



COPPE/UFRJ

APLICAÇÃO DE TÉCNICAS DE DATA MINING E FERRAMENTAS OLAP NA BASE
DE DADOS INTEGRADA PIATAM

Ricardo de Souza Alencar

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Civil.

Orientador(es): Alexandre Gonçalves Evsukoff

Fernando Pellon de Miranda

Rio de Janeiro

Junho 2009

Livros Grátis

<http://www.livrosgratis.com.br>

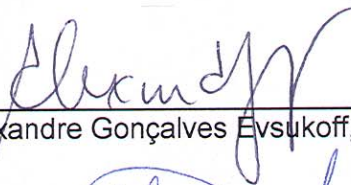
Milhares de livros grátis para download.

APLICAÇÃO DE TÉCNICAS DE DATA MINING E FERRAMENTAS OLAP NA BASE
DE DADOS INTEGRADA PIATAM

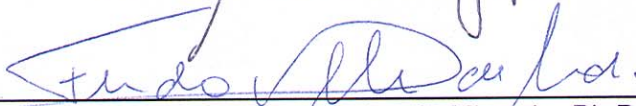
Ricardo de Souza Alencar

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM
ENGENHARIA CIVIL.

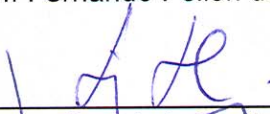
Aprovada por:



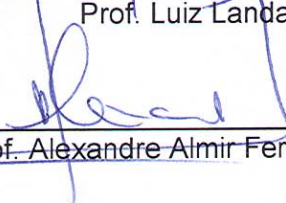
Prof. Alexandre Gonçalves Evsukoff, Ph.D.



Prof. Fernando Pellon de Miranda, Ph.D.



Prof. Luiz Landau, D.Sc.



Prof. Alexandre Almir Ferreira Rivas, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 2009

Alencar, Ricardo de Souza

Aplicação de Técnicas de Data Mining e Ferramentas OLAP na Base de Dados Integrada PIATAM/ Ricardo de Souza Alencar. – Rio de Janeiro: UFRJ/COPPE, 2009.

XIII, 129 p.: il.; 29,7 cm.

Orientador: Alexandre Gonçalves Evsukoff

Fernando Pellon de Miranda

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Civil, 2009.

Referencias Bibliográficas: p. 101-108.

1. Data Warehouse. 2. Olap. 3. Regras de Associação. I. Evsukoff, Alexandre Gonçalves *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

Dedico este trabalho aos meus pais, Alencar e Sheila, que me deram todo o suporte necessário ao meu desenvolvimento pessoal e acadêmico.

AGRADECIMENTOS

Aos meus pais, um muito obrigado seria pouco pelo apoio dado a mim, pelos ensinamentos, orientações dos caminhos certos a serem seguidos em momentos difíceis, sempre me apoiando em todas as decisões.

Agradeço ao meu orientador, Prof. Alexandre Evsukoff, pela tranquilidade passada durante a elaboração do trabalho, além de estar sempre disponível para trocar idéias e dar sugestões.

Ao Prof. Fernando Pellon, meu co-orientador, que sempre demonstrava interesse em saber como o trabalho estava ficando e quando me via perguntava “como anda o trabalho?”

Ao Prof. Luiz Landau por participar desta banca e por ter me dado a oportunidade de trabalhar no LAMCE, sempre me tratando com profissionalismo e de igual para igual. O sentimento de gratidão será eterno.

Ao Prof. Alexandre Almir Ferreira Rivas por ter aceitado se deslocar de Manaus até o Rio de Janeiro para participar de minha banca.

À família LAMCE (Laboratório de Métodos Computacionais em Engenharia – COPPE/UFRJ), onde trabalho e que me abriu as portas para o Mestrado, em especial, aos grandes amigos de trabalho e de vida, Fábio Andrezo, Humberto Buffoni, Victor Santiago, Vítor Santos, Rodrigo Freitas, Diego Mano, Fernando Taboada, Patrick Muller, Thiago Esteves, Priscila Tobias e Leandro Ouriques.

Um agradecimento em especial ao Luis Fernando Nunes Melo, por sempre ter me incentivado a ingressar no mestrado e por acreditar no meu potencial. Também serei eternamente grato a você.

Agradeço também ao meu irmão Rodrigo, minha prima Verônica, a sempre prestativa Marcília e, em especial, à minha prima Mônica pelo suporte em todos os sentidos e também pela amizade.

Ao Fábio Roque que, apesar do pouco convívio, sempre demonstrou muita seriedade e interesse em me ajudar no processo de aquisição de dados para utilização nesta pesquisa.

Agradeço, ainda, às muitas pessoas que não têm seus nomes citados, mas que contribuíram para mais uma realização em minha vida.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

APLICAÇÃO DE TÉCNICAS DE DATA MINING E FERRAMENTAS OLAP NA BASE
DE DADOS INTEGRADA PIATAM

Ricardo de Souza Alencar

Junho / 2009

Orientadores: Alexandre Gonçalves Evsukoff
Fernando Pellon de Miranda

Programa: Engenharia Civil

O Projeto PIATAM (Potenciais Impactos Ambientais no Transporte Fluvial de Gás Natural e Petróleo na Amazônia), idealizado para minimizar possíveis efeitos relacionados com as atividades da Indústria do Petróleo na área de transporte de óleo e gás no trecho do Rio Solimões entre Urucu-Coari-Manaus, dispõe de dados de natureza ambiental e sócio-econômica, coletados periodicamente em campo por pesquisadores de diferentes áreas.

O presente estudo objetiva a construção de um repositório de dados (Data Warehouse) para o tema “Ictiofauna” da Base de Dados Integrada PIATAM e, a partir desta construção, a aplicação de ferramentas OLAP (*On-Line Analytical Processing*) visando uma maior interação dos gerentes e participantes do projeto com os dados existentes através de agregações e sumariações. Além disto, foram geradas regras de associação objetivando a descoberta de conhecimento interessante na base de dados estudada.

Os resultados mostraram que a construção de um repositório de dados aliado a utilização de ferramentas OLAP permitiu que fossem feitas análises globais das informações visando a tomada de decisão, além de auxiliar no processo de descoberta de conhecimento.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

APPLICATION OF DATA MINING TECHNIQUES AND OLAP TOOLS IN THE PIATAM
INTEGRATED DATABASE

Ricardo de Souza Alencar

June / 2009

Advisors: Alexandre Gonçalves Evsukoff
Fernando Pellon de Miranda

Department: Civil Engineering

The PIATAM Project (Potential Environmental Impacts in the Fluvial Transport of Natural Gas and Petroleum in Amazonia), idealized to minimize possible effects related with the activities of the Petroleum Industry in the area of oil and gas transportation in Solimões River route in the stretch Coari-Manaus, has data of environmental and socioeconomic nature, collected periodically in field by researchers of different areas.

The present study aims at the construction of a Data Warehouse for the theme "Ictiofauna" of the PIATAM integrated database and, starting from this construction, the application of OLAP (*On-line Analytical Processing*) tools seeking a larger interaction of the managers and participants of the project with the existent data through aggregations and summarizations. In addition, association rules were generated aiming at the discovery of interesting knowledge in the studied database.

The results showed that the construction of a Data Warehouse coupled with the use of OLAP tools allowed the performance of global analyses of the information with the purpose of decision making, besides aiding in the process of knowledge discovery.

ÍNDICE

RESUMO.....	vii
ABSTRACT.....	viii
ÍNDICE DE FIGURAS.....	xi
ÍNDICE DE TABELAS.....	xiii
CAPÍTULO 01 – INTRODUÇÃO.....	01
1.1 – Apresentação.....	01
1.2 – Objetivo do Trabalho.....	03
1.3 – Organização do Texto.....	03
CAPÍTULO 02 – BDI PIATAM – BASE DE DADOS INTEGRADA PIATAM...05	
2.1 – O Projeto PIATAM.....	05
2.1.1 PIATAM I.....	11
2.1.2 PIATAM II.....	11
2.1.3 PIATAM III.....	12
2.1.4 PIATAM IV.....	13
2.2 – As Áreas de Pesquisa do PIATAM.....	14
2.3 – BDI-PIATAM – Base de Dados Integrada PIATAM.....	20
CAPÍTULO 03 – DATA WAREHOUSE E OLAP.....24	
3.1 – Introdução.....	24
3.2 – Data Warehouse.....	30
3.2.1 – As Principais Ferramentas Utilizadas em um Data Warehouse..31	
3.2.2 – Arquitetura de um Data Warehouse.....	33
3.2.3 – Integração dos Dados.....	35
3.2.4 – Modelagem de Dados.....	38
3.2.5 – Modelagem Multidimensional.....	39
3.3 – Olap (On-Line Analytical Processing).....	43
3.3.1 – Conceitos Básicos sobre Olap.....	43
3.3.2 – Tipos de Sistemas Olap.....	45
3.3.3 – Funções.....	46
3.3.4 Integração de OLAP com SIG.....	48
CAPÍTULO 04 – REGRAS DE ASSOCIAÇÃO.....50	

4.1 – Introdução.....	50
4.2 – Regras de Associação.....	53
4.2.1 – Decomposição da Geração de Regras.....	55
4.2.2 – Algoritmo Apriori.....	56
4.2.3 – Geração de Regras.....	58
4.2.4 – Regras Interessantes.....	59
CAPÍTULO 05 – CONSTRUÇÃO DO DATA WAREHOUSE DE ICTIOFAUNA E GERAÇÃO DAS REGRAS DE ASSOCIAÇÃO.....	62
5.1 – Construção do Data Warehouse de Ictiofauna.....	62
5.1.1 – Construção do Banco de Dados Relacional.....	63
5.1.2 – Construção do Data Warehouse.....	68
5.1.2.1 – Fato Pesca.....	69
5.1.2.2 – Fato Peixe.....	71
5.1.2.3 – Fato Amostra Larva.....	74
5.1.3 Implementação do Data Warehouse.....	76
5.1.4 Análise dos Resultados Obtidos com a Construção do Data Warehouse.....	78
5.2 – Regras de Associação.....	83
5.2.1 Análise dos Resultados Obtidos com as Regras de Associação.....	90
CAPÍTULO 06 – CONSIDERAÇÕES FINAIS.....	98
6.1 Conclusão.....	98
6.2 Trabalhos Futuros.....	99
REFERÊNCIAS BIBLIOGRÁFICAS.....	101
ANEXO 01.....	109
ANEXO 02.....	116
ANEXO 03.....	121
ANEXO 04.....	126

ÍNDICE DE FIGURAS

Figura 1.1 – Área de Estudo do Projeto PIATAM. Fonte: Adaptada de (MOREIRA, 2005).....	06
Figura 1.2 – BDI PIATAM. Fonte: Website PIATAM (http://www.piatam.ufam.edu.br/).....	21
Figura 3.1 – Orientação dos dados no ambiente operacional e no DW. Fonte: Adaptada de (MACHADO, 2004).....	29
Figura 3.2 – Arquitetura do projeto de DW. Fonte: Adaptada de (KIMBALL, 1998).....	33
Figura 3.3 – Integração dos dados no DW. Fonte: Adaptada de (Machado, 2004).....	36
Figura 3.4 – Variação do tempo nos dados operacionais e no DW. Fonte: Adaptada de (Machado, 2004).....	37
Figura 3.5 – Volatilidades dos dados nos sistemas operacionais e no Data Warehouse. Fonte: Adaptada de (Machado, 2004).....	38
Figura 3.6 – Diagrama de um modelo Entidade-Relacionamento. Fonte: (MARQUES et al., 2000).....	39
Figura 3.7 – Exemplo de esquema estrela sobre vendas. Fonte: Adaptado de (MARQUES et al., 2000).....	40
Figura 3.8 – Um exemplo de cubo de dados sobre vendas. Fonte: (MARQUES et al., 2000).....	42
Figura 3.9 – Modelo de Banco de Dados Multidimensional – Tridimensional. Fonte: (BISPO, 1998).....	46
Figura 3.10 – Operação de Pivot. Fonte: (BISPO, 1998).....	47
Figura 3.11 – Operação de Slice. Fonte: (BISPO, 1998).....	47
Figura 3.12 – Operação de Drill-down/up. Fonte: (BISPO, 1998).....	48
Figura 4.1 – O processo de Descoberta de Conhecimento em Bases de Dados Fonte: Adaptada de Business Intelligence, KDD and Data Mining News (http://www.kmining.com/).....	52
Figura 4.2 – Tabela T. Fonte: Adaptada de (AGRAWAL et al, 1993).....	54
Figura 5.1 – Fluxograma de Procedimentos para a Construção do Data Warehouse para o Tema Ictiofauna.....	63

Figura 5.2 – Modelagem ER do Banco de Dados Relacional. Fonte: Adaptada do Relatório de Manutenção do Modelo de Dados da Base de Dados Integrada do PIATAM (versão ATECH.439.03.0002/A).....	64
Figura 5.3 – Exemplo de Script de Criação de Tabelas no SGBD.....	65
Figura 5.4 – Estrutura dos Flat Files.....	66
Figura 5.5 – Exemplo de Script de Inclusão de Dados no SGBD.....	66
Figura 5.6 – Banco de Dados Povoado.....	67
Figura 5.7 – Modelo Multidimensional do Fato Pesca.....	71
Figura 5.8 – Modelo Multidimensional do Fato Peixe.....	73
Figura 5.9 – Modelo Multidimensional do Fato Amostra Larva.....	76
Figura 5.10 – Ambiente de Desenvolvimento do SQL Server Business Intelligence Development Studio.....	77
Figura 5.11 – Resposta da Consulta “Qual a quantidade de coletas realizadas por Local de Coleta nos últimos 03 anos?”.....	78
Figura 5.12 – Resposta da Consulta “Qual a quantidade de coletas realizadas em cada mês de 2007?”.....	79
Figura 5.13 – Resposta da Consulta “Qual a quantidade de pescas realizadas por Local de Pesca nos últimos 03 anos?”.....	80
Figura 5.14 – Resposta da Consulta sobre as medidas dos peixes (Nome Científico e Nome Vulgar) coletados nas pescas realizadas na localidade do Paraná do Iauara em 2007.....	81
Figura 5.15 – Resposta da Consulta “Qual a quantidade de larvas coletadas por localidade no ano de 2007 ?”.....	82
Figura 5.16 – Dados de Variável da Água.....	83
Figura 5.17 – Dados de Coleta de Peixe.....	84
Figura 5.18 – Histograma de Espécies de Peixe.....	85

ÍNDICE DE TABELAS

TABELA 3.1 – Diferenças entre OLTP e OLAP.....	27
TABELA 3.2 – aracterísticas de um Data Warehouse.....	28
TABELA 5.1 – Tabelas e Registros do Banco de Dados.....	68
TABELA 5.2 – Legenda do Histograma de Espécies de Peixe.....	85
TABELA 5.3 – Intervalo de Dados do Atributo Peso.....	86
TABELA 5.4 – Intervalo de Dados do Atributo Comprimento.....	87
TABELA 5.5 – Intervalo de Dados do Atributo PH.....	87
TABELA 5.6 – Intervalo de Dados do Atributo Oxigênio de Superfície.....	88
TABELA 5.7 – Intervalo de Dados do Atributo Oxigênio de Profundidade.....	88
TABELA 5.8 – Intervalo de Dados do Atributo Temperatura de Superfície (oC).....	89
TABELA 5.9 – Intervalo de Dados do Atributo Temperatura de Profundidade (oC).....	89
TABELA 5.10 – Regras de Associação Referentes ao Lago Ananá.....	91
TABELA 5.11 – Regras de Associação Referentes ao Lago Araçá.....	92
TABELA 5.12 – Regras de Associação Referentes ao Lago Aruã.....	93
TABELA 5.13 – Regras de Associação Referentes ao Lago Baixo.....	93
TABELA 5.14 – Regras de Associação Referentes ao Lago Iauara.....	94
TABELA 5.15 – Regras de Associação Referentes ao Lago Maracá.....	94
TABELA 5.16 – Regras de Associação Referentes ao Lago Poraquê.....	95
TABELA 5.17 – Regras de Associação Referentes ao Lago Preto.....	96

Capítulo 01

INTRODUÇÃO

Neste primeiro capítulo, serão apresentados os assuntos a serem tratados nesta dissertação. Serão mostrados também, o objetivo do trabalho e a sua organização.

1.1 Apresentação

O projeto PIATAM (Potenciais Impactos e Riscos Ambientais da Indústria do Petróleo e Gás no Amazonas), envolve uma rede de instituições de pesquisa e ensino, a maioria, da Amazônia, com o interesse comum de gerar informações no campo das ciências naturais, sociais e tecnológicas.

Considerando a necessidade de aumentar a abrangência das informações disponíveis para o projeto, bem como reunir uma quantidade maior de dados relacionados aos impactos ambientais e sociais decorrentes das atividades da Petrobrás na planície aluvial do Rio Solimões, foi idealizada e implementada uma base de dados geo-espacial heterogênea que inclui dados de natureza tabular, vetorial e pictorial. Esta base integra dados oriundos de diversas linhas de pesquisa do PIATAM e é fruto de discussões entre os vários participantes do projeto (TOCANTINS *et al*, 2006).

Na análise de grande volume de dados, normalmente são utilizadas técnicas que compreendem a realização de consultas SQL (*Structured Query Language*) ou mecanismos de visualização de dados. Porém, tais técnicas não possibilitam que sejam expressas algumas questões importantes para tomada de decisão, além de, muitas vezes, serem dispendiosas (em termos financeiros e de tempo), subjetivas e inviáveis, quando aplicadas a grandes bases de dados.

Diante disto, diversas pesquisas têm sido direcionadas ao desenvolvimento de tecnologias de extração automática de conhecimento a partir de dados. Esse campo de pesquisa é chamado de Descoberta de Conhecimento em Bases de Dados, também conhecido como KDD (*Knowledge Discovery in Databases*). Esse processo, de natureza iterativa e interativa, é bastante utilizado em diversos domínios, na extração de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados (TANIAR, 2007; CHEN & LIU, 2005; APTE *et al*, 2002; LEE & SIAU, 2001; FAYYAD *et al*, 1996).

Uma das técnicas de mineração de dados que vêm se destacando é a Regras de Associação. Em linhas gerais, uma regra de associação caracteriza o quanto a presença de um conjunto de itens nos registros de uma base de dados implica na presença de algum outro conjunto distinto de itens nos mesmos registros (AGRAWAL & SRIKANT, 1994). Além da utilização dessa técnica nas pesquisas acadêmicas, os resultados das mesmas têm sido utilizados no comércio, em contratos de seguro, na saúde, no geoprocessamento, na biologia molecular, em bases de dados de exsudações de óleo, em sistemas militares, entre outras áreas (DOMINGUES, JORGE, & SOARES, 2006; METWALLY *et al*, 2005; AGGELIS, 2004; FONSECA *et al*, 2003; SEMENOVA *et al*, 2001; SILVA, 2008, ALVES, 2007).

A definição clássica de data warehouse é: “uma coleção de dados orientada por assuntos, integrada, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão” (INMON, 1992). Estes repositórios de dados históricos possuem uma tabela fato e várias tabelas dimensão para armazenar dados de um determinado assunto. Assim, o data warehouse forma uma base de dados que permite efetuar um tratamento adequado da informação, o qual pode habilitar a descoberta e exploração de tendências empresariais importantes.

Uma tendência no processo de Descoberta de Conhecimento em Bases de Dados é o uso conjunto das tecnologias de mineração de dados (*data mining*) e *data warehouse*, de maneira que o processo de mineração seja beneficiado pelo pré-processamento que os dados passaram ao popular o data warehouse, além da facilidade para a extração dos dados a serem utilizados no processo de descoberta de conhecimento.

1.2 Objetivo do Trabalho

O objetivo do trabalho é o de, através da construção de um Data Warehouse para o tema “Ictiofauna” da Base de Dados Integrada PIATAM, sejam utilizadas ferramentas OLAP para que analistas, gerentes e executivos sintetizem informações através de comparações, visões personalizadas, análise histórica e projeção dos dados em vários cenários.

A busca de associação entre variáveis é, freqüentemente, um dos propósitos das pesquisas empíricas. A possível existência de relação entre variáveis podem orientar análises, conclusões e auxiliar possíveis tomadas de decisão. Portanto, visando a descoberta de conhecimento interessante na base de dados, serão executados algoritmos de regras de associação.

1.3 Organização do Texto

Esta dissertação está dividida em 07 capítulos. Nesse primeiro capítulo foi apresentado o contexto em que se insere este trabalho, bem como seus objetivos.

No Capítulo 2, é descrito o Projeto PIATAM (Potenciais Impactos e Riscos Ambientais da Indústria do Petróleo e Gás no Amazonas), além da descrição do BDI PIATAM (Base de Dados Integrada PIATAM) visando a contextualização do trabalho.

No Capítulo 3, são abordados os conceitos de *Data Warehouse*, de sua arquitetura aos conceitos de DataMart. Também serão expostos os conceitos de OLAP (*On-line Analytical Processing*).

No Capítulo 4, será descrito o processo de mineração de dados, sendo apresentados suas etapas e os usuários que participam desse processo. A associação, uma das tarefas da mineração de dados, também é descrita nesse capítulo, visto que o foco deste trabalho está nas regras de associação.

No Capítulo 5, será descrita a construção do *Data Warehouse* para o tema “Ictiofauna”. Em seguida, será feita a análise dos resultados da construção do Data

Warehouse através das consultas OLAP. Também serão geradas as regras de associação visando a descoberta de conhecimento interessante na base de dados. Por fim, as regras geradas serão analisadas.

No Capítulo 06, são apresentadas as conclusões deste trabalho e propostas para trabalhos futuros.

Após as referências bibliográficas, serão apresentados o Dicionário de Dados do Banco de Dados utilizado no presente estudo, Scripts de Criação de Tabelas e Inclusão de Dados no Banco de Dados, além dos Histogramas dos Atributos Utilizados na Geração das Regras de Associação.

Capítulo 02

BDI PIATAM – BASE DE DADOS INTEGRADA PIATAM

Este capítulo apresentará o projeto PIATAM (Potenciais Impactos e Riscos Ambientais da Indústria do Petróleo e Gás no Amazonas), bem como a BDI PIATAM (Base de Dados Integrada PIATAM) que consiste num repositório centralizado que combina dados de interesse do projeto PIATAM.

2.1 O Projeto PIATAM

O Projeto “Monitoramento das Áreas de Atuação da Petrobras: Potenciais Impactos e Riscos Ambientais da Indústria do Petróleo e Gás no Amazonas” (PIATAM) abrange as rotas fluviais de transporte de petróleo, gás e derivados da Petrobras e suas contratadas no Estado do Amazonas, nos trechos Manaus (Terminal do Solimões) – Lago Coari – Coari – Urucu. A metodologia proposta no Projeto prevê a construção de Sistemas de Informações Ambientais Georreferenciadas, associados aos levantamentos de campo e aplicados ao planejamento, à prevenção e ao controle das atividades da indústria de petróleo, incluindo a modelagem das variações sazonais do nível das águas das bacias hidrográficas envolvidas.

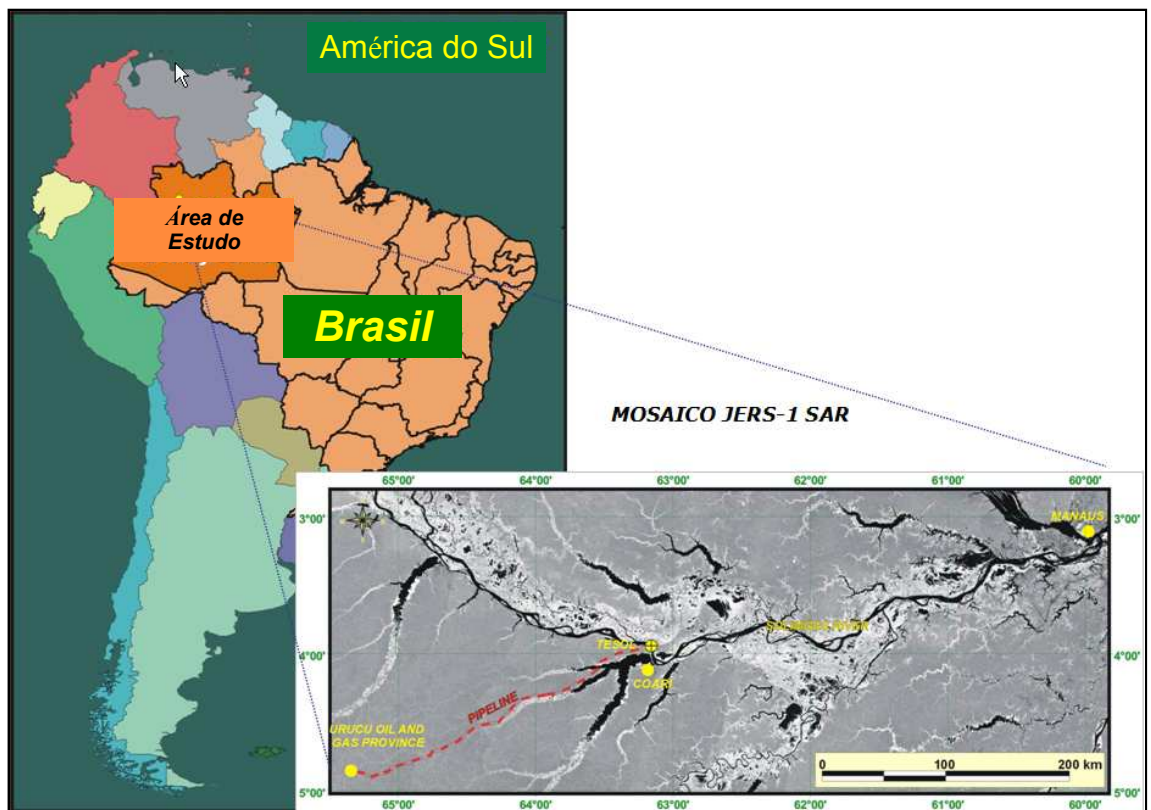


Figura 1.1 – Área de Estudo do Projeto PIATAM. Fonte: Adaptada de (MOREIRA, 2005).

Durante a excursão, nove comunidades são visitadas pelo barco de pesquisa PIATAM. São elas:

- Bom Jesus - A Comunidade Bom Jesus do Paraná do Iauara, está localizada à margem direita do rio Solimões, na costa interna da ilha Iauara, possuindo coordenadas geográficas de latitude $03^{\circ} 36' 39'' S$ e longitude de $6^{\circ} 17' 24'' W$. O relevo é de várzea baixa com freqüentes alagações anual, composto de seis lagos de água branca. A população é composta de 29 famílias com o total de 169 pessoas. A produção se divide em comercial: com o cultivo da malva e juta; de subsistência: com o cultivo de banana, feijão, macaxeira, milho e pescado (na entressafra da malva, surgindo como alternativa econômica); e extrativismo vegetal: com o taxí e a jabuticaba. A Organização social se dá através de Comunidade de base, por meio da religião católica, que

ocorre doze vezes por mês, embora existam pessoas da Assembléia de Deus (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- Esperança II - Esta comunidade localiza-se à margem direita do Rio Solimões, primeira comunidade abaixo do Terminal Solimões – TESOL, município de Coari, possuindo coordenadas geográficas de latitude 030 57'59"S e longitude 630 08'52"W. Tem uma população composta por 46 pessoas divididas em 10 famílias. O relevo é constituído de várzea baixa, com frequentes alagações anuais (onde as casas dos moradores estão localizadas) e, de terra firme (destinada às áreas de plantio). Os produtos mais importantes para a renda da maioria das famílias são: castanha, açaí, mandioca e a pesca. Porém, para subsistência, essas famílias utilizam também a criação animal (galinha e boi) e a caça (queixada e paca). A pesca no período de seca é praticada para comercialização, sendo o peixe o principal produto da dieta alimentar. A organização social da comunidade é ativa, pois existe a Associação de Moradores, Associação dos Agricultores, Sindicato dos Trabalhadores Rurais e Colônia de Pescadores. As famílias costumam se reunir para o culto religioso, semanalmente, e as reuniões da comunidade são semestrais (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- Lauro Sodré - A comunidade Lauro Sodré foi fundada no ano de 1984, localiza-se a margem esquerda do rio Solimões, na confluência do Paraná do Nazária, possuindo coordenadas geográficas de latitude 030 51'33"S e longitude 620 35'08"W, no município de Coari. o relevo é de várzea alta, com raras alagações anuais. A população é composta de 110 famílias sendo que a maioria é de produtores, dando um total de 470 pessoas. A produção é diversificada tendo a agricultura e a pesca como principal atividade. Os produtos cultivados são: banana, cheiro-verde, cebola, manga, abacate, cacau, mandioca, macaxeira e milho. Os produtos são comercializados na feira do produtor de Coari e na comunidade. A Organização Social se dá através da associação de produtores que contam com a participação de 40 famílias tendo duas reuniões por mês e Associação Comunitária Lauro Sodré (ACOLASO), fundada em 16 de agosto de 2003 com uma reunião uma vez ao mês, bem como cultos religiosos que são os mais frequentes realizados

quatro vezes ao mês (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- Matrinchã - Localiza-se à margem esquerda do Rio Solimões, na costa da Ilha Matrinchã. Possuindo coordenadas geográficas de latitude 030 46'44"S e longitude de 620 21'54"W, no município de Codajás. A comunidade apresenta relevo de várzea baixa com freqüentes alagações anuais, possuindo quatro lagos de várzea. Sua população é de 31 pessoas distribuídas em 6 famílias, do qual retiram sua subsistência através da plantação de maxixe, cebola, tomate, pimenta, feijão, jerimum, macaxeira e milho; com a criação de galinha, pato; com a caça de mutum e guariba; e com a pesca de algumas espécies como: ruelo, pacu, curimatá e tucunaré. Para fins comerciais, os principais produtos são provenientes da Agricultura e da Pesca dos produtos citados acima. A comunidade organiza-se por meio de comunidade de base. As atividades sociais mais praticadas são o culto religioso que se repete seis vezes ao ano e a reunião da escola que ocorre uma vez ao mês (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- Nossa Senhora das Graças - A Comunidade Nossa Senhora das Graças da Costa do Pesqueiro II, localizada às margens direita do rio Solimões, em frente ao município de Manacapuru, coordenadas geográficas de latitude 030 20'37"S e longitude 600 35'34"W, possui uma população de 312 pessoas, distribuídas em 66 famílias, todas estas residentes em relevos de várzea baixa, onde está situada a comunidade. As famílias moradoras da referida localidade estão organizadas em comunidade de base, onde a igreja promove eventos religiosos e sociais. A religião predominante da comunidade é o catolicismo, havendo também famílias pertencentes a outras religiões. A Comunidade apresenta duas Associações, sendo uma Associação de Pais e Mestres e outra de Produtores. As reuniões da comunidade e reuniões da associação são realizadas uma vez ao mês, enquanto que os cultos religiosos são realizados com uma frequência maior, chegando a oito vezes ao mês. Os principais produtos para a obtenção de renda são: o pescado; a mandioca (para a produção de farinha); a malva e o milho. Grande parte da população alimenta-se do pescado sendo esta atividade praticada com certas restrições, respeitando assim

o período para pesca comercial e os locais onde esta prática pode ser realizada, assim como os métodos realizados, como o arrasto de fundo que é uma técnica permitida ao passo que a pesca por arrastão é extremamente proibida, respeitando desta forma a principal atividade (o pescado) da referida comunidade (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- Nossa Senhora de Nazaré - A Comunidade Nossa Senhora de Nazaré da Costa do Paratari I, está localizada à margem esquerda do rio Solimões, possuindo coordenadas geográficas de latitude 030 35'04"S e longitude 600 56'03"W, no município de Manacapuru. O relevo é caracterizado por várzea alta com alagações pouco freqüentes. A população existente nesta comunidade é de 43 famílias, o equivalente a 160 pessoas no total. A comunidade possui uma produção bastante diversificada com agricultura e pesca, onde os produtos mais comercializados pela maioria das famílias desta comunidade são: a Pesca (com o surubim, o tucunaré e o caparari), a Agricultura (com o mamão, o maracujá, a banana, a mandioca, a pimenta de cheiro, a chicória e a manga). A organização social das famílias se subdivide em reuniões através de cultos religiosos que acontecem oito vezes ao mês nas igrejas existentes na comunidade e reuniões comunitárias da Associação de Desenvolvimento Agrícola (fundada em 28 de Agosto de 2000) realizada uma vez ao mês (Website PIATAM - <http://www.piatam.ufam.edu.br/>).
- Santo Antônio - A Comunidade Santo Antônio da Costa da Terra Nova localiza-se à margem esquerda do rio Solimões, no município de Anori, possuindo latitude 030 50'41"S e longitude 610 39'47"W. O relevo é de Várzea baixa com alagações e de subsistência, na agricultura os plantios de ciclo curto e médio se direcionam aos cultivos de mandioca, maxeira, milho, melancia e hortaliças em geral para o abastecimento das famílias, tendo também o pescado como um produto de primeira necessidade durante todo o ano. Além dos produtos supracitados, os mesmos exercem outras atividades extrativistas como: a caça e as retiradas de madeiras (Jacareúba e Muiratinga). Estes produtos são em parte utilizados para a subsistência e para a comercialização sendo esta última geradora de renda para a maioria das famílias que

comercializam seus produtos no município de Anamá, Anori e Manaus. A Organização social se dá através dos cultos religiosos com frequência de quatro vezes ao mês (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- Santa Luzia da Costa do Buiuçuzinho - A comunidade de Santa Luzia da Costa do Buiuçuzinho, fundada em 1990, localiza-se à margem direita do lago de Coari, próximo a foz do Rio Urucu, com latitude de 040 11'60"S e longitude 630 42'33"W. O relevo é de terra firme em sua totalidade. Possui uma população de 14 famílias e um total de 91 pessoas residentes na comunidade. A sua produção destina-se à subsistência e ao comércio, onde os principais produtos utilizados no local são: mandioca, macaxeira, milho, cará, banana, cana-de-açúcar, castanha, açaí e bacaba. A pesca é incipiente, praticada apenas para subsistência, os principais produtos comercializados são: castanha, açaí, farinha e o pato. A comunidade Santa Luzia se organiza através de culto religioso, que acontece quatro vezes ao mês, bem como reuniões da comunidade, realizada duas vezes ao mês (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- Santa Luzia do Baixio - Localiza-se à margem direita do paraná da Ilha do Baixio, margem esquerda do Rio Solimões, município de Iranduba, possuindo coordenadas geográficas de latitude 030 17'18"S e longitude 600 04'42"W. O relevo é caracterizado por várzea baixa, com alagações em área de plantio, durante alguns meses do ano. A população é composta de 106 famílias, o equivalente a 401 pessoas no total. A comunidade possui uma produção dividida em: subsistência e comercial. A produção comercial baseia-se no plantio de hortaliças, cujo ciclo é curto e o produto é facilmente comercializado na capital. A produção de subsistência é centralizada nos cultivos agrícolas mais comuns, como: cebolinha, pepino, feijão-de-metro, tomate, jerimum, malva e melancia. A organização social se dá através da Associação Comunitária de Santa Luzia, que está organizada em Associação de Produtores, Associação de Pais e Mestres e SAT – Sistema de Aprendizagem Tutorial, com participação de 26 famílias. Os comunitários se reúnem em cultos religiosos quatro vezes ao mês, tanto famílias da religião católica, quanto famílias da Assembléia de Deus Tradicional. As

reuniões comunitárias são realizadas uma vez por mês, para tratar de assunto relacionados à comercialização e compra de produtos, as reuniões das Associações ocorrem uma vez ao ano (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

2.1.1 PIATAM I

O Projeto iniciou no mês de janeiro de 2001, com o PIATAM I – Potenciais Impactos Ambientais do Transporte Fluvial de Gás Natural e Petróleo na Amazônia: Análises de Risco, Planos de Contingência e Estratégias de Recuperação de áreas Impactadas, tendo como executora a Universidade Federal do Amazonas – UFAM, por meio do Centro de Ciências do Ambiente – CCA e como financiadora a FINEP (Financiadora de Estudos e Projetos do Ministério da Ciência e Tecnologia), com recursos do Fundo Nacional de Desenvolvimento Científico e Tecnológico (FNDCT) – CTPETRO. (MANUAL DE GESTÃO PIATAM, 2006)

O Projeto apresentou como objetivo identificar os principais pontos de risco no transporte fluvial de gás natural e petróleo, bem como identificar e avaliar os principais impactos nas áreas potencialmente afetadas, como forma de subsidiar a elaboração de planos de contingência e de recuperação de áreas impactadas. (MANUAL DE GESTÃO PIATAM, 2006)

2.1.2 PIATAM II

O PIATAM II, financiado pela Petrobras e pela FINEP, iniciou suas atividades em janeiro de 2002, e se constituiu no segundo ciclo do PIATAM. Esse projeto apresentou como objetivo principal, estruturar e disponibilizar informações sobre a produção, transporte e refino de gás e petróleo no Estado do Amazonas, através do monitoramento e da avaliação de riscos ambientais dessas atividades, visando à redução de danos ao ambiente e às populações humanas provocados por eventuais vazamentos. (MANUAL DE GESTÃO PIATAM, 2006)

Concluído em junho de 2004, o PIATAM II foi o resultado do esforço interdisciplinar e interinstitucional de pesquisadores, técnicos e bolsistas da Universidade Federal do Amazonas – UFAM, do Instituto Nacional de Pesquisas da Amazônia – INPA, do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro – COPPE/UFRJ e do Instituto de Tecnologia da Amazônia – UTAM, instituições interessadas na aplicação de modernas técnicas de monitoramento das atividades da Petrobras na Amazônia, com base nos princípios de gestão, integrando conceitos de meio ambiente, de qualidade, de saúde e de segurança operacional em projetos de pesquisa. (MANUAL DE GESTÃO PIATAM, 2006)

2.1.3 PIATAM III

O Projeto PIATAM III, sequência do PIATAM I e PIATAM II, dá continuidade à implementação da construção da série histórica socioambiental, devendo subsidiar as ações de monitoramento e controle da indústria do petróleo no Estado do Amazonas, onde se destaca a Petrobrars.

É proposta do PIATAM III contribuir para o monitoramento das atividades da Petrobras, em tempo real. Nos ciclos anteriores, várias ações foram iniciadas no sentido de explorar as possibilidades tecnológicas disponíveis ou a serem disponibilizadas. Essas possibilidades incluem, entre outras, a infra-estrutura de monitoramento ambiental propiciada pelo Sistema de Proteção da Amazônia – SIPAM e as iniciativas de mídia desenvolvidas em outros projetos. (MANUAL DE GESTÃO PIATAM, 2006)

O Projeto tem sido um bom exemplo de execução de iniciativas de pesquisa na Amazônia, pois a experiência adquirida, ao longo de seus ciclos, possibilita aperfeiçoamentos constantes em prol de uma maior efetividade nas suas ações. Para tanto, utiliza-se uma forma padrão de gestão, por meio do emprego das diversas ferramentas e monitoramento dos processos e indicadores definidos para o Projeto.

As metas do PIATAM III são:

- Continuar a implementação da construção da série histórica socioambiental para subsidiar as ações de intervenção, monitoramento e controle da indústria;
- Aperfeiçoar e ampliar o sistema de gerenciamento de dados e informação para implementar interações com outros sistemas e sociedade;
- Implementar a utilização de sistemas de monitoramento e modelagem ambiental através de recursos tecnológicos de última geração;
- Identificar ações e oportunidades que visem à promoção da melhoria da qualidade de vida das populações humanas na área de atuação da indústria e;
- Implementar conceitos e ferramentas gerenciais necessários para suportar atividades do Projeto e contribuir para a excelência das ações da Petrobras.

Além das metas acima definidas, o PIATAM III também pretende assessorar cientificamente a Petrobras na melhoria de seu desempenho empresarial na Amazônia, capacitando-a como instituição econômica inovadora e líder em projetos de promoção do desempenho regional.

2.1.4 PIATAM IV

O quarto ciclo do Projeto PIATAM (Inteligência Socioambiental Estratégica da Indústria do Petróleo e Gás na Amazônia) envolve importantes instituições de ensino e pesquisa que atuam na região, como a Universidade Federal do Amazonas - Ufam, o Instituto Nacional de Pesquisas da Amazônia - INPA e a Fundação Centro de Análise, Pesquisa e Inovação Tecnológica - Fucapi, assim como instituições de outros estados

como o Instituto de Pós-Graduação e Engenharia da Universidade Federal do Rio de Janeiro - COPPE/UFRJ e a Universidade do Estado de São Paulo – UNESP (CASTRO *et al*, 2008).

Neste ciclo, o projeto atua mais incisivamente nas comunidades estudadas, pesquisando a questão da geração de benefícios para elas, resultando na criação de novos projetos de geração de renda para os habitantes destas comunidades. Além disto, o quarto ciclo garante a continuidade da confecção das séries históricas de dados socioambientais da região.

2.2 As Áreas de Pesquisa do PIATAM

A seguir, serão apresentadas as áreas de pesquisa estudadas. São elas:

- **Comunicação, Design e Multimídia** - Para que o conceito do programa PIATAM estabeleça um paralelo com a linha institucional adotada pelas empresas financiadoras, torna-se fundamental um planejamento de comunicação adequado às demandas e a possíveis ampliações da rede de pesquisa. Assim, o grupo de Comunicação Design e Multimídia tem somado esforços para reformular o ambiente informacional, virtual e de multimídia do programa e gerar recursos didáticos interativos. Além disso, o grupo realiza uma divulgação em grande escala do programa PIATAM através de todas as mídias disponíveis, para que ele se consolide como um investimento ecologicamente aceitável e empresarialmente responsável, deixando em evidência o conceito de responsabilidade social e ambiental da empresa financiadora e solidificando a identidade visual do PIATAM. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).
- **Doenças Tropicais** - A malária é uma doença tropical muito comum na Amazônia e o seu controle faz-se necessário como medida de saúde pública. O projeto está estruturado para estudar os anofelinos, família de mosquitos transmissores da malária, e os mecanismos de transmissão da doença nas comunidades ao longo da área de atuação do PIATAM, determinando as estratégias de controle entomológico e epidemiológico com base em medidas de rotina e em novos processos

biotecnológicos e reduzindo os riscos de transmissão dessa doença. O grupo de Doenças Tropicais também complementa as ações de controle com informações em Educação e Saúde e assistenciais, que possibilitem a melhoria das condições e qualidade de vida das comunidades. Atualmente, está trabalhando na concepção de um modelo estocástico que estabeleça associação entre a chuva e a malária, buscando identificar similaridades espaciais e temporais na incidência da doença. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Modelagem** - A previsão e gerenciamento dos impactos socioambientais de eventuais derramamentos de petróleo no trecho da várzea entre Coari e Manaus requer o desenvolvimento de modelos capazes de prever a trajetória de manchas de petróleo e a distribuição e sensibilidade ambiental dos habitats aquáticos potencialmente atingíveis por estas manchas. Para serem eficazes no gerenciamento de acidentes, estes modelos devem ser integrados num ambiente interativo (sala de situação), onde o gerente pode obter informações sobre a trajetória de uma mancha e seus possíveis impactos em tempo real. Assim, o objetivo central do componente de modelagem é de desenvolver modelos espaço-temporais capazes de prever: 1) a trajetória de manchas de petróleo no canal e na várzea do rio Solimões/Amazonas e 2) a distribuição e sensibilidade ambiental dos habitats aquáticos potencialmente atingíveis durante um evento de derramamento. O grupo pretende, assim, aperfeiçoar modelos já desenvolvidos e combiná-los para produzir um modelo integrado de deriva para o sistema fluvial. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Limnologia** - Esta área de estudo do PIATAM soma esforços no sentido de elaborar um diagnóstico da qualidade das águas do sistema hídrico na planície de inundação do rio Solimões entre Coari e Manaus, visando fornecer subsídios para a identificação e avaliação de possíveis impactos ambientais decorrentes de atividades antrópicas na várzea amazônica, bem como para a elaboração de um programa de monitoramento da qualidade das águas e do meio ambiente do Sistema

Solimões/Amazonas no trecho estudado. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **História do Homem** - O Grupo de História do Homem é dividido em dois sub-grupos: o de Arqueologia e o de História e Meio Ambiente.
Arqueologia - As áreas adjacentes à calha do rio Solimões eram densamente povoadas nos últimos séculos antecedentes ao início da colonização europeia. O trabalho de levantamento arqueológico na área de pesquisa do PIATAM permitirá saber mais sobre a história e a cultura desses povos, realizando atividades de pesquisa e educação em dois sítios recentemente identificados (Lauro Sodré e Marrecão) e implementando e assessorando o laboratório de arqueologia do PIATAM em Manaus, além de delimitar e elaborar mapas topográficos digitais e abrir escavações e sondagens para a verificação da profundidade dos depósitos nos sítios Lauro Sodré e Marrecão.
História e Meio Ambiente - A partir do interesse em investigar as relações entre a história e a natureza na Amazônia, nasceu o Projeto Rios e Homens. Este é vinculado ao Núcleo de Pesquisa em Política, Instituições e Práticas Sociais - POLIS, da UFAM, que estabeleceu uma parceria com o PIATAM, formando o grupo de História e Meio Ambiente. O Projeto Rios e Homens é bastante amplo, articulando quatro sub-projetos temáticos, cada um com uma sub-coordenação própria: Saúde, História e Políticas Públicas; Etnias, Identidades e Fronteiras; Imagens da Amazônia: Histórias, Memórias e Iconografia Científica; Rios, Cidades e Homens: Trajetórias Coloniais e Pós-coloniais. Estes verticalizam dimensões das relações entre história e natureza e utilizam o rio como tema e como estratégia de pesquisa. Assim, pretende-se articular pesquisas que tenham como tema a Amazônia a partir de um espaço concreto – o rio – e que possibilitem estudos comparativos, privilegiando esses eixos temáticos e incorporando problemas de pesquisa ligados à história da região. Além disso, o grupo também visa buscar a interlocução com linhas temáticas da historiografia que discutam as articulações entre a história e a natureza e fortalecer as atividades de investigação multidisciplinares, articulando-se a outros grupos de pesquisa dentro do PIATAM e construindo bases para um diálogo, de caráter transversal, a partir das

análises da área de História e Meio Ambiente. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Tecnologia da Informação** - O PIATAM transformou-se em um programa estratégico para as instituições que dele participam. Esta premissa torna-se verdadeira para a equipe de Tecnologia da Informação do programa, pois o que caracteriza o atual nível de evolução tecnológica do projeto não é a centralidade de conhecimento e informação, e sim o ciclo de realimentação cumulativo entre a inovação e seu uso, gerando informações que são utilizadas em planos de contingência e nas análises de risco da atividade de transporte petrolífero no Amazonas. O programa PIATAM vem agregado subsídios para projetos de políticas públicas e privadas na região e para projetos que envolvem tecnologias de ponta. Um dos objetivos da equipe de Tecnologia da Informação é a manutenção, a ampliação, a integração e o compartilhamento da Base de Dados Integrada PIATAM (BDI) como parte da Gestão da Informação do projeto que serve como instrumento para transferência de conhecimento. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Ictiofauna** - Considerando a possibilidade de ocorrer acidentes a partir dos processos de produção, refino e transporte de petróleo e gás, os organismos bentônicos, incluindo peixes de hábitos bentônicos e os peixes de hábito pelágico, desde a fase larval até a fase adulta, estão sujeitos aos efeitos negativos advindos dessas substâncias, direta ou indiretamente. Para monitorar o ecossistema aquático de forma a preservar suas condições naturais, o grupo de Ictiofauna, dividido em três subgrupos, caracteriza a composição e estrutura das assembléias de peixes, organismos bentônicos e ictioplancton em lagos localizados na área de influência da construção e transporte de petróleo e gás, e realiza estudos acerca dos hábitos alimentares das espécies de peixes que formam as comunidades de lagos de várzea para construir um modelo ecotrófico. O conhecimento da composição específica da ictiofauna de lagos expostos ao efeito antrópico como possíveis acidentes com petróleo subsidiará ações mitigadoras, principalmente quando se fizerem necessárias a sua proteção e recomposição. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Gestão** - Como o próprio nome diz, este grupo é responsável pelo gerenciamento do programa. O desenvolvimento de um modelo de gestão passa sempre por uma melhoria contínua e sistemática, visando à solução de todos os entraves e caminhos críticos de um programa. O aperfeiçoamento será sempre perseguido, pois se deseja contribuir para que a eficiência e eficácia sejam conseguidas de modo constante. Para tanto, o grupo de Gestão implementa conceitos e ferramentas gerenciais necessárias para suportar as atividades do programa e contribuir para a excelência das ações da Petrobras. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Solos** - A degradação do solo, seja em consequência de atividades agrícolas ou da exploração de outros recursos naturais, resulta em prejuízos significativos para todo o ecossistema. O sucesso de sua recuperação depende, entre outros fatores, do nível de conhecimento do solo e do impacto sofrido. Por isso, este grupo de pesquisa cria um conjunto de informações que permite, em casos de acidentes, identificar o nível de impacto sobre o solo e a possibilidade de reversão, bem como avaliar a capacidade do solo em resistir a determinados impactos e estabelecer classes de vulnerabilidade, elaborando um diagnóstico dos solos da área de atuação do programa PIATAM. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Flora** - A Amazônia brasileira é constituída por um mosaico de tipos vegetacionais, na sua maioria restrita a áreas específicas e com endemismos particulares, o que pode, com o corte continuado da floresta, levar à extinção de inúmeras espécies vegetais. A manutenção da biodiversidade de plantas e animais silvestres é vital à sobrevivência humana. Nesse sentido, o grupo de Flora está trabalhando no mapeamento e caracterização das tipologias vegetacionais, além de verificar o status de conservação da vegetação, identificar forófitos (árvore hospedeira) e suas epífitas (plantas que se estabelecem diretamente sobre o tronco, galhos ou ramos), utilizando-as como indicador de conservação de hábitat e produzir listagem de espécies por hábitat e sua importância econômica, coletando e identificando espécies nos diversos ambientes, principalmente aquelas com valor

comercial e medicinal. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Socioeconomia** - A questão do valor dos recursos naturais ainda é e será por muito tempo um tema controverso. Se um acidente com petróleo ocorrer na área de atuação do PIATAM, muitas perguntas irão ser feitas pela sociedade e uma delas será com certeza: qual o valor do dano causado ao ambiente, às pessoas e economia da área afetada? Neste sentido, o grupo promove a caracterização da produção, consumo e uso dos ambientes das comunidades existentes no trecho do transporte de petróleo e gás entre Coari e Manaus e difunde técnicas de processamento da produção agropecuária e pesqueira aos habitantes das comunidades. O grupo também trabalha na caracterização do ambiente sociocultural das comunidades ribeirinhas no trecho entre Coari e Manaus. Além disso, o grupo desenvolve estudos de valoração ambiental utilizando fundamentos econômicos e técnicas computacionais econométricas que poderão ajudar a dar respostas mais bem informadas para a tomada de decisão. Auxiliariamente, os indicadores econômicos construídos a partir da realidade dos ribeirinhos fornecerão informações importantes sobre o que se costuma chamar de "linha de base", ou seja, a situação sem distúrbio. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).
- **Macrófitas Aquáticas** - As macrófitas aquáticas (plantas aquáticas) são essenciais para a manutenção da produtividade dos lagos, pois participam ativamente da ciclagem de nutrientes e produção de matéria orgânica. Servem também de refúgio e abrigo para diversos organismos, desempenhando um papel fundamental para a manutenção dos estoques naturais de peixes, pois aumentam a sobrevivência e o recrutamento de juvenis. As macrófitas aquáticas são, ainda, grandes armazenadoras de nutrientes e toleram ambientes impactados, apresentando elevadas taxas de crescimento, mesmo em ambientes poluídos por diferentes tipos de efluentes, sendo assim consideradas excelentes bioindicadoras. Assim, o grupo referente a essa linha de pesquisa visa gerar informações sobre ecologia, sistemática e composição química de macrófitas aquáticas dos lagos de várzea do rio Solimões, localizados na área de influência de atuação da

Petrobras, relacionar a estrutura da assembléia de peixes que habita os bancos de macrófitas aquáticas com a biomassa e diversidade dessas plantas e determinar espécies de macrófitas como bioindicadoras da toxicidade de petróleo. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

- **Entomologia** - É sabido que algumas espécies de insetos representantes das famílias Sphingidae (Lepidoptera) e Tephritidae (Diptera) apresentam alto grau de afinidade hospedeira em relação a algumas famílias botânicas, podendo constituir-se em possíveis bioindicadores do status de conservação florística de ecossistemas tropicais. Por sua vez, algumas espécies de esfingídeos são polinizadoras específicas de determinadas espécies botânicas, representativas da flora de várzea e terra firme. Assim, a captura dessas espécies representativas pode auxiliar na caracterização do “status de conservação” de um determinado ecossistema, possibilitando gerar indicadores que subsidiarão ações de monitoramento ambiental. Nesse sentido, o grupo de entomologia avalia e correlaciona a diversidade de mariposas esfingídeas (Sphingidae; Lepidoptera) e moscas-das-frutas (Tephritidae; Diptera) com a diversidade da flora em ambiente de várzea e terra firme na região do Médio Solimões. (Website PIATAM - <http://www.piatam.ufam.edu.br/>).

2.3 BDI-PIATAM – Base de Dados Integrada PIATAM

Considerando a necessidade de aumentar a abrangência das informações disponíveis para o projeto, bem como reunir uma quantidade maior de dados relacionados aos impactos ambientais e sociais decorrentes das atividades da Petrobras na bacia do rio Solimões, foi idealizada e implementada uma base de dados geo-espacial heterogênea, que inclui dados de natureza tabular, vetorial e pictorial, associada ao desenvolvimento de uma ferramenta web que permite acesso e gestão remotos. Esta base integra dados oriundos de diversas linhas de pesquisa do PIATAM e é fruto de discussões entre os vários participantes do projeto, na elaboração do modelo de dados.



Figura 1.2 – BDI PIATAM. Fonte: Website PIATAM (<http://www.piatam.ufam.edu.br/>).

A propriedade intelectual das informações que povoam a base de dados é assegurada pelo termo de confidencialidade firmado entre pesquisadores, entidade financiadora e entidade gestora da Base de Dados Integrada – BDI PIATAM. Cada usuário da BDI possui senha individualizada e estão enquadrados em diferentes combinações de perfis de uso e acesso (TOCANTINS, 2006).

A multidisciplinaridade do projeto e a conseqüente heterogeneidade dos dados elevaram a complexidade da solução no que se referiu à concepção de um eficaz modelo de base de dados. Deste modo, optou-se por explorar os aspectos espaço-temporais que permitam a construção de séries históricas necessárias ao monitoramento ambiental da região. Este sistema é essencialmente composto de dados coletados em campo por pesquisadores de diversas áreas temáticas, imagens de satélite, de radares e mosaicos georreferenciados.

Na BDI estão disponíveis mecanismos para combinar as várias informações, assim como para consultar e recuperar o conteúdo existente.

O desenvolvimento e manutenção da BDI PIATAM é realizado de forma iterativa, acompanhando a evolução do entendimento e do conhecimento gerados pelo projeto. Cada uma das fases contempla três etapas distintas: (i) Concepção; (ii) Modelagem e Projeto; e (iii) Conversão e Inclusão de dados (TOCANTINS, 2006).

Na concepção da base de dados são analisados todos os dados a serem integrados com o objetivo de detectar eventuais falhas a serem corrigidas, identificar relacionamentos e pontos comuns existentes, além de estabelecer a área de abrangência adequada para os dados espaciais. Para garantir a interação espaço-temporal, as excursões para coletas de dados são realizadas no mesmo período, envolvendo todos os temas, nos mesmos marcos PIATAM. No total são realizadas quatro excursões anuais (cheia, vazante, enchente e seca), em nove comunidades ao longo do rio Solimões, entre Manaus e Coari (TOCANTINS, 2006).

Na etapa de concepção (i), os dados analisados são classificados quanto ao formato (convencionais ou espaciais) e quanto à origem (cartografia básica e temática, dados de sensores e/ou dados advindos da coleta dos trabalhos de campo). Na seqüência, é selecionado o escopo do modelo a ser definido. (TOCANTINS, 2006).

Na etapa de modelagem e projeto (ii) é definido o esquema de banco de dados a ser adotado, utilizando um elemento de correlação para representar os relacionamentos existentes em função do tempo e do espaço, entre os diferentes tipos de dados integrados (TOCANTINS, 2006).

O elemento de correlação espaço-temporal utilizado representa as informações dos pontos de coleta visitados durante as excursões realizadas por meio de dados como coordenadas geográficas e data/hora de coleta, organizados a partir de uma semântica relevante ao projeto, ou seja, o Marco PIATAM e Período Hidrológico.

Na modelagem e projeto da base foram considerados inicialmente os seguintes temas, coletados em campo, referentes às áreas de pesquisa do PIATAM: documentação, entomologia, flora, arqueologia, varzeometria, herpetofauna, ictiofauna, IDH-C, crustáceos, limnologia, malária, socioeconomia e solos. A cada ciclo, o escopo temático é revisto, incorporando novos temas de interesse do projeto.

A etapa de conversão e inclusão de dados (iii) adota os procedimentos de validação da modelagem, de correção de falhas detectadas em dados de fases anteriores à implementação da BDI e de restrição dos dados espaciais à abrangência territorial de interesse para o PIATAM (TOCANTINS, 2006).

A cada ciclo iterativo, envolvendo as três etapas de desenvolvimento, são disponibilizados dados para consulta e análises geográficas, sendo estes organizados por temas. As informações disponibilizadas permitem realizar trabalhos georreferenciados e, em primeira análise, efetuar validações espaciais dos pontos de coleta a partir das coordenadas de localização cadastradas (geração de mapas).

A tendência natural da BDI PIATAM é configurar-se com um eficiente Sistema de Informações Geográficas – SIG frente à diversidade de objetivos com que os dados existentes são utilizados e à necessidade de evidenciar visualmente, numa única base de dados, informações espaciais oriundas do meio físico-biótico, dados censitários, cadastrais, coordenadas de localização, imagens de satélite, entre outras.

Capítulo 03

DATA WAREHOUSE E OLAP

Neste capítulo, serão abordadas as principais características das ferramentas Data Warehouse e OLAP, presentes na nova geração de sistemas de apoio à tomada de decisão. São apresentados alguns dos conceitos que balizam estas ferramentas e como tais conceitos podem ser aplicados.

3.1 INTRODUÇÃO

As primeiras edições de SGBDs (Sistemas Gerenciadores de Banco de Dados) concentravam-se em um único banco de dados que atendia todos os propósitos dos usuários destes sistemas. No entanto, as diferentes características dos dados exigem uma melhor adaptação dos bancos de dados a essas peculiaridades. Então, foram desenvolvidos novos modelos de sistemas de banco de dados e ferramentas computacionais mais adequadas a essas áreas de conhecimento específicas, de modo a atender com maior eficiência e qualidade, gerando, por exemplo, bancos de dados específicos para sistemas geográficos, suporte a decisão e multimídia.

Processamento informacional ou analítico é o processamento que atende às necessidades dos gerentes durante as tomadas de decisões (INMON, 1997).

Comparando dois tipos de sistemas de banco de dados que se relacionam com a área de *Data Warehouse* (DW), temos os bancos de dados operacionais e os bancos de dados analíticos. Os bancos operacionais dão suporte a todas as operações de uma organização, sendo utilizado com grande frequência e registrando a situação momentânea da organização, seja por exemplo, no cadastramento de clientes à baixa de um produto no estoque. Os bancos de dados analíticos visam registrar a história da organização de modo a dar suporte em momentos de decisão, geralmente, utilizado com baixa frequência.

Essa divisão fica clara quando se avalia as diferenças entre as características físicas dos dados operacionais (específico) e analíticos (sumarizado). A tecnologia de suporte aos dados operacionais é diferente da tecnologia de análise. As comunidades que utilizam os sistemas são distintas, normalmente os dados operacionais são manipulados por operadores do sistema, enquanto que os dados analíticos são acessados por gerentes/administradores da organização. E, finalmente, as características do processamento do ambiente operacional (maior segurança, controle e integridade) e do ambiente de informação (desnormalizado, velocidade de acesso) são fundamentalmente diferentes.

Um sistema de suporte à decisão provê informação para os usuários que podem analisar uma situação e construir uma decisão. O processamento analítico examina os aspectos dos dados buscando encontrar tendências e/ou relações entre as informações (POE *et al*, 1998).

No processamento analítico, os tempos de respostas são amplamente atenuados em comparação ao processamento operacional. A tecnologia analítica não dá suporte a bloqueios e disputa de dados, transações, segurança, como na operacional, daí uma das razões da sua superior velocidade nas buscas por dados.

Diferente dos sistemas de banco de dados operacionais que sofrem constantes atualizações em seus registros (individualmente), pois regem a operação da organização, os sistemas analíticos têm uma frequência de atualização bem inferior, podendo ser por semanas ou meses. Devido à própria característica do Data Warehouse, que visa a sumarização dos dados, os registros são constantemente acessados e seus conteúdos agrupados para análise, sem praticamente ocorrer alterações dos registros individuais.

Os termos OLTP (*On-Line Transaction Processing* - Processamento de Transações em Tempo-Real) e OLAP (*On-Line Analytical Processing* - Processamento Analítico em Tempo-Real) descrevem o modo de processamento nos sistemas de bancos de dados.

Os bancos de dados operacionais que tratam as operações diárias de uma empresa, baseados em OLTP, atingem proporções de centenas de megabytes e até mesmo gigabytes. Consistência e capacidade de recuperação de dados são críticas, e a maximização do poder de processar transações é requerida para minimizar os problemas que podem ser causados pela concorrência de processos.

No caso do processamento analítico, baseado em OLAP, sistemas de suporte à decisão ajudam os analistas a identificar tendências, estabelecer comparações e a prever resultados futuros, devendo-se dar maior importância aos dados históricos, totalizados e consolidados em detrimento dos dados detalhados ou individualizados. As funções e manipulações em OLAP não podem ser executadas de imediato nos sistemas OLTP.

Os sistemas baseados em OLTP mantêm apenas um instantâneo das informações mais recentes, pois estão constantemente sendo atualizados. Esse modelo processa de milhares até milhões de transações diariamente. O usuário manipula na maioria das vezes um registro por vez, tendo que fazer a mesma tarefa inúmeras vezes.

O modelo OLAP requer um histórico de transações em longos períodos de tempo. Requer dados estáticos e completos, ajustados para evitar erros. Os sistemas baseados em OLAP processam apenas uma transação por ciclo de transferência, é o caso de quando os dados são carregados para o sistema, mas podem processar milhares de consultas diárias simultâneas. O modelo OLAP observa os mecanismos da organização, lida raramente com um único registro, gera resumos e não listagens.

Essas diferenças funcionais se traduzem em diferenças na concepção de seus modelos. Comumente, os sistemas OLTP utilizam o modelo entidade-relacionamento. O modelo entidade-relacionamento normaliza os dados de forma a minimizar a redundância e é ideal para a entrada e atualizações de dados.

Os sistemas OLAP são raramente atualizados e são utilizados, principalmente, para consultas.

A organização dos dados é uma das principais diferenças entre os modelos de dados (KIMBALL 1998). O modelo entidade-relacionamento gera uma grande complexidade no banco de dados; isso se reflete, por exemplo, nas ações de busca. Tem baixa eficiência em consultas pois dificilmente é entendido pelo usuário e não é navegado de forma útil pelo software de SGBD. KIMBALL (1998) define que modelo entidade-relacionamento não deve ser utilizado como base para sistemas de suporte a decisão.

O OLAP, por exigir grandes recursos, geralmente é implementado em uma máquina diferente da que contém o banco de dados operacional (OLTP). Além do

recurso requerido, há o motivo do OLAP integrar dados de vários sistemas OLTP e das estruturas de dados serem diferentes.

A tabela 3.1 ilustra as diversidades apresentadas pelos dois tipos de sistemas.

TABELA 3.1 - Diferenças entre OLTP e OLAP.

Dados Primitivos / Dados Operacionais - OLTP	Dados Derivados/ Dados Sistemas de Apoio a Decisão - OLAP
Baseados em Aplicações	Baseados em Assuntos
Detalhados	Resumidos ou Refinados
Exatos em Relação ao Momento do Acesso	Representam valores de momentos já decorridos ou instantâneos (histórico)
Atendem a comunidade funcional	Atendem a comunidade gerencial
Podem ser atualizados	Não são atualizados
São processados repetitivamente	Processados de forma heurística
Requisitos de processamento conhecidos com antecedência	Requisitos de processamento não são conhecidos com antecedência
A performance é fundamental	Performance atenuada
Acessados uma unidade por vez	Acessados um conjunto por vez
Voltados para transações	Voltados para análise
O controle de atualizações é atribuído a quem tem posse	O controle de atualização não é problema
Alta disponibilidade	Disponibilidade atenuada
Gerenciados em sua totalidade	Gerenciados por subconjuntos
Não contemplam a redundância	A redundância é permitida
Estrutura fixa	Estrutura flexível
Pequena quantidade de dados usados em um processo	Grande quantidade de dados usados em um processo
Atende às necessidades cotidianas	Atende às necessidades gerenciais
Alta probabilidade de acesso	Baixa ou modesta probabilidade de acesso

O Data Warehouse (DW) é uma ferramenta de sistemas analíticos e pode ser definido como:

”Um Data Warehouse é uma coleção de dados orientados a assuntos, integrados, variáveis com o tempo, não voláteis, para suporte ao processo gerencial de tomada de decisão.”(INMON, 1992)

Pode-se destacar as principais características de um DW como: a orientação a assunto, a padronização dos dados (integração), os dados não se perdem nos processos do banco de dados (volatilidade), armazenamento dos registros em uma série temporal, os dados são sumarizados (resumidos), a não necessidade de normalização e um conjunto de informações sobre os dados armazenados (metadados). Essas características podem ser observadas na tabela 3.2.

TABELA 3.2 - Características de um Data Warehouse.

Característica	Descrição
Orientação a assunto	Dados organizados de modo como o usuário se refere a ele
Integração	Inconsistências e informações conflitantes são removidas, isto é, os dados são "limpos"
Não volátil	Somente leitura de dados, dados não são atualizados por usuários
Séries históricas	Os dados estão em séries históricas, não somente na situação corrente
Resumidos	Os dados operacionais estão/são agregados
Abrangência	Mantêm uma série temporal, considerando que mais dados serão armazenados
Não normalizado	Dados podem ser redundantes
Metadados	Dados sobre dados

O DW deve ter seu planejamento focado nos assuntos mais importantes da organização, fazendo com que o projeto seja orientado a assuntos. Isso contrasta com as aplicações mais tradicionais que são orientadas a processos/funções. A figura 3.1 ilustra essa relação. No ambiente operacional, o planejamento focaliza a aplicação e suas características específicas. Enquanto que no DW a abordagem busca a uma

área de interesse, um assunto, seja por exemplo: "Cliente" ou "Qualidade de serviço" ou, em nosso caso, um tema de estudo do PIATAM, por exemplo, "Ictiofauna" ou "Limnologia".

Essa diferença entre a orientação das aplicações a processos e funções e a orientação por assuntos do DW aparece no conteúdo dos dados. O DW utiliza somente os dados que serão usados no processamento do sistema de suporte a decisão, enquanto que os dados orientados a aplicação possuem detalhes que atendem aos requisitos imediatistas do processamento funcional, que podem ou não ter relevância ou interesse ao analista de SAD (Sistema de Apoio a Decisão).

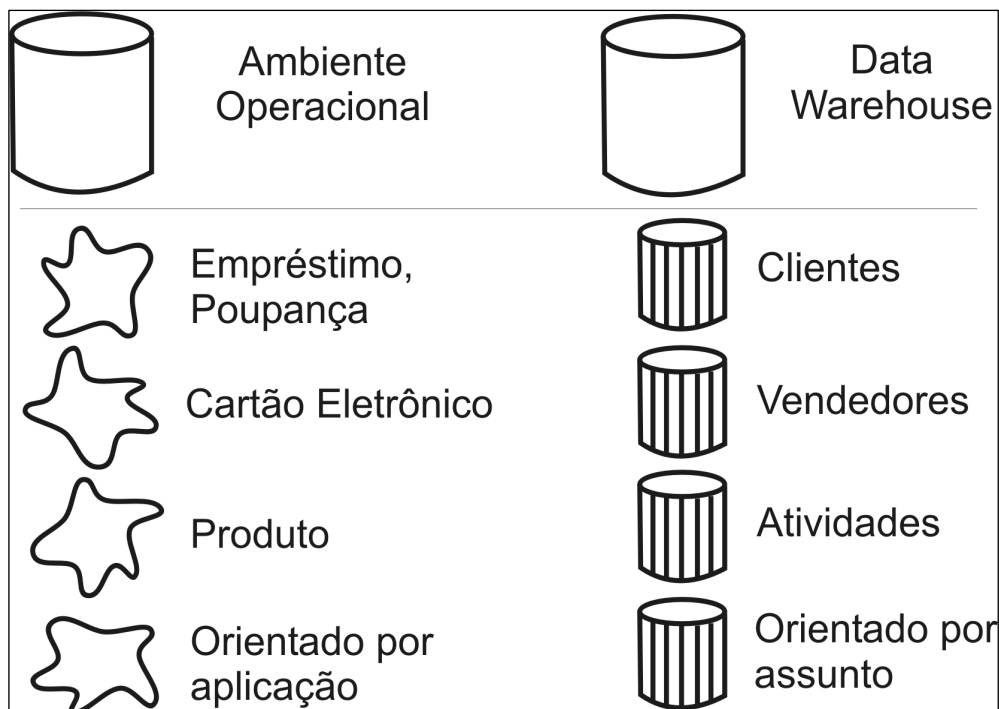


Figura 3.1 - Orientação dos dados no ambiente operacional e no DW. Fonte: Adaptada de (MACHADO, 2004).

3.2 Data Warehouse

O Data Warehouse (armazém de dados) possui como objetivo satisfazer as necessidades dos usuários (normalmente executivos, gerentes, analistas ou pessoas relacionadas com processos decisórios) quanto ao armazenamento dos dados que servirão de base para a realização de consultas e análises necessárias ao gerenciamento dos negócios.

Nos últimos anos, o conceito de “Data Warehouse” – DW evoluiu de um considerável conjunto de idéias relacionadas com armazenamento de dados para uma arquitetura voltada para a extração de informação especializada a partir de dados operacionais.

O objetivo fundamental de um Data Warehouse é servir de base para análise e consulta. Principalmente análise do tipo estratégica, quando se busca um apoio objetivo no conjunto de informações geradas por qualquer organização para se tomar uma decisão de razoável relevância (KIMBAL 1996).

Segundo INMON (1992) a ferramenta é:

- Orientada aos principais assuntos ou negócios da empresa como clientes, vendas, produtos, apólices, tratamentos, seguros, viagens, etc., enquanto os sistemas de informações tradicionais são orientados a processos como estoques, entradas e saídas de materiais, compras e vendas, faturamento e contabilidade;
- Integrada, ou seja, trabalha de forma a padronizar os termos e as estruturas técnicas que são utilizados nos sistemas de informações tradicionais, por exemplo, nestes diversos sistemas o sexo pode ser armazenados como: "m" ou "f", "0" ou " 1 ", "x" ou "y", "macho" ou "fêmea", "homem" ou "mulher", "dama" ou "cavalheiro" e outras formas; no DW apenas uma destas formas poderá aparecer padronizando esta referência;
- Não volátil, ou seja, no ambiente operacional, os dados sofrem as alterações necessárias como: incluir, alterar ou excluir dados. Porém,

no DW, os dados permitem apenas duas atividades: a sua carga para o banco de dados e as consultas; os dados armazenados não são alterados;

- Variante no tempo, ou seja, a estrutura dos dados do DW sempre contém algum elemento de tempo, enquanto nos sistemas de informações tradicionais isso não ocorre obrigatoriamente; Nestes sistemas, o horizonte de tempo é normalmente de 2 a 3 meses, enquanto no DW este horizonte pode chegar a 10 anos.

Um dos benefícios proporcionados pelo DW é a diminuição do tempo que os usuários levam para obter as informações necessárias aos seus processos decisórios, uma vez que diminui o tempo gasto com tarefas operacionais, como pesquisa e identificação dos dados necessários. Ao reunir informações dispersas em diversos bancos de dados e plataformas distintas, o DW permite que sejam feitas análises bastante eficazes, transformando dados esparsos em informações estratégicas, antes inacessíveis ou pouco aproveitadas (TAURION, 1997).

3.2.1 As Principais Ferramentas Utilizadas em um Data Warehouse

Segundo pesquisa realizada por BISPO (1998), as principais ferramentas utilizadas em um Data Warehouse são:

- Ferramenta para armazenamento: são os bancos de dados, considerados o coração do Data Warehouse e parte imprescindível do projeto;
- Ferramenta para a extração de dados: busca, nas bases de dados operacionais, os dados que vão ser armazenados no Data Warehouse;
- Ferramenta para a transformação de dados: ajusta os dados operacionais para o formato do Data Warehouse. Este formato auxilia as futuras pesquisas;

- Ferramenta para o refinamento ou limpeza de dados: faz os ajustes necessários nos dados, fazendo correções, desmembramento e fusões de dados, quando necessário, visando melhorá-los para facilitar as futuras pesquisas;
- Ferramenta para transferência de dados e replicação: pode ser considerada um subconjunto da ferramenta de extração. Não faz nenhum tipo de processamento ou transformação, apenas transfere um dado de um lugar "A" para "B". Geralmente, é utilizada para facilitar e dar uma resposta mais rápida às consultas ou análises, movendo os dados para um lugar apropriado e permitindo agilizar o serviço solicitado;
- Ferramenta para gerenciamento e administração: o gerenciamento é o monitoramento dos bancos de dados, por exemplo, quanto ao desempenho, integridade e segurança de dados; enquanto a administração é o monitoramento do suporte ao sistema, tais como, os recursos humanos e os esquemas de manutenção preventiva e corretiva dos equipamentos;
- Ferramentas para gerenciamento de consultas: fazem consultas e/ou geram relatórios, extraindo os dados do Data Warehouse, resumindo-os e apresentando-os em um formato apropriado.

Em poucas palavras, este conjunto de ferramentas que formam o DW tem por finalidade permitir a extração de informação a partir de dados operacionais, armazenando estes dados em um formato que torne as futuras pesquisas mais ágeis e provendo o usuário de ferramentas para a realização destas pesquisas.

3.2.2 Arquitetura de um Data Warehouse

Um exemplo de arquitetura de um Data Warehouse pode ser visto na figura 3.2 com seus componentes descritos a seguir:

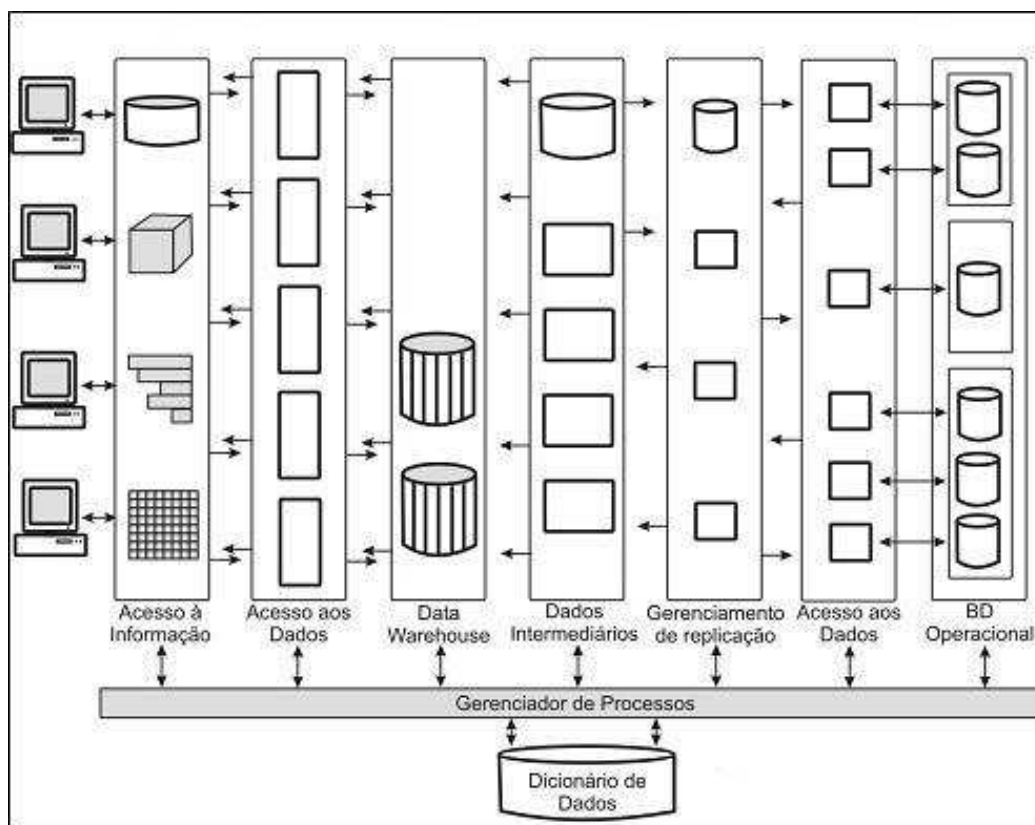


Figura 3.2 - Arquitetura do projeto de DW. Fonte: Adaptada de (KIMBALL, 1998).

- **Dicionário de Dados (ou Metadados):** São as informações que descrevem os dados utilizados pela empresa, isto envolve informações como descrições de registros, comandos de criação de tabelas, diagramas Entidade/Relacionamentos (E-R), dados de um dicionário de dados, etc. É necessário que exista uma grande variedade de metadados no ambiente de data warehouse para que ele mantenha sua funcionalidade e os usuários não precisem se preocupar onde residem os dados ou a forma com que estão armazenados;

- Camada BD Operacional: Corresponde aos dados dos sistemas operacionais da organização, junto com dados provenientes de outras fontes externas que serão tratados e integrados para compor o DW. Pode ser, por exemplo, o banco de dados dos produtos de uma empresa, do departamento de RH, ou dos fornecedores;
- Camada Acesso aos Dados: Esta camada é responsável pela ligação entre as ferramentas de acesso à informação e os bancos de dados operacionais (comunica-se com diferentes SGBDs). Pode ser a representação dos dados fontes no ambiente de DW ou o ponto de comunicação entre as ferramentas de manipulação de dados com os dados do DW;
- Camada Gerenciamento de Replicação: Esta camada inclui todos os processos necessários para selecionar, editar, resumir, combinar e carregar o DW e as correspondentes informações de acesso a partir das bases operacionais. Exemplificando, seriam os módulos de integração dos dados, podendo ser um cálculo de média de um dado em uma tabela fonte para ser armazenado no DW;
- Camada Dados Intermediários: Gerencia o transporte de informações pelo ambiente de redes. É usada para isolar aplicações, operacionais ou informacionais, e entregá-las em locais e tempos determinados;
- Camada Data Warehouse: O DW corresponde aos dados usados para fins "informacionais" (analíticos). Em alguns casos, DW é simplesmente uma visão lógica ou virtual dos dados, podendo de fato não envolver o armazenamento destes dados. Como exemplo, poderia ser um DW que visa analisar o Clima de uma região, tendo como fonte de dados a média da temperatura, o tempo e as regiões estudadas (seguindo uma hierarquia: bacia hidrográfica, sub-bacias, ponto de medição);
- Camada de acesso à informação: É a camada com a qual os usuários finais interagem. Representa as ferramentas que o usuário utiliza para navegar pelos dados de DW. Podemos citar uma ferramenta, produzida pela Oracle, Oracle 9i Discoverer. Também envolve o hardware e software utilizado para obtenção de relatórios, planilhas, gráficos e outros;

- Camada de gerenciamento de processos: Está envolvida com o controle das diversas tarefas a serem realizadas para construir e manter as informações do dicionário de dados e do DW. Contribui para manter o DW atualizado e consistente.

3.2.3 Integração dos Dados

As diferentes decisões de projetistas de aplicações refletem estilos distintos, originando centenas de maneiras de codificação, estruturas e convenções de nomes.

No Data Warehouse, a forma como o dado é representado não é considerada um fator de vital importância. O que realmente importa é se o dado está integrado e consistente, independente de sua fonte. Quando os dados são movidos do ambiente operacional (orientado a aplicações) para o ambiente analítico, eles são integrados antes de entrarem no DW. Assim, os dados no DW precisam ser armazenados de forma única, aceitável globalmente, mesmo quando as aplicações operacionais subjacentes armazenam os dados de modo diferente.

INMON (1997) classifica a integração como a mais importante característica dos sistemas de DW. Essa característica aparece na consistência das convenções de nomes, na consistência das variáveis de medidas, consistência de atributos físicos de dados e assim por diante.

A figura 3.3 ilustra o processo de integração. Os dados existentes no ambiente operacional (à esquerda da figura), apesar de terem representação lógica correta, não estão padronizados. Através do processo de integração, os dados, que podem até ser originados de diferentes ambientes operacionais, passam para o ambiente de DW (à direita) de forma padronizada e integrada. Ou seja, a aplicação I.A possui uma representação física do conceito Masculino e Feminino representada pelos caracteres "M" e "F" respectivamente.

A aplicação I.B também possui essa representação física com os números "1" e "0" e a aplicação I.C usa os conjuntos de caracteres "masculino" e "feminino".

Podemos observar que na Aplicação II da figura 3.3, a integração de dados ocorre de forma análoga ao da Aplicação I.

A integração vem para padronizar essa representação física dentro do DW.

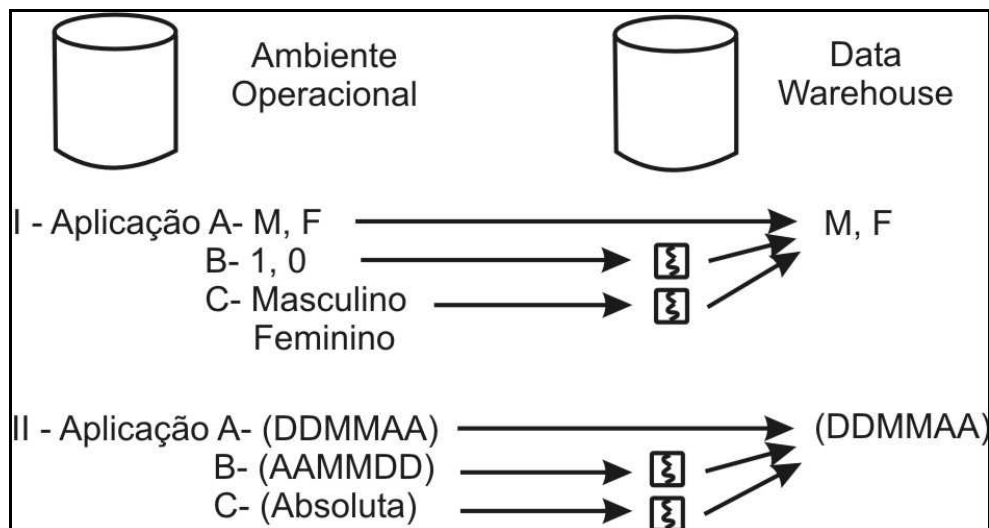


Figura 3.3 - Integração dos dados no DW. Fonte: Adaptada de (Machado, 2004).

Ao contrário do ambiente de dados operacional, em que os dados estão corretos no momento do acesso, todos os dados no DW são preciosos em algum instante no tempo (não sendo necessário o “agora”), ou seja, os dados variam com o tempo, tornando intuitiva a associação com um histórico de dados.

Em sistemas de DW esse elemento de tempo sempre estará presente; assim, o banco de dados armazena um histórico dos dados operacionais a que ele está ligado. Essas características relacionadas à variação de tempo podem ser observadas na figura 3.4, onde o horizonte de tempo em que os dados ficam armazenados no banco de dados é muito superior nos sistemas de DW.

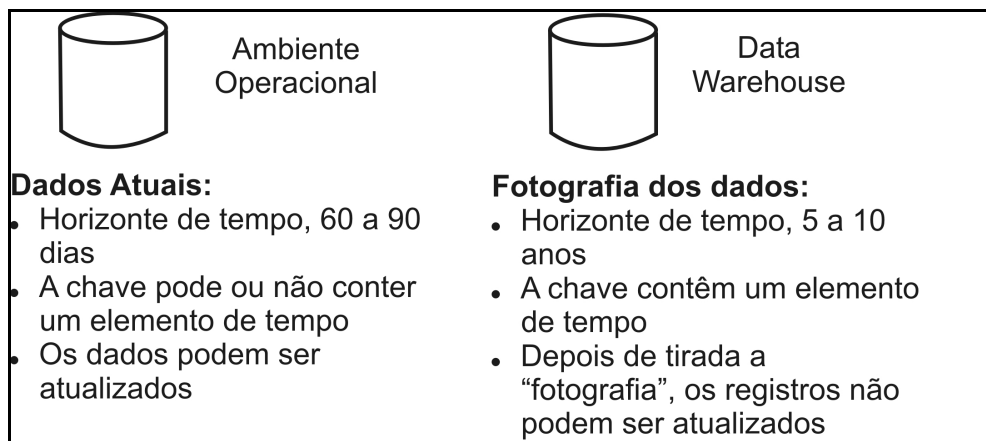


Figura 3.4 - Variação do tempo nos dados operacionais e no DW. Fonte: Adaptada de (Machado, 2004).

Os dados do DW não são voláteis como em ambientes de dados operacionais, ou seja, no ambiente OLTP, operações de atualização, inserções, remoções e alterações são executadas regularmente (registro a registro), alterando valores e, conseqüentemente, perdendo os valores armazenados antes da alteração. No DW a manipulação dos dados é mais simples, incluindo apenas dois tipos de operações: a carga inicial e o acesso (consulta) aos dados. Não existe atualização de dados no DW como parte de processos normais; então, não há perda de informações. Podemos observar essas características na figura 3.5, onde os dados operacionais sofrem várias ações que alteram seus estados e no ambiente DW, os dados são somente carregados e acessados.

Se a "fotografia" dos dados a serem armazenados no DW for tirada incorretamente, então ela poderá ser corrigida. Assumindo que os dados armazenados estão corretos, eles não são mais alterados.

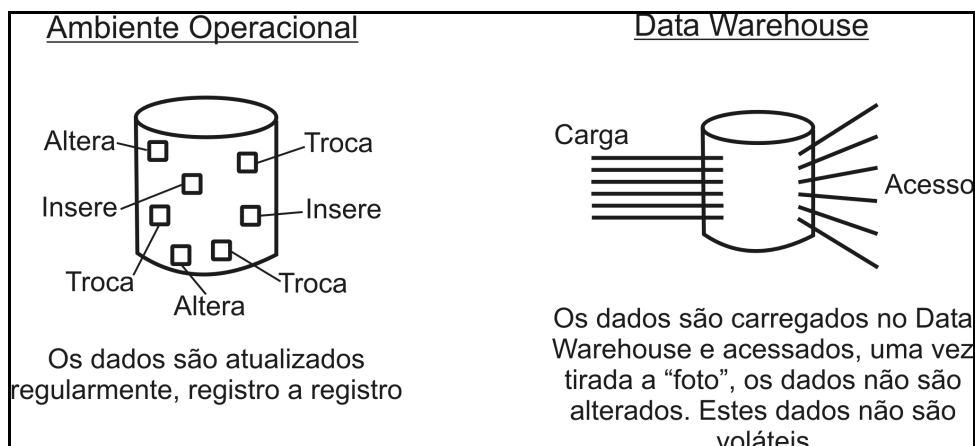


Figura 3.5 - Volatilidades dos dados nos sistemas operacionais e no Data Warehouse. Fonte: Adaptada de (Machado, 2004).

3.2.4 A Modelagem dos Dados

A modelagem de dados é a prática de elaborar um banco de dados usando modelos de dados já consagrados (MACHADO, 2004).

Segundo MACHADO (2004), o processo de modelagem pode ser descrito da seguinte maneira: primeiro, desenvolve-se um modelo conceitual de alto-nível do processo ou atividade que se deseja modelar. Depois, usa-se este modelo para derivar um modelo lógico, no qual os dados são abordados com mais detalhes. Finalmente, a partir do modelo lógico elabora-se o modelo físico que provê todos os detalhes da implementação do banco de dados.

As duas principais técnicas de modelagem de dados são: a Modelagem Entidade-Relacionamento (ER) e a Multidimensional.

O modelo Entidade-Relacionamento divide os dados em diversas tabelas, que se relacionam entre si, formando um complexo diagrama, como exemplificado na figura 3.6. Sua estrutura é importante para a eficiência e o desempenho no ambiente operacional, onde aplicativos específicos trabalham com esta estrutura e, normalmente, não são necessárias consultas que extrapolem o âmbito desses aplicativos. Quando consultas extras tornam-se necessárias, são feitas atualizações nos referidos aplicativos para que incorporem as novas consultas, porém, isto pode demandar muito tempo de trabalho.

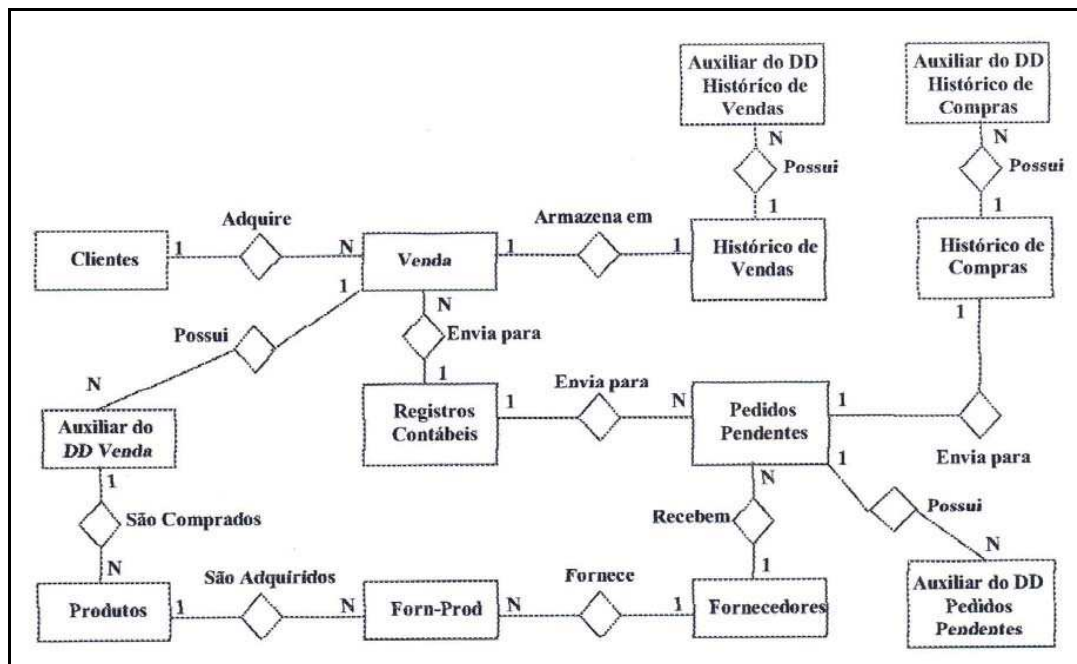


Figura 3.6 - Diagrama de um modelo Entidade-Relacionamento. Fonte: (MARQUES *et al.*, 2000).

De acordo com MACHADO (2004), em um ambiente operacional, a modelagem ER é a única opção a ser escolhida. Entretanto, com o advento do DW, houve a necessidade de uma técnica como suporte ao ambiente de análise multidimensional de dados, uma vez que o foco deixa de ser o controle operacional e passa a ser a disponibilização de um histórico de dados para consultas e análises.

3.2.5 Modelagem Multidimensional

A modelagem multidimensional é uma técnica utilizada para a conceitualização de modelos de negócio como um conjunto de medidas descritas por aspectos comuns (BARQUINI, 1996). Para entender melhor essa definição, é importante descrever alguns conceitos relacionados com a modelagem multidimensional, independentemente da tecnologia de banco de dados utilizada para implementá-la.

O esquema estrela é capaz de modelar as múltiplas dimensões de um repositório de dados através de **tabelas de dimensão**, que estão relacionadas com uma tabela central, também chamada de **tabela de fatos**. Os fatos são valores ou índices que podem ser medidos em um determinado processo de negócio, as dimensões são classes que descrevem as medidas numéricas, sendo que cada

dimensão é descrita por um conjunto de atributos que muitas das vezes formam uma hierarquia (GATZIU & VAVOURAS, 1999; GOLFARELLI *et al*, 1998; KIMBALL, 1997).

Como exemplo, podemos tomar uma tabela de fatos como sendo as vendas realizadas por uma empresa. Nesse exemplo, as tabelas de dimensão devem armazenar informações como o tipo do produto envolvido na venda, a data em que a venda foi efetuada e o consumidor envolvido. Os atributos ano, mês, semana e dia da dimensão Data formam uma hierarquia. O esquema estrela desse exemplo é mostrado na Figura 3.7. Neste exemplo, as chaves primárias Consumidor e Produto são representadas pelas siglas Cid e Pid respectivamente.

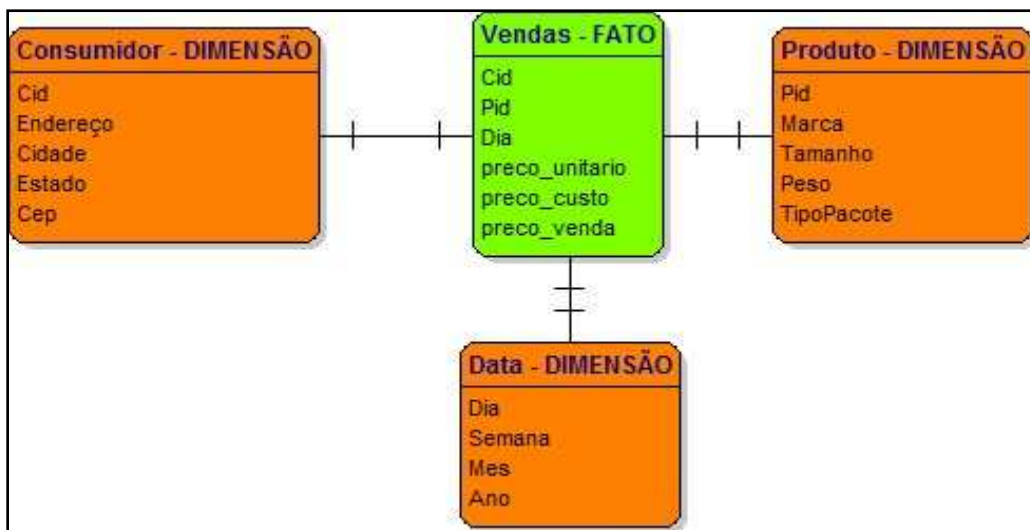


Figura 3.7 - Exemplo de esquema estrela sobre vendas. Fonte: Adaptado de (MARQUES *et al.*, 2000).

O esquema estrela objetiva a desnormalização dos dados, visando um melhor desempenho no ambiente de apoio à tomada de decisão em relação às estruturas altamente normalizadas das bases de dados operacionais.

O segredo para se obter esse desempenho é limitar o número de uniões que terão de ser realizadas e a complexidade de cada união. Esse esquema objetiva também a criação de um modelo de dados que seja mais compreensível ao usuário final, procurando representar a maneira natural de como ele enxerga o seu negócio, uma vez que os esquemas ER (Entidade-Relacionamento) são de difícil interpretação por parte dos usuários finais, além de não representarem a maneira natural de como eles visualizam seu negócio (KIMBALL, 1997; TODMAN, 2001).

Uma variação do esquema estrela, chamado de esquema floco de neve, é utilizada para representar as hierarquias das dimensões através da normalização das tabelas de dimensão. No esquema floco de neve, as tabelas de dimensão podem se tornar tabelas de fatos de outras tabelas obtidas. A vantagem desse tipo de esquema é que se torna mais fácil a manutenção das tabelas de dimensão, já que há uma diminuição na redundância dos dados. Entretanto, uma estrutura não normalizada é mais eficiente no momento de execução das consultas, um requisito indispensável em sistemas de DWs (GATZIU & VAVOURAS, 1999; GOLFARELLI *et al.*, 1998; KIMBALL, 1997).

Quando o esquema de um DW é composto por mais de uma estrutura do tipo estrela, este pode ser chamado de esquema constelação (BARQUINI, 1996).

As tabelas de dimensão são caracterizadas por vários aspectos gerais. Normalmente elas são altamente desnormalizadas. Embora, freqüentemente, fale-se que o esquema estrela é desnormalizado, na verdade somente as tabelas de dimensão são desnormalizadas. As tabelas de dimensão possuem mais colunas do que as tabelas do banco de dados operacional e geralmente possuem menos registros do que as tabelas de fatos.

Em certas situações é necessário utilizar uma chave substituta para a dimensão, pois ela permite a captura do histórico da dimensão e em outros casos fornecem um desempenho de união melhor do que as chaves operacionais. As dimensões também possuem referências aos registros correspondentes nas tabelas de origem, além de campos de data adicionais e flags indicando se um determinado registro da dimensão está ativo ou não (COREY *et al.*, 2001).

Há uma outra maneira de considerar a multidimensionalidade de um repositório, onde as múltiplas dimensões do Data Warehouse são representadas por meio de cubos de dados. Cada eixo do cubo corresponde a uma dimensão.

Considerando o esquema estrela apresentado anteriormente, as dimensões produto, data e consumidor vão constituir os eixos do cubo de dados. Os pontos de interseção entre todas as dimensões são chamados células e representam uma visão do cubo. A maior parte das visões podem ser computadas em função de outras. Diz-se que tais visões são dependentes. A Figura 3.8 ilustra um cubo de dados.

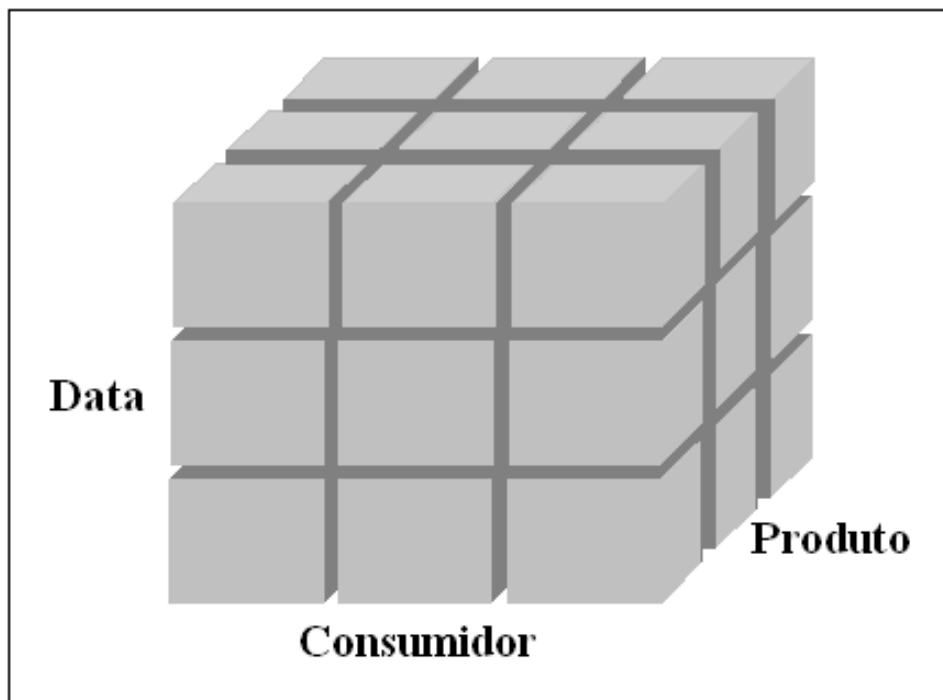


Figura 3.8 - Um exemplo de cubo de dados sobre vendas. Fonte: (MARQUES *et al.*, 2000).

Projetistas de bancos de dados podem e devem separar o conceito de visão multidimensional dos dados, obtida através da modelagem multidimensional, do conceito de armazenar os dados de forma multidimensional.

Os resultados da modelagem multidimensional podem ser implementados diretamente utilizando a tecnologia de banco de dados multidimensional ou através do esquema estrela, em um banco de dados relacional. A falta de um modelo de dados multidimensional convencional, tal como o esquema estrela, para bancos de dados relacionais e a falta de um método de acesso padrão, tal como o SQL, acabaram influenciando a utilização da tecnologia de bancos de dados relacional para representar e armazenar dados multidimensionais (KIMBALL, 1997).

3.3 OLAP (*On-line Analytical Processing*)

O Data Warehouse, como foi mostrado, é basicamente uma ferramenta que visa o armazenamento de dados, porém, não dispõe de recursos para realizar consultas e análises muito sofisticadas ou cálculos mais complexos. Para esses casos, existe uma outra ferramenta de *Business Intelligence*, o OLAP, que possui a capacidade de processamento dos dados.

O grande diferencial do OLAP está principalmente na sua capacidade de permitir consultas e análise dos dados de forma consistente, interativa, rápida e com uma grande variedade de possíveis visões dos dados. No OLAP, as respostas não são automáticas. Trata-se de um processo interativo em que o usuário formula hipóteses, faz consultas, recebe informações, verifica um dado específico em profundidade e faz comparações.

O OLAP auxilia os usuários a sintetizarem as informações sobre a empresa por meio de comparações, visões personalizadas, análises estatísticas, previsões e simulações.

3.3.1 Conceitos Básicos sobre OLAP

Segundo MACHADO (2004), OLAP é o conjunto de ferramentas que possibilita efetuar a exploração dos dados de um Data Warehouse.

OLAP também pode ser entendido como um conjunto de ferramentas especialmente projetadas para dar suporte ao processo decisório através de consultas, análises e cálculos por parte dos seus usuários.

Esta ferramenta tornou-se a sucessora dos Sistemas de Informações para Executivos (EIS - Executive Information Systems), sistemas desenvolvidos nas décadas de 80 e 90 com o intuito de auxiliar na obtenção de informações úteis para o gerenciamento de crises, ou seja, períodos de emergência. Atualmente, o OLAP

possui um âmbito mais amplo que a citada ferramenta, uma vez que, além de dar suporte ao processo decisório, também suporta os processos de análise estratégica. (MACHADO, 2004).

A ferramenta OLAP permite que seus usuários tenham acesso aos dados que descrevem uma determinada atividade ou negócio, permitindo-lhes uma melhoria na compreensão, gerenciamento e planejamento destes negócios. Permite, ainda, analisar as múltiplas dimensões dos dados através de muitos ângulos e combinações, além de identificar tendências e descobrir o que está conduzindo os negócios.

A ferramenta pode ser usada em diversas atividades:

- Departamentos de Finanças - para planejar orçamentos e realizar análises financeiras;
- Departamento de Vendas - para fazer análises e estimativas de vendas;
- Departamento de Marketing - para realizar pesquisas e análises de mercado, estimativas, análises de clientes e segmentação de mercado;
- Manufatura - para realizar o planejamento, análises da produção e análises de falhas ou defeitos;
- Gerenciamento de energia e utilidades - para auxiliar na análise e interpretação de dados provenientes de medições, elaboração de indicadores de desempenho e correlação entre as grandezas medidas, viabilizando estudos que tenham como foco o uso racional da energia.

As ferramentas OLAP permitem aos usuários analisar os dados em dimensões múltiplas, como região, produto, tempo e vendedor, por exemplo. Cada dimensão também pode conter hierarquias, por exemplo, a dimensão tempo pode conter as hierarquias ano, trimestre, mês, semana ou dia. A dimensão região pode ter as hierarquias: continente, país, estado, cidade ou bairro. Os dados, nestas dimensões, são agregados, ou seja, são reunidos, mas pode-se navegar livremente de uma hierarquia para outra até chegar-se na mínima granularidade dos dados.

Essa navegabilidade é importante por permitir a obtenção de informações com diferentes níveis de sumarização dos elementos e de detalhes disponíveis nos dados.

Desta maneira, o usuário pode fazer análises compatíveis com o horizonte pretendido, ou seja, pode obter informações provenientes da visualização de comportamentos e tendências ao longo da janela de tempo desejada.

3.3.2 Tipos de Sistemas OLAP

BISPO (1998) identificou quatro tipos de estruturas OLAP:

- ROLAP – Relational On-Line Analytical Processing: O ROLAP é uma simulação da tecnologia OLAP feita em banco de dados relacionais que, por utilizar a estrutura relacional, possui a grande vantagem de não ter restrições no volume de armazenamento de dados;
- MOLAP – Multidimensional On-Line Analytical Processing: O MOLAP é uma classe de sistemas que permite a execução de análises bastante sofisticadas, usando banco de dados multidimensionais. A estrutura multidimensional é chamada de cubo, onde as dimensões do cubo representam os componentes dos negócios da empresa, como modelo, departamento, tamanho, etc., e a interseção das dimensões é chamada de medida, como custo ou quantidade vendida. Apesar do termo “cubo”, o MOLAP pode trabalhar com diversas dimensões, inclusive mais que três. Este tipo de estrutura propicia ao sistema um ótimo desempenho no acesso a qualquer dado, com rapidez e um rico e complexo conjunto de funções de análise. A figura 3.9 ilustra uma estrutura multidimensional;

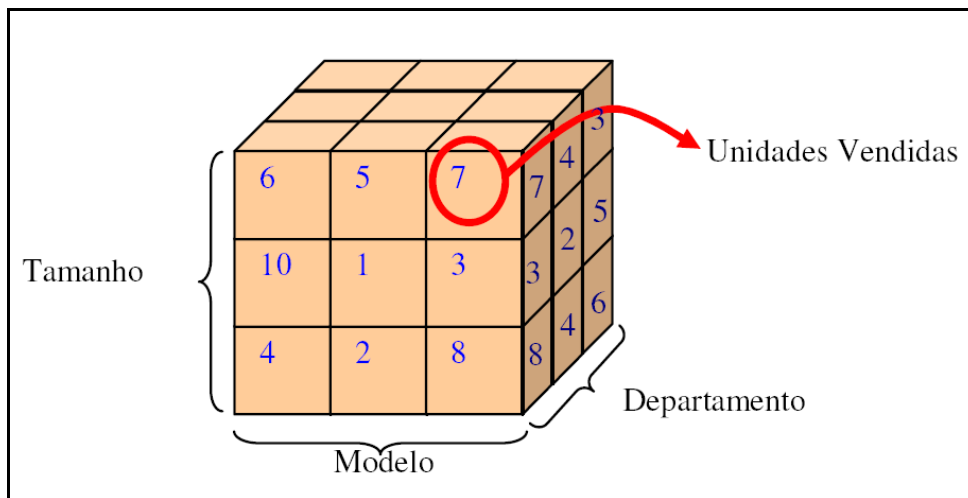


Figura 3.9 - Modelo de Banco de Dados Multidimensional – Tridimensional.
Fonte: (BISPO, 1998).

- HOLAP – Hybrid OLAP: As ferramentas HOLAP utilizam um sistema híbrido de MOLAP e ROLAP. São inteligentes e selecionam automaticamente a tecnologia mais adequada, de acordo com a atividade que será executada, proporcionando um máximo desempenho;
- WOLAP ou Web OLAP: Já existe a migração da tecnologia OLAP para o ambiente da Internet, e a nova versão da ferramenta está sendo chamada de Web OLAP ou WOLAP.

3.3.3 Funções

A fim de permitir uma visualização e manipulação multidimensional dos dados, as ferramentas OLAP oferecem diferentes funções (Codd, 1993; Inmon & Hackathorn, 1997). São elas:

- **Pivot:** muda a orientação dimensional de uma pesquisa. Por exemplo, pivot pode consistir na troca de linhas e colunas, ou mover uma das

dimensões da linha, para a dimensão da coluna, como ilustrado na figura 3.10;

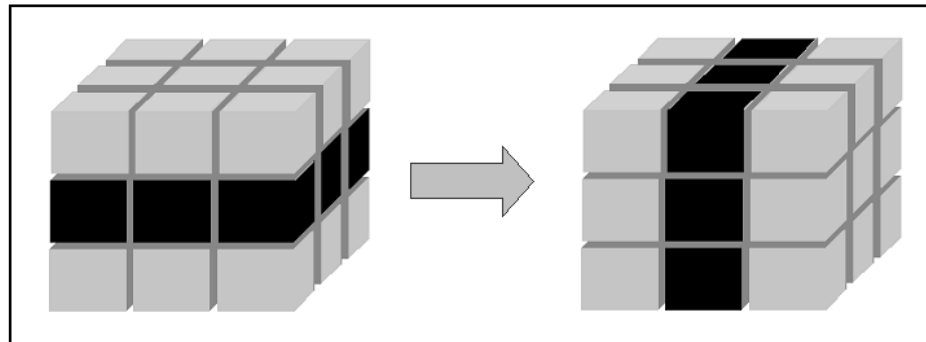


Figura 3.10 - Operação de Pivot. Fonte: (BISPO, 1998).

- **Roll-up:** as bases de dados multidimensionais geralmente têm hierarquias ou relações de dados baseadas em fórmula dentro de cada dimensão. Então, a execução do roll-up computa todas essas relações para uma ou mais dimensões;
- **Slice:** um slice é um subconjunto da estrutura multidimensional que corresponde a um valor simples em lugar de um ou mais atributos das dimensões. É como fixar um valor de uma das dimensões de um cubo e considerar para pesquisa o subcubo formado por esse valor e pelas outras dimensões do cubo inicial, como mostrado na figura 3.11;

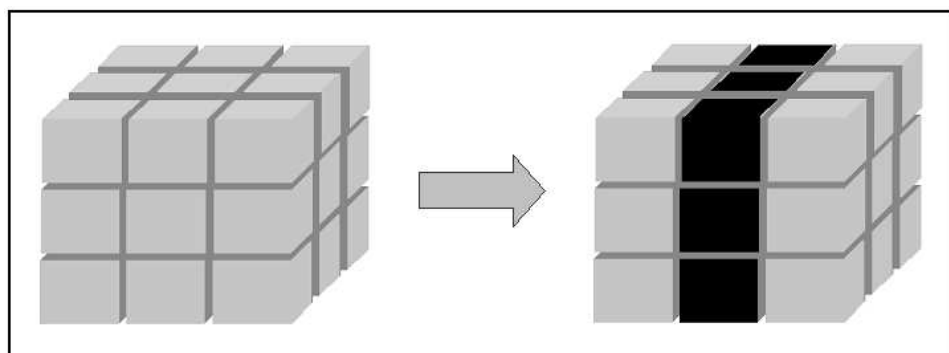


Figura 3.11 - Operação de Slice. Fonte: (BISPO, 1998).

- **Drill-down/up:** consiste em fazer uma exploração em diferentes níveis de detalhe das informações, como por exemplo, analisar uma

informação por continente, país ou estado, partindo da mesma base de dados. Essa função é ilustrada na figura 3.12;

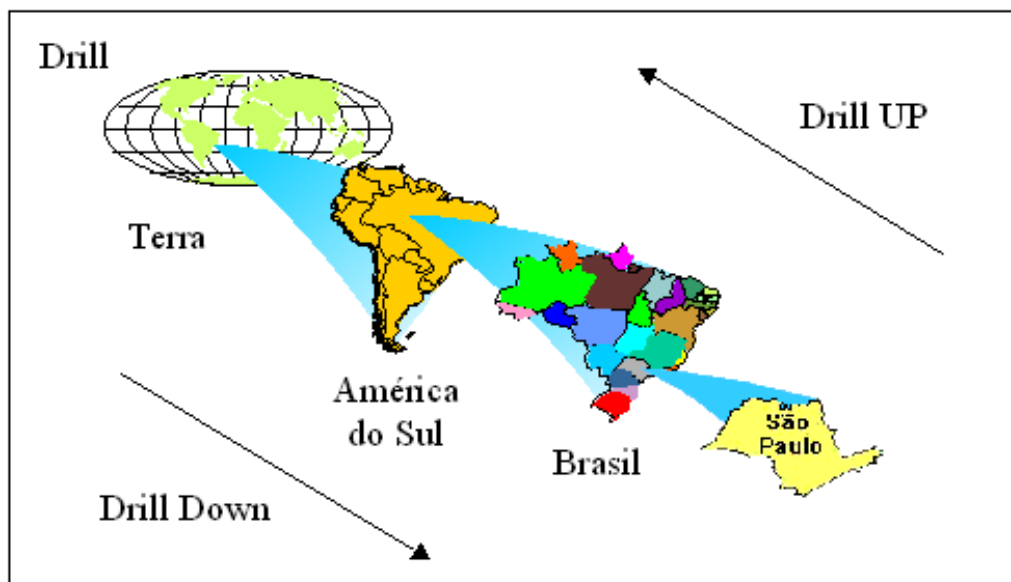


Figura 3.12 - Operação de Drill-down/up. Fonte: (BISPO, 1998).

- **Drill-across:** é o processo de unir duas ou mais tabelas fatos de mesmo nível de detalhes, ou seja, tabelas com o mesmo conjunto de colunas e restrições dimensionais.

3.3.4 Integração de OLAP com SIG (Sistema de Informações Geográficas)

Atualmente, a integração entre as tecnologias de suporte à decisão convencional (*DW* e *OLAP*) e os *SIG* têm despertado interesse especial tanto da comunidade científica quanto das organizações comerciais, pois, se do lado do *SIG*, *OLAP* incrementa o poder das consultas espaciais, do lado do *DW/OLAP*, existem muitas informações de natureza geográfica que precisam de um suporte espacial para serem melhor discernidas.

No contexto dos sistemas de suporte a decisões estratégicas que envolvem *SIG*, sistemas *OLAP* e *SIG-OLAP* integrados, pode-se classificar as linguagens de consulta encontradas na literatura em três principais grupos. Esta classificação pode

ser visualizada conforme aspectos relacionados ao contexto de aplicação da linguagem, à natureza dos dados que serão recuperados e quanto aos tipos de operadores que a linguagem permite aplicar sobre os referidos dados.

O primeiro grupo consiste no espacial, bastante utilizado em Sistemas de Informações Geográficas (SIG), permitindo a recuperação de dados armazenados em uma base geográfica. Neste tipo de linguagem de consulta, é possível a utilização de operadores espaciais (OPEN GIS CONSORTIUM, 2001; CLARKE, 1997) para calcular distância entre feições geográficas, adjacência, verificar se uma feição geográfica intercepta outra, entre outras operações.

Outro grupo de linguagem de consulta é aquele voltado para consulta a dados multidimensionais, comumente empregado em sistemas OLAP. Neste grupo de linguagem, o aspecto principal é a recuperação e análise de dados multidimensionais, normalmente armazenados em um DW através da utilização de operadores analíticos (SILVA & TIMES, 2004). Estes operadores possibilitam a implementação de visões configuráveis dos dados em diferentes ângulos e níveis de agregação devido à estrutura multidimensional do DW.

Por fim, tem-se o grupo das linguagens de consulta geográfica multidimensional. Neste grupo, estariam as linguagens com características e operadores pertinentes aos dois grupos citados anteriormente, permitindo assim, a utilização simultânea tanto de operadores espaciais como multidimensionais. O principal foco de aplicação deste tipo de linguagem seria na consulta de dados armazenados em um DWG (Data Warehouse Geográfico) (FIDALGO *et al*, 2004), que é uma das tecnologias essenciais para o provimento de um ambiente de suporte à decisão geográfico-multidimensional.

Apesar destes trabalhos, as propostas existentes não provêm uma integração completa entre as operações geográficas e multidimensionais, o que torna um desafio a proposição de linguagens com tais características.

No próximo capítulo, serão abordados os conceitos de Data Mining e, em especial, as Regras de Associação.

Capítulo 04

REGRAS DE ASSOCIAÇÃO

Serão compreendidos neste capítulo, os conceitos de Mineiração de Dados e, em especial, as Regras de Associação. A descoberta de regras de associação é uma técnica de mineração de dados que procura identificar determinados padrões de dados em bases de dados de grande dimensão, permitindo, após a sua interpretação, adquirir conhecimento específico acerca do problema em análise.

4.1 INTRODUÇÃO

A Descoberta de Conhecimento em Bases de Dados (DCBD), também conhecida como KDD (*Knowledge Discovery in Databases*), é resultado de um longo processo de pesquisa que se iniciou devido ao grande volume de dados armazenados nos bancos de dados e à complexidade destes dados. Seja pela existência de equipamentos que transformaram de forma expressiva a capacidade de coletar e armazenar dados ou seja pelo acesso e disponibilidade desses equipamentos em larga escala a preços compatíveis com a valia dos processos, a habilidade de armazenar os dados superou em muito a habilidade de analisar e entender os dados armazenados. Desta forma, exigem-se novas técnicas computacionais que suportem a extração automática de conhecimento dos grandes bancos. Essas técnicas, que auxiliam, inteligente e automaticamente, na tarefa de analisar dados na busca de conhecimentos úteis, são o objeto de estudo da área de pesquisa de DCBD.

A DCBD é um conjunto de procedimentos pelo qual se analisa e transforma um conjunto de dados em conhecimento, em padrões interessantes, fazendo utilização de técnicas automáticas para a extração destes padrões.

Diversas etapas compõem a DCBC, onde os dados são manipulados até que a informação útil seja revelada. Para que o processo seja iniciado é necessária a

compreensão do domínio da aplicação e dos objetivos a serem explorados. Fazem parte deste processo as seguintes etapas (FAYYAD *et al*, 1996):

- Seleção: Nesta etapa ocorre a seleção da base de dados relevantes que servirá para todo o processo;
- Pré-processamento: Fase utilizada para a limpeza dos dados, retirada de ruídos ou aberrações dos dados, para integração de dados heterogêneos e para tratar dados incompletos;
- Transformação: Esta etapa serve para que os dados sejam convertidos para o formato adequado à sua utilização pela fase de mineração de dados. Pode-se realizar, também, a redução de dados, permitindo um número menor de variáveis sob consideração na mineração;
- Mineração dos dados: Pode ser considerada o núcleo da DCBD, consistindo na aplicação de algoritmos para extração de padrões dos dados. Em virtude de sua importância, o termo mineração de dados é, por vezes, utilizado para identificar todo o processo de DCBD;
- Interpretação: Fase onde ocorre a interpretação correta dos resultados obtidos pela mineração para posterior consolidação do conhecimento e a viabilização de uma utilização prática do mesmo.

Esta apresentação das atividades pode sugerir que exista uma trajetória linear do processo de DCBD. No entanto, isso geralmente não se verifica, uma vez que em cada etapa pode ser identificada a necessidade de retorno para cada uma das etapas anteriores, sendo portanto um processo iterativo e interativo, também, pois necessita da intervenção de um especialista em todas as fases da descoberta. As etapas descritas anteriormente estão representadas na figura 4.1.

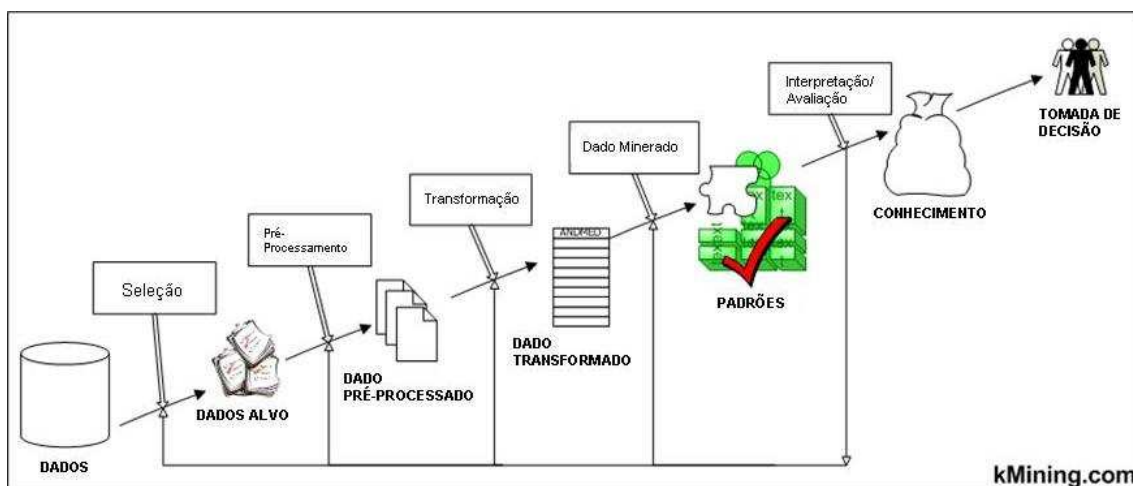


Figura 4.1 – O processo de Descoberta de Conhecimento em Bases de Dados Fonte: Adaptada de Business Intelligence, KDD and Data Mining News (<http://www.kmining.com/>).

Uma importante tarefa de mineração de dados é a descoberta de regras de associação. Essa tarefa é capaz de identificar grupos de atributos de um banco de dados tipicamente relacionados, representando a probabilidade de que um grupo apareça em uma transação visto que outro está presente na mesma.

Embora a descoberta de regras de associação possa ser aplicada a qualquer base de dados, foi inicialmente proposta para a análise de cesta de produtos (*market basket analysis*). Este tipo de base de dados é composta pelos itens adquiridos pelo cliente em uma determinada cesta de compras, ou seja, em uma mesma transação.

Muitas decisões de *marketing* podem ser tomadas após o estudo dessas transações, auxiliando na reorganização das lojas, na produção de catálogos customizados, na seleção de produtos para promoção, na determinação de itens para uma propaganda mais inovadora, na seleção de artigos concorrentes, no fomento de *cross-selling*, na identificação de oportunidades de vendas de pacotes de produtos ou serviços, etc. (BRUSSO, 2000; JORGE, 2003; GONÇALVES, 2001).

Além da análise utilizada para o *marketing*, diversos estudos já foram realizados propondo a utilização da tarefa de regras de associação em diversas áreas: serviços bancários, telecomunicações, saúde, análise de páginas da *WEB*, entre outros (BRUSSO, 2000; JORGE, 2003; SRIKANT & AGRAWAL, 1997). Informalmente, as regras de associação podem ser vistas como um tipo de regra SE-ENTÃO (SIEBES, 2003).

Com o desenvolvimento dos Sistemas de Informações Geográficas, um novo tipo de dados (dados geo-referenciados) despontou, pois adicionam a riqueza do fator localização às restantes variáveis de caracterização. Mais recentemente, tem-se dado crescente atenção a outra característica interessante que é a introdução da dimensão tempo, que assim permite estudar a evolução dos acontecimentos, que pode assumir um caráter preditivo ou regressivo em relação a um dado momento temporal de um acontecimento.

O Data Mining Geo-Espacial (GAHEGAN, 2001; OPENSHAW, 1999) deve ser entendido como um tipo especial de mineração de dados, que procura realizar funções genéricas semelhantes às da mineração de dados convencional, modificadas por forma a atender os aspectos especiais da geo-informação (OPENSHAW 1994, 1999). No entanto, o que há de verdadeiramente importante na Mineração de Dados Geo-Espacial são as distribuições e padrões espaciais e espaço-temporais, pois proporcionam uma visão única da realidade e conferem-lhe as características singulares que possui. A importância da Mineração de Dados Geo-Espacial aumentou com a crescente divulgação e conseqüente massificação dos dados geo-referenciados, oriundos sobretudo de mapas digitais, imagens de sensoriamento remoto e, em alguns países, com a geo-referenciação dos dados censitários.

4.2 Regras de Associação

Em sua forma original, a tarefa de descobrir regras de associação foi definida para um tipo especial de dados, freqüentemente chamado de “*market basket data*”. Este tipo de banco de dados é composto por um conjunto de transações onde cada transação é composta por um conjunto de itens. Traduzindo para o ambiente relacional, este banco é representado por uma tabela T composta de atributos binários. Os atributos correspondem aos itens e os registros na tabela correspondem às transações. Os atributos binários assumem valor 1, caso o item esteja presente na transação, ou valor 0, caso o item não esteja presente na transação (SIEBES, 2003).

A Figura 4.2, representa a tabela T composta do atributo tID , identificador da transação, e de atributos binários $I1, \dots, I12$, que representam os itens.

t_{ID}	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}
1	1	1	0	0	1	0	0	0	0	0	0	0
2	0	1	0	0	1	1	0	0	0	1	1	1
3	0	0	0	0	0	1	0	0	1	0	0	0
4	0	1	1	0	0	1	0	0	1	0	0	0
5	0	0	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0	1	0
7	0	1	0	0	0	1	0	0	0	0	0	0
8	1	0	0	0	0	1	0	0	0	1	0	0
9	0	0	0	0	0	1	0	0	1	0	0	0

Figura 4.2 – Tabela 7. Fonte: Adaptada de (AGRAWAL *et al*, 1993).

Uma regra de associação consiste em uma expressão da forma $X \rightarrow Y$, onde X e Y são conjuntos de itens (AGRAWAL *et al*, 1993). Ou seja, $X_1 \wedge \dots \wedge X_m \rightarrow Y_1 \wedge \dots \wedge Y_n$, onde $X_i (i \in \{1, m\})$ e $Y_j (j \in \{1, n\})$ representam os itens, ou seja, são os atributos do banco de dados. Intuitivamente, as regras significam que as transações do banco que contêm X tendem a conter Y .

Um modelo formal para representar o problema de descoberta de regras de associação foi proposto por Agrawal, Imielinski e Swami em 1993 (AGRAWAL *et al*, 1993). Considere um conjunto de atributos binários $I = \{I_1, I_2, \dots, I_m\}$, chamado de itens. Seja D um conjunto de transações, onde cada transação $t \subset I$ é um conjunto de itens, chamado *itemset*. Cada transação é associada a um identificador único, chamado t_{ID} . Seja $X \subset I$ um conjunto de itens. Uma transação t contém X se $X \subset t$. Uma regra de associação é um relacionamento na forma $X \rightarrow Y$, onde $X \subset I$, $Y \subset I$ e $X \cap Y = \phi$.

A regra $X \rightarrow Y$ possui suporte s no conjunto de transações D se $s\%$ das transações em D contêm X e Y (AGRAWAL *et al*, 1993; AGRAWAL *et al*, 1996). O suporte é a probabilidade de uma transação em D conter $X \cup Y$, ou seja, ele indica a frequência da regra. O suporte $s(X \cup Y)$ é dado, então, por:

$$\text{quantidade} \frac{(X \cup Y)}{\text{tamanho}(D)} = \frac{|\{t \in D \mid X \subset t, Y \subset t\}|}{|D|}$$

A regra $X \rightarrow Y$ é válida no conjunto de transações D com confiança c se $c\%$ das transações do banco de dados que contêm X também contêm Y . A confiança é a probabilidade de Y ocorrer em uma transação de D visto que X ocorre, ou seja, indica a “força” da regra. A confiança é dada por:

$$\frac{\text{Suporte}(X \cup Y)}{\text{Suporte}(X)}$$

4.2.1 Decomposição da Geração de Regras

A tarefa de descobrir regras de associação consiste em gerar todas as regras que possuam suporte maior ou igual a um suporte mínimo e uma confiança maior ou igual a uma confiança mínima, sendo o suporte mínimo e a confiança mínima especificados pelo usuário. Esses parâmetros servem como medida capaz de filtrar os padrões inúteis, sendo trabalhados apenas padrões significativos (LOPES, 1999).

O problema de geração de regras de associação é geralmente decomposto em dois subproblemas (AGRAWAL *et al*, 1993; AGRAWAL *et al*, 1996) :

1) Descobrir todas as combinações de itens que tenham suporte maior ou igual ao mínimo previamente especificado. Essas combinações de itens são chamadas de conjunto de itens freqüentes (*large itemsets*), e as outras combinações de conjunto de itens raros (*small itemsets*). Se a combinação de itens freqüentes possui k itens, então, é designada por *k-itemset*;

2) Gerar as regras de associação do banco de dados utilizando os conjuntos de itens freqüentes. A regra só será válida se a confiança da mesma for maior ou igual à mínima predeterminada.

A performance da tarefa de mineração de regras de associação é de fato determinada pelo subproblema 1. As regras de associação correspondentes podem ser derivadas de uma maneira direta. Logo, a descoberta dos conjuntos de itens freqüentes é o foco maior dos trabalhos encontrados (PARK & CHEN, 1997).

4.2.2 Algoritmo Apriori

O algoritmo a seguir, descreve o algoritmo *Apriori*, proposto por (AGRAWAL & SRIKANT, 1994).

- 1) $F_1 = \{1 - \text{itemset}\}$
- 2) Para cada $k=2; L_{k-1} \neq 0$; (passo = 1) faça
- 3) $C_k = \text{Novos_Candidatos}(F_{k-1})$
- 4) Para todas as transações $t \in D$ faça
- 5) $C_t = \text{subconjunto}(C_k, t)$; // São os candidatos contidos em t
- 6) Para todos os candidatos $c \in C_t$
- 7) $c.\text{suporte} = c.\text{suporte} + 1$
- 8) $F_k = \{c \in C_k \mid c.\text{sup orte} \geq \text{sup orte}_{\text{minimo}}\}$
- 9) $F = F \cup F_k$

Inicialmente, o conjunto F_1 é identificado. F_1 é a lista de conjuntos de itens freqüentes com apenas um elemento. Em seguida, em cada passo k do algoritmo, C_k é gerado utilizando a lista F_{k-1} , onde F_{k-1} é a lista de conjuntos de itens freqüentes com $k-1$ elementos. Nesta etapa, a base de dados é percorrida para calcular o suporte de cada candidato. Todos os candidatos com suporte maior que o mínimo são incluídos na lista de itens freqüentes.

A rotina *Novos_Candidatos* do algoritmo *Apriori* gera os candidatos utilizando somente os conjuntos de itens freqüentes encontrados no passo anterior, sem considerar as transações no banco de dados. A percepção básica do *Apriori* é que qualquer subconjunto de um conjunto de itens freqüentes deve ser freqüente. Portanto, o conjunto de candidatos contento k itens pode ser gerado fazendo uma combinação dos conjuntos de itens freqüentes de tamanho $k-1$, e anulando aqueles que contenham algum subconjunto que não seja freqüente (AGRAWAL & SRIKANT, 1994).

Podemos observar abaixo, o algoritmo utilizado pelo *Apriori* para a geração do conjunto de itens candidatos.

- 1) $C_k \leftarrow \text{select } p.\text{item}_1, p.\text{item}_2, \dots, p.\text{item}_{K-1}, q.\text{item}_{K-1}$
 From $F_{k-1}p, F_{k-1}q$
 Where $p.\text{item}_1 = q.\text{item}_1, \dots, p.\text{item}_{K-2} = q.\text{item}_{K-2}, p.\text{item}_{K-1} < q.\text{item}_{K-1}$
- 2) Para todos os conjuntos de itens $c \in C_k$ faça
 Para todos os subconjuntos s de tamanho $k-1$ de c faça
 Se ($s \notin L_{k-1}$) então
 Retire c de C_k

No passo 1, é realizada a combinação dos conjuntos em F_{k-1} para a geração de C_k . No passo 2, ocorre a retirada de C_k dos candidatos que possuem algum subconjunto de tamanho $k-1$ que não seja freqüente, ou seja, não pertençam à lista F_{k-1} . No algoritmo *Apriori*, assume-se que os itens de cada transação do banco de dados estão em ordem lexicográfica, logo $p.\text{item}_1, p.\text{item}_2, \dots, p.\text{item}_k$ e $q.\text{item}_1, q.\text{item}_2, \dots, q.\text{item}_k$ estão assim ordenados.

A mineração de regras de associação é uma tarefa freqüentemente difícil no mínimo por duas razões: primeiro, existem muitos dados e, segundo, os dados possuem muitos atributos.

O tempo de execução do *Apriori* é linear em relação ao tamanho da amostra, logo, o número de passagens pela base de dados é um fator que reflete neste tempo. No *Apriori*, são realizadas n passagens pela base de dados, onde n é a quantidade de itens do maior conjunto de candidatos. Cada passagem pelo banco de dados significa uma rotina de E/S.

O número de atributos do banco de dados é um fator de aumento da complexidade do *Apriori*. O pior caso ocorre quando todos os subconjuntos do conjunto de atributos são freqüentes. Seja $|T|$ a quantidade de atributos da tabela T , que representa o banco de dados D . Neste caso, a complexidade é $O(2^{|T|})$. Entretanto, assumindo que T seja uma tabela esparsa (a maioria dos valores dos atributos é 0, vide Figura.16), então é esperado que o conjunto de itens freqüentes tenha tamanho máximo de k , onde $k \ll |T|$. Se esta expectativa for satisfeita, no pior caso a complexidade será (SIEBES, 2003) :

$$O\left(\sum_{j=1}^k \binom{|T|}{j}\right) = O(|T|^k) \ll O(2^{|T|})$$

Se considerarmos uma tabela densa, onde existam muitos conjuntos freqüentes longos (com muitos itens), a performance do *Apriori* é sensivelmente degradada. Isto ocorre porque são realizadas tantas passadas pela base de dados quanto é o tamanho do mais longo conjunto de itens freqüentes. Além disso, verificar um conjunto grande de candidatos é computacionalmente caro, principalmente para conjuntos longos. Quando a quantidade de itens dos conjuntos freqüentes é grande, o *Apriori* é degradado por causa do custo computacional e não pelo excesso de E/S.

Quanto menor for o suporte mínimo especificado, maiores são as chances de serem considerados freqüentes os conjuntos candidatos. Esse parâmetro é portanto também um fator para a complexidade do *Apriori* que pode se refletir tanto no custo de E/S quanto no custo de CPU. Pode aumentar a E/S, já que com um suporte pequeno conjuntos com quantidades maiores de itens poderão vir a se tornar freqüentes e exigir mais passadas pelo banco. Quanto ao custo de CPU, maior número de candidatos deverão ser testados.

4.2.3 Geração de Regras

A geração de regras se efetua a partir do conjunto de itens freqüentes. Nesta etapa, é utilizado o parâmetro de confiança mínima, que mede a força com que um grupo de itens depende de outro grupo. Na geração das regras, a base de dados não necessita ser percorrida, pois o cálculo da confiança da regra pode ser feito somente com o suporte do antecedente e do conseqüente da regra sendo analisada. Na fase de geração dos itens freqüentes, o suporte de todos os itens já é calculado.

Para a geração de regras a partir dos conjuntos de itens freqüentes selecionados pode ser utilizado o procedimento abaixo:

- 1) Para cada conjunto de itens frequente f de F faça
- 2) Para cada subconjunto não nulo s de f faça
- 3) $confiança(f) = (suporte(f) / suporte(f-s))$
- 4) Se $confiança(f) \geq confiança_mínima$

A regra $(f - s) \rightarrow s$ é válida com $confiança(f)$ e $suporte(f)$

Para cada um dos conjuntos de itens freqüentes, são extraídos os seus subconjuntos não nulos e a confiança de cada regra é calculada. A regra será selecionada se a confiança da mesma for maior ou igual à confiança mínima previamente estabelecida.

A complexidade da geração de regras é também exponencial, pois para cada f são considerados $2^{|f|} - 1$ subconjuntos não nulos de f . Entretanto, como é considerado que $|f| \leq k \ll |T|$, este não se torna um problema. Em alguns casos, somente as regras com apenas um conseqüente são geradas, o que torna a complexidade do algoritmo de geração de regras linear (SIEBES, 2003).

4.2.4 Regras Interessantes

Uma regra é dita interessante se ela proporciona alguma informação útil. Além do suporte e da confiança, existem diversas medidas para avaliar o interesse de uma regra: sustentação (*lift*), interesse (*interest*), convicção (*conviction*), força coletiva (*collective strength*), ganho (*gain*), entropia (*entropy*), cobertura (*coverage*), influência (*leverage*), etc. (SIEBES, 2003; JORGE, 2003; BAYARDO & AGRAWAL, 1999; WIZSOFT, 2003; MAGNUM, 2003). Tais medidas podem ser utilizadas para a decisão de quais regras devem ser mantidas e quais devem ser descartadas. Ao usuário, são exibidas apenas aquelas mais interessantes. Desta forma, reduz-se o tempo de análise das regras pelo usuário final.

O *lift* de uma regra de associação revela de que maneira uma regra prediz o conseqüente melhor que uma predição aleatória:

$$lift(X \rightarrow Y) = \frac{sup\ orste(XY) / sup\ orste(Y)}{sup\ orste(Y) / |D|} = confiança(X \rightarrow Y) \times \frac{|D|}{sup\ orste(Y)}$$

onde $|D|$ é o tamanho da base de dados, como $|D|$ refere-se à 100% da base de dados, consideramos que $|D|=1$.

A medida interesse (*interest*) de uma regra mede a dependência entre X e Y . É o quociente entre a probabilidade conjunta observada e a probabilidade conjunta sob independência. Na verdade, a medida interesse é igual ao *lift*:

$$\frac{P(X, Y)}{P(X) \times P(Y)} = \frac{sup\ orste(XY)}{sup\ orste(X) \times sup\ orste(Y)}$$

O interesse ou *lift* mínimo geralmente é igual a 1. Se a regra possui *lift* menor do que 1, deve ser descartada. Tomemos como exemplo o seguinte caso: 80% das pessoas compram leite, 2% das pessoas compram salmão e 1,5% das pessoas compram leite e salmão. O *lift* da regra *salmão* \rightarrow *leite* será $1,5\% / 80\% \times 2\% = 0,9375$. Esta regra pode ser descartada pois o *lift* é menor do que 1. Na verdade, leite não é dependente de salmão. O que ocorre é que geralmente as pessoas que levam salmão também levam leite, mas somente porque elas levam o leite na maioria das vezes, comprando ou não o salmão.

O interesse não mede se X causa Y . Para uma percepção deste tipo, utiliza-se a convicção (*conviction*) das regras que é dada por:

$$\frac{P(X) \times P(\neg Y)}{P(X, \neg Y)} = \frac{sup\ orste(X) \times (|D| - sup\ orste(Y))}{sup\ orste(X) - sup\ orste(XY)} = \frac{|D| - sup\ orste(Y)}{|D| (1 - confiança(X \rightarrow Y))}$$

ou

$$\frac{P(X) \times P(\neg Y)}{P(X, \neg Y)} = \frac{1}{interesse(X \rightarrow \neg Y)}$$

Quanto mais alta a convicção, mais freqüentemente Y ocorre junto de X . Quando for igual a 1, indica independência. Se possuir um valor muito alto, pode identificar regras pouco interessantes.

Vejamos outro exemplo. Se 1% da população é militar, se 50% das pessoas são adultas e se todos os militares devem ser adultos obrigatoriamente, então a regra *militar* \rightarrow *adulto* possui

$$\text{interesse} = \frac{1\%}{1\% \times 50\%} = 2$$

e

$$\text{convicção} = \frac{1}{\text{interesse}(\text{militar} \rightarrow \neg \text{adulto})} = \frac{1}{0} = \infty$$

A convicção é sensível à direção, já o interesse não é. A convicção de $X \rightarrow Y$ não é igual à convicção de $Y \rightarrow X$, mas o interesse de $X \rightarrow Y$ é igual ao interesse de $Y \rightarrow X$.

Após a geração de todas as regras de associação, utilizando as medidas de avaliação de interesse das regras, aquelas regras que não atingirem determinados limiares de interesses poderão ser eliminadas.

Além das medidas acima, consideradas medidas objetivas de interesse porque dependem apenas dos dados e parâmetros utilizados, existem as medidas subjetivas de interesse. As medidas subjetivas não podem ser utilizadas para filtrar automaticamente as regras interessantes após o processamento das mesmas, pois essas medidas dependem diretamente da avaliação do usuário final da aplicação de mineração de dados. Como são medidas subjetivas, uma regra pode ser interessante para um usuário e não interessante para outro. Do ponto de vista do usuário, uma regra é interessante se ela possui um grau elevado de utilidade ou de inesperabilidade. Em relação à utilidade, um padrão é interessante se o usuário pode construir algo, tomar decisões a partir dele. A inesperabilidade, por sua vez, deve auxiliar na descoberta de regras surpreendentes, deve ser capaz de contradizer as expectativas do usuário (BRUSSO, 2000).

Capítulo 05

CONSTRUÇÃO DO DATA WAREHOUSE DE ICTIOFAUNA E GERAÇÃO DAS REGRAS DE ASSOCIAÇÃO

No presente capítulo, serão apresentados os passos para a construção do Data Warehouse para o tema Ictiofauna e as Regras de Associação encontradas. Também será feita a análise dos respectivos resultados.

5.1 Construção do Data Warehouse de Ictiofauna

Para construção do Data Warehouse para o tema Ictiofauna do projeto PIATAM, foram necessárias a realização de etapas que envolviam, por exemplo, a construção de um banco de dados relacional e a definição dos fatos relevantes do negócio.

Para um melhor entendimento, estas etapas serão descritas através de um fluxograma (Figura 5.1) que mostrará a ordem em que elas foram executadas.

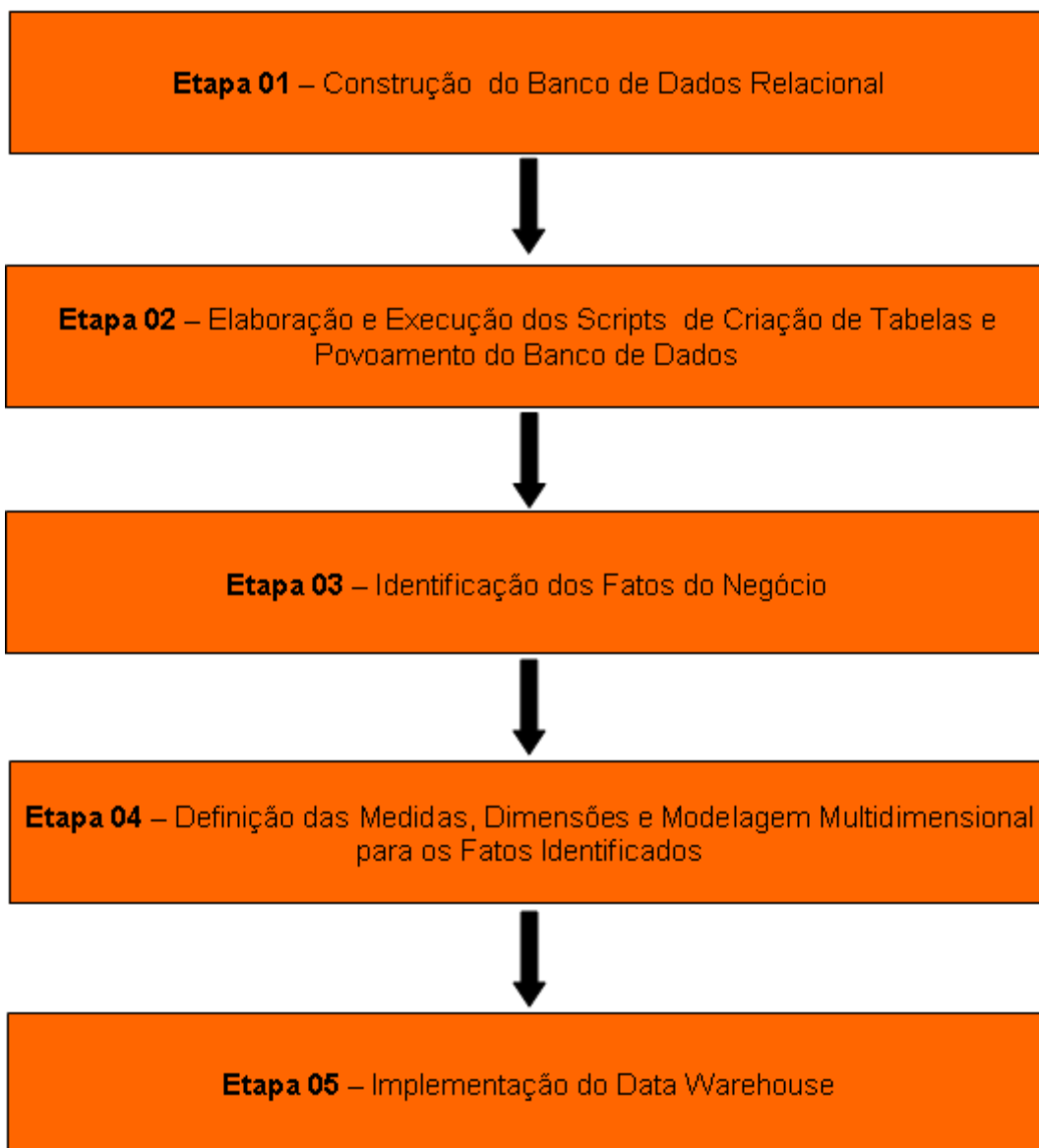


Figura 5.1 – Fluxograma de Procedimentos para a Construção do Data Warehouse para o Tema Ictiofauna.

5.1.1 Construção do Banco de Dados Relacional

Para a criação do banco de dados relacional, foi feita a modelagem ER (Entidade-Relacionamento) com ajuda de uma Ferramenta Case. Esta modelagem do banco de dados relacional foi desenvolvida utilizando o mesmo modelo de dados do Relatório de Manutenção do Modelo de Dados da Base de Dados Integrada do PIATAM (versão ATECH.439.03.0002/A).

A figura 5.2 apresenta a modelagem do Banco de Dados Relacional.

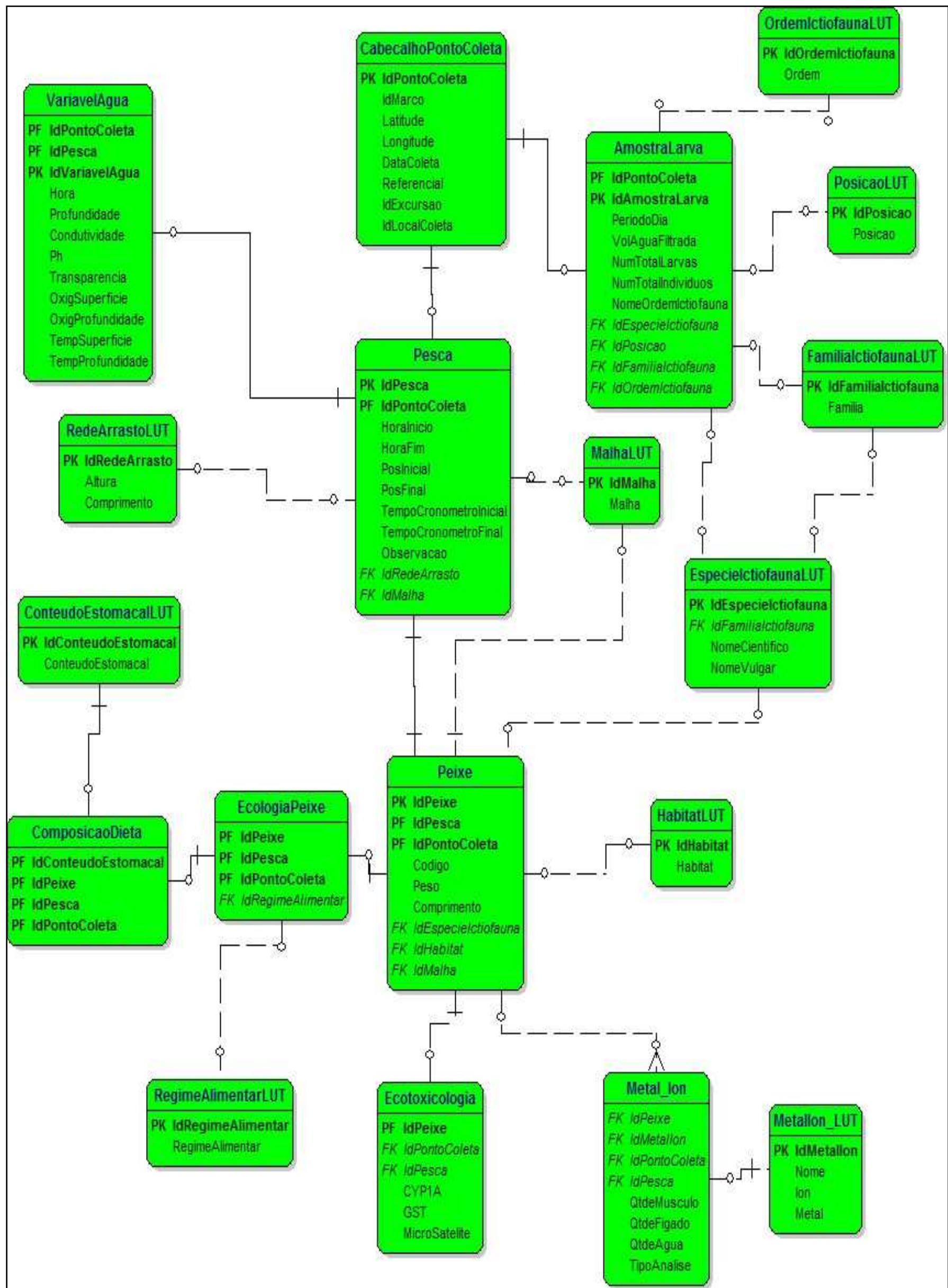


Figura 5.2 – Modelagem ER do Banco de Dados Relacional. Fonte: Adaptada do Relatório de Manutenção do Modelo de Dados da Base de Dados Integrada do PIATAM (versão ATECH.439.03.0002/A).

O próximo passo para a criação do Banco de Dados Relacional, foi a criação de *scripts* que, tendo como base a modelagem ER criada, adicionassem as tabelas ao Sistema Gerenciador de Banco de Dados (SGBD) escolhido para o trabalho (*Microsoft SQL Server 2005*), obedecendo os relacionamentos e atributos que os dados traziam da BDI PIATAM.

A figura 5.3 apresenta um exemplo de script de inclusão de dados no SGBD. Através deste script são criadas as tabelas no Banco de Dados.

```
CREATE TABLE VariavelAgua (
  IdPontoColeta NUMERIC(11) NOT NULL,
  IdVariavelAgua NUMERIC(11) NOT NULL,
  Hora DATE,
  Profundidade NUMERIC(17,2),
  Condutividade NUMERIC(17,2),
  Ph NUMERIC(17,2),
  Transparencia NUMERIC(17,2),
  OxigSuperficie NUMERIC(17,2),
  OxigProfundidade NUMERIC(17,2),
  TempSuperficie NUMERIC(17,2),
  TempProfundidade NUMERIC(17,2),
  PRIMARY KEY (IdPontoColeta, IdVariavelAgua)
);

CREATE TABLE Pesca (
  IdPontoColeta NUMERIC(11) NOT NULL,
  IdPesca NUMERIC(11) NOT NULL,
  HoraInicio DATE,
  HoraFim DATE,
  PosInicial VARCHAR(50),
  PosFinal VARCHAR(50),
  TempoCronometroInicial VARCHAR(50),
  TempoCronometroFinal VARCHAR(50),
  Observacao VARCHAR(300),
  IdRedeArrasto NUMERIC(11),
  IdMalha NUMERIC(11),
  PRIMARY KEY (IdPontoColeta)
);

CREATE TABLE RedeArrastoLUT (
  IdRedeArrasto NUMERIC(11) NOT NULL,
  Altura NUMERIC(10,2),
  Comprimento NUMERIC(10,2),
  PRIMARY KEY (IdRedeArrasto)
);

CREATE TABLE MalhaLUT (
  IdMalha NUMERIC(11) NOT NULL,
  Malha VARCHAR(20),
  PRIMARY KEY (IdMalha)
);
```

Figura 5.3 – Exemplo de Script de Criação de Tabelas no SGBD.

Os dados de Ictiofauna provenientes da BDI PIATAM foram adquiridos em formato texto (*flat files*). Na figura 5.4 é apresentada a estrutura de um destes arquivos.

```

1380:32:Geophagus altifrons (Heckel,1840);
1381:34:Henmiodus immaculatus (Kner,1858);
1382:15:Hoplerythrinus unitaeniatus (Spix & Agassiz,1829);
1406:32:Satonoperca sp.;Curimatã
1384:30:Loricariichthys maculatus (Bloch,1794);
1385:30:Loricariichthys nudirostris (Kner,1858);
1387:31:Plagioscion auratus (Castelnau,1855);
1388:35:Prochilodus nigricans (Spix & Agassiz,1829);
1390:23:Pterodoras lentiginosus (Eigenmann,1917);
1395:19:Serrasalmus robertsoni;
1397:19:Astyanax sp.;
1398:32:Cichla sp.;
1399:19:Mylesinus sp.;
1400:19:Mylossoma sp.;
1401:25:Pinelodella sp.;
1402:30:Pterygoplichthys sp.;
13:11:Leporinus affinis (Günther,1864);Aracú
14:11:Leporinus fasciatus (Bloch,1794);Aracú amarelo, aracu flamengo
15:11:Leporinus friderici (Bloch,1794);Aracú cabeça-gorda, Aracú comum
16:11:Leporinus trifasciatus (Steindachner,1876);Aracú cabeça gorda
17:11:Schizodon fasciatus (Spix & Agassiz,1829);Aracú-comum
18:15:Erythrinus erythrinus (Bloch & Schneider,1801);Jejú
19:15:Hoplias malabaricus (Bloch,1794);Traira
20:19:Acestrorhynchus falcatus (Bloch,1794);Peixe cachorro
21:19:Brycon cephalus (Günther,1869);Matrinxã
22:19:Triportheus auritus (Valenciennes,1840);Sardinha comprida
23:37:Raphiodon vulpinus (Spix & Agassiz,1829);Peixe-cachorro
24:35:Semaprochilodus insignis (Jardine & Schomburgk,1841);Jaraqui-escama-grossa
25:35:Semaprochilodus taeniurus (Valenciennes,1817);Jaraqui-escama-fina
26:10:Colossoma macropomum (Cuvier,1818);Tambaqui, ruelo
27:10:Metynnis argenteus (Ahl,1923);Pacú
28:10:Mylossoma duriventre (Cuvier,1818);Pacú comum
29:10:Piaractus brachyomus (Cuvier,1818);Pirapitinga
30:10:Pygocentrus nattereri (Kner,1858);Piranha-cajú
31:10:Serrasalmus elongatus (Kner,1858);Piranha-mucura
32:10:Serrasalmus rhombeus (Linnaeus,1766);Piranha-preta
33:10:Serrasalmus spilopleura (Kner,1858);Piranha-tucupi, Piranha-amarela
34:10:Pristobrycon calmoni (Steindachner,1908);Piranha

```

Figura 5.4 – Estrutura dos Flat Files

Para a etapa de povoamento do banco de dados, scripts de inclusão de dados no SGBD foram desenvolvidos e executados. Podemos visualizar um exemplo deste script na figura 5.5.

```

BULK
INSERT RegimeAlimentarLUT
FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de Dados/DadosRicardo/CSVs/Ictiofauna/ict_regimealimentarlut.csv'
WITH
(
    FIELDTERMINATOR = ';',
    ROWTERMINATOR = '\n'
)
go

BULK
INSERT HabitatLUT
FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de Dados/DadosRicardo/CSVs/Ictiofauna/ict_habitatlut.csv'
WITH
(
    FIELDTERMINATOR = ';',
    ROWTERMINATOR = '\n'
)
go

BULK
INSERT MalhaLUT
FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de Dados/DadosRicardo/CSVs/Ictiofauna/ict_malhaLUT.csv'
WITH
(
    FIELDTERMINATOR = ';',
    ROWTERMINATOR = '\n'
)
go

BULK
INSERT OrdemIctiofaunaLUT
FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de Dados/DadosRicardo/CSVs/Ictiofauna/ict_ordemictiofaunalut.csv'
WITH
(
    FIELDTERMINATOR = ';',
    ROWTERMINATOR = '\n'
)

```

Figura 5.5 – Exemplo de Script de Inclusão de Dados no SGBD.

Após a execução do *script* de inclusão de dados, o banco de dados foi povoado, como podemos ver no exemplo da figura 5.6.

ALDEBARAN.Icti...na - dbo.Peixe		ALDEBARAN.Icti...eIctiofaunaLUT		Object Explorer Details					
IdPeixe	IdPesca	IdPontoColeta	Codigo	Peso	Comprimento	IdEspecieIctiof...	IdHabitat	IdMalha	
8288	8192	24973	5389	370.00	37.50	10	1	34	
8289	8192	24973	5390	735.00	30.00	87	1	34	
8290	8192	24973	5391	650.00	26.50	26	3	34	
8291	8192	24973	5392	500.00	24.50	26	3	34	
8292	8192	24973	5393	240.00	23.00	1363	2	31	
8293	8192	24973	5394	290.00	21.00	53	3	31	
8294	8192	24973	5395	190.00	18.50	53	3	31	
8295	8192	24973	5396	280.00	22.00	87	1	31	
8296	8192	24973	5397	85.00	12.50	33	1	31	
8297	8192	24973	5398	60.00	12.00	33	1	31	
8298	8192	24973	5399	75.00	11.50	33	1	31	
8299	8192	24973	5400	75.00	12.50	33	1	31	
8300	8192	24973	5401	50.00	12.00	34	1	31	
8301	8192	24973	5402	120.00	18.00	69	1	31	
8302	8192	24973	5403	125.00	19.00	69	1	31	
8303	8192	24973	5406	445.00	21.00	61	1	31	
8304	8192	24973	5407	265.00	17.00	61	1	31	
8305	8192	24973	5408	450.00	28.50	1363	2	31	
8306	8192	24973	5409	365.00	26.50	1363	2	31	
8307	8192	24973	5410	1255.00	58.00	10	1	31	
8308	8192	24973	5411	1085.00	54.50	10	1	31	
8309	8192	24973	5412	1070.00	52.00	10	1	31	
8310	8192	24973	5413	85.00	12.50	33	1	31	
8311	8192	24973	5414	85.00	15.00	34	1	31	
8312	8192	24973	5415	100.00	15.00	34	1	31	
8313	8192	24973	5416	180.00	15.50	30	3	31	
8314	8192	24973	5417	205.00	16.00	30	3	31	
8315	8192	24973	5418	150.00	15.50	30	3	31	
8316	8192	24973	5419	155.00	15.50	30	3	31	
8317	8192	24973	5420	125.00	15.00	30	3	31	
8318	8192	24973	5421	190.00	16.00	30	3	31	
8319	8192	24973	5422	175.00	16.00	30	3	31	

Figura 5.6 – Banco de Dados Povoado.

Após o povoamento do banco de dados, foi possível ter uma melhor visualização dos dados disponíveis. A tabela 03 mostra as tabelas do Banco de Dados assim como a quantidade de registros para cada tabela.

TABELA 5.1 – Tabelas e Registros do Banco de Dados

<i>Tabela</i>	<i>Registros</i>
AmostraLarva	16.383
CabecalhoPontoColeta	24.921
ComposicaoDieta	20.254
ConteudoEstomacalLUT	55
EcologiaPeixe	20.158
Ecotoxicologia	1.021
EspecielctiofaunaLUT	244
FamilialctiofaunaLUT	32
HabitatLUT	5
MalhaLUT	13
Metal_Ion	91
Metallom_LUT	4
OrdemlctiofaunaLUT	10
Peixe	16.383
Pesca	204
PosicaoLUT	12
RedeArrastoLUT	2
RegimeAlimentarLUT	13
VariavelAgua	184

A descrição de cada uma das tabelas pode ser encontrada no Anexo 1.

5.1.2 Construção do Data Warehouse

Após a construção e povoamento do Banco de Dados Relacional, podemos iniciar a construção do Data Warehouse para o tema Ictiofauna.

O primeiro passo para a criação deste repositório de dados foi a identificação dos fatos do negócio para, em seguida, definirmos as medidas e dimensões para cada fato identificado, além da construção dos modelos multidimensionais.

5.1.2.1 Fato Pesca

Este fato objetiva a manipulação de dados referentes às pescas realizadas para estudos do Projeto PIATAM.

- **Medidas**

As medidas consideradas para o Fato Pesca foram:

- Freqüência da Pesca – Esta medida corresponde a quantidade de pescas realizadas;
- Localização da Pesca – Indica os locais onde as pescas foram realizadas;
- Data da Pesca – Mostra a data de realização da pesca;
- Média do PH – Corresponde a média do Ph da água onde as pescas foram realizadas;
- Média do Oxigênio na Superfície – Mostra a média do Oxigênio na Superfície onde foram realizadas as pescas;
- Média do Oxigênio na Profundidade – Indica a média do Oxigênio na Profundidade onde as pescas foram realizadas;
- Média da Temperatura na Superfície – Corresponde a média da Temperatura na Superfície no local da pesca;
- Média da Temperatura na Profundidade – Mostra a média da Temperatura na Profundidade onde a a pesca se realizou.

Apesar de implementadas no modelo multidimensional, medidas referentes à “profundidade da água”, “condutividade da água”, “transparência da água”, “tipo de rede utilizada na pesca” e “tipo de malha utilizada na pesca” não foram consideradas, pois estas informações não possuíam registros significativos na época em que os dados foram adquiridos.

• Dimensões

As dimensões consideradas para o Fato Pesca foram:

- Ponto de Coleta – Com esta dimensão, é possível visualizar os pontos de coleta onde as pescas foram realizadas;
- Marco PIATAM – Indica o marco PIATAM quando a pesca foi realizada;
- Excursão – Através desta dimensão, podemos visualizar as informações referentes à excursão, além dos períodos hidrológicos que a pesca foi realizada (Seca, Enchente, Cheia ou Vazante);
- Variável da Água – Esta dimensão permite a visualização dos dados de acordo com o tipo de variável da água escolhido.

• Modelo Multidimensional

O tipo de modelo multidimensional utilizado para a modelagem do Data Warehouse proposto neste estudo é o modelo estrela. No modelo estrela, todas as tabelas relacionam-se diretamente com a tabela de fatos, sendo assim, as tabelas dimensionais devem conter todas as descrições que são necessárias para definir uma classe.

Podemos visualizar na (Figura 5.7) que a tabela Fato_Pesca se relaciona com as Tabelas de Dimensão: Dim_Marco, Dim_PontoColeta, Dim_Excursão e Dim_VariavelAgua.

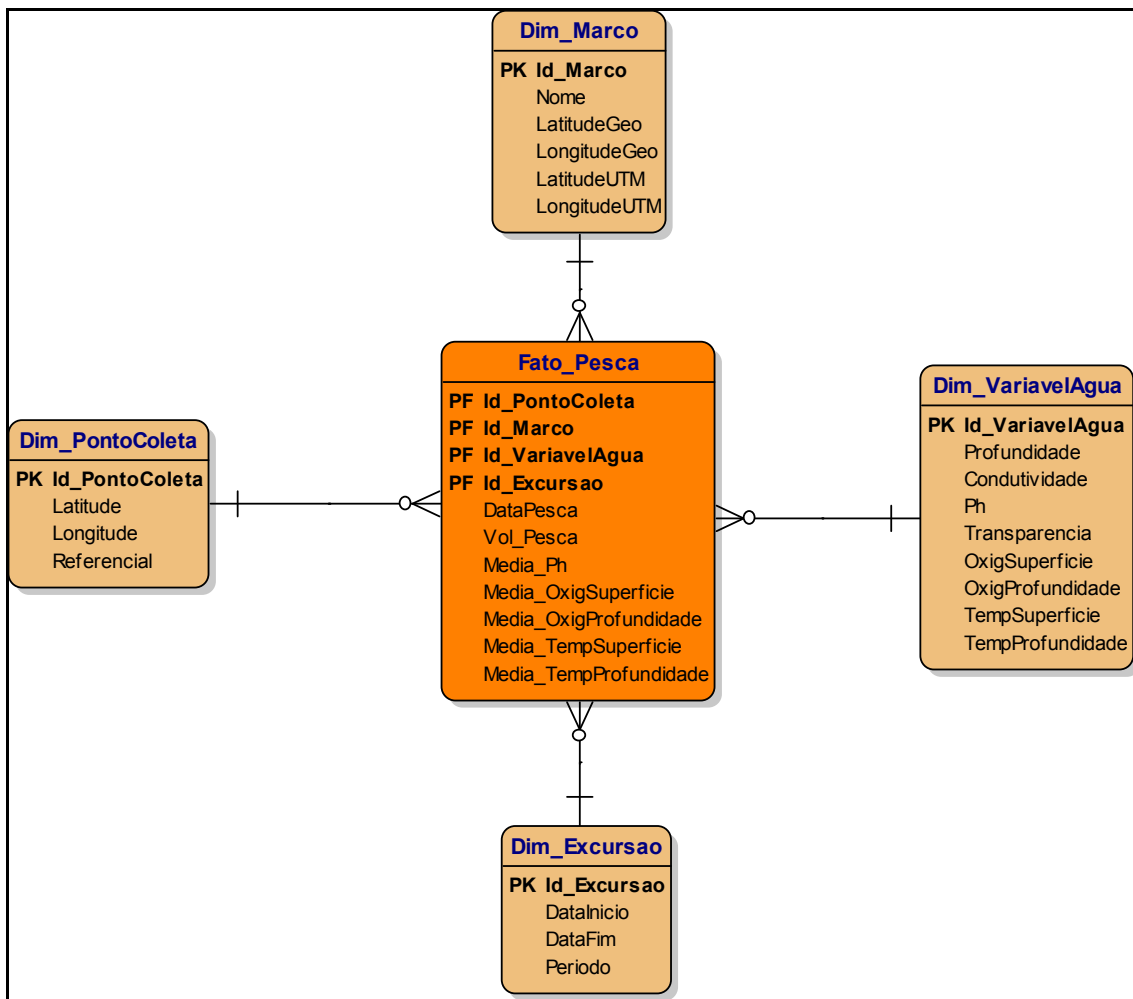


Figura 5.7 – Modelo Multidimensional do Fato Pesca.

5.1.2.2 Fato Peixe

Este fato objetiva a manipulação de dados referentes aos peixes coletados.

- **Medidas**

As medidas consideradas para o Fato Peixe foram:

- Localização Coleta – Esta medida corresponde ao local onde o peixe foi coletado;

- Data da Coleta – Mostra a data de realização da coleta do peixe;
- Comprimento Máximo – Indica o comprimento máximo de um peixe coletado;
- Comprimento Mínimo – Mostra o comprimento mínimo de um peixe coletado;
- Média do Comprimento – Corresponde à média do comprimento de um peixe;
- Peso Máximo – Indica o peso máximo dentre os peixes coletados;
- Peso Mínimo – Corresponde ao peso mínimo dentre os peixes coletados;
- Média do Peso – Apresenta a média do peso dos peixes coletados.

• Dimensões

As dimensões consideradas para o Fato Peixe foram:

- Ponto de Coleta – Com esta dimensão, é possível visualizar os pontos de coleta dos peixes;
- Marco PIATAM – Indica o marco PIATAM quando a coleta do peixe foi realizada;
- Excursão – Através desta dimensão, podemos visualizar as informações referentes à excursão, além dos períodos hidrológicos que o peixe foi coletado (Seca, Enchente, Cheia ou Vazante);
- Nome Científico – Utilizando esta dimensão, é possível visualizar o nome científico do peixe;
- Nome Vulgar – Visualizamos o nome vulgar de uma espécie;
- Família – É mostrada a família do peixe coletado;

- Espécie – Pode ser visualizada a espécie de um peixe coletado.

- **Modelo Multidimensional**

O modelo estrela sugerido para o Fato peixe está apresentado na próxima figura (Figura 5.8).

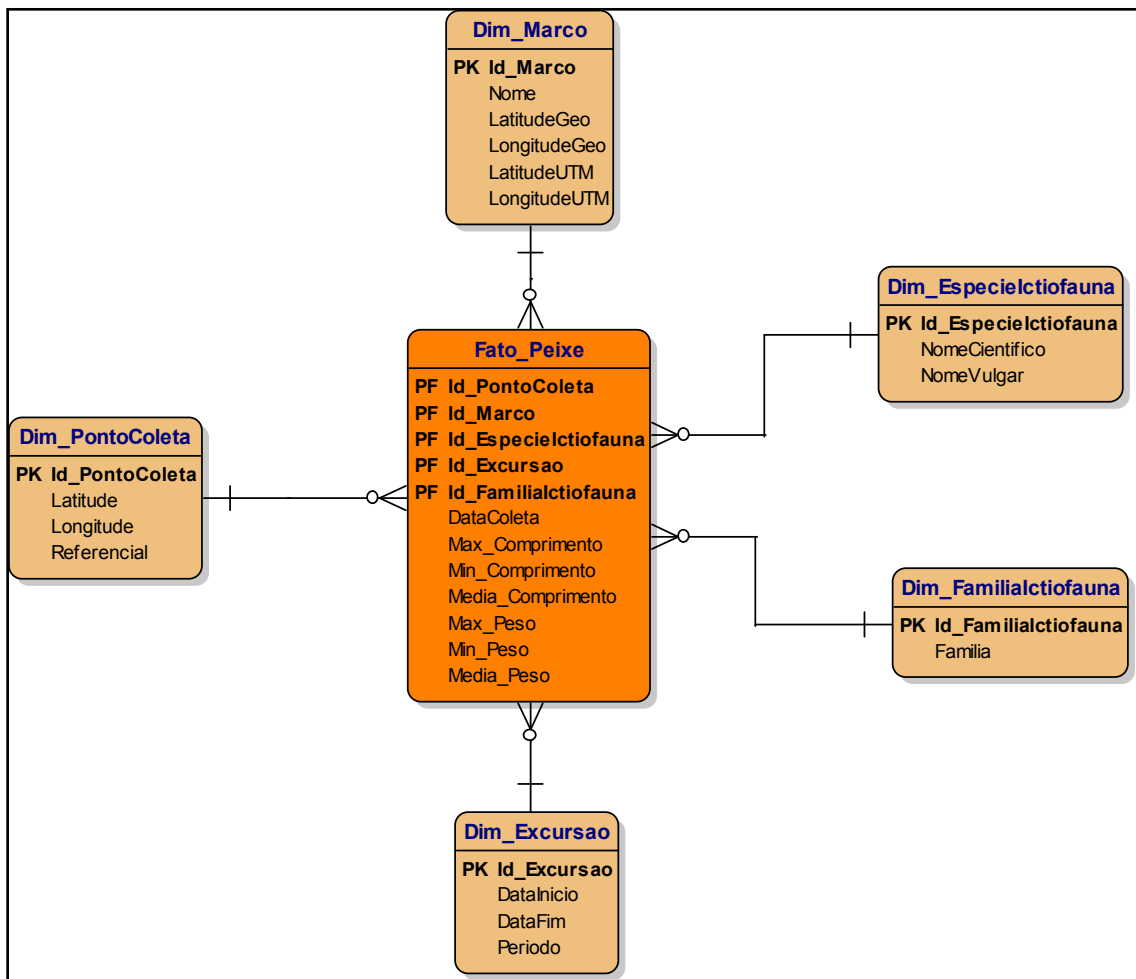


Figura 5.8 – Modelo Multidimensional do Fato Peixe.

Neste modelo, a tabela Fato_Peixe se relaciona com as Tabelas de Dimensão: Dim_Marco, Dim_PontoColeta, Dim_Excursoão, Dim_Familialctiofauna e Dim_Especielctiofauna.

5.1.2.3 Fato Amostra Larva

Este fato objetiva a manipulação de dados referentes as amostras de larvas de peixe coletadas pelo PIATAM.

- **Medidas**

As medidas consideradas para o Fato Amostra Larva foram:

- Localização Coleta – Esta medida corresponde ao local onde a amostra foi coletada;
- Data da Coleta – Mostra a data de realização da coleta da amostra;
- Volume de Água Filtrada – Indica o volume de água filtrada (em metros cúbicos) em uma amostra;
- Média do Volume de Água Filtrada – Corresponde à média do volume de água filtrada (em metros cúbicos) em uma amostra;
- Número Total de Larvas – Indica o número total de larvas coletadas em uma amostra;
- Média do Número Total de Larvas – Podemos visualizar a média do número total de larvas em uma amostra.

- **Dimensões**

As dimensões consideradas para o Fato Amostra Larva foram:

- Ponto de Coleta – Com esta dimensão, é possível visualizar os pontos de coleta onde as amostras de larvas foram coletadas;

- Marco PIATAM – Indica o marco PIATAM quando a coleta da amostra de larva foi realizada;
- Excursão – Através desta dimensão, podemos visualizar as informações referentes à excursão, além dos períodos hidrológicos que a amostra de larva foi coletada (Seca, Enchente, Cheia ou Vazante);
- Nome da Ordem da Larva – Utilizando esta dimensão, é possível visualizar o nome da ordem da larva. Estas possuem os seguintes domínios: Perciforme, Siluriforme, Characiforme, Osteoglossiforme, Beloniforme, Clupeiforme, Gymnotiforme, Tetraodontiforme, Cypriniforme, Pleuronectiforme;
- Posição – Refere-se a posição onde a larva foi coletada. Possui os seguintes valores: RCS (rio/centro/superfície), RCF (rio/centro/fundo), RMS (rio/margem/superfície), RMF (rio/margem/fundo), LCS (lago/centro/superfície), LCF (lago/centro/fundo), LMS (lago/margem/superfície), LMF (lago/margem/fundo);
- Família – É mostrada a família da larva coletada;
- Espécie – Pode ser visualizada a espécie de uma larva coletada.

• Modelo Multidimensional

Podemos verificar que na modelagem multidimensional do Fato Amostra Larva (Figura 5.9), a tabela Fato_AmostraLarva se relaciona com as Tabelas de Dimensão: Dim_Marco, Dim_PontoColeta, Dim_Excursão, Dim_PosiçãoIctiofauna, DimOrdemIctiofauna, DimFamiliarIctiofauna e Dim_EspecieIctiofauna.

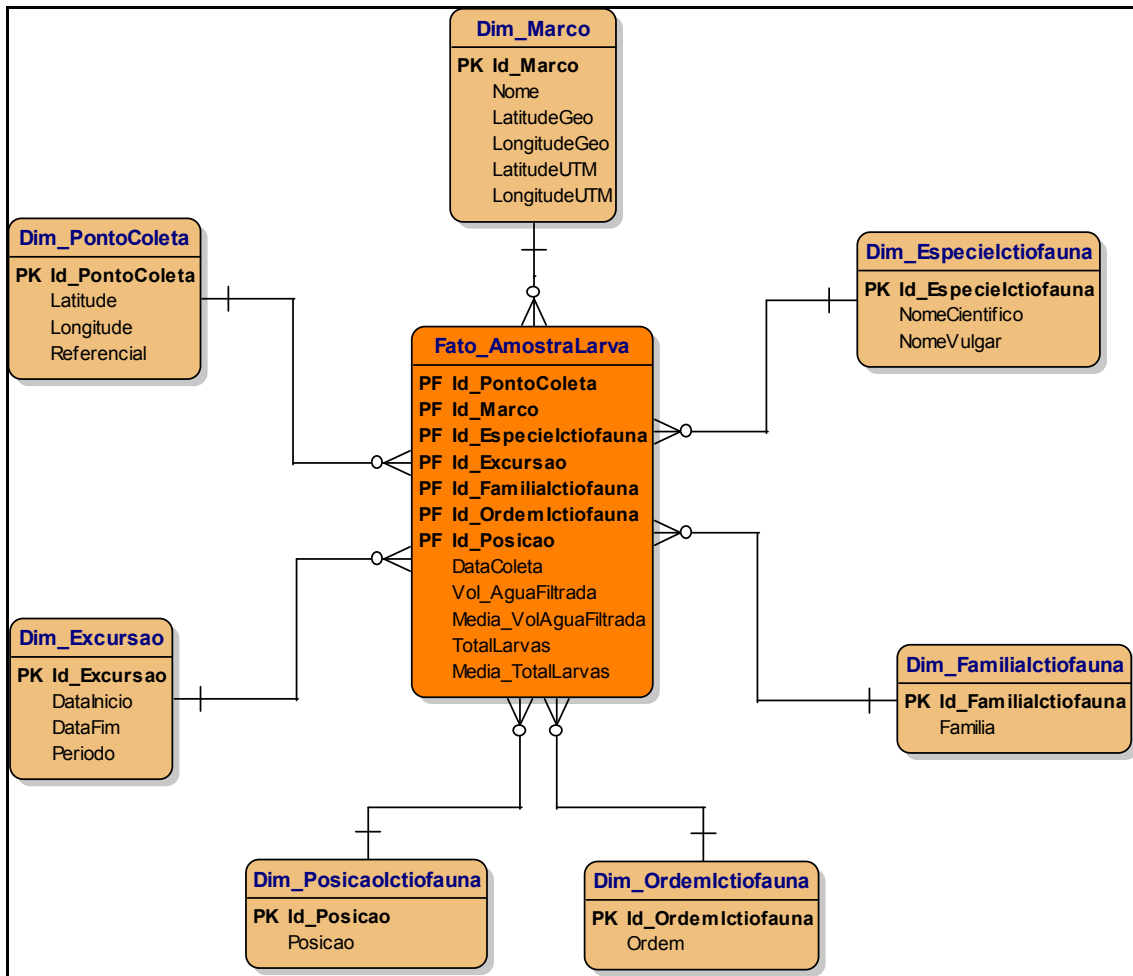


Figura 5.9 – Modelo Multidimensional do Fato Amostra Larva

5.1.3 Implementação do Data Warehouse

O Data Warehouse foi implementado através do *software* SQL Server *Business Intelligence Development Studio*, que provê uma plataforma de desenvolvimento integrada para desenvolvedores de sistemas de *Business Intelligence*.

Podemos destacar as seguintes características:

- Permite a execução de integrações complexas de dados, transformações e sínteses em uma alta velocidade para grandes volumes de dados;
- Facilita a manipulação entre dados relacionais e dados multidimensionais;

- Possui funcionalidades analíticas avançadas que auxiliam a manipulação dos cubos multidimensionais;
- Permite a geração de relatórios personalizados e entregues em diferentes formatos e com várias formas de interatividade e opções de impressão.

Podemos visualizar abaixo (Figura 5.10) uma imagem do ambiente de desenvolvimento do *SQL Server Business Intelligence Development Studio*.

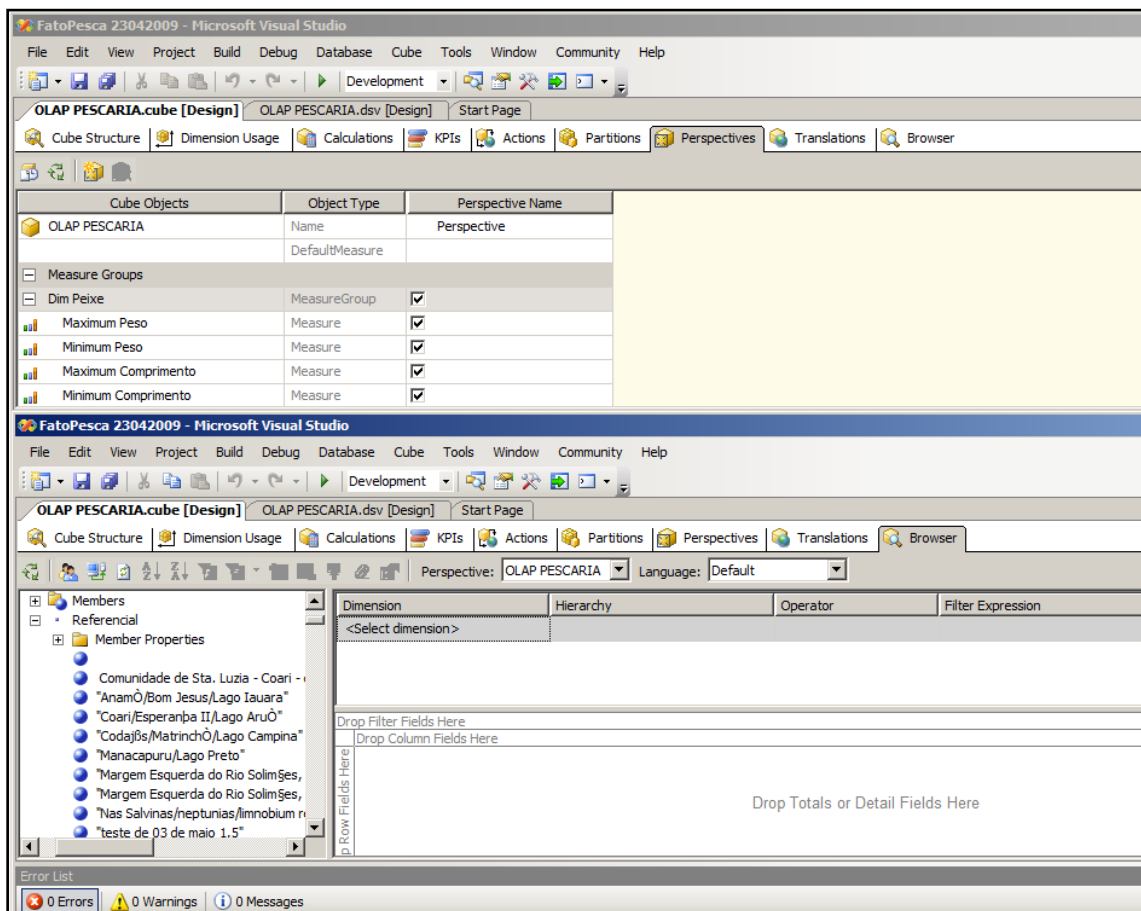


Figura 5.10 – Ambiente de Desenvolvimento do *SQL Server Business Intelligence Development Studio*.

5.1.4 Análise dos Resultados Obtidos com a Construção do Data Warehouse

Conforme mencionado no item 3.3 do Capítulo 03, o Data Warehouse é, basicamente, uma ferramenta que visa o armazenamento de dados, porém, não dispõe de recursos para realizar consultas e análises sofisticadas ou cálculos mais complexos.

Existem várias maneiras de recuperar informações de um Data Warehouse, por exemplo, o uso de ferramentas de consulta e emissão de relatórios ou a utilização de um EIS - Executive Information Systems (Sistema de Informação Executiva).

Para as consultas realizadas neste trabalho, foi utilizada a ferramenta Microsoft SQL Server *Analysis Services* (SSAS), que é uma ferramenta de processamento analítico (OLAP) embutida no SQL Server *Business Intelligence Development Studio*.

Na consulta realizada abaixo (Figura 5.11), foi solicitada a pergunta “Qual a quantidade de coletas realizadas por Local de Coleta nos últimos 03 anos?”

LOCAL COLETA	2006	2007	2008	TOTAL
Comunidade Santa Luzia do Baixio	0	0	1	1
Lg Anamá	4	3	1	8
Lg Araçá	4	2	0	6
Lg Aruã	4	2	1	7
Lg Baixio	4	4	0	8
Lg Campina	0	2	0	2
Lg Iauara	4	2	1	7
Lg Maracá	4	4	1	9
Lg Poraquê	4	3	1	8
Lg Preto	4	4	1	9
Paraná do Iauara	0	1	0	1
Rio Urucu	0	2	0	2
TOTAL	32	29	7	68

Figura 5.11 – Resposta da Consulta “Qual a quantidade de coletas realizadas por Local de Coleta nos últimos 03 anos?”.

A consulta retornou a quantidade de pescas solicitadas, assim como os totais de pescas realizadas por cada Local de Coleta e por ano.

Através das Ferramentas OLAP, podemos realizar análises de várias perspectivas. Em uma outra consulta, podemos obter um resultado de outro ponto de vista.

LOCAL COLETA	2007												TOTAL
	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	
Lg Anamá	0	0	1	0	0	1	0	0	1	0	0	0	3
Lg Araçá	0	0	0	0	0	0	0	0	1	0	0	1	2
Lg Aruã	0	0	1	0	0	0	0	0	1	0	0	0	2
Lg Baixio	0	0	1	0	0	1	0	0	1	0	0	1	4
Lg Campina	0	0	1	0	0	1	0	0	0	0	0	0	2
Lg Iauara	0	0	0	0	0	1	0	0	1	0	0	0	2
Lg Maracá	0	0	1	0	0	1	0	0	1	0	0	1	4
Lg Poraquê	0	0	1	0	0	1	0	0	1	0	0	0	3
Lg Preto	0	0	1	0	0	1	0	0	1	0	0	1	4
Paraná do Iauara	0	0	1	0	0	0	0	0	0	0	0	0	1
Rio Urucu	0	0	1	0	0	0	0	0	1	0	0	0	2
TOTAL	0	0	9	0	0	7	0	0	9	0	0	4	29

Figura 5.12 – Resposta da Consulta “Qual a quantidade de coletas realizadas em cada mês de 2007?”.

Na consulta acima (Figura 5.12), podemos visualizar as coletas realizadas em cada mês do ano de 2007. A função utilizada para a realização desta consulta foi o *drill down* (item 3.3.3 do Capítulo 03), que consiste em fazer uma exploração em diferentes níveis de detalhe das informações. Neste caso, partimos de um nível “por ano” para um nível “por mês”.

Foi solicitada a pergunta “Qual a quantidade de pescas realizadas por Local de Pesca nos últimos 03 anos?” para a consulta abaixo (Figura 5.13).

LOCAL PESCA	2006	2007	2008	TOTAL
Comunidade Santa Luzia do Baixio	0	0	66	66
Lg Anamá	1161	1095	167	2423
Lg Araçá	1463	1291	424	3178
Lg Aruã	434	122	25	581
Lg Baixio	340	397	58	795
Lg Campina	0	204	0	204
Lg Iauara	738	386	129	1253
Lg Maracá	851	600	150	1601
Lg Poraquê	145	400	50	595
Lg Preto	1403	529	235	2167
Paraná do Iauara	0	80	0	80
Rio Urucú	0	147	0	147
TOTAL	6535	5251	1304	13090

Figura 5.13 – Resposta da Consulta “Qual a quantidade de pescas realizadas por Local de Pesca nos últimos 03 anos?”.

Como podemos ver, a resposta foi obtida assim, como os totais de pescas realizadas por ano e por localidade.

Na próxima consulta (Figura 5.14), foram solicitados os peixes (Nome Científico e Nome Vulgar) coletados nas pescas realizadas na localidade do Paraná do Iauara em 2007, assim como as seguintes medidas: Quantidade, Comprimento Mínimo, Comprimento Máximo, Média do Comprimento, Peso Mínimo, Peso Máximo e Média do Peso.

NOME CIENTÍFICO	NOME VULGAR	QTDE	COMPRIMENTO MÍNIMO	COMPRIMENTO MÁXIMO	MÉDIA DO COMPRIMENTO	PESO MÍNIMO	PESO MÁXIMO	MÉDIA DO PESO
Ageneiosus brevifilis (Valenciennes, 1840)	Mandubé, palmito, fidalgo	1	28.50	28.50	28.50	440.00	440.00	440.00
Ageneiosus sp	Mandubé	1	11.50	11.50	11.50	40.00	40.00	40.00
Anodus elongatus (Agassiz, 1829)	Cubiu	1	19.50	19.50	19.50	125.00	125.00	125.00
Auchenipterus nuchalis (Spix & Agassiz, 1829)	Mandi peruano	10	14.00	18.00	15.20	40.00	85.00	53.00
Loricaria cataphracta (Linnaeus, 1758)	Acarí	1	25.50	25.50	25.50	130.00	130.00	130.00
Mylossoma aureum (Agassiz, 1829)	Pacú manteiga	1	13.50	13.50	13.50	65.00	65.00	65.00
Não Disponível	Não Disponível	1	13.00	13.00	13.00	55.00	55.00	55.00
Pellona castelnaeana (Valenciennes, 1847)	Apapá amarelo, sardinhão	7	16.50	23.00	19.71	19.00	140.00	94.14
Pinirampus pirinampu (Spix & Agassiz, 1829)	Piranambú, barbado, barba-chata	4	18.00	35.00	26.25	85.00	766.00	352.75
Platynemathichthys notatus (Jardine & Schomburgk, 1841)	Cara de gato	2	26.00	28.00	27.00	310.00	390.00	350.00
Potamorhina altamazonica (Cope, 1878)	Branquinha-cabeça-lisa	10	11.50	18.50	16.05	55.00	180.00	134.50
Potamorhina latior (Spix & Agassiz, 1829)	Branquinha-comum	1	15.50	15.50	15.50	75.00	75.00	75.00
Pterophyllum scalare (Valenciennes, 1823)	Acará-bandeira	2	6.00	6.50	6.25	25.00	25.00	25.00
Pygocentrus nattereri (Kner, 1858)	Piranha-cajú	18	12.50	15.00	13.94	95.00	185.00	139.17
Rhytiodus microlepis (Kner, 1858)	Aracu	6	18.00	26.00	21.33	95.00	210.00	143.33
Schizodon fasciatus (Spix & Agassiz, 1829)	Aracú-comum	1	19.50	19.50	19.50	165.00	165.00	165.00
Serrasalmus elongatus (Kner, 1858)	Piranha-mucura	3	12.00	18.50	14.33	40.00	55.00	45.00
Serrasalmus rhombeus (Linnaeus, 1766)	Piranha-preta	6	8.50	13.50	10.25	20.00	100.00	40.00
Serrasalmus sp.	---	2	8.00	11.50	9.75	25.00	55.00	40.00
Triportheus albus (Cope, 1872)	Sardinha comum	1	11.50	11.50	11.50	30.00	30.00	30.00
Triportheus auritus (Valenciennes, 1840)	Sardinha comprida	1	19.00	19.00	19.00	80.00	80.00	80.00
TOTAL		80	---	---	---	---	---	---

Figura 5.14 – Resposta da Consulta sobre as medidas dos peixes (Nome Científico e Nome Vulgar) coletados nas pescas realizadas na localidade do Paraná do Iauara em 2007.

Como último exemplo, a figura abaixo (Figura 5.15) mostra o resultado da consulta para a pergunta “Qual a quantidade de larvas coletadas por localidade no ano de 2007 ?”

LOCAL COLETA	QUANTIDADE	TOTAL LARVAS COLETADAS
Canal do Rio Solimões	155	19671
Paraná Baixo	241	246503
Paraná do laudara	113	5204
TOTAL	509	271378

Figura 5.15 – Resposta da Consulta “Qual a quantidade de larvas coletadas por localidade no ano de 2007 ?”.

Através destes exemplos, podemos constatar que as ferramentas OLAP poderiam auxiliar analistas, gestores e colaboradores a sintetizarem informações sobre os dados de Ictiofauna, através de comparações, análises históricas, etc.

Em aplicações de análise de dados, um dos fatores mais críticos é o tempo de resposta ao usuário devido ao grande volume de dados envolvido nas consultas desse tipo de aplicação. A maneira mais usada para reduzir o tempo de execução das consultas de forma consistente é agregando os dados em totais e subtotais, através das dimensões envolvidas no assunto em questão.

Podemos agrupar cidades em estados, regiões e países, meses em trimestres e anos, etc. Apesar das hierarquias não serem partes necessárias das dimensões, as aplicações que refletem negócios do mundo real com um mínimo de complexidade, geralmente, apresentam algumas hierarquias dimensionais a exemplo das listadas acima.

A utilização de sistemas OLTP é indicada para usuários que tem por finalidade a criação, atualização e recuperação de informações sobre registros individuais (visão tabular). Porém, o OLAP permite que usuários realizem análises globais das informações (visão conceitual multidimensional dos dados), visando a tomada de decisões estratégicas.

5.2 Regras de Associação

Para a geração das regras de associação, procurou-se trabalhar com tabelas que apresentassem um povoamento satisfatório de dados (maioria dos atributos preenchidos e nenhum/poucos dado(s) faltante(s)), além de utilizar tabelas que já estivessem armazenadas no Data Warehouse, facilitando a extração dos dados.

As tabelas que apresentaram um melhor povoamento dos dados foram as referentes às Variáveis da Água (Figura 5.16) e Coleta de Peixe (Figura 5.17).

IdPontoColeta	IdPesca	IdVariavelAgua	Hora	Profundidade	Condutividade	Ph	Transparencia	OxigSuperficie	OxigProfundidade	TempSuperf
24980	8199	344	1/6/2007 5:00:0...	NULL	NULL	5.80	NULL	0.45	0.29	28.60
24981	8200	345	1/6/2007 7:00:0...	NULL	NULL	6.70	NULL	0.70	0.58	28.80
24981	8200	346	1/6/2007 5:00:0...	NULL	NULL	6.20	NULL	1.07	0.26	31.80
24982	8201	347	1/6/2007 8:20:0...	NULL	NULL	7.60	NULL	1.15	0.90	28.70
24982	8201	348	1/6/2007 5:30:0...	NULL	NULL	6.40	NULL	2.20	0.30	28.60
24983	8202	349	1/6/2007 8:20:0...	NULL	NULL	7.10	NULL	1.10	0.80	28.40
24983	8202	350	1/6/2007 5:00:0...	NULL	NULL	6.30	NULL	2.40	0.30	29.50
24984	8203	351	1/6/2007 7:00:0...	NULL	NULL	6.80	NULL	1.30	0.90	26.80
24984	8203	352	1/6/2007 5:00:0...	NULL	NULL	6.70	NULL	2.50	1.40	28.00
24985	8204	353	1/6/2007 7:20:0...	NULL	NULL	6.50	NULL	0.70	0.30	27.50
24985	8204	354	1/6/2007 5:00:0...	NULL	NULL	6.30	NULL	2.30	0.40	28.30
24987	8206	355	1/6/2007 6:00:0...	NULL	NULL	5.60	NULL	4.00	4.00	28.90
24987	8206	356	1/6/2007 5:00:0...	NULL	NULL	6.70	NULL	5.40	4.30	31.90
24989	8208	357	1/6/2007 6:40:0...	NULL	NULL	6.69	NULL	6.60	5.80	30.80
24989	8208	358	1/6/2007 4:30:0...	NULL	NULL	6.88	NULL	5.30	0.05	32.90
24990	8209	359	1/6/2007 7:00:0...	NULL	NULL	6.30	NULL	1.86	0.97	29.00
24990	8209	360	1/6/2007 5:00:0...	NULL	NULL	6.50	NULL	5.50	0.07	32.60
24991	8210	361	1/6/2007 6:50:0...	NULL	NULL	6.40	NULL	0.13	0.08	29.10
24991	8210	362	1/6/2007 5:00:0...	NULL	NULL	4.85	NULL	2.30	0.04	34.20
24992	8211	363	1/6/2007 6:00:0...	NULL	NULL	5.25	NULL	3.04	0.06	30.60
24992	8211	364	1/6/2007 5:00:0...	NULL	NULL	5.75	NULL	2.30	0.30	33.60
24993	8212	365	1/6/2007 6:00:0...	NULL	NULL	5.40	NULL	2.10	0.93	31.30
24993	8212	366	1/6/2007 5:00:0...	NULL	NULL	5.70	NULL	4.70	3.20	33.50
24994	8213	367	1/6/2007 6:00:0...	NULL	NULL	4.00	NULL	4.00	2.50	30.80
24995	8214	368	1/6/2007 7:00:0...	NULL	NULL	4.40	NULL	3.00	2.10	28.40
24995	8214	369	1/6/2007 4:00:0...	NULL	NULL	4.20	NULL	2.60	1.80	31.30
24996	8215	370	1/6/2007 5:00:0...	NULL	NULL	5.60	NULL	2.77	1.83	31.80
24998	8217	387	1/6/2007 7:00:0...	NULL	NULL	7.60	NULL	2.90	2.70	28.60
24998	8217	388	1/6/2007 5:00:0...	NULL	NULL	7.30	NULL	5.40	4.80	31.70
24999	8218	389	1/6/2007 7:20:0...	NULL	NULL	7.80	NULL	3.80	3.40	29.40
24999	8218	390	1/6/2007 5:00:0...	NULL	NULL	7.50	NULL	3.60	3.50	29.60
25000	8219	391	1/6/2007 7:00:0...	NULL	NULL	8.00	NULL	0.30	0.30	28.60
25000	8219	392	1/6/2007 4:00:0...	NULL	NULL	7.40	NULL	2.30	1.20	31.30

Figura 5.16 – Dados de Variável da Água.

ALDEBARAN.Icti...na - dbo.Peixe		ALDEBARAN.Icti...o.VariavelAgua		Object Explorer Details			
IdPeixe	IdPesca	IdPontoColeta	Codigo	Peso	Comprimento	IdEspecieIctiof...	
10323	8211	24992	7037	120.00	14.00	30	
10324	8211	24992	7038	190.00	15.50	30	
10325	8211	24992	7039	165.00	14.50	30	
10326	8211	24992	7040	155.00	15.00	30	
10327	8211	24992	7041	225.00	16.50	30	
10328	8211	24992	7042	245.00	16.50	30	
10329	8211	24992	7043	170.00	15.00	30	
10330	8211	24992	7044	190.00	16.00	30	
10331	8211	24992	7045	175.00	15.50	30	
10332	8211	24992	7046	125.00	14.00	30	
10333	8211	24992	7047	125.00	13.50	30	
10334	8211	24992	7049	210.00	16.50	30	
10335	8211	24992	7050	125.00	15.50	33	
10336	8211	24992	7051	130.00	15.50	33	
10337	8211	24992	7052	100.00	13.00	33	
10338	8211	24992	7053	115.00	13.50	33	
10339	8211	24992	7054	90.00	13.00	33	
10340	8211	24992	7055	75.00	13.50	1365	
10341	8211	24992	7056	100.00	13.50	34	
10342	8211	24992	7057	45.00	10.00	34	
10343	8211	24992	7058	40.00	10.00	34	
10344	8211	24992	7059	100.00	15.00	1365	
10345	8211	24992	7060	485.00	28.00	16	
10346	8211	24992	7061	50.00	11.50	99	
10347	8211	24992	7062	65.00	11.00	28	
10348	8211	24992	7063	55.00	11.00	34	
10349	8211	24992	7064	385.00	25.00	16	
10350	8211	24992	7065	410.00	26.00	16	
10351	8211	24992	7066	305.00	23.00	92	
10352	8211	24992	7067	290.00	22.00	92	
10353	8211	24992	7068	260.00	21.00	92	
10354	8211	24992	7069	255.00	21.00	92	
10355	8211	24992	7070	265.00	22.00	92	
10356	8211	24992	7071	295.00	22.00	92	

Figura 5.17 – Dados de Coleta de Peixe.

Apesar do bom povoamentos destas tabelas, alguns atributos da tabela referente à Variável da Água possuíam muitos dados ausentes. Desta forma, os atributos Profundidade, Condutividade e Transparência não foram considerados.

O próximo passo foi a escolha do peixe a ser analisado. O critério utilizado para esta escolha foi a quantidade de peixes pescados por cada espécie.

Para verificar esta distribuição de dados, foi feito um histograma (Figura 5.18), que consiste em um gráfico de barras das freqüências de uma variável; neste caso, as variáveis são os peixes coletados.

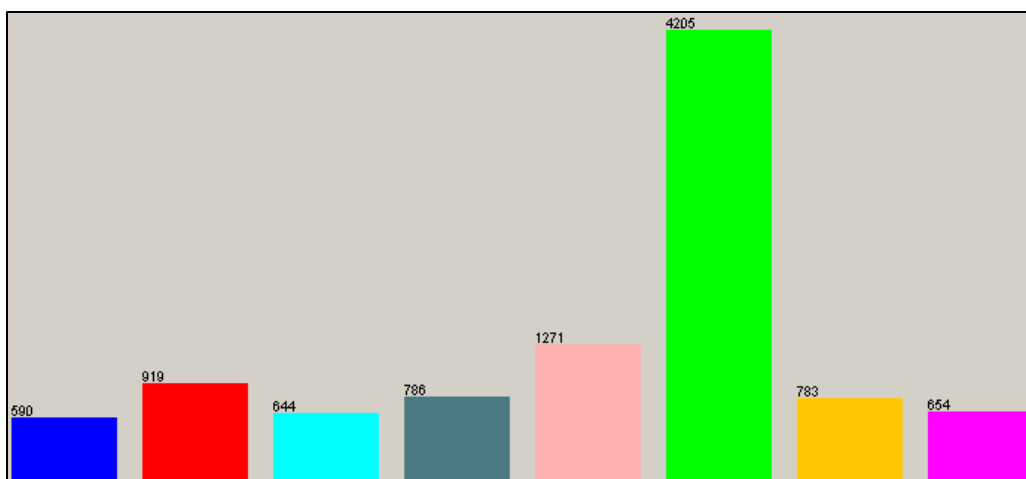


Figura 5.18 – Histograma de Espécies de Peixe.

TABELA 5.2 – Legenda do Histograma de Espécies de Peixe.

COR	ESPECIE	QUANTIDADE
Verde-limão	<i>Pygocentrus nattereri</i>	4205
Rosa	<i>Pterygoplichthys pardalis</i>	1271
Vermelho	<i>Potamorhina altamazonica</i>	919
Cinza	<i>Psectrogaster rutiloides</i>	786
Amarelo	<i>Serrasalmus rhombeus</i>	783
Magenta	<i>Serrasalmus spilopleura</i>	654
Ciano	<i>Potamorhina latior</i>	644
Azul	<i>Mylossoma duriventre</i>	590
TOTAL		9852

Através do histograma, verificamos que o peixe *Pygocentrus nattereri* (Piranha-caju) representa aproximadamente 50% dos peixes coletados e foi o escolhido para a análise.

Os atributos escolhidos para extração através do Data Warehouse foram: Local da Coleta do Peixe, Peso, Comprimento, Média do PH, Média do Oxigênio de Superfície, Média do Oxigênio de Profundidade, Média da Temperatura de Superfície e Média da Temperatura de Profundidade.

As regras de associação, geralmente, apresentam uma abordagem booleana do problema, onde importa apenas se o item está presente (“1” ou SIM) ou não (“0” ou NÃO) na transação, sendo desconsiderados os atributos quantitativos. Porém, normalmente, bancos de dados não são restritos a atributos binários, mas também

contém atributos com valores numéricos. A BDI PIATAM se enquadra nessa condição, uma vez que grande parte de suas informações são variáveis quantitativas.

Para que as regras de associação fossem geradas, tornou-se necessária a transformação dos dados quantitativos em subconjuntos (intervalos de dados). Foram utilizados intervalos eqüidistantes, levando-se em conta o valor máximo e mínimo de cada atributo. Os intervalos são:

- **Peso** – Apresentou valores que iniciavam em 0,5 e terminavam um pouco acima de 4,50 (Valores em kg, de acordo com dicionário de dados fornecido para realização do estudo).

TABELA 5.3 – Intervalo de Dados do Atributo Peso.

Intervalo	Denominação Qualitativa
0,5 Kg à 1,50 Kg	Muito Leve
1,501 Kg à 3 Kg	Leve
3,001 Kg à 4,50 Kg	Pesado
4,501 em diante	Muito Pesado

- **Comprimento** – Apresentava valores que iniciavam em 5 e terminavam um pouco acima de 20 (Valores em centímetros, de acordo com dicionário de dados fornecido para realização do estudo).

TABELA 5.4 – Intervalo de Dados do Atributo Comprimento.

Intervalo	Denominação Qualitativa
5 cm à 10 cm	Muito Curto
10,1 cm à 15 cm	Curto
15,1 cm à 20 cm	Longo
20,1 cm em diante	Muito Longo

-
- PH (Potencial Hidrogeniônico) – Apresentou valores entre 1 e 8,9 .

TABELA 5.5 – Intervalo de Dados do Atributo PH

1 a 2
2,1 a 3
3,1 a 4
4,1 a 5
5,1 a 6
6,1 a 7
7,1 a 8
8,1 a 9

- Oxigênio de Superfície – Apresentou valores entre 0,1 e 5,5 . Porém, a maiorias destes valores estavam no intervalo de 0,1 à 4. (Valores em miligrama de oxigênio por litro, de acordo com dicionário de dados fornecido para realização do estudo).

TABELA 5.6 – Intervalo de Dados do Atributo Oxigênio de Superfície.

0,1 à 1
1,1 à 2
2,1 à 3
3,1 à 4
4,1 em diante

- Oxigênio de Profundidade – Apresentou valores entre 0,1 e 5,5 . Porém, a maiorias destes valores estavam no intervalo de 0,1 à 4. (Valores em miligrama de oxigênio por litro, de acordo com dicionário de dados fornecido para realização do estudo).

TABELA 5.7 – Intervalo de Dados do Atributo Oxigênio de Profundidade.

0,1 à 1
1,1 à 2
2,1 à 3
3,1 à 4
4,1 em diante

- Temperatura de Superfície – Apresentou valores entre 25 e 34 (Valores em grau Celsius, de acordo com o dicionário de dados fornecido para realização do estudo).

TABELA 5.8 – Intervalo de Dados do Atributo Temperatura de Superfície (°C).

Intervalo	Denominação Qualitativa
25 à 28	Baixa
28,1 à 30	Intermediária
30,1 à 34	Alta

- Temperatura de Profundidade – Apresentou valores entre 25 e 34 (Valores em grau Celsius, de acordo com o dicionário de dados fornecido para realização do estudo).

TABELA 5.9 – Intervalo de Dados do Atributo Temperatura de Profundidade (°C)

Intervalo	Denominação Qualitativa
25 à 28	Baixa
28.1 à 30	Intermediária
30.1 à 34	Alta

Após esta etapa, foi iniciada a geração das regras de associação utilizando o algoritmo *Apriori*.

5.2.1 Análise dos Resultados Obtidos com as Regras de Associação

Para a geração das regras de associação, foi utilizado o software livre Weka (*Waikato Environment for Knowledge Analysis* – Disponível para Download em: <http://www.cs.waikato.ac.nz/ml/weka/>), no qual é formado por um conjunto de implementações de algoritmos de diversas técnicas de Mineração de Dados.

Para uma primeira análise, o algoritmo foi executado utilizando valor de suporte mínimo 0.3 (30%) e confiança mínima 0.9 (90%).

Com essa parametrização, foi gerado um número muito grande de regras (acima de 1.000), sendo estas, em sua maioria, regras óbvias.

Mesmo após as mudanças na parametrização do algoritmo e utilização de medidas de interesse, tal como, *lift* e *conviction*, o número de regras geradas não diminuiu, inviabilizando a análise dos resultados.

Como as variáveis escolhidas para o estudo eram quantitativas e necessitavam de uma conversão para a geração das regras de associação, o número de variáveis aumentou (de 08 para 32), contribuindo para o grande número de regras geradas.

Apesar disto, com o intuito de descobrir alguma informação interessante, foram geradas regras de associação para cada localidade onde o peixe *Pygocentrus nattereri* (Piranha-cajú) foi coletado. Foram considerados apenas os peixes com peso “entre 0.5 Kg e 1.5 Kg” (Muito Leve) e “entre 1.51 Kg e 3 Kg” (Leve) pois dos 4.205 registros deste peixe, apenas 339 não estavam neste intervalo.

Estas regras foram geradas utilizando como valor de suporte mínimo 0,3 e confiança mínima 0,6.

A seguir serão apresentadas as regras interessantes encontradas.

TABELA 5.10 - Regras de Associação Referentes ao Lago Ananá.

Regra	Antecedente	Conseqüente	Suporte (%)	Confiança (%)
01	Peso=MuitoLeve	PH=entre 7.1 e 8	47,7	67,5
02	Peso=Leve	PH=entre 5.1 e 6	35,1	63,9
03	Peso=MuitoLeve	OxigenioSuperficie=2.1 e 3	40,3	64,9
04	Peso=Leve	OxigenioSuperficie=2.1 e 3	61,7	83,8
05	Peso=MuitoLeve	OxigenioProfundidade=0.1 e 1	49,2	68,2
06	Peso=Leve	OxigenioProfundidade=1.1 e 2	44,8	67,2
07	Peso=MuitoLeve	TemperaturaSuperficie=Intermediária	40,2	65,6
08	Peso=Leve	TemperaturaSuperficie=Intermediária	55,2	77,6
09	Peso=MuitoLeve	TemperaturaProfundidade=Intermediária	82,7	92,3
10	Peso=Leve	TemperaturaProfundidade =Intermediária	52,7	75,4

TABELA 5.11 - Regras de Associação Referentes ao Lago Araçá.

Regra	Antecedente	Conseqüente	Suporte (%)	Confiança (%)
01	Peso=MuitoLeve	PH=entre 7.1 e 8	78,4	83,1
02	Peso=Leve	PH=entre 7.1 e 8	49,2	72,3
03	Peso=MuitoLeve	OxigenioSuperficie=2.1 e 3	67,9	74,8
04	Peso=Leve	OxigenioSuperficie=2.1 e 3	58,4	75,9
05	Peso=MuitoLeve	OxigenioProfundidade=0.1 e 1	37,6	65,9
06	Peso=Leve	OxigenioProfundidade=2.1 e 3	33,6	62,7
07	Peso=MuitoLeve	TemperaturaSuperficie=Intermediária	67,3	72,3
08	Peso=Leve	TemperaturaSuperficie=Intermediária	56,2	69,4
09	Peso=MuitoLeve	TemperaturaProfundidade=Intermediária	64,1	70,7
10	Peso=Leve	TemperaturaProfundidade =Intermediária	66,7	71,3

TABELA 5.12 - Regras de Associação Referentes ao Lago Aruã.

Regra	Antecedente	Conseqüente	Suporte (%)	Confiança (%)
01	Peso=MuitoLeve	PH=entre 7.1 e 8	65,7	73,2
02	Peso=Leve	PH=entre 7.1 e 8	71,2	80,2
03	Peso=MuitoLeve	OxigenioSuperficie=2.1 e 3	72,5	77,4
04	Peso=Leve	OxigenioSuperficie=2.1 e 3	72,7	78,3
05	Peso=MuitoLeve	OxigenioProfundidade=1.1 e 2	68,3	71,5
06	Peso=Leve	OxigenioProfundidade=1.1 e 2	77,2	81,5
07	Peso=MuitoLeve	TemperaturaSuperficie=Baixa	69,4	77,6
08	Peso=Leve	TemperaturaSuperficie=Baixa	71,9	75,3
09	Peso=MuitoLeve	TemperaturaProfundidade=Baixa	58,7	68,4
10	Peso=Leve	TemperaturaProfundidade =Baixa	69,7	76,8

TABELA 5.13 - Regras de Associação Referentes ao Lago Baixo.

Regra	Antecedente	Conseqüente	Suporte (%)	Confiança (%)
01	Peso=MuitoLeve	PH=entre 6.1 e 7	52,8	65,4
02	Peso=MuitoLeve	OxigenioSuperficie=Acima de 4	33,7	63,6
03	Peso=MuitoLeve	OxigenioProfundidade=1.1 e 2	45,1	67,8
04	Peso=MuitoLeve	TemperaturaSuperficie=Alta	71,5	89,3
05	Peso=MuitoLeve	TemperaturaProfundidade=Alta	69,8	79,4

TABELA 5.14 - Regras de Associação Referentes ao Lago Iauara.

Regra	Antecedente	Conseqüente	Suporte (%)	Confiança (%)
01	Peso=MuitoLeve	PH=entre 7.1 e 8	53,6	68,7
02	Peso=Leve	PH=entre 5.1 e 6	52,7	67,4
03	Peso=MuitoLeve	OxigenioSuperficie=2.1 e 3	61,8	73,3
04	Peso=Leve	OxigenioSuperficie=1.1 e 2	61,3	74,7
05	Peso=MuitoLeve	OxigenioProfundidade=2.1 e 3	57,9	71,8
06	Peso=Leve	OxigenioProfundidade=0.1 e 1	78,7	90,4
07	Peso=MuitoLeve	TemperaturaSuperficie=Alta	49,1	62,3
08	Peso=Leve	TemperaturaSuperficie=Alta	58,3	72,4
09	Peso=MuitoLeve	TemperaturaProfundidade=Intermediária	82,3	92,1
10	Peso=Leve	TemperaturaProfundidade =Intermediária	61,2	73,5

TABELA 5.15 - Regras de Associação Referentes ao Lago Maracá.

Regra	Antecedente	Conseqüente	Suporte (%)	Confiança (%)
01	Peso=MuitoLeve	PH=entre 6.1 e 7	43,9	60,3
02	Peso=MuitoLeve	OxigenioSuperficie=1.1 e 2	62,9	71,7
03	Peso=MuitoLeve	OxigenioProfundidade=1.1 e 2	71,9	88,7
04	Peso=MuitoLeve	TemperaturaSuperficie=Intermediária	73,3	89,3
05	Peso=MuitoLeve	TemperaturaProfundidade=Baixa	54,7	68,5

TABELA 5.16 - Regras de Associação Referentes ao Lago Poraquê

Regra	Antecedente	Conseqüente	Suporte (%)	Confiança (%)
01	Peso=MuitoLeve	PH=entre 7.1 e 8	78,7	87,6
02	Peso=Leve	PH=entre 7.1 e 8	87,5	93,6
03	Peso=MuitoLeve	OxigenioSuperficie=1.1 e 2	65,4	72,5
04	Peso=Leve	OxigenioSuperficie=1.1 e 2	88,6	78,4
05	Peso=MuitoLeve	OxigenioProfundidade=1.1 e 2	92,3	94,5
06	Peso=Leve	OxigenioProfundidade=1.1 e 2	97,5	95,8
07	Peso=MuitoLeve	TemperaturaSuperficie=Intermediária	76,3	84,7
08	Peso=Leve	TemperaturaSuperficie=Intermediária	89,7	92,3
09	Peso=MuitoLeve	TemperaturaProfundidade=Baixa	83,7	77,2
10	Peso=Leve	TemperaturaProfundidade =Baixa	92,9	93,7

TABELA 5.17 - Regras de Associação Referentes ao Lago Preto

Regra	Antecedente	Conseqüente	Suporte (%)	Confiança (%)
01	Peso=MuitoLeve	PH=entre 7.1 e 8	68,7	74,9
02	Peso=Leve	PH=entre 6.1 e 7	91,5	94,7
03	Peso=MuitoLeve	OxigenioSuperficie=0.1 e 1	52,1	68,4
04	Peso=Leve	OxigenioSuperficie=2.1 e 3	52,2	68,7
05	Peso=MuitoLeve	OxigenioProfundidade=0.1 e 1	72,7	75,9
06	Peso=Leve	OxigenioProfundidade=1.1 e 2	73,4	76,7
07	Peso=MuitoLeve	TemperaturaSuperficie=Intermediária	63,2	72,4
08	Peso=Leve	TemperaturaSuperficie=Intermediária	67,3	74,9
09	Peso=MuitoLeve	TemperaturaProfundidade=Baixa	56,8	67,3
10	Peso=Leve	TemperaturaProfundidade =Intermediária	76,5	78,5

A partir das informações adquiridas em cada localização, podemos fazer algumas observações em relação à espécie de peixe *Pygocentrus nattereri* (Piranha-cajú). São elas:

- Os Peixes com peso entre 0,5 Kg e 1,5 Kg (muito leve), geralmente, foram coletados na presença de PH “entre 6.1 e 7” e “entre 7.1 e 8” . Já os peixes com peso entre 1,51 Kg e 3 Kg, na maioria das vezes, foi coletado quando o PH está “entre 7.1 e 8”. Existem poucos registros de coleta desta espécie de peixe em águas com PH entre 1 e 5.
- Tanto os peixes com peso entre 0,5 Kg e 1,5 Kg, quanto os de peso entre 1,51 Kg e 3 Kg, foram coletados, geralmente, quando o Oxigênio de Superfície estava “entre 2,1 e 3” . Existem pouquíssimos registros de

coleta desta espécie de peixe quando o Oxigênio de Superfície registra valores entre 3.1 e 4.

- Tanto os peixes com peso entre 0,5 Kg e 1,5 Kg, quanto os de peso entre 1,51 Kg e 3 Kg, foram coletados, geralmente, quando o Oxigênio de Profundidade estava “entre 0,1 e 1” e “entre 1,1 e 2” . Da mesma forma que ocorreu com o Oxigênio de Superfície, existem pouquíssimos registros de coleta desta espécie de peixe quando o Oxigênio de Profundidade registra valores entre 3,1 e 4.
- Tanto os peixes com peso entre 0,5 Kg e 1,5 Kg, quanto os de peso entre 1,51 Kg e 3 Kg, foram coletados, geralmente, quando a Temperatura de Superfície estava “28,1 e 30” (intermediária) e “30,1 e 34” (alta), ou seja, em temperaturas que variam de 28⁰C a 34⁰C. Em temperaturas de superfície mais baixas, esta espécie de peixe foi pouco coletada.
- Tanto os peixes com peso entre 0,5 Kg e 1,5 Kg, quanto os de peso entre 1,51 Kg e 3 Kg, foram coletados, de uma maneira geral, em todas as Temperaturas de Profundidade, ou seja, em temperaturas que variam de 25⁰C a 34⁰C, não havendo diferença entre temperaturas de profundidade baixas, intermediárias ou altas.

Para melhor interpretação das informações obtidas com o estudo, faz-se necessária uma maior interação com especialistas da área de Ictiofauna, desta forma, poderiam ser analisadas algumas questões mais específicas deste tema além da verificação dos resultados encontrados.

Capítulo 06

CONSIDERAÇÕES FINAIS

6.1 Conclusão

O presente estudo apresentou a construção de um Data Warehouse (DW) para o tema “Ictiofauna” da Base de Dados Integrada PIATAM e, a partir desta construção, a aplicação de ferramentas OLAP (*On-Line Analytical Processing*) visando uma maior interação dos gerentes e participantes do projeto com os dados existentes através de agregações e sumarizações. Com a construção do DW, foram geradas regras de associação objetivando a descoberta de conhecimento interessante no repositório de dados construído.

Inicialmente, para a construção do DW, foi necessária a criação de um banco de dados relacional que serviria de apoio à criação do DW. Para isto, foi feita a modelagem Entidade-Relacionamento e foram criadas as tabelas que iriam ser povoadas com os dados provenientes da BDI-PIATAM.

Após a criação do Banco de Dados Relacional foram definidas as Tabelas Fato e as Medidas e Dimensões do Data Warehouse. Após estas definições, ferramentas OLAP foram aplicadas nos dados gerando exemplos de consultas que poderiam ser feitas por integrantes do projeto.

Após a demonstração do uso das ferramentas OLAP, foram escolhidos os fatos que seriam analisados através das regras de associação. Após a extração dos dados escolhidos, foi necessária a transformação dos dados quantitativos em subconjuntos de dados. Para isto, foram definidos limites para cada atributo quantitativo.

Os algoritmos de regras de associação foram gerados. Porém, o número de regras geradas inviabilizou a análise dos resultados. Uma das razões para o grande número de regras geradas foi que, com a transformação dos dados quantitativos para subconjuntos de dados, o número de atributos aumentou consideravelmente.

Porém, visando ainda a descoberta de algum conhecimento no repositório de dados, foram geradas regras de associação para uma espécie de peixe nas localidades em que ele foi coletado.

Conclui-se que a construção de um Data Warehouse para os dados da BDI-PIATAM, não apenas para dados de Ictiofauna mas para todos os temas estudados, é de grande valia, pois os participantes do projeto poderiam acessar informações sintetizadas, fazer comparações, realizar consultas personalizadas e análises históricas em vários cenários.

Além disto, pode-se concluir que a construção do Data Warehouse auxiliou no processo de descoberta de conhecimento pois através dele foram identificados possíveis relacionamentos entre os fatos, além da extração dos dados utilizados na geração das regras de associação.

6.2 Trabalhos Futuros

Para trabalhos futuros propõem-se:

- Conforme dito anteriormente, poderia ser construído um DW-PIATAM que armazenasse os dados de todos os temas, desde o início do projeto até os dias atuais. Com isto, as consultas aos dados poderiam ser facilitadas. O DW-PIATAM construído poderia ser acessado via web, como acontece com a BDI-PIATAM, facilitando o acesso aos dados de qualquer localidade;
- Inclusão do Sistema de Indicadores PIATAM como medida do Data Warehouse. No início do estudo pensou-se nessa possibilidade, porém, não haviam dados disponíveis para realização dos cálculos na época de sua aquisição;
- Seria de grande proveito a integração do Data Warehouse com a ferramenta de Data Mining. As ferramentas comerciais já possuem esta funcionalidade, porém, o usuário não possui total controle dos algoritmos utilizados pelo software;

- Conforme mencionado durante o trabalho, regras de associação, geralmente, apresentam uma abordagem booleana do problema, onde importa apenas se o item está presente (“1”) ou não (“0”) na transação, sendo desconsiderado os atributos quantitativos. Entretanto, normalmente, os bancos de dados não são restritos a atributos binários, mas também contém atributos com valores numéricos. A BDI PIATAM se enquadra nessa condição uma vez que grande parte de suas informações são variáveis quantitativas. Para resolver este problema, poderia ser utilizado conjuntos fuzzy combinado com regras de associação. A idéia é que os conjuntos fuzzy atuem como uma interface entre a escala numérica e uma escala simbólica (i.e. termos linguísticos), onde as regras descobertas são apresentadas de um modo linguístico e por consequência mais compreensivo e amigável;
- Adotar outros critérios para definição dos intervalos dos atributos utilizando métodos estatísticos;
- Outro estudo interessante seria a integração de um SIG (Sistema de Informação Geográfica) à um Data Warehouse com funcionalidades OLAP, possibilitando uma visão espacial dos problemas. Como dito durante o trabalho, atualmente, esta é uma área de grande evidência.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGGELIS, V. (2004). Association rules model of e-banking services. *Data Mining V – Information and Communication Technologies*.
- AGRAWAL, R. & R. SRIKANT (1994). Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke, & C. Zaniolo (Eds.), *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB 1994*, pp. 487–499. Disponível em: <http://citeseer.ist.psu.edu/agrawal94fast.html>.
- AGRAWAL, R.; AGGARWAL, C.; PRASAD, V.V.V. Depth First Generation of Long Patterns. In *7th Int'l Conference on Knowledge Discovery and Data Mining, 2000*.
- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. Mining Association Rules between Sets of Items in Large Databases. *Proceedings of the ACM SIGMOD Conference*. Washington, USA, May 1993.
- AGRAWAL, R.; MANNILA, H.; SRIKANT, R.; TOIVONEN, H.; VERKAMO, A. INKERI. Fast Discovery of association Rules. In *Advances in Knowledge Discovery and Data Mining*, Fayyad et al. (Eds.). AAAI Press/The MIT Presss, 1996.
- AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules. *Proc. Int'l Conf. Very Large Data Bases*, pp. 487-499, Santiago, Chile, Sept. 1994.
- Alves, A. S., *Regras De Associação e Classificação em Ambiente de Computação Paralela Aplicadas à Sistemas Militares*. Tese (Doutorado), COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2007.
- APTE, C., B. LIU, E. P. D. PEDNAULT, & P. SMYTH (2002). Business applications of data mining. *Communications of the ACM*.
- BARQUIN, R. C.; EDELSTEM, H. A. *Building, using and managing the Data Warehouse*. New Jersey: Prentice Hall PTR, 1997.
- BARQUINI, R. (1996). *Planning and Designing the Warehouse*. New Jersey: Prentice-Hall.

- BAYARDO JR., R. J.; AGRAWAL, R. Mining the Most Interesting Rules. In: ACM SIGKDD INT'L CONF. ON KNOWLEDGE DISCOVERY AND DATA MINING, 5, 1999, San Diego. Proceedings... New York: ACM Press, 1999.
- BAYARDO, R. J. Efficiently mining long patterns from databases. In ACM SIGMOD Conf. of Management of Data, 1998.
- BISPO, C. A. F. Uma Análise da Nova Geração de Sistemas de Apoio à Decisão. Dissertação (Pós-Graduação) - Escola de Engenharia de São Carlos, Universidade de São Paulo. São Carlos, 1998.
- BRIN, S.; MOTWANI, R.; ULLMAN, J. - Dynamic Itemset Counting and Implication Rules for Market Basket Data. In Proc. of the ACM Int'l Conf. on Management of Data, 1997.
- BRUSSO, M. J. - Access Miner: uma proposta para a extração de regras de associação aplicada à mineração do uso da Web. Dissertação de Mestrado, Instituto de Informática, UFRGS, 2000.
- BURDICK, D.; CALIMLIM, M.; GEHRKE, J. - MAFIA: a maximal frequent itemset algorithm for transactional databases. In Intl. Conf. on Data Engineering, Apr. 2001.
- CASTELO, R.; FEELDERS, A.; SIEBES, A. - Mambo: Discovering Association Rules Based on Conditional Independencies. In Lecture Notes in Computer Science, vol. 2189, pages 289-298. 2001.
- CASTRO, I. X., SILVA, J. C., RIVAS, A. A. F., FREITAS, C. E. C. PIATAM – INTELIGÊNCIA SOCIOAMBIENTAL ESTRATÉGICA NA AMAZÔNIA: GOVERNANÇA GERANDO CIÊNCIA & TECNOLOGIA. T&C Amazônia, Ano IV, Número 15, 2008.
- CHEN, S. Y. & X. LIU (2005). Data mining from 1994 to 2004: an application-orientated review. International Journal of Business Intelligence and Data Mining.
- CLARKE, K. C. Getting Started with Geographic Information Systems, Prentice Hall, USA, 1997.
- CODD, E. F. (1993). Providing OLAP (On-Line Analytical Processing) to User-Analyst: an IT Mandate. E. F. Codd and Assoc.

- COREY, M., M. ABBEY, I. ABRAMSON, & B. TAUB (2001). Oracle8i Data Warehouse. Rio de Janeiro-RJ: Editora Campus.
- COSTA, C. E.; NEIS, J. F. Desenvolvendo um Middleware para a Integração de Servidores OLAP e Servidores de Mapa na Web. Tese (Graduação) — Universidade Federal de Santa Catarina, Florianópolis, 2007.
- DOMINGO, C.; GAVALDÀ, R.; WATANABE, O. - On-line Sampling Methods for Discovering Association Rules. Universitat Politècnica de Catalunya, Departament de Llenguatges i Sistemes Informàtics, Report Number: LSI-99-4-R, 1998.
- DOMINGUES, M. A., A. M. JORGE, & C. SOARES (2006). Using association rules for monitoring meta-data quality in web portals. In Proceedings of the II Workshop em Algoritmos e Aplicações de Mineração de Dados (WAAMD-2006) – SBBD/SBES.
- FAYYAD, U. M., G. PIATETSKY-SHAPIRO, & P. SMYTH (1996). Knowledge discovery and data mining: Towards a unifying framework. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, KDD 1996. AAAI Press.
- FAYYAD, U.; SHAPIRO, G.; SMYTH, P. - The KDD process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM (Special Issue on Data mining). November, 1996.
- FERREIRA, A. C. Um Modelo para Suporte à Integração de Análises Multidimensionais e Espaciais. Tese (Mestrado) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2002.
- FIDALGO, R. N., TIMES, V. C., SILVA, J., SOUZA, F.F., GeoDWFrame: A Framework for Guiding the Design of Geographical Dimensional Schemas. Proc. Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK), 2004.
- FIGUEIREDO A.M.C.M. Molap x Rolap: Embate de Tecnologias para Data Warehouse. Developer's Magazine, n. 18, p 25-25, fev, 1998.
- FONSECA, B. M., P. B. GOLGHER, E. S. DE MOURA, & N. ZIVIANI (2003). Using association rules to discover search engines related queries. In Proceedings of the First Conference on Latin American Web Congress – LA-WEB'03, pp. 66–71. IEEE Computer Society.

- GAHEGAN, M. (2001) "Data mining and knowledge discovery in the geographical domain", White Paper: National Academies Computer Science and Telecommunications Board.
http://www7.nationalacademies.org/cstb/wp_geo_gahegan.pdf
- GATZIU, S. & A. VAVOURAS (1999). Data Warehousing: Concepts and Mecanismos. Heidelberg - Germany: DMDW.
- GOLFARELLI, M., D. MAIO, & S. RIZZI (1998). Conceptual Design of Data Warehouses from E/R Schemes. In Hawaii International Conference on System Sciences, Kona - Hawaii.
- GONÇALVES, L. Modelos Neuro-Fuzzy Hierárquicos BSP para Classificação de Padrões e Extração e Regras Fuzzy em Bancos de Dados. Dissertação de Mestrado, Departamento de Engenharia Elétrica, PUC-Rio, 2001.
- GOUDA, K.; ZAKI, M. - Efficiently mining maximal frequent itemsets. In Proc. of the 1st IEEE Int'l Conference on Management of Data, 2001.
- Gunolulos, D.; Mannila, H.; Saluja, S. - Discovering all the most specific sentences by randomized algorithms. In Int'l Conf. on Database Theory, 1997.
- HAN, J. P.; YIN, Y. - Mining Frequent Patterns without candidate generation. In Proc. of the ACM SIGMOD Int'l Conf. on Management of Data, 2000.
- HAN, MAO, R. - Closet: An effecient algorithm for mining frequent closed itemsets. In SIGMOD Int'l Workshop on Data Mining and Knowledge Discovery, 2000.
- HOUTSMA, M.; SWAMI, A. - Set-oriented Mining of Association Rules. Research Report RJ 9567, IBM Almaden Ressearch Center, San Jose, California, 1993.
- INMON, W. Building the Data Warehouse. QED Press/ John Wiley, 1º edição. 1992.
- INMON, W. H. & R. D. HACKARTHORN (1997). Como usar o Data Warehouse. Rio de Janeiro: Infobook / IBPI Press.
- INMON, W. H. Como construir Data Warehouse. Rio de Janeiro: Campus, 1997.
- JORGE, A. M. – Material de Apoio do Mestrado em Análise de Dados e Sistema de Apoio à Decisão. Faculdade de Economia, Universidade do Porto. Disponível em: www.niaad.liacc.up.pt/~amjorge/Aulas/madsad/ecd2.

- KIMBALL, R. (1997). Data Warehouse Toolkit. São Paulo: Makron Books.
- KIMBALL, R. Data Ware House toolkit. São Paulo: Makron Books, 1998.
- KIMBALL, R. The Data Warehouse Toolkit. Wiley, 1º edição, 1996.
- LIN, D-I.; DUNHAM, M. H. - Pincer-search: A new algorithm for discovering de maximum frequent set. In 6th Int'l Conf. Extending Database Technology, 1998.
- LIU, B.; HSU, W.; MA, Y. - Mining Association Rules with Multiple Minimum Supports. Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99).
- LOPES, C.H.P. - Classificação de registros em banco de dados por evolução de regras de associação utilizando algoritmos genéticos. Dissertação de Mestrado, Dept. de Engenharia Elétrica, PUC/RJ, 1999.
- MACDONALD, G. C.; RUBIK, J. R. Pesquisa e Seleção de Ferramentas Livres e Baseadas em Padrões de Sistemas Abertos para a Elaboração de Interfaces OLAP sobre a Web. Tese (Graduação) — Universidade Federal de Santa Catarina, Florianópolis, 2007.
- MACHADO, F. N. R. Tecnologia e projeto de Data Warehouse: uma visão multidimensional. Érica, 1º edição. São Paulo, 2004.
- MAGNUM OPUS. Disponível em: www.giwebb.com.
- MARQUES, V. F., R. H. SUZUQUI, G. S. ARAÚJO, & J. C. C. NETO (2000, Fevereiro). Projeto e Construção de um Sistema de Banco de Dados que Permita Acesso Pela Internet Usando o SGBD Oracle. Projeto de Graduação apresentado no curso de Bacharelado em Ciência da Computação - DCT - CCET - UFMS.
- METWALLY, A., D. AGRAWAL, & A. E. ABBADI (2005). Using association rules for fraud detection in web advertising networks. In Proceedings of the 31st international conference on very large data bases – VLDB'05.
- MOREIRA, F. R. S. Uso de Regras de Associação Fuzzy na Base de Dados Integrada PIATAM/SIPAM. 2005. Disponível em:<http://www.labhidro.ufsc.br/Eventos/V%20EHA/Regras%20de%20Associacao%20Fuzzy%20na%20Base%20PIATAM-final2.ppt>

- OPEN GIS CONSORTIUM. Filter encoding implementation specification. Technical report, Open GIS Consortium, 2001.
- OPENSHAW, S. (1994), "Two exploratory space-time-attribute pattern analysers relevant to GIS", In: Fotheringham, A. S., Rogerson, P. A. (eds) Spatial analysis and GIS, Taylor and Francis Ltd. 1900 Frost Road, Suite 101, Bristol PA 19007, pp 83-103
- OPENSHAW, S. (1999), "Geographical data mining: key design issues", GeoComputation 99, Mary Washington College in Fredericksburg, VA, USA, on 25-28 July 1999 http://www.geovista.psu.edu/sites/geocomp99/Gc99/051/gc_051.htm
- PARK, J.S.; CHEN, M.-S. - Using a Hash-Based Method with Transaction Trimming for Mining Association rules. IEEE Transactions on Knowledge and Data Engineering, Vol. 9, No 5, 1997.
- PASQUIER, N.; BASTIDE, Y.; TAOUIL, R.; LAKHAL, L. - Discovering frequent closed itemsets for association rules. In Proc of 7th Int'l Conf. on Database Theory (ICDT), Verlag, 1999.
- PENNA, R. A. C.; REIS, A. S.; CAVALCANTI, M. Informações Geográficas Integradas à Inteligência do Negócio. Disponível em: <<http://www.gpi.ufrj.br/pdfs/artigos/Penna,%20Reis%20Jr,%20Cavalcanti%20-%20Informacoes%20Geograficas%20Integradas%20a%20Inteligencia%20do%20Negocio%20-%20SUCESU%20-%202004.pdf>>
- POE, V.; KLAWER, P.; BROBST, S. Building a Data Warehouse for decision support. New Jersey: Prentice Hall PTR, 1998.
- RAO, F. Spatial hierarchy and olap-favored search in spatial data warehouse. DOLAP, Louisiana, Novembro 2003.
- RIANI, A. P. da C. A questão do uso de banco de dados Relacional e Multidimensional em consultas em Data Warehouses. 2002. Disponível em: <[http://www.comp.ufla.br/monografias/ano2002/A questão do uso de banco de dados relacional e multidimensional em consultas em DataWarehouses.pdf](http://www.comp.ufla.br/monografias/ano2002/A%20questao%20do%20usode%20banco%20de%20dados%20relacional%20e%20multidimensional%20em%20consultas%20em%20DataWarehouses.pdf)>.
- SAVASERE, A.; OMIELINSKI, E.; NVATHE, S. - An Efficient Algorithm for Mining Association Rules in Large Databases. In Very Large Databases, 1995.

- SEMENOVA, T., M. HEGLAND, W. GRACO, & G. WILLIAMS (2001). Effectiveness of mining association rules for identifying trends in large health databases. In F. J. Kurfess & M. Hilario (Eds.), Workshop on Integrating Data Mining and Knowledge Management, The 2001 IEEE International Conference on Data Mining, ICDM 2001. Disponível em: <http://cui.unige.ch/~hilario/icdm-01/DM-DM-Final/Semenova.pdf>
- SHIGUNOV, F. Uma Aplicação OLAP sobre a Web para Análise dos Dados do Vestibular da UFSC e Diretrizes para sua integração com GIS. Monografia (Graduação) Universidade Federal de Santa Catarina, Florianópolis, 2007.
- SIEBES, A. - Association Rules. Institute of Information and Computing Sciences, Department of Mathematics and Computer Science, Universiteit Utrecht. Disponível em www.cs.uu.nl/docs/vakken/dm/assoc-complete.pdf.
- SILVA, A. F. T. Identificação de Regras de Associação Interessantes em uma Base de Dados sobre Exsudações de Óleo no Golfo do México. Dissertação (Mestrado), COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- SILVA, J. Propondo uma linguagem de consulta geográfica multidimensional. GEOINFO, 2004.
- SILVA, J; TIMES, V.C; BARROS, R. S. M, Integrando Serviços Analíticos e Geográficos para Suporte à Decisão na Web. Dissertação de Mestrado – Universidade Federal de Pernambuco (UFPE), 2004.
- SRIKANT, R.; VU, Q.; AGRAWAL, R. - Mining Association Rules with Item Constraints. IBM Almaden Research Center & American Association for Artificial Intelligence, 1997.
- TANIAR, D. (2007). Research And Trends in Data Mining Technologies And Applications (Advances in Data Warehousing and Mining Series). Idea Group Publishing.
- TAURION C. Data Warehouse: Estado de arte e estado de prática. Developer's Magazine, n. 6, p 10-11, fev, 1997.
- TOCANTINS, C. A.; BOTELHO, A. .; CAVALCANTE, K. V; COSTA, M. G; MIRANDA, F. P.; MOREIRA, F. R. S.; PRADO, W. F.; NETO, M. A. A. Base de dados Integrada PIATAM: aplicação multidisciplinar e integração espaço temporal. 1º Simpósio de

Geotecnologias no Pantanal, Campo Grande, Brasil, 11-15 novembro 2006, Embrapa Informática Agropecuária/INPE, p.757-762.

- TODMAN, C. (2001). Designing a Data Warehouse - Supporting Customer Relationship Management. New Jersey: Prentice-Hall.
- TOINIVEN, H. - Sampling Large Databases for Association Rules. Proceedings of the 22th VLDB Conference, 1996.
- WIZSOFT. Disponível em www.wizsoft.com.
- ZAKI, M. J.; PARTHASARATHY, S.; LI, W. - New Algorithms for Fast Discovery of Association Rules. In Proc. of the 3th ACM Intl. Conf. On Knowledge Discovery and Data Mining, 1997.
- ZAKI, M.J.; HSIAO, C.-J. - CHARM: An Efficient Algorithm for Closed Itemset Mining. Proceedings of the Second SIAM International Conference on Data Mining, USA, 2002.

Anexo 01

Dicionário de Dados do Banco de Dados

Descrição das Tabelas

Entidade	Descrição
Cabeçalho.PontoColeta	Ver seção Cabeçalho.
AmostraLarva	Tabela usada para registrar informações de amostras de larva de peixe.
ComposicaoDieta	Tabela usada para registro da composição da dieta dos peixes.
ConteudoEstomacalLUT	Lookup table referente ao conteúdo estomacal dos peixes. Usada para complementar dados de Ictiofauna.
EcologiaPeixe	Tabela usada para registrar informações da Ecologia do Peixe.
Ecotoxicologia	Tabela usada para registrar informações gerais de ecotoxicologia.
EspecielctiofaunaLUT	Lookup table referente à espécie do peixe (nome científico e vulgar). Usada para complementar dados de Ictiofauna.
FamilialctiofaunaLUT	Lookup table referente à família das larvas. Usada para complementar dados de Ictiofauna.
HabitatLUT	Lookup table referente ao habitat dos peixes. Usada para complementar dados de Ictiofauna.
MalhaLUT	Lookup table referente às malhas de pesca. Usada para definir o apetrecho utilizado na pesca..
Metal_Ion	Tabela usada para armazenar a contagem de metais e íons (Cadmio, Chumbo e Cobre para Metais. Cadmio, Cobre e Manganês para Ions)
MetallonLUT	Lookup table referente à metais ou íons. Usada para complementar dados de Ictiofauna.
OrdemlctiofaunaLUT	Lookup table referente à ordem das larvas. Usada para complementar dados de Ictiofauna.
Peixe	Tabela usada para registrar informações do Peixe.
Pesca	Tabela usada para registrar informações da coleta de peixes.
PosicaoLUT	Lookup table referente à posição onde são efetuadas as coletas. Usada para complementar dados de Ictiofauna.
RedeArrastoLUT	Lookup table referente às redes de arrasto. Usada para definir o apetrecho utilizado na pesca..
RegimeAlimentarLUT	Lookup table referente ao regime alimentar. Usada para complementar dados de Ictiofauna.
VariavelAgua	Tabela usada para registrar variáveis que servem para fazer análises físico-químicas da água. Atualmente são feitas duas coletas dessas variáveis para cada coleta de peixes.

AmostraLarva

Coluna	Tipo de dado	Descrição
IdPontoColeta	NUMBER(11, 0)	Identificador do ponto de coleta.
IdAmostraLarva	NUMBER(11, 0)	Identificador da amostra.
PeriodoDia	VARCHAR2(20)	Período do dia em que foi efetuada a coleta (diurno, noturno).
VolAguaFiltrada	NUMBER(5, 2)	Volume de água filtrada. unidade: m3 (metros cúbicos).
NumTotalLarvas	NUMBER(7, 0)	Número total de larvas (NTL).
NumTotalIndividuos	NUMBER(7, 0)	Número total de indivíduos por [ordem, gênero, família, espécie].
IdEspeciePesca	NUMBER(11, 0)	Identificador da espécie do peixe.
IdPosicao	NUMBER(11, 0)	Identificador da posição.
IdFamiliarctiofauna	NUMBER(11, 0)	Identificador da família do peixe.
IdOrdemlctiofauna	NUMBER(11, 0)	Identificador da ordem da larva.

ComposicaoDieta

Coluna	Tipo de dado	Descrição
IdConteudoEstomacal	NUMBER(11, 0)	Identificador do conteúdo estomacal.
IdPeixe	NUMBER(11, 0)	Identificador da coleta do peixe.
IdPontoColeta	NUMBER(11, 0)	Identificador do ponto de coleta.

ConteudoEstomacalLUT

Coluna	Tipo de dado	Descrição
IdConteudoEstomacal	NUMBER(11, 0)	Identificador do conteúdo estomacal.
ConteudoEstomacal	VARCHAR2(50)	Descrição do conteúdo estomacal.

EcologiaPeixe

Coluna	Tipo de dado	Descrição
IdPeixe	NUMBER(11, 0)	Identificador da coleta do peixe.
IdPontoColeta	NUMBER(11, 0)	Identificador do ponto de coleta.

Coluna	Tipo de dado	Descrição
IdRegimeAlimentar	NUMBER(11, 0)	Identificador do regime alimentar da ictiofauna.

Ecotoxicologia

Coluna	Tipo de dado	Descrição
IdPeixe	NUMBER(11, 0)	Identificador da coleta do peixe.
IdPontoColeta	NUMBER(11, 0)	Identificador do ponto de coleta.
CYP1A	NUMBER(13, 2)	CYP1A. Atributo de definição da Expressão Gênica. Unidade de medida: %
GST	NUMBER(3, 0)	GST. Atributo que define a atividade enzimática absoluta unidade: μ moles/min/gwt (micromoles por minuto por grama de tecido úmido)
MicroSatelite	VARCHAR2(500)	Descritivo das seqüências de DNA da espécie.

EspecieIctiofaunaLUT

Coluna	Tipo de dado	Descrição
IdEspecieIctiofauna	NUMBER(11, 0)	Identificador da espécie do peixe.
IdFamilialIctiofauna	NUMBER(11, 0)	Identificador da família do peixe.
NomeCientifico	VARCHAR2(100)	Nome da espécie do peixe.
NomeVulgar	VARCHAR2(100)	Nome vulgar da espécie do peixe.

FamiliaIctiofaunaLUT

Coluna	Tipo de dado	Descrição
IdFamilialIctiofauna	NUMBER(11, 0)	Identificador da família das larvas.
Familia	VARCHAR2(100)	Nome da família da larva.

HabitatLUT

Coluna	Tipo de dado	Descrição
IdHabitat	NUMBER(11, 0)	Identificador do habitat.
Habitat	VARCHAR2(100)	Descrição do habitat.

MalhaLUT

Coluna	Tipo de dado	Descrição
IdMalha	NUMBER(11, 0)	Identificador da malha de pesca.
Malha	VARCHAR2(20)	Tipo de malha de pesca.

Metal_Ion

Coluna	Tipo de dado	Descrição
IdPeixe	NUMBER(11, 0)	Identificador de coleta da ictiofauna.
IdMetallon	NUMBER(11, 0)	Identificador do metal ou íon.
IdPontoColeta	NUMBER(11, 0)	Identificador do ponto de coleta.
QtdeMusculo	NUMBER(15, 4)	Quantidade de Metal ou íon nas brânquias do peixe. unidade: µg/g tecido (micrograma por grama de tecido) unidade: micro grama/grTec
QtdeFigado	NUMBER(15, 4)	Quantidade do metal ou ion no fígado do peixe. unidade: µg/g tecido (micrograma por grama de tecido)
QtdeAgua	NUMBER(15, 4)	Quantidade do metal ou ion na água.. unidade: µg/g tecido (micrograma por grama de tecido)
TipoAnalise	VARCHAR2(1)	M-Metal I-Ion

MetallonLUT

Coluna	Tipo de dado	Descrição
IdMetallon	NUMBER(11, 0)	Identificador do metal ou íon.

Coluna	Tipo de dado	Descrição
Nome	VARCHAR2(15)	Nome do metal ou íon. Domínio: Cádmió, Chumbo e Cobre para Metais. Cádmió, Cobre e Manganês para Íons.
Ion	NUMBER(1, 0)	Indica se o elemento é um íon. Domínio: 0-não; 1-sim
Metal	NUMBER(1, 0)	Indica se o elemento é um metal. Domínio: 0-não; 1-sim

OrdemIctiofaunaLUT

Coluna	Tipo de dado	Descrição
IdOrdemIctiofauna	NUMBER(11, 0)	Identificador da ordem da larva.
Ordem	VARCHAR2(100)	Nome da ordem da Larva. Domínio: Perciforme, Siluriforme, Characiforme, Osteoglossiforme, Beloniforme, Clupeiforme, Gymnotiforme, Tetraodontiforme, Cypriniforme, Pleuronectiforme

Peixe

Coluna	Tipo de dado	Descrição
IdPeixe	NUMBER(11, 0)	Identificador da coleta do peixe.
IdPontoColeta	NUMBER(11, 0)	Identificador do ponto de coleta.
Codigo	NUMBER(11, 0)	Código seqüencial de identificação dos peixes coletados.
Peso	NUMBER(10, 2)	Peso do peixe. unidade: kg (quilograma)
Comprimento	NUMBER(10, 2)	Comprimento do peixe. unidade: cm (centímetro)
IdEspecieIctiofauna	NUMBER(11, 0)	Identificador da espécie do peixe.
IdHabitat	NUMBER(11, 0)	Identificador do habitat.

Pesca

Coluna	Tipo de dado	Descrição
IdPontoColeta	NUMBER(11, 0)	Identificador do ponto de coleta.

Coluna	Tipo de dado	Descrição
Horainicio	DATE	Hora de início da coleta (hh:mm).
HoraFim	DATE	Hora de término da coleta (hh:mm).
PosInicial	VARCHAR2(50)	Posição Inicial onde é jogado a rede de arrasto. Usado apenas para peixe bentônicos.
PosFinal	VARCHAR2(50)	Posição Final onde é recolhida a rede de arrasto. Usado apenas para peixe bentônicos.
TempoCronometroInicial	VARCHAR2(50)	Marcação do valor inicial do tempo para coleta de peixes bentônicos.
TempoCronometroFinal	VARCHAR2(50)	Marcação do valor final do tempo para coleta de peixes bentônicos.
Observacao	VARCHAR2(300)	Observações gerais da Pesca.
IdRedeArrasto	NUMBER(11, 0)	Identificador da rede de arrasto.
IdMalha	NUMBER(11, 0)	Identificador da malha de pesca.

PosicaoLUT

Coluna	Tipo de dado	Descrição
IdPosicao	NUMBER(11, 0)	Identificador da posição.
Posicao	VARCHAR2(50)	Posição onde foi efetuada a coleta: RCS (rio/centro/superfície); RCF (rio/centro/fundo); RMS (rio/margem/superfície); RMF (rio/margem/fundo); LCS (lago/centro/superfície); LCF (lago/centro/fundo); LMS (lago/margem/superfície); LMF (lago/margem/fundo).

RedeArrastoLUT

Coluna	Tipo de dado	Descrição
IdRedeArrasto	NUMBER(11, 0)	Identificador da rede de arrasto.
Altura	NUMBER(10, 2)	Altura da rede. Unidade: metros (m)
Comprimento	NUMBER(10, 2)	Comprimento da Rede. Unidade: metros (m)

RegimeAlimentarLUT

Coluna	Tipo de dado	Descrição
IdRegimeAlimentar	NUMBER(11, 0)	Identificador do regime alimentar da ictiofauna.
RegimeAlimentar	VARCHAR2(50)	Descrição do regime alimentar.

VariavelAgua

Coluna	Tipo de dado	Descrição
IdVariavelAgua	NUMBER(11, 0)	Identificador do ponto de coleta.
IdPontoColeta	NUMBER(11, 0)	Identificador do ponto de coleta.
Hora	DATE	Hora em que foi coletado as variáveis. Format: HH:MM
Profundidade	NUMBER(17, 2)	Profundidade em que foi coletado as variáveis. unidade: m (metro)
Condutividade	NUMBER(17, 2)	Condutividade. unidade: μ S25/cm (microsiemens a 25 graus Centígrados por centímetro)
Ph	NUMBER(17, 2)	Potencial hidrogeniônico (pH). domínio: 0 a 14.
Transparencia	NUMBER(17, 2)	Transparência da água. unidade: metros.
OxigSuperficie	NUMBER(17, 2)	Oxigênio de superfície. unidade: mgO2/l (miligrama de oxigênio por litro)
OxigProfundidade	NUMBER(17, 2)	Oxigênio de profundidade. unidade: mgO2/l (miligrama de oxigênio por litro)
TempSuperficie	NUMBER(17, 2)	Temperatura de Superfície. unidade: °C (grau Centígrado ou Celsius)
TempProfundidade	NUMBER(17, 2)	Temperatura de Profundidade. unidade: °C (grau Centígrado ou Celsius)

Anexo 02

Scripts de Criação de Tabelas no Banco de Dados

```
#=====
# Project Filename: C:\PIATAM\Estudos\Base de Dados\Modelo\Ictiofauna#
# Author: ALENCAR, R. S. #
# DBMS: SQL SERVER #
# Generated on: 20/06/2008 18:24:31 #
#=====
#
```

```
#=====
# Tables #
#=====
#
```

```
CREATE TABLE VariavelAgua (
  IdPontoColeta NUMERIC(11) NOT NULL,
  IdVariavelAgua NUMERIC(11) NOT NULL,
  Hora DATE,
  Profundidade NUMERIC(17,2),
  Condutividade NUMERIC(17,2),
  Ph NUMERIC(17,2),
  Transparencia NUMERIC(17,2),
  OxiSuperficie NUMERIC(17,2),
  OxiProfundidade NUMERIC(17,2),
  TempSuperficie NUMERIC(17,2),
  TempProfundidade NUMERIC(17,2),
  PRIMARY KEY (IdPontoColeta, IdVariavelAgua)
);
```

```
CREATE TABLE Pesca (
  IdPontoColeta NUMERIC(11) NOT NULL,
  IdPesca NUMERIC(11) NOT NULL,
  HoraInicio DATE,
  HoraFim DATE,
  PosInicial VARCHAR(50),
  PosFinal VARCHAR(50),
  TempoCronometroInicial VARCHAR(50),
  TempoCronometroFinal VARCHAR(50),
  Observacao VARCHAR(300),
  IdRedeArrasto NUMERIC(11),
```

```

    IdMalha NUMERIC(11),
    PRIMARY KEY (IdPontoColeta)
);

CREATE TABLE RedeArrastoLUT (
    IdRedeArrasto NUMERIC(11) NOT NULL,
    Altura NUMERIC(10,2),
    Comprimento NUMERIC(10,2),
    PRIMARY KEY (IdRedeArrasto)
);

CREATE TABLE MalhaLUT (
    IdMalha NUMERIC(11) NOT NULL,
    Malha VARCHAR(20),
    PRIMARY KEY (IdMalha)
);

CREATE TABLE CabecalhoPontoColeta (
    IdPontoColeta NUMERIC(11) NOT NULL,
    IdMarco NUMERIC(11),
    Latitude NUMERIC(13,4),
    Longitude NUMERIC(13,4),
    DataColeta DATE,
    Referencial VARCHAR(200),
    IdExcursao NUMERIC(11),
    IdLocalColeta NUMERIC(11),
    PRIMARY KEY (IdPontoColeta)
);

CREATE TABLE AmostraLarva (
    IdAmostraLarva NUMERIC(11) NOT NULL,
    PeriodoDia VARCHAR(20),
    VolAguaFiltrada NUMERIC(5,2),
    NumTotalLarvas NUMERIC(7),
    NumTotalIndividuos NUMERIC(7),
    IdPontoColeta NUMERIC(11) NOT NULL,
    IdOrdemlctiofauna NUMERIC(11),
    IdPosicao NUMERIC(11),
    IdFamiliarlctiofauna NUMERIC(11),
    IdEspecielctiofauna NUMERIC(11),
    PRIMARY KEY (IdAmostraLarva, IdPontoColeta)
);

CREATE TABLE OrdemlctiofaunaLUT (
    IdOrdemlctiofauna NUMERIC(11) NOT NULL,
    Ordem VARCHAR(100),
    PRIMARY KEY (IdOrdemlctiofauna)
);

CREATE TABLE PosicaoLUT (
    IdPosicao NUMERIC(11) NOT NULL,
    Posicao VARCHAR(50),
    PRIMARY KEY (IdPosicao)
);

```

```

CREATE TABLE FamiliarctiofaunaLUT (
  IdFamiliarctiofauna NUMERIC(11) NOT NULL,
  Familia VARCHAR(100),
  PRIMARY KEY (IdFamiliarctiofauna)
);

CREATE TABLE EspecielctiofaunaLUT (
  IdEspecielctiofauna NUMERIC(11) NOT NULL,
  NomeCientifico VARCHAR(100),
  NomeVulgar VARCHAR(100),
  IdFamiliarctiofauna NUMERIC(11),
  PRIMARY KEY (IdEspecielctiofauna)
);

CREATE TABLE Peixe (
  IdPeixe NUMERIC(11) NOT NULL,
  IdPontoColeta NUMERIC(11) NOT NULL,
  Codigo NUMERIC(11),
  Peso NUMERIC(10,2),
  Comprimento NUMERIC(10,2),
  IdEspecielctiofauna NUMERIC(11),
  IdHabitat NUMERIC(11),
  PRIMARY KEY (IdPeixe, IdPontoColeta)
);

CREATE TABLE HabitatLUT (
  IdHabitat NUMERIC(11) NOT NULL,
  Habitat VARCHAR(100),
  PRIMARY KEY (IdHabitat)
);

CREATE TABLE ConteudoEstomacalLUT (
  IdConteudoEstomacal NUMERIC(11) NOT NULL,
  ConteudoEstomacal VARCHAR(50),
  PRIMARY KEY (IdConteudoEstomacal)
);

CREATE TABLE ComposicaoDieta (
  IdConteudoEstomacal NUMERIC(11) NOT NULL,
  IdPeixe NUMERIC(11) NOT NULL,
  IdPontoColeta NUMERIC(11) NOT NULL,
  PRIMARY KEY (IdConteudoEstomacal, IdPeixe, IdPontoColeta)
);

CREATE TABLE EcologiaPeixe (
  IdRegimeAlimentar NUMERIC(11),
  IdPeixe NUMERIC(11) NOT NULL,
  IdPontoColeta NUMERIC(11) NOT NULL,
  PRIMARY KEY (IdPeixe, IdPontoColeta)
);

CREATE TABLE RegimeAlimentarLUT (
  IdRegimeAlimentar NUMERIC(11) NOT NULL,
  RegimeAlimentar VARCHAR(50),
  PRIMARY KEY (IdRegimeAlimentar)
);

```

```
);
CREATE TABLE Ecotoxicologia (
  CYP1A NUMERIC(13,2),
  GST NUMERIC(3),
  MicroSatelite VARCHAR(500),
  IdPeixe NUMERIC(11) NOT NULL,
  IdPontoColeta NUMERIC(11) NOT NULL,
  PRIMARY KEY (IdPeixe, IdPontoColeta)
);
```

```
CREATE TABLE Metal_Ion (
  QtdeMusculo NUMERIC(15,4),
  QtdeFigado NUMERIC(15,4),
  QtdeAgua NUMERIC(15,4),
  TipoAnalise VARCHAR(1),
  IdMetallon NUMERIC(11) NOT NULL,
  IdPeixe NUMERIC(11) NOT NULL,
  IdPontoColeta NUMERIC(11) NOT NULL,
  PRIMARY KEY (IdMetallon, IdPeixe, IdPontoColeta)
);
```

```
CREATE TABLE Metallon_LUT (
  IdMetallon NUMERIC(11) NOT NULL,
  Nome VARCHAR(15),
  Ion NUMERIC(1),
  Metal NUMERIC(1),
  PRIMARY KEY (IdMetallon)
);
```

```
#=====
# Foreign Keys #
#=====
#=====
```

```
ALTER TABLE VariavelAgua
  ADD FOREIGN KEY (IdPontoColeta) REFERENCES Pesca (IdPontoColeta);
```

```
ALTER TABLE Pesca
  ADD FOREIGN KEY (IdRedeArrasto) REFERENCES RedeArrastoLUT
  (IdRedeArrasto);
```

```
ALTER TABLE Pesca
  ADD FOREIGN KEY (IdMalha) REFERENCES MalhaLUT (IdMalha);
```

```
ALTER TABLE Pesca
  ADD FOREIGN KEY (IdPontoColeta) REFERENCES CabecalhoPontoColeta
  (IdPontoColeta);
```

```
ALTER TABLE AmostraLarva
  ADD FOREIGN KEY (IdPontoColeta) REFERENCES CabecalhoPontoColeta
  (IdPontoColeta);
```

```
ALTER TABLE AmostraLarva
```

```

    ADD FOREIGN KEY (IdOrdemlctiofauna) REFERENCES OrdemlctiofaunaLUT
    (IdOrdemlctiofauna);

ALTER TABLE AmostraLarva
    ADD FOREIGN KEY (IdPosicao) REFERENCES PosicaoLUT (IdPosicao);

ALTER TABLE AmostraLarva
    ADD FOREIGN KEY (IdFamiliarlctiofauna) REFERENCES FamiliarlctiofaunaLUT
    (IdFamiliarlctiofauna);

ALTER TABLE AmostraLarva
    ADD FOREIGN KEY (IdEspecielctiofauna) REFERENCES EspecielctiofaunaLUT
    (IdEspecielctiofauna);

ALTER TABLE EspecielctiofaunaLUT
    ADD FOREIGN KEY (IdFamiliarlctiofauna) REFERENCES FamiliarlctiofaunaLUT
    (IdFamiliarlctiofauna);

ALTER TABLE Peixe
    ADD FOREIGN KEY (IdEspecielctiofauna) REFERENCES EspecielctiofaunaLUT
    (IdEspecielctiofauna);

ALTER TABLE Peixe
    ADD FOREIGN KEY (IdHabitat) REFERENCES HabitatLUT (IdHabitat);

ALTER TABLE Peixe
    ADD FOREIGN KEY (IdPontoColeta) REFERENCES Pesca (IdPontoColeta);

ALTER TABLE ComposicaoDieta
    ADD FOREIGN KEY (IdConteudoEstomacal) REFERENCES
    ConteudoEstomacalLUT (IdConteudoEstomacal);

ALTER TABLE ComposicaoDieta
    ADD FOREIGN KEY (IdPeixe,IdPontoColeta) REFERENCES EcologiaPeixe
    (IdPeixe, IdPontoColeta);

ALTER TABLE EcologiaPeixe
    ADD FOREIGN KEY (IdRegimeAlimentar) REFERENCES RegimeAlimentarLUT
    (IdRegimeAlimentar);

ALTER TABLE EcologiaPeixe
    ADD FOREIGN KEY (IdPeixe,IdPontoColeta) REFERENCES Peixe (IdPeixe,
    IdPontoColeta);

ALTER TABLE Ecotoxicologia
    ADD FOREIGN KEY (IdPeixe,IdPontoColeta) REFERENCES Peixe (IdPeixe,
    IdPontoColeta);

ALTER TABLE Metal_Ion
    ADD FOREIGN KEY (IdMetallon) REFERENCES Metallon_LUT (IdMetallon);

ALTER TABLE Metal_Ion
    ADD FOREIGN KEY (IdPeixe,IdPontoColeta) REFERENCES Ecotoxicologia
    (IdPeixe, IdPontoColeta);

```

Anexo 03

Scripts de Povoamento do Banco de Dados

```
BULK
INSERT Metallon_LUT
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_metalionlut.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
```

go

```
BULK
INSERT RegimeAlimentarLUT
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_regimealimentarlut.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
```

go

```
BULK
INSERT HabitatLUT
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_habitatlut.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
```

go

```
BULK
INSERT MalhaLUT
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_malhalut.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
```

```

)
go

BULK
INSERT OrdemIctiofaunaLUT
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_ordemictiofaunalut.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

BULK
INSERT PosicaoLUT
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_posicaolut.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

BULK
INSERT FamiliarctiofaunaLUT
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_familiaictiofaunalut.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

BULK
INSERT RedeArrastoLUT
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_redearrastolut.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

BULK

```



```

INSERT ConteudoEstomacaLUT
  FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/lctiofauna/ict_conteudoestomacallut.csv'
  WITH
  (
    FIELDTERMINATOR = ';',
    ROWTERMINATOR = '\n'
  )

```

go

```

BULK
INSERT EspecielctiofaunaLUT
  FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/lctiofauna/ict_especieictiofaunalut.csv'
  WITH
  (
    FIELDTERMINATOR = ';',
    ROWTERMINATOR = '\n'
  )

```

go

```

BULK
INSERT Pesca
  FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/lctiofauna/ict_pesca.csv'
  WITH
  (
    FIELDTERMINATOR = ';',
    ROWTERMINATOR = '\n'
  )

```

go

```

BULK
INSERT Peixe
  FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/lctiofauna/ict_peixe.csv'
  WITH
  (
    FIELDTERMINATOR = ';',
    ROWTERMINATOR = '\n'
  )

```

go

```

BULK
INSERT AmostraLarva
  FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/lctiofauna/ict_amostralarva.csv'
  WITH
  (

```

```

        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

BULK
INSERT ComposicaoDieta
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
    Dados/DadosRicardo/CSVs/Ictiofauna/ict_composicaodieta.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

BULK
INSERT EcologiaPeixe
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
    Dados/DadosRicardo/CSVs/Ictiofauna/ict_ecologiapeixe.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

BULK
INSERT Ecotoxicologia
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
    Dados/DadosRicardo/CSVs/Ictiofauna/ict_ecotoxicologia.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

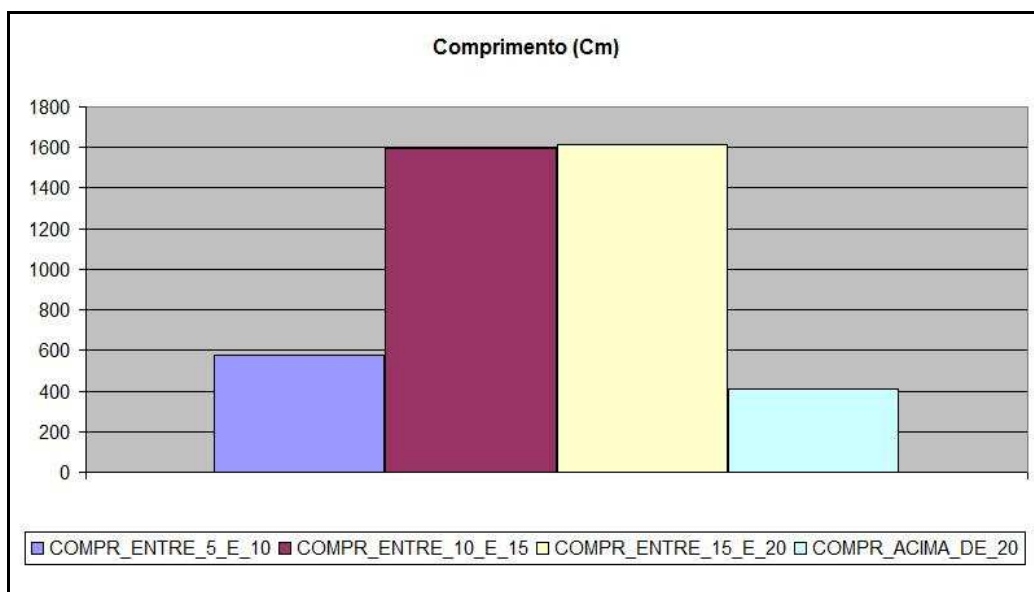
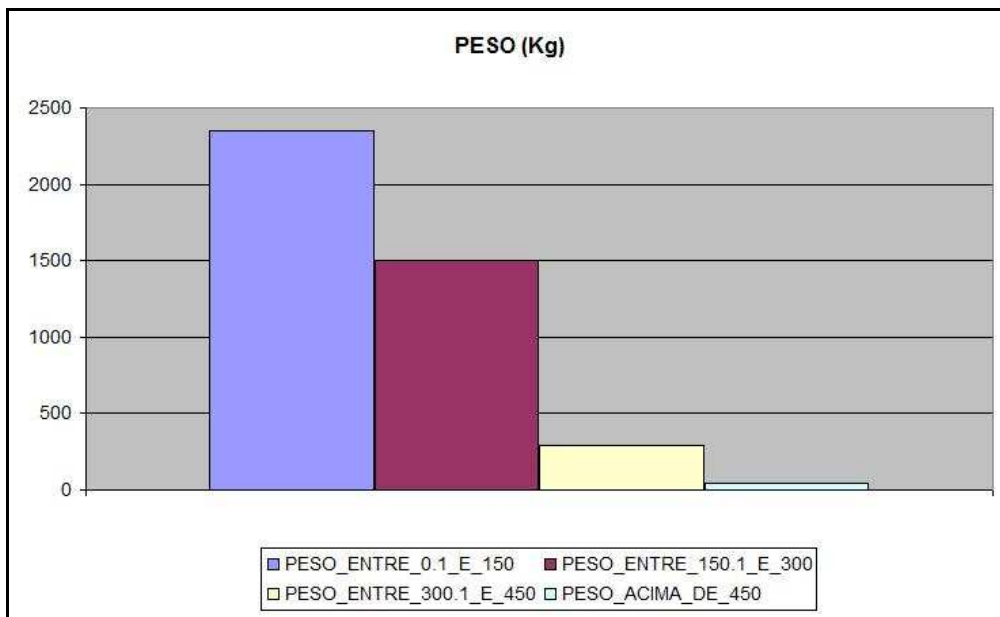
BULK
INSERT Metal_Ion
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
    Dados/DadosRicardo/CSVs/Ictiofauna/ict_metal_ion.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
go

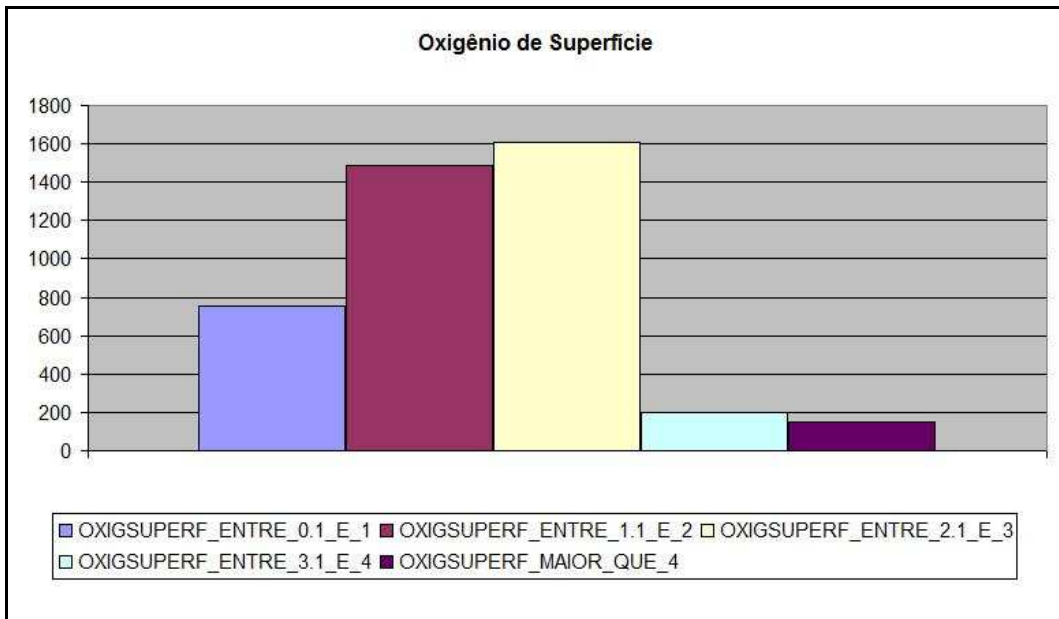
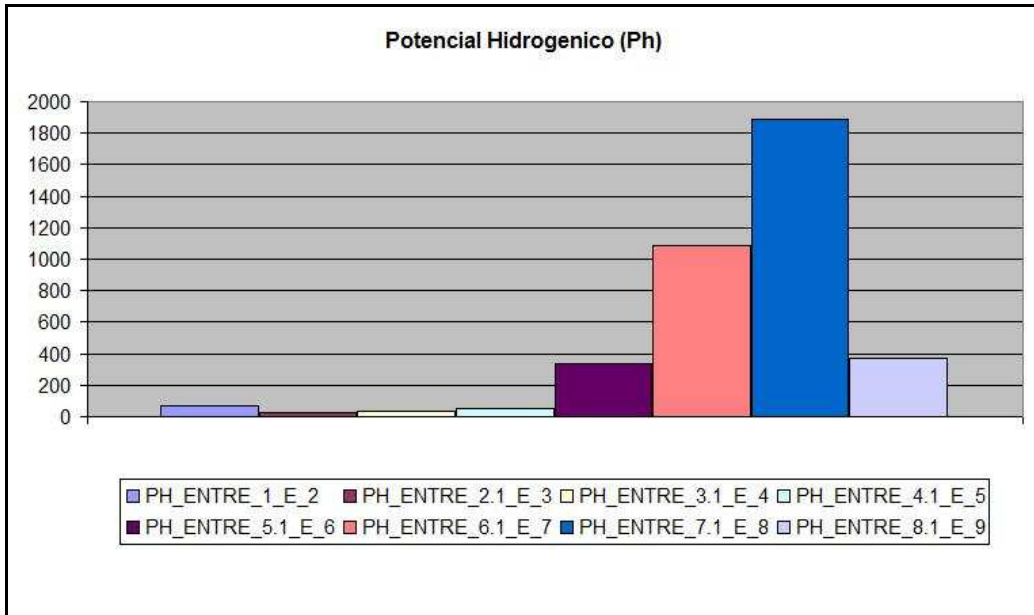
```

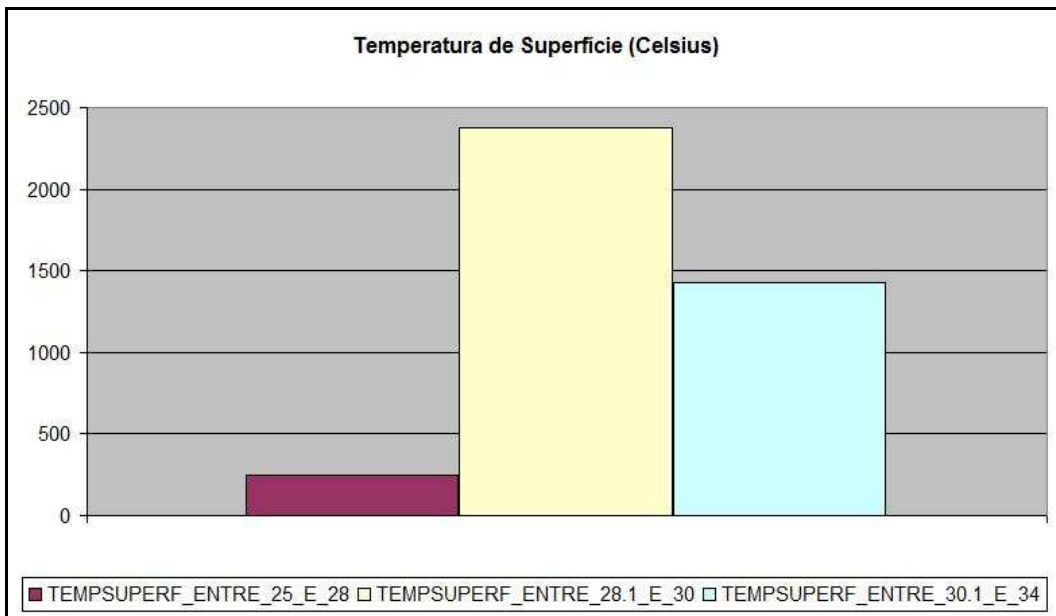
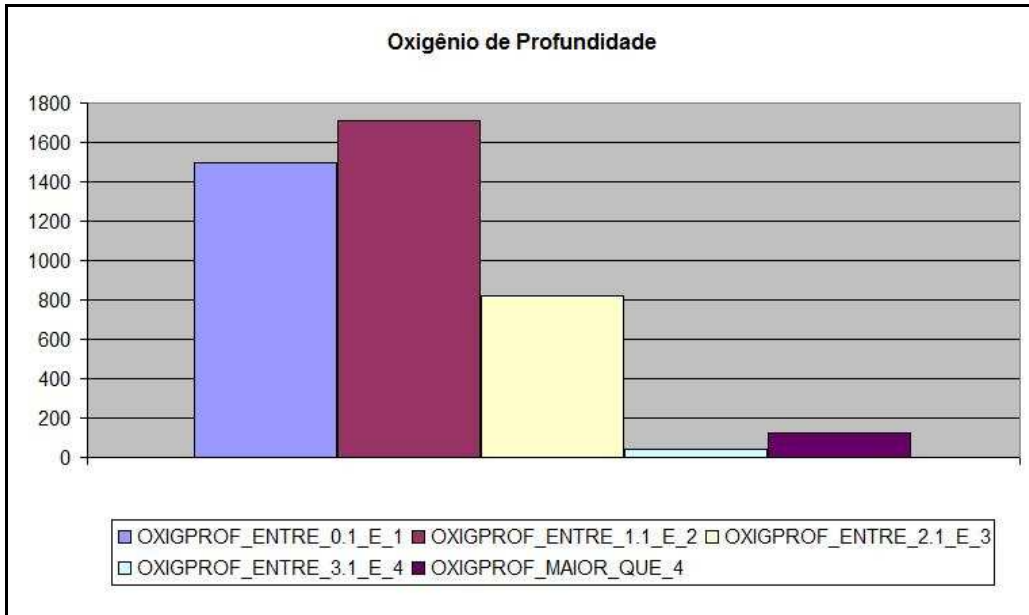
```
BULK
INSERT VariavelAgua
    FROM 'c:/Ricardo/Mestrado/PIATAM/Estudos/Base de
Dados/DadosRicardo/CSVs/Ictiofauna/ict_variavelagua.csv'
    WITH
    (
        FIELDTERMINATOR = ';',
        ROWTERMINATOR = '\n'
    )
```

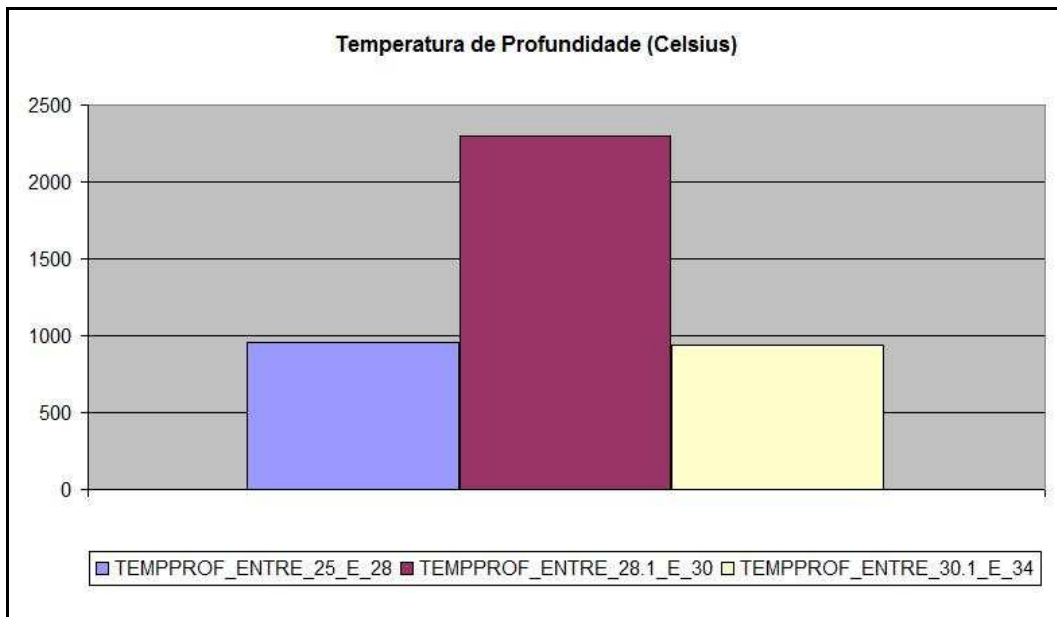
Anexo 04

Histogramas dos Atributos Utilizados para Geração das Regras de Associação









Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)