



Disciplina: Engenharia de Dados

Profa. Dra. Angélica Félix de Castro

Orientações para Projeto da 3ª unidade – Vale 10,0

- 1. Você pode usar qualquer ferramenta tecnológica que quiser: tanto para fazer o ETL quanto para fazer o Dashboard final**
- 2. Escolha um tema qualquer (pode ser na área de educação, saúde, negócios, esporte...)**
- 3. Levantamento de Necessidades e Definição dos objetivos: Finja ser o dono ou gestor desse tema e pense em 3 perguntas analíticas, que serão respondidas com análise de dados.**
- 4. Requisitos da Informação: Faça um ou mais modelos estrela para responder as perguntas acima.**
- 5. Obtenção das fontes de dados e Staging Area: obtenha fontes de dados distintas com dados relacionados ao assunto escolhido. Tem que haver dados nesses 4 formatos:**
 - **Tabela simples – equivalente a um BD transacional**
 - **Arquivo .CSV**
 - **Arquivo .XML**
 - **Um JSON**

6. Após coletar os dados do item anterior, fazer a limpeza e preparação deles. Alteração dos dados de acordo com as regras de negócios. Existem algumas funções de pré-processamento e talvez você precise usar essas funções. Por exemplo:
 - Padronizar informações. Ex.: Sexo: Masculino/Feminino, M/F, 1/2)
 - Agregar dados. Ex.: agrupar totais de venda por estado
 - Deletar informações nulas ou que não são uteis nesse momento
 - Demais operações que forem necessárias
7. Planejamento: Crie uma estrutura (pode ser tabela ou Data Lake ou Data Warehouse) que consiga responder as questões pensadas no item 3.
8. Aplicar processamento em batch (hora marcada para acontecer) ou em stream (processamento em tempo real).
9. Fazer um relatório final com tudo que você fez desde o início até agora. Uma nova engenheira de dados vai entrar na empresa (Angélica) e ela precisa entender todo o processo que você fez!
10. Fazer um dashboard para mostrar ao seu chefe ou seu gestor, respondendo todas as perguntas que ele fez no item 3.