

Hadoop

Uma Plataforma de Dados Não Convencional para Apoiar a Implantação da Cultura de Orientação a Dados nas Empresas

Patrícia Cunha de Lima¹, Fernando Soares de Siqueira²

Sumário. 1. Introdução. 2. Gestão orientada a dados. 2.1 Orientação à dados. 3. Banco de dados não convencional. 3.1 Histórico de banco de dados não relacionais. 3.2 Características de banco de dados não relacionais. 3.3 Modelo de dados banco de dados não relacional. 4. Hadoop. 4.1 Modelo de dados do hadoop. 4.2 HDFS e Mapreduce. 4.2.1 HDFS (Hadoop distributed file system). 4.2.2 Mapreduce. 4.3 outros componentes. 5. Resultados. 6. Considerações finais. 7. Referências .

Resumo: A Cultura de Orientação a Dados tem por objetivo uma forma de gestão empresarial focada na tomada de decisões a partir de dados. Isto garante uma maior assertividade e rapidez, ao comparar com decisões, nas quais o fator empírico é predominante. Essa cultura está alavancando a competitividade do mercado nos últimos anos. Devido a esta mudança, faz-se necessária uma estrutura que possa ser capaz de coletar, integrar e analisar os dados dos mais diversos formatos, oferecendo ao usuário final a possibilidade de tirar o máximo proveito desses em

¹Graduanda em Sistemas de Informação. UNI7 – Centro Universitário 7 de Setembro, Fortaleza – CE. vippatriciac@gmail.com

²Mestre em Informática Aplicada. UNI7 – Centro Universitário 7 de Setembro, Fortaleza – CE. fernando.siqueira@uni7.edu.br

prol da empresa. Como um dos elementos desta arquitetura computacional, surge uma plataforma de dados voltada para armazenamento de dados não estruturadas, o Hadoop. Essa plataforma auxilia na captação e análise de dados em nível de Big Data. Por meio de pesquisa bibliográfica, este artigo tem o objetivo de mostrar como a orientação a dados é importante para manter a competitividade das empresas e como o Hadoop, como ferramenta de gestão de dados, pode auxiliar na arquitetura necessária para que haja essa gestão de dados, dando ênfase às suas características e sua estrutura.

Palavras-chave: análise de dados, orientação a dados, big data, banco de dados não convencional, Hadoop.

1 Introdução

Quando falamos a respeito de competitividade de mercado, estamos nos referindo a um contexto que está em evolução constante. Juntamente desse contexto de competição, faz-se necessário que inovações e mudanças sejam implantadas pelas empresas para que possam estar aptas a atuar de forma que façam concorrência às demais e obtenham uma vantagem competitiva.

Existem empresas que já implantaram mudanças que agregam esse diferencial competitivo através da inovação, não apenas em seu contexto tecnológico, mas também na reformulação de sua organização e de negócios. Também existem aquelas empresas que já foram fundadas com base nessas novas mudanças que o mercado está propondo.

Um elemento fundamental dessas mudanças é a gestão orientada a dados. Nos últimos anos, as empresas vêm focando seus esforços para construir uma cultura organizacional que seja baseada na orientação de dados, isto é, utilizar os dados como ponto chave de sucesso e diferencial no processo de tomada de

decisão da empresa, possibilitando a geração de um valor agregado para o seu cliente, interno ou externo (GABRIEL, 2018).

As empresas orientadas a dados precisam de muita assertividade e agilidade em suas informações, para tal é imprescindível investimentos em ferramentas que façam o tratamento dos dados de forma eficiente e inteligente. Novas estratégias de dados e de análises prometem acelerar a transformação digital, mas o sucesso dependerá da variedade de arquiteturas complementares. Os profissionais técnicos precisam mudar de arquiteturas fixas e rígidas para portfólios flexíveis de dados e análises para melhor se adaptarem à demanda futura. (GARTNER, 2019).

Entende-se então que é elementar conhecer a demanda que as empresas vão ter quanto aos dados e o seu processamento, mas para que possam ser capazes de atender a estas demandas, será necessária a implantação de arquiteturas que consigam absorver aos dados que são precisos para a organização.

Nesse ponto, as organizações podem ter seus objetivos frustrados ao não conseguir obter retorno compatível aos investimentos que está fazendo, mas para que haja esse retorno, é necessária que a arquitetura escolhida seja a correta para atender as necessidades que a empresa apresenta.

Assim se faz necessário um alinhamento entre as estratégias de negócio da empresa e sua arquitetura de TI. Trazendo novas tendências para o mercado e desafiando os métodos e processos tradicionais.

Por estes desafios, este trabalho objetiva, principalmente, demonstrar como um software de armazenamento e processamento distribuído de grandes conjuntos de dados, pode se encaixar em arquiteturas de empresas orientadas a dados, no caso o Hadoop.

2 Gestão Orientada a Dados

Na Era Digital, as formas de produção e valor passam a depender de dados oriundos dos diversos sistemas de informação que permeiam a cadeia de negócios das organizações. Os dados se tornam o elemento principal dentro das empresas, indústrias e demais setores produtivos da economia, levando ao surgimento de uma economia em que o valor se encontra em sua extração adequada, manipulação eficiente e transformação em algo relevante – denominado isso de data capital (capital de dados) (GABRIEL, 2018).

Obter *insights* (percepções, ideias, sugestões, indicativos, correlações, etc.) que apoiem a gestão a ser mais efetiva dentro de um modelo de sistema complexo, que envolve um grande volume de dados, inúmeros agentes e que exiba comportamentos emergentes não triviais e auto-organizados, não é tão simples. Contudo, é imprescindível para a organização ter uma diferenciação entre seus concorrentes e aumentar a sua efetividade de atuação, conseqüentemente, aumentar a sua vantagem competitiva. Dessa forma, nessa era digital não se pode deixar de ter um modelo de gestão com esse novo valor econômico.

2.1 Orientação à Dados

Quando se fala a respeito de uma organização que tem gestão orientada a dados, falamos de uma empresa que preza pelo alto rendimento, focando as suas tomadas de decisões tomando como base evidências e um caráter objetivo. Esse tipo de mentalidade empresarial pode englobar todos os tipos de decisões nos mais diversificados ramos (SALVADOR, 2017).

A cultura de orientação de dados tem características, que a transformam em um modelo para tornar a empresa que a prática competitiva no mercado. Essa cultura estimula a criatividade na obtenção dos dados, pois podem haver dados perdidos ou não utilizados e isso pode dar a ideia errônea de que aqueles dados não tem valor, quando na verdade podem ser importante para outras áreas da empresa.

Outra característica seria que ela é voltada para as pessoas, por mais que a gestão a dados pareça ser fria e fundamentada na lógica e focada nos cálculos. A análise de dados tem necessidades que estão ligadas com competências de gestão de pessoas. E, geralmente, as empresas que têm essa cultura de orientação a dados focam em treinamentos de colaboradores.

Em uma entrevista feita pelo Economist Intelligence Unit , feita em 2012, a 530 empresas executivos seniores da América do Norte, Ásia/Pacífico, Europa Ocidental e América Latina, de diferentes segmentos industriais. A pesquisa descobriu que as empresas mais bem sucedidas adotaram uma cultura orientada a dados, estas conseguiram intensificar a utilização de dados por meio do fornecimento do treinamento necessário e do compartilhamento de dados por todos os níveis de funcionários e departamentos.

Para que se possa conseguir essa melhoria significativa, é necessário que se faça uma análise da arquitetura necessário para disponibilizar esse compartilhamento e processamento dos dados. Cada empresa tem uma demanda diferente a respeito de seus dados, porém com o advento da internet e IoT, é necessária uma arquitetura que suporte várias tipos de dados, até mesmo os não estruturados e que possa fazer o tratamento correto e eficiente dessas fontes de dados tão diversificadas.

É preciso que os componentes dessa arquiteturas sejam robustos e flexíveis para conseguir englobar e tratar de uma quantidade grande de dados produzidos e organizar suas várias fontes de maneira a entregar informações valiosas para a organização.

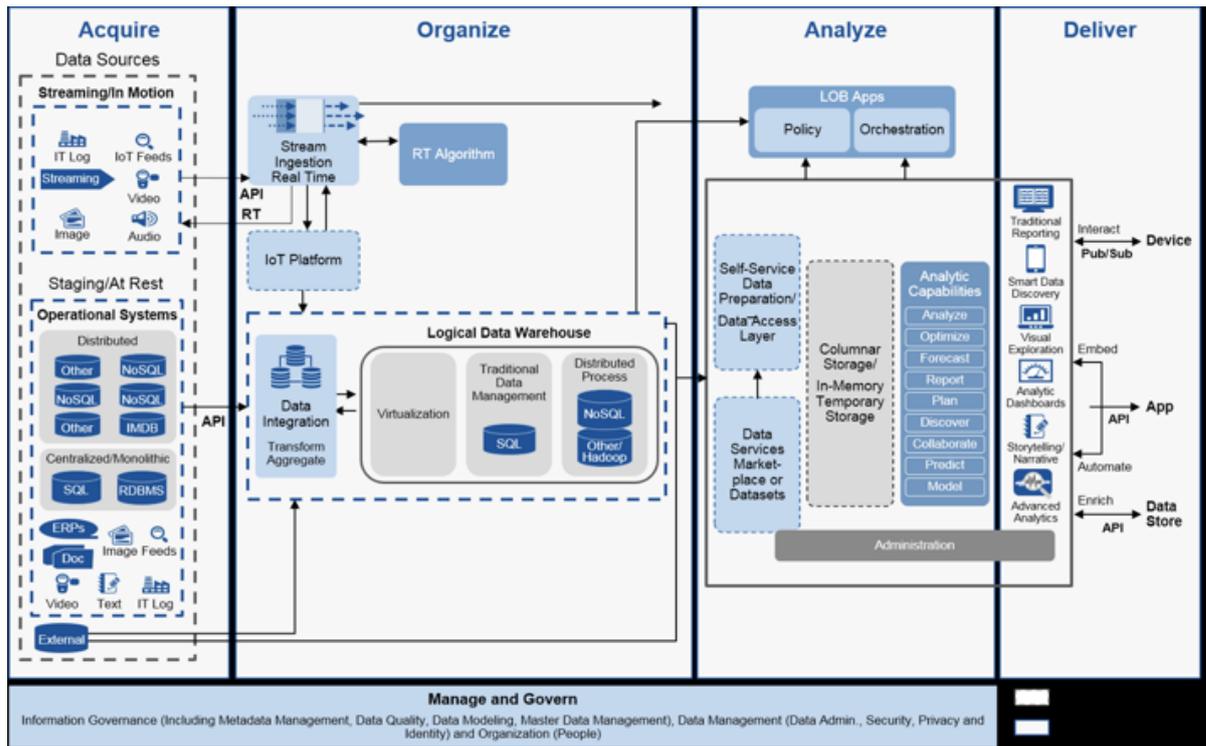


Figura X – Arquitetura Genérica de *Data & Analytics* (GARTNER, 2019)

3 Banco de Dados Não Convencional

3.1 Histórico de Banco de Dados Não Relacionais

O avanço das tecnologias como a internet, redes sociais, internet das coisas e muitos outros elementos que compõem o ecossistema do Big Data, têm imposto um cenário que amplia o universo de dados. E para esse novo cenário a representação e mapeamento dos dados requerem uma abordagem apropriada que ofereça desempenho e vantagens similares às que são encontradas no armazenamento dos dados com estruturas e associações bem definidas e conhecidas para as aplicações tradicionais (sistemas de folha de pagamento, sistemas bancários, sistemas contábeis, dentre outros que automatizam os processos das empresas).

A ampliação do uso da internet acarretou no aparecimento de diversos novos tipos de dados e com estes também um grande volume de geração de mais e mais dados a todo o momento. Resultando em um imenso repositório de dados que chegou um determinado momento que tratar e armazenar esses dados de maneira relacional tornou-se cada vez mais complexo e caro, o que não trazia vantagens para as empresas. Impulsionando a necessidade de se haver um banco de dados que considerasse essas novas especificidades que os bancos de dados relacionais não suportavam com a mesma eficiência em que atuavam com as aplicações tradicionais.

Em 2006, foi usado do termo NoSQL para os bancos de dados não relacionais que visava diferenciar das soluções do tipo relacional, como a exemplo MySQL e PostgreSQL. Porém, já se sabe que o termo NoSQL e não-relacional não auxilia muito na definição do que esses bancos de dados são realmente. As tecnologias NoSQL são recentes e ainda é necessário muito estudo para se estabelecer em qual contexto melhor se enquadram, principalmente em relação a auxiliar as organizações no cenário do Big Data.

Apesar dos termos não conseguirem passar o que é esse tipo de banco de dados, temos na maioria deles características em comum como: não relacional, distribuído, de código aberto e escalável horizontalmente, ausência de esquema ou esquema flexível, suporte à replicação nativo e acesso via APIs simples (NOSQL 2010).

Em meio a estas características, destaca-se principalmente a escalabilidade pelo fato dos bancos NoSQL terem sido criados a partir da natureza dos dados da web. A necessidade de se alcançar altos graus de paralelismo unido ao processamento de grandes volumes de dados, juntamente com a distribuição de sistemas em escala global. Essa própria natureza dos dados da web impulsionou a criação de soluções otimizadas e que pudessem tratá-los de acordo com sua particularidade, a falta de estrutura nos dados da web ou a necessidade de uma

flexibilidade de estrutura. Algo que se tornou inviável para os bancos relacionais atuarem de forma eficiente, tanto no armazenamento quanto no tratamento.

3.2 Características de Banco de Dados Não Relacionais

Mesmo com o pouco entendimento do contexto e dos limites que esses bancos podem atuar, existem vários relatos de empresas reportando sucesso na utilização desses bancos de dados uma vez que, em conjunto com os bancos de dados relacional, consegue oferecer a agilidade e assertividade necessária para as empresas construírem uma base de dados formada por dados internos e dados externos de formato heterogêneo e a utilizarem para exploração e análise de dados, tirando proveito de insights obtidos desse processo de análise de dados.

Importante destacar que, essa natureza dos bancos de dados não relacionais envolver a representação de uma imensa base de dados e não necessariamente estruturada impõe a necessidade de se ter uma solução para escalabilidade. Para tal, existem dois tipos de escalabilidade oferecidas para essa arquitetura de dados: vertical e horizontal. A vertical trata da ampliação do tamanho do servidor com mais CPU e memória RAM, uma solução não muito elegante, mas necessária ou até mesmo a única opção em algumas situações. Quanto a horizontal aborda a adição de mais servidores e o balanceamento da carga entre eles, este modelo é mais elegante.

Quando se fala em banco de dados não relacionais, fala-se de escalabilidade horizontal. Nos bancos NoSQL essa é feita com mais facilidade em comparação aos relacionais por não ser afetado pela alta concorrência que é gerada e que precisa ser tratada para não gerar alguma inconsistência. Um meio muito usado, mas alternativo, para alcançar a escalabilidade horizontal é o *Sharding*. Essa técnica divide os dados em múltiplas tabelas a serem armazenadas ao longo de diversos nós na rede, que na realidade é romper a cadeia de relacionamentos, forte característica encontrada nos bancos relacionais.

Outra característica particular de bancos NoSQL é que, com base no teorema de CAP (*Consistency, Availability and Partition tolerance*), não é sempre que a consistência do banco pode ser mantida. Esse teorema explica que só dois dos estados podem ser alcançados ao mesmo tempo e para a realidade são priorizadas a disponibilidade e tolerância à partição. Logo, as propriedades do A.C.I.D não são respeitadas ao mesmo tempo. Então, faz-se necessário haver um planejamento para que o sistema possa tolerar inconsistências temporárias com o objetivo de priorizar a disponibilidade.

3.3 Modelo de Dados Banco de Dados Não Relacional

Existem alguns modelos ofertados no mercado para a escolha de qual o melhor se encaixa com as necessidades da empresa. Cada um desses modelos possui opções de ferramentas que os têm como base para o armazenamento e tratamento dos dados recebidos (Figura 1). Os mais conhecidos são:

- Chave-valor (*key-value*);
- Banco de Dados Orientado a Documento;
- Orientado a Grafos;
- Orientado a Coluna (*column family*);

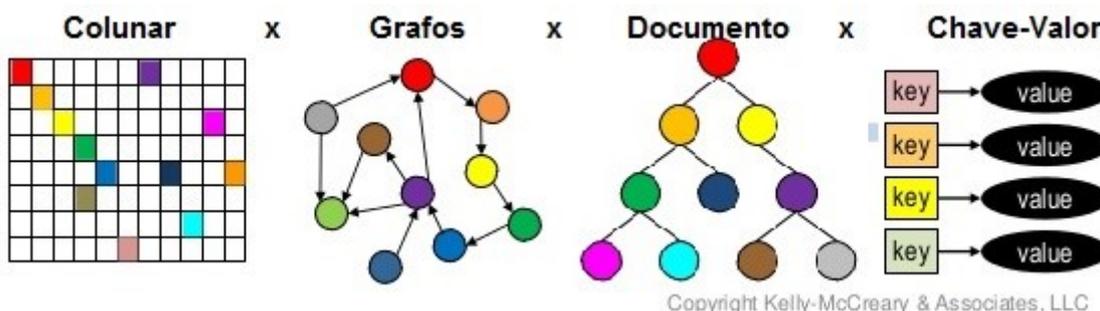


Figura 1 - Modelos de dados de bancos NoSQL

O modelo de Chave-valor (*key-value*) é conhecido por ser simples e permite a sua visualização através de uma tabela de hash, no qual há uma chave única e um indicador de determinado dado, podendo ser uma String ou um binário(KOKAY, 2012).

Quanto ao modelo Orientado a Documento, este modelo armazena coleções e documentos. Isto é, um documento normalmente é um objeto identificador único e um conjunto de campos que podem ser strings, listas ou documentos aninhados. Ao contrário do que foi apresentado no banco de dados chave-valor onde se cria uma única tabela *hash*, aqui temos um agrupamento de documentos sendo que em cada um destes documentos temos um conjunto de campos e o valor deste campo (KOKAY, 2012).

Já o modelo Orientado a Grafos tem três componentes básicos: nós (vértices dos grafos), os relacionamentos (arestas) e as propriedades (conhecidos também como atributos). Visto como multigrafo rotulado e direcionado, onde cada par de nós pode ser conectado por mais de uma aresta. A utilização deste modelo é muito útil quando é necessário fazer consultas demasiadamente complexas (KOKAY, 2012).

Dentre estes, o presente trabalho abordará o Modelo Orientado a Coluna, de modo mais aprofundado, exemplificando com o banco de dados não relacional Hadoop, também objeto deste trabalho e descrito na próxima seção.

4 HADOOP

4.1 Modelo de dados do Hadoop

O Modelo de Dados Orientado a Coluna foi inicialmente desenvolvido com base no Big Table do Google. Uma importante vantagem desse modelo é a sua capacidade de armazenar e processar uma quantidade grande de dados em diversas máquinas. Aqui existem as chaves, mas neste caso elas apontam para atributos ou colunas múltiplas. Os dados são indexados por uma tripla (coluna, linha e *timestamp*), a coluna e linha são identificadas por chaves e o *timestamp* habilita a diferenciação de múltiplas versões de um mesmo dado.

As colunas são distribuídas por família da coluna. Dando destaque que as operações de escrita e leitura são atômicas, então, os valores associados a uma linha são considerados em sua execução, independente das colunas que estão

sendo lidas ou escritas. O conceito utilizado para este modelo é o de família de colunas, com a intenção de combinar colunas que armazenam o mesmo tipo de informação.

Como a busca neste tipo de banco de dados é atômica, mesmo que o interesse seja buscar uma informação específica de uma linha, todas as colunas serão retornadas quando esta mesma linha for consultada. Este modelo permite ainda o particionamento de dados, oferecendo forte consistência.

Tabela			
Família de coluna 1		Família de coluna 2	Família de coluna 3
Coluna 1	Coluna 2	Coluna 3	Coluna 4
#1 {Chave: Valor, Chave: Valor}			
#2 {Chave: Valor, Chave: Valor}			

Figura 2 - Modelo de dados orientados a colunas (Tipos de bancos de dados NoSQL. Micreiros.com, 2017. Disponível em: <<http://micreiros.com/tipos-de-bancos-de-dados-nosql/>>). Acesso em: 16 de maio de 2019.

O uso do Hadoop está com foco, principalmente, para redução de mapas e Jobs. O Hadoop armazena os dados em disco e utiliza a técnica de replicação para garantir tolerância a falhas. Ele foi projetado para funcionar desde um único servidor até um cluster com milhares de máquinas. Entendemos que ele é uma solução concebida para tratar falhas na camada de aplicação, fornecendo um serviço de alta disponibilidade baseado em um grid de computadores. No entanto o Hadoop possui uma grande latência para as consultas. Há dois componentes principais do Hadoop, que abordaremos no próximo ponto, são estes: *Hadoop Distributed File System* (HDFS) e o *MapReduce*.

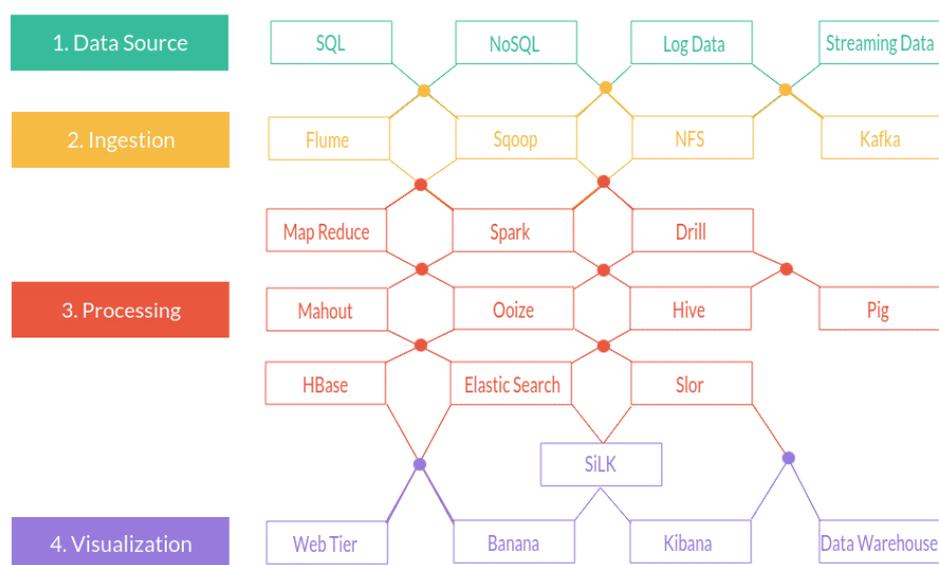


Figura 3 - Modelo de dados Hadoop (Tipos de bancos de dados NoSQL. Micreiros.com, 2017. Disponível em: <<http://micreiros.com/tipos-de-bancos-de-dados-nosql/>>). Acesso em: 16, maio de 2019.

4.2 HDFS e MapReduce

4.2.1 HDFS (Hadoop Distributed File System)

O Hadoop foi projetado para armazenar grandes massas de dados, do porte de terabytes e petabytes, e usa o HDFS como sistema de armazenamento. O HDFS possibilita a conexão de nós (computadores pessoais padrão) contidos nos clusters através dos quais os arquivos de dados são distribuídos. É possível acessar e armazenar os arquivos de dados como um sistema de arquivos contínuo. O acesso aos arquivos de dados é gerenciado de um modo em fluxo, o que quer dizer que aplicativos ou comandos são executados por meio do modelo de processamento

MapReduce, que veremos mais à frente. O HDFS é tolerante a falhas e proporciona acesso de alto rendimento a grandes conjuntos de dados.

Um diferencial notável, entre os outros sistemas de arquivos, é o modelo WORM (*write-once-read-many*) do HDFS que abrandava as exigências do controle de simultaneidade, simplifica a persistência de dados e habilita acesso de alto rendimento. Outra característica única do HDFS é o conceito que, normalmente, é melhor localizar a lógica de processamento próxima dos dados, ao invés de mover os dados para o espaço do aplicativo. O HDFS delimita a gravação dos dados rigorosamente a um gravador por vez. Os bytes são sempre anexados ao final do fluxo e há a garantia de que os fluxos de bytes serão armazenados na ordem gravada.

4.2.2 MapReduce

MapReduce é uma técnica de processamento e um modelo de programa para computação distribuída baseada em *Java*. O algoritmo *MapReduce* se divide em duas tarefas importantes que são mapear e reduzir. O mapa busca um conjunto de dados e os converte em outro conjunto de dados, onde os elementos individuais são divididos em tuplas (pares chaves / valor). Num segundo momento, reduzir tarefa, que recebe a saída de um mapa como uma entrada e combina essas tuplas de dados em um conjunto menor de tuplas. Assim como diz o nome *MapReduce*, a redução é sempre executada após o mapa.

A principal vantagem do *MapReduce* é ser fácil de escalar o processamento de dados em vários nós de computação. Sob este modelo, as primitivas de processamento de dados são chamadas de mapeadores e redutores. A decomposição, de um aplicativo de processamento de dados em

mapeadores e redutores, às vezes, não é tão simples assim. Mas, uma vez que escrevemos um aplicativo no formulário *MapReduce*, dimensionar o aplicativo para executar centenas, milhares ou até dezenas de milhares de máquinas em um cluster é simplesmente uma mudança na configuração. Essa escalabilidade simples é o que atraiu muitos programadores a usar o modelo *MapReduce*.

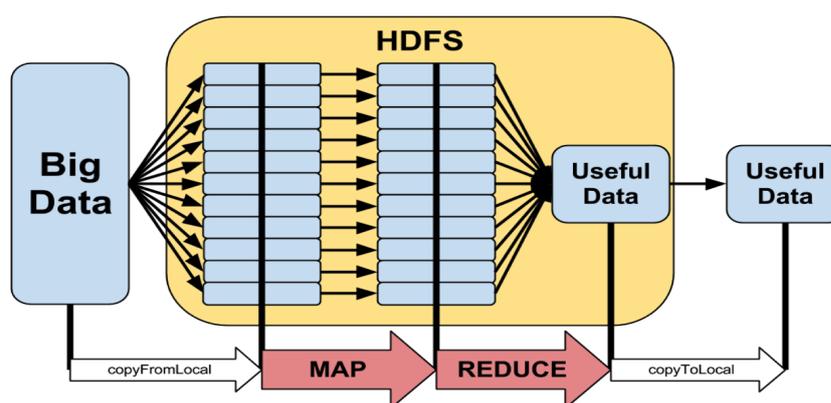


Figura 4 - Modelo de funcionamento do HDFS e MapReduce (Conceptual Overview of Map-Reduce and Hadoop. Glennklockwood.com, 2015. Disponível em: <<https://www.glennklockwood.com/data-intensive/Hadoop/overview.html>>). Acesso em: 16 de maio de 2019.

4.3 Outros componentes

É importante entendermos que para ter esse alto rendimento em relação a processamento de *Big Data*, o Hadoop conta com outros componentes em sua arquitetura para auxiliá-lo. A seguir, apresenta-se destaques e se faz uma breve explicação desses componentes:

- *ZooKeeper*: é um serviço de coordenação distribuída para gerenciar grandes conjuntos de Clusters.
- *Oozie*: é um sistema de agendamento de *WorkFlow*, usado para

gerenciar principalmente os Jobs de *MapReduce*.

- *Pig*: é uma linguagem de procedimentos de alto nível para consultar grandes conjuntos de dados semiestruturados usando Hadoop e a Plataforma *MapReduce*
- *Sqoop*: é um projeto do ecossistema Hadoop, cuja responsabilidade é importar e exportar dados do banco de dados de dados relacionais.
- *Spark*: é uma ferramenta *Big Data* para o processamento de grandes conjuntos de dados. Foi desenvolvido para substituir o *MapReduce*, pois processa 100x mais rápido que o *MapReduce*.
- *HBase*: é um banco de Dados não relacionais, projetado para trabalhar com grande conjunto de dados (*Big Data*). É o banco de dados oficial do Hadoop.
- *Flume*: é um serviço que permite enviar dados diretamente para o HDFS. É um serviço que funciona em ambiente distribuído (em *cluster*) para coletar, agregar e mover grandes quantidades de dados de forma eficiente.
- *Mahout*: é dedicado a *Machine Learning – Data Science*. Ele permite a utilização dos principais algoritmos de clustering, testes de regressão e modelagem estatística.
- *Kafka*: é foi desenvolvido pelo *Linkedin* e liberado como projeto *Open-source* em 2011. O *Apache Kafka* é um sistema para gerenciamento de fluxo de dados em tempo real, gerados a partir de websites, aplicações e sensores.
- *Ambari*: tem como objetivo tornar o gerenciamento do Hadoop mais simples. O *Ambari* fornece uma interface de usuário da *Web* de gerenciamento do Hadoop intuitiva e fácil de usar.

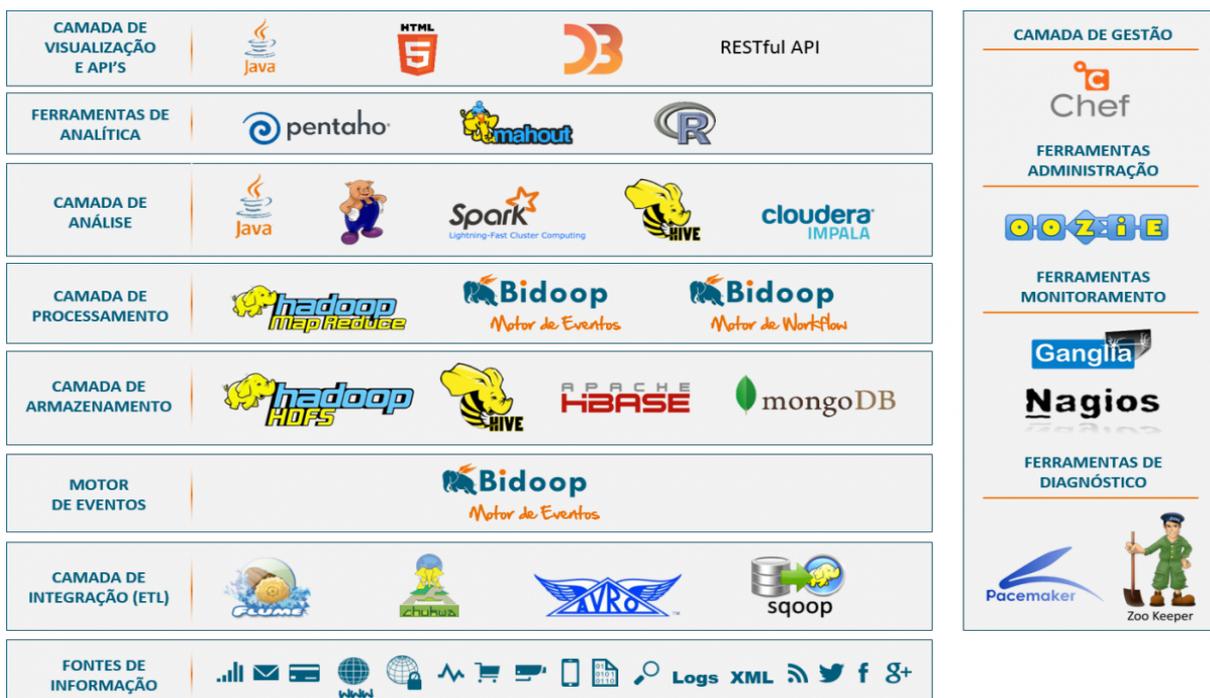


Figura 5 - Aplicações que compõem o Ecossistema Hadoop (Apache Hadoop: Tudo o que você precisa saber. Cetax.com.br. Disponível em: <<https://www.cetax.com.br/apache-Hadoop-tudo-o-que-voce-precisa-saber/>>). Acesso em: 16, maio de 2019.

5 Resultados

Com base no que foi explanado na seção anterior, é possível entender que o Hadoop tem várias características que podem vir a atender uma significativa lacuna para tratar os dados não estruturados, dando escalabilidade e, conseqüentemente, desempenho para que seja aplicados algoritmos que permitam os usuários explorarem e analisarem os dados para obterem *insights* e assim tomarem decisões baseadas nos dados ali armazenados.

Essas características nos levam ao resultado de que o Hadoop é uma ferramenta que podem agregar muito no valor da empresa e, principalmente, se esta organização tem seu foco na orientação a dados. Juntamente com o uso do *Machine*

Learning e da Inteligência Artificial, esta ferramenta é capaz de entregar um alto desempenho em análise de dados.

Se pensarmos nessa arquitetura voltada para os vários tipos de dados e a quantidade que a empresa necessita para fazer suas análises, o Hadoop atenderia empresas de médio e grande porte sem nenhuma dificuldade. Sua atuação lhe torna uma opção muito viável, já que esta flexibilidade faz dele uma ferramenta com um baixo custo, tendo em vista o seu alto retorno para as organizações.

6 Considerações Finais

De acordo com o que podemos observar, o Hadoop tem características e componentes que forma sua arquitetura com foco para atender as necessidades de armazenamento e processamento de dados. Esses fatores fazem dele uma ferramenta versátil e extremamente útil para as empresas que desejam ou têm seu foco orientado a dados.

Chegamos por fim ao resultado de que, o Hadoop é uma plataforma que tem recursos e características que podem vir a auxiliar as organizações como uma parte de sua arquitetura para focar em alto grau de desempenho e poder de processamento e baixo custo.

7 Referências

GABRIEL, M. **Você, eu e os robôs: pequeno manual do mundo digital**. São Paulo: Atlas, 2018.

GARTNER Group. **2019 Planning Guide for Data and Analytics**. Gartner, 2019.

OLAVSRUD, T. **Empresas orientadas a dados são mais competitivas**. Disponível em: <https://cio.com.br/empresas-orientadas-a-dados-possuem-melhor-desempenho/>. Acesso em: 16/05/2019

SALVADOR, CE. **Sete características de empresas orientadas a dados**. Disponível em: <https://computerworld.com.br/2017/01/20/sete-caracteristicas-de-empresas-orientadas-dados/>. Acesso em: 16/05/2019

SOARES, David Jose. **Empresas Orientadas a Dados e Análises: a tecnologia está a serviço da tomada de decisão?**. 2017. 23 f. Trabalho de Conclusão de Curso (Graduação em Gestão da Informação) – Universidade Federal de Uberlândia, Uberlândia, 2017. Acesso em: 16/05/2019

SOARES, Marília Cavalcante. **Banco de dados NoSQL: Um novo paradigma - Revista SQL Magazine 102**. 2012. Disponível em: <https://www.devmedia.com.br/banco-de-dados-nosql-um-novo-paradigma-revista-sql-magazine-102/25918>. Acesso em: 16/05/2019