

Sistemas de Apoio à Decisão

Processo de Extração

-

Extração a partir de Tabelas

Extração a partir de Tabelas

- ◆ Cenário
- ◆ Estruturas de dados
- ◆ Alterações na package *pck_extract*
 - Procedimento *table_extract()*
 - Procedimento *main()*
- ◆ Testes

Extração a partir de Tabelas

◆ Cenário

- Extração dos dados das tabelas fonte da BD operacional
 - Tabelas T_*
- O acesso aos dados vai ser feito através de vistas, recorrendo a um *database link* público, *dblink_sadsb*
- Exemplo
 - Dados da tabela fonte T_PRODUTOS

*select * from view_produtos @dblink_sadsb;*

Extração a partir de Tabelas

◆ Estruturas de Dados

- Técnica *CDC*: *Timestamps*
 - Adicionada uma coluna nas tabelas fonte para registar a hora/data de alteração de cada registo
 - Para preencher a coluna são utilizados *triggers* que disparam automaticamente sempre que são inseridos ou atualizados regtos
- Técnica de extração incremental
- Estruturas de dados necessárias na *DSA*
 - Tabelas *T_DATA*
 - Tabela de controlo (*T_INFO_EXTRACTIONS*)

Extração a partir de Tabelas

◆ Estrutura de Dados de Controlo

- Tabela *T_INFO_EXTRactions*
- Permite controlar o processo de extração de dados a partir de tabelas de uma BD operacional
- Se a informação desta tabela se perder, ou for corrompida, pode comprometer todo o *DW*
- Constituída por duas colunas
 - *source_table_name*: nome da tabela fonte
 - *last_timestamp*: *timestamp* (data/hora) do registo mais recente extraído da tabela fonte

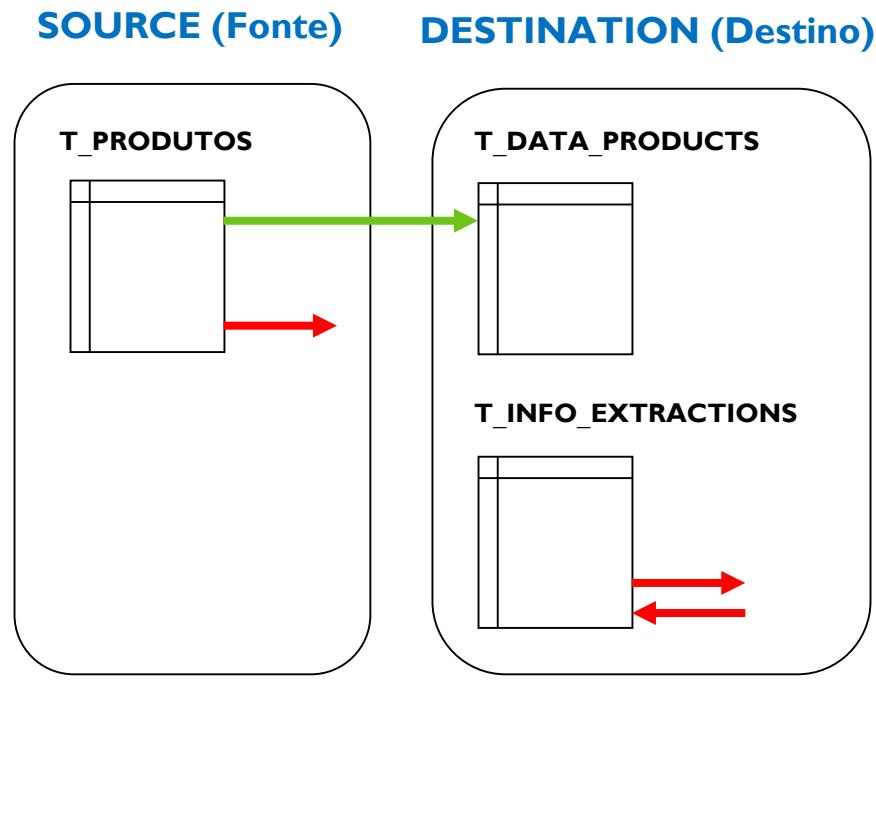
Extração a partir de Tabelas

◆ Alterações na package *pck_extract*

- Procedimento *table_extract()*
 - A extração a partir de tabelas é uma extração incremental
 - Duas situações distintas
 - 1^a extração: extrair todos os dados
 - 2^a extração e seguintes: extrair apenas os dados novos ou modificados desde a extração anterior
- Procedimento *main()*
 - Chamar o procedimento *table_extract()* para extrair os dados a partir das tabelas fonte

Extração a partir de Tabelas

◆ Esquema para a tabela dos Produtos



Extração a partir de Tabelas

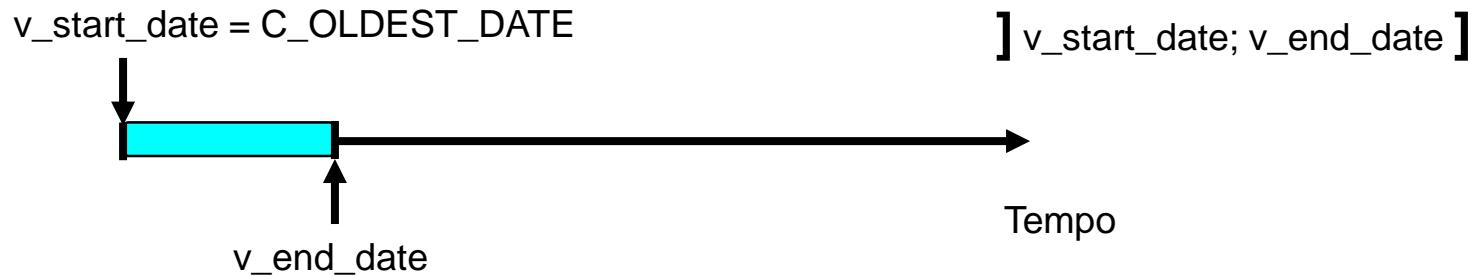
◆ Passos do Algoritmo

1. Limpar a tabela destino T_DATA
2. Obter a data do último registo extraído da tabela fonte a partir da $T_INFO_EXTRACTIONS$
3. Obter a data do registo mais recente na tabela fonte
 - **Se** há registos para extrair **Então**
 - 1^a Extração: **TODOS** os registos da tabela fonte
 - 2^a Extração e seguintes: **APENAS** os registos novos ou modificados
 - 4. Extrair os registos da tabela fonte
 - 5. Atualizar a tabela $T_INFO_EXTRACTIONS$

Extração a partir de Tabelas

◆ Procedimento *table_extract()*

- Variáveis *v_start_date* e *v_end_date*
- 1^a Extração: **TODOS** os registos



Extração a partir de Tabelas

◆ Procedimento *table_extract()*

- Variáveis *v_start_date* e *v_end_date*
- 2^a Extração e seguintes: **APENAS** os registos novos ou modificados



Extração a partir de Tabelas

◆ Testes

- Cenários para testar a extração a partir de tabelas
 - 1^a extração
 - 2^a extração e seguintes