



UNIVERSIDADE DO SUL DE SANTA CATARINA
BRUNA CORRÊA GOMES
VINICIUS FRITZEN DE OLIVEIRA

**OS CURSOS DA ÁREA DAS CIÊNCIAS DA COMPUTAÇÃO NO BRASIL:
UMA ANÁLISE EXPLORATÓRIA DE DADOS.**

Palhoça
2007

BRUNA CORRÊA GOMES
VINICIUS FRITZEN DE OLIVEIRA

**OS CURSOS DA ÁREA DAS CIÊNCIAS DA COMPUTAÇÃO NO BRASIL:
UMA ANÁLISE EXPLORATÓRIA DE DADOS.**

Trabalho de conclusão de curso apresentado ao Curso de graduação em Sistemas de Informação, da Universidade do Sul de Santa Catarina, como requisito parcial para obtenção do título de Bacharel.

Orientador: Prof^o. Dr. Aran Bey Tcholakian Morales

Palhoça
2007

**BRUNA CORRÊA GOMES
VINICIUS FRITZEN DE OLIVEIRA**

**OS CURSOS DA ÁREA DAS CIÊNCIAS DA COMPUTAÇÃO NO BRASIL:
UMA ANÁLISE EXPLORATÓRIA DE DADOS.**

Este trabalho de Conclusão de Curso foi julgado adequado à obtenção do título de Bacharel em Sistemas de Informação e aprovado em sua forma final pelo Curso de Sistemas de Informação, da Universidade do Sul de Santa Catarina.

_____, ____ de _____ de 20____.
Local dia mês ano

Prof. e Orientador Aran Bey Tcholakian Morales, Dr.
Universidade do Sul de Santa Catarina

Prof. e Coordenador Mauro Madeira.
Universidade do Sul de Santa Catarina

Prof^ª. Maria Inés Castiñeira, Dra.
Universidade do Sul de Santa Catarina

Dedico este trabalho aos meus pais Alberto e Raquel, aos meus irmãos Filipe e Artur, a minha avó Helena, aos meus tios e primos, pelo incentivo, cooperação e apoio em todos os momentos desta importante etapa da minha vida.

Bruna

Dedico este trabalho aos meus pais, porque sem eles nada disso seria possível, por me ensinarem a ter caráter, humildade, amor, carinho e respeito, gestos e valores cada vez mais raros de se encontrar, e principalmente, por sempre acreditarem que eu era capaz de vencer, tanto esta, como todas as outras caminhadas que já fiz, e que ainda farei. À minha namorada pelo apoio e compreensão, meu irmão pelos sempre bons e alegres momentos que me oferece, e a todos os meus familiares, amigos e colegas, que direta ou indiretamente, ajudaram-me a dar mais esse passo na vida.

Vinicius

AGRADECIMENTOS

Ao nosso orientador Prof^o Dr. Aran Bey Tcholakian Morales pela sua dedicação, disponibilidade, interesse e orientação. À nossa coordenadora do projeto de conclusão de curso Prof^a Maria Inés Castiñeira. E a todos que, direta ou indiretamente, colaboraram para que este trabalho fosse realizado.

Agradeço primeiramente a Deus, pois sem Ele, nada seria possível. Agradeço ao meu colega Vinicius Oliveira pela sua dedicação e companheirismo. Aos meus pais por todo o apoio, incentivo, amor e educação que eles sempre me deram. A minha avó por todo o seu carinho e incentivo. Aos meus irmãos pela cooperação e apoio; a todos os meus tios, primos e amigos que me ajudaram de alguma forma.

Bruna

A minha colega de equipe Bruna Gomes, pelo apoio e dedicação, a todos os bons professores que estiveram à minha frente e que me engrandeceram com seu conhecimento, em especial ao professor e orientador desta monografia Dr. Aran Bey Tcholakian Morales, pela paciência, empenho e paixão pelos assuntos abordados, e aos meus pais, irmão e namorada, pelo incentivo constante.

Vinicius

RESUMO

Com base nos dados disponibilizados pelo IBGE e INEP, este estudo, por meio da análise exploratória de dados (AED), tem como objetivo identificar padrões e mostrar determinadas relações, entre cursos, docentes, discentes e empresas da área de TI. Por meio da técnica de Business Intelligence, os dados foram coletados de diferentes fontes, depois tratados e carregados em um banco de dados, formando o *Data Warehouse*. As relações e cruzamentos foram feitos através do Access[®], e a visualização das análises utilizando AED. As informações aqui levantadas podem ser utilizadas para estudos do futuro, por meio de cenários e suposições, ajudando assim, a melhorar os cursos da área de TI, e auxiliando os acadêmicos na hora da escolha por determinada Instituição de Ensino, curso ou região para sua formação. A formação de um conceituado profissional em TI depende, além de outras coisas, de uma boa Instituição de Ensino, que por sua vez precisa de docentes (professores) qualificados, com boa graduação, titulação e experiência. Saber que tipo de profissional o mercado está “comprando”, também ajuda a abrir novos cursos ou focalizar melhor determinadas disciplinas.

Palavras-chave: Análise Exploratória de Dados. Business Intelligence. Data Warehouse. Cursos da área das Ciências da Computação.

ABSTRACT

Based on data provided by IBGE and INEP, this research, done through the exploratory analysis of data (AED), aims at identifying standards and showing specific relationships between courses, professors, students and companies of IT. Through the Business Intelligence technique, the data was collected, forming the Data Warehouse. The relationships and crossings were made through Access, and the display of the analysis through AED. The information gathered can be used for future studies, through scenery and assumptions, helping the improvement of IT courses, helping the students when they need to choose an Educational Institution, a course or a region for their educational development. The development of a highly qualified professional of IT depends, among other things, on a good Educational Institution, which, on the other hand, needs good professors, who are titled and experienced. Being aware of the type of professional the market is interested in also helps the start of new courses or to emphasizes the focus on certain subjects.

Key-words: Exploratory Analysis of Data. Business Intelligence. Data Warehouse. Computer Science Area Courses.

LISTA DE FIGURAS

Figura 1 – Proposta de Solução.....	24
Figura 2 – Abrangência de um sistema de <i>BI</i>	28
Figura 3 – Arquitetura de <i>BI</i>	29
Figura 4 – Exemplo integração.....	34
Figura 5 – A questão da não-volatilidade.....	34
Figura 6 – Arquitetura Global.....	38
Figura 7 – Arquitetura de Data Marts Integrados.....	40
Figura 8 – Abordagem “ <i>Top-Down</i> ”.....	42
Figura 9 – Implementação <i>Bottom Up</i>	43
Figura 10 – Modelo Dimensional.....	47
Figura 11 – Modelo Estrela.....	48
Figura 12 – Dimensões do cubo.....	55
Figura 13 - Técnicas utilizadas pela mineração e dados.....	56
Figura 14 – Visões de um sistema de software.....	59
Figura 15 - Relações entre os diagramas da UML.....	61
Figura 16 - Diagrama de caso de uso.....	62
Figura 17 - Ator, caso de uso e relacionamento.....	63
Figura 18 - Diagrama de Classes.....	64
Figura 19 - Diagrama de Atividade.....	65
Figura 20 - Diagrama de Seqüência.....	66
Figura 21 - Elementos do diagrama de robustez.....	67
Figura 22 - Diagrama de robustez.....	67
Figura 23 - Processo ICONIX, mostrando a contribuição dos “três amigos”.....	69
Figura 24 - ICONIX - Atividades da análise de requisitos.....	72
Figura 25 - ICONIX - Atividades da análise e projeto preliminar.....	73
Figura 26 - ICONIX - Atividades do projeto.....	74
Figura 27: Diagrama de Casos de uso.....	75
Figura 28: Diagrama de classe do <i>ETL</i>	76
Figura 29: Diagrama de caso de uso do <i>ETL</i>	77
Figura 30: Diagrama de robustez do <i>ETL</i>	78
Figura 31: Diagrama de seqüência do <i>ETL</i>	79
Figura 32: Diagrama de caso de uso do <i>Front-End</i>	80
Figura 33: Diagrama de robustez do <i>Front-End</i>	80
Figura 34: Diagrama de seqüência do <i>Front-End</i>	81
Figura 35: Modelo de dados.....	83
Figura 36 – Proposta de Solução.....	84

LISTA DE GRÁFICOS

Gráfico 1 – Análises cursos.....	89
Gráfico 2 – Cursos por titulação do coordenador e organização acadêmica (total por titulação).	90
Gráfico 3 – Cursos por Categoria Administrativa e Titulação do Coordenador	90
Gráfico 4 – Cursos por Categoria Administrativa e Ano de início do Curso (total por ano de início do curso).	91
Gráfico 5 – Cursos por Organização Acadêmica e Ano de início do Curso.	91
Gráfico 6 – Cursos por Região e Ano de início do Curso.	92
Gráfico 7 – Total de Docentes por Titulação e Região (Total por Região) (a), e Docentes (em %) por	94
Titulação e Região (total por Região) (b).	94
Gráfico 8 – Titulação Máxima dos Docentes por Curso - Região.....	95
Gráfico 9 – Relação (em %) por Titulação e Organização Acadêmica (total por titulação) (a), Docentes (em %) por Titulação e Organização Acadêmica (total por Organização Acadêmica) (b) e Relação Titulação Máxima dos Decentes por curso (total por Organização Acadêmica) (c).....	96
Gráfico 10 – Docentes (em %) por titulação e categoria administrativa total por categoria administrativa (b) e por titulação (a), e Relação titulação máxima dos docentes por curso por categoria administrativa (c).	98
Gráfico 11 – Análises por tempo médio dos docentes.	99
Gráfico 12 – Discentes (em %) por região e vagas,inscritos,ingressos,matriculas,concluintes e cursos – região.	102
Gráfico 13 – Relação vagas, inscritos, ingressos, matriculas e concluintes por curso – região..	103
Gráfico 14 – Discentes (em %) por organização acadêmica e vagas,inscritos,ingressos,matriculas,concluintes e cursos.	104
Gráfico 15 – Relação vagas, inscritos, ingressos, matriculas, concluintes por curso – organização acadêmica.....	105
Gráfico 16 – Discentes (em %) por categoria administrativa e vagas, inscritos, ingressos, matriculas, concluintes e cursos.....	106
Gráfico 17 – Relação vagas, inscritos, ingressos, matriculas, concluintes por curso – categoria administrativa.....	106
Gráfico 18 – Discentes por ano de início do curso e vagas, inscritos, ingressos, matriculas, concluintes e cursos.	107
Gráfico 19 – Relação vagas, inscritos, ingressos, matriculas, concluintes por curso – ano de início do curso.....	108
Gráfico 20 – Relações incritos/vagas.	111
Gráfico 21 – Relação inscritos/vagas por área específica.	112
Gráfico 22 – Relações ingresso/concluinte.	113
Gráfico 23 – Relação ingresso/concluínite por área específica.....	114
Gráfico 24 – Análises total de empresas.	116
Gráfico 25 – Número de empresas por CNAE.....	117
Gráfico 26 – Soma de salários por região.	118
Gráfico 27 – Relação soma salário/número de empresas.....	118

Gráfico 28 – Relação soma salário/número de pessoas assalariadas – região.....	119
Gráfico 29 – Relação salário/número de pessoas assalariadas - CNAE.....	120
Gráfico 30 – Análises Empresas.....	120
Gráfico 31 – Soma pessoas ocupadas por CNAE.....	121
Gráfico 32 – Soma de pessoas salário por CNAE.	122
Gráfico 33 – Relação número de pessoas assalariadas/empresas.....	122

LISTA DE QUADROS

Quadro 1 – Comparação entre modelo Relacional – E/R e modelo Dimensional.	45
Quadro 2 - Vantagens e desvantagens dos métodos de prospecção.	53
Quadro 3 – Questões e Técnicas.....	69
Quadro 4 – Fases e Marcos do Iconix.	71
Quadro 5 - Tecnologias Utilizadas.	86

LISTA DE TABELAS

Tabela 1 – Tabela de discentes (em %) por área específica e vagas, inscritos, ingressos, matriculas, concluintes e cursos.....	109
Tabela 2 – Tabela de relação vagas, inscritos, ingressos, matriculas, concluintes por curso – área específica.	110

LISTA DE SIGLAS

AED – Análise Exploratória de Dados

BD – Banco de Dados

BI – *Business Intelligence*

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CIEE – Centro de Integração Empresa Escola

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

DCU – Diagrama de Caso de Uso

DSS – *Decision Support System*

DW – *Data Warehouse*

EIS – *Executive Information System*

ERP – *Enterprise Resource Planning*

ER – Entidade-Relacionamento

ETL – *Extraction, Transformation and Loading*

IBGE – Instituto Brasileiro de Geografia e Estatística

IES – Instituição de Ensino Superior

INEP – Instituto Nacional de Estudos e Pesquisas

MCT – Ministério da Ciência e Tecnologia

ODS – Operational Data Store

OLAP – *Online Analytical Processing*

OLTP – *Online Transaction Processing*

PIB – Produto Interno Bruto

RUP – *Rational Unified Process*

SAD – Sistema de Apoio à Decisão

SQL – *Structured Query Language*

TI – Tecnologia da Informação

UML – *Unified Modeling Language*

SUMÁRIO

1 INTRODUÇÃO	19
1.1 APRESENTAÇÃO DO PROBLEMA.....	19
1.2 JUSTIFICATIVA	21
1.3 OBJETIVOS.....	22
1.3.1 Objetivo Geral	22
1.3.2 Objetivos específicos.....	23
1.4 DELIMITAÇÕES.....	23
1.5 PROPOSTA DE SOLUÇÃO.....	23
1.5.1 Fonte de Dados.....	24
1.5.2 ETL	24
1.5.3 Data Warehouse.....	24
1.5.4 Análise dos Dados	24
1.6 ESTRUTURA DO TRABALHO	25
2 REVISÃO BIBLIOGRÁFICA – SISTEMAS DE INFORMAÇÃO	26
2.1 BUSINESS INTELLIGENCE	26
2.1.1 Conceito	26
2.1.2 Sistema de BI.....	27
2.1.3 Arquitetura de BI	28
2.1.3.1 Stage ou ETL (<i>Extraction, Transformation and Loading</i>).....	29
2.1.3.2 Data Warehouse.....	30
2.1.3.3 Front – End	30
2.2 DATA WAREHOUSE	31
2.2.1 Conceito	31
2.2.2 Definições.....	32
2.2.3 Características	32
2.2.3.1 Orientado por assunto.....	33
2.2.3.2 Integrado.....	33

2.2.3.3 Não Volátil	34
2.2.3.4 Variável em relação ao tempo	35
2.2.4 Justificativas.....	35
2.2.5 Arquitetura de um Data Warehouse	36
2.2.5.1 Arquitetura Global.....	37
2.2.5.2 Arquitetura de Data Mart Independente	38
2.2.5.3 Arquitetura de Data Mart Integrados.....	39
2.2.6 Abordagens de Implementação	40
2.2.6.1 Abordagem “ <i>Top-Down</i> ”	40
2.2.6.2 Abordagem Bottom Up	42
2.2.6.3 Arquitetura BUS	43
2.2.7 Modelo Dimensional.....	44
2.2.7.1 Fatos	45
2.2.7.2 Dimensões	45
2.2.7.3 Medidas (variáveis)	46
2.2.7.4 Vantagens da Modelagem Dimensional.....	47
2.2.7.5 Modelo Estrela.....	48
2.2.8 Granularidade de Dados.....	49
2.3 PROSPECÇÃO DE DADOS	50
2.3.1 Vantagens e desvantagens dos diferentes métodos e técnicas de prospecção	51
2.3.2 Análise Exploratória de Dados.....	54
2.3.2.1 Olap	54
2.3.2.1.1 Operadores OLAP	55
2.3.2.2 Data Mining.....	56
2.4 CONCLUSÃO	57
3 MODELAGEM.....	58
3.1 UML	58
3.1.1 Visões de um sistema	59
3.2 DIAGRAMAS DA UML	60
3.2.1 Diagramas de casos de uso (Use-Cases).....	61

3.2.1.1 Atores	62
3.2.2 Diagramas de Classe.....	63
3.2.3 Diagramas de atividades	64
3.2.4 Diagramas de seqüência.....	65
3.2.5 Diagrama de robustez	66
3.3 METODOLOGIA ICONIX.....	68
3.3.1 Principais características do ICONIX.....	70
3.3.2 Fases e Marcos	70
3.3.2.1 Análise de Requisitos	71
3.3.2.2 Análise e Projeto Preliminar.....	72
3.3.2.3 Projeto.....	73
3.3.2.4 Implementação.....	74
3.4 MODELAGEM DO SISTEMA	74
3.4.1 Casos de uso	74
3.4.1.1 ETL	76
3.4.1.1.1 Diagrama de classes	76
3.4.1.1.2 Diagrama de caso de uso	76
3.4.1.1.3 Diagrama de Robustez.....	77
3.4.1.1.4 Diagrama de Seqüência	78
3.4.1.2 Font-End	79
3.4.1.2.1 Diagrama de caso de uso	79
3.4.1.2.2 Diagrama de Robustez.....	80
3.4.1.2.3 Diagrama de Seqüência	81
3.4.2 Implementaçã.....	82
3.5 CONCLUSÃO.....	82
4 DESENVOLVIMENTO.....	83
4.1 PROPOSTA DE SOLUÇÃ.....	84
4.1.1 Fonte de Dados.....	84
4.1.2 ETL	85
4.1.2 Data Warehouse.....	85

4.1.3 Análise dos Dados	85
4.2 TECNOLOGIAS UTILIZADAS	85
4.3 ANÁLISES	87
4.3.1 Análises de Cursos	87
4.3.1.1 As análises	89
4.3.2 Análises de Docentes.....	92
4.3.2.1 As análises	93
4.3.2.1.1 Por região.....	94
4.3.2.1.2 Por organização acadêmica	96
4.3.2.1.3 Por categoria administrativa	97
4.3.2.1.4 Por tempo médio do docente	99
4.3.3 Análises de Discentes	100
4.3.3.1 As análises	102
4.3.3.1.1 Por região.....	102
4.3.3.1.2 Por organização acadêmica	104
4.3.3.1.3 Por categoria administrativa	105
4.3.3.1.4 Por ano de início do curso	107
4.3.3.1.5 Por área específica do curso	108
4.3.3.1.6 Relação inscritos/vagas.....	111
4.3.3.1.7 Relação ingressos/concluintes	113
4.3.4 Análises das Empresas	115
4.3.4.1 As análises	116
4.3.4.1.1 Pelo número de empresas	116
4.3.4.1.2 Pela soma de salários	117
4.3.4.1.3 Pelo número de pessoas ocupadas	120
4.3.4.1.4 Pelo número de pessoas salário	121
4.4 CONCLUSÃO	123
5 CONCLUSÃO E TRABALHOS FUTUROS.....	124
5.1 CONCLUSÕES	124
5.2 TRABALHOS FUTUROS	125

REFERÊNCIAS BIBLIOGRÁFICAS	126
---	------------

1 INTRODUÇÃO

O mundo atual está sendo marcado por grandes avanços científicos e tecnológicos em um momento que o acesso à informação e ao conhecimento se expande para toda a sociedade. Todos os dias novas pesquisas são iniciadas com a finalidade de melhorar a qualidade de vida do ser humano e continuar suprindo as necessidades já existentes. A Tecnologia da Informação (TI) e as áreas afins são peças fundamentais para o auxílio de tais pesquisas.

Nos dias de hoje, a sociedade está vivendo a era da informação e do conhecimento, e a TI é essencial nesta nova era, sendo impossível pensar em educação, comunicação, saúde, segurança, entre outras áreas, sem a TI. Assim, nota-se que a TI está fortemente ligada ao desenvolvimento sócio-econômico do país, sendo visível e inquestionável sua importância em nossa sociedade.

Conseqüentemente, com o aumento da tecnologia e do uso do computador, aumenta a necessidade da existência de profissionais graduados e especializados em áreas relevantes a TI. Neste sentido os cursos oferecidos pelas instituições de ensino superior (IES) na formação de recursos humanos, produção técnico-científica, as empresas da área de TI e suas demandas, e os financiamentos em pesquisa e inovação, são variáveis essenciais para a inserção do Brasil na sociedade da informação e do conhecimento.

1.1 APRESENTAÇÃO DO PROBLEMA

O Ministro Sergio Rezende afirmou dia 2 de março de 2007 na 1ª. Conferência Rio de Janeiro International Software & Services Outsourcing, que as ações e os investimentos, que vêm sendo feitos nas áreas de formação de especialistas e de incentivos à empresa de tecnologia da informação, asseguram o apoio que o país necessita para se tornar um grande exportador de softwares e serviços. (MINISTÉRIO DA CIÊNCIA E TECNOLOGIA, 2007b).

Esteve também presente no evento, o ministro do Desenvolvimento, Indústria e Comércio Exterior, Luiz Fernando Furlan, que ressaltou a importância da área de software para o aumento das exportações, elogiou os investimentos em capital de risco que vêm sendo feitos pelo

Ministério da Ciência e Tecnologia (MCT) para o setor, e enumerou uma série de programas que serão adotados para melhorar o nível educacional no país. (MINISTÉRIO DA CIÊNCIA E TECNOLOGIA, 2007b).

Furlan lembrou que para os empresários estrangeiros, a educação e a logística são os maiores entraves ao aumento das exportações de Tecnologia da Informação. E citou a Índia como um exemplo de mudança bem sucedida de um país que soube se comunicar e mostrar ao mundo que é capaz de produzir conhecimento de qualidade. Em suas palavras: "Nós precisamos mudar a imagem de nosso País e a área de tecnologia é importante para que o mundo entenda o que somos". (MINISTÉRIO DA CIÊNCIA E TECNOLOGIA, 2007b).

Este projeto faz uma análise exploratória dos dados sobre as variáveis que atuam na formação dos recursos humanos especializados na área de TI. A análise exploratória de dados (AED) tem por objetivos tornar mais clara a descrição dos dados, por meio da construção de indicadores, como distribuição, comparação e análises. Dependendo dos dados disponíveis, a AED permite análises geográficas, temporais, por ator ou setor, assim como análises de cenários futuros, identificando demandas futuras e potenciais, baseadas em tendências históricas e eventos hipotéticos.

A AED envolve visualização gráfica dos dados, indicadores e índices que permitem ao analista identificar padrões, comportamentos, relações e dependências.

Entre essas variáveis destacam-se os cursos oferecidos pelas Instituições de Ensino Superior (IES), a formação de pesquisadores, a produção científica e tecnológica, a formação de recursos humanos especializados, as empresas do setor, assim como os investimentos, tantos públicos quanto privados, diretos e indiretos.

Para construir os instrumentos de análises pretendidos por meio das análises exploratórias, é necessário ter um repositório de dados, ou seja, uma base de dados consistente, uniforme e confiável que contenha todas as informações necessárias para realizar as análises requeridas. A técnica adequada para estabelecer uma infra-estrutura de bases de informação, agregada segundo as diversas dimensões de análise, viabilizando a gestão integrada das informações, é a concepção, projeto e implantação de um *data warehouse (DW)*.

Tendo como base um repositório de dados (DW), o próximo passo, é a utilização de algoritmos e métodos avançados para extrair informações relevantes e conhecimento dos dados armazenados no repositório. Os métodos oferecem a possibilidade de cruzamento de variáveis

capazes de gerar informações com indicadores gerenciais, indicadores de avaliação e de desempenho, permitindo realizar análises que possibilitam a visualização quantitativa e o perfil das informações.

1.2 JUSTIFICATIVA

Na largada do século 21, as bases para o desenvolvimento soberano e sustentável de um país não são apenas aquelas que prevaleciam até meados do século passado, riquezas naturais, extensão territorial, ou até mesmo a capacidade militar. Desenvolvidos são os países que detêm conhecimento técnico-científico e, portanto, a capacidade de decidir sobre o seu destino. Aos demais, resta o papel de fornecedores de matérias-primas e de mercados importadores de produtos de maior valor agregado, condenados ao subdesenvolvimento. É o rumo de um Mundo cada vez mais globalizado, onde, nos últimos 50 anos, os avanços e as descobertas científicas superaram todo o saber acumulado, até então, pela Humanidade. (MINISTÉRIO DA CIÊNCIA E TECNOLOGIA, 2007a).

Tal transformação da economia mundial nas últimas décadas resultou em novos paradigmas de desenvolvimento, baseados no segmento de serviços, que tem como ingrediente fundamental a produção tecnológica.

Entre o segmento de serviços, destaca-se nas últimas décadas a indústria de tecnologias da informação fortemente associada com a formação de recursos humanos altamente especializados e com a pesquisa científica e tecnológica.

O Brasil investe 1% do seu PIB em ciência e tecnologia, sendo o país do hemisfério sul que mais investe neste setor. O presidente do Conselho de Administração do Centro de Integração Empresa-Escola (CIEE), Paulo Nathanael Pereira, disse, no 8º seminário promovido pela entidade em parceria com o jornal Gazeta Mercantil, em São Paulo, o seguinte "Devemos considerar a educação como insumo e o desenvolvimento econômico como produto. Afinal, o progresso não se dá sem escolaridade qualificada". O evento abordou o desenvolvimento educacional como principal arma para o crescimento econômico brasileiro. (MINISTÉRIO DA CIÊNCIA E TECNOLOGIA, 2007c).

Marcos Troyjo, presidente internacional da Gazeta Mercantil, destacou que educação e integração entre empresas e escolas é fundamental para o impulso econômico do País. "É através da educação que se pode aumentar lucro, competitividade e capacitação do terceiro setor", disse. Citando o colunista norte-americano do *The New York Times*,

Thomas Friedman, abordou as diferentes fases da globalização. "Já passamos pela globalização dos estados nacionais, das empresas e neste momento vivemos uma terceira fase, a globalização dos indivíduos". Para tornar o País mais competitivo, destaca a necessidade de aumentar a capacitação profissional. (MINISTÉRIO DA CIÊNCIA E TECNOLOGIA. 2007c).

Estabelecer relações entre as variáveis que tem influência no setor de TI, conhecer seus principais atores, o dinamismo e relação entre eles, seu comportamento atual e futuro, são essenciais no processo de planejamento para a área de desenvolvimento tecnológico no país. Responder quais são as áreas de maiores demandas e os nichos de mercado existentes para elas, quais são as regiões "importadoras" e as "exportadoras" de mão de obra na área, fornecer indicadores para auxiliar nas ações que devam ser tomadas, tanto na área de ensino, de pesquisa, investimentos em recursos humanos ou outros, são perguntas que o trabalho pretende responder.

1.3 OBJETIVOS

Os objetivos deste projeto são:

1.3.1 Objetivo Geral

Por meio da Análise Exploratória de Dados, o objetivo deste trabalho é ter um panorama sobre os principais fatores que influenciam a área das ciências da computação no Brasil; o que consiste em captar, organizar e analisar os dados, de modo que se consiga estabelecer as relações existentes entre as diversas variáveis que podem influir no meio acadêmico e empresarial do setor de TI.

1.3.2 Objetivos específicos

- Construir indicadores nacionais das variáveis envolvidas que permitam analisar a TI em seu estado atual.
- Identificar lacunas de conhecimento existentes nos atuais cursos de graduação da área de ciência da computação.
- Fornecer visões e perfis das análises, de acordo com as diferentes variáveis em estudo.
- Construir estruturas de informação que efetivarão as análises especificadas por meio de métodos e algoritmos adequados.
- Documentar a especificação das estruturas de informação e das análises especificadas.

1.4 DELIMITAÇÕES

As delimitações do trabalho aconteceram em dois sentidos:

- Dependeram das variáveis coletadas. Foram analisadas somente aquelas que continham informações suficientes.
- A visualização das conclusões baseadas na AED dependeu das relações existentes entre as variáveis analisadas e as ferramentas disponíveis para realizar tais análises.

1.5 PROPOSTA DE SOLUÇÃO

A figura a seguir mostra a proposta de solução:

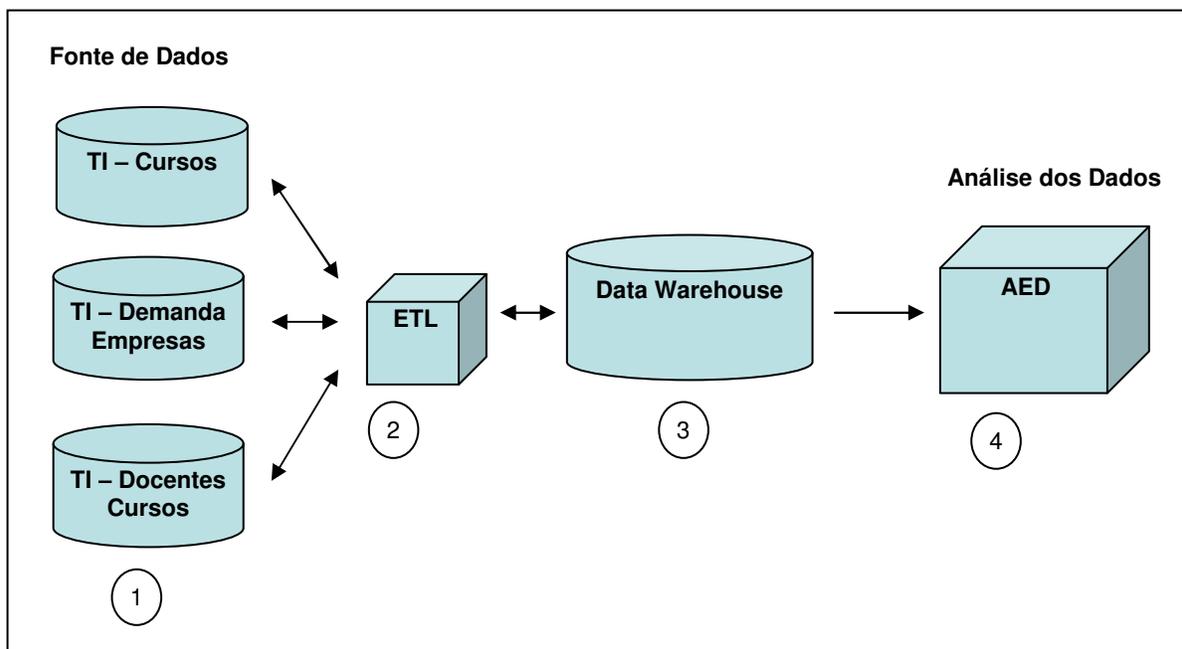


Figura 1 – Proposta de Solução.

Fonte: Elaboração dos Autores, 2007.

- 1.5.1 Fonte de Dados** – Fontes de dados que são oriundas de três diferentes vertentes, os dados sobre os cursos de Tecnologia da Informação (TI), os dados sobre a formação dos docentes destes cursos e os dados sobre a demanda das empresas de TI. Estas fontes de dados formam a base de dados dos sistemas operacionais de onde serão extraídos os dados para a formação do *Data Warehouse (DW)*.
- 1.5.2 ETL** – Extração dos dados da base de dados dos sistemas operacionais, transformação destes dados para um único formato e carga dos mesmos no *DW*.
- 1.5.3 Data Warehouse** – Repositório de dados baseado em assuntos, integrado, não volátil e variável em relação ao tempo. Será utilizado este repositório para a análise de dados e sistemas de apoio à decisão.
- 1.5.4 Análise dos Dados** – Análise exploratória dos dados (AED) sobre os dados armazenados no *DW* para estabelecer relações e indicadores que esclareçam como os cursos de Ciência da Computação estão distribuídos em relação as variáveis analisadas, como docentes, discentes, entre outras.

1.6 ESTRUTURA DO TRABALHO

O trabalho está dividido em 5 capítulos, onde o capítulo 1 apresenta o problema, a justificativa, os objetivos, as delimitações, a proposta de solução para a realização deste, e o método de pesquisa utilizado.

O capítulo 2 encontra-se dividido em duas partes, uma que trata de Prospecção e a outra dos Sistemas de Apoio a Decisão. Dentro de Prospecção são vistas as análises prospectivas, que é a metodologia de Análise Exploratória de Dados. Em Sistemas de Apoio a Decisão são abordados os conceitos e elementos da arquitetura de *Business Intelligence (BI)* e definições, objetivos, características, arquiteturas e modelos de um *Data Warehouse (DW)*.

O capítulo 3 mostra os principais conceitos da *UML* e da Metodologia ICONIX. Bem como, apresenta parte da modelagem desenvolvida neste trabalho.

O capítulo 4 aborda a coleta e extração de dados para um *DW*, assim como as análises exploratórias de tais dados para formação de possíveis cenários futuros. Os dados coletados são sobre os cursos da área das ciências da computação, formação profissional dos docentes destes cursos e demanda do mercado de trabalho na área de Tecnologia da Informação (TI).

O capítulo 5 apresenta as conclusões obtidas com a realização deste trabalho e propostas para trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA – SISTEMAS DE INFORMAÇÃO

Com a crescente aceleração das TI's e da competitividade em todos os mercados, análises das variáveis e indicadores sobre o setor se tornam necessários para tomar algumas decisões no presente, visando caminhos alternativos para o futuro próximo. Várias técnicas e ferramentas são citadas na literatura para realizar tal abordagem. Assim sendo, o capítulo 2 encontra-se dividido em duas partes, uma que trata da Prospecção de dados e a outra dos Sistemas de Apoio a Decisão.

Dentro da prospecção de dados serão abordadas as análises exploratórias de dados que é a metodologia escolhida para o presente trabalho. Em Sistemas de Apoio a Decisão serão apresentados os conceitos e elementos da arquitetura de *Business Intelligence (BI)* e definições, objetivos, características, arquiteturas e modelos de um *Data Warehouse*.

2.1 BUSINESS INTELLIGENCE

Esta seção descreve conceitos, características e arquiteturas do *BI*.

2.1.1 Conceito

“O conceito de *BI – Business Intelligence*, de forma mais ampla, pode ser entendido como a utilização de variadas fontes de informação para se definir estratégias de competitividade nos negócios da empresa”. (BARBIERI, 2001).

Para Herring (1997), o BI é um processo organizacional de coleta e análise sistemática de informação, que depois é disseminada como suporte para apoiar a tomada de decisão.

Atualmente as empresas perdem tempo e dinheiro por não conseguirem extrair informações relevantes de imensas montanhas de dados. A parte gerencial das companhias fica impotente mediante a pilhas de tabelas, arquivos e relatórios que se ligam, se relacionam de maneira transacional, tornando-os inacessíveis ou no mínimo de difícil entendimento, atravancando as tomadas de decisão.

O grande desafio de todo indivíduo que gerencia qualquer processo é a análise dos fatos relacionados a seu dever. Ela deve ser feita de modo que, com as ferramentas e dados disponíveis, o gerente possa detectar tendências e tomar decisões eficientes no tempo correto. Com essa necessidade, surgiu então o conceito de *Business Intelligence*. (SERRA, 2002, p.77).

As principais características de uma solução de BI são:

- Analisar dados que tenham contexto;
- Trabalhar com múltiplas alternativas (hipóteses);
- Buscar e mesclar dados de várias fontes;
- Usar a experiência;
- Tratar as informações e dados coletados para ajudarem na tomada decisão da empresa.

2.1.2 Sistema de BI

No ambiente de BI cada caso é único forçando a construção da ferramenta, ou ao menos, modificando bastante alguma ferramenta já existente, principalmente quando a inteligência envolve recolhimento e análise de largas quantidades de dados desestruturados. A quantidade de softwares para gerenciamentos destas informações é alta, visto que se faz quase necessário a construção de um novo para cada caso. Obviamente, que se o banco de dados for pequeno ou simples, ferramentas já existentes podem servir perfeitamente. Uma pesquisa nas ferramentas de busca da internet pode revelar diversas informações sobre cada um destes programas, facilitando a escolha. Com o crescimento do mercado de ferramentas de Data Warehouse, que faz parte do mercado de BI, ferramentas melhores e mais sofisticadas foram desenvolvidas para apoiar a estrutura do Data Warehouse e sua utilização.

Segundo Tyson (1997), Sistema de *Business Intelligence* – BI é um processo que envolve a coleta, análise e validação de informações sobre concorrentes, fornecedores, candidatos potenciais à aquisição, candidatos a *joint-ventures* e alianças estratégicas. Essas informações também podem incluir eventos econômicos, reguladores e políticos que tenham impacto potencial sobre os negócios da empresa. O processo de BI avalia e valida todas essas informações e as transforma em conhecimento estratégico.

As estratégias das empresas funcionam como base para o sistema de BI.

O principal objetivo de um Sistema de Inteligência de Negócios é aprender sobre o ambiente competitivo externo, visando o conhecimento do posicionamento competitivo da empresa, o que impulsionará mudanças internas e facilitará decisões estratégicas na empresa. (COMPETE.NET, 2007) .



Figura 2 – Abrangência de um sistema de BI.
Fonte: Adaptado de HERRING, 1997.

2.1.3 Arquitetura de BI

A figura abaixo exemplifica um típico modelo de arquitetura de *Business Intelligence*.



Figura 3 – Arquitetura de BI.
Fonte: ALCANTARA, 2006.

2.1.3.1 Stage ou ETL (*Extraction, Transformation and Loading*)

O ETL tradicional é um processo orientado por lotes que se concentra na junção e integração dos dados durante o tempo de ociosidade do negócio. (GORDON; GRIGG; HORNE; THURMAN, 2007).

Segundo o Data Warehouse (2007), o *ETL* (Extração, Transformação e Carga) serve para:

- Buscar os dados nas fontes de dados da empresa (dados externos, ERP's e outros dados operacionais), tomando o devido cuidado, pois trabalha com a movimentação dos dados.
- Limpar e transformar os dados uma vez que estão em diferentes formatos e saíram de diferentes tipos de fontes possuindo muito lixo e inconsistência. Em alguns sistemas, a informação sobre o sexo do cliente pode estar como “H” de Homem ou “M” de Masculino, por exemplo.
- Carregar os dados, mas por possuir uma enorme complexidade, os fatores como integridade dos dados, criação de rotinas de carga para atender determinadas situações e a parte de histórico das tabelas deve ser levado em consideração. No momento da carga é

necessário checar os campos que são chaves estrangeiras com suas respectivas tabelas para certificar-se de que os dados existentes na tabela da chave estrangeira estão de acordo com a tabela da chave primária.

2.1.3.2 Data Warehouse

É um banco de dados que consolida dados extraídos de diversos sistemas de produção e operacionais em um grande banco de dados que pode ser utilizado para relatórios e análises gerenciais (LAUDON & LAUDON, 1999).

O *Data Warehouse* ou Armazém de Dados, também proporciona ferramentas para busca e análise, colocando informação de todos os setores da organização ao alcance de quem toma as decisões. (INMON, 2000).

“É derivado da produção ou das bases de dados transacionais e usado para automatizar os processos do negócio” (LIAUTAD, 2000, p.60).

Certamente, a idéia não é tanto que os dados residam fisicamente em um único sistema computadorizado, mas que todos os dados estejam armazenados e acessíveis através de uma rede de sistemas distribuídos de modo que se apresentem como uma coleção sem emenda de dados incorporados. (Bigus, 1996, p.7-9, tradução nossa).

No item 2.2 os tópicos vinculados ao DW serão abordados com maior profundidade.

2.1.3.3 Front – End

Front-end pode ser entendido como uma interface gráfica, na qual o usuário possa interagir com programas em modo texto, relatórios e gráficos. Dentro das aplicações de *Front-End* estão, as ferramentas de processamento e análises analítico (OLAP) e as ferramentas de extração de conhecimento ou mineração de dados. Estes tópicos serão abordados no item 2.3.

2.2 DATA WAREHOUSE

Devido aos avanços tecnológicos e a globalização ocorridos nas últimas décadas, as organizações vem sofrendo mudanças estruturais e organizacionais, tornando-se cada vez mais competitivas. A informação é entendida como insumo básico, trocando a velha economia, baseada em recursos materiais, linha de montagem e tomada de decisão centralizada, por uma economia global, com tomada de decisões descentralizadas e conduzida pela informação (NETO, 2001).

Segundo Kimball (2002, p.2), um dos bens mais preciosos de qualquer empresa são suas informações. O grande problema encontrado atualmente nas empresas, é que a maioria delas possuem as suas informações armazenadas em sistemas transacionais (sistemas de apoio à operação), que dão suporte ao dia-a-dia da organização, ou seja, elas possuem a informação disponível, porém a um custo muito elevado de acesso em termos tecnológicos, comprometendo a confiabilidade das informações obtidas e penalizando os sistemas transacionais em termos de performance. Além disso, para extração dessas informações precisa-se de recursos humanos especializados, o que acaba dificultando o acesso às informações por parte dos tomadores de decisão das organizações ou usuários do negócio. (TCHOLAKIAN, 2007).

Para solucionar este problema surgiu o *Data Warehouse*. O *DW* preocupa-se em extrair, integrar, limpar e dar consistência a dados vindos tanto de sistemas operacionais da organização quanto de dados externos, tais como pesquisas de mercado e dados sobre a concorrência. Em outras palavras, ele procura integrar e consolidar dados disponíveis em diferentes acervos, aumentando assim o conteúdo informacional de tais acervos para desta forma, poder atender às necessidades e expectativas de nível estratégico da empresa. O *DW* também dimensiona e consolida os dados, organizando-os de forma a melhorar a performance das consultas. (CENTENARO, 2003).

2.2.1 Conceito

O *Data Warehouse* é o ponto central da arquitetura de processamento de informações para sistemas de informática modernos, suportando o processamento informacional de Sistemas de Apoio à Decisão (SAD) por meio de um alicerce sólido de integração de dados corporativos e históricos para a realização de análises gerenciais. (INMON e HACKATHORN, 1997 apud CENTENARO,2003).

2.2.2 Definições

Kimball (1998, p.19 apud NETO, 2001), define o *Data Warehouse* como sendo:

Uma fonte de dados que possui a capacidade de consultar as informações da organização. O *data warehouse* é na realidade uma união de todas as bases de dados constituintes da organização. O data warehouse é alimentado por uma área especial que prepara os dados para serem utilizados. O gerenciamento do data warehouse é responsável tanto pela busca das informações desejadas quanto pela preparação destas informações.

Para Machado (2000, p.6), construir um *DW* significa construir armazéns de dados nos quais a história da empresa, seus clientes, fornecedores e operações se mantêm disponíveis e acessíveis para consultas e análises.

Laudon e Laudon (1999, p. 137) definem um *Data Warehouse* como: “[...] um banco de dados que consolida dados extraídos de diversos sistemas de produção e operacionais em um grande banco de dados que pode ser utilizado para relatórios e análises gerenciais.”

Inmon (1997, p.33), um dos precursores da ferramenta, diz que: "Um *data warehouse* é um conjunto de dados baseado em assuntos, integrado, não-volátil, e variável em relação ao tempo, de apoio às decisões gerenciais.”

2.2.3 Características

Baseado na definição do Inmon, um *DW* possui quatro características básicas: é orientado por assunto, integrado, não volátil e variante no tempo.

2.2.3.1 Orientado por assunto

“[...] armazena as informações agrupadas por assuntos de interesse da empresa que são mais importantes, em contraste com os sistemas operacionais que são orientados a processos desenvolvidos para manter as transações realizadas diariamente” (MACHADO, 2000, p.14).

2.2.3.2 Integrado

Representação única para os dados oriundos dos diversos sistemas que irão compor a base de dados do *Data Warehouse*. Para Inmon (1997), de todos os aspectos do *DW*, esse é o mais importante.

“[...] ao projetar-se o modelo de dados do *data warehouse*, tem-se o cuidado de eliminar as redundâncias e as possibilidades de respostas ambíguas. O modelo é construído e organizado de forma a obter-se respostas únicas e certas, normalmente de forma independente da forma como os dados estão organizados nos sistemas operacionais;” (SELL, 2001,p.17).

Por meio da integração, os dados são representados de uma única forma, ou seja, os dados de todos os sistemas que formarão a base de dados do *Data Warehouse* são padronizados. Dessa forma a maior parte do trabalho de construção de um *DW* está na análise dos sistemas transacionais e dos dados que eles contêm. (TCHOLAKIAN, 2007).

Um exemplo clássico de integração é a representação do estado civil de uma pessoa. Muitos sistemas tratam como: “S” solteiro, “C” casado e assim por diante. Outros sistemas, por sua vez, tratam como: 1 solteiro, 2 casado e assim consecutivamente. Para solucionar este problema, estabelece-se um padrão e dentro do *DW* utiliza-se somente uma das formas, ou seja, a forma padronizada. (CENTENARO, 2003).

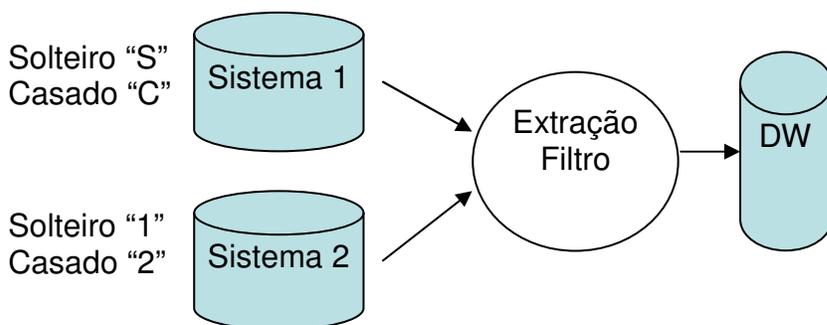


Figura 4 – Exemplo integração.
Fonte: Adaptado de MACHADO, 2000.

2.2.3.3 Não Volátil

“[...] a principal idéia na alimentação de dados no *data warehouse* é a de que em determinados períodos, sejam extraídos dados dos sistemas operacionais e armazenados no *data warehouse*. Uma vez armazenado, o dado não sofrerá alterações;” (SELL, 2001, p.17)

Para Machado (2000, p.15), um *DW* possui duas operações básicas: “a carga dos dados (inicial e incremental) e o acesso a esses dados em modo de leitura”. Estes dados se originam dos sistemas transacionais, são transformados e purificados e em seguida carregados no *DW*. (MELLO, 2002, p.28).

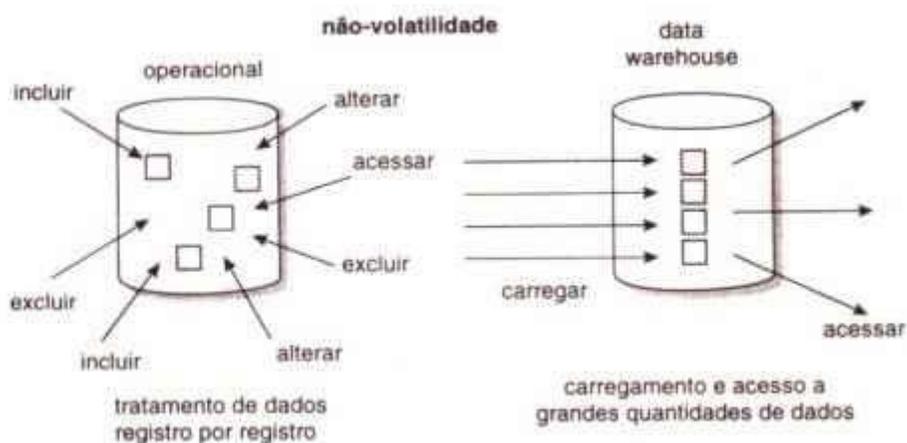


Figura 5 – A questão da não-volatilidade.
Fonte: INMON, 1997, p.36.

2.2.3.4 Variável em relação ao tempo

Segundo Machado (2000, p.15), “os dados de um *Data Warehouse* são preciosos em relação ao tempo e representam resultados operacionais em determinado momento de tempo, o momento em que foram capturados”. Isto quer dizer que os dados de um *DW* não podem sofrer alterações.

“[...] com o acúmulo de dados sobre diversos períodos, o *data warehouse* fornecerá subsídios para análises do negócio em tempos diferentes, possibilitando análises de regressões, tendências, etc” (SELL, 2001, p.17).

Inmon (1997), entende que os diversos modos pelos quais a variação em relação ao tempo se manifesta, são: a) O horizonte de tempo: o dos sistemas operacionais é, normalmente, de 60 a 90 dias, enquanto o do *DW* é muito maior, de 5 a 10 anos. b) Os BD operacionais podem ser atualizados (sofrer alterações em seus registros), por isso contêm dados de “valor corrente”, ou seja, dados cuja exatidão é válida apenas para o momento de acesso. O *DW* por sua vez, contém dados que não passam de uma série sofisticada de instantâneos, capturados num determinado momento. c) A estrutura de chaves dos dados operacionais pode conter ou não, elementos de tempo, como mês, ano, dia, etc. Já a estrutura de chave do *DW* sempre contém algum elemento de tempo.

2.2.4 Justificativas

Machado (2000, p.12), diz que as principais justificativas para a aplicação de tecnologia de *DW* em uma organização são a existência na empresa de:

- Várias plataformas de hardware e software.
- Constantes alterações nos sistemas transacionais corporativos.
- Dificuldade acentuada na recuperação de dados históricos em períodos superiores ao ano atual de operações.
- Existência de sistemas “pacotes” de fornecedores diferentes.
- Falta de padronização e integração dos dados existentes nos diversos sistemas.
- Carência de documentação e segurança no armazenamento dos dados.

- Dificuldade de aplicação de sistemas EIS ou DSS devido a dependências múltiplas de sistemas corporativos.

Segundo Machado (2000, p.11), o principal objetivo do *DW* é disponibilizar informações para apoio a decisões da empresa. Para ele o resultado de um projeto *Data Warehouse*, se resume em:

- Informação disponível para a gestão
- Visão de curvas de comportamento.
- Agilidade de ferramentas para apoio a decisão
- Segurança de informações para a decisão.
- Maior abrangência de visão de indicadores.
- Recursos mais abrangentes para a análise de negócios.
- Necessidades e expectativas executivas atendidas por tecnologia da informação.

“Sua crescente utilização pelas empresas está relacionada à necessidade do domínio de informações estratégicas para garantir respostas e ações rápidas, assegurando a competitividade de um mercado altamente competitivo e mutável” (MACHADO, 2000, p.11).

Além dessas quatro características ressaltadas por Inmon, Machado (2000, p.13) acrescenta que o *DW* possui um conjunto de características que o distingue de outros ambientes de sistemas convencionais. São elas:

- Extração de dados de fontes heterogêneas (existentes ou externas);
- Transformação e integração dos dados antes de sua carga;
- Normalmente requer máquina e suporte próprio;
- Visualização dos dados em diferentes níveis. Os dados do *Data Warehouse* podem ou não ser extraídos para um nível mais específico, os *Data Marts*, e a partir destes, para um banco de dados individual;
- Utilização de ferramentas voltadas para acesso com diferentes níveis de apresentação;
- Dados somente são inseridos, não existindo atualização, ou melhor, *updates*.

2.2.5 Arquitetura de um Data Warehouse

Segundo Machado (2000, p. 31), a escolha de uma arquitetura é uma decisão gerencial do projeto, e está normalmente baseada nos fatores relativos à infra-estrutura disponível, ao ambiente de negócios ou porte da empresa, simultaneamente com o escopo de

abrangência desejado, assim como a capacitação dos empregados e dos recursos disponibilizados ou projetados para o investimento.

A escolha da abordagem de implementação é uma decisão que pode gerar dramáticos impactos quanto ao sucesso de um projeto de *DW*. São diversas as variáveis que afetam a escolha da implementação e arquitetura, entre elas o tempo para a execução do projeto, o retorno do investimento a ser realizado, a velocidade dos benefícios da utilização das informações, a satisfação do usuário executivo e os recursos necessários à implementação de uma arquitetura. (MACHADO, 2000).

Para Machado (2000, p.32), a arquitetura determinará ou será determinada por onde o *Data Warehouse* ou *Data Marts* estiverem residindo.

Data Mart é sub-conjunto de dados de um *Data Warehouse*. Geralmente são dados referentes a um assunto em especial, como por exemplo: vendas, estoque, controladoria; ou diferentes níveis de sumarização, como: vendas anual e vendas mensal, que focalizam uma ou mais áreas específicas. Seus dados são captados no *DW*, desnormalizados e indexados para suportar intensa pesquisa (DATA MART, 2007).

A seguir, serão apresentadas três arquiteturas básicas: arquitetura Global, arquitetura de *Data Mart* Independente e arquitetura de *Data Marts* Integrados, definidas por Kimball (2002), Machado (2000) e DWBrasil (2005).

2.2.5.1 Arquitetura Global

Na arquitetura global o *DW* é projetado e construído baseado nas necessidades da empresa como um todo. É considerado como um repositório comum de dados disponível para toda empresa e baseado no suporte a decisão. (MACHADO, 2000, p. 32).

Os dados podem estar fisicamente centralizados ou fisicamente distribuídos nas instalações de uma empresa. Utiliza-se a centralização física quando a empresa existe em um único local e o *DW* é administrado por um departamento de tecnologia da informação. A distribuição física de um *DW* é utilizada caso a empresa possua vários locais físicos (instalações)

e os dados estejam em múltiplas instalações físicas, também com a administração feita por um departamento de tecnologia da informação. (MACHADO, 2000).

Para Machado (2000), esta arquitetura possibilita os usuários a utilizar visões corporativas de dados, que geralmente são requisitos de negócio; porém, este tipo de ambiente consome muito tempo de desenvolvimento e administração, além de seu custo de implementação ser muito elevado.

A figura 6 apresentada a seguir, permite uma melhor compreensão desta arquitetura.

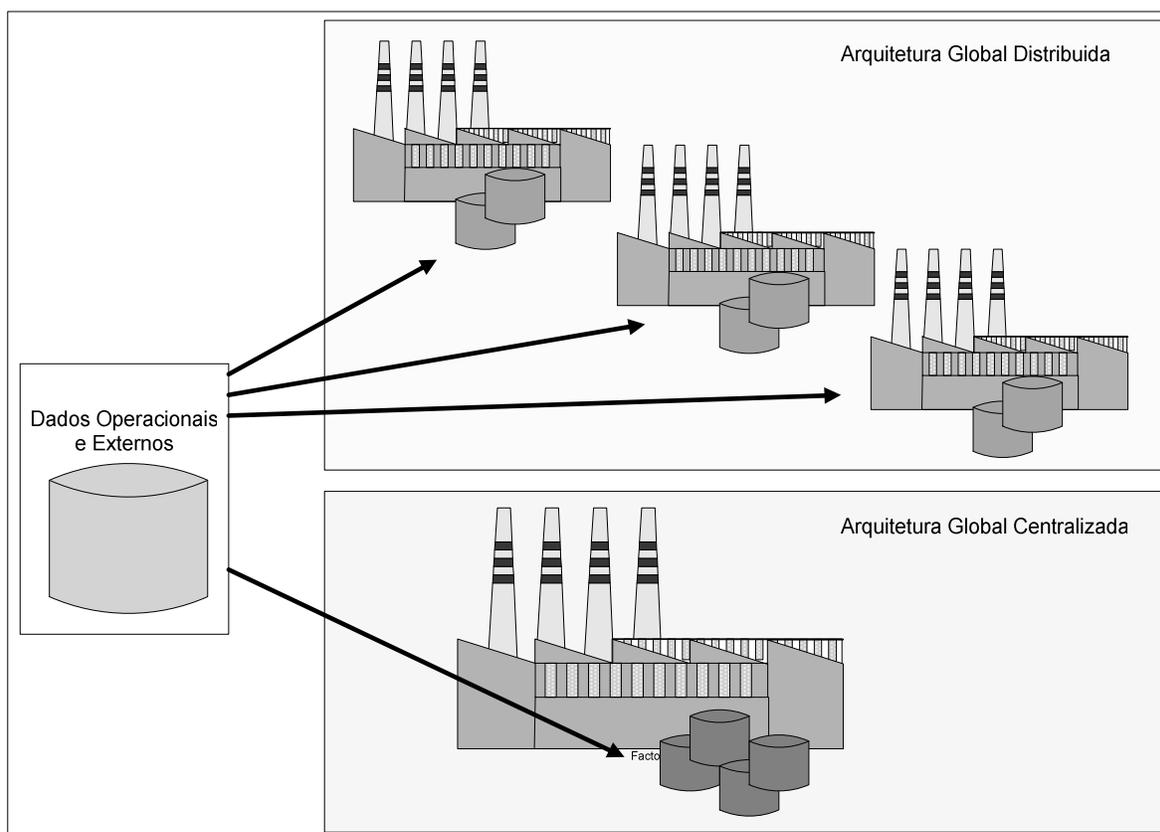


Figura 6 – Arquitetura Global.
Fonte: MACHADO, 2000.

2.2.5.2 Arquitetura de Data Mart Independente

A arquitetura independente implica em *data marts stand-alone* controlados por um grupo específico de usuários, e que atende somente às suas necessidades específicas e departamentais, sem foco corporativo algum. Sendo assim, não existe nenhuma conectividade

desses *Data Marts* com os *Data Marts* de outros departamentos ou áreas de negócio (MACHADO, 2000).

Esta arquitetura raramente tem impacto nos recursos de tecnologia da informação e sempre resulta em implementação rápida. Em contrapartida, possui o mínimo de interação corporativa, não permitindo nenhuma visão global do negócio. Este tipo de *Data Mart* está disponível somente ao pessoal do departamento específico “proprietário” do *Data Mart* (MACHADO, 2000).

2.2.5.3 Arquitetura de Data Mart Integrados

A arquitetura de *Data Marts* integrados é basicamente uma distribuição de implementação. Apesar de os *DM* serem implementados separadamente por grupos de trabalho ou departamentos, eles são integrados ou interconectados, provendo uma visão corporativa maior dos dados e informações. (MACHADO, 2000).

O alto nível de integração desta arquitetura é similar ao da arquitetura global, porém neste caso os usuários de um departamento podem acessar e utilizar os dados de um *DM* de outro departamento.

Em relação à arquitetura independente também apresenta vantagens, possuindo muito mais funções e capacidades de informação. Conseqüentemente, por meio desta integração, aumenta sensivelmente o nível de complexidade de requisitos.

Machado (2000) também comenta que: “neste caso a atuação da área de tecnologia da informação deve ser bem maior que na arquitetura independente, ficando sob sua responsabilidade o controle e administração dos *Data Marts*”. O lado positivo desta arquitetura é que aumenta a capacidade e qualidade de visão corporativa de informação.

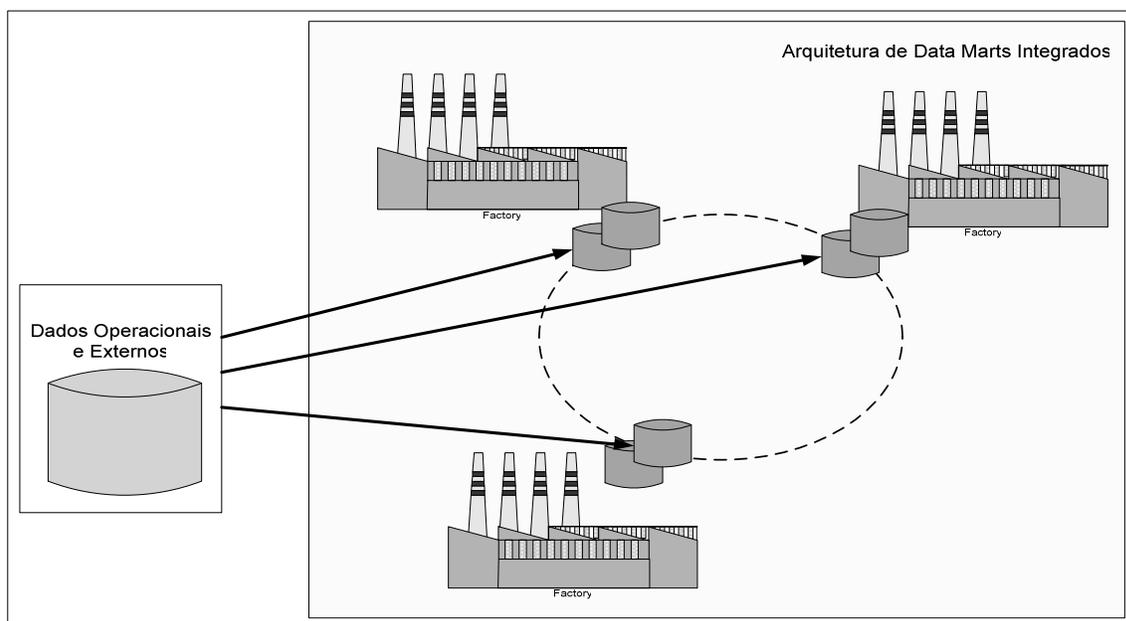


Figura 7 – Arquitetura de Data Marts Integrados.
Fonte: MACHADO, 2000.

2.2.6 Abordagens de Implementação

Vários tipos de implementação das arquiteturas apresentadas podem ser utilizados. Nas palavras de Machado (2000, p.36):

A opção por um tipo de implementação é influenciada por fatores como a infra-estrutura de tecnologia da informação, a arquitetura escolhida, o escopo da implementação, os recursos disponíveis e principalmente a necessidade ou não de acesso corporativo dos dados, retorno de investimento desejado e velocidade de implementação.

“As abordagens de implementação se referem à forma como o *DW* é projetado e implementado em relação às necessidades globais da empresa”. (TCHOLAKIAN, 2007) As três abordagens de implementação que serão descritas neste trabalho são a abordagem *top-down*, a abordagem *bottom-up* e a arquitetura *BUS*.

2.2.6.1 Abordagem “Top-Down”

Abordagem *Top-Down* é a abordagem proposta por Inmon (1997), a qual baseia-se em um *data warehouse* corporativo central, baseado no modelo relacional e totalmente normalizado.

O processo de extração, transformação e carga e, conseqüentemente, a área de estágio de dados são implementados de forma única e integrada. Este *DW* corporativo contém todos os dados organizacionais e tem o propósito de servir de base de dados para os diversos *data marts* departamentais implementados com base no modelo dimensional (MELLO, 2002).

Esta implementação requer maior planejamento e trabalho de definições conceituais de tecnologia completos antes de iniciar-se o projeto de *DW*. Além disso, envolve todas as pessoas da empresa que irão participar do projeto corporativamente, todos os departamentos (MACHADO, 2000).

“O processo inicia-se com a extração, a transformação e a integração dos sistemas operativos e dados externos para um *ODS* ou até diretamente das fontes operacionais. Em seguida os dados e metadados são transferidos para o *DW*”. (MACHADO, 2000, p. 36) Então, a partir do *DW* são extraídos os dados e metadados para os *Data Marts*.

Nesta abordagem, a base de dados é corporativa única, homogênea e totalmente integrada, porém traz consigo problemas relativos ao alto custo de implementação e demora na apresentação de resultados (VASCONCELLOS, 1999, BALLARD et al. 2001 apud MELLO 2002).

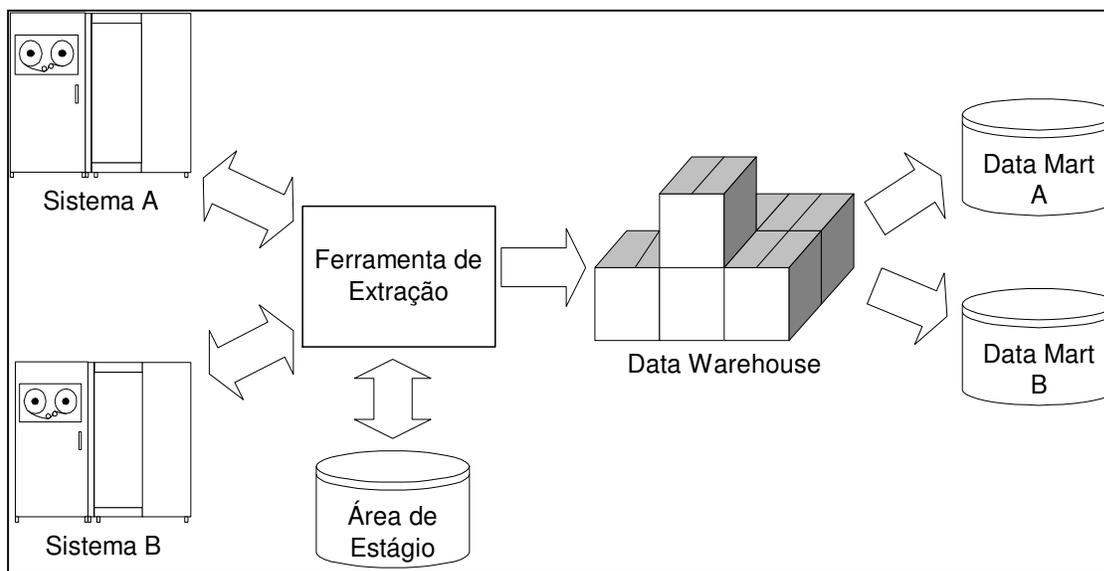


Figura 8 – Abordagem “*Top-Down*”.
 Fonte: SELL, 2001.

2.2.6.2 Abordagem Bottom Up

Segundo Machado (2000, p.39), o propósito da abordagem “*bottom-up*” é: “[...] a construção de um *DW* incremental a partir do desenvolvimento de *Data Marts* independentes”.

O processo dessa abordagem inicia-se com a extração, a transformação e a integração dos dados para um ou mais *Data Marts*, sendo estes baseados em um modelo dimensional (MACHADO, 2000).

A falta de um gerenciador que garanta padrões únicos de metadados é um dos grandes problemas desta implementação, sendo esta falta de padronização a responsável pela falha na elaboração incremental do *DW*. O planejamento, a monitoração e o estabelecimento de regras de desenvolvimento, ou seja, metodologias, podem minimizar a ocorrência de redundâncias de dados e inconsistência entre os *Data Marts*. Por outro lado, uma grande vantagem apresentada pela abordagem *bottom-up*, é a de possuir um rápido retorno do investimento, razão pela qual, tal abordagem vem se tornando cada vez mais popular. (MACHADO, 2000).

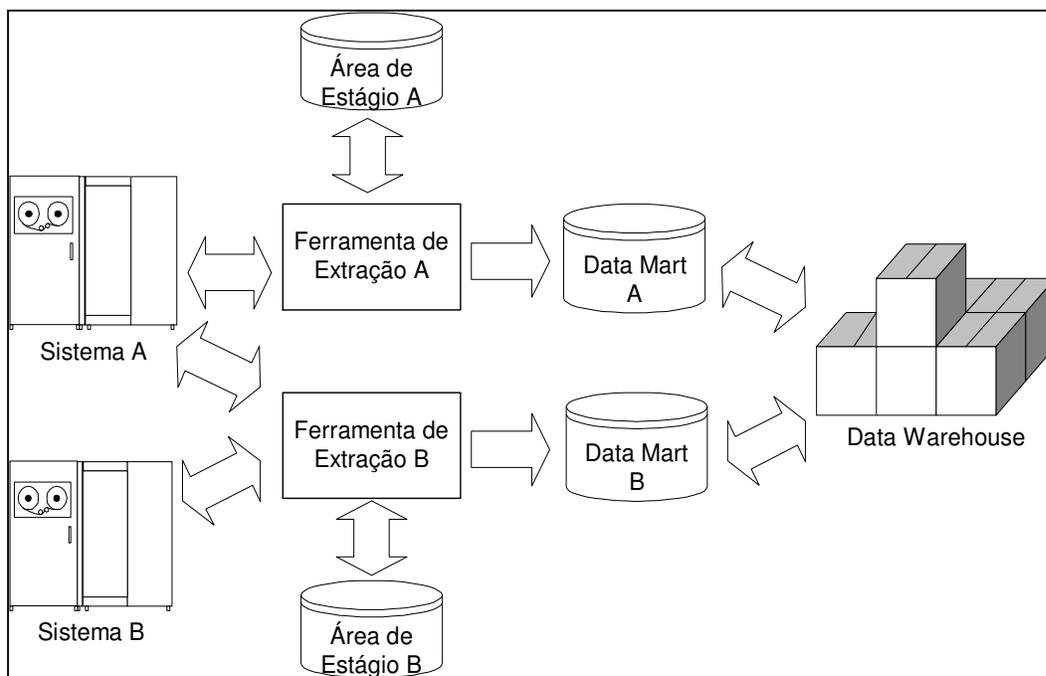


Figura 9 – Implementação *Bottom Up*.
Fonte: SELL, 2001.

2.2.6.3 Arquitetura BUS

A arquitetura *BUS* baseia-se no princípio de que os requisitos devem ser bem definidos e que uma integração entre os *data marts* deve ser planejada antes de começar a sua construção, sendo esta construção um processo contínuo de desenvolvimento (CENTENARO, 2003).

Tal implementação tem o propósito de integrar a abordagem “*top-down*” com a “*bottom-up*”. Nela efetua-se a modelagem de dados do *Data Warehouse* de visão macro, e em seguida implementa-se partes desse modelo, as quais são escolhidas por área de interesse e constituem os *Data Marts*. Cada *Data Mart* gerado a partir do macromodelo de dados do *DW* é integrado ao modelo físico do *DW* (MACHADO, 2000).

A garantia da consistência dos dados é a principal vantagem desta abordagem. Ela é obtida em virtude do modelo de dados para os *Data Marts* ser único, possibilitando realizar o mapeamento e o controle dos dados. (SOARES, 1998).

2.2.7 Modelo Dimensional

Em comparação à modelagem ER, a modelagem dimensional é mais simples, mais expressiva e mais fácil de ser entendida. Porém, por ser um conceito relativamente novo, ainda não há firmeza em seus detalhes de técnicas de desenvolvimento, como já existe no modelo ER. (FELÍCIO, 2005)

Machado (2000, p. 63), conceitua a modelagem dimensional como sendo: “uma técnica de concepção e visualização de um modelo de dados de um conjunto comum de medidas que descrevem aspectos comuns de negócios”. É usada especialmente para a sumarizar e reestruturar dados e apresentá-los em visões que suportem a análise dos valores desses dados.

Para Tcholakian (2007):

O modelo dimensional proporciona uma representação do banco de dados consistente com o modo como o usuário visualiza e navega pelo *DW*, combinando tabelas com dados históricos em séries temporais, cujo contexto é descrito através de tabelas de dimensões.

O modelo dimensional baseia-se em três elementos:

- Fatos
- Dimensões
- Medidas (variáveis).

Modelo Dimensional	Modelo Relacional - E/R
Padrão de estrutura mais fácil e intuitiva	Modelo mais complexo
Anterior ao Modelo de Entidade e Relacionamento, anos 60	Ênfase nos Bancos de Dados Relacionais, anos 70
Tabelas Fato e tabelas Dimensão	Tabelas que representam Dados e Relacionamentos
Tabelas Fato são o núcleo-normalizadas	Todas as tabelas são comumente normalizadas

Tabelas Dimensão são os pontos de entrada	As tabelas são indistintamente acessadas e de filtro inicial
Tabelas Dimensão opcionalmente normalizadas	Todas as tabelas são comumente normalizadas
Modelo mais facilmente " <i>joined</i> "	Maior dificuldade de " <i>join</i> " pelo número maior de tabelas
Leitura mais fácil do modelo por usuários não especializados	Maior dificuldade de leitura pelo usuário não especializado

Quadro 1 – Comparação entre modelo Relacional – E/R e modelo Dimensional.
Fonte: ZHAOFU, 2007.

2.2.7.1 Fatos

“Fato é uma coleção de itens de dados, composta de dados de medidas e de contexto”.(MACHADO, 2000, p.63).

Cada fato representa um item, transação ou evento do negocio e é utilizado para analisar o processo de negocio de uma empresa. O fato é representado por valores numéricos e implementado em tabelas de fato (*fact table*). (MACHADO, 2000)

Uma tabela fato é a principal tabela de um modelo dimensional em que as medições numéricas de desempenho da empresa estão armazenadas. (KIMBALL, 2002)

2.2.7.2 Dimensões

“Dimensões são conceitualmente os elementos que participam de um fato, assunto de negócios”. Elas determinam o contexto de um assunto de negócio. (MACHADO, 2000, p. 64).

Segundo Kimball (2002, p. 24) as tabelas dimensões estão sempre acompanhando uma tabela de fatos. Elas contêm descritores textuais da empresa. Machado (2000), ressalta também que dimensões normalmente não possuem atributos numéricos, pois são somente

descritivas e classificatórias dos elementos que participam de um fato. Um exemplo citado por ele para esclarecer o conceito é de um banco de dados que analisa as vendas de produtos. As dimensões que participam deste fato vendas de produtos são:

- Tempo
- Localização
- Clientes
- Vendedores
- Cenários (realizado, projetado)

2.2.7.3 Medidas (variáveis)

Medidas são os atributos numéricos que representam um fato. Elas representam a performance de um indicador de negócios relativo às dimensões que participam desse fato. (MACHADO, 2000)

Os números atuais são denominados de variáveis, por exemplo, medidas são o valor em reais das vendas, o número de unidades de produtos vendidos, a quantidade em estoque, o custo de venda, entre outros.

Determina-se uma medida pela combinação das dimensões que participam de um fato e estão localizadas como atributos de um fato.

O modelo dimensional permite a visualização de dados na forma de um cubo, no qual cada dimensão do cubo representa o contexto de um determinado fato, e a intersecção entre as dimensões representa as medidas do fato (Figura 10). Matematicamente o cubo possui apenas três dimensões, porém, no modelo dimensional a metáfora do cubo pode possuir quantas dimensões forem necessárias para representar um determinado fato (MACHADO, 2000 apud MELLO 2002). Isso acontece porque um modelo dimensional consiste em mais de três dimensões sendo por essa razão definido como um hipercubo. Porém visualizar graficamente em hipercubo é muito difícil, desta forma utiliza-se a referência a cubo (MACHADO, 2000, p. 65).

A figura 10 demonstra a representação de um fato vendas por meio de um cubo, no qual a medida é o volume de vendas, o que é determinado pela combinação de três dimensões: Localização, Produto e Tempo. A dimensão Localização e a dimensão Produto possuem dois níveis de hierarquia. Por exemplo, a dimensão Localização tem o nível cidade e o nível estado. Assim cada cubo possui o valor de medida de quantidade de vendas de celulares e *paggers*.

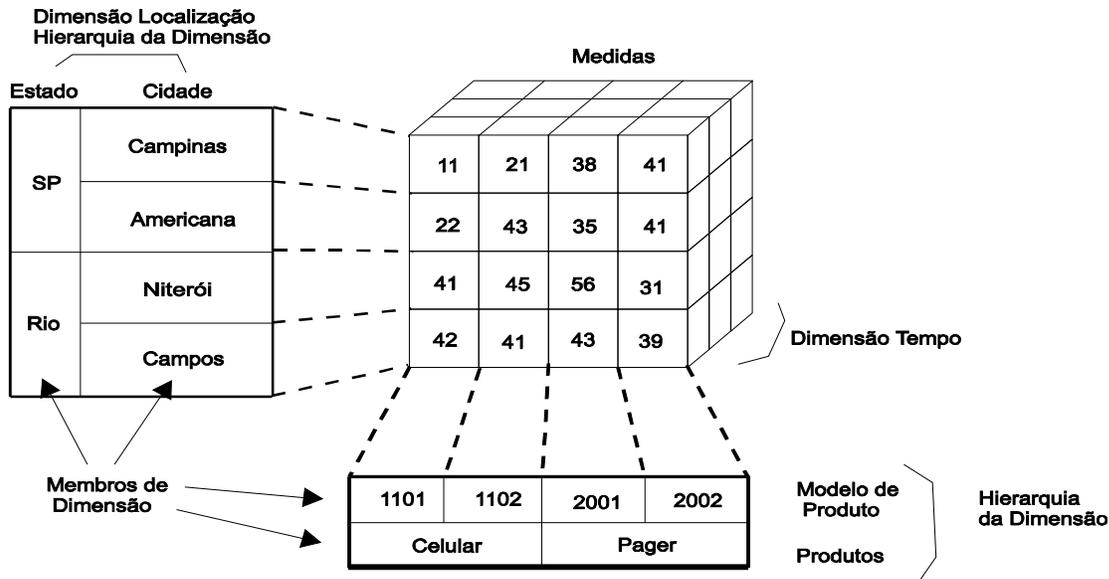


Figura 10 – Modelo Dimensional.
Fonte: MACHADO, 2000, p. 66.

2.2.7.4 Vantagens da Modelagem Dimensional

Para Tcholakian (2007), as vantagens da modelagem dimensional são:

- Facilmente entendido pelos técnicos, usuários finais e ferramentas de consulta;
- Processamento das consultas com maior eficiência;
- Flexível a mudanças: as tabelas existentes podem ser modificadas adicionando novos e inesperados fatos e dimensões, adicionar atributos e quebrar dimensões mais facilmente.
- As mudanças no modelo não impactam sobre as ferramentas de consulta;
- Existência de muitos modelos publicados e de técnicas para tratamento de problemas específicos para vários ambientes de negócio;
- Surgimento de softwares específicos para administração do modelo e processamento de consultas utilizando “agregados”.

2.2.7.5 Modelo Estrela

O modelo estrela é a estrutura básica de um modelo de dados dimensional. Este modelo é composto por uma grande entidade central denominada fato (*fact table*) e um conjunto de entidades menores denominadas dimensões (*dimension tables*), arranjadas ao redor desta entidade central, formando assim uma estrela, como mostra a figura 11. (MACHADO, 2000).

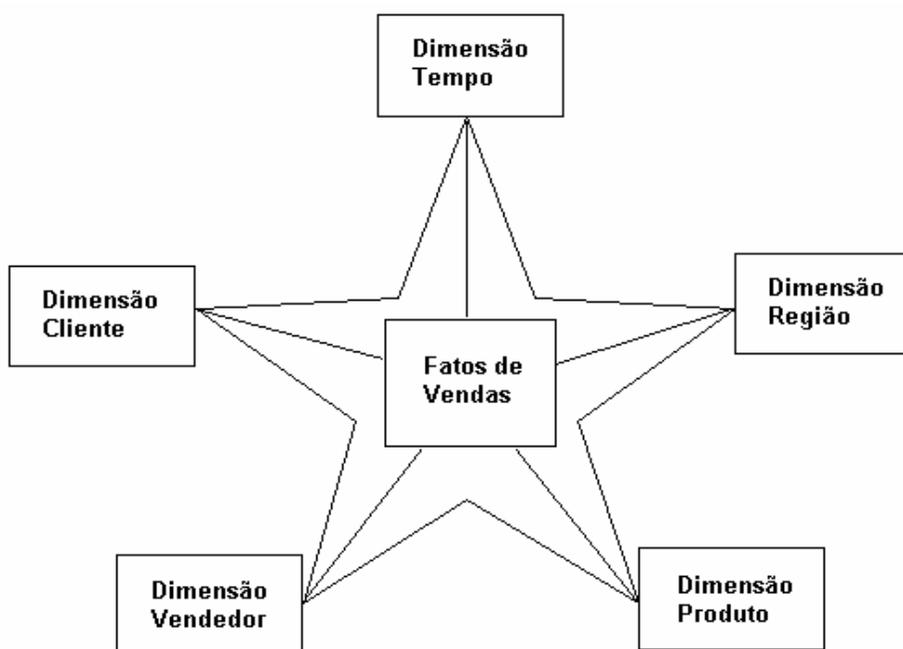


Figura 11 – Modelo Estrela.
Fonte: MACHADO, 2000.

O centro da estrela da figura 11 é o fato Vendas, ao seu redor estão dispostas as dimensões que participam deste fato: Tempo, Região, Produto, Vendedor e Cliente.

O relacionamento entre a entidade fato e as suas dimensões é uma simples ligação entre as duas entidades em um relacionamento de um para muitos no sentido da dimensão para o fato. Cada tabela de dimensão possui uma única ligação com a tabela de fato, a qual é feita através de chaves externas (TCHOLAKIAN, 2007).

Tabelas dimensionais guardarão em sua maioria informações textuais, as quais ajudam a definir um componente da dimensão do negócio, elas possuem dados sobre produtos, mercados, períodos de tempo. As tabelas de fatos, por sua vez, possuem os dados temporais direcionados ao foco do negócio, além de conter as informações decorrentes das tabelas de

dimensões, isso garante a precisão do acesso aos dados por meio de uma estrutura de chave completa. Desta forma, elimina as pesquisas em tabelas e resulta em maior desempenho possível (HARRISON, 1998 apud CENTENARO, 2003).

“Devido à forma simples de organizar os dados, o modelo dimensional além de agilizar o processamento das consultas, permite uma melhor visualização dos dados”. (TCHOLAKIAN, 2007).

2.2.8 Granularidade de Dados

A granularidade refere-se ao nível de detalhe ou resumo com o qual serão armazenados os dados no *data warehouse* (INMON, 1997). A sua definição afeta diretamente o volume de dados do *DW* (tamanho do banco de dados) e, portanto, a qualidade das consultas que poderão ser feitas. Quanto maior o nível de detalhes dos dados, menor o nível de granularidade, conseqüentemente, quanto menor o nível de detalhes, maior é a granularidade.

Para Machado (2000, p.43), “a granularidade é importante, pois servirá como parâmetro de avaliações do tipo: quanto maior o volume de dados, menor a performance”.

Uma granularidade alta garante maior rapidez nas consultas feitas, porém diminui a riqueza de informações que se pode extrair, enquanto uma menor granularidade possibilita a extração de qualquer informação, mas acarreta maior volume de dados, conseqüentemente maior tempo de resposta à consulta e maior investimento em *hardware* (CENTENARO, 2003).

Segundo Sell (2001), “o nível adequado de granularidade deve ser definido de tal forma que atenda as necessidades do usuário, tendo como limitação os recursos disponíveis, ou seja, é necessário encontrar um ponto de equilíbrio”.

Um *Data Warehouse* pode ser implementado em níveis duais de granularidade ao longo do tempo. Os dados mais recentes são armazenados em um nível de granularidade menor (maior detalhe) e, com o passar do tempo, a medida que os dados vão ficando obsoletos, estes são resumidos, criando um nível de granularidade maior. Por exemplo, pode-se definir que dados terão grão mensal para os anos mais recentes e, grão anual, para os anos menos recentes. Somente

quando o volume de dados for relativamente pequeno, não serão necessários níveis duais de granularidade (INMON, 1997).

2.3 PROSPECÇÃO DE DADOS

Imaginar, debater e modelar o futuro são atividades muito antigas. Para descrever o futuro há necessidade de saber mais do que já é conhecido. Compartilhar dúvidas e permitir a entrada de novas idéias e atitudes, bem como encontrar linguagem e crença comuns, são requisitos essenciais para que possa ser estabelecida uma linha de pensamento que possibilite construir o caminho da mudança.

“Estudos prospectivos constituem poderosos auxiliares do planejamento e do gerenciamento dos níveis de incerteza, porém precisam estar inseridos em um contexto planejado, isto é, estar embasados em diretrizes e necessidades pré-estabelecidas”. (PROSPECÇÃO TECNOLÓGICA EM ENERGIA, 2005, p.33).

Como prospecção é um assunto novo, de muita pesquisa e interesse nos últimos anos, ainda não há um consenso de suas abordagens, porém, já é possível identificar algumas delas, mesmo que muitas vezes usadas sem distinção:

- *Foresight* – pode ser definido como “um processo pelo qual pode-se chegar a um entendimento mais completo das forças que moldam o futuro a longo prazo e que devem ser levadas em consideração na formulação de políticas, planejamento e tomada de decisão. *Foresight* inclui meios qualitativos e quantitativos para monitorar pistas e indicadores das tendências de desenvolvimento e seu desenrolar, e é melhor e mais útil quando diretamente ligado à análise de políticas e suas implicações [...]”.(COATES, 1985 apud CGEE, 2005, p.33).
- *Forecasting* – Trabalha com informações de evolução histórica, modelagem matemática de tendências e análise de projeção de situações futuras, executadas normalmente de forma periódica.
- Monitoração Tecnológica - acompanhamento de evolução e identificação de sistemas de mudança, realizados de forma mais ou menos sistemática e contínua. (MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO EXTERIOR, 2007b).

2.3.1 Vantagens e desvantagens dos diferentes métodos e técnicas de prospecção

A realização de estudos prospectivos, em âmbito internacional, é relativamente recente e decorre de um contexto de grandes mudanças no cenário internacional, especialmente no que diz respeito à globalização da economia e à aceleração das mudanças tecnológicas. A capacidade de antecipar tornou-se um elemento de extrema importância para garantir a competitividade de empresas e países. Dessa forma, novas técnicas e ferramentas foram surgindo no decorrer dos últimos anos, tentando utilizar os conhecimentos tácitos e explícitos existentes para tentar não prever como o futuro será, mas sim entender quais são os fatores condicionantes e identificar os melhores caminhos para a construção do futuro pretendido (CGEE, 2007).

No Brasil, a realidade não é diferente. Estudos voltados para compreensão de forças que orientam o futuro são bastante recentes. Até pouco tempo atrás o foco era o presente ou o estado de determinada tecnologia, não olhavam para o futuro e não tinham como foco construir uma ferramenta de apoio à decisão (CGEE, 2007).

A utilização de mais de um método ou técnica na realização de um estudo prospectivo é uma necessidade apontada por diversos autores. Isto ocorre pelas dificuldades inerentes a esse tipo de atividade e pelo fato de nenhum método atender a todas as necessidades envolvidas. De maneira geral, os métodos quantitativos são combinados com os qualitativos e os conhecimentos explícitos acrescentam-se aos conhecimentos tácitos na busca de complementaridade ou de visões diferenciadas (CGEE, 2007).

É importante ressaltar que cada método, técnica ou ferramenta apresenta vantagens e desvantagens. Por exemplo, os métodos quantitativos deparam-se com a necessidade de séries históricas confiáveis ou da existência de dados padronizados. Já os métodos qualitativos, muitas vezes, têm problemas devido ao limite do conhecimento dos especialistas, de suas preferências pessoais e parcialidades (CGEE, 2007).

Assim sendo, a escolha correta da metodologia a ser utilizada está fortemente ligada a qualidade dos resultados dos estudos. Especialistas na área recomendam o uso de mais de uma técnica, método ou ferramenta (CGEE, 2007).

As vantagens e desvantagens de alguns métodos e técnicas serão apresentadas no Quadro 2.

Método	Pontos Fortes	Pontos Fracos
Monitoramento & Sistemas de Inteligência	<p>Fornecer uma grande quantidade de informação, oriunda de um diversificado número de fontes.</p> <p>Pode resultar no excesso de informação, não seletiva e não analisada.</p>	<p>Pode ser usada no início da prospecção, como contextualização inicial do tema e, ao final, como forma de manter os temas críticos permanentemente atualizados.</p> <p>As informações, por si, estão mais relacionadas ao passado e ao presente, portanto, só a análise pode dar a perspectiva do futuro.</p>
Tendências	<p>Fornecer previsões substanciais, baseadas em parâmetros quantificáveis.</p> <p>É particularmente precisa no curto prazo.</p>	<p>Requer dados históricos consistentes e coletados ao longo de um período razoável de tempo.</p> <p>Só funciona para parâmetros quantificáveis.</p> <p>É vulnerável a mudanças bruscas e descontinuidades.</p> <p>Pode ser perigosa quando se faz projeções de longo prazo.</p>
Opinião de Especialistas	<p>Permite a identificação de muitos modelos e percepções internalizados pelos especialistas que os tornam explícitos.</p> <p>Permite que a intuição encontre espaço na prospecção.</p> <p>Incorpora à prospecção aqueles que realmente entendem da área que está sendo prospectada.</p>	<p>Muitas vezes é difícil identificar os especialistas.</p> <p>Muitas vezes as projeções que fazem são erradas ou preconceituosas.</p> <p>Às vezes são ambíguas e divergentes entre especialistas da mesma área.</p>
Cenários	<p>Apresentam retratos ricos e complexos dos futuros possíveis.</p> <p>Incorporam uma grande variedade de informações qualitativas e quantitativas produzidas por meio de outros métodos de prospecção.</p> <p>Normalmente incorporam elementos que permitem à quem decide definir a ação.</p>	<p>Algumas vezes são mais fantasia do que prospecção, quando se identifica o futuro desejado sem considerar as restrições e barreiras que se tem que ultrapassar para chegar até lá.</p>

<p>Métodos descritivos e matrizes; Métodos estatísticos; Modelagem e simulação</p>	<p>Modelos podem exibir comportamento de sistemas complexos simplesmente pela separação de aspectos importantes dos detalhes desnecessários.</p> <p>Alguns sistemas oferecem possibilidades de incorporação do julgamento humano.</p> <p>Fornecem excelentes percepções e análises sobre o comportamento de sistemas complexos.</p> <p>Possibilitam o tratamento analítico de grandes quantidades de dados.</p>	<p>Técnicas sofisticadas podem camuflar falsos pressupostos e apresentar resultados de má qualidade.</p> <p>Alguns modelos e simulações contêm pressupostos essenciais que devem ser testados para ver sua aplicabilidade ao estudo.</p> <p>Todos os modelos requerem adaptações antes de serem usados e devem ser validados.</p> <p>O sucesso na previsão de um comportamento histórico não garante a previsão bem sucedida do futuro.</p> <p>As fontes de dados usadas em data e <i>text mining</i> devem ter um certo grau de padronização para que a análise não induza a erros.</p>
<p>Criatividade</p>	<p>Aumenta a habilidade de visualizar futuros alternativos. Diminui as visões preconcebidas dos problemas ou situações.</p> <p>Encoraja a criação de um novo padrão de percepção.</p> <p>É excelente para ser usado no início do processo.</p>	<p>O coordenador ou líder do grupo deve ter capacidade de condução do processo para evitar descaminhos.</p> <p>Se mal conduzido, pode levar à futurologia e descrédito do processo.</p>
<p>Avaliação / Decisão</p>	<p>Ajudam a reduzir a incerteza no processo decisório.</p> <p>Auxiliam no estabelecimento de prioridades quando há um número grande de variáveis a serem analisadas.</p>	<p>É preciso ter consciência que os métodos reduzem, mas não eliminam a incerteza no processo decisório.</p>

Quadro 2 - Vantagens e desvantagens dos métodos de prospecção.
Fonte: COELHO, 2003 apud CGEE, 2007.

2.3.2 Análise Exploratória de Dados

Problemas complexos exigem a coleta de muitos dados e observações, contendo cada uma delas, inúmeros atributos e variáveis. A análise exploratória é chamada assim, pois utiliza métodos estatísticos para explorar informações destes dados (JOHNSON E WICHERN, 1998).

A análise exploratória de dados (AED) do termo *Exploratory Data Analysis*, é uma técnica da área de estatística descrita por John W. Tukey em 1977, e visa obter o máximo de informações ocultas entre os dados (HOAGLIN; MOSTELLER; TUKEY, 1992).

A análise exploratória de dados (AED) tem por objetivo tornar mais clara a descrição dos dados por meio da construção de indicadores, sua distribuição, comparação e análises. A AED procura estabelecer relações existentes entre as diversas variáveis disponíveis, fazendo análises e conclusões relevantes sobre o assunto em questão. Dependendo dos dados disponíveis, a AED permite análises geográficas e temporais, por ator ou setor. Ela também envolve visualização gráfica dos dados, indicadores e índices, além da procura de boas descrições de dados, a fim de ajudar o analista a desenvolver algumas hipóteses sobre o assunto e modelos apropriados para tais dados, o que contribui para identificar padrões, comportamentos, relações e dependências (BAILEY E GATRELL, 1995).

As técnicas de visualização são geralmente gráficas. Elas incluem, entre outras coisas, Box-Plot, histogramas, gráficos de probabilidades e gráficos residuais (EDA, 2007, tradução nossa).

Este trabalho apresenta exclusivamente o método de histogramas, pois tal método mostra graficamente o resumo de um conjunto de dados distribuídos.

2.3.2.1 Olap

A sigla *OLAP* foi criada como uma ligeira variação de um termo tradicional em bancos de dados, *OLTP* (do inglês *On Line Transaction Processing*, ou processamento de transações on-line). *OLAP* é a sigla em inglês para *Online Analytical Processing*, ou

processamento analítico on-line. É uma abordagem tecnológica para gerar respostas rápidas a consultas analíticas de natureza tipicamente dimensional. A tecnologia *OLAP* é uma parte do processo de *Business intelligence*, que também inclui *Data warehouse*, *ETL* e *Data mining*.

Com a tecnologia *OLAP* faz-se possível manipular, examinar e visualizar, de diferentes perspectivas e com interatividade, grandes quantidades de dados contidos em bases de dados multidimensionais. Tal tecnologia também envolve a análise de relacionamentos mais complexos sobre centenas de dados contidos em bases de dados, com o intuito de descobrir tendências e padrões. Algumas aplicações típicas de *OLAP* são relatórios de negócios, *marketing*, relatórios gerenciais e previsão, relatórios financeiros e áreas similares.

2.3.2.1.1 Operadores OLAP

Para Machado(2000) os modelos Multidimensionais, devem suportar as operações OLAP, tais como:

- *Drill Down* – operador que deixa mais detalhada a visão dos dados, nos níveis hierárquicos, ou seja, muda a granularidade dos dados. Ex.: País → Região.
- *Roll Up* – contrário do *Drill Down*. Ex.: Região → País.
- *Slice* – operador usado para navegar pelas dimensões, cortando o cubo (Figura 12), porém mantendo a perspectiva de visualização dos dados.
- *Dice* – gira o cubo, permitindo a mudança da perspectiva de visão, mudando a rotação do cubo e invertendo as dimensões.

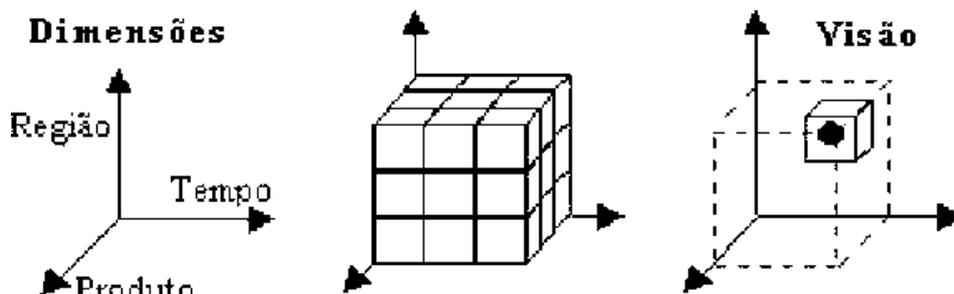


Figura 12 – Dimensões do cubo.
Fonte: Adaptado de TCHOLAKIAN, 2007.

Os caminhos de navegação são determinados pelas hierarquias de dimensão. Pode-se utilizar as operações combinando operadores, por exemplo *Drill Down* com *Slice*.

2.3.2.2 Data Mining

“O *Data Mining* é uma tecnologia usada para revelar informação estratégica escondida em grandes massas de dados”. (BOGHI, SHITSUKA. 2002).

Com o passar do tempo, os bancos de dados das empresas, tornam-se cada vez mais preciosos, pois acumulam enormes quantidades de dados, que agrupados e ordenados, geram padrões e informações precisas para tomada de decisão. Porém em alguns casos, fica difícil encontrar tais informações, pois algumas delas encontram-se em bancos diferentes como se pode ver na Figura 13. “A mineração de dados é uma área extensa, interdisciplinar e envolve o estudo de diversas técnicas”. (VIANA, 2004, v. 10, p. 16).



Figura 13 - Técnicas utilizadas pela mineração e dados.
Fonte: Adaptado de VIANA, 2004.

Para Boghi e Shitsuka (2002, p. 193), essa tecnologia é usada em áreas diversas, como análise de riscos, análise de dados e marketing direcionado.

Para se encontrar padrões formados pelos dados ordenados, é preciso que os dados brutos sejam simplificados, desconsiderando aquilo que é isolado, específico e privilegiar aquilo que é genérico, corriqueiro. Um supermercado que tenha vendido uma quantidade imensa de um determinado produto em uma única data para um cliente, pode apenas significar que esse cliente

procurava em particular, uma grande quantidade desse produto naquele exato momento. O que não implica, em nenhuma tendência de mercado.

2.4 CONCLUSÃO

Certamente que o conhecimento diminui a dúvida, e por isso houve aqui uma tentativa de englobar de maneira sucinta, o maior número de informações, sobre o processo de tomada de decisão, por meio de ferramentas e técnicas do *Business Intelligence*. Dados e informações soltos, não revelam muito, mas se ordenados e pareados com outros, apresentam facilmente algumas tendências, o que levam a agir e viver melhor agora, seguindo a visão do amanhã.

3 MODELAGEM

Um sistema de informação é a junção de processos, pessoas, dados, softwares e tecnologias, que se integram para adicionar valor à empresa na qual esse sistema será utilizado. Para se desenvolver softwares com eficiência e rapidez, é necessário um grupo de pessoas e ferramentas adequadas.

A modelagem é uma das principais atividades que levam à implantação de um software bem sucedido. Ela é uma simplificação da realidade, onde construímos modelos para comunicar a estrutura e o comportamento desejados do sistema, para visualizar e controlar a arquitetura do sistema, para gerenciar os riscos, e para uma melhorar a compreensão em sua totalidade do sistema que estamos desenvolvendo. (BOOCH; RUMBAUGH & JACOBSON, 2000).

3.1 UML

Surgida na metade da década de 90, a *UML*, que é a abreviação de Linguagem de Modelagem Unificada (*Unified Modeling Language*), veio com o objetivo de formalizar e padronizar os modelos de um projeto de software por meio de uma notação gráfica.

A *UML* foi criada por Grady Booch, Ivar Jacobson e James Rumbaugh, que são conhecidos como “os três amigos”. Possuidores de um vasto conhecimento na área de modelagem, foram eles que desenvolveram a *UML*, juntando as três mais conceituadas metodologias de modelagem orientada a objetos e adicionado novos conceitos e visões da linguagem.

A *UML* é uma linguagem visual para modelar sistemas orientados a objetos. Isso quer dizer que a *UML* é uma linguagem constituída de elementos gráficos (visuais) utilizados na modelagem que permitem representar os conceitos do paradigma da orientação a objetos. (BEZERRA, 2002, pg.14).

A *UML* pode ser utilizada para a modelagem de sistemas não importando o tipo de linguagem de programação ou processo de desenvolvimento utilizado, ou seja, ela é uma

linguagem livre, o que se torna um fator importante frente à necessidade de diferentes sistemas de software precisarem de diferentes abordagens de desenvolvimento. (BEZERRA, 2002, p. 14).

Segundo Booch (2000, p.13), a “UML é apenas uma linguagem e, portanto, é somente uma parte de um método para desenvolvimento de software”.

3.1.1 Visões de um sistema

O desenvolvimento de um software raramente é compreendido no primeiro contato. Mesmo os desenvolvedores e programadores de sistemas dividem os projetos em subsistemas para um melhor entendimento. Cada divisão é descrita por um ou mais modelos alternativos e notações gráficas que servirão para a criação de visões dos elementos do modelo do software.

“Na prática porém, aparecerá um pequeno número de combinações comuns, que são consistentes com as cinco visões mais úteis da arquitetura de um sistema complexo de software”.(BOOCH; RUMBAUGH & JACOBSON, 2000, p. 25).

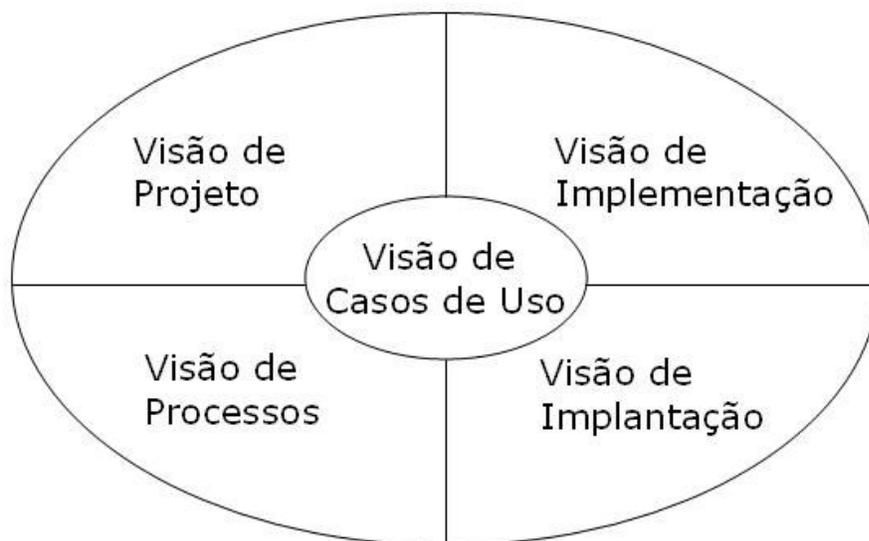


Figura 14 – Visões de um sistema de software.
Fonte: BEZERRA, 2002, p. 16.

Cada uma destas visões, engloba diferentes características do software, e são detalhadas da seguinte forma:

- Visão de Caso de Uso: descreve o sistema com um olhar externo dos relacionamentos entre o software e os agentes externos ao software. É uma das primeiras visões a ser criada, e serve de base o desenvolvimento das demais visões do sistema.
- Visão de Projeto: detalha as funcionalidades encontradas para a implementação, a partir de uma visão externa ao sistema.
- Visão de Implementação: Utilizado mais por desenvolvedores, essa visão detalha a implementação de versões do sistema, construídas com o agrupamento de componentes (módulos) e subsistemas.
- Visão de Implantação: descreve a distribuição física do sistema (computadores, impressoras, e outros periféricos) e a conexão entre eles.
- Visão de Processo: mostra o desempenho, aspectos de paralelismo (concorrência) e sincronização do sistema.

3.2 DIAGRAMAS DA UML

Quando se utiliza a linguagem *UML* como sustentação à modelagem, necessita-se criar vários documentos, gráficos ou textuais.

Para Bezerra (2002, p.16):

[...] esses documentos são denominados artefatos de software, ou simplesmente artefatos. São os artefatos que compõem as visões do sistema. Os artefatos gráficos produzidos durante o desenvolvimento de um sistema de software são definidos através da utilização dos diagramas da *UML*.

Todos estes diagramas se relacionam de alguma forma (Figura 15), e cada um permite a visão parcial do software sendo modelado.

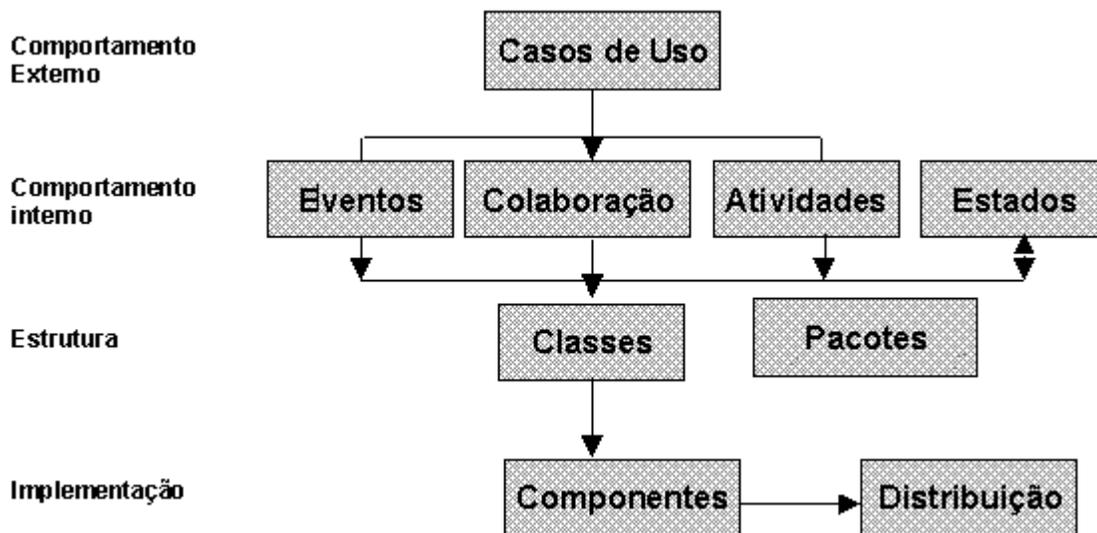


Figura 15 - Relações entre os diagramas da UML.
Fonte: DEBONI, 2007.

3.2.1 Diagramas de casos de uso (Use-Cases)

Para Carlson (2002, p.34), “os diagramas de caso de uso modelam o comportamento de um sistema, subsistema ou de uma classe, ilustrando as relações entre um conjunto de casos de uso e seus atores”. Como demonstrado na Figura 16.

Esses comportamentos são, na verdade, recursos para visualizar, especificar, construir e documentar o que se espera do sistema na fase de captação e análise de requisitos. Um caso de uso nada mais é, do que uma funcionalidade do sistema em sua totalidade. Como por exemplo, o processamento de empréstimos em um banco, é um caso de uso central. (BOOCH; RUMBAUGH & JACOBSON, 2000).

O diagrama de casos de uso (DCU) corresponde a uma visão externa do sistema e representa graficamente os atores, casos de uso e relacionamentos entre esses elementos. O diagrama de casos de uso tem o objetivo de ilustrar em um nível alto de abstração quais elementos externos interagem com que funcionalidades do sistema. Nesse sentido, a finalidade de um DCU é apresentar um tipo de “diagrama de contexto” que representa os elementos externos de um sistema e as maneiras segundo as quais eles utilizam. (BEZERRA, 2002, p.57).

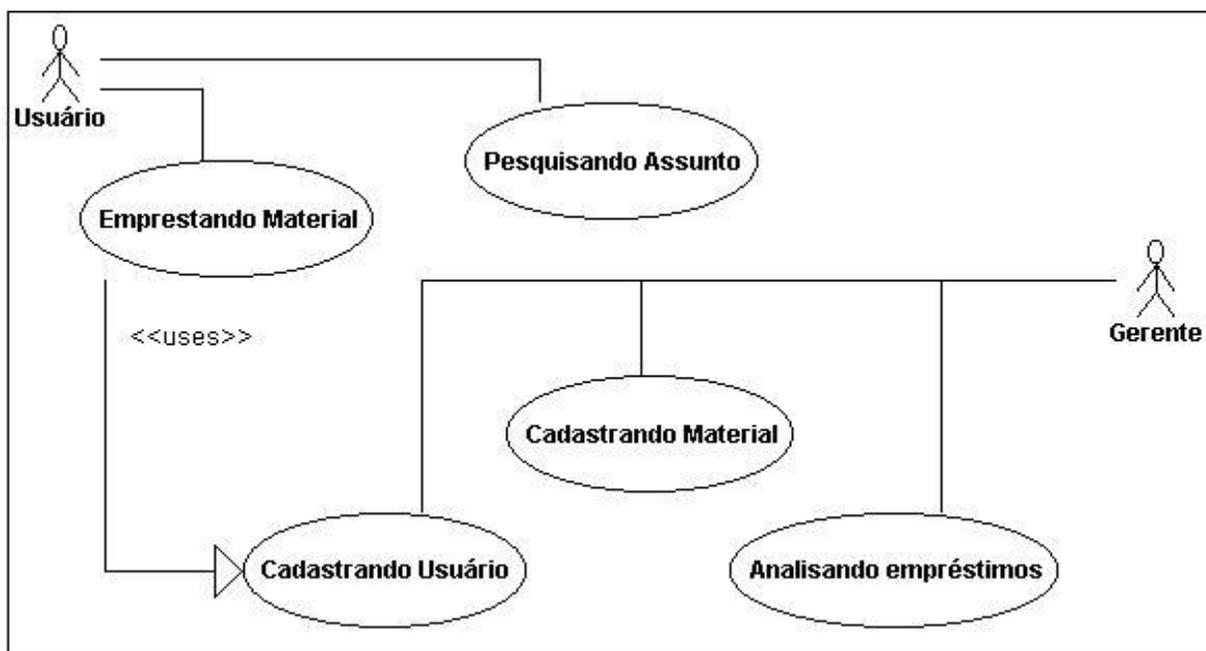


Figura 16 - Diagrama de caso de uso
 Fonte: DEBONI, 2007.

3.2.1.1 Atores

Um ator representa um conjunto de atividades coerentes que um ser humano, um periférico ou até outro software desempenharia com o software.

“Uma instância de um ator, portanto, representa uma interação individual com o sistema de uma maneira específica. Embora você utilize atores em sua modelagem, os atores não são de fato, parte do sistema. Eles residem fora do sistema”. (BOOCH; RUMBAUGH & JACOBSON, 2000, p. 221).

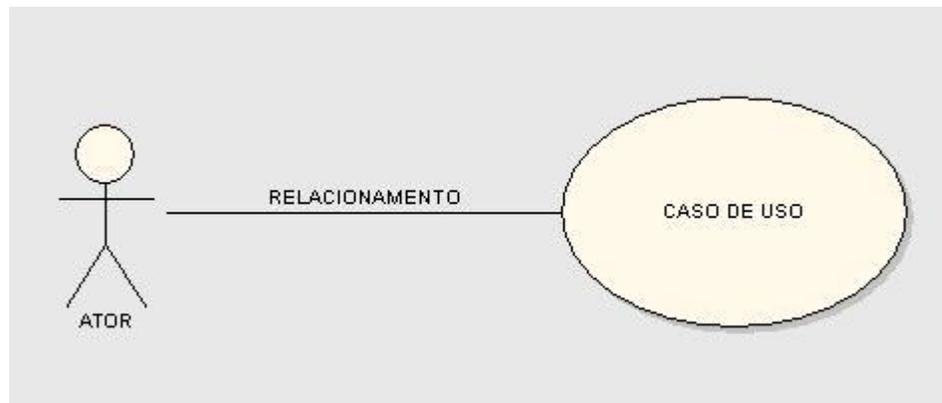


Figura 17 - Ator, caso de uso e relacionamento.
 Fonte: Elaborado pelos Autores, 2007.

3.2.2 Diagramas de Classe

O diagrama de classe oferece a visualização das especificações para as classes de software e de interfaces de uma aplicação. (Larman, 2002).

Esses diagramas são os que aparecem mais vezes na modelagem de sistemas orientados a objetos, e servem para ilustrar um conjunto de classes, interfaces, colaborações e suas respectivas interações.

“Use os diagramas de classe para fazer a modelagem da visão estática do projeto de um sistema. Na maioria dos casos, isso envolve a modelagem do vocabulário do sistema, a modelagem de colaboração ou a modelagem de esquemas. [...] Os diagramas de classe são importantes não só para a visualização, a especificação e a documentação de modelos estruturais, mas também para a construção de sistemas executáveis por intermédio de engenharia de produção reversa”.(BOOCH; RUMBAUGH & JACOBSON, 2000, p. 104).

É comum encontrar os seguintes itens, nos diagramas de classe, como na Figura 18:

- Classes;
- Interfaces;
- Colaborações;
- Relacionamentos.

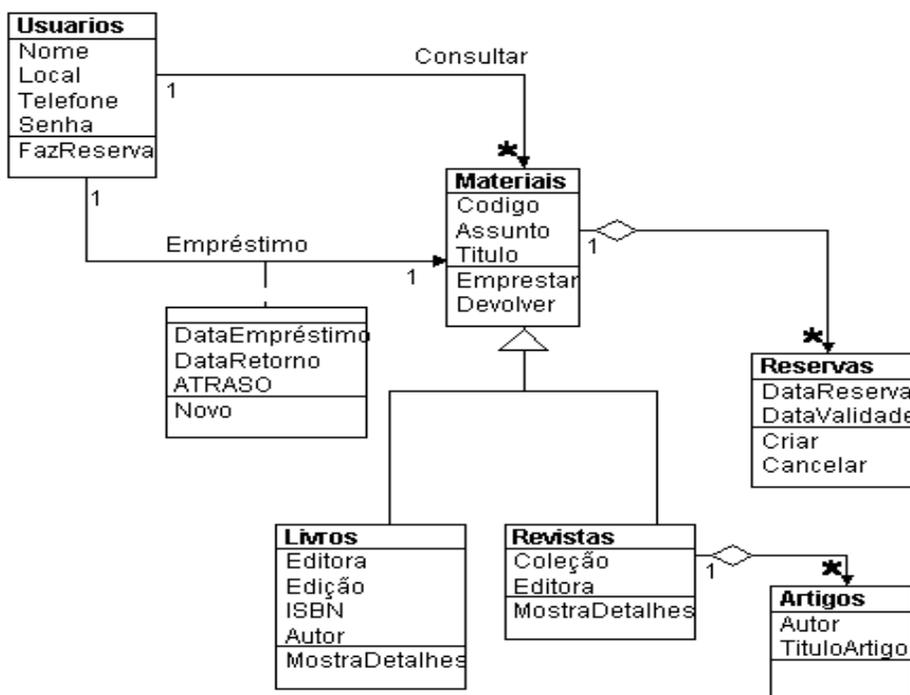


Figura 18 - Diagrama de Classes.
Fonte: DEBONI, 2007.

3.2.3 Diagramas de atividades

Um diagrama de atividades mostra os aspectos dinâmicos do sistema, como o fluxo de controle entre as atividades.

“As atividades efetivamente resultam em alguma ação, formada pelas computações executáveis atômicas que resultam em uma mudança de estado do sistema ou o retorno de um valor”. (BOOCH; RUMBAUGH & JACOBSON, 2000, p. 257).

São características dos diagramas de Atividades, as Transições, os Objetos e os Estados de atividade e de ação, como se pode ver na Figura 19.

Para Bezerra (2002, p. 228):

“[...] o leitor que trabalha com computação a mais de uma década deve se lembrar de uma ferramenta bastante utilizada no passado: o *fluxograma*. Na verdade, o diagrama de atividade pode ser visto como uma extensão dos fluxogramas. Além de possuir toda semântica existente em um fluxograma (com notação ligeiramente diferente), o diagrama de atividade possui notação para representar ações concorrentes (paralelas), juntamente com a sua sincronização”.

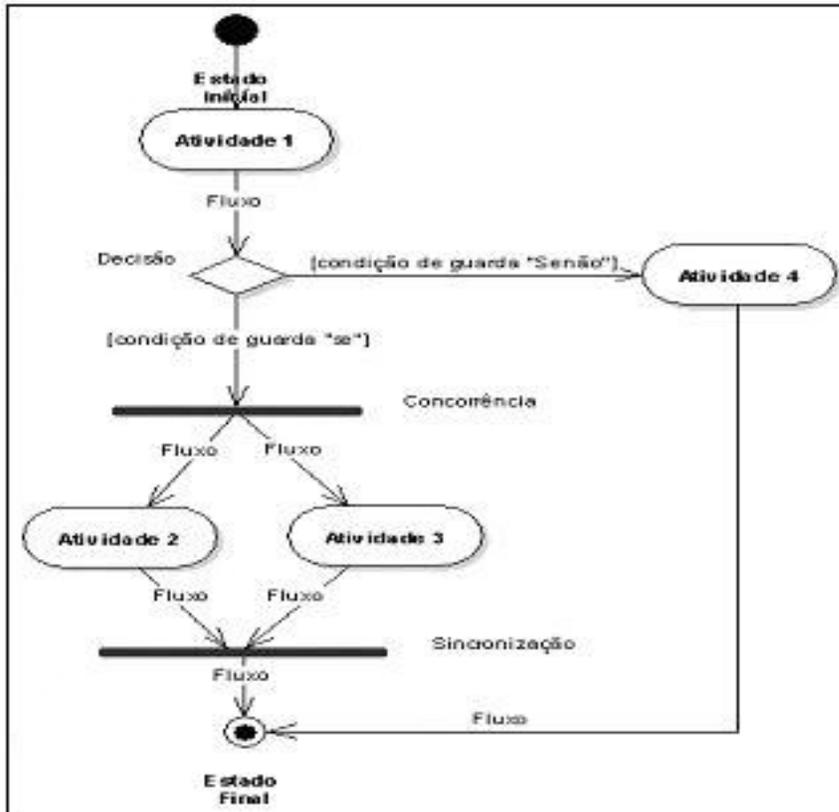


Figura 19 - Diagrama de Atividade
 Fonte: BIZZARRO, 2007.

3.2.4 Diagramas de seqüência

A *UML* disponibiliza uma visualização das interações entre atores e as ações iniciadas por eles, o diagrama de seqüência, que assim como os outros diagramas, possui um conjunto de elementos gráficos, como visto na Figura 20. “O enfoque do diagrama de seqüência está em como as mensagens são enviadas no decorrer do tempo”.(BEZERRA, 2002, p.148).

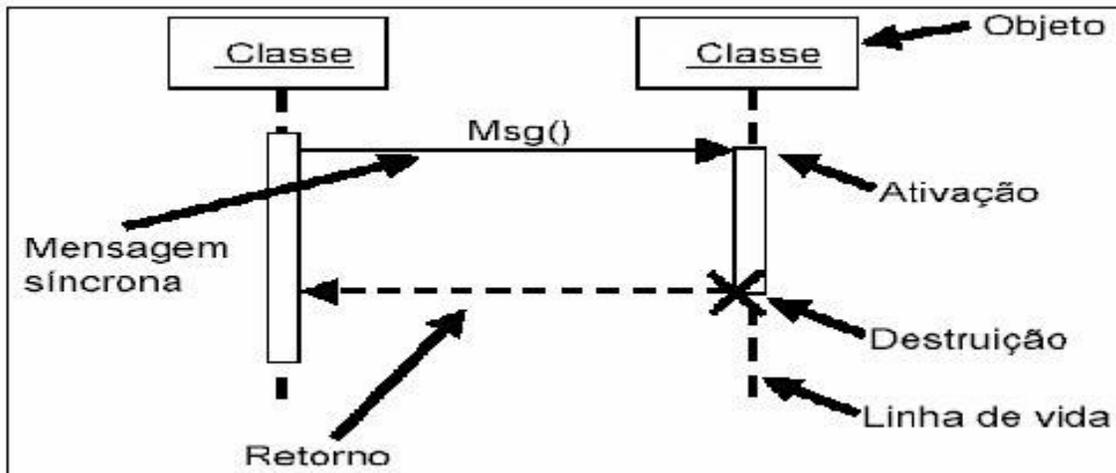


Figura 20 - Diagrama de Seqüência.

Fonte: BIZZARRO, 2007.

O Diagrama de Seqüência pode ser compreendido da seguinte forma, conforme Macoratti (2007):

- Objetos são representados por retângulo com seus nomes sublinhados.
- As flechas horizontais são rotuladas com as mensagens.
- A ordem das mensagens no tempo é indicada pela posição vertical, com a primeira mensagem aparecendo no topo.
- As linhas de vida dos Objetos são representadas por linhas verticais tracejadas.
- As interações entre Objetos são indicadas por flechas horizontais que são direcionadas da linha vertical que representa o Objeto cliente para a linha que representa o Objeto fornecedor.
- A numeração é opcional e baseada na posição vertical.

“Um diagrama de seqüência do sistema mostra, para uma particular seqüência de eventos dentro de um caso de uso, os atores externos que interagem diretamente com o sistema, o sistema (como uma caixa-preta) e os eventos de sistema que os atores geram”. (LARMAN, 2002, p.143).

3.2.5 Diagrama de robustez

É o diagrama que melhor auxilia o desenvolvedor a aprimorar as especificações dos casos de uso. Disponibiliza a rastreabilidade (*traceability*) das necessidades funcionais, estabelecidas nos casos de uso, para o detalhamento do projeto detalhado e implementação.

É formado pelos seguintes elementos (Figura 21):

- Objeto de Limite (interface) – *boundary* – mostra a relação entre o ator e o início do fluxo das atividades do caso de uso.
- Objeto de Controle – *control* – estão geralmente relacionados com as regras de negócio, o processamento e ou associados ao modelo de domínio do problema.
- Objeto de Entidade – *entity* – objetos que armazenam informações sobre o domínio do problema, e também responsáveis por realizar algum tipo de persistência.



Figura 21 - Elementos do diagrama de robustez.

Fonte: Elaboração dos Autores, 2007.

O diagrama de robustez, apesar de não ser um diagrama oficial da *UML*, preenche a lacuna entre o caso de uso e o diagrama de seqüência.

Os atores somente se comunicam com objetos limite, que por sua vez só se comunicam com atores e controladores (Figura 22).

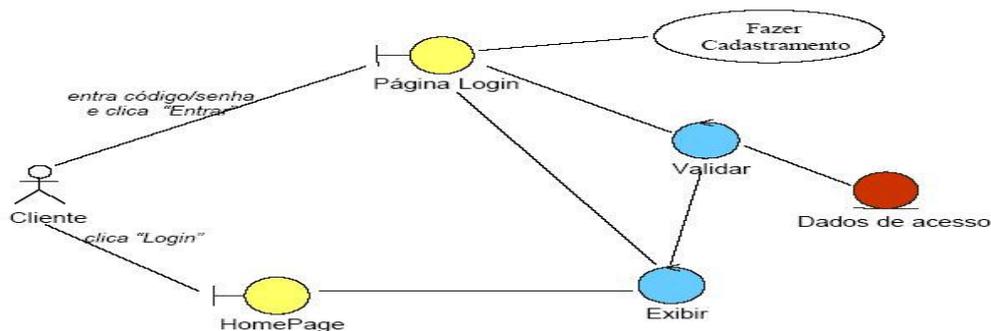


Figura 22 - Diagrama de robustez

Fonte: MARTINS, 2007.

3.3 METODOLOGIA ICONIX

A Metodologia ICONIX foi elaborada por Doug Rosenberg e Kendall Scott a partir da síntese do processo unificado pelos “três amigos” - Booch, Rumbaugh, e Jacobson o qual tem dado suporte e conhecimento a mesma desde 1993 (ROSENBERG & SCOTT, 1999 apud BONA, 2002).

O ICONIX é “um processo simplificado que unifica conjuntos de métodos de orientação a objetos em uma abordagem completa, com o objetivo de dar cobertura ao ciclo de vida”. (BONA, 2002, p.60).

É considerado uma metodologia pura, prática e simples, mas também poderosa e com um componente de análise e representação dos problemas sólido e eficaz, por isso, a metodologia ICONIX é caracterizada como um Processo de Desenvolvimento de Software desenvolvido pela ICONIX Software Engineering. (MAIA, 2007).

Para Rosenberg & Scott (2001 apud FLORÃO, 2004), o ICONIX é um processo que se situa entre o *RUP – Rational Unified Process* (qualificado pelos autores como “imenso”) e o *XP- eXtreme Programming* (qualificado por eles como “minúsculo”). É processo dirigido por casos de uso, como o *RUP*, mas não é tão burocrático, ou seja, não gera tanta documentação. O ICONIX também é um processo relativamente compacto, como o *XP*, mas não descarta a análise e projeto (design) como o *XP* faz.

O ICONIX utiliza a linguagem de modelagem *UML* e possui uma característica exclusiva chamada “Rastreabilidade dos Requisitos” (*Traceability of Requirements*), mais precisamente, ele permite “obrigatoriamente” por meio de seus mecanismos, verificar em todas as fases, se os requisitos estão sendo atendidos. (MAIA, 2007).

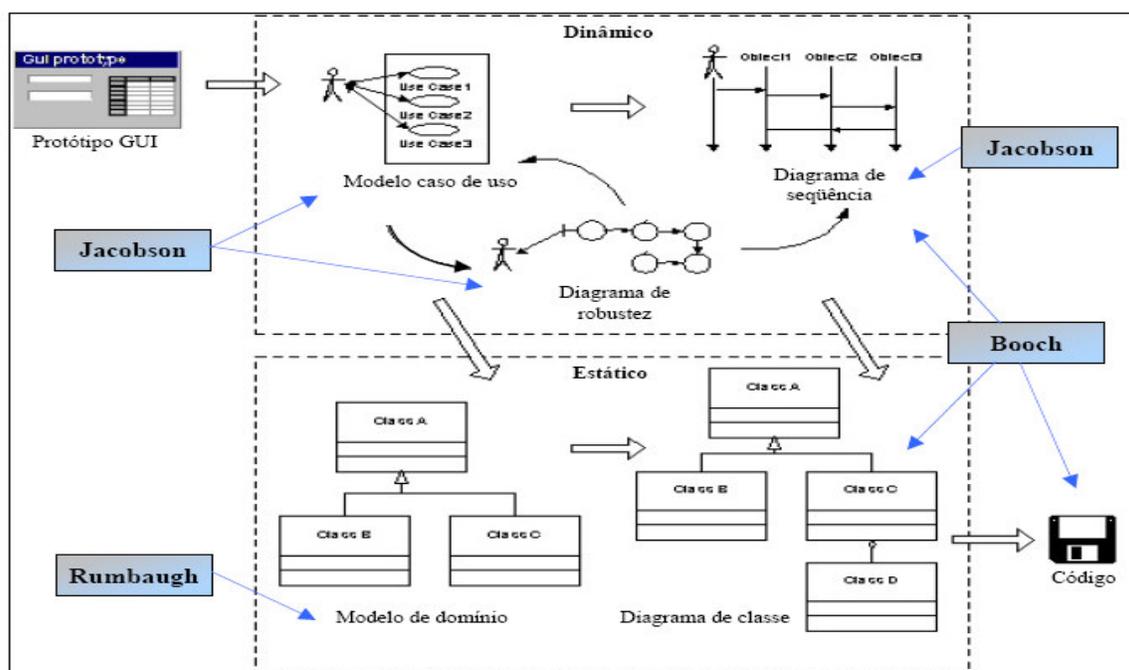


Figura 23 - Processo ICONIX, mostrando a contribuição dos “três amigos”.
Fonte: BONA, 2002.

De acordo com Rosenberg & Scott (1999 apud BONA, 2002), o ICONIX tem como base responder algumas questões fundamentais sobre o software. Assim, utiliza técnicas da *UML* que auxiliam a prover a melhor resposta. As questões e as técnicas são:

Questões	Técnicas
1 - Quem são os usuários do sistema (ou atores), e o que eles estão tentando fazer?	Utilizar casos de uso
2 - O que são, no "mundo real" (chamado domínio de problema), os objetos e as associações entre eles?	Utilizar diagrama de classe de alto nível
3 - Que objetos são necessários para cada caso de uso?	Utilizar análise de robustez
4 - Como objetos estão colaborando e interagindo dentro de cada caso de uso?	Utilizar diagrama de seqüência e de colaboração
5 - Como será manipulado em tempo-real aspectos de controle?	Utilizar diagrama de estado
6 - Como realmente será construído o sistema em um nível prático?	Utilizar diagrama de classe de baixo nível

Quadro 3 – Questões e Técnicas.

Fonte: Elaboração dos autores, 2007.

O processo ICONIX é dividido em dois grandes setores, o Modelo Estático e o Modelo Dinâmico, que podem ser divididos em paralelo e de modo recursivo (Figura 23).

O Modelo Estático é formado pelos Diagramas de Domínio e o Diagrama de Classe que pertencem à divisão estática e modelagem de funcionamento do sistema sem nenhuma interação do usuário. Ao contrário do Modelo Estático, o Modelo Dinâmico sempre mostra o usuário interagindo com o sistema por meio de ações onde o sistema apresenta alguma resposta ao usuário em tempo de execução. (MAIA, 2007).

O processo ICONIX trabalha a partir de um protótipo de interface onde se desenvolvem os diagramas de caso de uso baseado nos requisitos do usuário. Com base nos diagramas de caso de uso se faz à análise robusta para cada caso de uso e com os resultados obtidos, é possível desenvolver o diagrama de seqüência e posteriormente, povoar o modelo de domínio já revisado com métodos e atributos descobertos. (MAIA, 2007).

3.3.1 Principais características do ICONIX

Segundo Rosenberg & Scott (2001 apud FLORÃO, 2004), as características-chave do ICONIX são:

- **Uso enxuto da UML:** os passos do processo estabelecem apenas um conjunto mínimo e suficiente para que um projeto orientado a objetos seja bem sucedido.
- **Alto grau de rastreabilidade:** a cada passo do processo retorna-se de alguma forma aos requisitos. Nenhuma parte do processo permite que o analista se afaste muito das necessidades do usuário.
- **Iterativo e incremental:** múltiplas iterações acontecem entre o desenvolvimento do modelo de domínio e a identificação e análise dos casos de uso. O modelo estático é incrementalmente refinado durante as sucessivas iterações que acontecem no modelo dinâmico.

3.3.2 Fases e Marcos

Rosenberg & Scott (1999 apud FLORÃO, 2004) oferecem uma visão geral da abordagem ICONIX ao organizá-la em quatro grandes fases ou passos principais. Cada fase possui marcos (milestones) associados. Estes marcos são pontos definidos do processo no qual o

trabalho da equipe torna-se visível para os gerentes. As fases e os respectivos marcos são mostrados a seguir:

Fases	Marcos
Análise de requisitos	Marco 1: Revisão de requisitos
Análise e projeto preliminar	Marco 2: Revisão do projeto preliminar
Projeto	Marco 3: Revisão crítica do projeto
Implementação	Marco 4: Entrega

Quadro 4 – Fases e Marcos do Iconix.
Fonte: FLORÃO, 2004.

Essas quatro fases, junto com seus respectivos marcos, oferecem uma visão abrangente do processo, que será detalhada na seqüência.

3.3.2.1 Análise de Requisitos

As atividades realizadas durante a fase de análise de requisitos são as seguintes (BONA, 2002); (FLORÃO, 2004):

- **Modelo de Domínio** – Um diagrama de classes em alto nível (macro) para identificar os objetos do "mundo real" e os relacionamentos de agregação e generalização entre eles;
- **Prototipação** – Se possível apresentar uma prototipação rápida de interface do sistema ou diagramas de navegação para que assim, o cliente possa compreender melhor o sistema proposto;
- **Modelo de Caso de Uso** - Identificar os casos de uso (use cases) do sistema, mostrando os atores envolvidos;
- **Diagrama de Pacote** – Utilizar o diagrama de pacote para organizar os caso de uso em grupos.
- Associar os requisitos funcionais aos casos de uso e objetos de domínio.

Depois de concluídas estas atividades, chega-se ao **Primeiro Marco** do processo: **A Revisão dos Requisitos**.

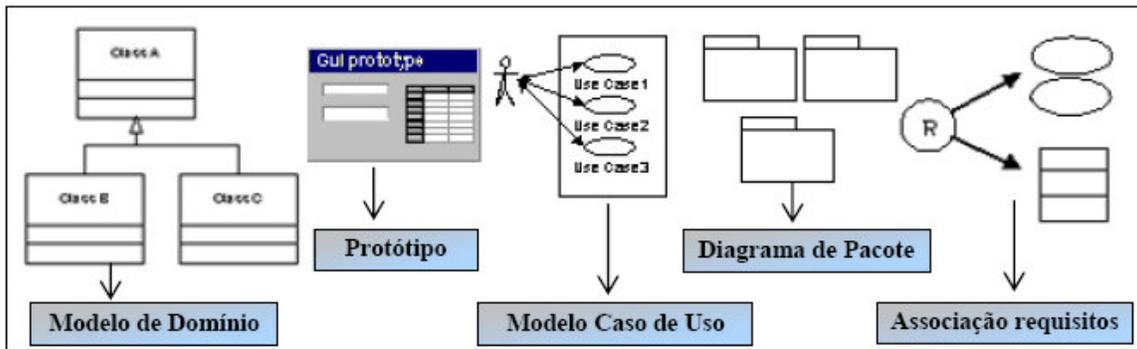


Figura 24 - ICONIX - Atividades da análise de requisitos
 Fonte: BONA, 2002.

Um importante aspecto do ICONIX é que um requisito se distingue explicitamente de um caso de uso (SILVA & VIDEIRA, 2001 apud BONA, 2002). Neste sentido, um caso de uso descreve um comportamento; um requisito descreve uma regra para o comportamento. Além disso, um caso de uso satisfaz um ou mais requisitos funcionais; um requisito funcional pode ser satisfeito por um ou mais casos de uso.

3.3.2.2 Análise e Projeto Preliminar

Durante a fase de Análise e Projeto Preliminar são realizadas as seguintes atividades (BONA, 2002); (FLORÃO, 2004):

- **Caso de Uso** - Escrever os casos de uso, com o fluxo principal das ações, podendo também conter o fluxo alternativo e o fluxo de exceções;
- **Análise de Robustez** - Apresentar a análise de requisitos. Para cada caso de uso:
 - Identificar um conjunto de objetos que executam o cenário descrito;
 - Atualizar o diagrama de classes do modelo de domínio;
- **Diagrama de classe** – Finalizar sua atualização.

Uma vez concluídas estas atividades, alcança-se o **Segundo Marco** do processo, no qual é realizada a **Revisão do Projeto Preliminar**, descrito na figura 25.

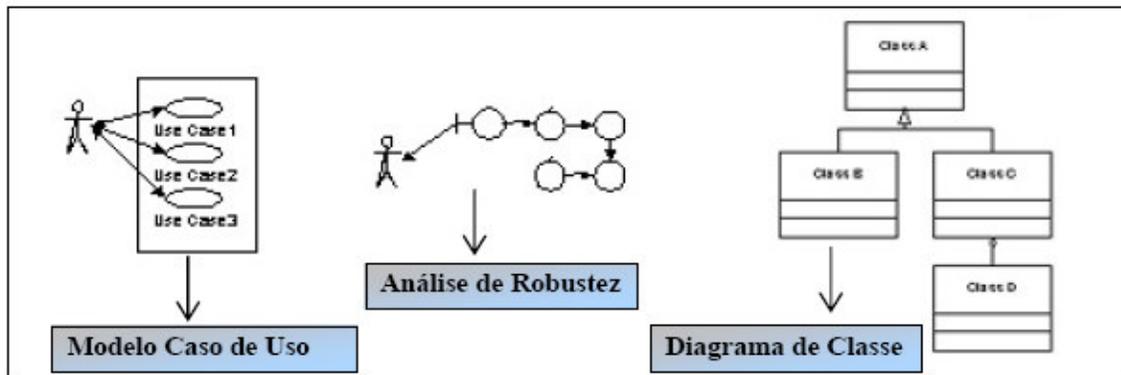


Figura 25 - ICONIX - Atividades da análise e projeto preliminar.
Fonte: BONA, 2002.

3.3.2.3 Projeto

Durante a fase de projeto são realizadas as seguintes atividades (Figura 26):

- **Diagrama de Seqüência** - Especificar o comportamento do projeto por meio do diagrama de seqüência. Onde para cada caso de uso identificar as mensagens entre os diferentes objetos. Caso necessário utilizar diagrama de colaboração para representar as transações entre objetos. Também pode-se utilizar o diagrama de estado para mostrar o comportamento em tempo real;
- **Diagrama de classe** - Terminar o modelo estático, adicionando detalhes ao projeto no diagrama de classe;
- Fazer uma verificação junto com a equipe para ver se o projeto satisfaz todos os requisitos identificados. (BONA, 2002).

Após serem concluídas as tarefas, chega-se ao **Terceiro Marco** do processo, no qual é realizada a **Revisão Detalhada/Crítica do Projeto**, descrita na Figura 26.

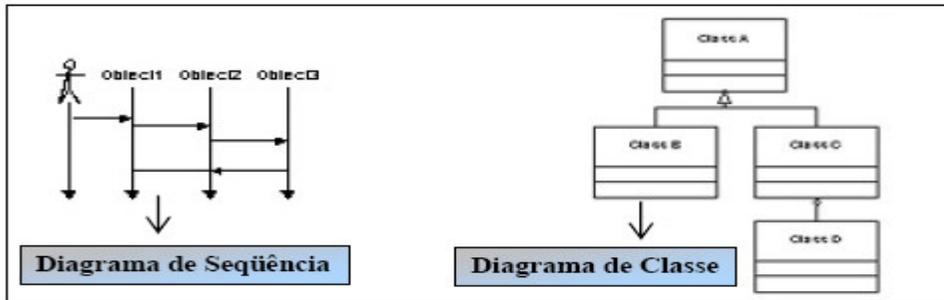


Figura 26 - ICONIX - Atividades do projeto.
Fonte: BONA, 2002.

3.3.2.4 Implementação

As atividades que dão suporte a fase de implementação, são (MAIA, 2002):

- **Diagrama de Componente** - Se for preciso utilizar diagramas de componente para apoiar a fase de desenvolvimento;
- **Código** - Escrever/Gerar o código;
- **Teste de Unidade e Integração** - Realizar testes de unidade e de integração;
- **Teste de Aceitação do Usuário** - Realizar testes de aceitação do usuário.

Ao serem concluídas estas atividades, alcança-se o **Quarto** e último **Marco** do processo, no qual é realizada a **entrega** do produto.

3.4 MODELAGEM DO SISTEMA

3.4.1 Casos de uso

O diagrama abaixo expõe o funcionamento do sistema. São mostrados dois casos de uso, o *ETL* e o *Front-End*. Antes de aplicar a ferramenta *ETL*, há necessidade de definir as

informações que são de interesse para o usuário, e o administrador deve adicioná-las ao banco de dados. Após esse passo, aplica-se a ferramenta de *ETL* para buscar os dados na base de dados, filtrá-los e tratá-los, para que sejam carregados no *Data Warehouse*. A seguir, aplica-se a ferramenta de *Front-End*, para gerar os relatórios que auxiliaram na tomada de decisão.

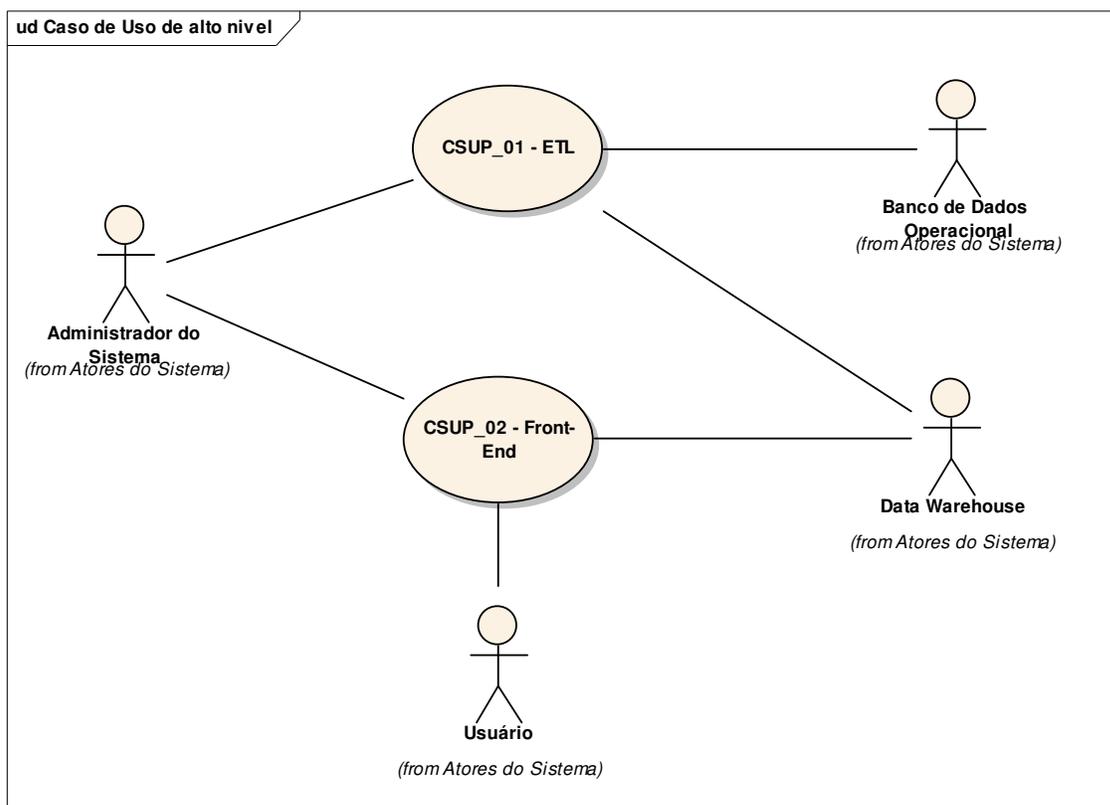


Figura 27: Diagrama de Casos de uso
Fonte: Elaboração dos Autores, 2007.

No diagrama de casos de uso, os atores podem estar representados por:

- a) Banco de dados Operacional: dados dos sistemas *OLTP* (transacionais), os quais contém os dados que serão utilizados para a alimentação (carga) do *Data Warehouse*.
- b) Administrador: pessoa responsável por carregar no *Data Warehouse*, usando o *ETL*, os dados disponibilizados no banco de dados operacional.
- c) Usuário: ator que visualizará as informações geradas por meio da ferramenta de *Front-End*.

d) *Data Warehouse*: Repositório de dados, formado por um modelo de dados dimensional.

3.4.1.1 ETL

O *ETL* é uma ferramenta utilizada para extrair dados de um banco de dados, transformá-los e inseri-los em outro banco de dados especial, o *Data Warehouse*.

3.4.1.1.1 Diagrama de classes

Este diagrama mostra como o *ETL* é formado.

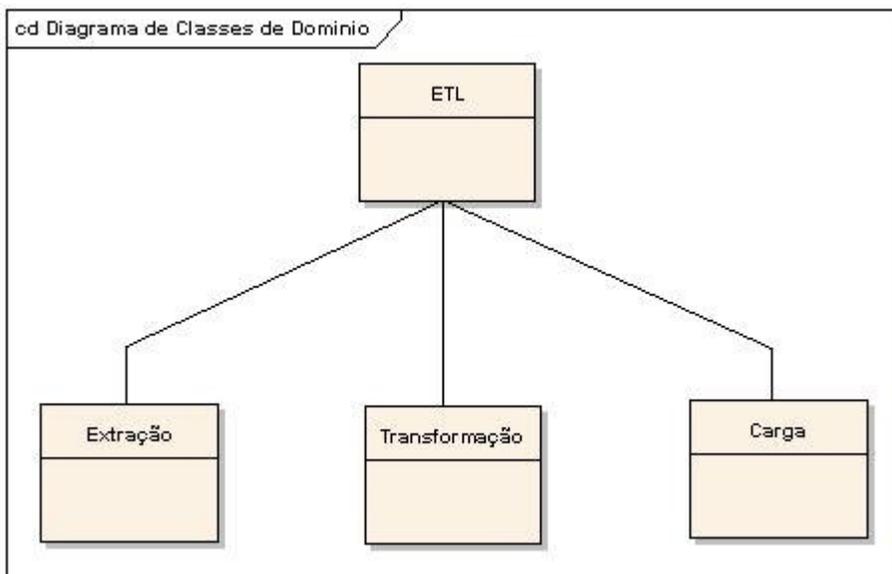


Figura 28: Diagrama de classe do *ETL*.
Fonte: Elaboração dos Autores, 2007.

3.4.1.1.2 Diagrama de caso de uso

O diagrama abaixo auxilia na visualização das principais funcionalidades da parte de *ETL*.

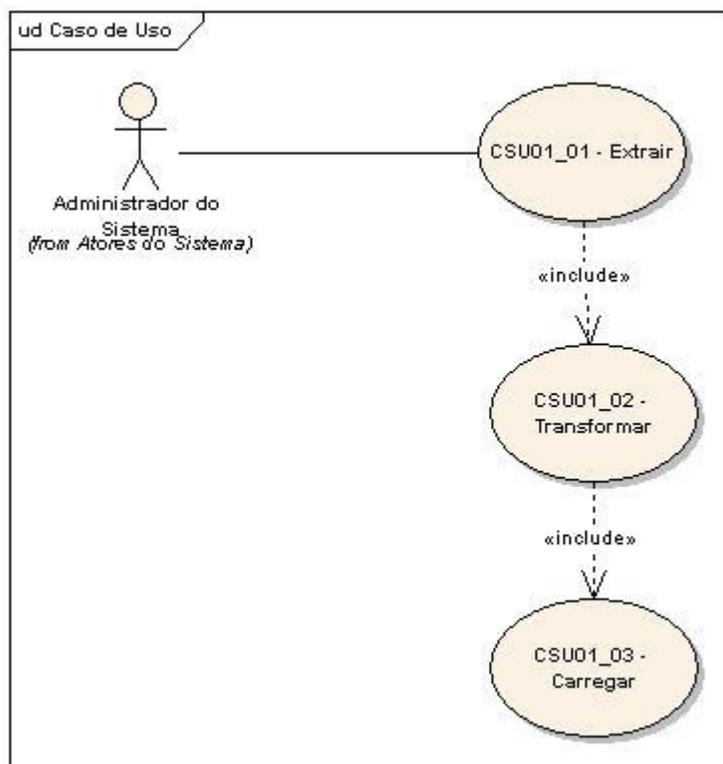


Figura 29: Diagrama de caso de uso do *ETL*.
Fonte: Elaboração dos autores, 2007.

A finalidade do *ETL* é extrair os dados dos sistemas operacionais ou base de dados, transformar (limpar e tratar) esses dados, e carregá-los no *Data Warehouse*, onde os dados estarão a disponíveis para consulta.

3.4.1.1.3 Diagrama de Robustez

Diagrama que explica como funcionará o *ETL*.

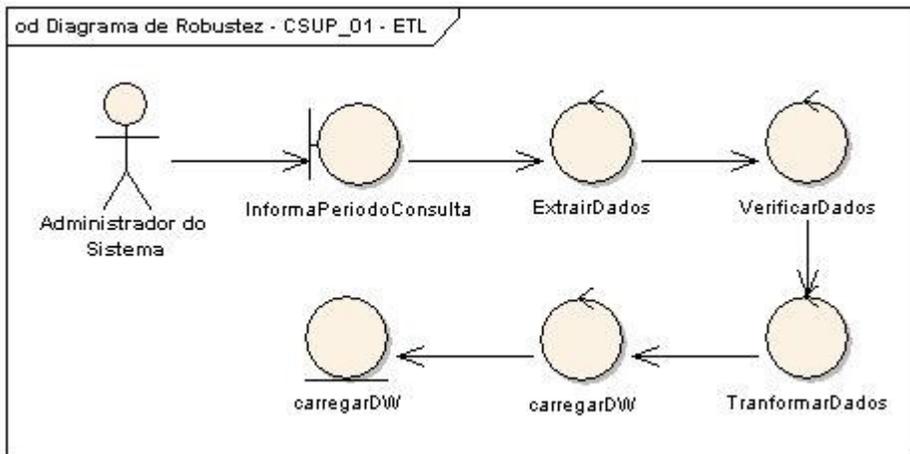


Figura 30: Diagrama de robustez do *ETL*.
 Fonte: Elaboração dos Autores, 2007.

Percebe-se que o administrador do sistema informa ao software, por meio de uma janela, o período da consulta que deseja, o *ETL* pega os dados necessários pra tal, verifica, filtra e limpa esses dados, transformando-os em dados prontos para o *Data Warehouse*. Esses dados são carregados no *Data Warehouse*, e uma nova janela é mostrada ao administrador, confirmando a operação.

3.4.1.1.4 Diagrama de Seqüência

O diagrama de seqüência abaixo demonstra o “ciclo de vida” dos processos de *ETL*, e como eles interagem.

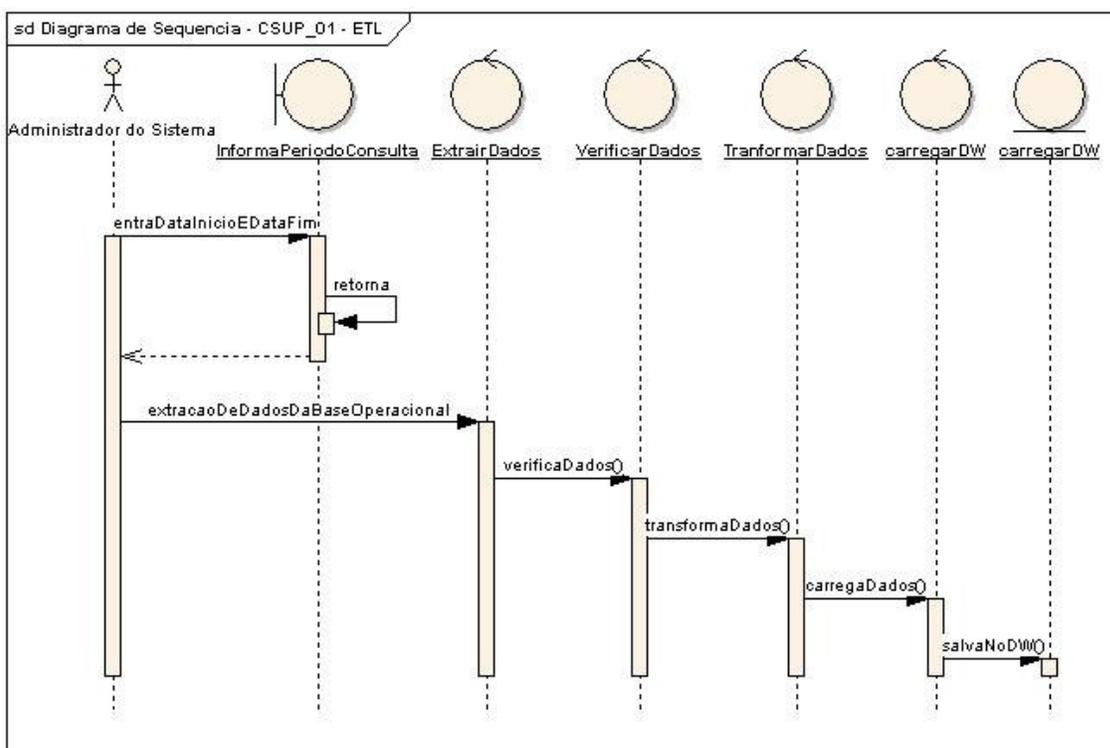


Figura 31: Diagrama de seqüência do *ETL*.
 Fonte: Elaboração dos Autores, 2007.

3.4.1.2 Font-End

O *Front-End* é a parte do sistema que interage diretamente com o usuário, permitindo que ele visualize os dados disponíveis no *Data Warehouse*.

3.4.1.2.1 Diagrama de caso de uso

O diagrama de caso de uso auxilia na visualização das principais funcionalidades da parte de *Front-End*, conforme demonstrado na figura abaixo:

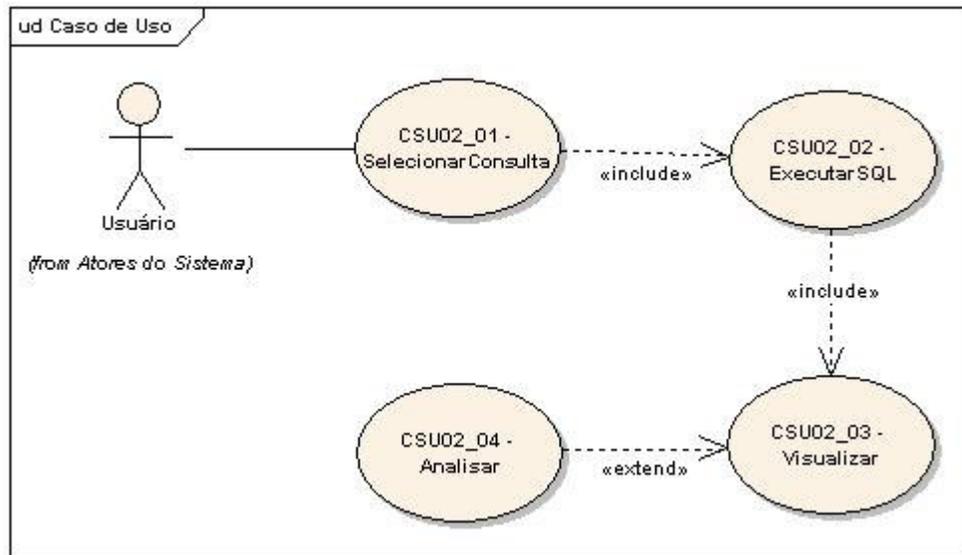


Figura 32: Diagrama de caso de uso do *Front-End*.

Fonte: Elaboração dos autores, 2007.

Após a análise desses dados, o usuário possui informações para a criação dos cenários, e somado a isso, o surgimento da prospecção.

3.4.1.2.2 Diagrama de Robustez

O diagrama de robustez demonstrado abaixo descreve como funcionará o *Front-End*.

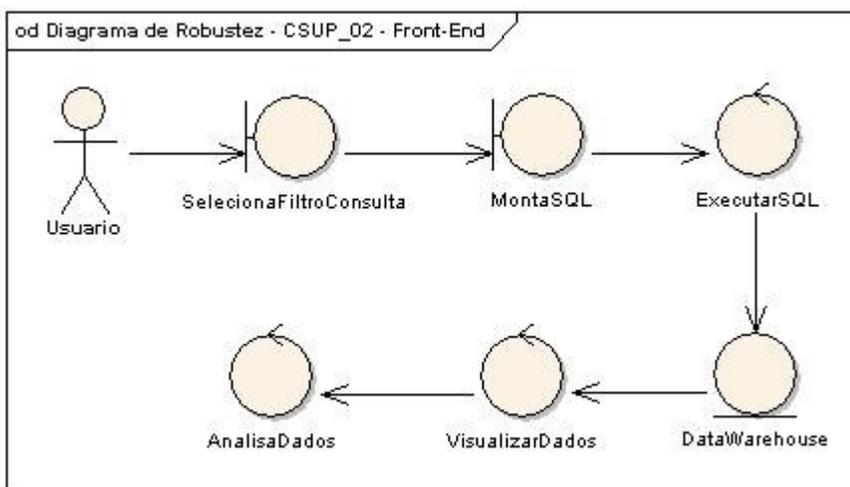


Figura 33: Diagrama de robustez do *Front-End*.

Fonte: Elaboração dos autores, 2007.

Nota-se por meio do diagrama, que em uma interface gráfica, o usuário seleciona o filtro de consulta e monta a SQL. Em seguida manda executar essa SQL, que vai ao *Data Warehouse* e busca as informações relativas ao filtro. Essas informações são visualizadas e analisadas, permitindo a criação dos relatórios que auxiliaram na tomada de decisão.

3.4.1.2.3 Diagrama de Seqüência

Este diagrama demonstra o “ciclo de vida” dos processos de *Front-End*, e como eles interagem.

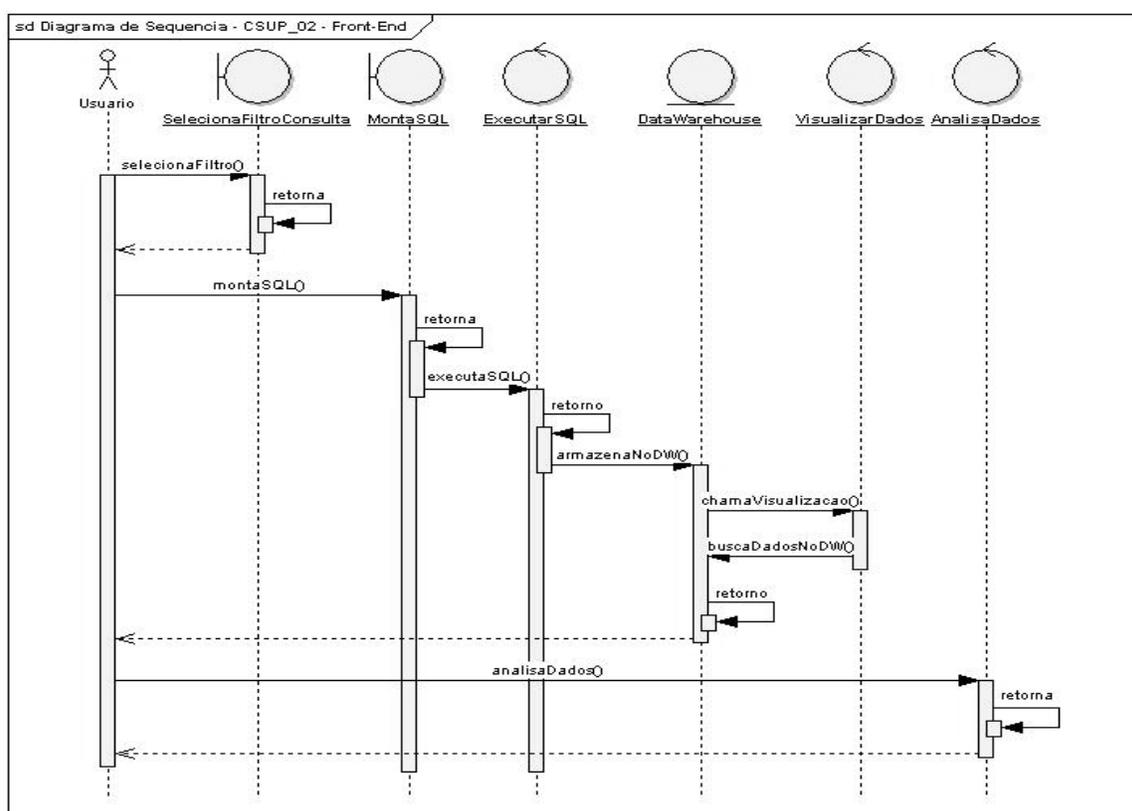


Figura 34: Diagrama de seqüência do *Front-End*.

Fonte: Elaboração dos autores, 2007.

3.4.2 Implementação

Na implementação o projeto é testado. Esta etapa é realizada no próximo capítulo.

3.5 CONCLUSÃO

Alguns problemas podem ser resolvidos no mesmo instante que aparecem, porém outros necessitam de alguma estratégia ou reflexão prévia. Da mesma forma, softwares e arquiteturas precisam deste planejamento para serem resolvidos com maior rapidez, clareza, certeza e coerência. Sendo assim, tentou-se aqui, abranger um pouco da história da *UML* e da metodologia de modelagem *ICONIX*, suas características, definições e diagramas, para que com isso, se consiga um melhor raciocínio e compreensão da solução do problema.

4 DESENVOLVIMENTO

Neste trabalho, o foco principal foi dado no campo da pesquisa, assim, o desenvolvimento acontece na busca por dados encontrados no *DW* e na construção das análises, que geram informações pertinentes ao tema do trabalho.

Os dados foram captados nos sites do IBGE e do INEP.

Foram encontradas algumas dificuldades no momento da averiguação dos dados, pois algumas informações estavam somadas a outras, e não conferiam com as informações da outra fonte de dados. Também na parte do referencial teórico, houve a necessidade de buscar material em outras universidades, pois além de se tratar de um assunto novo, a biblioteca não oferece uma diversidade de autores e títulos necessários para um bom embasamento bibliográfico.

Depois de todas as tabelas de dados conferidas, elas foram organizadas e padronizadas. Foi elaborado o modelo de dados (tabelas fato e dimensão), conforme mostra a figura abaixo.

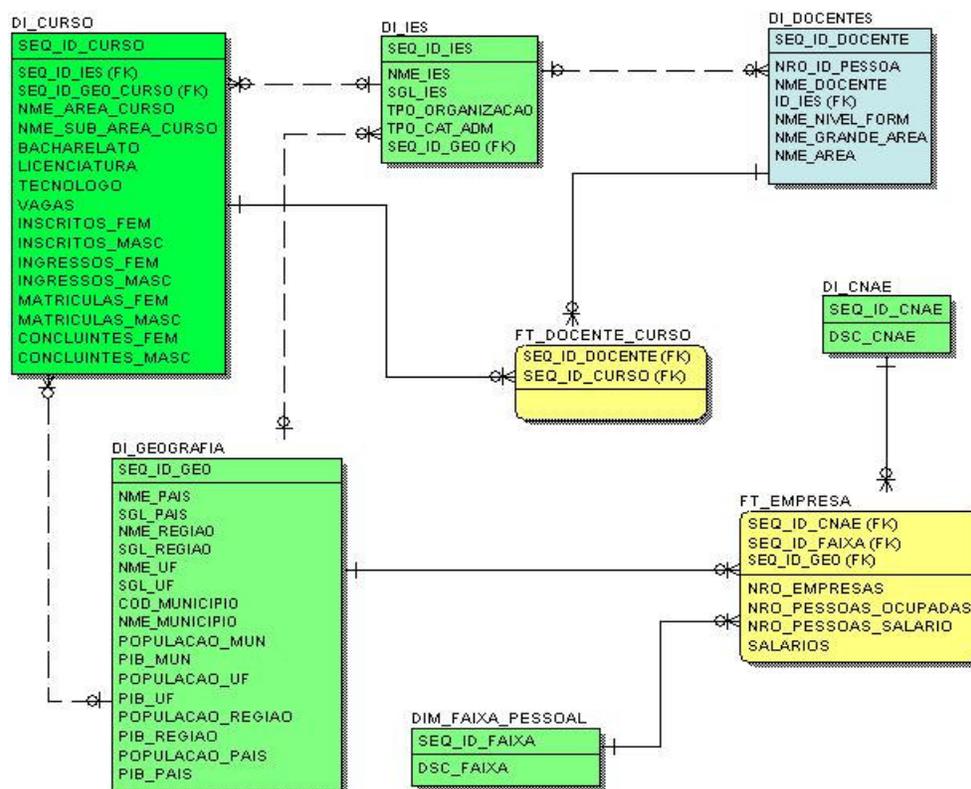


Figura 35: Modelo de dados

Fonte: Elaboração dos Autores, 2007.

Feito isso, os dados foram carregados no *Data Warehouse* e assim, iniciado o processo de cruzamento de informações, com o objetivo de deixar estas tabelas menores e ao mesmo tempo mais objetivas, limpas e completas.

Depois de concluído o DW, foi utilizado como ferramenta de *Front-end* o software Microsoft Excel. Por meio dele, foram montadas tabelas dinâmicas. A partir destas tabelas e utilizando fórmulas e gráficos, foram elaboradas respostas e comparações pertinentes, bem como caminhos mais adequados para serem traçados no futuro.

4.1 PROPOSTA DE SOLUÇÃO

A proposta da solução para a análise exploratória dos dados está representada na figura abaixo.

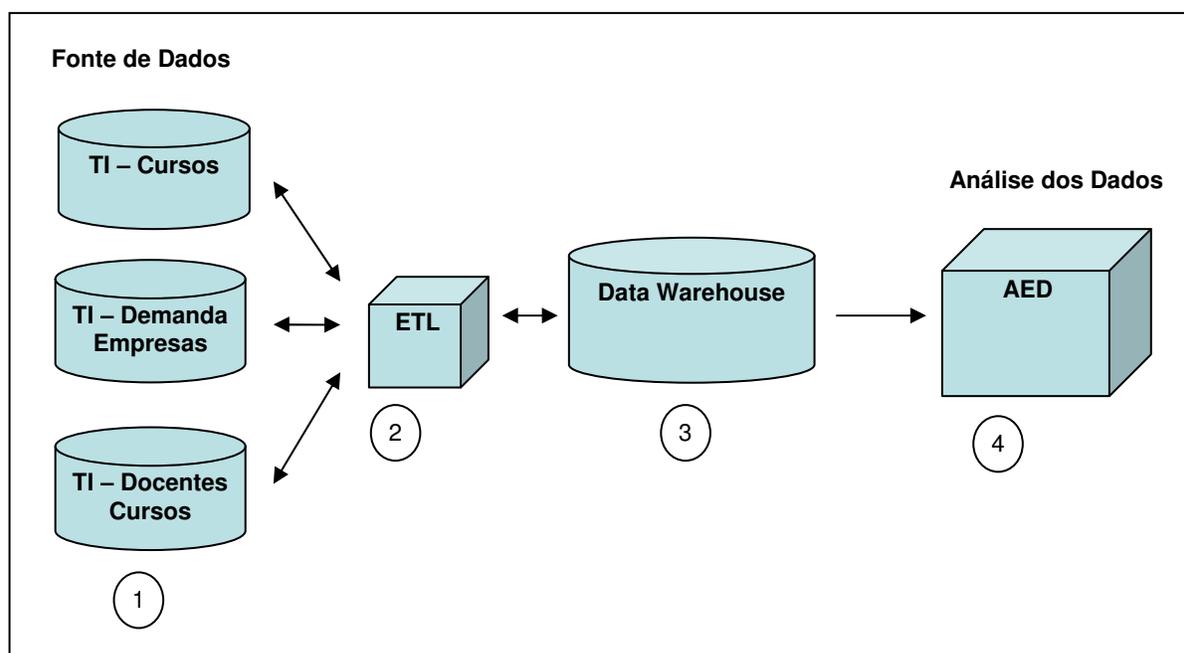


Figura 36 – Proposta de Solução.

Fonte: Elaboração dos Autores, 2007.

4.1.1 Fonte de Dados – Os dados foram retirados de tabelas e arquivos disponibilizados pelos órgãos IBGE e INEP. Estes arquivos trazem informações sobre os cursos da área de Ciência da Computação, como os docentes e discente, fala também das Instituições de

Ensino que oferecem cursos nesta área, e sobre as empresas ligadas a Tecnologia da Informação. Somadas estas fontes de dados formam a base de dados, a qual foram extraídos os dados para a formação do *Data Warehouse (DW)*.

- 4.1.2 ETL** – Extração dos dados da base de dados, realizada com ajuda do Excel e do Word. Em seguida, foi feita a limpeza dos dados, tirando o que não tinha valor. A transformação destes dados para um único formato foi obtida por meio do Access, pois além de ser utilizado para montagem do *DW* propriamente dito, é fácil de fazer as consultas. Feito a carga dos dados no *DW*, as análises e cruzamentos de informações puderam ser iniciados.
- 4.1.2 Data Warehouse** – Repositório de dados baseado em assuntos, é integrado e não volátil e pode sofrer variação no decorrer do tempo. Utilizou-se este repositório para a análise de dados e como sistemas de apoio à decisão. Por meio da linguagem SQL, faz-se o cruzamento de tabelas e dados, com o propósito de visualizar os dados confrontados que formam novas tabelas que serão analisadas exploratoriamente.
- 4.1.3 Análise dos Dados** – Análise exploratória dos dados (AED) armazenados no *DW* para esclarecimento de dúvidas e dogmas entorno das IES, discentes, docentes e cursos. Utilizando o Excel, montou-se os histogramas e gráficos para visualizar as informações.

4.2 TECNOLOGIAS UTILIZADAS

A maioria das ferramentas utilizadas faz parte do Pacote *Microsoft Office*[®]. Como visto na Figura 38, foi utilizado o software da *Sparxsystems*[®], o *Enterprise Architect*[®], para a construção de diagramas. Os sites do IBGE e do INEP serviram como base de busca para adquirir informações e tabelas necessárias para o desenvolvimento do *Data Warehouse*. Os dados recolhidos na Internet foram transformados, ou seja, limpados e padronizados por meio do Excel e depois carregados no banco de dados Access, formando assim o *Data Warehouse*. Com o *DW* pronto no Access, foi utilizada a linguagem SQL (*Structured Query Language*) para realizar as relações e consultas, e com base nestas consultas, se tornar possível as análises de dados. A SQL é a linguagem mais usada para acessar banco de dados.

Ferramenta	Uso
 IBGE – Instituto Brasileiro de Geografia e Estatística	Fonte dos Dados.
 INEP – Instituto Nacional de Estudos e Pesquisas	Fonte de Dados.
 Enterprise Architect®	Criação dos Diagramas
 EXCEL®	Transformação dos dados.
 ACCESS®	Criação do DW, e fonte das consultas.
 Structured Query Language	Linguagem de Consulta Estruturada, usada para extrair as informações e fazer o relacionamento das tabelas.
 WORD®	Editor de texto utilizado na monografia.
 Windows ^{xp} ®	Plataforma utilizada em todas as partes do trabalho.

Quadro 5 - Tecnologias Utilizadas.

Fonte: Elaboração dos Autores, 2007.

4.3 ANÁLISES

Com o *DW* devidamente carregado e pronto, extraiu-se as informações pertinentes, e para a melhor visualização dos resultados, usando o método de histogramas e gráficos, foram feitas as análises.

Para facilitar, os integrantes deste trabalho dividiram as análises em quatro partes, sendo elas: Cursos, Docentes, Discentes e Empresas.

4.3.1 Análises de Cursos

As variáveis utilizadas para as análises de Cursos foram:

a) **Região Geográfica:** São as 05 (cinco) regiões geográficas do Brasil. Sendo elas:

- Norte (NO)
- Nordeste (NE)
- Centro-Oeste (CO)
- Sudeste (SE)
- Sul (SU)

b) **Organização Acadêmica:** Organização acadêmica da IES. Neste trabalho estas categorias são divididas em:

- Centro
- Faculdade
- Universidade

Onde:

Centro representa todos os: Centros Federais de Educação Tecnológica, Centros Universitários e Centros Universitários Especializados.

Faculdade representa todas as: Faculdades Integradas, Faculdades, Faculdades de Tecnologia, Institutos Superiores e Escolas Superiores.

Universidade representa: Universidades.

c) **Categoria Administrativa:** Categoria administrativa da IES. Sendo estas categorias divididas em:

- **Federal** (Mantidas e administradas pelo Governo Federal).
- **Estadual** (Mantidas e administradas pelos governos dos estados).
- **Municipal** (Mantidas e administradas pelo poder público municipal).
- **Privada** (São as mantidas e administradas por pessoas físicas ou jurídicas de direito privado).

d) **Ano de Início:** Representa o ano de início do curso na IES. Sendo dividido em:

- 1969 - 1978
- 1979 - 1988
- 1989 - 1998
- 1999 – 2008

e) **Área do curso:** Representa a área geral do curso. Sendo dividida em:

- Ciência da Computação
- Processamento da informação
- Uso do computador

f) **Área Específica:** Representa a área específica do curso. Sendo dividida em:

- Administração de redes
- Análises de sistemas
- Banco de Dados
- Ciência da Computação
- Comércio Eletrônico
- Computação Gráfica
- Engenharia de Computação (hardware)
- Engenharia de softwares
- Informática (ciência da computação)
- Informática Educacional
- Linguagens de Programação (visual basic, c++ etc)
- Processamento de dados
- Segurança da Informação
- Sistemas de Informação
- Sistemas Operacionais
- Tecnologia da informação
- Tecnologia em desenvolvimento de softwares

- Tecnologia em informática
- Uso da Internet

g) **Curso:** Representa o total de cursos.

4.3.1.1 As análises

A seguir serão apresentados alguns dos gráficos gerados nas análises dos cursos com um breve comentário a respeito de cada gráfico.

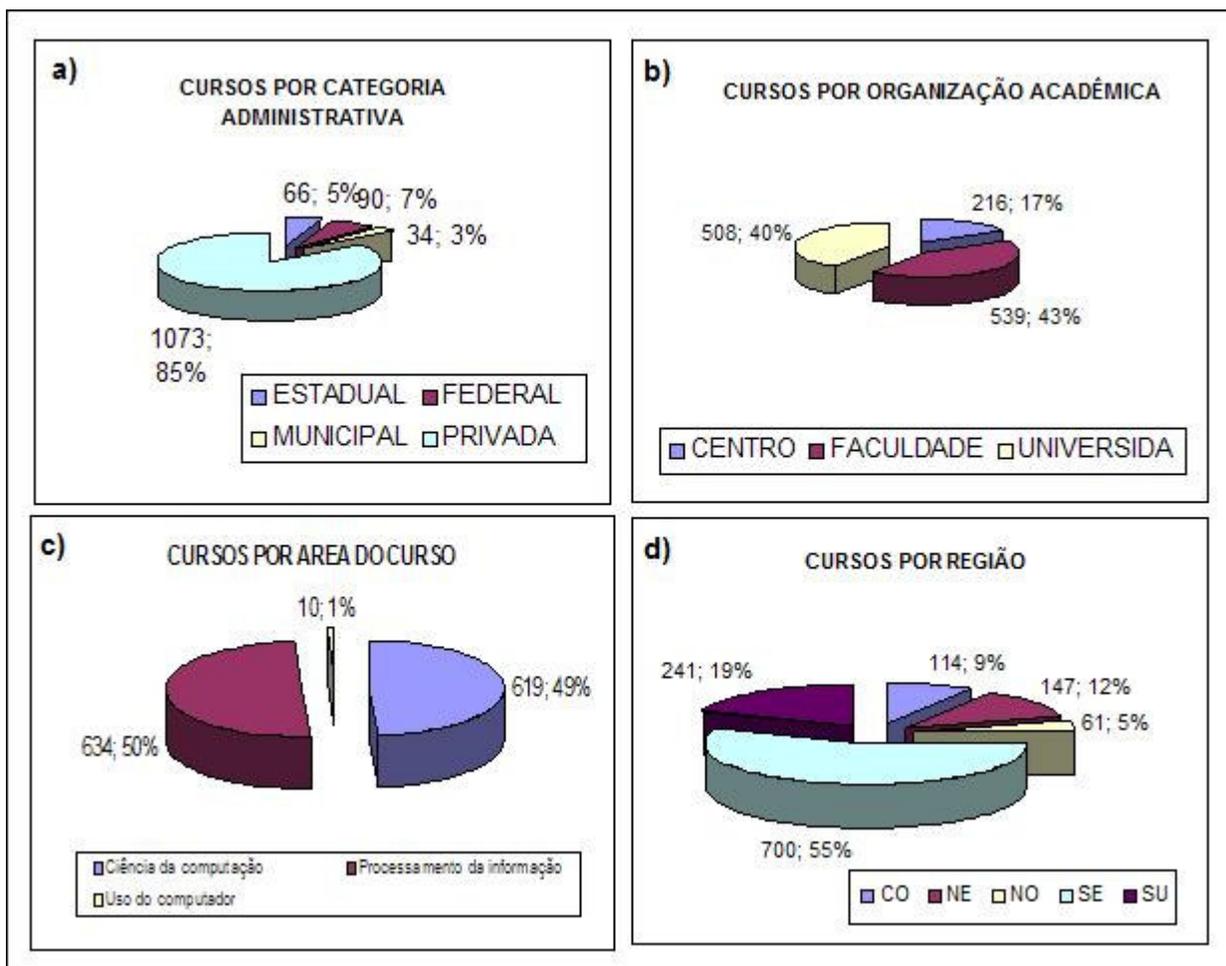


Gráfico 1 – Análises cursos.

Fonte: Elaboração dos Autores, 2007.

A maioria dos 1.263 cursos é de instituições Privadas, estão concentrados em Faculdades e Universidades, e localizados na região Sudeste. A região Sul aparece em segundo lugar com 19% do total. Os cursos estão divididos basicamente entre as áreas de Ciência da Computação e Processamento da informação (Gráfico 1).

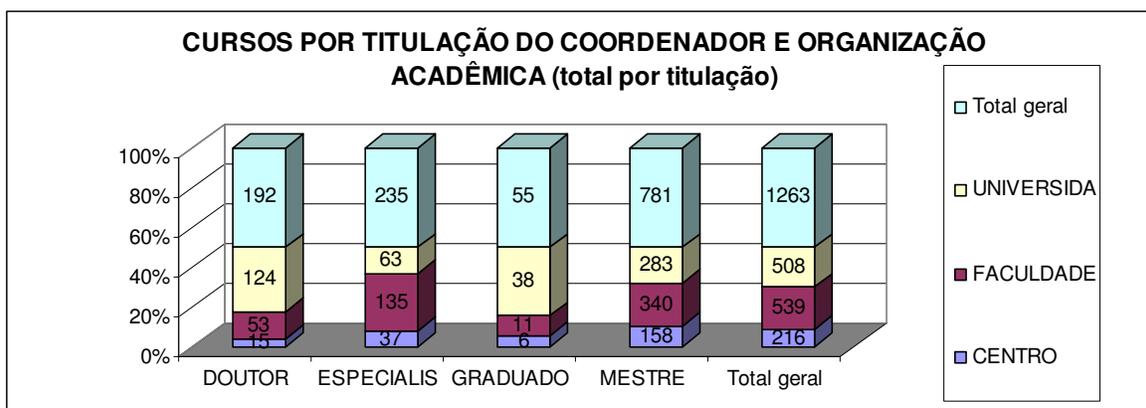


Gráfico 2 – Cursos por titulação do coordenador e organização acadêmica (total por titulação).

Fonte: Elaboração dos Autores, 2007.

Baseando-se nos gráficos 2 e 3, nota-se que a maioria dos coordenadores doutores e graduados estão nas universidades, e dos especializados e mestres estão nas faculdades. Porém, em todas as categorias administrativas a maioria dos coordenadores possuem a titulação de mestre, principalmente entre as IES privadas. Os coordenadores com mestrado representam 62% do total, enquanto os que têm titulação máxima de doutor 15%, inferior aos 19% de especialistas.

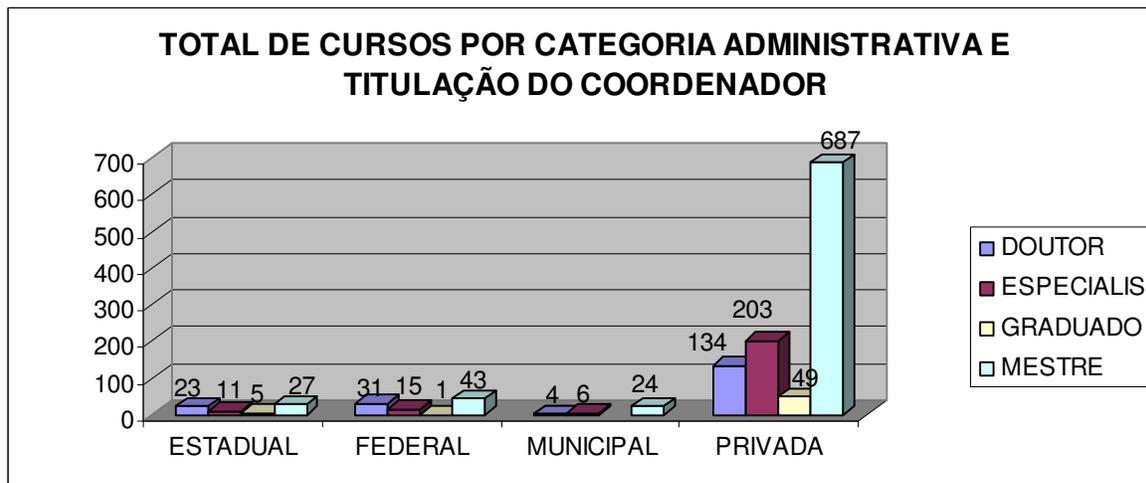


Gráfico 3 – Cursos por Categoria Administrativa e Titulação do Coordenador

Fonte: Elaboração dos Autores, 2007.

Por meio do gráfico abaixo (gráfico 4) nota-se que nos últimos anos o número de cursos vêm crescendo muito, principalmente nas IES Privadas, e que elas são as que predominam no país, ao contrário das IES Municipais. Os cursos iniciados depois de 1998, representam 74% do total de cursos existentes.

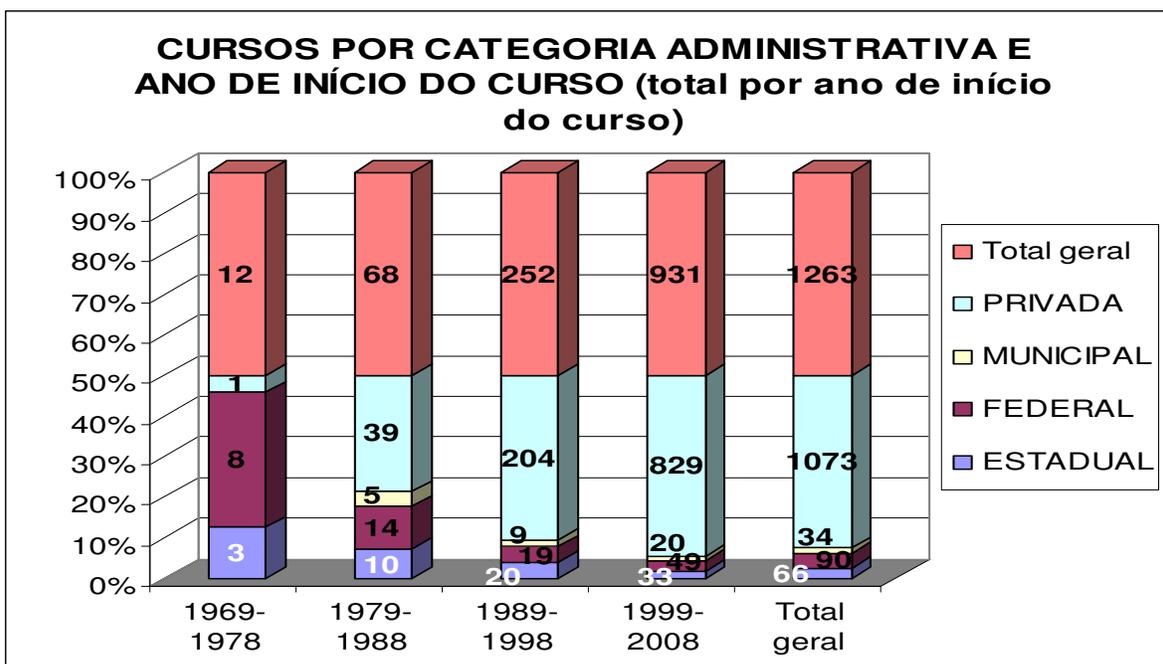


Gráfico 4 – Cursos por Categoria Administrativa e Ano de início do Curso (total por ano de início do curso).
Fonte: Elaboração dos Autores, 2007.

Nota-se que antigamente eram as Universidades que predominavam e atualmente são as Faculdades. (Gráfico 5)

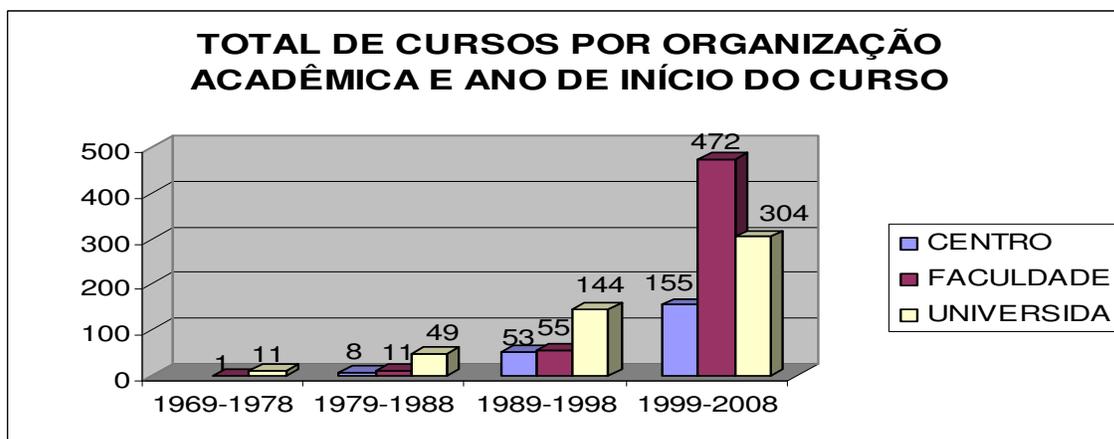


Gráfico 5 – Cursos por Organização Acadêmica e Ano de início do Curso.
Fonte: Elaboração dos Autores, 2007.

O gráfico 6 mostra que duas regiões não possuíam IES com cursos na área de ciência da computação antes de 1979, são elas: as regiões Norte e Centro-Oeste.

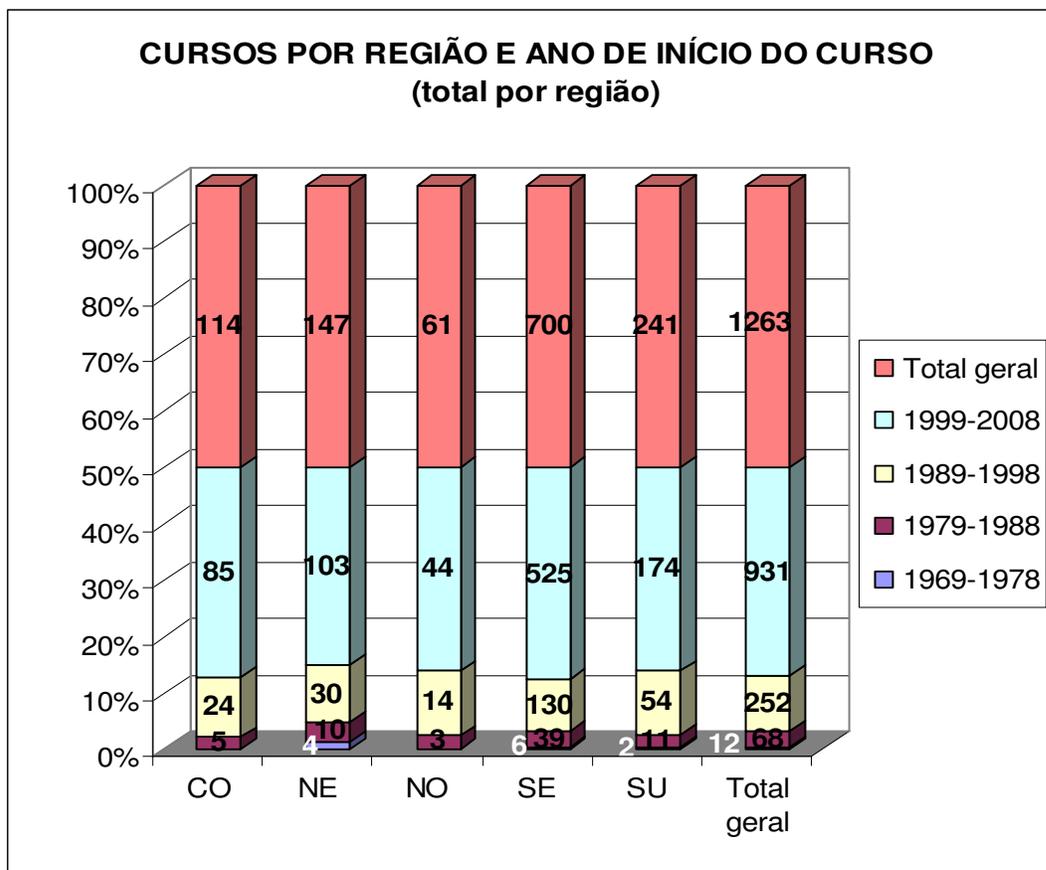


Gráfico 6 – Cursos por Região e Ano de início do Curso.
Fonte: Elaboração dos Autores, 2007.

4.3.2 Análises de Docentes

As variáveis utilizadas para as análises de Docentes foram:

a) **Região Geográfica:** São as 05 (cinco) regiões geográficas do Brasil. Sendo elas:

- Norte (NO)
- Nordeste (NE)
- Centro-Oeste (CO)
- Sudeste (SE)
- Sul (SU)

b) Organização Acadêmica: Organização acadêmica da IES. Neste trabalho estas categorias são divididas em:

- Centro
- Faculdade
- Universidade

Onde:

Centro representa todos os: Centros Federais de Educação Tecnológica, Centros Universitários e Centros Universitários Especializados.

Faculdade representa todas as: Faculdades Integradas, Faculdades, Faculdades de Tecnologia, Instituto Superior e Escola Superior.

Universidade representa: Universidades.

c) Categoria Administrativa: Categoria administrativa da IES. Sendo estas categorias divididas em:

- **Federal** (Mantidas e administradas pelo Governo Federal).
- **Estadual** (Mantidas e administradas pelos governos dos estados).
- **Municipal** (Mantidas e administradas pelo poder público municipal).
- **Privada** (São as mantidas e administradas por pessoas físicas ou jurídicas de direito privado).

d) Graduação: Representa os docentes com graduação.

e) Especialização: Representa os docentes com especialização.

f) Mestrado: Representa os docentes com mestrado.

g) Doutorado: Representa os docentes com doutorado.

h) Tempo médio ES: Representa o tempo médio do docente no Ensino Superior.

i) Tempo médio IES: Representa o tempo médio que o docente está na IES.

j) Curso: Representa o total de cursos.

4.3.2.1 As análises

As análises de docentes foram divididas em três partes, sendo elas pelas variáveis: região, organização acadêmica e categoria administrativa. A subseção 4.3.2.1.4 apresenta os gráficos de tempo médio dos docentes nas IES e no Ensino Superior. Alguns dos gráficos gerados serão apresentados a seguir.

4.3.2.1.1 Por região

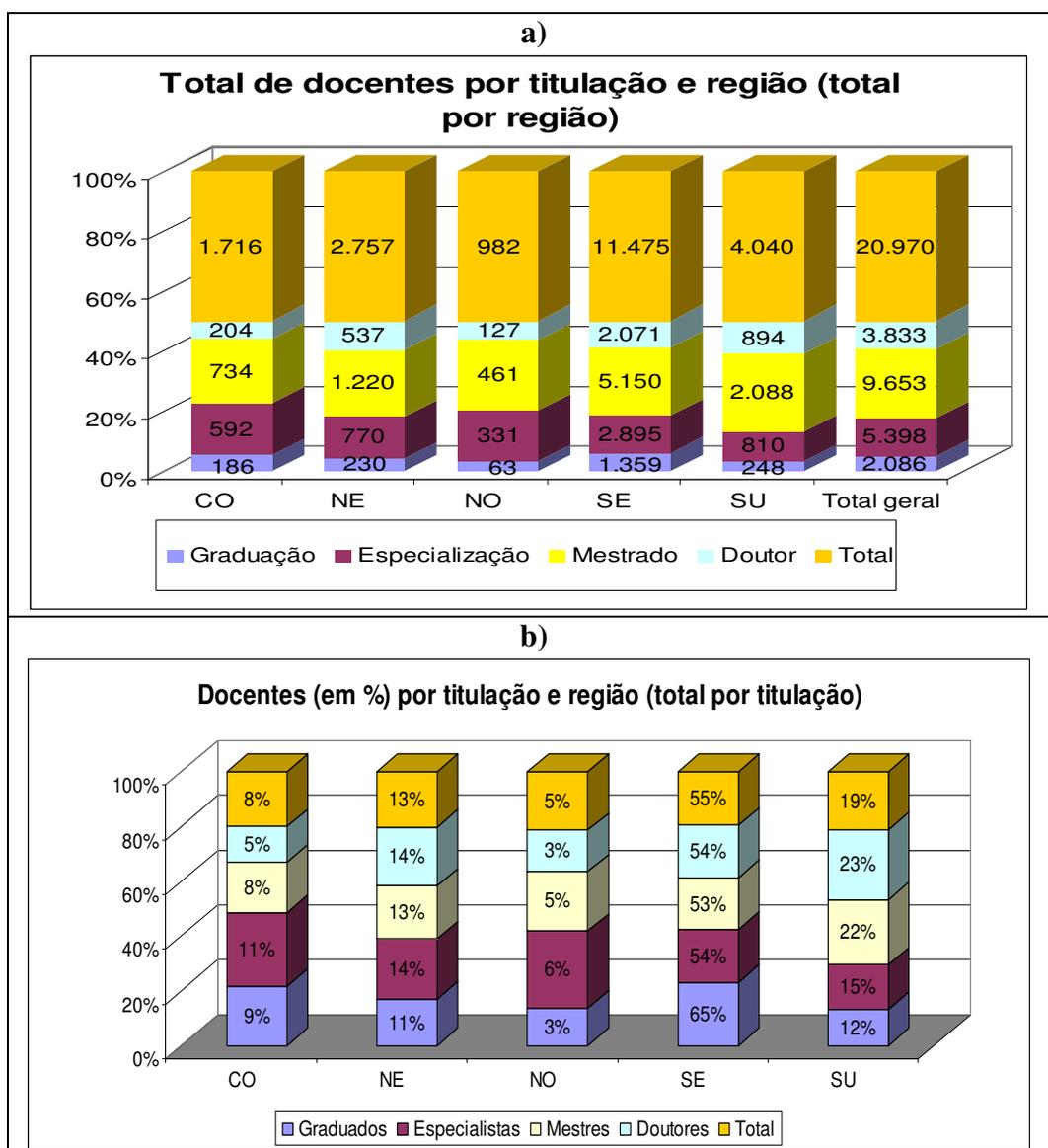


Gráfico 7 – Total de Docentes por Titulação e Região (Total por Região) (a), e Docentes (em %) por Titulação e Região (total por Região) (b).

Fonte: Elaboração dos Autores, 2007.

Nota-se que a região Sudeste é a que possui mais docentes, e retém metade dos docentes com mestrado do país, porém a região Sul apresenta quase $\frac{1}{4}$ (um - quarto) dos docentes com mestrado. Também pode-se ver que apenas 5% do total de docentes do país está na região

Norte e que mais da metade (55%) está na região Sudeste, região esta que detém 54% dos professores doutores do país. A região Sul apresenta também grande número de professores com doutorado se comparado com as outras regiões, representando 23% do total. A região Norte aponta a maioria dos professores com titulação de especialista, porém estes especialistas somente representam 6% do total de docentes especialistas do país.

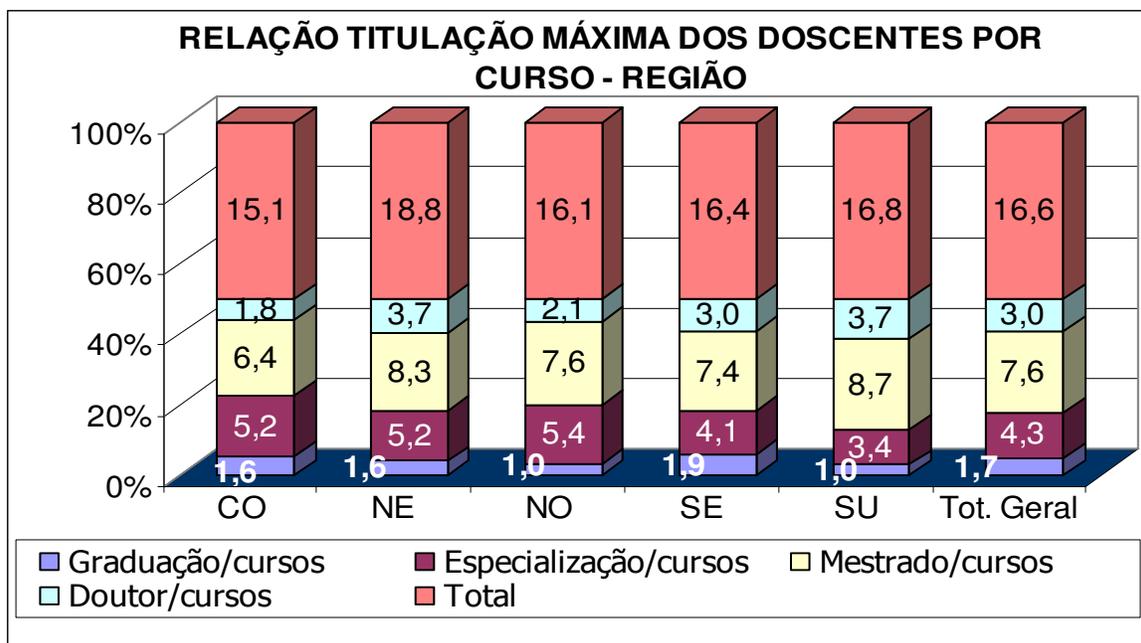


Gráfico 8 – Titulação Máxima dos Docentes por Curso - Região.
Fonte: Elaboração dos Autores, 2007.

Por meio do gráfico acima (gráfico 8) pode-se perceber que na maioria das regiões a média é de 16 (dezesesseis) professores por curso, e que a região que possui o maior número de professores é a Nordeste, com quase 19 (dezenove) por curso.

A região sul é a que menos possui docentes com especialização, ao contrário da região Norte, mas é a região que possui mais docentes mestres, superando até a região Sudeste. As regiões Sul e Nordeste são as que possuem a maior média de doutores, com quase 04 (quatro) por curso, mas somente na região SU a média de doutores é superior a média dos especialistas. As regiões Norte e Sul são as que possuem menos docentes com titulação máxima de graduação, tendo como média apenas 1 (um) docente por curso.

4.3.2.1.2 Por organização acadêmica

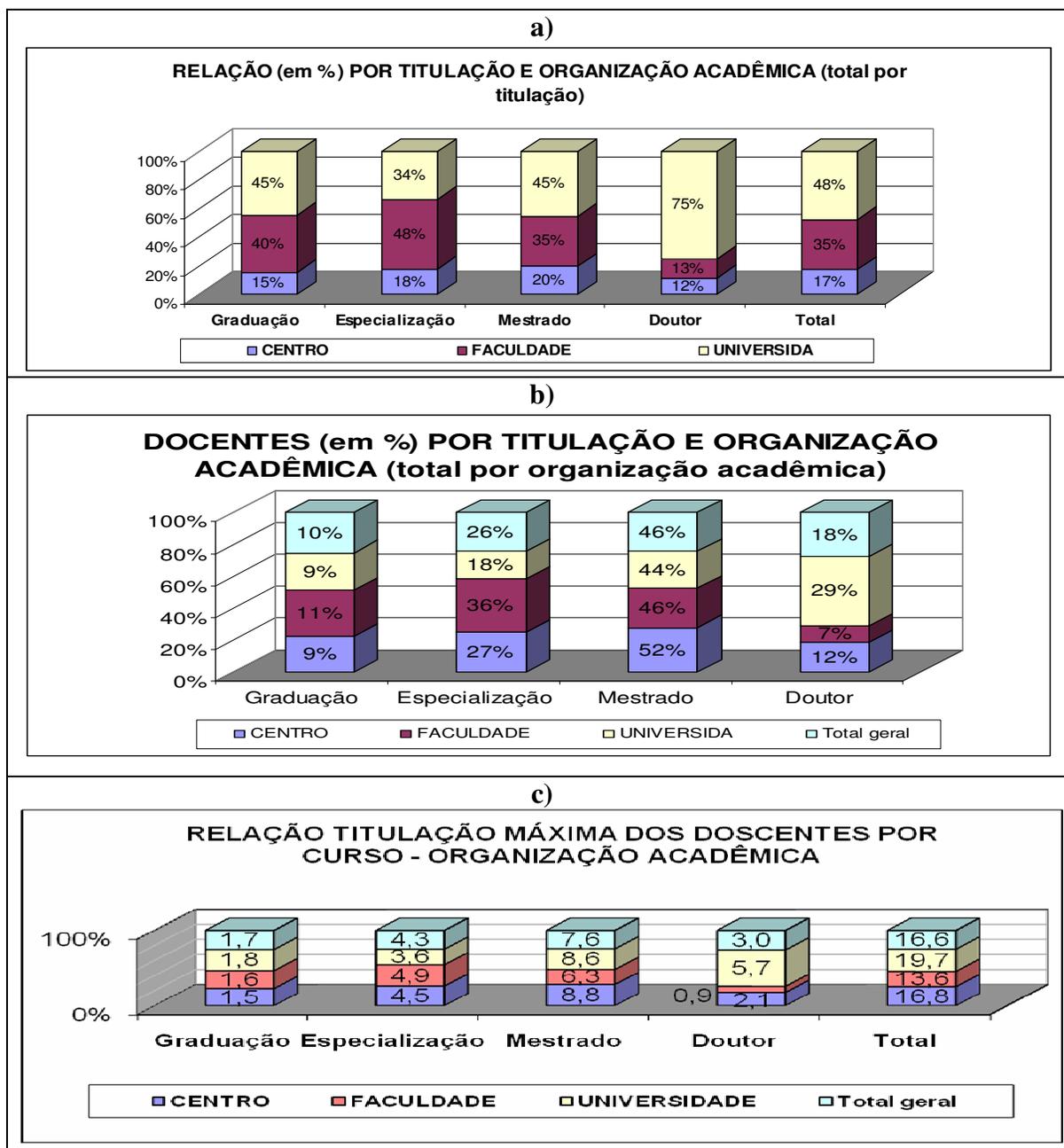


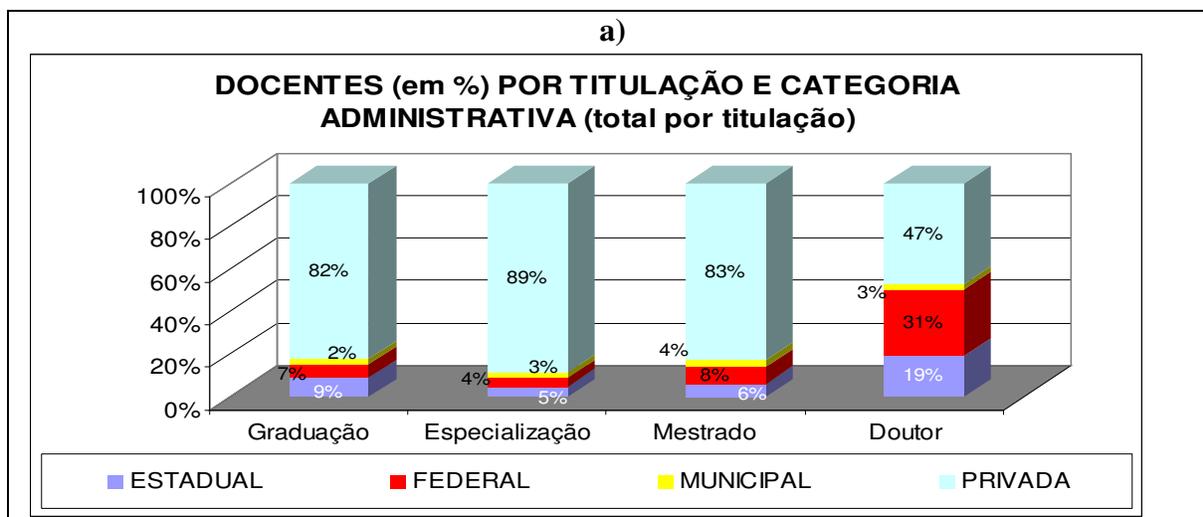
Gráfico 9 – Relação (em %) por Titulação e Organização Acadêmica (total por titulação) (a), Docentes (em %) por Titulação e Organização Acadêmica (total por Organização Acadêmica) (b) e Relação Titulação Máxima dos Docentes por curso (total por Organização Acadêmica) (c).

Fonte: Elaboração dos Autores, 2007.

A maioria dos docentes com mestrado, com graduação e/ou com doutorado, está nas universidades (gráfico 9a), indicando que somados, estão nas universidades 48% (9a) dos doutores do país, que são 18% (9b) dos docentes do país, e que também possuem mais docentes por curso, sendo de quase 20 (vinte) docentes para cada curso (9c). Os doutores correspondem a média de 5,7 docentes por curso (9c) nas universidades, o oposto das faculdades, que possuem uma média de menos de 1 (um) de docente com doutorado por curso (9c). Porém, quando se fala de docentes com especialização, as faculdades é que levam a vantagem, com 4,9 docentes por curso, contra os 3,6 das universidades (9c).

Nos centros acadêmicos e faculdades tem-se mais mestres que doutores, o contrário do que ocorre nas universidades (9a). Nas faculdades aparecem mais docentes com especialização, representando quase 50% do total do país (9a), e a minoria destes docentes está nos centros acadêmicos. No entanto, os centros acadêmicos são os que possuem a maior média de docentes com mestrado (9c). As universidades possuem a maioria de seus docentes com a titulação de mestre, representando 44% (9b). Em relação aos docentes apenas com graduação nota-se que existe um equilíbrio entre os três tipos de organizações acadêmicas, com a média ficando entre 1,5 e 1,8 de docentes por curso (9c).

4.3.2.1.3 Por categoria administrativa



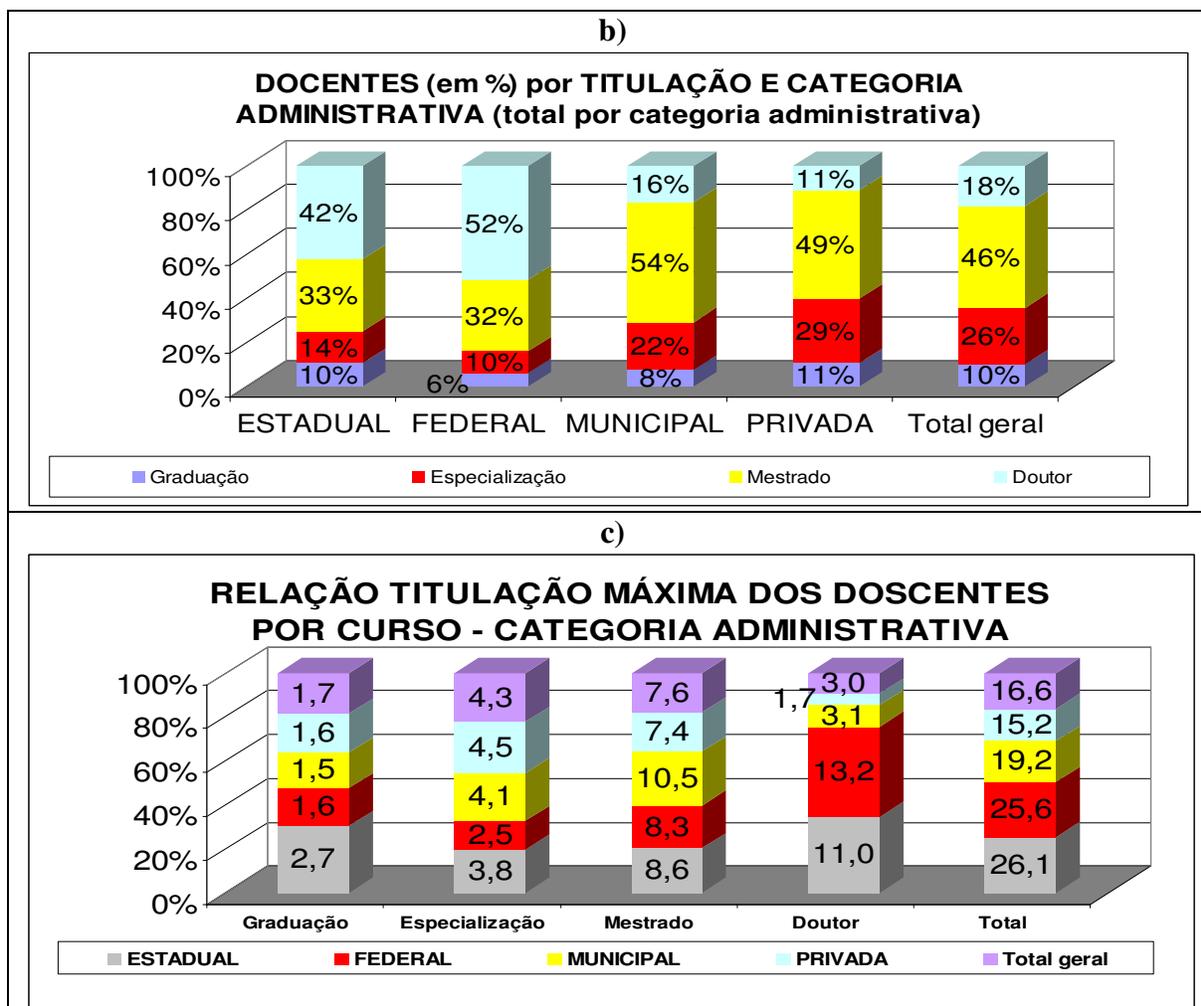


Gráfico 10 – Docentes (em %) por titulação e categoria administrativa total por categoria administrativa (b) e por titulação (a), e Relação titulação máxima dos docentes por curso por categoria administrativa (c).
Fonte: Elaboração dos Autores, 2007.

Por meio da análise da relação docentes/curso por categoria administrativa pode-se ver que as instituições Estaduais são as que possuem maior média de professores por curso (26) e que as instituições Privadas são as que possuem a menor média (15), sendo que a diferença entre as duas é de mais de 10 docentes. Nota-se que as instituições Federais e Estaduais possuem muito mais docentes com doutorado do que as instituições Municipais e Privadas, e ainda que as Federais são as que menos possuem docentes com especialização.

As instituições Municipais possuem a maioria dos seus docentes com mestrado, mas têm apenas 2% de docentes com graduação. As instituições Privadas possuem 82% dos docentes com graduação (gráfico 10a), além da grande maioria dos docentes com especialização, com mestrado, e quase a metade dos docentes com doutorado (10a) e, portanto são as que possuem mais docentes no geral.

4.3.2.1.4 Por tempo médio do docente

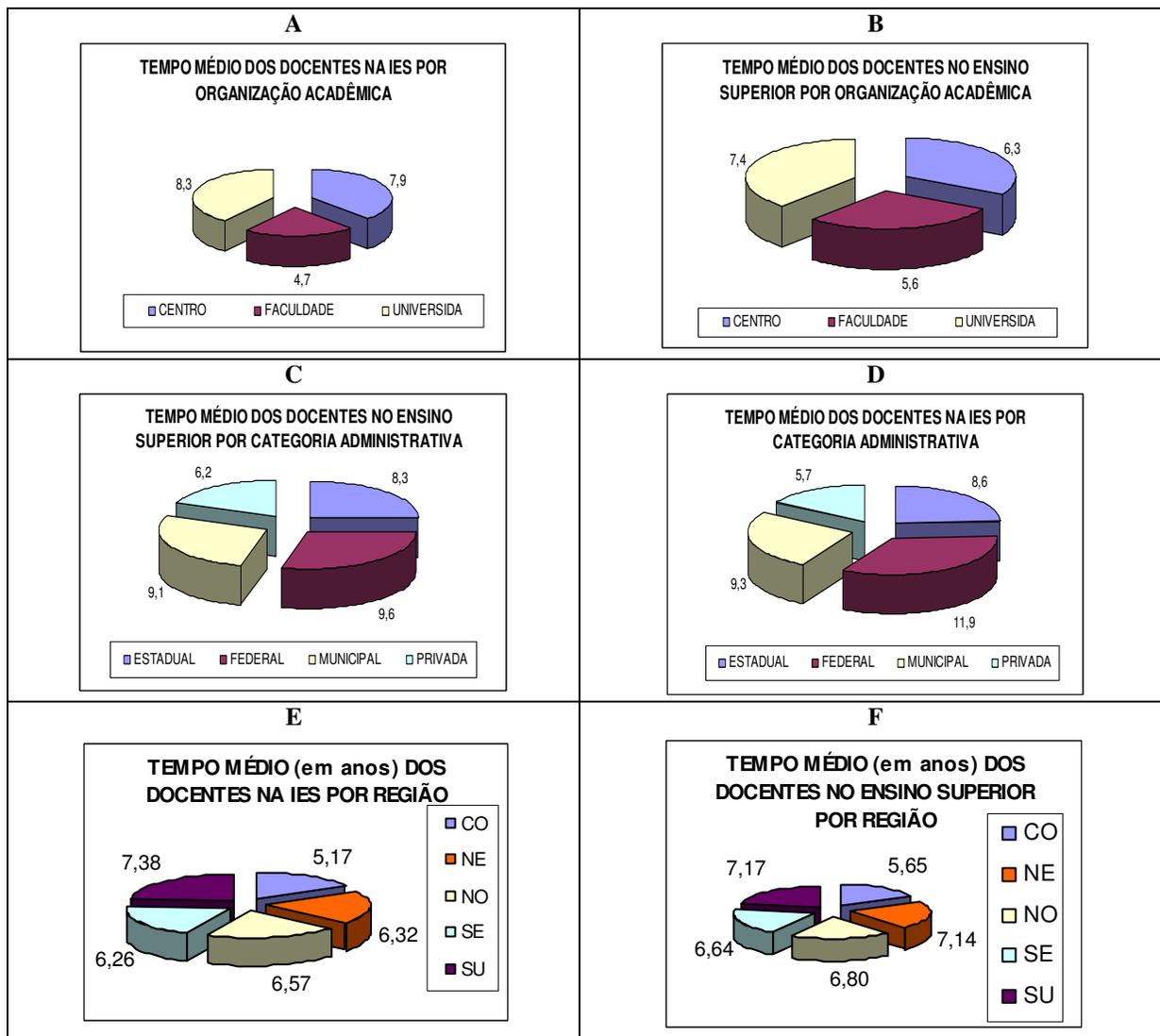


Gráfico 11 – Análises por tempo médio dos docentes.
Fonte: Elaboração dos Autores, 2007.

Através dos gráficos acima (gráfico 11) nota-se que os docentes da região Sul são os que permanecem mais tempo no ensino superior, e na mesma instituição, sendo de 7,38 a média de anos na mesma IES. Pode-se ver que os docentes que possuem a maior média de tempo no Ensino Superior, assim como na mesma IES, são os das instituições Federais e Municipais. As Universidades Federais são as que possuem docentes que estão mais tempo na mesma IES, e com

maior tempo médio no Ensino Superior, ou seja, são as que possuem docentes com mais experiência no ensino.

4.3.3 Análises de Discentes

As variáveis utilizadas para as análises de Discentes foram:

a) **Região Geográfica:** São as 05 (cinco) regiões geográficas do Brasil. Sendo elas:

- Norte (NO)
- Nordeste (NE)
- Centro-Oeste (CO)
- Sudeste (SE)
- Sul (SU)

b) **Organização Acadêmica:** Organização acadêmica da IES. Neste trabalho estas categorias são divididas em:

- Centro
- Faculdade
- Universidade

Onde:

Centro representa todos os: Centros Federais de Educação Tecnológica, Centros Universitários e Centros Universitários Especializados.

Faculdade representa todas as: Faculdades Integradas, Faculdades, Faculdades de Tecnologia, Institutos Superiores e Escolas Superiores.

Universidade representa: Universidades.

c) **Categoria Administrativa:** Categoria administrativa da IES. Sendo estas categorias divididas em:

- **Federal** (Mantidas e administradas pelo Governo Federal).
- **Estadual** (Mantidas e administradas pelos governos dos estados).
- **Municipal** (Mantidas e administradas pelo poder público municipal).
- **Privada** (São as mantidas e administradas por pessoas físicas ou jurídicas de direito privado).

- d) **Vagas:** Representa o total de vagas oferecidas pelos cursos.
- e) **Inscritos:** Representa o total de inscritos para ocupar as vagas oferecidas.
- f) **Ingressos:** Representa o total de ingressos que vão ocupar as vagas oferecidas.
- g) **Matrículas:** Representa o total de matrículas.
- h) **Concluintes:** Representa o total de concluintes.
- i) **Ano de Início:** Representa o ano de início do curso na IES. Sendo dividido em:

- 1969 - 1978
- 1979 - 1988
- 1989 - 1998
- 1999 – 2008

- j) **Área Específica:** Representa a área específica do curso. Sendo dividida em:

- Administração de redes
- Análises de sistemas
- Banco de Dados
- Ciência da Computação
- Comércio Eletrônico
- Computação Gráfica
- Engenharia de Computação (hardware)
- Engenharia de softwares
- Informática (ciência da computação)
- Informática Educacional
- Linguagens de Programação (visual basic, c++ etc)
- Processamento de dados
- Segurança da Informação
- Sistemas de Informação
- Sistemas Operacionais
- Tecnologia da informação
- Tecnologia em desenvolvimento de softwares
- Tecnologia em informática
- Uso da Internet

- k) **Curso:** Representa o total de cursos.

4.3.3.1 As análises

As análises de discentes foram divididas em cinco partes. Sendo elas pelas variáveis: por região, por organização acadêmica, por categoria administrativa, por ano de início e por área específica do curso. As subseções 4.3.3.1.6 e 4.3.3.1.7 apresentam as relações inscritos/vagas e ingressos/concluintes respectivamente, com gráficos relacionados a todas estas cinco variáveis.

4.3.3.1.1 Por região

A partir do gráfico 12, nota-se que a região Sudeste é a região que possui o maior número de vagas, inscritos, ingressos, matrículas, concluintes e cursos. Representando mais de 50% em todas as variáveis acima mencionadas. Outra região que se destaca, representando a segunda região que mais possui cursos e vagas é a região Sul.

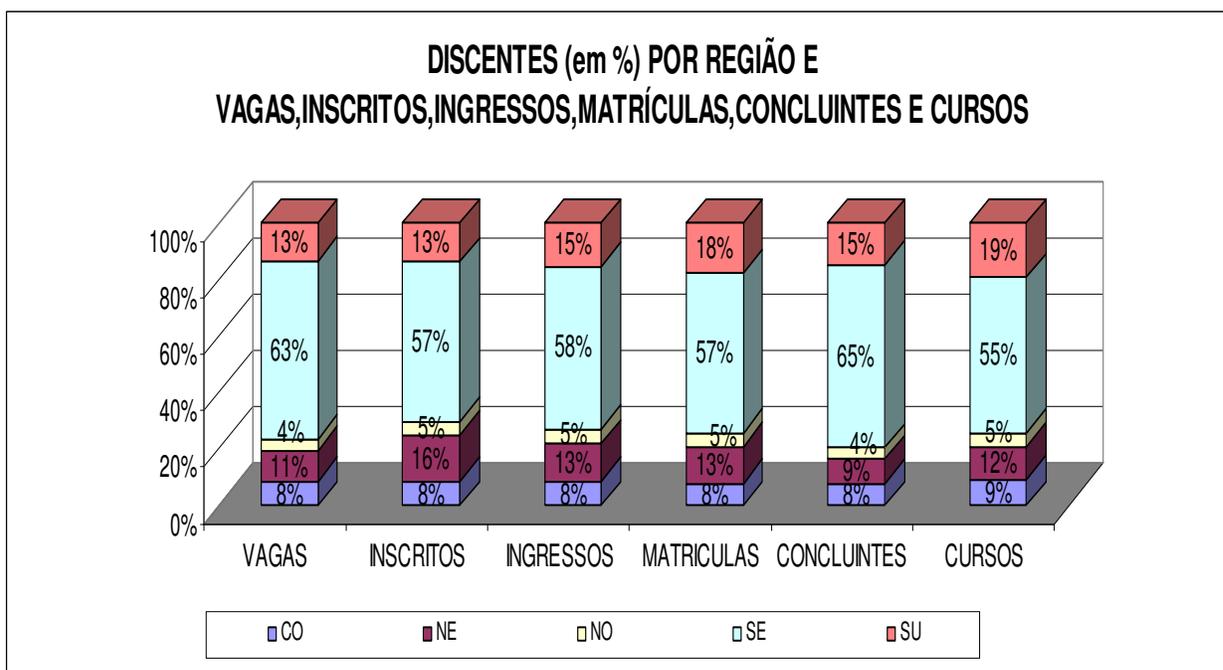


Gráfico 12 – Discentes (em %) por região e vagas, inscritos, ingressos, matrículas, concluintes e cursos – região.
Fonte: Elaboração dos Autores, 2007.

O gráfico 13 apresenta dados sobre as médias vagas/curso, inscritos/curso, ingressos/curso, matrículas/curso, conluíntes/curso por região e ainda uma média geral do país. Nota-se que enquanto a região Sudeste possui a maior média de vagas por curso, a região Sul possui a menor média. Percebe-se também que a região Nordeste é a que possui mais inscritos por curso e que a região Sul é a que menos possui.

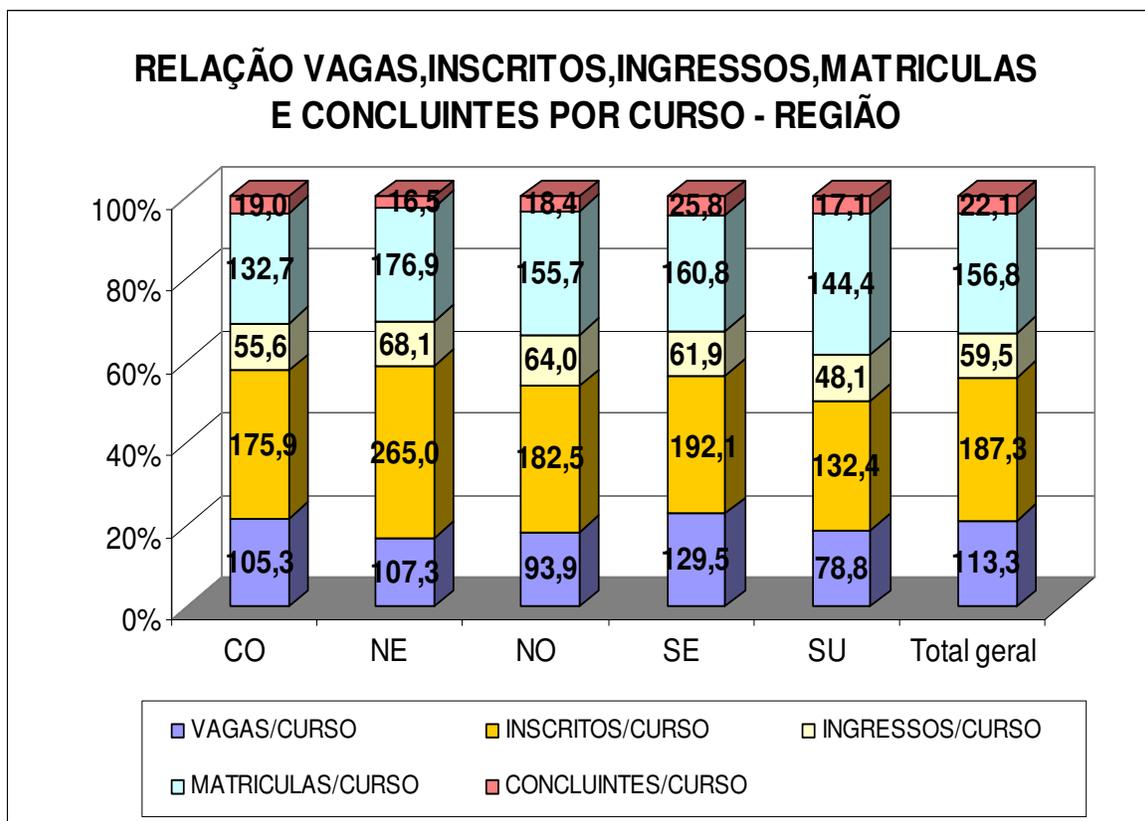


Gráfico 13 – Relação vagas, inscritos, ingressos, matrículas e conluíntes por curso – região.
Fonte: Elaboração dos Autores, 2007.

É importante notar no gráfico 13, que o número de inscritos supera amplamente o de vagas oferecidas, no entanto os ingressos são pouco mais da metade das vagas oferecidas, e o número de conluíntes é aproximadamente 35% dos ingressos.

4.3.3.1.2 Por organização acadêmica

Nota-se no gráfico 14 que a maioria das vagas estão nas Faculdades e Universidades. Apesar de as Faculdades obterem o maior número de vagas, a maioria dos inscritos está nas Universidades, representando quase 50% do total do país. Porém apesar de menos inscritos, são as Faculdades que possuem o maior número de ingressos.

Em relação ao número de matrículas são as Universidades que possuem o maior número, com 47% do total. Assim como possuem o maior número de matrículas, as Universidades também possuem o maior número de concluintes.

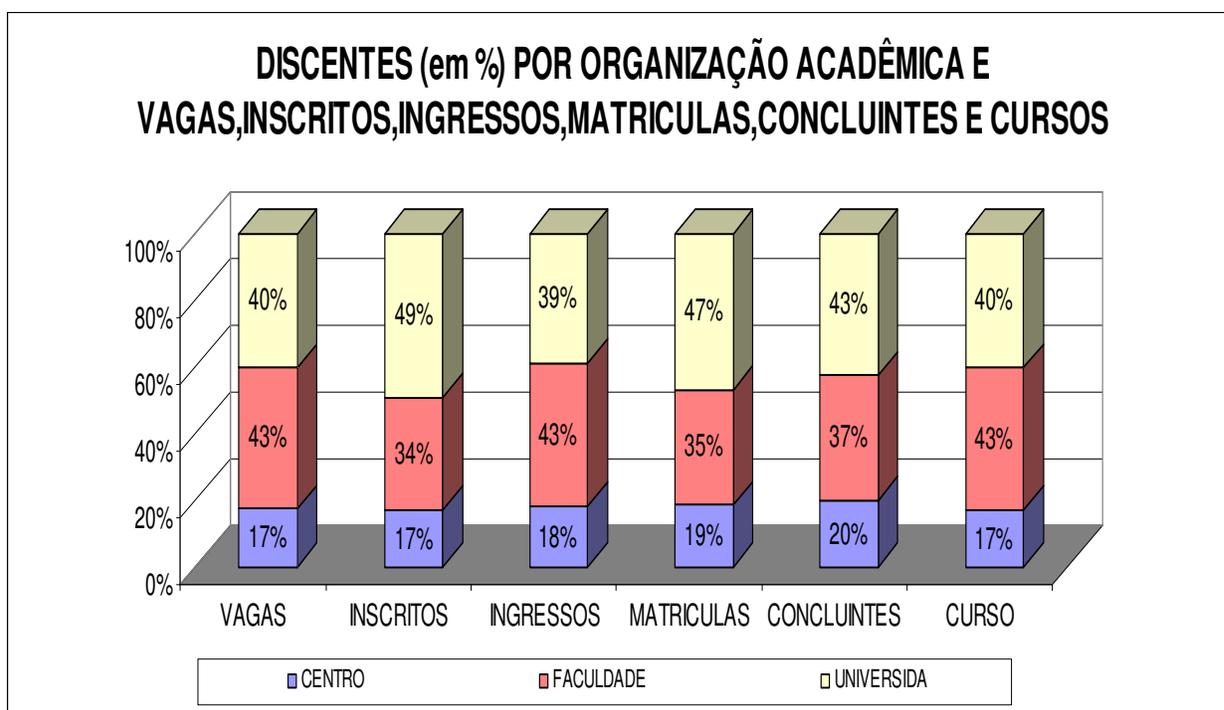


Gráfico 14 – Discentes (em %) por organização acadêmica e vagas, inscritos, ingressos, matrículas, concluintes e cursos.

Fonte: Elaboração dos Autores, 2007.

Por meio do gráfico 15 percebe-se que existe um equilíbrio na média de vagas/curso por organização acadêmica e que as Universidades possuem mais inscritos por curso do que as Faculdades e Centros Acadêmicos.

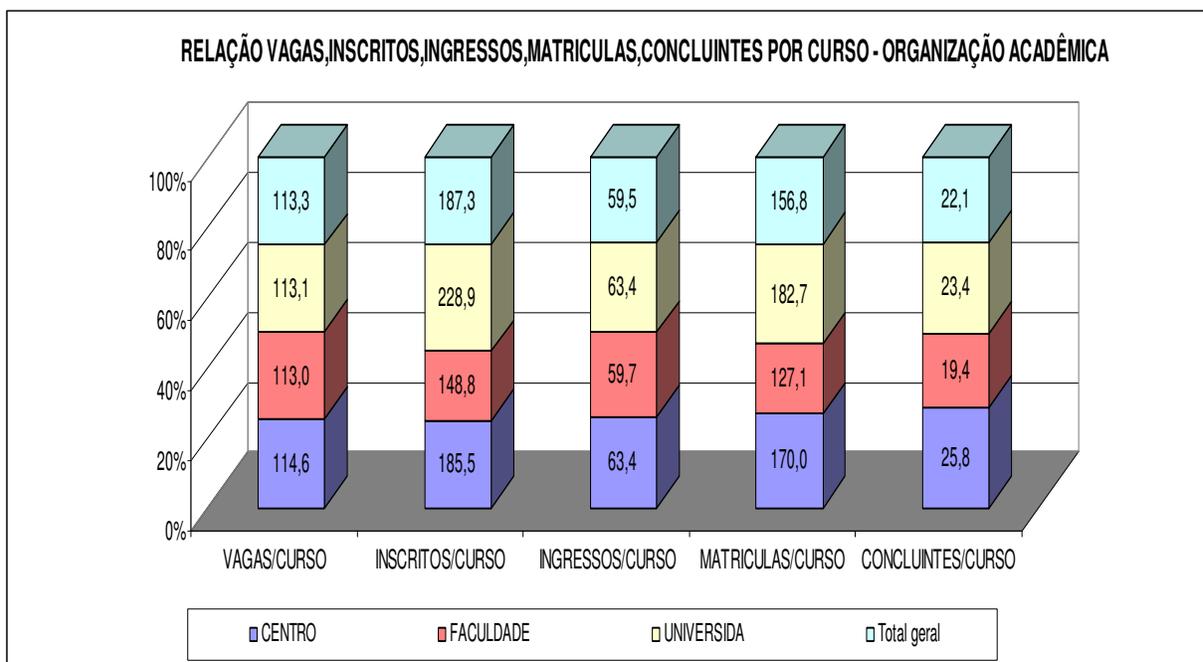


Gráfico 15 – Relação vagas, inscritos, ingressos, matrículas, concluintes por curso – organização acadêmica.
Fonte: Elaboração dos Autores, 2007.

4.3.3.1.3 Por categoria administrativa

Nota-se explicitamente, no gráfico 16, que a grande maioria das vagas está nas instituições privadas. Sendo que somente 8% das vagas pertencem às instituições públicas (3% em Federais, 3% em Estaduais e 2% Municipais).

A maioria dos inscritos também encontram-se nas instituições privadas, representando 62% do total. Percebe-se que apenas 2% são instituições Municipais. 85% dos discentes que ingressam pertencem a instituições Privadas, ou seja, a grande maioria que entra em uma IES está em instituições Privadas. Assim também como os ingressos, as matrículas estão todas em grande maioria nas instituições Privadas. (Gráfico 16)

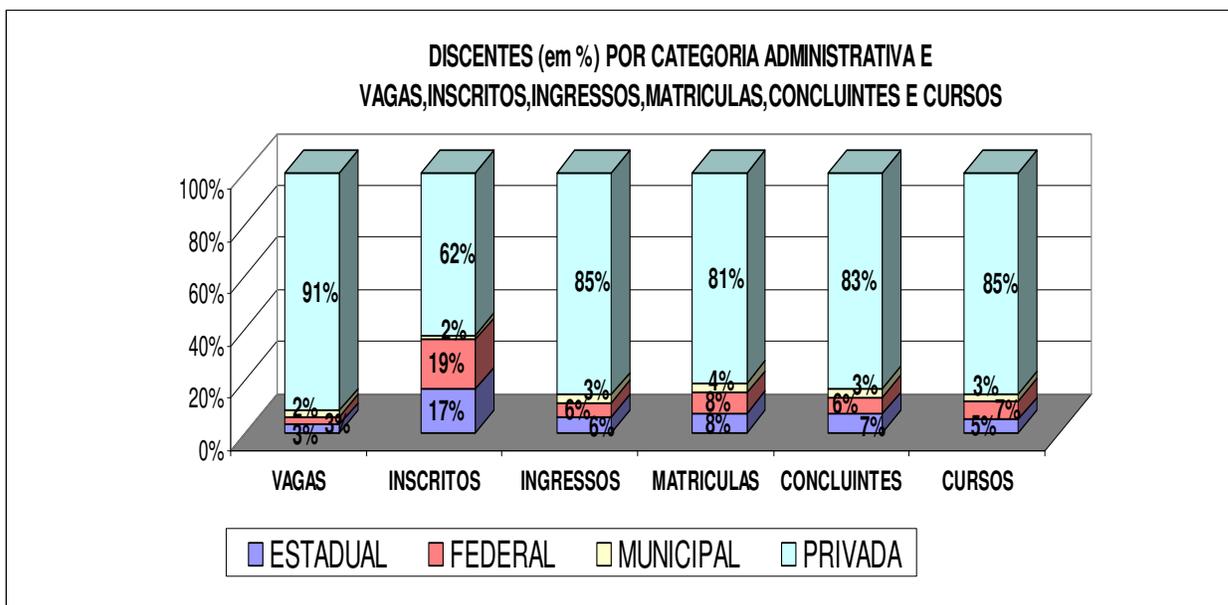


Gráfico 16 – Discentes (em %) por categoria administrativa e vagas, inscritos, ingressos, matrículas, concluintes e cursos.

Fonte: Elaboração dos Autores, 2007.

As instituições estaduais são as que possuem maior média de inscritos, ingressos e matrículas por curso. As privadas e municipais apresentam o maior número de vagas por curso e as federais possuem a média mais baixa de concluintes por curso. (Gráfico 17)

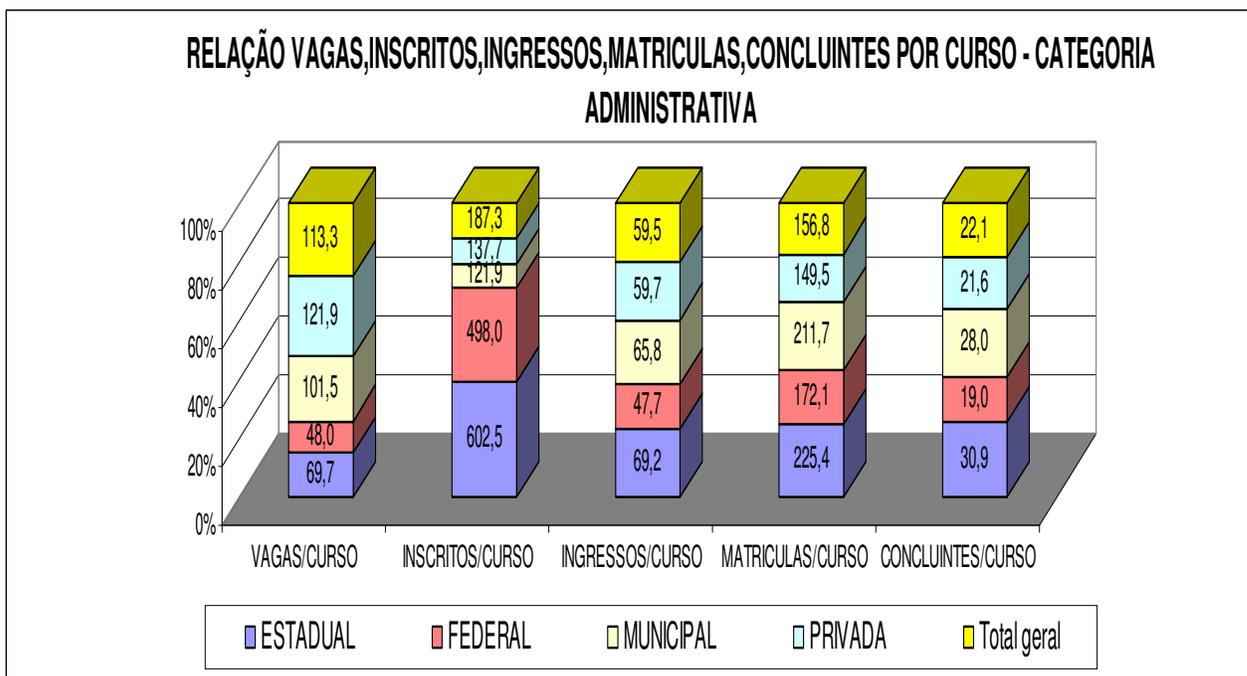


Gráfico 17 – Relação vagas, inscritos, ingressos, matrículas, concluintes por curso – categoria administrativa.

Fonte: Elaboração dos Autores, 2007.

4.3.3.1.4 Por ano de início do curso

Por meio do gráfico abaixo (gráfico 18) observa-se que a maioria das IES que possuem cursos na área das ciências da computação no Brasil são IES novas, possuindo elas menos de 09 (nove) anos de existência. Apenas 1% dos cursos nesta área existem há mais de 30 (trinta) anos.

A grande maioria dos inscritos, ingressos, matrículas e concluintes são de IES com menos de 09 (nove) anos de existência. Isso significa que atualmente muito mais pessoas cursam o ensino superior na área de ciência da computação do que se comparado a 30/40 anos atrás.

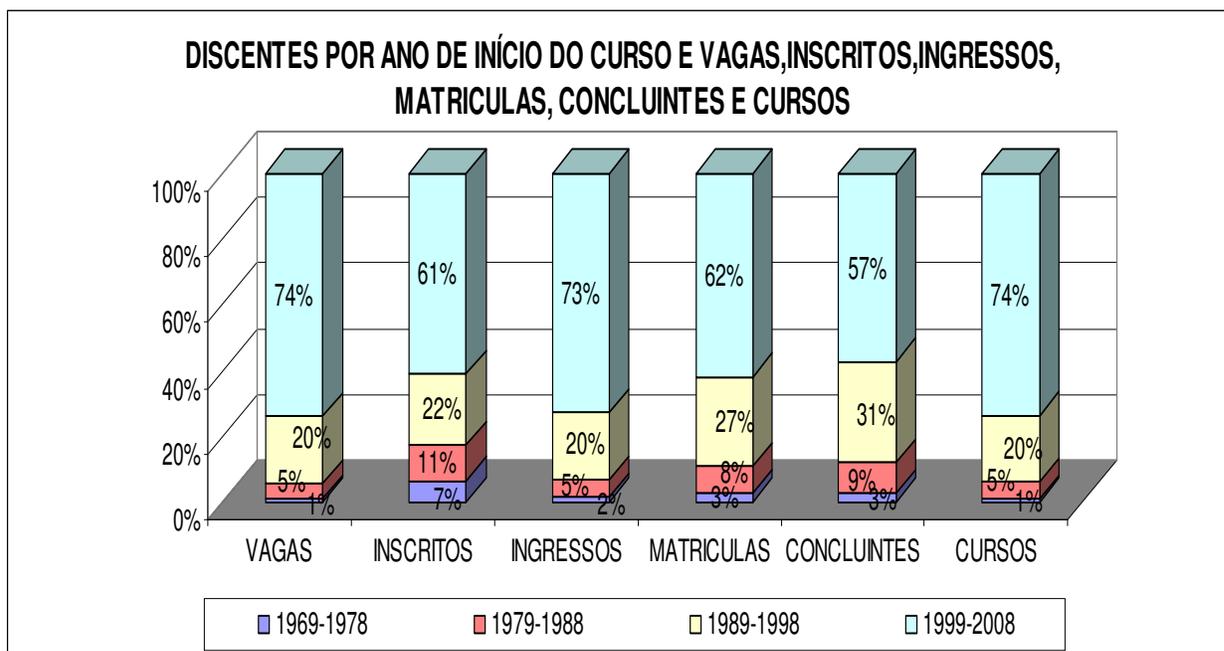


Gráfico 18 – Discentes por ano de início do curso e vagas, inscritos, ingressos, matrículas, concluintes e cursos. Fonte: Elaboração dos Autores, 2007.

Analisando o gráfico 19, nota-se que a relação de inscritos por curso é absurdamente maior nas instituições mais antigas, que possuem entre 30 e 40 anos de existência.

A relação vagas/curso é praticamente a mesma tanto nos cursos mais antigos quanto nos cursos mais novos. As instituições mais antigas possuem maior média de ingressos, matrículas e concluintes por curso.

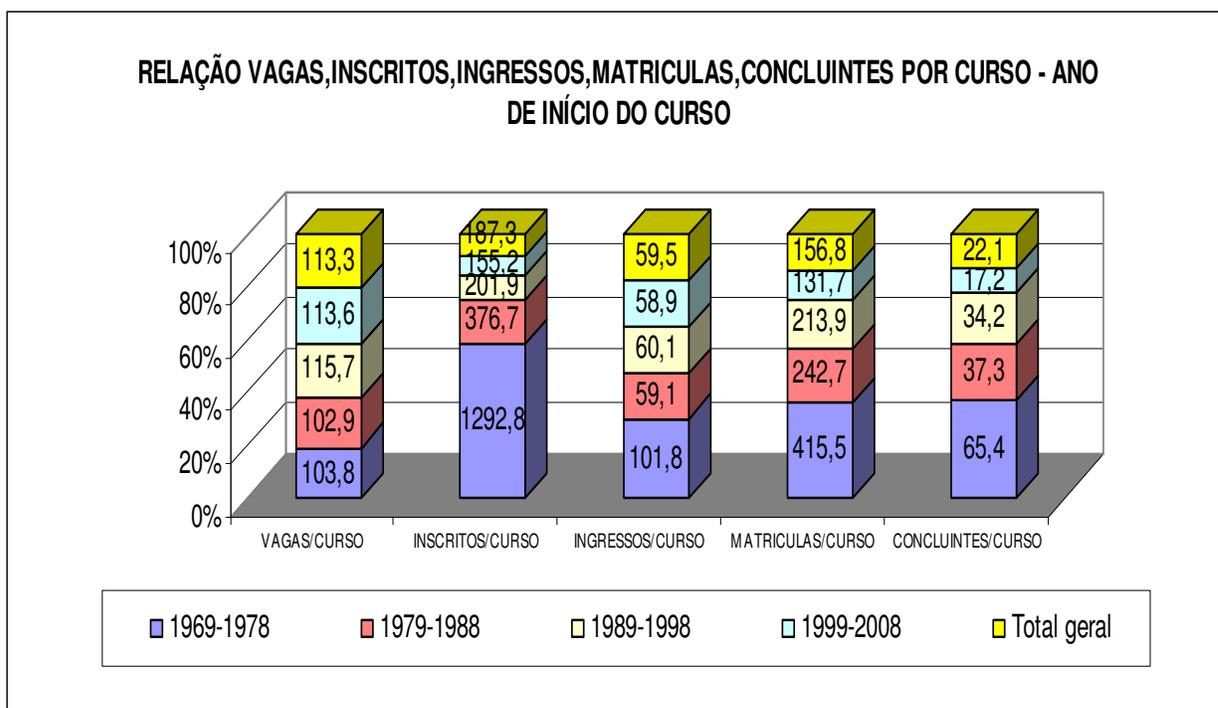


Gráfico 19 – Relação vagas, inscritos, ingressos, matrículas, concluintes por curso – ano de início do curso.
Fonte: Elaboração dos Autores, 2007.

4.3.3.1.5 Por área específica do curso

Por meio da tabela 1 nota-se que a maioria das vagas estão distribuídas basicamente entre os cursos com área específica de: Sistemas de Informação, Ciência da Computação, Tecnologia da Informação e Administração de Redes. Sendo que a maior parte dos inscritos é de Sistemas de Informação e Ciência da Computação.

Os ingressos também são bem maiores em Sistemas de Informação e Ciência da computação. Porém Sistemas da Informação tem uma porcentagem bem maior que Ciência da computação, tendo quase o dobro dos ingressos de Ciência da Computação.

Assim como o número de matrículas, o número de concluintes também é maior em Ciência da Computação e Sistemas de Informação. Percebe-se assim que estes cursos são os predominantes no país.

Também nota-se que existem pouquíssimos cursos de Sistemas Operacionais e Informática Educacional, existindo somente um curso de cada em todo o país.

ÁREA ESPECÍFICA	VAGAS	INSCRITOS	INGRESSOS	MATRICULAS	CONCLUINTES	CURSOS
Administração de redes	9,81%	7,91%	10,45%	5,28%	6,86%	9,03%
Análise de sistemas	2,35%	3,40%	3,06%	3,22%	3,43%	2,53%
Banco de dados	1,28%	0,45%	0,57%	0,31%	0,53%	1,19%
Ciência da computação	26,06%	31,50%	22,60%	29,52%	28,15%	21,06%
Comércio eletrônico	0,22%	0,03%	0,05%	0,08%	0,16%	0,40%
Computação gráfica	0,46%	0,12%	0,25%	0,18%	0,07%	0,63%
Engenharia de computação (hardware)	1,12%	2,00%	1,19%	1,35%	0,76%	1,43%
Engenharia de softwares	0,02%	0,10%	0,04%	0,06%	0,04%	0,08%
Informática (ciência da computação)	1,08%	1,63%	1,32%	2,34%	1,81%	1,43%
Informática educacional	0,03%	0,01%	0,03%	0,05%	0,20%	0,08%
Linguagens de programação (visual basic, c++ etc)	0,17%	0,06%	0,16%	0,12%	0,03%	0,16%
Processamento de dados	3,86%	6,48%	4,23%	5,18%	7,59%	5,62%
Segurança da informação	0,35%	0,29%	0,28%	0,18%	0,36%	0,40%
Sistemas de informação	39,48%	33,11%	41,48%	40,91%	37,61%	41,57%
Sistemas operacionais	0,04%	0,02%	0,04%	0,02%	0,00%	0,08%
Tecnologia da informação	0,38%	0,10%	0,17%	0,13%	0,39%	0,48%
Tecnologia em desenvolvimento de softwares	2,15%	2,20%	2,72%	1,48%	0,82%	2,14%
Tecnologia em informática	10,74%	10,25%	10,95%	9,29%	10,74%	11,32%
Uso da internet	0,39%	0,34%	0,40%	0,29%	0,46%	0,40%
Total geral	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Tabela 1 – Tabela de discentes (em %) por área específica e vagas, inscritos, ingressos, matriculas, concluintes e cursos.

Fonte: Elaboração dos Autores, 2007.

ÁREA ESPECÍFICA	VAGAS	INSCRITOS	INGRESSOS	MATRICULAS	CONCLUINTES
Administração de redes	123,2	164,0	69,0	91,8	16,8
Análise de sistemas	105,0	250,9	72,0	199,3	29,9
Banco de dados	122,2	70,6	28,5	41,5	9,8
Ciência da computação	140,2	280,1	63,9	219,7	29,5
Comércio eletrônico	62,0	15,0	7,8	32,2	9,2
Computação gráfica	83,1	36,1	23,3	45,1	2,4
Engenharia de computação (hardware)	89,1	262,3	49,9	148,2	11,8
Engenharia de softwares	30,0	241,0	31,0	120,0	11,0
Informática (ciência da computação)	85,9	214,6	55,3	257,3	28,0
Informática educacional	50,0	19,0	19,0	106,0	55,0
Linguagens de programação (visual basic, c++ etc)	125,0	75,5	60,5	117,0	3,5
Processamento de dados	77,7	215,8	44,8	144,5	29,8
Segurança da informação	100,4	138,8	42,2	70,4	20,0
Sistemas de informação	107,6	149,2	59,4	154,3	20,0
Sistemas operacionais	60,0	57,0	32,0	45,0	0,0
Tecnologia da informação	91,7	39,3	21,5	43,5	18,2
Tecnologia em desenvolvimento de softwares	114,2	192,8	75,8	108,4	8,5
Tecnologia em informática	107,5	169,5	57,6	128,6	21,0
Uso da internet	112,0	159,4	59,8	116,6	25,4
Total geral	113,3	187,3	59,5	156,8	22,1

Tabela 2 – Tabela de relação vagas, inscritos, ingressos, matrículas, concluintes por curso – área específica.
Fonte: Elaboração dos Autores, 2007.

Os cursos com área específica em Ciência da Computação são os que apresentam maior número de vagas, inscritos e matrículas por curso.

Outra informação que chama atenção é em relação ao número de vagas oferecidas por Engenharia de Software, sendo de 30 vagas por curso, número muito menor do que os cursos de outras áreas específicas.

Em relação ao número de ingressos, a área específica que possui mais ingressos é Tecnologia em desenvolvimento de softwares e a que menos possui é Comércio eletrônico.

A respeito dos concluintes nota-se que Informática Educacional é o que possui a maior média; e Sistemas Operacionais, Computação Gráfica e Linguagens de Programação são os que menor média de concluintes por curso.

4.3.3.1.6 Relação inscritos/vagas

O gráfico 20 apresenta a relação inscritos/vagas por meio de quatro perspectivas diferentes. Nele nota-se que a região que possui maior índice candidato/vaga, ou seja, a concorrência é maior para entrar em um curso, é a região Nordeste com 2,47 candidatos para uma vaga. Percebe-se também que a concorrência é muito maior nas Universidades, sendo ela de dois candidatos por uma vaga, e nas instituições Federais chegando a 10,4 candidatos para uma única vaga. Numero este que se comparado às instituições Privadas e Municipais é quase 10 (dez) vezes maior. Em relação ao ano de início do curso, o gráfico 20c mostra que a concorrência por uma vaga em IES que oferecem cursos na área de ciência da computação foi diminuindo com o passar do tempo, pois como nota-se antigamente a concorrência por uma vaga era muito maior do que é hoje em dia.

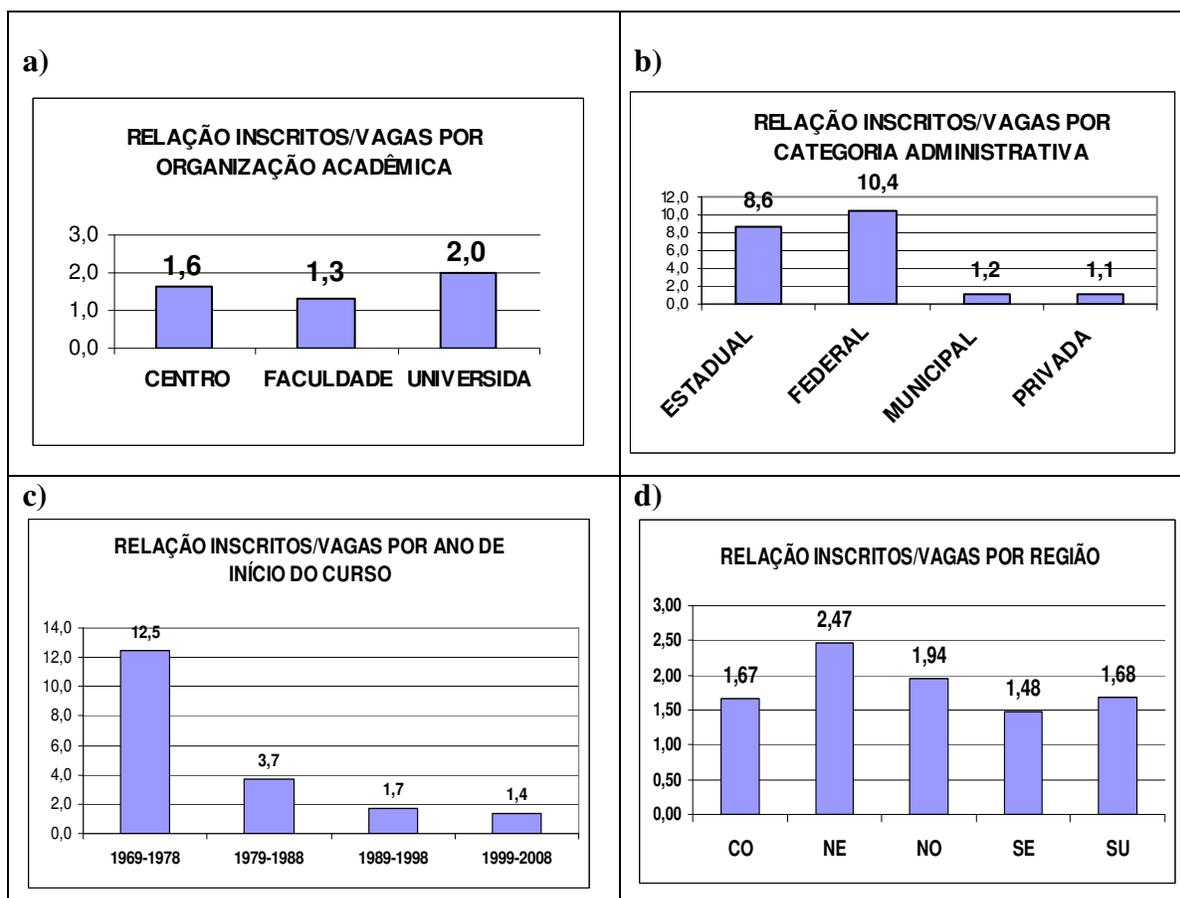


Gráfico 20 – Relações inscritos/vagas.
Fonte: Elaboração dos Autores, 2007.

Através do gráfico abaixo (gráfico 21) podemos notar que a maior concorrência está nos cursos de Engenharia de Software, tendo um índice de 08 (oito) candidatos para uma vaga.



Gráfico 21 – Relação inscritos/vagas por área específica.
Fonte: Elaboração dos Autores, 2007.

4.3.3.1.7 Relação ingressos/concluintes

Por meio do gráfico 22b percebe-se que as instituições privadas possuem a média ingressos/concluintes mais alta, o que significa que em relação ao número de ingressos as instituições privadas possuem menos concluintes. O gráfico 22a mostra que são as faculdades que possuem menos concluintes em relação ao número de inscritos.

No gráfico 22c, pode-se ver que a maioria dos discentes que ingressavam entre 1969 e 1989, concluíam o curso, e que a região Nordeste é a que apresenta mais discentes que desistem do curso, pois a relação ingressos/concluintes é a mais alta (Gráfico 22d).

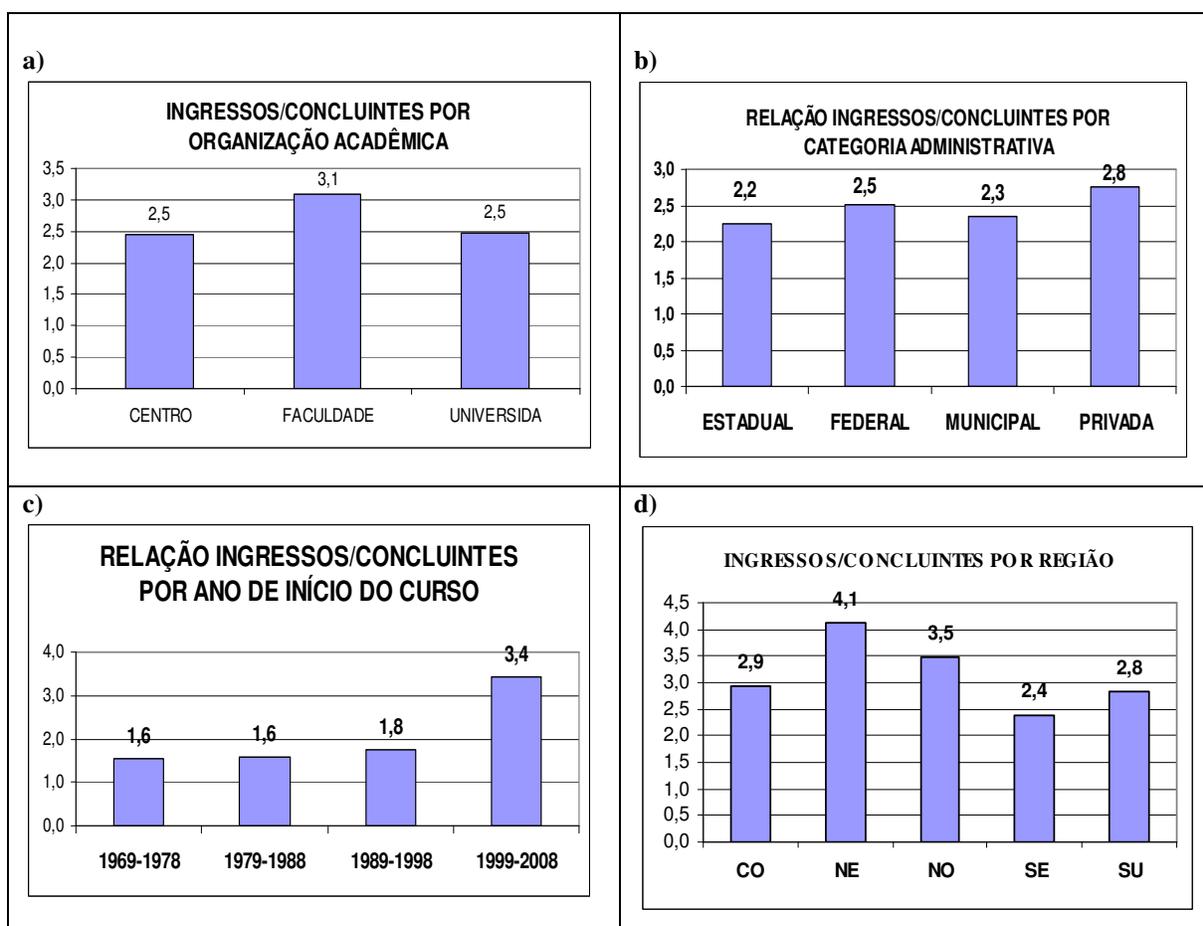


Gráfico 22 – Relações ingresso/concluinte.
Fonte: Elaboração dos Autores, 2007.

O gráfico 23 mostra que a maioria das desistências estão em sistemas operacionais, nas linguagens de programação, computação básica e tecnologia de desenvolvimento de softwares.

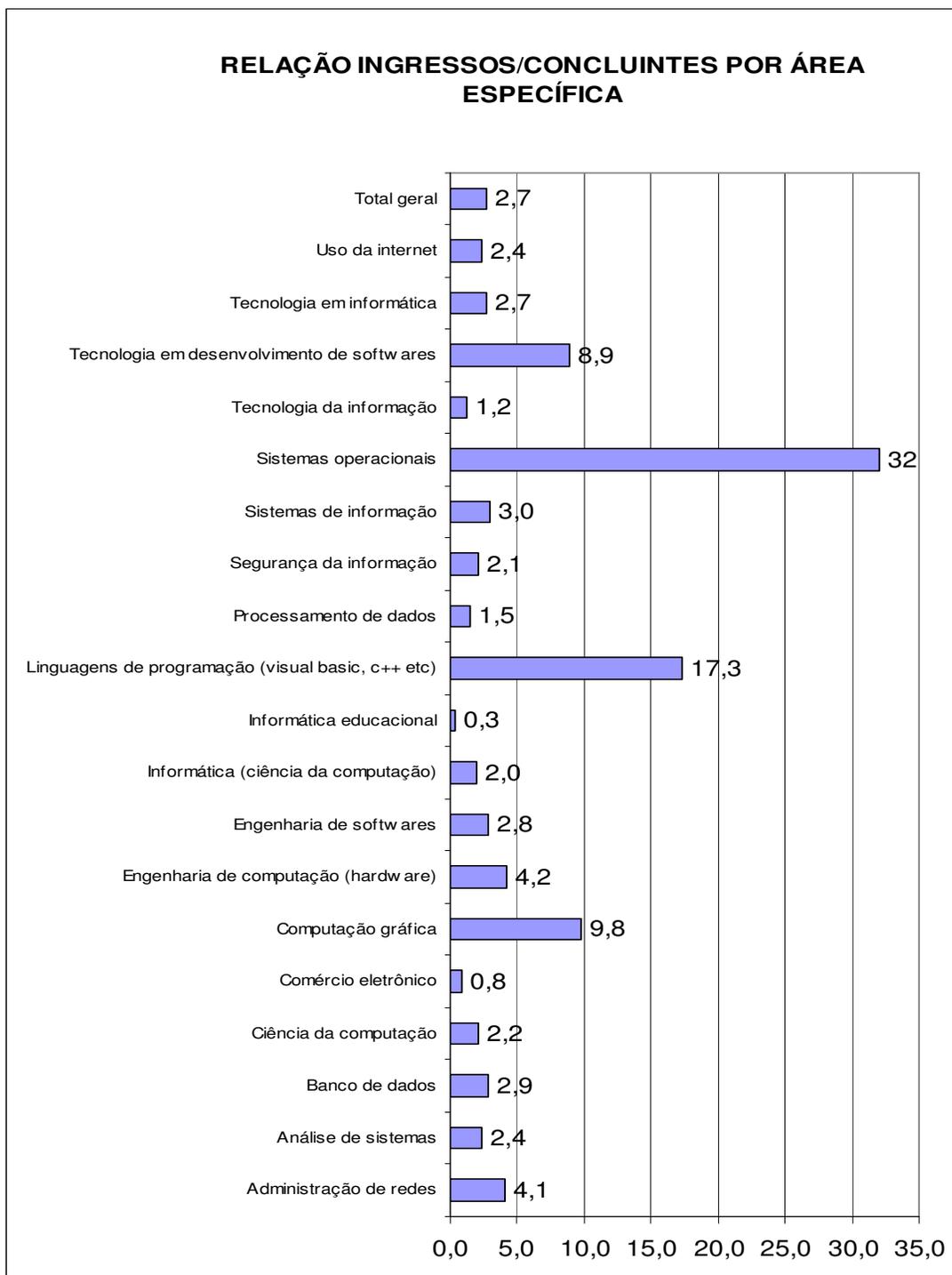


Gráfico 23 – Relação ingresso/concluinte por área específica.

Fonte: Elaboração dos Autores, 2007.

4.3.4 Análises das Empresas

As variáveis utilizadas para as análises Empresas foram:

- a) **Número de Empresas:** Representa o total de empresas.
- b) **Código CNAE:** É uma classificação padronizada utilizada pelo IBGE para identificação das unidades produtivas do Brasil. Significa classificação nacional de atividades econômicas. No caso deste trabalho foi somente levado em consideração as Atividades de informática e serviços relacionados (correspondente ao CNAE 72). São elas:
 - Consultoria em hardware (72.1)
 - Consultoria em software (72.2)
 - Processamento de dados (72.3)
 - Atividades de banco de dados e distribuição on-line de conteúdo eletrônico (72.4)
 - Manutenção e reparação de máquinas de escritório e de informática (72.5)
 - Outras atividades de informática, não especificadas anteriormente (72.9)
- c) **Região Geográfica:** São as 05 (cinco) regiões geográficas do Brasil. Sendo elas:
 - Norte (NO)
 - Nordeste (NE)
 - Centro-Oeste (CO)
 - Sudeste (SE)
 - Sul (SU)
- d) **Faixa Pessoal:** É uma classificação utilizada pelo IBGE para classificar empresas pelo número de funcionários.
- e) **Salários:** Representa o total dos salários.
- f) **Número de Pessoas Ocupadas:** Representa o total de pessoas da empresa.
- g) **Número de Pessoas Salário ou Número de Pessoas Assalariadas:** Representa o total de pessoas assalariadas da empresa.

4.3.4.1 As análises

As análises dos dados das empresas foram divididas em quatro categorias, sendo elas: pelo número de empresas, pela soma de salários, pelo número de pessoas ocupadas e pelo número de pessoas salário, como apresentada a seguir.

4.3.4.1.1 Pelo número de empresas

O gráfico 24 apresenta o total de empresas em duas perspectivas diferentes, uma por meio da variável faixa pessoal e a outra por meio da variável região. Assim, nota-se que a grande maioria das empresas de TI encontram-se concentradas na região Sudeste do país. Em seguida está a região Sul, que se comparada com as outras três regiões (Nordeste, Centro-Oeste e Norte) representa um número consideravelmente maior (gráfico 24b). A maior parte destas empresas são formadas por poucas pessoas (de 0 a 4) e que existem um numero reduzido de empresas com mais de 500 pessoas em nosso país. (Gráfico 24a)

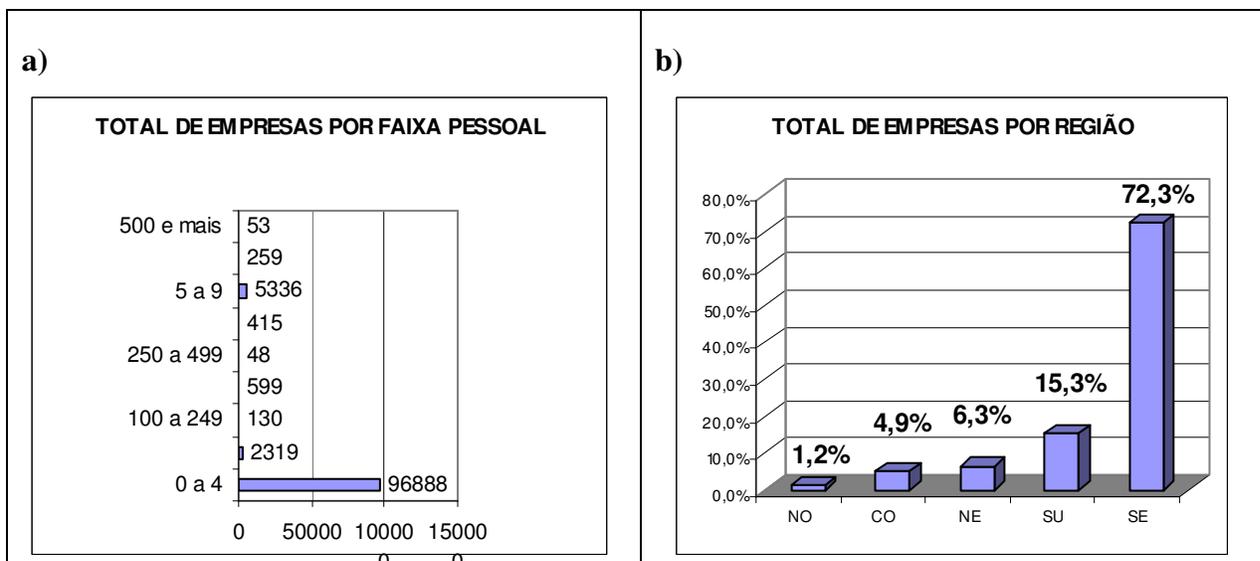


Gráfico 24 – Análises total de empresas.
Fonte: Elaboração dos Autores, 2007.

Quanto ao número de empresas por CNAE pode-se notar uma grande diferença na área de Banco de Dados e distribuição On-line de conteúdo eletrônico em relação às outras áreas. (Gráfico 25)

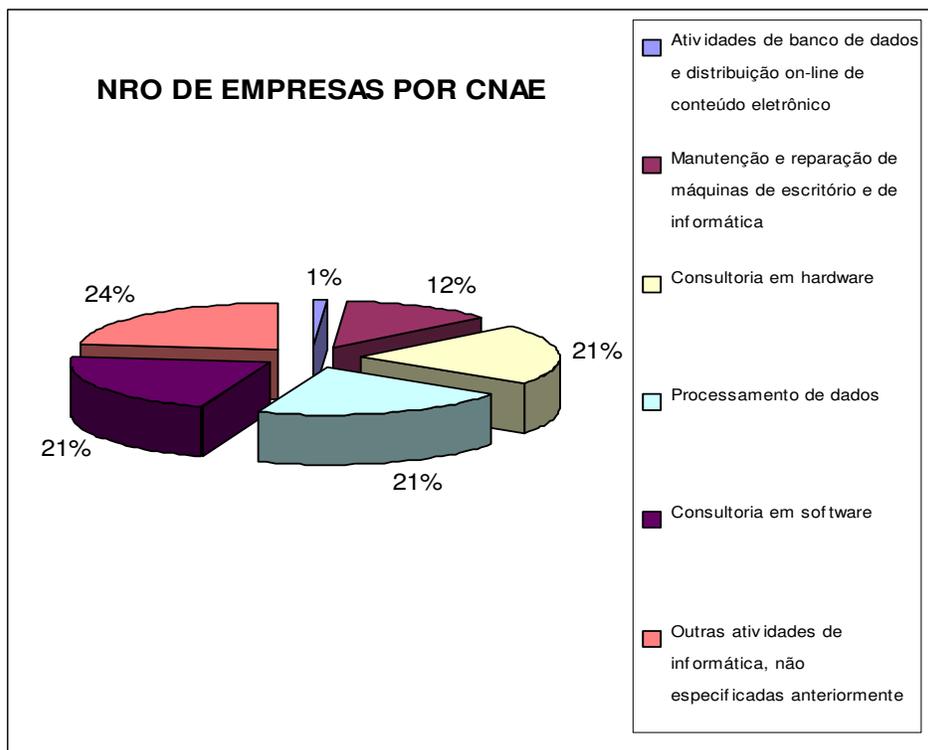


Gráfico 25 – Número de empresas por CNAE.
Fonte: Elaboração dos Autores, 2007.

4.3.4.1.2 Pela soma de salários

A região Sudeste assim como representa o maior número de empresas também representa a maior soma de salários. Porém nota-se que a região Norte, apesar de possuir menos empresas, é a que apresenta a segunda maior soma de salários. (Gráfico 26)

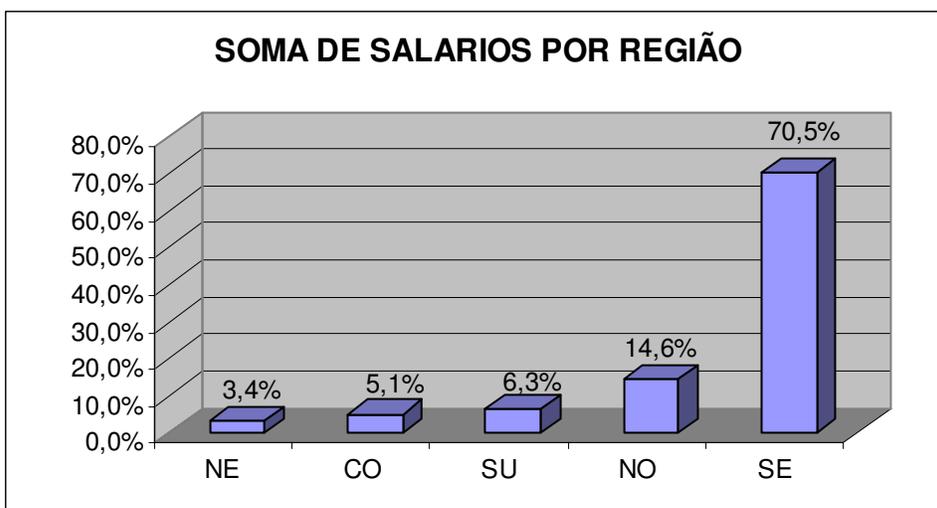


Gráfico 26 – Soma de salários por região.
Fonte: Elaboração dos Autores, 2007.

Por meio do gráfico 27 notamos que a região Norte possui um índice muito maior em relação às outras regiões do país. Isso pode significar que a região Norte tem salários mais altos ou que a região norte tem empresas que possuem mais funcionários. Apenas através do gráfico abaixo nada se pode afirmar que justifique esta relação ser muito maior do que o das outras regiões.

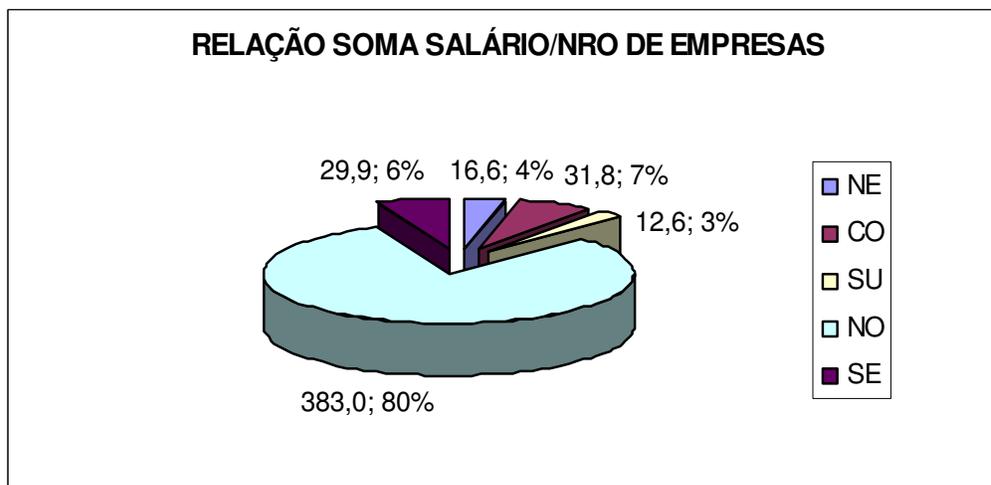


Gráfico 27 – Relação soma salário/número de empresas.
Fonte: Elaboração dos Autores, 2007.

Por meio do gráfico de relação salário por número de pessoas ocupadas e assalariadas, pode-se perceber que a região Norte é a região que possui a média de salários mais altos, seguido pela região Sudeste. A região Sul e região Nordeste são as que possuem a menor média de salários (Gráfico 28).

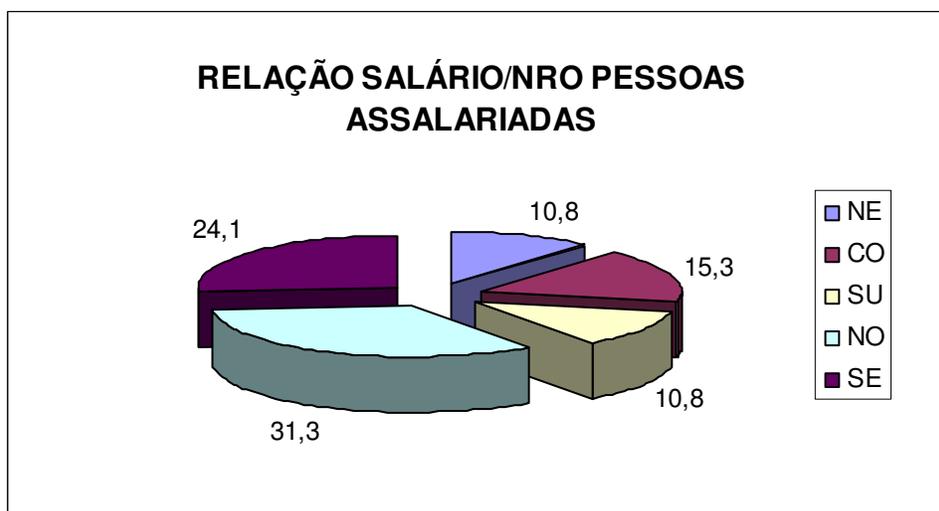


Gráfico 28 – Relação soma salário/número de pessoas assalariadas – região.
Fonte: Elaboração dos Autores, 2007.

Pelo gráfico abaixo (gráfico 29), nota-se que as atividades de Consultoria em Software são as que possuem a média de salários mais altos, seguido pelas atividades de Processamento de Dados e atividades de Banco de Dados e distribuição on-line de conteúdo eletrônico.

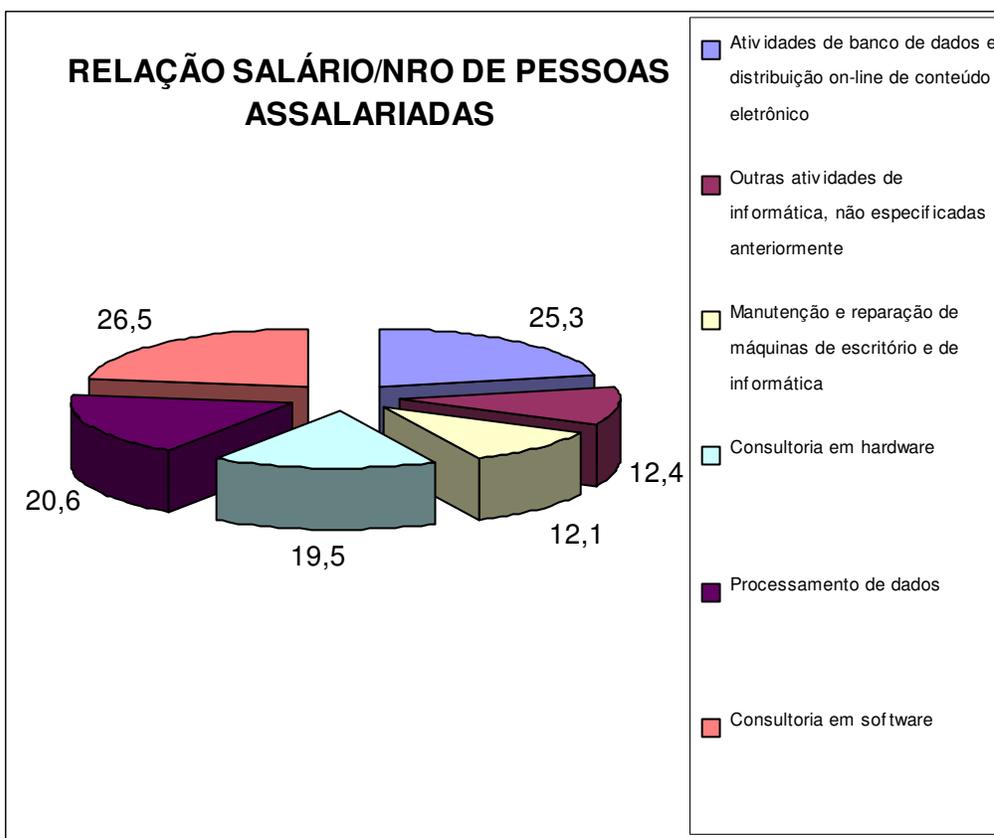


Gráfico 29 – Relação salário/número de pessoas assalariadas - CNAE.

Fonte: Elaboração dos Autores, 2007.

4.3.4.1.3 Pelo número de pessoas ocupadas

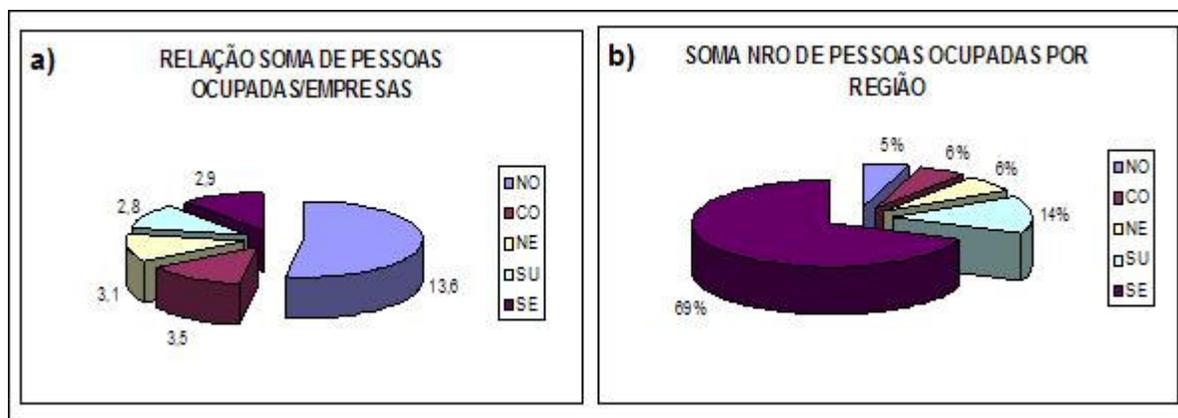


Gráfico 30 – Análises Empresas.

Fonte: Elaboração dos Autores, 2007.

Nota-se que a maioria das pessoas que trabalham da área de TI se encontram concentradas na região Sudeste e Sul do país, mas que é a região Norte que possui a maior média de pessoas por empresa, apresentando um número muito maior em relação aos outros estados. (Gráfico 30)

Por meio do gráfico abaixo (Gráfico 31) percebe-se que não existem muitos profissionais atuando da área de banco de dados e distribuição on-line de conteúdo. Também nota-se que as atividades de consultoria de software e processamento de dados são as que possuem maior porcentagem de profissionais.

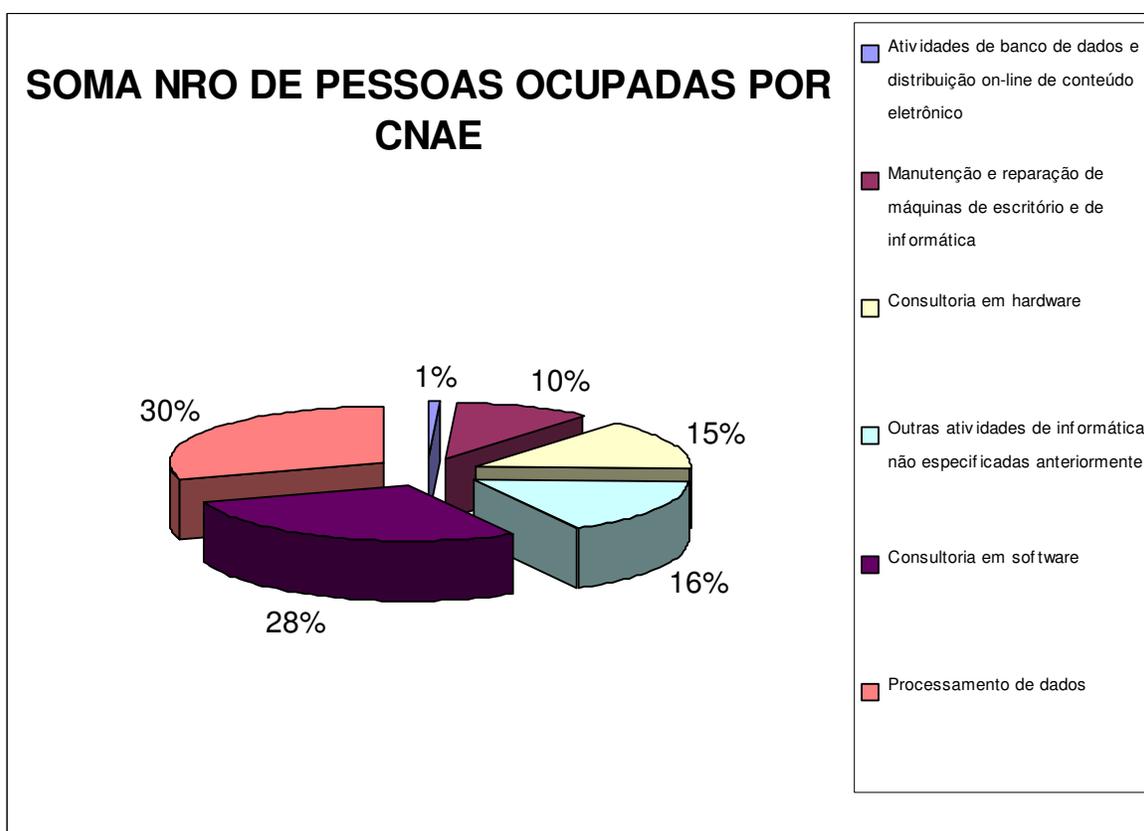


Gráfico 31 – Soma pessoas ocupadas por CNAE.
Fonte: Elaboração dos Autores, 2007.

4.3.4.1.4 Pelo número de pessoas salário

Nota-se que a maior parte das pessoas assalariadas são as que trabalham com atividades de processamento de dados, representando 40% do total. Em seguida estão as atividades de consultoria em software, representando 34% do total. (Gráfico 32)

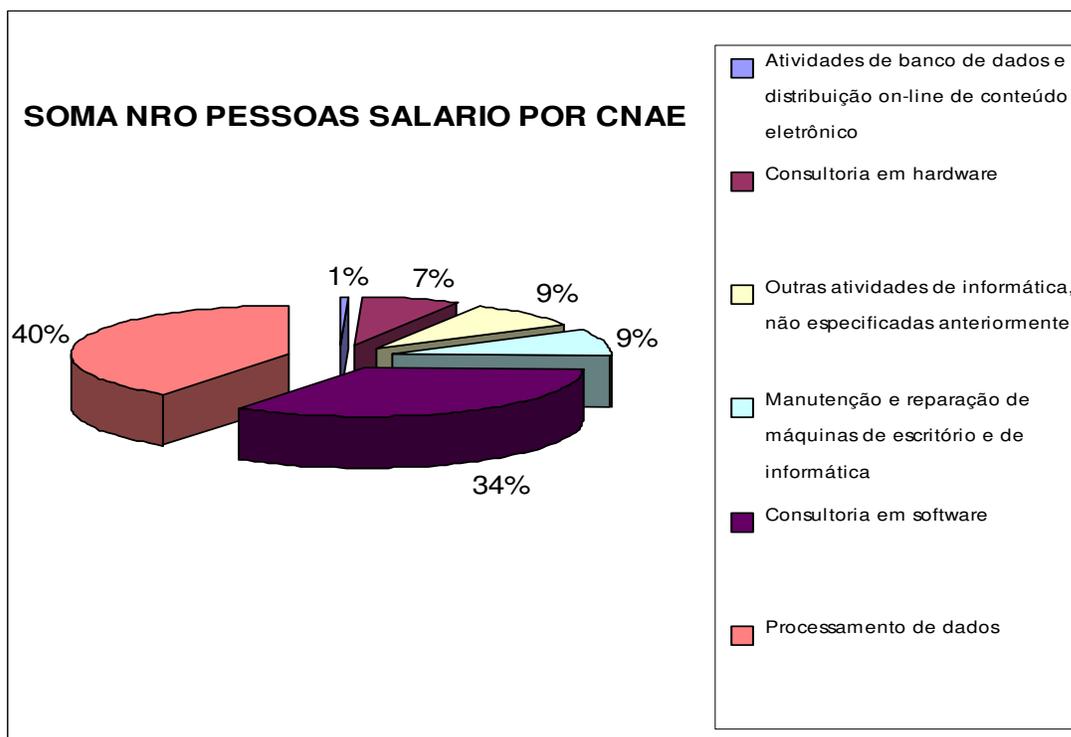


Gráfico 32 – Soma de pessoas salário por CNAE.
Fonte: Elaboração dos Autores, 2007.

No gráfico abaixo (gráfico 33) pode-se notar que a região Norte possui em média 12 funcionários assalariados por empresa, sendo esta média muito acima das médias apresentadas nas demais regiões.

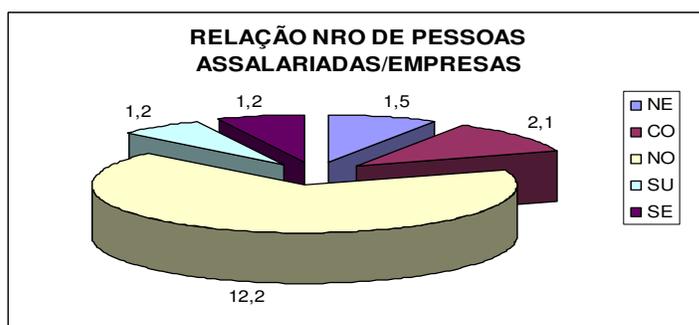


Gráfico 33 – Relação número de pessoas assalariadas/empresas.
Fonte: Elaboração dos Autores, 2007.

4.4 CONCLUSÃO

Por meio da Análise Exploratória de Dados, montou-se aqui, gráficos e histogramas que facilitam a visualização das informações obtidas do *DW*. Com essas informações, é possível que um analista consiga demarcar onde estão as IES melhores preparadas para fornecer um ensino de melhor qualidade, assim como o mercado que assimile com mais facilidade – rapidez e abrangência – e que pague melhor os recém graduados na área de TI. Porém, é notório as regiões que apresentam o maior número de doutores, mestres e graduados, assim como o maior número de inscritos, matriculados e concluintes.

Cabe ainda ressaltar aqui, que a intenção deste projeto não é formar opinião ou trazer respostas, mas apresentar informações que possam ser analisadas por especialistas, e ai sim, apontar os motivos e tendências destas informações.

5 CONCLUSÃO E TRABALHOS FUTUROS

5.1 CONCLUSÕES

Todos os indivíduos que pensam em ter uma carreira de sucesso, buscam formação em uma IES que seja capaz de lhe proporcionar uma base de conhecimento adequado para a área que se pretende atuar. Para isso, uma boa Instituição de Ensino necessita de docentes (professores) qualificados, com boa graduação, titulação e experiência. Com base nesse pensamento, buscou-se os dados para a construção do *Data Warehouse* a fim de extrair as informações necessárias para elaboração das análises exploratórias.

Algumas idéias projetadas sobre e para o tema, tanto na parte de análises quanto no momento de montagem do *Data Warehouse*, ficaram inviáveis e outras foram sendo melhoradas ao longo do desenvolvimento do projeto. Mas de uma maneira geral, o trabalho atingiu seu objetivo principal, uma vez que as análises e gráficos formulados e mostrados no capítulo 4 demonstram as informações que se desejava averiguar.

Depois de pronto o DW, e de feitas as análises, pode-se ver que certezas antes inquestionáveis, perderam seu valor, ou no mínimo, tornaram-se uma incógnita. Em Ciência da Computação, onde se encontram diversos cursos, e que foi a área de ensino aqui analisada, notou-se que existe pouco investimento em áreas que o mercado necessita há algum tempo, como Comércio Eletrônico e Segurança da Informação. Não se pode esquecer, que na área de TI, a velocidade dos acontecimentos e da evolução da tecnologia, tanto em hardware quanto em software, é muito alta. Cinco anos, é muito tempo para dar-se conta, que o mercado necessita de outros campos e de profissionais formados em cursos com focos diferentes dos oferecidos nas IES atualmente.

Foram analisados um total de 236.515 inscritos, 143.131 vagas, 75.199 ingressos, 27.903 concluintes, em 1263 cursos, dos 26 estados mais o Distrito Federal, que formam as 5 regiões. Isso mostra que o número de inscritos é maior que o de vagas, mas este é menor que o de ingressos. Também se pode ver, por exemplo, dados como os das IES estaduais e federais que possuem uma média de 26 docentes por curso, enquanto que nas IES privadas tem apenas 15

docentes por curso, assim como o fato da região nordeste ser a 3ª em números de cursos da área de ciências da computação, mas a 1ª em número de inscritos por curso.

O motivo destes resultados, não se pode nem se tentou explicar, cabe aos analistas esta tarefa, a intenção neste projeto, foi além de levantar os dados e apresentar as informações, mostrar que a AED pode ser útil antes de desenvolver um Sistema de Informação, pois visa descobrir que tipo de informações devem ser solicitadas para realizar as análises desejadas.

5.2 TRABALHOS FUTUROS

Uma das maiores dificuldades encontradas foi achar uma base de dados que pudesse servir para o estudo e posterior desenvolvimento do DW, de forma a atingir todos os objetivos propostos.

Depois de concluído o desenvolvimento do DW e começado as análises, notou-se que o objetivo inicial, que era a prospecção tecnológica, não poderia ser concluído porque a base de dados sobre as empresas de TI não era suficiente para isto. Desta forma, sugere-se a aplicação destas técnicas quando disponível uma base de dados mais robusta sobre empresas de setor de TI, capaz de oferecer condições para prospecção tecnológica.

Sugere-se também a aplicação deste mesmo trabalho para análises específicas estaduais e por IES.

Incrementar o DW com informações sobre os curso de pós-graduação em ciências da computação que podem ser obtidos do site da CAPES, informações sobre a produção científica e tecnológica dos docentes das IES, informações dos grupos de pesquisa do CNPq.

REFERÊNCIAS BIBLIOGRÁFICAS

ALCÂNTARA, I. **Business Intelligence: Elevando a gestão dos negócios a um novo patamar.**

Disponível em:

<<http://www.microsoft.com/brasil/technet/colunas/IgorAlcantara/ConceitosBI.msp>>. Acessado em: 02 abr. 2007.

ANTUNES, P. Cenários em Gestão do Ambiente. Universidade Nova de Lisboa. Disponível em:

<<http://ecomana.dcea.fct.unl.pt/disciplinas/gestao/files/cenarios.pdf>>. Acessado em: 02 maio 2007.

BAILEY, T.; GATRELL, A. **Interactive Spatial Data Analysis.** Longman Scientific and Technical, London, 1995.

BARBIERI, C. **BI Business Intelligence. Modelagem e Tecnologia.** Rio de Janeiro: Axcel Books, 2001.

BEZERRA, Eduardo. **Princípios de análise e projeto de sistemas com UML.** Rio de Janeiro: Campus, 2002.

BIGUS, Joseph P. **Data Mining with Neural Networks: solving business problems from application development to decision support.** New York: McGraw-Hill, 1996.

BIZZARRO, Fernanda Oviedo. **Engenharia de Software II – Artigos de UML.** Disponível em: <<http://www.inf.unisul.br/~oviedo>> Acessado em: 03 jun. 2007.

BOGHI, C.; SHITSUKA, R. **Sistemas de Informação - Um Enfoque Dinâmico.** São Paulo: Érica, 2002.

BONA, Cristina. **Avaliação de Processos de Software: Um estudo de caso em XP e ICONIX.** 2002. 122 f. Dissertação (Mestrado em Engenharia de Produção)- Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2002.

BOOCH, Grady; RUMBAUGH, James & JACOBSON, Ivar. **UML Guia do usuário.** Rio de Janeiro: Campus, 2000.

CARLSON, D. **Modelagem de Aplicações XML com UML: Aplicações Práticas de e-business**. São Paulo: Pearson Education, 2002.

CENTENARO, A. C. **Desenvolvimento e Implantação de um *Data Warehouse* corporativo com *Data Marts* distribuídos em uma cooperativa agroindustrial**. 2003. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2003.

CGEE - CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS. **Estudos Temáticos e de Futuro**. Disponível em: <<http://www.cgee.org.br/prospeccao/index.php?show=1>>. Acessado em: 30 abr. 2007.

CGEE - CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS. **Metodologia de Construção de Cenários**. Disponível em: <<http://www.cgee.org.br/arquivos/pro0103.pdf>>. Acessado em: 30 abr. 2007b.

CHRISPINO A. **Cenários em Educação Química: Instrumentos necessários. 2000**. Disponível em: <<http://www.fq.unam.mx/sitio/edquim/111/111-ani2.pdf>>. Acessado em: 02 maio 2007.

COMPETE.NET. **Sistema de Inteligência de Negócios**. Disponível em: <http://www.competenet.org.br/evento/ana_valeria.pdf> Acessado em 12 abr. 2007.

DATA MART. **Artigo**. Disponível em: <http://pt.wikipedia.org/wiki/Data_Mart>. Acessado em: 25 nov 2007.

DATAWAREHOUSE. **Artigos**. Disponível em: <<http://www.datawarehouse.inf.br/artigos/etl.asp>>. Acessado em: 23 abr. 2007

DEBONI, J. E. Z. **Breve Introdução aos Diagramas da UML**. Disponível em: <<http://www.voxxel.com.br/pages/introdiauml.html>>. Acessado em: 03 jun. 2007.

DESENVOLVIMENTO.GOV.BR 2. Disponível em: <<http://www.desenvolvimento.gov.br/arquivo/sti/proAcao/proTecnologica/plaAcao.pdf>>. Acessado em: 24 abr. 2007.

ECOMAN - *ECOLOGICAL ECONOMICS AND ENVIRONMENTAL MANAGEMENT CENTRE*. **Arquivos**. Disponível em: <<http://ecoman.dcea.fct.unl.pt/disciplinas/gestao/files/cenarios.pdf>>. Acessado em 26 maio 2007.

EDA. Artigo. Disponível em: <<http://www.itl.nist.gov/div898/handbook/eda/eda.htm>>. Acessado em: 28 set 2007.

FELÍCIO, F. L. **Desenvolvimento de um software gerenciador de Sistemas SAD**. 2005. Monografia (graduação em Sistemas de Informação) – UNOESTE, Presidente Prudente, 2005.

FLORÃO, M. R. **Processos de Desenvolvimento de Software Orientados a Objeto: Um Estudo do ICONIX**. 2004. 99 f. Trabalho de Conclusão do Curso de Especialização em Desenvolvimento de Software para Web apresentado a Universidade do Vale do Itajaí, como requisito para obtenção do título de Especialista em Desenvolvimento de Software para a Web, 2004.

GORDON, S, GRIGG, R, HORNE, M E THURMAN, S. **Inteligência de Negócio Orientada a Serviços**. Microsoft. Disponível em: <<http://www.microsoft.com/brasil/msdn/arquitetura/Journal/InteligenciaServicos.msp>> Acessado em: 24 abr. 2007.

HERRING, J. P. What is intelligence analysis? **Competitive Intelligence Magazine**. v.1, n.2, Jul./ Aug. p.13-16, 1997.

HERRING, J. P. **Producing CTI that Meets Management Needs and Expectations**. SCIP Competitive Technology Intelligence Symposium. Boston, 1997.

HOAGLIN, David C.; MOSTELLER, Frederick; TUKEY, John W. **Análise exploratória de dados: técnicas robustas: um guia**. Lisboa: John Wiley, 1992.

INMON, W. H. **Como Construir o Data Warehouse**. Rio de Janeiro: Campus, 1997.

INMON, W. H. **Como Construir o Data Warehouse**. Rio de Janeiro: Campus, 2000.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 4 ed. New Jersey, Prentice-Hall, inc., 1998.

JOHNSTON, R. Estudos Prospectivos. **Experiências Nacionais de Estudos Prospectivos: Reflexões da Austrália.** Elaborado para o Seminário Internacional sobre Estudos Prospectivos em Ciência e Tecnologia. Brasília. 27-28 setembro de 2000.

KIMBALL, R. **The Data Warehouse Toolkit: guia completo para modelagem dimensional.** Rio de Janeiro: Campus, 2002.

LARMAN, Craig. **Utilizando UML e padrões: uma introdução à análise e ao projeto orientados a objetos e ao processo unificado.** Porto Alegre: Bookman, 2002.

LAUDON, K. C.; LAUDON, J. P. **Sistemas de Informações: com Internet.** Rio de Janeiro: LTC, 1999.

LIAUTAUD, Bernard; HAMMOND, Mark. **e-Business intelligence: turning information into knowledge into profit.** New York: McGraw-Hill, 2000.

MACHADO, F. N. R. **Projeto de Data Warehouse: uma visão multidimensional.** São Paulo: Érica, 2000.

MAIA, José Anízio. Formado em Sistemas de Informação, certificado Java. Um dos fundadores do JugNanaus. **Construindo Software com Qualidade e Rapidez Usando ICONIX.** Disponível em: <<http://www.jugmanaus.com>>. Acessado em: 26 maio 2007.

MACORATTI, José Carlos. **Artigos.** Disponível em: <http://www.macoratti.net/uml_vb.htm>. Acessado em: 01 jun 2007.

MARTINS, Cláudio Roberto. **UML e Programação Orientada a Objetos.** Disponível em: <<http://www.cci.unama.br/clauidiomartins/cinbesa/UMLppt.pdf>>. Acessado em: 05 jun. 2007.

MELLO, J. A. B. **Uma proposta de modelo de dados para suporte ao processamento transacional e de apoio informacional simultaneamente.** 2002. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2002.

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA. **Notícias MCT**. Disponível em:
<<http://agenciact.mct.gov.br/index.php/content/view/31046.html>>. Acessado em: 02 mar. 2007a.

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA. **Notícias MCT**. Disponível em:
<http://agenciact.mct.gov.br/index.php?action=/content/view&cod_objeto=43366>. Acessado em:
03 mar. 2007b.

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA. **Clipping Online**. Disponível em:
<<http://agenciact.mct.gov.br/index.php/content/view/28542.html>>. Acessado em: 04 mar. 2007c.

MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO EXTERIOR.
Secretaria de Tecnologia Industrial: Programas e Ação. **Prospectiva Tecnológica**. Disponível em:
<<http://www.desenvolvimento.gov.br/sitio/sti/proAcao/proTecnologica>>. Acessado em: 24
abr. 2007a.

MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO EXTERIOR.
Secretaria de Tecnologia Industrial: Programas e Ação. **Programa Brasileiro de Prospectiva
Tecnológica**. Disponível em: <
<http://www.desenvolvimento.gov.br/arquivo/sti/proAcao/proTecnologica/plaAcao.pdf>>.
Acessado em: 23 abr. 2007b.

NETO, J. B. **Tecnologia da Informação para o gerenciamento do conhecimento obtido das
bases de dados de uma organização**. 2001. Dissertação (Mestrado em Engenharia de Produção)
– Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa
Catarina, Florianópolis, 2001.

PROSPECÇÃO TECNOLÓGICA EM ENERGIA. **Centro de Gestão e Estudos Estratégicos**.
Brasília: CGEE, 2005.

SELL, D. **Uma arquitetura para distribuição de componentes tecnológicos de sistemas de
informações baseados em Data Warehouse**. 2001. Dissertação (Mestrado em Engenharia de
Produção) – Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de
Santa Catarina, Florianópolis, 2001.

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

SOARES, V. J. A. **Modelagem Incremental no Ambiente de Data Warehouse**. 1998. Dissertação (Mestrado em Informática) – Instituto de Matemática – Núcleo de Computação Eletrônica – NCE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1998.

TCHOLAKIAN, A. B. **Sistemas de Apoio a Decisão**. Disponível em:
<http://www2.stela.ufsc.br/aran/sad/sad_aulas.htm>. Acessado em: 25 abr. 2007.

TYSON, K. W. M. **Competition on 21th Century**. St Lucie Press, EUA, 1997.

VIANA, Reinaldo. Mineração de dados: Introdução e Aplicações. **Revista SQL Magazine**, v. 10, p.16-25, São Paulo: 2004.

WIKIPÉDIA. **Business Intelligence**. Disponível em:
<http://pt.wikipedia.org/wiki/Business_Intelligence>. Acessado em 10 abr. 2007.

ZHAOFU, Xieli. **As aplicações de B.I. em vida pessoal**. Trabalho de Formatura Supervisionado - Monografia. Disponível em:
<<http://www.linux.ime.usp.br/~cef/mac499-06/monografias/rec/xieli/mono2.pdf>>. Acesso em: 25 abr. 2007.