

# PYSPARK ANALISANDO DADOS COM SPARK

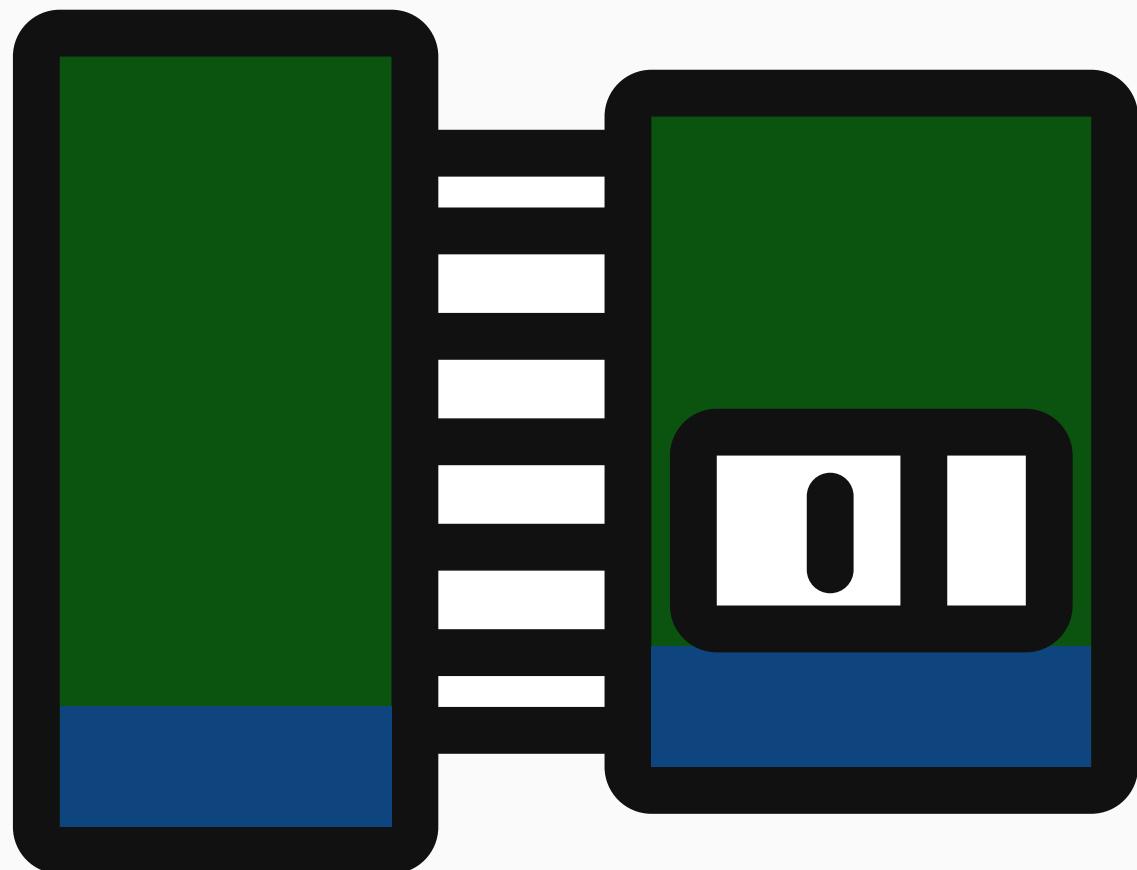
Grimaldo Oliveira

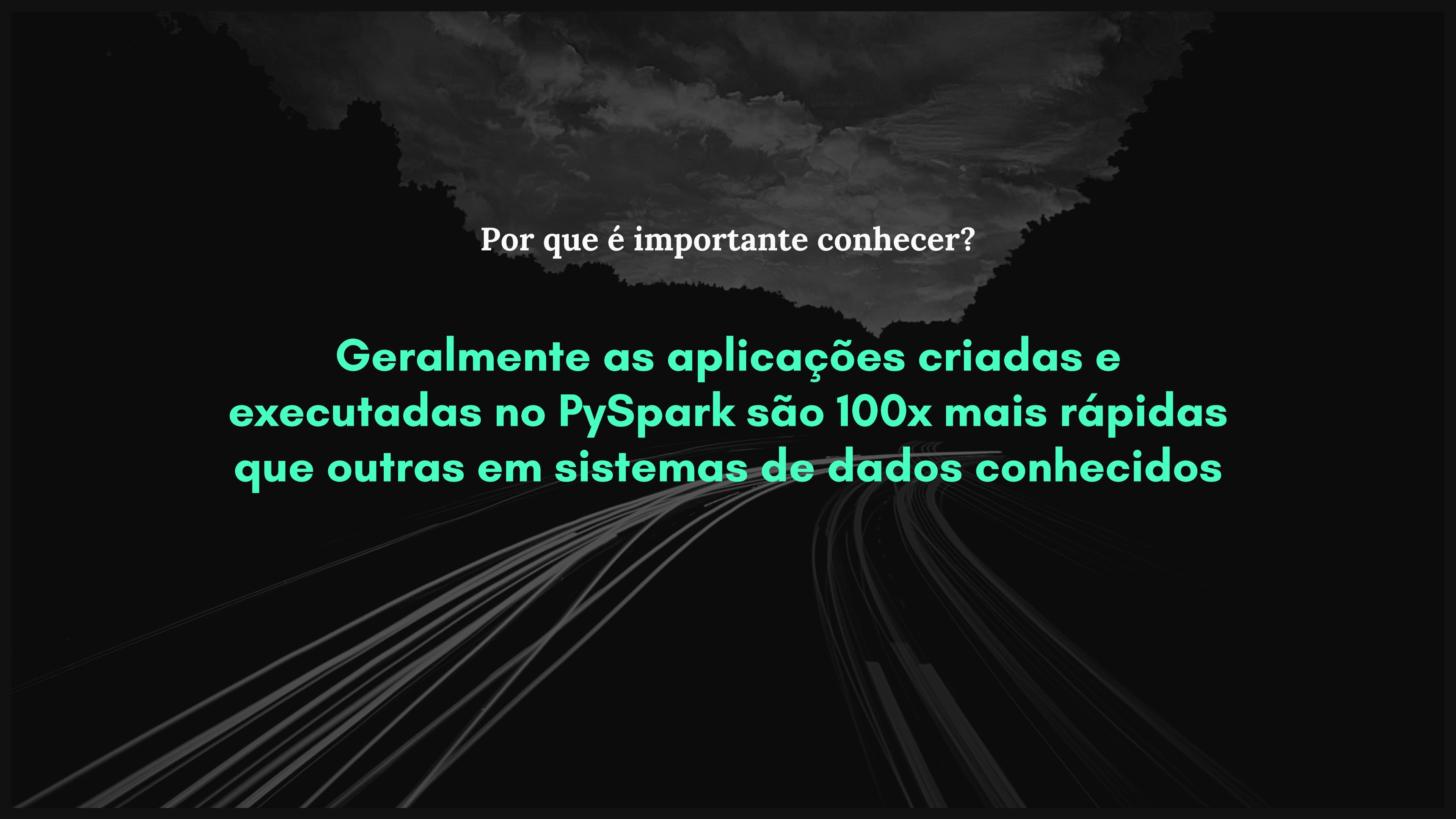


# Sobre

## PYSPARK

PySpark é uma API Python para Apache SPARK que é denominado como o mecanismo de processamento analítico para aplicações de processamento de dados distribuídos em larga escala e aprendizado de máquina, ou seja, para grandes volumes de dados.

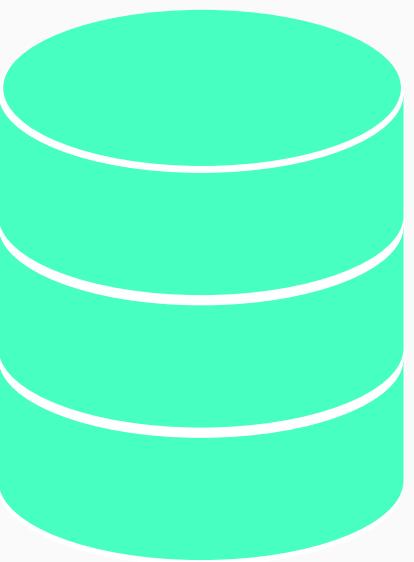




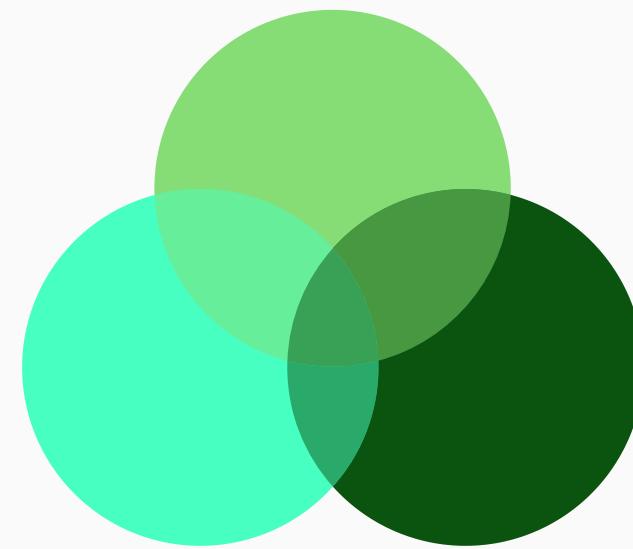
Por que é importante conhecer?

**Geralmente as aplicações criadas e executadas no PySpark são 100x mais rápidas que outras em sistemas de dados conhecidos**

# COMPONENTES DE APRENDIZAGEM



Dados em RDD e  
DATAFRAME



SPARK SQL

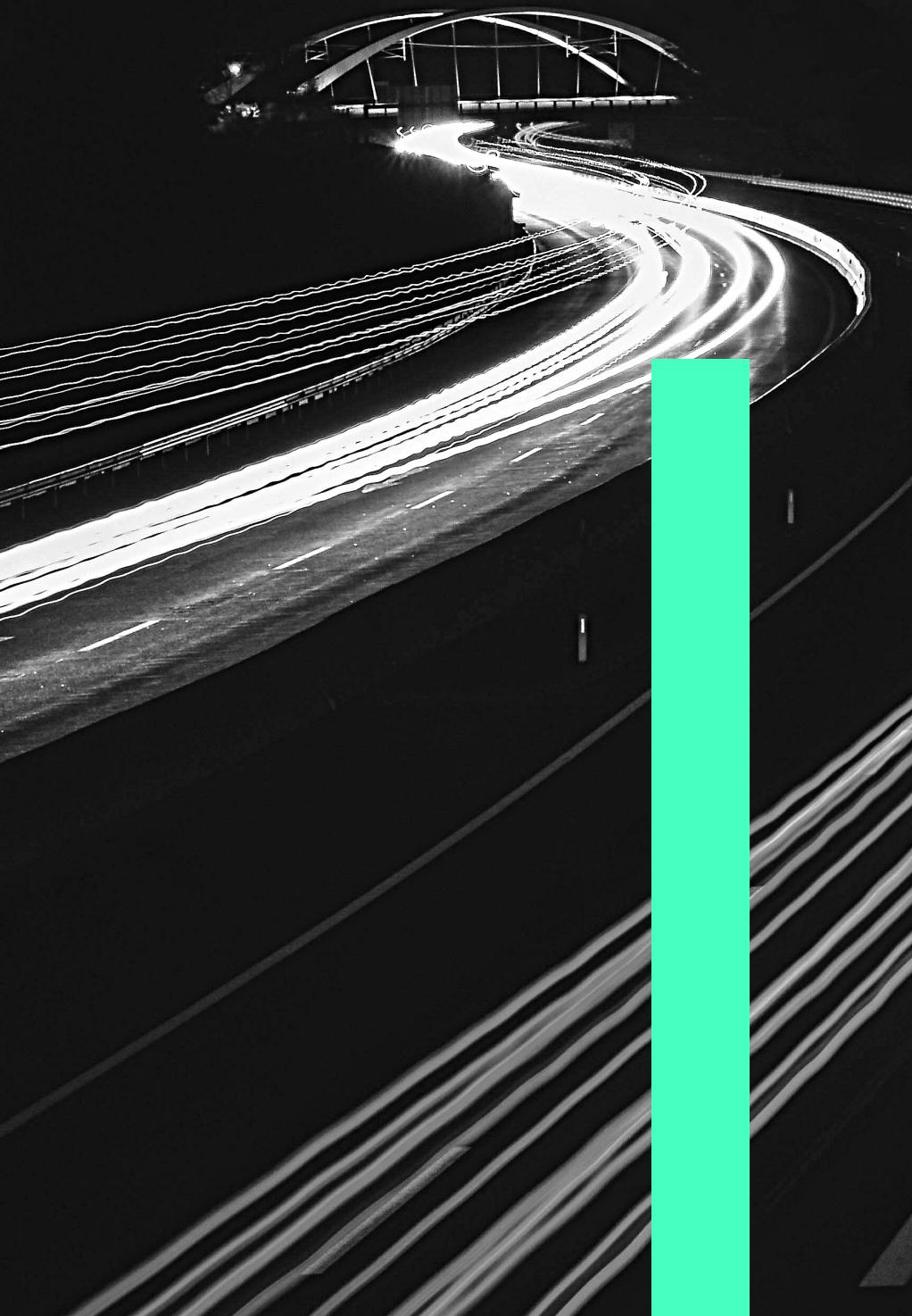


SPARK STREAMING

# Como é conhecido no mercado

O uso do PySpark está diretamente ligado a ferramenta PYTHON que é a mais utilizada no mundo para a área de Ciência de Dados. Graças a isso, a busca por profissionais que saibam manipular SPARK e PYTHON é muito grande.

Há uma corrida para que profissionais da área de dados conheçam e saibam construir projetos na nuvem, sendo o SPARK e a linguagem PYTHON as mais requeridas em praticamente todos os projetos.



# Venha conhecer!

**Site com todos os cursos**

[< CLIQUE -> TODOS OS CURSOS >](#)

**Site link direto ao treinamento**

[< CLIQUE -> APRENDAPYSPARK >](#)