

PROJETO BI-UFRGS: do modelo de dados corporativo ao data warehouse.

Dr. Hubert Ahlert

Centro de Processamento de Dados – Universidade Federal do Rio Grande do Sul

Hubert@cpd.ufrgs.br

Resumo

Este trabalho visa mostrar a trajetória da UFRGS na construção de seu data warehouse a partir da adaptação de uma metodologia anteriormente proposta e utilizada para a modelagem de dados corporativa da Universidade. São explorados os aspectos da adaptação da metodologia para permitir uma modelagem multidimensional e os resultados iniciais do uso da tecnologia do BI na construção dos sistemas de informações gerenciais.

Palavras-chave: Administração de Dados, Modelagem, Data Warehouse, Business Intelligence

1. Introdução

Desde a modernização dos sistemas de informações da UFRGS, que começou a ser projetada em julho de 1995, o Centro de Processamento de Dados da Universidade vem trabalhando na integração de seus sistemas através de um modelo de dados corporativo.

Em 1995 foi adquirido um acervo de software para garantir a instalação da plataforma cliente-servidor como arquitetura computacional básica dos sistemas mantidos pelo CPD para fins administrativos e acadêmicos da Universidade. Esse acervo estava constituído por produtos de banco de dados, ferramenta CASE para modelagem dos sistemas e ferramentas de desenvolvimento.

A migração dos sistemas legados, residentes em mainframe, para a nova plataforma e também a demanda pela construção de novos aplicativos revelou a necessidade de investir em uma estratégia de análise e projeto de sistemas que considerasse um levantamento das necessidades do usuário de forma rápida e com a participação ativa da comunidade afetada pelo projeto. Para isto, foi desenvolvida uma metodologia [CPD 97] que combinou a idéia de análise e planejamento da informação (modelagem corporativa) pregada pela engenharia de informações ([KIP 93], [FEL 88]) e os conceitos de modelagem de dados [CHE 90] com a técnica JAD de condução de reuniões ([COS 94], [MOR 91]). Junto com a metodologia surgiu a necessidade de reunir um grupo de profissionais responsáveis pela difusão e aplicação dessa metodologia e pelo gerenciamento dos efeitos que ela trazia em relação ao desenvolvimento dos sistemas de informação. Na época foi instalado, no CPD, o grupo de Administração de Dados da Universidade para tratar desse assunto.

A metodologia proposta exigia uma participação efetiva do usuário em todo o processo de modelagem. A participação desse usuário se baseava na formação de um grupo de trabalho no qual havia uma ampla interação entre os responsáveis por dados e rotinas da Universidade e os técnicos em informática.

Aproveitando os conhecimentos obtidos a partir da adoção da metodologia de modelagem de dados [CPD 97] desenvolvida no âmbito da Administração de Dados do CPD, o processo de modelagem foi adaptado para adequá-lo a um paradigma de modelagem de informações gerenciais baseada na tecnologia Data Warehouse ([INM 94], [KIM 96],[CHA 97]). Se por um lado o processo de obtenção do data warehouse seria facilitado pela extração dos dados de uma base única que implementa um modelo de dados corporativo, a própria metodologia de modelagem contribuiria para que o usuário, agora em âmbito gerencial, tivesse uma participação efetiva na construção dos sistemas de informações gerenciais.

O presente artigo procura mostrar esse processo de construção do data warehouse da UFRGS e como a tecnologia do BI, Business Intelligence, ([HAR 96], [WEL 95], [WEL 96]) foi

naturalmente introduzida em um passo que sucedeu o processo de obtenção do modelo de dados corporativo da Universidade.

2. Obtenção do modelo corporativo

Na instalação da plataforma cliente-servidor a primeira questão discutida pela equipe técnica do CPD foi como agilizar a prestação de serviços para o usuário e como tornar mais rápido o processo de construção dos sistemas de informação. Concluiu-se que uma das atividades da análise e projeto que consumia muito tempo no cronograma de desenvolvimento era justamente o contato com os usuários para definir os requisitos do sistema. As entrevistas individuais com todos os usuários envolvidos de alguma forma com a área de negócios do sistema representavam um enorme tempo no ciclo de desenvolvimento.

Em função disso foi proposta uma metodologia de modelagem de dados [CPD 97] que incorporou as técnicas JAD (Joint Application Design) ([COS 94] ,[MOR 91]) de condução de reuniões. Com essa metodologia pretendia-se substituir as entrevistas tradicionais, realizadas individualmente com cada usuário, por sessões de trabalho que reuniam representantes do pessoal que, de alguma forma, estavam envolvidos com a área de negócios sendo investigada. Esse grupo de trabalho (GT) discutia o assunto e, por consenso, definia o modelo. O critério de escolha dos participantes do grupo de trabalho baseou-se no conhecimento que cada integrante tinha da estrutura da Universidade em relação à área de negócios que estava sendo modelada. Os representantes dos usuários foram selecionados em função da importância das tarefas que realizavam junto a Pró-reitorias e Unidades. Na área de informática, o grupo contou com técnicos das áreas de administração de dados e de desenvolvimento de sistemas de informação. A composição formal de um GT era a seguinte:

- Um gerente do projeto (responsável pela organização dos trabalhos e registro dos resultados das reuniões – documentação da modelagem)
- Um mediador (guia imparcial das reuniões, para mediar conflitos e garantir consenso nas decisões)
- Representantes dos usuários
- Técnicos de informática (administração de dados e analista de negócio)

A metodologia utilizada prevê a realização de reuniões formais do grupo de trabalho para análise dos dados e conseqüente análise dos requisitos dos sistemas de informação, com o objetivo de buscar decisões de consenso para a construção do modelo de dados (objetivo principal) e do modelo de funções (objetivo complementar). Além dos integrantes do GT, as sessões de trabalho poderiam eventualmente contar com a presença de um especialista no assunto sendo discutido com vistas a um aprofundamento de detalhes técnicos ou operacionais.

Os trabalhos de modelagem do GT sempre iniciavam através de um levantamento das principais atividades realizadas dentro da área de negócios sendo investigada com o intuito de identificar as funções ligadas ao assunto. Isto gerava um Diagrama Hierárquico de Funções (DHF). A partir do DHF era estabelecida uma ordem de prioridade com a intenção de adotar uma abordagem sistemática de análise das funções pela sua importância e, assim, estabelecer um roteiro para a modelagem de dados. Cada função era analisada sob o enfoque dos dados por ela sendo manipulados. Desta forma realizava-se o levantamento dos dados ligados a cada função e, paulatinamente, era construído um Diagrama Entidade-Relacionamento (ER) que ilustrasse as entidades e relacionamentos representativos do modelo de dados associado a função. A cada nova função sendo investigada, incrementava-se o modelo de dados através da adição de novos atributos às entidades já modeladas anteriormente ou através da adição de novas entidades e relacionamentos necessários para contemplar a função em estudo.

Concluído e revisado o modelo de dados relativo a área de negócios, o GT complementava o trabalho através de uma análise das responsabilidades pelos dados que estavam representados no

modelo com o objetivo de estabelecer os direitos de acesso e a responsabilidade pela sua atualização.

Para os sistemas legados do mainframe, que atendiam a uma parcela da área de negócios sendo investigada, foi realizada uma engenharia reversa para obter o modelo de dados correspondente. Cabia a equipe de administração de dados, com o auxílio do analista de sistemas responsável pelo sistema legado, realizar uma compatibilização desse modelo para adequá-lo ao modelo de dados gerado durante as reuniões de modelagem do GT.

Cada GT que concluía seus trabalhos gerava como resultado um diagrama que posteriormente deveria ser ajustado para integrar o modelo de dados corporativo da Universidade.

Considerando que a integração imediata de todos os diagramas se mostrou muito complexa para ser conseguida em um único passo, optou-se pela integração progressiva a partir das áreas de interesse da Universidade. Assim, considerou-se modelos parciais para:

- a) Área Acadêmica – como atividade fim da Universidade ela engloba graduação, pós-graduação, pesquisa, extensão e biblioteca
- b) Área Administrativa – como atividade meio da Universidade ela engloba recursos humanos, patrimônio, almoxarifado, protocolo etc

O modelo corporativo se tornou essencial para a Universidade pois permitiu uma posterior integração de sistemas e evitou que dados fossem armazenados de forma redundante, dispersos em diferentes sistemas. O modelo corporativo não impede uma posterior distribuição dos dados, em diferentes locais e máquinas, através dos mecanismos de bancos de dados distribuídos. Pelo contrário, conhecendo-se todos os dados manipulados na empresa é mais fácil identificar os pontos de melhor distribuição por responsabilidades de atualização.

No modelo corporativo gerado duas entidades se mostraram fundamentais para a integração de visões parciais do modelo e facilitar o processo de integração entre os diferentes sistemas de informação. São as entidades PESSOA e ÓRGÃOS. No caso da Universidade, os diferentes perfis da comunidade, independente de seu vínculo com a UFRGS (alunos, técnico-administrativos e docentes), tem dados pessoais comuns que são representados como atributos da entidade PESSOA. Da mesma forma, toda estrutura organizacional da Universidade é utilizada e compartilhada por vários sistemas de informação e o seu registro se dá através de uma entidade que foi denominada ÓRGÃOS.

3. Resultados da modelagem

Em meados de 1997 foi realizada a modelagem de duas áreas de negócio iniciais da Universidade: Área de Recursos Humanos e Área de Ensino e Pesquisa. Os trabalhos de modelagem dessas duas áreas permitiu validar as idéias incorporadas na metodologia proposta e consolidar a versão inicial do modelo de dados corporativo da UFRGS. Foram criados diagramas ER num total de 129 entidades, 192 relacionamentos e 1010 atributos para a área de Recursos Humanos e um total de 200 entidades, 249 relacionamentos e 515 atributos para a área de Ensino e Pesquisa.

A partir dessa versão inicial do modelo de dados corporativo, foi gerado o banco de dados correspondente e os dados da área de Recursos Humanos começaram a ser carregados das fitas espelho do sistema SIAPE (sistema de pagamento central da União mantido pelo SERPRO). Isto permitiu o nascimento do banco de dados da UFRGS na nova plataforma.

Em 1998 foram desenvolvidas as primeiras versões de aplicativos cliente-servidor para a área de Recursos Humanos e a área de Pesquisa. Nessa mesma época aplicações para a Internet foram liberadas para estas duas áreas.

Na área do controle acadêmico de graduação foi dada uma solução de contorno para disponibilizar rapidamente os dados dos alunos de graduação na Internet através de uma “clonagem” do banco de dados (DMSII) do mainframe para correspondente imagem em um banco de dados relacional. Foi adquirido um software denominado DataBridge para permitir essa

replicação de dados em um banco de dados SYBASE. Um novo sistema de graduação, já inteiramente na plataforma cliente-servidor e segundo o modelo de dados corporativo construído em 1997, foi implantado em 2004.

Na segunda metade de 1998, a Pró-Reitoria de Planejamento (PROPLAN) também foi modelada e incorporada ao modelo de dados corporativo. Hoje já existem diversos módulos do sistema de informação da PROPLAN implantados segundo esse modelo de dados.

No segundo trimestre de 2001 foi realizado um trabalho de modelagem da área de extensão que produziu 135 entidades, 137 relacionamentos e 770 atributos. Essa área também está incorporada no modelo corporativo e aplicações operacionais para a área também já foram desenvolvidas através de formulários da Internet.

Para o sistema da Bibliotecas, que mantém seu repositório em um banco de dados ORACLE, foi criada um mecanismo de carga periódica de dados do banco corporativo da UFRGS que sincroniza a base local com o banco de dados da Universidade. Este mecanismo de sincronização foi uma solução adotada porque o sistema é mantido por um software adquirido de terceiros (Sistema Aleph da ExLibris).

Hoje tem outros sistemas, como o de acesso ao Restaurante Universitário (RU), o de acesso a Portarias de Unidades, o de Ensino à Distância (EAD) e o de Serviços de Diretórios (LDAP), que utilizam semelhante mecanismos de sincronização para compatibilizar suas bases locais com o banco de dados corporativo da UFRGS. Todos esses mecanismos de sincronização estão baseados em um arquivo de “LOG” que registra alterações específicas no banco da UFRGS, detectadas através de “triggers” do banco associadas a estruturas de dados (tabelas) que precisam ser sincronizadas.

Oito anos após a conclusão dos trabalhos realizados pelos GTs de modelagem, o modelo de dados corporativo construído na oportunidade continua sendo bastante consistente. Novos sistemas de informação atualmente em desenvolvimento utilizam o modelo proposto sem a necessidade de grandes esforços de modelagem. Eventuais pequenos ajustes tornam-se necessários, como a inclusão de algum atributo novo aqui ou a inclusão de uma nova entidade ali, mas sem grandes mudanças no modelo original.

4. A experiência da modelagem corporativa aplicada no BI

Com a experiência acumulada durante a obtenção do modelo de dados corporativo da Universidade foi possível pensar em uma adaptação da metodologia, originalmente proposta para a modelagem tradicional de dados (que considera a representação do dado na forma como será implantada no banco), para uma versão que permitisse a modelagem multidimensional (que considera a representação da informação na forma como o usuário a utiliza).

Já que a modelagem tradicional de dados permitiu a construção do banco de dados da Universidade, a adaptação dessa modelagem, incorporando a tecnologia BI no processo, certamente permitiria a construção do data warehouse e, por conseguinte, os cubos para os sistemas gerenciais.

A primeira providência tomada foi procurar uma forma participativa do usuário na construção do modelo multidimensional de maneira semelhante a atuação dos grupos de trabalhos (GTs) durante a obtenção do modelo de dados corporativo. A diferença entre os dois processos de modelagem estava no fato de que, durante a obtenção do modelo de dados, estava sendo avaliada a necessidade de dados para cada atividade (função) analisada, onde pessoas de nível operacional traziam a forma como os dados eram manipulados em seus postos de trabalho. Já durante a obtenção do modelo multidimensional seria avaliada a necessidade de informações para os gestores da Universidade, onde pessoas de nível gerencial traziam a forma como as informações seriam interpretadas para produzir indicadores de gestão, consolidando-os na inteligência do negócio (BI).

Como representantes de usuários no GT, as pessoas de nível operacional deveriam ser substituídas por pessoas de nível estratégico/gerencial.

Como roteiro de modelagem, a análise de funções e atividades executadas deveria ser substituída pela análise dos indicadores de gestão.

Feitas essas adaptações preliminares, foi nomeada uma comissão para definir o modelo multidimensional que daria suporte aos sistemas gerenciais da Universidade. O grupo estava composto por representantes de gestores das áreas de ensino, recursos humanos, avaliação institucional e planejamento. Como representantes da área de informática, estavam no grupo pessoas da administração de dados e analistas de negócio que atuavam nas áreas específicas (data marts) sendo exploradas durante as reuniões.

A cada novo tema (fato) sendo discutido nas reuniões de modelagem foi apresentado para o grupo um fragmento do modelo de dados corporativo que ilustrava os prováveis dados que originariam os indicadores relacionados ao tema em estudo. O modelo de dados serviu de fonte de investigação dos atributos que compõem o modelo mutlidimensional (CUBO).

Dentre os diversos atributos de entidades do modelo de dados foram identificados aqueles que dão origem a dimensões do cubo (utilizadas para permitir a filtragem de dados durante as consultas ao cubo) e aqueles que dão origem as medidas (utilizadas para somatórios, médias, percentuais ou outros cálculos sobre os valores no agrupamento ou cruzamento das diferentes dimensões selecionadas no cubo durante as consultas).

A experiência com as reuniões de modelagem tem comprovado que o modelo de dados corporativo permite um fácil mapeamento entre os atributos de entidades do modelo de dados e as dimensões do modelo multidimensional.

Em alguns casos o mapeamento é direto. O exemplo abaixo mostra isto.

ATRIBUTO DA ENTIDADE	DIMENSÃO DO CUBO
PESSOA.Sexo	Sexo
PESSOA.Corpele	Corpele

Em outros casos o mapeamento é resultado de uma função de conversão como aparece no exemplo a seguir.

ATRIBUTO DA ENTIDADE	FUNÇÃO DE CONVERSÃO	DIMENSÃO DO CUBO
PESSOA.DataNascimento	Calcula_Idade(d)	Idade
PESSOA.DataNascimento	Calcula_FaixaEtaria(d)	FaixaEtária

A função Calcula_Idade(d), por exemplo, pode facilmente ser expressa em SQL como ilustrado a seguir:

```
select      convert(int,convert(float,
                DATEDIFF(dd, PESSOA.DataNascimento, GETDATE()))/365.25)
from PESSOA where .....
```

A figura 1 ilustra um exemplo dos mapeamentos diretos e dos realizados por funções de conversão.

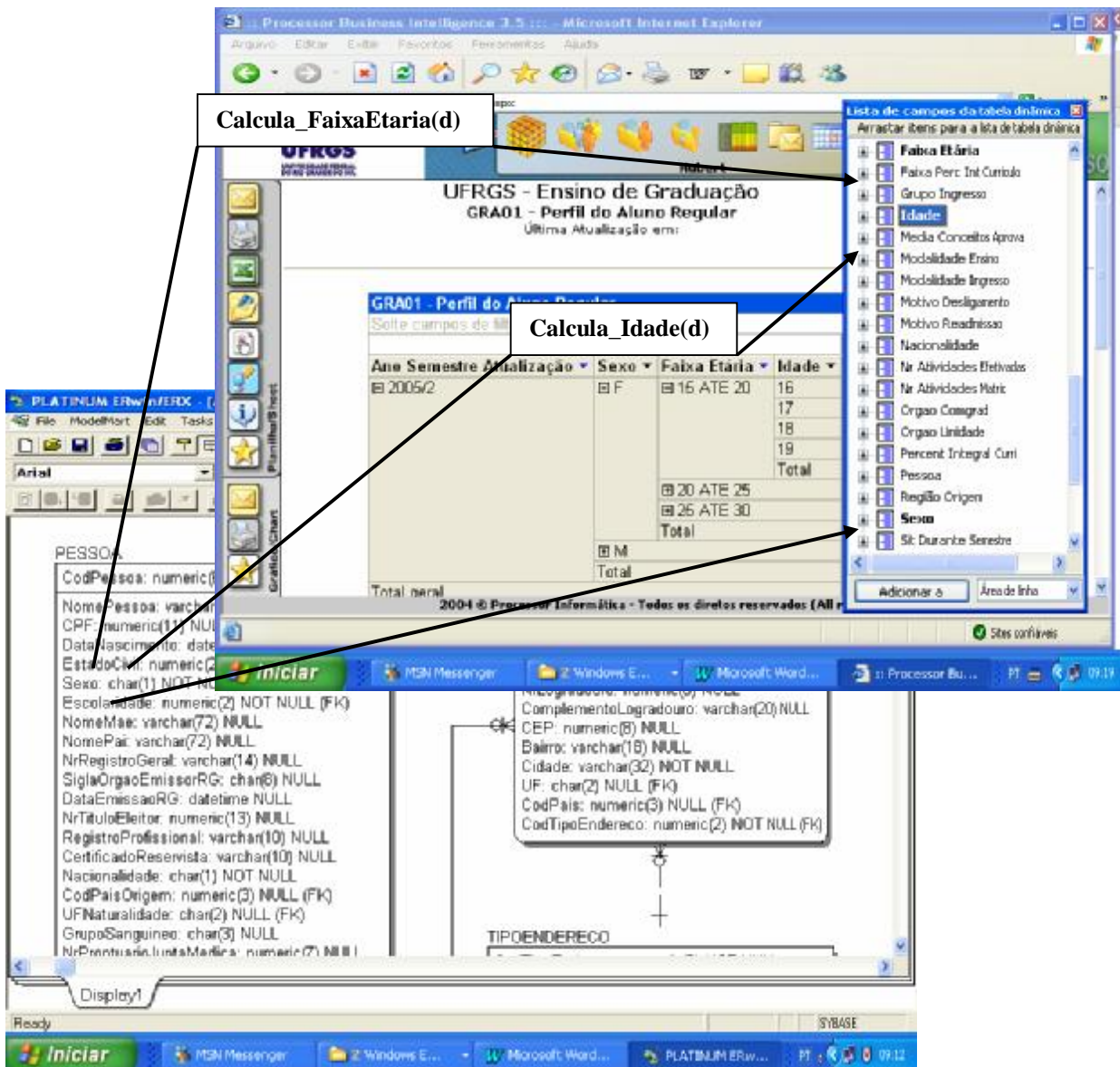


Figura 1. Mapeamento de atributos do Diagrama ER para dimensões do CUBO

Da mesma forma como o modelo de dados foi finalizado pelo GT através da análise e determinação das responsabilidades pela atualização dos dados e, por conseguinte, da configuração dos direitos de acesso, o modelo multidimensional deve ser finalizado determinando-se quais gestores poderão manipular, com ou sem restrições, o cubo construído a partir do modelo.

Já que o data warehouse mantém uma base temporal de dados extraídos do ambiente de produção da empresa, torna-se fundamental estabelecer também, no final do processo de modelagem de cada cubo, a periodicidade das extrações de dados (momentos de coleta) e a identificação das estruturas (tabelas) do banco de dados operacional que deverão alimentar as estruturas (tabela fato e tabelas dimensões) do data warehouse.

5. Os primeiros resultados do BI

O grupo de modelagem BI escolheu a área de recursos humanos para iniciar a análise de indicadores de gestão. Da mesma forma como havia sido realizado na modelagem para obter o banco de dados corporativo, a área de recursos humanos foi escolhida para iniciar todo o processo de modelagem por ser considerada a área central a partir da qual as pessoas são distribuídas para as demais áreas da empresa. A gestão de pessoas na empresa sempre, de alguma forma, passa pelo RH.

Como tema inicial do processo de modelagem foi escolhido o “PERFIL FUNCIONAL DO SERVIDOR”. Sendo assim, os indicadores de gestão relacionados ao servidor da UFRGS, utilizados pela Pró-reitoria de Recursos Humanos e pela Secretaria de Avaliação Institucional da Universidade, serviram de roteiro de modelagem.

Ao grupo de modelagem BI foi apresentado o fragmento do modelo de dados corporativo que representa os dados manipulados pelo sistema de informação utilizado pela Pró-reitoria de Recursos Humanos e, a partir de indicadores de gestão de pessoal sendo avaliados, atributos das entidades do modelo foram analisados e identificados aqueles que seriam fonte de informação para compor os indicadores. Do conjunto de atributos identificados no modelo foram determinadas as dimensões e medidas do cubo que implementariam o tema em discussão.

Para o cubo “PERFIL FUNCIONAL DO SERVIDOR” foram definidas dimensões de filtragem de consultas como Sexo, Idade, CorPele e TempoServiçoUFRGS, em um total de 28 dimensões. O atributo TempoServiçoUFRGS, calculado a partir da data de ingresso do servidor na Universidade e a data da extração dos dados do banco de dados operacional para o data warehouse, foi utilizado como dimensão para permitir filtros do tipo “seleção de servidores com mais de 25 anos de serviço” e também gerou uma medida (MédiaTempoServiçoUFRGS) para determinar, por exemplo, a média de tempo de serviço dos servidores em uma determinada Unidade Acadêmica ou em toda a Universidade. Atributos como NomePessoa, Email e RamalUFRGS, que normalmente não aparecem como informações gerenciais, foram incorporados a dimensão Pessoa do cubo para permitir que, a partir do resultado exibido de uma consulta específica ao cubo, fosse possível rastrear um servidor (pessoa) que integrasse esse resultado.

O segundo tema explorado pelo grupo de modelagem foi o “PERFIL DO CANDIDATO DO VESTIBULAR”. Desta vez a justificativa da escolha foi a de traçar o perfil da clientela da Universidade que ingressa pelo vestibular e que dá origem ao aluno de graduação. Nesse tema, indicadores como “percentual de candidatos, estratificados por sexo, que ingressaram na UFRGS nos últimos 5 anos” serviram como roteiro de modelagem. Foram criadas 40 dimensões e 10 medidas para esse cubo.

Dimensões que são oriundas de características de dados pessoais como Sexo, Idade e CorPele foram comuns aos dois temas iniciais estudados. Isto reforçou a nossa convicção anterior, estabelecida durante processo de modelagem de dados, que uma entidade PESSOA em um modelo é sempre um objeto de integração entre visões distintas do modelo.

Concluída a modelagem dos dois temas, foi realizada a implantação dos dois cubos como protótipos para validar toda a idéia do data warehouse e da tecnologia BI que estavam sendo introduzidos na UFRGS.

Como sistema de gerência de banco de dados para hospedar o data warehouse foi escolhido o MS SQL Server. Justificou a escolha o fato que todas as ferramentas que acompanham o banco permitem também dar suporte a tecnologia do BI. Ferramentas como DTS (Data Transformation Services), para extração de dados, e Analysis Services, para construir os cubos, estão incorporadas ao pacote do MS SQL Server.

6. Trabalhos futuros

Concluída a implantação inicial do data warehouse, pretende-se que a administração de dados, com o apoio do grupo de modelagem BI, documente todo esse processo e crie uma política de uso da ferramenta de consulta ao cubo. Será imprescindível um bom treinamento dos usuários da ferramenta (gestores) para que informações gerenciais obtidas das consultas sejam resultantes de

filtros consistentes (seleção correta dos itens das dimensões envolvidas na consulta) que não distorçam as medidas exibidas ou se traduzam em interpretações errôneas dos dados manipulados no cubo.

Após consolidados os procedimentos de extração de dados e de geração do data warehouse e também desenvolvidas algumas aplicações gerenciais que justificarão todo investimento e esforços realizados, o passo seguinte é avançar na tecnologia BI para incorporar mecanismos de mineração de dados (data mining ([BRA 96], [FAY 96], [KIM 98])).

Queremos, por exemplo, comprovar uma antiga constatação de um velho e experiente professor de cálculo da Universidade. Em turmas de disciplinas compartilhadas pelos cursos de matemática e engenharia ele sempre costumava dizer ironicamente para alunas que não alcançavam bons conceitos: "... as meninas da matemática só vem às minhas aulas para procurar marido engenheiro!". Queremos verificar se havia alguma verdade na afirmação do velho mestre. Talvez a mineração de dados nos permita realmente comprovar essa tese do professor de cálculo. O futuro dirá se havia um padrão consistente entre as notas (negativas) das alunas do curso de matemática e um relacionamento (conjugal) com alunos da engenharia.

7. Conclusões

Desde a época em que iniciou a migração de seus sistemas legados do mainframe para a plataforma cliente-servidor, o CPD da UFRGS vem apostando no modelo de dados corporativo como uma forma de facilmente conseguir uma integração entre os sistemas de informação da Universidade. Na época o que justificou a construção do modelo foi justamente a busca de uma maneira de conhecer todas as áreas de negócio da Universidade através dos dados que elas necessitam e manipulam.

Um fator importante que alavancou a construção do modelo corporativo foi a adoção de uma metodologia que agilizasse o levantamento dos requisitos dos sistemas de informação através de GTs de modelagem que, por consenso entre usuários e técnicos de informática, procuraram determinar os dados necessários a cada atividade vinculada a área de negócio. O presente artigo mostrou esta metodologia como uma adaptação da modelagem corporativa sugerida na literatura sobre Engenharia da Informação, incorporando a técnica JAD para condução de reuniões em grupo. Usou-se uma combinação de idéias para compor uma metodologia modificada.

Existe uma velha lei que diz: "... na natureza nada se cria, tudo se transforma." (Lei de Lavoisier). Esta lei certamente poderia ser aplicável para a informática, mais especificamente na área de Engenharia de Software. É tudo uma questão de modificar, adaptar e evoluir. Um passo de cada vez. O exemplo trazido neste artigo vai ao encontro dessa constatação. A metodologia utilizada para a modelagem de dados corporativa na UFRGS, que já havia sido modificada anteriormente a partir da proposta original da Engenharia de Informação, foi agora submetida a adaptações para contemplar um levantamento de informações gerenciais. A modelagem corporativa utilizava, como roteiro de modelagem, o conjunto de atividades das áreas operacionais da empresa para gerar um modelo de dados. Já a modelagem multidimensional utiliza o conjunto de indicadores das áreas estratégicas para gerar um modelo multidimensional (Cubo) que permite aos gestores visões múltiplas sobre os dados que darão origem às informações gerenciais. O que tem de comum nos dois processos de modelagem é a forma participativa do usuário interagir com o time de informática na construção dos modelos.

8. Bibliografia

- [AHL 98] AHLERT, H.: Avaliação do Processo de Implantação da Administração de Dados no CPD da UFRGS, Porto Alegre, Escola de Administração /UFRGS, jul 1998, (trabalho de conclusão de curso)

- [ANU 93] ANUNI. Administração de dados: em busca da eficiência, Publicação do 39º ENANUNI, Ano VII, nº, abr 1993.
- [BAR 92] BARBLERI, C. Administração de Dados: Para onde vamos?, Datanews, ago 1992.
- [BRA 96] Brachman, R.J., Khabaza, T., Kloesgen, W., Simoudis, E., "Mining Business Databases", Communications ACM, Special Issue on Data Mining, V.39, N.11, novembro 1996.
- [CHA 97] Chaudhuri, S., Dayal, U. "An Overview of Data Warehousing and OLAP Technology", SIGMOD Record , V.26, N.1, março 1997.
- [CHE 90] CHEN, P.: Modelagem de Dados: a abordagem Entidade-Relacionamento para Projeto Lógico. São Paulo: Makron, McGraw-hill, 1990.
- [COS 94] COSTA, O. W. DIAS DA, JAD - Joint Application Design, Infobook, Rio de Janeiro, RJ, 1994.
- [CPD 97] CPD-UFRGS. Projeto de Democratização da Informação na UFRGS: Metodologia para a Modelagem Conceitual, Publicação Interna CPD-UFRGS, 1997.
- [CPD 04] CPD-UFRGS. A travessia digital: do processamento de dados para a tecnologia da informação 1996 a 2004, Universidade Federal do Rio Grande do Sul. Centro de Processamento de Dados, Porto Alegre, 2004.
- [FAY 96] Fayad, U., Shapiro, G.P., "Data Mining and Knowledge Discovery in Databases: An overview", Communications ACM, Special Issue on Data Mining, V.39, N.11, novembro 1996.
- [FEL 88] FELICIANO NETO, A., et alii. Engenharia da informação: Metodologia, Técnicas e Ferramentas. São Paulo, McGraw-Hill, 1988.
- [FIN 95] Finklestein, R., "Multidimensional Databases: Where Relational Fears to Tread", Database Programming & Design, V.8, N.4, abril 1995.
- [HAN 92] Han, J., Cai, Y., and Cercone, N., "Knowledge Discovery in Databases: An Attribute-oriented Approach", Proceedings of the 18th VLDB Conference, Vancouver, British Columbia, Canada, 1992.
- [HAR 96] Harinarayan, V., Rajaraman, A., Ullman, J., "Implementing Data Cubes Efficiently", SIGMOD Record, V.25, N.2 , junho 1996.
- [INM 94] Inmon, W.H., Hackathorn, R.D., Using the Data Warehouse, Wiley-QED Publication, 1994.
- [KIM 96] Kimball, R., The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, John Wiley & Sons, Inc., New York, 1996.
- [KIM 98] KIMBALL, R.: Preparando-se para o Data Mining, DBMS nº 9, Fev-Mar/98.
- [KIP 93] KIPPER, A., F. et alii. Engenharia de Informações. Conceitos, Técnicas e Métodos. Sagra: D. C. LUZATTO, Porto Alegre, 1993.
- [MOR 91] MORAIS, F., Técnicas de Reunião JAD (Joint Application Design) VI Congresso Brasileiro de Metodologias de Desenvolvimento de Sistemas São Paulo, nov 1991.
- [WEB 90] WEBSTER, R. Na Administração de Dados é preciso conhecer o 'business'. Datanews, mai 1990.
- [WEL 95] Weldon, J.L., "Managing Multidimensional Data: Harnessing the Power", Database Programming and Design, V.8, N.8, agosto 1995.
- [WEL 96] Weldon, J.L., "Choosing Tools for Multidimensional Data", Database Programming and Design, V.9, N.2, fevereiro 1996.