

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO

2ª PROVA DE BD III – PROF. SERGIO SERRA

LEIA AS INSTRUÇÕES E O CADERNO DE PERGUNTAS COM ATENÇÃO ANTES DE FAZER A PROVA

OBSERVAÇÃO 1 - Esta prova é teórico-prática, pode ser feita em casa, sendo entregue EXCLUSIVAMENTE através do SIGAA, conforme informado em sala de aula. Você pode consultar softwares a vontade. NÃO COPIE E COLE de outras fontes (Internet ou de colegas). As provas serão corrigidas por comparação. Provas assemelhadas serão zeradas. Essa prova vale 10 pontos.

OBSERVAÇÃO 2 - Você tem várias horas para fazer a prova, pode entregar **até dia 16/07/2022 as 17:59:00**. Ela deve ser anexada ao SIGAA, **NÃO envie por email**, não serão corrigidas provas enviadas por e-mail. É responsabilidade EXCLUSIVA do aluno garantir que a prova foi submetida e recebida pelo professor. Certifique-se disso. A cada HORA DE ATRASO será atribuído 1,0 ponto de desconto da nota final da prova. Para isso usarei o relógio do Google. Não deixa para a última hora. Faça a sua prova com calma. Além disso, você deverá entregar a prova ASSINADA! Para isso basta salvar sua assinatura junto ao arquivo da prova anexado, o arquivo final deve ter **formato PDF**. **NÃO ENVIE VERSÕES DISTINTAS**. Arquivos com vírus terão automaticamente a nota ZERO!

OBSERVAÇÃO 3 - A frequência do dia da prova será atribuída quando o professor receber o material enviado pelo aluno (este caderno de perguntas e respostas de cada questão da prova). Caso não emita resposta, a frequência não será atribuída e a nota será ZERO. Caracterizará uma falta na avaliação.

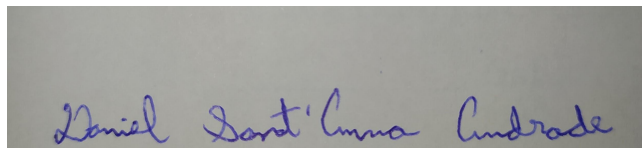
OBSERVAÇÃO 4 - **NÃO TENTE BURLAR O SISTEMA**. Seja criativo, a prova não é pesada nem difícil, responda na ordem das perguntas e descreva em DETALHES AS TODAS as soluções encaminhadas para cada questão.

PROVA INDIVIDUAL, RESPOSTAS SOMENTE NA COR AZUL

Aluno: Daniel Sant' Anna Andrade

Matrícula: 20200036904

Assinatura:



1ª QUESTÃO (2.5 pontos por item)

Disserte tecnicamente sobre:

- a) O que é processo ETL e ELT? Quais são as suas principais características técnicas e diferenças. Na prática, como e onde são utilizados?

R = ETL é um forma de integração de dados realizada em três etapas: extração (extract), transformação (transform) e carregamento (load). Ele é utilizado em contexto de banco de dados distribuídos pois permite combinar dados de diversas fontes, e é geralmente utilizado para criar data warehouses. Seguindo as três etapas, os dados são extraídos de um sistema de origem, transformados em um formato em que possam ser analisados e carregados em um outro sistema.

ELT é uma modernização do processo de ETL, também em 3 etapas, porém ocorre em ordem diferente, utilizada para aprimorar a performance. Ele se torna mais

rápido para o carregamento e o processamento de dados apenas por inverter a ordem das etapas. Assim, ele extrai dados de um sistema de origem, os carrega em um sistema de destino e só após isso ele realiza a transformação, sendo assim mais simples para lidar com grandes volumes de informação, reduzindo o tempo de carregamento e sendo mais eficiente por aproveitar todo o processamento desenvolvido na configuração do data warehouse, diminuindo a transferência de dados.

Outras diferenças são que nos ETL os profissionais focam em todas as 3 etapas, enquanto no ELT os profissionais só focam na etapa de extração e carregamento. Nos ETL possuem altas taxas de manutenção, enquanto nos ELT os analistas possuem mais independência para o manuseio de dados brutos. Nos ETL a abordagem é muito massiva não sendo viável para empresas pequenas e médias, nos ELT o processo é escalável, adaptável para todos os tamanhos de empresas.

Apesar das diferenças, os dois processos são utilizados para atender o mesmo propósito, organizar os dados para tomada de decisão. Em suma, o ELT é utilizado para processos mais econômicos pois utiliza menos recursos e leva menos tempo, enquanto o ETL é mais robusto, aguentando maior quantidade de informação.

- b) O que é tabela de fatos em modelos multidimensionais? Quais suas características? Existe alguma relação com tabela de fato e cubo multidimensional? Explique.

R = Uma tabela de fatos é uma tabela que armazena valores detalhados de medidas em um data warehouse. Ela geralmente armazena uma grande quantidade de dados históricos e redundantes e resume os relacionamentos entre as dimensões. Essas tabelas de fatos se relacionam com tabelas dimensionais que são adjacentes a elas.

2ª QUESTÃO (2 pontos)

Quais as principais diferenças entre o modelo estrela e floco de neve? Explique cada um deles e apresente e discuta um exemplo de cada um.

R = No modelo estrela as tabelas relacionam-se com a tabela de fatos, dessa forma, a tabela de fatos fica ao centro cercada de tabelas dimensionais, com o seu formato se assemelhando a uma estrela.

No modelo floco de neve, assim como no anterior, as tabelas dimensionais relacionam-se com a tabela de fatos, porém, algumas dimensões relacionam-se apenas entre elas, isto ocorre para que seja possível normalizar as tabelas dimensionais, diminuindo o espaço utilizado por essas tabelas.

Enquanto o modelo floco de neve reduz o espaço de armazenamento, ele adiciona várias tabelas ao modelo, deixando o mais complexo, tendo mais tabelas para serem utilizadas em uma consulta. Já no modelo estrela, ele é mais simples e fácil de implementar, porém repete muitas informações, o que aumenta muito o espaço utilizado.

3ª QUESTÃO (3 pontos)

A tecnologia de Data Warehouse oferece suporte às ferramentas OLAP, que apresentam visões multidimensionais de dados permitindo a análise das operações de negócio para facilitar a tomada de decisões. Estas ferramentas suportam algumas operações de maneira a dar aos analistas o poder de observar os dados de várias maneiras em níveis diferentes. Considere duas destas operações mostradas nas figuras abaixo.

Ano	Dados	Região			
		Ásia	Europa	América do Norte	Total Geral
2010	Soma de Hardware	97	23	198	318
	Soma de Software	83	41	425	549
2011	Soma de Hardware	115	28	224	367
	Soma de Software	78	65	410	553
2012	Soma de Hardware	102	25	259	386
	Soma de Software	55	73	497	625
Soma de Hardware Total		314	76	681	1071
Soma de Software Total		216	179	1332	1727

Região	Dados	Ano			
		2010	2011	2012	Total Geral
Ásia	Soma de Hardware	97	115	102	314
	Soma de Software	83	78	55	216
Europa	Soma de Hardware	23	28	25	76
	Soma de Software	41	65	73	179
América do Norte	Soma de Hardware	198	224	259	681
	Soma de Software	425	410	497	1332
Soma de Hardware Total		318	367	386	1071
Soma de Software Total		549	553	625	1727

Figura 1

Região	Variação de vendas
Africa	105%
Ásia	57%
Europa	122%
América do Norte	97%
Pacífico	85%
América do Sul	163%

Pais	Variação de vendas
China	123%
Japão	52%
Índia	87%
Cingapura	95%

Figura 2

Explique e descreva os tipos de operação aplicadas nas figuras.

R = A figura 1 se trata da operação Pivot (rotação), onde os dados são vistos sob uma nova perspectiva. Assim, antes a organização principal era através de cada ano e era detalhado a quantidade de hardware e software através das regiões, na segunda imagem da figura 1 ele organiza através da região e detalha através dos anos. O resultado final (soma de Hardware total e soma de Software Total), continua com os mesmos valores, ou seja, não é alterada a análise da informação, é sim apenas colocada em uma nova perspectiva.

A figura 2 se trata da operação Drill Down onde se aumenta o nível de detalhe da informação aumentando a granularidade, aumentando a velocidade do acesso às informações e o volume de dados armazenados. Na Figura 2, a tabela demonstrava a variação geral de vendas em cada uma das regiões, ao aumentar a granularidade da informação referente a Ásia, ela demonstra a variação geral de vendas em cada um dos países daquela região.

4ª QUESTÃO (até 0,5 ponto EXTRA)

Insira aqui o resumo descritivo (1 lauda completa) do que foi apresentado em uma das quatro palestras da oficina de dados do consórcio GO-FAIR AGRO (links no SIGAA). Explique claramente qual a relação do que foi apresentado na palestra com o conteúdo desta disciplina.

BOA SORTE