Anuário da Produção de Iniciação Científica Discente

Vol. 0, N. 0, Ano 2015

Uso de bases meteorológicas relacionais em banco de dados nosql para gerar informações que apoiem a tomada de decisão de agricultores

Anhanguera Educacional

João rafael Campos da Silva

1099477936

ResuMO

Com a grande popularização da internet, o mundo está ficando cada vez mais digital e com isso a quantidade de dados existentes sobre diversos ramos cresce a cada dia. Um dos setores que se informatizou há um bom tempo e acumula quantias gigantescas de dados hoje em dia é a meteorologia. Para o processamento desse grande volume de dados de uma forma rápida foram surgindo várias tecnologias que antes não eram tão importantes, como banco de dados NoSQL e modelos de processamento distribuído como o Hadoop. O projeto abordado nos próximos capítulos visa fundamentar essas tecnologias para que seja possível armazenar uma grande massa de dados meteorológicos obtidas de bases relacionais como PROTIN e BDMEP em um banco de dados NoSQL a fim de gerar informações precisas com maior rapidez que possam auxiliar agricultores na tomada de decisão. Para que haja uma integração entre esses dois tipos de banco de dados será utilizada uma tecnologia para a construção de WebService chamada REST, que através de recursos fornecidos é possível que outras aplicações obtenham dados por protocolo HTTP de forma rápida e segura. Acoplando todas essas tecnologias em uma única infraestrutura será possível realizar a construção de uma aplicação que analise dados obtidos por estações de meteorologia ao longo dos anos e transforma-las em informações valiosas para auxiliar agricultores na hora de decidir o momento certo para efetuar o seu plantio ou se será necessário investir em algum sistema de irrigação podendo evitar grandes perdas por falta de um simples planejamento.

Palavras-Chave: Big Data, Hadoop, NoSQL, WebService, Agricultura, Banco de Dados, Meteorologia.

Introdução

Busca-se fundamentar nesse trabalho tecnologias relacionadas a construção de uma infraestrutura para o Big Data. Serão abordados conceitos de Big Data assim como a introdução a sua infraestrutura. Serão discutidas opções de bancos de dados NoSQL, que são compatíveis com o modelo distribuído do Big Data, e webservices REST para a coleta dos dados de bancos de dados meteorológicos relacionais.

Bases de dados meteorológicos como PROTIN e BDMEP armazenam diariamente em forma digital uma infinidade de informações, referentes as séries históricas da rede de estação do INMET, são 291 estações meteorológicas convencionais que armazenam cerca de 3 milhões de informações sobre o clima (BDMEP, 2015).

Atualmente esses dados se encontram em bancos de dados relacionais, onde a busca por alguma informação de valor pode ser demorada devido a grande massa de dados. Uma alternativa seria realizar uma modelagem especial com bancos de dados não relacionais, permitindo uma escalabilidade mais barata e menos trabalhosa além da rapidez em que esses dados poderiam ser acessados por um sistema analítico fornecendo informações valiosas que em bancos relacionais talvez não seriam possíveis.

Isso tudo se encaixa no termo Big Data que de forma geral consiste em um grande volume de dados não estruturados que se processados com agilidade e com um objetivo em mente podem servir de grande apoio a tomadas de decisões (DEVMEDIA, 2015). Dados meteorológicos armazenados diariamente desde 1961, se processados com essa rapidez, poderiam gerar uma previsão das melhores épocas para que um agricultor realize um plantio seguro (BDMEP, 2015).

1. OBJETIVO

Nesse artigo busca-se elucidar algumas tecnologias muito importantes para se trabalhar com o Big Data, e como as mesmas podem ser úteis para a construção de uma infraestrutura que analise um grande volume de informações meteorológicas garimpando informações valiosas.

Com essas informações em mãos será possível auxiliar agricultores na tomada de decisão baseada em uma base de dados meteorológicos não estruturada, gerada a partir de um banco de dados relacional como BDMEP e PROTIN.

1. metodologia

Foi necessário vasculhar a internet para encontrar grandes volumes de dados meteorológicas que sejam confiáveis e acessíveis por meio de download ou webservices.

Buscar em livros, artigos e tutoriais na internet meios de realizar a construção de uma base não estruturada a partir de dados relacionais de um modo que seja aproveitado o maior volume de informações possíveis.

Estudar como é realizada a construção de web services para a interlocução de dados de uma forma segura e com performance admissível. Depois de entendido o procedimento, será desenvolvido um webservice de forma genérica para aceitar informações de vários lugares garantindo a variedade de dados.

Foi realizado um estudo sobre alguns bancos de dados NoSQL oferecidos pelo mercado e identificado os prós e contras de cada um. Depois de decidido qual banco utilizar, será realizada uma modelagem especial baseada nos dados que serão obtidos pelo webservice.

1. Big data

Com o grande desenvolvimento da internet no mundo, muitas coisas estão sendo digitalizadas. Nesse capitulo serão levantados conceitos de um novo sistema de tratamentos para esse grande volume de dados gerados diariamente.

Em meio ao grande crescimento da tecnologia no mundo as grandes organizações preveem que o grande volume de dados gerados pela internet servirá de grande ajuda no apoio a decisões baseada em estatísticas. Os dados acumulados são oriundos de redes sociais, sensores, e-mails, vídeos, websites, sensores, câmeras e smartphones, entre outros. (PARAIZO, 2013)

Apesar de todo esse conteúdo estar disponível na internet para qualquer empresa utilizar, a maioria deles não é tratado e analisado de uma forma que forneça informações uteis para a organização. Por esse motivo o Big Data tem chamado a atenção de vários analistas de dados, o mesmo não possui uma definição exata, depende da situação em que é aplicado.

O Big Data pode ser definido mais simplificadamente por “Big Data = volume + variedade + velocidade + veracidade, gerando algum valor”. Quando falamos em volume, estamos nos referindo a grande massa de dados disponíveis na internet que pode ser utilizada de modo oportuno por qualquer organização que se interesse. Por se tratar de dados originários de diversos lugares, a variedade de informações obtidas é imensa.

Para que a empresa possa obter algum lucro do investimento aplicado no processamento desses dados, o retorno de informações uteis deve ser ágil e os mesmos devem ser confiáveis, caso contrário, as informações se tornam lixo acumulado em servidores caros, prejudicando a organização. (TAURION, 2013)

* 1. Etapas do processo Big Data

Para que a implantação do conceito Big Data seja implantada com sucesso em uma empresa, devem ser seguidas algumas etapas de processamento.

A primeira fase do Big Data é a coleta de dados, deve-se coletar o maior volume e variedade de informações e armazená-las para futuramente serem processadas. Nessa fase também é necessário que dados quebrados ou sem sentido sejam descartados, para que não influenciem em uma análise posterior. (CHEDE, 2012)

Depois disso vem a parte de integração, agregação e representação dos dados obtidos. Cada tipo de dado deve receber um tratamento diferente, esse objetivo pode ser alcançado com mais facilidade se forem elaboradas categorias para cada um dos tipos de dados encontrados. (CHEDE, 2012)

A próxima etapa é realizar modelagem desses dados, por se tratar de uma variedade enorme de categorias requer um conhecimento maior sobre as regras de negócio, aqui entra o “Cientista de Dados”.

A última fase propõe a realização de uma pesquisa sobre novas formas de visualização, que irão aliciar na interpretação dos dados processados e a obtenção de informações úteis que ajudem de alguma forma no desenvolvimento da empresa.

* 1. Vantagens

Compreender o que os clientes realmente necessitam, estudando seus hábitos de consumo. Quando se entende exatamente o que o cliente quer, o resultado obtido pode ser mais claro, ganhando a confiança do mesmo. (HABITZREITER, 2014)

Encontrar potenciais compradores a partir da mensuração em tempo real das redes sociais. O desenvolvimento da tecnologia permite que pessoas de diversas localidades geográficas conheçam o produto e ofertas em tempo real, com isso pode ocorrer expansão nas vendas (HABITZREITER, 2014)

Evitar que haja imprevistos no mercado graças a análise em tempo real de informações de diversas variáveis do mercado. Tudo pode ser analisado em tempo real, tornando possível a prevenção e antecipação a possíveis dificuldades no mercado e tomar decisões para contornar esses problemas. (ZUPPO, 2014)

É possível monitorar a concorrência sem que ela perceba para criar ofertas especiais. Conhecer o concorrente e extremamente necessário para aumentar lucros. (ZUPPO, 2014)

Estas são as principais vantagens de se utilizar Big Data, existem outras, pois como disse anteriormente, o Big Data se trata de vários conceitos.

1. APACHE HADOOP

Neste capítulo é abordado um framework da Apache que vem sendo a principal escolha de grandes empresas na hora de montar uma infraestrutura para trabalhar com Big Data.

Estamos vivendo a era da informação onde a geração de dados digitais é cada vez maior, alcançando os petabytes diários caracterizando o Big Data.

No entanto, é necessária uma implementação que analise esse grande volume de dados e processe pesados cálculos para identificar comportamentos e fornecer sugestões.

Para solucionar esse problema surgiu o Apache Hadoop capaz de processar muitos dados em um sistema computacional distribuído. O framework também contém um modelo de programação e um sistema de armazenamento de dados que permitem um processamento muito rápido.

O Haddop possui um modelo de programação chamado de MapReduce que permite o processamento distribuído de uma grande tarefa. Além do MapReduce o framework também contém um sistema de armazenamento chamado de HDFS (Hadoop Distributed File System) que como o próprio nome diz, também trabalha de forma distribuída facilitando o trabalho do MapReduce. (GASPAROTTO, 2013)

Ao longo do tempo foram incorporadas outras ferramentas ao Hadoop como o Hive, Pig, ZooKeeper, Flume, Mahout entre outras que buscam solucionar problemas específicos na arquitetura agregando ainda mais o framework Hadoop. (GASPAROTTO, 2013)

O Haddop foi criando para trabalhar de forma distribuída, mas também é possível usá-lo de forma independente ou pseudo distribuído facilitando para o desenvolvedor na hora da implementação das funcionalidades. (MARTINS, 2013)

* 1. Introdução ao MapReduce

Hoje em dia é um desafio armazenar, manipular e analisar uma grande quantidade de dados de forma inteligente. Sistemas corporativos ou sociais geram milhares de dados por dia que não podem ser processados por tecnologias tradicionais.

Pensando nesse contexto surgiu o termo Big Data, que é definido pelos "3v", volume, velocidade e variedade. (ALECRIM, 2013)

Diante desse desafio, dois engenheiros do Google desenvolveram uma ferramenta chamada MapReduce que permite separar uma grande massa de dados em pequenos pedaços e distribuir o processamento dos mesmos em vários computadores de baixo custo, tornando o buscador da empresa mais rápido e eficiente. (ALECRIM, 2013)

O MapReduce se mostrou adequado para trabalhar com problemas que podem ser divididos em subproblemas. Ele é um modelo de programação que une o Map, mapeamento de dados, que serve como entrada para a função Reduce que minimiza esses dados em informações úteis para a empresa. (ANDRADE, 2013)

Vários Maps e Reduces podem ser espalhados em maquinas de baixo custo tornando o processamento ágil e escalável. Para a utilização do MapReduce é necessário que a entrada de dados esteja organizada um uma lista contendo chave e valor para que possam ser dividas para o processamento paralelizado. (ANDRADE, 2013)

* 1. Sistema de Arquivo HDFS

O Hadoop possui seu próprio sistema de arquivos distribuído, esse sistema é chamado de HDFS (Hadoop Distributed File System), ele é responsável pela organização, armazenamento, localização, compartilhamento e proteção de arquivos que ficam distribuídos em uma rede de servidores. (HANSON, 2013)

Em um sistema distribuído é essencial que haja um controle de concorrência, ou seja, quando diversos usuários tentam gravar o mesmo arquivo. Tal controle é necessário para que haja uma operação atômica e não prejudique a consistência nas informações obtidas. Um sistema que trabalha de forma distribuída deve garantir a atomicidade nas operações de leitura, escrita, criação ou remoção de um arquivo, de forma transparente para quem manipula os dados. (HANSON, 2013)

Segundo HANSON o HDFS faz parte da camada de armazenamento do Hadoop, o mesmo foi otimizado para operações de leitura e escrita visando o processamento de arquivos grandes na escala de Terabytes que ficam localizados em computadores de um cluster.

O sistema de arquivos do Haddop é altamente escalável e disponível, graças a replicação de dados em vários computadores e um processo de tolerâncias a falhas. O HDFS se encarrega dividir os arquivos em blocos, normalmente de 64MB, e espalha-los por várias maquinas (Em caso de cluster no mínimo 3 e em modo local no mínimo 1) ficando assim tolerante a falhas tanto de software quanto de hardware. O HDFS é dividido estruturado em um padrão mestre-escravo com dois processos principais, NameNode e DataNode. (MICROSOFT, 2014)

1. bancos de dados nosql

O termo NoSQL surgiu em 1998 como alternativa de um banco de dados não relacional de código aberto. Segundo Carlo Strozzi, seu autor, alega que a iniciativa NoSQL "é completamente distinto do modelo relacional e portanto deveria ser mais apropriadamente chamado "NoREL" ou algo que produzisse o mesmo efeito". (IANNI, 2012)

Com o grande avanço da internet e o surgimento de novas tecnologias foram aparecendo novos tipos de dados mais complexos que se tornaram uma dificuldade para o modelo relacional e com uma manutenção muito cara. Em 2006 o termo NoSql voltou a ser comentado. (PORCELLI, 2010).

O nome NoSQL era uma tentativa de caracterizar diversos bancos de dados não relacionais que estavam surgindo tentando fazer referência aos nomes de bancos de dados relacionais populares como MySQL, MS SQL, PostgreS0QL, etc. (ITEXTO, 2012)

Com a popularização das redes sociais e telefonia móvel, a internet se tornou uma fonte imensa de dados não estruturados que se encaixaram perfeitamente no modelo não relacional, ou seja, bancos NoSQL. (ITEXTO, 2012)

Portanto, os bancos de dados NoSQL, foram ficando mais populares entre as poderosas empresas pois reúnem as particularidades de poder trabalhar com dados semiestruturados ou não estruturados vindos de variadas origens (arquivos multimídia, arquivos de log, websites, etc.). (WEISSMANN, 2015)

* 1. Comparando Bancos NoSQL com o Modelo Relacional

Por muito tempo vem sendo usado bancos de dados relacionais como a principal escolha para o armazenamento de dados corporativos, porem grandes empresas vem adotando uma nova forma de tratar esses dados.

Com o grande aumento no volume de dados coletados diariamente e a previsão de que isso só tem a crescer, bancos relacionais perdem em escalabilidade.

Como alternativa para solucionar esse problema surgiu o NoSql, que vem do termo Not Only SQL que proporciona a possibilidade de trabalhar em um ambiente distribuído, aumentando a escalabilidade e oferecendo um alto desempenho a baixo custo. (WEISSMANN, 2012)

Para solucionar o problema de escalabilidade o modelo relacional geralmente são utilizadas duas alternativas, o aumento do número de servidores (escalabilidade horizontal) ou o upgrade dos servidores (escalabilidade vertical). Mas se o volume de dados continuar a crescer rapidamente essa alternativa se torna muito custosa. Uma melhor solução seria trabalhar de forma distribuída, no entanto, o sistema estruturado em linhas e colunas dos modelos relacionais gera uma enorme dificuldade no manuseio desses dados quando as tabelas se encontram em computadores diferentes. Por conta disso foram surgindo as soluções não relacionais. (SOUZA, 2012)

O NoSQL busca uma forma nova de modelagem, modificando a rígida e tradicional estrutura utilizada nos bancos relacionais. Grandes empresas tinham como objetivo flexibilizar os dados conforme as suas necessidades, essa flexibilidade é extremamente importante para o acesso distribuído, fornecendo alta escalabilidade e dispensabilidade para as aplicações. (SOUZA, 2012)

A intenção do NoSQL não é substituir o modelo relacional tradicional, mas sim fornecer uma alternativa para as empresas que tratam de um grande volume de dados e que os mesmos possam ser um pouco mais flexíveis. (BRITO, 2012)

Os bancos de dados NoSQL são divididos da maneira que armazenam as informações, os principais tipos são: chave-valor, colunas, grafos e documentos. (BRITO, 2012)

Por tanto o NoSQL vem para somar como uma alternativa altamente escalável e com grande disponibilidade de dados consistentes, portanto vale analisar bem e escolher a melhor alternativa de Banco de dados para a sua empresa.

* 1. Características de Bancos NoSQL

Segundo o portal de artigos DEVMEDIA, essas são as principais características dos bancos de dados NoSQL.

Bancos de dados NoSQL visão a utilização de processamento paralelo para a manipulação de informações. Para que o processamento de um grande volume de dados ocorra com uma performance razoável, é mais eficaz que o trabalho seja dividido em várias tarefas menores e que podem assim ser executadas de forma assíncrona utilizando todo o poder de processamento de cada máquina em uma infraestrutura distribuída.

O uso de vários processadores pequenos trabalhando ao mesmo tempo se torna mais eficiente do que um processador maior que pode ficar ocioso executando uma única função de cada vez, além de ser uma alternativa muito mais econômica evitando que empresas fiquem presas as poucas fornecedoras que disponibilizam hardwares poderosos.

Para atender seus usuários de forma mais apropriada as empresas distribuem seus data centers por diversas partes do país ou do mundo, visando solucionar problemas de disponibilidade e performance. (KOKAY, 2012)

A distribuição deles combinada com o hardware barato, obrigando ao sistema a necessidade de ser sólido o suficiente para tolerar falhas constantes e imprevisíveis, seja a mesma relacionada ao hardware ou a infraestrutura do local onde o data center se localiza. (KOKAY, 2012)

Pensando nessas necessidades foram surgindo diversos tipos de bancos NoSQL buscando atender a diversas particularidades, como por exemplo os Bancos de Dados Chave/Valor, orientados a documentos, famílias e colunas e grafos.

1. WEBSERVICE REST

O REST é fortemente baseado no protocolo HTTP que é basicamente a base da Web como a conhecemos, sendo que a navegação desse protocolo pode ser considerado REST. Porém o REST não é tão simples quanto o protocolo HTTP, ele é constituído por várias regras que devem ser seguidas para que se possa utilizar o protocolo. (SAUDATE, 2013)

* 1. Semânticas de Recursos

Basicamente todos os serviços e operações executados pelo REST são baseados em recursos, que são entidades bem definidas em softwares, que possuem endereços próprios e identificadores. (XAVIER, 2013)

Como exemplo podemos referenciar um sistema de comercio de camisetas e a camiseta é um recurso. Assim, segundo as regras do REST dizem que camisetas devem ter uma URL própria e que esta URL deve ser significativa. Desta forma uma URL que deixaria claro o recurso camisetas seria /camisetas. (XAVIER, 2013)

De acordo como modelo REST, esta URL realizará interação com todas as camisetas do sistema. Para tratar de camisetas específicas, são usados identificadores. Estes identificadores podem ter qualquer formato, por exemplo, suponha uma camiseta Abercombie com uma chave primária 123456. Note que o nome da camiseta não serve como identificador, por existirem várias dessa marca, mas a chave primaria sim, por ser única em todo o banco de dados. Desta forma, deve ser possível buscar esta camiseta através da sua chave primária, com a URL /camisetas/123456. Lembre-se, regras semelhantes aplicam-se a qualquer recurso. O identificador a ser utilizado na URL pode ser qualquer coisa que você assim desejar, e mais de um tipo de identificador pode ser usado para alcançar um recurso. (SAUDATE, 2013)

* 1. Interação por Métodos

Para interagir com as URL’s, os métodos HTTP são utilizados. A regra de ouro para esta interação é que URL’s são substantivos, e métodos HTTP são verbos. Isto quer dizer que os métodos HTTP são os responsáveis por provocar alterações nos recursos identificados pelas URL’s. (TILKOV, 2008)

Estas modificações são padronizadas, de maneira que a função GET recupera os dados identificados pela URL, a POST cria um novo recurso, a PUT atualiza um recurso e a DELETE apaga um recurso.

Note que estes quatro métodos principais podem ser diretamente relacionados a operações de bancos de dados. Assim, para recuperar o cliente de número 1 do banco de dados, basta utilizar o método GET em conjunto com a URL /camiseta/1; para criar um novo cliente, basta utilizar o método POST sobre a URL /camiseta (o identificador será criado pelo banco de dados); para atualizar este cliente, utilize o método PUT sobre a URL /camiseta/1 e, finalmente, para apagar o cliente, utilize o método DELETE sobre a URL /camiseta/1. (COUTINHO, 2013)

Note que todas estas operações são lógicas. Isto quer dizer que utilizar o método DELETE, por exemplo, não significa necessariamente excluir o dado do banco de dados - significa apenas indisponibilizar o recurso para consumo pelo método GET. Ou seja, DELETE pode apenas marcar o dado no banco de dados como desativado. (COUTINHO, 2013)

1. dEscrição do protótipo

O projeto será executado em quatro passos, na primeira parte será realizada a construção de uma modelagem NoSQL que consiga armazenar um grande volume de dados obtido das bases meteorologias BDMEP e PROTIM.

Será construído um software para garimpar dados, através de um webservice REST, das bases relacionais de dados meteorológicos e armazena-los em uma infraestrutura especial NoSQL, para que o sistema analítico possa processar os dados obtidos mais rapidamente fornecendo um resultado mais preciso para o usuário.

Com os dados em mãos é preciso analisa-los para obter um resultado refinado que sirva de alguma forma como apoio para os agricultores que farão uso do sistema. Para essa função utilizaremos uma infraestrutura Hadoop processando uma grande massa de dados em um ambiente distribuído, de forma rápida e eficiente.

Por fim será implementada uma aplicação utilizando as facilidades da plataforma Java EE para que o agricultor possa consultar gráficos e projeções do tempo, baseado nos dados analisados, de um determinado local do estado de Mato Grosso do Sul. O software fornecerá sugestões ao agricultor, como por exemplo, se o plantio realizado precisará de um sistema de irrigação devido à falta de chuvas no local ou alertar o risco de grandes geadas e temporais naquele período de tempo provocando, em um caso mais crítico, a perca da lavoura.

Com toda a infraestrutura montada e trabalhando de forma conjunta, o agricultor degustara de uma aplicação de grande valor para o apoio a decisões sobre futuros investimentos na área.

1. considerações finais

Nesse trabalho abordamos conceitos para a construção de uma infraestrutura compatível com o Big Data. Foram fundamentados e exemplificados os tipos de bancos de dados NoSQL que trabalham de forma distribuída fornecendo uma integração muito grande com o HDFS, o sistema de arquivos do Hadoop.

Conhecemos os componentes principais do framework Hadoop para a o processamento do Big Data, como o MapReduce e o HDFS. Também exploramos alguns softwares integrados ao Haddop que auxiliam no processamento analítico dos dados fornecendo informações úteis para as empresas.

Foi realizada uma breve introdução do que é o tipo de webservice REST que é fundamental para a busca de informações em grandes bases de dados meteorologias relacionais existentes na internet.

Concluímos que com tecnologias poderosas e de código aberto pode ser realizada a construção de uma infraestrutura moderna e eficiente para a utilização do Big Data.

Este trabalho foi muito importante para ampliar meus conhecimentos e a melhor compreensão dessas novas tecnologias que estão em alta nos dias de hoje. Com os conhecimentos obtidos será mais simples a construção prática do meu projeto assim como elucidar a resolução de alguns problemas em meus objetivos.

Referências

ALECRIM, E. Hadoop MapReduce: Introdução a Big Data. Disponível em: <http://www.devmedia.com.br/hadoop-mapreduce-introducao-a-big-data/30034>. Acesso em: 25 de Maio de 2015.

ALGARTECH. Cientista de dados. Disponível em: <http://www.algartech.com/portugues/noticias/em-noticia/mercado/cientista-de-dados-3-habilidades-que-universidades-devem-ensinar/>. Acesso em: 25 de Maio de 2015.

ANDRADE, T. P. C. MapReduce: Conceitos e Aplicações. Disponível em: <www.ic.unicamp.br/~cortes/mo601/trabalho\_mo601/tiago\_cruz\_map\_reduce/relatorio.pdf>. Acesso em: 25 de Maio de 2015.

BDMEP. Banco de Dados Meteorológicos para Ensino e Pesquisa. Disponível em: <http://www.inmet.gov.br/projetos/rede/pesquisa/>. Acesso em: 12 de Abril de 2015.

BRITO, R. W. Bancos de Dados NoSQL x SGBDs Relacionais: Análise Comparativa. Disponível em: <http://www.infobrasil.inf.br/userfiles/27-05-S4-1-68840-Bancos%20de%20Dados%20NoSQL.pdf>. Acesso em: 25 de Maio de 2015.

CHEDE, C. Estamos definitivamente entrando na era do Big Data. Disponível em: <https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/estamos\_definitivamente\_entrando\_na\_era\_do\_big\_data?lang=en>. Acesso em: 25 de Maio de 2015.

COUTINHO, P. C. Primeiros passos com os serviços REST. Disponível em: <http://www.infoq.com/br/articles/rest-introduction>. Acesso em: 25 de Maio de 2015.

DEVMEDIA. Big Data: Armazenamento, análise e gerenciamento. Disponível em: <http://www.devmedia.com.br/big-data-armazenamento-analise-e-gerenciamento/30 918> Acesso em: 9 de Maio de 2015.

DEVMEDIA. Comparando o NoSQL ao modelo relacional. Disponível em: <http://www.devmedia.com.br/comparando-o-nosql-ao-modelo-relacional/30917>. Acesso em: 12 de Abril de 2015.

GASPAROTTO, H. M. Hadoop MapReduce: Introdução a Big Data. Disponível em: <http://www.devmedia.com.br/hadoop-mapreduce-introducao-a-big-data/30034>. Acesso em: 25 de Maio de 2015.

HABITZREITER, P. Vantagens da utilização do Big Data. Disponível em: <http://www.administradores.com.br/artigos/tecnologia/vantagens-da-utilizacao-do-big-data/79939/>. Acesso em: 25 de Maio de 2015.

HANSON, J. J. Uma Introdução ao Hadoop Distributed File System. Disponível em: <http://www.ibm.com/developerworks/br/library/wa-introhdfs/#icomments>. Acesso em: 25 de Maio de 2015.

IANNI, V. Introdução aos bancos de dados NoSQL. Disponível em: <http://www .devmedia.com.br/introducao-aos-bancos-de-dados-nosql/26044> Acesso em: 9 de Maio de 2015.

ITEXTO. Desconstruindo NoSQL: em busca de melhores termos. Disponível em: <http://www.itexto.net/devkico/?p=1523>. Acesso em: 25 de Maio de 2015.

JONES, M. T. Processamento de Dados Distribuídos com Hadoop – Parte 01. Disponível em: <http://imasters.com.br/artigo/18122/desenvolvimento/processamento\_de\_dados\_distribuidos\_com\_hadoop\_parte\_01/>. Acesso em: 25 de Maio de 2015.

KNXISAAC. Como aproveitar as informações da web e adequar as melhores para sua empresa. Disponível em: <http://www.konics.com.br/blog/?p=2131>. Acesso em: 25 de Maio de 2015.

KOKAY, M. C. Banco de dados NoSQL: Um novo paradigma. Disponível em: <http://www.devmedia.com.br/banco-de-dados-nosql-um-novo-paradigma-revista-sql-magazine-102/25918>. Acesso em: 25 de Maio de 2015.

MARTINS, C. R. L. Portal de Tecnologia da Informação para Meteorologia. Disponível em: <http://bancodedados.cptec.inpe.br/estatisticas/cobertura.jsp>. Aces-so em: 12 de Abril de 2015.

MICROSOFT. Introdução ao Hadoop no HDInsight: Processamento e análise de big data na nuvem. Disponível em: <http://azure.microsoft.com/pt-br/documentation/articles/hdinsight-hadoop-introduction/>. Acesso em: 25 de Maio de 2015.

MOURA, T. G. C. Estudo Banco de Dados espaciais distribuído. Disponível em: <http://www.terralab.ufop.br/dokuwiki/lib/exe/fetch.php?media=terralab:equipe:mourathiago:relatorio.pdf>. Acesso em: 25 de Maio de 2015.

PARAIZO, L. Estudo da IBM aponta nível de uso do Big Data no Brasil. Disponível em: <http://www.noticenter.com.br/?modulo=noticias&caderno=tecnologia&noticia=02497-estudo-da-ibm-aponta-nivel-de-uso-do-big-data-no-brasil#.VWM24EYj4bg>. Acesso em: 25 de Maio de 2015.

PERERA, S. Considerações sobre o Banco de Dados Apache Cassandra. Disponível em: <http://www.ibm.com/developerworks/br/library/os-apache-cassandra/>. Acesso em: 25 de Maio de 2015.

PORCELLI, A. O que é noSQL. Disponível em: <http://www.devmedia.com.br/o-que-e-nosql-java-magazine-86/18777>. Acesso em: 25 de Maio de 2015.

PROTIM. Portal de Tecnologia da Informação para Meteorologia. Disponível em: <http://bancodedados.cptec.inpe.br/estatisticas/cobertura.jsp>. Acesso em: 12 de Abril de 2015.

SATO, P. M. GraphDB Series: O que é um banco de dados de grafos. Disponível em: <http://www.dev.mayogax.me/o-que-e-banco-de-dados-de-grafos>. Acesso em: 25 de Maio de 2015.

SAUDATE, A. REST: Construa API's inteligentes de maneira simples. Editora Casa do Código, vol. 1, nº 314, São Paulo, 2013.

SILVA, B. A. Apache Hadoop. Disponível em: <http://www.dcomp.sor.ufscar.br/verdi/topicosCloud/Hadoop.pdf>. Acesso em: 25 de Maio de 2015.

SOUZA, C. Comparativo Técnico entre tecnologias de banco de dados relacional, NoSQL e NewSQL. Disponível em: <http://www.academia.edu/6559619/Comparativo\_T%C3%A9cnico\_entre\_tecnologias\_de\_banco\_de\_dados\_relacional\_NoSQL\_e\_NewSQL>. Acesso em: 25 de Maio de 2015.

TAURION, C. Você realmente sabe o que é Big Data?. Disponível em: < https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce\_realmente\_sabe\_o\_que\_e\_big\_data?lang=en>. Acesso em: 25 de Maio de 2015.

TILKOV, S. Uma rápida Introdução ao REST. Disponível em: <http://www.infoq.com/br/articles/rest-introduction>. Acesso em: 25 de Maio de 2015.

VARDANYAN, M. Um olhar sobre alguns bancos de dados NoSQL. Disponível em: <http://imasters.com.br/artigo/22034/banco-de-dados/um-olhar-sobre-alguns-bancos-de-dados-nosql/>. Acesso em: 25 de Maio de 2015.

WEISSMANN, H. L. Entendendo NoSQL. Disponível em: <http://www.devmedia.com.br/entendendo-nosql-revista-sql-magazine-111/27851>. Acesso em: 25 de Maio de 2015.

XAVIER, O. C. Serviços Web Semânticos Baseados em RESTful. Disponível em: <http://www.portal.inf.ufg.br/mestrado/sites/www.inf.ufg.br.mestrado/files/uploads/Dissertacoes/OtavioCala%C3%A7a.pdf>. Acesso em: 25 de Maio de 2015.

ZHU, J. Data Modeling For Big Data. Disponível em: <http://www.ca.com/ us/~/media/Files/Articles/ca-technology-exchange/data-modeling-for-big-data-zhu-wa ng.pdf>. Acesso em: 12 de Abril de 2015.

ZUPPO, D; COSTA, L. S; FERNANDES, S. Big Data: Estudo do Ambiente, Desafios e Analise Estratégica para o Brasil. Disponível em: <http://portal.crie.coppe.ufrj.br/portal/data/documents/storedDocuments/%7B93787CAE-E94C-45C7-992B-9403F6F40836%7D/%7B53B680A0-DBBE-48B4-AB48-9D0F77B736D1%7D/RJ24\_Projeto%20Big%20Data.pdf>. Acesso em: 25 de Maio de 2015.