

**Desenvolvimento de um Projeto de Integração de Dados (ETL)
1/2024**

Grupo: o trabalho deve ser desenvolvido em grupo de acordo com a divisão da disciplina Projeto em Ciência de Dados II

Valor: 20 pontos

Observações:

- 1) atenção às entregas parciais para cada uma das etapas.
- 2) o envolvimento de cada aluno no desenvolvimento do projeto será considerado para a atribuição de sua nota.
- 3) nas apresentações de cada fase, cada integrante do grupo deve participar, ativamente, explicando partes do projeto. Alunos ausentes ou que não participarem da apresentação terão seus pontos reduzidos.

O desenvolvimento de um projeto de ETL é um processo complexo que envolve pessoas, atividades e tecnologias. Para permitir a vivência desta complexidade, este trabalho propõe a execução deste projeto, conforme as instruções a seguir.

Cada grupo deverá definir um negócio e estabelecer uma demanda de desenvolvimento de um projeto de integração de dados, que atenda a especificidade do negócio. Para isto, o projeto deverá ser executado em etapas (fases) compostas de Reuniões de Trabalho de acompanhamento e de apresentação dos resultados de cada fase. O cronograma é apresentado a seguir.

Fase	Reunião de Trabalho	Apresentação
1 - Definição do negócio, do problema e da demanda a ser atendida	09/04/2024	
2 - Identificação da base de dados e da arquitetura de ETL	16/04/2024	23/04/2024
3 - Modelagem Dimensional e Implementação do DW	02/05/2024	09/05/2024
4 - Implementação do processo de ETL	16/05/2024	28/05/2024
5 - Homologação dos dados	06/06/2024	13/06/2024
6 - Apresentação do Projeto		25/06/2024

Fase 1: Definição do negócio, do problema e da demanda a ser atendida

Você foi contratado por uma organização para desenvolver uma solução de integração para um processo de BI.

Nesta fase, defina o cliente e descreva, brevemente, o problema de negócio que ele objetiva resolver. Não deve ser descrito como o problema será abordado, apenas qual é o problema a ser enfrentado, apresentando o cenário atual, a(s) área(s) a ser atendida(s) e os requisitos de dados/informação.

Data da Reunião de Trabalho: 09/04/2024 (terça-feira)

Fase 2: Identificação da base de dados e da arquitetura de ETL

Nesta fase, deve ser identificada a base de dados que será utilizada no Projeto e a definição da arquitetura do processo de ETL, que levará os dados de sua origem ao repositório final (DW). Informe as fontes de dados, o formato em que estão disponíveis, o dicionário dos dados e a maneira como eles estão disponibilizados.

Apresente a arquitetura de ETL, indicando as ferramentas que serão utilizadas para acessar e manipular os dados, as integrações necessárias e o tratamento dos dados, bem como o SGBD que será utilizado para a implementação do *Data Warehouse* (DW).

Elabore, também, um diagrama que represente esta arquitetura.

Data da Reunião de Trabalho: 16/04/2024 (terça-feira)

Data da Apresentação da Fase 2: 23/04/2024 (terça-feira)

Fase 3: Modelagem Dimensional e Implementação do DW

Nesta fase, elabore o modelo dimensional do DW. Defina as métricas que responderão as perguntas sobre o problema/negócio do cliente. A partir delas, defina as tabelas fato e dimensão, especificando as chaves primárias e estrangeiras.

Em seguida, implemente o modelo físico, apresentando os scripts SQL utilizados.

Data da Reunião de Trabalho: 02/05/2024 (quinta-feira)

Data da Apresentação da Fase 3: 09/05/2024 (quinta-feira)

Fase 4: Implementação do processo de ETL

Nesta fase, implemente o processo de ETL descrevendo as etapas de extração, transformação e carga. Disponibilize os códigos elaborados em um repositório e compartilhe no Projeto.

Registre, com imagens, as etapas de execução.

Data da Reunião de Trabalho: 16/05/2024 (quinta-feira)

Data da Apresentação da Fase 4: 28/05/2024 (terça-feira)

Fase 5: Homologação dos dados

Nesta fase, o objetivo é mostrar que os dados que chegaram ao DW são equivalentes àqueles disponíveis na base de dados original. Para isso, escolha algumas amostras dos dados com valores agregados comparando os seus valores a partir dos dados de origem e aqueles que estão no DW.

Aqui pode ser utilizada uma ferramenta de visualização.

Data da Reunião de Trabalho: 06/06/2024 (quinta-feira)

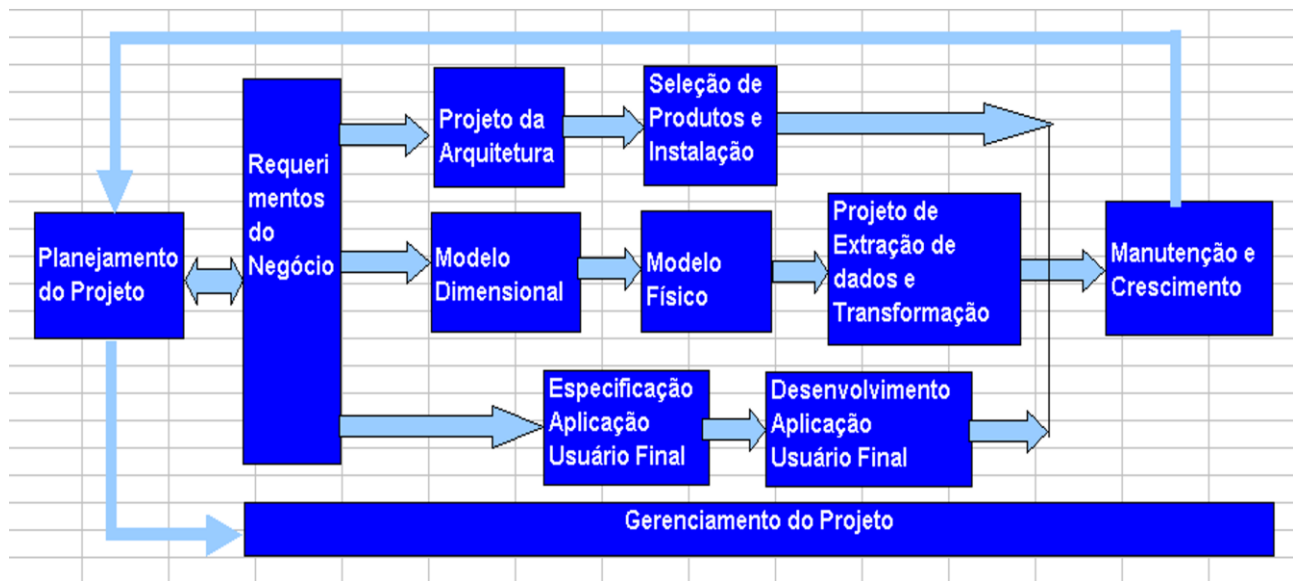
Data da Apresentação da Fase 5: 13/06/2024 (quinta-feira)

Fase 6: Apresentação da solução de Integração

Nesta fase, o grupo deve apresentar o projeto desenvolvido, descrevendo o seu percurso e demonstrando, se possível, a sua execução.

Data da Apresentação do Projeto Final: 25/06/2024 (terça-feira)

O Ciclo de Vida Kimball



Grupo	Alunos
1	André Luiz Santos Moreira da Silva Carla Luiza de Oliveira Daniel Rocha Franca Lígia Ferreira de Carvalho Gonçalves
2	Patrick Leandro da Silva Pedro Henrique Azevedo de Medeiros Pedro Henrique Rodrigues da Silva
3	Marco Tulio Gontijo de Sousa Paulo Victor Bonomo Nascimento Rafael Romualdo Pinto Rodrigues

Bruno Petrocchi de Sena Azevedo