

Desafio 03

ME315-2S2025

brotto

2025-09-24

Parquet

O Apache Parquet é um formato de armazenamento de dados orientado a colunas, gratuito e de código aberto, no ecossistema Apache Hadoop. É semelhante ao RCFile e ao ORC, os outros formatos de arquivo de armazenamento em colunas do Hadoop, e é compatível com a maioria das estruturas de processamento de dados do Hadoop. Ele oferece esquemas eficientes de compactação e codificação de dados com desempenho aprimorado para lidar com dados complexos em massa. – Wikipedia

Função de leitura é `read_parquet` do pacote `arrow`

```
#lendo o arquivo parquet
iris_parquet <- read_parquet("./iris.parquet")

kable(head(iris_parquet))
```

sepal.length	sepal.width	petal.length	petal.width	variety
5.1	3.5	1.4	0.2	Setosa
4.9	3.0	1.4	0.2	Setosa
4.7	3.2	1.3	0.2	Setosa
4.6	3.1	1.5	0.2	Setosa
5.0	3.6	1.4	0.2	Setosa
5.4	3.9	1.7	0.4	Setosa

JSON

Em computação, **JSON** (pronúncia [ˈd ejzən], J-son em inglês), um acrônimo de JavaScript Object Notation, é um formato compacto, de padrão aberto independente, de troca de dados simples e rápida (parsing) entre sistemas, especificado por Douglas Crockford em 2000, que utiliza texto legível a humanos, no formato atributo-valor (natureza auto-descritiva). Isto é, um modelo de transmissão de informações no formato texto, muito usado em *web services* que usa transferência de estado representacional (REST) e aplicações AJAX, substituindo o uso do XML. – Wikipedia

Função de leitura é `fromJSON` do pacote `jsonlite`

```
iris_json <- fromJSON("./iris.json")

#lendo o arquivo JSON
kable(head(iris_json))
```

sepal_width	sepal_length	petal_width	petal_length	species
5.1	3.5	1.4	0.2	setosa

sepal_width	sepal_length	petal_width	petal_length	species
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

Refs

JSON. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2025. Disponível em: <https://pt.wikipedia.org/w/index.php?title=JSON&oldid=70826464>. Acesso em: 8 set. 2025.

Apache Parquet. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2025. Disponível em: https://en.wikipedia.org/w/index.php?title=Apache_Parquet&oldid=1305948399. Acesso em: 8 set. 2025.