

Saberes e práticas contemporâneas em gestão e inovação na Educação Profissional e em Sistemas Produtivos

Arquitetura de Bancos de Dados da área da Saúde – Uma Revisão Sistemática

Giovana Fadini¹, Marcia Ito ²; Marcelo Okano³

Resumo

Com o avanço da tecnologia é possível armazenar um volume cada vez maior de dados, principalmente na saúde, na qual é preciso além de armazenar fazer um bom gerenciamento deles. Projetar a arquitetura destes bancos de dados é primordial para o adequado gerenciamento dos dados. Esse artigo tem como objetivo identificar e avaliar pesquisas relevantes no contexto da arquitetura dos bancos de dados na área de saúde. O método utilizado foi a revisão sistemática da literatura, na qual foram selecionados 60 artigos para uma pré-leitura e posteriormente 5 para leitura completa. Conclui-se que não há um crescente interesse de pesquisas acadêmicas nesse contexto pelo fato de que foram encontrados apenas métodos para otimização de arquiteturas existentes.

Palavras-chave: Banco de dados, Saúde, Revisão Sistemática

Abstract

With the advancement of technology it is possible to store an increasing volume of data, especially in health, in which it is necessary besides storing to make a good management of them. Designing the architecture of these databases is paramount for proper data management. This article aims to identify and evaluate relevant research in the context of health database architecture. The method used was the systematic review of the literature, in which 60 articles were selected for a pre-reading and 5 later for complete reading. It is concluded that there is not a growing interest in academic research in this context due to the fact that only methods for optimizing existing architectures were found.

Keywords: Database, Health, Systematic Review

¹ Mestrando do Programa de Mestrado Profissional em Gestão e Tecnologia em sistemas produtivos do Centro Paula Souza – gifadini@yahoo.com.br

² Research Staff Member, IBM Research Brasil – IBM-BRL, São Paulo, SP, Brasil – marciaito@br.ibm.com

³ Professor do Programa de Mestrado Profissional em Gestão e Tecnologia em sistemas produtivos - marcelo.okano@fatec.sp.gov.br

1. Introdução

Zhang(2015) explica que o constante desenvolvimento da área de tecnologia da informação (TI), e os avanços em pesquisas para solucionar problemas em dispositivos de armazenamento de dados com grandes capacidades, permitem que cada vez mais grandes volumes de dados sejam armazenados, resultando assim em volumosos bancos de dados.

De acordo com Raghupathi (2014), o setor de saúde tem gerado uma grande quantidade de dados, devido a manutenção de registros dos pacientes (EHR – *Eletronic Health Records* ou EMR – *Eletronic Medical Records* ou PEPs – Prontuários Eletrônicos dos Pacientes), imagens de exames de raios X, tomografias, ressonância, exames laboratoriais, tais como sangue, urina, que geralmente apresentam-se nos mais diversos formatos, pdf, doc, txt, além dos documentos que são manuscritos, tais como relatórios, receitas, atestados, dentre outros.

Toda essa informação é armazenado em bancos de dados que de acordo com Date (2000), por si só pode ser considerado como um repositório ou recipiente para uma coleção de arquivos de dados computadorizados, que possuem uma arquitetura adequada aos dados que será manipulado.

De acordo com Elmasri (1989), um sistema de banco de dados é apenas um sistema computadorizado de manutenção de registros (conjunto de dados) que por sua vez podem ser inseridos, selecionados, excluídos ou até mesmo alterados pro meio de seus usuários. Esses dados são gerenciados através de um Sistema Gerenciador de Banco de Dados (SGBD).

Segundo Feldman (2012), os avanços no gerenciamento de dados, particularmente na virtualização e na computação em nuvem, permitem o desenvolvimento de outros sistemas de gerenciamento de dados, que por sua vez possuem uma arquitetura e modelagem específicas, a fim de facilitar o armazenamento e manipulação mais eficientes de grandes volumes de dados.

Lohr (2012) explica que os dados são acumulados em tempo real com rapidez. O fluxo constante de novos dados que se acumulam a taxas sem precedentes apresenta novos desafios. Assim como o volume e a variedade de dados que são coletados e armazenados mudou, também tem-se a velocidade em que é gerado e que é necessário para recuperar, analisar, comparar e tomar decisões.

Meeker (2016) afirma que se faz necessário a utilização de um sistema de banco de dados com uma arquitetura, que seja capaz de trabalhar com todos esses dados, facilitando a sua integração e promovendo uma melhor organização devido as diversas fontes as quais se originam.

Visto que vários dados são gerados na área da saúde e que se faz necessário que haja um bom gerenciamento dos mesmos e que para isso é primordial que o banco de dados sejam projetados com uma arquitetura que

permita além de receber e armazenar, também gerenciar de forma efetiva e otimizada, é o que justifica esse trabalho de pesquisa e que tem como objetivo identificar e avaliar pesquisas relevantes no contexto da arquitetura dos bancos de dados na área de saúde.

2. Materiais e Métodos

A fim de orientar o mapeamento da revisão sistemática da literatura desse artigo, levantou-se as questões de pesquisa que atuam como fator de inclusão e exclusão dos estudos encontrados. As questões de pesquisa que guiam esse trabalho são:

Quais são os bancos de dados utilizados na área da saúde?

Como é a organização e arquitetura dos bancos de dados dos sistemas de informação em saúde?

Posteriormente, definiu-se as bases de busca utilizados na pesquisa. Buscou-se incluir as principais fontes de trabalhos nas especialidades de ciência da computação e de saúde. As bases selecionadas são listadas na Tabela 1.

Tabela 1. Tabela das bases de busca disponíveis para revisões sistemáticas em Computação e Medicina

Base	Especialidade	Endereço
ACM Digital Library	Ciência da Computação	http://dl.acm.org
IEEE Xplore	Ciência da Computação	http://ieeexplore.ieee.org
PubMed	Saúde	www.ncbi.nlm.nih.gov/pubmed
LILACS	Saúde	http://lilacs.bvsalud.org/

Após a definição das bases onde são efetuadas as buscas, definiu-se os termos sinônimos para cada item das questões de pesquisa, além de agrupar os termos conforme a Tabela 2. Posteriormente, foram definidos os termos de busca (termos concatenados com operadores booleanos). É preciso representar os conjuntos de termos encontrados nas questões de pesquisa a fim de recuperar os trabalhos relevantes nas bases pesquisadas. O Grupo 1 descreve sinônimos relacionados a área da saúde. O Grupo 2 trata de bancos de dados, por fim o Grupo 3 ilustra sinônimos de Sistemas de Informação

Tabela 2. Tabela com os termos sinônimos e grupos de pesquisa

	Grupo1	Grupo2	Grupo3

Termo 1	<i>Eletronic Medical Record</i>	<i>Database</i>	<i>Information Systems</i>
Termo 2	<i>Health</i>	<i>Database Design</i>	<i>Information Technology</i>
Termo 3	<i>Healthcare</i>	<i>Database architecture</i>	<i>Systems</i>
Termo 4	<i>Eletronic Health Record</i>		<i>Information Management Systems</i>
Termo 5	<i>Hospital</i>		<i>Hospital information Systems</i>

A fim de responder as questões de pesquisa levantadas, combinou-se os grupos de sinônimos e definiu-se as Strings de Busca conforme Tabela 3. Os Grupos 1 e 2 buscam identificar trabalhos que tratam de banco de dados e sua arquitetura em sistemas de saúde, bem como registros médicos e registros de saúde. Por outro lado, os Grupos 2 e 3 buscam identificar trabalhos sobre arquitetura de bancos de dados de Sistemas de Informação em saúde.

Tabela 3. Tabela com a combinação de Grupos e as Strings de Busca

Grupos Combinados	String de Busca
Grupo1 e Grupo2	<i>("Eletronic Medical Record") AND ("database design" OR "database architecture")</i>
Grupo2 e Grupo3	<i>("Hospital Information Systems") AND ("database design" OR "database architecture")</i>
Grupo1 e Grupo3	<i>("healthcare") AND ("information Management systems")</i>

Após a utilização das strings de busca nas bases de publicações já descritas, alguns critérios foram utilizados para selecionar os artigos para essa revisão sistemática da literatura, tais critérios adotados foram:

- Artigos publicados em revistas ou conferências revisadas que respondam uma ou mais questões de pesquisa;
- Artigos publicados nos últimos 10 anos, de 2006 até 2016;
- Trabalhos escritos em língua inglesa, por esse motivo foram descartados os artigos selecionados nas bases LILACS;
- Estudos disponíveis para *download* nas bases de buscas definidas.
- Estudos duplicados

- Artigos que não atendiam o escopo desse trabalho, ou seja trabalhos que não respondiam a todas as perguntas que orientam essa pesquisa.
- Após a seleção dos artigos com os critérios já mencionados, todos os resumos foram enviados aos 3 autores desta revisão sistemática da literatura, a fim de que cada um deles desse seu parecer final a respeito dos artigos, sendo esta a ultima fase de seleção.

Por fim, estudos que não respondem às questões de pesquisa e aquelas que não se encontravam disponíveis para *download* são excluídos da análise.

4. Resultados e Discussão

Com as pesquisas nas bases, foi percebido que combinando os grupos 2 e 3, com os termos *Hospital Information Systems e database design ou database architecture* os resultados das buscas eram mais satisfatórios, visto que em outras combinações, não obtinha-se um retorno de artigos, como no caso das combinações das Strings *healthcare AND information Management systems*, que em duas bases das três bases pesquisadas não foi obtido nenhum retorno de artigos que atendiam a esse critério de busca. Por esse motivo encontrou-se um total de 189 artigos. Fazendo um análise refinada nesse universo de 189 artigos, foram descartados os que se repetiam, bem como os que não estavam no inglês, restando assim um universo de 60 artigos.

Esses 60 artigos então foram analisados pelos 3 autores desse trabalho, a fim de ler os seus resumos e verificar se atendiam e respondiam diretamente as perguntas de pesquisa desse artigo. Após a análise dos autores, chegou-se a um número de 5 artigos que foram selecionados como escopo desse trabalho.

Vale ressaltar que foram encontrados vários artigos que abordavam a questão tecnológica em saúde, trazendo temas atuais como desenvolvimento mobile, novos frameworks e novas propostas de ferramentas de visualização de dados o que demonstra um grande interesse e preocupação para esses temas tanto na área médica como em tecnologia. Entretanto, não pertencem ao escopo deste artigo. A Tabela 4 apresenta os artigos que foram encontrados após a utilização das Termos de Busca que estão na Tabela 3, sendo que esse número foi a posteriori reduzido para atender as perguntas de pesquisa. Um resumo em termos quantitativos, encontra-se na Tabela 4.

Tabela 4. Tabela com os resultados encontrados nas bases de busca

Grupos	ACM	IEEE	PUBMED
Grupo1 e Grupo2	0	144	0
Grupo2 e Grupo3	2	172	15

Grupo1 e Grupo3	0	144	0
-----------------	---	-----	---

Para a busca dos Grupos 1 e 2 encontraram-se 144 artigos, sendo todos os artigos descartados, visto que no Grupo2 e Grupo3 tem-se a ocorrência e abrangência maior na seleção dos artigos. Do Grupo 1 e 3 encontra-se 144 artigos, repetindo a seleção realizada no Grupo1 e Grupo2, esses também foram descartados. Os artigos considerados para fins de seleção foram 2 da base da ACM, 172 do IEEE e 15 da base da PubMed.

Uma análise temporal da evolução da quantidade de artigos publicados é realizada, na qual verifica-se um aumento expressivo de número de artigos a partir do ano de 2010 e uma linha de tendência crescente nas duas áreas conforme mostra na Figura 1. Fato que reforça a atual relevância dos temas na pesquisa acadêmica.

Figura 1. Gráfico de evolução do número de artigos por ano



Pode-se perceber que com o passar dos anos, a quantidade de artigos publicados com essa temática foi reduzida, mesmo sendo um tema relevante, visto que a área da saúde de acordo com Raghupathi (2014) é o setor que mais gera um grande volume de dados, sendo necessária uma estrutura adequada para suportar e gerenciar esses dados.

Pela leitura dos artigos, pode-se perceber que é nulo o relato em relação a quais bancos de dados são efetivamente utilizados na área da saúde, nenhum dos 5 artigos selecionados responde a esse questionamento, mas apontam outros detalhes sobre arquitetura de banco de dados utilizada, desenvolvimento de algoritmos para bancos de dados distribuídos, a transação de dados em ambiente mobile por meio do SOA, a integração dos dados médicos multimídias em sistemas distribuídos e também a implementação de um novo método para acrescentar dados que estejam faltando quando se fala em registros sobre anestésias. Uma análise inicial, dos 5 artigos selecionados encontram-se na Tabela 5.

Tabela 5. Tabela com a análise inicial dos artigos selecionados

Artigo	Autores - Tema	Base	Ano	Método	Ferramenta
--------	----------------	------	-----	--------	------------

1	<p>Antoine Lamer, Julien De Jonckheere, Romaric Marcilly, Benoît Tavernier, Benoît Vallet, Mathieu Jeanne, Régis Logier</p> <p>A substitution method to improve completeness of events documentation in anesthesia records</p>	PUBMED E LILACS	2015	Estudo de Caso	DIANE
2	<p>Eunjeong Park, Hyo Suk Nam</p> <p>A Service-Oriented Medical Framework for Fast and Adaptive Information Delivery in Mobile Environment</p>	IEEE	2009	qualitativo	Desenvolvimento de framework
3	<p>Jimison Lavindrasana, Adrien Depeursinge, Patrick Ruch, Stéphane Spahni, Antoine Geissbuhler, Henning Müller</p> <p>Design of a Decentralized Reusable Research Database Architecture to Support Data Acquisition in Large Research Projects</p>	LILACS e PUBMED	2007	Estudo de caso com 6 instituições.	Web-services
4	<p>Liviu Constantinescu, Jinman Kim, (David) Dagan Feng</p> <p>SparkMed: A Framework for Dynamic Integration of Multimedia Medical Data Into Distributed m-Health Systems</p>	IEEE	2012	qualitativo	SparkMed
5	<p>Umut Tosun - Distributed Database Design using Evolutionary Algorithms</p>	IEEE	2014	experimental	Proposta de um algoritmo evolutivo

No primeiro artigo selecionado, de acordo com Lamer et al (2015), os sistemas de informação hospitalares (HIS) são sistemas muito importantes nas organizações em saúde, dentre eles destaca-se os sistemas de gerenciamento de informações de anestésias (AIMS), onde são armazenadas informações

clínicas desde a pré-intervenção anestésica, até a alta do paciente para as salas de recuperação.

Ele relata que apesar do alto volume de dados coletados nos sistemas AIMS, o banco de dados desse sistema, não facilita a análise estatística, pois sua arquitetura é organizada em um esquema transacional. Um banco de dados transacional é organizado para ser alimentado e consultado através de um software específico associado. É desenvolvido de forma a fornecer respostas rápidas sobre as operações do dia-a-dia ou exibição de dados e curvas.

A organização normalmente limita as consultas a um registro particular por vez. Além disso, na maioria dos casos, esse tipo de esquema não adere a formatos padronizados e só pode ser lido e não modificado. Como resultado, a capacidade de adicionar funcionalidades em tempo real, bem como extrair e tratar dados para suporte de decisão ou projetos de pesquisa é limitada com a maioria dos AIMS, devido a arquitetura do banco de dados transacional. (LAMER et al, 2015).

Essa limitação no sistema AIMS permite que possa ser visualizado os dados sobre uma intervenção específica, mas não são análises retrospectivas sobre um enorme volume de intervenções.

A pesquisa do Lamer et al (2015), propõe o desenvolvimento de um data warehouse e uma nova arquitetura adequada para análise dos dados, ele descreve um novo método que permite a detecção de eventos ou dados não existentes nos registros de anestésias.

Park e Nam (2009) apresentam no segundo artigo a importância da utilização dos dispositivos móveis em pacientes que necessitem de um tratamento rápido e que estejam em condições críticas, a fim de estabelecer uma comunicação rápida e consistente entre hospitais e médicos.

Eles propõe, uma estrutura que suporta o acesso onipresente a sistemas médicos usando serviços móveis integrados, além de desenvolver um framework orientado a serviços que gerencia dados precisos do paciente por meio da integração de sistemas médicos locais, usando padrões de informação médica como imagens digitais.

Apresentam no artigo um protótipo que permite a visualização em dispositivos móveis de imagens de tomografia computadorizada, recuperação de imagens para tratamento de AVC, por meio de rede sem fio, de acordo com Park e Nam (2009), a estrutura proposta por eles, reduz os departamentos hospitalares e facilita tratamentos nas ausências de médicos especialistas.

Lavindrasana et al (2007) apresentam em seu artigo um projeto de arquitetura de banco de dados de pesquisa reutilizável descentralizada para apoiar a aquisição de dados em grandes projetos de pesquisa. No artigo buscam projetar uma arquitetura de banco de dados descentralizada, escalável e reutilizável com menores custos de manutenção para gerenciar e integrar dados

heterogêneos distribuídos necessários como base para um projeto de pesquisa de grande escala.

O projeto proposto foi implementado dentro de seis centros clínicos que participam do projeto @neurlST como parte de um sistema maior de integração de dados e reutilização para tratamento de aneurisma.

Constantinescu et al. (2012) por sua vez propõe o SparkMed, que é uma estrutura para integração dinâmica de dados médicos multimídia em sistemas distribuídos de saúde. O framework por eles desenvolvido permite que aplicativos de dados médicos compartilhem dados com hosts móveis através de uma rede sem fio (como WiFi e 3G), vinculando-se a sistemas de software existentes e implantando-os como aplicativos de m-Health.

Desenvolveram um protótipo do framework SparkMed para avaliação de uma simulação de fluxo de trabalho radiológico, que usa o SparkMed para implantar um visualizador de imagens radiológicas como um aplicativo m-Health para uso telemédico por radiologistas e partes interessadas.

No último artigo selecionado, Tosun (2014) desenvolve um projeto de banco de dados distribuídos utilizando algoritmos evolutivos.

De acordo com Tosun (2014) o desempenho de um sistema de banco de dados distribuído depende, em particular, da alocação do site dos fragmentos. As consultas acessam diferentes fragmentos entre os sites e existe um site de origem para cada consulta. Um algoritmo de alocação de dados deve distribuir os fragmentos para minimizar os custos de transferência e liquidação da execução dos planos de consulta.

O custo principal para um algoritmo de alocação de dados é o custo da transmissão de dados em toda a rede, e os algoritmos evolutivos escaláveis foram desenvolvidos para minimizar os custos de execução dos planos de consulta. Os algoritmos propostos encontram soluções quase ideais para o problema de alocação de dados.

Destacamos neste item uma descrição dos pontos principais abordados por cada autor em seus respectivos artigos, o propiciou um maior entendimento sobre os artigos selecionados para essa revisão sistemática.

5. Considerações finais

Nesse trabalho, apresenta-se uma revisão sistemática da literatura atual no contexto de bancos de dados na área da saúde, bem como a arquitetura utilizada. É possível com essa revisão sistemática, mostrar que com o passar dos anos houve um descréscimo considerável nas pesquisas acadêmicas nesse contexto.

Há uma grande necessidade que mais pesquisas sejam realizadas na área de banco de dados, bem como de sua arquitetura, visto a importância dos dados armazenados, além da quantidade e variedade.

Diversos trabalhos abordam o uso de banco de dados, mas não apontam estudos sobre novas arquiteturas e também não relatam qual banco de dados estão utilizando para esse armazenamento.

Dos trabalhos selecionados percebe-se que a preocupação em se otimizar processos já conhecidos por meio de novos algoritmos, frameworks e desenvolvimento de aplicações em dispositivos móveis é bem explorado nas pesquisas, deixando uma lacuna quando fala-se em características dos bancos de dados no quesito estrutura, sendo um assunto a ser ainda explorado

Referências

CONSTANTINESCU, J. Kim and D. Feng, **SparkMed: A Framework for Dynamic Integration of Multimedia Medical Data Into Distributed m-Health Systems,**" in IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 1, pp. 40-52, Jan. 2012. doi: 10.1109/TITB.2011.2174064

DATE, C. J. Introdução aos Sistemas de Bancos de Dados. Campus, 2000.

ELMASRI, R.; NAVATHE, S. B. Fundamentals of Databases Systems. Benjamin – Cummins, 1989.

FELDMAN B, MARTIN EM, SKOTNES T: Big Data in Healthcare Hype and Hope. October 2012. Dr. Bonnie 360. 2012. Disponível em: <http://www.west-info.eu/files/big-data-in-healthcare.pdf>. Acesso em: 20/06/2017

IAVINDRASANA J, Depeursinge A, Ruch P, Spahni S, Geissbuhler A, Müller H., Design of a Decentralized Reusable Research Database Architecture to Support Data Acquisition in Large Research Projects. Stud Health Technol Inform 129 (2007), 325-9

LAMER, A., De Jonckheere, J., Marcilly, R. et al. J Clin Monit Comput (2015) 29: 741. <https://doi.org/10.1007/s10877-015-9661-3>

MEEKER, MARY. THE DATA: Next Big Wave - Leveraging this unlimited connectivity & storage to collect / aggregate / correlate / interpret all of this data to improve people's lives & enable enterprises to operate more efficiently, from Internet Trends Report 2016. Disponível em: http://dq756f9pzlyr3.cloudfront.net/file/2016_internet_trends_report_final.pdf. Data de acesso: 15/06/2017

PARK E, Nam HS. A service-oriented medical framework for fast and adaptive information delivery in mobile environment. IEEE Trans Inf Technol Biomed. 2009; 13:1049–1056.

RAGHUPATHI, W.; RAGHUPATHI, V. Big data analytics in healthcare: promise and potential (2014). Disponível em: <http://hisjournal.biomedcentral.com/articles/10.1186/2047-2501-2-3>. Data de acesso: 15/07/2017

TOSUN. U., Distributed database design using evolutionary algorithms, in Journal of Communications and Networks, vol. 16, no. 4, pp. 430-435, Aug. 2014. doi: 10.1109/JCN.2014.000073

ZHANG. J; ML, Huang, and Z, Meng, Visual analytics for BigData variety and its behaviours, Computer Science and Information System, Vol. 12, Issue 4, pp. 1171-1191, Nov 2015