

Título: Monitoramento da Abundância de Indivíduos

O presente roteiro descreve o monitoramento da abundância de indivíduos durante as etapas de “implantação” e “operação” de um projeto sobre espécies de animais de uma determinada região do país. O objetivo é coletar dados sobre a presença e quantidade de indivíduos de cada espécie em diferentes datas e horários, visando obter informações relevantes para a compreensão da dinâmica populacional e a avaliação do impacto.

1. Instale o Pentaho Data Integration em um ambiente apropriado.
2. Importe o conjunto de dados fornecido pela Brintell.
3. Crie um novo projeto no Pentaho Data Integration.
4. Adicione uma nova transformação ao projeto.
5. Adicione uma etapa de leitura de dados para ler o conjunto de dados fornecido.
6. Configure os parâmetros da etapa de leitura para acessar o arquivo CSV.
7. Adicione etapas de transformação necessárias para processar os dados, como filtragem, ordenação, limpeza ou agregação.
8. Utilize a funcionalidade de agrupamento para criar uma coluna com a soma da ABUNDANCIA_INDIVIDUOS por CAMPANHA, ETAPA e MODULO. Mostre o resultado do cálculo sem ocultar as demais colunas.
9. Faça o mesmo para agrupar por NOME_CIENTIFICO.
10. Crie duas colunas para mostrar a data de registro mais antiga e a data de registro mais nova, respectivamente.
11. Cria uma coluna booleana para mostrar 1 quando o NOME_CIENTIFICO for igual a LIPAUGUS VOCIFERANS e 0 quando for outro.
12. Realize as transformações nos campos de data e hora, transformando-os em somente 1 campo do tipo Timestamp.
13. Transforme os registros da coluna PERIODO_SAZONAL para minúsculo e retire os espaços das extremidades.
14. Mostre quantos dias distintos existem no campo DATA_REGISTRO.
15. Faça um cruzamento dos IDs dos dois arquivos fornecidos para acrescentar a coluna STATUS_EXTINCAO.
16. Inclua uma coluna contendo a data de execução da transformation.
17. Adicione uma etapa de carga para gravar os dados transformados em um banco de dados de sua preferência.
18. Defina o ID como chave primária.
19. Execute a transformação e verifique se os dados são processados corretamente.
20. Verifique se os dados são carregados com sucesso no destino definido.
21. Realize consultas ou análises nos dados transformados para confirmar a precisão e integridade das informações.
22. Crie um job e agende para executar de hora em hora.

23. Avalie a performance da transformação em termos de tempo de execução e recursos utilizados.
24. Identifique possíveis melhorias ou otimizações no processo de ETL.
25. Analise a escalabilidade do processo para lidar com grandes volumes de dados.
26. Documente o processo de criação da transformação, incluindo as etapas realizadas, configurações, transformações aplicadas e qualquer informação relevante.