

# Processamento Massivo Paralelo

BI e Big Data  
UNIFACS

# Robespierre Pita

Bacharel em Sistemas de Informação - UNIFACS

Especialista em Redes de Computadores e Telecom - UNIFACS

Mestre em Ciência da Computação - UFBA (integração de dados em Big Data)

Doutorando em Ciência da Computação - UFBA (machine learning)

Cientista de Dados no CIDACS.

# E você?

- O que é Big Data?
- O que é aprendizado de máquina?
- O uso de dados já melhorou alguma atividade sua?
- Qual sua formação?
- Já trabalha com dados grandes? Onde? Como?
- Já trabalha com aprendizado de máquina? Onde? Como?
- Quais as expectativas com a disciplina?



# Expectativas...

- Entender o que é Big Data e quais as principais soluções para lidar com grandes volumes de dados hoje.
- Entender a manipulação básica de dados para uso em data analytics.
- Entender fundamentos do aprendizado de máquina e validação dos modelos.
- Usar esse entendimento para liderar ou implementar projetos de Big Data Analytics.

- 
- Se tornar expert:
    - nas ferramentas utilizadas na disciplina.
    - em aprendizado de máquina.
    - em soluções big data.

# Big Data Analytics

**Ementa:** Conceitos, tecnologias. Mineração de dados. Machine learning. Tomada de decisões. Data Analytics. Ferramentas. Desafios na utilização de Big Data na indústria avançada: cenário, desafio, benefícios.

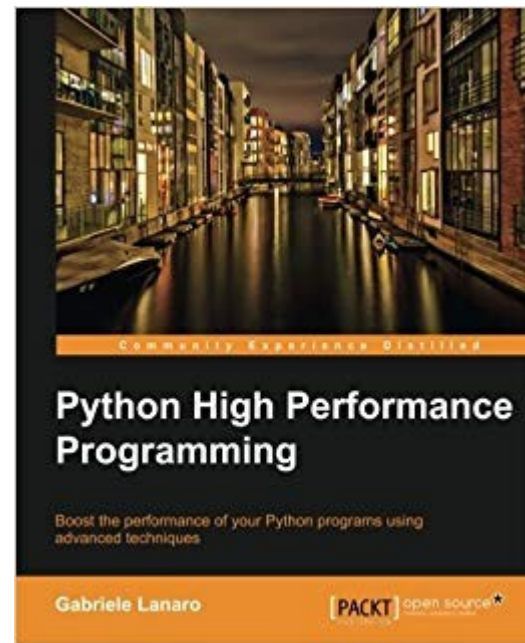
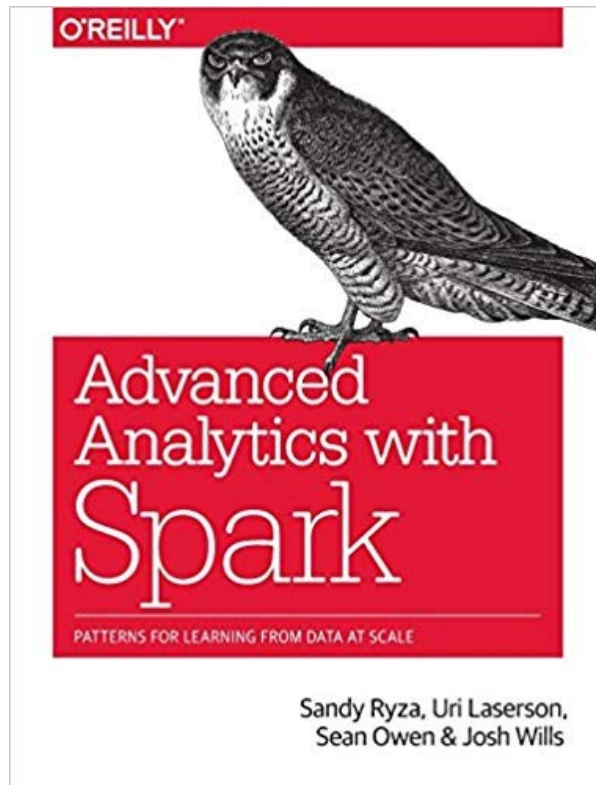
## Conteúdo programático:

Módulo I - 22 a 24/07	Módulo II - 24 e 27/07
Apresentação da disciplina	PRÁTICA – Preprocessamento de dados
Fundamentos de computação paralela e distribuída	PRÁTICA – Análise de dados
Ferramentas de Big Data	PRÁTICA – Machine Learning e Visualização
Fundamentos de aprendizado de máquina	Avaliação.

# Avaliação

- Os alunos serão separados em trios ou duplas (preferencialmente duplas).
- Alguns datasets serão disponibilizados para a execução de modelos de machine learning. O resultado do trabalho será um relatório contendo:
  - O pré-processamento de dados feito pela equipe;
  - Modelos de aprendizado de máquina utilizados;
  - Os resultados de desempenho dos modelos (tempo de execução e acurácia);
  - O código fonte usado para rodar os modelos no Jupyter notebook.
- O que será avaliado?
  - Justificativa sobre o pré-processamento adotado (sugestão: experimentar diferentes técnicas).
  - Cuidados contra overfitting.
  - Interpretação dos resultados.
- Qual o tempo de avaliação:
  - As últimas 3 horas do curso.

# Principais referências



# Dúvidas?

