

# Resumo

Este documento trata do conceito, técnicas e tecnologias relacionadas ao Business Intelligence. Após uma introdução aos conceitos e há um pouco da história do Business Intelligence no primeiro capítulo, são explorados as técnicas e tecnologias do capítulo 2 ao 4.

No capítulo 5, chegamos à tendência atual do Business Intelligence que é o chamado Self-Service BI.

No sexto e último capítulo é apresentada uma conclusão, levando em consideração as perspectivas tecnológicas e gerenciais.

# Sumário

1- Introdução .....	3
1.1 - Um pouco da história e conceitos de Business Intelligence .....	3
2 - Data Mining .....	7
2.1 - Definição.....	7
2.2 - Etapas do processo.....	7
2.2.1 - Setup das bases de dados.....	8
2.2.2 - Mineração dos dados .....	8
2.2.3 - Validação dos resultados .....	8
2.3 - Data Minings e sua aplicabilidade em BI.....	9
3 - Data Warehouse .....	10
3.1 - O que é?.....	10
3.3 - Processos .....	12
3.4 - Vantagens e Desvantagens .....	13
4 – OLAP.....	15
4.1 – Definição .....	15
4.2 – Conceito.....	15
4.2.1 – Cubo .....	16
4.2.2 – Dimensão .....	16
4.2.3 – Medidas .....	17
4.3 - Por que é tão rápido? .....	17
4.4 - Arquitetura.....	17
4.4.1 – MOLAP .....	18
4.4.2 - ROLAP .....	19
4.4.3 – HOLAP .....	19
5 - Self-Service BI .....	20
5.1 – Motivação .....	20
5.2 - Os prós e contras .....	21
6 – Conclusões.....	23
7- Referências.....	24

# 1- Introdução

## 1.1 - Um pouco da história e conceitos de Business Intelligence

Uma definição formal e simples de Business Intelligence (BI) seria a seguinte:

*BI é o conjunto de técnicas que utilizam Tecnologia da Informação para auxiliar na tomada de decisão. Este conjunto de técnicas tem como objetivo a análise e mineração de dados, assim como gestão de desempenho e benchmarking.*

Na prática, para entender Business Intelligence e suas aplicações, precisamos entender a tomada de decisões. Numa empresa, decisões de negócio precisam ser tomadas a todo instante para assegurar a presença da mesma no mercado. Tais decisões são tomadas mediante a um contexto de negócio, que nada mais é do que o contexto interno e externo (mercado) da empresa.

É importante salientar que decisões de negócio possuem um forte fator humano envolvido e isso torna o processo de tomada de decisões menos determinístico. Com isso podemos concluir que pessoas tomam decisões, e não sistemas.

Sun Tzu, em seu livro A Arte da Guerra, diz que para vencer a pessoa deve deter todo o conhecimento de suas fraquezas e virtudes e, além de tudo, as fraquezas e virtudes de seu inimigo. A falta deste conhecimento pode resultar na derrota. Trazendo este conceito para o mundo dos negócios, uma empresa precisa ter conhecimento de suas capacidades, suas fraquezas, as tendências de mercado e, além disso, o conhecimento sobre as empresas competidoras.

No processo de aquisição deste conhecimento, surgem questões pertinentes ao contexto de mercado, úteis para as tomadas de decisão. Quais são as necessidades de nossos clientes? Qual é a demanda deles? O que fazemos certo? Em que pontos deixamos os clientes desejando? O que outras empresas estão fazendo para solucionar questões recorrentes? Estas são algumas das perguntas que precisam ser respondidas para importantes tomadas de decisão, como balanceamento de produção, evolução de produtos e mudança de estratégia de marketing, entre outras.

Como mencionado anteriormente, pessoas tomam decisões. Mas baseado no que se faz uma boa decisão? Informação.

Por toda a história, a informação foi usada pelo homem para propagar o conhecimento, aprender e manter registro de fatos. Em suma, a informação é usada pra encontrar respostas para questões que levam o homem a uma melhor compreensão do mundo a sua volta.

Então, informação é sobre respostas e respostas permitem que pessoas tomem decisões. E como empresários conseguem informação? De dados.

Estes provêem informações capazes de responder à perguntas como as anteriormente citadas, têm muito valor e são preciosos, logo precisam ser guardados.

Antes da década de 60, dados eram guardados em gabinetes de arquivos. Esses passaram a ser guardados em formatos de mídia como disquetes, com o advento dos computadores.

Ambas as abordagens supracitadas (gabinete e mídia) eram arriscadas e difíceis de gerenciar. Em 1969, Edgard Codd inventou o banco de dados, fato que revolucionou a maneira como dados eram armazenados. Porém, para popular um banco de dados era necessária muita experiência e especialização.

O mundo dos negócios precisava de um modo mais fácil de popular um banco de dados. Devido a essa grande demanda, meios foram criados para ficar entre o usuário e o banco de dados, provendo o poder de inserção de dados. Estes meios surgiram na década de 1970 e eram chamados de Aplicações de Negócio.

Então bancos de dados vieram como uma solução para o problema de armazenamento de dados e as aplicações de negócio vieram como uma maneira mais fácil de inserir dados em um banco de dados. Mas e como fica o acesso? Uma primeira abordagem para tentar facilitar o acesso aos dados foi a da geração de relatórios. A primeira tentativa acabou por não dar muito certo. Relatórios eram unidimensionais e havia um outro problema emergente: a fragmentação de dados.

Extrair dados de um banco ainda era uma tarefa difícil por que esses vinham de múltiplas fontes. Uma empresa geralmente possui várias Aplicações de Negócio, gerindo dados de diferentes contextos. Integrar tais fontes de dados não era um processo simples. Eis que no início da década de 1980 surgiram os Data Warehouses, que traziam uma solução para a fragmentação dos dados.

Os Data Warehouses tornaram o acesso e a gerência dos dados mais fácil. Com dados acessíveis e bem gerenciados eles podiam ser bem servidos Com isso inicio-se o consumo dos mesmos em grande escala.

Em 1981 o termo Business Intelligence foi inventado por Howard Dresner, como um termo descrito como "*conceitos e métodos para melhorar a tomada de decisão de negócio utilizando sistemas de suporte baseados em fatos*".

Na época, ferramentas de BI podiam gerar relatórios e analisar dados. Com o passar do tempo novas ferramentas foram surgindo e a demanda de dados crescendo. Com o crescimento da demanda, começou a necessidade de adquirir dados mais rapidamente e, neste ponto, Business Intelligence tornou-se também uma questão de desempenho.

Com o advento da internet, o número de fontes de informação passou a crescer rapidamente. Com novas fontes, empresas queriam mais dados e os queriam na mesma velocidade. Os anos de 1990 trouxeram mais ferramentas de BI que proviam mecanismos para consulta, relatórios, análise e apresentação de dados. Contudo, elas ainda não funcionavam bem, pois a maior parte da informação acabava em tabelas Excel, que não traziam uma boa maneira de apresentar os dados.

Eventualmente, as ferramentas de BI trouxeram um novo desafio: *a consistência*. As ferramentas facilitaram o acesso e este abriu caminho para múltiplas versões de dados, introduzindo mais um desafio para a gerência dos dados.

As ferramentas existentes eram muito caras e difíceis de manter. Isto levou os desenvolvedores de ferramentas de BI a buscar maneiras de adicionar funcionalidades à custos reduzidos. Foi neste ponto que as plataformas de BI começaram a surgir e com isso, vieram às aquisições das empresas desenvolvedoras, que acabaram nas mãos de grandes empresas da área de TI como Oracle, SAP, IBM e Microsoft.

Agora, existiam mais ferramentas e mais funcionalidades. Business Intelligence estava consolidado no mercado e tornou-se uma prioridade para o mundo dos negócios.

Ainda assim, havia algo errado. As pessoas de negócio ainda não conseguiam obter respostas para suas decisões. O problema não eram as pessoas, mas sim as pessoas acessando os sistemas de BI.

O problema era que as ferramentas não eram intuitivas, logo o acesso a elas era difícil. Quando pessoas de negócio têm problemas com sistemas elas chamam

quem os mantém: o departamento de TI, desviando-os de suas tarefas de desenvolvimento para prestar suporte ao usuário.

Business Intelligence, neste ponto, tornou-se uma questão de usabilidade. Era sobre transformar os dados em formatos que pudessem ser facilmente consumidos por pessoas de negócio, justamente por que são eles quem precisam dos dados. BI precisava ser focado nas pessoas, porque *pessoas tomam decisões*.

O novo desafio do Business Intelligence agora é tornar o acesso aos dados mais amigável.

## 2 - Data Mining

### 2.1 - Definição

Tomando-se por base a conjuntura do mundo contemporâneo, onde tem-se um enorme volume de dados em crescimento exponencial, faz-se claramente indispensável a utilização de algum recurso que possibilita aos interessados obter o máximo proveito de toda esta informação. Surge daí um dos principais “pilares” do processo de Business Intelligence: O Data mining.

De acordo com Usama Fayyad, Ph.D. em Ciência da Computação pela Universidade de Michigan, Data Mining (ou Mineração de dados) assim pode ser definida:

*“Data Mining é o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e, por fim, compreensíveis.”*

*Fayyad (1996)*

Em outras palavras, trata-se de um processo analítico de dados, com vasta utilização de técnicas de estatística, inteligência artificial, recuperação da informação e reconhecimento de padrões, que objetiva delinear (evidenciando ou extraíndo tais padrões) estruturas de conhecimento, de modo a conduzir, do melhor modo possível, as decisões de negócio.

### 2.2 - Etapas do processo

Pode-se dizer que o processo de Data Mining consiste de três etapas: setup das bases de dados, mineração dos dados e validação dos resultados.

Segue abaixo, uma melhor explicação de cada uma destas etapas. Este artigo não irá aos pontos mais fundos destes conceitos por se tratar de um artigo sobre Business Intelligence, e não estruturas de banco de dados e afins.

### 2.2.1 - Setup das bases de dados

Primeiramente, é necessário, no processo de Data Mining, a seleção de uma amostra de dados, que venham a ser limpas e “pluralizadas”. Entenda-se por limpeza a manutenção da consistência dos dados, preenchimento de informações, remoção de redundâncias, entre outros; e por “pluralização” o ato de valorizar os dados genéricos em detrimento dos dados mais específicos (uma vez que dados específicos têm um padrão muito pontual, que não pode ser compartilhado).

Daí, tem-se o surgimento dos Data Warehouses. Pode, ainda, ocorrer a criação de um conjunto de treino (da rede neural do Data Mining) e um conjunto de testes, para verificação da precisão dos algoritmos.

### 2.2.2 - Mineração dos dados

Nesta fase, utilizam-se algoritmos de clusterização, classificação, regressão e associação, de modo a efetuar o agrupamento de dados de certa forma similares, verificando o padrão que seguem, e o relacionamento entre suas variáveis.

### 2.2.3 - Validação dos resultados

Nesta etapa, observa-se se os padrões encontrados na amostra selecionada podem ser extrapolados para um conjunto maior de dados no qual se esteja interessado, o que não necessariamente irá ocorrer. Sendo este o caso, deve-se reavaliar a eficácia da etapa de setup.

Tendo as etapas acima sido cumpridas com sucesso, o processo de data mining agora torna possível uma análise e interpretação muito mais eficiente dessas informações. Após tanto, pode-se dizer que uma enorme massa de dados tornou-se conhecimento, chave para decisões bem tomadas.

## 2.3 - Data Minings e sua aplicabilidade em BI

Conhecimento é poder. Assim sendo, tendo em mãos os resultados (válidos) de um data mining, empresas podem focar melhor seus “ataques” comerciais, por exemplo. Imaginemos um mercado que possua um site com suas promoções. A partir do conhecimento gerado pelo data mining, o mercado pode ver, por exemplo, que sexta-feira é o dia em que mais se vende um dado produto, como cerveja, e dar um destaque maior a isto. Pode ainda, classificando seu público-alvo, verificar que tipo de produto apetece a que tipo de pessoa, dentre outras infinitas possibilidades.

Pode-se ainda verificar sua aplicabilidade em evidenciar uma nova tendência de mercado, como, por exemplo, a queda no interesse de adultos do sexo masculino entre 22 e 40 anos por desktops, e aumento no interesse por portáteis.

As possibilidades são infinitas, sendo este, portanto, um investimento de retorno praticamente garantido.

## 3 - Data Warehouse

### 3.1 - O que é?

Surgido no final da década de 80 como um conceito criado por pesquisadores da IBM, os Data Warehouses tornaram-se possíveis com o avanço dos sistemas de informações mais robustos juntamente com o aumento da necessidade de análise de dados. Uma breve definição seria:

*O Data Warehouse (DW) representa o armazenamento de dados relativos às atividades de uma organização em um sistema de informação, cujo objetivo será auxiliar a tomada de decisões de maneira segura e rápida.*

Sistemas tradicionais não conseguiam cumprir o objetivo de análise dos dados com a simples geração de relatórios. Assim, a demanda por melhores ferramentas para auxílio abriu portas para que a implementação de Data Warehouses virassem realidade. Devido à possibilidade de manipular e analisar grandes volumes de dados, o Data Warehouse, hoje, é o principal responsável pelo apoio às decisões das principais soluções de Business Intelligence do mercado. A ferramenta mais popular para exploração de um Data Warehouse é a Online Analytical Processing (OLAP), mas muitas outras podem ser usadas.

Data Warehousing consiste em organizar informações corporativas de maneira confiável, consolidada, única e integrada. Permitindo que decisões sejam tomadas embasadas em fatos concretos e não em intuições, cruzando informações de diversas fontes, agilizando o processo e diminuindo os erros.

### 3.2 – Características

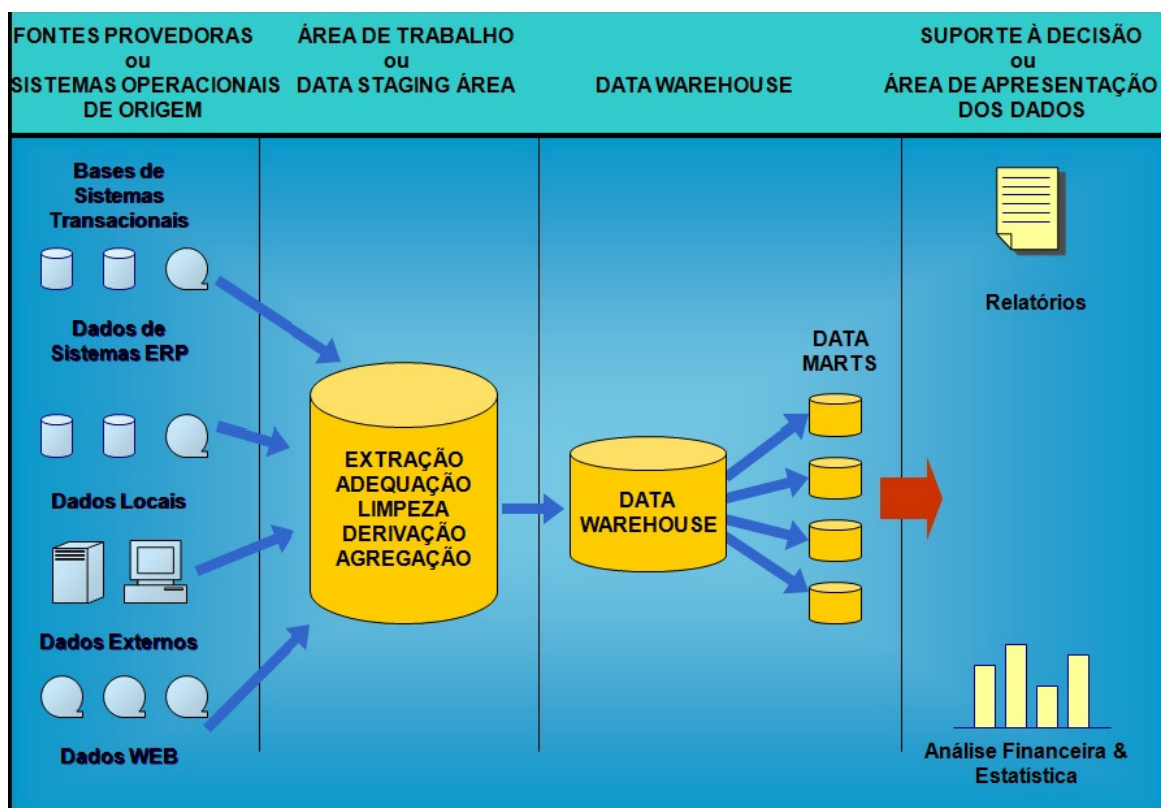
A primeira característica é ser orientado a assunto, ou seja, estará orientado ao redor do principal assunto da organização. Em contra-partida, o ambiente de negócio é organizado por aplicações funcionais, uma organização bancária terá aplicações para empréstimos, investimentos e seguros.

Outra característica, talvez a mais importante, é o fato que um DW deve ser integrado. A integração mostra-se em diferentes maneiras como na convenção consistente de nomes, na forma consistente das variáveis, na estrutura consistente de códigos, nos atributos físicos consistentes dos dados, por assim se dizer, na uniformidade da informação.

O Data Warehouse é, também, não-volátil. Permitindo apenas a carga inicial dos dados e consultas a estes dados. Após serem integrados e transformados, os dados são carregados em bloco, para que estejam disponíveis aos usuários. No ambiente operacional, ao contrário, os dados são atualizados registro a registro, em múltiplas transações. Essa volatilidade requer um trabalho considerável para assegurar integridade e consistência através de atividades de rollback, recuperação de falhas, commits e bloqueios. Não é requerido esse grau de controle dos sistemas orientados a transações.

Deve ser variante no tempo. Significa que o dado em um DW representa algum momento específico, ou seja, ele não é atualizável, ao passo que o dado em ambiente de produção é atualizado de acordo com mudanças de estado. A cada ocorrência de uma mudança, uma nova entrada é criada, para marcar esta mudança.

### 3.3 - Processos



[Adaptado de *SunExpert Magazine*, Outubro 1998.]

Sistemas operacionais de origem – Aplicações de negócio que capturam as transações da empresa. Estes devem ser considerados como externos ao DW, pois se presume que tenha pouco ou nenhum controle sobre o conteúdo e o formato dos dados. Os sistemas de origem também são chamados Sistemas Legados ou OLTP (Online Transaction Processing).

Data Staging Area – É tanto uma área de armazenamento como um conjunto de processos, normalmente denominada de ETL (Extract – Transformation - Load).

Data Warehouse/Data Mart – A área de apresentação dos dados é o local em que os dados ficam organizados, armazenados e tornam-se disponíveis para serem consultados diretamente pelos usuários para a aplicação de análise. Um Data Mart trata de problema departamental ou local e é definido como um subconjunto altamente agregado de dados, normalmente escolhido para responder a uma questão de negócio específica ao invés da corporação inteira.

Ferramenta de Acesso a Dados – O último componente principal do ambiente de Data Warehouse é a ferramenta de acesso a dados. Por definição, toda ferramenta de acesso a dados consulta os dados na área de apresentação do DW.

### 3.4 - Vantagens e Desvantagens

- Vantagens:
  - Um Data Warehouse fornece um modelo de dados comum para todas as informações de interesse, independentemente da fonte dos dados. Isto facilita quando há a necessidade para relatar e analisar as informações.
  - Antes de carregar os dados, as inconsistências são identificadas e resolvidas. Simplificando bastante a elaboração de relatórios e análise.
  - As informações estão sob o controle de usuários do DW, para que, mesmo se os dados do sistema-fonte é removida ao longo do tempo, as informações do armazém podem ser armazenadas com segurança por períodos prolongados de tempo.
  - Como são separados dos sistemas de negócios, Data Warehouses recuperam dados sem prejudicá-los.
  - Podem trabalhar em conjunto com aplicações de gestão de relacionamento com clientes (CRM), aumentando o valor das aplicações de negócio.
  - Facilitam os pedidos de apoio à decisão, tais como relatórios de tendências, relatórios de exceção e relatórios que mostram o desempenho real versus objetivos.

- Desvantagens
  - Não são o ambiente ideal para dados não-estruturados.
  - Como os dados devem ser extraídos, transformados e carregados, há um elemento de latência nos dados.
  - Durante sua vida, Data Warehouses podem ter custos elevados.
  - Podem ficar ultrapassados com relativa rapidez. Há um custo de fornecer informações que não são as melhores para a organização.
  - Muitas vezes existe uma linha tênue entre Data Warehouses e sistemas de negócio. Duplicações de funcionalidades de alto custo podem ser desenvolvidas ou a funcionalidades desenvolvidas no DW, deveriam ter sido desenvolvidas nos sistemas.

## 4 – OLAP

### 4.1 – Definição

OLAP é um acrônimo para *On-line Analytical Processing*. Como se pode notar, o nome é pomposo, porém diz muito pouco sobre o que e como é esta tecnologia de suporte a tomadas de decisão.

O termo *On-line* serve para indicar que, independente da (grande) quantidade de dados a serem analisados, o tempo de resposta de uma consulta deve ser baixíssimo, possibilitando assim um grau de interatividade elevado, interatividade esta, fundamental como iremos ver no decorrer do capítulo.

### 4.2 – Conceito

Diferentemente da mineração de dados, onde os dados são extraídos com o auxílio de algoritmos computacionais complexos, o foco do OLAP é, em posse de uma extensa base de dados, possibilitar que o setor de negócios de uma empresa a manipule diretamente de maneira simples, concisa e rápida.

O OLAP inova ao trazer um modelo de dados multidimensional, modelo este muito diferente do tradicional *linha x coluna* a qual estamos habituados a ver nos Sistemas de Gerenciamento de Banco de Dados (SGBDs) atuais. Justamente devido a este novo modelo que o OLAP é tão eficiente na tarefa em que se propõe a cumprir.

Pouco importa durante o processo da decisão um dado individual de um certo cliente ou de um certo produto, pois nenhum desses dados aponta uma tendência ou um problema. O conjunto desses dados é que é importante. E é nesse tipo de situação que o OLAP notabiliza-se. A sua natureza multidimensional favorece a análise granular e agregada de muitos dados ao mesmo tempo.

Por trás do modelo multidimensional existem três conceitos chave, conceitos estes que serão esmiuçados nos próximos parágrafos.

### 4.2.1 – Cubo

O cubo pode ser definido como uma estrutura capaz de analisar dados sob diferentes perspectivas de maneira rápida e eficiente. A disposição dos dados em cubos supera o problema de analisar grandes massas de dados existentes em bancos de dados bidimensionais. Pode-se pensar no cubo como uma extensão do conceito de planilhas. Em uma planilha as linhas representam uma dimensão (mês, por exemplo) e as colunas representam uma segunda dimensão (lucro, por exemplo). No cubo podemos ter duas ou mais dimensões.

Há que se dizer que é complicado imaginar algo com mais de três dimensões, porém lidamos com dados multidimensionais frequentemente ao emitir relatórios. Normalmente precisamos organizar a data por ano, por mês, por loja, por setor, etc.

### 4.2.2 – Dimensão

Uma dimensão é uma lista de tipos similares sob a perspectiva do usuário. Um bom exemplo seria dizer que anos, trimestres, meses, semanas, etc, formam um dimensão chamada tempo. O mesmo se aplica para pais, estado, cidade, bairro que podem formar uma dimensão chamada localização geográfica.

Pode-se perceber que dimensões normalmente exibem um comportamento hierárquico. Tiremos novamente para exemplo a dimensão tempo. Um ano é pai de trimestres, que é pai de meses e assim por diante. Da mesma forma ano pode ser pai também de mês, fornecendo assim um grande número de combinações possíveis durante a criação de um relatório.

Essa hierarquia possibilita o que é chamado em OLAP de *drill-down* e *drill-up*, que basicamente pode ser definido como o poder de navegar pelos níveis da hierarquia, aumentando ou diminuindo o detalhamento dos dados.

As dimensões atuam como índices para identificarmos valores dentro de um cubo.

### 4.2.3 – Medidas

Cubos armazenam medidas ou valores quantitativos. Como dito anteriormente, para podermos identificar um medida dentro do cubo precisamos de suas dimensões.

### 4.3 - Por que é tão rápido?

Mesmo com todos estes conceitos ainda não ficou claro o motivo pelo qual abordagem multidimensional do OLAP é tão eficiente.

Como foi dito acima, cubos armazenam medidas, medidas estas que representam valores que na maioria das vezes foi pré-consolidado na base de dados. Mas o que isso significa?

Como os valores, em quase sua totalidade já foram previamente calculados, fazer uma consulta que me retorne o número de alunos, por nível (graduação/pós-graduação), por curso, tipo do curso caso seja de graduação (bacharelado/licenciatura/etc), por situação da sua matrícula é um simples acesso as células do cubo indexadas pelas dimensões adequadas.

Há que se dizer que existe um motivo para nem todos os valores serem pré-calculados: o tamanho físico (em disco) ocupado pelo cubo seria muito grande, devido ao grande número de dimensões existentes. Para decidir qual campo será consolidado de antemão, vários fatores devem ser levados em consideração como, por exemplo, o tamanho ocupado pela agregação ou o custo de atualizar as células quando o banco de dados for atualizado.

### 4.4 - Arquitetura

Ferramentas OLAP foram criadas baseando-se no modelo cliente-servidor. É importante frisar que, em teoria, as duas partes podem ser distribuídas

separadamente, podendo assim o usuário utilizar o cliente de um fabricante e o servidor de outro.

O servidor tem que ser capaz de aceitar diversas conexões de diversos usuários simultaneamente sem comprometer a integridade dos dados (ou seja, bloquear recursos do cubo quando necessário) e fornecer confidencialidade (permissionamento, etc.).

Dependendo de como o servidor armazena os dados ele pode ser classificado de 3 maneiras: MOLAP (Multidimensional OLAP), ROLAP (Relational OLAP) ou HOLAP (Hybrid OLAP).

#### 4.4.1 – MOLAP

É a forma clássica do OLAP e é muitas vezes chamado apenas de OLAP. Esse tipo de servidor armazena os dados numa estrutura multidimensional otimizada. Devido a este fato é necessário executar uma tarefa conhecida como *processing* que consiste em pré-computar valores e armazená-los no cubo para uso futuro.

- Vantagens:
  - Consultas mais rápidas devidas a estrutura própria para o tipo de operação sendo realizado, indexação e cache.
  - Menor utilização de espaço em disco devido aos algoritmos de compressão utilizados.
- Desvantagens:
  - O *processing* pode demorar muito tempo se uma quantidade muito grande de dados tiver que ser processada (pode ser contornado se utilizarmos um processo incremental depois da primeira carga)
  - Algumas implementações desse tipo são suscetíveis a um fenômeno conhecido como *database explosion*, que significa, basicamente, que o banco de dados começa a consumir muito espaço em disco devido a certas circunstâncias (número elevado de dimensões, por exemplo)

#### 4.4.2 - ROLAP

Esse tipo de servidor guarda os dados em um banco de dados relacional e gera consultas para calcular as informações apropriadas quando o usuário as requisita. É bem verdade que mesmo o ROLAP usando um banco de dados relacional para armazenar os dados, a modelagem de suas tabelas tem que ser cuidadosamente elaborada (utiliza-se o conceito de *star-schemas*) impossibilitando assim que um banco planejado para uso convencional (Online Transaction Processing) funcione como repositório OLAP.

- Vantagens:
  - Muitos consideram o ROLAP mais escalável (lida melhor com um maior número de informação)
  - Pode ser acessado por qualquer ferramenta de *reporting* que suporte SQL
  - Menor tempo de *ETL*
- Desvantagens:
  - Consultas são mais lentas devido a estrutura de dados utilizada.
  - Tabelas de agregação devem ser criadas manualmente (mais código para ser mantido)

#### 4.4.3 – HOLAP

Tentativa de unir o melhor dos mundos MOLAP e ROLAP. Armazenando dados de agregação em um *backend* MOLAP e dados detalhados/descriptivos num *backend* ROLAP, o servidor utiliza a melhor característica de cada estrutura de dados de forma ótima baseado na consulta requisitada.

## 5 - Self-Service BI

### 5.1 – Motivação

Existem muitas instituições que são dependentes de sistemas de informação específicos que as auxiliam em sua gerência. Na maioria dos sistemas de informação, nem todas as consultas necessárias aos analistas de negócio estão cobertas, criando a necessidade de criação de consultas ad hoc.

O problema da geração de consultas ad hoc é que analistas de negócio geralmente não possuem capacidade técnica para desenvolver consultas ao banco de dados, o que acaba resultando numa requisição de desenvolvimento de consulta para a equipe de Tecnologia da Informação da instituição.

Parte da equipe de TI, que possui as capacidades técnicas para o desenvolvimento de uma consulta acaba por ser desviada de suas atividades de desenvolvimento e manutenção do Sistema para criar a consulta. O mesmo ocorre com requisições de relatórios.

Além da redução na produtividade, é notado também um problema mais grave, que é relacionado ao tempo levado para a geração das consultas pela equipe de TI. A demora do desenvolvimento das consultas pode atrasar a análise dos dados e comprometer uma tomada de decisão. A demora, muitas vezes ocorre por falta de conhecimento profundo do domínio de negócio por parte dos desenvolvedores.

Geralmente, analistas de negócio possuem um conhecimento muito maior do domínio do que os desenvolvedores e muitas vezes numa especificação de consulta podem haver termos e regras de negócio que são mal interpretadas por parte dos desenvolvedores, levando a dados errôneos ou imprecisos que conseqüentemente causam retrabalho no desenvolvimento da consulta.

Tendo em vista tais desvantagens, podemos concluir que ao adicionarmos a capacidade dos próprios analistas de negócio produzirem suas consultas ad hoc, teremos ganhos na produtividade, tanto no setor de TI quanto no setor de Gestão, com a geração de consultas e relatórios mais rápidas e precisas, devido ao maior conhecimento do domínio pelo analista de negócios e uma menor interrupção da equipe de TI. Neste ponto, entra o conceito de Self-Service BI.

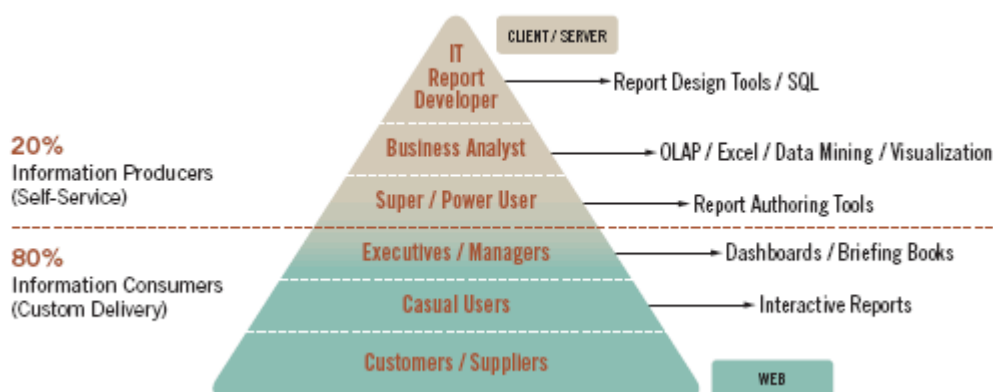
## 5.2 - Os prós e contras

Ser menos interrompido para criar consultas e relatórios é o sonho dourado de qualquer equipe de TI. Ter poder e acesso total e direto aos dados também é o sonho dourado de qualquer analista de negócio. Nesta hora é conveniente lembrar da frase "*grandes poderes trazem grandes responsabilidades*".

Qualquer equipe de TI sabe que muito poder na mão de muitos usuários não é muito bom. Usuários com muito poder acabam por ter muitas responsabilidades de gerar as informações e relatórios que precisam para realizar seus trabalhos. A verdade é que a maioria dos usuários não quer esta responsabilidade e nem é parte de seu trabalho tê-la.

Uma pesquisa[1] mostra que a maior parte dos relatórios gerados são, na verdade, personalizações pontuais de uma quantidade pequena de temas, além disso, são usados uma quantidade mínima de vezes e acabam por ficarem esquecidos, consumindo espaço desnecessário em disco. A grande quantidade de relatórios parecidos gerados pelos usuários torna muito difícil a manutenção e gerência dos mesmos, reduzindo a usabilidade do sistema pelos próprios usuários, caindo no que se chama de *reporting chaos*.

Uma pesquisa mais aprofundada revela que 80% dos usuários apenas consomem informações geradas pelos 20% restantes dos usuários.



Estes 80% de consumidores são usuários casuais, que consistem de trabalhadores de "linha de frente", como vendedores e gerentes.

Podemos concluir que mesmo com toda a evolução das ferramentas de BI, tornar a interface mais amigável e tornar os dados mais acessíveis, surgiu um novo problema gerencial, como vimos com o *reporting chaos*.

[1] <http://tdwi.org/articles/2007/10/18/the-myth-of-selfservice-bi.aspx>

## 6 – Conclusões

A tecnologia sempre trouxe soluções para os mais variados problemas que surgiram com o Business Intelligence, porém, como vimos anteriormente, nenhuma tecnologia resolve todos os problemas. O resultado da pesquisa que mostra que apenas 20% dos usuários de um sistema de BI são produtores de informação, junto com o problema do *reporting chaos* nos diz que existe um problema na definição de papéis dos usuários no sistema.

Uma empresa precisa definir regras estritas de permissionamento e acesso fazendo jus à proporção de consumidores e produtores de informação. A definição de tipos de relatórios básicos, complexos e genéricos também deve ser levada em consideração a fim de evitar a criação desnecessária e pontual de relatórios.

A criação de consultas e relatórios ad hoc pode ser atenuada pelas ferramentas de Self-Service BI, mas não irão desaparecer. Sempre haverá um caso especial que está além das capacidades do sistema e adicionar tal capacidade não justifica o alto custo de implementação, quando comparado ao custo de desenvolvimento do relatório ou consulta.

Cabe à gestão da empresa balancear o uso das ferramentas de BI, alcançando o equilíbrio na desempenho de seus processos que requerem acesso e controle dados.

## 7- Referências

WIKIPÉDIA. In: Data Warehouse. 2010. Disponível em:  
<[http://pt.wikipedia.org/wiki/Data\\_Warehouse](http://pt.wikipedia.org/wiki/Data_Warehouse)>. Acesso em: 15 de Julho de 2010

DW. In: Data Warehouse. 2010. Disponível em:  
<<http://www.datawarehouse.inf.br/dw.htm>> e  
<[http://www.datawarehouse.inf.br/Academicos/A%20PUBLICAR\\_DATA\\_WAREHOUSE\\_MARCELL\\_OLIVEIRA.pdf](http://www.datawarehouse.inf.br/Academicos/A%20PUBLICAR_DATA_WAREHOUSE_MARCELL_OLIVEIRA.pdf)>. Acesso em: 15 de Julho de 2010

Soares, Ismael In: Palestra pela empresa BlueSoft, 2010 Disponível em:  
<<http://vimeo.com/6962558>>. Acesso em: 19/07/2010

Microsoft BI TV In: History of Business Intelligence, 2010. Disponível em:  
<[http://www.youtube.com/watch?v=\\_1y5jBESLPE&feature=related](http://www.youtube.com/watch?v=_1y5jBESLPE&feature=related)> Acesso em:  
19/07/2010

Eckerson, Wayne W. In: The Myth of Self-Service BI, 2010. Disponível em:  
<<http://tdwi.org/articles/2007/10/18/the-myth-of-selfservice-bi.aspx>>. Acesso em:  
20/07/2010

WIKIPÉDIA. In: Data Warehouse. 2010. Disponível em:  
<[http://pt.wikipedia.org/wiki/Data\\_mining](http://pt.wikipedia.org/wiki/Data_mining)>. Acesso em: 17 de Julho de 2010

Frاند, Jason In:DataMining. 2010. Disponível em:  
<[http://www.anderson.ucla.edu/faculty/jason.frاند/teacher/technologies/palace/data\\_mining.htm](http://www.anderson.ucla.edu/faculty/jason.frاند/teacher/technologies/palace/data_mining.htm)>. Acesso em: 17 de Julho de 2010

WIKIPEDIA. In: Online\_analytical\_processing  
2010 [http://en.wikipedia.org/wiki/Online\\_analytical\\_processing](http://en.wikipedia.org/wiki/Online_analytical_processing). Acesso em:  
15/07/2010

WIKIPEDIA. In: Hybrid OLAP 2010 <http://en.wikipedia.org/wiki/HOLAP>. Acesso em:  
15/07/2010

WIKIPEDIA. In: Relational OLAP 2010 <http://en.wikipedia.org/wiki/ROLAP>. Acesso em:  
15/07/2010

WIKIPEDIA. In: Multidimensional OLAP 2010 <http://en.wikipedia.org/wiki/MOLAP>.  
Acesso em: 15/07/2010

WIKIPEDIA. In: OLAP cube 2010 [http://en.wikipedia.org/wiki/OLAP\\_cube](http://en.wikipedia.org/wiki/OLAP_cube). Acesso em:  
15/07/2010

WIKIPEDIA. In: Hybrid OLAP 2010 <http://en.wikipedia.org/wiki/HOLAP>. Acesso em:  
15/07/2010

WIKIPEDIA. In: Hybrid OLAP 2010 <http://en.wikipedia.org/wiki/HOLAP>. Acesso em: 15/07/2010

Pendse, Nigel. In: What is OLAP? 2010 <http://www.bi-verdict.com/fileadmin/FreeAnalyses/fasmi.htm>. Acesso em: 16/07/2010

Mailvaganam, Hari. In: Introduction to OLAP [http://www.dwreview.com/OLAP/Introduction\\_OLAP.html](http://www.dwreview.com/OLAP/Introduction_OLAP.html). Acesso em 16/07/2010

OLAP Concil. In: <http://www.geoplan-systems.de/Dokumente/OLAP-Glossary.pdf>. Acesso em: 18/07/2010

Swanhart, Justin. In: <http://www.mysqlperformanceblog.com/2010/07/12/intro-to-olap/>. Acesso em: 16/07/2010