

## Introdução ao Curso

### Professores:

- Rafael Erlacher
- Elias Theodoro
- Jéssica Mirely de Sousa Carvalho
- Elivelton Repolho
- Bruno Trevisan
- Willian Ribeiro



## **Tópicos abordados**

Apache Spark + Databricks

Databricks - Apache Spark™ Certified Developer

## **Nível de conteúdo**

Desde o nível básico até avançado!

## **Modelo de Ensino**

Ensino a distância (somente online)

## **Videoaulas no Youtube**

[https://www.youtube.com/playlist?list=PLcCdpkLIh\\_PEAfNMf0ETZSo6VzDM\\_ydI](https://www.youtube.com/playlist?list=PLcCdpkLIh_PEAfNMf0ETZSo6VzDM_ydI)

## **Meetings para revisão de conteúdo e esclarecimento de dúvidas**

Todas as terças-feiras as 20h no Google Meet.

## **Duração do curso**

20 semanas (até a segunda quinzena de setembro).

## **Disponibilização dos conteúdos**

Todos os sábados no Youtube + GitHub

## **Material de apoio + Desafios semanais + Slides + Bibliografias Recomendadas**

<https://github.com/erlacherDev/Blueshift-Databricks-Spark/>

## Principal Referência para o curso!

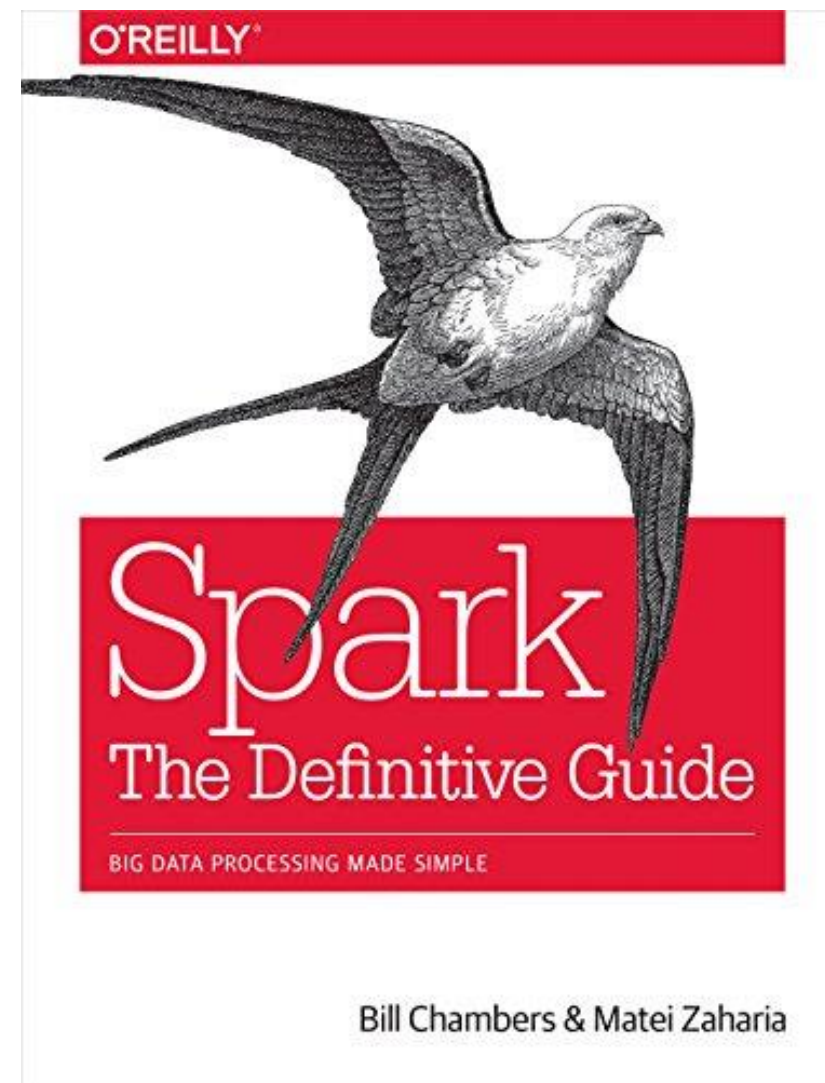
Link para compra do livro:

[https://www.amazon.com/\\_/dp/1491912219?tag=oreilly20-20](https://www.amazon.com/_/dp/1491912219?tag=oreilly20-20)

Material de Apoio:

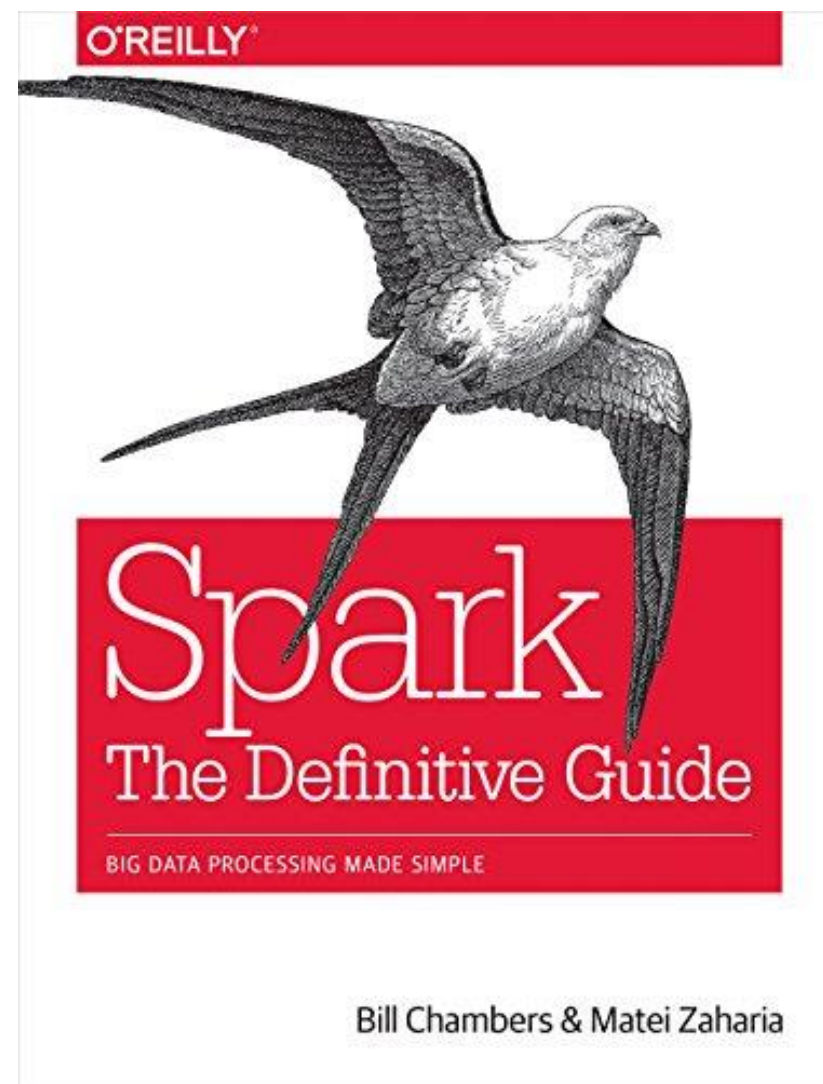
<https://github.com/databricks/Spark-The-Definitive-Guide/tree/master/data>

## Livro disponibilizado no GitHub!



## Principais Capítulos Abordados

1. What is Apache Spark?
2. A gentle Introduction to Spark
3. A Tour of Spark's Toolset
4. Structured API Overview
5. Basic Structured Operations
6. Working With Different Types of Data
7. Aggregations
8. Joins
9. Data Sources
10. Spark SQL
12. RDDs
13. Advanced RDDs
14. Distributed Shared Variables
18. Monitoring and Debugging
19. Performance and Tuning
21. Structured Streaming Basics
24. Advanced Analytics and Machine Learning Overview
30. Graph Analytics





## Materiais extras disponibilizados pelo Databricks:

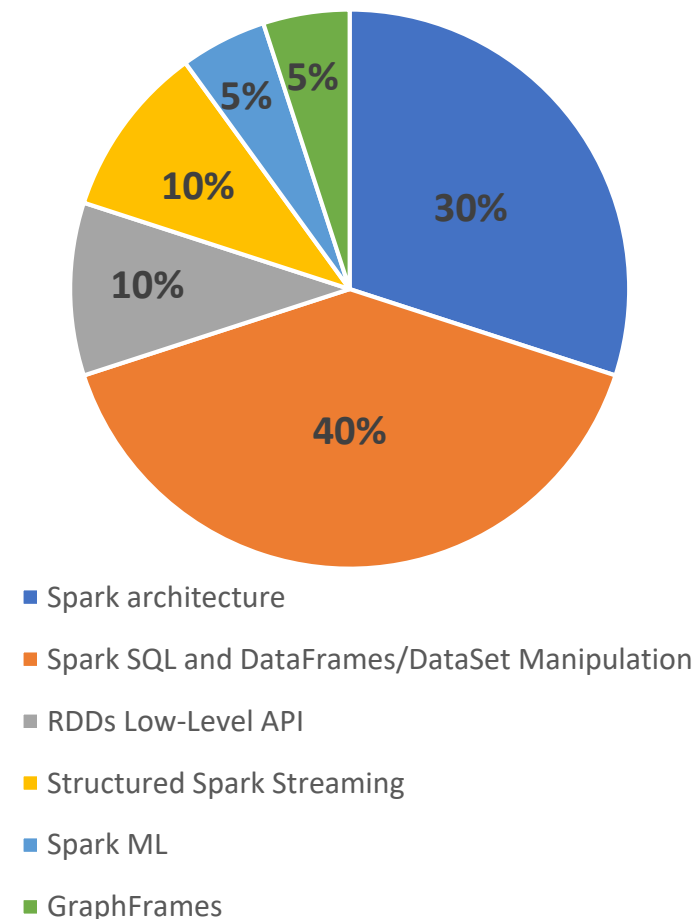


**Disponíveis no GitHub!**

- Introdução ao Apache Spark
- Introdução ao Databricks
- RDDs and Low-Level APIs
- Spark SQL and DataFrames Manipulation
- Spark Architecture and Run-time Behavior
- Structure Streaming
- Spark ML
- GraphFrames

**Não será lecionado Python e Scala!**

## Databricks: Spark Developer Certification



## CRT010: Databricks Certified Developer: Apache Spark™ 2.X

- Preço da certificação: \$300 com direito a duas tentativas
- Duração da prova: 180 minutos
- Prova em Inglês
- 40 Questões
- Pode ser feita em Python ou Scala (Your choice!)
- Passing Score: 65%
- Em caso de aprovação a BlueShift reembolsará a prova!

**Disponível até 31 de outubro de 2019!**



<https://github.com/vivek-bombatkar/Databricks-Apache-Spark-2X-Certified-Developer> -> Guia para a certificação

<https://databricks.com/sparkaisummit/north-america/sessions> -> Palestras dos últimos Spark Summit

[https://www.youtube.com/watch?v=WaxMj5\\_SLUI](https://www.youtube.com/watch?v=WaxMj5_SLUI) -> Introduction to Databricks

<https://www.youtube.com/watch?v=p8FGC49N-zM> -> What is Apache Spark?