

XIV SEMINÁRIO NACIONAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA

PRIORIZAÇÃO DE OBRAS USANDO SOLUÇÃO DE DATA WAREHOUSE

FERNANDO ANTONIO GRUPPELLI JUNIOR
EDIVAR GARCIA AZONI

COPEL - COMPANHIA PARANANENSE DE ENERGIA

Palavras-chave: data warehouse, armazém de dados, priorização de obras,
planejamento da distribuição

Foz do Iguaçu, 19 a 23 de novembro de 2000

PRIORIZAÇÃO DE OBRAS USANDO SOLUÇÃO DE DATA WAREHOUSE

SUMÁRIO

	página
1. Introdução	2
2. O Problema	2
3. Justificativa	3
3.1. Histórico	3
3.2. O que é um Armazém de Dados ou Data Warehouse	3
3.3. Proposta	4
4. Metodologia	5
5. Desenvolvimento	6
6. Conclusões	9
REFERÊNCIAS BIBLIOGRÁFICAS	10

1)INTRODUÇÃO

Os investimentos em rede de distribuição primária podem ser considerados como uma parcela significativa do orçamento anual da empresa. Desta forma , a preocupação em investir corretamente é fundamental. O planejamento local ocupa-se de propor obras adequadas a manter a qualidade do serviço e atender ao mercado consumidor. Estas obras, após sua proposição, são priorizadas com o objetivo de minimização do custo global, o que significa basicamente prestar um serviço adequado a um custo mínimo. Nos métodos atuais calcula-se a relação custo/benefício de cada obra sendo que o custo é obtido facilmente através do levantamento de custos médios unitários e os benefícios provêm das variações no custo global que a intervenção no sistema provoca. Entretanto, a avaliação destes benefícios envolve normalmente grande quantidade de dados históricos do sistema, especialmente quando se refere a interrupções do sistema. Estes dados normalmente estão em bases diferentes e fazem parte de diversos sistemas transacionais da empresa.

A tecnologia da informação tem se mostrado, nos últimos anos, particularmente eficaz na forma de gerenciar quantidades cada vez maiores de dados e a transformá-los em informação útil para a empresa.

Um dos conceitos mais difundidos atualmente em grandes grupos tem sido o Data Warehouse ou Armazém de Dados. A utilização de Data Warehouse permite reunir dados de diversas fontes diferentes integrando os de forma adequada a sua utilização , o que não era possível em sistemas convencionais .

Este trabalho propõe a utilização de tecnologia de Data warehouse para auxiliar a reunião de informações adequadas e confiáveis e permitir os cálculos das relações custo benefício de obras na rede primária de distribuição. Pretende sugerir uma maior integração e aproveitamento das possibilidades de um data warehouse e sua futura integração com sistemas de geoprocessamento.

2)PROBLEMA

O problema mais comum na determinação da relação custo/benefício de obras da distribuição era a estrutura extremamente complexa de dados históricos a serem consultados. O desafio era tornar grande quantidade de dados em informação útil ao negócio da empresa. Não adianta ter montanhas de dados e não conseguir acessá-los, ou simplesmente mostrá-los adequadamente.

Percebemos também que a cada ano aumentava exponencialmente a quantidade de informações necessárias aos cálculos: causas e componentes de interrupções, novos conjuntos com metas e penalidades aplicáveis, dados por consumidor, dados por chave e alimentador, etc. Cada mudança exigia a solicitação para que um analista modificasse relatórios e bancos de dados. Existia uma necessidade de informações disponíveis em tempo real e não em relatórios via mainframe e que só seriam recebidos após sua solicitação. Além do mais existiam ilhas de informação, ou seja, informações que estavam em sistemas isolados e não eram disseminadas. Alguns sistemas também tinham dificuldade em guardar históricos.

Isto tornava o planejador um especialista em garimpar informações dentro de diversos sistemas da empresa, sendo que a maior parte estava disponível apenas via mainframe. Alguns tipos de dados geravam polêmica e dúvidas quanto a sua utilização levando à necessidade de padronização através de manuais detalhados. Pilhas de papéis com números de chaves, dados de interrupções de alimentadores, históricos, etc, tinham que ser utilizados e muitas vezes acabavam existindo erros na reunião das informações .

O resultado era a imprecisão e dúvida na tomada de decisão.

3)JUSTIFICATIVA

3.1)Histórico

Foi durante as décadas de 1960 e 1970 que as empresas passaram a investir em uma revolucionária tecnologia de armazenamento informatizado das suas informações: os Bancos de Dados. Talvez o maior benefício dos Bancos de Dados tenha sido a possibilidade de centralizar as informações de forma a torná-las íntegras e não redundantes na empresa. Já nessa época os sistemas de informação que faziam uso de Bancos de Dados possuíam também um conjunto de relatórios ditos *gerenciais* ou *estatísticos*. Tais relatórios tinham a função de consolidar as transações registradas no Banco de Dados de forma a responder perguntas como: *Qual foi o volume de vendas das Filiais da Região Sul no mês passado ?*

Na década de 1980 duas novas tendências se consolidaram no ambiente empresarial. A primeira delas dizia respeito à necessidade de descentralização das decisões corporativas. E junto a isso uma nova tecnologia colocava em xeque o conceito dos Bancos de Dados centralizados em grandes computadores: era a arquitetura *cliente/servidor* que colocava um computador pessoal em cada mesa de trabalho. Obviamente a indústria de informática acompanhou esses movimentos, tornando comum a proliferação de pequenos Bancos de Dados departamentais empresa afora. Havia a mala direta do Departamento de Marketing; o Sistema de Pequenos Gastos da área Financeira, etc.

A década de 1990 nos mostrou empresas com um imenso volume de dados em mãos e uma pergunta a ser respondida: *E agora, o que fazemos com essas informações ?* Pareceu até um movimento natural que dessa grande quantidade de dados (acumulada nos 30 anos de história dos Bancos de Dados) surgisse a resposta: *Vamos consolidá-las em um outro Banco de Dados, o qual chamaremos de Armazém de Dados*. Inicia então a era do *conteúdo* e do *conhecimento*. Não basta apenas ter um grande volume de dados armazenado. É imperativo que possamos obter informações rápidas e precisas que nos auxiliem a gerir a empresa de forma estratégica.

3.2) O que é um Armazém de Dados ou Data Warehouse .

Segundo W. H. Inmon, um dos principais autores nesta área, um Armazém de Dados é definido como: *Um conjunto de dados organizados por assuntos, integrados, não volátil e variável em relação ao tempo, de apoio às decisões gerenciais.*

Vejamos a seguir alguns aspectos sobre essa definição:

Organizado por assuntos: Significa que os dados são agrupados de acordo com suas características comuns. Há no Armazém de Dados informações de Recursos Humanos, Marketing, Produção, Financeiros, etc.

Integrados: Apesar de serem organizados por assunto, os dados são inter-relacionados. Tal integração é obrigatória e através dela é possível, por exemplo, cruzarmos informações de RH com volume de vendas dos vendedores, por exemplo.

Não volátil: Uma vez que um dado entra no Armazém de Dados, ele (salvo raras exceções) não mais é removido. Dessa forma, informações históricas são mantidas à disposição para consulta durante anos.

Variável em relação ao tempo: O registro do que aconteceu no período (mês, ano, década) passado é importante para a execução de comparações com o que ocorre hoje. Praticamente todas as informações no Armazém de Dados são analisadas em função do tempo.

3.2.1)Armazém de Dados e Intranet

Cronologicamente falando, a grande difusão das tecnologias associadas à Internet teve sua expansão na mesma época do surgimento dos Armazéns de Dados. A partir dessa coincidência histórica, a indústria de informática atentou para o seguinte: *A Internet (e sua variante empresarial, a Intranet) é um excelente local para a divulgação das informações contidas no Armazém de Dados.*

Essa afirmação está solidamente assentada sobre as características técnicas que compreendem essas duas tecnologias. A Intranet caracteriza-se por ser um meio rápido de difusão de informações por toda a empresa. A interface tradicional em uma Intranet (o *browser*) mostra-se um excelente meio para se efetuar consultas em um Armazém de Dados (são raros os casos em que o usuário final faz cadastramento manual de informações no Armazém de Dados). E finalmente, há de salientar que os custos de implantação e de treinamento no uso de um Armazém de Dados numa Intranet caem vertiginosamente.

3.2.2) Como as informações são estruturadas no BD

Em um Armazém de Dados, há basicamente dois tipos de dados armazenados. O primeiro deles é chamado de *Dimensão*. Uma Dimensão é o meio pelo qual os dados são analisados. Por exemplo, uma Dimensão típica em um Armazém de Dados é o Tempo, pois os dados são avaliados por Ano, Trimestre, Mês, etc. Outras dimensões seriam: Regiões Geográficas, Produtos, Equipes de Atendimento, etc.

O segundo tipo de dado em um Armazém de Dados é chamado de *Fato*. O Fato nada mais é que o dado propriamente dito que será avaliado. Por exemplo: Total de Vendas, Valor Médio do Produto, Tempo Médio de Reposição, Tempo Médio de Atendimento, etc.

É do cruzamento destes dois tipos de dados que o Armazém de Dados produz respostas rápidas a perguntas do tipo:

Qual foi o volume de vendas da Filial Sul no último trimestre em relação às outras Filiais ?

Qual a performance dos meus vendedores nesse mês em relação ao mesmo período do ano passado ?

3.3) Proposta

Tornava-se impossível fazer investimentos em sistemas novos que saberíamos que ficariam obsoletos e com os mesmos problemas em pouco tempo. Entretanto, a tecnologia citada anteriormente tornou possível capturar dados em grandes quantidades, armazená-los de forma cada vez mais detalhada e torná-los informação precisa.

Optou-se então pela utilização de um Data Warehouse ou Armazém de Dados, tecnologia já existente e disponível no mercado e que difere dos padrões atualmente utilizados pois é criado um banco de dados paralelo aos sistemas operacionais da empresa, onde novos métodos de armazenamento, geração, análise e recuperação de informações são utilizados.

Esperava-se do sistema :

- acesso a dados corporativos.
- dados consistentes, isto é, todos recebem as mesmas informações.
- dados separados ou combinados conforme a necessidade.
- conjunto de ferramentas para análise e apresentação das informações.
- acesso rápido de qualquer local da empresa, com a utilização de intranet.

4)METODOLOGIA

Tudo começa com a definição de quais dados serão armazenados no Data Warehouse. Esta etapa é particularmente importante pois define os níveis básicos do processo a ser analisado. São na verdade as questões que o sistema deverá responder. Depois é necessário fazer a modelagem dimensional, determinando os fatos e as dimensões. Para cada dimensão escolhida devem ser definidos seus atributos. O tempo é sempre uma dimensão a considerar em dados históricos como estes que estamos trabalhando. Os fatos são itens a medir ou a analisar. A distinção entre fatos e dimensões nem sempre é fácil.

O próximo passo é a identificação dos sistemas em que os dados estão armazenados para que a carga seja feita no Data Warehouse. Os dados são então extraídos e após isto checados. Esta também é uma fase importante do processo, pois não se pode liberar os dados sem ter certeza de que estão corretos.

Passamos então a utilização e análise, que é feita através de ferramenta OLAP(On-Line Analytic Processing), e que permite o acesso e visualização da informação da forma desejada e via Intranet. Até aqui temos praticamente todas as informações necessárias e reunidas para proceder o cálculo da relação custo/benefício de uma determinada obra, mas o Sistema de Suporte a Decisão ainda não está terminado pois é necessário um aplicativo desktop que através de uma interface transfira os dados para onde os cálculos são realizados.

O último passo é o aplicativo que vai realizar o cálculo das variações do custo antes e depois da intervenção no sistema. Neste caso optamos por um aplicativo que roda no próprio microcomputador e que a partir de informações próprias e externas faz o cálculo da relação custo benefício de cada obra baseado no Método Payoff [1] .

5)DESENVOLVIMENTO

Na figura abaixo podemos visualizar todas as etapas do processo já implementadas.

FIGURA 1 - SISTEMA IMPLANTADO

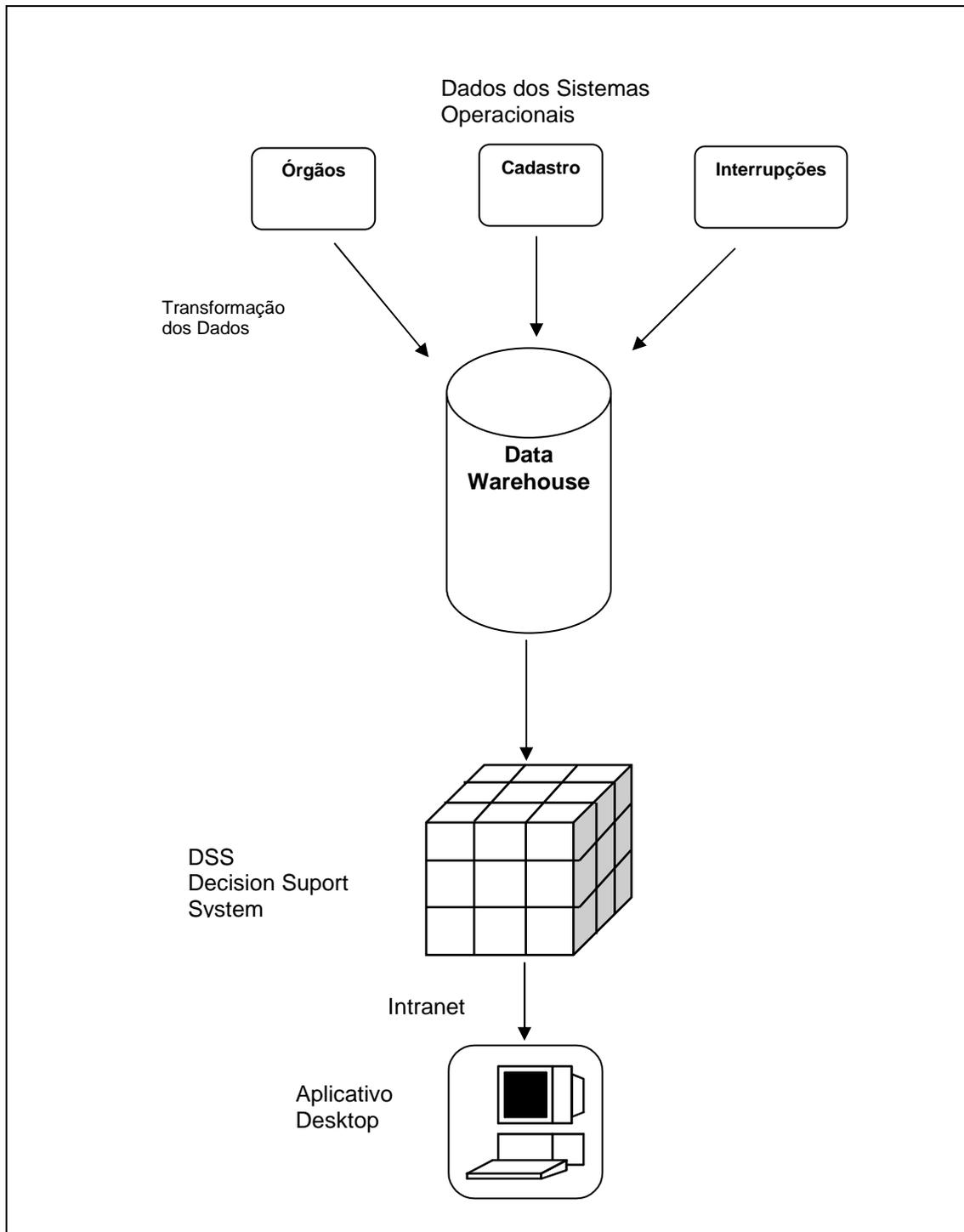


FIGURA 1

A partir da definição dos sistemas operacionais mais utilizados no cálculo da relação custo/benefício, foi realizada a carga dos dados referentes ao período de um ano e posteriormente, após a verificação e correção destes dados, o sistema foi carregado com outros períodos.

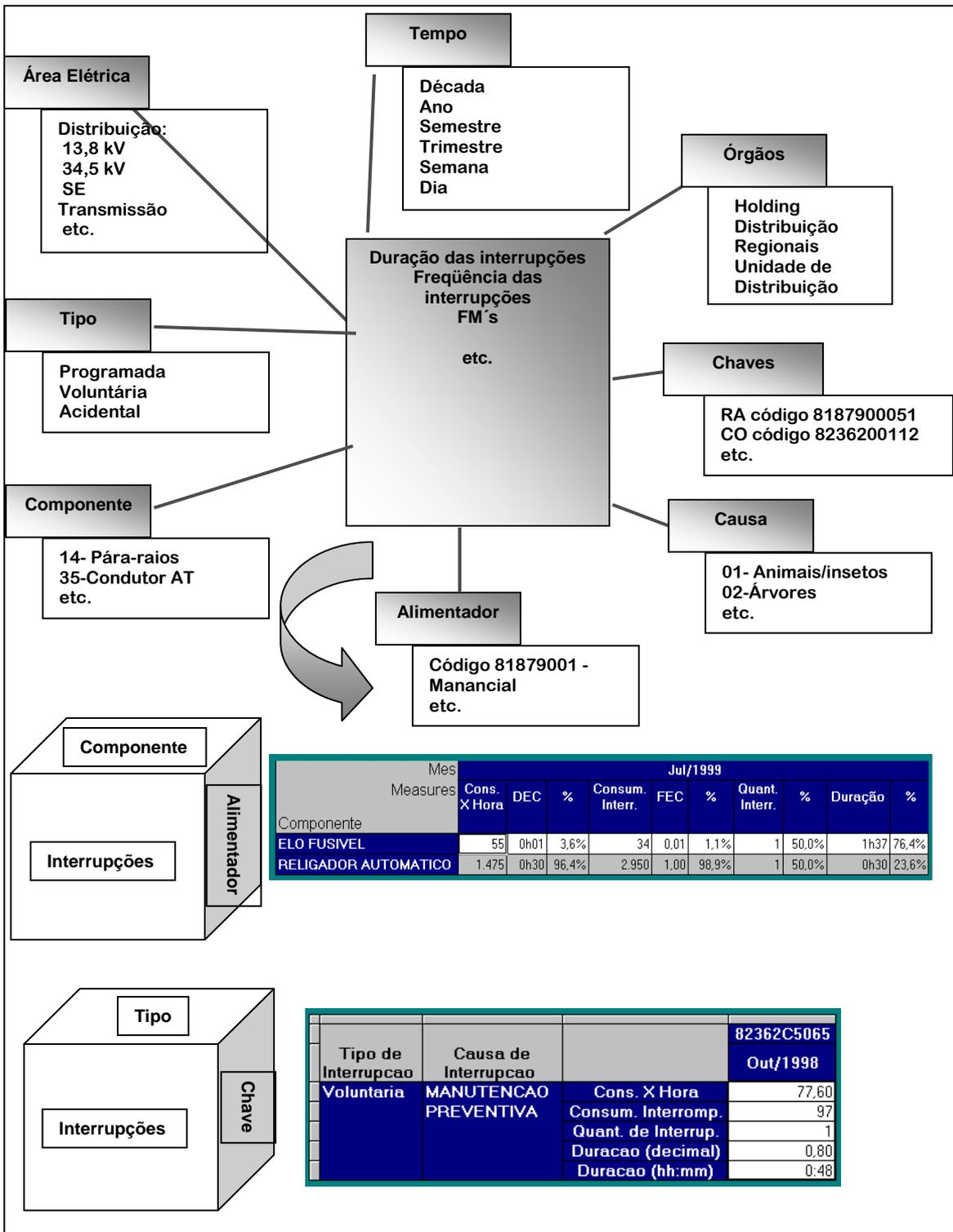
A tabela fato e algumas dimensões do modelo podem ser vistas na figura 2. As durações e as frequências das interrupções são colocadas na tabela fato pois são os dados mais importantes a serem avaliados.

Através de algumas ferramentas, os principais relatórios e gráficos são elaborados de acordo com a necessidade do planejador, podendo fornecer qualquer tipo de relação entre eles conforme mostrado na figura 2, onde podemos ver dois exemplos de relatórios. No primeiro, é possível identificar as componentes de todas as interrupções de um determinado alimentador e no segundo todos os tipos e causas de interrupção de uma determinada chave.

Os dados podem ser consultados via Intranet, ferramenta que já está disponível em qualquer local da empresa. Dessa forma, o planejador consegue consultar apenas as informações que lhe interessam no momento, podendo ser de um alimentador ou de uma chave ou até mesmo de um órgão. Estes dados específicos são transferidos para um aplicativo no seu microcomputador e através da escolha de alguns parâmetros pode ser rapidamente calculada a relação custo benefício.

Por exemplo, um caso em que o planejador verifica a aplicação de rede compacta em determinado alimentador, requeria pelo sistema anterior a consulta ao histórico de interrupções do Religador Automático da subestação, verificando quais as interrupções que poderiam ter sido evitadas. Isto envolvia a consulta de dezenas de interrupções em determinado período de tempo (normalmente um ano), verificando combinações de causas e componentes que poderiam ser utilizadas. Agora, esta mesma informação pode ser rapidamente conseguida através de consulta para qualquer chave, RA, alimentador ou órgão da empresa. Da mesma forma para outros tipos de obras qualquer combinação de dados, que antes requeria a pesquisa em diversos sistemas, está disponível para consulta. É importante ressaltar que os dados brutos já vem tratados e otimizados, não exigindo do usuário tempo extra para realizar esta tarefa.

FIGURA 2- MODELO DIMENSIONAL DO DATA WAREHOUSE



6) CONCLUSÕES / RECOMENDAÇÕES

O sistema foi implantado no final de 1999 e os primeiros resultados obtidos indicam que o acesso às informações melhorou, reduziu-se o excesso de relatórios pedidos via mainframe e obteve-se ganho de produtividade dos planejadores. Dessa forma, o planejador que antes passava 90% de seu tempo na preparação dos dados e apenas 10% na análise da viabilidade das obras, pode agora inverter isto e dedicar-se a uma melhor elaboração do programa de obras. Na verdade, o sistema poderia ser usado também para indicar locais onde o atendimento está mais crítico e até mesmo quais são as causas. Esta é uma etapa a ser mais explorada na utilização deste tipo de solução que passa de um data warehouse passivo para ativo, onde o sistema toma decisões ou vem a sugerir-las.

Percebemos também, a importância da confiabilidade nos sistemas transacionais existentes, pois a implantação de um Data Warehouse não pode ser encarada como um produto, mas como um processo onde os dados já existentes são transformados. Evidentemente que se estes dados estiverem errados, os resultados finais também estarão comprometidos. Por isso a necessidade de sistemas de qualidade para controle de informações.

Uma das grandes vantagens após a implantação é a divisão da informação até seu nível atômico, o que não acontecia em relatórios que já vinham prontos. Dessa forma o usuário passa a escolher exatamente o que deseja e acaba descobrindo que tradicionalmente usava relatórios que forneciam informações inadequadas a sua finalidade.

À medida que os usuários do sistema acostumam-se com o real valor dos dados passam a exigir informações mais refinadas, o que basicamente significa maior quantidade de consultas, resultados intermediários e maior processamento analítico. Isto acaba tomando mais tempo no processamento das informações e tornando o sistema mais lento, portanto no momento de dimensionar o sistema é necessário estar preparado para isto.

Mesmo com todas estes melhoramentos, percebemos que existe mais espaço para aprimoramento nos sistemas de priorização de obras e tomada de decisão, principalmente na sua integração aos sistemas de geoprocessamento aplicados ao planejamento da distribuição. Os primeiros testes indicam que este tipo de integração é adequado e irá trazer mais produtividade sistemas existentes.

O objetivo final é fazer com que a tomada de decisão seja feita da forma mais adequada, no menor tempo possível e com dados confiáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1) COPEL. Priorização de obras de Distribuição - Método do Payoff. Curitiba, maio 1995, rev. março 2000.
- 2) KIMBALL, Ralph., "Data Warehouse Toolkit", São Paulo: Makron Books, 1998
- 3) HARRISON, Thomas H., "Intranet Data Warehouse", Editora Berekeley 1998
- 4) MICROSTRATEGY INCORPORATED, "The Case for Relational OLAP", White Paper.
- 5) MICROSTRATEGY INCORPORATED, "Relational OLAP: An Enterprise-Wide Data Delivery Architecture", White Paper