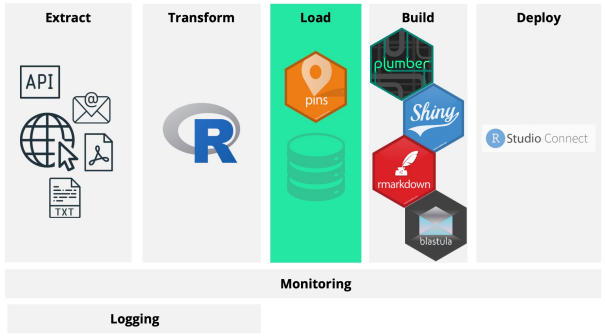




Logging

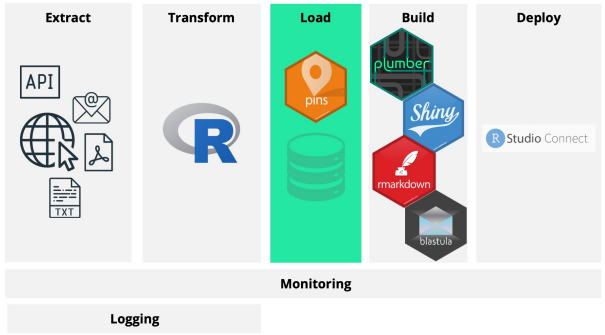
Make your logs valuable

- Execution time ⏰
- Which script was executed 📍
- Where problems occurred 👀



Storage

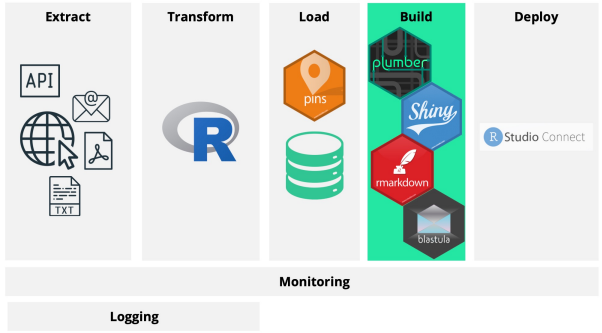
- AWS MySQL database
- Azure SQL Server databases
- Azure fileshare
- Pins



The beauty of pins

A **pin** can store **any resource**

- data.tables, data.frames, models, you name it!
- Share resources by publishing to **R Studio Connect**, GitHub, Azure, Google Cloud, S3, and more
- Great to use in Shiny apps
- It's fast

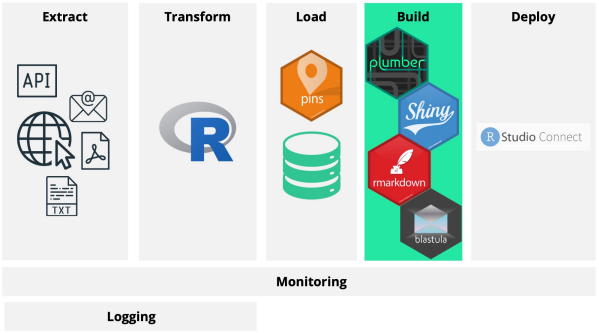


Plumber

Plumber helps you to turn your R code into a **REST API**

REpresentational **S**tate **T**ransfer **A**pplication **P**rogramming **I**nterface

aka **a way to retrieve information from the internet**



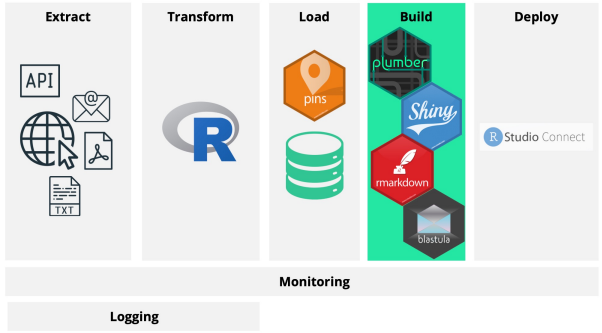
Plumber

All you need to do:

1. **Import** the plumber package
2. Make a **plumber.R** file with your endpoints
3. **Run** your API

```
# plumber.R

#* Echo back the input
#* @param msg The message to echo
#* @get /echo
function(msg="") {
  list(msg = paste0("The message is: '", msg, "'"))
}
```

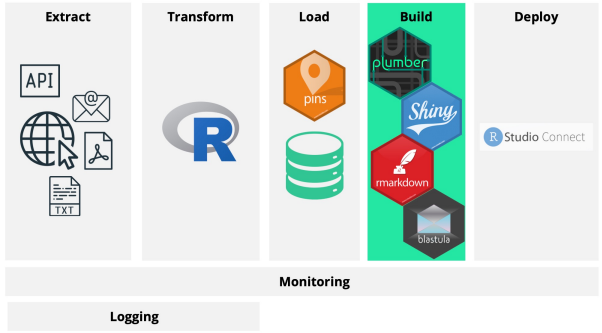
Shiny

Interactive **web applications**

Keep your app updated by using **database connections** or **pins,** avoid sourcing of static assets

Manage your connections to the app:

- Manage database connections to your instance by using the **pool** package
- Manage number of users allowed per session



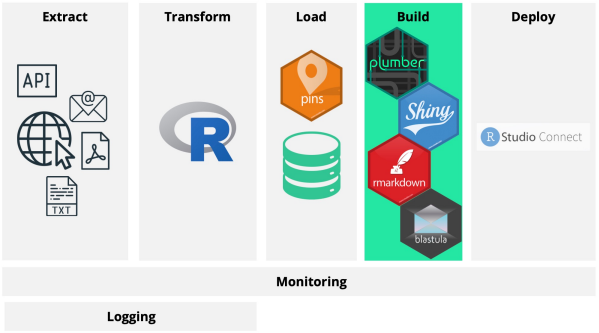
R Markdown

Interactive reports

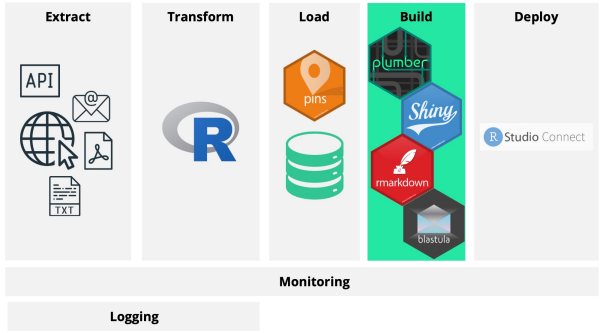
Send beautiful email reports with **blastula**

Can also use Shiny elements!

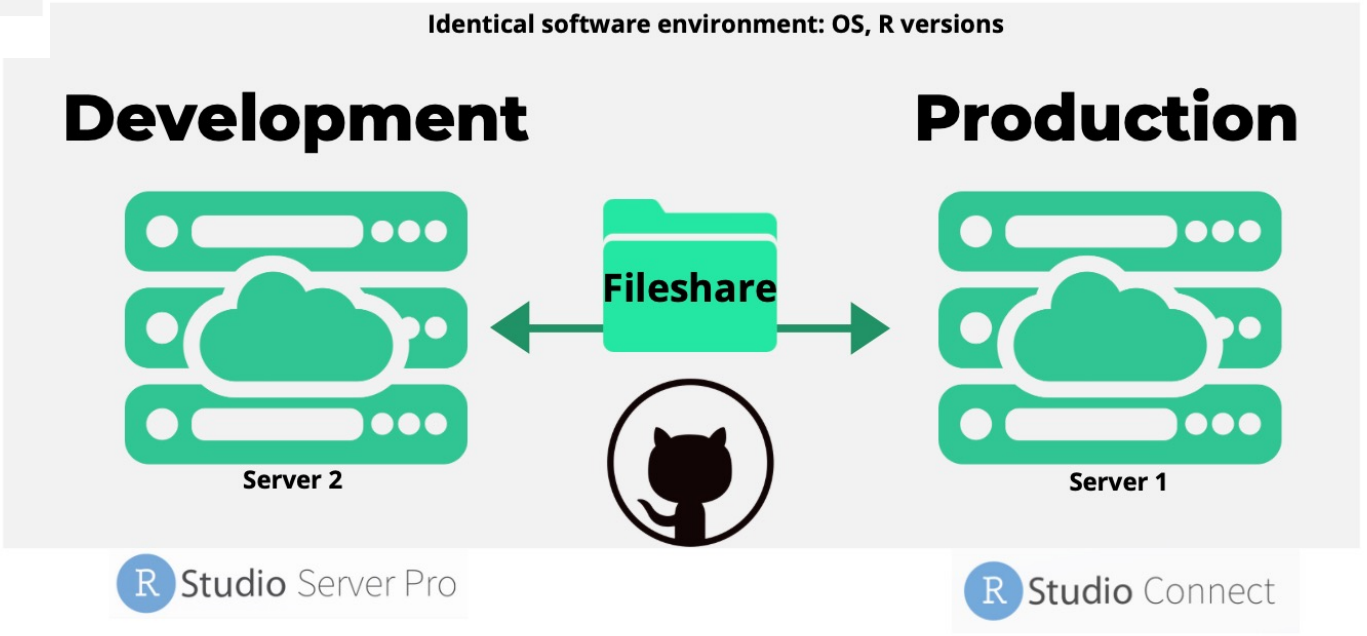
Convenient for communicating the most important changes each month/week



Separated environments



Separated environments



Dev

Development environment for Shiny Apps, Plumber APIs and ETL processes

Deployed Shiny Apps
Deployed Plumber APIs

Live

Live scheduled ETL processes

R Studio Connect

R Studio Connect

Hosting platform for all your R (and Python) related projects:

Applications (Shiny), Documents (R Markdown reports), pins, APIs

Makes it easy to distribute your products

Compared to shinyapps.io (for example):

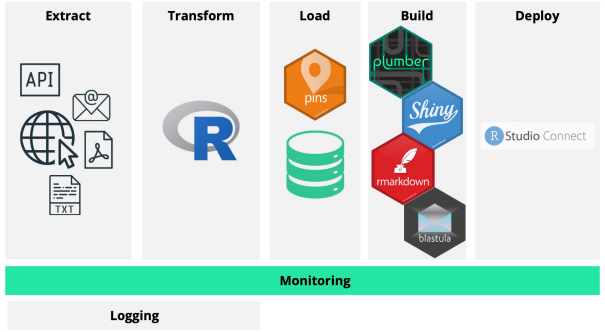
Higher level of control

Higher level of customization

R Studio Connect

R Studio Connect

Let’s take a look!



Waking up with our Data Pipeline Report

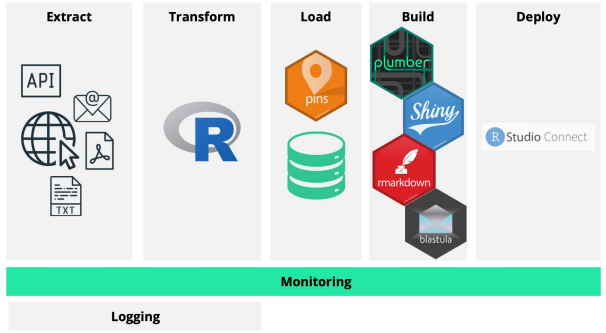
Data Pipeline Report

Welcome to the new Data Pipeline Email Report!

Recent updates (last 3 days)

etl code	dataset	update
etl_030	hospital_england_scmd	Fri 18 Jun
etl_009	concessionary_update_1	Fri 18 Jun
etl_031	concessionary_report	Thu 17 Jun
etl_005	prescribing_england	Thu 17 Jun

Using **R Markdown** with **Blastula**



And... Our Server Monitor Report

Server Monitor Report

Hi Greg and Veerle,

This is the Server Monitor Report for Server 1 and Server 2 Both servers are checked for problems every day. You will only get an email when it is Monday, or when there are issues.

Runtime of report: Monday, June 21, 2021 at 6:01 AM (UTC).

Server logging

Not a date that is the same as the runtime of the report? Please check crontab!

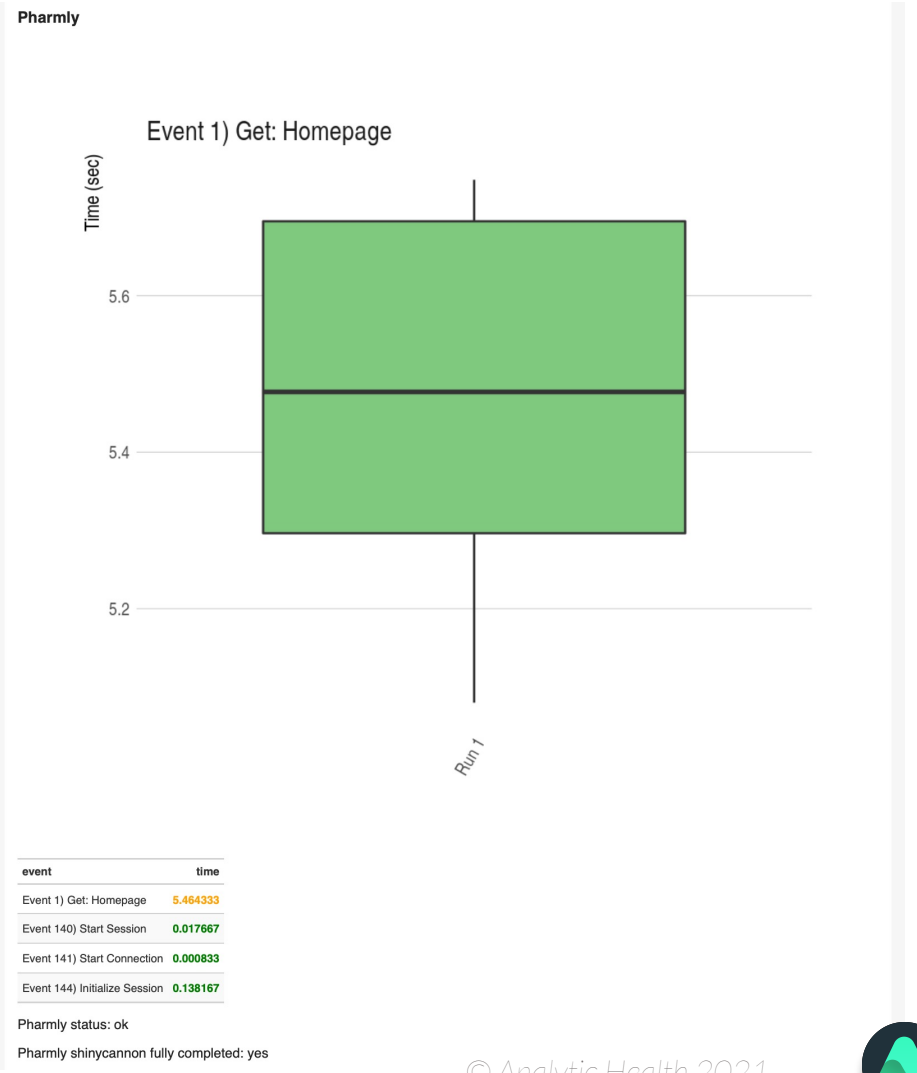
server	latest_log_date
Server 1	Mon Jun 21 05:05:02 UTC 2021
Server 2	Mon Jun 21 05:04:02 UTC 2021

Cron job: executed

Disk space

server	partition	filesystem	percentage_used
Server 1	/	sdc1	68%
Server 1	/data-drive	sdb1	2%
Server 1	/mnt	sda1	1%
Server 2	/	sdc1	86%
Server 2	/mnt	sda1	1%
Not applicable	/ahshare1	// 'ahshare1	54%

Using **shinycannon** to check if our apps are running accordingly



Our top tips

How to keep it manageable?

1. Version control: always, on everything 
2. Split your code up: smaller scripts are less overwhelming! ✂️
3. Package your code: gather and maintain functions 📦
4. Fixed project set-up: consistency brings clarity 💡
5. Documentation: for yourself and for your fellow developers 📄
6. Logging: know what's going on! 🔦
7. Monitoring: don't be surprised 🔍

R in Production – yes you can!

Questions?

Let’s keep in touch!

https://www.linkedin.com/in/veerlevanleemput/

Package overview & Resources

Mentioned packages & other resources

Packages

- cronR (https://cran.r-project.org/web/packages/cronR/vignettes/cronR.html)
- Pins (https://pins.rstudio.com)
- Plumber (https://www.rplumber.io)
- R markdown (https://rmarkdown.rstudio.com)
- Blastula (https://pkgs.rstudio.com/blastula/)
- Shiny (https://shiny.rstudio.com)
- Pool (https://github.com/rstudio/pool)
- Shinycannon (https://github.com/rstudio/shinycannon)

Other resources

- Engineering Production-Grade Shiny apps (https://engineering-shiny.org)
- R Studio Connect (https://www.rstudio.com/products/connect/)
- R Studio Server Pro (renamed to R Studio Workbench, https://www.rstudio.com/products/workbench/):