

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 que devem ser entregues em duas metas, com apresentações/defesas diferentes.**

A **primeira parte** consiste num estudo teórico, que só poderá ser realizada por frequência, 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 universitário, 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 e uma apresentação, corresponderá a 15% da nota final**, ou seja, três valores. A segunda parte, que envolve a criação e exploração de um *Data Warehouse*, **representará os restantes 20%, equivalendo aos remanescentes quatro valores**. Em resumo o **trabalho valerá 35%, sendo 15% para a primeira parte e 20% 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

- Existiram duas metas. 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 da primeira meta é **9 de novembro pelas 23:55**, não sendo admitidas entregas à posterior (Valido para todas as épocas)
 - As apresentações desta primeira meta é obrigatória e será realizadas nas aulas teóricas seguintes mediante inscrição em slots.
- A data de entrega é da segunda meta é **7 de dezembro pelas 23:55**, não sendo admitidos não sendo admitidas entregas à posterior (Valido para todas épocas de frequência)
- 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 e realizarem uma apresentação em aula.

A apresentação deverá ter um tempo de 30 minutos e será apresentado nas aulas teóricas da UC.

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 17 de novembro.

Temas sugeridos para a parte 1:

1. Arquiteturas e Componentes do Data Warehouse. Definição, comparação e enquadramento num caso de estudo.
2. Arquiteturas em nuvem para Data Warehouses: Comparação detalhada entre os principais (AWS, Azure, GCP) e seus serviços específicos para Data Warehouses.
3. Data Warehouses distribuídos: Análise de tecnologias como Apache Hadoop e Spark, e seus impactos na escalabilidade e desempenho.
4. Impacto do Regulamento Geral sobre a Proteção de Dados nos Data Warehouses: Analisar como as empresas estão a adaptar seus Data Warehouses para cumprir as exigências da GDPR, incluindo a gestão de consentimento, anonimização de dados e segurança.
5. O papel do Data Warehouse na transformação digital das empresas portuguesas.
6. Data Warehouses no setor da saúde: Analisar o uso de Data Warehouses para melhorar a gestão de dados clínicos, otimizar processos e apoiar a tomada de decisões na área da saúde.
7. Data Warehouses no setor financeiro: Investigar como os bancos e outras instituições financeiras utilizam Data Warehouses para gerenciar riscos, detectar fraudes e personalizar produtos e serviços.
8. OLAP e as suas ferramentas.
9. Ferramentas de BI e o seu sucesso/insucesso.
10. Conceito e ferramentas de Data Mining. Porquê utilizar Data Mining?
11. Estado de arte sobre algoritmos de Data Mining.
12. Ciclos e abordagens de desenvolvimento do Data Warehouse. Qual utilizaria?
13. Ferramentas, modelos e abordagens de monitorização do Data Warehouse. Quais as mais apropriadas.
14. Casos de sucesso vs casos de insucesso do Data warehouse.
15. Instalação de um motor de base de dados preparado para Data Warehouse (que não seja o SQL

Server).

16. Exploração de uma ferramenta de BI (que não seja o Power BI).
17. Integração de Fontes de Dados Externas em Power BI: Desafios e Soluções – Exemplo prático.
18. Utilização de python para realizar ETL.
19. Python para análise de dados.
20. Resumo e comparação de metodologias para realização de ETL..
21. Apresentação e exploração da Linguagem DAX (Data Analysis Expressions).
22. Criação e Monitorização de KPIs com DAX – Um exemplo prático.
23. Casos de Sucesso e Insucesso - Data Warehouses em setores específicos: Estudo de casos de sucesso e insucesso em setores na saúde.
24. Casos de Sucesso e Insucesso - Data Warehouses em setores específicos: Estudo de casos de sucesso e insucesso na indústria.
25. Inteligência Artificial e Data Mining: Como a IA está transformando o campo do Data Mining, com foco em técnicas como aprendizado de máquina e deep learning
26. Lessons learned de projetos de Data Warehouse: Análise de projetos de Data Warehouse que falharam e identificação das principais causas de fracasso.

Trabalho a Desenvolver

A avaliação da parte 1 assenta em duas componentes. Um relatório redigido em formato de artigo e uma apresentação em aula.

O relatório deve ter uma estrutura perfeitamente definida e capaz de esquematizar os assuntos a versar.

A apresentação em aula terá de ter uma duração compreendida entre 20 e 30 minutos. Todos os elementos devem apresentar e devem responder a perguntas individuais.

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*
- Analise/descrição tema escolhido
- Conclusões
- Referências

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

Trabalho Prático – Parte 2: Desenvolvimento de um Sistema de Apoio à Decisão com Data Warehouse

A universidade em foco possui cerca de 20 mil alunos, distribuídos por 30 cursos de licenciatura que abrangem áreas tão diversas como gestão, saúde, tecnologia, letras e desporto.

A intenção da instituição é criar um sistema que permita uma compreensão profunda e uma monitorização eficaz do desempenho académico dos seus estudantes. O objetivo principal é desenvolver um sistema que facilite a tomada de decisões estratégicas, visando a melhoria contínua do ensino e o sucesso dos alunos.

Para alcançar este objetivo, o *Data Warehouse* deverá integrar e processar dados provenientes de várias fontes dentro da universidade, incluindo bases de dados relacionais e ficheiros CSV.

Esta abordagem abrangente oferecerá uma visão dos diversos percursos académicos dos estudantes e quais os cursos com maior rentabilidade.

Os alunos deverão conceber um modelo dimensional que capture adequadamente as complexidades do ambiente académico. Este modelo deve incluir, no mínimo, informações sobre os alunos (incluindo dados demográficos e socioeconómicos), cursos, disciplinas, matrículas e situação financeira e sobre os diversos cursos (maior adesão, maior sucesso de saída).

É crucial que o sistema permita análises detalhadas por diferentes períodos, como semestres e anos letivos dos últimos cinco anos.

Uma componente fundamental do sistema será a sua capacidade de identificar padrões associados ao sucesso ou insucesso académico. Isto implica considerar variáveis como o regime de frequência (tempo inteiro ou parcial) e as condições socioeconómicas dos estudantes. Além disso, o sistema deverá monitorizar eficazmente o estado das matrículas, identificando estudantes com matrículas pendentes ou não concluídas. A análise da situação financeira dos alunos é outro aspeto crucial. O sistema deve ser capaz de revelar informações sobre propinas em atraso ou apoios financeiros recebidos.

Esta funcionalidade permitirá à universidade identificar padrões que possam influenciar o sucesso académico, proporcionando a oportunidade de ajustar políticas de apoio e intervenção de forma mais precisa e eficaz.

Para além da monitorização do desempenho individual, o sistema oferecerá à universidade uma visão mais ampla das tendências e padrões que afetam o sucesso académico em geral. Isto possibilitará à instituição tomar decisões mais informadas sobre estratégias pedagógicas, distribuição de recursos e medidas de apoio. Consequentemente, a universidade não só promoverá o sucesso dos seus alunos, como também melhorará a qualidade dos seus cursos e aumentará a eficiência no acompanhamento académico. É fundamental que a solução desenvolvida ofereça uma plataforma robusta para a exploração de dados e geração de relatórios. O resultado deve ser um sistema completo e funcional que permita à administração da universidade explorar os dados de forma intuitiva e eficiente. Os dashboards e relatórios criados no Power BI e Microsoft Analysis Services devem oferecer visualizações claras e informativas, com capacidade de drill-down para análises mais detalhadas. O sistema final deve ser capaz de responder a questões complexas sobre o desempenho académico, auxiliando na identificação de tendências, desafios e oportunidades de melhoria nos processos educativos da instituição. Esta capacidade analítica avançada será crucial para orientar as decisões estratégicas da universidade, promovendo um ambiente académico mais eficiente e centrado no sucesso dos estudantes

Tarefas propostas:

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 existem cerca de 30 cursos. Contemple apenas cursos de licenciatura (3 anos). Suponha que uma turma só pode ter 40 alunos no máximo.
7. Implemente o modelo em estrela no SQL Server
8. Introduza valores no modelo em estrela do SQL, simulando uma operação de ETL com base em dados de uma base de dados SQL e de um ficheiro csv. Os dados devem ser criados pelos alunos.
9. Apresentar 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. 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.