

## A UTILIZAÇÃO DE TECNOLOGIAS DE VISÕES MULTIDIMENSIONAIS PARA APOIO À ANÁLISE DOS PERFIS SOCIOECONÔMICOS DE ALUNOS DO ENSINO BÁSICO

Ana Carolina Soares e Soares <sup>1</sup>

**RESUMO** - O Ministério da Educação (MEC) a cada dois anos realiza o Sistema de Avaliação da Educação Básica (SAEB). Nesta avaliação, são empregados formulários socioeconômicos, além de provas de Língua Portuguesa e Matemática para alunos do ensino fundamental w médio de escolas públicas e privadas.

Em virtude do tamanho da base de dados do SAEB 2015, se faz necessário a utilização de ferramentas próprias para a devida exploração e descoberta de conhecimento. O objetivo deste artigo é analisar os padrões que podem motivar um desempenho inferior ou superior dos alunos de escolas públicas, referente ao aprendizado na matérias básicas do ensino fundamental do Estado do Rio de Janeiro.

No presente trabalho, apresenta-se, no primeiro momento, uma revisão bibliográfica sobre *Business Intelligence*. Onde, serão abordados os principais tópicos do tema e o desenvolvimento do projeto de análise multidimensional. Por meio de softwares específicos, serão gerados alguns relatórios, que servirão de base para uma análise dos fatores positivos e negativos influentes no desempenho dos alunos.

Palavras-Chave: *Business Intelligence*; OLAP, SAEB; *Data Warehouse*

---

<sup>1</sup> Graduada em Engenharia da Computação pela Universidade Veiga de Almeida, Rio de Janeiro, RJ, carolsoaresrj@hotmail.com

## **The use of multidimensional vision technologies to support the analysis of socioeconomic profiles of elementary school students.**

**ABSTRACT** - The Ministry of Education (MEC) every two years carries out the Basic Education Assessment System (SAEB). In the evaluation, are chosen socioeconomic, as well as tests of Portuguese Language and Mathematics for the students of fundamental fundamental education of public and private schools.

Due to the size of the SAEB 2015 database, it is necessary to use tools for dynamic exploration and knowledge discovery. The monograph is a study about the patterns that can motivate the performance of elementary and middle school students of elementary and middle school students in the State of Rio de Janeiro.

In the present work, a bibliographic review of Business Intelligence is presented, without a first moment. Topics are covered topics and the multidimensional project. Specific software, generated a few reports, that will serve as the basis for a collection of positive and negative data influencing student performance.

Keywords: Business Intelligence; Mining; OLAP, SAEB; Data warehouse

## **INTRODUÇÃO**

O Ministério da Educação (MEC) a cada dois anos realiza o Sistema de Avaliação da Educação Básica (SAEB) também conhecido como Prova Brasil. Nesta avaliação, participam todos os estudantes de escolas públicas urbanas do 5º e do 9º e 3º ano do ensino médio de turmas com mais de 20 alunos de todo o Brasil. Este censo visa a coleta de indicadores de proficiência em Língua Portuguesa e Matemática dos alunos do ensino básico, além de informações sobre as escolas, diretores, professores e alunos. Estas informações são armazenadas em uma ampla base de dados disponibilizada a pesquisadores pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). No Brasil, com a

implantação do SAEB, foi produzido um conjunto de informações que possibilitou o desenvolvimento de pesquisas sobre o sistema escolar do país. Essas informações evidenciaram quais eram os principais fatores, extra e intraescolares, relacionados ao desempenho escolar. De acordo com Ferrão (2001), o nível socioeconômico dos alunos, ao lado de outros fatores, tais como o atraso escolar e a cor/etnia, se mostrou significativamente associado ao desempenho obtido pelos estudantes em testes cognitivos.

Para Alves (2014), devido ao seu grande volume de dados, esta base nem sempre é explorada em toda sua magnitude, deixando de se aproveitar informações valiosas sobre a situação e o desenvolvimento escolar dos alunos do ensino básico da rede pública. Por isso, a definição dos indicadores importantes, a formulação dos itens com clareza e precisão, a extensão dos questionários, a adequação à faixa etária e a possibilidade de itens total ou parcialmente não terem tido respostas, são elementos críticos a serem considerados. Os pesquisadores necessitam de informações relevantes e fidedignas, o que nem sempre estão disponíveis da forma ideal.

Grandes bases de dados operacionais impossibilitam uma análise mais profunda e detalhadas sobre seus dados, pois as consultas em *Structured Query Language* (SQL) se tornam complexas e de difícil programação, além de suas limitações.

Sendo assim, serão analisadas duas visões no desenvolvimento do presente trabalho: visões multidimensionais de dados. A tecnologia será utilizada na descoberta de conhecimento na base de dados INEP, fornecendo um modelo ideal para avaliações analíticas e preditivas.

## MODELO DIMENSIONAL

Modelos de dados multidimensionais tiram proveito dos relacionamentos inerentes nos dados, preenchendo os dados em matrizes multidimensionais chamadas de cubos de dados. Para montagem dos esquemas foi necessária a execução de algumas etapas prévias.

Como ponto de partida, foi desenvolvida a seleção dos dados disponibilizados no portal do INEP: [portal.inep.gov.br/microdados](http://portal.inep.gov.br/microdados). Em anos intercalados, o Governo Federal promove o SAEB, prova de língua portuguesa e matemática aplicada às classes do quinto e nono ano das escolas públicas. Além da prova, os alunos também respondem um questionário com questões diversas sobre seu perfil socioeconômico. Da mesma forma, professores e diretores igualmente são convidados a responder questionários relacionados ao próprio perfil e da instituição.

Os seguintes arquivos foram selecionados para o presente trabalho: a tabela TS\_ALUNO\_5EF e TS\_ALUNO\_9EF, que correlacionam a identificação do aluno, suas notas em língua portuguesa e matemática e suas respostas do formulário socioeconômico; TS\_PROFESSOR, tabela com os dados do questionário do professor; TS\_DIRETOR, respostas do questionário do Diretor de cada Escola e ,finalmente, TS\_ESCOLA, as respostas do questionário aplicado à Escola.

Inicialmente, devido ao grande volume de dados, foi inserida uma restrição a somente os dados dos alunos de escolas públicas do Estado do Rio de Janeiro (RJ). Com isso, reduziram-se a quantidade para 157.425 registros de alunos do quinto ano e 116.835 registros de alunos do nono ano respondentes. Como o objetivo do presente trabalho é demonstrar o potencial das ferramentas de BI na análise dos dados do SAEB 2015, o resultado não foi prejudicado.

Em seguida, para garantir a consistência dos dados, realizou-se a limpeza e transformação dos dados. O arquivo TS\_ALUNO\_5EF, por exemplo, possuía diversos registros sem correspondência nas tabelas TS\_PROFESSOR, TS\_DIRETOR e TS\_ESCOLA. Ou seja, alguns registros da tabela que continha as notas dos alunos não possuía correspondente nas tabelas com os dados socioeconômicos do professor, diretor e escola.

Sendo assim, não eram informações que deveriam ser consideradas, já que o objetivo do trabalho é descobrir padrões ocultos.

As tabelas de Alunos (TS\_ALUNO\_5EF e TS\_ALUNO\_9EF) possuíam dois campos referentes ao preenchimento do censo e da prova. São considerados participantes os alunos que preencheram o teste (IN\_PREENCHIMENTO\_PROVA = 1) e que estavam declarados no Censo Escolar 2015 (IN\_SITUACAO\_CENSO = 1). A taxa de participação é a razão entre o total de alunos participantes na SAEB 2015 e o total de alunos declarados no Censo Escolar 2015, que eram o público alvo dessa avaliação. Senso assim, inicialmente foram selecionados 157.425 registros com os dados de proficiência dos alunos do 5º ano do Estado do Rio (ano de 2015). Destes, 26.170 (cerca de 17% do total) possuíam um desses campo em branco ou com o valor zero (0) , registros esses que foram eliminados. Resultando, então, em 131.255 registros em conformidade para a análise.

Posteriormente, analisando as tabelas, foi chegada a conclusão que as tabelas de alunos deveriam ser divididas em duas cada uma, pois misturavam dados do questionário socioeconômico e os campos referentes às notas. Em uma tabela foram inseridos os dados referentes às notas em língua portuguesa e matemática juntamente com o campo identificador do aluno, dando origem à tabela fato. Um fato é uma observação do negócio. Muitos fatos no mundo do negócio são numéricos e aditivos. Usamos o termo “fato” para representar uma medida do negócio. Uma medida é obtida pela intersecção de todas as dimensões. Uma linha em uma tabela Fato corresponde a uma medida. Todas as medidas em uma tabela Fato devem ser do mesmo grão. De acordo com Navathe (2011, p.724),

Uma tabela de fatos pode ser imaginada como tendo tuplas, uma para cada fato registrado. Esse fato contém alguma(s) variável(is) observada(s) e a(s) identifica com ponteiros para tabelas de dimensão. A tabela de fatos contém os dados, e as dimensões identificam cada tupla nesses dados.

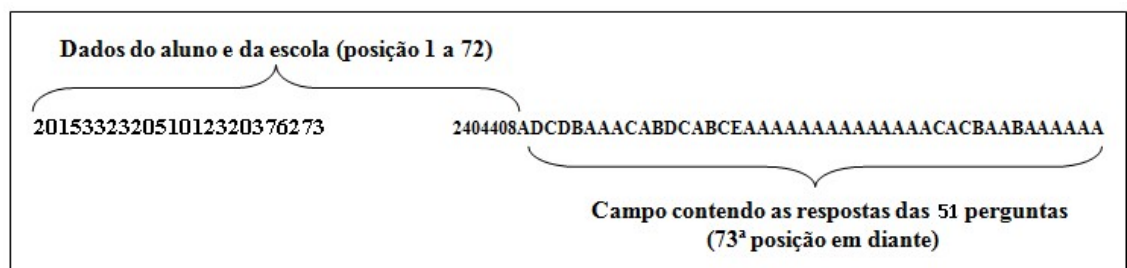
Outra tabela também foi criada com o intuito de armazenar as respostas e o identificador do aluno, originando a tabela dimensão do aluno. Para as tabelas dimensões, também, serão utilizados os dados das tabelas TS\_PROFESSOR, TS\_DIRETOR e TS\_ESCOLA. As tabelas dimensões são responsáveis por qualificar as informações disponíveis nas tabelas fato. São por elas que é possível analisar os dados sob diversas óticas. Além de, também, possibilitar a análise de diversos níveis de granularidade e detalhamento de suas informações.

No próximo passo, a tarefa de transformação mais onerosa foi converter os dados referentes às respostas dos questionários em informações que pudessem ser compreendida mais facilmente por parte do analista. Para exemplificar, a resposta da primeira questão do formulário do aluno era exibida na coluna TX\_RESP\_Q001, a resposta para a segunda pergunta, era exibida na coluna TX\_RESP\_Q002, e assim por diante. Assim como as respostas, as perguntas também precisavam ser traduzidas para melhor entendimento. Questões como o sexo do aluno, por exemplo, foram respondidas com a letra “A” e com a letra “B”, sendo “A” para sexo masculino e “B” para sexo feminino. Essa mesma lógica se replica para todas as demais questões dos questionários do aluno.

A disponibilização da base de dados realizada pelo INEP, foi realizada por meio de arquivos do tipo *Comma-Separated Values (CSV)*. Todos os campos são separados por vírgula, sendo a primeira linha destinada ao cabeçalho da tabela. Nos campos em branco, em que o respondente não apresentou resposta, não foram inseridos nenhum caractere a fim de manter o arquivo estruturado.

A Figura abaixo ilustra um exemplo de um registro referente ao questionário do aluno do 5º ano. As 72 posições iniciais possuem o código identificador do aluno, sua série, e dados sobre a escola. Da 73ª posição em diante, localiza-se o campo contendo as 51 respostas do respectivo aluno.

Figura 1 - Representação de um registro contendo os dados do questionário de um aluno do 5º ano.



Fonte: Elaborada pelo autor

Pode se verificar que essa representação não proporciona um entendimento direto das respostas do aluno. Sendo assim, foi realizado um processo de transformação dos dados onde foi substituída cada letra pela resposta correspondente com o auxílio dos questionários apresentados ao respondente, também baixados do portal INEP. Na mesma lógica, também

foram substituídos os nomes das colunas para as perguntas correspondentes. Como mostrado na figura abaixo:

Figura 2 - Registros após transformação para inserção no *data mart*

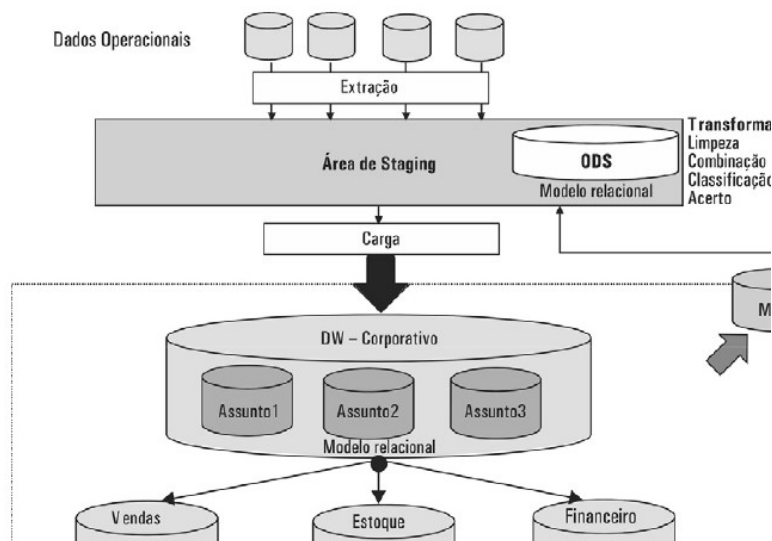
TX_RESP_Q002	Como você se considera?
a	Branco(a)
a	Branco(a)
d	Amarelo(a)
f	Indígena(a)
c	Preto(a)
c	Preto(a)
a	Branco(a)

Fonte: Elaborada pelo autor

No passo seguinte, os dados foram importados para um *data mart*. Um *data mart* é uma coleção de dados talhados para as necessidades do processamento de suporte a decisão de um departamento. Segundo Inmon (2000), é um subconjunto do DW que foi personalizado para atender as necessidades do departamento, mas pode ser usado por mais de um departamento.

De acordo com Barbieri (2011), o data warehouse poderá ser usado para fornecer os insumos para a carga dos data marts departamentais, por assuntos ou linhas de negócios, numa abordagem tipicamente *top-down*.

Figura 3 - Topologia Top-Down

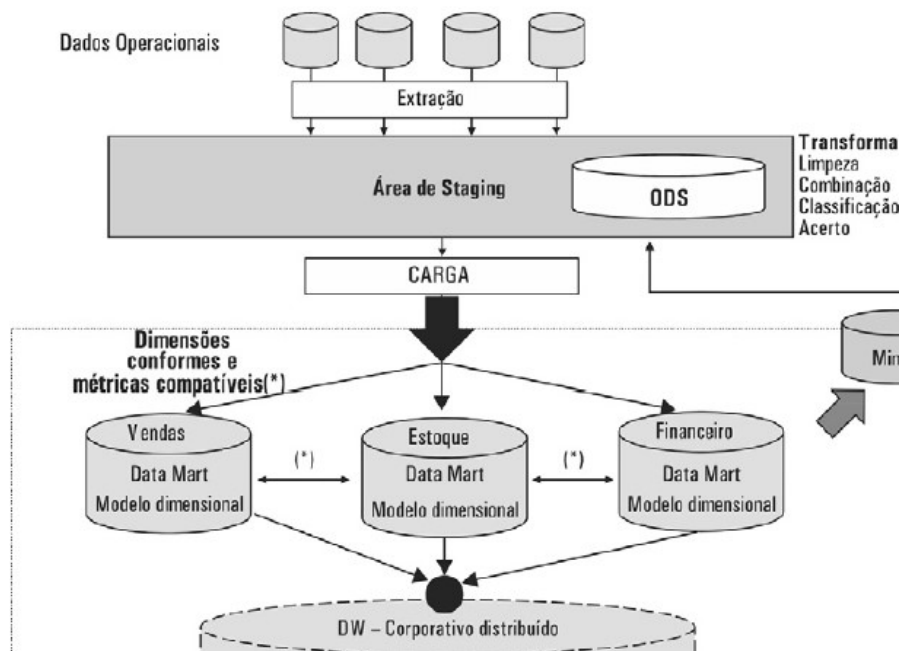


Fonte: Barbieri (2011, p.111)

Tipicamente, o dado é desnormalizado, selecionado e resumido quando passa do DW para o DM.

Outra alternativa arquitetural de construção de data warehouse, caracterizada pela criação inicial dos *data marts* integrados, que serão criados separadamente a partir dos dados operacionais trabalhados pela camada de *Staging*. Nessa arquitetura, os *data marts* serão criados de forma integrada, observando-se a conformidade entre as dimensões e a coerência das suas métricas. Essa forma, *bottom-up*, permite que o data warehouse corporativo seja formado de forma distribuída através da criação dos data marts separadamente.

Figura 4 - Topologia Bottom-Up



Fonte: Barbieri (2011, p.112)

Na modelagem do *data mart*, foi utilizado o esquema estrela, composto de uma tabela fato ligada a quatro dimensões.

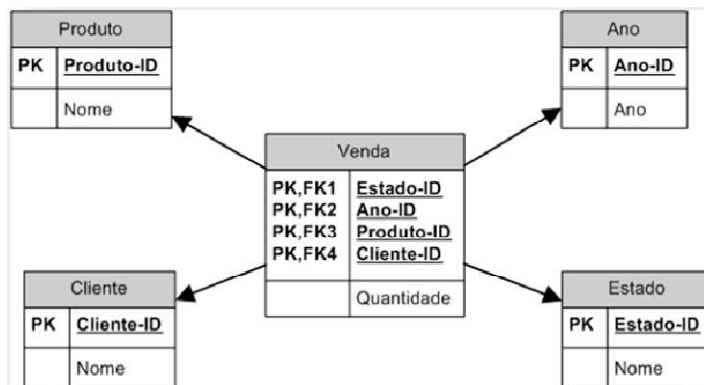


## CRIAÇÃO DO MODELO MULTIDIMENSIONAL

A Modelagem Multidimensional é uma técnica de projeto lógico que busca apresentar dados em um *framework* intuitivo e com acesso de alto desempenho. *Framework* intuitivo significa dizer que os dados são organizados de uma forma que muito se assemelha a um dos nossos principais processos cognitivos: analisar fatos por diferentes perspectivas. Alto desempenho significa que os dados são organizados e armazenados de modo que consultas analíticas podem ser executadas com uma menor quantidade de cruzamentos (*joins*) entre as tabelas. Segundo Kimball (1998), a abordagem de modelagem dimensional, utilizada para o desenvolvimento do projeto lógico, é a melhor maneira de modelar dados de suporte à decisão e oferece os melhores resultados em termos de facilidade de uso e alto desempenho.

O modelo multidimensional é inerentemente dimensional e usa o modelo relacional com algumas importantes restrições. Todo modelo dimensional é composto de uma tabela com chaves compostas, chamada Fato, e um conjunto de tabelas menores, chamadas Dimensões. Cada tabela Dimensão tem uma chave simples que corresponde exatamente a uma das partes da chave composta da tabela Fato. Essa característica estrutural, parecida com uma estrela, é chamada de *Star Schema*, ilustrada na Figura abaixo:

Figura 5 - Star Schema



Fonte: Inmon (2000)

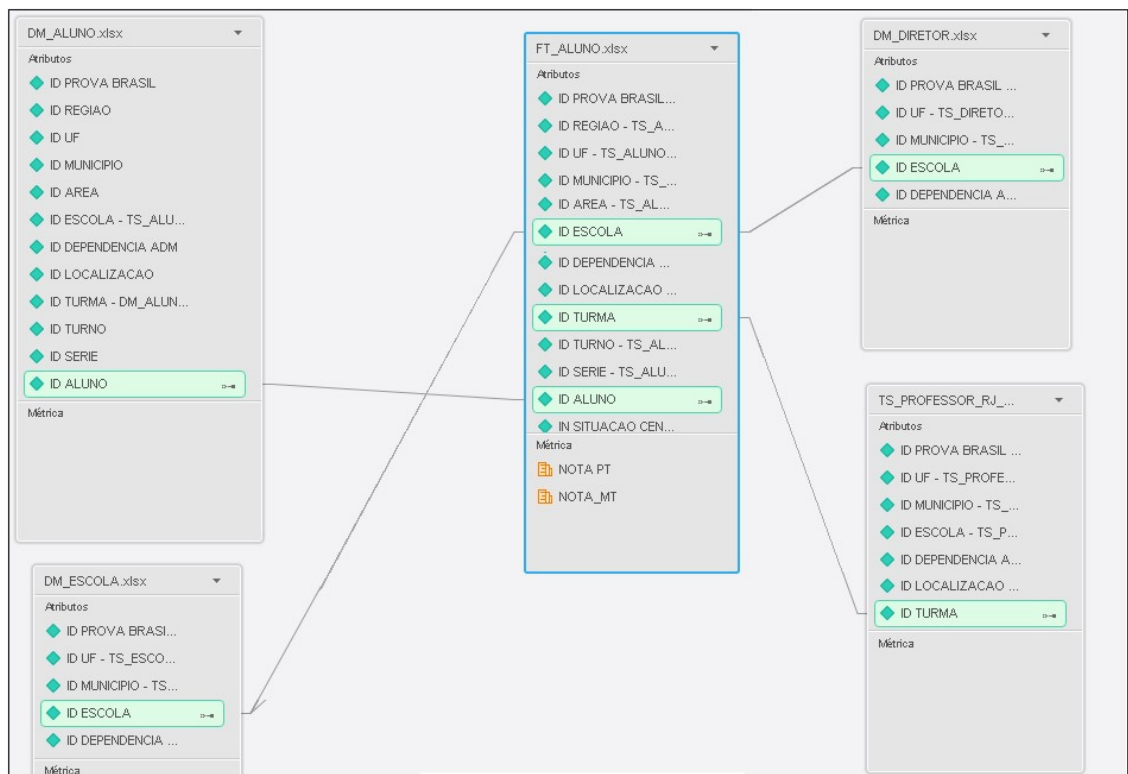
Uma tabela Fato expressa um relacionamento muitos-para-muitos, visto que ela é composta por um conjunto de, no mínimo, duas chaves estrangeiras. As tabelas Fato mais úteis contêm um ou mais fatos numéricos que ocorrem para cada combinação das chaves que identificam cada registro.

Os fatos de uma tabela Fato são numéricos e aditivos. Aditividade é crucial, uma vez que as aplicações DW quase nunca recuperam um único registro da tabela Fato. Em vez disso, elas retornam centenas ou milhares de registros de uma vez e a única coisa útil a se fazer com tantos registros é somá-los. Tabelas Dimensão, em contraste, normalmente contêm informação textual descritiva. As tabelas Dimensão são o ponto de entrada do DW.

O projeto de banco de dados do modelo dimensional tende a ser mais legível para o usuário do negócio, visto que um único diagrama ER se parte em múltiplos diagramas Dimensionais, cada um para atender requisitos analíticos de áreas de negócio diferentes.

Disponibilizar um ambiente rápido, com boa usabilidade e flexível é um grande desafio para os analistas de B.I. Para criar o ambiente de B.I., foi utilizado o *software MicroStrategy* que possui alto desempenho em análises de alto nível. Foram desenvolvidos dois cubos, um para o modelo dimensional dos dados referentes aos alunos do 5º ano e outro para os dados dos alunos do 9º ano. Essa distinção foi realizada para melhor análise e por existirem duas tabelas fatos (TS\_ALUNO\_5EF\_FT e TS\_ALUNO\_9EF\_FT). O modelo multidimensional resultante para os alunos do 5º ano, por exemplo, é exibido abaixo:

Figura 6 – Modelo Multidimensional dos dados dos alunos do 5º ano



Fonte: Elaborada pelo autor

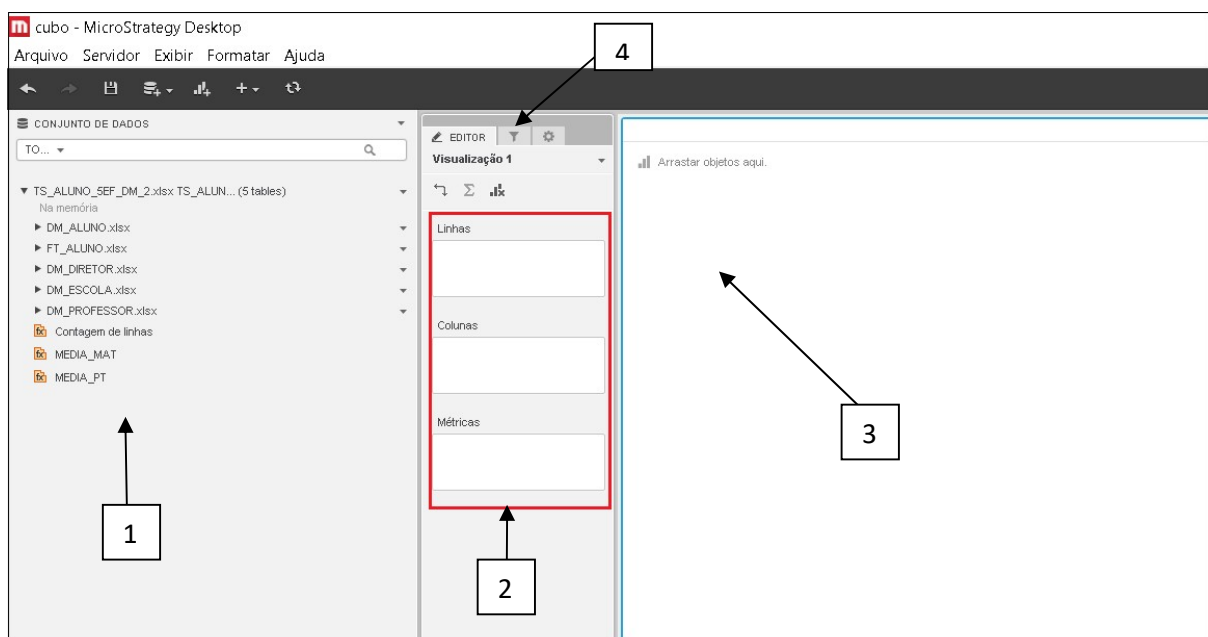
## MONTAGEM DE RELATÓRIOS

Na obtenção de indicadores de qualidade na área da educação, é primordial que se tenha disponível um ambiente de fácil uso, que permita a montagem rápida de relatórios e que possibilite a rápida extração de informações e do conhecimento. As ferramentas corretas com dados íntegros possibilitam o educador/analista diagnosticar e monitorar a evolução do quadro educacional brasileiro.

As tecnologias discutidas no presente artigo possibilitam ao analista desenvolver um trabalho de exploração de novas informações e conhecimentos a partir da grande base de dados de educação, disponibilizada pelo INEP.

Dentro dessa perspectiva, inicialmente serão apresentadas algumas figuras para ilustrar o ambiente de geração de relatórios. A Figura 16 apresenta a interface de geração de relatórios que utiliza, como base, o *data mart* com os dados do SAEB 2015 dos alunos do ensino básico do Estado do Rio de Janeiro, com foco nos resultados das médias adquiridas em Língua Portuguesa.

Figura 7 - Interface para geração de relatórios



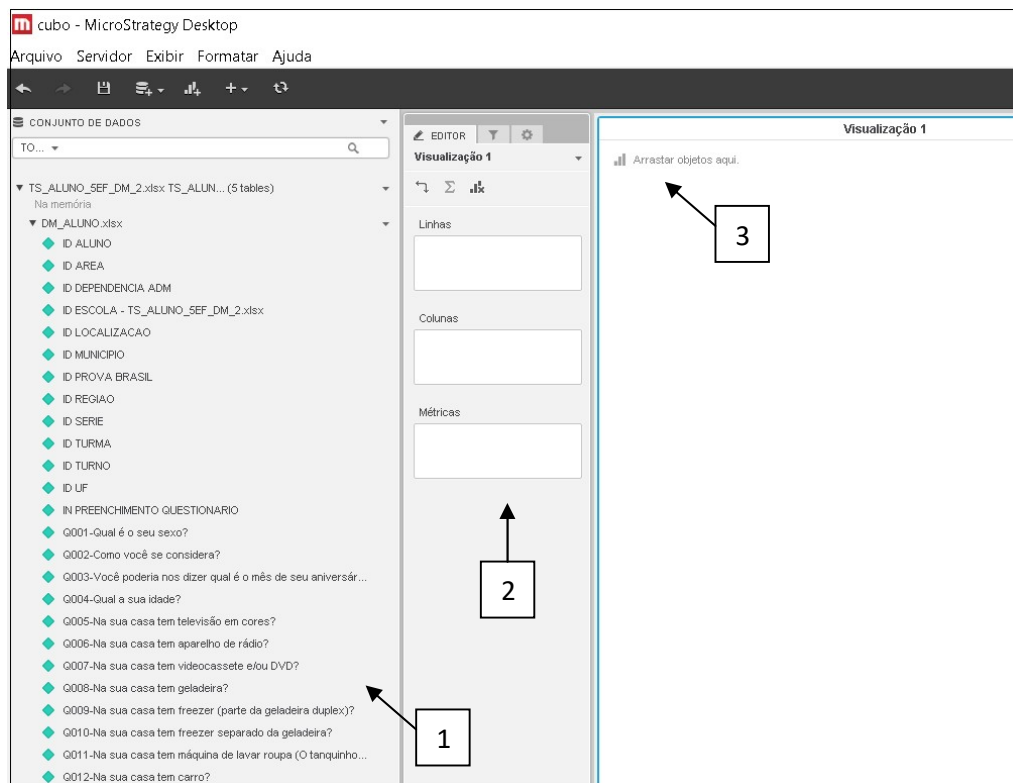
Fonte: Elaborada pelo autor

A área 1, localizada no quadrante esquerdo da tela, contém o cubo, as tabelas dimensões e fato, além das métricas ou medidas a serem empregadas. Foram criadas as dimensões DM\_ALUNO (contém as respostas do questionário socioeconômico do aluno), DM\_DIRETOR(questionário do diretor), DM\_ESCOLA(questionário da escola) e DM\_PROFESSOR(questionário do professor). As métricas MEDIA\_MAT e MEDIA\_PT representam, respectivamente, as médias das notas nas provas de Matemática e Língua Portuguesa. Elas foram criadas utilizando o agregador AVG nas notas dos alunos na tabela fato. Já a métrica Contagem de linhas foi criada pelo sistema e representa o somatório do número de registros.

A área 2, localizada na parte do meio da tela, é destinada à inserção dos campos que se deseja consultar. Nela deverão ser inseridos os atributos das linhas, das colunas e as métricas para cálculo. A área 4 possibilita a inserção de parâmetros para filtragem dos dados. Por exemplo, como o *data mart* possui os dados de todo o Estado do Rio, essa área poderia ser utilizada pelo usuário para fazer a filtragem dos dados somente do município de Arraial. Finalmente, na parte inferior à direita, localiza-se o espaço onde são construídos os relatórios (área 3).

As dimensões apresentadas na área 1 podem ser expandidas na janela. Ao efetuar a expansão (clique no botão de seta, localizado ao lado esquerdo do nome de cada dimensão), as perguntas dos questionários podem ser visualizadas. A Figura abaixo apresenta a janela com a dimensão correspondente aos alunos do 5º ano expandida (DM\_ALUNO).

Figura 8 - Expansão da dimensão correspondente aos alunos do 5º ano do ensino fundamental



Fonte: Elaborada pelo auto

Nota-se que os atributos da dimensão estão nomeados a partir do seu número, seguido da questão, propriamente dita. Para criar qualquer relatório, basta clicar e arrastar os campos existentes na área 1 para a área 2, onde é feito o cruzamento dos dados. Ressalta-se que a criação de qualquer novo relatório, envolvendo, uma, duas, ou mais dimensões, pode ser feita em poucos segundos. Os valores agregados (médias e totais) são calculados automaticamente e apresentados para cada dimensão considerada na geração dos relatórios.

A Figura abaixo apresenta um exemplo de relatório que possui as informações das médias dos alunos do 5º ano do ensino fundamental do Estado do Rio de Janeiro, com visualização a partir da perspectiva do aluno (se ele já foi ou não reprovado).

Figura 9 - Exemplo de relatório gerado pela ferramenta

Q045-Você já foi reprovado?	MEDIA_MAT	MEDIA_PT	Contagem de linhas
Não.	225,40	216,22	63,23%
Sim, uma vez.	202,55	190,44	22,59%
Sim, duas vezes ou mais.	202,07	188,47	9,15%
Em branco.	201,56	187,42	5,01%
<b>Total</b>	<b>216,90</b>	<b>206,41</b>	<b>100,00%</b>

Fonte: Elaborada pelo autor

Para obtenção deste relatório, foi necessário, inicialmente, expandir a dimensão Aluno, selecionar o atributo referente à questão 45 do questionário do aluno (Q045-Você já foi reprovado?) e arrastá-lo para a área 2 (de cruzamento). Em seguida, as métricas MEDIA\_MAT, MEDIA\_PT e Contagem de linhas foram arrastadas para, também, a região 2.

Fazendo uma análise inicial do relatório apresentado, pode-se perceber que alunos que nunca foram reprovados tendem a obter melhores resultados (maiores médias) nas provas.

Outro relatório gerado como exemplo, é o relatório que evidência o grau de incentivo e interesse dos pais com os alunos. Podemos cruzar os dados da pergunta “Seus pais ou responsáveis incentivam você a ler?” (Questão 29) com os dados da pergunta “Com qual frequência seus pais, ou responsáveis por você, vão à reunião de pais?” (Questão 26). Abaixo o resultado do relatório:

Figura 10 - Relatório da freqüência dos pais, ou responsáveis vão à reunião de pais.

Metrics	MEDIA_MAT			MEDIA_PT			Contagem de linhas		
	Sim	Não	Em branco	Sim	Não	Em branco	Sim	Não	Em branco
Q026-Com qual frequência seus pais, ou responsáveis por você, vão à reunião de pais?									
Sempre ou quase	220,86	215,02	206,68	211,64	201,54	193,27	51,31%	2,33%	0,78%
De vez em quando.	215,97	212,26	200,20	205,46	198,59	187,30	29,55%	2,22%	0,49%
Nunca ou quase nunca.	209,90	202,00	191,34	198,66	184,86	174,21	7,77%	1,32%	0,22%
	196,39	186,59	202,45	181,32	160,59	188,79	0,87%	0,12%	2,95%

Fonte: Elaborada pelo autor

Podemos notar que as médias maiores alcançadas foram pelos alunos que os pais ou responsáveis estão sempre ou quase sempre presentes nas reuniões de pais e recebem dos mesmos incentivos para leitura. Também é possível constatar a expressiva variação negativa das médias dos alunos em que os pais nunca ou quase nunca se fazem presentes nas reuniões de pais. Diferentemente, a variação de média dos alunos que não recebem o incentivo de pais ou responsáveis não se faz tão expressiva quanto a anterior, apesar de existir.

Para Garcia (2011), é substancial a importância do respeito e acolhimento das diferenças culturais, da inclusão do conhecimento familiar no trabalho educativo e do estabelecimento de canais de comunicação. O contato com a diversidade, além de ampliar os horizontes das crianças e do professor, fundamenta uma postura ética e de - mocrática nas relações humanas. Sendo assim, a escola deve possibilitar diferentes momentos de contato com as famílias, como reuniões periódicas com os pais de um mesmo grupo de alunos. De modo geral, indica que a postura da escola deve ser de abertura, sintetizada na afirmação de Brasil (1998, p.77): “as instituições de educação infantil, por intermédio de seus profissionais, devem desenvolver a capacidade de ouvir, observar e aprender com as famílias”.

Evidentemente, essa é apenas uma constatação inicial, a partir de uma primeira análise, que poderia ser aprofundada para identificar causas, fatores e condições que levaram a esses resultados. De qualquer forma, pode-se concluir que a ferramenta permite a fácil geração de relatórios que descortinem novas perspectivas aos educadores; que permitam,

conforme já mencionado, a descoberta de conhecimento a partir das bases de dados de educação.

Um fator relevante que pode ser investigado, também, é se o nível de escolaridade dos pais ou responsáveis dos alunos influencia ou não nas médias finais. Para isso, foi gerado um relatório sobre as respostas dos alunos do 9º ano e suas médias em Língua Portuguesa:

Figura 11 – Relatório do grau de escolaridade da mãe ou responsável dos alunos de 9º ano

Q19 - Até que série sua mãe, ou a mulher responsável por você, estudou?	MEDIA_PT	Porcentagem do total global (Contagem de linhas)
Completo a Faculdade.	252,72	29,84%
Completo o Ensino Médio, mas não completou a Faculdade.	252,21	8,35%
Completo a 8.ª série/9.º ano do Ensino Fundamental, mas não completou o E. M.	249,94	12,97%
Completo a 4.ª série/5.º ano, mas não completou a 8.ª série/9.º ano do E.F.	246,84	13,72%
Não completou a 4.ª série/5.º ano do Ensino Fundamental.	239,37	8,82%
Não sei.	237,23	22,20%
	234,92	2,97%
Nunca estudou.	230,74	1,13%
<b>Total</b>	<b>28,37</b>	<b>100,00%</b>

Fonte: Elaborada pelo autor

Figura 12 – Relatório do grau de escolaridade do pai ou responsável dos alunos de 9º ano

Q23 - Até que série seu pai, ou o homem responsável por você, estudou?	MEDIA_PT	Porcentagem do total global (Contagem de linhas)
Completo o Ensino Médio, mas não completou a Faculdade.	253,67	22,28%
Completo a Faculdade.	253,41	7,21%
Completo a 4.ª série/5.º ano, mas não completou a 8.ª série/9.º ano do E.F.	249,52	10,42%
Completo a 8.ª série/9.º ano do Ensino Fundamental, mas não completou o E. M.	246,75	11,46%
Não sei.	241,51	35,53%
Não completou a 4.ª série/5.º ano do Ensino Fundamental.	240,03	8,03%
	238,58	3,27%
Nunca estudou.	231,49	1,80%
<b>Total</b>	<b>28,37</b>	<b>100,00%</b>

Fonte: Elaborada pelo autor

Pode-se perceber que alunos que possuem pais ou responsáveis com maior grau de escolaridade, agregam um melhor desempenho. Para Pinheiro (2015), as diferenças educacionais no Brasil tendem a ser transmitidas de forma direta para os resultados escolares das gerações futuras. Quanto maior o nível de escolaridade dos pais maior a proficiência escolar do filho. Quanto a isso, viu-se que essa maior escolaridade do pai e/ou da mãe se traduz num maior desempenho do aluno através de canais que podem ser: na escolha de melhores colégios, incentivá-lo ao estudo ou, até mesmo, inibi-lo de trabalhar fora durante a adolescência.



Pode-se perceber que o relatório contém todas as características, já apresentadas, da utilização da tecnologia de visão multidimensional de dados. Mais ainda, fica clara a facilidade de manipulação e criação de relatórios de auxílio à decisão, mesmo que se feito por usuários sem maiores conhecimentos técnicos na área de computação ou bancos de dados. A construção de relatórios que representem os resultados dos alunos nas provas de Língua Portuguesa, abrangendo informações que caracterizam os perfis de alunos, professores, diretores e escolas, seja isoladamente, seja em diferentes combinações das dimensões, é possibilitada a partir de interação com uma interface gráfica simples que provê o cálculo automatizado de médias e totais de alunos, com resposta rápida e eficiente.

## CONCLUSÕES

O Sistema de Avaliação da Educação Básica -Saeb, criado, em 1990, tem como objetivo avaliar a qualidade do ensino ministrado nas escolas de ensino fundamental das redes públicas. Seu propósito é alcançado por meio de questionários socioeconômicos e provas de conhecimento de Língua Portuguesa e Matemática, aplicadas aos alunos de 5º ano, 9º ano e 3º ano do ensino médio.

O presente trabalho teve como objetivo desenvolver um modelo multidimensional, apresentando o poder das tecnologias utilizadas. A primeira etapa do trabalho contou com a criação de um *data mart*, que foi fundamental para a utilização consistente das ferramentas de mineração e visão multidimensional nos microdados do SAEB 2015. As tarefas de seleção, limpeza, transformação e carga dos dados formam o alicerce que possibilitou a posterior descoberta de conhecimento em bancos de dados (*KDD*). Esse processo envolveu elevado nível de trabalho e tempo, mas garantiu posterior facilidade e agilidade para geração de relatórios de apoio à decisão.

No decorrer do trabalho, foram desenvolvidos diversos relatórios, levando em consideração, por exemplo, o grau de escolaridade de professores, pais e responsáveis. Para Calhado (2012), a variável “grau de escolaridade dos pais” apresenta um alto grau de correlação com as melhores notas no SAEB. O autor realizou uma pesquisa comparando os resultados das redes de ensino estadual e municipal do município de Ribeirão Preto/ SP com o desempenho dos alunos. Também, destacou em seu trabalho que: o Índice Socioeconômico

das Escolas (ISE) consegue discriminar satisfatoriamente as características socioeconômicas dos alunos atendidos pelas escolas e podem ser utilizados para análises dessa natureza.

Nos relatórios gerados, também, foi possível constatar que, alunos que realizavam freqüentemente atividades como cópia de extensos textos do quadro obtiveram uma média menor se comparado aos alunos que freqüentemente exerciam atividades, por exemplo, de discussões de contos, crônicas, poesias ou romances. De acordo com Ferreira (2013), essa constatação é válida, pois a cópia de texto por ser algo meramente mecânico, pode acabar fazendo com que os alunos sejam moldados, ou seja, condicionados a copiar. A prática da cópia pode ser relacionada com a concepção de ensino comportamentalista, na qual, segundo Mizukami (1986) o professor molda o comportamento do aluno através de exaustivas atividades repetitivas.

O artigo desenvolvido abordou alguns possíveis relatórios sob os dados do SAEB 2015, atingindo assim, o objetivo proposto de desenvolver alguns indicadores do ensino fundamental em escolas do Rio de Janeiro.

Como sugestões para trabalhos futuros seria válido a implantação das ferramentas de B.I. em toda a base de dados do SAEB 2015 e avaliar a possibilidade de complementar o estudo com outras base de dados como o Censo Escolar e ENCCEJA.

## REFERÊNCIAS

ALVES, T.; PASSADOR, C.S. **Educação Pública no Brasil: condições de oferta, nível socioeconômico dos alunos e avaliação.** São Paulo: Annablume.

BARBIERI, Carlos. **Business Intelligence: modelagem e qualidade.** Rio de Janeiro: Elsevier, 2011.

BRASIL Ministério da Educação e do Desporto. Secretaria de Educação Fundamental. Referencial curricular nacional para a educação infantil. Brasília, 1998.

CALHADO, G.C; PASSADOR C.S. **INFRAESTRUTURA ESCOLAR, PERFIL SOCIOECONÔMICO DOS ALUNOS E QUALIDADE DA EDUCAÇÃO PÚBLICA EM RIBEIRÃO PRETO/SP.** Ribeirão Preto, 2012.

DATE, C. J. **Introdução a sistemas de banco de dados.** 8 ed. São Paulo: Campus, 2003.

ELMASRI, R.; NAVATHE, S.B. **Sistemas de banco de dados**. 6 ed. São Paulo: Pearson, 2011.

FERREIRA, L. G. F. **A presença da cópia na escola: Estratégia para ensinar ou ritual mecânico?**. São Paulo, 2013.

GARCIA, H. H. G. O.; MACEDO, L. **Reuniões de Pais na Educação Infantil: Modos de Gestão**. Cadernos de Pesquisa. 2011.

INMON, W.H. **Como construir o data warehouse**. 2 ed. Rio de Janeiro: Campus, 1997.

MEC/INEP. **Microdados do Sistema de Avaliação da Educação Básica 2015 - Manual do Usuário**. Brasília: Ministério da Educação - INEP, 2015.

MIZUKAMI, M. G. N. **Ensino: as abordagens do processo**. São Paulo, 1986.

NAMEN, Anderson Amendoeira ; SOARES, Ana Carolina Soares. **UTILIZAÇÃO DE CLASSIFICADORES BAYESIANOS PARA ANÁLISE DOS FATORES RELACIONADOS AO DESEMPENHO DOS ALUNOS DA QUARTA SÉRIE DO ENSINO FUNDAMENTAL**. XV Encontro de Modelagem Computacional, 2012, Uberlândia - MG. XV Encontro de Modelagem Computacional e III Encontro de Ciência e Tecnologia de Materiais, 2012.

NAMEN, Anderson Amendoeira ; SOARES, Ana Carolina Soares . **Mineração de Dados Relacionados ao Aprendizado de Língua Portuguesa, Um Estudo Exploratório**. XIV Encontro de Modelagem Computacional, 2011, Nova Friburgo. Anais do XIV Encontro de Modelagem Computacional XIV EMC e do II Encontro de Ciência e Tecnologia de Materiais II ECTM. Rio de Janeiro: Rede SIRIUS - UERJ, 2011.

PINHEIRO, D. M. S. **Efeito da Educação dos Pais Sobre o Desempenho Dos Filhos: Uma Simulação com Países Similares ao Brasil**. Rio de Janeiro, 2015.

KIMBAL, R. et al. **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses**. New York: John Wiley & Sons, 1998.

RODRIGUES, I. C. **PERFIL DA ESCOLA PÚBLICA NA VISÃO DE PAIS, ALUNOS, PROFESSORES. UM OLHAR SOBRE RELACIONAMENTO PROFESSOR/ ALUNO**. Paraná, 20