

SISTEMAS DE INFORMAÇÃO II

CAPÍTULO 1.1

SISTEMAS DE DATA WAREHOUSING: INFRA-ESTRUTURA INFORMACIONAL E FUNCIONAL

FILIPESÁ
FILIPES.A@ISEC.PT

1.1. Data warehouse como infraestrutura informacional de uma organização

1 Data Warehouse

Definição

(1) “Um Data Warehouse é simplesmente um repositório de dados, único, completo e consistente, dados esses obtidos de uma variedade de fontes e tornado disponível a utilizadores finais de uma forma que eles possam compreendê-los e usá-los no contexto do negócio.”

Barry Devlin, IBM Consultant

(2) “Um Data Warehouse é uma colecção de dados
orientada por temas,
Integrada,
não volátil,
variante no tempo,
para suporte do processo de tomada de decisões de gestão.”

Inmon, W.H., Building the Data Warehouse, 1993/96

1.1. Data warehouse como infraestrutura informacional de uma organização

1 Data Warehouse

Definição

(3) “ É um processo em andamento que aglutina dados de fontes heterogêneas, incluindo dados históricos e dados externos para atender às necessidades de consultas estruturadas e ad-hoc, relatórios analíticos e de suporte a decisão ”

[Harjinder]

(4) “ É uma coleção de técnicas e tecnologias que juntas disponibilizam um foco pragmático e sistemático para tratar com o problema do utilizador final de aceder a informações que estão distribuídas em vários sistemas da organização ”

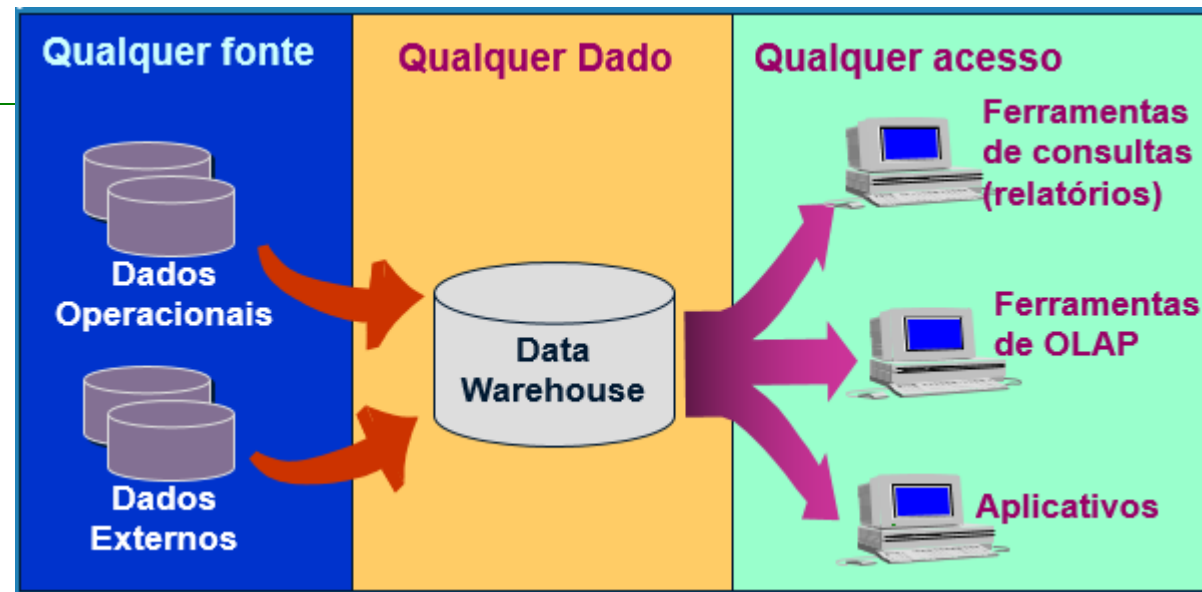
[Barquini]

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.1. Definição

Data Warehouse não é o fim, é um meio que as empresas dispõem para analisar informações podendo utilizá-las para a melhoria dos processos atuais e futuros



1. Sistemas de Data Warehousing: Infraestrutura Informacional e Funcional

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. O que é uma DW?

- Sistema que armazena dados históricos usados no processo de tomada de decisão
- Integra os dados corporativos de uma empresa num único repositório
- Para que serve?
 - Para criar uma visão única e centralizada dos dados que estavam dispersos em diversos BDs
 - Permite que utilizadores finais executem consultas, criem relatórios e façam análises
- Nota: A definição de DW será aprofundado no capítulo 1.1.2

1. Sistemas de Data Warehousing: Infraestrutura Informacional e Funcional

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

- Informação
 - Melhor recurso do qual uma empresa pode dispor para tomar decisões estratégicas
 - Obtida analisando dados históricos sobre vendas, produção, clientes, etc.
- Análise dos dados
 - Fornece informações vitais para a empresa
 - Pode aumentar a competitividade da empresa
 - Era feita intuitivamente pelos gestores

1. Sistemas de Data Warehousing: Infraestrutura Informacional e Funcional

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

- Dificuldades para obter informação
 - Quantidade de dados a serem analisados crescem com a expansão dos negócios e com o passar dos anos
 - Dados oriundos de fontes diferentes podem criar informações dispares
 - Impossível para um ser humano manter e analisar todos os dados
 - Informação não é mais mantida por gestores devido à mobilidade no mercado de trabalho

1. Sistemas de Data Warehousing: Infraestrutura Informacional e Funcional

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. DW deve:

- Integrar dados de múltiplas fontes
- Facilitar o processo de análise sem impacto para o ambiente de dados operacionais
- Obter informação de qualidade
- Atender diferentes tipos de utilizadores finais
- Ser flexível e ágil para responder a novas análises

1. Sistemas de Data Warehousing: Infraestrutura Informacional e Funcional

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Introdução (1)

In the beginning were applications. Users thought that applications would provide them with information. And insofar as the stated requirements of the applications were concerned, the applications sufficed. But over time the business requirements changed. **Keeping the applications in sync with the changing business requirements was a difficult thing to do.**

Along the way in trying to keep up with changing requirements, the end user encountered some other limitations to the world of applications. Those limitations were the need for integration, and historical data.

The applications that the corporations had created or otherwise acquired **had no notion of integration**. One application thought a customer was one thing. Another application has yet another interpretation of what a customer was. When it came to the corporate understanding of data, there was – simply stated – no corporate understanding. From a corporate perspective the manager could not answer such basic questions as: Who is a customer? What is a product? What is a transaction?

In short, the different applications were never designed to work together in an integrated manner.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Introdução (2)

The **2nd issue** was that **applications focused inevitably on very current data**. Applications could reveal: 1) how much money a customer has in the bank **right now**; or 2) where a shipment was **right now**; 3) what the status of an insurance policy was **right now**; 4) what the quota of a salesperson was **right now**; 5) and so on.

The applications were designed to keep track of what is going on right now. But when it came to a sense of need and importance of historical information, the applications treated historical data with no respect at all.

Unfortunately, **integration and history a very important component of information**. And applications simply did not measure up (*estavam à altura*).

The end user's first reaction was to **rewrite the applications of yesterday**. But this idea quickly fell by the wayside. The end user found that – as far as the applications were concerned – **the clock could not be turned back**. There were simply too many applications, too much undocumented code, too much fragile code, too much complexity to even attempt to roll back the tide of applications. Thus **was born the notion of a data warehouse**, an alternative to the **dilemma** of the end user who **needed information but could not impose change on the legacy applications environment**.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Introdução (3)

Like all radical and fresh concepts, the notion of a data warehouse was derided and scorned (*escarnecida e refutada*) by academics and theoreticians. Since the idea of a data warehousing had not risen among their ranks, it could not possibly be a valid concept. Today data warehousing is no longer a theory. It is conventional wisdom, and corporations around the world recognize that the road forward leads to the data warehouse.

Data warehousing forms the center of a wide universe. From the corporate data warehouse, with its granular, corporate integrated data, spring many different kinds of decision support activity. The data warehouse form the basis for such DSS processing as 1) data mart, department processing; 2) simple reporting of corporate information; 3) exploration processing; 4) data mining; 5) operational data stores; 6) project warehousing, and so forth.

But data warehousing did not happen all at once. Like a jigsaw puzzle, data warehousing had been put together a piece at a time. The world of data warehousing has been led by writers and by practitioners who became writers. These leaders have described from their experience what works and what does not.

W. H. Inmon, foreword from Objected-Oriented Data Warehouse Design

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Evolução Histórica: 4 Gerações de Ambientes Computacionais

Um Data Warehouse (DW) é um ambiente, não um produto.

Um DW é um artefato arquitectural dos sistemas de informação que proporciona aos utilizadores informação de suporte à decisão actual e histórica a que é difícil de aceder ou não está presente nas bases de dados operacionais.

De uma perspectiva de negócio, de forma a sobreviver e prosperar no ambiente de negócios global, os gestores necessitam de respostas relativas às suas empresas, porque:

- As decisões têm de ser tomadas com rapidez e de forma correcta, utilizando os dados disponíveis;
- Os utilizadores são especialistas de negócio e não profissionais de SI/TI;
- Os dados disponíveis duplicam a cada 18 meses, o que afeta o tempo de resposta e a capacidade de compreender o seu conteúdo.

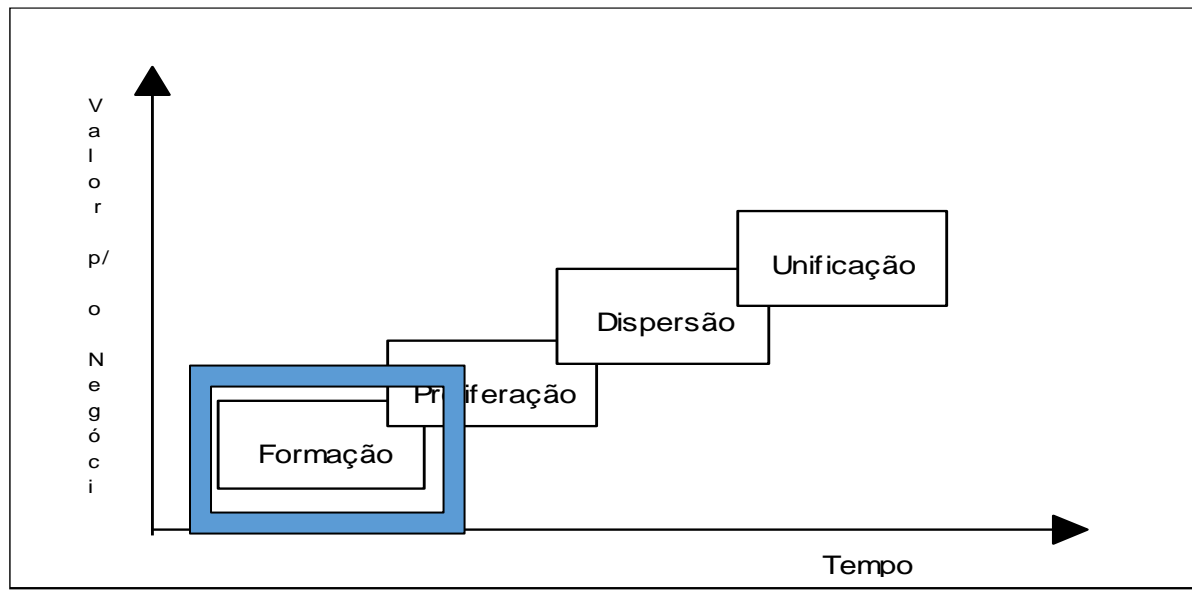
Também as organizações distribuíram o controlo para fora do nível médio que, tradicionalmente, proporcionava e monitorizava a informação de negócio. Os utilizadores dependem cada vez mais da informação – desde as medidas relativas aos fatores críticos até à informação vital relacionada com os eventos de negócio – assim, a necessidade para proporcionar um information warehouse, para os restantes, tornou-se mais crítica.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Evolução Histórica: 4 Gerações de Ambientes Computacionais

O Data Warehousing é resultado de evolução havida nos ambientes computacionais.

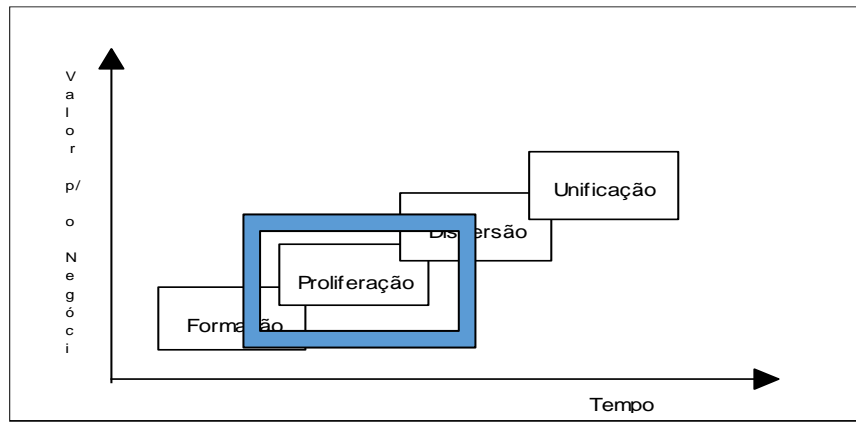


1.^a - Formação: O desenho das aplicações é modelado pelo conhecimento dos requisitos do negócio, sendo focado nas necessidades imediatas da empresa, sendo adquiridas ou desenvolvidas uma de cada vez, levando assim à sua completa independência entre si

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Evolução Histórica: 4 Gerações de Ambientes Computacionais

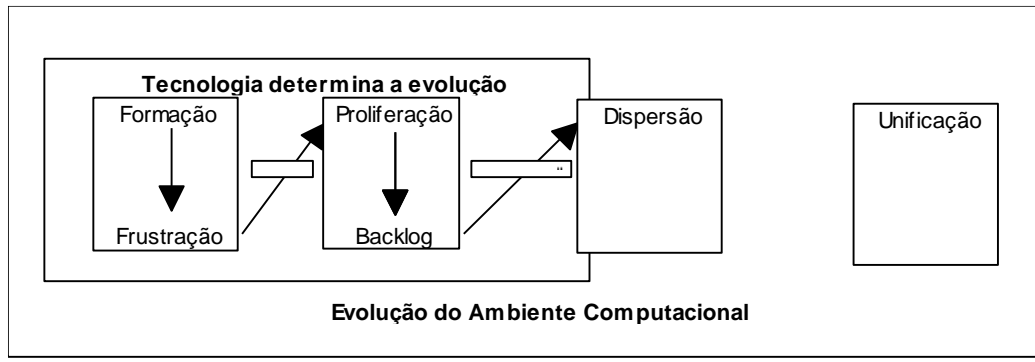


2.^a - Proliferação: A ênfase é colocada mais do que na eficiência dos processos, na eficácia e gestão do trabalho, proporcionando funcionalidade adicional ao valor das tarefas executadas. A mudança operou-se a nível do reconhecimento do valor da informação sobre os dados e surgem as bases de dados relacionais. Em vez de ilhas de automação característica da 1.^a geração de ambientes computacionais, as empresas enfrentam agora a existência de uma miríade de sistemas interligados, empregando múltiplas tecnologias, uma variedade de linguagens de programação e sistemas de gestão de base de dados e documentação incompleta ou não precisa, relativa ao ambiente e seus componentes.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Evolução Histórica: 4 Gerações de Ambientes Computacionais



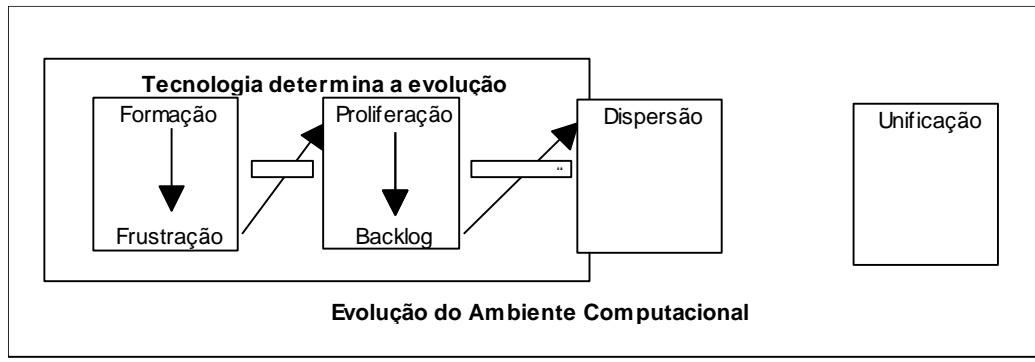
3.^a - Dispersão: A comunidade de utilizadores começa a usar as “suas” ferramentas para satisfazer as suas necessidades de informação, muitas vezes à revelia do departamento de processamento de dados, levando ao ambiente, dito de “Dispersão”.

A vulgarização do computador pessoal (PC) e o advento dos ambientes tipo cliente/servidor proporcionaram às empresas um conjunto de novas opções para satisfação das necessidades de informação. Esta geração é caracterizada pelo crescimento explosivo do número de opções disponíveis para atingir a quase insaciável necessidade de informação na empresa e pela dispersão dessas opções pela companhia. A informação é reconhecida como tendo um valor estratégico.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Evolução Histórica: 4 Gerações de Ambientes Computacionais



3.^a - Dispersão

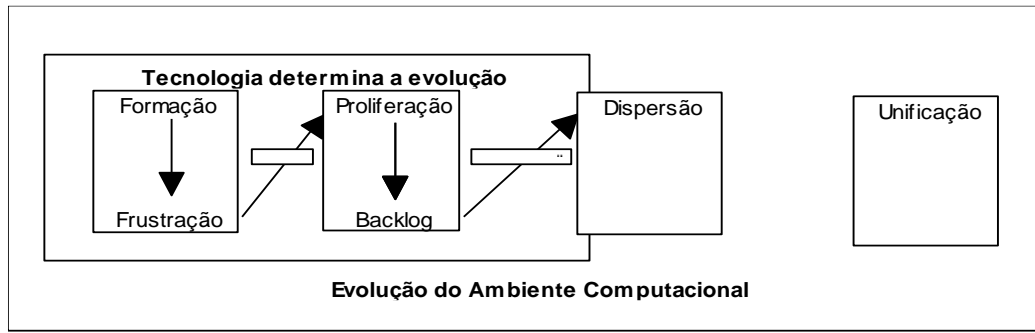
Mas, devido à procura crescente de informação por parte dos “knowledge workers”, originaram-se situações desagradáveis.

- como a execução das consultas obrigava à junção, ordenação e agrupamento de uma imensidão de registos, eram tarefas muito demoradas e consumidoras de recursos, que entrando imediatamente em conflito com o processamento transaccional em linha, que teria prioridade absoluta;
- o intervalo de tempo destinado a este tipo de actividade foi assim relegado para o final do dia ou mesmo para a noite;
- a estrutura dos dados estava otimizada para o processamento transaccional e não para consultas, levando a tempos de resposta muito alargados.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Evolução Histórica: 4 Gerações de Ambientes Computacionais



3.ª - Dispersão

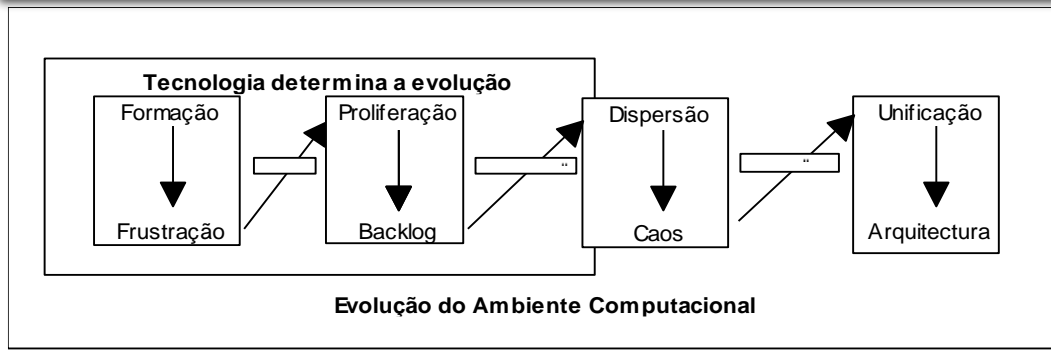
A solução encontrada para parte dos problemas foi mover os dados para os PC, utilizando os chamados programas de extracção, para aí serem analisados.

- Mas:** a) os dados não eram credíveis: o mesmo relatório, obtido de fontes diferentes, dava resultados díspares, devido às clássicas falhas de integração;
- b) a utilização de dados operacionais, obtidos em timings diferentes, contribui para as dificuldades de integração já referidas;
- c) dados originados externamente, uma vez integrados, perdem a sua identidade (faltam metadados - dados acerca dos dados);
- d) caso uma análise necessite de dados existente em vários sistemas operacionais, é difícil localizá-los, verificar a sua estrutura, resolver as polissemias e sinónimos e, finalmente, procurar os programadores e analistas para criar os programas de extracção;
- e) o horizonte temporal histórico é quase sempre insuficiente para a maioria das análises.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Evolução Histórica: 4 Gerações de Ambientes Computacionais



4.^a - Unificação

2 Soluções:

- Radical
- Adaptativa

Solução radical - Reformular completamente os SI das empresas com custos enormes, não só financeiros, mas, e principalmente, de tempo e risco. Foi tentada, em muitos casos, sem grandes resultados.

Solução adaptativa - Abandono da velha máxima: um sistema, duas funções; uma base de dados que deve servir o processamento de carácter operacional e o processamento de suporte às decisões. Passaríamos a ter **duas categorias de sistemas distintos**:

1. o sistema que suportaria o **processamento operacional** e
2. o sistema de **suporte à decisão**

Esta dicotomia de sistemas na empresa foi inicialmente preconizada por W.H. Inmon que baptizou o novo sistema unificador de "**Data Warehouse**"

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Solução Adaptativa: 1) Integração dos Dados Conduzida pelas Interrogações

Mas porquê ter um repositório físico separado e não **integrar os dados a pedido (abordagem lazy)**?

✧ **PRÓS**

- Acesso a dados mais actuais (todas as fontes de dados directamente acessíveis);
- Nenhuma duplicação de dados.

✧ **CONTRAS**

- Atrasos no processamento de interrogações;
- Fontes de dados lentas ou indisponíveis;
- Filtragens e processo de integração complexo;
- Compete com o processamento local nas fontes;
- Dados perdidos nas fontes (e.g., dados históricos) não podem ser recuperados.

✧ **Esta abordagem não teve aceitação, apesar de ser, porventura, apelativa.**

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

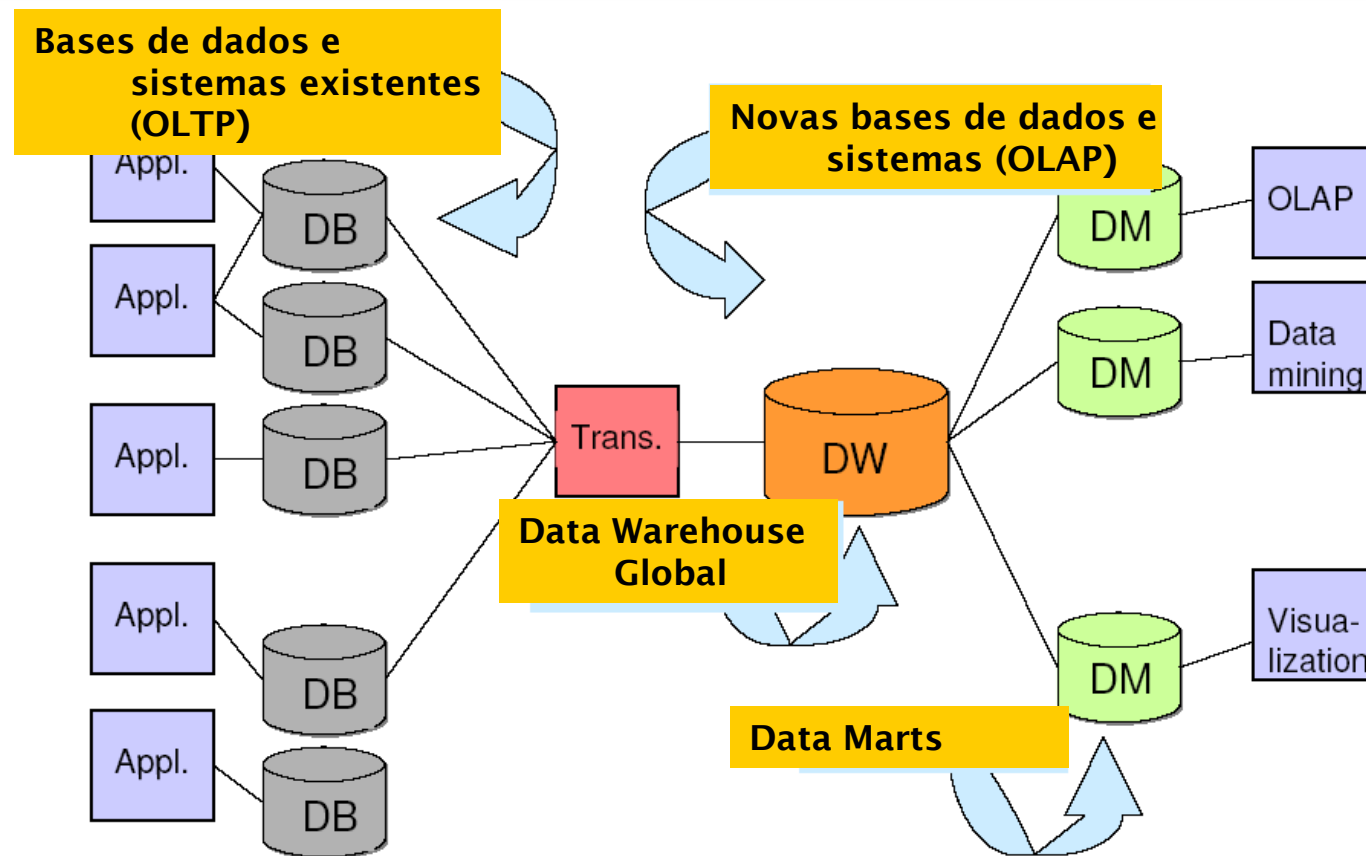
Solução Adaptativa: 2) Integração dos Dados Conduzida pelo DW

- ✧ Os dados são integrados e preparados em avanço (abordagem *eager*);
- ✧ Os dados são armazenados no DW para resposta a interrogações e análises.
- ✧ PRÓS
 - Alto desempenho em interrogações;
 - Não interfere com o processamento local nas fontes;
 - Assume que a actualização do DW é possível durante o período de não utilização do processamento local;
 - As interrogações complexas correm no DW;
 - As interrogações OLTP(Online Transaction Processing) correm nos sistemas fonte.
- ✧ CONTRAS
 - Alguma duplicação de dados;
 - Só dados mais actualizados (possivelmente do próprio dia) não estão ainda disponíveis.
- ✧ Esta solução foi adoptada pela indústria.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Solução Adaptativa: 2) Integração dos Dados Conduzida pelo DW



1.1. Data warehouse como infraestrutura informacional de uma organização

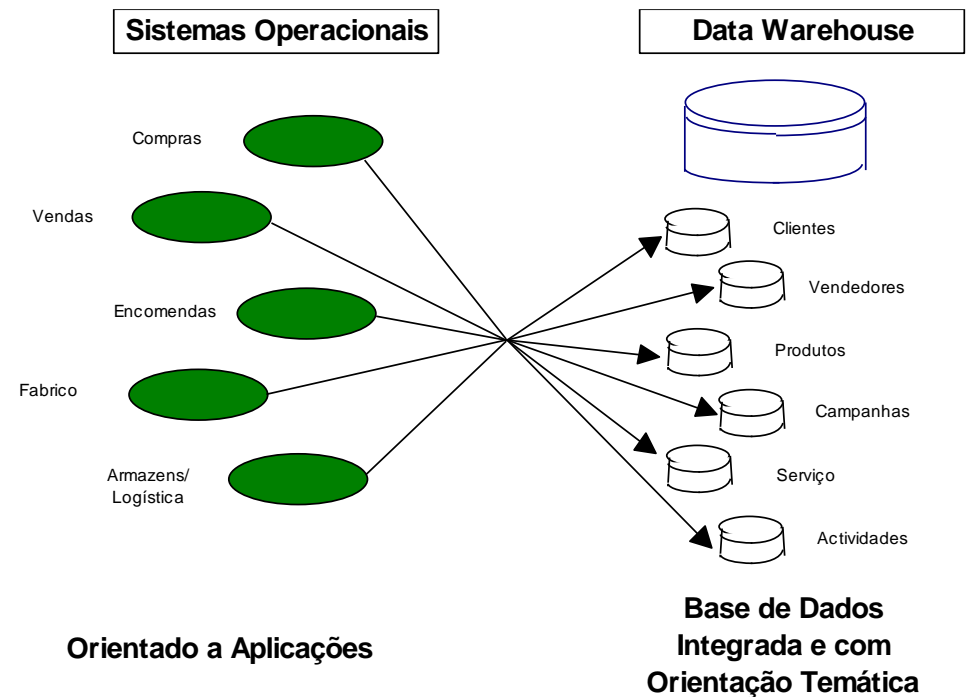
1.1.1. Porquê um DW?

Bases de Dados Operacionais e Informacionais

A evolução dos sistemas de informação nas empresas, reportando-nos às três primeiras gerações, pode resumir-se numa frase:

a nossa capacidade de recolher e armazenar dados de todos os tipos ultrapassou de longe a nossa possibilidade de os analisar, sumarizar e extrair conhecimentos desses dados.

Enquanto que a tecnologia das bases de dados nos forneceram as ferramentas básicas que permitiram um armazenamento e pesquisa eficiente de grandes volumes de dados, a questão de como **ajudar a compreender e analisar esses dados** não teve resposta eficaz.



1.1. Data warehouse como infraestrutura informacional de uma organização

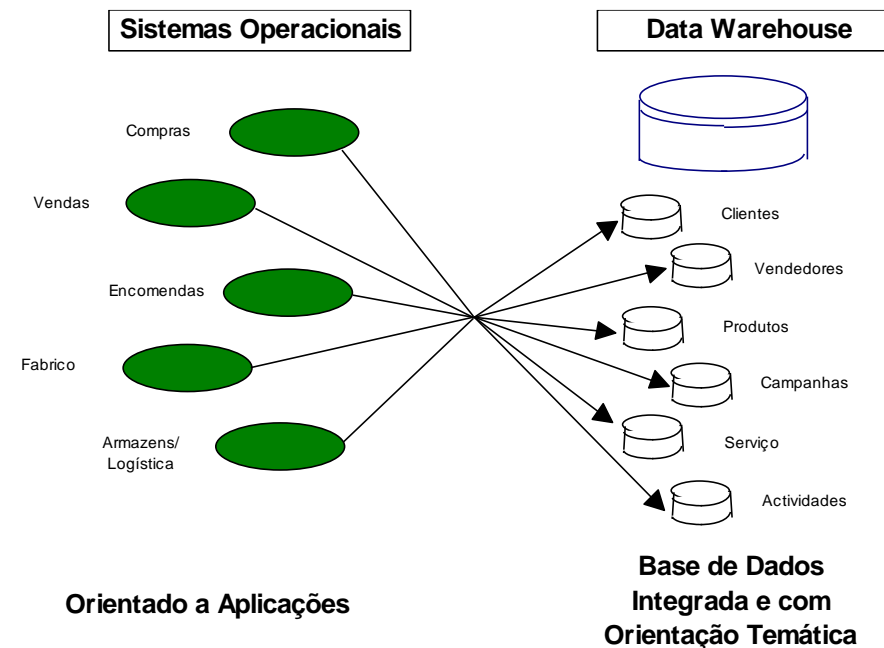
1.1.1. Porquê um DW?

Bases de Dados Operacionais e Informacionais

O problema tem ainda duas facetas mais:

- Uma consiste na diversidade de sistemas de gestão de bases de dados e modelos de dados existente numa mesma empresa. Qualquer tentativa de obter uma visão global do negócio esbarrava com grandes dificuldades ou mesmo impossibilidades. Há, assim, uma necessidade clara de isolar o negócio das diferenças em tecnologias e modelos de dados.

- Outra faceta da questão em que importa assentar, desde já, (e que contribuía para o desencanto dos gestores do negócio, face à informação de que dispunham) consistia **na visão distinta que os gestores têm do negócio, relativamente ao modelo de dados habitual nos sistemas transacionais**. O analista de negócio vê a empresa de forma temática (por assuntos) e faz a sua análise com base num pequeno número de valores (medidas) referentes a factos que são avaliados, segundo um conjunto de perspectivas (dimensões), que fornecem um contexto aos valores.



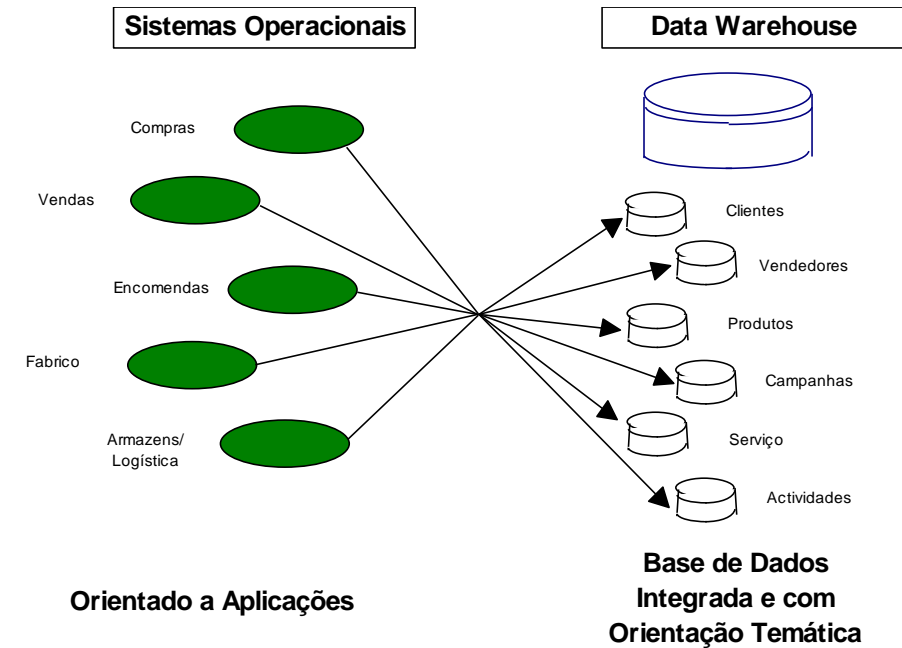
1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Bases de Dados Operacionais e Informacionais

A forma característica de ver a empresa por parte dos analistas de negócio, traduz-se num modelo dimensional, que será, adiante, abordado de forma mais aprofundada.

Este modelo é radicalmente diverso dos modelos de dados altamente normalizados, típicos das aplicações OLTP. Isto torna a visualização da informação numa perspectiva multidimensional, impossível ou, no mínimo, difícil e especialmente demorada.



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Bases de Dados Informacionais vs. Operacionais

- **Acesso aos dados:** para os dados informacionais, tende a ser had-hoc, não pré-definido e estruturado;
- **Modelo de dados:** reflete as necessidades analíticas dos utilizadores, enquanto o modelo de dados operacionais é normalizado para suportar as propriedades da atomicidade, consistência, isolamento e durabilidade (ACID);
- **Base temporal:** Dados recentes e históricos, agregados e derivados, enquanto os dados operacionais tendem a ser correntes ou um snapshot de dados recentes;
- **Alterações aos dados:** Para os dados informacionais são periódicas, agendadas por actualizações em batch, enquanto que os dados operacionais estão sujeitos a alterações contínuas de elevada frequência.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

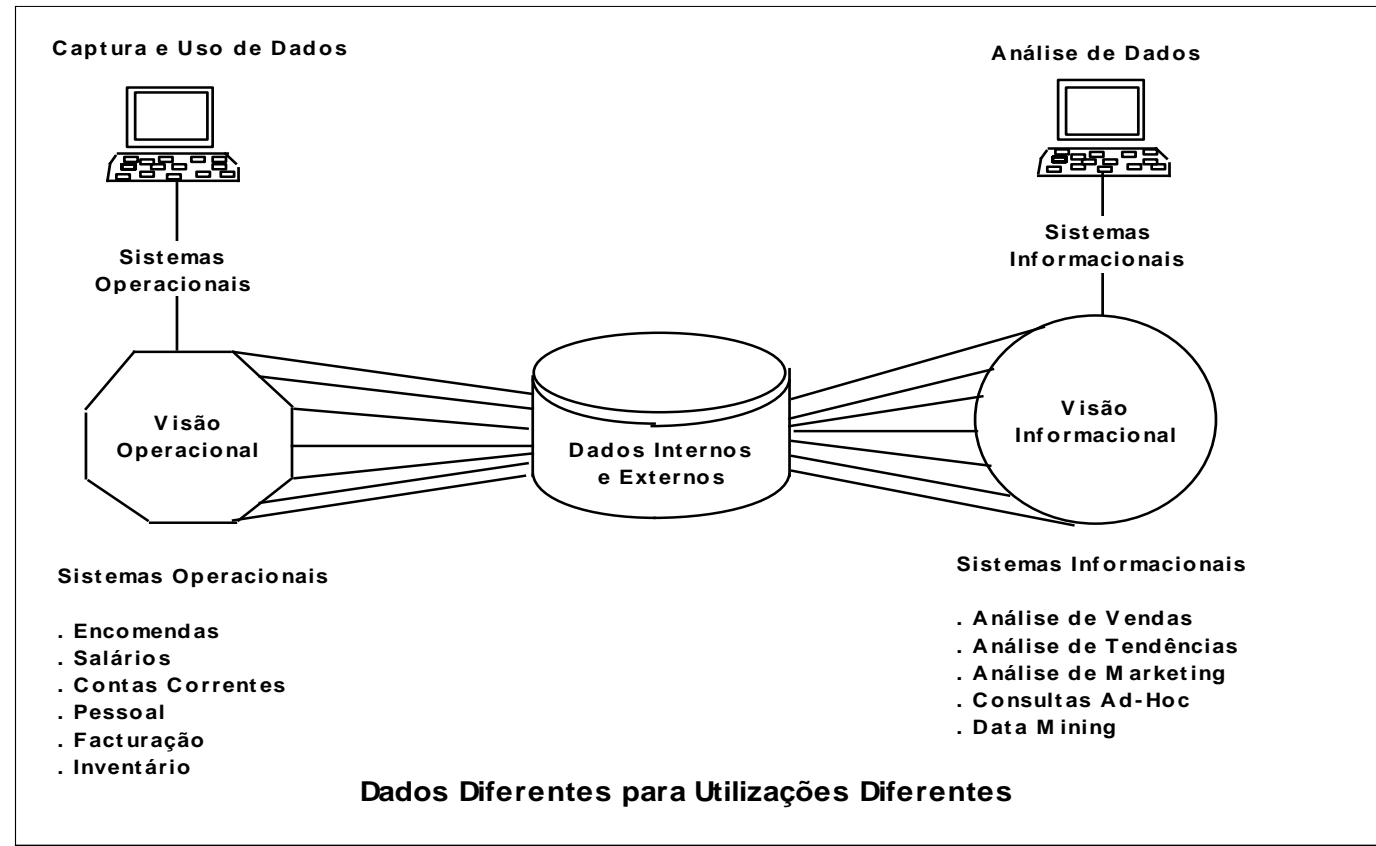
Bases de Dados Informacionais vs. Operacionais

- **Unidade de trabalho:** os dados informacionais são objecto de queries, enquanto que os dados operacionais são objecto de inserções, actualizações e remoções concorrentes;
- **Registos acedidos por transacção:** Milhões, para dados informacionais, versus dezenas para operacionais;
- **Número de utilizadores concorrentes:** Tipicamente, dezenas ou centenas para dados informacionais *versus* milhares para operacionais
- **Volume de transacções:** Relativamente baixo para dados informacionais, mas alto para dados operacionais;
- **Tipo de utilizadores:** Analíticos ou de gestão *versus* operacionais;
- **Número de índices:** Geralmente muitos e complexos *versus* poucos e simples.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

Bases de Dados Operacionais vs Informacionais



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW?

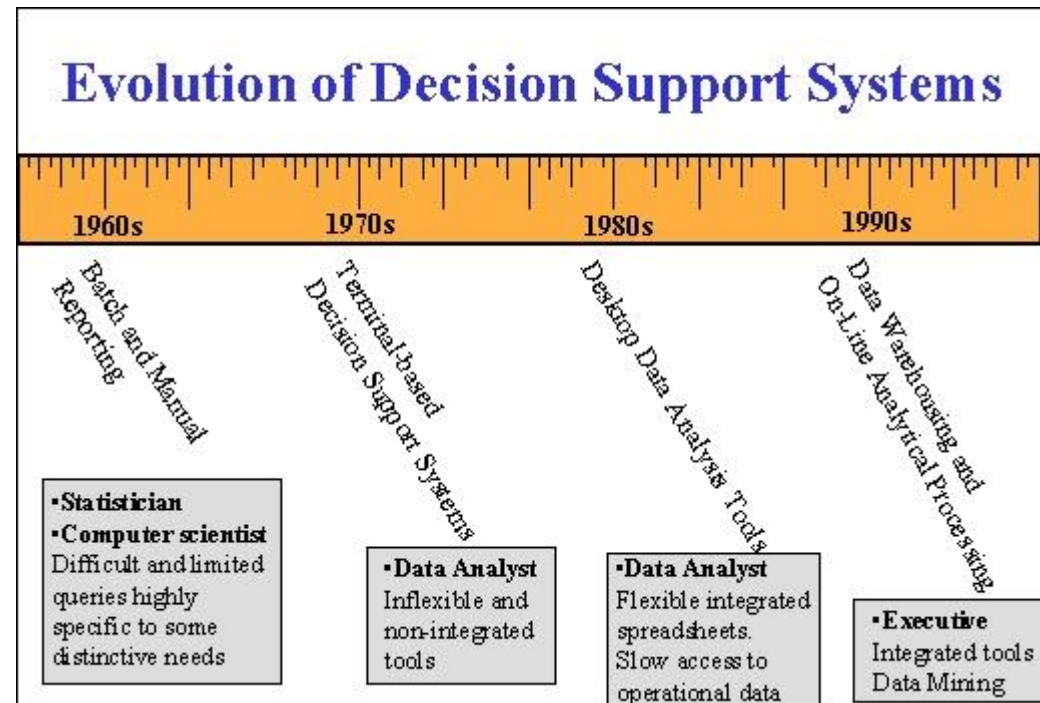
Bases de Dados Operacionais e Informacionais (resumo)

	Dados operacionais	Dados informacionais
Conteúdo	Valores actuais	Sumarizados, arquivados, derivados
Organização	Por aplicação	Por assunto
Estabilidade	Dinâmicos	Estáticos até ao refrescamento
Estrutura	Optimizada para transacções	Optimizada para consultas complexas
Frequência de acesso	Alta	Média a baixa
Tipo de acesso	Leitura/actualização/eliminação campo a campo	Leitura/agregação, grandes transacções
Utilização	Previsível e repetitiva	Ad hoc, não pre-definida, de heurística não estruturada
Tempo de resposta	Subsegundo (<1s) até 2-3 s	Alguns segundos a minutos ou mesmo mais

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.1. Porquê um DW? (resumo)

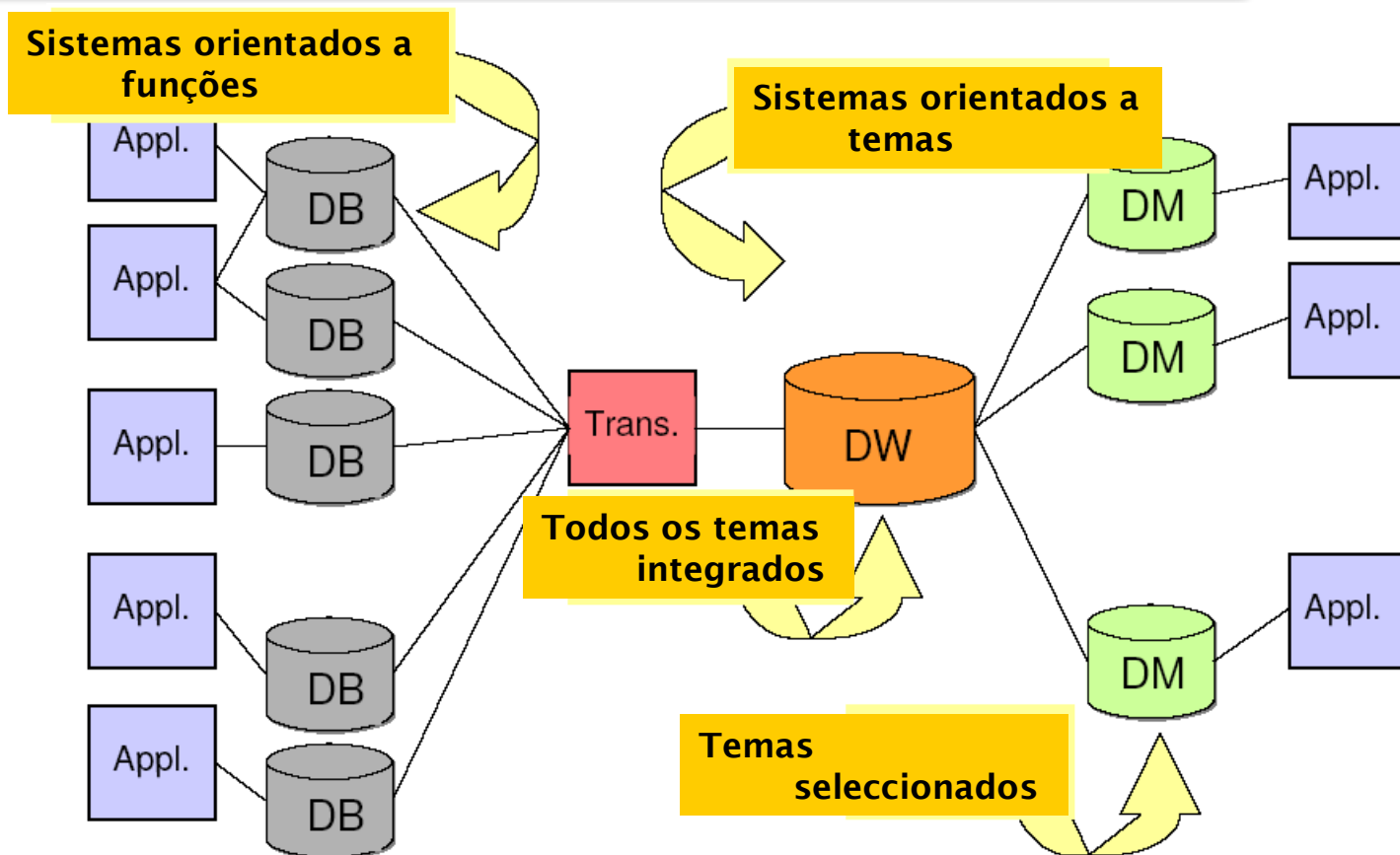
- As empresas possuem uma imensa quantidade de dados, dados operacionais e factos;
- Estes dados normalmente estão em uma diversidade de bases de dados e localizações físicas;
- Os dados estão disponíveis ou arquivados, mas em localizações diversas, heterogéneas e distribuídas;
- Os decisores necessitam de aceder a informação (dados sumarizados) virtualmente num local único;
- Este acesso necessita de ser rápido, independentemente do tamanho da base de dados e da respectiva “idade”.



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Orientação a funções vs. orientação a temas



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Orientação por temas

- **Nos sistemas operacionais:**

- os dados são orientados por aplicações: empréstimos, depósitos, cartões de crédito, etc. (no sector bancário); produção, compras, entregas, facturação, cobranças (num ambiente industrial);
- o foco da atenção são as transacções diárias características da operação normal do negócio;
- já vimos, os dados estão estruturados de forma a otimizar o processamento de transacções;
- os dados estão distribuídos por várias aplicações e, em cada uma, por muitas tabelas.

- **O sistema operacional trata dos processos do negócio.**

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Orientação por temas

- No DW:**

- Para conseguir uma visão de um assunto, ter-se-ia de aceder a muitas tabelas e de várias aplicações (se usados dados operacionais);

- Assim:

- os dados são orientados por assuntos / temas, tais como clientes, campanhas, vendedores, fornecedores, empregados, produtos;

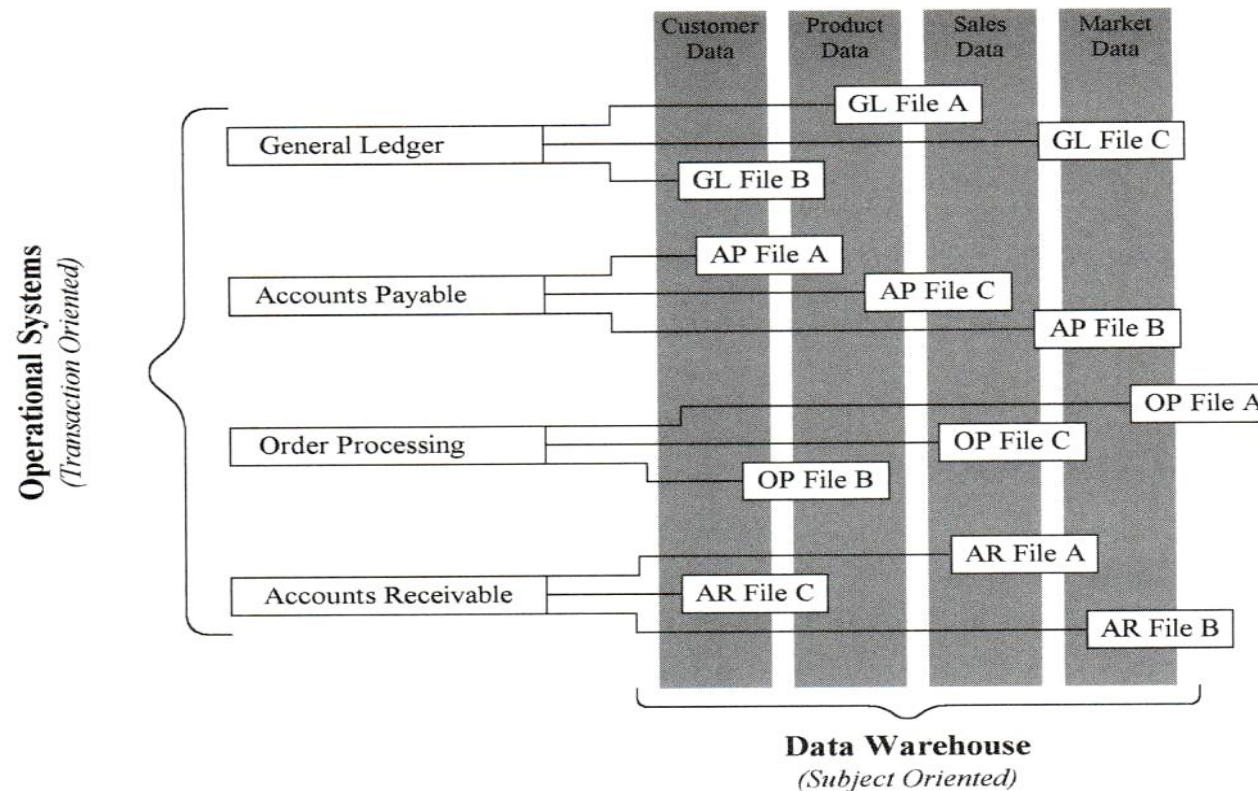
- lida-se com coisas que, no ambiente de negócio, influenciam as transações.

- O DW é focado nas entidades de alto nível do negócio.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Orientação por temas



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Integrada

“Um dos principais problemas para estabelecer um Warehouse é conseguir que os dados sejam consistentes... Embora a solução, segundo as regras, devesse consistir na modificação dos sistemas operacionais, para obter uma maior consistência, tentar limpar sistemas antigos é como “fazer ferver o oceano”. Se a este facto juntarmos o de que o Sistema Operacional não é a única fonte de dados, a responsabilidade de tornar os dados consistentes recai no D.W., sob o que se denomina de integração: incorporação dos dados em termos consistentes.”

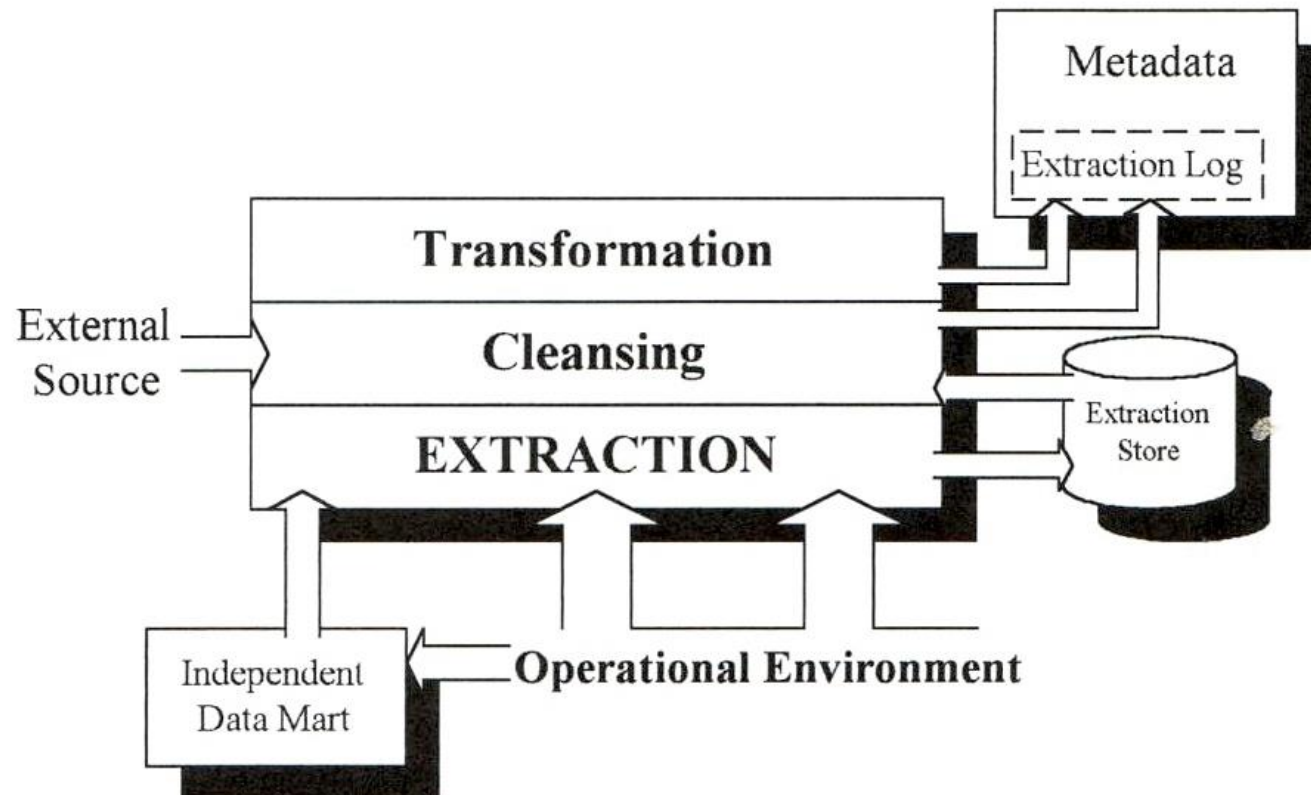
Enrique Peres del Campo

- A integração ocorre quando os dados migram do sistema operacional para o DW
- O problema resulta das decisões diversas de concepção (não existência de padronização) havidas na criação dos sistemas operacionais quanto à codificação, convenções de nomeação, atributos, medidas, etc.
- Os dados integrados no DW devem desfazer as muitas inconsistências entre os dados provindos das várias fontes

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

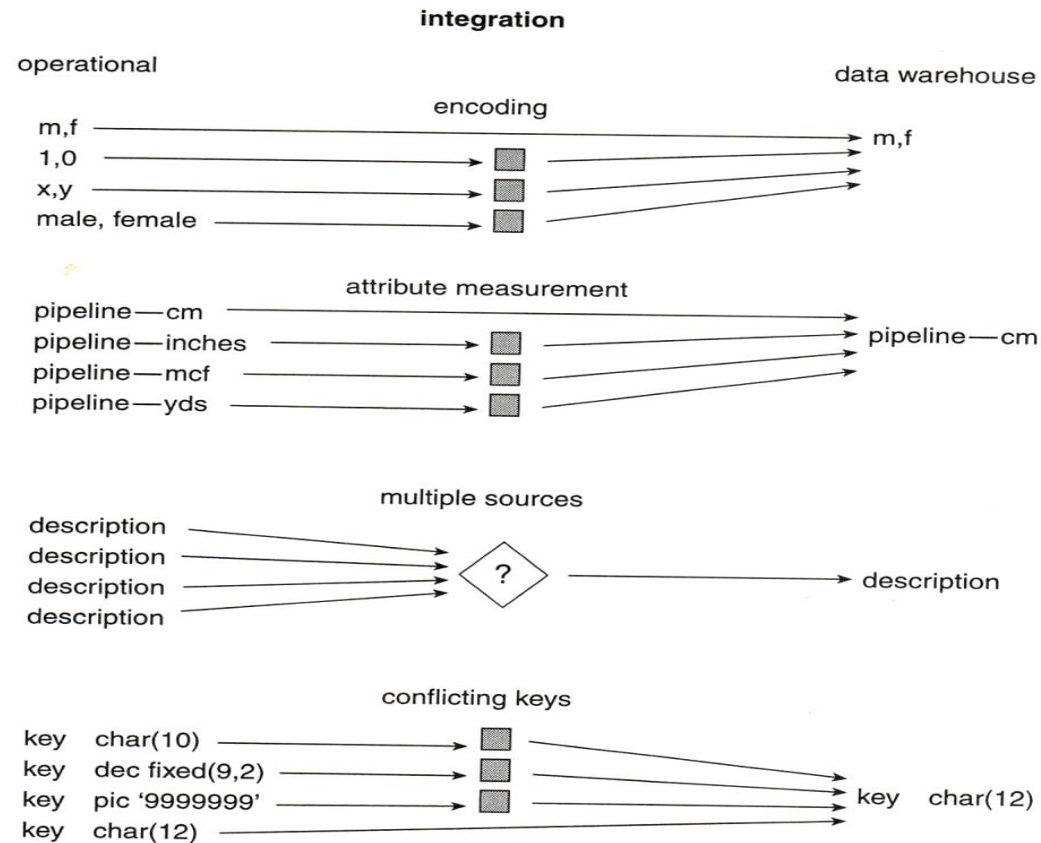
1.1.2.2. Análise detalhada da definição(2): Integrada



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Integrada

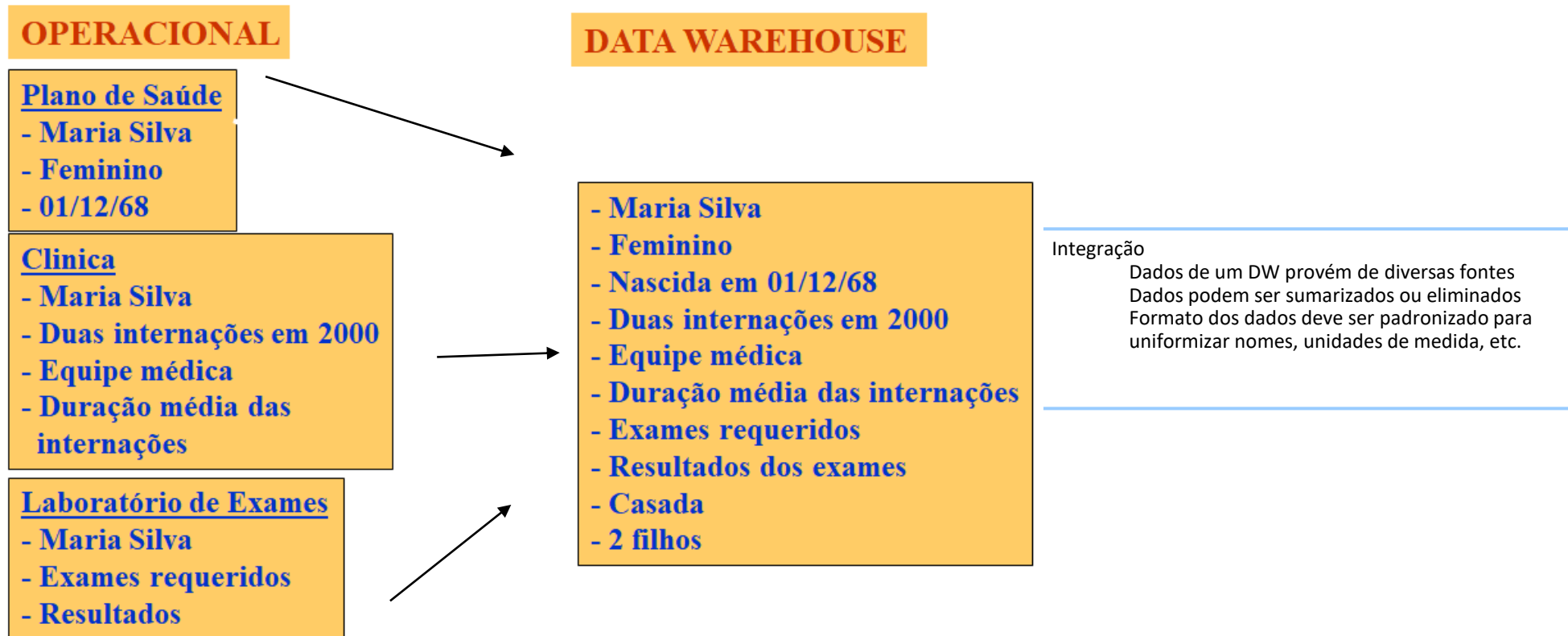


Retirado de: *Building the Data Warehouse*, Bill Inmon

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Integrada



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Não volátil

Os dados no DW são não voláteis:

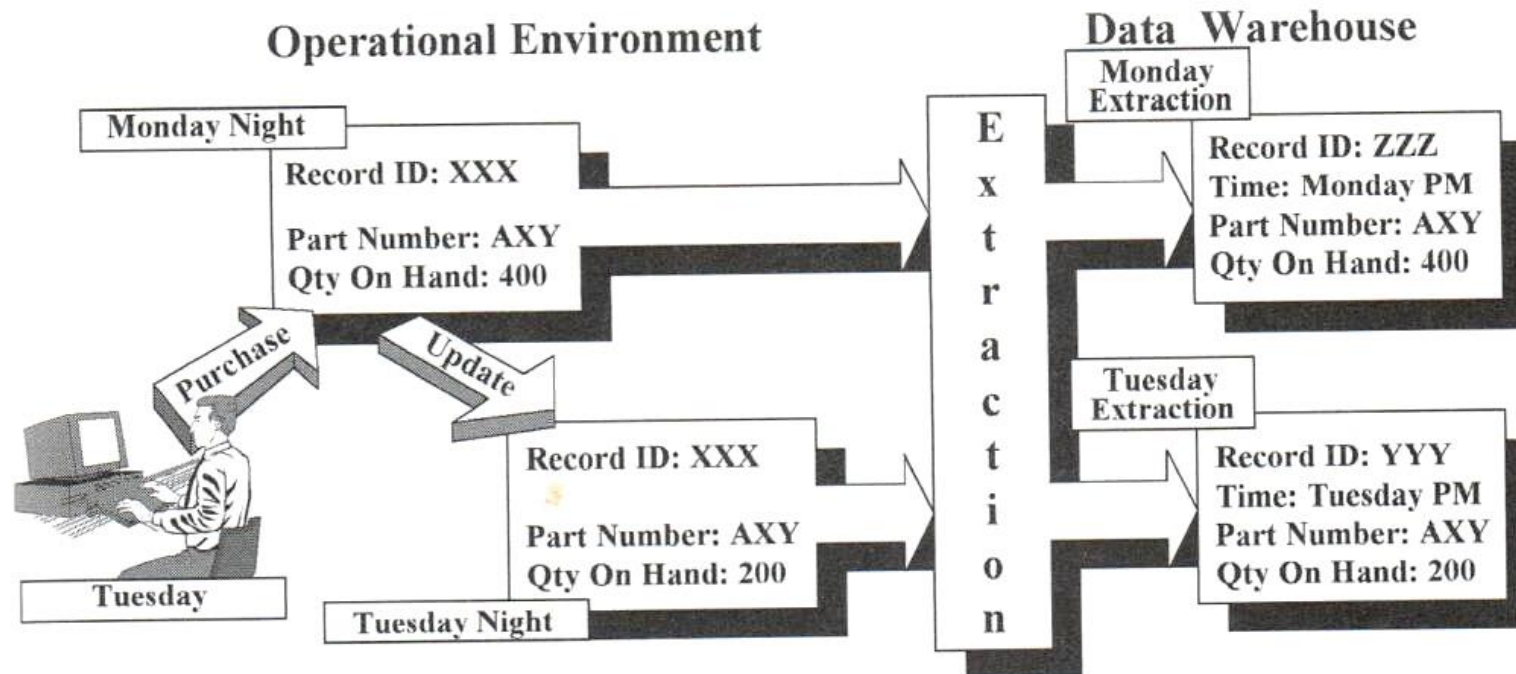
- depois de serem carregados no DW, as alterações são restringidas a situações onde sejam detectados problemas de qualidade dos dados ou na actualização de tabelas resultantes de sumarizações ou em situações de tratamento de dimensões em alteração lenta;
- a informação é assim acrescentada ou removida da base de dados, mas nunca alterada.

Desta forma, uma consulta efetuada no mês anterior, mostrará, hoje, resultados idênticos.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

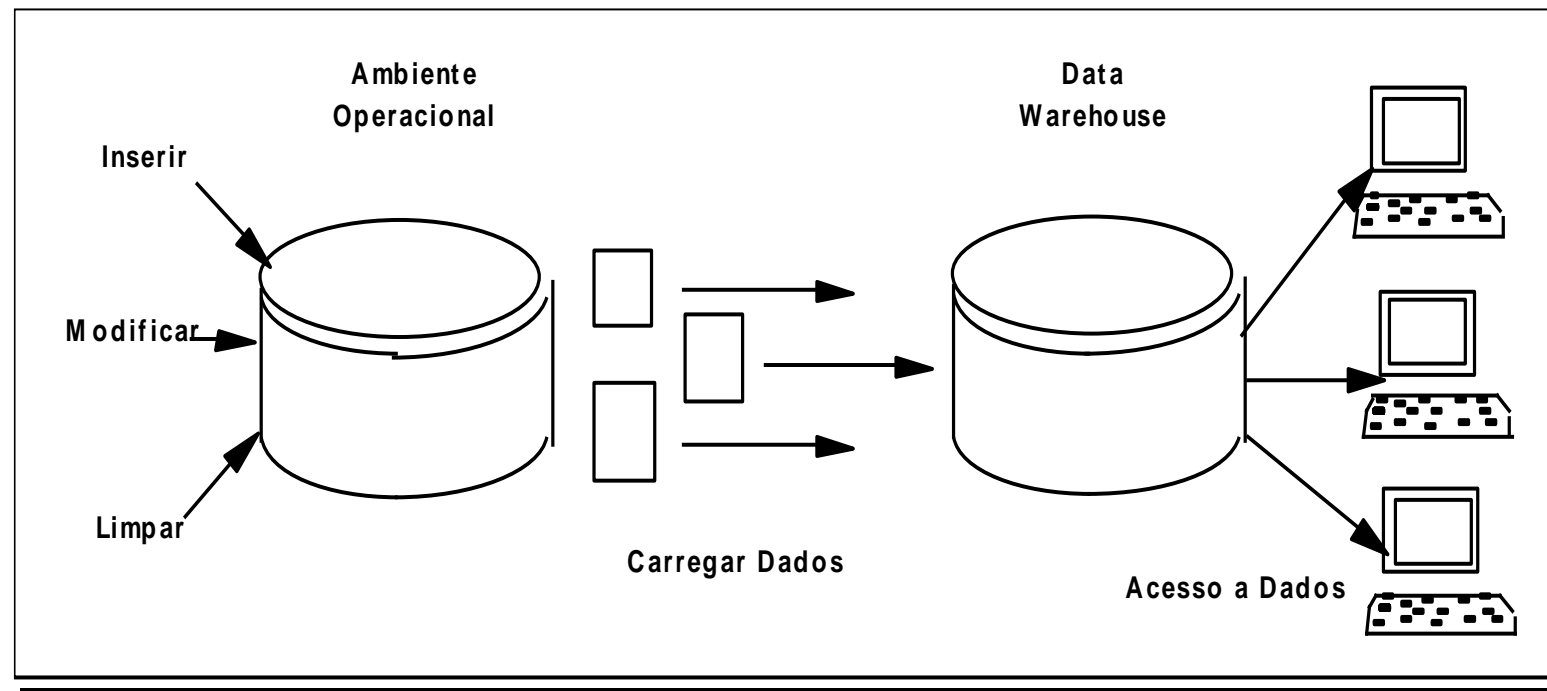
1.1.2.2. Análise detalhada da definição(2): Não volátil



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Não volátil



1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Variante no tempo

É consequência da não volatilidade:

- O data warehouse tira “fotografias” (snapshots) do estado do sistema operacional e outros ambientes externos;
- Essas fotografias são arquivadas no data warehouse como frames de um filme;
- O tempo não é uma variável numa frame, mas é estático;
- O tempo permite ver os dados através do tempo (tal como o filme mostra as frames numa sucessão temporal).

Desta forma é permitida a visualização de padrões ou tendências ao longo do tempo.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.2. Análise detalhada da definição(2): Para suporte do processo de tomada de decisões de gestão

O utilizador dum DW é o estratega de negócio: gestores de produto, marketing, gestores de departamento e CEOs.

Ao examinar o resultado de uma consulta percebem-se coisas que determinam nova consulta e assim sucessivamente:

- A produtividade é determinada pela rapidez da obtenção de respostas às sucessivas interrogações;
- A concepção do DW deve levar em conta esta forma típica de utilização dos dados de um DW;
- É crítico que o sistema mostre um desempenho apropriado e que a interface permita a exploração dos dados.

O desafio para o analista de negócio deve ser compreender os dados não como os obter.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.3. Objectivos (1)

O Data Warehouse deve:

- tornar facilmente acessível a informação da organização: o conteúdo do DW deve ser compreensível, sendo os dados intuitivos e óbvios para o utilizador, sendo os tempos de resposta curtos, utilizando ferramentas de acesso simples e fáceis de utilizar;
- apresentar a informação da organização de forma consistente: os dados do DW devem ser credíveis, assemblados de uma variedade de fontes, integrados, purificados, assegurando a qualidade e só disponibilizados quando adequados à forma de utilização dos analistas de negócio;
- ser adaptativo e capaz de suportar mudanças: a mudança é inevitável e está na ordem do dia. Mudam as necessidades do utilizador, as condições do negócio, dados e tecnologia. O DW deve ser concebido para suportar as mudanças inevitáveis. As mudanças não podem implicar a invalidação de dados ou aplicações existentes.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.3. Objectivos (2)

O Data Warehouse deve:

- ser um bastião seguro que protege os nossos recursos informacionais. A informação do DW é valiosa. Deve haver um controlo eficaz do acesso à informação confidencial da organização.
- servir como a base de tomada de melhores decisões. O DW deve conter os dados certos que permitam a toma de decisões. O objectivo último e verdadeiro do DW é o o seu valor para a tomada de decisões e o impacto no negócio.
- ser aceite pela sua utilidade efectiva. Se não há opção quanto à utilização dos sistemas operacionais, já o DW pode ser ou não utilizado. Se 6 meses após o início de funcionamento os utilizadores não continuarem a utilizá-lo activamente, é porque o teste de aceitação falhou.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.4. Benefícios

1. Tem uma orientação para assuntos;
2. Integra dados de fontes múltiplas diversificadas;
3. Permite análises de dados no tempo;
4. Permite relatórios e consultas ad hoc;
5. Proporciona capacidades analíticas aos decisores;
6. Liberta a sobrecarga de desenvolvimento no I/S;
7. Proporciona desempenho melhorado em consultas analíticas complexas;
8. Evita a sobrecarga nas bases de dados de processamento transaccional;
9. Converte os dados da empresa em informação estratégica;
10. Permite a reengenharia dos processos do negócio.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.4. Benefícios (1)

1. Tem uma orientação para assuntos:
 - Os dados estão ordenados por assunto e não por aplicação, tornando assim a “navegação” mais intuitiva. Esta situação está mais próxima, em termos conceptuais, da forma como os decisores pensam acerca do seu negócio.
2. Integra dados de fontes múltiplas diversificadas:
 - Uma das motivações principais que leva a implementação dum DW é que as fontes de dados são normalmente múltiplas e diversificadas.
3. Permite análises de dados no tempo:
 - As bases de dados operacionais possuem informação detalhada de horizontes temporais limitados (no máximo, ano corrente mais o anterior), mas as consultas analíticas necessitam de muito mais. Gráficos de tendências e procura de relações são possibilitados pelos snapshots tirados a intervalos definidos.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.4. Benefícios (2)

4. Permite relatórios e consultas ad hoc:

- Um DW é desenhado, tendo em mente a sua utilização, por parte dos utilizadores finais. Estão disponíveis um conjunto de ferramentas com extensiva utilização de objectos, que escondem completamente dos utilizadores a complexidade da base de dados (tabelas detalhe, sumarizações, índices), permitindo assim a maior liberdade aos utilizadores finais.

5. Proporciona capacidade analítica aos decisores:

- Um conjunto de ferramentas de análise OLAP estão disponíveis actualmente, ou utilizando directamente o DW, normalmente com um desenho físico especialmente adaptado e um middleware multidimensional, ou incorporando um motor multidimensional intrínseco. “Há tipicamente um número de diferentes dimensões a partir das quais um dado conjunto de dados pode ser analisado. Esta perspectiva plural ou Visão Conceptual Multidimensional traduz a forma como a maioria dos analistas de negócio vêem, de modo natural, a sua empresa.”[CoddOLAP]

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.4. Benefícios (3)

6. Liberta a sobrecarga de desenvolvimento de SI:

- Dadas as ferramentas disponibilizadas aos utilizadores finais e o seu carácter intrinsecamente intuitivo, estes obtêm por si os resultados pretendidos, sem necessidade de obrigar à intervenção do pessoal de SI e, na maioria das vezes, acedendo a informação que, de outra maneira, estaria completamente inacessível.

7. Proporciona desempenho melhorado em consultas analíticas complexas

- “Consultas agregadas são frequentes em aplicações de suporte à decisão, onde tabelas históricas muito extensas são "juntas" com outras tabelas e agregadas. Como as tabelas são grandes, uma melhor optimização de consultas agregadas tem o potencial de resultar em enormes ganhos de desempenho.”;
- Em termos de desenho do DW, for forma a conseguir-se um melhor desempenho, e já que a maioria das consultas analíticas colocadas são de natureza repetitiva e conhecida (por análise do perfil de utilização do DW), as agregações mais frequentemente utilizadas são pré-calculadas. O “segredo” do elevado desempenho dos produtos OLAP multidimensionais, reside precisamente neste pré-cálculo. Na sua versão relacional, assumem a forma de vistas pré-calculadas e materializadas.

1.1. Data warehouse como infraestrutura informacional de uma organização

1.1.2. Data Warehouse

1.1.2.4. Benefícios (4)

8. Evita a sobrecarga nas bases de dados de processamento transaccional:

- Tratava-se de uma situação duplamente lesiva: os utilizadores DSS achavam o sistema extremamente lento e não permitindo minimamente a execução de consultas em on-line, o que lesava extraordinariamente a produtividade dos analistas. Por outro lado, as filas de atendimento aos clientes, no balcão, crescia, apesar do contínuo melhoramento do hardware disponível. O que se passa é que a optimização necessária a uma das aproximações é lesiva da outra.

9. Converte os dados da empresa em informação estratégica:

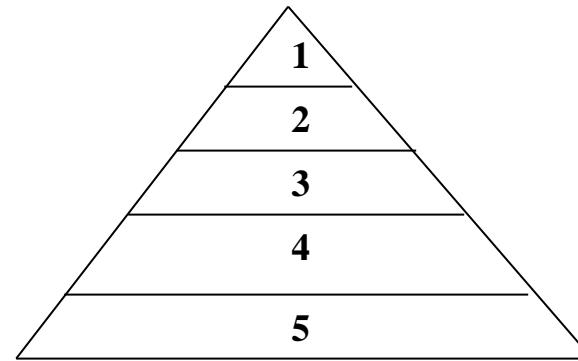
- “A informação é uma arma estratégica, no ambiente de negócio de mudanças rápidas dos dias de hoje. As empresas estão a compreender que a chave do sucesso para a competição e crescimento reside na sua capacidade de rapidamente obterem a informação correcta para avaliar tendências, prever mudanças nos mercados e analisar o desempenho.”

10. Favorece a reengenharia dos processos do negócio:

- Permite o acesso à informação subjacente aos processos do negócio, originando ímpetus baseados em factos geradores de oportunidades de reengenharia.

Data Warehousing, OLAP e Data-Mining

- Generalização de dados
 - Um processo que abstrai um grande conjunto de dados relevantes para tarefas em um banco de dados de níveis conceituais dos mais baixos para níveis mais altos



Conceptual levels

- Abordagens:
 - Abordagem de cubos (OLAP approach/[data warehouses](#))
 - Abordagem de Orientada a Atributos([AOI/data-mining](#))

Data-Mining

- **Consiste numa série de algoritmos que são aplicados a conjuntos de dados**
- **Os algoritmos geram modelos que realizam tarefas como Classificação de novos itens (desconhecidos)**
- **Agregação de itens de acordo com as suas características comuns Geração de regras de associação entre itens**
- **Previsão de valores futuros de um item Detecção de padrões sequenciais Detecção de valores atípicos**
- **(para detalhar em outras apresentações)**

QUESTÕES

Filipe Sá

Filipe.sa@isec.pt