



# Formação Engenheiro de Dados

Clássico Versus Big Data

# Analisar Dados



DATAWAREHOUSE  
TRADICIONAL NÃO É  
SUBSTITUÍDO PELO  
DATALAKE



NOSQL NÃO SUBSTITUI  
MODELO RELACIONAL



TECNOLOGIA DIFERENTES  
PARA PROBLEMAS  
DIFERENTES!

# Comparação

VS



Clássico



Big Data

## Velocidade

D+I



Streaming / Batch

## Volume

Terabytes



Petabytes

## Variedade

Dados Estruturados



Dados não Estruturados,  
Semi-Estruturados e  
Estruturados

## Arquitetura

Centralizada/Vertical



Distribuída ou  
Horizontal



**PROF.  
FERNANDO  
AMARAL**  
[www.datascientist.com.br](http://www.datascientist.com.br)

# Comparação

VS



Clássico

Big Data

## Tráfego de Dados

Cliente - Servidor



Entre Servidores

## Cargas

Dados Limpos  
e de Valor



Todo Dado  
Coletados

## Processo de Carga

ETL



ELT

## Hardware

Especializado/  
Appliance



Commodity



**PROF.  
FERNANDO  
AMARAL**  
[www.datascientist.com.br](http://www.datascientist.com.br)



# Data Lakes

- ◇ Armazém de Dados Corporativo
- ◇ Pode ser sobre HDFS
- ◇ Dados são carregados no formato nativo
- ◇ Pode ter áreas de:
  - ◇ Descarga (formato nativo)
  - ◇ Consumo (tratados)
  - ◇ Cópia Temporal

# Data Lakes

