



Apresentação

Bem vindo(a) à Aceleração PySpark da Capgemini. O objetivo é introduzir o ecossistema Hadoop e os conceitos básicos de processamento de dados usando a tecnologia Spark com Python (PySpark). Nesta primeira semana, o objetivo é aprendermos juntos com base em um curso de 10h disponível pela própria Capgemini. Abaixo é apresentado o roteiro que seguiremos ao longo da semana.



Dia 1

- ☐ Week 1 e 2 - Videos (Introduction to BigData + Hadoop)
- ☐ Reading
- ☐ Practice Quiz
- ☐ Quiz

Dia 2

- ☐ Week 3
- ☐ Hands-on Lab (Introduction to Apache Spark)
- ☐ Reading
- ☐ Practice Quiz
- ☐ Quiz

Dia 3

- ☐ Week 4 - Videos (Introduction to DataFrames & SparkSQL)
- ☐ Hands-on Lab (DataFrames & SparkSQL)
- ☐ Reading
- ☐ Practice Quiz
- ☐ Quiz

Dia 4

- ☐ Week 5 - Videos (Spark Architecture + Runtime Environments)
- ☐ Hands-on Lab
- ☐ Reading
- ☐ Practice Quiz
- ☐ Quiz

Dia 5

- ☐ Week 6 - Monitoring & Tuning
- ☐ Final Exam
- ☐ Course Wrap-Up