



**Cecília Junqueira Sartini**

Jundiaí  
2016

**Cecília Junqueira Sartini**

**BIG DATA**

Projeto apresentado ao Curso de  
Ciência da Computação da universidade  
Anhanguera Educacional.  
Orientador: Mariana M. Nunes.

**Jundiaí  
2016**

# SUMÁRIO

1 - INTRODUÇÃO.....	5
1.1 - O Problema.....	5
2 - OBJETIVOS.....	6
2.1 Objetivo Geral ou Primário.....	6
2.2 Objetivos Específicos ou Secundários.....	6
3 JUSTIFICATIVA.....	6
4 FUNDAMENTAÇÕES TEÓRICA.....	6
4.1 Big Data – O TRATAMENTO MapReduce + Hadoop.....	7
4.2 Big Data – volume, variedade e velocidade.....	7
4.3 Big Data – Importância e potencial.....	8
4.4 Big Data – INFLUÊNCIAS.....	8
4.4.1 Big Data – INFLUÊNCIA NA Saúde.....	9
4.4.2 Big Data – INFLUÊNCIA NA Educação.....	9
4.4.3 Big Data – INFLUÊNCIA em Bancos.....	9
4.4.4 Big Data – INFLUÊNCIA para o Governo.....	9
4.4.5 Big Data – INFLUÊNCIA na Produção.....	10
4.5 BIG DATA – DESAFIOS.....	10
4.5.1 BIG DATA – DESAFIO da Rapidez.....	10
4.5.2 BIG DATA – DESAFIO da Qualidade.....	10
4.5.3 BIG DATA – DESAFIO das Informações com sentido.....	10
4.5.4 BIG DATA – DESAFIO da Interpretação.....	11
4.5.5 BIG DATA – DESAFIO da Representação Gráfica.....	11
5 METODOLOGIAS.....	11
5.1 Bases de dados.....	11
5.2 Limites de Tempo.....	11
5.3 Idiomas.....	11
5.4 Principais termos buscados.....	11
5.5 Dos problemas e diferenças nos processos de indexação.....	11
6 CRONOGRAMA DE DESENVOLVIMENTO.....	12

7 REFERÊNCIAS.....	13
--------------------	----

# 1 INTRODUÇÃO

O conceito Big Data cresceu no fim do século XX e início do século XXI, é algo que se vem usando há muito tempo na área de tecnologia exemplos disso são as aplicações em Data Mining, Business Intelligence e CRM (Customer Relationship Management)

Big Data é um termo usado para extrair e guardar doses grandes de volume, grandes variedades de dados, permitindo alta velocidade de resposta, além da variabilidade.

Uma das funcionalidades que se utiliza o Big Data é para tomarem melhores medidas e rumos estratégicos de negócio.

A sua importância não segue apenas em sua quantidade de dados, se consegue verificar bases de seja qual for a fonte para localizar a resposta que se procura, ajudando assim a reduzir custo, reduzir tempo, tomar as melhores soluções e criação de novos projetos e boas ofertas. Além de ser usado a potência do Big Data com outros programas é capaz de realizar outras tarefas, trazendo benefícios e pontos positivos.

## 1.1 O Problema

O termo Big Data descreve grandes armazenadores de dados, por esta razão ocupam muito espaço, acarretando na dificuldade em gerir quanto verificar esses dados, fazendo com que tenham espaço para resposta imprecisa.

Com ações de tempo resposta rápidas para explorar algum dado ocorrem análises incompletas trazendo algo que não estaria pesquisando, além de que, um grande número de conhecimento dificultaria assim, para localizar algo importante.

## 2 OBJETIVOS

### 2.1 Objetivo Geral ou Primário

Manipular, investigar e refletir como é a estrutura e a operação através de Big Datas.

### 2.2 Objetivos Específicos ou Secundários

- Apresentar os grandes resultados em captar novas tendências através de uma ampla quantia de dados inúteis.
- Conceituar o Big Data: característica, processamento de dados, aplicabilidade.

## 3 JUSTIFICATIVA

“Big Data”, como o nome inspira, é uma grande quantia de dados, que é capaz de inventada ou não, e também, que podem ser obtidos de fontes controladas ou não. Sobretudo, quando se fala em Big Data, expressa-se em obter e processar uma enorme porção de dados a fim de obter dados importantes para sua relação, independentemente do bloco.

Ademais o Big Data é Open Source. E através do Hadoop (núcleo fundamental de uma plataforma para tratar do Big Data) podendo ser executado em um servidor de internet chamado Apache, que é responsável por concretizar e controlar tal ambiente. Por sua vez, o Apache é um mecanismo que trabalha em cima do Linux OS, que além de ser gratuito, também possui muitas empresas trabalhando para adequar as funcionalidades para chegar a um trabalho gradativamente mais profissional.

## 4 FUNDAMENTAÇÕES TEÓRICA

*“Esta nova era de Big Data tem um significado muito mais amplo, ainda, do que já foi mencionado. Estamos redefinindo o jeito com que a área de TI trabalha com os principais desafios da atualidade e antevê cenários a partir de sistemas computacionais, processos, conexões e bancos de dados. Sem dúvida, tal impacto será sentido nas empresas em curto prazo – desde a forma com que são estruturadas, como o operacional é realizado, até mesmo*

na gestão de investimentos e talentos”, (FILADORO, DIRETOR DE TECNOLOGIA DA ONLINE DATA CENTER, 2016)

*“O que irá realmente popularizar o Big Data será a sua capacidade de conectar não apenas cientistas de dados e tecnólogos, mas profissionais de todas as áreas. E, sem dúvida, uma das principais formas de fazer isso é usar visualizações que não se limitam a mostrar dados, números ou mesmo gráficos, mas que fazem todos esses elementos ganharem vida”.*

(CHRIS SELLAND, VICE-PRESIDENTE DE MARKETING E DESENVOLVIMENTO COMERCIAL, 2016).

#### 4.1 Big Data – O TRATAMENTO MapReduce + Hadoop

Tais aplicações são um modelo de programa usado para diferenciar companhias envolvidas com progresso mundial de software, como Google e Yahoo. O modelo viabiliza suporte para computação paralela. Para isso, implementa-os no Big Data.

Em uma das teses de graduação do Centro de Pesquisa Científica e Tecnológica da Escola Superior Politécnica del Litoral, Guayaquil, Equador, sugeriu-se que o MapReduce é uma programação que concede recursos aos dados distribuídos. Também, que é propício a falhas para processamento em grande escala.

Até por ter sido baseado nas linguagens de programação já existentes, é discutido e considerado, também, como concorrente de implementação.

#### 4.2 Big Data – volume, variedade e velocidade

Ainda que a expressão "Big Data" seja parcialmente atual, a prática de compilar e guardar grandes quantias de informes para uma análise ocasional de dados é ascendente. A concepção estimulou-se no começo dos anos 2000, quando um analisador renomado, Doug Laney, expressou a descrição de Big Data entre três principais: O Volume, que é o que grandes Instituições fazem, coletando dados de diversas fontes, como operações públicas, redes sociais e notações de sensores ou dados emitidos de máquina a máquina. A Velocidade, onde os dados trafegam na velocidade sem preâmbulo e deve-se ser examinado em tempo ágil. E a Variedade, que os dados são estabelecidos, e em todos os gêneros. De dados organizados, dados numéricos em bancos de

dados usuais, até títulos de texto não desenvolvidos, e-mail, vídeo, áudio, dados de cotações da bolsa e processos financeiros.

No entanto, mesmo tendo criado três, já existem mais duas perspectivas que estão sendo consideradas: A Variabilidade, que além da velocidade e variedade de dados que gradualmente aumenta, as movimentações de dados podem ser excessivamente volúveis com picos constantes. E a Complexidade, com que os dados de hoje surgem. Como vem de várias fontes, torna-se difícil satisfazer uma relação, adaptar-se, limpar e mudar dados entre sistemas distintos.

#### 4.3 BIG DATA – Importância e potencial

Sua importância não é somente à quantidade de dados que se tem, mas em como que você usa tais dados. Pode-se localizar dados de indefinidas fontes para ter *feedbacks* que possibilitam a diminuição de custos, ou de tempo, o avanço de produtos e propostas potencializadas, e deliberações mais inteligentes.

Ao combinar Big Data com uma alta potência, suporta-se tarefas referentes a negócios, como: Entender razão de falhas, questões e erros em tempo quase real; elaborar vales no local de venda com base no registro de gastos dos clientes; recalcular carteiras de risco inteiras com rapidez, e também, detecção comportamentais fraudulentas antes que eles afetem a organização.

A parcela de dados que está sendo criada e guardada em um nível mundial é extraordinária, e segue aumentando. Isso revela que há ainda mais capacidade para transcrever concentrações de negócios provenientes de tal conhecimento.

#### 4.4 Big Data – INFLUÊNCIAS

Em teoria, o termo "Big Data" inclui conjuntos de dados que são grandes a ponto de serem não controláveis, ou ilegíveis por humanos sem auxílio de uma ou mais máquinas.

Entretanto, com a assistência delas, pode influenciar e expandir muito, a rotina de diversas áreas.



#### 4.4.1 Big Data – INFLUÊNCIA na Saúde

Na Saúde, o Big Data não é somente uma influência, mas também surge para acrescer evolução, visto que com a digitalização dos dados, a divisão da Saúde é favorecida com a facilidade para o progresso para traçar novas técnicas, posto que a informação se expandirá. É utilizado para as informações de pacientes, e nos planos de convênio, e também, nas receitas e laudos. Quando se aborda o assunto de responsabilidade com a Saúde, é necessário que tudo seja feito com agilidade, exatidão e, em alguns casos, com razoável nitidez para atender as padronizações rigorosas deste ramo.

#### 4.4.2 Big Data – INFLUÊNCIA na Educação

Na Educação, profissionais possuindo uma visão voltada a dados são capazes de ter um abalo notável sobre os sistemas acadêmicos, alunos e currículos. Conhecendo e sabendo usar o Big Data, eles podem reconhecer estudantes em risco, certificar que os alunos estão melhorando de forma apropriada, e podem efetuar um retorno avaliativo de apoio melhor aos docentes.

#### 4.4.3 Big Data – INFLUÊNCIA em Bancos

Os serviços financeiros como bancos, com grandes porções de referências transcorrendo a partir de inúmeras fontes, tem como desafios descobrir maneiras novas e empreendedoras de gerenciamento para Big Data. Ao passo que Big Data é considerável para entender os clientes e aumentar seu bem-estar, é semelhantemente importante para diminuir os riscos e falhas enquanto conserva um acordo regulador. Entretanto, para o caso de bancos, também demanda que as firmas financeiras estejam um passo à frente neste jogo, com pesquisas avançadas.

#### 4.4.4 Big Data – INFLUÊNCIA para o Governo

Para o Governo, quando as instituições governamentais são aptas de aproveitar e aplicar Big Data, elas progridem consideravelmente ao que se refere a administrar serviços públicos, lidar com a paralisação ou a prevenir a

criminalidade. Porém, embora existam ganhos com o uso de Big Data, os comandantes também devem tratar das questões de nitidez e privacidade de conhecimento.

#### 4.4.5 Big Data – INFLUÊNCIA na Produção

Na manufatura, pensando em que Big Data pode proporcionar, os trabalhadores podem progredir em qualidade e produção, diminuindo o desperdício. Por fim, o varejo, que com a criação de contato com o cliente é essencial para o setor, e o melhor modo de gerenciar este trato é Big Data. Para os varejistas, é indispensável descobrir a melhor forma de vender aos clientes, o modo mais competente de lidar com transações, e a melhor estratégia de aumento do número de negócios repetidos.

#### 4.5 BIG DATA – DESAFIOS

Muitos impedimentos diários envolvem a pesquisa, dos quais profissionais de TI têm de lidar a todo tempo para fazer da análise de Big Data, uma ferramenta útil para os negócios.

Igualmente, para se obter benefícios da análise de Big Data, tem-se o desafio de salientar o que releva-se em meio a tantos dados.

##### 4.5.1 BIG DATA – DESAFIO da Rapidez

Análise e esclarecimento de dados não é mais um diferencial. O tempo em que isso é feito, sim. Como espera-se respostas rápidas, dos dados, as ações criadas à partir destas devem ser igualmente ágeis.

Todavia, o desafio realmente se multiplica quando um conhecimento mais preciso é requerido.

##### 4.5.2 BIG DATA – DESAFIO da Qualidade

Cumprido o primeiro desafio, do tempo, percebe-se que há o da confiabilidade. Afinal, de nada adianta astúcia, se não houver credibilidade.

##### 4.5.3 BIG DATA – DESAFIO das Informações com sentido

E então, surge o terceiro desafio: Ao verificar tamanha parcela de dados, categorias e fontes, o resultado deve surgir de forma compreensível.

#### 4.5.4 BIG DATA – DESAFIO da Interpretação

Há também o desafio da contextualização, pois sem ela, dados não fazem sentido algum, e ficam parecendo irrelevantes.

#### 4.5.5 BIG DATA – DESAFIO da Representação Gráfica

E então, tem-se o desafio que dados não estruturados não podem ser representados graficamente.

## 5 METODOLOGIAS

O seguinte trabalho se classifica como uma pesquisa acadêmico e está dividido em 4 passos a saber:

### 5.1 Bases de Dados

As buscas foram realizadas nas bases de dados bibliográficas do site SCIELO.

### 5.2 Limites de Tempo

Foram selecionados artigos publicados no intervalo de um ano.

### 5.3 Idiomas

Foram selecionados artigos escritos em inglês, português ou espanhol.

### 5.4 Principais termos buscados

O principal tema buscado foi “Big Data”

### 5.5 Dos problemas e diferenças nos processos de indexação

Como não houve problema na busca das bases de dados bibliográficas optou-se pelo uso exclusivo de sites.

Com essa estratégia, houve uma recuperação de um número maior de referências, garantindo a detecção da maioria dos trabalhos publicados dentro dos critérios pré-estabelecidos.

# 6 CRONOGRAMA DE DESENVOLVIMENTO

**Quadro 1** – Cronograma de execução das atividades do Projeto e do Trabalho de Conclusão de Curso.

ATIVIDADES	2016					2017						
	AG O	SE T	OU T	NO V	DE Z	JA N	FE V	MA R	AB R	MA I	JU N	JU L
Escolha do tema. Definição do problema de pesquisa		x	X									
Definição dos objetivos, justificativa.		x	X									
Definição da metodologia.			X	x								
Pesquisa bibliográfica e elaboração da fundamentação teórica.	x	x	X	x								
Entrega da primeira versão do projeto.				x	X							
Entrega da versão final do projeto.				x	x							
Revisão das referências para elaboração do TCC.							x	x	X			
Elaboração do Capítulo 1.							X					
Revisão e reestruturação do Capítulo 1 e								x				

elaboração do Capítulo 2.												
Revisão e reestruturação dos Capítulos 1 e 2. Elaboração do Capítulo 3.								x	x			
Elaboração das considerações finais. Revisão da Introdução.									x	X		
Reestruturação e revisão de todo o texto. Verificação das referências utilizadas.										X		
Elaboração de todos os elementos pré e pós-textuais.										X		
Entrega da monografia.										X	x	
Defesa da monografia.											X	

## 7 REFERÊNCIAS

APACHE HADOOP. **Welcome to apache™ hadoop®!**. Disponível em: <<http://hadoop.apache.org/>>. Acesso em: 24 nov. 2016.

BIG DATA BUSINESS. **Big Data analytics: você sabe o que é?**. Disponível em: <<http://www.bigdatabusiness.com.br/voce-sabe-o-que-e-big-data-analytics/>>. Acesso em: 24 nov. 2016.

BROADTEC. **Desafios diários do Big Data**. Disponível em: <<http://www.broadtec.com.br/desafios-diarios-do-big-data/>>. Acesso em: 24

nov. 2016.

COMPUTERWORLD. **Cinco problemas que vem com o Big Data.** Disponível em: <<http://computerworld.com.br/cinco-problemas-que-vem-com-o-big-data>>. Acesso em: 24 nov. 2016.

DEVELOPERWORKS. **desafios de segurança e privacidade em Big Data.** Disponível em: <<https://www.ibm.com/developerworks/community/blogs/tlcbr/entry/mp256?lang=en>>. Acesso em: 24 nov. 2016.

ENDEAVOR. **Big Data: descubra o que é e como usar na sua empresa.** Disponível em: <<https://endeavor.org.br/big-data-descubra-o-que-e-e-como-usar-na-sua-empresa/>>. Acesso em: 24 nov. 2016.

INFO WESTER. **O que é Big Data?** Disponível em: <<http://www.infowester.com/big-data.php>>. Acesso em: 24 nov. 2016.

PROFISSIONAIS TI. **Como enfrentar os desafios do Big Data?** Disponível em: <<https://www.profissionaisiti.com.br/2013/06/como-enfrentar-os-desafios-do-big-data/>>. Acesso em: 24 nov. 2016.

SAS. **Big Data: o que é e por que é importante.** Disponível em: <[http://www.sas.com/pt\\_br/insights/big-data/what-is-big-data.html](http://www.sas.com/pt_br/insights/big-data/what-is-big-data.html)>. Acesso em: 24 nov. 2016.

SCIELO. **Acerca de la aplicación de mapreduce + hadoop en el tratamiento de Big Data.** Disponível em: <[http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=s2227-18992015000300004&lang=pt](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=s2227-18992015000300004&lang=pt)>. Acesso em: 24 nov. 2016.

SCIELO. **Big Data em Saúde: estamos perto?** Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=s0103-507x2016000100008&lang=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=s0103-507x2016000100008&lang=pt)>. Acesso em: 24 nov. 2016.

VERT. **Por que o Big Data é tão importante para as empresas?** Disponível em: <<http://www.vert.com.br/blog-vert/por-que-o-big-data-e-tao-importante-para-as-empresas/>>. Acesso em: 24 nov. 2016.