

(<https://databricks.com>)

Etapa 1 - Criação da base de dados em formato parquet

Essa etapa de carregamento é oferecida pelo próprio databricks community edition. Aqui mantenho o comentário original. O Databricks utiliza do *Pyspark* para a leitura e gravação dos dados:

This notebook will show you how to create and query a table or DataFrame that you uploaded to DBFS. DBFS (<https://docs.databricks.com/user-guide/dbfs-databricks-file-system.html>) is a Databricks File System that allows you to store data for querying inside of Databricks. This notebook assumes that you have a file already inside of DBFS that you would like to read from.

```
# File location and type
file_location = "/FileStore/tables/amazon.csv"
file_type = "csv"

# CSV options
infer_schema = "false"
first_row_is_header = "true"
delimiter = ","

# The applied options are for CSV files. For other file types, these will be ignored.
df = spark.read.format(file_type) \
    .option("inferSchema", infer_schema) \
    .option("header", first_row_is_header) \
    .option("sep", delimiter) \
    .load(file_location)

display(df)
```

Table

	product_id	product_name	category
1	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Charging and Data Sync Cable Compatible for iPhone 13, 12,11, X, 8, 7, 6, 5, iPad Air, Pro...	Computers&Accessories Accessori...
2	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5m Braided Type C Cable for Smartphones, Tablets, Laptops & other Type C devices, ...	Computers&Accessories Accessori...
3	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync USB Cable Compatible for iPhone 13, 12,11, X, 8, 7, 6, 5, iPad Air, Pro, Mini & iOS Devices	Computers&Accessories Accessori...
4	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB Stress Resistant, Tangle-Free, Sturdy Cable with 3A Fast Charging & 480mbps Data Tr...	Computers&Accessories Accessori...
5	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 Pin USB Cable with Charge & Sync Function for iPhone, iPad (Grey)	Computers&Accessories Accessori...
6	B08Y1TFSP6	pTron Solero TB301 3A Type-C Data and Fast Charging Cable, Made in India, 480Mbps Data Sync, Strong and Durable 1.5-Meter Nylo...	Computers&Accessories Accessori...
7	B08WRWPM22	boAt Micro USB 55 Tangle-free, Sturdy Micro USB Cable with 3A Fast Charging & 480mbps Data Transmission (Black)	Computers&Accessories Accessori...
8	B08DDRGWTJ	MI Usb Type-C Cable Smartphone (Black)	Computers&Accessories Accessori...
9	B008IFXQFU	TP-Link USB WiFi Adapter for PC(TL-WN725N), N150 Wireless Network Adapter for Desktop - Nano Size WiFi Dongle Compatible with...	Computers&Accessories Networkir...
10	B082LZGK39	Ambrane Unbreakable 60W / 3A Fast Charging 1.5m Braided Micro USB Cable for Smartphones, Tablets, Laptops & Other Micro USB ...	Computers&Accessories Accessori...
11	B08CF3D7QR	Portronics Konnect L POR-1081 Fast Charging 3A Type-C Cable 1.2Meter with Charge & Sync Function for All Type-C Devices (Grey)	Computers&Accessories Accessori...
12	B0789LZTCJ	boAt Rugged v3 Extra Tough Unbreakable Braided Micro USB Cable 1.5 Meter (Black)	Computers&Accessories Accessori...
13	B07KSMBL2H	AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot	Electronics HomeTheater,TV&Vide...
14	B085DTN6R2	Portronics Konnect CL 20W POR-1067 Type-C to 8 Pin USB 1.2M Cable with Power Delivery & 3A Quick Charge Support, Nylon Braide...	Computers&Accessories Accessori...

685+ rows

| Truncated data due to byte limit

```
permanent_table_name = "amazon__csv"
df.write.format("parquet").saveAsTable(permanent_table_name)
```

O comando acima salva o dataset original em parquet. Ele aparentemente, uma vez executado, não necessita executar novamente. Porém mantenho aqui o comando para deixar claro como foi escrito o DF em parquet no Databricks.

