

Mauro Luis da Silva

Vilma de Lira Maia

PROPOSTA DE IMPLEMENTAÇÃO DE DATA WAREHOUSE NA CTEEP

Monografia apresentado à Escola
Politécnica da Universidade de São
Paulo para obtenção do Título de
Especialização em Tecnologia da
Informação.

São Paulo

2002

Mauro Luis da Silva

Vilma de Lira Maia

PROPOSTA DE IMPLEMENTAÇÃO DE DATA WAREHOUSE NA CTEEP

Dissertação apresentada a Escola
Politécnica da Universidade de São
Paulo para obtenção de Especialização
em Tecnologia da Informação.

Área de Concentração:
Tecnologia da Informação

Orientador:
Prof. Dr. Jorge Rady

São Paulo
2002

RESUMO

As informações contidas nos tradicionais sistemas orientados à transação não suprem as necessidades de consultas onde os gerentes 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.

Sabemos que os bancos de dados são de vital importância para a empresa, e também estamos cientes de que sempre foi difícil analisar o conteúdo neles existentes, devido ao enorme volume de informação e o fato de estarem espalhados nos diversos sistemas, entre eles sistemas técnicos e administrativos. O Data Warehouse funciona como um processo de aglutinação dos dados de fontes heterogêneas, incluindo dados atuais, históricos e externos para atender a necessidades de consultas estruturadas, relatórios analíticos e de suporte ao processo de tomada de decisão. Sendo assim o objetivo desse trabalho é abordar e os principais conceitos de Data Warehouse enfocando a necessidade de sua implementação na CTEEP, definindo conceitos, estratégias de implantação, arquitetura e estrutura dessa tecnologia.

ABSTRACT

The information handle by traditional transaction-oriented systems does not comply with the needs of data inquiry and extraction when the managerial team has to access information about long history periods, sometimes covering data from many previous years which certainly will not be available in the system used to handle the daily operations of the company.

Databases are really important to the companies, however it is really hard to analyze the available information on these databases due to different reasons as for example the huge information volume and also the fact of these databases are spread out in different systems and environments, including technical and administrative systems.

Data Warehouse works as an agglutination process of data from different sources, including operational data, history and external data in order to attend the needs of structured inquiries, analytical reports and the called decision support processes. Therefore, the main goal of this work (essay) is explain the main concepts of Data Warehouse technology focusing the necessity of its implementation at CTEEP, defining concepts, implementation strategies, architecture and structure of this specific technology.

SUMÁRIO

1 Introdução	1
1.1 Justificativa	1
1.2 Objetivo do trabalho	4
2 Data Warehouse	6
2.1 Principais Conceitos	6
2.2 Comparação OLTP X OLAP	7
2.3 Definição de Data Warehouse	9
2.3.1 Orientado por assunto	10
2.3.2 Integração	10
2.3.3 Variação no tempo	12
2.3.4 Não volatilidade	12
2.3.5 Credibilidade dos dados	13
2.3.6 Granularidade	13
2.3.7 Esquemas do tipo Estrela e Floco de Neves	13
2.4 Vantagens do Data Warehouse	16
2.5 Transferência de dados	16
3 Estratégias de implantação	18
3.1 Múltiplos Servidores	18
3.2 Estrutura utilizada	19
3.3 Objetivos da Organização	20
3.4 Identificar informações	21
3.5 Ambiente físico e recursos humanos	21
3.6 Tempo de Armazenamento dos dados	22
3.7 Transformação dos dados	22
4 Arquitetura do Data Warehouse	24
4.1 Arquitetura de Dados	24
4.2 Arquitetura centralizada X arquitetura distribuída	25
4.3 Processo de Data Warehouse	26
4.4 Ambiente de processamento	28

5	Estrutura de um Data Warehouse	30
5.1	<i>Níveis de detalhe e sumarização</i>	30
5.2	<i>Metadados</i>	31
5.3	<i>Aspectos de uma estrutura de Data Warehouse</i>	33
6	Necessidades do Data Warehouse na CTEEP	34
6.1	<i>História da Empresa</i>	34
6.2	<i>Principais Sistemas de Informação</i>	34
6.3	<i>Aplicação do Data Warehouse na CTEEP</i>	38
7	Conclusão	40

1 Introdução

1.1 Justificativa

Vivemos hoje uma época de globalização dos negócios e a informação é palavra chave do momento. Sendo assim prover a informação certa, para a pessoa certa, no momento certo é cada vez mais garantia de sucesso do negócio ou até mesmo de sobrevivência da empresa. O mundo competitivo dos negócios está exigindo cada vez mais a tomada de decisão no menor espaço de tempo possível e os sistemas de informação não conseguem atender essas necessidades dos tomadores de decisão que requerem acesso rápido e fácil às informações.

Cada vez mais as empresas se deparam com a necessidade de acelerar o processo de tomada de decisão, mas, para isso, devem reagir com velocidade às mudanças do ambiente, o que se dá através de análise, planejamento e execução de ações práticas e estratégias adequadas. Infelizmente os dados disponíveis dentro de sistemas corporativos quase sempre se encontram espalhados dentro da empresa em plataformas independentes, o que acaba dificultado a integração e a análise dos dados tornando complicado o acesso à essas informações e dificultando as atividades de análise gerencial.

A tecnologia de Data Warehouse foi desenvolvida justamente com a finalidade de auxiliar os administradores, gerentes e executivos na tomada de decisões, oferecendo às organizações uma forma extremamente eficiente e flexível de se obter informações que possibilitem a tomada de decisões no momento apropriado.

O objetivo do desenvolvimento de um Data Warehouse é prover suporte à decisão, não a nível setorial, mas decisões estratégicas dos processos de negócios de interesse da organização. Sendo assim, é importante que outros aspectos sejam considerados durante o seu desenvolvimento, além de processos e dados. A dispersão dos dados, por exemplo, deve ser considerada, pois esses dados estão distribuídos e muitas vezes dispersos geograficamente.

É importante que se compreenda as diferenças entre as razões para construção de um Data Warehouse e os objetivos ditados para o desenvolvimento de sistemas

transacionais. Estes últimos se destinam à produção de relatórios, a consultas e à atualização de seus dados, assegurando a execução de tarefas operacionais de forma otimizada dentro de um setor, enquanto numa solução de Data Warehouse, a real motivação é mais abrangente e deve enfatizar os requisitos de negócio, não se destinando a substituir os relatórios mantidos pelos sistemas transacionais. A solução Data Warehouse possibilita ao usuário a formulação de consultas de uma forma amigável, utilizando para tal a terminologia do negócio.

Um Data Warehouse oferece os fundamentos e os recursos necessários para um Sistema de Apoio à Decisão (SAD) eficiente, fornecendo dados integrados e históricos que servem desde a alta direção até as gerências de nível mais baixo, auxiliando os executivos a obterem o mais rápido possível respostas para perguntas que não são encontradas nos sistemas operacionais e com isso tomar decisões baseadas em fatos.

Sistemas de Apoio à Decisão (SAD) [INM97], também conhecidos como Decision Support Systems (DSS) [POE98], são sistemas que realizam o processamento analítico e provêm as informações necessárias ao usuário, permitindo a análise de situações e a tomada de decisões.

Antes do surgimento do Data Warehouse, o processamento dos Sistemas de Apoio à Decisão ocorria em um único banco de dados ou plataforma, que servia a todos os propósitos. No mesmo ambiente em que as informações eram processadas para atender às necessidades operacionais, eram implementados os Sistemas de Apoio à Decisão, tornando o processamento pouco eficiente para extração e manipulação dos dados. A necessidade de se criar um ambiente próprio para manipular informações vem como resposta a muitos fatores, tanto tecnológicos como econômicos e organizacionais [INM97]. Desta forma, um Data Warehouse provê um banco de dados especializado, que gerencia o fluxo de informações a partir de banco de dados corporativo e fontes de dados externas à empresa.

É comum encontrarmos uma diversidade de sistemas de informação funcionando nas empresas. No caso da CTEEP temos o sistema de gestão empresarial (ERP), o sistema de apoio à manutenção (MANTEC) e o sistema de supervisão e controle (SSC), sistemas estes que suprem as necessidades de cada um dos setores da

organização. Além destes sistemas que estão em uso, existem ainda aqueles que já foram desativados e que, por um longo período de tempo, geraram e armazenaram grandes quantidades de dados que não podem ser simplesmente arquivados e esquecidos para sempre. Abaixo é apresentada a descrição resumida de cada um desses sistemas:

ERP – São sistemas de gestão empresarial caracterizados principalmente por abranger um amplo escopo de funcionalidades, pela integração de seus dados e pela capacidade de adaptação a vários tipos de organização. Tais sistemas melhoram o fluxo de dados nas empresas, facilitam o acesso e propiciam maior produtividade e velocidade de resposta. As funcionalidades dos sistemas ERP são geralmente organizadas em módulos referentes às áreas funcionais das empresas tais como Suprimentos, Recursos Humanos e Financeiro.

MANTEC – É um sistema corporativo de gerenciamento da manutenção de equipamentos que permite um acompanhamento dinâmico da manutenção, visando reduzir os desligamentos de linhas de transmissão e minimizar o número de equipamentos em manutenção. Permite também elaborar planejamento e programações otimizadas com o intuito de cumprir contratos firmados entre a CTEEP (Companhia de Transmissão de Energia Elétrica Paulista), ONS (Operador Nacional do Sistema), ANEEL (Agência Nacional de Energia Elétrica), geradoras e concessionárias de distribuição de energia elétrica.

SSC – É o sistema responsável pela supervisão e controle de toda a rede de transmissão de energia elétrica da empresa. Processa os valores das grandezas elétricas telemedidas através de unidades terminais remotas instaladas nas usinas e subestações do sistema elétrico. Este ambiente é organizado em torno de um gerenciador de base de dados de tempo real de alto desempenho, onde são tratados alarmes (sobreposição, filtragem, organização em listas) e eventos (registro em arquivo). Ele também é responsável pelo controle da ativação e desativação dos processos e pelo tratamento das falhas do sistema elétrico.

Desta forma, a finalidade do Data Warehouse na CTEEP é criar um repositório único de dados com as principais informações dos sistemas ERP, SSC e MANTEC, fornecendo os recursos necessários a um sistema de apoio à decisão.

Os objetivos tradicionais, sob a ótica da Tecnologia da Informação, para a implementação de um projeto de Data Warehouse são disponibilizar acesso à informação, ou seja, através de consultas pré estabelecidas os usuários finais não precisam conhecer o modelo da base de dados, melhorar a qualidade das decisões, utilizando para isso informações confiáveis em um menor espaço de tempo, redução permanente de custos no ciclo de recuperação de informações onde o usuário tem a flexibilidade de elaborar as suas próprias consultas à base de dados, não ficando dependente da área de sistemas.

A solução de Data Warehouse pode trazer maior eficiência ao processo de recuperação de informações quando há a necessidade de incorporar novas informações. Tradicionalmente os processos de extração de informações são muito trabalhosos e, na melhor das hipóteses, têm um custo linear. Quando tratamos de legados mais antigos esse custo aumenta de forma exponencial, crescendo a cada nova necessidade de extração.

1.2 Objetivo do trabalho

Esse trabalho tem como objetivo descrever os principais conceitos de Data Warehouse e como essa tecnologia seria aplicada no ambiente corporativo, sendo que a estrutura está dividida em sete capítulos.

No capítulo 2 são definidos alguns conceitos de Data Warehouse como orientação, integração, volatilidade e granularidade. No capítulo 3 são citadas algumas estratégias de implantação utilizando múltiplos servidores, qual a estrutura utilizada, a importância de se definir previamente as necessidades e os objetivos da organização.

No capítulo 4 é abordado o tipo de arquitetura de um Data Warehouse, focando o tipo de arquitetura de dados, as diferenças entre uma arquitetura centralizada e descentralizada, quais os processos de um Data Warehouse e algumas características do ambiente de processamento. No capítulo 5 é comentado sobre a estrutura de um Data Warehouse, seus principais aspectos, os diferentes níveis que os dados se encontram armazenados e a importância dos metadados.

No capítulo 6, é abordado sobre a necessidade do Data Warehouse na empresa, dando uma visão geral dos sistemas de informação utilizados e como a implantação de um Data Warehouse ajudaria na tomada de decisão.

Finalizando, no capítulo 7 é apresentada a conclusão do trabalho.

2 Data Warehouse

2.1 Principais Conceitos

Um Data Warehouse é um banco de dados que contém dados históricos e atuais, usados para análise e decisões, permitindo que se dê respostas às mais diferentes questões realizadas por executivos. Com o processamento desses dados os executivos podem tomar decisões baseadas em fatos e não apenas em intuições e especulações.

Para entender melhor o que é um *Data Warehouse* convém fazer uma comparação entre ele e os bancos de dados operacionais [DAL99], conforme apresentado na tabela 1.

Características	Bancos de dados Operacionais	Data Warehouse
Objetivo	Operações diárias do negócio Utilizado para dar suporte aos processos do dia a dia da organização.	Analisar o negócio Utilizado para o processo de suporte à decisão.
Uso	Operacional Voltado para rotinas operacionais da organização.	Informativo Informar a situação de um determinado negócio.
Tipo de processamento	OLTP Processamento de transações	OLAP Processamento analítico
Unidade de trabalho	Inclusão, alteração, exclusão. Processos efetuados por registro.	Carga e consulta Processos efetuados por grupo de registros.
Número de usuários	Centenas Usuários sem poder de decisão	Dezenas Usuários com poder de decisão
Tipo de usuário	Operadores	Comunidade gerencial
Granularidade	Detalhados Mantém detalhes de todas as transações diárias.	Detalhados e resumidos Mantém dados sumariados.

Redundância	Não ocorre	Ocorre
	Devido a integridade dos dados.	Pode ocorrer para agilizar as consultas
Integridade	Transação	A cada atualização

Tabela 1 – Comparação entre Banco de Dados Operacionais e *Data Warehouse*.

2.2 Comparação OLTP X OLAP

A idéia de Data Warehouse é integrar os dados internos e externos de uma organização em uma estrutura única permitindo uma melhor utilização dos dados pelos analistas, gerentes e executivos. Uma vez obtida a integração, sistemas OLAP (On-Line Analytical Processing) fornecem mecanismos sofisticados para análise dos dados, tais como: análise de relacionamento entre muitos itens de negócio, comparação de dados agregados através de períodos de tempo hierárquicos e apresentação de dados em diferentes perspectivas.

Os Data Warehouse são projetados para processamento on-line analítico (OLAP) ao invés de processamento transacional on-line (OLTP).

Os sistemas OLTP executam as atividades básicas de inserção, atualização, consulta e deleção de dados em um banco de dados operacional. Devido a essas características os bancos de dados operacionais são ditos de leitura e escrita, sendo permitido ao usuário ler e gravar dados [POE98].

O objetivo de aplicações OLAP que é utilizado, principalmente, em Data Warehouse permite ao usuário comparar os dados de qualquer parte do negócio com os de qualquer outra e definir novas análises, conforme a necessidade, sem precisar acessar vários bancos de dados. Segundo Kimball [KIM98], OLAP é uma tecnologia projetada para permitir acesso e análise multidimensional sobre os vários níveis de negócios da empresa.

As características que diferenciam os sistemas OLAP e OLTP são apresentadas através da tabela 2.

OLTP	OLAP
Baseado em aplicações	Baseado em assuntos ou negócios
Processamento baseado em transações diárias, como inclusões, alterações e consultas.	Processamento baseado em consultar informações sobre temas específicos para o negócio da empresa
Atendem a comunidade funcional	Atendem a comunidade gerencial
A grande maioria dos usuários são funcionários com poder de decisão menor.	Usuários com maior poder de decisão.
Contínua atualização (dados voláteis)	Dados não são atualizados (dados não voláteis)
No processamento diário inclui rotinas como inclusão, alteração e exclusão de informações.	Os dados só são atualizados quando se faz necessário nova carga.
Acessados um registro por vez	Acessados um conjunto por vez
Na operação do sistema o usuário na maioria das vezes está manipulando apenas um registro.	Para efetuar análises, são executadas consultas em um conjunto de registros.
Voltados para transações	Voltados para análise
Tem como objetivo efetuar as diversas transações operacionais (inserção, alteração, exclusão).	O foco principal é prover consultas para análise gerencial.
Não contempla a redundância	É permitida a redundância
Visando manter a integridade dos dados não permite redundância.	Data Warehouses distribuídos usualmente envolvem dados redundantes e como consequência os processos de carga e atualização são mais complexos.
Normalmente relacional	Normalmente multidimensional
Processamento centrado em entidades, relacionamentos, decomposição funcional e análise das transações	Centrado em fatos, dimensões e hierarquias, sobre os quais são realizadas diversas agregações com o objetivo de responder questões como total de vendas para cada produto por trimestre.

Dados atuais	Dados atuais e históricos
Processam geralmente dados mais atuais.	Para gerar informações para tomada de decisão, utilizam dados atuais e históricos.
Orientados ao processo	Orientados ao negócio
Usado tradicionalmente para coletar, processar e gerar relatórios de dados operacionais.	Destinam-se a disponibiliza informações sobre o negócio da empresa, para a tomada de decisão.

Tabela 2 - Resumo das diferenças entre OLTP e OLAP [DAL99]

2.3 Definição de Data Warehouse

Um Data Warehouse é projetado para suportar informações escondidas na enorme quantidade de dados de uma empresa. Ao longo do tempo os sistemas de informações são desenvolvidos e implementados visando o controle de um determinado processo na empresa.

O objetivo do Data Warehouse é servir de base para as ferramentas de suporte à decisão, concentrando todas os dados, estratégicos e históricos, extraídos dos sistemas transacionais.

O armazenamento de dados ganhou muito impulso a partir do sucesso de software de sistemas de gerenciamento de dados relacionais (RDBMS), combinando custos cada vez menores e desempenho cada vez melhor.

Abaixo são apresentados alguns conceitos de Data Warehouse na visão de alguns autores.

Segundo Kimball [KIM98], Data Warehouse é a fonte de dados de consulta do empreendimento.

Segundo Inmon [INM97], que é considerado o pai do conceito, Data Warehouse é uma coleção de dados orientados por assunto, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo gerencial de tomada de decisão.

Para Barquini [BAR96], Data Warehouse é uma coleção de técnicas e tecnologias que juntas disponibilizam um enfoque pragmático e sistemático para tratar com o problema do usuário final de acessar informações que estão distribuídas em vários sistemas da organização.

A seguir estão detalhadas as principais características do Data Warehouse no conceito de Inmom [INM97].

2.3.1 Orientado por assunto

É uma característica bastante marcante uma vez que toda a modelagem é baseada em torno dos principais assuntos da empresa, enquanto todos os sistemas corporativos estão voltados para processos e aplicações específicas. Os assuntos são o conjunto de informações relativas a determinada área estratégica da empresa.

Deve-se discutir com o usuário quais os seus objetivos, definir as informações relevantes para o processo de análise, além de se preocupar com os tipos de consultas que serão realizadas sobre os dados do Data Warehouse antes de construí-lo.

2.3.2 Integração

A integração talvez seja a parte mais importante desse processo, pois ela será responsável por sincronizar os dados de todos os sistemas existentes na empresa, e colocá-los no mesmo padrão. O Data Warehouse deve armazenar dados de vários sistemas da empresa, e em alguns casos, dados externos.

Geralmente os dados não estão padronizados, devido a diversos problemas, e é necessário integrá-los antes da carga para o Data Warehouse. Um exemplo clássico é o do sexo, onde num sistema esse dado está guardado no formato **m** para masculino, e **f** para feminino, já no outro, o mesmo dado está guardado, como **0** para masculino e **1** para feminino (Figura 01). Isso geraria um grande problema na hora da análise, porém na fase de Extração, Transformação e Carga, todos os formatos devem ser convertidos num único padrão, e então carregado no Data Warehouse.

O Data Warehouse permite ter uma base de dados integrada e histórica para análise dos dados, e isso pode e deve se tornar um diferencial competitivo para as empresas. Tendo uma ferramenta desse porte na mão, o executivo pode decidir com muito mais eficiência qual a melhor solução para um determinado problema.

Quanto mais forem os diferentes sistemas transacionais a fazerem parte do Data Warehouse, mais complicada e trabalhosa será a integração dos dados.

Independente das dificuldades encontradas, os dados precisam ser armazenados em um formato único e aceitável, para que o usuário, ao acessar o Data Warehouse, não precise se preocupar com a credibilidade e consistência dos dados armazenados, mas sim com a análise das informações obtidas.

INTEGRAÇÃO

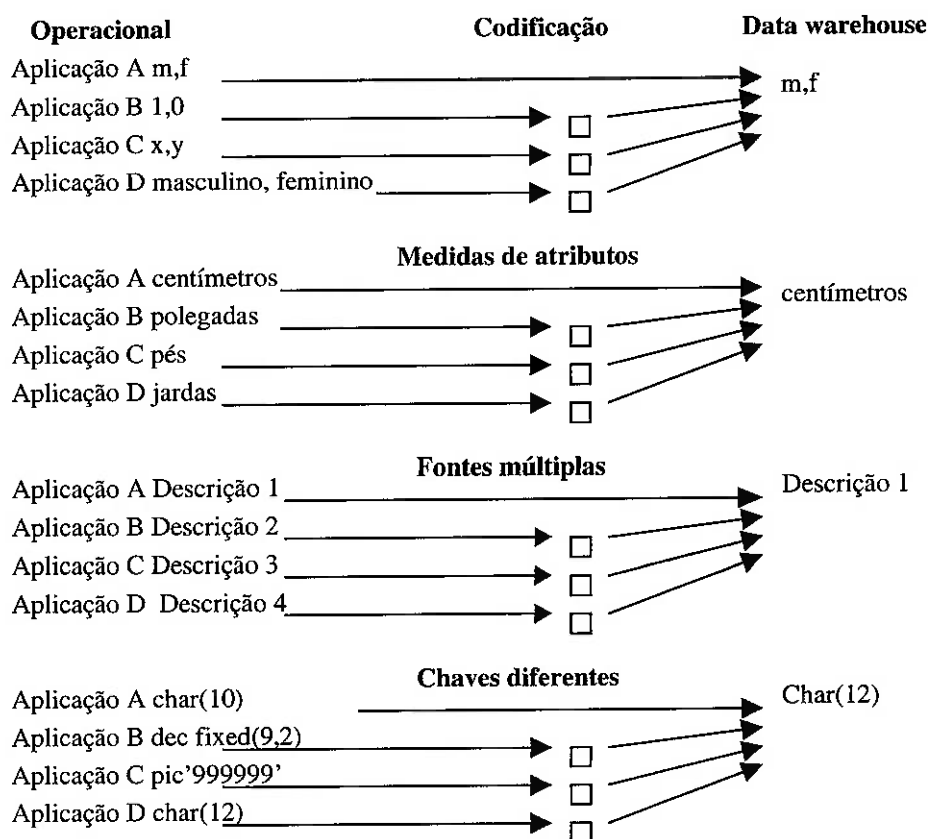


Figura 01 – A questão da integração dos dados [HARR01]

2.3.3 *Variação no tempo*

De acordo com Inmom [INM97] todos os dados no Data Warehouse têm validade em algum instante no tempo. Por isso é que se diz que esses dados “variam com o tempo”.

Em um Data Warehouse é normal que as informações sejam representadas em horizontes de tempo maiores que cinco anos chegando até o limite da idade dos dados ou em um período considerado satisfatório, conforme a sua aplicação. Normalmente nas aplicações operacionais o período de tempo é muito mais curto, pois existe a necessidade de uma resposta rápida para atender as exigências das tarefas diárias, o que só se consegue com o processamento de poucos dados.

Assim como os dados, os metadados, que incluem definições dos itens de dados, rotinas de validação, algoritmos de derivação, etc, também possuem elementos temporais para que, com eventuais mudanças nas regras do negócio a empresa não perca dados históricos. Dados armazenados corretamente no Data Warehouse não são mais atualizados, obtendo-se uma imagem fiel da época em que foram gerados.

2.3.4 *Não volatilidade*

Significa que o *Data Warehouse* permite apenas a carga inicial e as consultas dos dados. Após serem integrados e transformados, os dados são carregados em bloco para o *Data Warehouse*, para que estejam disponíveis aos usuários para acesso. No ambiente operacional, ao contrário, os dados são, em geral, atualizados registro a registro, em múltiplas transações. Esta volatilidade requer um trabalho considerável para assegurar integridade e consistência. Um *Data Warehouse* não requer este grau de controle típico dos sistemas orientados a transações [CAM99]

A maneira como os dados são carregados e tratados é completamente diferente dos sistemas operacionais. Nesses sistemas existem vários controles e atualizações de registros. No Data Warehouse existem somente inserções de dados e consultas.

2.3.5 Credibilidade dos dados

A credibilidade dos dados é de longe o fator mais importante para o sucesso de qualquer projeto. Mesmo a mais simples divergência pode causar sérios problemas quando se trata de tomada de decisão. Dados não confiáveis podem resultar em relatórios inúteis. Se dados de péssima qualidade são disponibilizados em um Data Warehouse o seu resultado final será um suporte à decisão de baixo nível e com alto risco para o negócio da empresa.

2.3.6 Granularidade

O mais importante aspecto do projeto de um Data Warehouse é a questão da granularidade. A granularidade diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no Data Warehouse. Quanto mais detalhes, mais baixo o nível de granularidade, quanto menos detalhes, mais alto o nível de granularidade.

Quando se tem um nível de granularidade muito alto, o espaço em disco e o número de índices necessários se tornam bem menores, mas, em consequência disso há uma correspondente diminuição da possibilidade de utilização dos dados para atender a consultas detalhadas.

Quanto menor o nível de granularidade maior a possibilidade de responder às consultas. No entanto, a quantidade de recursos computacionais aumenta quando as perguntas são muito específicas. Em virtude disso, o balanceamento do nível de granularidade é um dos aspectos mais críticos do planejamento de um Data Warehouse, uma vez que na maior parte do tempo há uma grande demanda por eficiência no armazenamento e no acesso aos dados.

2.3.7 Esquemas do tipo Estrela e Floco de Neves

O modelo lógico mais comum utilizado em um Data Warehouse é o modelo dimensional, também chamado de star join schema.

Em um esquema do tipo estrela as instâncias são armazenadas em uma tabela contendo o identificador de instância (com uma chave de múltiplas partes), valores

das dimensões descritivas para cada instância, e valores dos fatos, ou medidas, para aquela instância (tabela de fatos).

Além disso, pelo menos uma tabela é usada, para cada dimensão, para armazenar dados sobre a dimensão (tabela de dimensão). Este esquema é chamado de estrela, por apresentar a tabela de fatos “dominante” no centro do esquema e as tabelas de dimensões nas extremidades. A tabela de fatos é ligada às demais tabelas por múltiplas junções, enquanto as tabelas de dimensões se ligam apenas à tabela central por uma única junção.

A tabela de fatos é onde as medidas numéricas do fato representado estão armazenadas. Cada uma destas medidas é tomada segundo a interseção de todas as dimensões. A figura 2 ilustra o esquema tipo estrela.

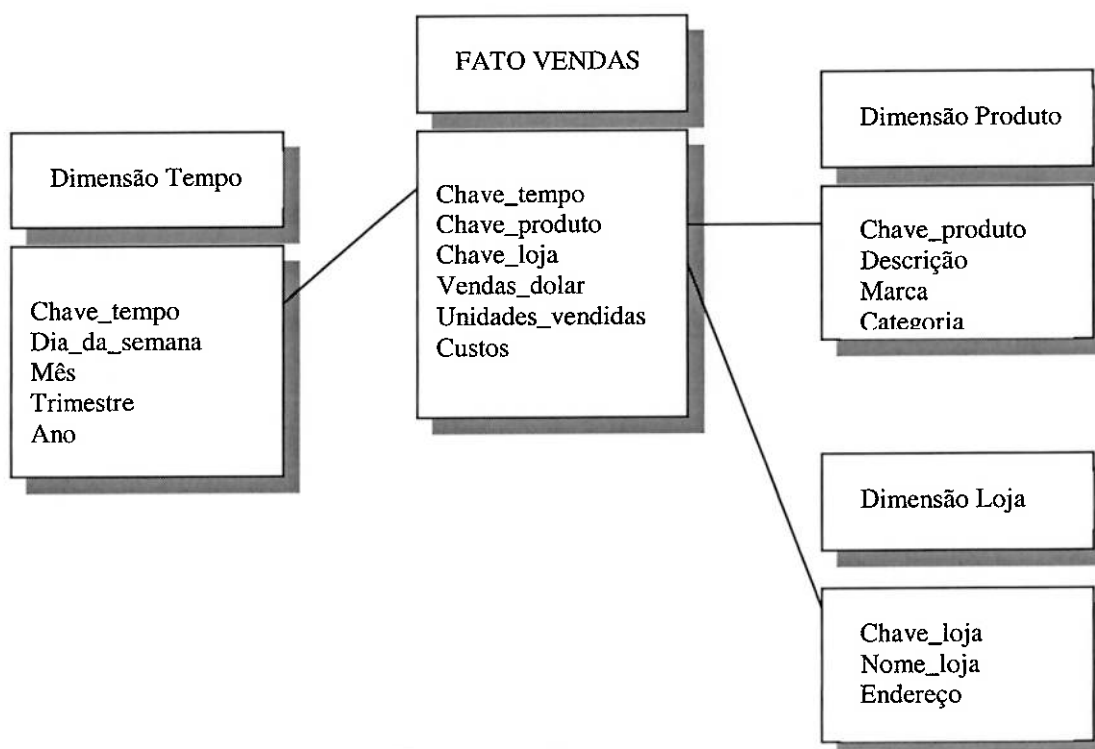


Fig.2 - Esquema Estrela [CAM99]

Outro tipo de estrutura bastante comum é o esquema do tipo floco de neve ou “snow flake”, que consiste em uma extensão do esquema estrela onde cada uma das “pontas” da estrela passa a ser o centro de outras estrelas. Isto porque cada tabela de dimensão seria normalizada, dividindo a tabela original ao longo de hierarquias

existentes em seus atributos. Na fig. 3, a dimensão produto possui uma hierarquia definida onde categoria se divide em marca, e marca se divide em produto. Da mesma forma, a dimensão tempo inclui ano que contém mês, e mês que contém dia-mês. Cada um deste relacionamento muitos-para-1 geraria uma nova tabela em um esquema floco de neve.

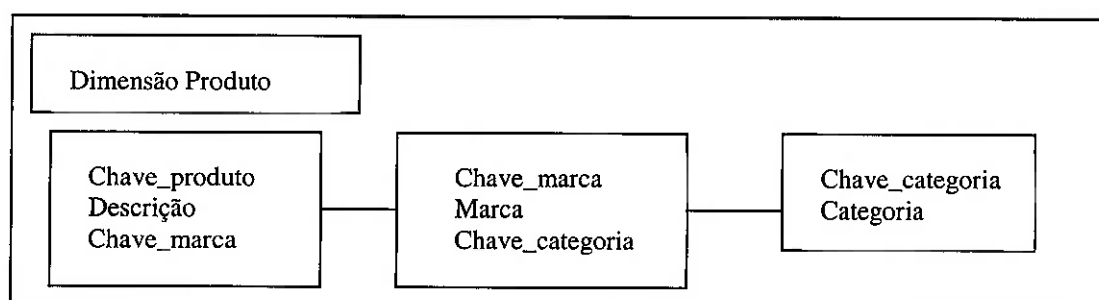


Fig.3 - Esquema Floco de Neve [CAM99]

A principal desvantagem do esquema snowflake em relação ao star é a relativa complexidade da estrutura de dados snowflake normalizada. Se os usuários estiverem gerando consultas com ferramentas de consulta típicas e ad hoc, será mais difícil navegar através do snowflake. Adicionalmente, programas de carga e a manutenção geral ficam mais difíceis de administrar à medida que o modelo de dados se torna mais complexo.

2.4 Vantagens do Data Warehouse

A grande vantagem de um Data Warehouse é permitir a tomada de decisões baseada em fatos. Esta tecnologia começa a se delinear como uma ferramenta de gestão obrigatória para as empresas sobreviverem nos próximos anos. O ambiente de negócios é crescentemente dinâmico e, à medida que as regras de negócio são incorporadas às aplicações, exige-se uma rapidez cada vez maior nas respostas.

Alguns dos benefícios proporcionados pelo Data Warehouse são: maior tempo de análise para a tomada de decisão; eliminação de tarefas operacionais como pesquisa e identificação dos dados necessários ao processo decisório; melhor confiabilidade das informações e uma democratização das informações sobre o desempenho dos negócios.

Uma solução de Data Warehouse bem desenhada objetiva satisfazer às necessidades de análise de informações dos usuários, como monitorar e comparar as situações atuais com as passadas, e prever as futuras. Ao reunir as informações dispersas pelos diversos bancos de dados operacionais da empresa, possibilita-se que sejam feitas análises bastante eficazes, transformando os dados armazenados em informações que antes eram inacessíveis.

Ao contrário de um banco de dados operacional, o objetivo é que um Data Warehouse contenha somente os dados necessários para o suporte ao processo decisório. Os dados são armazenados e usados para fazer comparações, verificar tendências e fazer previsões.

Os administradores de dados devem estabelecer as diretrizes para administrar os recursos de dados a serem incorporados. Estas diretrizes incluem definir padrões para os elementos e para a modelagem de dados e definir políticas de compartilhamento dos mesmos com grupos diferentes dentro e fora da organização.

2.5 Transferência de dados

À medida que os dados vão sendo inseridos nos sistemas transacionais, ou seja, nos sistemas que armazenam as informações das transações diárias da empresa, o Data

Warehouse vai sendo municiado destas informações nas mais diversas formas, onde estes dados devem passar por alguns processos antes da carga, tais como extração, transformação e agregação :

Extração – As várias alternativas para extração permitem balancear desempenho, restrições de tempo e de armazenamento. Por exemplo, se a fonte for um banco de dados on-line, pode-se submeter uma consulta diretamente ao banco para criar os arquivos de extração. As rotinas de extração devem ser capazes de isolar somente aqueles dados que foram inseridos e atualizados desde a última extração.

Transformação e filtros – Uma vez que os dados são extraídos e colocados na área de trabalho temporária, eles devem passar por uma série de tratamentos. O primeiro destes tratamentos refere-se à limpeza ou filtragem dos dados, onde o objetivo é garantir a integridade dos mesmos através de rotinas que tentam identificar anomalias e resolvê-las, deixando os dados em um estado consistente antes de serem inseridos no Data Warehouse. A correção de erros de digitação, a descoberta de violação de integridade, a substituição de caracteres desconhecidos e a padronização de abreviações, podem ser exemplos de limpeza de dados.

O próximo passo é colocar os dados em uma forma homogênea, verificando alguns conflitos típicos de domínio de atributo como: diferenças de unidades (quando as unidades utilizadas diferem, embora forneçam a mesma informação), diferença de precisão (quando a precisão escolhida varia de um ambiente para o outro), diferenças em código ou expressões (quando o código utilizado difere um do outro, como no caso de sexo representado por m ou f e por 1 e 0) [DAL99].

Agregação - Processo pelo qual os dados de nível baixo de detalhe são previamente sumarizados e incluídos em tabelas que armazenam informações sumarizadas. [INM97]. Para dados comumente acessados a pré-sumarização é bastante útil. Ela permite usar os resultados intermediários ou “agregados” reduzindo significativamente os recursos requeridos para fornecer o resultado final [HAR01].

3 Estratégias de implantação

Um Data Warehouse não pode ser simplesmente comprado e instalado e por isso alguns passos lógicos para a implantação de um Data Warehouse devem ser seguidos.

Deve-se coletar e analisar requisitos do negócio, ou seja, o Data Warehouse deve ir de encontro às necessidades de negócio da empresa, criar um módulo de dados e um plano físico para o desenho do Data Warehouse.

O módulo de dados seria uma área auxiliar onde os dados recebidos dos outros sistemas transacionais seriam tratados antes de serem migrados para o Data Warehouse.. Devem ainda se definir a origem dos dados, ou seja, quais os bancos de dados operacionais que fornecerão dados para serem carregados no Data Warehouse, escolher a tecnologia do banco de dados e a plataforma a ser utilizada, extrair dados do banco de dados operacional, transformá-los, limpá-los e carregá-los para o Data Warehouse.

Por último deve-se escolher o acesso ao banco de dados e ferramentas de acesso, escolher o software de conectividade de banco de dado, escolher o tipo de análise de dados e o software de apresentação e como será feita a atualização dos dados.

Nos próximos itens são descritos os principais aspectos envolvidos na implantação de um Data Warehouse.

3.1 Múltiplos Servidores

Uma estratégia de implantação é a de divisão do Data Warehouse em múltiplos servidores. Essa divisão pode ser por região geográfica, unidade, tempo e função.

Deve-se tomar muito cuidado com a replicação de dados. Quando uma grande parte dos usuários necessita uma pequena parte do banco de dados corporativo, a replicação de dados ou movê-los para um banco de dados localizado pode ser uma solução mais útil do que um Data Warehouse.

3.2 Estrutura utilizada

Para que a implementação de um Data Warehouse seja um sucesso é necessário que se tenha em mente a estrutura que será utilizada. O processo é semelhante à construção de uma casa, ou seja, nunca devemos construir uma casa sem que haja uma planta da mesma a ser seguida. Sendo assim o gerente do projeto de Data Warehouse deve ficar bastante atento e considerar com muito zelo qual estrutura será utilizada na sua construção, considerando principalmente os aspectos abaixo descritos:

Planejamento - Devem ser identificados quais os problemas de negócios a serem resolvidos e fornecer um processo estruturado, que é o primeiro passo crítico na construção de um Data Warehouse. Estes serviços podem ser independentes ou feitos concorrentemente.

Objetivo - Este passo é necessário para proteger o negócio da tentativa de implementar uma solução para a qual ela não está preparada ou que pode influenciar outras áreas funcionais da companhia não incluídas no planejamento. O plano do projeto de implementação deve ser ajustado com base nos resultados das avaliações.

Suporte e Incremento - O suporte e o incremento do Data Warehouse incluem uma série de processos operacionais e de valores que dão suporte às operações e à manutenção de um Data Warehouse.

Estes processos têm como propósito dar suporte diário aos usuários, esclarecendo dúvidas quanto à sua utilidade e desempenho, ajudar a ampliar o uso da solução e expandir o sistema, possibilitando incluir novas aplicações, novos usuários ou novos dados, aumentando o uso da solução através de orientação aos usuários finais e, finalmente, ajudar a manter o sistema continuamente atualizado e em crescimento, dando melhor suporte às decisões de negócios no planejamento e meios de controle para a entrega dos valores de negócios.

Necessidades do Negócio - A implantação de um Data Warehouse deve vir de encontro às necessidades de negócio da empresa. Uma estratégia recomendada é a criação de um Data Warehouse central a partir de Data Marts (que são um subconjunto do Data Warehouse, desempenhando o papel de um Data Warehouse

departamental [HAR01]) sendo que cada um seria desenvolvido à medida das necessidades de negócio.

Projetar um Data Warehouse que atenda às necessidades corporativas não é uma tarefa fácil. Com toda a evolução tecnológica atual, ainda existem muitas limitações que devem ser respeitadas. A chave para o sucesso na elaboração de um Data Warehouse está em se estabelecer adequadamente os objetivos. É necessário que sejam definidas as metas reais a serem alcançadas com sua implantação.

As necessidades de informações por parte dos gerentes variam com o passar do tempo, sempre se adequando às necessidades que são impostas pelas interferências ocorridas através dos diversos eventos internos e externos que ocorrem e afetam a empresa. A elaboração do projeto do Data Warehouse deve prever esta dinamicidade das necessidades de informações, ou seja, deve ser bastante flexível e permitir modelar os dados de acordo com as necessidades da empresa.

3.3 Objetivos da Organização

É fundamental o uso de um método para levantamento e modelagem dos indicadores de desempenho da empresa, sendo necessário monitorar as metas propostas que justificarão o investimento a ser realizado e definir se o Data Warehouse está sendo bem projetado.

O projeto de um Data Warehouse deve ser dividido em fases. Primeiro identificam-se os objetivos da organização, uma vez que, toda empresa deve ter uma missão, que é a sua razão de ser e determina o seu ramo de negócio. Em seguida devem-se definir os objetivos executivos, os quais representam a situação futura que a organização deseja atingir e por último definem-se os objetivos estratégicos que são objetivos relacionados às áreas funcionais da empresa e subordinados aos objetivos executivos. Os objetivos estratégicos têm como meta otimizar, aumentar, maximizar, expandir ou aprimorar o negócio da empresa.

3.4 Identificar informações

Identificar as informações que são necessárias para dar o suporte necessário aos interesses corporativos é uma tarefa fundamental para o sucesso de um projeto de Data Warehouse. Deve-se identificar onde essas informações serão obtidas, como serão modeladas, armazenadas e recuperadas.

Como os negócios estão se tornando cada vez mais dinâmicos e haverá uma acumulação gradativa de dados armazenados, assim como poderá haver um aumento também gradativo de consultas no Data Warehouse, deve-se prever os gastos com a evolução tecnológica que darão suporte às novas necessidades.

A equipe do projeto de um Data Warehouse deve ser constituída de pessoas tanto da área de negócios quanto da área de informática (desenvolvimento de sistemas ou similar) que devem trabalhar em conjunto para encontrar a melhor solução que atenda aos propósitos para os quais o Data Warehouse está sendo projetado.

Um projeto de Data Warehouse deve ser implementado gradualmente, sendo constantemente refinado à medida que a área de tecnologia e a própria organização aprendam a usá-lo. Alguns problemas inesperados poderão aparecer.

O problema da segurança das informações é um fator que não pode deixar de ser abordado na elaboração de um projeto de Data Warehouse. Nele devem estar previstas normas de acesso e de utilização dos dados, os termos de compromisso que serão assumidos pelos usuários e pelo pessoal de suporte técnico, e os esquemas de auditoria do sistema.

3.5 Ambiente físico e recursos humanos

Outro aspecto a ser abordado no projeto do Data Warehouse é com relação ao ambiente físico no qual serão instalados os equipamentos, tanto os servidores quanto os terminais de consulta e pesquisa. Deve-se ter a preocupação de fornecer as condições recomendadas pelo fabricante para o perfeito funcionamento de todo o sistema, sem gerar futuras surpresas desagradáveis.

Um outro assunto a ser abordado no projeto de um Data Warehouse deve ser o preparo dos recursos humanos que utilizarão o sistema, assim como os que darão suporte ao mesmo. Muitas vezes, as empresas não fazem esta previsão no projeto, não alocam recursos financeiros para tal, e depois se vêem em dificuldades para proporcionar o conhecimento e o treinamento necessário para a utilização ou suporte eficaz do sistema.

3.6 Tempo de Armazenamento dos dados

Com relação ao tempo em que os dados devem ser armazenados percebe-se que quando dados muito antigos são utilizados para análise pode-se produzir uma resposta que não é necessariamente condizente com a realidade atual da empresa e induzir a uma decisão que não é a mais acertada. Porém, o armazenamento somente com os dados mais recentes também pode levar a resultados errôneos. Portanto esta definição é muito importante, e deverá ser amplamente discutida com os projetistas do Data Warehouse, com os analistas de negócios e com os futuros usuários do sistema.

3.7 Transformação dos dados

O Data Warehouse recupera dados de uma variedade de bancos de dados operacionais heterogêneos. Os dados são transformados e entregues ao armazém de dados baseado em um modelo selecionado (ou definição traçada). A transformação dos dados e a movimentação dos processos são executadas sempre que uma atualização para os dados de warehouse for requisitada e por isso deve existir alguma forma de automatização para administrar e executar estas funções. As informações que descrevem o modelo e a definição dos elementos de fonte de dados são chamadas metadados.

A fase de limpeza tem que ser tão dinâmica quanto possível para acomodar todos os tipos de consultas até mesmo aquelas que podem requerer informação de baixo nível.

O dados devem ser extraídos de fontes de produção em intervalos regulares e serem agrupados centralmente, mas o processo de limpeza tem que remover duplicação e reconciliar diferenças entre vários estilos de coleção de dados. Uma vez que os dados estejam limpos, eles devem então serem transferidos para o Data Warehouse.

O volume de informações é outro aspecto importante de Data Warehouse por causa da complexidade envolvida no processamento de consulta ad-hoc e por causa das vastas quantidades de dados que a organização deseja usar no Data Warehouse.

4 Arquitetura do Data Warehouse

O Data Warehouse deve ser capaz de responder a consultas complexas de maneira rápida sem deixar de mostrar detalhes relevantes, mas, para isso ele deve possuir uma arquitetura que permita coletar, manipular e apresentar os dados de forma eficiente e rápida.

4.1 Arquitetura de Dados

A figura 04 ilustra vários componentes de uma arquitetura de dados que devem ser considerados antes de se implementar qualquer Data Warehouse, identificando como os dados serão movidos ao longo do sistema.

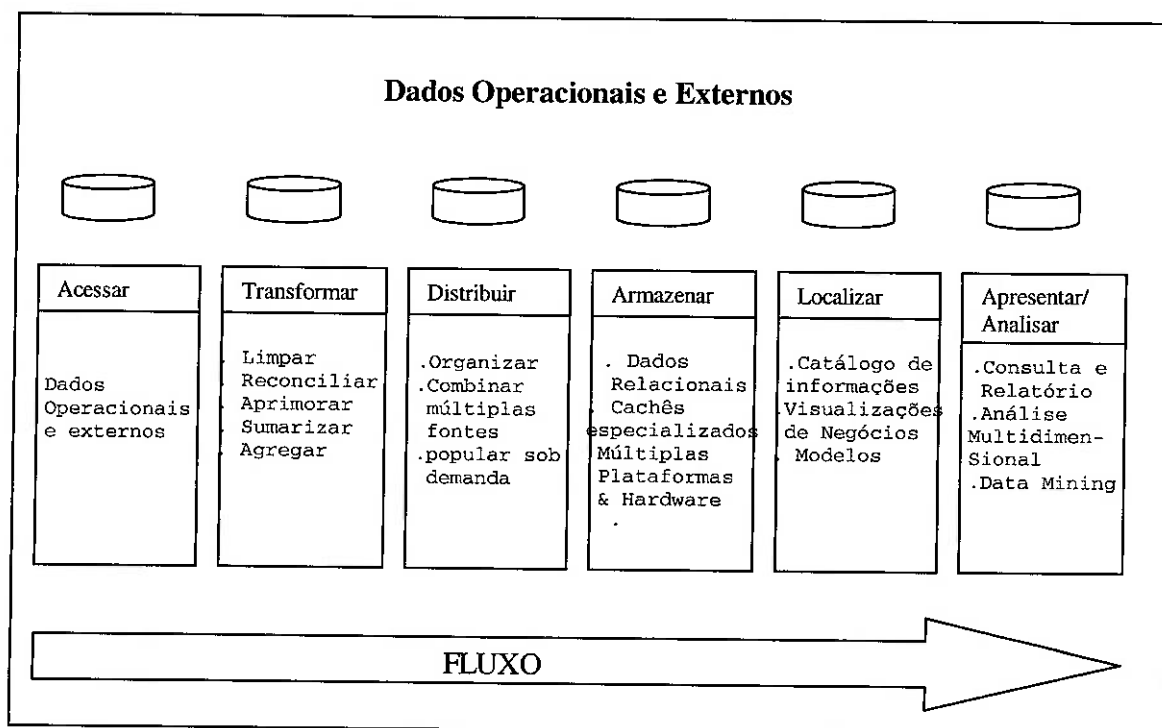


Figura 04 – Componente do Data Warehouse [WILL97]

Construir um Data Warehouse que seja eficiente e que sirva de suporte às decisões da empresa não é uma tarefa fácil, isso exige muito mais do que simplesmente descarregar ou copiar dados dos bancos de dados operacionais atuais para um banco de dados maior.

Para saber como o Data Warehouse faz para armazenar, integrar, comunicar, processar e apresentar os dados que os usuários necessitam, deve ser fazer um estudo de uma arquitetura que atenda suas necessidades.

4.2 Arquitetura centralizada X arquitetura distribuída

Os primeiros projetos sobre Data Warehouse referiam-se a uma arquitetura centralizada. Embora seja interessante, fornecendo uniformidade, controle e maior segurança, a implementação desta abordagem não é uma tarefa fácil. Requer uma metodologia rigorosa e uma completa compreensão dos negócios da empresa. Esta abordagem pode ser longa e dispendiosa e por isto sua implementação exige um planejamento bem detalhado.

Com a utilização de Data Marts, a abordagem distribuída passou a ser uma das opções de arquitetura de Data Warehouse. Os Data Marts podem ser mostrados de duas maneiras, a primeira é *top-down* e a outra é a *botton-up*.

Quando a empresa cria um Data Warehouse e depois parte para a segmentação, ou seja, divide o Data Warehouse em áreas menores gerando assim pequenos bancos orientados por assuntos, está se empregando a técnica *top-down*. Quando a situação é inversa, ou seja, a empresa cria primeiro um banco de dados para uma única uma área e a partir da visualização dos primeiros resultados parte para outra área e assim sucessivamente até resultar num Data Warehouse corporativo, está se utilizando a técnica de *botton-up*.

A tecnologia usada tanto no Data Warehouse como no Data Mart é a mesma, as variações que ocorrem são mínimas, concentrando-se no volume de dados e na complexidade de carga. A principal diferença é a de que os Data Marts são voltados somente para uma determinada área enquanto que o Data Warehouse é voltado para os assuntos da empresa toda. Sendo assim cabe a cada empresa avaliar a sua demanda e optar pela melhor solução.

Os altos custos de implementação de um Data Warehouse acabam limitando o seu uso por grandes companhias, que muitas vezes não estão dispostas a correr riscos no investimento em um empreendimento que não se tem certeza do sucesso e,

conseqüentemente, do retorno do investimento, tornando os Data Marts, nesse caso, uma boa alternativa e de custo menor.

4.3 Processo de Data Warehouse

A arquitetura do Data Warehouse é baseada num sistema de banco de dados relacional. Quando um dado entra em um Data Warehouse ele é transformado em uma estrutura integrada. O processo pode envolver filtragem e compactação de dados. O Data Warehouse guarda informações por cinco ou mais anos; portanto é um grande banco de dados que além dos dados gerais armazena também várias camadas de dados sobre dados (metadados) e ainda vários níveis de consolidações.

A seguir são citados alguns dos processos envolvidos na criação de um sistema de Data Warehouse, consistindo basicamente no processo de extrair informações de banco de dados distintos, integrar e armazenar estas informações de uma forma íntegra e possibilitar consultas rápidas e inteligentes sobre a base de dados gerada.

Para um melhor entendimento do processo, podemos dividi-lo em quatro grandes atividades, que são:

Extração dos Dados – Consiste em coletar os dados relevantes dos diversos bancos de dados de uma corporação, agrupar e transformar estes dados para uma única forma integrada. “Aproximadamente 60% do tempo do Data Warehouse é despendido nesta etapa” [MAC 00].

Armazenamento dos Dados - Inserção dos dados extraídos das fontes externas no Data Warehouse, através do processamento analítico destas informações.

Nesta etapa, os dados já limpos e filtrados são armazenados em várias estruturas de um banco de dados multidimensional. Neste tipo de banco de dados não existe preocupação com a replicação dos dados e sim com a velocidade de consulta dos dados, de forma que os dados aparecem replicados em diversos lugares e muitos índices são gerados.

Consulta dos Dados - Compreende todo o ambiente fornecido para geração de consultas “ad-hoc”, que são consultas personalizadas construídas por usuários que não são especialistas em computação.

Deve-se utilizar interfaces mais refinadas de consulta, com apoio gráfico, possibilitando operações de zoom e partição das informações.

Um consulta no Data Warehouse não representa somente uma seleção de dados, segundo alguns critérios como nos bancos de dados convencionais, ela é muito mais complexa pois, além de processar um enorme conjunto de registros ela precisa realizar um pré-processamento destes dados antes de entregá-los para a interface de visualização. Neste pré-processamento são executadas funções matemáticas de soma, multiplicação, média, percentagem, desvios e muitas outras.

Apresentação dos Dados - A principal preocupação nesta etapa é apresentar o resultado das consultas da melhor maneira possível para a análise do usuário. Para tanto, são necessários recursos mais aprimorados que a representação textual dos dados, como o uso de gráficos, planilhas para simulação e até mesmo animações. A forma como os dados são apresentados são de fundamental importância neste tipo de sistema. Uma simples operação de troca dos eixos de um gráfico ou tabela podem fazer grande diferença.

A arquitetura do Data Warehouse tem basicamente três componentes, conforme Figura 05.

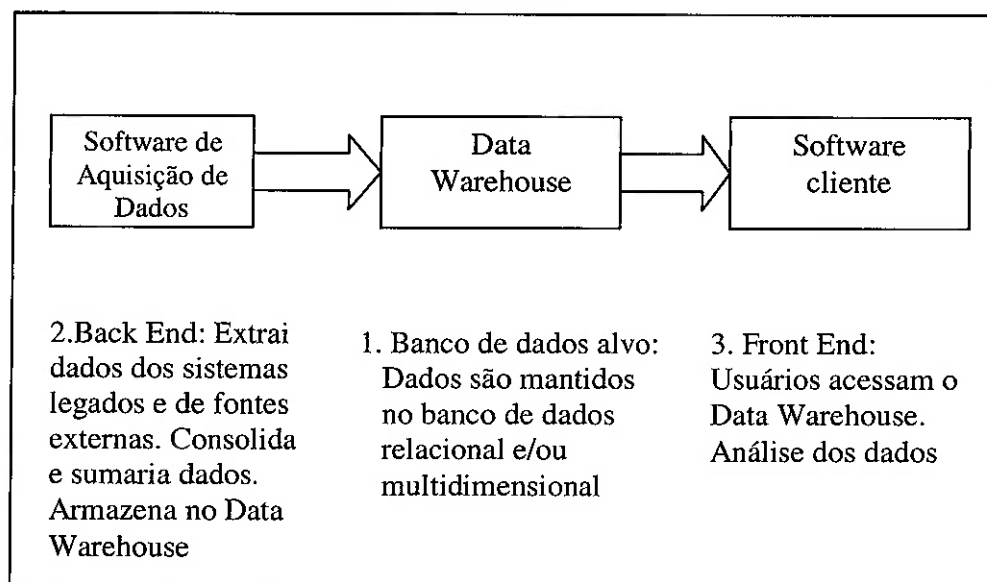


Figura 05 – Arquitetura básica

1. O Data Warehouse contém os dados e o software do banco de dados associados. Ele é freqüentemente referenciado como o “banco de dados alvo”.
2. O software de aquisição de dados (back-end) que extrai dados dos sistemas herdados e fontes externas, consolida e sumaria os dados e os armazena dentro do Data Warehouse.
3. O software do cliente (front-end) permite usuários e aplicações (como DSS – Sistemas de Suporte a Decisão e EIS – Sistemas de Informações Executivas) acessar e analisar dados no Data Warehouse.

Todos os componentes acima podem estar em uma mesma plataforma ou em plataformas diferentes, mas, independente da combinação de plataformas os três componentes são obrigatórios.

4.4 Ambiente de processamento

Existem algumas razões pelas quais o Data Warehouse seja separado do sistema de processamento on-line, que são:

Desempenho - Os picos e vales de requerimentos degradam o desempenho do sistema de transação on-line.

Acesso aos dados - Organizações freqüentemente mantêm banco de dados múltiplos, que servem funções OLTP (Processamento de transação on-line) diferentes.

O Data Warehouse, sendo uma integração de todos os dados da empresa, combina todas estas fontes de dados e adiciona fontes de dados externas.

O dado para EIS (Sistema de Informações Executivas), DSS (Sistema de Suporte a Decisão) e outras aplicações de suporte a decisão usam estas fontes múltiplas. O usuário típico do Data Warehouse não se preocupa onde os dados estão armazenados. Eles precisam e querem o acesso irrestrito aos dados que o sistema OLTP (Processamento de transação on-line) possui.

Formato dos dados - A informação está guardada num formato padrão porque os dados no Data Warehouse são integrados.

Qualidade dos dados - Os dados no Data Warehouse são validados e propriamente agregados. Pelo fato de os dados do Data Warehouse possuírem credibilidade, quando as pessoas os usam, elas gastam o seu tempo tentando entender o que os dados significam e não discutindo sobre os valores corretos dos dados.

Como qualquer produto de software, Data Warehouse é de importância para a organização somente se é usado regularmente. Como resultado o Data Warehouse cria e armazena sumários de dados acessados mais freqüentemente. O pré-processamento não somente acelera respostas para consultas que perguntam por dados sumariados, mas também para respostas para outras consultas porque reduzem o tempo ocupado do Data Warehouse.

5 Estrutura de um Data Warehouse

Um Data Warehouse contém basicamente cinco tipos de dados, que são: dados correntes (atuais) detalhados, dados mais antigos (históricos) detalhados, dados levemente sumariados, dados altamente sumariados e metadados. Esses dados não são necessariamente armazenados no mesmo servidor.

5.1 Níveis de detalhe e sumarização

Sumariar dados é um processo de entrada de dados primitivos ou detalhes de dados para armazenamento mais compacto numa forma útil para análise numa aplicação particular de registro de dados. Sumariar dados seleciona, filtra, combina, reorganiza e manipula dados detalhados ou atômicos, dados estes que servem para produzir categorias predeterminadas e específicas, totais e comparações.

Dados correntes detalhados - Esses dados refletem os acontecimentos mais recentes, podendo tornar-se muito volumosos se armazenados no mais baixo nível de granularidade. Frequentemente, eles são simplesmente uma replicação dos dados de transações correntes que foram limpos e então armazenados. Entretanto, nem todos os campos guardados nos sistemas de transação devem ser transferidos para o Data Warehouse. Note que, embora referenciados como “correntes”, estes dados são atualizados somente no tempo em que forem removidos do sistema de transação.

Muitas questões de suporte de decisões envolvem o uso de dados que podem ser obtidos diretamente dos registros detalhados nas transações. A razão dos dados detalhados serem replicados no Data Warehouse é possibilitar a execução de consultas sem deixar o desempenho do sistema de transação cair ou parar em grandes períodos de tempo, enquanto que consultas legítimas são respondidas.

Dados mais antigos detalhados - A maioria dos dados do Data Warehouse tem regras em que, quando as informações detalhadas alcançam uma certa idade, o estado é mudado do disco para algum agente de armazenagem de massa. Embora ainda recuperável na forma detalhada, o tempo de acesso é um pouco mais longo porque um agente mais lento é usado.

Dados levemente sumariados - A experiência mostra que, sumariando o dado numa forma que antecipe requerimentos para uma quantidade padrão, a sensibilidade e o uso do warehouse são melhorados. Do ponto de vista do projetista, duas decisões são requeridas: seleção dos atributos e seleção da unidade de tempo para sumariar.

As duas questões envolvem processos em que as transações não precisam ser realizadas repetidamente, mas mais espaço de armazenagem é necessário. Atributos e combinações de atributos que são freqüentemente usados para consultas deveriam ser sumariados, enquanto que aqueles que são raramente usados não deveriam. Uma vez que os atributos são selecionados, a próxima questão é como freqüentemente cada particular atributo deveria ser sumariado.

Dados altamente sumariados - Algumas informações, particularmente aquelas requeridas por muitos usuários, deveriam estar disponíveis na forma compactada e facilmente acessível. Esta informação tipicamente inclui informações que são consultadas repetidamente. Esta informação vai além de sumariar os dados de transação que são mantidos. Ela inclui a capacidade para manter sumários de dados por períodos de tempo longos assim que as tendências sejam estabelecidas. Com o armazenamento de dados altamente sumariados, tempos de resposta são também melhorados.

5.2 Metadados

Os Metadados são definidos como “dados sobre dados” e têm um papel muito importante na administração dos dados, pois é através deles que as informações são processadas, atualizadas e consultadas. Eles devem ao menos conter a estrutura dos dados, o algoritmo usado para sumarizar os dados e a desenho do ambiente operacional para o Data Warehouse.

Em um banco de dados relacional, os metadados são as representações dos objetos definidos na base de dados como por exemplo as definições de suas tabelas, colunas, base de dados, visões e alguns outros objetos. No Data Warehouse, metadados se referem a algo que define um objeto, tal como uma tabela, uma coluna, um consulta, um relatório, uma regra de negócio, ou um algoritmo de transformação.

Segundo Inmon [INM97] os metadados mantêm informações sobre “o que está onde” no Data Warehouse.

Os metadados servem como um mapa para localizar o local onde a informação está armazenada, mas, para isso acontecer o Data Warehouse deve conter um componente que cumpre as funções catalogadas para a informação que gerencia.

Esses catálogos devem ter a função de mapear o caminho em que um tipo de informação é derivado de outro, fornecer mapas para extração de dados do sistema operacional e carregá-los para o Data Warehouse. Armazenar as regras de negócio construídas, armazenar o controle de acesso e as regras de segurança ao suporte do gerenciamento de segurança, possibilitar para os metadados localizar as mudanças durante todo o tempo, organizar os metadados para serem supervisionados para capturar seu histórico de mudanças, armazenar a estrutura e o conteúdo do Data Warehouse. Identificar clara e formalmente o sistema de registro para o Data Warehouse, tornar disponível a integração e transformação lógica como uma parte regular dos metadados do warehouse e armazenar métricas.

Com essas informações o usuário final pode determinar se uma requisição será grande ou pequena antes de submetê-la.

A figura 06 [WILL97] mostra a estrutura de dados em um Data Warehouse.

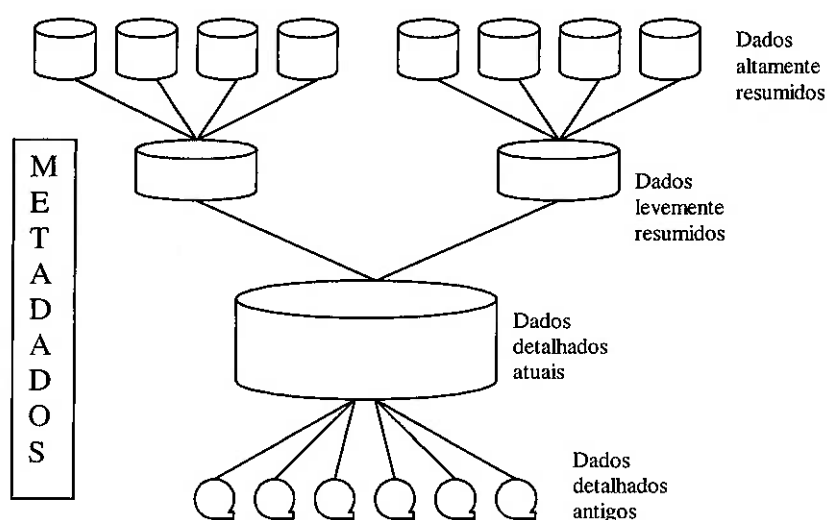


Figura 06 – Estrutura de dados dentro de um Data Warehouse

5.3 Aspectos de uma estrutura de Data Warehouse

Segundo Harry [HARR01], a estrutura de um Data Warehouse possui cinco aspectos-chave:

- *Política.* O primeiro passo do desenvolvimento de um Data Warehouse é determinar quais informações serão incluídas e qual será a frequência de atualização, backup e proteção dos dados.
- *Transformação.* Antes que os dados brutos sejam armazenados eles devem ser filtrados para que tenham significado para os usuários que irão acessá-los.
- *Armazenamento.* À medida que os dados são transferidos dos vários sistemas aplicativos para o Data Warehouse, devem ser armazenados para maximizar a flexibilidade e administração do sistema e o acesso como um todo.
- *Análise.* É o momento onde se examinam as necessidades da empresa quanto ao modelo analítico, cálculos complexos e análises de grandes volumes de dados.
- *Acesso.* Selecionar, visualizar e manipular com habilidade os dados que compõem a categoria geral referida como acesso aos dados. Planilhas, ferramentas de consulta e geradores de relatório atendem a esse requisito apresentando uma interface gráfica para navegar os dados do Data Warehouse.

6 Necessidades do Data Warehouse na CTEEP

6.1 História da Empresa

A CTEEP foi criada em abril de 1999, dispõe de 11.000 quilômetros de linhas de transmissão, com as seguintes classes de tensão: 88, 138, 230, 345 e 440 quilovolts (kV). São mais de 17.000 quilômetros de circuitos e 75 subestações, distribuídas praticamente por todo o território do Estado de São Paulo. Um sistema que soma uma capacidade de transformação de 18.000 megavolts-ampère (MVA).

A competência da CTEEP já conta com aval internacionalmente reconhecido. A gestão do Centro de Operação do Sistema São Paulo (COS-SP), localizado no município de Jundiaí, está inteiramente dentro dos padrões da norma ISO 9002. Em novembro de 1999 este Centro recebeu um novo Sistema de Supervisão e Controle – SSC, o mais moderno do País, que em seguida foi instalado nos Centros Regionais de Operação (CRO) de Bauru e Cabreúva para que, com uma atuação integrada, fossem ampliadas a confiabilidade e a qualidade da coordenação, supervisão e controle em tempo real do sistema elétrico.

6.2 Principais Sistemas de Informação

Atualmente a CTEEP possui vários sistemas de informação, sistemas estes que atendem aos principais processos do negócio da empresa. O objetivo deste trabalho é identificar a necessidade da implementação de um Data Warehouse integrando as informações do sistema de manutenção (MANTEC), do sistema de supervisão e controle (SSC) e do sistema de gestão empresarial (ERP) visando criar uma base de dados única que será fundamental para a construção de um sistema de suporte à decisão (SAD).

MANTEC

O sistema MANTEC possibilita documentar e registrar históricos dos eventos, movimentação de equipamentos, composição de custos, geração de subsídios para

estoque de peças, periodicidade de manutenção, sendo uma ferramenta necessária para controlar e otimizar a função de manutenção de equipamentos.

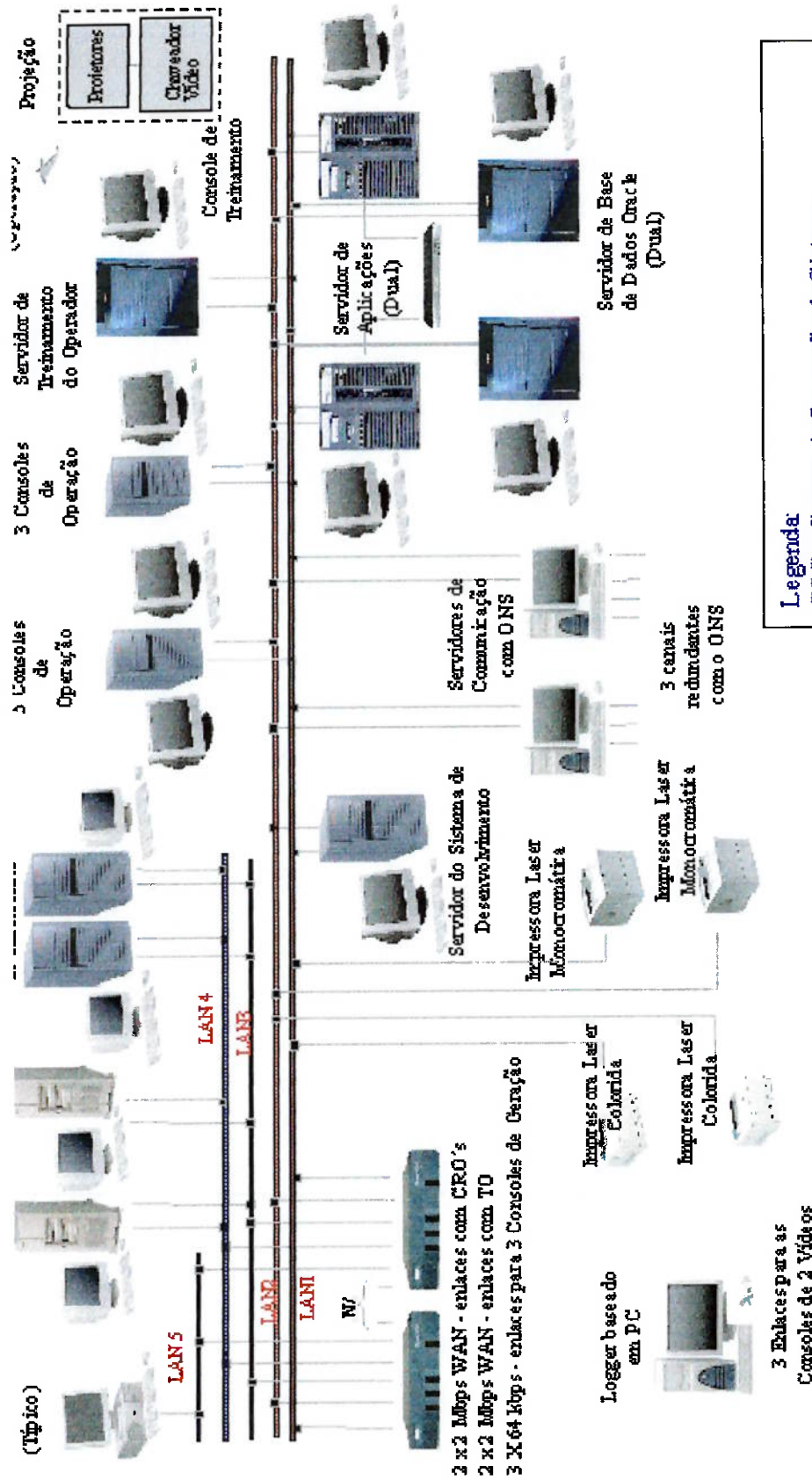
Dentre as principais funcionalidades do sistema encontra-se a emissão de relatório que permite o controle sobre a gestão do próprio sistema de manutenção, como por exemplo: quantidades de ordens de serviço emitidas por área de manutenção. Possibilita também a emissão de relatórios gráficos sobre a parada de equipamentos considerando os tempos totais da ocorrência e os tempos de atendimento das ocorrências.

SSC (Sistema de Supervisão e Controle)

O sistema SSC processa os valores das grandezas elétricas telemedidas através de unidades terminais remotas (UTRs) que ficam instaladas nas usinas e subestações do sistema elétrico. Essas medidas feitas ciclicamente pelas unidades remotas são retransmitidas por canais de microondas, diretamente para os centros regionais a que pertencem.

Cada centro retransmite essas informações para a matriz no município de Bom Jardim, no interior de São Paulo. A partir desse ponto, o SSC redireciona parte dessas informações para o Centro Nacional de Operação do Sistema, em Brasília/DF, por meio de três canais de comunicação.

A arquitetura do SSC está representada pela figura 07.



Legenda:

COS – Centro de Operação do Sistema
 CRO – Centro Regional de Operação
 ONS – Operador Nacional do Sistema
 TO – Departamento da Operação
 LAN – Rede de Área Local
 WAN – Rede de Área Estendida
 GPS – Sistema de Posicionamento Global

Centro da Operação do Sistema COS - BOM JARDIM

Figura 7 - A arquitetura do SSC

ERP – Sistema de Gestão Empresarial

ERP é definido como uma arquitetura de software que facilita o fluxo de informações entre todas as atividades da empresa como suprimentos, finanças e recursos humanos. É um sistema amplo de soluções e informações. Um banco de dados único, operando em uma plataforma comum que interage com um conjunto integrado de aplicações, consolidando todas as operações do negócio em um simples ambiente.

O ERP emprega a tecnologia cliente/servidor. Isto significa que o usuário do sistema (cliente) roda uma aplicação (rotina de um módulo do sistema) que acessa as informações de um sistema de gerenciamento de uma base de dados única (servidor). Isto, ao contrário do antigo sistema de mainframe, reflete o conceito de computação descentralizada.

O sistema de ERP implantado na CTEEP foi o Oracle Application com os módulos de Suprimentos, Financeiro e Recursos Humanos/Folha de Pagamento (figura 08).

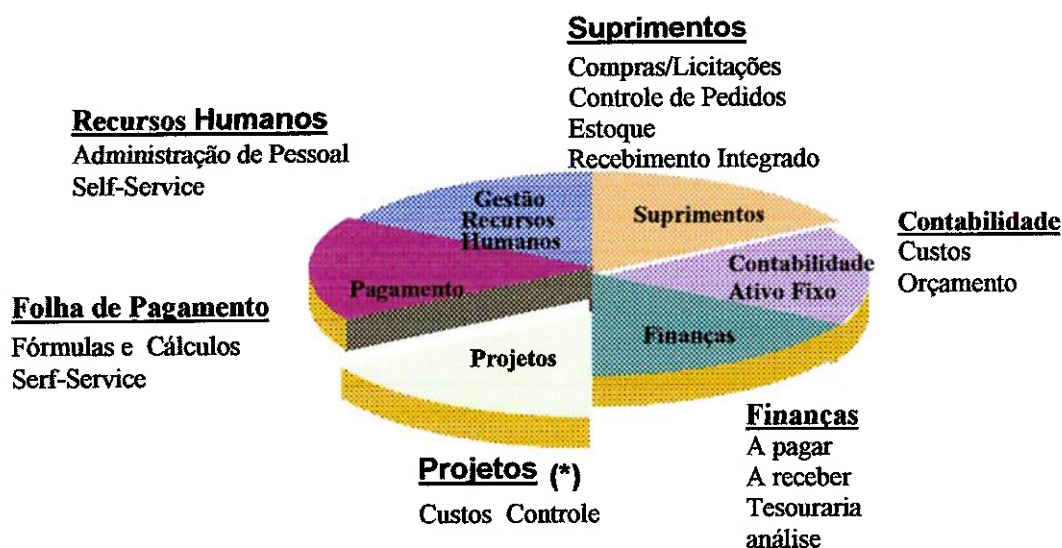


Figura 08 - sistema de ERP implantado na CTEEP

6.3 Aplicação do Data Warehouse na CTEEP

Abaixo são apresentadas algumas sugestões de análises cruzadas que podem ser obtidas com a implementação de Data Warehouse na empresa:

Custo de projetos - Hoje a estrutura da conta contábil da empresa já prevê a contabilização por projetos. O Data Warehouse entraria com a integração de todos os dados do projeto que estão espalhados nos sistemas técnicos (MANTEC) e no ERP (dados financeiros).

Simulações - Através de simulações seria possível saber, por exemplo, quanto representa para a empresa os gastos com funcionários de determinadas áreas e quanto representaria para ela o pagamento de determinado benefício/adicional, hoje somente com o módulo de Folha de Pagamento do ERP não é possível fazer esse tipo de simulação.

Qualidade - Trazer dados da ANEEL e dados estatísticos do setor elétrico, comparando com índices de qualidade. Mapear o estado de todos equipamentos do sistema elétrico, plano de manutenção, visando melhorar o nível de qualidade do sistema de transmissão, utilizando os dados do MANTEC e do SSC.

Saúde - Relacionar informações de afastamento de funcionários com período anual, identificando possíveis doenças sazonais, podendo com isso elaborar campanhas de prevenção através da análise de dados do ERP-Folha de Pagamento e ERP-Recursos Humanos.

Custo de estoque – Através de informações obtidas através do ERP-Suprimentos e ERP-Financeiro, pode-se analisar quanto custa manter os locais de estoque da empresa, será que estocar é a melhor solução?

Prevenção de falhas no sistema de transmissão - Mapear consumo das localidades (visando evitar black-out) identificando necessidade de construção de futuros canais de comunicação no sistema de transmissão através da análise dos dados do MANTEC e SSC.

A figura 09 representa a integração das informações dos sistemas ERP, SSC e MANTEC.

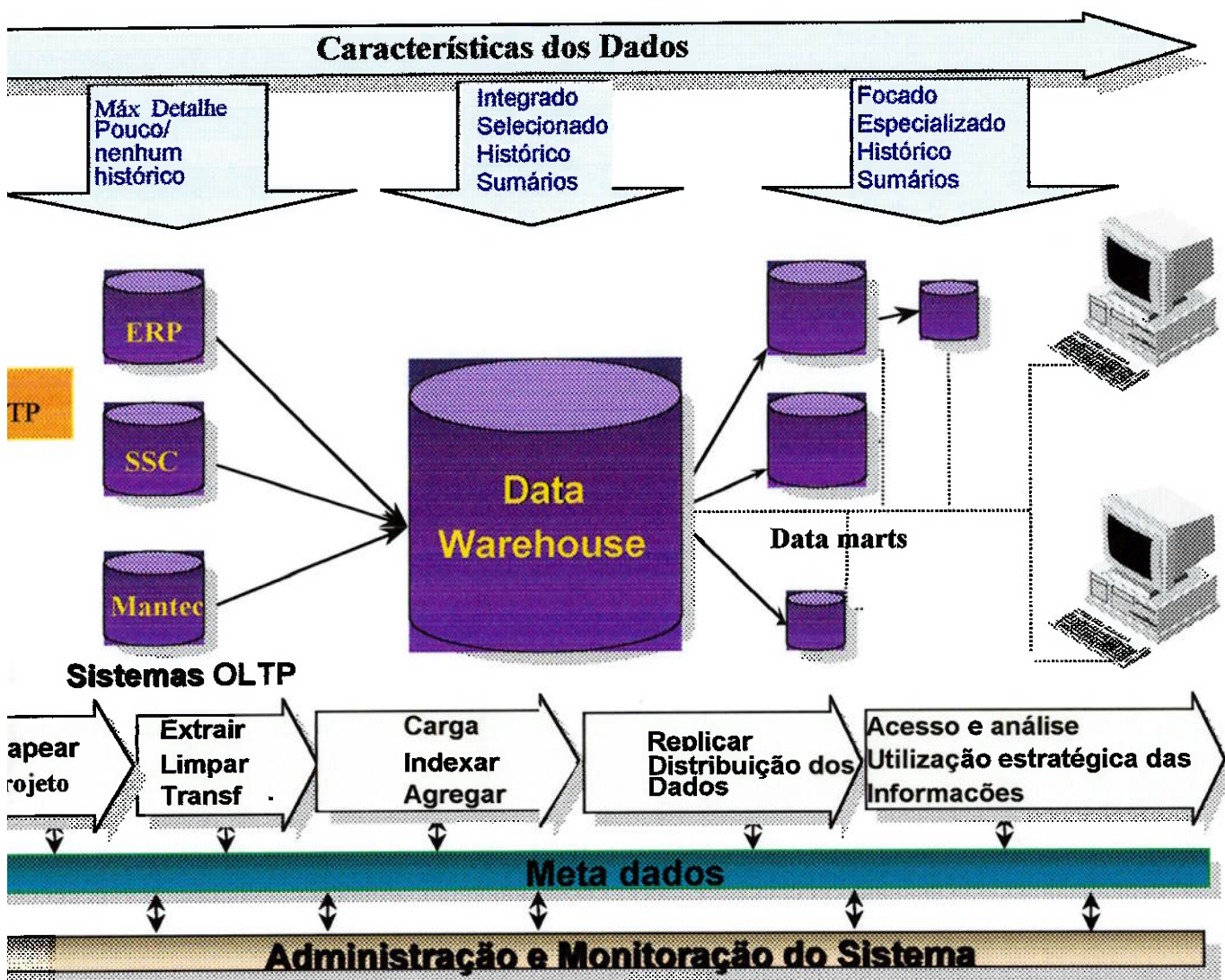


Figura 09 – Integração dos Sistemas da CTEEP

7 Conclusão

Esse trabalho propõe a análise da implementação de um ambiente de Data Warehouse na CTEEP, disponibilizando informações com qualidade e em menor tempo para a alta direção da empresa. Essa implementação visa desvincular a diretoria da dependência de seus subordinados para obter informações que possam subsidiar a tomada de decisão.

Não basta simplesmente acumular grandes volumes de dados na empresa, mas, é preciso transformá-los em informações úteis ao dia-a-dia dos negócios de forma a tornar eficaz o processo de tomada de decisões.

Com a proposta de implementação do Data Warehouse, através de uma Metodologia confiável e de ferramentas adequadas, a empresa obterá um repositório que será de importância vital para a implantação de um sistemas de suporte à decisão, convertendo dados em resultados.

BIBLIOGRAFIA

- [INM97] Inmon, William H. *Como construir o Data Warehouse*. Rio de Janeiro, Editora Campus, 1997.
- [KIM98] Kimball, Ralph. *Data Warehouse Toolkit*. São Paulo, Makron Books, 1998.
- [WILL97] Inmon, William H. *Como usar o Data Warehouse*. Rio de Janeiro, Infobook, 1997.
- [HARR01] Harry Singh, *Data Warehouse*. São Paulo, Makron Books, 2001.
- [BAR96] Barquini, Ramon, *Planning and designing the Warehouse*, New Jersey, Prentice-Hall, 1996, 311 pg.
- [POE98] Poe, Vidette; Klauer, Patricia & Brobst, Stephen. *Building a Data Warehouse for decision support*. New Jersey, Prentice Hall PTR, 1998.
- [DAL99] Dal'Alba, Adriano. *Um estudo sobre Data Warehouse*. Via Internet. <http://www.geocities.com/siliconvalley/port/5072/>.
- [CAM99] Campos, Maria Luiza & FILHO, Arnaldo V. Rocha. *Data Warehouse*. via Internet. <http://genesis.nce.ufrj.br/dataware/tutorial/indice.html>.