

# SANTUARIO DOS DADOS



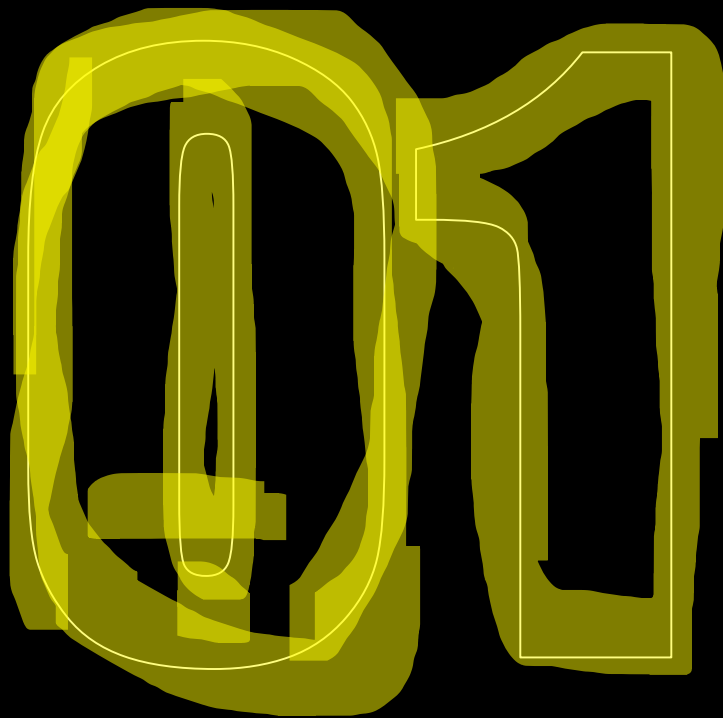
## EXPLORANDO DATABRICKS

# DATABRICKS ESSENTIALS

## INTRODUÇÃO AO DATABRICKS

Databricks é uma plataforma unificada de análise de dados baseada no Apache Spark, que facilita a construção e o gerenciamento de pipelines de dados, aprendizado de máquina e análises em larga escala. Este ebook apresenta os principais conceitos e funcionalidades do Databricks com exemplos práticos de código.





# CONFIGURANDO UM WORKSPACE NO DATABRICKS

---

Um workspace no Databricks é onde você armazena e organiza notebooks, bibliotecas, e outros recursos. Para começar, você precisa criar um workspace

```
# Exemplo de criação de um notebook no Databricks  
dbutils.notebook.run("path/to/your_notebook", 60)
```



# TRABALHANDO COM NOTEBOOKS

---

Notebooks são documentos interativos onde você pode escrever código, texto, e visualizar resultados. No Databricks, notebooks suportam múltiplas linguagens, como Python, SQL, Scala e R.

```
# Exemplo de célula de código em um notebook  
data = spark.read.csv("/path/to/csv")  
display(data)
```





# MANIPULAÇÃO DE DADOS COM DATAFRAMES

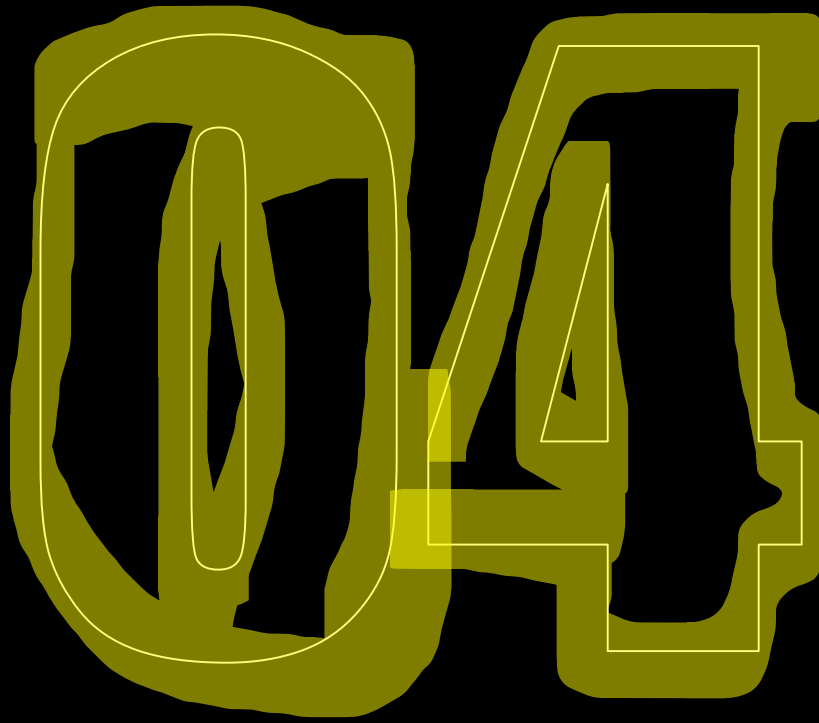
---

DataFrames são estruturas de dados distribuídas, semelhantes a tabelas em bancos de dados relacionais, que facilitam a manipulação e a análise de grandes volumes de dados.

```
# Leitura de um arquivo CSV em um DataFrame
df = spark.read.format("csv").option("header", "true").load("/path/to/file.csv")

# Exibindo as primeiras linhas do DataFrame
df.show(5)

# Filtrando dados
filtered_df = df.filter(df['age'] > 30)
filtered_df.show()
```



# ESCREVENDO E EXECUTANDO CONSULTAS SQL

---

No Databricks, você pode escrever e executar consultas SQL diretamente nos notebooks. Isso é útil para analistas de dados que preferem SQL ao invés de linguagens de programação.

```
-- Exemplo de consulta SQL em um notebook Databricks  
SELECT name, age FROM people WHERE age > 30
```

# 05

## VISUALIZAÇÕES DE DADOS

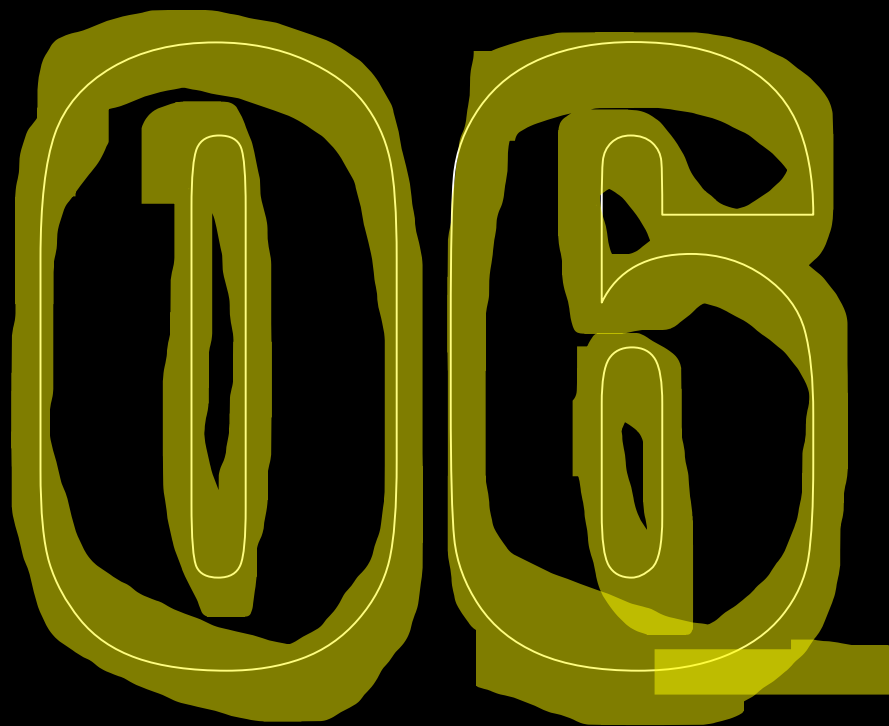
### I

Databricks permite criar visualizações interativas para explorar e entender seus dados. Você pode usar bibliotecas como Matplotlib, Seaborn ou usar as visualizações integradas do Databricks.

```
# Exemplo de criação de um gráfico de barras usando Matplotlib
import matplotlib.pyplot as plt

data = df.groupby("age").count().toPandas()
data.plot(kind="bar", x="age", y="count")

plt.show()
```



# MACHINE LEARNING COM DATABRICKS

---

O Databricks oferece suporte completo para o desenvolvimento de modelos de aprendizado de máquina, desde a preparação dos dados até a implantação dos modelos.

```
# Exemplo de pipeline de Machine Learning usando PySpark
from pyspark.ml.feature import VectorAssembler
from pyspark.ml.classification import LogisticRegression

# Preparação dos dados
assembler = VectorAssembler(inputCols=["feature1", "feature2"], outputCol="features")
data = assembler.transform(df)

# Treinamento do modelo
lr = LogisticRegression(featuresCol="features", labelCol="label")
model = lr.fit(data)

# Previsão
predictions = model.transform(data)
predictions.show()
```





# GERENCIAMENTO DE TAREFAS COM JOBS

---

Jobs no Databricks permitem agendar a execução de notebooks, scripts ou pipelines de dados. Isso é útil para tarefas de ETL ou treinamento de modelos que precisam ser executadas periodicamente.

```
# Exemplo de agendamento de um Job via API Databricks
import requests

response = requests.post(
    'https://<databricks-instance>/api/2.0/jobs/runs/submit',
    headers={'Authorization': 'Bearer <your-token>'},
    json={
        "run_name": "Example Job",
        "existing_cluster_id": "<cluster-id>",
        "notebook_task": {
            "notebook_path": "/path/to/your_notebook"
        }
    }
)
print(response.json())
```

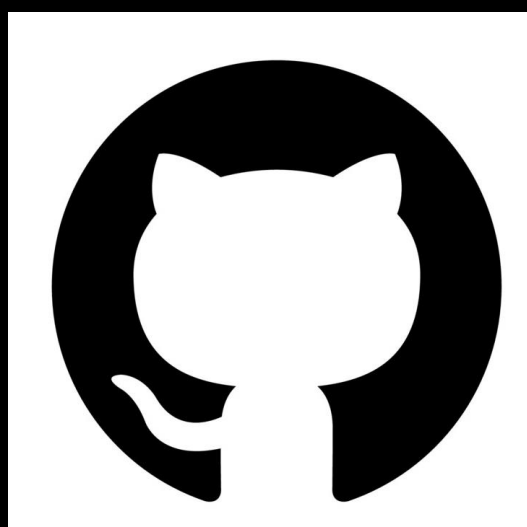
# AGRADECIMENTOS

---

OBRIGADO POR LER ATÉ AQUI

Esse Ebook foi gerado por IA, e diagramado por humano.  
O passo a passo se encontra no meu GitHub

Esse conteúdo foi gerado com fins didáticos de construção, não foi realizado uma validação cuidadosa humana no conteúdo e pode conter erros gerados por uma IA.



<https://github.com/DiegoPablo2021/>