



# THE DEVELOPER'S CONFERENCE

## Trilha Big Data


Cientistas e Engenheiros de Dados... preciso dos dois perfis em um projeto Big Data?

**Ana Sanchez**  
Engenheira de Dados na  **LOGICALIS**

# Quem sou eu?



## Ana Sanchez

Engenheira de Dados na  LOGICALIS

- Formada em:

- Análise e Desenvolvimento de Sistemas
- Análise de Big Data



- 7 anos de experiência:

- Programação Web
- Programação de Jogos
- Sistemas de Gerenciamento de Redes
- Business Intelligence (ETL)
- Big Data (Engenharia de Dados)
- Formação de Engenheiros de Dados



Certified for

 | **Analytics**

Cognos 10 BI Administrator

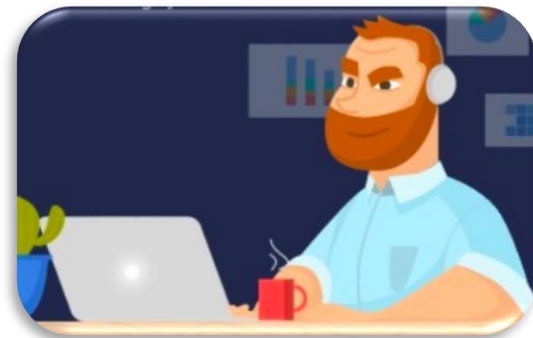


ADMINISTRATOR





Cientistas e Engenheiros de Dados...  
preciso dos dois perfis em um projeto  
Big Data?



2012



THE  
DEVELOPER'S  
CONFERENCE



DATA

# Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

“If “sexy” means having rare qualities that are much in demand, data scientists are already there. They are difficult and expensive to hire and, given the very competitive market for their services, difficult to retain.”

# Boom da Ciência de Dados



**FEATURED PROGRAM**  
**Data Science**

[Learn more](#)

Professional Certificate Program

HARVARD UNIVERSITY

JOHNS HOPKINS UNIVERSITY

**Data Science**

Johns Hopkins University

**Applied Data Science with Python**

University of Michigan

MICROMASTERS® PROGRAM

UCSanDiegoX  
**Data Science**  
Gain the critical skills needed to become a data scientist, rated...

coursera

Cursos



**Data Scientist**  
Microsoft · Redmond, WA, US

[Save](#) [Apply](#)



**Data Scientist**  
Amazon Web Services (AWS) · Baltimore, MD, US

[Save](#) [Apply](#)



**Senior Data Scientist**  
The New York Times · New York City, NY, US

[Save](#) [Apply](#)



**Data Scientist**  
Facebook · Menlo Park, CA, US

[Save](#) [Apply](#)



**Data Scientist**  
Cisco · San Jose, CA, US

[Save](#) [Apply](#)



**Data Scientist**  
General Motors · Detroit, MI, US

[Save](#) [Apply](#)



**DATA SCIENTIST**  
Walmart · Carlsbad, CA, US

[Save](#) [Apply](#)



**Data Scientist**  
Pearson · Centennial, CO, US

[Save](#) [Apply](#)

Vagas

# Cientista de Dados



- Amostra de Dados
- Limpeza dos Dados
- Análise Exploratória dos Dados (Estatística)
- Modelos de Machine Learning



- Modelos de Inteligência Artificial
- Programação básica à intermediária (Python/R)
- Apresentação dos resultados obtidos
- Experimentação

# Estava indo tudo bem até que...



THE  
DEVELOPER'S  
CONFERENCE

Experimentação,  
Prova de Conceito  
(POC)



Produção



**Ciências de Dados**

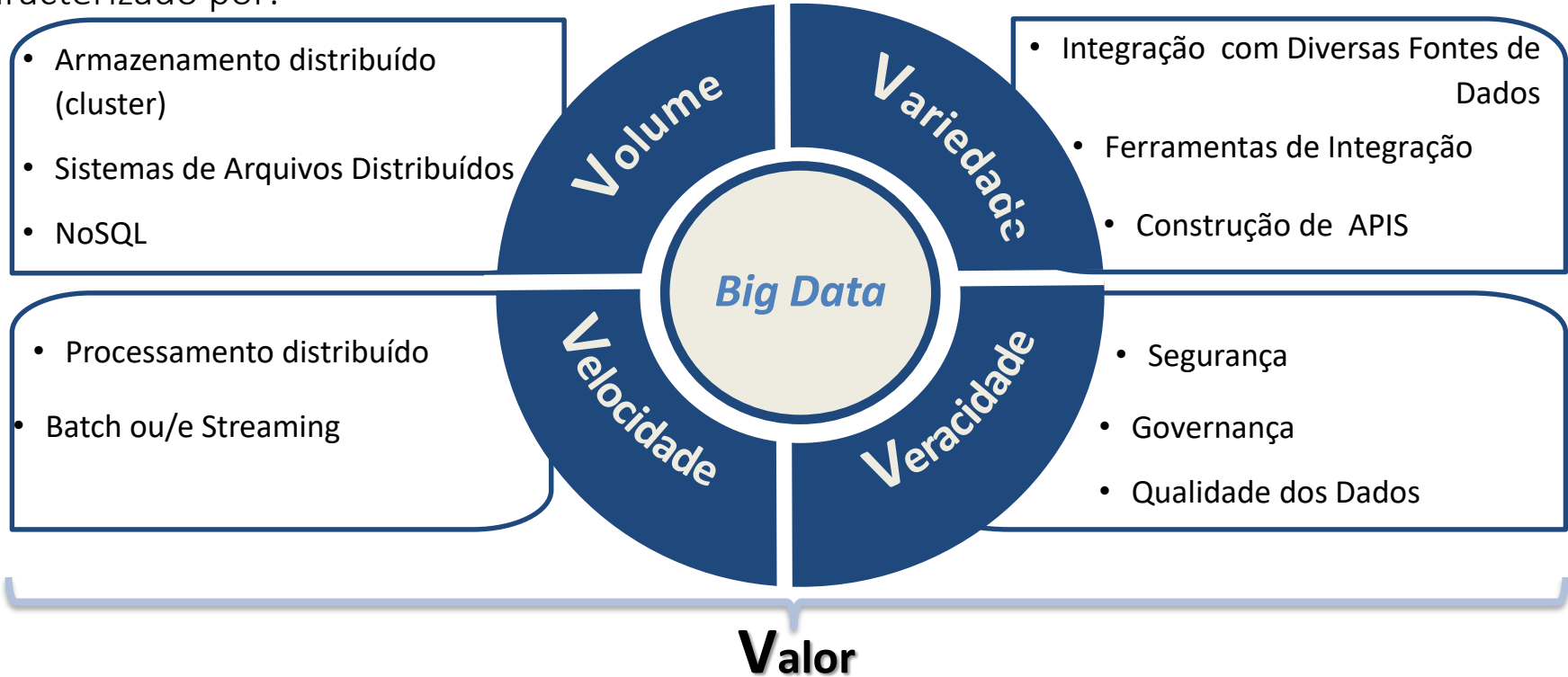
**Big Data**

# Os 5 Vs...



THE  
DEVELOPER'S  
CONFERENCE

Big Data é um conjunto de tecnologias que integradas formam uma solução que consegue lidar com o armazenamento, processamento e análises de dados que são caracterizado por:

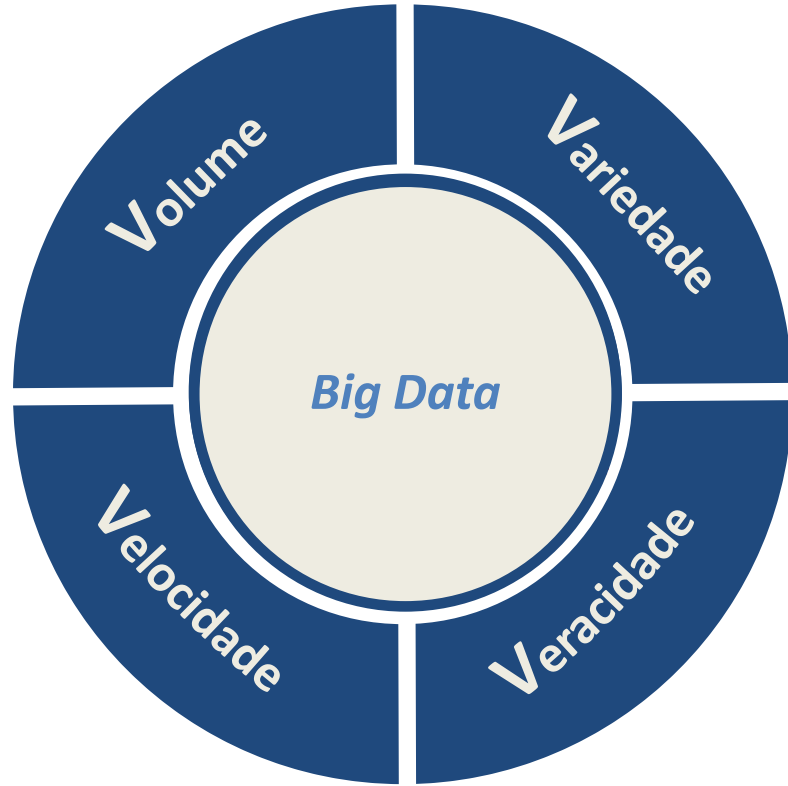




# Os 5 Vs...



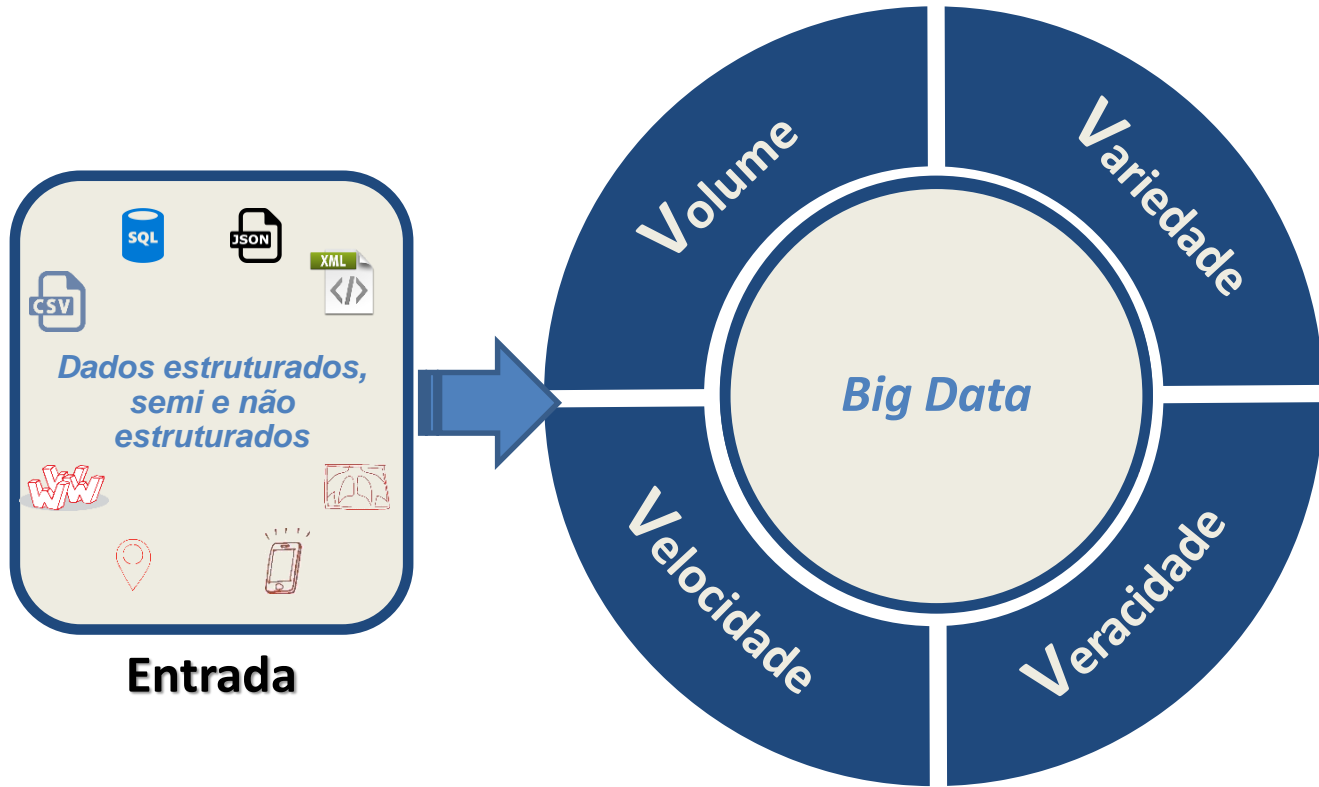
THE  
DEVELOPER'S  
CONFERENCE



# Os 5 Vs...



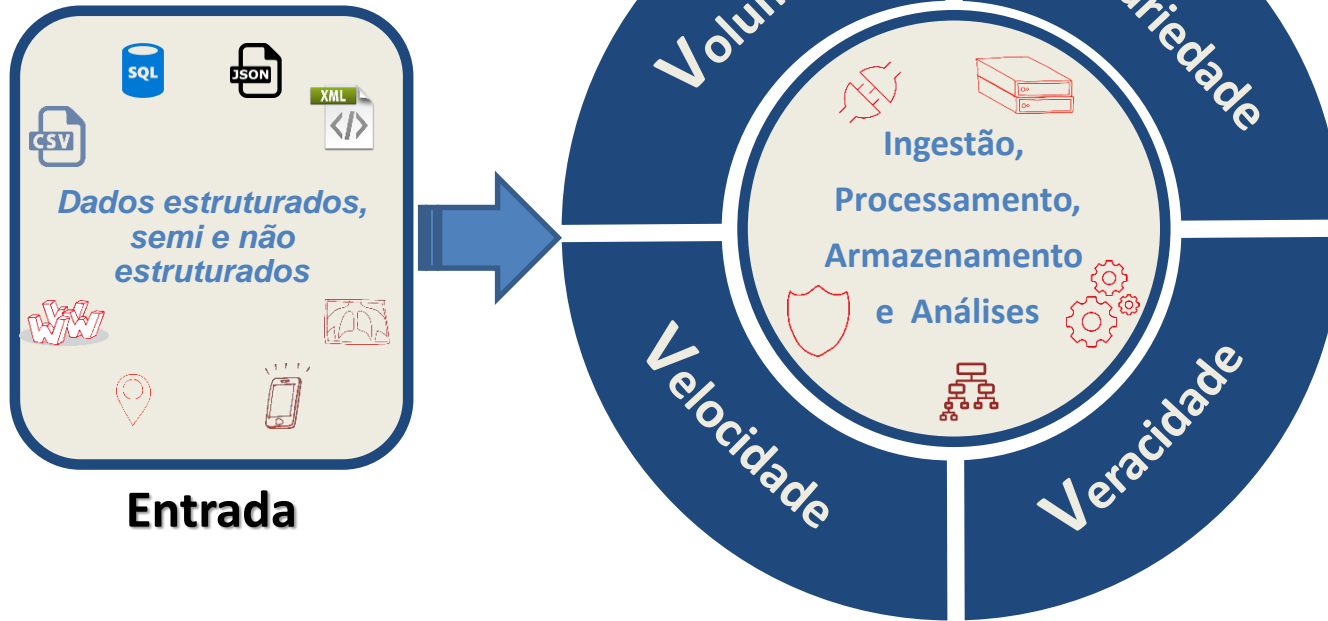
THE  
DEVELOPER'S  
CONFERENCE



# Os 5 Vs...



THE  
DEVELOPER'S  
CONFERENCE



# Os 5 Vs...



THE  
DEVELOPER'S  
CONFERENCE



**Entrada**

**Saída**

# Novo Cenário em Ambiente de Produção



## ➤ Integração de Fontes de Dados

- Dados Externos (APIS)
- Streaming de Dados

## ➤ Computação Distribuída

- Armazenamento
- Processamento

## ➤ Pipeline

## ➤ Segurança

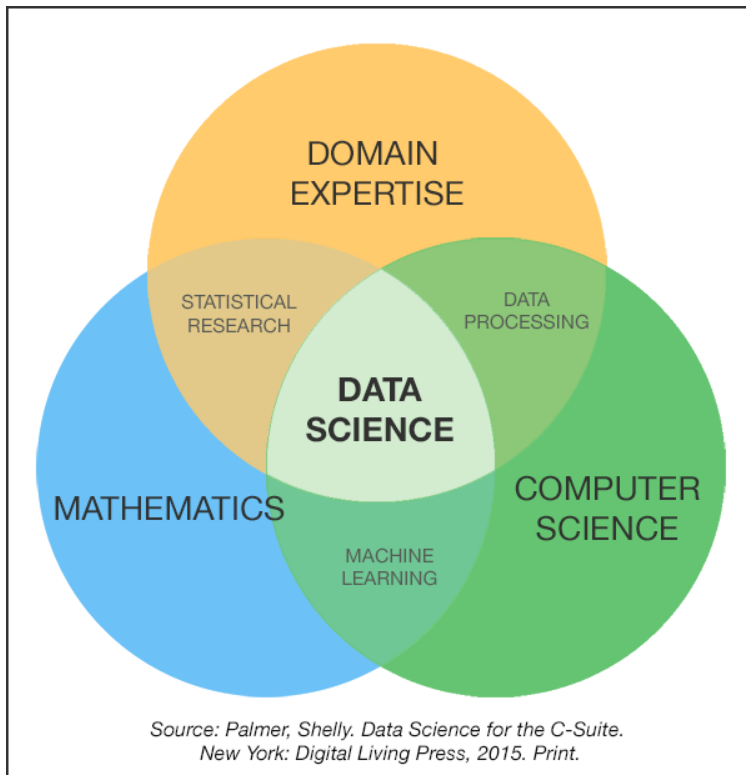
**Computação/  
Sistemas**



# Atuação do Cientista de Dados?



THE  
DEVELOPER'S  
CONFERENCE



# Saindo do Foco

- Matemática e Estatística Avançada
- Machine Learning, Inteligencia Artificial e Deep Learning
- Habilidades Analíticas Avançadas



# Cientistas fazendo o papel do Engenheiro de Dados...



THE  
DEVELOPER'S  
CONFERENCE

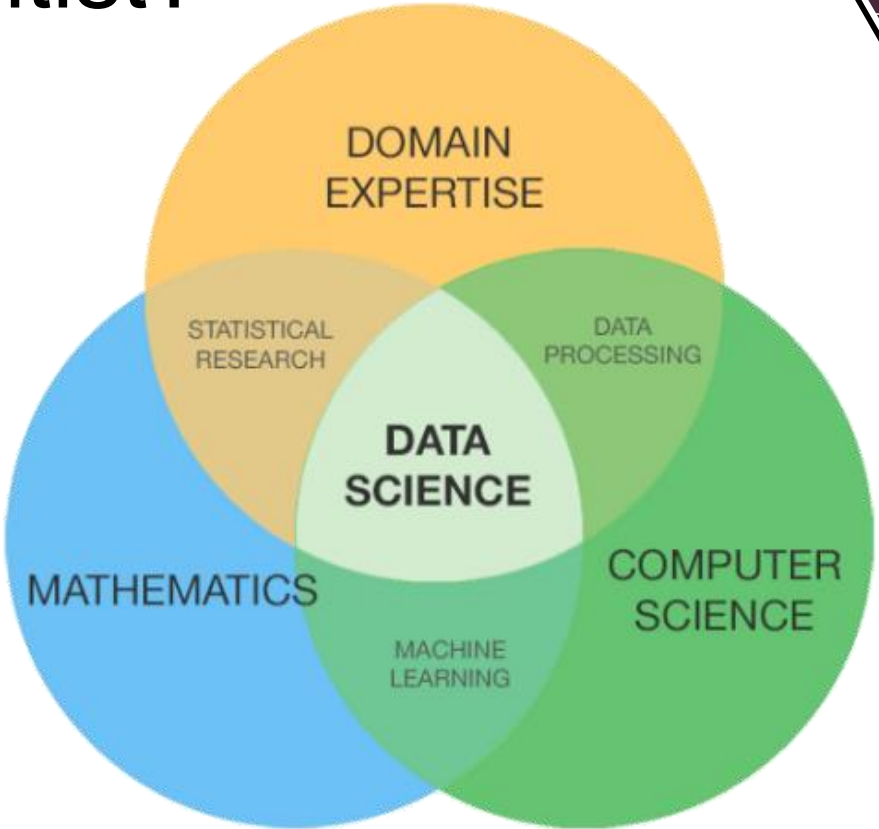
- Only **15 percent** of businesses reported **deploying their big data project to production**, effectively unchanged from last year (14 percent). – Gartner (2016)
- “I’ve seen companies task their data scientists with things you’d have a data engineer do. The **data scientists** were running at **20-30% efficiency**.” – Jesse Anderson (2018)
- In 2017, **24.0% of Data Scientists have changed job**.  
Christopher Bolard (2018)



# Data Scientist?



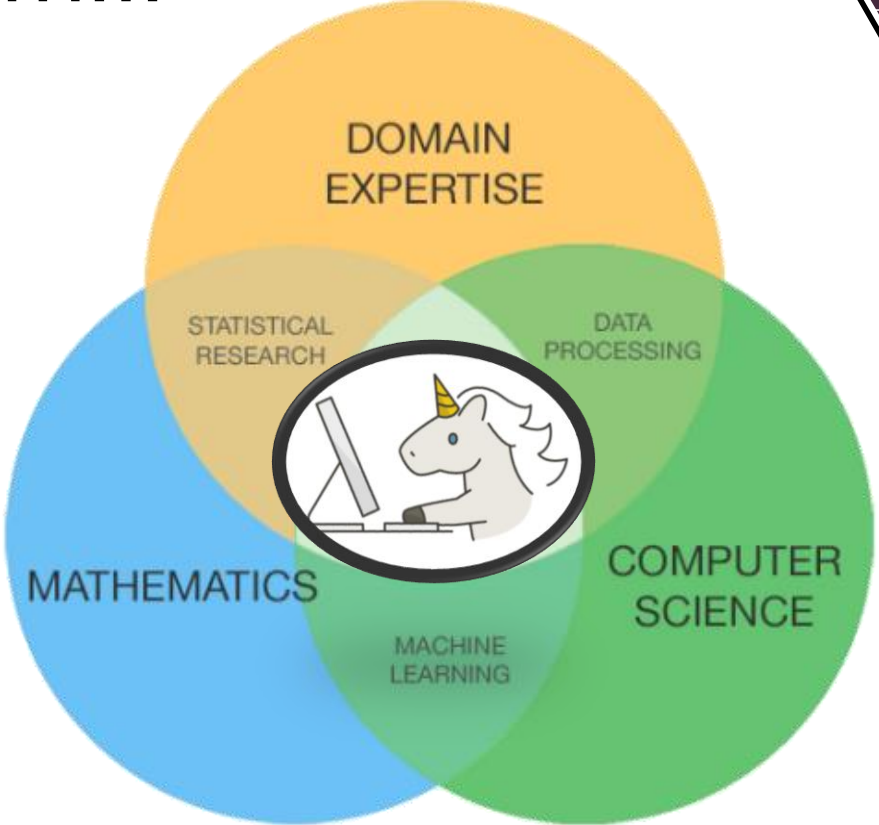
THE  
DEVELOPER'S  
CONFERENCE



# Data Unicorn!!!



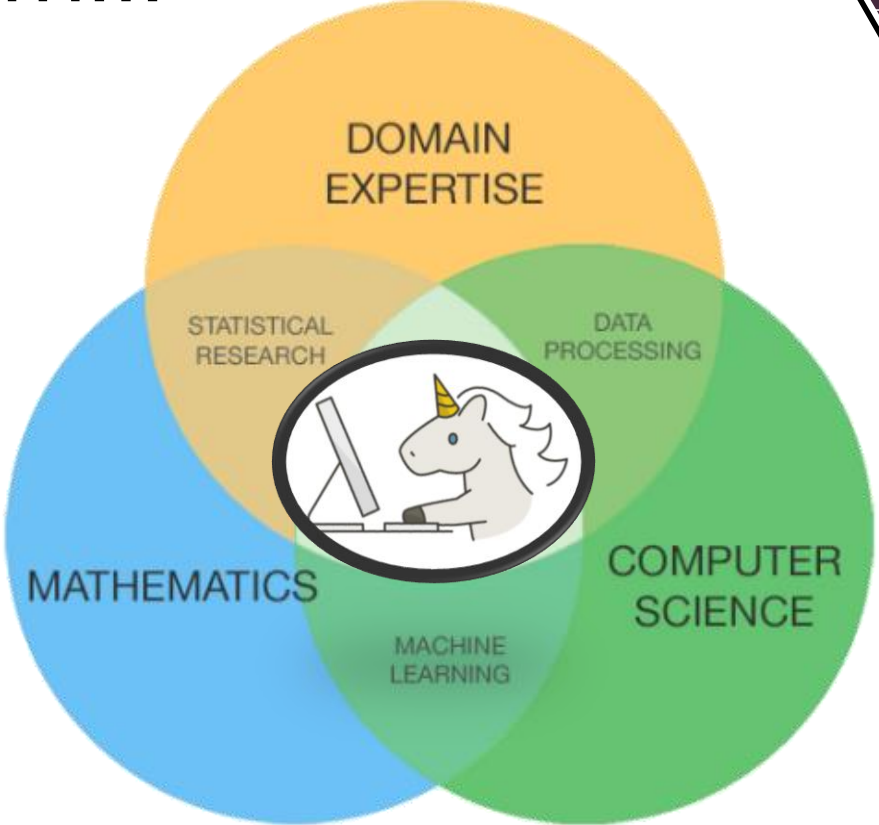
THE  
DEVELOPER'S  
CONFERENCE



# Data Unicorn!!!



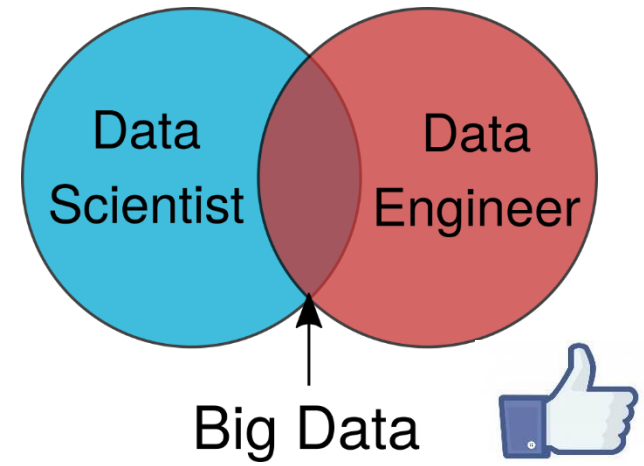
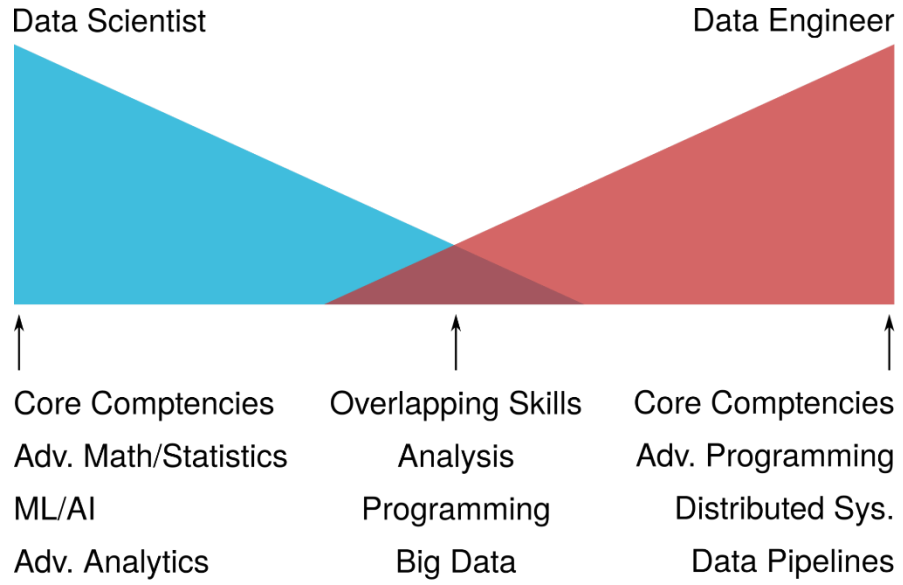
THE  
DEVELOPER'S  
CONFERENCE



# Ciência + Engenharia de Dados = Big Data



THE DEVELOPER'S CONFERENCE



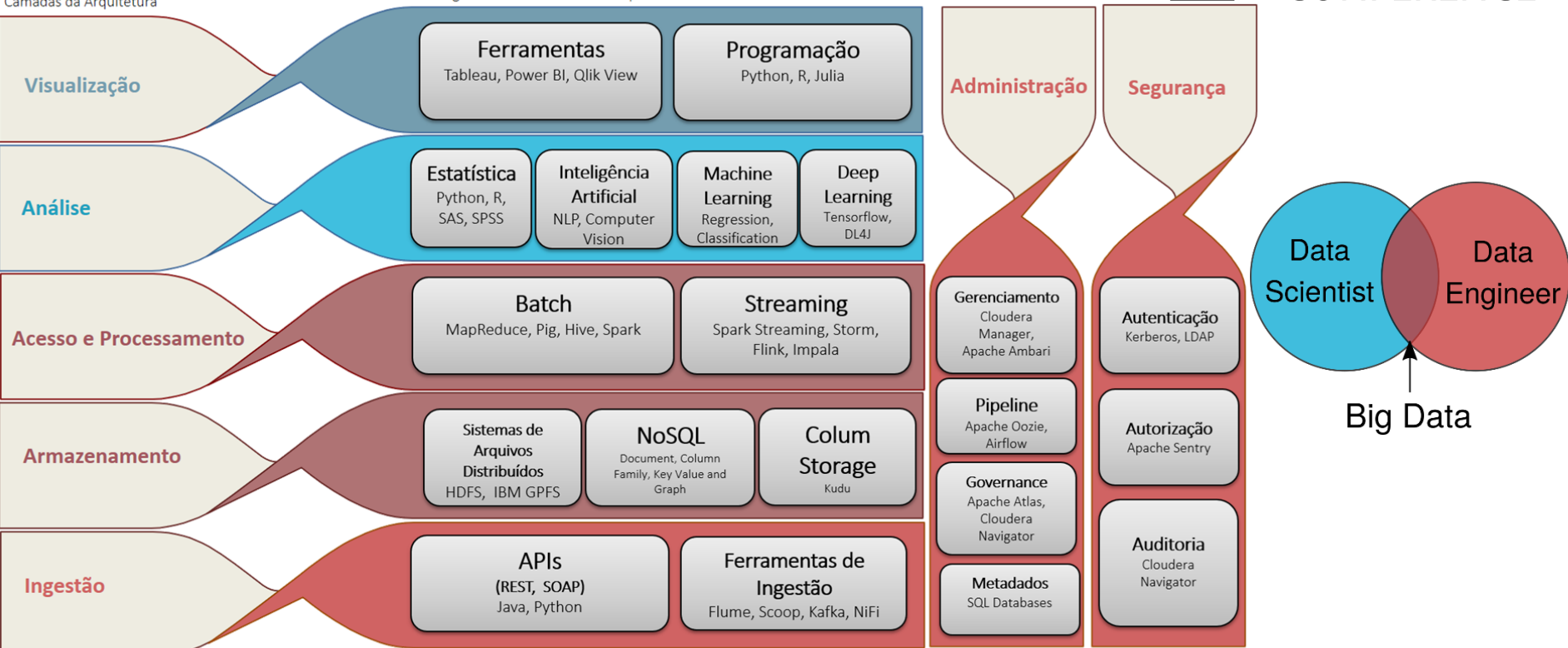
# Camadas da Arquitetura Big Data que cada um atua



THE DEVELOPER'S CONFERENCE

Camadas da Arquitetura

Tecnologias e Ferramentas de Exemplo



Dados Não Estruturados



Dados Semi-estruturados



Dados Estruturados



# Quem ganha mais?

NY



THE  
DEVELOPER'S  
CONFERENCE

➤ Lei da oferta e procura!!!

According to Glassdoor, the average salary of a Data Scientist in New York as of May **2016** 10th, 2016 is **\$108,659**.

According to Glassdoor, the average salary of a Data Engineer in New York as of March 10, 2016 is **\$95,526**.

## Data Scientist Salaries in New York City, NY Area

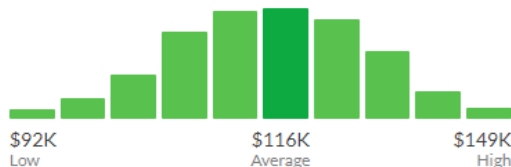
491 Salaries Updated Jul 14, 2019

Industries  Company Sizes  Years of Experience

Average Base Pay

**\$115,815** /yr

1% below national average



How much does a Data Scientist make in New York City, NY?

The average salary for a Data Scientist is \$115,815 in New York City, NY. Salaries estimates are based on 491 salaries submitted anonymously to Glassdoor by Data Scientist employees in New York City, NY. [Less](#)

**2019**

Additional Cash Compensation

Average **\$11,423**

Range **\$4,147 - \$26,854**

## Data Engineer Salaries in New York City, NY Area

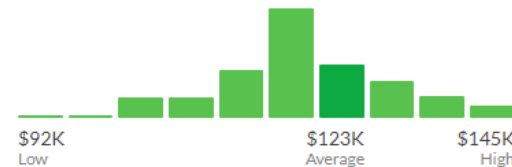
169 Salaries Updated Jul 11, 2019

Industries  Company Sizes  Years of Experience

Average Base Pay

**\$123,070** /yr

6% above national average



How much does a Data Engineer make in New York City, NY?

The average salary for a Data Engineer is \$123,070 in New York City, NY. Salaries estimates are based on 169 salaries submitted anonymously to Glassdoor by Data Engineer employees in New York City, NY. [Less](#)

Additional Cash Compensation

Average **\$10,650**

Range **\$2,903 - \$25,432**

# Quem ganha mais?

➤ Lei da oferta e procura!!!

SP  
2019



THE  
DEVELOPER'S  
CONFERENCE

## Data Scientist Salaries in Sao Paulo, Brazil Area

66 Salaries Updated Jul 15, 2019

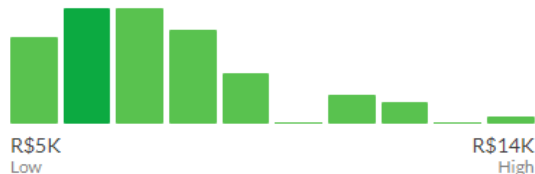
Industries

Company Sizes

Years of Experience

Average Base Pay

R\$7,725 /mo



Additional Cash Compensation ?

Average R\$20,867

Range R\$4,618 - R\$39,193

How much does a Data Scientist make in Sao Paulo, Brazil?  
The average salary for a Data Scientist is R\$7,725 in Sao Paulo, Brazil. Salaries estimates are based on 66 salaries submitted anonymously to Glassdoor by Data Scientist employees in Sao Paulo, Brazil. [Less](#)

## Data Engineer Salaries in Sao Paulo, Brazil Area

26 Salaries Updated Jul 11, 2019

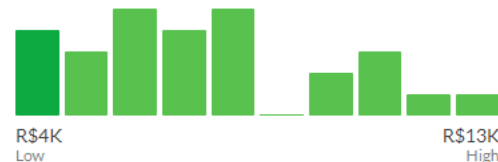
Industries

Company Sizes

Years of Experience

Average Base Pay

R\$6,936 /mo



Additional Cash Compensation ?

Average R\$38,750

Range R\$17,594 - R\$126,536

How much does a Data Engineer make in Sao Paulo, Brazil?

The average salary for a Data Engineer is R\$6,936 in Sao Paulo, Brazil. Salaries estimates are based on 26 salaries submitted anonymously to Glassdoor by Data Engineer employees in Sao Paulo, Brazil. [Less](#)

# Quem ganha mais?

➤ Lei da oferta e procura!!!

Brasil  
2019



THE  
DEVELOPER'S  
CONFERENCE

## Salários de Data Scientist

96 Salários Atualizado em 15 de jul 2019

Salário médio

R\$ 7.725 /mês

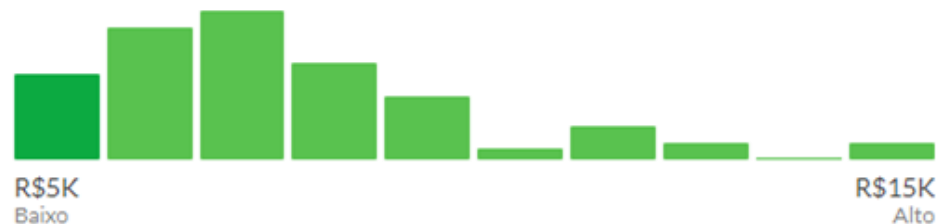
## Salários de Data Engineer

25 Salários Atualizado em 11 de jul 2019

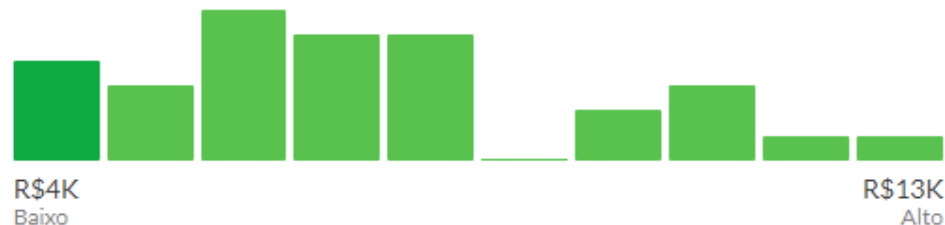
Salário médio

R\$ 7.160 /mês

Confiança  
muito alta



Confiança  
muito alta

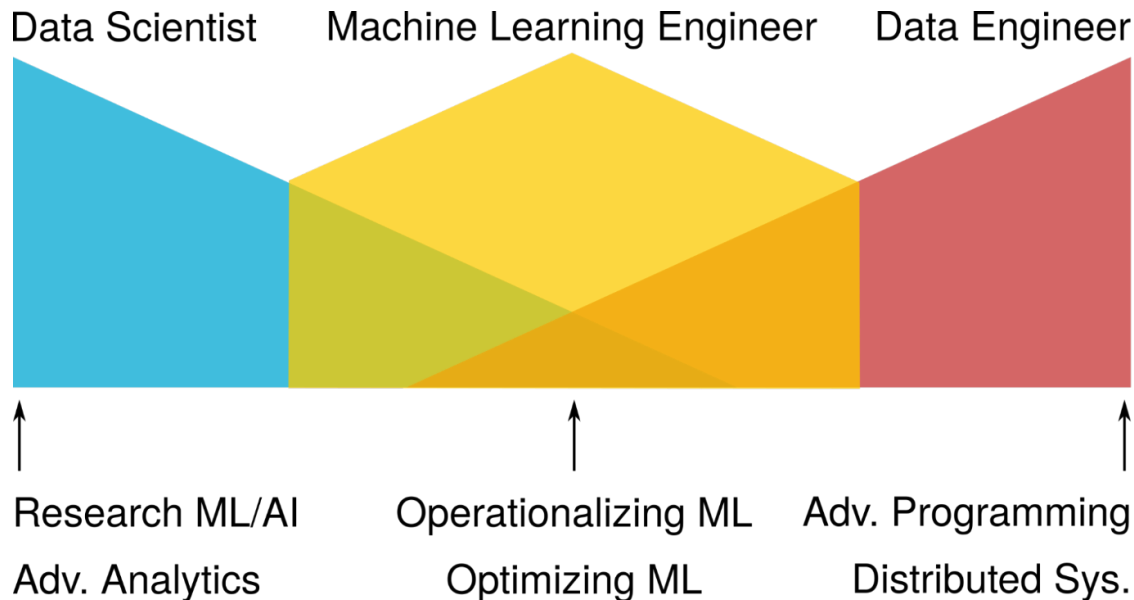




# Novo Perfil: Engenheiro de Machine Learning



THE  
DEVELOPER'S  
CONFERENCE



BDI

BIG DATA INSTITUTE

For more information go to <http://bigdatainstitute.io>

Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) Licensed

# Engenheiro de Machine Learning



THE  
DEVELOPER'S  
CONFERENCE

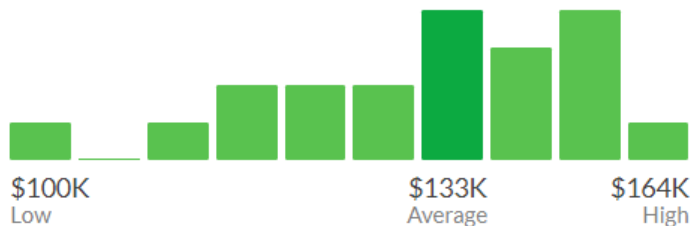
## Machine Learning Engineer Salaries in New York City, NY Area

19 Salaries Updated Jul 15, 2019

Industries  Company Sizes  Years of Experience

Average Base Pay

**\$132,500** /yr



Additional Cash Compensation [?](#)

Average \$7,865

Range \$4,214 - \$20,809

How much does a Machine Learning Engineer make in New York City, NY?

The average salary for a Machine Learning Engineer is \$132,500 in New York City, NY. Salaries estimates are...

[More](#)

# Dicas: Por onde começo a estudar?



THE  
DEVELOPER'S  
CONFERENCE

## Ciência de Dados

- Statistics 101 <https://cognitiveclass.ai/courses/statistics-101/>
- Introduction to Data Science <https://cognitiveclass.ai/courses/data-science-101/>
- Data Science Methodology <https://cognitiveclass.ai/courses/data-science-methodology-2/>
- Python for Data Science - <https://cognitiveclass.ai/courses/python-for-data-science/>
- Data Analysis with Python <https://cognitiveclass.ai/courses/data-analysis-python/>
- Machine Learning with Python <https://cognitiveclass.ai/courses/machine-learning-with-python/>
- Data Visualization with Python - <https://cognitiveclass.ai/courses/data-visualization-with-python/>
- Deep Learning - <https://cognitiveclass.ai/courses/introduction-deep-learning/>

## Engenharia de Dados

- Big Data - <https://cognitiveclass.ai/courses/what-is-big-data/>
- Hadoop - <https://cognitiveclass.ai/courses/introduction-to-hadoop/>
- Hbase - <https://cognitiveclass.ai/courses/using-hbase-for-real-time-access-to-your-big-data/>
- Hive - <https://cognitiveclass.ai/courses/hadoop-hive/>
- Flume/Scoop - <https://cognitiveclass.ai/courses/flume-sqoop-moving-data-into-hadoop/>
- Kafka - <https://cognitiveclass.ai/courses/simplifyingdatapipelines/>
- Spark I - <https://cognitiveclass.ai/courses/what-is-spark/>
- Spark II - <https://cognitiveclass.ai/courses/spark-rdd/>
- Spark Streaming I - <https://www.udemy.com/apache-spark-streaming-with-python-and-pyspark/>
- Solr - <https://cognitiveclass.ai/courses/introduction-to-solr/>
- Oozie - <https://cognitiveclass.ai/courses/controlling-hadoop-jobs-using-oozie/>

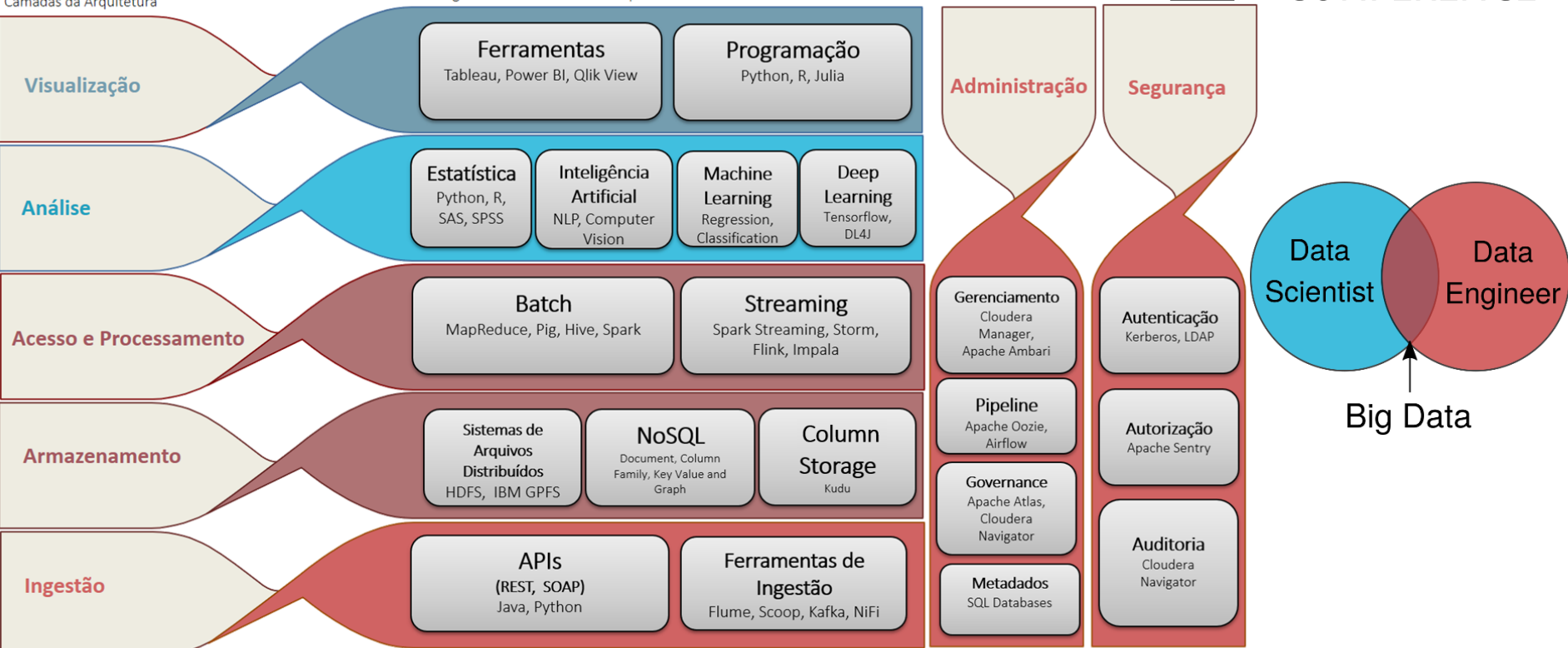
# Camadas da Arquitetura Big Data que cada um atua



THE DEVELOPER'S CONFERENCE

Camadas da Arquitetura

Tecnologias e Ferramentas de Exemplo



Dados Não Estruturados



Dados Semi-estruturados



Dados Estruturados



# Quantos engenheiros e cientistas em um projeto?



THE  
DEVELOPER'S  
CONFERENCE

- “A common starting point is 2-3 data engineers for every data scientist. For some organizations with more complex data engineering requirements, this can be 4-5 data engineers per data scientist. This includes organizations where data engineering and data science are in different reporting structures. You need more data engineers because more time and effort is needed to create data pipelines than to create the ML/AI portion.” – Jesse Anderson (2018)

# Case Real LOGICALIS

Otimização de Rotas de Caminhões de Lixo (Granada –ES)



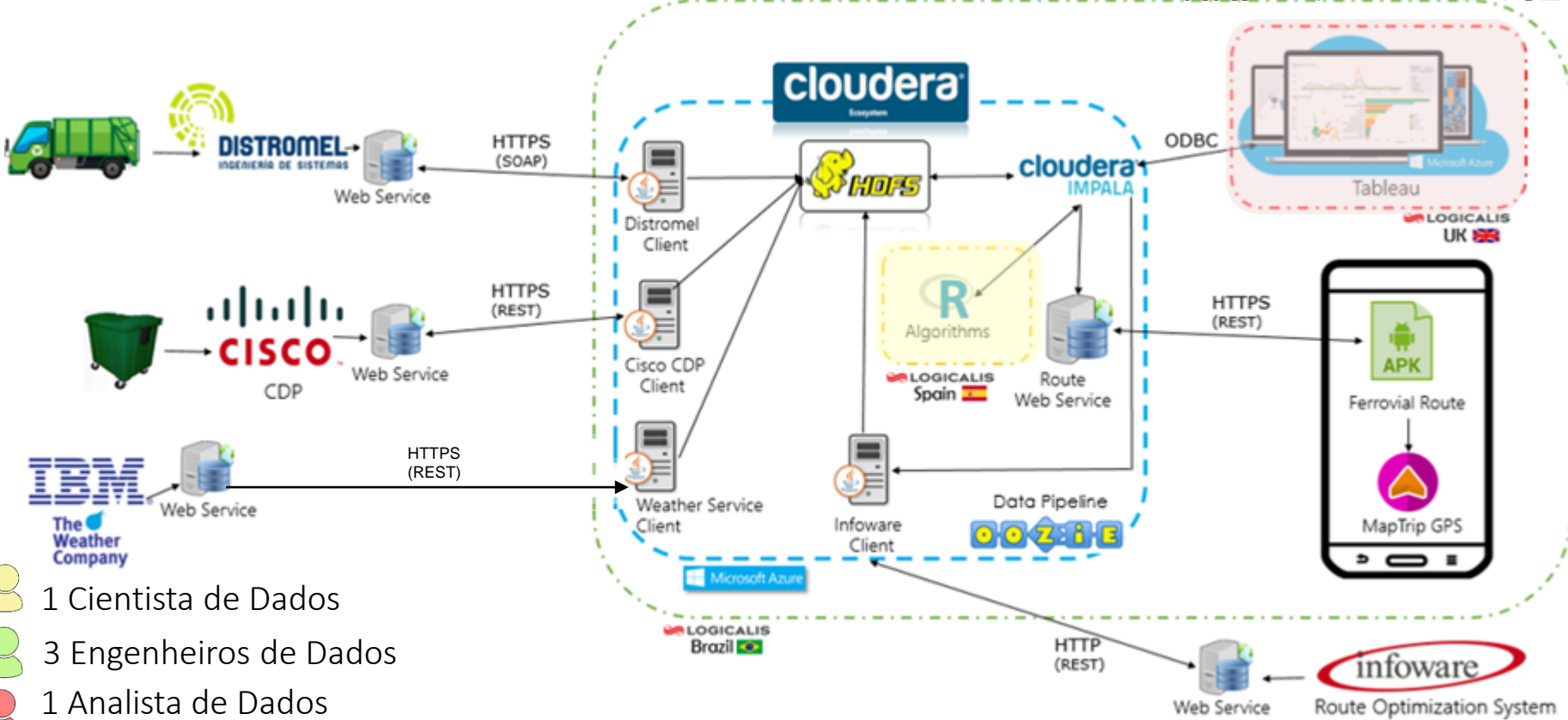
THE  
DEVELOPER'S  
CONFERENCE




- Evitar passar em lixeiras vazias
- Evitar super super lotação das lixeiras
- Criar rotas inteligentes



# Case Real LOGICALIS

Otimização de Rotas de Caminhões de Lixo (Granada –ES)



-  1 Cientista de Dados
-  3 Engenheiros de Dados
-  1 Analista de Dados

# Case Real LOGICALIS

Otimização de Rotas de Caminhões de Lixo (Granada –ES)



**30%** KM rodados a menos

**35%** a menos de tempo gasto

**35%** a menos de lixeiras visitadas

**30%** a menos de custo operacional

**30%** a menos de emissões de CO2





# E tudo isso só foi possível porque?



THE  
DEVELOPER'S  
CONFERENCE

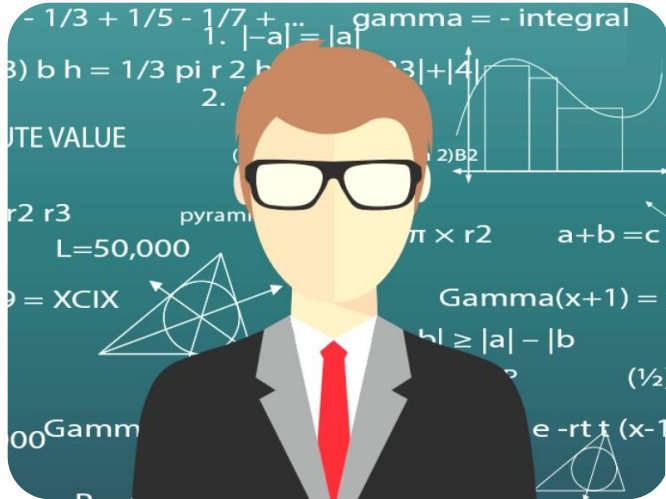


Working better together

# E tudo isso só foi possível porque?



THE  
DEVELOPER'S  
CONFERENCE



Ciência de Dados

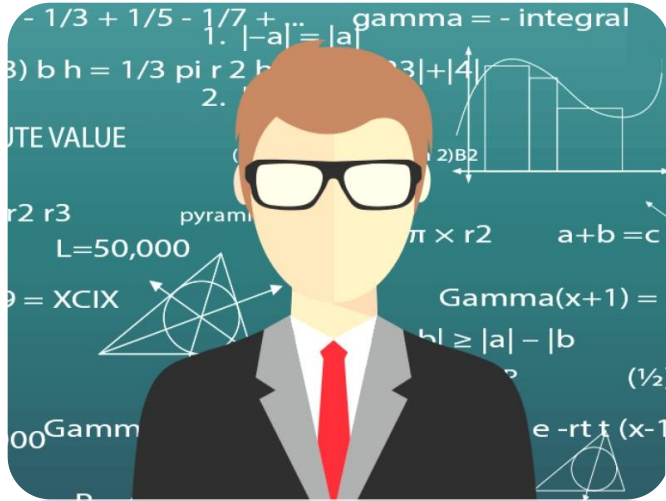


Working better together

# E tudo isso só foi possível porque?



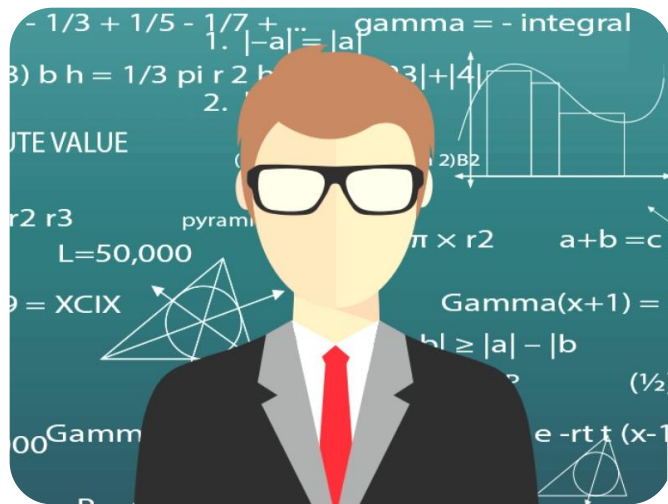
THE  
DEVELOPER'S  
CONFERENCE



Ciência de Dados



# E tudo isso só foi possível porque?



Ciência de Dados



Engenharia de Dados



# Cientistas e Engenheiros de Dados



THE  
DEVELOPER'S  
CONFERENCE

- Não há um perfil mais importante que o outro, eles são complementares: só existe projeto de Big Data de sucesso se existir, pelo menos, os dois!



# Dúvidas?



THE  
DEVELOPER'S  
CONFERENCE



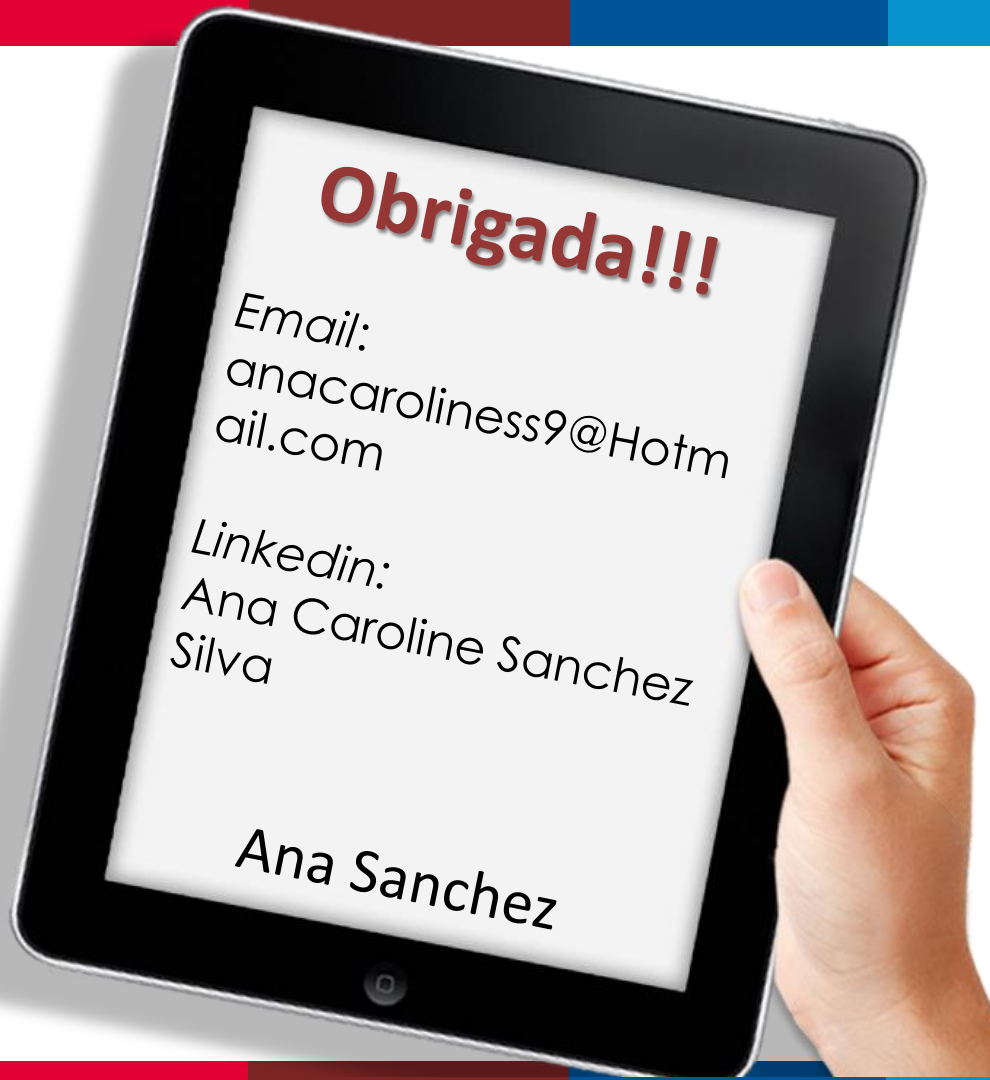
# Referências



THE  
DEVELOPER'S  
CONFERENCE

- › Data Scientist: The Sexiest Job of the 21st Century [https://edisciplinas.usp.br/pluginfile.php/4561873/mod\\_resource/content/1/Data%20Scientist%20The%20Sexiest%20Job%20of%20the%2021st%20Century.pdf](https://edisciplinas.usp.br/pluginfile.php/4561873/mod_resource/content/1/Data%20Scientist%20The%20Sexiest%20Job%20of%20the%2021st%20Century.pdf)
- › Gartner Survey Reveals Investment in Big Data Is Up but Fewer Organizations Plan to Invest <https://www.gartner.com/en/newsroom/press-releases/2016-10-04-gartner-survey-reveals-investment-in-big-data-is-up-but-fewer-organizations-plan-to-invest>
- › Top 5 Reasons Most Big Data Projects Never Go Into Production <https://www.infoworks.io/top-5-reasons-big-data-projects-never-go-production/>
- › Professional Certificate in Data Science <https://online-learning.harvard.edu/series/professional-certificate-data-science>
- › Data Science Specialization <https://www.coursera.org/specializations/jhu-data-science>
- › Applied Data Science with Python Specialization <https://www.coursera.org/specializations/data-science-python>
- › MicroMasters® Program in Data Science <https://www.edx.org/micromasters/uc-san-diegox-data-science>
- › Data Engineer VS Data Scientist <https://towardsdatascience.com/data-engineer-vs-data-scientist-bc8dab5ac124>
- › Data engineer vs. Data scientist- What does your company need? <https://www.dezyre.com/article/data-engineer-vs-data-scientist-what-does-your-company-need/262>
- › Data engineers vs. data scientists <https://www.oreilly.com/ideas/data-engineers-vs-data-scientists>
- › Ferrovial Services e Logicalis - uma parceria tecnológica (Otimização de rotas de caminhões de lixos) <https://www.youtube.com/watch?v=rHbhVoh0vY>
- › Learn How to Create and Manage Big Data Teams <https://www.bigdatainstitute.io/books/data-engineering-teams-book/>
- › Salários de Cientistas e Engenheiros de Dados <https://www.glassdoor.com.br/index.htm>





THE  
DEVELOPER'S  
CONFERENCE