



COPPE/UFRJ

APLICAÇÃO DE FERRAMENTAS ROLAP EM BANCO DE DADOS AMBIENTAL
NA ZONA COSTEIRA AMAZÔNICA

Luiz Fernando Machado de Oliveira e Souza

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: Nelson Francisco Favilla Ebecken

Rio de Janeiro
Abril de 2009

Livros Grátis

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

Milhares de livros grátis para download.

APLICAÇÃO DE FERRAMENTAS ROLAP EM BANCO DE DADOS AMBIENTAL
NA ZONA COSTEIRA AMAZÔNICA

Luiz Fernando Machado de Oliveira e Souza

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) 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. Nelson Francisco Favilla Ebecken, D. Sc.

Prof. Luiz Landau, D.Sc.

Prof. Fernando Pellon de Miranda, D.Sc.

Prof. Valéria Menezes Bastos, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

ABRIL DE 2009

Souza, Luiz Fernando Machado de Oliveira e

Aplicação de Ferramentas ROLAP em Banco de Dados Ambiental na Zona Costeira Amazônica/Luiz Fernando Machado de Oliveira e Souza. – Rio de Janeiro: UFRJ/COPPE, 2009.

XII, 85 p.: 29,7 cm.

Orientador: Nelson Francisco Favilla Ebecken

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

Referencias Bibliográficas: p. 81-85.

1. Mineração de Dados. 2. Ferramentas OLAP. 3. Descoberta de Conhecimento. 4. Classificação. 5. Agrupamento. I. Ebecken, Nelson Francisco Favilla. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

Agradecimentos

Em primeiro lugar, agradeço enormemente a Deus, aos meus pais (em memória), pelos ensinamentos que semearam e que perduram até hoje e para sempre, a minha mãe Iá que me criou, amou e ama por toda sua vida, aos meus irmãos (em especial ao Pedro e Cezar) e a minha tia Teca pelo apoio e carinho incondicionais.

Agradeço à Cris minha amiga, namorada, mulher, pela paciência, apoio, e por ter transformado todo o difícil e longo caminho em algo mais prazeroso de ser vivido.

Agradeço ao meu orientador, Prof. Nelson Francisco Favilla Ebecken, pela tranquilidade transmitida, mesmo nos momentos de quase desespero. Obrigado pela orientação e incentivo em todos os momentos, fazendo com que todas as dúvidas se esclarecessem de maneira clara e objetiva.

Aos meus cinco sobrinhos lindos, cada um com seu encanto, pelos momentos de pureza e alegria incomparáveis.

A toda a família LAIT (Laboratório de Análise de Imagens do Trópico Úmido / UFPA) e em especial ao coordenador Prof. Dr. Pedro Walfir Martins e Souza Filho, meu irmão amado e maior incentivador na busca desse título.

Agradeço por tudo ao meu grande amigo Alexandre Silva. Dentre tantos outros que de uma forma ou de outra contribuíram para o êxito desta jornada.

A Petrobras pelo Projeto PIATAM mar que foi o facilitador de toda essa realização.

Ao professor Fernando Pellon de Miranda pela confiança depositada nos projetos da rede PIATAM, na Amazônia e em mim.

Ao professor Luiz Landau e toda equipe do LAMCE (Laboratório de Métodos Computacionais em Engenharia) e ao Henrico e toda a equipe do LABSAR, em especial ao técnico Humberto Buffoni.

Agradeço aos amigos e companheiros de estudo Jorge Morgado, Bruno Vilela, Alexandre Politano, Cristiano Mariotti e Jonatas Vasconcelos, por transformarem o trabalho em algo mais prazeroso e principalmente, pelo conhecimento compartilhado.

Agradeço também, aos funcionários do programa pelo apoio a educação e pela excelência no atendimento.

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 FERRAMENTAS ROLAP EM BANCO DE DADOS AMBIENTAL
NA ZONA COSTEIRA AMAZÔNICA

Luiz Fernando Machado de Oliveira e Souza

Abril/2009

Orientador: Nelson Francisco Favilla Ebecken

Programa: Engenharia Civil

Este trabalho apresenta uma proposta metodológica na análise de uma base de dados relacional, com dados ambientais da zona costeira amazônica visando principalmente à busca de conhecimento escondido, bem como, a ratificação e/ou retificação de conhecimentos existentes. Na realização do estudo foram utilizadas técnicas de mineração de dados como classificação e agrupamento com ferramentas de *Data Warehouse* e *OLAP*, idealizando uma maior interatividade dos pesquisadores com as informações contidas na base de dados do Projeto PIATAM mar. Os resultados experimentais observados contribuíram com o conhecimento pré-existente e despertaram para um estudo futuro mais aprofundado e diversificado.

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

APPLICATION OF ROLAP TOOLS IN ENVIRONMENTAL DATABASE IN THE
AMAZON COASTAL ZONE

Luiz Fernando Machado de Oliveira e Souza

Abril/2009

Advisor: Nelson Francisco Favilla Ebecken

Department: Civil Engineering

This work presents a methodological research to database relational analysis of environmental data of the Amazon coastal zone. The main aim is to get the hidden knowledge, as well as, the ratification and/or rectification of the existing knowledge. During the development of this work, it was used data mining techniques, as classification and cluster analysis with data warehouse tools and OLAP. This way, we intended to get a greater interaction between researchers with the information storage in the PIATAM marine database. The experimental results contributed with the pre-existing knowledge and awoke to a future more detailed and diverse.

Título

APLICAÇÃO DE FERRAMENTAS ROLAP EM BANCO DE DADOS AMBIENTAL NA ZONA COSTEIRA AMAZÔNICA

Sumário

1. Introdução	01
1.1. Motivação	02
1.2. Objetivo	02
1.3. Estado da Arte	02
1.4. Relevância	02
1.5. Descrição do Trabalho	03
2. Contexto Teórico	04
2.1. O Projeto Piatam Mar	04
2.2. Banco de Dados	05
2.2.1.1. Contextualização	05
2.3. Data Warehouse	07
2.3.1. Arquitetura	08
2.3.1.1. Armazenamento	08
2.3.1.2. Modelagem Multidimensional	08
2.3.1.3. Metadado	09
2.3.1.3.1. Tipo de Informação considerada Metadado	09
2.4. OLAP	11
2.4.1. Funcionamento	11
2.4.2. Tipos de OLAP	12
2.4.2.1. Multidimensional (MOLAP)	12
2.4.2.2. Relacional (ROLAP)	12
2.4.2.3. Híbrido	12
2.4.2.4. Comparação	12
2.5. Mineração de Dados (Data Mining)	15
2.5.1. Classificação	16
2.5.2. Agrupamento	16
2.5.2.1. Métodos de Agrupamento	17
2.5.2.1.1. K-means	17
2.5.2.1.2. Método Hierárquico	17
3. Metodologia Proposta	18
3.1. Análise da base de dados	19
3.2. Exportação/Importação da Base de Dados	19
3.3. Construção do DW	19

3.3.1. Operacionalização da ferramenta ETL (OWB - Oracle Warehouse Builder)	19
3.4. Construção do OLAP	21
3.5. Análise dos Resultados	21
4. Ambiente Computacional	22
4.1. Arquitetura do Projeto	22
4.1.1. Banco de Dados Oracle 10g	22
4.1.2. Ferramenta ERwin Data Modeler 4.1	22
4.1.3. SQL Navigator 5.0	24
4.1.4. Oracle Warehouse Builder 11g	24
4.2. Construção do Modelo do Projeto DW Piatam Mar	25
4.2.1. Modelagem do DW	25
4.2.2. Importação do Metadado	26
4.2.3. Mapeamentos e Transformações do DW	28
5. Estudo de Caso e Resultados	32
5.1. Objetivo	32
5.2. Identificação dos Fatos de Negócio	32
5.2.1. Fato Coleta de Dados	32
5.2.1.1. Medidas	33
5.2.1.2. Dimensões	34
5.2.1.2.1. Localização	34
5.2.1.2.2. Tema	34
5.2.1.2.3. Período	34
5.2.1.3. Modelo Relacional	34
5.2.2. Fato Coleta Química da Água	37
5.2.2.1. Medidas	37
5.2.2.2. Dimensões	38
5.2.2.2.1. Localização	38
5.2.2.2.2. Tempo	38
5.2.2.3. Modelo Relacional	38
5.2.3. Fato Coleta Climatologia	39
5.2.3.1. Medidas	39
5.2.3.2. Dimensões	39
5.2.3.2.1. Localização	39
5.2.3.2.2. Tempo	39
5.2.3.3. Modelo Relacional	39
5.2.4. Fato Coleta Geologia	40
5.2.4.1. Medidas	40

5.2.4.2.	Dimensões	41
5.2.4.2.1.	Localização	41
5.2.4.2.2.	Tempo	41
5.2.4.2.3.	Granularidade	41
5.2.4.2.4.	Seleção	41
5.2.4.2.5.	Curtose	41
5.2.4.2.6.	Assimetria	42
5.2.4.3.	Modelo Relacional	42
5.2.5.	Fato Coleta Peixes Estuarinos	43
5.2.5.1.	Medidas	43
5.2.5.2.	Dimensões	43
5.2.5.2.1.	Localização	43
5.2.5.2.2.	Tempo	44
5.2.5.2.3.	Utilização Habitat	44
5.2.5.2.4.	Sexo	44
5.2.5.2.5.	Nome Científico	44
5.2.5.2.6.	Família	44
5.2.5.2.7.	Espécie	44
5.2.5.3.	Modelo Relacional	44
5.3.	Operacionalização do Data Warehouse	45
5.3.1.	Oracle Data Mart Designer	46
5.3.2.	Oracle Warehouse Builder	46
5.3.3.	BaseViews e MetaViews	46
5.3.4.	Oracle Interprise Edition	47
5.3.5.	Oracle Web Application Server	47
5.3.6.	Oracle Reposts	47
5.4.	Análise de resultados	48
5.4.1.	Peixes Estuarinos	50
5.4.2.	Climatologia	51
5.4.3.	Química da Água	52
5.4.4.	Relação entre todas as Dimensões do Cubo	53
6.	Conclusões e Trabalhos Futuros	53
6.1.	Conclusões	54
6.2.	Trabalhos Futuros	55
	Referências Bibliograficas	56

Lista de Figuras

Figura 2.1: Área de abrangência do Projeto PIATAM Mar	04
Figura 2.2: Fluxo de informações da base de dados do Projeto PIATAM Mar	07
Figura 2.3: Modelo multidimensional	09
Figura 2.4: Cubo lógico de dados onde as arestas representam as dimensões	15
Figura 3.1: Etapas da metodologia proposta	18
Figura 3.2: Espaço de trabalho do Design Center	20
Figura 4.1: página do modelo FatoColetaPeixesEstuarinos	23
Figura 4.2: recurso de Engenharia Reversa do ERWin	23
Figura 4.3: recursos de Engenharia Remissiva do ERWin	24
Figura 4.4: Espaço de Trabalho do Design Center	25
Figura 4.5: fatos principais do Projeto DW Piatam Mar	26
Figura 4.6: criação do módulo que receberá os objetos importados	27
Figura 4.7: processo de importação dos objetos	27
Figura 4.8: criação de uma sequence	28
Figura 4.9: disponibilização de uma sequence	28
Figura 4.10: mapeamento do arquivo externo	29
Figura 4.11: criação da tabela externa TB_ASSIMETRIA	30
Figura 4.12: Mapeamento Editor configurando a dimensão DM_ASSIMETRIA	31
Figura 5.1: Modelo coleta de dados	33
Figura 5.2: Modelo coleta química da água	37
Figura 5.3: Modelo coleta climatologia	38
Figura 5.4: Modelo coleta geologia	41
Figura 5.5: Modelo coleta peixes estuarinos	44
Figura 5.6: Ambiente do Design Center	47
Figura 5.7: Visualizador de Dados Relacional: DM_LOCALIZAÇÃO	48
Figura 5.8: Visualizador de Dados Relacional: DM_PERIODO	48
Figura 5.9: Visualizador de Dados Relacional: DM_TEMA	48
Figura 5.10: Visão do SQL Navigator Fato Coleta Peixes Estuarinos	49
Figura 5.11: Visão do SQL Navigator Fato Climatologia	50
Figura 5.12: Visão do SQL Navigator Fato Química da Água	51
Figura 5.13: Visão do SQL Navigator com todas as dimensões do cubo	52

Lista de Tabelas

Tabela 5.1: Visão I: Coleta de dados	34
Tabela 5.2: Visão II: Coleta de dados	35

Lista de Anexos

Dicionário de Dados	59
1. PONTO DE COLETA	59
Cadastro Doc – Cartográfica	59
Documentação	59
Cadastro Excursão	59
Cadastro Marco - Piatam – Mar	59
Cadastro Período – Hidrológico	60
Ponto de coleta	60
Cadastro Sazonalidade Anual	60
2. ABIÓTICO	61
Cadastro Abiótico – Climatologia	61
Cadastro Abiótico – Executor	62
Cadastro Abiótico – Geofísica	62
Cadastro Abiotico – Geologia	63
Cadastro Abiótico - Geomorfologia - Ambiente – Estuarino	64
Cadastro Abiotico - Geomorfologia - Ambiente – Fluvial	67
Cadastro Abiótico - Geomorfologia - Ambiente – Lagos	70
Cadastro Abiotico - Geomorfologia - Ambiente - Linha – Costa	72
Cadastro Abiotico - Geomorfologia - Ambiente – Plataforma	72
Cadastro Abiotico - Geomorfologia - Ambiente – Praia	74
Cadastro Abiotico - Geomorfologia - Ambiente – Recifes	77
Cadastro Abiótico - Geomorfologia - Ambiente – Talude	78
3. BIÓTICO	79
Cadastro Biótico – Peixes Estuarinos	79

1. Introdução

O grande volume de informações e a multidisciplinaridade das bases de dados ambientais, aliado à interdisciplinaridade e ao anseio pela descoberta de conhecimento contido nesses bancos de dados consome valioso tempo de especialistas de diversas áreas do conhecimento, tanto na alimentação das bases como na busca de informações.

Com isso, o processo convencional de consultas criadas a partir apenas da visão do especialista faz com que se obtenham resultados limitados ao conhecimento individualizado e/ou compartilhado desses pesquisadores, não possibilitando a descoberta de conhecimento “escondido” no grande volume de dados propiciando maior interatividade contida na diversidade de informação disponível.

Para atender este novo contexto, surge uma nova área denominada Descoberta de Conhecimento em Bases de Dados (Knowledge Discovery in Databases – KDD), que vem despertando grande interesse junto às comunidades científica e industrial. A expressão Mineração de Dados (Data Mining), mais popular, é, na realidade, uma das etapas da Descoberta de Conhecimento em Bases de Dados [Passos & Goldschmidt 2005].

Uma base de dados geográfica possibilita a interação de diversas ciências a cerca da mesma área de pesquisa, seja ela pontual ou uma área de abrangência considerável em relação ao projeto.

Os bancos de dados ambientais tem em comum, uma variedade de “temas” em estudo e pouca associação realizada entre eles. Em alguns casos é de extrema importância buscar essa interação varrendo a base com um todo. Desta maneira, o resultado pode ser analisado e validado por especialistas ou mesmo por um comitê de pesquisadores dos temas e área de estudo envolvidos.

Além disso, faz-se necessário o estudo e o planejamento adequado da modelagem dos dados, bem como, a utilização de ferramentas OLAP (*On Line Analytical Processing*), com o intuito de gerar a maior quantidade de pesquisas possível na base, processando e analisando de forma rápida o banco de dados.

A tecnologia OLAP fornece uma forma de análise simples, permitindo fatiar a informação em subconjuntos de dados ou cubos inter-relacionados. Desta forma, pode-se analisar os dados usando recursos OLAP padrão, tais como *page-by*, *pivot*, classificação, filtro e *drill up/down* para ter acesso a uma série de visões.

Essas visões são o principal objetivo desse trabalho, pois através delas e da análise dos especialistas, busca-se a descoberta e a validação de conhecimento contido no banco de dados.

1.1. Motivação

Estudar a importância e diferenciação da modelagem espacial de dados ambientais interdisciplinares no escopo do projeto PIATAM mar (Potenciais Impactos Ambientais no Transporte de Petróleo e Derivados na Zona Costeira Amazônica), que abrange toda a costa norte do Brasil e busca avaliar potenciais impactos ambientais na região.

Buscar conhecimento escondido dentro da base de dados com a utilização de ferramentas OLAP na mineração dos dados.

1.2. Objetivo

O objetivo principal deste trabalho é estudar a importância da utilização de técnicas computacionais de descoberta de conhecimento em bases de dados ambientais, visando classificar e/ou agrupar dados buscando informações consistentes entre diversas áreas do conhecimento agregadas a uma determinada localização geográfica na zona costeira amazônica.

1.3. Estado da Arte

Mineração de dados é um conjunto de técnicas computacionais dentro de uma grande área denominada Descoberta do Conhecimento em Bases de Dados, também conhecida por KDD (*Knowledge Discovery in Databases*). Estas técnicas são empregadas em diversas áreas de estudo com uma aplicação maior em *Marketing* e em processos decisórios empresariais. Em relação a aplicação de ferramentas OLAP e *Data Warehouse* (DW), não foi encontrado nenhum trabalho aplicado em banco de dados ambientais na zona costeira amazônica.

1.4. Relevância

A contribuição deste trabalho está baseada na exploração e descoberta de conhecimento em bases de dados ambientais na costa amazônica. Sua importância persiste na análise de aplicações OLAP e DW objetivando a busca de conhecimentos, sejam eles relevantes e/ou inesperados ou ainda a ratificação ou retificação dos mesmos.

1.5. Descrição do Trabalho

Este trabalho está dividido em seis capítulos da seguinte forma:

- O capítulo 1 apresenta a introdução, objetivo, motivação, estado da arte e relevância que serviram como base para a pesquisa da modelagem de dados e aplicação de ferramentas OLAP e DW no banco de dados;
- O segundo capítulo apresenta os conceitos em que se baseia a base de dados, os processos de KDD utilizados e as ferramentas aplicadas, bem como, a escolha dos temas do banco de dados a serem relacionados e estudados nesta dissertação;
- O capítulo 3 trata da metodologia proposta e define as etapas para o desenvolvimento do trabalho;
- O quarto capítulo trata do ambiente computacional, descrevendo suas funcionalidades e implementação;
- No capítulo 5, são identificados alguns estudos de casos realizados, o que demonstra a aplicação da ferramenta e os resultados obtidos;
- Finalmente, o capítulo 6 apresenta as considerações finais, identificando as conclusões do trabalho e sugestões de desenvolvimento futuro que pode ser acrescentado.

2. Contexto Teórico

2.1. O Projeto Piatam mar

Este projeto é uma grande parceria entre o setor produtivo e a comunidade científica para otimizar a gestão ambiental dos negócios da Petrobras na Zona Costeira Amazônica.

São mais de 150 pesquisadores, técnicos e bolsistas investigando uma área de 1.100km de extensão, da baía de São Marcos (MA) até o Cabo Orange (AP), onde se localizam as capitais dos três estados envolvidos no Projeto – Amapá, Pará e Maranhão.

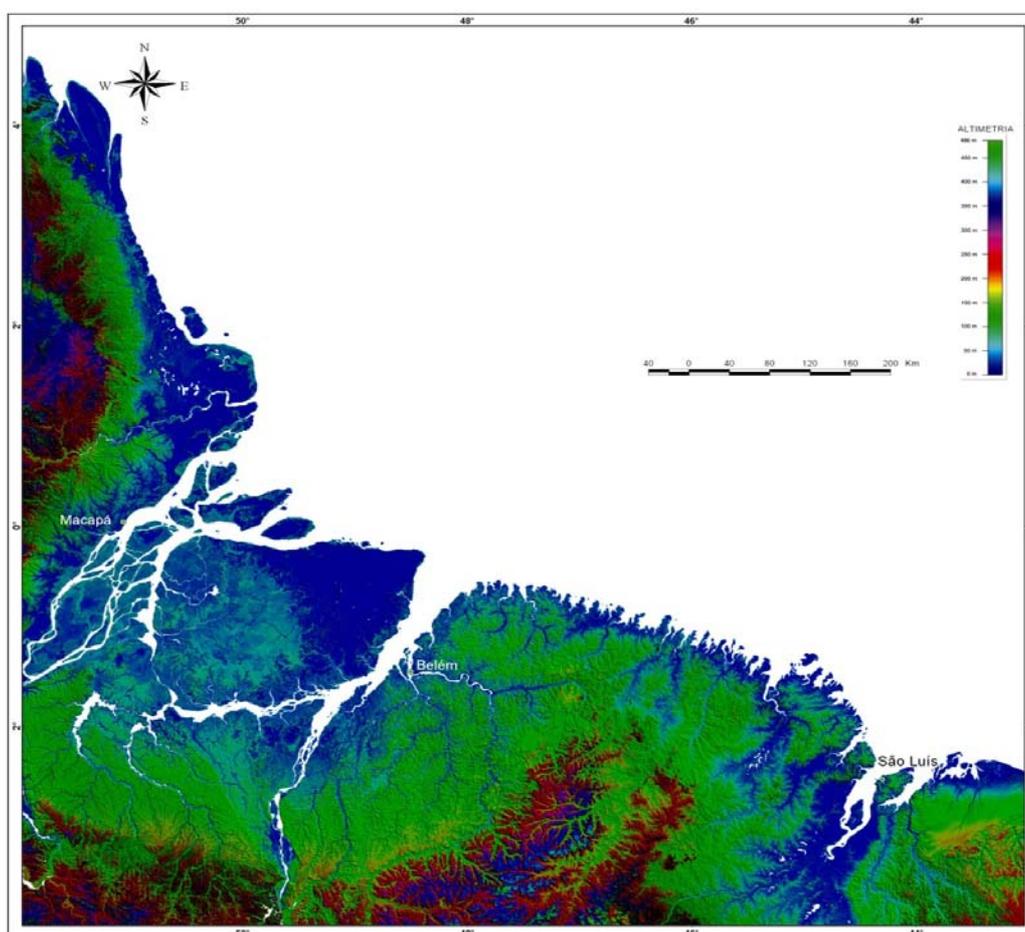


Figura 2.1: Área de abrangência do Projeto PIATAM Mar

Participam do projeto importantes instituições de pesquisa, tais como: universidades federais do Pará (UFPA), Rural da Amazônia (UFRA) e do Maranhão (UFMA), além do Museu Paraense Emílio Goeldi (MPEG), Universidade Estadual do Maranhão (UEMA), dos institutos Evandro Chagas (IEC), de Pesquisas Científicas e

Tecnológicas do Estado do Amapá (IEPA), Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia da Universidade Federal do Rio de Janeiro (COPPE/UFRJ).

Os grupos de pesquisa estão produzindo estudos sobre geologia costeira, oceanografia, hidrogeologia, meteorologia, geoquímica, modelagem, aves, bentos, peixes, plâncton, pólen, vegetação, socioeconomia, arqueologia, sensoriamento remoto, malária e outros temas.

2.2. Banco de Dados

Inicialmente, vamos fazer uma diferenciação básica entre dado e informação. Pois o dado em sua forma original, não nos traz conteúdo algum que não um valor abstrato. Este valor tratado ou agregado de outros atributos sim, nos trazem uma informação.

Assim, [Oliveira 1998], define dado como sendo qualquer elemento identificado em sua forma bruta que, por si só não conduz a uma compreensão de determinado fato ou situação e informação como sendo o dado trabalhado que permite ou colabora com a tomada decisões.

Já, [Falkenberg 1998], complementa ao afirmar que, informação é um incremento do conhecimento pessoal obtido através de uma ação de recebimento em uma transferência de mensagem, ou seja, é a diferença entre a concepção interpretada de uma mensagem recebida e o conhecimento pessoal antes da ação de recepção.

O banco de dados pode ser visto como o equivalente eletrônico de um armário de arquivamento. Em outras palavras, é um repositório de dados ou recipiente dada uma coleção de arquivos de dados computadorizados [Date 2000].

2.2.1. Contextualização da Base de Dados Relacional

O modelo relacional de banco de dados foi desenvolvido pelo cientista inglês Dr. Edgar Frnk “Ted” Codd, em junho de 1970 com a publicação do artigo “A Relational Model of Data for Large Shared Data Banks” na revista Communications of the ACM, v. 13, n. 6, pp. 377-387, publicada por Association fot Computing Machinery Inc [Manzano 2007].

A linguagem de consulta estruturada SQL foi desenvolvida primeiramente pela empresa IBM (Internacional Business Machine), e apresentada em sua primeira versão no ano de 1974, com o nome de Structured English QUery Language (SEQUEL). A linguagem de consulta SEQUEL foi desenvolvida pelo PH. D. Donald D. Chamberlin e outros profissionais da IBM.

Em 1979, um grupo de engenheiros que havia participado do projeto de desenvolvimento do SYSTEM/T (banco de dados relacional da IBM), fundou uma empresa denominada Relational Software, Inc. que disponibilizou o primeiro sistema de gerenciamento de banco de dados relacional comercialmente viável, baseado em linguagem SQL, denominado Oracle. Tal sistema será utilizado em todo o desenvolvimento desta dissertação [Manzano 2007].

As linguagens de Quarta Geração aparecem durante a metade da década de 1970 como linguagens com um elevado nível de abstração. Essa geração de linguagem é a mais próxima que se conseguiu chegar da linguagem humana. Segundo [Pressman 1992], “as linguagens de quarta geração combinam características procedimentais e não-procedimentais”, possibilitando ao usuário especificar “condições e as correspondentes ações (...), encorajando, ao mesmo tempo, o usuário a indicar o resultado desejado (...) e então aplicar seu conhecimento específico do domínio para preencher os detalhes procedimentais”. As linguagens de consultas utilizadas em bancos de dados são consideradas de quarta geração por acomodarem o conceito com bastante propriedade. Das linguagens existentes para consulta de dados, a que mais se destacou e tornou-se um padrão universal é a SQL [Martin 1986] [Manzano 2007].

O conjunto de dados aqui trabalhados traz informações de localização geográfica, peixes estuarinos, geologia, química da água, climatologia, bentos, fitoplâncton e zooplâncton. Abaixo, segue informações acerca das tabelas da base de dados e ao final, nos anexos, todo o dicionário de dados e a base relacional.

O principal critério na seleção das tabelas a serem estudadas foi o da alimentação da base de dados. Assim, mostra-se em Anexos, as tabelas alvo do estudo por terem um volume de dados considerável mínimo para demonstrar algum resultado prático do trabalho.

Por questões de direito de propriedade não mostraremos a modelagem completa do projeto, apenas a modelagem pertinente ao DW construído. Entretanto, mostraremos o fluxo de informações (figura 2.1).

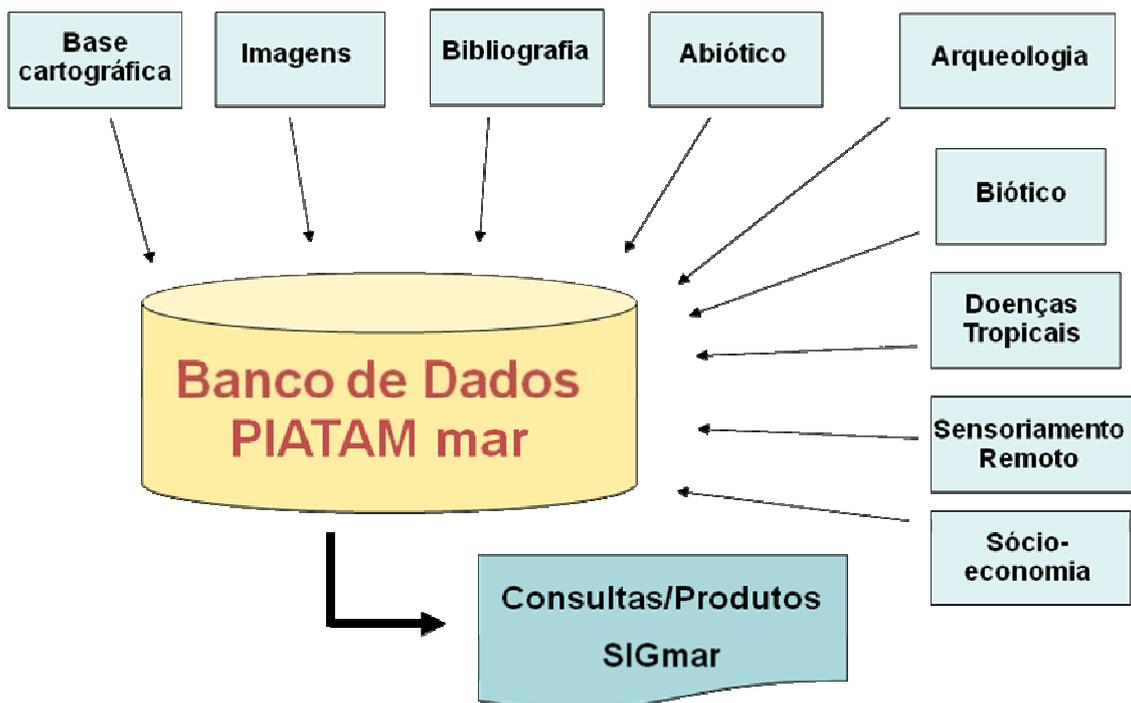


Figura 2.2: Fluxo de informações da base de dados do Projeto PIATAM Mar

2.3. Data Warehouse

A realização de Data Warehousing (DW) [Gardner 1998] é considerado como um dos primeiros passos para tornar factível a análise de grande quantidade de dados no apoio ao processo decisório. O objetivo básico é criar um repositório, conhecido por Data Warehouse, que contenha dados limpos, agregados e consolidados que possam ser analisados por ferramentas OLAP (*On-Line Analytical Processing*). Essas ferramentas apresentam facilidades para a realização de consultas complexas em Base de Dados multidimensionais. As ferramentas utilizadas para analisar um Data Warehouse, normalmente, são orientadas às consultas, ou seja, são dirigidas pelo usuário, os quais possuem hipóteses que gostariam de comprovar, ou simplesmente, executam consultas aleatórias. Essa abordagem dependente do usuário pode impedir que padrões escondidos nos dados sejam encontrados de forma “inteligente”, uma vez que o usuário não terá condições de imaginar todas as possíveis relações e associações existentes em um grande volume de dados. Por isso, faz-se necessária a utilização de técnicas de análise dirigidas por computador que possibilitem a extração automática (ou semi-automática) de novos conhecimentos, a partir de um grande repositório de dados [Bradley, Fayyad & Mangasarian 1998] e [Rezende 2005].

2.3.1. Arquitetura

2.3.1.1. Armazenamento

Um *Data Warehouse* pode armazenar grandes quantidades de informação, às vezes divididas em unidades lógicas menores que são chamadas de Data Marts. O esquema de dados mais utilizado em bancos de dados relacionais para a Modelagem Multidimensional é o *Star Schema* (esquema estrela). Apesar de bastante utilizado, não existe um padrão na indústria de software para o armazenamento de dados. Existem, na verdade, algumas controvérsias sobre qual a melhor maneira para estruturar os dados em um *Data Warehouse*.

Geralmente, o *Data Warehouse* não armazena informações sobre os processos correntes de uma única atividade de negócio, mas sim cruzamentos e consolidações de várias unidades de negócios de uma empresa [Kimball 2002].

2.3.1.2. Modelagem Multidimensional

A modelagem multidimensional é a técnica estruturada desenvolvida para obtenção de modelos de dados de simples entendimento e alta performance de acesso aos dados.

Como um *Data Warehouse* é um banco de dados orientado somente para a consulta de dados, a orientação da técnica criou os denominados modelos estrela [Machado 2006].

Existem outros modelos, tais como: flocos de neve e o de constelação, mas que não serão abordados neste trabalho.

No modelo estrela (figura 2.2), há uma grande tabela fato, contendo as medidas e é rodeada por tabelas dimensionais. Assim, as chaves das dimensões conectam-se com a tabela fato através de chaves estrangeiras [Fonseca 2007].

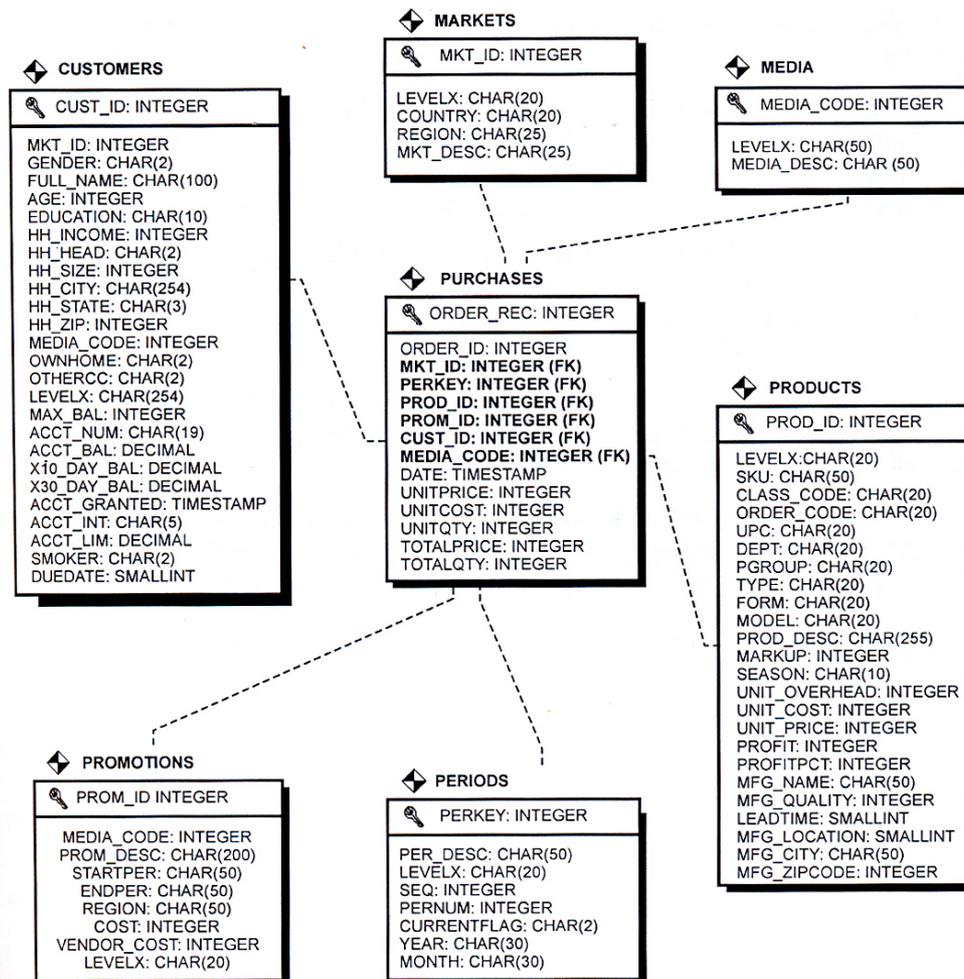


Figura 2.3: Modelo multidimensional [Machado 2006]

2.3.1.3. Metadado

O conceito Metadado é considerado como sendo os "dados sobre dados", isto é, os dados sobre os sistemas que operam com estes dados. Um repositório de metadados é uma ferramenta essencial para o gerenciamento de um *Data Warehouse* no momento de converter dados em informações para o negócio. Entre outras coisas, um repositório de metadados bem construído deve conter informações sobre a origem dos dados, regras de transformação, nomes e alias, formatos de dados, etc. Ou seja, esse "dicionário" deve conter muito mais do que as descrições de colunas e tabelas: deve conter informações que adicionem valor aos dados [Kimball 1998].

2.3.1.3.1. Tipo de Informação considerada Metadado

Os metadados são utilizados normalmente como um dicionário de informações e, sendo assim, devem incluir [Kimball 2002]:

- **origem dos dados:** todo elemento de dado precisa de identificação, sua origem ou o processo que o gera. Esta identificação é muito importante no caso da necessidade de saber informações sobre a fonte geradora do dado. Esta informação deve ser única, ou seja, cada dado deve ter uma e somente uma fonte de origem.
- **fluxo de dados:** todo elemento de dado precisa ter identificado os fluxos nos quais sofre transformações. É importante saber que dados servem de base para que processos.
- **formato dos dados:** todo elemento de dados deve ter identificado seu tamanho e tipo de dado.
- **nomes e alias:** todo elemento de dados deve ser identificado por um nome. Este nome pode ser da Área de Negócios ou um nome técnico. No caso de serem usados alias para os nomes, pode-se ter os dois. Devem existir padrões para criação de nomes e alias (ex.: convenções para abreviações), evitando assim ambigüidades.
- **definições de negócio:** estas definições são as informações mais importantes contidas nos metadados. Cada elemento de dado deve ser suportado por uma definição do mesmo no contexto da Área de Negócio. O método de manutenção destas informações também deve ser muito consistente, de forma que o usuário possa obter facilmente definições para as informações desejadas. Nestas definições devem ser evitadas referências a outros metadados que necessitem de uma segunda pesquisa para melhor entendimento.
- **regras de transformação:** são consideradas como sendo as Regras de Negócio codificadas. Estas regras são geradas no momento da extração, limpeza e agrupamento dos dados dos Sistemas Operacionais. Cada regra de transformação codificada deve estar associada a um elemento de Metadado. Se mais de uma aplicação contiver a mesma regra de transformação, deverá ser garantido que estas sejam idênticas.
- **atualização de dados:** o histórico das atualizações normalmente é mantido pelo próprio banco de dados, mas definir um elemento de metadado, indicando as datas de atualização dos dados, pode ajudar o usuário no momento de verificar a atualidade dos dados e a consistência da dimensão tempo do Data Warehouse.

- **requisitos de teste:** identifica os critérios de julgamento de cada elemento de dado. Valores possíveis e intervalos de atuação. Deve conter também padrões para procedimentos de teste destes dados.
- **responsabilidade sobre informações:** deve ser identificado o responsável por cada elemento de dados do Data Warehouse e também o responsável pela entrada de metadados.
- **acesso e segurança:** os metadados devem conter informação suficiente para que sejam determinados os perfis de acesso aos dados. Deve-se poder identificar que usuários podem ler, atualizar, excluir ou inserir dados na base. Deve haver, também, informações sobre quem gerencia estes perfis de acesso e como se fazer contato com o Administrador da Base de Dados.

2.4. OLAP

O termo processamento analítico on-line (ou OLAP, de *online analytical processing*) foi cunhado em um *white paper* escrito para a *Arbor Software Corp.* em 1993 [21.10], embora (como ocorre com o termo *data warehouse*), o conceito seja muito mais antigo. Ele pode ser definido como “o processo interativo de criar, administrar, analisar e gerar relatórios sobre dados” – e é habitual acrescentar que os dados em questão são percebidos e manipulados como se estivessem armazenados em um *array multidimensional* [Date 2000].

2.4.1. Funcionamento

Desde o início da tecnologia OLAP [Codd,1993], duas diferentes abordagens podem ser usadas com a tecnologia OLAP. A primeira estende o modelo relacional para garantir os requisitos OLAP, sendo que surgiram outras propostas [Kimball, 1998] e todas incluem extensões para operar o cubo de dados. A segunda abordagem trata sobre a implantação de uma estrutura multidimensional dos dados, e alguns autores propuseram modelos sobre a mesma [Molina et al, 2006]. Alguns trabalhos relativos a implementação OLAP foram realizados como em [Burdick et al, 2007], onde o modelo OLAP é estendido para representar a imprecisão e incerteza dos dados [Fonseca, 2007].

2.4.2. Tipos de OLAP

2.4.2.1. Multidimensional (MOLAP)

Usam arrays multidimensionais como estrutura de dados e as operações OLAP são realizadas sobre essas estruturas. O modelo multidimensional é implantado em matrizes tendo as dimensões como os índices, e a medida como conteúdo da célula, nesta implementação apenas a medida armazenada [Fonseca, 2007].

2.4.2.2. Relacional (ROLAP)

Usam tabelas relacionais como estrutura básica de dados e uma célula do espaço multidimensional é uma tupla. Esta tupla contém atributos que indentificam a localização da célula no espaço multidimensional e outros atributos que contêm as medidas. As operações OLAP são realizadas através de SQL [Fonseca, 2007].

2.4.2.3. Híbrido

Recentemente, muitos produtos MOLAP ou ROLAP têm tentado se firmar no mercado como HOLAP, ou OLAP híbrido, alegando algum fator de integração entre os dois tipos de arquitetura. Os três argumentos mais usados pelos fabricantes dessas ferramentas são [Leitão 2000]:

- a) Oferecer escolha entre um banco de dados relacional e um banco de dados multidimensional;
- b) Carregar resultados de consultas relacionais em um banco de dados multidimensional ou estrutura proprietária que simule a multidimensionalidade dos dados, em tempo de execução.
- c) Usar um banco de dados multidimensional para fazer um cache dos dados com um nível maior de agregação e usar um banco relacional para fazer um acesso dinâmico aos dados detalhados.

2.4.2.4. Comparação

O local de armazenamento pode ser um banco de dados multidimensional, como um MOLAP; um banco de dados relacional, como em HOLAP, ou ainda arquivos em base de dados locais ou mesmo em memória de estações clientes, como em WOLAP [Oliveira, 2002].

Na arquitetura MOLAP os dados ficam armazenados em um banco de dados multidimensional, onde o servidor MOLAP atua e o usuário trabalha, monta e manipula os dados diferentes no servidor [Carvalho, 2004]. Os dados de um banco multidimensional são armazenados em um espaço menor que o utilizado para

armazenar os mesmos dados em um banco de dados relacional. No banco multidimensional, os dados são mantidos em estruturas de dados do tipo array de maneira a prover um melhor desempenho ao acessá-los. Além de ser uma arquitetura rápida uma outra vantagem é o rico e complexo conjunto de funções de análises presentes nos bancos multidimensionais [Carvalho, 2004].

Uma de suas limitações é a possibilidade dos dados serem esparsos (nem todo cruzamento das dimensões contém dados), ocorrendo a chamada explosão de armazenamento de dados, ou seja, um imenso banco de dados multidimensional contendo poucos dados armazenados. Outras limitações dessa ferramenta estão relacionadas ao fato dos bancos multidimensionais serem sistemas proprietários que não seguem padrões, ou seja, cada desenvolvedor cria a sua própria estrutura para o banco e as próprias ferramentas de suporte [Carvalho, 2004].

Já a arquitetura ROLAP é uma simulação da tecnologia OLAP feita em banco de dados relacionais que, por utilizar a estrutura relacional, possui a vantagem de não restringir o volume de armazenamento de dados [Carvalho, 2004]. Essa ferramenta não utiliza cubos pré-calculados como a MOLAP. À medida que o usuário monta sua consulta em uma interface gráfica, a ferramenta acessa os metadados ou quaisquer outros recursos que possua, para gerar uma consulta SQL (Structured Query Language) [Thomsem, 2002].

A sua principal característica é a possibilidade de fazer qualquer consulta, atendendo melhor os usuários que não têm um escopo de análise bem definido. Essa ferramenta tem a vantagem de utilizar tecnologia estabelecida, de arquitetura aberta e padronizada, beneficiando-se da diversidade de plataformas, escalabilidade e paralelismo de hardware. Sua desvantagem é o conjunto pobre de funções para análises dimensionais e o baixo desempenho da linguagem SQL na execução de consultas pesadas [Carvalho, 2004].

A arquitetura HOLAP, ou processamento híbrido, está se tornando a mais popular para os produtos atuais, porque consegue combinar a capacidade e a escalabilidade das ferramentas ROLAP com o desempenho superior dos bancos de dados multidimensionais [Thomsem, 2002]. Por exemplo, suponha uma base de 50.000 clientes distribuídos em 500 cidades, 23 estados, 5 regiões e um total geral. Até o nível de cidades o armazenamento multidimensional resolveria as consultas para levantar totais de vendas. Porém, se fosse necessário consultar o total de vendas de um determinado cliente, o banco de dados relacional responderia com muito mais rapidez à solicitação. Essa situação é típica para a indicação da arquitetura HOLAP [Oliveira, 2002].

Uma outra arquitetura é a DOLAP que é uma arquitetura desktop do OLAP, ou seja, é uma ferramenta para usuários que possuam uma cópia da base multidimensional ou de um subconjunto dela ou ainda, que queiram acessar um repositório de dados central localmente. O usuário ao acessar este repositório, dispara uma instrução SQL e acessa os cubos já existentes no banco de dados multidimensional residente no servidor OLAP e obtém de volta um para ser analisado em sua estação de trabalho [Cavalcanti, Oliveira e Monteiro, 1998].

A vantagem dessa arquitetura é a redução da sobrecarga no servidor de banco de dados uma vez que todo o processamento OLAP acontece na máquina cliente e a desvantagem é o tamanho do microcubo que não pode ser muito grande, caso contrário, a análise pode ser demorada e a máquina do cliente não suportar [Cavalcanti, Oliveira e Monteiro, 1998].

Finalmente, a arquitetura WOLAP é a utilização de uma ferramenta OLAP a partir de um browser¹. Essa arquitetura tem duas tecnologias que estão em constante evolução, a primeira é a Web e a segunda são as ferramentas OLAP. A diferença desta ferramenta para as outras é que ela utiliza a Web, facilitando assim, a distribuição da ferramenta, o acesso remoto dos dados a serem analisados e a utilização da aplicação independente de plataforma [Andreato, 1999].

Essa ferramenta representa a migração da tecnologia OLAP para o ambiente da Internet. Atualmente o uso de Web browsers para acesso OLAP está sendo muito divulgado, entretanto, poucos são os sites que utilizam esta ferramenta. O WOLAP, segundo pesquisas realizadas, será a chave para aplicações na Internet e deverá oferecer um caminho simples e barato no acesso a dados do DW [Andreato, 1999].

A partir do modelo estrela pode-se, através de ferramentas OLAP, construir cubos de dados OLAP (figura 2.3), e realizar buscas nesse espaço multidimensional (figura 2.2), mostrada anteriormente. Os cubos de dados são visões lógicas multidimensionais dos dados com referência hierárquica. As ferramentas OLAP fornecem funcionalidades para análise interativa de dados em diferentes visões e granularidades, permitindo visualizar as hierarquias e navegar pelas dimensões [THONSEN, 2002].

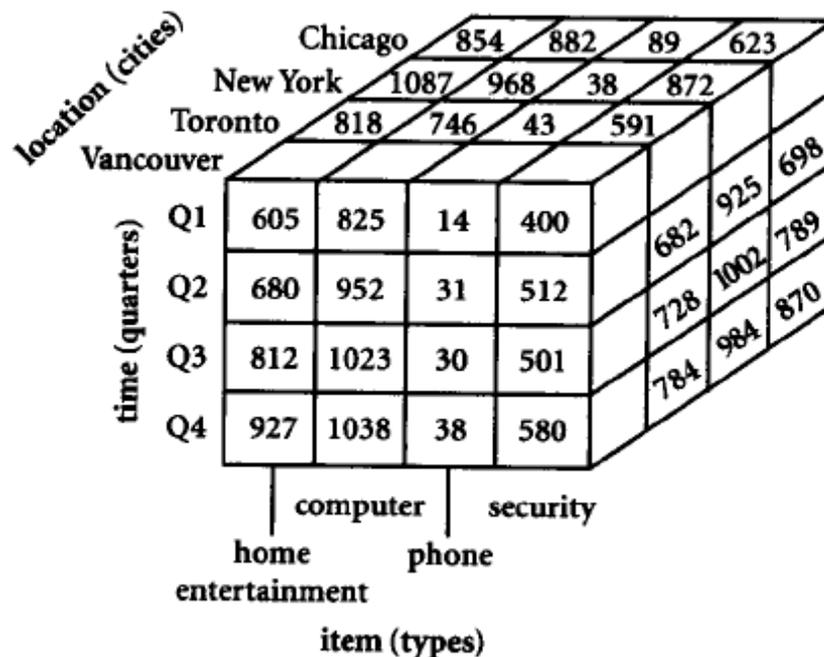


Figura 2.4: Cubo lógico de dados onde as arestas representam as dimensões [HAN, 2001].

A essência do OLAP se dá pela eficiente computação das agregações dos diversos conjuntos de dimensões do cubo de dados. Estas agregações são referenciadas em SQL como “group by” [HAN, 2001].

2.5. Mineração de Dados (*Data Mining*)

A definição aceita por diversos pesquisadores de Mineração de Dados foi elaborada por Fayyad, Piatetsky, & Smyth (1996a) como sendo: “Extração de Conhecimento de Base de Dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados”. [Rezende 2005].

A Mineração de Dados é a principal etapa do processo de KDD. Nesta etapa ocorre a busca efetiva por conhecimentos novos e úteis a partir dos dados. Por este motivo, diversos autores referem-se à Mineração de Dados e ao Processo de KDD de forma indistinta, como se fossem sinônimos [Passos & Goldschmidt, 2005].

A capacidade de se produzir e armazenar dados cresce de maneira ainda mais acelerada nos últimos anos, e as perspectivas demonstram que a tônica deste crescimento deve permanecer ou mesmo perenecer com o passar do tempo. A expansão da utilização do computador e a internet tem sido fatores preponderante neste sentido.

Entretanto, o principal objetivo da utilização do computador ainda tem sido o de resolver problemas operacionais das organizações. Estas, em sua grande maioria,

ainda não utilizam recursos computacionais na tomada de decisões, apesar de possuírem vastas quantidades de dados que ficam, muitas vezes, retidas em “cofres” de dados. Estas organizações têm dificuldades na identificação de formas de exploração dos seus dados, e mais ainda na transformação desses vastos repositórios de dados em conhecimento [FERRARI, 1997].

Data Mining tem sido reconhecida como uma nova área de pesquisa interdisciplinar, envolvendo aprendizado de máquina, estatística, banco de dados e outras áreas [Berson & Smith 97]. Tal área pode ser definida como "descoberta eficiente de regras interessantes em grandes volumes de dados" [Srikant & Agrawal 95].

2.5.1. Classificação

Uma das tarefas de KDD (Descoberta de Conhecimento em Base de Dados) mais importantes e mais populares é a de classificação. Informalmente, essa tarefa pode ser compreendida como a busca por uma função que permita associar corretamente cada registro de um banco de dados a um único rótulo categórico denominado classe. Uma vez identificada, essa função pode ser aplicada a novos registros de forma a prever a classe em que tais registros se enquadram [Passos & Goldschmidt 2005].

A classificação de dados é um processo composto de duas etapas. Na primeira um modelo é construído para descrever um conjunto pré-determinado de dados classificados. O modelo é construído pela análise das entradas de cada atributo considerado. Cada entrada é tratada como se pertencesse a apenas uma classe, determinada por um dos atributos denominado de classe do atributo. As amostras utilizadas para treinar o modelo são selecionadas aleatoriamente do conjunto de dados e são denominadas de conjunto de treinamento. Como a classe de cada instância de treinamento também é fornecida, esta etapa também é conhecida como aprendizagem supervisionada. A classificação supervisionada contrasta com a aprendizagem não supervisionada (cluster), na qual a classe de cada instância de treinamento não é conhecida, ou muitas vezes, nem mesmo o número de classes é conhecido a priori [Han & Kamber, 2001].

2.5.2. Agrupamento

A tarefa de “clusterização”, também chamada de Agrupamento, é usada para particionar os registros de uma base de dados em subconjuntos ou *clusters*, de tal

forma que elementos em um *cluster* compartilhem um conjunto de propriedades comuns que os distingam dos elementos de outros *clusters*. O objetivo dessa tarefa é maximizar similaridade *intercluster*. Diferente da classificação que tem rótulos predefinidos, a “clusterização” precisa automaticamente identificar os rótulos. Por esta razão, a “clusterização” é também denominada indução não supervisionada [Goldschmidt & Passos, 2005].

2.5.2.1. Métodos de Agrupamento

Os métodos de agrupamento mais conhecidos e utilizados são os métodos por particionamento e os métodos hierárquicos.

2.5.2.1.1. K-means

O algoritmo *k-means* toma um parâmetro de entrada, *k*, e divide um conjunto de *n* objetos em *k* clusters tal que a similaridade *intracluster* resultante seja alta, mas a similaridade *intercluster* seja baixa. A similaridade em um cluster é medida em respeito ao valor médio dos objetos neste cluster (centro de gravidade do cluster).

A execução do algoritmo *k-means* consiste em, primeiro, selecionar aleatoriamente *k* objetos, que inicialmente representam cada um a média de cluster. Para cada um dos objetos remanescentes, é feita a atribuição ao *cluster* ao qual o objeto é mais similar, baseado na distância entre o objeto e a média do *cluster*. A partir de então, o algoritmo computa as novas médias para cada *cluster*. Este processo se repete até que uma condição de parada seja atingida [Goldschmidt & Passos, 2005].

2.5.2.1.2. Método Hierárquico

Os algoritmos de clusterização hierárquicos criam uma decomposição hierárquica da base de dados. A decomposição hierárquica é representada por um dendrograma, uma árvore que iterativamente divide a base de dados em subconjuntos menores até que cada subconjunto consista de somente um objeto [Goldschmidt & Passos, 2005].

3. Metodologia Proposta

A metodologia proposta neste trabalho define as etapas necessárias para o desenvolvimento das tarefas de classificação e/ou agrupamento em uma base de dados ambiental na zona costeira amazônica.

Tendo como objetivo prático o enriquecimento da metodologia de mineração de dados, foi desenvolvido um ambiente com ferramentas OLAP e DW a partir de aplicativos da Oracle, visando a busca de conhecimento na base de dados.

O trabalho evoluiu a partir da seleção de parte da base de dados do Projeto PIATAM Mar e posteriormente, da exportação da base de dados originalmente em MySQL para uma base Oracle.

Assim, a primeira etapa deste trabalho foi a análise da base de dados que possui vertentes de suma importância, destacando-se a localização, o tempo, a coleta de dados e os temas pesquisados.

A partir de então, podemos dividir o trabalho em cinco etapas (figura 3.1):

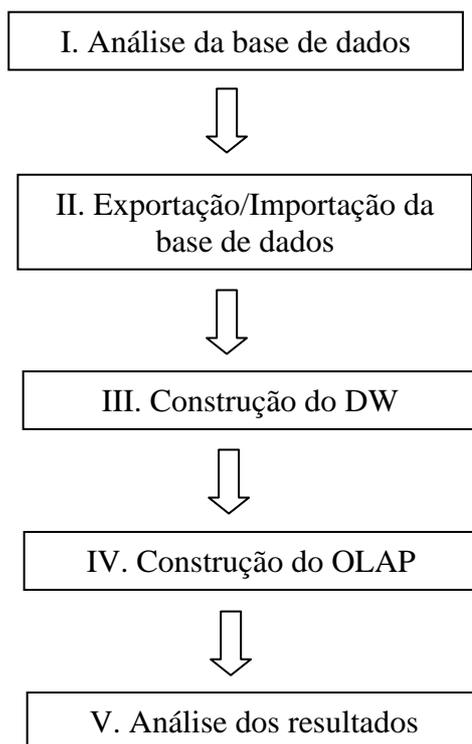


Figura 3.1: Etapas da metodologia proposta

As etapas são desenvolvidas seqüencialmente e ao final de cada uma é realizada a análise para a verificação de desempenho e resultado.

3.1. Análise da base de dados

Nesta etapa, foi realizada verificação na base quanto a consistência dos dados e principalmente pelo povoamento da base de dados.

A partir de então, foi construída uma aplicação simples apenas com a finalidade de ler a base de dados em MySQL e exportar para um arquivo texto ou para uma planilha Excel. Tal procedimento foi facilitado pelo baixo povoamento do banco de dados.

3.2. Exportação/Importação da Base de Dados

Nesta fase, os dados são migrados para uma base Oracle 10g, onde os dados são novamente analisados e verificados quanto a sua consistência de relacionamentos e integridade.

É de suma importância que os dados mantenham a integridade, pois todo o trabalho será comprometido caso a base não esteja tal e qual antes da importação.

3.3. Construção do DW

Nesta etapa, é realizada a construção do DW que permite análise multidimensional dos dados em diferentes granularidades, e pode auxiliar o processo de mineração de dados na busca de conhecimento. Os Datawarehouses são usados em diferentes áreas de conhecimento e sendo integrado a outras tecnologias como realizado em SAMPAIO et al. (2006), onde uma nova área de pesquisa é abordada, sendo chamada de Spatial Data Warehousing (SDW), que integra as tecnologias de Datawarehousing e Sistemas de Informações Geográficas (GIS) [Fonseca 2007].

3.3.1. Operacionalização da ferramenta ETL (OWB - Oracle Warehouse Builder)

No OWB, o desenvolvimento de data warehouses é orientado a projeto, ou seja, para iniciar-se a modelagem e construção de um novo warehouse é necessário criar-se um novo projeto. As informações trabalhadas em cada projeto são armazenadas no próprio repositório do OWB (repositório que fica armazenado em um banco de dados Oracle). Em cada projeto criado no OWB existe a possibilidade de se definir dois tipos de módulos:

- Módulos Fontes de Dados: basicamente permitem mapear as bases de dados de onde são extraídos os dados operacionais;

- Módulos Warehouse: permitem modelar os dados que ficam disponíveis para análise no data warehouse.

A ferramenta do OWB utilizada para modelagem, mapeamento, transformação e disponibilização dos dados é denominada Design Center (figura 3.2).

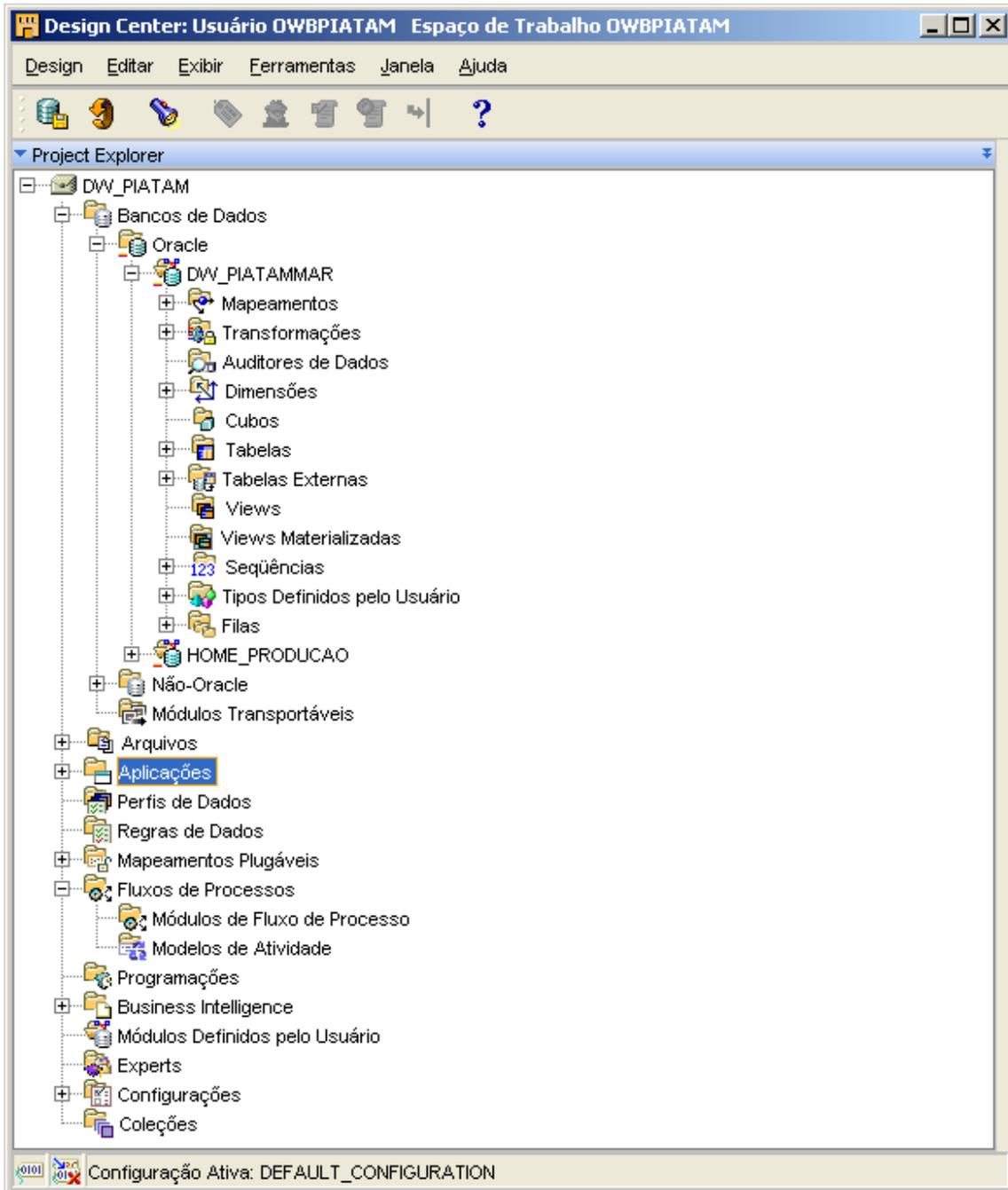


Figura 3.2: Espaço de trabalho do Design Center

3.4. Construção do OLAP

Atualmente, a maioria dos Datawarehouses são usados para OLAP, entretanto, com o crescimento do uso de Data Mining pode-se esperar que os dados dos Datawarehouses sirvam como importante fonte para o processo de Data Mining (HAN, 2001).

Além disso, OLAP é parte da tecnologia de Datawarehouse que habilita usuários a explorar dados do Datawarehouse interativamente. As ferramentas OLAP fornecem funcionalidades para análise interativa de dados em diferentes visões e granularidades (HAN, 2001).

Então, nesta etapa, buscamos facilitar a maior interatividade dos pesquisadores em gerar suas próprias visões e granularidades, permitindo assim uma maior transparência (aqui com o conceito de abstração do que está sendo realizado, mas mantendo a consistência e integridade dos resultados) do que realmente está sendo realizado.

3.5. Análise dos Resultados

Todas as etapas tem grande importância no desenvolvimento da metodologia e são avaliadas a cada término de processo. Ainda assim, faz-se necessária uma análise conjunta e interdisciplinar com os pesquisadores especialistas em cada área de pesquisa, com a finalidade de ratificar ou retificar o conhecimento adquirido.

Nesta etapa, faz-se a análise do que foi obtido através do DW, onde, caso haja informação suficiente, isto quantitativamente, serão aplicadas ferramentas OLAP sobre o DW e feita nova análise.

4. Ambiente Computacional

Para o desenvolvimento deste trabalho foi montado o ambiente computacional utilizando as ferramentas *CASE* de larga utilização no mercado, com o objetivo de facilitar futuras implementações.

4.1. Arquitetura do Projeto

O projeto foi realizado utilizando a ferramenta ERwin Data Modeler 4.1 para a modelagem e documentação da arquitetura do DW, SQL Navigator 5.0 em diversas etapas do projeto, auxiliando desde a modelagem até a visualização dos fatos de negócio, Oracle Data Base 10g utilizado para criação da base de dados operacional, Oracle Warehouse Builder 11g para os processos OLAP (Mapeamento e Transformação).

4.1.1. Banco de Dados Oracle 10g

A base de dados relacional foi criada utilizando os recursos do SGBD Oracle 10g. Sistema de alta robustez e performance, expressivamente utilizado no mercado em aplicações que exigem tais requisitos.

4.1.2. Ferramenta ERwin Data Modeler 4.1

A modelagem seguiu os mesmos princípios de modelagem de dados Entidade-Relacionamento. O ERwin, neste caso, oferece uma gama vasta de recursos que suportam esse tipo de trabalho (figura 4.1).

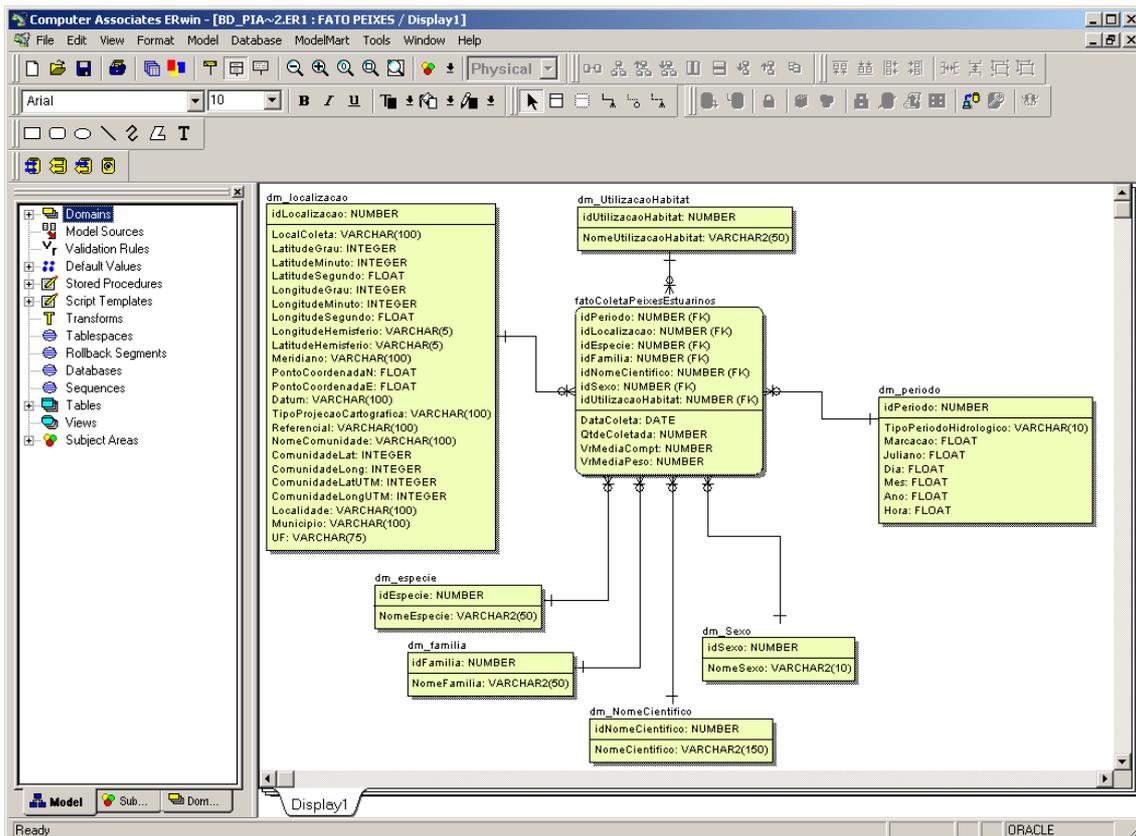


Figura 4.1: página do modelo FatoColetaPeixesEstuarinos

Para início da modelagem foi utilizado o recurso de Engenharia Reversa da ferramenta para importar metadados de estruturas do banco de dados operacional (figura 4.2).

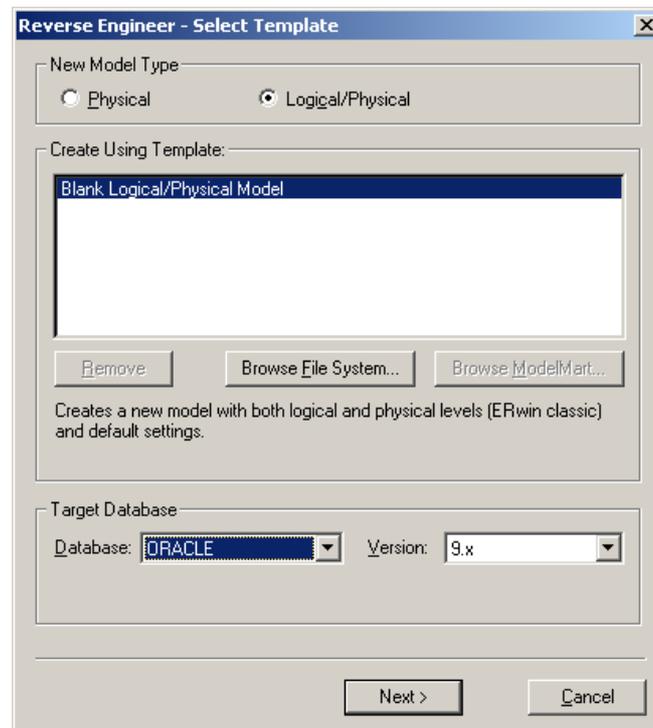


Figura 4.2: recurso de Engenharia Reversa do ERWin

Outro recurso muito útil do ERWin é a Engenharia Remissiva, que consiste na capacidade de geração ou atualização de um esquema a partir do modelo gráfico da ferramenta (figura 4.3).

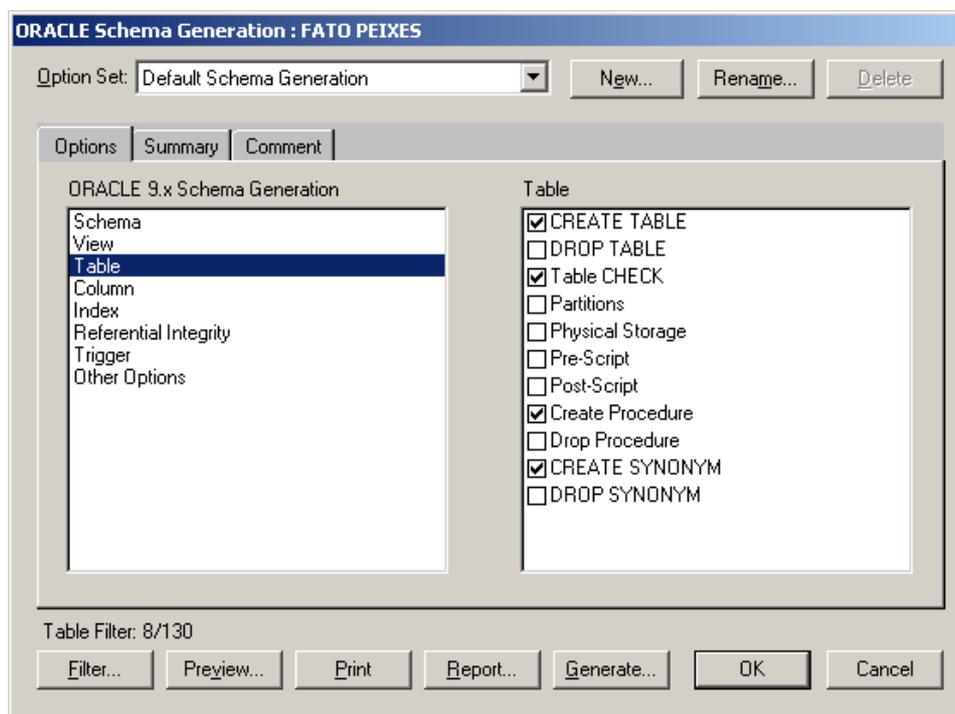


Figura 4.3: recursos de Engenharia Remissiva do ERWin

4.1.3. SQL Navigator 5.0

O Navigator é uma ferramenta direcionada a DBAs e programadores em bancos de dados. Totalmente direcionada a bancos de dados Oracle, a versão 5 da ferramenta, conecta em bancos de dados das versões 8i ou superiores, possibilitando ao usuário um arsenal de recursos para manipulação de esquemas de bancos de dados operacionais.

4.1.4. Oracle Warehouse Builder 11g

A principal ferramenta utilizada neste ambiente foi o Oracle Warehouse Builder 11g. Responsável pelos processos de modelagem e construção de um novo warehouse da base operacional, para tal, utilizou-se o aplicativo Design Center que consiste em uma interface gráfica para importação, mapeamento, modelagem dimensional, cubos de dimensões, jobs (schedules de objetos) e quase todos os tipos de objetos presentes em um esquema de banco de dados baseado no Oracle (figura 4.4).

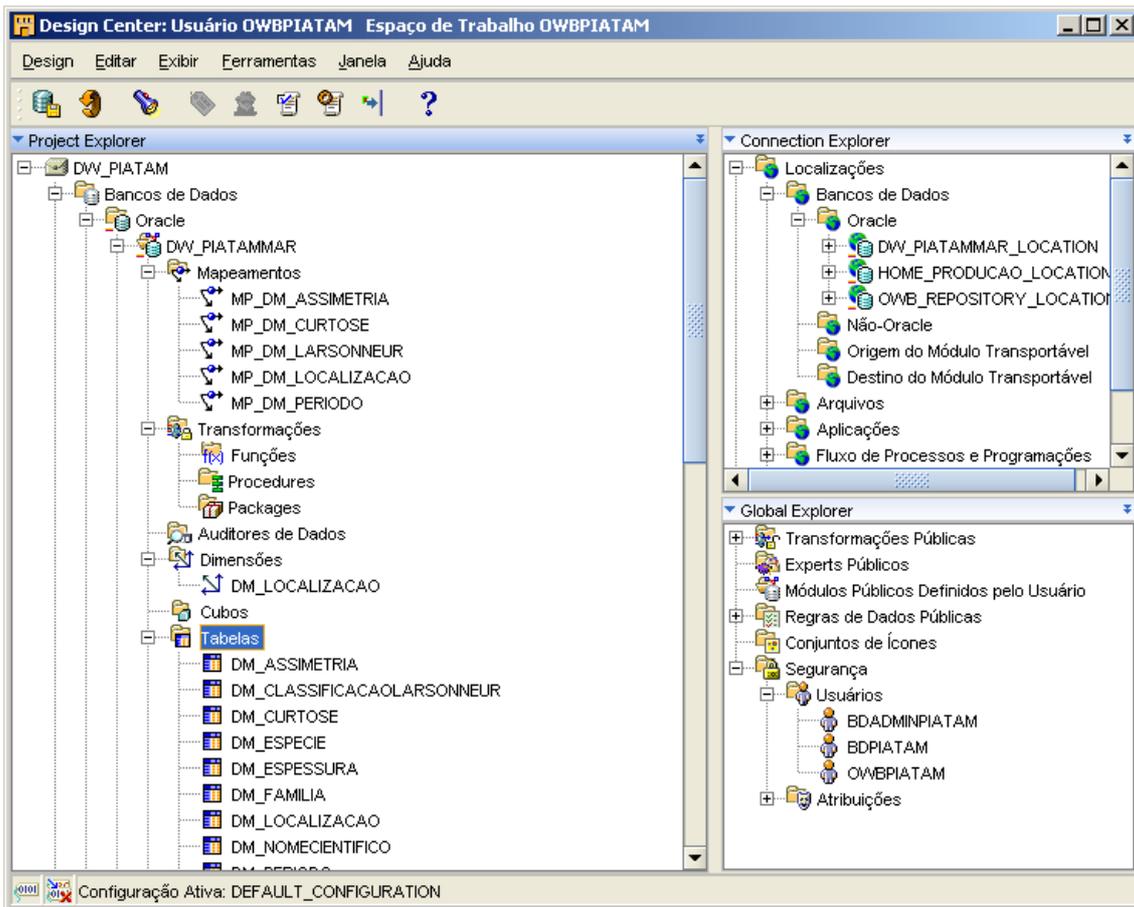


Figura 4.4: Espaço de Trabalho do Design Center

4.2. Construção do Modelo do Projeto DW Piatam Mar

4.2.1. Modelagem do DW

O modelo multidimensional do projeto foi descrito em detalhes anteriormente. Sua construção considerou os fatos de negócio mais relevantes da base operacional Piatam mar, contidas no esquema BDPIATAM (figura 4.5).

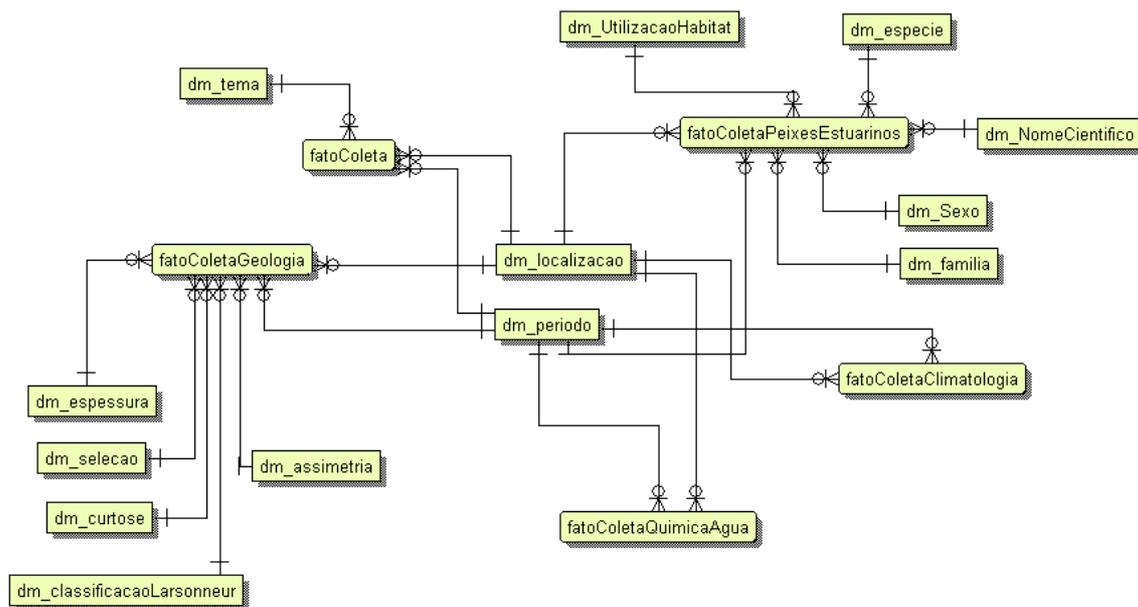


Figura 4.5: fatos principais do Projeto DW Piatam Mar

O ERWin, conforme já mencionado, foi a ferramenta utilizada nesta primeira etapa de modelagem. O passo seguinte foi a geração dos objetos para o DW em um esquema diferente denominado DWPIATAM, utilizando, para tal, o recurso Engenharia Remissiva do ERWin.

4.2.2. Importação do Metadado

O próximo passo foi a criação do DW propriamente dito, utilizando a ferramenta OWB 11g. O Oracle Warehouse Builder necessita de um repositório de trabalho, descrito na seção 3.3.1, que serve como base estrutural na construção de qualquer DW. Para o projeto Data Warehouse Piatam Mar foi criado o repositório OWBPIATAM e sobre este, foi criado o projeto DW_PIATAM.

Como os objetos já estavam modelados e criados no esquema DWPIATAM, utilizou-se o recurso de importação do Design Center. A ferramenta utiliza o conceito de módulo para organizar os objetos importados. Dessa forma, deve-se criar ou indicar qual módulo receberá os objetos importados (figura 4.6).

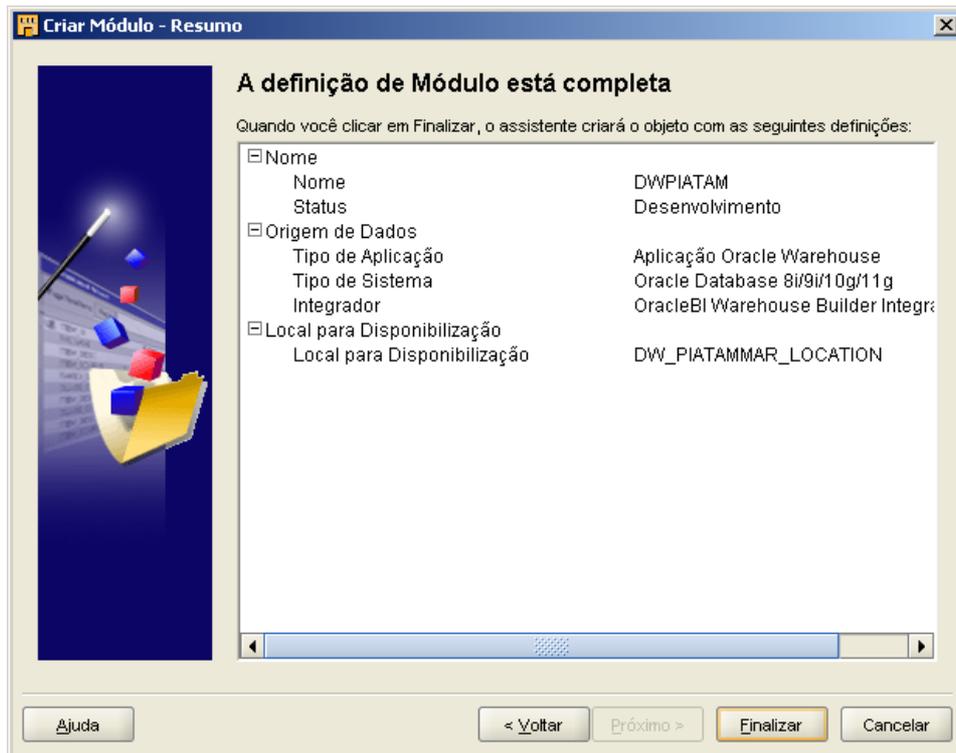


Figura 4.6: criação do módulo que receberá os objetos importados

Após organizados os módulos, iniciou-se o processo de importação dos objetos da base operacional (figura 4.7).

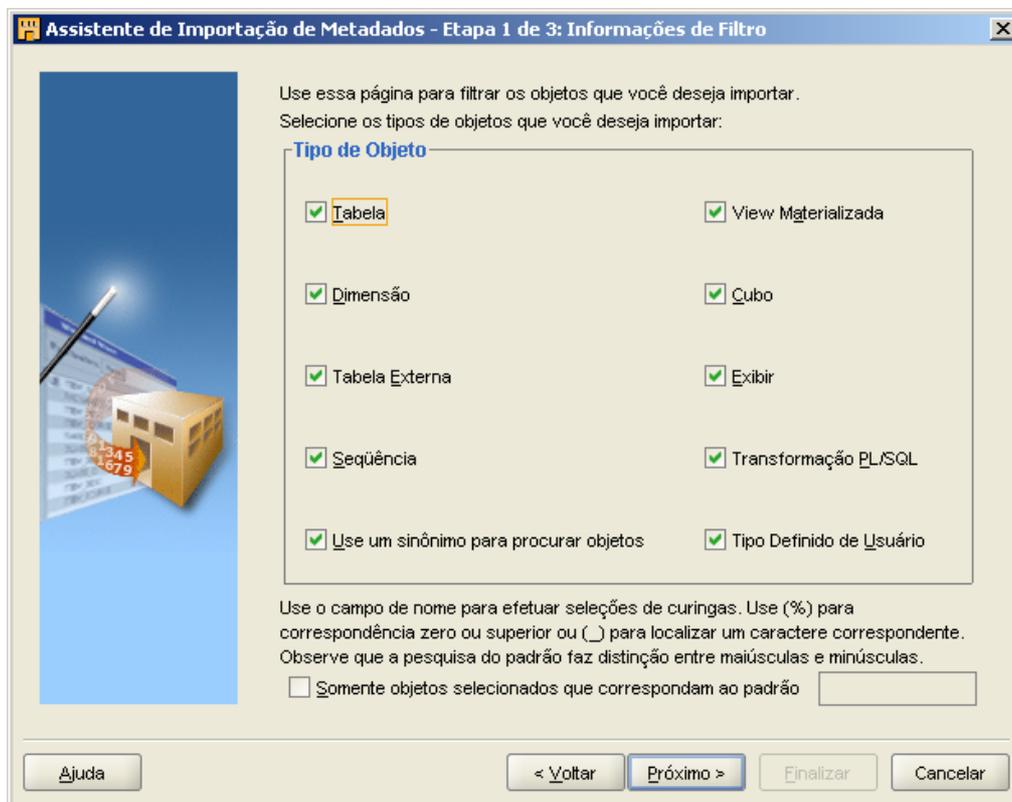


Figura 4.7: processo de importação dos objetos

Os objetos foram importados para um módulo chamado DW_PIATAMMAR, onde foram criados e processados os mapeamentos e transformações do Warehouse.

4.2.3. Mapeamentos e Transformações do DW

Como as dimensões foram modeladas, criadas e importadas para o repositório do projeto, o passo seguinte foi mapear os dados importantes para a análise dentro do DW. Assim, foram criados os objetos de apoio para suporte aos processos de Mapeamento e Transformação do Design Center.

Um dos objetos que mostrou-se necessário para suporte aos mapeamentos, foi a criação de sequences (figura 4.8). As sequences, no OWB, são praticamente iguais aos mesmos objetos de uma base relacional, com o diferencial que após criados, para serem utilizados, faz-se necessário a “Disponibilização” do objeto (figura 4.9).

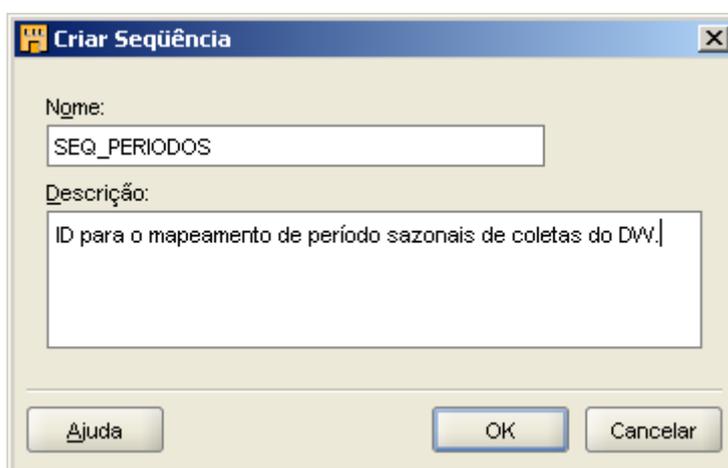


Figura 4.8: criação de uma *sequence*.

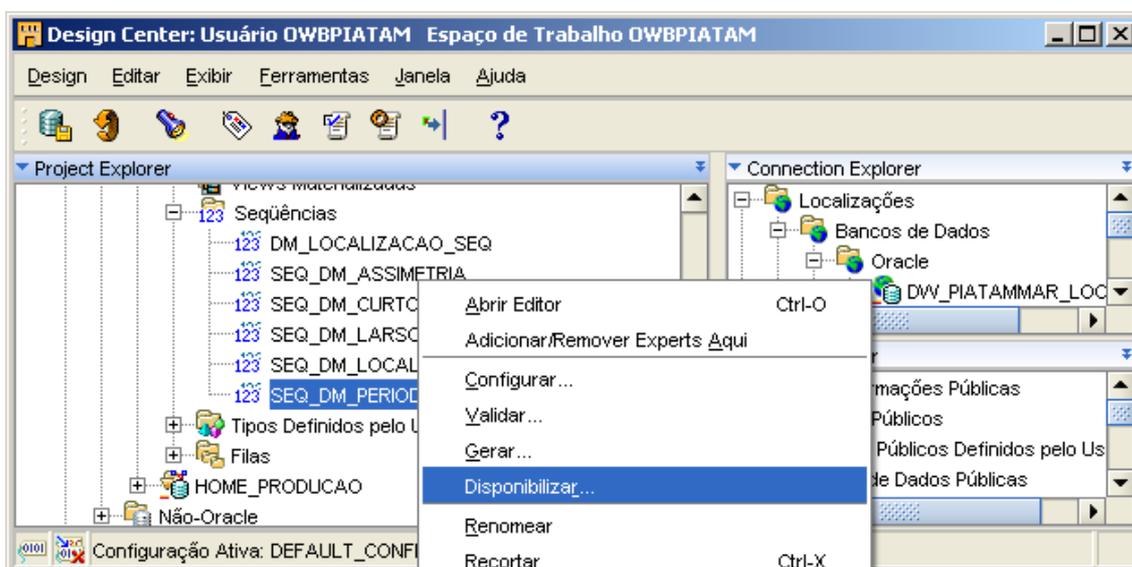


Figura 4.9: disponibilização de uma *sequence*.

Outro objeto muito importante para o suporte aos mapeamentos do Warehouse foi a Tabela Externa. Consiste em um objeto, que aponta para um arquivo de origem dos dados externo ao repositório. Faz-se necessário previamente, mapear o arquivo externo, antes de criar a tabela externa.

O mapeamento consiste em criar uma conexão do tipo “Arquivos” que será utilizada para localizar a pasta de origem do arquivo físico, e a configuração de diversos parâmetros de propriedades do arquivo externo (figura 4.10).

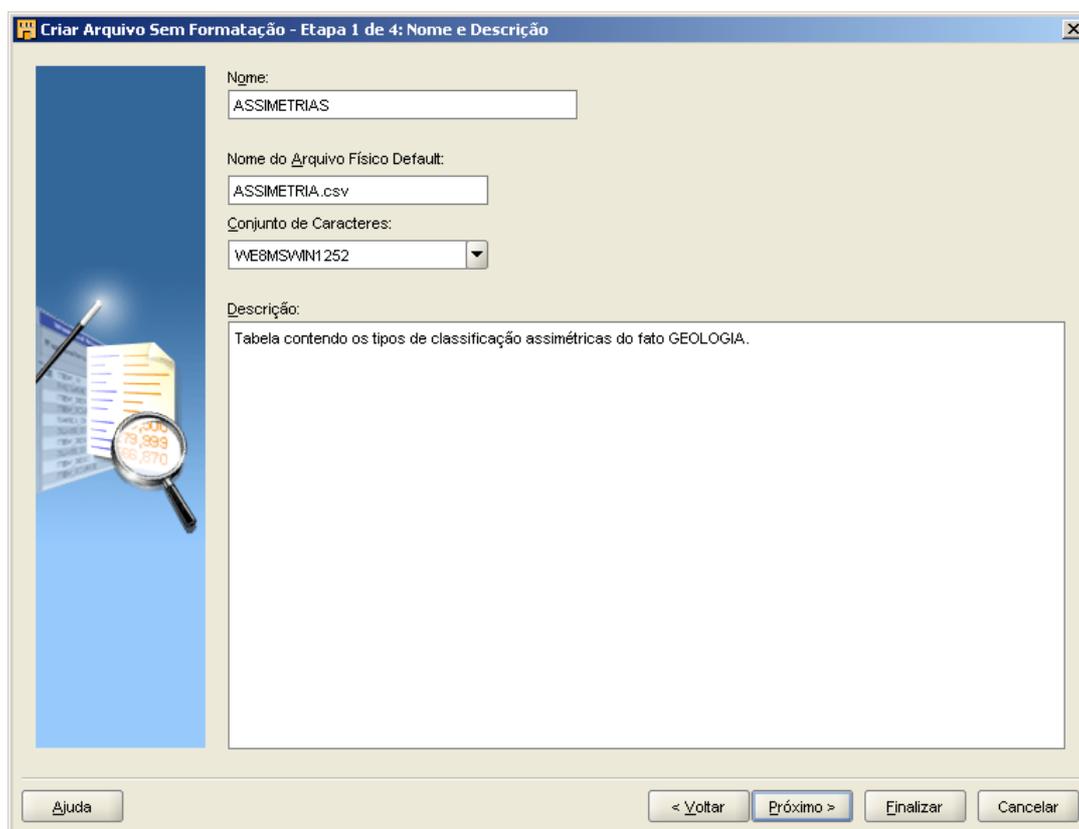


Figura 4.10: mapeamento do arquivo externo.

Após mapeados, os arquivos externos podem ser utilizados para criação das tabelas externas (figura 4.11).

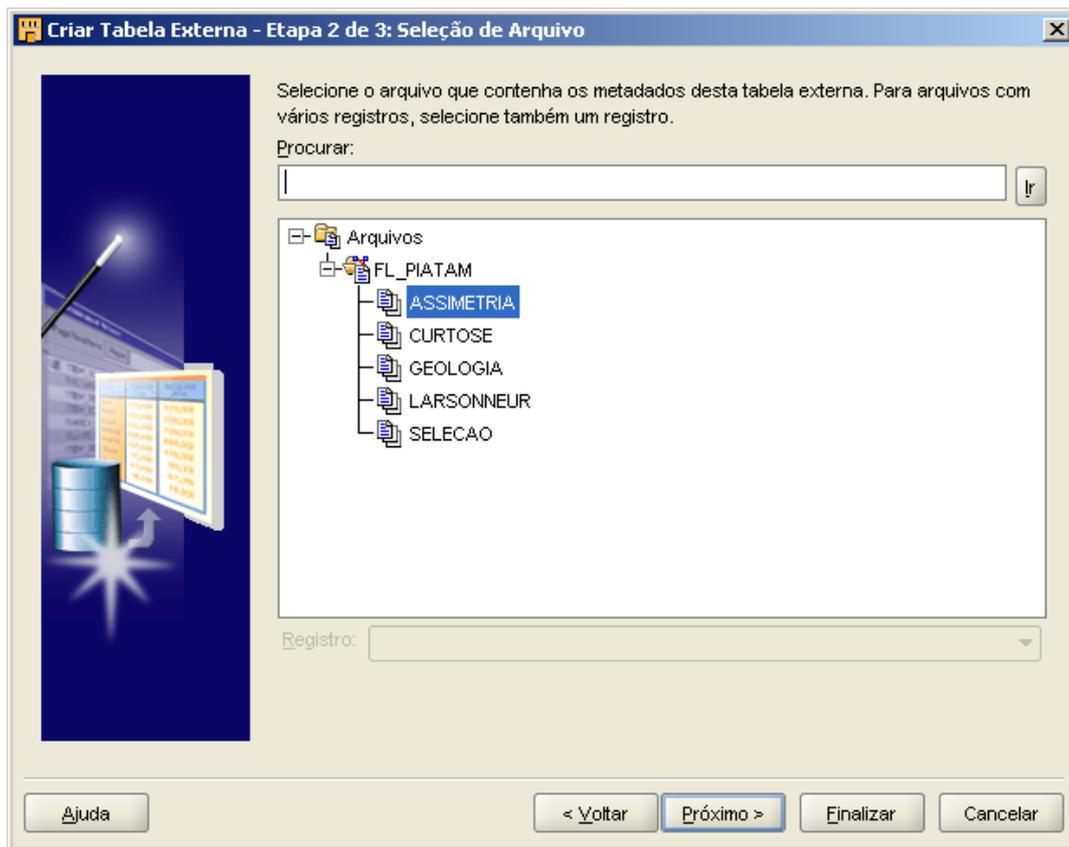


Figura 4.11: criação da tabela externa TB_ASSIMETRIA.

Criados todos os objetos que darão suporte aos mapeamentos e transformações, iniciou-se a construção dos mesmos. O recurso que garante essa tarefa é o Mapeamento Editor, ferramenta gráfica que oferece uma sucessão de recursos para se construir mapeamentos consistentes e robustos (figura 4.12).

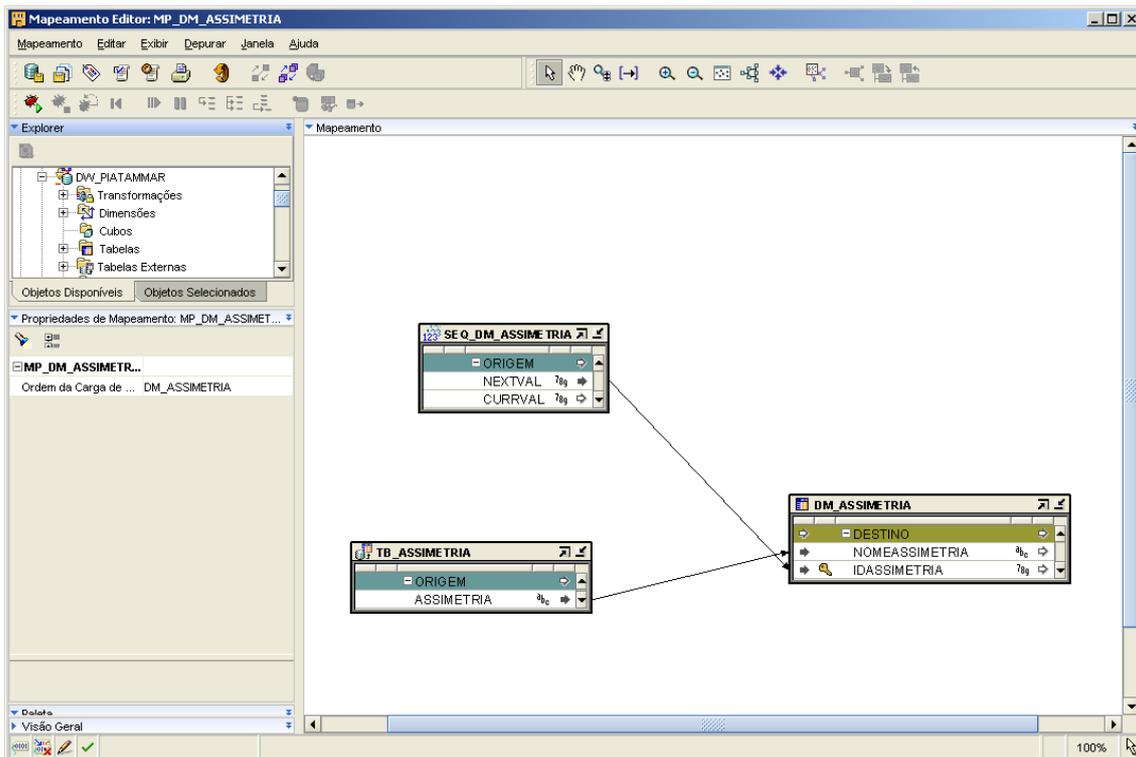


Figura 4.12: Mapeamento Editor configurando a dimensão DM_ASSIMETRIA.

Dessa forma, todo o modelo do projeto foi aplicado e devidamente mapeado (algumas transformações fizeram-se necessárias durante este processo de mapeamento), com seus respectivos dados de origem na base operacional.

5. Estudo de Caso e Resultados

5.1. Objetivo

Quando falamos em um projeto de um *Data Mart* (DM) ou um *Data Warehouse* (DW), não referimos simplesmente em projetar um banco de dados. O objetivo é pesquisar um grande repositório de informações de cunho científico de apoio à decisão, o qual apresente séries históricas, que reflitam a evolução de fatos do cotidiano da Pesquisa do Projeto Piatam Mar que devido a grande área de estudo e a diversidade do conhecimento gera um grande volume de dados.

Este volume de dados e a diversidade da pesquisa são o foco principal deste estudo que visa buscar conhecimento/informação nesta base de dados aplicando ferramentas de busca e descoberta de conhecimento.

Assim, utilizamos neste projeto o Modelo *Star* ou Estrela. *Star schema* é o termo comum para representação de modelos de dados multidimensionais, mais usualmente denominado modelo estrela. É a estrutura básica de modelos de dados multidimensionais, sua composição típica possui uma grande entidade central denominada fato e um conjunto de entidades menores denominadas dimensões, dispostas ao redor formando uma estrela, conforme veremos adiante.

5.2. Identificação dos Fatos de Negócio

O Projeto PIATAM mar é constituído por Base de Dados de natureza ambiental e sócio-econômica, desta forma, podemos identificar os seguintes processos de negócio:

5.2.1. Fato Coleta de Dados

Todos os dados do Projeto, com exceção dos referentes a bibliografia, são georreferenciados, significando que estão ligados a um destino geográfico referenciado na base de dados, ou seja, são dados coletados em um local específico previamente selecionado como de interesse do projeto. O fato analisado poderá demonstrar como evolui a coleta de dados de pesquisa, o quantitativo coletado, as regiões de maior ou menor grau de exploração, assim, esse processo representa o fato Coletas (Tabela Fato Coletas).

5.2.1.1. Medidas

A operacionalização de um Data Mart tem como objetivo fundamental a consulta de dados históricos, normalmente totalizados por períodos de tempo, por região e outras combinações de classificação de uma informação. No final, o que se deseja é obter valores numéricos e sua evolução ou não no tempo.

A esses valores denominamos de medidas ou métricas. As medidas se classificam em dois tipos:

- **Valores aditivos:** referem-se ao fato sobre os quais podem ser aplicadas as operações de soma, subtração e média.
- **Valores não aditivos:** referem-se aos fatos que não podem ser manipulados livremente, como valores de percentuais ou relativos. Na realidade são os indicadores de desempenho dos fatos.

Para tais valores, os cálculos devem ser realizados com os dados absolutos que os baseiam. Por exemplo, nível de intensidade, índices ou indicadores de performance são não aditivos e por isso estáticos. Sua validade limita-se ao momento em que são obtidos e sua soma ao longo do tempo não tem significado, contudo, são úteis em futuras manipulações.

Existem várias medidas que podem estar associando ao fato Coletas, como: volume de coleta, locais de coletas e período da coletas.

- Volume de coleta, variável que corresponde a quantidade de coletas realizadas;
- Locais de coleta, variável que indica os principais locais de coletas de dados dentro da região delimitada do projeto;
- Período da coleta, indicador do tempo em que ocorrem as coletas de dados.

Em um DW, uma determinada medida pode possuir uma hierarquia de composição de seu valor. À entidade fato, independente de possuir ou não todos os elementos dessa hierarquia, devemos dar fundamental importância para à sua composição.

Neste trabalho os elementos básicos (as medidas) são valores exclusivamente aditivos. Devemos entender que todo o fato tem medidas aditivas em sua origem e que os não aditivos não podem existir sem os dados de origem.

5.2.1.2. Dimensões

São os elementos que participam de um fato, assunto de negócio. São as possíveis formas de visualizar os dados, ou seja, são os “por” dos dados: “por período”, “por região”, “por tema”, “por pesquisador”, etc.

As dimensões determinam o contexto do assunto de negócio, conforme segue:

5.2.1.2.1. Localização

Essa dimensão é formada pelos seguintes níveis hierárquicos: Marco PIATAM Mar e Ponto de Coleta. Onde os Marcos são sub-áreas de pesquisas pré-determinadas como de interesse para pesquisa, sendo três ambientais e três portuárias. Já os pontos de coleta são pontos, linhas ou polígonos onde são coletadas amostras e realizadas pesquisas.

5.2.1.2.2. Tema

Compreende os diversos tipos de dados coletados como:

Abiótico: Química da Água, Climatologia, Geologia;

Biótico: Bentos, Fitoplâncton, Peixes Estuarinos, Zooplâncton.

5.2.1.2.3. Período

São os diferentes períodos que são coletados os dados nos Pontos de Coletas dos Marcos Piatam Mar. Tais períodos são complexos devido a grande influência da vazão da foz do rio Amazonas, do oceano Atlântico, dos ventos, dentre outros fatores naturais.

5.2.1.3. Modelo Relacional

O DW é modelado para ambientes de banco de dados relacionais. Para a construção e implementação do nosso modelo foram aplicados os mesmos princípios de modelagem de dados Entidade-Relacionamento, conforme podemos constatar no modelo do fato Coleta de Dados (figura 5.1):

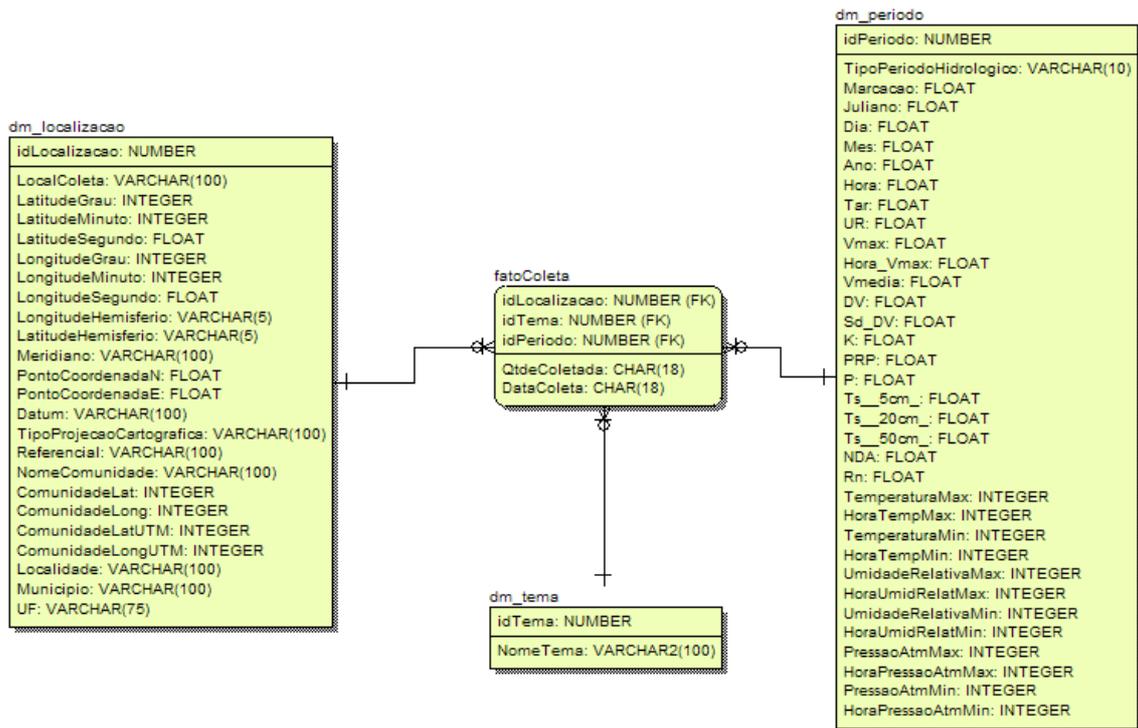


Figura 5.1: Modelo coleta de dados

O Data Mart é criado a partir do modelo de dados operacional, onde as dimensões que importam são aquelas que caracterizam cada coleta realizada e registrada no banco de dados e que interesam manter dados históricos.

Com o modelo podemos analisar as coletas ao longo do tempo.

Podemos obter as coletas realizadas em cada ponto de coleta e ainda com os períodos classificados por tema (tabela 5.1).

TEMA	MUNICÍPIO	UF	JANEIRO	FEVEREIRO	MARÇO
Química da Água	BREU BRANCO	PA	6	4	2
.	CURRALINHO	PA	10	0	2
TOTAL Química da Água			16	4	4
Climatologia	BREU BRANCO	PA	3	0	4
.	CURRALINHO	PA	5	2	4
TOTAL Climatologia			8	2	8
Geologia	BREU BRANCO	PA	3	3	1
.	CURRALINHO	PA	2	1	1
TOTAL Geologia			5	4	2
Bentos	BREU BRANCO	PA	11	1	6
.	CURRALINHO	PA	4	4	6
TOTAL Bentos			15	5	12
Fitoplâncton	BREU BRANCO	PA	5	5	0
.	CURRALINHO	PA	0	0	0
TOTAL Fitoplâncton			5	5	0
Peixes Estuarinos	BREU BRANCO	PA	0	7	11
.	CURRALINHO	PA	7	10	11
TOTAL Peixes Estuarinos			7	17	22
Zooplâncton	BREU BRANCO	PA	2	12	5
.	CURRALINHO	PA	1	5	5
TOTAL Zooplâncton			3	17	10

Tabela 5.1: Visão I: Coleta de dados

Esta consulta permite analisar totais por tema e por período em uma determinada localidade.

O modelo multidimensional permite realizar análises de várias perspectivas da visão dos dados do banco de dados. Em uma outra consulta podemos obter uma listagem de outro ponto de vista (tabela 5.2):

MUNICÍPIO	UF	TEMA	JANEIRO	FEVEREIRO	MARÇO
CURRALINHO	PA	Bentos	4	4	6
.	.	Climatologia	5	2	4
.	.	Fitoplâncton	0	0	0
.	.	Geologia	2	1	1
.	.	Peixes Estuarinos	7	10	11
.	.	Química da Água	10	0	2
.	.	Zooplâncton	1	5	5
BREU BRANCO	PA	Bentos	11	1	6
.	.	Climatologia	3	0	4
.	.	Fitoplâncton	5	5	0
.	.	Geologia	3	3	1
.	.	Peixes Estuarinos	0	7	11
.	.	Química da Água	6	4	2
.	.	Zooplâncton	2	12	5

Tabela 5.2: Visão II: Coleta de dados

As informações podem ser analisadas agora de um outro ponto de vista. Essa mudança na perspectiva da análise é denominada, segundo os conceitos OLAP, *dice*, que é uma forma de mudança das dimensões a serem visualizadas [Machado 2006].

Concluimos que as dimensões (classificações, agrupamentos em que se deseja analisar os fatos) em um modelo *star schema* são os balizadores de análise dos dados em nosso banco de dados.

Os demais fatos de negócio seguem o mesmo raciocínio e serão apenas descritos no decorrer da análise do trabalho.

ABIÓTICO

É um tema onde estão armazenadas tabelas fatos com informações abióticas. Essas informações caracterizam-se pelos dados coletados de caráter não biológicos.

5.2.2. Fato Coleta Química da Água

Este fato armazena os dados químicos da água coletada em determinada localização.

5.2.2.1. Medidas

Vejamos as diversas medidas que podem estar associadas a coleta química da água, como:

- Volume de coleta, variável que corresponde a quantidade de coletas realizadas de dados abióticos;
- Locais de coleta, variável que indica os principais locais de coletas de dados dentro da região delimitado do projeto;
- Período da coleta, indicador do tempo em que ocorrem as coletas de dados.
- Nitrogênio Total, é a somatória de todo material de origem do nitrogênio (Standard Methods 1985);
- Sólidos Totais, somatória dos sólidos totais e dos dissolvidos (Standard Methods 1985);
- Coliformes Total, são as medidas das bactérias de origem exclusivamente fecal (Standard Methods 1985);
- Carbono Total, são as medidas totais de carbono.

5.2.2.2. Dimensões

5.2.2.2.1. Localização

Essa dimensão é formada pelos seguintes níveis hierárquicos: Marco Piatam Mar e Ponto de Coleta de seus respectivos dados abióticos.

5.2.2.2.2. Tempo

São os diferentes período que são coletados os dados nos Pontos de Coletas dos Marcos Piatam Mar.

5.2.2.3. Modelo Relacional

Aplicaremos novamente os princípios de modelagem de dados Entidade-Relacionamento, conforme podemos constatar no modelo do fato Coleta Química da Água (figura 5.2):

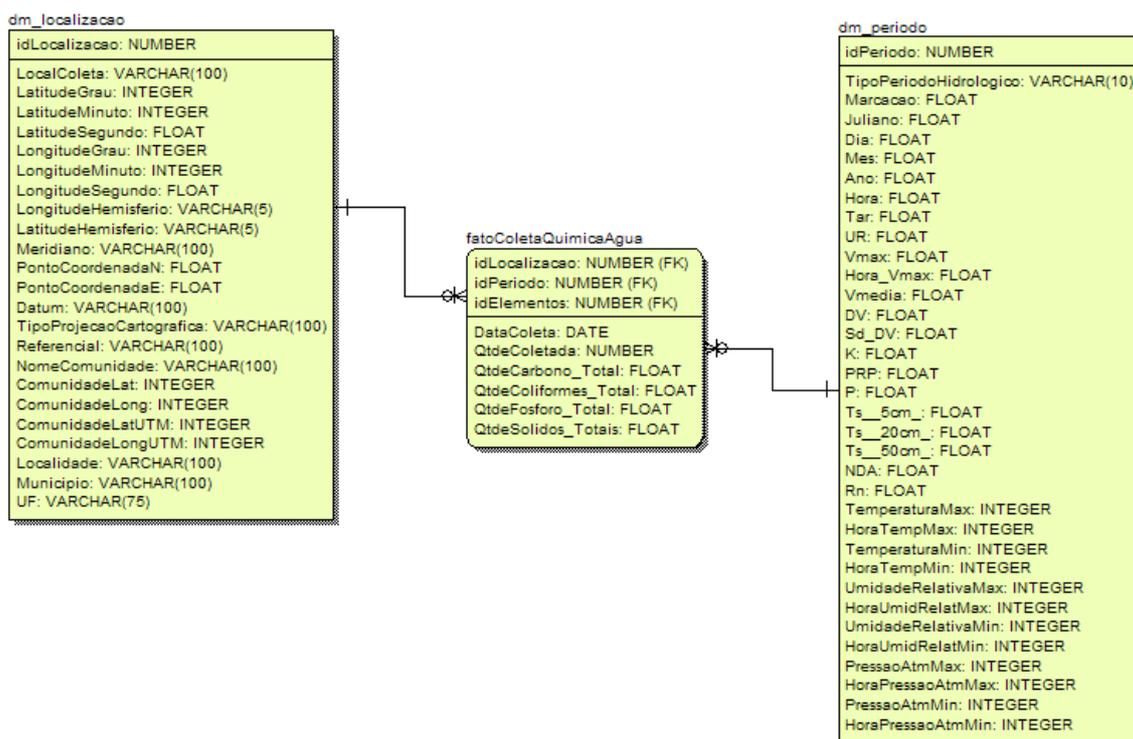


Figura 5.2: Modelo coleta química da água

Neste modelo podemos analisar as coletas ao longo do tempo. Também podemos obter as coletas realizadas em cada ponto de coleta classificados por períodos.

Conforme vimos no fato Coleta de Dados, o modelo multidimensional permite a análise em vários ângulos de visão do negócio, resumizando os dados conforme a necessidade que se deseje avaliar.

5.2.3. Fato Coleta Climatologia

Este fato armazena os dados de climatologia coletados em determinada localização.

5.2.3.1. Medidas

Há diversas medidas que podem estar associando ao fato Climatologia, como:

- Volume de coleta, variável que corresponde a quantidade de coletas realizadas de dados climatológicos;
- Locais de coleta, variável que indica os principais locais de coletas de dados dentro da região delimitado do projeto;
- Período da coleta, indicador do tempo em que ocorrem as coletas de dados.

5.2.3.2. Dimensões

5.2.3.2.1. Localização

Essa dimensão é formada pelos seguintes níveis hierárquicos: Marco PIATAM Mar e Ponto de Coleta de seus respectivos dados abióticos de climatologia.

5.2.3.2.2. Tempo

São os diferentes período que se realizam as escursoes para coleta de dados nos Marcos Piatam Mar.

5.2.3.3. Modelo Relacional

Aplicaremos novamente os princípios de modelagem de dados Entidade-Relacionamento, conforme podemos constatar no modelo do fato Coleta Climatologia (figura 5.3):

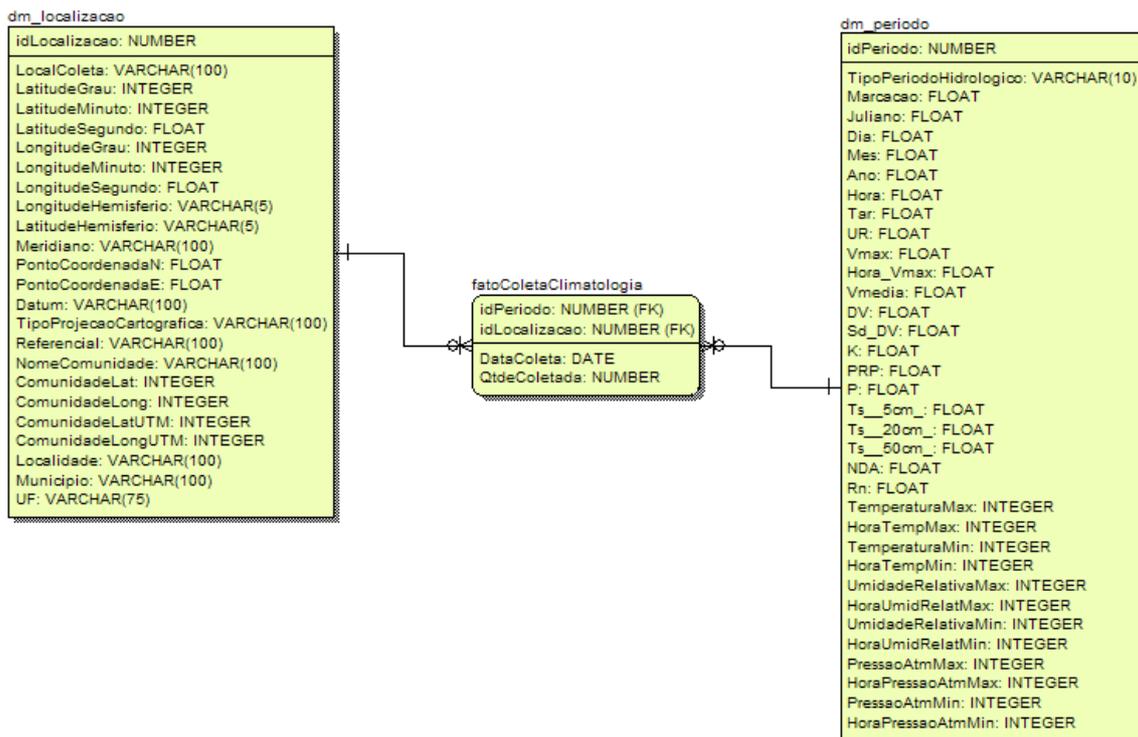


Figura 5.3: Modelo coleta climatologia

Para este modelo também podemos analisar as coletas ao longo do tempo. Podemos obter as coletas realizadas em cada ponto de coleta classificados por períodos.

Essas diferentes visões dos dados totalizados são possíveis graças a modelagem multidimensional do negócio, onde é possível sumarizar os dados do ângulo necessário conforme a necessidade do pesquisador.

5.2.4. Fato Coleta Geologia

Este fato armazena os dados geológicos coletados em determinada localização.

5.2.4.1. Medidas

Há diversas medidas que podem estar associadas ao fato Geologia, como:

- Volume de coleta, variável que corresponde a quantidade de coletas realizadas de dados climatológicos;
- Locais de coleta, variável que indica os principais locais de coletas de dados dentro da região delimitado do projeto;

- Período da coleta, indicador do tempo em que ocorrem as coletas de dados.
- Valor Média, variável de totalização das médias de classificação da amostra.
- Valor Seleção, variável de totalização de seleções da amostra.
- Valor Assimetria, variável de totalização da assimetria da amostra.
- Valor Curtose, variável de totalização do tipo de classificação textural da amostra.

5.2.4.2. Dimensões

5.2.4.2.1. Localização

Essa dimensão é formada pelos seguintes níveis hierárquicos: Marco PIATAM Mar e Ponto de Coleta de seus respectivos dados abióticos geológicos.

5.2.4.2.2. Tempo

São os diferentes período que se realizam as escursoes para coleta de dados nos Marcos Piatam Mar.

5.2.4.2.3. Granularidade

Compreende as características de classificação de granularidade das amostra de dados coletados: Areia fina, Areia grossa, Areia média, Areia muito fina, Silte grosso e Silte médio.

5.2.4.2.4. Seleção

São os tipos de seleção textural das amostra de dados coletados: moderadamente selecionado, muito bem selecionado, muito pobremente selecionado e pobremente selecionado.

5.2.4.2.5. Curtose

Características de classificação de textura que pode ser: Leptocúrtica, Extremamente leptocúrtica, Mesocúrtica, Muito leptocúrtica, Muito platicúrtica e Platicúrtica.

5.2.4.2.6. Assimetria

Compreende a características de classificação de textura das amostra de dados coletados, que pode ser: Muito negativa e Muito positiva.

5.2.4.3. Modelo Relacional

Aplicaremos outra vez os princípios de modelagem de dados Entidade-Relacionamento, conforme podemos constatar no modelo do fato Coleta Geológica (figura 5.4):

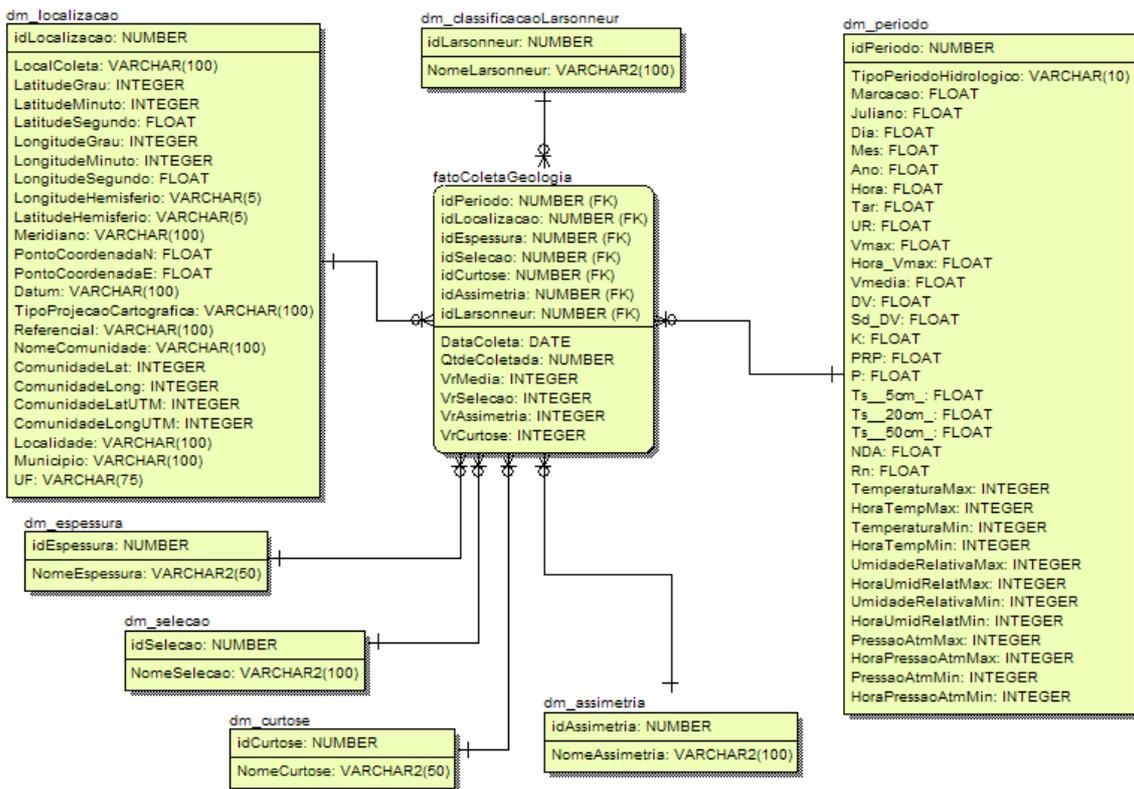


Figura 5.4: Modelo coleta geologia

Neste Data Mart, criado a partir do modelo de dados operacional, destacamos as dimensões que importam para o tema pesquisado, são aquelas que caracterizam cada coleta Geológica realizada e registrada no banco de dados e que interessam manter dados históricos.

Com o modelo podemos analisar as coleções de dados ao longo do tempo, como as coletas realizadas em cada ponto de coleta e ainda com os períodos classificados por espessura, e ou seleção, e ou curtose, e ou assimetria e ou classificação Larsonneur.

Este modelo multidimensional permite realizar análises de várias perspectivas da visão dos dados do banco de dados. Essas diferentes visões dos dados totalizados são possíveis uma vez que o modelo utiliza as técnicas de modelagem multidimensional. O pesquisador poderá utilizar várias combinações (*dices*) de características da geologia do ponto de coleta, em um determinado período desejado.

BIÓTICO

Caracteriza-se pelos dados coletados de caráter biológicos.

5.2.5. Fato Coleta Peixes Estuarinos

Este fato armazena os dados de peixes de estuários coletados em determinada localização.

5.2.5.1. Medidas

As medidas que podem estar associadas ao fato Coleta Peixes Estuarinos, são:

- Volume de coleta, variável que corresponde a quantidade de coletas realizadas de dados bióticos;
- Locais de coleta, variável que indica os principais locais de coletas de dados dentro da região delimitado do projeto;
- Período da coleta, indicador do tempo em que ocorrem as coletas de dados.
- Média do Comprimento, variável que indica a média de comprimento dos espécimes coletados;
- Média do Peso, indicador de média do peso dos espécimes coletados.

5.2.5.2. Dimensões

5.2.5.2.1. Localização

Essa dimensão é formada pelos seguintes níveis hierárquicos: Marco PIATAM Mar e Ponto de Coleta de seus respectivos dados abióticos geológicos.

5.2.5.2.2. Tempo

São os diferentes período que se realizam as escursoes para coleta de dados nos Marcos Piatam Mar.

5.2.5.2.3. Utilização Habitat

Indica como os espécimes utilizam um determinado local no momento da coleta de dados nos Marcos Piatam Mar.

5.2.5.2.4. Sexo

Classificador do sexo da espécie, pode ser: M – Macho, F – Fêmea e I - Indeterminado.

5.2.5.2.5. Nome Científico

São as denominações científicas atribuídas a uma determinada espécie.

5.2.5.2.6. Família

São as classificações de agrupamento científicas com características em comum, atribuídas a uma determinada espécie.

5.2.5.2.7. Espécie

São as denominações atribuídas a indivíduos com características semelhantes.

5.2.5.3. Modelo Relacional

Aplicaremos os princípios de modelagem de dados Entidade-Relacionamento, conforme podemos constatar no modelo do fato Coleta Peixes Estuarinos (figura 5.5):

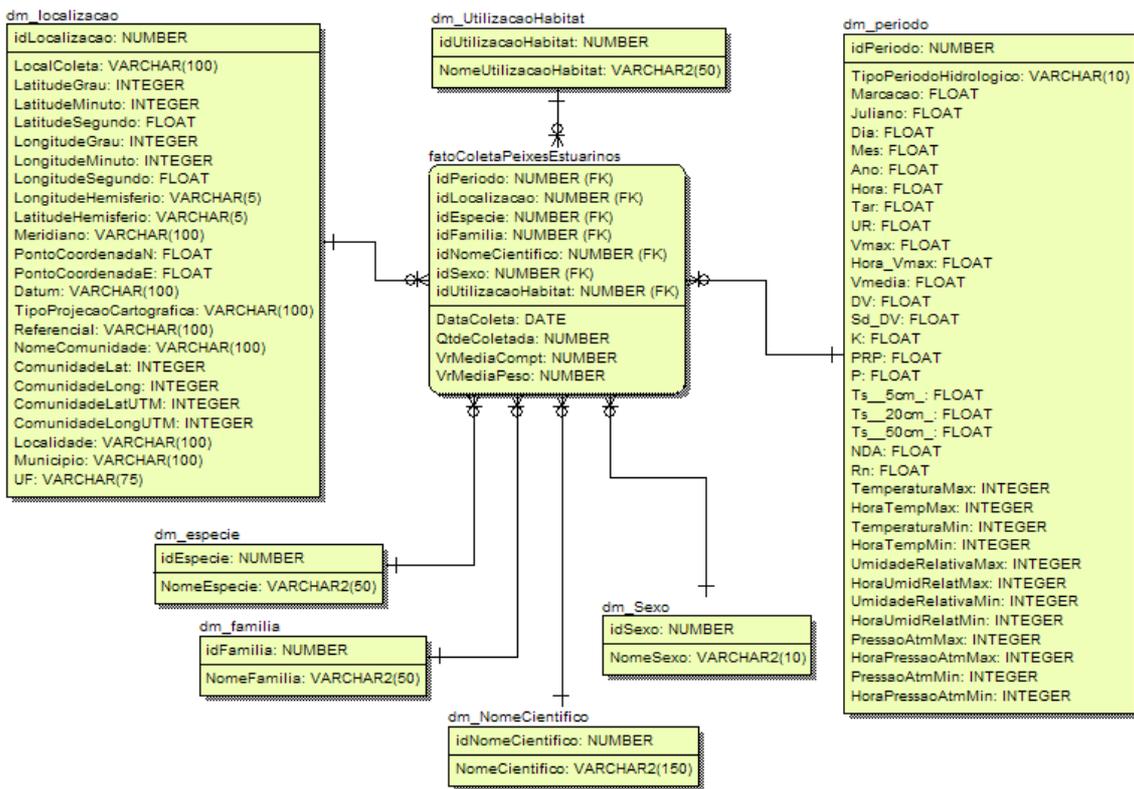


Figura 5.5: Modelo coleta peixes estuarinos

Podemos identificar, neste Data Mart, as dimensões que importam para o tema pesquisado, as quais caracterizam cada coleta de Peixes Estuarinos realizada e registrada no banco de dados e que interessam manter séries históricas.

Com já percebemos, o modelo nos permite analisar as coleções de dados ao longo do tempo, quantificar as coletas realizadas em cada ponto de coleta e classificá-las e sumariá-las por Uso do Habitat, Sexo, Nome Científico, Família e Espécie.

Como já vimos, este é um modelo multidimensional, e nos permitirá realizar análises de várias perspectivas de interesse sobre o tema. Com essas diferentes visões, o pesquisador poderá utilizar várias combinações para analisar séries históricas conforme a necessidade e interesse da pesquisa.

5.3. Operacionalização do *Data Warehouse*

Vimos anteriormente como é feita a modelagem multidimensional do negócio, como o pesquisador pode consultar seus dados em diversas dimensões, nas diversas perspectivas dos dados.

Ainda restou a dúvida de como manter as informações nos modelos analisados. Para tal, são utilizados os índices cruzados, estruturas de hierarquia e

dimensões de tempo. Essas estruturas permitem a navegação pelas dimensões através de, principalmente, tabelas que permitem a visualização de tendências e comportamentos do tema pesquisado.

As ferramentas que permitem a operacionalização do DW são conhecidas como SGB Multidimensionais. Em particular utilizaremos o *Oracle Data Mart Suite* (ODMS), uma completa ferramenta para construir e manter *Data Marts*. Composto de um conjunto de produtos que incluem *software* de banco de dados, ferramenta de modelagem, ferramenta de extração, transformação e transporte (ETT) de dados e ferramenta de análise e relatório.

Adiante, serão apresentadas cada ferramenta e uma descrição de sua participação em cada etapa da construção e manutenção do projeto.

5.3.1. Oracle Data Mart Designer

Ferramenta de construção das estruturas do Data Mart, possui a capacidade de realizar engenharia reversa no banco de dados fonte. Provê as ferramentas necessárias para projetar graficamente o destino na base de dados para o Data Mart. Objetos tabelas, visões e chaves sem a necessidade de interação manual.

5.3.2. Oracle Warehouse Builder

O aplicativo permite e facilita o processo de projetar e executar a extração, transformação e transporte dos dados operacionais (OLTP) do banco de dados fonte, para uma base de dados de um *Data Mart*.

5.3.3. BaseViews e MetaViews

As representações dos bancos de dados físicos são denominadas *BaseViews*, para tal, utiliza-se a ferramenta gráfica de edição *BaseView*.

Os *Metaviews* permitem a criação e representação do negócio, em nosso caso, dos temas de interesse da pesquisa. Componentes da ferramenta podem incluir colunas e fórmulas customizadas.

A ferramenta *Oracle Data Flow Editor* permite a criação de planos que representam o processo de acesso, carregamento e transformação dos dados. O *Data Collection Agent* processa e executa esses planos.

5.3.4. Oracle Interprise Edition

A central de administração do ODMS é o *Oracle Interprise Edition* que provê alto desempenho nas consultas e robusto gerenciamento de dados. Comparando com aplicações OLTP, os *Data Marts* são criados usando técnicas diferentes. São técnicas cuja característica principal é a complexidade de consultas *ad hoc* executadas em volumes muito grandes de dados.

5.3.5. Oracle Web Application Server

Essa ferramenta atende aos requisitos necessários para acessar um *Data Mart* na *web*. É possível criar HTML dinâmico Utilizando o *Data Mart Designer* o qual disponibiliza automaticamente programas PL/SQL executados no banco de dados *Oracle*. Também provê estrutura técnica necessária para liberar relatórios na *web*.

5.3.6. Oracle Reports

Ferramenta que proporciona um ambiente para criação de relatórios personalizados e otimizados. É composto pela seguinte estrutura:

- *Report Builder*: componente com assistente de criação de relatórios sofisticados;
- *Graphics Builder*: componente declarativo para exibição gráfica dos dados do relatório.
- *Reports Runtime* e *Graphics Runtime*: componente para desenvolvimento em ambiente cliente/servidor;
- *Reports Server*: componente que possibilita o desenvolvimento de relatórios *multi-tier*. Através desse componente é possível liberar a aplicação em ambiente *multi-tier* que usa a tecnologia *caching* avançada para prover balanceamento de carga dinâmico.

Cada ferramenta participou de uma etapa do processo de criação e manutenção do DW e são de fundamental importância para o projeto, sem as quais não seria possível manter as informações no tempo em que elas são úteis para a análise, inviabilizando a operacionalização do DW.

5.4. Análise de resultados

Concluindo o estudo de caso e após a construção de todo o ambiente de DW no OWB e suas devidas validações através da ferramenta *Design Center* (figura 5.6), pode-se observar que a insuficiência de dados não viabilizou a análise criteriosa dos resultados. Visto que, não há como ocorrer a sumarização dos fatos de negócio, e conseqüentemente, visualizar funcionalmente componentes vitais de um DW como os Cubos. Suas dimensões, principalmente Localização, Tempo e Temas ficam sem uma análise mais detalhada.

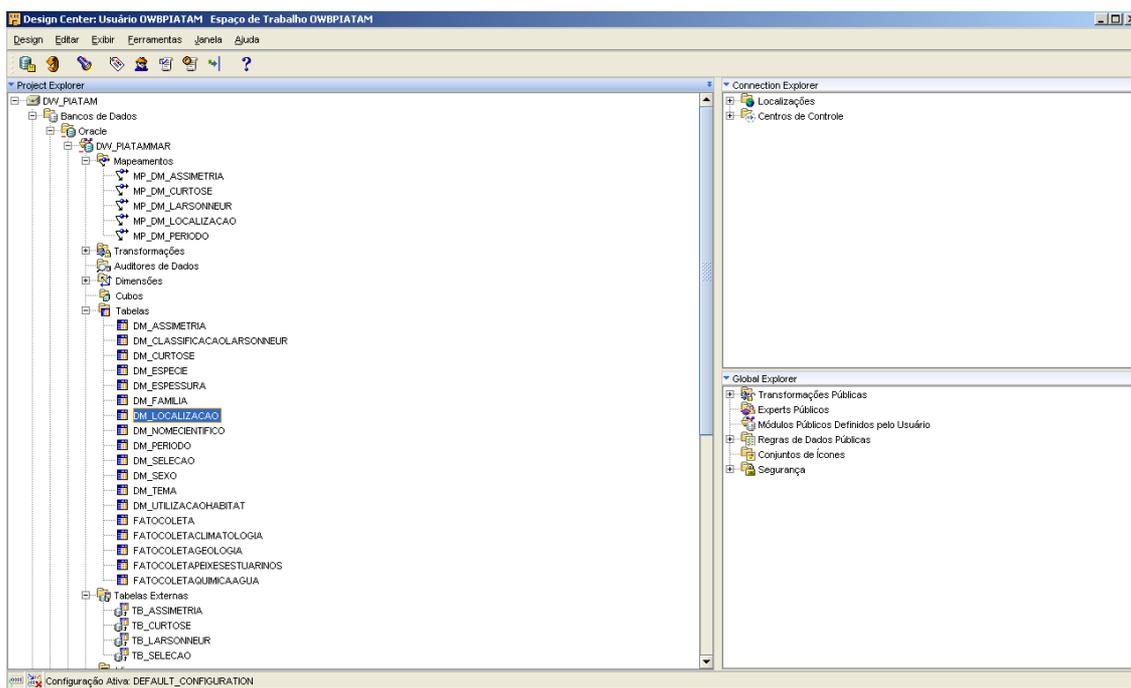


Figura 5.6: Ambiente do *Design Center*

Utilizando o *Design Center*, acessamos as tabelas do DW que, através dos mapeamentos e transformações, interagem com as tabelas da base de dados operacional. A partir desta ferramenta, podemos acessar e manipular todas as funcionalidades do DW executando, dessa forma, a classificação e sumarização das informações que são o objetivo deste trabalho.

Assim, mostramos algumas tabelas do DW com pouco ou nenhum povoamento o que inviabiliza uma análise mais prática e detalhada dos resultados deste trabalho. Na seqüência, as dimensões de Localização (figura 5.7), de Período (figura 5.8), e um exemplo de Tema (figura 5.9), que são as visualizações de suma importância para os demais fatos de negócio, que são o foco este trabalho.

IDLOCALIZA...	NOME COMUN...	LOCAL COLE...	COMUNIDAD...	LATITUDEGR...	COMUNIDAD...	LATITUDEMIN...	COMUNIDAD...	LATITUDESE...	COMUNIDAD...	LONGITUDEG...	LONGITUDEM
1	N.S. APAREC...	INTERIOR	0	0	0	0	0	0	0	0	0
2	CRISTO RED...	URBANO	0	0	0	0	0	0	0	0	0
3	SAGRADO	URBANO	0	0	0	0	0	0	0	0	0

Exibindo 3 Linhas de 3

Figura 5.7: Visualizador de Dados Relacional: DM_LOCALIZAÇÃO

IDPERIODO	MARCAÇÃO	TIPO PERIODO HIDROLOGICO	JULIANO	DIA	MES	ANO	HORA
1		cheia		1	1	2008	8
2		enchente		5	1	2008	12
3		vazante		10	1	2008	18
4		seca		20	1	2008	23
5		cheia		1	2	2008	8
6		enchente		5	2	2008	12
7		vazante		10	2	2008	18
8		seca		20	2	2008	23
9		cheia		1	3	2008	8
10		enchente		5	3	2008	12
11		vazante		10	3	2008	18
12		seca		20	3	2008	23

Exibindo 12 Linhas de 12

Figura 5.8: Visualizador de Dados Relacional: DM_PERIODO

NOMETEMA	IDTEMA	
1	Química da Água	1
2	Climatologia	2
3	Geologia	3
4	Bentos	4
5	Fitoplâncton	5
6	Peixes Estuarinos	6
7	Zooplâncton	7

Exibindo 7 Linhas de 7

Figura 5.9: Visualizador de Dados Relacional: DM_TEMA

Sendo assim, partimos para um povoamento aleatório de parte da base de dados e por conseguinte das tabelas fato, gerando um conteúdo mínimo para a comprovação da funcionalidade e interessabilidade da aplicação do DW na base.

Partindo dessa premissa, analisamos agora os fatos e dimensões do DW em alguns relacionamento envolvendo as dimensões de Tema com a Localização e o Tempo. Lembramos que a dimensão Tempo ficou prejudicada devido a falta de dados na coleta. Então estamos utilizando apenas parte do período hidrológico como dimensão.

5.4.1. Peixes estuarinos

Os resultados apresentaram uma relação interessante entre o local de coleta, espécie, período hidrológico, sexo, utilização do habitat, quantidade coletada e valores médios de comprimento e peso por espécie (figura 5.10).

The screenshot shows the SQL Navigator interface with a query executed. The query selects data from several tables: localcoleta, categoriataxonmica, bdpiatam, dm_sexo, and dm_utilizacaohabitat. The results are displayed in a table with 13 rows and 9 columns.

Row #	LOCALCOLETA	ESPECIE	PERIODOHIDROLOGICO	SEXO	UTILIZACAOHABITAT	QTDECOLETADA	VRMEDIACOMPT	VRMEDIAPESO
1	Lagoa Salina - Região Bragantina	Oligopites saurus	enchente	F	Reproduzindo	1	20	0,11
2	Região Estuarina do Município de Amapá	seca	seca	I	Berçário	2	197,5	0,078
3	Estuário do Rio Curuçá	Colomesus psittacus	cheia	F	Reproduzindo	2	23,5	0,235
4	Estuário do Rio Curuçá	Mugil curema	cheia	F	Reproduzindo	2	22	0,225
5	Lagoa Salina - Região Bragantina	Moenkhausia colletti	enchente	F	Berçário	1	24	0,26
6	Estuário de São Caetano de Odívelas	Gobioides broussonnetii	seca	F	Berçário	1	21	0,16
7	Estuário do Rio Curuçá	Achirus lineatus	cheia	I	Berçário	1	23	0,21
8	Bacias do Igarapé da Fortaleza e do Rio Curiaú	Satanoperca jurupari	cheia	F	Berçário	1	195	0,084
9	Bacias do Igarapé da Fortaleza e do Rio Curiaú	Triportheus curtus	cheia	F	Berçário	1	180	0,072
10	Estuário de São Caetano de Odívelas	Ophisthomena oglinum	seca	F	Berçário	1	20	0,15
11	Lagoa Salina - Região Bragantina	Cynoscion acoupa	enchente	F	Berçário	1	20	0,11
12	Região Estuarina do Município de Sucuriú	Bagre bagre	cheia	F	Reproduzindo	2	409	519,335
13	Bacias do Igarapé da Fortaleza e do Rio Curiaú	Serrasalmus calmoni	cheia	F	Reproduzindo	1	230	0,126

The interface also shows the SQL query used to generate these results:

```

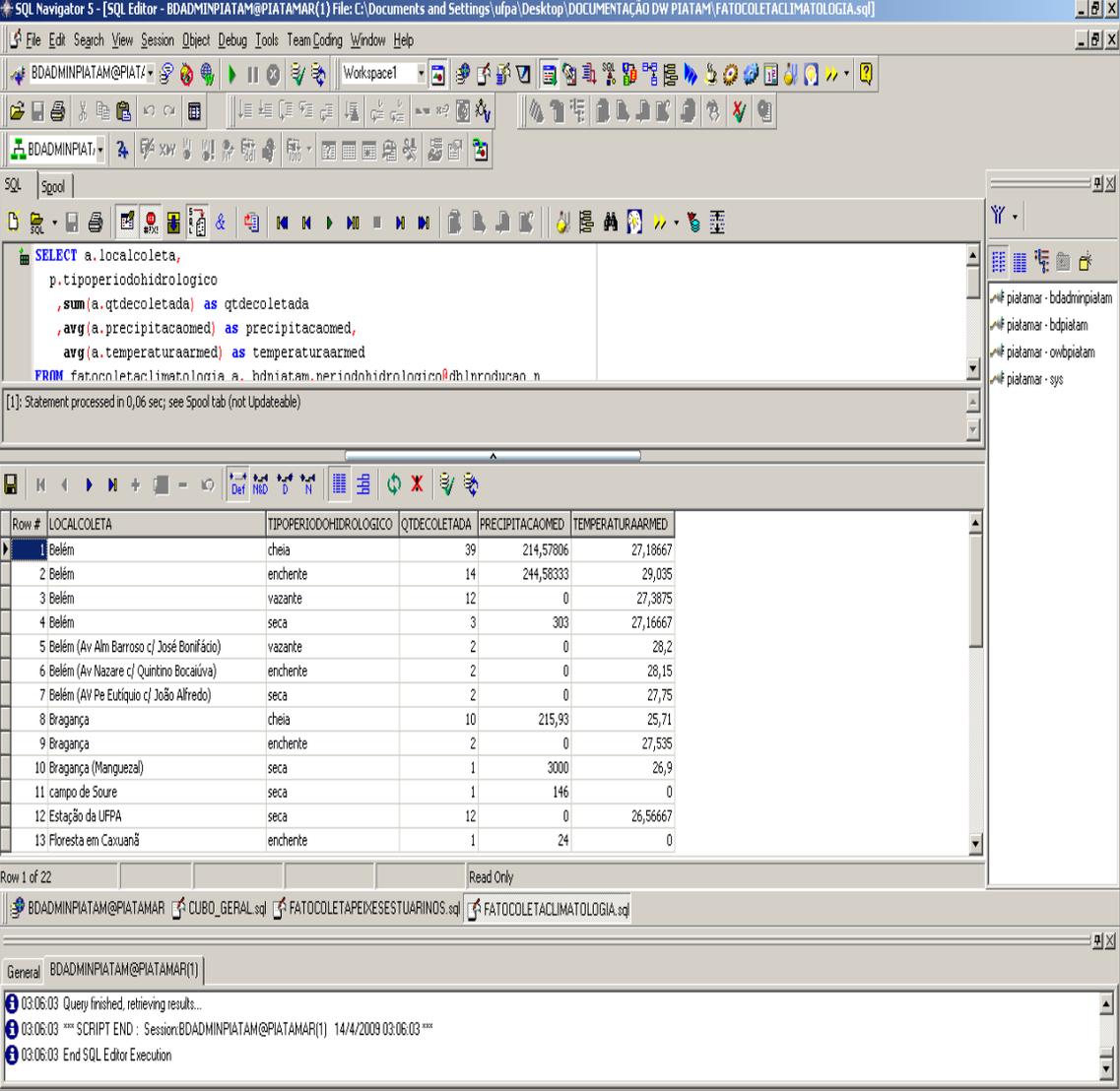
SELECT a.localcoleta,
(SELECT descricao FROM categoriataxonmica@dblproducao where idcategoriataxonmica = a.idespecie) as especie,
(select p.tipoperiodohidrológico from bdpiatam.periodohidrológico@dblproducao p where p.idperiodohidrológico = a.idperiodo) as peri
(select nomesexo from dm_sexo where idsexo = a.idsexo) as sexo,
(SELECT nomeutilizacaohabitat FROM dm_utilizacaohabitat where idutilizacaohabitat = a.idutilizacaohabitat) as utilizacaohabitat.
  
```

Log messages at the bottom indicate the query finished at 02:46:26.

Figura 5.10: Visão do SQL Navigator Fato Coleta Peixes Estuarinos

5.4.2. Climatologia

Apesar do baixo povoamento da base de dados, os resultados também apresentaram confiável relação entre o local de coleta, período hidrológico, quantidade coletada e valores médios de precipitação e temperatura do ar (figura 5.11).



The screenshot displays the SQL Navigator interface with the following SQL query and its results:

```
SELECT a.localcoleta,
       p.tipoperiodohidrologico
       ,sum(a.qtdecoletada) as qtdecoletada
       ,avg(a.precipitacaomed) as precipitacaomed,
       avg(a.temperaturaarmed) as temperaturaarmed
FROM fatocoletaclimatologia a, bdpiatam.periodohidrologico bdbprodacaon
```

Row #	LOCALCOLETA	TIPOPERIODOHIDROLOGICO	QTDECOLETADA	PRECIPTACAOMED	TEMPERATURAARMED
1	Belém	cheia	39	214,57806	27,18667
2	Belém	enchente	14	244,58333	29,035
3	Belém	vazante	12	0	27,3875
4	Belém	seca	3	303	27,16667
5	Belém (Av Alm Barroso c/ José Bonifácio)	vazante	2	0	28,2
6	Belém (Av Nazare c/ Quintino Bocaiuva)	enchente	2	0	28,15
7	Belém (AV Pe Eutiquio c/ João Alfredo)	seca	2	0	27,75
8	Bragança	cheia	10	215,93	25,71
9	Bragança	enchente	2	0	27,535
10	Bragança (Manguezal)	seca	1	3000	26,9
11	campo de Soure	seca	1	146	0
12	Estação da UFPA	seca	12	0	26,56667
13	Floresta em Caxuanã	enchente	1	24	0

Row 1 of 22

General: BDADMINPIATAM@PIATAMAR(1)

03:06:03 Query finished, retrieving results...

03:06:03 *** SCRIPT END : Session:BDADMINPIATAM@PIATAMAR(1) 14/4/2009 03:06:03 ***

03:06:03 End SQL Editor Execution

Figura 5.11: Visão do SQL Navigator Fato Climatologia

5.4.3. Química da água

Novamente, os resultados apresentaram confiabilidade relacional entre o local de coleta, período hidrológico, quantidade coletada e valores médios de dureza, PH, temperatura, turbidez, alcalinidade e condução elétrica (figura 5.12).

The screenshot shows the SQL Navigator interface with a query executed. The query is as follows:

```
SELECT a.localcoleta,
(SELECT p.tipoperiodohidrologico from bdpiatam.periodohidrologico@db1producao p where p.idperiodohidrologico = a.idperiodo) as peri
SUM(a.qtdecoletada) AS qtdecoletada,
ROUND (NVL (AVG (a.vrdurezamed), 0), 5) AS vrdurezamed,
ROUND (NVL (AVG (a.vrphmed), 0), 5) AS vrphmed.
```

The result table is displayed below the query:

Row #	LOCALCOLETA	PERIODOHIDROLOGICO	QTDECOLETADA	VRDUREZAMED	VRPHMED	VRTEMPERATURA	VRTURBIDEZ	VRALCALINIDADAMED	VRCOND_ELETRICAMED
1	Água subterrânea do município de Belém, Distrito de Ico vazante		18	0	4,6	29,1	0	122,6	192,5
2	Baía de Inajá	cheia	5	0	7	29,2	458,2	0	487,4
3	Barcarena	cheia	22	0	5,6	28,1	11	62,6	4191,3
4	Belém	seca	30	0	6,1	28,1	64,1	74,9	181,5
5	Belém	cheia	1	0	6	26	0	0	852
6	Belém	vazante	15	0	6,5	28,7	45,1	27,2	257,6
7	Bragança	enchente	41	42,8	5	28,4	26	12,6	166
8	Canal de drenagem da região metropolitana de Belém	vazante	50	0	6,4	27,7	22,9	0	204,8
9	Estuário de São Caetano de Odrievelas	seca	4	0	7,3	28,5	63	2	39300

The interface also shows the status bar indicating 'Row 1 of 9' and 'Read Only'. The bottom panel shows the execution log with the following messages:

```
03:16:02 Query finished, retrieving results...
03:16:02 *** SCRIPT END : Session:BDADMINPIATAM@PIATAMAR(1) 14/4/2009 03:16:02 ***
03:16:02 End SQL Editor Execution
```

Figura 5.12: Visão do SQL Navigator Fato Química da Água

5.4.4. Relação entre todas as dimensões do Cubo

A análise dos resultados relacional entre todas as dimensões do cubo mostrou confiabilidade, bem como, abriu um leque de opções analíticas que carecem de apoio do conhecimento específico de pesquisadores de cada tema. Além disso, sugere a aplicação de um SIG (Sistema de Informação Geográfica), o que será tratado nas considerações finais e trabalhos futuros.

Entretanto, vale ressaltar a amplitude de conhecimento que pode ser adquirido a partir do estudo dos resultados proporcionados pelo DW, pela velocidade de obtenção dos dados e pela interatividade proporcionada aos usuários/pesquisadores do Projeto (figura 5.13).

The screenshot shows the SQL Navigator interface with a query executed and results displayed in a table. The query is as follows:

```
SELECT c.localcoleta
,p.tipoperiodohidrologico
,(SELECT descricao FROM categoriataxonmica@dbIproducao where idcategoriataxonmica = pe.idespecie) as especie
,(select nomesexo from dm sexo where idsexo = pe.idsexo) as sexo
,(SELECT nomeutilizacaohabitac FROM dm_utilizacaohabitac where idutilizacaohabitac = pe.idutilizacaohabitac) as utilizacaohabitac
,SUM(pe.qtdcoletada) AS qtdcoletada neixes_estua
```

The results table contains 8 rows of data:

Row #	LOCALCOLETA	TIPOPERIODO...	ESPECIE	SEXO	UTILIZACAOH...	QTDDECOLETADA_PE...	VRMEDIACOMPT	VRMEDIAPESO	QTDDECOLETADA_CLIMA...	PRECIPITACAOMED	TEMPERATURAARMED
1	Belém	seca	Anchoa hepsetus	M	Berçário	3	25	0,3	9	303	27,16667
2	Belém	enchente	Anchoa hepsetus	M	Berçário	3	25	0,3	42	244,58333	29,035
3	Belém	enchente	Elops saurus	F	Reproduzindo	3	15	0,07	42	244,58333	29,035
4	Belém	cheia	Elops saurus	F	Reproduzindo	3	15	0,07	117	214,57806	27,16667
5	Belém	cheia	Anchoa hepsetus	M	Berçário	3	25	0,3	117	214,57806	27,16667
6	Belém	seca	Elops saurus	F	Reproduzindo	3	15	0,07	9	303	27,16667
7	Belém	vazante	Anchoa hepsetus	M	Berçário	3	25	0,3	36	0	27,3875
8	Belém	vazante	Elops saurus	F	Reproduzindo	3	15	0,07	36	0	27,3875

Figura 5.13: Visão do SQL Navigator com todas as dimensões do cubo

6. Conclusões e Trabalhos Futuros

6.1. Conclusões

No presente trabalho foi utilizado um conjunto de técnicas e ferramentas já existentes e consagradas para a solução de problemas relacionados com a descoberta de conhecimento em base de dados, mais enfaticamente aplicados em banco de dados ambiental na zona costeira amazônica. Assim, foram descritas técnicas de classificação e agrupamento com a finalidade de buscar especificidades na base de dados em estudo.

O problema, antes de todo o volume de dados a ser trabalhado, foi como organizar o banco de dados ambientais da zona costeira amazônica para possibilitar o uso de técnicas computacionais de descoberta do conhecimento; tendo em vista a singularidade da região estudada.

O pouco povoamento de algumas tabelas fato não permitiu uma validação ou análise mais crítica da aplicação das ferramentas. Tal dificuldade na análise, refere-se a verificação da base de dados e seus resultados. Em alguns casos,, fez-se necessário um povoamento aleatório da base de dados, replicando a informação a partir de um padrão analisado nas relações com pouca informação contida na base.

A partir deste povoamento, foi possível analisar os resultados obtidos pelo DW, e assim, validarmos a modelagem do DW e ratificarmos o bom desempenho. Com o estudo foi possível verificar diversos relacionamentos entre as dimensões do cubo, tais como: temas (peixes estuarinos, climatologia, química da água), localização e tempo. Entretanto, a dimensão tempo ficou prejudicada devido a baixa densidade de dados no banco.

Assim, os bons resultados obtidos e as dificuldades nos remetem a trabalhos futuros sugeridos a seguir.

A utilização de ferramentas OLAP sobre um DW é também, eficiente para uma base de dados ambiental. No entanto, devemos ter cuidados específicos no modelamento do banco de dados, bem como, o modelo de DW a ser utilizado. Neste trabalho, optamos por modelar um DW de acordo com o Projeto e importarmos para o Oracle, ao invés de criarmos dentro do próprio SGBD que oferece suporte para tal implementação.

Esta peculiaridade na construção trouxe benefícios devido a particularidades do Projeto quanto as dimensões criadas para a formação de Cubos.

A metodologia aplicada atendeu plenamente os objetivos propostos. Os testes e avaliações da modelagem mostraram aplicabilidade e eficiência da solução

proposta para a mineração de dados a partir de um Data Warehouse em soluções de banco de dados ambientais na zona costeira amazônica, abrindo assim, a aplicabilidade destas ferramentas em pesquisas interdisciplinares e em áreas geográficas de grande abrangência.

6.2. Trabalhos Futuros

Para trabalhos futuros, propõe-se experimentos em outras áreas de pesquisa ampliando a possibilidade de descoberta de conhecimento, bem como, o maior povoamento da base de dados visando uma varredura mais ampla das informações armazenadas. Além disso, seria bastante interessante a aplicação em outras bases de dados e com outras ferramentas disponíveis no mercado. Podendo também ser feita uma análise de performance entre os SGBD's mais robustos existentes no mercado e das aplicações utilizadas.

Outro estudo interessante seria a aplicabilidade de um SIG (Sistema de Informação Geográfica) agrado ao DW e/ou ao OLAP, dando assim, uma melhor visibilidade espacial do problema.

Referências Bibliográficas

- [Agrawal 1994] Agrawal R. and Srikant R., “**Fast algorithms for mining association rules**”. In *VLDB-94*, 1994.
- [Andreatto, 1999] Andreatto, R. **Construindo um Data Warehouse e Analisando suas Informações com Data Mining e OLAP**. Monografia Final de Curso. Faculdade de Ciências Administrativas, Faculdade de Valinhos. 1999.
- [Bradley, Fayyad & Mangasarian 1998] Bradley, P., U. Fayyad, & O. Mangasarian. **Data Mining: Overview and Optimization Opportunities**. Technical Report MSR-TR-98-04, Microsoft Research Report, Redmond, WA.
- [Burdick 2007] Burdick, D. et al. **OLAP over uncertain and imprecise data**. The VLDB Journal (2007), n° 16, pag. 123-144.
- [Carvalho 2004] Carvalho, B. F. **Arquiteturas de Ferramentas OLAP**. SQL Magazine, Rio de Janeiro, ano 1, ed. 9, p.12-16, 2004.
- [Carvalho 2006] Carvalho, L. A. V. de; **Data Mining – A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. Ciência Moderna LTDA, Rio de Janeiro, 2006.
- [Cavalcanti 1998] Cavalcanti, M. C.; Oliveira, P. C. S.; Monteiro, R. R.; SOARES, V. J. **Análise Comparativa de Ferramentas OLAP**. UFRJ, 1998.
- [Chatterjee & Price, 1977] Chatterjee, S.; Price, B. **Regression analysis by example**. John Wiley e Sons, New York, 1977.
- [Codd 1970] Codd, E. F.; **A Relational Model of Data for Large Shared Data Banks**. Communications of the ACM, v. 13, n. 6, pp. 377-387, Association for Computing Machinery Inc, 1970.
- [Codd 1993] Codd, E. F.; **Providing OLAP (On-Line Analytical Processing) to User Analysts: Na IT Madate**. E. F. Codd and Assoc., 1993.
- [Date 2000] Date, C. J. **Introdução a Sistemas de Banco de Dados**. Campus. Rio de Janeiro, 2000.
- [Dubois et al. 2006] Dubois, D.; Hüllermeier, E.; Prade, H. **A systematic approach to the assessment of fuzzy association rules**. Data Mining Know Disc (2006) 13:167–192 DOI 10.1007/s10618-005-0032-4
- [Fayyad 1996a] Fayyad, U., G. Oiatetsky-Shapiro, & P. Smyth (1996a). **From data mining to knowledge discovery: an overview**. Em *Advances in Knowledge Discovery & Data Mining*, pp. 1-34.
- [Falkenberg 1998] Falkenberg, E. D.; et al. **A framework of information systems concepts FRISCO Report**. International Federation for Information Processing & Department of Computer Science, University of Leiden: The Netherlands, 1998.

- [Ferrari 1997] Ferrari, R. **Viagem ao SIG: planejamento estratégico, visualização, implantação e gerenciamento de sistemas de informação geográfica.** Curitiba: Editora Sagres, 1997.
- [Fonseca 2007] Fonseca, M. P. A. da. **Classificação Bayesiana de Grandes Massas de Dados em Ambientes ROLAP.** COPPE/UFRJ, 2007.
- [FONTELES-FILHO 1989] FONTELES-FILHO, A. A. **Recursos Pesqueiros: Biologia e Dinâmica Populacional.** Fortaleza: Imprensa Oficial do Ceará. 1989, 296p.
- [Gardner 1998] Gardner, S. R. **Building the Data Warehouse.** Communications of the ACM 41(9), 52-60.
- [Han & Kamber 2001] Han, J.; Kamber, M. **Data Mining: Concepts and Techniques.** Academic Press, USA, 2001.
- [Han & Kamber, 2006] Han, J.; Kamber, M. **Data Mining: Concepts and Techniques.** Elsevier, USA, 2006.
- [Kimball 2002] Kimball, R. **The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling.** Second Edition. Wiley, 2002.
- [Kimball 1998] Kimball, R. et al. **The Data Warehouse Lifecycle Toolkit.** Wiley, 1998.
- [Leitão 2000] Leitão, C. L. **Construção de Aplicações com o Uso de Ferramentas OLAP.** Universidade Federal do Rio de Janeiro. Projeto Final do Curso de Graduação de Bacharelado em Informática, 2000.
- [Machado 2006] Machado, F. N. R. **Tecnologia e Projeto de Data Warehouse: uma visão multidimensional.** 2ª edição. São Paulo. Ed. Érica, 2006.
- [Manzano 2007] Manzano, J. A. N. G.; **Oracle Database 10g: Express Edition – Interativo: guia básico de orientação e desenvolvimento.** 1ª edição. São Paulo. Ed, Érica, 2007.
- [Martin 1986] Martin, J.; **4GL: Fourth-Generation Languages.** Vol. III. USA. James Martin and The ARBEN Group, 1986.
- [Molina 2006] Molina et al. **A New Fuzzy Multidimensional Model.** IEEE Transactions on Fuzzy System, vol. 14, nº 6, Dezembro 2006.
- [Neter & Wasserman 1996] Neter, J.; kutner, M.; Nachtsheim, C.; Wasserman W. **Applied Linear Statistical Models.** Local: WB McGraw - Hill, 1996.
- [Oliveira 1998] Oliveira, D. de P. R. **Sistemas de informações gerenciais: estratégias, táticas, operacionais.** 5. ed.- São Paulo: Atlas, 1998.
- [Oliveira 2002] Oliveira, W, J. **Data Warehouse.** 2. ed. Florianópolis: Visual Books, 2002.
- [Passos & Goldschmidt 2005] Passos, E.; Goldschmidt, R. **Data mining: um guia prático.** Elsevier, Rio de Janeiro, 2005.

- [Pressman 1992]** Pressman, R. S. **Software Engineering: A Practitioner's Approach**. 3ª ed. USA: McGraw-Hill, 1992.
- [Rezende 2005]** Rezende, Solange Oliveira et al. **Mineração de dados**. In: Rezende, Solange Oliveira (coord). **Sistemas Inteligentes: fundamentos e aplicações**. São Paulo: Manole, 2005. Cap. 12, p. 307-335.
- [Sampaio 2006]** Sampaio M. C. ET AL, **Towards a Logical Multidimensional Model for Spatial Data Warehousing and OLAP**. DOLAP'06, November 10, 2006.
- [Sparre & Venema 1998]** SPARRE, P. & VENEMA, S. C. **Introduction to tropical fish stock assessment**. FAO Fish. Tech. Pap. 361/1. 1998. 376 p.
- [Thomsem 2002]** Thomsem, E. **Construindo Sistemas de Informações Multidimensionais**. 2ª ed. São Paulo: Campus, 2002.

Anexos

Dicionário de Dados

1. PONTO DE COLETA

Cadastro Doc – Cartográfica

Coluna	Tipo de dado	Descrição
Id Documento Cartográfica	NUMBER(11, 0)	Identificador da tabela de documentação cartográfica
Marco Piatam	NUMBER(11, 0)	Identificador de tabela de marco Piatam mar (escolhe-se entre os marcos preestabelecidos)
IdPontoColeta	NUMBER(11, 0)	Identificador da tabela de ponto de coleta

Documentação

Coluna	Tipo de dado	Descrição
Id Documentação	NUMBER(11, 0)	Identificador da tabela de documentação
Id Ponto de Coleta	NUMBER(11, 0)	Identificador da tabela de ponto de coleta
Id Marco Piatam Mar	NUMBER(11, 0)	Identificador da tabela de marco Piatam mar
Tipo Doc Produzida	VARCHAR2(50)	Tipo de documentação produzida
Qtde Doc Produzida	VARCHAR2(50)	Quantidade de documentação produzida

Cadastro Excursão

Coluna	Tipo de dado	Descrição
Id Excursão	NUMBER(11, 0)	Identificador da tabela de excursão
Id Período Hidrológico	NUMBER(11, 0)	Identificador da tabela de período
Data Início Excursão	DATE	Data do início da excursão (dia, mês, ano).
Data fim Excursão	DATE	Data final da excursão (dia, mês, ano).

Cadastro Marco - Piatam – Mar

Coluna	Tipo de dado	Descrição
Id Marco Piatam Mar	NUMBER(11, 0)	Identificador do Marco Piatam mar.
Nome Comunidade	VARCHAR2(100)	Designação do local onde será coletada a informação
Comunidade Latitude	NUMBER(11, 0)	São as coordenadas da localidade
Comunidade Longitude	NUMBER(11, 0)	São as coordenadas da localidade
Comunidade Latitude U T M	NUMBER(11, 0)	São as coordenadas da localidade
Comunidade Longitude U T M	NUMBER(11, 0)	São as coordenadas da localidade
Localidade	VARCHAR2(100)	Nome da localidade
Município	VARCHAR2(100)	Nome do município a qual pertence a localidade
U F	VARCHAR2(75)	Escolhe-se dentre os Estados onde o projeto atua

Cadastro Período – Hidrológico

Coluna	Tipo de dado	Descrição
Id Período Hidrológico	NUMBER(11, 0)	Identificador da tabela de período hidrológico
Tipo Período Hidrológico	VARCHAR2(10)	Tipo de período hidrológico (cheia, seca, enchente ou vazante)

Ponto de coleta.

Coluna	Tipo de dado	Descrição
Ponto de coleta	NUMBER(11, 0)	Identificador da tabela de ponto de coleta
Local coleta	VARCHAR2(100)	Neste campo é colocado o nome do local da coleta
Marco PIATAM mar	NUMBER(11, 0)	Identificador da tabela de marco Piatam mar (mostra em qual encontra-se inserido este dado)
Tipo Projeção Cartográfica	VARCHAR2(100)	Tipo de Projeção Cartográfica
Latitude Grau	NUMBER(2, 0)	Latitude Grau
Latitude minuto	NUMBER(2, 0)	Latitude minuto
Latitude Segundo.	NUMBER(2, 5)	Latitude Segundo.
Longitude Grau	NUMBER(2, 0)	Longitude Grau
Longitude minuto	NUMBER(2, 0)	Longitude minuto
Longitude Segundo.	NUMBER(2, 5)	Longitude Segundo.
Latitude Hemisfério	VARCHAR2(5)	Latitude Hemisfério
Longitude Hemisfério	VARCHAR2(5)	Longitude Hemisfério

Cadastro Sazonalidade Anual.

Coluna	Tipo de dado	Descrição
Id sazonalidade Anual	NUMBER(11, 0)	Identificador da tabela de sazonalidade anual
Ano	NUMBER(11, 0)	Ano da da coleta dos dados
Janeiro	CHAR(3)	Ocorrência no mês de janeiro
Fevereiro	CHAR(3)	Ocorrência no mês de fevereiro
Março	CHAR(3)	Ocorrência no mês de março
Abril	CHAR(3)	Ocorrência no mês de abril
Mai	CHAR(3)	Ocorrência no mês de maio
Junho	CHAR(3)	Ocorrência no mês de junho
Julho	CHAR(3)	Ocorrência no mês de julho
Agosto	CHAR(3)	Ocorrência no mês de agosto
Setembro	CHAR(3)	Ocorrência no mês de setembro
Outubro	CHAR(3)	Ocorrência no mês de outubro
Novembro	CHAR(3)	Ocorrência no mês de novembro
Dezembro	CHAR(3)	Ocorrência no mês de dezembro

2. ABIÓTICO

Cadastro Abiótico – Climatologia.

Coluna	Tipo de dado	Descrição
Id climatologia	VARCHAR(255)	Identificador da tabela de aves
Precipitação Mínima	NUMBER(10, 5)	precipitação mínima medida em mm/tempo
Precipitação Média	NUMBER(10, 5)	precipitação média medida em mm/tempo
Precipitação Máxima	NUMBER(10, 5)	precipitação máxima medida em mm/tempo
Temperatura Ar Mínima	NUMBER(10, 5)	Temperatura mínima do ar medida em oC
Temperatura Ar Média	NUMBER(10, 5)	Temperatura média do ar média em oC
Temperatura Ar Máxima	NUMBER(10, 5)	Temperatura máxima do ar medida em oC
Insolação	NUMBER(10, 5)	Tempo de exposição à radiação solar medido em hora
Direção Vento	NUMBER(10, 5)	Direção do vento medida em graus
Intensidade Vento Mínima	NUMBER(10, 5)	Intensidade mínima do vento medida em metro por segundo
Intensidade Vento Média	NUMBER(10, 5)	Intensidade máxima do vento média em metro por segundo
Intensidade Vento Máxima	NUMBER(10, 5)	Intensidade máxima do vento medida em metro por segundo
Pressão Atmosférica Mínima	NUMBER(10, 5)	Peso mínimo da coluna de ar de seção reta unitária sobre um ponto na superfície.
Pressão Atmosférica Média	NUMBER(10, 5)	Peso médio da coluna de ar de seção reta unitária sobre um ponto na superfície.
Pressão Atmosférica Máxima	NUMBER(10, 5)	Peso máximo da coluna de ar de seção reta unitária sobre um ponto na superfície.
Umidade Relativa Ar Mínima	NUMBER(10, 5)	Quantidade mínima de vapor d'água na atmosfera em percentuais.
Umidade Relativa Ar Média	NUMBER(10, 5)	Quantidade média de vapor d'água na atmosfera em percentuais.
Umidade Relativa Ar Máxima	NUMBER(10, 5)	Quantidade máxima de vapor d'água na atmosfera em percentuais.
Umidade Solo Mínima	NUMBER(10, 5)	Umidade mínima do solo medida em percentual
Umidade Solo Média	NUMBER(10, 5)	Umidade média do solo medida em percentual
Umidade Solo Máxima	NUMBER(10, 5)	Umidade máxima do solo medida em percentual
Saldo Rad	NUMBER(10, 5)	Diferença entre a energia solar recebida na superfície e a perda para a atmosfera na forma de ondas longas, expresso em Watt/m ² .
Nebulosidade	NUMBER(10, 5)	Nebulosidade medida em oitavos.
Sazonalidade Diária (Qtd Dias)	NUMBER(10, 5)	Sazonalidade Diária (Qtd Dias)

Coluna	Tipo de dado	Descrição
Sazonalidade Diária (Qtd Mês)	NUMBER(10, 5)	Sazonalidade Diária (Qtd Mês)
Sazonalidade Diária (Qtd Ano)	NUMBER(10, 5)	Sazonalidade Diária (Qtd Ano)

Cadastro Abiótico – Executor

Coluna	Tipo de Dado	Descrição
id Executor	NUMBER(11, 0)	Identificador da tabela de documentação
Nome	VARCHAR(100)	Identificador da tabela de ponto de coleta
Sigla	VARCHAR(50)	Identificador da tabela de marco Piatam mar

Cadastro Abiótico - Geofísica.

Coluna	Tipo de Dado	Descrição
Id Geofísica	NUMBER(11, 0)	Identificador da tabela de Geofísica
Executor	NUMBER(11, 0)	Este campo já vem preenchido com os nomes que foram inseridos no campo nome do Cadastro Abiotico – Executor.
Batimetria	VARCHAR(50)	Levantamento topográfico do fundo de corpo aquoso, cuja profundidade é obtida pela medição do tempo de ida e volta de um impulso sonoro emitido (eco).
Sísmica	VARCHAR(50)	Fornecer dados sobre a disposição estrutural das camadas sedimentares abaixo do fundo marinho. Através de um registro sísmico é possível avaliar parâmetros como espessura de camadas, mergulho, presença de falhamentos, ocorrências de acumulações rasas de gás biogênico e deslizamentos submarinos.
Magnetometria	VARCHAR(50)	É o método de prospecção geofísica, cuja a finalidade consiste em pesquisar as estruturas geológicas subsuperficiais utilizando-se o magnetômetro
Gravimetria	VARCHAR(50)	Medida do campo gravitacional terrestre utilizando-se gravímetros. No mar, os gravímetros especialmente instalados em navios ou submarinhos (Sugiyo, 1998). A unidade de medida utilizada é o mgal (1/1000gal).
Sonografia	VARCHAR(50)	Método de mapeamento da topografia do fundo ao longo de uma faixa de largura variável através de sinais que são direcionados lateralmente para ambos os lados, ou seja, trabalham com dois feixes de sinais acústicos.

Cadastro Abiotico – Geologia

Coluna	Tipo de Dado	Descrição
Id Geologia	NUMBER(11, 0)	Identificador da tabela de Geologia.
Tipo Sistema	VARCHAR(50)	Tipo de sistema geológico
Formação Geológica	VARCHAR(50)	Unidade litogenética fundamental da classificação local das rochas. A sua individualização é geralmente determinada por modificações litológicas, quebras nas continuidades de sedimentação ou outras evidências de importância
Litologia	VARCHAR(50)	Caráter físico de uma rocha, em geral determinado macroscopicamente ou com ajuda de lupa de bolso.
Métodos Data	VARCHAR(50)	Determinação de idade de amostras geológicas, em números de anos, por um dos vários métodos baseados na velocidade de desintegração de elementos químicos radioativos contidos nesses materiais.
Pb 2 1 0	VARCHAR(50)	Método de determinação de idade baseado na atividade no chumbo 210(meia-vida = 22,2 anos) comparando-se a atividade medida com a de amostras atuais.
C 1 4	VARCHAR(50)	Determina o tempo absoluto em função da meia-vida relativamente curta de carbono 14. Este se forma a partir do nitrogênio do ar pela reação nuclear provocada pelos raios cósmicos, na estratosfera.
Cronologia Absoluta	NUMBER(11, 0)	Consiste na determinação temporal de eventos geológicos em números de anos.
Textura	NUMBER(11, 0)	Aspecto menor inerente à rocha, que depende do tamanho, da forma, do arranjo e da distribuição dos seus componentes.
Grau Esfericidade	VARCHAR(50)	Grau de aproximação de um fragmento de uma esfera perfeita.
Classes Granulométricas	VARCHAR(50)	Especificação da dimensão dos diâmetros dos materiais detríticos, porém, segundo Jacques Boucarte, os fatores de ordem química são levados em consideração para caracterizar o material (Suguio, 1972).
Matacão	VARCHAR(50)	Termo aplicado a fragmentos rochosos arredondados jazendo sobre a superfície do terreno ou no interior do solo solto de composição diferente das rochas circunjacentes e de dimensões maiores do que 256mm de diâmetro.

Calhau	VARCHAR(50)	Fragmento mineral ou lítico, de origem detrítica com diâmetro entre 64 e 256 mm, sendo portanto maior que o seixo e menor do que o matacão, arredondado por desgaste durante os processos de transporte aquoso, eólico ou glacial.
Seixo	VARCHAR(50)	Designação geral para fragmentos minerais ou líticos arredondados como, por exemplo, de quartzo a quartzito com diâmetros entre 4 e 64mm de diâmetros.
Grânulos	VARCHAR(50)	Segundo a escala granulométrica de Wentworth (1922), corresponde ao elemento clástico cujo o diâmetro situa-se entre 2 e 4 mm. Uma rocha sedimentar clástica formada por agregado com predominância de grânulos deve se chamado de conglomerado fino ou microconglomerado
Areia Muito Grossa	VARCHAR(50)	Granularidade da areia
Areia Grossa	VARCHAR(50)	Granularidade da areia
Areia Fina	VARCHAR(50)	Granularidade da areia
Areia Muito Fina	VARCHAR(50)	Granularidade da areia
Silte	VARCHAR(50)	1. Sedimento clástico composto de partículas entre 1/16 a 1/256mm de diâmetro. 2. solo formado por 80% ou mais de silte(0,05 – 0,002m) e menos de 12% de argila(menor que 0,002mm).
Argila - Silte	VARCHAR(50)	Argila - Silte
Grau de Arredondamento - Silte	VARCHAR(50)	Grau de Arredondamento - Silte

Cadastro Abiótico - Geomorfologia - Ambiente – Estuarino.

Coluna	Tipo de Dado	Descrição
Id Amb Est	NUMBER(11, 0)	Identificador do Marco Piatam mar.
Dep Canal	VARCHAR(100)	Designação do local onde será coletada a informação
Tipo A	CHAR(3)	Classificação quanto ao padrão de circulação de estuários. Ocorrem em costas submetidas a forte influencia fluvial e amplitudes de maré inferiores a 2 m. A estratificação das massas de água é bem definida
Tipo B	CHAR(3)	Classificação quanto ao padrão de circulação de estuários. Ocorrem em costas submetidas ao domínio de meso-marés (amplitude entre 2 e 4 m) e apresentam mistura parcial entre as águas por efeito de turbulência.

Tipo C	CHAR(3)	Classificação quanto ao padrão de circulação de estuários. Dominados por macro-marés e verticalmente homogêneos. Uma corrente em direção ao mar aberto é formada, enquanto que próximo ao fundo uma corrente de água salina está presente e dirige-se para o interior do estuário.
Tipo D	CHAR(3)	Classificação quanto ao padrão de circulação de estuários. Homogeneidade vertical e lateral, ou seja, não há um gradiente vertical de salinidade.
Estuário Positivo	CHAR(3)	Estuário onde se verifica a diluição mensurável da salinidade da água do mar, em função do excesso de drenagem continental e/ou precipitação.
Estuário Negativo	CHAR(3)	Estuário onde a evaporação excede o influxo de água doce (continental + precipitação), de modo que a salinidade do estuário é maior que a do mar.
Funil Estuarino	CHAR(3)	Escolhe-se dentre os Estados onde o projeto atua
Segto Meandrante	CHAR(3)	Padrão de canal fluvial característico de rios maduros de baixo gradiente, com ampla planície de inundação, por onde divaga o rio com trajetória mais ou menos sinuosa.
Areia	CHAR(3)	Fração granulométrica de sedimentos detriticos ou clásticos compreendida entre 0,062 (1/16) e 2 mm.
Argila	CHAR(3)	Fração granulométrica de sedimentos cujos componentes predominantes são os argilominerais, todos com tamanhos menores do que 0,004mm de diâmetro.
Sistema Vales Afogados	CHAR(3)	Baía formada por submersão de um vale costeiro por subsidência do continente ou elevação do nível marinho. A sedimentação no interior do vale pode progredir rapidamente por efeito combinado da baixa energia e aporte de sedimentos terrígenos por um rio.
Ilhas Barreira	CHAR(3)	Ilha arenosa, paralela ao litoral, separada do continente por uma laguna. É construída pela ação da deriva litorânea dos sedimentos, mas pode estar associada as variações relativas do nível do mar.

Laguna Costeira	CHAR(3)	Corpo de águas rasas e calmas situado em planícies costeiras, em geral mantendo comunicação restrita com o mar. Forma, com freqüência, sistema barreira/laguna, geneticamente ligado a dinâmica costeira.
Estuário Tectônico-São	CHAR(3)	considerados estuários restantes, pois se formam com os eventos geológicos e geomorfológicos, tais como, falhas tectônicas, erupção vulcânicas tremores e deslizamento de terra.
Silte	CHAR(3)	Fração granulométrica de sedimentos clásticos composto por partículas entre 1/16 a 1/256 mm de diâmetro.
Barras Est Longitudina	CHAR(3)	Barra de sedimentos arenosos formada após o sulco longitudinal e disposta aproximadamente paralela a linha de costa. Pode ficar exposta na maré baixa, ou ocorrer abaixo do nível das águas na zona de costa afóra.
Planície Maré Arenosa	CHAR(3)	São as coordenadas da localidade.
Planície Maré Lamosa	CHAR(3)	São formados no lado do oceano aberto e no interior de de uma laguna. Os sedimentos são transportados para dentro na maré enchente e para fora na maré vazante.
Delta Maré	VARCHAR(50)	Deltas formados pelo acúmulo de sedimentos durante as marés enchente e vazante, respectivamente.
Delta Maré Enchente	CHAR(3)	Escolhe-se dentre os Estados onde o projeto atua.
Delta Maré Vazante	CHAR(3)	Escolhe-se dentre os Estados onde o projeto atua.
Pantano Salino	CHAR(3)	Zona pantanosa periodicamente inundada por águas salobras, em geral vegetada por arbustos e gramíneas.
Dominado Mares	VARCHAR(50)	Recebem sedimentos dos rios na cabeceira do estuário, e, da plataforma continental adjacente.
Altura Maré	NUMBER(11, 0)	Nome do município a qual pertence a localidade.
Estuário Micromare	VARCHAR(50)	Os estuários de tipo cunha salina são típicos de regiões de micromaré e lugares em que predominam condições de grandes descargas fluvial. Portanto, são estuários dominados pela descarga fluvial e pelo processo de entranhamento, que é responsável pelo aumento de salinidade da camada superficial, e a mistura por difusão turbulenta é desprezível.
Estuário Macromare	VARCHAR(50)	Existência ou não do Estuário.
Estuário Meso Maré	VARCHAR(50)	Existência ou não do Estuário.

Dominado Ondas	VARCHAR(50)	Quando os efeitos de turbulência causados pelas ondas no interior do estuário são dominantes, o estuário é classificado como verticalmente bem misturado (tipos C e D). Esse tipo de estuário forma-se em geral em canais rasos e estreitos forçados por descargas fluviais pequenas.
----------------	-------------	---

Cadastro Abiotico - Geomorfologia - Ambiente – Fluvial

Coluna	Tipo de Dado	Descrição
Id Amb Fluvial	NUMBER(11, 0)	Identificador da tabela de geomorfologia ambiente fluvial
Areia	CHAR(3)	Fração granulométrica de sedimentos detríticos ou clásticos compreendida entre 0,062 (1/16) e 2 mm.
Argila	CHAR(3)	Fração granulométrica de sedimentos cujos componentes predominantes são os argilominerais, todos com tamanhos menores do que 0,004mm de diâmetro.
Silte	CHAR(3)	Fração granulométrica de sedimentos clásticos composto por partículas entre 1/16 a 1/256 mm de diâmetro.
Mista	CHAR(3)	Existência mista dos sedimentos
Campos Perm Inundados	CHAR(3)	Permanência ou não da inundação dos campos (sim ou não).
Campos Periódicamente Inundados	CHAR(3)	Os campos não ou não inundados periodicamente.
Canais Retilíneos	CHAR(3)	Em geral eles exibem uma sinuosidade desprezível devida ao desenvolvimento de barras laterais. Os trechos mais retilíneos são muito limitados, e sugerem que essas partes jamais ultrapassem 10 vezes a largura do canal.
Canais Meandrantés	CHAR(3)	Quando a sinuosidade do canal for superior a 1,5 pode ser denominado de meandrante. Esta é uma propriedade inerente ao rio meandrante, que reflete a intensidade de meandramento do canal. A sinuosidade do canal aumenta, em geral, montante para a jusante, em consonância com a diminuição da declividade e aumento da frequência de sedimentos pelíticos na carga sedimentar.

Canais Braided	CHAR(3)	Padrão de canal fluvial caracterizado por uma grande subdivisão em canais menores separados entre si por barras longitudinais de canais. São normalmente compostos por areia, cascalho e suas misturas. Esse padrão é característico de rios ou trechos de rios em que o volume de sedimentos carreados é maior que a capacidade de transporte do rio.
Canais Anastomosados	CHAR(3)	Canais fluviais caracterizados por vários canais curvilíneos menores fluindo à baixa velocidade, ao redor de ilhas permanentes e cobertas por vegetação.
Barras Longitudinais	CHAR(3)	Barra de sedimentos arenosos formada após o sulco longitudinal e disposta aproximadamente paralela a linha de costa.
Barras Transversais	CHAR(3)	Barra arenosa ou cascalhosa total ou parcialmente submersa pelas águas conforme a estação do ano, acumuladas e orientadas em relação às correntes fluviais em diferentes partes do canal fluvial.
Barras Pontal	CHAR(3)	Crista, em geral arenosa, longa e curva formando séries paralelas, formadas próximo a margem interna (convexa) de canais meandrantés.
Depósitos Planície Inundação	CHAR(3)	Depósito de sedimentos finos acumulados durante as enchentes, por transbordamento do canal.
Depósitos Preenchimento Canal	CHAR(3)	Depósito sedimentar predominantemente grosseiro acumulado em um canal fluvial, onde a capacidade da corrente aquosa foi insuficiente para remover a areia e outros detritos a medida em que eram supridos pela área fonte.
Depósitos Pântanos	CHAR(3)	Depósitos Pântanos de água doce ou salobra, comuns em regiões rebaixadas e irregulares da zona litorânea ou margens de rios e lagos. Compostos predominantemente de lamas ricas em MO, contém óxidos e carbonatos de ferro, localmente areia e marga, passando lateralmente para depósitos marinhos ou lacustres.

Depósitos Dique Marginal	CHAR(3)	São normalmente sem estratificação, é composto pela fração mais grosseira dos sedimentos depositados pelo transbordamento, quando a água perde competência. É uma elevação natural de seção transversal assimétrica, mais abrupta na face voltada para o canal e mais suave para a planície de inundação que margeia cursos fluviais, confinando-os aos seus canais.
Depósitos Rompimento Canal	CHAR(3)	Depósito de sedimentos mais grosseiros que os da planície de inundação e mais finos do que os do canal fluvial. Tem a forma de um leque aluvial, é originado pelo fluxo divergente de água que extravasa através de uma ruptura no dique marginal de um rio durante a enchente.
Depósitos Res Canal	CHAR(3)	Depósito de material grosseiro localizado na porção basal do depósito de preenchimento de canal, que foi deixado para trás pela eliminação dos sedimentos finos para jusante, no processo de transporte fluvial.
Igapós	CHAR(3)	As matas banhadas pelas águas pretas e claras costumam ser chamadas de florestas de igapós.
Varzeas	CHAR(3)	As matas banhadas pelas águas brancas costumam ser chamadas de florestas de várzea. A vegetação da várzea é muito mais rica do que a vegetação dos igapós, por causa da fertilidade das águas brancas e dos solos aluvionais por elas trazidos.
Predomina Erosão	CHAR(3)	Onde os processos de remoção física e de transporte são ativos e predominantes.
Predomina Deposição	CHAR(3)	Onde os agentes de transporte são menos eficazes, permitindo a deposição de materiais sedimentares.

Cadastro Abiótico - Geomorfologia - Ambiente – Lagos

Coluna	Tipo de Dado	Descrição
Id Amb Lago	NUMBER(11, 0)	Identificador da tabela de Lagos Ambiente Sedimentar Lagos
Acidotrofico	VARCHAR(50)	Lago cuja acidez pode provir da putrefação de matéria orgânica vegetal.
Alcalitrofico	VARCHAR(50)	Lagos com pH acima de 9.
Distrofico	VARCHAR(50)	Caracterizado por águas rasas com alto teor de MO, baixa disponibilidade de nutrientes e altas DBO. É um estágio intermediário entre o lago eutrófico e um pântano.
Eutrofico	VARCHAR(50)	Lago raso, com alta produtividade primária com abundante vegetação litorânea e densa população planctônica. Tende a empobrecer em Oxigênio durante o verão.
Oligotrofico	VARCHAR(50)	Lago com considerável de OD material nutriente limitado (Baixa produtividade primária).
Salgado	VARCHAR(50)	Lago com salinidade superior a 500 mg/L.
Lago Afundamento	CHAR(3)	Corpo de águas estacionárias alojados em uma depressão fechada de região cárstica.
Lago Antropico	CHAR(3)	Lagos que resultam das alterações impostas à topografia e à vegetação locais, tais como charcos, em áreas de baixada e lagoas, em pontos onde os cursos d' água foram interrompidos.
Lago Barragem	CHAR(3)	Lago originado pela barragem de um vale, por razões variadas (variações do nível do mar, corridas de lama, tectônica, climáticas).
Lago Cratera	CHAR(3)	Lago alojado no interior de uma cratera, em geral formado por atividade vulcânica/ impactos meteoritos.
Lago Crescente	CHAR(3)	Lago em forma de meia-lua, situado em posição bem definida na faixa de meandro, formado pelo abandono de trechos do meandro durante a migração lateral do canal fluvial.
Lago Deflação	CHAR(3)	Lagos que ocupam depressões formadas pela depressão eólica, encontrados principalmente em regiões de clima áridos e semi-áridos, formam corpos de águas muito rasos que se ressecam em estiagem prolongadas.
Lago Deltáico	CHAR(3)	Lago formado ao longo da margem ou no interior de um delta, embaixamentos rasos ou pelo aprisionamento de parte do mar pelo crescimento de depósitos deltáicos.

Lago Estrutural	CHAR(3)	Originado pelo acumulo de água em depressão formada por movimentos tectônicos. Na maioria das vezes a margem é aproximadamente retilínea e a estrutura é bastante simples, podendo exibir grandes profundidades.
Lago Reliquiar	CHAR(3)	Lagos formados em regiões costeiras submetidas a transgressões e regressões.
Lago Suspenso	CHAR(3)	Lago permanente cujo nível das águas situa-se bem acima dos outros corpos d'água incluindo os aquíferos associados ao lago.
Tipo Substrato	VARCHAR(50)	Identificar o tipo de substrato existente. Os dados inseridos são caracteres do tipo varcha contendo ate 100 algarismos.
Efêmero	VARCHAR(50)	Lago que contem água apenas em uma parte do ano (ou que se torna seco sazonalmente).
Holomitico	VARCHAR(50)	Lago caracterizado pela circulação total de suas águas durante o inverno, devido aos processos de convecção.
Meromitico	VARCHAR(50)	Lago onde a água circula apenas uma vez.
Oligomitico	VARCHAR(50)	Lago onde a água circula poucas vezes, porem sem regimes fixos.
Temperado	VARCHAR(50)	Lago onde a T da água superficial, deve ao menos uma vez ao ano ser inferior a 4°C.
Tropical	VARCHAR(50)	Lago com estratificação térmica normal, isto é com temperatura sempre superior a 4°C.
Nome Depósito	VARCHAR2(100)	Nome do depósito

Cadastro Abiotico - Geomorfologia - Ambiente - Linha – Costa

Coluna	Tipo de Dado	Descrição
Id Amb Linha Costa	VARCHAR2(50)	Identificador da tabela de geomorfologia Ambiente linha de costa
Substr Rochoso	CHAR(3)	Rocha eólica em geral ígnea ou metamórfica, exposta a superfície terrestre ou recoberta por materiais inconsolidados.
Substr Arenoso	CHAR(3)	Presença ou ausência de Substrato arenoso
Substr Lamoso	CHAR(3)	Presença ou ausência de Substrato lamoso
Costa Erosiva	CHAR(3)	A costa erosiva é caracterizada pela presença das seguintes feições: falésias, pancadas de barasão, chaminés marinha e arcos em cavernas, que apresenta testemunhos residuais de antigas falésias e bancadas de abrasão.
Presença Falésia	CHAR(3)	Presença ou ausência de falésia
Falesia Ativa	CHAR(3)	É a falésia que está sendo atacada pela ação das ondas, isto é está em formação.
Falesia Inativa	CHAR(3)	Ocorre mais ou menos afastada da linha costeira atual, pode ter se formado pelo progradação ou descida do nível do mar.
Presença Plataforma Abrasão	CHAR(3)	Plataformas geradas pela abrasão marinha.
Costa Acrecionaria	CHAR(3)	Constitui uma costa em avanço, isto é, linha costeira que avança mar adentro e que segundo Valentin (1952), pode resultar de processos orgânicos e inorgânicos.

Cadastro Abiotico - Geomorfologia - Ambiente – Plataforma

Coluna	Tipo de Dado	Descrição
Id Amb Plataformal	NUMBER(11, 0)	Identificador da tabela Ambiente Plataformal Geomorfologia
Plataforma Externa	NUMBER(10, 5)	É a porção externa da plataforma continental situada abaixo da base das ondas inicia se a 30m de profundidade podendo chegar a 200m, exibindo em geral fundo lamacento.
Plataforma Interna	NUMBER(10, 5)	Porção interna da plataforma continental se inicia no nível de maré baixa, estende-se até 30m de profundidade.

Largura	VARCHAR2(50)	A largura média é de ordem de 75km, podendo atingir poucos quilômetros, como no caso da plataforma continental, em frente da cidade de Salvador (9 km), ou mais de 350km na região da Antártida
Recifes Corais	CHAR(3)	Recife orgânico calcário, composto de coral sólido e areia. As algas podem ser responsáveis por mais da metade do CaCO ₃ recifal. São encontrados em águas onde a temperatura média mensal é maior ou igual a 18°C.
Calc Bioclast Biodetríticos	CHAR(3)	Designação utilizada por Larsouner, são sedimentos marinhos recentes com 50% a 70% de calcário (CaCO ₃) biogênico.
Espessura Depósitos	CHAR(3)	Espessura Depósitos
Plataf Autóctone	CHAR(3)	Plataf Autóctone
Plataf Alóctone	CHAR(3)	Plataf Alóctone
Sedim Terrigenos	CHAR(3)	Sedimentos formados por fragmentos minerais derivados do intemperismo, erosão e deposição de rochas ígneas, sedimentares ou metamórficas preexistentes.
Sedim Biogênicos	CHAR(3)	Sedimentos contendo material produzido por plantas ou animais, tais como recifes de coral, fragmentos de conchas, carapaças de diatomáceas, radiolários, foraminíferos e coccolitoforídeos.
Sedim Autigenicos	CHAR(3)	Podem ser exemplificados pelos nódulos polimetálicos e phillipsita. Os nódulos polimetálicos também conhecidos por nódulos de manganês são matérias de precipitação química singenética de cores preta a acinzentada, compostos principalmente de Fe ⁺² e Mn ⁺² mal cristalizado. Eles apresentam núcleos de fragmentos de conchas, corais, dentes de tubarão, ossos e rochas. A phillipsita, por outro lado é um silicato hidratado de Ca, Na, K e Al, que se cristaliza como mineral esbranquiçado de brilho vítreo de hábito acicular.
Sedim Reliquiares	CHAR(3)	Caracterizado por ter sido formado em ambiente diferente em que se encontra atualmente. Não está em equilíbrio com as condições atuais.
Sedim Palimpsest	CHAR(3)	Apresentam feições representativas de diferentes ciclos de sedimentação, p. ex. fluvial e marinho raso.

Plataf Dom Tempestades	CHAR(3)	Estão associadas as correntes meteorológicas onde há intensa atividade dos ventos e das ondas ocasionando o desenvolvimento de grandes correntes de fundo.
Plataf Dom Mares	CHAR(3)	O sistema regular de correntes associado aos ciclos de mares em plataformas adjacentes e regiões costeiras submeditas a regimes de macromarés, estas fortes e regulares correntes disponibilizam sedimentos, podendo assim criar e fazer migrar formas de leito de grande porte junto ao fundo, exemplo: sandwaves.

Cadastro Abiotico - Geomorfologia - Ambiente – Praia

Coluna	Tipo de Dado	Descrição
Id Amb Praia	NUMBER(11, 0)	Identificador da tabela de geomorfologia ambiente de praia
Arrebe Progres Deslizante	CHAR(3)	Modo de arrebentação de onda que se rompem gradualmente percorrendo uma distancia mais ou menos longa, em virtude do fundo plano e declividade suave.
Arrebe Mergulhante	CHAR(3)	Modo de arrebentação de ondas cujas cristas se rompem após enrolar-se em espiral, com estampido, quando o fundo apresenta declividade intermediaria. Sinônimo: arrebentação mergulhante.
Arrebe Ascendente	CHAR(3)	Modo de arrebentação de onda, que se rompem em vagalhões, em virtude do fundo de forte declividade, isto é, a onda se forma muito rapidamente e se quebra diretamente sobre a praia. Sinônimo: Vagalhão.
Presença Ante Praia	CHAR(3)	Parte da praia situada entre o limite superior da preamar e a linha de baixamar.
Presença Face Praia	CHAR(3)	Setor da praia normalmente exposto a ação do espraiamento das ondas.
Areia	CHAR(3)	Fração granulométrica de sedimentos detríticos ou clásticos compreendida entre 0,062 (1/16) e 2 mm.
Argila	CHAR(3)	Fração granulométrica de sedimentos cujos componentes predominantes são os argilominerais, todos com tamanhos menores do que 0,004mm de diâmetro.
Silte	CHAR(3)	Fração granulométrica de sedimentos clásticos composto por partículas entre 1/16 a 1/256 mm de diâmetro.

Presença Pos Praia	CHAR(3)	Porção superior da praia, além do alcance das ondas e marés ordinárias, ou que se estende desde a crista praiial, construída pelo nível de preamar de sizígia até o sopé da escarpa praiial.
Presença Dunas Fixas	CHAR(3)	Duna protegida da ação do vento por uma cobertura vegetal ou por cimentação da areia.
Presença Dunas Moveis	CHAR(3)	Duna ativa que sofre deslocamento como unidade independente, sob incessante ação eólica.
Presença Cord Litorâneo	CHAR(3)	Feição deposicional alongada, em geral de composição arenosa ou, por vezes, cascalhosa ou conchífera, disposta paralelamente a paleolinhas praiiais e separada uma das outras por depressões.
Presença Ilha Barreira	CHAR(3)	Ilha essencialmente arenosa que se estende paralelamente ao litoral. Separada do continente por uma laguna. Ela é tipicamente construída pela ação da deriva litorânea de sedimentos, mas pode ser parcialmente relacionada a mudanças no nível do mar.
Presença Chenier	CHAR(3)	Cordões de sedimentos arenosos, cascalhosos ou conchíferos, isolados ou unidos entre si, dispostos paralelamente ao litoral sobre sedimentos argilosos de pântanos costeiros. Representam fases estacionárias ou retrogradantes em costas regressivas e formam-se durante as grandes tempestades.
Presença Zona Arrebentação	CHAR(3)	Faixa em geral estreita, onde as ondas se rompem pela diminuição da profundidade das águas até abaixo de um determinado valor igual ou inferior à base das ondas. A movimentação dos sedimentos, em ambas as direções, ocorre no sentido da zona de arrebentação.
Presença Zona Surf	CHAR(3)	Área situada entre o limite externo de arrebentação e o limite de espraiamento das ondas.

Presença Zona Espraimento	CHAR(3)	Zona em que ocorre o espalhamento da água do mar, após a arrebentação das ondas, seguindo acima sobre a superfície praial. Os sedimentos se movem obliquamente à praia e paralelamente às cristas das ondas durante o avanço das águas de espraimento e no retorno por gravidade, as areias são transportadas perpendicularmente à praia. Assim o padrão de deriva é zigzagueante.
Altura do Berma	CHAR(3)	Porção praticamente horizontal da praia ou pós-praia formada pela sedimentação por ação de ondas.
Largura do Estirancio	CHAR(3)	É a região entre marés, ou seja, entre nível da maré baixa e maré alta. É, portanto, a porção da praia que sofre normalmente a ação das marés e os efeitos do espraimento e o refluxo da água.
Altamente Dissipativa	CHAR(3)	Altamente Dissipativa
Dissipativa	CHAR(3)	Tipo de praia de baixa declividade (menos de 20) em que a energias das ondas, na face litorânea, é dissipada acentuadamente pelo atrito na zona de surfe, que apresenta grande largura.
Reflexiva	CHAR(3)	Tipo de praia de alta declividade (60 a 80) em que a energia das ondas é predominantemente refletida pelos leitos da antepraia e face litorânea, porque as ondas se quebram muito perto da praia.
Intermediaria	CHAR(3)	Praia com características morfológicas e dinâmicas intermediaria entre dissipativa e reflexiva.
Corrente Retorno	CHAR(3)	Forte corrente superficial que flui do litoral par o mar aberto, correspondendo ao movimento de retorno das águas acumuladas na zona costeiras pela chegada dos sucessivos trens de onda. O comprimento dessas correntes pode variar de 70 a 830 m e a velocidade pode atingir 2 nós.

Cadastro Abiótico - Geomorfologia - Ambiente – Recifes

Coluna	Tipo de Dado	Descrição
Id Amb Recife	NUMBER(11, 0)	Identificador da tabela de Ambiente recife de corais geomorfologia
Largura Recifes	NUMBER(12, 0)	Largura dos recifes
Comprimento Recifes	NUMBER(12, 0)	Comprimento dos recifes
Biogênico	CHAR(3)	Relacionado a um depósito sedimentar resultante da atividade fisiológica dos organismos.
Ooids	CHAR(3)	Componente individual(diâmetro igual ou inferior a 2mm) de uma rocha oolítica.
Intraclastos	CHAR(3)	fragmento calcário retrabalhados, muitas vezes arredondados, de deposição penecontemporânea, que foi erodido e redepositado nas cercanias incorporado a calcários de idade mais nova.
Pellets	CHAR(3)	Partícula composta por micrito , em geral, sem qualquer estrutura interna visível. A maioria é considerada de origem fecal.
Lama Carbonatica	CHAR(3)	Os tamanhos de silte e argila dos componentes de sedimentos carbonáticos são coletivamente designados como lama carbonática. A maioria da lama carbonática consiste de pequenos indicadores e plaquetas de cristais de carbonato que também são pequenos para a resolução com microscópio binocular ou petrográfico. Conforme, interpretações das partículas de lama carbonática foram baseadas na noção que elas partem de uma origem comum chamada precipitação inorgânica da água salina.
Recife Franja	CHAR(3)	Recife coralino paralelo e em contato direto com a linha costeira de um continente ou ilha, tendo a separá-los apenas um estreito canal. Forma uma barreira com declividade suave para o continente e brusca queda para o mar.
Recife Barreira	CHAR(3)	Recife coralino alongado, de disposição paralela a linha de costa de um continente ou ilha, do qual é separado por uma laguna com água demasiadamente profunda para o desenvolvimento de recifes.
Atol	CHAR(3)	Ilha ou conjunto de ilhas calcárias disposta em anel, encerrando uma laguna interna(30 – 100m de profundidade)de forma em geral circular.

Backreef	CHAR(3)	É a parte do ecossistema do recife de coral que se estende da costa à crista do recife.
Laguna	CHAR(3)	Corpo de água rasa e camas, situadas em planícies costeiras, em geral mantendo uma comunicação restrita com o mar.

Cadastro Abiótico - Geomorfologia - Ambiente – Talude

Coluna	Tipo de Dado	Descrição
Id Am bTalude	NUMBER(11, 0)	Identificador da tabela Talude Geomorfologia Ambiente Talude
Tipo Depósito	VARCHAR(50)	Depósitos acumulados na base de uma escarpa. O material pode ser trazido pela erosão do lençol de escoamento superficial, ou pelo efeito da gravidade. Esta última, a maior responsável pela formação de grandes depósitos de talude(material de creep, de desmonoronamento, de colúvio, etc.) (Suguio, 1972).
Hemipelagicos	VARCHAR(50)	Depósitos que ocupam uma posição intermediária entre os depósitos neríticos da plataforma continental e os depósitos pelágicos de fundos submarinos profundos (Suguio, 2003)
Sope Continental	VARCHAR(50)	Representa uma superfície submarina além da base do Talude Continental, em geral com gradientes menores que 1:1000, ocorrendo a profundidades entre 1375 e 5185 metros com largura entre 300 e 400 Km e...(Suguio, 2003).
Planície Abissal	VARCHAR(50)	Superfície plana, com baixa declividade que ocupa as porções mais profundas das bacias oceânicas (Suguio, 2003).
Pelágicos	VARCHAR(50)	
Turbiditos	VARCHAR(50)	São depósitos originários da ação das correntes de turbidez, e que se deslocam como uma avalanche de sedimentos através do talude continental depositando-o sob a forma de leques submarinos gerando a chamada sequência de Bouma.

3. BIÓTICO

Cadastro Biótico – Peixes Estuarinos.

Coluna	Tipo de dado	Descrição
IdPeixesEstuarinos:	INTEGER (11, 0)	Identificador da tabela de peixes estuarinos
Fonte	VARCHAR (255)	Fonte das informações (deve ser dada preferência a fontes bibliográficas)
Localizacao	VARCHAR (20)	Local de ocorrência dos táxons, expressa em coordenadas geográficas
ProtLegalCateg	VARCHAR (50)	Indica se a área em que a espécie foi registrada encontra-se dentro de uma área de proteção ambiental. Categorias de proteção legal segundo a Lei Federal 9985 de 18.07.2000
ProtLegalEsfera	VARCHAR (10)	Indica a esfera de administração da unidade de conservação definida no campo anterior. Segundo definição da Lei Federal 9985 de 18.07.2000
StatusConservacao	VARCHAR (10)	Categorias definidas na Lista nacional das espécies da fauna brasileira ameaçadas de extinção. Para preenchimento consultar www.mma.gov.br/port/sbf/fauna/grupos3.htm . Categorias: Planilha I anexa - Linha 7
StatusExploracao	VARCHAR (20)	Classificação do recurso segundo o estágio de exploração. Categorias: Inexplorado, Em desenvolvimento, Limite máximo de exploração, Sobre-explorado, Colapsado, Em recuperação.
ConcQuant	VARCHAR (10)	Representa a quantidade presente. A unidade de medida deverá ser inserida no campo ConcUnid
ConcUnid	VARCHAR (20)	Unidade na qual é expressa a concentração (ex: ind/m, g/cm, cel/ml, etc)
TipoDesova	VARCHAR (20)	Freqüência com que uma população como um todo, ou seus indivíduos isoladamente, são capazes de liberar os gametas (Fonteles-Filho, 1989). Categorias: Desova única, Desova total, Desova parcelada, Desova intermitente
InicioDesova	VARCHAR (10)	Refere-se ao período de início do ato da desova (liberação dos gametas).
FimDesova	VARCHAR (10)	Refere-se ao período de término do ato da desova (liberação dos gametas).
InicioMigracao	VARCHAR (10)	Início do movimento sistemático dos indivíduos pertencentes a um estoque (Sparre & Venema, 1998). Categoria 1: Larva, Juvenil e Adulto.

FimMigracao	VARCHAR (10)	Final do movimento sistemático dos indivíduos pertencentes a um estoque (Sparre & Venema, 1998). Categoria 1: Larva, Juvenil e Adulto.
InicioLarva	VARCHAR (10)	Refere-se ao período no qual registra-se a presença de larva (embrião que se torna independente, antes de assumir os aspectos característicos da espécie) no entanto no ambiente.
FimLarva	VARCHAR (10)	Refere-se ao período no qual não mais se registra a presença de larvas (embrião que se torna independente, antes de assumir os aspectos característicos da espécie) no ambiente.
InicioJuvenil	VARCHAR (10)	Refere-se ao período no qual se registra a presença de juvenis (organismos antes de amadurecer sexualmente pela primeira vez) no ambiente.
FimJuvenil	VARCHAR (10)	Refere-se ao período no qual não mais se registra a presença de juvenis (organismos antes de amadurecer sexualmente pela primeira vez) no ambiente.
InicioAdulto	VARCHAR (10)	Refere-se ao período no qual se registra a presença de adultos (organismos capazes de reproduzir) no ambiente.
FimAdulto	VARCHAR (10)	Refere-se ao período no qual não se registra a presença de adultos (organismos capazes de reproduzir) no ambiente.
Tam1aMaturação	INTEGER(11,0)	Tamanho em que pelo menos metade dos indivíduos potencialmente capazes de se reproduzir tenha atingido esta condição (Fonteles-Filho, 1989). Em centímetro.
HabitoAlimentar	VARCHAR (10)	Tipo de alimento preferencialmente ingerido. Categorias: Carnívoro, Herbívoro, Onívoro e Detritívoro
ComAssintLinf	INTEGER(11,0)	Comprimento teórico máximo (ou assintótico) que a espécie atingiria se vivesse eternamente. Em centímetro.
CrecimentoK	INTEGER(11,0)	É uma medida da taxa na qual o tamanho máximo assintótico é atingido. Em centímetro/ano ou centímetro/mês.
Lmax	INTEGER(11,0)	Tamanho máximo observado para o táxon. Em centímetro.
Tmax	INTEGER(11,0)	Longevidade máxima observada para o táxon. Em anos ou meses.
MortM	INTEGER(11,0)	Mortalidade devido à todas outras causas, excluso a pesca. Isto inclui a predação, fome, doença e morte devido à velhice (Sparre & Venema, 1998).
MortF	INTEGER(11,0)	Mortalidade devido à atividade pesqueira (Sparre & Venema, 1998).

MortZ	INTEGER(11,0)	Soma da mortalidade natural com a mortalidade total (Sparre & Venema, 1998).
Comentários	VARCHAR (255)	Espaço para comentários do pesquisador

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](#)