

## O uso do PostgreSQL e MongoDB no gerenciamento de dados de monitoramento de consumo de Cannabis na região da fronteira sul do Brasil

GILBERTO KREISLER<sup>1</sup>; WINE SILVA<sup>2</sup>; ALINE NEUTZLING<sup>3</sup>; BEATRIZ FRANCHINI<sup>4</sup>

<sup>1</sup>Universidade Federal de Pelotas – gkfneto@inf.ufpel.edu.br

<sup>2</sup>Universidade Federal de Pelotas – wdssilva@inf.ufpel.edu.br

<sup>3</sup>Universidade Federal de Pelotas – neutzling@live.de

<sup>4</sup>Universidade Federal de Pelotas – beatrizfranchini@hotmail.com

### 1. INTRODUÇÃO

As áreas de saúde necessitam de recursos computacionais adequados às suas especificidades. Quanto à análises dos dados, necessitam de ferramentas, tais como banco de dados, que possibilitem a extração de informações que direcionem as pesquisas. Um banco de dados é uma coleção de dados relacionados, sendo os dados fatos que podem ser gravados e possuam um significado implícito (ELMASRI, 2003). Informação é uma coleção de dados que, quando apresentada de determinada forma e em determinado momento, melhora o conhecimento do indivíduo que a recebe, de modo a que este indivíduo se torne mais capaz de realizar a ação ou decisão a que se propõe (GALLIERS, 1987).

Segundo Elmasri, (2005, p. 4), uma coleção de dados pode ser considerada um banco de dados. Entretanto o uso do termo *banco de dados* é geralmente mais restrito, sendo necessário cumprir determinadas propriedades implícitas:

- Representa alguns aspectos do mundo real, chamado de **minimundo**. As mudanças no **minimundo** são refletidas no banco de dados;
- Um banco de dados é uma coleção lógica e coerente de dados com algum significado inerente. Uma organização de dados ao acaso não pode ser corretamente interpretada como um banco de dados;
- Um banco de dados é projetado, construído e povoado por dados, atendendo uma proposta específica.

O projeto “Identificação de indicadores para o monitoramento e avaliação dos impactos da nova política uruguaia de regulação do mercado de Cannabis sobre a saúde pública e o consumo de drogas na zona de fronteira entre Brasil e Uruguai” (Projeto Cannabis Fronteira) optou pela utilização de dois Sistemas Gerenciadores de Banco de Dados (SGBD), que, segundo Elmasri, (2003), são uma coleção de programas que permitem criar e manter um banco de dados, facilitando os processos de definição, construção, manipulação e compartilhamento de bancos de dados entre vários usuários.

O Projeto Cannabis Fronteira utiliza dois modelos para o gerenciamento dos dados, um modelo relacional, através do PostgreSQL, e outro não-relacional, através do MongoDB.

PostgreSQL é um sistema de banco de dados relacional de código aberto, onde por relacional pode se entender que a estrutura interna do banco é representada por tabelas; o software possui mais de 15 anos de desenvolvimento ativo, apresentando uma boa reputação quanto a confiabilidade, integridade dos dados e exatidão, também sendo compatível com as propriedades ACID (Atomicidade, Consistência, Isolamento, Durabilidade) (POSTGRESQL, 2016).

MongoDB é um banco de dados não-relacional orientado à documentos, onde a ideia de linhas de uma tabela de um banco relacional é substituída por uma estrutura mais flexível: um documento. Um documento, junto com seus subdocumentos, pode representar diferentes relacionamentos diminuindo o tempo de consulta ao banco (CHODOROW,2010), esta estrutura flexível permite que as informações coletadas tenham uma estrutura mais aberta, onde cada informação pode possuir sua organização e atributos diferentes tendo a mesma natureza.

## 2. METODOLOGIA

Durante os meses de janeiro, fevereiro e março de 2016 foram realizadas saídas à campo pela região da fronteira Sul do Brasil, onde questionários quantitativos acerca de práticas de consumo de *Cannabis* foram preenchidos pela equipe de campo que observou grupos que faziam uso de *Cannabis*. Estes questionários geraram dados para alimentação do banco de dados primário utilizando o PostgreSQL.

Após a obtenção dos dados primários, parte da equipe focou na busca de informações nacionais e internacionais a respeito das práticas de consumo de Cannabis, tendo sido reunidas informações relevantes sobre estas práticas as quais foram armazenadas em um banco não-relacional, sendo feito o uso do MongoDB.

### 2.1 Utilização do PostgreSQL

Inicialmente, os dados coletados a partir dos questionários eram armazenados em uma planilha Excel por uma questão de praticidade devido a sua fácil manipulação e difundido conhecimento, compondo uma tabela com 81 atributos (colunas).

A utilização inicial desta planilha tornou necessário um processo de normalização para a eliminação de redundâncias, inconsistências e para o devido enquadramento no modelo relacional, para então ser feita a inserção no banco. Segundo Elmasri (2003), o processo de normalização pode ser visto como um processo de análise do relacionamento entre os dados a fim de garantir a minimização de redundâncias e minimização das anomalias geradas a partir de inserções, deleções e atualizações de dados no banco.

O processo inicial de normalização consistiu na descrição do significado de cada variável do questionário para então agrupar cada atributo de acordo com seu tema; os atributos referentes ao observador agrupados em uma tabela, os referentes ao local onde foi feita a observação, em outra e assim sucessivamente, dando forma a cada entidade do minimundo; sendo observada as relações entre cada entidade, buscando representar as relações do minimundo.

## 2.2 Utilização do MongoDB

A obtenção dos dados que compõem o banco de dados secundário se deu a partir dos dados disponíveis de saúde no Brasil, como o DATASUS, dados provenientes de pesquisas internacionais feitas por ONG e órgãos governamentais de países como Irlanda.

Uma vez obtidos os dados, foi possível observar a não uniformidade dos aspectos considerados em cada pesquisa: para uma mesma temática, uma pesquisa poderia classificar os participantes por sexo, ou por idade, por exemplo. Essa não uniformidade culminou na escolha de um banco de dados não-relacional para o armazenamento, uma vez que bancos relacionais requerem um esquema rígido para os dados devido ao seu armazenamento em tabelas.

Após a escolha das variáveis pela equipe, iniciou-se o processo de armazenamento na base de dados; devido ao fato dos dados não serem brutos e em um formato não importável toda a inserção precisou ser feita de forma manual através de um software de gerenciamento para o MongoDB, o software *Robomongo*.

Posterior à fase de inserção de dados na base de dados, verificou-se a necessidade da exportação dos mesmos para um modelo relacional para sua disponibilização, pois a utilização de bancos relacionais tem maior difusão. Essa necessidade se mostrou um desafio devido à não existência de ferramentas automatizadas para a migração do modelo orientado à documentos para o modelo relacional, o que precisou ser feito de forma manual através da utilização de scripts, pequenos trechos de códigos que executava diretamente no banco para cada coleção no banco.

## 3. RESULTADOS E DISCUSSÃO

O uso dos dois modelos de banco de dados possibilitou o armazenamento de dados sem perda informações, a unificação dos dados em um só modelo relacional resultaria em muitas tabelas em um banco relacional, que não seria o mais indicado para a representação da informação, uma vez que cada fonte de informação possui uma estrutura diferente, haveriam tabelas para cada tipo de estrutura, ou tabelas gigantescas com vários campos não utilizados; já a unificação dos dados no modelo não-relacional, dificultaria a análise dos dados primários, pois os dados no formato relacional puderam ser facilmente exportado para softwares estatísticos, como o IBM SPSS.

Identificou-se a necessidade de uma posterior utilização do serviço de nuvens para a hospedagem dos bancos devido a sua flexibilidade, pois possui

a característica de pagamento conforme utilização de recursos, também permitindo que os dados sejam acessados e modificados em uma base única por todos os usuários, tendo como um ponto de análise futura mais aprofundada. Com isso, melhorando a qualidade no acesso aos dados, garantindo consistência e confiabilidade (ZANG, 2010).

#### 4. CONCLUSÕES

A modelagem e documentação dos bancos de dados primários se dará após normalização do banco, para assim começar a implementação e posterior incorporação dos dados ao banco. Ainda está sendo analisado qual o grau de normalização será necessário para os dados primários, a fim de se adequar melhor as necessidades do projeto.

Os dados primários, após a análise, também serão importados no banco não-relacional, onde irão compor uma grande base de dados a respeito de práticas de consumo de *cannabis*, esta base de dados será armazenada em nuvem para o acesso remoto aos dados.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

ELMASRI, R.; NAVATHE, S.B.. **Sistemas de Banco de Dados**, São Paulo: Pearson, 2005.

CHODOROW, Kristina. **MongoDB: the definitive guide**. O'Reilly Media, Inc., 2010.

POSTGRESQL. **About**. Acessado em 14 jun. 2016. Online. Disponível em: <https://www.postgresql.org/about/>.

ZHANG, Q.; CHENG, L.; BOUTABA, R. Cloud computing: state-of-the-art and research challenges. **The Brazilian Computer Society**, Published Online, n.1, p.7- 18, 2010.

GALLIERS, R.. **Information Analysis – Selected Readings**. Sidney: Addison-Wesley Publishing Co, 1987.