

Universidade Federal de Santa Catarina
Programa de Pós-Graduação em
Engenharia de Produção (PPGEP)

Heleno Fülber

**PROPOSTA DE UM MODELO DE SISTEMAS DE INFORMAÇÃO
BASEADO EM *DATA WAREHOUSE* PARA ALAVANCAR A
GERÊNCIA DO ACOMPANHAMENTO DE ALUNOS
EM AMBIENTES DE *E-LEARNING***

Dissertação de Mestrado

Florianópolis
2004

HELENO FÜLBER

**PROPOSTA DE UM MODELO DE SISTEMAS DE INFORMAÇÃO
BASEADO EM *DATA WAREHOUSE* PARA ALAVANCAR A
GERÊNCIA DO ACOMPANHAMENTO DE ALUNOS
EM AMBIENTES DE *E-LEARNING***

Dissertação submetida ao curso de Pós-Graduação em Engenharia de Produção e Sistemas no Centro Tecnológico da Universidade Federal de Santa Catarina, como requisito parcial para obtenção do grau de Mestre em Engenharia de Produção e Sistemas.

Orientador: Prof. Dr. Alejandro Martins Rodrigues

Florianópolis
2004

Heleno Fülber

**PROPOSTA DE UM MODELO DE SISTEMAS DE INFORMAÇÃO
BASEADO EM *DATA WAREHOUSE* PARA ALAVANCAR A
GERÊNCIA DO ACOMPANHAMENTO DE ALUNOS
EM AMBIENTES DE *E-LEARNING***

Esta dissertação foi julgada e aprovada para a obtenção do título de **Mestre em Engenharia de Produção** no **Programa de Pós-Graduação em Engenharia de Produção** da Universidade Federal de Santa Catarina.

Florianópolis, 30 de junho de 2004.

Prof. Edson Pacheco Paladini, Dr.
Coordenador do Curso

BANCA EXAMINADORA

Prof. Alejandro Martins Rodrigues, Dr.
Universidade Federal de Santa Catarina
Orientador

Prof. Aran Bey Tcholakian Morales, Dr.
Universidade Federal de Santa Catarina

Profa. Elizabeth Sueli Specialsk, Dr.^a
Universidade Federal de Santa Catarina

AGRADECIMENTOS

Ao professor Alejandro Martins agradeço a orientação e a atenção dispensada na elaboração deste trabalho. Muito mais que um orientador, mostrou-se um amigo.

Agradeço também aos amigos Luis Henrique Bogo, Raphael Winckler de Bettio, Fábio Ferreira da Silva, Cloves das Neves, Jaime Prochnow e a todos os outros que sempre estiveram a meu lado quando deles precisei.

Em especial, agradeço a minha família, minhas irmãs e a meus pais, Osmar e Eliane Fülber, estejam sempre ao meu lado, apoiando-me e incentivando-me, durante todos esses anos.

RESUMO

FÜLBER, Heleno. **Proposta de um modelo de sistemas de informação baseado em *data warehouse* para alavancar a gerência do acompanhamento de alunos em ambientes de *e-learning***. 2004. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis.

Este trabalho apresenta um modelo de aplicação de Sistemas de Informação, baseado em *data warehouse* e *data mining*, para a alavancagem dos sistemas de gerência de aprendizagem – Learning Management System (LMS). Para tanto, disponibiliza informações estratégicas sobre a logística do processo de aprendizagem, que ajudam a compreender a logística do *e-learning*, o que possibilita constante melhora do desenho educacional. Assim, mostra-se como um sistema de apoio à decisão, para que os agentes envolvidos nos processos do *e-learning* possam balizar suas ações. Primeiramente, faz-se uma revisão dos conceitos de sistemas de gerência de aprendizagem, sistemas de informação, *data warehouse* e Mineração de Dados. Em seguida, apresenta-se então o modelo, que foi desenvolvido no Laboratório de Ensino a Distância (LED) da Universidade Federal de Santa Catarina (UFSC) e aplicado no curso de Formação em Gestão Pública no Brasil, da Caixa Econômica Federal.

Palavras-chave: *E-learning*, Sistemas de Gerência de Aprendizagem, Sistemas de Apoio à Decisão, *Data Warehouse*, Mineração de Dados.

ABSTRACT

This work presents an Information Systems application model, based on Data Warehouse and Data Mining, for leveraging the Learning Management System (LMS). Making available strategic information on the learning process logistics, which help to understand the e-learning logistics, allowing a constant improvement of the educational design. Presenting itself as a decision support system so the agents involved in the e-learning processes may outline their actions. A revision is done on the learning management systems, information systems, Data Warehouse and Data Mining concepts, then presenting the model, developed with the Distance Education Laboratory (LED) of the Federal University of Santa Catarina (UFSC) and applied to the Caixa Econômica Federal (Brazilian Bank) course, Public Management Formation in Brazil.

Key-words: E-learning, Learning Management System, Decision Support Systems, Data Warehouse, Data Mining.

LISTA DE FIGURAS

Figura 1: Componentes de um SI	33
Figura 2: Níveis de granularidade.....	37
Figura 3: Cubo com três dimensões.....	39
Figura 4: As duas personalidades do Webhouse	42
Figura 5: Diagrama de caso de uso do modelo.....	50
Figura 6: Diagrama de classes do data warehouse	52
Figura 7: Projeto da área de transição.....	53
Figura 8: Modelo entidade relacionamento da aplicação.....	58
Figura 9: Interação dos componentes de extração.....	60
Figura 10: Arquivo de relacionamento aluno tutor.....	61
Figura 11: Modelo do cubo de dados Participação <i>chat</i>	63
Figura 12: Modelo do cubo de dados Envio dúvida	63
Figura 13: Exemplo de utilização da ferramenta Analysis Service.....	64
Figura 14: Operação de drill down no cubo Participação chat.....	65
Figura 15: Visualização do cubo Participação chat nas dimensões tempo e tutor	66
Figura 16: Visualização do cubo Participação chat na dimensão endereço aluno	67
Figura 17: Visualização do cubo Envio dúvida nas dimensões tempo e aluno.....	68
Figura 18: Visualização do cubo Envio dúvida nas dimensões tipo dúvida e tempo.....	68
Figura 19: Filtro de visualização nas dimensões não selecionadas	69
Figura 20: Visualização do cubo Acesso sistema nas dimensões tempo e horário	70

LISTA DE QUADROS

Quadro 1: Tipos de sistemas de informação	34
Quadro 2: Casos de uso do modelo	50
Quadro 3: SQL de seleção dos fatos participação <i>chat</i>	61
Quadro 4: Comandos SQL de verificação, inserção e atualização dos fatos participação <i>chat</i>	62

SUMÁRIO

1 INTRODUÇÃO	11
1.1 JUSTIFICATIVA	14
1.2 OBJETIVO GERAL	14
1.3 OBJETIVOS ESPECÍFICOS	14
1.4 ORGANIZAÇÃO DO ESTUDO.....	15
2 SISTEMAS DE GERÊNCIA DE APRENDIZAGEM	16
2.1 CARACTERÍSTICAS DE UM LMS	16
2.2 LMS X LCMS	18
2.2.1 Objetos de Aprendizagem	18
2.3 PADRÕES E ESPECIFICAÇÕES PARA LMS	21
2.4 AVALIAÇÃO DE UM LMS	23
2.4.1 Exemplo de Avaliação de Ferramentas LMS.....	25
2.4.2 Critérios de Avaliação Edutech.....	27
2.4.3 Conclusões da Avaliação	28
2.4.4 Considerações sobre a Avaliação	29
2.5 LICENCIAMENTO DE UM LMS.....	29
2.6 CONSIDERAÇÕES FINAIS	30
3 DATA WAREHOUSE	32
3.1 SISTEMAS DE INFORMAÇÕES (SI)	32
3.2 DATA WAREHOUSE (DW).....	35
3.2.1 Características de um <i>Data Warehouse</i>	36
3.2.2 Granularidade e Particionamento	37
3.2.3 Processo Analítico <i>On-line</i> (OLAP).....	38
3.2.4 Cubo de Decisão.....	39
3.2.5 Projeto de um <i>Data Warehouse</i> Dimensional.....	40

3.2.6 <i>Data Webhouse</i>	41
3.3 <i>DATA MINING</i> (DM)	43
3.3.1 <i>K-Means</i>	45
3.4 CONCLUSÃO.....	45
4 SOLUÇÃO PROPOSTA.....	47
4.1 MODELO DO SISTEMA.....	47
4.1.1 Identificação dos Processos.....	47
4.1.2 Definição dos Fatos a Serem Analisados.....	48
4.1.3 Dimensões a Serem Analisadas e Granularidade	48
4.2 PROJETO.....	49
4.3 ÁREA DE TRANSIÇÃO.....	53
4.4 CONSIDERAÇÕES SOBRE O MODELO.....	54
4.5 RESULTADOS ESPERADOS	54
5 APLICAÇÃO: SISTEMA DE APOIO À DECISÃO PARA LMS.....	57
5.1 MODELO APLICADO.....	57
5.2 EXTRAÇÃO DE DADOS PARA O AMBIENTE	59
5.3 ÁREA DE APRESENTAÇÃO.....	62
6 CONCLUSÕES E SUGESTÕES.....	71
6.1 CONCLUSÕES	71
6.2 SUGESTÕES PARA EXTENSÕES	72
6.3 CONSIDERAÇÕES FINAIS	72
REFERÊNCIAS	73

1 INTRODUÇÃO

O mundo contemporâneo tornou-se globalizado e altamente tecnificado, o que fez surgir uma competição quase ‘selvagem’ entre as nações, a qual demanda cada vez mais pessoas qualificadas, tanto culturalmente como em escolaridade formal. As habilidades exigidas são cada vez mais sofisticadas, pois a evolução do modo de produção despreza o trabalho desqualificado e repetitivo, que é substituído eficientemente pela automatização e robotização. Devido a essas exigências, a educação é mais do que nunca um dos pilares para a construção de qualquer sociedade que pretenda ser desenvolvida (FRANCO, 1998).

Nesse horizonte, mais do que uma alternativa ao ensino que obriga professores e alunos a encontrar-se no mesmo espaço e no mesmo tempo, o ensino a distância (EAD) apresenta-se hoje como resposta às necessidades emergentes de uma sociedade caracterizada por elevados níveis de competitividade, em que o ‘tempo’ é fator crítico no desenvolvimento dos indivíduos e das instituições. Comprova isso o fato de que o acesso ao conhecimento deve ser possível em qualquer momento, em qualquer lugar e, sobretudo, quando for considerado necessário e oportuno.

Boff (2000) define EAD como uma forma de educação, na qual alunos e professores se encontram separados fisicamente, sendo o processo de interação multidirecional, apoiado por tecnologia de comunicação, em que o aluno é protagonista de seu aprendizado e o professor, um facilitador dele. É um sistema baseado no uso seletivo de meios instrucionais, tanto tradicionais quanto inovadores, que promovem o processo de auto-aprendizagem, para se atingir objetivos educacionais específicos, com potencial de maior cobertura geográfica do que a dos sistemas educativos tradicionais (LANDIN, 1997). De forma sucinta, pode-se afirmar que EAD é o ensino que ocorre quando o ensinante e o aprendente (aquele a quem se ensina) estão separados (no tempo e no espaço).

De acordo com Dalmau *et al.* (2002), devido às características de flexibilidade de espaço e tempo, o EAD pode se adaptar a diversas demandas. Com metodologia voltada para atender a adultos com compromissos pessoais e profissionais, o EAD permite ao aluno a continuação dos estudos, sem o abandono de outras atividades. Seu grande diferencial é proporcionar ao aluno a opção de escolher o próprio local e horário de estudo. A possibilidade de gerar produtos customizados, ajustados e adaptados às necessidades dos clientes permite ganhos em tempo e adequação no atendimento a demandas específicas, que não estejam contempladas a contento em estruturas educacionais tradicionais.

Outra das vantagens do EAD em relação à educação presencial está na economia. Ao se deslocar um grande grupo de funcionários para participar de aulas presenciais em outras localidades, tem-se, na substituição desses profissionais, bem como no pagamento dos seus salários, sem que eles estejam efetivamente trabalhando, considerável aumento de custos fixos (ROSSETT, 2001).

O *e-learning* – modalidade de ensino-aprendizagem que combina a filosofia de educação a distância com as mais avançadas tecnologias de comunicação, em que o treinamento acontece através de uma rede, usualmente Internet ou em uma intranet de uma instituição –, segundo Costa e Peralta (2001), desponta, rapidamente como a principal opção de ensino-aprendizagem, para grande porcentagem das pessoas de todo o mundo. O desenvolvimento das novas tecnologias tem promovido, no panorama mundial, um significativo incremento do ensino a distância, quer em termos do número de alunos envolvidos, quer em termos do número de universidades que incluíram essa modalidade de ensino na sua oferta curricular (COSTA e PERALTA, 2001).

Dentre as vantagens da utilização de plataformas de *e-learning* como instrumento de formação para pessoas, destacam-se (WEBSCHOOL, 2003):

- Diversificação e ampliação da oferta de cursos;
- Oportunidade de atualização adaptada às exigências atuais, a pessoas que não puderem freqüentar o curso tradicional;
- Eficaz combinação de estudo e trabalho;
- Formação fora do contexto da sala de aula;
- Respeito ao ritmo de aprender do aluno, que é o centro do processo de aprendizagem e sujeito ativo de sua formação;
- Comunicação bidirecional freqüente, garantindo aprendizagem dinâmica e inovadora;
- Redução das despesas pessoais para treinamentos presenciais, como transporte, hospedagem etc.

É importante compreender que *e-learning* não é apenas o treinamento a distância, pois, isso, os velhos cursos por correspondência já faziam. Nem é, tampouco, simplesmente o ensino *on-line*, algo já disseminado em várias universidades e empresas norte-americanas, canadenses, européias e até sul-americanas. Então, *e-learning* não é *e-training*, é uma nova

pedagogia ou logística, em que o treinamento e o aprendizado têm seu foco na tela do computador e no novo papel assumido pelos professores.

Abaixo, citam-se alguns números que, segundo o Instituto Gartner (2003), traduzem o crescimento atual do *e-learning*:

- O mercado global para *e-learning* crescerá de \$2,1 bilhões em 2002 para \$33,6 bilhões em 2005;
- Perto de 4.000 universidades americanas agora oferecem cursos via Internet ou usam a Web para auxiliar nos cursos presenciais;
- Cerca de dois milhões de estudantes fazem cursos *on-line* nas instituições de ensino superior americanas, e esse número subirá para cinco milhões em 2006;
- Em 2005, 70% das grandes corporações terão seus próprios sistemas de *e-learning*.

Sendo o *e-learning* uma realidade consolidada e irreversível, o desafio deixa de ser implantar tecnologias que permitam a realização do *e-learning*, passando a ser a gerência do aprendizado e de seu conteúdo. Um sistema de gerência de aprendizagem (Learning Management System – LMS) é uma solução estratégica, responsável pelo planejamento, pela distribuição e pelo controle dos eventos de aprendizagem, e o LCMS (Learning Content Management System) é a parte dele responsável pelo conteúdo.

Com a rápida evolução e as mudanças tecnológicas, tornou-se fundamental que as pessoas tomem suas decisões com maior rapidez, sendo assim, elas necessitam ter à mão informações atuais e precisas. No entanto, as informações precisam ser correlacionadas, de modo que as pessoas possam tomar decisões mais facilmente e trabalhar com cenários futuros. Nesse sentido, a utilização de Sistemas de Informação baseados em *data warehouse* pode facilitar a ação das pessoas no processo decisório, no que tange à obtenção de dados e de conteúdos relevantes para as organizações das mais diversas áreas.

Assim, com a realização deste estudo, que leva em conta a identificação e disponibilização de informações estratégicas em tempo real, pretende-se possibilitar a melhoria constante do modelo pedagógico. Acredita-se que isso pode ser obtido, tornando-se mais efetivo o acompanhamento dos acadêmicos e dos conteúdos ministrados, identificando-se possíveis falhas no modelo de aprendizagem e permitindo uma administração estratégica do estudo acadêmico.

1.1 JUSTIFICATIVA

A utilização de ambientes Web para educação, tanto em cenários acadêmicos como corporativos, tem sido constante nos últimos anos, e tende a aumentar cada vez mais, devido à redução de custo que tais ambientes proporcionam às corporações e à possibilidade de independência geográfica proporcionada. Entretanto, um dos grandes desafios para essas plataformas tem sido avaliar a logística do *e-learning*, entender a correlação entre agentes, como professores, tutores, monitores e alunos, seus perfis pessoais e de utilização das ferramentas de interatividade com o conteúdo e sua compreensão.

No Laboratório de Ensino a Distância da UFSC, desenvolveu-se um ambiente virtual de aprendizagem denominado VIAS-K e os desafios acima relacionados também foram constatados em relação a ele. Uma solução que pode amenizar a falta de parâmetros para o entendimento da logística do *e-learning* é a abordagem de Sistemas de Informação baseada em *data warehouse*, cujo objetivo é correlacionar informações para que um especialista, nesse caso, um tutor ou professor, possa realizar análises desestruturadas e adquirir o conhecimento que está implícito.

1.2 OBJETIVO GERAL

O objetivo geral desta pesquisa é propor um modelo de aplicação de Sistemas de Informação, baseado em *data warehouse* e *data mining*, que permita a alavancagem dos sistemas de gerência de aprendizagem (Learning Management System – LMS), com a disponibilização de informações estratégicas, em tempo real, sobre a logística do processo de aprendizagem. Dessa forma, quer-se possibilitar constante melhora no desenho educacional do *e-learning*, com a reestruturação do projeto pedagógico e seus conteúdos.

1.3 OBJETIVOS ESPECÍFICOS

A pesquisa tem quatro objetivos específicos:

- Adquirir conhecimento implícito, a partir da correlação e análise desestruturada dos dados de utilização de sistemas de *e-learning*, objetivando criar uma nova forma de avaliação de cursos;
- Incorporar outros atores na análise do *e-learning*, permitindo uma análise mais abrangente, uma vez que o conhecimento gerado pela análise dos dados depende do conhecimento e da visão de quem os recebe;

- Flexibilizar a tomada de decisão, disponibilizando as informações em tempo real, de maneira mais dinâmica, e não somente ao final dos cursos, por meio dos tradicionais formulários de avaliação;
- Gerar grupos, ou seja, classificação dos alunos em 'n' classes, baseando-se em variáveis do interesse de quem estiver analisando, a fim de possibilitar ações como colocar um grupo de alunos que não interagem com um tutor que estimula a interação e vice-versa.

1.4 ORGANIZAÇÃO DO ESTUDO

Este trabalho está disposto em seis capítulos. O primeiro capítulo traz a introdução, justificativa e os objetivos da pesquisa. O segundo define Sistema de Gerência de Aprendizagem (Learning Management System – LMS), apresentando suas características e diferenciando LMS de LCMS, bem como objetos de conhecimento e os principais padrões e especificações para LMS. O terceiro capítulo descreve sistemas de informação, suas características e seus tipos, contextualiza o *data warehouse*; em seguida, enfoca o *data warehouse*, abordando suas características, fases de desenvolvimento, a tecnologia OLAP, granularidade e cubo de decisão; por fim, aborda *Data Mining* e o algoritmo de classificação *k-means*. O quarto capítulo apresenta a solução proposta, ou seja, o modelo de aplicação de sistemas de informações desenvolvido, seu projeto, definição da área de transição, considerações e resultados esperados. O quinto capítulo mostra a aplicação do modelo em um caso prático, no curso de Formação em Gestão Pública no Brasil, realizado pela Caixa Econômica Federal, por meio do Laboratório de Ensino a Distância da Universidade Federal de Santa Catarina. O último capítulo traz as conclusões, sugestões para novos trabalhos e considerações finais da pesquisa.

2 SISTEMAS DE GERÊNCIA DE APRENDIZAGEM

Um sistema de gerência de aprendizagem (Learning Management System – LMS) é o *software* que automatiza a administração de eventos do treinamento e simplifica a administração dos programas de treinamento e aprendizagem em uma organização. Também, ajuda a planejar o progresso de aprendizagem dos estudantes e a comunicação e colaboração deles com seus colegas. Para administradores, ajuda no planejamento, na análise e no acompanhamento do processo de aprendizagem. As potencialidades da base de dados do LMS estendem-se a funções adicionais, tais como gerência da companhia, avaliações *on-line*, personalização e outros recursos (BRENNAN, 2003).

De acordo com Dean (2002), a definição de que LMS é um ‘pacote’ de *software* que suporta o gerenciamento do aprendizado em uma organização é muito geral. Segundo o autor, os LMSs diferem dos *softwares* que controlam recursos em centros de aprendizagem, e podem ser divididos, de acordo com suas características, em três tipos: sistemas que suportam a aprendizagem baseada em sala de aula; sistemas que suportam a aprendizagem a distância *off-line*, como o treinamento baseado cópia, vídeo etc.; e sistemas que suportam o treinamento *on-line*.

2.1 CARACTERÍSTICAS DE UM LMS

Há duas características que todo LMS deve ter, ou seja, o controle de estudantes e informações sobre o curso. As facilidades típicas para suportá-las são:

- Registro de estudantes: Inclui registro de detalhes pessoais e a geração de *login* e senha;
- Registro de cursos: Registros como pré-requisitos, qualificações obtidas na conclusão e no tempo do estudo;
- Acompanhar o progresso dos estudantes: Isso tipicamente envolverá gravar os módulos terminados, as sessões e os cursos, resultados de testes e se o estudante passou ou não;
- Controle de custos e pagamentos: Serão necessários em muitas circunstâncias;
- Relatórios: É essencial ter bons relatórios, fornecidos aos usuários frequentemente e com muita flexibilidade na abstração de dados e nas maneiras de apresentá-los.

Controlar o treinamento em sala de aula requer considerar fatores adicionais referentes à alocação e aos custos dos recursos necessários, o que inclui:

- Agendamento das salas;
- Alocação dos instrutores;
- Registro do equipamento de suporte: vídeo, projetores, PCs etc.;
- Agendamento dos estudantes;
- Relatórios, que acontecem anteriormente ao curso e incluem agendamentos, instruções, listas de espera, realocações etc.;
- Utilização de *e-mail*, facilidade que pode ser fornecida para comunicação entre instrutores e alunos.

A principal característica da aprendizagem *off-line* é a possibilidade de cada módulo dos cursos ser iniciado pelo estudante a qualquer momento, que dependerá da disponibilidade individual de tempo. Nesse tipo de aprendizagem, a ajuda do tutor deve ter suporte pelo LMS.

As características que um LMS *off-line* deve ter incluem:

- Distribuição do material; e
- Alocação dos tutores.

A área de sistemas que suportam o treinamento *on-line* está em crescimento, devido à crescente disponibilidade de treinamentos via Internet ou intranets. LMSs que suportam a aprendizagem *on-line* tendem a ser sofisticados e bastante complexos. Dentre as características adicionais que eles podem ter, estão:

- Distribuição *on-line*: Algumas instituições oferecem somente seus próprios cursos, mas poderiam disponibilizar cursos e material padronizado de outras instituições;
- Mecanismos de cobrança de usuários e métodos de pagamento de fornecedores;
- Conferência *on-line* em grupo;
- Marcações, para que os estudantes possam reiniciar um módulo em que pararam na última vez que o acessaram;
- Tutoriais de apoio *on-line*;
- *Download* do material de apoio, para estudos *off-line*.

Segundo Dean (2002), algumas características que principalmente permitem a ajuda aos gerentes de recursos humanos e a empregados, com seus planos de treinamento, podem ser apresentadas pelos diferentes tipos de LMSs. Elas incluem:

- Recomendar cursos baseados no perfil do estudante e em treinamentos já encerrados;
- Planejar o treinamento para que o estudante atinja habilidades desejadas;
- Certificação;
- Identificação de tendências e implicações delas nos recursos.

2.2 LMS X LCMS

De acordo com Greenberg (2002), é importante definir o LCMS (Learning Content Management System), pois alguns fornecedores estão posicionando os LCMSs como a nova onda de LMSs e, na verdade, LMS e LCMS são sistemas complementares, mas diferentes.

Enquanto um LMS é uma solução estratégica para o planejamento, a distribuição e o controle de todos os eventos de aprendizagem de uma organização que tem como foco gerenciar os alunos, mantê-los a par de seus progressos e desempenho, em contraste, o foco de um LCMS é o conteúdo da aprendizagem, pois busca permitir a autores e a *designers* instrucionais meios mais eficientes de elaborar o conteúdo de *e-learning*.

IDC define LCMS como um sistema que cria, armazena, reúne e entrega o conteúdo personalizado para o *e-learning* no formato de objetos de aprendizagem. Embora um LMS controle e administre todas as formas de aprendizagem em uma organização, o foco de um LCMS concentra-se no aprendizado *on-line*, geralmente, nos objetos de aprendizagem (GREENBERG, 2002).

2.2.1 Objetos de Aprendizagem

O Institute of Electrical and Electronic Engineers (IEEE, 2003) define os objetos de aprendizagem (OA) como “Qualquer entidade, digital ou não digital, que possa ser usada, reutilizada ou referenciada durante o uso de tecnologias que suportem ensino”. Já o padrão Instructional Management Systems (IMS, 2003) define que “Qualquer espécie de entidade digital a qual tenha a capacidade de exprimir algum conhecimento pode ser considerada um objeto de aprendizado”.

Segundo Singh (2001), um objeto de aprendizado deve ser bem estruturado e dividido em três partes bem definidas:

- **Objetivos:** Essa parte do objeto tem como intenção demonstrar ao aprendiz o que ele poderá aprender, a partir do estudo desse objeto. Também, poderá conter uma lista dos conhecimentos prévios necessários para um bom aproveitamento de todo o conteúdo disponível. Ainda, pode ser comparado, de forma grosseira, à ementa de uma disciplina;
- **Conteúdo instrucional:** Aqui deverá ser apresentado todo o material didático necessário, para que, no término, o aluno possa atingir os objetivos citados no item anterior, utilizando qualquer tipo de mídia;
- **Prática e *feedback*:** Uma das características importantes do paradigma Objetos de Aprendizado é que, a cada final de aprendizado, julga-se necessário que o aprendiz verifique se o seu desempenho atingiu as expectativas. Em caso negativo, o aprendiz deve ter liberdade para voltar a utilizar o objeto quantas vezes achar necessário.

Além de definir esses três aspectos que os objetos de aprendizado devem ter, a IMS também definiu um padrão de armazenagem de informações necessárias para indexação dos objetos de aprendizado, de modo que a característica 'indexação e procura' possa ser cumprida. A maneira encontrada pelos pesquisadores envolvidos no consórcio IMS foi a criação de metadados, os quais podem ser considerados dados sobre dados, ou seja, informação sobre informação.

De acordo com Longmire (2001), os objetos de aprendizado têm características para procurar resolver diversos problemas existentes atualmente, quanto ao armazenamento e à distribuição de informação, por meios digitais. As características enfocadas por Longmire são:

- **Flexibilidade:** Como os objetos de aprendizado são construídos de maneira que tenham início, meio e fim, eles já nascem flexíveis, podendo ser reutilizados sem necessitar de nenhum tipo de manutenção. Essa capacidade de reutilização evidencia as vantagens desse novo paradigma. A criação de novos cursos, utilizando conhecimentos já armazenados e consolidados, não é mais utopia, pode ser considerada um objetivo que está próximo de ser alcançado;
- **Facilidade para atualização:** Como os mesmos objetos são utilizados em diversos momentos, a atualização deles em tempo real é relativamente simples. Desde que

todos os dados relativos a esse objeto estejam em um mesmo banco de informações, não há necessidade de atualizar esse conhecimento em todos os ambientes que o utilizam. Dessa forma, a pessoa que apenas utilizou o conhecimento de um autor poderá contar com correções e aperfeiçoamentos, sem ter de se preocupar em fazer isso;

- Customização: A mesma característica que proporciona ao objeto flexibilidade, também proporciona customização jamais encontrada em outro paradigma educacional. Como os objetos são independentes, utilizá-los em um curso, uma especialização ou qualquer outro tipo de qualificação torna-se real, sendo que cada entidade educacional pode utilizar os objetos e arranjá-los da maneira que mais lhe convier. Também, os indivíduos que necessitarem poderão montar seu próprio aprendizado;
- Conteúdos programáticos avançando: Assim, cria-se um novo paradigma, o *on-demand learning*;
- Interoperabilidade: A criação de um padrão para armazenamento de objetos de aprendizado traz mais uma vantagem, a interoperabilidade, ou seja, reutilização dos objetos não apenas em nível de plataforma de ensino, mas também em nível mundial. O fato de um OA ser criado e poder ser utilizado em qualquer plataforma de ensino do mundo aumenta ainda mais as vantagens desses objetos. Assim que a barreira lingüística for quebrada, a interoperabilidade entre bancos de objetos de todo o mundo será selada, trazendo vantagens jamais vistas na educação;
- Aumento do valor do conhecimento: A partir do momento em que um objeto é reutilizado várias vezes e em diversas especializações, esse objeto, ao longo do tempo, melhora, assim, sua consolidação cresce de maneira espontânea. A melhora significativa da qualidade do ensino é mais uma vantagem que pode ser considerada, ao se pensar em objetos de aprendizado;
- Indexação e procura: A padronização dos objetos também facilita a procura por um objeto necessário. Quando um conteudista necessitar de determinados objetos para completar seu conteúdo programático, a padronização deles e a utilização de assinaturas digitais tenderão a gerar facilidade de encontrar objetos com as mesmas características, em qualquer banco de objetos que esteja disponível para eventuais consultas.

Conforme Greenberg (2002), um LCMS armazena objetos de aprendizagem em um repositório central, para que os *designers* instrucionais recuperem e montem cursos personalizados. Isso beneficia desenvolvedores e estudantes, porque os cursos tradicionais tendem a ter mais conteúdo do que um único estudante pode absorver ou necessita absorver sobre um tópico, dividindo o conteúdo do curso em objetos de aprendizagem e disponibilizando-os em tempo real, conforme as necessidades e capacidade de absorção. O resultado final é o aumento de produtividade, porque os estudantes não desperdiçam tempo com matéria irrelevante.

Um LMS pode ter impacto direto no trabalho de milhares de alunos e controlar todos os aspectos da aprendizagem organizacional. Especialistas no assunto recomendam utilizar um LMS que permita a integração com um LCMS (GREENBERG, 2002).

2.3 PADRÕES E ESPECIFICAÇÕES PARA LMS

Existem muitos padrões e especificações sendo criados para controlar aspectos de treinamentos multimídia e gerência computadorizada da aprendizagem. Nesta pesquisa, concentra-se em alguns que, de acordo com Dean (2002), são os mais importantes e proeminentes hoje. É relevante diferenciar padrões e especificações. Há os padrões que são produzidos e oficialmente reconhecidos por corporações, como o IEEE e ISO. Há também outras corporações que desenvolvem especificações, que são padrões que fornecem parâmetros de desenvolvimento para desenvolvedores trabalharem. IMS, AICC e ADLNet são as corporações que produzem especificações. A maioria dos comitês de desenvolvimento comunicam-se e há participação extensiva da indústria nessas corporações.

O comitê de padrões de tecnologias de aprendizagem do Institute of Electrical and Electronic Engineers (IEEE) é a principal corporação educacional que pode estabelecer padrões reconhecidos para a gerência de aprendizagem nos EUA. A maioria dos grupos mencionados abaixo trabalha com o IEEE no desenvolvimento de padrões que geralmente são aceitos. O IEEE LTSC tem 20 grupos cobrindo todas as áreas de tecnologia de aprendizagem. Nenhum padrão formal na gerência de aprendizagem foi emitido ainda pelo comitê.

International Standards Organisation¹ (ISO) tem uma subcomissão SC36 trabalhando na padronização na área de tecnologias de aprendizagem, em conjunto com o IEEE LTSC.

¹ <http://www.iso.org/>

O Aviation Industry CBT Committee² (AICC) projetou padrões para que a indústria da aviação permitisse que o treinamento interativo fosse entregue por meio das diferentes plataformas de *software* e *hardware*. Seu trabalho inclui um modelo para a gerência de instrução de computador (Computer Managed Instruction – CMI) que atualmente está sendo estendido à entrega de treinamento via Web.

O IMS Global Consortium, Inc³ (IMS) é outra corporação influente, situada nos EUA, mas com centros colaboradores no Reino Unido, na Austrália, no Canadá e em Singapura. Trabalha com instituições educacionais e fornecedores do *software* para definir especificações. Produziu especificações para objetos de aprendizagem (metadados de objetos de aprendizagem), armazenamento de conteúdo, perguntas e testes de interoperabilidade, informação sobre o estudante. Detalhes sobre isso podem ser encontrados na seção de especificações do *website* do IMS. Outras especificações estão em estágios diferentes do desenvolvimento. Os grupos do IMS trabalham próximos a grupos relevantes do IEEE.

European Community Information Society Standardization System for Learning Technology⁴ (CEN/ISSS LT) é um grupo europeu que está examinando padrões emergentes, avaliando-os, para ver como se encontram em relação às exigências multilingual e multicultural da comunidade européia. Somente onde o grupo considera essencial, ele cria novos padrões.

A Microsoft desenvolveu um recurso para intercâmbio de aprendizagem (Learning Resources Interchange⁵ – LRN), que é uma execução da especificação de armazenamento do IMS.

A Advanced Distributed Learning Network⁶ (ADLNet) foi criada pelo governo dos EUA, principalmente, para suportar aprendizagem distribuída para as forças armadas e outras agências de governo. Produziu a especificação SCORM (*shareable content object reference model*).

Segundo Hall (2003), o IEEE é o responsável pela padronização geral dos padrões de objetos de aprendizagem. Esses padrões são baseados nas especificações SCORM de modelo de referência para compartilhamento do conteúdo de objetos, e são influenciados por grupos da indústria, tais como o IMS e o AICC. Em termos simples, esses grupos de padrões da indústria estão criando especificações técnicas para permitir e suportar um modelo

² <http://www.aicc.org/>

³ <http://www.imsproject.org/>

⁴ <http://www.cenorm.be/iss/worksop>

⁵ <http://www.microsoft.com/elearn/support.asp>

⁶ <http://www.adlnet.org/>

padronizado para a aprendizagem baseada na Web. Quando se montam conteúdos de múltiplas fontes e diferentes redes, a grande tarefa é fazer com que os *softwares* e as ferramentas se comuniquem, ou seja, fazer dados de importação e exportação de diferentes LMSs responderem aos mesmos comandos operacionais.

2.4 AVALIAÇÃO DE UM LMS

A seleção de um sistema de gerência de aprendizagem é freqüentemente uma das decisões o mais críticas que serão feitas. Selecionar um LMS que se adapte às exigências, assim como às necessidades futuras, pode ser um processo difícil (BRENNAN, 2003).

Para Hall (2003), um sistema de gerência de aprendizagem deve:

- Consolidar iniciativas de treinamento em plataformas escaláveis de baixo custo, baseadas na Web;
- Disponibilizar o conteúdo rapidamente e em múltiplas línguas;
- Medir a eficácia das iniciativas de treinamento;
- Mesclar o aprendizado de sala de aula com o *on-line*;
- Integrar-se a outras soluções de aplicações da organização;
- Centralizar e automatizar a administração;
- Utilizar serviços do tipo *self-service* sempre que possível;
- Suportar portabilidade e os padrões AICC, IMS e SCORM;
- Personalizar o conteúdo e permitir reuso do conhecimento.

De acordo com Greenberg (2002), as seguintes capacidades devem ser procuradas em um LMS:

- Suporte para aprendizagem combinada: Pessoas aprendem de maneiras diferentes e um LMS deve oferecer um currículo que misture a sala de aula e cursos virtuais facilmente, o que, combinado às características de LCMS, permite treinamento personalizado;
- Integração a sistema de recursos humanos: Quando os sistemas são integrados, um empregado pode entrar com uma nova requisição no sistema de recursos humanos e automaticamente se matricular para o treinamento preparado para sua função na empresa;
- Ferramentas de administração: O LMS deve permitir aos administradores controlar os registros dos usuários e os perfis, definir papéis, ajustar currículos, representar

os caminhos para certificação, atribuir tutores, autores de cursos, gerenciar conteúdo, administrar orçamentos internos, pagamentos de usuários e cobranças. Os administradores necessitam de acesso completo à base de dados do treinamento, de modo a criar padrões e customizar relatórios de desempenho individuais ou em grupos. Nesse sentido, o sistema deve também permitir a construção de programações para alunos, instrutores e salas de aula. A mais importante característica deve ser a utilização de interfaces amigáveis;

- Integração de conteúdo: É importante para um LMS fornecer sustentação nativa em larga escala para conteúdos externos. Ao comprar um LMS, deve-se considerar que alguns LMSs são compatíveis somente com os conteúdos do próprio fornecedor. Um fornecedor de LMS deve certificar que o conteúdo externo trabalhará dentro do sistema e que se terá fácil acesso às funcionalidades do sistema;
- Aderência aos padrões: Um LMS deve tentar suportar padrões como SCORM e AICC. A sustentação de padrões significa que o LMS poderá importar e controlar o conteúdo externo, sem se preocupar com as ferramentas que foram utilizadas para produzi-lo. Deve-se ter atenção, a menos que o fornecedor certifique que o índice trabalhará, para não se ter custos adicionais;
- Capacidade de avaliação: Deve fornecer avaliações, testes, ter colaboradores que melhorem as avaliações com o passar do tempo. É uma boa idéia ter como característica do produto que ele permita ao autor criar avaliações dentro do ambiente e as incluir como parte do curso;
- Gerência de habilidades: Um componente de gerência de habilidades permite às organizações medir as necessidades de treinamento e identificar as áreas que necessitam de melhorias específicas, com base na competência coletiva dos funcionários. Essa característica também pode ser utilizada para procurar empregados com habilidades especializadas.

Segundo Hall (2003), os fatores críticos para avaliar um LMS são disponibilidade, escalabilidade, usabilidade, interoperabilidade, estabilidade e segurança:

- Disponibilidade elevada: O LMS deve ser robusto e servir simultaneamente às necessidades diversas dos estudantes, administradores e instrutores;

- Escalabilidade: A infra-estrutura deve suportar expansão futura quanto ao volume de instruções e à quantidade de estudantes;
- Usabilidade: Para suportar serviços automatizados e personalizados, tais como aprendizagem em ritmo individual, o acesso, a entrega e apresentação do material devem ser facilitados e bastante intuitivos;
- Interoperabilidade: Para suportar conteúdo de diferentes fontes e múltiplos vendedores, as soluções de *hardware/software* do LMS devem ser baseadas em padrões abertos da indústria para desenvolvimento Web (XML, SOAP ou AQ) e ser compatíveis com os principais padrões de aprendizagem (AICC, SCORM, IMS e IEEE);
- Estabilidade: A infra-estrutura do LMS deve ser confiável e permitir a execução de modo eficaz, funcionar 24x7 (24 horas do dia, nos sete dias da semana);
- Segurança: Como toda solução colaborativa, o LMS deve limitar e controlar seletivamente o acesso *on-line* ao conteúdo, os recursos e as funções administrativas, interna e externamente, para sua diversa comunidade de usuários.

No passado, muitas companhias sobreprojetaram suas soluções de aprendizagem, gastando muito dinheiro em consultorias e customização de suas aplicações, porque os modelos pedagógicos do aprendizado *on-line* ainda estavam na infância. Baseado nisso, Hall (2003) sugere que seja terceirizada a solução e que o LMS escolhido seja construído em arquitetura aberta, que suporte os padrões de aprendizagem emergentes (IMS, AICC e SCORM).

2.4.1 Exemplo de Avaliação de Ferramentas LMS

O Edutech é um grupo de trabalho suíço apoiado pelo Federal Office of Education and Science⁷, pela Swiss University Conference⁸ e University of Fribourg⁹, e trabalha com aplicações de novas tecnologias de informação e comunicação nas universidades suíças e instituições de ensino superior. Também, serve como referência para pessoas e instituições que trabalham nessa área.

⁷ <http://www.admin.ch/bbw/>

⁸ <http://shkwww.udnibe.ch/>

⁹ <http://www.unifr.ch/home/welcome.html>

Prestando consultoria para o projeto SVC (Swiss Virtual Campus¹⁰) que planeja disponibilizar em nível nacional, a partir de 2004, uma plataforma baseada na Web para projetos do ensino superior suíço, o Edutech realizou uma avaliação das ferramentas de LMS disponíveis, selecionadas por pré-seleção, com os seguintes critérios:

- A plataforma é extensamente utilizada por instituições de ensino superior;
- A plataforma já é utilizada em algum dos projetos do SVC;
- A plataforma consta nas listas de avaliações de Baumgartner¹¹ ou Schulmeister¹²;
- O grupo Edutech teve experiências positivas com a plataforma ou ouviu comentários positivos sobre ela de outras instituições.

As ferramentas pré-selecionadas e avaliadas foram:

- Blackboard ML (Blackboard Inc¹³)
- Clix 5.0 (imc AG¹⁴)
- Globalteach (TWI AG¹⁵)
- IBT Server 6.1 (Time4You¹⁶)
- Qualilearning/Luvit 3.5 (Qualilearning/Luvit AB¹⁷)
- WebCT Vista 1.2 (WebCT Inc¹⁸)

Segundo o Edutech (2003), os sistemas de gerência de aprendizagem, bem como as plataformas baseadas na Web, são sistemas complexos que têm grande quantidade de funções. Uma plataforma tem larga escala de usuários (estudantes, autores, tutores, administradores) e cada grupo de usuário tem exigências específicas, sendo difícil estabelecer uma lista formal das exigências, o que conseqüentemente torna uma avaliação uma tarefa extremamente complexa e cara.

¹⁰ <http://virtualcampus.ch/>

¹¹ P. Baumgartner, H. Häfele, K. Maier-Häfele, Auswahl von Lernplattformen, StudienVerlag, 2002. Disponível em: <<http://www.virtual-learning.at/ergebnisse.htm>>.

¹² R. Schulmeister, Lernplattformen für das virtuelle Lernen, Oldenbourg Verlag, 2002. Disponível em: <<http://www.oldenbourg.de/frame0.htm?http://www.oldenbourg.de/cgi-bin/rotitel?T=27250>>.

¹³ <http://www.blackboard.com/>

¹⁴ <http://www.im-c.de/>

¹⁵ <http://www.globalteach.com/>

¹⁶ <http://www.time4you.de/>

¹⁷ <http://www.qualilearning.com/>

¹⁸ <http://www.webct.com/>

2.4.2 Critérios de Avaliação Edutech

Com base em experiências anteriores com avaliações, o número de critérios foi reduzido tanto quanto possível, e as avaliações agora são mais subjetivas do que em outros estudos, sendo testado completamente cada produto. Abaixo, apresentam-se, de maneira sucinta, os critérios adotados, que podem ser encontrados de forma detalhada em Edutech (2003).

Critérios de avaliação do ambiente do estudante:

- Facilidade de utilização: o que inclui lista de favoritos do sistema, impressão da página corrente, busca de palavras etc.;
- Suporte a tecnologias padrões da Web: Netscape 6, IE 6, outros navegadores etc.;
- Ambiente funcional: busca por palavras-chave, ambiente configurável, ajuda sensível do contexto, coleções de ferramentas etc.

Critérios para avaliação de tutoria e didáticas:

- Facilidade de utilização;
- Comunicação: ferramentas de fórum e *chat* com grupos privados de trabalho, *e-mail* etc.;
- Gerência de estudantes: criação de grupos por tutores, criação de grupos por estudantes etc.;
- Acompanhamento de atividades: atividades individuais, atividades em grupos, testes individuais de desempenho, testes de *performance* das aulas etc.

Critérios para avaliação do desenvolvimento do curso:

- Facilidade de utilização: editores *on-line* e *off-line* para estrutura e conteúdos do curso, interface para *download* e *upload* de recursos etc.;
- Estrutura de desenvolvimento flexível: arquitetura extensível, suporte a APIs SCORM, customização da navegação etc.;
- Suporte para os desenvolvedores: tutoriais básicos para autores, manuais avançados para programadores, completa documentação de APIs etc.;
- Compatibilidade entre ferramentas de autoria padrões da Web: compatibilidade entre conteúdos, *download* e *upload* via ftp etc.;
- Avaliação: editor *on-line* e *off-line* de questionários, suporte a vários tipos de questionários etc.;

- Ambiente multilíngüe: francês, alemão, italiano;
- Suporte a XML;
- Migração dos cursos atuais do WebCT.

Critérios para avaliação do sistema e administração:

- Gerência do sistema: *performance*, disponibilidade, manutenção e segurança;
- Administração: um servidor para várias instituições, autenticação flexível, documentação etc.;
- Modelo de licenciamento flexível;
- Arquitetura técnica.

2.4.3 Conclusões da Avaliação

A partir dos resultados obtidos, o grupo Edutech fez as seguintes recomendações ao SVC:

- Adquirir licença de dois anos do WebCT – O WebCT tem uma grande base de usuários nas instituições de ensino superior suíças. A arquitetura tem potencial para suportar milhares de estudantes em diversas instituições. Embora sua capacidade de extensão seja limitada, ela enquadra-se em projetos como o SVC, devido a suas ferramentas de autoria de fácil utilização, a sua eficiência e ao seu grande número de ferramentas didáticas.
- Optar por uma plataforma extensível mais tarde – Nenhuma das plataformas avaliadas foi completamente convincente quanto a essa capacidade, entretanto, há anúncios de fabricantes a respeito disso para próximas versões. Também há plataformas adicionais que não puderam ser avaliadas, mas devem ser consideradas – IBM/Lotus Learning Space, OLAT, Oracle iLearn, dentre outras (EDUTECH, 2003).

2.4.4 Considerações sobre a Avaliação

Os resultados da avaliação não podem ser generalizados, pois o próprio Edutech enfatiza que os critérios de avaliação foram deliberadamente inclinados para o contexto do SVC.

Ao analisar os relatórios detalhados de cada ferramenta, em relação aos critérios adotados pelo Edutech, que podem ser obtidos em Edutech (2003), pode-se perceber o porquê de o grupo recomendar nova análise posterior, pois as ferramentas apresentam diferentes pontos fortes e fracos. Assim, não foi possível apontar uma como melhor, pois essa observação fica bastante dependente do contexto em que a ferramenta será utilizada.

2.5 LICENCIAMENTO DE UM LMS

Segundo Hall (2003), os sistemas de gerência de aprendizagem podem ser comparados a um sistema de ERP, em termos da complexidade e do custo da execução. Como toda grande aplicação da empresa, a instalação requer perícia e acompanhamento progressivo. Basicamente, há quatro modelos de licenciamento disponíveis no mercado:

- Compra do *software* e instalação na própria organização;
- A organização compra o *software*, mas ele é abrigado e controlado remotamente por um terceiro. Os administradores, instrutores e estudantes operam o sistema pela Internet;
- A organização compra o *software* e instala-o em sua própria plataforma, mas a manutenção e os melhoramentos são controlados pelo vendedor de LMS ou por outro terceiro integrador do sistema;
- Para quem tem preocupação com custos, há uma quarta opção. Alguns vendedores LMS operam seu próprio sistema de aprendizagem e oferecem o acesso compartilhado, no qual se aluga um ‘espaço’ no sistema do vendedor.

Para determinar qual é o modelo correto para uma organização, devem-se consultar os profissionais de TI envolvidos e traçar as potencialidades e necessidades técnicas da organização. Não há solução de LMS que se adapte a qualquer organização, mas sim diferentes fornecedores, o que permite escolhas. A decisão da organização se terá ou não infra-estrutura que permita a hospedagem ou instalação é igualmente importante. Uma vez feita a escolha, deve-se ter flexibilidade para a fazê-la outra vez, pois as necessidades de

aprendizagem mudam, geralmente crescem, não podendo acontecer de se travar na configuração inicial. Se for instalado um LMS, ele deve ser flexível, extensível e baseado em padrões abertos. Para que não haja perda de funcionalidades, como perda das informações arquivadas, o LMS deve suportar modalidades, tecnologias e ferramentas de aprendizagem novas e emergentes (HALL, 2003).

2.6 CONSIDERAÇÕES FINAIS

De acordo com Greenberg (2002), um LMS fornece um único ponto de acesso às diferentes fontes de conteúdo. Ele automatiza a administração da aprendizagem e oferece oportunidades sem precedentes para o desenvolvimento de recurso humano. Ainda, identifica as pessoas que necessitam de um curso em particular e diz a elas como o curso se encaixa em suas carreiras profissionais, quando e como está disponível (sala de aula, *on-line*, CD-ROM), se há pré-requisitos, como e quando podem cumpri-los. Uma vez que o estudante completa o curso, o LMS pode administrar os testes de proficiência, relatar os resultados e recomendar as etapas seguintes. Com essa capacidade, LMSs são instrumentos para assegurar que as organizações encontrem rígidas certificações em mercados verticalizados, finanças e governo. De maneira resumida, Hall (2003) define que, da melhor forma possível, um LMS consolidará as diversas mídias, automatizará a seleção e administração dos cursos, reunirá e entregará os conteúdos, calculará o índice e a eficácia de aprendizagem.

Devido à grande quantidade de plataformas e fornecedores de diversas nacionalidades, apresentar uma relação das principais ferramentas de mercado torna-se tarefa difícil. Em julho de 2003, seguindo alguns critérios particulares, a Virtual Learning Community¹⁹ publicou uma lista com as 16 melhores plataformas de LMS do mercado, que, abaixo, é apresentada em ordem alfabética:

- Blackboard (<http://www.blackboard.com/>)
- CLIX (<http://www.im-c.de>)
- Distance Learning System (<http://www.ets-online.de>)
- Docent Enterprise (<http://www.docent.com>)
- eLearning Suite (<http://www.hyperwave.com>)
- IBT Server (<http://www.time4you.de>)
- iLearning (<http://www.oracle.com>)
- ILF (<http://www.mit.de>)

- ILIAS (<http://www.ilias.uni-koeln.de>)
- Learning Management System (<http://www.lotus.com>)
- Sun Enterprise Learning System (<http://www.sun.de/Schulung/>)
- Saba Enterprise Learning Suite (<http://www.saba.com>)
- Sitos Cubix (<http://www.bitmedia.cc>)
- Top Class (<http://www.topclass.ch>)
- VIVERSA (<http://www.viwis.de>)
- WebCT (<http://www.webct.com>)

Berntal (2003), em seu relatório *The State of E-learning*, conclui que o *e-learning* é mais eficaz quando as organizações usam aprendizagem de portal ou sistema de gerência de aprendizagem (LMS). Aproximadamente 40% das organizações da amostra utilizam *learning* portal ou um LMS. Uma análise revelou que a eficácia percebida do *e-learning* na organização era significativamente mais elevada quando a organização tinha usado um *learning* portal ou um LMS, entretanto, é pouco provável que as organizações que se arriscam pela primeira vez no *e-learning* usem essas ferramentas, embora elas puderam ser indicativas de ‘desenvolvimento mais desenvolvido’ do *e-learning*. Para chegar a essa conclusão, Berntal pesquisou 139 organizações de 15 países, sendo 56% americanas e aproximadamente 50% com menos de 400 empregados.

Segundo o Instituto Gartner (2003), em seu relatório *E-learning LMS Magic Quadrant 2003*, em 2005, 70% das grandes corporações mundiais já terão adquirido ou desenvolvido suas plataformas para gerenciamento dos processos de aprendizagem, o que demonstra claramente o potencial desse mercado.

¹⁹ <http://iol3.uibk.ac.at/virtualllearning/>

3 DATA WAREHOUSE

Neste capítulo, apresentam-se o conceito, as características e a metodologia do *data warehouse*, porém, antes, é apresentada uma introdução aos Sistemas de Informações, para que haja contextualização do universo onde o *data warehouse* está inserido. Por fim, faz-se uma introdução ao *data mining*, que trabalha em conjunto com o *data warehouse*, apesar de seus conceitos serem dissociados.

3.1 SISTEMAS DE INFORMAÇÕES (SI)

De acordo com Dalfovo e Amorim (2000), a não-utilização de informações como recursos estratégicos leva o executivo, muitas vezes, a administrar por impulsos ou baseado em modismos. Nesse sentido, os sistemas de informação surgiram como forma de manter o executivo preparado, com visão integrada de todas as áreas, sem gastar muito tempo ou precisar ter conhecimento aprofundado de cada área.

Conforme Oliveira (1992), sistemas de informação são um método organizado de prover informações passadas, presentes e futuras, relacionadas às operações internas e ao serviço de inteligência externa. Ele serve de suporte para as funções de planejamento, controle e operação de uma empresa via fornecimento de informações no padrão de tempo apropriado para assistir o tomador de decisão.

De acordo com Alter (1992), sistemas de informação são a combinação de práticas de trabalho, informações, pessoas e informações tecnológicas organizadas para o acompanhamento de metas numa organização. Os componentes dos sistemas de informação são:

- a) Informações: Sistemas de informação podem incluir dados formatados, textos, imagens e sons. Dados são fatos, imagens ou sons que podem ou não ser pertinentes ou importantes para uma tarefa em particular;
- b) Pessoas: Exceto quando uma tarefa é totalmente automatizada, os sistemas de informação também podem necessitar de pessoas para dar entrada nos dados, processá-los ou usá-los;

- c) Informações tecnológicas: Incluem *hardware* e *software* para executar uma ou mais tarefas de processamento de dados, tais como captura, transmissão, armazenamento, recuperação, manipulação ou apresentação dos dados;
- d) Práticas de trabalho: São os métodos usados por pessoas e a tecnologia para executar os trabalhos;
- e) Objetivos: São as metas a serem alcançadas, definidas pela empresa.

A relação existente entre os componentes de um sistema de informação é apresentada na Figura 1.

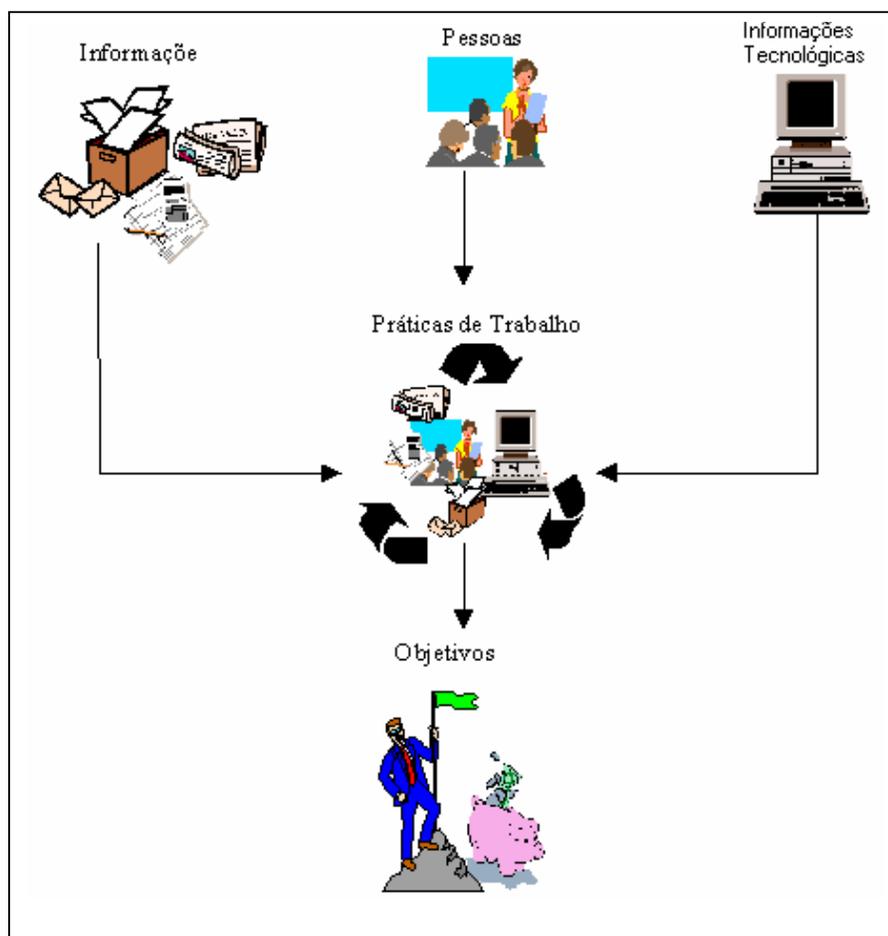


Figura 1: Componentes de um SI

Fonte: Adaptado de Alter (1992).

French (*apud* MECHELN, 1997) apresenta os estágios pelos quais as informações são processadas e classificadas. São eles:

- Dados: fatos relacionados e seus valores;
- Informação: relacionamento de entidades para formular um contexto;
- Conhecimento: relacionamento de informação ligada às regras de negócio para formular a perícia;
- Julgamento: sobre decisões ligadas ao contexto global para evidenciar regras proveitosas.

A tecnologia da informação auxilia nos três primeiros estágios, ou seja, dados, informação e conhecimento, sendo que o julgamento acerca das decisões cabe principalmente ao analista das informações (CRUZ, 1996).

Devido ao uso da informação em vários tipos de sistemas, os sistemas de informação foram classificados de acordo com sua finalidade, divididos de acordo com seu nível de influência em uma organização. Assim, tem-se: sistemas de nível operacional; sistemas de nível de conhecimento; sistemas de nível gerencial e sistemas de nível estratégico (LAUDON e LAUDON, 1996). Os sistemas de nível operacional são os que monitoram as atividades e transações elementares da organização, direcionados aos procedimentos ditos rotineiros. Os sistemas de nível de conhecimento têm como finalidade auxiliar a integrar novos conhecimentos e os negócios e auxiliar a controlar o fluxo de papéis. Os sistemas de nível gerencial são projetados para monitoração, controle, tomada de decisão e atividades administrativas dos gerentes médios. Sistemas de nível estratégico auxiliam o gerente sênior a planejar suas atividades e suportam planejamento a longo prazo.

No Quadro 1, Laudon e Laudon (1996) relacionam os tipos específicos de sistemas com cada nível organizacional.

Quadro 1: Tipos de sistemas de informação

Tipos de sistemas	Informação de entrada	Processamento	Informação de saída	Usuários	Nível
Sistema de Suporte a Executivo	Dados agregados	Gráficos; simulações; interativos	Projeções; respostas às perguntas	Gerentes sênior	Estratégico
Sistema de Suporte a Decisão	Baixo volume de dados; modelos analíticos	Interativo; simulações; análises	Relatórios especiais; análise da decisão; resposta às perguntas	Profissionais e gerentes de <i>staff</i>	Nível gerencial
Sistema de	Resumo dos	Relatórios	Sumários e	Gerentes	Nível

Informação Gerencial	dados; alto volume dos dados; modelos simples	rotineiros; modelos simples; análise de baixo nível	relatórios de exceção	médios	gerencial
Sistema Especialista	Especificação de projeto; base de conhecimento	Modelagem; simulações	Modelos; gráficos	Profissionais; <i>staff</i> técnico	Nível de conhecimento
Sistema de Automação de Escritório	Documentos	Documentos; gerenciamento; particionamento; comunicação	Documentos; <i>mail</i>	Trabalhadores de escritório	Nível de conhecimento
Sistema de Processamento de Transações	Transações; eventos	Classificação; listagem; junção; atualização	Relatórios detalhados; listas; sumários	Pessoal de operações; supervisores	Nível operacional

De acordo com Kimball e Ross (2002), um *data warehouse* é o ponto central da arquitetura de processamento de informações estratégicas para tomada de decisões e fornece suporte informacional para os sistemas de apoio à decisão (SAD) e para os sistemas de informação executiva (EIS). Freitas e Pozzebom (2003) defendem mudança do foco dos sistemas de informações executivas para sistemas de informação empresarial, em que os tomadores de decisão de todos os níveis possam se beneficiar da oferta de informações.

3.2 DATA WAREHOUSE (DW)

Os dados armazenados em um ou mais sistemas operacionais de uma organização são um recurso, mas, de modo geral, raramente servem como recurso estratégico no seu estado original. Os sistemas convencionais de informática não são projetados para gerar e armazenar as informações estratégicas, o que torna os dados vagos e sem valor para apoio ao processo de tomada de decisões (OLIVEIRA, 1998).

De acordo com Inmon (1997), um sistema de *data warehouse* é composto, dentre outras ferramentas, por um banco de dados, para onde somente as informações necessárias para a tomada de decisões são carregadas, vindas de bancos operacionais. Como esse novo banco de dados contém apenas as informações necessárias, as pesquisas feitas sobre ele são rápidas e podem responder a questões complexas.

Segundo Inmon (1997), um *data warehouse* pode ser definido como um banco de dados especializado, o qual integra e gerencia o fluxo de informações, a partir dos bancos de dados corporativos e de fontes de dados externas à empresa. Um *data warehouse* é construído

para que tais dados possam ser armazenados e acessados, de forma que não sejam limitados por tabelas e linhas, estritamente relacionais. A função do *data warehouse* é tornar as informações corporativas acessíveis para o seu entendimento, gerenciamento e sua utilização.

Uma boa solução de *data warehouse*, de acordo com Inmon (1997), tem como finalidade atender às necessidades de análise de informações dos usuários, como monitorar e comparar as operações atuais com as passadas, e prever situações futuras. Ao transformar, consolidar e racionalizar as informações dispersas por diversos bancos de dados e plataformas, ele permite que sejam feitas análises estratégicas bastante eficazes em informações antes inacessíveis ou subaproveitadas.

3.2.1 Características de um *Data Warehouse*

De acordo com Inmon (1997), um *data warehouse* é um conjunto de dados baseado em assuntos, integrado, não-volátil, variável em relação ao tempo, de apoio às decisões gerenciais.

Segundo Oliveira (1998), os dados usados pelo *data warehouse* devem ser:

- a) Orientados por assunto: Devem se orientar de acordo com os assuntos que trazem maior número de informações da organização, como, por exemplo, clientes, produtos, atividades, contas. Os assuntos são implementados com uma série de tabelas relacionadas em um *data warehouse*;
- b) Integrados: Os *data warehouses* recebem os dados de um grande número de fontes. Cada fonte contém aplicações, que têm informações, que normalmente são diferentes de outras aplicações em outras fontes. O filtro e a tradução necessários para transformar as muitas fontes em um banco de dados consistente são chamados integração;
- c) Não-voláteis: Os dados no sistema operacional são acessados um de cada vez, são cadastrados e atualizados. Já, no *data warehouse*, é diferente, pois a atualização é em massa e só acontece de tempo em tempo;
- d) Histórico: Os dados do sistema operacional podem ou não conter algum elemento de tempo, já, para o *data warehouse*, o elemento tempo é fundamental.

3.2.2 Granularidade e Particionamento

De acordo com Inmon (1997), a questão da granularidade é o mais importante aspecto do projeto de um *data warehouse*, pois se refere ao nível de detalhe ou de resumo contido nas unidades de dados existentes no *data warehouse*. Quanto menos detalhe, mais alto será o nível de granularidade. Conforme Oliveira (1998), granularidade envolve o nível de detalhamento para a sumarização de cada unidade de dados. Mais detalhes são caracterizados por um baixo nível de granularidade, enquanto menos detalhes descrevem um alto nível de granularidade. A decisão sobre o nível de granularidade das informações do *data warehouse* afeta tanto o volume contido quanto o tipo de pesquisa a que pode ser respondido.

Quando se tem um nível de granularidade muito alto, o espaço em disco e o número de índices necessários tornam-se bem menores, porém, há correspondente diminuição da possibilidade de utilização dos dados para atender às consultas detalhadas. Com nível de granularidade muito baixo, é possível responder a praticamente qualquer consulta, mas uma grande quantidade de recursos computacionais é necessária para responder a perguntas muito específicas. Esses conceitos podem ser observados na Figura 2.

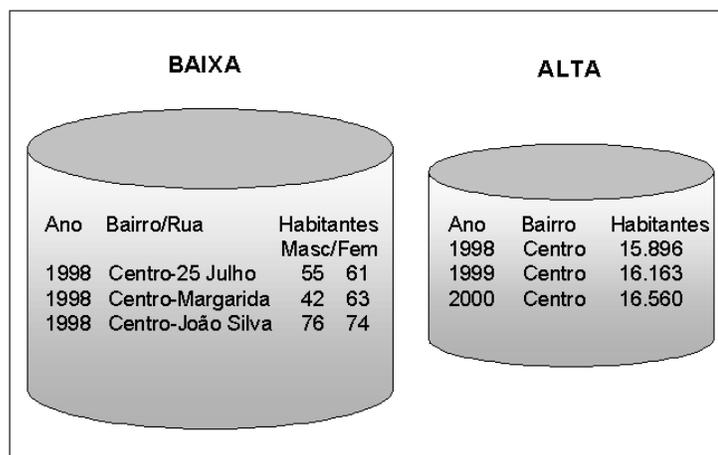


Figura 2: Níveis de granularidade

Fonte: adaptado de Inmon (1997).

De acordo com Inmon (1997), o ponto principal da definição de um *data warehouse* é descobrir o equilíbrio entre a necessidade real do nível de granularidade do usuário final do projeto e o custo envolvido nessa implantação.

Segundo Inmon (1997), depois da granularidade, o particionamento é a questão mais importante em um projeto de *data warehouse*. Particionamento refere-se à repartição dos dados em unidades físicas separadas, que podem ser tratadas independentemente.

Concordando com Inmon, de acordo como Oliveira (1998), particionamento diz respeito à divisão de dados em unidades físicas separadas, que podem ser manipuladas independentemente. Quanto menores forem as unidades físicas, mais rápido será o acesso. Uma unidade de dado é única para cada partição. O particionamento é acompanhado pela aplicação dos seguintes critérios: data, linha de negócios, geografia, unidade organizacional e todos os anteriores. Inmon (1997) acrescenta que, no *data warehouse*, as questões referentes ao particionamento de dados não enfocam a necessidade de o particionamento ser feito ou não, mas como ele deve ser feito.

3.2.3 Processo Analítico *On-line* (OLAP)

On-Line Analytical Processing, ou seja, Processo Analítico *On-line* (OLAP), é um método importante na arquitetura do *data warehouse*, mediante o qual os dados podem ser transformados em informação. OLAP é um conjunto de funcionalidades que tenta facilitar a análise multidimensional (MDA) – *Multidimensional Analysis* –, que é a habilidade de manipular dados que tenham sido agregados a várias categorias ou dimensões. O propósito da análise multidimensional é auxiliar o usuário a sintetizar informações empresariais por meio da visualização comparativa, personalizada, e também por meio da análise de dados históricos e projetados (INMON, WELCH, e GLASSEY, 1999).

De acordo com Cielo e Paz (2000), as principais vantagens de uma ferramenta OLAP referem-se à sua característica de permitir a visualização das informações de várias formas, conforme a necessidade de detalhamento. As principais características OLAP são:

- a) *Drill across*: Permite ao usuário pular um nível intermediário dentro da mesma dimensão. Por exemplo, a dimensão período é composta por ano, semestre, trimestre, mês e dia. O usuário executará um *drill across* quando ele passar diretamente para o semestre, mês ou dia;
- b) *Drill down*: Permite aumentar o nível de detalhe da informação, diminuindo o grau de granularidade;
- c) *Drill up*: Ao contrário do *drill down*, possibilita aumentar o grau de granularidade, diminuindo o detalhamento da informação;

- d) *Drill through*: Ocorre quando o usuário passa de uma informação contida em uma dimensão para outra. Por exemplo, quando se está na dimensão período e, no próximo passo, a análise da informação é feita pela dimensão bairro;
- e) *Slice and dice*: É uma das principais características de uma ferramenta OLAP. Corresponde à técnica de mudar a ordem das dimensões, mudando assim a orientação segundo a qual os dados são visualizados. Altera linhas por colunas, de maneira a facilitar a compreensão dos usuários.

3.2.4 Cubo de Decisão

De acordo com Inmon (1997), cubo de decisão refere-se a um conjunto de componentes de suporte a decisões, que podem ser utilizados para cruzar tabelas de um banco de dados, gerando visões por meio de planilhas ou gráficos, envolvendo o cálculo de dados que o usuário virá a solicitar, mas que podem ser derivados de outros dados.

Segundo Cielo e Paz (2000), os cubos são massas de dados que retornam das consultas feitas ao banco de dados e podem ser manipulados e visualizados por inúmeros ângulos (*slice and dice*) e diferentes níveis de agregação (*drill down/up*). Um cubo pode ter 'n' dimensões, sendo cada dimensão um tipo de informação. A Figura 3 mostra um cubo com três dimensões:

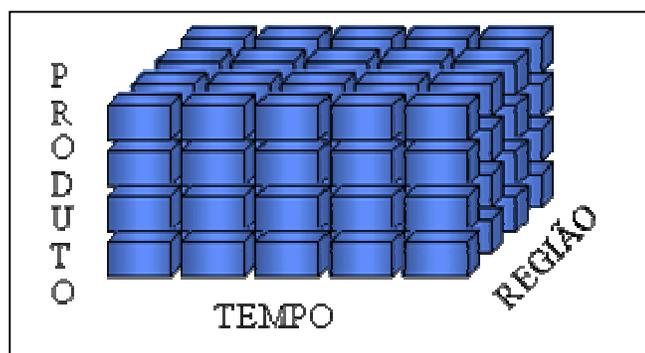


Figura 3: Cubo com três dimensões

Fonte: Adaptado de Inmon, Welch e Glassey (1999).

De acordo com Inmon, Welch e Glassey (1999), a estrutura de projeto necessária para gerenciar grandes quantidades de dados residentes em uma entidade contida no *data warehouse* é denominada *star join* (junção em estrela). Nela, a entidade que está no centro do

star join é chamada de *fact table* (tabela de fato) e será altamente povoada, pois é gerada pela combinação das informações. Em torno da tabela de fatos, estão as tabelas de dimensões.

3.2.5 Projeto de um *Data Warehouse* Dimensional

De acordo com Kimball e Ross (2002), para construir um *data warehouse*, há um processo de combinação das necessidades de informações de uma comunidade de usuários com os dados que realmente estão disponíveis. O projeto fundamenta-se em nove etapas de decisão, que são direcionadas pelas necessidades do usuário e pelos dados disponíveis. A metodologia não consiste em abordagens pré-formuladas, que podem ser aplicadas a qualquer organização, pois sempre devem ser vistas as necessidades mais importantes da organização, de forma eficiente, e se o *data warehouse* que está sendo construído é simples o suficiente para ser utilizado pelos usuários e pelo *software*. Essas nove etapas serão utilizadas neste projeto, e consistem em:

- a) Identificar os processos que se pretende modelar, correspondendo, a cada processo escolhido, uma tabela de fatos;
- b) Definir a granularidade de cada tabela de fatos para cada processo, especificando o nível de detalhe a ser representado pelo fatos;
- c) Definir as dimensões de cada tabela de fatos;
- d) Especificar os fatos;
- e) Analisar os atributos das dimensões, de modo a estabelecer descrições completas e terminologia apropriada;
- f) Decidir sobre projeto físico: agregações, dimensões heterogêneas, minidimensões etc.;
- g) Preparar dimensões para suportar evoluções (mudanças);
- h) Definir a duração do banco de dados (previsão do histórico);
- i) Definir a frequência com que os dados devem ser extraídos e carregados no *data warehouse*.

Kimball e Ross (2002) recomendam que as nove etapas da decisão sejam tomadas na ordem apresentada. As tabelas de fatos são construídas a partir da identificação dos processos. A granularidade da tabela de fato é feita a partir do nível de detalhamento das informações da tabela. As tabelas de dimensões são identificadas após ter-se a tabela de fatos, a granularidade das informações. Os fatos pré-calculados descarregarão os fatos mensuráveis na tabela de

fatos, bem como o preenchimento dos registros das tabelas de dimensões. Com a criação do modelo físico, que inclui o rastreamento de dimensões de modificações lentas, como adicionar agregados, dimensões heterogêneas, minidimensões e modos de consultas e outras decisões de armazenamento físico, a amplitude do tempo deve ser indicada para se saber quanto tempo de informações vai ficar armazenado. O tempo de extração será para indicar de quanto em quanto tempo as informações serão carregadas para as tabelas.

3.2.6 *Data Webhouse*

Conforme Kimball e Merz (2000), com o advento do *data warehouse*, surgiu a necessidade de mapear melhor as questões gerenciais dos *sites* da Web. O aumento significativo de páginas dedicadas a *e-busines* fez com que fosse pensada uma maneira de criar um armazém de dados que seguisse conceitos e características semelhantes aos do *data warehouse*, criou-se assim uma espécie de Armazém de Dados para Web, ao qual se pode chamar de *data webhouse*.

Seguindo as mesmas características do *data warehouse*, o *data webhouse* deve ser orientado por assunto, não-volátil, variável no tempo e integrado. O *data webhouse* permite analisar todo o caminho realizado por um visitante em um *site* da Web, permite mapear cada clique, conhecer melhor cada cliente via informação disponibilizada nos arquivos de *log* do servidor e *cookies*. Assim, torna muito mais fácil e real uma análise de cada evento realizado dentro do *site* pelo visitante (cliente), segmentando todas as informações para o uso adequado.

Trazer a Web para o *warehouse* significa trazer comportamentos para o *data warehouse*, pois já há bastante alimentação vinda de sistemas de processamento de transações, e parte dessa alimentação provém de transações capturadas por meio de interfaces da Web, porém, capturar somente transações não é suficiente, deve-se capturar, analisar e entender o comportamento dos usuários que clicam nos *sites* da Web.

Na Web, há uma seqüência de clique, que é literalmente um *log* de cada gesto efetuado por cada visitante a cada *site* da Web. A seqüência de clique é potencialmente um registro muito melhor de comportamento do que outras fontes detalhadas de dados mais tradicionais. Por exemplo, os dados de registro de detalhe de chamada (*call detail record* – CDR) das companhias de telecomunicações empalidecem quando comparados à seqüência de clique. Os dados de CDR somente podem mostrar que uma pessoa fez uma chamada para outra pessoa e, com sucesso, estabeleceu uma conexão por certos minutos. Não permitem saber por que uma pessoa fez uma chamada para outra pessoa. Nesse caso, devem-se indagar algumas perguntas

do tipo: Houve envolvimento de uma transação comercial? Ambas as partes ficaram satisfeitas? Quem sabe? Responder a essas perguntas será mais fácil se houver um *log* em que esteja registrado tudo o que essa pessoa fez, na seqüência de clique.

A seqüência de clique, por outro lado, é uma série cronológica de ações que pode ser agrupada em sessões. A trajetória das ações que conduziram a uma compra ou a outro comportamento em que se está interessado pode ser analisada e entendida. Pode-se ter muito mais confiança em como os indivíduos chegaram até o fornecedor, qual era seu propósito e qual a qualidade de experiência, conforme Kimball e Merz (2000).

De acordo com Kimball e Merz (2000), o *data webhouse* tem duas personalidades. A primeira, que foi vista até agora, é trazer a Web para o *warehouse* (por meio dos dados da seqüência de cliques) e a segunda personalidade do *data webhouse* é trazer o *data warehouse* existente para a Web (dispondo os serviços do *warehouse* na Web), conforme demonstra a Figura 4.

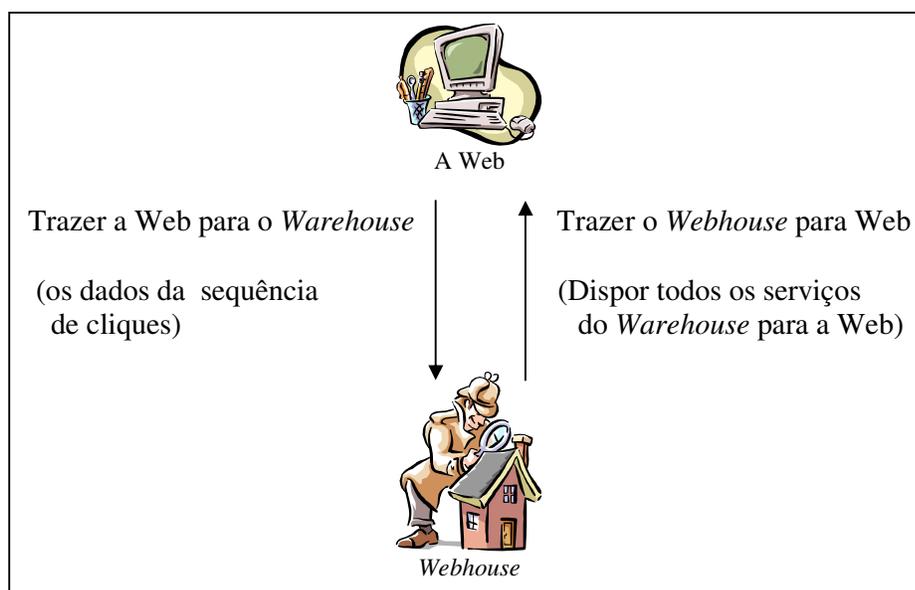


Figura 4: As duas personalidades do Webhouse

Fonte: Adaptado de Kimball e Merz (2000).

Conforme Kimball e Merz (2000), trazer o *warehouse* para a Web significa fazer com que todas as interfaces do *data warehouse* sejam disponibilizadas via navegadores da Web. Isso amplia o espectro da mera entrada de dados, para consultas *ad hoc* simples, passando por informações complexas e desenvolvimento sério de aplicativo, até chegar à administração de bancos de dados e sistemas.

Trazer o *warehouse* para a Web também significa tratar, de uma vez por todas, as questões relativas a um ambiente completamente distribuído. O *data webhouse* é uma alternativa profunda à abordagem completamente centralizada do *data warehouse*. O *data webhouse* não pode mais ser centralizado, assim como a própria Internet não pode sê-lo (KIMBALL e ROSS, 2000).

Maiores detalhes sobre *data warehouse* e *data webhouse* podem ser encontrados em Inmon (1997), Kimball e Ross (2002) e Kimball e Merz (2000).

3.3 DATA MINING (DM)

Data mining é o uso de técnicas automáticas de exploração de grandes quantidades de dados, de forma a descobrir novos padrões e relações que, devido ao volume de dados, não seriam facilmente descobertos a olho nu pelo ser humano (CARVALHO, 2001). Então, são chamadas de *data mining* todas as técnicas que permitem extrair conhecimento de uma massa de dados que, de outra maneira, permaneceria escondido nas grandes bases.

Na fase anterior ao processo do DM, tem-se o pré-processamento, no qual são executadas as fases de coleta, armazenagem e ‘limpeza’ dos dados. Para realizá-lo com sucesso, é necessário conhecimento da base, incluindo-se o entendimento dos dados e a limpeza e sua preparação, para não haver duplicação de conteúdo por erros de digitação, abreviações diferentes, valores omissos, dentre outros (QUONIAM *et al.*, 2004).

De acordo com Carvalho (2001), para o aprendizado ocorrer, uma série de informações de diferentes formatos e fontes precisa ser organizada de maneira consistente, na grande memória empresarial denominada *data warehouse*, pois as técnicas de *data mining* necessitam de bancos de dados limpos, padronizados e organizados. Após obtido isso, métodos de análise estatísticas e de inteligência artificial precisam ser aplicados sobre esses dados e relações novas e úteis devem ser descobertas.

A mineração de dados pode ser realizada de três diferentes formas (metodologias), em função do nível de conhecimento que se tem do problema estudado:

- a) Descoberta não supervisionada de relações: Se nada se sabe sobre o comportamento do fenômeno, pode-se simplesmente deixar que as técnicas automáticas do *data mining* procurem nos dados relações ‘novas’ escondidas, e que, a olho nu, não se poderiam localizar com facilidade;

- b) Testagem de hipótese: Quando se tem algum conhecimento sobre o campo de atuação da empresa ou alguma idéia sobre que relação nova se busca, pode-se definir uma hipótese e verificar sua confirmação ou refutação;
- c) Modelagem de dados: Quando se tem um nível maior de conhecimento da área e da relação que se deseja estudar.

Qualquer uma das três possíveis metodologias necessita basicamente das mesmas técnicas para ser realizada. As técnicas são de caráter genérico e podem ser implementadas por meio de ferramentas diferentes, como redes neuronais artificiais, estatística ou inteligência artificial simbolista. Apesar de alguns autores apresentarem grande número de técnicas básicas, cinco técnicas gerais abraçam didaticamente todas as outras formas de apresentação e permitem visão global e apropriada para uma introdução ao assunto:

- a) Classificação: Classificar um objeto é determinar com que grupo de entidades, já classificadas anteriormente, esse objeto apresenta mais semelhança. Redes neuronais artificiais, estatística e algoritmos genéticos são algumas das ferramentas muito utilizadas para classificar dados;
- b) Estimativa: Estimar uma grandeza é avaliá-la, tendo como base casos semelhantes, nos quais essa grandeza esteja presente. Redes neuronais artificiais, estatística, algoritmos genéticos e *simulated annealing* são algumas ferramentas muito utilizadas para estimar grandezas;
- c) Previsão: A previsão consiste na determinação do futuro de uma grandeza. Redes neuronais artificiais e estatística são ferramentas utilizadas em previsão;
- d) Análise de afinidade: A análise de afinidade preocupa-se em descobrir que elementos dos eventos têm relações no tempo e determinar que fatos ocorrem simultaneamente com probabilidade razoável (co-ocorrência). A inteligência artificial simbolista é utilizada em análise de afinidade;
- e) Análise de agrupamentos: Agrupar é, com base em medidas de semelhança, definir quantas e quais classes existem em um conjunto de entidades. Redes neuronais artificiais, estatística e algoritmos genéticos são ferramentas utilizadas para a análise de agrupamentos (CARVALHO, 2001).

Neste estudo, utiliza-se o método *k-means*, que cria um modelo de agrupamento, isto é, descobre semelhanças e agrupa os dados de forma a maximizar a similaridade dentro do grupo e a maximizar a diferença entre grupos (BATISTA, 2004).

3.3.1 *K-Means*

O algoritmo *k-means* é um método de classificação em que a distância entre um ou mais pontos centrais para todos elementos serve como medida para criar os agrupamentos. Ele é capaz de descobrir semelhanças nos dados originais e agrupá-los de forma a maximizar a similaridade dentro do grupo e a maximizar a diferença entre grupos (HALMENSCHLAGER e VALIATI, 2004).

De acordo com Coelho e Ebecken (2004), o algoritmo de *k-means* inicia-se selecionando-se aleatoriamente *k* objetos para representar cada um dos *clusters*. Cada um dos objetos restantes na base de dados é agrupado no *cluster* que tem o elemento representante mais similar, utilizando-se a média dos objetos dentro de cada *cluster* para representá-lo.

Esse algoritmo é muito simples e tem custo computacional razoável, entretanto, também apresenta dois grandes problemas. Primeiramente, é bastante inconveniente requerer o valor de *k* como parâmetro. Acontece que, para dados reais, normalmente essa informação é desconhecida. O outro problema dos algoritmos de segmentação baseados em particionamento é que eles são muito sensíveis ao particionamento inicial que é feito (pois essa é uma escolha aleatória de *k* objetos para representar inicialmente cada *cluster*). Os algoritmos realizam a busca de um ponto de máximo para o seu critério de parada, a partir dessa solução inicial. Não existem garantias de que os algoritmos encontrarão o máximo global. É inclusive possível que os algoritmos apresentem diferentes soluções, quando executados mais de uma vez (COELHO e EBECKEN, 2004). Essa técnica é utilizada, em geral, quando se deseja mapear dados em classes predefinidas.

Maiores detalhes sobre *data mining* e o algoritmo de *k-means* podem ser encontrados em Carvalho (2001) e Coelho e Ebecken (2004).

3.4 CONCLUSÃO

Neste capítulo, apresentou-se uma revisão teórica sobre *data warehouse*, mostrando seus principais conceitos, com o propósito de fundamentar a elaboração e implementação do modelo proposto neste trabalho. Iniciou-se pela contextualização do *data warehouse* no universo dos sistemas de informações, apresentando-se a definição, as características e os tipos de sistemas, e concluiu-se que o *data warehouse* é o ponto central da arquitetura de processamento de informações estratégicas para tomada de decisão.

Após, foi apresentado o conceito, as características e a metodologia de desenvolvimento do *data warehouse*, abordando-se aspectos de suma importância, como granularidade, particionamento, processo analítico *on-line* (OLAP), cubo de decisão e projeto dimensional.

Finalizou-se com uma introdução à mineração de dados – *data mining* – que pode incrementar o modelo, ao ajudar a descobrir informações que não seriam facilmente descobertas a olho nu. Apresentaram-se suas metodologias e técnicas, com maior enfoque no algoritmo de classificação supervisionado *k-means* e demonstrou-se também que, para se aplicar técnicas de mineração, é necessário ter a memória empresarial organizada de forma consistente, o que se obtém com a criação de um *data warehouse*.

4 SOLUÇÃO PROPOSTA

O modelo de aplicação de sistemas de informações desenvolvido neste estudo visa à alavancagem dos sistemas de gerência de aprendizagem, com a disponibilização de informações estratégicas em tempo real, sobre a logística do processo de aprendizagem. Ele foi desenvolvido no Laboratório de Ensino a Distância da UFSC, com base na experiência de utilização de seu ambiente virtual de aprendizagem (VIAS-K). O modelo proposto a seguir é o resultado de um extenso trabalho de análise com diversos agentes envolvidos, como coordenadores pedagógicos, professores, tutores, monitores, em que se levantaram seus critérios e informações necessárias para uma avaliação estratégica da logística do processo de aprendizagem. Com isso, visava-se a uma abordagem clara que pudesse ser aplicada a outros sistemas de gerência de aprendizagem, com o mínimo de esforços.

4.1 MODELO DO SISTEMA

Para criação do modelo proposto, seguiu-se o modelo de projeto de *data warehouse* dimensional visto no capítulo 3, apresentado por Kimball e Ross (2002), que se baseia em um processo de combinação das necessidades de informações de uma comunidade de usuários com os dados que realmente estão disponíveis.

4.1.1 Identificação dos Processos

Após análise realizada com os diversos agentes envolvidos, observaram-se os seguintes processos chave para análise da logística do processo de aprendizagem:

- a) Interação: A interação por si gera conhecimento, sendo o objetivo a análise para estímulo da interação entre os próprios acadêmicos, que deixam de ser apenas receptores e passam a agentes ativos na criação de conhecimento;
- b) Suporte: Objetiva a análise de o quão claro está o conteúdo, verificação de se o processo de interação entre acadêmicos está surtindo efeito na compreensão do conteúdo e análise das dificuldades que o acadêmico encontra em relação à plataforma;
- c) Desempenho: Assim como suporte, a análise de desempenho busca avaliar o conteúdo e a eficácia do processo de interação, além de servir como índice para verificação de possíveis medidas adotadas;

- d) Utilização da plataforma: Compreendendo-se os hábitos de utilização da plataforma, podem-se planejar melhores atualizações e horários de atividades.

4.1.2 Definição dos Fatos a Serem Analisados

Usualmente, cada processo que se deseja avaliar torna-se uma tabela de fato, porém, para o processo de interação, foram selecionadas as ferramentas de *chat* e fórum, que, apesar de serem direcionadas à interatividade, representam informações distintas. Essas ferramentas são amplamente difundidas entre as plataformas de ensino e representam exatamente a informação esperada, a interação acadêmica. Os demais processos geram uma tabela de fatos cada. Assim ficaram definidas as tabelas de fatos:

- participação fórum;
- participação *chat*;
- envio de dúvida;
- avaliação;
- acesso ao sistema.

Os três primeiros fatos representam uma participação quantitativa. Depois de apurado um fato, a análise qualitativa das interações deve ser possível, por meio da plataforma de ensino. A avaliação é representada pelo desempenho obtido, nota. Apesar de o acesso também ser uma informação quantitativa, sua análise fica voltada aos horários de utilização.

4.1.3 Dimensões a Serem Analisadas e Granularidade

As dimensões comuns a todos os fatos utilizadas para análises associativas entre eles são:

- tempo (ano, mês, quartil do mês, final de semana);
- aluno (nome, CPF, *e-mail*, sexo e data de nascimento), que representa a figura do acadêmico;
- curso (título), turma em que o aluno está matriculado;
- endereço aluno (região e estado);
- conteúdo (objeto de conhecimento): Visando a facilitar as análises associativas em relação ao conteúdo, recomenda-se utilizar os objetos de conhecimento em níveis

que representem as disciplinas. Essa dimensão não está presente somente para o fato de acesso ao sistema.

Dimensões para análises específicas dos fatos:

- tutor (nome, CPF, *e-mail*, sexo, data de nascimento, região e estado): Representa o responsável pelo acompanhamento dos alunos, pelo estímulo às discussões sobre o conteúdo. Está presente nos fatos participação fórum e participação *chat*;
- fórum (título): Fórum ao qual as participações estão vinculadas. Está presente no fato participação fórum;
- sala *chat* (título): Sala em que a conversação foi realizada, presente no fato participação *chat*;
- tipo dúvida (título): Tipo da dúvida enviada. Sugere-se uma classificação simples, como dúvidas técnicas, administrativas ou de conteúdo. Está presente no fato envio dúvida;
- professor (nome, CPF, *e-mail*, sexo, data de nascimento, região e estado): Representa a figura do responsável pela elaboração do conteúdo e montagem das aulas. Está presente nos fatos avaliação e envio dúvida;
- atividade (título): Atividade de avaliação realizada, presente no fato avaliação;
- horário (período, hora): Horário do dia, presente no fato acesso sistema.

4.2 PROJETO

A Linguagem Unificada de Modelagem (UML) é uma linguagem padrão para especificar, visualizar, documentar e construir artefatos de um sistema, e pode ser utilizada com todos os processos, ao longo do ciclo de desenvolvimento e por meio de diferentes tecnologias de implementação. O diagrama de caso de uso fornece um modo de descrever a visão externa do sistema e suas interações com o mundo exterior, representando uma visão de alto nível de funcionalidade, mediante o recebimento de um tipo de requisição do usuário. O diagrama de classes é a essência da UML. Trata-se de uma estrutura lógica estática, que mostra uma coleção de elementos declarativos de modelo, como classes, tipos e seus respectivos conteúdos e relações (FURLAN, 1998).

A Figura 5 demonstra o diagrama de caso de uso do modelo proposto, no qual se tem os atores e seus casos de uso, que são suas interações com o ambiente.

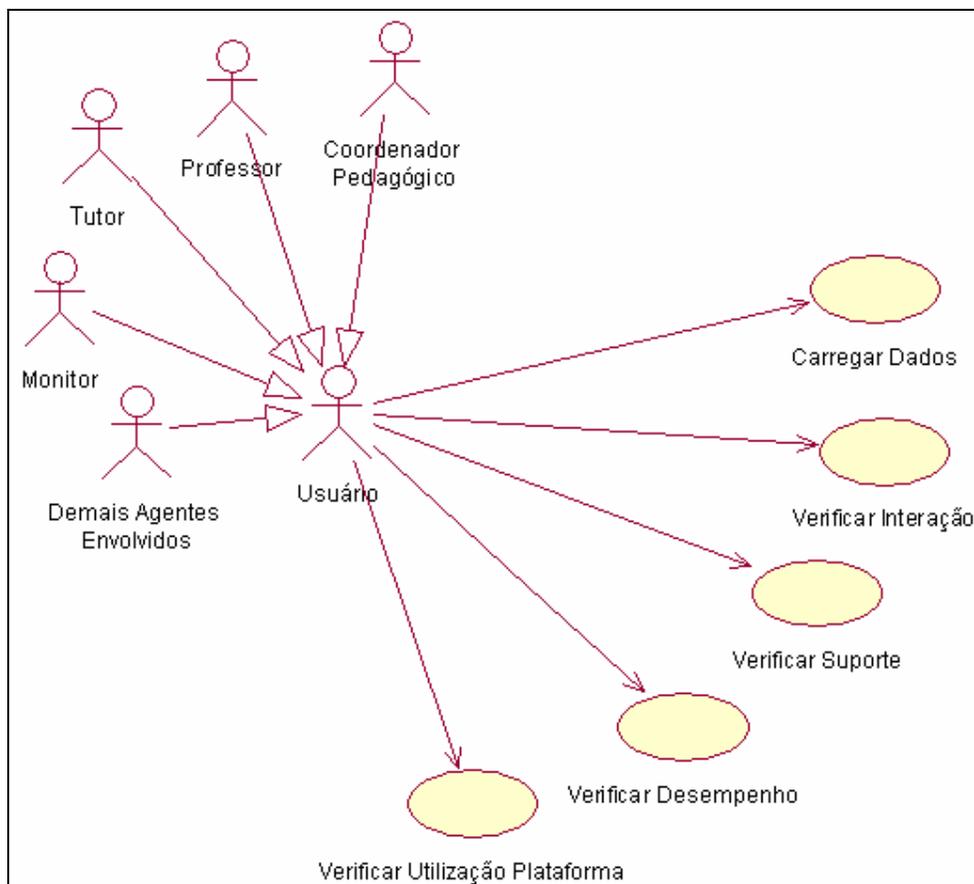


Figura 5: Diagrama de caso de uso do modelo

O Quadro 2 traz informações sobre cada caso de uso, nome, ator que inicia a ação e descrição de sua funcionalidade. Nele, o ator, sempre denominado usuário, representa todos os agentes envolvidos no processo, beneficiando todos os possíveis tomadores de decisão com a oferta de informações.

Quadro 2: Casos de uso do modelo

Caso de uso	Ator	Descrição
Carregar Dados	Usuário	Com os dados operacionais da plataforma de ensino e informações externas, caso a plataforma em questão não gere todas as informações necessárias, é realizada a carga dos dados para o <i>data warehouse</i> .
Verificar Interação	Usuário	Consulta da interação dos acadêmicos nas ferramentas de <i>chat</i> e fórum por sala de <i>chat</i> ou fórum, tutor, conteúdo, curso, tempo e aluno.
Verificar Suporte	Usuário	Consulta sobre a solicitação de suporte, envio de dúvidas por aluno, tempo, curso, conteúdo,

		professor e tipo de dúvida.
Verificar Desempenho	Usuário	Consulta do desempenho nas avaliações por conteúdo, curso, professor, tutor, tempo e aluno.
Verificar Utilização Plataforma	Usuário	Consulta sobre o tempo de acesso à plataforma de ensino por aluno, curso e tempo.

Na Figura 6, mostra-se o diagrama de classes do *data warehouse*.

Para utilização de um banco de dados relacional e a especificação orientada a objetos, deriva-se o modelo de objetos em tabelas relacionais, criando-se uma tabela para cada classe, em que os atributos formam as colunas. No modelo, optou-se pela não-representação de uma classe abstrata pessoa, para as dimensões tutor, professor e aluno. No caso da representação da classe abstrata pessoa, que seria herdada pelas classes tutor, professor e aluno, sendo necessário gerenciar objetos descendentes de forma genérica, os atributos herdados devem ser copiados para as tabelas que representam as classes descendentes, não sendo criada tabela para a classe abstrata. Cada instância de classe é representada por uma linha na tabela de sua classe.

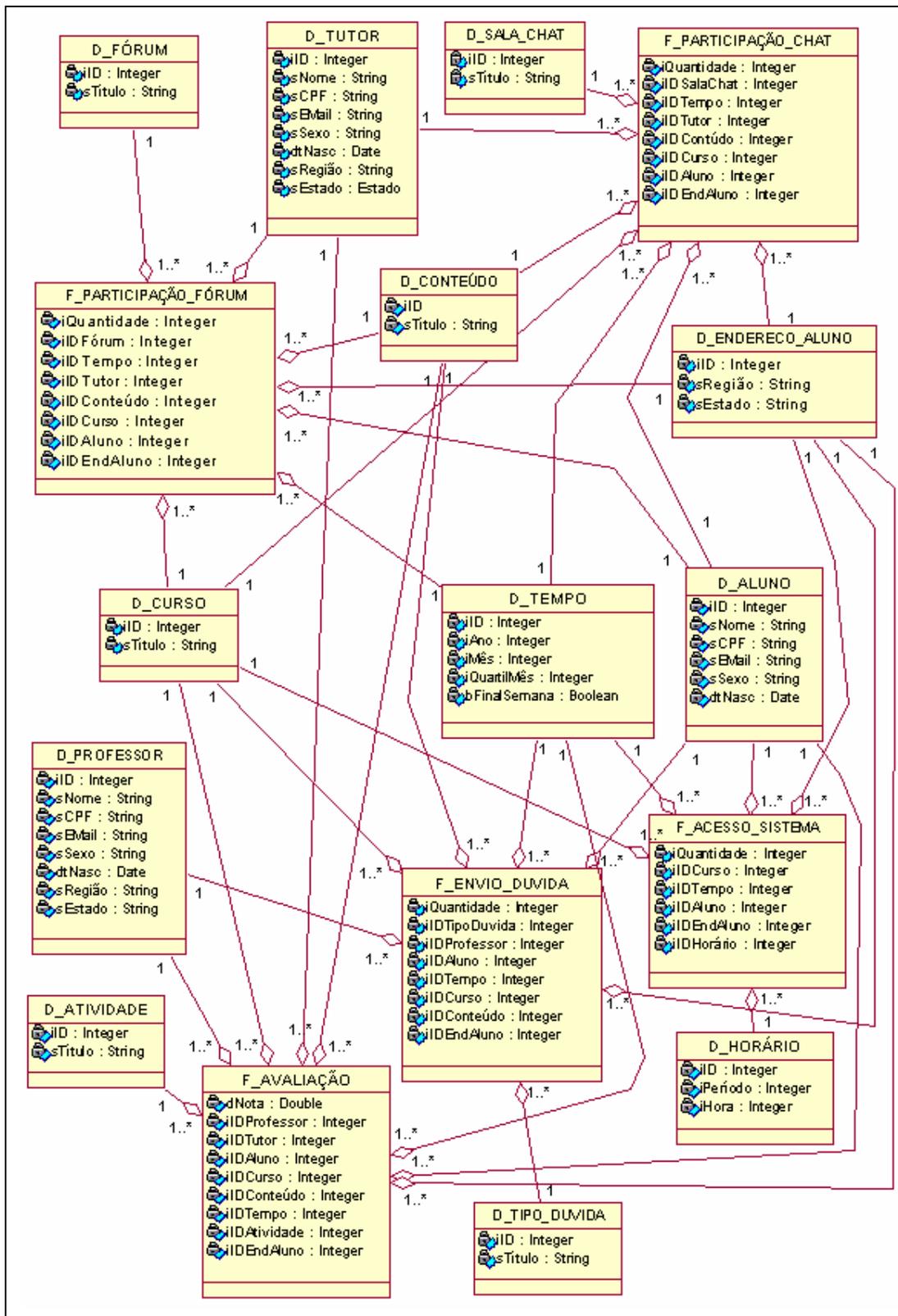


Figura 6: Diagrama de classes do data warehouse

4.3 ÁREA DE TRANSIÇÃO

A área de transição abrange tudo entre os sistemas operacionais de origem e o *data warehouse*, sendo tanto uma área de armazenamento como um conjunto de processos visando à extração, limpeza, transformação e migração dos dados dos sistemas existentes para o *data warehouse*, aquisição dos dados.

A Figura 7 apresenta o projeto de área de transição para o modelo proposto.

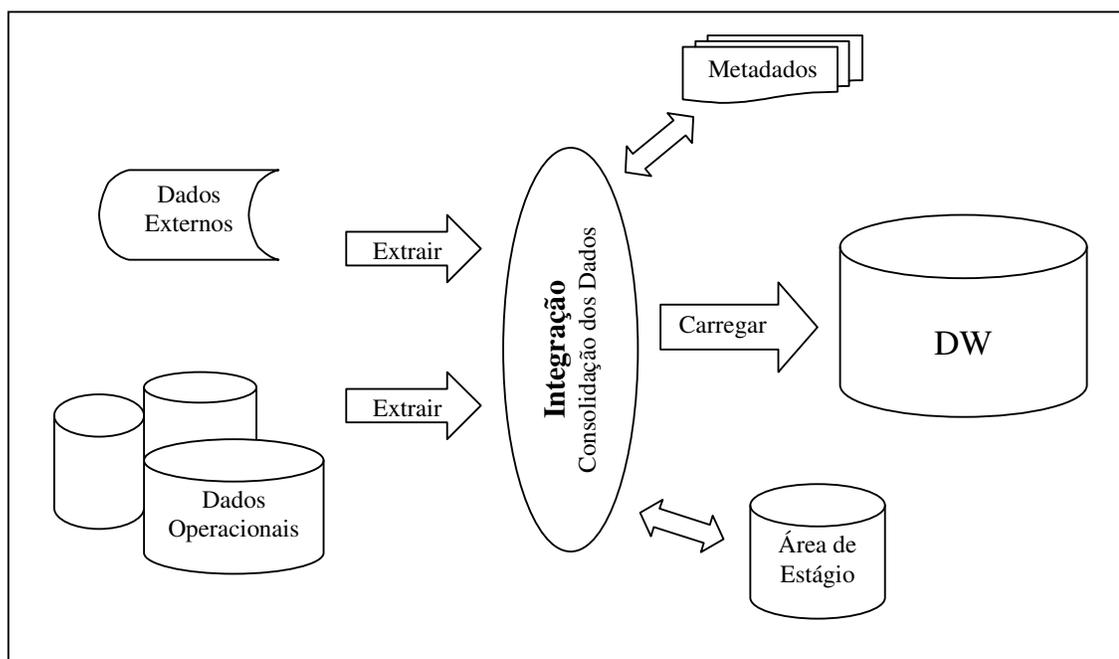


Figura 7: Projeto da área de transição

Com os dados operacionais coletados da plataforma de ensino e os dados externos, realiza-se a integração e o armazenamento no *data warehouse*. A integração de dados apresenta-se como fator primordial para aplicação do modelo proposto em qualquer sistema de gerência de aprendizado, em que informações não disponíveis pela plataforma em questão podem ser incorporadas por meio de controles externos.

O termo metadados refere-se a todas as informações no ambiente de *data warehouse* que não são os dados propriamente ditos, são dados sobre os dados. Na área de transição, os metadados visam a orientar os processos de transformação e a carga dos dados, com regras de transformação, definições de dimensão e fato em conformidade, resultados de *log* de execução e cronogramas de extração, transformação e limpeza. Até código de programação personalizado, criado na área de transição, são metadados (KIMBALL e ROSS, 2002).

4.4 CONSIDERAÇÕES SOBRE O MODELO

A criação do modelo não visa a esgotar, finalizar o assunto, mas sim servir de referência para a construção de sistemas, exemplo de aplicação. As dimensões representam uma sugestão de trabalho, manuseio da informação, e poderiam ser alteradas de acordo com a necessidade de análise, sendo que alterações podem e devem ser feitas na busca de aproximá-lo a realidade em que for aplicado. Por exemplo:

- Criar novas dimensões: poder-se-ia aplicar uma dimensão de endereço para todos os atores envolvidos, não apenas ao aluno, como no modelo, caso eles também estejam distribuídos em diferentes cidades, estados, regiões etc.;
- Aumentar ou diminuir o nível de detalhamento, granularidade das dimensões, de acordo com a necessidade em questão;
- Alterar o nível de sumarização de dimensões, como a de conteúdo, que, por utilizar objetos de conhecimento, pode ser representada em diversos níveis, desde estruturas amplas, como uma disciplina, ou menos, como seus itens, tópicos específicos dela.

Apesar de não se demonstrar isto explicitamente no modelo, a utilização da mineração de dados pode incrementar, refinar a oferta de informações. Como exemplo disso, tem-se os algoritmos de agrupamento, que visam a descobrir semelhanças nos dados originais e agrupá-los, de forma a maximizar a similaridade dos dados. Dentre eles, cita-se o método de classificação *k-means*, que, por trabalhar de modo supervisionado, permite, além da classificação em cima de dados relevantes, de interesse de quem os estiver analisando, a alteração do número de agrupamentos, podendo, por exemplo, diminuir a quantidade de grupos até se conseguir visualizar e interpretar os resultados.

4.5 RESULTADOS ESPERADOS

Com a implementação do modelo, buscam-se informações sobre a logística da aprendizagem, por meio da análise desestruturada dos processos de interação, suporte e avaliação. Não necessariamente, buscam-se conclusões, mas sim indicativos que, associados ao conhecimento prévio de quem estiver realizando as análises, ajudem na compreensão do processo de aprendizagem.

Abaixo, são apresentados exemplos de questionamentos. Tendo em vista as informações e os indicativos obtidos pela correlação dos fatos através das dimensões, ter-se-ão mais subsídios para sua compreensão:

- Os alunos que mais interagem são os que consultam o suporte? Se estimulada a interação, diminui a quantidade de consultas?
- Isso serve para o desempenho, qual a sua relação com a interatividade? Realmente, quem interage mais alcança melhor desempenho?
- Quais tutores estimulam mais a interatividade dentre os acadêmicos? Identificando e disseminando suas práticas de trabalho, a interatividade aumenta?
- Quanto a fatores culturais, acadêmicos de determinadas regiões interagem mais ou menos? O desempenho é uniforme ou varia, dependendo do conteúdo que é estudado?
- Quais conteúdos geraram menos dúvidas? Quais professores o fizeram? Adotando-se a maneira de estruturação e os tipos de mídias para conteúdos que geram muitas dúvidas, elas diminuem?
- Conteúdos sem dúvidas são conteúdos com boas notas?
- Qual o horário de maior acesso ao sistema? Direcionando-se as atividades colaborativas para ele, a interatividade aumenta?

Como essas questões, poderiam ser levantadas muitas outras, mas a idéia não é esgotar as possibilidades, e sim demonstrá-las, pois o objetivo é permitir a análise desestruturada, em que o tomador de decisão, usuário que estiver explorando os dados, decide o que é importante e de que maneira ele quer visualizar isso. Por exemplo, se o interesse fosse verificar as áreas de conteúdo que estimulam discussão, essa informação poderia ser obtida de várias maneiras:

- conteúdo – participação;
- curso – conteúdo – participação;
- alunos – conteúdo – participação;
- curso – alunos – conteúdo – participação;
- endereço alunos (região, estado) – conteúdo – participação;
- tutor – conteúdo – participação (alunos);
- conteúdo – tempo (mês, semana) – participação;
- etc.;

Isso seria feito, demonstrando-se a análise de um fato (participação), privilegiando-se uma dimensão (conteúdo). Imagine-se isso sob a ótica de todas as dimensões ou com o cruzamento dos fatos. Então, ao se obter os dados de maneira consolidada e agrupada, o usuário poderia explorá-los, correlacionando-os da maneira que lhe fosse conveniente, de acordo com seu conhecimento prévio sobre o assunto. Assim, ele não ficaria preso a meia dúzia de relatórios pré-formatados.

5 APLICAÇÃO: SISTEMA DE APOIO À DECISÃO PARA LMS

A Caixa Econômica Federal, por meio do Laboratório de Ensino a Distância, em parceria com a Universidade Federal de Santa Catarina, produziu e executou o curso de Formação em Gestão Pública no Brasil, cujo objetivo principal foi habilitar e qualificar seus participantes para a leitura, síntese e análise do cenário político brasileiro, bem como para a avaliação de temas relacionados à gestão pública, sobretudo, aqueles voltados para a Lei de Responsabilidade Fiscal.

O curso contou com a participação de 1.600 alunos, distribuídos por todas regiões e estados brasileiros, sendo 1.500 deles superintendentes de negócio, gerentes de mercado do segmento estados e municípios, gerentes gerais, gerentes e técnicos das áreas, gestores de produtos da matriz da Caixa e os demais foram convidados de órgãos externos. A duração do curso foi 180 horas de aula, sendo iniciado em 20 de agosto de 2002 e terminado em 20 de dezembro do mesmo ano.

Sendo esse o contexto de aplicação do modelo proposto no capítulo anterior, em seguida, apresentam-se detalhes da implementação.

5.1 MODELO APLICADO

Dos processos levantados pelo modelo, somente o de avaliação não pôde ser aplicado ao curso em questão – fato que será melhor explicado adiante –, o qual resulta na exclusão do caso de uso verificar desempenho, apresentado na Figura 5, no capítulo anterior, diagrama de casos de uso do modelo.

Para a geração da base de dados, a não-implementação do processo de avaliação implica a retirada do fato avaliação e da dimensão atividade, única utilizada somente por esse fato. A Figura 8 apresenta o Modelo Entidade Relacionamento (MER) da aplicação, em que estão evidenciadas todas as tabelas contidas no banco de dados utilizado e seus relacionamentos.

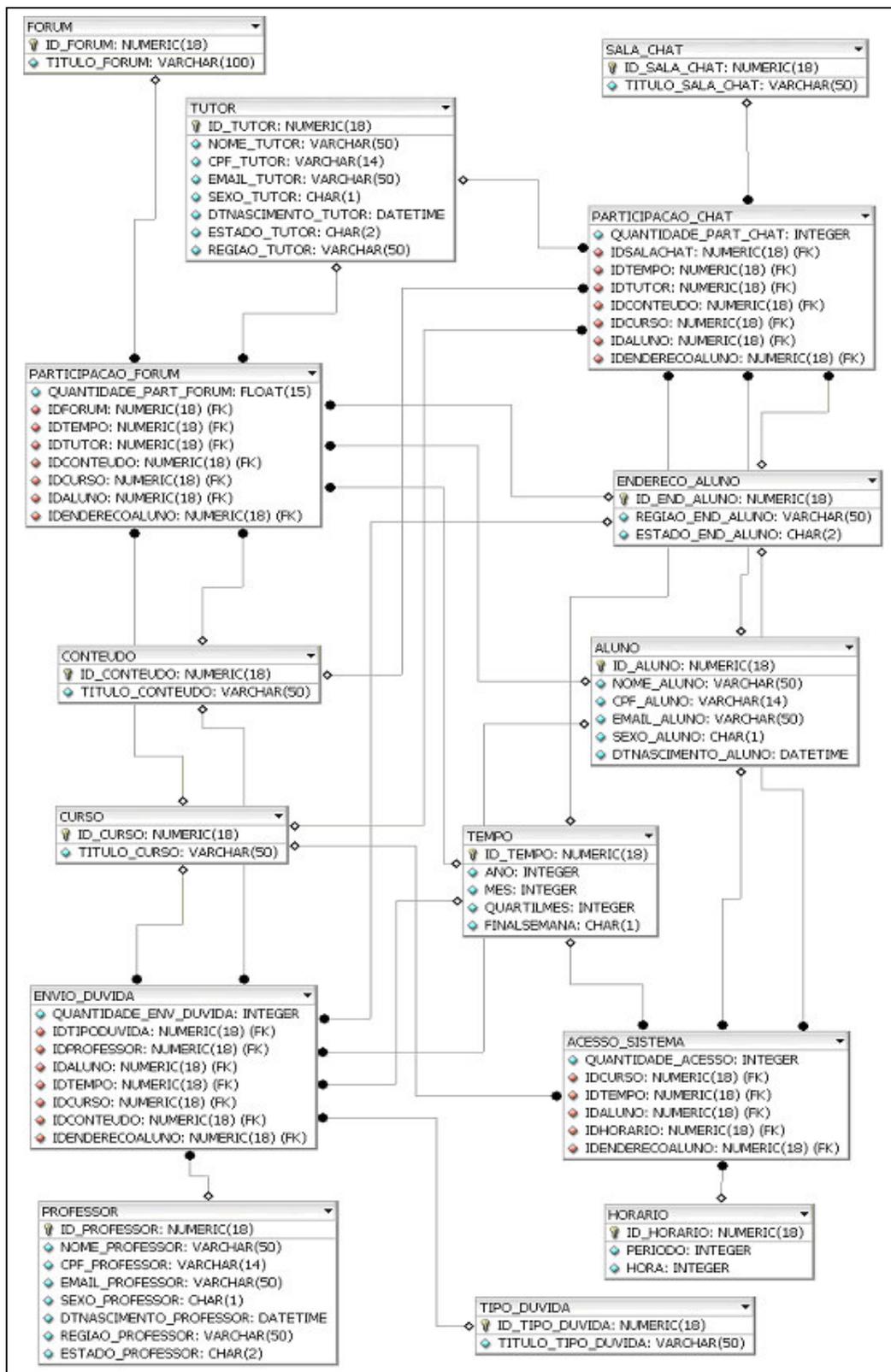


Figura 8: Modelo entidade relacionamento da aplicação

O banco de dados utilizado para o desenvolvimento da aplicação foi o Microsoft SQL Server – Enterprise Edition®.

5.2 EXTRAÇÃO DE DADOS PARA O AMBIENTE

O modelo para área de transição apresentado no capítulo anterior, Figura 7, representa com exatidão o processo de extração, limpeza, transformação e migração dos dados do sistema existente para o *data warehouse*, processo de aquisição dos dados.

A princípio, imaginou-se a utilização de dados externos, com o intuito de generalizar o modelo ao ser aplicado em outras plataformas, uma vez que a plataforma em questão possui ferramentas que controlam e geram informações a respeito dos processos (dados) propostos para análise no modelo, mas não foi essa a realidade encontrada. Não se sabe exatamente o porquê, se por falta de conhecimento da plataforma ou simplesmente opção gerencial, o fato é que vários controles referentes ao curso foram realizados de forma externa à plataforma, o que acabou se mostrando interessante para aplicação do modelo, pois comprovou a idéia de generalização que se propôs.

Como mencionado, dos processos levantados pelo modelo, somente o de avaliação não pôde ser implementado, devido ao fato de o controle de entrega das avaliações ter sido implementado à parte da plataforma e não se ter acesso a eles. As avaliações consistiram em auto-avaliações ao final de cada capítulo e entrega de artigo ao final de cada módulo, sendo a entrega dos artigos realizada diretamente nos *e-mails* particulares dos tutores.

Para coleta de dados, desenvolveu-se uma aplicação com a linguagem Java, que se conecta à base da plataforma ou interpreta arquivos externos exportados para arquivos de texto e carrega os dados desejados para a área de estágio, onde os dados são integrados e posteriormente carregados para o *data warehouse*. Isso pode ser observado na Figura 9.

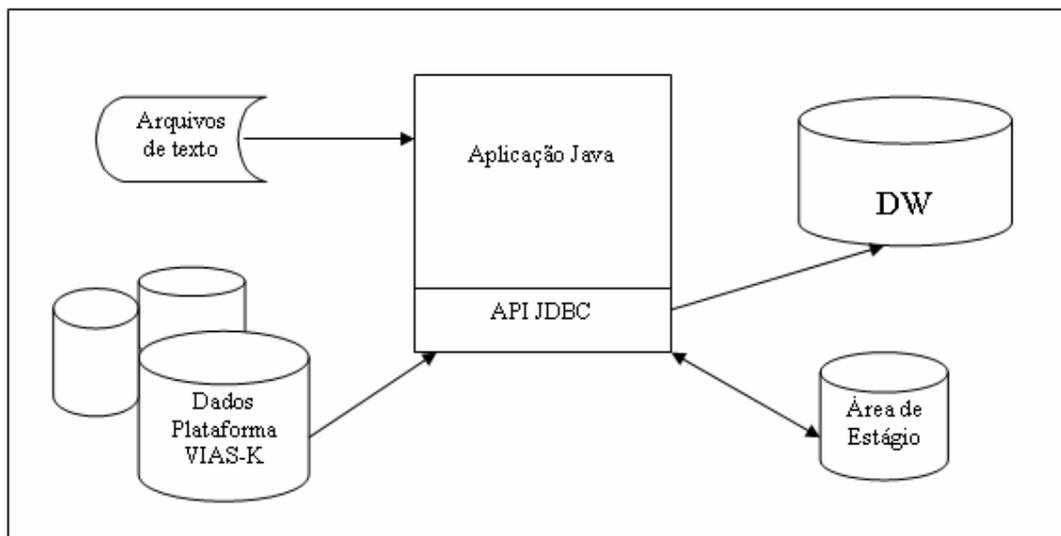


Figura 9: Interação dos componentes de extração

JDBC é uma API (Application Program Interface) para acesso a SGBD (Sistemas Gerenciadores de Banco de Dados) relacionais, por meio de comandos SQL (Structured Query Language). A API para programação do sistema é a mesma para qualquer SGBD.

A primeira tarefa, ao se realizar a carga, é a extração e transferência dos dados referentes às dimensões para área de estágio. Para as dimensões fórum, sala *chat*, curso, aluno, tutor, professor e tipo dúvida, os dados foram apenas copiados da base da plataforma para a área de estágio, preservando-se inclusive o valor dos identificadores. Para a dimensão conteúdo, criou-se uma tabela de transição, responsável pelo mapeamento do identificador na base da plataforma para o identificador utilizado no DW. Para as dimensões endereço aluno, tempo e horário, criaram-se rotinas de criação na aplicação, sendo que o campo período da dimensão horário varia de 1 a 4, representando períodos de seis horas, que se iniciam à zero hora.

Após a carga das dimensões, realiza-se a carga dos fatos, seguindo a mesma seqüência que será exemplificada mediante a carga dos fatos participação *chat*. Inicialmente, selecionam-se os fatos que serão carregados juntamente com dados suplementares, utilizados na identificação das dimensões dos fatos. Nesse caso, realizou-se uma união entre as tabelas *chat* e *chat* mensagem, para a turma a ser carregada, o que pode ser visto no Quadro 3.

Quadro 3: SQL de seleção dos fatos participação *chat*

```
"select cm.id_salachat as idsala, cm.horaenvio as hora, cm.id_remetente as
idusu, sc.data as data from chatmensagem cm, salachat sc where
cm.id_salachat = sc.id and sc.id_turma = "+iIdTurma
```

Tendo identificadas as dimensões sala *chat*, curso e aluno. Para identificação da dimensão tutor, como a ferramenta de grupos não foi utilizada por todos os tutores do curso, utilizaram-se os dados de uma planilha de controle dos próprios tutores, de onde se gerou um arquivo de texto contendo o CPF do aluno e o nome ou apelido do tutor, com mostra a Figura 10.

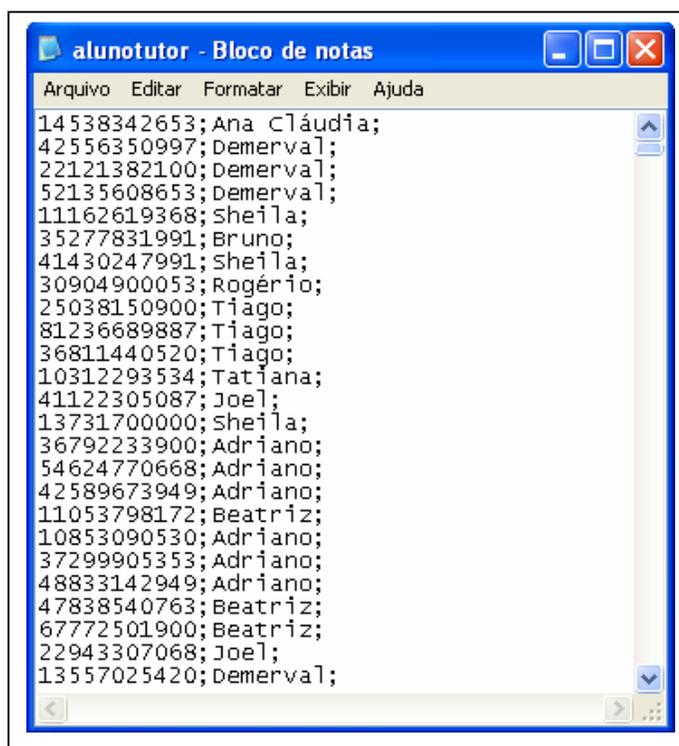


Figura 10: Arquivo de relacionamento aluno tutor

Assim, resta à aplicação identificar o nome ou apelido do tutor pelo CPF do aluno e mapeá-lo para o identificador da dimensão tutor. Para a dimensão endereço aluno, é recuperado o endereço dele na base da plataforma de ensino e localizado o identificador da dimensão no DW. De maneira similar, o identificador da dimensão tempo é recuperado, baseando-se na data de ocorrência do fato. Em decorrência de não se ter selecionado o conteúdo na criação dos *chats*, na maioria dos casos, a identificação dessa dimensão também

foi baseada nas datas de ocorrência dos fatos, sendo que, para aumentar a confiabilidade dos dados, essa dimensão foi restrita às duas grandes áreas do curso. Criou-se também uma tabela de transição, mapeando-se o identificador na base da plataforma para o identificador utilizado no DW, o que se utilizará em futuras cargas, em que as informações referentes ao conteúdo estiverem completas.

Após a identificação de todos os fatos, verifica-se a existência do fato em questão. Caso o fato ainda não exista, ele é inserido, caso contrário, é atualizado. O Quadro 4 apresenta os comandos SQL de verificação, inserção e atualização dos fatos participação *chat*.

Quadro 4: Comandos SQL de verificação, inserção e atualização dos fatos participação *chat*

```
"select quantidade from participacao_chat where idsalachat = "+iIdSalaChat+" and
idtempo = "+iIdTempo+" and idtutor = "+iIdTutor+" and idconteudo = "
+iIdConteudo+" and idcurso = "+ iIdCurso +" and idaluno = "+iIdAluno

"insert into participacao_chat (quantidade, idsalachat, idtempo, idtutor, idconteudo,
idcurso, idaluno) values(1,"+iIdSalaChat+", "+iIdTempo+", "+iIdTutor+",
"+iIdConteudo+", "+iIdCurso+", "+iIdAluno+" )"

"update participacao_chat set quantidade = "+ (iQuantidade+1)+" where idsalachat =
"+iIdSalaChat+" and idtempo = "+iIdTempo+" and idtutor = "+iIdTutor+" and
idconteudo = " +iIdConteudo+" and idcurso = "+ iIdCurso +" and idaluno = "+iIdAluno
```

Realizadas a carga dos demais fatos, os dados agora integrados e consolidados são carregados para a base do *data warehouse*.

5.3 ÁREA DE APRESENTAÇÃO

A ferramenta Microsoft Analysis Management foi utilizada para os serviços de análise. Trata-se de uma ferramenta para processamento OLAP, que exporta relatórios para Excell, HTML e Power Point, além de ter funcionalidades para construção de modelos para *data mining*.

Por meio dessa ferramenta, são construídos cubos multidimensionais de dados, em que se setam fatos, dimensões e valores mensuráveis. As figuras 11 e 12 apresentam os modelos dos cubos Participação chat e Envio dúvida.

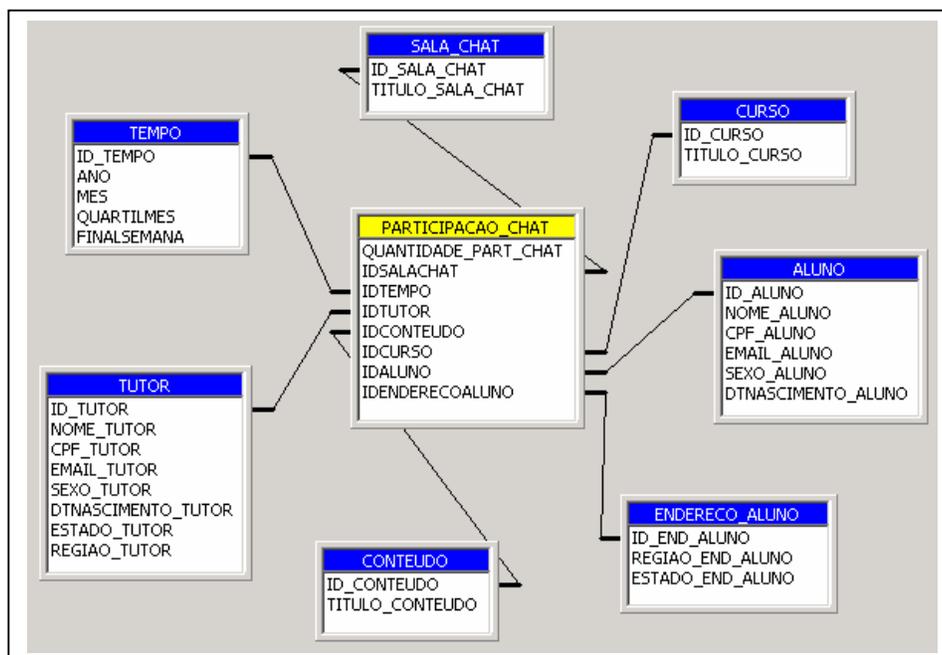


Figura 11: Modelo do cubo de dados Participação chat

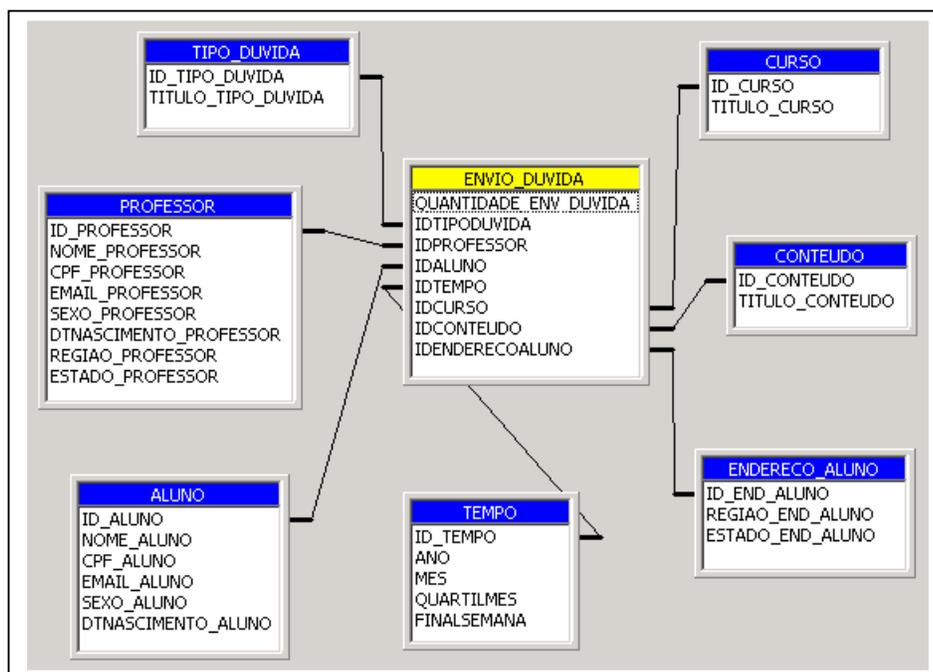


Figura 12: Modelo do cubo de dados Envio dúvida

Com a configuração para conexão com a base do *data warehouse*, que se encontra no Microsoft SQL 2000, o cubo é processado e disponibilizado para acesso.

A Figura 13 traz um exemplo de apresentação dos resultados, exploração dos dados, utilizando-se o Microsoft Analysis Service.



Figura 13: Exemplo de utilização da ferramenta Analysis Service

A ferramenta tem duas áreas distintas. Acima, encontra-se a área para seleção de dimensões e, abaixo, a área de apresentação dos dados. Nesse exemplo, visualiza-se o cubo Participação *chat*, com a dimensão tutor selecionada, apresentando a quantidade total de participações dos alunos e a quantidade agrupada pelo sexo dos tutores.

Mediante um duplo clique em um campo da área de apresentação, podem-se realizar operações de *drill up* ou *drill down*, aumentando-se ou diminuindo-se o nível de detalhamento da informação. Assim como a Figura 13, a Figura 14 apresenta a visualização do cubo Participação *chat* com a dimensão tutor selecionada, com a diferença de ter sido realizada uma operação de *drill down* na dimensão tutor, aumentando-se o nível de detalhamento, agrupando-se a participação não só pelo sexo dos tutores, mas também pelo nome de cada um.

Double-click a member to drill up or down.

		MeasuresLevel
- Sexo Tutor	Nome Tutor	Quantidade Part Chat
- F	Fabiola Vieira	1.886
	Isadora Montibeller	2.169
	Karina N. Figueira	3.116
	Márcia Machado	2.808
	Maria Augusta Tridapalli	2.514
	Raquel Costa	2.666
	Sheila Maria Neves Martins	2.579
	Tatiana Rodrigues Borges	1.762
	Walevska da S. Vieira	1.940
	M Total	42.384
- M	Adriano de Amarante	2.429
	Alessandro N. de Carvalho	2.515
	Bruno D'alessio Pereira	3.299
	César Augusto dos Santos	3.577
	Demerval Ribeiro Vianna Fi	2.150

Figura 14: Operação de drill down no cubo Participação chat

Na área de seleção, clicando-se em uma dimensão e arrastando-a para área de apresentação, acrescentam-se outras dimensões à área de apresentação, podendo-se não só incluir novas dimensões na área de apresentação como também alterar a ordem delas, mudando-se assim a orientação segundo a qual os dados são visualizados. Essa é uma operação de *slice and dice*, que é uma das principais características de uma ferramenta OLAP, ou seja, alterar a ordem de apresentação de maneira a facilitar a compreensão dos usuários.

A Figura 15 apresenta o cubo Participação chat com as dimensões tempo e tutor selecionadas, a informação, a participação dos alunos para cada tutor e sumariza também pelo sexo do tutor, mês e ano.

Cube Browser - Participação Chat

Conteúdo: All Conteúdo | Curso: All Curso
 Endereço Aluno: All Endereço Aluno | Sala Chat: All Sala Chat
 Aluno: All Aluno

- Ano	+ Mes	- Sexo Tutor	Nome Tutor	MeasuresLevel	
- 2002	+ 10	- M	Rodrigo Valverde da Silva	1.179	
			Rogério Enderle	748	
			Sidney T. Oliveira	1.259	
			Tiago N. V. Leitão	1.141	
			All Tutor	All Tutor Total	19.334
			+	Total	26
	+ 11	- F		F Total	7.928
			Ana Cláudia da Silva	732	
			Beatriz Lopez Pereira	512	
			Caroline Sartori Velloso	513	
			Débora Thais Klein Nunes	488	
			Fabiola Vieira	506	
			Isadora Montibeller	602	
			Karina N. Figueira	730	
			Márcia Machado	810	
			Maria Augusta Tridapalli	726	
		Raquel Costa	759		
		Sheila Maria Neves Martins	692		
		Tatiana Rodrigues Borges	357		

Double-click a member to drill up or down.

Close Help

Figura 15: Visualização do cubo Participação chat nas dimensões tempo e tutor

Na Figura 16, o cubo Participação chat pode ser observado por meio da dimensão endereço aluno, tendo sido realizada uma operação de *drill down* na região nordeste.

		MeasuresLevel
- Regiao End Aluno	Estado End Aluno	Quantidade Part Chat
All Endereço Aluno	All Endereço Aluno Total	72.900
+ CENTRO-OESTE	CENTRO-OESTE Total	14.914
	NORDESTE Total	9.413
	AL	852
	BA	2.286
	CE	1.267
- NORDESTE	MA	684
	PB	1.072
	PE	2.062
	PI	647
	RN	543
+ NORTE	NORTE Total	2.720
+ SUDESTE	SUDESTE Total	27.575
+ SUL	SUL Total	18.278

Figura 16: Visualização do cubo Participação chat na dimensão endereço aluno

A Figura 17 apresenta o cubo Envio dúvida nas dimensões tempo e aluno, resumindo a informação pelo ano, mês e sexo do aluno. Já, na Figura 18, o mesmo cubo é apresentado nas dimensões tipo dúvida e tempo.

Cube Browser - Envio Duvida

Conteúdo: All Conteúdo | Curso: All Curso
 Endereço Aluno: All Endereço Aluno | Professor: All Professor
 Tipo Duvida: All Tipo Duvida

			MeasuresLevel
- Ano	+ Mes	+ Sexo Aluno	Quantidade Env Duvida
All Tempo	All Tempo Total	+ M	395
- 2002	2002 Total	All Aluno	518
		+ F	123
		+ M	395
	+ 10	All Aluno	140
		+ F	41
		+ M	99
	+ 11	All Aluno	82
		+ F	21
		+ M	61
	+ 12	All Aluno	26
		+ F	4
			+ M

Double-click a member to drill up or down.

Close Help

Figura 17: Visualização do cubo Envio dúvida nas dimensões tempo e aluno

Cube Browser - Envio Duvida

Conteúdo: All Conteúdo | Curso: All Curso
 Endereço Aluno: All Endereço Aluno | Professor: All Professor
 Aluno: All Aluno

			MeasuresLevel
Titulo Tipo Duvida	- Ano	+ Mes	Quantidade Env Duvida
Conteúdo	- 2002	+ 10	120
		+ 11	70
		+ 12	19
		+ 8	59
		+ 9	128
		All Tempo	All Tempo Total
Técnica	- 2002	2002 Total	122
		+ 10	20
		+ 11	12
		+ 12	7
		+ 8	35
		+ 9	48

Double-click a member to drill up or down.

Close Help

Figura 18: Visualização do cubo Envio dúvida nas dimensões tipo dúvida e tempo

Outra maneira de refinar as consultas é aplicar filtros nas dimensões não selecionadas na área de apresentação. Por exemplo, na Figura 19, é apresentada a mesma visualização demonstrada na Figura 18, o cubo Envio de dúvida com as dimensões tipo dúvida e tempo selecionadas. A diferença é a filtragem dos dados na dimensão endereço aluno, sendo apresentadas somente as informações dos alunos que atendem à opção selecionada.

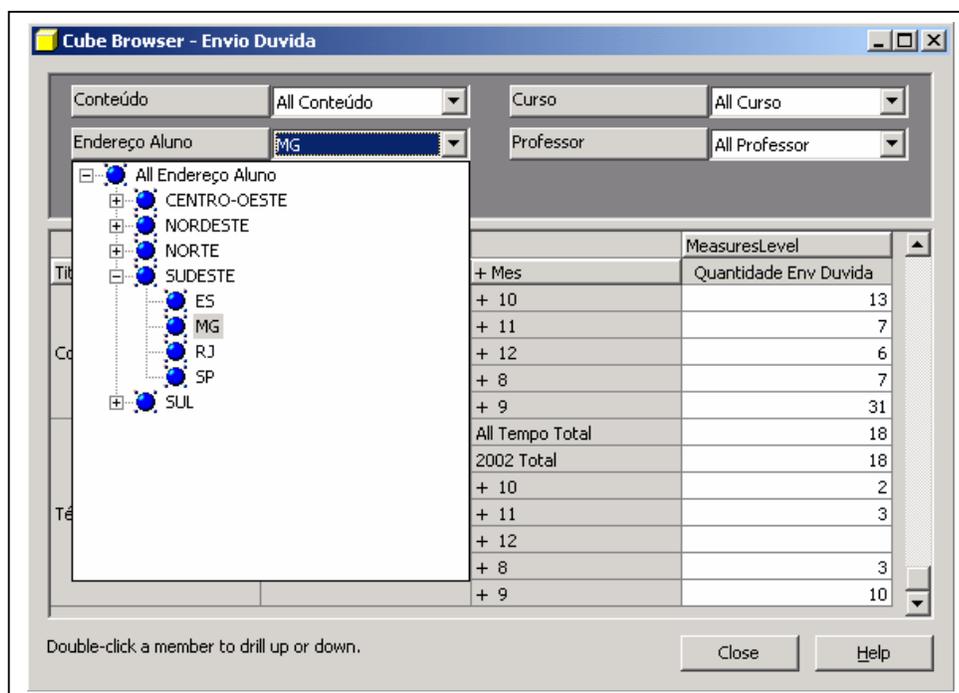


Figura 19: Filtro de visualização nas dimensões não selecionadas

A Figura 20 mostra o cubo Acesso sistema com as dimensões tempo e horário selecionadas, com os dados de acesso ao sistema pelos alunos sumarizados por ano, mês e horário de acesso, que representa períodos de seis em seis horas, com início à zero hora.

- Ano	+ Mes	+ Período	MeasuresLevel Quantidade Acesso
- 2002	+ 10	All Horário	15.702
		+ 1	224
		+ 2	4.445
		+ 3	5.664
		+ 4	5.369
	+ 11	All Horário	12.714
		+ 1	183
		+ 2	3.699
		+ 3	4.601
		+ 4	4.231
	+ 12	All Horário	9.152
		+ 1	154
+ 2		2.460	

Figura 20: Visualização do cubo Acesso sistema nas dimensões tempo e horário

Com as figuras de 13 a 20, procuraram-se exemplos de exploração dos dados, apresentando as maneiras possíveis de correlação das informações, que se referem à característica das ferramentas OLAP de permitir a visualização das informações de várias formas, conforme a necessidade de detalhamento. Assim, permite-se a exploração, a análise desestruturada, que, juntamente com a competência na análise da informação e a correta definição do cenário que a informação descreve, será o fator crítico de sucesso.

6 CONCLUSÕES E SUGESTÕES

Este capítulo apresenta as conclusões obtidas com o desenvolvimento do trabalho e sugestões para extensões.

6.1 CONCLUSÕES

Retornando ao objetivo geral deste trabalho – propor um modelo de aplicação de sistemas de informação que permita a alavancagem dos sistemas de gerência de aprendizagem (LMS), mediante a disponibilização de informações estratégicas sobre a logística do processo de aprendizagem –, conclui-se que ele foi alcançado. Com a aplicação do modelo em uma plataforma de *e-learning* de grande porte, como a plataforma VIAS-K, e em um curso de abrangência nacional, com 1.600 alunos, como o de Formação em Gestão Pública no Brasil, da Caixa Econômica Federal, a eficácia da arquitetura foi comprovada. Assim, também se atingiram os objetivos específicos:

- Adquirir conhecimento implícito a partir da correlação e análise desestruturada dos dados de utilização de sistemas de *e-learning*, visando a uma nova forma de avaliação do curso. Isso foi demonstrado na aplicação, pela utilização de uma ferramenta OLAP, que tem como principal característica permitir a visualização da informação de várias formas, mudando a orientação segundo a qual os dados são visualizados, conforme a necessidade de detalhamento;
- Incorporar outros atores na análise do *e-learning*, permitindo uma análise mais abrangente, uma vez que o conhecimento gerado pela análise dos dados depende do conhecimento e da visão de quem os recebe. O sistema não tem relatórios pré-formatados e direcionados a determinadas áreas, os dados estão disponíveis para quem tiver acesso explorá-los, de acordo com o interesse do explorador;
- Flexibilizar a tomada de decisão, disponibilizando as informações em tempo real, de forma mais dinâmica, não somente ao final dos cursos, por meio dos tradicionais formulários de avaliação. Apesar de o curso em questão estar terminado quando foi aplicado o modelo, essa característica evidencia-se na aplicação, bastando realizar-se as cargas de dados durante a execução dos cursos;

- Geração de grupos: Classificação dos alunos em ‘n’ classes, baseando-se em variáveis do interesse de quem estiver analisando os dados, possibilitando ações como colocar um grupo de alunos que não interagem com um tutor que estimula a interação e vice-versa. Apesar de não se ter implementado essa funcionalidade, seus conceitos foram introduzidos e sua aplicação apresentada no modelo, pretendendo-se realizar a implementação em seqüência à realização desta pesquisa.

6.2 SUGESTÕES PARA EXTENSÕES

Na seqüência deste estudo, pretende-se implementar um *front end* com ferramentas que facilitem a interpretação das informações, apresentando-se gráficos dinâmicos diretamente dos cubos OLAP pré-processados e implementando-se, também, a geração de grupos por meio do algoritmo de classificação supervisionado *k-means*, apresentado no modelo.

Outra possibilidade em análise é a construção de componentes baseados em técnicas de *data mining* para extrair conhecimento, novos padrões e relações não facilmente descobertos pela simples exploração dos dados.

6.3 CONSIDERAÇÕES FINAIS

Este trabalho apresenta-se como um sistema de apoio à decisão, para que os agentes envolvidos nos processos do *e-learning* possam balizar suas ações. Assim, como qualquer sistema de apoio à decisão, a interpretação dele está sujeita ao tomador de decisão, em sua competência na montagem de cenários e análise da informação.

Mesmo havendo o fato de a base de dados do curso selecionado apresentar alguns problemas de falta de informação, devido à utilização de controles externos à plataforma de ensino – o que, por um lado acabou sendo vantajoso para o trabalho, uma vez que a utilização de dados externos comprovou a idéia proposta de generalização do modelo –, a aplicação nesse contexto mostrou-se válida, principalmente, pela abrangência do curso, 1.600 alunos espalhados por todas regiões e estados brasileiros.

REFERÊNCIAS

- ALTER, Steven. **Information systems: a management perspective**. USA: Addison Publishing Company, 1992.
- BATISTA, Paulo; SILVA, Mário J. **Prospecção dos Dados de Acesso a um Servidor de Notícias na Web**. Universidade de Lisboa. Lisboa. Disponível em: <<http://www.fccn.pt/crc1999/FINAIS/artigo28/ARTIGO28.pdf>>. Acesso em: 12 jan. 2004.
- BERNTHAL, Paul. **The state of e-learning: Developing soft skills**. Disponível em: <http://www.ddiworld.com/pdf/ddi_stateofe-learning_rr.pdf>. Acesso em: 10 jun. 2003.
- BOFF, Elisa. **Ambiente para Construção Cooperativa de História em Quadrinhos**. 2000. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciências da Computação, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre.
- BRENNAN, Michael. **Learning management systems**. Disponível em: <<http://www.elearningbrokers.org/lms/index.htm>>. Acesso em: 10 jun. 2003.
- CARVALHO, Luís Alfredo Vidal de. **Data mining: a mineração de dados no marketing, medicina, economia, engenharia e administração**. São Paulo: Érica, 2001.
- CIELO, Ivã Rafael; PAZ, Luiz Cláudio. **Arquiteturas OLAP**. Brasília, dez. 2000. Disponível em: <<http://www.datawarehouse.inf.br/>>. Acesso em: 27 ago. 2001.
- COELHO, Paulo Sérgio de Souza; EBECKEN, Nelson F. F. **Segmentação de dados em um número desconhecido de grupos usando algoritmos genéticos**. Universidade Federal do Rio de Janeiro. Rio de Janeiro. Disponível em: <<http://www.sobrapo.org.br/simposios/XXXIII/artigos/130-ST295.pdf>>. Acesso em: 12 jan. 2004.
- COSTA, Fernando Albuquerque; PERALTA, Maria Helena. *e-learning: Formação de formadores para a construção de contextos de aprendizagem significativa*. In: A. Estrela & J. Ferreira (Eds.). **Tecnologias em educação: Estudos e investigações**. Faculdade de Psicologia e de Ciências da Educação, Universidade de Lisboa. Lisboa: AFIRSE, 2001.
- CRUZ, Tadeu. **Sistemas de Informações Gerenciais**. São Paulo: Atlas, 1996.
- DALFOVO, Oscar; AMORIM, Sammy Newton. **Quem tem informação é mais competitivo**. Blumenau: Acadêmica, 2000.
- DALMAU, Marcos Baptista Lopes *et al.* A Educação Profissional, a EAD e as Universidades Corporativas: Um Mercado Emergente. **Associação Brasileira de Educação a Distância**. 2002. Disponível em: <<http://www.abed.org.br>>. Acesso em: 20 nov. 2002.
- DEAN, Christopher. **Technology based training & on-line learning: An overview of authoring systems and learning management systems available in the UK**. 2002. Disponível em: <<http://www.peak.co.uk/AuthoringSystem.pdf>>. Acesso em: 6 jun. 2003.

EDUTECH. **Evaluation of Learning Management Systems**. Disponível em: <<http://www.edutech.ch/edutech/tools/ev2.php>>. Acesso em: 8 jul. 2003.

FRANCO, Marcelo A. **A Magnitude da Informação Digital**. 1998. Disponível em: <<http://www.revista.unicamp.br/revista/infotec/educacao/educacao.html>>. Acesso em: 15 mai. 2003.

FREITAS, Henrique. e POZZEBOM, Marlei. **Características Desejáveis de um EIS**. Enterprise Information System. Rumo a Proatividade. PPGA. Universidade Federal do Rio Grande do Sul. Disponível em: <<http://read.adm.ufrgs.br/read05.artigo.eis.htm>>. Acesso em: 10 nov. 2003.

FURLAN, José Davi. **Modelagem de objetos através da UML**. São Paulo: Makron Books, 1998.

GREENBERG, Leonard. **LMS and LCMS: What's the difference?**. 2002. Disponível em: <<http://www.learningcircuits.org/2002/dec2002/greenberg.htm>>. Acesso em: 13 mai. 2003.

HALL, John. **Assessing Learning Management Systems**. 2003. Disponível em: <http://www.clomedia.com/content/templates/clo_feature.asp?articleid=91&zoneid=29>. Acesso em: 7 jun. 2003>.

HALMENSCHLAGER, Carine; VALIATI, João Francisco. **Aplicação da descoberta de conhecimento em dados de Mortalidade do RS**. Universidade Federal do Rio Grande do Sul. Porto Alegre. Disponível em: <<http://www.inf.ufrgs.br/~carineh/ArtigoSIM.pdf>>. Acesso em: 12 jan. 2004.

IEEE – Institute of Electrical and Electronic Engineers. Disponível em: <<http://www.ieee.org>>. Acesso em: jun.2003.

IMS – Instructional Management Systems. Disponível em: < <http://www.imsproject.org/>>. Acesso em: jun.2003.

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

INMON, William H.; WELCH, J. D.; GLASSEY, Katherine L. **Gerenciando data warehouse**. São Paulo: Makron Books, 1999.

INSTITUTO GARTNER. Disponível em: <<http://www4.gartner.com/Init>>. Acesso em: jan. 2003.

KIMBALL, Ralph; ROSS, Margy. **The data warehouse toolkit**. Rio de Janeiro: Campus, 2002. Tradução de Ana Beatriz Tavares e Daniela Lacerda.

KIMBALL, Ralph; MERZ, Richard. **Data webhouse**. Rio de Janeiro: Campus, 2000. Tradução de Edson Furmankiewicz e Joana Figueiredo.

LANDIM, Cláudia Maria M. P. Ferreira. **Educação a distância**: algumas considerações. Rio de Janeiro: [s.n], 1997.

LAUDON, K. e LAUDON, J. **Management Information Systems-Organization and Technology**. Macmillan Publishing Company, EUA, 1996, 818p.

LONGMIRE, W. **A Primer On Learning Objects**. American Society for Training & Development, Virginia, USA, 2001.

MECHELN, Pedro José von. **SAIIP1-GI - Sistema de Apoio ao Planejamento no Processo de Tomada de Decisão do Jogo de Empresas GI-EPS**. 1997. Dissertação (Mestrado em Engenharia de Produção) – Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis.

OLIVEIRA, Adelize Generini de. **Data warehouse: conceitos e soluções**. Florianópolis: Advanced, 1998.

OLIVEIRA, Djalma de Pinho Rebouças. **Sistemas de informações gerenciais**. São Paulo: Atlas, 1992.

QUONIAM, Luc; TARAPANOFF, Kira; ARAÚJO, Rogério Henrique de; ALVARES, Lillian. **Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil**. Disponível em: <<http://www.ibict.br/cionline/300201/3020104.pdf>>. Acesso em: 12 jan. 2004.

ROSSETT, Allison. **The ASTD e-learning Handbook: Best Practices, Strategies, and Case Studies for a Emerging Field**. New York: McGraw-Hill, 2001. 500p.

SINGH, H. **Introduction to learning objects**. 2001. Disponível em: <<http://www.elearningforum.com/july2001/singh.ppt>>. Acesso em: 17 jun. 2003.

WEBSCHOOL. **Educação a Distância**. Vantagens. Disponível em: <http://www.webschool.com.br/ead_vantagens.php3>. Acesso em: 07 jan. 2003.