

H2O e Sparkling Water - Tópicos em Estatística 1

2022.1 - UnB - Professor Guilherme Rodrigues

Bruno Gondim, Eraldo Jair, José Felipe, Paulo Henrique e Pedro Aguiar

12/09/2022

Sumário I

- 1 Introdução
- 2 *H2O - Flow Web*
- 3 *H2O - R*
- 4 *H2O - Google Colab*
- 5 *Sparkling Water*
- 6 Fontes

Section 1

Introdução

H2O e *H2O.ai*: o que é?

The logo for H2O.ai is displayed within a large, rounded yellow rectangle. The text "H2O.ai" is centered in a bold, dark gray sans-serif font. The "H2O" part is significantly larger and bolder than the ".ai" domain extension.

H2O.ai

A H2O.ai é uma empresa de software de machine learning e inteligência artificial, sediada no Vale do Silício e reconhecida como visionária pelo Gartner. Formam um time de “Makers” que trouxe para o mercado novas plataformas e tecnologias que impulsionaram o movimento de IA.

São os idealizadores da H2O, plataforma líder em algoritmos de código aberto (open-source) para machine learning e inteligência artificial, utilizada por aproximadamente metade das empresas listadas na Fortune 500 e confiada por mais de 20 mil organizações, além de centenas de cientistas de dados ao redor do mundo.

É possível efetuar o *download* tanto localmente quanto em nuvem.

Ambientes em nuvem suportados:

- Instâncias *Amazon EC2* e *Storage S3* (*RedHat AMI*, *Amazon Linux AMI*, e *Ubuntu AMI*)
- *Amazon AWS*
- *Microsoft Azure*
- *Databricks*
- *IBM Power 9*
- *Nimbix Cloud*
- *Google Cloud*

Em h2o.ai/download, é possível fazer o *download*

H2O

H2O works with R, Python, Scala on Hadoop/Yarn, Spark or your laptop.

H2O is licensed under the [Apache License, Version 2.0](#)

Direct Downloads

[Latest Stable Release](#)[Nightly Bleeding Edge](#)[Prior Releases](#)

Cloud Downloads



Section 2

H2O - Flow Web

H2O - Flow Web

H2O Flow é uma interface de usuário de código aberto para *H2O*. É um ambiente interativo baseado na web que permite combinar execução de código, texto, matemática, gráficos e mídia avançada em um único documento.

H2O Flow envia comandos para *H2O* como uma sequência de células executáveis. As células podem ser modificadas, reorganizadas ou salvas em uma biblioteca. Cada célula contém um campo de entrada que permite inserir comandos, definir funções, chamar outras funções e acessar outras células ou objetos na página. Quando você executa a célula, a saída é um objeto gráfico, que pode ser inspecionado para visualizar detalhes adicionais.

Embora o *H2O Flow* suporte *API REST*, *scripts R* e *CoffeeScript*, nenhuma experiência de programação é necessária para executar o *H2O Flow*. Você pode clicar em qualquer operação *H2O* sem nunca escrever uma única linha de código. Você pode até mesmo desabilitar as células de entrada para executar o *H2O Flow* usando apenas a *GUI*. O *H2O Flow* foi projetado para guiá-lo em todas as etapas, fornecendo prompts de entrada, ajuda interativa e fluxos de exemplo.



Version 3.36.1.4

Fast Scalable Machine Learning API
For Smarter Applications

DOWNLOAD AND RUN

INSTALL IN R

INSTALL IN PYTHON

INSTALL ON HADOOP

USE FROM MAVEN

KUBERNETES

DOWNLOAD H₂O

Get started with H₂O in 3 easy steps

1. Download H₂O. This is a zip file that contains everything you need to get started.
2. From your terminal, run:

```
cd ~/Downloads
unzip h2o-3.36.1.4.zip
cd h2o-3.36.1.4
java -jar h2o.jar
```



3. Point your browser to <http://localhost:54321>

Observação:

ao rodar o código no terminal, o comando “*unzip*” não foi reconhecido pelo *Windows PowerShell*.

Utilizando o comando “*Expand-Archive*” no lugar, funcionou.

Ambiente de trabalho - H2O Flow

The screenshot displays the H2O Flow web interface. At the top, the menu bar includes "H2O FLOW" followed by "Flow", "Cell", "Data", "Model", "Score", "Admin", and "Help". Below the menu, the title "Untitled Flow" is shown above a toolbar with icons for file operations and execution. The main workspace area is labeled "assist" and contains an "Assistance" panel with a table of routines.

Routine	Description
importFiles	Import file(s) into H ₂ O
importSqlTable	Import SQL table into H ₂ O
getFrames	Get a list of frames in H ₂ O
splitFrame	Split a frame into two or more frames
mergeFrames	Merge two frames into one
getModels	Get a list of models in H ₂ O
getGrids	Get a list of grid search results in H ₂ O
getPredictions	Get a list of predictions in H ₂ O
getJobs	Get a list of jobs running in H ₂ O
runAutoML	Automatically train and tune many models
buildModel	Build a model
importModel	Import a saved model
predict	Make a prediction

On the right side, a "Help" sidebar is visible with the heading "Using Flow for the first time?" and a "Quickstart Videos" button. Below this, it says "Or, view example Flows to explore and learn H₂O." and "STAR H2O ON GITHUB!". A "Star" button is present. Under the "GENERAL" section, there is a list of links: "Flow Web UI ...", "... Importing Data", "... Building Models", "... Making Predictions", "... Using Flows", and "... Troubleshooting Flow". Under the "EXAMPLES" section, it states: "Flow packs are a great way to explore and learn H₂O. Try out these Flows and run them in your browser."

At the bottom left, a status bar shows a green dot and the word "Ready". At the bottom right, it shows "Connections: 0" and the H₂O logo.

Na próxima vez que você quiser iniciar o *Flow*, rode o seguinte código no terminal:

```
cd ~/Downloads/h2o-3.36.1.4  
java -jar h2o.jar
```

E acesse localhost/flow novamente.

Em tutorial.flow, está presente todo o passo a passo de como usar o *H2O Flow*. Alguns pontos importantes da interface são:

- Fluxos de exemplo: *HELP* -> *view example Flows*;
- Começar um novo fluxo: “*nem flow*”;
- Modos de célula: “editar” ou “comandar”
 - Células de comando:
 - MD (*markdown*)
 - CS (código padrão)
 - RAW (comentários de código)
 - H[1-6] (níveis de cabeçalho)
 - *Clips*: salvar células

Section 3

H2O - R

H2O - R

Código para instalar o pacote “H2O” no R:

```
# The following two commands remove any previously
# installed H2O packages for R.
if ("package:h2o" %in% search())
{ detach("package:h2o", unload=TRUE) }
if ("h2o" %in% rownames(installed.packages()))
{ remove.packages("h2o") }

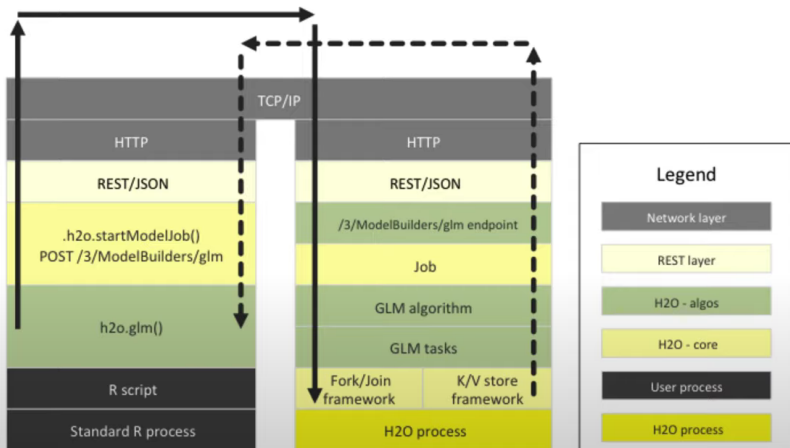
# Next, we download packages that H2O depends on.
pkgs <- c("RCurl","jsonlite")
for (pkg in pkgs) {
  if (! (pkg %in% rownames(installed.packages())))
  { install.packages(pkg) }}
```

```
# Now we download, install and initialize
# the H2O package for R.
install.packages("h2o", type="source",
repos=
"https://h2o-release.s3.amazonaws.com/h2o/rel-zumbo/4/R")

# Finally, let's load H2O and start up an H2O cluster
library(h2o)
h2o.init()
```

Workflor - Arquitetura do H2O.AI no R

H2O.AI - ARQUITETURA



Resumidamente, Comando R \rightarrow h2o \rightarrow Ambiente R

Exemplo de comando:

h2o_df=h2o.importFile("./path/arquivo.csv")

- se transforma internamente num arquivo do tipo h2oFrame

Vantagens:

- Até 100x mais veloz que o Scikit learn.
- Escalabilidade (quanto mais máquinas no cluster, melhor.)
- Interface UI para monitoramento em tempo real

Por que usar o h2o:

- superar limitações do R, tais como:
- Leitura de arquivos grandes (5GB+)
- Mau gerenciamento de memória

Problemas de usar o H2o no R:

- Modelos em R (**e python também**) não lidam tão bem assim com aplicações Java e Scala

Section 4

H2O - Google Colab

Section 5

Sparkling Water

Section 6

Fontes

Fontes

- <https://docs.h2o.ai/h2o-tutorials/latest-stable/index.html>
- <https://spark.rstudio.com/guides/h2o.html>