



INSTITUTO TECNOLÓGICO DE AERONÁUTICA

CE-240 PROJETO DE SISTEMAS DE BANCO DE DADOS

Prof. Dr. Adilson Marques da Cunha

### **Lista de Exercícios 3**

#### **Terceira Forma Normal (3NF) do Aplicativo de Banco de Dados**

Antônio Magno Lima Espescht

[antonio.espescht@gmail.com](mailto:antonio.espescht@gmail.com)

São José dos Campos - SP

13 de abril de 2009

# Sumário

<b>1</b>	<b>Introdução .....</b>	<b>1</b>
1.1	Título .....	1
1.2	Motivação .....	1
1.3	Objetivo .....	1
<b>2</b>	<b>Conteúdo.....</b>	<b>1</b>
2.1	Contextualização.....	2
2.2	Definição das Entidades .....	2
2.3	Definição dos Atributos das Entidades .....	3
2.4	Normalização .....	3
2.4.1	Primeira Forma Normal (1NF).....	3
2.4.2	Segunda Forma Normal (2NF).....	4
2.4.3	Terceira Forma Normal (3NF).....	5
2.5	Exemplos de Dados.....	6
2.6	Dicionarização no ERWin 4.0 .....	8
2.7	Modelo Entidade-Relacionamento.....	9
2.8	Versão 1.0 do Modelo Lógico .....	9
2.9	Versão 1.0 do Modelo Físico.....	10
<b>3</b>	<b>Conclusão e Comentários.....</b>	<b>11</b>
<b>4</b>	<b>Referências .....</b>	<b>11</b>

# 1 Introdução

Em atendimento à Lista de Exercícios (ListEx) 3, foram realizadas as atividades:

- Desenvolver um Protótipo de Aplicativo de Banco de Dados versão 1.0 do módulo **ADE** (Aplicação de Filtros e Cálculo de Estatísticas) do *Sistema de Aquisição, Tratamento, Monitoramento e Difusão de Dados Hidrológicos (Sis ATMH)*.
- Normalizar as entidades [1] desde a Primeira Forma Normal (1NF) até a Terceira Forma Normal (3NF).
- Definir exemplos de dados que preenchem as tabelas definidas.

Em atendimento à sugestão da página 6.1d.7 da aula “A 10ª Técnica de BD – Trigramação” [1], foram realizadas as atividades:

- Elaborar a versão 1.0 do Dicionário de Dados do módulo **ADE** na ferramenta Erwin 4.0.
- Elaborar o Modelo Entidade-Relacionamento do módulo **ADE** na 3NF.

Este texto sumaria os resultados obtidos.

## 1.1 Título

*Obtenção da Terceira Forma Normal (3NF) do Módulo de Aplicação de Filtros e Cálculo de Estatísticas (AFE).*

## 1.2 Motivação

Ao exercitar as técnicas de normalização e modelagem do curso CE-240 iremos obter um modelo de dados mais robusto e menos susceptível a defeitos, erros e falhas na inclusão, na atualização e na exclusão de dados.

É uma oportunidade para aplicarmos, na prática, os conceitos teóricos aprendidos em sala de aula. Além disso, é mais um passo na direção da montagem do Sis ATMH, que é o objetivo final do curso.

## 1.3 Objetivo

Desenvolver a versão 1.0 do modelo de dados do *Módulo de Aplicação de Filtros e Cálculo de Estatísticas (AFE)*, na *Terceira Forma Normal (3NF)*, visando melhorar os tempos de acesso e reduzir anomalias de inclusão, atualização e exclusão de dados.

# 2 Conteúdo

Deve-se ressaltar que as informações de contextualização abaixo descritas foram arbitradas pelo autor e não representam necessariamente a realidade da Agência Nacional de Águas. São apenas de conjecturas para permitir a resolução desta Lista de Exercícios.

## 2.1 Contextualização

A Agência Nacional de Águas (ANA) instalou *Pontos de Coleta de Dados* (PCDs) em diversos locais da bacia Amazônica.

Cada PCD possui um código de identificação e integra diversos tipos de medidores como, por exemplo:

- **GPS**[4]: registra a localização do PCD via latitude, longitude e altitude medidas de acordo com a projeção *Universal Transverse Mercator* (UTM) e o datum WGS84 [5].
- **Linnígrafo**[2]: registra a altura da lâmina d'água de um rio, em metros.
- **Pluviógrafo**[3]: registra a altura pluviométrica acumulada, em milímetros, ao longo do tempo. A partir da altura pluviométrica pode-se inferir a precipitação pluviométrica numa certa área e num certo intervalo de tempo.
- **Termógrafo**: registra a variação da temperatura, em graus Celsius, com o tempo.

Em intervalos de tempo regulares, que podem variar de 1 a 4 semanas, um operador visita o PCD para coletar dados e dar manutenção nos medidores.

Ao retornar para seu escritório, o operador preenche formulários padronizados a partir dos dados coletados em vários PCDs.

Mensalmente, o operador envia os formulários preenchidos para um escritório central no qual são digitados em um sistema da ANA que consiste e consolida estes dados em séries históricas de dados hidrológicos.

O *Módulo de Aplicação de Filtros e Cálculo de Estatísticas* (AFE) deve propiciar o armazenamento destes dados em um Banco de Dados Relacional, permitindo a aplicação de filtros, a identificação de *outliers*[6], o cálculo de estatísticas e o tratamento padronizado de dados espúrios.

## 2.2 Definição das Entidades

A partir da contextualização, definimos as entidades:

- **PCD**: identifica o Ponto de Coleta de Dados.
- **MEDIDOR**: identifica o equipamento que fez a medida. Um PCD contém um ou mais medidores. Cada medidor mede um determinado parâmetro como altura da lâmina d'água, precipitação pluviométrica ou temperatura.
- **OPERADOR**: identifica quem coletou o dado medido. Cada operador realiza medidas numa certa data em um determinado PCD.
- **SERIE**: armazena séries históricas dos dados medidos.
- **FILTRO**: identifica os filtros que podem ser aplicados a um subconjunto da série de dados. Cada filtro corresponde a um algoritmo codificado na forma de "stored procedure" no Banco de Dados.
- **ESTATISTICA**: identifica estatísticas (isto é, valores mínimo, média, máximo e desvio-padrão) calculadas para uma série de dados em um certo período de tempo.

- **PERIODO:** identifica o período (dia, mês ou ano) para o qual foram calculadas as estatísticas para a série de dados.

## 2.3 Definição dos Atributos das Entidades

Em seguida, definimos os atributos de cada entidade:

- **OPERADOR:** identificador do operador, nome, contato.
- **PCD:** identificador do PCD, data de instalação, localização.
- **MEDIDOR:** identificador do medidor, parâmetro medido, descrição.
- **SERIE:** identificador do OPERADOR, identificador do PCD, identificador do MEDIDOR, data da leitura, valor lido, outlier.
- **FILTRO:** identificador do filtro, nome, descrição.
- **ESTATISTICA:** identificador do PCD, identificador do período, dia, mês, ano, mínimo, média, máximo, desvio-padrão.
- **PERIODO:** identificador do período, nome, descrição.

## 2.4 Normalização

Os dados hidrológicos poderiam ser armazenados nesta tabela:

```
DADOS_HIDROLOGICOS_LISTAO (dados_operador, dados_pcd, dados_medidor,  
data_leitura, valor_medido, outlier)
```

Os dados estatísticos, calculados a partir dos dados hidrológicos, podem ser armazenados nesta:

```
DADOS_ESTATISTICOS_LISTAO (identificador_pcd, dados_medidor, dados_periodo,  
dia, mes, ano, minimo, media, maximo, desvio_padrao).
```

Finalmente, os filtros poderiam ser definidos em uma terceira tabela:

```
FILTRO (identificador_filtro, nome_filtro, descricao_filtro)
```

### 2.4.1 Primeira Forma Normal (1NF)

Diz-se que uma tabela ou relação está na 1NF quando **todos os seus registros possuem o mesmo conjunto de atributos e esses atributos são atômicos, isto é, possuem itens indivisíveis.**

Logo, a tabela DADOS\_HIDROLOGICOS\_LISTAO apresentada não está na Primeira Forma Normal pois possui dados não atômicos:

- **dados\_operador:** pode ser subdividido em (chapa\_operador, nome\_operador, contato\_operador). Podemos considerar que contato\_operador não é atômico, obtendo (chapa\_operador, nome\_operador, telefone\_operador, endereco\_operador).

- **dados\_pcd**: pode ser subdividido em (id\_pcd, data\_inst\_pcd, local\_pcd). Considerando que local\_pcd não é atômico, obtemos (id\_pcd, data\_inst\_pcd, latitude\_pcd, longitude\_pcd, altitude\_pcd).
- **dados\_medidor**: pode ser subdividido em (id\_medidor, parametro\_medido).

Dependendo do contexto, alguns atributos poderiam ser ainda subdivididos. Por exemplo: telefone\_operador poderia ser subdividido em ddd\_telefone\_operador e telefone\_operador. No entanto, por simplicidade, iremos considerar um contexto no qual todos os atributos apresentados são atômicos. Logo, a tabela DADOS\_HIDROLOGICOS abaixo descrita contém os mesmos dados que a DADOS\_HIDROLOGICOS\_LISTAO, porém a tabela DADOS\_HIDROLOGICOS está na 1NF:

```
DADOS_HIDROLOGICOS (chapa_operador, nome_operador, telefone_operador,
endereco_operador, id_pcd, data_inst_pcd, latitude_pcd, longitude_pcd,
altitude_pcd, id_medidor, parametro_medido, data_leitura, valor_medido,
outlier)
```

Analogamente, a tabela DADOS\_ESTATISTICOS\_LISTAO contém o atributo dados\_periodo que pode ser subdividido em id\_periodo, nome\_periodo, descricao\_periodo e o atributo dados\_medidor que pode ser subdividido em id\_medidor, parametro\_medido. Realizando estas subdivisões, obtemos a tabela DADOS\_ESTATISTICOS na 1NF:

```
DADOS_ESTATISTICOS (id_pcd, id_medidor, parametro_medido, id_periodo,
nome_periodo, descricao_periodo, dia, mes, ano, minimo, media, maximo, desvio-
padrao).
```

A tabela FILTRO já está na 1NF:

```
FILTRO (id_filtro, nome_filtro, descricao_filtro)
```

## 2.4.2 Segunda Forma Normal (2NF)

A tabela DADOS\_HIDROLOGICOS possui uma anomalia que prejudica as operações de atualização: para identificarmos um registro temos que informar o valor de todos os campos chave (os quais aparecem sublinhados).

A atualização seria menos trabalhosa se tivéssemos que informar apenas a chave da qual o atributo depende funcionalmente. Por exemplo, para alterar nome\_operador poderíamos informar apenas chapa\_operador.

Podemos remover esta anomalia convertendo a tabela para a Segunda Forma Normal (2NF) na qual **todos os atributos não-chave referem-se à chave inteira e não a partes da chave**. Para obter este efeito, subdividimos DADOS\_HIDROLOGICOS em diversas tabelas.

```
OPERADOR(chapa_operador, nome_operador, telefone_operador,  
endereco_operador)
```

```
PCD(id_pcd, data_inst_pcd, latitude_pcd, longitude_pcd, altitude_pcd)
```

```
MEDIDOR(id_medidor, descricao_medidor, parametro_medido)
```

```
SERIE_HISTORICA(chapa_operador, id_pcd, id_medidor, data_leitura,  
valor_medido, outlier)
```

Analogamente, na tabela DADOS\_ESTADISTICOS seria interessante atualizar nome\_periodo informando apenas a chave id\_periodo. Para remover esta anomalia, convertemos a tabela para 2NF obtendo:

```
PERIODO (id_periodo, nome_periodo, descricao_periodo)
```

```
ESTADISTICA (id_pcd, id_medidor, id_periodo, dia, mes, ano, minimo, media,  
maximo, desvio_padrao)
```

A tabela FILTRO já está na 2NF:

```
FILTRO (id_filtro, nome_filtro, descricao_filtro)
```

### 2.4.3 Terceira Forma Normal (3NF)

Um conjunto de tabelas está na Terceira Forma Normal (3NF) quando **estiver na 2NF e cada um de seus atributos não-chave estiverem referindo-se diretamente à sua respectiva chave.**

Analisando as tabelas criadas na 2NF, constatamos que estão na 3NF. Portanto, definimos as tabelas como:

```
OPERADOR(chapa_operador, nome_operador, telefone_operador,  
endereco_operador)
```

```
PCD(id_pcd, data_inst_pcd, latitude_pcd, longitude_pcd, altitude_pcd)
```

```
MEDIDOR(id_medidor, descricao_medidor, parametro_medido)
```

```
SERIE_HISTORICA(chapa_operador, id_pcd, id_medidor, data_leitura,  
valor_medido, outlier)
```

```
PERIODO (id_periodo, nome_periodo, descricao_periodo)
```

```
ESTADISTICA (id_pcd, id_medidor, id_periodo, dia, mes, ano, minimo, media,  
maximo, desvio_padrao)
```

```
FILTRO (id_filtro, nome_filtro, descricao_filtro)
```

Consideramos que os valores filtrados serão retornados por funções e/ou tabelas temporárias na forma dado\_medido, dado\_filtrado. Portanto, os dados filtrados não são armazenados em tabelas.

Mais adiante na integração com outros módulos poderemos reconsiderar esta decisão.

## 2.5 Exemplos de Dados

A título de exemplo, informamos abaixo 5±2 registros para cada entidade de nosso modelo.

**Tabela 1: Exemplo de dados para a Tabela OPERADOR.**

chapa_operador	nome_operador	telefone_operador	endereco_operador
1	José Silva	55 92 1234 5678	Igarapé, 5 – Lagoa – Manaus, AM
2	Maria Silva	55 68 4321 8765	Rua Riozinho, 123 – Bairro das Enchentes – Bujari, AC
3	Cacique Juruna		Tribo dos Piraraquaras, Alto Xingu, PA

**Tabela 2: Exemplo de dados para a Tabela PCD.**

id_pcd	data_inst_pcd	latitude_pcd	longitude_pcd	altitude_pcd
10	01/01/2000	-22,6625	-43,9567	370
10	01/01/2001	-22,5389	-43,7775	389
20	01/01/2000	-22,7489	-44,1247	440

Em nosso modelo, um mesmo PCD pode estar em locais diferentes desde que em datas diferentes. É o que acontece, por exemplo, com o PCD 10 que em 01/01/2000 estava em uma coordenada e em 01/01/2001 foi movido para outra coordenada.

**Tabela 3: Exemplo de dados para a Tabela MEDIDOR.**

id_medidor	descricao_medidor	parametro_medido
1	XPTO1	Altura pluviométrica (mm)
2	XPTO2	Temperatura (Celsius)
3	XPTO3	Lâmina D'Água (m)

Em nosso modelo, consideramos apenas 3 tipos de medidores. Na prática, haveria dezenas de tipos de medidores sendo talvez necessário especificar melhor cada medidor, armazenando, por exemplo: erro médio esperado, dados do fabricante, dados de calibração etc.



**Tabela 4: Exemplo de dados para a Tabela SERIE\_HISTORICA.**

chapa_operador	id_pcd	id_medidor	data_leitura	valor_medido	outlier
1	10	1	01/01/2000 10:30:35	34	F
1	10	1	01/01/2000 10:32:35	35	F
1	10	1	01/01/2000 10:34:40	42	T
2	20	1	01/01/2000 10:30:35	34	F
2	20	2	01/01/2000 10:30:35	50	F
2	20	3	01/01/2000 10:30:35	1,25	F
3	20	3	01/02/2000 12:30:35	1,35	F

Nota-se que o terceiro registro deste exemplo é um outlier pois seu valor está completamente fora do esperado, sendo pouco provável que a temperatura aumentasse 7 °C em apenas 2min.

**Tabela 5: Exemplo de dados para a Tabela PERIODO.**

id_periodo	nome_periodo	descricao_periodo
D	Dia	Cálculos estatísticos considerando períodos de 1 dia.
M	Mês	Cálculos estatísticos considerando períodos de 1 mês.
A	Ano	Cálculos estatísticos considerando períodos de 1 ano.

**Tabela 6: Exemplo de dados para a Tabela ESTATISTICA.**

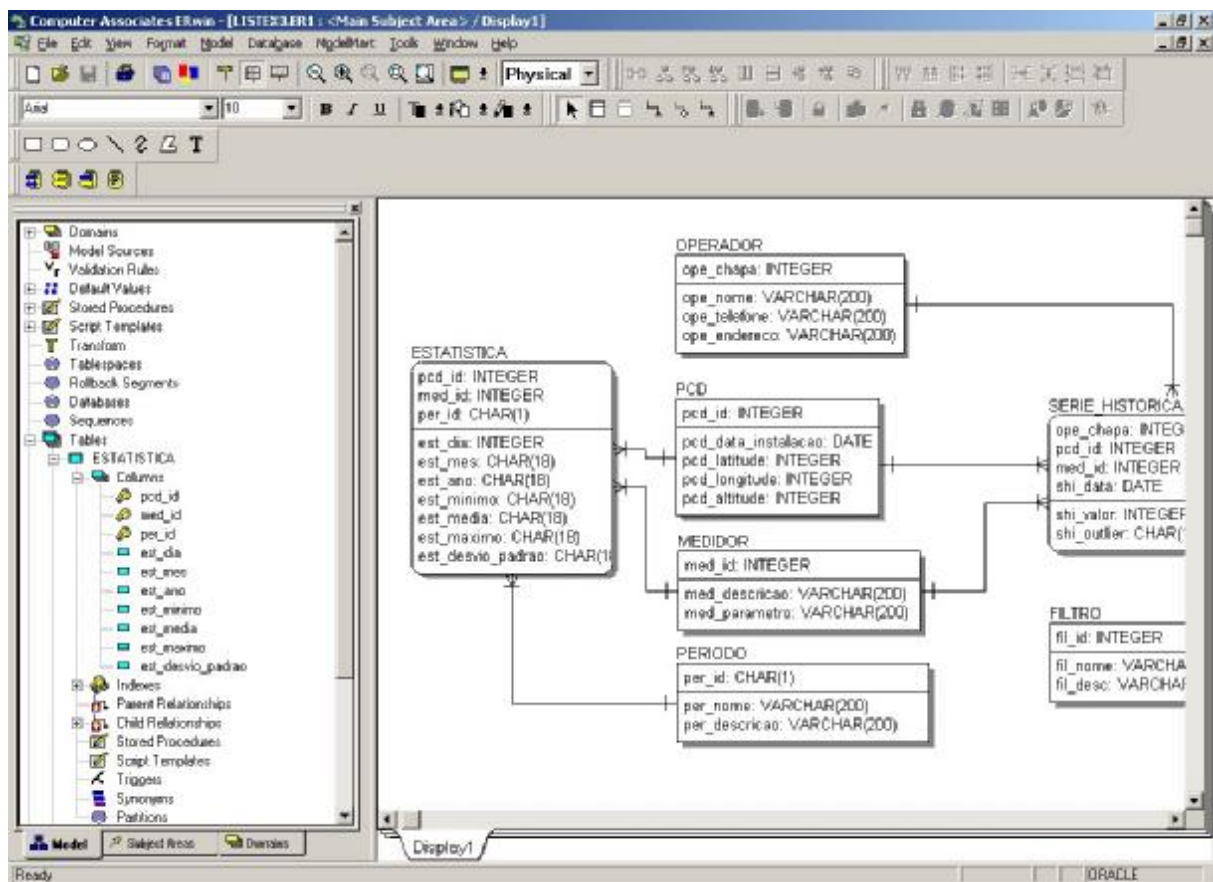
id_pcd	id_medidor	id_periodo	dia	mes	ano	minimo	media	maximo	desvio_pad
1	10	D	01	01	2000	34,0	34,5	35,0	1,0
1	10	M	NULL	01	2000	34,0	34,5	35,0	1,0
1	10	A	NULL	NULL	2000	34,0	34,5	35,0	1,0
2	20	D	01	01	2000	34,0	34,0	34,0	0
2	20	M	NULL	01	2000	34,0	34,0	34,0	0
2	20	A	NULL	NULL	2000	34,0	34,0	34,0	0

**Tabela 7: Exemplo de dados para a Tabela FILTRO.**

id_filtro	nome_filtro	descricao_filtro
1	Média Móvel	Retorna o valor filtrado como sendo a média aritmética dos N valores anteriormente lidos numa série histórica.
2	Outlier	Preenche o campo outlier para uma determinada série histórica.
3	Mínimos	Retorna o valor filtrado como sendo o valor mínimo dos N valores anteriormente lidos numa série histórica. Valores com outlier=True são ignorados.
4	Máximos	Retorna o valor filtrado como sendo o valor máximo dos N valores anteriormente lidos numa série histórica. Valores com outlier=True são ignorados.

## 2.6 Dicionarização no ERWin 4.0

A definição das tabelas na 3NF foi inserida na ferramenta ERWin 4.0 gerando a versão 1.0 do Dicionário de Dados que é parcialmente visível na parte esquerda da Figura 1.



**Figura 1: Vista parcial do dicionário de dados no ERWin 4.0**

## 2.7 Modelo Entidade-Relacionamento

Com o auxílio do ERWin 4.0 foi obtido o Modelo Entidade-Relacionamento do módulo ADE (Figura 2).

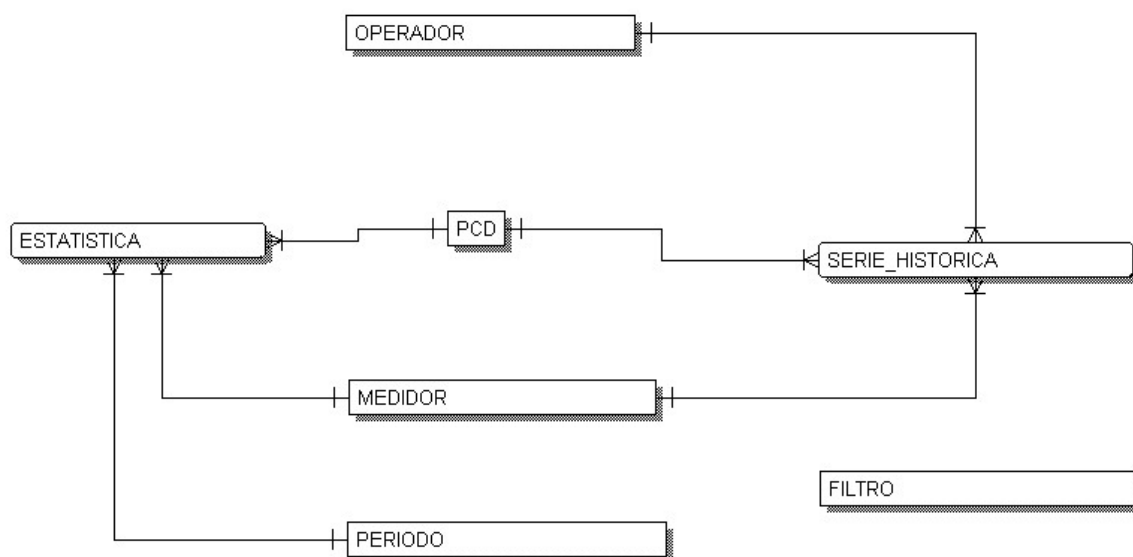


Figura 2: Modelo Entidade-Relacionamento do Módulo ADE no ERWin 4.0

## 2.8 Versão 1.0 do Modelo Lógico

Com o auxílio do ERWin 4.0 foi obtido o Modelo Lógico do módulo ADE (Figura 3).

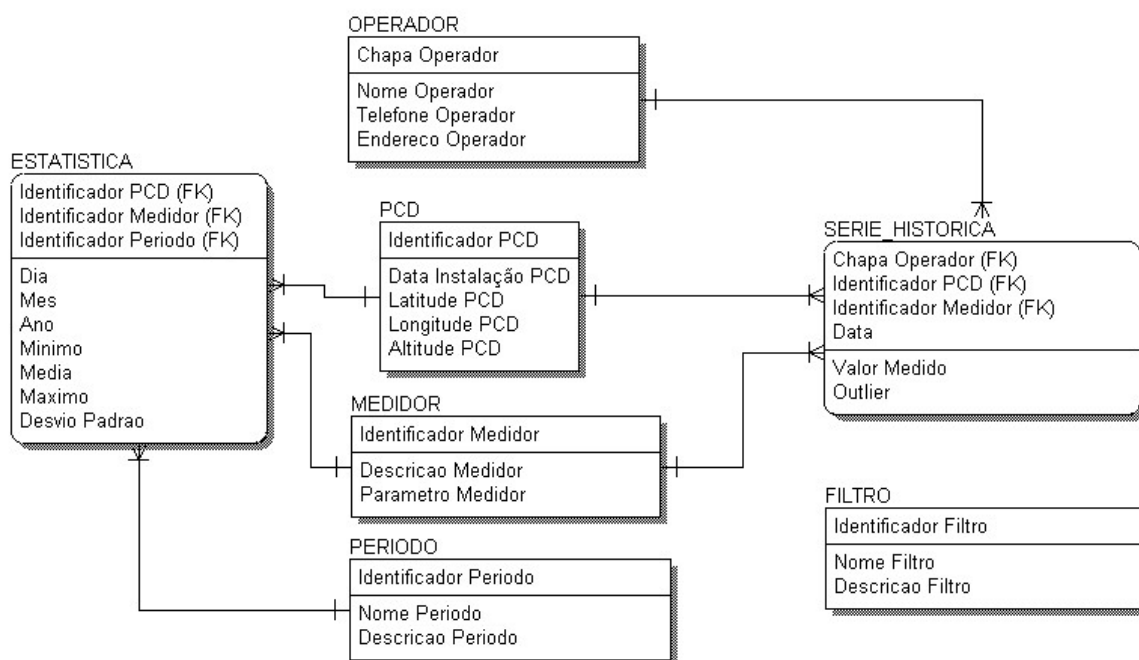


Figura 3: Modelo Lógico do Módulo ADE no ERWin 4.0

## 2.9 Versão 1.0 do Modelo Físico

A partir do Modelo Lógico e com o auxílio do ERWin 4.0, foi utilizada a Técnica da Trigramação para obter o Modelo Físico do módulo ADE (Figura 4).

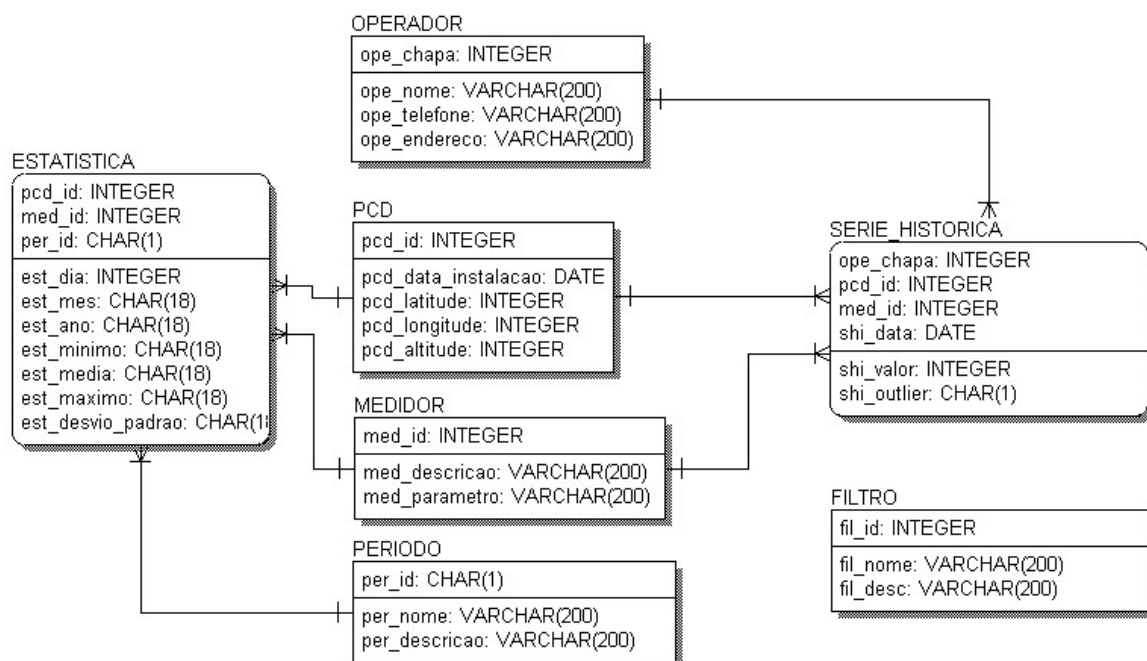


Figura 4: Modelo Físico do Módulo ADE no ERWin 4.0

### 3 Conclusão e Comentários

As técnicas de normalização e modelagem do curso CE-240 foram aplicadas ao módulo **ADE** (*Aplicação de Filtros e Cálculo de Estatísticas*) do *Sistema de Aquisição, Tratamento, Monitoramento e Difusão de Dados Hidrológicos (Sis ATMH)* obtendo-se sete entidades.

O número relativamente alto de entidades é justificado pelo módulo **ADE** ser o resultado da união dos módulos **ADF** (*Aplicação de Filtros*) e **CDL** (*Cálculo de Limites*) que em 23/03/2009 foram agrupados num único módulo visando compatibilizar o número de módulos a desenvolver com a quantidade de alunos desenvolvedores.

As técnicas de normalização e modelagem mostram-se muito eficientes, resultando na criação de tabelas na 3NF de modo praticamente direto. Houve certa dificuldade apenas para reconstruir as tabelas nas formas 2NF e 1NF, para fins didáticos, conforme solicitado na ListEx 3.

A definição da tabelas na 3NF foi inserida na ferramenta ERWin 4.0 o que permitiu gerar o Modelo Lógico e, quase simultaneamente, o Modelo Físico do módulo ADE. A técnica da trigramação foi aplicada, renomeando-se os atributos físicos criados automaticamente pelo ERWin 4.0.

Como sugestão, acredito que seria interessante exibir aos alunos a contextualização real da ANA bem como exemplos de dados reais para que estes tivessem uma visão mais prática do problema a ser resolvido.

### 4 Referências

- [1] CUNHA, ADILSON MARQUES DA. Notas de Aula de “CE-240 Projeto de Sistemas de Bancos de Dados” no Primeiro Período de 2009. ITA – Instituto Tecnológico de Aeronáutica. Disponível em: <http://www.comp.ita.br/~cunha>. Acessado em: 04 de abril de 2009.
- [2] MARTINS, JOSÉ RODOLFO S. e FADIGA JR., FRANCISCO M. **Hidrologia Básica**. Notas de Aula de “Capacitação Tecnológica e Transferência de Tecnologia em Drenagem Urbana”. 2003. EPUSP – Escola Politécnica da Universidade de São Paulo. Disponível em: [www.fcth.br/public/cursos/praiagrande/1%20-%20Hidrologia%20Basica.ppt](http://www.fcth.br/public/cursos/praiagrande/1%20-%20Hidrologia%20Basica.ppt). Acessado em: 05 de abril de 2009.
- [3] MEDEIROS, YVONILDE. **Precipitação**. Apostila do Curso “ENG-371 Hidrologia” do Grupo de Recursos Hídricos (GRH). Cap. 03. 2005. UFBA – Universidade Federal da Bahia. Disponível em: [http://www.grh.ufba.br/download/2005.2/Apostila\(Cap3%20-%20Parte1\).pdf](http://www.grh.ufba.br/download/2005.2/Apostila(Cap3%20-%20Parte1).pdf). Acessado em: 04 de abril de 2009.
- [4] RODRIGUES, ALEX PINHEIRO MACHADO. **Portal GPS**. Disponível em: <http://www.portalgps.com.br>. Acessado em: 06 de abril de 2009.
- [5] WIKIMEDIA FOUNDATION, INC. Wikipedia (*Datum*). Disponível em: <http://pt.wikipedia.org/wiki/Datum>. 2009. Visitado em: 04 de abril de 2009.
- [6] WIKIMEDIA FOUNDATION, INC. Wikipedia (*Outlier*). Disponível em: <http://en.wikipedia.org/wiki/Outlier>. 2009. Visitado em: 28 de março de 2009.