

Estruturação de um Banco de Dados Espacial como Repositório dos Dados do Projeto SIGAreias - ANE

Raphael Cordeiro Pedrosa de Almeida

rcpa@a.recife.ifpe.edu.br

Aida Ferreira Araújo

aidaferreira@recife.ifpe.edu.br

Ioná Maria Beltrão Rameh

ionarameh@recife.ifpe.edu.br

RESUMO

Este estudo consiste em relatar a estruturação de um banco de dados espacial para o desenvolvimento do Sistema SIGAreias e sua possível evolução para um banco multidimensional. O Sistema de Informações Geográficas denominado SIGAreias foi criado para viabilizar as análises espaciais das informações obtidas em campo pelo projeto “Águas de Areias”. Sua estruturação, assim como a do banco de dados, consiste em facilitar o processamento dos dados e obtenção das informações processadas para a tomada de decisões (melhor gestão dos recursos hídricos e áreas degradadas, entre outras). O processo foi descrito em fases, que vão desde o tratamento dos dados para inserção no repositório até o desenvolvimento do sistema de informação geográfica em ambiente Web (SIG-Web), bem como verificar como este se comporta junto ao referido repositório para a execução das consultas e funcionalidades do sistema SIGAreias

Palavras-chave: Banco de dados, SIG, Processamento dos Dados, Informações Espacializadas, Modelagem Multidimensional.

ABSTRACT

The aim of this paper is report a structuring of a spatial database for the development of the SIGAreias system and its own evolution to a multidimensional database. The Geographic Information System (GIS) called SIGAreias was created to make spatial analysis of the information obtained in the fieldwork by the Águas de Areias project. Its structuring, as well as that of the database, was aimed at facilitating the processing of data and obtaining the information processed for decision making. The process has been described in phases, ranging from the treatment of the data for insertion in the repository to the development of the geographic information system in the Web environment (SIG-Web) a whole and how it behaves next to this repository for the execution of the queries and features of the SIGAreias system

Keywords: Databases, GIS, Data Processing, Spatial Information, Multidimensional Modeling.

1 INTRODUÇÃO

A Associação Águas do Nordeste - (ANE)¹, ONG criada em 13 de dezembro de 2011 sem fins lucrativos tem como missão institucional contribuir para o conhecimento, a conservação e a gestão das águas na Região Nordeste do Brasil, atuando na educação e comunicação para a conservação e gestão das águas, bem como desenvolver iniciativas acadêmicas na pesquisa, ensino e extensão, voltadas para o meio ambiente e em especial. A referida Associação desenvolve trabalhos e promove ações nos municípios de Brejo da Madre de Deus, Santa Cruz do Capibaribe, Poção e Jataúba relacionados a construção de poços/reservatórios para suprir necessidades dos moradores em regiões afetadas pela falta d'água, bem como relatar os riscos que a extração da areia dessas aluviões proporcionam para o desgaste ou até mesmo o desaparecimento desses reservatórios naturais. Uma das ações tomadas consistiu em um cadastramento de residências e do quantitativo de poços existentes ao longo da calha do Rio Capibaribe (trecho que abrange aproximadamente 79 km ao longo do rio, sendo 1km a partir do eixo do rio, 500m na sua margem esquerda e direita) a fim de contabilizar a população que utiliza esse recurso. Esse cadastramento gerou uma série de planilhas contendo todos os elementos coletados (identificação dos moradores da área;

tipo de reservatório e/ou ponto de captação existente na localidade daquele usuário; quantidade de pessoa que se utilizam deste recurso; entre outros). A partir dessa coleção de dados armazenados em papéis, surgiu a necessidade de criar um sistema computacional que pudesse gerenciar de alguma forma esse material, tornando possível a exibição destes dados sobre mapas de forma digital. Assim, a solução encontrada foi o desenvolvimento de um SIG (Sistema de Informação Geográfica) para o monitoramento da área estudada, correlacionando as informações espaciais disponíveis nos bancos de dados de empresas e entidades públicas (mapeamento das mesorregiões do estado; das bacias hidrográficas que abrangem a área do projeto) aos dados coletados em campo pela equipe da ANE, proporcionando a visualização espacial do contexto analisado, bem como a criação de um repositório de dados para consultas e processamento de informações. Os sistemas SIG's são utilizados para este propósito, unificando grandes quantidades de dados acerca de um determinado tema e/ou localidade com o objetivo de tornar mais simples a manipulação, processamento e captura de informações, bem como indicar respostas às várias questões sobre planejamento urbano e regional, meio rural e levantamento de recursos renováveis, “[...] integrando operações convencionais de base de dados, com possibilidades de seleção e busca de informações (Query) e análise estatística, conjuntamente com a possibilidade de visualização oferecida pelos mapas” (PINA E SANTOS, 2000, p.15). A partir dos questionamentos para elaboração deste sistema e do

¹ Associação Águas do Nordeste (ANE). Disponível em: <http://www.aguasdonordeste.org.br/website/>. Acesso em 03 de abril de 2016

processo de desenvolvimento da aplicação “SIGAreias”, foi sugerida uma evolução na estrutura do repositório de dados, possibilitando ao usuário o acesso às novas ferramentas de análise dos dados (a partir da utilização de softwares para manipulação e exibição web de cubos OLAP (Online Analytical Processing), bem como o processamento dessas informações geradas para a tomada de decisões acerca dos problemas trabalhados pela ANE na região estudada (gerenciamento dos recursos hídricos da região afetada ou mesmo gestão/criação de hotspots para mitigação da falta de água em determinadas regiões, tomando como referência a extração ilegal da areia de aluvião dos respectivos pontos de coleta da água, por exemplo).

2 MATERIAIS E MÉTODOS

A modelagem e estruturação do repositório para persistência dos dados fornecidos pela ANE, foi realizada de acordo com a formatação utilizada no formulário de cadastro dos usuários de poços (o qual foi utilizado pelos recenseadores em seu trabalho de campo), a fim de possibilitar a manipulação desses dados pelo sistema. O referido formulário é composto por:

1. Dados gerais do usuário entrevistado (nome, endereço, município, telefone, quantidade de pessoas morando na localidade, se tem algum tipo de criação ou cultivo na localidade), bem como informações acerca do tipo de manancial existente e/ou utilizado na respectiva localidade recenseada (Rio, Riacho, Barragens, Açudes, Tanque de Pedra, Águas Subterrâneas, Olho D'Água);

2. Como ocorre a extração dessa água do reservatório (extração mecanizada ou elétrica) e para qual fim é utilizada essa água;

3. Se o usuário recenseado faz uso de água de chuva ou não, bem como se o mesmo é fornecedor de água para algum outro indivíduo, seja ele cadastrado ou não.

A partir destes aspectos, o modelo Entidade-Relacionamento foi gerado com o objetivo de organizar os dados para o sistema e de facilitar a criação de funcionalidades de busca e manipulação dos dados inseridos no repositório do sistema. Durante esse processo, também foram definidas estratégias de confecção e organização da estruturação do sistema como um todo (padrão MVC de arquitetura de software², cuja estrutura separa a representação da informação da interação com o usuário; metodologias voltadas para o desenvolvimento ágil (Agile Development³), com a finalidade de criar uma estrutura organizacional

² O padrão arquitetural Model-View-Controller (MVC) é uma forma de quebrar uma aplicação, ou até mesmo um pedaço da interface de uma aplicação, em três partes: o modelo, a visão e o controlador, onde o Controlador (Controller) que interpreta as entradas do mouse e teclado (ações do usuário) e envia para o Modelo (Model), que por sua vez gerencia um ou mais elementos de dados, resolve os problemas e responde as questões. A Visão (View) é responsável apenas por exibir os dados para o usuário através da combinação de gráficos e textos. (DEV MEDIA, 2016)

³ A metodologia de Desenvolvimento Ágil (Agile Development) descreve uma série de princípios para o desenvolvimento de software, onde as requisições e soluções propostas envolvem a auto-organização de seus participantes (não há uma delegação de responsabilidades, mas sim cada um ocupa o espaço e a responsabilidade a qual possui mais aptidão ou empatia) e o esforço colaborativo (todos os participantes do time devem corroborar para o bom desenvolvimento do projeto, auxiliando nas dificuldades uns dos outros, bem como conhecendo o processo como um todo para tentar mitigar os possíveis impedimentos existentes). (AGILE SOFTWARE DEVELOPMENT, 2017);

para conduzir o processo do desenvolvimento da aplicação SIGAreias; aplicação de padrões de projetos (design pattern) para resolução de possíveis problemas que possam ocorrer ao longo do desenvolvimento da aplicação). Este processo inicialmente abrangeu as principais necessidades da ANE, ligadas principalmente a persistência dos dados para utilização em trabalhos acadêmicos voltados para às áreas ambientais, de estudos estatísticos, entre outras. Entretanto, foi verificado a necessidade de um melhor planejamento para o sistema, como também para o repositório de dados, tendo em vista o desenvolvimento de uma aplicação SIG, disponível para as diversas plataformas existentes (Mobile, Desktop e Web). A confecção e modelagem do repositório é baseado no modelo entidade-relacionamento de estruturação de banco de dados proposto por Peter P. Chen, em 1976, o qual é tido como um conjunto de ferramentas conceituais que descrevem os dados e como estes são persistidos em um banco de dados, relacionados entre si, bem como a semântica e restrições de consistência existentes (DATE, 2004, p.355). Esses dados, por sua vez, são representados através das tabelas existentes no repositório, como também a partir dos relacionamentos existentes entre eles, visando atender o chamado “Princípio da Informação” onde “[...] todo o conteúdo de informação do banco de dados é representado de um e somente um modo, ou seja, como valores explícitos em posições de colunas em linhas de tabelas.” (DATE, 2004, p.53). Silberschatz também discorre sobre essa característica complementando sobre as inúmeras aplicações comerciais existentes que agregam esse modelo ao processamento de dados e como este

modelo, devido a simplicidade de implementação, proporciona vantagens aos desenvolvedores se comparado a outros modelos existentes (modelo de rede ou o modelo hierárquico) (SILBERSCHATZ, 2006, p.25).

Assim, a escolha do banco de dados para o sistema foi iniciada, tendo em vista características e serviços que possibilitassem a persistência dos dados de forma segura e que o acesso a este ocorresse de forma rápida e confiável. Um banco de dados pode ser definido como uma coleção de dados relacionados (derivados de alguma fonte), registrados (de algum evento ou situação do mundo real), que possuem alguma razão para existir (razão/razões estas pelas quais o indivíduo observador necessita daqueles dados registrados) e um público alvo a ser atingido (que situação ou problema o indivíduo observador quer ou deseja resolver). Das várias formas de se manter ou gerenciar um banco de dados, a forma computadorizada é a mais comum atualmente, visto que este sistema/estrutura “[...] pode ser criado e mantido por um grupo de programas de aplicações escritos especificamente para essa tarefa [...]” (NAVATHE, 2010, p.3).

O PostgreSQL foi a ferramenta escolhida como Sistema Gerenciador de Banco de Dados (SGBD) para manter e gerir as informações que foram disponibilizadas para o sistema em produção. Este sistema possui código aberto e é otimizado para aplicações mais complexas, isto é, que envolvem grandes volumes de dados ou que tratam de informações críticas. Junto ao seu módulo geográfico PostGIS, o qual permite a identificação, manipulação e inserção de dados com coordenadas geográficas, este é considerado a

melhor alternativa de repositório de dados em virtude das bibliotecas e aplicações adjacentes ao SGBD (para exibição espacial, bem como exibição desses dados de forma interativa na aplicação). Navath afirma sobre as vantagens da aplicação dessas inúmeras ferramentas de gerenciamento de dados e como elas podem contribuir para a evolução do chamado banco de dados tradicional (comumente empregado onde a maior parte dos dados existentes é do tipo numérico ou texto) para um repositório mais complexo, de diferentes tipos de dados, sendo denominado banco de dados multimídia, cuja possibilidade abrange o armazenamento de outros tipos existentes (áudio, vídeo, imagens, etc) (NAVATHE, 2010, p.2).

3 METODOLOGIA

A primeira modelagem ER foi estruturada de forma a abranger todos os aspectos analisados no formulário cadastral. Este possui uma característica central: os dados correspondentes ao usuário entrevistado para os respectivos locais de coleta da água de aluvião. Os dados acerca desta entidade foram registrados no modelo proposto de forma a servir de ponto de acesso a outras informações complementares (dados do recenseador, se o usuário é consumidor da água ou fornece de água à outra família, se tem alguma produção ou criação animal/vegetal, se o recenseado possui interesse em obter cisterna), bem como determinar os principais pontos de coleta da água de aluvião registrados para aquela determinada residência.

Figura 1. Modelo Entidade-Relacionamento do Banco de Dados do SIGAreias (modelagem inicial).



Fonte: Projeto SIGAreias - Associação Águas do Nordeste; Raphael Almeida, 2016

A figura 01 representa o modelo proposto em reunião pelo grupo de desenvolvimento da aplicação, cuja estrutura tenta repetir a mesma utilizada no formulário físico de recenseamento. Neste, é a partir da entidade Formulário (cujo os campos abrangem: nome do usuário, latitude e longitude, quantidade de pessoas no domicílio, contato, etc) que outras entidades são ligadas e complementam as informações acerca do usuário que está sendo cadastrado (as entidades: tipo_de_uso que corresponde às finalidades de uso da água dos mananciais e formulários criados; recenseador que corresponde aos recenseadores responsáveis pelo cadastramento naquele determinado local; marcação de ponto com latitude e longitude, que corresponde a localização da casa principal do recenseado).

A entidade Formulário também serve como chave primária para outras características acerca da usabilidade do manancial (entidades relacionadas aos mananciais cadastrados (manancial que corresponde às características do manancial, de que forma ele é mantido, qual o mecanismo utilizado para extração da água; detalhamento_manancial, esta utilizada como complemento para os mananciais advindos de nascentes e olhos d'água; agua_subterranea, cujas informações são concernentes a

utilização de águas advindas de rios subterrâneos; cisterna referente a utilização de cisternas para captação da água das chuvas)), como também para as características e formas de uso dessa água coletada (entidades `criacao_animal`, `outros_usos`, `irrigacao`, `cultura_vegetal`, que foram pensadas para comportar as informações pertinentes ao tempo de cultivo ou criação animal naquela localidade, especificando qual a utilidade do manancial para o desenvolvimento dessa criação/cultura). Essa estrutura foi pensada como forma de evitar o desgaste, extravio do material utilizado para a coleta, bem como preservar todo e qualquer aspecto de informação coletado pelo trabalho de campo desenvolvido.

A estrutura, no início da produção no sistema, foi viável e apresentou poucos impactos na produção das funcionalidades de consultas e localização espacial dos pontos cadastrados pelos recenseadores do projeto. Para cada entidade principal do sistema ou as quais se desejava espacializar no formato de mapa digital (formulário, locais, denúncia, poços) foi criada uma nova coluna denominada “`the_geom`”, onde os campos latitude e longitude são convertidos e o hash4 proveniente desse processamento é armazenado e utilizado pelas aplicações espaciais ligadas ao SIGAreias (GeoServer5 e OpenLayers6), junto ao seu módulo geográfico PostGIS (a coluna “`the_geom`” só é possível por causa deste recurso que, ao ser instalado junto ao SGBD, disponibiliza bibliotecas para identificação e conversão de valores numéricos em coordenadas geográficas no PostgreSQL). Para ANE o processamento dos dados de latitude e longitude coletados proporcionou a utilização destes na confecção de

mapas digitais, possibilitando uma melhor utilização de seus recursos logísticos e financeiros (acesso aos dados cadastrados no repositório e visualização destes no contexto geral da área do projeto), bem como a criação de estratégias de atuação em determinadas localidades próximas aos pontos coletados para aplicação de ações de manejo sustentável e conscientização acerca da utilização dos recursos da região pelos moradores (quantidade de pessoas nas localidades que possuem ou utilizam recursos provenientes das suas residências ou de residências vizinhas; a partir da cultura existe naquela determinada área, quais as formas de mitigar os problemas e aumentar a produtividade sem prejuízos ao ambiente; estratégias de empreendedorismo sustentável voltadas para a construção de reservatórios e poços de aluvião).

Figura 2. Mapa Digital gerado pelo sistema a partir das informações cedidas pela ANE.



Fonte: Projeto SIGAreias - Associação Águas do Nordeste; Raphael Almeida, 2016

Entretanto, com a perspectiva de criação de novas funcionalidades para cadastros e consultas espaciais das informações no sistema, voltadas para a tomada de decisões dos principais órgãos responsáveis, foi observado a necessidade de uma melhor estratégia visual do software (front-end do sistema, cujas modificações podem ser inicialmente observadas na figura 2, onde a ferramenta se utiliza das informações contidas em seu repositório para espacializar os dados no mapa-base), como também de uma

voláteis (termo que implica na característica de não poderem ser alterados, mas ainda sim com a possibilidade de serem excluídos), variáveis no tempo (principal característica de um data warehouse), utilizados para o suporte às decisões gerenciais. Essa característica abrange a nova estrutura, tendo em vista apenas a leitura dos dados existentes para processamento dos dados, bem como a inserção de novos dados, dando um “refresh” nas análises e atualizando os aspectos processados aquelas novas informações (DATE, 2004, p.603).

Para a estruturação do novo repositório, o processo de modelagem compreende desde o desenvolvimento do modelo teórico, implementação das estruturas junto ao SGBD escolhido (instalação de bibliotecas e recursos para o processamento de dados), bem como o armazenamento dos dados para processamento (criação da massa de dados). Este processo deve ser feito de modo a representar de forma satisfatória, os relacionamentos entre entidades (projeto lógico do modelo do banco de dados), “[...] garantindo, assim, que as operações relacionais funcionarão como se deseja e não produzirão resultados surpreendentes” (DATE, 2004, p.593), tornando as informações geradas compreensíveis e navegáveis, sem a ocorrência de erros ou discrepância entre as informações geradas e reais da região. A eficiência e o desempenho do banco atrelado ao sistema também é um fator importante e deve ser observado durante o processo de elaboração do novo modelo, para que o sistema não seja sobrecarregado com requisições desnecessárias, bem como sejam estruturados processos de transformação dos dados que preservem a informação gerada (DATE, 2004, p.593).

Outro aspecto importante é com relação ao modelo multidimensional, caracterizado pelo fato dos dados estarem armazenados em mais de uma dimensão no banco de dados (eixo X e Y para atributos e tuplas + eixo Z para o critério tempo) o qual deve ser observado, principalmente com relação ao SGBD utilizado, já que este demandará uma série de funcionalidades para gerenciamento do repositório de dados (técnicas de armazenamento, gestão e controle das funcionalidades e execuções de processos) e como estes recursos serão consumidos pelo sistema (DATE, 2004, p.612).

Essa estrutura gera possibilidades para criação de novas funcionalidades de inserção de dados; a utilização de ferramentas OLAP; criação da informação a ser utilizada em processos analíticos para tomada de decisão; até mesmo a utilização do chamado NoSQL7, em determinadas partes do sistema, com o objetivo de auxiliar nas requisições que podem vir a ocorrer (transformando parte dos dados do repositório relacional para a estrutura NoSQL, criando assim vantagens para os desenvolvedores no que concerne ao processamento das informações e otimização dos processos ao longo do uso do sistema pelo usuário).

4 CONCLUSÃO

A oportunidade de elaborar o software SIGAreias, bem como a estruturação de seu repositório de dados proporcionou não apenas a ampliação dos conhecimentos que envolvem o desenvolvimento de uma plataforma WEB, mas também o entendimento da organização de um repositório de dados relacional (estruturação das entidades, atributos e tuplas), forma como os dados são utilizados pelo sistema e como esses

são trabalhados para atender as necessidades do usuário (relacionamentos e conexões entre os dados existentes, exibição e utilização destes na estrutura da aplicação em si), proporcionando uma melhor análise e visualização das inferências espaço-temporais para a tomada de decisão a partir dos dados espacializados.

Além desses, entender como o processamento dos dados pode ser estruturado e realizado pelo sistema gerenciador de banco de dados proporciona um leque de possibilidades, não apenas acerca da

evolução do sistema (abrindo vazão para a criação de novas funcionalidades para análise dos dados), como também um maior controle sobre a atualização dos já existentes e inserção de novos, que virão a contribuir para a evolução do repositório no aspecto temporal, trazendo consigo a oportunidade de estudos acerca dos aspectos físicos e sociais da região junto ao critério de tempo (evolução do processo de consumo, conservação de recursos naturais e monitoramento da região em virtude de possíveis processos de degradação).

REFERENCIAS

ANE - Associação Águas do Nordeste. Pernambuco. Disponível em: <<http://www.aguasdonordeste.org.br/website/>>. Acesso em 16 de fevereiro de 2016.

BARREIROS, Daniel Bion; FERREIRA, Aida A; BARBOSA, Ioná Maria Beltrão Rameh. **SIGWeb para Dados de Qualidade de Água em Bacias Hidrográficas**. CIENTEC: Revista de ciência, tecnologia e humanidades do IFPE, v. 7, p. 43-55, 2015.

DATE C. J.; VIEIRA, A. L. S. **Introdução a sistemas de bancos de dados**. 8a edição. ed. Rio de Janeiro: Editora Elsevier, 2004. 896 p.

ELMASRI, R.; NAVATHE, S. B.. **Sistemas de Banco de Dados**. 4a ed., Pearson-Addison-Wesley, 2005.

FUNÇÃO HASH. In: WIKIPEDIA: the free encyclopedia. [San Francisco, CA: Wikimedia Foundation, 2010]. Disponível em: https://pt.wikipedia.org/wiki/Função_hash. Acesso em: 04 de out. de 2016.

GEOSERVER. In: Open Geospatial Consortium. [Wayland, MA]. Página inicial. Disponível em: <http://geoserver.org/>. Acesso em: 11 de fev. de 2016.

HOSTING DATA. **NoSQL DATABASES**. Página inicial. Disponível em: <http://nosql-database.org/>. Acesso em: 28 de out. de 2016.

OLAP. *In:* WIKIPEDIA: the free encyclopedia. [San Francisco, CA: Wikimedia Foundation, 2010]. Disponível em: <https://pt.wikipedia.org/wiki/OLAP>. Acesso em: 23 de out. de 2016.

OPENLAYERS. *In:* Open Geospatial Consortium. [Wayland, MA]. Disponível em: <http://openlayers.org/>. Acesso em: 14 fev. de 2016.

PINA, M. F.; SANTOS, S. M.: **Conceitos básicos de sistemas de informação geográfica e cartografia aplicados à saúde.** Departamento de Informações em Saúde, 2000.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **SISTEMA DE BANCO DE DADOS.** São Paulo: Makron Books, 1999.