



Proposta Comercial

Curso:

Python and Spark for Big Data and Machine Learning

Código do Curso: 4522



Carga Horária:
40 horas



Oferecido nas modalidades:

- Presencial em SP
- Online: Live Class ou Agile Class
- In Company

4.Linux Open Software Specialists™

Empresa líder na formação de profissionais Linux e open software.

Mais de 70.00 alunos treinados.

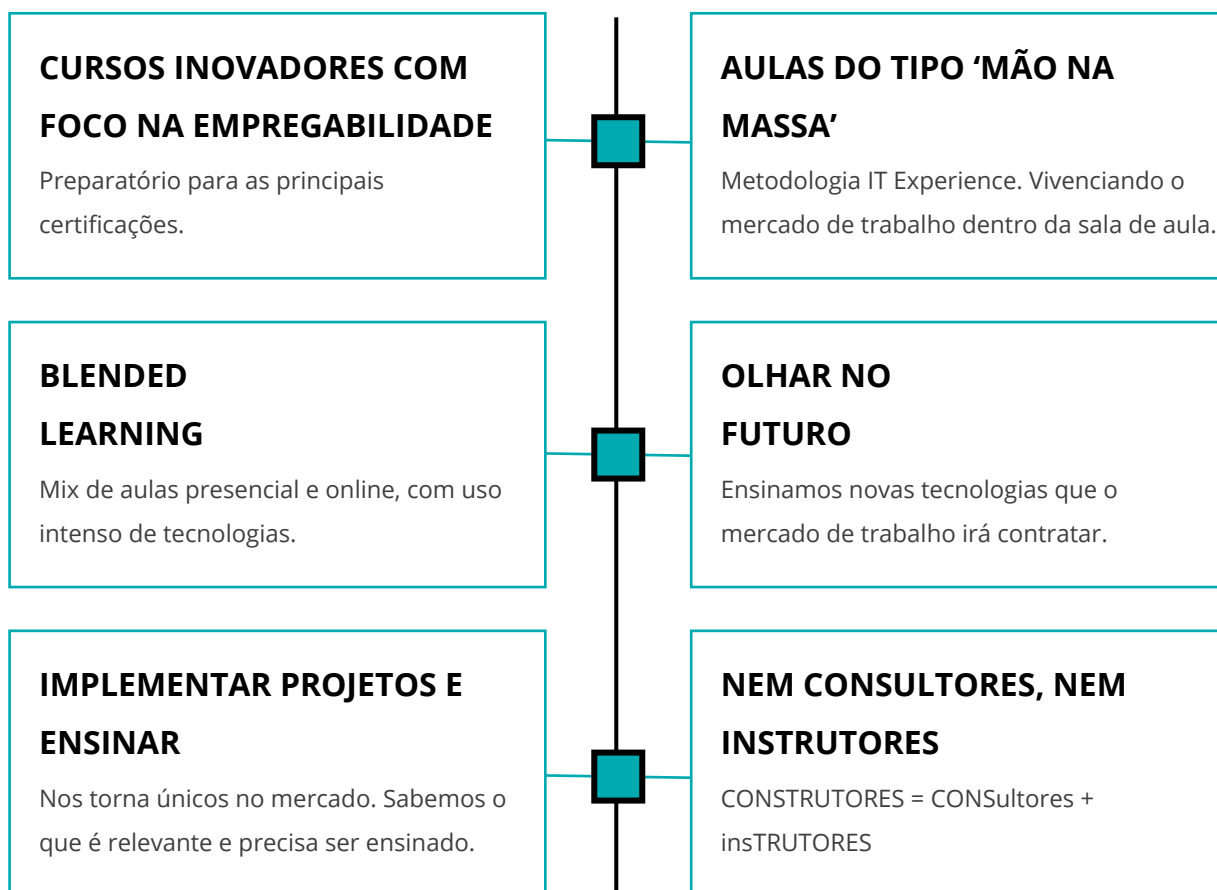
Mais de 4800 empresas atendidas.

Muito Prazer, somos a 4Linux.

Fundada em 2001, a 4Linux é líder de mercado em cursos de Linux e open source com números que impressionam: mais de 70.000 alunos treinados, mais de 4.800 empresas atendidas e mais de 40 diferentes cursos altamente especializados. Somos uma das poucas escolas de TI que também atua em consultoria e isso traz inúmeros benefícios aos nossos alunos, com uma metodologia de ensino única.

Localizada a apenas 3 minutos do metrô da Vila Mariana, na cidade de São Paulo - onde ministramos nossos cursos presenciais - também oferecemos os cursos nas modalidades online e In Company.

Veja abaixo por que nossos cursos transformam carreiras e nossos alunos são disputados pelas empresas:





Quem deve fazer este curso:

Proporcionar uma visão geral dos principais algoritmos de Machine Learning e, de forma crítica, discutir como aplicá-los usando ferramentas Open Source, tais como Spark e Python; Entregar ao aluno um RoadMap por meio da transmissão de conhecimentos a respeito das áreas: Ciência de Dados, Machine Learning e Processamento de Alto Desempenho.

Alguns números deste curso:

+70.000

Alunos foram treinados pela
4Linux

+246

Alunos assistiram este
curso

R\$7.009,00

Média salarial de quem concluiu
o curso *

+19

Empresas contrataram este
curso

* Valor médio aproximado com base nas pesquisas dos maiores portais de empregos: Catho, Indeed, TrabalhaBrasil, Glassdoor e Apinfo.



Após fazer este curso, o aluno estará apto a:

- Aplicar técnicas de mapeamento/redução nos dados;
- Extrair e carregar tabelas a partir de arquivos, bases de dados e outras fontes (ETL);
- Preparar infraestrutura para trabalhar com APIs de Machine Learning em Python;
- Acelerar o processamento de grandes bases de dados;
- Usar os principais algoritmos de Machine Learning com Python e suas APIs;
- Trabalhar com Pyspark;
- Planejar e implementar um projeto de Big Data para Análise e Visualização de Dados;
- Aplicar os princípios de Ciência de Dados, Machine Learning e Big Data;

Veja os diferenciais do curso:

01 Big Data, Machine Learning, Computação de Alto Desempenho e Análise de Dados se relacionam

02 Todo material em língua portuguesa (a maioria dos cursos de terceiros está no idioma inglês)

03 Oportunidade de conhecer as experiências reais de empresas que aplicaram as práticas ensinadas

04 Demonstra na prática as diferenças de papéis do Engenheiro de Dados e do Cientista de Dados



Ementa do curso

Introdução a Big Data

- Python para Big Data
- Jupyter Notebook
- Google Colab

Python para Big Data

- O modelo de programação MapReduce
- Introdução à Programação funcional com Python: Funções Lambda em Python
- Trabalhando com grafos em Python
- Numpy
- Pandas

Análise exploratória de Dados

- Tipos de Variáveis
- Normalização de Dados com Python
- Visualização de Dados com Python (Matplotlib, Seaborn, Bokeh)
- Medidas de similaridade

Matemática e Dados

- Introdução à Estatística Descritiva
- Média, Mediana, Desvio Padrão, Variância
- Teorema de Bayes e Inferência Bayesiana
- Teste Estatístico de Hipótese
- Exemplo: Executando um Teste A/B

Spark e Pyspark

- Operações básicas com DataFrames
- Operações avançadas com DataFrames
- Operações com Dados Faltantes e Datas



Ementa do curso

Introdução ao Machine Learning

- Introdução a Machine Learning
- Overview dos Algoritmos de Machine Learning
- K-nearest neighbors – kNN
- Construção de Pipeline de Machine Learning com modelo CRISP DM
- Introdução a Sistemas de Recomendação e Filtros Colaborativos

Machine Learning

- Regressão Linear
- Regressão Logística x kNN – Avaliação de Modelos
- Fundamentos de Redes Neurais Artificiais
- Regressão Logística x kNN – Avaliação de Modelos
- Introdução a Processamento Natural de Linguagem (Modelos n-grams)
- Word2Vec com Gensim
- Clusterização usando Kmeans
- Classificação usando Naive Bayes
- Classificação usando Árvores de Decisão e Random Forest
- Classificação usando Support Vector Machine

Introdução ao Machine Learning Spark MLlib

- Introdução ao Spark MLlib
- Tokenização de documentos de textos para tarefas de classificação
- Regressão Logística com Spark MLlib
- KNN com Spark MLlib
- Spark Streaming



Ementa do curso

Deep Learning

- Introdução a Deep Learning
- Redes Neurais Convolucionais
- Redes Neurais Recorrentes
- Introdução ao TensorFlow
- Conhecendo a base de dados MNIST
- Aplicação Prática de Redes Neurais MLP para Reconhecimento de Imagem
- Aplicação Prática de Redes Neurais Convolucionais para Reconhecimento de Imagem
- Aplicação Prática de Redes Neurais Recorrentes para Análise de Sentimentos

Processamento de Alto Desempenho com Python para Big Data

- A revolução das GPUs em Machine Learning
- Arquitetura Cuda e Modelo de Programação em GPU
- Introdução ao Numba
- Comparação de desempenho entre Numpy, Numpy assíncrono e Numpy com Numba em Operações estatísticas
- Conhecendo os principais frameworks Python para Machine Learning que usam GPUs



Pré-requisitos

Para o aluno

- > Desejável conhecer comandos básicos do Linux;
- > Desejável conhecer comandos SQL básicos;
- > Desejável conhecer Lógica de Programação;
- > Desejável conhecer o básico da linguagem Python (variáveis, estruturas de repetição, estruturas condicionais, lambdas) ou ter realizado curso [Python Fundamentals \(520\)](#);

Computacionais presencial/EAD/EAD AO VIVO

- > É necessário que o aluno tenha um computador (Notebook ou Desktop) com no mínimo 8GB de memória RAM, com processador com suporte à 64bits pois será necessário emular máquinas virtuais para realizar os laboratórios práticos
- > Ter instalado o VirtualBox com o Extension Pack em seu sistema operacional (Linux, MacOS X, Windows) pois será necessário emular máquinas virtuais para realizar os laboratórios práticos
- > Caixas de Áudio ou Fones de ouvido
- > Monitor configurado com resolução mínima de 1024x768
- > Navegador de Internet Google Chrome/Chromium
- > Sistema Operacional Linux, Windows ou MacOS X
- > Recomendado 5MB de velocidade de conexão internet banda larga
- > Alunos com computadores da Apple de arquitetura ARM não conseguirão realizar nossos cursos que necessitam de virtualização (VirtualBox, KVM, VMWare, Parallels), pois até o momento não há suporte oficial e/ou estável nestas plataformas para a virtualização de máquinas com arquitetura x86_64

In Company

- > Sala equipada com projetor, Quadro Branco ou FlipChart
- > Acesso à internet por Banda larga, utilizando Rede Ethernet
- > Caso exista algum proxy ou bloqueio na rede, a 4Linux deverá ser informada para providenciar com antecedência o download dos arquivos necessários

FICOU COM ALGUMA DÚVIDA?

Converse agora com nossos consultores para
informações de datas e valores

FALE COM A GENTE

SP

T: +55 11. 2125-4747

T: +55 11. 2125-4748

W: +55 11. 96429-0501

Rua Vergueiro, 3057

Vila Mariana, SP

04101-300