



**MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO TOCANTINS
CAMPUS PORTO NACIONAL
LICENCIATURA EM COMPUTAÇÃO**

DANIELLE NASCIMENTO DE SANTANA

**SISTEMAS DE GERENCIAMENTO DE BANCOS DE DADOS: Um estudo sobre
sua utilização em empresas que manipulam grandes volumes de dados.**

**Porto Nacional - TO
2018**



**MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO TOCANTINS
CAMPUS PORTO NACIONAL
LICENCIATURA EM COMPUTAÇÃO**

DANIELLE NASCIMENTO DE SANTANA

**SISTEMAS DE GERENCIAMENTO DE BANCOS DE DADOS: Um estudo sobre
sua utilização em empresas que manipulam grandes volumes de dados.**

Trabalho de Conclusão de Curso
apresentado à Coordenação do Curso
de Licenciatura em Computação do
Instituto Federal de Educação,
Ciência e Tecnologia do Tocantins –
Campus Porto Nacional, como
exigência à obtenção do grau de
Licenciado em Computação.

Orientador: Professor Mestre Elvis
Nascimento da Silva.

**Porto Nacional - TO
2018**

FICHA CATALOGRÁFICA

DANIELLE NASCIMENTO DE SANTANA

SISTEMAS DE GERENCIAMENTO DE BANCOS DE DADOS: Um estudo sobre sua utilização em Redes Sociais e Similares.

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Licenciatura em Computação do Instituto Federal de Educação, Ciência e Tecnologia do Tocantins – *Campus Porto Nacional*, como exigência à obtenção do grau de Licenciado em Computação.

Aprovado em: ____/____/____

BANCA AVALIADORA

Professor Mestre Elvis Nascimento da Silva – Orientador
IFTO – Porto Nacional

Professora Mestre Ordália Dias Da Silva Guilherme
IFTO – Porto Nacional

Professor Mestre Renato de Oliveira Bastos
IFTO – Porto Nacional

Dedico esse trabalho aos meus filhos: Ana Luísa, Joaquim Júnior e Maria Júlia que sempre me acompanharam (literalmente) e me motivaram a permanecer estudando.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me dar força e coragem para prosseguir em mais uma etapa da minha vida.

Ao meu esposo e aos meus filhos, que sempre estiveram ao meu lado e me motivaram a continuar nessa caminhada.

Ao meu orientador, Prof. Elvis Nascimento da Silva e sua esposa Lucielle Ferreira Alves, pela orientação, paciência, dedicação, e principalmente amizade na elaboração deste trabalho. Se não fossem suas palavras de motivação, as madrugadas a fio, a hospedagem em sua residência, seu acolhimento, nada disso seria possível.

Aos demais professores e mestres que tive o prazer de conhecer ao longo do Curso de Licenciatura em Computação, sentirei muita saudade.

A todos os meus amigos pelos momentos de convívio, apoio e incentivo. A todos que, de alguma forma, me ajudaram e fizeram parte dessa conquista.

"O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo. Mesmo não atingindo o alvo, quem busca e vence obstáculos, no mínimo fará coisas admiráveis."

José de Alencar

RESUMO

A presente Monografia traz a realização de um estudo acerca da utilização de Sistemas de Gerenciamento de Bancos de Dados (SGBD) em instituições que manipulam grande volume de dados, como as Redes Sociais. Baseando-se na revisão bibliográfica buscou-se uma base dos principais autores que abordam os temas dos tipos de Bancos de Dados como também o seu armazenamento, para assim estabelecer uma compreensão dos objetivos propostos. Através de pesquisa de campo, objetivou verificar quais são as implementações das arquiteturas dos SGBD's utilizados por empresas como Facebook, Dropbox e Wikipedia. A metodologia utilizada foi descritiva, quantitativo-qualitativa, as informações extraídas foram embasadas de acordo com a temática abordada, utilizando um questionário enviado para organizações que foram selecionadas de maneira aleatória. Como resultados observou-se que o SGBD Relacional MySQL ainda é o modelo mais utilizado, embora nos projetos implementados utilizam-se grandes porções de dados, através do conceito de Big Data e no paradigma da computação em nuvem. No entanto, há um desafio em lidar em manipular os dados de um sistema, quando se trata desse modelo. Os SGBD's NoSQL (Não somente relacionais) surgem como solução de modelagem que provê de forma mais simplificada, ganhos de desempenho (processamento), um grau elevado de escalabilidade, flexibilidade e disponibilidade. O Cassandra surge com destaque, como um modelo distribuído baseado no *BigTable* do *Google*. Os grandes centros de dados (Data Centers) encontram-se distribuídos mundialmente, com destaque na costa Norte Americana, parte da Europa e em Singapura, localizada no sudoeste da Ásia.

Palavras-Chave: SGBD's. Banco de dados. BIG DATA.

ABSTRACT

This monograph presents a study about the use of Database Management Systems (DBMS) in institutions that manipulate a large volume of data, such as Social Networks. Based on the bibliographic review, a search was made for a database of the main authors that deal with the subjects of the types of Data Banks as well as their storage, in order to establish an understanding of the proposed objectives. Through field research, the objective was to verify the implementations of the DBMS architectures used by companies such as Facebook, Dropbox and Wikipedia. The methodology used was descriptive, quantitative-qualitative, the information extracted was based on the subject matter, using an for organizations that were randomly selected. As a result, it was observed that the MySQL Relational DBMS is still the most used model, although in the implemented projects large portions of data are used, through the concept of Big Data and in the paradigm of the cloud computing. However, there is a challenge in dealing with manipulating a system's data when it comes to this model. NoSQL (not just relational) DBMSs emerge as a simplified modeling solution, performance (processing) gains, a high degree of scalability, flexibility and availability. Cassandra comes out prominently as a distributed model based on Google's Big Table. The large data centers are distributed worldwide, with emphasis on the North American coast, part of Europe and Singapore, located in southwest Asia.

Keywords: DBMS. Database. BIG DATA.

LISTA DE ILUSTRAÇÕES

Quadro 1- Dado x Informação	16
Quadro 2 - Tipos de Bancos de Dados	24
Quadro 3 - Relação entre ACID x BASE	30
Figura 1 - Sistemas com arquivos próprios	17
Figura 2 - Exemplo de Sistemas com o uso de BD	19
Figura 3 - Estrutura Geral de um SGBD.....	21
Figura 4 - Teorema CAP	28
Figura 5 - Etapas de elaboração da trabalho.	Erro! Indicador não definido.
Figura 6 - Localização geográfica dos Datacenters	42
Gráfico 1- SGBD's SQL e NoSQL	40
Gráfico 2- Linguagens de programação utilizadas.....	41

LISTA DE TABELAS

Tabela 1- Campos da planilha para extração de informações dos artigos	35
Tabela 2 - Base de Dados e Linguagens Utilizadas.....	38

LISTA DE ABREVIATURAS E SIGLAS

ACID	Atomic, Consistent, Isolation and Durable
BD	Banco de Dados
BDOO	Banco de Dados Orientado a Objeto
DBMS	Database Management System
ERP	Enterprise Resource Planning,
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
MRP	Material Requirement Planning
MYSQL	My Structured Query Language
NoSQL	Not Only SQL
OLAP	Online Analytical Processing
RDBMS	Relational Database Management System
SGBD	Sistemas de Gerenciamento de Banco de Dados
SGBD-R	Sistemas Gerenciadores de Bancos de Dados Relacionais
SQL	Structured Query Language
SUS	Sistema Único de Saúde
TI	Tecnologia da Informação
XML	eXtensible Markup Language

SUMÁRIO

1. INTRODUÇÃO	13
1.1. Organização da Monografia	14
2. REFERENCIAL TEÓRICO	15
2.1. Dado x Informação	15
2.2. Preservação da Informação	16
2.3. Origem dos Bancos de Dados	16
<i>2.3.1. Sistemas de Gerenciamento de Banco de Dados</i>	<i>19</i>
2.4. Evolução dos Bancos de Dados	22
<i>2.4.1. Modelos de Bancos de Dados</i>	<i>23</i>
2.5. Bancos de dados nas Organizações	32
3. METODOLOGIA	34
3.1. Procedimentos Metodológicos	34
3.2. Tipos de Pesquisa	34
4. RESULTADOS	38
5. CONCLUSÃO	46
5.1. Limitações	46
5.2. Trabalhos Futuros	47
REFERÊNCIAS	48
APÊNDICES	52
Apêndice A – Modelo do e-mail enviado para as empresas	52

1. INTRODUÇÃO

O crescimento exponencial das redes de computadores e serviços de internet nos diversos setores faz com que cada vez mais tenhamos dados de vital importância trafegando sob a competência de novas tecnologias, envolvendo diferentes áreas como: telefonia móvel, redes de computadores domésticas e corporativas, serviços via satélite etc. Utiliza-se de serviços e recursos computacionais de forma remota em pontos antes nunca alcançados através de dispositivos portáteis e nômades como PDAs¹, computadores pessoais, palmtops e aparelhos celulares. Toth (2011) afirma que nas grandes aplicações Web, desktops, ou até mesmo móveis, está-se a tornar cada vez mais comum lidar com grande volume de dados, sendo que no futuro a situação será mais exigente em termos de armazenamento de dados.

A motivação para a realização desse trabalho veio inicialmente da curiosidade da pesquisadora em estudar como algumas instituições lidam com a grande quantidade de informações geradas por usuários, no que diz respeito às formas de armazenamento, com o foco inicial direcionado para as Redes Sociais, tendo em vista que milhares de dados como fotos, vídeos, arquivos de áudio, são gerados e compartilhados a todo momento. A partir daí definiu-se o tema: “Sistemas de Gerenciamento de Bancos de Dados: Um estudo sobre sua utilização em empresas que manipulam grandes volumes de dados.

O Objetivo Geral da pesquisa é verificar quais são as implementações das arquiteturas dos Bancos de Dados utilizadas por instituições que lidam com amplas massas de dados.

Como objetivos específicos apresentam-se:

- Identificar as bases de dados utilizadas nas instituições pesquisadas;
- Mapear as iniciativas de armazenamento de dados das referidas empresas, na esfera mundial;
- Mostrar as características (singularidades) das soluções encontradas.

Em consonância com os objetivos, buscam-se respostas para as seguintes indagações:

¹ *Personal Digital Assistant* – Tipo de computador de bolso.

Q1: Quais os modelos de Bancos de dados implementados em empresas que trabalham com grandes volumes de dados?

Q2: Onde estão localizadas/armazenadas, no cenário mundial, as informações desses bancos de dados?

Q3: Quais as peculiaridades das implementações encontradas?

Do ponto de vista acadêmico e profissional esse trabalho tem importância, tendo em vista que a pesquisa possui respostas absolutas e de cunho científico, sendo, portanto de relevância para os alunos, para a academia e para a comunidade científica em geral, pois deixará registros do estudo realizado que servirão de base para outras pesquisas na área, bem como poderá ser utilizado também por desenvolvedores de softwares interessados em saber quais soluções de armazenamento estão sendo implementadas por algumas empresas que trabalham com grande volume de dados, podendo assim, ter uma referência das tendências de mercado.

Do ponto de vista pessoal é inegável a satisfação da pesquisadora em ter a possibilidade de deixar suas contribuições para estudos e trabalhos futuros.

1.1. Organização da Monografia

Esta monografia encontra-se organizada em V capítulos. A seguir, apresenta-se uma breve descrição de cada um dos capítulos:

Capítulo I: Introdução: dissertar sobre a apresentação do tema, que fornece uma contextualização do assunto e os objetivos desta pesquisa.

Capítulo II: Referencial Teórico: Apresenta os conceitos relacionados ao objeto de estudo;

Capítulo III: Metodologia: apresenta procedimentos metodológicos, Universo da pesquisa.

Capítulo IV: Resultados do desenvolvimento e da pesquisa.

Capítulo V: Conclusão, Limitações e Trabalhos Futuros.

2. REFERENCIAL TEÓRICO

Neste capítulo são abordados os principais conceitos que serviram de base para o desenvolvimento da pesquisa.

Com a finalidade de alinhar esta pesquisa ao cenário da atualidade no que diz respeito ao armazenamento de dados em instituições que lidam com grande volume destes, e mostrar o que dizem os principais teóricos relacionados ao tema, a fundamentação abordará os seguintes assuntos: Dados x Informação; Preservação da informação; Origem e evolução dos Bancos de Dados; Tipos de Bancos de Dados (conceitos e características).

2.1. Dado x Informação

Antes de adentrar-se no universo dos Bancos de Dados, é importante reconhecer a diferença entre dados e informação. Para Stair (2011), a informação é um dos recursos valiosos em uma organização. Além de fornecer um conjunto de fatos que, organizados, fornecem subsídios para administração, por ser baseada na representação real de dados obtidos sobre um determinado assunto ou fato. Já o dado, é um elemento que mantém a sua forma bruta, que sozinho não dá uma compreensão de uma determinada situação, ou seja, o termo “Dado” envolve fatos, imagens, sons que podem ou não serem úteis para determinado fim. Ainda de acordo com Stair (2011), a informação não pode ser confundida com dados, pois estes representam fatos de um mundo real, como por exemplo, total de horas trabalhadas em uma semana, número de produtos em estoque, quantidade do produto vendido no dia.

Baseada na literatura sabe-se que a informação é quando esses dados são organizados de forma significativa e pode ser muito valiosa por contribuir para que pessoas e suas organizações possam desempenhar tarefas de forma mais eficiente e eficaz, tomando decisões e alcançando metas. O **Quadro 1** ilustra a informação obtida através de dados de acordo com (SANTANA, 2018).

Quadro 1- Dado x Informação

Dado	Informação
Rendimento Familiar	Financiamentos; Programas Sociais; etc.
Altura, Peso	Índice de Massa Corporal (IMC); Serviço Militar; etc.
Data de Nascimento	Idade; menor ou não de idade; etc.

Fonte: SANTANA (2018).

2.2. Preservação da Informação

O uso do computador facilita a produção de um grande volume de dados e documentos que podem desaparecer na mesma velocidade em que são produzidos. Segundo Innarelli (2006) isto ocorre por que:

“(...) a humanidade ainda não tem prática e nem experiência para preservar a memória digital. Memória que está sendo perdida a cada dia em virtude da obsolescência das tecnologias, da deterioração das mídias digitais e principalmente pela falta de políticas de preservação digital.”

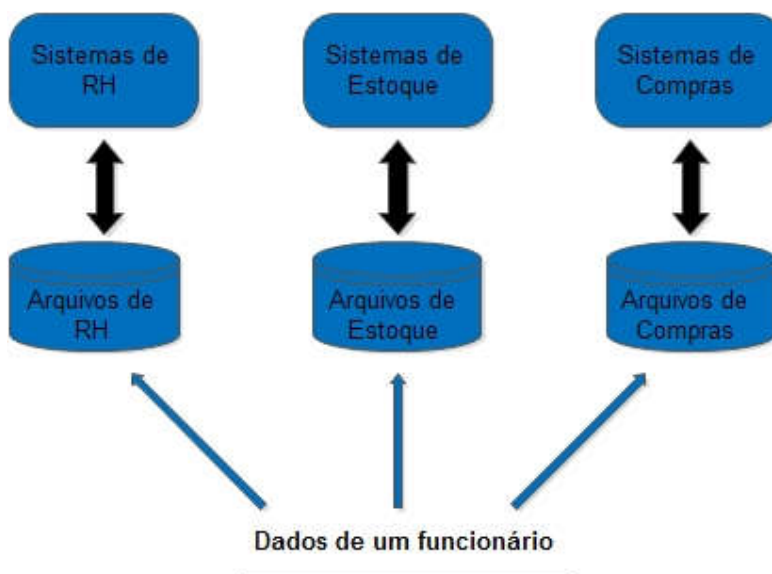
Na atualidade, a preservação digital é um dos grandes desafios enfrentados por profissionais da informação, pois o material digital carrega consigo problemas estruturais que colocam em risco sua longevidade(FERREIRA, 2006).

O material digital, apesar das imensas vantagens inerentes à sua aplicação, acarreta um problema estrutural que coloca em risco a sua longevidade (FERREIRA, 2009). O documento digital pode ser copiado infinitas vezes sem perda de qualidade, mas precisa de um contexto tecnológico para que possa ser interpretado. Essa dependência torna-o particularmente vulnerável à rápida obsolescência a que a tecnologia está sujeita, é ditada pela volatilidade do mercado informático (BARBEDO et al., 2007), Outros autores como (CAMARGO; BELLOTTO, 2005)salientam que nunca é demais destacar que fontes históricas jamais serão reencontradas se forem destruídas. Como também Innarelli (2006) complementa que a perda destas informações pode deixar uma grande lacuna histórica, uma vez que muitas pesquisas podem deixar de existir por não haver material suficiente para consulta.

2.3. Origem dos Bancos de Dados

Antes da existência dos bancos de dados, qualquer sistema, programa ou aplicação que precisasse armazenar e manipular dados fazia uso de um sistema de arquivos. Ou seja, cada sistema, programa ou aplicação desenvolvido tinha seus próprios arquivos de armazenamento dos dados (SOUZA, 2013). Os programas eram escritos em respostas às necessidades do momento e, geralmente, por programadores diferentes. O problema dessa situação é que, desenvolvedores com metodologias de trabalho diferentes, pensam e programam de forma diferente. Ou seja, entre os diversos sistemas desenvolvidos gradualmente para uma determinada empresa ou instituição poderia haver dados em comum em vários bancos de dados(SIEBRA, 2010). A Figura 1 apresenta um exemplo de um ambiente em que existem sistemas com seus próprios arquivos.

Figura 1 - Sistemas com arquivos próprios



Fonte: Adaptado(SIEBRA, 2010)

Nessa organização, seria necessário um processo de replicação de dados em cada sistema de arquivos, o que gera redundância e inconsistência de dados. O uso de sistemas de arquivos possuiu problemas tais como: a) Isolamento dos dados, em que os dados estão armazenados em arquivos distintos, que não possuem relações, e ainda, podem conter dados com formatos distintos; b) Problemas com acesso concorrente, em que a consistência dos dados é

comprometida, uma vez da dificuldade de permitir que múltiplos usuários possam ter acesso aos dados de forma simultânea(SIEBRA, 2010).

Assim temos o conceito de BD (Banco de dados) é um sistema de manipulação de registro com o propósito geral de manter os dados e torná-lo disponível sempre que necessário, normalmente armazena os dados relacionados a um sistema de computador (FOSTER; GODBOLE, 2016). Sarkis (2001) afirma que existem dois tipos de BD: os operacionais (transacionais) e os analíticos. Os operacionais são usados no dia, quando existir a necessidade de coletar, armazenar e modificar dados. Têm a característica de armazenar dados dinâmicos que se modificam constantemente e refletem a informação instantaneamente. Como exemplo, podemos citar o caso de controle de estoques, em que o estoque atual deve ser preciso e representar fielmente a quantidade de produtos existentes fisicamente a todo o instante. Já os analíticos, manipulam multidimensões. O autor ainda ressalva que os Sistemas Analíticos atuam na análise de dados e na tomada de decisões. Eles possuem um enfoque diferente do operacional, eles têm frequentemente um campo de ação distinta. Enquanto os sistemas operacionais necessitam ser normalmente enfocados em uma única área, os analíticos necessitam frequentemente ser estendidos a várias áreas diferentes e necessitam de grandes quantidades de dados operacionais relacionados.

Para Sarkis (2001), esses Bancos de Dados possuem a característica de não serem voláteis, eles não atualizam continuamente as informações, e as mantêm como snapshot (instantâneo) de dados, que é um registro específico do tempo, ocasionando um grande armazenamento de dados históricos. Essa característica define esse tipo de banco de dados como somente de leitura, diferente dos bancos operacionais (escrita e leitura) que executam as atividades básicas de inserção, atualização, consulta e deleção de dados.

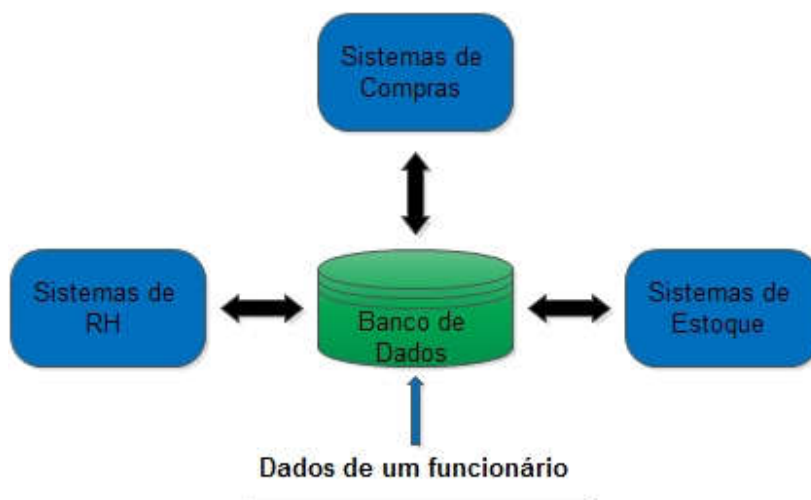
Um exemplo de sistemas analíticos, segundo Cougo (2013), são as ferramentas OLAP², voltadas para a análise analítica dos dados, diferente da operação com dados transacionais, trata da capacidade de analisar grandes volumes de informações nas mais diversas perspectivas dentro de um

²Software cuja tecnologia de construção permite aos analistas de negócios, analisar e visualizar dados corporativos de forma rápida, consistente e principalmente interativa.

DataWarehouse³. Nesse formato, a informação não é volátil, existe um denso histórico de dados e as consultas SQL's são focadas em desempenho extremo, voltada a auxiliar as organizações na tomada de decisão.

A Figura 2 exemplifica esse tipo de sistema.

Figura 2 - Exemplo de Sistemas com o uso de BD



Fonte: Adaptado(SIEBRA, 2010)

Com os BD's, os dados só necessitam ser armazenados uma única vez e passam a ser acessados por todos os sistemas (SIEBRA, 2010).

2.3.1. Sistemas de Gerenciamento de Banco de Dados

De acordo com O'brien (2007), SGBD é um “conjunto de programas de computador que controla a criação, manutenção e uso dos bancos de dados por uma organização e seus usuários finais”. Nessa mesma linha, Siebra (2010), afirma que um SGBD é uma coleção de programas que habilitam usuários a criar e manter um banco de dados. Ele é um software de propósito geral, que facilita o processo de definição, construção e manipulação de um banco de dados:

- Definição: Envolve especificar estruturas e tipos de dados para serem gravados no banco de dados, com uma descrição detalhada de cada tipo de dado;

³ Depósito de dados digitais que serve para armazenar informações detalhadas de uma empresa ou organização, criando e organizando relatórios através de históricos que são depois utilizados pela empresa para ajudar a tomar decisões importantes com base nos fatos apresentados.

- Construção: Processo de consistir e gravar inicialmente dados no banco de dados;
- Manipulação: Inclui funções como consulta por dados específicos e atualização para refletir as alterações no mundo real.

Alguns autores defendem que a finalidade dos SGBD's é implementar uma visão do BD para os usuários; criar e alterar as estruturas de dados; armazená-los fisicamente, manipulá-los e fornecer a geração de relatórios, portanto é um sistema de software de propósito geral que facilita os processos de definição, construção, manipulação e compartilhamento de bancos de dados entre vários usuários e aplicações (ELMASRI; NAVATHE, 2005).

De acordo com Laudon (1999), as vantagens da utilização de um SGBD são:

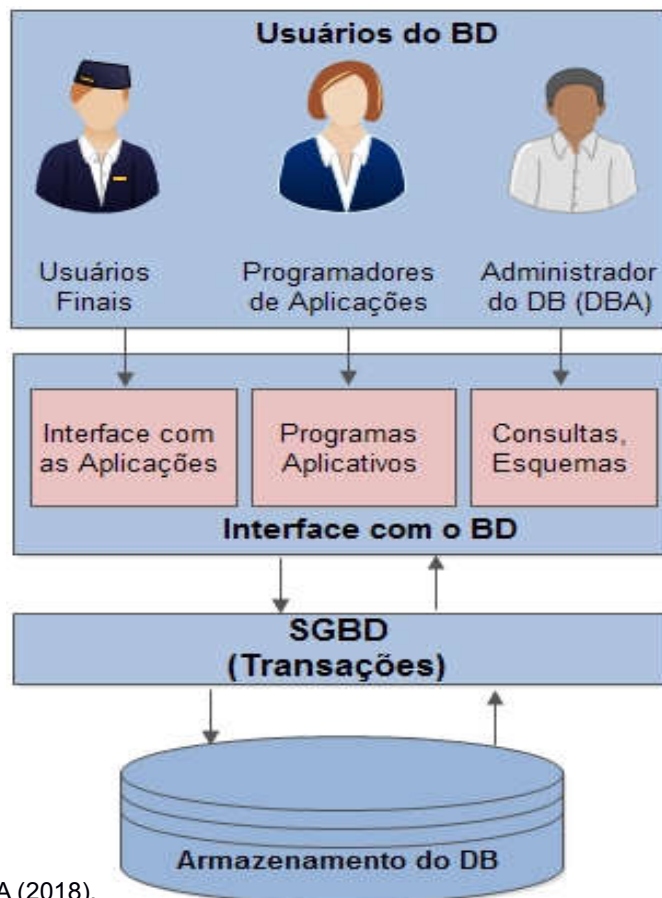
- Os dados são independentes dos programas aplicativos. O SGBD distingue entre visões lógicas e físicas de dados de modo que muitos programas aplicativos podem utilizar dados de um mesmo banco de dados compartilhado;
- A redundância e a inconsistência de dados são reduzidas. Como os dados são independentes dos programas aplicativos, não há necessidade de construir arquivos isolados em que os mesmos elementos de dados são repetidos cada vez que uma nova aplicação é chamada. Os dados são mantidos em apenas um lugar;
- A complexidade é reduzida pelo gerenciamento consolidado de dados, acesso e utilização por meio do SGBD;
- As informações são mais fáceis de acessar e usar. O banco de dados estabelece relacionamentos entre diferentes fragmentos de informações. Os dados de diferentes registros e aplicações podem ser mais facilmente acessados e combinados.

Esses recursos possibilitam que múltiplos usuários possam manipular a mesma base de dados simultaneamente de forma com que os dados permaneçam íntegros, e facilita o processo de criação dos sistemas por parte dos desenvolvedores. Os SGBD's oferecem outros recursos, como garantias de integridade dos dados, recuperação de falhas, segurança, controle de concorrência, controle de transações, dentre outros (LAUDON, 1999). No que diz respeito a transações, Siebra (2010) afirma que Controle de Transações é um conjunto de operações sobre oBD que devem ser executadas integralmente e sem falhas ou interrupções. Caso contrário, devem ser desfeitas, e que Transação é um grupo de operações que desempenha uma função lógica única dentro de uma aplicação do BD. O autor ainda assegura que uma transação deve ter as seguintes propriedades que formam a sigla ACID:

- A – Atomicidade: ou todas as operações envolvidas na transação ocorrem, ou nenhuma delas deve ter efeito sobre o banco de dados. Assegurá-la é tarefa do SGBD.
- C – Consistência: ao final da execução da transação a consistência dos dados no banco deve ter sido mantida. Assegurá-la é tarefa do programador.
- I – Isolamento: uma transação deve ter sua execução realizada de forma isolada em relação à execução de outras transações.
- D – Durabilidade: depois que uma transação é executada com sucesso, as modificações por ela realizadas devem ser mantidas no sistema, mesmo na ocorrência de falhas. Assegurá-la é tarefa do SGBD.

E reforça que qualquer sistema não está totalmente seguro, no conceito de transações, em muitas aplicações é importante assegurar que, uma vez detectado um problema, o dado deve ser salvo em seu último estado . Um exemplo disso suponha que um cliente de uma instituição financeira deseja transferir um valor de R\$ 20,00 (vinte reais) de sua conta para a outra conta. Caso ocorra uma falha durante o procedimento, é possível que o valor seja debitado da conta do usuário, e não creditado na outra conta destino. Isso cria um estado de inconsistência no BD. Assim, é importante, através das Transações de um BD, que ocorra consistência de dados, ou seja, que os procedimentos da transferência ocorram por completo, ou não ocorram. De acordo com Santana (2018) essa é a representação da Estrutura de um SGBD:

Figura 3 - Estrutura Geral de um SGBD



Fonte: SANTANA (2018).

2.4. Evolução dos Bancos de Dados

O século XIX foi um dos propulsores para o desenvolvimento de formas de armazenamento de dados, devido à enorme quantidade de dados que foram surgindo para serem armazenados (UNIPVIRTUAL, [s.d.]). Um exemplo foi o censo americano realizado em 1880 que levou quase sete anos para ser totalmente reunido a mão. Os engenheiros americanos constataram que devido ao grande crescimento da população, o próximo censo não estaria concluído antes que o outro fosse começar. Para solucionar o problema do censo americano, um engenheiro chamado Herman Hollerith⁴ criou uma forma de armazenar os dados do censo em cartões perfurados. Com a utilização do equipamento de Hollerith a contagem do censo de 1890 foi realizada em apenas um mês.

⁴ Hermann Hollerith - funcionário do *United States Census Bureau* - inventou, em 1880, uma máquina para realizar as operações de recenseamento da população.

De acordo com o artigo (EARLY OFFICE MUSEUM, 2000), em 1896 Hollerith formou a *TabulatingMachineCompany* com o intuito de produzir e comercializar o seu equipamento. A combinação de sua companhia com várias outras originou a que hoje chamamos de IBM.

Com o passar do tempo, as formas de armazenamento foram evoluindo de maneira constante: Em 1900 utilizavam cartões perfurados como meio de armazenamento. Até as décadas de 1920, 30 e 40 os cartões perfurados foram o único meio de armazenamento dos dados utilizados nesta época (ELMASRI; NAVATHE, 2005). Ainda contando a história dessa evolução, os autores supracitados contam que no início de 1930a era do armazenamento magnético apagável surgiu através da utilização de fitas magnéticas para armazenamento do som. A utilização da fita magnética acabou originando a criação dos discos magnéticos. De 1950 a 1960 as fitas magnéticas eram unanimidade quando se falava em armazenamento de dados em computadores.

Então em 1960 aconteceu à conversão das fitas magnéticas para os discos magnéticos e o crescimento da capacidade de armazenamento. Iniciava-se o armazenamento de dados computacionais estruturado no formato hierárquico, chamado de banco de dados hierárquico. Já em 1970, começava a era de banco de dados relacional e *Database Management System* (DBMS), com base em cálculo relacional, coleta de dados relacionados logicamente e uma descrição destes dados. Logo em 1980 iniciavam-se os estudos de banco de dados orientados a objetos (BDOO), que visavam criar um tipo de banco de dados que pudesse interagir de maneira mais eficaz com a programação orientada a objeto como: Java, C++ ou *Smalltalk*,⁵ que armazenavam objetos complexos, como textos, imagens e gráficos(ELMASRI; NAVATHE, 2005).

2.4.1. Modelos de Bancos de Dados

Os SGBD's mais utilizados ou atualmente mais implementados em diversos segmentos (Indústria, entretenimento, Mercado Financeiro, Educação, etc.). Para melhor representar uma classificação, Brito (2010), apresenta uma comparação de uma análise de possibilidade de se optar estratégias do chamado SGBD NoSQL em detrimento de um SGBD tradicional (SQL) e que é preciso levar em

⁵Linguagem de programação orientada a objeto.

consideração algumas questões básicas, como, por exemplo, os critérios de escalonamento, consistência de dados e disponibilidade. Há na literatura vários trabalhos que comparam o desempenho de SGBD's SQL e NOSQL, utilizando técnicas de *benchmarks*⁶ entre os modelos, comparando consultas, tempode resposta, armazenamento, etc. Portanto outros autores (CROWTHER; LAKE, 2013) e (POKORNY, 2013) definem características desses dois modelos, como mostra o

Quadro 2 :

Quadro 2 - Tipos de Bancos de Dados

TIPO	CARACTERÍSTICAS
Bancos de Dados Relacionais	O modelo relacional de banco de dados é fundamentado em princípios matemáticos, principalmente da álgebra relacional. É representado por uma coleção de tabelas (entidade) e um conjunto de linhas (tuplas) e uma lista de valores de atributos. Em um banco de dados relacional uma série de tabelas de registros com atributos são ligadas entre si por uma ou mais relações. Essas relações são criadas usando chaves estrangeiras que são atributos que contém os mesmos dados em ambas as tabelas (CROWTHER; LAKE, 2013)
NoSQL ou não Relacionais	Esses sistemas estão sendo denominados de NoSQL, cuja tradução mais aceita no momento é Not Only SQL ou sistemas pós relacionais (POKORNY, 2013) . Grande parte deles não utiliza a SQL como linguagem de consulta e, ao contrário dos sistemas relacionais, não há a necessidade de criação de esquemas rígidos para armazenamento dos dados.

Fonte: SANTANA (2018).

2.4.1.1. Bancos de Dados Relacionais

O modelo relacional de banco de dados é fundamentado em princípios matemáticos, principalmente da álgebra relacional. Representado por uma coleção de tabelas (entidade) e um conjunto de linhas (tuplas), uma lista de valores de atributos. Álgebra relacional consiste em um conjunto de operações sobre as relações, onde cada operação produz uma nova relação a partir de um ou mais relações já existente(COUGO, 2013).

Em um banco de dados relacional uma série de tabelas de registros com atributos são ligadas entre si por uma ou mais relações. Essas relações são criadas

⁶Um benchmark (literalmente, banco de teste) é uma análise comparativa de produtos ou serviços da concorrência para um produto existente ou para um setor de uma determinada atividade no quadro da concepção de um novo produto(CARLOS, 2017).

usando chaves estrangeiras que são atributos que contém os mesmos dados em ambas as tabelas(CROWTHER; LAKE, 2013).

Para a manipulação de dados em um banco de dados foi criado o padrão de linguagem bastante difundido no mercado conhecido pela sigla SQL (*Structured Query Language*) que possibilita aos usuários operações de inclusão, consulta, alteração e deleção de dados. O SQL é a linguagem utilizada em diversos produtos do mercado tais como banco de dados Oracle, SQL Server, e o PostgreSQL(FOSTER; GODBOLE, 2016). Os autores ainda afirmam que uma característica dos bancos de dados que utilizam o SQL como linguagem é a forma de organizar os dados, todos são bancos de dados relacionais, e isso quer dizer que os dados estão organizados em linhas e colunas de tabelas relacionadas entre si.

Os Sistemas Gerenciadores de Bancos de Dados Relacionais (SGBD-R) apoiam-se em uma fundamentação teórica sólida introduzida por (BOSCARIOLI et al., 2006) denominada de Modelo Relacional de Dados. Um dos principais objetivos deste modelo é prover independência física dos dados para as aplicações, de forma que estas não tenham que obrigatoriamente conhecer detalhes de como os dados encontram-se organizados no meio de armazenamento. Esta independência é obtida através da manutenção do esquema do banco de dados, isto é, de metadados contendo definições de tabelas, visões e índices, junto com o mapeamento destes elementos no meio de armazenamento. A existência de esquemas para as tabelas do banco de dados é uma forte característica dos SGBD-R. Para cada tabela, é mantido um registro do seu nome e informações de suas colunas. Cada coluna possui um nome e o tipo de dados dos valores que podem estar presentes nas linhas destas colunas. Conseqüentemente, as linhas de uma mesma tabela compartilham a mesma definição, ou seja, os mesmos tipos de atributos. Além do modelo relacional, várias funcionalidades ajudaram a popularizar este tipo de sistema, Queiroz (2013) destaca:

a) métodos de indexação baseados em árvores-b (COMER, 1979) ou tabelas *hash*⁷ (CORMEN et al., 1990), capazes de acelerar a execução de consultas e operações de atualização;

⁷ Estrutura de dados especial, que associa chaves de pesquisa a valores. Seu objetivo é, a partir de uma chave simples, fazer uma busca rápida e obter o valor desejado.

- b) Existência de uma linguagem declarativa de alto nível para consulta e manipulação de dados, como a SQL (MELTON, 2003);
- c) um modelo de programação baseado no conceito de transações que correspondem a unidades de trabalho, as quais podem agrupar uma ou mais operações e cuja execução deve ser realizada de forma atômica, satisfazendo algumas propriedades conhecidas por ACID: Atomicidade, Consistência, Isolamento e Durabilidade (FOSTER; GODBOLE, 2016). Reforçado por Elmasri;Navathe (2005), que explicam que a atomicidade garante que todas as operações encapsuladas pela transação sejam realizadas, ou, no caso de falhas, o efeito de nenhuma delas seja aplicado sobre o banco. A consistência garante que, ao final da execução de uma transação, as regras de integridade do banco sejam respeitadas. O isolamento garante que o resultado da execução de transações concorrentes seja o mesmo de quando executadas de forma isolada (seriais). A durabilidade garante que, após a confirmação de uma transação, os resultados sejam refletidos no banco de dados, mesmo se ocorrerem falhas após a confirmação de sua execução.
- d) Mecanismos de replicação de dados geralmente baseados em um modelo mestre/ escravo, no qual consultas de leitura podem ser submetidas a qualquer nó, e de escrita, apenas no nó mestre.

2.4.1.2. Sistemas NoSQL ou Não Relacionais

Tecnologias recentes de armazenamento, gerenciamento e processamento de grandes volumes de dados têm sido criadas por dois grupos distintos: a dos gigantes da internet, como Google, Amazon e Facebook, e a dos grupos de pesquisa em bancos de dados associados a comunidades científicas, principalmente, de Astronomia e Sensoriamento Remoto (QUEIROZ, et al., 2013). O primeiro grupo, segundo o autor, tem se preocupado com a criação de sistemas para atenderem aplicações web que lidam com grande número de usuários, como máquinas de busca, redes sociais e demais aplicações da Web 2.0. Nestas aplicações, existe a necessidade de realizar análises sobre petabytes⁸ de dados provenientes de fontes não estruturadas, como documentos HTML, PDF, arquivos

⁸O petabyte é um múltiplo da unidade de informação byte. O prefixo peta indica a décima quinta potência de 1000 e significa 10^{15} .

texto, mensagens eletrônicas, entre outras. O segundo grupo, segundo o autor, tem buscado criar sistemas mais adequados à comunidade científica, introduzindo novos modelos de dados e linguagens de consultas baseados em matrizes (*arrays*).

Trata-se de uma tentativa de resposta à crescente demanda deste meio, que, com os avanços nos equipamentos de coletas de dados, como telescópios, sensores a bordo de satélites, Geo-Sensores e GPS, tem enfrentado dificuldades para o armazenamento, processamento e análise de quantidades gigantescas de dados. Por isso alguns autores já falam em tsunami de dados (BERRIMAN; GROOM, 2011). Esses novos sistemas estão sendo denominados de NoSQL, cuja tradução mais aceita no momento é *Not only* SQL ou sistemas pós relacionais (POKORNY, 2013). Grande parte deles não utiliza a SQL como linguagem de consulta e, ao contrário dos sistemas relacionais, não há a necessidade de criação de esquemas rígidos para armazenamento dos dados. Os sistemas que surgiram motivados pelos trabalhos publicados pela Google funcionam de forma distribuída e são altamente tolerantes à falhas (CHANG et al., 2008). Estes sistemas encontram-se em operação em grandes centros de dados onde é comum a ocorrência de falhas em alguns nós devido a problemas de disco, CPU ou até mesmo de comunicação de rede. Em consequência, eles possuem recursos para replicação de dados em várias máquinas, aumentando a disponibilidade do sistema: quando um nó cai, outro passa a responder.

2.4.2. NoSQL e SQL

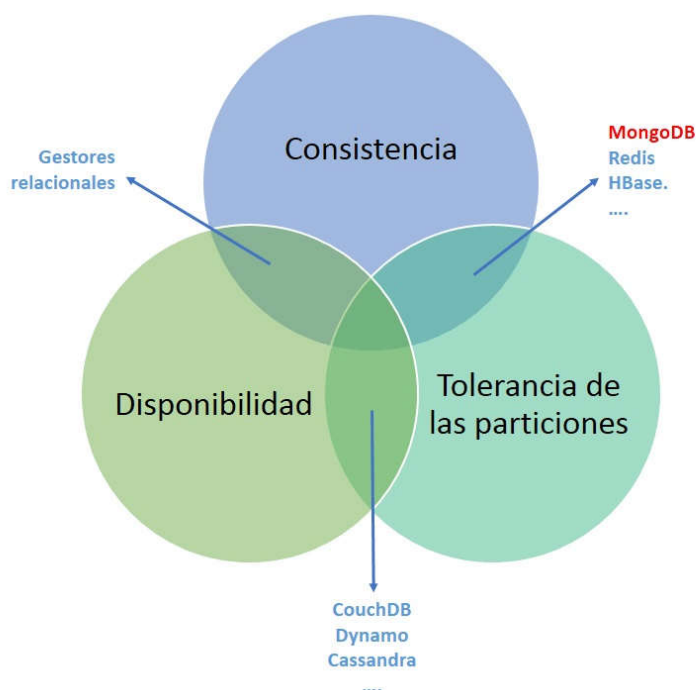
De modo geral, NoSQL não é um termo depreciativo contra SQL, de modo que as duas tecnologias podem coexistir dentro de um aplicativo específico. NoSQL possui tecnologia de banco de dados escalável horizontalmente, que reduz os problemas encontrados em sistemas de saúde atuais que utilizam RDBMS (*Relational Database Management System*) (STRAUCH et al., 2011). Os bancos de dados relacionais adotam uma estrutura de armazenamento estático, onde são apresentados conceitos básicos de relações, atributos e tuplas⁹, os quais são relativamente de fácil compreensão, enquanto NoSQL utiliza estruturas dinâmicas

⁹Cada linha formada por uma lista ordenada de colunas representa um registro, ou tupla.

que indicam semânticas do conteúdo do documento, como por exemplo, arquivos multimídia, que exigem uma categoria de banco de dados que suporte essa especificação. Assim, NoSQL veio como “uma solução para a questão da escalabilidade no armazenamento e processamento de grandes volumes de dados na Web 2.0 (DIANA; GEROSA, 2010). Para visualizar melhor algumas das diferenças entre um modelo NoSQL e o Relacional, é sabido que o modelo clássico é baseado em transações, o que fornece garantia na integridade e consistência. Conhecemos, na literatura, que essas características transacionais são chamadas de ACID: Atomicidade; Consistência; Isolamento e Durabilidade. Porém, em se tratando de um ambiente distribuído, os modelos representados pela ACID sofrem conflitos quanto a alta disponibilidade, pois a informação tem de estar sempre disponível, independentemente do tipo de tecnologia utilizada.

NoSQL possui características que vão ao encontro com o teorema de *Brewer* ou Teorema CAP (BROWNE, 2017). O autor apresenta três funcionalidades (Consistência, Alta Disponibilidade e Tolerância ao particionamento) para bancos de dados distribuídos e que somente duas dessas funcionalidades, um sistema distribuído, pode comportar simultaneamente, ou seja, de acordo com o Teorema, um banco de dados distribuído, pode ter:

Figura 4 - Teorema CAP



- **Disponibilidade e Tolerância a partição (AP)** – Garante que o sistema permaneça disponível, se houver falha, o sistema continua funcionando com alguma redundância de recursos; A tolerância a partições mantém o sistema operando mesmo se falhar a rede, pois o processamento é dividido entre os *hosts*¹⁰ de forma independente. Exemplo: Cassandra, CouchDB (BROWNE, 2017);
- **Consistência e Disponibilidade (CA)** - Parte da base de dados não está preocupada com a tolerância e, principalmente, com o uso de replicação, afim de que se garanta a consistência dos dados e disponibilidade (BROWNE, 2017);
- **Consistência e Tolerância a partição (CP)** – Garante o armazenamento nos hosts de um sistema de banco de dados distribuído e a consistência dos dados, porém não é bom o suficiente para a disponibilidade. Exemplos são: *BigTable, HBase e MongoDB* (BROWNE, 2017).

Contudo, o Teorema contribui para especificar a implementação mais adequada a ser adotada com banco de dados NoSQL.

2.4.3. ACID VS BASE

Os SGBD de bancos de dados relacionais implementam as propriedades ACID (Atomicidade, Consistência, Isolamento e Durabilidade), enquanto os SGBD de modelos de dados NoSQL não garantem as propriedades ACID, mas garantem propriedades BASE (Basicamente Disponível, Estado Leve, Eventualmente Consistente) (NAYAK et al., 2013). Enquanto o ACID fornece um conjunto de propriedades que garantem que as transações das bases de dados são processadas com confiança, o oposto modelo BASE, que é derivado diretamente do teorema CAP, aponta para fornecer um diferente conjunto de propriedades das que o ACID fornece (SILVA, 2010). O Quadro 3 a seguir mostra algumas características desses dois conceitos.

¹⁰Host ou hospedeiro é qualquer máquina ou computador conectado a uma rede.

Quadro 3- Relação entre ACID x BASE

ACID	BASE
Consistência forte	Fraca consistência
Isolamento	Foco em disponibilidade
Conservador (pessimista)	Agressivo (otimista)
Evolução difícil (esquema, por exemplo)	Evolução mais fácil

Fonte: Adaptado de (BREWER, 2000)

2.4.4. BIG DATA

O volume exponencial de dados nos últimos anos tornou-se extremamente difícil de ser mensurado e analisado, mas também passou a ser visualizado por muitos especialistas como novas oportunidades de negócios, sejam eles econômicos ou políticos. Através dessa situação, exigiu-se o surgimento de novas soluções tecnológicas, que pudessem mapear o grande volume de dados em tempo real e em alta velocidade (ANTONIUTTI; ALBAGLI, 2014a). Essa gama de informações diz respeito a um tipo de dados não estruturados (imagens, vídeos, sons e textos). Segundo Gartner (2018), dessa realidade surgiu o conceito de big data, que pode ser caracterizado basicamente por meio dos “3 V’s”: (volume, velocidade e variedade respectivamente). No entanto, outras duas características foram adicionadas atualmente ao conceito que se trata do *valor* e *veracidade* o que na literatura encontra-se o termo “5 V’s” (DEMCHENKO, 2013).

De acordo com Furlan [s.d.], O *Volume* trata da gigantesca quantidade de dados que são criados; A *Velocidade* é indispensável para atuar quase em tempo real sobre o enorme volume de dados; A *Variedade* diz respeito ao formato em que os dados se apresentam (forma estruturada e não estruturada); A *Veracidade* vai da certeza que os dados tem significado e autenticidade; Ainda para o autor, O *Valor* trata da necessidade da instituição provedora dos dados, de que os projetos de Big Data implementados (ou que venha a implementar) obtenham retornos e investimentos.

Outros autores também conceituam os “5 V’s” como:

- **Volume:** Se refere à grande quantidade de dados a serem analisados. Não há uma definição da medida, em bytes, da quantidade necessária de dados para afirmar que alguma organização está lidando com big data (ANTONIUTTI; ALBAGLI, 2014b). Outros garantem que a questão do volume depende do caso e da natureza do dado, pois algumas centenas de gigabytes podem caracterizar como big data devido à velocidade ou tempo necessário para processamento”

(VILLARS et al., 2011). Portanto, a grande massa de dados é um aspecto relativo; depende de outros fatores;

- **Velocidade:** Os dados fluem em uma velocidade sem precedentes e devem ser tratados em tempo hábil. Sensores, celulares e contadores inteligentes estão impulsionados a necessidade de lidar com imensas quantidades de dados em tempo real, ou quase real(SAS, 2018);
- **Variedade:** De acordo com Furlan [s.d.], os dados podem provir de fontes diferentes, e que possuem características que fogem dos sistemas de armazenamentos convencionais. Neste aspecto, a grande variedade de dados, é uma nova promessa de integração de várias tecnologias e soluções de comunicação. A variedade se refere à forma em que os dados aparecem de forma estruturada, que são minoria, e não estruturados, que são a maioria;
- **Veracidade:** Trata da qualidade dos dados e informações, que é característica essencial para que os usuários interessados (executivos, gestores públicos e a sociedade em geral) usem e (re)usem os dados de maneira apropriada e real, gerando informações críveis para eles mesmos(RIBEIRO, 2014);
- **Valor:** Alecrim (2015) afirma que a informação não tão somente poder, mas também é patrimônio. Combinando "*volume + velocidade + variedade + veracidade*", que caracterize uma solução de Big Data, se mostrará inviável se o resultado não trazer benefícios significativos e que compensem o investimento.

Como solução para o gerenciamento do crescente número de dados, surge o termo *Big Data* trata de um conceito de solução tecnológica, no qual o foco é a capacidade de lidar com o grande armazenamento de dados digitais em volume,variedade, maior velocidade e agregando valor. Na prática, seria uma tecnologia que permite analisar qualquer tipo de informação digital em tempo real, sendo fundamental para a tomada de decisões, fornecendo a capacidade de compreender e obter valor, o que ajuda as organizações a operar de forma mais eficiente e rentável(TAURION, 2013).

2.4.5. Gerenciamento de Dados nas Nuvens

As tecnologias oferecidas pela “Nuvem” são utilizadas em grande escala associadas ao conceito de *Big Data*. Atualmente, a utilização de SGBD's vem

ganhando grande espaço em uma plataforma já bastante utilizada por diversos aplicativos ou serviços: a “Computação em Nuvem”. Para o termo “nuvem” existem algumas definições na área da computação. Esse termo foi inspirado em ilustrações de livros-texto que mostravam ambientes remotos, como por exemplo, a Internet, como imagens de nuvens, com a finalidade de esconder a complexidade que está por trás deles (LAPSIA et al., 2012). Uma tecnologia de provedor de “nuvem” que fornece infraestrutura de hardware, software, ou um aplicativo (como um serviço aos seus clientes), em um cenário mais simples, que permite a seus usuários a capacidade de gerir virtualmente e dinamicamente o processamento, recursos de rede e armazenamento (ROSENTHAL et al., 2010). Embora seja um termo recente, há anos, vários serviços de computação em nuvem são oferecidos gratuitamente aos usuários como contas de *e-mails* (*Yahoo.com, Gmail.com, etc.*) ou serviços de armazenamento (*OneDrive, Dropbox, Google Drive, etc.*), porém alguns serviços possuem planos pagos (VILLARS et al., 2011).

2.5. Bancos de dados nas Organizações

Matsumoto (2008), diz que o conjunto de informações que são utilizadas pelos administradores de uma organização diminui as incertezas e tem a denominação de “informação para os negócios”. Essas informações podem se subdividir em: mercadológicas, financeiras, estatísticas, sobre empresas e produtos, jurídicas e outras. A metodologia utilizada para manter e acessar os dados armazenados é realizada através de um Sistema de Banco de Dados que define os tipos de dados que estão armazenados e qual a sua classificação quanto ao seu formato.

De acordo com Laurindo (2002), os dados e os produtos originados dos sistemas de informação não são de propriedade do analista de sistemas ou do centro de processamento de dados, pertence aos usuários. E para Machado (1996), os dados são baseados no princípio de serem estáveis no decorrer da vida de uma determinada empresa ou organização, possuindo características voláteis por refletir uma realidade a ser automatizada e não apenas um momento a ser modificado:

“A modelagem de dados (ou de informações) está baseada no princípio de que, comprovadamente, os dados são estáveis no decorrer da vida de uma empresa ou organização, não tendo

volatilidade dependente de fatores pessoais, governamentais e temporais. Já os procedimentos possuem esta característica de volatilidade, pois sofrem constantes alterações, seja por fatores pessoais, quando existe troca de pessoas e métodos, por decisões governamentais e legislativas, por fatores de calamidade, e outros, externos às atividades normais da empresa.”

Informação que é reforçada por Elmasri; Navathe (2005):

“É correto afirmar que os bancos de dados desempenham um papel crítico em quase todas as áreas em que os computadores são usados, incluindo negócios, comércio eletrônico, engenharia, medicina, genética, direito, educação e biblioteconomia”

Cougo (2013) afirma que o trabalho do banco de dados de forma geral se concentra em armazenamento e consulta de dados. Com a evolução das aplicações e o grau de importância que a informação ganhou dentro das organizações, o banco de dados passou a ter não só a necessidade de armazenar e consultar, mas também de ser capaz de transformar a informação de modo que permita o uso de dados históricos para ser utilizado como base na tomada de decisões estratégicas (NASCIMENTO, 2017)

3. METODOLOGIA

3.1. Procedimentos Metodológicos

Apresenta-se a seguir as descrições metodológicas referentes ao tipo de pesquisa, instrumentos e coletas de dados e o universo onde as informações foram coletadas.

A pesquisa científica é “um procedimento formal, com método de pensamento reflexivo, que requer um tratamento científico e se constitui no caminho para conhecer a realidade ou para descobrir verdades parciais” (MARCONI; LAKATOS, 2003).

3.2. Tipos de Pesquisa

Quanto à natureza, trata-se de uma pesquisa básica, com abordagem quali-quantitativa,). “A pesquisa qualitativa não se preocupa com representatividade numérica, mas, sim, com o aprofundamento da compreensão de um grupo social, de uma organização, etc. sendo assim tem como foco os aspectos da realidade, e não quantitativo” (GERHARDT; SILVEIRA, 2009).

Quanto aos objetivos classifica-se como exploratória, “Este tipo de pesquisa tem como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses” (GERHARDT; SILVEIRA, 2009).

Quanto aos procedimentos técnicos enquadra-se como bibliográfica. A pesquisa bibliográfica constituiu a primeira etapa do trabalho, onde a fundamentação teórica teve os estudos baseados na leitura e análise de livros, revistas, sites acadêmicos e artigos científicos. As informações tiradas foram embasadas de acordo com a temática abordada, buscando uma base dos principais autores que abordam os temas dos tipos de Bancos de Dados como também o seu armazenamento, para assim estabelecer uma compreensão dos objetivos propostos. Todos os estudos selecionados foram lidos integralmente. Um mecanismo utilizado para auxílio no processo de leitura dos estudos foi uma planilha eletrônica com 10 campos delimitados que faz referência a determinado assunto ou informação contida no estudo. A Tabela 1 apresenta os campos com sua respectiva descrição.

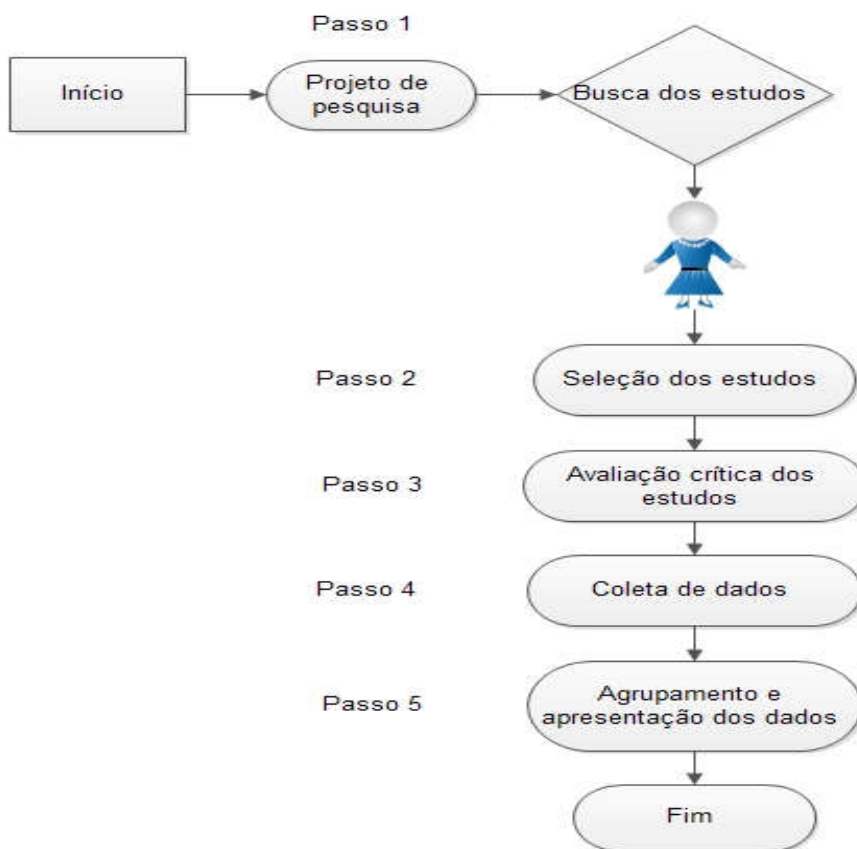
Tabela 1- Campos da planilha para extração de informações dos artigos

Variável	Descrição
Ferramentas e Linguagens	Ferramentas tecnológicas utilizadas, como Linguagens de programação, servidores web, frameworks, etc.
Abordagem Utilizada	(Distribuída, centralizada ou híbrida)
Ambiente em Nuvem	Caso utilizem <i>Cloud Computer</i> para armazenamento e/ou processamento de informações
Projeto/Produto	Nome do produto ou projeto utilizado no Estudo
Licença	Qual tipo de licença do produto (Open Source/Proprietária)
Banco de Dados	Mecanismo (s) de armazenamento utilizado
Tamanho do Banco	Tamanho em GB.
Modelo de Dados	Estrutura de armazenamento dos dados.
Objeto de Consulta	Linguagem de Consulta utilizada ((SQL, AQL, XQuery, etc.)
Plataforma Mobile?	Caso a proposta/sistema/produto utiliza alguma funcionalidade em dispositivos móveis

Fonte: SANTANA (2018).

As informações foram organizadas criteriosamente e posteriormente avaliadas, a Figura 5 apresenta a descrição das etapas que constituíram o processo de elaboração do trabalho

Figura 5- Etapas de elaboração do trabalho.



Fonte: SANTANA (2018).

A pesquisa bibliográfica foi complementada com os seguintes procedimentos:

1. Verificação da lista de “Referências Bibliográficas”: a cada finalização da leitura dos estudos e artigos que possuíam afinidade com o tema foram visualizados e baixados.
2. Contato por e-mail com o desenvolvedor de softwares empresariais, Sr. Paulo Coutinho, proprietário da empresa PR Soluções inteligentes, pesquisador, que prontamente enviou referências sobre uma pesquisa similar que desenvolveu e publicou em seu blog¹¹. O material recebido foi devidamente analisado e contribuiu positivamente para esse trabalho.

Como complemento da pesquisa bibliográfica utilizou-se também a pesquisa de Campo que se caracteriza pelas investigações em que, além da pesquisa

¹¹pcoutinho.com

bibliográfica e/ou documental, se realiza coleta de dados junto a pessoas, com o recurso de diferentes tipos de pesquisa (GERHARDT; SILVEIRA, 2009).

Na pesquisa de Campo utilizou-se como recurso para coleta de dados, a aplicação de questionário (Apêndice A), (MARCONI; LAKATOS, 2003) afirmam que o questionário é um instrumento de coleta de dados. Foi aplicado a instituições que manipulam e são responsáveis pelo armazenamento de grande volume de dados, através do envio de e-mails. O questionário apresenta campos para preenchimento de acordo com a tabela 1, já apresentada. A relação das instituições foi aleatória, feita após um estudo que resultou em uma lista de empresas/instituições que manipulam grande volume de dados.

4. RESULTADOS

Os resultados obtidos foram baseados no enfoque do estudo que abordou as seguintes questões:

Q1: Quais os modelos de SGBD's implementados em empresas que trabalham com grande volume de dados?

Os resultados obtidos estão exibidos na **tabela 2**:

Tabela 2- Base de Dados e Linguagens Utilizadas

		Facebook	Google	TripAdvisor	Foursquare	Twitter	Quora	OpenStreetMa ps	Dropbox	Wikipédia	Σ
Banco de Dados	MySQL	•	•	•		•	•	•	•	•	8
	PostgreSQL							•			1
	Oracle										0
	SQL Server			•							1
	MegaStore		•								1
	MongoDB				•						1
	Cassandra	•				•					2
	Lucene					•					1
	HBase					•					1
	MariaDB								•		1
	BaseX										0
	Mencached	•				•			•		3
	BigTable		•								1
	FlockDB						•				1
	Gizzard						•				1
	Não definido										-
Linguagens de programação	PHP	•								•	2
	Scala				•	•					2
	Dart		•								1
	Go		•								1
	Python						•		•		2
	Java							•			1
	Não definido			•							1

Fonte: SANTANA (2018).

Dentre os modelos de Bancos de Dados citados o *MySQL*¹² é o mais utilizado (n=8) pelas empresas pesquisadas. Tendo em vista que é um servidor e gerenciador de banco de dados do tipo relacional, e que atualmente atende a aplicações de todos os portes.

¹²O MySQL é um sistema de gerenciamento de banco de dados (SGBD), que utiliza a linguagem SQL (Linguagem de Consulta Estruturada, do inglês Structured Query Language) como interface.

Memcached (n=3) funciona como intermediário na frente da camada de acesso do cliente MySQL e gerencia informações apenas na memória sem opção de armazená-lo persistentemente.

O Cassandra (n=2) é um tipo de banco NoSQL que originalmente foi criado pelo Facebook e que atualmente é mantido pela Apache¹³ e outras empresas. Ele é um sistema de banco de dados distribuído baseado no modelo BigTable do Google e no sistema de armazenamento Dynamo da Amazon.com.

PostgreSQL (n=1) é um SGBDR (Sistema de Gestão de Banco de Dados Relacional) funcionando em sistemas do tipo UNIX (exemplo: Linux, FreeBSD, AIX, HP-UX, IRIX, Solaris, etc.). Uma das principais qualidades de PostgreSQL é de ser um software livre, ou seja, gratuito e cujas fontes estão disponíveis.

O Microsoft SQL Server (n=1) é um sistema SGBDR, produto de software cuja principal função é a de armazenar e recuperar dados solicitados por outras aplicações de software, seja aqueles no mesmo computador ou aqueles em execução em outro computador através de uma rede (incluindo a Internet).

O MongoDB (n=1) é um banco de dados que armazena documentos no estilo JSON (fala-se “Jeison”). Isso quer dizer que o desenvolvedor pode carregar para o banco de dados qualquer documento que tenha a sintaxe exigida pelo JSON, sem se preocupar com a estrutura de dados em si.

O Apache Lucene (n=1) é uma API de busca e indexação de documentos, escrito em Java. Ele é composto por basicamente duas etapas: indexação e pesquisa. Dado o texto primeiro passo é a indexação que processa os dados originais e gera uma estrutura que facilita a busca e gera palavras-chaves, em seguida vem à busca que visa estar buscando a partir das palavras-chaves indexadas e retorna pela semelhança do texto com a consulta.

HBASE (n=1) É o banco de dados que roda no sistema HDFS (*Hadoop Distributed File System* – sistemas de arquivos distribuído) por isso dá aos usuários a capacidade única de trabalhar diretamente com dados armazenados no *Hadoop*¹⁴. As características incluem grande escalabilidade.

¹³[Apache Software Foundation](#), responsável por mais de uma dezena de projetos envolvendo tecnologias de transmissão via web, processamento de dados e execução de aplicativos distribuídos.

¹⁴*Hadoop* é uma plataforma de software em Java de computação distribuída.

O MariaDB (n=1) surgiu como uma continuação do projeto MySQL, algum tempo após este ter sido comprado pela Oracle¹⁵.

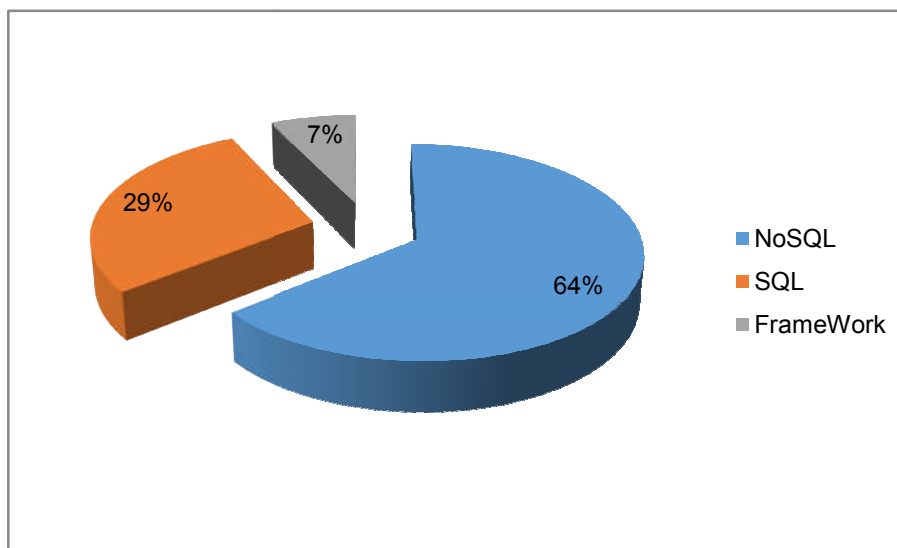
O BigTable (n=1) é um SGBD orientado a colunas criado pelo Google e usado pelo GFS (Google File System)¹⁶ para gerenciar Petabytes¹⁷ de informações.

O FlockDB (n=1) tem uma estrutura otimizada para armazenar e operar sobre grafos, que são estruturas compostas por vértices ligados por arestas com aplicação diversa na medicina, genética, economia e matemática.

Embora o SGBD Oracle seja caracterizado por ser um banco relacional robusto, seguro e com capacidade de gerenciar enormes quantidades de dados, não foi mencionado em nenhuma das implementações pesquisadas.

O **Gráfico 1** exibe o resultado obtido no que diz respeito a classificação dos tipos de Banco de Dados em relação ao modelo relacional e não relacional.

Gráfico 1- SGBD's SQL e NoSQL



Fonte: SANTANA (2018).

Analisando os dados obtidos, observou-se que as instituições que manipulam grande volume de dados, principalmente as redes sociais que trabalham com dados do tipo não estruturados, optam por utilizar os Sistemas NoSQL como solução de armazenamento.

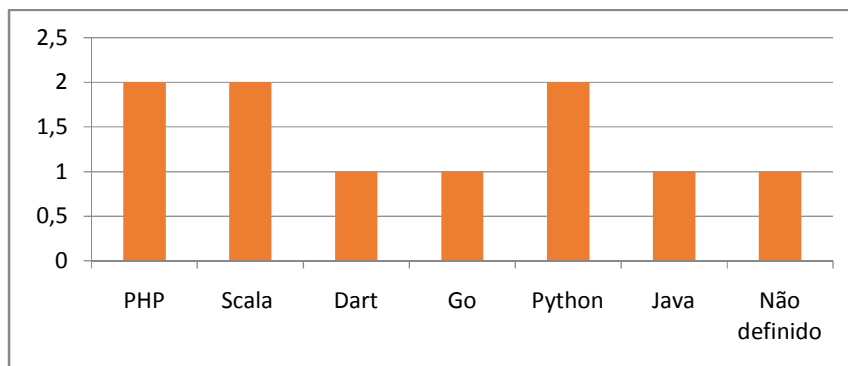
¹⁵ Oracle Corporation, líder mundial na área de banco de dados.

¹⁶ Sistema de arquivos escalável para aplicações de distribuição intensiva de dados.

¹⁷ Múltiplo da unidade de informação byte. O prefixo peta indica a décima quinta potência de 1000 e significa 10^{15} .

No que diz respeito às linguagens de programação utilizadas (citadas) pelas instituições pesquisadas o resultado está exibido no **Gráfico 2**:

Gráfico 2- Linguagens de programação utilizadas



Fonte: SANTANA (2018).

Nota-se que as linguagens PHP, Scala e *Python* tiveram maior expressividade (n=2), seguidas de Dart, Go e Java (n=1).

Q2: Onde estão localizadas/armazenadas, no cenário mundial, as informações desses bancos de dados?

Um Data Center é um ambiente projetado para concentrar servidores, equipamentos de processamento e armazenamento de dados. Com a finalidade de abrigar milhares de servidores e bancos de dados e processar grandes quantidades de informação, esses locais são estrategicamente escolhidos, inclusive com acesso restrito por questões de segurança, tendo em vista que abrigam dados e informações de milhares de usuários.

Os Datacenters das Redes Sociais pesquisadas estão mapeados quanto a sua localização na **Figura 6**:

Figura 6 - Localização geográfica dos Datacenters



Fonte: Adaptado(WIKIPÉDIA, 2017)

Nota-se que a maioria está localizada nos Estados Unidos. Em seguida a Europa em países como: Alemanha, Holanda, Londres e Suécia. Destaca-se na Ásia o polo em Singapura, e registra-se a presença na China, Japão, Tailândia e Taiwan. Na América do Sul encontra-se um data Center do Google, localizado no Chile. Não foram encontrados registros nos demais continentes.

Q3: Quais as peculiaridades das implementações encontradas?

Algumas particularidades foram observadas durante o estudo, dentre elas destacam-se:

O *Google* investe fortemente em energia renovável para seus centros de dados em todo o mundo e onde pode inovar para tornar seus centros de dados tão eficientes quanto possível. Mantendo a sua promessa de emissões de carbono zero, também compensa as emissões investindo em empresas que também podem se beneficiar de novas medidas de eficiência.

Singapura também possui uma economia de internet vibrante e está localizada no centro de um dos mercados de Internet de mais rápido crescimento

no mundo, no Sudeste da Ásia, onde milhões de usuários estão online todos os dias para informações e entretenimento, novas oportunidades de negócios. Fatores que contribuem para que o país seja classificado como um pólo de localização de data centers no mundo, tendo em vista que já abriga estruturas de centro de dados de outras grandes companhias (DropBox e Facebook).

O Dropobox possui uma grande quantidade de datacenters, devido à enorme quantidade de dados que manipula e por possuir usuários espalhados por todo o globo.

O Facebook possui um centro de dados localizado em Luleå, na Suécia, A cidade possui uma população de 46 mil pessoas e está perto da Finlândia. As temperaturas são predominantemente baixas, o mês mais quente, julho, tem uma temperatura média de 16 graus Celsius. Em fevereiro, esse número cai para 6 graus. No entanto, esse clima é um grande aliado para as empresas de data center, que gastam milhões de dólares para esfriar os seus servidores.

Na América do Sul, encontra-se um data Center da empresa Google localizado no Chile, e no Brasil, em São Paulo, já existem escritórios da Google e Facebook.

Grande parte das implementações encontradas, utilizam-se não somente de um SGBD, mas um grupo de mecanismos de armazenamento que são utilizados de acordo com sua característica essencial, como por exemplo, consultas simples para um retorno simples de informações ou uma resposta a uma tomada de decisão, que utiliza um poder de processamento em uma grande massa de dados. Ou seja, SGBD's NoSQL e SQL podem trabalhar juntos.

O Cassandra inicialmente foi criado pelo Facebook, que abriu seu código-fonte para a comunidade em 2008. Agora é mantido por desenvolvedores da fundação Apache e colaboradores de muitas empresas.

Na maioria dos aspectos o MariaDB vai funcionar exatamente como o MySQL: todos os comandos, interfaces, bibliotecas que existem no MySQL também existem no MariaDB. Não há nenhuma necessidade de converter um banco de dados para migrar para o MariaDB.

MongoDB mostra-se como um novo paradigma para quem quer fugir do modelo relacional. É altamente flexível e escalável, ou seja, altamente expansível de acordo com o crescimento de dados, de acordo com alguns autores, sua

simplicidade o torna um SGBD excelente para quem pretende iniciar um novo projeto.

Embora encontrado nos resultados obtidos, apenas um (n=1) nas implementações, no processo de pesquisa, viu-se que o HBase, é largamente utilizado para gerenciar grandes porções de dados em um *cluster*, porém está diretamente ligado a um projeto chamado *Hadoop*, um framework para processamento distribuído. É uma ferramenta que trabalha com grandes volumes de dados, principalmente nas nuvens, esse framework promove o desenvolvimento de software de código aberto e fornece uma estrutura para o desenvolvimento de aplicações altamente escaláveis em computação distribuída, alcançando confiabilidade através da replicação dos dados nos vários nós da rede. Possui sistema de arquivos próprio, o HDFS (*Hadoop Distributed File System*) que armazena grandes volumes de dados em vários *hosts*.

O SGBD MySQL recebe destaque no estudo, pois possui versões que atendem aos princípios do Teorema CAP. Todos os bancos de dados distribuídos conseguem armazenar os dados com técnicas que visam atender a esses requisitos, além de executar consultas com baixa latência, ou seja, o tempo de resposta de um evento até chegar a um destino e retornar, e serem flexíveis a documentos XML ou JSON.

Os resultados mostraram a linguagem de programação JAVA, que é disponibilizada gratuitamente, a mais utilizada. JAVA é uma linguagem robusta, com suporte à programação distribuída, possui qualidades como o poder de processamento por suportar grande número de usuários.

Percebemos que foram utilizadas várias fontes de dados para armazenamento de informações e que na implementação de um ambiente distribuído de informações é necessária uma integração desses dados. A integração de dados é objeto fundamental de um sistema distribuído, através dela os especialistas podem obter uma gama de informações, antes armazenados em diferentes fontes de dados, tais como servidores MySQL, servidores JDBC, Oracle, etc. Isso favorece o diagnóstico preciso. A integração possui um objetivo natural, na presença de várias fontes de dados sobre um determinado tema é fornecer uma interface com única consulta que recupera dados de todas estas fontes.

Importante ressaltar que, em algumas das implementações, utilizam-se a nuvem para “hospedar” seus serviços de hardware e software, ou seja, o uso do armazenamento, memória e servidores conectados, via internet. Isso é consequência do novo paradigma de serviço de computação em nuvem. Esse serviço vem mudando a forma de infraestrutura de computação utilizada atualmente visto a proliferação do número de aplicações que utilizam a nuvem como plataforma.

Por final, dar-se ênfase quanto a tendência entre os autores da utilização de tecnologias livres e sem custos para a implantação de seus eficientes sistemas.

5. CONCLUSÃO

As instituições que já atuam ou pretendem desenvolver com SGBD's em seus sistemas de informações, lidam com uma variedade de dados de variadas fontes, e é necessária uma velocidade na transmissão dos dados produzidos para atender as demandas. No entanto, há um desafio em lidar em manipular os dados de um sistema, quando se trata de Big Data. Os SGBD's NoSQL surgem como solução de modelagem que provê de forma mais simplificada, ganhos de desempenho (processamento), um grau elevado de escalabilidade, flexibilidade e disponibilidade.

Para isso, a escolha entre SGBD's relacionais *opensource*, tais como MySQL ou MariaDB ou soluções não-relacionais como MongoDB, Hbase ou Cassandra, depende não somente do conhecimento da ferramenta por parte do desenvolvedor ou do custo, pois muitos não são pagos, mas sim de necessidades intrínsecas de cada projeto a ser desenvolvido. A opção de uma solução recai sobre vários fatores, como a compatibilidade com outras plataformas, o manejo da equipe envolvida no projeto, possuir ferramentas de administração com interface gráfica amigáveis, possibilidade de gerenciamento mais fácil e produtivo, ser escalável, a disponibilidade de recursos para investir na ferramenta, etc.

Pelo exposto, compreende-se que passamos por uma grande transformação tecnológica e social, marcada pela enorme quantidade de dados que surgem de forma exponencial (Big Data). Cabeas organizações adotarem as soluções para seus projetos que atenderem a grande demanda de usuários.

5.1. Limitações

O estudo realizado apresentou algumas dificuldades encontradas que devem ser levadas em consideração.

No que diz respeito à aplicação do questionário enviado por e-mail, além da demora para obter resposta, a grande maioria das instituições pesquisadas, não deu retorno, e as que o fizeram deixaram de preencher muitos campos, tornando a tabulação dos dados muito difícil. A definição da lista das empresas também pode ser considerada um fator limitante, tendo em vista o fato de que a escolha foi aleatória.

5.2. Trabalhos Futuros

Contudo, a pesquisa levantou alguns questionamentos e um maior interesse para outros estudos acerca do assunto. O conhecimento adquirido é importante para que se continue buscando novos modelos de gerenciamento de banco de dados, apontando as riquezas de soluções existentes. Novos frutos desse trabalho serão desenvolvidos por parte da pesquisadora, alunos, professores e interessados em utilizar material para disseminar informações sobre o universo dos Bancos de Dados.

Pretende-se trabalhar na construção de um Banco de Dados, para ser utilizado como repositório das informações obtidas por meio dessa pesquisa, bem como o armazenamento de outros trabalhos da academia, proporcionando a interação entre os alunos (tendo em vista, que os alunos de outras instituições poderão contribuir), e principalmente mantendo uma fonte de pesquisa dentro do *Campus* Porto Nacional.

REFERÊNCIAS

ALARCÓN, José Manuel. **Fundamentos de bases de dados NoSQL: MongoDB**. 2014. Disponível em: <<https://www.campusmvp.es/recursos/post/fundamentos-de-bases-de-datos-nosql-mongodb.aspx>>. Acesso em: 13 jan. 2018.

ALECRIM, Emerson. **O que é Big Data?** 2015. Disponível em: <<https://www.infowester.com/big-data.php>>. Acesso em: 9 jan. 2018.

ANTONIUTTI, Cleide; ALBAGLI, Sarita. **Uso do big data em campanhas políticas eleitorais**. [s. l.], 2014. a.

BARBEDO, Francisco et al. RODA: **Repositório de Objectos Digitais Autênticos**. Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas, [s. l.], v. 0, n. 9, 2007. Disponível em: <<https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/535>>. Acesso em: 22 jan. 2018.

BERRIMAN, G. Bruce; GROOM, Steven L. **How will astronomy archives survive the data tsunami?** Communications of the ACM, [s. l.], v. 54, n. 12, p. 52–56, 2011.

BOSCARIOLI, Clodis et al. **Uma reflexão sobre banco de dados orientados a objetos**. In: CONGRESSO DE TECNOLOGIAS PARA GESTÃO DE DADOS E METADADOS DO CONE SUL, PARANÁ, BRASIL 2006, Anais... [s.l.: s.n.] Disponível em: <<http://conged.deinfo.uepg.br/artigo4.pdf>>. Acesso em: 5 out. 2017.

BREWER, Eric A. **Towards robust distributed systems**. In: PODC 2000, Anais... [s.l.: s.n.]

BRITO, Ricardo W. **Bancos de dados NoSQL x SGBDs relacionais: análise comparativa**. Faculdade Farias Brito e Universidade de Fortaleza, [s. l.], 2010.

BROWNE, Julian. **Brewer's CAP Theorem** <=: julianbrowne. 2017. Disponível em: <<http://www.julianbrowne.com/article/brewers-cap-theorem>>. Acesso em: 10 jan. 2018.

CAMARGO, A. M...; BELLOTTO, H. **Dicionário brasileiro de terminologia arquivística**. [s. l.], 2005.

CARLOS, Vialfa. **Benchmark - Definição**. 2017. Disponível em: <<http://br.ccm.net/faq/13986-benchmark-definicao>>. Acesso em: 24 jan. 2018.

CHANG, Fay et al. Bigtable: **A distributed storage system for structured data**. ACM Transactions on Computer Systems (TOCS), [s. l.], v. 26, n. 2, p. 4, 2008.

COUGO, Paulo. **Modelagem conceitual e projeto de banco de dados**. [s.l.] : Elsevier Brasil, 2013.

CROWTHER, P.; LAKE, P. **Concise Guide To Databases - A Practical Introduction (Undergraduate Topics in Computer Science)**. [s. l.], 2013.

DE QUEIROZ, Gilberto Ribeiro et al. **Bancos de Dados Geográficos e Sistemas NoSQL: onde estamos e para onde vamos**. Revista Brasileira de Cartografia, [s. l.], n. 65/3, 2013. Disponível em: <<http://www.rbc.lsie.unb.br/index.php/rbc/article/viewFile/594/577>>. Acesso em: 5 out. 2017.

DIANA, Mauricio De; GEROSA, Marco Aurélio. **Nosql na web 2.0**: Um estudo comparativo de bancos não-relacionais para armazenamento de dados na web 2.0. In: IX WORKSHOP DE TESES E DISSERTAÇÕES EM BANCO DE DADOS 2010, Anais... [s.l.: s.n.] Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/wtdbd/2010/sbbd_wtd_12.pdf>. Acesso em: 5 out. 2017.

EARLY OFFICE MUSEUM. **Tabulating Machines**. 2000. Disponível em: <http://www.officemuseum.com/data_processing_machines.htm>. Acesso em: 27 out. 2016.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. [s. l.], 2005.

FERREIRA, Miguel. **Introdução à preservação digital conceitos, estratégias e actuais consensos**. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>. Acesso em: 18 dez. 2017.

FERREIRA, Miguel. **Preservação de longa duração de informação digital no contexto de um arquivo histórico**. [s. l.], 2009. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/9563>>. Acesso em: 22 jan. 2018.

FOSTER, Elvis C.; GODBOLE, Shripad. **Database systems: a pragmatic approach**. [s.l.] : Apress, 2016.

FURLAN, Matheus Batista. **DESMISTIFICANDO O MUNDO DO BIG DATA**. [s. l.], [s.d.].

GARTNER, 2018. **What Is Big Data?** - Gartner IT Glossary - Big Data. 2018. Disponível em: <<https://www.gartner.com/it-glossary/big-data/>>. Acesso em: 6 jan. 2018.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo. **Métodos de pesquisa**. [s.l.] : Plageder, 2009.

INNARELLI, Humberto Celeste. **Preservação de documentos digitais: confiabilidade de mídias de CD-ROM e CD-R**. [s. l.], 2006.

LAPSIA, Vijay et al. **Where should electronic records for patients be stored?** International Journal of Medical Informatics, [s. l.], v. 81, n. 12, p. 821–827, 2012.

LAUDON, KC. KC, LAUDON, JP **Sistemas de Informação**. Rio de Janeiro. 4ª Edição. Editora LTC, [s. l.], 1999.

LAURINDO, Fernando José Barbin. **Tecnologia da informação: eficácia nas organizações.** [s.l.] : Futura, 2002.

MACHADO, Felipe Nery Rodrigues. **Projeto de banco de dados: uma visão prática.** [s.l.] : Ed. Érica, 1996.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos da Metodologia Científica.** [s.l.] : Atlas, 2003.

MATSUMOTO, Cristina Yoshie. **A Importância do Banco de Dados em uma Organização.** Maringá Management, [s. l.], v. 3, n. 1, 2008.

NASCIMENTO, Jean Rodrigues. **Comparação entre Servidores de Banco de Dados Tradicionais e Sistemas Dedicado: um estudo de caso sobre a aplicação do Oracle Exatada Software.** Revista de Engenharia e Pesquisa Aplicada, [s. l.], v. 2, n. 2, 2017. Disponível em: <<http://revistas.poli.br/index.php/repa/article/view/552>>. Acesso em: 4 jan. 2018.

NAYAK, Ameya et al. **Type of NOSQL databases and its comparison with relational databases.** International Journal of Applied Information Systems, [s. l.], v. 5, n. 4, p. 16–19, 2013.

O'BRIEN, James A. **Sistemas de informação e as decisões gerenciais na era da internet.** [s.l.] : Saraiva, 2007.

POKORNY, Jaroslav. **NoSQL databases: a step to database scalability in web environment.** International Journal of Web Information Systems, [s. l.], v. 9, n. 1, p. 69–82, 2013.

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

ROSENTHAL, Arnon et al. **Cloud computing: A new business paradigm for biomedical information sharing.** Journal of Biomedical Informatics, [s. l.], v. 43, n. 2, p. 342–353, 2010.

SARKIS, Laura Costa. **Data warehouse: o processo de migração de dados.** [s. l.], 2001.

SAS. **O que é Big Data?** | SAS. 2018. Disponível em: <https://www.sas.com/pt_br/insights/big-data/what-is-big-data.html>. Acesso em: 6 jan. 2018.

SIEBRA, Sandra de Albuquerque. **Banco de Dados.** [s. l.], 2010.

SILVA, Carlos André Reis Fernandes Oliveira. **Data modeling with NoSQL: How, when and why.** [s. l.], 2010.

SOUZA, Alexandre Morais De. **Critérios de seleção de sistemas de gerenciamento de banco de dados não relacionais em organizações privadas.** 2013. Universidade de São Paulo, [s. l.], 2013.

STAIR, Ralph M. **Princípios de Sistemas de Informação**. [s. l.], 2011.

STRAUCH, Christof et al. **NoSQL databases**. Lecture Notes, Stuttgart Media University, [s. l.], v. 20, 2011.

TAURION, Cezar. **Big data**. [s.l.] : Brasport, 2013.

TOTH, Renato Molina. **Abordagem NoSQL-Uma real Alternativa**. Sorocaba, São Paulo, Brasil: Abril, [s. l.], v. 13, 2011. Disponível em: <http://www.dcomp.sor.ufscar.br/verdi/topicosCloud/nosql_artigo.pdf>. Acesso em: 5 out. 2017.

VILLARS, Richard L. et al. **Big data: What it is and why you should care**. White Paper, IDC, [s. l.], v. 14, 2011.

WIKIPÉDIA. **Fronteiras dos continentes** – Wikipédia, a enciclopédia livre. 2017. Disponível em: <https://pt.wikipedia.org/wiki/Fronteiras_dos_continentes>. Acesso em: 13 jan. 2018.

APÊNDICES

Apêndice A – Modelo do e-mail enviado para as empresas

Caro Desenvolvedor de Software,

Meu nome é Danielle Nascimento de Santana, sou estudante de Graduação do Curso de Licenciatura em Computação do Instituto Federal de Educação, Ciência e Tecnologia do Tocantins. Estou conduzindo uma pesquisa para o meu trabalho de conclusão de curso, que visa identificar as arquiteturas, modelos de bancos de dados, tipos de linguagens utilizadas, etc.

Agradeço se você puder me fornecer mais informações sobre seu produto, como as seguintes perguntas / informações:

1. Qual é o nome do projeto / produto?
2. Qual o Ano do Desenvolvimento?
3. Que linguagens de programação são usadas?
4. O armazenamento / processamento de dados é executado de forma centralizada ou distribuída? Ou ambos?
5. Existe um servidor Web? Qual?
6. Que banco de dados é usado?
7. Você usa uma estrutura para desenvolvimento de software? Se sim, qual?
8. Ele usa um ambiente de nuvem?
9. Qual é a Tecnologia de Serviços da Web (REST, SOAP?) Usada? Ou outro (RMI, CORBA, etc.)?
10. Quais são os mecanismos de segurança utilizados?;
11. O que é o Modelo de Referência utilizado?
12. Que padrões são usados?
13. Existe uma licença para usar? Qual?
14. Qual é o tamanho do banco de dados (GB)?
15. Que idioma de consulta (SQL, SQL, Query, SQL + XPath) é usado?
18. O que é a Solução de Persistência?
19. Existe uma avaliação de desempenho? Qual?;
20. Existe uma plataforma móvel?
21. Quaisquer outras informações importantes?
22. Há publicações acadêmicas?

Responsável por esta informação: