1. PROCESSO TDMINING

O processo TDMINING é a proposta de um experimento com objetivo de ; **desenvolver uma abordagem para auxiliar o desenvolvedor a conhecer seus comportamentos em relação à dívida técnica do tipo *code smell***.

Para cada uma das atividades identificadas no processo é necessário responder as seguintes perguntas, com base no 5W2H:

* *What*? O que deve ser feito? Passos ou tarefas e descrição da ação;
* *Why*? Por que deve ser feito? Razão, justificativa;
* *Where*? Onde deve ser feito? Área, local;
* *When*? Quando deve ser feito? Data, prazo;
* *Who*? Quem deve fazer? Quem é o responsável pela ação?
* *How*? Como a ação deve ser feita? Qual o método, processo?
* *How Much*? Quanto deve custar para fazer? Custo envolvido na ação.

No entanto, as perguntas “Quando deve ser feito?” e “Quanto deve custar para fazer? o executor da atividade deve determinar. Já a pergunta “Onde deve ser feito?“ para as atividades não informada o executor da atividade deve determinar.

O processo TDMINING, representado na Figura 4‑3 demonstra a raia Desenvolvimento. Embora uma empresa, possam ter uma área de infraestrutura, as atividades de infraestrutura foram consideradas em um subprocesso. No entanto, nada impede que interessados de forma autônoma executem o processo, desde que tenham acesso aos recursos necessários para sua replicação.

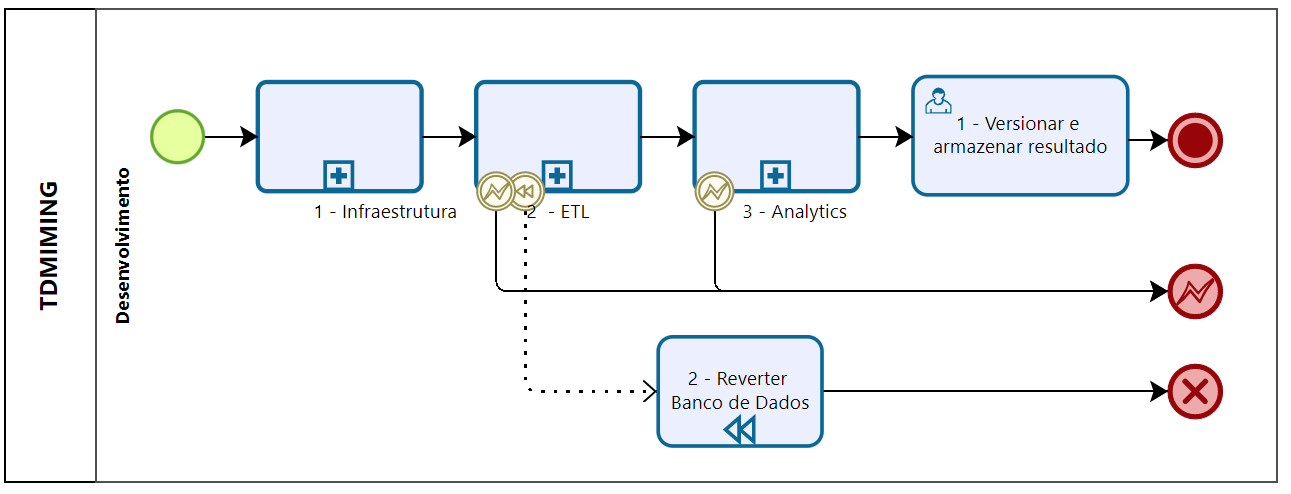


Figura 4‑3. Processo TDMINING. Fonte: o Autor.

Para todas as atividades do processo TDMINING na pergunta “**quem deve fazer?”** será informado um responsável. As responsabilidades de cada uma das atividades em relação ao perfil, pode ser observada na matriz RACI no Quadro 4‑1.

Quadro 4‑1: Matriz RACI do processo TDMINING. Fonte: o Autor.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | INFRAESTRUTURA | BANCO DE DADOS | INTERESSADO | ESPECIALISTA |
| SUBPROCESSO 1 - INFRAESTRUTURA | | | | |
| ATIVIDADE 1 - SOLICITAR INFRAESTRUTURA |  |  | R/A |  |
| ATIVIDADE 2 - PREPARAR INFRAESTRUTURA | R |  | A |  |
| ATIVIDADE 3 - CONCEDER ACESSOS | R |  | A |  |
| ATIVIDADE 4 - PREENCHER DOCUMENTO INFRAESTRUTURA E ACESSOS | R |  | A |  |
| ATIVIDADE 5 - PREPARAR INFRAESTRUTURA DE BANCO DE DADOS | I | R | A |  |
| SUBPROCESSO 2- ETL | | | | |
| ATIVIDADE 1 - IDENTIFICAR REPOSITÓRIO DE CÓDIGO-FONTE |  |  | R/A |  |
| ATIVIDADE 2 - PARAMETRIZAR ARQUIVO PYTHON |  |  | R/A |  |
| ATIVIDADE 3 - PARAMETRIZAR ARQUIVO DE ETL SQL |  |  | R/A |  |
| ATIVIDADE 4 - EXECUTAR ARQUIVO PYTHON |  |  | R/A |  |
| ATIVIDADE 5 - EXECUTAR SCRIPT DE BANCO DE DADOS |  | R | R/A |  |
| ATIVIDADE 6 - IMPORTAR DADOS |  |  | R/A |  |
| ATIVIDADE 7 - ENTREGAR DADOS |  |  | R/A |  |
| ATIVIDADE 8 - VALIDAR ERRO |  |  | R/A |  |
| SUBPROCESSO 3 - ANALYTICS | | | | |
| ATIVIDADE 1 - PREPARAR ARQUIVOS PARA EXECUÇÃO |  |  | R/A |  |
| ATIVIDADE 2 - IMPORTAR ARQUIVOS JUPYTER |  |  | R/A |  |
| ATIVIDADE 3 - EXECUTAR SCRIPTS JUPYTER |  |  | R/A |  |
| ATIVIDADE 4 - DOCUMENTAR RESULTADOS |  |  | R/A |  |
| ATIVIDADE 5 - VALIDAR INFORMAÇÕES |  |  | C/I | R/A |
| ATIVIDADE 6 - INFORMAR RESTRIÇÕES |  |  | C/I | R/A |
| ATIVIDADE 7 - PREENCHER RESTRIÇÕES |  |  | C/I | R/A |
| ATIVIDADE 8 - EXECUTAR SCRIPT DE RESTRIÇÕES |  |  | R/A | C/I |
| PROCESSO PRINCIPAL - TDMINING | | | | |
| ATIVIDADE 1 - VERSIONAR E ARMAZENAR RESULTADO |  |  | R/A |  |
| ATIVIDADE 2 - REVERTER BANCO DE DADOS |  | R | A/C/I |  |

O detalhamento dos conhecimentos necessários para cada um dos perfis de responsabilidade pode ser observado no Quadro 4‑2, observado que para cada conhecimento é atribuído um nível de requerimento. O nível de requerimento é tratado conforme a RFC2119[[1]](#footnote-2).

Quadro 4‑2: Conhecimentos necessários por responsável. Fonte: o Autor.

|  |  |  |  |
| --- | --- | --- | --- |
| RESPONSÁVEL | CONHECIMENTOS NECESSÁRIOS | PERFIL | POSSÍVEIS FUNÇÕES |
| INFRAESTRUTURA | Instalação de softwares (*MUST*), administração de acessos em banco de dados (*MUST*), administração de acessos em aplicações (*MUST*). | JUNIOR (*MUST NOT*), PLENO (*SHOULD*), SENIOR (*SHOULD*) | Analista de Infraestrutura, Administradores de servidor, Técnico de Informática |
| BANCO DE DADOS | Administração de acessos em banco de dados (*MUST*), execução de scripts SQL (*MUST*). | JUNIOR (*MUST NOT*), PLENO (*SHOULD*), SENIOR (*SHOULD*) | Administrator de infraestrutura de banco de dados; Administrador de dados. |
| INTERESSADO | Conhecimento em linguagem de programação Python (*SHOULD*), conhecimento em linguagem SQL (*SHOULD*). | JUNIOR (*SHOULD*), PLENO (*SHOULD*), SENIOR (*SHOULD*) | Desenvolvedor, Líder Técnico; Arquiteto de *Software*, Analista de Sistemas, Analista de Qualidade de Software, Lider Técnico, Gerente de Desenvolvimento de *Software*, Engenheiro de *Software*. |
| ESPECIALISTA | Conhecimento sobre o produto de software (*MUST*) , conhecimento sobre qualidade de código fonte (*SHOULD*), conhecimento em linguagem de programação Python (*SHOULD*), conhecimento em linguagem SQL (*SHOULD*), conhecimento da arquitetura do software analisado (*MUST*), conhecimento sobre estatística (*MAY*), conhecimento sobre ciência de dados (*MAY*), conhecimento em boas práticas de programação (*SHOULD*) | JUNIOR (*MUST NOT*) , PLENO (*SHOULD*), SENIOR (*SHOULD*) | Desenvolvedor, Líder Técnico; Arquiteto de *Software*, Analista de Sistemas, Analista de Qualidade de *Software*, Lider Técnico, Gerente de Desenvolvimento de *Software*, Engenheiro de *Software*. |

Conforme observado no Quadro 4‑2, o que determina cada um dos papéis responsáveis pela execução das tarefas é o conhecimento. Se um profissional possui todos os conhecimentos necessários para os quatro papéis, logo o mesmo, pode atuar na execução das tarefas. No entanto, caso no processo não seja possível identificar um profissional que possa atuar como ESPECIALISTA, o processo pode seguir o fluxo DC1 “Não há especialista” conforme será abordado no item 1.3. O ônus em não ter o papel do ESPECIALISTA está em relação que possíveis *code smells* poderiam ser desconsiderados da análise, exemplo, *code smell* aceitos por ser inerente a arquitetura.

## SUBPROCESSO 1 – INFRAESTRUTURA

O subprocesso Infraestrutura está representado na Figura 4‑4.

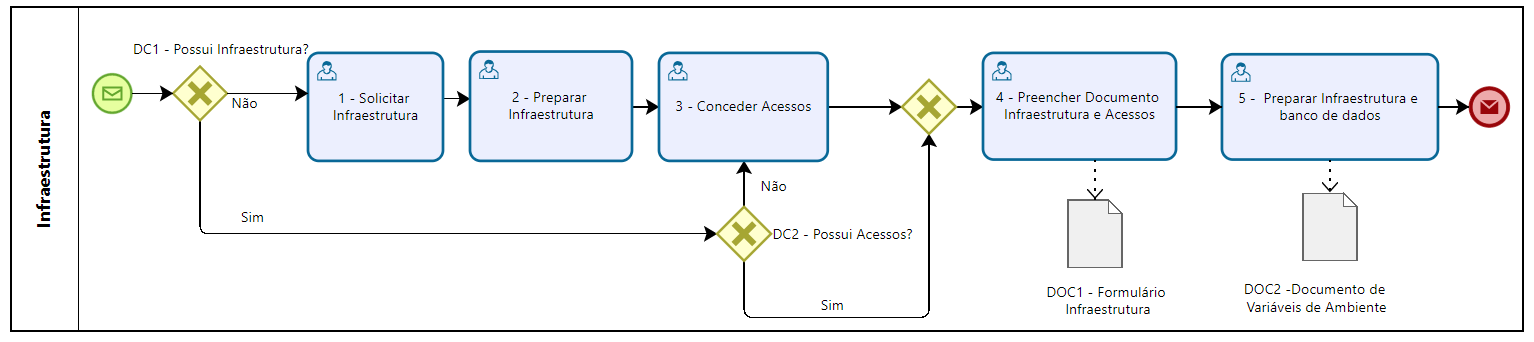


Figura 4‑4. Subprocesso Preparar Infraestrutura. Fonte: o Autor.

#### **SUBPROCESSO 1 - ATIVIDADE 1 – SOLICITAR INFRAESTRUTURA**

**O que deve ser feito?**

* Comunicar ao responsável pela infraestrutura a necessidade de acesso a servidores de SonarQube, Jupyter Notebook e Banco de dados.

**Por que deve ser feito?**

* Para viabilizar a infraestrutura para executar o subprocesso de ETL e Analytics.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante envio de solicitação (e-mail, sistema de helpdesk, e/ou comunicação informal). Acesso a e-mail, sistema de helpdesk ou outro meio de comunicação para o executor.

#### **SUBPROCESSO 1 - ATIVIDADE 2 – PREPARAR INFRAESTRUTURA**

**Caso DC1 – Se a resposta for “NÃO”.**

**O que deve ser feito?**

* Preparar servidor com SonarQube;
* Preparar no mesmo servidor o Jupyter Notebook;
* Preparar no mesmo servidor o banco de dados Microsoft SQL Server\* (pode ser substituído por outro SGBD);
* Instalar no mesmo servidor Git e Python;
* Instalar o *SonarScanner* no servidor;
* Garantir que os recursos computacionais estão devidamente configurados e os acessos corretamente configurados.
  + Se houver qualquer erro de acesso ou infraestrutura para o SUBPROCESSO 2 – ETL e SUBPROCESSO 3 – Analytics o processo TDMINING será cancelado por ERRO.

**Por que deve ser feito?**

* Para conceder acessos e infraestrutura para o solicitante executar o subprocesso de ETL e Analytics.

**Quem deve fazer?**

* PERFIL INFRAESTRUTURA.

**Como a ação deve ser feita?**

* A instalação do SonarQube deve seguir os procedimentos determinados pelo fornecedor, disponível em <https://docs.sonarqube.org/latest/> e <https://docs.sonarqube.org/latest/requirements/hardware-recommendations/> ;
* A instalação do Jupyter Notebook deve seguir os procedimentos determinados pelo fornecedor, disponível em <https://jupyter.org/install>;
* A instalação do banco de dados deve considerar a recomendação do fornecedor;
* Instalação do GIT, disponível em <https://git-scm.com/book/en/v2/Getting-Started-Installing-Git> ;
* Instalação do Python, disponível em <https://docs.python.org/3/> ;
* Instalação do SonarScanner no servidor, disponível <https://docs.sonarqube.org/latest/analysis/scan/sonarscanner/> .

#### **SUBPROCESSO 1 - ATIVIDADE 3 – CONCEDER ACESSOS**

**O que deve ser feito?**

* Criar *login* e senha ao servidor com SonarQube;
* Criar *login* e senha ao servidor com Jupyter Notebook;
* Criar *login* e senha ao servidor com banco de dados Microsoft SQL Server \* (pode ser substituído por outro SGBD).

**Por que deve ser feito?**

* Para conceder acessos e infraestrutura para o solicitante executar o subprocesso de ETL e Analytics.

**Quem deve fazer?**

* PERFIL INFRAESTRUTURA.

**Como a ação deve ser feita?**

* Mediante acesso do executor aos servidores para criação de perfil e usuário de acesso aos sistemas.

#### **SUBPROCESSO 1 - ATIVIDADE 4 – PREENCHER DOCUMENTO INFRAESTRUTURA E ACESSOS**

**O que deve ser feito?**

* Preencher no documento DOC1 os acessos para solicitante, detalhando login e senha, endereço para acesso ao SonarQube, Jupyter Notebook e banco de dados.
* Enviar documento ao solicitante.

**Por que deve ser feito?**

* Para o solicitante ter conhecimento dos seus acessos aos servidores solicitados.

**Onde deve ser feito?**

* A descrição de onde será realizada no documento DOC1.

**Quem deve fazer?**

* PERFIL INFRAESTRUTURA.

**Como a ação deve ser feita?**

* Preencher o documento DOC1.
* Enviar o documento para o solicitante.

#### **SUBPROCESSO 1 - ATIVIDADE 5 – PREPARAR INFRAESTRUTURA DE BANCO DE DADOS**

**O que deve ser feito?**

* Aplicar o *script* CREATE\_DDL.sql em banco de dados. O objetivo do *script* CREATE\_DDL.sql é criar as estruturas de dados de dimensões e fatos conforme exibido no Quadro 4‑3. Na Figura 4‑5, a cor vermelha é definida para as tabelas que representam o *starschema* e em azul as tabelas que complementam para a estrutura de *snowflake*.
* Preenchimento do DOC2.

**Por que deve ser feito?**

* Para criar o modelo de dados para persistência de dados;
* Para preparar execução da coleta de dados.

**Onde deve ser feito?**

* No servidor de banco de dados descrito no documento DOC1.

**Quem deve fazer?**

* PERFIL BANCO DE DADOS.

**Como a ação deve ser feita?**

* Mediante o acesso ao banco de dados descrito no DOC1.

Quadro 4‑3. Dicionário de dados TDMINING. Fonte: o Autor.

| TABELA | COLUNA | TIPO | DESCRIÇÃO |
| --- | --- | --- | --- |
| DM\_ACTION | DESCRIPTION | texto | Descrição da ação |
| DM\_ACTION | ID\_ACTION | número | Identificador da ação |
| DM\_COMMITS | COMMIT\_DATE | Data, hora e minuto | Data do *commit* |
| DM\_COMMITS | COMMIT\_HASH | texto | Identificador do *commit* |
| DM\_COMMITS | ID\_COMMIT | número | Identificador do *commit* |
| DM\_COMMITS | ID\_DEVELOPER | número | Identificador do desenvolvedor |
| DM\_COMMITS | ID\_PROJECT | número | Identificador do projeto |
| DM\_COMMITS | ID\_TIME | número | Identificador do tempo |
| DM\_DEVELOPER | DEVELOPER\_NAME | texto | Nome do desenvolvedor ou identificador ou e-mail |
| DM\_DEVELOPER | ID\_DEVELOPER | número | Identificador do desenvolvedor |
| DM\_PROJECT | ID\_PROJECT | número | Identificador do projeto |
| DM\_PROJECT | PROJECT\_NAME | texto | Nome do projeto |
| DM\_RULE | ID\_RULE | número | Identificador da regra |
| DM\_RULE | ORIGIN\_RULE\_ID | texto | Identificador da origem da regra |
| DM\_RULE | RULE\_DESCRIPTION | texto | Descrição da regra |
| DM\_RULE | RULE\_NAME | texto | Nome da regra |
| DM\_SEVERITY | DESCRIPTION | texto | Descrição da severidade |
| DM\_SEVERITY | ID\_SEVERITY | número | Identificador da severidade |
| DM\_TDITEM | DESCRIPTION | texto | Descrição do tipo de dívida técnica |
| DM\_TDITEM | ID\_TDITEM | número | Identificador do tipo de item de dívida técnica |
| DM\_TIME | DATE | Data | Data no formato dd-mm-yyyy |
| DM\_TIME | ID\_TIME | número | Identificador da data |
| DM\_VERSION | ID\_COMMIT | número | Identificador do *commit* |
| DM\_VERSION | ID\_PROJECT\_VERSION | número | Identificador da versão para o projeto |
| DM\_VERSION | ID\_VERSION | número | Identificador da versão |
| FT\_TECHNICALDEBT | EFFORT | número | Esforço para correção da dívida técnica em minutos |
| FT\_TECHNICALDEBT | ID\_ACTION | número | Identificador da ação |
| FT\_TECHNICALDEBT | ID\_DEVELOPER | número | Identificador do desenvolvedor |
| FT\_TECHNICALDEBT | ID\_FTECHNICALDEBT | número | Identificador do fato de dívida técnica |
| FT\_TECHNICALDEBT | ID\_PROJECT | número | Identificador do projeto |
| FT\_TECHNICALDEBT | ID\_RULE | número | Identificador da regra |
| FT\_TECHNICALDEBT | ID\_SEVERITY | número | Identificador da severidade |
| FT\_TECHNICALDEBT | ID\_TDITEM | número | Identificador do tipo de dívida técnica |
| FT\_TECHNICALDEBT | ID\_TIME | número | Identificador da data de criação da dívida técnica |
| FT\_TECHNICALDEBT | ID\_VERSION | número | Identificador da versão do produto que foi criado a dívida técnica |
| FT\_TECHNICALDEBT | IGNORE | número | Identificador se o dado deve ser ignorado nas análises, 1 -SIM , 0- NÃO |
| ORGANIZATION | ID\_ORGANIZATION | número | Identificador da organização |
| ORGANIZATION | ORGANIZATION\_NAME | texto | Nome da organização ou empresa |
| ORGANIZATION\_PROJECT | ID\_ORGANIZATION | número | Identificador da organização |
| ORGANIZATION\_PROJECT | ID\_ORGANIZATION\_PROJECT | número | Identificador do projeto na organização |
| ORGANIZATION\_PROJECT | ID\_PROJECT | número | Identificador do projeto |
| PROJECT\_VERSION | DATE\_VERSION | Data, hora e minuto | Data da versão |
| PROJECT\_VERSION | DATE\_VERSION\_STR | texto | Data da versão em formato texto |
| PROJECT\_VERSION | ID\_PROJECT | número | Identificador do projeto |
| PROJECT\_VERSION | ID\_PROJECT\_VERSION | número | Identificador do verão do projeto |
| PROJECT\_VERSION | VERSION\_NAME | texto | Nome da versão |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | EFFORT | número | Esforço para correção da dívida técnica em minutos |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_ACTION | número | Identificador da ação |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_DEVELOPER | número | Identificador do desenvolvedor |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_FTECHNICALDEBT | número | Identificador do fato de dívida técnica |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_PROJECT | número | Identificador do projeto |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_RULE | número | Identificador da regra |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_SEVERITY | número | Identificador da severidade |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_TDITEM | número | Identificador do tipo de dívida técnica |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_TIME | número | Identificador da data de criação da dívida técnica |
| FT\_TECHNICALDEBT\_IGNORE\_CONDITION | ID\_VERSION | número | Identificador da versão do produto que foi criado a dívida técnica |

Diagrama

Descrição gerada automaticamente

Figura 4‑5. Diagrama entidade relacionamento – dimensões x fato. Fonte: o Autor.

## SUBPROCESSO 2 – ETL

O subprocesso ETL tem o propósito entregar os dados necessários para a mineração, conforme tarefas detalhadas na Figura 4‑6.

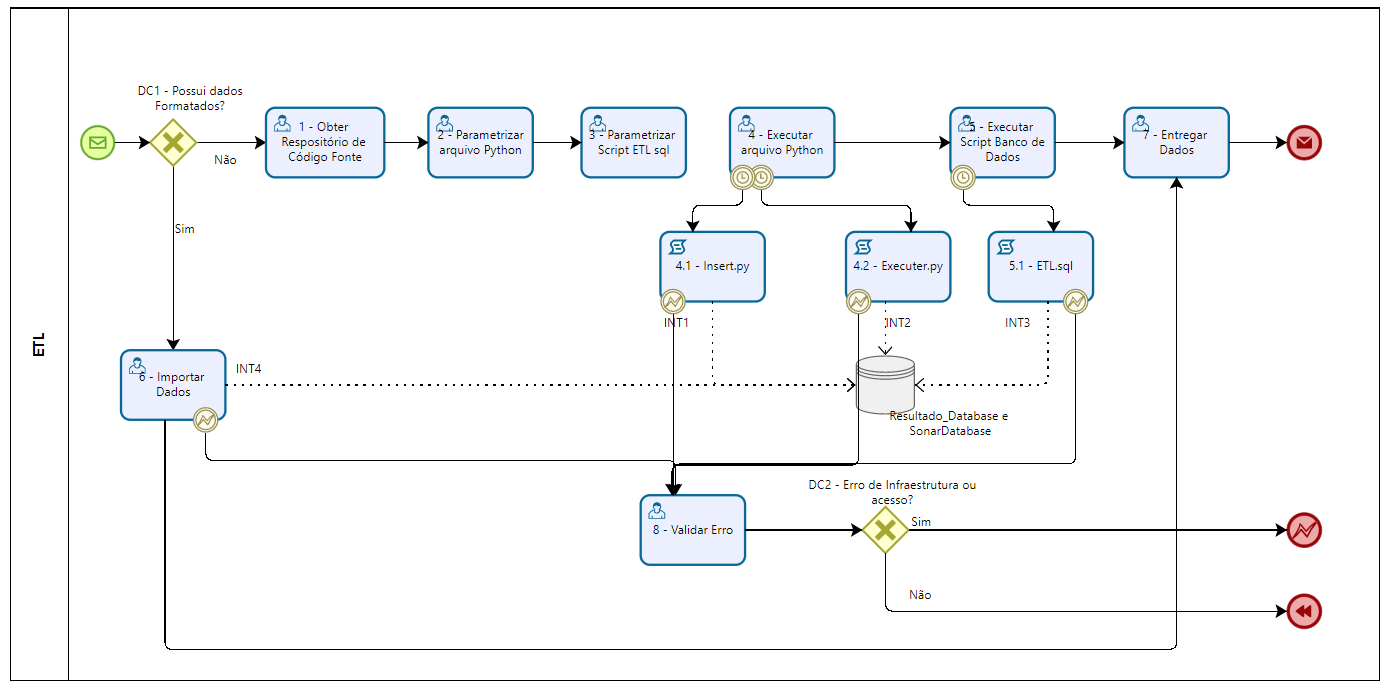


Figura 4‑6. Subprocesso ETL. Fonte: o Autor.

#### **SUBPROCESSO 2 - ATIVIDADE 1 – OBTER REPOSITÓRIO DE CÓDIGO-FONTE**

**O que deve ser feito?**

* Identificar e obter repositório de código-fonte;
* Realizar cópia do repositório por meio de GIT CLONE.

**Por que deve ser feito?**

* Para obter o histórico do controle de versão do produto selecionado.

**Onde deve ser feito?**

* No servidor identificado no DOC1.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante o acesso ao servidor descrito no DOC1 deve-se criar um diretório para realizar cópia do repositório de código-fonte. Por meio da linha de comando GIT CLONE, conforme documentação disponível em https://github.com/git-guides/git-clone#:~:text=The%20git%20clone%20command%20is,your%20own%20machine%20by%20cloning.

#### **SUBPROCESSO 2 - ATIVIDADE 2 – PARAMETRIZAR ARQUIVO PYTHON**

**O que deve ser feito?**

* Alterar o arquivo config.ini informado no DOC2, o diretório do repositório de código e endereço do banco de dados.

**Por que deve ser feito?**

* Para preparar os arquivos insert.py e executer.py para execução.

**Onde deve ser feito?**

* No servidor informado no DOC1.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante alteração de variáveis de ambiente, utilizando de editor de código-fonte/texto de preferência do executor.

**Quanto deve custar para fazer?**

* A definir pelo PERFIL INTERESSADO.

#### **SUBPROCESSO 2 - ATIVIDADE 3 – PARAMETRIZAR ARQUIVO DE ETL SQL**

**O que deve ser feito?**

* Alterar no arquivo ANALISYS\_TEMP.sql o nome dos *schemas,* conforme informado no DOC2.

**Por que deve ser feito?**

* Para realizar a transformação e carga dos dados no formato do StarSchema.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante alteração de código-fonte em linguagem SQL, utilizando de editor de código-fonte /texto de preferência do executor.

**Quanto deve custar para fazer?**

* A definir pelo PERFIL INTERESSADO.

#### **SUBPROCESSO 2 - ATIVIDADE 4 – EXECUTAR ARQUIVO PYTHON**

**O que deve ser feito?**

* Executar o arquivo insert.py, aguardar a conclusão do processamento;
* Executar o arquivo executer.py, aguardar a conclusão do processamento.

**Por que deve ser feito?**

* Para realizar a criação dos dados para análise.

**Onde deve ser feito?**

* Os arquivos devem ser executados no servidor informado no DOC1.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante a execução dos arquivos por linha de comando;
* Python <nome\_do\_arquivo.py>

#### **SUBPROCESSO 2 - ATIVIDADE 5 – EXECUTAR *SCRIPT* DE BANCO DE DADOS**

**O que deve ser feito?**

* Executar o arquivo ANALISYS\_TEMP.sql. A tarefa é concluída quando todos os procedimentos no arquivo foram executados com sucesso.

**Por que deve ser feito?**

* Para realizar a transformação de dados para o modelo STARCHEMA proposto.

**Quem deve fazer?**

* Interessado, desenvolvedor ou pesquisador.

**Como a ação deve ser feita?**

* Mediante a execução do conteúdo do arquivo em interface para execução de comandos SQL.

#### **SUBPROCESSO 2 - ATIVIDADE 6 – IMPORTAR DADOS**

**O que deve ser feito?**

* Importar dados formatados para a base de dados. Dados formatados conforme demonstrado no Quadro 4‑3. Para essa tarefa não é obrigatório que os dados que serão persistidos tenham origem do SonarQube, podendo ser originados de qualquer ferramenta de análise estática, desde que sigam o modelo de dados proposto.

**Por que deve ser feito?**

* Para realizar a análise dos dados.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Por meio de *script* ou *software* identificados pelo interessado.

**Quanto deve custar para fazer?**

* A definir pelo PERFIL INTERESSADO.

#### **SUBPROCESSO 2 - ATIVIDADE 7 – ENTREGAR DADOS**

**O que deve ser feito?**

* Exportar os dados das tabelas criadas para o formato CSV ou fornecer a *string* de conexão com o banco de dados.

**Por que deve ser feito?**

* Para permitir a execução da análise dos dados.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Os arquivos com extensão CSV podem ser extraídas por tabela, conforme IDE de preferência do executor. A *string* de conexão pode ser informada por *e-mail*, texto ou qualquer canal de comunicação de preferência do executor.

#### **SUBPROCESSO 2 - ATIVIDADE 8 – VALIDAR ERRO**

**O que deve ser feito?**

* Analisar os erros gerados na execução dos scripts ou importação de dados.

**Por que deve ser feito?**

* Para poder reiniciar o processo em caso de erro.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Se algum dos erros são referentes a infraestrutura (DC2 - Erro de Infraestrutura ou acesso?) igual a SIM. O Processo deve ser terminado com erro. Os problemas de infraestrutura devem ser verificados com o responsável. Senão, o processo de erro gera um evento de compensação. O evento de compensação é tratado a nível do processo principal TDMINING com a atividade “2 – Reverter banco de dados”.

## SUBPROCESSO 3 – *ANALYTICS*

O subprocesso *Analytics,* ilustrado na Figura 4‑7, tem por objetivo entregar o resultado da mineração de dados meio de gráficos e tabelas.

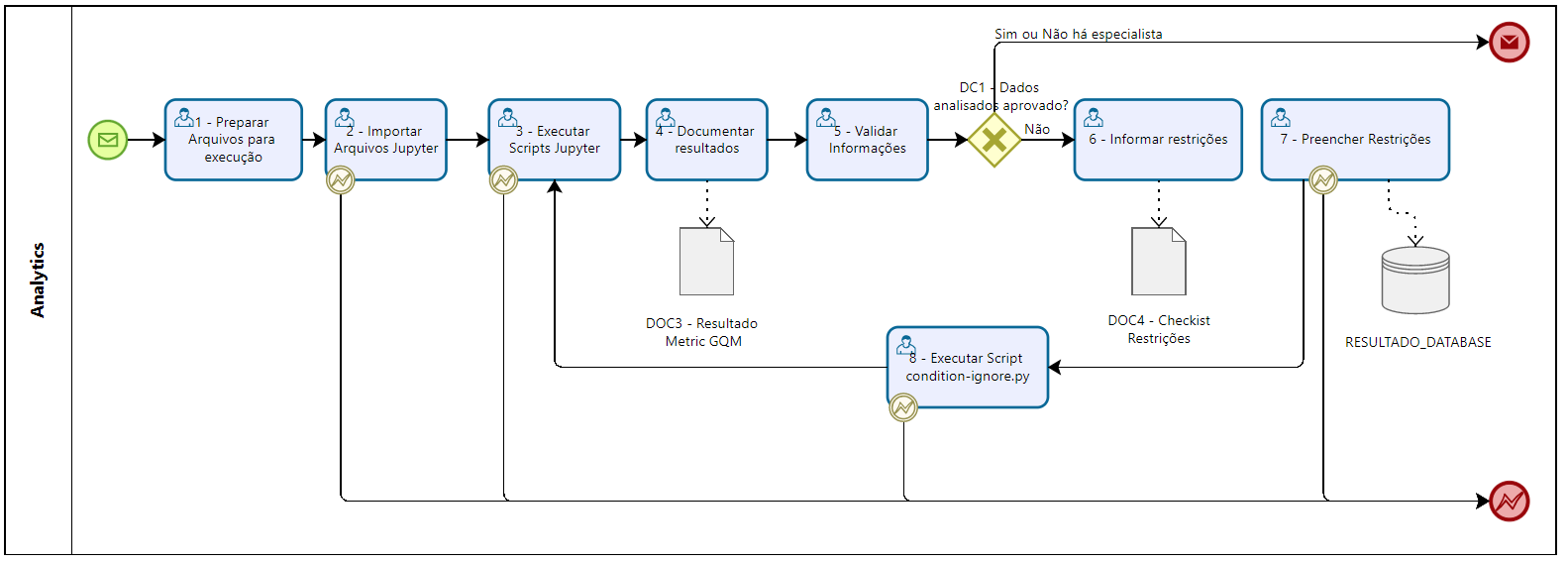


Figura 4‑7. Subprocesso Analytics. Fonte: o Autor.

#### **SUBPROCESSO 3 – ATIVIDADE 1 – PREPARAR ARQUIVOS PARA EXECUÇÃO**

**O que deve ser feito?**

* Garantir que o arquivo config.ini está com as variáveis de ambiente conforme DOC2.

**Por que deve ser feito?**

* Garantir que os *scripts* consigam conectar ao banco de dados.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante utilização de editor de texto ou código-fonte.

#### **SUBPROCESSO 3 – ATIVIDADE 2 – IMPORTAR ARQUIVOS JUPYTER**

**O que deve ser feito?**

* Importar arquivos com a extensão IPYNB para a aplicação do Jupyter Notebook.

**Por que deve ser feito?**

* Para poder executar os arquivos no servidor Jupyter Notebook.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante a utilização da interface gráfica do Jupyter Notebook.

#### **SUBPROCESSO 3 – ATIVIDADE 3 – EXECUTAR *SCRIPTS* JUPYTER**

**O que deve ser feito?**

* Executar no Jupyter Notebook os *scripts:*
  + teste-normalidade.py
  + teste-hipotese.py
  + question-1-metric-1.ipynb
  + question-1-metric-2.ipynb
  + question-2-metric-1.ipynb
  + question-3-metric-1.ipynb
  + question-4-metric-1.ipynb
  + question-5-metric-1.ipynb.

Os scripts apontados estão disponibilizados na internet através do endereço <https://github.com/manoelvsneto/tdmining>, para cada uma das tarefas tem relação a um script Jupyter Notebook. A relação dos scripts em relação e seus objetivos podem ser observados no APÊNDICE H, onde também é possível verificar quais os tipos de informação que a execução de cada *script* retorna como resultado.

**Por que deve ser feito?**

* Para obter o resultado da métrica de interesse.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante a utilização da interface gráfica do Jupyter Notebook.

#### **SUBPROCESSO 3 – ATIVIDADE 4 – DOCUMENTAR RESULTADOS**

**O que deve ser feito?**

* Capturar o resultado da execução dos *scripts* da Tarefa 3 e documentar no arquivo DOC3.

**Por que deve ser feito?**

* Para tomar decisões sobre o resultado das métricas executadas.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Mediante preenchimento do documento DOC3.

#### **SUBPROCESSO 3 – ATIVIDADE 5 – VALIDAR INFORMAÇÕES**

**O que deve ser feito?**

* Analisar os dados apresentados no DOC3.

**Por que deve ser feito?**

* Para avaliar a existência de dados que devem ser desconsiderados. Pois em detrimento do conhecimento do ESPECIALISTA, pode ocorrer que determinada regra de *code smell* deva ser desconsiderada.

**Quem deve fazer?**

* PERFIL ESPECIALISTA.

**Como a ação deve ser feita?**

* Leitura e interpretação dos resultados apresentados no DOC3 .

#### **SUBPROCESSO 3 – ATIVIDADE 6 – INFORMAR RESTRIÇÕES**

**O que deve ser feito?**

* Preencher o documento DOC4 – Checklist de restrições com as condições que devem ser ignoradas para tratamento dos dados.

**Por que deve ser feito?**

* Por motivo de ser um *code smell* ser aceito no projeto, ou até mesmo ser inerente a arquitetura do sistema. Ou em condições que deva ser removida da análise um determinado desenvolvedor ou data.
  + Idealmente é esperado que não haja viés do ESPECIALISTA em relação aos dados.

**Quem deve fazer?**

* PERFIL ESPECIALISTA.

**Como a ação deve ser feita?**

* Preenchimento de formulário.

#### **SUBPROCESSO 3 – ATIVIDADE 7 – PREENCHER RESTRIÇÕES**

**O que deve ser feito?**

* Preencher o formulário conforme o exemplo no DOC4 – CHECKLIST DE RESTRIÇÕES.

**Por que deve ser feito?**

* Para ignorar dados desnecessários na análise.

**Quem deve fazer?**

* PERFIL ESPECIALISTA.

**Como a ação deve ser feita?**

* Inserção de dados em banco de dados na tabela FT\_TECHNICALDEBT\_IGNORE\_CONDITION.

#### **SUBPROCESSO 3 – ATIVIDADE 8 – EXECUTAR *SCRIPT* DE RESTRIÇÕES**

**O que deve ser feito?**

* Executar o *script* condition-ignore.py.

**Por que deve ser feito?**

* Para ignorar os dados que devem ser desconsiderados das análises.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Por meio de execução de *script* Python em ambiente de desenvolvimento.

## PROCESSO PRINCIPAL - ATIVIDADE 1 – VERSIONAR E ARMAZENAR RESULTADO

**O que deve ser feito?**

* Atribuir versão ao documento DOC3 e armazenar em sistema de controle de versão, de gestão de conteúdo ou local de conhecimento dos interessados.

**Por que deve ser feito?**

* Manter histórico e fonte de referência para tomada de decisão.

**Quem deve fazer?**

* PERFIL INTERESSADO.

**Como a ação deve ser feita?**

* Preenchimento do documento DOC3.
* Armazenamento do documento DOC3 em local definido pelo PERFIL INTERESSADO.

**Quanto deve custar para fazer?**

* A definir pelo PERFIL INTERESSADO.

## PROCESSO PRINCIPAL - ATIVIDADE 2 – REVERTER BANCO DE DADOS

**O que deve ser feito?**

* Rever banco de dados TDMINING;

**Por que deve ser feito?**

* Para permitir que seja possível realizar o experimento após um erro.

**Quem deve fazer?**

* PERFIL BANCO DE DADOS.

**Como a ação deve ser feita?**

* Execução do *script* revert-database.py.

1. RELAÇÃO DOS SCRIPTS PYTHON COM AÇÕES DE BANCO DE DADOS

|  |  |  |  |
| --- | --- | --- | --- |
| QUESTÃO | OBJETIVO | SCRIPT | APRESENTAÇÃO / ESTATISTICA |
|  | Recuperar os dados de commit de um repositório de código fonte e persistir os dados em banco de dados. | insert.py | Não se aplica |
|  | Processar individualmente os commits por meio de análise estática para identificar code smell em relação ao commit anterior. | executer.py | Não se aplica |
| TESTE DE NORMALIDADE | Realizar teste de normalidade com os dados de code smell. | teste-normalidade.py | Histograma, Shapiro-Wilk, Kolmogorov-Smirnov |
| TESTE DE HIPOTESE | Realizar teste de hipótese com os dados de code smell. | teste-hipotese.py | Gráfico de dispersão, SpearmanR |
| QUESTÃO 1 METRICA 1 | Responder à questão: Qual o desenvolvedor mais cria code smells? | question-1-metric-1.ipynb | Gráfico de pizza, Tabela , Sumarização |
| QUESTÃO 1 METRICA 2 | Responder à questão: Qual o desenvolvedor mais remove code smells? | question-1-metric-2.ipynb | Gráfico de pizza, Tabela , Sumarização |
| QUESTÃO 2 METRICA 1 | Responder à questão: Qual o code smell que foi mais criado pelos desenvolvedores? | question-2-metric-1.ipynb | Tabela, Sumarização. |
| QUESTÃO 3 METRICA 1 | Responder à questão: Quais os code smells que ocorrem em comum pelo desenvolvedor? | question-3-metric-1.ipynb | Regra de associação (APRIORI). Itens Frequentes |
| QUESTÃO 4 METRICA 1 | Responder à questão: Qual a tendência de code smell criado? | question-4-metric-1.ipynb | Gráfico de linha (Média Móvel (SMA) ), Tabela |
| QUESTÃO 5 METRICA 1 | Responder à questão: Qual a previsão de criação de code smell? | question-5-metric-1.ipynb | Tabela. Histograma. Teste de estacionariedade (ADF e KPSS). Previsão para series temporais (ARIMA) |
|  | Ignorar registros da tabela fato. Condições informadas pelo ESPECIALISTA no DOC4 - CHECKLIST DE RESTRIÇÕES | condition-ignore.ipynb | Não se aplica. |
|  | Limpar o banco de dados para uma nova carga de dados | reset-database. py | Não se aplica |

1. Disponível em: <https://datatracker.ietf.org/doc/html/rfc2119> e https://www.ietf.org/rfc/rfc2119.txt. [↑](#footnote-ref-2)