

Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação
Disciplina SCC0245 - Processamento Analítico de Dados
Disciplina SCC5968 - Processamento Analítico de Dados em Larga Escala

docente

Profa. Dra. Cristina Dutra de Aguiar (cdac@icmc.usp.br)

Aluno PAE

João Pedro de Carvalho Castro (jp.carvalhocastro@alumni.usp.br)

Primeiro Trabalho Prático

Este trabalho tem por objetivo reforçar o conhecimento do aluno com relação à modelagem de um problema do mundo real e ao desenvolvimento de uma aplicação de *data warehousing*.

O trabalho deve ser feito em grupos de até 3 alunos. A solução deve ser proposta exclusivamente pelo grupo com base nos conhecimentos adquiridos nas aulas. Consulte as notas de aula e o material indicado na página da disciplina quando necessário.

Definições a Serem Realizadas

Os grupos devem definir os seguintes aspectos relacionados ao trabalho:

- Tema geral do trabalho.
- Pelo menos 3 assuntos de interesse (ou processos de negócio).
 - Sinônimos para assuntos de interesse: foco de interesse, métrica ou medida numérica (no nível conceitual), fatos (no nível lógico).
- Pelo menos 2 esquemas estrela, usando como base os assuntos de interesse. Na proposta dos esquemas estrela, deve-se ter em mente a posterior criação de pelo menos 1 constelação de fatos.
- Pelo menos 2 fontes de dados, os quais representam fontes de dados que serão usados para popular a aplicação a ser desenvolvida. É permitida a geração de dados sintéticos caso necessário. Entretanto, é fortemente indicado o uso de dados reais e volumosos.

- É necessário usar, pelo menos, duas fontes de dados.
- As fontes de dados podem ser obtidas da mesma plataforma/repositório.
- Pelo menos 5 principais consultas a serem respondidas.
 - As consultas devem ser especificadas de acordo com os aspectos dinâmicos do modelo de dados multidimensional.

Estrutura Geral do Trabalho

Capa

As seguintes informações devem ser fornecidas na capa do trabalho: o nome da instituição, o nome do curso, o nome da disciplina, o nome do professor responsável, o nome do trabalho, o nome dos participantes e os respectivos números USP, e a data de entrega do trabalho.

Índice

O índice deve listar os nomes das seções que compõem o trabalho e as suas respectivas páginas de início.

Descrição do Problema

Esta seção deve descrever detalhadamente as características do problema do mundo real a ser modelado. Deve ser feita uma descrição explicativa dos aspectos descritos a seguir. Na descrição explicativa, devem ser apresentadas justificativas para a escolha e para a importância.

- Qual o tema geral do trabalho.
- Quais os processos de negócio que serão modelados.

- Para cada processo de negócio, qual o grão do processo de negócio, quais as dimensões que aplicam a cada linha da tabela de fatos e quais os fatos propriamente ditos.
- Para cada dimensão, quais seus atributos descritivos e como esses atributos se relacionam entre si.
- Para cada fato, se ele é aditivo, semiaditivo ou não aditivo. Se o fato for derivado, explicar porque ele é derivado e como ele pode ser obtido.
- Para cada atributo e cada fato, como ele será povoado. Ou seja, de qual(quais) fonte(s) o dado vai ser obtido, se é necessário fazer a integração e, em caso positivo, como essa integração será realizada.
- A integração é necessária quando o mesmo dado é obtido de dois ou mais provedores diferentes. Por exemplo, pode-se obter de um provedor dados de sexo representados como 'masculino' e 'feminino' e de outro provedor dados de sexo representados como 0 e 1. No processo de integração, deve ser definido como esse valor vai ser representado no data warehouse.

Observação 1. Essa seção requer a descrição explicativa de todos os aspectos. Por exemplo, para cada dimensão, explicar o seu significado, bem como todos os seus atributos. Para cada atributo, também explicar o seu significado e identificar quais atributos se relacionam entre si e quais atributos são apenas descritivos, não se relacionando com outros atributos.

Observação 2. Lembre-se que o trabalho será analisado por uma pessoa que não participou do desenvolvimento do mesmo. Assim, quaisquer particularidades ou características interessantes ao problema do mundo real devem ser descritas nessa seção.

Esquemas

Essa seção deve exibir os esquemas estrela para o problema do mundo real sendo analisado, de acordo com a descrição do problema realizada anteriormente. Esses esquemas devem ser descritos da seguinte forma:

- Para cada esquema estrela, fazer o desenho do esquema estrela.
- Para cada dimensão, definir os seus atributos. Especificar também a(s) hierarquia(s) de atributos usando o operador de ordenação parcial.
- Para cada fato, mostrar qual a função de agregação que será usada para o fato ser agregado. Por exemplo, se o fato é aditivo, ele deve ser agregado usando-se a função de agregação SOMA.

Essa seção também deve exibir as constelações de fatos para o problema do mundo real sendo analisado, de acordo com a descrição dos esquemas estrela. Para cada constelação de fatos:

- Fazer o desenho da constelação de fatos.
- Explicar o processo que foi necessário para projetar a constelação de fatos. Essa explicação pode ser feita de forma descritiva ou na forma de desenho.

Observação 1. Certifique-se de que os esquemas realmente refletem as características do problema do mundo sendo modelado.

Observação 2. Lembre-se que, na sala de aula, foram discutidos aspectos de modelagem adicionais, incluindo: (i) para as tabelas de dimensão, campos artificiais como chave primária; (ii) tabela bugiganga; (iii) dimensão degenerada; e (iv) tabela de fatos sem fato.

Observação 3. Os esquemas devem seguir, obrigatoriamente, a mesma notação utilizada em sala de aula.

Consultas

Essa seção deve especificar as principais consultas a serem suportadas pela aplicação. Nesse primeiro trabalho prático, as consultas devem ser especificadas somente na forma textual. Devem ser especificadas consultas caracterizadas da seguinte forma:

- Pelo menos 2 consultas *roll-up* ou *drill-down*.
- Pelo menos 1 consulta *slice and dice*, podendo ser apenas *slice*, apenas *dice* ou *slice and dice*.
- Pelo menos 1 consulta *pivot*.
- Pelo menos 1 consulta *drill-across*.

Observação 1. Deve ser feita uma descrição da importância e da justificativa de cada consulta.

Observação 2. Lembre-se que os esquemas propostos devem oferecer suporte para as consultas a serem realizadas. Portanto, essas consultas devem ser levadas em consideração quando da definição desses esquemas.

Observações Adicionais

[1] As correções realizadas em uma parte do trabalho devem ser realizadas pelo grupo para a próxima parte do trabalho. Na entrega da segunda parte do trabalho, o documento referente à primeira parte do trabalho que contém as correções propostas também deve ser entregue.

[2] Os integrantes do grupo devem constar na capa do trabalho. Nomes de integrantes que não constarem na capa receberão a nota zero (0).

[3] Em caso de plágio, as notas dos trabalhos envolvidos serão zero (0).

[4] Podem cair questões na prova relativas ao desenvolvimento do trabalho, caso o critério de avaliação do oferecimento corrente inclua provas.

Instruções de Entrega

Material a ser entregue.

- (1) Um documento no formato .pdf.
- (2) Uma apresentação feita pelos integrantes do grupo, a qual deve explicar o trabalho desenvolvido. Ou seja, o grupo deve apresentar: uma motivação geral para a escolha do tema do trabalho e dos assuntos de interesse. Adicionalmente, devem ser apresentados os seguintes resultados: as fontes de dados, os esquemas estrela, a constelação de fatos e as consultas, além de destacados os aspectos de decisão tomados na especificação desses resultados. Todos os integrantes do grupo devem participar da apresentação. Participantes que não participarem da apresentação ficarão com a nota final igual a zero. O tempo de apresentação dos integrantes deve ser homogêneo. Durante a apresentação, serão realizadas perguntas sobre o trabalho desenvolvido.

Local de entrega: Conforme especificado no critério de avaliação do oferecimento corrente.

Data de entrega: Especificada na página da disciplina.

Bom Trabalho !