



MATERIAL *Didático*

BIG DATA E GESTÃO DE DADOS



**CRENCIADA JUNTO AO MEC PELA PORTARIA
N 3.455 DO DIA 19/11/2003**

www.facuminasead.com.br

  31 3842-3838

Sumário

EVOLUÇÃO DAS INFORMAÇÕES E DOS DADOS	4
BIG DATA	6
EXEMPLOS DE USO DE BIG DATA.....	10
CARACTERÍSTICAS DO BIG DATA	11
TECNOLOGIAS PARA O BIG DATA.....	14
BIG DATA APLICADO AS EMPRESAS	18
GESTÃO DA INFORMAÇÃO	21
GESTÃO DO CONHECIMENTO	23
GESTÃO DE DADOS.....	25
GERENCIAMENTO DE DADOS	27
REFERENCIAS.....	30

FACUMINAS

A história do Instituto Facuminas, inicia com a realização do sonho de um grupo de empresários, em atender a crescente demanda de alunos para cursos de Graduação e Pós-Graduação. Com isso foi criado a Facuminas, como entidade oferecendo serviços educacionais em nível superior.

A Facuminas tem por objetivo formar diplomados nas diferentes áreas de conhecimento, aptos para a inserção em setores profissionais e para a participação no desenvolvimento da sociedade brasileira, e colaborar na sua formação contínua. Além de promover a divulgação de conhecimentos culturais, científicos e técnicos que constituem patrimônio da humanidade e comunicar o saber através do ensino, de publicação ou outras normas de comunicação.

A nossa missão é oferecer qualidade em conhecimento e cultura de forma confiável e eficiente para que o aluno tenha oportunidade de construir uma base profissional e ética. Dessa forma, conquistando o espaço de uma das instituições modelo no país na oferta de cursos, primando sempre pela inovação tecnológica, excelência no atendimento e valor do serviço oferecido.

EVOLUÇÃO DAS INFORMAÇÕES E DOS DADOS

Desde o início do século XXI ocorreram mudanças significativas no âmbito das Tecnologias da Informação e Comunicação (TIC), neste contexto pode-se citar a computação em nuvem, internet das coisas e as redes sociais. O acesso e o uso destas tecnologias fizeram com que a quantidade de dados aumentasse de uma forma contínua e a uma velocidade sem precedentes (CAI; ZHU, 2015).

As tecnologias vêm revolucionando a maneira como as pessoas lidam e interagem com o mundo a seu redor de forma cada vez mais drástica. Desde o surgimento da Internet até sua popularização, o ser humano depende cada vez mais das máquinas para realizar suas tarefas diárias. Com a criação dos smartphones passa-se a estar conectados à Internet 24 horas por dia, 7 dias por semana. Com a Internet das Coisas, muito em breve, vários objetos de nossas casas e trabalho estarão, também, inteiramente conectados.

Todos esses aparatos rastreiam cada passo do que é feito na Web e cada um desses passos é um potencial gerador de dados. Como passa-se cada vez mais tempo online, gera-se progressivamente mais dados, de forma mais rápida. Qualquer tarefa realizada quando se está conectado, desde enviar um e-mail, escrever uma mensagem em uma rede social ou até mesmo deixar o GPS do celular ligado enquanto se movimenta, deixa rastros imperceptíveis que transformam-se em dados.

Gera-se atualmente aproximadamente 1 zettabyte de dados por dia mas, até então, a imensa maioria desse volume era ignorada. Porém, percebeu-se em determinado momento que esses dados podem ser valiosíssimos quando bem aproveitados, principalmente para as empresas, que podem obter informações em tempo real, geradas diretamente por seus clientes.

Para o governo, significam informações sobre como a sociedade se comporta e formas mais fáceis de monitorar a população. Por outro lado, cria-se uma fina barreira entre os benefícios da exploração de dados e a invasão de privacidade.

Neste mundo globalizado e constantemente conectado, surge então uma nova fonte de poder que não pode mais ser ignorada. Os especialistas a chamam de Big Data

De acordo com Furlan e Laurindo (2017), o crescimento dos dados gerados demandou o desenvolvimento de novas soluções e tecnologias que auxiliassem na sua gestão. E diante desta necessidade surge o big data, propondo novas abordagens para a geração, seleção e manipulação destes grandes volumes dados. O termo big data está relacionado com grandes quantidades de dados, que possuem características distintas, são heterogêneos, providos de diferentes fontes, com controles distribuídos e descentralizados (MCAFEE; BRYNJOLFSSON2012).

Dada a importância das informações disponibilizadas, a qualidade dos dados que geram estas informações tornou-se um dos grandes desafios para as organizações se manterem em um mercado cada vez mais competitivo, sendo que a partir da década de 90 iniciaram-se diversos estudos sobre metodologias e ferramentas para auxiliar no processo de gestão da qualidade dos dados dentro das organizações, e uma das proposições mais relevantes foi o Total Data Quality Management (TDQM), feita por Madnick e Wang em 1992 (ZHU et al., 2012).

O programa desenvolvido pelo Massachusetts Institute of Technology - MIT é baseado na estrutura de Gerenciamento de Qualidade Total (TQM) para melhoria da qualidade no domínio da fabricação, proposto por William Edwards Deming em 1982. Suas pesquisas iniciais desenvolveram um modelo que defende a melhoria contínua da qualidade dos dados, seguindo ciclos de definição, medição, análise e melhoria.

A partir do TDQM, várias outras proposições relacionadas as dimensões e atributos da qualidade dos dados foram feitas, porém, a definição de quais critérios a serem adotados depende do contexto em que os mesmos serão aplicados (BATINI et al., 2009).

Outra questão a se considerar remete à diversidade dos dados disponíveis, uma vez que os mesmos são originados a partir de diferentes fontes, causando uma sobrecarga de informação para a sociedade, gerando inúmeras oportunidades de atuação para os profissionais que atuam na área da gestão da informação (RIBEIRO, 2014).

BIG DATA

Para melhor compreender o significado de Big Data, traduzido literalmente para o português como “Grandes Dados”, é necessário definir o que é Dado. Na Ciência da Informação, a conceituação deste termo está quase sempre ligada aos termos Informação e Conhecimento. Por se tratarem de definições com significados muito similares, não há consonância entre os autores da área para defini-los, considerando o aspecto contextual como essencial para distinção, por vezes tênue, dos termos discutidos. (RUSSO, 2010).

Dado é definido por Setzer (2001, não paginado) como “[...] uma seqüência de símbolos quantificados ou quantificáveis [...] Com essa definição, um dado é necessariamente uma entidade matemática e, desta forma, é puramente sintático”. Já Angeloni (2003, p.18) afirma que “Os dados são elementos brutos, sem significado, desvinculados da realidade.” Davenport (1998, p.18) define dado como uma “Simple observação sobre o estado do mundo”. Sendo ele facilmente estruturado e obtido por máquinas, além de ser quantificado com frequência. O dado é facilmente transferível. Conclui-se à partir das elucidações de Russo (2010, p.15), que “[...] dados são sinais que não foram processados, correlacionados, integrados, avaliados ou interpretados de qualquer forma, e, por sua vez, representam a matéria prima a ser utilizada na produção de informações”.

A informação, segundo Setzer (2001, não paginado) “[...] é uma abstração informal (isto é, não pode ser formalizada através de uma teoria lógica ou matemática), que está na mente de alguém, representando algo significativo para essa pessoa”. Na definição de Angeloni (2003, p.18) “As informações são dados com significado.” A informação é considerada como “[...] dados processados e contextualizados”. Davenport (1998, p.18) diz que informação são “Dados dotados de relevância e propósito”. A informação requer uma unidade de análise e exige consenso com relação ao seu significado e é imprescindível a mediação humana. Para Russo (2010, p.15) “[...] informação pode ser entendida como dados processados e contextualizados”.

Conhecimento é definido por Davenport (1998, p.19 apud RUSSO, 2010, p.17) como “[...] a informação mais valiosa [...] é valiosa precisamente porque alguém deu à informação um contexto, um significado, uma interpretação”. Setzer (2001, não paginado) caracteriza conhecimento como “[...] uma abstração interior, pessoal, de algo que foi experimentado, vivenciado, por alguém”. Angeloni (2003, p.18) define:

O conhecimento pode então ser considerado como a informação processada pelos indivíduos. O valor agregado à informação depende dos conhecimentos anteriores desses indivíduos. Assim sendo, adquirimos conhecimento por meio do uso da informação nas nossas ações. Desta forma, o conhecimento não pode ser desvinculado do indivíduo; ele está estritamente relacionado com a percepção do mesmo, que codifica, decodifica, distorce e usa a informação de acordo com suas características pessoais, ou seja, de acordo com seus modelos mentais. (ANGELONI, 2003, p.18).

O conhecimento de acordo com Davenport (1998, p.18) é uma “Informação valiosa da mente humana. Inclui reflexão, síntese, contexto”. É de difícil estruturação, transferência e captura em máquinas. É frequentemente tácito. O conhecimento tácito é capital dos seres humanos, dificilmente transmitido e capturado, dependendo das experiências e vivências da pessoa (RUSSO, 2010, p.19). Russo (2010, p.18) distingue os termos analisados como apresentado na figura a seguir:

Figura 1: Dado x Informação X Conhecimento

Dado	Informação	Conhecimento
Conjunto de letras, números ou dígitos que não contém significado claro.	Dado trabalhado, útil, tratado, com valor significativo atribuído ou agregado a ele com um sentido natural e lógico.	Informação trabalhada por pessoas e pelos recursos computacionais, possibilitando a geração de cenários, simulações e oportunidades.

Fonte: Russo, 2010, pag. 18.

Analisando as definições acima, percebe-se que o dado, por si só, não possui significado claro e precisa portanto, ser capturado e analisado, para que se transforme em informação. Uma característica importante a ser analisada sobre os dados é que antes do Big Data, os dados coletados pelas empresas provinham de fontes internas e eram, majoritariamente, dados estruturados, ou seja, “[...] dados

formatados em linhas e colunas numéricas organizadas” (DAVENPORT, 2014, p.113). Esse tipo de dado já vem sendo explorado há um tempo pelas organizações, estando presente em bancos de dados, arquivos sequenciais e com relação de importância. (CIO, 2012 apud CANARY, 2013).

Big Data pode ser definida, de maneira mais simplista, como um conjunto de técnicas capazes de se analisar grandes quantidades de dados para a geração de resultados importantes que, em volumes menores, dificilmente seria possível. Em tese, podemos definir o conceito de Big Data como um conjunto de dados extremamente amplos que, por isto, necessitam de ferramentas especiais para comportar o grande volume de dados que são encontrados, extraídos, organizados, transformados em informações que possibilitam uma análise ampla e em tempo hábil.

Big Data é a análise e a interpretação de grandes volumes de dados de grande variedade. Para isso são necessárias soluções específicas para Big Data que permitam a profissionais de TI trabalhar com informações não-estruturadas a uma grande velocidade. As ferramentas de Big Data são de grande importância na definição de estratégias de marketing. Com elas é possível, por exemplo, aumentar a produtividade, reduzir custos e tomar decisões de negócios mais inteligentes.

Alguns DBAs e programadores que conheço migraram seus currículos instantaneamente de “Especialistas em Bancos de Dados” para “Especialistas em Big Data” afinal de contas trabalhavam com bancos de dados de alguns Terabytes (TBs) e teoricamente isso é Big Data. Na verdade não é simplesmente o volume de dados que define o Big Data, o conceito é calcado em um pilar de 5 itens.

Para Erl,Khattak e Buhler (2016), big data é um campo que se dedica à análise, ao processamento e armazenamento de grandes datasets, e que as suas soluções e práticas são geralmente necessárias quando as tecnologias e técnicas tradicionais não são suficientes para a execução destas atividades.

Para os autores, big data não é apenas uma tecnologia, é também sobre como as tecnologias podem impulsionar uma organização cada vez mais as organizações estão gerando enormes quantidades de dados que são provenientes de fontes distintas e armazenados de diferentes maneiras, o que demanda um

processo de gestão específico para garantir a sua qualidade. Vianna, Dutra e Frazzon (2016, p.193), enfatizam a importância de se fazer uma gestão efetiva e a [...] “necessidade de transformar esses dados em informações de qualidade, que possam ser utilizadas para direcionar os negócios e as estratégias das organizações, minimizando riscos, e apoiando o processo de tomada de decisões”.

Quando falamos de Big Data, temos basicamente dois tipos de estruturação da informação: os dados estruturados e não estruturados. Vamos entender um pouco melhor sobre cada um:

Dados estruturados

Os dados estruturados são aqueles que possuem uma estrutura determinada, com categorias, clusters e definições, como localização, vendas e informações sobre o perfil de clientes, contatos entre outros.

Dados estruturados são encontrados nos bancos de dados que, para armazenar qualquer dado, precisam ter muito bem definidos onde cada informação estará. Softwares de empresas como ERP, CRM, sistemas financeiros, sistemas de RH entre outros possuem dados estruturados.

Dados não estruturados

Os dados não estruturados são os mais complexos de se trabalhar, pois não existe neles uma estruturação sequer, sendo necessária intervenção humana para sua preparação. Estamos falando dos dados de mídias sociais, como YouTube, Facebook, Instagram, portais de notícias etc. Todas essas mídias que citamos lidam com dados em vídeos, imagens, textos e até mesmo áudio, por isso a complexidade para o tratamento desses dados é maior.

As soluções de Big Data são feitas para lidar com um grande volume de dados não-estruturados. Isso significa que eles não têm relação entre si e nem uma estrutura definida. São, por exemplo, posts no Facebook, vídeos, fotos, tweets, geolocalização, comportamento.

Teoricamente, os dados não-estruturados só poderiam ser analisados por humanos. Por outro lado, os dados estruturados podem facilmente ser colocados em uma tabela do Excel, por exemplo. Eles possuem regularidade entre si.

Ferramentas "comuns", feitas para analisar dados estruturados, como os preços dos supermercados de uma região, não são capazes de analisar dados não-estruturados. As ferramentas de Big Data não devem só dar conta da grande quantidade de dados variáveis, mas devem fazer isso a uma grande velocidade.

As ferramentas de armazenamento de Big Data também são diferentes das usadas para armazenar dados comuns. Ao contrário dos bancos de dados comuns, os bancos usados no Big Data devem ter elasticidade, pois precisam suportar não só grandes volumes, mas grandes volumes que crescem muito em pouco tempo. Eles também precisam ser flexíveis para aceitar vários tipos de mídias.

CURIOSIDADES: QUAL O VOLUME DE DADOS GERADO DIARIAMENTE?

- O Facebook armazena, acessa e analisa mais de 50 petabytes de informações geradas pelos usuários, a cada mês são gerados mais de 700 milhões de minutos por mês.
- A cada minuto são feitos uploads de 48 horas de vídeos no Youtube, ou seja, nunca ninguém conseguirá assistir todos os vídeos do Youtube.
- Diariamente mais de 500 milhões de mensagens são enviadas pelo Twitter, com uma média de 5700 TPS (Twittes per Second ou Mensagens por Segundo), o recorde é de 143.199 TPS.
- O Google processa diariamente mais de 3 bilhões de pesquisas em todo o mundo, sendo desse total 15% totalmente inéditas. Seu “motor” de pesquisa rastreia 20 bilhões de sites diariamente, armazenando 100 petabytes de informação. Sem contar todas as informações que as companhias geram diariamente, sejam elas estruturadas ou não.

EXEMPLOS DE USO DE BIG DATA

As empresas usam o Big Data para realizar os desejos dos clientes antes que eles peçam. Um exemplo disso são as sugestões de sites de compras, como a Amazon, e as recomendações de serviços como Netflix e Spotify.

Big Data também pode ser usado para satisfazer os clientes de outras maneiras como, por exemplo, dando mais conforto a ele em situações incômodas, mas que não são desagradáveis o suficiente para chegar a um canal de reclamações da empresa. A companhia aérea Delta identificou que uma das maiores preocupações de seus clientes durante as viagens era com a bagagem. Com base nessa informação, criaram um recurso no aplicativo da companhia que ajuda os viajantes a rastrear suas malas, o Track My Bag.

A Microsoft anunciou uma parceria com a Universidade Federal de Minas Gerais para realizar pesquisas que podem ajudar a prever engarrafamentos com até uma hora de antecedência. O Projeto de Previsão de Trânsito deve usar dados como números históricos do departamento de transportes, câmeras nas ruas, mapas de trânsito do Bing e posts nas redes sociais.

A análise de Big Data também pode ser usada para dar suporte a ações do governo. Durante a epidemia de gripe suína, uma equipe da Telefônica usou dados da rede de celulares para entender como as pessoas estavam se locomovendo durante o surto da doença. Suas descobertas ajudaram a validar ações do governo para combater a crise.

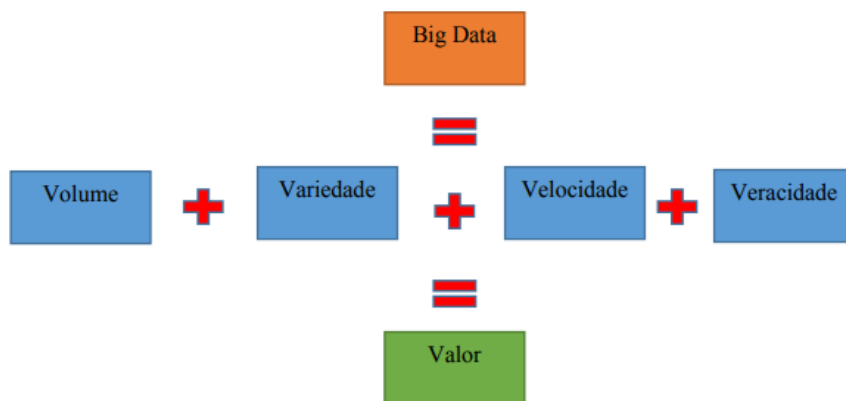
CARACTERÍSTICAS DO BIG DATA

Existem na literatura, diferentes pontos de vista em relação as características que compõem um ambiente big data. Três delas foram inicialmente identificadas por Doug Laney no início de 2001, quando o autor publicou um artigo descrevendo o impacto do volume, da velocidade e da variedade dos dados em data warehouses corporativos (LANEY, 2001).

Anos mais tarde novas características foram incorporadas ao conjunto de aspectos relacionados a este conceito, são elas: veracidade, valor, variabilidade (ZIKOPOULOS et al., 2012; GANDOMI; HAIDER, 2015). Buscando auxiliar em um melhor entendimento acerca das características dos ambientes big data, a seguir será apresentado o conceito de cada uma delas.

O V de Veracidade é sobre as fontes e a qualidade dos dados, pois eles devem ser confiáveis. Já o V de Valor é relacionado aos benefícios que as soluções de Big Data vão trazer para uma empresa. Cada instituição precisa checar se os benefícios trazidos pela análise de Big Data compensam o alto investimento nas soluções específicas para isso.

Figura 2: Os 5V's do Big Data



Volume – Volume dos dados:

Passamos a falar muito rápido de Gigabytes para Terabytes e agora estamos falando de Petabytes e outros volumes. Hoje são contabilizados em média 12 Terabytes de Tweets diariamente, em 2012 foram gerados cerca de 2.834 Exabytes (que são milhões de Gigabytes) a previsão é que em 2020 se gerem anualmente 40.026 Exabytes de informações. Refere-se a grandes quantidades de dados e informações que são geradas a partir de fontes variadas. De acordo com Erl, Khattak e Buhler (2016), o volume de dados que é processado pelas soluções big data é substancial e cada vez maior e impõem exigências distintas de armazenamento e processamento, apresentando um grande desafio às estruturas de TI convencionais,

pois grandes volumes de dados requerem armazenamento escalonável e uma abordagem distribuída para possibilitar a consulta aos mesmos.

Velocity – Velocidade:

Hoje para alguns negócios, 1 minuto pode ser muito tempo, detecção de fraudes, liberações de pagamentos, análises de dados médicos ou qualquer outra informação sensível a tempo. A maior parte dos projetos de DW/BI (Data Warehouse e Business Intelligence) ainda tem latência em D-1, ou seja, carregamos o dia anterior. Ainda acreditamos que essa solução se aplique a muitos negócios, porém, para algumas análises, quanto mais próximo do tempo real, maior pode ser o incremento de negócio.

Possui também relação com o tempo de resposta para determinada requisição. Com as comunicações em tempo real, cada vez mais vem sendo possível atingir uma maior velocidade para troca de dados e informação (RIBEIRO, 2014). McAfee e Brynjolfsson (2012) afirmam que para muitas aplicações, a velocidade de criação de dados é ainda mais importante do que o volume dos mesmos, uma vez que as informações em tempo real ou quase em tempo real, tornam possível que uma empresa seja mais ágil que seus concorrentes.

Variety – Variedade:

Big Data também poderia ser considerado como Any Data (qualquer dado), hoje temos capacidade de capturar e analisar dados estruturados e não estruturados, texto, sensores, navegação Web, áudio, vídeo, arquivos de logs, catracas, centrais de ar condicionado, entre outros. Uma nova série de ferramentas está se tornando parte dos projetos, são as ferramentas de machine data, dados de máquina, quase qualquer aparelho eletrônico hoje em dia tem uma estrutura de dados ou programação, o ponto é que nem todos podem ser acessados (ainda..).

Ela também diz respeito a diversidade dos dados e informações. Kaisler et al. (2013), afirma que a partir de uma perspectiva analítica, fazer a gestão da variedade de dados é provavelmente o maior obstáculo para utilização efetiva de grandes volumes de dados. Formatos de dados incompatíveis, estruturas de dados não

alinhadas e semânticas inconsistentes representam desafios significativos que se não vencidos podem levar ao insucesso de um projeto.

Valor:

Está relacionado com o retorno do investimento e é o resultado da combinação dos aspectos citados anteriormente. Esta característica está intuitivamente relacionada com a característica de veracidade, pois quanto maior a qualidade dos dados, mais valor ela possui para o negócio (KAISLER et al., 2013). Da mesma forma, pode-se considerar que valor e velocidade possuem relações inversas, pois quanto mais tempo se leva para que os dados sejam transformados em informações relevantes, menos valor terá para o negócio, uma vez que resultados obsoletos prejudicam a qualidade e a rapidez na tomada de decisão. •

Variabilidade

Refere-se aos dados que estão em constante variação, como por exemplo, dados meteorológicos. Gandomi e Haider (2015), afirmam que a variabilidade possui relação com a variação nos fluxos de dados.

TECNOLOGIAS PARA O BIG DATA

Como previamente discutido, ainda existem divergências sobre o termo Big Data, apesar de sua crescente popularidade. Sendo assim, tornar este conceito aplicável, ou seja, transformar os dados e “[...]criar soluções de negócios que agreguem valor para as companhias” (TAURION, 2013a, não paginado) ainda apresenta-se como uma tarefa complexa e controversa.

Criam-se então, novas tecnologias capazes de administrar o imenso volume de dados gerados pelo Big Data, que excedem a capacidade das antigas soluções, preocupadas em analisar dados estruturados gerados pelos sistemas internos das empresas. Uma tecnologia, para decifrar o fluxo veloz do Big Data, deve ser capaz de criar rapidamente modelos estatísticos, que se encaixem aos dados, otimizando-

os e prevendo-os. Para tal, são interessantes tecnologias com aprendizado de máquina, que são flexíveis e adaptáveis. (DAVENPORT, 2014, p.111).

O aprendizado de máquina, também chamado de modelagem automatizada, significa que o software consegue adaptar diferentes modelos de dados a fim de obter o melhor ajuste possível. Esse processo é muito rápido, porém pode gerar resultados difíceis de interpretar e explicar. (DAVENPORT, 2014, p.116).

Apesar do Big Data usualmente gerar uma discussão em torno do grande volume de dados não-estruturados, Davenport (2014, p.113) acredita que a grande novidade relacionada à nova tecnologia é como ela irá gerar valor às organizações, seja reduzindo os custos, aumentando a velocidade de processamento de dados, desenvolvendo novos produtos ou serviços e proporcionando um melhor suporte ao processo decisório, informando novos dados e modelos.

A grande vantagem das tecnologias de Big Data é que a maioria delas é gratuita ou barata, não sendo necessário grandes investimentos por parte das empresas neste sentido. No entanto, para arquitetar e programar esses softwares faz-se necessário uma equipe extremamente especializada e bem treinada, o que demanda tempo e um custo elevado. (DAVENPORT, 2014, p.115).

A Figura 3, a seguir, apresenta uma visão geral das tecnologias do Big Data, segundo Davenport (2014, p.112).

Figura 3: Visão geral das tecnologias do Big Data

Tecnologia	Definição
Hadoop	Software de código aberto para o processamento de big data em uma série de servidores paralelos
MapReduce	Um framework arquitetônico no qual o Hadoop se baseia
Linguagens de script	Linguagens de programação adequadas ao big data (por exemplo, Python, Pig Hive)
Aprendizado de máquina	Software para identificar rapidamente o modelo mais adequado ao conjunto de dados
Visual Analytics	Apresentação dos resultados analíticos em formatos visuais ou gráficos
Processamento de Linguagem Natural (PLN)	Software para análise de texto – frequências, sentido, etc.
In-memory analytics	Processamento de big data na memória do computador para obter maior velocidade

Fonte: Davenport, 2014, Pag. 112.

As tecnologias de Big Data são divididas sob duas óticas:

- a) As tecnologias envolvidas com análise de dados, cujos principais representantes são o Hadoop MapReduce (HMR);
- b) As tecnologias de infraestrutura, responsáveis pelo armazenamento e processamento dos dados. São os bancos de dados Not Only Structured Query Language (NoSQL).

Alguns exemplos de softwares banco de dados NoSQL que tratam dados estruturados e não-estruturados são o Big Table, usado pela Google, o DynamoDB, usado pela Amazon, o Neo4j, entre outros.

O Hadoop é uma das principais tecnologias relacionadas ao Big Data. Criado em 2005 pelo Yahoo, o Hadoop é um projeto da comunidade Apache, dividido em duas partes: a primeira, o HMR é baseado no MapReduce, software utilizado pelo Google para acelerar as pesquisas realizadas em seu buscador, a segunda é o Hadoop Distributed File System (HDFS), que é “[...] um sistema de arquivos distribuídos otimizados para atuar em dados nãoestruturados” (TAURION, 2013a, não paginado). Na prática, os dados precisam primeiramente estar armazenados no HDFS para então serem processados pelo HMR.

Além destas duas tecnologias, existe também o Hadoop Common que é um “[...] conjunto de bibliotecas e utilitários que suportam os projetos Hadoop” (TAURION, 2013a, não paginado). Davenport (2014, p.113) define o Hadoop como “[...] um conjunto de ferramentas de software de código aberto e um framework para distribuir os dados em vários computadores; trata-se de um ambiente unificado de armazenamento e processamento altamente escalonável para grandes e complexos volumes de dados”.

Outra tecnologia de extrema importância para o Big Data é o Stream Process. Sua principal diferença com relação ao Hadoop é que este último analisa os dados que são primeiramente armazenados em seu sistema, enquanto o Stream Process faz a análise de dados em tempo real. A vantagem desta tecnologia é que a análise dos dados no momento que os mesmos estão sendo criados permite a realização de ações imediatas em resposta às questões levantadas. (TAURION, 2013a).

O grande desafio para os gestores que pretendem se arriscar a trabalhar com Big Data é escolher a opção de banco de dados mais adequada, dentre tantas e, especialmente, qual o melhor componente de análise de dados a ser utilizado, sendo este último o principal responsável por transformar estes dados em valor para as empresas. Outros aspectos a serem considerados, quando se analisa as tecnologias de Big Data são o conceito de stream process, que permite o tratamento de dados em tempo real e as tecnologias que permitem a visualização destes dados. Alguns exemplos são as nuvens de tags, clustergramas, History Flow e Spatial Information Flow

A utilização da computação em nuvem é outro fator tecnológico interessante, tendo em vista sua capacidade de suportar imensos volumes, sua elasticidade e menor custo se comparado ao armazenamento em servidores físicos.

Para obter sucesso tirando proveito do Big Data, faz-se essencial para as empresas começarem a investir em tecnologias que o comportem, tendo sempre em vista que um dos atributos essenciais quando se fala deste novo conceito é a resiliência. O volume de dados, sua natureza, a velocidade com que são criados novos dados, está em constante mudança e somente as organizações que

souberem se adaptar prontamente e tirar proveito dessas variações, é que serão bem sucedidas. (TAURION, 2013a).

BIG DATA APLICADO AS EMPRESAS

O crescimento exponencial do volume de informações gerado é visível. Ignorar os dados disponíveis é um erro que pode levar o negócio ao fracasso. Afinal, é por meio deles que você pode prever tendências e mudanças no mercado de atuação.

Por isso, mais que ter acesso a uma grande quantidade de dados, é necessário tratá-los de maneira correta. O ideal é fazer uma análise aprofundada para encontrar respostas certas, que visem a quatro principais objetivos:

- Redução de custos;
- Economia de tempo;
- Desenvolvimento de produtos e otimização de ofertas;
- Otimização das tomadas de decisão.

A ideia é que essa ferramenta ajude a:

- Identificar falhas, defeitos e problemas praticamente em tempo real;
- Criar promoções a partir dos hábitos de compra do cliente;
- Detectar comportamentos fraudulentos antes de prejudicarem o negócio;
- Recalcular carteiras de risco.

Com essa explicação, fica claro que o propósito do Big Data é fazer a análise de dados — e é aí que está a inteligência do processo. Sem uma avaliação metódica e precisa, é impossível saber qual é o melhor caminho. Mas o que efetivamente significa a análise de dados? A ideia é encontrar padrões comportamentais para monitorá-los e, sempre que houver alguma alteração, tomar a decisão certa. É o caso de sua empresa vender o produto X, mas identificar que o

número de comercializações está em queda porque o item Y é similar, porém mais resistente. Nesse caso, antes de perder fatia de mercado, você pode substituir a mercadoria para evitar outros prejuízos.

O trabalho que permite cruzar os dados e, a partir disso, interpretá-los é o Big Data Analytics. Essa ferramenta permite extraí-los, organizá-los, tratá-los e entendê-los para, então, transformá-los em informações úteis para o negócio.

Vários são os exemplos que demonstram o uso dos dados no seu negócio. Primeiramente, você precisa entender que há três tipos:

- **Social data:** são aqueles derivados das pessoas que ajudam a identificar comportamentos para atuar de forma direcionada. Por exemplo: pesquisas feitas no Google e comentários nas redes sociais;
- **Enterprise data:** são gerados por organizações, como nos setores de RH, financeiro, produção etc. Contribuem para mensurar a produtividade e identificar gargalos;
- **Personal data ou data of things:** são originários de dispositivos conectados à internet, como TVs, carros e geladeiras. Tem como base a Internet das Coisas (IoT, ou Internet of Things). Um exemplo seria usar as informações de um aplicativo de trânsito para alimentar painéis eletrônicos na cidade a fim de informar sobre o tráfego.

Diante desse contexto, a aplicação do Big Data depende da realidade da sua empresa. É fundamental contar com a ajuda da tecnologia para coletar e filtrar os dados, com o propósito de utilizá-los conforme as suas necessidades. Por exemplo: se você tem um site ou aplicativo de saúde e bem-estar, pode cruzar os dados dos usuários — como localização, idade, peso, altura, sexo etc. — para verificar qual doença tem ocorrência mais provável em determinada região. Assim, pode traçar estratégias específicas, caso dos artigos sobre esse assunto para o blog.

Já uma empresa fabricante de aviões tem a possibilidade usar sensores para identificar quando é necessário que a aeronave faça alguma parada para reabastecimento ou se é possível chegar ao destino sem interromper a viagem.

Veja, a seguir, outros exemplos de uso do Big Data em diferentes setores, com cases de sucesso.

Marketing

Essa área depende da identificação do comportamento do consumidor para criar ações específicas e encontrar o cliente certo para cada produto. A análise depende de diferentes fatores, como sazonalidade, canal de vendas, abordagens de marketing (características, preço etc.).

Em uma confecção, você pode usar o Big Data para descobrir se há uma preferência nacional ou regional por suéteres de lã ou blusas de algodão. Da mesma forma, a visualização desses dados podem indicar oportunidades.

Um exemplo claro dessa ação foi o case do McDonalds. Diariamente, a empresa coleta e combina os dados de seus maiores restaurantes no mundo para padronizá-los e entender a reação dos públicos. Com isso, já foram criados sanduíches com base nas análises de sentimento feitos em redes sociais, lançadas promoções em tempo real e modificada a logística do drive thru a partir da reação dos consumidores de cada país.

Controle de qualidade

As variáveis abrangidas nesse critério são muitas, como defeitos por unidade, rendimentos de primeira passagem, taxas de preenchimento e outras. Elas podem até incluir os fornecedores, a capacitação dos colaboradores e as condições ambientais no momento da produção.

Todos esses itens podem ser analisados com o Big Data. Os dados de produção, inclusive, são passíveis de integração a outros, o que facilita a identificação e a eliminação de gargalos e fatores de lentidão.

Um exemplo de controle da qualidade pode ser verificar com a FarmLogs. A empresa de gerenciamento de produções agrícolas faz a mineração de dados para prever cenários de colheitas futuras, estado da vegetação e condições de plantio. Assim, os fazendeiros conseguem aumentar significativamente sua produtividade.

Financeiro

Esse setor está relacionado a todos os outros da empresa e a visualização de seus números é fundamental para verificar a saúde do negócio. É o caso do setor de vendas, que geralmente desconhece o custo de produção para cada um dos itens. Ou do marketing, que nem sempre sabe qual mercadoria tem menor margem de lucro para empregar esforços naquelas que trazem mais lucratividade.

Por meio de algoritmos complexos, o banco dos Estados Unidos JP Morgan, por exemplo, prevê tendências para indicar aos seus investidores qual é o melhor momento para comprar e vender ações. A mesma lógica pode ser aplicada internamente para a empresa, com a análise de dados de outros setores e comparação com os índices financeiros.

GESTÃO DA INFORMAÇÃO

Esta fundamentação conceitual discorre diferentes aspectos sobre a gestão da informação com o propósito de serem referenciais para a análise dos dados de pesquisa. A literatura reporta variadas formas de gerenciar a informação em uma organização. McGee e Prusak (1994, p. 114) formulam um modelo de processo de gerenciamento de informação com foco no valor estratégico da informação, definido em três tarefas:

- I. Identificação de necessidades e requisitos de informação;
- II. Classificação e armazenamento de informação/tratamento e apresentação da informação; e
- III. Desenvolvimento de produtos e serviços de informação.

Davenport (2000) também relata que o fenômeno da gestão da informação ocorre de forma processual, estabelecendo o Modelo Ecológico para o gerenciamento da informação. Este modelo está sustentado no modo holístico de pensar e em quatro atributos-chave:

- I. Integração dos diversos tipos de informação;

- II. Reconhecimento de mudanças evolutivas;
- III. Ênfase na observação e na descrição; e
- IV. Ênfase no comportamento pessoal e informacional.

Envolve três ambientes:

O ambiente informacional que é o centro da abordagem ecológica e abrange seis componentes: estratégia da informação, política da informação, cultura e comportamento em relação à informação, equipe da informação, processos de administração informacional e arquitetura da informação;

O ambiente organizacional, que abrange todas as atividades organizacionais, sendo composto: pela situação dos negócios, investimentos em tecnologia e distribuição física;

O ambiente externo, que consiste em informações sobre três tópicos fundamentais: mercados de negócios, mercados tecnológicos e mercados da informação.

Em harmonia com a ecologia informacional, Davenport (2000, p. 173) define o gerenciamento da informação como processos, isto é, “um conjunto estruturado de atividades que incluem o modo como as empresas obtêm, distribuem e usam a informação e o conhecimento”. Assim, estabelece quatro passos a seguir para a realização do processo de gerenciamento da informação.

Primeiro, determinação das exigências, uma vez que é importante compreender o contexto nas quais as atividades gerenciais são realizadas, identificar quais as fontes necessárias e as informações para a gerência. Por meio dessa compreensão é possível realizar o mapeamento da informação disponível na organização, registrar os recursos informacionais existentes, as unidades responsáveis e os serviços e sistemas disponíveis.

Segundo, obtenção de informação, que se constitui em uma atividade contínua, a qual acompanha o desenvolvimento organizacional. O fornecimento dessa informação precisa proporcionar aos usuários as informações necessárias em produtos e serviços informacionais. Essa atividade é composta de diversas tarefas, não necessariamente sequenciais, a saber: exploração do ambiente informacional,

classificação da informação em uma categoria, a formatação e a estruturação da informação.

Terceiro, distribuição, que se refere à disseminação da informação aos gerentes e funcionários que delas necessitam. No entanto, é necessário definir as estratégias dessa distribuição, podendo-se optar tanto pela divulgação às pessoas autorizadas como pela disponibilização a todos.

E quarto, utilização da informação, que pode ser considerada a etapa mais importante desse processo, uma vez que todos os esforços das demais etapas convergem e se justificam para proporcionar o uso da informação em seu contexto organizacional.

De acordo com Costa (2003), a informação é concebida como matéria-prima para gerar o conhecimento. A literatura sobre gestão do conhecimento coloca o conhecimento tácito e as informações de caráter informal como vitais para a sobrevivência em mercados cada vez mais competitivos. Nessa análise, a gestão da informação se expande para gestão do conhecimento e os sistemas são requisitados para processar tanto as informações informais como os produtos das atividades intelectuais. Tais sistemas necessitam abranger informações externas e internas, coletadas sistematicamente, analisadas e disseminadas para toda a organização, com a missão de transformar informações em conhecimento estratégico.

GESTÃO DO CONHECIMENTO

Em contra partida, a gestão do conhecimento trabalha tanto com o conhecimento explícito quanto com o conhecimento tácito, que é “o acúmulo de saber prático sobre um determinado assunto, que agrega convicções, crenças, sentimentos, emoções e outros fatores ligados à experiência e à personalidade de quem detém.” (MIRANDA, 1999 apud VALENTIM, 2002). O conhecimento tácito é o conhecimento que cada um tem como resultado de seu aprendizado prévio, tornando-o difícil de ser repassado.

“O conceito da gestão do conhecimento surgiu nos anos 90, não somente associado ao processo operacional, mas principalmente à estratégia organizacional” (SANTOS, 2001; SVEIBY, 1998; apud FELIX, 2003). E vem sendo discutido cada vez mais como uma tendência para o mercado econômico no presente e no futuro. Principalmente nos tempos atuais em que as organizações “com um considerável grau de flexibilidade e de capacidade inovadora, não se preocupam mais em somente armazenar dados e informação”. (BERBE, 2005) Ainda de acordo com Berbe, 2005, “a competitividade, a informação e a necessidade de diferencial competitivo fizeram com que as empresas percebessem a importância do conhecimento e da sua gestão”.

E apesar da compreensão por parte das empresas da extrema importância na gestão do conhecimento nas organizações, ainda encontram grandes dificuldades na sua aplicação. Como afirma Wilson, (2006, p. 45), quando diz que “o uso do termo gestão de conhecimento é, na maioria dos casos, sinônimo de gestão da informação”, corroborando as barreiras encontradas para sua aplicação.

“A gestão do conhecimento, de forma abrangente, refere-se ao planejamento e controle de ações (políticas, mecanismos, ferramentas, estratégias e outros) que governam o fluxo do conhecimento, em sua vertente explícita – e para isso englobam práticas da gestão da informação – e sua vertente tácita.” (LEITE; COSTA, 2007)

Deste modo podemos dizer que “a gestão do conhecimento é um processo complexo e intimamente relacionado com processo de comunicação nas organizações.” (SMOLIAR, 2003; IVES et al., 1998; THEUNISSEN, 2004; apud LEITE; COSTA, 2007). Que de acordo com Dawson, (2000, apud BERBE, 2005), trata-se “de todos os aspectos relacionados com a forma como as pessoas desempenham funções baseadas em conhecimento”.

Berbe, (2005), afirma que:

“empresas mais inovadoras, que se voltam para a Gestão do Conhecimento, necessitam de uma abordagem que veja a organização como uma comunidade humana, cujo conhecimento coletivo representa um diferencial competitivo em relação aos seus mais diretos concorrentes. O conhecimento coletivo é valorizado, criando-se redes informais de pessoas que realizam trabalhos diversos com pessoas que eventualmente estão dispersas em diferentes unidades de negócio.”

Sendo assim, a gestão do conhecimento utiliza ferramentas em busca da inovação com o objetivo de fazer as organizações se desenvolverem. É o que explica Willian James (1907, apud NONAKA; TAKEUCHI, 1997, p. 31), quando diz que “ideias não tem valor exceto quando passam para as ações que rearrumam e reconstroem de alguma forma, em menor ou maior medida, o mundo no qual vivemos.” E essas ações serão determinadas pela gestão do conhecimento em concordância com as necessidades da instituição que atua.

Wilson, (2003, p. 54), define gestão do conhecimento como habilidades de competência informacional, ou seja, aprender a aprender, quando afirma que o gestor não tem controle sobre o conhecimento. Partindo do pressuposto que o conhecimento tácito “é muito pessoal e difícil de ser codificado, ou seja, expresso por palavras” (NONAKA; TAKEUCHI, 1997). Portanto, o gestor só pode auxiliar no uso da informação, sua organização e disponibilização, e esperando que o usuário compreenda e a partir desse entendimento possa resultar em conhecimento.

Pelo ponto de vista, Wilson (2003), está correto na sua afirmação, pois o conhecimento é intangível. Mas quando aplicamos gestão do conhecimento estaremos direcionando-a de acordo com um objetivo. E quando direcionamos as informações para o ponto certo (equipe ou as pessoas certas), pode resultar em conhecimento.

Basta pensarmos no papel dos educadores universitários. Direcionam os alunos quando fornecem informações que irão resultar em novo conhecimento e, consequentemente, será expresso nos trabalhos acadêmicos de conclusão de curso. Portanto, podemos entender a gestão da informação como a identificação e organização das informações existentes na instituição e a gestão do conhecimento como utilização dessas informações junto à percepção dos usuários.

GESTÃO DE DADOS

A gestão de dados tem o objetivo de gerenciar e zelar pelos dados das empresas, tratando-os como um recurso valioso, de modo que as informações possam ser transformadas em valor empresarial e embasar decisões estratégicas. E, para isso, a gestão de dados utiliza processos, profissionais, metodologias e ferramentas.

Atualmente, no contexto da Transformação Digital, um gerenciamento inteligente das informações se torna ainda mais importante. À medida que as organizações são cada vez mais orientadas por processos digitais, a quantidade de dados cresce exponencialmente. Nunca foi tão fácil captar informações de mercado, mas, por outro lado, nunca foi tão crucial protegê-las. Hoje, a segurança dos dados é uma prioridade.

Além disso, empresas que falham em organizar suas grandes quantidades de informações acabam gastando desnecessariamente com armazenagem. Também tendem a arcar com gastos extras com compliance e recursos humanos — o tempo gasto para procurar informações, gerenciar processos e cumprir tarefas aumenta de forma drástica. Entretanto, mais do que tudo isso, a gestão de dados é um alicerce da Transformação Digital. Como sabemos, ela vai muito além da tecnologia: ela está diretamente ligada à cultura organizacional e às operações.

A gestão de dados eficiente gera vantagem competitiva para as empresas. Apesar de a maioria dos benefícios serem diferentes em cada empresa, alguns ainda são comuns à maioria. Entre eles podemos destacar:

- Melhor alinhamento entre as áreas de tecnologia e de negócio.
- Conhecimento dos dados utilizados na empresa através da adoção de um vocabulário único sobre as definições dos dados que circulam na empresa.
- Entendimento das principais necessidades de dados e informações da empresa, fornecendo um importante subsídio para estabelecer o planejamento para absorção, criação e/ou transformação de novos dados e informações.

- Melhoria na qualidade e confiabilidade dos dados e informações através do uso de dados cada vez mais claros, precisos, íntegros, integrados, pertinentes e oportunos.
- Criação da cultura do uso de indicadores de processo e qualidade dos dados.
- Reutilização de dados considerados corporativos, contribuindo dessa forma para a melhoria da qualidade dos dados e também reduzindo os esforços, tempos e custos do desenvolvimento de novas aplicações.
- Redução dos riscos e falhas no desenvolvimento dos sistemas e aplicações.
- Eliminação ou redução drástica na quantidade de informações redundantes, contribuindo para reduzir os esforços em manter íntegras as informações que antes eram redundantes.
- Estabelecimento de mecanismos formais de segurança, acesso e disponibilização de dados e informações a quem realmente necessita.
- Aumento da produtividade das pessoas que utilizam os dados e as informações.

Sem uma gestão efetiva dos dados, a evolução desta cadeia não é atingida, portanto, para atingir os objetivos é fundamental a disciplina atuar nos estágios iniciais da cadeia. Por esta razão o nome da disciplina é Gestão de Dados e não Gestão das Informações ou Gestão do Conhecimento. Porém, valem ressaltar que, dependendo do nível da maturidade da empresa, as ações de gestão para a evolução da cadeia podem se estabelecer em outros níveis.

GERENCIAMENTO DE DADOS

A análise de dados tem sido cada vez mais importante para o sucesso, sendo você um pequeno empresário ou uma grande indústria. Para que a sua empresa se destaque e deixe a concorrência para trás, é necessário reunir e analisar informações sobre clientes e o mercado como um todo. Sendo assim, em diversos

aspectos, fazer o gerenciamento de dados de forma inteligente é absolutamente essencial para o sucesso dos negócios e para atingir um nível sustentável de crescimento.

Se puder resumir em uma palavra o que é gerenciamento de dados, podemos definir como Organização. Esse processo é baseado em coletar, validar, armazenar e garantir a segurança dos dados para poderem ser transformados, de fato, em informações úteis. Decisões importantes fundamentadas em dados crus podem ser muito perigoso, por isso deve-se ir além e analisar cada detalhe da base de dados para qualificar o processo de tomada de decisões, que não deve ser realizado de maneira precipitada.

Quando todas as informações das quais você precisa estão bem organizadas e devidamente armazenadas, de forma integrada e ordenada, arquivos como projetos, balanços financeiros, prospecções de vendas e outros documentos importantes são encontrados com maior facilidade no sistema toda vez que você precisar analisá-los.

A segurança de dados é um ponto de extrema importância para qualquer organização. Quando a proteção das informações é abalada, seja no caso de um vazamento, seja diante de um roubo de dados, o próprio futuro do negócio fica comprometido.

É inevitável que, para fazer bons negócios e sustentar o crescimento, a empresa tenha a internet como uma ferramenta de trabalho no dia a dia. Entretanto, quem está sempre conectado também está em constante risco de perder seus dados, por diversos motivos.

A seguir, elencamos algumas das principais ameaças à segurança da informação, passíveis de ocorrerem quando não há um adequado gerenciamento de dados. Confira:

1. **Ciber ataques** :Infelizmente, toda e qualquer empresa está sujeita aos ataques de pessoas de má-fé e cibercriminosos em geral. Entre outras possibilidades, eles podem invadir a rede da sua organização, prejudicando a integridade dos dados. Outra ação maléfica causada

por esses criminosos é o envio de vírus, que podem danificar por completo os seus softwares.

2. **Configurações de segurança desatualizadas:** Existem diversas soluções em sistemas de dados que possibilitam a personalização de medidas de segurança e restrição de acessos por nível de usuário, a fim de proteger sua estrutura de dados e permitir que só usuários autorizados acessem determinadas pastas e arquivos importantes. Porém, infelizmente muitas empresas simplesmente não se aproveitam dessa funcionalidade, ficando mais vulneráveis a problemas e falhas de segurança.
3. **Desastres:** Em 2017, em um dos mais graves atentados cibernéticos mundiais, o ransomware WannaCry fez milhares de vítimas em mais de 150 países, entre pequenos negócios a grandes corporações. O ransomware é um tipo de vírus que sequestra os dados de um computador ou notebook e não libera o acesso a eles até que um resgate seja pago.

Nesse sentido, se uma empresa não conta com um programa de recuperação de desastres, por meio do qual ela se prepara para catástrofes como desastres naturais, falha humana (acidental ou não), ataques cibernéticos, quebra de equipamentos, entre outros, os dados gerados no decorrer das atividades são perdidos.

Para auxiliar você a garantir a segurança das informações de sua empresa, é necessário ações que para garantir a eficiência do gerenciamento de dados. Como backups, diversas empresas adotam uma periodicidade mensal (ou maior) para seus backups. Entretanto, considere a imensa quantidade de dados que são gerados e processados durante esse tempo: o mais adequado é realizar backups diários, ao final de cada jornada de trabalho. Isso pode ser feito de forma programada, sem que você precise executar o processo manualmente todos os dias.

O uso da cloud computing tem se tornado uma tendência cada vez mais presente no mercado. A praticidade e o dinamismo que essa tecnologia oferece vem

beneficiando empresas de todos os portes e segmentos. De fato, é compreensível que muitas empresas julguem necessário ter alguma forma de backup físico (como alternativa e uma segunda opção de garantia), porém, além de dispensar o uso de aparelhos físicos (como HDs), o backup na nuvem realmente garante a preservação e integridade dos dados.

Outro elemento essencial que é proporcionado pelo gerenciamento de dados na nuvem é a disponibilização de modo mais fácil, dinâmico e rápido das informações, inclusive, para um número mais amplo de pessoas se necessário. Compartilhar informações entre todas as pessoas na empresa e agentes externos, fica mais simples com os dados disponibilizados na nuvem. O tempo que se ganha com o uso da tecnologia em nuvem é gigantesco. Com as informações sincronizadas e acessíveis praticamente de modo instantâneo, agiliza-se o fluxo de atividades, otimizando o tempo.

Implemente ferramentas de monitoramento e controle, as ferramentas para monitoramento e controle serão, de fato, muito úteis para averiguar as suspeitas de ataques ou vazamentos e, além disso, reforçar o sigilo interno, ou seja, garantir que as pessoas só tenham disponibilidade às informações de acordo com seu nível de autorização de acesso.

REFERENCIAS

REDAÇÃO. **O que é Big Data?** Disponível em<:<https://canaltech.com.br/big-data/o-que-e-big-data/>>. Acesso em 03 de maio de 2020.

BATINI, Carlo; SCANNAPIECA, Monica. **Data Quality: Concepts, Methodologies and Techniques**. New York. Springer, 2006

BATINI, Carlo et al. **Methodologies for Data Quality Assessment and Improvement**. ACM Computing Surveys, n.3, v.41, 2009, p. 1-52.

BATINI, Carlo. et al. **From Data Quality to Big Data Quality**. Journal of Database Management, v. 26, n. 1, 2015, p. 60–82.

BECKER, David; MCMULLEN, Bill; KING, Trish Dunn. Big Data, Big Data Quality Problem. In: **IEEE INTERNATIONAL CONFERENCE ON BIG DATA**, 2015, Santa Clara. Anais eletrônicos... Santa Clara: 2015. p.2644-2653

CAI, Li; ZHU, Yangyong. **The Challenges of Data Quality and Data Quality Assessment in the Big Data Era**. Data Science Journal, v. 14, n. 0, 2015, p. 2. Dis

CIANCARINI, Paolo; POGGI, Francesco; RUSSO, Daniel. **Big Data Quality: a Roadmap for Open Data**. **2ND IEEE INTERNATIONAL CONFERENCE ON BIG DATA COMPUTING SERVICE AND APPLICATIONS**, 2., 2016, Oxford. Anais eletrônicos... Praga: 2016.

ENDLER, Gregor; BAUMGAERTEL, Philipp; LENZ, Richard. Pay-as-you-go data quality improvement for medical centers. In: **CONFERENCE ON EHEALTH - HEALTH INFORMATICS MEETS EHEALTH**, 2013, Vienna. Anais eletrônicos... Vienna: 2013.

RDBCI: **Revista Digital Biblioteconomia e Ciência da Informação RDBCI** : Digital Journal of Library and Information Science DOI 10.20396/rdbci.v16i1.8650412 © RDBCI: Rev. Digit. Bibliotecon. Cienc. Inf. Campinas, SP v.16 n.1 p. 194-210 jan./abr. 2018 [207] p.13-18. Disponível em: . Acessoem: 7 jul. 2017.

ERL, Thomas; KHATTAK, Wajid; BUHLER, Paul. **Big Data Fundamentals: Concepts, Drivers & Techniques**. Boston: Prentice Hall, 2016.

FREITAS, Patrícia Alves de et al. **Information Governance, Big Data and Data Quality**. In: IEEE 16TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND ENGINEERING (CSE), 16., 2013, Sydney. Anais eletrônicos... Sydney: 2013. p.1142- 1143.

FURLAN, Patricia Kuzmenko; LAURINDO, Fernando José Barbin. **Agrupamentos epistemológicos de artigos publicados sobre big data analytics**. Transinformação, v. 29, n. 1, 2017, p. 91-100.

GANAPATHI, Archana; CHEN, Yanpei. **Data Quality: Experiences and Lessons from Operationalizing Big Data**. **4TH IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA)**, 4., 2016, Washington. Anais eletrônicos... Washington: 2016.

GANDOMI, Amir; HAIDER, Murtaza. **Beyond the hype: Big data concepts, methods, and analytics**. International Journal of Information Management, v. 35, n. 2, 2015, p. 137– 144.

HARYADI, AdiskaFardaniet al. **Antecedents of Big Data Quality An Empirical Examination in Financial Service Organizations**. **4TH IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA)**, 4., 2016, Washington. Anais eletrônicos... Washington: 2016.

HAZEN, Benjamin T. et al. **Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications**. International Journal of Production Economics, v. 154, 2014, p. 72–80.

JUDDOO, Suraj. **Overview of data quality challenges in the context of Big Data**. In: INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND SECURITY (ICCCS), 2015, Pamplemousses. Anais eletrônicos... Pamplemousses : 2015..

KAISLER, Stephen et al. **Big Data: Issues and Challenges Moving Forward**. In: XLVI HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 46., Maui, 2013.

KELLING, Steve et al. **Taking a 'Big Data' approach to data quality in a citizen science project**. AMBIO, v. 44, n. 4, 2015, p. S601–S611.

KWON, Ohbyung; LEE, Namyoon; SHIN, Bongsik. **Data quality management, data usage experience and acquisition intention of big data analytics**. International Journal of Information Management, v. 34, n. 3, 2014, p. 387–394.

LANEY, Doug. **Application Delivery Strategies**. META Group, 2001. Disponível em: . Acesso em: 7 jul. 2017.

MCAFEE, Andrew; BRYNJOLFSSON, Erik. Big Data. **The management revolution**. Harvard Business Review, v. 90, n. 10, 2012 p. 61–68.

MERINO, Jorge et al. **A Data Quality in Use model for Big Data**. Future Generation Computer Systems, v. 63, 2016, p.123-130.

PAIM, Isis; NEHMY, Rosa Maria Quadros, GUIMARÃES, César Geraldo. **Problematização do conceito "Qualidade" da Informação**. Perspectivas em Ciência da Informação, v. 1, n. 1, 1996, p. 111–119.

PORTAL DE PERIÓDICOS DA CAPES/MEC. 07 jun. 2017.

RAO, Dhana; GUDIVADA, Venkat N.; RAGHAVAN, Vijay V. **Data Quality Issues in Big Data**. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA, 2015, Santa Clara. Anais eletrônicos... Santa Clara: 2015.

RIBEIRO, Claudio José Silva. **Big Data: os novos desafios para o profissional da informação**. Informação & Tecnologia, v. 1, n. 1, 2014, p. 96–105.

RDBCI: Revista Digital Biblioteconomia e Ciência da Informação RDBCI : **Digital Journal of Library and Information Science** DOI 10.20396/rdbci.v16i1.8650412 © RDBCI: Rev. Digit. Bibliotecon. Cienc. Inf. Campinas, SP v.16 n.1 p. 194-210 jan./abr. 2018 [209]

SADIQ, Shazia; PAPOTTI, Paolo. Big Data Quality - Whose problem is it? 32ND **IEEE INTERNATIONAL CONFERENCE ON DATA ENGINEERING (ICDE)**, 32., 2016, Helsinki. Anais eletrônicos... Helsinki: 2016.

SAHA, Barna; SRIVASTAVA, Divesh. Data Quality: The other Face of Big Data. In: **IEEE 30TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING (ICDE)**, 30., 2014, Chicago. Anais eletrônicos...Chicago: 2014.

SOMASUNDARAM, G.; SHRIVASTAVA, Alok. **Armazenamento e Gerenciamento de Informações: Como armazenar, gerenciar e proteger informações digitais**. Porto Alegre: Bookman. 2011. 472p.

TALEB, Ikbalet al. **Big Data Quality: A Quality Dimensions Evaluation**. 13TH IEEE INT CONF ON UBIQUITOUS INTELLIGENCE AND COMP, 13., 2016, Toulouse. Anais eletrônicos... Toulouse: 2016. Disponível em: . Acesso em: 7 jul. 2017.

VALENTE, Nelma T. Zubek; FUJINO, Asa. **Atributos e dimensões de qualidade da informação nas Ciências Contábeis e na Ciência da Informação**: um estudo comparativo. *Perspectivas em Ciência da Informação*, v. 21, n. 2, 2016, p. 141–167. Disponível em: . Acesso em: 16 mar. 2017.

VIANNA, William Barbosa; DUTRA, Moisés Lima; FRAZZON, Enzo Morosini. **Big data e gestão da informação: modelagem do contexto decisional apoiado pela sistemografia**. *Informação & Informação*, v. 21, n. 1, 2016, p. 185.