



CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS

COMUNIDADE EVANGÉLICA LUTERANA "SÃO PAULO"
Recredenciado pela Portaria Ministerial nº 3.607 - D.O.U. nº 202 de 20/10/2005

JÚLIO CÉSAR PEREIRA DE OLIVEIRA

**DESENVOLVIMENTO DE UM SISTEMA DE APOIO A DECISÃO PARA O
PROGRAMA BOLSA UNIVERSITÁRIA**

PALMAS

2011



CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS

COMUNIDADE EVANGÉLICA LUTERANA "SÃO PAULO"
Recredenciado pela Portaria Ministerial nº 3.607 - D.O.U. nº 202 de 20/10/2005

JÚLIO CÉSAR PEREIRA DE OLIVEIRA

DESENVOLVIMENTO DE UM SISTEMA DE APOIO A DECISÃO PARA O PROGRAMA BOLSA UNIVERSITÁRIA

Trabalho apresentado como requisito parcial da disciplina de Trabalho de Conclusão de Curso (TCC) do curso de Sistemas de Informação, orientado pelo Professor Mestre Fernando Luiz de Oliveira.

Palmas

2011

JÚLIO CÉSAR PEREIRA DE OLIVEIRA

**DESENVOLVIMENTO DE UM SISTEMA DE APOIO A DECISÃO PARA O
PROGRAMA BOLSA UNIVERSITÁRIA**

Trabalho apresentado como requisito parcial da disciplina de Trabalho de Conclusão de Curso (TCC) do curso de Sistemas de Informação, orientado pelo Professor Mestre Fernando Luiz de Oliveira.

Aprovada em _____ de 2011.

BANCA EXAMINADORA

Prof. M.Sc. Fernando Luiz de Oliveira
Centro Universitário Luterano de Palmas

Prof. M.Sc. Cristina D'Ornelas Filipakis
Centro Universitário Luterano de Palmas

Prof. M.Sc. Fabiano Fagundes
Centro Universitário Luterano de Palmas

Palmas

2011

DEDICATÓRIA

“Post partum virgo inviolatta permansisti, Deo genitrix intercede pro nobis”.
Santo Antonio de Santana Galvão (Frei Galvão)

*Ave, Maris Stella,
Dei mater alma,
Atque semper Virgo,
Felix caeli porta.*

*Sumens illud Ave,
Gabrielis ore,
Funda nos in pace
Mutans Evae nomen.*

*Solve vincla reis,
Profer lumen caecis,
Mala nostra pelle,
Bona cuncta posce.*

*Monstra te esse Matrem,
Sumat per te preces,
Qui pro nobis natus
Tulit esse tuus.*

*Virgo singularis,
Inter omnes mitis,
Nos, culpis solutos,
Mites fac et castos.*

*Vitam praesta puram,
Iter para tutum:
Ut, videntes Jesum,
Semper collaetemur.*

*Sit laus Deo Patri,
Summo Christo decus
Spiritus Sancto,
Tribus honor unus. Amen.*

AGRADECIMENTOS

Agradeço a Deus pelo dom da vida e seu amor que me fez superar as dificuldades durante todos os anos da graduação. Agradeço por sua providência em toda a caminhada.

Aos meus pais que muito abdicaram em suas vidas pela minha, e também por saber o que significa essa graduação para vocês. Aos demais familiares, que acompanharam todas as dificuldades enfrentadas e testemunharam que com Deus tudo é possível.

Agradeço a minha amada esposa Adna pela extrema compreensão, paciência, carinho, amor e incentivos oferecidos. Sua confiança em meu sucesso foi fundamental para que eu pudesse continuar nos momentos difíceis.

Agradeço aos meus filhos Júlio Vítor e José Lucas, pois a presença de vocês sempre me faz prosseguir.

Ao meu orientador Fernando, pela atenção e paciência dispensada durante todo o trabalho de conclusão.

A todos os professores: Cristina, Mária, Parcilene, Fabiano, Edeílson, Jackson, Ricardo Marx, Michael Schuenck e Andrés Lázaro.

Aos amigos que comigo formaram a “turma do nível”, Fábio, Hugo e Iroilton.

Aos amigos e compadres Charles e Maria Helena que entraram na minha vida e da minha família em um momento delicado e foram expressões da misericórdia de Deus.

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO DE LITERATURA.....	14
2.1	SISTEMAS DE INFORMAÇÃO (SI)	14
2.2	TIPOS DE SISTEMAS DE INFORMAÇÃO	16
2.2.1	Sistema de Apoio à Decisão (SAD)	20
2.3	<i>DATA WAREHOUSE</i>	24
2.3.1	Características de um <i>Data Warehouse</i>	28
2.3.1.1	Orientado a Assuntos	28
2.3.1.2	Integrado	28
2.3.1.3	Variante no Tempo	29
2.3.1.4	Não Volátil	30
2.3.1.5	Granularidade.....	30
2.3.1.6	Credibilidade dos Dados	32
2.3.2	Arquitetura de Desenvolvimento	32
2.3.2.1	Arquitetura Global	33
2.3.2.2	Arquitetura Independente	33
2.3.2.3	Arquitetura Integrada.....	33
2.3.3	Abordagem de Implementação do DW	34
2.3.3.1	Abordagem <i>Top-Down</i>	34
2.3.3.2	Abordagem <i>Bottom-Up</i>	36
2.3.4	Data Mart	37
2.4	MODELAGEM DE <i>DATA WAREHOUSE</i>	38
2.4.1	Tabela de Fatos	40
2.4.2	Tabela de Dimensões	42
2.4.3	Modelo Estrela	42
2.4.4	Modelo <i>Snowflake</i>	43
2.5	PROCESSO ETL (<i>EXTRACTION, TRANSFORMATION AND LOAD</i>)	45
2.6	ON-LINE ANALYTICAL PROCESSING (OLAP)	47
2.6.1	Operações Olap.....	50
2.6.2	Arquitetura	50
3	MATERIAIS E MÉTODOS.....	53
3.1	<i>HARDWARE</i>	53
3.2	<i>SOFTWARE</i>	53
3.3	METODOLOGIA DO TRABALHO	53
3.3.1	Definição da Abordagem de Desenvolvimento do <i>Data Warehouse</i>	55
3.3.2	Definição do modelo dimensional	55
3.3.3	Definição do modelo de Sistema de Apoio à Decisão	56
4	RESULTADOS E DISCUSSÕES	57
4.1	DOMÍNIO	57
4.2	ESTRUTURA E EXEMPLOS DOS DADOS OPERACIONAIS.....	58
4.2.1	Processo ETL	61
4.3	MODELO DIMENSIONAL	63
4.3.1	Tabela Dimensão Curso	63
4.3.2	Tabela Dimensão Renda	64

4.3.3 Tabela Dimensão Tempo.....	65
4.3.4 Tabela Dimensão Família	66
4.3.5 Tabela Dimensão Bolsista	67
4.3.6 Tabela Fato.....	68
4.4 DESENVOLVIMENTO DO PROJETO <i>ANALYSIS SERVICES</i> E OLAP	69
4.4.1 Definição dos Cubos.....	73
4.4.2 Processando e Visualizando Cubos	80
4.4.3 Criando <i>Measure</i> (medidas) de processamento no cubo	84
4.4.4 Partições.....	89
4.4.5 Carregando as informações e gerando gráficos	94
4.4.6 Gráficos Gerados a partir da Base de Dados OLAP	102
4.4.6.1 Perfil Sócio Econômico.....	108
4.4.7 Conclusões	110
5 CONSIDERAÇÕES FINAIS	112
5.1 TRABALHOS FUTUROS	113
6 REFERÊNCIAS BIBLIOGRÁFICAS	114

LISTA DE FIGURAS

Figura 1: Tipos de Sistemas de Informação.....	16
Figura 2: Estrutura de um SIG.....	18
Figura 3: Característica de um SIE	19
Figura 4: Apresentação da arquitetura de um SAD.....	22
Figura 5: Componentes de BI ou SAD	24
Figura 6: Níveis de granularidade	31
Figura 7: Método Top-down	35
Figura 8: Método Bottom-up.....	36
Figura 9: Exemplo de Dimensões e Fato	39
Figura 10: Modelo dimensional de negócio.....	40
Figura 11: Exemplo de tabela fato.....	41
Figura 12: Exemplo de Modelo Estrela	43
Figura 13: Exemplo de Modelo Snowflake	44
Figura 14: Atuação de OLAP e OLTP em SI.....	49
Figura 15: Dados disponibilizados sobre programa de governo.....	59
Figura 16: Base de dados MS Access criada para programa de governo.	60
Figura 17: Dados armazenados em Microsoft Access.	61
Figura 18: Modelo Relacional.....	62
Figura 19: Modelagem dimensional (Modelo Estrela).	63
Figura 20: Configuração da dimensão tempo sobre a tabela D_Tempo.	66
Figura 21: Configurando conexão com base de dados para projeto.	70
Figura 22: Criação do <i>Data Source</i>	71
Figura 23: Seleção das tabelas para <i>Data Source</i>	71
Figura 24: Nomeação do <i>Data Source</i>	72
Figura 25: Apresentação do Data Source View.....	72
Figura 26: Definição do tipo e construtor a ser utilizado no cubo.	73
Figura 27: Seleção da fonte de dados.....	74
Figura 28: Identificação da tabela de fatos e dimensão.	75
Figura 29: Definição das colunas que serão usadas como medidas.....	76
Figura 30: Definição das colunas que serão usadas como dimensão.....	76
Figura 31: Construtor do cubo.....	77
Figura 32: Operação <i>Business Intelligence</i>	78
Figura 33: Configurando atributos da dimensão BI.	79
Figura 34: Criação de operações BI.....	80
Figura 36: Visualização do cubo.	81
Figura 37: Acesso as configurações de processamento.	82
Figura 38: Processando o projeto.	82
Figura 39: Browse do cubo com eixos e colunas e medidas.....	83
Figura 40: Criação de medidas.	84
Figura 41: Seleção da operação da medida.....	85
Figura 42: Processamento das medidas.	86
Figura 43: Medidas criadas sobre as dimensões	87
Figura 44: <i>Dimension Usage</i>	88
Figura 45: Definição do tipo de armazenamento.....	89
Figura 46: Seleção do tipo de armazenamento.....	90
Figura 47: Acessando projeto por SQL Server.....	90
Figura 48: Selecionando projeto para processamento.....	91

Figura 49: Processando projeto por completo.....	92
Figura 50: Status ao final do processamento do projeto.	93
Figura 51: Tela do Microsoft Excel.	94
Figura 52: Seleção do tipo de base de dados para conexão.	95
Figura 53: Configurando conexão com servidor.....	95
Figura 54: Seleção da base de dados e do objeto de conexão.....	96
Figura 55: Seleção de como importar dados e em quais células começar.	96
Figura 56: Apresentação de área dinâmica para emissão de relatórios.....	97
Figura 57: Visão geral de relatório emitido.	98
Figura 58: Relatório emitido com dados numéricos.	99
Figura 59: Gráfico gerado a partir dos dados numéricos.	100
Figura 60: Tabela de montagem de relatórios com base em OLAP.....	101
Figura 61: Mensagem de pesquisa em base de dados OLAP para emissão de relatórios dinâmicos.	102
Figura 62: Total de bolsistas contemplados por ano.	103
Figura 63: Percentual de área de conhecimento.....	104
Figura 64: Bolsistas por ano, área de conhecimento e IES.....	105
Figura 65: Total de bolsistas por UF.	106
Figura 66: Bolsistas oriundos do estado do Tocantins preferência por área de conhecimento.	107
Figura 67: Identificação do perfil dos bolsistas.....	108
Figura 68: Valores empregados em bolsas por área de conhecimento.	109
Figura 69: Arquitetura do SAD.	110

LISTA DE TABELAS

Tabela 1: Resumo das diferenças entre sistema OLTP e <i>Data Warehouse</i>	25
Tabela 2: Diferenças entre OLTP e OLAP.	48
Tabela 3: Representação dos atributos da tabela D_Curso.....	64
Tabela 4: Representação dos atributos da tabela D_Renda.....	64
Tabela 5: Representação dos atributos da tabela D_Tempo	65
Tabela 6: Representação dos atributos da tabela D_Familia	67
Tabela 7: Representação dos atributos da tabela D_Bolsista	67
Tabela 8: Representação dos atributos da tabela F_PerfilBolsista	68

LISTA DE ABREVIATÖES

BI – *Business Intelligense*

DM – *Data Mart*

DMX – *Multidimensional Expressions*

DOLAP – *Desktop On Line Analytical Processing*

DW – *Data Warehouse*

ER – Entidade Relacionamento

ETL – *Extract, Transformation and Load*

HOLAP – *Hibrid On Line Analytical Processing*

MOLAP – *Multidimensional On Line Analytical Processing*

ODBC -- *Open Data Base Connectivity*

ODS – *Operational Data Storage*

OLAP – *On-line Analytical Processing*

OLTP - *On-line Transaction Processing*

ROLAP – *Relational On Line Analytical Processing*

SAD -- Sistema de Apoio à Decisão

SAE -- Sistema de Apoio à Estratégia

SGBM – Sistema de Gerência de Banco de Modelos

SGDB – Sistema de Gerência de Banco de Dados

SI – Sistemas de Informação

SIG -- Sistema de Informação Gerencial

SQL – *Structured Query Language*

SSIS – *SQL Server Integration Service*

TI - Tecnologia da Informação

WOLAP – *Web On Line Analytical Processing*

RESUMO

Diante da competitividade do mercado e da presença de Sistemas de Informação (SI) no auxílio administrativo e estratégico das organizações, dois fatores passaram a ter maior atenção: o conhecimento e a informação. Esses sistemas têm responsabilidades de coletar, organizar, analisar, compartilhar e monitorar as informações que oferecem suporte a gestão de negócios e, concomitantemente, ao processo decisório. Um tipo de SI seria o Sistema de Apoio à Decisão (SAD), que pode ser desenvolvido com base em uma metodologia avançada para armazenamento dos dados, baseado em ambiente *Data Warehouse* (DW), que permite utilizar recursos de limpeza e armazenamento dos dados relevantes para o processo de extração de conhecimento. Para auxiliar no processo de análise os dados utiliza-se de ferramentas OLAP, que possibilita a emissão de relatórios sobre os dados explorados. Esse trabalho tem por objetivo realizar a revisão de literatura sobre as tecnologias citadas, bem como suas especificações e os conceitos envolvidos, para então auxiliar na construção de um SAD que vai auxiliar no processo decisório, abrangendo um programa de governo para disponibilização de bolsas de ensino superior, realizado pelo estado do Tocantins.

Palavras-chave: *Data Warehouse*, Tomada de decisão, OLAP.

1 INTRODUÇÃO

Diante da competitividade do mercado, as organizações acabam gerando um grande volume de dados que precisam ser trabalhados para gerar algum conhecimento para as mesmas. Nesse cenário, os Sistemas de Apoio à Decisão auxiliam no sentido de organizar esses dados de forma estratégica considerando sua relevância para os diversos setores da organização (GOMES, 2010, p.14). Para os SI que atuam no nível estratégico é comum utilizarem bases de dados que possibilitem na construção um histórico das informações. Para tanto, utilizam a tecnologia de *Data Warehouse* (DW), que servem para integrar e gerenciar o fluxo de informações a partir dos bancos de dados e fontes de dados da organização. O ambiente de DW permite que os dados armazenados sejam acessados e possibilita o entendimento para utilização no suporte aos SADs (ESPIRITO SANTO, 2006, p. 16). Esse trabalho tem o objetivo de construir um Sistema de Apoio à Decisão baseado em *Data Warehouse* para o programa bolsa universitária do Governo do Estado do Tocantins.

Esse trabalho está estruturado da seguinte forma: o próximo capítulo contém a revisão de literatura, onde são apresentados os conceitos relacionados a sistemas de informação, bem como os seus tipos disponíveis e aplicações, e ainda, os conceitos de *Data Warehouse* e sua aplicação. No terceiro capítulo são apresentados os materiais e métodos, onde são apresentadas as ferramentas utilizadas e a metodologia para o desenvolvimento do mesmo. No quarto capítulo são apresentados os resultados e discussões, contendo os passos para construção e a análise dos relatórios emitidos, que auxiliarão a coordenação do programa na tomada de decisão. No quinto capítulo são apresentadas as considerações finais e por fim, as referências bibliográficas que auxiliaram na construção deste trabalho.

2 REVISÃO DE LITERATURA

Neste capítulo serão abordadas as definições de Sistemas de Informação (SI), os tipos existentes, assim como ferramentas e tecnologias como *Data Warehouse* e OLAP. Estas tecnologias, por ser a base deste trabalho, serão detalhadas, abordando suas especificações, características e arquitetura.

2.1 Sistemas de Informação (SI)

O termo Sistema de Informação é utilizado para definir um sistema que coleta, organiza, processa e transmite dados a fim de auxiliar no processo decisório de um ambiente corporativo (REZENDE, 2005, p. 26). Assim, a partir da exploração dos dados da organização e por meio de um sistema eficiente para extrair a informação, é possível a compreensão de fatos ou situações na organização. Isto porque, segundo Padoveze (2000, p 43), “a informação é o dado que foi processado e armazenado de forma compreensível para seu receptor e que apresenta valor real percebido para suas decisões correntes ou prospectivas”.

Considerando a presença da informação e os esforços para possuí-la, é possível verificar o quão valiosa é a informação para a organização. Assim, percebendo que da informação dependem os resultados, segundo Padoveze (2000, p. 44), o valor da informação está relacionado a:

- redução da incerteza no processo de tomada de decisão;
- relação dos benefícios gerados pela informação versus o curso de produção;
- aumento da qualidade da decisão.

Então, para mensurar seu valor, é necessário considerar sua atuação e o modo como reduz as dúvidas pertinentes ao domínio durante o processo decisório, pois conseqüentemente, aumenta a qualidade da decisão e reduz riscos. Assim sendo, o sistema é a base que facilita a identificação dos dados e informações a serem explorados e assim, traçar metas e objetivos.

Considerando o contexto corporativo, com forte ênfase no mercado econômico, a Tecnologia da Informação (TI) tem sua parcela de participação no sucesso das organizações, visto que auxilia no processo de competitividade das indústrias, por participar diretamente do processo estratégico da empresa. O emprego de SI se dá nas mais diversas áreas, mas em caráter estratégico, visa originalmente a redução de custos, qualidade de produtos e serviços, automação de processos e otimização de operações. Segundo Pereira (1997, p 239), “a Tecnologia da Informação surgiu da necessidade de se estabelecer estratégias e instrumentos de captação, organização, interpretação e uso das informações”. Para Batista (2004, p 59), “Tecnologia da Informação é todo e qualquer dispositivo que tenha a capacidade para tratar dados e/ou informações, tanto de forma sistêmica como esporádica, independente da maneira como é aplicada”.

A partir do planejamento de TI, notam-se então mudanças na forma do exercício gerenciador no ambiente organizacional. Isto porque com a grande atuação e difusão dos microcomputadores bem como as telecomunicações, a Tecnologia da Informação se aplica a nível estratégico das organizações, com sistemas voltados ao nível gerencial, apoio a executivos e sistemas estratégicos a fim de orientar sua atuação no mercado competitivo (FARIAS, 2007, p 45). Com isso, é possível então considerar alguns efeitos da atuação da TI, que auxiliando na estratégia das organizações, visa à eficiência e ganhos no processo administrativo. Esses efeitos estão classificados em cinco níveis (YONG, 1992, p 79):

- Nível 1: redução de custos;
- Nível 2: melhoramentos em qualidade;
- Nível 3: novos produtos;
- Nível 4: administração estratégica via TI;
- Nível 5: novos conceitos e paradigmas via TI.

Assim sendo, na seção 2.2 são apresentados os tipos de sistemas de informação que auxiliam no processo de gestão das organizações.

2.2 Tipos de Sistemas de Informação

Existem alguns tipos de sistemas de informação que atendem níveis diferentes em uma organização. Esses níveis de auxílio são classificados como: nível gerencial, nível de conhecimento e nível estratégico. Os sistemas de informação foram classificados conforme suas funções administrativas e tipos de dados que manipulam, que conforme suas especificações auxiliam os executivos nos diferentes níveis hierárquicos da organização, auxiliando no processo de tomada de decisões. Na Figura 1, são apresentados os tipos de Sistemas de Informação.

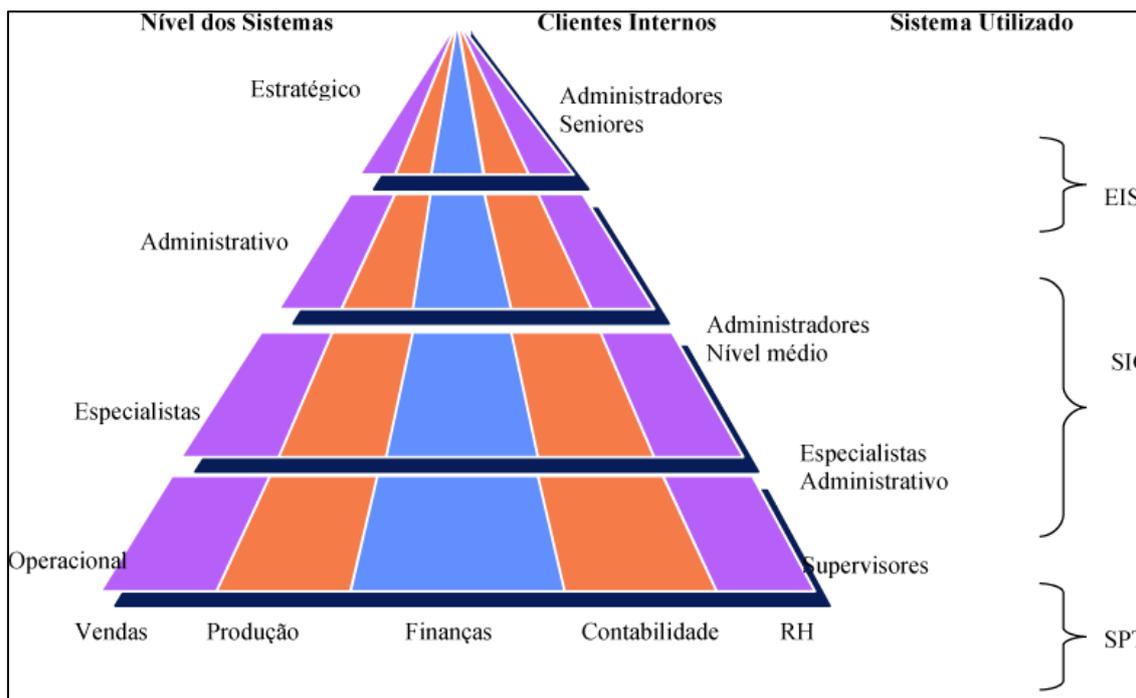


Figura 1: Tipos de Sistemas de Informação (LAUDON e LAUDON, 1997 *apud* MACCARI, 2006, p. 6).

Na Figura 1, orientando-se da esquerda para a direita, apresenta-se a relação dos tipos de sistemas de informação com os níveis de atuação. Em seguida, observam-se as áreas em que o sistema atua no âmbito da

organização, bem como os profissionais que utilizam as informações geradas. Neste caso, nota-se também, marcada pelas chaves, os tipos de sistemas e em quais níveis de atuação esses se enquadram na organização. Na base da pirâmide são apresentadas as áreas em que os sistemas podem auxiliar. A seguir, apresentam-se a descrição dos tipos de sistemas de informação.

- **Sistemas de Informação de Suporte às Transações (SPT)** – são sistemas de informação que tratam das rotinas da organização. Auxiliam nas funções operacionais, mais propriamente relacionadas às atividades do negócio envolvido. Em geral, são os primeiros a serem implantados, pelo fato de coletarem dados e também porque em relação aos outros sistemas de informação têm custos mais reduzidos. Esses sistemas servem de base para os sistemas, pois sua finalidade é processar dados, realizar cálculos, armazenar, recuperar dados e apresentar dados a usuários, otimizando as rotinas (CUNHA, 2005, p. 51). Esses sistemas podem ser exemplificados como: sistemas CRUD (*Create, Ready, Update e Delete*) de domínios diversos, balanços financeiros, de estoque e sistema de vendas.
- **Sistemas de Informação Gerenciais (SIG)** - esses sistemas dão suporte às atividades de planejamento, controle e organização. Segundo Oliveira (2000, p. 59), “o Sistema de Informação Gerencial é representado pelo conjunto de subsistemas, visualizados de forma integrada e capaz de gerar informações necessárias ao processo decisório”. Esse sistema auxilia a empresa a alcançar suas metas, e auxiliam seus gestores com detalhes sobre as atividades de rotinas para que de forma eficiente possam controlar, organizar e planejar suas ações (OLIVEIRA, 2000, p.59). Suas bases de dados possuem dados de origem interna e externa, pois sua função é prover o efeito ação-reação entre o planejamento e seus resultados. Como fontes externas podem ser consideradas: popularização do produto, aumento de demanda, recorde de produção, entre outras. Como fontes internas, consideram-se: as bases de dados presentes nos diversos setores da organização, planilhas, textos, dentre outros que podem depender do domínio relacionado. Para otimização do processo gerencial, os sistemas atuam sobre os problemas organizacionais e, assim, auxiliam na solução. A seguir, na Figura 2 é apresentada a estrutura do SIG.

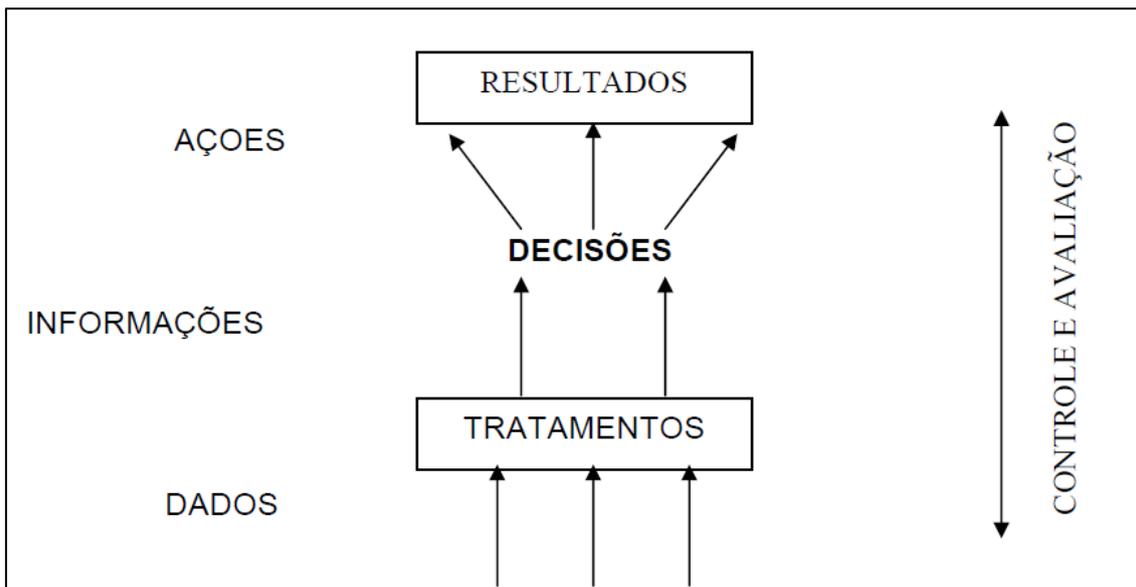


Figura 2: Estrutura de um SIG (TRABSI, 2011, ON-LINE).

Na Figura 2 é apresentada a estrutura de um SIG. A partir dos dados existentes sobre determinado problema identificado, esses passam por um processo de tratamento (pode-se considerar um processo de padronização dos dados), que são transformados em informações pertinentes ao domínio. A partir desse ponto, as informações padronizadas são utilizadas no processo decisório, sendo que as mesmas podem orientar as decisões estratégicas da organização.

- **Sistema de Informação Executiva (SIE)** – para Furlan (1994, p. 13), um SIE é “um mecanismo computadorizado que fornece aos executivos as informações necessárias para gerenciar o negócio. Este tipo de SI consiste em: combinar dados internos e externos; na utilização de menus gráficos, no acesso a banco de dados internos e externos dos setores diretamente envolvidos nas operações. Os dados são apresentados em forma de relatórios comprimidos”. Para Pozzebon (1996, p. 29), o SIE “é uma solução em termos de informática que disponibiliza informações corporativas e estratégicas para as decisões de uma organização, de forma a otimizar sua habilidade para tomar decisões de negócios importantes”. Um SIE propõe a integração das informações existentes nas organizações, sejam elas quais forem e estejam onde estiverem, seja em ambiente interno ou externo à organização. Esse sistema oferece de forma prática o acesso a todos os dados e, conseqüentemente, informações presentes na organização, de

forma rápida e facilitada. Auxilia ao executivo na tarefa de acompanhar diariamente os resultados, promover a interação entre os sistemas de áreas diversas. Este sistema possui também a característica de fornecer os estágios da informação, desde os níveis mais simples ao máximo detalhamento. Na Figura 3, são apresentadas as características de um SIE.

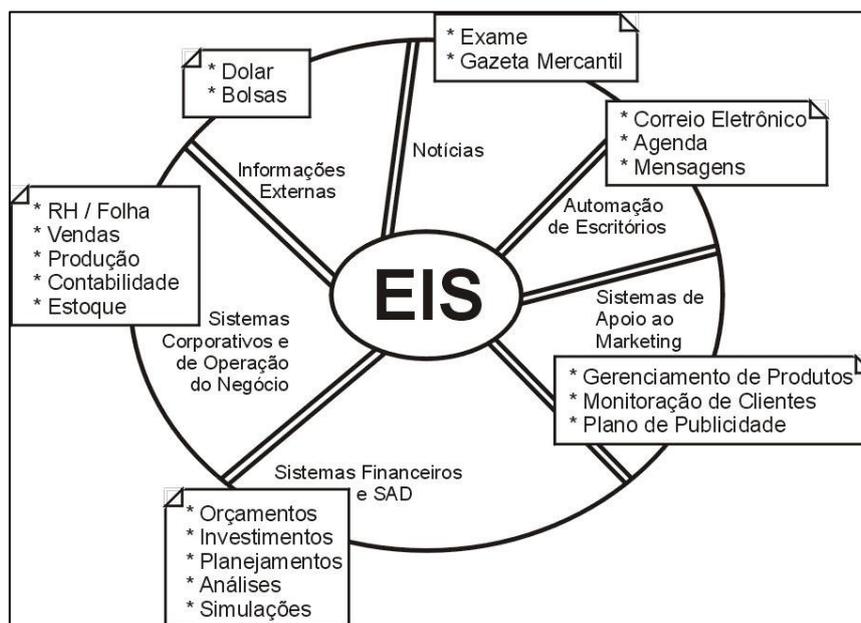


Figura 3: Característica de um SIE (ALVES, 2006, p. 6).

Na Figura 3 é apresentado o que propõe um SIE. Verifica-se a existência de fontes diversas representadas nas caixas de textos. Essas fontes auxiliam SI's específicos, ou seja, SI's desenvolvidos para atividades de setores da organização. Esses SI's produzem informação que, por sua vez, servem de entrada para um SIE e assim auxiliam no processo estratégico. Cabe ressaltar que esse processo engloba toda a organização, assim as informações são centralizadas a fim de auxiliar no processo decisório e estratégico.

- **Sistemas de Apoio à Decisão (SAD)** – para Bortolin (2005, p. 1), Sistemas de Apoio à Decisão são “programas capazes de manipular um grande volume de informações, obterem dados de fontes externas e serem baseados nas abordagens de otimização e heurística”. A seguir, a seção 2.2.1 abordará com maiores detalhes esse tipo de SI, pois o mesmo fundamenta este trabalho.

2.2.1 Sistema de Apoio à Decisão (SAD)

Para Laudon e Laudon (2005, p. 238) "um Sistema de Apoio à Decisão (SAD) auxilia o processo de decisão gerencial combinando dados, ferramentas e modelos analíticos sofisticados e software amigável ao usuário em um único e poderoso sistema que pode dar suporte à tomada de decisão, semiestruturada e não estruturada". Desta forma, considera-se que um SAD auxilia o processo decisório sobre problemas específicos, com foco sobre a eficiência da tomada de decisão sobre o domínio envolvido. Esses sistemas são aplicados em situações que as informações necessárias são de difícil descoberta nas bases de dados, que conseqüentemente a torna também complexa para manipulação.

Um SAD fornece aos usuários um conjunto flexível de ferramentas e capacidades para analisar dados importantes. Esses sistemas possuem a função de oferecer a organização informações para a tomada de decisão. Para tanto, o sistema oferece como saída informações que possam esclarecer alguns pontos relacionados ao assunto para apoiar o processo decisório.

O SAD auxilia no processo das decisões não rotineiras da organização, pois são aplicados no auxílio da avaliação e monitoramento dos processos. Por isto, este tipo de sistema auxilia a organização quanto ao processo de competitividade, pois disponibiliza a empresa recursos tecnológicos para armazenar e buscar informações de formas ágeis e eficientes. O processo decisório é bastante complexo para a organização, então, vale ressaltar que o auxílio do SAD não se limita até o momento da tomada de decisão, mas auxilia também a analisar as conseqüências das decisões tomadas bem como confrontar com o cenário atual da organização (MARTINS, 2007, p. 16).

Para realizar as atividades um SAD reúne os procedimentos em modelos para realizar o processamento dos dados e também uma avaliação dos problemas para então atuar no processo de decisão. Pode-se então considerar que na atuação do SAD, ou seja, dentre os problemas solucionados há uma classificação de três tipos: problemas estruturados, semiestruturados e não estruturados.

- Problemas estruturados – quando as fases operacionais estão definidas e os resultados esperados estão claros. Estão relacionados aos problemas de rotinas e ou atividades repetitivas na organização.
- Problemas semiestruturados – são problemas onde as operações são conhecidas, porém, não se tem conhecimento total das variáveis de interferência nos resultados, assim, deseja-se conhecê-las totalmente para então replanejar as atividades.
- Problemas não estruturados – neste caso, os critérios de decisão não estão definidos ou ainda, são desconhecidos sendo ainda necessário reunir os dados e realizar o levantamento dos pontos pelo qual se deseja iniciar o estudo e partir desse ponto, reunir os dados pertinentes para reconhecimento e assim possibilitarem a definição.

A definição do modelo a ser utilizado está diretamente relacionada à complexidade do problema. Para tanto, são considerados alguns pontos como: tempo para atuação da solução e o custo da solução tecnológica. Observa-se então a responsabilidade e importância de um SAD para a organização, pois possibilita a mesma a uma larga vantagem competitiva sendo que o sistema atua com intuito de agrupar as informações para auxiliar na identificação dos padrões e tendências para o processo decisório (LIMA, 2007, p. 14).

Assim, confrontando a situação de uma organização que não possui sistemas que auxiliem no processo de tomada de decisão, verifica-se que as mesmas são baseadas nas experiências individuais, e ainda em dados históricos. Observadas as responsabilidades e os benefícios destes sistemas, é importante ressaltar as características que possibilitam realizar tais atividades no âmbito da organização. Segundo LIMA (2007, p. 24), um SAD possui as seguintes características:

- Suporte a manipulação de grande volume de dados;
- Capacidade de acesso a dados externos e integrá-los a sistemas internos;
- Flexibilidade e dinamismo na geração de relatórios (textos, planilhas, gráficos);
- Capacidade de oferecer propostas com análise de melhor solução nas situações apresentadas;

- Antecipar situações e trabalhar com hipóteses de problemas e decisões a fim de auxiliar na verificação de impactos e resultados;
- Oferecer suporte a decisões em níveis diferenciados no âmbito da organização (do operacional, passando pelo estratégico até o decisório).
- Agilidade no desenvolvimento por ter participação dos usuários (responsável) no processo;
- Adaptabilidade de novas ferramentas de apoio à decisão;
- Dinamismo sobre a manipulação dos dados e informações;
- Direcionamento e configuração da disponibilização das informações conforme perfil do profissional.

Observando as características e as expectativas sobre a implantação de um SAD, é necessário então ter conhecimento sobre sua arquitetura. Esta é composta por três partes, sendo: Sistema de Gerência de Banco de Dados (SGDB), Sistema de Gerência de Banco de Modelos (SGBM) e um Sistema de Interface com o Usuário.

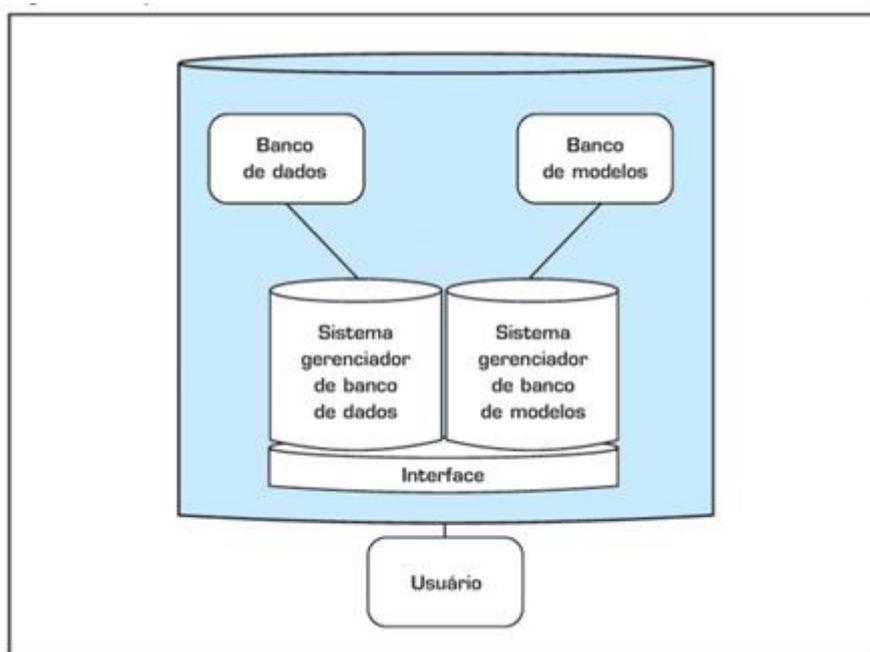


Figura 4: Apresentação da arquitetura de um SAD (SPRAGUE e WATSON, 1989 *apud* CLERICUZI, 2006, *on-line*).

A Figura 4 apresenta a arquitetura ou componentes de um SAD. Conforme a figura, o banco de dados é responsável pelo armazenamento dos dados e associado a seu funcionamento está o SGDB que realiza a

manutenção e possibilita o acesso aos dados. Por meio do SGDB são realizadas a combinação dos dados estruturados e não estruturados e a inclusão e processamento das consultas nas bases. Em paralelo está o banco de modelos que possui as ferramentas de análise de dados e a manipulação dos modelos de apoio a decisão. Esses modelos realizam a interação com os dados e bem como as atividades pertinentes ao sistema para disponibilização das informações referentes ao domínio e de interesse dos responsáveis pelo processo decisório. O SGBM realiza o gerenciamento dos modelos e suas atividades são bastante semelhantes as do SGDB, mas, sobre os modelos. Por fim, a interface com usuário, que é um subsistema que possibilita a acessibilidade do usuário ao sistema. Vale ressaltar que o sucesso do SAD está relacionado à usabilidade apresentada na interface, sendo que este deve preocupar-se em disponibilizar de maneira simples e direta as funcionalidades, possibilitando a navegabilidade pelas funcionalidades do sistema.

Para a construção do SAD e também permitir a análise dos dados, o SAD pode ser composto por ferramentas como: *Data Warehouse* (DW), *Data Marts* (DM), ferramentas OLAP, entre outras tecnologias. Atualmente, o termo SAD tem sido substituído por *Business Intelligence* (BI), que em sua definição é “a habilidade de estruturar, acessar e explorar informações, normalmente armazenadas em um DW ou DM, com o objetivo de desenvolver percepções, entendimentos, conhecimentos, os quais podem produzir um melhor processo de tomada de decisão” (BARBIERI, 2001, *apud* FELBER, 2006, p. 18). Na Figura 5 são apresentados os componentes do SAD ou BI, enfatizando a atuação dessas ferramentas.

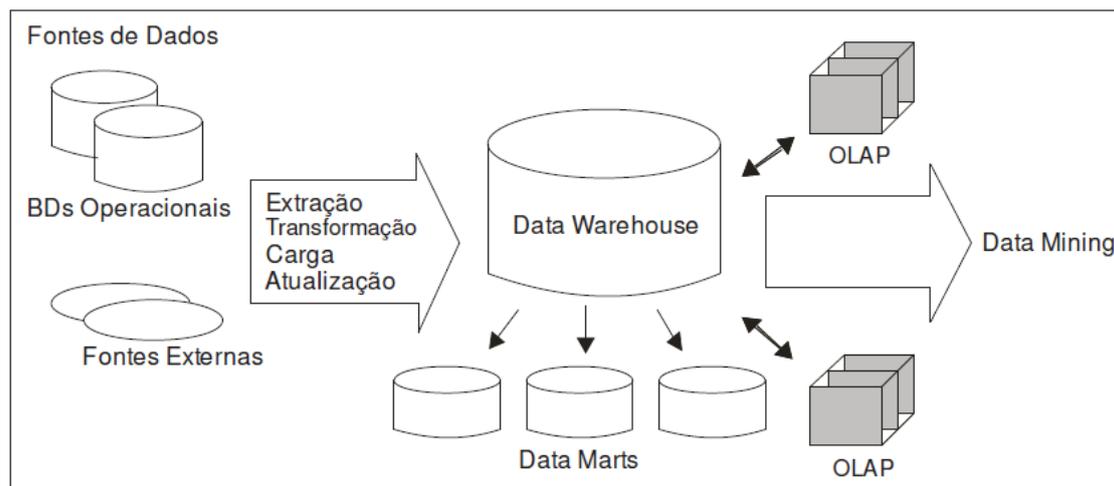


Figura 5: Componentes de BI ou SAD (MARTINS, 2007, p. 19).

Na Figura 5 também apresenta a arquitetura de um SAD, mas destaca-se a presença do DW e da ferramenta OLAP. A seguir, na seção **Erro! Fonte de referência não encontrada.** são apresentados os conceitos e especificações de *Data Warehouse* e posteriormente, na seção 2.6 sobre ferramentas OLAP.

2.3 *Data Warehouse*

Data Warehouse (ou armazém de dados) é um sistema de computação com finalidade de organizar os dados de uma organização, possibilitando que, a partir desses dados, ferramentas possam trabalhar para dar suporte à tomada de decisão. Inmon (1997, p.33) define *Data Warehouse* como sendo uma “coleção de dados orientada por assuntos, integrada, variante no tempo, e não volátil, cujo objetivo é dar suporte aos processos de tomada de decisão”. Porém, outros autores apresentam suas definições com foco nas necessidades do usuário final. Por exemplo, para Barquini (1996, p. 36), DW “é uma coleção de técnicas e tecnologias que juntas disponibilizam um enfoque pragmático e sistemático para tratar com o problema do usuário final que precisa acessar informações que estão distribuídas em vários sistemas da organização”. Já para Singh (2001, p. 12), DW é “um processo de integração dos dados corporativos de uma empresa em um único repositório a partir do qual os usuários finais podem facilmente executar consultas, gerar relatórios e fazer

análises”. No entanto, todas as definições apresentadas possuem como ponto comum a utilização da tecnologia como solução de reunir dados dispersos de uma organização, concentrá-los em uma base de dados e oferecer mecanismos de consultas para que o usuário encontre novas rotas de ações motivadas pelas informações extraídas dos dados.

A organização que opta por utilizar um *Data Warehouse* bem estruturado possibilita ao executivo a redução de tempo em atividades de análise de informações, onde por meio de geração de relatórios, auxiliarão na tomada de decisão (DALFOVO, 1998, p. 12). Um DW é um banco de dados com função de armazenar dados relacionados ao ambiente organizacional. Dados esses que foram definidos, tratados e padronizados para que em determinado momento, passando por um processo de processamento, forneça informações sobre assuntos específicos. Para Harjinder e Rao (1996, *apud* KORUGI, 2001, p. 12) um DW fornece uma “imagem única da realidade do negócio”. Em geral, DW consiste em um conjunto de programas computacionais para extrair informações da base de dados operacional da organização. A seguir, a Tabela 1 apresenta a diferença entre um sistema de processamento *Online Transaction Processing* (OLTP) e um DW.

Tabela 1: Resumo das diferenças entre sistema OLTP e *Data Warehouse* (LIMA, 2006, p. 15).

Características	Banco de Dados Operacionais	<i>Data Warehouse</i>
Objetivo	Operacional	Informativo
Tipo de processamento	OLTP	OLAP
Tipo de operação	Inclusão, alteração, exclusão	Carga e consulta
Número de usuários	Milhares	Centenas
Tipo de usuários	Operadores	Tomadores de decisão
Interação com usuário	Somente pré-definida	Pré-definida e ad-hoc
Condição dos dados	Dados operacionais	Dados analíticos
Volume	Megabytes – gigabytes	Gigabytes – Terabytes
Histórico	60 a 90 dias	5 a 10 anos

Granularidade	Detalhados	Detalhados e resumidos
Estrutura	Estática	Variável
Acesso a registros	Dezenas	Milhares
Atualização	Contínua (em tempo real)	Periódica (em batch)
Integridade	Transação	A cada atualização
Número de índices	Poucos/simples	Muitos/complexos
intenção dos índices	Localizar um registro	Aperfeiçoar consultas

A partir das informações da Tabela 1 é possível perceber que um sistema OLTP é responsável por armazenar os dados referentes à rotina da organização, e um DW prepara a organização para o processo decisório a partir de SI. Com a implementação de um DW a organização possui um conjunto de subsídios que auxiliam na análise de negócios e, com isso, otimiza o processo de tomada de decisão, uma vez que o DW utiliza de ferramentas computacionais para extração de informações de sistemas e base de dados já utilizados na organização. Com isso, pode-se então destacar as vantagens da utilização de um DW no ambiente organizacional: redução de tempo de alcance e identificação da informação, alteração no fluxo de tarefas e atividades operacionais, otimização da geração de dados, segurança e procedência das informações e qualidade da informação para tomada de decisões (KUHLEN e VIEIRA, 2005, p. 3).

Considerando então a utilidade de um DW, pode-se então destacar alguns objetivos para sua implementação no âmbito organizacional (KIMBALL, 2002, p. 75):

- facilidade no acesso as informações corporativas, pois permite acesso aos dados com facilidade e respondem rapidamente as consultas realizadas. Esta facilidade não se refere somente ao acesso, mas também quanto à compreensão das informações disponibilizadas;
- confiabilidade e consistência da informação, visto que os dados são obtidos de formas diversas, mas oferece mecanismos de filtragem e tratamento dos dados para então as disponibilizar;

- dinâmico, por permitir adaptações tanto na disponibilização das informações quanto da integração de novas bases, conforme as necessidades do domínio;
- segurança das informações, pois possibilita o controle de acesso às informações de suma importância para o processo decisório da organização;
- auxiliar na tomada de decisões, uma vez que as informações do DW são disponibilizadas e padronizadas para auxiliar no processo decisório. Portanto, deve-se buscar ao máximo explorar todas as bases de dados, arquivos e planilhas da organização.

Considera-se então que uma organização que busca explorar sua base de dados para então produzir conhecimento, deseja alcançar a excelência quanto ao serviço ou produto disponibilizado e, assim, alcançar o topo no processo de competitividade (LERMEN, 2006, p.2). De acordo com Gonçalves (2003, p 13), “o grande diferencial das empresas para permanecerem no mercado é o investimento em recursos tecnológicos”, pois utilizam dos sistemas e recursos investidos não somente para possuir os dados, mas também mantê-los de formas organizada e estruturada. Assim sendo, a organização que opta por utilizar um *Data Warehouse* bem estruturado, possibilita ao executivo a redução de tempo em atividades de análise de informações, onde por meio de geração de relatórios, auxiliarão na tomada de decisão (DALFOVO, 1998, p. 12).

Para Gonçalves (2006, p. 13), “o grande diferencial das empresas para permanecerem no mercado é o investimento em recursos tecnológicos”. Essas características oferecem à organização a possibilidade de se manter com um nível alto de competitividade, pois permite adequação em seus processos e essencialmente a tomada de decisões são apoiadas em informações e relatórios baseados na realidade da organização. A seguir, na seção 2.3.1 são apresentadas as características de um DW.

2.3.1 Características de um *Data Warehouse*

Nas próximas seções serão abordadas as características de um DW, possibilitando a visualização de seus componentes.

2.3.1.1 Orientado a Assuntos

A orientação a assunto no DW se dá pelo fato que os dados são armazenados e organizados conforme o interesse da organização. Os sistemas operacionais têm sua orientação para os processos e tem o foco no armazenamento dos dados resultantes da rotina. A orientação a assuntos auxilia na organização e análise dos dados, pois assim permite armazenar um montante maior de dados e quando a organização necessitar explorar o assunto ou ainda ter auxílio no processo decisório, os dados obedecerão a um padrão de organização de especificação personalizada, oferecendo assim informações concisas sobre o que a organização necessita.

Para exemplificar essa característica pode-se supor uma empresa representante de diversas marcas que fornecem materiais esportivos que, em determinado momento, precisa traçar o perfil de seus clientes. Neste contexto, os dados podem ser organizados pelo assunto “Clientes”, facilitando traçar o perfil e tomar as decisões necessárias (GONÇALVES 2003, p. 13).

2.3.1.2 Integrado

Essa característica é de suma importância para organizações que possuem bases de dados distintas para os sistemas operacionais de rotinas. Essa característica de DW possibilita centralizar em um único local os dados envolvidos por meio dos processos de extração, transformação e carga destes dados. Desta forma, independente da base de dados e a forma como os dados estão armazenados ou ainda representados, esta característica possibilita a integração dos dados (MARTINS, 2007, p. 22). Com isso, ressalta-se então a

flexibilidade do DW, tendo em vista a capacidade de padronização e transformação dos dados para integração.

Em uma organização podem existir mais de uma base de dados responsável pelo armazenamento dos dados operacionais. Nesta situação é possível que dados semelhantes estejam representados de formas distintas, ou em uma base de dados relacional é possível realizar consultas sobre os dados, porém, no ambiente de DW é possível realizar a integração dos dados para possíveis extrações de conhecimento. O DW no processo de transformação dos dados realiza a junção das bases. (ZAJACZKOWSKI, 2005, p. 16).

Essa característica pode ser exemplificada sobre o atributo sexo que pode estar representado como: “m/f ou M/F”, ou “1/0”, ou ainda, “masculino/feminino” e no ambiente de DW, são convertidos obedecendo a um padrão.

2.3.1.3 Variante no Tempo

Na estrutura dos dados de um DW, existe um atributo tempo. O tempo é um componente indispensável, operando como chave. No DW a cada alteração nos dados, uma nova entrada é realizada, assim, os dados anteriores são mantidos, pois os mesmos são precisos em relação a um intervalo de tempo.

Em *Data Warehouse* os dados armazenados fazem referência a um intervalo de tempo sendo que, a cada alteração nos dados, será registrada uma nova entrada, sem que os dados já existentes sejam alterados (BONI, 2009, p. 48). Segundo Machado, (2000, p. 15), “os dados de um DW são precisos em relação ao tempo e representam resultados operacionais em determinado momento de tempo: o momento em que foram capturados”. Desta forma, pode-se salientar que a data (tempo) é um fator preponderante, pois é utilizado como dimensão do projeto.

2.3.1.4 Não Volátil

A característica da não volatilidade em DW está relacionada com o fato dos dados não serem editados em seu ambiente, ou seja, uma vez inseridos, os dados são disponibilizados somente para leitura (consulta).

Considera-se então que os dados são inseridos no ambiente do DW de tempo em tempo, e a partir desta inserção, a base é atualizada com um novo período de dados, porém, os dados não são sobrepostos ou alterados (LIMA, 2006, p. 19).

2.3.1.5 Granularidade

Essa característica refere-se ao nível de detalhamento dos dados, controlando as unidades dos dados mantidas no DW. O DW pode conter um alto nível de detalhamento ou ainda resumido. Nesse sentido, quanto maior o nível de detalhamento, menor o nível de granularidade e, conseqüentemente, quando maior o nível de granularidade, menor o detalhamento. Vale ressaltar que com baixo nível de granularidade, o fluxo de dados será elevado no DW, exigindo maior infraestrutura de processamento e armazenamento (BONI, 2009, p. 49).

Essa questão é de extrema relevância no projeto do DW, pois tem relação com o volume de dados, desempenho e eficiência do projeto. O volume elevado de dados exige recursos de infraestrutura, ou seja, poder de processamento e de armazenamento. Definir sobre o nível de granularidade é um ponto preponderante do projeto, pois requer análise dos recursos de infraestrutura, conforme descrito, porém, há também a preocupação sobre sua utilidade, pois o DW deve atender as necessidades da organização (LIMA, 2006, p. 19). A seguir, na Figura 6, é exemplificado o conceito de granularidade.

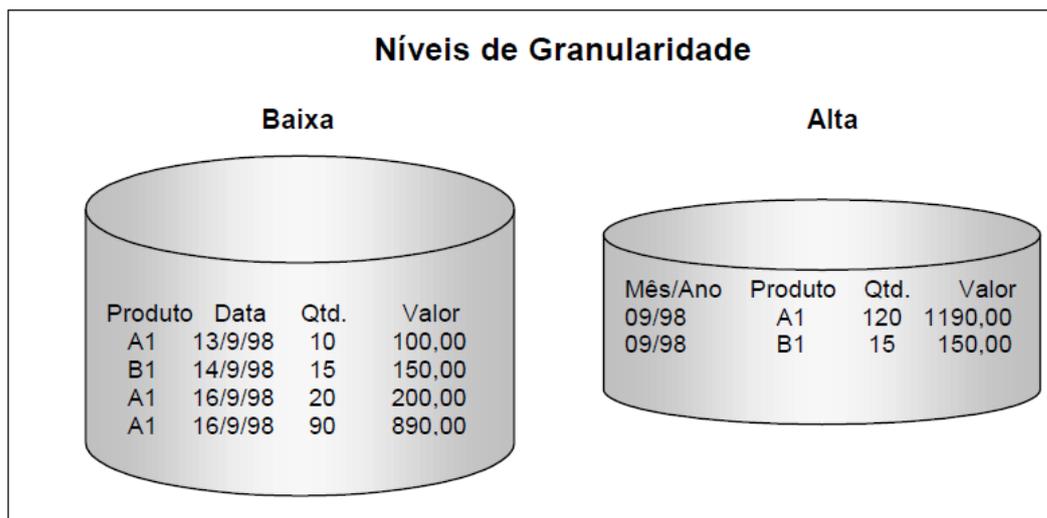


Figura 6: Níveis de granularidade (KUROGI, 2001, p. 19).

Na Figura 6 percebem-se dois projetos referentes ao fluxo de vendas de determinado produto. O primeiro projeto possui um baixo nível e granularidade, onde contém um alto detalhamento e quantidade de dados referentes à venda do produto. No segundo projeto, observam-se as mesmas características sendo detalhadas, porém a detalhamento refere-se à quantidade de registros que pode estar relacionada ao tempo de atualização dos dados, fazendo com que a quantidade de registros tornem-se detalhes para a base de dados em questão.

A decisão sobre o nível de granularidade do projeto de DW é de grande importância, pois se considera toda a demanda e eficiência do armazenamento bem como a disponibilização de dados, pois influenciará no que será disponibilizado para análise.

Em uma organização podem existir dois níveis de granularidade, denominado níveis duais de granularidade. O primeiro nível é de alta granularidade, que por sua vez, oferece uma descrição resumida dos dados. O segundo nível, de baixa granularidade, contém um maior detalhamento (INMON, 1997, p. 51).

Abordando um nível duplo de granularidade no DW, a organização contempla as duas abordagens de dados, e poderá disponibilizar informações conforme a necessidade da situação.

2.3.1.6 Credibilidade dos Dados

A credibilidade dos dados é fundamental para o sucesso do projeto de DW, visto que não pode haver a menor incoerência nos dados, pois compromete todo o processo decisório e servem de base para a descoberta de informações que, por sua vez, pode levar a organização à tomada de decisões equivocadas.

Há também a preocupação em implementar relatórios úteis para organização. Os relatórios gerados são utilizados em situações diversas da organização, especialmente no processo de tomada de decisão e, por isso, faz-se necessário que a fonte de dados seja confiável, pois do contrário, põe em risco o processo da organização envolvido. Além disso, vale ressaltar outras atividades importantes para o processo que é a manutenção e atualização da base de dados, visto que antes dos dados serem integrados ao DW, esses passam pelos processos de transformação para possibilitar a integração e coerência (SOUZA, 2005, p. 3). Na próxima seção, serão abordadas as arquiteturas de desenvolvimento do DW.

2.3.2 Arquitetura de Desenvolvimento

A definição da arquitetura a ser utilizada no desenvolvimento de um DW é uma atividade importante para o projeto. Para isso, devem ser considerados pontos como: infraestrutura disponível, porte da empresa e os recursos necessários para a construção. Isto porque a arquitetura está relacionada ao conjunto de regras adotadas para a construção do DW. Existem três tipos de arquitetura de implementação que, por sua vez, podem ser implementadas utilizando as abordagens *top down* ou *bottom up*. Nas subseções a seguir são apresentados os tipos de arquiteturas disponíveis para a construção de um DW.

2.3.2.1 Arquitetura Global

A arquitetura global consiste na construção de um DW que aborde todo o domínio da empresa. Esse DW embarca todos os setores que conseqüentemente concentram as informações para o processo decisório. Nessa arquitetura os dados são extraídos das fontes e passam por um processo de limpeza onde são classificados o que são considerados úteis a organização. Depois de classificados os dados entram em um processo de transformação a fim de garantir a consistência para então serem carregados à base do DW.

Essa arquitetura possibilita aos usuários acesso as informações de toda a organização, que por sua vez, oferece o relatório de forma mais consistente, uma vez que os dados são acessados de forma centralizada e estão em uma base de dados (MACHADO, 2004, p. 48). Por outro lado, para a implementação utilizando esta arquitetura, será exigido um prazo de tempo maior e também de maior risco, pois haverá manipulação de maior montante de dados para a construção do DW, e por isso também aumentam os riscos.

2.3.2.2 Arquitetura Independente

A arquitetura independente consiste na criação de *Data Marts* isolados, ou seja, pequenos *Data Warehouse* setoriais. Assim, é possível focar em setores da organização e, por isso, o processo de armazenamento dos dados também obedece a essa característica setorial. Essa arquitetura é mais simples no processo de implementação e o prazo também é menor. Porém, possui a limitação de ser acessado somente nos setores interessados, não estando disponível para toda a organização (MACHADO, 2004, p. 50).

2.3.2.3 Arquitetura Integrada

A arquitetura integrada consiste em integrar os *Data Mart's* dos setores da organização. Essa arquitetura compreende em interligar e possibilitar a

integração dos dados. Cada *Data Mart* é correspondente a um setor e atende a demanda do mesmo. Porém, com a integração entre os *Data Mart's*, é possível que cada setor acesse os dados dos demais setores, possibilitando que o processo decisório possa ser gerado baseado na visão global da organização.

2.3.3 Abordagem de Implementação do DW

Para a implementação de DW existem dois tipos: *top down* e *bottom-up*. Para o processo, a definição sobre qual utilizar sofre influências diversas como: infraestrutura disponibilizada, a arquitetura definida, variáveis da organização relacionadas ao setor ou área interessada ou ainda o tempo disponível para execução (ALMEIDA, 2006, p. 18).

2.3.3.1 Abordagem Top-Down

A abordagem *top-down* foi a primeira a ser utilizada em um ambiente de DW e é considerada a abordagem padrão, além de ser a mais utilizada no processo de implementação de DW. A implementação utilizando *top down* envolve no processo todos os setores da organização. A interação com os setores possibilita a definição dos padrões e também a relevância dos dados para o processo decisório (MARTINS, 2006, p.25).

Tendo os dados submetidos ao processo de transformação e integração, são carregados para o ambiente do DW que, por sua vez, são resumidos, dimensionados e distribuídos para os DM dependentes, conforme apresentado na Figura 7.

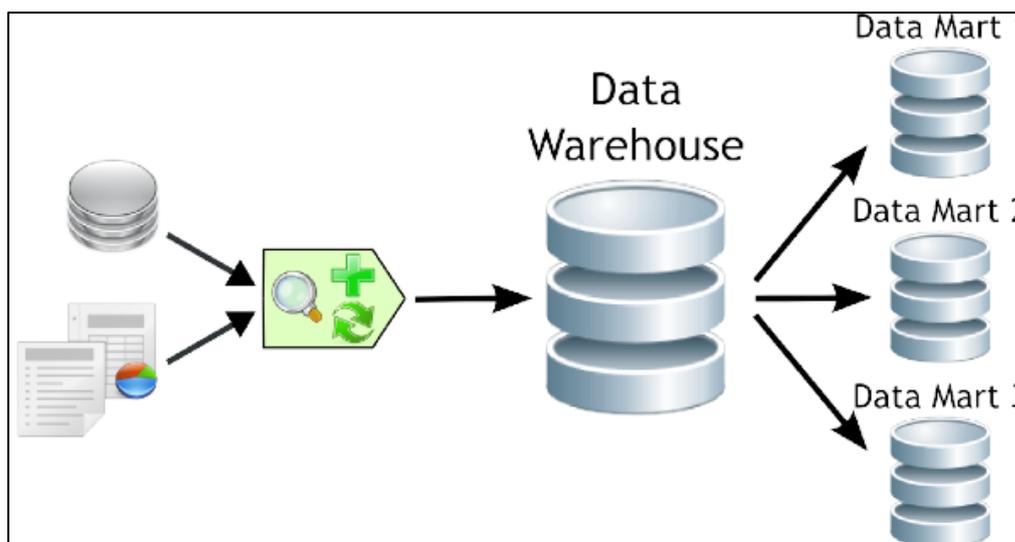


Figura 7: Método Top-down (DATAPRIX, 2010, on-line).

Na Figura 7 é apresentada a abordagem *top-down*. Observam-se as fontes dos dados, sendo os sistemas operacionais ou ainda documentos que passam pelo processo de transformação e integração. Após a transformação, os dados são inseridos na base do DW. Neste momento, os dados são distribuídos para os DM que, por sua vez, são dependentes pois derivam do DW central.

Sobre essa abordagem, podem-se destacar vantagens como: sua arquitetura é integrada, possibilita a inclusão de novos assuntos e atender mais setores, pois a fonte para os setores é o DW e com isso, os dados passam por um processo rigoroso para alcançar consistência e padronização (MARTINS, 2006, p. 25).

Um DW pode ser utilizado para a descoberta de informações a fim de auxiliar o processo decisório com dados estatísticos, emissão de relatórios ou ainda, suporte a ODS (*Operational Data Storage*). ODS “é o armazenamento intermediário dos dados que facilita a integração dos dados no ambiente antes de serem utilizados” (SOARES, 1998 *apud* NOLETO, 2008, p. 23).

Uma desvantagem a ser observada é que essa abordagem torna o processo de desenvolvimento demasiadamente longo, pois a solução é implementada integralmente. Só depois de todo o DW implementado é que são construídos os DM para atender as necessidades dos setores. A implementação a partir dessa abordagem tem um custo elevado, pois de início aborda um modelo bastante detalhado. Nessa abordagem o processo de DW já

inicia robusto e exige alto investimento em infraestrutura de armazenamento e processamento (MARTINS, 2006, p.26).

2.3.3.2 Abordagem *Bottom-Up*

Essa abordagem segue uma sequência inversa à abordagem *Top-down*, ou seja, são criados os DM para os setores da organização para então iniciar o processo de DW.

Bottom-Up tem sido muito utilizada pelo fato que os resultados são alcançados em prazos mais curtos, pois o tempo de implementação é menor e a solução é aplicada diretamente no setor (TAVARES, 2008, p.33). A seguir, na Figura 8 é apresentada a abordagem *bottom-up*.

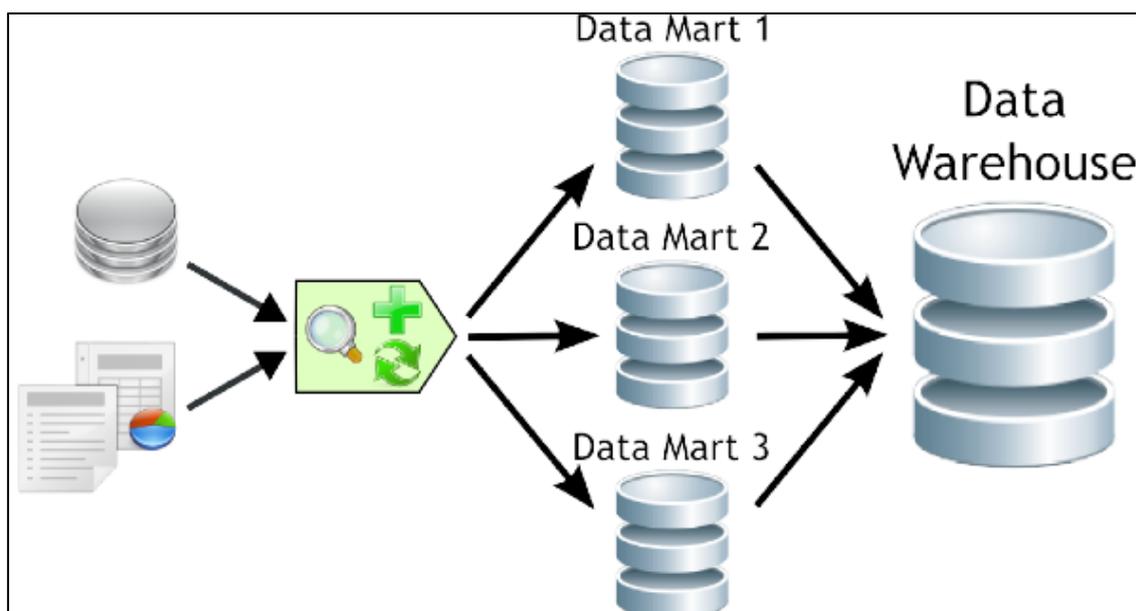


Figura 8: Método Bottom-up (DATAPRIX, 2010, on-line)

Na Figura 8 é apresentada a abordagem *bottom-up*. Inicialmente são observadas as fontes de dados, os sistemas operacionais e documentos que passam pelo processo de extração e transformação. Essa transformação prepara os dados e, a partir de então, são migrados para um DM. A partir desse ponto, as bases dos DM's já estão disponibilizados para o apoio aos setores. Depois de comporem o DM os dados alimentam a base de um DW que centralizará os dados para assim apoiar o processo decisório da

organização. Nessa abordagem nota-se uma redução de esforços para a construção do DW, pois os dados já estarão formatados de acordo com a necessidade dos setores.

Sobre a abordagem *bottom-up*, podem ser destacadas as seguintes vantagens: tempo de implementação reduzido, resultados apresentados mais rápidos e pode dar foco a setores de maior relevância e necessidade para a organização. Como desvantagem, pode-se destacar a dificuldade em padronizar os DM, pois precisa atender não somente o DM, mas estar num padrão consistente para o DW que utilizará de toda a base para o processo decisório. (MARTINS, 2006, p. 26). A seguir, seção 2.3.4, é apresentado o conceito de *Data Mart*.

2.3.4 Data Mart

Um *Data Mart* (DM) é parte de um *Data Warehouse*, ou seja, “é uma coleção de assuntos de uma área, organizado para apoio à decisão, baseado nas necessidades de um determinado departamento ou setor” (INMON 1998, *apud* SOARES, 1998, p. 35). Assim, pode-se então considerar que um DM é um *Data Warehouse* específico para um setor da organização, que apoia o processo de tomada de decisão daquele setor no qual gera informações.

A principal vantagem de um DM é que exige um investimento reduzido em comparação com o necessário para a implementação de um DW. Fora a redução de custo, deve-se também considerar o tempo para implementação de um DM que também é menor que na implementação do DW e assim pode oferecer suporte a situações específicas em setores da organização mais rapidamente e já auxiliar o processo decisório que, por sua vez, pode iniciar uma orientação da organização no processo de utilização de um SI para auxílio do processo decisório (MARTINS, p. 23, 2007).

Pelo fato de ser direcionado a um setor específico da organização, durante o período de implementação o setor envolvido no processo oferece suporte oferecendo informações sobre pontos importantes na tomada de decisão, bem como das situações de conflitos e/ou ambiguidades no processo decisório. De acordo com a arquitetura utilizada no processo de construção de

DW, é possível realizar a construção de DM para cada setor da organização e posteriormente realizar a integração em um único DW. Essas especificações e informações são apresentadas na seção 2.4, onde se trata da modelagem de DW.

2.4 Modelagem de *Data Warehouse*

A modelagem no processo de construção de um DW é essencial, pois é neste momento que é possível aplicar as técnicas de construção, verificar a presença de redundâncias bem como fazer as análises necessárias sobre o domínio envolvido e as necessidades da organização. Também é nesse momento que há uma definição sobre os pontos relevantes e necessários para a tomada de decisão. Nesse processo o objetivo é uma modelagem precisa de modo que não comprometa a sequência de atividades do processo de construção do DW.

Nesse momento, um ponto bastante delicado é a comunicação, pois haverá uma grande necessidade de interação entre o responsável pelo desenvolvimento e a organização, pois por parte do responsável pela implementação há o interesse em conhecer todo o negócio e reconhecer com riqueza de detalhes o problema bem como se dá o processo de tomada de decisão da organização. Por parte da organização, há preocupação em oferecer suporte a todo o processo, conforme as solicitações do responsável pela implementação, pois é a partir desse *feedback* que será reconhecido os detalhes para a construção do DW (ZILKOSKI, 2003, p. 21).

Nesse processo, é construído o modelo dimensional. Esse modelo é utilizado para construção do DW e, posteriormente, para inclusão dos dados originados dos sistemas operacionais, das rotinas da organização. Esse modelo é composto de uma tabela central (tabela fatos) e de tabelas de dimensões. As tabelas dimensões são construídas conforme e quantidade necessária no projeto. Vale ressaltar que a tabela de fatos possui chaves que a relacionam com as tabelas dimensionais (MARTINS, 2007, p. 38). Na Figura 9 é apresentado um exemplo contendo a tabela fato e tabelas dimensões.

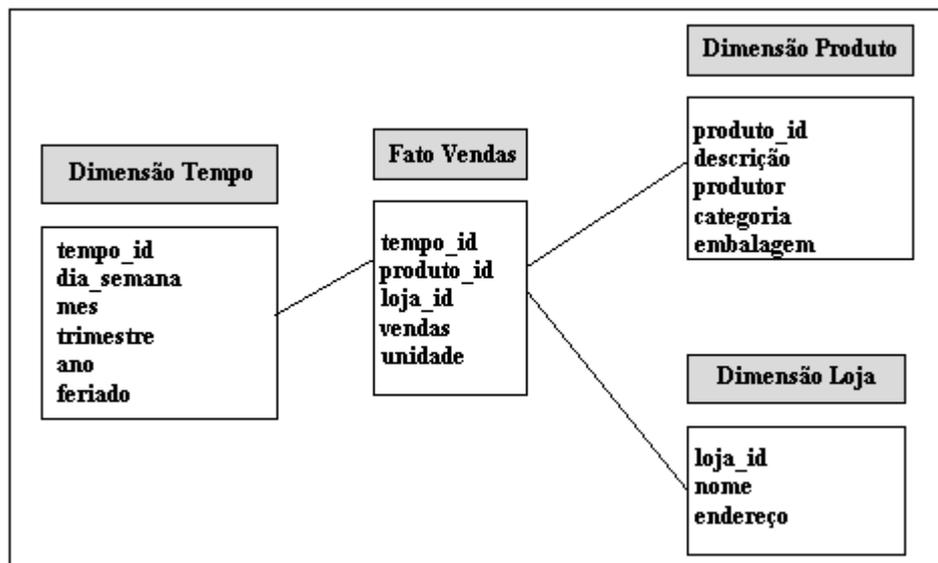


Figura 9: Exemplo de Dimensões e Fato (KURIKE e DIAS, 1999, *on-line*).

Na Figura 9 verifica-se a tabela “Fato Vendas”, que é o assunto e que vai orientar os relatórios. Verifica-se também as dimensões “Tempo”, “Produto” e “Loja”. A partir da dimensão “Tempo”, é possível criar associações das vendas com o tempo, e nos relatórios serão considerados o fato “venda” em determinado “tempo”. Conseqüentemente, a dimensão “Produto” possibilitará associações entre “Venda” e “Produto” e a dimensão “Loja”, entre o fato “Venda” e “Loja”.

Esta modelagem é a técnica utilizada para construir modelos de negócio ou modelos lógicos e tem por finalidade apresentar os dados de forma padronizada e intuitiva de modo que atenda a demanda de acessos. O modelo dimensional representa os dados do negócio como um cubo de dados, de modo que os valores estão inseridos nas células e em seus lados estão definidas as dimensões naturais dos dados, conforme apresentado na Figura 10.

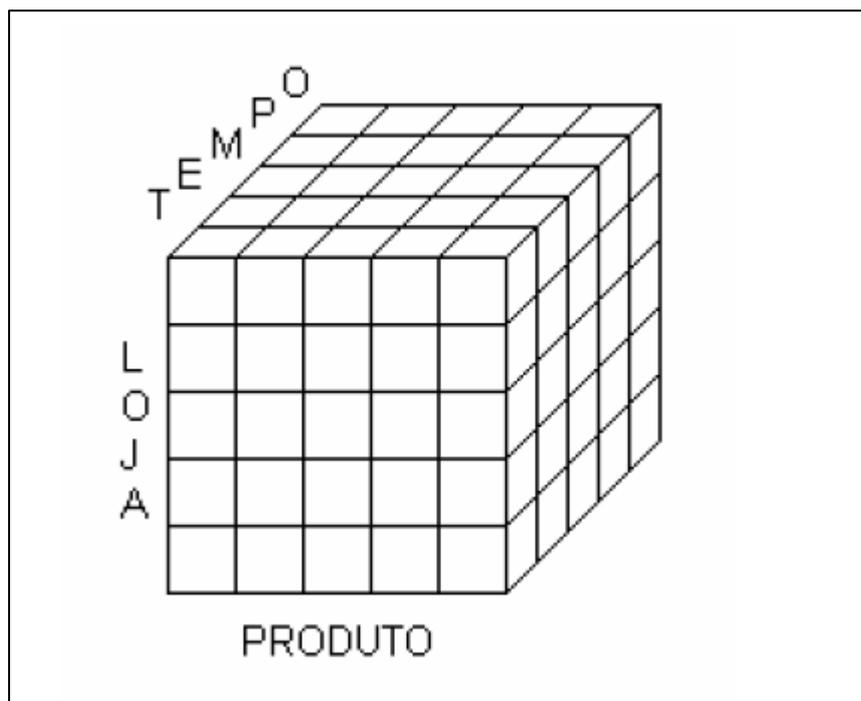


Figura 10: Modelo dimensional de negócio (KIMBALL ,1996, *apud* ZIULKOSKI, 2003, p. 22).

Na Figura 10 é apresentada a representação do fato “produto” (conjunto de produtos disponibilizados) relacionado às dimensões “loja” (áreas de emprego) e “tempo” (período de tempo de atuação). Então, é possível verificar dados referentes ao produto considerando uma loja e um intervalo de tempo. Por isso, a denominação dimensional, pois o cubo agrupa os dados com uma estrutura integrada que possibilita uma análise completa do fato envolvido. Nas próximas seções são abordados componentes do processo de modelagem para a construção de um DW.

2.4.1 Tabela de Fatos

A tabela de fatos é a tabela principal do modelo dimensional. Nessa tabela são armazenadas as informações responsáveis pelo relacionamento de desempenho que comunica o assunto do negócio. Seus registros estão ligados às tabelas dimensões que, por sua vez, apresentam a granularidade dos fatos. Também armazenam os identificadores que proporcionam a relação entre a tabela fatos e as tabelas dimensões.

A tabela fatos possui vários registros (fatos) que, por sua vez, cada um armazena medidas numéricas. Possui uma chave primária que resulta da combinação das chaves das dimensões. A seguir, na Figura 11 é apresentado um exemplo de tabela fato.

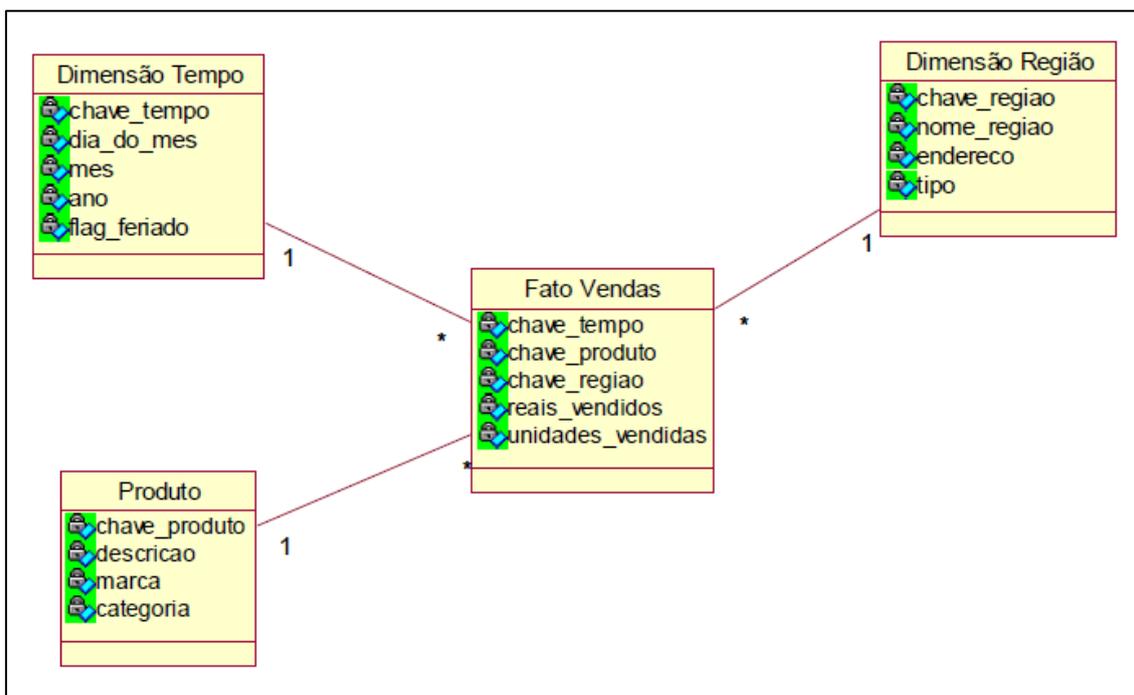


Figura 11: Exemplo de tabela fato (NOLETO, 2008. p. 31).

Na Figura 11 é apresentado um exemplo de tabela fato. A tabela fato Vendas possui as chaves das tabelas dimensões e a partir dessas chaves que as consultas realizadas sobre a tabela acessarão as dimensões e buscarão as informações para responder a busca. Todas as consultas em DW são realizadas a partir da Tabela Fatos. Também possui os campos *reais_vendidos* e *unidades_vendidas*. Esses dados são relevantes ao fato “vendas”, portanto, esses campos também compõe a tabela Fatos.

Segundo Kimball (2002, *apud* COSTA E LIMA, 2006, p. 24), “a tabela fatos é a principal tabela que compõe um modelo dimensional, responsável por armazenar medições (linha em uma tabela de fatos), numéricas ou métricas de desempenho que está relacionada a um assunto ou processo de negócio”. O fato que a tabela se refere está relacionado com o problema que o DW se propõe a solucionar.

2.4.2 Tabela de Dimensões

As tabelas de dimensões são utilizadas para representar as informações e detalhes do fato analisado. Cada tabela dimensão é composta por uma chave primária e um conjunto de atributos, que, por sua vez, definem as regras para concentração dos dados para auxiliar as consultas realizadas na tabela fato. Os atributos tem papel de auxiliar na construção das restrições e conteúdo dos relatórios, e essa atividade é fundamental para que a eficiência na construção do DW bem como sua compreensão (MARTINS, 2007, p.36).

Os dados armazenados nas tabelas dimensões são carregados da base de dados operacionais. Sobre as tabelas, é possível utilizar o recurso de relacionamento entre os atributos de modo que sua estrutura seja organizada de forma hierárquica (KIMBALL 2002 *apud* LIMA, 2006, p. 24). Essa estrutura hierárquica auxilia nas buscas dimensionais e conforme a estrutura e função do DW facilitam as consultas, pois os dados podem vir da base de origem e o DW podendo assim aproveitar a estrutura. O relacionamento auxilia o DW sobre a navegação dos dados, pois pode estar armazenado na tabela de forma resumida, porém, esse recurso permite acessos a outras tabelas o que disponibiliza mais informações na consulta. A seguir serão apresentados os modelos que podem ser utilizados na modelagem de um DW.

2.4.3 Modelo Estrela

O modelo estrela é a abordagem básica da modelagem dimensional e é composto por uma tabela central (tabela fatos), e as tabelas dimensões a ela relacionadas. Essa estrutura possibilita uma leitura fácil por parte dos usuários sobre os dados, pois possui os relacionamentos bem definidos e simplificados. Essa estrutura é geralmente empregada quando usuários não são familiarizados com a estrutura de banco de dados, pois tem grande semelhança com modelos de negócio (LIMA, 2006, p.23). A seguir, na Figura 12, é apresentada a estrutura desse modelo.

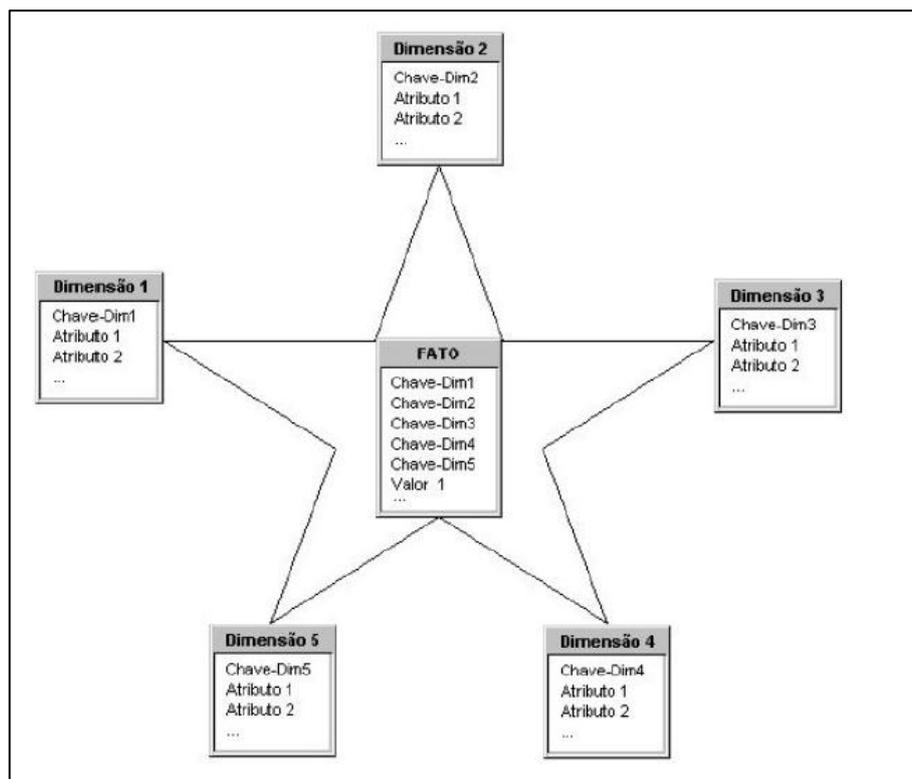


Figura 12: Exemplo de Modelo Estrela (HOKAMA, 2004 apud Martins, 2007, p. 37).

A Figura 12 apresenta um exemplo do modelo estrela. Observa-se neste modelo a tabela central (tabela fato) que possui chaves de cada tabela dimensão e atributos, bem como os relacionamentos, interligando as tabelas. Esse nome é atribuído devido à disposição das tabelas e forma de ligação (relacionamentos) da tabela fato com as tabelas dimensões. Sobre essa estrutura não é utilizada a técnica de normalização das tabelas.

2.4.4 Modelo *Snowflake*

O modelo *snowflake* (flocos de neve), ao contrário do modelo estrela, recomenda a normalização das tabelas até a 3ª forma normal, de modo que sejam retirados os atributos que possuem dependência de outros atributos. Outro fator que aumenta a complexidade dessa abordagem é o fato de poder possuir mais que uma tabela fato e as dependências possibilitarem a navegação entre as mesmas.

A normalização torna essa abordagem mais complexa, aumentando a possibilidade do usuário compreendê-lo na condição de uso, visto que a normalização cria dependência entre as tabelas e conseqüente, também aumenta o número de tabelas (KIMBALL, 1998, p. 32). A seguir, a Figura 13 apresenta um exemplo do modelo *snowflake*.

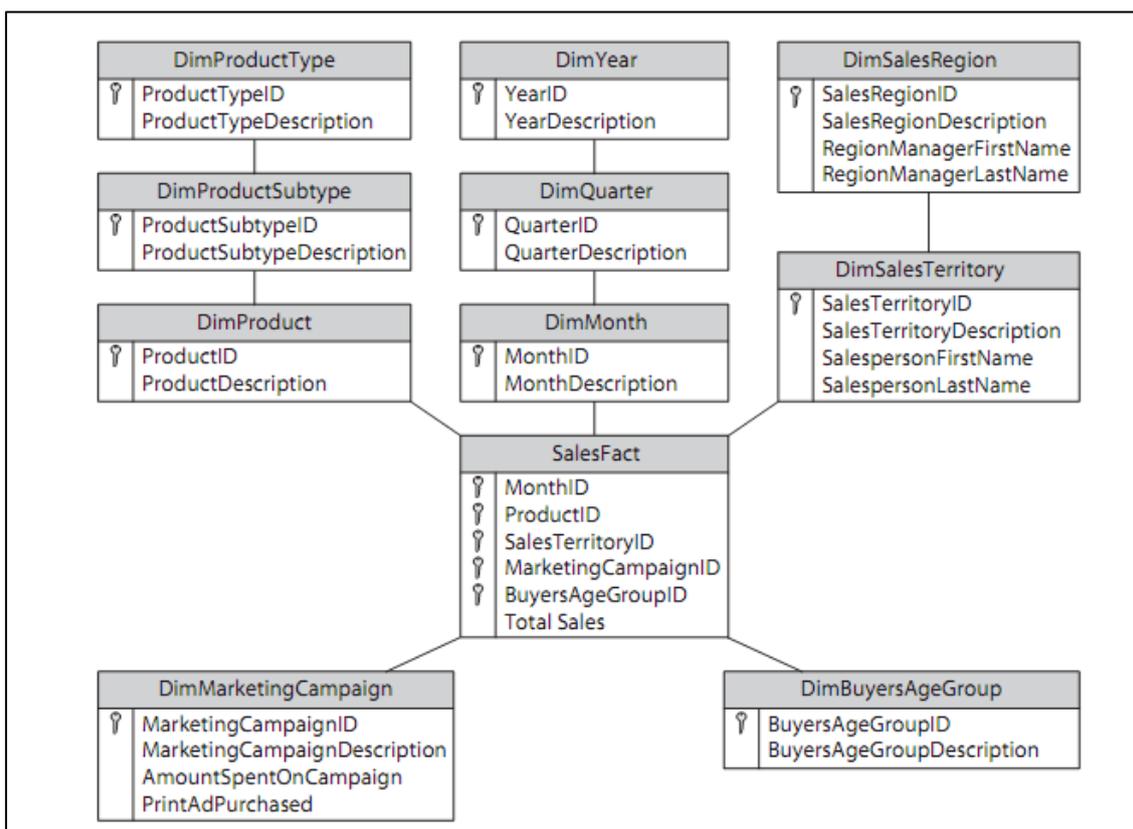


Figura 13: Exemplo de Modelo Snowflake (MORENO, 2010, *on-line*).

Na Figura 13 é apresentado um exemplo do modelo *snowflake*. Observa-se a figura a partir da tabela “*SalesFact*”, que é a tabela Fato Vendas, que possui as tabelas dimensões a ela relacionadas. Esse exemplo utiliza a partir de uma tabela dimensional, a dimensão produto e as tabelas ligadas por relacionamentos (*subtype* e *productype*) foram criadas a partir da aplicação das regras de normalização e, conseqüentemente, cria-se esse efeito de hierarquia das tabelas dimensões.

2.5 Processo ETL (*Extraction, Transformation and Load*)

O processo ETL é responsável por extrair dados de uma base de dados e transformar o mesmo a partir de um padrão configurado e útil e, por fim, carregá-lo (inserir) em outra base. Esse processo é aplicável no processo de implementação de DW, pois permite consolidar dados de fontes diferentes, possibilitando a comunicação com diferentes bases e realizar a leitura de arquivos de diversos formatos.

Esse processo é um momento crítico na implementação de DW, por envolver atividades como: movimentação, modelagem e padronização de um volume considerável de dados (ABREU, 2008, p. 1). O processo é dividido em três etapas:

- Extração – nessa fase é realizada a seleção dos dados úteis e os mesmos são carregados para a base de dados do DW.
- Transformação – nessa fase é realizada a limpeza dos dados a fim de construir um padrão dentre os dados que tenham a mesma representação semântica, porém estão representados em outro formato.
- Carga – nessa fase realiza a integração dos dados garantindo sua integridade.

Esse processo é de suma importância, especialmente para DW, pois as informações nele extraídas atuam no processo decisório da organização. Nesse processo é necessário que os dados sejam devidamente manipulados, caso contrário, oferece grande risco, uma vez que as decisões da organização serão baseadas nessas informações, afetando diretamente os negócios (FERREIRA, 2010, p. 2).

No processo ETL deve-se buscar as informações mais importantes nas bases de dados existentes, úteis (internas e externas) e de interesse da organização para o processo de implementação do DW. Durante a fase de extração dos dados podem ocorrer problemas com os acessos as bases, tendo em vista a heterogeneidade das fontes bem como das plataformas o que pode exigir a aplicação de formas distintas de extração dos dados das bases envolvidas.

A próxima fase é a transformação dos dados. Nesse momento a atividade realizada não é somente a transformação, mas também a limpeza dos dados, para que os dados estejam livres de erros de digitação e da presença de caracteres que possam comprometer o valor e conteúdo para a base do DW. Nessa fase também se toma precauções sobre a duplicidade de informações, cuidado para não descaracterizar os dados, ambiguidade da informação e consistência do dado. Tudo isso a fim de garantir uma homogeneidade para que os arquivos sejam inseridos na base de dados do DW (PALITOT, 2007, p. 20). Essa fase, pode ser exemplificada da seguinte forma: os atributos dataDeNascimento em determinada fonte, podem estar no formato YYYY/MM/DD. Em outra tabela pode estar no formato, DD/MM/YY. O processo realizará a carga e transformação e passará a ser representado no formato DD/MM/YYYY. Essa atividade é realizada para evitar os conflitos da modelagem e para tanto, são criadas regras de conversão para os padrões que são descritos no ambiente de DW.

Finalizando o processo, é a fase de carga. Nesta fase os dados são carregados no ambiente de DW ou DM conforme as necessidades da organização. Nessa fase ocorrem mapeamentos que respeitam as restrições e integridade, para criar a visão unificada das bases, obedecendo às descrições sintáticas e semânticas do dado original. Essa fase possui um grau de complexidade elevado, pois da sua eficiência depende a heterogeneidade dos dados.

Ao finalizar a construção do DW e o alimentar sua base de dados, a primeira etapa estará concluída. O ambiente proporcionado pelo DW centraliza os dados na organização, mas ainda é necessário o emprego de ferramentas adequadas que possibilitem a extração de conhecimento a partir dos dados armazenados, sendo que, muitas ferramentas podem ser empregadas para acessar e auxiliar no processo de análise dos dados denominado On Line Analytical Processing (OLAP). A seguir, não seção 2.6 são abordados os conceitos de OLAP.

2.6 On-Line Analytical Processing (OLAP)

OLAP é uma ferramenta *BI* utilizada para apoiar organizações na atividade de análise dos dados, a fim de construir base de conhecimentos para auxiliar no processo de tomada de decisão. Por meio de uma ferramenta OLAP é possível acessar um ambiente de DW e possibilitar a realização de análise dos dados. O termo OLAP faz referência ao conjunto de ferramentas integradas para análise de dados e a fim de auxiliar os processos de descoberta de informações e decisório. OLAP atua sobre a base de dados aplicando cálculos, fórmulas e utilizando algoritmos de inteligência artificial, para que os dados armazenados sejam transformados em informações úteis a organização (MACHADO, 2004, p.85).

As ferramentas OLAP auxiliam na atividade de análise de dados em *real-time* a fim de auxiliar a solucionar questões nos âmbitos gerenciais e executivos da organização. Essas ferramentas geram soluções para extração de dados das bases operacionais e na construção de relatórios. Essas ferramentas surgiram em paralelo aos SADs para explorar os dados oriundos de DM e DW.

OLAP analisa os dados a partir de um ambiente de DW, porém, os DW e DM, por sua vez, tem suas bases de dados alimentadas a partir de sistemas operacionais que utilizam bases de dados que utilizam o modelo OLTP (*On-Line Transaction Processing*). Esse modelo auxilia no armazenamento dos dados operacionais, resultantes das rotinas da organização e oferece suporte a um ambiente normalizado e com consultas simples, com baixo fluxo de dados (PETRINI, 1999 *apud* PETRINI, 2006, p. 26). Conforme o uso, esses sistemas começam a possuir uma base de dados com uma grande quantidade de dados e, por isso, o sistema tem uma queda de performance, devido o aumento do tempo para percorrer os dados na base. Mas essas não são as únicas diferenças entre OLTP e OLAP, portanto, a seguir, na Tabela 2 são apresentadas as diferenças dos modelos.

Tabela 2: Diferenças entre OLTP e OLAP (LIMA, 2006, p.16).

Características	Transacional OLTP	Analítico OLAP
Usuários típicos	Usuário em geral	Gerentes e analistas de negócio.
Aplicação do sistema	Operações de rotinas	Análises de negócio
Interação do usuário	Pré-determinado	<i>Ad-hoc</i>
Características do trabalho	Leitura/Gravação	Leitura
Unidade de trabalho	Transação	Consulta
Processamento	Orientado a processos	Orientado a assuntos
Atualização	Um registro por vez	Vários registros por vez

Conforme Tabela 2, nota-se que os sistemas que utilizam OLAP e OLTP tem finalidades distintas. Sistemas com OLTP em geral, são sistemas do tipo STP, apresentado anteriormente nesse trabalho. Esses sistemas não possuem funcionalidades de análise de dados, mas de rotinas e, por isso, possuem essa classificação. Com isso, percebe-se que os usuários desses sistemas desempenham atividades para alimentar a base de dados da organização. Os sistemas possuem uma interação limitada com o usuário, ou seja, todas as atividades são predeterminadas, pois os usuários não desempenham atividades com responsabilidades de gerar relatórios ou dados e suas atividades são transacionais, o que gera um fluxo também pré-determinado. São orientados a processos, ou seja, são atividades relacionadas a setores da organização, que seu fluxo no nível horizontal da organização. E esses sistemas não possuem necessidade de manter um histórico dos dados, mas somente registrar as atividades.

Por sua vez, sistemas OLAP são voltados para o processo de tomada de decisão e, por isso, atendem a usuários responsáveis por analisar e gerenciar atividades. Desta forma, as atividades realizadas sobre eles são somente de leitura na base de dados e atuam sobre assuntos definidos pelo usuário. São projetados para reunir e processar um alto nível de dados, pois esses sistemas realizam atividades como relação entre informações, montagem de relatórios e

cálculos, para oferecer consistência e clareza sobre o conhecimento extraído. A seguir, na Figura 14 são apresentados os tipos de SI's que utilizam os modelos OLAP e OLTP.

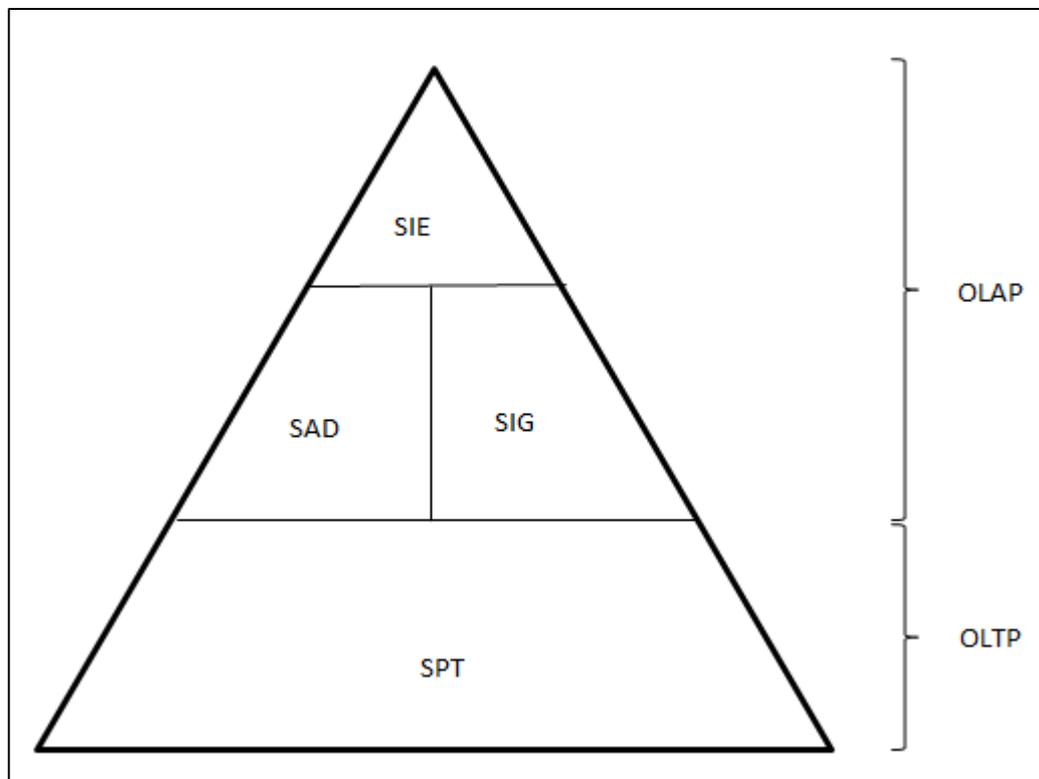


Figura 14: Atuação de OLAP e OLTP em SI (Adaptada de LAUDON e LAUDON, 1997 *apud* MACCARI, 2006, p. 6).

Conforme apresentado na Figura 14, os SPT que tratam da parte operacional, responsável por armazenar dados referentes à rotina utilizam o modelo OLTP, pois as transações desses sistemas são consultas básicas e o ponto principal é o armazenamento de informações. Os SI's que possuem montante de dados é bastante elevado e as operações realizadas são de maior complexidade, faz-se necessário à utilização de ferramentas que auxiliem no processo de extração de dados. No processo de análise o modelo OLAP é eficiente, pois permite que os usuários adquiram conhecimento na exploração dos dados. Observando o processo de BI, o emprego de OLAP é na fase final do processo, pois OLAP possibilita aos usuários acessar e extrair dados de uma base dimensional. Desta forma, oferece suporte ao processo decisório, analisando os dados do DW. Na próxima seção são abordadas as

características de OLAP. Na seção a seguir são apresentadas as operações de OLAP (LUSTOSA, 2009, p. 67).

2.6.1 Operações Olap

As ferramentas OLAP disponibilizam algumas operações para o acesso a dados (LIMA, 2006, p. 17). As principais são:

- **Drill Down** – operação que oferece ao usuário a possibilidade de aumentar o nível de detalhamento e diminuição do nível de granularidade. Assim, possibilitando um alto nível de detalhamento de dados.
- **Drill Up ou Roll Up** – esta operação possibilita ao usuário reduzir o nível de detalhamento do dado e por isso, aumentar a granularidade.
- **Drill Cross** – possibilita ao usuário utilizar níveis de detalhamentos específicos. Com essa operação, pode-se ir diretamente para o detalhamento desejado, sem antes passar pelas fases ou detalhes que orientam a base de dados.
- **Drill Throught** – essa operação permite ao usuário a mudança de uma dimensão para outra. Exemplificando essa situação, em determinado momento os dados da dimensão tempo estão sendo analisados, mas posteriormente, pode-se analisar qualquer outra dimensão.
- **Slice and Dice** – essa operação permite que o usuário defina sobre quais partes da base deseja realizar a busca, ou seja, essa operação permite limitar a consulta.

As operações em OLAP estão relacionadas com a arquitetura definida, assim sendo, na seção 2.6.2 são abordadas as arquiteturas de OLAP para aplicação sobre bases de dados para apoiar o processo de tomada de decisão.

2.6.2 Arquitetura

A definição da arquitetura de OLAP está relacionada à forma de armazenamento (PALITOT, 2007, p. 17). As especificações das arquiteturas

disponíveis bem como sua utilização são empregadas conforme as necessidades do projeto. Conforme a estrutura do projeto, e o local de armazenamento for um banco de dados multidimensional, podem ser utilizados os casos MOLAP e/ou ROLAP. Se um banco de dados relacional, HOLAP e DOLAP ou ainda, arquivos em base de dados local, DOLAP e WOLAP. A seguir, são apresentadas as características e especificações das arquiteturas.

- *Multidimensional On Line Analytical Processing (MOLAP)*: essa é a arquitetura tradicional OLAP. Utiliza os dados armazenados em cubos dimensionais e não oferece suporte para banco de dados relacionais. Possibilita que os usuários manipule os dados do cubo diretamente no ambiente do servidor. Esse fato proporciona um desempenho elevado e conseqüentemente, possibilita uma ágil recuperação de dados na construção dos cubos. Todo o processamento é efetuado na criação do cubo o que disponibiliza os dados integrais em menos tempo. Porém, o alto desempenho faz com que a quantidade de dados gerados seja limitada.
- *Relational On Line Processing (ROLAP)*: essa arquitetura armazena os dados em banco de dados relacionais e atua sobre os dados utilizando as funções da cláusula '*WHERE*'. Não possui limites na quantidade de dados analisados, porém há o limite do próprio banco de dados utilizado. Por não haver restrições sobre o fluxo de dados, torna seu desempenho baixo e os relatórios se comparam às consultas realizadas em SQL.
- *Hybrid On Line Analytical Processing (HOLAP)*: essa arquitetura é uma combinação de MOLAP e ROLAP, ou seja mescla os pontos positivos de cada uma. Trabalha com a alta performance da MOLAP com a escalabilidade da ROLAP. O desempenho é otimizado porque utiliza cubos dimensionais para armazenar as chaves das informações e o detalhamento no banco de dados relacional.
- *Desktop On Line Analytical Processing (DOLAP)*: quando o servidor recebe uma instrução SQL de um cliente, o retorno é um microcubo de informações. Nessa arquitetura o processamento OLAP acontece no cliente e assim, não sobrecarrega o servidor no processamento dos dados. E essa característica também é uma limitação, pois como os dados são analisados no cliente, as informações no cubo são reduzidas e por isso é denominado como estrutura de resposta um microcubo.

- *Web On Line Analytical Processing* (WOLAP): essa arquitetura utiliza os recursos de OLAP a partir do *browser*. Vale ressaltar a evolução do próprio ambiente de processamento (*web*) e a geração de novas ferramentas OLAP para o ambiente. Nessa arquitetura, trabalha em conjunto com as outras já citadas, pois aproveita da arquitetura para análise e armazenamento.

3 MATERIAIS E MÉTODOS

Nesta seção apresentam-se os detalhes sobre o conceito *Data Warehouse*, bem como a descrição e a configuração das ferramentas que compõe o processo de desenvolvimento.

3.1 Hardware

Para o desenvolvimento deste projeto foi utilizado computador pessoal, com a seguinte configuração: Intel *Core 2 Duo* 2.00 Ghz 3 Gb de RAM. Não foi necessário nenhum equipamento específico para o desenvolvimento da aplicação.

3.2 Software

Para a realização deste trabalho foram utilizados os seguintes softwares e ferramentas:

- *Microsoft Access 2010*;
- *Microsoft Excel 2010*
- *Microsoft SQL Server 2005*;
- *Microsoft SQL Server 2008*;
- *Microsoft Visual Studio 2008*;
- Ferramenta OLAP *Microsoft Analysis Services*.

3.3 Metodologia do Trabalho

Para o desenvolvimento deste trabalho foram realizadas pesquisas em trabalhos de conclusão de curso, dissertações de mestrado e fóruns, a fim de coletar maior quantidade de informações sobre Sistemas de Informação,

Sistemas de Apoio à Decisão, construção de *Data Warehouse* e ferramentas OLAP e, assim, orientar o processo de descoberta de informações e processo decisório. Posteriormente, foi realizada a pesquisa bibliográfica, com base em artigos científicos, dissertações, fóruns e especificações de fabricantes de ferramentas e tecnologias relacionadas à DW.

Para conhecimento do domínio, foi necessário realizar entrevistas com o diretor de mobilização da juventude, coordenadores e pedagogos responsáveis pelo programa bolsa universitária, a fim de obter informações sobre o programa, condições e requisitos de ingresso, análise de cadastros e classificadores. Também foram identificados todos os documentos (arquivos de textos e planilhas), bem como as fontes de dados, cadastros e planilha de valores das bolsas.

A partir da base de dados, textos e planilhas, além das entrevistas, puderam-se entender as necessidades dos gestores, bem como a estrutura de organização dos dados nos documentos. Partindo da identificação dessa estrutura, foi construída uma base de dados relacional para iniciar o processo de padronização dos dados e facilitar o processo de construção do DW. Para tanto, foi utilizado o *Microsoft SQL Server 2008*. A definição dessa ferramenta se deu pela facilidade em integrar os arquivos fontes de dados com a ferramenta SGDB, considerando os recursos nativos da ferramenta, sendo que, utilizar outra ferramenta seria necessário implementar um mecanismo de leitura dos dados na base de origem e posteriormente, a inserção dos dados para a base de dados. Posteriormente, foi necessário realizar a inserção dos dados dos arquivos disponibilizados na base de dados OLTP construído para, na sequência, realizar o processo ETL e inserir os dados no DW. Para realização do processo ETL foi utilizado o recurso de *Data Transformation Services (DTS)* do *Microsoft SQL Server 2008*.

Por fim, tendo a base de dados alimentada, iniciou o processo de análise dos dados e, para tanto, foi utilizado a *SQL Server Analysis Services (SSAS)*, como ferramenta OLAP. Esta ferramenta possibilitou a realização e emissão de relatórios para a extração de conhecimento e, conseqüentemente, ofereceu suporte ao processo decisório sobre áreas de investimento, alocação de recursos, verificação da aplicação das bolsas e a atuação do programa no âmbito da Instituição de Ensino Superior com relação aos cursos atendidos.

3.3.1 Definição da Abordagem de Desenvolvimento do *Data Warehouse*

Para a construção do *Data Warehouse* para a base deste trabalho, foi definida a utilização da abordagem *Bottom-Up*. Essa decisão foi baseada na estrutura da Secretaria da Juventude e dos Esportes e a atuação do programa Bolsa Universitária. O programa é de responsabilidade de uma Diretoria Específica sendo que a Secretaria possui outros setores e responsabilidades e o SAD baseado em DW em questão, contempla as responsabilidades e oferecerá resultados de um local específico. Assim sendo, esse será um *Data Mart* que auxiliará os coordenadores do programa Bolsa Universitária e caso haja interesse por parte dos chefes de setores e do gestor, podem ser construídos outros DM's que serão integrados e comporão o *Data Warehouse* que contemplará todo o processo da Secretaria Estadual dos Esportes e da Juventude.

3.3.2 Definição do modelo dimensional

Para o desenvolvimento desse trabalho, foi definida a utilização do Modelo Estrela. Essa decisão foi baseada na estrutura de armazenamento dos dados encontrados nos documentos de textos, planilhas eletrônicas e bases de dados do Microsoft Access. Sobre os dados dos bolsistas, não havia preocupação de armazenamento detalhado das informações o que nas especificações e conceitos de DW, referem-se ao detalhamento do dado. O detalhamento no conceito de DW define o nível de granularidade dos dados. O outro modelo utilizado é o modelo conhecido como *Snowflake*. Esse modelo é utilizado quando os dados estão com um nível de detalhamento maior e as tabelas de armazenamento dos mesmos podem passar por um processo de normalização. Os dados relacionados ao programa de governo disponibilizados possuem baixo nível de detalhamento, desta forma, o modelo adotado para construção do DW foi o Modelo Estrela.

3.3.3 Definição do modelo de Sistema de Apoio à Decisão

Para a realização deste trabalho, foi identificado que o SAD desenvolvido irá atuar na resolução de problemas semiestruturados, pois, baseando nos objetivos levantados pela coordenação do programa, verificou-se que a mesma não possuía um elevado conhecimento das variáveis de interferência. Por isso, esse sistema inclusive tem por objetivo oferecer essa visão e, conhecendo-as, iniciar um processo de construção de uma cadeia de conhecimentos e auxiliar no processo decisório.

4 RESULTADOS E DISCUSSÕES

Esta seção apresentará as etapas realizadas para a construção do ambiente DW e a aplicação da ferramenta OLAP, que possibilitará a análise sobre os dados e extração de conhecimento, formando um Sistema de Apoio à Decisão.

4.1 Domínio

O programa bolsa universitária é um programa de governo estadual que oferece bolsa de estudos a alunos carentes matriculados em instituições de ensino superior não gratuito. Assim, têm-se o objetivo de ajudar o aluno a concluir seus estudos sem prejudicar a renda familiar. Além de receber a bolsa-auxílio, o jovem terá a oportunidade de usar os seus conhecimentos em trabalhos comunitários, em contrapartida aos benefícios recebidos. Os cursos atendidos pelo programa devem estar credenciados e regularizados pelo Conselho Estadual de Educação.

Como exigência do programa, para manutenção da bolsa, o bolsista deve estar regularmente matriculado com bom desempenho acadêmico, não podendo haver reprovação por faltas em uma disciplina sequer ou ainda, reprovação em duas disciplinas por notas. Como contrapartida, o bolsista deverá prestar serviços comunitários em entidades e instituições orientados pelo programa conforme a natureza de sua área de formação, ou ainda estar inserido em projetos de pesquisa da instituição de ensino.

A coordenação do programa levantou a necessidade de haver uma ferramenta tecnológica que auxiliasse na gestão do programa e proporcionasse a extração do conhecimento a partir dos dados armazenados sobre os bolsistas contemplados. Assim, foram disponibilizados documentos de textos, planilhas eletrônicas com os dados dos bolsistas para que, por meio deles,

fosse implementado o Sistema de Apoio à Decisão. Também foram realizadas reuniões com a coordenação do programa, que permitiu coletar algumas de suas necessidades de informação:

- Qual preferência de curso segundo classificação socioeconômica?
- Qual preferência de curso segundo candidatos naturais do TO e verificação sobre atuação?
- Qual preferência de curso segundo candidatos naturais de outros estados?
- Qual o perfil socioeconômico dos candidatos?
- Qual o perfil socioeconômico dos candidatos e preferência por área (humanas - exatas)?
- É eficiente a aplicação dos recursos (bolsas) sobre os candidatos?
- Qual a classificação de faixa etária e preferência de cursos?

A partir da identificação das necessidades por parte da coordenação do programa de governo, iniciou a análise dos dados. Estes, em sua grande maioria, foram disponibilizados por meio de planilhas eletrônicas, contendo os dados dos bolsistas. Além disso, também havia dados em arquivos de textos e também em arquivos do Microsoft Access. A partir desse momento, iniciou-se o processo de junção desses dados para um só formato, no caso uma nova base de dados operacional. Nesta fase, os arquivos de textos e as planilhas eletrônicas foram trabalhadas manualmente. Esta tarefa foi necessária para padronização tanto no formato dos dados quanto na estrutura do arquivo, uma vez que não existia um padrão. Os passos posteriores à estruturação e unificação dos dados na planilha são apresentados na seção 4.2.1 que apresenta o processo ETL. A seguir, na seção 4.2 são apresentadas a estrutura e exemplos dos dados disponibilizados sobre o programa de governo que se encontrava em planilhas eletrônicas.

4.2 Estrutura e Exemplos dos Dados Operacionais

Para a iniciação do processo de construção do *Data Warehouse*, foi necessário reunir os arquivos que continham os dados dos bolsistas do programa bolsa

universitária. Assim sendo, foram disponibilizadas planilhas eletrônicas, que continham os dados dos bolsistas, bem como os dados sobre a instituição de ensino superior na qual o bolsista encontrava-se matriculado, conforme apresentado na Figura 15.

	A	B	C	D
1	NOME	ACADEMICODATANASC	ACADEMICONATURALIDADE	IES
2	FERNANDO ANTONIO GARIBALDI	08/01/1951	GOIANIA	CENTRO UNIVERSITARIO LUTERANO DE PALMAS
3	EDSON PEREIRA DA SILVA	05/09/1966	IPOORA	Instituto de Ensino Superior de Porto Nacional
4	EDICLEIA ALVES BARBOSA	30/10/1976	TOCANTINÓPOLIS	universidade catolica do tocantins
5	FRANCISCA DE LIMA	12/01/1964	NAZARÉ DO PIAUÍ	universidade gurupi
6	DORALICE M DA SILVA SOUSA	29/01/1966	BABAÇULÂNDIA	Instituto Tocantinense antonio carlos
7	LUCIREIS OLIVEIRA LIMA	29/04/1974	ARAPOEMA	FUNDAÇÃO EDUCACIONAL DE PARAÍSO
8	MARIA ANDREIA DA FONSECA	21/07/1973	OUVIDOR	FUNDAÇÃO EDUCACIONAL DO BICO DO PAPAGAIO
9	RAGNER FRANKLIN MEDEIROS	26/01/1989	CONCEIÇÃO DO ARAGUAIA	FUNDAÇÃO EDUCACIONAL DO BICO DO PAPAGAIO
10	MIRYAM CRISTINA ALCANTARA QUEIROZ	21/08/1983	TAGUATINGA	FUNDAÇÃO EDUCACIONAL DOM ORIONE
11	JULIANA VIEIRA DA SILVA	02/03/1984	ALVORADA	FUNDAÇÃO UNIVERSIDADE DO TOCANTINS
12	ANTONIA VISGUEIRA DIAS	31/03/1975	DUIRE	FUNDAÇÃO DE ENSINO SUPERIOR DO SUDESTE DO TOCAN
13	HEMERSON PIRES DE FREITAS	14/08/1984	PEIXE	FACULDADE SERRA DO CARMO
14	KESSYA RIBEIRO NOLETO LOPES	14/11/1988	ARGUAINA	sociedade civil de educação continuada
15	MARIA NALVA DE ARAUJO RIBEIRO	01/03/1975	NOVO ALEGRE	sociedade civil de educação continuada
16	ANA AMELIA PIRES FEITOSA	18/12/1969	BABAÇULÂNDIA	sociedade civil de educação continuada
17	GLEUZENIR FERREIRA LEMOS	08/11/1974	GOIANIA	sociedade civil de educação continuada
18	IRISMAR LIMA DE OLIVEIRA	08/06/1985	COLINAS	sociedade civil de educação continuada
19	DIEGO DE MORAES	18/02/1989	IMPERETRIZ	sociedade civil de educação continuada
20	IVONE DA SILVA FERREIRA	11/11/1972	MARANHÃO	sociedade civil de educação continuada
21	AURILENE CASSIMIRO ALENCAR	14/06/1984	GOIATINS	sociedade civil de educação continuada
22	VALERIA LEANDRO MOURÃO	28/08/1989	ARAGUAINA	sociedade civil de educação continuada
23	ANA PRISCILLA FEITOSA RODRIGUES	17/07/1989	ARAGUAINA	sociedade civil de educação continuada
24	ALEXANDRE PEREIRA DIAS	03/05/1986	PORTO NACIONAL	sociedade civil de educação continuada
25	ROSANGELA CRISTINO GOMES	16/05/1978	UBEPLÂNDIA	sociedade civil de educação continuada
26	ALEXANDRE FRANCELINO DE MOURA	30/01/1986	NOVA OLINDA	sociedade civil de educação continuada
27	ELIENNE ROSA SANTANA	31/03/1978	GONÇALVES DIAS	sociedade civil de educação continuada
28	WELTON PEREIRA FRAGOSO	03/12/1980	GURUPI	sociedade civil de educação continuada
29	POLIANNE DE CASSIA CORRÊA	05/06/1988	ANANIAS	sociedade civil de educação continuada
30	MARIA ELENICE DE OLIVEIRA FERNANDES	23/03/1977	COLINAS	sociedade civil de educação continuada
31	MARCIAVAM MIRANDA SILVA	28/08/1977	AXIXÁ	sociedade civil de educação continuada
32	JANAINA GONÇALVES DE ARAUJO FRANCA OLIVEIRA	05/05/1982	MIRACEMA	sociedade civil de educação continuada
33	MATILDA DE PAULA E SILVA	02/02/1989	GOIANIA	sociedade civil de educação continuada

Figura 15: Dados disponibilizados sobre programa de governo.

Na Figura 15 são apresentados os dados dos bolsistas do programa, contendo nome, data de nascimento, naturalidade e instituição de ensino superior. Os dados coletados por meio das planilhas foram organizados e unificados de acordo com essa estrutura. Porém, ainda existiam dados desse programa de governo em outro formato, mais precisamente no formato do banco de dados do *Microsoft Access*. A estrutura dessa tabela é apresentada na Figura 16.

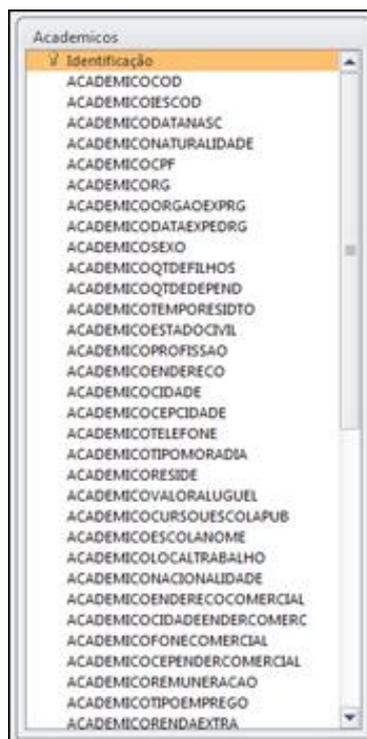


Figura 16: Base de dados *MS Access* criada para programa de governo.

Na Figura 16 é apresentada a estrutura da base de dados criada para armazenamento dos dados dos bolsistas com os campos: ACADEMICOCOD, ACADEMICOSEXO, ACADEMICODATANASC, ACADEMICONATURALIDADE etc. Essa base de dados encontrava-se alimentada e desnormalizada. A seguir, Figura 17, são apresentados exemplos dos dados existentes nesta tabela.

ID	Curso	Data	Local	Nota	Outros
1	63 FASEC	28/08/1989 00:00	ARAGUAINA	030.74	1005 SSP-TO
2	64 FUNDAÇÃO ED	17/07/1989 00:00	ARAGUAINA	023.32	9106 TO
3	65 EDUCON	03/05/1986 00:00	PORTO NACIOI	008.17	7099 SSP-TO
4	66 EDUCON	16/05/1978 00:00	UBERLANDIA	840.77	197.1 SSP-TO
5	67 EDUCON	30/01/1986 00:00	NOVA OLINDA	011.79	4574 SSPGO
6	68 CEULP/ULBRA	31/03/1978 00:00	GONÇALVES DI	872.57	1579 SSP/MA
7	69 EDUCON	03/12/1980 00:00	GURUPI	889.50	315.1 SSP/TO
8	70 ITPAC - ARAGU	05/06/1988 00:00	ANANAS	001.07	5913 POLICIA CIVIL/PA
9	71 CEULP/ULBRA	23/03/1977 00:00	COLINAS	663.36	302.1 SSP/TO
10	72 CATOLICA DO	28/08/1977 00:00	AXIXÁ	853.78	2533 SSP/TO
11	73 EDUCON	05/05/1982 00:00	MIRACEMA	000.95	298.1 SSP/TO
12	74 UNIRG	03/10/1989 00:00	GOIANIA	027.07	7472 SSP-TO
13	75 EDUCON	15/03/1981 00:00	MIRANORTE	935.77	273.1 SSP-TO
14	76 EDUCON	27/07/1984 00:00	TAGUATINGA	023.41	8512 SSP/TO
15	77 EDUCON	09/08/1980 00:00	GOIANIA	938.43	3988 SSP-TO
16	78 EDUCON	07/03/1985 00:00	PORTO NACIOI	007.77	7014 SSP
17	33 ITPAC - PORTO	08/01/1951 00:00	GOIANIA	051.74	3192 SSP/GO
18	39 EDUCON	05/09/1966 00:00	IPORÁ	364.74	4567 SSP-TO
19	44 EDUCON	30/10/1976 00:00	TOCANTINÓPC	983.59	2322 SSP/TO
20	46 EDUCON	12/01/1964 00:00	NAZARÉ DO PI	083.09	1880 SSP-TO
21	47 EDUCON	29/01/1966 00:00	BABAÇULÂNDI	387.15	1728 SSP/TO
22	48 EDUCON	29/04/1974 00:00	ARAPOEMA	765.90	2585 SSP/TO
23	49 UNIRG	21/07/1973 00:00	OUIVIDOR	576.63	1682 SSP-DF
24	50 CEULP/ULBRA	26/01/1989 00:00	CONCEIÇÃO DI	017.89	8604 SSP/TO
25	51 EDUCON	21/08/1983 00:00	TAGUATINGA	003.64	450.1 SSP/TO
26	52 CEULP/ULBRA	02/03/1984 00:00	ALVORADA	973.38	423.1 SSP-TO
27	53 CEULP/ULBRA	31/03/1975 00:00	DUERE	817.87	1834 SSP-TO

Figura 17: Dados armazenados em Microsoft Access.

Na Figura 17 são apresentados os dados armazenados no arquivo do *Microsoft Access* referentes ao programa bolsa universitária. Assim sendo, pode-se observar a disposição dos dados no arquivo.

Os dados coletados nessa fase do desenvolvimento serviram para que fosse planejado e criado um novo banco de dados operacional (apresentado na seção 4.2.1), de modo a organizar melhor estes dados e servir como base para a reativação do programa de governo e, posteriormente, alimentar a base do *Data Warehouse* para manutenção e auxílio no processo decisório referentes ao referido programa. A seguir, na seção 4.2.1 são apresentados o modelo relacional criado e os passos para envio dos dados dos arquivos a este novo modelo.

4.2.1 Processo ETL

O processo ETL no desenvolvimento desse trabalho se deu em duas etapas. Primeiramente, os dados disponibilizados por meio de planilhas eletrônicas e arquivos de textos foram estruturados e organizados em uma única planilha eletrônica. Assim, foi possível conhecer os padrões dos dados úteis à coordenação do programa de bolsa universitária e, conseqüentemente, os

dados necessários para armazenamento sobre os bolsistas. Com isso, foi identificada a necessidade de construção de um modelo relacional, conforme apresentado na Figura 18. Esse modelo foi construído para estruturar e organizar os dados do programa e, desta forma, prover com facilidade o processo de transformação e carga desta base para o ambiente de DW.

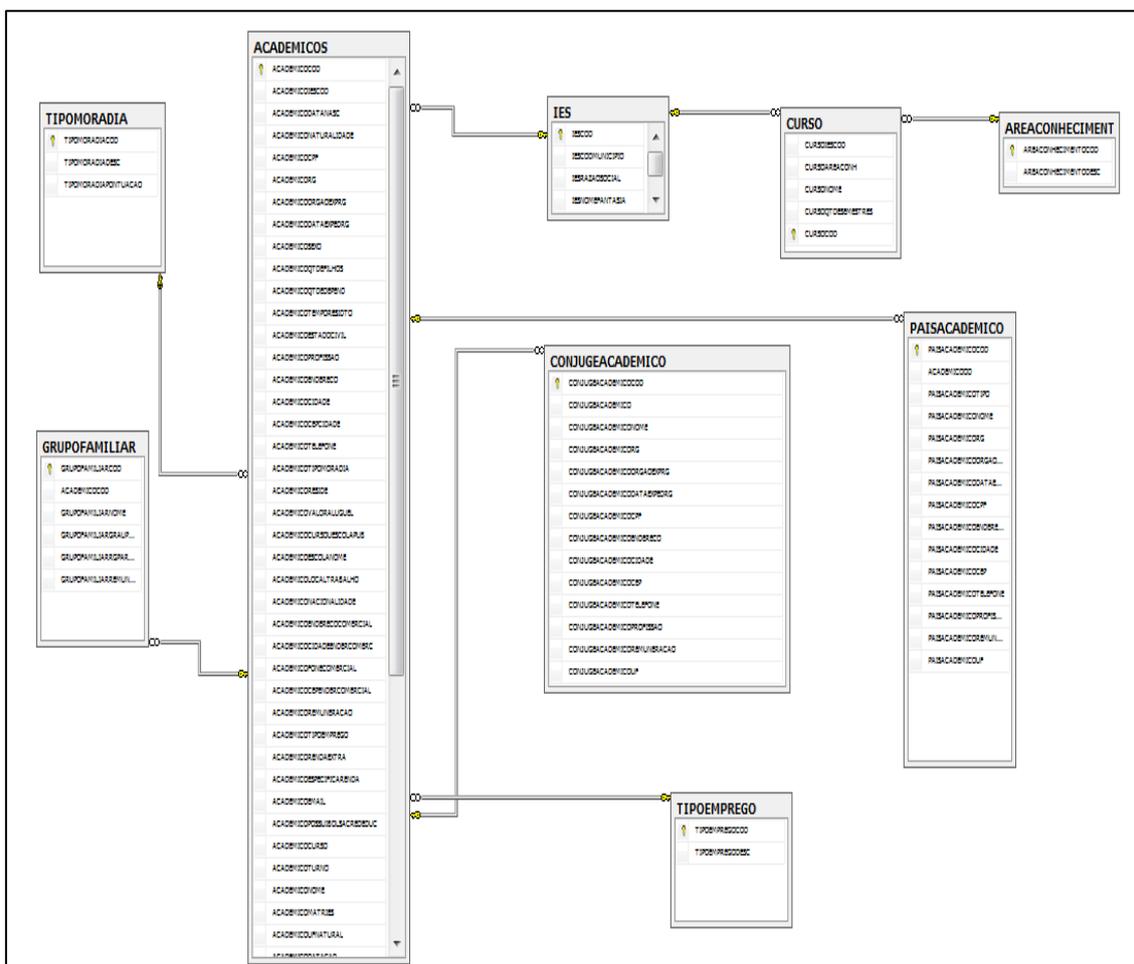


Figura 18: Modelo Relacional.

Na **Erro! Fonte de referência não encontrada.** é apresentado o modelo relacional criado a partir dos dados disponibilizados do programa de governo. Esse modelo operacional possibilitou a organização dos dados e também o processo de tipagem para armazenamento em uma base de dados. Esse modelo possibilitou melhor visualização sobre os dados armazenados para então a construção do modelo dimensional do *Data Warehouse*. A seguir, na seção 4.3 é apresentado o modelo dimensional. O processo de carga dos dados tanto para a base de dados operacional, quanto para a base do *Data Warehouse* encontram-se no ANEXO I deste trabalho.

4.3 Modelo Dimensional

Baseado nas questões levantadas pela equipe coordenadora do programa e com base nos dados disponibilizados foi proposto o seguinte modelo dimensional, conforme apresentado na Figura 19.

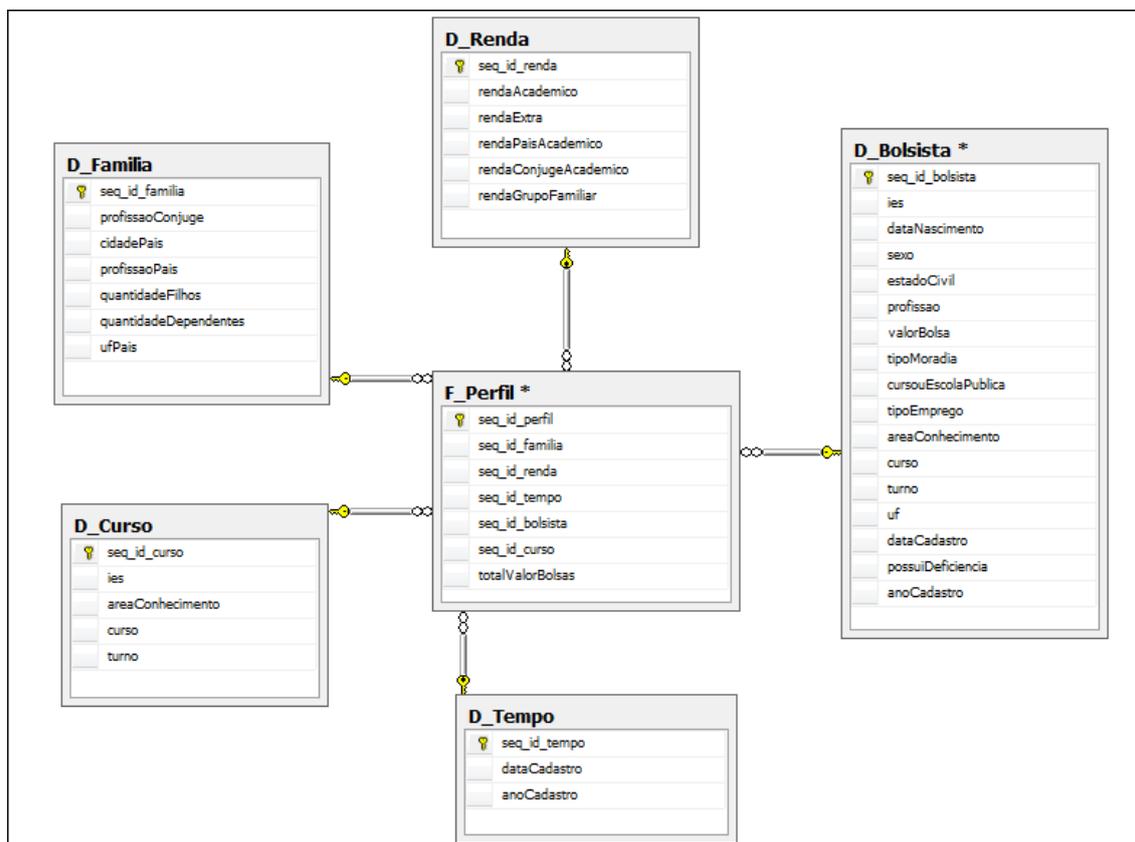


Figura 19: Modelagem dimensional (Modelo Estrela).

Na Figura 19 é apresentado o modelo dimensional criado como parte do processo de desenvolvimento do *Data Warehouse*. Pode-se observar que foram criadas as seguintes tabelas dimensões: D_Curso, D_Familia, D_Bolsista, D_Renda e D_Tempo; e a tabela fato, F_PerfilBolsista. As subseções a seguir apresentam as tabelas do modelo dimensional upracitado.

4.3.1 Tabela Dimensão Curso

A tabela D_Curso possui os dados dos cursos dos bolsistas assistidos no programa. Esta tabela dimensão foi criada para atender a demanda sobre os

curso assistido e ainda reconhecer os cursos e áreas de conhecimento mais atendidas pelo programa. A Tabela 3 apresenta os atributos desta dimensão.

Tabela 3: Representação dos atributos da tabela D_Curso

D_Curso	
Atributos	Descrição
seq_id_curso	Chave primária de D_Curso
Nome	Nome do curso
areaConhecimento	Área de conhecimento relacionada
IES	Instituição de Ensino Superior
Turno	Turno do curso

A seguir, na seção 4.3.2 serão abordados o objetivo bem como a representação da tabela dimensão renda.

4.3.2 Tabela Dimensão Renda

A tabela D_Renda possui os dados relacionados às rendas, a fim de auxiliar no conhecimento do perfil do bolsista. Esta tabela dimensão auxiliará no processo de reconhecimento do perfil socioeconômico, e as rendas declaradas para inscrição são de suma importância para obtenção desta informação. A Tabela 4 apresenta os atributos desta tabela dimensão.

Tabela 4: Representação dos atributos da tabela D_Renda

D_Renda	
Atributos	Descrição
seq_id_renda	Chave primária de D_Renda
rendaConjuge	Renda do Cônjuge do bolsista
rendaBolsista	Renda do Bolsista
Renda_extra_bolsista	Renda extra do bolsista
rendaPais	Renda dos pais do bolsista
rendaGrupoFamiliar	Renda familiar do bolsista

A seguir, na seção 4.3.3 serão abordados o objetivo bem como a representação da tabela dimensão tempo.

4.3.3 Tabela Dimensão Tempo

A tabela D_Tempo possui dados relacionados às datas dos cadastros e das informações e atividades realizadas relacionadas ao programa bolsa universitária. Essa tabela auxiliará a reconhecer as alternâncias do cenário da disponibilização das bolsas e disposição dos recursos considerando intervalos de tempo e as variações entre as disciplinas atendidas considerando esses mesmos intervalos. Na Tabela 5 são apresentados os atributos desta tabela dimensão.

Tabela 5: Representação dos atributos da tabela D_Tempo

D_Tempo	
Atributos	Descrição
seq_id_tempo	Chave primária de D_Tempo
Ano	Atributo que armazena os anos referentes às datas da base
Data	Atributo que armazena as datas

A tabela dimensão tempo utilizada no projeto foi criada manualmente baseada nos dados disponibilizados, porém, as operações sobre o tempo no ambiente do DW, foram criadas automaticamente, com o auxílio do *Microsoft SQL Server Business Intelligence Development Studio* (SSBIDS). Criando as operações sobre o tempo com o auxílio do assistente do SSBIDS. Assim sendo na Figura 20 são apresentadas as operações desta dimensão.

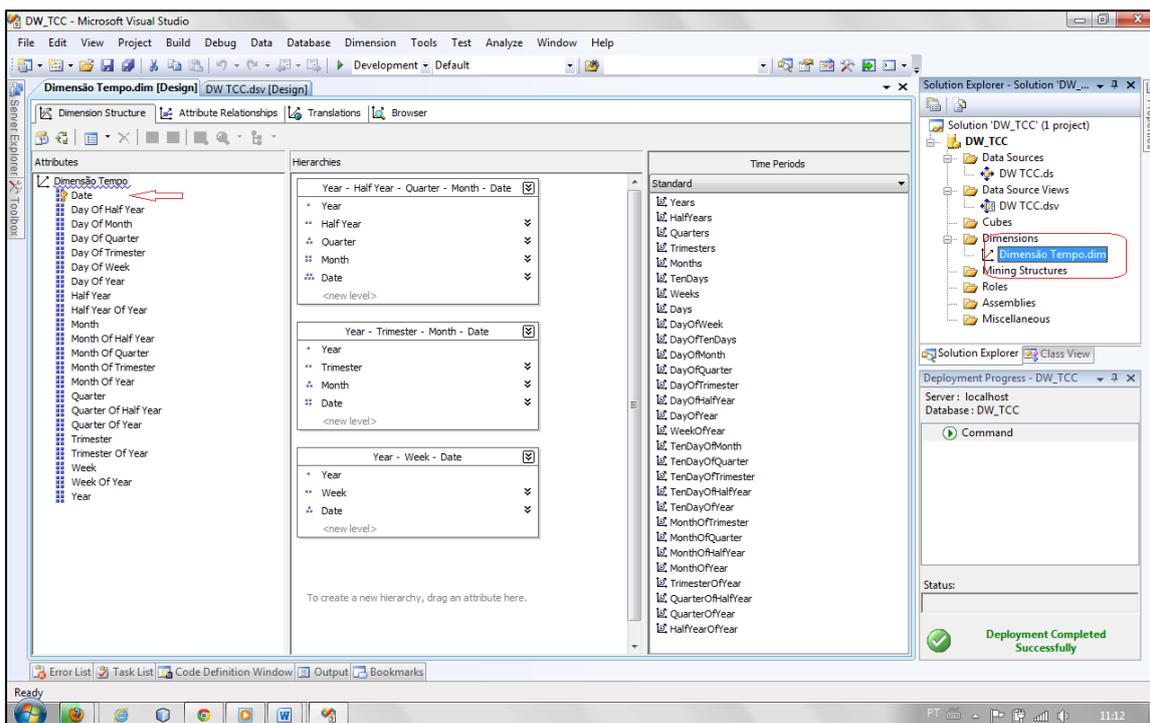


Figura 20: Configuração da dimensão tempo sobre a tabela D_Tempo.

Na Figura 20 são apresentadas as operações e classificações sobre a tabela dimensão tempo. Ali se apresentam a estrutura de organização e as subdivisões do tempo para processamento dos dados (dias, semanas, meses, trimestres etc). Todas essas classificações auxiliarão na extração de conhecimento baseando-se no tempo de inserção dos dados. Assim, o DW poderá emitir dentro dessas expectativas de tempo os relatórios, bem como as marcações para instâncias dos dados inseridos. A seguir, na seção 4.3.4 serão abordados o objetivo bem como a representação da tabela dimensão família.

4.3.4 Tabela Dimensão Família

A tabela D_Familia possui dados relacionados aos familiares dos bolsistas assistidos pelo programa bolsa universitária. Essas características serão utilizadas para agregar dados para extração de conhecimento a respeito do perfil do bolsista. Na Tabela 6 são apresentados os atributos desta tabela dimensão.

Tabela 6: Representação dos atributos da tabela D_Familia

D_Familia	
Atributos	Descrição
seq_id_familia	Chave primária de D_Familia
profissaoPais	Descreve a profissão dos pais do bolsista
profissaoConjuge	Profissão do cônjuge do bolsista
quantidadePessoas	Quantidade de pessoas que residem com o bolsista
quantidadeTrabalham	Quantidade de pessoas residentes com o bolsistas que trabalham
quantidadeMenores	Quantidade de menores que residem com o bolsista

A seguir, na seção 4.3.5 serão abordados o objetivo bem como a representação da tabela dimensão bolsista.

4.3.5 Tabela Dimensão Bolsista

A tabela D_Bolsista possui os dados relacionados ao bolsista. Esses dados representam os dados diretamente relacionados aos bolsistas e tais dados auxiliaram na definição do perfil do bolsista. Na Tabela 7 são apresentados os atributos desta tabela dimensão.

Tabela 7: Representação dos atributos da tabela D_Bolsista

D_Bolsista	
Atributos	Descrição
seq_id_bolsista	Chave primária de D_Bolsista
dataNascimento	Descreve a data de nascimento do bolsista
Sexo	Descreve o sexo do bolsista
tipoResidencia	Descreve o tipo de residência do bolsista

estadiCivil	Descreve o estado civil do bolsista
profissao	Descreve a profissão do bolsista
UF	Descreve a UF de origem de nascimento do bolsista
portadorDeficiencia	Descreve se o bolsista é portador de deficiência
cursouEscolaPublica	Descreve se o bolsista estudou em escola pública
valorBolsa	Descreve o valor da bolsa recebida
tipoEmprego	Descreve o tipo de emprego do bolsista
areaConhecimento	Descreve a área de conhecimento do acadêmico
curso	Descreve o curso do acadêmico
turno	Descreve o turno do acadêmico
anoCadastro	O ano de cadastro no programa

A seguir, na seção 4.3.6 serão abordados o objetivo bem como a representação da tabela fato.

4.3.6 Tabela Fato

A tabela fato F_PerfilBolsista é composta pelas chaves das tabelas dimensões, como chaves estrangeiras e, por meio dessas chaves, relacionam-se com as mesmas, bem como com seus parâmetros de aplicação. Além disto, para facilitar as consultas, esta tabela apresenta as totalizações dos dados referentes a bolsas fornecidas e rendas dos familiares. Isso influenciará diretamente na identificação do perfil do público desse programa de governo. Na Tabela **8** **Erro! Fonte de referência não encontrada.** apresenta os atributos desta tabela fatos.

Tabela 8: Representação dos atributos da tabela F_PerfilBolsista

F_PerfilBolsista	
Atributos	Descrição

seq_id_perfil	Chave primária de perfil bolsista
seq_D_tempo	Chave estrangeira da Tabela D_Tempo
seq_D_Curso	Chave estrangeira da Tabela D_Curso
seq_D_Renda	Chave estrangeira da Tabela D_Renda
seq_D_Bolsista	Chave estrangeira da Tabela D_Bolsista
Seq_D_Familia	Chave estrangeira da Tabela D_Familia
Seq_D_Bolsa	Chave estrangeira da Tabela D_Bolsa
totalValorBolsas	Total do valor das bolsas

Tendo sido criadas as tabelas em um ambiente de banco de dados, faz-se necessário a inserção dos dados nesse ambiente, para que possa iniciar o processo de exploração e emissão de relatórios para a tomada de decisão. Essa atividade está apresentada na seção **Erro! Fonte de referência não encontrada..**

4.4 Desenvolvimento do projeto *Analysis Services* e OLAP

Após a criação do *Data Warehouse*, a sequência do processo consiste em construir os cubos de decisão para efetuar as consultas. Com os cubos construídos, é possível criar relacionamentos entre as dimensões para então aplicar os recursos de OLAP e, então, exportar dados para emissão de relatórios. Para tanto, é necessário iniciar um novo projeto de *Analysis Services* no *SQL Server Business Intelligence Development Studio*. Para tanto, na aba *Solution Explorer*, é necessário adicionar um novo *Data Source*, configurando a conexão com a base de dados e logo após nomear esta operação. Na Figura 21 é apresentado o processo de configuração da base de dados. É necessário selecionar o servidor, configurar o usuário para autenticação e, por fim, selecionar a base de dados.

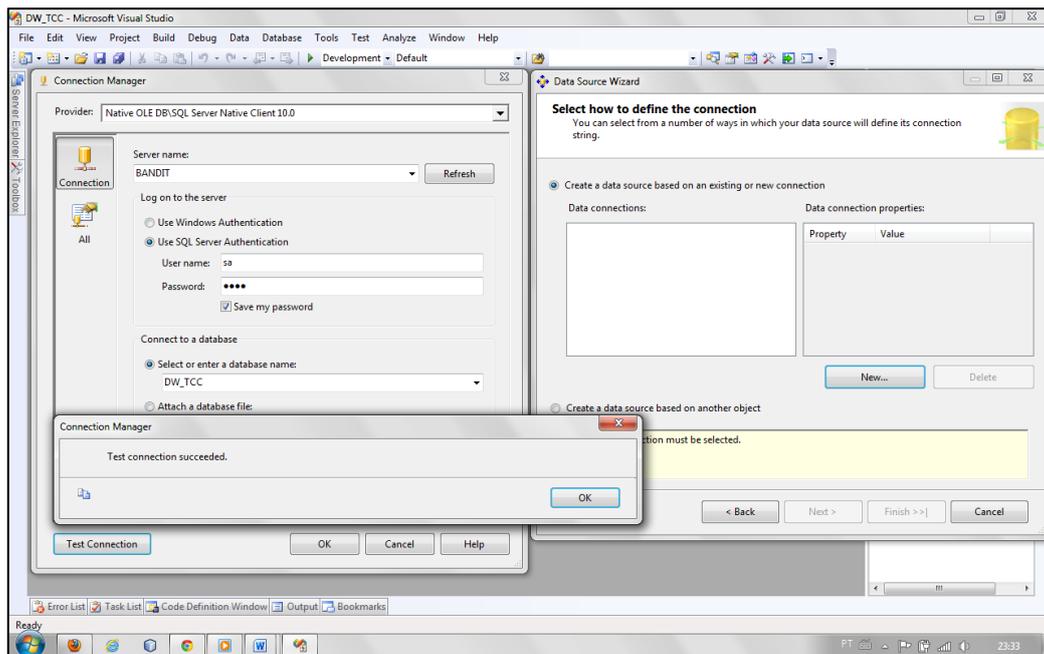


Figura 21: Configurando conexão com base de dados para projeto.

O próximo passo é adicionar um *Data Source View* – que funcionará como um espelho da base de dados e, também, apresentará todos os relacionamentos entre as tabelas, caso ainda não tenham, é possível criá-la. Para tanto, basta adicioná-lo e configurá-lo através do *Solution Explorer*. Em sua configuração deverá ser escolhido o *Data Source* anteriormente criado como fonte de dados, em seguida, escolhe-se as tabelas disponíveis em tal base de dados, restando apenas nomeá-lo e logo após poder-se-á observar o resultado, conforme demonstrado na Figura 22.

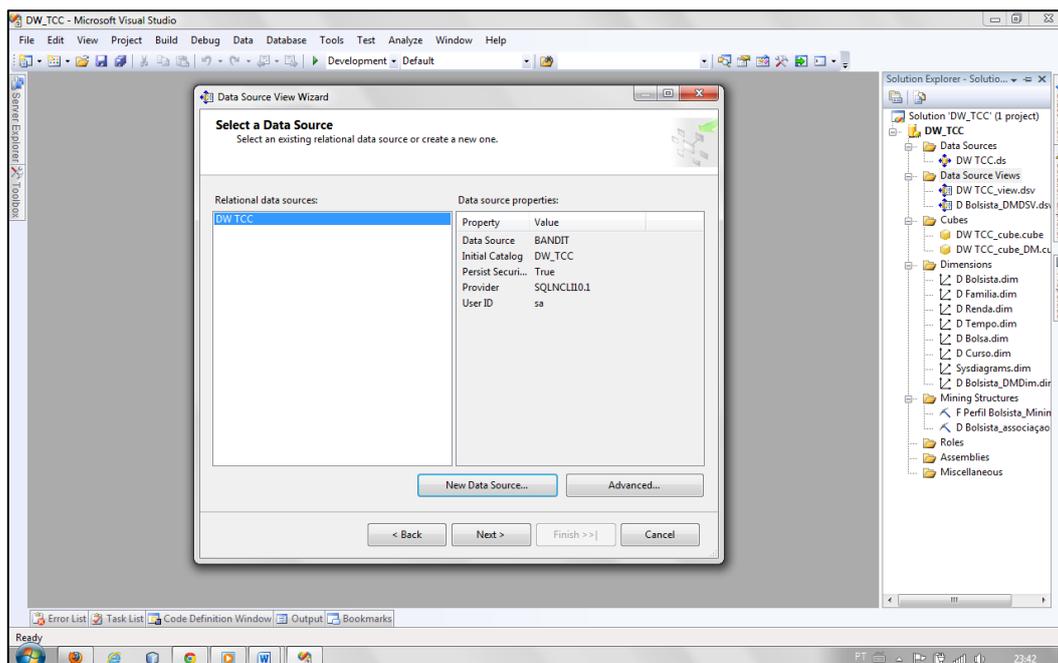


Figura 22: Criação do *Data Source*.

Na Figura 22 é apresentada a definição da fonte de dados para o projeto de *Analysis Services*. Avançando no processo, são listadas as tabelas da base de dados selecionada disponibilizadas para o processo, conforme apresentado na Figura 23.

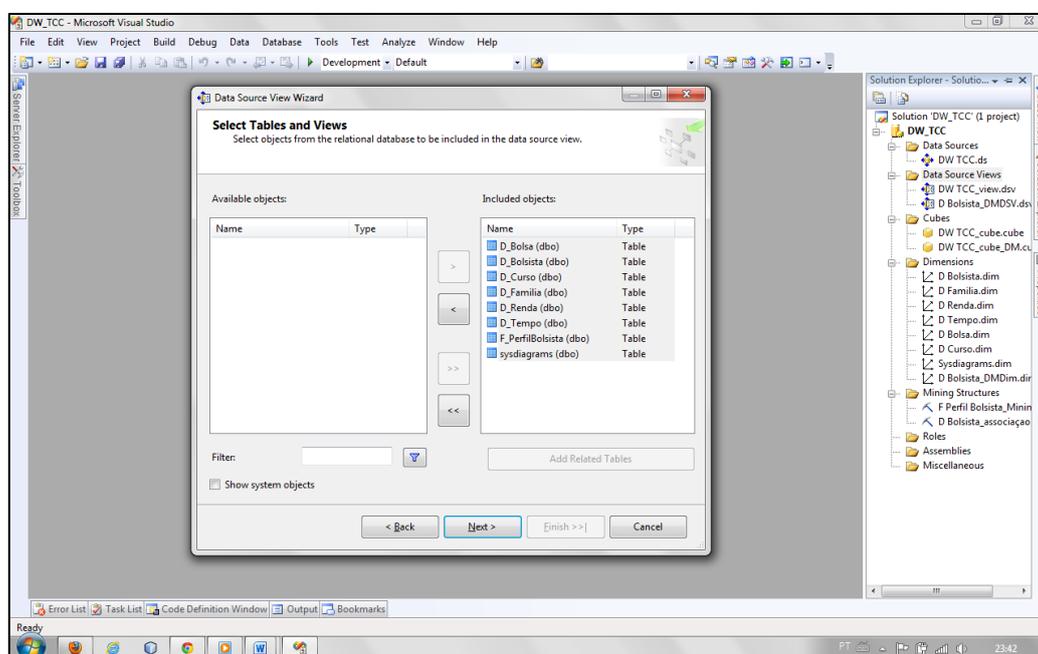


Figura 23: Seleção das tabelas para *Data Source*.

Na Figura 23 é apresentada a opção de seleção das tabelas a serem espelhadas por ele. Neste caso, as tabelas selecionadas estarão disponíveis

para exploração, ou seja, serão utilizadas na construção dos cubos de decisão. Por fim, é necessário nomear o *Data Source View*, conforme apresentado na Figura 24.

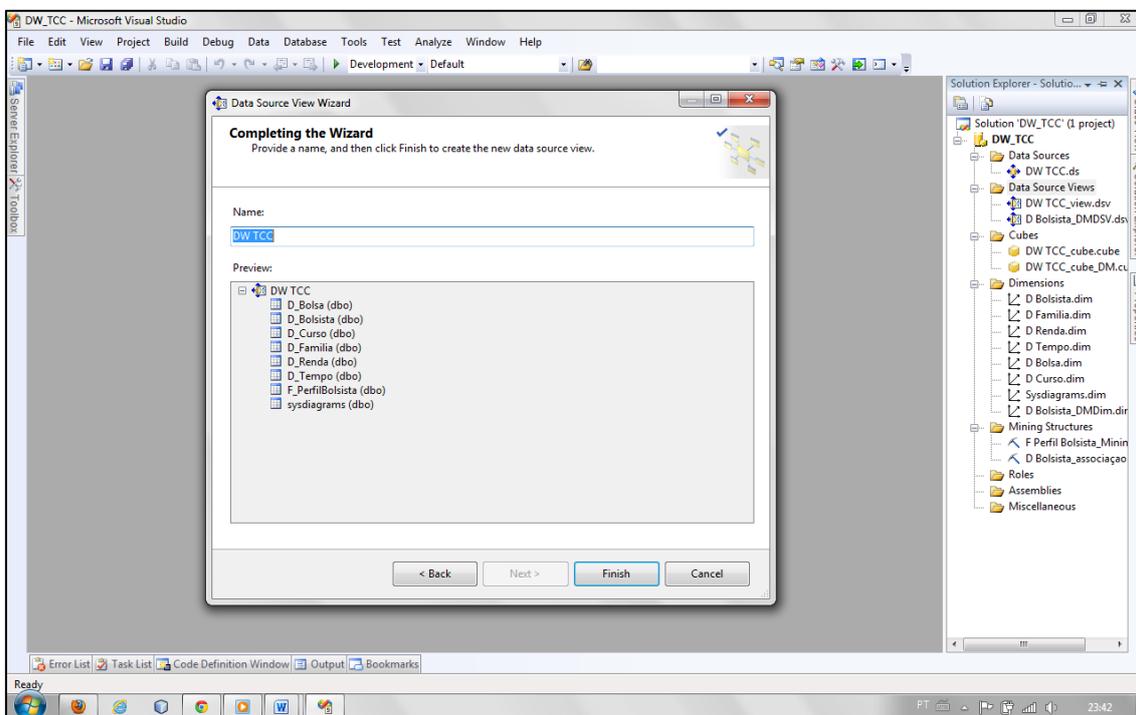


Figura 24: Nomeação do *Data Source*.

O resultado desta operação é a apresentação das tabelas e seus relacionamentos, conforme Figura 25.

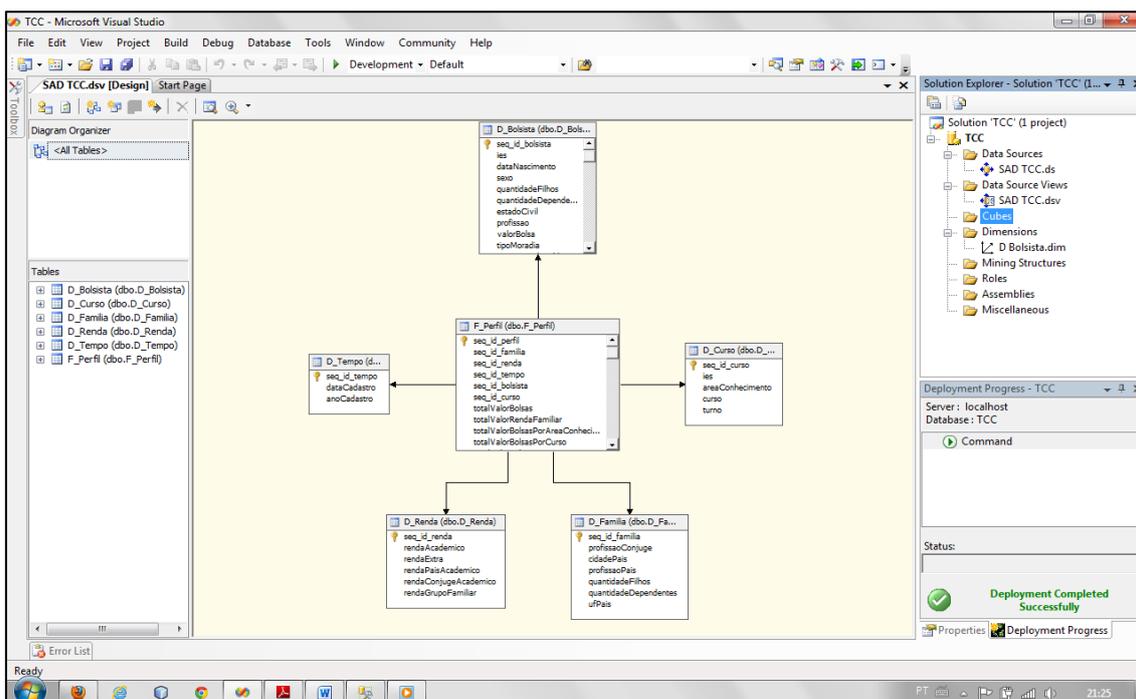


Figura 25: Apresentação do *Data Source View*.

Na Figura 25 é apresentado o resultado da criação do cubo, mostrando as tabelas envolvidas e os respectivos relacionamentos. A seguir, na seção 4.4.1 é apresentada a definição dos cubos que auxiliarão nas consultas na base de dados.

4.4.1 Definição dos Cubos

Na criação dos cubos, definem-se as consultas que serão realizadas sobre a base. A definição dos cubos é criada ao selecionar no projeto a opção *cubes* e com o botão direito a opção *New Cube*. O início da criação do cubo se dá por escolher se utilizará as fontes de dados anteriormente criadas ou *templates*. Na opção de utilizar os *Data Sources* criados, a opção *auto build* permite que o próprio serviço crie os cubos com dimensão e agregações, mas não é recomendado, pois utilizará todas as colunas da tabela fato (até as agregações) e todas as colunas das tabelas de apoio, como dimensões, gerando cubos grandes e com dados desnecessários. A seguir, a partir da Figura 26, são apresentados os passos para criação do cubo.

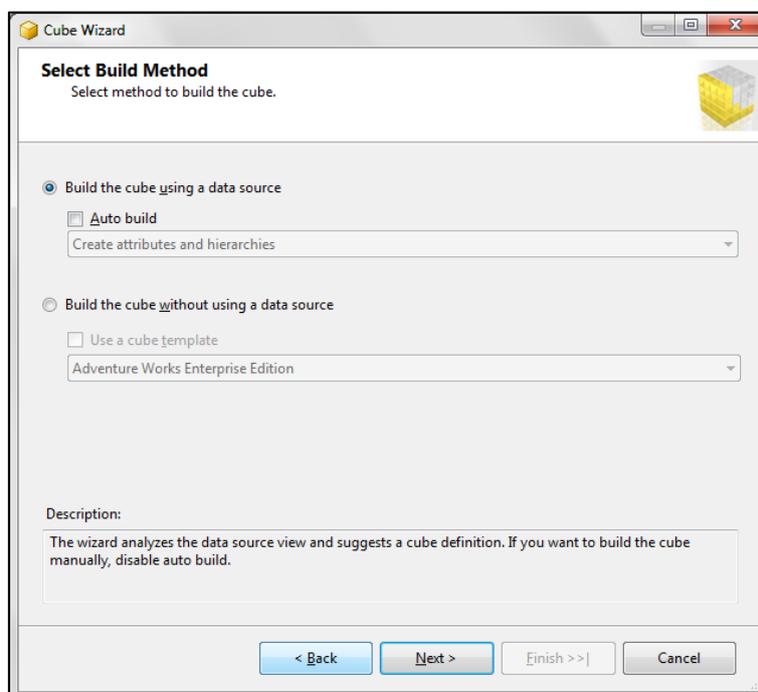


Figura 26: Definição do tipo e construtor a ser utilizado no cubo.

O passo seguinte é definir dentre as apresentações de dados (*data source view*) qual será utilizada como fonte para a geração do cubo, conforme apresentado na Figura 27.

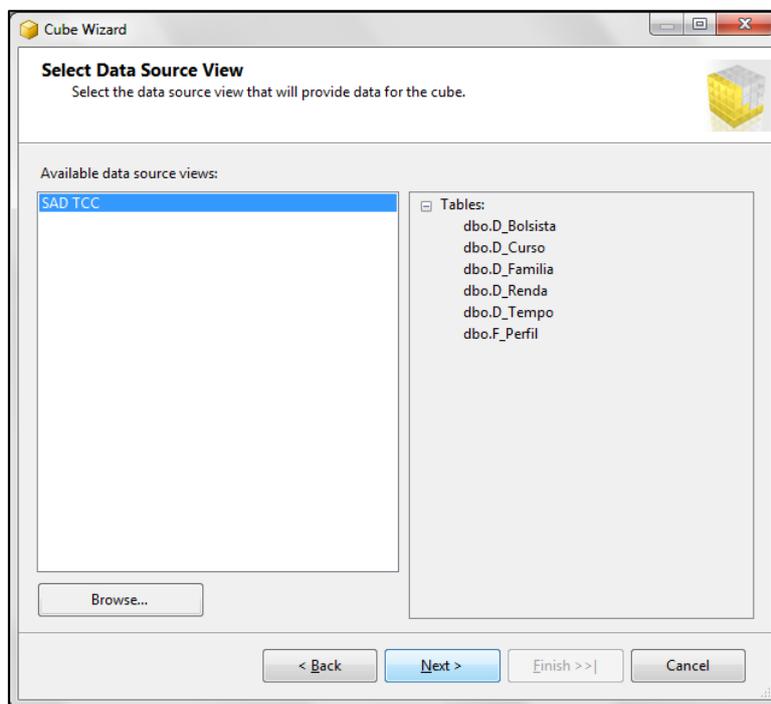


Figura 27: Seleção da fonte de dados

Na Figura 27 é apresentado o processo de criação de um cubo. O cubo é criado sobre o espelhamento da base de dados, ou seja, sobre o *Data Source View* criado no projeto. Assim sendo, na criação do cubo é necessário selecionar este objeto criado para disponibilização das tabelas e em consequência os dados. A seguir, na Figura 28 é apresentado o processo de definição das tabelas fato, dimensão tempo e demais dimensões do *Data Warehouse* para o cubo.

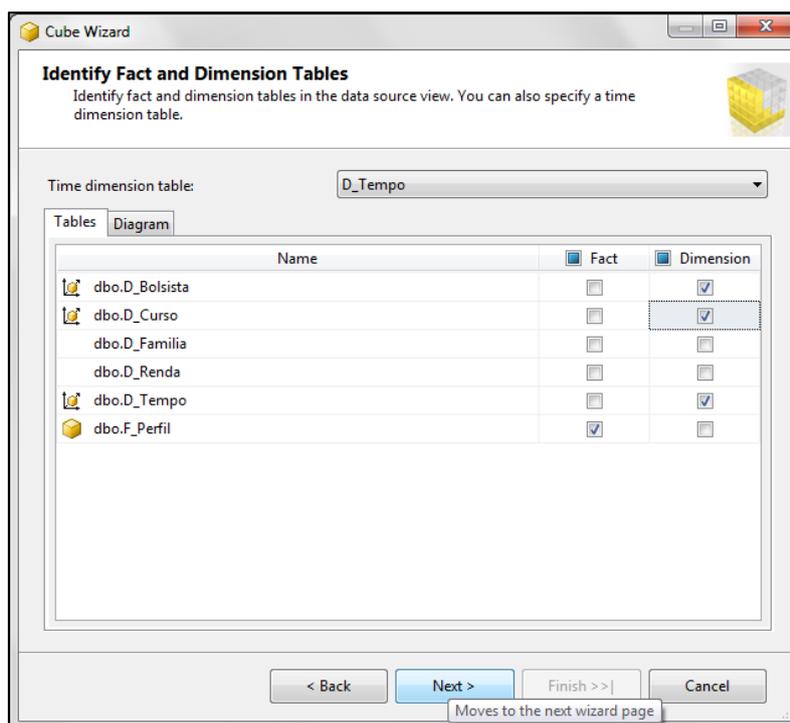


Figura 28: Identificação da tabela de fatos e dimensão.

Na Figura 28 são apresentadas as tabelas e as opções de seleção da tabela fato e as dimensões para processamento. Após a identificação da tabela de fatos é necessário selecionar as colunas que serão agregadas. Sugere-se neste passo a seleção de colunas que contenham dados que realmente irão servir para análise, ou que permitam verificar uma tendência ou outro desvio nos fatos pesquisados. Neste momento é possível definir também a tabela tempo e as operações sobre essa tabela para processamento dos dados. A seguir, na Figura 29 são apresentadas as colunas das tabelas selecionadas e, nesse momento, selecionam-se as colunas que contêm dados úteis a serem explorados e processados para extração de conhecimento.

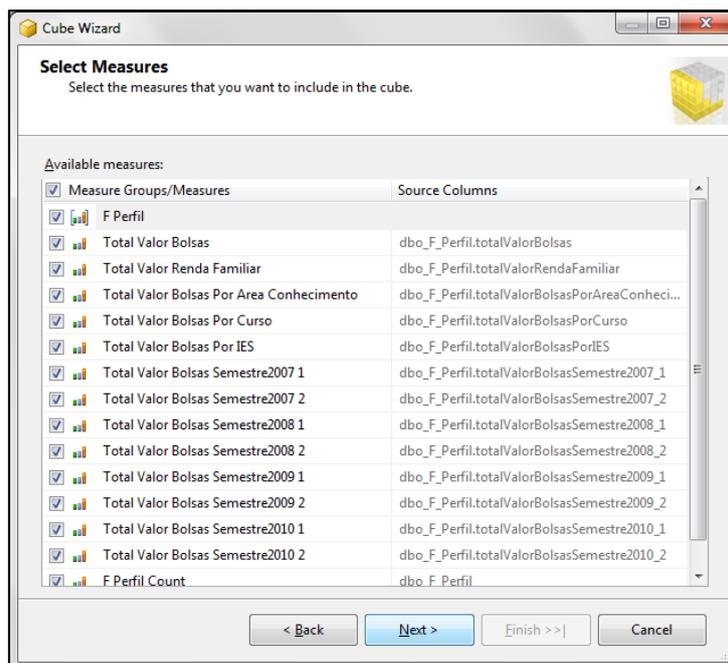


Figura 29: Definição das colunas que serão usadas como medidas.

Definidos os dados agregados definem-se quais eixos, ou subgrupos, que essas agregações serão efetuadas. Assim como no caso das agregações, não se deve escolher todas as colunas de todas as tabelas, como dimensões, pois quanto mais dimensões forem escolhidas, mais dados precisam ser gerados, conforme Figura 29. A seguir, na Figura 30 são apresentadas as colunas das dimensões disponibilizadas para efetuar o processamento.

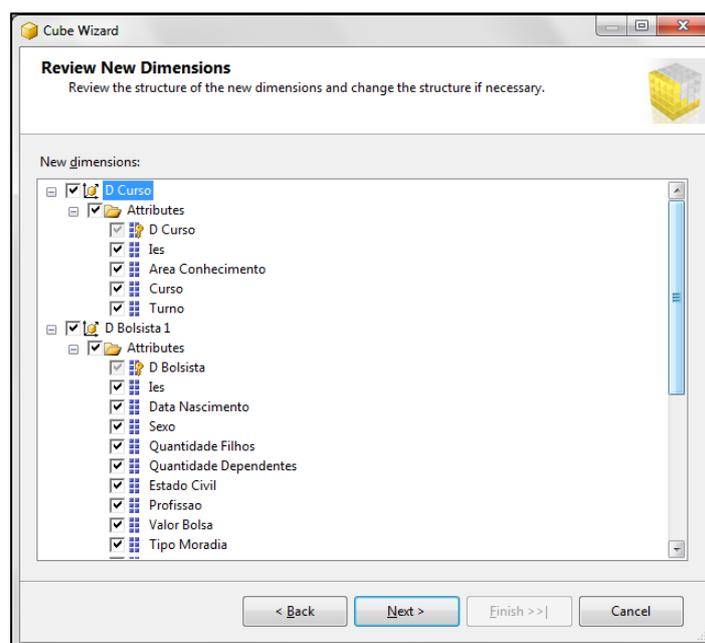


Figura 30: Definição das colunas que serão usadas como dimensão.

A Figura 30 apresenta as opções de seleção das colunas e ao término da seleção, ao avançar, finaliza-se esse processo e está completa a criação do cubo, sendo necessário nomear o objeto e observar no centro da janela que as tabelas que compõe o cubo serão apresentadas com seus respectivos relacionamentos, conforme Figura 31.

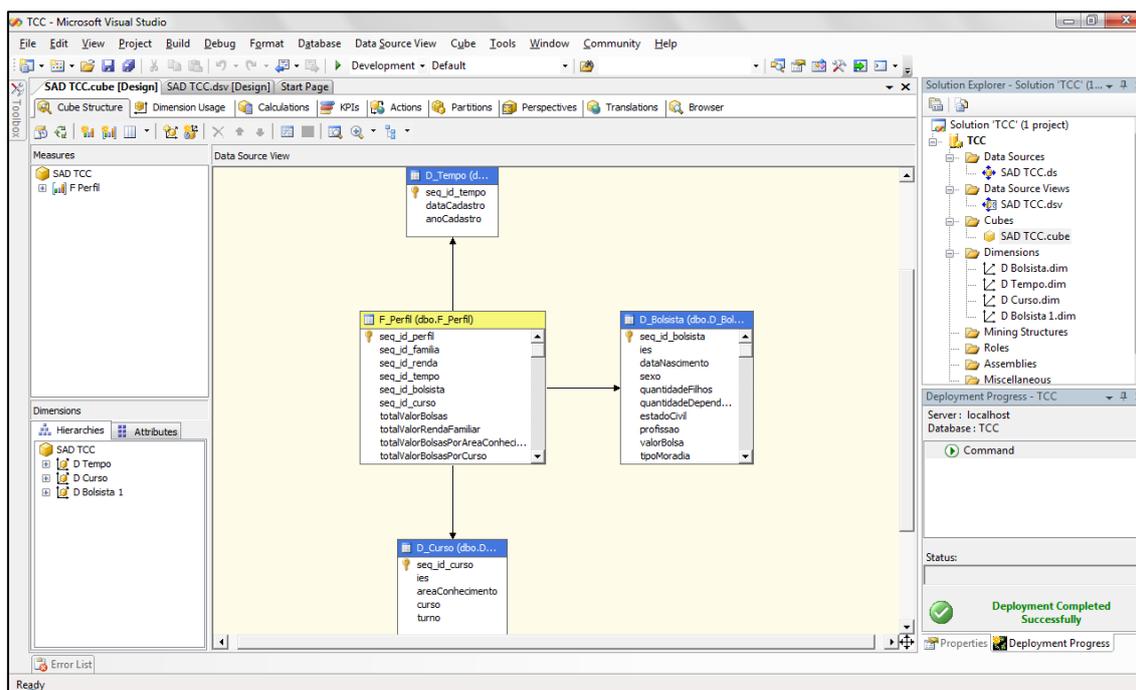


Figura 31: Construtor do cubo.

Na Figura 31 é apresentado o cubo bem como o relacionamento entre as tabelas sobre as quais o cubo foi configurado. Neste cubo foram utilizadas as tabelas fato (F_Perfil), D_Tempo (Dimensão Tempo), D_Bolsista (Dimensão Bolsista) e D_Curso (Dimensão Curso). Sobre esse cubo serão efetuadas operações sobre os dados e posteriormente serão apresentados e acessados, por meio do *Microsoft Excel*, para então realizar a emissão de relatórios gráficos. Com a conclusão da criação do cubo, percebe-se que sob a pasta *dimensions* na aba *Solution Explorer*, são criadas as dimensões com os nomes das tabelas utilizadas e configuradas como dimensão na criação do cubo. A seguir, na Figura 32 é apresentado o processo de criação de uma operação *Business Intelligence*, sobre os dados do cubo criado.

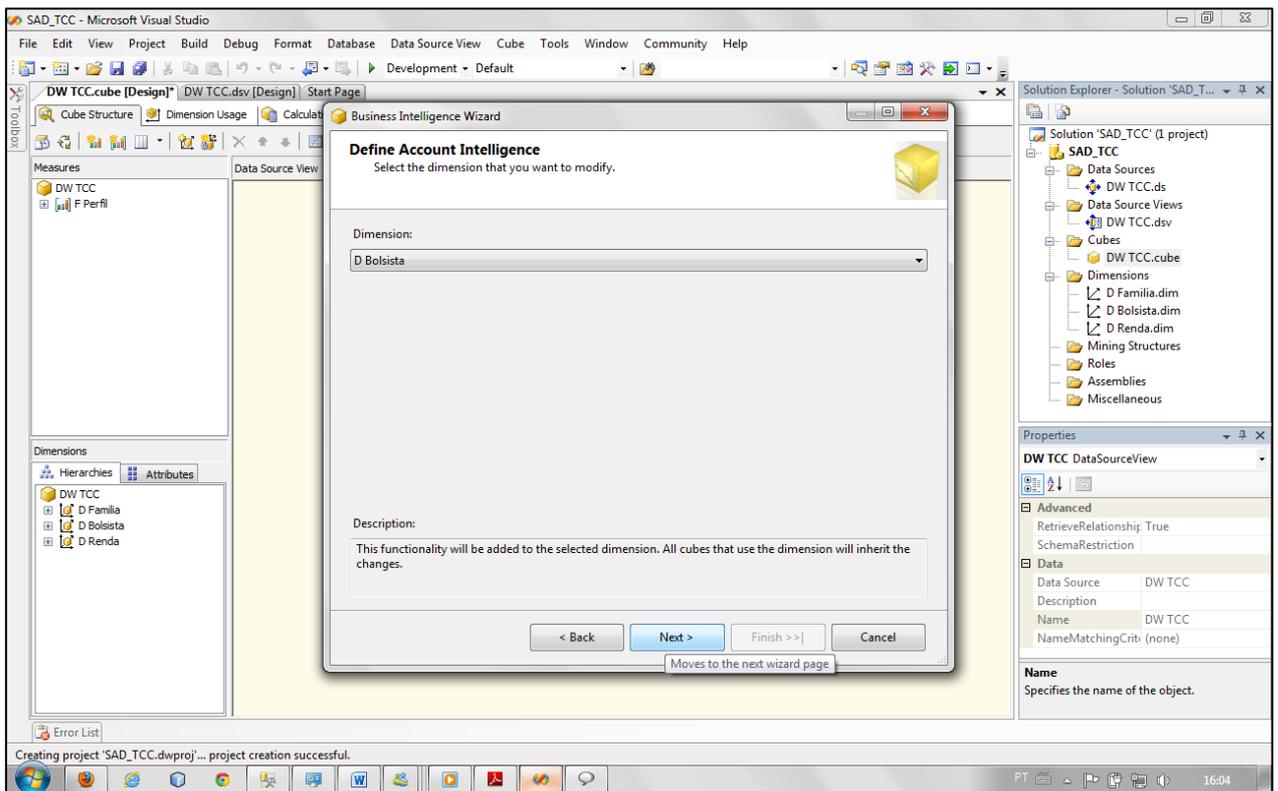


Figura 32: Operação *Business Intelligence*.

Na Figura 32 é apresentado o processo de criação de uma operação *Business Intelligence*. Essas operações exploram os dados das tabelas. Tendo sido selecionada a tabela, na sequência da configuração da operação devem ser selecionadas as colunas da tabela dimensão sobre as quais será efetuado o processamento, conforme apresentado na Figura 33.

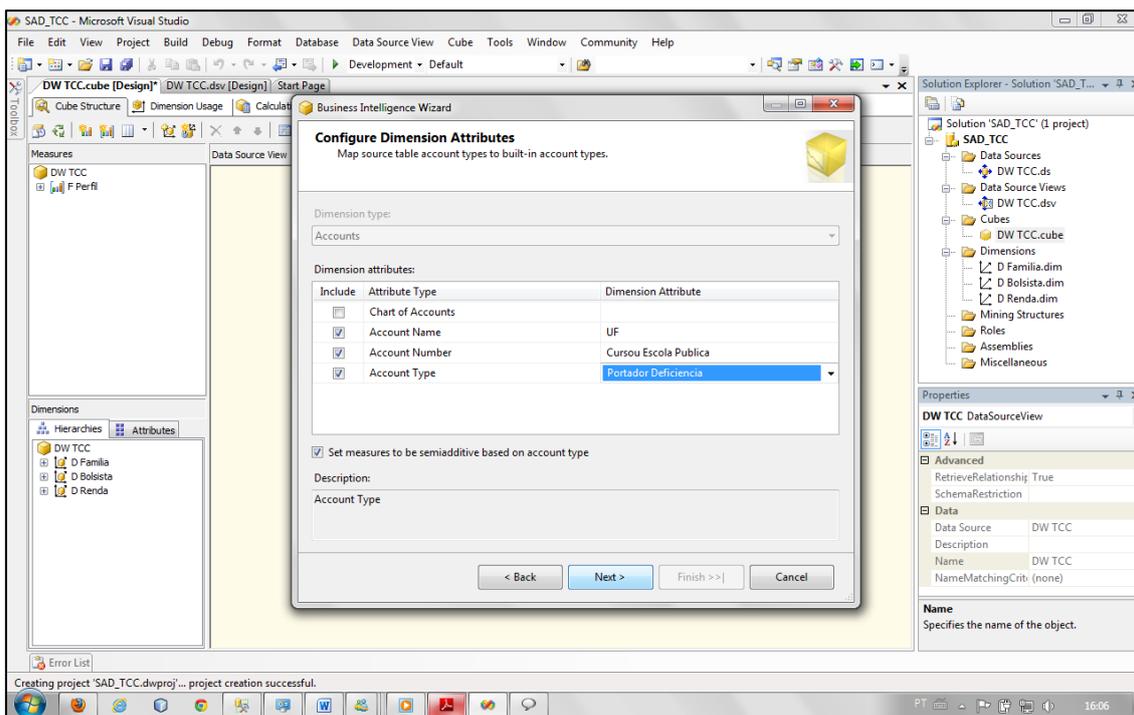


Figura 33: Configurando atributos da dimensão BI.

Na Figura 33 são apresentadas as colunas da tabela e as operações de processamento sobre os dados da coluna. Também é oferecida a opção de realizar alterações sobre os dados na tabela fato. Ao avançar no processo, são apresentadas a lista de operações a serem efetuadas sobre os dados selecionados, conforme apresentada na Figura 34.

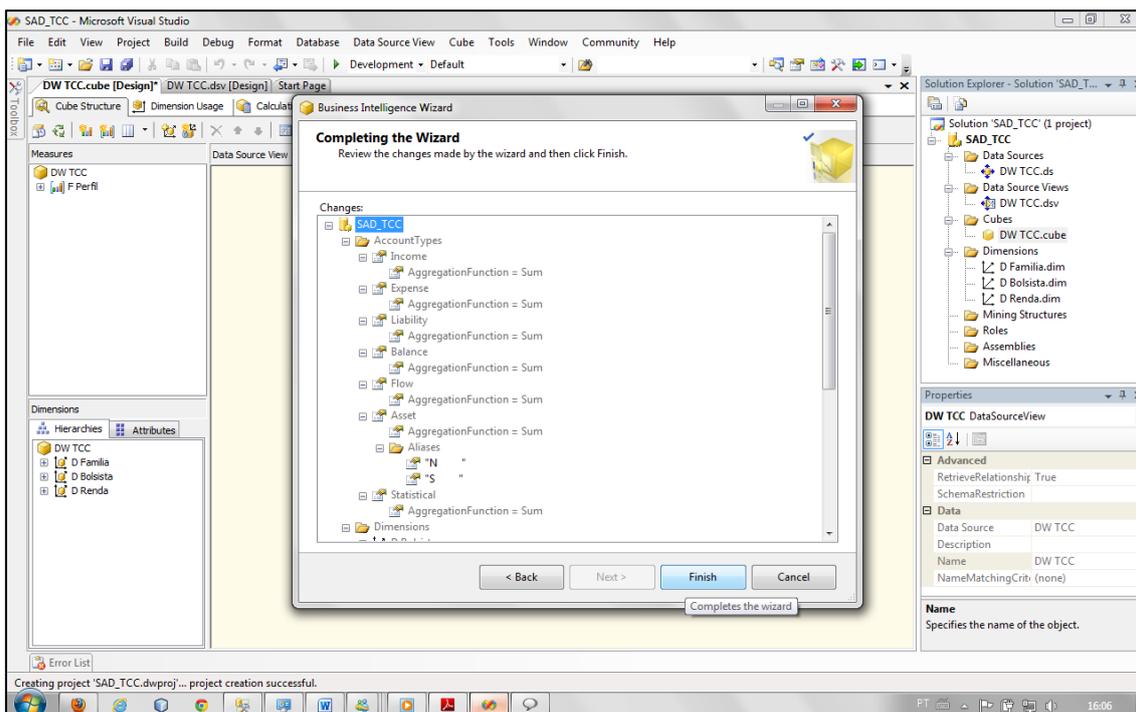


Figura 34: Criação de operações BI.

Nas Figura 34, apresenta a finalização do processo de criação de operações BI sobre o cubo. Assim sendo, o cubo está pronto a ter seus dados processados e esse processo é apresentado na seção 4.4.2.

4.4.2 Processando e Visualizando Cubos

Para processar o cubo e verificar os resultados clica-se em *Process* (o segundo na barra de ferramentas do cubo). Caso o projeto ainda não esteja salvo, o *Microsoft Visual Studio* irá gerar um aviso e salvará o projeto para o processo. A seguir, a Figura 35 apresenta o processamento do cubo.

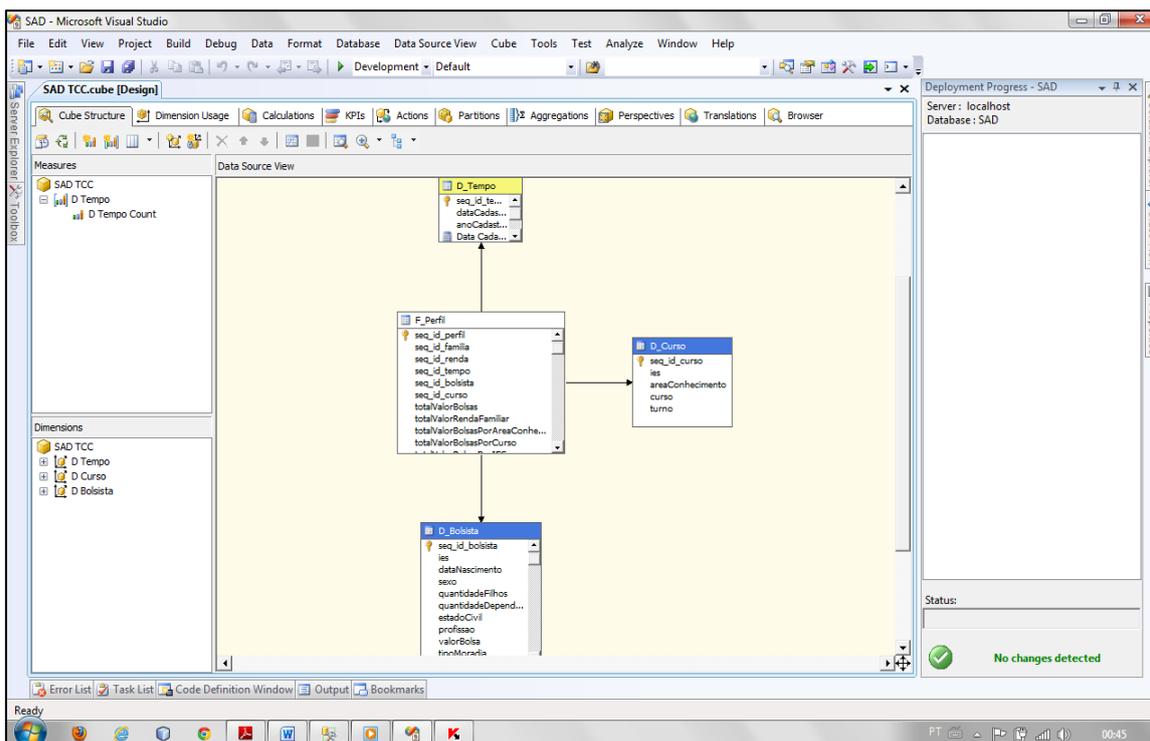


Figura 36: Visualização do cubo.

Iniciado o processamento do cubo, o primeiro dado informado é o tipo de processamento. Os processamentos podem ser *Full* (completo) ou *Incremental* (apenas alterações). Normalmente utiliza-se o modo *full* para processar todos os objetos criados no projeto *Analysis Services*. Caso haja alterações sobre as tabelas dimensões ou fatos, o processamento deve ser realizado acessando projeto por meio da função *Analysis Services* no Microsoft SQL Server. A seguir, na Figura 37 é apresentado o processamento dos objetos criados sobre o cubo de consulta.

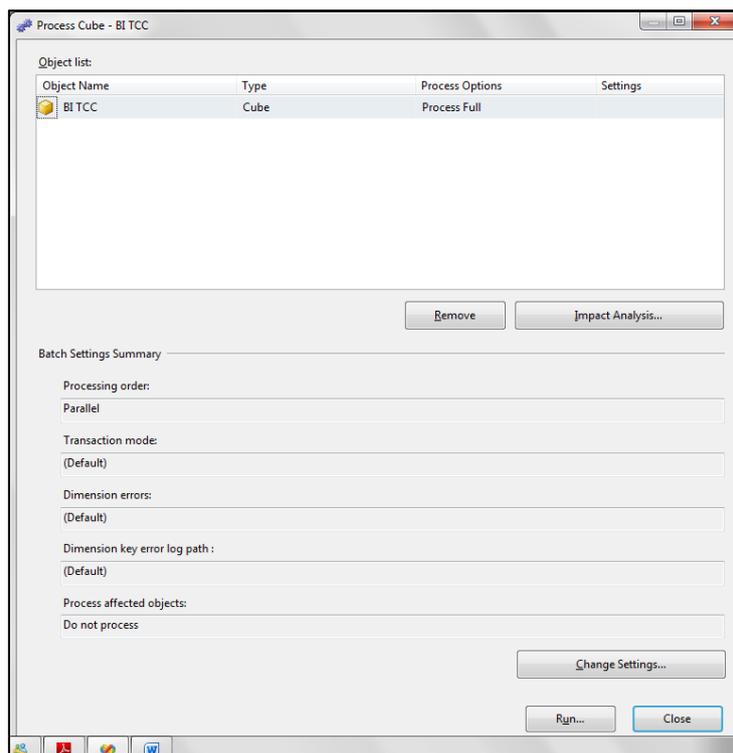


Figura 37: Acesso as configurações de processamento.

Escolhendo o modo de processamento pode-se executar o cubo (*Run*). O *Visual Studio* emite uma tela de aviso sobre o andamento do processamento, conforme apresentado a seguir, na Figura 38. Não havendo erros no processamento, ele é apresentado, juntamente os nomes das dimensões processadas e a *query* (salva) executada.

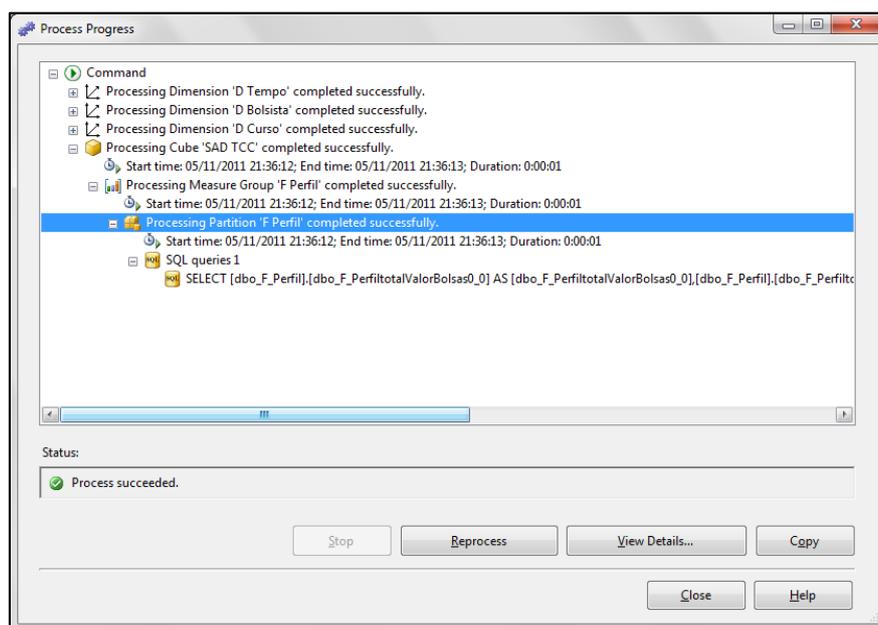


Figura 38: Processando o projeto.

Com o cubo já processado é possível visualizar seus dados e assim ter uma ideia de como os dados estarão disponíveis para análise. Para isso, na aba *Browser* do *Visual Studio*, do lado direito, apresenta uma árvore com as medidas e as dimensões separadas por tabela. Para visualizar os dados, basta arrastar os dados para o centro da tela, lembrando que nas colunas e linhas são inseridas as dimensões e no centro as medidas, conforme pode ser observado na Figura 39.

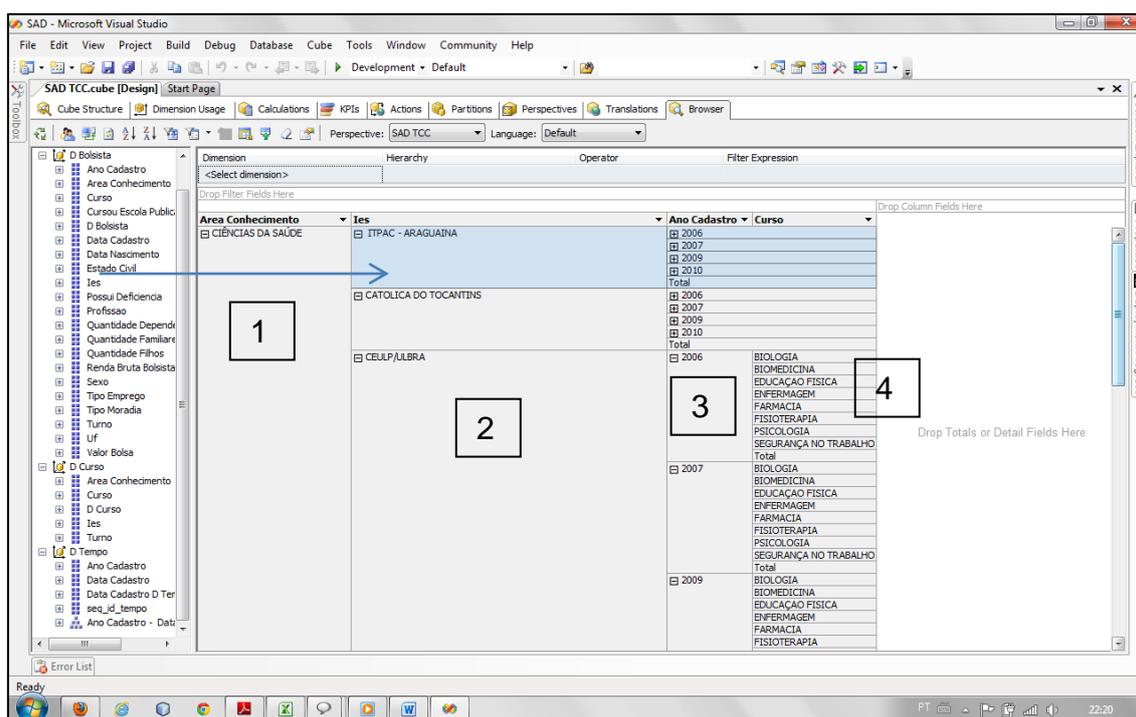


Figura 39: Browse do cubo com eixos e colunas e medidas.

Na Figura 39 é apresentada a organização dos dados inseridos no cubo criado. Assim sendo, pode-se perceber a seguinte organização na disponibilização dos dados: na coluna 1 a classificação de área de conhecimento, na coluna 2 são apresentadas as IES. Na coluna 3 o ano do cadastro e, por fim, na coluna 4, os cursos da IES com bolsistas do programa. Nota-se que o visualizador (*pivot table*) gera sinais (+ e -) para que possam ser verificados os detalhes do curso, ou seja, clicando nestes sinais podemos ver apenas o curso e o ano selecionado. Essa visualização também pode ser realizada criando uma perspectiva do cubo, selecionando por meio das abas superiores do cubo na opção *perspective*. Na seção 4.4.3, serão apresentados os passos para criação das medidas para processamento dos dados.

4.4.3 Criando *Measure* (medidas) de processamento no cubo

No projeto de *Analysis Services*, medidas são os dados que já existem nos cubos e que podem sofrer cálculos ou serem processados. Neste caso, as tabelas que possuem valores precisam ser mensuradas, sendo que por meio desse recurso são resolvidos, processados e inseridos na base OLAP. Assim sendo, a partir da Figura 40, será apresentado o processo da criação de medidas.

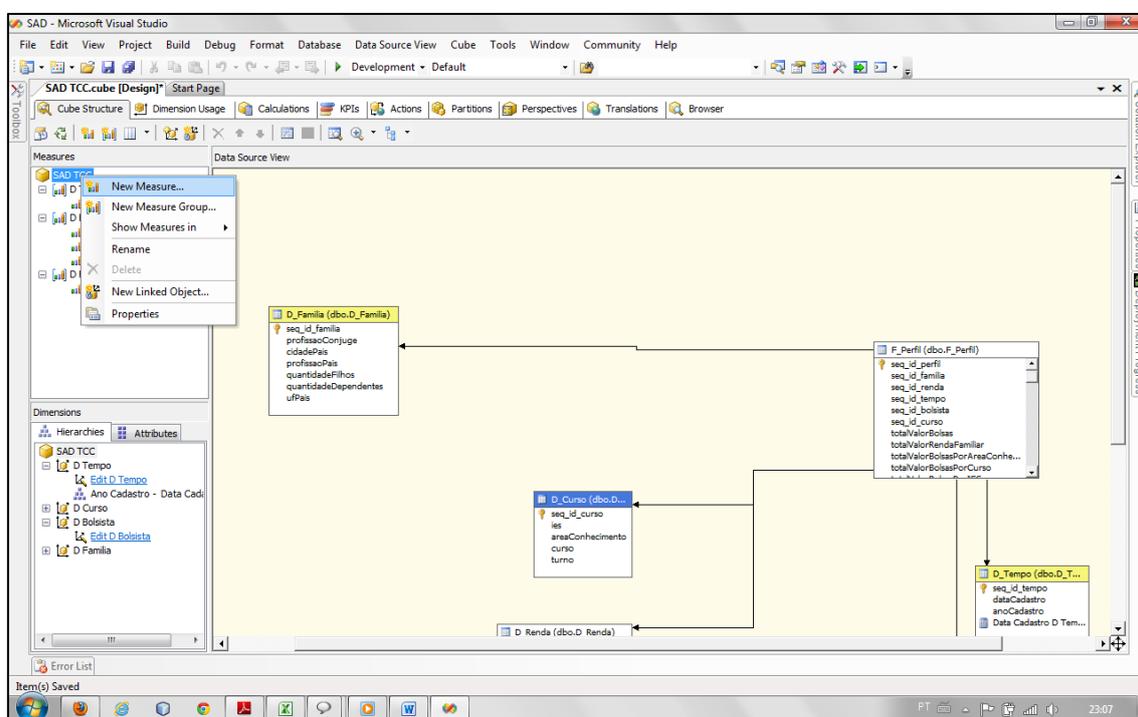


Figura 40: Criação de medidas.

Conforme apresentado na Figura 40, para a criação de uma medida, é necessário dar um clique duplo sobre o cubo, para acessar as propriedades das tabelas do mesmo. Então, na aba *Measure*, solicitar a criação de nova *measure*. Ao selecionar a opção, serão apresentadas as opções conforme a Figura 41.

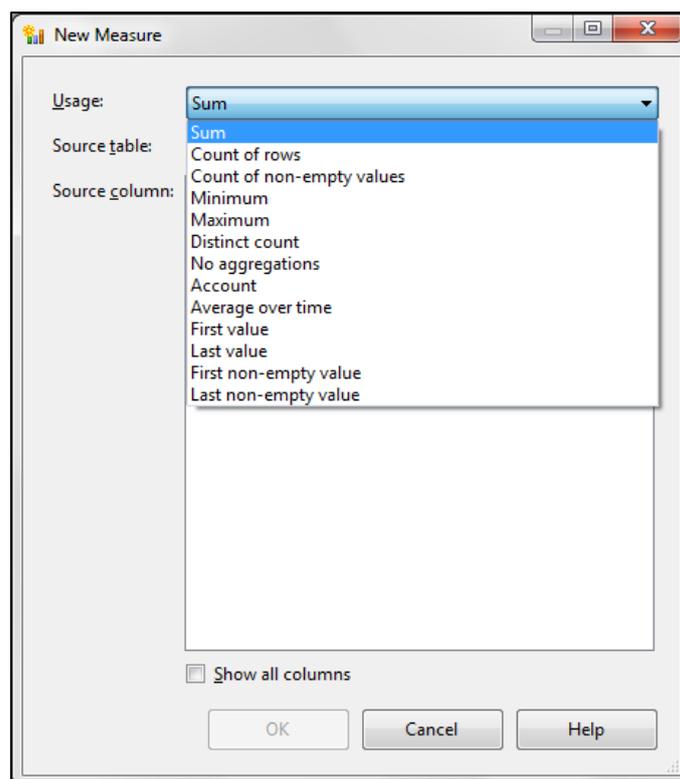


Figura 41: Seleção da operação da medida.

Como se pode observar, na Figura 41 é possível selecionar o tipo de cálculo (tipo de medida) para o processamento das tabelas dimensões. São apresentadas as opções de soma, contagem de linhas, contagem de valores em branco, valores mínimos e máximos, etc. Essas operações são realizadas sobre as colunas da base de dados do DW. Tendo sido selecionada a operação, é necessário selecionar a tabela que contém os dados que deseja-se processar e, por fim, a tabela desejada. Finaliza-se o processo clicando em “ok” e ao finalizar é possível verificar a medida na aba *Measure*. Ao finalizar a criação das medidas, é necessário salvar o projeto e solicitar um novo processamento para que essas operações sejam carregadas e criadas as *queries* (consultas) referentes às mesmas para posterior processamento e acesso.

Para esse projeto BI, foram criadas as medidas sobre as dimensões D_Bolsista, D_Curso, D_Tempo, D_Familia e sobre a tabela Fatos F_Perfil. Medidas são cálculos ou processos executados sobre os valores existentes nos cubos. As medidas criadas sobre as tabelas dimensões são processadas e apresentadas suas respectivas *queries*, conforme a seguir, na Figura 42.

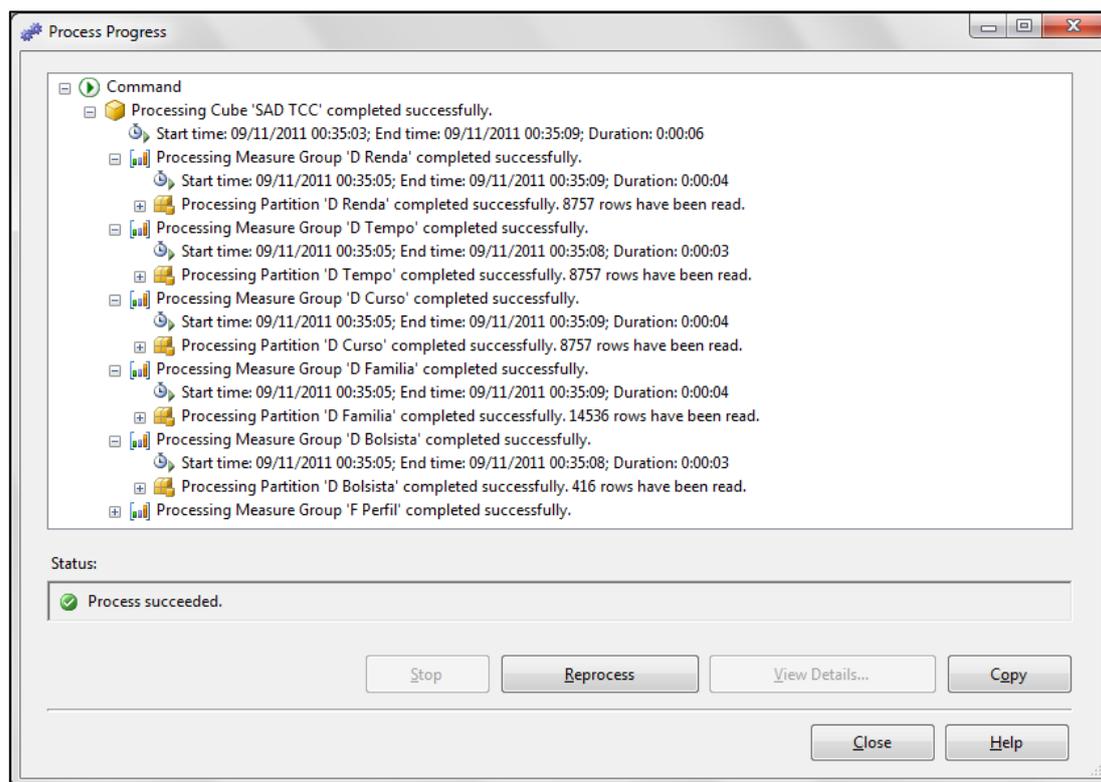


Figura 42: Processamento das medidas.

A criação das medidas varia conforme as necessidades métricas do projeto BI. Assim sendo, para este projeto foram criadas medidas do tipo *Sum*, *Count of rows*, *distinct count* e *No Aggregations*, para respectivamente, somar os valores de registros de determinada coluna (valor total de bolsas agregado), contagem de linhas, que automaticamente, processa o total de registros (total de bolsistas), contagem de valores distintos com alguma semelhança na significância para o sistema (servidores públicos municipais, estaduais ou federais) A partir das medidas criadas sobre todas as dimensões do projeto, foram combinados os processamentos entre elas para iniciar a emissão de relatórios. A seguir, na Figura 43 são apresentadas partes das medidas criadas e as variáveis utilizadas sobre as mesmas para processamento.

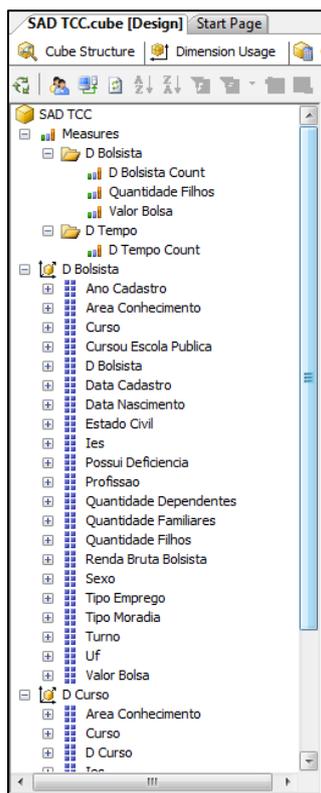


Figura 43: Medidas criadas sobre as dimensões

Conforme apresentado na Figura 43, foram criadas as medidas “D Bolsista”, “D Tempo” e ‘D Curso”. Essas medidas foram criadas para realizar os cálculos e processamento das informações contidas nas respectivas tabelas. Assim sendo, podem-se observar na figura que sob a medida criada das tabelas dimensões, são listados abaixo os atributos das medidas, que consequentemente, são as colunas da tabela dimensão.

Ainda sobre as medidas, é necessário relacionar as medidas sobre as quais se deseja analisar e realizar o cruzamento dos dados. Essa operação é realizada na aba *Dimension Usage*. Nela serão apresentadas as dimensões do projeto e as relações entre as mesmas para obtenção de informações a partir do projeto BI, conforme Figura 44.

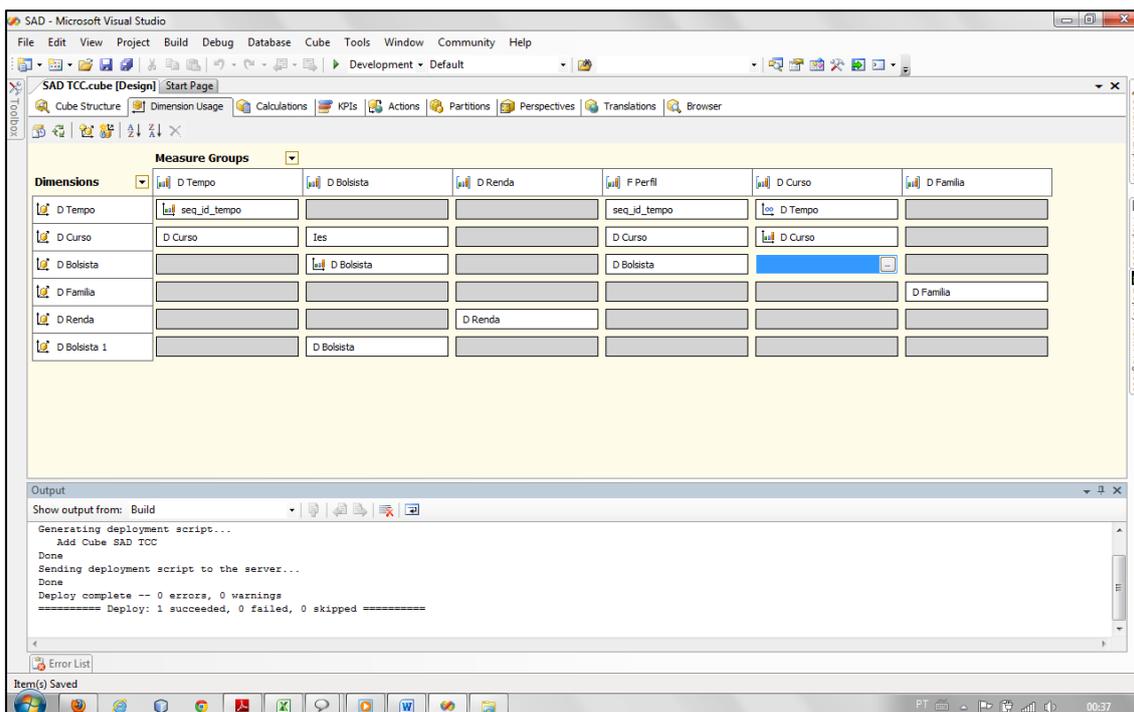


Figura 44: Dimension Usage.

Na Figura 44 são apresentados os usos das dimensões, ou seja, as formas com as quais se deseja visualizar as dimensões, quais relações se deseja realizar com esta dimensão e ainda mais, quais os campos dessa dimensão se deseja explorar. Nessa fase do projeto, é necessário relacionar os atributos do cubo, para que quando processadas as medidas, a partir do cruzamento das informações, seja otimizada a emissão dos relatórios. Assim sendo, nota-se que são apresentadas as tabelas dimensões do projeto nas guias horizontais e verticais. E nas colunas são disponibilizados os atributos das respectivas colunas para relacionamento. Com isso, relaciona-se o atributo da coluna horizontal com a coluna vertical, por exemplo: na segunda coluna, referente à dimensão curso, o atributo “!es”, é relacionado com a dimensão bolsista. A partir desse relacionamento, será realizado um mapeamento entre as tabelas e os respectivos atributos para a emissão de relatórios. A seguir, na seção 4.4.4 serão apresentadas as partições de armazenamento OLAP criadas para esse projeto.

4.4.4 Partições

Para o de armazenamento de OLAP podem ser utilizados: MOLAP (total), HOLAP (parcial) ou ROLAP (leitura *online*). A seguir, na Figura 45 é apresentado o processo de definição do tipo de armazenamento OLAP.

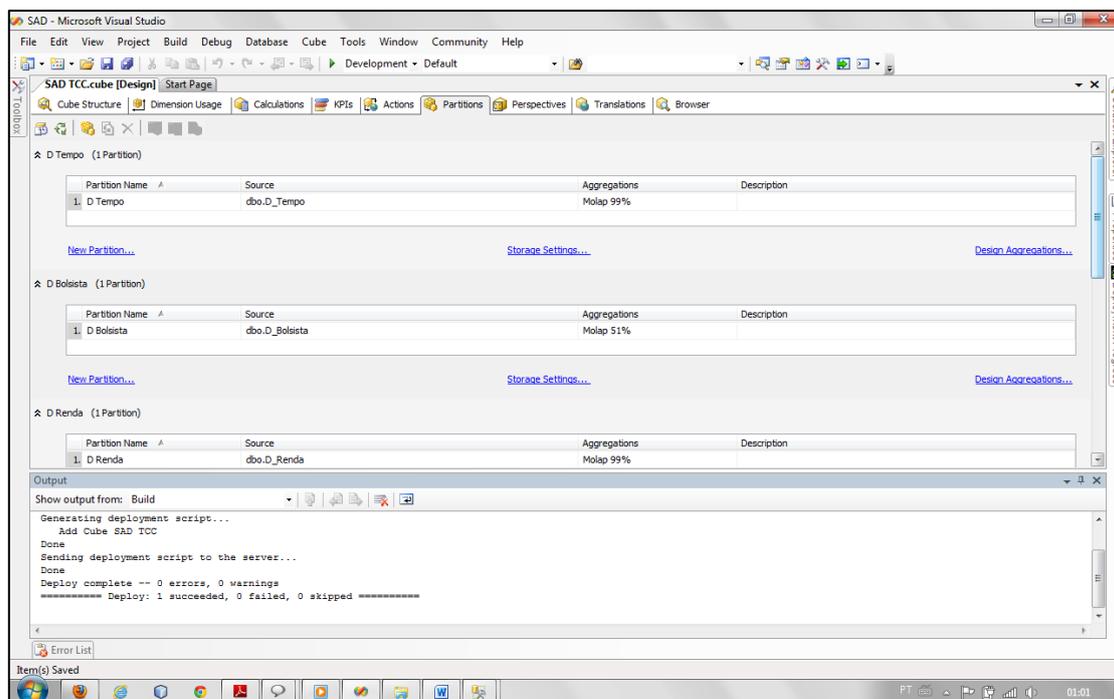


Figura 45: Definição do tipo de armazenamento.

Na Figura 45 são apresentadas as partições criadas a partir das dimensões do projeto. Assim sendo, para a definição da forma de armazenamento para a dimensão é necessário clicar sobre o *link storage settings*.

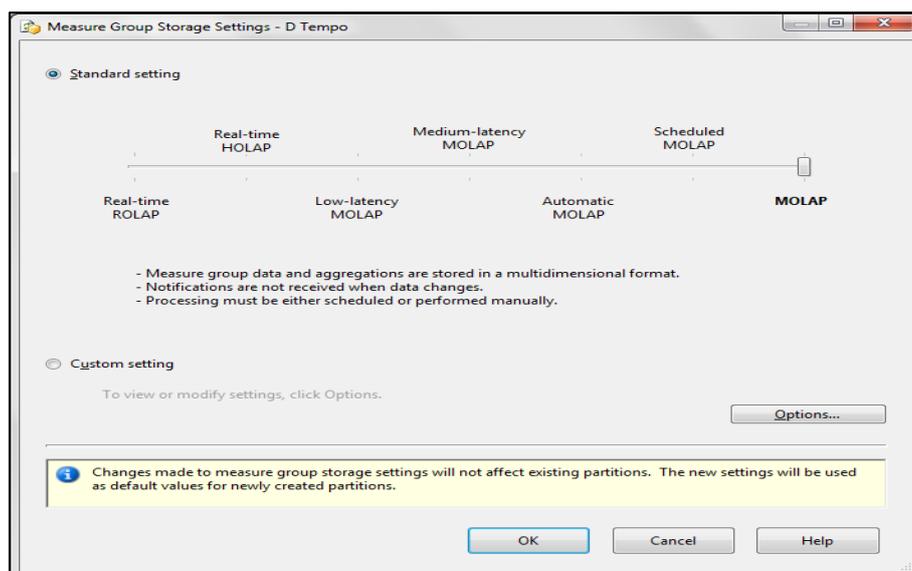


Figura 46: Seleção do tipo de armazenamento.

Na Figura 46 são apresentadas as opções de armazenamento disponíveis para o projeto que, neste caso, foi selecionada a opção MOLAP, para realizar o armazenamento total das informações.

Por fim, para finalizar o projeto, é necessário acessar o projeto para o meio do SQL Server nas funções de *Analysis Services* e processar o projeto para que todas as operações realizadas sejam salvas e carregadas para a base de dados.

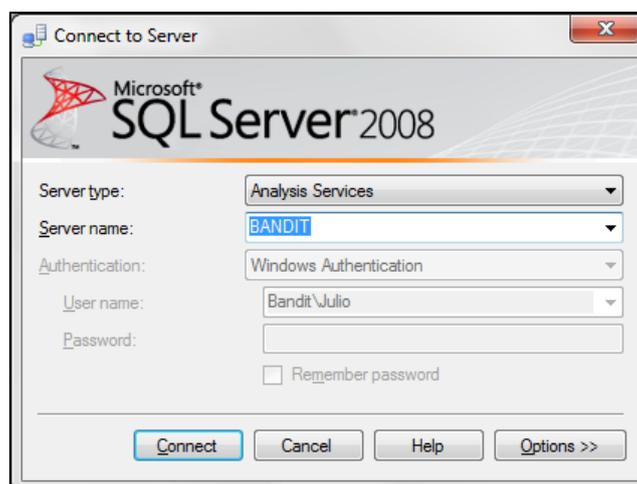


Figura 47: Acessando projeto por SQL Server

Na Figura 47 é apresentada a tela de *login* do *Microsoft SQL Server 2008*. Para realizar o acesso do *Analysis Services*, é necessário selecionar em *Server Type*: *Analysis Services*, depois, o nome do servidor e solicitar a

conexão. Após a conexão, é necessário selecionar a base de dados do projeto e clicar em *Process*.

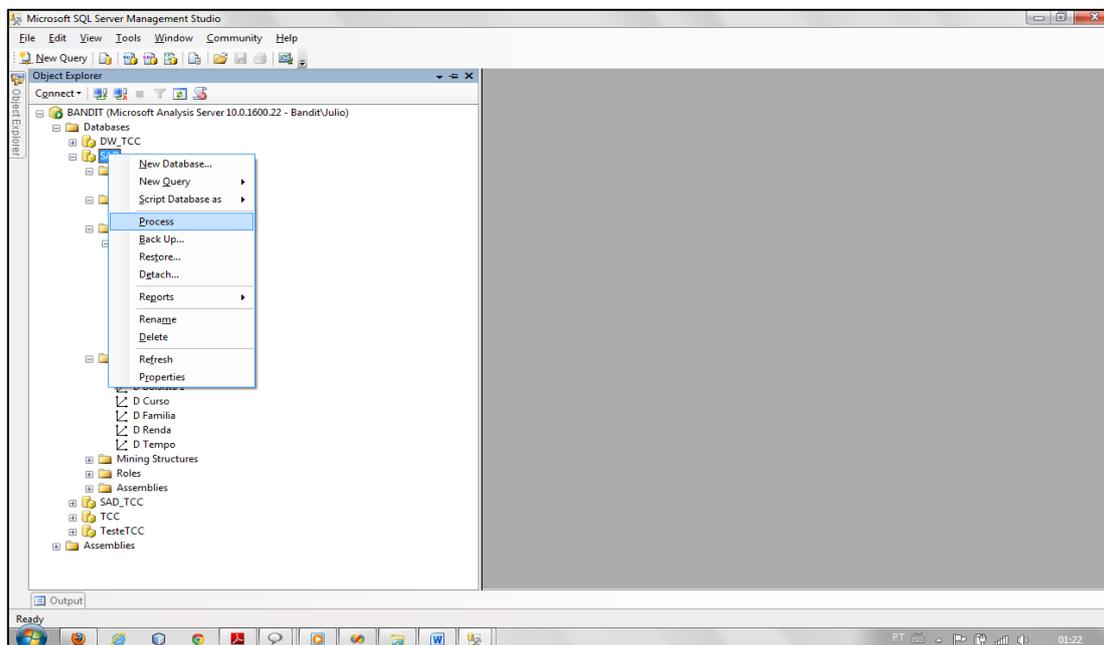


Figura 48: Selecionando projeto para processamento.

Na Figura 48 é apresentada a opção de processamento do projeto *Analysis Services* por meio do *SQL Server*. Essa operação necessariamente deve ocorrer por meio do *SQL Server*, pois neste processamento serão carregadas todas as ações e alterações realizadas sobre a base de dados do projeto. Diferente dos processamentos realizados por meio do *Microsoft SQL Server Business Intelligence*, que processa as últimas alterações e/ou configurações realizadas sobre os objetos do projeto. Assim sendo, na Figura 49 são apresentadas as opções de processamento integral do projeto AS.

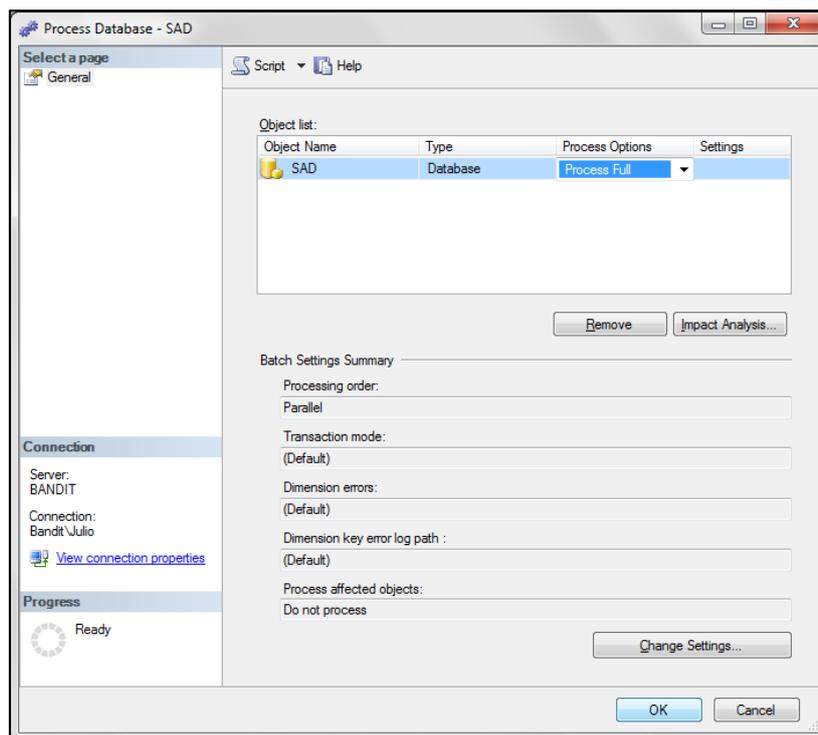


Figura 49: Processando projeto por completo.

Na Figura 49 é apresentada a opção de processamento completo do projeto AS. Na lista de objetos do projeto a opção de processamento, onde constam as seguintes opções: *Default Process*, *Process Full* e *Unprocess*. Na opção *Default* o processamento é sobre as últimas alterações realizadas e somente sobre o objeto manipulado, ou seja, caso seja atualizada a base de dados do projeto com a inserção de novos registros e simultaneamente seja criado um novo cubo, o processamento ocorrerá somente sobre o cubo e, conseqüentemente, serão apresentados erros de processamento, sendo necessário um processamento completo. A opção *Default* é a mesma realizada por meio do *SQL Server Business Intelligence Development Studio*. A opção *Full* é o processamento completo, ou seja, todo o projeto será processado, desde os mapeamentos do(s) cubo(s), relacionamentos, registros, medidas, *scripts*, ou seja, todo o projeto. Por fim, a opção *Unprocess* (não processados), verifica as alterações do projeto, ele realiza a comparação da versão anterior do projeto e realiza o processamento sobre as alterações. Essa opção é válida para situações de tempo real e pode ser exemplificada pela seguinte situação: caso esse projeto seja um projeto de consulta online e sejam realizadas alterações sobre o projeto, para que o projeto não fique indisponível por algum tempo, essa opção levanta as alterações e conforme as configurações do

Server, o projeto é reconectado automaticamente e as alterações carregadas, sem haver necessidade de parar o serviço.

Para o processamento do projeto, é necessário selecionar o tipo de processamento e clicar em *ok*, e ao término do processamento, é apresentado o *status* e os objetos do projeto processados, conforme apresentado na Figura 49. A seguir, na Figura 50 apresentado o status do processamento.

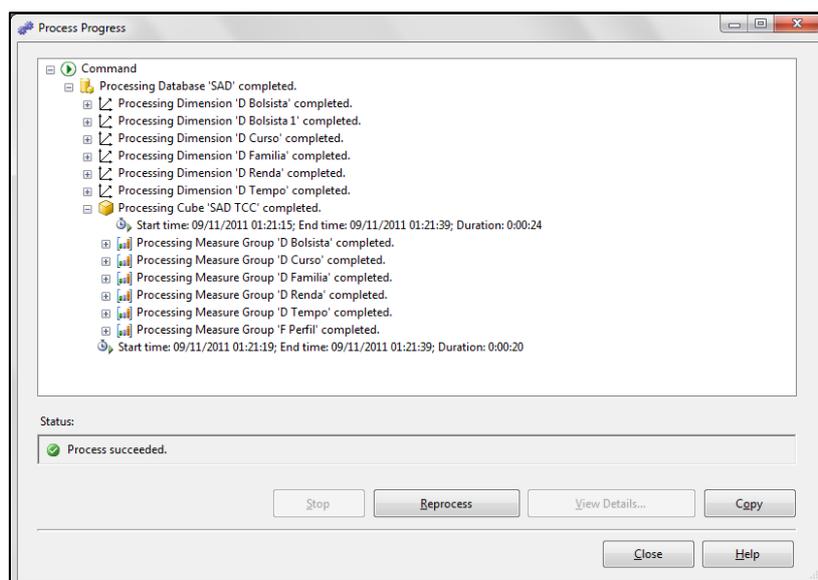


Figura 50: Status ao final do processamento do projeto.

Na Figura 50 é apresentado o resultado (*status*) do processamento do projeto. Assim sendo, é possível observar primeiramente a base de dados processada e sob a base as dimensões carregadas e o *status* do processamento, neste caso: *completed*. Posteriormente, ainda obedecendo a hierarquia de objetos da base de dados, é apresentado o cubo criado para o projeto e sobre o mesmo, as medidas de processamento criados para processamento e transformações sobre os dados. Por fim, o *status* geral do projeto, com a opção *Process Succeeded*, informando que não houve erros sobre qualquer objeto do projeto, podendo realizar as consultas e emissão de relatórios. A seguir, na seção 4.4.5 é apresentado o processo de emissão de relatórios a partir de OLAP.

4.4.5 Carregando as informações e gerando gráficos

A visualização dos dados e emissão de relatórios OLAP pode ser realizada de diversas formas. Podem ser emitidos por meio da ferramenta *Cristal Reports*, *Microsoft Excel* ou ainda utilizando alguma ferramenta WOLAP, que disponibilizará acesso por meio do *browser (web)*. Para demonstrar os resultados deste trabalho, será utilizada a ferramenta *Microsoft Excel 2010*. Essa ferramenta possibilita a emissão de relatórios e para facilitar a análise dos mesmos, possui um gerador de gráficos dinâmicos que são gerados sobre os dados da base OLAP. A partir da Figura 51: Tela do Microsoft Excel será apresentado o processo de conexão com a base de dados OLAP, bem como a disposição dos dados para emissão dos relatórios dinâmicos e, ainda, serão apresentados alguns gráficos gerados a partir da base OLAP para comprovação e validação deste trabalho.

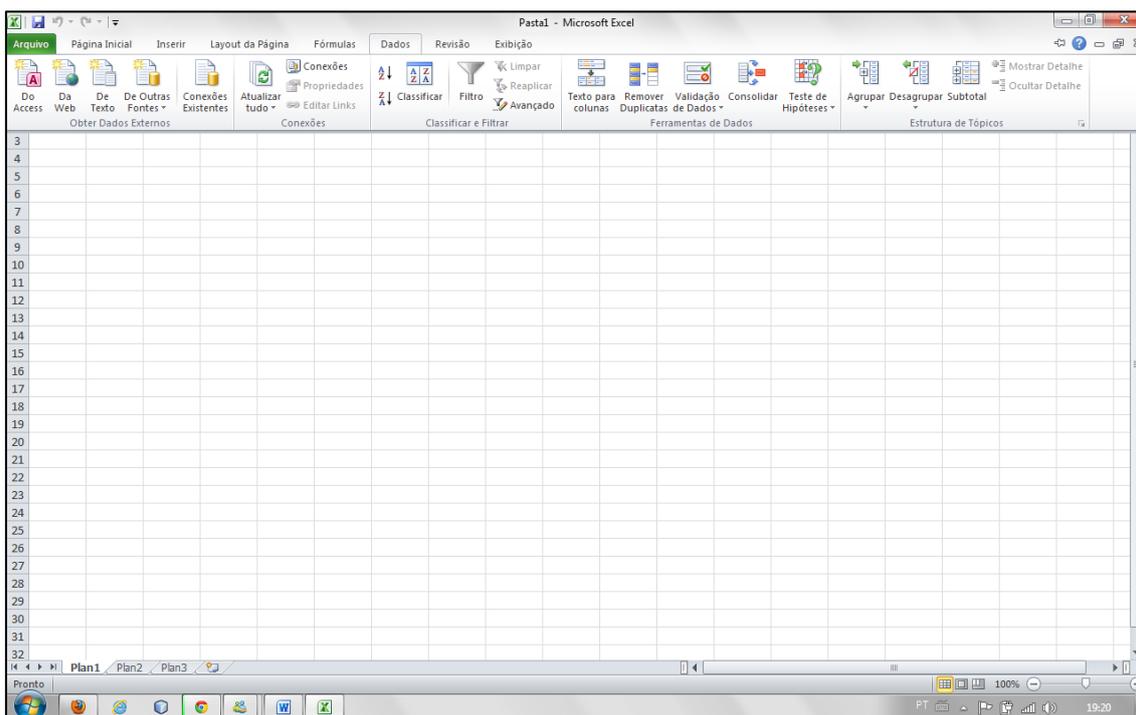


Figura 51: Tela do Microsoft Excel.

Na Figura 51 é apresentada uma planilha em branco, criada para suportar uma conexão com base de dados. Para conexão com uma base de dados do projeto *Analysis Services* por meio do Microsoft Excel, é necessário a

partir de uma nova planilha, selecionar a aba Dados. Conforme apresentado na Figura 52.

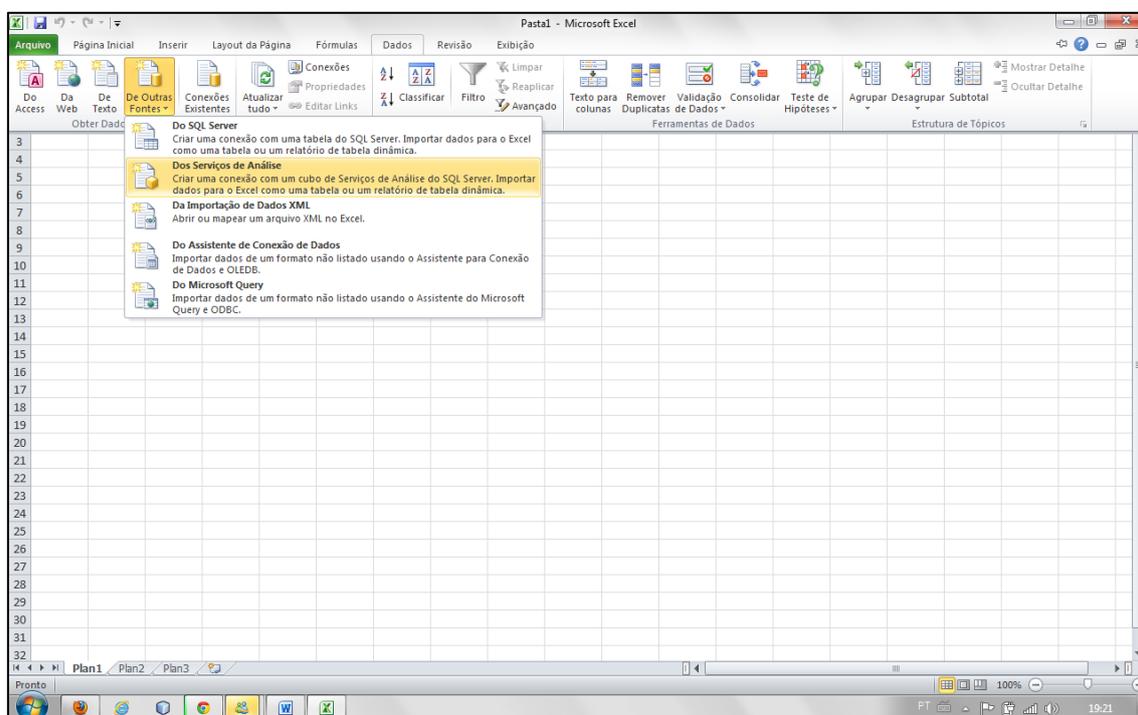


Figura 52: Seleção do tipo de base de dados para conexão.

A partir da seleção da aba dados, são disponibilizadas as opções dos tipos de conexões. Para esse processo, é necessário selecionar conexão com outras fontes, conforme apresentado na Figura 52. Selecionado o tipo de conexão, é necessário configurar a conexão, conforme apresentado na Figura 53.

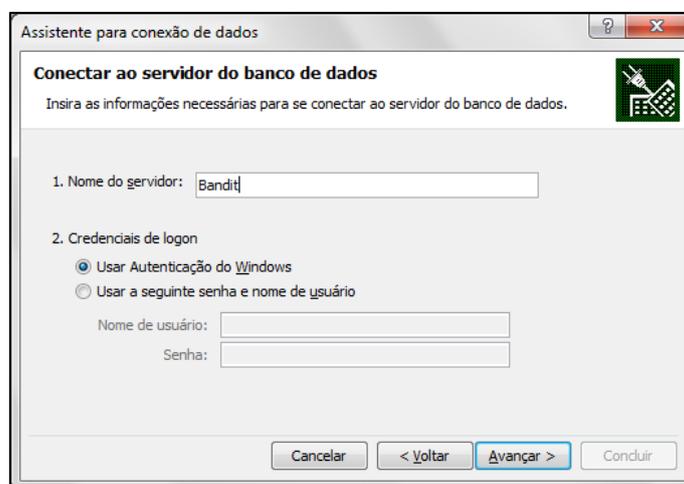


Figura 53: Configurando conexão com servidor.

A Figura 53 apresenta as opções para preenchimento dos dados do servidor de base de dados para conexão, bem como usuário e senha para autenticação. As informações utilizadas para essa configuração (login e senha) devem ser as mesmas utilizadas para conexão no *SQL Server Analysis Service*, tendo em vista a instância da base e as permissões do usuário. A seguir, na Figura 54 são apresentadas as opções de conexão.

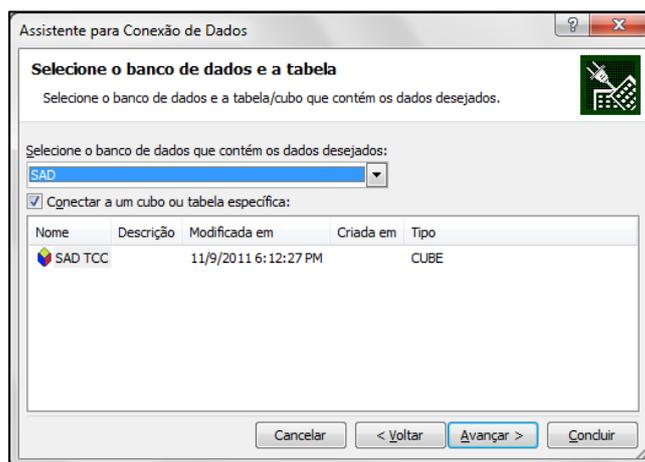


Figura 54: Seleção da base de dados e do objeto de conexão.

Na Figura 54 são apresentadas as opções de base de dados *Analysis Service* e, logo abaixo, a opção de conexão no cubo (objeto) de consulta do projeto. Devidamente selecionadas as base de dados e o cubo de consulta, é necessário avançar e selecionar a melhor opção de apresentação conforme apresentado na Figura 55.

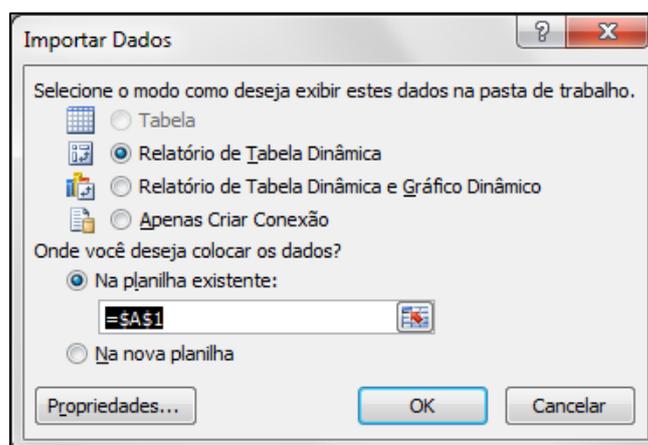


Figura 55: Seleção de como importar dados e em quais células começar.

Na Figura 55 são apresentadas as opções do modo como se deseja exibir os dados na planilha. As opções disponibilizadas são: criar relatório de tabela dinâmica, que disponibilizará as colunas das tabelas configuradas no cubo para consulta; relatório de tabela dinâmica e gráfico dinâmico, que se diferencia da primeira por já disponibilizar a estrutura do gráfico no ato da importação da estrutura e automaticamente, conforme as seleções das tabelas, os gráficos são montados em tempo real; e, por fim, apenas criar a conexão para possibilitar a busca de dados em outro momento. Pode-se verificar também que é possível selecionar a partir de qual célula se deseja disponibilizar os dados, e por fim, a possibilidade de criar em uma nova planilha. Clicando em *ok*, a estrutura é disponibilizada em uma área de planilha eletrônica conforme apresentado na Figura 56.

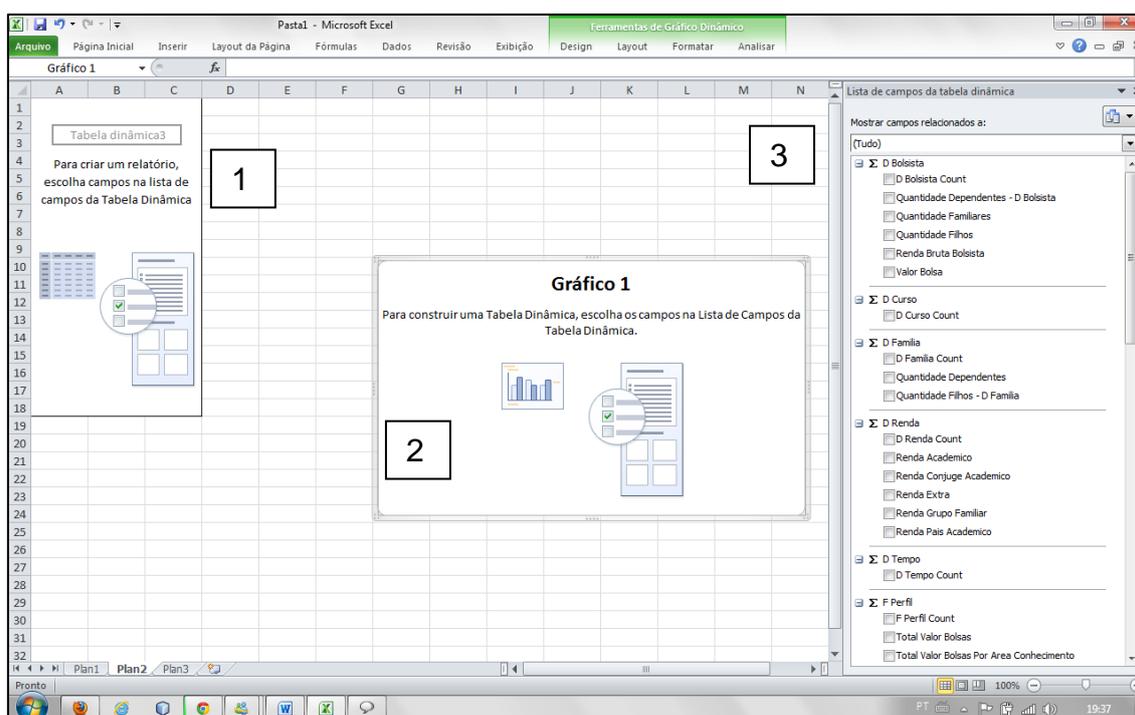


Figura 56: Apresentação de área dinâmica para emissão de relatórios.

Na Figura 56 é apresentada a estrutura disponibilizada na planilha para emissão dos relatórios com consultas OLAP, a partir da definição de criação de relatório de tabela dinâmica com gráfico dinâmico. Conforme a Figura 56, na área definida com o número 1, são apresentados os valores da tabela, ou seja, os cálculos gerados sobre os valores das colunas configurados nas medidas.

Na área 2 é disponibilizado o gráfico criado em tempo real e na área 3 os campos de seleção disponibilizados a partir do cubo de decisão para emissão dos relatórios e montagem dos gráficos. A seguir, na Figura 57 é apresentada uma visão geral das áreas preenchidas e com os relatórios emitidos a partir da consulta no cubo de decisão e, em seguida, serão apresentadas as partes desta planilha para melhor visualização dos detalhes da busca sobre o cubo para emissão dos relatórios.

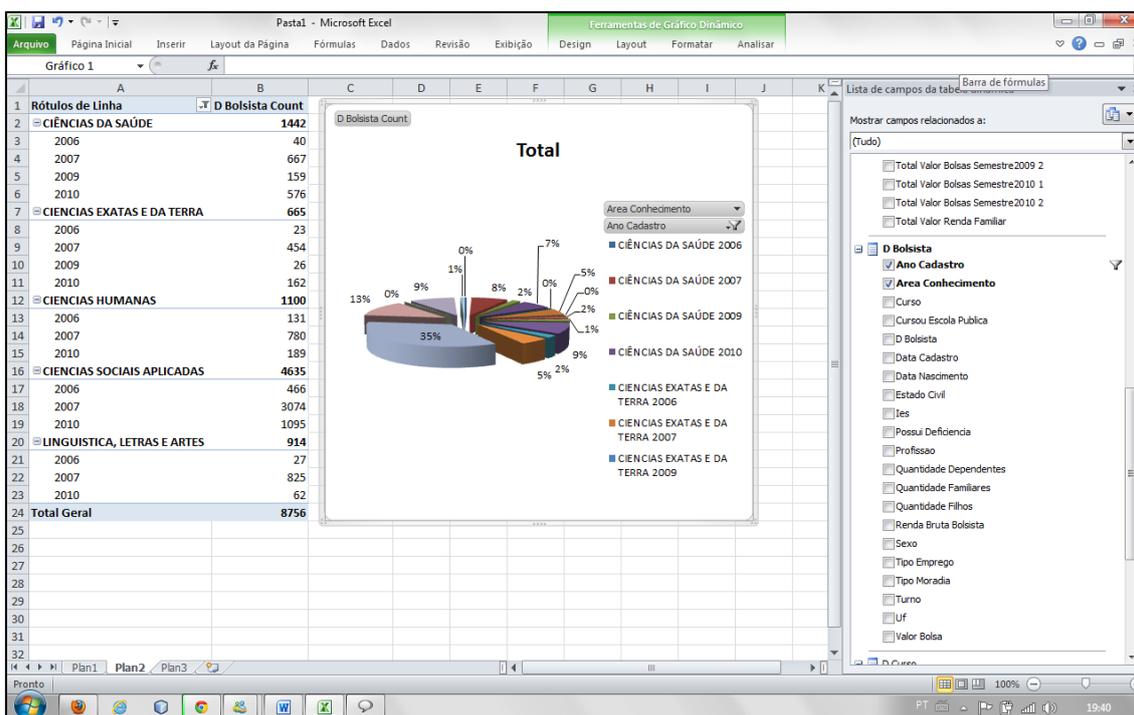


Figura 57: Visão geral de relatório emitido.

Na Figura 57 observam-se os campos disponibilizados a partir da consulta na base OLAP. Observando a figura da direita para a esquerda, a primeira área possui os campos de seleção baseado nas medidas do cubo para geração de relatórios dinâmicos. No outro extremo, lado esquerdo, os valores desses campos, sendo que a partir deles, são gerados os gráficos, presente na parte central da planilha. A seguir, na Figura 58 é apresentado parte do relatório, mais precisamente a região onde são apresentados os valores presente na base de dados.

	A	B
1	Rótulos de Linha	D Bolsista Count
2	CIÊNCIAS DA SAÚDE	1442
3	2006	40
4	2007	667
5	2009	159
6	2010	576
7	CIENCIAS EXATAS E DA TERRA	665
8	2006	23
9	2007	454
10	2009	26
11	2010	162
12	CIENCIAS HUMANAS	1100
13	2006	131
14	2007	780
15	2010	189
16	CIENCIAS SOCIAIS APLICADAS	4635
17	2006	466
18	2007	3074
19	2010	1095
20	LINGUISTICA, LETRAS E ARTES	914
21	2006	27
22	2007	825
23	2010	62
24	Total Geral	8756

Figura 58: Relatório emitido com dados numéricos.

Na Figura 58 são apresentados os resultados do processamento entre as medidas aplicadas sobre as tabelas do cubo. Este relatório foi gerado a partir das áreas de conhecimento e o total de bolsas contempladas a cada ano. Pode-se observar que o quantitativo de bolsistas referentes aos anos de 2006, 2007, 2009 e 2010 considerando as áreas de conhecimento. Ao final, o total parcial e abaixo, o total geral dos bolsistas. A seguir, na Figura 59, será apresentado o gráfico gerado a partir desses dados dinâmicos, bem como os filtros de seleção para geração do gráfico.

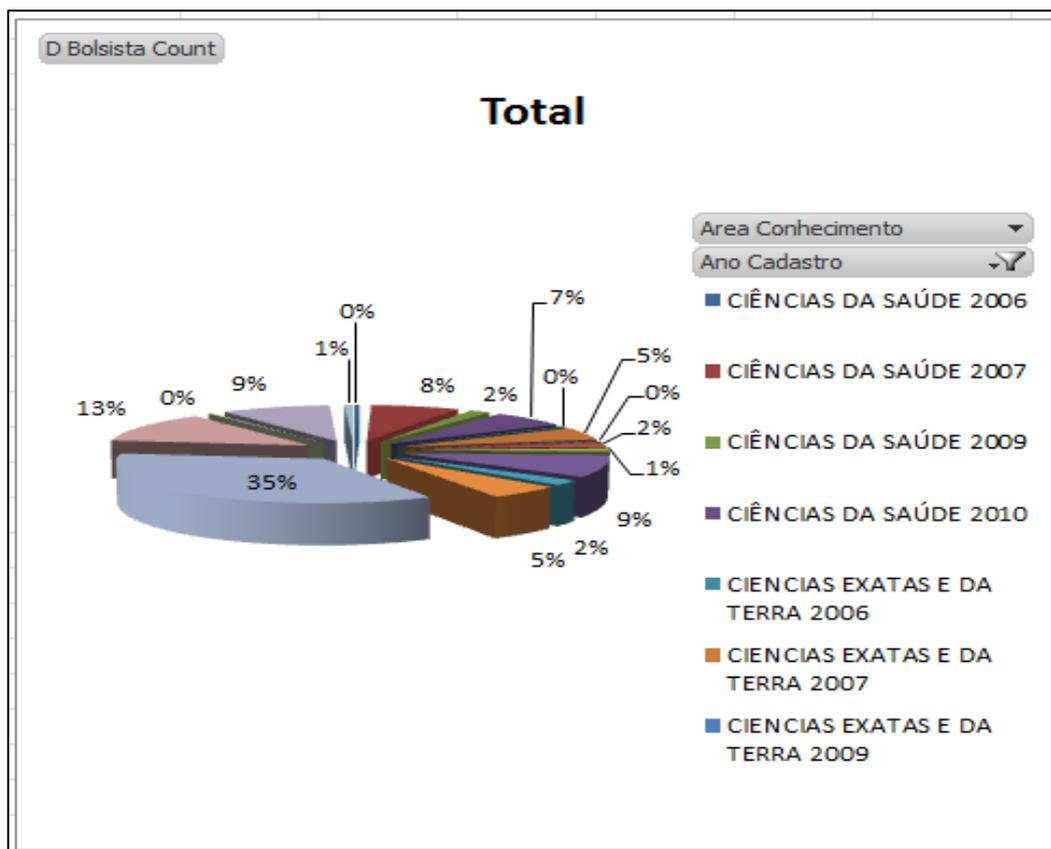


Figura 59: Gráfico gerado a partir dos dados numéricos.

Na Figura 59 é apresentado o gráfico gerado a partir dos dados relacionados ao cadastro de bolsistas por área de conhecimento separado por anos, conforme os valores quantitativos apresentados na Figura 58. No gráfico dinâmico é possível observar que ao gerar o gráfico, os filtros são transportados para a área do gráfico de modo que outras seleções e filtragem possam ser realizadas sobre a base de modo que seja explorada em todos os pontos. Sendo selecionado um dado ou informação da base de dados, o gráfico é atualizado em tempo real. A seguir, na Figura 60 será apresentada a área da planilha responsável por apresentar as áreas disponíveis para filtragem e montagem dos relatórios.

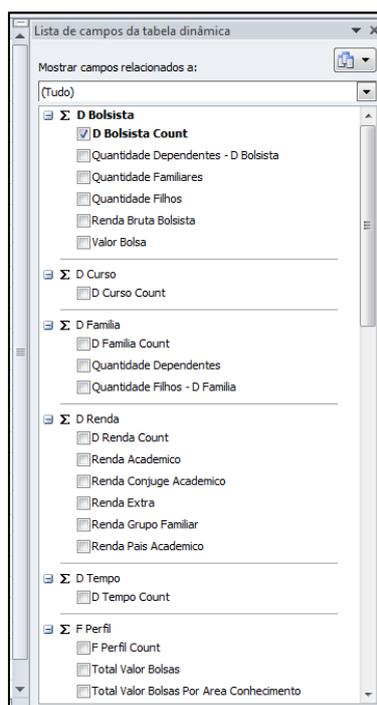


Figura 60: Tabela de montagem de relatórios com base em OLAP.

Na Figura 60 são apresentados os campos da base de dados na qual se pode construir os relatórios. A partir desses valores é construído o relatório quantitativo e, conseqüentemente, os gráficos para análise. Esses valores são disponibilizados a partir das medidas sobre as dimensões. Assim, é possível verificar as colunas das tabelas dimensões. Ao selecionar um campo diferente para ser inserida informação no relatório, é realizada uma consulta na base de dados OLAP, conforme apresentado na Figura 61.

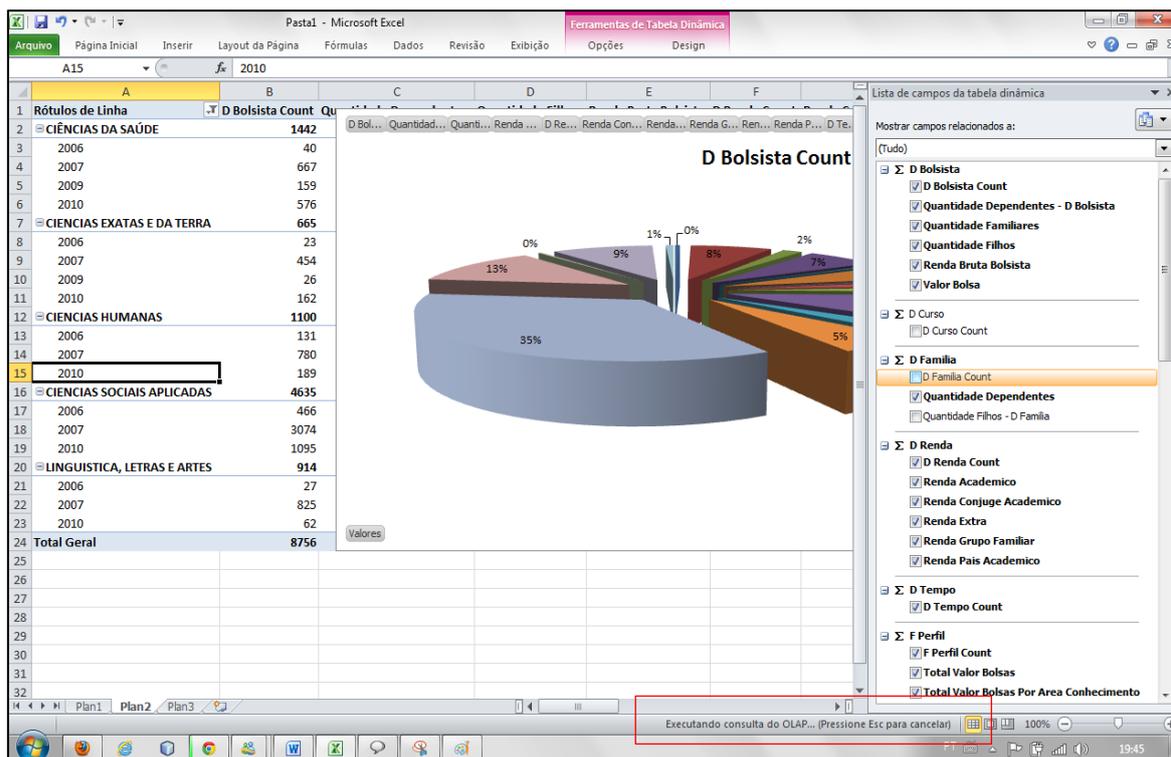


Figura 61: Mensagem de pesquisa em base de dados OLAP para emissão de relatórios dinâmicos.

Na Figura 61 é apresentada novamente a visão geral, porém, o que se deseja destacar é que ao marcar o campo de contagem de perfil (F Perfil Count), por ser um campo com um número elevado de registros, é possível observar conforme marcado, a mensagem “Executando consulta do OLAP...”. Isso para confirmar o alcance do objetivo desse trabalho, na construção de um SAD baseado em DW utilizando uma ferramenta OLAP, e a partir desse ponto, fica a cargo do gestor a definição dos dados a serem cruzados para emissão de relatórios e extração de conhecimento. A seguir, na seção 4.4.6 são apresentados os gráficos gerados a partir da consulta na base OLAP.

4.4.6 Gráficos Gerados a partir da Base de Dados OLAP

Nesta seção serão apresentados os relatórios das questões levantadas junto a coordenação do programa, de modo que sejam representados graficamente os valores processados a partir da base de dados do *Data Warehouse* com o emprego da ferramenta OLAP para composição do Sistema de Apoio à Decisão. Para tanto, serão apresentados os gráficos para explorar os dados e

a partir das informações possam ser observadas algumas variáveis que respaldem e orientem o processo decisório sobre o programa. Assim sendo, na Figura 62, são apresentados dentro do montante dos bolsistas do programa, o percentual de assistência por ano, e assim, possibilitar uma visão geral sobre a assistência.

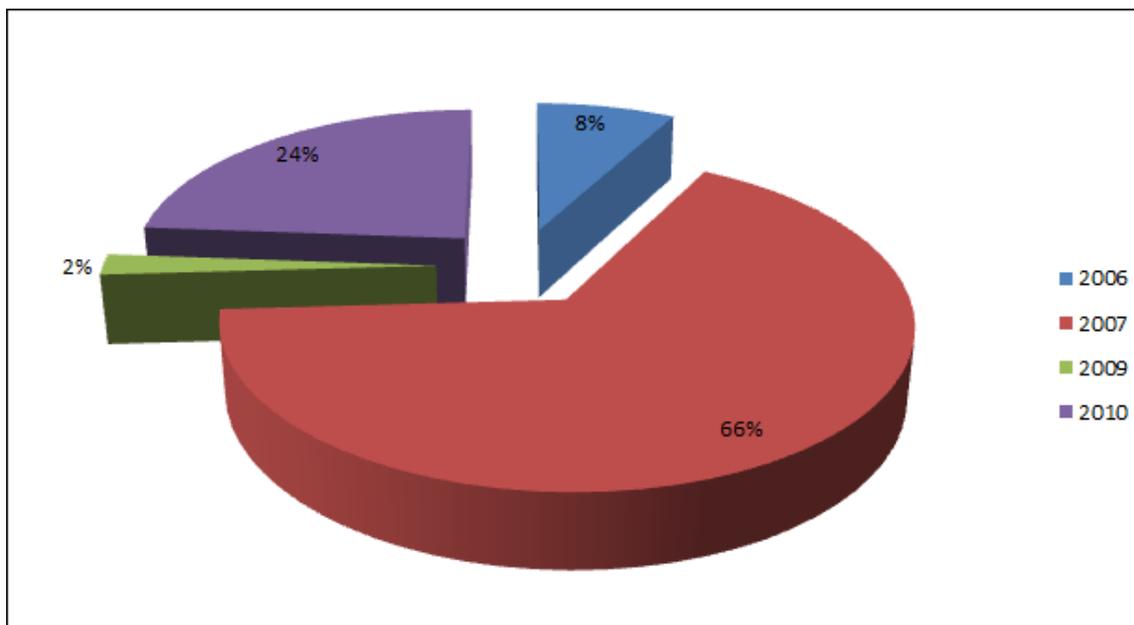


Figura 62: Total de bolsistas contemplados por ano.

Conforme a Figura 62, pode-se observar que dentro do montante geral dos assistidos no programa bolsa universitária, são apresentados dados referentes os anos de 2006, 2007, 2009 e 2010. Sobre o ano de 2008 a coordenação do programa não disponibilizou dado, o que impossibilitou o processo de tratamento e, conseqüentemente, a carga na base de dados do *Data Warehouse*. Nesse sentido, é possível perceber que o ano de 2007 possui o maior percentual de assistência, apresentando 66% dos assistidos seguido de 2010 com 24%, 2006 com 8% e 2009 com 2%. Pode-se considerar nesse momento um equívoco no planejamento na concessão de bolsas podendo ocorrer um estouro no orçamento disponível, impossibilitando a concessão para o ano seguinte. Essa afirmação não pode ser baseada somente nesse índice, porém, a discrepância entre os valores percentuais deixam margem para averiguação, podendo ser verificado junto ao sistema financeiro e, no caso, realizar o levantamento de novos índices que comprovem essa indicação. Ainda considerando o montante de bolsas concedidas, a seguir, na Figura 63 são apresentados os índices de assistência por área de

conhecimento por ano, ou seja, serão apresentados gráficos exclusivos da assistência para cada ano do programa.

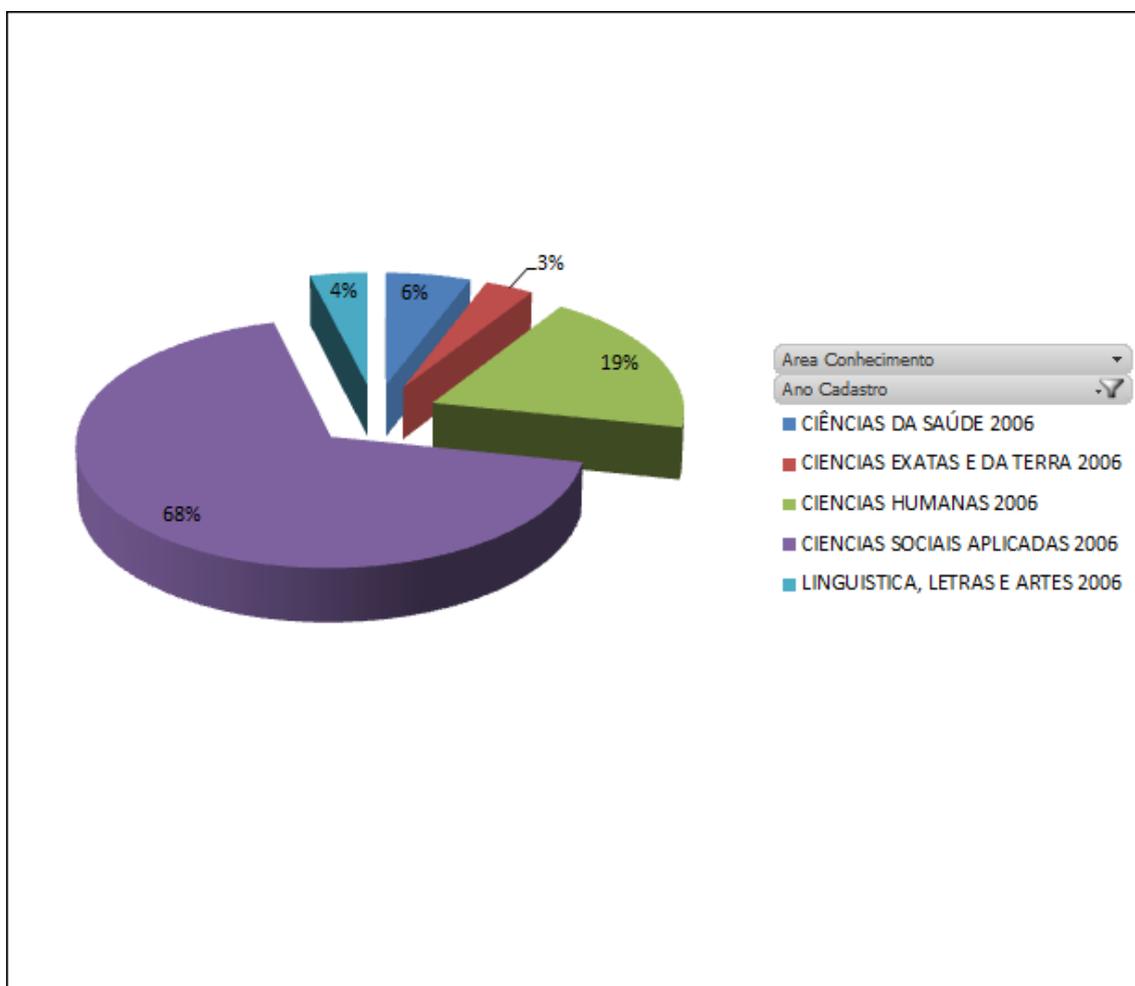


Figura 63: Percentual de área de conhecimento.

Na Figura 63 são apresentados os dados percentuais das bolsas concedidas no ano de 2006, considerando as áreas de conhecimento. Nota-se que na região dos gráficos são disponibilizadas as opções de filtros sobre os dados utilizados no cruzamento para emissão do relatório. Com isso, o gestor, ao gerar o relatório tem a possibilidade de filtrar os anos de interesse e realizar o cruzamento dos dados com as áreas de conhecimento. Conforme apresentado na figura, 68% dos bolsistas, são de cursos da área das ciências sociais aplicadas. Essa informação tem sua natureza em situações diversas, entre elas: ao iniciar uma IES, os primeiros cursos disponibilizados estão relacionados a essa área, pois não requerem a montagem de laboratórios específicos e também, não possui dificuldade de encontrar docentes para essa

área e com isso, pode-se projetar a IES até que se possa disponibilizar cursos de outras áreas de conhecimento. A seguir, a Figura 64 apresentará os valores da área das ciências sociais aplicadas considerando uma IES, a fim de validar a demanda de concessão de bolsas nesta área.

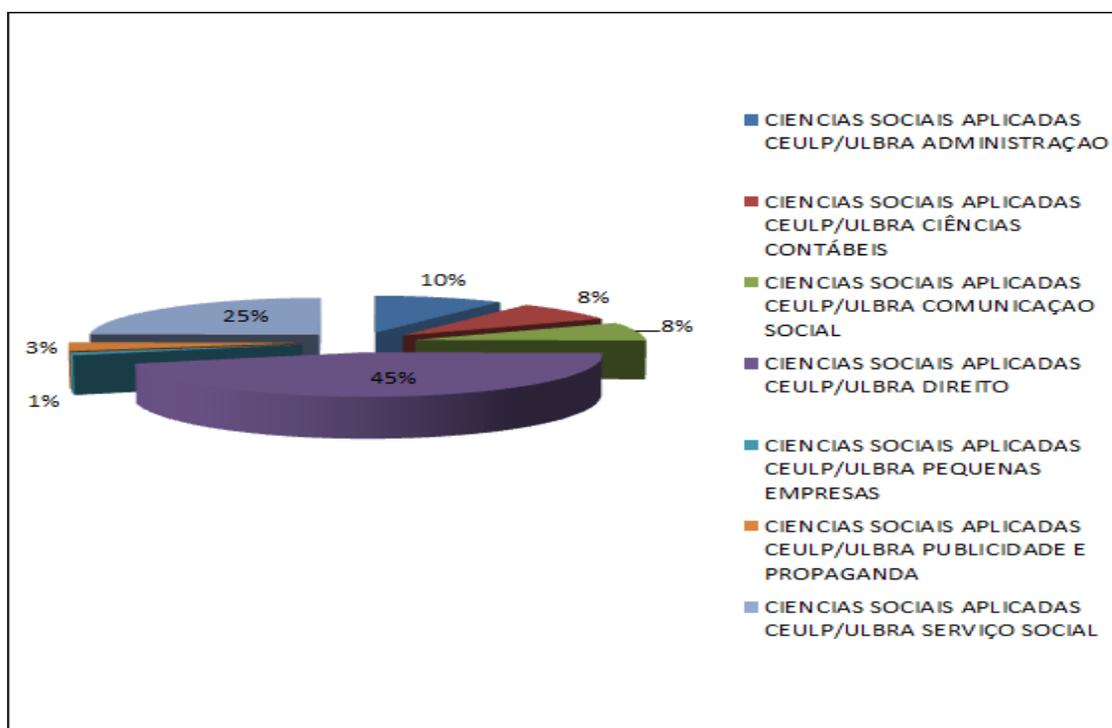


Figura 64: Bolsistas por ano, área de conhecimento e IES.

Para emissão do relatório apresentado na Figura 64, foi considerado o montante bolsistas da área de ciências sociais aplicadas e da IES CEULP-ULBRA. Esse gráfico foi emitido para validar a demanda de bolsas concedidas aos acadêmicos dessa área de conhecimento. Neste caso, foi considerado apenas uma IES, tendo em vista a disponibilização de informações no gráfico o que torna ilegível. Mas também, comprovando que a grande massa das bolsas concedidas são de áreas relacionadas à gestão pública, o que fortalece a visualização da vocação administrativa do estado do Tocantins. Essas informações podem ser emitidas considerando as IES cadastradas e contempladas com o programa e ainda, considerando os recursos alocados sobre cursos das mesmas. A seguir, na Figura 65 serão apresentados os percentuais de bolsistas contemplados por UF.

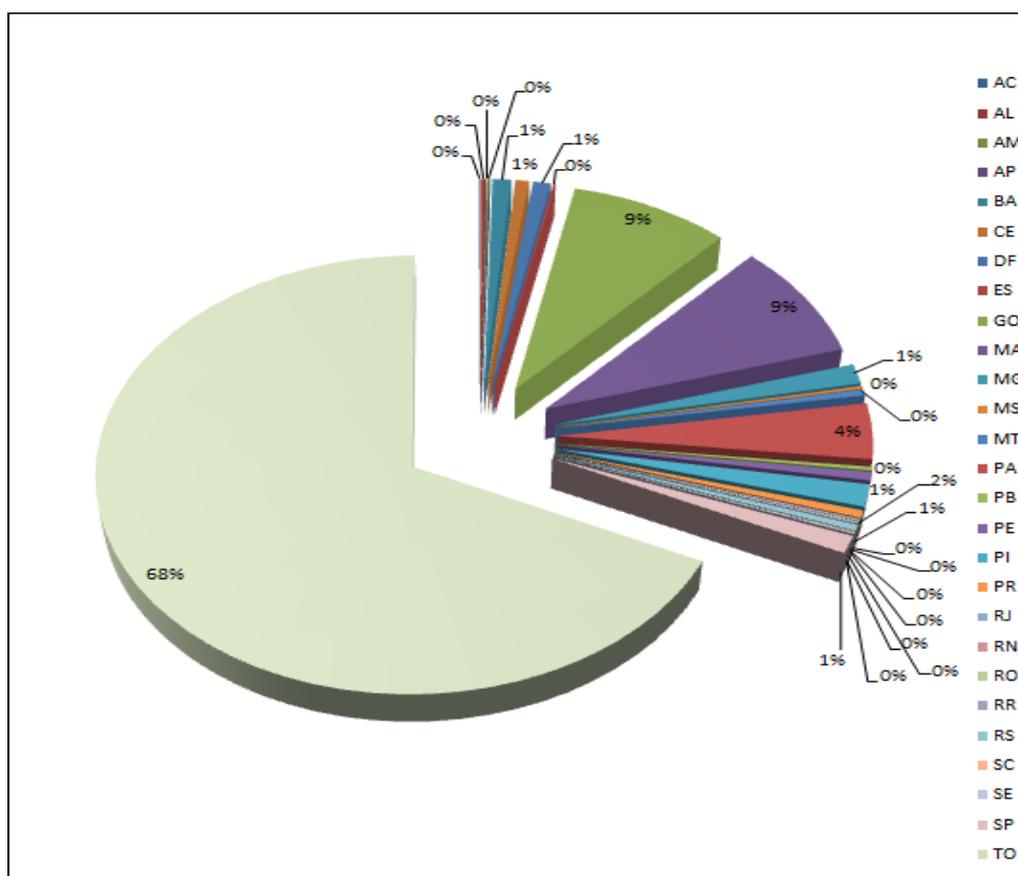


Figura 65: Total de bolsistas por UF.

Na Figura 65, são apresentados os valores percentuais referentes às bolsas concedidas a acadêmicos e a classificação por unidade federativa, de modo que sejam identificados os naturais do estado do Tocantins. Isso para verificar a efetividade do programa sobre para a população do estado. Conforme apresentado, verifica-se que 68% dos bolsistas são naturais deste estado, com isso, comprovando a eficiência do programa em contribuir para o aumento do ingresso e formação do estudante tocantinense no ensino superior. Verifica-se ainda a diferença é bastante significativa em relação às outras unidades federativas e que houve empenho na atenção e também na procura por parte do estudante tocantinense no auxílio à formação superior. A seguir, na Figura 66 são apresentados dados percentuais referentes bolsistas naturais do estado do Tocantins, considerando as áreas de conhecimento.

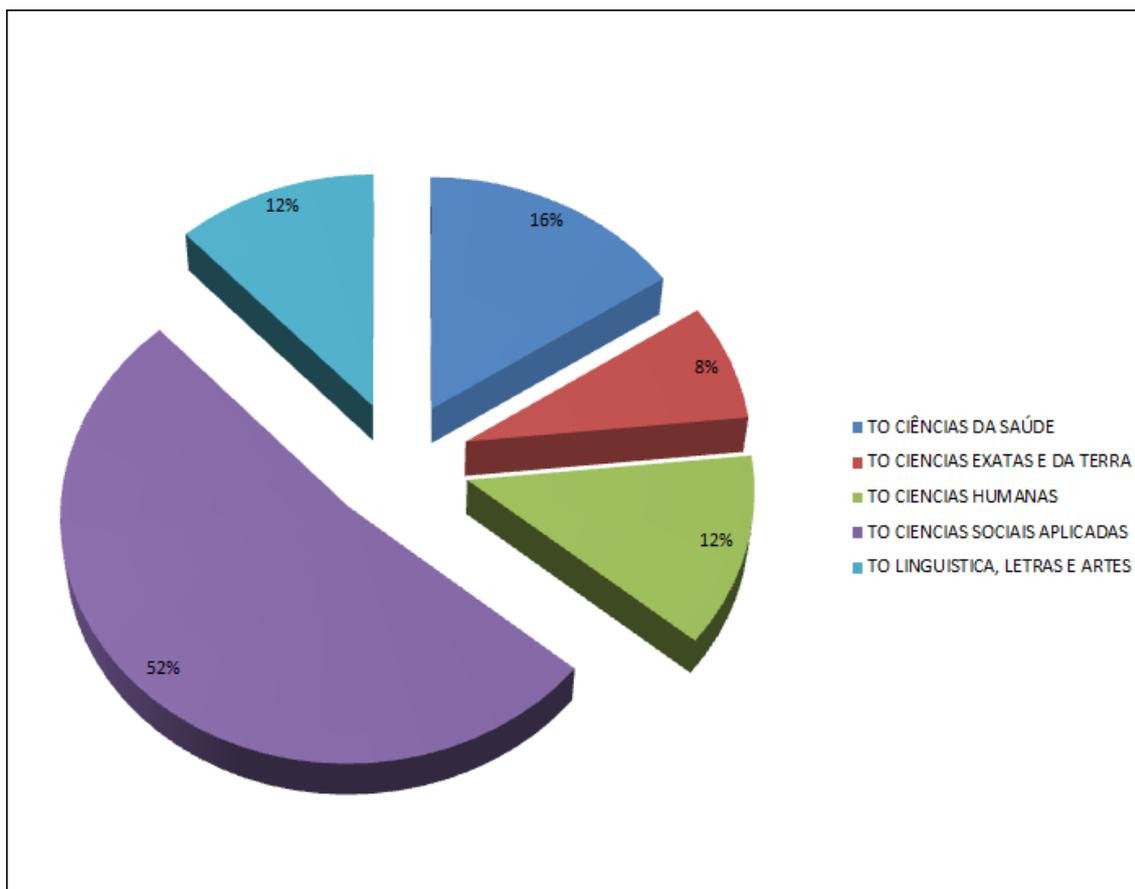


Figura 66: Bolsistas oriundos do estado do Tocantins preferência por área de conhecimento.

Conforme Figura 66, pode-se observar que dos bolsistas contemplados, a grande maioria são de cursos da área das ciências sociais aplicadas, com 52%. Em seguida, ciências da saúde com 16%, linguística, letras e artes e ciências humanas com 12% cada e ciências exatas com 8%. Esse gráfico foi gerado para além de confirmar a área de conhecimento de maior cobertura pelo programa, a preferência para os bolsistas naturais do estado do Tocantins. A partir desses dados a coordenação do programa pode se orientar no que tange a disposição de vagas, e planejamento sobre valores, para estimar os investimentos semestrais e ainda propor soluções para regiões específicas do estado, a fim de aumentar atuação do programa na atenção a estudantes do ensino superior. A seguir, na 4.4.6.1 serão apresentados gráficos a fim de orientar no conhecimento do perfil socioeconômico do bolsista.

Na Figura 67, é apresentado o relatório gráfico considerando as seguintes informações: ano de ingresso no programa, área de conhecimento, curso, estado civil, número de dependentes e o sexo do bolsista. Percebe-se que o relatório gráfico ficou bastante fracionado e os percentuais baixos. Assim sendo, para traçar o perfil é interessante cruzar esses valores percentuais com o montante total e, nesse caso, verificar o nível de interferência da quantidade de dependentes com o tempo para formação. Estas informações podem auxiliar na definição do período de concessão da bolsa para candidatos com características (perfil) similares.

A ferramenta possibilitou o cruzamento de dados para extração de conhecimento, possibilitando a coordenação do programa a verificar os dados sobre a alocação dos recursos do programa. A seguir, na Figura 68 é apresentado o relatório referente à alocação de recursos considerando as áreas de conhecimento.

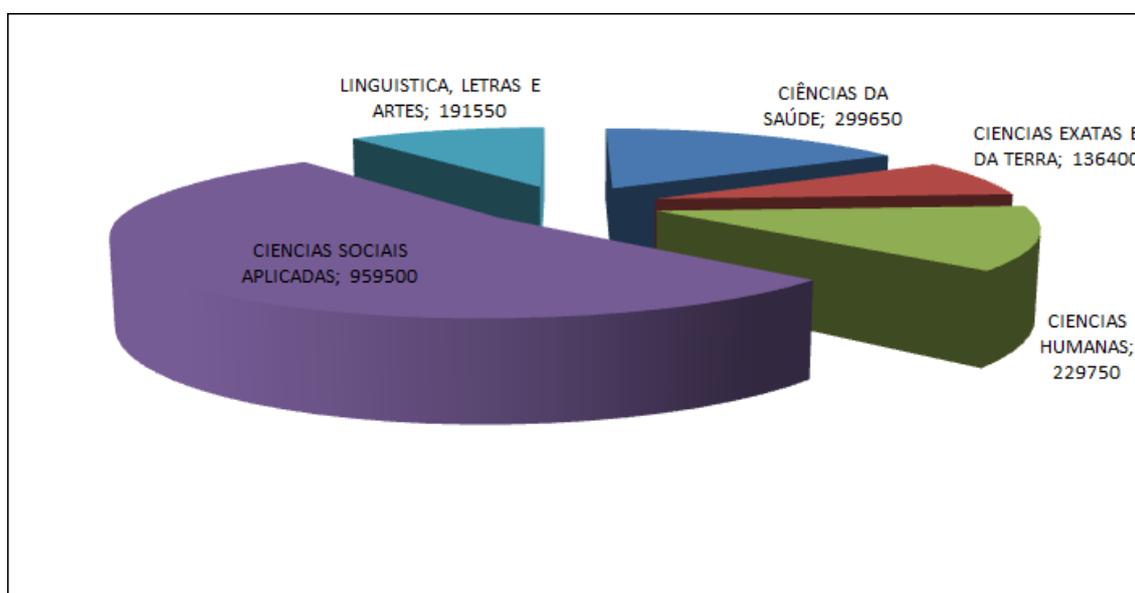


Figura 68: Valores empregados em bolsas por área de conhecimento.

Conforme apresentado na Figura 68, pode-se observar os recursos alocados por área de conhecimento. Nesse caso, com base nesse relatório gráfico a equipe coordenadora pode verificar a eficiência na atuação do programa na assistência a bolsistas, e ainda, confirmar a disposição das bolsas

por área considerando os valores empregados. Nesse caso, pode-se também relacionar esses valores verificando a efetividade dos recursos para atender bolsistas de determinado perfil e assim propor mudanças para melhor assistência do programa. A seguir na seção **Erro! Fonte de referência não encontrada.** são apresentadas as considerações finais deste trabalho.

4.4.7 Conclusões

Nesta seção serão apresentadas as considerações sobre o processo de construção do Sistema de Apoio à Decisão baseado em *Data Warehouse*, considerando todo o processo apresentado, além da facilidade de integração e utilização da ferramenta para geração de gráficos e emissão de relatórios, utilizando o *Microsoft Excel*. Nesse sentido, inicialmente é necessário apresentar a arquitetura da aplicação desenvolvida, que é apresentada na Figura 69.

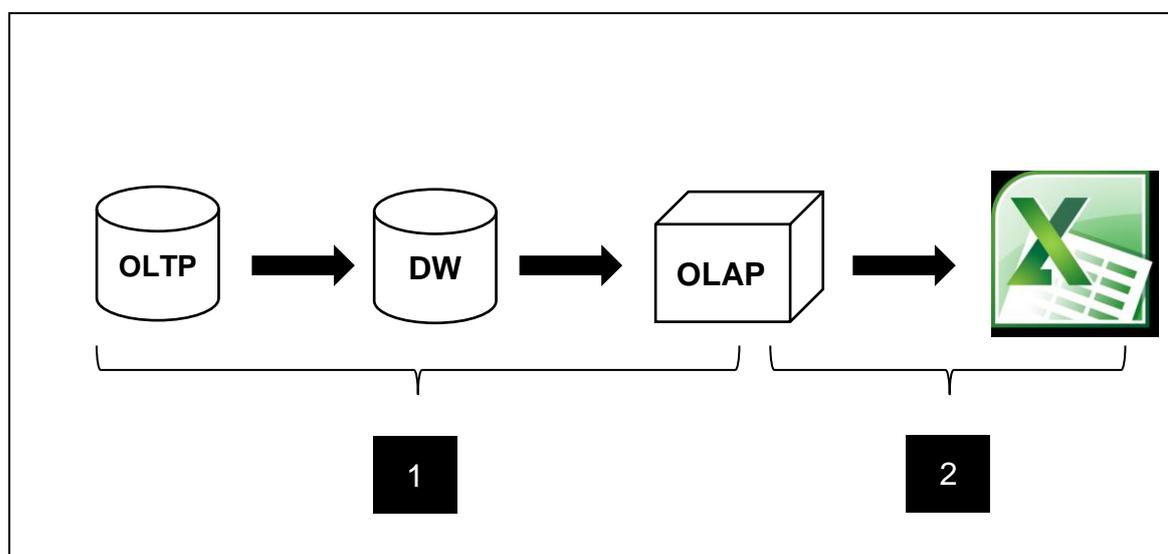


Figura 70: Arquitetura do SAD.

Conforme apresentado na Figura 70, o programa bolsa universitário não possuía uma base de dados operacional de modo que fosse possível armazenar os dados dos bolsistas do programa. Desta forma, foi implementada uma base de dados que possibilitasse armazenar e organizar os dados dos candidatos, de forma a permitir que uma aplicação pudesse ser construída.

Isso porque os dados encontravam-se dispersos em arquivos e planilhas eletrônicas. Com esse modelo relacional construído e a junção dos dados dispersos neste ambiente, o processo de construção do *Data Warehouse*, bem como a carga dos dados foi facilitada. A partir da construção e carga dos dados no ambiente do *Data Warehouse*, iniciou o processo de implementação do SAD, a partir de um projeto *Analysis Services*. Nesse projeto foram criados cubos e métricas sobre os valores das tabelas dimensões do *Data Warehouse*.

O processo de implementação do Sistema de Apoio à Decisão requereu grande trabalho, tendo em vista o processo de estruturação e organização dos dados, construção do modelo para *Data Warehouse*, o projeto *Analysis Services* e a integração da ferramenta OLAP. Assim sendo, buscou-se nesse momento algo que favorecesse a operação do mesmo, de modo que ficasse a cargo do gestor o cruzamento das informações conforme a necessidade e consequentemente a emissão dos relatórios. Então, foi integrado ao projeto o MS Excel, que permite fazer conexão aos projetos *Analysis Services*.

Assim sendo, por meio do Excel, foi viabilizado o controle total da geração do relatório a cargo do gestor, não requerendo a presença do profissional de TI, o que além de facilitar a utilização da ferramenta, possibilitou ao gestor a extração da informação conforme a necessidade e praticamente em tempo real, sendo atribuição direta no processo decisório. Então, ainda sobre a **Erro! Fonte de referência não encontrada.**, as atividades marcadas por número 1, são atribuições do profissional de TI e a de número 2, podem ser realizadas integralmente pelo gestor.

5 CONSIDERAÇÕES FINAIS

O apoio de um Sistema de Informação no processo decisório oferece agilidade e segurança à instituição, pois possibilita a exploração dos dados para extração de conhecimento, podendo, subsidiar traçagem de metas e estratégias para a tomada de decisão. Tais características se tornam significantes para uma organização, pois o crescimento no volume de dados a serem manipulados é exponencial. Com isso, aumentam-se também as variações entre as consultas a serem realizadas, as combinações e relações das informações.

O ambiente de *Data Warehouse* auxilia no processo de agrupamento dos dados. Nesse ambiente, os dados se encontram padronizados e detalhados, conforme interesse e utilidade para a organização. A forma de disponibilização dos dados a partir do ambiente de Data Warehouse favorece a extração de informações sobre fatos específicos, apoiando favoravelmente o processo decisório.

Porém, o ambiente de DW não é suficiente para orientar a organização no processo decisório e, para tanto, faz-se necessário à agregação de ferramentas OLAP para que o processo seja eficiente, pois o DW armazena os dados e ferramentas OLAP são utilizadas para recuperá-las, explorá-las e gerar informações. Assim, pode-se afirmar que as tecnologias são complementares, pois a partir de um DW bem elaborado os relatórios produzidos serão de qualidade e orientarão a organização com exatidão e segurança.

É crescente a evolução das ferramentas OLAP para o processo de recuperação e análise de dados e oferece grande vantagem por seu funcionamento interativo, o que permite que a partir das informações extraídas, novas áreas sejam exploradas e identificadas, auxiliando no aperfeiçoamento do processo da organização como um todo. OLAP e DW compõe o conceito de

BI e oferecem recursos para que a organização defina a partir dos requisitos e especificações do negócio subsídios para o processo decisório e, com isso, possibilita a organização criar métricas nas ações para alcançar seus objetivos de expansão e produção.

Este trabalho foi finalizado e seu objetivo foi alcançado, porém, não houve tempo hábil para validação por parte da equipe coordenadora do programa. Mas os relatórios emitidos apresentam valores e métricas do programa que podem facilitar a extração de informações e construção de conhecimento para então auxiliar no processo estratégico e decisório para reativação do programa de concessão de bolsas, tendo em vista a disposição de dados do programa para cruzamento. A verificação de alguns dados referentes a valores abordados no projeto não pôde ser comprovada, tendo em vista a necessidade de consultar o sistema financeiro, sendo uma atividade trabalhosa onde a coordenação do programa poderá verificar todos os extratos e comprovantes para auxiliar no processo de tomada de decisão. Essas consultas em sistemas auxiliares, estão relacionadas a política interna para acesso e por isso a consulta não pode ser realizada para construção do sistema. A seguir, a seção 5.1 é apresentada a sugestão de trabalhos futuros.

5.1 Trabalhos futuros

Como proposta de trabalhos futuros, é sugerido a mineração do *Data Warehouse*, utilizando técnicas de *Data Mining*, para então, gerar novas consultas e assim, refinando o processo decisório a partir do detalhamento e extração de padrões da base. Sugere-se também, a criação de um processo ETL, de modo que periodicamente os dados da base operacional sejam carregados no ambiente do *Data Warehouse*.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, Fabio Silva Gomes da Gama. **Desmistificando o conceito de ETL**. Faculdade Salesiana Maria Auxiliadora. Macaé – RJ, 2008. Revista de Sistemas de Informação. Disponível em: http://www.fsma.edu.br/si/Artigos/V2_Artigo1.pdf. Acessado em: junho de 2011.

ALMEIDA, Alexandre Marques. **Proposição de indicadores para avaliação técnica de projetos de Data Warehouse**: um estudo de caso no *Data Warehouse* da plataforma Lattes. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Santa Catarina, UFSC – 2006. Disponível em: <http://www.ufpe.br/gepec/exemplos/ex%20tese%2014975.pdf>. Acessado em: maio de 2011.

ALVES, Celso; Mauro; Schaffer, Samuel. **Sistema de Apoio à Decisão. EIS – Executive Information System**. Universidade do Vale do Rio dos Sinos, São Leopoldo – RS, 2006. Disponível em: <http://www.datawarehouse.inf.br/Academicos/eis.pdf>. Acessado em: março de 2011.

BARBIERI, Carlos. **Business Intelligence: modelagem e tecnologia**. Rio de Janeiro: Axcel Books, 2001. 424 p.

BARQUINI, Ramon. **Planning and designing the Warehouse**, Prentice Hall. New Jersey, 1996. 311 p.

BATISTA, Emerson de Oliveira. **Sistema de Informação: o uso consciente da tecnologia para o gerenciamento**. Editora Saraiva. São Paulo, 2004

BONI, Laercio Aparecido. **Proposta de um Data Warehouse para tomada de decisão na gestão da produção em uma indústria farmacêutica**. Dissertação (Mestrado em Engenharia de Produção), 2009. Universidade Metodista de Piracicaba. Disponível em: http://www.unimep.br/phpg/bibdig/pdfs/docs/15032011_115434_laercio_boni.pdf. Acessado em: maio de 2011.

BORTOLIN, Sérgio Antônio Martini Júnior. **Sistemas de Apoio à Decisão**. Dissertação (Trabalho de Conclusão de Curso), 2005. Universidade da Região da Campanha, URCAMP. Alegrete, RS. Disponível em: <<http://www.al.urcamp.tche.br/infocamp/edicoes/nov05/Apoio%20a%20Deci sao.pdf>> Acessado em: abril de 2011.

CLERICUZI, Adriane Zenaide; Almeida, Adiel Teixeira; Seixas Costa, Ana Paula Cabral. **Aspectos relevantes do SAD nas organizações: um estudo exploratório**. Dissertação (Mestrado em Computação). Universidade Federal de Pernambuco, 2006. Disponível em: http://www.scielo.br/scielo.php?pid=s0103-65132006000100002&script=sci_arttext . Acessado em: maio de 2011.

CUNHA, Nara Inês Pereira Vieira. **Proposta de um modelo de dados dimensional como contribuição ao processo de gestão de estoques**. Dissertação (Mestrado em Engenharia de Produção e Sistemas). Pontifícia Universidade Católica do Paraná, 2005. Disponível em: <http://www.produtronica.pucpr.br/sip/conteudo/dissertacoes/pdf/NaraCunha.pdf> Acessado em: junho de 2011.

DALFOVO, Oscar. **Desenho de um modelo de sistema de informação estratégico para a tomada de decisão nas pequenas e médias empresas do setor têxtil de Blumenau**. Dissertação (Tese Doutorado em Ciência da Computação). Universidade Regional de Blumenau – Santa Catarina, 1998. Acessado em: Abril de 2011. Disponível em: <http://www.inf.furb.br/~dalfovo/tese/dissfinad.pdf>

DATAPRIX Information Technologies. **Knowledge Is Goal**. Disponível em: <<http://www.dataprix.net/pt-pt/24-data-mart>>. Acessado em: maio de 2011.
FARIAS, Gabriela Belmont. **O bibliotecário – Gestor da Informação: Representações do segmento imobiliário sobre competências**. Dissertação (Trabalho de Conclusão de Curso). Universidade Federal de Santa Catarina, 2007. Disponível em: < <http://www.cin.ufsc.br/pgcin/GabrielaFarias.pdf>>. Acessado em: abril de 2011.

ESPÍRITO SANTO, Frederico Guilherme Mariani. **Construção de um Data Mart para apoio às tomadas de decisão das empresas PROEMBARQUE e CASACON**. Dissertação (Trabalho de Conclusão de Curso em Ciência da Computação), 2006. Universidade do Vale do Itajaí. Itajaí – SC. Disponível em: < <http://siaibib01.univali.br/pdf/Frederico%20Guilherme%20Mariani%20do%20Esp%C3%ADrito%20Santo.pdf>> Acessado em: junho de 2011.

FELBER, Edmilson J. W. **Proposta de Uma Ferramenta OLAP em um Data Mart Comercial**: Uma Aplicação prática na Indústria Calçadista. Novo Hamburgo: 2006. 102 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Instituto de Ciência Exatas e Tecnológicas, Centro Universitário Feevale, Novo Hamburgo.

FURLAN, José Davi; Ivo, Ivonildo da Mota; Amaral, Francisco Piedade. **Sistemas de Informação Executiva. IES - Executive Information System**. Editora Makron Books, São Paulo, 1994. 155 p.

GARCIA, Elias; GARCIA, Osmarina Pedro Garcia. **A importância do sistema de informação gerencial para a gestão empresarial**. Revista Ciências Sociais em Perspectiva, do Centro de Ciências Sociais Aplicadas de Cascavel - PR, v.2 , n.1, pags. 21-32,. 2003.

GOMES, João Augusto Mescua. **Apoio tecnológico à estratégia organizacional**: fatores para o sucesso no uso de Business Intelligence. Dissertação (Mestrado em Administração), 2010. Universidade FUMEC, Belo Horizonte, MG. Disponível em: <http://www.face.fumec.br/cursos/mestrado/dissertacoes/completa/joao_augusto_mescua_gomes.pdf> . Acessado em: junho de 2011.

GONÇALVES, Márcio. **Extração de dados para Data Warehouse**. Editora Axcel Books, 1ª edição. Rio de Janeiro, 2003

INMON, William H. **Como construir o Data Warehouse**. Rio de Janeiro, 1997. Editora Campus. 392 p.

JÚNIOR, Ary Ferreira. **Proposta de implantação de um “Sistema de informação empresarial”, como uma ferramenta estratégica de desenvolvimento, crescimento e geração de emprego e rendas nas micro e pequenas empresas**. Dissertação (Trabalho de Conclusão de Curso) Universidade Federal do Rio de Janeiro, UFRJ. Rio de Janeiro, 1999.

KIMBALL, Ralph; Ross Margy; Thornthwalte, Warren; Mundy, Joy; Becher, Bob. . **The Data Warehouse toolkit**. Editora John Wiley & Sons Inc, 2002. 436 p.

KUHNEN, Cleomir; Vieira, Valter Afonso. **Vantagens obtidas com a utilização da implantação do Data Warehouse**: um estudo de caso na UNIMED de Francisco Beltrão. Dissertação (Trabalho de Conclusão de Curso). Universidade de São Paulo, 2005. Disponível em:

<http://www.ead.fea.usp.br/semead/8semead/resultado/trabalhosPDF/20.pdf>

Acessado em: abril de 2011.

KURIKE, Rafael Z; Dias, Leonardo Renato. **Uma pesquisa em constante desenvolvimento na área de Mineração de Dados**. Universidade Estadual de Maringá. Paraná, 1999. Disponível em: <
http://www.din.uem.br/ia/a_multid/mineracao/> Acessado em: maio de 2011.

KUROGI, Adelina Machie Hashimoto; OLIVEIRA, Raquel Rida Valente de; MORAIS, Rosângela Damasceno. **Desenvolvimento de um Data Warehouse para a área de telecomunicações móveis**. Dissertação (Trabalho de Conclusão de Curso em Processamento de Dados), 2001. Universidade da Amazônia – UNAMA.

Disponível em <
http://www.nead.unama.br/site/bibdigital/monografias/DATA_WAREHOUSE.pdf
>. Acessado em: maio de 2011.

LAUDON, Kenneth; Laudon, Jane P. **Sistemas de informação gerenciais, administrando a empresa digital**. Prentice Hall, 5ª edição. São Paulo, 2005.

LERMEN, Roberto. **Gestão estratégica e o uso do Balanced Scorecard**. Revista eletrônica de contabilidade, vol III nº 1. Jan – Jun / 2006. Revista Ciência Contábil. Disponível em:
<<http://w3.ufsm.br/revistacontabeis/anterior/artigos/vIIIIn01/a05vIIIIn01.pdf>>

Acessado em: abril de 2011.

LIMA, Claudiomir Cirino. **Gestão da Tecnologia da Informação: uma análise em empresas de máquinas e implementos agrícolas**. Dissertação (Trabalho de Conclusão de Curso). Universidade de Passo Fundo, RS, 2007. Disponível em:
http://www.upf.br/computacao/images/stories/TCs/arquivos_20071/claudio_cirino.pdf. Acessado em: abril de 2011.

LIMA, Helena Carolina Costa. **Utilização de Data Warehouse e Data Mining no acompanhamento das atividades de pesquisa do CEULP/ULBRA**. Dissertação (Trabalho de Conclusão de Curso). Centro Universitário Luterano de Palmas – Tocantins, 2006.

LUSTOSA, Rodrigo Bastos. **Processo de desenvolvimento participativo de sistema de Data Warehouse: uma aplicação no PROGER**. Dissertação (Tese de Mestrado em Computação). Universidade Federal da Paraíba – João Pessoa, 2009. Disponível em: <
http://www.ccsa.ufpb.br/ppga/site/arquivos/dissertacoes/dissertacao_61.pdf>

Acessado em: junho de 2011.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e Projeto de Data Warehouse**: uma visão multidimensional. 1ª Edição, São Paulo. Editora Érica, 2000. 320 p.

MARTINS, Daniela Siqueira. **Proposta de desenvolvimento de um Data Mart na área de gestão de pessoas e aplicação de uma ferramenta OLAP**. Dissertação (Trabalho de Conclusão de Curso). Centro Universitário FEEVALE. Novo Hamburgo, RS. 2007 Disponível em: <<http://tconline.feevale.br/tc/files/1212.pdf>>. Acessado em: maio de 2011.

MORENO, Luan. **Data Warehouse e BI**. Disponível em: <<http://luanmorenodba.wordpress.com/2010/09/15/data-warehouse-e-bi-%E2%80%93-parte-ii/>> Acessado em: junho de 2011.

OLIVEIRA, Djalma de Pinho Rebouças. **Sistemas de Informações Gerenciais**: Estratégias e táticas operacionais. Editora Atlas, 8ª edição. São Paulo, 2000.

PADOVEZE, Clóvis Luís. **Contabilidade gerencial**: um enfoque e sistemas de informação contábil. Editora Atlas, São Paulo, 2000. 648 p.

PALITOT, Álvaro Alencar Barbosa. **Análise comparativa de arquiteturas de distribuição de Data Warehouse**. Dissertação (Trabalho de Conclusão de Curso em Ciência da Computação). Universidade Federal de Pernambuco. 2007. Disponível em: <<http://www.cin.ufpe.br/~tg/2006-2/aabp.pdf>>. Acessado em: maio de 2011.

POZZEBON, Marlei e Freitas, Henrique M. R. **Construindo um EIS (Enterprise Information System) da (e para) empresa**. Revista de Administração, São Paulo, 1996. Ed out/dez 1996.

REZENDE, Denis Alcides. **Engenharia de Software e Sistemas de Informação**. Editora Brasport livros e multimídia., Rio de Janeiro, 2005. 344 p.
SINGH, Harry. **Data Warehouse: Conceitos, Tecnologias, Implementação e Gerenciamento**. Editora Makron Books. São Paulo, 2001, 406 p.

SOARES, Vânia Jesus de Araújo. **Modelagem Incremental no Ambiente de Data Marts**. Dissertação (Trabalho de Mestrado em Informática). Universidade Federal do Rio de Janeiro – UFRJ, 1998.

SOUZA, Antônio Artur; Passalongo, Cristiani. **Avaliação de Sistemas de Informação Contábeis**: Estudo de Casos Múltiplos. Encontro da Associação

Nacional de Pós-Graduação e pesquisa em Administração (EnANPAD), Brasília – DF, 2005. Disponível em: http://www.anpad.org.br/evento.php?cod_evento_edicao=9. Acessado em: maio de 2011.

TAVARES, Renato. **Proposta de um sistema de apoio à decisão para controle e gerenciamento agrícola em usinas de açúcar e álcool**. Dissertação (Mestrado em Engenharia de Produção). Universidade de São Paulo, 2008. Disponível em: http://www.datawarehouse.inf.br/Academicos/Tese_Final_Renato.pdf. Acessado em: maio de 2011.

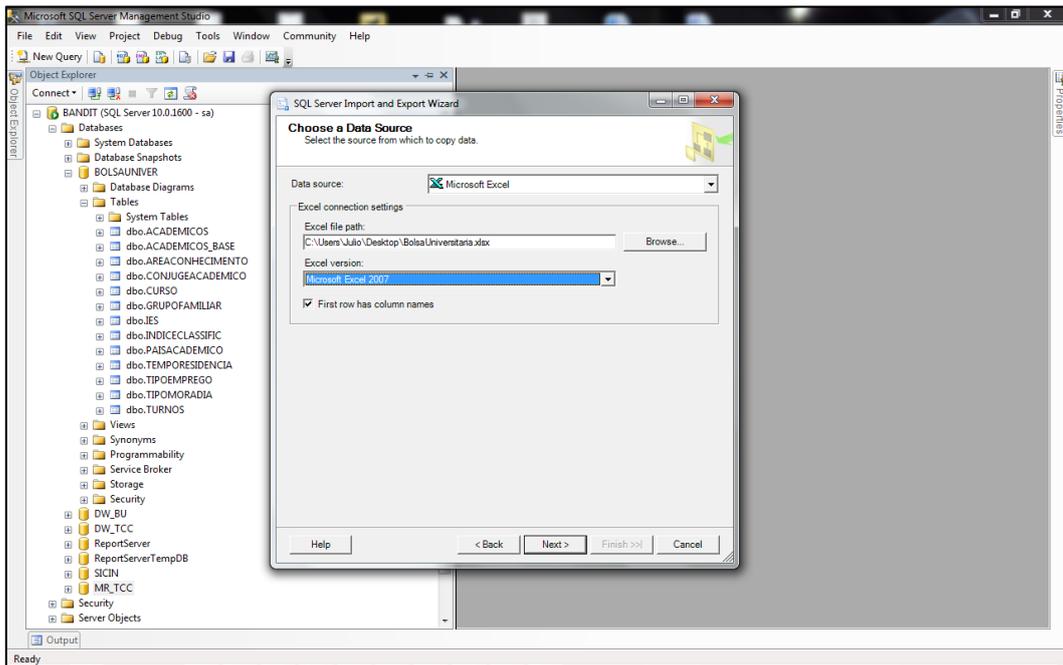
TRABSI. **Sistema de Informação Gerencial - SIG**. Disponível em: <http://trabsi.wordpress.com/2011/>. Acessado em: abril de 2011.

YONG, Chu Shao. **Tecnologia da Informação**. Fundação Getúlio Vargas, São Paulo. Disponível em: < <http://www16.fgv.br/rae/artigos/775.pdf>>. Acessado em: abril de 2011.

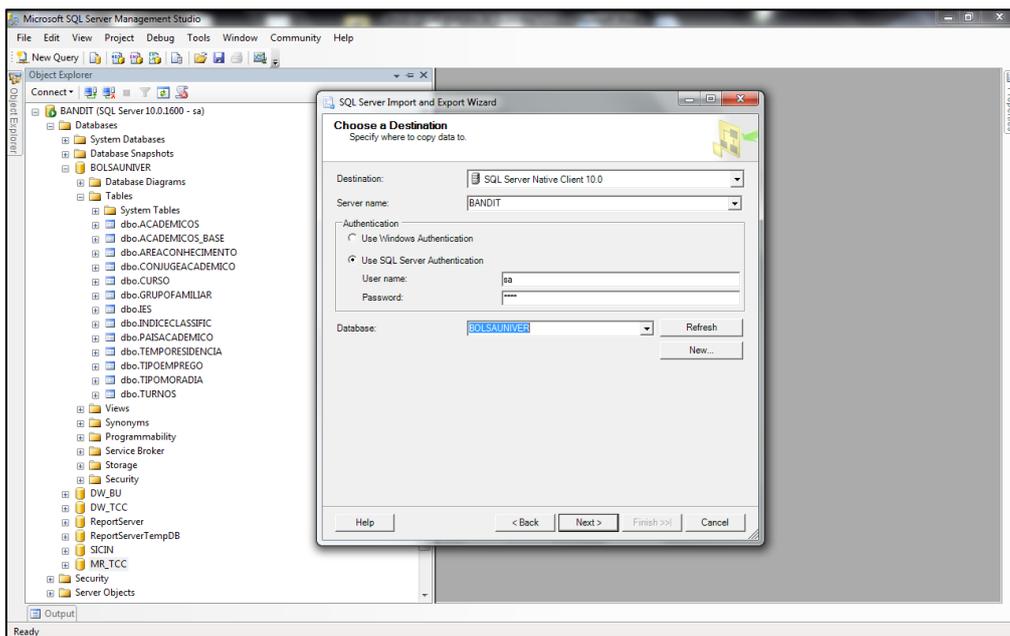
ZAJACZOKOWSKI, Fábio Luís. **A utilização de um “Data Warehouse” clínico no grupo hospitalar Conceição**. Dissertação (Trabalho de Conclusão de Curso de SI). Fundação Oswaldo Cruz. Porto Alegre – RS, 2005. Disponível em:< <http://157.86.8.50/bitstream/icict/1599/1/Trabalho%20final%20%20F%C3%A1bio.pdf>> Acessado em maio de 2011.

ZILKOSKI, Luís Cláudio Chaves. **Coleta de requisitos e modelagem de dados para Data Warehouse**: um estudo de caso utilizando técnicas de aquisição de conhecimento. Universidade Federal do Rio Grande do Sul. Porto Alegre – RS, 2003. Disponível em: < http://www.inf.ufrgs.br/bdi/administrator/components/com_jresearch/files/publications/Monografia-LuisZiulkoski.pdf> Acessado em: maio de 2011.

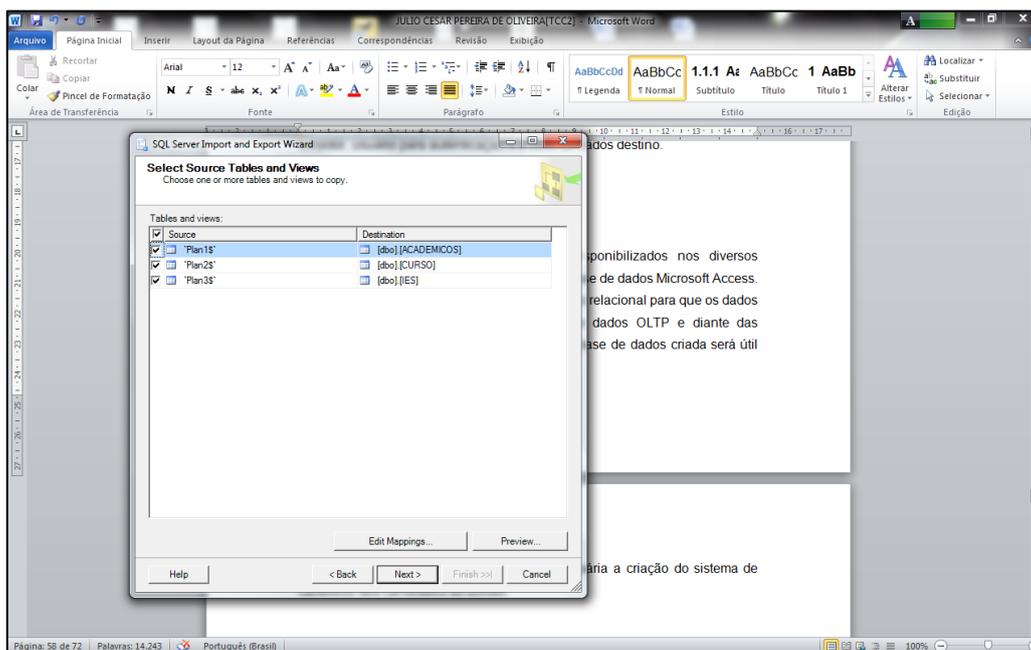
ANEXO I



Conforme apresentado na upracitada, na opção *Data Source*, é necessário selecionar o tipo de fonte de dados para a importação dos mesmos, neste caso, o arquivo *BolsaUniversitaria.xlsx*. A próxima operação é configurar a base de dados que receberá o conteúdo, nesse caso, a *BolsaUniver*, conforme apresentada na **Erro! Fonte de referência não encontrada.**a seguir.



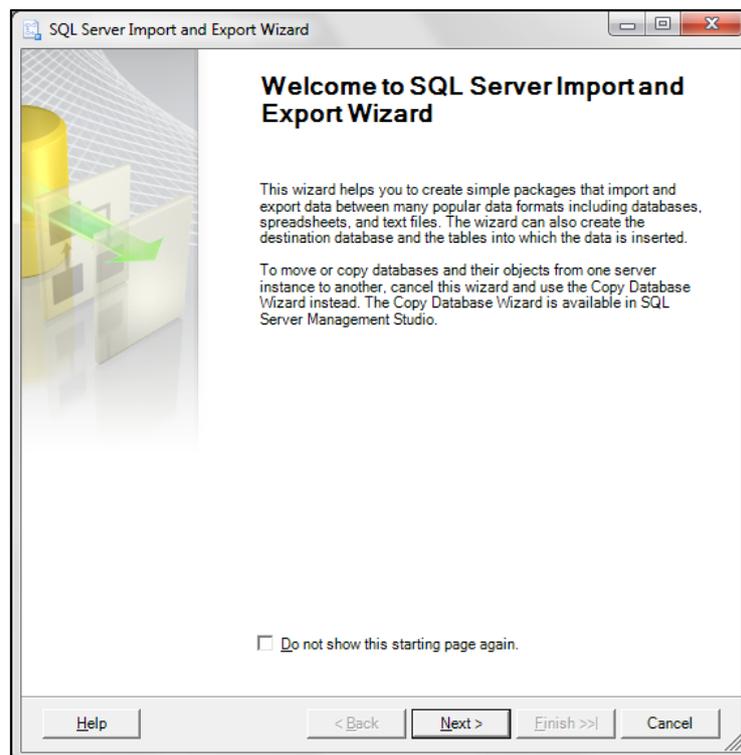
Tendo sido selecionado o *Data Source*, avançando no processo, são apresentadas as opções, onde se seleciona o tipo de fonte de dados destino, a instância do servidor, usuário para autenticação e a base de dados de destino. Essa operação permite relacionar a fonte de origem e a base de destino, além de permitir a especificação do tipo dos dados, conforme apresentado na Figura a seguir.



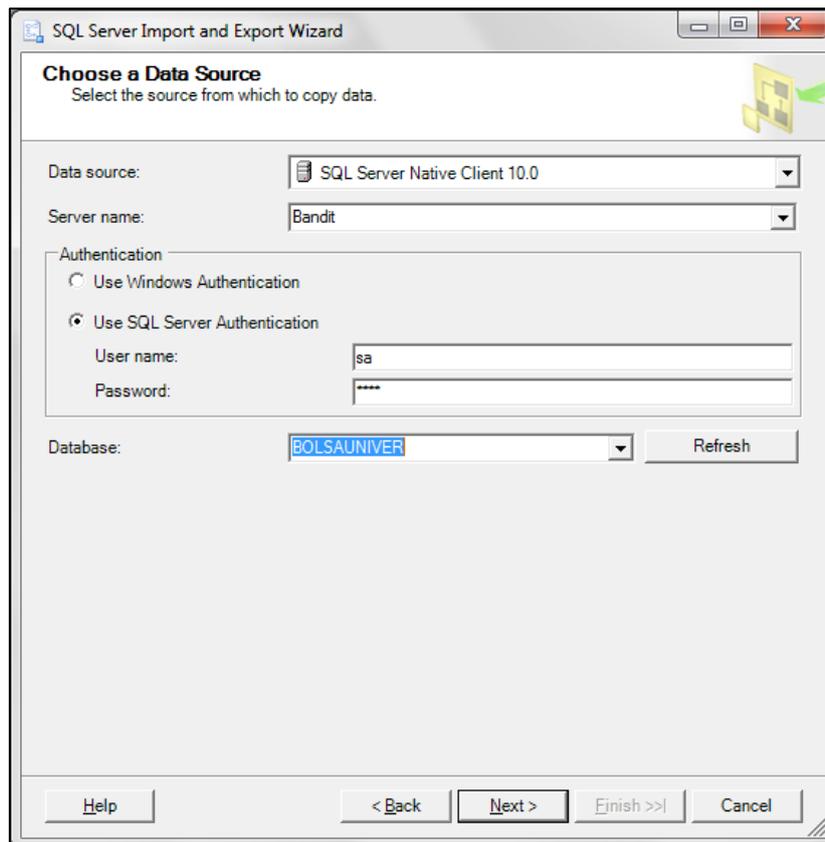
Na **Erro! Fonte de referência não encontrada**.são apresentadas as opções de planilhas do arquivo e as tabelas as quais se deseja armazenar os respectivos dados. Assim sendo, os dados das planilhas foram enviados para a base *BOLSAUNIVER (BOLSAUNIVER.mdb)*, que ainda não é o ambiente de DW, mas a base de dados construída para organização dos dados e construção da relacionamento entre os mesmos. Tendo sido feita a relação entre planilhas e tabela da base de dados, finaliza-se o processo (*finish*), e encerra o processo de envio dos dados. Esse processo foi novamente executado para a importação dos dados contidos em arquivo do *Microsoft Access*.

Importando dados para a base do *Data Warehouse*

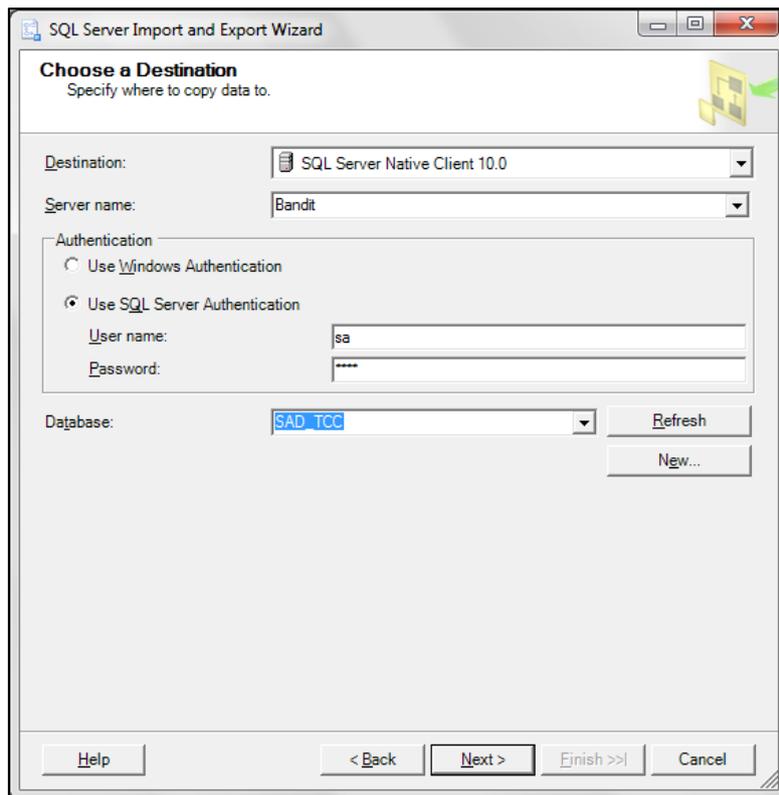
O fato de ter sido criado o modelo relacional referente ao programa de governo facilitou tanto a criação do modelo dimensional, quanto à alimentação da base de dados do *Data Warehouse*. Para tanto, foi utilizado o Microsoft DTS, do Visual Studio 2008. Com esse recurso, os dados foram transferidos de uma base para outra atendendo a estrutura projetada e somente os dados úteis à tabela destino. Na **Erro! Fonte de referência não encontrada.**a seguir é apresentado o início do processo de transferência dos dados utilizando o DTS.



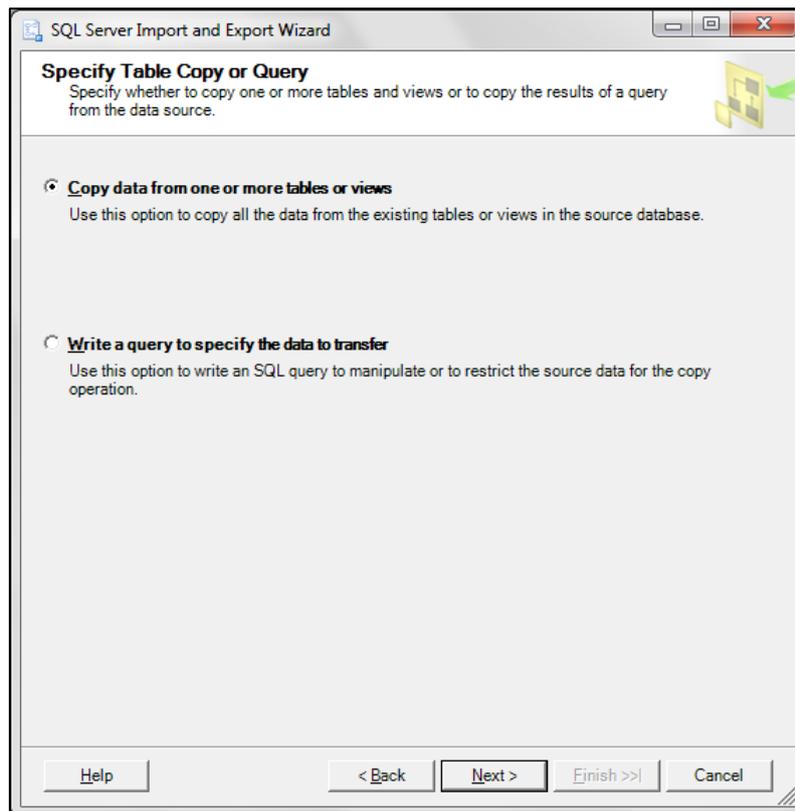
Na **Erro! Fonte de referência não encontrada.** foi apresentada a tela de abertura do Microsoft DTS que é acessado por meio do recurso *Import and Export Data* disponível no *Microsoft SQL Server 2008*. Avançando no processo, são apresentadas as opções conforme a próxima Figura.



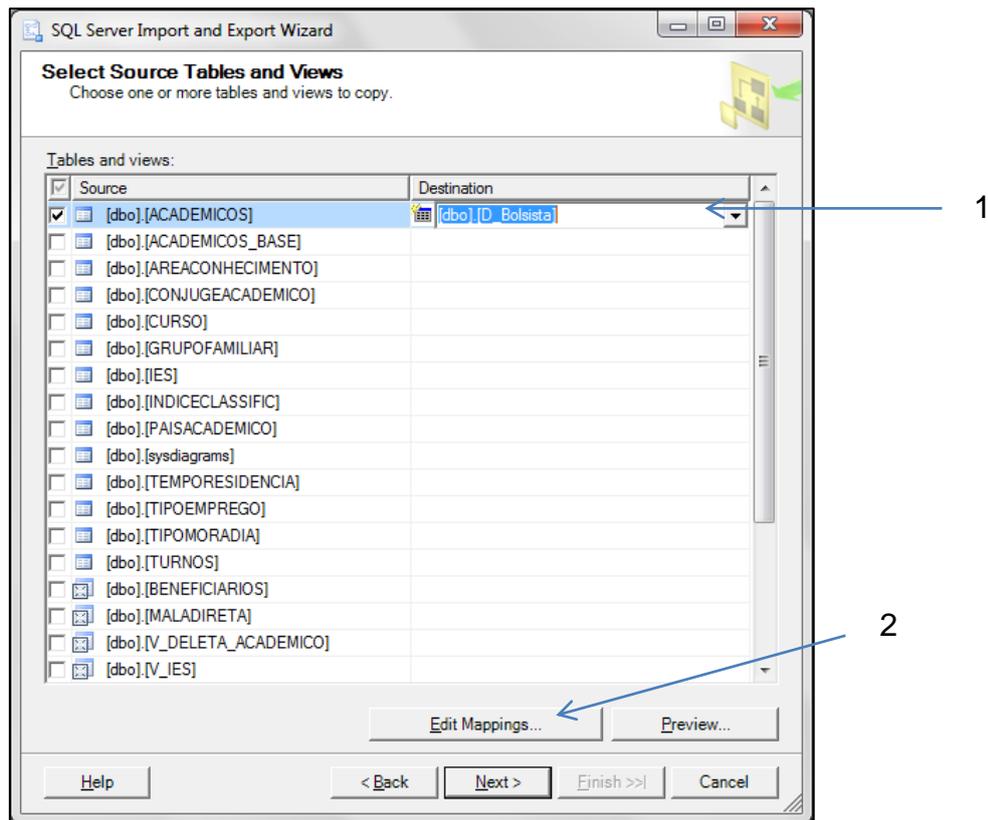
Conforme apresentado na **Erro! Fonte de referência não encontrada.**, para o início do processo de importação dos dados para configuração da fonte de dados origem é necessário preencher os dados configurando o *Data Source* (fonte de dados) e o *Server name* (nome do servidor), no campo *Authentication* define-se o usuário e senha para efetuar a operação, sugere-se utilizar o usuário “sa” e por fim, na opção *Database* (base de dados), seleciona-se a base de dados origem para transferência. Avançando no processo, é necessário realizar a configuração para a base de dados destino, conforme apresentado **Erro! Fonte de referência não encontrada.** a seguir.



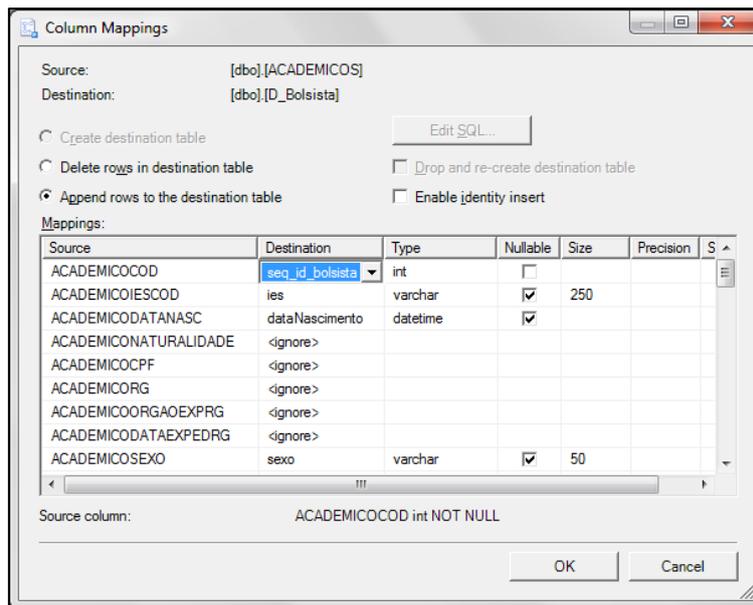
São apresentadas as opções para configuração da base de dados destino para a importação dos dados. As opções são as mesmas da opção origem e foram preenchidos da mesma forma com os mesmos dados, sendo que, o que difere neste caso é na opção *Database*, onde a base de dados selecionada é a destino. Ao avançar no processo apresenta-se a tela conforme próxima **Erro! Fonte de referência não encontrada.** com as opções para cópia para uma ou mais tabela (1ª opção), ou utilizando uma *query SQL* (2ª opção).



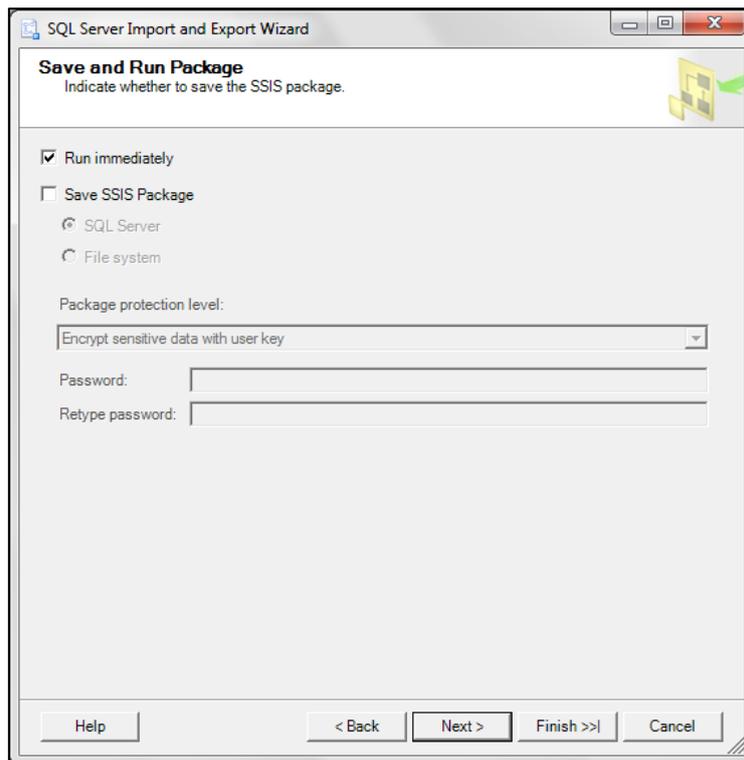
São apresentados os tipos de transferências de dados disponíveis para realização. Seleciona-se a opção de cópia de dados para uma ou mais tabelas. Essa opção foi possível pelo fato da base de dados origem estar com os dados organizados para serem inseridos na base do DW. Assim sendo, no passo seguinte são realizados os mapeamentos de tabelas origem e destino para importação dos dados, conforme **Erro! Fonte de referência não encontrada.** que se segue.



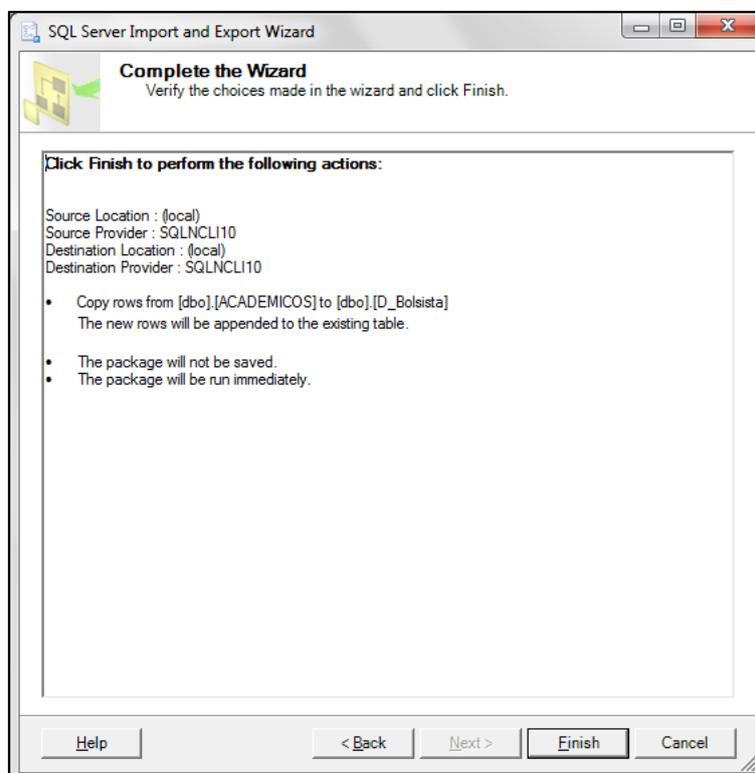
Na **Erro! Fonte de referência não encontrada**.são apresentadas as opções de mapeamento entre tabela origem e destino. Neste caso, a tabela origem “Academicos” estará importando os dados para a tabela destino “D_Bolsista”, marcados pela seta 1. Na opção da seta 2, são realizados os mapeamentos entre as colunas da tabela, que ao selecionar, são apresentadas as opções conforme a **Erro! Fonte de referência não encontrada**.que se segue.



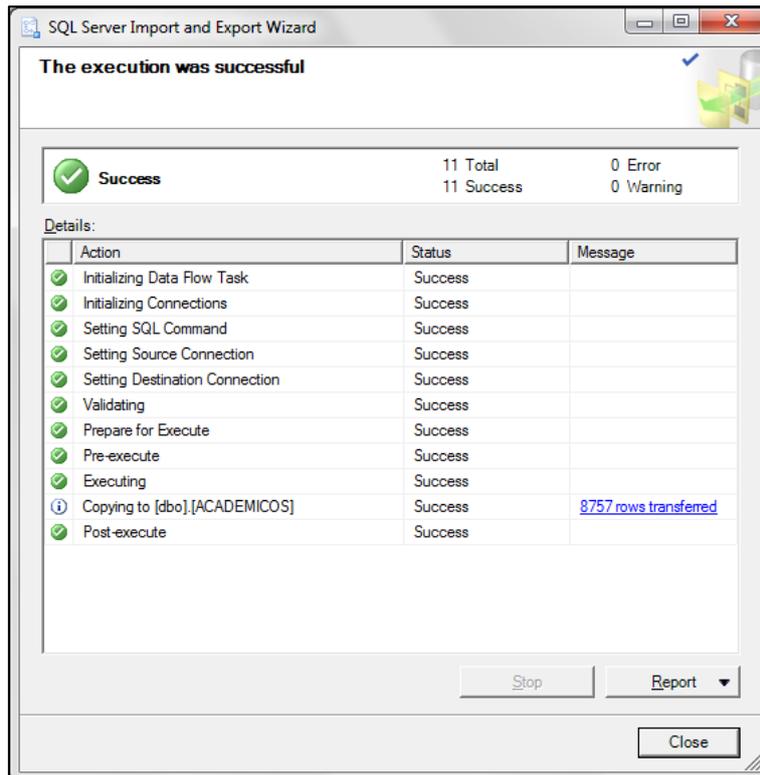
A seguir, são apresentadas as opções de mapeamento entre colunas. Na coluna *Source* são apresentadas as colunas da tabela origem e na opção *Destination* as opções de colunas para a tabela de destino dos dados. Nesta operação, são importados os dados da tabela “Academicos” para a tabela “D_Bolsista”. Tendo sido selecionados os dados úteis no processo, os dados considerados não úteis (não selecionados) sejam ignorados. Tendo sido finalizado o processo de mapeamento entre as colunas, seleciona-se “Ok”. O próximo passo é a seleção do momento de execução da operação, podendo ser realizado imediatamente ou em um projeto SSIS e salvo em um pacote. Conforme as necessidades desse projeto, as atividades serão executadas imediatamente.



Estando selecionada a opção para execução imediata, o próximo passo é a finalização do processo.



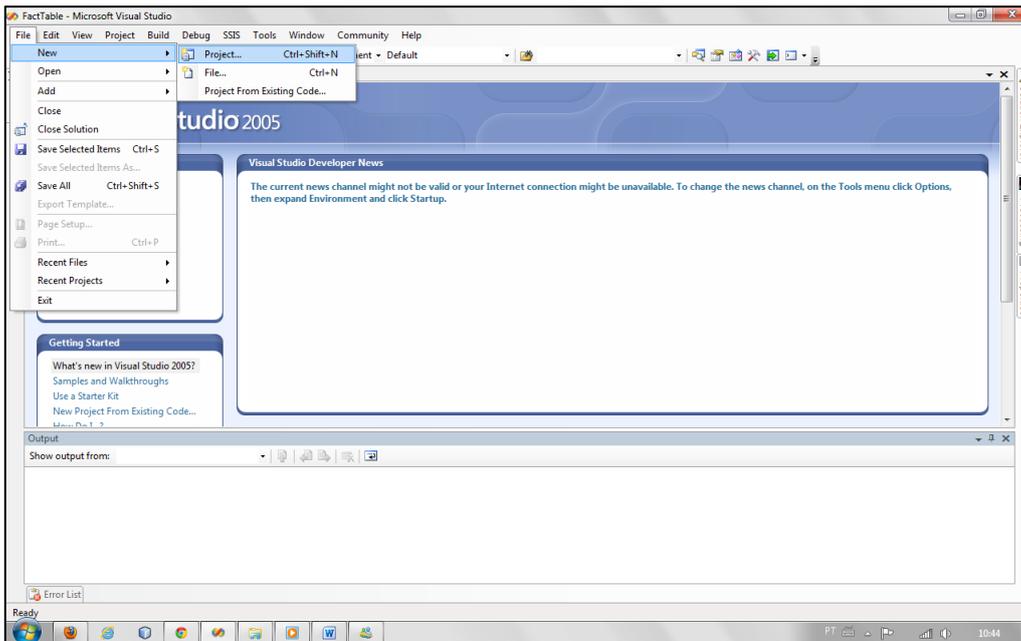
Tendo sido finalizado o processo, é executada a importação dos dados configurados no processo do DTS.



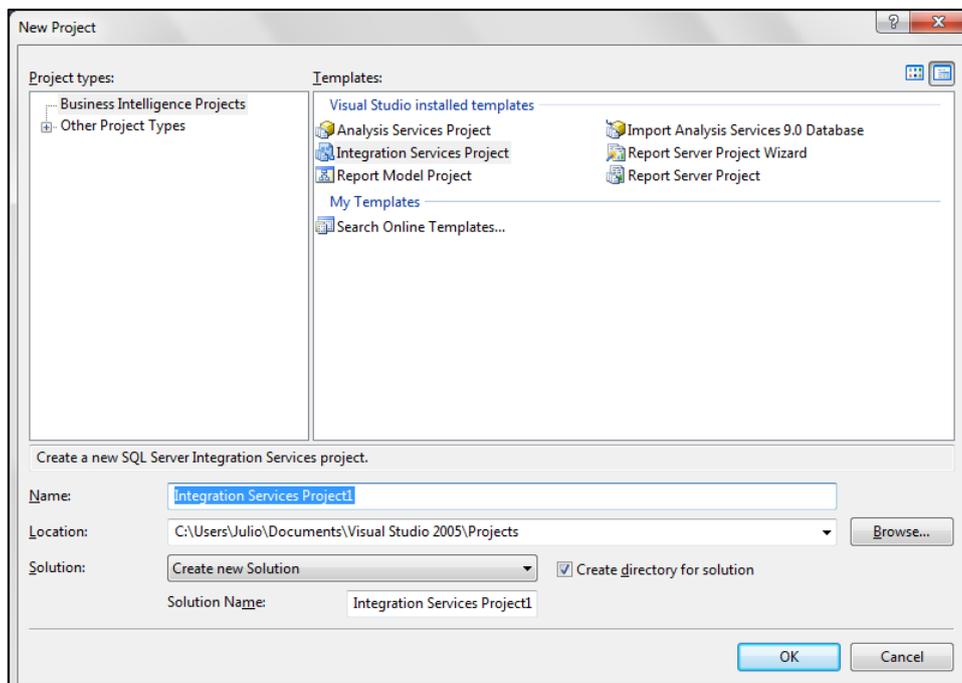
De acordo com o apresentado na **Erro! Fonte de referência não encontrada.**, para a tabela “*D_Bolsista*”, selecionada como tabela destino, foram copiados 8757 registros da tabela “*Academicos*”. Esse processo foi realizado para todas as tabelas dimensões do DW. Assim sendo, estando às tabelas dimensões do *Data Warehouse* alimentadas, o próximo passo foi a inserção dos dados na tabela de fatos.

Inserção dos dados na Tabela Fatos

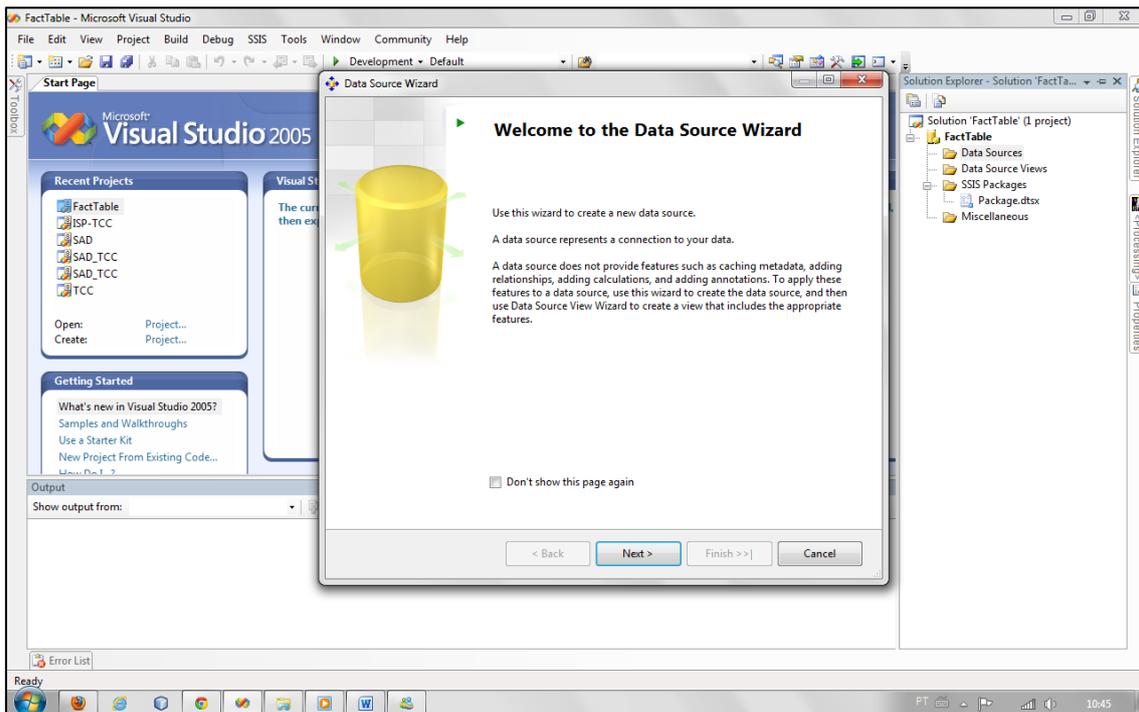
Para a criação e importação dos dados na tabela fatos é necessária a criação de um projeto *Integration Services*. Essa necessidade é devida as características da tabela fatos, que contém as chaves primárias das tabelas dimensões e a inserção desses dados precisa ser realizada de uma forma integral em um processo único. Neste caso, a tabela fatos com os respectivos relacionamentos foi criada no *SQL Server*, esta sequência demonstra a carga dos dados.



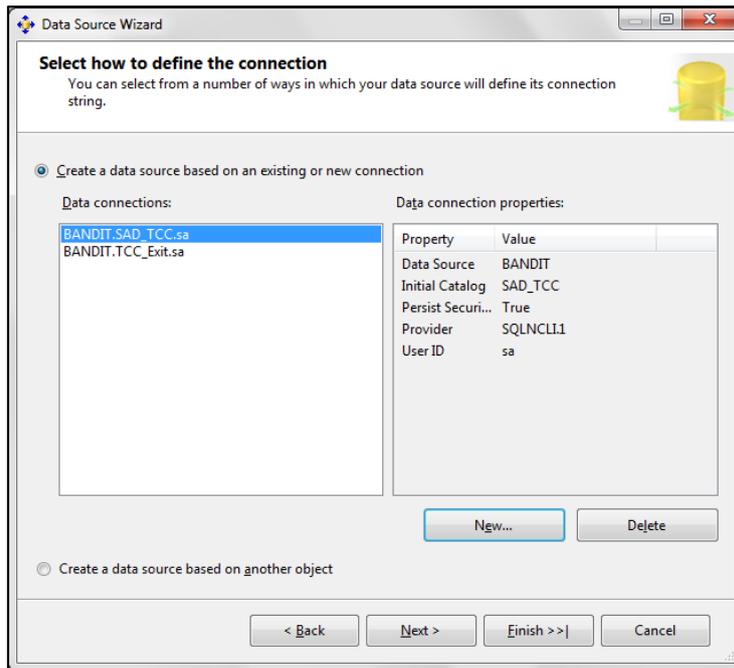
Na **Erro! Fonte de referência não encontrada.** é apresentado o início do processo, onde é criado o projeto no *SQL Server Business Intelligence Development Studio*. Na opção *file*, seleciona-se *New* e posteriormente, *Project*, para a criação de um novo projeto. A seguir, é apresentada a opção de projeto a ser criado para essa operação.



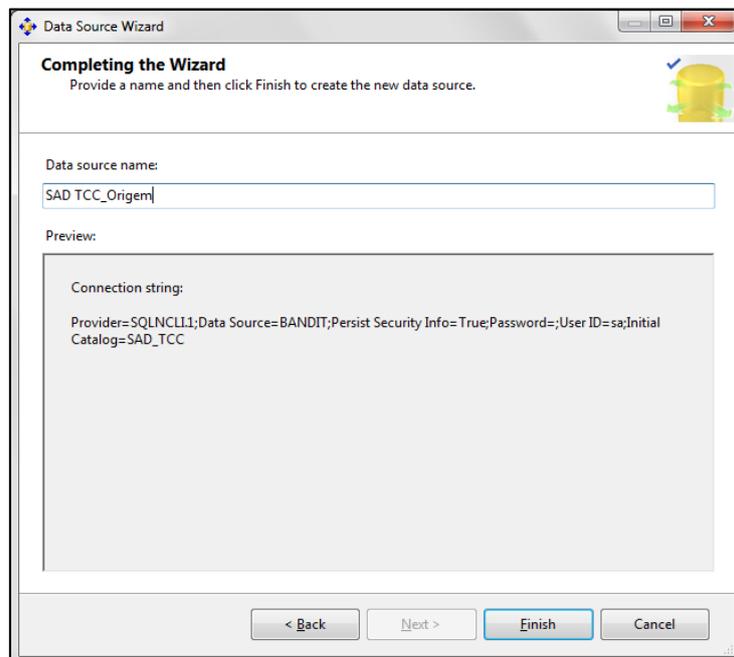
Tendo sido selecionada a opção de criação de novo projeto, é apresentada a tela, onde são apresentadas as opções de projeto. Para esse projeto, seleciona-se a opção *Integration Services Project* e abaixo estão as opções para o nome do projeto e local de armazenamento. A seguir, é apresentado o processo da configuração da fonte de dados para o processo de carga dos dados.



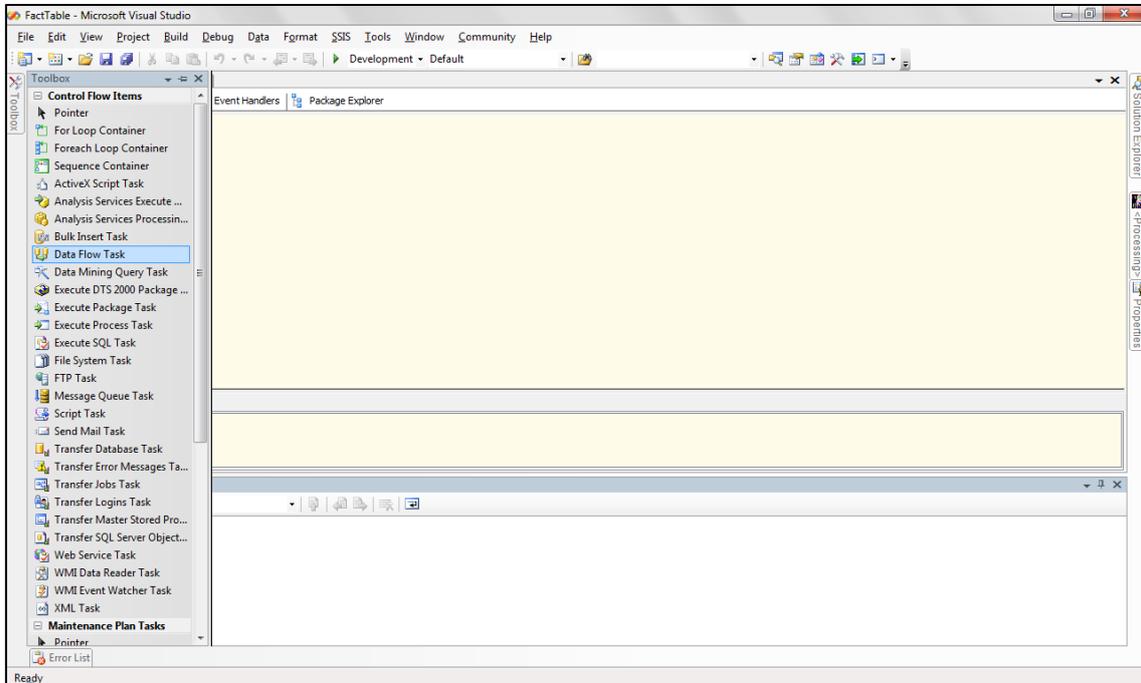
Na **Erro! Fonte de referência não encontrada.** é apresentado o início do processo de configuração da fonte de dados para carga. Para tanto, a partir da aba *Solution Explorer*, clicando com o botão direito sobre a pasta *Data Source* e solicitando a criação de um *New Data Source*. A seguir, é apresentado o processo de configuração do *Data Source*.



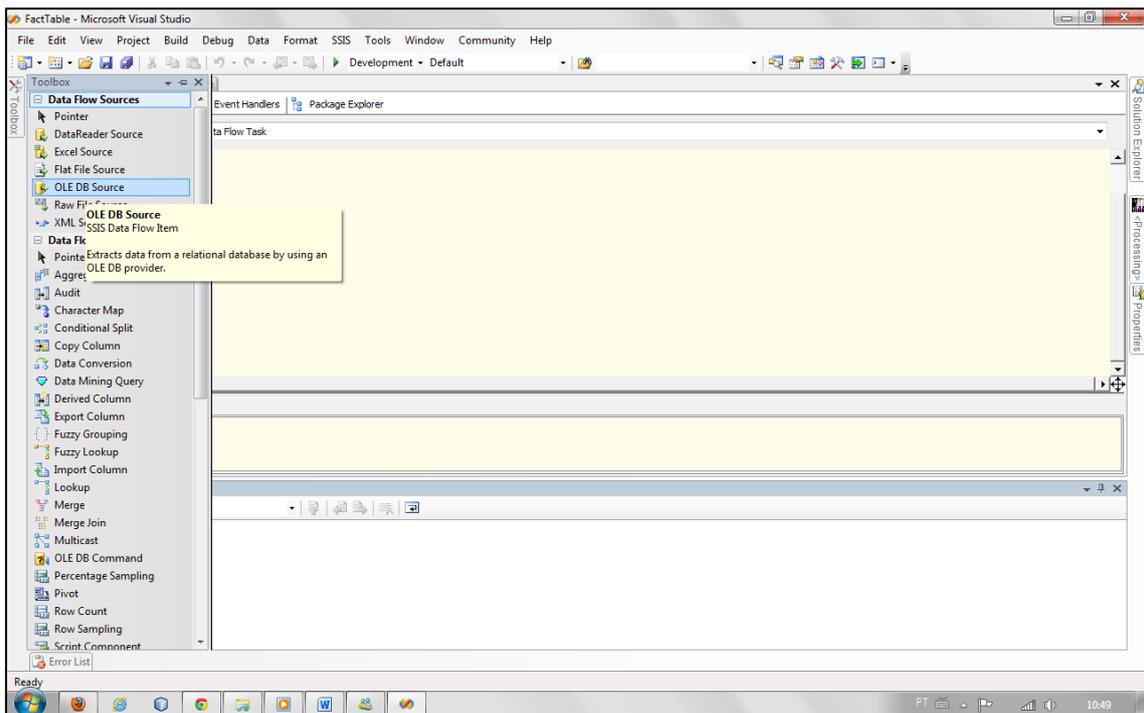
Na configuração do *Data Source*, deve ser selecionada as fontes de dados origem e destino. Caso as bases não se encontrem listadas já nas conexões, é necessário realizar a configuração a partir da opção *New*. Este projeto já possui as fontes configuradas e, por isso, foi necessário somente selecionar a fonte. Assim sendo, selecionada a fonte, é necessário nomear a fonte e finalizar o processo.



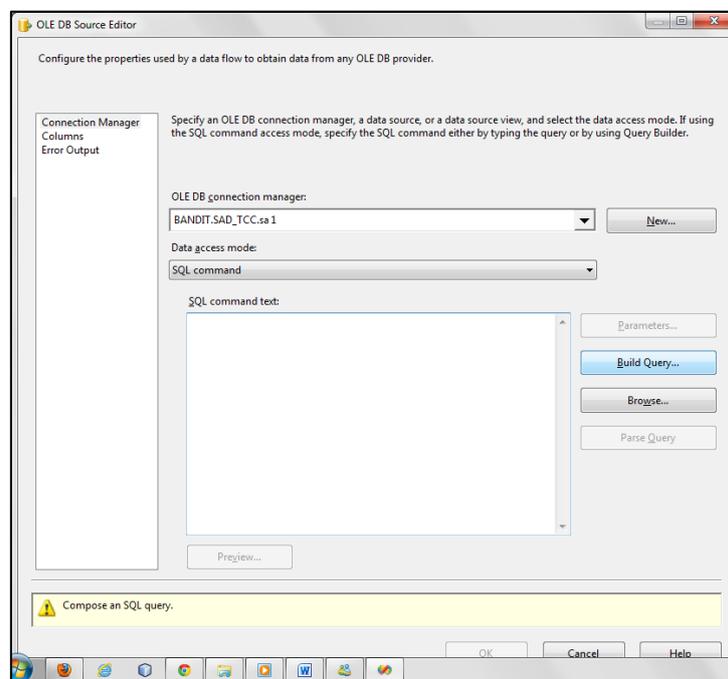
É disponibilizada a opção de nomear o *Data Source* para então finalizar o processo de criação. A seguir, é apresentado o início do processo de adição dos componentes para inserção dos dados.



Para iniciar o processo de inclusão dos componentes é necessário acessar a aba *Solution Explorer* e abrir o arquivo *Package.dtsx* e, em seguida, na aba *Toolbox* inserir um componente *Data Flow Task*, conforme apresentado na **Erro! Fonte de referência não encontrada..** A seguir, é apresentado o passo seguinte, a adição do componente *OLE DB Source*.

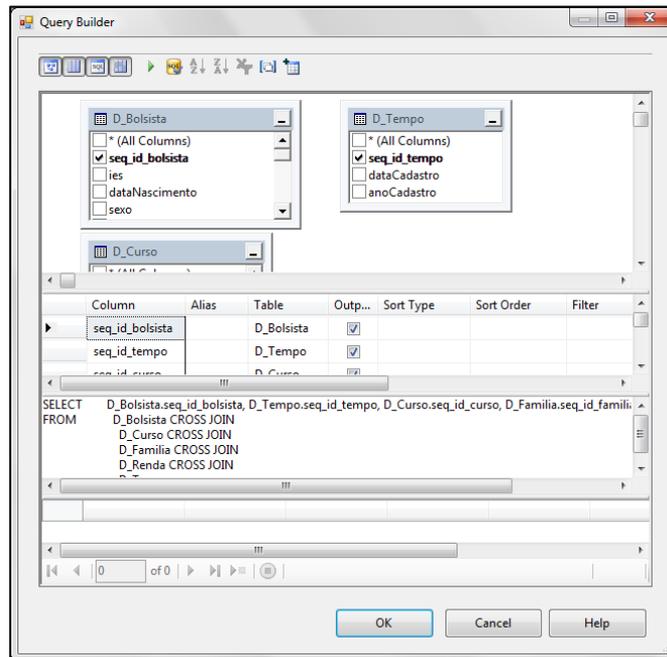


Tendo acessado o componente *Data Flow Task*, seja por meio de um clique duplo ou pela barra superior, insere-se na área deste componente outro componente da barra *Tolbox*, o *OLE DB Source*.

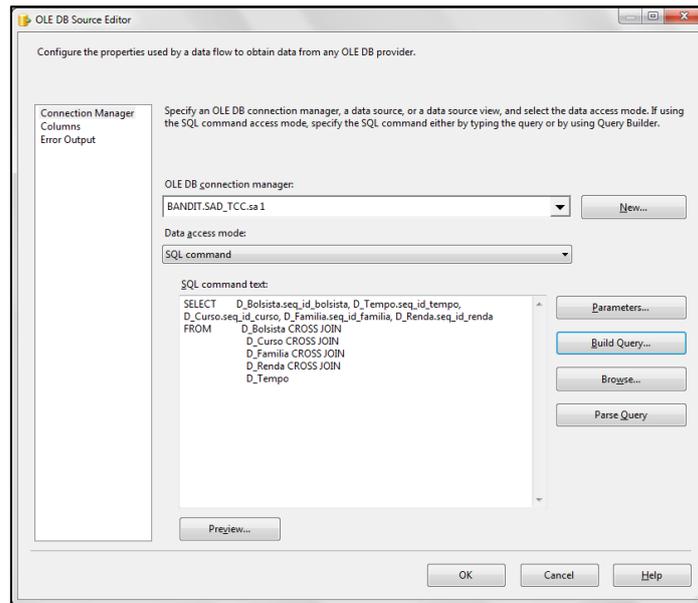


Para configurar esse componente é necessário selecionar a base de dados na opção *OLE DB Connection Manager* e na opção *Access Mode*.

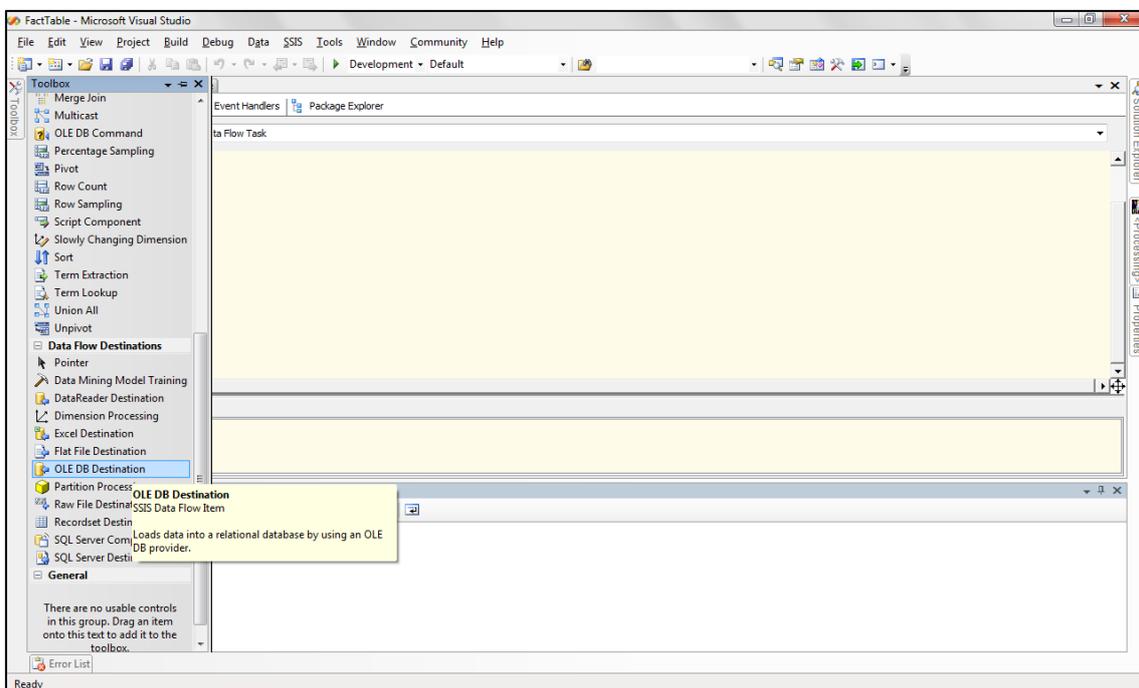
Então, a partir do botão *Build Query*, é realizada a seleção das tabelas e das colunas a serem utilizados, conforme apresentado a seguir.



Na **Erro! Fonte de referência não encontrada.** é apresentada a sequência de seleção das colunas das tabelas para inserção. Assim, é possível selecionar as colunas e, logo abaixo, os campos são relacionados e, ainda, é apresentada a *Query SQL* gerada a partir desta seleção. A seguir, é apresentado o preenchimento da opção *SQL Command* com a *Query SQL* gerada a partir da seleção nas tabelas.

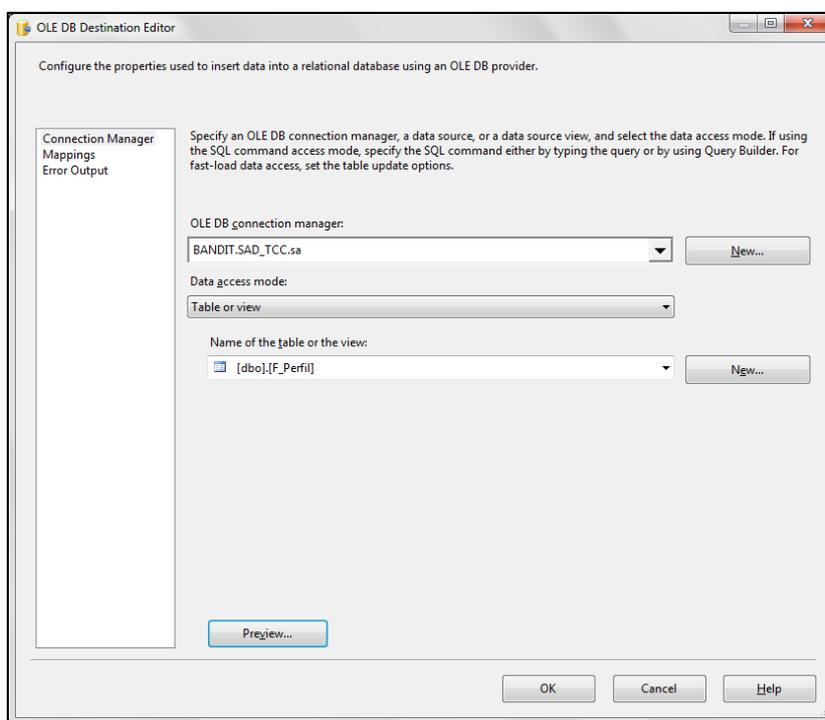


Tendo sido selecionadas as colunas das tabelas do *Data Source*, a partir dessa configuração, para a sequência é clicar em “ok” e finalizar a configuração do componente *OLE DB Source*. A seguir, é apresentada a inserção do componente *OLE DB Destination*.

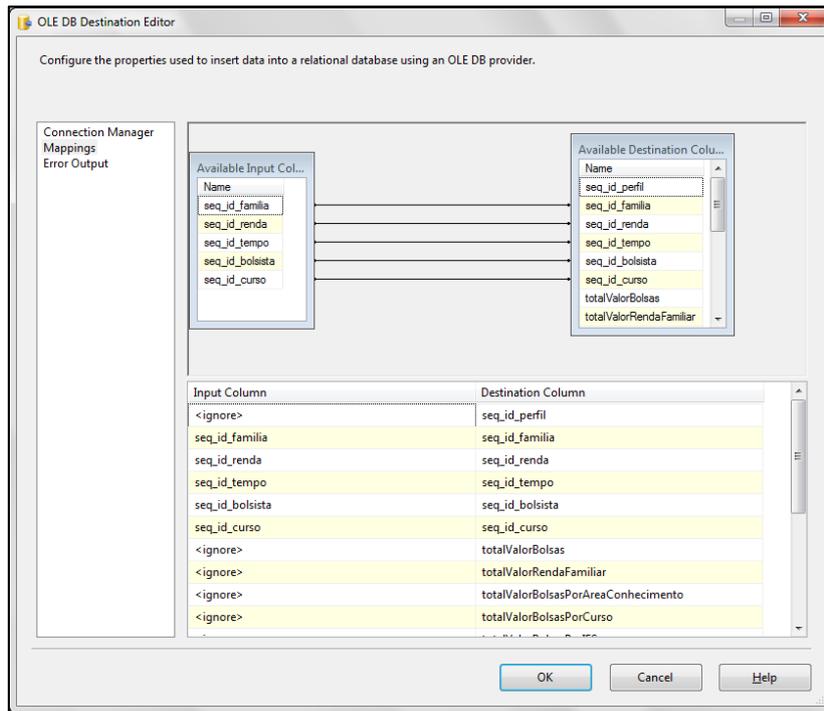


Na **Erro! Fonte de referência não encontrada.** é apresentado o processo de inserção de outro componente que vai ser adicionado ao projeto e, por meio de um relacionamento, vai ser configurado o destino dos dados

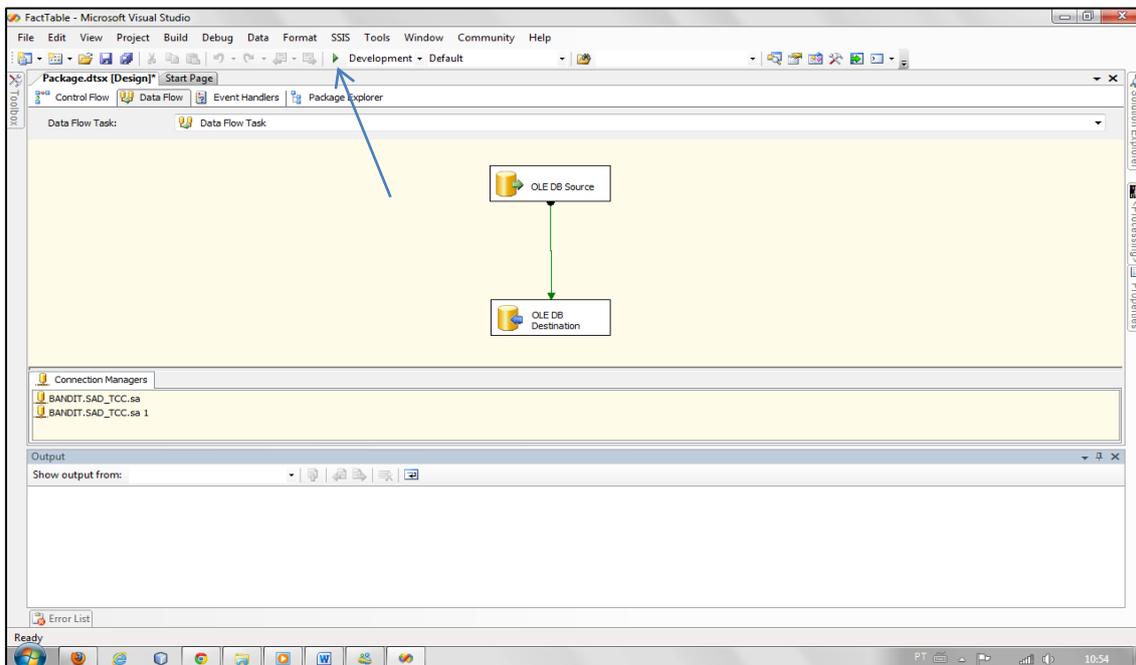
selecionados na componente *OLE DB Source*. Neste caso, será inserido o componente *OLE DB Destination*. Inserido o componente, a partir de um clique duplo, são disponibilizados as opções de configuração desse componente.



Na **Erro! Fonte de referência não encontrada.**, são apresentadas as opções de configuração do componente *OLE DB Destination*. Inicialmente, é necessário selecionar a base de dados que possui a tabela destino dos dados selecionados e configurados no componente *OLE DB Source*. Na opção *Data Access Mode*, por meio do tipo *Table or View*, pois a inserção vai ser realizada em uma tabela única e esta forma vai ser possível realizar a configuração por seleção da tabela, na opção posterior, *Name of the table or the view*. Então, é necessário realizar ainda o mapeamento das colunas com os campos, para tanto, é necessário definir a opção *mappings* e será disponibilizada a tabela construída a partir dos dados selecionados como origem e a tabela destino selecionada.

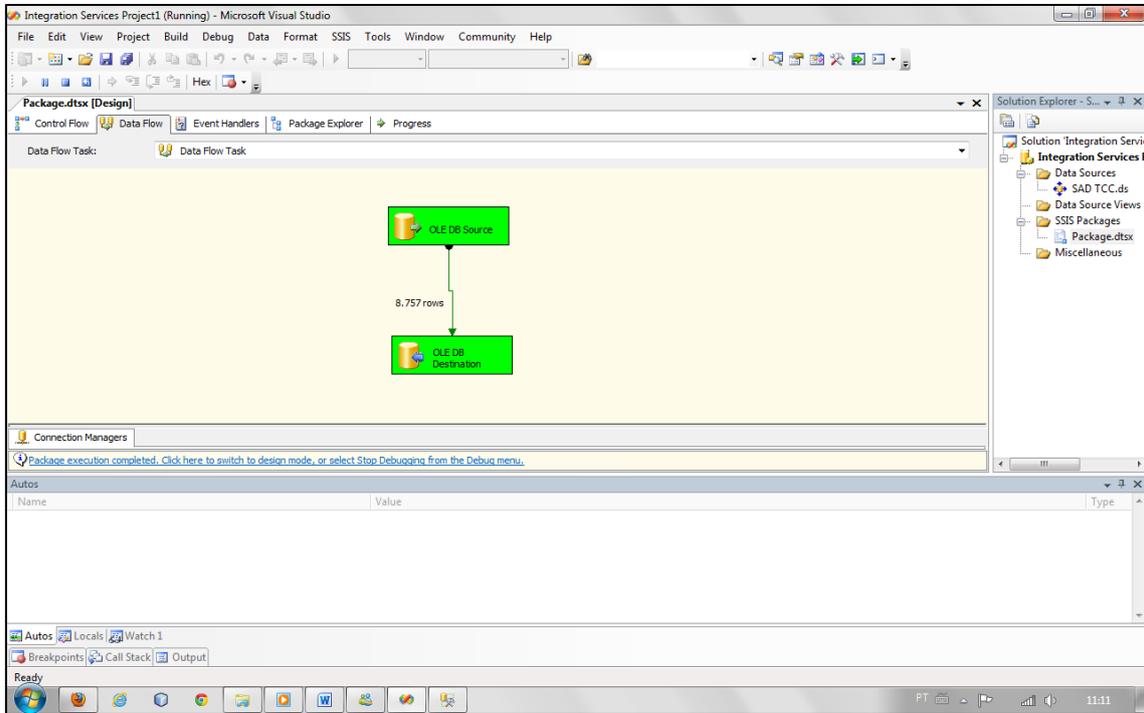


Na **Erro! Fonte de referência não encontrada.** é apresentada as opções de mapeamento dos dados origem e destino definidos nos componentes. Tendo sido definidos os mapeamentos, por meio da ligação dos campos origem e destino, finaliza-se esse processo clicando na opção “ok”.



Na **Erro! Fonte de referência não encontrada.** é apresentada a opção de processamento do projeto. Para tanto, é necessário efetuar o comando de processamento do projeto, conforme apontado na figura. Então iniciará o

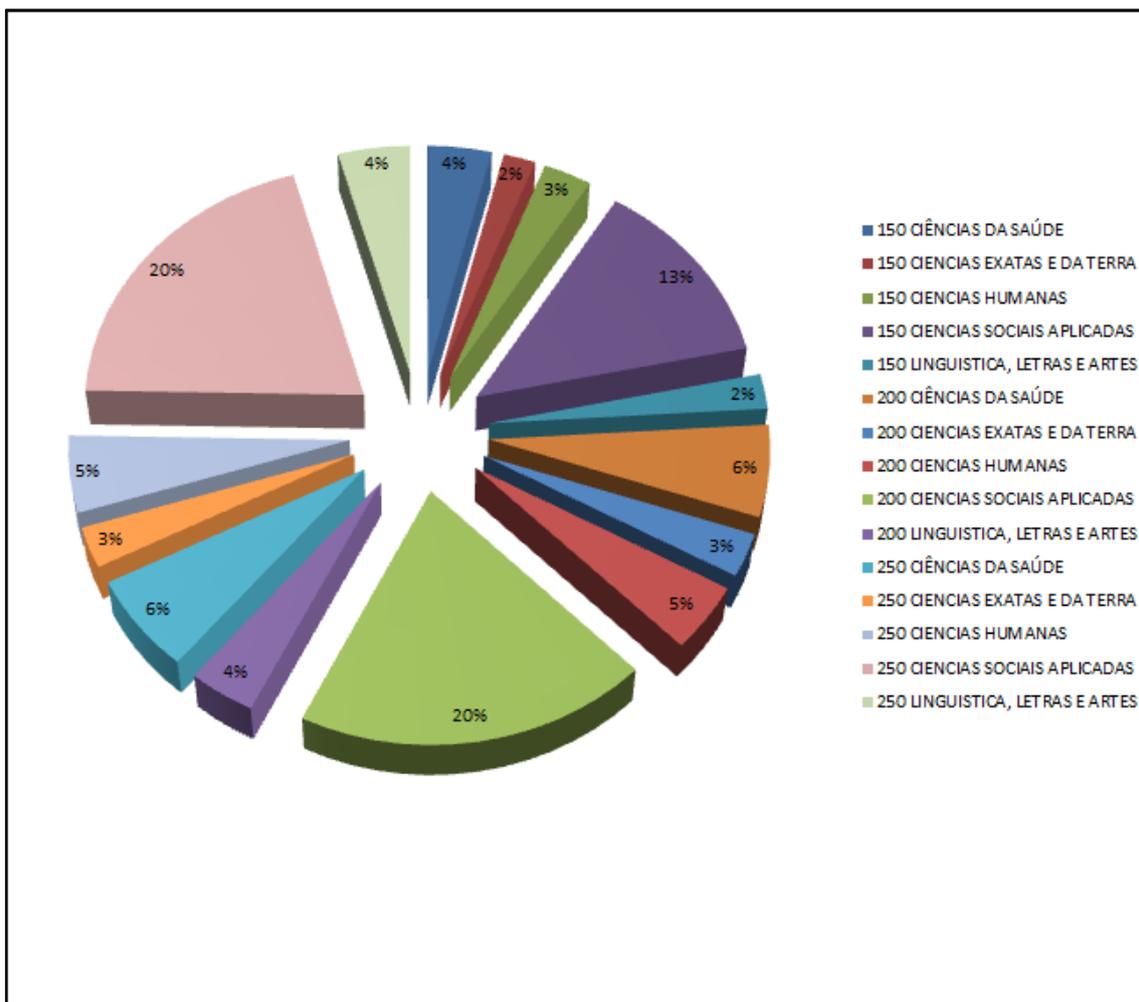
processo de carga dos dados e estando corretamente configurados e não havendo erros, ao término do processamento, os componentes configurados serão apresentados conforme **Erro! Fonte de referência não encontrada..**



Na **Erro! Fonte de referência não encontrada.** é apresentado o fim do processo de carga dos dados e entre os componentes, o total de linhas processadas e inseridas na tabela destino. Neste caso, foram inseridos 8757 linhas na tabela fatos do *Data Warehouse*.

ANEXO II

Os relatórios gráficos apresentados em anexo, servem para orientar a emissão dos relatórios. Assim sendo, serão apresentados relatórios resultantes do cruzamento de informações sobre o programa bolsa universitária para demonstração das diversas opções a disposição do gestor para extração de informações.

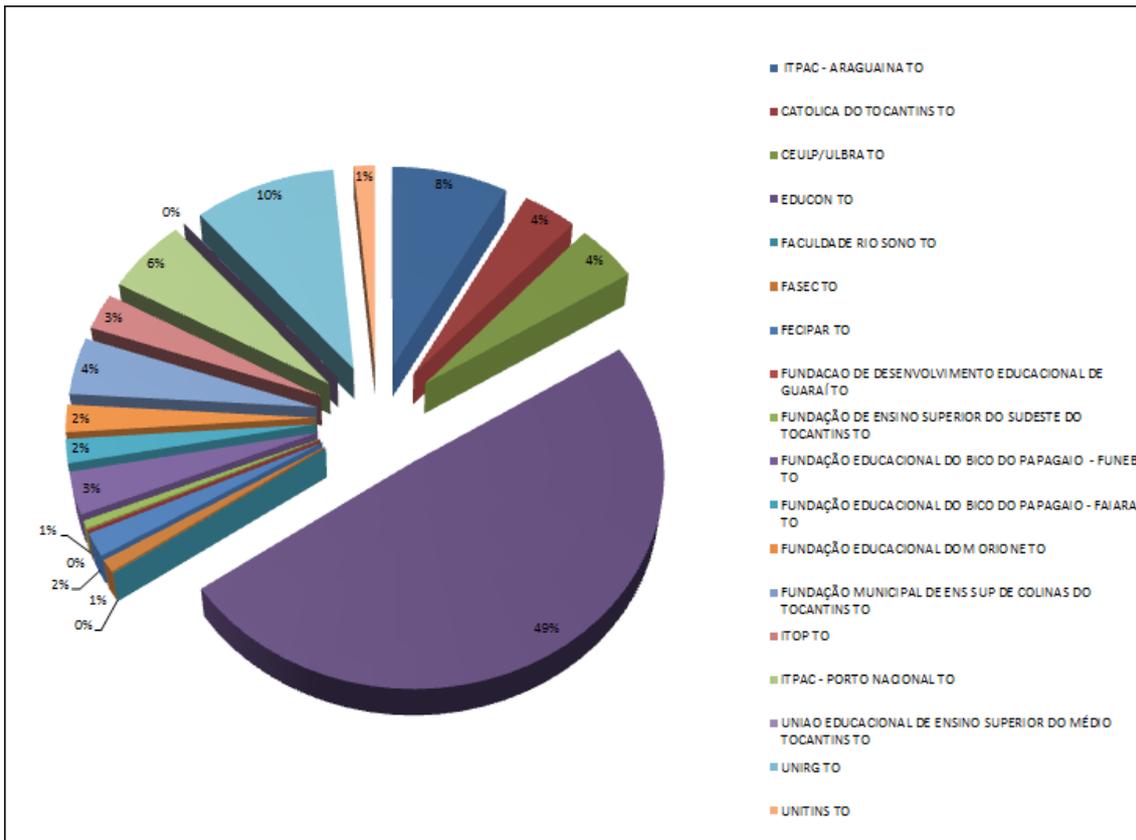


Valores:

1º – valor da bolsa

2º - área de conhecimento

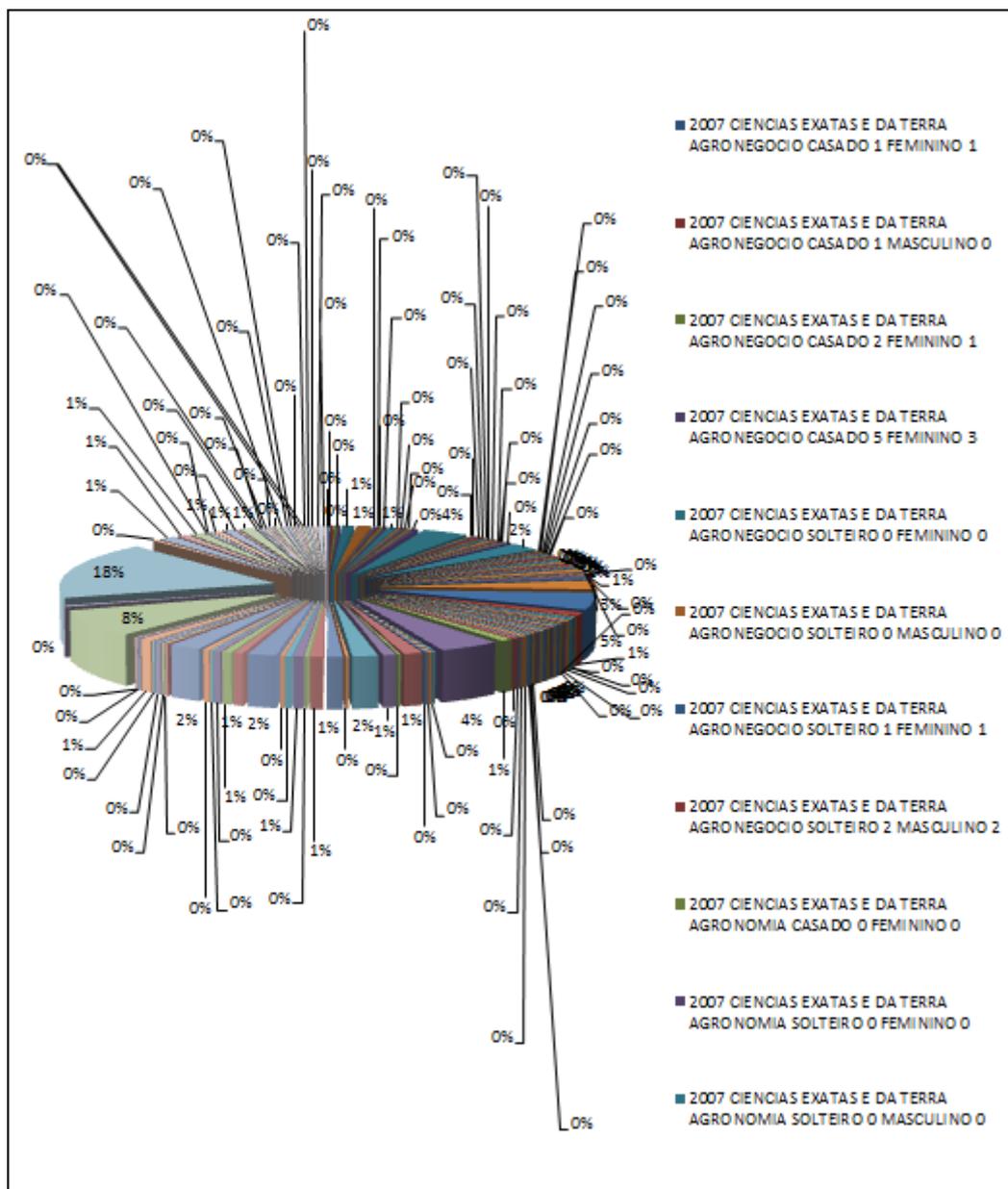
Exemplo: Referente à área das ciências da saúde com bolsistas no valor de R\$ 150,00, equivalem a 4% do valor total.



Valores:

1º – Percentual de bolsistas por IES

Exemplo: Referente às IES, 49% dos bolsistas equivale a Educon.



Valores:

1º – Ano de cadastro

2º - área de conhecimento

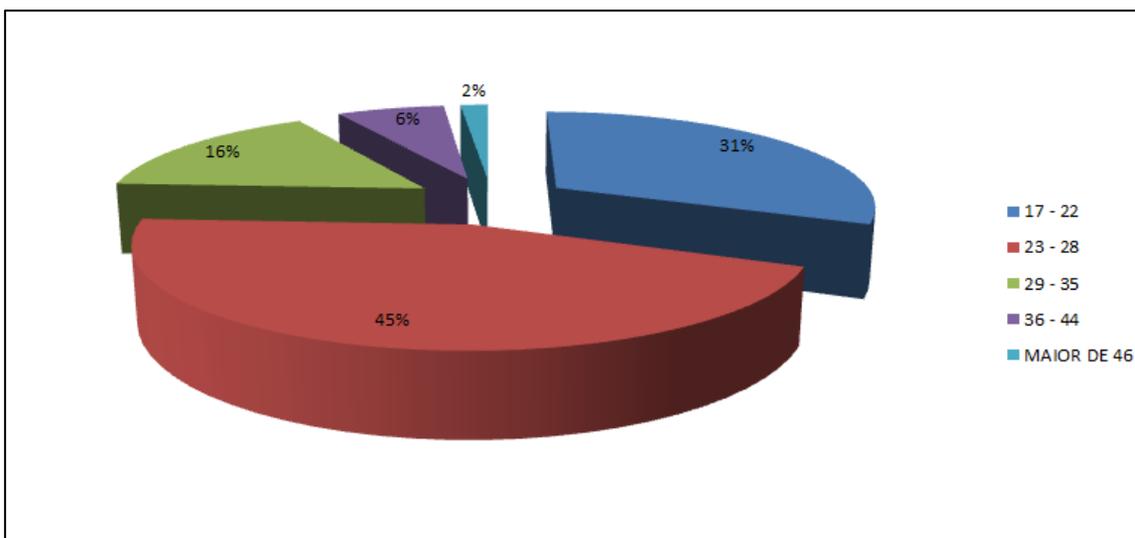
3º - Curso

4º - Estado civil

5º - Número de dependentes

6º - Sexo

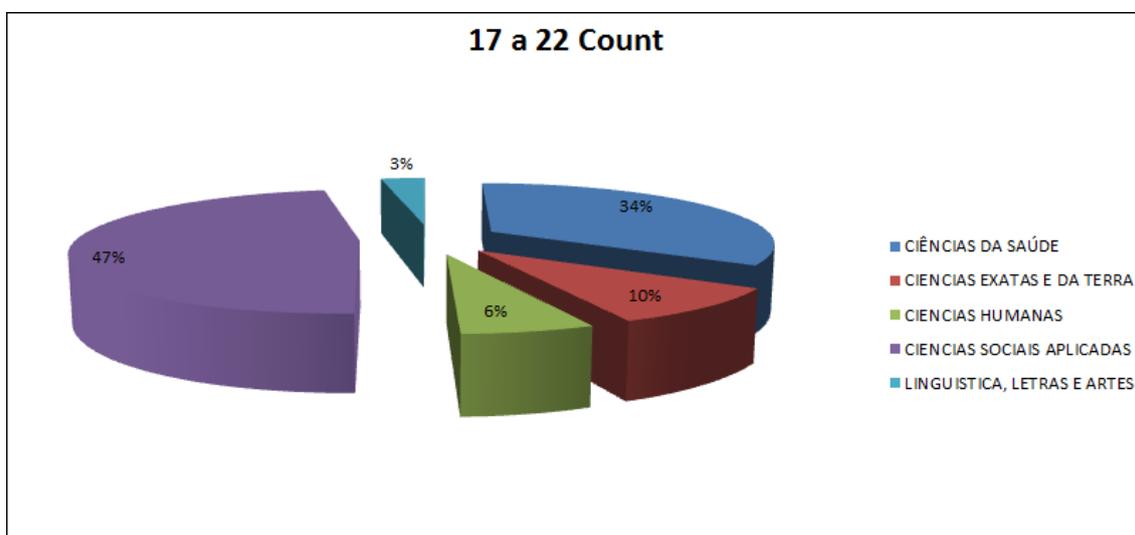
7º - Número de familiares (mora sozinho)



Valores:

1º – Faixa etária do total de bolsistas

Exemplo: Bolsistas da faixa etária entre 17 e 22 anos representam 31% do total dos bolsistas.

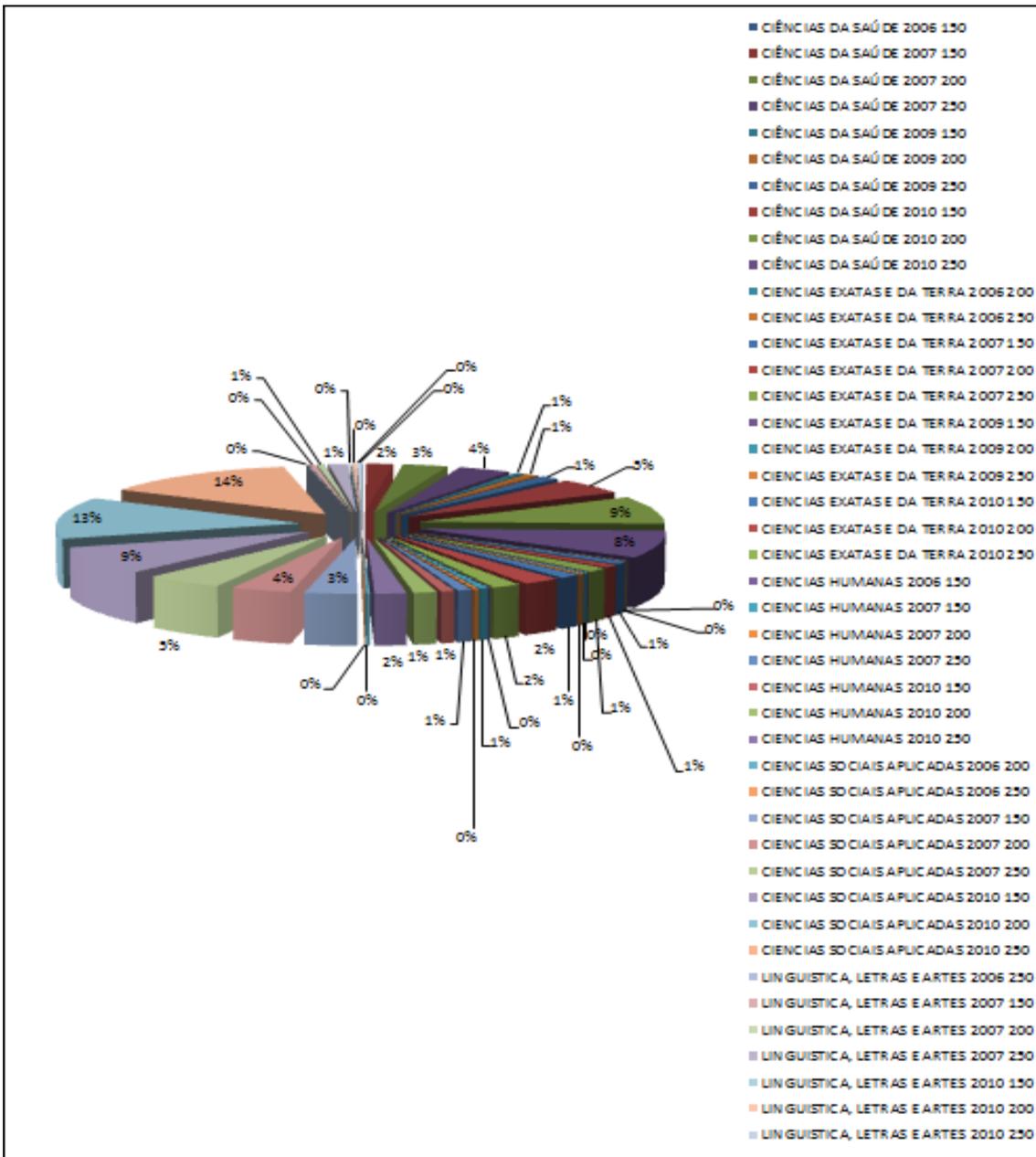


Valores:

1º – Classificação de bolsistas de faixa etária entre 17 e 22 anos

2º - área de conhecimento

Exemplo: Considerando os bolsistas da faixa etária entre 17 e 22 anos, da área das ciências da saúde, representam 34% do total dos bolsistas atendidos.



Valores:

1º – Classificação de bolsistas de faixa etária entre 17 e 22 anos

2º - área de conhecimento

3º - ano de cadastro

4º - valor da bolsa

Exemplo: Considerando os bolsistas da faixa etária entre 17 e 22 anos, da área das ciências sociais aplicadas, no ano de 2007, atendido por bolsa no valor de R\$ 150,00, representam 13% do total dos bolsistas atendidos.