JORNADA FINANÇAS PESSOAIS

<DATA MAPPING - EXTRAÇÃO DOS DADOS - B3>

<Elaborado por: Flávio Cavalcante Uyemura>

<Data: 21/11/2023>Sumário

Sumário

[**1.** **Sumário Executivo** 3](#_Toc421785590)

[**1.1.** **Objetivo** 3](#_Toc421785591)

[**1.2.** **Escopo** 3](#_Toc421785592)

[**2.** **Fluxo Processo ETL** 4](#_Toc421785593)

[**2.1.** **AS IS** 4](#_Toc421785594)

[**2.2.** **TO BE** 4](#_Toc421785595)

[**3.** **Requisitos** 4](#_Toc421785596)

[**3.1.** **Requisitos de Negócio** 4](#_Toc421785597)

[**3.2.** **Requisitos de Segurança** 5](#_Toc421785598)

[**3.3.** **Requisitos de Desenvolvimento ETL** 5](#_Toc421785599)

[**4.** **Servidores, Bancos de Dados e Acessos** 5](#_Toc421785600)

[**4.1.** **Servidores Envolvidos** 5](#_Toc421785601)

[**4.1.1.** **Bancos de dados** 5](#_Toc421785602)

[**4.1.2.** **File System** 6](#_Toc421785603)

[**4.2.** **Informação para acesso aos dados / servidores pelo ETL** 6](#_Toc421785604)

[**5.** **ETL Mapping** 7](#_Toc421785605)

[**5.1.** **Integração de Empresas/Filiais** 7](#_Toc421785606)

[**5.2.** **Integração de Professores** 8](#_Toc421785607)

[**5.3.** **Integração de Atividades dos Professores** 9](#_Toc421785608)

[**5.4.** **Confirmação de Integração** 9](#_Toc421785609)

[**6.** **Glossário** 11](#_Toc421785610)

[**7.** **Anexos** 11](#_Toc421785611)

[**7.1.** **Regras para configuração de Leitura / Escrita para arquivos de Texto** 11](#_Toc421785612)

[**7.2.** **Identificação de atualizações do campo na versão atual do Mapping** 11](#_Toc421785613)

[**7.3.** **Regras de transformação para o ETL Mapping** 11](#_Toc421785614)

[**7.4.** **Estratégias de Carga** 12](#_Toc421785615)

[**8** **Histórico de Alteração do Documento** 12](#_Toc421785616)

[9 Aprovação 12](#_Toc421785617)

1. **Sumário Executivo**
   1. **Objetivo**

O objetivo é fazer um processo de ETL para carregar os dados diários gerados das cotações dos ativos da B3.

* 1. **Escopo**Desenvolver um processo de ETL para carregar os dados do arquivo COTAHIST.AAAA.TXT. O foco é simular no On-Premise a arquitetura de medalhão da Cloud Computing. Mais informações sobre esta arquitetura podem ser consultadas em <https://www.dataside.com.br/dataside-community/big-data/afinal-o-que-e-a-arquitetura-medalhao>

1. **Fluxo Processo ETL**
   1. **AS IS**

ETL

DESTINO

ORIGEM

* 1. **TO BE**

1. **Requisitos**
   1. **Requisitos de Negócio**

N/A.

* 1. **Requisitos de Segurança**

N/A

* 1. **Requisitos de Desenvolvimento ETL**

Fazer o download, instalação e configuração das seguintes ferramentas:

|  |  |
| --- | --- |
| Ferramentas | Download |
| MySQL Community | <https://www.mysql.com/products/community/> |
| Python | <https://www.python.org/downloads/> |
| VS Code | <https://code.visualstudio.com/download> |
| Pycharm Community | <https://www.jetbrains.com/pycharm/download/?section=windows> |
| GIT | <https://git-scm.com/downloads> |

1. Criar conta no GitHub (<https://github.com/>)
2. Fazer o clone do repositório Jornada Finanças Pessoais (<https://github.com/fuyemura/Jornada_financas_pessoais.git>)

Utilizar os seguintes arquivos/documentos:

|  |  |
| --- | --- |
| Arquivos/Documentos | Observações |
| SeriesHistoricas\_Layout.pdf | Possui informações detalhas sobre o arquivo COTAHIST.AAAA.TXT |
| COTAHIST.2023.TXT | Arquivo para testar o processo de ETL |
| Jornada\_financas\_pessoais.mwb | Arquivo com a estrutura das tabelas e schemas do MySQL |

1. **Servidores, Bancos de Dados e Acessos** 
   1. **Servidores Envolvidos**
      1. **Bancos de dados**

|  |  |  |
| --- | --- | --- |
|  | | |
| **DESENVOLVIMENTO** | | **PRODUÇÃO** |
| ID\_Servidor |  |  |
| Tipo Banco | MySQL | MySQL |
| Host | Localhost | localhost |
| Banco |  |  |
| Porta |  |  |

* + 1. **File System**

|  |  |  |
| --- | --- | --- |
|  | | |
| **DESENVOLVIMENTO** | | **PRODUÇÃO** |
| ID\_Servidor |  |  |
| Host | localhost | localhost |
| Pasta | \jornada\_financas\_pessoais\data | \jornada\_financas\_pessoais\data |
| Delimitador | ( X ) Separador ( )Largura Fixa | ( X ) Separador ( )Largura Fixa |
| Delimitador |  |  |

* 1. **Informação para acesso aos dados / servidores pelo ETL**

1. **ETL Mapping** 
   1. **Mapeamento de carga**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Processo(s)** | | Extracao\_cotahist\_b3.py | | | | | | | | | |  | | | | | | | | |
|  | **Descrição** | | Processo de extração dos dados das cotações históricas da B3 | | | | | | | | | |  | | | | | | | | |
|  | **Dependência** | |  | | | | | | | | | |  | | | | | | | | |
|  | **Source** | | COTAHIST.AAAA.TXT | | | | | | | |  |  |  | | | |  |  |  |  |  |
|  | **Source Owner** | |  | | | | | | | |  |  |  | | | |  |  |  |  |  |
|  | **Target** | | COTAHIST | | | | | | | |  |  |  | | | |  |  |  |  |  |
|  | **Target Owner** | | BRONZE | | | | | | | |  |  |  | | | |  |  |  |  |  |
|  |  |  |  |  | | |  | | | |  |  |  | | | |  |  |  |  |  |
|  | **Mapping**  **(I, A, E)** | **EXTRACT  ( Source )** | | | | | | | | **TRANSFORMATION** | | | **LOAD  ( Target )** | | | | | | | | **Comentários** |
|  | **Campo** | | | **Tipo e Tamanho** | **Formato** | | **Aceita Nulo** | **Chave** | **Transformações, Qualidade ou Exceções** | | | **Campo** | **Tipo e Tamanho** | **Formato** | **Aceita Nulo** | | | | **Chave** |  |
|  |
|  |  | tipo\_registro | | |  |  | |  |  | Mapeamento direto | | | tipo\_registro |  |  |  | | | |  |  |
|  |  | data\_pregao | | |  |  | |  |  | Mapeamento direto | | | data\_pregao |  |  |  | | | |  |  |
|  |  | codigo\_bdi | | |  |  | |  |  | Mapeamento direto | | | codigo\_bdi |  |  |  | | | |  |  |
|  |  | codigo\_negociacao | | |  |  | |  |  | Mapeamento direto | | | codigo\_negociacao |  |  |  | | | |  |  |
|  |  | tipo\_mercado | | |  |  | |  |  | Mapeamento direto | | | tipo\_mercado |  |  |  | | | |  |  |
|  |  | nome\_resumido\_empresa | | |  |  | |  |  | Mapeamento direto | | | nome\_resumido\_empresa |  |  |  | | | |  |  |
|  |  | especificacao\_papel | | |  |  | |  |  | Mapeamento direto | | | especificacao\_papel |  |  |  | | | |  |  |
|  |  | prazo\_dias\_mercado | | |  |  | |  |  | Mapeamento direto | | | prazo\_dias\_mercado |  |  |  | | | |  |  |
|  |  | moeda\_referencia | | |  |  | |  |  | Mapeamento direto | | | moeda\_referencia |  |  |  | | | |  |  |
|  |  | preco\_abertura\_papel | | |  |  | |  |  | Mapeamento direto | | | preco\_abertura\_papel |  |  |  | | | |  |  |
|  |  | preco\_maximo\_papel | | |  |  | |  |  | Mapeamento direto | | | preco\_maximo\_papel |  |  |  | | | |  |  |
|  |  | preco\_minimo\_papel | | |  |  | |  |  | Mapeamento direto | | | preco\_minimo\_papel |  |  |  | | | |  |  |
|  |  | preco\_medio\_papel | | |  |  | |  |  | Mapeamento direto | | | preco\_medio\_papel |  |  |  | | | |  |  |
|  |  | preco\_ultimo\_negocio | | |  |  | |  |  | Mapeamento direto | | | preco\_ultimo\_negocio |  |  |  | | | |  |  |
|  |  | preco\_melhor\_oferta\_compra | | |  |  | |  |  | Mapeamento direto | | | preco\_melhor\_oferta\_compra |  |  |  | | | |  |  |
|  |  | preco\_melhor\_oferta\_venda | | |  |  | |  |  | Mapeamento direto | | | preco\_melhor\_oferta\_venda |  |  |  | | | |  |  |
|  |  | numero\_negocios\_efetuados | | |  |  | |  |  | Mapeamento direto | | | numero\_negocios\_efetuados |  |  |  | | | |  |  |
|  |  | quantidade\_total\_titulos | | |  |  | |  |  | Mapeamento direto | | | quantidade\_total\_titulos |  |  |  | | | |  |  |
|  |  | volume\_total\_titulos | | |  |  | |  |  | Mapeamento direto | | | volume\_total\_titulos |  |  |  | | | |  |  |
|  |  | preco\_exercicio\_opcoes | | |  |  | |  |  | Mapeamento direto | | | preco\_exercicio\_opcoes |  |  |  | | | |  |  |
|  |  | indicador\_correcao\_precos | | |  |  | |  |  | Mapeamento direto | | | indicador\_correcao\_precos |  |  |  | | | |  |  |
|  |  | data\_vencimento\_opcoes | | |  |  | |  |  | Mapeamento direto | | | data\_vencimento\_opcoes |  |  |  | | | |  |  |
|  |  | fator\_cotacao\_papel | | |  |  | |  |  | Mapeamento direto | | | fator\_cotacao\_papel |  |  |  | | | |  |  |
|  |  | preco\_exercicio\_pontos | | |  |  | |  |  | Mapeamento direto | | | preco\_exercicio\_pontos |  |  |  | | | |  |  |
|  |  | codigo\_papel\_sistema | | |  |  | |  |  | Mapeamento direto | | | codigo\_papel\_sistema |  |  |  | | | |  |  |
|  |  | numero\_distribuicao\_papel | | |  |  | |  |  | Mapeamento direto | | | numero\_distribuicao\_papel |  |  |  | | | |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Processo(s)** | | Carga\_cotacoes\_historicas\_b3.py | | | | | | | | | |  | | | | | | | | |
|  | **Descrição** | | Processo de carga dos dados das cotações históricas da B3 | | | | | | | | | |  | | | | | | | | |
|  | **Dependência** | |  | | | | | | | | | |  | | | | | | | | |
|  | **Source** | | COTAHIST | | | | | | | |  |  |  | | |  |  |  | |  |  |
|  | **Source Owner** | |  | | | | | | | |  |  |  | | |  |  |  | |  |  |
|  | **Target** | | COTACOES\_HISTORICAS | | | | | | | |  |  |  | | |  |  |  | |  |  |
|  | **Target Owner** | | SILVER | | | | | | | |  |  |  | | |  |  |  | |  |  |
|  |  |  |  |  | | |  | | | |  |  |  | | |  |  |  | |  |  |
|  | **Mapping**  **(I, A, E)** | **EXTRACT  ( Source )** | | | | | | | | **TRANSFORMATION** | | | **LOAD  ( Target )** | | | | | | | | **Comentários** |
|  | **Campo** | | | **Tipo e Tamanho** | **Formato** | | **Aceita Nulo** | **Chave** | **Transformações, Qualidade ou Exceções** | | | **Campo** | **Tipo e Tamanho** | **Formato** | | | | **Aceita Nulo** | **Chave** |  |
|  |
|  |  | tipo\_registro | | |  |  | |  |  | Mapeamento direto | | | tipo\_registro |  |  | | | |  |  |  |
|  |  | data\_pregao | | |  |  | |  |  | Converter para data com o formato ‘YYYYMMDD’ | | | data\_pregao |  |  | | | |  |  |  |
|  |  | codigo\_bdi | | |  |  | |  |  | Mapeamento direto | | | codigo\_bdi |  |  | | | |  |  |  |
|  |  | codigo\_negociacao | | |  |  | |  |  | Mapeamento direto | | | codigo\_negociacao |  |  | | | |  |  |  |
|  |  | tipo\_mercado | | |  |  | |  |  | Mapeamento direto | | | tipo\_mercado |  |  | | | |  |  |  |
|  |  | nome\_resumido\_empresa | | |  |  | |  |  | Mapeamento direto | | | nome\_resumido\_empresa |  |  | | | |  |  |  |
|  |  | especificacao\_papel | | |  |  | |  |  | Mapeamento direto | | | especificacao\_papel |  |  | | | |  |  |  |
|  |  | prazo\_dias\_mercado | | |  |  | |  |  | Mapeamento direto | | | prazo\_dias\_mercado |  |  | | | |  |  |  |
|  |  | moeda\_referencia | | |  |  | |  |  | Mapeamento direto | | | moeda\_referencia |  |  | | | |  |  |  |
|  |  | preco\_abertura\_papel | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | preco\_abertura\_papel |  |  | | | |  |  |  |
|  |  | preco\_maximo\_papel | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | preco\_maximo\_papel |  |  | | | |  |  |  |
|  |  | preco\_minimo\_papel | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | preco\_minimo\_papel |  |  | | | |  |  |  |
|  |  | preco\_medio\_papel | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | preco\_medio\_papel |  |  | | | |  |  |  |
|  |  | preco\_ultimo\_negocio | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | preco\_ultimo\_negocio |  |  | | | |  |  |  |
|  |  | preco\_melhor\_oferta\_compra | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | preco\_melhor\_oferta\_compra |  |  | | | |  |  |  |
|  |  | preco\_melhor\_oferta\_venda | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | preco\_melhor\_oferta\_venda |  |  | | | |  |  |  |
|  |  | numero\_negocios\_efetuados | | |  |  | |  |  | Converter para numérico | | | numero\_negocios\_efetuados |  |  | | | |  |  |  |
|  |  | quantidade\_total\_titulos | | |  |  | |  |  | Converter para numérico | | | quantidade\_total\_titulos |  |  | | | |  |  |  |
|  |  | volume\_total\_titulos | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | volume\_total\_titulos |  |  | | | |  |  |  |
|  |  | preco\_exercicio\_opcoes | | |  |  | |  |  | Converter para numérico com 2 casas decimais | | | preco\_exercicio\_opcoes |  |  | | | |  |  |  |
|  |  | indicador\_correcao\_precos | | |  |  | |  |  | Mapeamento direto | | | indicador\_correcao\_precos |  |  | | | |  |  |  |
|  |  | data\_vencimento\_opcoes | | |  |  | |  |  | Converter para data com o formato ‘YYYYMMDD’ | | | data\_vencimento\_opcoes |  |  | | | |  |  |  |
|  |  | fator\_cotacao\_papel | | |  |  | |  |  | Mapeamento direto | | | fator\_cotacao\_papel |  |  | | | |  |  |  |
|  |  | preco\_exercicio\_pontos | | |  |  | |  |  | Converter para numérico com 6 casas decimais | | | preco\_exercicio\_pontos |  |  | | | |  |  |  |
|  |  | codigo\_papel\_sistema | | |  |  | |  |  | Mapeamento direto | | | codigo\_papel\_sistema |  |  | | | |  |  |  |
|  |  | numero\_distribuicao\_papel | | |  |  | |  |  | Mapeamento direto | | | numero\_distribuicao\_papel |  |  | | | |  |  |  |

1. **Evidências**
2. **Anexos**
   1. **Regras para configuração de Leitura / Escrita para arquivos de Texto**

|  |  |
| --- | --- |
| **Delimitador** | **Racional** |
| Separado por Caractere | Indica que a leitura/gravação de um arquivo de texto deve ser através de um Caractere Delimitador especifico, por exemplo “ ; “, cada ocorrência deste caractere no arquivo indica uma nova coluna de dados. |
| Largura Fixa | Indica que a leitura/gravação de um arquivo de texto deve ser através de largura fixa, neste caso deve ser preenchida a largura no campo “Tipo e Tamanho” do ETL Mapping informando a coluna inicial e final correspondente ao campo Ex.: Campo1 de 1 a 10, Campo2 de 11 a 30, etc |

* 1. **Identificação de atualizações do campo na versão atual do Mapping**

|  |  |
| --- | --- |
| **Atualização** | **Racional** |
| I | Identifica uma inclusão de um campo no mapa de ETL desde a última versão do documento |
| A | Identifica uma alteração no campo no mapa de ETL desde a última versão do documento |
| E | Identifica a remoção de um campo no mapa de ETL desde a última versão do documento |
|  | Identifica que não houve alteração no campo desde a última versão do documento |

* 1. **Regras de transformação para o ETL Mapping**

|  |  |  |
| --- | --- | --- |
| **Regra** | **Notação a ser utilizada no Mapping** | **Notação técnica** |
| Condicional | Se...Então...Senão | IF...THEN...ELSE |
| Conversão para Data | Converter para data com o formato ‘dd/mm/yyyy hh24:mi:ss’ | TO\_DATE() |
| Conversão para Texto | Conveter para caracter | TO\_CHAR() |
| Conversão para Número | Converter para numérico | TO\_NUMBER() |
| Remoção de Espaços | Remover espaços em branco | TRIM() |
| Relação entre Tabelas  Obrigatória | Relação entre as tabelas *TAB1* e *TAB2* onde *CAMPO1* = *CAMPO2* | TAB1.CAMPO1 = TAB2.CAMPO2 |
| Relação entre Tabelas  Opcional | Relação entre as tabelas *TAB1* e *TAB2* onde *CAMPO1* = *CAMPO2* retornando todos os registros da tabela *TAB1* | TAB1.CAMPO1 = TAB2.CAMPO2 (+) |
| Tratamento de Nulos | Se *Campo* = Nulo Então -1 senão Campo. | NVL(Campo, -1) |
| Conversão para Maiúsculas | Converter *Campo* para maiúsculas | UPPER(Campo) |
| Conversão para Minúsculas | Converter *Campo* para minúsculas | LOWER(Campo) |
| Lookup | Trazer na tabela *TAB1* onde: *CAMPO1* = *CAMPO2* retornando: *CAMPO3* | LOOKUP |
| Arredondamento | Arredondar *Campo* para 6 casas decimais | ROUND(Campo,6) |
| Concatenação | Concatenar *Campo1* com *Campo2* | Campo1||Campo2 |
| Somar | Somar *Campo1* | SUM(Campo1) |
| Maior Valor | Maior valor do *Campo1* | MAX(Campo1) |
| Menor Valor | Menor valor do *Campo1* | MIN(Campo1) |
| Truncar Valor | Truncar *Campo1* | TRUNC(Campo1) |
| Parse | Realizar Parse no *Campo1* usando como separador a primeira “,” encontrada , considerar a informação antes da vírgula. | SUBSTR() |

* 1. **Estratégias de Carga**

|  |  |
| --- | --- |
| **Estratégia** | **Racional** |
| Truncate e Insert | Trunca a tabela Target e realiza um insert simples de todas as informações do Source |
| Delete e Insert | Deleta na tabela Target os dados com base na chave do Source e realiza um insert simples de todas as informações do Source |
| Delete | Deleta na tabela Target os dados com base na chave do Source |
| Insert | Realiza simples insert direto na tabela Target das informações do Source |
| Delta c/ Insert | Realiza o Delta ( Verifica as mudanças ) e insere os registros do Source que não existem no Target |
| Delta c/ Insert e Update | Realiza o Delta ( Verifica as mudanças ) e insere os registros do Source que não existem no Target e Atualiza os registros do Source que já existem no Target |
| Delta c/ Insert, Update e Delete | Realiza o Delta ( Verifica as mudanças ) e insere os registros do Source que não existem no Target, Atualiza os registros do Source que já existem no Target e Deleta do Target registros que não vieram do Source |

1. **Histórico de Alteração do Documento**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Versão** | **Descrição** | **Responsável** |
| 21/11/2023 | 1.0 | Criação do documento | Flávio Uyemura |

# 9. Aprovação

|  |  |  |
| --- | --- | --- |
| **Data** | **Aprovado por** | **Validação** |
| dd/mm/aaaa |  |  |
|  |  |  |