

# CONIC SEMESP

16º Congresso Nacional de Iniciação Científica

**TÍTULO:** UM ESTUDO SOBRE A INFRAESTRUTURA PARA BIG DATA

**CATEGORIA:** EM ANDAMENTO

**ÁREA:** CIÊNCIAS EXATAS E DA TERRA

**SUBÁREA:** COMPUTAÇÃO E INFORMÁTICA

**INSTITUIÇÃO:** CENTRO UNIVERSITÁRIO DAS FACULDADES METROPOLITANAS UNIDAS

**AUTOR(ES):** MÔNICA BEATRIZ GOBBO CASTILHO

**ORIENTADOR(ES):** MILKES YONE ALVARENGA

Realização:

**SEMESP**  
sindicato das mantenedoras de ensino superior



Apoio:

 **ENIAC**  
Educação Básica e Superior

## **1.RESUMO**

Nos últimos anos, os avanços significativos que ocorreram nas Tecnologias da Informação propiciaram novas formas de interação e comunicação entre empresas, organizações e pessoas. Isto tem resultado em um volume exponencial de dados e informações a serem processadas e armazenadas. Neste cenário, o Big Data se apresenta como um modelo eficiente para tirar proveito destes dados e informações de modo a gerar análises cujos resultados podem ser extremamente importantes na tomada de decisões e estabelecimento de estratégias para empresas e organizações. No presente trabalho deseja-se investigar a infraestrutura necessária para adoção do Big Data e sua relação com Computação em Nuvem. Pretende-se estudar aspectos desta infraestrutura e avaliar como a mesma pode contribuir na composição de soluções eficientes com uso de Big Data.

## **2.INTRODUÇÃO**

As Tecnologias da Informação que vem surgindo nos últimos anos influíram diretamente na comunicação das pessoas, empresas, instituições e no mundo dos negócios e permitiram novas formas de pessoas e organizações interagirem entre si, o que resultou em um aumento exponencial do volume de dados e informações que circulam na internet e nas redes de computadores [2][5].

Atualmente, milhões de mensagens são enviadas diariamente por e-mail, Smartphones, Tablets, Notebooks, entre outros. Rede Sociais, que recebem uma quantidade inimaginável de postagens a todo momento, também ganharam um papel importante nas empresas e passaram a afetar direta e indiretamente seus processos e formas de fazer negócios [5][6].

Além das informações geradas diretamente pela “atuação humana”, há também aquelas produzidas a todo momento por sensores, máquinas, dispositivos automáticos, radares, entre outros, que percorrem trafegam em redes computacionais o tempo todo gerando informações e novos dados que são armazenados e compartilhados de forma autônoma [6].

É neste cenário de multiplicação de informações e de conseqüente gigantesco armazenamento de dados que surgem as figuras do Big Data e da Mineração de Dados. O Big Data congrega um conjunto de tecnologias que permite que as organizações analisem dados, cujo acesso antes não possuíam para auxiliar na tomada decisões ou gerenciar atividades de maneira muito mais eficiente [1][4].

### **3.OBJETIVOS**

No presente trabalho deseja-se investigar a infraestrutura necessária para adoção do Big Data e sua relação com Computação em Nuvem. Pretende-se estudar aspectos desta infraestrutura e avaliar como a mesma pode contribuir na composição de soluções eficientes com uso de Big Data.

### **4.METODOLOGIA**

Este estudo está sendo realizado por meio de Pesquisa Bibliográfica na literatura relacionada ao tema do mesmo incluindo livros, artigos, periódicos e publicações técnicas para fundamentação teórica deste trabalho. Adicionalmente têm sido feitas consultas a conteúdos específicos disponíveis na Internet para se ampliar o conhecimento de tecnologias e sobre soluções já implementadas na adoção de Big Data.

### **5.DESENVOLVIMENTO**

O Big Data possui pelo menos quatro características [4]:

- *volume*: capacidade de lidar com a imensa quantidade de dados que são gerados a todo o momento.

- *velocidade*: diz respeito à velocidade em que os dados são armazenados e processados e, em alguns casos, esta tarefa é feita em tempo real.

- *variedade*: como consequência da diversidade de informações geradas hoje, dados em formato estruturado e não estruturado oriundos de inúmeras fontes devem ser tratados como parte de um todo - um tipo de dado pode ser inútil se não for associado a outros. Trata-se da necessidade de integrar e aproveitar cada informação de modo que ela se relacione e faça sentido junto as demais coletadas.

- *veracidade*: os dados devem ser confiáveis. É necessário que haja mecânicos assegurem o máximo possível à coerência dos dados.

Muitas empresas e organizações ainda possuem dificuldades com alguns aspectos básicos relacionados à administração de dados e sua efetiva exploração: elas ainda não sabem como extrair valor das informações que reúnem [6]. A utilização de ferramentas do Big Data abre um novo leque de oportunidades. Sendo realizada de forma correta, a análise desta gigantesca massa de dados pode revelar padrões de comportamento, correlações entre dados até então desconhecidos, significados antes ocultos e ajudar a prever tendências de consumo. Isto representa grandes vantagens competitivas para empresas e organizações [2][3].

A infraestrutura necessária para adoção do Big Data envolve a utilização de Computação de Alto desempenho com uso de Clusters trabalhando com Computação em Nuvem [2]. O conceito de computação em nuvem (do inglês, *cloud computing*) refere-se à ao processamento e armazenamento de dados e informações realizados em computadores remotos interligados via internet, seguindo o princípio da computação em grade [1].

O armazenamento de dados é feito em serviços que poderão ser acessados de qualquer lugar do mundo, a qualquer hora, não havendo necessidade de instalação de programas ou de armazenar dados. O acesso a programas, serviços e arquivos é remoto, através da Internet - daí a alusão à nuvem. O uso desse modelo (ambiente) é mais viável do que o uso de unidades físicas [1] [3].

## 6. RESULTADOS PRELIMINARES

Os resultados parciais obtidos até o momento indicam a importância da utilização do Modelo de Computação em Nuvem e sua infraestrutura para implementar soluções com Big Data não somente pela necessidade de se armazenar grandes volumes de informações mas também pela alta demanda de processamento requerida. O Modelo de Computação em Nuvem pode empregar diferentes configurações e padrões de hardware na sua implementação. No entanto, grande parte dos provedores de Computação em Nuvem e Data Centers que operam com este modelo para soluções com Big Data necessitam de hardwares suficientemente robustos para suportarem não somente aplicações que envolvam serviços que requerem alto desempenho computacional mas também que sejam capazes de suportar alta demanda de acesso de inúmeros clientes da nuvem.

## 7. FONTES CONSULTADAS

- [1] TAURION, C. **Big Data** Edição digital, Editora Brasport, Rio de Janeiro, 2013.
- [2] MCAFEE, A; BRYNJOLFSSON, E. **Big Data: The Management Revolution**. Harvard Business Review, edição de outubro de 2012.
- [3] PEW RESEARCH INTERNET PROJECT. **The Future of Big Data**. Disponível em: <http://www.pewinternet.org/2012/07/20/the-future-of-big-data/> (Acesso em 05 de março 2016).
- [4] INFO WESTER. **O que é Big Data?** Disponível em: <http://www.infowester.com/big-data.php> (Acesso em 05 de abril 2016).
- [5] O GLOBO. **Estudo da EMC prevê que volume de dados virtuais armazenados será seis vezes maior em 2020**. Disponível em: <http://oglobo.globo.com/sociedade/tecnologia/estudo-da-emc-preve-que-volume-de-dados-virtuais-armazenados-sera-seis-vezes-maior-em-2020-12147682#ixzz39V1Sn2Yw> (Acesso em 26 de março 2016).
- [6] EIU – *Economist Intelligence Unit*. **Big data - Lessons from the leaders**. Disponível em: [http://www.economistinsights.com/sites/default/files/downloads/EIU\\_SAS\\_BigData\\_4.pdf](http://www.economistinsights.com/sites/default/files/downloads/EIU_SAS_BigData_4.pdf) (Acesso em 26 de março 2016).