

Unidade Curricular: Sistemas de Informação II

Trabalho Teórico-Prático de Avaliação

Objectivo:

No âmbito do trabalho teórico-prático de avaliação da unidade curricular de Sistemas de Informação II, o objetivo é que os alunos desenvolvam competências no planeamento, criação e exploração de sistemas *Data Warehouse* na vertente de apoio à tomada de decisão realizando, inicialmente um trabalho de investigação e terminando com a implementação prática de um sistema de informação desenvolvido em *SQL Server* e *Power BI*.

O trabalho será **dividido em duas partes distintas**. A **primeira parte** consiste em uma abordagem teórica, na qual os alunos devem elaborar um relatório de pesquisa que aborde um tema relacionado com Sistemas de Informação sob a perspetiva de um *Data Warehouse*. O tema a ser abordado será selecionado a partir de uma lista disponibilizada no sistema inforestudante.

A **segunda parte do trabalho** envolve a criação de um *Data Warehouse* que ofereça suporte a um gestor empresarial, utilizando o *SQL Server* e *Power BI* como plataforma. O tema será fornecido pelo docente, e incluirá as etapas de criação de dados operacionais, Extração, Transformação e Carregamento (ETL), criação do modelo de dados e apresentação de *dashboards* para apoio à tomada de decisão.

Avaliação

Conforme mencionado na primeira aula, este trabalho contribui com **35% da nota final da unidade curricular**, o que equivale a **sete em vinte valores**. Com o intuito de simplificar a avaliação, e dado que o trabalho está dividido em duas partes, estabelece-se que **a primeira parte, a componente teórica, na qual os alunos devem elaborar um relatório de pesquisa, corresponderá a 20% da nota final**, ou seja, quatro valores. A segunda parte, que envolve a criação e exploração de um *Data Warehouse*, **representará os restantes 15%, equivalendo aos remanescentes três valores**. Em resumo o **trabalho valerá 35%, sendo 20% para a primeira parte e 15% para a segunda parte**.

Os trabalhos são sujeitos a defesa individual, durante a qual será verificada a autoria e conhecimentos, podendo haver mais do que uma defesa caso subsistam dúvidas. A nota final do trabalho é diretamente proporcional à qualidade da defesa. Elementos do mesmo grupo podem ter notas diferentes consoante o desempenho e grau de participação individuais que demonstraram na defesa.

A falta à defesa implica automaticamente a perda da totalidade da nota do trabalho.

Nota importante, no dia da defesa, os alunos devem trazer o software devidamente instalado e a funcionar com o trabalho desenvolvido e entregue.

Plágios e trabalhos feitos por terceiros: o regulamento da escola descreve o que acontece nas situações de fraude.

Data e regras de submissão

- Só existirá a meta final. As datas e requisitos das metas são indicados mais abaixo. A entrega é feita via nónio (inforestudante) através da submissão de um único arquivo zip¹ cujo **nome** respeita o seguinte padrão²:

SI2_2324_tp_nome1_numero1_nome2_numero2_nome3_numero3.zip

(nomes e números 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*.
- A data de entrega é **19 de dezembro pelas 23:55**, não sendo admitidos à posterior
- 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 ao(s) outro(s) aluno(s) do grupo.
- **É necessário que ambos estejam inscritos em turmas práticas (mesmo que seja em turmas diferentes)**

Constituição do Grupo

Os grupos serão constituídos por um máximo de três elementos. O número de elementos do grupo será considerado na avaliação do trabalho. Sempre que se justificar poderão ser desenvolvidos trabalhos individuais. Todos os elementos do grupo devem participar na realização do trabalho e estarem preparados para responder às perguntas que surgirem na avaliação do mesmo.

¹ Leia-se “**zip**” - **não é arj, rar, tar**, ou outros. O uso de outro formato será **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**.

Parte 1 - Relatório de pesquisa

A parte 1 do trabalho final consiste num trabalho de investigação e exploração de um tema relacionado com sistemas de informação na vertente de apoio à decisão. Para esse fim os alunos, devem redigir um relatório, que apresente um estado de arte do tema selecionado para o trabalho. De forma a preparar a defesa, além do relatório, os alunos devem preparar uma apresentação, resumo, do trabalho.


Nos capítulos seguintes, serão apresentados os temas a escolher e a estrutura “mínima” do trabalho a realizar na primeira parte.

Escolha do tema

O tema do relatório de pesquisa deve ser selecionado de uma lista prévia que se encontra na plataforma inforestudante. O mesmo tema poderá ser selecionado, apenas por dois grupos iguais, ou seja, o tema deixará de estar disponível quando tiver sido selecionado duas vezes. No caso de não quererem selecionar um tema previamente sugerido, poderão apresentar um outro tema, no entanto este terá de ser validado previamente pelo docente.

A escolha do tema deve ser realizada obrigatoriamente na plataforma inforestudante até ao fim da aula prática de 10 de novembro.

Temas sugeridos para a parte 1:

1. Arquiteturas e Componentes do Data Warehouse. Definição, comparação e enquadramento num caso de estudo.
2. OLAP e as suas ferramentas.
3. Ferramentas de BI e o seu sucesso/insucesso.
4. Conceito e ferramentas de Data Mining. Porquê utilizar Data Mining?
5. Estado de arte sobre algoritmos de Data Mining.
6. Ciclos e abordagens de desenvolvimento do Data Warehouse. Qual utilizaria?
7. Ferramentas, modelos e abordagens de monitorização do Data Warehouse. Quais as mais apropriadas.
8. Casos de sucesso vs casos de insucesso do Data warehouse.
9. Instalação de um motor de base de dados preparado para Data Warehouse (que não seja o SQL Server).
-  10. Exploração de uma ferramenta de BI (que não seja o Power BI).
11. Utilização de python para realizar ETL.
12. Resumo de metodologias para realização de ETL.

Trabalho a Desenvolver


A avaliação da parte 1 assenta em duas componentes. Um relatório redigido em formato de artigo e uma apresentação no dia da respetiva defesa. O relatório deve ter uma estrutura perfeitamente definida e capaz de esquematizar os assuntos a versar.

Independentemente do tamanho (em páginas) do relatório, a estruturação impõe-se em qualquer situação. Como decerto já devem ter reparado, os modelos dos relatórios, sem imporem uma estrutura pré-definida, incluem sempre a divisão em capítulos e secções, sendo um dos capítulos iniciais a introdução ou

enquadramento, que inclui, além da descrição do problema a solucionar/a investigar, convenientemente descrito, também a sua contextualização. Este capítulo pode ou não incluir referência e descrição breve de outros trabalhos na área que permitam mostrar ao leitor onde pode obter informação adicional no domínio, e eventualmente, depois, efectuar-se uma avaliação comparativa da solução encontrada face a outras. Segue-se, em norma, a descrição do conjunto de conceitos ou teorias que suportam a proposta de solução para o problema em mãos.

Uma última nota! Nunca esquecer que a inclusão de referências é obrigatória e que a existência de figuras, tabelas ou outros elementos adicionais deve ser convenientemente etiquetada e referenciada no texto.

Em termos de capítulo, devem pelo menos incluir, os seguintes capítulos:

- 
- Resumo
 - Introdução
 - Contextualização
 - Estado de arte genérica sobre datawarehouse
 - Análise/descrição tema escolhido
 - Conclusões
 - Referências

Nota - podem ser incluídos outros capítulos que os alunos achem pertinentes

Parte 2 – Criação de uma datawarehouse

A segunda parte do trabalho prático envolve a exploração de ferramentas utilizadas nas aulas práticas, nomeadamente o SQL Server, o Visual Studio na vertente do analysis services e o Microsoft PowerBI, com o objetivo de desenvolver um sistema de data warehouse capaz de implementar um sistema de apoio à decisão para uma empresa que possui várias lojas de retalho.

Para este propósito, no próximo capítulo, descreve-se o cenário desejado, seguido de dez questões-chave que se espera que os alunos resolvam e expliquem. Nesta segunda parte, os alunos devem ainda elaborar um relatório contendo as respostas e as devidas justificações das opções tomadas e dos resultados obtidos.

É crucial salientar que as respostas devem ser sempre justificadas e explicadas, uma vez que respostas diretas não serão consideradas

Tema para a parte 2 - Caso: Cadeia de Lojas de Retalho

Uma cadeia de lojas de retalho tem estabelecimentos lojas em todo o território nacional. Cada uma das lojas é um moderno e típico supermercado com um conjunto completo de departamentos que incluem mercearia, alimentos congelados, frescos, carne, padaria, licores e vinhos, artigos de limpeza, máquinas e eletrodomésticos, maquinaria, artigos para casa e jardim, etc. Cada loja tem aproximadamente 60000 produtos individuais nas suas prateleiras.

Os produtos individuais são chamados SKUs (stock keeping units). Cerca de 40000 destes SKUs provêm de fabricantes externos e têm códigos de barras impressos na embalagem do produto. Esses códigos de barras são denominados UPCs (universal product codes). Os UPCs estão ao mesmo nível de granularidade dos SKUs individuais. Cada embalagem diferente de um produto tem um UPC separado e então é um SKU separado.

Os restantes 20000 SKUs provêm de departamentos como talho, padaria, frescos e não dispõem de códigos UPC reconhecidos a nível nacional. A loja atribui códigos SKU a esses produtos e coloca etiquetas de códigos de barras nesses itens. Embora os códigos de barras não sejam UPCs, são, decerto, códigos SKU.

Os dados são recolhidos em diversos locais na loja, mas principalmente nas caixas registadoras (TPVs), quando os clientes adquirem os produtos. Os TPVs usam scanners para ler os códigos de barras, sendo o local onde as compras dos clientes são medidas. O armazém, onde recebem os produtos dos fabricantes ou outros grandes distribuidores, é outro ponto interessante de recolha de dados.


Na loja, a gestão está especialmente interessada com a logística das encomendas, armazenamento nas prateleiras e vendas dos produtos, maximizando os lucros em cada loja. O lucro final provém de sobrecarregar tanto quanto possível cada produto, baixando os custos de aquisição e sobretaxas, e, ao mesmo tempo, atrair o máximo de clientes possível.

Algumas das decisões de gestão mais importantes têm a ver com preços e promoções. Quer a gestão da loja, quer os escritórios centrais, passam muito tempo a lidar com os preços e lançar promoções. As promoções numa loja incluem reduções temporárias de preços, cupões em jornais e cartazes na loja. A forma mais direta e eficaz de criar um grande aumento nas vendas de produtos é baixar o preço de forma dramática. Mas, infelizmente, uma tão grande redução de preço não é sustentável, pois que, provavelmente estarão a ser vendidos a um preço abaixo do preço do custo. Por esse motivo, a visibilidade de todas as formas de promoção é uma parte importante das operações de análise realizadas na loja.

Tarefas propostas: Fazer todas

1. Selecionar o processo de negócio a modelar. É certo que o primeiro modelo a construir será aquele que tenha o maior impacto – deve responder às questões de negócio mais prementes e cujos dados sejam imediatamente disponíveis.
2. Avaliar a granularidade: qual o nível de detalhe mais fino que será disponibilizado no modelo dimensional.
3. Selecionar as dimensões e os atributos que serão incluídas no modelo dimensional.
4. Selecionar os factos que serão incluídos no modelo dimensional
5. Construir o modelo em estrela concebido, incluindo os atributos que considerar mais significativos. Deve também elaborar um dicionário de dados que inclua também explicação sobre as relações entre tabelas.
6. Efetue cálculos aproximados do tamanho que ocupará o modelo concebido. Suponha que as condições seguintes serão observadas: terá de registar dados de 5 anos e a cadeia de supermercados tem 25 lojas. Só 5000 produtos, em média, são vendidos em cada dia e, das 3000 combinações de promoções, só 1, em média, ocorre em cada dia. No geral, considere-se que, em média são emitidos, em cada loja, 1000 talões com 20 linhas cada.
7. Implemente o modelo em estrela no SQL Server
8. Introduza valores no modelo em estrela do SQL, simulando uma operação de ETL. Os dados devem ser criados pelos alunos.
9. Apresentar um conjunto de cinco vistas no Visual Studio Data Warehouse de apoio a um gestor de topo. Justificar a escolha de cada vista de acordo com o processo de negócio.
10. Apresentar um dashboard no Power BI que utilize o modelo em estrela criado anteriormente. O dashboard deve apresentar, pelo menos, 5 vistas diferenciadas. Justificar a escolha de cada vista de acordo com o processo de negócio.

Notas importantes:

- 
- Deve realizar um relatório que explique todas as opções tomadas
 - No dia da defesa, devem trazer o sistema preparado e funcional de forma a poderem mostrar o trabalho realizado.