

Sistemas de Apoio à Decisão

Carregamentos de manutenção num DW

-

Change Data Capture

Change Data Capture

- ◆ Cenário
- ◆ O que é?
- ◆ Como funciona?
 - *Timestamps*
 - *Partições*
 - *Triggers*
- ◆ Extracção

Change Data Capture: cenário

- ◆ O DW já teve um primeiro carregamento de dados, pretendendo gerir-se os carregamentos posteriores
- ◆ Pretende-se que a fase de extracção use apenas os novos dados fonte ou aqueles que tenham sofrido alterações
- ◆ As fontes de dados têm obrigatoriamente que ser bases de dados operacionais relacionais

Change Data Capture: o que é?

- ◆ Nome técnico genérico que representa os vários métodos para detectar quais os novos dados ou os dados alterados nos sistemas fonte
- ◆ Alguns sistemas proprietários (ex: Oracle) têm ferramentas integradas de raiz para fazer esta detecção
- ◆ Em SAD: construção «manual» de um sistema de CDC

Change Data Capture: como funciona?

- ◆ Existem 3 métodos de detectar quais os dados fonte que foram alterados ou adicionados
 - *Timestamps*
 - Alguns SGBD permitem que as suas tabelas tenham uma coluna na qual é registada qual a hora/data da alteração de cada registo
 - Partições
 - As tabelas de dados são divididas em partições e cada partição representa um espaço temporal (por exemplo, 1 partição=1 dia)
 - *Triggers*
 - Permitem simular o mecanismo por *timestamps*.

Change Data Capture por triggers

◆ Quando usar

- Na impossibilidade de aplicar um dos outros dois métodos

◆ Objectivo

- Simular o método por *timestamps*
- Quando um registo na fonte de dados for alterado/inserido, marcar esse registo com a data/hora do sistema

◆ Construção

1. Adicionar às tabelas **do sistema fonte** (por DDL) uma coluna que permita guardar hora/data (DATE)
2. Criar *triggers* **no sistema fonte** que sejam «disparados» por cada INSERT ou UPDATE nessas tabelas
3. Após cada extracção a partir de uma fonte, guardar a hora/data do último registo extraído dessa fonte

Change Data Capture por triggers

◆ Exemplo

SISTEMA FONTE

t_produtos

<u>Cod</u>	Preço	Nome	Marca	Peso	<i>Last_changed</i>
ARR1	1.30	Arroz Carolino	Cigala	1kg	29-10-05 12:10
FAR1	0.65	Farinha T55	Nacional	1kg	30-10-05 06:19
ARR2	1.25	Arroz Carolino	Saludões	1kg	10-03-2009 14:00

1º
O preço de um
produto é
alterado



2º
O *trigger* de
UPDATES é
disparado



3º
O *timestamp*
do registo é
actualizado

TRIGGER

Sadsedfdfsdf
Sadsfdfs
Ddsfd

:NEW.timestamp:=SYSDATE;

sdfdsf
jdjdjd

Change Data Capture: extracção

◆ Extracção

- Aquando de uma extracção da tabela t para a *Data Staging Area*, são extraídos apenas os registos r em que

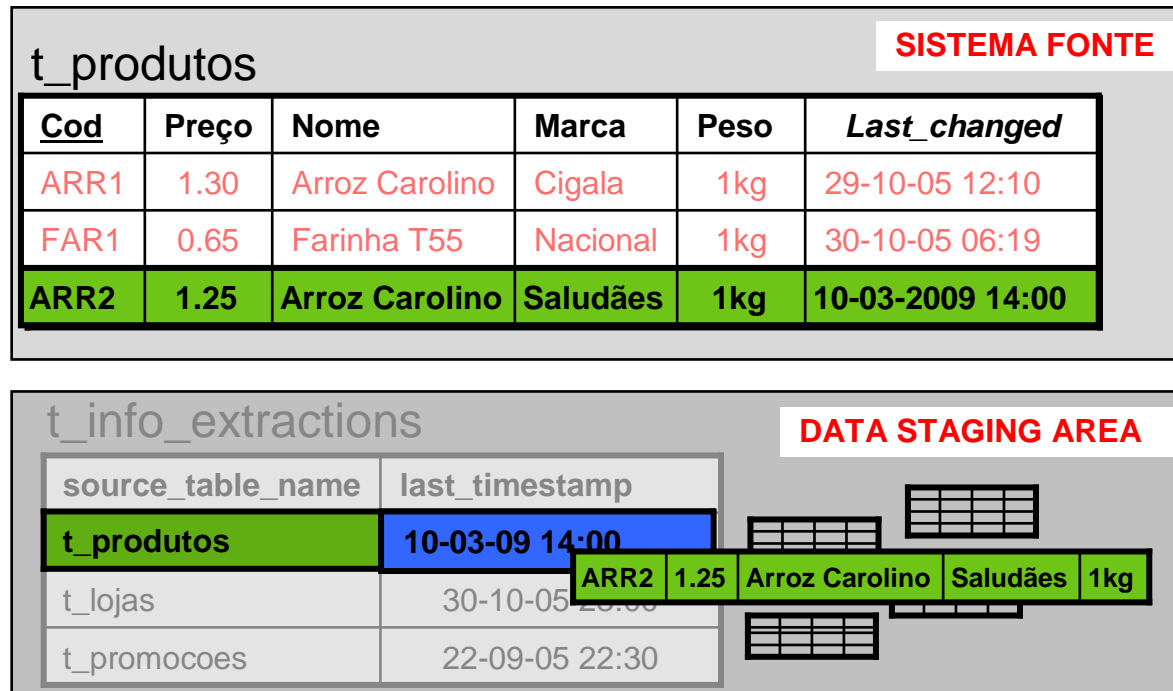
$Tabela(t).registro(r).last_changed > \text{MAX}(tabela(t).last_changed)$
dos registos da última extracção

◆ Implicações

- Após cada extracção tem que guardar-se o maior valor *last_changed* de cada tabela interveniente (para futuras extracções)

Change Data Capture: extracção

◆ Exemplo



1º
A extracção é iniciada para **t_produtos**



2º
Seleccção dos registos



3º
Envio dos registos para a DSA



4º
Actualização da tabela **t_info_extractions**