

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO DE GESTÃO E ECONOMIA
ESPECIALIZAÇÃO EM MBA EM GESTÃO EMPRESARIAL

LILIAN KAISS

**ROTEIRO DE ANÁLISE DE DADOS E BIG DATA APLICADO A UM PROBLEMA
DE LOGÍSTICA**

MONOGRAFIA DE ESPECIALIZAÇÃO

CURITIBA

2020

LILIAN KAISS

**ROTEIRO DE ANÁLISE DE DADOS E BIG DATA APLICADO A UM PROBLEMA
DE LOGÍSTICA**

Monografia de especialização apresentado ao Curso MBA em Gestão Empresarial, do Departamento Acadêmico de Gestão e Economia, da Universidade Tecnológica Federal do Paraná – UTFPR, como requisito parcial para obtenção do certificado de Especialista em Gestão Empresarial

Orientador: Dr. Rodrigo Alves

CURITIBA

2020

TERMO DE APROVAÇÃO

ROTEIRO DE ANÁLISE DE DADOS E BIG DATA APLICADO A UM PROBLEMA DE LOGÍSTICA

Esta monografia foi apresentada no dia 29 de maio de 2020, como requisito parcial para a obtenção do título de Especialista em MBA em Gestão Empresarial – Universidade Tecnológica Federal do Paraná. A candidata Lilian Kaiss apresentou o trabalho para a Banca Examinadora composta pelos professores abaixo assinados. Após a deliberação, a Banca Examinadora considerou o trabalho aprovado.

Dr. Rodrigo Alves
Orientador

Msc. Egon Bianchini Calderari
Banca

Dr^a. Luciana Vieira de Lima
Banca

Visto da coordenação:

Dr. Paulo Daniel Batista de Sousa

RESUMO

KAISS, Lilian. **Roteiro de análise de dados e Big Data aplicado a um problema de logística.** 2020. 41 f. Monografia de Especialização em MBA em Gestão Empresarial – Programa de Pós-Graduação em Administração-PPGA, Universidade Tecnológica Federal do Paraná. Curitiba, 2020.

O mundo passa por uma transformação com relação aos dados que estão sendo gerados num volume crescente, nos mais variados formatos e esse fenômeno também acontece nas organizações. A ciência dos dados volta-se para a logística pela sua complexidade e importância na performance das empresas e pelos desafios que todos os elementos da cadeia de suprimentos apresentam. É determinante que as organizações tenham a capacidade de extrair valor dos dados e para tal não existem formatos prontos. Na elaboração de um roteiro para análise de dados e Big Data aplicável ao problema de logística de falta de peças no ponto de uso da linha de produção de uma fábrica é importante compreender as características gerais da análise de grande volume de dados, as principais etapas de produção, armazenamento, análise e visualização dos dados e as práticas da gestão da informação que analisam grandes volumes de dados. Além das técnicas é relevante compreender como a estrutura organizacional e as práticas da gestão da informação favorecem o sucesso do Big Data na organização. O perfil do cientista de dados dentro das organizações é um profissional com conhecimento técnico básico e conhecimento aprofundado em gestão de projetos, pois geralmente lidera equipes multifuncionais. Há uma tendência de aproximação das estruturas de TI às outras áreas de negócios, atuando de forma mais integrada. O Big Data traz vários aspectos positivos para as empresas, suportando a interpretação de dados para a tomada de decisão em tempo real, melhorando a performance e a eficiência interna e adequando os produtos e serviços às necessidades dos clientes. Os desafios da implementação do Big Data nas organizações está relacionado à questões operacionais de acesso, armazenamento, busca e compartilhamento do dado que é tratado pelo Big Data de uma forma holística na organização, o que também pode esbarrar na cultura organizacional da empresa. A revisão bibliográfica do artigo científico de WANG e colaboradores (2016, p.98-110) é feita para indicar possíveis métodos aplicáveis ao problema de logística de falta de peças no ponto de uso da linha de produção de uma fábrica. Este problema está relacionado ao planejamento operacional da logística e uma possível abordagem seria a aplicação de Big Data no inventário, otimizando e determinando os níveis adequados de inventário. Outra abordagem possível é a aplicação de Big Data na logística, utilizando todos os dados gerados na cadeia para projetar a flexibilidade nas operações logísticas. Uma abordagem adequada para desenvolver o Big Data na cadeia logística neste problema é uma estrutura colaborativa com o compartilhamento dos dados na cadeia desde os fornecedores de peças até a fábrica. A aplicação de análise preditiva ajuda a entender os padrões na falta de peças e a prever esses eventos antecipadamente, diminuindo os gastos com ações de remediação. O uso da técnica de otimização ajuda a entender as relações e conexões nos dados trazendo percepções para melhoria do processo. Uma análise mais aprofundada requer aplicação prática do Big Data nos dados, fora do escopo deste trabalho.

Palavras-chave: Big Data. Logística. Cadeia de Suprimentos.

ABSTRACT

KAISS, Lilian. **Data analysis and Big Data roadmap applied to a logistics problem**. 2020. 41 f. Monografia de Especialização em MBA em Gestão Empresarial – Programa de Pós-Graduação em Administração-PPGA, Universidade Tecnológica Federal do Paraná. Curitiba, 2020.

The world is facing a transformation related to the data being generated in an increasing volume, in the most varied formats and this phenomenon happens inside organizations. Data science focuses on logistics due to its complexity and importance in the performance of companies and the challenges that all supply chain elements present. It is crucial that organizations have the capacity to extract value from the data and for that, there are no ready formats. In the development of a roadmap for data analysis and Big Data applicable to the logistics problem of missing parts at the point of use of the production line of a factory it is important to understand the general characteristics of the analysis of Big Data, such as production, storage, analysis and visualization of the data and management practices that analyze large volumes of data. In addition to the techniques, it is relevant to understand how the organizational structure and information management practices contribute to the success of Big Data in the organization. The profile of the data scientist within organizations is a professional with basic technical knowledge and deep knowledge in project management, as he usually leads multifunctional teams. There is a tendency to bring IT structures closer to other business areas, acting in a more integrated manner. Big Data brings several positive aspects to companies, supporting the interpretation of data for real time decisions, improving performance and internal efficiency and adapting products and services to customers' needs. The challenges of implementing Big Data in organizations are related to operational issues such as access, storage, search and sharing data that is treated by Big Data in a holistic way in the organization, which can also step into the organizational culture of the company. The bibliographic review of the scientific article by WANG and collaborators (2016, p.98-110) is done to indicate possible methods applicable to the logistic problem of missing parts at the point of use of a factory production line. This problem is related to the operational planning of logistics and a possible approach would be the application of Big Data in the inventory, optimizing to define the appropriate levels of inventory. Another possible approach is the application of Big Data in logistics, using all the data generated in the chain to project flexibility in logistics operations. A suitable approach to develop Big Data in the logistics chain in this problem is a collaborative structure sharing data in the chain from the parts suppliers to the factory. The application of predictive analysis helps to understand the patterns in the lack of parts and to predict these events in advance, reducing the costs with remediative actions. The use of the optimization technique helps to understand the relationships and connections in the data, bringing insights to improve the process. Further analysis requires practical application of Big Data to the data, which is out of scope of this project.

Keywords: Big Data. Logistics. Supply Chain.

SUMÁRIO

1. INTRODUÇÃO	6
1.1 TEMA DA PESQUISA	6
1.2 PROBLEMA DE PESQUISA	6
1.3 JUSTIFICATIVA	7
1.4 OBJETIVO GERAL E ESPECÍFICO	8
1.4.1 Objetivo Geral	8
1.4.2 Objetivo Específico	8
1.5 PROCEDIMENTOS METODOLÓGICOS	8
2. FUNDAMENTAÇÃO TEÓRICA	10
2.1 CIÊNCIA DE DADOS	10
2.2 CIENTISTA DE DADOS	11
2.3 BIG DATA	13
2.4 ETAPAS DO BIG DATA E ANALYTICS	17
2.4.1 Produção de Dados	17
2.4.2 Armazenamento de Dados	18
2.4.3 Análise de Dados	21
2.4.4 Visualização de Dados	24
3. METODOLOGIA	27
3.1 BIG DATA APLICADO À LOGÍSTICA E CADEIA DE SUPRIMENTOS	27
3.2 SUPPLY CHAIN ANALYTICS NA ESTRATÉGIA LOGÍSTICA E DA CADEIA DE SUPRIMENTOS	28
3.2.1 Fornecimento Estratégico	28
3.2.2 Design da Cadeia de Suprimentos	29
3.2.3 Design e Desenvolvimento de Produtos	30
3.3 SUPPLY CHAIN ANALYTICS NA OPERAÇÃO LOGÍSTICA E DA CADEIA DE SUPRIMENTOS	31
3.3.1 Planejamento da Demanda	31
3.3.2 Compras	31
3.3.3 Produção	32
3.3.4 Inventário	32
3.3.5 Logística	33
3.4 TIPOS DE ANÁLISE DE SUPPLY CHAIN ANALYTICS	34
3.5 TÉCNICAS DE SUPPLY CHAIN ANALYTICS	35
3.6 ESTRUTURA DE MATURIDADE DE SUPPLY CHAIN ANALYTICS	36
3.6.1 Supply Chain Analytics Funcional	36
3.6.2 Supply Chain Analytics Baseado Em Processo	37
3.6.3 Supply Chain Analytics Colaborativo	37
3.6.4 Supply Chain Analytics Ágil	37
3.6.5 Supply Chain Analytics Sustentável	38
4. CONSIDERAÇÕES FINAIS	39
REFERÊNCIAS	41

1. INTRODUÇÃO

1.1 TEMA DA PESQUISA

Segundo Amaral (2016, p. XIII):

O mundo está passando por uma transformação sem precedentes e o protagonista desta mudança são os dados. Há poucas décadas, produzi-lo, armazená-lo e analisá-lo era privilégio para raros e gigantescos equipamentos, hoje, a abundância de produção e armazenamento estão tornando esta matéria-prima igualmente farta. O dado vem em quantidades nunca antes imaginadas, nos mais diversos formatos, caótico, rápido e extremamente valioso. Uma miríade de padrões, tecnologias, sistemas, modelos, algoritmos, conceitos e equipamentos são necessários para produzi-lo, armazená-lo, tratá-lo, analisá-lo, visualizá-lo e mantê-lo íntegro e seguro. Chegamos então a uma nova ciência, a ciência de dados, a fim de estudar este elemento que está se tornando o protagonista de uma nova era da história humana (AMARAL, 2016, p. XIII).

A aplicação da ciência de dados dentro das organizações está relacionada à melhoria de performance, que traz mais competitividade para as empresas. A ciência de dados aplicada à logística e cadeia de suprimentos tem chamado a atenção devido a sua complexidade e ao papel de destaque da logística e cadeia de suprimentos na melhoria da performance dos negócios (WANG et al., 2016, p. 99).

A logística e toda a cadeia de suprimentos enfrentam os desafios mais significativos que podem resultar em ineficiências e desperdícios na cadeia, como embarques atrasados, aumento dos custos de combustível, fornecedores inconsistentes e as expectativas crescentes dos clientes (WANG et al., 2016, p. 99).

A ciência de dados aplicada à logística e cadeia de suprimentos impactam as organizações melhorando a visibilidade, flexibilidade e integração das cadeias de suprimento globais e processos logísticos, gerenciando de uma forma mais efetiva a volatilidade da demanda e as flutuações de custo (WANG et al., 2016, p. 99).

1.2 PROBLEMA DE PESQUISA

Estamos vivendo a era da informação, na qual diariamente são gerados dados numa quantidade difícil de mensurar. Segundo Amaral (2016, p. 8-9):

Um smartphone de hoje tem maior capacidade que o melhor computador de 1985; temos mais de seis bilhões de pessoas com telefones celulares; mais de 1,7 bilhões de pessoas estão usando redes sociais; um disco para armazenar toda a música do mundo custa US\$ 600; quase três milhões de e-mails enviados por segundo; cem horas de

vídeo são carregados no Youtube por minuto; quinhentos milhões de tweets por dia; 92% dos dados do mundo foram criados nos últimos dois anos; 2,3 trilhões de GB de dados são criados por dia; cem terabytes de dados são carregados por dia no Facebook (AMARAL, 2016, p. 8-9).

Segundo o conteúdo didático disponível no website do Serviço Nacional de Aprendizagem Industrial (SENAI, 2019): “90% dos dados disponíveis hoje foram gerados nos últimos dois anos, o que demonstra a velocidade da geração dos dados na sociedade atual”. De acordo com o material didático disponível no website da instituição de ensino (SENAI, 2019):

O fator principal para a indústria é a capacidade de processar e avaliar as informações relevantes, pois de nada serve possuir grandes volumes de dados se não puder fazer uso deles. É preciso extrair conhecimentos úteis e valiosos, de modo que se faz necessário o uso de ferramentas e técnicas de gestão para processar grande volume de dados, em diversos formatos e em velocidade adequada (SENAI, 2019).

No contexto organizacional, como definir o caminho a ser seguido para trabalhar com dados e extrair informações que agregam valor para a empresa? Como deve ser trabalhada e tratada a questão dos dados de uma fábrica que diariamente recebe peças para montagem e coleta a informação de suprimento e de falta de peças numa linha de montagem de modo que conhecimentos úteis sejam gerados ao longo do processo com os dados coletados diariamente?

1.3 JUSTIFICATIVA

No contexto onde as organizações geram cada vez mais dados e numa velocidade cada vez maior, como as organizações podem se preparar para ter a capacidade de analisar e aproveitar os dados gerados na cadeia, seja no processo produtivo ou serviços (clientes, fornecedores, parceiros, etc.), é uma questão de sobrevivência e competitividade aprender a trabalhar com os seus dados. Segundo o conteúdo didático disponível no website do Serviço Nacional de Aprendizagem Industrial (SENAI, 2019): “na era digital, a informação é considerada um ativo, mas sem a capacidade de processamento e análise o dado é um item sem valor”.

O Big Data aplicado à logística e cadeia de suprimentos está nos seus estágios iniciais e ainda há muito espaço para ser explorado, dificilmente se encontra um modelo pronto para ser implementado nas questões logísticas. Estudos recentes sobre as aplicações de Big Data na logística e cadeia de suprimentos concentraram-se principalmente nas definições de análise e

diferentes perspectivas e na identificação de oportunidades para pesquisa e educação na cadeia de suprimentos (WANG et al., 2016, p. 99).

Além da metodologia de trabalho com os dados em si, é importante entender de que forma a estrutura organizacional e as práticas de gestão da informação influenciam o êxito da aplicação do Big Data dentro das organizações.

1.4 OBJETIVO GERAL E ESPECÍFICO

1.4.1 Objetivo Geral

Elaborar um roteiro para Análise de Dados e Big Data aplicável ao problema de logística de falta de peças no ponto de uso da linha de produção de uma fábrica.

1.4.2 Objetivo Específico

- a) Estudar as características gerais de análise de grandes volumes de dados e os principais elementos envolvidos neste processo: produção, armazenamento, análise e visualização dos dados
- b) Levantar as práticas de gestão da informação de organizações que analisam grandes volumes de dados
- c) Indicar possíveis métodos de Big Data aplicáveis ao problema de logística de falta de peças no ponto de uso da linha de produção de uma fábrica. A falta da peça na estação de montagem pode ser ocasionada por problemas de qualidade, atraso nas entregas ou problemas de capacidade do fornecedor e também é ocasionada por mudanças no programa de produção que flutua conforme as variações de demanda dos clientes. Dados de peças faltantes são coletados diariamente ao longo dos anos mas não são usados para identificar padrões para entendimento e até mesmo previsões antecipadas de falta.

1.5 PROCEDIMENTOS METODOLÓGICOS

O trabalho será embasado em uma revisão bibliográfica das características gerais de análise de grandes volumes de dados e os principais elementos envolvidos neste processo e das

práticas de gestão da informação de organizações que analisam grandes volumes de dados segundo os autores referenciados neste trabalho.

As possíveis abordagens de Big Data aplicáveis ao problema de logística de falta de peças no ponto de uso da linha de produção de uma fábrica serão descritas nas considerações finais deste trabalho com base na revisão bibliográfica de um artigo científico de Big Data e Analytics aplicado a logística e cadeia de suprimentos de WANG e colaboradores (2016, p. 98-110).

2. FUNDAMENTAÇÃO TEÓRICA

2.1 CIÊNCIA DE DADOS

Dado pode ser definido como o fato coletado e armazenado que pode estar em formato analógico ou digital ou ainda existir em um formato não eletrônico (AMARAL, 2016, p. 3). Os dados podem ser gerados por humanos ou por máquinas (MARQUESONE, 2016, p. 20). Dados gerados por humanos são aqueles cujo conteúdo foi gerado a partir de um pensamento, são os dados gerados para documentar algo e também os dados que refletem a interação das pessoas no mundo digital (MARQUESONE, 2016, p. 20). Os dados gerados por máquinas são os dados digitais produzidos por processos de computadores, aplicações e outros mecanismos além dos computadores como sensores e etiquetas de objetos (MARQUESONE, 2016, p. 22). O dado analisado e com algum significado gera a informação que interpretada, entendida e aplicada para um fim gera o conhecimento (AMARAL, 2016, p. 3).

Nas organizações a ciência de dados se dedica a estudar o dado em todo seu ciclo de vida, da produção ao descarte e abrange outras ciências, modelos e tecnologias, processos e procedimentos, mas normalmente está associada somente aos processos de análise dos dados correspondente à parte estatística do processo (AMARAL, 2016, p. 4).

Dispositivos como computadores ou sensores produzem o dado em formato digital que deve ser preservado em um dispositivo eletrônico para utilização, a produção do dado só fará sentido se o mesmo for mantido em algum tipo de mídia (AMARAL, 2016, p. 4). Após esta etapa, o dado passa por processos de transformação necessários para adequá-los ao modelo de utilização que tem estrutura diferente do modelo de armazenagem de dados (AMARAL, 2016, p. 4).

Dados produzidos, armazenados e transformados partem então para a etapa de análise de dados que consiste na execução operações para extrair informação e conhecimento dos dados (AMARAL, 2016, p. 5). Outra etapa importante na vida do dado é a visualização, que geralmente não altera a estrutura do dado mas o torna mais adequado para que o consumidor do dado o compreenda de forma clara e intuitiva (AMARAL, 2016, p. 5).

No final de sua vida útil, o dado passa pelo processo de descarte, mas ainda existem mais elementos relacionados ao ciclo de vida do dado como segurança, privacidade, qualidade e políticas estabelecidas dentro de cada organização (AMARAL, 2016, p. 5).

2.2 CIENTISTA DE DADOS

Seguindo a linha de quem cursou engenharia é engenheiro, quem estudou medicina é médico, é difícil definir o que um cientista de dados estudou (ILHÉU, 2020). A profissão de cientista de dados é uma profissão nova, com um mercado vasto mas com a demanda desorganizada (ILHÉU, 2020). Pelo fato de existir pouca oferta de profissionais com formação específica, os cursos existentes no Brasil foram abertos recentemente, e matemáticos, físicos, analistas de sistema e até engenheiros ocupam as vagas de cientista de dados (ILHÉU, 2020). Profissionais que atuam como cientista de dados sugerem que as especializações nessa área são um bom caminho de aperfeiçoamento até mesmo para profissionais com outras formações (ILHÉU, 2020).

Quando o Big Data se tornou um termo em evidência, emergiu também o cientista de dados (AMARAL, 2016, p. 13). A maioria dos especialistas descrevia o cientista de dados como um profissional com conhecimentos muito técnicos, especialista em tudo, porém o cenário que está se consolidando no mercado aponta para um perfil profissional um pouco diferente (AMARAL, 2016, p. 14).

Na prática, o que o mercado precisa de fato são profissionais multidisciplinares, que tenham conhecimento técnico básico para que atuem como um líder, cercado de profissionais especializados em suas respectivas áreas de conhecimento, estes sim são extremamente especializados no que fazem (AMARAL, 2016, p. 14). O cientista de dados juntamente com sua equipe multidisciplinar vai propor a implementação de soluções para os desafios apresentados, que serão desenvolvidos dentro do conceito de projetos, contendo escopo, prazos e custos (AMARAL, 2016, p. 14). O cientista de dados precisa sim do conhecimento técnico para o sucesso no projeto, mas é fundamental que ele domine o gerenciamento de projetos, dessa forma o perfil do cientista de dados pode ser definido com um profissional com conhecimento multidisciplinar e conhecimento no gerenciamento de projetos (AMARAL, 2016, p. 15).

O Big Data é implementado por equipes multidisciplinares lideradas pelo cientista de dados com especialistas em diversas áreas como equipes de extração, DBA, programadores, especialistas no assunto, estatístico, minerador de dados, especialistas em ferramentas específicas, arquiteto, coordenador de infraestrutura, analista de negócio, designer, entre outros (AMARAL, 2016, p. 16).

Além do papel do cientista de dados, também é interessante discutir o papel e a estrutura de TI dentro da organização, pois ela influencia diretamente na função do cientista de dados, facilitando ou dificultando seu trabalho. A posição de TI nas estruturas organizacionais ainda sofre grande influência de como os dirigentes vêem seu papel na criação de valor e em estratégias competitivas, entretanto esta visão tem se modificado nos últimos anos e os dirigentes tem que estar atentos a estas mudanças (LAURETTI, 2019).

Nas últimas décadas, muitas empresas passaram a adotar uma estrutura matricial, na qual as unidades de negócio organizam as atividades primárias e as atividades de suporte atendem as necessidades das unidades de negócio (LAURETTI, 2019). Empresas com esta visão se apoiam em conceitos ultrapassados, onde as empresas são voltadas para dentro, focando nas suas próprias atividades, com estruturas hierárquicas pesadas e rígidas (LAURETTI, 2019). Este perfil de empresa tem dificuldades com a inovação, já que estão focadas nos seus produtos e processos atuais e não percebem as mudanças que estão acontecendo no mercado (LAURETTI, 2019).

O que se observa atualmente nas empresas de alto desempenho é o fenômeno que se chama de “tombamento” (*tilting*) das estruturas da organização, ou seja, corresponde ao movimento figurado de virar a estrutura organizacional da empresa (LAURETTI, 2019). Ao invés das empresas se organizarem em atividades, organizam-se por processos, implicando na adoção de novas maneiras de trabalhar e de gerenciar o trabalho (LAURETTI, 2019).

A gestão por processos organizacionais agrupa funcionários e recursos para produzir um trabalho completo e a informação segue o fluxo para onde é necessária, sem o filtro da hierarquia, diferente da gestão tradicional por funções (LAURETTI, 2019).

Nesse contexto, os profissionais que analisam, organizam, estruturam e aperfeiçoam os processos organizacionais têm um papel-chave (LAURETTI, 2019). Profissionais de TI com uma sólida visão de processos e de integração de toda a organização podem conduzir estas mudanças organizacionais, pois são capazes de se tornar excelentes gestores de alguns processos fundamentais para as empresas, como por exemplo, o da gestão do conhecimento organizacional (LAURETTI, 2019).

Nesta perspectiva, todas as indústrias dependem da tecnologia para funcionarem de forma mais eficiente, inovadora e com qualidade e a posição de *Chief Technology Officer* (CTO) é definida como o profissional que administra essa tecnologia junto à gestão (BATTISTELLI, 2019). Entre as funções do CTO estão prestar suporte aos desenvolvedores, alinhar a tecnologia à estratégia de gestão, criar a infraestrutura tecnológica, inovação, treinamento e desenvolvimento das equipes (BATTISTELLI, 2019). O CTO precisa estar

sempre atualizado nas condições do mercado para implementar novas tecnologias que se encaixem com os propósitos e necessidades da organização (BATTISTELLI, 2019).

O exemplo do ex-presidente dos Estados Unidos, Barack Obama é citado por Battistelli (2019) para evidenciar a importância da função do CTO em qualquer organização: em 2009 Aneesh Chopra foi nomeado como o primeiro CTO a servir o governo do país. Ele foi contratado para proporcionar as tecnologias que suprissem problemas sociais, como o acesso a postos de trabalho, além de promover melhorias nos sistemas digitais de departamentos públicos de saúde e de segurança (BATTISTELLI, 2019). Ampliar o acesso à banda larga era uma tarefa de fundo social, na opinião de Obama, motivo pelo qual essa também foi uma missão dada a Chopra enquanto ele ocupava tal colocação (BATTISTELLI, 2019).

2.3 BIG DATA

Big Data pode ser caracterizado como um fenômeno associado a grandes volumes de dados, embora a definição formal seja caracterizada por um conjunto de três a cinco “Vs”, sendo os três primeiros “Vs”: volume, velocidade e variedade e os outros dois “Vs”: veracidade e valor (AMARAL, 2016, p. 7). Além destas características, o Big Data pode ser definido como o fenômeno em que dados são produzidos em vários formatos e armazenados por uma grande quantidade de dispositivos e equipamentos (AMARAL, 2016, p. 7). A evolução tecnológica avançou muito em termos de miniaturização e capacidade de processamento, e os insumos de tecnologia, como CPUs, memórias e unidades de armazenamento vêm se tornando cada vez mais baratos (AMARAL, 2016, p. 7). O barateamento, a miniaturização e o aumento da capacidade de processamento levam à disseminação de equipamentos, dispositivos e processos capazes de produzir e armazenar dados, virtualização, computação na nuvem e internet (AMARAL, 2016, p. 7).

O Big Data vai além da dimensão de volume de dados, está relacionado principalmente à diversidade: dados em todos os formatos, por todos os lados, de todas as maneiras (AMARAL, 2016, p. 9). Big Data também traz uma mudança sociocultural, é uma nova fase da revolução industrial, é um fenômeno e não somente uma tecnologia e o tamanho do impacto social, cultural e empresarial é difícil de mensurar, mas já se pode afirmar que vai mudar o mundo como se conhece hoje (AMARAL, 2016, p. 9). O Big Data não impõe o fim dos modelos tradicionais, que funcionaram muito bem até hoje mas não são mais adequados para resolver

problemas de dados para os quais quando foram desenhados não existiam (AMARAL, 2016, p. 9).

Do ponto de vista tecnológico, o elemento principal associado ao Big Data é o registro de qualquer fenômeno, natural ou não, em dados que são armazenados para reprodução ou análise imediata ou futura (AMARAL, 2016, p. 10). O Big Data pode ser definido ainda como o fenômeno da massificação de elementos de produção e armazenamento de dados, bem como os processos e tecnologias para extraí-los e analisá-los (AMARAL, 2016, p. 12).

O Big Data faz com que os eventos sejam armazenados cada vez mais para serem reproduzidos e analisados (AMARAL, 2016, p. 10). Antes, armazenar dados era considerado caro, por isso apenas os dados com valor imediato eram mantidos, as empresas guardavam somente dados operacionais de vendas, compras e contabilidade e os demais dados eram descartados o mais rápido possível. Outra característica que o Big Data traz é a capacidade de armazenar dados de forma indiscriminada (AMARAL, 2016, p. 10).

O Big Data também traz uma mudança do conceito de amostragem (AMARAL, 2016, p. 10). A amostragem torna os processos mais rápidos e baratos, mas ela tem um preço: a margem de erro (AMARAL, 2016, p. 10). O Big Data pode reduzir drasticamente a necessidade de processos de análise de dados por amostragem, pois a análise se dará sob 100% dos dados, eliminando o risco da transação mais crucial ao processo ser descartada durante o procedimento de amostragem (AMARAL, 2016, p. 10).

O Big Data envolve o uso de diversos tipos de conceitos e tecnologias, como computação nas nuvens, virtualização, internet, estatística, infraestrutura, armazenamento, processamento, governança e gestão de projetos (AMARAL, 2016, p. 11). Para as empresas o Big Data traz muitas oportunidades, tanto na vantagem competitiva como na criação de produtos e serviços orientados a dados (AMARAL, 2016, p. 11).

As empresas utilizam a análise de dados principalmente para melhorar sua atividade principal e o que está relacionado diretamente ao seu negócio (AMARAL, 2016, p. 11). O Big Data permite que os dados sejam utilizados para tornar a empresa muito mais competitiva e eficiente, coletando e analisando dados que não estão apenas relacionados diretamente a sua atividade principal (AMARAL, 2016, p. 11). A empresa pode utilizar a análise de dados para fazer melhorias em outras áreas melhorando os processos de recrutamento, melhorando a produtividade dos colaboradores e até mesmo melhorando a eficiência da linha de produção (AMARAL, 2016, p. 11). A análise de dados também pode contribuir no relacionamento com os clientes, a empresa será capaz de entender melhor as necessidades de seus clientes, aumentando a satisfação e a fidelização de clientes através (AMARAL, 2016, p. 11). Os dados

também ajudam as empresas a selecionar seus fornecedores baseados em índices mais precisos (AMARAL, 2016, p. 11). A análise de dados também traz a possibilidade das empresas analisarem os concorrentes, descobrindo em quais aspectos os clientes estão satisfeitos, buscando estratégias para atraí-los e trabalhando no lançamento de novos produtos (AMARAL, 2016, p. 11). O Big Data é uma ferramenta vital para a inovação de produtos e serviços, fundamental para a sobrevivência das empresas e as que não souberem usá-lo vão desaparecer, engolidas pelas concorrentes que serão mais eficientes, com custos menores, com produtos de mais qualidade e clientes mais satisfeitos (AMARAL, 2016, p. 11).

Há um cenário amplo de aplicação do Big Data nas organizações (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 49). Empresas de segmentos que há tempos utilizam-se dos dados disponíveis tendem a avançar na vanguarda do Big Data: empresas on-line, empresas de bens de consumo, seguradoras, empresas do segmento de turismo e transporte ou empresas de cartões de crédito (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 49). Outros segmentos tem alto potencial de uso do Big Data: companhias bancárias, telecomunicações, mídia e entretenimento, varejo e empresas de utilidade pública, como concessionárias de energia elétrica (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 49). Outros setores, como o de saúde, empresas B2B (*business-to-business*) e produtos industriais são caracterizados por organizações que por diversos motivos tinham menor acesso aos dados e agora podem fazer uso daqueles provenientes dos novos meios de produção e análise de informações (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 49).

Um volume maior de dados captados também pode significar dados mais específicos (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 54). O Big Data aplicado em estudos de marketing com grande volume de dados de clientes pode melhorar os serviços e torná-los mais personalizados, uma maior capacidade analítica pode fornecer informações úteis às organizações sobre cada um de seus clientes (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 54).

Nas organizações há a necessidade de tomada de decisão em tempo real, gerando a necessidade de se obter dados de maneira on-line para prestar apoio às decisões em tempo hábil (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 54). Personalização, decisão em tempo real, estrutura de serviço e o *service-oriented thinking* são tendências em forte crescimento (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 54). Para as empresas, o Big Data traz ganho de escala na análise de dados, que é fundamental devido à quantidade de dados disponíveis para serem analisados (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 54). Além disso, dados disponíveis para serem analisados em tempo real tem o potencial de tornar os profissionais e as

decisões mais eficientes e eficazes, contribuindo para uma redução no tempo de tomada de decisão (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 54).

Quando uma organização implementa uma solução Big Data, significa que adotou mecanismos para lidar, analisar e retirar vantagens do fato de manter um grande volume de informação, exigindo preparação, trabalho e estratégia (VIEIRA; PEDROSA; SOARES, 2019, p. 1). A fase de análise de dados traz novos conhecimentos para as organizações através de conclusões úteis, observação de tendências, diferenças e variações na informação obtida (VIEIRA; PEDROSA; SOARES, 2019, p. 1). Os avanços com o Big Data e Analytics além de permitirem melhor tomada de decisão e aumento da vantagem competitiva da organização, impulsionam a necessidade de encontrar significado nos dados que estão em constante mudança e de encontrar relações entre esses dados (VIEIRA; PEDROSA; SOARES, 2019, p. 1).

O Big Data também traz alguns desafios para as empresas, se por um lado há uma grande quantidade de dados para descobrir o que antes não era possível, por outro há desafios em lidar com esta enorme quantidade de dados (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 55). Questões relacionadas ao acesso, armazenamento, busca e compartilhamento dos dados ainda restringem e dificultam sua operacionalização nas organizações, muitas vezes contornar estes desafios é mais complicado do que trabalhar os dados em si (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 55). Outro ponto é que as organizações ainda se mostram muito céticas quanto à adoção de soluções de Big Data e o conceito de cultura orientada para os dados ainda não faz parte da cultura das empresas (VIEIRA; PEDROSA; SOARES, 2019, p. 1).

Outro desafio que o Big Data traz para as empresas é o custo em lidar com essa grande quantidade de informação (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 55). Além disso, reunir o conhecimento com uma produção de dados crescente também é desafiador, pois requer ferramentas que precisam ser mais inteligentes e sofisticadas para filtrar todos os dados de pesquisa disponíveis mas que não podem se tornar obsoletas com o aumento na quantidade de dados (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 55). Ainda há desafios importantes relacionados aos recursos humanos como treinamento e capacitação que não são facilmente encontrados (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 55).

Ainda há os desafios inerentes ao funcionamento do Big Data, embora haja uma série de ferramentas e técnicas com objetivo de limpar os dados, às vezes não há tempo disponível para analisar os volumosos conjuntos de dados, tornando a inconsistência dos dados ainda um desafio a ser superado (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 55). Outra questão é a privacidade, as fontes de dados são propriedade de terceiros e a aquisição ou compartilhamento

deve ser negociado, com o risco de não ser disponibilizado quando forem considerados fonte de vantagem competitiva (SILVEIRA; MARCOLIN; FREITAS, 2015, p. 55).

2.4 ETAPAS DO BIG DATA E ANALYTICS

Da produção ao descarte o dado passa por uma série de etapas (AMARAL, 2016, p. 17). Alguns dados não sofrem nenhuma transformação após serem produzidos, podem ser descartados após a produção ou podem ser armazenados por tempo indeterminado (AMARAL, 2016, p. 17). Normas corporativas e a legislação vigente determinam por quanto tempo o dado deve ser armazenado e quando será descartado por processos internos (AMARAL, 2016, p. 17). As etapas do dado no Big Data dependem da finalidade da análise, mas geralmente o ciclo padrão abrange seis etapas: produção, armazenamento, transformação, armazenamento analítico, análise e descarte (AMARAL, 2016, p. 17).

Uma vez definido o foco do projeto de Big Data e as respostas que se desejam obter através dos dados, os dados que deverão ser capturados podem ser identificados. Alguns questionamentos podem ser usados como apoio para identificar quais e como os dados serão coletados: esses dados já existem ou precisam ser gerados? São internos ou externos? Em qual formato eles estão? (MARQUESONE, 2016, p. 31).

2.4.1 Produção de Dados

A captura dos dados é definida como a etapa inicial do Big Data e nesta etapa deve ser traçada uma estratégia para definir como os dados capturados serão armazenados (MARQUESONE, 2016, p. 31).

Os dados internos são dados que a empresa é dona e possui controle como dados de sistemas de gerenciamento da empresa, arquivos e documentos gerados por colaboradores, sensores, registros de eventos, dados de servidores e demais registros internos (MARQUESONE, 2016, p. 31). Uma recomendação para empresas que pretendem iniciar sua jornada em Big Data é identificar formas de organizar, analisar e utilizar seus dados internos para melhoria dos negócios, além de esses dados serem mais fáceis de serem adquiridos do que

os externos, eles podem revelar informações importantes para as decisões da empresa (MARQUESONE, 2016, p. 33).

A captação de dados internos através de sensores pode ser feita com uma solução IoT (*Internet of Things* – Internet das Coisas), na qual os objetos possuem capacidade de comunicação com outros objetos e pessoas (MARQUESONE, 2016, p. 33). Essa solução requer um meio de transmissão de dados entre os sensores e um servidor para prover o armazenamento devido ao baixo poder de armazenamento da maioria dos sensores (MARQUESONE, 2016, p. 33).

Os dados externos são dados coletados de fontes externas como dados de domínio público, dados de sites de terceiros e mídias sociais online (MARQUESONE, 2016, p. 33). Neste tipo de dado, a captura, armazenamento e análise são um desafio devido à variedade dos dados (MARQUESONE, 2016, p. 33). A principal forma de captura de dados de mídias sociais online é por meio de uma API (*Application Programming Interface*), que pode ser definida como um conjunto de instruções e padrões de programação para que usuários tenham acesso aos dados de um aplicativo ou plataforma. (MARQUESONE, 2016, p. 34).

O acesso aos dados de domínio público tem sido facilitado pelo conceito de dados abertos (*open data*) (MARQUESONE, 2016, p. 35). São considerados dados abertos aqueles que qualquer pessoa pode livremente usar, reutilizar e redistribuir, sem restrições legais, tecnológicas ou sociais e o livre acesso a esses dados contribui trazendo informação de qualidade para a sociedade, facilitando a compreensão sobre a situação de um contexto público (MARQUESONE, 2016, p. 35).

2.4.2 Armazenamento de Dados

Por algumas décadas o banco de dados relacional foi o padrão mundial para armazenagem de dados e como o nome sugere, neste modelo os dados são armazenados em estruturas de tabelas que podem estar relacionadas com outras da mesma base de dados (MARQUESONE, 2016, p. 36). Na criação de bancos de dados neste modelo, surgiram diferentes SGBDRs (Sistema de Gerenciamento de Bancos de Dados Relacionais) como Oracle, PostgreSQL e MySQL (MARQUESONE, 2016, p. 36). Os SGBDRs são caracterizados pela alta integridade aos dados armazenados e o uso da *Structured Query Language* (SQL) para operações de criação e manipulação dos dados (MARQUESONE, 2016, p. 36). Os SGBDRs

trouxeram mudanças significativas no gerenciamento de dados, trazendo integridade e consultas complexas dos dados, sendo aplicado em diversos segmentos (MARQUESONE, 2016, p. 36).

Os SGBDRs também trouxeram alguns desafios com relação à escalabilidade, disponibilidade e flexibilidade desse modelo para grande volume e variedade de dados, principalmente para os dados gerados nos últimos anos por aplicações Web (MARQUESONE, 2016, p. 37). A escalabilidade está relacionada à capacidade de manter o mesmo desempenho com adição de nova carga e a inicialmente os SGBDRs conseguiam garantir esse desempenho com a adição de mais recursos computacionais de infraestrutura ao servidor que hospeda o banco de dados, mas à medida que o volume de dados aumentou consideravelmente, esse modelo ficou inviável (MARQUESONE, 2016, p. 37). Outro desafio do SGBDRs está relacionado à capacidade do sistema manter a integridade de dados, o que pode tornar o serviço indisponível quando alguma transação violar uma regra, pois para algumas aplicações manter o sistema disponível é mais importante que a própria integridade (MARQUESONE, 2016, p. 37). Com relação à flexibilidade, para utilizar um SGBDR é necessário ter inicialmente planejado toda a modelagem dos dados antes de armazená-los, mas para muitas soluções atuais devido à característica não estruturada dos dados não é viável o conhecimento antecipado dos dados. Por essas razões os SGBDRs deixaram de ser a solução ideal para aplicações e serviços que necessitavam de escalabilidade, alta disponibilidade e flexibilidade, originado o termo NoSQL (*not only SQL*) (MARQUESONE, 2016, p. 37).

Diferente do banco de dados relacional, os modelos NoSQL sacrificam algumas propriedades de integridade para um maior desempenho e escalabilidade para soluções que trabalham com grande volume de dados (MARQUESONE, 2016, p. 38). Os modelos NoSQL são classificados conforme a estrutura que os dados são armazenados em modelo orientado a chave-valor, orientado a documentos, orientado a colunas e orientado a grafos (MARQUESONE, 2016, p. 38).

O banco de dados orientado a chave-valor possui a estrutura mais simples entre todos os modelos que utilizam como estratégia o armazenamento de dados utilizando chaves como identificadores de informações gravadas em um campo identificado valor (MARQUESONE, 2016, p. 39). Banco de dados orientados a chave-valor são adequados para aplicações que realizam leituras frequentes mas apresentam algumas limitações para consultas mais complexas (MARQUESONE, 2016, p. 39). Este tipo de banco de dados pode ser aplicado em diversos cenários: armazenamento de imagens e de documentos, criação de cache de objetos,

armazenamento de dados de sessões do usuário e rastreamento de atributos transientes (MARQUESONE, 2016, p. 39).

O banco de dados orientado a documentos é uma extensão do banco de dados orientado a chave-valor e atualmente o modelo NoSQL mais popular (MARQUESONE, 2016, p. 41). Além de oferecer simplicidade e flexibilidade no gerenciamento dos dados, também oferece meios de criação de índices sobre os valores dos dados armazenados, criando mais possibilidades de consultas definindo documentos como estruturas flexíveis que podem ser obtidas por meio de dados semiestruturados no formato XML e JSON (MARQUESONE, 2016, p. 41). A vantagem deste modelo com relação ao modelo chave-valor é a criação de consultas e filtros sobre os valores armazenados (MARQUESONE, 2016, p. 42). A disponibilidade deste banco permite trabalhar com a replicação dos dados em um cluster, garantindo que o dado ficará disponível mesmo com a ocorrência de falha em um dos servidores (MARQUESONE, 2016, p. 42). Esse modelo é indicado para aplicações que trabalham diretamente com coleções de documentos como o armazenamento de conteúdo de páginas Web, a catalogação de documentos de uma empresa e o gerenciamento de inventário de um *e-commerce* (MARQUESONE, 2016, p. 42).

O banco de dados orientado a colunas é o mais complexo de todos os modelos de armazenamento NoSQL, é considerado uma extensão do armazenamento orientado a chave-valor e possui conceitos similares ao do modelo relacional, como a criação de linhas e colunas (MARQUESONE, 2016, p. 42). Esse tipo de banco trata de resolver problemas de escalabilidade e flexibilidade no armazenamento de dados, sendo a solução mais adequada para aplicações com volumes imensos de dados, com necessidade de alto desempenho e de alta disponibilidade na leitura e escrita dos dados (MARQUESONE, 2016, p. 43). A característica flexibilidade deste tipo de banco de dados está ligada à modelagem por “famílias de colunas”, ao invés de definir antecipadamente as colunas necessárias para armazenar um registro, o responsável pela modelagem de dados define famílias de colunas que são organizadas em grupos de itens de dados frequentemente usados em conjunto em uma aplicação (MARQUESONE, 2016, p. 44). Esse modelo oferece flexibilidade e grande escalabilidade, o registro de um item pode ter informações gravadas em diversas famílias de colunas que podem estar armazenadas em diferentes servidores (MARQUESONE, 2016, p. 44).

O banco de dados orientado a grafos é recomendado para situações com foco no relacionamento dos dados, há casos em que a descoberta de como os dados estão relacionados é mais importante do que os dados em si (MARQUESONE, 2016, p. 46). Este modelo é o mais especializado entre os quatro tipos de armazenamento NoSQL (MARQUESONE, 2016, p. 46).

Diferente dos outros modelos, em vez dos dados serem modelados utilizando o formato de linhas e colunas, eles possuem uma estrutura definida na teoria dos grafos, usando vértices e arestas para armazenar os dados dos itens coletados e os relacionamentos entre esses dados, respectivamente (MARQUESONE, 2016, p. 46).

Na prática, a tendência é que empresas adotem soluções híbridas, com diferentes modelos de banco de dados, relacionais e NoSQL, para as diferentes aplicações (MARQUESONE, 2016, p. 47). Um estudo de comparação é uma ótima estratégia na escolha do banco de dados mais adequado para cada situação (MARQUESONE, 2016, p. 47).

2.4.3 Análise de Dados

A fase de processamento inicia-se após a captura e armazenamento de dados, na qual algoritmos capazes de processar terabytes ou até pentabytes de dados deverão ser aplicados, representando um desafio no Big Data (MARQUESONE, 2016, p. 62).

Quando se trata do processamento de grande volume de dados é necessário garantir a escalabilidade da solução através da gestão adequada dos recursos computacionais utilizados e do monitoramento contínuo da execução para identificar quedas de desempenho e criar mecanismos que impeçam que a solução esgote algum recurso e seja interrompida (MARQUESONE, 2016, p. 62). Caso esses controles não sejam realizados, um projeto de Big Data com um ótimo propósito pode ser invalidado (MARQUESONE, 2016, p. 62).

Dentro do projeto de Big Data é necessário que as organizações façam um planejamento de escalabilidade de acordo com a demanda, sem deixar de manter a disponibilidade mesmo com a ocorrência de falhas, que possivelmente vão ocorrer com o uso de inúmeras máquinas (MARQUESONE, 2016, p. 62).

Uma das primeiras tecnologias de Big Data que até hoje continua sendo utilizada é o Hadoop, inicialmente projetado por Doug Cutting e Mike Cafarella como uma ferramenta de busca na Web, tal como o serviço do Google, porém com o código de programação aberto lançado oficialmente em 2006 (MARQUESONE, 2016, p. 65). O Hadoop tem muitas vantagens em aplicações com grande volume de dados como baixo custo, escalabilidade, tolerância a falhas e aplicação em novas análises (MARQUESONE, 2016, p. 67). Desde o início o Hadoop foi projetado para o armazenamento e processamento de dados em servidores tradicionais, reduzindo custos com infraestrutura, além de ser gratuito e embora tenha sido desenvolvido

para um propósito específico, grandes empresas passaram a usar o Hadoop para em diversas aplicações de Big Data (MARQUESONE, 2016, p. 67).

A fase de análise inicia-se após a captura, armazenamento e processamento de dados (MARQUESONE, 2016, p. 97). Na prática, raramente a análise inicia-se com dados estruturados, muitas vezes a base de dados tem dados incompletos, inconsistentes, corrompidos, duplicados, em formatos inadequados, com caracteres indesejados, etc. (MARQUESONE, 2016, p. 97). Por esse motivo é necessário realizar o tratamento dos dados antes da análise e geralmente a preparação e limpeza dos dados consomem 80% do tempo total do processo de análise de dados (MARQUESONE, 2016, p. 97). Como cada base de dados possui sua peculiaridade, muitas operações de tratamento de dados são manuais e de difícil automatização, porém apesar de consumir recursos e tempo, o tratamento dos dados é muito importante para evitar inconsistências nos resultados das análises (MARQUESONE, 2016, p. 98).

É muito comum a utilização de dados sem refinamento no Big Data, o problema é que sem o processo de limpeza, dados incorretos podem não ser descartados ou corrigidos e caso sejam usados no modelo analítico, o resultado obtido pode não representar a realidade dos fatos ou ainda pior, se uma organização tomar uma decisão orientada por esses resultados, pode desencadear uma série de ações baseadas em fatos inconsistentes (MARQUESONE, 2016, p. 98).

Para se obter sucesso na análise de dados, é preciso seguir um processo de análise que geralmente passa pelas seguintes etapas (MARQUESONE, 2016, p. 100):

1. Entendimento do negócio: etapa de definição das perguntas, do objetivo da análise de dados e do plano a ser seguido;
2. Compreensão dos dados: etapa de coleta e exploração dos dados para compreensão da estrutura, atributos e contexto;
3. Preparação dos dados: etapa de limpeza, filtragem, estruturação, redução e integração dos dados;
4. Modelagem dos dados: etapa de seleção dos dados, definição e construção do modelo;
5. Validação do modelo: etapa de avaliação dos resultados gerados pelo modelo;
6. Utilização do modelo: etapa de utilização e monitoramento dos resultados do modelo.

O processo de limpeza requer uma inspeção minuciosa dos dados e operações de correção e remoção e alguns questionamentos podem ajudar a identificar quais operações devem ser realizadas: existem dados duplicados, incompletos, com erros de digitação, dados

iguais representados de diferentes formas ou até mesmo dados que violam as regras de negócio (MARQUESONE, 2016, p. 102)? Algumas linguagens que possuem pacotes com funções específicas para tratamento de dados podem ajudar nessas operações como R e Python (MARQUESONE, 2016, p. 102).

Após os dados serem preparados para a análise, inicia-se a fase de modelagem dos dados (MARQUESONE, 2016, p. 107). As atividades relacionadas à mineração dos dados podem ser divididas em duas categorias: descritiva, que tem como foco caracterizar e apresentar as propriedades de um conjunto de dados de maneira concisa e informativa; e preditiva, cujo objetivo é construir um modelo para prever as propriedades e tendências de um conjunto de dados desconhecido (MARQUESONE, 2016, p. 107).

Após os dados serem tratados e ter sido construído o modelo de acordo com a análise desejada, inicia-se a fase de validação do modelo com o objetivo de avaliar o desempenho do modelo por meio de dados reais (MARQUESONE, 2016, p. 114). Existem diversas formas para medir a qualidade de um modelo, as mais comuns são (MARQUESONE, 2016, p. 114):

1. Utilização de medidas estatísticas para validação do uso correto dos dados e do modelo;
2. Separação da base de dados em treinamento e teste para avaliar o modelo antes da utilização em um ambiente de produção;
3. Avaliação de profissionais especializados em análise de dados e na área de negócio em que o modelo foi aplicado para determinar se os resultados são coerentes e significativos.

Na prática, é comum que mais de uma forma seja utilizada para assegurar a qualidade do modelo (MARQUESONE, 2016, p. 115). Outros aspectos dos resultados obtidos também são validados como acurácia, que avalia quão bem o modelo faz a correlação de um resultado com os atributos nos dados de entrada; confiabilidade, que avalia como o modelo é executado em diferentes conjuntos de dados; e utilidade do modelo, que avalia o quanto o modelo oferece informações significativas ao propósito da análise (MARQUESONE, 2016, p. 115).

Com a quantidade massiva de dados gerados atualmente, novos desafios foram surgindo à análise de dados e métodos estatísticos, algoritmos de aprendizado de máquina e técnicas de mineração de dados precisaram ser adaptados para suportar modelos de processamento paralelo e distribuídos (MARQUESONE, 2016, p. 118). Dessa forma, tecnologias de processamento de Big Data culminaram em formas inovadoras de obtenção de insights sobre dados e a junção das

técnicas tradicionais de Analytics com as tecnologias de Big Data originou o termo Big Data Analytics, que tem como foco a extração de informação a partir de uma avalanche de dados (MARQUESONE, 2016, p. 118).

As técnicas de Analytics podem ser divididas em quatro diferentes categorias, as diferenças estão no resultados que elas produzem e nos algoritmos e técnicas usados (MARQUESONE, 2016, p. 119):

1. Análise descritiva: é a forma mais básica de se obter indicadores e a mais usada pelas empresas, tem como objetivo principal sumarizar os dados passados das informações coletadas.
2. Análise diagnóstica: tem como foco respostas relacionadas a dados históricos mas busca identificar informações relacionadas aos fenômenos ocorridos na empresa.
3. Análise preditiva: considerada um divisor de águas entre os quatro tipos de análise porque além de somente compreender o passado, ela traz informações sobre os riscos e as oportunidades do futuro.
4. Análise prescritiva: sugere ações para as predições utilizando algoritmos programados com um mínimo de intervenção humana. O algoritmo deve ser capaz de se adaptar conforme os parâmetros, otimizando a predição.

2.4.4 Visualização de Dados

A visualização dos dados pode ser definida como a comunicação da informação utilizando representações gráficas para oferecer uma maior compreensão do que os dados representam (MARQUESONE, 2016, p. 131). A visualização de dados é essencial na transmissão adequada da mensagem (MARQUESONE, 2016, p. 131).

É importante definir desde o início o propósito da visualização, ela pode ser exploratória, na qual os dados precisam ser avaliados minuciosamente; ou pode ser explanatória, na qual os resultados obtidos são considerados corretos pelo analista dos dados e ele já compreende o que os dados estão representando e está pronto para demonstrar os resultados obtidos para um grupo de pessoas (MARQUESONE, 2016, p. 132). Na visualização exploratória não é necessário nenhum refinamento visual dos gráficos, muitas vezes quem visualiza é a própria pessoa que criou a análise (MARQUESONE, 2016, p. 132).

A visualização de dados melhora a comunicação, pois apresenta os resultados de modo efetivo; melhora o monitoramento do desempenho da empresa, pois permite resumir os resultados obtidos e também suporta o processo de tomada de decisão, pois permite revelar tendências e desvios das tendências (MARQUESONE, 2016, p. 135).

No processo de geração da interface visual é importante decidir qual gráfico é mais adequado e algumas perguntas podem apoiar esta escolha (MARQUESONE, 2016, p. 138):

- A comparação é entre valores? Nesse caso os gráficos de colunas, barras, áreas circulares, linhas e de dispersão devem ser considerados.
- É relevante mostrar como os dados estão distribuídos? Nesse caso os gráficos de dispersão, histogramas e gráficos de área 3D devem ser considerados.
- A composição dos dados é importante? Nesse caso os gráficos de pizza, área, barras empilhadas e colunas empilhadas devem ser considerados.
- É interessante que a audiência identifique tendências? Nesse caso os gráficos de linha, linha de dois eixos e coluna devem ser considerados.
- O relacionamento entre os dados deve ser evidenciado? Nesse caso os gráficos de bolha, linha ou dispersão devem ser considerados.

Existem outras formas além dos gráficos convencionais que podem ser consideradas, mapas podem ser usados para apresentar dados que tratam de informações geográficas, como nome de cidades, estados, países, códigos postais, latitude e longitude; uma nuvem de palavras pode ser usada para representar dados baseados em texto, representando com o tamanho da palavra a frequência de ocorrência de cada palavra da base de dados; um formato circular pode representar grafos direcionados, facilitando a compreensão sobre como cada elemento está relacionado com os demais (MARQUESONE, 2016, p. 138).

As etapas necessárias para gerar uma visualização são as seguintes, sendo que muitas delas fazem parte das etapas de armazenamento, processamento e análise de dados (MARQUESONE, 2016, p. 145):

1. Aquisição: etapa de captura dos dados para análise.
2. Estruturação: etapa de definição de uma estrutura padrão para os dados.
3. Filtragem: etapa de eliminação de dados incorretos, incompletos e não relevantes.

4. Mineração: parte da análise de dados, etapa técnica de extração de informações dos dados.
5. Representação: início da representação visual dos dados, etapa de geração do modelo visual básico dos dados.
6. Refinamento: aperfeiçoamento da visualização dos dados utilizando técnicas gráficas para torna-la mais eficiente.
7. Interação: melhoria da visualização dos dados inserindo funcionalidades que oferecem mais experiências ao leitor.

Com a alta demanda de dados que precisam ser analisados e visualizados atualmente é importante que existam ferramentas e técnicas que nos permitam automatizar o processo de visualização, pois além de facilitar o processo, agilizam a criação das visualizações, atendendo às necessidades de muitas soluções que precisam visualizar as informações em tempo real (MARQUESONE, 2016, p. 146).

As ferramentas atuais para visualização de dados oferecem recursos para atuar principalmente nas etapas de refinamento e interação de dados (MARQUESONE, 2016, p. 146). Além da possibilidade de gerar diferentes representações visuais com o mesmo conjunto de dados também é possível interagir com os dados, obtendo informações mais detalhadas e fazendo diferentes comparações entre os dados (MARQUESONE, 2016, p. 146).

3. METODOLOGIA

Nesta seção serão elencados os elementos e conceitos considerados relevantes na revisão bibliográfica do artigo científico *Big data analytics in logistics and supply chain management: Certain investigations for research nad applications* de WANG e colaboradores (2016, p. 98-110) publicado no *International Journal of Production Economics* para indicar possíveis métodos de Big Data aplicáveis ao problema de logística de falta de peças no ponto de uso da linha de produção de uma fábrica. As possíveis abordagens para o problema serão descritas nas considerações finais deste trabalho.

A exploração do Big Data aplicado à logística e cadeia de suprimentos ainda está em seus estágios iniciais e geralmente não há roteiros definidos para aplicar nas questões do cotidiano das indústrias. O problema de falta de peças no ponto de uso da linha de produção de uma fábrica envolve vários aspectos logísticos de toda a cadeia de suprimentos como qualidade do fornecedor, atrasos do fornecedor, capacidade do fornecedor, flutuações na demanda e gerenciamento de inventário ao mesmo tempo.

O artigo de WANG e colaboradores foi selecionado por contemplar os vários aspectos logísticos envolvidos no problema em questão, além de apresentar vários métodos possíveis e a sua relação com cada aspecto logístico. Como o objetivo deste trabalho limita-se somente aos conceitos teóricos e resultados práticos estão fora do escopo, o artigo apresentou-se como o mais adequado para a revisão bibliográfica.

A seguir, apresentam-se os elementos, conceitos e métodos relevantes e aplicáveis ao problema de logística de falta de peças no ponto de uso da linha de produção de uma fábrica extraídos do artigo de WANG e colaboradores (2016, p. 98-110).

3.1 BIG DATA APLICADO À LOGÍSTICA E CADEIA DE SUPRIMENTOS

Na fase estratégica do planejamento da cadeia de suprimentos, o Big Data e Analytics desempenham um papel vital ajudando as empresas a tomarem decisões estratégicas no fornecimento, no projeto da cadeia de suprimentos e no desenvolvimento e design do produto (WANG et al., 2016, p. 101). O Big Data e Analytics ajuda as lideranças a entender as mudanças

nas condições de marketing, identificar e avaliar os riscos e formular estratégias, melhorando a flexibilidade e a lucratividade da cadeia de suprimentos (WANG et al., 2016, p. 101).

Na fase de planejamento operacional da cadeia de suprimentos, o Big Data e Analytics auxilia a gerência na tomada de decisões operacionais da cadeia de suprimentos, no planejamento da demanda, em compras, na produção, no estoque e na logística (WANG et al., 2016, p. 101). Através da análise do desempenho da cadeia de suprimentos, o Big Data e Analytics é útil para melhorar a eficiência das operações das empresas, reduzir a variabilidade do processo e implementar as melhores estratégias nos níveis tático e operacional (WANG et al., 2016, p. 101). Essas melhorias são alcançadas através da conexão entre todas as operações dos processos da cadeia de suprimentos, dos fornecedores de matéria-prima aos clientes finais (WANG et al., 2016, p. 101).

A aplicação de Big Data e Analytics na logística e cadeia de suprimentos pode ser definida pelo termo Supply Chain Analytics (WANG et al., 2016, p. 99).

3.2 SUPPLY CHAIN ANALYTICS NA ESTRATÉGIA LOGÍSTICA E DA CADEIA DE SUPRIMENTOS

As aplicações de Supply Chain Analytics na estratégia logística e da cadeia de suprimentos são as seguintes (WANG et al., 2016, p. 101):

3.2.1 Fornecimento Estratégico

O fornecimento estratégico é colaborativo, com foco no gerenciamento do relacionamento com fornecedores, analisando os custos e adquirindo commodities e serviços de maneira econômica (WANG et al., 2016, p. 101). O fornecimento estratégico ajuda as empresas a otimizar o desempenho financeiro, minimizar o custo das operações e melhorar a performance de seus fornecedores (WANG et al., 2016, p. 101).

O Supply Chain Analytics pode ajudar a alcançar esses objetivos através da análise de perfil de gastos, processos de compra e demanda futura para garantir que as estratégias de fornecimento estejam alinhadas às metas e objetivos estratégicos da organização; e facilitando

o desenvolvimento de estratégias de terceirização, avaliando as tendências do mercado de suprimentos e insumos e economia dos fornecedores (WANG et al., 2016, p. 101).

Outro aspecto importante do fornecimento estratégico é a avaliação e seleção de fornecedores (WANG et al., 2016, p. 101). O Supply Chain Analytics permite que as organizações comparem as melhores práticas do setor, definam metas de desempenho e implementem métricas personalizadas (WANG et al., 2016, p. 101).

A falha de fornecedores no fornecimento de bens ou serviços no prazo de entrega, na quantidade ou com relação à qualidade tem um grande impacto na empresa (WANG et al., 2016, p. 101). Para os fornecedores que regularmente falham nas entregas, as organizações costumam manter uma quantidade maior de produtos em estoque, comprometendo o capital de giro (WANG et al., 2016, p. 101). Neste caso o Supply Chain Analytics é usado para prever interrupções no fornecimento através do mapeamento da cadeia de suprimentos e do relacionamento com os fornecedores para identificar as fontes de incerteza no fornecimento e gerenciar o relacionamento colaborativo contínuo com os fornecedores (WANG et al., 2016, p. 101). Evitar interrupções no fornecimento significa primeiramente selecionar cuidadosamente os fornecedores com boa reputação, trabalhar de forma colaborativa com eles para manter seu desempenho e monitorar eventos, como desastres naturais, para evitar ou mitigar rapidamente qualquer interrupção; e segundo, proteger a organização de perdas financeiras e ter a capacidade de trocar de fornecedor (WANG et al., 2016, p. 101).

3.2.2 Design da Cadeia de Suprimentos

O design da cadeia de suprimentos está relacionado à configuração física e à infraestrutura da cadeia, que envolve decisões sobre número, localização e tamanho das fábricas, centro de distribuição e armazéns que servem como estoque intermediário e pontos de expedição entre as fábricas e varejistas existentes (WANG et al., 2016, p. 101).

Os problemas de design da cadeia de suprimentos podem ser classificados em duas categorias em termos de informações disponíveis sobre demanda: aqueles com demanda conhecida e aqueles com flutuações ou incertezas na demanda e o Supply Chain Analytics auxilia nos problemas de design da cadeia de suprimentos em ambas situações (WANG et al., 2016, p. 101).

A quantidade de dados envolvidos no design da cadeia de suprimentos é enorme, contém a demanda agregada por produto em cada unidade de varejo, a capacidade da fábrica, os custos de remessa por unidade entre cada local e o custo fixo das operações em cada local potencial (WANG et al., 2016, p. 101). A maior parte da literatura tenta determinar a configuração da rede da cadeia de suprimentos minimizando a soma de vários componentes de custo como um objetivo único (WANG et al., 2016, p. 101).

3.2.3 Design e Desenvolvimento de Produtos

Para serem competitivas, capturar participação no mercado e melhorar a lucratividade as empresas precisam garantir que seus produtos sejam de alta qualidade e confiabilidade, fortalecer suas capacidades para produzir produtos diferenciados e de baixo custo (WANG et al., 2016, p. 101). Além de decisões críticas como custo versus qualidade, o design e o desenvolvimento de produtos influenciam na eficiência e pontualidade do lançamento de novos produtos geradores de receita (WANG et al., 2016, p. 103).

O Supply Chain Analytics ajuda as empresas a produzirem produtos de alta qualidade a preços competitivos, otimizando as trocas de produtos e aumentando a receita de vendas além de permitir que as empresas superem seus concorrentes aproveitando ao máximo as oportunidades de mercado (WANG et al., 2016, p. 103). Dados facilmente acessíveis permitem um rápido processo de design e desenvolvimento (WANG et al., 2016, p. 103). As empresas também podem monitorar e analisar as substâncias nos componentes fornecidos, usando dados em tempo real de processos e fornecedores internos, determinando com facilidade se os componentes dos fornecedores estão em conformidade com os padrões de qualidade, os regulamentos governamentais e os requisitos dos clientes e desempenho, evitando atrasos por problemas de não conformidade (WANG et al., 2016, p. 103). O Supply Chain Analytics avalia possíveis trocas, executando uma análise com cenários hipotéticos para avaliar os efeitos do design do produto e dos custos de desenvolvimento, alcançando o design mais econômico que atende aos critérios de qualidade e confiabilidade (WANG et al., 2016, p. 103).

3.3 SUPPLY CHAIN ANALYTICS NA OPERAÇÃO LOGÍSTICA E DA CADEIA DE SUPRIMENTOS

As aplicações de Supply Chain Analytics na logística e operações da cadeia de suprimentos são as seguintes (WANG et al., 2016, p. 103):

3.3.1 Planejamento da Demanda

Um aspecto fundamental no gerenciamento da cadeia de suprimentos é o gerenciamento de processos e operações para atender à demanda (WANG et al., 2016, p. 103). O planejamento da demanda não é apenas uma previsão, ele envolve o planejamento de vendas e operações e analisa diferentes segmentos de clientes em termos de canais, marcas e produtos até os níveis de estoque e desenvolve modelos usados para moldar a demanda e criar planos de receita, que é a base para o planejamento e previsão colaborativos com os principais parceiros da cadeia de suprimentos (WANG et al., 2016, p. 103).

Usando técnicas de previsões de demanda, vendas e otimização, o planejamento de vendas e operações fornece um recurso integrado de gerenciamento multifuncional para o gerenciamento de marketing, produção e estoque para gerenciar componentes operacionais e garantir os compromissos do cliente (WANG et al., 2016, p. 103).

3.3.2 Compras

Uma grande quantidade de dados em compras é gerada a partir de várias fontes ou aplicativos por meio de gastos, avaliações de desempenho de fornecedores e negociações internas ou externas, facilitando o uso de análises avançadas (WANG et al., 2016, p. 103). O Supply Chain Analytics fornece aos tomadores de decisão de compras análises consistentes e baseadas em dados para uma ampla variedade de decisões importantes e questões comerciais, como problemas de qualidade e disponibilidade de material (WANG et al., 2016, p. 103). Segundo a literatura, a aplicação do Supply Chain Analytics divide-se em gerenciamento de

riscos de suprimento e gerenciamento do desempenho de fornecedores (WANG et al., 2016, p. 103).

3.3.3 Produção

A análise da cadeia de suprimentos permite que os fabricantes entendam os diferentes custos de produção envolvidos e como eles influenciam os resultados finais (WANG et al., 2016, p. 103-104). A aplicação do Supply Chain Analytics pode fornecer informações úteis sobre os níveis de capacidade de produção e informar às lideranças se são necessárias melhorias para maximizar a produtividade (WANG et al., 2016, p. 104). Além disso, o Supply Chain Analytics pode apoiar as empresas a ajustar a produção para garantir que a combinação certa de recursos seja alocada nas linhas de produção certas (WANG et al., 2016, p. 104). O Supply Chain Analytics também é usado pelos analistas de produção para identificar o desperdício de material e técnicas e processos de fabricação que podem reduzir ou até eliminar esse desperdício (WANG et al., 2016, p. 103-104).

3.3.4 Inventário

As empresas estão continuamente acumulando conjuntos de dados gigantescos nos sistemas ERP (*Enterprise Resource Planning*) através da internet, dispositivos eletrônicos e aplicativos de software (WANG et al., 2016, p. 104). Os dados gerados nos sistemas ERP incluem dados históricos de demanda e previsão de demanda, prazos de reposição, o nível de serviço desejado, o custo de manutenção e o custo fixo de fazer um pedido de reposição (WANG et al., 2016, p. 104). Desafios como diferentes necessidades organizacionais e flutuações de oferta e demanda têm impacto nos níveis de estoque (WANG et al., 2016, p. 104). O Supply Chain Analytics pode ajudar as organizações a projetarem sistemas modernos de otimização de inventário necessários para lidar com os desafios mais complexos no gerenciamento (WANG et al., 2016, p. 104).

O uso de Supply Chain Analytics em sistemas de inventário gerenciado pelo fornecedor, conhecido por VMI (*Vendor Managed Inventory*), permite a coleta, o processamento e o

relatório dos dados do inventário, podendo informar as decisões relacionadas à melhoria do desempenho do inventário (WANG et al., 2016, p. 104). Supply Chain Analytics também pode ajudar na previsão precisa das necessidades de estoque e na resposta às mudanças nas demandas dos clientes (WANG et al., 2016, p. 104). Além disso, o Supply Chain Analytics ajuda a determinar os níveis de inventário apropriados e a obter uma visão holística dos níveis de estoque em toda a cadeia de suprimentos (WANG et al., 2016, p. 104).

3.3.5 Logística

Os dados de logística são gerados a partir de diferentes fontes nas redes de distribuição, previsões sobre a capacidade de fornecimento nas fábricas dos fornecedores, previsões de demanda em pontos de demanda e capacidade da rede (WANG et al., 2016, p. 104). A logística global gera uma enorme quantidade de dados e Big Data resultante de etiquetas RFID (*Radio-Frequency Identification*), dispositivos móveis e transações EDI (*Electronic Data Interchange*) que podem ser aproveitados no planejamento logístico (WANG et al., 2016, p. 104). Devido a interrupções no fornecimento e a incerteza na demanda, as ferramentas de análise preditiva são essenciais para projetar a flexibilidade da cadeia de suprimentos nas operações de logística (WANG et al., 2016, p. 104).

No planejamento logístico é essencial o roteamento da equipe e do equipamento mas diferentes variáveis como número de veículos, capacidade do veículo, restrições de duração de viagem e janelas de tempo de entrega e coleta, entre outros, complicam o planejamento das operações de transporte e distribuição na rede logística global (WANG et al., 2016, p. 104). Metodologias e técnicas de análise são usadas para otimizar roteamento de mercadorias, veículos e tripulação a fim de equilibrar os custos e as margens de transporte levando em consideração a manutenção e a segurança (WANG et al., 2016, p. 104).

Além de determinar se o Supply Chain Analytics será implementado na fase estratégica do planejamento da cadeia de suprimentos ou na fase de planejamento operacional da cadeia de suprimentos e quais os segmentos estarão envolvidos na aplicação do Big Data, é importante determinar a análise e a técnica apropriadas.

O método de análise e a técnica utilizados no projeto tem uma influência significativa na eficácia e eficiência do desempenho do Supply Chain Analytics (WANG et al., 2016, p.

100). A aplicação de Big Data e Analytics na logística e cadeia de suprimentos contempla as seguintes formas de análise e técnicas (WANG et al., 2016, p. 101).

3.4 TIPOS DE ANÁLISE DE SUPPLY CHAIN ANALYTICS

A análise de *Supply Chain Analytics* pode ser caracterizada em três categorias principais (WANG et al., 2016, p. 101):

- a) Análise descritiva: ocorre em períodos padronizados ou quando necessária utilizando técnicas como *online analytical processing* (OLAP) e visa identificar problemas e oportunidades nos processos e funções existentes.
- b) Análise preditiva: envolve o uso de algoritmos matemáticos e programação para descobrir padrões explicativos e preditivos nos dados. O objetivo desse tipo de análise é projetar com precisão o que acontecerá no futuro e fornecer razões para que isso aconteça.
- c) Análise prescritiva: envolve o uso de dados e algoritmos matemáticos para determinar e avaliar decisões alternativas que envolvem objetivos e requisitos caracterizados por alto volume e complexidade, com o objetivo de melhorar o desempenho dos negócios. A análise prescritiva inclui tomada de decisão, otimização e simulação com vários critérios.

As análises preditiva e prescritiva desempenham um papel vital para ajudar as empresas a tomarem decisões eficazes na direção estratégica da organização (WANG et al., 2016, p. 101). Elas podem ser aplicadas para resolver problemas relacionados às mudanças na cultura organizacional, decisões de fornecimento, configuração da cadeia de suprimentos e design e desenvolvimento de produtos ou serviços (WANG et al., 2016, p. 100).

A análise descritiva responde às perguntas relacionadas a questões como: “o que aconteceu?”, “o que está acontecendo?” (WANG et al., 2016, p. 100). Essas decisões também podem envolver análise de desempenho, empregando modelos, técnicas e ferramentas para ajudar as empresas a tomarem decisões rápidas, eficientes e eficazes (WANG et al., 2016, p. 100).

3.5 TÉCNICAS DE SUPPLY CHAIN ANALYTICS

As técnicas de Supply Chain Analytics podem ser caracterizadas da seguinte forma (WANG et al., 2016, p. 104-105):

1. **Análise estatística:** as análises estatísticas incluem dois tipos de técnicas, qualitativa e quantitativa. Métodos qualitativos, baseados julgamento subjetivo de consumidores ou especialistas, são apropriados quando dados anteriores não estão disponíveis. Abordagens quantitativas são usadas para previsões a partir de dados passados. As características do Big Data de velocidade, volume e variedade de dados trazem alguns desafios, o volume acumula ruído de dados, criando altos custos computacionais e instabilidade algorítmica; a alta variedade de dados requer a aplicação de diferentes técnicas e metodologias. Os métodos estatísticos tradicionais foram projetados para tamanhos moderados de amostra, não para dados massivos, por esse motivo procedimentos estatísticos eficazes têm recebido crescente atenção para explorar o seu uso no Big Data.
2. **Simulação:** o Big Data traz mais desafios para modelagem e simulação, mas ambas podem se beneficiar. O Supply Chain Analytics possibilita análise e processamento mais aprofundados e novos métodos para os problemas de simulação com volume de dados. Além disso, possibilita a modelagem e simulação de sistemas complexos, pois se concentra no relacionamento entre as operações da cadeia de suprimento, dessa forma, o Supply Chain Analytics pode agregar dados desintegrados de diferentes operações e alcançar a otimização global.
3. **Otimização:** o uso de técnicas de otimização no Supply Chain Analytics ajuda a melhorar a precisão da previsão da demanda e do planejamento da cadeia de suprimentos, mas seu uso é problemático em larga escala. A otimização do Big Data pode ser cara e instável e dificultar o êxito das técnicas tradicionais no Supply Chain Analytics. Por outro lado, a otimização ajuda a analisar sistemas dinâmicos altamente complexos com grandes volumes de dados, com múltiplas restrições e fatores obtendo informações que permitem aos gerentes tomar decisões apropriadas. Além disso, a otimização ajuda a analisar as medidas de desempenho da cadeia de suprimentos, como redução de custos e atendimento à demanda, entre outras. Outro benefício associado à otimização é sua flexibilidade, pois pode descobrir novas conexões de dados e transformá-las em insights e extrair valor comercial de grandes quantidades de dados.

Além de determinar o método de análise e a técnica que serão implementados no Supply Chain Analytics, é importante levar em conta mais alguns aspectos. O Supply Chain Analytics tem um papel extremamente importante nas operações logísticas e da cadeia de suprimentos, portanto deve estar integrado as outras atividades de negócios (WANG et al., 2016, p. 106). Para esse fim, as lideranças precisam entender o impacto das práticas de Supply Chain Analytics em toda a organização procurar adaptar essas práticas considerando o seu impacto na organização e em outros parceiros para que o Supply Chain Analytics traga contribuições nas análises holísticas de negócios (WANG et al., 2016, p. 106).

Muitos desafios relacionado à cultura e política organizacional são enfrentados na implementação do Supply Chain Analytics, por isso é importante que os recursos relevantes da organização participem do processo de elaboração das estratégias de negócio e da seleção das metodologias e técnicas apropriadas de acordo com as metas e objetivos de negócios (WANG et al., 2016, p. 106).

A estrutura de Supply Chain Analytics desenvolve-se com base na maturidade e nos objetivos da cadeia de suprimentos, que podem ser caracterizados em cinco níveis diferentes: Supply Chain Analytics funcional, Supply Chain Analytics baseado em processo, Supply Chain Analytics colaborativa, Supply Chain Analytics ágil e Supply Chain Analytics sustentável (WANG et al., 2016, p. 105). O Supply Chain Analytics baseada em processo e funcional estão relacionados às operações logísticas e da cadeia de suprimentos, enquanto o Supply Chain Analytics sustentável, ágil e colaborativo estão relacionados às estratégias logísticas e da cadeia de suprimentos.

3.6 ESTRUTURA DE MATURIDADE DE SUPPLY CHAIN ANALYTICS

3.6.1 Supply Chain Analytics Funcional

As cadeias de suprimentos organizadas funcionalmente e não totalmente integradas apresentam desafios relacionados à duplicação de atividades e processos e à falta de coordenação entre os parceiros da cadeia de suprimentos (WANG et al., 2016, p. 105). Para este cenário, o Supply Chain Analytics funcional é utilizado para analisar e resolver problemas relevantes para as funções da cadeia de suprimentos, melhorando a eficiência das operações de toda a cadeia (WANG et al., 2016, p. 105).

3.6.2 Supply Chain Analytics Baseado em Processo

Essa estrutura de Supply Chain Analytics se concentra principalmente em ajudar as empresas a alcançar eficácia operacional nos processos da cadeia de suprimentos utilizando técnicas e ferramentas para resolver os problemas relacionados com a integração de processos internos da cadeia de suprimentos na organização (WANG et al., 2016, p. 105). Nas cadeias de suprimentos organizadas de forma multifuncional, de fluxo contínuo entre as funções e que estão totalmente alinhadas com os objetivos do negócio o Supply Chain Analytics pode ser aplicado para ajudar na integração de processos (WANG et al., 2016, p. 105).

3.6.3 Supply Chain Analytics Colaborativo

A colaboração na cadeia de suprimentos permite e protege o compartilhamento de informações e conhecimentos, enquanto reduz a complexidade operacional através da padronização de processos e interfaces e da racionalização de produtos (WANG et al., 2016, p. 105-106). Esta técnica lida com situações no nível estratégico, nas quais uma organização colabora com parceiros de negócios externos para executar operações da cadeia de suprimentos (WANG et al., 2016, p. 105-106). Os dados externos de fornecedores ou parceiros são combinados com o processo de tomada de decisão para ajudar as empresas a tomar melhores decisões e alcançar a integração da cadeia de suprimentos (WANG et al., 2016, p. 105-106).

3.6.4 Supply Chain Analytics Ágil

As mudanças são constantes dentro das empresas, podem ocorrer alterações no produto ou serviço em relação ao design e ao desenvolvimento, aos requisitos do cliente ou ao mix de marketing que uma organização oferece e a capacidade de responder a essas mudanças é extremamente importante no ambiente incerto de hoje (WANG et al., 2016, p. 106). O Supply Chain Analytics possibilita o monitoramento em tempo real das cadeias de suprimentos contribuindo para a redução da incerteza, dando flexibilidade e velocidade no atendimento às

mudanças dos clientes e prazos curtos relacionados à transformação das cadeias de suprimento (WANG et al., 2016, p. 106). As empresas precisam desenvolver uma estrutura de Supply Chain Analytics ágil para lidar com altas incertezas nas operações da cadeia de suprimentos e obter vantagem competitiva (WANG et al., 2016, p. 106).

3.6.5 Supply Chain Analytics Sustentável

A literatura destaca a necessidade das organizações de gerenciar e colaborar estreitamente com fornecedores e clientes em questões de sustentabilidade para alcançar um melhor controle de riscos e sustentabilidade organizacional (WANG et al., 2016, p. 106). O Supply Chain Analytics pode ser definido como o uso da análise de negócios na coleta, análise e disseminação de dados relacionados à sustentabilidade, cujo objetivo é fornecer as informações apropriadas que podem ser usadas para a tomada de decisão eficaz e eficiente em questões de sustentabilidade (WANG et al., 2016, p. 106).

Para esse fim, o Supply Chain Analytics pode reunir e analisar dados relacionados à sustentabilidade de maneira eficiente e eficaz, apoiando assim uma variedade de necessidades de informações que incluem previsão, análise e avaliação de questões econômicas, ambientais e sociais (WANG et al., 2016, p. 106). O Supply Chain Analytics sustentável requer pensamento e alinhamento mais amplos entre os objetivos estratégicos e análise de Big Data, além de cultura organizacional de suporte e estudiosos também destacam a relação entre objetivos estratégicos, cultura, transparência e gerenciamento de riscos (WANG et al., 2016, p. 106). É importante que os líderes entendam o papel do Supply Chain Analytics como o agente que permite que as informações sejam transformados no formato necessário para a tomada de decisões estratégicas relacionadas à sustentabilidade e o comprometimento de todos os níveis gerenciais é essencial para as organizações e cadeias de suprimentos que adotam práticas sustentáveis (WANG et al., 2016, p. 106).

4. CONSIDERAÇÕES FINAIS

A sociedade atual gera cada vez mais dados numa velocidade cada vez maior e nos mais diferentes formatos e este fenômeno está presente também nas organizações. O dado pode ser considerado um ativo e extrair conhecimento útil deles é questão de sobrevivência e competitividade, caso contrário ele é apenas um item sem valor.

Nas empresas, o perfil da pessoa que trabalha com dados não é somente alguém com conhecimento em estatística e programação. O mercado tem indicado que o perfil desse profissional é alguém com algum conhecimento técnico mas principalmente com conhecimento em gestão de projetos, pois geralmente lidera uma equipe multidisciplinar com outras especialidades. O mercado também aponta uma tendência nas estruturas de TI, que estão atuando de forma mais integrada a outros setores, pois é fundamental que a gestão de dados esteja alinhada com a estratégia da empresa.

É importante que a organização compreenda e aplique as técnicas e métodos mais adequados ao seu contexto para cada etapa do ciclo de vida do dado desde a sua produção, passando pelo armazenamento, análise e visualização até o descarte do dado.

O Big Data, caracterizado de uma forma muito resumida como o fenômeno em que dados são produzidos em vários formatos e armazenados por uma grande quantidade de dispositivos e equipamentos, traz muitos aspectos positivos para as empresas. No que diz respeito a tomada de decisão, o Big Data suporta a interpretação de dados para a tomada de decisão em tempo real, e não no dia seguinte, quando geralmente acontece a análise dos resultados do dia anterior nas empresas. O Big Data traz oportunidades para as organizações, tanto na vantagem competitiva por analisar melhor a sua eficiência e performance interna, tanto pela criação de produtos e serviços orientados pelos dados através do melhor entendimento das necessidades de seus clientes. Mas o Big Data também traz alguns desafios para as organizações, o acesso, armazenamento, busca e compartilhamento dos dados internamente são muitas vezes mais difíceis do que se trabalhar com os dados em si. Outra dificuldade enfrentada é eliminar a parte operacional e manual da informação quando falta governança de dados e integração de sistemas na organização. Muitas vezes a implementação do Big Data pode esbarrar na cultura organizacional, pois o dado é tratado de uma forma holística além das fronteiras dos setores e departamentos.

O Supply Chain Analytics, definição de Big Data Analytics aplicado à logística e cadeia de suprimentos, tem ganhado importância devido à complexidade e relevância desses setores

na performance das organizações, mas ainda se encontra em seus estágios iniciais. Não existem modelos prontos para as questões do cotidiano das empresas, mas sim diferentes abordagens que podem ser aplicadas para diferentes conceitos, objetivos e características no contexto de cada organização.

A questão da falta de peças no ponto de uso da linha de produção de uma fábrica está relacionada ao planejamento operacional da logística, pois envolve decisões mais operacionais do que estratégicas. Entre as possíveis soluções, pois outras abordagens também se aplicam ao contexto, uma possível seria a aplicação de Big Data no inventário. O Supply Chain Analytics pode ajudar na otimização do inventário, determinando os níveis de inventário apropriados para lidar com os desafios mais complexos no gerenciamento e obter uma visão holística dos níveis de estoque em toda a cadeia. Outra possível aplicação é na logística em si, pois os dados gerados ao longo de toda a cadeia podem ser aproveitados para responder melhor às variações no programa de produção e no fornecimento e o Big Data também auxilia a projetar a flexibilidade nas operações logísticas devido à interrupções no fornecimento e variabilidade na demanda.

Uma abordagem adequada para desenvolver o Supply Chain Analytics no problema da falta de peças é uma estrutura de Big Data colaborativa, com o compartilhamento e proteção de dados em toda a cadeia, desde os diferentes fornecedores de peças até a fábrica. Esses dados combinados ajudam na tomada de decisão e na integração de toda a cadeia. A aplicação de análise preditiva, usando algoritmos para descobrir padrões explicativos e preditivos nos dados ajudaria a entender se há padrões nas peças faltantes relacionados a tipos de peças, fornecedores, sazonalidade e ajudaria também a prever esses eventos de forma que a fábrica pode tomar ações proativas e mais baratas antecipadamente, como negociação com fornecedores, alteração da sequência de produção e não ações mais caras como fretes extras e parada de linha. O uso da técnica de otimização é bastante adequado para entender melhor as relações e conexões nos dados, trazendo percepções que podem agregar valor e trazer melhorias neste processo de suprimento de peças.

As possíveis abordagens e técnicas de Supply Chain Analytics citadas acima podem contribuir muito na análise holística da questão da falta de peças no ponto de uso da linha de produção de uma fábrica. Para uma análise mais aprofundada do problema é necessária a aplicação prática do Big Data nos dados e avaliação dos resultados, que estão fora do escopo deste trabalho.

REFERÊNCIAS

AMARAL, F. **Introdução à Ciência de Dados: mineração de dados e Big Data**. 1ª ed. Rio de Janeiro-RJ: Alta Books, 2016.

BATTISTELLI, J. **Como se tornar um CTO?**. Disponível em: <<https://blog.mastertech.com.br/negocios/como-se-tornar-um-cto/>>. Acesso em: 25 out. 2019.

ILHÉU, T. **O que faz um cientista de dados e como se tornar um**. Disponível em: <<https://guiadoestudante.abril.com.br/orientacao-profissional/o-que-faz-um-cientista-de-dado-s-e-como-se-tornar-um/>>. Acesso em: 06 jun. 2020.

LAURETTI, C. M. **Repensando TI na estrutura organizacional**. Disponível em: <<https://www.tiespecialistas.com.br/repensando-ti-na-estrutura-organizacional/>>. Acesso em: 25 out. 2019.

MARQUESONE, R. **Big Data – Técnicas e tecnologias para extração de valor dos dados**. Casa do Código, 2016.

REHMAN, M. H.; CHANG, V.; BATOOL, A.; WAH, T. Big data reduction framework for value creation in sustainable enterprises. **International Journal of Information Management**, p. 917–928, 2016.

RESUMO SOBRE INDÚSTRIA 4.0: entenda rapidamente os conceitos e benefícios. Disponível em: <<https://www.pollux.com.br/blog/>>. Acesso em: 08 mai. 2019.

SENAI. SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL. **Curso EAD Desvendando a Indústria 4.0**. Disponível em: <<https://senaiead.timetoknow.com/pr/login>>. Acesso em: 08 mai. 2019.

SILVEIRA, M.; MARCOLIN, C. B.; FREITAS, H. M. R. Uso Corporativo do Big Data. **Revista de Gestão e Projetos - GeP**, v. 6, n. 3, set./dez 2015. Disponível em: <<http://www.revistagep.org/ojs/index.php/gep/article/view/369/pdf>>. Acesso em: 7 out. 2019.

VIEIRA, V.; PEDROSA, I.; SOARES, B. **Big Data & Analytics: uma abordagem usando entrevistas a especialistas em Auditoria**. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/7976069>>. Acesso em: 7 out. 2019.

WANG, G; GUNASEKARAN, A.; NGAI, E.; PAPADOPOULOS, T. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. **International Journal of Production Economics**, p. 98–110, 2016.