

INTRODUÇÃO AO CLUSTERIZAÇÃO EM BIG DATA

Universidade de Fortaleza
- UNIFOR

MBA EM GESTÃO
ANALÍTICA COM
BUSINESS INTELLIGENCE
E BIG DATA

Prof. Manoel Ribeiro

Conteúdo

- Aula
 - Carreira em Big Data
 - Avaliação

Carreira

$$\Delta x = v_0 t$$
$$\Delta x = v_0 t + \frac{at^2}{2}$$
$$v = v_0 + at$$
$$v^2 = v_0^2 + 2a \Delta x$$

$$\nabla \cdot \vec{E} = \frac{1}{\epsilon_0} \rho$$
$$\nabla \cdot \vec{B} = 0$$
$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

$$\nabla \times \vec{B} = \mu_0 \vec{J} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t}$$

$$r = r_x \hat{i} + r_y \hat{j} + r_z \hat{k}$$
$$r = (r_x, r_y)$$
$$z = \int r_x^2 + r_y^2 + r_z^2$$
$$y = \int r_x^2 + r_y^2 + r_z^2$$
$$r = \sqrt{r_x^2 + r_y^2}$$
$$r_{gx} = \frac{r_x}{r_z}$$

$$v_m = \frac{v + v_0}{2}$$

$$h = \frac{v^2 - v_0^2}{2g}$$

$$V = \frac{\Delta S}{\Delta T} = \frac{5-5_0}{T}$$



Data Scientist

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

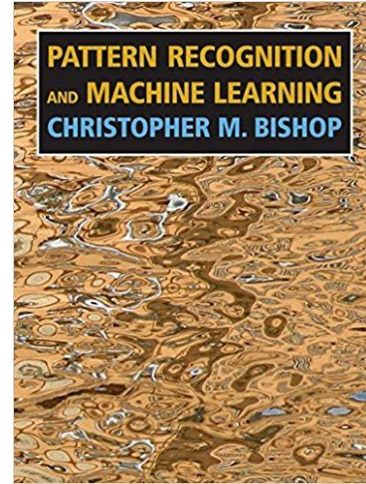


PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

Skill of Data Scientist

1. Matemática e Estatística avançada
 - a. Machine Learning
 - b. Modelagem Estatística
 - c. Inferência Bayesiana
 - d. Aprendizado supervisionado: Random Forest, Decision tree e regressão logística
 - e. Aprendizado não supervisionado: agrupamento e redução de dimensionalidade
 - f. Otimização: gradient descent and variant



Skill of Data Scientist

1. Computação e banco de dados
 - a. Fundamentos da ciência de computação
 - b. Linguagem de Script (Python)
 - c. Ferramenta estatística (R studio)
 - d. Banco de dados SQL e NoSQL (MongoDB)
 - e. Processamento Paralelo (MapReduce)
 - f. Hadoop
 - g. Spark
 - h. Experiência com Nuvem (AWS)
 - i. Visualização de dados (Tableau)



Data Engineer

DATA ENGINEER

SOFTWARE ENGINEERS BY TRADE

Role

Develops, constructs, tests and maintains architectures (such as databases and large-scale processing systems)

Mindset

All-purpose everyman



HIRED BY



Languages

SQL, Hive, Pig, R, Matlab, SAS, SPSS, Python, Java, Ruby, C++, Perl

Skills & Talents

- ✓ Database systems (SQL & NO SQL based)
- ✓ Data modeling & ETL tools
- ✓ Data APIs
- ✓ Data warehousing solutions

Data Engineer Responsibilities

1. Concepção, construção, instalação, teste e manutenção de sistemas de gerenciamento de dados altamente escaláveis
2. Garantir que os sistemas atinjam os requisitos comerciais e as práticas do setor.
3. Construa algoritmos, protótipos, modelos preditivos e prova de conceitos de alto desempenho
4. Oportunidades de pesquisa para aquisição de dados e novos usos para dados existentes
5. Desenvolver processos de conjunto de dados para modelagem de dados, mineração e produção

Data Engineer Responsibilities

1. Integre novas tecnologias de gerenciamento de dados e ferramentas de engenharia de software em estruturas existentes
2. Crie componentes de software personalizados (por exemplo, UDFs especializados) e aplicativos de análise
3. Utilize uma variedade de idiomas e ferramentas (por exemplo, linguagens de script) para casar sistemas juntos
4. Instale e atualize procedimentos de recuperação de desastres
5. Recomenda formas de melhorar a confiabilidade, eficiência e qualidade dos dados
6. Colabore com arquitetos de dados, modeladores e membros da equipe de TI em objetivos de projeto

Skill of Data Engineer

1. Computação e banco de dados
 - a. Arquitetura de BigData
 - b. ETL/BI/Data warehouse
 - c. Instalação de cluster (Hadoop, Spark, MongoDB)
 - d. Sistema distribuído (HDFS)
 - e. Expertise em Nuvem (AWS, OpenStack)
 - f. Banco de dados NoSQL (Todos)
 - g. Banco de dados avançado (APIs, data load)
 - h. Banco de dados em memória
 - i. Aplicações escaláveis



Salary of Data Engineer

1. Data Engineer

Glassdoor

Average Salary (2015): \$95,936 per year

Minimum: \$66,000

Maximum: \$117,000

2. Big Data Engineer

Robert Half Technology 2015 Salary Guide

Average Salary (2014): \$110,250 – \$152,750

Average Salary (2015): \$119,250 – \$168,250

Average Salary (2016): \$129,500 – \$183,500



Google Trends

Comparar



data scientist

Termo de pesquisa



data engineer

Termo de pesquisa



Adicionar comparação

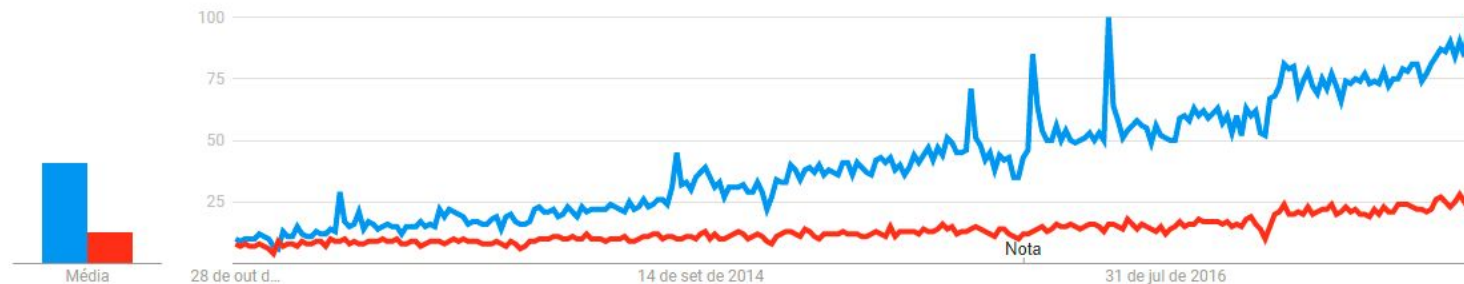
Todo o mundo ▼

Nos últimos 5 anos ▼

Todas as categorias ▼

Pesquisa na Web ▼

Interesse ao longo do tempo ⓘ





Início



Minha rede



Vagas



Mensagens



Notificações



Eu ▼

Todos

Pessoas

Vagas

Empresas

Grupos

Instituições de ensino



data scientist



Brasil

Data do anúncio ▼

Empresa ▼

Nível de experiência ▼

Localidade ▼

Recursos do LinkedIn ▼

[Você é profissional de TI - Faça a sua inscrição no evento referência de Infraestrutura Digi](#)

Exibindo 139 resultados

Classificar por: Relevância ▼



Oferta De Emprego - Cientista De Dados

Confidential via buscojobs.com.br

Rio de Janeiro, BR

Ver todos os empregos de Não esp. Voltar à página de pesquisa. Enviar esta oferta para um amigo. 2017-10-10 / Viewed 10 times, RJ, Rio de Janeiro. Adicionar emprego aos favoritos. www.buscojobs.com.br



Seja um dos 10 primeiros candidatos

há 2 semanas



Todos

Pessoas

Vagas

Empresas

Grupos

Instituições de ensino

📁 data engineer

📍 Brasil

Data do anúncio ▼

Empresa ▼

Nível de experiência ▼

Localidade ▼

Recursos do LinkedIn ▼

[Green Belt em Fortaleza](#) - Participe da melhor qualificação Green Belt em Fortaleza. Início em

Exibindo 45 resultados

Classificar por: Relevância ▼



Data Engineer

LexisNexis Risk Solutions

São Paulo e Região, Brasil

Good time management skills. BS Engineering/Computer Science or equivalent experience required Key Responsibilities. Experience . Ability to work with simple data models.



1 ex-aluno trabalha aqui

há 2 semanas



Data Engineer

VS



Data Scientist

Vamos Trabalhar?



Tiago Aguirre

Fim