

Sistemas Informação II 25/26

Trabalho Prático - Desenvolvimento de um Sistema de Apoio à Decisão com Data Mart para o Porto Marítimo da Figueira da Foz

Regras gerais

O trabalho prático tem como principal objetivo criar um pequeno **data mart** para analisar viagens de **navios comerciais** com destino final no **Porto da Figueira da Foz** e apresentar um **dashboard** em Microsoft Power BI para a análise da informação. Apenas **viagens concluídas** entram na análise. Eixo temporal padrão: **data_chegada**. Moeda: **EUR**

Em termos técnicos, abrange conhecimentos de **data warehouse**, **Power BI**, **SQL** e **Python**, e deve ser concretizado utilizando as linguagens, tecnologias, APIs e mecanismos trabalhados nas aulas. O foco está no uso correto dos mecanismos e recursos lecionados; não será atribuída preferência a escolhas de implementação alheias ao âmbito da disciplina.

Não é permitida uma abordagem baseada na mera colagem de excertos de outros programas ou exemplos. Todo o código e todas as implementações apresentadas têm de ser compreendidos e explicados por quem os apresenta; caso contrário, não serão contabilizados.

1. Descrição geral, conceitos principais e elementos envolvidos

Administração do Porto da Figueira da Foz pretende monitorizar a sua atividade através de dashboards em Microsoft Power BI, suportados por um data mart de baixa complexidade e atualização frequente. Basicamente pretende-se analisar informações sobre barcos que já concluíram a viagem. (esta informação é útil, para quando estiver a fazer o etl, ter em atenção quais valores irá tratar/importar).

O objetivo é disponibilizar indicadores estratégicos claros que permitam responder a questões de gestão muito objetivas, como compreender quanto se arrecada em taxas por país de origem das viagens, perceber a contribuição das chegadas por país de destino para a receita global, avaliar quanto se cobra em média por viagem, observar o volume de contentores movimentados ao longo do tempo e por tipo de condutor (homem/mulher) e analisar o peso total movimentado. Os dados operacionais são fornecidos a partir de um SGBD MySQL que suporta o ERP do Porto Marítimo. A forma como este ERP funciona não interessa para o objetivo do trabalho. O SGBD onde estão os dados, referem-se às tabelas viagem, contentores, taxas, barco, condutor, empresabarcos, empresasdonaccontentor e localizacao, podendo existir um ficheiro CSV auxiliar com atributos de referência.

Em termos práticos o trabalho a realizar consiste em conceber um modelo dimensional em estrela simples com uma única tabela de factos, bem como um conjunto reduzido de indicadores de apoio à decisão, com visualização em Power BI. O sistema deve integrar dados provenientes de uma base de dados relacional e de um ficheiro CSV, recorrendo a um processo de ETL em Python para extrair, transformar e carregar a informação, garantindo consistência e qualidade dos dados.

Como a informação está distribuída por duas fontes, da base de dados relacional deve ir buscar apenas os dados genéricos que definem/descrevem os barcos, os ids de cada viagem assim como as localizações de origem e destino. (figura 1)

Os restantes dados devem ser tirados do ficheiro csv, que se encontra no infordocente, mas que tem o seguinte cabeçalho:

idviagem,data_partida,data_chegada,tipoviagem,idcondutor,nome_condutor,certificacao,perfil_condutor,receita_taxas_eur, num_contentores,peso_total_contentores_kg,teu_total,observacoes

Se achar necessário, os restantes dados podem ser simulados através de gerados de dados do Mockaroo.

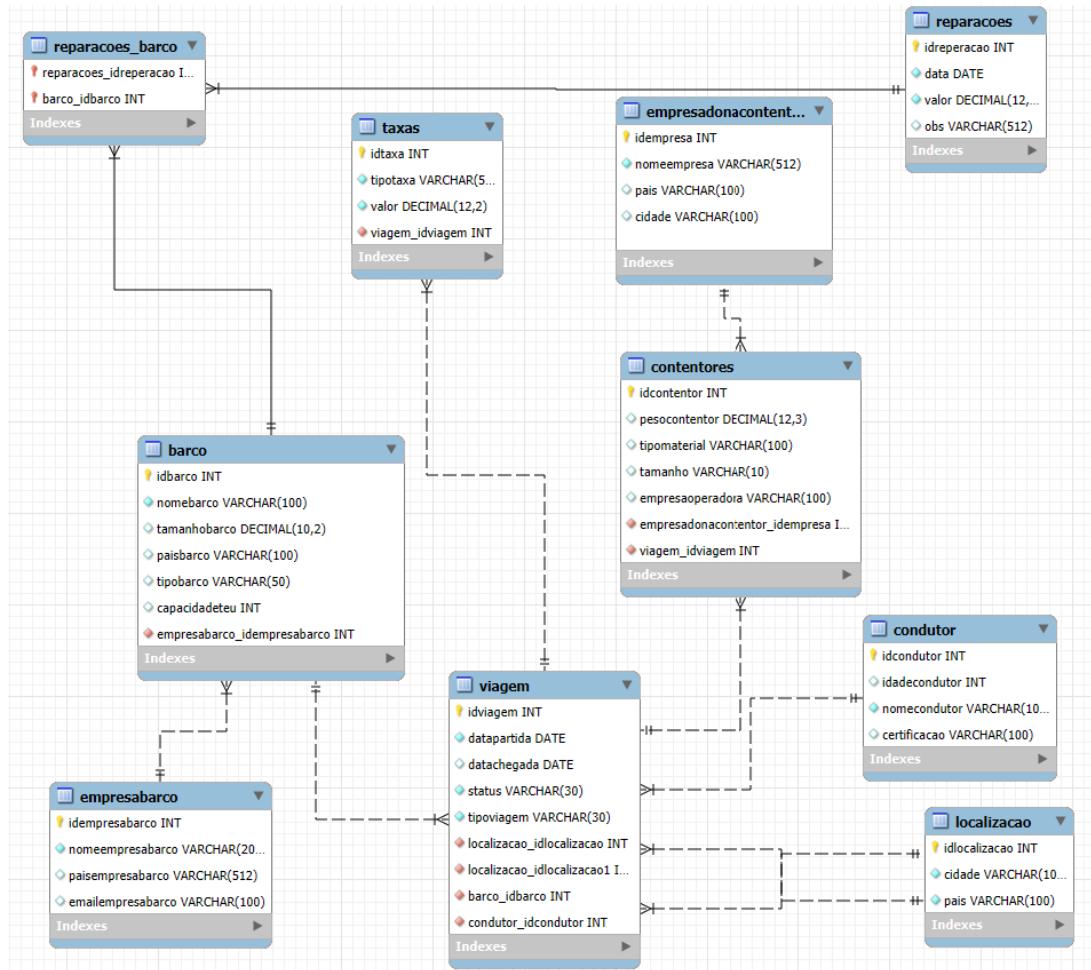


Figura 1 - BD Mysql - Dados operacionais

Notas: na tabela viagem o campo **localizacao_idlocalizacao** é o porto de origem, o **campo localizacaoo_idlocalizacao1** é o porto de destino;

Pretende-se um data mart e não um data warehouse corporativo, com foco na simplicidade e na rapidez de implementação.

2. Fonte de dados

Os dados a utilizar têm ser provenientes de três fontes, Em resumo:

- **MySQL (ERP):** A figura 1 representa o modelo de dados operacional. O script SQL para mySQL com a criação das tabelas e com os valores está disponível para download no infordocente. Deve analisar o modelo e recolher apenas os dados necessários para construir o seu modelo em estrela. Nota a moeda utilizada no mysql é euros.
- **CSV (Infordocente):** O ficheiro CSV é outra fonte de dados. Deve analisar o mesmo e recolher os dados necessários para introduzir no modelo em estrela. Nota a moeda utilizada no csv é dólares.
- **Dados gerados por si. (Se necessário, simular valores no Mockaroo.)**

3. KPI a desenvolver no dashboard

Para manter o objetivo do trabalho, definem-se cinco KPI obrigatórios, todos suportados por uma única tabela de factos ao nível da viagem:

- Viagens por país de origem com filtros por data completa e semestre.
- Receita total de taxas de viagens com destino à figueira da foz com filtros por país de origem e sincronizado com mapa.
- Número total de viagens realizadas por condutor e por tipo de certificado do condutor e por idade.
- Análise de barcos, tamanho e tipo de barco.
- Deve acrescentar 2 KPI's a sua escolha.

Nota: como referido anteriormente, só deve contabilizar viagens que tem o destino o porto da figueira da foz "figfoz". Para cada KPI deve aplicar filtros.

4. TAREFAS PROPOSTAS

1. Desenhar a arquitetura da solução.
2. Selecionar o processo de negócio, definir KPI e granularidade.
3. Construir o modelo em estrela e o dicionário de dados.
4. Implementar o modelo no SQL Server.
5. Realizar um script em python que realize o etl a partir dos dados de origem (base de dados mysql) para uma área staging ou diretamente no SQLServer;
6. Preencher o Data Mart com dados do CSV (completados com o mokaroo) e do MySQL.
7. Estimar espaço para 5 anos: apresente suposições e cálculos.
8. Apresentar um dashboard no Power BI que utilize o modelo em estrela criado anteriormente e responda aos kpis que identificou no processo. Justificar a escolha de cada vista de acordo com o processo de negócio.

Notas importantes:

- Deve realizar e entregar obrigatoriamente um relatório que explique todas as opções tomadas;
- Deve preparar uma pequena apresentação em powerpoint para 20 minutos;
- Em anexo deve entregar todo o código/scripts utilizados;
- Apresentar os dashboards criados no powerBI.
- No dia da defesa, devem trazer o sistema preparado e funcional de forma a poderem mostrar o trabalho realizado;
- Pode substituir os pontos 7 e 8 por dados por um unico dataset gerado por si, mas terá um desconto de 20% na nota do trabalho final.

5. Regras gerais do trabalho

Aplicam-se as seguintes regras, descritas na primeira aula e na ficha da unidade curricular (FUC):

- O trabalho pode e deve ser realizado em grupos de três alunos.
- Apresentação: 20 minutos, com demonstração do sistema e defesa obrigatória, que é individual e presencial. Podem existir moldes adicionais a definir e anunciados através dos canais habituais na altura em que tal for relevante (por exemplo, ser ou não necessário trazer computador).

- A entrega do trabalho prático é feita via nónio (inforestudante) através da submissão de um único **arquivo zip**¹ cujo nome respeita o seguinte padrão²:

si2_2526_tp_nome1_numero1_nome2_numero2_nome3_numero3.zip

(nome e número são os nomes e números de aluno dos elementos do grupo)
- A não adesão ao formato de compressão indicado (.zip) ou ao padrão do nome do ficheiro será penalizada, podendo levar a que o trabalho nem sequer seja visto.
- Do arquivo zip deverão constar:
 - Todo o **código fonte** desenvolvido;
 - Um **relatório** (em pdf) com o conteúdo relevante e que deverá ser da exclusiva autoria dos membros do grupo.
- Cada grupo submete o trabalho uma vez, sendo indiferente qual dos alunos o faz.
- **É obrigatório** que o aluno que faz a submissão **associe no nónio a entrega também aos outros membros do grupo**.
- **É conveniente que ambos estejam inscritos em turmas práticas (mesmo que seja em turmas diferentes). A não inscrição impede o registo do trabalho na plataforma, e, por conseguinte, pode levar à perda da nota.**

Data de entrega: **Domingo, 7 de dezembro, 2025.**

Importante: O trabalho deve ser realizado por ambos os elementos do grupo. Não são aceites divisões herméticas em que um elemento faz uma parte e apenas essa, nada sabendo do restante. Se num grupo existir uma participação desigual, o professor que faz a defesa deve ser informado do facto no início da defesa. Não tendo sendo informado e sendo detetada essa desigualdade, ambos os alunos ficarão prejudicados em vez de apenas aquele que trabalhou menos.

¹ Leia-se “zip” - não é arj, rar, tar, ou outros. O uso de outro formato poderá ser **penalizado**. Há muitos utilitários da linha de comando UNIX para lidar com estes ficheiros (zip, gzip, etc.). Use um.

² O não cumprimento do formato do nome causa atrasos na gestão dos trabalhos recebidos e será **penalizado**.