

**POLITÉCNICO
DE LEIRIA**

SISTEMAS DE APOIO À DECISÃO

Curso de Engenharia Informática, Regimes Diurno e Pós-Laboral

3º Ano, 1º Semestre, Ano Letivo 2023/2024

Prova Escrita I – Avaliação Periódica

Data: 16 de novembro de 2023

Duração: 01h00

Sem Consulta

Nas perguntas de escolha múltipla as respostas erradas têm uma penalização de $1/N$, em que N é o número de opções de resposta erradas da pergunta, mas a cotação na pergunta nunca será inferior a zero.

1. [0,75 valores]

Das seguintes opções, escolher a(s) que corresponde(m) a etapa(s) do processo de tomada de decisão:

- ☐ Comparação de alternativas
- ☐ Execução e avaliação
- ☐ Análise e identificação da situação
- ☐ Classificação dos riscos de cada alternativa
- ☐ Escolher a melhor alternativa
- ☐ Desenvolvimento de alternativas

2. [0,75 valores]

Faça a ligação entre as propriedades das Bases de Dados Operacionais (coluna da esquerda) por oposição às propriedades dos *Data Warehouses* (coluna da direita).

Propriedade		Opção de escolha	
A	Acessos de leitura/escrita	<i>Data-driven</i>	1
B	Pesquisas pré-definidas	Carregamentos periódicos	2
C	Dados atualizados em tempo-real	Consultas <i>ad-hoc</i> e relatórios	3
D	Acessos a poucos registos de cada vez	Estrutura otimizada para consultas complexas	4
E	<i>Event-driven</i>	Muitos registos em cada acesso	5
F	Estrutura otimizada para atualizações	Acessos só de leitura	6

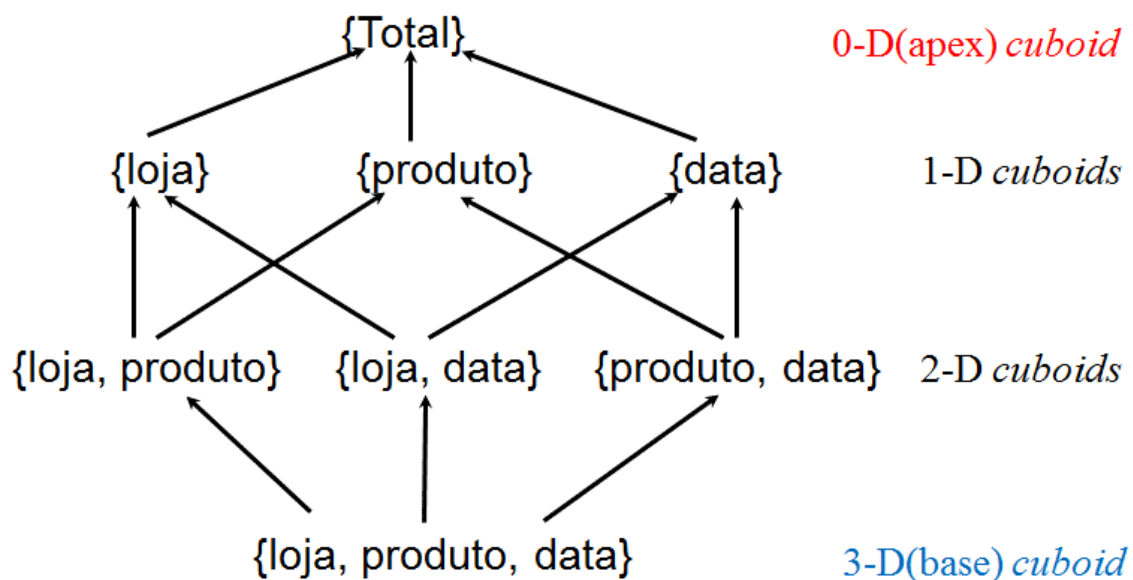
3. [0,75 valores]

Das seguintes opções, escolher a(s) correta(s) relativamente às características dos *Data Warehouses*.

- ☐ As suas estruturas estão otimizadas para poupança do espaço de armazenamento
- ☐ Têm uma forte dependência com o tempo
- ☐ As suas estruturas de dados podem conter redundância
- ☐ São concebidos sobretudo para a produção de *reports* estáticos
- ☐ Armazenam grandes quantidades de dados
- ☐ Os seus dados são bastante voláteis

4. [0,75 valores]

A imagem seguinte apresenta um grafo de cuboides (cubos de dados).

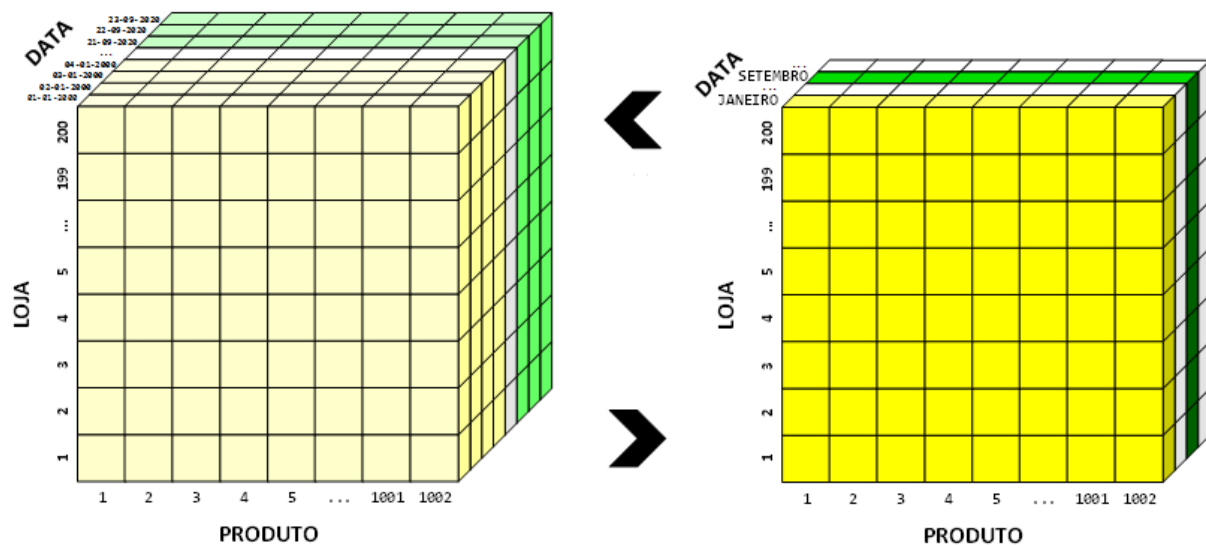


De entre as seguintes opções, escolher a(s) correta(s).

- ☐ O ápex cuboide contém o nível mais baixo de agregação
- ☐ O base cuboide contém o nível mais baixo de agregação
- ☐ O base cuboide contém o nível mais elevado de agregação
- ☐ A operação OLAP CUBE calcula todas as agregações possíveis
- ☐ O ápex cuboide contém o nível mais elevado de agregação
- ☐ A operação OLAP ROLLUP calcula todas as agregações possíveis

5. [1 valor]

A figura seguinte ilustra operações típicas realizadas num sistema OLAP.



Das seguintes opções, escolher a(s) que se referem às operações da figura.

- ☐ *Pivoting*
- ☐ *Roll-up*
- ☐ *Slicing*
- ☐ *Drill-down*
- ☐ *Dicing*
- ☐ *Cube*

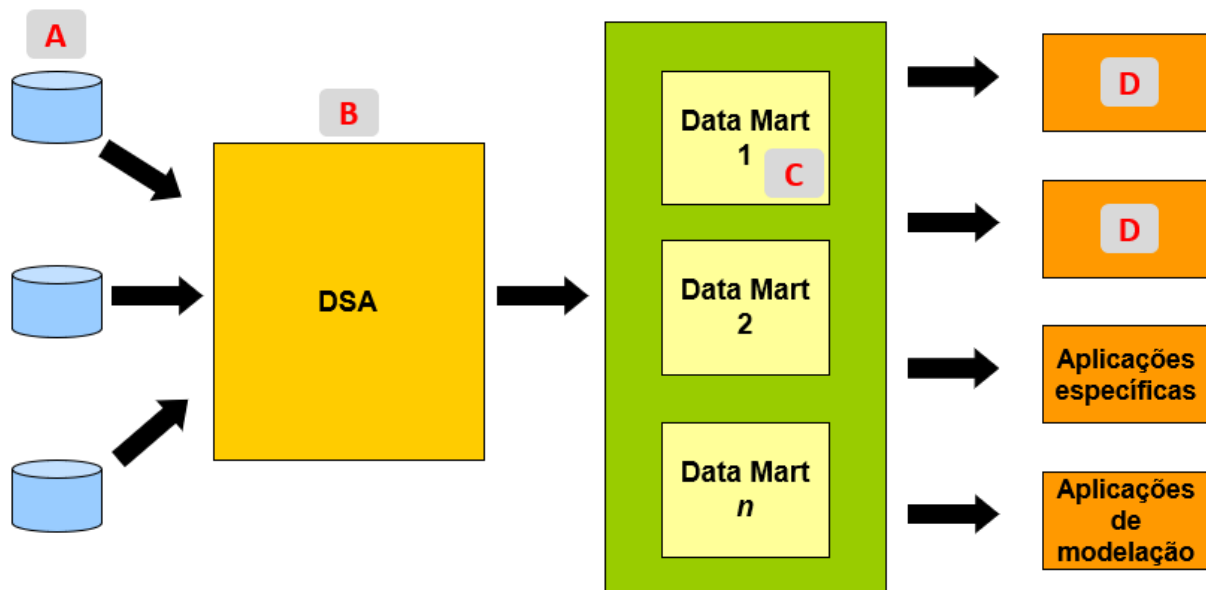
6. [1 valor]

Fazer a correspondência entre as propriedades dos sistemas OLAP (coluna da esquerda) e os valores adequados (coluna da direita).

Propriedade	Opção de escolha	
A	Modelo de dados	Poucos
B	Utilizadores	Normalizado
C	Período dos dados	Desnormalizado
D	Atualizações/inserções	Em tempo real
E	Atividade	Sobretudo atual
F	Objetivo crítico	Analítica
		Disponibilidade
		Histórico
		Operacional
		Performance elevada
		Batch
		Muitos

7. [1 valor]

A imagem seguinte é referente à abordagem de Ralph Kimball para um *Data Warehouse*.



Para cada conceito (coluna da esquerda), escolher a letra da imagem onde esse conceito pode ser encontrado/implementado.

Conceito	Letra na Imagem
Concatenação de <i>strings</i>	
<i>Screens</i>	
<i>Queries ad-hoc</i>	
Sistema OLTP	
Ferramenta OLAP	
Estruturas de otimização do desempenho de <i>queries</i>	

8. [0,75 valores]

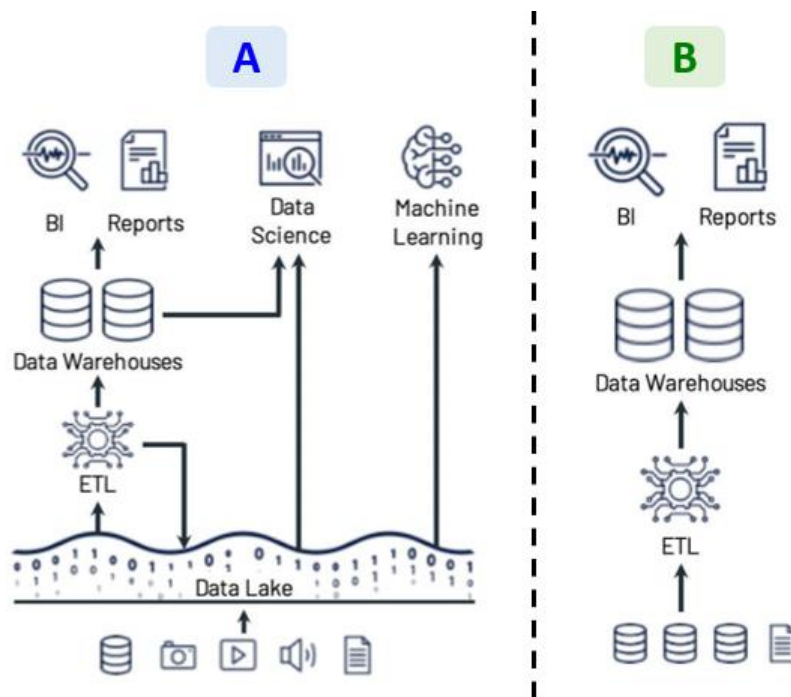
As quatro operações seguintes (A, B, C e D) são algumas das necessárias para a construção de um *Data Warehouse*:

- A. Afinação do desempenho das *queries*
- B. Identificar os objetivos analíticos a atingir
- C. Desenhar o modelo dimensional
- D. Implementar *screens*

Indique a ordem certa pela qual estas etapas devem ser realizadas.

9. [1 valor]

A figura seguinte apresenta diferentes arquiteturas (A e B) para o processo de *Data Warehousing*.



Das seguintes afirmações acerca destas arquiteturas, escolher a(s) verdadeira(s):

- ☐ A arquitetura B é a preferida dos *data scientists*
- ☐ A arquitetura A é mais recente do que a arquitetura B
- ☐ A abordagem clássica de Ralph Kimball é representada pela arquitetura A
- ☐ Na arquitetura A, os dados fonte são fortemente estruturados
- ☐ Destas arquiteturas, a arquitetura B é a mais próxima da plataforma usada nas aulas práticas
- ☐ As atividades de *Machine Learning* nunca usam dados tratados

10. [1 valor]

Das seguintes opções, escolher a(s) correta(s) relativamente ao processo de extração de dados.

- ☐ São registados os problemas de qualidade encontrados nos dados
- ☐ São utilizadas técnicas SCD
- ☐ O processo de eliminação pode, em algumas circunstâncias, ser usado para extrair dados de tabelas
- ☐ Permite extrair os dados fonte para a área de tratamento de dados
- ☐ Geralmente são utilizadas técnicas CDC
- ☐ O processo de eliminação permitirá, numa outra etapa, detetar quais os dados novos e/ou modificados

11. [1 valor]

A seguir, apresentam-se os passos típicos do algoritmo de uma técnica de extração incremental.

- A. Obter a data do último registo extraído da tabela fonte
- B. Limpar a tabela destino
- C. Guardar a data do último registo extraído da tabela fonte
- D. Extrair os dados da tabela fonte
- E. Obter a data do registo mais recente da tabela fonte

Indique a ordem correta dos passos do algoritmo.

12. [1,25 valores]

A figura seguinte ilustra um cenário após a etapa de extração de dados.

SISTEMA FONTE			DATA STAGING AREA	
PRODUCT			T_INFO_EXTRACTIONS	
serial_nr	name	src_last_changed	source_table_name	last_timestamp
164425	Monitor 23", LG	2023-10-10 09:15:07	PRODUCT	2023-10-13 14:15:20
234234	Monitor 25 inch., Samsung	2023-10-10 19:40:01
233324	Pen Drive 64GB	2023-10-13 14:15:20		
552424	USB Pen 32GB	2023-10-12 09:10:45		
776859	USB Pen 128GB	2023-10-13 10:12:10		

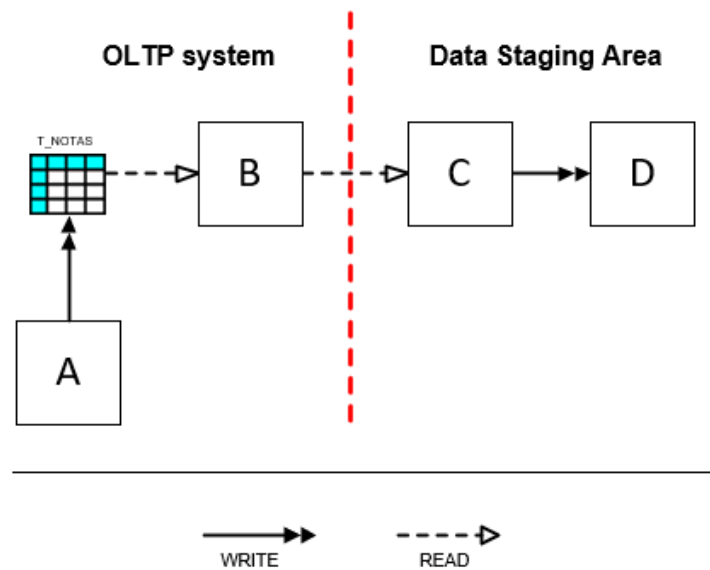
T_DATA_PRODUCTS	
id	name
233324	Pen Drive 64GB
776859	USB Pen 128GB

Selecione a(s) opção(ões) corretas para o cenário apresentado.

- ☐ Já ocorreu pelo menos uma extração antes da atual
- ☐ Foi utilizada uma técnica de extração completa
- ☐ O *timestamp* da tabela T_INFO_EXTRACTIONS não está correto
- ☐ Diz respeito à primeira extração
- ☐ Foi utilizada uma técnica de extração incremental
- ☐ A extração foi realizada no dia 13 de outubro de 2023







13. [1,25 valores]

Observe a seguinte imagem, representando a etapa de extração num *Data Warehouse* a partir de uma tabela relacional T_NOTAS.



Sabendo que a extração de dados a partir da tabela T_NOTAS é incremental e usa *timestamps*, escolha os objetos certos a colocar nos locais A/B/C/D.

Nota: Quando não houver equivalência, escolha a opção «objeto sem equivalência».

Objeto	Letra na Imagem
 trigger	
T_DATA_NOTAS 	
 transform_notas	
 load_notas	
VIEW 	
 table_extract	

14. [1 valor]

De entre as seguintes afirmações, escolher a(s) verdadeira(s) acerca do processo de transformação de dados.

- ☐ Podem ocorrer pesquisas em tabelas auxiliares (por exemplo, dicionários de dados)
- ☐ Os tipos de alterações SCD são implementados nesta etapa
- ☐ Rejeitar linhas com problemas de qualidade é a abordagem mais aconselhada
- ☐ Cada coluna vinda da etapa anterior terá sempre uma coluna idêntica no final da transformação
- ☐ Realizar junções entre tabelas é uma tarefa comum desta etapa
- ☐ Nesta etapa é gerada informação com interesse analítico para as futuras etapas de transformação

15. [1,25 valores]

A imagem seguinte apresenta os dados de uma tabela para onde foram extraídos dados de clientes, nomeadamente o código postal onde residem e a localidade associada ao código postal.

Nota: para cada código postal só é permitida uma localidade.

T_DATA_CUSTOMERS

<u>id</u>	nome	cod_postal	localidade	...
1	José Teixeira	3030-105		...
2	Maria Rita Violeta	1020-573	LISBOA	...
3	Antónia S. Teles	2430-120	LEIRIA	...
4	João Vasconcelos	1020-573	PORTO	...
5	Joana Maria Crava	2430-120	leiria	...
...

Escolher o(s) critério(s) de qualidade de dados que a coluna localidade não respeita:

- ☐ Conformidade
- ☐ Correção
- ☐ Consistência
- ☐ Coerência
- ☐ Certeza
- ☐ Completude

16. [1 valor]

Das seguintes opções, escolher o(s) documento(s) chave usado(s) pelos arquitetos de um *Data Warehouse* durante a implementação dos métodos da etapa de transformação:

- ☐ Requisitos do negócio
- ☐ Mapa Lógico de Dados
- ☐ Casos de uso
- ☐ Mapa Local da Arquitetura
- ☐ Modelos Lógicos das fontes de dados
- ☐ Lista de pesquisas *ad-hoc*

17. [1,25 valores]

A figura seguinte apresenta a especificação parcial de um dos documentos de um *Data Warehouse*.

source			target		target			operation
-	-	-	-	-	-	-	-	-
LOJAS.CSV	Nomeação da loja	varchar	t_data_stores_new	name	T_CLEAN_STORES	name	VARCHAR2(100)	-
	Referência da loja	char(6)	t_data_stores_new	reference		reference	CHAR(6)	-
	Nome do edifício	varchar	t_data_stores_new	building		address	VARCHAR2(250)	concatenate (IF (nome_edificio=' ') THEN NULL ELSE concatenate(nome_edificio,' - '); endereço; ' / ' ; código_postal; ', ' ; localidade)
	Endereço	varchar	t_data_stores_new	address				
	Código Postal	char(8)	t_data_stores_new	zip_code				
	Localidade	varchar	t_data_stores_new	location		location	VARCHAR2(40)	-
	Localidade	varchar	t_data_stores_new	location				
	distrito	varchar	t_data_stores_new	district		district	VARCHAR2(30)	-
	Código Postal	char(8)	t_data_stores_new	zip_code		zip_code	CHAR(8)	-
	Telefones	varchar	t_data_stores_new	telephones		telephones	CHAR(9)	SUBSTR(telefones sem espaços e sem pontos,1,9)
	Fax	varchar	t_data_stores_new	fax		fax	CHAR(9)	-
	Data de encerramento	date	t_data_stores_new	closure_date		status	VARCHAR2(8)	IF (data_encerramento IS NULL) THEN 'ACTIVE' ELSE 'INACTIVE'
MANAGERS.CSV	Referência da loja	char(6)	t_data_stores_new	reference	-	-	(JOIN OPERATION) (t_data_stores_new MINUS t_data_stores_old) JOIN (t_data_managers_new MINUS t_data_managers_old)	
	Referência da loja	char(6)	t_data_managers_new	reference	-	-		
	Director da loja	varchar	t_data_managers_new	manager_name	manager_name	VARCHAR2(100)		
	Data de entrada	date	t_data_managers_new	manager_since	manager_since	DATE		

Das seguintes afirmações, escolher a(s) verdadeira(s) acerca do documento.

- ☐ Todos os dados fonte são transformados em *strings*
- ☐ Das colunas-fonte visíveis no documento, algumas não estão a ser usadas na etapa de transformação
- ☐ Todos os dados fonte são modificados
- ☐ Todas as etapas ETL estão presentes no documento
- ☐ São usadas 16 colunas-fonte na etapa de extração
- ☐ Os dados fonte relativos à Data de encerramento sofrem uma mudança de tipo de dados

18. [1 valor]

De entre as seguintes opções, escolher a(s) correta(s) relativamente ao processo de carregamento de dados.

- ☐ A tabela de factos é carregada antes das tabelas de dimensão
- ☐ São carregados muitos registos de uma só vez
- ☐ As técnicas SCD permitem lidar com as alterações nas dimensões
- ☐ As tabelas de dimensão são carregadas antes da tabela de factos
- ☐ Nos carregamentos periódicos, os problemas com atualizações nas dimensões não se colocam
- ☐ As estruturas de otimização podem atrasar o carregamento

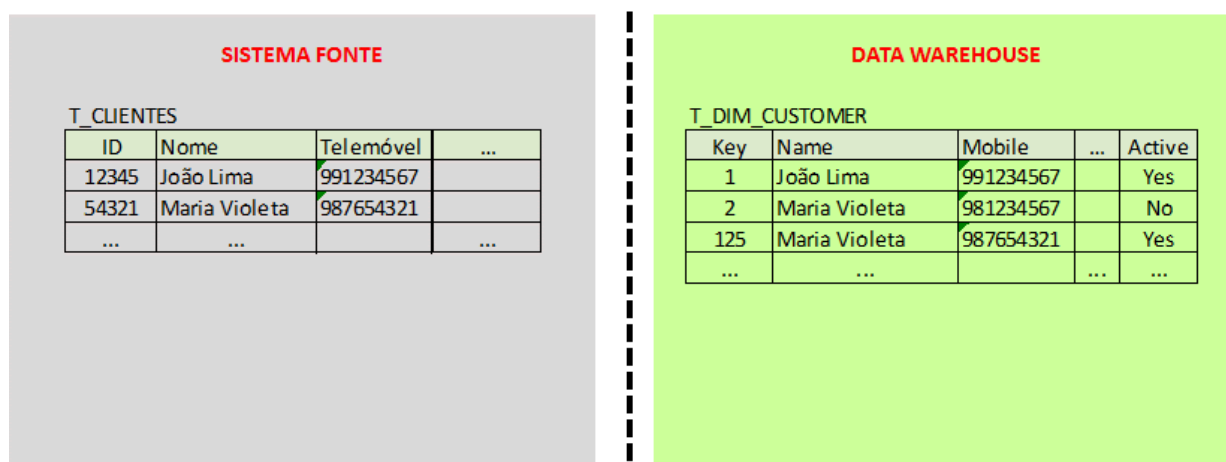
19. [1 valor]

Alguns mecanismos de otimização do desempenho do *Data Warehouse*, se estiverem ativos, podem prejudicar a rapidez da etapa de carregamento de dados. Escolher esse(s) mecanismo(s):

- ☐ Logs do sistema operativo
- ☐ Particionamento de tabelas
- ☐ Views
- ☐ Agregados
- ☐ Índices
- ☐ Desagregados

20. [1,25 valores]

A figura seguinte ilustra um cenário após a etapa de carregamento de dados.



Selecione a(s) opção(ões) corretas para o cenário apresentado.

- ☐ Na tabela de dimensão podem existir várias versões do mesmo registo
- ☐ O atributo Active não é necessário
- ☐ Já foi feito mais do que um carregamento para o *Data Warehouse*
- ☐ O atributo Mobile sofre alterações tipo 3 (SCD3)
- ☐ A chave primária da tabela de dimensão vem do sistema fonte
- ☐ O atributo Mobile sofre alterações tipo 2 (SCD2)