

BIG DATA PARA ARMAZENAMENTO E PROCESSAMENTO DE DADOS MASSIVOS PROVENIENTES DA INTERNET INDUSTRIAL
Dr. Eduardo Manuel de Freitas Jorge¹, Msc. Fábio Britto de Carvalho Almeida¹ e Regina Pereira Silva¹

¹SENAI/CIMATEC, E-mail: eduardo.jorge@fieb.org.br; falmeida@fieb.org.br; reginabiode@gmail.com

RESUMO

As indústrias de manufatura contam hoje com os sistemas de controle e aquisição de dados denominados de MES (Manufacturing Execution System). Soluções deste tipo são softwares que adquirem dados discretos de um processo, a partir de diversas fontes, os armazenam em um banco de dados e os disponibilizam através de relatórios analíticos. Proporcionando uma visão unificada de todo o processo e ajudando assim que as aplicações cliente tenham um controle mais preciso das etapas da produção. O objetivo deste artigo é apresentar uma solução de Big Data para a construção de uma aplicação analítica de MES na Indústria, detalhando softwares que atendam os requisitos não funcionais de Replicação, Sharding e Schema Design. A solução é representada através de uma arquitetura que possui na sua base os softwares MongoDB e Pentaho. O cenário de uso para teste de a arquitetura proposta objetiva unificar informações de Máquinas Injetoras volumosas, distribuídas e variadas do chão de fábrica em um contexto de uma Indústria com inúmeras filiais. A unificação das informações visa proporcionar ao gestor, indicadores de produção por tempo de funcionamento destas máquinas, com o monitoramento de Andon's. As tecnologias utilizadas para o desenvolvimento da solução proposta nesse trabalho conseguiram atender de forma positiva a arquitetura proposta e aos requisitos não funcionais de performance, desempenho e disponibilidade das informações, seja em aplicações da Web das Coisas e Internet das Coisas, como também no contexto da Internet Industrial.

Palavras-chaves: Tecnologia da Informação, Sistema de Execução da Manufatura, MongoDB

ABSTRACT

The manufacturing industries today rely on the control systems and data acquisition called MES (Manufacturing Execution System). Such solutions are software that acquire data from a discrete process, from various sources, store them in a database and make them available through analytical reports. Providing a unified view of the entire process and helping so that client applications have more precise control of the stages of production..The objective of this article is to present a solution of Big Data for building an analitic application of MES in Industries, detailing softwares that do not attend to non funcional requirement of Replication,

Sharding e Schema Design. This solution has its architecture based on MongoDB and Pentaho. The scenario for validating this architecture is intended to unify information from large, distributed and different Injector Machines. Those information will be used to generate production indicators by working time of those machines. The technologies used to develop the solution proposed in this work were able to answer positively to proposed architecture and nonfunctional requirements of performance, performance and availability of the information, whether in Web applications of Things and Internet of Things, as well as in the context of Industrial Internet.

Palavras-chaves: *Information Technology, Sistema de Execução da Manufatura, MongoDB*

1. INTRODUÇÃO

Uma das áreas de pesquisa de destaque na computação nestes últimos anos é a temática Big Data e *Cloud Computing* (armazenamento em nuvem) [1]. As evoluções tecnológicas na capacidade de armazenamento de dados, no poder de processamento computacional e na redução de custos de sensores eletrônicos, somado a alta conectividade têm propiciado surgimento de soluções de Big Data, pode-se citar exemplos de uso no Facebook[2], Google[3], Ebay[4], Amazon[5] e IBM Watson[6]. O contexto deste artigo está em analisar como estes avanços também se refletem no setor industrial, conhecido pelo termo de Internet Industrial.

Ambientes da Internet Industrial são compostos por máquinas inteligentes com inúmeros sensores eletrônicos. Estes sensores estão conectados com a internet por meio de protocolos específicos gerando uma grande quantidade de dados que podem ser analisados em tempo real. A coleta destes dados objetiva o desenvolvimento de soluções analíticas, buscando eficiência nos processos industriais. Por exemplo, com o tratamento dessas informações podem ser realizadas manutenções preditivas antecipando possíveis problemas de operação. Segundo dados de uma pesquisa realizada pelo IDC Brasil e que se encontra no site da IBM [7], apontam que em 2011 foram produzidos 1,7 zettabytes de dados, em 2012 2,7 zettabytes e que em 2015, ano desta pesquisa, pode chegar a 8 zettabytes de dados, sendo que muitos destes dados serão produzidos pela Internet Industrial.

As indústrias de manufatura contam hoje com os sistemas de controle e aquisição de dados denominados de MES (*Manufacturing Execution System*) [8]. Soluções deste tipo são softwares que adquirem dados discretos de um processo, a partir de diversas fontes, os armazenam em um banco de dados e os disponibilizam através de relatórios analíticos. Proporcionando uma visão unificada de todo o processo e ajudando assim que as aplicações cliente tenham um controle mais preciso das etapas da produção. Já existem soluções de manutenção preditiva na Indústria, porém estruturadas com Banco de Dados Relacional

OLTP (*Online Transaction Processing ou Processamento*) ou OLAP (*On-line Analytical Processing*). Entretanto, será que uma estrutura de Big Data não seria a melhor opção para aplicações de Internet Industrial, que demandam dados volumosos e distribuídos. Outro ponto é que soluções deste tipo, não precisam assegurar todas as características dos Bancos de Dados convencionais como as propriedades ACID (*Atomicidade, Consistência, Isolamento, Durabilidade*). Em outro contexto, mas, com a problemática similar, o Facebook opta por uso de Big Data, o vice-presidente da engenharia de infraestrutura da empresa, Jay Parikh [9] ressalta que são mais de 500 terabytes de dados processados por dia, e mais de 2,7 bilhões de likes diários e 300.000.000 uploads de fotos.

O objetivo deste artigo é apresentar uma solução de Big Data para a construção de uma aplicação analítica de MES na Indústria, detalhando os seus componentes de software que atendam aos requisitos não funcionais de Replicação, *Sharding e Schema Design*. A solução é representada através de uma arquitetura que possui na sua base os softwares MongoDB [10] e Pentaho [11]. O cenário de uso para teste de a arquitetura proposta objetiva unificar informações de máquinas Injetoras [12] volumosas, distribuídas e variadas do chão de fábrica em um contexto de uma Indústria com inúmeras filiais. A unificação das informações visa proporcionar ao gestor, indicadores de produção por tempo de funcionamento destas máquinas, com o monitoramento de *Andon's* [13].

Para a construção desta solução a primeira etapa foi à seleção das ferramentas que atendessem os requisitos de Replicação, *Sharding, Schema Design*, consultas OLAP, e BA (*Business Analytics*). Após esta etapa foi realizada a criação de um protótipo ainda fora do contexto MES para testar de forma preliminar a arquitetura. Por fim, para analisar a efetividade da solução de MES, criou-se um protótipo com dados em um ODS (*Operational Data Store*) que agruparam dados das máquinas Injetoras e por fim montou-se a base de dados Big Data e foram desenvolvidas as consultas OLAP.

O artigo está organizado da seguinte forma. Na Seção 2, abordar-se os conceitos fundamentais de cada tecnologia que compõe esta solução. A Seção 3 apresenta o trabalho desenvolvido. E por fim a seção 4 apresenta a conclusão do presente artigo.

2. BIG DATA, MONGO DB E PENTAHO

À medida que à Internet das Coisas, a Web das coisas e a Internet Industrial avançam e são produzidos dados aceleradamente, buscam-se nas tecnologias de *Big Data Analytics* entender todas, essas informações. Nesta seção detalham-se brevemente os conceitos de Big Data e as duas ferramentas base da arquitetura desta pesquisa: MongoDB e Pentaho.

2.1. BIG DATA

Big Data refere-se às informações que estão além da capacidade das ferramentas de bancos de dados típicos de capturar, armazenar, tratar e analisar [14]. E ainda Big Data refere-se a um grande volume de dados com enorme velocidade [12]. Ou seja, é a transformação de uma enorme quantidade de dados oriundos de diversas localidades, em informações em tempo real. Big Data baseia-se no 5V's, que são volume, variedade, velocidade, veracidade e valor [15].

Volume: crescimento exponencial, 3X mais dados produzidos, que armazenados, dados estes, produzidos por seres humanos e Web das Coisas tudo que visualizamos através dos browsers pela Web (redes sócias, e-commerce), Internet das Coisas (IoT - *Internet of Things*), objetos físicos conectados a internet por meio de RFID's (*Radio-Frequency Identification*) (relógios, carros, Google Glass, tênis) e Internet Industrial máquinas equipadas com dispositivos embarcados e sistemas inteligentes, também, conectados à rede (turbinas de gás e petróleo, motores à jato, equipamentos hospitalares, máquinas injetoras);

Variabilidade: os tipos de dados suportados por uma estrutura de Big Data são variados, mas atualmente têm-se três classificações de estruturas básicas de dados, são elas:

Dados estruturados: dados com comprimento e tipo definido. Ex: números, datas, strings;

Dados não estruturados: dados sem um formato específico. Ex: imagens de satélite, dados científicos, fotos, vídeos, documento, mídia social, web;

Dados semiestruturados: que não são aderentes a um esquema fixo, mas podem ser autodefendidos. Ex: JSON, EDI, XML.

Velocidade: criação e captura de dados brutos a taxas muito rápidas, arquivos em lote, obtidos de bancos de dados, ou dados gerados em tempo real (em *streaming*), sendo preciso agir em tempo real;

Veracidade: nem todos os dados gerados possuem um valor ao negócio, podem ser possíveis dados incompletos e desatualizados, sendo assim é necessário desvincular o que é relevante e rico de conteúdo em meio ao iceberg de informações e dados;

Valor: os dados quando processados e analisados trazem consigo informações valiosas que podem transformar a forma como os dados são vistos, agregando novos valores e insights.

As empresas que agregam uma solução de Big Data ao seu negócio podem ganhar em competitividade com seus concorrentes. Seja conhecendo melhor o seu cliente ou oferecendo melhores serviços aos mesmos e ainda podem melhorar na gestão e manutenção de seus produtos. Um dos bancos de dados que consegue atender positivamente a estrutura de Big Data é o MongoDB, baseado no paradigma não relacional (NoSQL - *Not Only SQL*), este suporta variados tipos de dados, um das cinco características de Big Data, por exemplo.

2.2. MONGODB

MongoDB é um Banco de Dados Open-Source, NoSQL (não relacional) e orientados a Documentos, que foi desenvolvido na Linguagem C++. Um Documento em MongoDB é uma estrutura de dados composta por Chave e Valor. As principais características deste Banco não relacional são: alta performance, alta disponibilidade, escalabilidade automática e sem modelagem (*Schema Design*), além de recursos como busca textual (*Full Text Search*), Aggregation Framework, Índices Espaciais (geográficos), *Sharding* e *Replica Set* (replicação de dados), No site oficial do mesmo é possível obter uma documentação atualizada que enumera todos os recursos da solução. A seguir detalha-se os componentes do MongoDB relevantes para esta pesquisa. A seguir detalha-se os componentes do MongoDB relevantes para esta pesquisa.

Schema Design em MongoDB são esquemas dinâmicos e implícitos, não existindo a necessidade de criar toda estrutura do banco de dados (chamado de *Schema*), a criação de coleções é realizada dinamicamente, existem documentos de diferentes tipos na mesma

coleção, os campos podem ser alterados em cada documento a qualquer momento sem afetar a estrutura da coleção e estas não precisam ter um conjunto idêntico de campos.

MongoDB, também, proporciona replicação de dados, que é o processo de sincronização das informações entre vários servidores, existindo redundância dos mesmos. Com isso tem-se um aumento da disponibilidade, já que existem várias cópias dos dados em diferentes servidores, assim os mesmos são protegidos em caso de perda de um servidor. Há recuperação de falha de hardware e interrupções de serviço. Com cópias adicionais dos dados, pode-se dedicar um para recuperação de desastres, elaboração de relatórios, ou backup. O MongoDB trabalha com o conceito de mestre/escravo, no qual uma réplica primária recebe todas as operações de escrita e as secundárias replicam essas operações mantendo assim um mesmo conjunto de dados idênticos. Sendo que, o primário (mestre) é quem recebe as operações de escrita e os secundários (escravos) replicam aceitando apenas operações de leitura diretamente, sem a necessidade de passar pelo mestre.



5

Figura 1. Exemplo de Arquitetura do MongoDB para Replicação de Dados¹.

O *Mongo Shell* é um shell interativo que é baseado em comandos da Linguagem *JavaScript* oferecendo flexibilidade no gerenciamento do banco de dados, e possibilitando a execução de operações administrativas.

Os documentos em MongoDB são no formato JSON (ver um exemplo na figura 2), que é um acrônimo para “*JavaScript Object Notation*”, é um formato leve para intercâmbio de dados computacionais. JSON é um subconjunto da notação de objeto de *JavaScript*, mas seu uso não requer o uso da linguagem exclusivamente. Documentos JSON suportam: números, strings e valores booleanos, bem como arrays e hashes e outros JSON’s aninhados, e o MongoDB usa documentos JSON para armazenar registros na sua base de dados com esse formato.

MongoDB representa documentos JSON em formato binário codificado chamado BSON (*Binary JSON*) nos bastidores. BSON estende o modelo de JSON para fornecer os tipos de

⁵ Disponível em: <http://www.engineering.ons.ee/2013/10/onsee-implementa-alta-disponibilidade-em-seu-banco-de-dados/>. Acesso em 15 fevereiro 2015.

dados adicionais e para ser eficiente para codificação e decodificação em diferentes idiomas. BSON é uma versão binária do JSON, guardando os dados através do uso de pares chave/valor.

A escolha pela utilização do MongoDB, dentre outros bancos NoSQL e com capacidade de suportar uma solução de Big Data, e que ele possui características como citadas acima e outras funcionalidades: escalabilidade, performance, não está preso a uma modelagem definitiva, processamento em tempo real, dados com crescimento exponencial, drivers para quase todas as linguagens de programação, é possível utilizar ao mesmo tempo um banco relacional e banco NoSQL como MongoDB. Empresas como FourSquare, Easytaxi, Globo.com e EA Games utilizam este banco em suas aplicações [10].

2.2. PENTAHO

O funcionamento de uma plataforma de BI pode ser exemplificado com o Pentaho. Os conceitos presentes na arquitetura do mesmo tais como: OLAP, Data Mart e Data Warehouse são bem presentes na literatura [1].

Segundo Howard Dresner, BI é um termo genérico, usado para descrever um conjunto de conceitos e métodos para aperfeiçoar a tomada de decisões de negócios utilizando sistemas de suporte baseados em fatos.

O valor real dos sistemas de BI está na sua utilização para apoiar as organizações para tomar decisões bem informadas que levará a uma maior rentabilidade, custos reduzidos, eficiência, crescimento da quota de mercado [16].

O termo OLAP é definido como uma interface com o usuário que proporciona a capacidade de ter ideias sobre os dados, permitindo analisá-los profundamente em diversos ângulos. Suas principais funções são: visualização multidimensional dos dados, exploração, rotação. O *Data Warehouse* armazena as informações de forma eficiente, o OLAP deve recuperá-las com a mesma eficiência, porém com mais rapidez. As duas tecnologias se complementam, ao ponto que um DW para ser bem sucedido, já na sua concepção, deve levar em consideração o que se deseja apresentar na interface OLAP. Ou seja, OLAP é uma interface com o usuário e não uma forma de armazenamento de dados, porém se utiliza do armazenamento (*Data Warehouse*) para poder apresentar as informações. E *Data Marts* são subconjuntos de *Data Warehouse*, ou seja, também são repositórios de dados [1].

Uma das ferramentas mais utilizadas no mercado para a análise de dados é o Pentaho, a versão utilizada para este projeto é a Pentaho Community, que é *Open Source* (código-aberto).

Pentaho é um poderoso *Business Intelligence Suite* oferece recursos como: relatórios, tabelas dinâmicas de OLAP, Dashboards entre outros [16]. Ou seja, o conjunto de ferramentas integradas do Pentaho proporciona um ambiente de desenvolvimento para a criação de soluções BI, com alto desempenho e disponibilidade para seus usuários para a tomada de decisões. Pentaho é desenvolvido em Java e necessita da JVM (*Java Virtual Machine*) para rodar suas aplicações, e é *Open Source*.

Para entender melhor todo o conjunto de ferramentas disponibilizadas pelo Pentaho, será descrito a sua arquitetura para melhor compreender o seu funcionamento. Como foi dito anteriormente, o Pentaho é composto por um conjunto de ferramentas, e algumas delas constam na arquitetura. Agora, iremos descrever sobre cada etapa desta arquitetura:

Camada Fonte de Dados: Pentaho suporta dados provenientes de diversas bases de dados como PostgreSQL, MySQL, Oracle, arquivos de texto e/ou arquivos com extensão .xml, .xlsx e Big Data como Hadoop, MongoDB, Cloudera, entre outros;

Camada Dados e Integração de Aplicações ETL: Para realizar transformações em Pentaho, seja para extrair dados de uma base de dados ou salvar, utilizamos a plataforma do PDI (*Pentaho Data Integration*) que contém a ferramenta de ETL (*Extract Transform Load*) Kettle. ETL é o processo de extração de dados oriundos de fontes externas como as citadas no tópico anterior, e que visa à transformação, a extração ou importação dos dados para ambientes como Data Warehouse;

Camada de Business Intelligence: esta camada tem uma biblioteca principal, o JPivot, que é responsável por customizar tabelas e gráficos OLAP. Tendo opções de *drill down e slice-and-dice* e usa a ferramenta Mondrian, por exemplo, como engine de consultas OLAP;

Camada de Apresentação: na camada de apresentação há um conjunto de ferramentas (o CTools) que o Pentaho disponibiliza para a criação de *dashboards* dinâmicos, e há também a possibilidade de integração com o *Bootstrap*, estilizador de componentes CSS. O principal ambiente de desenvolvimento é o CDE (*Comunity Dashboard Editor*) no qual pode-se conectar aos *Data Warehouse* e/ ou *Data Marts* para a extração dos dados, como também, extrai-los diretamente das chamadas *Kettles Queries*, que são as transformações feitas no PDI.

A plataforma Pentaho, apesar de toda essa divisão em camadas, consegue integrar desde a base de dados até a interface online com o usuário, dispondo as informações de forma dinâmica e acessível, sendo de fácil entendimento. Existe uma comunidade ativa que utiliza e discute sobre todas as funcionalidades disponibilizadas pela plataforma e que está em constante atualização, mantendo também documentação de como integrar com diversos bancos, sejam eles relacionais ou não, inclusive com o MongoDB.

3. BIG DATA PARA ARMAZENAMENTO E PROCESSAMENTO DE DADOS MASSIVOS PROVENIENTES DA INTERNET INDUSTRIAL

Como mencionado na introdução este artigo descreve uma arquitetura para o projeto coletar dados gerados pelas Máquinas Injetoras distribuídas na matriz e filial, através dos protocolos de conexão com a internet. Objetiva-se, armazená-los em coleções no MongoDB, e em sequência dispor essas informações em painéis (*dashboards*), feitos com *Pentaho BI Server*. Com isso, o processo foi dividido em três etapas importantes:

Análise dos recursos do banco MongoDB, seu funcionamento, aplicabilidade dos seus principais conceitos de NoSQL, a criação de uma base de dados, ou seja, uma coleção ou coleções, realizações de consultas, inserções, e exclusões, replicação de dados;

Análise da ferramenta Pentaho e integração da mesma com o banco, realizando transformações e as disponibilizando nos painéis gráficos para consultas para tomada de decisões pelos seus usuários, disponíveis via internet;

A coleta dos dados das Injetoras, com a criação de um ODS (*Operational Data Store*) feito na linguagem Java capaz de integrar todos os dados e inseri-los para o banco para armazená-los.

Duas das etapas acima foram realizadas como forma de protótipos, a exemplo, foi criada uma base de dados no ManogoDB e deste foram realizadas consultas, (*queries*), para fins de testes, assim como etapa 2, na qual foi realizada a integração dos dados com o Pentaho. Algumas das especificações técnicas sobre a ferramenta já foram descritas na seção 2.3 bem como o seu funcionamento com base na sua arquitetura. O Pentaho já vem com drivers para a integração com diversas bases de dados e entre elas o MongoDB. Foi realizada uma transformação na ferramenta, Kettle, que faz parte do conjunto de ferramentas do Pentaho, que seria a extração de dados. Ou seja, entrada ou saída de diversos formatos dados de uma base transformadas em outros tipos de arquivos de dados como arquivos de texto (csv, txt), planilhas (xml, xls), ou importadas para *Data Marts* e/ou *Data Warehouse*, integrando sistemas de empresas distribuídas. Posteriormente foi feita a criação dos dashboards realizados na interface User Console com CDE (*Community Dashboard Edition*) com base nas informações extraídas na transformação feita no Kettle. Restando agora à etapa C para conclusão desta solução em Big Data com base na aplicabilidade dos conceitos de *Replica Set*, *Sharding* e *Schema Desing*.

3.1. ARQUITETURA

A figura a seguir retrata da arquitetura deste projeto. Sendo uma visão geral de todo o projeto, desde a coleta de dados até a ilustração de painéis gráficos das informações.

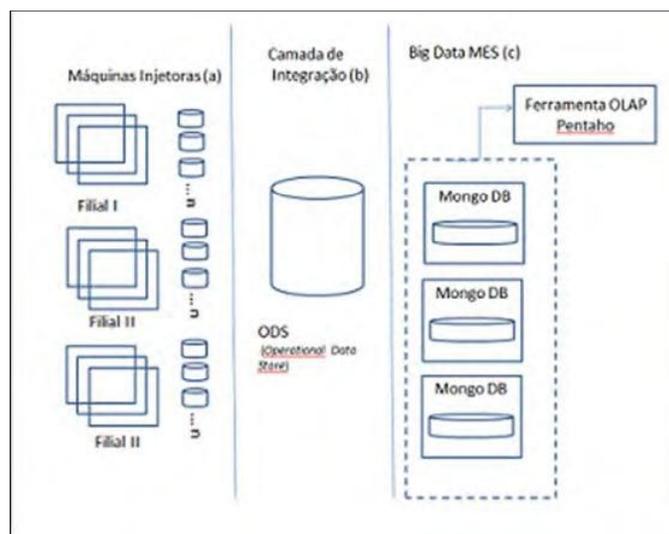


Figura 2. Modelo da Arquitetura do Projeto.

Agora, iremos descrever sobre cada etapa desta arquitetura:

Camada A: esta camada representa a distribuição de máquinas Injetoras por filial e tendo os seus dados armazenados em um conjunto n de banco dados igualmente distribuído;

Camada B: esta representa a integração dos dados em uma única base, centralizando-os, para isso, foi realizada a implementação de um ODS, responsável por integraliza-los;

Camada C: e na última camada a representação de replicação com o MongoDB e a integração deste com o Pentaho, ferramenta pela qual os dados serão apresentados graficamente.

4. CONCLUSÃO

As tecnologias utilizadas para o desenvolvimento desta solução em Big Data conseguem atender de forma positiva a arquitetura proposta e aos requisitos não funcionais de performance, desempenho e disponibilidade das informações, seja em aplicações da Web das Coisas e Internet das Coisas, como também no contexto da Internet Industrial.

Observa-se que apesar desta solução estar no contexto industrial, ela poderá ser reutilizado para outras necessidades similares. Além disso, os elementos de conexão, apresentados na solução MES, entre os software MongoDB e Pentaho podem ser utilizados por outros desenvolvedores interessados na temática Big Data. Já no contexto industrial, a relevância desta pesquisa é apoiar a criação de soluções para a Internet Industrial.

REFERÊNCIAS

¹Kimball, Ross. The Data Warehouse Toolkit: the Complete Guide to Dimensional Modeling. Second Edition, Wiley, 2002.

²Wikipédia. Conceito de Cloud Computing. Disponível em : < http://pt.wikipedia.org/wiki/Computa%C3%A7%C3%A3o_em_nuvem>. Acesso em 15 de fevereiro 2015.

³Facebook. Informações sobre o uso de Big Data Analytics pelo Facebook. Disponível em: <http://siliconangle.com/blog/2012/02/21/facebooks-big-data-story-where-behavior-meets-business/> >. Acesso em 15 de fevereiro 2015.

⁴Google. Platarform Google Analytics. Disponível em: < <http://www.google.com/analytics/>>. Acesso em 15 de fevereiro 2015.

⁵Ebay. Como grande poderes dados a viagem do cliente eBay. Disponível em: <http://www.computerweekly.com/news/2240219736/Case-Study-How-big-data-powers-the-eBay-customer-journey> . Acesso em 15 de fevereiro 2015.

⁶Site da Datafloq. How Amazon Is Leveraging Big Data. Disponível em: <<https://datafloq.com/read/amazon-leveraging-big-data/517>>. Acesso em 15 de fevereiro 2015.

⁷IBM Watson Foundations. A Big Data and Analytics Platform. Disponível em: <<http://www.ibmbigdatahub.com/ibm-watson-foundations>>. Acesso em 16 fevereiro 2015.

⁸Site da IBM. Pesquisa realizada pelo IDC Brasil. Disponível em: <http://www.ibm.com/midmarket/br/pt/infografico_bigdata.html>. Acesso em 10 novembro 2015.

⁹Da Costa, D. S.; Lima, J. J. L.; Santos, V. L. R. Tecnologias da Informação no Chão de Fábrica. Bahia: Salvador, 2011. 36 p.

¹⁰MongoDB. The MongoDB 2.6 Manual. Disponível em: <<http://docs.mongodb.org/manual/>>. Acesso em: 12 agosto 2014.

¹¹Site do Pentaho. Tutorias, Downloads, Fóruns da Comunidade Pentaho. Disponível em: <<http://www.pentaho.com>>. Acesso em 19 agosto 2014.

¹²Ribeiro, L. dos S. Evolução Tecnológica e Automoção das Máquinas Injetoras. São Paulo: Zona Leste, 2009. 78p.

¹³Termos de Manufaturas. ANDON - Definição. Disponível em: <<http://www.manufacturingterms.com/Andon>>. Acesso em 8 Janeiro 2015.

¹⁴Manyika, J. et al. Big Data: the next frontier for innovation, competition and productivity. McKinsley Global Institute, 2011.

¹⁵Datastorm. 5 Vs: a estrutura do Big Data. Disponível em: <<http://datastorm.com.br/blog/artigos/5-vs-a-estrutura-do-big-data/>>. Acesso em 25 Fevereiro 2015.

¹⁶Bouman, R.; Dongen J. Von (2009). Pentaho Soluções. Business Intelligence e Data Warehousing with Pentaho and MySQL. (Wiley Publishing, Inc.10475 Boulevard Crosspoint Indianapolis, IN 46256).