

# Sistemas de Apoio à Decisão

---

## Processo de Extração

-

### Extração a partir de Ficheiros

# Extração a partir de Ficheiros

---

- ◆ Cenário
- ◆ Estruturas de dados
- ◆ Alterações na package *pck\_extract*
  - Procedimento *file\_extract( )*
  - Procedimento *main( )*
- ◆ Testes

# Extração a partir de Ficheiros

---

## ◆ Cenário

- Extração dos dados do ficheiro Excel “Lojas.xls” que contém duas folhas
  - LOJAS (ficheiro “*stores.csv*”)
  - DIRETORES (ficheiro “*managers.csv*”)
- Os ficheiros “\*.csv” encontram-se numa diretoria no servidor Oracle
- Estão disponíveis cópias dos ficheiros na pasta de instalação da plataforma ETL^2
- A extração será feita a partir dos ficheiros de texto do servidor

# Extração a partir de Ficheiros

---

## ◆ Estruturas de Dados

- Técnica CDC: Processo de eliminação
  - Todos os dados fonte são carregados
  - Os dados da extração anterior são guardados na DSA
  - Os dados são comparados e as diferenças passam para o processo de transformação
- Técnica de extração completa
- Estruturas de dados necessárias para cada ficheiro
  - Tabela externa (T\_EXT\_\*)
  - Tabela NEW (T\_DATA\_\*\_NEW)
  - Tabela OLD (T\_DATA\_\*\_OLD)

# Extração a partir de Ficheiros

---

## ◆ Estrutura de Dados *External Table*

- Utilizada para ler os dados de ficheiros externos para a BD e vice-versa
- Uma tabela externa não é mais do que uma vista para um ficheiro
- Constituída por duas partes
  - 1<sup>a</sup> Parte: idêntica a um *CREATE TABLE* normal
  - 2<sup>a</sup> Parte: específica das tabelas externas, que permite dizer ao Oracle (*SQL Loader*) como interpretar os dados do ficheiro

# Extração a partir de Ficheiros

---

## ◆ Estrutura de Dados *External Table*

- Passos para criar uma Tabela Externa
  - Criar o objeto: cria os metadados e diz ao Oracle como aceder aos dados
  - Extrair os dados: o objeto é preenchido com dados
- Comando para extrair os dados do ficheiro para o objeto Tabela Externa

```
select * from T_EXT_ *;
```
- Se algo não correr bem devem ser consultados os ficheiros de log, diretoria *src\_files*, no servidor Oracle

# Extração a partir de Ficheiros

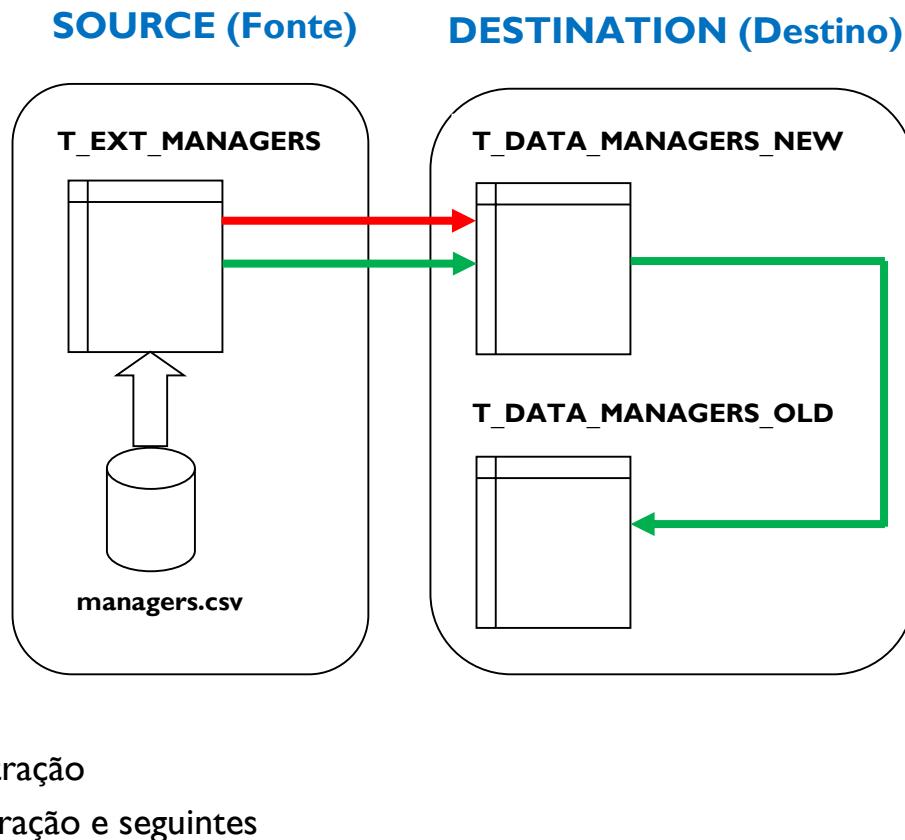
---

## ◆ Alterações na package *pck\_extract*

- Procedimento *file\_extract( )*
  - A extração a partir de ficheiros é sempre uma extração total, ou seja, todos os dados do ficheiro são extraídos
  - Duas situações distintas
    - 1<sup>a</sup> extração (tabelas destino vazias)
    - 2<sup>a</sup> extração e seguintes
- Procedimento *main( )*
  - Chamar o procedimento *file\_extract( )* para extrair os dados a partir dos ficheiros
    - “*stores.csv*”
    - “*managers.csv*”

# Extração a partir de Ficheiros

## ◆ Esquema para o ficheiro “managers.csv”



# Extração a partir de Ficheiros

---

## ◆ Passos do Algoritmo

- Técnica de extração completa
  1. Apagar o que está na tabela *OLD*
  2. Copiar o que está na tabela *NEW* para a tabela *OLD*
  3. Apagar o que está na tabela *NEW*
  4. Copiar o que está na tabela externa para a tabela *NEW*
- Implementação dos passos do algoritmo com código PL/SQL dinâmico

# Extração a partir de Ficheiros

---

## ◆ Testes

- Cenários para testar a extração a partir de ficheiros
  - 1<sup>a</sup> extração
  - 2<sup>a</sup> extração e seguintes