



COMO O **BIG DATA** PODE AJUDAR  
NA TRANSFORMAÇÃO DIGITAL DO  
SEU NEGÓCIO?

<b>Introdução .....</b>	<b>Pg. 3</b>
<b>O que é Big Data? .....</b>	<b>Pg. 5</b>
<b>Arquitetura do Big Data .....</b>	<b>Pg. 7</b>
<b>Dados estruturados x não estruturados .....</b>	<b>Pg. 8</b>
<b>Como o Data Lake atua no Big Data .....</b>	<b>Pg. 9</b>
<b>Solução ou Problema? .....</b>	<b>Pg. 11</b>
<b>Big Data não está distante da nossa realidade .....</b>	<b>Pg. 14</b>
<b>Conclusão .....</b>	<b>Pg. 17</b>
<b>Sobre a Programmer's .....</b>	<b>Pg. 19</b>

# INTRODUÇÃO

Os dados vêm mudando o mundo que conhecemos. Somos afetados por eles em tudo o que fazemos, deixando um rastro digital simplesmente ao carregar um celular no bolso.

Todo dia, a cada minuto, onde estamos, o que vemos, acessamos e em tudo o que fazemos: dados são gerados. E este volume só cresce com as mídias sociais, máquinas e transações.

Um **estudo** prevê que até 2020, 1,7 MB de dados serão criados por segundo para cada pessoa conectada no mundo. Resumindo, esta é a era do **Big Data**.

Por isso, não é à toa que os dados estão se tornando um dos principais ativos de negócios. Eles estão no centro de muitas inovações tecnológicas transformadoras, incluindo análise preditiva, inteligência artificial, aprendizado de máquina e Internet das Coisas. Afinal, o que muitas empresas já entenderam é que os dados podem dizer - e muito - sobre o comportamento humano e, ainda, **extrair valor para que se desenvolva um novo mercado ou melhorar o atendimento.**

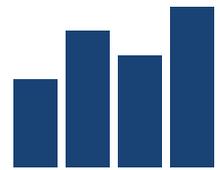
Neste e-book, vamos ajudar você a descobrir como um grande volume de dados bem estruturado pode ajudar sua organização a ter mais sucesso e impulsionar o desempenho nos negócios.

**Boa leitura!**

## O QUE É BIG DATA?

De um modo geral, Big Data é um conceito amplo, mas, em resumo, podemos dizer ser um conjunto de dados com um volume tão grande ao ponto que você não conseguiria gerenciá-los em uma base de dados relacional.

Para tanto, o conceito possui características bem específicas, conhecidas como **5Vs**:



**VOLUME:** A inimaginável quantidade de dados gerados a cada segundo e também dados já criados que precisam ser analisados;



**VELOCIDADE:** A velocidade na qual os dados precisam gerar algum valor, como na detecção de fraudes com cartões de crédito;



**VARIEDADE:** Os mais diferentes tipos e fontes de dados - desde dados financeiros (estruturados) até gravação de voz (não estruturados);



**VERACIDADE:** A confiabilidade dos dados, como postagens em mídias sociais, onde a qualidade ou precisão são menos confiáveis;



**VALOR:** Podemos dizer que este é o “V” mais importante, pois os dados só são úteis se você consegue extrair valor deles.

# ARQUITETURA DO BIG DATA

## CAMADA DE ARMAZENAMENTO E TRATAMENTO

## CAMADA DE CONSUMO

### Análises Rápidas (Speed Layer)



#### FERRAMENTAS:

Azure Stream Analytics  
Azure IoT Hub  
Spark Streaming  
Storm

### Análises Históricas (Batch Layer)



#### Data Lake

#### FERRAMENTAS:

Azure Data Factory  
HDInsight  
Hadoop  
Hive  
Azure DataBricks



#### Análises em tempo real

## CAMADA DE ANÁLISE



#### Banco de dados



#### Cubos



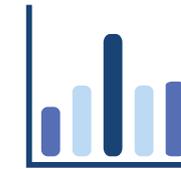
#### Machine Learning

#### FERRAMENTAS:

Azure Blob Storage  
Amazon DynamoDB  
Presto  
Cassandra  
Apache HBase  
Cosmos DB  
Azure Analysis Services



#### Análises Preditivas



Visualizações estratégicas  
Alertas diários  
Campanhas de marketing



Análise de sentimento  
Análise de crédito  
Segmentação de cliente  
Otimização de estoque

#### FERRAMENTAS:

Power BI  
Tableau  
QlikView  
Metabase

# DADOS ESTRUTURADOS X NÃO ESTRUTURADOS

Antes do boom do Big Data, apenas os dados que se encaixassem perfeitamente em tabelas ou bancos de dados tradicionais eram aproveitados e considerados de real valor. Com o advento da internet e das mídias sociais, um volume crescente de dados estruturados e não estruturados surgiu e, junto com eles, sua complexidade para organizá-los.

## **DADOS ESTRUTURADOS:**

São os dados que contêm uma organização para serem recuperados, que identificam diversos pontos sobre uma informação e tornam o trabalho da tecnologia mais simples. Ou seja, são informações armazenadas em CRM, como sistemas financeiros, de RH, entre outros.

## **DADOS NÃO ESTRUTURADOS:**

Esses dados precisam de intervenção humana para sua preparação, pois são mais complexos de trabalhar. Dentre eles, estão os dados de mídias sociais, como imagens, vídeos, textos e até áudios.

# COMO O DATA LAKE ATUA NO BIG DATA

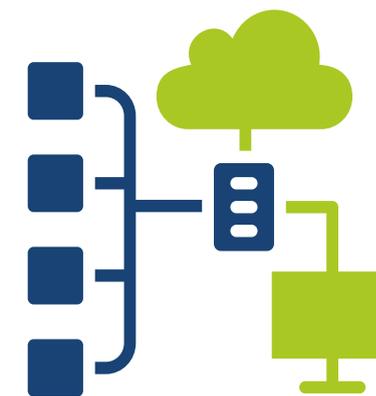
Data Lake - em tradução livre, “lago de dados” -, é um ambiente de armazenamento e processamento que permite reunir em um único local todos os dados, sejam eles estruturados ou não. Existem algumas variações de como organizar os dados no Data Lake, mas frequentemente são observadas as seguintes zonas:

- 1 LANDED OR TEMPORARY ZONE:** Esta é uma zona transitória onde o dado é trazido apenas para ser transportado para as outras zonas ou para algum outro fim (streaming, por exemplo);
- 2 RAW ZONE:** Nesta zona os dados são armazenados de forma crua, ou seja, da forma que são capturados com alguma técnica de compactação, que mantém boa parte das características originais da informação e economiza espaço em armazenamento;

**3 HISTORY ZONE:** Zona responsável por manter versões (quando aplicável) sobre os dados e facilitar seu consumo para os analistas e cientistas de dados;

**4 GOLD OR TRUSTED ZONE:** Esta zona contém dados transformados prontos para consumo de usuários finais e qualquer outro perfil que tenha interesse sobre a informação.

Estruturar um Data Lake assim como um ambiente de Big Data demanda da instituição diversos cuidados que serão devidamente abordados nos artigos futuros. Acompanhe as nossas postagens!



# SOLUÇÃO OU PROBLEMA?

Ter um grande volume de dados em mãos nem sempre é uma solução. Na maioria das vezes, pode ser um problema e o desafio está justamente em gerenciá-los.

Uma pesquisa realizada em 2018, pela NewVantage Partners, mostrou que, nos últimos cinco anos, 95% das empresas pesquisadas iniciaram um projeto em Big Data e que apenas 48,4% conseguiram “resultados mensuráveis” com seus investimentos.

Por isso, confira cinco desafios operacionais que a maioria das organizações enfrentam no dia a dia e que a fazem progredir na implementação de projetos com Big Data:



**ARMAZENAMENTO E PROCESSAMENTO:** Antes de pensar em coletar os dados, primeiro é preciso preparar o terreno: a infraestrutura de armazenamento e processamento. É preciso escolher o serviço especializado em gerenciamento de dados complexos, de alta performance, como a utilização de nuvem.

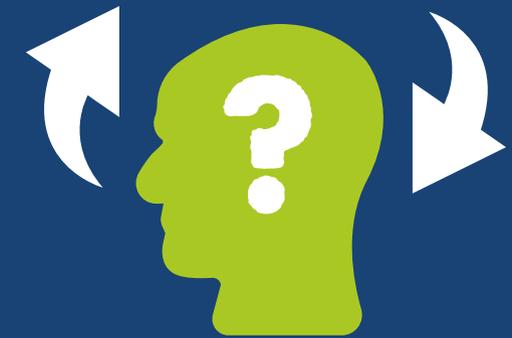
**ANÁLISE E CURADORIA DE DADOS:** É preciso realizar um pré-processamento para que os dados comecem a ser analisados.

Por exemplo, construir estratégias de negócios e buscar modelos de banco de dados, respondendo às perguntas centrais para estruturar, ordenar e interpretar esses dados.



**TIMES CAPACITADOS E PREPARADOS:** Para analisar Big Data, é preciso ter uma equipe multidisciplinar, que entenda de linguagens de programação em conjunto com matemática e estatística. Devido aos diversos pontos que dependem de dados nas organizações, muitas adotam a estratégias de SQUADS multidisciplinares para conseguir entregar valor de forma ágil e com maior qualidade.

**MAIS AGILIDADE NAS TOMADAS DE DECISÃO:** Os insights para as mudanças das análises dos dados precisam ser mais rápidos, acompanhando a velocidade de coleta de milhões de dados em segundos. Por isso, é preciso agir na mesma velocidade que os dados são gerados.

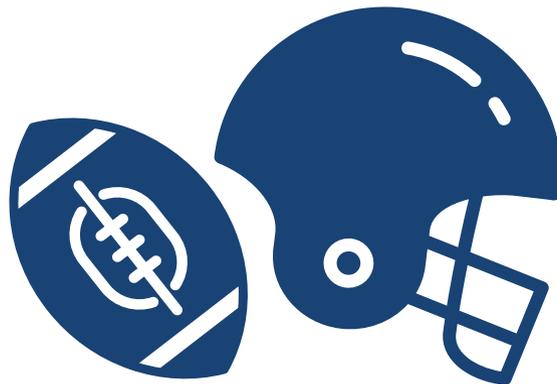


**PRIVACIDADE E SEGURANÇA:** Proteger a identidade dos usuários e identificar para qual fim esses dados serão usados são alguns dos questionamentos sobre a questão da segurança dos dados. É preciso assegurar que todos os seus dados, estruturados ou não, estão armazenados de forma segura.

# BIG DATA NÃO ESTÁ DISTANTE DA NOSSA REALIDADE

Os jogos ou os comerciais do Super Bowl atraem um público de todas as idades e em todo mundo, e toda essa emoção produz uma montanha de dados. Por isso, o **Super Bowl está usando o Big Data para mudar totalmente o jogo.**

Os sensores nos estádios e nos capacetes ajudam a coletar dados em tempo real, mostrando as mudanças de posição dos jogadores e até quando um jogador sofreu um golpe na cabeça.



**Os analistas esportivos mantêm o controle desde estatísticas dos jogadores individualmente e das equipes, até da distância em tempo real percorrida, posição em campo e como as condições climáticas afetam os jogadores.**

**Mas a coleta de dados vai além disso. Eles também coletam dados de diversas plataformas, inclusive nas escutas de mídias sociais, orientando os anunciantes a ter insights relacionados ao que as pessoas falam antes, durante e depois dos jogos.**



Com uma equipe antenada em data science, **a Netflix é focada em transformar dados em conteúdo criativo e lucrativo.**

Além de produzir um ótimo serviço recomendando filmes e séries que você pode gostar por meio dos dados coletados, eles também fazem um estudo de hábitos e de análise preditiva que orientam a produção de novos conteúdos originais.

Um exemplo é a análise dos dados de consumo em 2011, em que os analistas perceberam que os filmes de Kevin Spacey, os longas de David Fincher e uma minissérie britânica sobre política dos anos 90 faziam muito sucesso. Baseada nesse engajamento, a Netflix lançou a série original House of Cards em 2013, que foi muito bem-sucedida entre o público, chegando à sua sexta temporada em 2017.

# NETFLIX

## CONCLUSÃO

O ecossistema de **BIG DATA** evolui em um ritmo impressionante e pode ser empregado em qualquer área de negócio. O ponto crucial é ter a estrutura adequada para suportar este volume que só tende a crescer.

Mas uma coisa é certa, o **BIG DATA** teve e tem um papel importante na transformação digital de diversos negócios que o utilizaram muitas vezes para propósitos diferentes. As restrições até existem, mas mais por conta da nossa criatividade na hora de usar os dados e das perguntas que a eles fazemos, conseguimos extrair resultados fantásticos.

Que tal dar o primeiro passo? A Programmer's desenvolveu a solução **PILOTO EM CIÊNCIA DE DADOS** para identificar a viabilidade do seu projeto e, posteriormente, desenhar um roadmap completo para desenvolver modelos analíticos mais robustos e integrados às soluções corporativas ou aos produtos digitais das empresas.

Será que não está na hora de extrair todo o valor que os dados da sua empresa têm?



# **SOBRE A PROGRAMMER'S**

A Programmer's é uma empresa que desenvolve soluções de negócio e leva inovação, tecnologia especializada em desenvolvimento e inteligência de dados.

Nos últimos 30 anos, auxiliamos grandes marcas em sua transformação digital, acumulando expertise no desenvolvimento ágil por meio de tecnologias avançadas, como Big Data, Machine Learning, Inteligência Artificial, Analytics, Cloud Computing, entre outros.

Comprometidos com a qualidade de entrega e inovação que agregue valor, consolidamos parcerias concretas com clientes no mercado nacional e internacional, atuando na transformação digital de empresas do Brasil, EUA, Japão, Portugal, Itália e Austrália.

## **São Paulo**

Av. das Nações Unidas, 12901 | 25 and - Cidade Monções  
+ 55 (11) 3504-1100

## **Campinas**

Av. John Dalton, 301 – Edifício 3, Cj. 13A - Techno Park Campinas  
+ 55 (19) 3242-8033

## **Matão**

Av. XV de Novembro, 1368 - Centro  
+55 (16) 3384-3053

## **Araraquara**

Av. Dr. Gastão Vidigal, 139 - Jardim Primavera  
+55 (16) 3461-3088

contato: [comercial@programmers.com.br](mailto:comercial@programmers.com.br)

