



XI-077 – DETECÇÃO DE PERDAS APARENTES EM SISTEMAS DE DISTRIBUIÇÃO DE ÁGUA ATRAVÉS DE TÉCNICAS DE MINERAÇÃO DE DADOS

Roberta Macêdo Marques Gouveia⁽¹⁾

Mestranda concluinte do Programa de Pós-Graduação em Informática da Universidade Federal da Paraíba (PPGI/UFPB); Graduada em Ciência da Computação pela UFPB; Pesquisadora do Laboratório de Eficiência Energética e Hidráulica em Saneamento (LENHS)

Heber Pimentel Gomes⁽²⁾

Doutor em Hidráulica pela Universidade Politécnica de Madrid; Prof. do Programa de Pós-Graduação em Engenharia Urbana e Ambiental da Universidade Federal da Paraíba (PPGEUA/UFPB). Coordenador do Laboratório de Eficiência Energética e Hidráulica em Saneamento (LENHS)

Valéria Gonçalves Soares Elias⁽³⁾

Doutora em Ciências da Computação pela Universidade Federal de Pernambuco (UFPE); Prof^ª. e coordenadora do Programa de Pós-Graduação em Informática da Universidade Federal da Paraíba (PPGI/UFPB)

Moisés Menezes Salvino⁽⁴⁾

Doutorando do Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal da Paraíba (PPGEM/UFPB); Mestre em Engenharia Mecânica pela Universidade Federal da Paraíba (PPGEM/UFPB) e pesquisador do Laboratório de Eficiência Energética e Hidráulica em Saneamento.

Endereço⁽¹⁾: Jardim Universitário, s/n – Castelo Branco - João Pessoa - PB - CEP: 58051-900 - Brasil - Tel: +55 (83) 3216-7037 - e-mail: roberta@lenhs.ct.ufpb.br

RESUMO

O controle e o uso eficaz dos dados armazenados em sistemas gerenciadores de banco de dados é um dos grandes desafios enfrentados atualmente pelas empresas do setor de saneamento. Muitos softwares ainda são produzidos sem nenhuma preocupação com a geração futura de informações integradas e estratégicas. A solução constitui em reorganizar a maneira como os dados são armazenados, disponibilizados e acessados. Este cenário favorece o desenvolvimento do *Data Warehouse* e a aplicação das tecnologias OLAP e de Mineração de Dados. Portanto, diante deste contexto, é proposto neste trabalho um ambiente de suporte à decisão propício a unir dados armazenados em diferentes fontes, organizá-los no *Data Warehouse*, para em seguida aplicar OLAP e Mineração dos Dados, a fim de descobrir padrões e novos conhecimentos em um determinado setor do saneamento atendido pela Companhia de Abastecimento de Água e Esgoto da Paraíba.

PALAVRAS-CHAVE: Perdas Aparentes, Mineração de Dados, Sistema de Apoio à Decisão, Sistema de Distribuição de Água.

INTRODUÇÃO

As ações que buscam o controle e a redução das perdas de água delineiam-se na melhoria da qualidade da operação dos sistemas de abastecimento, conseqüentemente, na melhoria dos serviços prestados e do uso racional da água pela população. De acordo com o Programa de Modernização do Setor de Saneamento, de toda a água captada dos mananciais para abastecer os municípios brasileiros, quase a metade, 44,18%, se perde antes de chegar às casas e atender a população.

Este trabalho objetiva encontrar respostas para as questões relacionadas às irregularidades praticadas pelos consumidores, da qual a empresa de distribuição de água desconhece. Será apresentada e elaborada uma modelagem visando traçar o perfil do consumidor que contribui para os altos índices de perdas de água e de inadimplência, e a partir desta análise serão propostas medidas corretivas e preventivas para minimizar o problema. As tecnologias de *Data Warehouse* serão utilizadas como parte do processo de descoberta de conhecimento na base de dados e para obtenção dos resultados serão utilizadas as técnicas de Mineração de Dados a fim de detectar padrões e descobrir regras significativas na tentativa de reduzir os índices de Perdas Aparentes na distribuição de água do Estado da Paraíba.



As Perdas Aparentes ocorrem com a água que é tratada e fornecida pela companhia de abastecimento de água, e consumida pelos clientes, porém não é corretamente medida e, portanto não é faturada, nem gera arrecadação correspondente. De acordo com (GOMES, et al., 2007), as principais causas das Perdas Aparentes são os erros e desatualizações no cadastro de clientes; fraudes, violação ou danificação de medição nos hidrômetros e ligações clandestinas. Em virtude desta problemática, derivada das perdas acentuadas de água nos sistemas de distribuição de água, este trabalho empenha-se em estudar as Perdas Aparentes, através da análise dos dados obtidos de um setor referente à malha de distribuição da cidade de João Pessoa.

Um dos reflexos negativos das Perdas Aparentes são as perdas de faturamento por parte da companhia de abastecimento. Estudos apresentados no último relatório do (SNIS, 2007) mostraram que o valor médio das perdas de faturamento para todo o conjunto de prestadores de serviços do Brasil foi de 39,8%. O indicador de perdas de faturamento é calculado pela relação entre os volumes faturados e os disponibilizados para distribuição.

De acordo com (SNIS, 2007), o controle das perdas está diretamente relacionado aos índices de micromedição, ou seja, a medida que verifica o volume consumido pelos clientes da companhia, cujo valor será objeto da emissão da conta a ser paga pelo usuário. Em sistemas de abastecimento de água que possuem altos índices de micromedição as perdas existentes passam a ser reflexos das perdas físicas e não mais das perdas aparentes. Portanto, a motivação do presente artigo surge da necessidade de se propor uma abordagem para investigar as Perdas Aparentes de água através da análise dos dados obtidos de um setor referente à malha de distribuição da cidade de João Pessoa, Paraíba. A idéia geral é traçar o perfil do consumidor de acordo com seu nível, avaliando consumos e valores faturados por mês em um período de dois anos.

METODOLOGIA

Os Sistemas de Apoio à Decisão (SAD) visam à avaliação crítica das informações dos negócios, auxiliando a gerência a definir tendências, apontar problemas e absorver decisões inteligentes. Neste estudo são abordadas as tecnologias de *Data Warehouse* (DW), *On-Line Analytical Processing* (OLAP) e *Mineração de Dados* (*Data Mining* - DM).

O Processo de Descoberta de Conhecimento em Banco de Dados, mais conhecido como KDD, *Knowledge Discovery in Databases*, tem o objetivo de extrair da base de dados as informações desconhecidas, válidas e acionáveis, que poderão ser úteis para a tomada de decisão. O processo envolve três etapas iniciais: seleção, pré-processamento e transformação dos dados, as quais compõem a fase de preparação dos dados. Em seguida vem a etapa de Mineração de Dados, considerada essencial ao processo e que constitui o foco principal deste trabalho.

O emprego das técnicas de Mineração de Dados foi realizado com os dados fornecidos pela Companhia de Água e Esgota da Paraíba (CAGEPA/PB). O setor 64 da localidade 001 da Unidade de Negócio Litoral faz parte do estudo de caso de detecção de perdas aparentes proposto por este trabalho.

As Perdas Aparentes refletem diretamente no Indicador de Perdas de Faturamento. A relação entre o índice de micromedição e o índice de perdas de água revela os efeitos positivos que a correta hidrometração proporciona ao controle das perdas, em virtude do índice de perdas de faturamento ser inversamente proporcional ao índice de micromedição. Segundo (SNIS, 2007), a CAGEPA/PB possui índice de 45,3% em micromedição e 39,6% no índice de perdas de faturamento. Tais índices refletem-se de forma negativa, visto que as perdas de faturamento estão diretamente ligadas às perdas reais e aparentes. Desta forma, nos sistemas que possuem altos índices de micromedição, as ligações clandestinas deixam de ser um fator crítico para os problemas das perdas de água e as perdas existentes passam a ser reflexos das perdas físicas e não mais das aparentes.

O controle e o uso eficaz dos dados armazenados é um dos grandes desafios enfrentados atualmente pelas empresas. A solução constitui em reorganizar a maneira como a informação é armazenada, disponibilizada e acessada. Este cenário favorece o desenvolvimento do DW. Para (INMON, 2005) o termo é definido como um "depósito de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar as decisões da gerência".



Segundo (KIMBALL, et al., 2002), o DW deve fornecer os subsídios necessários para a transformação de uma base de dados que utiliza On-Line Transaction Processing (OLTP) para On-Line Analytical Processing (OLAP), onde OLTP significa os processamentos que executam as operações do dia-a-dia da organização e OLAP significa os processamentos que suportam a tomada de decisões, e que configura o objetivo do estudo em questão. O uso do *Data Warehouse* surgiu por duas razões: primeiro, pela necessidade de fornecer uma origem de dados única, limpa e consistente para fins de apoio à decisão; segundo, pela necessidade de fazê-lo sem causar impacto sobre os processos operacionais do dia-a-dia da companhia de abastecimento de água.

Segundo (PONNIAH, 2001), um *Data Warehouse* (DW) não é um simples software ou produto de hardware que se adquire para fornecer informações estratégicas. É, sim, um ambiente computacional onde os usuários podem encontrar informações estratégicas, e onde os usuários são colocados diretamente em contato com os dados que necessitam para tomar as melhores decisões.

De acordo com (INMON, 1992) o DW deve ser Orientado por Temas, Integrado, Variante no Tempo e Não Volátil. A propriedade “Orientado por Tema” refere-se à importância de organizar as informações pelos temas principais, onde cada tema pode envolver várias tabelas, dentre elas, tabelas acumulativas e detalhadas. A propriedade “Integrado” mostra a necessidade de acoplar, através de uma convenção padrão, dados de diferentes formatos. A terceira propriedade “Variante no Tempo” define que os dados não são atualizáveis e podem ser comparados ao longo do tempo. Eles são atribuídos como retratos da base de dados operacional atual, onde cada ocorrência e cada mudança são consideradas como um novo registro, pois a informação histórica não é perdida. A última propriedade “Não Volátil” corresponde ao DW ser disposto fisicamente para otimizações de inclusões e consultas. Ou seja, não deve ser um banco preparado para atualizações ou alteração.

A modelagem adotada foi o Esquema Constelação de Fatos. Tal esquema consiste em dispor as tabelas do modelo relacional para a modelagem multidimensional. Ele é formado por uma tabela central, denominada Tabela-Fato, a qual possui os dados principais da visão da análise. Nela ficam ligadas as tabelas de dimensão, onde cada uma é definida para conter uma única chave primária, baseada na integridade do relacionamento com a tabela de fatos. As tabelas dimensionais são desnormalizadas para aumentar o desempenho e representam as entidades que evoluem ao longo do tempo. Cada dimensão tem uma chave primária que corresponde a um dos campos da chave da tabela de fatos.

Na concepção do *Data Mart* é necessário definir a área de atuação, as tabelas-fato e suas dimensões. A área de atuação deste estudo é o setor comercial da companhia de saneamento. Quanto à definição das Tabelas-Fato e suas Dimensões, identificou-se duas Tabelas-Fato: Consumidor Potencial e Perda Aparente. A Tabela-Fato Consumidor Potencial é a entidade responsável pelo reconhecimento dos consumidores, através dos dados cadastrais referentes ao setor, que poderão gerar perda aparente no sistema. A Tabela-Fato Perda Aparente detém os atributos necessários para a realização das análises visando checar irregularidades na rede de distribuição de água.

RESULTADOS E DISCUSSÕES

A aplicação de algoritmos de mineração de dados para identificar Perdas Aparentes foi motivada por três motivos: a disponibilidade de grandes quantidades de dados, existência de dados históricos armazenados e possibilidade de encontrar um perfil de comportamento típico. Foram construídos modelos de mineração baseados em agrupamento e classificação visando traçar o perfil do consumidor que faz ligações clandestinas, para que medidas corretivas e preventivas sejam tomadas, objetivando a redução dessa não conformidade e, conseqüentemente, a redução das perdas de água.

Para facilitar os resultados, os atributos das tabelas foram classificados em: Cadastrais, Relacionamento e Padrão de Comportamento. As informações cadastrais são específicas do cliente e são praticamente estáticas ou se modificam pouco. As informações sobre o relacionamento do cliente com a empresa são do tipo: idade da ligação, idade do hidrômetro, padrão da ligação, característica do hidrômetro, pressão e vazão do trecho, etc. Já as informações referentes ao comportamento do cliente ao longo do tempo são do tipo: consumo mensal, percentual de variação de consumo (maior, média e a menor), irregularidades e multas aplicadas. Para a modelagem da mineração as variáveis de comportamento são as mais adequadas para achar um padrão de



comportamento que identifique Perdas Aparentes no sistema de abastecimento. Contudo, as outras não foram descartadas, visto que também auxiliam na interpretação dos resultados.

Dos vários modelos definidos na literatura, dois deles se mostraram adequados e serão utilizados neste trabalho. Os modelos são: 1) Análise de Agrupamento e 2) Classificação por Árvore de Decisão.

Uma Árvore de Decisão (AD) pode ser representada como um conjunto de regras, ou galhos, onde cada nó não terminal representa um teste ou decisão sobre o item considerado. O objetivo principal é separar as classes em que os dados estão agrupados. Um exemplo deste modelo seria a classificação dos consumidores em confiáveis ou não confiáveis junto à companhia de abastecimento, baseando-se na quantidade de pontos de utilização e o valor da conta. Com este modelo observa-se que uma regra de classificação terá sempre no seu conseqüente uma resposta ao fato das condições satisfazerem ou não a uma determinada classe previamente definida.

Tal conhecimento extraído da massa de dados de uma empresa permite ao gerente, por exemplo, tomar a decisão de realizar intervenção no abastecimento dos consumidores com baixo grau de confiança. O uso deste modelo para o setor de saneamento objetiva prever, ou seja, gerar dados futuros dos consumidores que possuem risco de infringir a companhia de saneamento através de inadimplências e/ou fraudes. A predição é a habilidade de elaborar cenários diferentes, para antecipar certos resultados.

Já com a aplicação do Modelo de Análise de Agrupamento pretende-se verificar a existência ou não de um grupo definido de consumidores potenciais que geram/geraram perdas aparentes para o setor de saneamento. A atividade de agrupamento, também denominada segmentação, consiste em identificar grupos naturais com base num conjunto de atributos. O Modelo de Agrupamento segmenta um conjunto de dados em grupos diferentes, cujos itens são semelhantes. Usando o agrupamento, a ferramenta de Mineração de Dados descobre grupos diferentes dentro dos dados.

As tecnologias OLAP foram projetadas para apoiar análises e consultas, além de auxiliar seus usuários a sintetizar informações através de comparações, visões personalizadas e análises históricas. Elas têm como característica principal permitir uma visão mais fácil e intuitiva dos dados multidimensionais, por meio de análises em diferentes perspectivas. De acordo com (HAN, et al., 2006), OLAP faz parte do processo que habilita usuários a explorar os dados em um *Data Warehouse*, fornecendo funcionalidades para análise interativa de dados em diferentes dimensões. A visualização multidimensional dos dados através das tecnologias OLAP favorece a análise de várias dimensões em única tela, em virtude da estrutura conceitual conhecida por cubos de dados. A visualização se dá através de configurações tridimensionais de linhas, colunas, funções *slice-and-dice* e gráficos, como mostra a Figura 1.

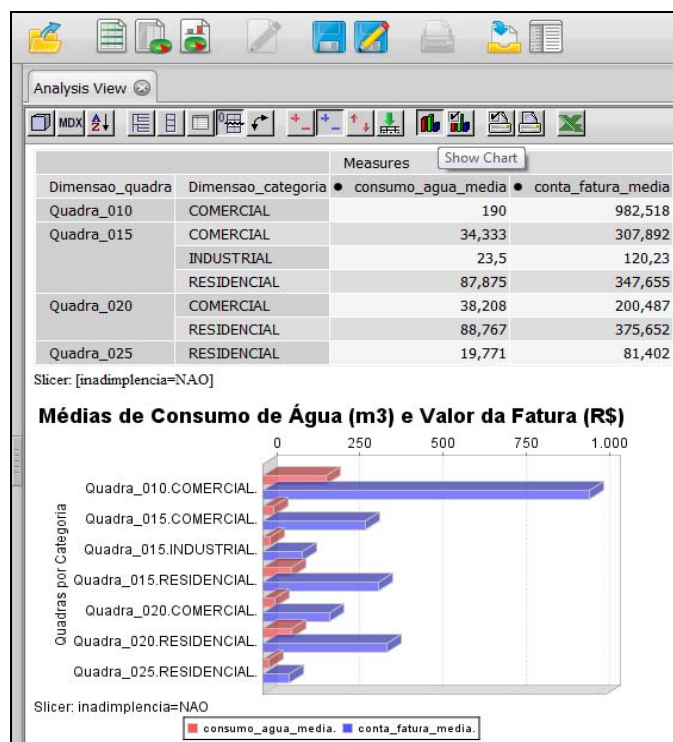


Figura 1: Visualização dos dados através de Ferramenta OLAP Pentaho Analysis View. Fonte: Dados do setor de saneamento de João Pessoa.

Os dados da Figura 1 foram obtidos através de uma consulta ao “Esquema Constelação de Fatos” implementado para o estudo de caso deste trabalho. O retorno desta consulta corresponde às médias de consumo de água em m³ e médias da fatura dos consumidores adimplentes agrupadas por quadra (010, 015, 020 e 025) e por categoria de consumo durante o período de 2007 a 2008. As novas informações úteis ao processo de tomada de decisão, dificilmente poderiam ser obtidas através de ambientes transacionais, ou seja, ambientes que não utilizam o DW, nem ferramentas OLAP.

A Figura 2 ilustra outra consulta executada sobre o cubo de dados “Perfil do Setor” do esquema Constelação de Fatos através da ferramenta *Pentaho Analysis View*. Esta consulta determina a quantidade de consumidores, pontos de utilização e quantidade de inadimplências da subcategoria FAVELA, associando-os com os agrupamentos das categorias de consumo (Comercial, Industrial, Público e Residencial), situações da ligação de água (Cortada, Ligada, Suprimida parcial e Suprimida total) e status de inadimplência (Adimplência e Inadimplência) dos consumidores.

As dimensões necessárias para execução da consulta acima foram: Categoria, Subcategoria, Água e Inadimplência. As métricas utilizadas foram: quantidade de consumidores, quantidade de pontos de utilização e quantidade de inadimplências.

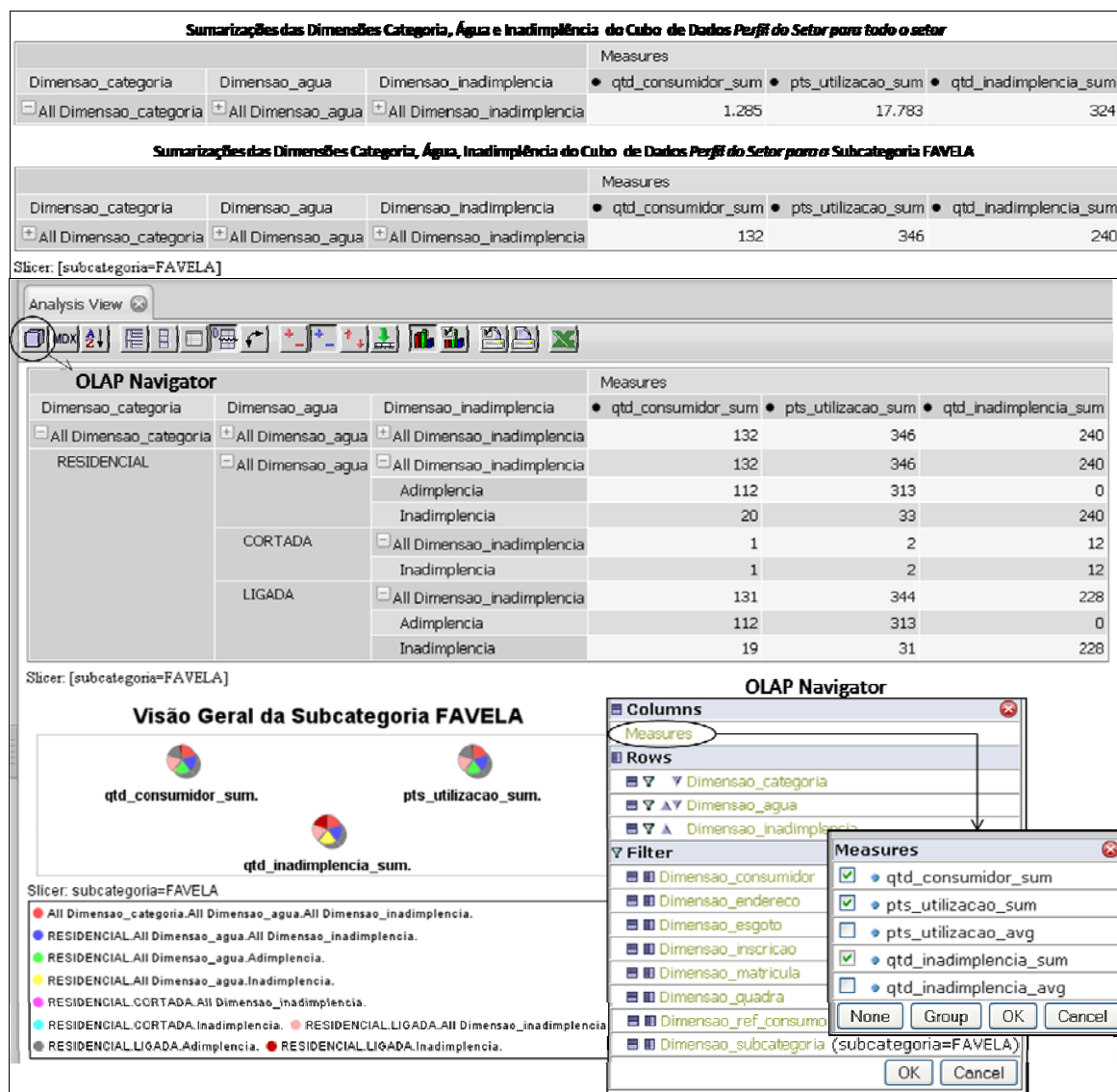


Figura 2: Consulta sobre o Perfil do Consumidor de baixa renda quanto à inadimplência

O resultado da consulta representado na Figura 2 informa que todos os consumidores da categoria FAVELA estão agrupados apenas pela categoria RESIDENCIAL e situação da água CORTADA e LIGADA. Dos 1.285 consumidores de todo o setor, 132 estão na subcategoria FAVELA. E das 324 inadimplências do setor, mais da metade, isto é 240, encontram-se na subcategoria FAVELA. Na situação da água CORTADA há um consumidor com 12 inadimplências. Já na ligação da água LIGADA, há consumidores inadimplentes (19) e adimplentes (112). Dos 346 pontos de utilização de água, 33 deles estão distribuídos entre os consumidores inadimplentes. Dos 15.420 consumidores de todo o setor estudado, 1.584 estão na subcategoria FAVELA. Dos 17.783 pontos de utilização de água de todo o setor, 346 pertencem a subcategoria FAVELA, sendo 33 de consumidores inadimplentes. Das 324 inadimplências encontradas em todo o setor, 240 estão na subcategoria FAVELA, ou seja, 74,1%.

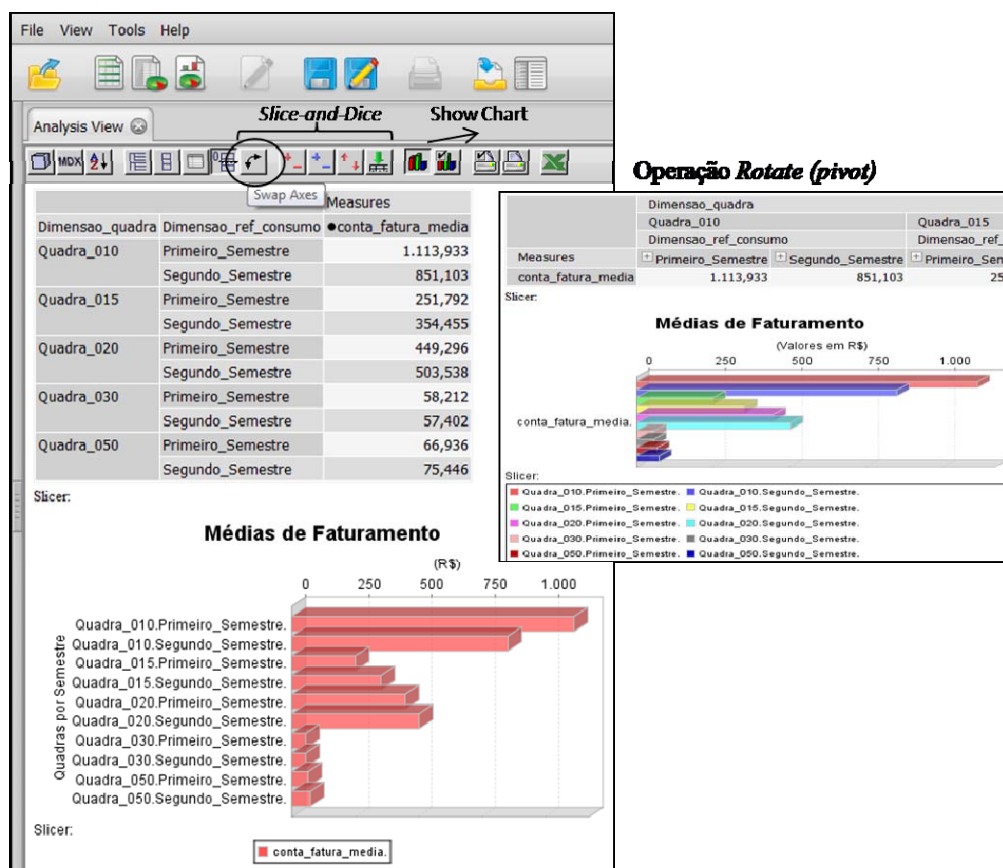


Figura 3: Consulta ao Esquema Constelação de Fatos. Fonte: Dados do setor de saneamento de João Pessoa.

A consulta da Figura 3 retorna as médias de FATURAMENTO das quadras (010, 015, 020, 030 e 050) agrupadas pela categoria de consumo COMERCIAL e SEMESTRES de referência. É destacado na Figura 3 o menu de navegação da ferramenta Pentaho Analysis View com as operações OLAP *slice-and-dice* e gráficas (Show Chart), estas que por sua vez, favorecem a exploração dos cubos de dados de forma mais intuitiva.

TRABALHOS RELACIONADOS

Um número significativo de trabalhos relacionados à área de pesquisa de *Data Warehouse*, OLAP e Mineração de Dados para o setor de Saneamento foi encontrado na literatura, dentre eles: (PASSINI 2002; TEIXEIRA 2006; QUEYROI 2007). Os resultados obtidos nos trabalhos mostram que a utilização de Mineração de Dados aplicada ao abastecimento de água acrescentou conhecimento ao especialista que antes era desconhecido. Um caso de sucesso de *Business Intelligence* foi a Companhia de Saneamento Básico do Estado de São Paulo (SABESP). A empresa implementou um *Data Warehouse* Corporativo e por meio dele obteve a capacidade de analisar o consumo de água com base em informações comparativas, permitindo melhorias e maior lucratividade na receita da ordem de R\$ 2 milhões por ano.

Informações deste tipo ainda precisam ser obtidas na região Nordeste, em especial, no Estado da Paraíba, foco e estudo de caso deste trabalho, por isso a importância e motivação deste estudo para região. Além disso, o uso de técnicas de Mineração de Dados junto ao *Data Warehouse* trará um ganho efetivo à entidade gestora do sistema de distribuição de água da Paraíba, haja vista que proporcionará um maior controle do comportamento dos consumidores e tomadas de decisões eficientes, visando à redução de perdas de água e das perdas econômicas, contribuindo para o uso racional e eficiente da água, esta que é um dos bens mais preciosos e essenciais para vida da população.



CONCLUSÃO

A análise estatística dos dados por meio das técnicas de Mineração de Dados em ambientes de *Data Warehouse* permite que se observem tendências e que se detectem regiões onde as Perdas Aparentes são mais frequentes nas diversas categorias de consumo. Portanto, as tecnologias apresentadas neste trabalho evidenciam o benefício e o valor agregado adquirido pela entidade gestora de sistema de abastecimento de água, pois obterá um maior controle do comportamento dos seus consumidores, proporcionando tomadas de decisões eficientes, que visam à redução de perdas de água e das perdas econômicas e desta forma contribuem para o uso racional e eficiente dos recursos hídricos.

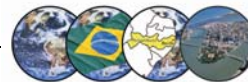
A principal fase do processo KDD, que é a Mineração de Dados, foi utilizada neste trabalho. Os dados foram analisados por meio de algoritmos a fim de produzir uma quantidade de padrões úteis, válidos e de fácil entendimento, com a finalidade de conduzir a tomadas de decisões inteligentes. A análise estatística dos dados, por meio da tecnologia de mineração permitiu que se observassem tendências, que se detectassem regiões onde as perdas aparentes são mais frequentes, que se estudassem os tipos de perdas que ocorrem com mais frequência nas diversas categorias de consumo, entre outros.

Através deste trabalho foi possível a detecção de Perdas Aparentes através da análise do comportamento do consumidor pelo uso de ferramentas de Mineração de Dados. A utilização de tecnologias de DM e DW para o setor de saneamento, tendo como principal motivação a questão do combate às perdas aparentes, com perspectiva nas crescentes irregularidades das ligações de água, evidencia-se a positividade e valor agregado para toda comunidade e segmento do saneamento urbano.

Portanto, neste trabalho foi implementado um sistema de extração de conhecimento em base de dados utilizando conjuntamente ferramenta de Mineração de Dados em ambiente de *Data Warehouse* com suporte às ferramentas OLAP. O DW proporcionou o armazenamento dos dados, transformando-os em informações confiáveis, que foram úteis ao processo de tomada de decisão. As ferramentas OLAP proporcionaram ao tomador de decisão, a formulação de hipóteses e a realização de consultas, visualizando o seu negócio através de várias dimensões. Enquanto que a utilização da Mineração de Dados apresentou-se como abordagem alternativa e automática de descobrir padrões nos dados, mostrando-se adequada para analisar grupos de dados, em virtude da dificuldade dos mesmos serem navegados ou explorados manualmente.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ALVES, Wolney Castilho, et al. PNCD A DTA D2 - 2004. Macromedição. Ministério das Cidades - Secretaria Nacional de Saneamento Ambiental, Programa Nacional de Combate ao Desperdício de Água - PNCD A. PNCD A DTA D2 - 2004. DTA D2 - Documento Técnico de Apoio nº D2.
2. ALVES, Wolney Castilho, et al. PNCD A DTA D3 - 2004. Micromedição. Programa Nacional de Combate ao Desperdício de Água - PNCD A, Secretaria Nacional de Saneamento Ambiental. PNCD A DTA D3 - 2004. DTA D3 - Documento Técnico de Apoio nº D3.
3. GOMES H. P., GARCÍA Rafael Pérez e REY Pedro L. Iglesias Abastecimento de Água - O estado da arte e técnicas avançadas (Livro). - 2007. - 1ª edição.
4. INMON W. H. Building the Data Warehouse (Livro). - (s.l.) : Editora: Wiley Publishing, Inc, 2005. - 4ª Edição.
5. KIMBALL Ralph e ROSS Margy The data warehouse toolkit : the complete guide to dimensional modeling (Livro). - (s.l.): John Wiley and Sons, 2002. - 2nd ed.
6. TANIAR David Data Mining and Knowledge Discovery Technologies (Livro). - (s.l.) : IGI Publishing, 2008.
7. GRAY Jim (et al.) (1996.). "Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals" (Periódico) / ed. Fayyad Usama.
8. INMON William H. (1992). "Building the Data Warehouse" (Livro). - NY: Editora: Wiley.
9. JAMES Kevin, CAMPBELL Stephanie L. e GODLOVE Christophe (2002) "Água e Energia: energia usada em sistemas de água". Editora: Alliance.
10. MARCKA, Estanislau, SILVA, Ricardo Toledo e CONEJO, João Gilberto Lotufo. Revisão 2004. Indicadores de Perdas nos Sistemas de Abastecimento de Água. Secretaria Nacional de Saneamento Ambiental, PNCD A - Programa Nacional de Combate ao Desperdício de Água. Revisão 2004. DTA A2 - Documento Técnico de Apoio.



11. MARQUES, Alfeu e SOUSA, Joaquim José de Oliveira. 2006. Hidráulica Urbana: Sistemas de Abastecimento de Água. 2006.
12. KIMBALL Ralph e ROSS Margy (2002) “The data warehouse toolkit: the complete guide to dimensional modeling” (Livro). - (s.l.): John Wiley and Sons, 2nd ed.
13. PASSINI Sílvia Regina Reginato (2002) “Data Mart para apresentação dos resultados econômico-financeiros da setorização. Dissertação de Mestrado.
14. PONNIAH Paulraj (2001) “Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals” (Livro). John Wiley & Sons.
15. QUEYROI Roberto (2007) “Aplicação de Modelo de Mineração de Dados em um Sistema de Apoio a Decisão para Empresas de Saneamento”. Dissertação de Mestrado / UFRJ - Rio de Janeiro.
16. SNIS (2007) “Sistema Nacional de Informações sobre Saneamento Diagnóstico dos Serviços de Água e Esgoto” / Ministério das Cidades - Brasília.
17. TEIXEIRA Cristina Josefa Santos (2006) “Descoberta de conhecimento em bases de dados como suporte a actividades de business intelligence: aplicação na área da distribuição da água”. Dissertação de Mestrado / Tecnologias e Sistemas de Informação – Universidade do Minho - Portugal.
18. ZIULKOSKI, L. C. C. 2003. Coleta de Requisitos e Modelagem de Dados para Data Warehouse: um estudo de caso utilizando Técnicas de Aquisição de Conhecimento. [Online] - Relatório, 2003. [Citado em: 04 de Abril de 2008.] <http://www.inf.ufrgs.br>.