

RODOLFO GRACIANO MAZARIN

ANÁLISE DE INFORMAÇÕES USANDO ORACLE DISCOVERER
EM UM DATA WAREHOUSE

ASSIS
2008

ANÁLISE DE INFORMAÇÕES USANDO ORACLE DISCOVERER EM UM DATA WAREHOUSE

RODOLFO GRACIANO MAZARIN

Trabalho de Conclusão de Curso apresentado ao
Instituto Municipal de Ensino Superior de Assis,
como requisito do Curso de Graduação, analisado
pela seguinte comissão examinadora:

Orientador : Prof. Dr. Alex Sandro Romeo de Souza Poletto

Analizador (1): Prof. Esp. Domingos de Carvalho Villela Júnior

Analizador (2): Prof. Dr. Luiz Ricardo Begosso

ASSIS
2008

RODOLFO GRACIANO MAZARIN

ANÁLISE DE INFORMAÇÕES USANDO ORACLE DISCOVERER
EM UM DATA WAREHOUSE

Trabalho de Conclusão de Curso apresentado ao
Instituto Municipal de Ensino Superior de Assis,
como requisito do Curso de Graduação, analisado
pela seguinte comissão examinadora:

Orientador: Prof. Dr. Alex Sandro Romeo de Souza Poletto

Área de Concentração: Sistemas de Bancos de Dados; Data Warehouse

ASSIS
2008

AGRADECIMENTOS

Agradeço primeiramente a DEUS que me deu a vida e a possibilidade de poder cursar novamente uma faculdade, depois de ter desperdiçado um tempo precioso buscando a realização onde não alcançaria.

Aos meus pais pelo apoio e incentivo, mesmo sem muitas condições, contudo mantiveram a fé que eu alcançaria meu objetivo.

À minha esposa e filhas pela paciência na compreensão da grandiosidade dessa atitude de cursar uma faculdade em meio a tantas dificuldades.

Aos mestres, pelo empenho e dedicação com que se comprometeram a compartilhar de seus conhecimentos, em especial ao meu orientador Alex Poletto, pela paciência e dedicação com que encarou esse desafio.

Foram sem duvida os cinco anos mais difíceis da minha vida, com muitas frustrações, alegrias, amizades, coisas que levarei para o resto da minha vida. Mas vejo isso como uma experiência de vida maravilhosa que sem dúvida me trouxe um grande crescimento individual e amigos que, com certeza, terei para sempre em minhas lembranças.

RESUMO

Existe hoje, uma diversidade infinita de formas de se levantar informações. Através de pesquisas ao público, pesquisas de mercado, etc. Analisar essas informações tem sido de crucial importância para o desenvolvimento de campanhas de atendimento a clientes, cobertura de mercado em vendas, e porque não dizer, redução nos custos de produção? Fazer uma boa análise das informações é de suma importância para o desenvolvimento dessas rotinas. Muitas empresas dispõem de informações importantíssimas em sua base de dados e porque não explorar essas informações a fim de traçar metas, controlar gastos e fazer com que, no mercado restrito que vivemos hoje, tirar o máximo de cada rotina das empresas? Para isso existem os SAD's (Sistemas de Apoio a Decisão) e saber manusear essas importantes ferramentas é o novo desafio para os analistas das empresas. A finalidade desse estudo é mostrar o funcionamento de uma ferramenta para extração de informações para tomada de decisões, de forma rápida, prática e objetiva.

Palavras Chave: *Data Warehouse. Data Mart. Oracle Discoverer. Business Intelligence*

ABSTRACT

It exists today, an infinite diversity of forms of information. Through research to the public, research of market, etc. To analyze these information has been of crucial importance for the development of attendance campaigns the customers, covering of market in vendas, and because not to say, reduction in the production costs? To make a good analysis of the information is of utmost importance for the development of these routines. Many companies you make use of important information in its database and because not to explore these information in order to trace goals, to control expenses and to make with that, in the restricted market that we live today, to take off the maximum of each routine of the companies? For this SAD's exist (Systems of Support the Decision) and to know to handle these important tools are the new challenge for the analysts of the companies. The purpose of this study is to show the functioning of a tool for extration of information for taking of decisions, of fast form, practical and objective.

Keywords: *It dates Warehouse. It dates Mart. Oracle Discoverer. Business Intelligence*

LISTA DE ILUSTRAÇÕES

Figura 1 – Elementos de um <i>Data Warehouse</i>	23
Figura 2 – A questão da não-volatilidade	25
Figura 3 – A visão das camadas	31
Figura 4 – Exemplo de Um Modelo Estrela (<i>STAR</i>)	33
Figura 5 – Exemplo de um Modelo Floco de Neve (<i>Snow Flake</i>).....	34
Figura 6 – Selecionando um banco de dados de origem	40
Figura 7 – Selecionando um banco de dados de destino.....	41
Figura 8 – Especifica como uma ou mais tabelas serão exportadas.....	41
Figura 9 – Escolha das tabelas	42
Figura 10 – Especifica se o trabalho é salvo, agendado ou replicado.....	42
Figura 11 – Transferencia de dados.....	43
Figura 12 – Criação da EUL	44
Figura 13 – Selecionando Usuário	45
Figura 14 – Criando nova área de negócio	45
Figura 15 – Selecionando usuário Oracle	46
Figura 16 – Selecionando tabelas que compõem a área de negocios	46
Figura 17 – Estabelecer itens que serão criados automaticamente	47
Figura 18 – Nomear área de negocios	47
Figura 19 – Criando Joins entre tabelas.....	48
Figura 20 – Área de trabalho finalizada.....	48
Figura 21 – Login no Discoverer Desktop Edition	49
Figura 22 – Criando novo caderno de trabalho	49
Figura 23 – Escolha de formas de exibição de resultados	50
Figura 24 – Definindo o layout.....	50
Figura 25 – Definindo Condições	51
Figura 26 – Classificando os dados	51
Figura 27 – Definindo Calculos	52
Figura 28 – Caderno de negocios finalizado	52
Figura 29 – Criando Pivoteamentos	53
Figura 30 – Trabalhando com Pivoteamentos.....	54

LISTA DE ABREVIATURAS E SIGLAS

DW	<i>Data Warehouse</i>
HCAA	<i>Health Care Administrators Association</i>
SADM	Sistemas de Apoio a Decisão em Marketing
B2B	Business to Business
COBOL	<i>Common Business Oriented Language</i>
DASD	<i>Direct Access Storage Device</i>
SGBD	Sistema de Gerenciamento de Banco de Dados
BI	<i>Business Intelligence</i>
SAD	Sistemas de Apoio a Decisão
DSS	<i>Decision Support System</i>
VLDB	<i>Very Large Databases</i>
OLAP	<i>On-Line Analytical Processing</i>
MDA	<i>Multidimensional Analysis</i>
ROLAP	<i>Relational On Line Analytic Processing</i>
MOLAP	<i>Multidimensional On Line Analytic Processing</i>
HOLAP	<i>Hybrid On Line Analytic Processing</i>
SQL	<i>Structure Query Language</i>
EUL	<i>End User Layer</i>
DTS	<i>Data Transformation Services</i>

SUMÁRIO

1	INTRODUÇÃO.....	10
2	BENEFÍCIOS DO SAD NA ÁREA DE NEGÓCIOS	14
3	DATA WAREHOUSE	17
3.1.	HISTÓRICO DO DATA WAREHOUSE	17
3.2.	DEFINIÇÃO DE DATA WAREHOUSE	20
3.3.	CARACTERÍSTICAS DO DATA WAREHOUSE	24
3.4.	DATA MART	26
3.5.	DATA WAREHOUSE X DATA MART	27
3.6.	ESTRUTURA DO DATA WAREHOUSE	28
3.6.1	Camada de Extração/Transformação.....	28
3.6.2	Camada de Metadados	29
3.6.3	Camada de Armazenamento dos Dados.....	30
3.6.4	Camada de Acesso aos Dados	30
3.6.5	Visão em Camadas	31
3.7.	Modelagem Multidimensional, Tabelas Fato, Dimensão e Medidas.....	31
3.7.1	Tabelas Fato	32
3.7.2	Tabelas Dimensão.....	32
3.7.3	Modelo Multidimensional	32
3.7.3.1	Modelo Estrela (<i>STAR</i>).....	33
3.7.3.2	Modelo Floco de Neve (<i>SNOW FLAKE</i>)	33
3.8.	Ferramentas de consulta e análise do <i>Data Warehouse</i>	34
3.8.1	OLAP (<i>On-Line Analytical Process</i>).....	34
3.8.1.1	Características do OLAP	35
3.8.2	<i>Data Mining</i>	36
3.9.	ARQUITETURA DO DATA WAREHOUSE	37

3.9.1 Arquiteturas de Acesso aos Dados	37
3.10. ORACLE DISCOVERER	38
4. IMPORTAÇÃO E EXPORTAÇÃO DOS DADOS	40
5. CRIAÇÃO E MANIPULAÇÃO DE ANÁLISES POR MEIO DO ORACLE DISCOVERER	44
6. CONCLUSÃO	55
REFERÊNCIAS BIBLIOGRÁFICAS	56

1. INTRODUÇÃO

O desenvolvimento de sistemas de *Data Warehouse* (DW) tem se tornado nos dias atuais uma grande área de estudo e aplicação nas empresas. A possibilidade de acessar informações confiáveis com boa velocidade e garantia de qualidade de dados está cativando os diretores das organizações que cada vez mais necessitam de um controle mais preciso dos dados da empresa, sem depender de intermediários para poder tomar suas decisões.

As informações contidas nos tradicionais sistemas orientados à transação não suprem as necessidades de consultas dos gerentes que precisam acessar longos períodos históricos, muitas vezes de vários anos, os quais certamente não estarão disponíveis nos sistemas utilizados para gerenciar as tarefas do dia-a-dia da empresa.

Os sistemas de apoio à tomada de decisão têm a finalidade de informar a alta direção da empresa e as gerências quando, como, porquê e onde ocorrem problemas ou oportunidades de melhoria dentro da organização, projetando certas características e "indicando" ao usuário situações que podem muitas vezes passar despercebidas (HCAA, 2003).

É comum encontrar uma diversidade de sistemas de informação funcionando nas empresas de grande, médio e até mesmo nas de pequeno porte, sistemas estes que suprem as necessidades de cada um dos setores das organizações. Além destes sistemas que estão em uso, existem ainda aqueles, que já foram retirados de circulação e que, por um longo período de tempo, armazenaram e geraram grandes quantidades de dados que não podem ser simplesmente arquivados e esquecidos para sempre.

Muitos dos resultados positivos e negativos de campanhas empreendidas podem estar ocultos num conjunto de dados que são analisados por relatórios inertes que mostram apenas o óbvio ao usuário.

Nos últimos anos, a tecnologia que mais se adequou à crescente necessidade das organizações de obterem e trabalharem com informações gerenciais que dão suporte às tomadas de decisão foi a de DW, a qual possui diversas características

que utilizam forma adequada e eficiente as ferramentas de desenvolvimento de modernos bancos de dados.

Através do uso de diversos recursos de extração e gerenciamento dos dados, a tecnologia de *DW* permite que um grande número de usuários possa realizar inferências em um ou mais bancos de dados modelados de uma forma especial, que agiliza o acesso às informações e também permite a formulação de consultas definidas a qualquer momento com um simples movimento de arrastar e soltar objetos em interfaces gráficas (HCAA, 2003).

O *DW* não é como um *software*, que pode ser comprado e instalado em todos os computadores da empresa em algumas horas, na realidade sua implantação exige a integração de vários produtos e processos. Graças aos avanços nos bancos de dados relacionais, no processamento paralelo e na tecnologia distribuída, finalmente a tecnologia da informação pode permitir que qualquer organização elabore um *DW*.

O mercado de *DW* tem se tornado um dos mais ativos, considerando-se que a maior parte dos fornecedores de sistemas de gerenciamento de banco de dados e ferramentas está lançando produtos mais novos e sofisticados. São nestes produtos de *software* que se espera que ocorram grandes avanços na área, especialmente com relação à melhoria de desempenho e qualidade das ferramentas para o usuário final. Como exemplo de avanço tecnológico, o conceito de *DW* também está viabilizando estes recursos no ambiente da *Internet*.

O *DW* permite conhecer melhor o perfil de cada cliente, fornecendo informações competitivas e essenciais a gerentes e responsáveis pelas decisões estratégicas resultando no aumento do conhecimento sobre os negócios, aumento dos lucros, estabilidade e competitividade da empresa.

Há quatro tipos de empresas: as que fazem as coisas acontecerem, as que acham que podem fazer as coisas acontecerem, as que admiram o que aconteceu, e as que não sabem que algo tenha acontecido.

A concorrência em um mercado global cada vez mais competitivo faz com que as empresas se comprometam cada vez mais em criar e manter consumidores satisfeitos.

Para serem bem-sucedidas e de alto desempenho, devem saber como se adaptar a um mercado em mudança contínua, praticando a arte do Planejamento Estratégico orientado para o mercado.

O Planejamento Estratégico orientado para o mercado é o processo gerencial de desenvolver e manter um ajuste viável entre os objetivos, experiências e recursos da organização e suas oportunidades de mercado mutantes. O propósito do Planejamento Estratégico é moldar e remodelar os negócios e produtos da empresa com o objetivo de crescimento e lucro.

Sua principal meta é ajudar as empresas a selecionar e organizar os negócios de maneira saudável, mesmo quando eventos inesperados descontrolam quaisquer de seus negócios ou linhas de produtos.

Uma das etapas do Planejamento Estratégico é a Formulação da Estratégia de Negócio.

Cada negócio deve estabelecer sua própria estratégia para atingir suas metas. Segundo Porter, Michael (1998), embora muitos tipos de estratégias estejam disponíveis, resume-se em três tipos genéricos que fornecem um bom ponto de partida para o pensamento estratégico: Liderança total em Custos, Diferenciação e Foco:

- Liderança total em custos: A empresa faz um grande esforço para reduzir ao máximo seus custos de produção e distribuição, podendo assim, oferecer preços menores do que seus concorrentes e obter uma maior participação de mercado.
- Diferenciação: Neste caso, a empresa concentra esforços para alcançar desempenho superior em uma importante área de benefício para o consumidor, valorizado por grande parte do mercado. Pode-se ser líder em serviços, em qualidade, em estilo, em tecnologias, etc.
- Foco: A empresa aborda um ou mais segmentos de mercado menores, em vez de ir atrás de um grande mercado. Ela deve conhecer as necessidades desses segmentos e obter liderança em custos ou encontrar uma forma de diferenciação dentro deste segmento-alvo.

Conforme Porter, Michael (1998), as empresas que adotam a mesma estratégia dirigida ao mesmo mercado ou segmento de mercado-alvo, formam um grupo

estratégico. Aquelas que aplicam melhor essa estratégia obtêm os maiores lucros, e aquelas que não definem uma estratégia clara serão mal sucedidas.

Criando uma estratégia eficaz de *e-business*, a empresa será capaz de:

- Aumentar a lealdade do cliente;
- Aumentar a rentabilidade;
- Diminuir o tempo de chegada de novos produtos ao mercado;
- Chegar até os clientes de maneira mais econômica, com ofertas específicas;
- Reduzir substancialmente os custos por transação;
- Reduzir drasticamente os custos do serviço ao cliente;
- Reduzir sensivelmente o tempo de atendimento ao cliente.

Para responder essas questões, tem-se hoje o auxílio de ferramentas de apoio à decisão para auxílio no Planejamento Estratégico e ações de *marketing*.

Este trabalho tem o objetivo de mostrar a utilidade de um Sistema de Apoio à decisão no nas várias situações de controle e gestão de informações das empresas.

O *Marketing* Digital poderá utilizar um Sistema de Apoio à Decisão de *Marketing* (SADM) para auxiliar seus gerentes de *marketing* na maximização das vendas, dos resultados e do vínculo com o cliente.

O Sistema de Apoio à Decisão de *Marketing* é um conjunto coordenado de dados, sistemas, ferramentas e técnicas com *software* e *hardware* de apoio pelos quais uma organização reúne e interpreta informações relevantes da empresa e do ambiente, transformando-as em base para ação de *marketing*.

2. BENEFÍCIOS DO SAD NA ÁREA DE NEGÓCIOS

Atualmente, em muitos casos, a empresa não está interessada somente em fechar uma venda: está interessada em conquistar um cliente específico para atendê-lo durante muito tempo. A empresa gostaria de demonstrar ao cliente que ela possui as condições necessárias para atender a suas necessidades de maneira superior, principalmente se ambas as partes puderem formar um compromisso de relacionamento.

O compromisso de trabalho conjunto em longo prazo exige o desenvolvimento de um conjunto completo de acordos. Assim, conforme Rackham, Neil (1998) o processo de venda movimenta-se desde as investigações preliminares sobre os problemas e necessidades do cliente, passando pela demonstração da capacidade do fornecedor, e finalmente chegando-se a um compromisso em longo prazo.

Hoje, mais empresas estão enfatizando a prática do *Marketing* de Relacionamento. Os clientes de hoje são grandes, e freqüentemente, globais. Eles preferem fornecedores que podem vender e entregar um conjunto coordenado de produtos e serviços em muitos locais; e que possam trabalhar em conjunto com as equipes do cliente para melhorar os produtos e os processos.

O *Marketing* de relacionamento é baseado na premissa de que os clientes importantes precisam receber atenção contínua. Os clientes não devem ser lembrados pelas empresas somente ao fechar um pedido, mas também em outras ocasiões.

O Atendimento ao Consumidor deve monitorar esses clientes, conhecer seus problemas e estar prontos para servi-los de inúmeras maneiras.

Para colocar em prática o *Marketing* de Relacionamento, a empresa deverá trabalhar as informações de seu Banco de Dados, de forma a aproveitá-lo para direcionar as ações de propaganda cliente a cliente, o que chamamos de *Marketing one-to-one*.

Assim, as empresas poderão personalizar as ações de *Marketing* de acordo com o grupo de pessoas, e em seguida, personalizar suas mensagens. Por exemplo: no caso da Loja Virtual, o cliente com perfil executivo, receberá um *e-mail marketing* ou mala direta de lançamentos de uma nova coleção somente com produtos direcionados a seu perfil.

Com isso, a empresa terá benefícios como:

- Ações podem ser programadas para atingir clientes potenciais no momento adequado permitindo que seus materiais de comunicação recebam maior índice de leitura porque são enviados às pessoas mais interessadas.
- Permite também o teste de mídias e mensagens alternativas para melhor definição da abordagem mais eficaz em termos de custo-benefício.
- Também torna a oferta e a estratégia da empresa menos visível aos concorrentes.
- Finalmente, as empresas podem mensurar as respostas de suas campanhas para decidir qual foi a mais rentável.

As empresas que conhecem seus clientes podem adaptar produtos, ofertas, mensagens, sistemas de entrega e métodos de pagamento para maximizar o resultado de suas campanhas.

Hoje, as empresas possuem uma ferramenta muito poderosa para reunir os nomes, endereços e outras informações sobre clientes atuais e potenciais: o Banco de Dados de Clientes.

Muitas empresas confundem *mailing list* de clientes com banco de dados de clientes. O *Mailing list* é simplesmente um conjunto de nomes, endereços e números de telefones.

Um banco de dados de clientes contém muito mais informações. Em *Marketing business to business (B2B)*, o perfil do cliente contém os produtos que ele compra, volume das compras e preços praticados, nomes dos contatos-chave, idades, datas de aniversários, *hobbies* e pratos favoritos, *ticket* médio de compras e avaliação qualitativa das forças e fraquezas da empresa no atendimento aos mesmos.

Um banco de dados de clientes bem desenvolvido é um ativo que pode dar à empresa uma vantagem competitiva, pois com as informações de seu Banco de Dados de Clientes, pode obter maior precisão para atingir seu mercado-alvo podendo identificar pequenos grupos de clientes para ofertas e comunicações de *marketing* aperfeiçoadas.

As empresas usam seus bancos de dados de quatro maneiras:

1. Identificação de clientes potenciais. Muitas empresas geram indicações de vendas para divulgar seus produtos e ofertas. Geralmente, o anúncio em algum tipo de dispositivo de resposta, como um cartão-resposta comercial, ou um número de telefone de discagem gratuita. O banco de dados é construído com base nessas respostas. A empresa identifica os melhores clientes potenciais em seu banco de dados, que depois, são contatados pelo correio, telefone ou visita pessoal, na tentativa de convertê-los em clientes reais.
2. Decisão sobre que clientes devem receber uma oferta específica. As empresas definem critérios para descrever o cliente-alvo ideal pra determinada oferta. Depois, procuram em seus bancos de dados de clientes aqueles que mais se ajustam ao tipo ideal. Ao anotar as taxas de resposta dos contatos, a empresa pode melhorar sua precisão de alvo no decorrer do tempo. Para acompanhar uma venda, pode estabelecer uma seqüência automática de atividades: após uma semana, enviar uma nota de agradecimento; após cinco semanas, enviar uma nova oferta; após dez semanas (se o cliente não respondeu), telefonar oferecendo um desconto especial.
3. Aprofundamento da lealdade do cliente. As empresas podem desenvolver o interesse e o entusiasmo dos clientes lembrando suas preferências, enviando brindes, cupons de descontos, etc.
4. Reativação das compras dos clientes. As empresas podem instalar programas de *mailing* para enviar cartões de aniversários, sugestões de compras de natal ou promoções fora de estação aos clientes de seus bancos de dados. Os bancos de dados podem ajudar a empresa a fazer ofertas mais atraentes, ofertas de substituição ou melhoria de produtos apenas quando os clientes estiverem preparados para agir.

Um *Data Warehouse* transforma conjuntos de dados diferentes em um banco de dados coerente e consolidado podendo extrair dados em qualquer nível de detalhe.

3. DATA WAREHOUSE

Como é sabido, o banco de dados é de vital importância para as empresas, porém existem dificuldades na geração de saídas para análise. Este fato se dá ao grande volume de dados distribuídos em diversos sistemas, dificultando a busca por informações para a tomada de decisões embasadas num histórico dos dados. Com a consolidação desses dados é possível identificar tendências de mercado e posicionar a empresa estrategicamente para ser mais competitiva e conseqüentemente maximizar os lucros diminuindo o índice de erros na tomada de decisão.

Com foco nestas necessidades, fornecedores de *software* produziram um novo conceito no mercado, o *Data Warehouse*. Esta tecnologia consiste em organizar os dados corporativos da melhor maneira, para dar subsídio de informações aos gerentes e diretores das empresas para tomada de decisão. Tudo isso num banco de dados paralelo aos sistemas transacionais da empresa.

3.1. HISTÓRICO DO *DATA WAREHOUSE*

Até a chegada do *Data Warehouse* houve toda uma evolução dos processos da área de tecnologia. A seguir, é apresentado parte dessa evolução, começando com o surgimento do ENIAC.

Em 1946, na Universidade da Pensilvânia, é construído um computador movido a 18.000 válvulas que executava 200 operações por segundo. Surgiu o ENIAC, um computador de grandes dimensões, ocupando uma grande sala (DW BRASIL, 2003).

Dez anos depois, é desenvolvido por um grupo de americanos um transistor que era capaz de executar a mesma tarefa das válvulas. Ganhava-se com o tamanho e a baixa potência dissipada. A partir daí, com o transistor, surgiram computadores bem menores e com capacidade incrementada. Em 1964, o *IBM System 360* surgiu para uso empresarial (DW BRASIL, 2003).

No final dos anos 60, para toda organização de grande porte era indispensável o uso de um computador. Era possível rodar apenas um aplicativo por vez, sempre executado sobre arquivos mestres. Tais aplicativos eram baseados em relatórios e programas desenvolvidos em COBOL (*Common Business Oriented Language*). O uso de cartões perfurados era comum. Os dados eram gravados em fitas magnéticas e acessados sequencialmente (DW BRASIL, 2003).

Em 1970, surge a tecnologia DASD (*Direct Access Storage Device*), substituindo as fitas magnéticas pelo armazenamento em disco. Com o DASD surgiu um novo tipo de *software* conhecido como SGBD ou sistema de gerenciamento de banco de dados, que tinha o objetivo de tornar o armazenamento e o acesso a dados no DASD mais fáceis para o programador. A partir daí, com essas duas novas tecnologias que surgiu a idéia de um “Banco de Dados” (DW BRASIL, 2003).

A idéia principal do banco de dados era armazenar dados estruturadamente, onde o computador seria um coordenador central para as atividades de toda empresa. Nesta linha, o banco de dados tornou-se um recurso corporativo básico. Foi um passo muito importante, pois os computadores passaram a ser vistos como uma poderosa ferramenta de negócios, em que as empresas alcançariam maior eficiência. A idéia dos sistemas de informação para os negócios começou a tomar forma.

Nas décadas de 70 e 80, os sistemas de informação foram aperfeiçoados, tornando-se menos caros e mais poderosos. Devido aos bancos de dados relacionais, as empresas viam-se bastante interessadas na informatização de seus negócios: os executivos tinham acesso a microcomputadores e a sua facilidade de uso aumentou muito.

Novos aperfeiçoamentos e tecnologias foram aparecendo. Facilidades como sistemas de reservas aéreas em nível mundial, sistemas bancários globais e cartões de crédito internacionais, tudo isso devido a um processamento de transações *online* de alta *performance*.

Na década de 80, entram em cena também os computadores pessoais e as linguagens de quarta geração, permitindo que o usuário final assumisse um papel mais ativo, controlando diretamente os sistemas e os dados, fora do domínio do clássico processamento de dados.

Assim, com toda essa evolução, as empresas se deparam com uma grande descoberta: de poder implementar seus recursos de *Business Intelligence*¹, pois com toda essa nova tecnologia era possível analisar de forma otimizada seus dados. Nasce um novo conceito para a tecnologia da informação e conforme Jayme Teixeira Filho, consultor da empresa Informal Informática, os sistemas informatizados passaram a pertencer a dois grupos:

- Sistemas de Processamento de transações: aqueles que têm como principal objetivo o registro acurado das operações e fatos relevantes das áreas de negócio. Sua ênfase é com a validação dos dados, visando maior qualidade e depuração das bases de dados.
- Sistemas de Suporte à Decisão: aqueles projetados para apoiar os gestores de negócio no processo de tomada de decisão numa perspectiva de mais longo prazo, no trato da informação, do que os sistemas de processamento de transações, e envolvendo um maior julgamento humano.

Novas ferramentas de análise de informação surgem, e não satisfeitos, os gerentes começaram a exigir dos sistemas transacionais respostas às suas solicitações, queriam ferramentas que pudessem gerar e armazenar informações estratégicas para dar suporte a um *Business Intelligence* eficiente e não apenas garantir a operação da empresa.

Com tal exigência, surge no mercado os Programas Extratores, programas que extraíam informações dos Sistemas Transacionais para trabalhá-las em outros ambientes, algumas vezes sendo tratadas, ocasionando certa falha na integridade das informações, acarretando na falta de credibilidade dos dados, queda de produtividade e a informação sendo publicada com valores diferentes. Além disso, os Sistemas Transacionais geravam um grande volume de dados e não possuíam uma forma de limpeza da base afim de torná-la mais leve, as pesquisas eram produzidas muito lentamente, devido à diversidade dos sistemas implantados nas empresas.

¹ **Business Intelligence** é o processo de consolidar e analisar dados visando obter informações que possibilitem o ganho de vantagem competitiva.

Outro ponto fraco que prejudicava as decisões era a falta de registros dos fatos históricos nos Sistemas Transacionais, pois estes trabalhavam com uma situação instantânea dos negócios.

Mesmo assim, com tantas adversidades, podemos dizer que é possível à prática de *Business Intelligence* com os sistemas operacionais da empresa, e com outras fontes de dados, como por exemplo, planilhas eletrônicas e dados em papel. Como tais dados não possuem entre si relação de integridade, pois provêm de várias fontes independentes podendo ocasionar alguns equívocos.

Em consequência, pesquisas foram iniciadas para solucionar esses problemas, começou-se a estudar uma forma de se armazenar a informação contida nos sistemas transacionais numa base de dados central, para que houvesse integração total dos dados da empresa e também manter o histórico das informações, onde o analista de negócios poderia ver os fatos de diversas dimensões. Surge daí, o *Data Warehouse*, uma modalidade de Sistema de Apoio à Decisão - SAD.

Traduzindo, *Data Warehouse* seria o mesmo que Armazém de Dados, com ele foram necessários novos métodos de estruturação de dados, tanto para armazenamento quanto para a recuperação de informações. Lembrando que as técnicas para se projetar um *Data Warehouse* são totalmente diferentes dos Sistemas Transacionais, assim como os usuários, o conteúdo dos dados, a estrutura, a forma de armazenamento (normalizado ou desnormalizado), as solicitações, o retorno e o volume de informações são diferentes.

O *Data Warehouse* é utilizado para fazer o gerenciamento do banco de dados e dar suporte à tomada de decisões, os tipos de dados e de informações que os gerentes precisam mudam conforme as transformações dos processos corporativos. Essas informações são bens valiosos, são armas que as empresas possuem para enfrentar a concorrência e se manterem no mercado, cada vez mais concorrido.

3.2. DEFINIÇÃO DE DATA WAREHOUSE

Não existe uma definição precisa do termo *Data Warehouse*, e diferentes autores atribuíram diferentes definições ao que é um *Data Warehouse*. Resumidamente, é um conjunto de dados e ferramentas que servem de base para a tomada de decisão dentro da empresa.

Algumas definições sobre *Data Warehouse*:

É um conjunto de ferramentas e técnicas de projeto, que quando aplicadas às necessidades específicas dos usuários e aos bancos de dados específicos permitirá que planejem e construam um *Data Warehouse*. (KIMBALL, 2002.)

Data Warehouse é um processo em andamento que aglutina dados de fontes heterogêneas, incluindo dados históricos e dados externos para atender à necessidade de consultas estruturadas e *ad-hoc*, relatórios analíticos e de suporte de decisão. (HAJJINDER; RAO, 1996)

Data Warehouse é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo de tomada de decisão. (INMON, 1997)

O *Data Warehouse* constitui a atual evolução dos Sistemas de Suporte à Decisão (DSS – *Decision Support Systems*), que surgiram do aprimoramento das tecnologias de rede e de banco de dados. É um sistema de gerenciamento de banco de dados relacional projetado, especificamente, para dar suporte à tomada de decisão gerencial, não para atender às necessidades dos sistemas transacionais. O esquema disponibiliza um banco de dados específico para suporte à decisão, que gerencia o fluxo de informação dos atuais bancos de dados corporativos e de fontes externas, para aplicações de suporte à decisão que atendam usuários finais.

Comparando o armazenamento nos bancos de dados transacionais com o armazenamento no *Data Warehouse*, temos:

Os bancos de dados transacionais armazenam as informações das transações diárias da empresa, são utilizados por todos os funcionários para registrar e executar operações pré-definidas, por isso seus dados podem sofrer constantes mudanças. Por não ocorrer redundância nos dados e as informações históricas não ficarem armazenadas por muito tempo, este tipo de banco de dados não exige grande capacidade de armazenamento.

No *Data Warehouse* são armazenados dados históricos extraídos de sistemas corporativos e de fontes externas, como mostra a Figura 1. Em geral, mantêm-se de 5 a 10 anos de dados (correntes e históricos). Estes dados operacionais e externos são “limpos”, para remover as inconsistências, e integrados, para criar um novo banco de dados mais ajustado à análise corporativa.

Com base nestes conceitos é possível concluir que o *Data Warehouse* é um conjunto de técnicas e bancos de dados integrados, projetados para suportar as funções dos Sistemas de Apoio à Decisão (SAD), onde cada unidade de dados está relacionada a um determinado assunto, ou fato. Esses bancos de dados darão informações de subsídio aos gerentes e diretores de empresas, para analisarem tendências históricas dos seus clientes e com isso melhorarem os processos que aumentem a satisfação e fidelidade dos mesmos.

Data Warehouses são modelados a partir de grandiosos banco de dados (VLDB – *Very Large Databases*), contendo milhões, e até mesmo centenas de milhões, de registros de dados. Conforme estes dados são coletados de vários sistemas em produção, é construída uma base histórica que o analista corporativo pode usar. Para permanecer atualizada e precisa, é necessário receber atualizações regulares e cuidar para que os dados antigos e desnecessários sejam excluídos. O processo de atualização precisa ser rápido, eficiente e automatizado, para não comprometer a confiabilidade da informação. Com as ferramentas de limpeza de dados, é possível: fundir os dados de muitas fontes num único banco de dados; automatizar a entrada e consistência dos dados; apagar os dados indesejáveis; e manter os dados num sistema de gerenciamento de banco de dados.

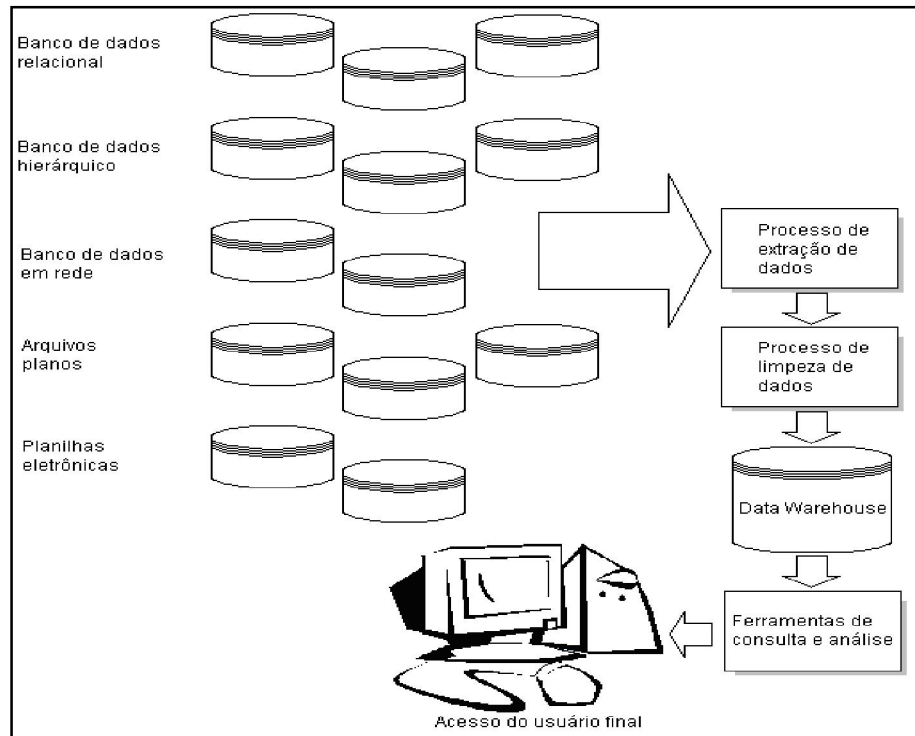


Figura 1. Elementos de um Data Warehouse (Inmon, 1999, p. 14)

Para organizar os dados, são necessários novos métodos de armazenamento, estruturação e novas tecnologias para a geração e recuperação dessas informações.

Essas tecnologias diferem dos padrões operacionais de sistemas de banco de dados em três maneiras:

- Dispõem de habilidade para extrair, tratar e agregar dados de múltiplos sistemas operacionais em *Data Marts*² ou *Data Warehouses* separados;
- Armazenam dados freqüentemente em formato de cubo (OLAP - *on-line analytical processing*) multidimensional permitindo rápida agregação de dados e detalhamento das análises (*drill down*);
- Disponibilizam visualizações informativas, pesquisando, reportando e modelando capacidades que vão além dos padrões de sistemas operacionais freqüentemente oferecidos.

² **Data Mart:** Subconjunto lógico do *Data Warehouse*, aplicado em departamentos da empresa.

3.3. CARACTERÍSTICAS DO DATA WAREHOUSE

Segundo Inmon (1997) “o *Data Warehouse* é um conjunto de dados baseados em assuntos, integrado, não-volátil, e variável em relação ao tempo, de apoio às decisões gerenciais”.

Citando Inmon (1997), a orientação por assunto é uma característica marcante de um *Data Warehouse*, pois toda modelagem está voltada em torno dos principais assuntos da empresa. Enquanto todos os sistemas transacionais clássicos estão voltados para processos e aplicações específicas, os *Data Warehouses* objetivam assuntos.

Assuntos são um conjunto de informações relativas à determinada área estratégica de uma empresa.

No caso de uma companhia de seguros, os sistemas transacionais abordados dizem respeito a automóvel, saúde, vida e perdas. Os principais assuntos ou negócios da empresa podem ser clientes, apólice, prêmio e indenização.

A segunda característica marcante do *Data Warehouse* é o fato de ser *integrado*. “De todos os aspectos, esse é o mais importante” (INMON, 1997). Através da integração é possível obter uma representação única para os dados de todos os sistemas que formarão a base de dados do *Data Warehouse*. Por isso, grande parte do trabalho na construção de um *Data Warehouse* está na análise dos sistemas transacionais e dos dados que eles contêm. Esses dados geralmente encontram-se armazenados em vários padrões de codificação, isso se deve aos inúmeros sistemas existentes nas empresas, e que eles tenham sido codificados por diferentes analistas. Isso quer dizer que os mesmos dados podem estar em formatos diferentes.

O processo de introdução dos dados no *Data Warehouse* é conduzido de forma que as muitas inconsistências das aplicações sejam desfeitas. É importante que a codificação do *Data Warehouse* seja feita de forma consistente e independente das aplicações de origem. Caso os dados de uma aplicação estejam codificados como x/y, eles serão convertidos à medida que forem transferidos para o *Warehouse*. As considerações sobre consistência são válidas para todas as questões de projeto de aplicações, como as convenções de atribuição de nomes, estruturas das chaves, unidades de medidas de atributos e características dos dados.

A terceira característica importante é a *não-volatilidade* dos dados. Como mostra a Figura 2.

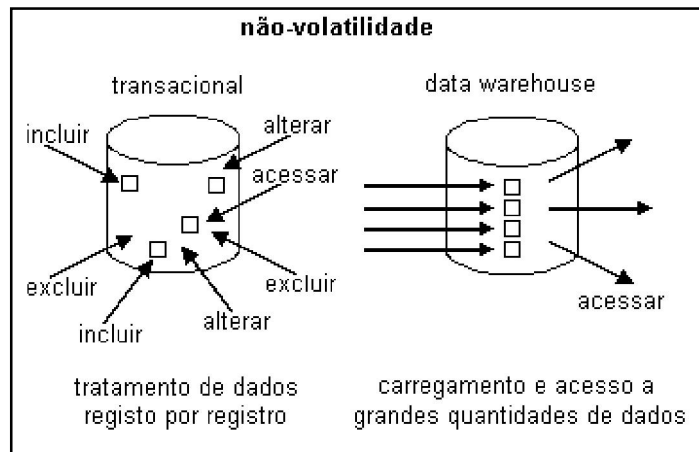


Figura 2. A questão da não-volatilidade (INMON, 1999, p. 36)

Os dados transacionais são regularmente acessados e tratados registro a registro. No ambiente transacional, os dados sofrem atualizações. Contudo, no ambiente do *Data Warehouse* existem somente duas operações, a carga e a consulta dos dados. Ambas as operações envolvem grande volume de dados. A atualização desses dados, geralmente não ocorre no ambiente do *Data Warehouse*.

Em outras palavras, a maior parte dos dados é física e radicalmente alterada quando passam a fazer parte do *Data Warehouse*. Do ponto de vista de integração, não são mais os mesmos dados do ambiente operacional. À luz destes fatores, a redundância de dados entre os dois ambientes raramente ocorre, resultando em menos de 1 por cento de duplicações. (INMON, 1997).

Essa definição dada por Inmon é muito válida porque os dados sempre passam por filtros antes de serem inseridos no *Data Warehouse*. Com isso muitos deles jamais saem do ambiente transacional, e outros são tão resumidos que não se encontram fora do *Data Warehouse*.

A última característica significativa do *Data Warehouse* diz respeito ao fato de ele ser variável em relação ao tempo, isso nada mais é do que manter o histórico dos dados durante um período de tempo muito superior ao dos sistemas transacionais.

Essa variação em relação ao tempo se manifesta de diversos modos:

- O horizonte de tempo válido para o *Data Warehouse* é de 5 a 10 anos, já nos sistemas transacionais o horizonte de tempo é de 60 a 90 dias (variando em função do escopo do projeto), isso porque num sistema transacional a finalidade

é de fornecer as informações no momento exato, já no *Data Warehouse*, o principal objetivo é analisar o comportamento das mesmas durante um período de tempo maior. Fundamentados nessa variação, os gerentes tomam as decisões em cima de fatos e não de intuições.

- Bancos de dados operacionais contêm dados de “valor corrente” – dados cuja exatidão é válida para o momento de acesso. Assim sendo, dados de valor corrente podem ser atualizados. Os dados existentes em um *Data Warehouse* são como fotografias que se refletem num determinado momento do tempo, por esse motivo não serão atualizados tendo-se assim uma imagem fiel da época em que foram gerados.
- A estrutura de chave dos dados transacionais pode conter, ou não, elementos de tempo, como ano, mês, dia etc. A estrutura de chave do *Data Warehouse* obrigatoriamente contém algum elemento de tempo, porque os dados sempre refletirão num determinado momento de tempo.

Assim como o dado, é importante frisar que os metadados, também possuem elementos temporais, porque mantêm um histórico das mudanças nas regras de negócio da empresa. Os metadados são responsáveis pelas informações referentes ao caminho do dado dentro do *Data Warehouse*.

3.4. DATA MART

O *Data Warehouse* pode constituir um único banco de dados corporativo, ao qual os usuários se conectam diretamente, ou pode ser composto de diversos sistemas menores chamados *Data Marts*, cada um voltado a atender as áreas específicas de negócio ou um departamento da empresa.

Um *Data Mart* pode ser visto como um *Data Warehouse* para um departamento ou unidade de negócio em particular, possui as mesmas características de um *Data Warehouse*, porém com mais rapidez na implementação, custo reduzido, controle local e, por ser menor, possibilita a Análise Multidimensional, com os cruzamentos e visões previamente calculados, visando aumentar a velocidade na consulta das informações.

Um *Data Warehouse* pode ser formado por vários *Data Marts* departamentais. Um *Data Mart* é um subconjunto lógico de um *Data Warehouse*.

Data Marts são úteis para isolar dados detalhados acerca de um assunto para efetuar estudos específicos, com parte do conjunto de dados do *Data Warehouse*.

3.5. DATA WAREHOUSE X DATA MART

As diferenças entre *Data Mart* e *Data Warehouse* são apenas com relação ao tamanho e ao escopo do problema a ser resolvido. Portanto, as definições dos problemas e os requisitos de dados são essencialmente os mesmos para ambos. Enquanto um *Data Mart* trata de problema departamental ou local, um *Data Warehouse* envolve o esforço de toda a companhia para que o suporte a decisões atue em todos os níveis da organização. Sabendo-se as diferenças entre escopo e tamanho, o desenvolvimento de um *Data Warehouse* requer tempo, dados e investimentos gerenciais muito maiores que um *Data Mart*.

Por muitos anos, todos os sistemas que extraíam dados de sistemas legados e os armazenavam de maneira utilizável para suporte à decisão eram chamados *Data Warehouses*. Recentemente, uma distinção tem sido feita entre os corporativos *Data Warehouses* e os departamentais *Data Marts*, mesmo que geralmente o conceito ainda continue sendo chamado de *Data Warehousing*.

Os *Data Marts* atendem necessidades de unidades específicas de negócio contrapondo os *Data Warehouses* que atendem demandas corporativas. Eles otimizam a entrega de informação de suporte à decisão e se focam na gerência e/ou dados exemplificativos ao invés do histórico de níveis gerais. Eles podem ser apropriados e gerenciados por pessoal fora do departamento de informática das corporações.

A crescente popularidade desses mal definidos *Data Marts* em cima da popularidade dos grandes sistemas de *Data Warehouses* Corporativos é baseada em muitos bons motivos:

- Os *Data Marts* têm diminuído drasticamente o custo de implementação e manutenção de sistemas de apoio à decisão e os têm posto ao alcance de um número muito maior de corporações.

- Eles podem ser prototipados muito mais rápido, com alguns pilotos sendo construídos entre 30 e 120 dias e sistemas completos sendo construídos entre 3 e 6 meses.
- Os *Data Marts* têm o escopo mais limitado e são mais identificados com grupos de necessidades dos usuários, o que se traduz em esforço/tempo concentrado.

Os departamentos autônomos e as pequenas unidades de negócio freqüentemente preferem construir o seu próprio sistema de apoio à decisão via abordagem *Data Marts*. Muitos departamentos de informática estão vendo a efetividade desta e estão construindo o *Data Warehouse* por assunto ou um *Data Mart* por vez. Gradualmente vão ganhando experiência e garantindo o suporte dos fatores chave de gerenciamento e venda. Começando com planos modestos e desenvolvendo na medida que se adquire mais conhecimento sobre as fontes de dados e as necessidades dos usuários fazendo com que as organizações justifiquem os *Data Marts* na medida em que os resultados tornam-se visíveis.

Algumas vezes, projetos que começam como *Data Warehouses* transformam-se em *Data Marts*. Quando as organizações acumulam grandes volumes de dados históricos para suporte à decisão que se mostram pouco ou nunca utilizados, elas podem reduzir o armazenamento ou arquivamento de informação e contrair o seu *Data Warehouse* em um *Data Mart* mais focado. Ou elas podem dividir o *Data Warehouse* em vários *Data Marts*, oferecendo tempos de resposta mais rápido, acesso mais fácil e menos complexidade para os usuários finais.

3.6. ESTRUTURA DO DATA WAREHOUSE

Pode-se definir a estrutura de um *Data Warehouse* em quatro camadas:

3.6.1 Camada de Extração/Transformação

É o processo automatizado que extrai, limpa e integra os dados dos sistemas operacionais de acordo com os padrões definidos no *Data Warehouse*, ou seja, garante a integridade dos dados através do tratamento dos mesmos. O processo de

transformação manipula os dados extraídos e formatados com o objetivo de prepará-los para que o modelo permita desempenho adequado durante o processo de consulta.

3.6.2 Camada de Metadados

Os metadados são componentes importantes na estrutura do *Data Warehouse*. Trata-se de um dicionário que armazena as especificações do *Data Warehouse*, incluindo dimensões utilizadas, hierarquia de cada dimensão, categorias elementares, regras de agregação, entre outros.

Os metadados, de maneira simplificada, são descrições sobre os dados. São dados que fazem referência a outros dados. Por exemplo, um catálogo telefônico pode ser considerado um catálogo de metadados por fazer referência a dados pessoais, no caso telefones e nomes. Todas as fases de um projeto de *Data Warehouse*, desde a modelagem até a visualização da informação, geram metadados. Neles estarão contidas informações como atributos das tabelas, agregadas utilizadas, cálculos necessários, descrições, periodicidade das cargas, histórico de mudança

Segundo Inmon (1997), os metadados mantêm informações sobre "o que está e onde está," no *Data Warehouse*. Tipicamente os aspectos que sobre os quais os metadados mantêm informações são:

- A estrutura dos dados, segundo a visão do programador;
- A estrutura dos dados, segundo a visão dos analistas de SAD;
- A fonte de dados que alimenta o *Data Warehouse*;
- A transformação sofrida pelos dados no momento de sua migração para o *Data Warehouse*;
- Modelo de dados;
- O relacionamento entre o modelo de dados e o *Data Warehouse*;
- O histórico das extrações de dados.

3.6.3 Camada de Armazenamento dos Dados

Constitui-se do armazenamento físico dos dados oriundos dos sistemas operacionais da empresa e externos, permitindo um acesso mais rápido e seguro aos dados do *DW*, além de prover maior flexibilidade de tratamento e facilidade de manipulação.

Esta camada é o componente principal do sistema, deve ser escalável e permitir a manipulação de grande volume de informações.

3.6.4 Camada de Acesso aos Dados

Representa a camada de aplicações a partir das quais os usuários acessam os dados disponíveis na camada de armazenagem (*Data Warehouse* ou *Data Marts*). É composta por uma infinidade de ferramentas que oferecem distintas capacidades de análise de dados, como extração de relatórios, análises estatísticas, análises quantitativas, simulações e outras.

Portanto, o *Data Warehouse* não pode ser considerado apenas um grande banco de dados, e sim o conjunto estruturado por camadas que viabilizam o armazenamento e recuperação de grandes volumes de dados.

3.6.5 Visão em Camadas

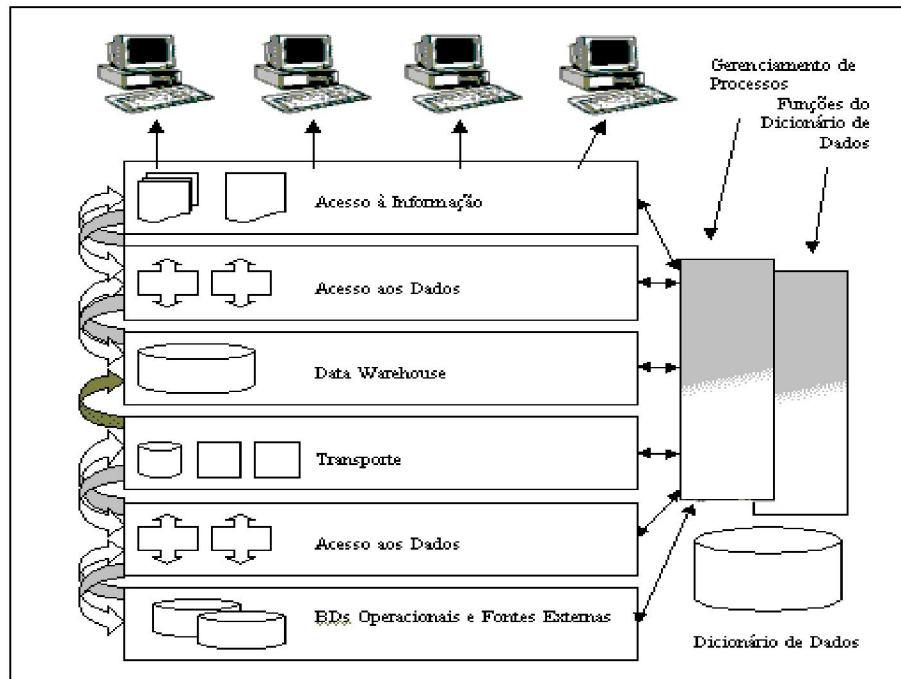


Figura 3: A visão das camadas (INMON, 1999)

3.7. Modelagem Multidimensional, Tabelas Fato, Dimensão e Medidas

Modelagem Multidimensional é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócio.

Um modelo multidimensional possui elementos básicos como:

- Fatos
- Dimensões

3.7.1 Tabelas Fato

Tabelas Fato (ou Tabelas de Fato) são consideradas como uma coleção de itens de dados, compostas de dados de medidas e de contextos, onde cada fato representa um item de negócio.

3.7.2 Tabelas Dimensão

Tabelas Dimensão são elementos que participam de um fato, assunto de negócio, como por exemplo, num banco de dados que analisa as vendas de produtos, teria as seguintes dimensões:

- Tempo
- Local
- Clientes
- Vendedores

Normalmente, essas tabelas não possuem atributos numéricos, pois são somente descritivas e classificatórias dos fatos.

3.7.3 Modelo Multidimensional

No modelo multidimensional, temos duas operações básicas, a *Drill Up* e a *Drill Down*.

Na operação *Drill Up*, tem-se a visualização dos dados num nível mais alto. Já na operação *Drill Down*, a visualização se dá num nível mais detalhado dos dados.

Existem dois modelos básicos na camada multidimensional. São eles o Modelo Estrela (*Star*) e o Modelo Floco de Neve (*Snow Flake*).

3.7.3.1 Modelo Estrela (Star)

É a estrutura básica de um modelo de dados multidimensional. Sua composição típica, possui uma grande entidade central denominada Fato e um conjunto de entidades menores denominadas Dimensões, arranjadas ao redor dessa entidade central, formando uma estrela, como demonstra a Figura 4.

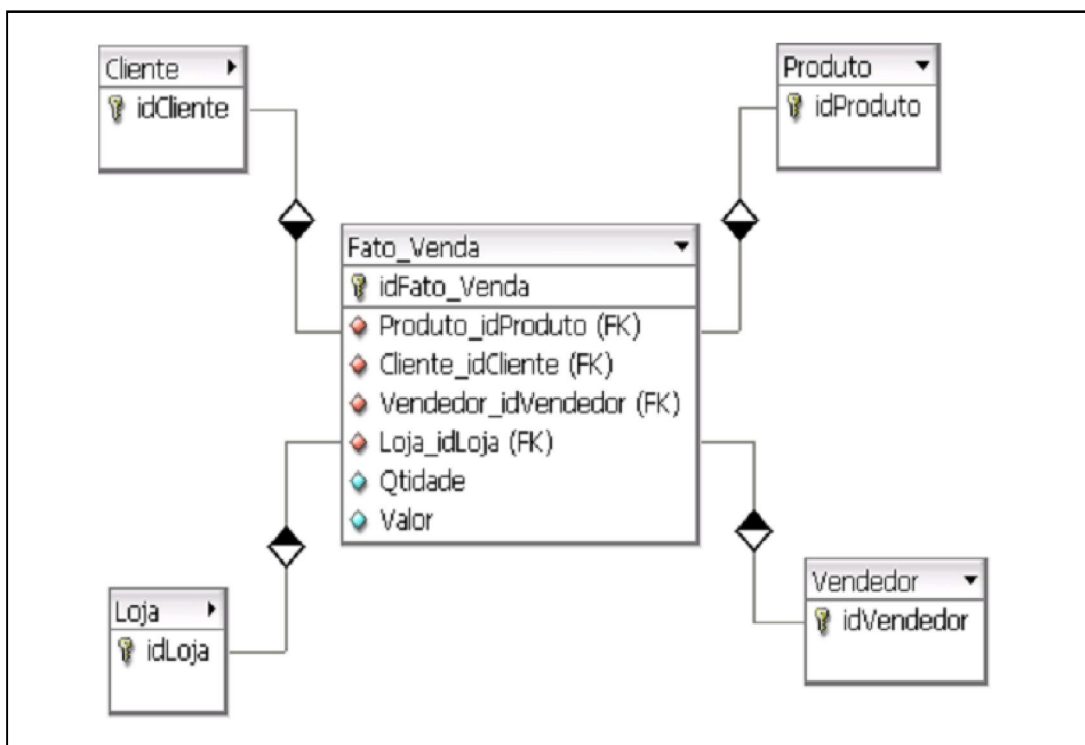


Figura 4 – Exemplo de um Modelo Estrela

3.7.3.2 Modelo Floco de Neve (Snow Flake)

O modelo Floco de Neve é o resultado da decomposição de uma ou mais dimensões que possuem hierarquias entre seus membros, conforme apresentado na Figura 5.

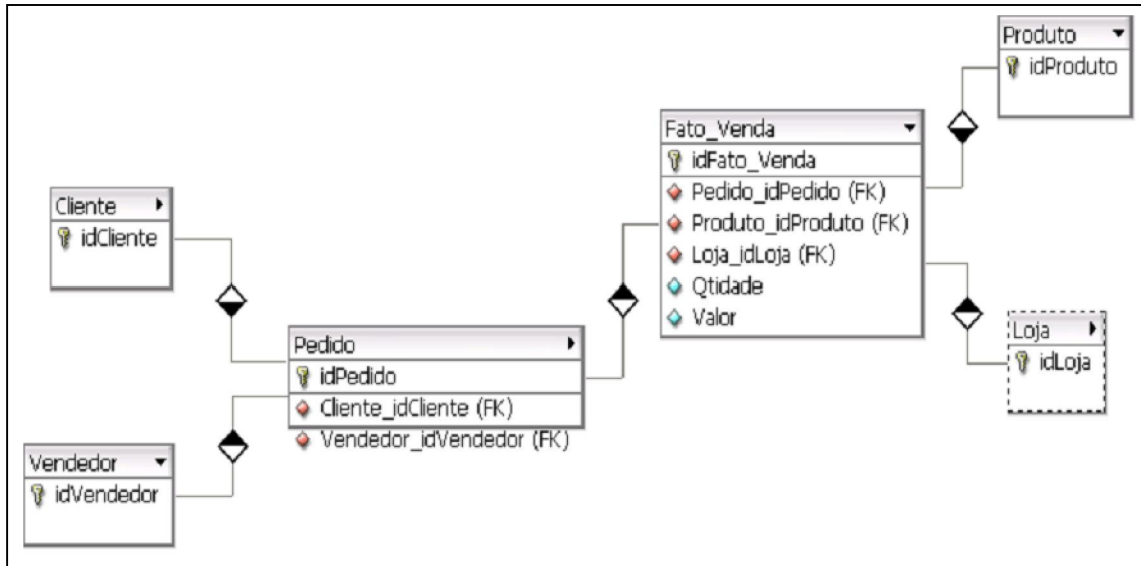


Figura 5 – Exemplo de um Modelo Floco de Neve (Snow Flake)

3.8. Ferramentas de consulta e análise do *Data Warehouse*

As ferramentas de acesso são importantes, pois a principal razão do *Data Warehouse* é prover informações aos usuários para que possam tomar decisões estratégicas. Conhecendo informações, possibilita a formulação de efetivas estratégias de *marketing*, negócios e vendas.

3.8.1 OLAP (*On-Line Analytical Process*)

O processamento analítico *on-line* (OLAP – *On-Line Analytical Processing*) é um método importante na arquitetura do *Data Warehouse* através do qual os dados podem ser transformados em informação.

De acordo com Inmon, OLAP (processamento analítico *On-Line*) é uma “categoria da tecnologia de *software* que permite que analistas, gerentes e executivos obtenham, de maneira rápida, consistente e interativa, acesso a uma variedade de visualizações possíveis de informação que foi transformada de dados puros para refletir a dimensão real do empreendimento do ponto de vista do usuário”. Ou seja,

OLAP é um conjunto de funcionalidades que tenta facilitar a análise multidimensional. A análise multidimensional (MDA – *Multidimensional Analysis*) é a habilidade de manipular dados que tenham sido agregados em várias categorias ou “dimensões”. De acordo com o conselho de OLAP, “o propósito da análise multidimensional é auxiliar o usuário a sintetizar informações empresariais através da visualização comparativa, personalizada, e também por meio da análise de dados históricos e projetados”. (INMON, 1999).

3.8.1.1 Características do OLAP

- Visão conceitual multidimensional: enfatiza a forma como o usuário “vê” os dados sem impor que os dados sejam armazenados em formato multidimensional;
- Transparência: localização da funcionalidade OLAP deve ser transparente para o usuário, assim como a localização e a forma dos dados;
- Facilidade de Acesso: acesso a fontes de dados homogêneas e heterogêneas deve ser transparente;
- Desempenho de consultas consistente: não deve ser dependente do número de dimensões;
- Arquitetura cliente/servidor: produtos devem ser capazes de operar em arquiteturas cliente/servidor;
- Dimensionalidade genérica: todas as dimensões são iguais;
- Manipulação dinâmica de matrizes esparsas: produtos devem lidar com matrizes esparsas eficientemente;
- Suporte multiusuário;
- Operações entre dimensões sem restrições;
- Manipulação de dados intuitiva;
- Relatórios/consultas flexíveis;
- Níveis de agregação e dimensões ilimitados: ferramentas devem ser capazes de acomodar 15 a 20 dimensões.

Em resumo, OLAP é uma tecnologia com recursos para disponibilização do suporte à Decisão e Estratégia de Negócio, obtendo de forma mais rápida, consistente e interativa uma variedade de visualizações possíveis da informação, facilitando a Análise Multidimensional.

3.8.2 *Data Mining*

Data Mining (ou mineração de dados) é o processo de extrair informações a partir de grandes bases de dados, usando-as para tomar decisões de extrema importância, revertendo em ganhos efetivos para a empresa. O *Data Mining* vai além de uma simples consulta a um banco de dados, ele permite explorar e inferir informação útil a partir dos dados, descobrindo informações “camufladas” no banco de dados.

Todos os sistemas de *Data Warehouse* possuem bases de dados bem organizadas e consolidadas que funcionam com boas ferramentas de exploração, por isso a tecnologia do *Data Mining* é considerada como um minerador de dados, o qual chamou a atenção dos construtores de *Data Warehouse* porque a mineração de dados tem alternativas de grandes descobertas de oportunidades de negócio e além disso, também pode traçar estratégias novas para o futuro.

Conforme Inmon (1997), o *Data Mining* analisa todos os dados possíveis para que possam descobrir todas as oportunidades escondidas nos relacionamentos, descobrindo o problema e diagnosticando os comportamentos dos negócios, que requerem o mínimo de intervenção dos usuários, que irá buscar somente os conhecimentos e produzir vantagens competitivas.

O conceito de colocar o *Data Mining* dentro do *Data Warehouse* reduz erros e custos. Antes eram necessários dias ou semanas para que uma análise de negócio ficasse pronta. Com o uso da *Data Mining* é possível construir o modelo imediatamente e de forma interativa.

Concluindo, a prática do *Data Mining* vem contribuir para uma maior eficácia das soluções de suporte à decisão permitindo conferir-lhes um caráter de “previsão do futuro”.

3.9. ARQUITETURA DO *DATA WAREHOUSE*

As arquiteturas de *Data Warehouse* ainda não estão bem definidas e sua classificação ainda é bastante confusa. Apesar disso, vários autores concordam com a classificação das arquiteturas funcionais e arquiteturas de acesso a dados.

3.9.1 Arquiteturas de Acesso aos Dados

Essas arquiteturas são baseadas nas formas como os dados são acessados e trabalhados pelas ferramentas analíticas. Os principais tipos de acesso são:

- *ROLAP (Relational On Line Analytic Processing)*: os dados são armazenados em sistemas de gerenciamento de banco de dados (SGBD) relacionais. ROLAP significa fazer consultas OLAP em SGBDs relacionais. Atualmente a solução mais utilizada.
- *MOLAP (Multidimensional On Line Analytic Processing)*: os dados e agregações são armazenados em SGBDs multidimensionais, otimizados para consultas analíticas. Teoricamente, é a melhor arquitetura, mas a falta de SGBDs multidimensionais mais consolidados dificulta sua aplicação.
- *HOLAP (Hybrid On Line Analytic Processing)*: é uma arquitetura híbrida, entre ROLAP e MOLAP. A idéia é armazenar os dados de maior granularidade do *Data Warehouse* em estruturas relacionais normalizadas e os agregados de menor granularidade em estruturas dimensionais nativas como cubos OLAP.

Essa granularidade é o nível de detalhe ou de resumo dos dados existentes em um *Data Warehouse*. Dados de maior granularidade são mais resumidos, requerem espaço em disco e número de índices bem menores. Já o nível de granularidade menor permite consultas mais detalhadas.

3.10. ORACLE DISCOVERER

Oracle Discoverer Plus é uma ferramenta de acesso a dados, através dela usuários ou profissionais de negócios podem localizar e analisar os exatos dados que precisam, de forma a tomar melhores decisões sobre seus negócios e sem a necessidade de ter nenhum conhecimento em Banco de Dados.

O Discoverer é ferramenta ideal para mineração de dados em *Data Warehouse*, ele suporta uma série de funções SQL (*Structure Query Language*), para *Business Intelligence* e aplicações de *Data Warehouse*. Trabalha com a idéia de Área de Negócios, que são compostas de *folders* (tabelas) e itens (colunas de tabelas) com a finalidade de aproximação com a linguagem do usuário.

O *Discoverer* permite aos usuários customizar relatórios, consultas, incorporar logotipos da empresa e imagens, e gerar gráficos para que se possa tomar as melhores decisões sobre o mercado de negócios.

Tem como benefícios, visualizar dados de maneira rápida, compartilhar dados com outras pessoas, orientar o *marketing* direcionado, analisar clientes e produtos, criar previsões com base em planos atualizados e resultados efetivos.

A camada final do usuário ou EUL (*End User Layer*), isola usuários finais da complexidade do banco de dados, fazendo que foque apenas nas questões de negócios e não de dados. A EUL gera declarações SQL, quando é executada uma consulta a EUL envia essas declarações para o Banco de Dados que retorna para *Discoverer Plus*, portanto o usuário não precisa ter conhecimento de SQL para poder manipular os dados. O importante da EUL é a preservação da integridade física do Banco de Dados, nada que o administrador ou o usuário faça através do *Discoverer* afetar o banco.

As Áreas de Negócios são grupamentos lógicos de tabelas ou visões do Banco de Dados que contém informações necessárias para análise, suporte à decisão e apresentação dos resultados. Possuem como característica, conter dados específicos para os usuários, onde o administrador do Banco de Dados disponibiliza esses grupamentos de dados para proporcionar ao usuário final, somente os dados necessários para análise, suporte à decisão e apresentação dos resultados.

Com as Áreas de Negócios definidas, o Administrador do Banco de Dados estará definindo a partir da necessidade do usuário os Cadernos de Trabalho.

Os Cadernos de Trabalho são compostos de uma ou mais folhas de trabalho. Uma folha de trabalho, nada mais é que o resultado de uma consulta que é visualizada. O usuário também tem a opção de trabalhar com vários cadernos de trabalho ao mesmo tempo, para se comparar valores entre eles.

4. IMPORTAÇÃO E EXPORTAÇÃO DOS DADOS

Em uma aplicação de *Data Warehouse* ou *Data Mart*, o analista de sistema trabalha com os dados do próprio cliente como já mencionado neste trabalho.

A empresa em questão utiliza com sistema gerenciador de banco de dados o *SQL-Server* e este trabalho foi desenvolvido em uma ferramenta *Olap* da *Oracle*, o *Discoverer*.

Para a resolução desse problema neste caso, ou outro que poderia vir conseqüentemente em outras aplicações utiliza-se um assistente de importação e exportação da própria ferramenta SQL que é a *DTS – Wizard*(*Data Transformation Services*).

É apresentado neste capítulo todo o processo para se exportar dados entre bancos de dados. As telas (Figuras) utilizadas na demonstração foram retiradas de <http://www.devmedia.com.br/articles>.

Cada figura representa um passo para conversão das informações de um banco hospedado em uma base de dados SQL Server para a base do Oracle.

O primeiro passo no processo de exportação dos dados é a seleção do qual banco de dados será exportado para o *data warehouse*, conforme ilustra a Figura 6.

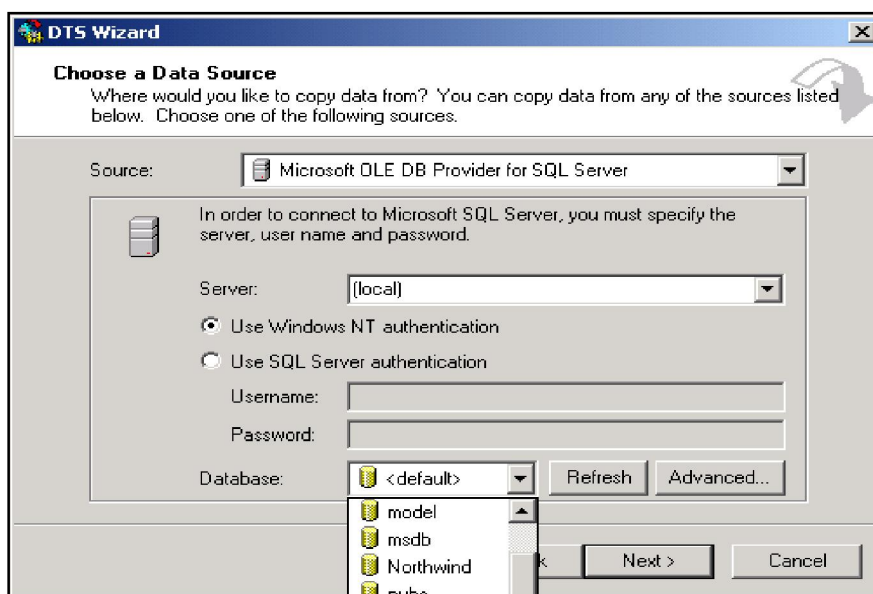


Figura 6. Selecionando banco de dados de origem

Após isso, seleciona-se o banco de dados de destino, conforme ilustra a Figura 7. O banco selecionado deve estar criado e configurado para receber as informações.

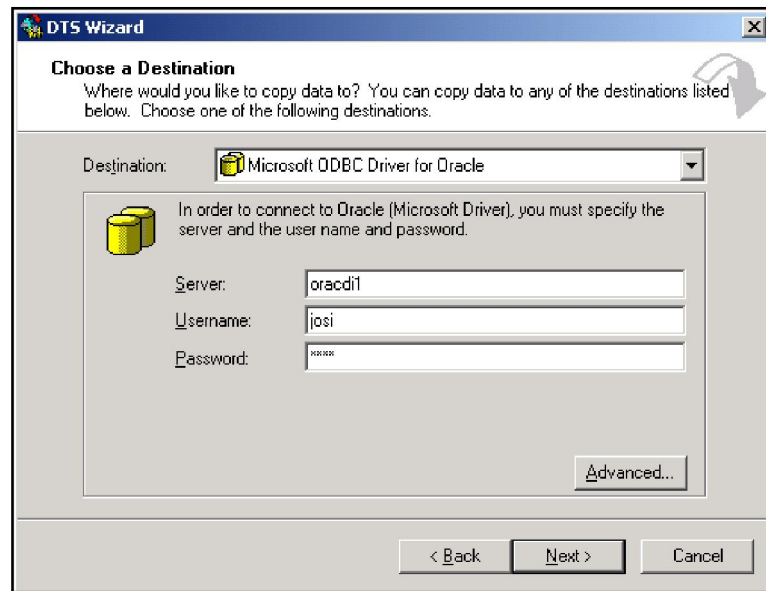


Figura 7: Selecionando banco de dados de destino.

Selecionado o banco, passamos para o passo seguinte que é a especificação de quais tabelas serão convertidas para a nova base de dados criada, conforme ilustra a Figura 8.

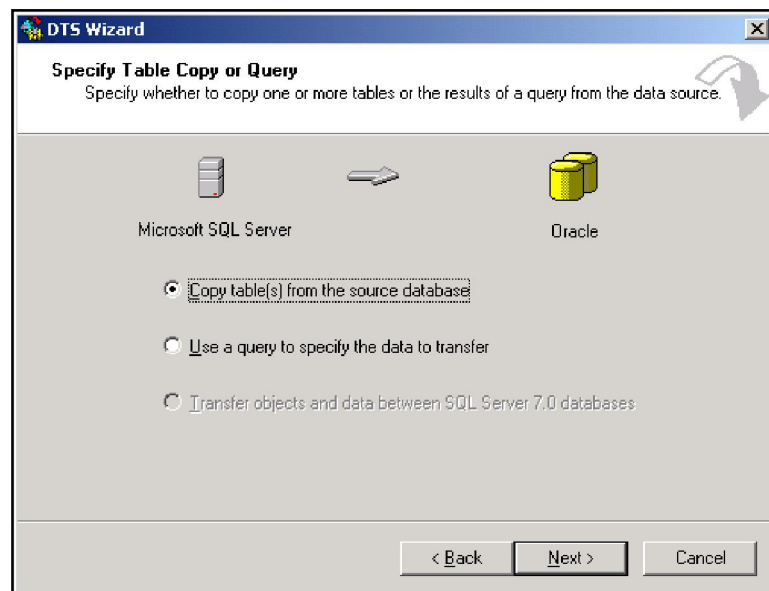


Figura 8: Especifica como uma ou mais tabela serão exportadas.

Após isso, o próximo passo é o de escolha das tabelas que serão exportadas, conforme nos mostra a Figura 9.

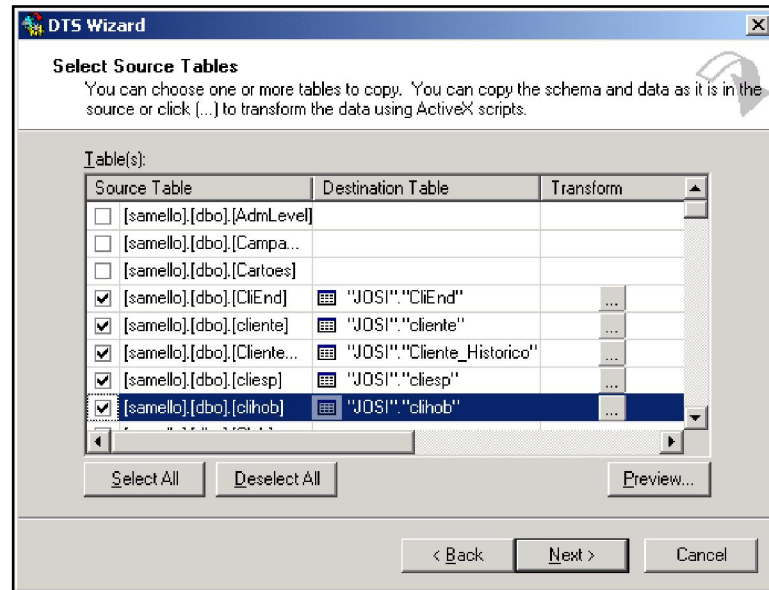


Figura 9: Escolha das Tabelas.

Escolhidas as tabelas, chega-se a hora de determinar se o trabalho será salvo, agendado ou replicado, segundo nos mostra a Figura 10.

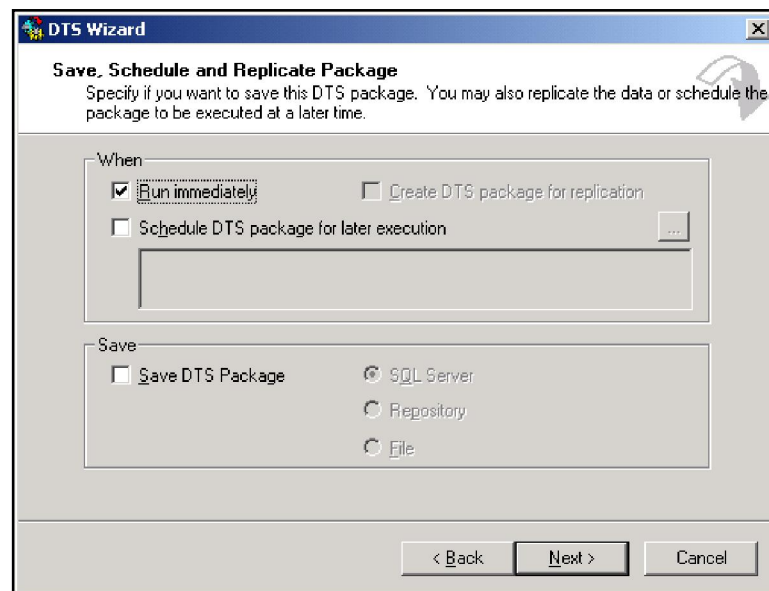


Figura 9: Especifica se o trabalho é salvo, agendado ou replicado.

Após serem feitas todas as configurações e seleções necessárias, passa-se à rotina de transferência de dados entre os bancos, conforme ilustra a Figura 11.

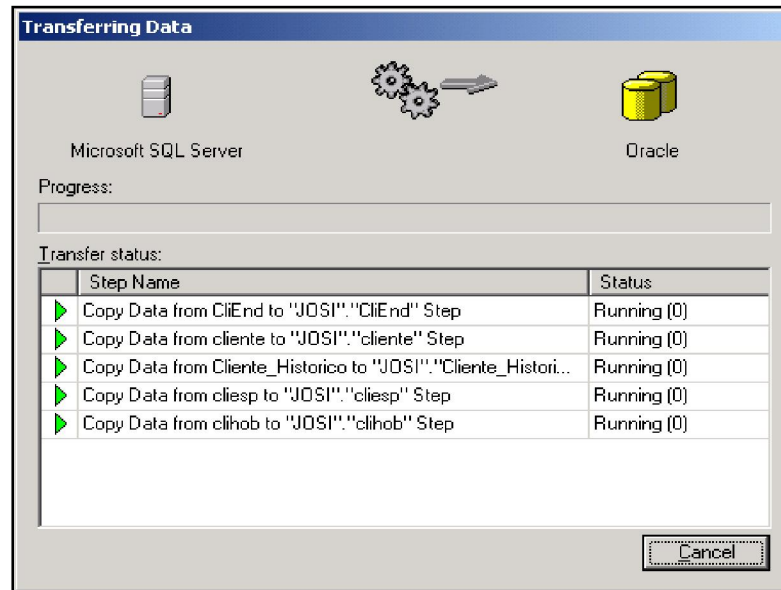


Figura 11: Transferência de dados.

5. CRIAÇÃO E MANIPULAÇÃO DE ANÁLISES POR MEIO DO ORACLE DISCOVERER

A partir do momento em que o usuário se loga, o *Discoverer* tomará posições diferentes dependendo da existência ou não de uma EUL.

Se não existe uma EUL criada, o *Discoverer Administration Edition* pergunta ao usuário se ele deseja criar uma EUL este momento.

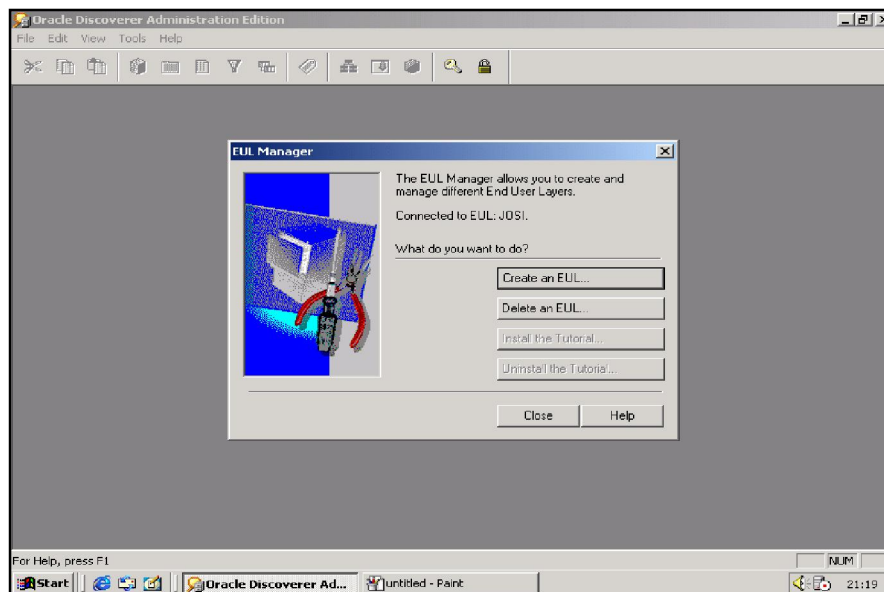


Figura 12: Criação da EUL.

Ao escolher a opção "Create na EUL", o assistente para criação da Camada final de usuário será iniciado, ou poderá escolher entre selecionar um usuário já criado, ou criar um novo usuário. Em seguida, informa o nome de usuário e senha.

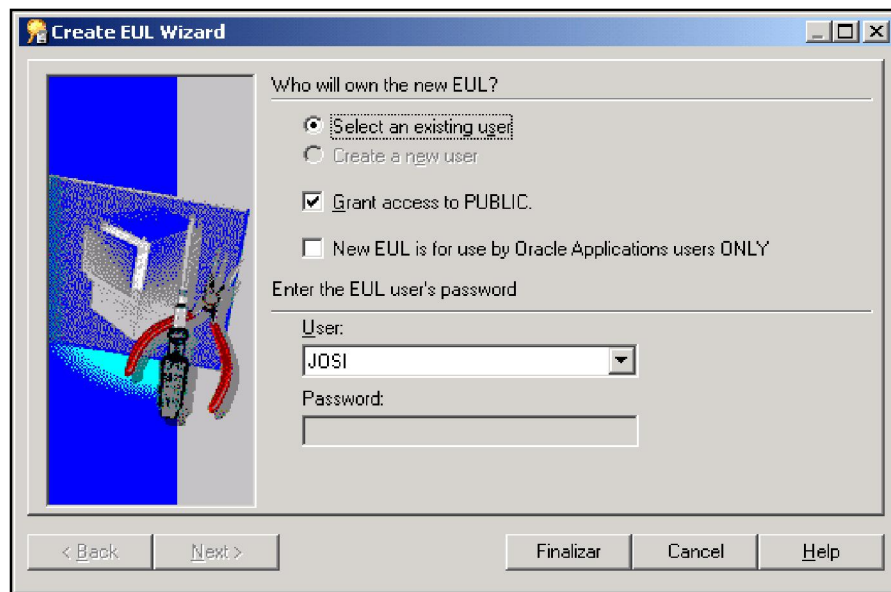


Figura 13: Selecionando Usuário.

Após a finalização da EUL, o analista deverá criar e preparar as Áreas de Negócios (*Business Area*), onde poderão ser agrupados os dados de maneira útil para que o usuário possa acessar suas consultas.

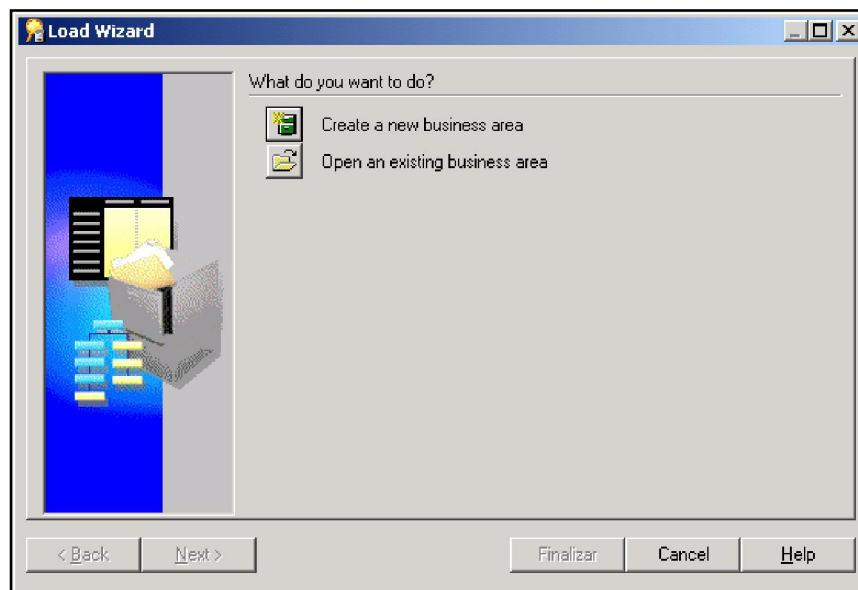


Figura 14: Criando nova área de negócio.

Após selecionar a opção para Criar a nova área de negócios, o analista deverá escolher um *Database link* e identificar o *user ID* que serão usadas nas áreas de negócios.

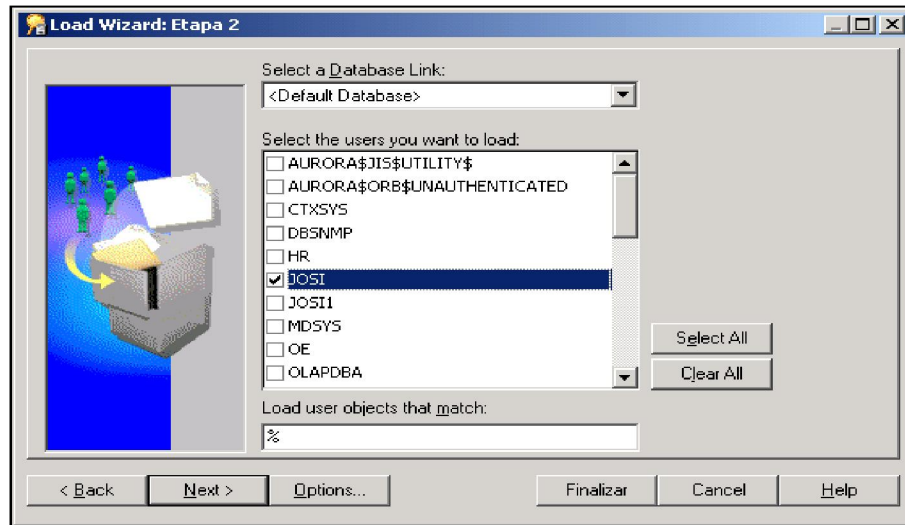


Figura 15: Selecionando usuário *Oracle*.

Na terceira etapa da criação da área de negócios, o usuário deverá escolher as tabelas e visões que serão carregados na área de negócios.

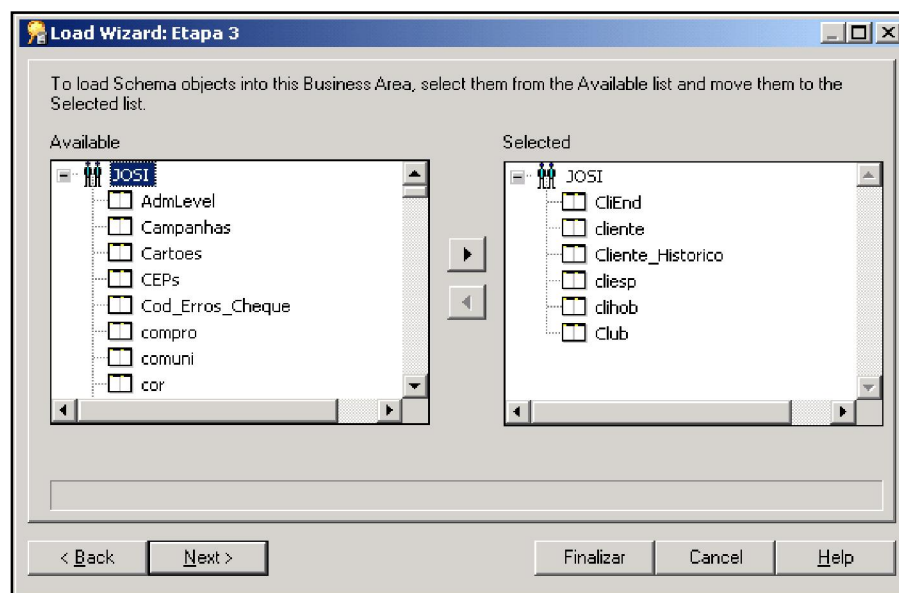


Figura 16: Selecionando tabelas que compõem a área de negócios.

O próximo passo oferece a oportunidade de se gerar automaticamente hierarquias, listas de valores e *joins* e também incluí-las na área de negócios.

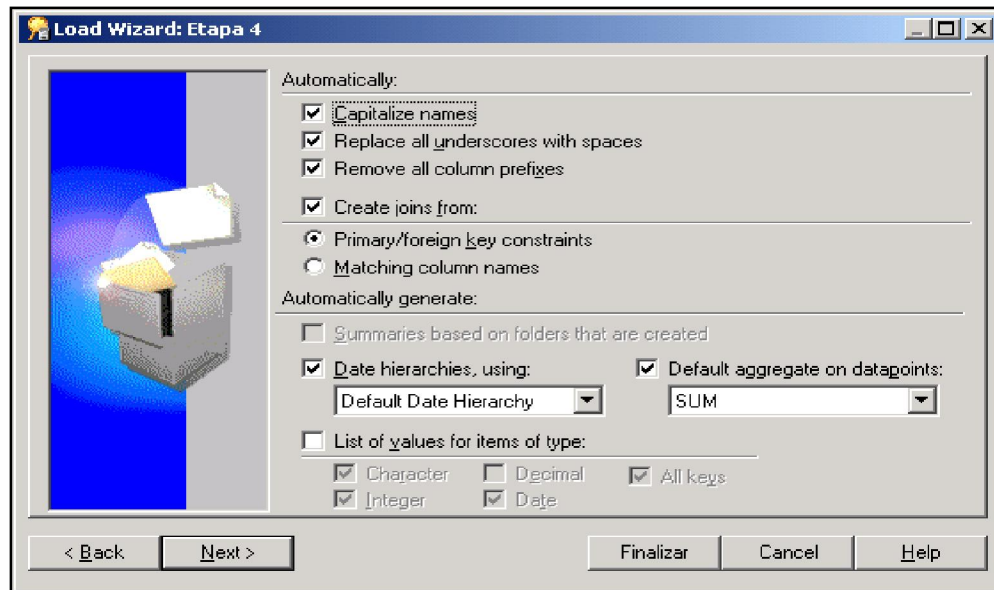


Figura 17: Estabelecer itens que serão criados automaticamente.

A última etapa consiste na nomeação da Área de Negócio. Ao finalizar, as *views* e tabelas serão criadas.

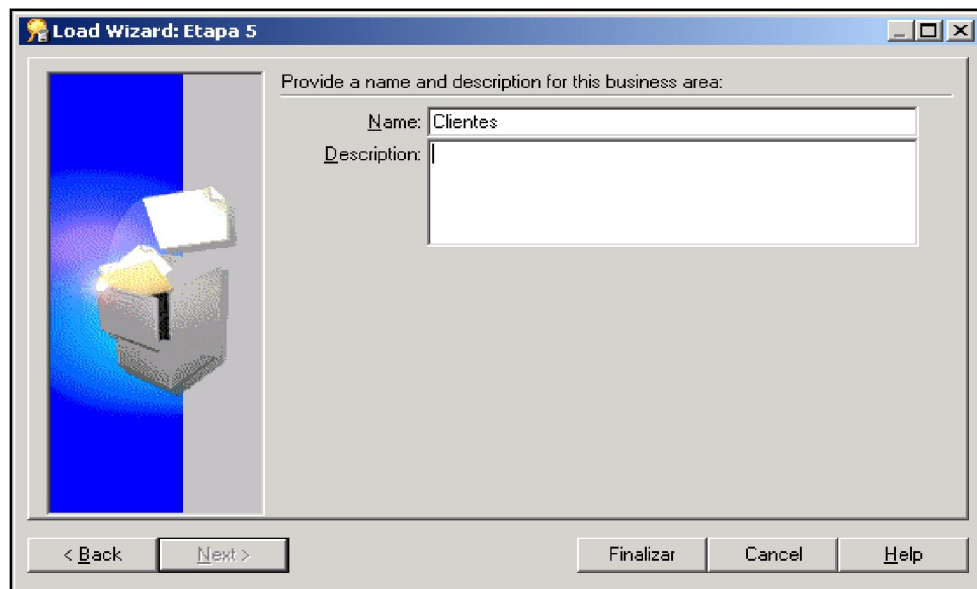


Figura 18: Nomear área de negócio.

Para que as tabelas se relacionem, é necessário criar *Joins* (relacionamentos) entre as tabelas.

Basta escolher o botão para criação do mesmo na barra de ferramentas do *Discoverer* e a Figura 19 será exibida para que o analista escolha a tabela *Master* e a *Detail*.

Podendo assim, criar Junções Mestres ou Junções Detalhes.

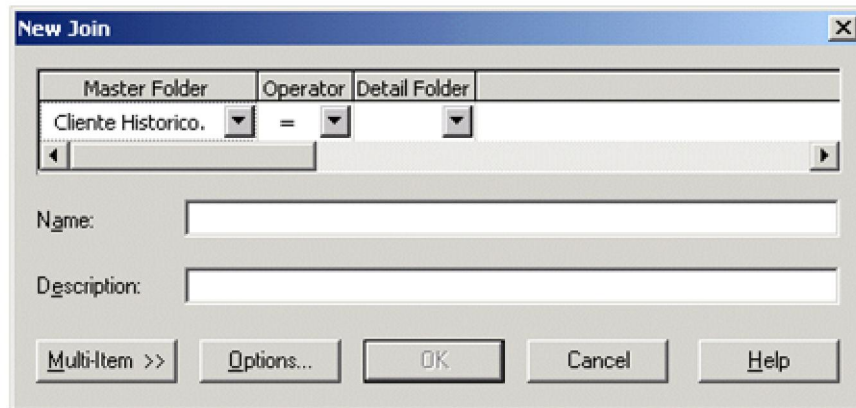


Figura 19: Criando *Joins* entre tabelas.

Sua Área de Negócios finalizada será composta pelas tabelas e *joins*.

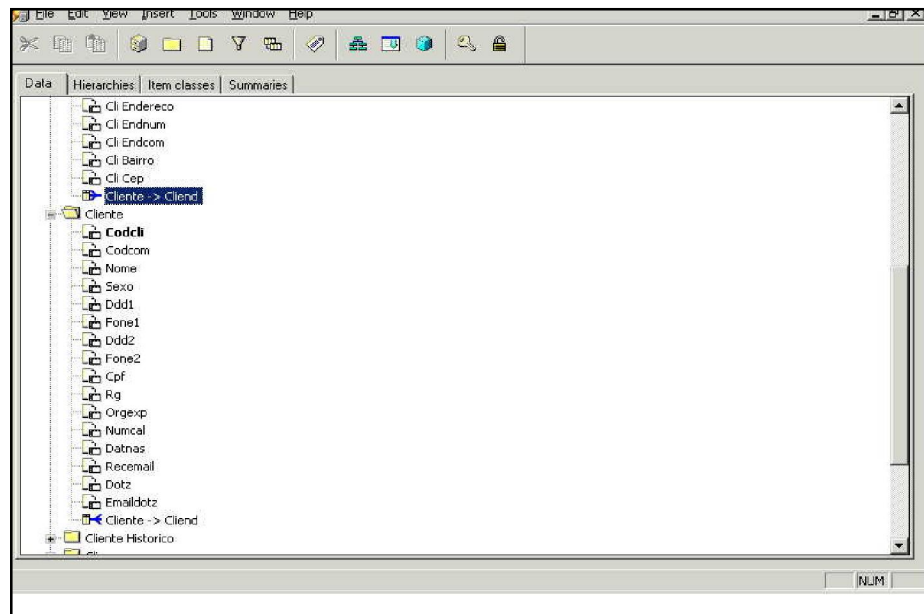


Figura 20: Área de Trabalho finalizada.

Após definir a Área de Negócios e os objetos que a compõe, o analista irá preparar o Caderno de Trabalho para o usuário final. Para isso deverá se logar no *Oracle Discoverer Desktop Edition*.



Figura 21: Login no Discoverer Desktop Edition.

Após se logar, o assistente para criação do Caderno de Trabalho será iniciado. Poderá ser criado um novo Caderno de Trabalho ou Abrir um Caderno já existente, como mostra a Figura 22.

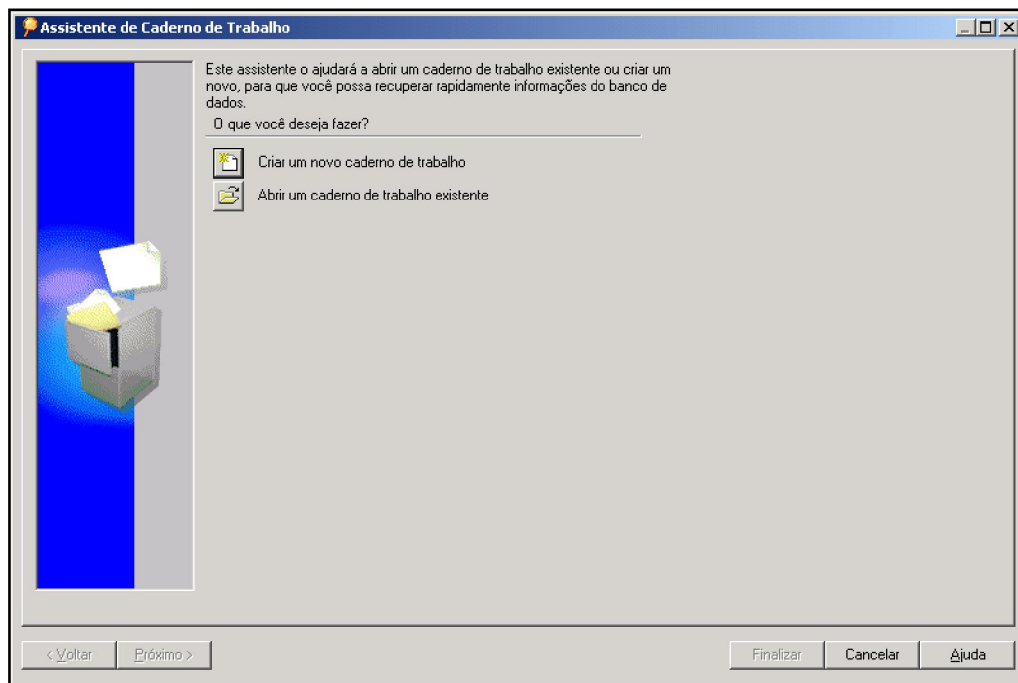


Figura 22: Criando novo caderno de trabalho.

Ao selecionar a opção para Criar um novo Caderno, poderá ser escolhida a forma em que os resultados serão exibidos para o usuário.

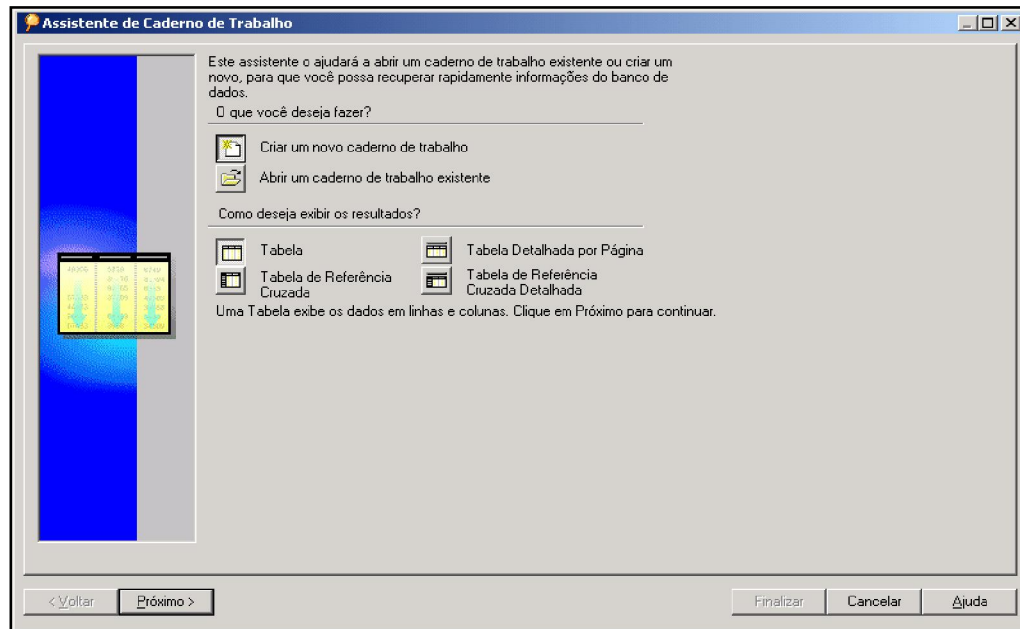


Figura 23: Escolha de formas de exibição de resultados.

A terceira etapa do assistente, permite que seja alterado o layout de itens de sua folha de trabalho.

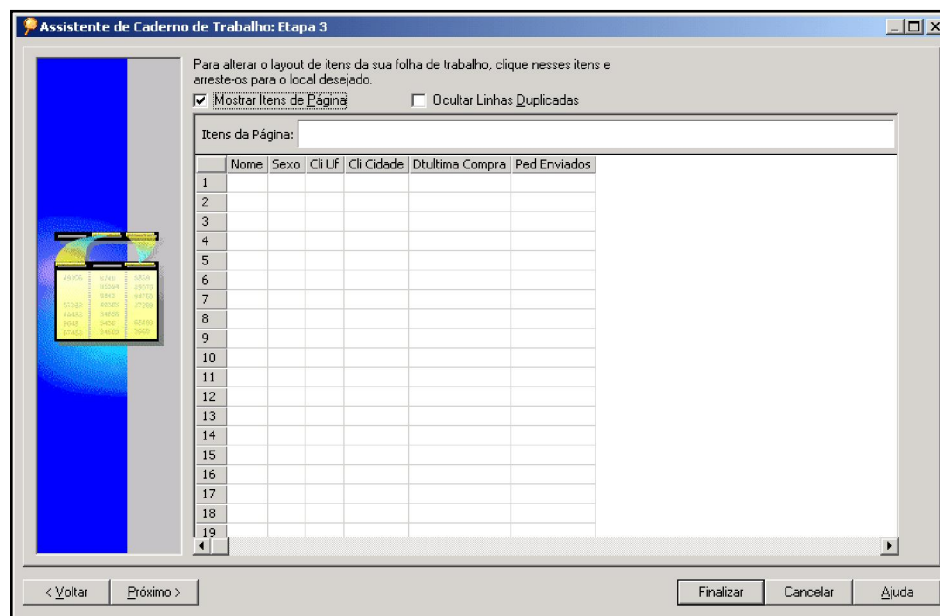


Figura 24: Definindo o Layout.

Em seguida, poderão ser definidas condições para limitar os resultados das folhas de trabalho através de critérios especificados.

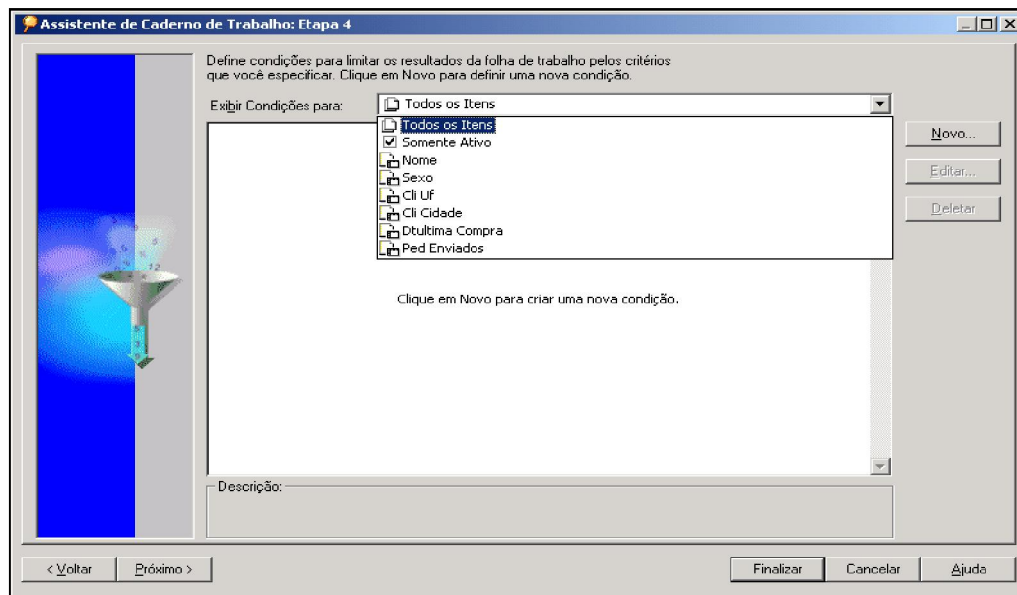


Figura 25: Definindo Condições.

Em seguida, o *Discoverer* possibilita que os dados sejam classificados em ordem crescente ou decrescente.

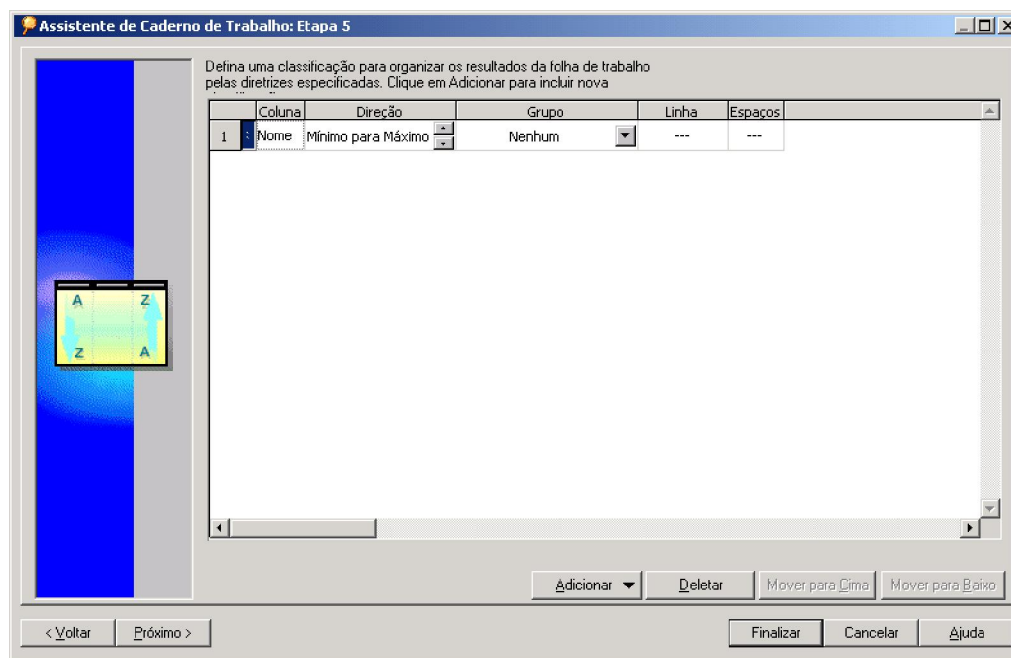


Figura 26: Classificando os dados.

A próxima etapa consiste em definir cálculos adicionando novas informações aos resultados das folhas de trabalho.

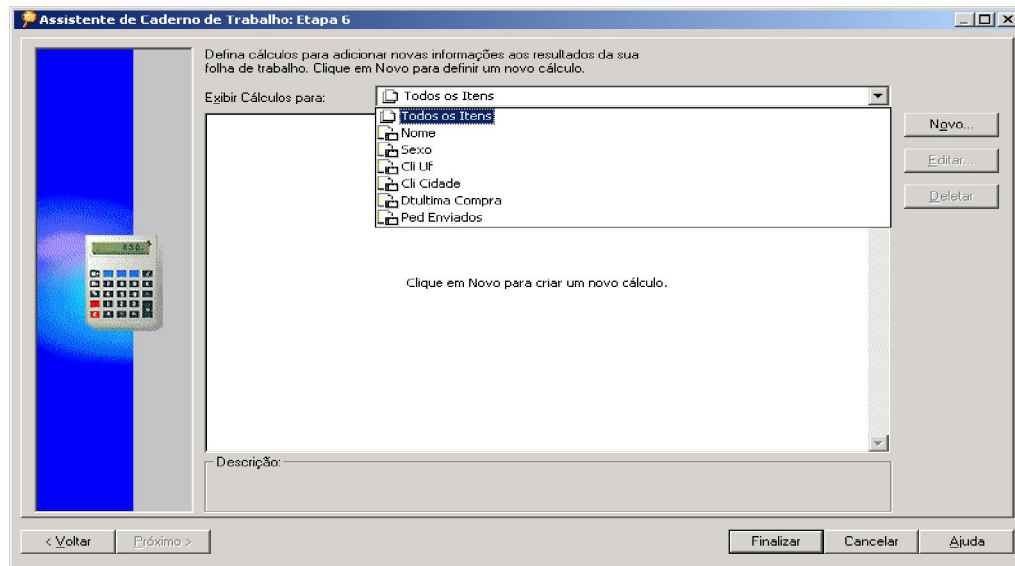


Figura 27: Definindo cálculos.

Finalizar a criação de seu Caderno de Trabalho, o mesmo será exibido como mostrado na Figura 28.

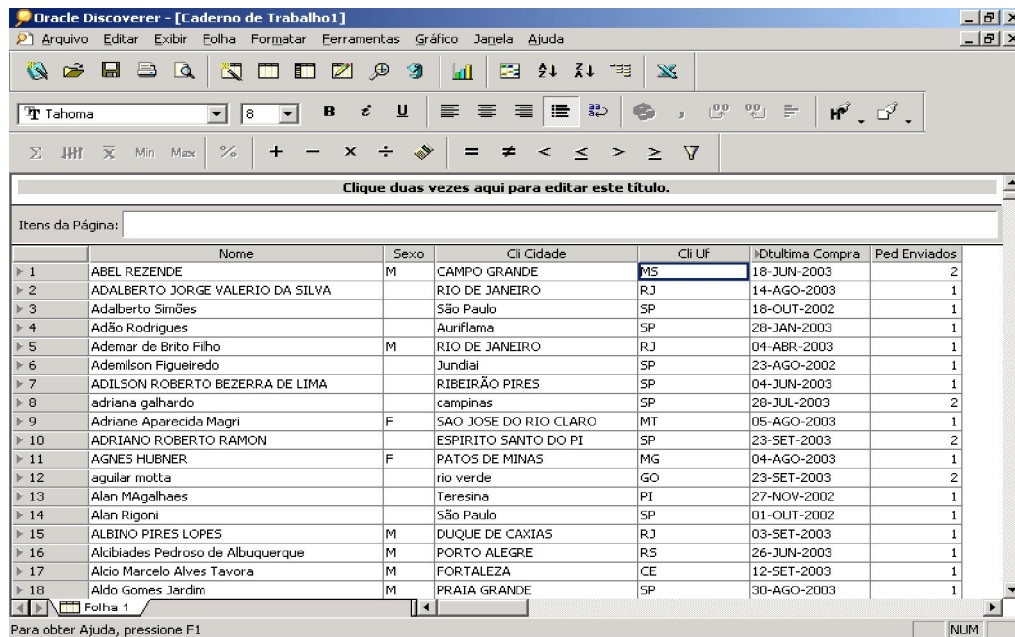


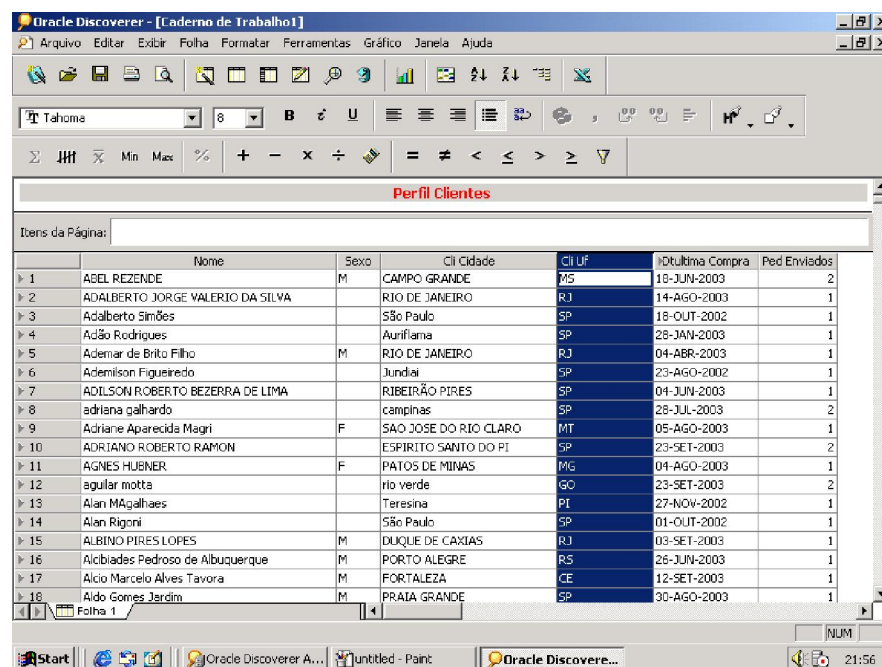
Figura 28: Caderno de Negócios finalizado.

Os Cadernos de trabalho poderão ser armazenados em disco para que o usuário possa utilizá-los posteriormente. Poderão também ser abertos vários cadernos de trabalho ao mesmo tempo, para comparação de valores, e melhor análise das tabelas. E poderão ser formatados, modificando a fonte, cor e tamanho do texto para melhor apresentação.

Para análise dos dados, pode se destacar três tipos básicos de operações que podem ser feitas no *Discoverer Desktop*:

- Pivoteamento (*Pivoting*)
- Ramificação (*Drilling*)
- Classificação (*Sorting Data*)

Abaixo, estaremos mostrando um exemplo do pivoteamento, onde basta arrastar a coluna Cli_UF para o eixo de página (Itens de Página)



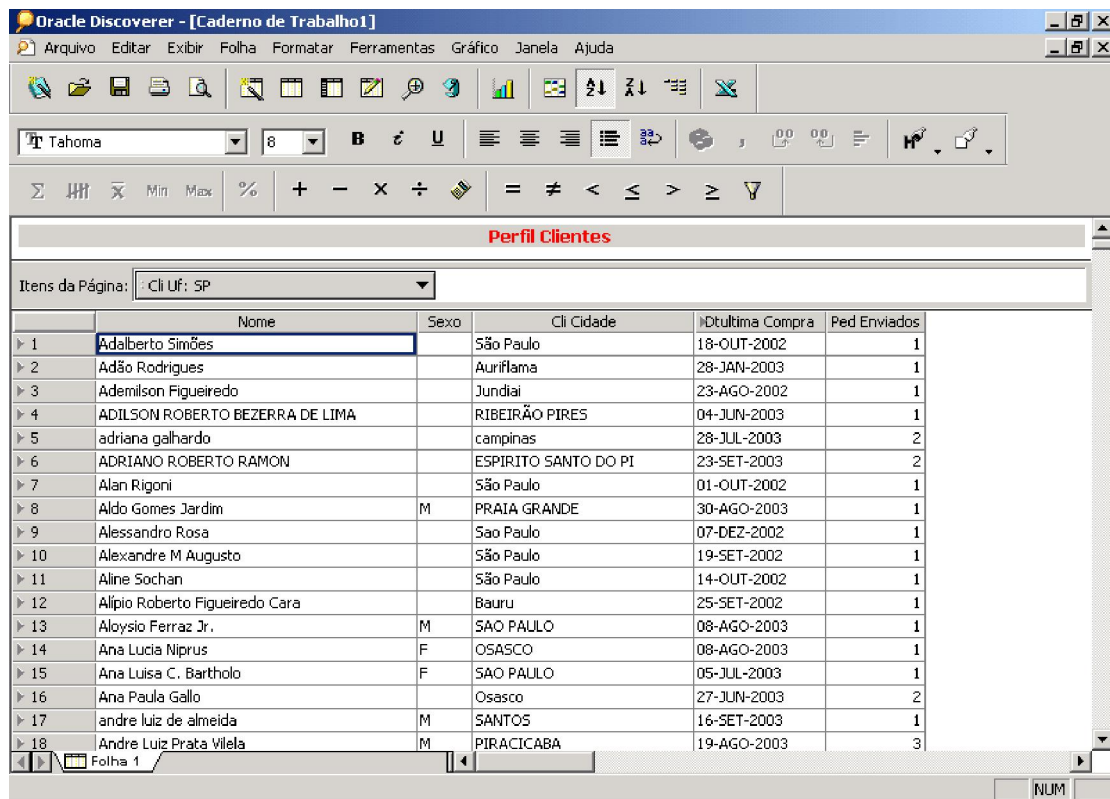
The screenshot shows the Oracle Discoverer Desktop interface. The main window displays a pivot table titled "Perfil Clientes". The columns are: Nome, Sexo, Cli Cidade, Cli UF, Última Compra, and Ped Enviados. The rows are numbered 1 to 18. The "Cli UF" column is highlighted in blue, indicating it is the active pivot field. The "Itens da Página:" (Items of Page:) section is empty, suggesting no specific items are selected for the pivot.

	Nome	Sexo	Cli Cidade	Cli UF	Última Compra	Ped Enviados
1	ABEL REZENDE	M	CAMPO GRANDE	MS	18-JUN-2003	2
2	ADALBERTO JORGE VALERIO DA SILVA		RIO DE JANEIRO	RJ	14-AGO-2003	1
3	Adalberto Simões		São Paulo	SP	18-OUT-2002	1
4	Adão Rodrigues		Auriflama	SP	28-JAN-2003	1
5	Ademar de Brito Filho	M	RIO DE JANEIRO	RJ	04-ABR-2003	1
6	Ademilson Figueiredo		Jundiaí	SP	23-AGO-2002	1
7	ADILSON ROBERTO BEZERRA DE LIMA		RIBEIRÃO PIRES	SP	04-JUN-2003	1
8	adriana galhardo		campinas	SP	28-JUL-2003	2
9	Adriane Aparecida Magri	F	SAO JOSE DO RIO CLARO	MT	05-AGO-2003	1
10	ADRIANO ROBERTO RAMON		ESPIRITO SANTO DO PI	SP	23-SET-2003	2
11	AGNES HUSNER	F	PATOS DE MINAS	MG	04-AGO-2003	1
12	agular mota		rio verde	GO	23-SET-2003	2
13	Alan Magalhães		Teresina	PI	27-NOV-2002	1
14	Alan Rigoni		São Paulo	SP	01-OUT-2002	1
15	ALBINO PIRES LOPES	M	DUQUE DE CAXIAS	RJ	03-SET-2003	1
16	Alcibiades Pedroso de Albuquerque	M	PORTO ALEGRE	RS	26-JUN-2003	1
17	Alcio Marcelo Alves Tavora	M	FORTALEZA	CE	12-SET-2003	1
18	Aldo Gomes Jardim	M	PRAIA GRANDE	SP	30-AGO-2003	1

Figura 29: Criando Pivoteamentos.

No Eixo de Página pode-se selecionar em uma caixa de listagem, o estado que se quer visualizar, como mostrado na Figura 30.

Neste caso, foi selecionado o estado de SP. Sem qualquer código SQL, o usuário acaba de criar uma consulta mostrando clientes residentes no estado de SP.



Oracle Discoverer - [Caderno de Trabalho1]

Arquivo Editar Exibir Folha Formatar Ferramentas Gráfico Janela Ajuda

Tahoma 8 B U

Σ HHI X Mini Max % + - x ÷ = ≠ < ≤ > ≥ V

Perfil Clientes

Itens da Página: Cli Uf: SP

	Nome	Sexo	Cli Cidade	Dtultima Compra	Ped Enviados
1	Adalberto Simões		São Paulo	18-OUT-2002	1
2	Adão Rodrigues		Auriflâma	28-JAN-2003	1
3	Ademilson Figueiredo		Jundiaí	23-AGO-2002	1
4	ADILSON ROBERTO BEZERRA DE LIMA		RIBEIRÃO PIRES	04-JUN-2003	1
5	adriana galhardo		campinas	28-JUL-2003	2
6	ADRIANO ROBERTO RAMON		ESPIRITO SANTO DO PI	23-SET-2003	2
7	Alan Rigoni		São Paulo	01-OUT-2002	1
8	Aldo Gomes Jardim	M	PRAIA GRANDE	30-AGO-2003	1
9	Alessandro Rosa		Sao Paulo	07-DEZ-2002	1
10	Alexandre M Augusto		São Paulo	19-SET-2002	1
11	Aline Sochan		São Paulo	14-OUT-2002	1
12	Alípio Roberto Figueiredo Cara		Bauru	25-SET-2002	1
13	Aloysio Ferraz Jr.	M	SAO PAULO	08-AGO-2003	1
14	Ana Lucia Niprus	F	OSASCO	08-AGO-2003	1
15	Ana Luisa C. Bartholo	F	SAO PAULO	05-JUL-2003	1
16	Ana Paula Gallo		Osasco	27-JUN-2003	2
17	andre luiz de almeida	M	SANTOS	16-SET-2003	1
18	Andre Luiz Prata Vilela	M	PIRACICABA	19-AGO-2003	3

Folha 1

NUM

Figura 30: Trabalhando com Pivoteamentos.

Como pode se verificar, uma consulta é criada facilmente sem que o usuário tenha a necessidade de conhecer linguagens de programação. Vale ressaltar que apesar de fácil o pivoteamento pode produzir resultados errôneos quando feito sem critério.

6. CONCLUSÃO

Conclui-se que *Data Warehouse* não é modismo, tornou-se uma necessidade para as organizações, não resolve os problemas da organização se não houver o fator humano para analisar os resultados e tomar as decisões, com isso é fundamental o envolvimento do usuário em todo o processo (usuário que atua na análise de informações para a tomada de decisão). São poucos os usuários com este perfil. O usuário que tentar atuar nessa área deve evitar trabalhar em nível operacional (detalhe) e sim em nível estratégico (dados consolidados / agregados), a não ser que seja muito necessária a análise detalhada.

O analista de Tecnologia da Informação deve possuir o conhecimento da tecnologia utilizada no projeto principalmente em: SGBD, Modelagem de Dados Relacionais e Multidimensionais.

A avaliação da necessidade de implantar tal sistema deve levar em conta:

- A necessidade da empresa de responder continuamente às mudanças de mercado.
- A existência ou não de uma base consistente de informações oriundas dos bancos de dados operacionais.
- A inviabilidade de obter as informações de suporte a decisão diretamente dos bancos de dados operacionais da empresa.

À medida que a infra-estrutura de informações das empresas amadurece, aumenta a necessidade de qualidade dos dados e de sistemas eficientes e eficazes de suporte à decisão. Esse tipo de sistema e o uso corporativo que se faz dele alavancam o conhecimento de negócio ou sua inteligência, resultando em vantagem competitiva.

No cenário atual onde os administradores das empresas precisam tomar decisões rápidas e certeiras em resposta aos diversos eventos que ocorrem a todo instante em suas empresas, faz-se necessário construir todo um mecanismo de suporte rápido e eficiente aos processos decisórios. O *Data Warehouse* é uma ferramenta apropriada para este tipo de situação. Se bem projetado, permite aos gerentes obter informações rápidas e confiáveis que lhe servirão de suporte para a solução de seus problemas e para atingir os objetivos propostos para os negócios.

REFERÊNCIAS BIBLIOGRÁFICAS

COREY, Michael; ABBEY, Michael; ABRAMSON, Ian; TAUB, Bem. Oracle 8i *Data warehouse*. Rio de Janeiro: Campus, 2001

HAIJINDER, G; RAO, P. C; *The official guide to data warehousing*, Nova York: Que Corporation, 1996.

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

INMON, W.H. *Gerenciando data warehouse*. São Paulo: Makron Books, 1999.

MACHADO, Felipe. *Projeto de data warehouse*. São Paulo: Érica, 2000.

KOTLER, Philip. *Administração de Marketing*. São Paulo: Atlas, 1998

KIMBALL, Ross; *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*", 2002.

PORTER, Michael E.; *Técnicas para Análise Industrial e da Concorrência*. 17ª Edição - Rio de Janeiro. Editora Campus, 1998.

RACKHAM, Neil; *Alcançando a Excelência em Vendas*. 1998. – São Paulo. Editora Makron Books

Fatec Jahu. Aula 9 . Sistemas de Suporte a Decisão. http://robson.fjaunet.com.br/fatec/sis_informacao/aulas/pdf/aula_09_si.pdf. Acessado em 05 Set. 2008

SQL Magazine. Instalando e configurando o Oracle Discoverer. <http://www.devmedia.com.br/articles>. Acessado em 11 Out. 2008

Healt Care Administrators Associates. <http://www.hcaa.org>. Acessado em 25 Jul. 2008