

Arquitetura de *Data Warehouse* Aula 02 - Desenvolvimento de Sistemas OLAP

Ambiente de *Data Warehouse*

- Evolução natural do Ambiente de Apoio à Decisão
- Contém DM – *Data Marts*
- Conceito – Exige o estudo e envolvimento da empresa e seus executivos na definição da base de dados corporativa.
- A construção exige transferência e transformação dos dados existentes em sistemas corporativos utilizados nas operações diárias de controle para uma base de dados independente.
- Essa base de dados ficará disponibilizada para os usuários, mantida por meio de processo diferenciado dos existentes para os sistemas em operação transacional normais à empresa

Ambiente de *Data Warehouse*

- Justificativas para utilização desta tecnologia:
 - Várias plataformas de hardware e software
 - Constantes alterações nos sistemas transacionais corporativos.
 - Dificuldade acentuada de recuperação de dados históricos em períodos superiores ao ano atual de operações
 - Existência de sistema de “pacotes” de fornecedores diferentes
 - Falta de padronização e integração dos dados existentes nos diversos sistemas;
 - Carência de documentação e segurança no armazenamento dos dados
 - Dificuldade de aplicação de sistemas de sistemas DSS ou EIS devido à dependências múltiplas de sistemas corporativos.

Ambiente de *Data Warehouse*

- Características
 - Extração de dados de fontes heterogêneas
 - Transformação e integração dos dados antes de sua carga final.
 - Normalmente requer máquina e suporte próprio.
 - Visualização dos dados em níveis diferentes. Os dados do DW podem ou não ser extraídos para um nível mais específicos – os DMs, e a partir deste para um banco de dados individual
 - Utilização de ferramentas voltadas para acessos com diferentes níveis de apresentação
 - Dados somente são inseridos – não há *updates*.

Ambiente de *Data Warehouse*

- Outras Características

- Orientação por Assunto

- ✓ Significa que um DW armazena informações agrupadas por assunto de interesse – em contraste com os sistemas que são orientados a processos desenvolvidos para manter as transações realizadas diariamente.

- Ambiente Operacional

Pedido, Nota Fiscal



Ordem de Produção , máquina



Falha, Produto



Data Warehouse

Vendas



Produção



Qualidade



Ambiente de *Data Warehouse*

- Variação de Tempo
 - Os dados de um DW são precisos em relação ao tempo, representam resultados operacionais em determinado momento de tempo , o momento em que foram capturados.
 - Os dados de um DW são um “snapshot”, um conjunto estático de registros de uma ou mais tabelas, capturados em um momento de tempo predeterminado.
 - Isso implica que os dados de um DW não possam ser atualizados.

Ambiente de *Data Warehouse*

- Não Volátil
 - Um Data warehouse tem duas operações básicas: a carga dos dados – inicial e incremental e o acesso a estes dados – no modo leitura.
 - Depois de carregado, um DW somente possui operações de consulta e sem necessidade de nenhum tipo de bloqueio por concorrência de usuários no acesso.



Ambiente de *Data Warehouse*

- Integração
 - As convenções de nomes, valores de variáveis e outros atributos físicos de dados como data types são formalmente unificados e integrados nessa base única.

Arquitetura do Ambiente

- A arquitetura do ambiente de DW, inclui, além da estrutura de dados, mecanismos de comunicação, processamento e apresentação da informação para o usuário final.
- De uma forma geral, as arquiteturas orientadas a esse ambiente são constituídas por um conjunto de ferramentas que respondem desde a carga até o processamento de consultas, assim como por repositórios de dados como o Data Warehouse e os Data Marts.

Arquitetura do Ambiente

- As ferramentas podem ser divididas em dois grupos:
 - Ferramentas relacionadas à **carga inicial** e às atualizações periódicas do DW que são responsáveis pela extração dos dados de múltiplos sistemas, assim como limpeza, transformação e integração dos dados.
 - Ferramentas relacionadas às **consultas orientadas** para o usuário final, que são responsáveis pela elaboração de relatórios, pesquisas informativas, análise de desempenho e mineração de dados.

Arquitetura do Ambiente

- Papéis do usuário:
 - Analistas responsáveis pela cargas dos dados
 - ✓ Representam os programadores que necessitam conhecer o mapeamento entre o DW e os sistemas operativos, além de todos os requisitos para filtragem dos dados
 - Usuários finais
 - ✓ São os especialistas, gerentes, executivos e analistas de negócio, que utilizam a informação para apoio à tomada de decisão.

Arquitetura do Ambiente

- Papéis do usuário:
 - Analistas responsáveis pelo desenvolvimento e manutenção dos Data Warehouse e Data Marts
 - ✓ Equivalem aos DBAs. Estabelecem o nível de preocupação com os metadados, com a arquitetura de armazenamento e com a estrutura de dados, visando, principalmente, melhorar o desempenho das consultas.
 - Administradores de Dados
 - ✓ Tem um papel fundamental neste novo contexto.
 - ✓ Seu papel é de integrador entre os ambientes e garantia de qualidade e existência íntegra do data warehouse por meio de metodologias de acompanhamento e administração dos metadados entre os sistemas transacionais e os processos de extração de dados.

Processos e Ferramentas do DW

- Processos de um DW
 - Consistem na extração dos dados dos sistemas, organização e integração dos dados de forma consistente para o Data Warehouse e no acesso aos dados para consultas, integrados de forma simples, eficiente e flexível
- A extração, organização e integração dos dados devem ser realizadas de forma que a consistência e a integridade das informações sejam garantidas.

Processos e Ferramentas do DW

- O sucesso de um DW depende da disponibilidade da ferramenta certa para as necessidades dos usuários.
 - Ferramentas para pesquisas e relatórios – interface gráfica – analisam o que aconteceu.
 - Ferramentas do tipo OLAP – permitem ao usuário analisar o porque dos resultados obtidos
 - ✓ ROLAP – OLAP Relacional
 - ✓ MOLAP – OLAP Multidimensional
 - ✓ HOLAP – OLAP Híbrida
 - ✓ DOLAP – OLAP Desktop

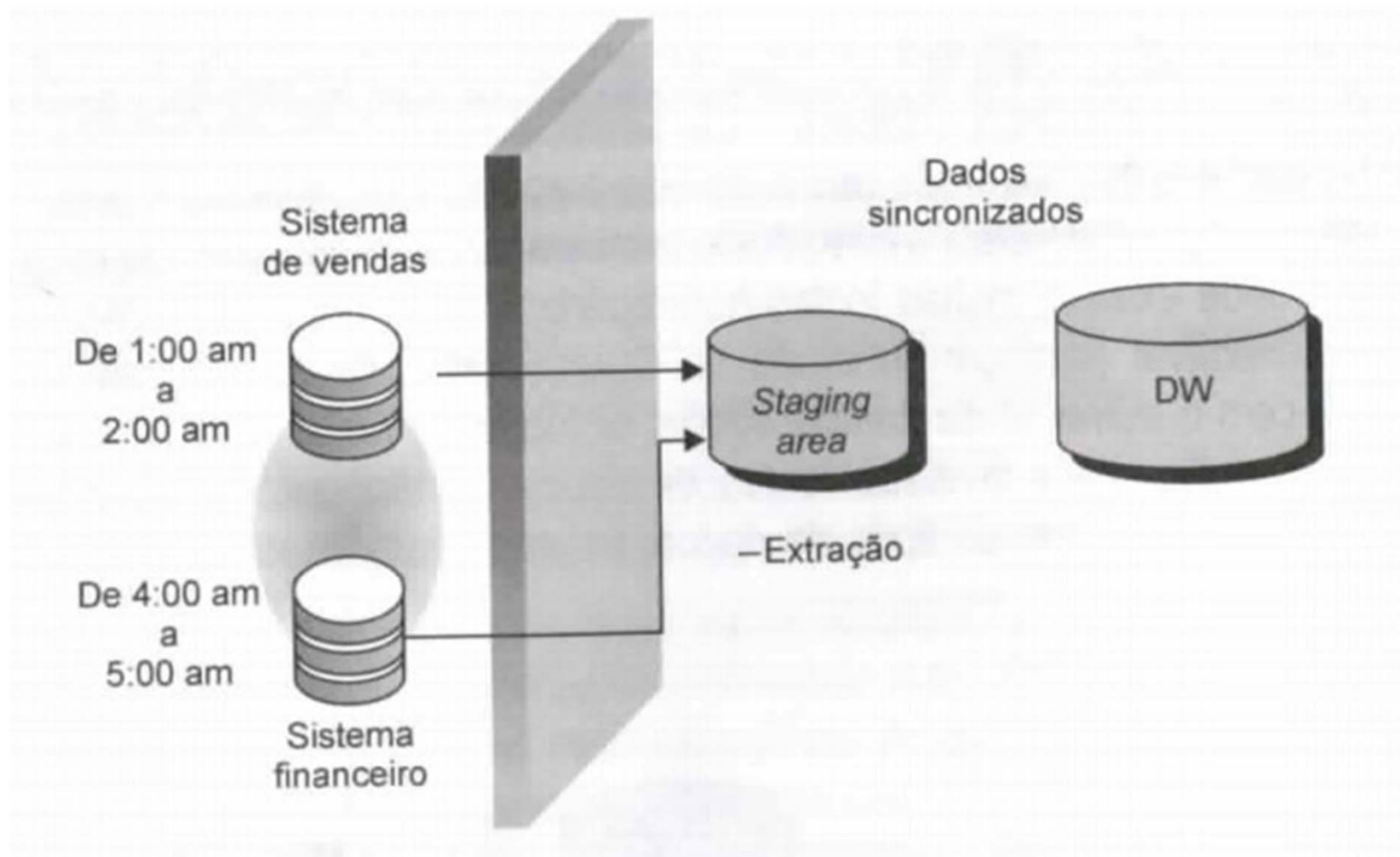
Processos e Ferramentas do DW

- Sistemas de Informações executivas – apresentam as informações de forma consolidada, em uma visualização mais simplificada.
- Data Mining – Uma categoria de ferramentas de análise denominada open-end – permite ao usuário avaliar tendências e padrões não conhecidos entre os dados. Utiliza-se de padrões modernos – redes neurais, algoritmos genéricos e lógica nebulosa.

Os Dados

- O DW pode apresentar os seguintes repositórios de dados:
 - ODS – Operacional Data Storage ou Staging Area
 - ✓ Representa um armazenamento intermediário dos dados, facilitando a integração dos dados do ambiente operativo antes da atualização do DW
 - ✓ Em sua proposta original, o ODS era um repositório temporário.
 - ✓ Atualmente, é chamado por alguns autores de DDS – Dynamic Data Storage
 - ✓ O DDS não é volátil, seus dados são armazenados ao longo do tempo e sofre alterações incrementais, desta forma com o decorrer do tempo, pode se tornar o DW.

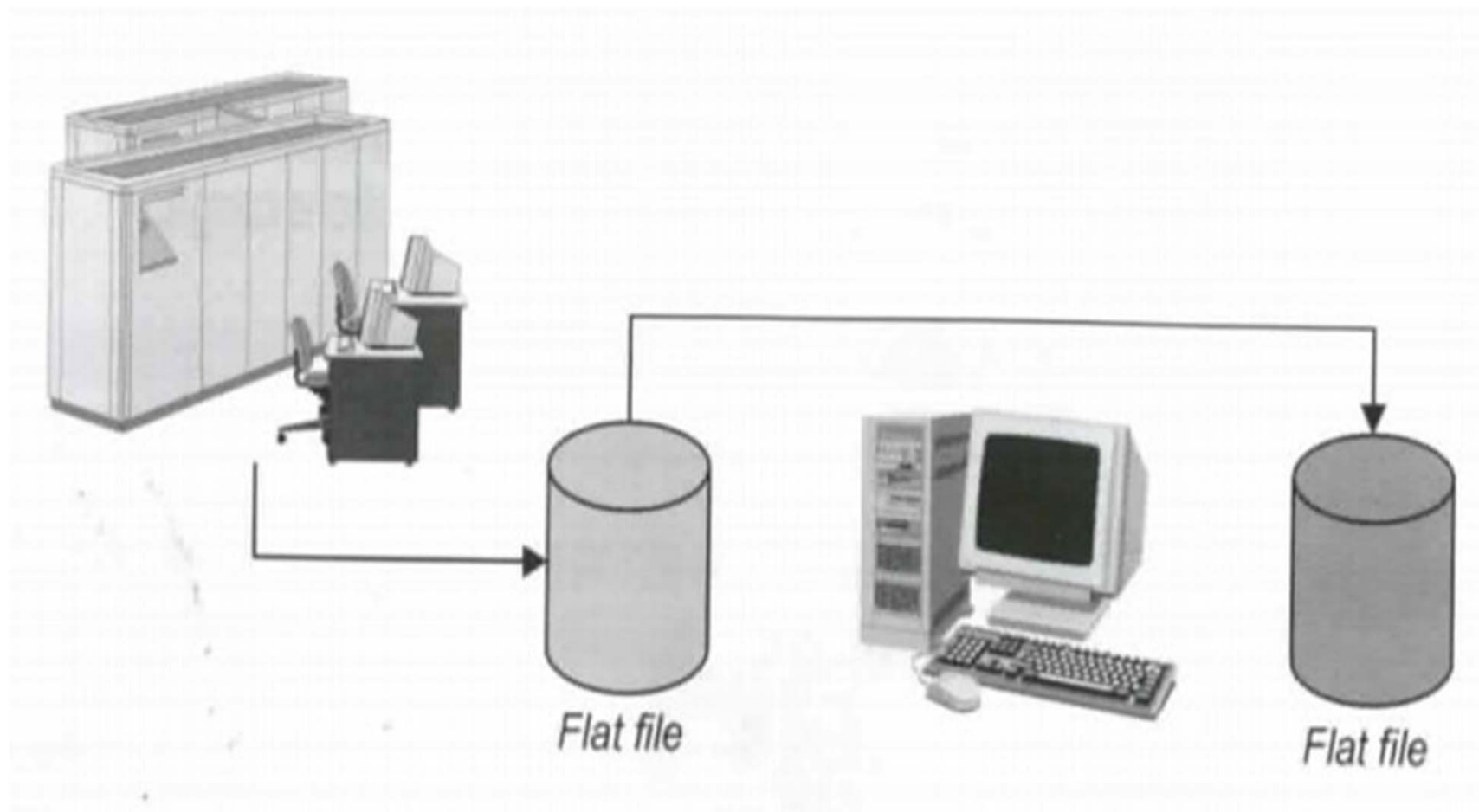
Os Dados



Os Dados

- ODS – *Operacional Data Storage* ou *Staging Area*
 - A utilização de *Staging Area* e o crescimento do DW, permite que este fique otimizado, pois garante a integração das informações, possibilitando uma fonte única e sincronizada de dados;
 - Sua função não é sumarizar dados, mas agilizar o processo de consolidação, proporcionando um melhor desempenho na fase de atualização dos dados.
 - A *Staging Area* é o único lugar para determinar os valores que vem efetivamente dos sistemas legados. Este deve ser o local para ser usado para “limpar” os dados “sujos”
 - Não é considerado um componente indispensável

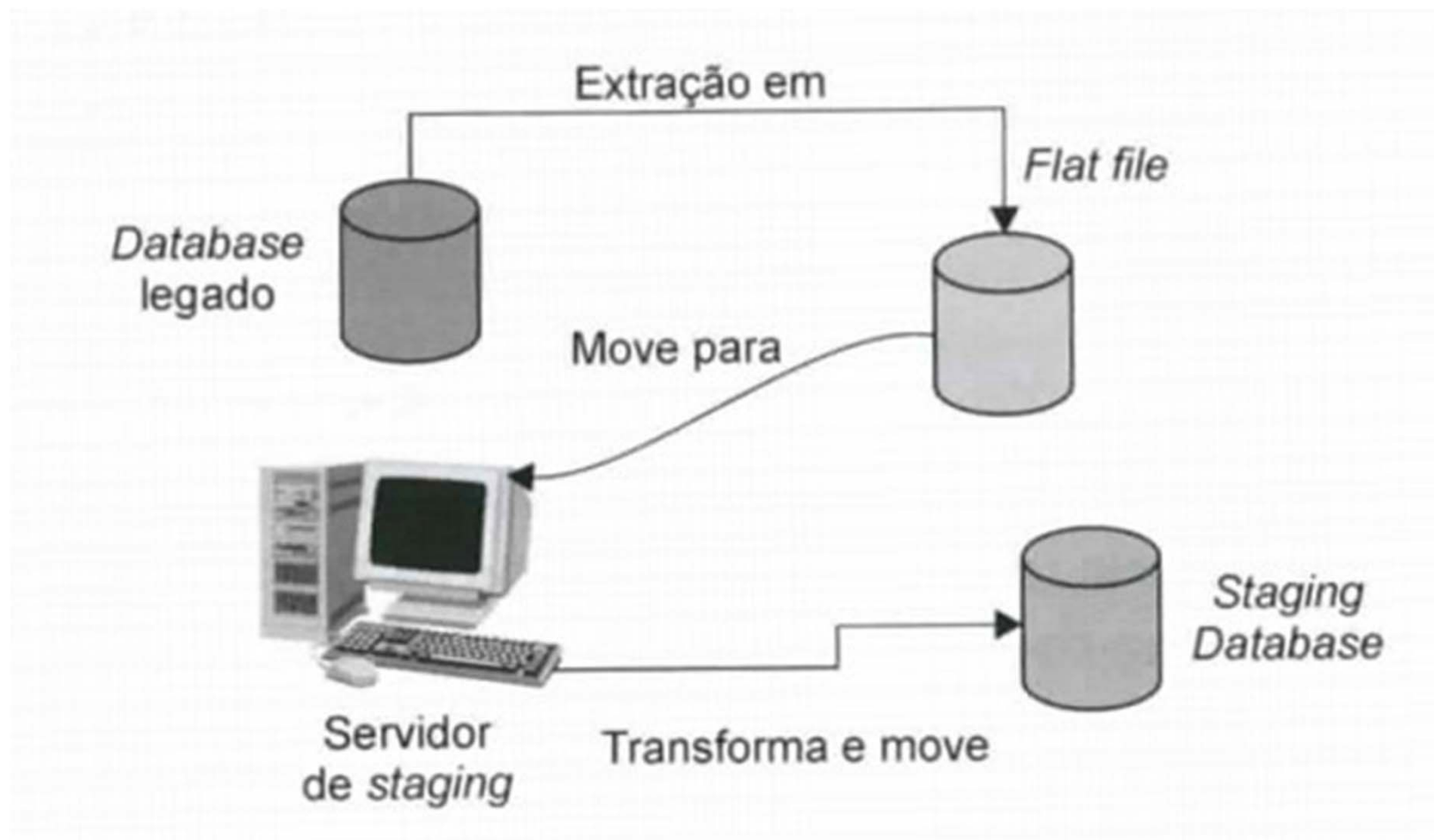
Os Dados



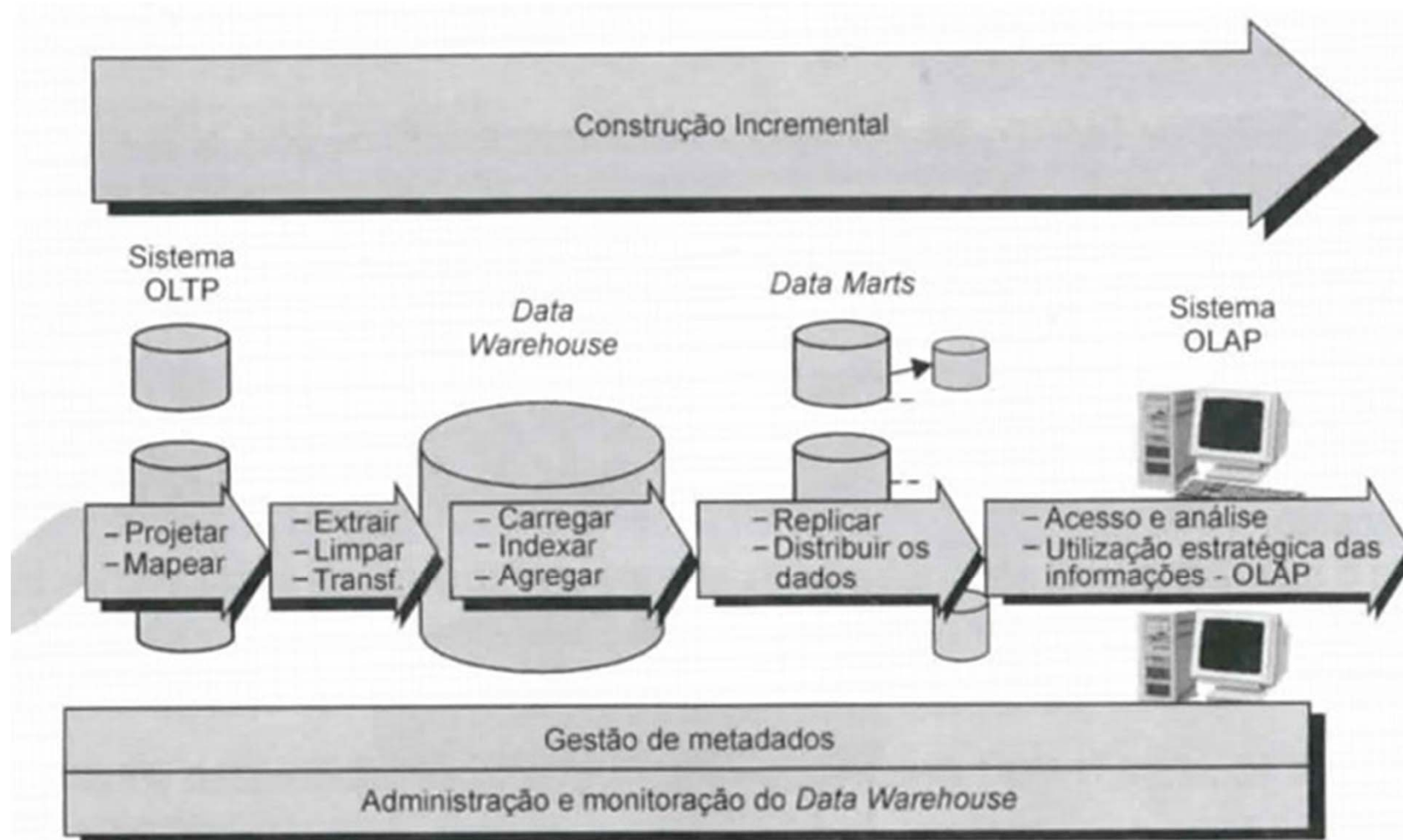
Os Dados

- A complexidade existente no projeto da Staging Area decorre - além da arquitetura- da escolha da forma de como os dados devem ser transferidos para ela:
 - Arquivos são gerados na origem como sequenciais e carregados na staging area como flat files para tratamento de transformação neste formato. Conjuntos de arquivos sequenciais serão utilizados no load do banco de dados do DW.
 - Carregar os flat files em um ODS, que sofrerá o processo de transformação dos dados e então fará a carga para o DW.
 - Utilização de backup e restore do banco de dados para geração de dados na staging area – neste caso os ambientes deverão ser idênticos.

Os Dados



Os Dados



Arquiteturas de DW

- Arquiteturas
 - Global
 - Independente
 - Integrada
- Abordagens de Implementação
 - Top Down
 - Bottom Up
 - Intermediária

Arquiteturas de DW

- Arquitetura Global
 - Considerada a que suporta toda ou a maior parte dos requerimentos ou necessidades de um Data Warehouse Integrado.
 - Projetado e construído com base nas necessidades de **toda** a empresa
 - Disponível para toda a empresa em toda a empresa
 - Pode ser fisicamente centralizado ou distribuído.

Arquiteturas de DW

- Arquitetura Independente
 - É a preferida dos fornecedores de software.
 - É isolada - permite que o trabalho de venda das soluções seja feito de forma isolada.
 - Permite que sejam criados Data Marts isolados na instalação da empresa sem conectividade uns com os outros.
 - A TI não controla a implementação e desenvolvimento.
 - A restrição da solução é que não há nenhuma integração entre os datamarts

Arquiteturas de DW

- Integrados
 - Neste caso, apesar dos data marts serem implementados de forma separadas, poderão ser compartilhados pelas áreas de negócio.
 - Esta arquitetura permite muitas funções e capacidades de informação em relação à arquitetura independente.
 - O nível de requisitos fica, no entanto, bastante complexo.
 - Neste caso, as ferramentas utilizadas para implementação deverão ser as mesmas.

Arquiteturas de DW

- Tipos de Implementações

Implementação TOP DOWN

- É conhecida como o padrão inicial do conceito de DW.
- Requer maior planejamento e trabalho antes de iniciar-se o projeto de DW propriamente dito.
- Tem como lado positivo forçar a empresa a definir regras de negócio de forma corporativa antes de iniciar o projeto de DW.

Arquiteturas de DW

- Vantagens
 - Herança de arquitetura: DW→DM
 - Visão de Empreendimento
 - Repositório de metadados centralizado e simples.
 - Controle e centralização de regras
- Desvantagens
 - Implementação muito longa
 - Alta taxa de risco
 - Heranças de Cruzamentos funcionais
 - Expectativas relacionadas ao ambiente

Arquiteturas de DW

- Implementação BOTTOM UP
 - Vem tornando-se muito popular
 - Permite que o planejamento e o desenho dos DataMarts possam ser realizados sem esperar que seja definida uma infraestrutura corporativa.
 - A infraestrutura corporativa poderá ser implementada incrementalmente, conforme forem produzidos os DMs
 - O grande problema desta implementação é que falta um gerenciador que garanta padrões únicos de metadados – mesmo com a independência dos DMs. (Padronização)

Arquiteturas de DW

- Vantagens
 - Implementação rápida
 - Retorno rápido
 - Manutenção no enfoque da equipe
 - Herança incremental
- Desvantagens
 - Perigo de legados gigantes (legadossauros)
 - Desafio de possuir a visão de empreendimento
 - Administrar e coordenar múltiplas equipes e iniciativas.
 - A maldição de sucesso

Arquiteturas de DW

- Implementação Combinada
 - Tem o propósito de integrar a arquitetura topdown e a bottom up.
 - Efetua-se a modelagem de dados do DW de visão macro, sendo o passo seguinte a implementação de partes deste modelo. Estas partes são escolhidas por processos.
 - Cada DM pode ser gerado a partir do macromodelo de dados do DW e integrado ao modelo físico do DW.
 - A principal vantagem desta abordagem é a garantia da consistência de dados, obtida em virtude do modelo de dados único.

Granularidade de Dados

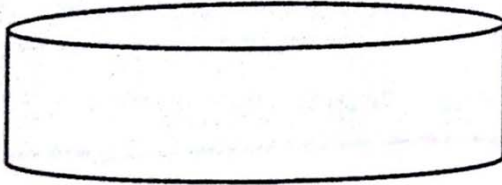
- Refere-se ao nível de sumarização dos elementos e de detalhe disponível nos dados
- Aspecto mais importante no projeto de DW
- Afeta profundamente o volume de dados que reside no DW e ao mesmo tempo o tipo de consulta que pode ser atendida.
- O volume de dados contido no DW é balanceado de acordo como o nível de detalhe de uma consulta desejada.

Granularidade de Dados

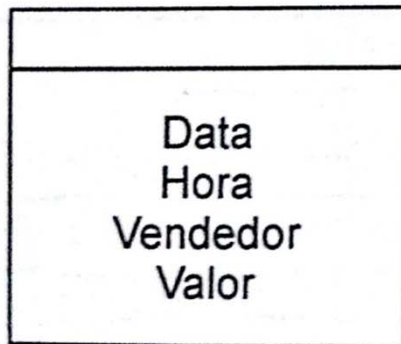
- Quanto mais detalhe temos nos dados, menor é a granularidade.
- Quanto menos nível de detalhe nos dados, maior é a granularidade.
- Quanto maior o volume de dados, menor a performance do DW.

Granularidade de Dados

Alto nível de detalhes
Baixo nível de granularidade

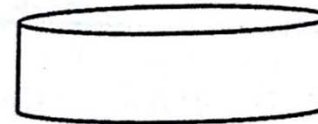


Exemplo
Detalhe de cada transação
de venda de um vendedor
realizada durante um mês

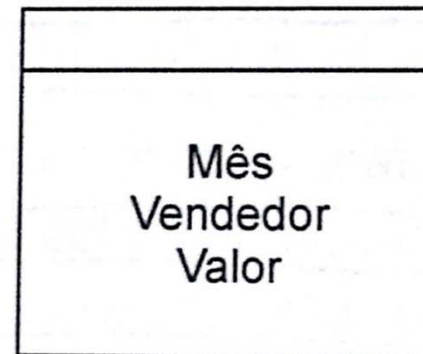


50 registros por mês

Baixo nível de detalhes
Alto nível de granularidade



Exemplo
Sumário das transações
de venda de um vendedor
realizadas durante um mês



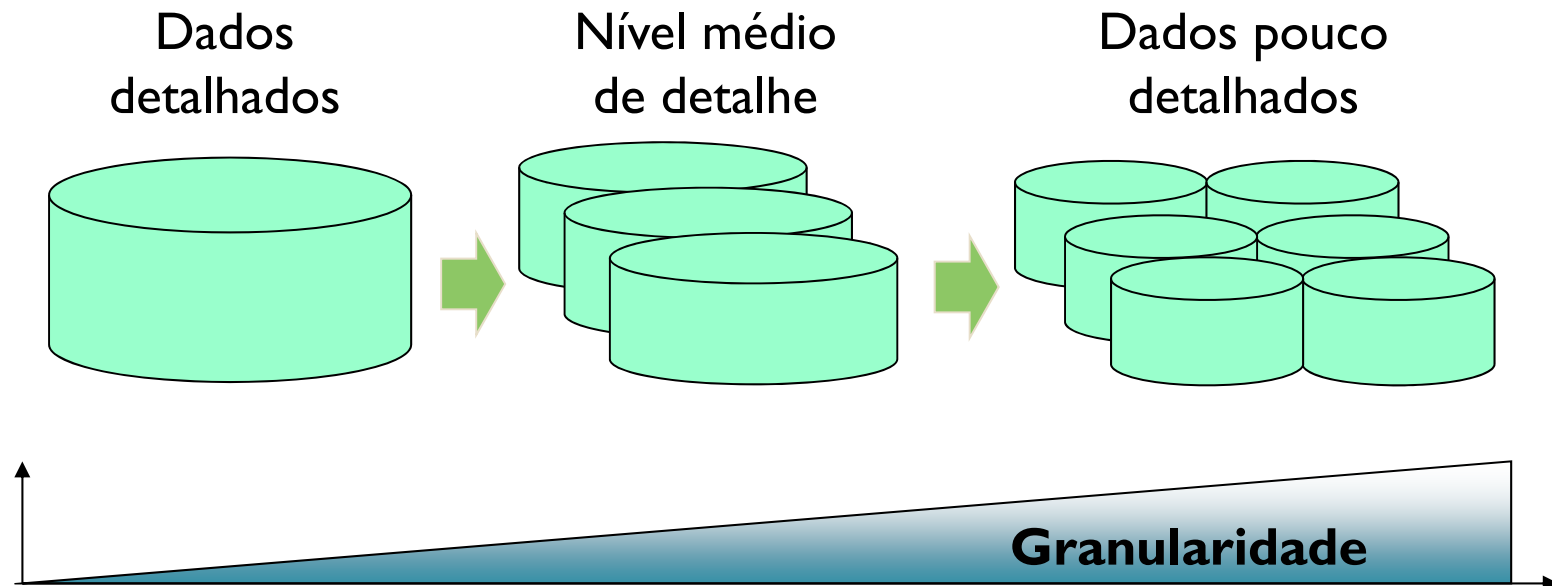
1 registro por mês

Granularidade de Dados

- Granularidade Baixa
 - É utilizada quando queremos manter o DW com base histórica de recuperação
 - À medida que o nível de granularidade se eleva há uma correspondente diminuição da possibilidade de utilização dos dados para atender a consultas
- Nível duplo de granularidade.

Granularidade de Dados

- Granularidade
 - Nível de detalhe dos dados
 - De extrema importância no projeto do DW



Granularidade de Dados

- Definir a granularidade adequada é vital para que o DW atenda seus objetivos
 - Mais detalhes → Mais dados → Análise mais longa → Informação mais detalhada
 - Menos detalhes → Menos dados → Análise mais curta → Informação menos detalhada
- Para evitar que se perca informação são criados vários níveis de granularidade

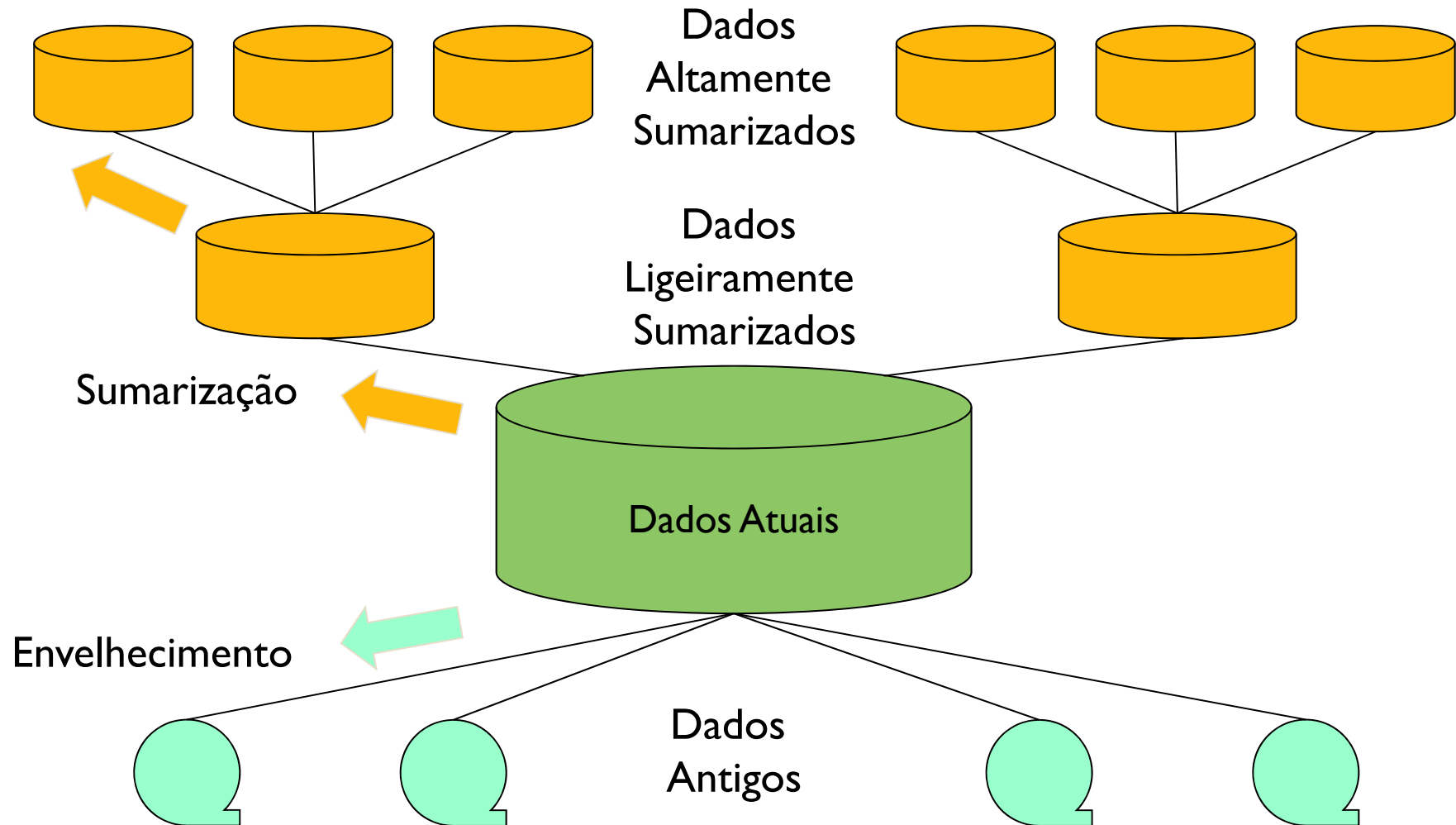
Granularidade de Dados

- Dados x Granularidade
 - Dados Atuais
 - ✓ Refletem acontecimentos recentes
 - ✓ Alto nível de detalhe (baixa granularidade)
 - Dados Sumarizados (1 ou + níveis)
 - ✓ Dados históricos condensados
 - ✓ Menor nível de detalhe (maior granularidade)
 - Dados Antigos
 - ✓ Dados históricos mantidos em fita, CD, etc
 - ✓ Alto nível de detalhe (baixa granularidade)

Granularidade de Dados

- Processo de sumarização
 - Aplica um novo esquema de modo a condensar os dados
 - Ex.: armazenar totais, médias etc.
- Processo de envelhecimento
 - Transfere os dados antigos do HD para fita, CD, etc.
 - Mantém o nível de detalhe para que nenhuma informação seja perdida

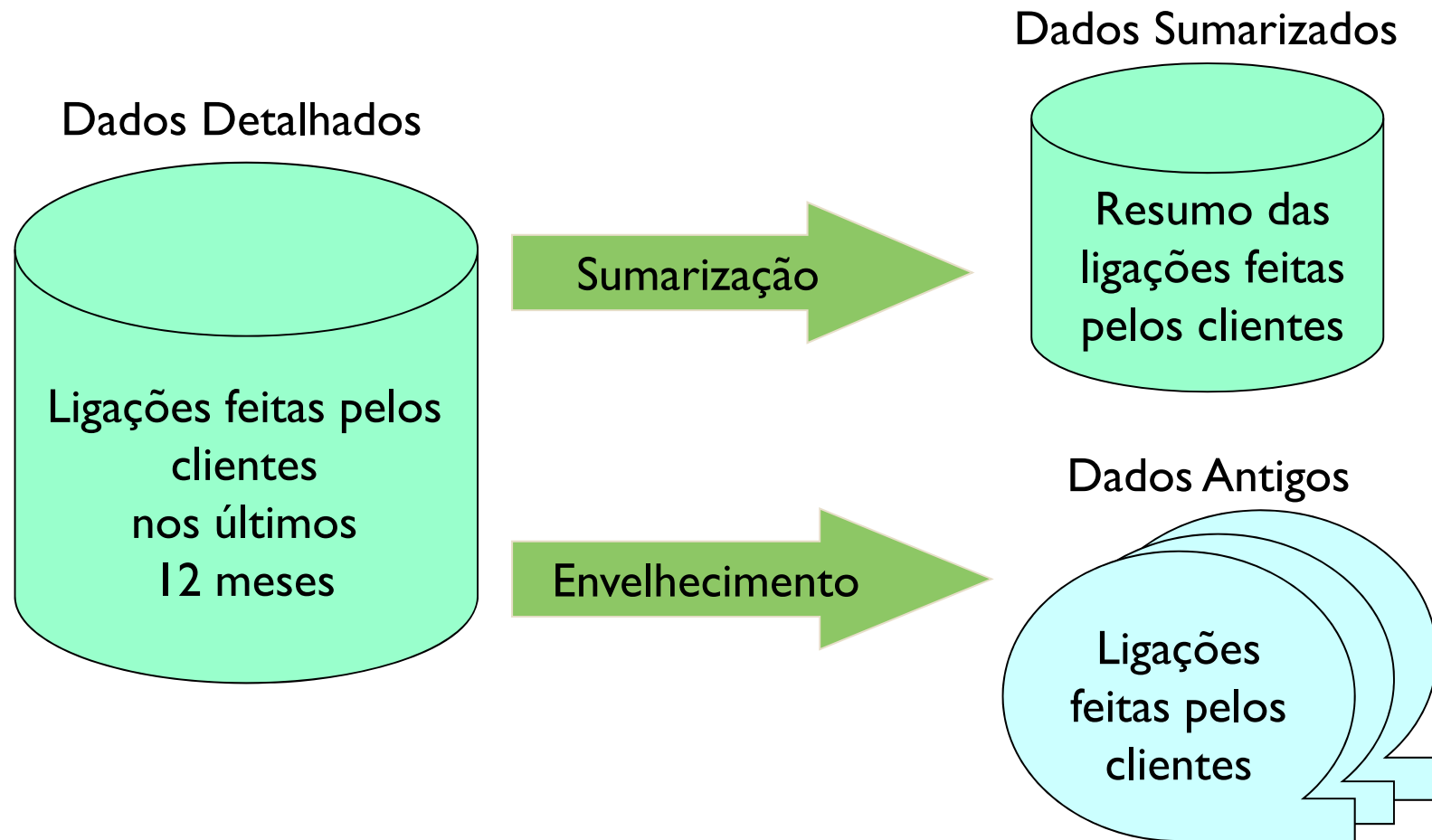
Granularidade de Dados



[Singh]

Granularidade de Dados

- Exemplo: Companhia Telefônica [Inmon]



Granularidade de Dados

- Exemplo: Companhia Telefônica (cont.)

Dados Detalhados

Ligações
Origem
Destino
Início
Fim
Tarifa
Status

Nº de registros:
ligações nos
últimos 12 meses

Dados Sumarizados

Ligações
Cliente
Mês
Pulsos
LongaDist
ValConta

Nº de registros:
contas emitidas
pela empresa

Dados Antigos

Ligações
Origem
Destino
Início
Fim
Tarifa
Status

Nº de registros:
ligações efetuadas
pela empresa

Granularidade de Dados

- Quanto menor a granularidade, mais detalhada é a informação disponível
 - No exemplo anterior, poderíamos determinar se o cliente A ligou para B na semana passada
 - Também poderíamos verificar se A faz muitas chamadas de longa distância
- Durante o processo de sumarização, algumas informações podem ser perdidas
 - Não seria possível saber se A ligou para B
 - É possível verificar o padrão de consumo de A

Modelagem de Dados para DW

- É completamente diferente daquela utilizada para sistemas OLTP
- Se movermos o modelo de dados transacional para um DW não funcionará porque ficará muito complexa.
- A modelagem de dados tradicional utiliza Modelagem Relacional

Modelagem de Dados para DW

OLTP X OLAP

- O OLTP (On-line Transaction Processing) captura as transações de negócios de um sistema e armazena no banco de dados.
- Esse tipo de base é utilizada em sistemas que registram pequenas transações (INSERT, UPDATE, DELETE) realizadas em tempo real e que ocorrem constantemente e de forma rápida.
- Os dados também pode ser alterados. Por não salvar histórico dos dados, isso não o qualifica como uma base de dados ideal para ajudar na tomada de decisões.
- Também é uma base que necessita de backup regularmente, pois caso o banco de dados seja perdido os dados não podem ser reaproveitados.

Modelagem de Dados para DW

- O principal objetivo da modelagem relacional em um sistema OLTP é eliminar ao máximo a redundância, de tal forma que uma transação que promova mudanças no estado do banco de dados, atue o mais pontualmente possível.
- Com isso, nas metodologias de projeto usuais, os dados são fragmentados por diversas tabelas (normalizados), o que traz uma considerável complexidade à formulação de uma consulta por um usuário final.
- Por isso, esta abordagem não parece ser a mais adequada para o projeto de um data warehouse, onde estruturas mais simples, com menor grau de normalização devem ser buscadas.

Modelagem de Dados para DW

OLAP

- O OLAP (On-line Analytical Processing) oferece uma alternativa diferente. Voltado para a tomada de decisões, proporciona uma visão dos dados orientado à análise, além de uma navegação rápida e flexível.
- O OLAP recebe dados do OLTP para que possa realizar as análises. Essa carga de dados acontece conforme a necessidade da empresa.
- Sendo um sistema para tomada de decisões, não realiza transações (INSERT, UPDATE, DELETE) pois sua finalidade são consultas.
- Possui dados atuais e históricos e não há necessidade de backups regularmente, sendo que ele possui informações do OLTP.
- Caso algo aconteça com a base OLAP basta fazer uma carga

novamente.

Modelagem de Dados para DW

Características	OLTP	OLAP
Operação típica	Atualização	Análise
Telas	Imutável	Definidas pelo Usuário
Nível de Dados	Atomizado	Altamente Sumarizado
Idade dos Dados	Presente	Histórico, Atual e Projetado
Recuperação	Poucos Registros	Muitos registros
Orientação	Registros	Arrays
Modelagem	Processo	Assunto

Obrigado!

...e agora suas perguntas?



ricardo.avila@outlook.com.br



@theavila