**UNIVERSIDADE PAULISTA**

**FLÁVIO ROCHA VALENÇA**

**RAFAEL DE SOUZA BATISTA**

**WESLLEY DO ESPIRITO SANTO**

**ANALISE DE GRANDE VOLUME DE DADOS COM D3JS**

**SANTOS**

**2015**

**FLÁVIO ROCHA VALENÇA**

**RAFAEL DE SOUZA BATISTA**

**WESLLEY DO ESPIRITO SANTO**

**Analise de grande volume de dados com D3js**

Trabalho de conclusão de curso para obtenção do título de graduação em Ciências da Computação apresentado à Universidade Paulista – UNIP.

Orientador: Luiz Guilherme

**SANTOS**

**2015**

**FLÁVIO ROCHA VALENÇA**

**RAFAEL DE SOUZA BATISTA**

**WESLLEY DO ESPIRITO SANTO**

**ANALISE DE GRANDE VOLUME DE DADOS COM D3JS**

Trabalho de conclusão de curso para obtenção do título de graduação em Ciências da Computação apresentado à Universidade Paulista – UNIP.

Aprovado em:

BANCA EXAMINADORA

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_/\_\_/\_\_\_

Nome do professor

Universidade Paulista – UNIP

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_/\_\_/\_\_\_

Nome do professor

Universidade Paulista – UNIP

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_/\_\_/\_\_\_

Nome do professor

Universidade Paulista – UNIP

**DEDICATÓRIA**

**AGRADECIMENTOS**

**RESUMO**

**ABSTRACT**

**LISTA DE ILUSTRAÇÕES**

**LISTA DE ABREVIATURAS E SIGLAS**

**SUMÁRIO**

# INTRODUÇÃO

O hábito de guardar informações sejam eles, documento, fotos, livros, estudos científicos e assim por diante, vem crescendo com a interação entre homem-máquina. E com isso, gerou reflexos na área de tecnologia, logo, armazenar as informações e dados se tornou algo do cotidiano, tendo em vista a facilidade do usuário em operar o sistema, que de forma mecânica deixa de perceber a quantidade de dados e informações armazenadas**.**

“No artigo publicado pelo “CCC (computing community consortium)” (Big-Data Computing: Criando revolucionário avanços no comércio, ciência e sociedade), “ a rede Wal-Mart recentemente contratou Hewlett Packard para construir um armazém capaz de armazenar 4 petabytes (4 trilhão de bytes) de dados, registrando cada compra registrada por seus terminais de ponto-de-venda (em torno de 267 milhões de transações por dia) em suas 6.000 lojas em todo o mundo. “Através da aplicação de aprendizagem de máquina a esses dados, eles podem detectar padrões que indicam a eficácia de suas estratégias de preços, publicidades, campanhas e gerenciar melhor suas cadeias de inventário e de abastecimento”.

**(artigo CCC (computing community consortium)” (Big-Data Computing: Criando revolucionário avanços no comércio, ciência e sociedade - Randal E. Bryant, Randy H. Katz, Edward D. Lazowska Version 8: December 22, 2008)**

Hoje, com o avanço e dependência da tecnologia, o homem manipula o que há de mais importante que permite que haja toda essa evolução, a informação.

Mas as histórias de como os dados se tornaram grandes começa muitos anos antes da história recente em torno do Big Data. A primeira tentativa de contabilizar essa grande massa de dados ocorreu em 1941, popularmente conhecida como a “Explosão da Informação” (termo usado pela primeira vez em 1941, de acordo com o Dicionário Oxford de Inglês). (PRESS, [s.d.])

De acordo com o artigo de Marcos Vieira, “O conceito de Big Data começou a ser discutido a cerca de 70 anos” devido ás enormes elevações de volume de dados tornando-se desafiante ter acesso e manipulação com o mesmo.(VIEIRA, [s.d.])

Ainda no artido de Marcos Vieira, embora já existisse o problema, “ o termo Big Data, no entanto, foi criado há 17 anos referindo-se à impossibilidade de armazenamento de grandes volumes de informações em data warehouses (armazém) únicos, já que o armazenamento digital ainda não era tão eficiente como hoje em dia.” (VIEIRA, [s.d.])

O “Big Data” forma um conjunto de dados definidos Velocidade, Volume e Variedade.

Velocidade: podemos defini-la como que seu principal objetivo seja conseguir com que toda a demanda gerada seja tratada em um espaço de tempo ideal pra cada uma delas.

Volume: trata-se do imenso volume de dados gerados todos os dias, seu desafio é o processo de armazenamento, transmissão ao mesmo tempo.

Variedade: são as variadas formas de informação processadas, e-mail, pagamentos, documentos, páginas web, vídeos e etc. O desafio aqui é colher todos esses dados, e para que haja uma boa performance de intepretação e analítica sobre eles.

**(**VIEIRA, M. **Entendendo Big DataNo Title**. Disponível em: <http://www.ecommercebrasil.com.br/artigos/entendendo-big-data>. Acesso em: 20 mar. 2015. **Já está referenciado)**

O artigo produzido pela revista “Ciência Hoje”, em um de seus trechos lembra também que junto com Big Data traz consigo também problemas éticos “Questões como a dificuldade de garantir a segurança e privacidade dos dados chegam a inviabilizar projetos, como uma base central de prontuários médicos, devido ao risco de essa informação ser utilizada de forma indevida”. (“Desafios do Big Data”, 2013)

Hoje em dia o grande problema não é a obtenção de dados e sim o seu tratamento, conseguir obter essa informação de maneira rápida e automatizada, já se parou pra pensar em como manipular e interagir com tantas informações geradas todos os dias?

Desenvolvido em 2011, tendo lançamento no em agosto de 2011, deu-se a origem ao D3js (Documents Data-Driven), uma biblioteca em JavaScript, uma linguagem de script para navegadores web, sendo a linguagem padrão para criação de páginas web com ótima interação com usuário, com o principal objetivo de se criar técnicas de visualização de dados interativos e dinâmicos, sendo aplicado em navegador web.

No site “Data Drive Journalism”, relata que o D3js vem “explodindo em popularidade, em parte porque além de ele ser extremamente flexível pelo fato de ele funcionar em navegadores modernos, incluindo Android e IOS dispositivos mobile, mas também porque é poderoso”.

Sua principal função é interligar os valores de uma massa de dados para documentar elementos, podendo assim manipular o documento a partir dessa obtenção de dados.

A flexibilidade é mesmo o maior trunfo dessa nova ferramenta, é ótimo até mesmo para jornalistas, pois não os limitam á uma forma de visual específica, sendo o próprio livre para fazer explorar o uso de gráficos através das informações, e a partir daí inventar novas formas visualização específica.

Ainda no site, de acordo com o “Data Drive Journalism”, “O The New York Times tem vindo gradualmente a eliminação gradual do Flash em favor de D3 e outras peças interativas baseadas em JavaScript, mesmo a contratação de Mike Bostock, o principal autor do D3, que agora está na equipe de gráficos do Times. O resultado foi uma série de gráficos interativos inovadores, incluindo os últimos [512 Caminhos para a Casa Branca](http://translate.googleusercontent.com/translate_c?depth=1&hl=pt-BR&prev=search&rurl=translate.google.com.br&sl=en&u=http://elections.nytimes.com/2012/results/president/scenarios&usg=ALkJrhig8dOygGN7-uHz6Z4msp6TPRxXWQ) , por Mike Bostock e Shan Carter”.

**(essa parte toda foi trazida daqui** [**http://datadrivenjournalism.net/resources/data\_driven\_documents\_defined - publicado 29-01-2013**](http://datadrivenjournalism.net/resources/data_driven_documents_defined%20-%20publicado%2029-01-2013) **- Scott Murray)**

# Objetivos

Nosso objetivo nesse trabalho, é apresentar essa ferramenta, apresentando as suas utilidades, pois, pode-se tirar grandes vantagens através de sua flexibilidade na manipulação de dados, podendo facilitar tanto para quem usa, para quem for analisar os gráficos gerados por ela, e a partir de então, podendo chegar a resultados satisfatórios rapidamente seja para uma empresa que queira evoluir-se, para o esporte gerando análises importantes, ou até mesmo para a ciência desenvolvendo e chegando a grandes descobertas.

Conforme o trabalho foi pesquisado, percebemos que no Brasil ainda não é muito popular, pode se aproveitar e expandir e apresentar para toda essa nova tecnologia.

O D3js parece ser o caminho para essa maneira rápida e eficaz de se manipular e interagir com tantas informações geradas num mundo que cada vez mais dependente da tecnologia, irá consequentemente irá gerar cada vez mais dados.

**REFERENCIAS**

BLEJMAN, M. **Como D3.js está mudando a forma de contar histórias com dados (e por que precisamos de uma hackatona)**. Disponível em: <http://ijnet.org/pt-br/blog/como-d3js-esta-mudando-forma-de-contar-historias-com-dados-e-por-que-precisamos-de-uma-hackaton>. Acesso em: 26 fev. 2015.

BOSTOCK, M. **Data-Driven Documents**. Disponível em: <http://d3js.org/>.

**D3.js – Gerando gráficos em Javascript**. Disponível em: <http://hanabrasil.com.br/2015/01/12/d3-js-gerando-graficos-em-javascript/>. Acesso em: 27 fev. 2014.

DE SOUSA, F. **Conheça a biblioteca D3js**. Disponível em: <http://revistaw.com.br/blog/conheca-a-biblioteca-d3-js/>.

Desafios do Big Data. p. 6, 2013.

EMERSON ALECRIM. **O que é Big Data?** Disponível em: <http://www.infowester.com/big-data.php>. Acesso em: 28 fev. 2015.

MURRAY, S. **Data-Driven Documents, Defined**. Disponível em: <http://datadrivenjournalism.net/resources/data\_driven\_documents\_defined>. Acesso em: 20 mar. 2015.

OTÁVIO GONÇALVES DE SANTANA. **Introdução ao Conceito de Big Data**. Disponível em: <http://www.devmedia.com.br/introducao-ao-conceito-de-big-data/27066>. Acesso em: 28 fev. 2014.

PRESS, G. **A Very Short History Of Big Data**. Disponível em: <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/>.

VIEIRA, M. **Entendendo Big DataNo Title**. Disponível em: <http://www.ecommercebrasil.com.br/artigos/entendendo-big-data>. Acesso em: 20 mar. 2015.