



ADRIANA AMORIN MEDEIROS

BRUNO HENRIQUE FLORINDO DA CUNHA

CECÍLIA JUNQUEIRA SARTINI

TAWANY TERESA OLIVEIRA BATISTA



BIG DATA

ADRIANA AMORIN MEDEIROS

BRUNO HENRIQUE FLORINDO DA CUNHA

CECÍLIA JUNQUEIRA SARTINI

TAWANY TERESA OLIVEIRA BATISTA

BIG DATA

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

Jundiaí

2016

SUMÁRIO

1 INTRODUÇÃO	3
1.1 O Problema	3
2 OBJETIVOS	4
2.1 Objetivo Geral ou Primário	4
2.2 Objetivos Específicos ou Secundários	4
3 JUSTIFICATIVA	5
4 FUNDAMENTAÇÃO TEÓRICA.....	6
5 METODOLOGIA.....	9
6 CRONOGRAMA DE DESENVOLVIMENTO	10
REFERÊNCIAS.....	11



1 INTRODUÇÃO

O conceito Big Data ganhou força no ano 2000, 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 armazenar grandes quantidades 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 decisões e direções estratégicas de negocio.

A sua importância não segue apenas em sua quantidade de dados, se consegue analisar dados de qualquer fonte para encontrar a resposta que se procura, ajudando assim a reduzir gasto, reduzir tempo, tomar as melhores decisões e desenvolvimento de novos projetos e boas ofertas. Além de ser usado a potencia do Big Data com outros programas pode 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 analisar esses dados, fazendo com que tenham espaço para resposta imprecisa.

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

2 OBJETIVOS

2.1 Objetivo Geral ou Primário

Manipular, analisar 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 grande quantidade de dados inúteis.
- Conceituar o big data: característica, processamento de dados, aplicabilidade.

3 JUSTIFICATIVA

“Big Data”, como o nome inspira, é uma grande quantidade de dados, que pode ser 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 quantidade de dados a fim de obter informações importantes para seu negócio, independentemente do bloco.

Ademais o Big Data é Open Source. E através do Hadoop (núcleo fundamental de uma plataforma para lidar com o Big Data) pode 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 cada vez mais profissional.



4 FUNDAMENTAÇÃO TEÓRICA

“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 – VOLUME, VARIEDADE E VELOCIDADE



Ainda que a expressão "big data" seja relativamente atual, a prática de compilar e armazenar grandes quantidades de informações para uma análise ocasional de dados é ascendente. A concepção estimulou-se no início dos anos 2000, quando um analista famoso, Doug Laney, articulou a definição de big data entre três principais: O Volume, que é o que grandes Instituições fazem, coletando dados de diversas fontes, como transações comerciais, redes sociais e informações de sensores ou dados transmitidos de máquina a máquina. A Velocidade, onde os dados trafegam na velocidade sem preâmbulo e devem ser tratados em tempo ágil. E a Variedade, que os dados são estabelecidos, e em todos os tipos de formatos. De dados estruturados, dados numéricos em bancos de dados tradicionais, até documentos de texto não estruturados, e-mail, vídeo, áudio, dados de cotações da bolsa e transações financeiras.

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 cada vez maiores, os fluxos de dados podem ser altamente volúveis com picos periódicos. 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 transformar dados entre diferentes sistemas.

4.2 BIG DATA – IMPORTÂNCIA E POTENCIAL

Sua importância não gira em torno da quantidade de dados que se tem, mas no que você faz com tais dados. Pode-se analisar dados de indefinidas fontes para encontrar respostas que permitam a redução de custos, ou de tempo, o desenvolvimento de novos produtos e ofertas otimizadas, e decisões mais inteligentes.

Ao combinar big data com uma alta potência, suporta-se tarefas relacionadas a negócios, como: Determinar a causa raiz de falhas, questões e defeitos em tempo

quase real; elaborar cupons no ponto de venda com base em hábitos de compra dos clientes; recalcular carteiras de risco inteiras com rapidez, e também, detecção comportamentais fraudulentos antes que eles afetem a organização.

A parcela de dados que está sendo criada e armazenada em um nível global é extraordinária, e continua aumentando. Isso revela que há ainda mais potencial para extrair concentrações de negócios provenientes de tais informações.

4.3 BIG DATA – INFLUÊNCIA

Big Data afeta organizações em quase todas as indústrias. Na Saúde, é utilizado para os registros de pacientes, e nos planos de tratamento, e também, nas informações de prescrição. Quando se trata de responsabilidade com a saúde, tudo precisa ser feito com agilidade, exatidão e, em alguns casos, com suficiente transparência para satisfazer as regulamentações rigorosas desta indústria. Na Educação, profissionais possuindo uma visão orientada a dados podem ter um impacto significativo sobre os sistemas escolares, estudantes e currículos.

Conhecendo e sabendo lidar com big data, eles podem identificar alunos em risco, assegurar que os estudantes estão progredindo de forma adequada, e podem implementar um sistema melhor de avaliação e apoio aos professores e diretores.

Os serviços financeiros como bancos, com grandes quantidades de informações fluindo partir inúmeras fontes, tem como desafios encontrar maneiras novas e inovadoras de gerenciamento para big data. Ao passo que big data é importante para entender os clientes e aumentar seu bem-estar, é semelhantemente importante para minimizar os riscos e fraudes enquanto mantém uma conformidade regulatória.

Entretanto, para o caso de bancos, também exige que as instituições financeiras estejam um passo à frente neste jogo, com análises avançadas. Para o Governo, quando as organizações governamentais são capazes de aproveitar e aplicar big data, elas progridem consideravelmente ao que se refere a gerenciar serviços públicos, lidar com o congestionamento ou a prevenir a criminalidade. Porém, embora existam vantagens com o uso de big data, os governos também devem tratar das questões de transparência e privacidade das informações. Na manufatura, pensando na visão que big data pode fornecer, os fabricantes podem progredir em qualidade e produção, diminuindo o desperdício. Por fim, o varejo, que com a construção de contato com o cliente é essencial para o setor, e a melhor maneira de

gerenciar este relacionamento é big data. Para os varejistas, é indispensável descobrir a melhor maneira de vender aos clientes, o modo mais eficaz de lidar com transações, e a forma mais estratégica de aumentar o número de negócios repetidos.



5 METODOLOGIA

Esta monografia englobará os conceitos relativos ao Problema. A definição de cada um destes tipos de sistemas está descrita na seção “Cronograma de desenvolvimento”. Para que o trabalho seja realizado, foram buscados artigos nas bases de dados de diversos sites, em um intervalo de um ano.

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

5.1 Bases de dados

As buscas foram realizadas em cinco bases de dados bibliográficas — PubMed, Web of Science, EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINAHL) e LILACS.

Ao finalizar as pesquisas em cada base, as referências duplicadas foram excluídas.

5.2 Limite de tempo

Foram selecionados artigos publicados entre 2005 e 2009 (incluindo aqueles disponíveis online em 2009 que poderiam ser publicados em 2010).

5.3 Idiomas

Foram selecionados artigos escritos em inglês, português ou espanhol.

5.4 Principais termos buscados

Há problemas e diferenças nos processos de indexação nas bases de dados bibliográficas; portanto, optou-se pela busca por termos livres, sem o uso de vocabulário controlado (descritores). 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. Um profissional graduado em biblioteconomia foi responsável por essa etapa do trabalho. Os termos >>>>>>>>>> foram combinados com as associações e desfechos de interesse.



6 CRONOGRAMA DE DESENVOLVIMENTO

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

[illegible]



REFERÊNCIAS

BIG DATA BUSINESS DISPONÍVEL EM <[HTTP://WWW.BIGDATABUSINESS.COM.BR/VOCE-SABE-O-QUE-E-BIG-DATA-ANALYTICS/](http://www.bigdatabusiness.com.br/voce-sabe-o-que-e-big-data-analytics/)> ACESSO EM 30 DE SETEMBRO DE 2016.

APACHE HADOOP DISPONÍVEL EM <[HTTP://HADOOP.APACHE.ORG/](http://hadoop.apache.org/)> ACESSO EM 30 DE SETEMBRO DE 2016.

TABLEAU DISPONÍVEL EM: < [HTTP://WWW.TABLEAU.COM/PT-BR/ASSET/TOP-8-TRENDS-BIG-DATA-2016?UTM_CAMPAIGN=PROSPECTING-BGDATA-ALL-ALL&UTM_MEDIUM=PAID+SEARCH&UTM_SOURCE=GOOGLE+SEARCH&UTM_LANGUAGE=PT&UTM_COUNTRY=BRA&KW=BIG%20DATA&ADGROUP=CTX-BIG+DATA-BIG+DATA+ALL-PT-P&ADUSED=120535072681&MATCHTYPE=P&PLACEMENT=&KCID=0AF8C491-F69A-42D6-A2BF-B9AA3AA02E62&GCLID=CMX37ZLU2S8CFQCIKQODOCCIVG](http://www.tableau.com/pt-br/asset/top-8-trends-big-data-2016?utm_campaign=prospecting-bgdata-all-all&utm_medium=paid+search&utm_source=google+search&utm_language=pt&utm_country=bra&kw=big%20data&adgroup=ctx-big+data-big+data+all-pt-p&adused=120535072681&matchtype=p&placement=&kcid=0af8c491-f69a-42d6-a2bf-b9aa3aa02e62&gclid=cmx37zlu2s8cfqciqkqodoccivg)>. ACESSO EM 10 DE OUTUBRO DE 2016.

INFO WESTER DISPONÍVEL EM < [HTTP://WWW.INFOWESTER.COM/BIG-DATA.PHP](http://www.infowester.com/big-data.php)> ACESSO EM 10 DE OUTUBRO DE 2016

COMPUTERWORLD DISPONÍVEL EM <[HTTP://COMPUTERWORLD.COM.BR/CINCO-PROBLEMAS-QUE-VEM-COM-O-BIG-DATA](http://computerworld.com.br/cinco-problemas-que-vem-com-o-big-data)> ACESSO EM 11 DE OUTUBRO DE 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/](https://endeavor.org.br/big-data-descubra-o-que-e-e-como-usar-na-sua-empresa/)> ACESSO EM 5 DE NOVEMBRO DE 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/](http://www.vert.com.br/blog-vert/por-que-o-big-data-e-tao-importante-para-as-empresas/)> ACESSO EM 5 DE NOVEMBRO DE 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 17 DE NOVEMBRO DE 2016

[HTTPS://WWW.GOOGLE.COM.BR/SEARCH?Q=INSIGHTS&OQ=INSIGHTS&AQS=CHROME..69I57J0L5.1997J0J7&SOURCEID=CHROME&IE=UTF-8](https://www.google.com.br/search?q=insights&oq=insights&aqs=chrome..69l57j0l5.1997j0j7&sourceid=chrome&ie=utf-8)