

Um estudo de caso sobre a tecnologia de *Business Intelligence* na área de tráfego de uma praça de pedágio

Joubert de Castro Lima¹, Luiz Alexandre H. S. Maciel¹, Celso Massaki Hirata¹,
Edgar Toshiro Yano¹, Adilson Marques da Cunha¹ e Maurício Micoski²

¹Instituto Tecnológico de Aeronáutica (ITA) - Praça Marechal Eduardo Gomes, 50 -
Vila das Acácias - CEP 12228-900 – São José dos Campos – SP – Brasil.

²Compsis – Computadores e Sistemas Industriais e Comerciais Ltda - Rua
Pindamonhangaba, 160 – Vila Nova Conceição - CEP 12231-090– São José dos
Campos – SP – Brasil.

{joubert, luizh, hirata, yano, cunha}@ita.br,
Mauricio.micoski@compsisnet.com.br

Abstract. *This article aims to implement a Business Intelligence (BI) system in the area of traffic of a toll plaza located between the cities of Delhi and Noida in India, in order to investigate how much the BI technology reduces the waste of resources involved in decision support of this market niche. The following items are included in the scope: (1) elicitation of functional and non-functional requirements of the collection process and the pouch process of a toll plaza, (2) identification of an architecture based on layers for BI solution, (3) definition of the set of techniques and services offered in each layer, and (4) assessment of the suite of a vendor when it is used in the implementation of a BI system for toll plaza described in this case study.*

Resumo. *Este artigo tem como objetivo implementar um sistema de Business Intelligence (BI) para a área de tráfego de uma praça de pedágio situada entre as cidades de Delhi e Noida na Índia, com o intuito de investigar o quanto a tecnologia de BI pode reduzir o desperdício de recursos envolvidos no suporte a decisão deste nicho de mercado. Pertencem ao escopo deste estudo: (1) o levantamento dos requisitos funcionais e não funcionais da arrecadação e dos malotes de uma praça de pedágio, (2) a identificação de uma arquitetura em camadas para uma solução de BI, (3) a definição do conjunto de técnicas e serviços oferecidos em cada camada, e (4) a avaliação da suíte de um vendedor quando utilizada na implementação de um sistema de BI para a praça de pedágio descrita no estudo de caso.*

1. Introdução

O processo de privatizações de rodovias de importância econômica cresce em todo Brasil, assim como em parte do mundo. A concessão de parte da malha rodoviária pode abranger várias praças de pedágio e várias rodovias. Para gerir praças de pedágio são utilizados Sistemas Inteligentes de Transporte - SIT (*Intelligent Transportation Systems - ITS*) compostos por tecnologias de comunicação e de controle integradas à infraestrutura de um sistema de transporte.

Uma das maiores dificuldades de uma praça de pedágio é não dispor de relatórios oriundos de diversas fontes de dados, que possam ser construídos, alterados e compilados pelo usuário final de maneira iterativa e incremental, com desempenho e com visualizações que enfatizem cores, texturas e ângulos, por exemplo. Uma outra dificuldade é a falta de uma solução computacional que permita inferir padrões comportamentais passados, não triviais, implícitos, previamente desconhecidos e potencialmente úteis, tanto de arrecadadores quanto de veículos, transportadoras, instituições financeiras e clientes.

O objetivo principal deste trabalho é implementar um sistema de Inteligência de Negócio - IN (*Business Intelligence BI*) para a área de tráfego de uma praça de pedágio situada entre as cidades de Delhi e Nodia na Índia. O objetivo secundário é avaliar um conjunto de ferramentas que compõe a suíte de *BI da IBM®*, quando utilizada para gerir a arrecadação e os malotes de uma praça de pedágio. Em consequência disso, acredita-se na possibilidade de investigar até que ponto a tecnologia de *BI* reduz o desperdício de recursos envolvidos como suporte à tomada de decisão neste nicho de mercado.

O restante deste artigo é dividido como se segue. A Seção 2 apresenta a descrição de uma praça de pedágio. A Seção 3 descreve os requisitos funcionais e não funcionais das subáreas de arrecadação e de malotes. A Seção 4 apresenta a arquitetura para o sistema de *BI*. A Seção 5 apresenta uma avaliação das ferramentas que compõem a suíte de *BI da IBM®* e, finalmente, a Seção 6 sintetiza as principais conclusões e propostas para trabalhos futuros.

2. Arrecadação e Malotes numa Praça de Pedágio

As descrições desta Seção foram elaboradas com base nas informações obtidas junto à COMPSIS, uma empresa brasileira de médio porte, que atua no desenvolvimento de soluções para gestão rodoviária em mais de seis países. O cliente da COMPSIS escolhido para realizar este estudo de caso foi uma empresa que possui uma praça de pedágio numa ponte entre as cidades Delhi e Noida, na Índia. Neste estudo de caso foi utilizada a base de dados operacional dos três primeiros anos desta praça. A seguir, passa-se a discutir como uma praça de pedágio é organizada.

Sentido da pista, pista, praça, rodovia e concessionária compõem a visão hierárquica das abstrações normalmente encontradas numa praça de pedágio. As pistas, do ponto de vista operacional, são agrupadas de acordo com o sentido, e podem ser reestruturadas a qualquer momento. Em cada pista, existe uma cabine de arrecadação que tarifa os veículos passantes, de acordo com a sua categoria. No sistema da COMPSIS, os veículos são divididos, normalmente, em oito categorias, variando desde motocicletas até carretas com 8 a 10 eixos.

Existe um serviço chamado Coletor Eletrônico de Pedágio (*Electronic Toll Collection -ETC*), em que o motorista paga uma mensalidade para que o seu veículo possa transitar livremente numa determinada praça, ou seja, o motorista utiliza uma pista automática onde não é necessário parar para efetuar o pagamento. Numa praça de pedágio, são normalmente aceitos pagamentos do tipo ETC, em débito, cartões de crédito, dinheiro ou cheque. Numa praça de pedágio, pode-se aceitar mais de um tipo de moeda, dependendo da política econômica do país.

Os funcionários são classificados como gerente, coordenador e arrecadador de praça. O arrecadador fica na cabine, classificando os veículos, de acordo com sua respectiva categoria, e coletando os valores dos pedágios. Considerando-se que tal classificação pode gerar erros, sensores realizam a detecção automática dos veículos para tarifação. Para os casos onde os valores detectados e classificados não coincidam, o coordenador da praça atribui o valor definitivo que deveria ter sido cobrado. Este valor definitivo, obviamente, é subjetivo e, desta forma, sujeito à fraude.

Periodicamente, são enviados malotes com os valores coletados de uma praça de pedágio para uma instituição financeira. As transportadoras são as responsáveis por este serviço. Cada malote registrado no sistema contém o valor declarado pelo funcionário da praça, o valor rejeitado pelo banco, e o valor definitivamente depositado. Este procedimento possibilita uma avaliação de hábitos de praças e bancos.

3. Requisitos de um Sistema de BI para a Área de Tráfego de uma Praça de Pedágio

A partir de entrevistas e pesquisas sobre soluções de BI já implementadas nos dois últimos anos, foram identificados os requisitos do sistema e realizadas entrevistas com especialistas em Tecnologia da Informação – TI (Information Technology – IT) da COMPSIS, ao longo de dois meses. A seguir, descreve-se os principais requisitos endereçados e já priorizados em dois grupos: funcionais e não funcionais.

3.1. Requisitos Não-Funcionais para uma solução de BI de pedágio

O sistema deve ser capaz de propiciar: o processamento de relatórios com múltiplas visões do negócio, a partir de milhões de registros, em menos de 5 minutos; a construção de relatórios, de maneira interativa e incremental, onde visões de negócio possam ser adicionadas ou suprimidas a qualquer momento; a configuração de abstrações tais como escalonadores de eventos, atividades de pré-processamento da base de dados operacional, esquemas multidimensionais, cubos, e técnicas de *data mining*; a importação e exportação de metadados pertinentes a um sistema de BI como, por exemplo, esquemas, cubos, medidas, dimensões e fatos (independente de ferramenta, proprietária ou não); e a importação de dados oriundos de diversas fontes relacionais, textuais ou orientadas a objeto, de diversos tipos, incluindo imagem e vídeo.

3.2. Requisitos Funcionais para arrecadação e malotes:

O sistema deve ser capaz de propiciar: consultas a relatórios que forneçam visões de cargos de uma praça de pedágio, funcionários, categorias de veículo, cliente ETC, data, hora, tipo de pagamento, moeda, abstrações de uma praça conforme descrição da Seção 2, e turnos de trabalho.

Com base nessas visões, o sistema deve ser capaz de propiciar também os seguintes tipos de mensurações sobre os arrecadadores: total, média, mínimo e máximo, e coeficiente padrão dos erros operacionais, do valor classificado pelo arrecadador, do valor detectado pelo sensor, e do tempo de atendimento; total, média, valores mínimos e máximos arrecadados, e valor do troco repassado ao cliente; total e média do número de veículos, e valores pagos pelos clientes.

O sistema deve ainda ser capaz de permitir: consultas a relatórios que forneçam visões de contas bancárias, agências bancárias, bancos, malotes, datas e horas sobre

criação e envio, correção dos malotes, transportadoras, cargos de uma praça de pedágio, funcionários e tipo de pagamento. Para essas visões, devem ser também propiciadas mensurações de: total; média; e mínimo e máximo do valor declarado, do valor depositado, e do valor recusado pelo banco.

O sistema deve também ser capaz de possibilitar consultas a relatórios que reúnam informações: sobre malotes, contas, transportadoras, bancos e seus respectivos valores declarados, depositados e recusados; sobre arrecadadores, coordenadores, abstrações de uma praça de pedágio, conforme descrito na Seção 2, com suas respectivas medidas, erros operacionais, valor detectado e classificado, e tempo de atendimento.

Desta forma, deve ser possível reunir-se, numa única visão, as áreas arrecadação e malotes. Por fim, o sistema deve ser capaz de propiciar também a capacidade de inferir o comportamento passado de veículos e arrecadadores.

4. Arquitetura de um Sistema de BI

A estratégia para endereçar uma possível solução para o problema descrito na introdução foi a de definir uma arquitetura em camadas (Figura 1). Inicialmente, mapeia-se os dados de uma fonte de dados original (camada 0), depois se extrai, transforma, carrega e modela, multidimensionalmente, os dados em repositórios unificados (camada 1) para que serviços de análise possam ser utilizados (camadas 2, 3).

4.1. Serviços e Técnicas Implementadas

Cada camada é composta de um conjunto de técnicas e serviços amplamente difundidos nos sistemas de *BI*, mas nunca aplicadas antes ao setor de pedágio.

A primeira técnica utilizada na camada 1 é a integração de dados oriundos de diversas fontes da camada 0. No estudo de caso do presente trabalho, a praça de pedágio utiliza o Banco de Dados Oracle, algumas tabelas independentes de SGBD e *flat files*. Das tabelas Oracle foram importadas a maioria dos dados que formam as dimensões e tabelas de fatos do sistema de *BI*. Os dados de meses, bimestres, trimestres e semestres do ano, além de dias do mês, quinzenas do mês e os feriados do ano e suas associações ao respectivo dia, mês e ano em questão foram obtidos de *flat files*.

Uma vez que os dados estejam integrados e mapeados a metadados numa base alvo, também chamada *target database*, faz-se necessário aplicar outras técnicas pertencentes à camada 1, mais especificamente ao módulo chamado *Extract Transform and Load* (ETL). Neste módulo, são aplicadas técnicas como *data cleaning*, *data transformation* e *data reduction*.

No sistema houve necessidade de limpar os dados, uma vez que atributos como valor pago, classe de veículo identificada pelos arrecadadores e sensores, tempo de atendimento, código do funcionário, data e hora da transação, conta bancária, código do malote e código do banco possuíam valores nulos que poderiam alterar significativamente a fase de análise. A eliminação de ruídos ocorreu principalmente no atributo tempo de atendimento. Valores discrepantes nos tempos não foram considerados para evitar erros de análise.

A transformação dos dados foi necessária para agregar, numa única tabela de fatos, os atributos valor pago, valor depositado e tempo de atendimento. A partir da agregação desses atributos foi possível reunir as medidas essenciais à arrecadação e aos malotes de uma praça de pedágio.

A redução dos dados foi necessária, e para isso, foi utilizada a discretização numérica. Atributos como tempo de atendimento, valores no malote e na arrecadação, já separados em dinheiro e cheque, foram discretizados em índices alto, médio e baixo, a fim de facilitar a identificação de padrões nas camadas 2 e 3.

Na camada 1, tem-se também os *Data Warehouses (DWs)*, repositórios que implementam uma coleção de dados orientada por assunto, integrada, variante no tempo e não volátil, que tem por objetivo dar apoio ao processo de tomada de decisão [Inmon 1997]. A partir de um *DW* é possível criar visões específicas, chamadas de *Data Marts (DM)*. O sistema de *BI* desenvolvido foi dividido em *DMs* para arrecadação e malotes que compõem o *DW* pedágio. Em cada *DM* foi implementado um conjunto de dimensões e medidas, de acordo com as necessidades da COMPSIS, relatadas na Seção 3.

O *DM* de arrecadação possui 6 tabelas de fatos, sendo 3 para registros válidos dos anos 2001, 2002 e 2003, e 3 para registros inválidos nos mesmos anos. Considera-se como registro válido, aquele em que o valor classificado é igual ao valor detectado. Foram implementadas 3 tabelas de fatos para o *DM* de malotes a partir dos mesmos anos. A granularidade dos fatos armazenados foi definida como cada veículo passante e cada malote enviado. Qualquer tipo de sumarização dos dados poderia trazer perdas na acurácia da análise proposta. A Figura 2 ilustra dois *DMs*, um para arrecadação e outro para malotes.

A análise de um *DM* pode ser feita a partir de sentenças SQL, computação de cubos multidimensionais, ambos métodos de tentativa e erro, ou por técnicas de mineração do *DM*. Na camada 2, tem-se o desenvolvimento de cubos que são arranjos (*lattice*) de *cuboids* [Gray 1997]. *Cuboids* são combinações de atributos dimensionais. No sistema desenvolvido, foram criados 27 diferentes cubos para as áreas de arrecadação e malotes. Alguns cubos continham dimensões e medidas pertinentes a ambas as áreas. Assim, foi possível verificar a flexibilidade de relatórios e a sofisticação da visualização do servidor *Online Analytical Processing – OLAP* da IBM® que mantém cubos em sua estrutura interna e os computa como objetos em memória *MOLAP (Multidimensional OLAP)* ou como tabelas num *SGBD* comercial - *ROLAP (Relational OLAP)* ou mesclando estas duas estratégias numa *HOLAP (Hybrid OLAP)*.

As técnicas aplicáveis à *DM*, ainda na camada 2, incluem associações/correlações, predições, classificações e análise de cluster. Neste trabalho, foi utilizada a técnica chamada de Regras de Associação, inicialmente formulada em [Agrawal 1993].

A partir de regras como Pagamento (X, “ETC”) => Pagamento (X, “Cartão”) [suporte = 34%, confiança = 79%] foi possível identificar tipos de pagamentos preferenciais de determinadas categorias de veículos, em épocas do ano e horas do dia distintas. De maneira análoga, foi possível identificar tipos de pagamentos preferenciais em pistas, praças, rodovias e até concessionárias. Com isso pôde-se implantar segmentações de *marketing*, reconfigurações de pistas, entre outras estratégias.

Foram geradas regras de associação para detectar combinações de arrecadadores com menor ou maior taxa de erros na arrecadação, maior ou menor tempo de atendimento para um turno específico ou para uma época específica do ano, horários habituais onde ocorriam maiores taxas de erros operacionais, maiores tempos de atendimento, horários habituais de tráfego de determinadas categorias de veículos, comportamentos suspeitos de arrecadadores e coordenadores, fabricantes de sensores com baixa acurácia, assumindo que os veículos detectados por estes quase nunca coincidiam com os classificados pelos arrecadadores num determinado turno do dia.

A camada 3 direciona a análise para a integração das tecnologias *OLAP* e *DM*. Faz-se uso do desempenho e da flexibilidade dos cubos, reunindo a isso a possibilidade de inferir o comportamento passado, com o intuito de prever o futuro. Infelizmente, a ferramenta de *BI* da *IBM* não possui uma licença única, que permita o acesso a toda suíte. Servidores *MOLAP*, *HOLAP* e *OLAPMining* não são oferecidos na licença *trial*.

O projeto do sistema descrito nesta seção foi planejado para 18 meses. Além do sistema, foi elaborada a especificação completa do mesmo, incluindo os esquemas multidimensionais, *ETLs*, cubos, regras de associação, documentação de instalação e operação da suíte da *IBM*®.

5. Avaliação da Solução de BI Utilizada

Os experimentos foram realizados na suíte de *BI IBM DB2* versão 8.1. O servidor de *BI* foi executado numa máquina PC com 2.1Ghz de *clock*, 80GB de disco (7200 RPM) e 1GB de memória RAM. A suíte da *IBM*® ao ser usada para implementar o estudo de caso da Seção 4, apresentou algumas conformidades e desconformidades em relação aos requisitos descritos na Seção 3. A tabela a seguir ilustra as considerações sobre a ferramenta utilizada.

Tabela 1. Avaliação da solução de BI da IBM

Característica	Camada
A ferramenta não é aberta. Os diversos metadados implementados em cada camada podem ser importados/exportados somente para soluções da <i>IBM</i> ®.	1, 2, 3
A solução <i>IBM</i> ® não implementa <i>cache</i> no cliente, ou seja, é fortemente centrada no servidor, impossibilitando em sua arquitetura que clientes realizem parte do processamento do servidor.	1, 2, 3
A suíte da <i>IBM</i> ® é flexível no que tange as diferentes configurações para <i>ETLs</i> , cubos, dimensões, medidas, hierarquias dimensionais e esquemas.	1, 2, 3
Os serviços de <i>DW</i> , <i>OLAP</i> ou <i>data mining</i> podem ser acessados de maneira transparente ao usuário.	1, 2, 3
A suíte é segura, permitindo a criação de usuários e grupos de usuários para cada camada da arquitetura proposta na Figura 1.	1, 2, 3
O módulo <i>ETL</i> do servidor <i>DW</i> é flexível, pois permite a importação de dados de fontes heterogêneas.	1
O módulo <i>ETL</i> do servidor <i>DW</i> pode ter baixo desempenho. Uma atividade de <i>ETL</i> utilizada para agregar três tabelas do banco operacional e criar novos atributos dimensionais na tabela de fatos demorou 27 horas, persistindo 73 milhões de tuplas numa primeira carga dos dados.	1
O módulo <i>ETL</i> do servidor <i>DW</i> permite a construção de <i>workflows</i> para carga, limpeza e transformação. É possível especificar dependências entre os fluxos e escalonadores a cada fluxo. Infelizmente, tais escalonadores não permitem a execução do fluxo em tempo real.	1
O servidor de <i>DW</i> não explicita a definição de dimensão e tabela de fatos. Não é possível, por	1

exemplo, definir várias tabelas e várias hierarquias para uma única dimensão.	
Os servidores <i>OLAP da IBM®</i> permitem <i>tuning</i> na base de dados, o que aumenta muito o desempenho das demais camadas de carregamento e análise de dados.	2
O servidor <i>OLAP Cube Views</i> oferece somente o armazenamento de cubos multidimensionais, a partir de bases relacionais (<i>ROLAP</i>)	2
O servidor <i>Cube Views</i> , a aplicação <i>Office Connect</i> , e o servidor <i>Intelligent Miner</i> não oferecem os resultados do desempenho da computação dos cubos e da geração de regras, respectivamente.	2
Infelizmente, os servidores <i>OLAP</i> requerem um nome único para cada membro de um cubo.	2
A estratégia de particionar cubos de maneira transparente ao usuário melhora o desempenho global. Existem 3 tipos de particionamento na suíte <i>IBM®</i> : <i>replicated</i> , <i>transparent</i> , e <i>linked</i> . Com exceção da partição <i>linked</i> , os servidores <i>OLAP</i> obrigam que cubos virtuais, oriundos de dois ou mais cubos particionados, tenham as mesmas dimensões. Isso reduz drasticamente a flexibilidade no desenvolvimento. Servidores <i>Cube Views</i> não oferecem o conceito de cubos virtuais.	2
Infelizmente, o servidor <i>IBM Intelligent Miner</i> não implementa <i>multi-dimensional association</i> , impossibilitando encontrar padrões extremamente úteis ao cliente.	2
A visualização de cubos pode ocorrer com a utilização de <i>plugins</i> para o Excel, ou com a confecção de páginas web com tabelas dinâmicas. Infelizmente, a estratégia de <i>plugins</i> não se aplica ao servidor <i>IBM Intelligent Miner</i> . Este faz uso de visualizadores próprios para regras de associação, os quais não oferecem ângulos, texturas, mas somente cores e textos na visualização.	2

De maneira geral, a suíte da *IBM®* atende às necessidades endereçadas na Seção 3. As maiores críticas são atribuídas: ao servidor de *DW*, uma vez que o mesmo não possui conceitos pertinentes à área; a impossibilidade de se minerar regras do tipo *multi-dimensional association*, por parte do servidor *IBM Intelligent Miner*; e a baixa flexibilidade de importação/exportação de metadados implementados na suíte *IBM®* para outros fornecedores de soluções de *BI*.

6. Conclusões e Trabalhos Futuros

O sistema de *BI* implementado conseguiu reunir, num único repositório, dados operacionais de uma praça de pedágio, facilitando com isso que relatórios pudessem ser construídos, alterados e compilados pelo usuário final, de maneira iterativa e incremental, com desempenho e com visualizações que enfatizam cores, texturas e ângulos, por exemplo. Além disso, as regras de associação implementadas proveram a capacidade de inferir padrões comportamentais passados de arrecadadores, fabricantes de sensores, veículos, clientes ETC, transportadoras e instituições financeiras.

O sistema atendeu a todos requisitos funcionais e a maioria dos não funcionais endereçados, e demonstrou enormes possibilidades para avanços na gestão de áreas complementares as de arrecadação e malotes, numa praça de pedágio.

Alguns trabalhos futuros incluem: estender o uso da tecnologia de *BI* para definir localizações estratégicas de veículos de emergência nas rodovias, para priorizar as manutenções futuras das rodovias, para ajustar semáforos e sinalizadores, para planejar e construir passarelas de pedestres, e para expandir trechos da malha rodoviária.

No âmbito governamental, soluções de *BI* podem trazer resultados que permitam os órgãos executores traçarem políticas que, por exemplo, incentivem o tráfego de cargas durante a madrugada, e escalonem horários mais adequados aos trabalhadores e às escolas da região.

O atual projeto para arrecadação e malotes está pronto, mas merece avanços contínuos na definição de novas dimensões e novas medidas. Além disso, novas técnicas de *data mining* e *OLAPMining* merecem ser avaliadas.

É recomendável realizar-se uma avaliação da aplicabilidade de outras soluções de *BI* na área de tráfego de uma praça de pedágio, com o intuito de fornecer um estudo comparativo com o presente trabalho. Como exemplos de outras soluções proprietárias de *BI*, pode-se citar as fornecidas pelas empresas *Oracle®* e *Microsoft®*. Como exemplos de soluções livres de *BI*, citam-se projetos como Mondrian [Mondrian 2005], Jpivot [Jpivot 2005] e Weka [Weka 2005].

Referências

- R. Agrawal, T. Imielinski, and A. Swami (1993). "Mining association rules between sets of items in large databases". In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D.C..
- J.Gray, S.Chaudhuri, A.Bosworth, A.Layman, D.Reichart, M.Venkatrao, F.Pellow e H.Pirahesh (1997). "Data Cube: A Relational Aggregation Operator Generalizing Goup-by, Cross-tabs, and Sub-totals". Data Mining and Knowledge Discover, 1:29-54.
- Inmon, W.H. (1997), Como Construir o Data Warehouse, Campus, Rio de Janeiro, 2ª edição.
- Jpivot Project (acessado em 22/07/2005), <http://jpivot.sourceforge.net/>.
- Mondrian Project (acessado em 22/07/2005), <http://perforce.eigenbase.org:8080/@md=d&cd=//open/mondrian/doc/&c=yJZ@//open/mondrian/doc/index.html>.
- Weka Project (acessado em 22/01/2005), <http://www.cs.waikato.ac.nz/~ml/weka/>.

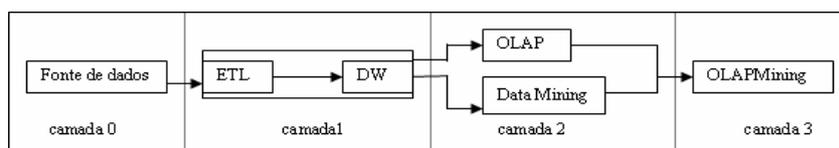


Figura 1. Uma típica arquitetura de um sistema de BI

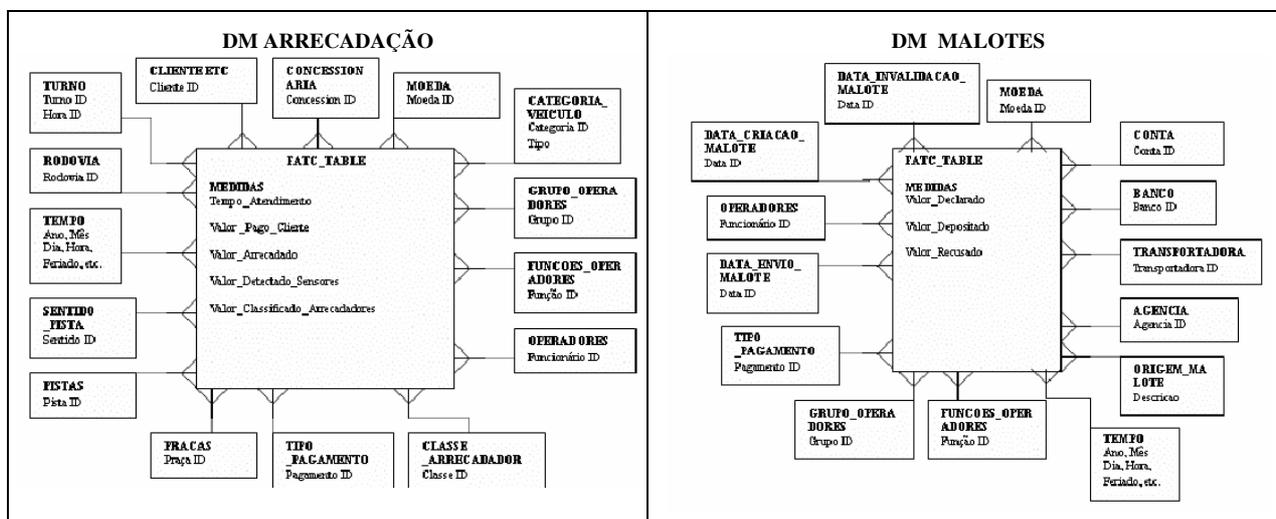


Figura 2. DMs arrecadação e malotes que compõem o DW pedágio