

# 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 registrar a hora/data de alteração de cada registro
- Para preencher a coluna são utilizados *triggers* que disparam automaticamente sempre que são inseridos ou atualizados registros

### ■ 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
  - 1ª extração: extrair todos os dados
  - 2ª extração e seguintes: extrair apenas os dados novos ou modificados desde a extração anterior

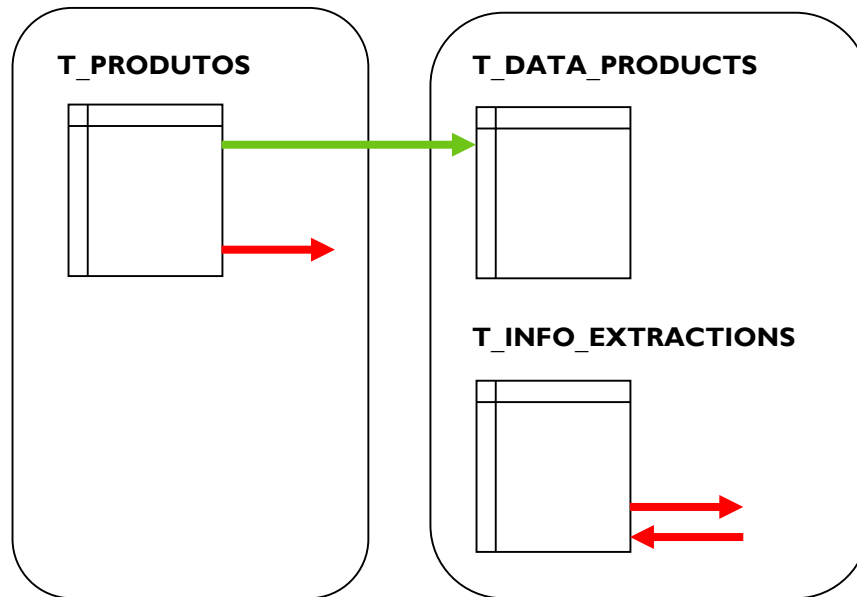
### ■ Procedimento *main( )*

- 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

**SOURCE (Fonte)**      **DESTINATION (Destino)**



— dados  
— *timestamp*

# Extração a partir de Tabelas

---

## ◆ Passos do Algoritmo

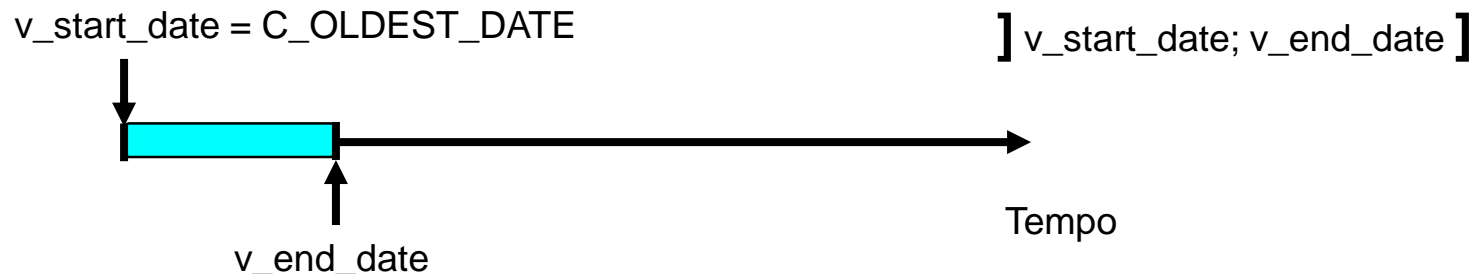
1. Limpar a tabela destino  $T\_DATA$
2. Obter a data do último registo extraído da tabela fonte a partir da  $T\_INFO\_EXTRACTIONS$
3. Obter a data do registo mais recente na tabela fonte
  - Se há registos para extrair Então
    - 1ª Extração: **TODOS** os registos da tabela fonte
    - 2ª Extração 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*
- 1ª Extração: **TODOS** os registos



# Extração a partir de Tabelas

## ◆ Procedimento *table\_extract*( )

- Variáveis *v\_start\_date* e *v\_end\_date*
- 2ª Extração e seguintes: **APENAS** os registos novos ou modificados



# Extração a partir de Tabelas

---

## ◆ Testes

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