

DESENVOLVIMENTO DE AMBIENTE PARA A GESTÃO DO CONHECIMENTO RELACIONADO AOS DADOS PRODUZIDOS PELO SISTEMA DE GERENCIAMENTO DE TRANSITO DA CIDADE DE JOINVILLE/SC - PARTE I**THIAGO L. DE OLIVEIRA^{1*}, THAYNARA A. M. DE JESUS², FERNANDO JOSÉ BRAZ³**¹ Instituto Federal Catarinense – Câmpus Araquari – IFC/ARA – Araquari/SC – thiagolocatellicdeoliveira@gmail.com; ² Instituto Federal Catarinense – Câmpus Araquari – IFC/ARA – Araquari/SC; ³ Orientador**ÁREA:** () Pesquisa; (X) Extensão **NÍVEL:** () Ensino médio; (X) Superior
CNPq-SETEC/MEC: Edital 17/2014**RESUMO**

A gestão de um ambiente passa pelo correto entendimento e interpretação dos dados que aquele ambiente produz. A tecnologia de banco de dados transacional permite o armazenamento e gerenciamento dos dados produzidos. Entretanto, se o foco da investigação acontece sobre resumos de dados e não detalhes individuais de dados, a ferramenta adequada para prover as capacidades de análise de dados é o *Data Warehouse*. Este projeto tem por objetivo principal o desenvolvimento de uma solução de Sistemas de Informação para o setor de Gestão de Trânsito da Cidade de Joinville. Neste documento serão apresentadas as principais ações desenvolvidas até o momento para a disponibilização deste ambiente de apoio ao processo de tomada de decisão na questão da gestão do trânsito do município de Joinville.

Palavras-chave: *Data Warehouse*; ETL; Pentaho.**INTRODUÇÃO**

O sistema do Departamento de Transito da cidade de Joinville (DETRANS) produz e armazena diariamente grande volume de dados gerados a partir do atendimento de ocorrências. Segundo informações do IBGE, a frota de veículos ativos na cidade no ano de 2014 envolvia cerca de 347.583 veículos. Fica evidente portanto, a importância de definir um ferramental que possa gerar informações do comportamento daquele ambiente e, a partir deles, oferecer suporte à tomada de decisão por parte dos gestores municipais. No processo de tomada de decisão estratégica e gerencial a granularidade dos dados a nível de registro não tem a mesma importância que o nível de resumos de dados. Para um gestor de trânsito, é muito mais importante ter conhecimento do montante de veículos envolvidos em irregularidades de trânsito em uma determinada região espaço-temporal da cidade, do que o conhecimento individual de cada um daqueles veículos. O conceito de *Data Warehouse* é utilizado exatamente com esta abordagem: servir de repositório de dados históricos e destinados a rápida produção de conhecimento a partir deles. A estrutura de *Data Warehouse* (DW), possibilita a análise de grande volumes de dados, através de medidas e dimensões. O presente projeto, que prevê o desenvolvimento de uma solução de descoberta do conhecimento relativo a área de gestão do trânsito de Joinville, envolve a etapa de

desenvolvimento de um *Data Warehouse* Espaço-temporal. O objetivo final desta estrutura de armazenamento de dados é servir como base para o desenvolvimento de ações de análise de dados. Estes processos de análise de dados poderão permitir a identificação de padrões e/ou tendências de ocorrências de eventos de trânsito em áreas espaço-temporais definidas pelo gestor do ambiente.

MATERIAL E MÉTODOS

O projeto foi iniciado em março de 2015 com término previsto para dezembro de 2016. A primeira etapa do projeto é a modelagem do *Data Warehouse* Espaço-Temporal que servirá como base para que o sistema de descoberta do conhecimento possa acontecer. Desde o ano de 2013 o NODES (Núcleo de Operacionalização e Desenvolvimento de Sistemas de Informação) do campus Araquari desenvolve o projeto de extensão denominado “Desenvolvimento de Sistema de Planejamento e Gestão de Trânsito para a Cidade de Joinville”. Naquele projeto, a proposta é o desenvolvimento de um ambiente que permita a automatização de rotinas da gestão do trânsito de Joinville. Aquela proposta envolve desde o desenvolvimento de uma solução de talonário eletrônico para o registro das autuações por parte dos agentes de trânsito. O banco de dados modelado para tal considera a normalização na sua terceira forma (3FN). Naquela proposta o foco era no registro individual de ocorrências, considerando também o georreferenciamento da ocorrência atendida. Embora um DW e banco de dados pareçam semelhantes, eles se distinguem no fator estrutural e funcional. Segundo Turban e Volonimo (2011, p.76) “um banco de dados é desenhado e otimizado para armazenar dados, enquanto os *Data Warehouses* são definidos para responder a questões de análise essenciais ao negócio.” Portanto, além da modelagem de um *Data Warehouse* ser orientada a fato e não a transação, é também preciso desenvolver uma atividade fundamental para que os dois ambientes (Transacional e DW) atuem em sincronismo: o processo de ETL (Extract, Transformation and Load) Extração, Transformação e Carga de dados. Este processo consiste, em uma definição bastante resumida, em transferir os dados de uma base transacional para um DW.

A arquitetura do DW utilizada no início do projeto foi a do tipo estrela (*Star Schema*). Este tipo de modelagem contém uma tabela de fatos central ligada diretamente à tabelas dimensionais. A tabela de fatos armazena medidas do fato que se investiga, que no caso deste projeto é “ocorrências de trânsito”. Dentre as medidas é possível citar quantidade e número médio de ocorrências. Enquanto que as tabelas de dimensão contém descrições sobre cada elemento que faz parte do processo. Isto oferece a possibilidade de aumento da capacidade de análise das medidas do fato.

Entretanto, devido alterações ocorrentes no banco de dados do DETRANS, a arquitetura utilizada passou a ser do tipo floco de neve (*Snowflake Schema*). Essa arquitetura consiste, em uma extensão do esquema estrela, onde algumas tabelas de dimensão experimentam mais um nível de normalização.

Após a definição da arquitetura, foi necessário utilizar uma ferramenta que facilitasse o desenvolvimento da solução proposta. Então foi considerado como melhor opção o software Pentaho (Pentaho Community Edition Business Intelligence e Business Analytics), versão 5.3 lançada em fevereiro de 2015, para a realização da atividade de ETL. Esse software simplifica a preparação, análise e visualização dos dados. Pentaho é uma plataforma completa de BI (Business Intelligence; Inteligência Empresarial), desenvolvido em Java com vantagens que vão desde sua distribuição gratuita a fácil integração com diferentes fontes de dados e aplicativos que utilizam padrões abertos. A ferramenta possibilita também o uso de APIs e tem flexibilidade nas opções de saída podendo ser em vários formatos, Adobe PDF, HTML, texto, etc. Além disso, é uma ferramenta Multiplataforma, suportada pelos sistemas operacionais Windows, Linux, Macintosh, Unix, etc.

De acordo com Oliveira (2015), o Pentaho é composto por várias aplicações, as utilizadas no projeto foram:

- BI Server: responsável pela interação com o usuário. Fornece serviços como a interface de visualização dos relatórios; área de administração do console; criação de relatórios; visão multidimensional dos dados;

- Pentaho Data Integration: Possibilita a atividade de ETL, preparo de fontes de dados para análise, mineração e a geração de relatórios.
- Pentaho Schema Workbench: Ferramenta gráfica desktop, que cria esquemas ROLAP (Relational On Line Analytical Processing) para análise.

RESULTADOS E DISCUSSÃO

A figura 1 representa uma parte do DER (Diagrama Entidade-Relacionamento) do Banco de Dados relacional do DETRANS utilizado com base para projeto.

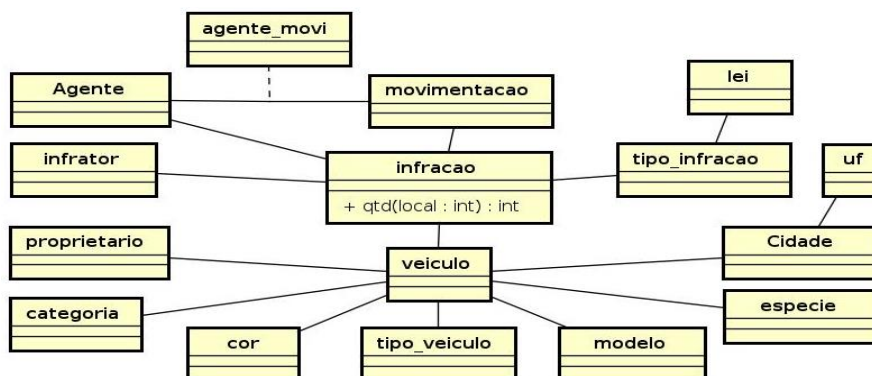


Figura 1 Representação de parte do DER do DETRANS

Esse processo gerou algumas versões do DW no decorrer do projeto. Na primeira versão foi utilizada a arquitetura do tipo estrela, com a tabela “infracao” definida como tabela de fatos e as tabelas: “tipo_infracao”, “agente”, “infrator” e “veiculo”, tabelas de dimensão.

Na segunda versão da modelagem foi identificado o problema de redundância de dados, o que inviabilizou a utilização do esquema. Por conta disso, uma alteração foi necessária, e a terceira versão, representada na figura 2, foi elaborada. Nesse momento houve uma importante alteração no esquema da base principal, que ocasionou a alteração da arquitetura do DW que passou a ser do tipo floco de neve.

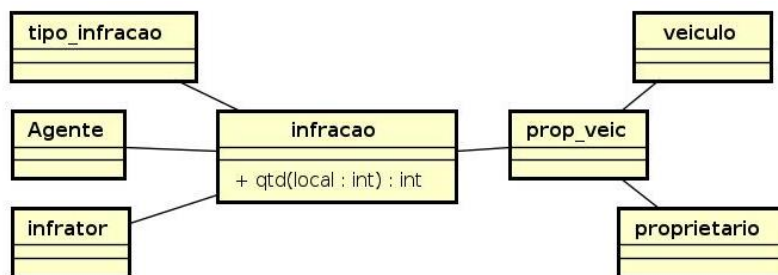


Figura 2 Terceira versão da modelagem

O modelo atual do DW permitirá resolver consultas do tipo:

- Quantidade de infrações de determinado tipo considerando região de ocorrência;
- Qual é o padrão das ocorrências de trânsito em Joinville;
- Quantidade de ocorrências de trânsito agrupadas por tipo de veículo, data, região;

Ao fim do projeto, estas e tantas outras consultas poderão ser realizadas pelo gestor do sistema considerando o georeferenciamento dos dados. A emissão da consulta poderá definir regiões de mapas digitais para sua execução. Esta funcionalidade acontecerá através do desenho de polígonos sobre um mapa digital apresentado na tela da solução.

CONCLUSÕES

O alto número de veículos na cidade de Joinville, implica na necessidade de um ambiente adequado tanto para o registro quanto para o gerenciamento dos dados relacionados com os eventos do trânsito. Entretanto, para isso é imprescindível que os dados estejam em formato e ambiente adequados. Considerando que um *Data Warehouse* é um repositório de dados históricos que são destinados a produção de conhecimento, este ambiente de armazenamento e gerenciamento de dados é fundamental para suportar um sistema de informação. Na primeira etapa deste projeto a proposta principal é o desenvolvimento e implantação de um *Data Warehouse* Espaço Temporal. Na segunda etapa do projeto acontecerá o desenvolvimento da solução de sistemas de informação visando revelar e apresentar o conhecimento. Na etapa inicial do projeto a utilização do software Pentaho foi de extrema valia para o desenvolvimento das atividades. A característica de dados georeferenciados, neste projeto, empresta um nível de complexidade maior para a modelagem do *Data Warehouse*. Entretanto, até o presente momento este procedimento tem sido executado sem grandes problemas para a equipe. Considerando que o projeto tem término previsto para o fim do ano de 2016, algumas etapas não foram satisfeitas até o momento. Esta primeira parte do projeto foi limitada ao estudo da base de dados transacional, definição do esquema e processos de cargas de dados do DW e, utilização dos conceitos e ferramentas de bancos de dados geográficos.

REFERÊNCIAS

- IBGE. 2014. *Frota Municipal de Veículos*. (<http://goo.gl/yc1xBv>). Acessado: 16/09/2015
- LAPENA, R e F. JAVIER. *Introducción al Desarrollo de Soluciones*, E.I.F. Madrid, AT&T Global Information Solutions, p.121995. In: RAMOS, Carla Sofia D. Moreira. *O Impacto das Tecnologias de Informação ao Nível da Gestão de Relacionamentos*. Universidade do Porto: 2002. p.320, 1995.
- LOPES, Maurício C.; OLIVEIRA, Percio A. *Ferremanta de Contrução de Data Warehouse*. 2007. 13 p. Departamento de Sistemas e Computação, Universidade Regional de Blumenau, Blumenau. 2007. In: NAVARRO, Maria C. *O que é Data Warehouse?* Brasília, 1996. (<https://www.serpro.gov.br/publicacao/tematec/1996/ttec27>). Acesso em: 13/04/2007
- OLIVEIRA, Vinícius. 2015. *Pentaho - Visão Geral*. Blnaprática: Tudo sobre Business Intelligence. (<http://www.binapratica.com.br/#!/visao-pentaho/c7gy>). Acessado: 04/09/2015
- PENTAHO CORPORATION, 2015. (<http://www.pentaho.com/>). Acessado: 27/05/15
- TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay E.; KING, David. *Business Intelligence: Um Enfoque Gerencial Para a Inteligência do Negócio*. Bookman. São Paulo: 2008. 207 p.
- TURBAN, Efraim; VOLONIMO, Linda. 2011. *Tecnologia da Informação para Gestão: Em Busca de um Melhor Desempenho - 8ªed*. Porto Alegre: Bookman. 123p.