



Pós-Graduação em Ciência da Computação

**MAILSON MELO DOS SANTOS FILHO**

**UMA ABORDAGEM MULTIDIMENSIONAL PARA OLAM COMO FERRAMENTA DE  
AVALIAÇÃO DE DESEMPENHO DE MODELOS DE CLASSIFICAÇÃO BINÁRIA**



Universidade Federal de Pernambuco

[posgraduacao@cin.ufpe.br](mailto:posgraduacao@cin.ufpe.br)

[www.cin.ufpe.br/~posgraduacao](http://www.cin.ufpe.br/~posgraduacao)

**RECIFE**

**2015**

**MAILSON MELO DOS SANTOS FILHO**

**UMA ABORDAGEM MULTIDIMENSIONAL PARA OLAM COMO FERRAMENTA DE  
AVALIAÇÃO DE DESEMPENHO DE MODELOS DE CLASSIFICAÇÃO BINÁRIA**

ESTE TRABALHO FOI APRESENTADO À PÓS-GRADUAÇÃO EM  
CIÊNCIA DA COMPUTAÇÃO DO CENTRO DE INFORMÁTICA DA  
UNIVERSIDADE FEDERAL DE PERNAMBUCO COMO  
REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE  
EM CIÊNCIA DA COMPUTAÇÃO

ORIENTADOR: PROF. DR. PAULO JORGE LEITÃO ADEODATO

**RECIFE**

**2015**

Catálogo na fonte  
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S237a Santos Filho, Mailson Melo dos  
Uma abordagem multidimensional para OLAM como ferramenta de avaliação de desempenho de modelos de classificação binária / Mailson Melo dos Santos Filho. – 2015.  
96 f.: il., fig., tab.

Orientador: Paulo Jorge Leitão Adeodato.  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Informação, Recife, 2015.  
Inclui referências e apêndice.

1. Mineração de dados. 2. Inteligência artificial. I. Adeodato, Paulo Jorge Leitão (orientador). II. Título.

006.312

CDD (23. ed.)

UFPE- MEI 2017-78

## FOLHA DE APROVAÇÃO

Dissertação de Mestrado apresentada por **MAILSON MELO DOS SANTOS FILHO** à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título “**Uma Abordagem Multidimensional para OLAM como Ferramenta de Avaliação de Desempenho de Modelos de Classificação Binária**” orientada pelo **Prof. Paulo Jorge Leitão Adeodato** e aprovada pela Banca Examinadora formada pelos professores:

---

Prof. Robson do Nascimento Fidalgo  
Centro de Informática / UFPE

---

Prof. Tiago Alessandro Espínola Ferreira  
Departamento de Estatística e Informática / UFRPE

---

Prof. Paulo Jorge Leitão Adeodato  
Centro de Informática / UFPE

Visto e permitida a impressão.  
Recife, 9 de setembro de 2015.

---

**Profa. Edna Natividade da Silva Barros**

Coordenadora da Pós-Graduação em Ciência da Computação do  
Centro de Informática da Universidade Federal de Pernambuco.

## AGRADECIMENTOS

Finalizando este curso de mestrado, agradeço a todas as pessoas que acreditaram em mim, me apoiando de diversas formas, a fim de vencer as dificuldades. Em especial à minha família.

Ao professor Paulo Jorge Leitão Adeodato, pela sua orientação, dedicação, paciência e por sempre acreditar que iríamos conseguir obter resultados relevantes durante a pesquisa.

A todos os membros da banca. É uma grande honra poder contar com a contribuição de todos.

Aos amigos Márcio, Paulo e Jorge pelo constante incentivo e apoio na elaboração desse trabalho.

À professora Mônica Ximenes Carneiro da Cunha, ex-orientadora enquanto graduação, por ter me ensinado muito sobre pesquisa e pelo grande incentivo.

Ao Centro de Informática da UFPE pela infraestrutura disponibilizada e pelo alto nível acadêmico de seus professores.

A todos que direta ou indiretamente contribuíram para a realização desta pesquisa. .

## RESUMO

Os sistemas de suporte à decisão são utilizados para auxiliar os gestores na tomada de decisão, inclusive, de alto nível. Eles implementam todas as etapas do processo de extração de conhecimento em banco de dados (KDD - *Knowledge-Discovery in Databases*) e o ambiente de consultas a dados e conhecimento, incluindo simulação de cenários sobre os indicadores do negócio (KPIs - *Key Performance Indicators*). As ferramentas OLAP (*Online Analytical Processing*) oferecem um ambiente para consultas que permite a análise de um grande volume de dados, por meio de cruzamento entre dados, apresentação de perspectiva de processamento multinível (hierarquias) para os dados e funções de perfuração e fatiamento dos conjuntos entre outras. Estas ferramentas têm interface amigável com o decisor humano apresentando gráficos, modelos e sumarizações, mas são limitadas a operações diretas e à validação de conhecimento humano sobre os dados do negócio. No processo de KDD, a mineração de dados usa inteligência artificial e estatística para extração de conhecimento dos dados e para apoiar a decisão humana em atividades de inferência e previsão. OLAM (*Online Analytical Mining*) é um conceito equivalente ao OLAP para navegação sobre o conhecimento extraído e sobre o efeito da sua aplicação na tomada de decisões sobre o desempenho do negócio tanto nos indicadores técnicos de inteligência artificial quanto nos KPIs. No entanto, não foi encontrado ferramental genérico que permita realizar de forma sistemática o OLAM. A proposta desse trabalho é mostrar que algumas operações fundamentais de OLAM podem ser implementadas por meio de ferramentas OLAP, mediante amostragem de dados com independência estatística submetidos à mineração e a certas transformações de dados. O trabalho é validado por meio de um problema de decisão binária, no qual a resposta do algoritmo de mineração de dados (escore de propensão) sobre a massa de dados estatisticamente independente passa a ser uma medida no *data warehouse*. O trabalho demonstra que essa abordagem permite a geração de cubos associado à classe-alvo e aos KPIs para avaliação de desempenho, tornando a ferramenta OLAP capaz de realizar OLAM. Ao final, o trabalho analisa as principais limitações das ferramentas OLAP que as impede de implementar OLAM com mais flexibilidade e as implicações conceituais e operacionais para serem superadas.

**Palavras-Chave:** OLAM. OLAP. Data Mining. Sistemas de Apoio a Decisão. Avaliação de Desempenho

## ABSTRACT

Decision support systems are used to assist managers inclusive in making high-level decisions. They implement all stages of knowledge extraction process in database (KDD - Knowledge Discovery in Databases) and the environment queries to data and knowledge, including simulation scenarios on the business indicators (KPIs - Key Performance Indicators). The OLAP (Online Analytical Processing) provides an environment for queries that allows analysis of large volumes of data through data merging, multilevel processing perspective presentation (hierarchies) for data and drilling functions and slicing of among other sets of functions. These tools have user-friendly interface with the human decision maker presenting graphics, models and summarization, but are limited to direct operations and human knowledge validation on business data. In the process of KDD, data mining uses artificial intelligence and statistical techniques for knowledge extraction in inference and forecasting activities for supporting human decision. OLAM (Online Analytical Mining) is a concept equivalent to OLAP to navigation on the extracted knowledge and the effect of applying them in making decisions about business performance both in technical indicators artificial intelligence as the KPIs. However, it was not found generic tools that allows perform systematically the OLAM. The purpose of this paper is to show that some fundamental operations of OLAM can be implemented by means of OLAP tools, sampled data with statistical independence subject to certain mining and data transformations. The job is validated by means of a binary decision problem, in which the response data mining algorithm (propensity score) statistically independent data about the mass becomes a measure in the data warehouse. The work demonstrates that this approach allows the generation of cubes associated with class-blalvo and KPIs for performance evaluation, making OLAP tool capable of performing OLAM. Finally, the paper analyzes the main limitations of OLAP tools that prevents them from implementing OLAM with more flexibility and the conceptual and operational implications for trying to overcome.

**Keywords:** OLAM. OLAP. Data Mining. Decision Support System. Performance evaluation

## LISTA DE FIGURAS

Figura 1 - Etapas de KDD (adaptado de Fayyad et al. 1996).....	15
Figura 2 – Modelos Dimensionais, Esquema Estrela e Snowflake (Fortulan e Gonçalves, 2005) .....	24
Figura 3 - Cubo de dados de vendas por trimestre, local e tipo de item (Han e Kamber, 2001) .....	27
Figura 4 - Cubo de Dados (adaptado de Microsoft, 2015) .....	31
Figura 5 - Exemplo da operação slice, dice, drill-down e roll-up .....	32
Figura 6. Arquitetura OLAM (Han, Cheen e Chiang, 1998).....	36
Figura 7. Curva ROC Modelo Geral .....	44
Figura 8. Modelo Dimensional.....	45
Figura 9. Criação da Medida Escore Máximo.....	47
Figura 10. Criação de Medida a partir de MDX .....	47
Figura 11. Visão Geral Seleção Dimensões e Medidas .....	48

## LISTA DE TABELAS

Tabela 1 - Diferenças entre BD transacionais e Data Warehouse (adaptado de Singh, 2001).....	21
Tabela 2. Tabela de Variáveis.....	42
Tabela 3. Modelo Reduzido.....	43
Tabela 4. Consulta variável Faixa Quantidade de Alunos.....	49
Tabela 5. Consulta variável Faixa Número de Funcionários.....	50
Tabela 6. Consulta variável Quantidade de Banheiros na Residência.....	51
Tabela 7. Consulta variável Escolaridade da Mãe.....	51

## LISTA DE ABREVIATURAS

<b>BI</b>	Business Intelligence
<b>DW</b>	<i>Data Warehouse</i>
<b>DM</b>	<i>Data Mart</i>
<b>EIS</b>	<i>Executive Information System</i>
<b>ERP</b>	<i>Enterprise Resource Planning</i>
<b>FAD</b>	Ferramentas de Apoio à Decisão
<b>KDD</b>	<i>Knowledge Discovery in Databases</i>
<b>MDX</b>	<i>Multidimensional Expressions</i>
<b>OLAM</b>	<i>OnLine Analytical Mining</i>
<b>OLAP</b>	<i>OnLine Analytical Processing</i>
<b>OLTP</b>	<i>OnLine Transaction Processing</i>
<b>ROC</b>	<i>Receiver Operating Characteristics</i>
<b>SAD</b>	Sistemas de Apoio à Decisão
<b>SGBD</b>	Sistema de Gerenciamento de Banco de Dados
<b>SQL</b>	<i>Structured Query Language</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>11</b>
1.1.	CONTEXTO .....	11
1.2.	OBJETIVOS .....	13
1.3.	ESTRUTURA DO TRABALHO .....	13
<b>2</b>	<b>ONLINE ANALYTICAL PROCESSING .....</b>	<b>14</b>
2.1.	<i>KNOWLEDGE DISCOVERY IN DATABASES</i> .....	14
2.1.1.	<i>Seleção</i> .....	16
2.1.2.	<i>Pré-processamento</i> .....	16
2.1.3.	<i>Formatação</i> .....	17
2.1.4.	<i>Mineração de dados</i> .....	17
2.1.5.	<i>Interpretação/Avaliação</i> .....	19
2.2.	DATA WAREHOUSE.....	19
2.2.1.	<i>Características de um data warehouse</i> .....	21
2.2.2.	<i>Modelagem Dimensional</i> .....	22
2.2.3.	<i>Granularidade</i> .....	25
2.2.4.	<i>ETL</i> .....	25
2.3.	O ANALYTICAL PROCESSING .....	26
2.3.1.	<i>Cubo de dados</i> .....	30
2.3.2.	<i>Conjunto de operações OLAP</i> .....	31
<b>3</b>	<b>ONLINE ANALYTICAL MINING .....</b>	<b>34</b>
3.1.	FERRAMENTAS OLAM ENCONTRADAS.....	36
3.2.	INTEGRAÇÃO PROPOSTA.....	38
<b>4</b>	<b>MODELAGEM.....</b>	<b>39</b>
4.1.	CONTEXTUALIZAÇÃO .....	39
4.2.	MEDIDA DE AVALIAÇÃO DE DESEMPENHO DO MODELO .....	40
4.3.	REGRESSÃO LOGÍSTICA .....	41
<b>5</b>	<b>CONTRIBUIÇÕES .....</b>	<b>45</b>
5.1.	<i>DATA WAREHOUSE</i> .....	45
5.2.	MODELO DIMENSIONAL .....	46
5.3.	MEDIDAS .....	46
5.4.	RESULTADOS .....	48
<b>6</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>52</b>
6.1.	CONTRIBUIÇÕES .....	53
6.2.	LIMITAÇÕES.....	53
6.3.	TRABALHOS FUTUROS .....	53
	<b>REFERÊNCIAS .....</b>	<b>55</b>
	<b>APÊNDICE .....</b>	<b>60</b>

## 1 INTRODUÇÃO

Este capítulo apresenta os motivos que levaram ao desenvolvimento desta dissertação e formaliza os seus objetivos. Em seguida, apresenta a estrutura em que os demais capítulos estão organizados.

### 1.1. CONTEXTO

O cenário corporativo mundial acompanhou grandes transformações no setor de Tecnologia da Informação (TI) recentemente, em que a TI passou a se apresentar como instrumento essencial para as organizações e, assim, vem agregando valor ao negócio. Tal transição é justificada fundamentalmente pela competição existente no mercado, o que fez da TI um meio de acesso e descoberta de informações mais precisas para a tomada de decisões. Dependendo do tipo de negócio que a organização realiza, é indispensável que os serviços prestados aos clientes também disponham de tecnologia (LAUDON & LAUDON, 2009).

A mudança de paradigma, causada pelos avanços tecnológicos, possibilitou extraordinária capacidade de coleta, processamento e armazenamento de grandes bases de dados. Essa superabundância de dados, que supera a capacidade humana de análise e extração do conhecimento contido ou “escondido” nos dados, impulsionou o surgimento de novo ramo da computação, a descoberta de conhecimento em bases de dados, do inglês *Knowledge Discovery in Databases* (KDD), (HAN & KAMBER, 2006; WITTEN & FRANK, 2005).

O Processo de KDD tem como objetivo principal encontrar uma maneira estruturada de, com o uso de Tecnologia da informação (TI), explorar essas bases de dados e reconhecer os padrões existentes pela modelagem de fenômenos do mundo real (FAYYAD, 1996).

Os sistemas de suporte à decisão são utilizados para auxiliar os gestores na tomada de decisão, inclusive naquelas de alto nível. Eles implementam todas as etapas do processo de extração de conhecimento em banco de dados (KDD) e o ambiente de consultas a dados e conhecimento, incluindo simulação de cenários sobre os indicadores do negócio (KPIs - *Key Performance Indicators*).

Para Drucker (1998) uma empresa deve ser fundamentada na informação e no conhecimento, para ser bem sucedida, trabalhando estes pontos como capital estratégico. É um grande desafio fazer com que as informações sejam coletadas, formatadas e usadas sistematicamente na tomada de decisões. Os indicadores de desempenho ajudam a definir e medir o progresso do negócio de acordo com o quão próximo as variáveis estão para satisfazer os KPIs atribuídos.

Em sistemas KDD, os conhecimentos para tomada de decisão são obtidos com a aplicação de técnicas de mineração de dados, que pode ser realizada sobre os dados contidos em bancos de dados de sistemas transacionais, após serem integrados e transformados. Porém, o ideal é que seja construído um *Data Warehouse* (DW), onde os dados já estarão preparados para servirem de entrada na aplicação dessas técnicas, podendo ser utilizados também por ferramentas *On-Line Analytical Processing* (OLAP) (DIAS, 2001).

As ferramentas de OLAP são importantes no processo decisório, permitindo analisar os dados de forma flexível e interativa, ajudando os analistas a sintetizarem informações sobre a organização, por meio de comparações, visões personalizadas, análises históricas e projeções de dados (KIMBALL *et al.*, 1998).

O OLAP tradicional é capaz de manipular apenas dados estruturados, portanto é importante estender essa visão e permitir analisar dados não estruturados e combiná-los entre si, possibilitando executar operações analíticas sobre todas as fontes de informação da organização (ZHANG *et al.*, 2009; NAPOLI *et al.*, 2006; NAPOLI; SELL, 2010).

Segundo Kimball (1998), enquanto OLAP é dedutivo e guiado por especialistas, Mineração de Dados é indutivo e guiado pelos próprios dados. Ambas necessitam de dados limpos e consistentes. E neste caso, o DW é capaz de fornecer dados para as duas tecnologias, o que o torna a principal fonte de dados para *Online Analytical Mining* (OLAM), cujo termo refere-se à junção de OLAP e Mineração de Dados.

Segundo Han e Kamber (2006), OLAM significa minerar interativamente em diferentes porções dos dados e em diferentes níveis de agregação, utilizando operações OLAP, podendo-se escolher as funções de Mineração de Dados e algoritmos dinamicamente, além de poder navegar pelos resultados da mineração.

No entanto, não foi encontrado ferramental genérico que permita realizar de forma sistemática o OLAM. A proposta desse trabalho é mostrar que algumas operações fundamentais de OLAM podem ser implementadas por meio de ferramentas OLAP, mediante amostragem de dados com independência estatística submetidos à mineração e a certas transformações de dados.

O trabalho é validado por meio de um problema de decisão binária, no qual a resposta do algoritmo de mineração de dados (score de propensão) sobre a massa de dados estatisticamente independente passa a ser um novo campo no DW. O trabalho demonstra que essa abordagem permite a geração de cubos associados à classe-alvo e aos KPIs para avaliação de desempenho, tornando a ferramenta OLAP capaz de realizar OLAM.

Ao final, o trabalho analisa as principais limitações das ferramentas OLAP que as impede de implementar OLAM com mais flexibilidade e as implicações conceituais e operacionais a serem superadas.

## 1.2. OBJETIVOS

Neste contexto, o principal objetivo do presente trabalho é propor uma abordagem para avaliação de desempenho de classificadores binários de forma dimensional.

A proposta assegura o desenvolvimento de uma abordagem OLAM a partir de dados de um DW com o acréscimo de dados oriundos da resposta do modelo de mineração de dados e dos impactos das decisões tomadas, com base nos KPIs do negócio.

## 1.3. ESTRUTURA DO TRABALHO

Este trabalho é composto por seis capítulos. No capítulo 2 é apresentado *Online Analytical Processing*, bem como os principais conceitos e abordagens utilizadas nas ferramentas analíticas.

O Capítulo 3 aborda os conceitos básicos sobre o processo de *Online Analytical Mining*. Outros aspectos relacionados ao desenvolvimento deste trabalho são detalhados, entre eles um breve histórico sobre ferramentas que implementam *Online Analytical Mining* e as abordagens para construção desse ferramental.

No Capítulo 4, é detalhada a metodologia experimental utilizada partindo desde a contextualização do caso utilizado para criação do modelo a ser avaliado.

No Capítulo 5 são apresentadas as principais contribuições obtidas no trabalho. A elaboração do data warehouse utilizado, como foram escolhidas as medidas, a criação do cubo de dados e apresentação dos resultados da abordagem.

Finalmente no Capítulo 6, são apresentadas as conclusões, dissertando sobre algumas limitações e trabalhos futuros relevantes para complementar este estudo.

## 2 **ONLINE ANALYTICAL PROCESSING**

Esse capítulo apresenta a revisão de literatura sobre *OnLine Analytical Processing* e suas tecnologias focando nos principais aspectos relativos à elaboração desse trabalho.

### 2.1. **KNOWLEDGE DISCOVERY IN DATABASES**

Os Sistemas de Apoio à Decisão (SAD) são sistemas computacionais que visam à sistematização e apoiar os processos decisórios empresariais, sendo comumente compostos pelos componentes: interface com o usuário, banco de dados e banco de modelos, visando fornecer apoio em decisões semiestruturadas e desestruturadas, estando sob controle do tomador de decisões e disponibilizando um conjunto de ferramentas para estruturar e aumentar a efetividade das decisões (JUNIOR et al., 2006).

De acordo com Date (2004), o processo de tomada de decisão com auxílio de computadores iniciou-se na década de 70, quando os processos começaram a ser informatizados e as informações passaram a ser pré-definidas e selecionadas por meio dos *Executive Information Systems* (EIS). Atualmente, o processo de tomada de decisão é informatizado, cabendo ao gestor definir atributos que sejam mais importantes ao processo decisório, tendo o apoio de informações oriundas do sistema de apoio à decisão, através de ferramentas OLAP, que será discutido o capítulo 5.

Como fonte primária para o SAD temos os sistemas de produção ou gestão, mais conhecidos como *Enterprise Resource Planning* (ERP), geralmente eles recuperam e atualizam um registro por vez, ou seja, um cliente, um produto, dentre outros. Conhecidos como sistemas transacionais, eles atuam em paralelo e garantem o acesso de muitos usuários aos recursos operacionais das organizações. Já os SAD lidam com poucos usuários por vez e não há necessidade crítica de acesso aos recursos operacionais, eles têm como fonte principal os dados oriundos dos sistemas ERP e que podem conter um *Data Warehouse*, que é responsável pela consolidação dos dados e auxílio a tomada de decisão.

Para que o SAD auxilie no processo de tomada de decisão é preciso que o mesmo seja composto de uma base de dados que auxilie o sistema, uma base de modelos que provê a capacidade de análise e uma interface visual que forneça o diálogo e a interação entre o usuário e o sistema (DUARTE et al., 2005). Os sistemas de apoio à decisão se utilizam de várias tecnologias, dentre elas, *Data Warehouse*, *Data Mart*, Processamento Analítico On-line (OLAP), Banco de Dados Multidimensionais, Mineração de Dados dentre outras.

As Ferramentas de Apoio à Decisão (FAD) fazem parte do conceito de *Business Intelligence* (BI), ou Inteligência de Negócios, e correspondem ao conjunto de tecnologias que permitem o cruzamento de informações e suporte a análise dos indicadores de desempenho de um negócio (COLAÇO, 2004).

As análises e recursos como tendências e comparações não são as únicas características que um SAD consegue fazer de forma eficiente. O método de fatiar e agrupar permite que o gerente disponha de muitas perspectivas diferentes das atividades que estão ocorrendo (INMON, 1997).

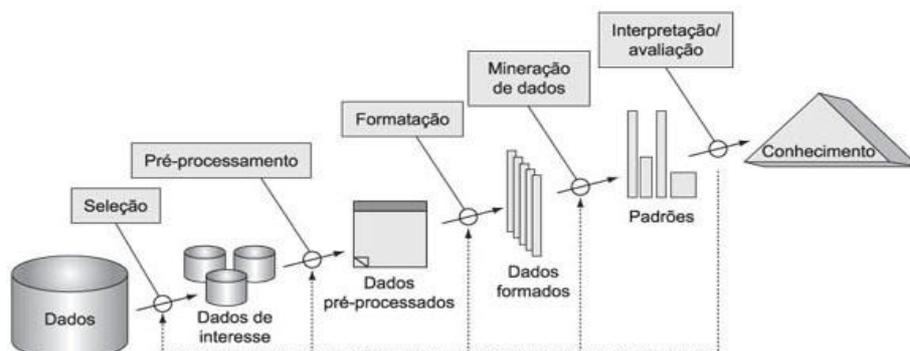
Fayyad et al. (1996), definem Knowledge Discovery in databases (KDD) ou descoberta de conhecimento em banco de dados como “processo não trivial de identificação, a partir de dados, de padrões que sejam válidos, novos, potencialmente úteis e compreensíveis.”

O termo não trivial no conceito de KDD significa a necessidade de utilização de mecanismos de busca ou inferência, o que difere de um processamento de dados direto em uma quantidade pré-definida. O conjunto de dados pode ser representado pelos fatos, enquanto os padrões expressam alguma linguagem capaz de descrever um subconjunto de dados. Estes devem ser válidos e possuir algum grau de precisão. Estes novos padrões gerados, dependendo de sua natureza podem ser considerados conhecimento, e, quando trazem algum benefício novo o qual seja compreendido rapidamente, podem ajudar os usuários na tomada de decisões.

Segundo Han e Kamber (2006), o processo de KDD pode ser visto como uma evolução natural da tecnologia de informação. Hoje é possível buscar e armazenar grandes quantidades de dados. Contudo, essa quantidade de dados gerados superou a capacidade humana de compreensão dos mesmos, o que criou a necessidade de ferramentas para realizar esta análise.

No contexto da execução e aplicação de resultados do KDD, Fayyad et al. (1996) propuseram um processo iterativo e iterativo envolvendo algumas etapas e decisões que devem ser feitas pelos usuários. As etapas sugeridas por Fayyad et al. (1996) para o processo de KDD foram as seguintes, conforme Figura 1 (FAYYAD et al., 1996):

Figura 1 - Etapas de KDD (adaptado de Fayyad et al. 1996)



Como demonstrado na figura 1, o processo é dividido em Seleção, Pré-processamento, Formatação, Mineração de dados e Interpretação/Avaliação.

#### 2.1.1. SELEÇÃO

Nesta fase é realizada a escolha das fontes de dados, podendo ser fontes internas ou externas da empresa ou mesmo fonte de conhecimento tácito dos especialistas. Ela visa a seleção dos dados a serem pré-processados, formatados, minerados e avaliados para obtenção do conhecimento. É a fase de escolha do objeto de estudo.

#### 2.1.2. PRÉ-PROCESSAMENTO

Nesta etapa são realizadas a aplicação de funções para remover registros duplicados, estratégias para o tratamento de campos em branco ou nulos, entre outras funções de suma importância para o resultado da mineração de dados.

Um dos principais problemas da mineração de dados é a qualidade dos dados que serão passados para o algoritmo, se os dados são passados com baixa qualidade para o algoritmo, o resultado também terá uma baixa qualidade. Além disso, há algoritmos que precisam dos dados em uma forma padrão, valores contínuos, sem ausência de valores, só aceitam dados categóricos ou binários, dentre outras características. O pré-processamento automatizado ou mesmo de forma manual tem como função resolver essas deficiências.

A ausência de valores também é uma situação a ser avaliada. Neste caso a causa pode estar na coleta dos dados, remoção de dados devido à inconsistência ou o próprio significado do atributo ser incompreensível. No caso de ausência de valores de atributos, existem as seguintes alternativas:

- Remover registros com valores faltantes;
- Prever o valor faltante com base nos valores de outros atributos;
- Lidar com os valores faltantes dentro do algoritmo de MD;
- Substituir os valores ausentes pela moda (valor mais frequente, no caso de atributos categóricos) ou pela média ou mediana (no caso de valores contínuos).

Para resolver esses entraves, visando à qualidade dos dados que serão utilizados como entrada para o algoritmo de mineração de dados é necessário haver o pré-processamento; processo que contempla:

1. Limpeza dos dados. Correções de erros, substituição de valores nulos/faltantes, conversão de dados nominais e numéricos

2. Seleção. Escolher os atributos relevantes para a tarefa em questão.
3. Redução de Domínio. Simplificar a distribuição de valores no espaço possível;
4. Discretização os dados. Transformação de atributos contínuos em categóricos para melhorar a compreensão do conhecimento por parte do algoritmo de mineração utilizado.

#### 2.1.3. FORMATAÇÃO

Esta etapa é responsável pela tarefa de seleção dos dados, que são armazenados em um DW, Data Mart ou um banco de dados relacional. Para auxiliar a efetivação desta etapa, deve ser mantido um catálogo de metadados sobre as fontes de dados, assim como, sobre o que está no DW, Data Mart e Data Set. Há de se ressaltar que a realização desta etapa exige pleno conhecimento dos dados principalmente em relação à qualidade dos dados de origem e o formato dimensional escolhido para armazenar os dados.

#### 2.1.4. MINERAÇÃO DE DADOS

“Mineração de dados” é um termo mais utilizado por profissionais da área de estatística, analistas de dados e pela comunidade que desenvolve sistemas de informações gerenciais, enquanto KDD tem sido mais utilizado por pesquisadores em Inteligência Artificial e Aprendizagem de Máquina (FAYYAD *et al.*, 1996).

Fayyad *et al.* (1996) diferencia os termos mineração de dados e KDD destacando que o componente de mineração de dados se refere apenas ao meio pelo qual padrões são extraídos e enumerados a partir dos dados, enquanto que KDD envolve a avaliação e interpretação dos padrões para decidir o que é conhecimento e o que não é incluindo a escolha do esquema de codificação, pré-processamento, amostragem e projeções realizadas antes da etapa de mineração de dados, bem como o pós-processamento naturalmente realizado depois da etapa de mineração de dados.

As tarefas do KDD por muitas vezes são retratadas como sendo as tarefas de mineração de dados. A etapa de mineração de dados especificamente acontece após a etapa de formatação dos dados, onde os dados já estão estruturados e prontos para a execução da etapa de mineração de dados. A seguir, são descritas as principais tarefas de mineração que envolvidas no KDD (GOLDSCHMIDT, 2005):

- *Descrição (Description)*, É utilizada para descrever os padrões e tendências a partir dos dados. Essa tarefa geralmente apresenta uma interpretação para os dados e resultados obtidos. Essa tarefa é muito utilizada em conjunto a técnicas de análise exploratória de dados, que tem dentre os seus objetivos o de comprovar a influência de variáveis no resultado.

- **Classificação (*Classification*)**, Uma das tarefas mais comuns, a Classificação, objetiva identificar a classe de determinado registro. No seu desenvolvimento, o modelo extrai conhecimento do conjunto de registros fornecidos, com cada registro já contendo o rótulo da classe à qual pertence, por meio de uma abordagem de aprendizado supervisionado. Por exemplo, categorizamos cada registro de um conjunto de dados contendo as informações sobre os alunos de uma escola: Perfil Iniciante, Perfil Médio e Perfil Avançado. O modelo analisa os registros e então é capaz de dizer em qual categoria um novo aluno se encaixa. Essa tarefa de classificação pode ser usada também, por exemplo, para diagnosticar quando uma transação de cartão de crédito pode ser uma fraude ou diagnosticar onde uma determinada doença pode estar presente.
- **Estimação (*Estimation*) ou Regressão (*Regression*)**. A estimação é similar à classificação, porém é usada quando o registro é rotulado com um valor numérico e não um categórico. Assim, pode-se estimar o valor da variável-resposta analisando-se os valores das demais. Por exemplo, considerando um conjunto de registros contendo os valores de venda de casas em um determinado bairro ou região e suas características, o modelo é capaz de estimar qual será o valor de uma casa a ser vendida naquele bairro ou região.
- **Predição (*Prediction*)**. A tarefa de predição é similar às tarefas de classificação e estimação, no entanto ela objetiva “prever” o valor futuro de um determinado atributo, por exemplo, prever o valor de uma ação seis meses adiante, prever o percentual que será aumentado de consumo de um item de um mercado se a determinada característica for acrescida ou decrescida ou prever o vencedor de um jogo baseando-se em dados estatísticos dos jogadores.
- **Agrupamento (*Clustering*)**. A tarefa de agrupamento visa a agrupar registros similares. Um agrupamento (ou cluster) é um conjunto de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. Esta tarefa difere da classificação, pois não necessita que os registros sejam previamente categorizados, numa abordagem de aprendizado não-supervisionado. Além disso, ela não tem a pretensão de classificar, estimar ou prever o valor de uma variável, ela apenas identifica os grupos de dados similares, por exemplo, encontrar grupos de consumidores com comportamento similar data uma base de dados grande contendo seus bens e compras passadas.
- **Associação (*Association*)**, A tarefa de associação objetiva identificar quais atributos possuem relações entre si. Apresentam a forma: SE atributo X ENTÃO atributo Y. É uma das tarefas mais conhecidas devido aos bons resultados obtidos, principalmente nas análises da "Cestas de

Compras" (Market Basket), onde o objetivo é identificar quais produtos são levados em conjunto pelos consumidores. Permite, por exemplo, determinar os casos onde o uso de um medicamento está associado ao aparecimento de efeitos colaterais ou Identificar os usuários que podem responder bem a uma oferta de novos serviços.

#### 2.1.5. INTERPRETAÇÃO/AVALIAÇÃO

Após a fase de mineração de dados, o processo de KDD passa à fase de avaliação e interpretação dos resultados. Esta fase envolve todos os participantes do projeto de extração de conhecimento. O analista de dados é incumbido de interpretar se o classificador atingiu as expectativas, avaliando os resultados de acordo com algumas métricas tais como taxa de erro, tempo de CPU e complexidade do modelo. O especialista no domínio irá verificar a compatibilidade dos resultados com o conhecimento disponível do domínio. E, por fim, o usuário final (*stakeholder*) é responsável por dar o julgamento final sobre a aplicabilidade dos resultados.

Ao final das etapas do KDD o resultado do processo deve ser compreensível. Entretanto, definir compreensibilidade não é uma tarefa trivial. Em certos contextos, compreensibilidade pode ser estimada pela simplicidade do modelo (como, por exemplo, número de nós de uma árvore de decisão). De acordo com Craven & Shavlik (1995), compreensibilidade é útil para validar o conhecimento (o especialista deseja inspecionar o conhecimento para verificar se o conhecimento é confiável), para descoberta de novos padrões, para a sugestão de melhores atributos e para o refinamento do conhecimento. Como a compreensibilidade de um modelo frequentemente não pode ser facilmente estimada, uma segunda medida chamada interessabilidade (SILBERSCHATZ & TUZHILIN, 1995) é frequentemente utilizada pela comunidade de KDD. Interessabilidade mede o valor de um padrão combinando validade, novidade, utilidade e simplicidade.

Atualmente, as pesquisas que envolvem o uso do KDD como ferramentas para o descobrimento de novos padrões de dados, ou até mesmo para o auxílio na tomada de decisão em ambientes organizacionais, obtiveram um crescimento significativo. O produto esperado do processo de KDD é, de fato, informações úteis e interessantes a serem utilizadas, por tomadores de decisão.

## 2.2. DATA WAREHOUSE

*Data Warehouse* pode ser traduzido como Armazéns de Dados. Neste trabalho será utilizado o termo *Data Warehouse* ou apenas DW, dado que a maioria dos autores o utiliza por considerarem mais intuitivo. De acordo com (INMON, 2005), o termo se refere a "um depósito de dados orientado por assunto, integrado, não volátil,

variável com o tempo, para apoiar as decisões da gerência”, onde “não volátil” significa que, uma vez inseridos, os dados não podem ser alterados, embora possam ser excluídos.

Para Kimball et al. (1998), um DW apresenta como objetivo principal a integração de dados de diferentes fontes e formatos. Ele não é construído com o objetivo de suportar o processo funcional ou operacional da empresa, ou seja, o DW não é o fim, mas o meio para facilitar o uso da informação.

Embora o conceito de *Data Warehousing* seja recente, ele se baseia em idéias que vinham sendo aplicadas em vários sistemas de informação há muitos anos (INMON, 1997).

O DW é uma ferramenta que oferece aos usuários, o acesso a qualquer nível de informação, que possa apoiar na sobrevivência num mercado competitivo (SINGH, 2001). Ainda SINGH (2001), considera *Data Mart* como subconjunto do DW que desempenha o papel de um DW departamental, regional ou funcional, podendo-se construir uma série deles ao longo do tempo e eventualmente vinculá-los através de um DW lógico para formar uma estrutura única onde os dados são unificados obtendo uma visão de empresa completa.

Segundo Serra (2002), os *data marts* não são componentes obrigatórios dos DWs; eles podem ser utilizados como testes por empresas que ainda não utilizam DW mas que têm interesse em implantar, ou até mesmo substituir um DW por iniciais *data marts* nos casos em que a empresa não possui recursos para implantação do mesmo.

Contudo há de se ressaltar que há uma distinção clara entre os bancos de dados tradicionais e os DWs, uma vez possuem finalidades diferentes. Os bancos de dados tradicionais são transacionais (relacional, orientado a objetos, de rede, ou hierárquico), enquanto que os DWs têm intuito o direcionamento principal para aplicações de apoio às decisões. Assim, os DWs são construídos e pensados visando a otimizar a recuperação de dados, e não o processamento das transações do dia a dia. Além disso, os DWs proporcionam acesso aos dados para análise analítica, histórica, para a descoberta de conhecimento e tomada de decisão; eles fornecem aos gestores e usuários finais suporte às demandas de alto desempenho de dados e informações de uma organização (ELMASRI e NAVATHE, 2005). Na Tabela 1 (SINGH, 2001) são listadas algumas diferenças entre os dados transacionais e do DW.

Tabela 1 - Diferenças entre BD transacionais e Data Warehouse (adaptado de Singh, 2001)

Característica	BD Transacional	Data Warehouse
Objetivo	Operações diárias	Analisar o negócio
Uso	Operacional	Informativo
Tipo de Processamento	OLTP	OLAP
Unidade de Trabalho	Inclusão, alteração e exclusão	Carga e consulta
Número de usuários	Milhares	Centenas
Tipo de usuário	Operadores	Tomador de Decisão
Interação do usuário	Somente pré-definida	Pré-definida e <i>ad-hoc</i>
Condições dos dados	Dados Operacionais	Dados Analíticos
Volume	<i>Megabytes-Gigabytes</i>	<i>Gigabytes-Terabytes</i>
Histórico	60 a 90 dias	5 a 10 anos
Granularidade	Detalhados	Detalhados e resumidos
Redundância	Evita-se	Ocorre
Estrutura	Estática	Variável
Manutenção desejada	Mínima	Constante
Acesso a registros	Dezenas	Milhares
Atualização	Contínua (Tempo Real)	Periódica (em <i>Batch</i> )
Integridade	Transação	A cada atualização
Número de Índices	Poucos/Simples	Muitos/Complexos
Intenção dos índices	Localizar um registro	Aperfeiçoar consultas

Segundo (PONNIAH, 2001), o *Data Warehousing* não é um software ou produto de hardware que se adquire para fornecer informações estratégicas. É um ambiente computacional onde os usuários são inseridos de modo direto em contato com os dados de que necessitam para tomar as melhores decisões.

O produto principal obtido de um projeto de *Data Warehousing* é o seu DW, cujo objetivo principal e básico é gerar um repositório que contenha dados limpos, agregados e consolidados, que podem ser acessados por ferramentas geradoras de relatórios gerenciais, do tipo OLAP (*On-Line Analytical Processing*) e Data Mining.

Temos como benefícios de um DW sua modelagem, multidimensional, onde os dados são estruturados de forma simplificada, ficando compreensível para um usuário de negócio ou não. Em conjunto com as técnicas de acesso que as ferramentas de análise oferecem, o DW permite um desempenho adequado a grande volume de dados e tem como objetivo o auxílio na tomada de decisão.

#### 2.2.1. CARACTERÍSTICAS DE UM DATA WAREHOUSE

De acordo com (INMON, 2005), o DW deve seguir quatro propriedades fundamentais, são elas:

- Orientado por Temas: Os dados de DW são focados nos assuntos relevantes ao processo de tomada de decisão de uma organização, enquanto que os dados transacionais focam no que acontece no seu dia a dia.

Alguns exemplos de assuntos utilizados em projetos de DW são: produtos, atividades, contas, clientes, entre outros.

- Integrado: A integração do DW diz respeito à consistência de denominações, das unidades dos valores contidos nas tabelas e outras padronizações aplicadas aos dados no sentido de que esses sejam transformados até estarem em um estado uniforme. Isso é devido ao fato que um DW em geral é composto por diversas fontes de dados, para que essas fontes possam interagir é necessária uma representação única para todos esses tipos de dados.
- Variante no Tempo: Os dados contidos em um DW são temporais, por estarem relacionados a períodos de tempo bem definidos, o que auxilia na análise e na confirmação de acontecimentos temporais dentro de uma determinada atividade ou ramo de negócio. A variação no tempo significa que todos os dados no DW são precisos em algum instante no tempo, enquanto que, no ambiente operacional, os dados estão corretos no momento de acesso. Dessa forma, os dados do DW mostram uma imagem fiel da época em que foram gerados.
- Não Volátil: Como já citado anteriormente, os dados no DW não sofrem modificações, são sempre inseridos e nunca alterados. Em contrapartida, os dados de um banco de dados transacional, sofrem frequentes alterações. Ou seja, não deve ser um banco preparado para atualizações.

### 2.2.2. MODELAGEM DIMENSIONAL

Segundo Kimball (1998), a modelação multidimensional tem o propósito de guardar os dados num formato que facilite o seu acesso e compreensão por parte dos utilizadores finais.

De acordo com Song et al. (2001), o modelo dimensional apresenta duas vantagens principais na sua utilização em DW. A primeira vantagem, diz respeito ao armazenamento, por parte do modelo, de uma análise espacial multidimensional em ambientes de bases de dados relacionais, sendo analisados dados respectivos a fatos temporais utilizando dimensões. A segunda vantagem está no fato de que um modelo dimensional apresenta seus dados de forma desnormalizado, contendo uma apresentação e estrutura simples, trazendo assim, uma simplificação do processamento de consulta de usuários finais e a elevação do desempenho.

Modelagem dimensional é a técnica em que os dados e modelo proporcionam uma representação do banco de dados consistente com o modo que o usuário visualiza e navega pelo DW, combinando tabelas com dados históricos em séries temporais, cujo contexto é descrito através de tabelas chamadas de dimensões (BALLARD *et al.*, 1998; HARRISON, 1998).

Essas vantagens permitem que as ferramentas OLAP analisem e gerenciem dados, proporcionando aos tomadores de decisões um grande ganho de desempenho através do rápido acesso a uma grande variedade de visões de dados organizados através da base de dados multidimensional (MENOLLI, 2004).

Segundo Menolli (2004), um modelo dimensional apresenta três elementos principais:

- Fatos: Coleção de dados, sendo que cada fato representa um item de negócio, uma transação ou evento de negócio. Machado (2000) afirma que “fato é tudo aquilo que reflete a evolução dos negócios do dia-a-dia de uma organização”.
- Dimensões: elementos que estão presentes em um fato, como exemplo, tempo, localização ou cliente. Podendo ser organizadas de maneira hierárquica, compostas por vários níveis, por exemplo, uma dimensão chamada “região” pode ser composta dos níveis “região”, “estado” e “cidade”. Ballard (1998) afirma que “Dimensão é uma coleção de itens do mesmo tipo que representa as visões do negócio”.
- Medidas: atributos numéricos que representam um fato, uma vez que cada medida é composta pela combinação de dimensões que estão em um fato. Um exemplo para elucidar o elemento “Medidas” é a quantidade em valor monetário das vendas de um determinado vendedor em um mês. Segundo Ballard (1998) uma métrica “é definida como um atributo numérico de um fato, e representa o comportamento do negócio para as dimensões”.

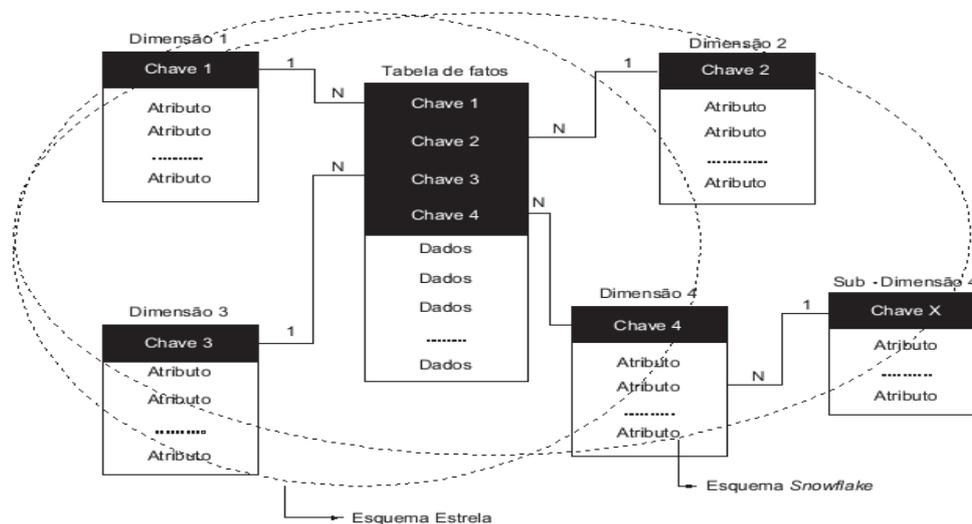
O modelo dimensional permite a visualização de dados na forma de um cubo, onde cada dimensão do cubo representa o contexto de um determinado fato e a intersecção entre as dimensões representa as medidas. Matematicamente, o cubo possui apenas três dimensões. Entretanto, no modelo dimensional, a metáfora do cubo pode possuir quantas dimensões forem necessárias para representar um determinado fato (MACHADO, 2000). Por este motivo este modelo também é chamado de modelo multidimensional.

Como principais técnicas de modelagem dimensional temos o modelo *Star-Schema* (Estrela) e o *Snowflake* (Floco de Neve). O mais comum é o modelo estrela, ele possui uma entidade central denominada tabela de fatos e um conjunto de entidades menores, as tabelas de dimensões, relacionadas com a tabela de fatos. Segundo Barquini (1996), as tabelas de fatos podem utilizar até 95% da área destinada ao DW.

Além do Star-Schema, outro modelo dimensional bastante difundido é o modelo Snowflake, sendo que este modelo pode ser conseguido pela decomposição de uma ou mais dimensões de maneira hierárquica (KIMBALL, 1998). É importante salientar que a diferença básica entre o modelo Star-Schema e Snowflake é que no último as dimensões são separadas em hierarquias, tendo uma tabela para cada nível, enquanto que no primeiro as mesmas dimensões com todos os níveis se dispõem em uma única tabela. A figura 2 (FORTULAN &

GONÇALVES, 2005) demonstra os modelos Estrela e Snowflake além de indicar a tabela de fatos e dimensões e subdimensões.

Figura 2 – Modelos Dimensionais, Esquema Estrela e Snowflake (Fortulan e Gonçalves, 2005)



Segundo Kimball (1998), no modelo estrela o tempo de acesso é mais eficiente do que no floco de neve, uma vez que, neste último, são necessárias mais junções entre as tabelas, portanto, um maior recurso temporal e processamento. Por outro lado, o modelo floco de neve é esteticamente mais semântico (melhor estruturado) para se visualizar se comparado ao modelo estrela, contudo deve-se salientar que a principal preocupação de um DW é realizar consultas com o máximo de eficiência possível.

Na modelagem dimensional existem operações que permitem a navegação entre os dados, a seguir estão descritas as operações básicas (MACHADO, 2000):

- *Drill Down*: navegação de um nível genérico para um nível mais detalhado.
- *Drill Up*: navegação de um nível detalhado para um nível mais genérico.
- *Drill Across*: combinação de várias tabelas de fato que compartilham as mesmas dimensões em um único relatório
- *Slice/Dice*: a operação *Slice* fatia o cubo para melhor visualização, por exemplo, onde diversos produtos são referenciados juntos, com o *Slice* seria possível visualizar separadamente, os tipos de produtos. Já

a operação *Dice* se refere à mudança de perspectiva de visão, ou seja, em um dado momento o foco é um determinado assunto, com o uso do *Dice* é possível mudar essa visão para outro assunto diferente, alterando o foco. Um exemplo, em um determinado momento é visualizado em um supermercado quantos produtos foram vendidos em certo período de tempo, com a operação *Dice* pode-se visualizar o número de produtos vendidos por esse supermercado, ou seja, o foco deixa de ser o supermercado e passa a ser a venda dos produtos;

- *Pivot*: alternância na ordem de apresentação das dimensões do cubo de dados, ou seja, alternar linhas e colunas.

### 2.2.3. GRANULARIDADE

A granularidade está ligada diretamente ao nível de detalhe dos dados que serão armazenados em um *data warehouse*, onde quanto maior a granularidade menor o nível de detalhe, portanto inversamente proporcional ao nível de detalhe. Ela desempenha um papel importante em um projeto de um DW, pois afeta o desempenho das consultas que serão feitas sob o DW. Nenhum outro aspecto de projeto terá importância se a granularidade não for conduzida de forma apropriada e se o particionamento não for cuidadosamente projetado e implementado (INMON, 1997).

O motivo pelo qual a granularidade é a principal questão de projetos de DW se dá no fato de que ela representa a profundidade dos dados e age diretamente relacionado ao volume de dados que será armazenado no DW, é com a definição do grão que a equipe vai definir o tipo de consulta que pode ser atendida. O volume de dados armazenado no DW deve ser equilibrado de acordo com o nível de detalhe de uma consulta.

### 2.2.4. ETL

O processo de ETL (extração, transformação e carga de dados) deve fazer parte de qualquer projeto de *data warehouse*. Essas etapas são responsáveis pela movimentação dos dados entre o ambiente transacional e o analítico com implementação de regras de transformação para garantir a consistência e qualidade dos dados. Esse processo pode ser desenvolvido manualmente (por código), entretanto atualmente existem diversas ferramentas de ETL que aceleram a construção e gerenciamento desse processo de forma transparente e eficaz.

O ETL pode ser comumente também aplicado em outras áreas de inteligência de negócios permitindo alimentação de dados para a análise de várias formas: Relatórios; *Dashboards* (painéis, *balanced scorecard*); Indicadores de Desempenho; Multidimensional (OLAP); Mineração de Dados (ATOL, 2007).

O grau de dificuldade, na integração dos dados entre estes ambientes - OLTP e DW - depende diretamente de como será o cenário de carga do DW e conseqüentemente do modelo dimensional e de como esses dados estão em seus sistemas de origem, que podem estar armazenados em esquemas comuns (homogêneos) ou em estruturas diferentes (heterogêneas).

### 2.3. O ANALYTICAL PROCESSING

Segundo Kimball *et al.* (1998), OLAP constitui-se das atividades gerais e específicas de consulta e apresentação de dados numéricos e textuais provenientes do DW. Os sistemas OLAP utilizam a visão multidimensional dos dados, sendo ela natural, fácil e intuitiva, permitindo a visualização dos dados em diferentes perspectivas de negócios de uma organização.

Inicialmente, surgiram as tecnologias que atendiam às necessidades das primárias de qualquer organização, operações transacionais, elas ficaram conhecidas como *On-Line Transaction Processing* (OLTP). Elas contemplam as transações diárias e visa o acesso aos registros de forma individual, ela tem como principais operações neste tipo de processo a alteração, inclusão, exclusão e consultas. Estas operações ocorrem muitas vezes em um mesmo dia e podem ser requisitadas ao sistema simultaneamente por muitos usuários, o que demanda uma resposta quase imediata do sistema. (AURÉLIO *et al.*, 2000).

As tecnologias *On-Line Analytical Processing* (OLAP), por sua vez, são projetadas para apoiar análises e consultas analíticas, elas auxiliam seus usuários a sintetizar informações através de comparações, visões e análises em um contexto histórico. As tecnologias OLAP têm como característica principal permitir uma visão mais fácil e intuitiva dos dados multidimensionais, por meio de análises em diferentes perspectivas (INMON, 2005).

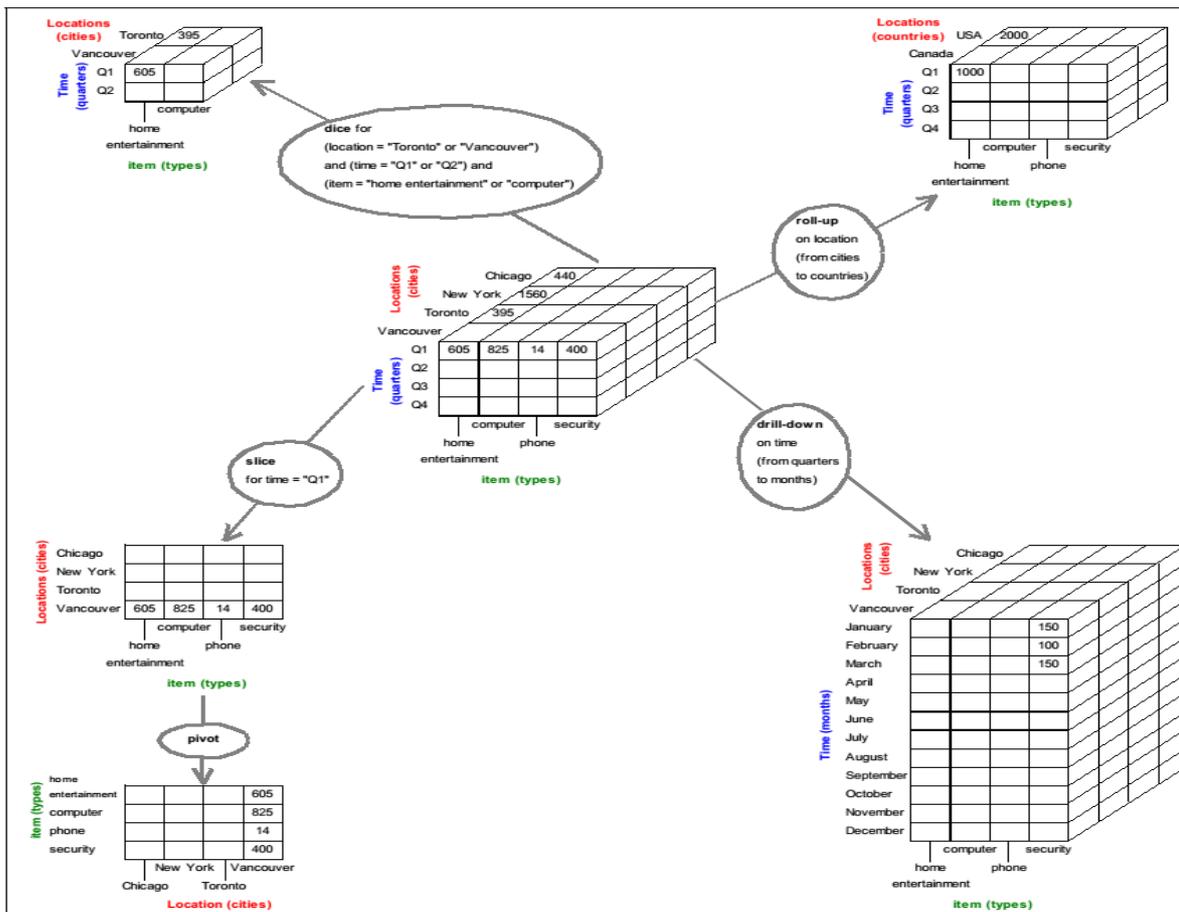
OLAP é uma categoria de software específica para realizar processamento analítico dos dados de um DW, de forma que este processamento deve: (1) ocorrer com alto desempenho e interatividade, e (2) auxiliar à tomada de decisão em uma organização, por meio da interpretação desses dados em uma variedade de visões multidimensionais.

De acordo com Han *et al.* (2006), OLAP faz parte do processo que habilita usuários a explorar os dados do DW, fornecendo funcionalidades para análise interativa de dados em diferentes dimensões e granularidades. As visões multidimensionais sobre os dados em OLAP provêm do fato de que questões típicas de análise de negócios em geral necessitam de uma visualização dos dados a partir de diferentes perspectivas, esse tipo de análise é o que os tomadores de decisões procuram, em uma empresa de varejo, por exemplo: "Qual a quantidade de itens vendidos, pontos de venda e quantidade de devoluções nas linhas de produtos, agrupados

pelos anos e meses, e filiais?”, ou ainda, “Quais os produtos mais vendidos em cada filial agrupado pela família de produto comparando com o mesmo período do ano passado?”.

Em uma consulta típica OLAP, se deseja obter informações, medidas de uma tabela de fatos, de várias perspectivas, dimensões, e em diferentes níveis de detalhe. Por exemplo, o valor de vendas pode ser analisado por filial, por produto e por mês. Um OLAP pode oferecer diversas funcionalidades para visualizar as informações de diferentes formas. A partir da figura 3 vamos realizar a análise de alguns conceitos que dão origem ao aspecto multidimensional.

Figura 3 - Cubo de dados de vendas por trimestre, local e tipo de item (Han e Kamber, 2001)



- Dimensões: São as diferentes perspectivas envolvidas, podem ser consideradas as características dos fatos envolvidos na entidade principal. Dimensões geralmente correspondem a campos não

numéricos em um *data warehouse* e fornecem informações descritivas. No caso da Figura 3, as dimensões são: Tempo (*Time*), Localização (*Location*), e Item (*Item*).

- **Medidas:** São as informações quantitativas que se deseja recuperar e analisar, isto é, campos numéricos da entidade principal, também podem ser chamadas de métricas. No caso da Figura 3, a medida/métrica é o total de vendas de produtos.
- **Cubos ou hipercubos:** Os dados que são extraídos do *data warehouse* e são organizados e armazenados em estruturas multidimensionais chamadas de cubo. No caso da Figura 3, temos um cubo com três dimensões: Tempo (*Time*), Localização (*Location*), e Item (*Item*), mas um cubo pode conter n dimensões, e é chamado de n-dimensional.
- **Membros:** São os elementos de uma dimensão. No caso da Figura 3, os membros da dimensão Item são *home entertainment*, *computer*, *phone* e *security*, enquanto que os membros da dimensão Localização (*Location*) são Vancouver, Toronto, New York e Chicago, os membros são os “valores” propriamente ditos das dimensões.
- **Hierarquias:** Os membros de uma dimensão são organizados (agregados) em níveis de hierarquias. Por exemplo, a dimensão Tempo (*Time*) da Figura 3, poderíamos organizar seus membros em níveis de granularidade, como Ano, Trimestre, Mês, Dia; ou seja, em uma hierarquia temporal. Em uma hierarquia, a granularidade de um membro de nível inferior é sempre menor do que a de um nível superior. Por exemplo, a granularidade de Dia é menor que a de Ano. A disposição dos membros de uma dimensão, nos níveis de uma hierarquia desta dimensão, deve ser feita respeitando-se o grau de agregação dos mesmos, pois um membro de menor granularidade sempre deve estar imediatamente abaixo do seu membro de granularidade maior seguindo-se a hierarquia.
- **Funções de Agregação:** Avaliam-se agregações das medidas segundo as dimensões do cubo por meio de funções de agregação. Alguns exemplos de funções de agregação são: soma, média, desvio padrão, mínimo, máximo. No caso da Figura 3, a função de agregação utilizada é soma, onde são apresentadas as somas dos totais das vendas segundo as dimensões, isto é, a soma de vendas de computadores (*computer*), no primeiro trimestre (Q1), na cidade de Chicago, por exemplo.

O OLAP também se caracteriza por possuir uma série de operações específicas para a manipulação/navegação dos dados multidimensionais. A seguir são discutidas brevemente algumas das operações mais populares, ilustradas com exemplos do cubo de vendas da Figura 3:

- *Drill-down*: desagregação/detalhamento dos dados para um nível com menor granularidade. Supondo-se que exista a hierarquia Tempo, definida como Ano→Trimestre→Mês→Dia no exemplo da figura 3, então, aplicando-se a operação *drill-down* sobre a dimensão Tempo, obter-se-ia uma nova visualização dos dados, com a dimensão Tempo disposta segundo o mês do ano da compra, conforme pode ser observado. Neste caso, mês (*months*) é um nível imediatamente abaixo de trimestre (*quarters*), na hierarquia definida de tempo.
- *Roll-up* ou *Drill-up*: agregação dos dados para um nível com maior granularidade. A operação de *roll-up* é exatamente o inverso da operação de *drill-down*, isto é, aplicando-se operações de *roll-up*, partindo-se de visualizações de dados em níveis de granularidade inferiores, chega-se a níveis de granularidade superiores. Na Figura 3, realizando-se a operação de *roll-up* na dimensão Localização (*Location*), que se encontra no nível de hierarquia Cidades (*cities*), esta passaria para o nível imediatamente acima, que neste caso é representado por Países (*countries*).
- *Slice/Dice*: seleção de parte de um cubo (fatiamento do cubo). Por exemplo, na Figura 3, uma operação de *slice* pode ser aplicada ao cubo, de modo que sejam selecionadas as células do cubo que fazem parte da dimensão Localização e Item para o primeiro trimestre (Q1) (fatia do cubo). A operação *dice* define um subcubo pela seleção de duas ou mais dimensões.
- *Pivoting/Rotate*: inversão/rotação dos eixos do cubo para visualização dos resultados de uma consulta. Na Figura 3, o cubo está sendo visualizado, dispondo-se a dimensão Item no eixo x e a dimensão Localização no eixo y. Por meio da operação de *pivoting* pode-se mudar a perspectiva de visualização, invertendo-se os eixos, isto é, dispondo-se a dimensão Localização no eixo x e a dimensão Item no eixo y, por exemplo.
- *Drill-across* - desagregação/detalhamento dos dados através de múltiplos níveis de dimensões diferentes. Por exemplo, na Figura 3, uma operação de *drill-across* aplicada ao cubo Países (*countries*) x Trimestres (*quarters*) x Tipos (*types*), poderia levar diretamente ao cubo Cidades (*cities*) x Meses (*months*) x Tipos sem passar pelo cubo intermediário Cidades x Trimestres x Tipos.

Apesar de aplicações de OLAP apresentarem os dados em visões lógicas multidimensionais, estes não necessariamente estão armazenados fisicamente em estruturas multidimensionais (banco de dados multidimensionais). Estruturas relacionais podem ser usadas para a representação e armazenamento de dados multidimensionais. De acordo com a forma como os dados estão armazenados no *data warehouse* que dá suporte aos sistemas OLAP, três tipos de arquitetura caracterizam as ferramentas OLAP (CHAUDHURI; DAYAL, 1996; PENDSE, 2000):

- ROLAP (*Relational OLAP*): Realiza seu processamento analítico em um DW com estrutura física relacional, e modelado dimensionalmente por meio de técnicas de modelagem chamadas Esquema Estrela ou Esquema Floco de Neve (KINBALL, 1996; INMON, 1996; CAMPOS; FILHO, 2000).
- MOLAP: (Multidimensional OLAP): Realiza seu processamento analítico em um DW cujo armazenamento físico usa tecnologia de banco de dados multidimensionais (GYSENS; LAKSHMANAN, 1997) com matrizes n dimensionais
- HOLAP (*Hybrid OLAP*)- Integra as características funcionais da ROLAP e MOLAP em uma única arquitetura híbrida. O armazenamento físico dos dados do DW é feito em tabelas relacionais, entretanto para implementar eficientemente as consultas, um cache dos níveis de agregação mais comuns é guardado na memória como uma matriz n dimensional.

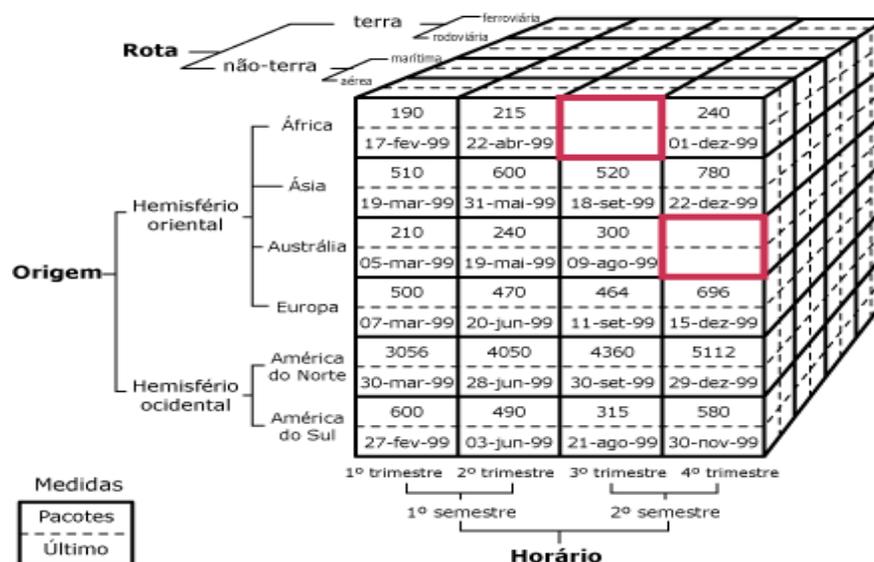
O ambiente OLAP permite aos usuários facilmente sumarizar e acessar dados, mas possuem algumas limitações na construção e manutenção de modelos analíticos complexos dos dados organizacionais. Para aumentar a sua eficácia, sistemas OLAP devem dar suporte a linguagens de consulta, possíveis em diferentes níveis de abstração (TORLONE; CABIBBO, 1998).

### 2.3.1. CUBO DE DADOS

A principal característica das tecnologias OLAP é permitir uma visão conceitual multidimensional dos dados. Um cubo de dados é uma estrutura que armazena os dados em formato dimensional. Uma dimensão é uma unidade de análise com dados agrupados.

Por exemplo, a dimensão tempo tem os dados agregados por meses, quadrimestres e semestres. A dimensão categoria tem os dados agregados em comercial, industrial, público e residencial, etc. A Figura 4 apresenta os dados modelados em um cubo dimensional, onde cada dimensão (Rota, Origem e Horário) representa um tema importante de uma empresa de transporte de encomendas para realização de análises e comparações. O cubo da Figura 4 é adaptado de (MICROSOFT, 2015) é referente a uma “Fato Envio de Pacotes”.

Figura 4 - Cubo de Dados (adaptado de Microsoft, 2015)



As tecnologias OLAP dispõem de operações para análise interativa de dados e com isso conseguem navegar em diferentes visões e granularidades, conseguindo então estruturar os dados de forma a visualizar as hierarquias e navegar pelas dimensões (THOMSEN, 2002).

As operações sobre os cubos de dados foram introduzidas por (GRAY et al., 1996) objetivando realizar múltiplas agregações em sistemas de Banco de Dados com suporte a OLAP. O operador Cubo é uma generalização n dimensional da operação *group by*, sendo capaz de executar diversos *group by* correspondentes a diversas combinações.

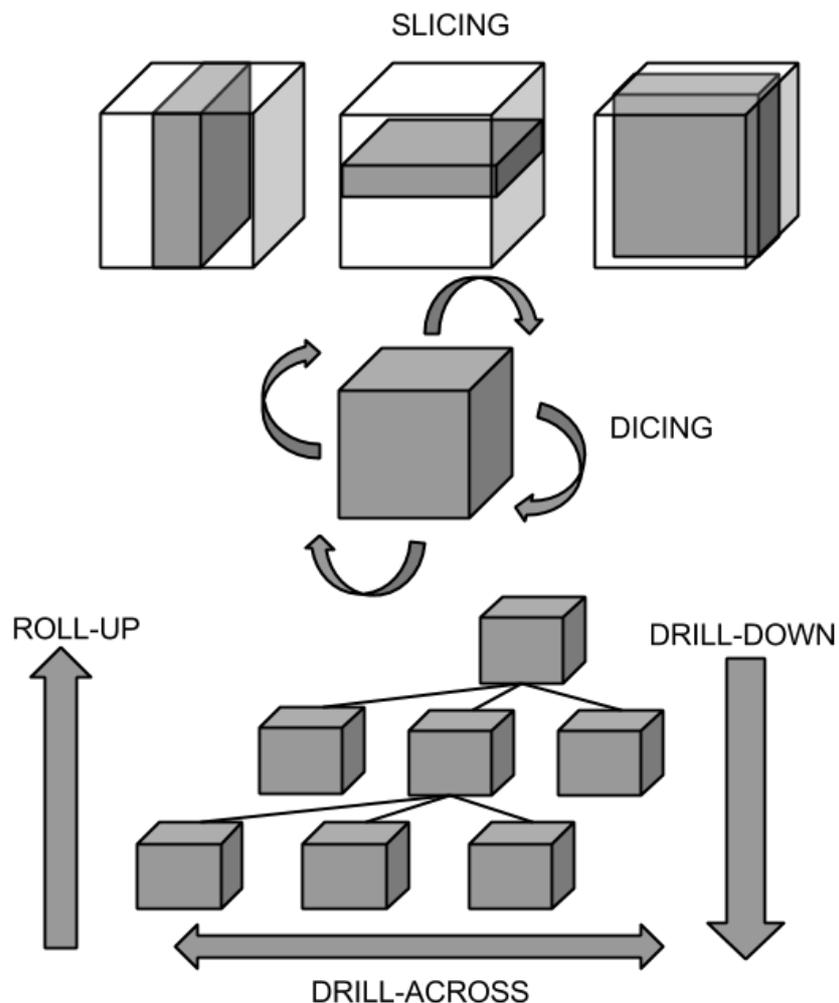
### 2.3.2. CONJUNTO DE OPERAÇÕES OLAP

Ao iniciar uma consulta a um DW é necessário traduzi-la de forma inteligível ao ambiente computacional. Assim, devem ser oferecidos aos analistas meios para realizar eficientemente uma consulta, a fim de obter resultados coerentes. Como solução, os desenvolvedores de ferramentas OLAP fornecem suporte para as operações de derivação de dados complexos.

Segundo (WREMBEL et al., 2007), o suporte às operações *Slice and Dice* é uma das principais características de uma ferramenta OLAP. A operação *Slice*, suportada pelas ferramentas OLAP, faz restrição de um valor ao longo de uma dimensão. Já a operação *Dice* é mais complexa, pois faz restrições de valores em nas dimensões.

O *Slice and Dice* compreende quatro operações, que são o *Ranging*, o *Drilling*, o *Rotation/Pivoting* e o *Ranking*. A Figura 5 ilustra de forma genérica a operação de *Slice*, *Dice*, *Drill-down* e *Drill-up/Roll-up*.

Figura 5 - Exemplo da operação slice, dice, drill-down e roll-up



A operação *Ranging* é responsável por, a qualquer momento, alterar o resultado das consultas, inserindo novas posições ou removendo as que estão em foco. Por exemplo, a inserção de um novo grupo de produtos em uma consulta representa uma operação de *Ranging*. O resultado dessa operação será considerado para todas as demais, e assim, pode-se encarar o resultado como um novo cubo gerado a partir do cubo original.

A operação *Drilling* consiste em escolher o foco no que se deseja analisar, o usuário ainda pode alterar o foco/escopo do que está analisando, além disso, os dados podem encontrar-se agregados em diversos níveis. A figura 5 ilustra a operação *Drilling*. O *Drilling* permite navegação por entre os níveis. Existem três operações OLAP que permitem mudar o escopo dos dados, são elas: *Drill-Down*, *Drill-Up* e *Drill-Across*.

A operação *Drill-Down* navega verticalmente na hierarquia, no sentido em que os dados são mais detalhados, em direção ao grão. Consiste em desagregar dimensões. A operação *Drill-Across* permite navegar entre os eixos da árvore hierárquica, ele é uma operação de grande utilidade, pois permite inserir e retirar posições do corrente cenário.

O *Drill-Up* ou *Roll-Up* faz parte da operação *Drilling* e realiza a função inversa do *Drill-Down*. Ela permite ao usuário uma visão mais agregada das informações. A operação *Rotation* ou *Pivot* além de permitir ao usuário inverter as posições das dimensões em foco, alterando linhas por colunas, com intuito de facilitar a compreensão ou dar uma outra visão aos dados.

Com a operação *Ranking* o analista pode filtrar as informações que se deseja obter. É possível fazer uma classificação dos dados adquiridos e operar diretamente sobre os valores das células. Todas as operações anteriores atuavam apenas sobre as posições ou dimensões dos dados, entretanto, através do *ranking*, o analista pode executar diversos tipos de filtros, eliminando assim os dados desnecessários e inconsistentes.

### 3 ONLINE ANALYTICAL MINING

Atualmente existe um campo emergente de pesquisa que se foca na integração das tecnologias de OLAP e Mineração de Dados. O termo Mineração Analítica On-Line (On-Line Analytical Mining – OLAM) é usado para se referir à integração destas duas tecnologias complementares em um ambiente de KDD foi introduzido por Han (1997).

Tipicamente, quanto mais tempo um DW está em uso, mais ele evoluirá (INMON, 1996). Esta evolução ocorre ao longo de um número de fases. Inicialmente, o DW é utilizado principalmente para a geração de relatórios e para responder consultas predefinidas. Inicialmente, é usado para analisar dados resumidos e detalhados, onde os resultados são apresentados na forma de relatórios e gráficos. Mais tarde, o DW é utilizado para fins estratégicos, realizando análise multidimensional e sofisticadas operações de *Slice and Dice*. Finalmente, o DW pode ser empregado na descoberta de conhecimento e tomada de decisão estratégica, utilizando ferramentas de Mineração de Dados. Neste contexto, as ferramentas para DW podem ser classificadas em ferramentas de acesso e recuperação, ferramentas de relatórios de banco de dados, ferramentas de análise de dados e ferramentas de Mineração de Dados (HAN & KAMBER, 2006).

Ao longo dos anos têm sido realizados diversos estudos nas áreas de OLAP e *data mining* com o objetivo de aperfeiçoar a exploração e análise de dados, e com isso a descoberta de padrões e relacionamentos entre os mesmos. Ainda que ambas as áreas forneçam boas ferramentas e soluções neste campo, apresentam algumas limitações. O OLAP tem poderosas ferramentas para organização e visualização de dados dentro do *data warehouse*, desde que esses dados sejam simples, pois não consegue lidar com dados complexos como imagens, texto, som ou vídeo. Por outro lado, as ferramentas de data mining não são adequadas para a organização e visualização, uma vez que são orientadas para esses fins. No entanto, estas técnicas apresentam bons resultados na análise e descoberta de conhecimento, tanto em dados simples como complexos (MESSAOUD et al., 2004; MESSAOUD et al., 2006).

As funções de mineração de dados são computacionalmente mais pesadas e complexas que as operações OLAP, logo uma eficiente integração das duas tecnologias pode se tornar um grande desafio. A mineração de conhecimento em bancos de dados multidimensionais é particularmente importante devido à complementaridade de OLAP e mineração de dados: enquanto OLAP fornece dados de alta qualidade (limpos, integrados, consistentes) necessários para a aplicação de ferramentas de análise e consultas OLAP, a mineração de dados, por sua vez, automatiza o processo de descoberta de padrões interessantes, análise e exploração do grande volume de dados disponíveis nos sistemas OLAP.

Segundo Kimbal (1997), enquanto OLAP é dedutivo e guiado por especialistas, Mineração de Dados é indutivo e guiado pelos próprios dados. Ambos necessitam de dados limpos e consistentes. E neste caso, o DW é capaz de fornecer dados para as duas tecnologias, o que o torna a principal fonte de dados para OLAM.

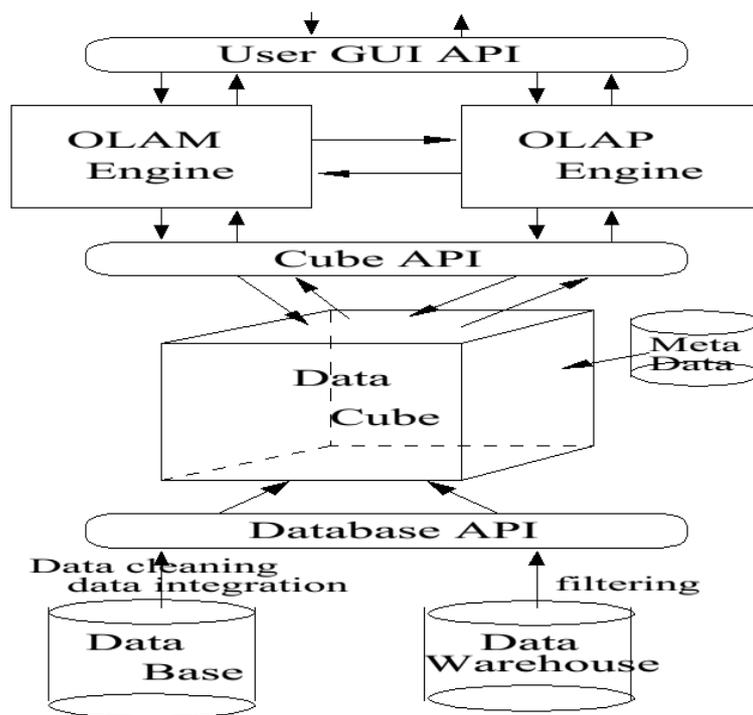
Segundo Han (1997), a mineração de estruturas multidimensionais é uma área ainda muito pouco explorada, mas que se poderá revelar bastante promissora devido principalmente às seguintes razões:

1. Em um DW os dados são integrados, consistentes e limpos. Características essenciais tanto para o OLAP como para a mineração de dados.
2. A integração permite aplicar técnicas de mineração de dados em diversos níveis de abstração e ainda a diferentes subconjuntos de dados, devido à sua utilização conjunta com técnicas analíticas como *slice*, *dice* e *pivoting*.
3. A interação entre utilizadores e sistemas torna-se mais fácil e flexível permitindo que estes selecionem e alterar funções de mineração dinamicamente

Segundo Han e Kamber (2006), Mineração de Dados frequentemente requer limpeza (data cleaning) e integração de dados (data integration). Eles ainda reportam que, a limpeza de dados é um importante problema para ambos os processos – DW e Mineração de Dados - visto que dados do mundo real tendem a ser incompletos e inconsistentes. Para Inmon (1996), a existência de um DW provê limpeza, integração e completude dos dados, permitindo que o processo de Mineração de Dados foque na sua principal tarefa: extrair conhecimento compreensível e útil.

Para Han et al. (2006), OLAM significa minerar interativamente em diferentes porções dos dados e em diferentes níveis de agregação, utilizando operações OLAP, podendo-se escolher as funções de Mineração de Dados e algoritmos dinamicamente, além de poder navegar pelos resultados da mineração. A arquitetura proposta por Han para os projetos de OLAM é apresentada pela figura 6, podemos analisar dois objetos para construção analítica uma para o OLAP Engine, capaz de realizar consultas OLAP, e outra para o OLAM Engine, provendo o conceito de mining para as consultas analíticas possuindo acesso ao cubo de dados e consequentemente ao DW.

Figura 6. Arquitetura OLAM (Han, Cheen e Chiang, 1998).



### 3.1. FERRAMENTAS OLAM ENCONTRADAS

Uma busca sistemática temporal orientada a ferramentas que possuem o conceito OLAM foi efetuada e foram encontrados os seguintes fatos:

- HAN (1997) explora as possíveis integrações entre OLAP e Data Mining e como podem dar ao usuário grande poder de exploração de dados, mostrando as vantagens de OLAM.
- TAN (1998) desenvolveu um sistema MOLAP próprio, no qual foram implementados algoritmos de mineração para dar suporte a OLAM, este sistema foi a base do módulo OLAM do DBMiner.
- HUA (1998) desenvolveu um módulo de mineração de regras de associação em cubos de dados, esse módulo foi implementado no DBMiner.
- Um classificador baseado em regras de associação é criado em LIU et al. (1998), onde um conjunto de regras são escolhidas baseadas em um suporte e confiança mínimos para então criar o classificador, essas regras são chamadas de regras de associação de classes

- CHEN (1999) fez um protótipo no qual o usuário explora exceções no cubo de dados, e minera regras de associações quantitativas. Neste trabalho o tratamento dos dados numéricos foi feito por inferência estatística.
- Em MAEDCHE et al. (2000) uma integração de OLAP e mineração de dados é realizada, em que dados de um cubo OLAP geram informações que alimentam um sistema de mineração de dados.
- Segundo KELKAR (2001) as áreas de integração de OLAP e Data Mining podem ser agrupadas como pré-processamento e pós-processamento do cubo.
- Em ORACLE (2005) um aplicativo que integra um algoritmo para importância de atributos é executado para fornecer as dimensões mais importantes para serem visualizadas em um ambiente OLAP.
- Em LIU et al. (2006) a análise das regras é feita usando operações OLAP, em um ambiente multidimensional.
- SHI et al. (2007) desenvolveu uma plataforma de mineração de dados para suporte a decisão chamada MSMiner. Neste sistema a integração dos processos é enfatizada e inclui mineração de dados e visualização OLAP. O sistema possui classificação, predição e clusterização, e é executado num ambiente ROLAP, baseado no esquema estrela.

As ferramentas OLAM encontradas se diferenciam por atuar em duas etapas as que são construídas em pré-processamento e em pós-processamento.

- Pré-processamento: A seleção de dimensões apropriadas para construção do cubo pode ser conduzida utilizando ferramentas de regras de associação e clusterização ou técnicas de componentes principais. Nesta etapa podem-se também aplicar algoritmos de Data Mining em dados gerados por outro algoritmo de Data Mining. Por exemplo, gerar clusters a partir de determinadas regras de associação encontradas e criar um cubo para cada cluster ou a criação de cubos de clientes com diferentes atributos onde cada cubo criado a partir de um cluster, com os atributos mais relevantes deste.
- Pós-processamento: Cubos possuem dados agregados em diferentes granularidades, logo diferentes algoritmos de Data Mining podem ser usados de maneira eficiente. Por exemplo, no caso de regras, estas requerem a contagem dos itens na busca das associações, o cubo possui as contagens pré-calculadas, acelerando assim o processo de busca. É possível ainda a partir de um subconjunto de dados visualizado no ambiente OLAP, submetê-lo a uma operação de Data Mining.

### **3.2. INTEGRAÇÃO PROPOSTA**

Este trabalho desenvolveu uma abordagem OLAM voltada em pós-processamento, nessa etapa o classificador foi treinado e testado. Após a fase de testes o resultado obtido, variáveis e escore de propensão são submetidos ao carregamento em um data warehouse.

A ferramenta tem o poder de avaliação do classificador a partir de consultas OLAP, garantindo a visualização dos dados em um ambiente com as características OLAP. Com a criação das medidas de avaliação de desempenho do classificador criadas em tempo de execução como permite uma consulta OLAP, garantindo que os subcubos gerados sejam avaliados em uma característica de um ambiente OLAM.

## 4 MODELAGEM

O capítulo apresentará as características do modelo que foi aplicado ao processo de OLAM, descrevendo o problema e os resultados.

### 4.1. CONTEXTUALIZAÇÃO

No Brasil, entre outros levantamentos sobre a educação, o INEP faz o Censo Escolar das escolas da educação básica e secundária e o Exame Nacional do Ensino Médio (ENEM) que avalia o desempenho do estudante ao fim da educação secundária. Em 2009, o ENEM passou a ser precisa fonte de informações sobre a educação secundária do país. Além do conhecimento técnico do aluno, o ENEM captura informações sócio-econômico-culturais sobre o seu perfil (INEP, 2012).

A pontuação das questões do Enem é feita por uma metodologia chamada Teoria da Resposta ao Item (TRI), que envolve psicologia, estatística e informática, e garante que todas as provas do Enem tenham o mesmo grau de dificuldade e possam ser comparadas (INEP, 2012).

A pontuação é diferente das avaliações clássicas, que somam os acertos. Na TRI, o item (a questão) mede níveis de conhecimentos diferentes. Há perguntas fáceis, medianas e difíceis, com pontuações diferentes. Além disso, esses itens têm de conseguir separar quem sabe o conteúdo de quem tenta acertar aleatoriamente. Pessoas que acertam o mesmo número de questões não têm notas iguais, porque acertaram e erraram questões diferentes.

Não há uma escala definida nas provas de múltipla escolha como há na redação. Em 2011, as notas máximas obtidas foram: Ciências Humanas, 793,1; Ciências da Natureza, 867,2; Linguagens, 795,5; e Matemática, 953.

A integração das informações do Censo Escolar com as do ENEM possibilita ao Governo Federal definir e validar as políticas públicas para a educação nacional (TRAVITZKI, 2013). O ENEM também é usado para gerar um ranking anual das escolas a partir do desempenho médio dos seus alunos nas provas das diversas áreas do conhecimento.

Este estudo demonstrará como desenvolver uma aplicação de OLAM para avaliar o desempenho de um classificador binário de uma solução de mineração de dados utilizada para identificar a qualidade da educação secundária privada no Brasil, a partir das bases do ENEM e do Censo Escolar do ano de 2011. As partes da solução de mineração de dados, como se chegou ao modelo computacional e os passos feitos para realizar o processo de construção da aplicação são descritos. Em seguida, detalha-se a criação do ambiente que resulta em uma ferramenta que aplica OLAM para avaliar o desempenho do classificador.

As bases do ENEM e Censo Escolar foram transformadas para o grão da decisão (escola) embutindo o conhecimento dos dados e especialistas, integradas e os formatos compatibilizados com a técnica de IA, regressão logística, que gerou o escore de propensão ao sucesso.

Para a análise OLAM, o desempenho do classificador foi avaliado por métricas calculadas em cada cubo de dados. Assim, as características/atributos das escolas e do classificador em segmentos ou subcubos em relação aos dados, podendo identificar tanto regiões onde o classificador possui um alto/baixo desempenho quanto onde as escolas são boas/fracas. Os resultados mostraram que essa abordagem de OLAM e consequentemente *domain-driven data mining* teve sucesso na validação de desempenho do classificador criado no estudo de caso.

#### **4.2. MEDIDA DE AVALIAÇÃO DE DESEMPENHO DO MODELO**

Para técnicas que produzem saídas contínuas, a decisão binária é tomada a partir de um limiar, abaixo do qual a decisão é feita para uma classe ou outra. Porém, a definição deste limiar é influenciada de acordo com o objetivo do problema, levando-se em consideração, geralmente, que os custos dos erros são diferentes para cada classe [Adeodato et al. 2004]. Para avaliar o desempenho do modelo desenvolvido será utilizada uma medida independente destes limiares. A métrica escolhida foi a Curva ROC (Receiver Operating Characteristics) (Provost & Fawcett, 1998;Fawcett, 2003).

A curva ROC é uma ferramenta poderosa para avaliação de classificadores binários de resposta contínua. A análise é feita por meio de um método gráfico simples e robusto, o qual permite estudar a variação da sensibilidade e especificidade do modelo, para diferentes valores de limiar de decisão (ponto de corte) (Provost & Fawcett, 1998;Fawcett, 2003).

As curvas ROC mostram a taxa verdadeiros positivos (VP) em função da taxas de falso positivos (FP), através da variação desse limiar. Esta relação prediz o comportamento dos classificadores independentemente dos custos e da distribuição das classes. Numa curva ROC, o eixo das ordenadas (y) representa VP e o eixo das abscissas (x) representa FP.

Para cada ponto de corte, a sensibilidade e o complemento da especificidade são calculados e colocados um em cada eixo de um gráfico bidimensional, produzindo a curva ROC. A área sob a curva (AUC) é a medida normalmente usada para comparar diferentes classificadores binários. A sensibilidade de um sistema de classificação é calculada a partir da razão entre os verdadeiros positivos sobre a soma dos verdadeiros positivos e falsos negativos, e a especificidade é calculada a partir da razão entre os verdadeiros negativos sobre a soma dos verdadeiros negativos e falsos positivos.

### 4.3. REGRESSÃO LOGÍSTICA

A etapa de modelagem computacional, regressão logística, deste trabalho teve como base os Microdados do ENEM 2011 que contêm informações sócio-econômico-culturais e de desempenho dos alunos ao final do ensino secundário, e o Censo Escolar 2011 que detalha as condições das escolas secundárias, da infraestrutura ao corpo docente. O conjunto de dicionário de dados utilizado no estudo de caso está nos anexos do trabalho. Este trabalho focará apenas nas escolas privadas.

Somente as escolas privadas foram selecionadas e as bases do ENEM e do Censo foram integradas pelo código único de identificação da escola. Anomalias que pudessem tanto desviar o foco quanto deteriorar a qualidade foram eliminadas. Os alunos que não realizaram todas as provas e os que não estavam concluindo o ensino médio foram descartados, bem como os atributos somente de escolas públicas (e.g. EJA). No grão escola, o trabalho preservou apenas as escolas privadas com 15 ou mais alunos para haver massa estatisticamente válida para as análises.

Para sistematizar este estudo, o objetivo foi definido como "binário". Nesse processo há dois pontos de alta relevância: 1) que métrica seria utilizada para avaliar a qualidade da escola e 2) que limiar seria adotado como critério para definir o que seria uma escola "boa". Mantendo o pragmatismo e isenção, este trabalho considerou a média aritmética das notas dos alunos como medida de qualidade das escolas, similar ao feito no ranking das escolas. A nota de cada aluno é a média aritmética das suas notas em cada prova.

Em seguida, definiu-se o quartil superior como limiar de binarização da nota para caracterizar o objetivo (escola forte ou fraca). Essa abordagem estatística evita polêmica e usa a robustez das separatrizes (quartis etc.) em relação a valores extremos (*outliers*), como discutido em (Adeodato 2015) e feito (Adeodato 2015 e SOUSA *et al.*, 2008) em outros domínios de aplicação. Assim, escola forte é aquela cuja média esteja no quartil superior das médias das escolas. A diferença na granularidade dos atributos aumenta a complexidade do trabalho. 1) Como atribuir uma renda familiar à escola a partir da renda familiar dos seus alunos? 2) Como atribuir uma formação docente à escola a partir da formação docente dos seus alunos? Esse aspecto dificulta equipes sem profissionais de mineração de dados desenvolverem projetos, tanto pelo grande volume de dados quanto pela necessidade de uso de inteligência artificial para embutir conhecimento dos especialistas em educação nas transformações dos atributos para mudança de grão, no processo *Domain-Driven Data Mining* (CAO, 2008).

Para evitar explosão combinatória, o estudo adotou a média e a moda para conjugar no grão escola os atributos numéricos e categóricos, respectivamente. Assim, as respostas às perguntas acima passam a ser 1) a

média da renda familiar dos alunos considerados e 2) a moda da formação docente dos professores considerados em cada escola.

Em geral, dados ausentes foram preenchidos pela média nos atributos numéricos e, nos categóricos, foi criada a categoria dado ausente, exceto em casos dependentes do domínio, em que havia relação de ordem, quando era preenchido por uma categoria extrema (e.g. educação dos pais ausente foi codificado como analfabeto). Nos numéricos, os *outliers* foram substituídos pelo extremo de três desvios-padrão.

Como fonte de dados, a base do ENEM (Contendo dados dos alunos e questionário socioeconômico) e Censo Escolar (contendo dados de professores e escolas). Para os oriundos da base do ENEM temos dados separados pelas categorias dados do inscrito, dados da escola, dados do município de aplicação da prova, dados da prova objetiva, dados da redação e dados de consistência com o censo da educação básica.

No final do processo de extração, transformação e carga, obtivemos uma base com 4803 escolas onde são registrado 24,98% escolas alvo. Na divisão para treinamento (75%) e teste (25%), ficamos com 3602 para o treinamento e 1201 para o teste garantindo a independência estatística dos data sets que serão submetidos aos processos. A primeira etapa após estar com as bases separadas, garantindo a independência estatística, foi extrair o conhecimento pelo modelo logístico na base de treinamento onde se encontra 91 variáveis explicativas, conforme apresentado na tabela 2.

**Tabela 2. Tabela de Variáveis**

Variáveis				
aluno_quantidade	m_q10	m_professores_idade	m_q1	m_q68
p_aluno_sexo_masculino	m_q12	qtd_mestrado	m_q3	m_q70
p_aluno_raca_branca	m_q14	qtd_nenhuma_pos	m_q5	m_q72
p_aluno_raca_parda	m_q16	p_professores_raca_n_decla	m_q7	qtd_especializacao
p_aluno_raca_indigena	m_q18	p_professores_raca_preta	m_q9	qtd_doutorado
uf_esc	m_q20	p_professores_raca_amarela	m_q11	p_professores_sexo_masculino
sala_diretoria	m_q22	total_professores	m_q13	rada p_professores_raca_branca
laboratorio_informatica	m_q24	aluno_media_idade	m_q15	p_professores_raca_parda
cozinha	m_q27	p_aluno_raca_n_informado	m_q17	p_professores_raca_indigena
sala_leitura	m_q30	p_aluno_raca_preta	m_q19	impressora
num_salas_existentes	m_q33	p_aluno_raca_amarela	m_q21	internet
videocassete	m_q58	p_alunos_necessidade_especial	m_q23	m_q8
parabolica	m_q60	localizacao	m_q26	m_q75
retroprojektor	m_q62	sala_professor	m_q28	Copiadora
num_computadores	m_q64	laboratorio_ciencias	m_q31	m_q65
num_funcionarios	m_q66	biblioteca	m_q57	
m_q2	m_q69	parque_infantil	m_q59	
m_q4	m_q71	tv	m_q61	
m_q6	m_q73	dvd	m_q63	

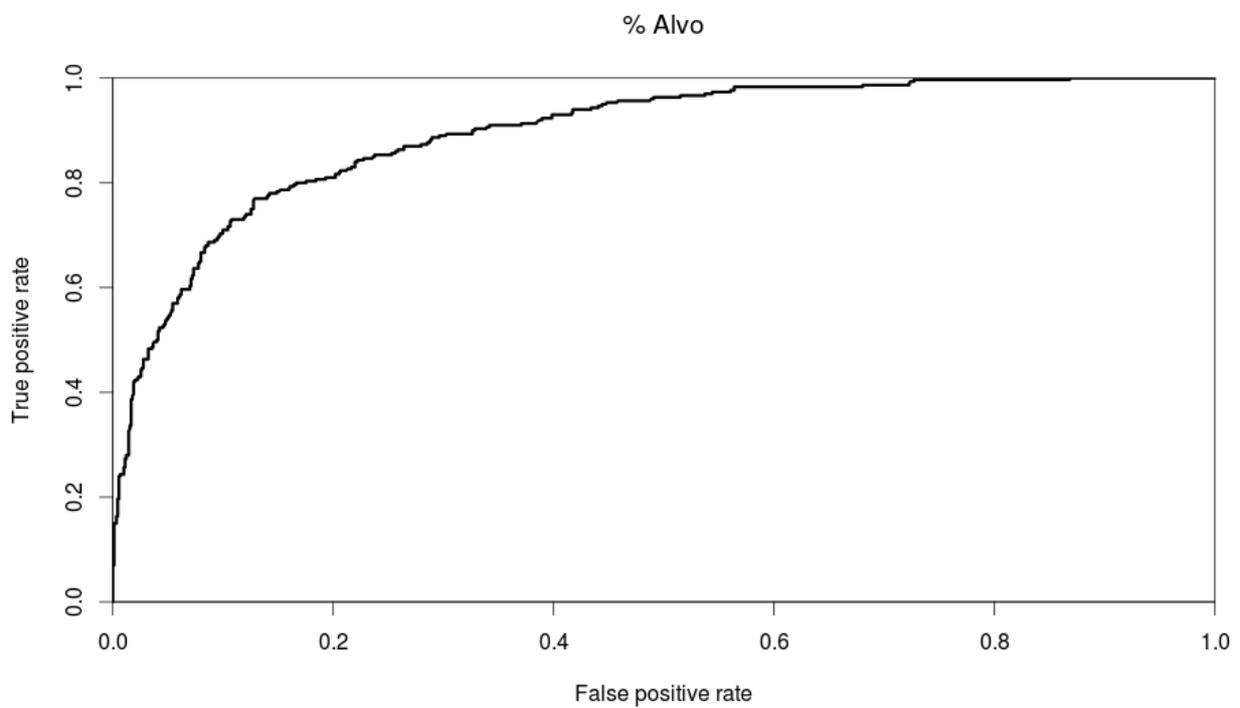
O modelo gerado detém 91 variáveis em sua instância original, para conseguir um modelo efetivo com um número menor de variáveis foi executado um processo de redução de modelo utilizando o teste estatístico Wald. Nesse teste cada coeficiente é avaliado usando o um teste de escore z, que é um teste individual aplicado a cada variável, com esse processo obtivemos 15 variáveis como resultado, conforme apresentado na tabela 3.

**Tabela 3. Modelo Reduzido**

Variáveis	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	3.683	2.412	1.527	0.13
aluno_quantidade	0.004	0.001	3.671	0.00
aluno_media_idade	-0.451	0.133	-3.399	0.00
p_aluno_raca_branca	0.010	0.003	3.325	0.00
p_alunos_necessidade_especial	-0.062	0.024	-2.622	0.01
laboratorio_informatica	-0.423	0.137	-3.081	0.00
num_salas_existentes	-0.014	0.005	-2.698	0.01
num_computadores	-0.002	0.001	-2.635	0.01
num_funcionarios	0.009	0.002	6.062	0.00
m_q28	-2.052	0.884	-2.322	0.02
m_q58	-1.269	0.109	-11.613	0.00
m_q68	0.360	0.138	2.601	0.01
m_q70	0.628	0.232	2.712	0.01
m_q75	1.626	0.158	10.313	0.00
m_professores_idade	0.070	0.011	6.125	0.00
qtd_mestrado	0.117	0.032	3.683	0.00

O modelo reduzido foi então submetido ao processo de avaliação de desempenho utilizando a técnica de área sob a curva ROC, o resultado da técnica foi de 0,84 a figura abaixo demonstra bem o comportamento da distribuição.

Figura 7. Curva ROC Modelo Geral



No final do processo temos a base de variáveis com o escore de propensão criado pelo classificador, esse data set gerado é o insumo necessário para carregar o modelo dimensional no processo de produção do DW, sendo carregadas as tabelas de dimensões e a tabela de fatos, para posterior análise do cubo de dados de desempenho do classificador, obtendo uma ferramenta que fornece capacidade OLAP sob um processo de avaliação do classificador.

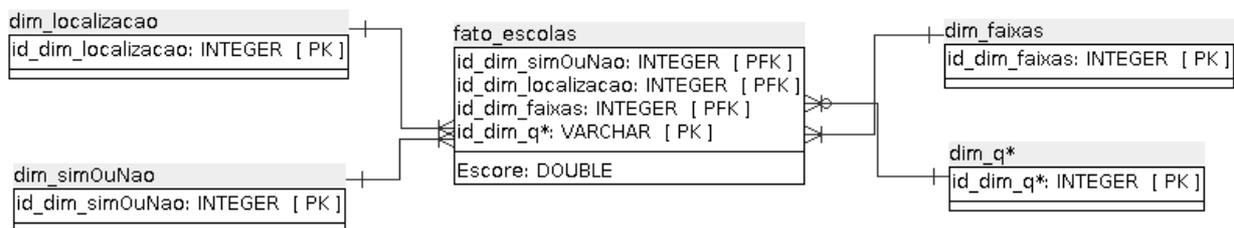
## 5 CONTRIBUIÇÕES

Esse capítulo apresenta as contribuições do trabalho, na visão de OLAM e como foram obtidos os resultados a partir de alterações no Data Warehouse e consultas OLAP.

### 5.1. DATA WAREHOUSE

Com a base de teste já com o escore de propensão, foi estruturado um modelo dimensional onde o escore de propensão fosse a medida de qualidade de cada escola a ser utilizada no modelo OLAP, apresentado na figura 8.

Figura 8. Modelo Dimensional



Esse modelo dimensional, formato estrela, apresenta as características das escolas obtidas nas na junção dos arquivos base, separando as dimensões em 4 tabelas físicas e 84 dimensões lógicas. Na tabela de fatos, temos a presença do escore de propensão para cada escola.

As Dimensões são os descritores dos dados oriundos dos fatos. Elas possuem o caráter qualitativo da informação e relacionamento de “um para muitos” com a tabela fato. São as dimensões que permitem a visualização das informações por diversos aspectos e perspectivas.

As dimensões oferecem características intrínsecas ao problema, para a abordagem OLAM ser efetiva em análise do desempenho do modelo nos dados. Dando possibilidade para serem usadas em um modelo n dimensional onde sendo utilizadas em conjunto podem responder questões como, por exemplo, “Qual a foi o desempenho do classificador em escolas das regiões Norte e Sul, que possuem sala de informática e faixa de quantidade de alunos vai de 10 a 40?”, “Qual o desempenho do classificador em escolas que possuem laboratório de informática, estão na região nordeste e a escolaridade do pai é superior ao ensino médio”.

Esse modelo de dimensão *versus* métrica para análise de desempenho está analisando a AUC ROC de cada conjunto de dimensões selecionadas na consulta OLAP em relação ao classificador, e consequentemente

subcubo gerado a partir das análises. Agregando então em um modelo OLAP uma visão de desempenho a partir da saída da mineração de dados.

## **5.2. MODELO DIMENSIONAL**

Foi criado um modelo dimensional em uma arquitetura ROLAP, onde é necessário um arquivo de entrada com uma estrutura de mapeamento dos dados físicos (5 tabelas, 4 referentes a dimensões e um referente a fato) em relação aos dados lógicos no nosso caso, contendo 86 elementos, 84 dimensões e 6 medidas. As dimensões contemplam a localização das escolas em hierarquia (Região, Estado e se é Rural ou Urbana), as questões binárias do quadro socioeconômico, as faixas das questões quantitativas e as questões categóricas do quadro socioeconômico.

A tabela de fatos contém apenas as dimensões e o escore de propensão para ser utilizado como medida geral em cada subcubo em relação ao classificador, além calcular o escore médio, escore máximo, escore mínimo, escore mediano e a área sob a curva ROC como medida de desempenho e o lift do conjunto.

A possibilidade de gerar a área sob a curva ROC em tempo de execução para cada subcubo utilizando uma abordagem MDX foi ponto chave no processo de transformar uma ferramenta tipicamente OLAP em uma ferramenta OLAM. Essa possibilidade tornou possível avaliar o classificador, podendo mesclar as características dos elementos avaliados (dimensões), em relação à(s) medida(s) que caracterizam o desempenho do modelo, especificamente a área sob a curva ROC.

## **5.3. MEDIDAS**

Apenas uma medida foi persistida fisicamente no data warehouse, o escore de propensão de cada registro obtido na etapa de teste, ele dá a base necessária para o cálculo de todas as outras medidas necessárias para os resultados esperados. As medidas então foram criadas utilizando uma técnica de medidas calculadas simples, cálculos de mínimo, máximo, média dentre outros, e por MDX, membros onde o cálculo envolve o conjunto de outros membros ou elaboração de conjuntos.

Então assim os cálculos são feitos em tempo de execução no momento da manipulação das dimensões acontece à criação das medidas, foram criadas utilizando membros calculados simples o escore médio, escore máximo, escore mínimo e o lift. A área sob a curva ROC pelo fato de manipular conjuntos e ter a necessidade de envolver campos auxiliares foi necessária à elaboração de mdx para a criação da medida em tempo de execução.

A figura 9 representa a criação do escore máximo, o processo para criação das demais medidas simples se dá da mesma forma, alterando apenas o tipo de agregação, a coluna selecionada “escore” representa o campo que o engine OLAP utilizará para criar a medida em tempo de execução utilizando a agregação do tipo “max”, então ele retornará o valor máximo do escore, do conjunto selecionado pelas dimensões características.

**Figura 9. Criação da Medida Escore Máximo**

Attribute	Value
name	Score Max
description	
aggregator	max
column	escore
formatString	
datatype	
formatter	
caption	
visible	<input type="checkbox"/>

A medida calculada por mdx é manipulada como um membro calculado, a exemplo da figura 10, o campo “formula” recebe a função que criará dinamicamente a medida, sendo inserida no conjunto de Measures disponíveis no cubo e sendo chamada de “auc roc”, ela será executada utilizando o conjunto filtrado pelas dimensões no ato da consulta.

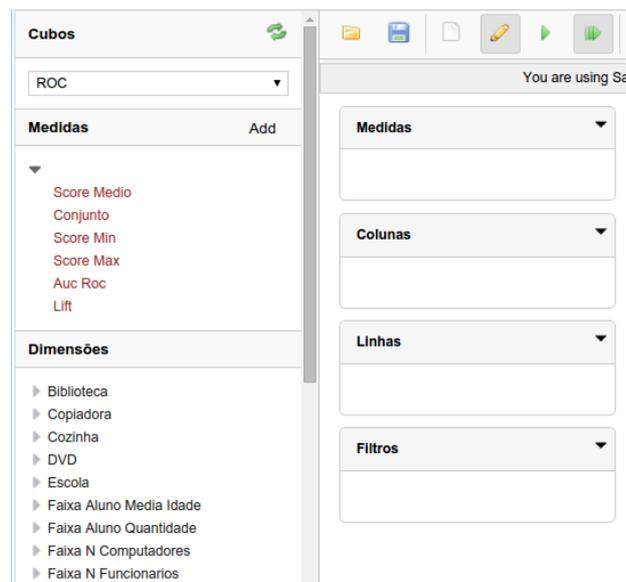
**Figura 10. Criação de Medida a partir de MDX**

Attribute	Value
name	auc roc
description	
caption	
dimension	Measures
hierarchy	
parent	
visible	<input checked="" type="checkbox"/>
formula   formulaElem...	WITH /* setting data row */ SET [linhas] AS nonEmptyCrossjoin(...
formatString	

## 5.4. RESULTADOS

A exploração dos resultados é dada pela execução das análises inferindo subcubos a partir da escolha das dimensões. A figura 11 apresenta o modelo de seleção das dimensões e medidas que podem ser utilizadas na análise, dando a possibilidade de escolher as características e medidas a serem utilizadas na análise e qual posição ela tomará no resultado, linha ou coluna.

Figura 11. Visão Geral Seleção Dimensões e Medidas



Os experimentos foram realizados em uma máquina com 16GB de memória RAM, Processador i7 com 8 núcleos com 2.20GHz e um disco rígido SSD de 120GB. Foi utilizado o PostgreSQL 9.3.10 exercendo o papel do SGBD que armazena o *data warehouse* em um sistema operacional Ubuntu AMD64. O engine OLAP utilizado foi o Mondrian 3.11 oferecendo a capacidade OLAP ao modelo dimensional apresentado pelo PostgreSQL.

Nos teste iniciais o tempo de execução das consultas de forma completa (calculando todo o conjunto de subcubos disponíveis), esgotava tempo de execução *default* do engine OLAP, 300 segundos. Dificuldade essa dada pelo nível de processamento de conjuntos (dimensões) avaliado nas consultas e pelo limitante da capacidade do engine OLAP a realizar cálculos interativos sob o mesmo conjunto, por exemplo, uma iteração sob um conjunto buscando um valor agrupado acumulativamente em relação à outra variável não é bem aceita pelas funções básicas oferecidas pelos engine OLAP dado que as consultas são criadas levando em consideração as hierarquias e níveis de relacionamento das dimensões propostas.

O primeiro passo foi alocar recurso temporal para validar a conclusão da análise e criação do resultado, alcançando 3000 segundos, fazendo com que a capacidade de hardware fosse extrapolada sem a resolução do cálculo.

Segundo um estudo feito sob a ótica de avaliação utilizando replica parcial (PAES; LIMA; MATTOSO, 2009), podemos particionar as consultas para obtendo um resultado eficaz sem oferecer perda no resultado, o foco principal é obter um desempenho maior nas consultas e processamento da OLAP, dividindo a consulta principal em n subconsultas de acordo com a consulta principal, no caso do trabalho quebrando as consultas nos subcubos em n subconsultas de acordo com o número de categorias avaliadas.

Além disso, foram removidas as consultas que eram oriundas de subcubos com menos de 30 registros em sua amostra devido à baixa amostragem e irrelevância estatística do subcubo com tais características. Como objeto resultante das consultas temos os campos medidas como escore mínimo, escore máximo, escore mediano, escore médio, lift, área sob a curva ROC, n (quantidade de registros no conjunto avaliado) e o tempo em segundos que foi utilizado na construção de cada subcubo avaliado em relação às categorias, construído em sua consulta OLAP.

A tabela 4 representa a variável “Faixa Quantidade de Alunos”, ela obteve sete subcubos válidos para avaliação de desempenho, subcubos com n maior do que 30, do classificador. Onde o melhor desempenho, ou seja, conseguiu maior assertividade e separabilidade sob o alvo, foi o de categoria “61\_70” que representa a faixa das escolas de 61 a 70 alunos na sala de aula, com uma área sob a curva ROC de 0,9637 a maior dentre os subcubos gerados pelas categorias para essa variável, essa categoria possui uma amostra de 63 unidades e foi a que obteve o maior escore e seu lift foi de 1,33.

**Tabela 4. Consulta variável Faixa Quantidade de Alunos**

Faixa Aluno Quantidade	ESCORE_MIN	ESCORE_MAX	ESCORE_MEDIANO	ESCORE_MEDIO	LIFT	AUC_ROC	N	TEMPO (s)
11_20	0,0051	84,3517	9,3398	18,2157	0,5323	0,8387	263	40,48
21_30	0,0172	94,7164	7,9496	18,3475	0,5976	0,8242	328	121,53
31_40	0,0022	89,3780	12,5179	23,9851	0,9718	0,8798	177	41,66
41_50	0,0019	96,5139	22,6997	32,3714	1,3667	0,9079	120	24,83
51_60	0,1101	83,7631	24,0039	31,4639	1,6623	0,8965	77	10,55
61_70	0,0867	99,0283	12,2522	33,9043	1,3333	0,9637	63	7,22
71_80	0,8037	89,8307	26,7408	34,5340	1,4444	0,9397	36	2,68

Para avaliação dos resultados foram avaliadas quatro variáveis, geradas suas subconsultas para realização do experimento e apresentação dos resultados.

Foram escolhidas duas variáveis que representam características da escola e duas variáveis que representam características dos alunos são elas respectivamente, Faixa Quantidade de Alunos, Faixa Número de Funcionários da Escola, Quantidade de Banheiros na casa dos Alunos e Escolaridade da mãe dos Alunos.

A tabela 5 apresenta a análise da variável “Faixa Número de Funcionários”, que representa a quantidade de funcionários que cada escola possui, para essa variável foi obtido dez subcubos válidos para a análise, o subcubo que alcançou melhor desempenho do classificador foi o representado pela categoria “11\_20” com uma área sob a curva ROC de 0,9416 e um lift de 0,4706. O subcubo onde o classificador obteve o pior desempenho foi o de categoria “21\_30” que representa as escolas que possuem de 21 a 30 funcionários na escola, obtendo uma área sob a curva ROC de 0,7502 e esse subcubo obteve um escore máximo com valor de 69,5996.

**Tabela 5. Consulta variável Faixa Número de Funcionários**

Faixa Numero Funcionarios	ESCORE_MIN	ESCORE_MAX	ESCORE_MEDIANO	ESCORE_MEDIO	LIFT	AUC_ROC	N	TEMPO (s)
11_20	0,1256	29,2354	5,6148	8,4952	0,4706	0,9416	34	1,54
21_30	0,0596	69,5996	6,4609	14,6598	0,4681	0,7502	94	9,9
31_40	0,0019	76,7172	6,2119	14,4074	0,3908	0,8851	174	32,69
41_50	0,0022	85,8155	6,8308	16,4494	0,7117	0,9207	163	28,89
51_60	0,1477	89,0583	9,2178	18,2363	0,6222	0,8641	180	39,25
61_70	0,1439	84,3517	9,2769	19,9360	0,5234	0,8625	107	15,47
71_80	1,6274	91,5545	14,7834	25,3294	0,9275	0,7912	69	6,65
81_90	1,4146	92,6942	23,7074	34,4457	1,3000	0,7816	80	10,07
91_100	1,5476	80,5908	25,7612	32,9885	1,2881	0,7684	59	5,74
100_ACIMA	0,0051	99,2295	62,7256	52,6319	2,2573	0,9046	241	120,43

O questionário socioeconômico utilizado na formação das variáveis dos data sets apresentam características dos alunos em relação às escolas onde podemos, por exemplo, avaliar o classificador e seu desempenho em relação aos subcubos gerados sob a ótica dos alunos.

A variável “Quantidade de Banheiros na Residência”, apresenta uma visão da estrutura da residência dos alunos, ela obteve três subcubos válidos para análise e é demonstrada pela tabela 6. O subcubo onde o classificador obteve o melhor desempenho foi o representado pela categoria “1” banheiro na casa com uma área sob a curva roc de 0,8345 e lift de 0,1735 contendo 461 registros no subcubo, o conjunto possui escore máximo de 64,1784, um escore médio de 5,3488 e um escore mediano de 3,3163. O subcubo que obteve o pior desempenho do classificador foi o da categoria “2” com uma área sob a curva roc de 0,7539 com um lift de 0,7363, esse subcubo possui um escore máximo de 99,2295 e um escore médio de 61,8122.

**Tabela 6. Consulta variável Quantidade de Banheiros na Residência**

Q75	ESCORE_MIN	ESCORE_MAX	ESCORE_MEDIANO	ESCORE_MEDIO	LIFT	AUC_ROC	N	Tempo (s)
1	0,0019	64,1784	3,3163	5,3488	0,1735	0,8345	461	153,23
2	0,9483	92,0327	14,0374	20,0158	0,7363	0,7539	402	234,36
3 ou Mais	8,8893	99,2295	67,5329	61,8122	2,4379	0,8182	338	140,36

E a última variável avaliada foi a “Escolaridade da mãe”, ela obteve três subcubos válidos, a categoria que obteve o maior desempenho no classificador foi o de categoria “Ensino Superior” com uma área sob a curva ROC de 0,8697 e um lift de 1,8850, esse conjunto teve um escore máximo de 99,22, um escore mediano de 45,34 e um escore médio de 44,69. O pior desempenho do classificador foi obtivo no subcubo que representa a categoria “Pos Graduacao” com uma área sob a curva ROC de 0,7453 e um lift de 1,76.

**Tabela 7. Consulta variável Escolaridade da Mãe**

Q3	ESCORE_MIN	ESCORE_MAX	ESCORE_MEDIANO	ESCORE_MEDIO	LIFT	AUC_ROC	N	Tempo (s)
Ensino Medio	0,0172	87,2206	6,3850	12,4425	0,3185	0,8147	628	427
Ensino Superior	1,1752	99,2295	45,3491	44,6905	1,8850	0,8697	452	400,5
Pos Graduacao	1,3079	94,1452	39,3748	42,2336	1,7600	0,7453	75	10,2

Durante o processo de consultas OLAM das variáveis escolhidas é perceptível à dinamicidade e a facilidade em encontrar os principais pontos onde o classificador obteve o melhor e o pior desempenho e fazer avaliações dos subcubos que estavam em foco na consulta, utilizando as medidas disponíveis nas consultas para avaliar possíveis correlações do desempenho do subcubo com o conjunto avaliado de dados avaliado, onde cada medida tem sua responsabilidade na distribuição de conhecimento do subcubo avaliado.

Para execução das demais consultas, nas demais variáveis, é necessária a execução da separação das consultas por categorias válidas, sendo um processo trivial e dinâmico, oferecendo uma praticidade ao tomador de decisão para avaliação do classificador. Continuando então com a flexibilidade que as ferramentas OLAP precisam ter sob os dados que estão sendo avaliados para garantir avaliação do classificador.

## 6 CONSIDERAÇÕES FINAIS

A motivação para este trabalho partiu da existência da deficiência em encontrar ferramentas genéricas que realizem atividades OLAM e/ou análise de desempenho de classificadores em um formato OLAM. A qualidade da solução desenvolvida e seus resultados apresentam características que apoiam os especialistas no domínio e mostraram a viabilidade em utilizar a Mineração de Dados em conjunto a operações OLAP para auxiliar os tomadores de decisão, estatísticos e especialistas a interpretar como o classificador em avaliação se comporta em relação a suas variáveis e subcubos formados nas consultas.

A maior parte dos trabalhos que abordam OLAM, apresentados no item 3.1, utilizam os modelos de arquitetura apresentados por Han, porém esse tipo de arquitetura provê apenas OLAM para mineração de dados em cubos, utilizando outro engine específico para as consultas OLAM e não utilizando o engine OLAP. Isso dificulta sua utilização de forma genérica, como por exemplo, na aplicação do conceito à avaliação de desempenho de classificadores, onde o objetivo é avaliar o comportamento do classificador a partir do resultado da mineração sobre a amostra testada. Assim a pesquisa de uma abordagem multidimensional para OLAM como ferramenta que utiliza conceitos de OLAP e DW de forma genérica para avaliação de modelos de classificação binária, possui grande relevância.

A abordagem utilizada no trabalho tem característica genérica pelo fato de não necessitar de nenhuma mudança na tecnologia; é possível utilizar qualquer *engine* OLAP, desde que permita a criação de membros calculados (comum na maioria dos *engines* OLAP) e tenha capacidade analítica, e o SGBD utilizado não possua nenhum recurso diferente do oferecido nativamente.

A partir da abordagem proposta é possível avaliar um classificador binário independente do problema em questão, pois o objeto de avaliação é o modelo. Construindo um *data warehouse* a partir do conjunto avaliado, inserindo o score dado pelo classificador fisicamente no modelo dimensional e criando o cubo de dados para o modelo dimensional a ser avaliado, criando a(s) medida(s) de avaliação do modelo.

Uma vez essas etapas realizadas é possível avaliar alterações no classificador simplesmente, alterando o classificador e executando a mesma base de testes ou avaliar o desempenho do classificador em outros testes, alterando a amostra (desde que com as mesmas variáveis e dimensões) e inserido-as no *data warehouse* para avaliação do classificador em relação a nova amostra de dados.

### 6.1. CONTRIBUIÇÕES

As principais contribuições oriundas deste trabalho são:

- Abordagem genérica para avaliação de classificadores binários pelo uso de ferramentas OLAP para realizar OLAM.
- Integração de Mineração de Dados e OLAP a partir do conjunto formado pelo *data warehouse*, cubo de dados e *engine* OLAP, utilizando consultas para visualizar o resultado dos subcubos.
- A abordagem não necessita de *engine* OLAM e nem alterações no *engine* OLAP para ser efetiva na avaliação de classificadores binários.
- Capacidade de detectar, explorar e avaliar o desempenho dos subcubos a partir das consultas geradas, utilizando e avaliando as medidas criadas pelos subcubos.
- Independência da equipe de mineração de dados para realizar avaliações de desempenho do classificador, sem necessidade de manipulações de dados via outra ferramenta se não as utilizadas na abordagem proposta pelo trabalho.

### 6.2. LIMITAÇÕES

As principais limitações da proposta oferecida pelo trabalho são:

- A capacidade limitada do *engine* OLAP em realizar interações sobre os conjuntos não oferece a flexibilidade necessária para calcular os conjuntos de subcubos de forma completa das variáveis sem a necessidade de separação das consultas, em máquinas de baixo desempenho.
- A abordagem necessita que as operações focadas em mineração de dados sejam executadas fora do ambiente proposto e apenas o resultado seja inserido na proposta, dando seu início pela carga do resultado no *data warehouse*.

### 6.3. TRABALHOS FUTUROS

A partir das considerações expostas acima, a primeira evolução natural para a abordagem e como continuação deste trabalho é desenvolver a capacidade analítica do *engine* OLAP permitindo a maior interação entre subconjuntos durante os cálculos analíticos em memória feitos pelo *engine* OLAP. Além de disponibilizar

essa capacidade como *feature* do *engine* OLAP alterando a forma como essa interação ocorre dentro das hierarquias e níveis do cubo de dados, fazendo assim uma navegação dinâmica durante os cálculos analíticos requisitados em memória pelo cubo de dados e conseguindo um desempenho aceitável do engine OLAP para esse tipo de cálculo analítico, hoje sendo um gargalo computacional da abordagem proposta no trabalho.

Pretende-se também adaptar o *engine* OLAP para a abordagem prover o cálculo automatizado pelo cubo de dados das principais métricas de avaliação de desempenho de classificadores binários. Oferecendo esse recurso a partir da criação de uma medida simples, criada trivialmente, e requisitando apenas a seleção das variáveis, trazendo maior dinamicidade ao processo de execução das consultas OLAP feitas pela abordagem.

## REFERÊNCIAS

- ADEODATO, P. L.; VASCONCELOS, G. C.; ARNAUD, A. L.; SANTOS, R. F., CUNHA, R. C. L.; MONTEIRO, D. S. M. **Neural Networks vs Logistic Regression: a Comparative Study on a Large Data Set**. International Conference on Pattern Recognition, Cambridge, pp. 355-358, 2004.
- ATOL CONSEILS et Développements. **LIVRE BLANC - Les ETL Open Source**. Disponível em: <www.atolcd.com>. Acesso em: 20 abr. 2015.
- AURÉLIO, Marco, Marley VELLASCO, e Carlos Henrique LOPES. **Descoberta de Conhecimento e Mineração de Dados**. Artigo, ICA – Laboratório de Inteligência Computacional Aplicada, Departamento de Engenharia Elétrica/PUC–Rio, 2000.
- BALLARD, Chuck, e Dirk HERREMAN. **Data Modeling Techniques for Data Warehousing**. IBM, International Technical Support Organization, 1998.
- BARQUINI, R. **Planning and designing the Warehouse**. Prentice-Hall, 1996.
- CAO, L.: **Introduction to Domain Driven Data Mining**. In: Cao, L., et al. (eds.) Data Mining for Business Applications, pp. 3–10. 2008.
- CHAUDHURI, S.; DAYAL, U. **Decision support, Data Warehouse, and OLAP**, in Tutorials of the Twenty-Second international Conf. On Very Large Data Base, Bombay, pages 295-306, 1996.
- CHEN Q., **Mining Exceptions and Quantitative Association Rules in OLAP Data Cube**, Tese de Mestrado , Simon Fraser University , Canadá , Julho, 1999.
- CODD, E. F.; CODD, S.B.; SALLEY, C.T. **Providing OLAP to user Analysts: An IT Mandate**, White Paper, Arbor Software Corporation, 1993.
- COLAÇO, Methanias Júnior. **Projetando Sistemas de Apoio a Decisão Baseados em Data Warehouse**. Rio de Janeiro: Editora: Axcel Books, 2004.
- CONOVER, W. J.; **Practical Nonparametric Statistics**, New York, John Wiley & Sons, 1999.
- Craven, M. W. e Shavlik, J. W. **Extracting comprehensible concept representations from trained neural networks**. In Working Notes on the IJCAI95 Workshop on Comprehensibility in Machine Learning. 1995
- DATE, C. J. **Introdução a Sistemas de Bancos de Dados**. 8a Edição. Editora: Campus, 2004
- DIAS, M. M. **Um modelo de formalização do processo de desenvolvimento de sistemas de descoberta de conhecimento em banco de dados**. Programa de Pós-Graduação em Engenharia de Produção. Tese de Doutorado. Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 2001.
- DRUCKER, Peter F. **Administrando para obter resultados**. SÃO PAULO: Pioneira, 1998.

DUARTE, M.D.O.; GUSMÃO, A.P.H.; ALMEIDA, A.T. **Sistema de apoio à decisão com modelo aditivo para priorização de sistemas de informação**. In: Encontro Nacional de Engenharia de Produção, XXV ENEGEP, 2005.

ELMASRI, R; NAVATHE, S. B. **Sistemas de Banco de Dados**. 4a ed. São Paulo: Addison Wesley, 2005.

ENEM. **ENEM**. Disponível em: <<http://portal.inep.gov.br/web/enem/sobre-o-enem>>. Acesso em: 18 Nov. 2014.

FAYYAD, U.M.; PIATETSKY-SHAPIRO, G. & SMYTH, P. **Advances in knowledge discovery & data mining**. Chapter 1: From data mining to knowledge discovery: an overview. AAAI/MIT. 1996.

FORTULAN, Marcos Roberto; GONCALVES FILHO, Eduardo Vila. **Uma proposta de aplicação de business intelligence no chão-de-fábrica**. Gestão da Produção, São Carlos, v. 12, n. 1, 2005.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: um guia prático**. Rio de Janeiro: Elsevier, 2005

GYSENS, M.; LAKSHMANAN, L. V. S. **A foundation for multi-dimensional databases**. In Twenty-third International Conf. On Very Large Bata Bases, Athens, pages 106-115, 1997.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**, 2nd edition. 2006.

HAN, J.; CHIANG, J.; CHEE, S.; CHEN, J.; CHEN, Q.; CHENG, S.; GONG, W.; KAMBER, M.; KOPERSKI, K.; LIU, G.; LU, Y.; STEFANOVIC, N.; WINSTONE, L.; XIA, B.;ZAIANE, O. R.; ZHANG, S.; ZHU, H. **DBMiner: A System for Data Mining in Relational Database and Data Warehouses**. Proc. CASCON'97: Meeting of Minds, Toronto, Canada, 1997.

HAN, J; CHEE, S; CHIANG, J. **Issues for on-line analytical mining of data warehouses**. In Proc. of the Workshop on Research Issues on Data Mining and Knowledge Discovery, Seattle, Washington, 1998.

HUA Z. , **On-Line Analytical Mining of Association Rules**, Tese de Mestrado ,Simon Fraser University , Canadá , Dezembro – 1998.

WITTEN, I, FRANK, E. **Data Mining: Pratical Machine Learning Tools and Techniques**. 2nd edition. In: Elsevier, USA. 2005.

INMON, W. H. **Como construir o Data Warehouse**. In: 2a edição. Rio de Janeiro: Campus. 1997.

INMON, W.H. **The Data Warehouse and Data Mining**. In: Communcation of the ACM, Vol. 39, No. 11. 1996.

INMON, W. H. **Building the Data Warehouse: Gettiing Started**. 4 a Edição. Editora: Wiley Publishing, Inc. 2005.

KELKAR B., **Exploiting symbiosis between data mining and OLAP for business insights**, DM Direct – online – dez/2001

KIMBALL, R. **The Data Warehouse Toolkit**. John Wiley, 1996.

KIMBALL, Ralph. **Digging into data mining: your data warehouse is your data mining platform**. DBMS and Internet Systems, 1997.

KIMBALL, Ralph et al. **The Data Warehouse Lifecycle Toolkit: Expert Methods For Designing, Developing and Deploying Data Warehouses**. New York: John Wiley & Sons Inc., 1998.

LAUDON, K. C.; LAUDON, J. P. **Management Information Systems**. 11. ed. PrenticeHall, 2009.

LIU B., HSU W., MA Y., “**Integrity classification and association rule mining**”, Proceeding of the Fourth Int. Conference on Knowledge Discovery and Data Mining, AAAI Press, New York, New York, 1998.

LIU B. et al, “**Rule Interestingness Analysis Using OLAP Operations**”, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining , 2006.

MACHADO, Felipe N. R. **Projeto de Data Warehouse: Uma visão Multidimensional**, Ed. Érica: São Paulo, 2000.

MAEDCHE A. et al, “**Enhancing Preprocessing in Data-Intensive Domains using Online-Analytical Processing Data Warehousing and Knowledge Discovery**”, Second International Conference, DaWaK 2000, London, UK, September 4-6, 2000, Proceedings.

MENOLLI, A. L. A. **Definição de uma Arquitetura de Data Warehouse para Gestão em Ciência e Tecnologia no Brasil**. Programa de Pós-Graduação em Ciência da Computação. Dissertação de Mestrado. Departamento de Informática. Universidade Estadual de Maringá, 2004.

MESSAOUD, R. B., BOUSSAÏD, O. & LOUDCHER-RABASEDA, S. 2006. **Mining association rules in olap cubes**. In International Conference on Innovations in Information Technology (ITT06), Dubai. 2006.

MESSAOUD, R. B., BOUSSAÏD, O. & RABASÉDA, S. 2004. **A new OLAP aggregation based on the AHC technique**. Proceedings of the 7th ACM international workshop on Data warehousing and OLAP. Washington, DC, USA: ACM. 2004

MICROSOFT CORPORATION. **Cube Cells**. Disponível em: <<https://msdn.microsoft.com/en-us/library/ms175449.aspx>>. Acesso em: 10 Abr. 2015.

NAPOLI, Marcio; SELL, Denilson; LIDUÁRIO, Luciano; BORGES, Rodrigo Ferreira; TODESCO, José Leomar; PACHECO, Roberto Carlos dos Santos. **Um Framework para Concepção de Ferramentas de Apoio a Decisão Baseadas em Ontologias**. In: Simpósio Brasileiro de Banco de Dados (SBBDD), Florianópolis. Anais, 2006.

NAPOLI, Marcio; SELL, Denilson. **Uma abordagem baseada em ontologias para realizar operações TextOLAP**. In: Ontobras, 2010, Florianópolis. 3º Seminário de Pesquisa em Ontologia no Brasil, 2010.

PAES, M. ; LIMA, Alexandre A. B. ; MATTOSO, Marta Lima de Queirós . **Processamento de Alto Desempenho em Consultas sobre Bases de Dados Geoestatísticos Usando Replicação Parcial**. In: Simpósio Brasileiro de Banco de Dados, 2009, Fortaleza. Simpósio Brasileiro de Banco de Dados- Trilha de Aplicações, 2009.

PENDSE, N. **OLAP Architectures**. Disponível em:<<http://www.olapreport.com/Architectures.htm>>. Acesso em: 10 mai. 2015.

PENTAHO, **Mondrian Architecture**. Disponível em: <<http://infocenter.pentaho.com/help/index.jsp>>. Acesso em: 15. Abr. 2015.

PORTAL INEP. **Censo Escolar**. Disponível em: <<http://portal.inep.gov.br/>>. Acesso em: 18 nov. 2014.

PONNIAH, Paulraj. **Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals**. John Wiley & Sons, Inc., 2001.

ROMÃO, W. **Descoberta de Conhecimento relevante em Bancos de Dados sobre Ciência e Tecnologia**. Tese Doutorado. Universidade Federal de Santa Catarina. Florianópolis, 2002.

SERRA, Laércio. **A essência do Business Intelligence**. São Paulo: Berkeley Brasil, 2002.

SILBERSCHATZ, A. & A. TUZHILIN. **On subjective measures of interestingness in knowledge discovery**. Proceedings of the First International Conference on Knowledge Discovery and Data Mining 1, 275–281. 1995.

SILVA, M. P. S.; ROBIN, J. R. **SKDQL – Uma Linguagem Declarativa de Especificação de Consultas e Processos para Descoberta de Conhecimento em Bancos de Dados e sua Implementação**. Dissertação de Mestrado. UFPE, 2002.

SINGH, Harry S. **Data Warehouse: Conceitos, Tecnologias, Implementação e Gerenciamento**. Tradução de Mônica Rosemberg. São Paulo: Makron Books, 2001.

SHI et al., “**MSMiner: a developing platform for OLAP**”, Decision Support Systems, nº 42, 2007, pag. 2016–2028.

SHIBA, S. K. **Desenvolvimento de modelo de processamento de extração de conhecimento em banco de dados para sistemas de suporte à decisão**. Dissertação de Mestrado, Escola Politécnica, Universidade de São Paulo. 2008.

SONG, Y.; ROWEN, W.; MEDSKER, C.; EWEN, E. **An Analysis of Many-to-Many Relationships Between Fact and Dimension Tables in Dimensional Modeling**. In Proceedings of the International Workshop on Design and Management of Data Warehouse (DMDW 2001), Interlaken, Suíça, 2001.

SOUSA, M.U.R.S. Silva, K.P. Adeodato, P.J.L. **Data Mining Applied To The Processes Celerity Of Pernambuco’s State Court Of Accounts**. In: Proc. CONTECSI 2008.

TAN Y. J. , **Datacube** : Its Implementation and Application in OLAP Mining, Tese de Mestrado, Simon Fraser University, Canadá, Setembro – 1998.

THOMSEN, E. **OLAP : Construindo Sistemas de Informações Multidimensionais**. 2o edição. Rio de Janeiro: Editora Campus, 2002.

TORLONE, Riccardo; CABIBBO, Luca. **A Logical Framework for Querying Multidimensional Data. Dipartimento di Informatica e Automazione**. Università di Roma. Rome, Italy, 1998.

TRAVITZKI, R. **ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar**. Tese doutorado, USP, São Paulo, 2013.

WEST, D. **Neural network credit scoring models**. Computers and Operations Research. vol 27, p. 1131-1152, 2000.

WREMBEL, Robert, e Christian KONCILIA. **Data Warehouses and OLAP: Concepts, Architectures and Solutions**. Edição: IRM Press. Hershey PA: Idea Group Inc., 2007.

ZHANG, Duo; Zhai, ChengXiang; Han, Jiawei; Srivastava, Asholk; Oza Nikunk. **Topic Modeling for OLAP on Multidimensional Text Databases**: Topic Cube and its Applications. Statistical Analysis and Data Mining vol. 2. 378-395. 2009.

## APÊNDICE

O Apêndice A fornece detalhes sobre o arquivo referente ao cubo, gerado no schema workbench, sua sintaxe e funcionamento.

```
<Schema name="Mestrado">
```

```
<Cube name="ROC" visible="true" cache="true" enabled="true">
```

```
<Table name="fato_escolas" schema="public">
```

```
</Table>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Biblioteca" description="Existe biblioteca na escola?">
```

```
<Hierarchy name="Biblioteca" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao" schema="public">
```

```
</Table>
```

```
<Level name="Desc Biblioteca" visible="true" column="biblioteca" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Copiadora" description="Existe copiadora na escola?">
```

```
<Hierarchy name="Copiadora" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao">
```

```
</Table>
```

```
<Level name="Desc Copiadora" visible="true" column="copiadora" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Cozinha" description="Existe cozinha na escola?">

  <Hierarchy name="Cozinha" visible="true" hasAll="true" primaryKey="id_dim_simounao">

    <Table name="dim_simounao">

      </Table>

      <Level name="Desc Cozinha" visible="true" column="cozinha" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="DVD" description="Existe aparelho de DVD na escola?">

      <Hierarchy name="DVD" visible="true" hasAll="true" primaryKey="id_dim_simounao">

        <Table name="dim_simounao">

          </Table>

          <Level name="Desc DVD" visible="true" column="dvd" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

            </Level>

          </Hierarchy>

        </Dimension>

```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Impressora" description="Existe Impressora na escola?">
```

```
<Hierarchy name="Impressora" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao">
```

```
</Table>
```

```
<Level name="Desc Impressora" visible="true" column="impressora" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Internet" description="Existe acesso a internet na escola?">
```

```
<Hierarchy name="Internet" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao">
```

```
</Table>
```

```
<Level name="Desc Internet" visible="true" column="internet" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Lab. Ciencias" description="Existe Lab. de Ci&#234;ncias na Escola?">
```

```
<Hierarchy name="Lab. Ciencias" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao">
```

```

</Table>

<Level name="Desc Lab Ciencias" visible="true" column="laboratorio_ciencias" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Lab. Informatica" description="Existe Lab. de Informatica na Escola?">

<Hierarchy name="Lab. Informatica" visible="true" hasAll="true" primaryKey="id_dim_simounao">

<Table name="dim_simounao">

</Table>

<Level name="Desc Lab Informatica" visible="true" column="laboratorio_informatica" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_localizacao" highCardinality="false"
name="Localizacao" description="Qual a localizacao da escola?">

<Hierarchy name="Localizacao" visible="true" hasAll="true" primaryKey="id_dim_localizacao">

<Table name="dim_localizacao" schema="public">

</Table>

<Level name="Regiao" visible="true" column="regiao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

```

```
<Level name="Estado" visible="true" column="estado" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
<Level name="Localizacao" visible="true" column="localizacao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Parabolica" description="Existe antena parabolica na escola?">
```

```
<Hierarchy name="Parabolica" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao" schema="public">
```

```
</Table>
```

```
<Level name="Desc Parabolica" visible="true" column="parabolica" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Parque Infantil" description="Existe parque infantil na escola?">
```

```
<Hierarchy name="Parque Infantil" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao">
```

```
</Table>
```

```
<Level name="Desc P. Infantil" visible="true" column="parque_infantil" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```

</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Retroprojektor" description="Existe retroprojektor na escola? ">
  <Hierarchy name="Retroprojektor" visible="true" hasAll="true" primaryKey="id_dim_simounao">
    <Table name="dim_simounao">
      </Table>
      <Level name="Desc Retroprojektor" visible="true" column="retroprojektor" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Sala Diretoria" description="Existe sala de diretoria na escola? ">
  <Hierarchy name="Sala Diretoria" visible="true" hasAll="true" primaryKey="id_dim_simounao">
    <Table name="dim_simounao">
      </Table>
      <Level name="Desc Diretoria" visible="true" column="sala_diretoria" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>

```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Sala Leitura" description="Existe sala de leitura na escola?">
```

```
<Hierarchy name="Sala Leitura" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao">
```

```
</Table>
```

```
<Level name="Desc Leitura" visible="true" column="sala_leitura" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Sala Professor" description="Existe sala de professores na escola?">
```

```
<Hierarchy name="Sala Professor" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao">
```

```
</Table>
```

```
<Level name="Desc Professor" visible="true" column="sala_professor" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="TV" description="Existe aparelho de tv na escola?">
```

```
<Hierarchy name="TV" visible="true" hasAll="true" primaryKey="id_dim_simounao">
```

```
<Table name="dim_simounao">
```

```

</Table>

<Level name="Desc TV" visible="true" column="tv" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_simounao" highCardinality="false"
name="Videocassete" description="Existe aparelho de videocassete na escola?">

<Hierarchy name="Videocassete" visible="true" hasAll="true" primaryKey="id_dim_simounao">

<Table name="dim_simounao">

</Table>

<Level name="Desc Videocassete" visible="true" column="videocassete" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa Aluno Media Idade" description="Faixa de idade media dos alunos.">

<Hierarchy name="faixa_aluno_media_idade" visible="true" hasAll="true" primaryKey="id_dim_faixas">

<Table name="dim_faixas" schema="public">

</Table>

<Level name="Faixa Aluno Media Idade" visible="true" column="faixa_aluno_media_idade" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

```

```

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa Aluno Quantidade" description="Faixa de quantidade de alunos">

  <Hierarchy name="Faixa Aluno Quantidade" visible="true" hasAll="true" primaryKey="id_dim_faixas">

    <Table name="dim_faixas" schema="public">

      </Table>

      <Level name="Faixa Aluno Quantidade" visible="true" column="faixa_aluno_quantidade" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa Professores Media Idade" description="Faixa de idade media dos professores">

      <Hierarchy name="Faixa Professores Media Idade" visible="true" hasAll="true" primaryKey="id_dim_faixas">

        <Table name="dim_faixas" schema="public">

          </Table>

          <Level name="Faixa Professores Media Idade" visible="true" column="faixa_m_professores_idade"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

            </Level>

          </Hierarchy>

        </Dimension>

        <Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa N Computadores" description="Faixa do numero de computadores ">

```

```

<Hierarchy name="Faixa N Computadores" visible="true" hasAll="true" primaryKey="id_dim_faixas">
  <Table name="dim_faixas" schema="public">
    </Table>
    <Level name="Faixa N Computadores" visible="true" column="faixa_num_computadores" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
  </Level>
</Hierarchy>
</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa N Funcionarios" description="Faixa da quantidade de funcionarios ">
  <Hierarchy name="Faixa N Funcionarios" visible="true" hasAll="true" primaryKey="id_dim_faixas">
    <Table name="dim_faixas" schema="public">
      </Table>
      <Level name="Faixa N Funcionarios" visible="true" column="faixa_num_funcionarios" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa N Salas" description="Faixa do numero de salas da escola">
      <Hierarchy name="Faixa N Salas" visible="true" hasAll="true" primaryKey="id_dim_faixas">
        <Table name="dim_faixas" schema="public">
          </Table>

```

```
<Level name="Faixa N Salas" visible="true" column="faixa_num_salas_existentes" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa P Aluno Raca Amarela" description="Faixa percentual de alunos de raca amarela">
```

```
<Hierarchy name="Faixa P Aluno Raca Amarela" visible="true" hasAll="true" primaryKey="id_dim_faixas">
```

```
<Table name="dim_faixas" schema="public">
```

```
</Table>
```

```
<Level name="Faixa P Aluno Raca Amarela" visible="true" column="faixa_p_aluno_raca_amarela"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa P Aluno Raca Branca" description="Faixa percentual de alunos de raca branca">
```

```
<Hierarchy name="Faixa P Aluno Raca Branca" visible="true" hasAll="true" primaryKey="id_dim_faixas">
```

```
<Table name="dim_faixas" schema="public">
```

```
</Table>
```

```
<Level name="Faixa P Aluno Raca Branca" visible="true" column="faixa_p_aluno_raca_branca"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa P Aluno Raca Indigena" description="Faixa percentual de alunos de raca indigena">

  <Hierarchy name="Faixa P Aluno Raca Indigena" visible="true" hasAll="true" primaryKey="id_dim_faixas">

    <Table name="dim_faixas" schema="public">

      </Table>

      <Level name="Faixa P Aluno Raca Indigena" visible="true" column="faixa_p_aluno_raca_indigena"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa P Aluno Raca N Informado" description="Faixa percentual de alunos de raca nao informado">

      <Hierarchy name="Faixa P Aluno Raca N Informado" visible="true" hasAll="true"
primaryKey="id_dim_faixas">

        <Table name="dim_faixas" schema="public">

          </Table>

          <Level name="Faixa P Aluno Raca N Informado" visible="true" column="faixa_p_aluno_raca_n_informado"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

            </Level>

          </Hierarchy>

        </Dimension>

        <Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa P Aluno Raca Parda" description="Faixa percentual de alunos de raca parda">

          <Hierarchy name="Faixa P Aluno Raca Parda" visible="true" hasAll="true" primaryKey="id_dim_faixas">

```

```

<Table name="dim_faixas" schema="public">

</Table>

<Level name="Faixa P Aluno Raca Parda" visible="true" column="faixa_p_aluno_raca_parda" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa P Aluno Raca Preta" description="Faixa percentual de alunos de raca preta">

<Hierarchy name="Faixa P Aluno Raca Preta" visible="true" hasAll="true" primaryKey="id_dim_faixas">

<Table name="dim_faixas" schema="public">

</Table>

<Level name="Faixa P Aluno Raca Preta" visible="true" column="faixa_p_aluno_raca_preta" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa P Aluno Sexo Masculino" description="Faixa percentual de alunos do sexo masculino">

<Hierarchy name="Faixa P Aluno Sexo Masculino" visible="true" hasAll="true" primaryKey="id_dim_faixas">

<Table name="dim_faixas" schema="public">

</Table>

<Level name="Faixa P Aluno Sexo Masculino" visible="true" column="faixa_p_aluno_sexo_masculino"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

```

```

</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa P Aluno Necessidade Esp" description="Faixa percentual de alunos com necessidade especial">
  <Hierarchy name="Faixa P Aluno Necessidade Esp" visible="true" hasAll="true"
primaryKey="id_dim_faixas">
    <Table name="dim_faixas" schema="public">
      </Table>
      <Level name="Faixa P Aluno Necessidade Esp" visible="true"
column="faixa_p_alunos_necessidade_especial" type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa Qtd Professores Dr" description="Faixa da Quantidade de Professores Dr">
      <Hierarchy name="Faixa Qtd Professores Dr" visible="true" hasAll="true" primaryKey="id_dim_faixas">
        <Table name="dim_faixas" schema="public">
          </Table>
          <Level name="Faixa Qtd Professores Dr" visible="true" column="faixa_qtd_doutorado" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
            </Level>
          </Hierarchy>
        </Dimension>

```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa Qtd Professores Esp" description="Faixa da Quantidade de Professores Especialistas">
```

```
<Hierarchy name="Faixa Qtd Professores Esp" visible="true" hasAll="true" primaryKey="id_dim_faixas">
```

```
<Table name="dim_faixas" schema="public">
```

```
</Table>
```

```
<Level name="Faixa Qtd Professores Esp" visible="true" column="faixa_qtd_especializacao" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa Qtd Professores Mestrado" description="Faixa da Quantidade de professores com Mestrado">
```

```
<Hierarchy name="Faixa Qtd Professores Mestrado" visible="true" hasAll="true"
primaryKey="id_dim_faixas">
```

```
<Table name="dim_faixas" schema="public">
```

```
</Table>
```

```
<Level name="Faixa Qtd Professores Mestrado" visible="true" column="faixa_qtd_mestrado" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa Qtd Professores N Pos" description="Faixa de Professores sem nenhuma p&#243;s
gradua&#231;&#227;o">
```

```
<Hierarchy name="Faixa Qtd Professores N Pos" visible="true" hasAll="true" primaryKey="id_dim_faixas">
```

```

<Table name="dim_faixas" schema="public">

</Table>

<Level name="Faixa Qtd Professores N Pos" visible="true" column="faixa_qtd_nenhuma_pos"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="id_dim_faixas" highCardinality="false"
name="Faixa Total de Professores" description="Faixa de Total de Professores">

<Hierarchy name="Faixa Total de Professores" visible="true" hasAll="true" primaryKey="id_dim_faixas">

<Table name="dim_faixas" schema="public">

</Table>

<Level name="Faixa Total de Professores" visible="true" column="faixa_total_professores" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q14" highCardinality="false"
name="Q14" description="Vc fez algum curso preparat&#243;rio para o trabalho que realiza ou realizou?">

<Hierarchy name="Q14" visible="true" hasAll="true" primaryKey="valor">

<Table name="dim_q14" schema="public">

</Table>

<Level name="Desc Q14" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

```

```

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q15" highCardinality="false"
name="Q15" description="Indique os cursos que voc&#234; frequentou ou frequenta? Curso
profissionalizante">

  <Hierarchy name="Q15" visible="true" hasAll="true" primaryKey="valor">

    <Table name="dim_q15" schema="public">

      </Table>

      <Level name="Desc Q15" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="m_q16" highCardinality="false"
name="Q16" description="Indique os cursos que voc&#234; frequentou ou frequenta? Curso preparat&#243;rio
para vestibular">

      <Hierarchy name="Q16" visible="true" hasAll="true" primaryKey="valor">

        <Table name="dim_q16" schema="public">

          </Table>

          <Level name="Q16" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

            </Level>

          </Hierarchy>

        </Dimension>

```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q17" highCardinality="false"
name="Q17" description="Indique os cursos que voc&#234; frequentou ou frequenta? Curso superior">
```

```
<Hierarchy name="Q17" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q17" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q17" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q18" highCardinality="false"
name="Q18" description="Indique os cursos que voc&#234; frequentou ou frequenta? Curso de l&#237;ngua
estrangeira">
```

```
<Hierarchy name="Q18" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q18" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q18" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q19" highCardinality="false"
name="Q19" description="Indique os cursos que voc&#234; frequentou ou frequenta? Curso de
computa&#231;&#227;o ou inform&#225;tica">
```

```
<Hierarchy name="Q19" visible="true" hasAll="true" primaryKey="valor">
```

```

<Table name="dim_q19" schema="public">
</Table>

<Level name="Desc Q19" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q2" highCardinality="false" name="Q2"
description="Ate Quando seu pai estudou?">
<Hierarchy name="Q2" visible="true" hasAll="true" primaryKey="valor">
<Table name="dim_q2" schema="public">
</Table>

<Level name="Desc Q2" visible="true" column="descricao" ordinalColumn="valor" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q20" highCardinality="false"
name="Q20" description="Indique os cursos que voc&#234; frequentou ou frequenta? Curso preparat&#243;rio
para concursos p&#252;blicos">
<Hierarchy name="Q20" visible="true" hasAll="true" primaryKey="valor">
<Table name="dim_q20" schema="public">
</Table>

<Level name="Desc Q20" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

```

```
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="m_q21" highCardinality="false"
name="Q21" description="Indique os cursos que voc&#234; frequentou ou frequenta? Outro curso">
  <Hierarchy name="Q21" visible="true" hasAll="true" primaryKey="valor">
    <Table name="dim_q21" schema="public">
      </Table>
    <Level name="Desc Q21" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" visible="true" foreignKey="m_q22" highCardinality="false"
name="Q22" description="Quantas horas semanais voc&#234; trabalha?">
    <Hierarchy name="Q22" visible="true" hasAll="true" primaryKey="valor">
      <Table name="dim_q22" schema="public">
        </Table>
      <Level name="Desc Q22" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q28" highCardinality="false"
name="Q28" description="Quantos anos voc&#234; levou para concluir o Ensino Fundamental?">
```

```
<Hierarchy name="Q28" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q28" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q28" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q3" highCardinality="false" name="Q3"
description="Ate quando sua mae estudou?">
```

```
<Hierarchy name="Q3" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q3" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q3" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q30" highCardinality="false"
name="Q30" description="Em que tipo de escola voc&#234; cursou o Ensino Fundamental?">
```

```
<Hierarchy name="Q30" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q30" schema="public">
```

```

</Table>

<Level name="Desc Q30" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q31" highCardinality="false"
name="Q31" description="Quantos anos voc&#234; levou para concluir o Ensino M&#233;dio?">

<Hierarchy name="Q31" visible="true" hasAll="true" primaryKey="valor">

<Table name="dim_q31" schema="public">

</Table>

<Level name="Desc Q31" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q33" highCardinality="false"
name="Q33" description="Em que tipo de escola voc&#234; cursou o Ensino M&#233;dio?">

<Hierarchy name="Q33" visible="true" hasAll="true" primaryKey="valor">

<Table name="dim_q33" schema="public">

</Table>

<Level name="Desc Q33" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q4" highCardinality="false" name="Q4"
description="Somando a sua Renda com a renda das pessoas que moram com vc, quanto &#233;, aprox., a
renda familiar mensal?">
```

```
<Hierarchy name="Q4" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q4" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q4" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q5" highCardinality="false" name="Q5"
description="Qual a sua renda mensal, aprox?">
```

```
<Hierarchy name="Q5" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q5" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q5" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q57" highCardinality="false"
name="Q57" description="Caso voc&#234; ingresse no Ensino Superior privado pretende recorrer aos
aux&#237;lios abaixo para custeio das mensalidades? FIES (Programa de Financiamento Estudantil)">
```

```
<Hierarchy name="Q57" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q57" schema="public">
```

```
</Table>
```

```
<Level name="Desc 57" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q58" highCardinality="false"
name="Q58" description="Caso voc&#234; ingresse no Ensino Superior privado pretende recorrer aos
aux&#237;lios abaixo para custeio das mensalidades? Pr&#243;-Uni (Programa Universidade para Todos)">
```

```
<Hierarchy name="Q58" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q58" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q58" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q59" highCardinality="false"
name="Q59" description="Caso voc&#234; ingresse no Ensino Superior privado pretende recorrer aos
aux&#237;lios abaixo para custeio das mensalidades? Bolsa de estudos da pr&#243;pria Institui&#231;&#227;o
de Ensino Superior">
```

```

<Hierarchy name="Q59" visible="true" hasAll="true" primaryKey="valor">
  <Table name="dim_q59" schema="public">
    </Table>
    <Level name="Desc Q59" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>

  <Dimension type="StandardDimension" visible="true" foreignKey="m_q6" highCardinality="false" name="Q6"
description="A casa onde vc mora &#233;?">
    <Hierarchy name="Q6" visible="true" hasAll="true" primaryKey="valor">
      <Table name="dim_q6" schema="public">
        </Table>
        <Level name="Desc Q6" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
          </Level>
        </Hierarchy>
      </Dimension>

      <Dimension type="StandardDimension" visible="true" foreignKey="m_q60" highCardinality="false"
name="Q60" description="Caso voc&#234; ingresse no Ensino Superior privado pretende recorrer aos
aux&#237;lios abaixo para custeio das mensalidades? Bolsa de estudos da empresa onde trabalho.">
        <Hierarchy name="Q60" visible="true" hasAll="true" primaryKey="valor">
          <Table name="dim_q60" schema="public">
            </Table>

```

```
<Level name="Desc Q60" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q61" highCardinality="false"
name="Q61" description="Voc&#234; tem em sua casa? TV em cores">
```

```
<Hierarchy name="Q61" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q61" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q61" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q62" highCardinality="false"
name="Q62" description="Voc&#234; tem em sua casa? Videocassete e/ou DVD">
```

```
<Hierarchy name="Q62" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q62" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q62" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q63" highCardinality="false"
name="Q63" description="Voc&#234; tem em sua casa? R&#225;dio">

  <Hierarchy name="Q63" visible="true" hasAll="true" primaryKey="valor">

    <Table name="dim_q63" schema="public">

      </Table>

      <Level name="Desc Q63" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="m_q64" highCardinality="false"
name="Q64" description="Voc&#234; tem em sua casa? Microcomputador">

      <Hierarchy name="Q64" visible="true" hasAll="true" primaryKey="valor">

        <Table name="dim_q64" schema="public">

          </Table>

          <Level name="Q64" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

            </Level>

          </Hierarchy>

        </Dimension>

        <Dimension type="StandardDimension" visible="true" foreignKey="m_q65" highCardinality="false"
name="Q65" description="Voc&#234; tem em sua casa? Autom&#243;vel">

          <Hierarchy name="Q65" visible="true" hasAll="true" primaryKey="valor">

```

```

<Table name="dim_q65" schema="public">
</Table>

<Level name="Desc Q65" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q66" highCardinality="false"
name="Q66" description="Voc&#234; tem em sua casa? M&#225;quina de lavar roupa">
<Hierarchy name="Q66" visible="true" hasAll="true" primaryKey="valor">
<Table name="dim_q66" schema="public">
</Table>

<Level name="Desc Q66" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q68" highCardinality="false"
name="Q68" description="Voc&#234; tem em sua casa? Freezer (aparelho independente ou parte da geladeira
duplex)">
<Hierarchy name="Q68" visible="true" hasAll="true" primaryKey="valor">
<Table name="dim_q68" schema="public">
</Table>

<Level name="Desc Q68" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

```

```
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="m_q69" highCardinality="false"
name="Q69" description="Voc&#234; tem em sua casa? Telefone fixo">
  <Hierarchy name="Q69" visible="true" hasAll="true" primaryKey="valor">
    <Table name="dim_q69" schema="public">
      </Table>
    <Level name="Desc Q69" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" visible="true" foreignKey="m_q7" highCardinality="false" name="Q7"
description="Sua casa est&#225; localizada em?">
    <Hierarchy name="Q7" visible="true" hasAll="true" primaryKey="valor">
      <Table name="dim_q7" schema="public">
        </Table>
      <Level name="Desc Q7" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q70" highCardinality="false"
name="Q70" description="Voc&#234; tem em sua casa? Telefone celular">
```

```
<Hierarchy name="Q70" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q70" schema="public">
```

```
</Table>
```

```
<Level name="Desc Q70" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q71" highCardinality="false"
name="Q71" description="Voc&#234; tem em sua casa? Acesso &#224; Internet">
```

```
<Hierarchy name="Q71" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q71" schema="public">
```

```
</Table>
```

```
<Level name="Q71" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q72" highCardinality="false"
name="Q72" description="Voc&#234; tem em sua casa? TV por assinatura">
```

```
<Hierarchy name="Q72" visible="true" hasAll="true" primaryKey="valor">
```

```
<Table name="dim_q72" schema="public">
```

```

</Table>

<Level name="Q72" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q73" highCardinality="false"
name="Q73" description="Voc&#234; tem em sua casa? Aspirador de p&#243;">

<Hierarchy name="Q73" visible="true" hasAll="true" primaryKey="valor">

<Table name="dim_q73" schema="public">

</Table>

<Level name="Q73" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q75" highCardinality="false"
name="Q75" description="Voc&#234; tem em sua casa? Banheiro">

<Hierarchy name="Q75" visible="true" hasAll="true" primaryKey="valor">

<Table name="dim_q75" schema="public">

</Table>

<Level name="Q75" visible="true" column="descricao" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

```

```

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q8" highCardinality="false" name="Q8"
description="Vc exerce ou j&#225; exerceu atividade remunerada?">

  <Hierarchy name="Q8" visible="true" hasAll="true" primaryKey="valor">

    <Table name="dim_q8" schema="public">

      </Table>

      <Level name="Q8" visible="true" column="valor" nameColumn="descricao" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="m_q9" highCardinality="false" name="Q9"
description="Grau de importancia para vc ter come&#231;ado a trabalhar. Ajudar meus pais nas despesas com
a casa.">

      <Hierarchy name="Q9" visible="true" hasAll="true">

        <Level name="Desc Q9" visible="true" column="m_q9" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

          </Level>

        </Hierarchy>

      </Dimension>

      <Dimension type="StandardDimension" visible="true" foreignKey="m_q1" highCardinality="false" name="Q1"
description="Quantas pessoas moram com vc?">

        <Hierarchy name="Q1" visible="true" hasAll="true">

```

```
<Level name="Desc Q1" visible="true" column="m_q1" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q10" highCardinality="false"
name="Q10" description="Grau de importancia para vc ter come&#231;ado a trabalhar. Sustentar minha
familia.">
```

```
<Hierarchy name="Q10" visible="true" hasAll="true">
```

```
<Level name="Desc Q10" visible="true" column="m_q10" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q11" highCardinality="false"
name="Q11" description="Grau de importancia para vc ter come&#231;ado a trabalhar. Ser independente.">
```

```
<Hierarchy name="Q11" visible="true" hasAll="true">
```

```
<Level name="Desc Q11" visible="true" column="m_q11" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
</Dimension>
```

```
<Dimension type="StandardDimension" visible="true" foreignKey="m_q12" highCardinality="false"
name="Q12" description="Grau de importancia para vc ter come&#231;ado a trabalhar. Adquirir
experi&#234;ncia">
```

```

<Hierarchy name="Q12" visible="true" hasAll="true">
  <Level name="Q12" visible="true" column="m_q12" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="m_q13" highCardinality="false"
name="Q13" description="Grau de importancia para vc ter come&#231;ado a trabalhar. Custear/pagar meus
estudos.">
  <Hierarchy name="Q13" visible="true" hasAll="true">
    <Level name="Q13" visible="true" column="m_q13" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" visible="true" foreignKey="m_q23" highCardinality="false"
name="Q23" description="Com que idade voc&#234; come&#231;ou a trabalhar?">
    <Hierarchy name="Q23" visible="true" hasAll="true">
      <Level name="Desc Q23" visible="true" column="m_q23" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="m_q24" highCardinality="false"
name="Q24" description="Indique o que levou voc&#234; a participar do ENEM: Testar meus conhecimentos">

```

```

<Hierarchy name="Q24" visible="true" hasAll="true">
  <Level name="Desc Q24" visible="true" column="m_q24" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>

<Dimension type="StandardDimension" visible="true" foreignKey="m_q26" highCardinality="false"
name="Q26" description="Indique o que levou voc&#234; a participar do ENEM: Obter a certifica&#231;&#227;o
do Ensino M&#233;dio ou acelerar meus estudos">
  <Hierarchy name="Q26" visible="true" hasAll="true">
    <Level name="Desc Q26" visible="true" column="m_q26" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>

  <Dimension type="StandardDimension" visible="true" foreignKey="m_q27" highCardinality="false"
name="Q27" description="Indique o que levou voc&#234; a participar do ENEM: Conseguir uma bolsa de
estudos (ProUni, outras)">
    <Hierarchy name="Q27" visible="true" hasAll="true">
      <Level name="Desc Q27" visible="true" column="m_q27" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>

    <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Escola">

```

```

<Hierarchy visible="true" hasAll="true">
  <Level name="Escola" visible="true" column="id_escola" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
    <Property name="alvo" column="alvo" type="Numeric">
      </Property>
    </Level>
  </Hierarchy>
</Dimension>
<Measure name="Score Medio" column="escore" aggregator="avg" description="Score Medio" visible="true">
</Measure>
<Measure name="aux_auc_roc" column="auc_roc" formatString="0.000" aggregator="sum"
description="aux_auc_roc" visible="false">
</Measure>
<Measure name="FDP_Bons" column="fdp_bons" formatString="0.00" aggregator="max" visible="false">
</Measure>
<Measure name="FDP_Maus" column="fdp_maus" formatString="0.00" aggregator="max" visible="false">
</Measure>
<Measure name="auxCount" column="id_escola" aggregator="count" visible="true">
</Measure>
<Measure name="Score Min" column="escore" aggregator="min" visible="true">
</Measure>
<Measure name="Score Max" column="escore" aggregator="max" visible="true">
</Measure>

```

```
<CalculatedMember                name="alvo"                formatString="#,##0.00"
formula="iif([Escola].Properties(&#34;alvo&#34;)=0,1,0)" dimension="Measures" visible="true">
</CalculatedMember>
<CalculatedMember    name="n_alvo"    formula="IIF([Measures].[alvo]=0,1,0)"    dimension="Measures"
visible="true">
</CalculatedMember>
<CalculatedMember    name="Count"    formula="sum([Measures].[auxCount])"    dimension="Measures"
visible="true">
</CalculatedMember>
</Cube>
</Schema>
```