

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
MBA EM GESTÃO DE NEGÓCIOS E TECNOLOGIA DA INFORMAÇÃO

Ronald Bertele

MAPEAMENTO DA PRODUÇÃO INTERNACIONAL SOBRE *BIG DATA*
ENTRE 2011 E 2013

Porto Alegre
2014

MAPEAMENTO DA PRODUÇÃO INTERNACIONAL SOBRE *BIG DATA* ENTRE 2011 E 2013

Ronald Bertele¹

José Jerônimo de Menezes Lima²

RESUMO

Este artigo tem por objetivo mapear a produção internacional sobre o tema *Big Data* entre janeiro de 2011 e dezembro de 2013. Para este mapeamento, a pesquisa coletou 298 publicações disponibilizadas pela *Elton Bryson Stephens Company* (EBSCOHost) nas bases de dados acadêmicas sobre o tema. Os artigos foram submetidos a uma seleção, sendo eliminadas as publicações em duplicidade ou que não tinham relação direta com o objetivo do estudo. A pesquisa bibliográfica analisou uma amostra final de 51 artigos. Observa-se nas publicações dos últimos três anos que, apesar desta tecnologia ser recente, o uso de *big data* vem crescendo em diversos segmentos de mercado, tornando-se um grande diferencial competitivo e cada vez mais essencial para o crescimento das empresas. A tecnologia ganhou força no mercado a partir do ano de 2012.

Palavras-chave: *Big Data*. Análise de Dados. *Data Mining*.

¹ Bacharel em Ciência da Computação pela UCS - Universidade de Caxias do Sul. E-mail: ronald.bertele@gmail.com.

² Doutor em Administração pela UNISINOS - Universidade do Vale do Rio dos Sinos. Email: jeronimol@unisinobr.

1 INTRODUÇÃO

Nos últimos anos o número de usuários com acesso à internet cresceu rapidamente. Este fato é explicado pela popularização dos dispositivos móveis como *smartphones* e *tablets* e pela praticidade de acesso aos serviços digitais. As informações geradas por estes usuários também estão crescendo pela troca de *e-mails*, participação em redes sociais, acesso a páginas de pesquisa, vídeos e muitas outras formas de compartilhar informações em meio digital. A análise desta massa de dados tem ganhado destaque no mundo em diversas áreas, pois o estudo deste “rastros digital” pode revelar tendências e comportamentos que até então passavam despercebidos ou simplesmente não eram avaliados pelas empresas (SCHONBERGUER, CUKIER. 2013).

A correlação destas informações pode revelar explicações sobre determinados eventos, como ajudar as organizações a melhorar determinadas características de seus produtos a fim de torná-los mais competitivos no mercado, ofertarem determinados tipos de serviços que gerem uma demanda ainda não explorada corretamente. Este tipo de análise também pode explicar o comportamento de uma determinada população perante um certo problema e ditar novas tendências de mercado encontrando um novo “oceano azul” (KIM; MAUBORGNE, 2005). Este grande volume de informações é definido como *Big Data*.

Mapeando a produção internacional entre janeiro de 2011 e dezembro de 2013 notou-se que o termo *big data* surgiu com o crescimento das informações digitais e da viabilidade de analisar estas informações por meio computacional. O avanço tecnológico dos equipamentos possibilitando maior capacidade de processamento e gerando respostas em pequenos intervalos de tempo está tornando o *big data* um grande diferencial de competitividade. As respostas precisam ser tão rápidas quanto a geração de novas informações.

A mineração de dados tradicional utiliza métodos estatísticos dos quais é selecionada uma amostra inicial e, aplicando-se determinados cálculos matemáticos, chega-se a resultados esperados com um índice de probabilidade (COURTNEY, 2012). Este índice de pode variar de acordo com o tamanho da amostra das informações. Com a tecnologia de *big data* foi criado um novo conceito de análise de dados. Segundo Fuchs (2013), neste tipo de análise são utilizados métodos estatísticos mais complexos, pois não se usa uma amostra e sim toda a massa de

dados. Com o conjunto completo de informações é possível identificar determinados comportamentos e situações que não podem não ser observados em uma pequena amostra de dados ou que passariam despercebidos usando as técnicas tradicionais de mineração de dados.

Para Gleick (2013), o excesso de informações sempre esteve presente na evolução da sociedade. Foi com a compreensão destas informações que a sociedade conseguiu entender o seu real significado e evoluir. Atualmente, o número de informações geradas cresce exponencialmente e um dos pontos mais importantes não é o tamanho destas informações, mas sim como compreender qual o significado desta avalanche e como ele pode contribuir para o contínuo crescimento da sociedade.

Dado o exposto, é necessário definir o que é *Big Data*.

De acordo com SAS (2013), *big data* é um termo popular usado para descrever o crescimento, a disponibilidade e o uso exponencial de informações estruturadas e não estruturadas. Logo, é o conceito que reúne um conjunto de informações que cresce rapidamente e que são armazenadas para serem analisados na expectativa de encontrar respostas ou comportamentos não mapeados. Conforme a Wikipédia (2013), o termo é usado para uma coleção de conjuntos de dados tão grande e complexo que se torna difícil para processar usando ferramentas de gerenciamento de banco de dados e aplicações de processamento de dados tradicionais.

A quantidade de informações digitais está crescendo a cada ano. Este tipo de dado é pouco analisado, pois muitas vezes não é estruturado, exigindo um alto poder de processamento e complexidade na sua interpretação, segundo Schonberger e Cukier (2013). Esta situação torna a possibilidade de correlação de informações inviável. Nos últimos anos houve grandes avanços tecnológicos na área de *hardware* e *software*, possibilitando equipamentos mais rápidos a um custo extremamente mais baixo, aliado a *software* capazes de incorporar e analisar dados não estruturados em um curto intervalo de tempo, disseminando assim uma nova cultura dentro das organizações. As empresas estão vendo este novo conceito de análise de dados como uma nova tecnologia fundamental para a tomada de decisões por permitir analisar qualquer tipo de informação digital em tempo real (GLOBO, 2013).

Este trabalho analisou os artigos publicados sobre *Big Data* nas bases de dados *Academic Source Complete (ASC)*, *Library*, *Information Science & Technology*

Abstracts (LISTA) e *Computers & Applied Sciences Complete* (CASC), fornecidas pela *Elton Bryson Stephens Company* (EBSCO). A pesquisa analisou artigos publicados entre 1º/01/2011 e 31/12/2013 que apresentam o termo "*big data*" no título.

Os resultados obtidos no mapeamento da literatura sobre o tema "*big data*" incluem a análise de um universo de 298 artigos publicados no período entre 2011 a 2013, que, por sua vez, resultou numa amostra final de 51 artigos. O objetivo da pesquisa é buscar os conceitos desta nova tecnologia, identificar quais os fatores que estão impulsionando a sua utilização bem como quais os projetos que estão utilizando esta nova tecnologia.

No capítulo 2 é apresentada a metodologia de pesquisa utilizada no trabalho.

No capítulo 3 é apresentado o referencial teórico, no qual foram expostos os conceitos sobre *Big Data*, bem como os termos que aparecem com mais frequência nos artigos pesquisados, dentro da temática principal, como mineração de dados (*Data Mining*), volume (*volume*), variedade (*variety*), velocidade (*velocity*) e Aplicações (*applications*). O referencial teórico foi obtido dos próprios artigos selecionados e de literatura de referência complementar também apresentada ao final do artigo.

No capítulo 4 é demonstrado o mapeamento realizado, de que forma foi feita a consulta aos artigos, como a publicação destes está distribuída por ano e por publicação de periódicos abordados dentro da temática pesquisada.

As considerações finais são apresentadas no capítulo 5.

2 METODOLOGIA

Este estudo identifica a produção internacional sobre o tema *big data*, sendo classificado como pesquisa qualitativa bibliométrica.

O método utilizado foi a pesquisa descritiva desenvolvida por meio da descrição das características de uma determinada população ou fenômeno, observando os fatos, registrando-os, analisando-os, classificando-os e interpretando-os, conforme orientam Gil (1999) e Andrade (2002).

Para realizar o mapeamento foi utilizado o software de gerenciamento de referências *Zotero*. Esta ferramenta tem como principais características a integração com navegadores, sincronização online, geração de citações em texto, rodapés e bibliografias, bem como a integração com os processadores de texto atuais (ZOTERO, 2013).

Araújo (2006, p.18-19) esclarece que a bibliometria, entre suas diversas funções, permite, por meio da análise das citações encontradas em publicações científicas, alcançar resultados.

[...] a identificação e descrição de uma série de padrões na produção do conhecimento científico. Com os dados retirados das citações pode-se descobrir: autores mais citados, autores mais produtivos, elite de pesquisa, frente de pesquisa, fator de impacto dos autores, procedência geográfica e/ou institucional dos autores mais influentes em um determinado campo de pesquisa; tipo de documento mais utilizado, idade média da literatura utilizada, obsolescência da literatura, procedência geográfica e/ou institucional da bibliografia utilizada; periódicos mais citados, “core” de periódicos que compõem um campo.

O estudo buscou auxílio da bibliometria para realizar o mapeamento das publicações disponibilizadas nas bases de dados de textos multidisciplinares fornecidas pela *Elton Bryson Stephens Company* (EBSCO) dentro do período de janeiro de 2011 e dezembro de 2013. A EBSCO é uma referência *online* acessível via internet que oferece uma variedade de base de dados com textos integrais e parciais disponíveis e acesso às fontes de pesquisas e publicações de renomados editores.

Para o mapeamento da produção acadêmica internacional sobre *big data* foram utilizadas as bases de dados *Academic Source Complete* (ASC), *Library, Information Science & Technology Abstracts* (LISTA) e *Computers & Applied Sciences Complete* (CASC).

A ASC é uma base de dados que oferece uma enorme coleção de periódicos em texto completo e multidisciplinar acadêmico mais valioso do mundo. Atualmente possui mais de 13.690 revistas indexadas e abstraídas, cerca de 9.100 periódicos em texto completo, mais de 7.900 revistas e jornais. Também possui conteúdos em *PDF* que datam de 1887. As referências citadas pesquisáveis são fornecidas para mais de 1.450 periódicos. Esta base é atualizada diariamente.

A base de dados LISTA está disponível gratuitamente para qualquer biblioteca. A base possui mais de 580 revistas do núcleo, cerca de 50 revistas prioritárias e 125 revistas mais seletivas, livros, relatórios de pesquisa e processos. Os textos incluem biblioteconomia, classificação, catalogação, bibliométrica, recuperação de informação *online* e gerenciamento de informações. O banco de dados se estende até meados dos anos 1960 e possui uma lista de tesouros em biblioteca, ciência da informação e tecnologia onde é possível pesquisar mais de 6.800 termos, 2.700 dos quais são termos preferenciais. A base também possui os perfis dos 5.000 autores indexados.

A base de dados CASC cobre as ciências aplicadas, representando o conhecimento coletado sobre os desafios tradicionais da engenharia e fornecendo material de pesquisa relativo às implicações comerciais e sociais das novas tecnologias. Esta base tem indexações e resumos de 2.000 periódicos científicos, publicações profissionais e outras fontes de referência de uma coleção diversificada. Também disponibiliza o texto completo para mais de 900 publicações periódicas. As áreas temáticas incluem as várias disciplinas de engenharia, teoria e sistemas de computadores, novas tecnologias, contexto social e profissional.

A pesquisa inicial ocorreu no dia 04/01/2014, identificando 298 artigos relacionados ao tema *big data*. Esta pesquisa buscou artigos nos quais a expressão *big data* estivesse presente no campo título (*TI Title*) da publicação e usando os limitadores de busca “texto completo” e “data de publicação” entre janeiro de 2011 e dezembro de 2013. Foram também utilizados os limitadores especiais “tipo de publicação” (*periodical*), “tipo de documento” (*article*) e “texto completo em *PDF*” para a base de dados ASP. Para as bases LISTA e CASC foram usados os

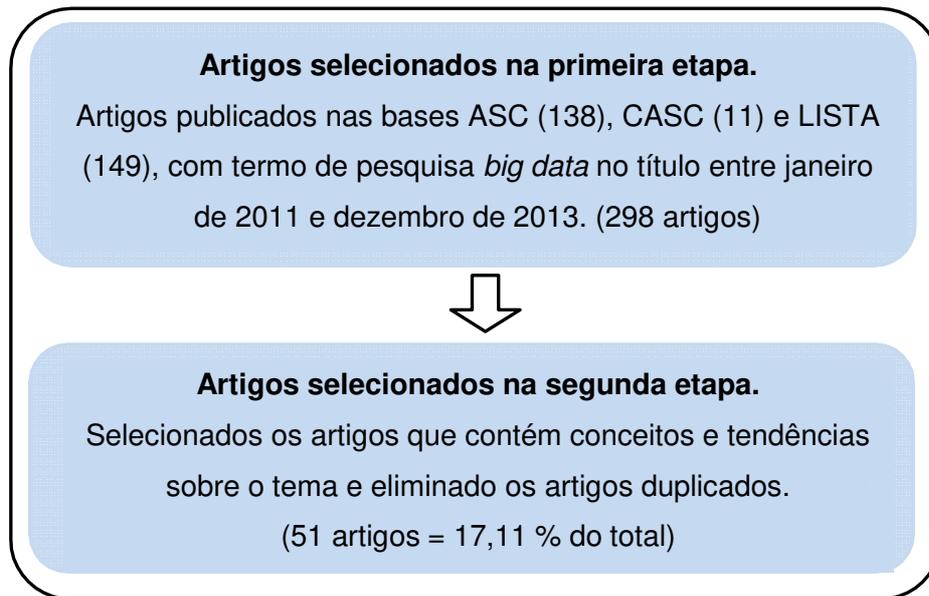
limitadores especiais “tipo de publicação” (*periodical*) e “tipo de documento” (*article*). Dos 298 artigos encontrados, 138 foram publicados na ASC, 11 na LISTA e 149 na CASC. Os tipos de documentos estão divididos em 190 revistas, 73 publicações comerciais e 35 revistas acadêmicas.

Embora o tema pesquisado seja recente, o número de publicações sobre o assunto vem crescendo nos últimos anos. Na pesquisa realizada não foram encontrados artigos acadêmicos, o que pode indicar algumas limitações no estudo realizado. O grande número de publicações hoje acontece em revistas especializadas em tecnologia e negócios.

O resultado inicial de 298 artigos foi submetido a um processo de seleção. Foram selecionados apenas os artigos que abordam o conceito da tecnologia e suas tendências na área de negócio. Havia publicações que estavam repetidas nas bases e estes artigos em duplicidade foram excluídos. Nesta seleção verificou-se que praticamente não existem artigos acadêmicos sobre este tema. Esta situação é explicada pelo fato da tecnologia ser recente e estar disponível para a grande maioria de empresas nos últimos anos devido aos avanços tecnológicos na área de *hardware* e *software* tornando viável a sua utilização. Deste processo de seleção, resultaram 51 artigos para compor a segunda etapa. Estes artigos representam 17,11% do total dos artigos e foram escolhidos para constituir a amostra final, tendo em vista a sua relação direta com a temática. Os artigos selecionados estão apresentados nas referências no capítulo 5.

A metodologia utilizada para a seleção dos artigos é apresentada na Figura 1.

Figura 1: Metodologia de seleção dos artigos.



Fonte: Elaborada pelos autores.

3 REFERENCIAL TEÓRICO

Este capítulo apresenta o referencial teórico que sintetiza o mapeamento realizado sobre o tema *Big Data* que embasa este estudo.

3.1 *Big Data*

A tecnologia *big data* foi criada para trabalhar com um volume grande de informações. A sua principal característica é que estas informações são dados pouco estruturados, havendo uma dificuldade na sua organização, exploração e análise de dados. Outro ponto importante é que estas informações precisam ser analisadas com agilidade, afim de que se possam buscar vantagens competitivas. Segundo Data (2012, p. 15-17) o conceito desta tecnologia está baseada sobre os 3 V que são velocidade, variedade e volume.

Nos últimos anos está havendo uma explosão de dados. Conforme Murnane (2012, p. 30-34) a explicação para este crescimento de informações está na acessibilidade e disponibilidade de novos dispositivos digitais que tornam o acesso a internet rápido e relativamente mais barato. Com este crescimento de usuários a lista de fontes digitais aumenta significativamente, pois a cada clique que os usuários da internet efetuam vão gerando um rastro digital e é justamente neste rastro digital que a tecnologia de *big data* pode ajudar a extrairmos informações e podem ser relevantes para uma determinada análise.

De acordo com Edward (2013, p. 20-30) uma definição mais pragmática define *big data* como a fronteira entre a equipe de tecnologia de informação e a equipe estratégica de negócios. Está na capacidade de uma empresa armazenar, processar e acessar todos os dados de que ela necessita para operar efetivamente, tomar decisões, reduzir riscos e servir os seus clientes com soluções de qualidade. Esta integração pode levar grandes vantagens competitivas. Mutter (2012, p. 20-21) define *big data* como um processo que envolve a coleta, corte e fragmentos de cubos de informações que podem ser rapidamente montados para identificar tendências macros ou criar novas perguntas diante a uma determinada situação.

Para Hasker (2012, p. 13-15), esta tecnologia proporciona oportunidades para as empresas olharem melhor para o seu negócio e fazer perguntas que até então

nunca foram pensadas ou que não foram possíveis de serem respondidas. A pergunta certa tem a resposta correta.

Novas ferramentas analíticas tendem a dar novas formas e maneiras de analisar as informações. Para Mahrt e Scharkow (2013, p. 20-33) o foco na análise de dados em *big data* leva a utilização de técnicas avançadas de análise, usando métodos estatísticos complexos para correlacionar os dados.

Analisando Marshal (2012, p. 213-224) pode-se resumir de maneira macro que uma análise de *big data* é feita através de um ciclo formado de 5 passos listados abaixo:

- 1) Buscar todas as informações de um determinado tema;
- 2) Identificar possíveis relações entre as informações;
- 3) Analisar as informações usando algoritmos matemáticos e estatísticos;
- 4) Em cima do resultado das análises iniciar novamente o processo de análise com os resultados e suas correlações;
- 5) Analisar os resultados e remover informações que não tem relação ou que não impactam outras informações.

Uma arquitetura de *big data* efetua estes passos milhares de vezes em poucos segundos. Isto é possível devido ao grande avanço tecnológico dos componentes computacionais permitem que sejam utilizados processos distribuídos para aumentar ainda mais a *performance* do resultado.

Na base de artigos analisados é citado que a meta da área de gestão do conhecimento ao longo dos anos é buscar a integração de informações a partir de múltiplas perspectivas internas e externas. O entendimento desta integração fornece a capacidade de tomada de decisões mediante alguns eventos. Segundo Lamont (2013, p. 8-11), no passado o custo da coleta e armazenamento gerava limitações para as empresas. As informações incompletas eram analisadas superficialmente e por isso muitas ligações de causa e efeito eram perdidas ou apontavam para situações que não eram verdadeiras. Somando a este fato as análises muitas vezes demoravam dias e no final o resultado gerado era desprezado.

No mundo digital moderno, a quantidade de informações digitais pode dobrar a cada 6 ou 7 meses em uma organização (ARNOLD, 2011). A análise e deste volume de dados pode conter respostas que explicam determinados fatos ou acontecimentos, bem como indicar tendências de fatos que ainda não foram observados. As organizações precisam entender que este volume de informações

pode ser comparado a um ativo de uma empresa. Quando este ativo não é utilizado de maneira eficaz e eficiente estamos forçando a depreciação do seu valor. A depreciação de um ativo da organização faz com que ela esteja em desvantagem perante outras organizações. A tecnologia *big data* tem como principal objetivo utilizar este ativo de forma eficaz e eficiente, gerando novos valores para as organizações segundo Pierce (2012, p. 28).

Esta nova tecnologia depende dos novos recursos computacionais de *hardware* e *software* para ser viável. Conforme Daugenti (2012, p. 50), outro fator importante para a mineração de dados é o conhecimento das pessoas. É preciso conhecer as informações e os conceitos, pois este conhecimento é essencial para entender e mapear as ligações entre os dados. Este conhecimento fica armazenado com as pessoas, sendo muito importante a utilização de ferramentas para armazenar este conhecimento. *Big data* exige o conhecimento de aspectos do negócio, atualidades e modelos matemáticos para que os modelos de análise possam extrair respostas mais completas e refinadas. O conhecimento precisa ser continuamente atualizado.

A utilização desta nova tecnologia traz benefícios em diversas áreas das organizações, porém como toda tecnologia é preciso avaliar quais os seus riscos, principalmente quando as informações possuem dados particulares. Conforme Koscieljew (2013) é importante avaliar os pontos positivos negativos. Atualmente estamos na era da informação, da disponibilização de dados e a utilização da tecnologia precisa ser utilizada com responsabilidade e ética, principalmente quando as análises envolvem dados de redes sociais, podendo expor os usuários a situações desagradáveis. Sua utilização deve se preocupar com as violações de privacidade. Um dos grandes desafios está na busca do ponto de equilíbrio entre a exploração de dados e garantias de privacidade das informações quando forem utilizadas informações disponíveis na internet.

Outro ponto citado em diversas publicações é sobre a privacidade dos dados. Goodendorf (2013, p. 29-33) destaca a preocupação com a privacidade dos dados quando são utilizadas informações das redes sociais ou informações pessoas sobre usuários disponibilizadas por serviços na *web*. Este autor destaca que a tecnologia ainda está em fase inicial de expansão e existem poucos padrões estabelecidos, mas neste tipo de serviço é preciso pesquisar sobre a credibilidade da fonte fornecedora de informações, evitando problemas futuros.

Analisando as publicações, Greenberg (2012, p. 46-47) anota a importância da utilização de *big data* nas organizações que buscam excelência na sua execução. Destaca que o que não se sabe pode prejudicar o negócio, por isso a tendência é que a tecnologia ganhe cada vez mais força e que todos os tipos de empreendimento cada vez mais tenham acesso a este tipo de análise. O diferencial da mineração de dados tende a ser uma vantagem no atual mercado competitivo.

Para Richards (2013, p. 18-24) as organizações precisam investir mais na segurança da informação. O conjunto de dados analisado precisa conter as informações precisas, sem violações de integridade para não haver distorções nas análises. É importante criar políticas de segurança para preservar todo o tipo de informação de negócio gerado na empresa.

A tecnologia de *big data* trabalha com grandes conjuntos de dados, porém para que as análises sejam mais efetivas e busquem uma correlação maior sobre as informações é preciso que as informações recebam determinados tratamentos e sejam agrupadas em pesquisas menores. Esta técnica é utilizada para que as ferramentas de análises possam interpretar o resultado destas pequenas análises e buscar a correlação entre as análises. Para Badke (2012, p. 47-49) esta é a maneira mais eficaz de buscar a velocidade da análise de grandes volumes de dados com qualidade. As pesquisas muito grandes tendem retornar muitos resultados, gerando um tempo maior de processamento e necessitando uma interação constante sobre os resultados, visto que as ferramentas encontram poucas correlações entre os dados, fazendo com que a análise muitas vezes não chegue a uma resposta esperada no tempo pré-estabelecido.

A pesquisa sobre o tema "*big data*", permite destacar os termos que aparecem com mais frequência dentro desta temática, destacando-se os três pilares da tecnologia que são a Velocidade (*Velocity*), Variedade (*Variety*) e o Volume de dados. Outro ponto que é abordado de forma constante na temática são as habilidades necessárias para gerir projetos deste tipo. Um ponto bastante abordado são os *cases* de sucesso que estão sendo realizados com esta nova tecnologia e as tendências de aplicações que utilizam estas análises. Estes termos são sintetizados a seguir.

3.1.1 Volume

A principal característica que diferencia *data mining* (mineração de dados) com *big data* é o volume de informações. A mineração de dados é o processo de descoberta de informações em conjuntos de dados. Esta técnica usa modelos matemáticos para derivar padrões e tendências que existem nos dados (MSDN, 2014). Para Skinner (2013, p. 18-19), o volume de dados nos impulsiona na direção de novas perspectivas e na maneira de como podemos interagir e entender melhor estes dados.

Em *big data* a análise de informações é feita sobre uma quantidade muito grande de informações. Com o contínuo crescimento das informações o grande desafio dos CIOs (*Chef Information Officer*) está em gerenciar este volume de informações otimizando seu armazenamento (INBAR, 2013). Para o gerenciamento ser efetivo é preciso compreender as informações existentes, identificar padrões e relações. A análise das informações pode conter novas informações o que torna este tipo de análise complexo.

No início de 2011 um grande volume de dados era medido em *petabytes* (10^{15}), já em 2012 surgiram medições em *zettabytes* (10^{21}) comprovando que o volume de dados está crescendo de forma exponencial (BIG, 2013).

3.1.2 Variedade

Segundo Royster (2013, p. 2-10) existem basicamente dois tipos de conjunto de dados utilizados nas análises, os dados estruturados e dados não estruturados. Os dados estruturados são números e palavras que podem ser facilmente categorizados e analisados. Os dados não estruturados incluem informações mais complexas como comentários de clientes a partir de sites comerciais, fotos e outros conteúdos multimídia, assim como comentários em redes sociais. Estes dados não podem ser facilmente separados em categorias ou analisados numericamente.

Uma das principais características do *big data* é a variedade dos dados. Diferente de outros sistemas onde as informações são armazenadas de forma estruturada nesta tecnologia grande parte das informações não são estruturadas, com isso sua análise e compreensão tornam-se bem mais difícil. Segundo Estes (2012, p. 12) para a análise deste tipo de informação é preciso ter algoritmos mais

robustos, que encontram padrões na linguagem humana. A partir destes padrões os algoritmos efetuam a mineração de dados para construir significados a partir de simetrias em dados não estruturados e usando métodos estatísticos. A análise de dados não estruturados utiliza informações de palavras chaves para categorizar a informação.

As empresas já utilizam sistemas ERP (*Enterprise Resource Planning* – Planejamento de Recursos Empresariais) e até mesmo de BI (*Business Intelligence*) que contém informações onde é possível consultar determinadas informações, mas existem outros tipos de informações que também poderiam ser consultadas e trazer novos resultados ou respostas a determinadas situações, estas informações estão em documentos, relatórios, desenhos, fotografias, vídeos, e-mails, blogs e *wikis*. Para White (2011, p. 21) estas informações não são estruturadas, mas complementam dados das organizações e por isso devem ser aproveitadas. As informações já estão na organização, o que é preciso são mecanismos de interpretação e análise de dados para interagir estes dados com as demais informações gerenciais das organizações.

Segundo Ball (2013, p. 25-27), a variedade de dados é a característica responsável por buscar dados, na maioria das situações, não estruturados e fazer um processamento onde estas informações tenham ligações, tenham significados e possam ser interpretados por um ser humano ou sirvam de entrada para um aplicativo que tenha como analisar estas informações. Esta tarefa é difícil e exige conhecimento sobre o tema das informações, é preciso fazer vários refinamentos até chegar a um resultado onde o modelo possa ser analisado para buscar novas informações.

Muitas empresas já armazenam um volume consideravelmente grande de informações. Estas informações são utilizadas em sistemas gerenciais provendo análises de BI. Conforme Greengard (2012, p. 20-23), a ascensão da mídia social e a crescente utilização e facilidade de acesso a *tablets* e *smartphones* está fazendo com que as organizações incorporem a mídia social aos seus modelos de dados. Com este novo modelo de dados é possível utilizar as análises atuais fazendo análises e cruzamentos com as redes sociais, sendo possível analisar os perfis dos usuários, preferências e muitas outras análises ampliando de maneira muito grande o campo de análise.

3.1.3 Velocidade

Um dos grandes desafios desta tecnologia é a velocidade de resposta (WAXER, 2013). Qualquer análise que seja feita hoje precisa chegar rapidamente ao resultado. Atualmente as informações estão crescendo rapidamente e com isso as análises precisam ser cada vez mais precisas e trazer respostas imediatas. As empresas não podem esperar um tempo de processamento de horas ou até mesmo dias para ter um retorno sobre os dados analisados. Preimesberger (2011) afirma que as novas ferramentas precisam lidar com *terabytes* e *petabytes* com agilidade, para que a informação possa ser analisada em tempo, antes que novos eventos aconteçam e mudem o cenário atual.

O volume de informações de uma empresa de cartões de crédito cresce rapidamente. Utilizando a tecnologia de *big data* é possível fazer uma série de análises sobre estas informações e encontrar novas oportunidades ou melhorar processos já existentes. No mundo globalizado as informações mudam rapidamente e para tirar benefícios desta nova tecnologia é fundamental que as análises sejam rápidas, pois com os resultados obtidos a empresa precisa tomar as devidas ações antes de seus concorrentes.

3.2 Competências

Para Stackpole (2012, p. 22-25), as organizações precisam cada vez mais de cientistas de dados devido a disseminação da tecnologia de mineração de dados. Este tipo de profissional é fundamental para que a empresa possa evoluir nas análises de grandes volumes de dados. As habilidades analíticas são fundamentais para a interpretação dos dados e na geração de novas análises.

No estudo realizado foi verificado que uma das principais características para o crescimento e sucesso de aplicações *big data* nas organizações são as competências que as pessoas precisam adquirir. Segundo Harbert (2013, p. 23-27), estas competências são essenciais para as análises e interpretação dos resultados. Este conhecimento diferenciado vai gerar novos tipos de profissionais no mercado.

O Quadro 1 faz um levantamento dos novos profissionais segundo as análises de Harbert (2013) sobre *big data*.

Quadro 1 – Profissionais de *big data* e suas funções.

Profissionais	Funções Desempenhadas
Cientista de Dados	Pessoas que contém sólidos conhecimentos em matemática, estatística e inteligência artificial
Arquiteto de Dados	Pessoas com habilidades sólidas em programação para fazer a integração de dados não estruturados
Analista de Dados	Pessoas com habilidades sólidas no negócio da organização, precisam ter a capacidade de analisar as informações e buscar relações nas informações com todas as áreas da empresa
Agente de Dados	Pessoas que recebem os resultados das análises e montam o plano de ação para efetuar as devidas alterações dos processos de negócio
Engenheiro de Dados	Pessoas com conhecimentos de infraestrutura e gestão de processos para montar o ambiente que dê suporte as análises

Fonte: Adaptado pelos autores a partir de Harbert (2013, p. 23-27).

Segundo Lowry (2013, p. 82-83), *big data* tem um mercado promissor, mas é importante que atrás desta imensa quantidade de informações estejam profissionais dedicados e treinados para que possam fazer as relações destes dados. Um computador pode analisar milhares de informações, mas ele não pode pensar e refletir o dado de forma não matemática e chegar à conclusão que as informações se complementam ou tem ligação entre si.

As análises de *big data* precisam de muitos dados e para a integração destes dados são necessárias várias estruturas de metadados. Segundo Kubick (2012, p. 26-28), outro fator importante para este tipo de projeto são as grandes ideias, devido a complexidade das informações é preciso ter agilidade para fazer a relação dos dados, buscar a integração de diferentes tipos de informações. Pessoas com bastante conhecimento do negócio têm mais facilidade em fazer estas relações. Esta é a etapa mais lenta de um projeto de mineração de dados, mas é fundamental para termos sucesso no projeto.

3.3 Cases de Big Data

O primeiro projeto utilizando os conceitos de *big data* foi o Projeto Genoma que iniciou em 1990 e terminou em 2003 (ROSE, 2012). Este projeto tinha produzido um mapa de cerca 20.000 a 25.000 genes que formam a espécie humana, ele foi decisivo na compreensão da humanidade. A coleta, análise e interpretação da grande quantidade de dados levou 12 anos exigindo um grande esforço e reunindo

brilhantes mentes humanas e o que havia de mais moderno em equipamentos. Inicialmente o projeto não havia tempo para terminar. A mente humana não mudou, mas os equipamentos evoluíram muito, de acordo com as pesquisas hoje este projeto levaria bem menos tempo, por isso esta tecnologia e tipo de análise de dados está sendo tão disseminada em diversas organizações.

Para Rogers (2011, p. 14-18), a convergência das novas tecnologias, a redução no custo total do projeto e tempo necessários para a análise de dados ajudou a tecnologia de *big data* a não ser apenas mais um sonho para as empresas, e sim uma realidade muito próxima e essencial para melhorar o alinhamento estratégico da empresa.

Cada organização possui a sua maneira de armazenar e gerenciar as suas informações. Segundo Five (2013, p. S2), para melhorar a mineração de dados de grandes volumes é preciso que as empresas melhorem a gestão de dados, com isso são citados alguns hábitos que melhoram esta gestão, tais como: os departamentos dentro das organizações precisam buscar algum software para manter arquivadas as informações geradas em conversas, reuniões e trocas de e-mails, outro fator importante é que uma portal interno, onde os colaboradores possam acessar e incluir informações é uma forma de centralizar a informações e obter detalhes que possam ficar despercebidos. A utilização deste portal deve ser fácil e estar acessível em todas as áreas. Os gestores precisam incentivar que os colaboradores usem o portal para esta troca de informações.

No estudo realizado foi verificado que para trabalhar com grandes volumes de dados é necessário ter uma arquitetura bem desenhada que possibilite esta análise. O custo de implantação de uma solução de *big data* ainda apresenta valores altos, porém existem muitas ferramentas *open source* que estão disponíveis no mercado e que inclusive são usadas e customizadas por grandes *players* de mercado segundo Baquero, Palacios e Molloy (2013, p. 29-35). Uma arquitetura *open source* muito utilizada nos projetos de *big data* é o *Apache Hadoop* que é um *software* para mapear as bases de dados e usar processamento paralelo para realizar as análises e interações dos dados. Conforme Farber (2013, p. 17-20), a concorrência de empresas que oferecem *hardware*, *software* e soluções para apoiar a análise e gestão de grandes volumes de dados é grande. O gerenciamento destas informações depende do nível de organização da organização, muitas vezes softwares mais simples, com um custo menor se adaptam mais facilmente a

realidade da empresa, é preciso fazer um levantamento de requisitos olhando para a estrutura atual da empresa.

Na base de artigos pesquisados neste estudo foi verificado que *big data* é uma tecnologia que está ganhando força nos últimos anos e as organizações de todos os tamanhos podem tirar proveito desta tendência tecnológica (BARTIK, 2013). A tecnologia ainda apresenta custos elevados na aquisição de suas ferramentas necessárias, mais com a popularidade da computação em nuvem e com as ferramentas certas e uma equipe dedicada é possível fazer uma organização nos dados e já utilizar esta tecnologia para ajudar nas decisões estratégicas e melhorar o desempenho dos negócios.

Segundo Matelski (2012, p. 32), *big data* representa um novo mercado de muitas oportunidades, mas é preciso que os executivos saibam gerenciar o rápido crescimento e acrescenta os desafios desta tecnologia, incluindo a complexidade e custo que ainda são elevados. Os líderes devem abordar as implicações do armazenamento de informações e da captura de informações por parte das empresas, a ascensão de multimídia, mídia social e da internet. Para Arnold (2012, p. 1-34), este grande volume de informações será analisado com métodos estatísticos ligados a sistemas de aprendizagem de máquina que juntos conseguem analisar os diversos cenários e fazer correlações. Os resultados serão analisados diversas vezes, pois a cada interação o sistema consegue aprender e buscar as correlações das informações até chegar a um cenário do tipo causa-efeito.

As análises sobre grandes volumes de dados podem oferecer novas oportunidades para as ciências sociais. Com estes conjuntos de informações os pesquisadores de diversas áreas podem aprofundar temas que até então não possuíam respostas suficientes para explicar determinados comportamentos. A possibilidade de avaliar pontos mais sutis podem revelar novas respostas que até então não eram possíveis de ser vistas, fazendo com que as tomadas de decisões sejam mais racionais (LUCRAFT, 2013).

Nos artigos estudados foi verificado que a utilização da tecnologia está sendo disseminada em montadoras. Conforme Noor (2013, p. 32-37), as fábricas estão usando a tecnologia de *big data* para aperfeiçoar o seu processo de produção. A análise de grande volume de dados em um curto espaço de tempo permite avaliar o processo de produção e prevenir a ocorrência de falhas, buscando uma qualidade maior do produto. Este tipo de análise é utilizado na fabrica de mísseis em

Huntsville, no Alabama nas operações de montagem dos armamentos. Quando é detectada alguma anomalia na montagem de algum componente todo o processo é parado, neste instante são identificados os processos origens que geraram o problema, garantindo produtos com uma qualidade excepcional. A montadora de automóveis Ford utiliza *big data* no seu novo modelo de carro, o Focus. Uma infinidade de sensores monitora o tempo inteiro todas as reações do carro, informando ao condutor diversas situações e muitas vezes fazendo automaticamente processos de correção no modo de condução do veículo para evitar possíveis acidentes.

Conforme Ann, Koonin e Shipp (2012, p. 4-7), as análises de *big data* estão trazendo uma revolução na ciência e tecnologia. Esta revolução será ainda mais profunda quando começar a interagir ainda mais com as decisões das pessoas. Estas mudanças vão afetar a forma de vida das pessoas impactando no estilo de vida da sociedade. As ciências sociais como a economia, sociologia, ciência comportamental e ciência política será totalmente transformada. O Quadro 2 mostra como o crescimento de dados pode ser utilizado na sociedade e mudar o seu comportamento.

Quadro 2 - Crescimento exponencial dos dados: alguns exemplos de uso potencial.

Disponibilidade de dados	O que você pode fazer com esses dados?
<p>Dados de localização</p> <ul style="list-style-type: none"> • Telefone celular • Câmeras de vigilância <p>Informação Comercial</p> <ul style="list-style-type: none"> • Transações com cartão de crédito • Vendas de imóveis • Pesquisas on-line <p>Informações sobre saúde</p> <ul style="list-style-type: none"> • Registros médicos eletrônicos • Internações hospitalares • Dispositivos para monitorar os sinais vitais • Vendas de Farmácia <p>Outras informações</p> <ul style="list-style-type: none"> • Imagem óptica, infravermelho e espectral • Dados Meteorológicos • Dados Sísmicos • Níveis de Radiações solares <p>Dados Projetados</p> <ul style="list-style-type: none"> • Dados administrativos • Censo da população 	<p>Localização</p> <ul style="list-style-type: none"> • Medir a migração urbana e rural • Montar mapa para rastrear os desastres naturais com os movimentos populacionais • Identificar bairros com maiores índices de problema nos serviços sociais <p>Saúde</p> <ul style="list-style-type: none"> • Mapear o comportamento humano e correlacionar com os resultados de doenças, tais como a diabetes e o câncer <p>Transporte</p> <ul style="list-style-type: none"> • Otimizar as operações para o fluxo de tráfego em determinados horários, regiões • Desenvolver planos de infraestrutura para o zoneamento e transporte público <p>Energia relacionada</p> <ul style="list-style-type: none"> • Prática de monitoramento, geração de relatórios e verificação das emissões de gases de efeito estufa • Detectar os riscos (por exemplo, vazamentos e plumas) <p>Gestão de emergências</p> <ul style="list-style-type: none"> • Estabelecer padrões de eficiência energética para edifícios e estabelecimentos comerciais • Utilizar o conhecimento dos comportamentos sociais para incentivar eficiência energética

Fonte: Adaptado pelos autores a partir de Ann, Koonin e Shipp (2012, p. 4-7).

Uma das grandes áreas de estudo que está utilizando a tecnologia de big data é a medicina. Conforme Conger (2012, p. 6-15), a ciência médica tem enormes bancos de dados biológicos relacionados aos pacientes. É um desafio para gerenciar toda esta quantidade de dados em termos de armazenamento, análise, compartilhamento e proteção. O uso da tecnologia do *big data* está ajudando os cientistas a comparar genomas de doenças similares para detectar vários fatores subjacentes, como efeitos colaterais e outros sintomas. Explorar estes dados complexos podem diminuir os custos e o desperdício de cuidados a saúde, no futuro estas informações serão essenciais nas consultas clínicas e vão ajudar os médicos a receitarem remédios mais eficazes (MITHAS, EARLEY, MURUGESAN e DJAVANSHIR. 2013).

Para Abbott (2013, p. 225-292), os dados sobre pacientes recolhidos das redes estaduais e federais de saúde têm o potencial de inaugurar uma nova era de regulação de medicamentos. A análise destas informações pode gerar infinitos resultados clínicos sobre o uso de drogas, variáveis demográficas, motivos de utilização dos entorpecentes e quais medicamentos são apropriados para determinada situação. Quanto maior o numero de informações analisadas, maior a probabilidade do medicamento utilizado ser mais adequado para determinado paciente. Este tipo de análise tende a revolucionar a área da saúde. É importante que os governos incentivem esta prática e troca de informações entre os setores da saúde.

MacDonald (2012, p. 14-15) cita que já existem aplicativos desenvolvidos em parceria com o *Google* que utilizam o *Tweeter* para verificar tendências emergentes para a saúde tais como surtos virais. Também já foram desenvolvidos sites onde o paciente informa um número de sintomas e são diagnosticadas em tempo real possíveis doenças baseadas na pontuação dos sintomas, dados epidemiológicos e tendências de saúde.

As análises de *big data* no campo da medicina precisam ser feitas com muito cuidado sendo fundamental a veracidade dos dados. Segundo Milstein e Jha (2013, p. 237-238), algumas análises que utilizam sintomas de pacientes e outros fatores para determinar doenças ou medicamentos passíveis de ser utilizados podem apresentar problemas quando o paciente omitir informações importantes ou mesmo informar dados errados, mesmo que seja por engano. Nestas situações as

correlações que dão origem “a novas descobertas” podem ser perdidas e resultar em dados inconsistentes.

Um projeto chamado *Durkheim* está fazendo uma coleta de dados com diversas informações de veteranos de guerra dos EUA. Estes dados serão analisados através de técnicas de *big data* buscando identificar padrões para ajudar os médicos especialistas em saúde mental a mapear os fatores de incidência de suicídios entre os veteranos (USING, 2013).

Segundo Daniels (2013, p. 19), *big data* representa mais uma tecnologia para ajudar as organizações a buscarem a excelência de seus serviços, por isso um ponto chave para o sucesso do projeto é conhecer bem os conceitos da tecnologia e procurar buscar a tecnologia mais aderente ao negócio. O custo deste tipo de projeto ainda é bem elevado, mas é essencial que os gestores tenham conhecimento do tipo de informação que eles possuem, bem como está sendo feito o seu gerenciamento de contenção. Este ponto é fundamental para que sejam utilizados *players* compatíveis com a sua arquitetura atual de sistemas, reduzindo os custos de operações desnecessárias.

Para Church e Dutta (2013, p. 23-31), a tecnologia de *big data* está inserida em vários ramos da sociedade, muitas vezes passando despercebida. Hoje a tecnologia está disponível em diferentes fontes que usam estas informações para oferecer serviços específicos tais como os buscadores de conteúdo na internet como o *Google* e *Bing*, nas trocas de mídia das redes sociais, nas transações de compra (via cartões de crédito, Atividade *Amazon* ou *Ebay* em *iPads*), fotos digitais e vídeos postados em sites como o *Flickr* e *Photobucket*, sensores meteorológicos e muitos outros dispositivos.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Este capítulo destina-se à apresentação e análise dos resultados da pesquisa.

4.1 Análise do Mapeamento

Os resultados obtidos no mapeamento da literatura sobre o tema “*big data*” incluem a análise de um universo de 298 artigos publicados no período entre 2011 a 2013, que, por sua vez, resultou numa amostra final de 51 artigos. O objetivo da pesquisa é buscar os conceitos desta nova tecnologia, identificar quais os fatores que estão impulsionando a sua utilização bem como quais os projetos que estão utilizando esta nova tecnologia.

A apresentação e a análise dos resultados estão descritas a seguir em forma de tabulações e gráficos, das quais consta a quantidade de artigos publicados por ano, o total de artigos por periódicos e a quantidade de autores por periódico.

Na Tabela 1 é exibido o número de publicações ao longo do ano. É possível identificar que nos últimos anos a tecnologia vem ganhando força no mercado.

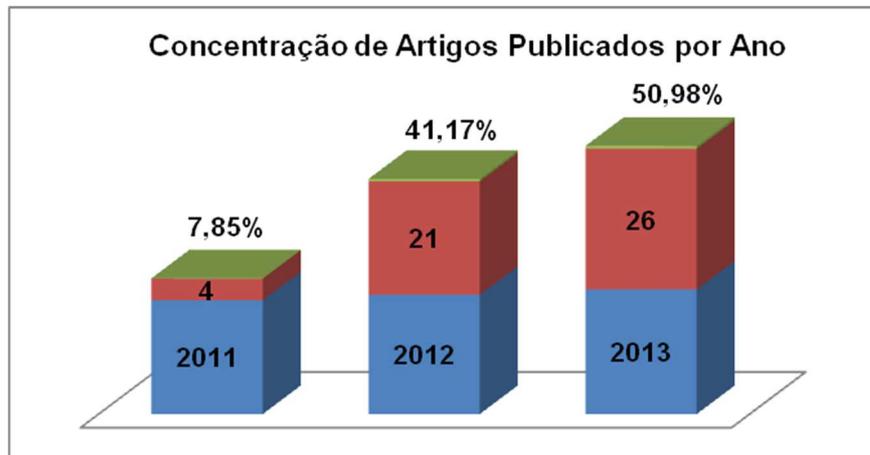
Tabela 1 – Concentração de artigos publicados por ano.

Ano	Publicações	%
2011	4	7,85 %
2012	21	41,17 %
2013	26	50,98 %
Total	51	100,00 %

Fonte: Elaborada pelos pesquisadores.

Verifica-se no Gráfico 1 que os anos que possuem maior volume de publicações e que compõem a amostra são 2012 e 2013, comprovando que apenas nos últimos anos é que o tema de *big data* começou a ganhar destaque no mercado.

Gráfico 1 - Concentração de artigos publicados por ano



Fonte: Elaborado pelos pesquisadores.

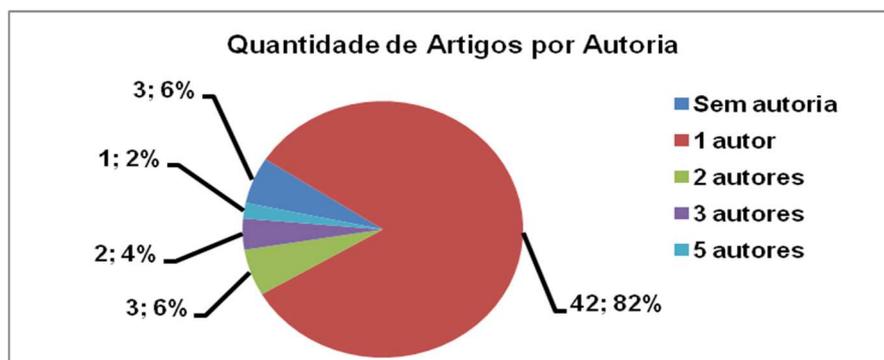
Destaca-se na Tabela 2 e no Gráfico 2 que, quanto à autoria dos artigos analisados nas publicações, que a mais utilizada é a autoria de apenas 1 autor com 82,35%. Apenas em 17,65 % dos artigos existe mais de autor na publicação ou ele não foi informado. Foi verificado que três artigos que fazem parte de duas revistas de tecnologia, a *Database Trends & Applications* e a *KM World* não apresentaram o autor em suas publicações.

Tabela 2 – Quantidade de autorias nos artigos selecionados.

Quantidade de Autores	Artigos	%
Sem autoria	3	5,88 %
1 autor	42	82,35 %
2 autores	3	5,88 %
3 autores	2	3,93 %
5 autores	1	1,96 %
Total	51	100,00 %

Fonte: Elaborada pelos pesquisadores.

Gráfico 2 – Quantidade de artigos por autoria.



Fonte: Elaborado pelos pesquisadores.

Na Tabela 3 é exibido o número de publicações por periódico. É possível identificar que o periódico com o maior número de artigos publicados, dentro do universo da amostra, é o Database Trends & Applications, representando 11,76% do total de publicações por periódicos.

Tabela 4 – Quantidade de publicações por periódico.

Periódico	Quantidade	%
Database Trends & Applications	6	11,76%
KM World	4	7,84%
Computerworld	3	5,88%
Significance	2	3,92%
Online	2	3,92%
Information Today	2	3,92%
Scientific Computing	2	3,92%
Information Management Journal	2	3,92%
Feliciter	2	3,92%
Information Security	2	3,92%
IT Professional	2	3,92%
Online	1	1,96%
Econtent	1	1,96%
Stanford Medicine	1	1,96%
Editor & Publisher	1	1,96%
Applied Clinical Trials	1	1,96%
Information Services & Use	1	1,96%
Engineering & Technology	1	1,96%
The Environmental Magazine	1	1,96%
Iowa Law Review	1	1,96%
American Journal of Managed Care	1	1,96%
Key Words	1	1,96%
Mechanical Engineering	1	1,96%
OD Practitioner	1	1,96%
Journal of Broadcasting & Electronic Media	1	1,96%
Occupational Outlook Quarterly	1	1,96%
eWeek	1	1,96%
CRM Magazine	1	1,96%
SC Magazine: For IT Security Professionals	1	1,96%
Baseline	1	1,96%
Siliconindia	1	1,96%
Research Information	1	1,96%
AALL Spectrum	1	1,96%
Total	51	100,00 %

Fonte: Elaborada pelos pesquisadores.

No estudo, identificaram-se conteúdos específicos que podem gerar construtos de pesquisa, conforme as temáticas destacadas na Tabela 4 e no Gráfico 3.

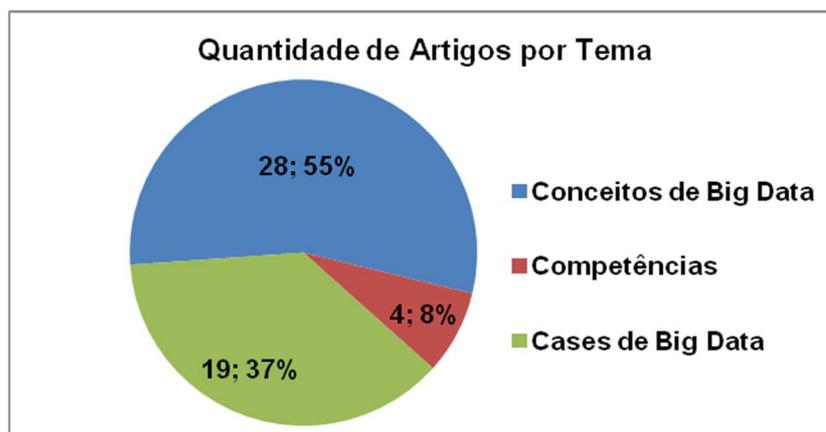
Tabela 4 – Quantidade de artigos por temática.

Conteúdo	Artigos	%
Conceitos de <i>Big Data</i>	28	54,90 %
Cases de <i>Big Data</i>	19	37,25 %
Competências	4	7,85 %
Total	51	100,00 %

Fonte: Elaborada pelos pesquisadores.

Analisando os artigos estudados observou-se que 28 abordam os conceitos desta nova tecnologia, representando 54,90% dos artigos analisados. Outro ponto importante foi o número de artigos empíricos apresentando *cases* utilizando esta tecnologia, e que representam 37,25% dos artigos pesquisados. Analisando os *cases* verificou-se que a área da Medicina concentra o maior número de aplicações desta tecnologia.

Gráfico 3 – Quantidade de artigos por tema.



Fonte: Elaborado pelos pesquisadores.

5 CONSIDERAÇÕES FINAIS

Para o presente artigo foi mapeada a produção internacional sobre *Big Data* no período de janeiro de 2011 e dezembro de 2013.

Os resultados da análise coletaram um universo de 51 artigos, os quais foram apresentados pelo número de publicações ao longo dos anos e pelo número de artigos por periódicos. O tema pesquisado ainda é recente, por isso houve dificuldades em encontrar artigos científicos. O número de publicações sobre o assunto vem crescendo nos últimos anos. O grande número de publicações mapeadas acontece em periódicos de revistas especializadas em tecnologia, negócios e acadêmicas.

A partir do mapeamento foi possível identificar que apesar da tecnologia ser recente nos últimos anos o número de empresas que estão iniciando projetos nesta área vem crescendo rapidamente. Esta é uma tendência, pois nos últimos anos o número de informações geradas é cada vez maior e esse número cresce de forma exponencial. A tecnologia de *big data* vem para analisar estes conjuntos de dados e buscar correlações existentes com outras informações estruturadas. Estas correlações podem mostrar uma situação que até então não era percebida pelos gestores, mas pode ajudar a melhorar o desempenho da organização, aumentando sua competitividade no mercado. As análises podem identificar uma tendência de mercado.

Analisando os periódicos foi identificado que a tecnologia está se disseminando, também, devido aos grandes avanços tecnológicos na área de hardware e software, possibilitando arquiteturas mais robustas e que apresentem resultados de forma mais rápida e com uma confiabilidade maior. A facilidade de acesso a internet com o uso de *tablets* e *smartphones* também incentiva a geração de várias informações nas de redes sociais. Outro ponto bastante mencionado é que o custo de projetos deste porte está diminuindo, com isso mesmo empresas de pequeno porte podem começar a pensar em analisar as suas informações. Existem grandes *players* de mercado, pagos e até mesmo *open source*, é preciso que as organizações verifiquem o tipo de projeto que pretendem implementar bem como qual é a arquitetura da empresa antes de escolher um fornecedor desta tecnologia. Este tipo de análise precisa analisar grandes conjuntos de dados e gerar

um resultado de forma rápida para que as devidas ações baseadas neste resultado possam ser aplicadas no tempo correto.

O interesse das organizações neste tipo de análise também requer que a empresa se prepare para poder implantar projetos deste porte. É preciso mapear as informações já existem e coletar os dados que ainda não são coletados. Segundo a análise do tema é fundamental que hajam profissionais dedicados a este tipo de análise com sólidos conhecimentos no negócio da organização e princípios de informática, matemática e estatística. O resultado deste tipo mineração de dados muitas vezes pode ficar confuso, por isso é importante que as pessoas envolvidas tenham a habilidade de analisar e interpretar o resultado.

Por fim, na pesquisa foi verificado que empresas de diversos segmentos já estão implantando projetos utilizando esta tecnologia. Os projetos de *big data* vão desde empresas de departamentos estratégicos como segurança e outros departamentos governamentais, como fábricas e principalmente na área da saúde. Muitos avanços da medicina estão acontecendo devido a este tipo de análise, pois é possível fazer o cruzamento de diversos sintomas do paciente, cruzando com as medicações e de acordo com o genoma de cada indivíduo chegar a um medicamento que apresente melhores resultados em um tempo menor. Assim como inúmeras informações são geradas diariamente é preciso que as organizações se preocupem no gerenciamento desta informação para utilizá-la da melhor forma possível.

REFERÊNCIAS

- ABBOTT, R. Big data and pharmacovigilance. **Iowa Law Review**, v. 99, p. 225-292, 2013.
- ANDRADE, M. M. **Como preparar trabalhos para cursos de pós-graduação: noções práticas**. 5. ed. São Paulo: Atlas, 2002.
- ANN, S. K.; KOONIN, S. E.; SHIPP, S. Big data and city living - what can it do for us?. **Significance**, v. 9, p. 4-7, 2012.
- ARAÚJO, C. A. **Bibliometria: evolução histórica e questões atuais**. Porto Alegre, v. 12, p. 11-32, 2006.
- ARNOLD, E. The big deal about big data. **Information Today**, v. 29, p. 1-34, 2012.
- ARNOLD, S. E. Big data: the new information challenge. **Information**, v. 35, p. 26-27, 2011.
- BADKE, W. Big search, big data. **Online**, v. 36, p. 47-49, 2012.
- BALL, S. Managing big data: what's relevant?. **AALL Spectrum**, v. 18, p. 25-27, 2013.
- BAQUERO, A. V.; PALACIOS, R. C.; MOLLOY, O. Business process analytics using a big data approach. **IT Professional**, v. 15, p. 29-35, 2013.
- BARTIK, D. Breaking down the five most common big data myths. **Database Trends & Applications**, v. 27, p. 32-33, 2013.
- BIG Data Demands Big Connectivity. **Database Trends & Applications**, v. 27, p. 37, 2013.
- CHURCH, A. H.; DUTTA, S. The promise of big data for OD. **OD Practitioner**, v. 45, p. 23-31, 2013.
- CONGER, K. Big data: what it means for our health and the future of medical research. **Stanford Medicine**, v. 29, p. 6-15, 2012.
- COURTNEY, M. The larging-up of big data. **Engineering & Technology**, v. 7, p. 72-75, 2012.
- DATA Management in the Era of Big Data. **Database Trends & Applications**, v. 26, p. 15-17, 2012.
- DANIELS, R. T. The key to smart big data: know the technology. **Information Today**, v. 30, p. 19, 2013.
- DAUGENTI, T. W. Big data requires new skills. **SC Magazine: For IT Security Professionals**, v. 23, p. 50, 2012.

- EDWARD, J. Big data in 2013: what to expect. **Information Management Journal**, v. 47, p. 20-30, 2013.
- ESTES, T. Automated understanding: the future of big data. **KM World**, v. 21, p. 10, 2012.
- FARBER, R. Big money for big data. **Scientific Computing**, p. 17-20, 2013.
- FIVE Big Data Habits of Highly Effective Organizations. **KM World**, v. 22, p. S2, 2013.
- FUCHS, H. **Big data, web analytics e data scientists**. Disponível em <<http://www.dp6.com.br/big-data-web-analytics-e-data-scientists/>>. Acesso em: 03 jan. 2013.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 1999.
- GLEICK, J. **A Informação**. São Paulo: Companhia das Letras, 2013.
- GLOBO. **Como funciona o big data (2013)**. Disponível em: <<http://oglobo.globo.com/infograficos/bigdata/>>. Acesso em: 27 dez. 2013.
- GOODENDORF, L. Managing big data privacy concerns. **Information Security**, v. 15, p. 29-33, 2013.
- GREENBERG, P. Big data, big deal. **CRM Magazine**, v. 16, p. 46-47, 2012.
- GREENGARD, S. BIG data unlocks business value. **Baselinemag**, v. 114, p. 20-23, 2012.
- HARBERT, T. Big data, big jobs?. **Computerworld**, v. 47, p. 23-27, 2013.
- HASKER, C. Big data opens new frontiers for the reporting and analytics industry. **Database Trends & Applications**, v. 26, p. 13-15, 2012.
- INBAR, D. Making the most out of your data: big data opportunities. **Siliconindia**, p. 34-35, 2013.
- LAMONT, J. Big data has big implications for knowledge management. **KM World**, v. 21, p. 8-11, 2012.
- LOWRY, D. Indexing and big data. **Key Words**, v. 21, p. 82-83, 2013.
- KIM, W. C.; MAUBORGNE, R. **A estratégia do oceano azul**. 19. ed. São Paulo: Campus, 2005.
- KOSCIEJEW, M. The era of big data. **Feliciter**, v. 59, p. 52-55, 2013.
- KOSCIEJEW, M. The perils of big data. **Feliciter**, v. 59, p. 32-35, 2013.
- KUBICK, W. R. Big data, information and meaning. **Applied Clinical Trials**, v. 21, p. 26-28, 2012.

- LUCRAFT, M. Can big data benefit society?. **Research Information**, v. 63, p. 6-7, 2013.
- MACDONALD, C. Using big data to improve health. **The Environmental Magazine**, v. 23, p. 14-15, 2012.
- MAHRT, M; SCHARKOW, M. The value of big data in digital media research. **Journal of Broadcasting & Electronic Media**, v. 57, p. 20-33, 2013.
- MARSHAL, C. Big data, the crowd and me. **Information Services & Use**, v. 32, p. 213-224, 2012.
- MATELSKI, J. Big data: challenge or opportunity? **Database Trends & Applications**, v. 26, p. 32, 2012.
- MILSTEIN, J. A.; JHA, A. K. Healthcare's big data challenge. **American Journal of Managed Care**, v. 19, p. 537-538, 2013.
- MITHAS, S.; LEE, M. R.; EARLEY, S.; MURUGESAN, S.; DJAVANSHIR, R. Leveraging big data and business analytics. **IT Professional**, v.15, p. 18-20, 2013.
- MSDN. **Conceitos de mineração de dados (2014)**. Disponível em: <<http://msdn.microsoft.com/pt-br/library/ms174949.aspx>>. Acesso em: 28 jan. 2014.
- MURNANE, L. G. Big data: a big opportunity for librarians. **Onlinemag**, v. 36, p. 30-34, 2012.
- MUTTER, A. D. Big data is a big deal for newspapers. **Editor & Publisher**, v. 145, p. 20-21, 2012.
- NOOR, A. Putting big data to work. **Mechanical Engineering**, v. 135, p. 32-37, 2013.
- PIERCEN, B. Why Big data is not a death sentence for data warehousing. **Database Trends & Applications**, v. 26, p. 28, 2012.
- PREIMESBERGER, C. Big ideas about big data. **eWeek**, v. 28, p. 34-37, 2011.
- RICHARDS, K. Big data analytics: new patterns emerge for security. **Information Security**, v. 15, p. 18-24, 2013.
- ROGERS, S. Big data is scaling BI and analytics. **Information Management Journal**, v. 21, p. 14-18, 2011.
- ROSE, S. Big data and the future. **Significance**, v. 9, p. 47-48, 2012.
- ROYSTER, S. Working with big data. **Occupational Outlook Quarterly**, v. 57, p. 2-10, 2013.
- SAS. **O que é big data? (2013)**. Disponível em: <<http://www.sas.com/offices/latinamerica/brazil/solucoes/bigdata/>>. Acesso em: 26 dez. 2013.

SCHONBERGUER, V. M.; CUKIER, K. **Big d: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**. Rio de Janeiro:

Campus, 2013.

SKINNER, D. Big data: it's not just big. **Scientific Computing**, v. 13, p. 18-19, 2013.

STRACKPOLE, B. Your big data to-do List. **Computerworld**, v. 46, p. 22-25, 2012.

USING big data to assess suicide risk. **KM World**, v. 22, p. 6, 2013.

WAXER, C. Big data blues. **Computerworld**, v. 47, p. 14-20, 2013.

WHITE, M. Big data, big challenges. **EContent**, v. 34, p. 21, 2011.

WIKIPÉDIA. **Big Data (2013)**. Disponível em:

<http://en.wikipedia.org/wiki/Big_data/>. Acesso em: 26 dez. 2013.

ZOTERO. **Zotero.org**. Disponível em:

<<http://www.zotero.org/>>. Acesso em: 26 dez. 2013.