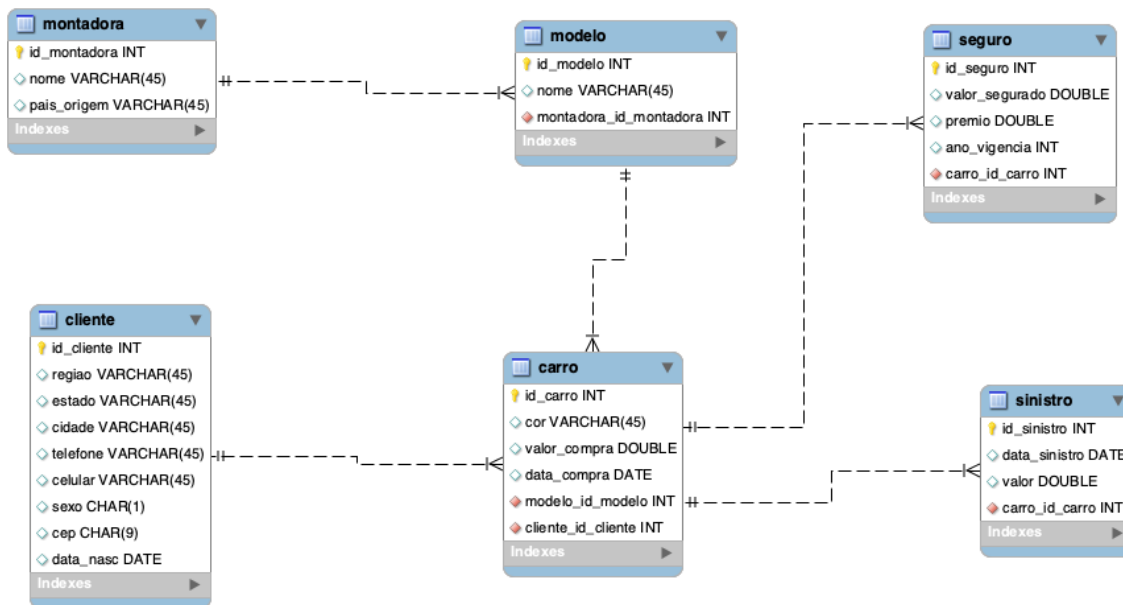


Trabalho 2

Este trabalho deverá ser feito pelos **mesmos grupos do trabalho 1**, e entregue até o dia **20/11/2019**. Não deixem para o último fim de semana. Comecem a fazer o trabalho desde já!

O trabalho consiste em analisar um banco de dados transacional de uma seguradora de veículos. O banco de dados será fornecido pelo professor com dados fictícios, e portanto não necessariamente fazem sentido em nossa realidade!

São arquivos texto com vários comandos INSERT de SQL. Ou seja, desde que sejam criadas as tabelas corretamente em um SGBD, os dados destes arquivos podem ser inseridos em um banco de dados relacional. Abaixo está um diagrama para este banco de dados relacional:



São 6 arquivos, cada um contendo comandos INSERT de uma das seguintes tabelas: montadora, modelo, carro, cliente, seguro e sinistro. Cada tabela possui sua chave primária (um “id”). Abaixo estão exemplos de comandos INSERT de cada arquivo (cada comando INSERT está em uma linha do arquivo):

```
INSERT INTO montadora(id_montadora, nome, pais_origem) VALUES(1, 'fiat', 'Italia');
```

```
INSERT INTO modelo(id_modelo, nome, montadora_id_montadora) VALUES(1, 'uno', '1');
```

```
INSERT INTO cliente(id_cliente, regioao, estado, cidade, telefone, celular,
sexo, cep, data_nasc) VALUES(1, 'norte', 'PA', 'placas', '', '70338553', 'F',
'40000-205', '1992-11-16');
```

```
INSERT INTO carro(id_carro, cor, valor_compra, data_compra, modelo_id_modelo,
cliente_id_cliente) VALUES(1, 'bege', 58795.00, '2014-03-25', 26, 1);
```

```
INSERT INTO seguro(id_seguro, valor_segurado, premio, ano_vigencia,
carro_id_carro) VALUES(1, '58795.00', 2939.75, 2014, 1);
```

```
INSERT INTO sinistro(id_sinistro, data_sinistro, valor, carro_id_carro)
VALUES(1, '2014-03-20', 1500.00, 1);
```

Detalhes importantes:

- observe que é um banco não necessariamente normalizado.
- os valores dos atributos “id” (chaves primárias) não refletem a “cronologia” das transacoes. Exemplo: um carro comprado em 2016 pode ter um “id” em sua tupla menor do que um carro comprado em 2012.
- considere que todos os carros foram comprados “zero quilômetro”.
- a forma de se “importar” estes dados para as ferramentas/SGBDs a serem utilizados faz parte do trabalho.

Um arquivo zipado será passado a um integrante de cada grupo pelo professor, sendo um arquivo **diferente** para cada grupo. Ou seja, cada grupo terá seus próprios arquivos (seu próprio banco transacional, com dados diferentes dos demais) e conseqüentemente os resultados das análises poderão ser diferentes por grupo!

O objetivo do trabalho é utilizar ferramentas de banco de dados/data warehouse/business intelligence e conceitos da disciplina sobre esses assuntos e sobre data mining para fazer análises sobre o banco de dados transacional fornecido.

Possíveis tarefas a serem feitas:

- 1 - modelagem dos dados no modelo estrela ou floco de neve, descrevendo todo o processo de transformação, limpeza e carga dos dados (ferramentas e/ou programas/técnicas utilizadas) e o modelo resultante (tabelas de fatos e dimensões e como e por que foram geradas desta maneira);
- 2 - mostrar o banco de dados no modelo dimensional que foi criado e o datawarehouse (cubos de dados) no business intelligence development studio do SQL server (ou outra ferramenta);
- 3 - mostrar uma aplicação OLAP sobre os dados do datawarehouse, usando os próprios recursos do SQL server. Formule possíveis perguntas do executivo sobre os dados e mostre as consultas com as respostas diretamente nos cubos de dados ou em outra aplicação ou front-end (excel?);
- 4 - modelar os dados e usar as ferramentas de tal forma a mostrar drill-down e roll-up em algumas dimensões (que façam sentido, é claro), ou pelo menos construir as hierarquias que façam sentido e mostrar consultas aos cubos filtrando por hierarquias.
- 5 - usar as ferramentas de relatórios do SQL Server (ou outra ferramenta) para produzir relatórios interessantes (respondendo perguntas dos executivos), que tenham parâmetros em tempo de execução e que utilizem nos parâmetros as hierarquias e métricas criadas.

6 - analisar os dados de acordo com os conceitos de Data Mining que estudamos, tentando descobrir conhecimentos (montando possíveis relações e tentando comprová-las através de consultas ao datawarehouse ou mesmo utilizando os recursos de Data Mining do SQL Server ou de outras ferramentas, como o Weka).

7 - fazer o mesmo datawarehouse e mostrar uma aplicação OLAP usando outras ferramentas (SQL server com outra aplicação OLAP ou outro *front-end* ou ainda outro SGBD com outra aplicação OLAP);

8 – comparar os resultados obtidos usando apenas o SQL Server e usando outras ferramentas, em conjunto com o SQL Server ou ainda apenas outras ferramentas.

Observações:

- para cada tarefa realizada, deve-se ter uma seção no relatório a ser entregue com a explicação sobre a tarefa. A tarefa também deve ser apresentada e discutida na aula de apresentação. Na apresentação devem ser mostrados cada um dos resultados resumidamente nos slides, bem como nas ferramentas (sendo executadas no momento da apresentação!);

- as tarefas de importação dos dados, possíveis formatações, etc, também devem ser documentadas e explicadas;

- o uso de cada ferramenta também deve ser explicado no relatório e na apresentação (principais recursos, formatos de arquivos, bancos de dados, etc).

- ferramentas já utilizadas por alunos: SQL Server 2008, SQL Server 2014, Business Intelligence Development Studio, Microsoft SQL Server Data Tools - Business Intelligence for Visual Studio 2013, Weka, rapidminer, Pentaho, Excel.

- outras ferramentas também podem ser utilizadas, desde que explicadas e mostradas em execução.

Não é obrigatório fazer todas as tarefas, desde que o grupo tenha uma boa justificativa para não fazer algumas e faça muito bem as demais. Lembrando que os trabalhos serão avaliados separadamente e por comparação com os demais (não copiem ideias).

Por outro lado, quanto mais tarefas forem feitas, mais pontos o grupo poderá ganhar, inclusive pontos extras!

Deverá ser entregue um **relatório**, contendo todo o processo de criação do data warehouse, do uso das ferramentas, de cada análise que foi feita, e como foi feita, e com a ajuda de quais ferramentas, ou seja, englobando todas as tarefas realizadas. Os dados deverão ser apresentados em texto e também em gráficos, sempre que possível, explicando no relatório como cada resultado foi obtido. É importante referenciar as figuras no texto. Imagine que ao ler esta seção, o professor fará o papel do executivo responsável pela seguradora.

A criatividade, o uso adequado de ferramentas, a qualidade, relevância e corretude dos resultados obtidos serão alguns dos critérios utilizados pelo professor na avaliação do trabalho, além é claro dos materiais entregues e da apresentação.

O relatório deverá ser entregue **em um arquivo PDF** até o dia **20/11/2019 até às 07h50min** (através do PVAnet). Deverá ser entregue também no mesmo dia e horário um arquivo **PDF contendo os slides da apresentação**. As apresentações serão feitas nos dias **20/11/2019 e 25/11/2019. Cada apresentação deverá ter entre 15 e 22 minutos.**

Bons estudos e bom trabalho!