

**UNIVERSIDADE DE CAXIAS DO SUL**  
**ÁREA DO CONHECIMENTO DE CIÊNCIAS EXATAS E**  
**ENGENHARIAS**

**PEDRO ANTONIO DELAGUSTINHI**

**ANÁLISE DE FERRAMENTAS DE MAPEAMENTO DE DADOS**

**CAXIAS DO SUL**

**2021**

**PEDRO ANTONIO DELAGUSTINHI**

**ANÁLISE DE FERRAMENTAS DE MAPEAMENTO DE DADOS**

Trabalho de Conclusão de Curso para  
obtenção do Grau de Bacharel em Sistemas  
de Informação da Universidade de Caxias  
do Sul.

Orientadora Prof. Dra. Maria de Fátima  
Webber do Prado Lima

**CAXIAS DO SUL**

**2021**

**PEDRO ANTONIO DELAGUSTINHI**

**ANÁLISE DE FERRAMENTAS DE MAPEAMENTO DE DADOS**

Trabalho de Conclusão de Curso para  
obtenção do Grau de Bacharel em Sistemas  
de Informação da Universidade de Caxias  
do Sul.

**Aprovado em 02/12/2021.**

**Banca Examinadora**

---

Prof. Dra. Maria de Fátima Webber do Prado Lima

Universidade de Caxias do Sul

---

Prof. Dra. Carine Geltrudes Webber

Universidade de Caxias do Sul

---

Prof. Dra. Elisa Boff

Universidade de Caxias do Sul

## **AGRADECIMENTOS**

Agradeço em primeiro lugar aos meus pais Vanius e Lilian e aos meus avós por acreditarem em mim e sempre estarem presentes me apoiando.

Agradeço a minha namorada Mycaella pela dedicação, paciência e o apoio cujo foi essencial durante toda a trajetória.

Agradeço também a Profa. Dra. Maria de Fátima Webber do Prado Lima responsável pela orientação deste trabalho, por todo suporte e dedicação durante esta etapa. Também sou grato aos demais professores.

Agradeço ainda aos colegas, amigos e demais pessoas que, de alguma forma, contribuíram para a concretização deste objetivo.

## RESUMO

Este trabalho tem como objetivo analisar e avaliar ferramentas de mapeamento de dados, com o intuito de identificar a ferramenta que mais tem aderências com a norma ABNT NBR ISO/IEC 27701 e a Lei Geral de Proteção de Dados Pessoais. Para fundamentar esta avaliação foram estudados alguns conceitos sobre classificação de dados, a Lei Geral de Proteção de Dados Pessoais, normas que gerem privacidade dos dados (ABNT NBR ISO/IEC 27701) e as normas que regem padrões para a avaliação da qualidade de softwares (ABNT NBR ISO/IEC 25050 e ABNT NBR ISO/IEC 25030). Com os estudos foram definidos critérios, métricas e casos de testes para a avaliação das ferramentas, de acordo com as funcionalidades essenciais para atendimento das conformidades que foram definidas pela norma ABNT NBR ISO/IEC 27701 e a Lei Geral de Proteção de Dados Pessoais. Foram avaliadas as ferramentas de extração, transformação e carregamento de dados Talend Open Studio, Pentaho Kettle e CloverDX. Ferramentas de mapeamento de dados desenvolvidas para apoiar as questões de conformidade de proteção aos dados não foram testadas, pois não são disponibilizadas suas versões de testes. Das ferramentas avaliadas a que obteve melhor resultado foi o Pentaho Kettle. Após realizar os testes e a análise das ferramentas de mapeamento de dados, conclui-se que as ferramentas de mapeamento de dados utilizadas para a extração, transformação e carregamento de dados não são as ferramentas apropriadas para serem utilizadas em conformidade com as Leis e normas.

**Palavras-chave:** Privacidade dos dados. Mapeamento de dados. Lei Geral de Proteção de Dados Pessoais.

## **ABSTRACT**

This work aims to analyze and evaluate data mapping tools, in order to identify the tool that most adheres to the ABNT NBR ISO/IEC 27701 standard and the General Law for the Protection of Personal Data. To support this assessment, some concepts about data classification, the General Law for the Protection of Personal Data, norms that manage data privacy (ABNT NBR ISO/IEC 27701) and the norms that govern standards for the evaluation of software quality were studied ( ABNT NBR ISO/IEC 25050 and ABNT NBR ISO/IEC 25030). With the studies, criteria, metrics and test cases were defined for the evaluation of the tools, according to the essential functionalities to comply with the conformity defined by the ABNT NBR ISO/IEC 27701 standard and the General Law for the Protection of Personal Data. The tools for extracting, transforming and loading data Talend Open Studio, Pentaho Kettle and CloverDX were evaluated. Data mapping tools developed to support data protection compliance issues have not been tested as no test versions of them have been made available. Of the tools evaluated, the one with the best result was the Pentaho Kettle. After performing the tests and analyzing them, it is concluded that the data mapping tools used for data extraction, transformation and loading are not the appropriate tools to be used in accordance with the Laws and standards.

**Keywords:** Data privacy. Data mapping. General Personal Data Protection Law.

## LISTA DE FIGURAS

Figura 1 - Ciclo de vida dos dados segundo a LGPD	16
Figura 2 - Momentos do ciclo dos dados	21
Figura 3 - Área e requisitos da ABNT NBR ISO/IEC 27001	35
Figura 4 - Modelo para medição de qualidade de produto de software	44
Figura 5 - Estrutura da divisão de Medições da Qualidade	45
Figura 6 - Estrutura da divisão de Medições da Qualidade	47
Figura 7 - Hierarquia dos requisitos do sistema e do software	48
Figura 8 - Exemplo de modelo de sistemas e modelos de qualidade	49
Figura 9 - Modelo de qualidade	51
Figura 10 - Características e Subcaracterísticas do Modelo de Qualidade	51
Figura 11 - Características de Qualidade em Uso	54
Figura 12 - Exemplo de tela do Talend Open Studio	59
Figura 13 - Exemplo de tela do Pentaho Kettle	60
Figura 14 - Exemplo de tela do Lumify	61
Figura 15 - Exemplo de tela do CloverDX	63
Figura 16 - Exemplo de tela do OneTrust	64
Figura 17 - Exemplo de tela do BigID	65
Figura 18 - Exemplo de tela do DataGrail	66
Figura 19 - Exemplo de tela do TrustEc	67
Figura 20 - Tela inicial Talend Open Studio	93
Figura 21 - Talend Open Studio Conexão ao MySQL	94
Figura 22 - Talend Open Studio Conexão ao Oracle	94
Figura 23 - Talend Open Studio parametrizado	95
Figura 24 - Talend Open Studio aplicando CRC	98
Figura 25 - Talend Open Studio configurando tScramcle	99
Figura 26 - Talend Open Studio dados criptografados	100
Figura 27 - Talend Open Studio dados criptografados	100
Figura 28 - Talend Open Studio armazenamento de logs	101
Figura 29 - Talend Open Studio resultado armazenamento de logs	101

Figura 30 - Talend Open Studio tratamento dos dados	103
Figura 31 - Talend Open Studio aplicar filtro aos dados	104
Figura 32 - Talend Open Studio configuração de anonimização	105
Figura 33 - Talend Open Studio resultado da anonimização	106
Figura 34 - Talend Open Studio apresenta travamentos com Oracle XE	107
Figura 35 - Talend Open Studio apresenta travamentos com Oracle XE	107
Figura 36 - Talend Open Studio uso de recursos	108
Figura 37 - Pentaho Kettle conectado com sucesso ao MySQL	110
Figura 38 - Pentaho Kettle conectado com sucesso ao SQL Server	110
Figura 39 - Pentaho Kettle conectado com sucesso ao Oracle XE	111
Figura 40 - Pentaho Kettle rotulação dos dados	113
Figura 41 - Pentaho Kettle criptografia de dados	114
Figura 42 - Pentaho Kettle resultado criptografia de dados	115
Figura 43 - Pentaho Kettle configurar armazenamento log	116
Figura 44 - Pentaho Kettle visualizando eventos	117
Figura 45 - Pentaho Data Integration tratamento dos dados	119
Figura 46 - Pentaho Kettle aplicar filtro aos dados	120
Figura 47 - Pentaho Kettle uso de recursos	122
Figura 48 - CloverDX conectado com sucesso ao MySQL	124
Figura 49 - CloverDX conectado com sucesso ao SQL Server	124
Figura 50 - CloverDX conectado com sucesso ao Oracle XE	125
Figura 51 - CloverDX definindo permissões ao grupo	126
Figura 52 - CloverDX criando usuário	127
Figura 53 - CloverDX utilizando usuario teste	127
Figura 54 - CloverDX armazenamento de eventos	129
Figura 55 - CloverDX tratamento dos dados	131
Figura 56 - CloverDX aplicar filtro aos dados	132
Figura 57 - CloverDX uso de recursos	134



## LISTA DE QUADROS

Quadro 1 - Níveis de classificação	16
Quadro 2 - Tratamento de informações	18
Quadro 3 - Requisitos específicos da SGPI com relação a ABNT NBR ISO/IEC 27001	30
Quadro 4 - Seções específicas pela ABNT NBR ISO/IEC 27002	31
Quadro 5 - Comparação entre ferramentas de mapeamento de dados	50
Quadro 6 - Comparação entre ferramentas de mapeamento de dados	62
Quadro 7 - Comparação dos serviços entre ferramentas de mapeamento de dados	63
Quadro 8 - Critérios para Avaliação	66
Quadro 9 - Métricas para avaliação de Qualidade Externa	69
Quadro 10 - Métricas para avaliação de Qualidade em Uso	70
Quadro 11 - Caso de Teste 1	74
Quadro 12 - Caso de Teste 2	74
Quadro 13 - Caso de Teste 3	75
Quadro 14 - Caso de Teste 4	75
Quadro 15 - Caso de Teste 5	76
Quadro 16 - Caso de Teste 6	76
Quadro 17 - Caso de Teste 7	76
Quadro 18 - Caso de Teste 8	77
Quadro 19 - Caso de Teste 9	77
Quadro 20 - Caso de Teste 10	78
Quadro 21 - Caso de Teste 11	78
Quadro 22 - Caso de Teste 12	78
Quadro 23 - Dicionário dos Dados Afastamento Remunerado	82
Quadro 24 - Qualidade dos Dados Afastamento Remunerado	82
Quadro 25 - Dicionário dos Dados Motoristas Habilitados	82
Quadro 26 - Funcionalidades Adequação	134
Quadro 27 - Critério Adequação	134
Quadro 28 - Tarefas Acurácia	135

Quadro 29 - Critério Acurácia	135
Quadro 30 - Tarefas Interoperabilidade	136
Quadro 31 - Critério Interoperabilidade	136
Quadro 32 - Conformidades	137
Quadro 33 - Critério Conformidade	138
Quadro 34 - Critério Recuperabilidade	139
Quadro 35 - Critério Operacionalidade	139
Quadro 36 - Critério Coexistência	140
Quadro 37 - Tarefas Efetividade	141
Quadro 38 - Critério Efetividade	141
Quadro 39 - Critério Produtividade	142
Quadro 40 - Conformidades	143
Quadro 41 - Classificação dos critérios	143

## **LISTA DE ABREVIATURAS E SIGLAS**

ABNT	Associação Brasileira de Normas Técnicas
ISO	Organizações Internacionais de Normalização
IEC	Comissão Eletrotécnica Internacional
NBR	Norma Brasileira
RGPD	Regulamento Geral sobre a Proteção de Dados
LGPD	Lei Geral de Proteção de Dados Pessoais
SGSI	Sistema de Gestão de Segurança da Informação
SGPI	Sistema de Gestão de Privacidade da Informação

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>14</b>
1.1 PROBLEMA DE PESQUISA	16
1.2 OBJETIVO	17
1.2.1 Objetivos Específicos	17
1.3 METODOLOGIA	18
1.4 ESTRUTURA DO TRABALHO	19
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>20</b>
2.1 CLASSIFICAÇÃO DOS DADOS ABNT NBR 16167	21
2.2 LEI GERAL DE PROTEÇÃO DOS DADOS PESSOAIS	25
2.2.1 Disposições preliminares	26
2.2.2 Requisitos para o tratamento de dados pessoais	26
2.2.3 Direitos do titular	27
2.2.4 Tratamento de dados pelo poder público	29
2.2.5 Transferência internacional de dados	30
2.2.6 Responsabilidade dos agentes de tratamento	31
2.2.7 Segurança e boas práticas	31
2.2.8 Fiscalização	33
2.3 ABNT NBR ISO/IEC 27701	33
2.4 ABNT NBR ISO/IEC 25020	42
2.5 ABNT NBR ISO/IEC 25030	45
2.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO	54
<b>3 FERRAMENTAS DE MAPEAMENTO DE DADOS</b>	<b>55</b>
3.1 CARACTERÍSTICAS DAS FERRAMENTAS DE MAPEAMENTO DE DADOS	55
3.2 EXEMPLOS E COMPARAÇÃO DE FERRAMENTAS DE MAPEAMENTO DE DADOS	57
3.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO	69
<b>4 PROPOSTA DE SOLUÇÃO</b>	<b>70</b>
4.1 DEFINIÇÃO DAS FERRAMENTAS DE MAPEAMENTO DE DADOS	70
4.2 CRITÉRIOS	71
4.3 MÉTRICAS	74
4.4 CASOS DE TESTES	79
4.5 VALIDAÇÃO DA PROPOSTA	85
4.6 QUALIDADE DOS DADOS	86
4.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO	89
<b>5 TESTE DAS FERRAMENTAS</b>	<b>91</b>

5.1 TALEND OPEN STUDIO	91
5.1.1 Caso de Teste 1: Parametrização da ferramenta	91
5.1.2 Caso de Teste 2: Avaliação de riscos dos dados	94
5.1.3 Caso de Teste 3: Segurança de Acesso	95
5.1.4 Caso de Teste 4: Identificação de dados sensíveis	95
5.1.5 Caso de Teste 5: Rotulação dos dados	96
5.1.6 Caso de Teste 6: Criptografia dos dados	96
5.1.7 Caso de Teste 7: Armazenamento de eventos	100
5.1.8 Caso de Teste 8: Resposta a incidentes de segurança	101
5.1.9 Caso de Teste 9: Análise crítica técnica do compliance	101
5.1.10 Caso de Teste 10: Entrega de dados ao titulares	102
5.1.11 Caso de Teste 11: Anonimização	103
5.1.12 Caso de Teste 12: Coexistência	105
5.2 PENTAHO KETTLE	108
5.2.1 Caso de Teste 1: Parametrização da ferramenta	108
5.2.2 Caso de Teste 2: Avaliação de riscos dos dados	110
5.2.3 Caso de Teste 3: Segurança de Acesso	111
5.2.4 Caso de Teste 4: Identificação de dados sensíveis	111
5.2.5 Caso de Teste 5: Rotulação dos dados	111
5.2.6 Caso de Teste 6: Criptografia dos dados	112
5.2.7 Caso de Teste 7: Armazenamento de eventos	114
5.2.8 Caso de Teste 8: Resposta a incidentes de segurança	116
5.2.9 Caso de Teste 9: Análise crítica técnica do compliance	117
5.2.10 Caso de Teste 10: Entrega de dados ao titulares	118
5.2.11 Caso de Teste 11: Anonimização	119
5.2.12 Caso de Teste 12: Coexistência	120
5.3 CLOVERDX	121
5.3.1 Caso de Teste 1: Parametrização da ferramenta	122
5.3.2 Caso de Teste 2: Avaliação de riscos dos dados	124
5.3.3 Caso de Teste 3: Segurança de Acesso	125
5.3.4 Caso de Teste 4: Identificação de dados sensíveis	127
5.3.5 Caso de Teste 5: Rotulação dos dados	127
5.3.6 Caso de Teste 6: Criptografia dos dados	127
5.3.7 Caso de Teste 7: Armazenamento de eventos	127
5.3.8 Caso de Teste 8: Resposta a incidentes de segurança	128
5.3.9 Caso de Teste 9: Análise crítica técnica do compliance	129
5.2.10 Caso de Teste 10: Entrega de dados ao titulares	130
5.3.11 Caso de Teste 11: Anonimização	131
5.3.12 Caso de Teste 12: Coexistência	132
5.4 CONSIDERAÇÕES FINAIS	133

<b>6 AVALIAÇÃO DAS FERRAMENTAS</b>	<b>135</b>
6.1 ADEQUAÇÃO	135
6.2 ACURÁCIA	136
6.3 INTEROPERABILIDADE	137
6.4 CONFORMIDADE	138
6.5 RECUPERABILIDADE	139
6.6 OPERACIONALIDADE	140
6.7 COEXISTÊNCIA	141
6.8 EFETIVIDADE	142
6.9 PRODUTIVIDADE	143
6.10 CONSIDERAÇÕES FINAIS	143
<b>7 CONCLUSÕES</b>	<b>146</b>
<b>REFERÊNCIAS</b>	<b>149</b>

## 1 INTRODUÇÃO

O aumento da influência da tecnologia vem crescendo de forma mais rápida nos últimos anos, o que vem gerando uma grande quantidade de informação. A informação é um conjunto de dados relacionados entre si, que levam determinado conhecimento (SEMIDÃO, 2014).

Como qualquer ativo, os dados devem estar seguros, respeitando os princípios da segurança da informação: confidencialidade, integridade e disponibilidade. A confidencialidade, determina que os dados fiquem protegidos de acessos não autorizados. A integridade é essencial para que os dados se mantenham em sua forma original, sem modificações indevidas. A disponibilidade é assegurar que os dados sejam acessíveis quando for necessário.

As organizações vêm acumulando maiores quantidades de dados com o passar do tempo. Até 2020 a previsão é que se alcance o volume de 44 zettabytes de informações armazenadas (COMPUTERWORLD, 2018). Esse grande acúmulo de dados acaba gerando um problema para as organizações que cada vez necessitam ter maior controle e segurança do que armazenam, e para o cidadão que pode ter suas informações acessadas de forma indevida ou vazadas. Conforme relatório da Kaspersky(2020), em 2019 ocorreu um aumento de 72% no aumento de vazamento dos dados através de malwares. Conforme um estudo conduzido pelo Instituto Ponemon, a violação dos dados ainda tem um impacto financeiro na média global de US \$3,8 milhões para as companhias, no período de agosto de 2019 e abril de 2020 (IBM, 2020).

Além das perdas financeiras, a exposição indevida de dados levaram a União Europeia a criar RGPD(Regulamento Geral sobre a Proteção de Dados) que entrou em vigor no dia 25 de maio de 2018. Esta que regulamenta a privacidade e proteção de dados pessoais, aplicável a todos os indivíduos na União Europeia, também é aplicável a empresas estabelecidas fora da União Europeia, que tratam de dados referentes a pessoas do bloco. O RGPD tem como objetivo dar aos cidadãos formas de controlar os seus dados pessoais e promover maior segurança.

No Brasil foi criada a LGPD (Lei Geral de Proteção de Dados Pessoais), que entrou em vigor dia 26 de agosto de 2020, e se fundamenta no respeito à privacidade. A LGPD cria novos conceitos jurídicos para os tipos de dados que são tratados e os separa em dados pessoais: dados sensíveis e dados anonimizados. As organizações que controlam esses dados devem garantir que o titular dos dados saiba para qual finalidade será utilizado, e que a transmissão e o controle de acesso aos dados sejam seguros, assim como a proteção ao vazamento dos mesmos e de sua integridade (BRASIL, 2018).

De acordo com a nova Lei, as empresas devem proteger os dados pessoais. No entanto, compreender e relacionar corretamente os dados se torna um processo importante para estar em conformidade com a lei, através da implementação de políticas, processos e softwares apropriados para gerenciar a forma de coletar, processar, analisar, armazenar, compartilhar, reutilizar e eliminar esses dados. A Figura 1 mostra uma tabela, onde resume esse ciclo dos dados que devem ocorrer conforme a LGPD. Esse ciclo ajuda a entender e garantir maior segurança dos dados.

Figura 1 - Ciclo de vida dos dados segundo a LGPD

CICLO DE VIDA DOS DADOS	
FASE DO CICLO	COM A LGPD
Coleta	Os dados pessoais coletados devem obedecer ao princípio da sua finalidade.
Processamento	O processamento de dados só pode ser realizado se o tratamento estiver dentro das hipóteses do Artigo 7º da LGPD.
Análise	A análise de dados deve levar em consideração a finalidade da coleta. Devem ser obedecidos os princípios de tratamento, com propósito legítimo, específico e explícito.
Compartilhamento	O compartilhamento de dados deve ser consentido pelo seus titulares (pessoa natural a quem se refere os dados pessoais).
Armazenamento	Os dados pessoais devem ser armazenados e mantidos por prazos definidos, ou seja, até que a finalidade seja alcançada ou deixem de ser necessários.
Reutilização	Quando ocorrer a troca de finalidade o titular deve ter consentimento da mudança.
Eliminação	Os dados pessoais devem ser eliminados após o término do seu tratamento.

Fonte: Próprio autor.

A quantidade de dados que as empresas trabalham é tão grande, que muitas vezes torna-se difícil realizar a identificação desses dados. Nesse contexto é



importante a realização do mapeamento de dados. Compreendendo o ciclo dos dados, é possível assimilar o que necessita ser protegido, e quais políticas de segurança aplicar a cada tipo de dado, protegendo as informações pessoais e tendo maior transparência.

Um estudo realizado pela Cisco (CISCO, 2019) sobre privacidade dos dados, mostrou que as empresas que investem em ferramentas de proteção dos dados, tiveram aumento em 41% de ganho em vantagem competitiva frente a outras organizações e ainda 39% mitigaram as perdas por vazamento de dados.

As ferramentas de mapeamento de dados são fundamentais para auxiliar nessas classificações dentro da organização. Elas fazem o processo de mapeamento por meio da comparação entre modelos de dados distintos. Assim, é possível verificar todo o fluxo dos dados, ou seja, em quais etapas os dados são coletados, armazenados e quais as suas classificações. Além disso, esse mapeamento permite verificar se cada tipo de dado está armazenado no local adequado, e quais controles de segurança são aplicados, facilitando a aderência das obrigações da LGPD.

A partir das informações coletadas no processo do mapeamento de dados, os mesmos podem ser utilizados para a elaboração de documentos como: relatório de impacto de proteção de dados, política de gestão de crises, manual de procedimentos e controles internos em proteção de dados. Atualmente, muitas ferramentas de mapeamento estão no mercado, dentre elas, algumas mais populares são de grandes empresas como por exemplo, Dell Boomi, Xplenty e CloverDX (G2, 2020) que atendem até mesmo outras funções relacionadas a banco de dados. Algumas ferramentas em código aberto conhecidas são DXMapper e Talend Open Studio (HOFFMAN, 2020).

## 1.1 PROBLEMA DE PESQUISA

Levando-se em consideração que juntamente com a transformação digital ocorrem uma grande quantidade de crimes cibernéticos, relacionados a vazamento de dados e informações, as organizações precisam entender uma forma correta de

se proteger e se adequar a Lei Geral de Proteção de Dados Pessoais(LGPD). Todavia, o aumento cada vez maior desses volumes de dados e dos processos de coleta e transformação, dificultam ainda mais para que as organizações tenham o controle total sobre eles.

Diante desse contexto, existem ferramentas de mapeamento de dados que são fundamentais para detectar os fluxos dos dados e realizar o inventário dos mesmos, assim sendo essencial para o auxílio da proteção da informação e da adequação a LGPD para a organização.

Entretanto, diante das possibilidades oferecidas e disponíveis a decisão de escolher a ferramenta de mapeamento de dados ideal para a organização deve ser tomada com muito cuidado. O projeto de implantação tem uma alta magnitude e envolve o engajamento de diversas partes do negócio, porque trata-se de manipular médios e grandes volumes de dados que necessitam ter processos de TI bem estruturados para resolverem lacunas encontradas no mapeamento, tornando-se necessário escolher a ferramenta correta.

**Questão de pesquisa:** Qual(is) a ferramenta(s) de mapeamento de dados, possui(em) aderência com a Lei Geral de Proteção de Dados(LGPD) e a norma ABNT NBR ISO/IEC 27701?

## 1.2 OBJETIVO

Analisar ferramentas destinadas a mapeamento de dados, que permitam o controle do ciclo de vida dos dados dentro da empresa, de acordo com a Lei Geral de Proteção de Dados(LGPD) e a norma ABNT NBR ISO/IEC 27701.

### 1.2.1 Objetivos Específicos

Os objetivos específicos a serem abordados no trabalho são:

- a) Aprofundar os conhecimentos sobre a adequação a LGPD;
- b) Definir critérios de avaliação de ferramentas de mapeamento de dados de acordo com a norma ABNT NBR ISO/IEC 27701;
- c) Testar e avaliar ferramentas de mapeamento de dados.

### 1.3 METODOLOGIA

A metodologia de desenvolvimento do trabalho consiste em um estudo teórico, abrangente e exploratório da literatura disponível sobre o tema a ser estudado.

Para isso, os estudos da literatura bibliográfica buscam aprofundar o conhecimento sobre a implementação de softwares de mapeamento de dados e seus funcionamentos, auxiliando a atingir os objetivos previamente definidos. A metodologia de desenvolvimento desse trabalho pode ser dividida em 4 etapas, sendo elas:

a) 1ª Etapa: Estudar e compreender os aspectos jurídicos da LGPD, paralelamente à norma de privacidade dos dados da ABNT NBR ISO/IEC 27701, que tem o objetivo de manter um padrão estabelecido e relacionado ao tratamento e a proteção dessas informações. Além de estudos referentes a ferramentas de mapeamento de dados;

b) 2ª Etapa: Definir quais ferramentas de mapeamento de dados desenvolvidos em código aberto a serem analisados e testados. A definição será feita a partir de estudos recentes que realizam a comparação das ferramentas desta categoria;

c) 3ª Etapa: Definir os critérios de avaliação das ferramentas. Para definição será utilizado as normas ABNT ISO/IEC 25020 e 25030 que tem o objetivo de normatizar os requisitos de qualidade e avaliação de produtos de software.

d) 4<sup>a</sup> Etapa: Teste e análise das ferramentas de acordo com os critérios definidos.

#### 1.4 ESTRUTURA DO TRABALHO

O Capítulo 2 apresenta um resumo da LGPD, abordando os requisitos, responsabilidades e deveres impostas pela mesma. Também trata sobre as normas relacionadas à privacidade dos dados (ABNT NBR ISO/IEC 27701), e a norma sobre a classificação de dados (ABNT NBR ISO/IEC 16167).

O Capítulo 3 aborda as ferramentas de mapeamento de dados, descreve suas características, e exemplifica ferramentas que podem ser aplicadas nas organizações

O Capítulo 4 trata sobre a proposta de solução para o problema abordado no Capítulo 1 deste trabalho, apresenta os estudos a serem realizados e as métricas e critérios para a avaliação das ferramentas de mapeamento de dados.

O Capítulo 5, expõe a execução dos casos de testes aplicados e cada ferramenta selecionada.

O Capítulo 6 apresenta os resultados obtidos na execução do trabalho, contendo as análises das avaliações de cada ferramenta de mapeamento de dados, e a que melhor se adere para solucionar o problema apresentado.

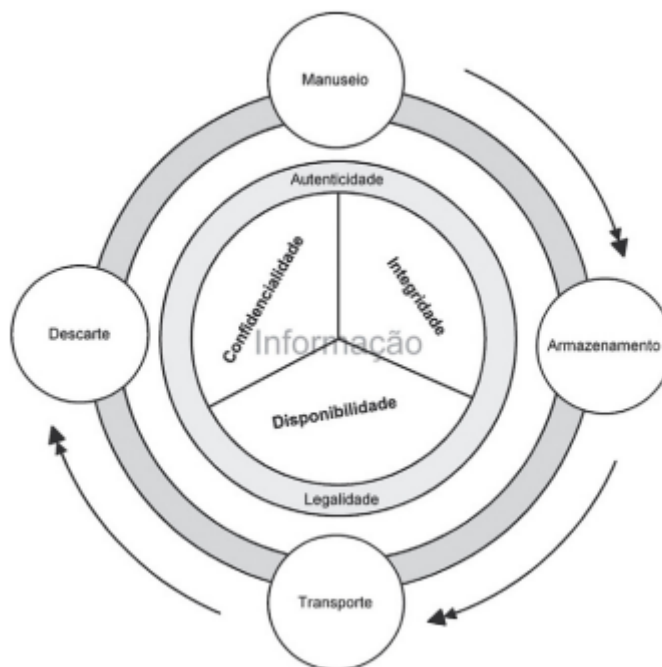
Por último, o Capítulo 7 apresenta a conclusão final deste trabalho.

## 2 FUNDAMENTAÇÃO TEÓRICA

Toda organização trabalha com um conjunto de variáveis para elaboração de suas estratégias corporativas. Dessa forma, quanto melhor o entendimento dos dados que possui e compõem suas atividades, maior será o conhecimento para a tomada de decisão (MATSUMOTO, 2006).

Os dados são elementos significativos para as organizações e essenciais para melhorar os resultados, obtendo um planejamento estratégico, por exemplo, todos os dados possuem um ciclo de vida, que é identificado pelos momentos vividos e que podem colocá-lo em risco. Os momentos vividos ocorrem quando são feitos o uso de dados, através de processos que fazem a operação da organização (SÊMOLA, 2003). Portanto, os momentos do ciclo dos dados são o manuseio, armazenamento, transporte e descarte (Figura 2).

Figura 2 - Momentos do ciclo dos dados



Fonte: Sêmola(2003).

O manuseio é caracterizado pela criação e manipulação dos dados, por exemplo, a coleta feita através do preenchimento de um formulário. O

armazenamento é o momento onde o dado é armazenado, seja em um banco de dados ou em alguma outra mídia. O transporte é a parte onde o dado é movido para outro local para ser acessado. O descarte é o momento final onde ele deve ser eliminado.

Como citado anteriormente, os dados devem ser protegidos em todo o ciclo de vida. Para proteger o dado, é necessário compreender alguns conceitos que são apresentados neste capítulo. A Seção 2.1 é referente a classificação dos dados através da ABNT NBR ISO/IEC 16167. A Seção 2.2 apresenta um resumo da LGPD(Lei Geral de Proteção de Dados Pessoais). A Seção 2.3 mostra a norma ABNT NBR ISO/IEC 27701 e os requisitos para implementar um sistema de gestão da privacidade dos dados. A Seção 2.4 apresenta a ABNT NBR ISO/IEC 25020 com modelos de qualidade de software. A Seção 2.5 mostra a ABNT NBR ISO/IEC 25030 e os requisitos de qualidade de software. Por fim, a Seção 2.6 apresenta considerações finais do capítulo.

## 2.1 CLASSIFICAÇÃO DOS DADOS ABNT NBR 16167

A norma ABNT NBR ISO/IEC 16167(ABNT,2013) estabelece diretrizes básicas para a classificação, tratamento, e rotulação das informações conforme a suas características de sensibilidade e criticidade para a organização. Tem como objetivo estabelecer níveis adequados para a proteção dos dados.

A grande quantidade de dados mantidos pelas organizações, são recursos de vários processos que mantêm as necessidades do negócio. Assim as pessoas envolvidas com os processos devem possuir somente acesso às informações necessárias, diretamente ou indiretamente, no desenvolvimento das atividades de trabalho e de suas responsabilidades.

Os dados que são de propriedade da organização, devem ser classificados de acordo com o nível de sensibilidade que representam para o negócio. Deve ser considerando os requisitos de legalidade, criticidade, vida útil, restrição e a análise dos riscos e do impacto para o negócio. Em todo momento que um dado é criado na organização ele já deve ter sua classificação correta. Caso ele passe por mudanças

durante o seu ciclo de vida, deve ocorrer mudanças em seus níveis de classificação também, para assim garantir a sua segurança.

Desse modo a ABNT NBR ISO/IEC 16167 sugere uma referência para a classificação dos dados na organização , dividido entre 4 níveis. No Quadro 1 se apresentam os 4 níveis de classificação existentes na norma ABNT NBR ISO/IEC 16167.

Quadro 1 - Níveis de classificação

<b>Níveis de classificação</b>	<b>Características básicas</b>
Nível 1	São os dados que podem ser divulgados publicamente. Normalmente a divulgação deste tipo de informação é de responsabilidade de áreas específicas que fazem interações com o público externo, como por exemplo, as áreas de comunicação e marketing.
Nível 2	São dados internos e que podem ser divulgados a todos os colaboradores e prestadores de serviços, desde que estejam comprometidos com a confiabilidade da informação.
Nível 3	Os dados já são restritos e devem ser divulgados somente a determinados grupos, áreas ou cargos.
Nível 4	São dados que requerem um tratamento especial e cuja divulgação não autorizada ou acessos indevidos podem gerar prejuízos financeiros, legais, normativos, contratuais ou na reputação da imagem, ou estratégias da organização.

Fonte: ABNT NBR 16167(2013)

A classificação serve como referência. A organização pode adicionar mais níveis e fazer alterações nos existentes. Não é recomendado fazer um esquema muito complexo devido a poder engessar o processo e conseqüentemente o fluxo

dos dados. Também não deve ser muito simples e com poucos níveis de classificação, pois podem levar a perda dos dados devido ao tratamento errado.

A norma recomenda que o nível de classificação atribuído aos dados, deve aparecer na rotulação, junto com o grupo de acesso. O processo de rotulação dos dados é de responsabilidade de quem realizou a criação ou alteração dos mesmos. Em caso de informações com múltiplas classificações dos dados, deve ser rotulado pela classificação mais restritiva que houver.

A rotulação sempre deve ser visível em arquivos físicos como documentos e pastas. Deve aparecer sempre na capa e também em cabeçalhos e rodapés, se possível. Nos meios digitais deve ser rotulado individualmente pelo nome do arquivo, seguido de sua classificação, ou pode ser rotulado diretamente no seu local de armazenamento.

Após o tratamento adequado dos dados, serão realizados os controles e a proteção adequada, assim garantindo a confidencialidade, a integridade e a disponibilidade. Nesse processo deve ser identificado cenários que acontecem nos processos diários da organização envolvendo os dados. Para cada tipo de cenário é estabelecido orientações básicas para o tratamento, conforme os níveis de classificação determinado.

Os cenários envolvem a produção, recepção, utilização, acesso, reprodução, transporte, transmissão, distribuição, destinação, arquivamento, armazenamento e eliminação dos dados. Todos os cenários formam um senso coletivo para o tratamento dos dados, de modo que qualquer área ou pessoa execute o mesmo tratamento para diversas situações.

A norma apresenta uma referência para o estabelecimento destas orientações de tratamento, considerando o cenário e os níveis de classificação. Alguns cenários podem ser observados no Quadro 2.



Quadro 2 - Tratamento de informações

(continua)

Cenários	Nível 1	Nível 2	Nível 3	Nível 4
Acesso lógico ou físico	Sem restrição	Somente para os colaboradores da Organização e prestadores de serviço	Somente pessoas do grupo de acesso	Somente pessoas do grupo de acesso
Armazenamento em arquivos digitais	Sem restrição	Somente nos servidores de arquivos na rede da Organização	Somente nos servidores de arquivos na rede da organização e com controle de acesso	Somente nos servidores de arquivos na rede da Organização e com controle de acesso. Convém que seja considerado o uso de criptografia.
Transporte Físico	Sem restrição	Sem restrições dentro das dependências da Organização. Somente com autorização do proprietário dos dados, se o transporte for para fora da Organização.	Somente com utilização de lacres, caso o transporte não seja realizado por alguém do grupo de acesso. Necessita de autorização do proprietário dos dados se for para fora das dependências da Organização. Armazenar em locais protegidos durante viagens	Somente com utilização de lacres. Caso o transporte não seja realizado por alguém do grupo de acesso, usar o serviço de entrega em mãos. O transporte para fora da Organização necessita autorização do proprietário dos dados. Armazenar em locais com chaves ou cofres durante viagens
Transmissão por e-mail	Sem restrição	Interno, sem restrições. Para fora da Organização é necessário autorização do proprietário dos dados.	Somente para o grupo de acesso. Para fora do grupo de acesso, é necessário a autorização do proprietário dos dados	Para o grupo de acesso. Para fora do grupo de acesso, é preciso autorização do proprietário da informação. Devem ser consideradas técnicas de proteção, como a criptografia.

(conclusão)

Eliminação de arquivos de computador	Sem restrição	Excluir da pasta onde está arquivada	Excluir da pasta onde está arquivada e da lixeira também	Excluir da lixeira dos dispositivos e adotar soluções tecnológicas visando garantir que os dados não possam ser recuperadas
--------------------------------------	---------------	--------------------------------------	--	---

Fonte: ABNT NBR ISO/IEC 16167(2013)

## 2.2 LEI GERAL DE PROTEÇÃO DOS DADOS PESSOAIS

A LGPD(Lei Geral de Proteção dos Dados Pessoais) ou Lei 13.709, que entrou em vigor dia 26 de agosto de 2020, visa fortalecer a proteção das informações pessoais e a transparência na forma de tratamento e armazenamento dos dados. A lei se fundamenta nos pontos do que se diz a respeito aos direitos fundamentais do cidadão, como privacidade, liberdade de expressão, informação, inviolabilidade da intimidade, honra e da imagem.

Para melhor compreensão dos deveres e responsabilidades que são apresentados, é importante conhecer as seguintes definições:

- a) titular: pessoa natural a quem se refere os dados pessoais;
- b) controlador: pessoa natural ou jurídica, a quem compete as decisões sobre o tratamento dos dados;
- c) operador: pessoa natural ou jurídica que realiza o tratamento dos dados em nome do controlador;
- d) encarregado: pessoa natural, indicada pelo controlador, que atua como canal de comunicação entre controlador e titulares, ou entre controlador e a autoridade nacional.

### **2.2.1 Disposições preliminares**

As normas gerais contidas na Lei Geral de Proteção de Dados Pessoais são de interesse nacional e devem ser observadas pela União, Estados, Distrito Federal e Municípios.

Para o cumprimento da Lei deve ser observado a finalidade da realização do tratamento de dados com propósitos legítimos. O titular precisa ser informado e consentir de forma explícita sobre a finalidade da utilização dos seus dados, o tratamento não pode ser além disto. O titular ainda deve ter o acesso sobre a forma e a duração do tratamento do seus dados, assim como, a qualidade dos dados deve ser garantida quanto a exatidão, clareza, relevância e atualização, conforme a finalidade proposta. Deve seguir as medidas técnicas e administrativas para se manter a segurança dos dados, assim protegendo dos ataques e dos vazamentos. Sempre que solicitado pelos agentes de tratamento devem ser mostradas as medidas de segurança empregadas, para validar o cumprimento dos requisitos da Lei.

### **2.2.2 Requisitos para o tratamento de dados pessoais**

A aplicação da Lei ocorre para qualquer tipo de tratamento de dados, seja por empresas nacionais ou não, tanto de direito público como privado, e até mesmo pessoa física, independentemente do meio, do país de sua sede ou do país onde estejam localizados os dados. Desde que o tratamento seja realizado em território nacional, ou a atividade de tratamento tenha por objetivo a oferta ou fornecimento de bens ou serviços de indivíduos localizados no território nacional, ou os dados pessoais objetos do tratamento tenham sido coletados no território nacional. Há exceção na aplicação da Lei para quando o tratamento é realizado por pessoa física, para fins particulares e não econômicos, ou fins exclusivamente jornalísticos, artísticos e acadêmicos, porém deve-se sempre existir o consentimento do titular. Também ocorre a exceção nos casos que a finalidade seja para a defesa nacional, segurança pública ou do Estado, e em atividade de investigação.

Caso ocorra uma necessidade em compartilhar os dados com outros controladores, que é a pessoa física ou jurídica que compete o tratamento de dados, deve ser solicitado novamente ao titular o consentimento de uso dos dados, deixando de forma clara a nova finalidade do uso.

Quando a coleta tem fins de pesquisa, os órgãos responsáveis devem efetuar o tratamento, desde que, sempre que possível seja garantido a anonimização dos dados pessoais, que é utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo

A lei ainda aborda o tratamento de dados pessoais sensíveis, que são aqueles que trazem elementos como opinião, ou caráter do titular como raça, etnia, religião, opinião política, dados referente a saúde ou vida sexual, dados genéticos ou biométricos. Estes dados têm requisitos de tratamento similares às regras de dados pessoais, porém nos casos de dados coletados para pesquisa, onde é solicitado a anonimização dos dados.

O tratamento dos dados pessoais deve possuir um estado que caracterize seu término. O término do tratamento deve ocorrer quando a finalidade desejada for alcançada, ou o período para o tratamento que foi estipulado quando coletado o dado se encerrar, ou ainda se houver a comunicação do titular expressando revogar o consentimento fornecido para o tratamento.

Depois do encerramento do tratamento, os dados pessoais devem ser eliminados, apenas podendo ser conservado em pesquisas, ou que sejam anonimizados para uso restrito do controlador.

### **2.2.3 Direitos do titular**

Qualquer momento o titular dos dados pessoais tem o direito de solicitar as seguintes informações perante o controlador do tratamento de dados:

- a) confirmação da existência do tratamento;

- b) acesso aos dados armazenados;
- c) correção ou atualização de dados armazenados;
- d) anonimização, bloqueio ou eliminação de dados desnecessários;
- e) portabilidade de dados, desde que exista requisição expressa para tal;
- f) eliminação dos dados, respeitando os requisitos dispostos na Lei;
- g) informação das entidades públicas e privadas com as quais existiu uso compartilhado de dados;
- h) informação sobre a possibilidade de não consentimento para uso dos dados e as consequências caso isso ocorra;
- i) revogação do consentimento, conforme regras dispostas na Lei.

Sempre que solicitado o acesso aos dados pessoais ou a informações sobre esses tratamento de dados, os mesmos devem ser disponibilizados de forma simples ou completa, incluindo a origem dos dados, critérios de tratamento, no prazo de quinze dias a contar da data de solicitação. O titular tem direito de realizar a solicitação sem nenhum custo por parte do controlador. Sempre que houver alterações ou atualizações dos dados, o responsável pelo compartilhamento deve informar imediatamente os agentes que fazem o uso dos dados.

O armazenamento de dados necessita ser feito de forma a favorecer o fornecimento das informações quando solicitadas pelo titular. As informações podem ser fornecidas de forma eletrônica ou impressa, sendo transmitidas de forma segura. O titular pode, também, solicitar de maneira integral, cópia dos dados pessoais em posse do controlador. Essa cópia precisa ser entregue em um formato que auxilie a utilização posterior, até mesmo em outras operações de tratamento.

Em casos de tratamento automatizado de dados, o titular pode solicitar correções dos critérios utilizados para tal, inclusive quando tratado de definições de perfil pessoal e profissional, ou perfil de consumo e de crédito. O controlador deve sempre providenciar informações claras referente aos critérios utilizados durante o tratamento automatizado dos dados. Os dados do titular jamais podem ser utilizados para causar o seu prejuízo.

#### **2.2.4 Tratamento de dados pelo poder público**

O tratamento de dados pessoais pelo poder público, deverá ser realizado para o atendimento da finalidade pública, no ato do seu interesse público. Com o objetivo de executar as competências legais ou cumprir as atribuições do serviço público.

Deve informar claramente a finalidade, procedimento adotados e a previsão legal e práticas utilizadas para o tratamento dos dados. As informações devem ser de fácil acesso, e de preferência no site da instituição relacionada.

Serviços de cartórios notariais e de registro que são empresas privadas mas a serviço do poder público, terão o mesmo tratamento de empresas de direito público. Os mesmos devem informar a administração pública por meio eletrônico, acesso aos seus dados.

Empresas públicas com sociedades de economia mista, que atuam no regime de concorrência, terão o mesmo tratamento disposto nas pessoas jurídicas de direito privado no termos da Lei.

Instituições de direito público devem manter os dados de forma interoperável e estruturados de forma a facilitar o compartilhamento. O compartilhamento de dados com empresas de direito privado só pode ocorrer quando a finalidade for exclusiva a execução de políticas públicas. Para este compartilhamento dos dados pessoais do poder público com empresas de direito privado, é preciso informar a autoridade nacional, e com a dependência do consentimento do titular dos dados.

A autoridade nacional poderá solicitar, aos órgãos do poder público a realização de operações de tratamento de dados pessoais, informações específicas

sobre a natureza dos dados e outros detalhes do tratamento realizado e poderá emitir parecer técnico para garantir o cumprimento desta Lei.

### **2.2.5 Transferência internacional de dados**

Quando ocorre a transferência de dados de forma internacional, a lei estipula os seguintes requisitos:

- a) Para países que proporcionem grau de proteção de dados conforme o previsto na Lei;
- b) Quando o controlador garantir e comprovar, através de contratos, normas corporativas ou selos, certificados e códigos de conduta, que cumpre os requisitos da Lei;
- c) Quando a finalidade do tratamento for de cooperação jurídica internacional ou a proteção da vida do titular ou terceiro;
- d) Sob autorização da autoridade nacional;
- e) Quando o resultado for compromisso assumido em acordo de cooperação internacional;
- f) Para cumprimento de políticas públicas ou atribuições legais;
- g) Sob o consentimento específico do titular.

A autoridade nacional define a avaliação do nível de proteção de dados do país estrangeiro alvo da transferência de dados, observando as normas em vigor.

Conforme a Lei, a responsabilidade dos agentes de tratamento, é de manter registros das operações de tratamento realizadas, que são a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração.

### **2.2.6 Responsabilidade dos agentes de tratamento**

O controlador, operador e encarregado que são os responsáveis pelo tratamento dos dados, possuem algumas responsabilidades específicas conforme a Lei. Todos precisam manter os registros das operações de tratamento de dados que realizam, especialmente quando baseados no legítimo interesse.

O controlador, quando solicitado pelo autoridade nacional, deverá elaborar um relatório de impacto à proteção dos dados pessoais, referente a suas operações de tratamento de dados. Este relatório deve conter no mínimo, a descrição dos dados coletados, a metodologia utilizada para a coleta, e para a garantia da segurança da informação, junto com a análise do controlador com relação às medidas adotadas para a mitigação dos riscos.

O operador deve realizar o tratamento de dados conforme as instruções fornecidas pelo controlador, não podendo agir de forma diferente, sob pena das sanções da Lei.

O encarregado é responsável pela comunicação com os titulares, recebendo reclamações e tomando as devidas providências. Também quando questionado pela autoridade nacional, deve tomar as providências cabíveis. Deve orientar os funcionários e contratados da organização sobre as práticas adotadas a respeito da proteção de dados pessoais. Ainda o encarregado deve ter sua identidade e informações de contato sempre disponíveis publicamente, de preferência no site do controlador.

### **2.2.7 Segurança e boas práticas**

A Lei também apresenta, alguns aspectos de segurança e boas práticas, que devem ser respeitados para a garantia da inviolabilidade dos dados pessoais, como a implantação de programa de governança de privacidade que tenha a finalidade de proteger os dados de acessos não autorizados, situações acidentais ou ilícitas de destruição de dados, perda, alteração, vazamento ou tratamento inadequado dos dados.



Os dados devem estar mantidos em uma infraestrutura com todas medidas técnicas e administrativas cabíveis para se manterem seguros, até mesmo após o término do tratamento de dados. Qualquer incidente de segurança envolvendo os dados armazenados deve ser comunicado à autoridade nacional. Deve informar as seguintes informações:

- a) qual a ocorrência;
- b) informações sobre os titulares envolvidos no incidente;
- c) as medidas de segurança utilizadas para proteção dos dados;
- d) os riscos relacionados com o incidente;
- e) motivos da demora para comunicação do incidente, se acaso ocorrer;
- f) meios para mitigar ou reverter os efeitos do prejuízo.

Também é permitido por Lei que os controladores e operadores, individualmente ou por meio de associações, formulem regras de boas práticas e governança. Estas estabelecem as condições de organização, regime de funcionamento, os procedimentos, as reclamações e petições de titulares, normas de segurança, padrões técnicos, as obrigações específicas para os envolvidos no tratamento, ações educativas e mecanismos de supervisão e mitigação dos riscos relacionados ao tratamento dos dados. Ao implementar tal programa deve ser respeitado no mínimo os seguintes itens:

- a) comprometimento do controlador em adotar políticas que assegurem o cumprimento da Lei;
- b) aplicabilidade a todo o conjunto de dados pessoais em seu poder;
- c) adaptação a sua estrutura;
- d) objetivo de estabelecer relação de confiança com o titular;
- e) integração com sua estrutura geral de governança
- f) existência de um plano de respostas a incidentes de segurança da informação;

- g) demonstrar efetividade em seu programa de governança.

### **2.2.8 Fiscalização**

Os agentes de tratamento dos dados, quando cometem infrações às normas previstas na Lei, ficam sujeitos a sanções administrativas que são aplicadas pela autoridade nacional. Estas que são aplicadas após o procedimento administrativo que possibilita a oportunidade de defesa.

As seguintes sanções podem ser aplicadas:

- a) advertência, com indicação de prazo para adotar medidas corretivas;
- b) multa simples, de até 2% do faturamento líquido da pessoa jurídica de direito privado, grupo ou conglomerado no Brasil no seu último exercício, limitada no total de R\$ 50.000.000,00 por infração;
- c) multa diária;
- d) tornar público a infração, após ser devidamente apurada e confirmada;
- e) bloqueio dos dados pessoais a que se refere a infração até sua regularização;
- f) eliminação dos dados pessoais a que se refere a infração;
- g) suspensão parcial do funcionamento do banco de dados em que se refere a infração pelo período máximo de 6 meses, até a regularização das atividades de tratamento;
- h) suspensão do exercício da atividade dos dados pessoais que se refere a infração pelo período de 6 meses;
- i) proibição parcial ou total do exercício de atividades relacionadas ao tratamento de dados.

### **2.3 ABNT NBR ISO/IEC 27701**

A grande maioria das organizações fazem o tratamento de dados pessoais. A necessidade de proteção da privacidade desses dados, se tornou uma necessidade para a sociedade moderna.

A ABNT NBR ISO/IEC 27701 é um complemento das normas ABNT NBR ISO/IEC 27001 e ABNT NBR ISO/IEC 27002 (ABNT, 2019).

A ABNT NBR ISO/IEC 27001(ABNT, 2013), representa um padrão de utilização para um gerenciamento de um sistema de gestão de segurança da informação(SGSI). A norma fornece e apresenta requisitos para que uma organização possa estruturar seu SGSI. Ela agrega um processo de escalonamento de riscos e valorização dos ativos, conduzindo para uma análise e identificação dos riscos e como implantar controles com o objetivo de minimizá-los. A Figura 3 apresenta um diagrama com os requisitos da norma.

Figura 3 - Área e requisitos da ABNT NBR ISO/IEC 27001



Fonte: ABNT NBR ISO/IEC 27001 (2013)

Enquanto a ABNT NBR ISO/IEC 27001 define a implementação de um SGSI, a ABNT NBR ISO/IEC 27701 adiciona requisitos com o objetivo de ampliar para um Sistema de Gestão de Privacidade da Informação (SGPI).

Algumas diretrizes para a implementação do SGPI são os mesmos requisitos que a ABNT NBR ISO/IEC 27001 determina. O Quadro 3 mostra a relação entre os requisitos das normas ABNT NBR ISO/IEC 27001 e ABNT NBR ISO/IEC 27701. É possível verificar através da primeira coluna, as seções que a norma estabelece. A coluna descrição apresenta um breve detalhamento sobre a seção da norma ABNT NBR ISO/IEC 27001. A coluna comentário especifica os requisitos que a ABNT NBR ISO/IEC 27701 complementa a norma ABNT NBR ISO/IEC 27001.

Quadro 3 - Requisitos específicos da SGPI com relação a ABNT NBR ISO/IEC 27001

(continua)

<b>Seção</b>	<b>Descrição</b>	<b>Comentário</b>
Contexto da Organização	Requisitos para entendimento de assuntos externos e internos e das partes interessadas.	Com a SGPI a organização deve compreender os fatores externos e internos dentro do seu contexto para poder alcançar o objetivo intencionado. É necessário determinar o papel de um controlador de dados ou operador de dados conforme determina a LGPD.
Liderança	Estabelece papéis e define responsabilidade	Sem requisitos específicos para a SGPI
Planejamento	Define requisitos para avaliação de risco, tratamento de riscos e define objetivos de segurança da informação	Com a SGPI deve ser assegurado que os dados pessoais participem de todos os processos de avaliação, identificação e gerenciamento dos riscos.
Apoio	Define requisitos de disponibilidade de recursos, competências, comunicação e controle de documentos	Sem requisitos específicos para a SGPI

(conclusão)

Operação	Implementações de avaliação e tratamento de riscos, assim como controles de processos necessários para atingir os objetivos de segurança da informação	Sem requisitos específicos para a SGPI
Avaliação de desempenho	Etapas de verificação e definição dos requisitos de monitoramento, medição análise, avaliação e auditoria	Sem requisitos específicos para a SGPI
Melhoria	Define requisitos para não conformidades e de ações para melhoria contínua	Sem requisitos específicos para a SGPI

Fonte: ABNT NBR ISO/IEC 27701(2019)

A norma ABNT NBR ISO/IEC 27002 é um código de práticas e um conjunto completo de controles para apoiar a aplicação de um SGSI. Tem o objetivo de estabelecer diretrizes e princípios para iniciar, implementar, manter e melhorar a gestão da segurança da informação em uma organização (ABNT, 2013; FERNANDES e ABREU, 2008). A norma é dividida em 14 áreas (Quadro 4).

Quadro 4 - Seções específicas pela ABNT NBR ISO/IEC 27002

(continua)

<b>Seção</b>	<b>Descrição</b>
Política de segurança da informação	Documento que deve ser criado sobre a política de informação da empresa e que deve conter e estabelecer os conceitos de segurança da informação.
Organização da segurança da informação	Estabelece uma estrutura para gerenciar a segurança da informação da organização, e define as responsabilidades das atividades relacionadas à segurança da informação.
Segurança em recursos humanos	Deixar funcionários e terceiros cientes das ameaças relacionadas à segurança da informação.

(conclusão)

Gestão de ativos	Identificação e classificação dos ativos. Assim mantendo uma estrutura de inventário para que se possa ter controle do que é protegido.
Controle de acesso	Prevenir acesso não autorizado aos recursos, e monitorar os acessos. Com o objetivo de assegurar danos e fraudes dos recursos.
Criptografia	Diretrizes para proteger dados através de técnicas de criptografia.
Segurança física e do ambiente	Equipamentos e instalações físicas devem ser mantidos em áreas seguras e apropriadas.
Segurança nas operações	Definição de procedimentos para gestão da operação dos recursos de processamento de informação. Tem o objetivo de planejar recursos para minimizar riscos.
Segurança nas comunicações	Definição de procedimento para a gestão da comunicação. Tem o objetivo de garantir segurança nos meios de comunicação.
Aquisitivo, desenvolvimento e manutenção de sistemas	Identificação dos requisitos do sistema de segurança da informação, visando a sua proteção durante os procedimentos de desenvolvimento, implementação ou manutenção.
Relacionamento na cadeia de suprimento	Diretrizes para o relacionamento com os fornecedores.
Gestão de incidentes de segurança da informação	Procedimentos para registro e notificações que devem ser estabelecidos para eventos relacionados à segurança da informação.
Aspectos da segurança da informação na gestão da continuidade do negócio	Estabelecimento de planos para continuidade dos negócios, visando impedir a interrupção das atividades do negócio.
Conformidade	Garantir a conformidade da organização com leis, contratos e obrigações.

Fonte: ABNT NBR ISO/IEC 27002(2013)

A ABNT NBR ISO/IEC 27701 adiciona diretrizes a alguns pontos já tratados pela ABNT NBR ISO/IEC 27002.

A seção de política de segurança da informação da ABNT NBR ISO/IEC 27002 acaba sendo estendida para que as organizações criem uma política específica para a privacidade dos dados, e que essa tenha comprometimento com a regulamentação de proteção dos dados pessoais.

Assim como a seção de organização da segurança da informação é incrementada pela necessidade de designar um responsável para comunicação entre a organização e o titular dos dados pessoais. Também precisa ser estabelecido uma equipe ou pessoa para ser responsável no desenvolvimento, implementação, manutenção e monitoramento do programa de privacidade da organização assegurando a sua aplicabilidade. Na LGPD esta pessoa ou equipe é chamada de encarregado de proteção de dados.

A seção de segurança em recursos humanos fica complementada para que a organização implemente medidas de conscientização, para possibilitar aos colaboradores que saibam sobre como realizar notificações de incidentes, e estejam cientes das possíveis consequências para a organização ou ao titular dos dados pessoais.

Na seção de gestão de ativos a classificação da informação da organização deve ser implementada para os dados pessoais. É de extrema importância para possibilitar entender os tipos de dados tratados e armazenados pela organização. Toda pessoa que possui envolvimento com os dados deve ter consciência da definição e identificação do que é um dado pessoal.

A seção de segurança física e de criptografia da ABNT NBR ISO/IEC 27002 também ficam estendidas pela ABNT NBR ISO/IEC 27701, no que diz sobre o uso de mídias físicas para armazenar os dados pessoais é recomendado a utilização da criptografia dos dados pessoais. No momento do descarte de qualquer dispositivo de armazenamento que contenham dados pessoais deve ser realizado processos que deixem os dados inacessíveis e não recuperáveis, além de ser documentado o processo. Quando a transferência dos dados pessoais é realizada por uma rede não segura, como por exemplo a internet pública, tudo deve ser criptografado.

O controle de acesso aos dados pessoais precisa ser rígido. Deste modo a seção fica estendida para que as organizações mantenham atualizado todos os perfis de usuários que tenham esse tipo de acesso, e garantir que cada pessoa tenha seu acesso individual. Usuários que possam estar comprometidos devem ter seu acesso cancelado, cabe também a organização fazer verificações com frequência em busca de credenciais não utilizadas. As pessoas que operam os dados pessoais precisam ter um acordo obrigatório de confiabilidade.

A seção de segurança nas comunicações da ABNT NBR ISO/IEC 27002 tem adicionado em sua norma, a necessidade de registrar os logs e estabelecer processos automáticos e manuais para identificar possíveis irregularidades. Quando possível, deve ser registrado quem, quando e qual dado pessoal foi acessado, alterado ou excluído. A retenção desses registros deve ser maior que o tempo necessário para o tratamento dos dados pessoais. Cabe à organização implementar um esquema de retenção para garantir conformidade com a legislação e não manter por tempo além do necessário.

Na gestão de incidentes de segurança, a organização deve estabelecer responsabilidades e normas internas para a identificação e registro das violações dos dados pessoais. Sempre que houver um incidente as autoridades devem ser notificadas conforme o tempo estabelecido pela regulamentação vigente. Nesta notificação deve conter e se manter registrado as seguintes informações:

- a) Contato para obtenção de maiores informações;
- b) Descrição da violação e probabilidade das consequências;
- c) Período de tempo que ocorreu;
- d) Relatório com indivíduos envolvidos e registros relacionados;
- e) Medidas e ações planejadas a serem tomadas.

Quando possível os titulares dos dados também devem ser notificados do ocorrido. Quando um operador dos dados identificar qualquer violação deve notificar imediatamente o controlador dos dados pessoais, para que o mesmo tome as ações devidas.



Como análise crítica a organização precisa prover auditorias para verificação das medidas de segurança, e testes de vulnerabilidades e invasões. O monitoramento contínuo deve ser feito para garantir que o tratamento dos dados esteja sendo executado dentro do permitido.

A norma ABNT NBR ISO/IEC 27701 também cria diretrizes próprias para o SGPI. A organização deve assegurar que os titulares dos dados pessoais tenham total clareza e entendimento de como será realizado o tratamento de dados. A declaração de consentimento, políticas e procedimentos deve ser clara e bem detalhada, não havendo ambiguidade. O consentimento deve ser registrado e documentado, de forma que possa ser fornecido em caso de eventuais situações.

Na coleta um limite precisa ser aplicado, proporcional para o propósito necessário.

O tratamento dos dados gera determinado risco ao titular dos dados pessoais. Nesse contexto precisa ser implementado uma avaliação de impacto da privacidade, através de elementos como tratamento em alta escala e dados sensíveis. A organização pode determinar mais elementos para realizar uma avaliação mais completa do impacto da privacidade.

O tratamento de dados necessita ser documentado e mantido em um inventário do tratamento de dados pessoais que a organização utiliza. Deve ser documentado os seguintes itens:

- a) Tipo de tratamento;
- b) Propósito para o tratamento;
- c) Descrição do tipo de categoria dos dados pessoais e seus titulares;
- d) Destinatários para quem os dados pessoais será divulgado, como filiais em outros países;
- e) Descrição das medidas de segurança tomadas;
- f) Relatório de avaliação de impacto de privacidade.

Quando solicitado pelo titular as informações precisam ser entregues a ele de forma clara e acessível. Sempre que os propósitos do tratamento forem mudados ou

estendidos devem ser informados ao titular, assim como transferências ou na utilização do uso de tomada de decisão automatizada.

Mecanismos que permitam que o titular dos dados possa cancelar ou modificar o consentimento deve ser implementado. Os titulares devem ficar cientes que podem cancelar o seu consentimento a qualquer momento, e é obrigação da organização proporcionar mecanismos para que isso aconteça. Por exemplo, se o método de coleta dos dados for por website ou e-mail, convém que o mecanismo de cancelamento seja o mesmo.

Também é necessário prover mecanismo que permite aos usuários dos dados pessoais alterar ou excluir determinados dados. Quando isso ocorrer deve ser replicado imediatamente para os demais utilizadores dos dados pessoais caso sejam compartilhados. Cabe à organização criar políticas para ter precisão ao informar os demais utilizadores.

Caso o titular solicite uma cópia dos dados deve ser disponibilizado para o mesmo. Em caso de dados que foram anonimizados não se deve realizar processos de re-identificação dos mesmos, pois podem causar risco à privacidade de demais dados pessoais. A organização precisa definir um padrão de documentos e procedimentos para a resposta de solicitações dos titulares dos dados pessoais.

Cabe a organização manter uma política e procedimento para o descarte dos dados pessoais. É necessário escolher técnicas de descarte de dados pessoais conforme a granularidade dos dados e da mídia física, a capacidade em se recuperar os dados excluídos, a natureza dos dados armazenados e as características físicas na qual os dados são armazenados.

A norma ABNT NBR ISO/IEC 27701 também apresenta em seu anexo A os controles necessários para que um controlador dos dados pessoais possa implantar o SGPI. O anexo B apresenta os controles fundamentais para a implantação do SGPI, caso a organização atue como operador dos dados pessoais.

O anexo C fornece um mapeamento de conformidade entre a norma ABNT NBR ISO/IEC 27701 com a norma ABNT NBR ISO/IEC 29100, está que trata sobre uma estrutura para a proteção de dados pessoais, firmando aspectos técnicos e organizacionais dentro da estrutura, ajudando a organização a determinar seus

requisitos de privacidade a partir de um entendimento sobre proteção de dados privados.

A norma ABNT NBR ISO/IEC 27701 também apresenta no seu anexo NA um mapeamento sobre a LGPD, a norma atende todos os requisitos solicitados pela LGPD, assim facilita o processo da organização em se adequar com a Lei. Usar a ISO 27701 para estender seus esforços de segurança no gerenciamento de privacidade, ajudará a demonstrar conformidade com a Lei em casos de auditoria pela agência nacional. Também é possível demonstrar a conformidade que a organização segue a norma para o mercado e a comunidade.

#### 2.4 ABNT NBR ISO/IEC 25020

A norma ABNT NBR ISO/IEC 25020(ABNT, 2009) faz parte de uma série de normas 25000 SQuaRE que abrange os processos complementares de especificação, medição e avaliação dos requisitos. O objetivo é auxiliar os produtos de software em processos de desenvolvimento e aquisição com especificações dos requisitos de qualidade. O SQuaRE estabelece critérios de especificação e avaliação dos requisitos de qualidade do produto. Inclui um modelo de qualidade para aderir as definições de qualidade do cliente com as características do produto de software. Esta série de normas SQuaRE são distribuídas nas seguintes divisões:

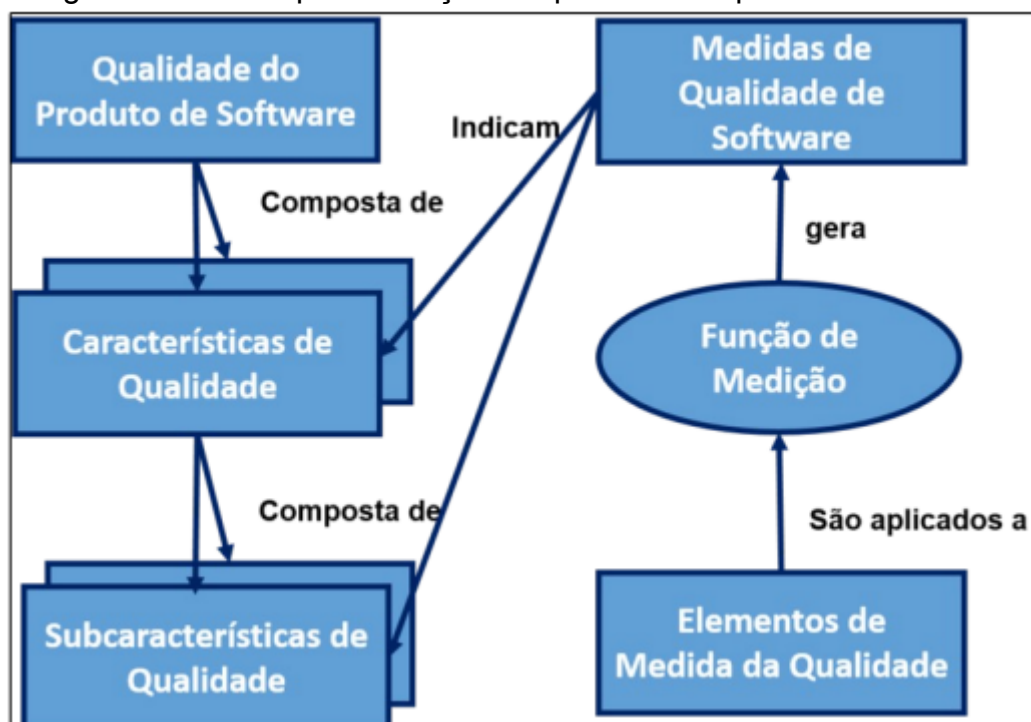
- a) Divisão de Gestão da Qualidade (ISO / IEC 2500n);
- b) Divisão de Modelo de Qualidade (ISO / IEC 2501n);
- c) Divisão de Medição de Qualidade (ISO / IEC 2502n);
- d) Divisão de Requisitos de Qualidade (ISO / IEC 2503n);
- e) Divisão de Avaliação de Qualidade (ISO / IEC 2504n).-

A norma ABNT NBR ISO/IEC 25010 define termos para características de qualidade do produto de software, e estas características são decompostas em sub características. A ABNT NBR ISO/IEC 25020 fornece informações para medição da qualidade de produto de software, através de suas características e subcaracterísticas das qualidades associadas e aos atributos de produto de

software. A norma fornece uma referência como modelo para a medição das características definidas na ABNT NBR ISO/IEC 25010.

A Figura 4 mostra que as medidas da qualidade de software são construídas pela aplicação de uma função de medição, a elementos de medida da qualidade ao longo do ciclo de vida do produto.

Figura 4 - Modelo para medição de qualidade de produto de software



Fonte: ABNT NBR ISO/IEC 25020 (2009)

Os critérios devem ser apresentados junto com o resultado da avaliação das medidas selecionadas em relação aos critérios. A ABNT NBR ISO/IEC 25030 fornece orientações com relação às especificações dos requisitos de qualidade de software.

São apresentados três tipos diferentes de medidas da qualidade do software para corresponder ao ciclo de vida da qualidade de produto de software. Medidas de qualidade de software internas são aplicadas a determinada parte do produto de software durante o estágio de desenvolvimento. Essa medição fornece aos usuários a habilidade de medir a qualidade de produtos intermediários, assim pode ser usado

para prever a qualidade de um produto final. Permite aos desenvolvedores tomarem ações de correção durante os ciclos de desenvolvimento.

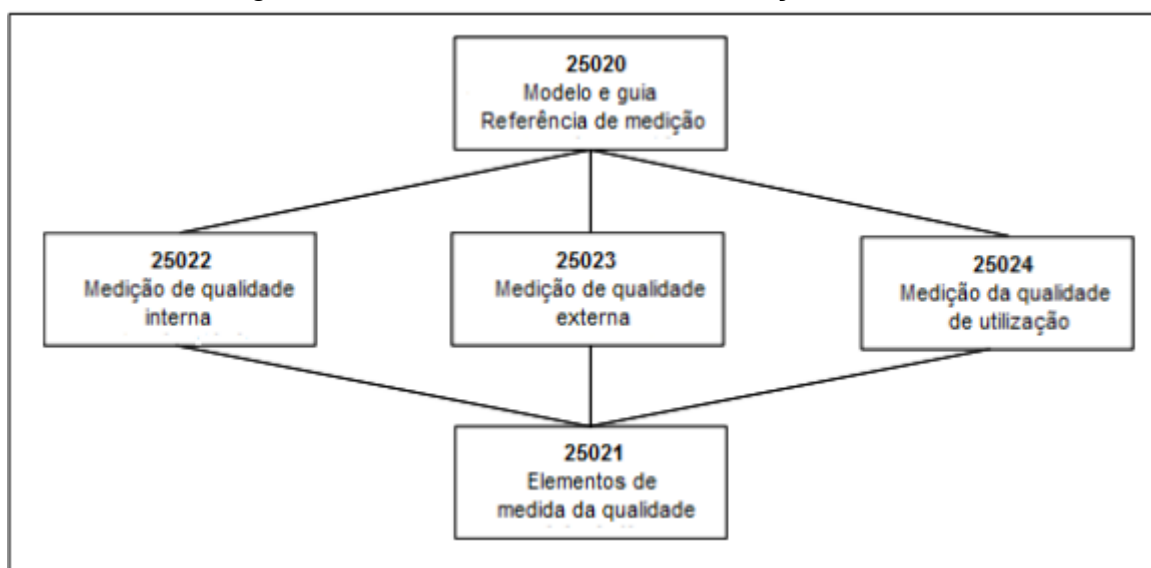
A forma de medida da qualidade externa do software, é utilizada para medir a qualidade do produto com base no comportamento do sistema em que o software faz parte. Esse tipo de medição é utilizado durante estágios de testes e de operação durante o ciclo de vida do produto.

As medidas da qualidade em uso medem quanto o produto satisfaz as necessidades dos usuários, em relação a suas metas específicas de negócios. Estas medidas são realizadas em um ambiente real de operação.

Durante o ciclo de vida do produto, essas medidas de qualidade de software são executadas e utilizadas para auxiliar o desenvolvimento do produto, apoio ou avaliação do produto.

A ABNT NBR ISO/IEC 25020 também é subdividida em algumas partes conforme a Figura 5. Um exemplo é a ABNT NBR ISO/IEC 25021 que oferece componentes de medida de qualidade que podem ser usados para construir medidas de qualidade de software. Estes componentes podem medir uma representação estática de software, um comportamento ou seus efeitos de utilização do software.

Figura 5 - Estrutura da divisão de Medições da Qualidade.



Fonte: ABNT NBR ISO/IEC 25020 (2009)

## 2.5 ABNT NBR ISO/IEC 25030

Como a norma ABNT NBR ISO/IEC 25020 a norma ABNT NBR ISO/IEC 25030 (ABNT, 2008) também faz parte da série de normas SQuaRE. Esta série contém normas distribuídas nas seguintes divisões:

- a) Divisão de Gestão da Qualidade (ISO / IEC 2500n);
- b) Divisão de Modelo de Qualidade (ISO / IEC 2501n);
- c) Divisão de Medição de Qualidade (ISO / IEC 2502n);
- d) Divisão de Requisitos de Qualidade (ISO / IEC 2503n);
- e) Divisão de Avaliação de Qualidade (ISO / IEC 2504n).

A ABNT NBR ISO/IEC 25030 leva a característica de qualidade de uso, que mensura a capacidade do software em proporcionar aos usuários atingirem suas metas com produtividade, satisfação, eficácia e segurança.

A identificação dos requisitos são necessários para o planejamento, especificação, desenvolvimento e avaliação do software. É de grande importância identificar os requisitos de qualidade de software como parte de um produto final. Este software geralmente faz parte de um sistema maior. Esta norma tem o objetivo nos requisitos de qualidade de software, através de uma perspectiva de sistema. Um sistema é definido como combinação de componentes interativos organizados que tendem a atingir uma ou mais finalidades.

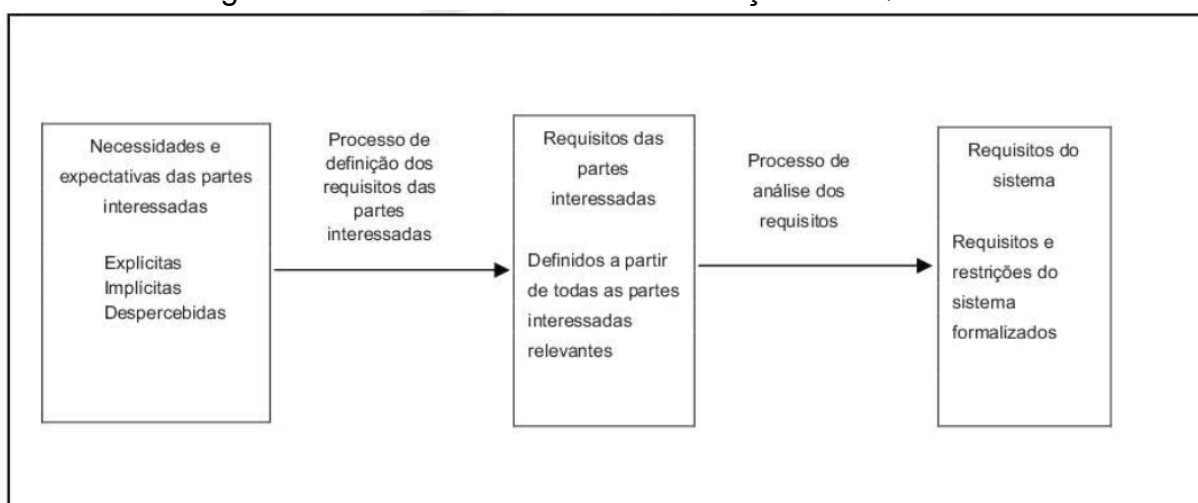
Durante o ciclo de vida de um sistema existe interesse de uma variedade de partes interessadas. Estas partes interessadas incluem todas as pessoas, organizações, que de alguma forma possuem interesse no sistema buscando atender suas expectativas e necessidades.

As necessidades podem ser indicadas de forma implícita ou explícita. As necessidades implícitas são entendidas como as que representam as expectativas com base no uso do produto de software, ou em rotinas de trabalhos existentes de operação do negócio. Já as necessidades explícitas são caracterizadas quando as partes interessadas não têm consciência de todas suas necessidades. Na maioria

das situações são expressadas somente quando o usuário testar o software ou algum protótipo.

As necessidades das partes interessadas são identificadas através do processo de definição e da análise dos requisitos, conforme é mostrado na Figura 5. Nesse processo é levado em consideração todas as vontades, desejos e expectativas das partes interessadas.

Figura 6 - Estrutura da divisão de Medições da Qualidade



Fonte: ABNT NBR ISO/IEC 25030 (2008)

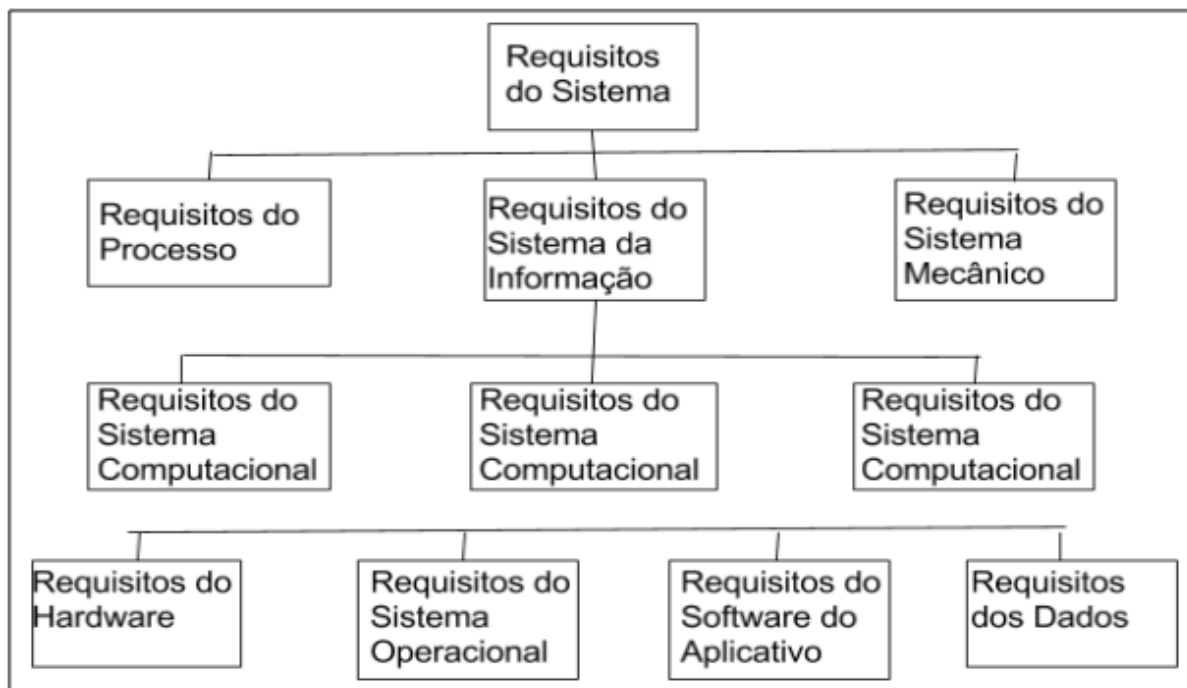
O resultado do processo de definição é chamado de requisitos das partes interessadas, e o resultado do processo de análise é chamado de requisitos do sistema.

Um método de análise irá transformar os requisitos das partes interessadas para uma visão técnica dos requisitos do sistema, que será utilizado para produzir o sistema desejado. Essa visão técnica é chamada de requisitos do sistema. Estes que indicarão quais características o sistema deve possuir para satisfazer os requisitos das partes interessadas.

Cada requisito do sistema deve ser formulado para cada componente diferente do sistema. Os requisitos não devem ser vistos isoladamente, mas de uma visão ampla onde se possa ver todos os requisitos e componentes do sistema.

Os requisitos das partes interessadas nem sempre podem implicar em todos os requisitos. Pode haver requisitos alternativos como de hardware, software ou de algum processo que ocorra de forma manual. A Figura 7 mostra a hierarquia de requisitos que deve ser atendida como base.

Figura 7 - Hierarquia dos requisitos do sistema e do software



Fonte: ABNT NBR ISO/IEC 25030 (2008)

Os requisitos de qualidade do software sempre devem estar relacionados com as características e subcaracterísticas de qualidade conforme determina o modelo aplicado. É necessário registrar para quais funções de software o requisito de qualidade vai ser aplicado. Deve ser documentado todos os critérios e medidas a serem utilizados, assim como os limites funcionais e limitações de implementação do software.

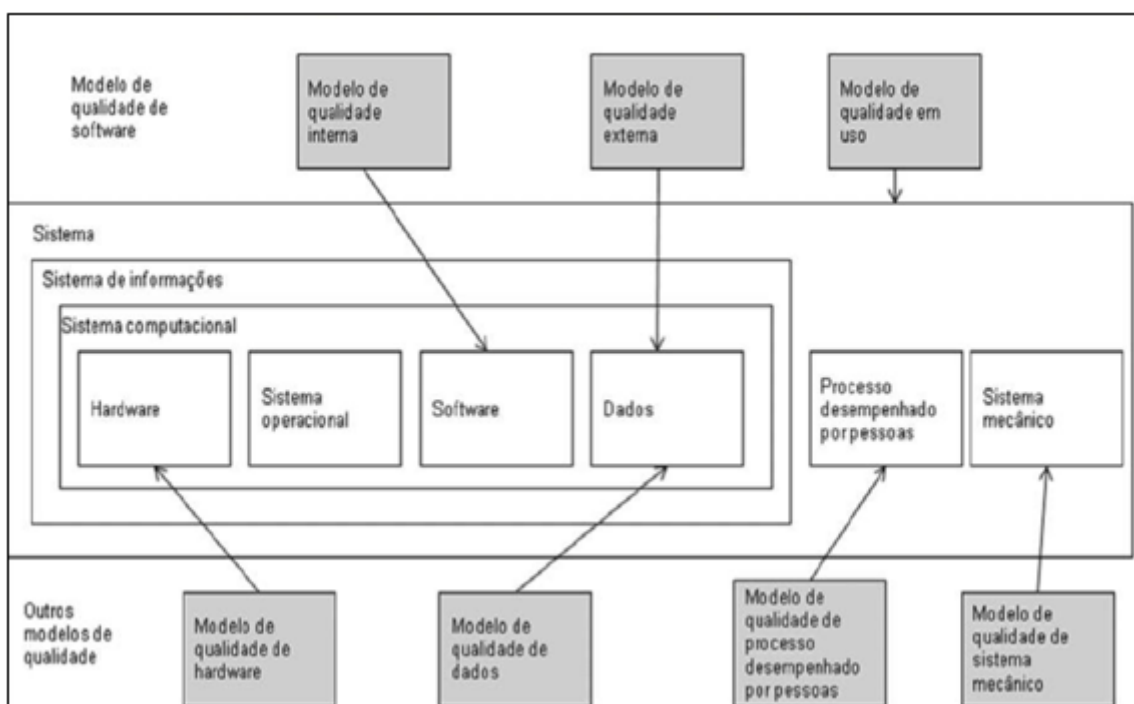
O modelo de qualidade define três diferentes perspectivas de qualidade:

- a) Qualidade do software em uso;
- b) Qualidade externa de software;
- c) Qualidade interna do software.



A qualidade de uso é quando o software está em seu ambiente operacional, realizando tarefas dos usuários. A qualidade externa discute propriedades relacionadas com a execução do software em hardware e aplicação de um sistema operacional. A qualidade interna está relacionada diretamente com as propriedades do software. A Figura 8 apresenta a interação entre os modelos apresentados.

Figura 8 - Exemplo de modelo de sistemas e modelos de qualidade



Fonte: ABNT NBR ISO/IEC 25030 (2008)

A sessão de medidas de qualidade do software são padrões derivativos das normas ISO/IEC 9126 e 14598. Desse modo as propriedades do software podem ser avaliadas através de medição direta, indireta ou medição das consequências, tudo depende dos critérios utilizados para a avaliação do produto de software, e nas definições da aplicação de medidas práticas de qualidade interna, externa e de uso.

A norma apresenta e indica o uso de qualidade de uso, que é o modelo para avaliação de qualidade utilizado para as definições dos critérios de avaliação dos softwares e das definições das metas de qualidade. Assim tende a mensurar a capacidade do software de atingir as metas dos usuários com produtividade, efetividade, satisfação e segurança.

Não é possível alocar todos os recursos para medir todas as características e subcaracterísticas, sendo elas internas e externas de todas partes de um software. Desta mesma forma não é funcional medir a qualidade em uso para todos os cenários de uso. Assim, os objetivos de negócio e os processos utilizados nas avaliações dos critérios e a serem utilizados devem definir o seguimento do projeto.

Deve ainda ser determinado métricas que relacionam as características do software. Qualquer propriedade interna ou externa, que interage com o ambiente do software pode se tornar uma métrica.

Existem as métricas internas que são os indicadores para avaliar o software, através de medições considerando as duas características próprias internas, sem a execução de programas. Já as métricas externas também servem para avaliar o software, de modo que a medição considere o seu comportamento do sistema ou o ambiente como um todo. As métricas de uso são indicadores do mesmo modo para a avaliar o software, através de medições com cenários e tarefas do dia a dia dos usuários.

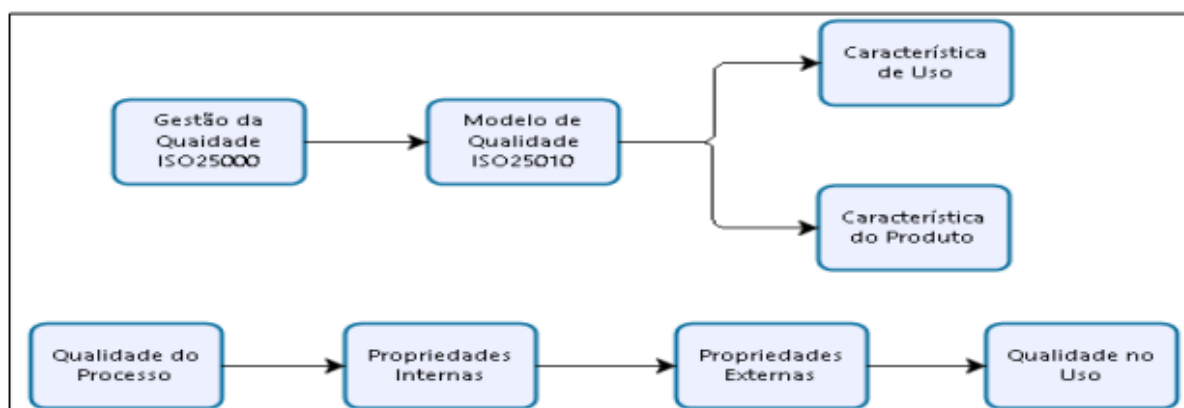
Para definir qual métrica usar como base, depende das metas selecionadas e das necessidades de avaliação do produto de software. Para a avaliação entre softwares precisa ser estabelecido métricas rigorosas, e os processos de mensuração devem possuir precisão suficiente para que seja estabelecido o processo de comparação entre eles.

O resultado das métricas devem estar em uma escala válida que seja conhecido e igual a todos. E que seja reprodutível, gerando resultados das mesmas medidas, independente que gerado por pessoas ou ocasiões diferentes. Para a comparação de softwares, o relatório dos resultados deve esclarecer se as métricas são objetivas ou empíricas. As métricas consideradas objetivas devem possuir um procedimento que demarca o número ou categoria do atributo. Para serem consideradas empíricas devem alcançar os dados através de observações ou questionários.

O recomendado para a correta avaliação de qualidade do produto de software, é a definição de um modelo de qualidade(Figura 9). Este modelo deve ser

utilizado para o estabelecimento dos critérios de avaliação dos softwares e de metas de qualidade.

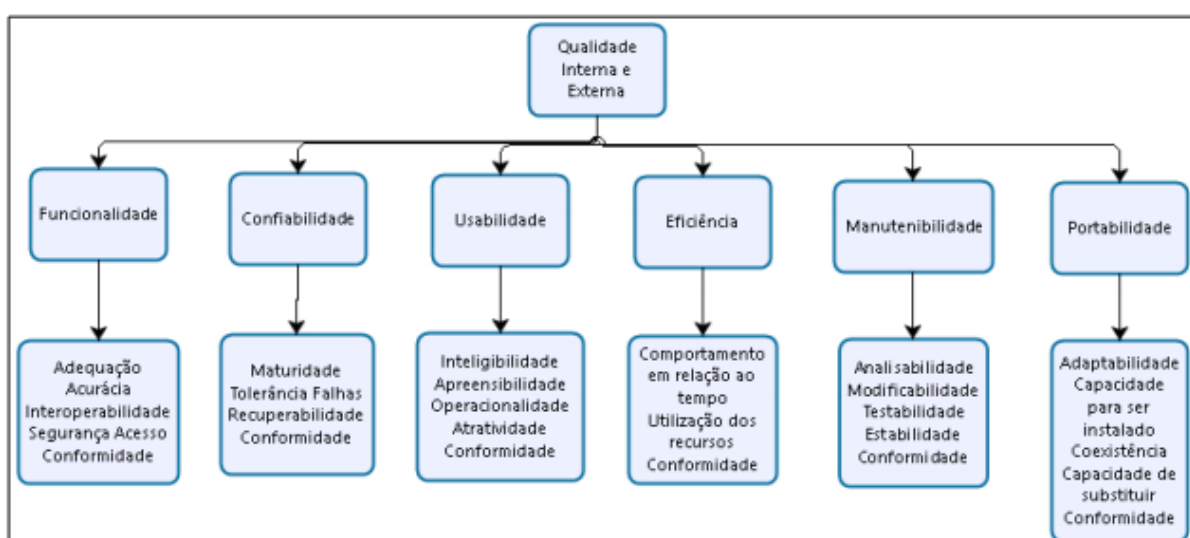
Figura 9 - Modelo de qualidade



Fonte: ABNT NBR ISO/IEC 25010 (2011)

Segundo a abordagem da ABNT NBR ISO/IEC 25010 (2011) há uma divisão de modelo de qualidade do produto de software definido em seis características de qualidade. A Figura 10 apresenta as características e subcaracterísticas de qualidade interna e externa.

Figura 10 - Características e Subcaracterísticas do Modelo de Qualidade



Fonte: ABNT NBR ISO/IEC 25010 (2011)

As características e subcaracterísticas de qualidade interna e externa podem ser entendidas pelos seguintes itens:

a) Funcionalidade: Capacidade do software em disponibilizar funções que satisfaçam as necessidades quando utilizado em condições específicas:

- Adequação: se o software faz o que ele deveria fazer;
- Acurácia: se o software faz o que ele promete corretamente;
- Interoperabilidade: se o software interage com os outros sistemas;
- Segurança de Acesso: se o software não permite acesso não autorizado a dados e programas;
- Conformidade: se o software está de acordo com as normas e leis.

b) Confiabilidade: Capacidade do software de manter um bom nível de desempenho quando utilizado em condições específicas:

- Maturidade: apresenta a frequência que o software apresenta falhas;
- Tolerância a Falhas: com que flexibilidade o software reage às falhas;
- Recuperabilidade: se o software consegue recuperar os dados em caso de falhas;
- Conformidade: se o software está de acordo com os padrões e normas de confiabilidade.

c) Usabilidade: Capacidade do software em ser entendido, assimilado, utilizado e atraente ao ponto de vista do usuário:

- Inteligibilidade: se é fácil de entender o conceito e a aplicação do software;
- Apreensibilidade: se é fácil de aprender a utilizar o software;
- Operacionalidade: se é fácil de operacionalizar e controlar o software;

- Atratividade: se o software é atrativo aos usuários;

- Conformidade: se o software está de acordo com os padrões e normas de usabilidade.

d) Eficiência: Capacidade do software de manter o desempenho adequado em condições explícitas:

- Comportamento em relação ao tempo: se o software tem um bom tempo de resposta e velocidade de execução;

- Utilização dos recursos: se o software utiliza muitos recursos;

- Conformidade: se o software está de acordo com os padrões e normas de eficiência.

e) Manutenibilidade: Capacidade do software em ser alterado. As alterações podem ser melhorias, correções ou adaptações do software:

- Analisabilidade: se é fácil de detectar as falhas do software;

- Modificabilidade: se é fácil modificar e adaptar o software;

- Testabilidade: se é fácil testar as mudanças feitas no software;

- Estabilidade: se há riscos ao fazer alterações no software;

- Conformidade: se o software está de acordo com os padrões e normas de manutenibilidade.

f) Portabilidade: Capacidade do software em ser migrado de ambiente:

- Adaptabilidade: se é possível adaptar o software a outros ambientes;

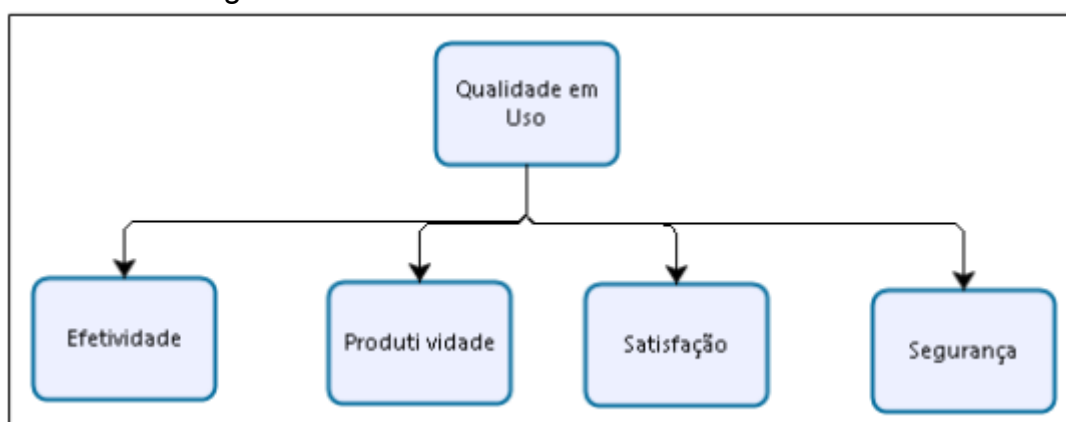
- Capacidade para ser instalado: se é fácil de instalar o software em outros ambientes;

- Coexistência: se pode coexistir com outros produtos independentes;

- Capacidade de Substituir: se é fácil substituir o software por outro;
- Conformidade: seu software está de acordo com padrões de portabilidade.

A ABNT NBR ISO/IEC 25030 também aborda e sugere o uso das características de qualidade em uso (Figura 11), que mensuram a capacidade do software em possibilitar que os usuários atinjam seus desejos com produtividade, efetividade e segurança.

Figura 11 - Características de Qualidade em Uso



Fonte: ABNT NBR ISO/IEC 25010 (2011)

- a) Efetividade é a capacidade que o software tem para fazer com que o usuário consiga alcançar seus objetivos de negócios, da melhor maneira e de forma completa.
- b) Produtividade é a capacidade do software em proporcionar que o usuário consiga utilizar a quantidade necessária de recursos com eficácia para que ele possa atingir um objetivo específico.
- c) Satisfação é a capacidade do software de satisfazer seu usuário enquanto ele o utiliza.
- d) Segurança é a capacidade do software de apresentar níveis aceitáveis de riscos para o cliente.

Tendo o conhecimento que é necessário alocar recursos para a avaliação dos softwares, nem sempre é possível medir todas as subcaracterísticas internas e externas de todas as partes do software.

Este capítulo auxiliará no processo de definição dos critérios de avaliação dos softwares testados com base nas características e subcaracterísticas, assim como nas formas de mensuração.

## 2.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

A norma ABNT NBR ISO/IEC 16167 auxilia a compreender e servir de referência para realização da classificação dos dados. Ela auxiliará na compreensão do ciclo de dados e para a classificação de cada tipo durante os testes com as ferramentas de mapeamento de dados .

A Lei Geral de Proteção de Dados Pessoais, refere-se à proteção, tratamento e uso de dados pessoais no Brasil. Ela objetiva um maior controle dos titulares sobre seus próprios dados pessoais que estão em poder das organizações. Fica possível verificar o que as organizações precisam implementar para se manter na regularidade. É necessário ter um panorama de como os dados são tratados e armazenados dentro das organizações. O estudo da LGPD será utilizado no trabalho de TCC nos processos de avaliação das ferramentas. Através das imposições e requisitos da Lei serão definidos critérios para realizar a avaliação das ferramentas de mapeamento de dados melhor.

A norma ABNT NBR ISO/IEC 27701 específica e fornece diretrizes para o estabelecimento, implementação, manutenção e melhoria contínua de um SGPI. Ela estabelece uma estrutura de privacidade dos dados, assim implementando técnicas e processos que auxiliam a adequação e aderência das organizações a LGPD. A compreensão da norma será utilizada no trabalho para definir os critérios que as ferramentas de mapeamento de dados devem atingir para entrar em conformidade com a LGPD.

As normas ABNT NBR ISO/IEC 25020 e 25030 definem as características desejadas para a seleção de um produto de software, e os processos e técnicas que

podem ser usados para garantir essas características. As normas ABNT NBR ISO/IEC 25020 e 25030 serão utilizadas no TCC para definir critérios e métricas para avaliação das ferramentas de mapeamento de dados.

Portanto, com o entendimento das normas citadas, o próximo capítulo apresenta as ferramentas de mapeamento de dados.

### **3 FERRAMENTAS DE MAPEAMENTO DE DADOS**

As organizações coletam dados de uma grande variedade de pontos. Com múltiplos pontos de coleta, os dados acabam armazenados em locais distintos, e em ferramentas que não se comunicam entre si. No passado, as organizações documentavam o mapeamento de dados em forma de papel, o que era suficiente para a época. Conforme os ambientes se tornaram mais complexos e em constante mudança, esse formato não conseguiu acompanhar o ritmo.

Neste capítulo é abordado as características e o funcionamento das ferramentas de mapeamento de dados (Seção 3.1). Na Seção 3.2 são apresentados exemplos e comparações de ferramentas. A seção 3.3 trata sobre as considerações finais do capítulo.

#### **3.1 CARACTERÍSTICAS DAS FERRAMENTAS DE MAPEAMENTO DE DADOS**

O mapeamento de dados é o processo de extrair campos de dados de um ou vários arquivos de uma origem e combiná-los com seus modelos de dados relacionados.

As ferramentas fazem o mapeamento de dados a partir da extração de dados de um ou mais repositórios de dados, e através da comparação entre modelos de dados consegue gerar uma visualização dos dados descobertos. O mapeamento de dados pode ser usado com um ampla variedade de tarefas de integração de dados, como (SHAHBAZ, 2016):

- a) Transformação de dados ou mediação de dados entre fontes e destino;
- b) Identificação de relacionamento de dados;



- c) Descoberta de dados sensíveis escondidos;
- d) Consolidação entre vários bancos de dados em um único banco de dados.

As tarefas de mapeamento de dados variam em complexidade, dependendo da hierarquia dos dados que estão sendo mapeados, bem como da disparidade entre os dados e o modelo de dados (CASTERS; BAUMAN; VAN DONGEN, 2010).

O maior desafio dos mapeadores de dados é descobrir como os dados fluem dos bancos de dados de origem até a interface do usuário final. Este fluxo determinará como os dados podem ser transformados até seu objetivo final.

As organizações possuem uma enorme gama de dados que são armazenados e gerenciados por vários tipos de banco de dados. As fontes de dados podem ser categorizados nas seguintes categorias (PRASAD; AGARWAL, 2016):

- a) Dados Estruturados: Gerados a partir de vários , CRM (*Customer Relationship Management*) e outros bancos de dados tradicionais;
- b) Dados semi estruturados: Dados formatados em XML, planilhas em Excel ou CSV.
- c) Dados Não-Estruturados: Não possuem estruturas bem definidas, alinhadas, padronizadas, podendo ser compostos por diversos elementos diferentes, desde texto e fotos.

O mapeamento pode ser realizado através de várias técnicas dependendo do software. Por exemplo, a técnica de mapeamento manual dos dados, requer que a equipe de dados codifique manualmente as conexões das fontes de dados para o software de mapeamento de dados. Normalmente, é escrito através de código em XSLT (*Extensible Stylesheet Language*). Conforme os dados crescem e se tornam mais complicados, o mapeamento manual não consegue acompanhar as necessidades e se torna preciso buscar outras soluções.

Outra técnica bastante comum é o mapeamento de dados por modelo, que é uma estratégia semiautomática que usa software para mapear os modelos de dados semelhantes (SHAHBAZ, 2016). O software compara as fontes de dados e o esquema de destino para gerar as conexões. Em seguida, é verificado o mapa e

realizado ajustes caso seja necessário. Geralmente estão presentes em softwares ETL que permitem que os usuários conectem os conjuntos de dados através de linhas.

Foram observadas no mercado ferramentas de mapeamento de dados que têm maior foco na integração dos dados, mesmo que o objetivo dessas ferramentas seja extrair, transformar e carregar dados de diversas fontes, podem ser utilizadas também com a finalidade de conformidade com a Lei. Esse grupo, por exemplo, é usado para facilitar o entendimento e construção de um *data warehouse*.

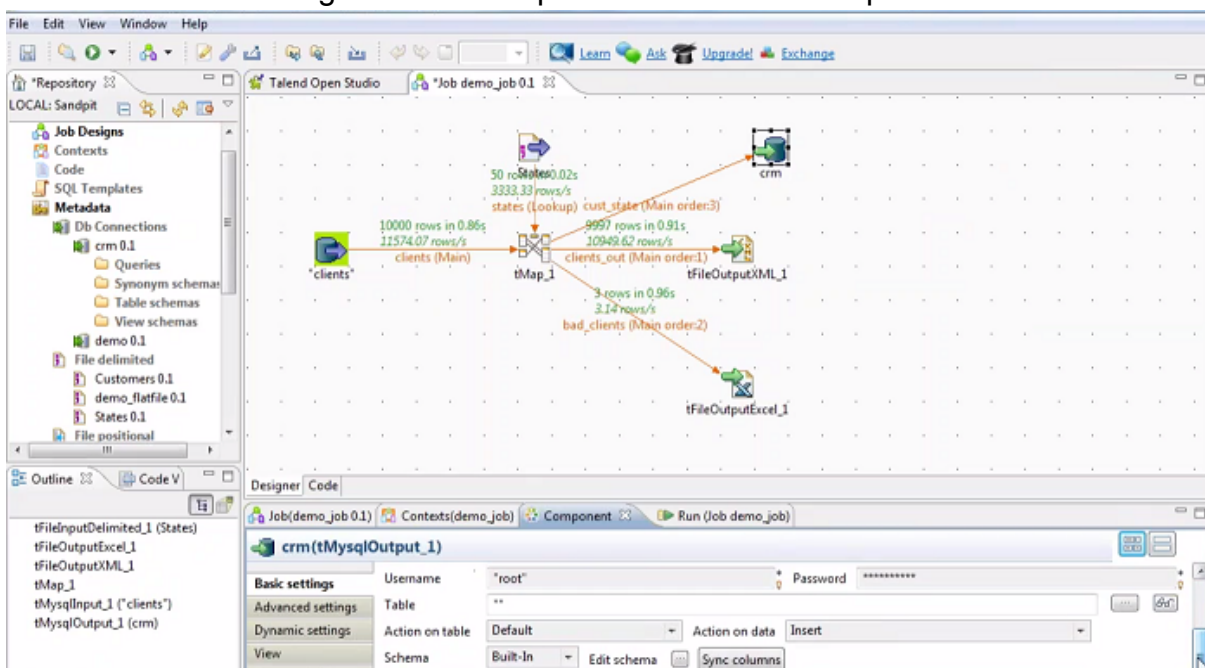
Já outro grupo de ferramentas de mapeamento de dados únicas ou que complementam um grupo de soluções com o objetivo de atender diretamente as conformidades impostas por normas e Leis, apresentam maior facilidade para atingir o objetivo final do trabalho. Muitas fornecedoras de soluções vendem suas ferramentas como um serviço, onde o cliente pode adquirir somente um módulo conforme sua necessidade, onde o mapeamento de dados é apenas um módulo que completa essa ferramenta (COMPUTERWORLD, 2020).

### 3.2 EXEMPLOS E COMPARAÇÃO DE FERRAMENTAS DE MAPEAMENTO DE DADOS

A demanda por ferramentas de mapeamento de dados vem crescendo, pois organizações de todos os tamanhos podem obter benefícios significativos. Nesta seção são citadas algumas ferramentas de mapeamento de dados disponíveis, tanto de forma gratuita e de código aberto como ferramentas pagas.

O Talend Open Studio é uma versão gratuita e de código aberto do conjunto comercial de ferramentas ETL da Talend. Ele fornece conectividade com fonte de dados como XML, Excel, JSON, CSV além de uma variedade de banco de dados SQL e NoSQL. Em vez de passar pelo longo processo de inserir manualmente os dados, o Talend Open Studio oferece uma ferramenta gráfica relativamente fácil de usar para mapear os dados (BOWEN, 2012). A Figura 12 demonstra um exemplo da tela de mapeamento de dados do Talend Open Studio.

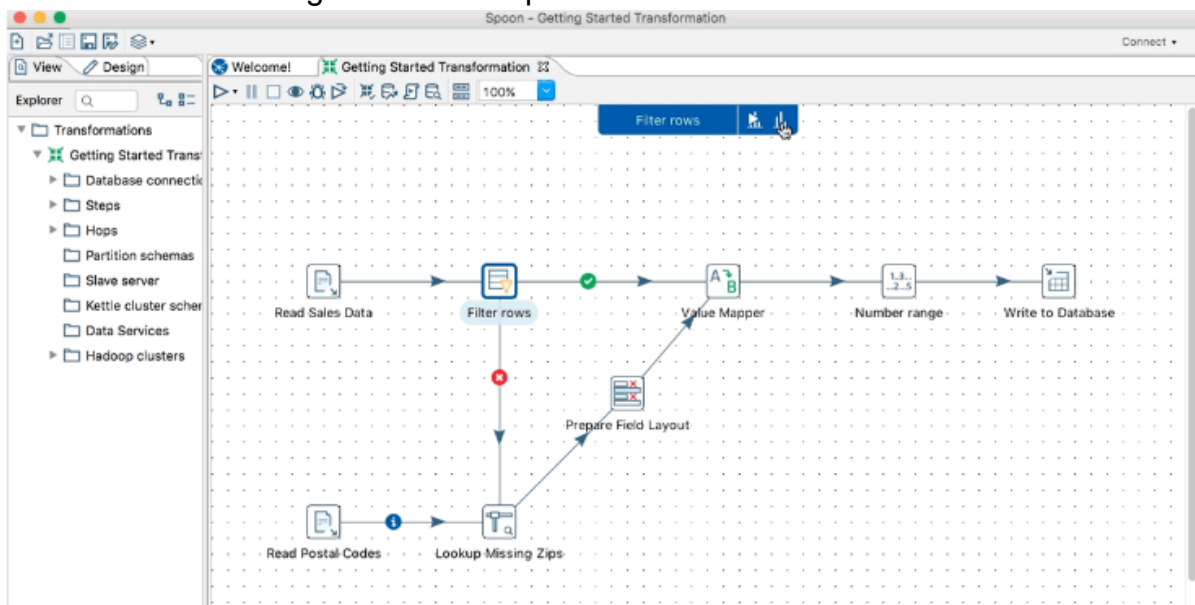
Figura 12 - Exemplo de tela do Talend Open Studio



Fonte: Bowen (2012)

Outra ferramenta é a Pentaho Kettle, que possui uma interface muito fácil de ser usada. Ela é uma ferramenta de código aberto desenvolvida pela Hitachi. Aceita diversas fontes de dados como XML, CSV, Excel, JSON e diversos banco de dados incluindo NoSQL. A ferramenta possui a facilidade de um sistema intuitivo onde permite conectar os modelos de dados para realização do mapeamento (CASTERS; BAUMAN; VAN DONGEN, 2010). A Figura 13 exemplifica a tela de mapeamento de dados do Pentaho Kettle.

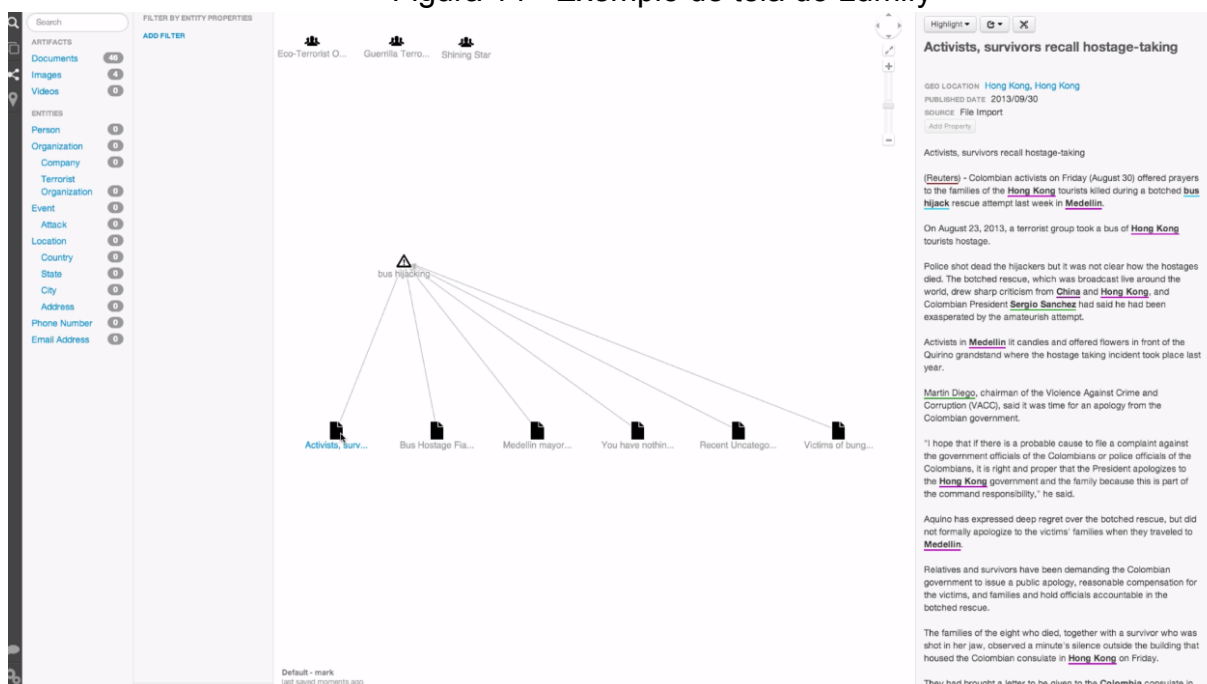
Figura 13 - Exemplo de tela do Pentaho Kettle



Fonte: Casters; Bauman; Van Dongen(2010)

A ferramenta Lumify também é desenvolvida em código aberto. Seu desenvolvimento é realizado pela Altamira Technologies. A ferramenta atende a diversas necessidades em um ambiente com grande volume de dados. O Lumify tem seu acesso baseado em navegador. A conexão com as fontes de dados acaba ficando limitada a somente a banco de dados SQL e NoSQL(ALTAMIRA TECHNOLOGIES CORPORATION, 2015). A Figura 14 demonstra um exemplo da tela de mapeamento de dados do Lumify.

Figura 14 - Exemplo de tela do Lumify



Fonte: Altamira Technologies Corporation (2015)

É possível visualizar os prós e os contras das ferramentas acima mencionados no Quadro 5, conforme os estudos de Iyer e Lakhtaria (2017) pode-se observar características parecidas entre as ferramentas, como todas serem de código livre.

Quadro 5 - Comparação entre ferramentas de mapeamento de dados

(continua)

Nome do Software	Prós	Contras
Talend Open Studio	Código livre Capaz de se conectar em diversas fontes de dados. Fácil utilização.	Personalização difícil.

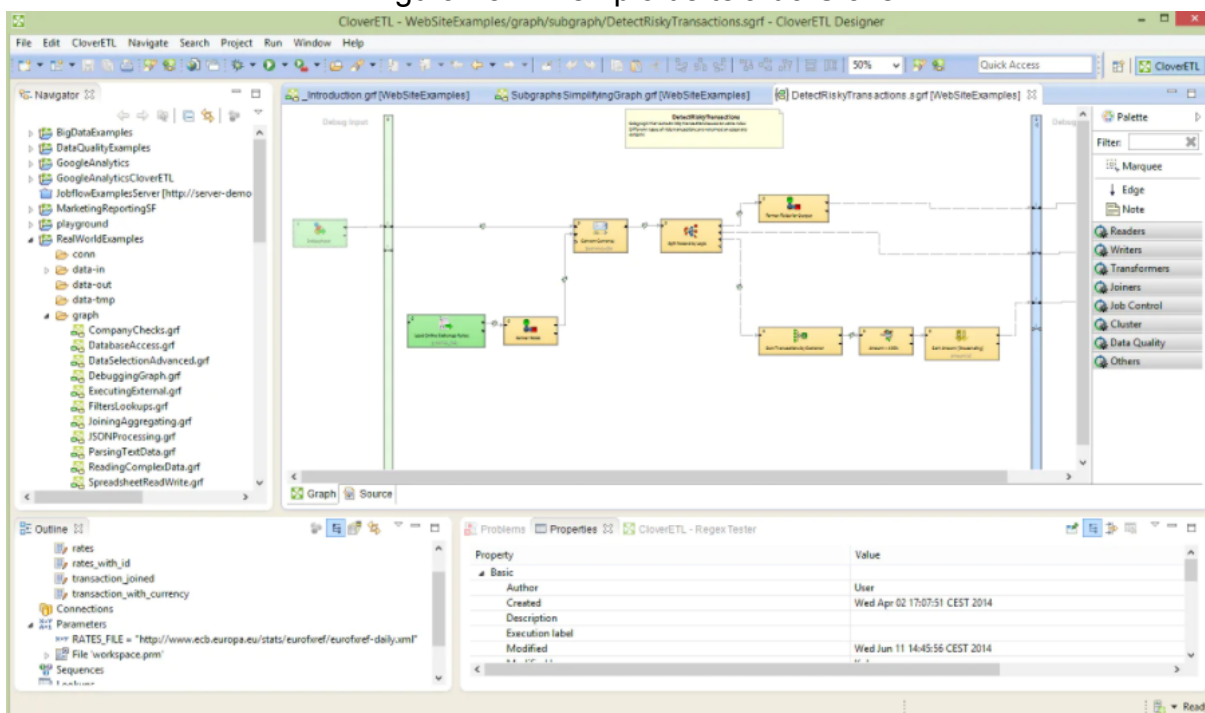
(conclusão)

Pentaho Kettle	Código livre Fácil utilização Capaz de se conectar em diversas fontes de dados.	Demora na análise dos dados.
Lumify	Código livre Utiliza ontologias para aprimorar o mapeamento.	Permite conexão com poucas fontes de dados.

Fonte: Iyer; Lakhtaria (2017)

Outra ferramenta é o CloverDX que é uma ferramenta paga desenvolvida pela Clover. O CloverDX é usado principalmente para a análise de dados para BI. O preço do licenciamento do CloverDX inicia em U\$5.000,00, em sua licença de forma perpétua, e pode chegar a valores maiores para obtenção de suporte. Não se tem uma versão gratuita do CloverDX, porém é oferecido um teste gratuito de 45 dias. Seu principal benefício é possuir um servidor de gerenciamento, onde permite que diversos usuários trabalhem em conjunto dentro de um ambiente corporativo. O CloverDX é integrado com diversos bancos de dados e arquivos, também permite a integração com API. A Figura 15 demonstra um exemplo da tela de mapeamento de dados do CloverDX.

Figura 15 - Exemplo de tela do CloverDX



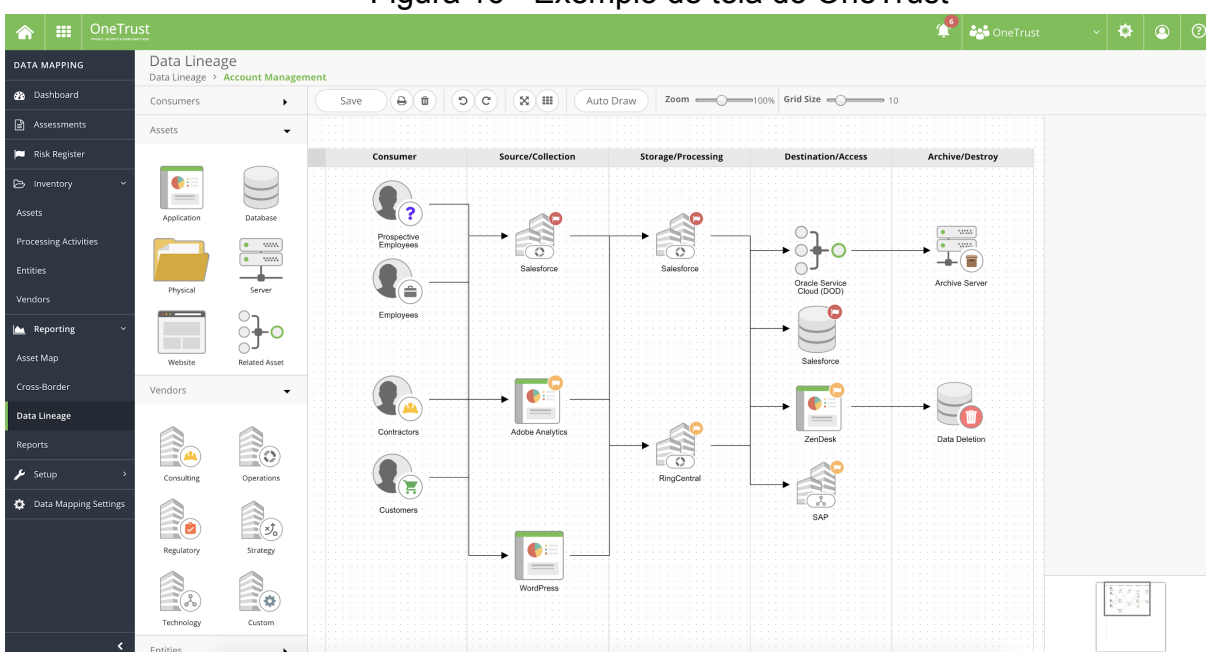
Fonte: CloverDX (2021)

As ferramentas Lumify, Pentaho, Talend e CloverDX são ferramentas de mapeamento de dados de caráter geral. Existem também ferramentas de mapeamento de dados que foram desenvolvidas para atender especificamente as conformidades da RGPD e LGPD. Essas ferramentas além de fazerem a parte de mapeamento de dados, fazem também outras tarefas de proteção à privacidade dos dados, como gerenciamento de políticas e termos consentimentos de privacidade, gerenciamento de risco internos e com fornecedores e a automatização de requisições e controles de direito dos dados dos titulares.

Um exemplo deste tipo de ferramenta é a OneTrust, que é uma ferramenta paga e que pode ser utilizada tanto localmente como em diretamente na nuvem. O software oferece desde rastreamento de geolocalização, controle de inventário, mapeamento de dados, gerenciamento de registros, auditoria e relatórios. Apresenta conexão com diversas fontes de dados, desde banco de dados SQL e NoSQL como

conexões diretamente à Microsoft Azure ou a Amazon AWS. A ferramenta já é compatível com as principais legislações como a *Health Insurance Portability and Accountability Act*(HIPAA), *Singapore's Personal Data Protection Act*(PDPA), LGPD, *California Consumer Privacy Act*(CCPA) e a RGPD. Além disso, atende as normas ABNT NBR ISO/IEC 27001 e 27701(ONE TRUST, 2020). A Figura 16 exibe a tela de mapeamento de dados do OneTrust.

Figura 16 - Exemplo de tela do OneTrust

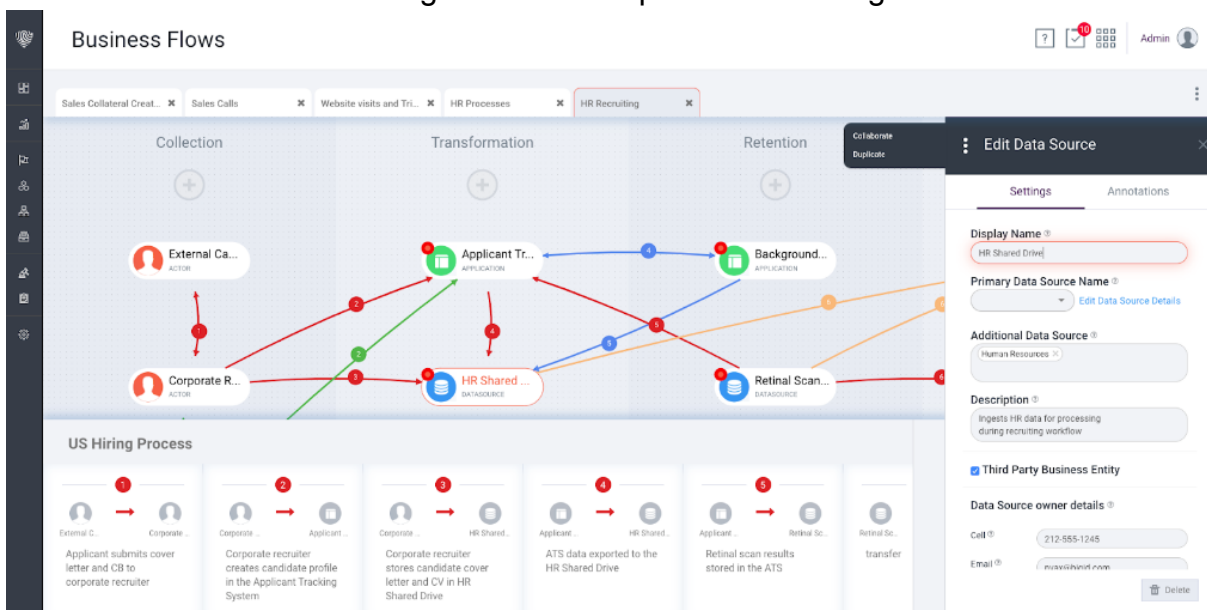


Fonte: Casters; Bauman; Van Dongen(2010)

Outra ferramenta que apresenta a característica para a conformidade é a BigID. A forma de licenciamento é paga e sua utilização é baseada somente em ambiente em nuvem. O software oferece desde recursos como mapeamento de dados, controle de acesso, gestão de consentimento, gestão de incidentes e identificação de dados sensíveis automaticamente. A BigID surgiu em 2016 com o objetivo de fazer as empresas entenderem os seus ativos de dados e tirar vantagem competitiva com isso (AMY-VOGT, 2021). Na Figura 17 pode-se observar a tela de mapeamento de dados do BigID em funcionamento na sua versão demo.



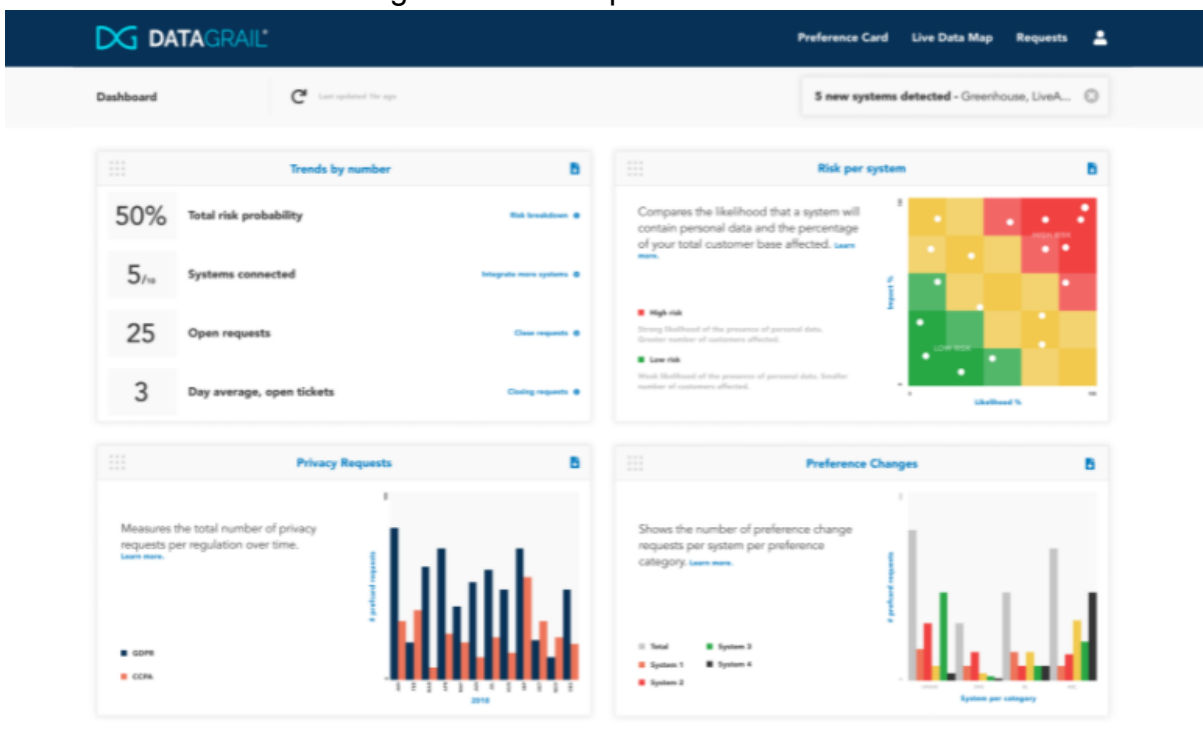
Figura 17 - Exemplo de tela do BigID



Fonte: Hospelhorn(2020)

Outra ferramenta disponível no mercado é a DataGrail, que é desenvolvida especificamente para as equipes de segurança gerenciarem os dados pessoais para as regulamentações de privacidade. A ferramenta suporta diversas conexões com variadas fontes de dados, e seu módulo principal é o de mapeamento de dados. A ferramenta promete classificar os dados para facilitar a localização dos dados de titulares quando são solicitados pelo mesmo. A Figura 18 mostra um *dashboard* centralizado com os mapas e *logs* de dados em tempo real.

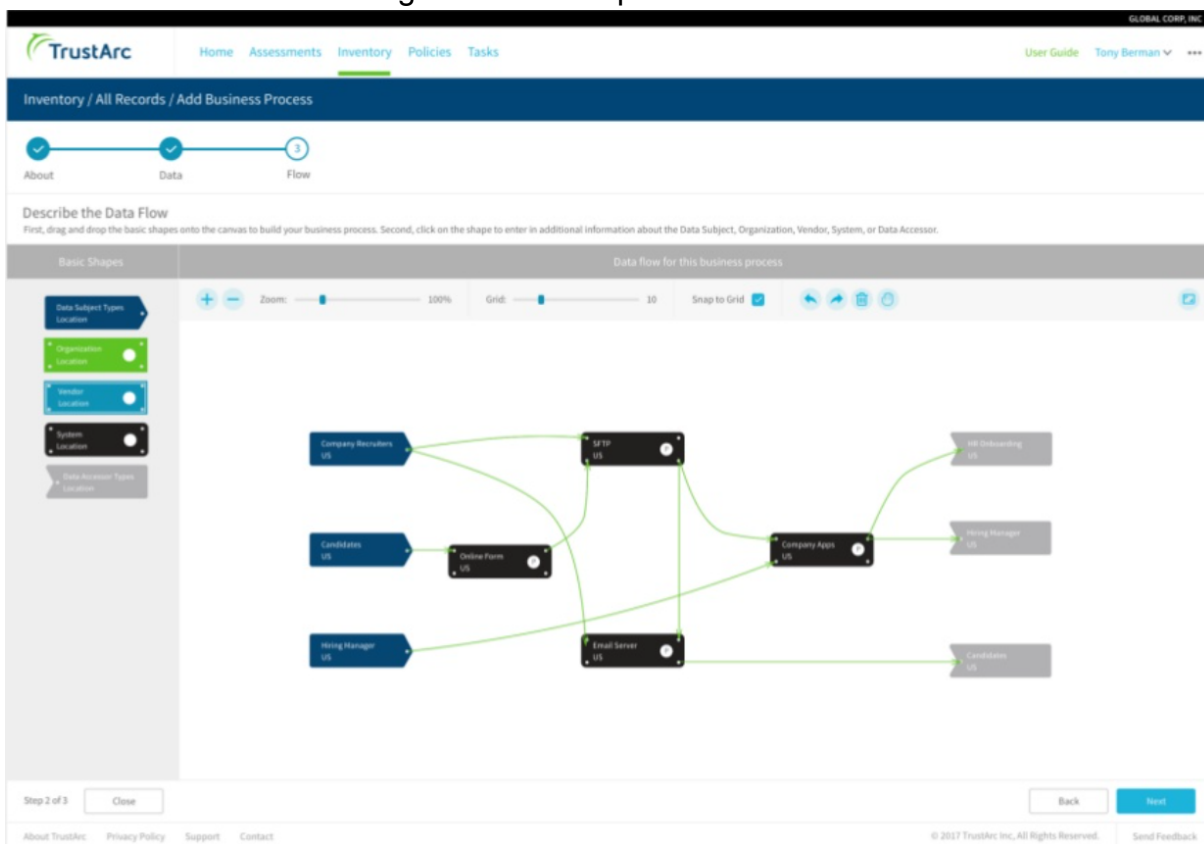
Figura 18 - Exemplo de tela do DataGrail



Fonte: DataGrail(2021)

A TrustArc é mais uma ferramenta de mapeamento de dados disponível, é comercializada como um serviço e executada em um ambiente web em nuvem. Tem o objetivo de gerenciar, proteger e obter informações sobre os dados em todos os repositórios de dados da organização. A TrustArc utiliza inteligência artificial para auxiliar na descoberta e classificação dos dados, além de fornecer padrões de políticas para atender diretamente as Leis, como a LGPD. A Figura 19 mostra a tela de configuração do processo de mapeamento de dados do TrustArc.

Figura 19 - Exemplo de tela do TrustArc



Fonte: DataGrail(2021)

As ferramentas citadas foram selecionadas após consultas em revistas digitais como SecurityMagazine (2021) e Cio (2018). O site de comparação de ferramentas G2<sup>1</sup> e Capterra<sup>2</sup> foram utilizados para comparar e verificar as características das ferramentas citadas.

Observando a documentação das ferramentas nos sites de seus fabricantes, foi desenvolvido o Quadro 6 e o Quadro 7 que comparam de forma breve as ferramentas apresentadas anteriormente. Esta comparação entre as ferramentas será essencial para a definição das ferramentas avaliadas no decorrer do trabalho.

O Quadro 7 reflete as principais funcionalidades referentes a LGPD:

- Anonimização dos dados: Os dados após anonimizados, sem possibilidade de reversão, não são mais tratados como dados pessoais. Assim podendo ser utilizados para fins diferentes ao consentimento do titular.

<sup>1</sup> G2. Disponível em: [www.g2.com](http://www.g2.com) Acesso em: 21 abril 2021.

<sup>2</sup> CAPTERRA. Disponível em: [www.capterra.com.br](http://www.capterra.com.br) Acesso em: 21 abril 2021.

- Avaliação de riscos: Importante para garantir que os dados pessoais estão em um ambiente seguro, e que garanta que os requisitos de segurança estejam adequados.
- Identificação de dados sensíveis: Dados sensíveis devem possuir camadas de segurança adicionais, identificar os dados é essencial para a organização.
- Gerenciamento de consentimento: Conforme o Artigo 18 da LGPD(BRASIL,2018) prevê que o titular dos dados pessoais a qualquer momento pode solicitar ao controlador dos dados, a confirmação da existência do seu consentimento.
- *Dashboard*: Funcionalidade que facilita o operador da ferramenta visualizar e gerenciar os dados.
- Classificação dos dados: Funcionalidade indispensável para a organização entender seus dados, e identificar o ciclo de vida destes dados.

Quadro 6 - Comparação entre ferramentas de mapeamento de dados

Ferramentas	Categoria de atuação	Licenciamento	Plataforma
Talend Open Studio	Integração de dados	Gratuita	Local
Pentaho Kettle	Integração de dados	Gratuita	Local
Lumify	Integração de dados	Gratuito	Local
CloverDX	Integração de dados	Pago	Nuvem e Local
BigId	Conformidade	Pago	Nuvem
Onetrust	Conformidade	Pago	Nuvem e Local
DataGrail	Conformidade	Pago	Nuvem
TrustArc	Conformidade	Pago	Nuvem

Fonte: Próprio autor.

Quadro 7 - Comparação dos serviços entre ferramentas de mapeamento de dados

	Talend Open Studio	Pentaho Kettle	Lumify	Clover DX	BigId	Onetrust	DataGrail	TrustArc
Anonimização dos dados	Sim	Não	Não	Não	Sim	Sim	Não	Não
Avaliação de riscos	Não	Não	Não	Não	Sim	Sim	Não	Sim
Identificação de dados sensíveis	Não	Não	Não	Não	Sim	Sim	Sim	Sim
Gerenciamento de consentimento	Não	Não	Não	Não	Sim	Sim	Não	Sim
Dashboard	Sim	Sim	Não	Sim	Não	Sim	Não	Sim
Classificação de dados	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Fonte: Próprio autor.

Pelas características apresentadas no Quadro 7, pode-se verificar uma diferença significativa nas funcionalidades presentes nas ferramentas desenvolvidas com foco na conformidade. Nessas ferramentas, nota-se a presença de funcionalidades específicas para o atendimento da RGPD ou LGPD, como gerenciamento de consentimento e avaliações de riscos. De forma semelhante, as ferramentas que têm sua forma de licenciamento paga, tiveram um número maior de funcionalidades disponibilizadas que as gratuitas.

De forma geral, a ferramenta Onetrust, ferramenta paga e com foco na conformidade, é a única que possui todas as funcionalidades selecionadas. Entre as ferramentas gratuitas, as que apresentaram maiores funcionalidades foram a Talend e o Pentaho Kettle.

### 3.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

O capítulo apresenta as ferramentas de mapeamento de dados, suas características e técnicas de funcionamento que são essenciais para o entendimento da proposta e avaliação.

Conforme pesquisado em diversos artigos e *sites* de revista, ainda existe uma baixa quantidade de ferramentas de mapeamento de dados que são gratuitas disponíveis no mercado. Ferramentas de mapeamento de dados que fazem parte de ferramentas de conformidade são todas de licenciamento pago. Além disso, as ferramentas com foco na conformidade das Leis ainda são recentes no mercado e a grande maioria é desenvolvido para atender empresas de grande porte(ESPINOZA, 2020). Essas afirmações puderam ser comprovadas no estudo de algumas ferramentas e apresentadas nos Quadros 6 e 7, onde a ferramenta OneTrust teve destaque.

## 4 PROPOSTA DE SOLUÇÃO

A proposta de solução deste trabalho baseia-se em analisar ferramentas de mapeamento de dados, avaliá-los quanto à conformidade com a LGPD e a capacidade para auxiliar as organizações a adequarem-se à Lei. Esta análise permitirá analisar quais ferramentas realizam o mapeamento de dados de acordo com os critérios definidos neste capítulo.

A seção 4.1 define as ferramentas de mapeamento de dados a serem avaliadas. A seção 4.2 estabelece os critérios para a realização de avaliação das ferramentas conforme as normas ABNT/NBR 25020 e 25030 e a LGPD. A seção 4.3 define as métricas utilizando o padrão de avaliação de software estudado neste trabalho.

Os casos de testes são apresentados na seção 4.4 para definir como base para a padronização dos testes e avaliações das ferramentas. A seção 4.5 descreve como os resultados serão validados. A seção 4.6 descreve a qualidade dos dados utilizados para a realização dos dados.

### 4.1 DEFINIÇÃO DAS FERRAMENTAS DE MAPEAMENTO DE DADOS

Observando os prós e contras definidos na seção 3.2, foram definidas três ferramentas de mapeamento de dados para serem avaliadas com base nos critérios e métricas definidos neste capítulo. As ferramentas BigID e OneTrust que possuem grande número de funcionalidades (Quadro 7) não apresentam versão gratuita de teste ou alguma demonstração para que se pudesse ser avaliado e testado. A decisão foi selecionar as ferramentas Talend Open Studio, Pentaho Kettle por serem ferramentas gratuitas que apresentam um número de funcionalidades maior que as demais ferramentas gratuitas comparadas. Também foi selecionada a ferramenta CloverDX que é uma ferramenta paga mas que disponibilizou uma versão de testes por 45 dias. A seleção dessas ferramentas possibilita a comparação de ferramentas pagas, com ferramentas gratuitas. Também pode ser comparado como as

ferramentas desenvolvidas para integração de dados podem ser utilizadas para auxiliar no cumprimento da LGPD.

#### 4.2 CRITÉRIOS

Os critérios para avaliação das ferramentas foram definidos baseados no entendimento do modelo de qualidade do produto de software apresentado na seção 2.5. O Quadro 8 detalha esses critérios.

Quadro 8 - Critérios para Avaliação

(continua)

Área	Critérios	Considerado	Descartado
Características de Qualidade Interna e Externa			
<b>Funcionalidade</b>	Adequação	X	
	Acurácia	X	
	Interoperabilidade	X	
	Segurança de Acesso		X
	Conformidade	X	
<b>Confiabilidade</b>	Maturidade		X
	Tolerância a Falhas		X
	Recuperabilidade	X	
	Conformidade		X
<b>Usabilidade</b>	Inteligibilidade		X
	Apreensibilidade		X
	Operacionalidade	X	
	Atratividade		X
	Conformidade		X



(conclusão)

<b>Eficiência</b>	Comportamento em Relação ao Tempo		X
	Utilização dos Recursos	X	
	Conformidade		X
<b>Manutenibilidade</b>			X
<b>Portabilidade</b>	Adaptabilidade		X
	Coexistência	X	
	Capacidade de Substituir		X
	Conformidade		X
<b>Características de Qualidade em Uso</b>			
<b>Efetividade</b>		X	
<b>Produtividade</b>		X	
<b>Satisfação</b>			X
<b>Segurança</b>			X

Fonte: ABNT ISO 25010 (2011)

A característica funcionalidade foi selecionada pois é importante que o sistema cumpra aquilo que ele propunha com êxito, por isso serão analisadas as subcaracterísticas de funcionalidade como adequação, acurácia, interoperabilidade, e conformidade.

A subcaracterística de conformidade é de extrema importância para alcançar o objetivo proposto, possui a capacidade de mensurar se o software avaliado atingiu a conformidade da LGPD e da norma ABNT NBR ISO/IEC 27701.

A confiabilidade é essencial para determinar se o software avaliado possui capacidade de se manter em um bom nível de desempenho, mesmo se utilizado em condições específicas. A subcaracterísticas recuperabilidade foi escolhida.

A usabilidade foi escolhida pois é necessário saber se o sistema é de fácil utilização perante o usuário, para tal é aplicado a subcaracterística operacionalidade.

A característica eficiência foi selecionada pois é importante entender se o software tem capacidade de manter o desempenho adequado em condições explícitas, deste modo é utilizado as subcaracterísticas como a utilização dos recursos.

O atributo portabilidade foi escolhido para entender se o software é hábil para ser utilizado em ambientes com outros softwares em execução, para isso serão utilizados a subcaracterística coexistência.

A característica manutenibilidade não é utilizada, devido ao objetivo do trabalho não ser alterar ou incluir funcionalidades novas ao software e sim testar as funcionalidade já impostas.

Das características de qualidade de uso, serão utilizados os critérios de efetividade e produtividade. Elas são as medições de um softwares considerando a qualidade do software em suas tarefas e cenários do dia a dia dos usuários.

A efetividade foi escolhida pois é capaz de mensurar se o software permite que o usuário consiga atingir os seus objetivos de forma correta e completa.

A produtividade foi selecionada pois ela é importante para mensurar se o software permite que o usuário consiga utilizar a quantidade de recursos necessários com eficácia para atingir um determinado objetivo.

Algumas características do modelo de qualidade em uso não serão consideradas. A característica de segurança não é utilizada por se tratar de testes em um ambiente interno, dessa forma não corre riscos de ataques externos, e assim não é possível mensurar testes dessa característica. Outra característica que não será utilizada é a satisfação, já que a forma de avaliação da característica seria através de questionários a grupos de usuários que utilizaram o software.

### 4.3 MÉTRICAS

Após a definição dos critérios a serem utilizados na avaliação dos softwares, é necessário definir as métricas para mensurar estes critérios.

As métricas internas são medições que consideram as características próprias internas, isto é, sem a execução do software avaliado. Como o objetivo do trabalho é analisar as funcionalidades do software, as métricas internas não são adequadas para serem utilizadas nesse trabalho.

As métricas que serão utilizadas no trabalho serão as métricas externas e de qualidade de uso. As métricas externas são as medições do software considerando a execução do software e de seus resultados no ambiente. As métricas de qualidade de uso são medições do software observando o comportamento e qualidade do seu sistema em tarefas e cenários do dia a dia do usuário. As métricas externas e de qualidade de uso a serem utilizadas estão definidas nos Quadros 9 e 10. A seleção das métricas considerou a relevância que elas possuem quanto às exigências pela norma ABNT NBR ISO/IEC 27701 e pela LGPD.

Quadro 9 - Métricas para avaliação de Qualidade Externa

(continua)

Característica	Subcaracterística	Métrica	Propósito da métrica	Método de aplicação	Medida e Fórmula	Interpretação	Tipo de medida
Funcionalidade	Adequação	Adequação Funcional	Quão adequadas são as funções avaliadas?	Número de funções que são adequadas para executar as tarefas especificadas em comparação com o número de funções avaliadas.	$X = 1 - A / B$ A = Número de funções nas quais os problemas são detectados na avaliação B = Número de funções avaliadas	$0 \leq X \leq 1$ Quanto mais perto de 1, mais adequado.	Quantitativa
Funcionalidade	Acurácia	Precisão (Resultado corretos)	Com que frequência os usuários finais encontram resultados com precisão inadequada?	Registre o número de resultados com precisão inadequada. Através dos relatórios emitidos pelas ferramentas e pelas consultas realizadas.	$X = A / T$ A = Número de resultados encontrados pelos usuários com um nível de precisão diferente do requerido T = tempo de operação	$0 \leq X$ O mais próximo de 0 é o melhor.	Quantitativa Tempo

(conclusão)

Funcionalidade	Interoperabilidade	Interoperabilidade disponível	Com que frequência o usuário encontra restrições ou falhas inesperadas ao trocar dados entre o software avaliado e demais softwares?	Use o software avaliado em conjunto com outro software que permita a troca de dados entre eles.	$X = 1 - A / B$ A = Número de softwares que apresentaram falhas ao trocar dados com o software avaliado. B = Número de softwares disponíveis.	$0 \leq X \leq 1$ Quanto mais perto de 1, melhor.	Quantitativa
Funcionalidade	Conformidade	Conformidade e de funcionalidade	Quão compatível é a funcionalidade do produto com os regulamentos, padrões e convenções aplicáveis?	Contar o número de itens que exigem conformidade que foram atendidos e comparar com o número de itens que exigem conformidade na especificação. Teste os casos de teste de acordo com os itens de conformidade. Realize testes funcionais. Contar o número de itens de conformidade que foram satisfeitos.	$X = 1 - A / B$ A = Número de itens de conformidade de funcionalidade especificados que não foram implementados durante o teste B = Número total de itens de conformidade de funcionalidade especificados	$0 \leq X \leq 1$ O mais próximo de 1,0 é o melhor.	Quantitativa
Usabilidade	Operacionalidade	Consistência operacional em uso	Quão consistente é o componente da interface do usuário?	Observe o comportamento do usuário e peça a opinião.	$Y = N / UOT$ N = Número de operações que o usuário encontrou inaceitavelmente inconsistente com a expectativa do usuário UOT = tempo de operação do usuário	$0 \leq Y$ O menor e mais próximo de 0,0 é o melhor.	Tempo
Confiabilidade	Recuperabilidade	Flexibilidade de recuperação	Em caso de falha o software é capaz de se restabelecer?	Observe o comportamento do software ao apresentar falhas e erros.	$X = A / N$ A = número de erros recuperados N = Número de erros encontrados	$X \geq$ Quanto menor, melhor.	Quantitativo
Portabilidade	Coexistência	Coexistência disponível	Com que frequência o usuário encontra restrições ou falhas inesperadas ao operar em simultâneo com outro software?	Use o software avaliado em simultâneo com outros softwares que o usuário geralmente usa.	$X = A / T$ A = Número de restrições ou falhas inesperadas que o usuário enfrenta ao operar em simultâneo com outro software T = tempo de operação simultânea de outros softwares	$0 \leq X$ O mais próximo de 0 é o melhor.	Quantitativo Tempo

Fonte: ABNT ISO 25010(2011)

Quadro 10: Métricas para avaliação de Qualidade em Uso

Característica	Métrica	Propósito da métrica	Método de aplicação	Medida e Fórmula	Interpretação	Tipo de medida
Efetividade	Frequência de Erro	Qual é a frequência de erros?	Casos de teste	$X = A / T$ A = número de erros tomados pelo usuário T = Tempo ou número de tarefas	$0 \leq X$ Quanto mais próximo de 0, melhor.	Quantitativo
Produtividade	Tempo da Tarefa	Quanto tempo demora para completar uma tarefa?	Casos de teste	$X = Ta / Tb$ Ta = Tempo ocioso do usuário Tb = tempo da tarefa	$X \geq$ Quanto menor, melhor.	Tempo

Fonte: ABNT ISO 25010 (2011)

Para fazer a avaliação das ferramentas Talend Open Studio, Pentaho Kettle e CloverDX serão utilizadas as medidas e as fórmulas definidas na ABNT ISO 25010 (2011) e mostradas nos Quadros 9 e 10. É importante ressaltar que a subcaracterística Conformidade é muito importante nesse trabalho, pois é ela que permite que seja verificado se as ferramentas atendem ou não a LGPD. Dessa forma, foram definidas as seguintes funcionalidades para serem testadas:

- a) Avaliação de riscos dos dados: verificar se a ferramenta permite estabelecer os critérios de aceitação de riscos, e como estes riscos são identificados. Também permite identificar as possíveis causas e consequências que ocorram esses riscos. Essa conformidade é exigida pela ABNT NBR ISO/IEC 27701 em sua seção de planejamento como mostra o Quadro 3. Ela atende o requisito da LGPD em seu Artigo 38 e descreve que a agência nacional poderá solicitar ao controlador dos dados um relatório de impacto à proteção dos dados pessoais.
- b) Segurança de acesso: averiguar se a ferramenta dispõe de controles de acessos de usuário, e, dessa forma, a ferramenta garante que o usuário só tenha acesso ao que é necessário para o usuário realizar suas funções. A conformidade é determinada pela ABNT NBR ISO/IEC 27701 em sua seção controle de acesso como mostra o Quadro 4. Essa funcionalidade atende o

requisito da LGPD em seu Artigo 46 e 49 e define o que um sistema seguro deve possuir e medidas de segurança para proteção dos dados pessoais.

- c) Classificação dos dados: testar se a ferramenta realiza a classificação dos dados pessoais conforme esses dados são coletados (ABNT NBR ISO/IEC 27701, seção gestão de ativos, apresentada no Quadro 4 deste trabalho). A finalidade dessa funcionalidade é atender os Artigos 5X , Artigo 6o, Artigo 46, Artigo 47 e Artigo 49 que apresentam regras para o tratamento de dados.
- d) Rotulação dos dados: apurar se a ferramenta realiza a rotulação dos dados pessoais, tornando de fácil visualização para o controlador, o que são dados pessoais, comuns e sensíveis. A norma ABNT NBR ISO/IEC 27701, na seção gestão de ativos apresentada no Quadro 4, estabelece essa necessidade. Essa funcionalidade visa atender os Artigos 5X , Artigo 6o, Artigo 46, Artigo 47 e Artigo 49 que apresentam regras para o tratamento de dados.
- e) Criptografia dos dados: aferir se a função de criptografia e descriptografia é aplicada aos dados pessoais armazenados, de acordo com o estabelecido na norma ABNT NBR ISO/IEC 27701 e apresentada no Quadro 4. Essa funcionalidade visa atender os Artigos 46 que exigem medidas de segurança para proteção dos dados pessoais.
- f) Armazenamento de eventos(*logs*): confirmar se a ferramenta armazena os eventos(*logs*) e permitindo o acesso ao histórico armazenado (norma ABNT NBR ISO/IEC 2770, seção segurança da operação). Essa funcionalidade visa atender os Artigos 46 que exigem medidas de segurança para proteção dos dados pessoais.
- g) Resposta a incidentes de segurança: certificar se a ferramenta auxilia na identificação dos incidentes de segurança, na identificação das

consequências e do período de tempo que ocorreu o incidente (ABNT NBR ISO/IEC 27701, seção gestão de incidentes de segurança da informação). Essa funcionalidade atende aos Artigos 48 e 50 que demandam a comunicação de incidentes de segurança.

- h) Análise crítica técnica do *compliance*: atestar se a ferramenta é capaz de monitorar e verificar se está sendo executado o tratamento permitido aos dados pessoais (norma ABNT NBR ISO/IEC 27701, seção *compliance*). Essa funcionalidade atende ao Artigo 50 que solicita boas práticas e transparência com o tratamento de dados.
- i) Entrega de dados aos titulares: Quando solicitado os dados pessoais de determinado titular, a ferramenta consegue localizar com precisão onde estão localizados estes dados. Essa funcionalidade atende ao Artigo 9 e Artigo 18 onde o titular tem o direito de acesso facilitado a suas informações sobre o tratamento dos seus dados pessoais.
- j) Anonimização: A anonimização dos dados pessoais é realizada pela ferramenta quando solicitado pelo usuário. Essa funcionalidade atende ao Artigo 16 onde o controlador dos dados deve anonimizar os dados pessoais caso deseje manter armazenado após o tratamento .

A fórmula definida na ABNT ISO 25010 (2011) para cálculo da conformidade ( $X = 1 - A / B$ ) considera o número de itens testados, onde A representa o número de itens não conformes e B representa o número de itens totais testados. O trabalho realiza a avaliação total de 10 itens de conformidade. Cada item testado será avaliado com o valor 0 se a ferramenta não implementar a funcionalidade, com o valor 0.5 se a ferramenta realiza parcialmente a funcionalidade ou com o valor 1 se ele implementa completamente a funcionalidade.

A Conformidade terá um peso maior do que as demais características avaliadas, pois esta subcaracterística mensuram as funcionalidades da ferramenta e

como elas se adequam às necessidades propostas pela Lei. Portanto, a divisão dos pesos de medidas é de 70% para a subcaracterística de Conformidade e os outros 30% são divididos igualmente pelas demais subcaracterísticas.

#### 4.4 CASOS DE TESTES

Escolhidos as ferramentas a serem avaliadas, definidos os critérios e as métricas para a avaliação destas ferramentas, é fundamental definir os casos de testes que servirão de base para padronização dos testes e avaliação das ferramentas.

Os casos de teste terão como base as métricas e funcionalidades definidas no capítulo 4, que foram definidos a partir dos modelos de qualidades, norma ABNT NBR ISO/IEC 27701 e da LGPD.

A relação de cada caso de teste com os critérios e métricas, primeiramente selecionados, se encontra no corpo de cada caso de teste, na linha Resumo das tabelas, onde é representado quais métricas atendem cada caso de teste.

Quadro 11 - Caso de Teste 1

<b>Caso de Teste</b>	Parametrização da ferramenta
<b>Resumo</b>	<p>Testar se a ferramenta permite parametrizar de acordo com fontes de dados e parâmetros necessários.</p> <p>Atende as métricas de: Adequação, Interoperabilidade, Coexistência, Tempo da Tarefa e Frequência de Erro.</p>
<b>Pré-condições</b>	Software parametrizado.
<b>Ação</b>	<p>Configurar a ferramenta.</p> <p>Configurar conexões com as fontes de dados.</p>
<b>Resultados Esperados</b>	Sistema parametrizado.

Fonte: Próprio autor.



Quadro 12 - Caso de Teste 2

<b>Caso de Teste</b>	Avaliação de riscos dos dados
<b>Resumo</b>	<p>Testar se a ferramenta consegue detectar os riscos existentes e medir a probabilidade dos riscos de acontecer e gerar relatório de avaliação dos riscos.</p> <p>Atende as métricas de: Adequação, Conformidade , Frequência de Erro e Tempo da Tarefa.</p>
<b>Pré-condições</b>	Detecção de risco.
<b>Ação</b>	<p>Avaliar a probabilidade dos riscos.</p> <p>Gerar relatório de avaliação de risco.</p>
<b>Resultados Esperados</b>	Relatório que apresente a probabilidade dos riscos.

Fonte: Próprio autor.

Quadro 13 - Caso de Teste 3

<b>Caso de Teste</b>	Segurança de Acesso
<b>Resumo</b>	<p>Testar se a ferramenta possui mecanismos que controlem o acesso de usuários.</p> <p>Atende as métricas de: Adequação, Conformidade, Tempo da Tarefa.</p>
<b>Pré-condições</b>	Sistema parametrizado.
<b>Ação</b>	<p>Cadastrar usuários.</p> <p>Cadastrar permissões de usuários.</p>
<b>Resultados Esperados</b>	Ferramenta negando acesso de usuários sem permissão.

Fonte: Próprio autor.

Quadro 14 - Caso de Teste 4

<b>Caso de Teste</b>	Identificação de dados sensíveis
<b>Resumo</b>	<p>Testar se a ferramenta realiza a identificação dos dados sensíveis(dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural).</p> <p>Atende as métricas de: Adequação, Acurácia, Coexistência, Conformidade, Tempo da Tarefa, Frequência de Erro.</p>
<b>Pré-condições</b>	Caso de testes 1.
<b>Ação</b>	<p>Cadastrar modelos de dados para dados pessoais e sensíveis.</p> <p>Gerar visualização dos dados classificados.</p>
<b>Resultados Esperados</b>	Dados classificados.

Fonte: Próprio autor.

Quadro 15 - Caso de Teste 5

<b>Caso de Teste</b>	Rotulação dos dados
<b>Resumo</b>	<p>Testar se a ferramenta possibilita adicionar rótulos aos seus dados pessoais existentes.</p> <p>Atende as métricas de: Adequação, Acurácia, Coexistência, Conformidade, Tempo da Tarefa e Frequência de Erro.</p>
<b>Pré-condições</b>	Caso de Teste 1.
<b>Ação</b>	<p>Criar rótulos.</p> <p>Correlacionar rótulos aos modelos de dados.</p> <p>Gerar visualização dos dados.</p>
<b>Resultados Esperados</b>	Dados rotulados podendo o usuário identificar um dados sensível apenas pela sua visualização.

Fonte: Próprio autor.

Quadro 16 - Caso de Teste 6

<b>Caso de Teste</b>	Criptografia dos dados
<b>Resumo</b>	<p>Testar se a ferramenta possibilita alguma forma de criptografia de dados pessoais.</p> <p>Atende as métricas de: Adequação, Conformidade, Tempo da Tarefa e Frequência de Erro.</p>
<b>Pré-condições</b>	<p>Caso de Teste 1.</p> <p>Caso de Teste 4.</p>
<b>Ação</b>	Executar criptografia em amostra de dados pessoais.
<b>Resultados Esperados</b>	Dados criptografados.

Fonte: Próprio autor.

Quadro 17 - Caso de Teste 7

<b>Caso de Teste</b>	Armazenamento de eventos
<b>Resumo</b>	<p>Testar se a ferramenta permite armazenar os eventos(<i>logs</i>) ocorridos.</p> <p>Atende as métricas de: Adequação, Operacionalidade , Conformidade, Tempo da Tarefa e Frequência de Erro.</p>
<b>Pré-condições</b>	Realização de eventos.
<b>Ação</b>	Armazenar os eventos( <i>logs</i> ).
<b>Resultados Esperados</b>	Visualização dos eventos armazenados.

Fonte: Próprio autor.

Quadro 18 - Caso de Teste 8

<b>Caso de Teste</b>	Resposta a incidentes de segurança
<b>Resumo</b>	<p>Testar se a ferramenta identifica as consequências e o período de tempo de incidentes.</p> <p>Atende as métricas de: Adequação, Operacionalidade , Coexistência, Conformidade, Tempo da Tarefa e Frequência de Erro.</p>
<b>Pré-condições</b>	Caso de Teste 1.
<b>Ação</b>	Gerar evento de segurança que cause danos aos dados.
<b>Resultados Esperados</b>	Visualização das consequências e do período de tempo que um evento de segurança ocorreu.

Fonte: Próprio autor.

Quadro 19 - Caso de Teste 9

<b>Caso de Teste</b>	Análise crítica técnica do <i>compliance</i>
<b>Resumo</b>	<p>Testar se a ferramenta identifica em que parte do ciclo de tratamento de dados está ocorrendo em determinado dado.</p> <p>Atende as métricas de: Adequação, Operacionalidade , Conformidade, Tempo da Tarefa e Frequência de Erro.</p>
<b>Pré-condições</b>	Seleção de duas amostras de dados pessoais.
<b>Ação</b>	<p>Ferramenta detecta em qual fase do ciclo de dados estão os dados pessoais.</p> <p>Determinar se a finalidade do tratamento dos dados pessoais já foi finalizada.</p>
<b>Resultados Esperados</b>	Identificar em que fase do ciclo de tratamento de dados estão os dados pessoais.

Fonte: Próprio autor.

Quadro 20 - Caso de Teste 10

<b>Caso de Teste</b>	Entrega de dados aos titulares
<b>Resumo</b>	<p>Testar se a ferramenta possibilita encontrar dados de determinado dado pessoal quando solicitado.</p> <p>Atende as métricas de: Adequação, Operacionalidade, Conformidade, Tempo da Tarefa e Frequência de Erro.</p>
<b>Pré-condições</b>	Caso de Teste 1.
<b>Ação</b>	Solicitar a ferramenta para visualizar dados pessoais de determinado titular.
<b>Resultados Esperados</b>	Visualização de todos os dados pessoais de determinado titular.

Fonte: Próprio autor.

Quadro 21 - Caso de Teste 11

<b>Caso de Teste</b>	Anonimização
<b>Resumo</b>	<p>Testar se a ferramenta possibilita anonimizar determinado grupo de dados pessoais, e que não possibilite a reversão.</p> <p>Atende as métricas de: Adequação, Operacionalidade, Conformidade, Tempo da Tarefa e Frequência de Erro.</p>
<b>Pré-condições</b>	Selecionar amostra de dados pessoais.
<b>Ação</b>	A ferramenta anonimiza os dados pessoais selecionados. Validar formas de reverter a anonimização pela própria ferramenta.
<b>Resultados Esperados</b>	Visualização dos dados pessoais anonimizados sem a possibilidade de reversão.

Fonte: Próprio autor.

Quadro 22 - Caso de Teste 12

<b>Caso de Teste</b>	Coexistência
<b>Resumo</b>	<p>Testar se a ferramenta executa sem apresentar restrições ou falhas ao operar em simultâneo com as ferramentas de banco de dados no mesmo ambiente.</p> <p>Atende as métricas de: Coexistência e Frequência de Erro.</p>
<b>Pré-condições</b>	Ferramentas instaladas.
<b>Ação</b>	Executar ferramentas em simultâneo com outras ferramentas.
<b>Resultados Esperados</b>	Ferramenta não apresenta restrições ou falhas ao ser executada simultaneamente com outras ferramentas.

Fonte: Próprio autor.

Os casos de testes têm como principal função contribuir para avaliar as ferramentas que foram selecionadas para a avaliação, por meio das métricas pré definidas, possibilitando que as ferramentas sejam avaliadas de forma justa e igual, testando todas as ferramentas da mesma maneira, assim tornando a avaliação das ferramentas mais precisa.

#### 4.5 VALIDAÇÃO DA PROPOSTA

Para realização das avaliações, por questões de segurança, optou-se em criar um ambiente de testes e não utilizar um estudo de caso real.

Ao criar o ambiente de testes para executar os casos de teste, é importante utilizar os mesmos dados, assim garantindo a padronização dos testes. Foram utilizados conjuntos de dados obtidos através do Dados.gov.br, que é um *site* disponibilizado pelo governo brasileiro que compartilha dados e informações públicas. Foram selecionados os conjuntos de dados sobre Afastamento Remunerado<sup>3</sup> e Motoristas Habilitados<sup>4</sup>. Esses conjuntos de dados foram

<sup>3</sup> **Dados.gov.br**. Disponível em: <https://dados.gov.br/dataset/afastamento-remunerado> Acesso em: 08 jun. 2021.

<sup>4</sup> **Dados.gov.br**. Disponível em: <https://dados.gov.br/dataset/motoristas-habilitados> Acesso em: 07 ago. 2021.

selecionados por possuírem dados pessoais em seu conteúdo, alguns dados são anonimizados para não causarem nenhum prejuízo.

Para testar as ferramentas selecionadas é necessário carregar estes dados em bancos de dados, para que as ferramentas de mapeamento de dados possam se conectar e extrair essas informações e assim realizar as análises.

Pelo estudo inicial, todas as ferramentas de mapeamento de dados suportam os bancos de dados Oracle Database XE, Microsoft SQL Server Express e MySQL. Desta forma, as ferramentas foram testadas individualmente, utilizando os casos de uso (seção 4.4) e os bancos de dados mencionados anteriormente.

A avaliação dos testes foi realizada utilizando os cálculos das métricas (seção 4.3). Após obter os resultados das métricas, foi aplicado o percentual sendo 70% para o critério de Conformidade e os demais critérios terão o peso dividido igualmente os 30% restantes. Após realizada a análise qualitativa, tendo como resultado a melhor ferramenta entre os critérios, assim atingindo o objetivo de identificar qual software possui maior aderência a LGPD.

#### 4.6 QUALIDADE DOS DADOS

A qualidade dos dados tem grande importância no processo de análise e classificação dos dados. É indispensável ter conhecimento dos dados utilizados para a criação dos padrões necessários.

O conjunto de dados do Afastamento Remunerado possui mais de 124.665 mil linhas de dados. O conjunto está dividido entre 11 colunas sendo elas nome, CPF, Descrição do cargo emprego, Nome do órgão de origem, UF da UPAG de vinculação, Cidade de residência, Nível de escolaridade, Ano/Mês início do afastamento, Valor rendimento líquido e Descrição do afastamento (Quadro 23).

Quadro 23 - Dicionário dos Dados Afastamento Remunerado

Campo	Tipo	Formato	Descrição
Nome	Texto(VARCHAR)		Nome do servidor afastado/licenciado.
CPF	Texto(VARCHAR)	***NNNNNN**	Número de inscrição no Cadastro de Pessoas Físicas da Receita Federal do Brasil.
Descrição do cargo emprego	Texto(VARCHAR)		Nome do Cargo ocupado de acordo com o plano de cargos e salários da carreira que estiver enquadrado / nome do emprego público ocupado.
Nome do órgão de origem	Texto(VARCHAR)		Nome do órgão público / Ente estatal Federal ao qual o servidor público está vinculado.
UF da UPAG de vinculação	Texto(VARCHAR)		Unidade da Federação da Unidade Pagadora.
Cidade de residência	Texto(VARCHAR)		Cidade declarada como de residência do servidor/empregado Público.
Nível de escolaridade	Texto(VARCHAR)		Nível escolaridade do cargo/emprego público.
Ano/Mês início do afastamento	Data	AAAAMM	Ano/Mês do início do Afastamento/Licença.
Ano/mês de referência	Data	AAAAMM	Ano/Mês de referência da extração do dado.
Valor rendimento líquido	Double		Remuneração servidor/empregado público
Descrição do afastamento	Texto(VARCHAR)		Afastamentos e Licenças descritos em lei.

Fonte: Brasil(2021).

Para realizar a análise desses dados foi verificada a quantidade de campos completos. Foram analisados os seguintes aspectos:

- a) Se haviam campos vazios, isto é, campos não preenchidos.
- b) Se haviam dados nulos ou seja, dados inconsistentes ou valores não esperados para a coluna especificada.

No conjunto de dados do afastamento remunerado tem a presença de dados vazios e nulos conforme o Quadro 24. Os dados vazios são considerados por não



afetarem a realização dos testes, já os dados nulos serão desconsiderados e removidos do conjunto de dados, já que podem interferir no resultado final dos testes aplicados.

Quadro 24 - Qualidade dos Dados Afastamento Remunerado

Coluna	Dados vazios	Dados nulos
Nome	0	0
CPF	0	0
Descrição do cargo emprego	10461	0
Nome do órgão de origem	0	0
UF da UPAG de vinculação	0	0
Cidade de residência	133	7
Nível de escolaridade	3487	0
Ano/Mês início do afastamento	0	0
Ano/mês de referência	0	0
Valor rendimento líquido	0	0
Descrição do afastamento	0	0

Fonte: Proprio autor.

O conjunto de dados sobre Motoristas Habilitados para a prestação do serviço de transporte rodoviário interestadual de passageiros possui 109.283 mil linhas de dados. O conjunto está dividido em 5 colunas sendo elas Razao\_social, Cnpj, Nome motorista, CPF e Vigência do cadastro (Quadro 25). Este conjunto não apresentou nenhum dado vazio ou nulo.

Quadro 25 - Dicionário dos Dados Motoristas Habilitados

<b>Campo</b>	<b>Tipo</b>	<b>Formato</b>	<b>Descrição</b>
razao_social	Texto(VARCHAR)		Razão social da empresa autorizada cadastrada na Receita Federal.
cnpj	Numérico(INT)	NN.NNN.NNN/NN NN-NN	Número da autorizatória no cadastro nacional de pessoa jurídica.
nome_motorista	Texto(VARCHAR)		Nome do motorista conforme consta na inscrição no Cadastro de Pessoas Físicas da Receita Federal do Brasil.
cpf	Numérico(INT)	NNN.***.***-NN	Número de inscrição no Cadastro de Pessoas Físicas da Receita Federal do Brasil.
vigencia_do_cadastro	Data	DD/MM/AAAA	Vigência do cadastro do motorista.

Fonte: Brasil(2021).

#### 4.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO

O capítulo apresentou a proposta de solução do trabalho, no que se refere a analisar ferramentas de mapeamento de dados, a fim de averiguar quais dessas ferramentas têm maior aderência com a norma a LGPD.

Na seção 4.1 são conhecidas as ferramentas que serão avaliadas pelo trabalho, foram estudadas diversas ferramentas e selecionadas três que são Talend Open Studio, Pentaho Kettle e CloverDX.

Após a definição das ferramentas que foram avaliadas, são selecionados os critérios na seção 4.2, os critérios utilizados na avaliação foram de características de qualidade externa e características de qualidade em uso. As métricas definidas na seção 4.3, definem a mensuração dos critérios selecionados, métricas externas são as medições do software considerando a execução do software e de seus resultados no ambiente. As métricas de qualidade de uso são medições do software

observando o comportamento e qualidade do seu sistema em tarefas e cenários do dia a dia do usuário.

Já os casos de testes definidos na seção 4.4 são necessários para reger um padrão de testes na avaliação das ferramentas. Eles funcionam como caminho para chegar no resultado esperado. A seção 4.5 apresenta a validação de como foi o ambiente da avaliação, e como atingir o objetivo proposto pelo trabalho. O capítulo tem importância para entender e desenvolver os testes realizados no Capítulo 5.

## 5 TESTE DAS FERRAMENTAS

Conforme o que foi definido na proposta de solução (Capítulo 4) , as ferramentas escolhidas para verificar a aderência a LGPD e a norma ABNT NBR 27701 foram o Talend Open Studio, Pentaho Kettle e CloverDX. Os testes realizados seguiram os casos de testes definidos na seção 4.4.

Para a utilização dos testes é importante a utilização dos mesmos dados. Os testes foram realizados no ambiente definido na seção 4.5.

### 5.1 TALEND OPEN STUDIO

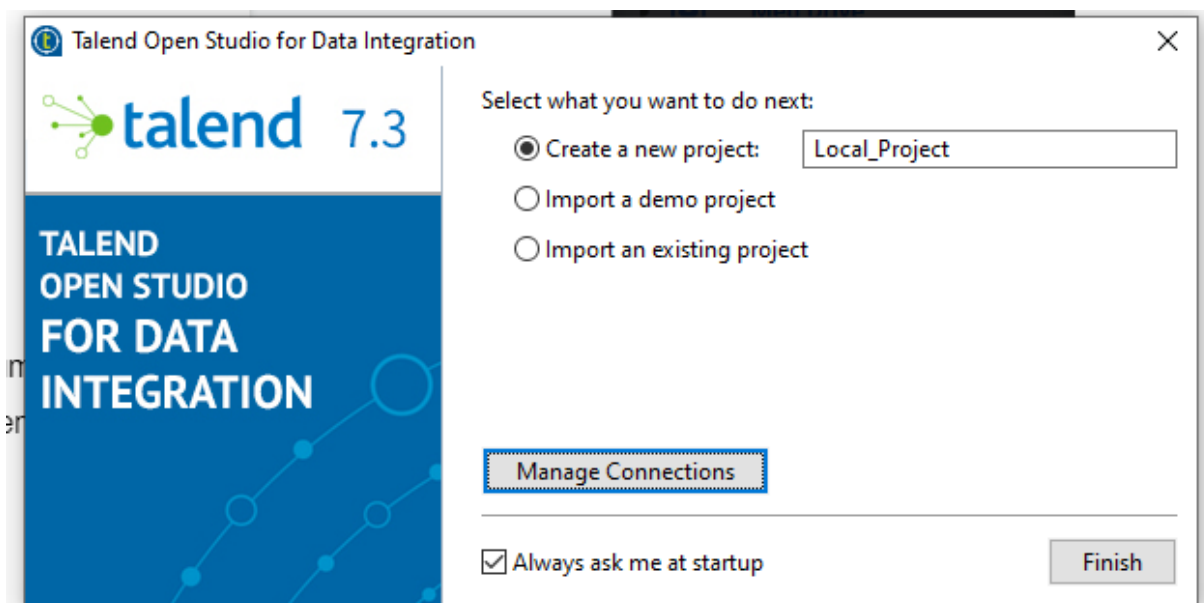
O Talend Open Studio é uma ferramenta desenvolvida pela Talend para a integração de dados de forma gratuita. A Talend disponibiliza outras ferramentas para a integração de dados como o Talend Data Fabric, que é uma opção de ferramenta paga.

O processo de instalação é bem simples. Os requisitos mínimos para um bom funcionamento é 4GB de memória RAM e ter o Java JRE 8 instalado previamente. A instalação consiste em executar o arquivo que foi baixado e que em seguida irá descompactar os arquivos da ferramenta no local desejado, após a descompactação já pode ser realizada a execução da ferramenta. Mesmo que a linguagem padrão da ferramenta seja definida como português, algumas telas podem aparecer no idioma inglês.

#### 5.1.1 Caso de Teste 1: Parametrização da ferramenta

O caso de teste 1 (Quadro 11) consiste em testar como a ferramenta pode ser parametrizada e conectada às fontes de dados (Seção 4.5). Após a instalação e execução da ferramenta foi necessário realizar as configurações primárias, na tela inicial criando um novo projeto conforme a Figura 20.

Figura 20 - Tela inicial Talend Open Studio

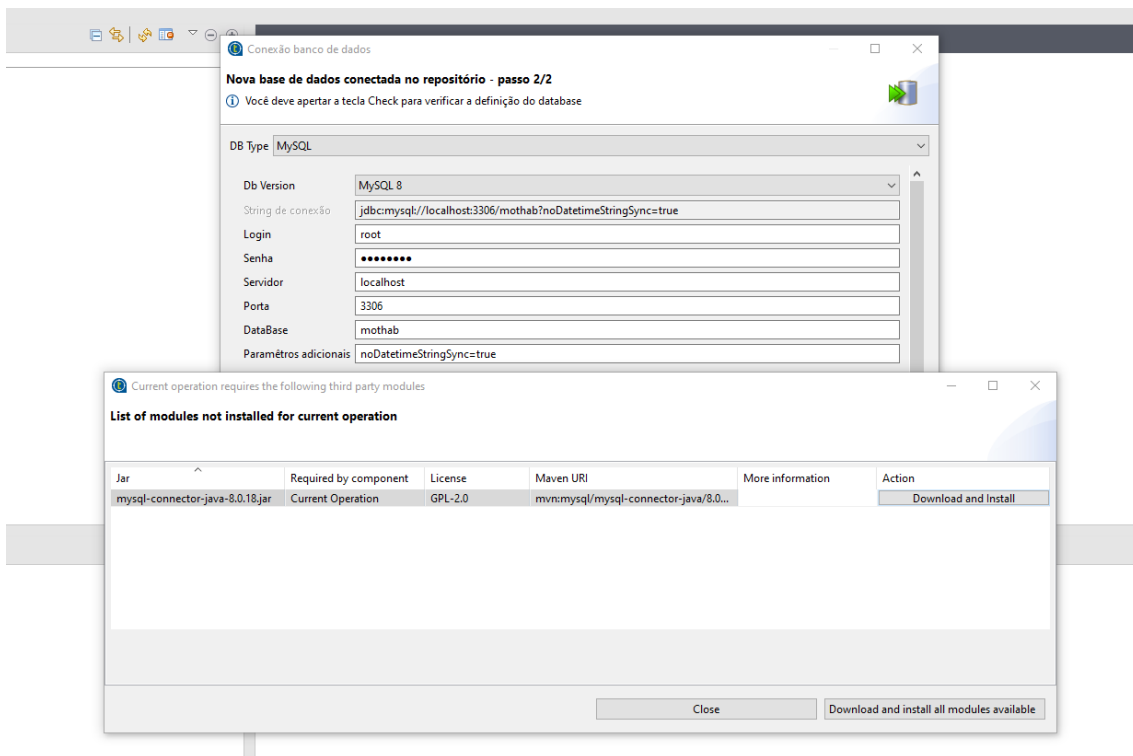


Fonte: Próprio autor.

Seguindo o teste, foi realizada a parametrização para se conectar às fontes de dados MySQL, Oracle XE e SQL Server definidas anteriormente na seção 4.5.

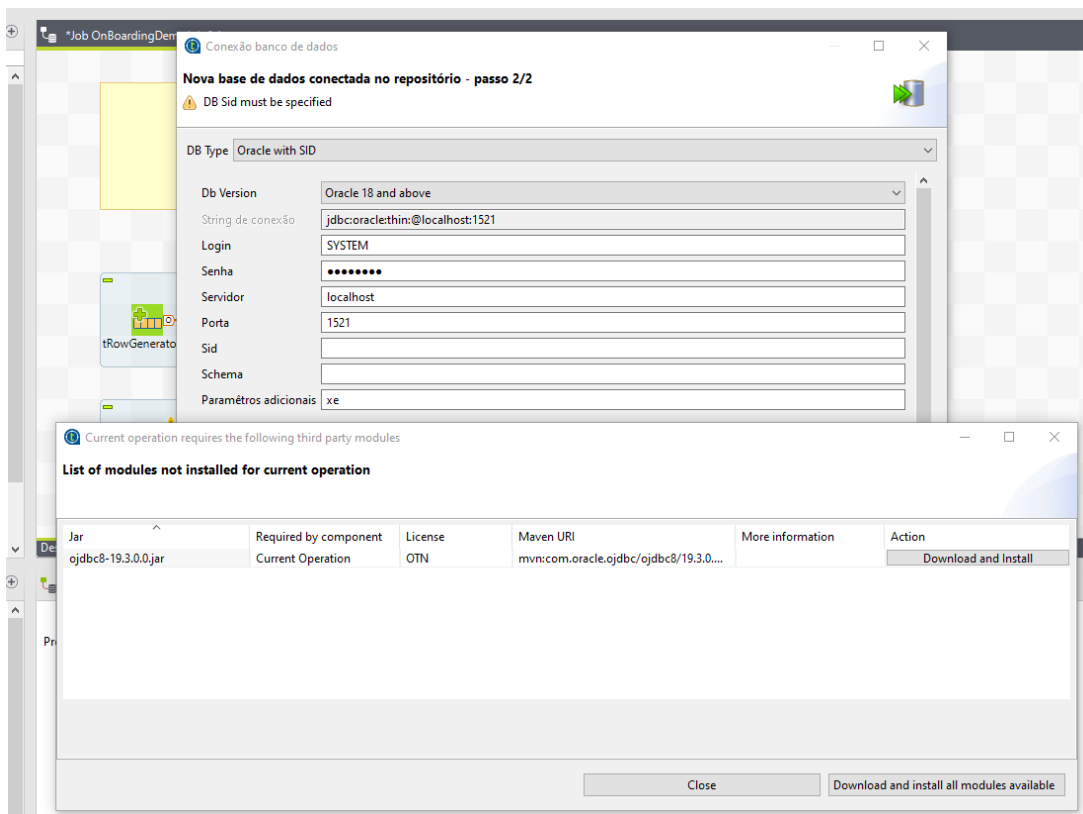
As Figuras 21 e 22 a seguir, mostram a conexão do Talend ao banco de dados MySQL e Oracle. Após preenchido os parâmetros necessários para a conexão, o Talend solicitou o *download* de um módulo adicional afim de realizar a conexão ao banco de dados.

Figura 21 - Talend Open Studio Conexão ao MySQL



Fonte: Próprio autor.

Figura 22 - Talend Open Studio Conexão ao Oracle

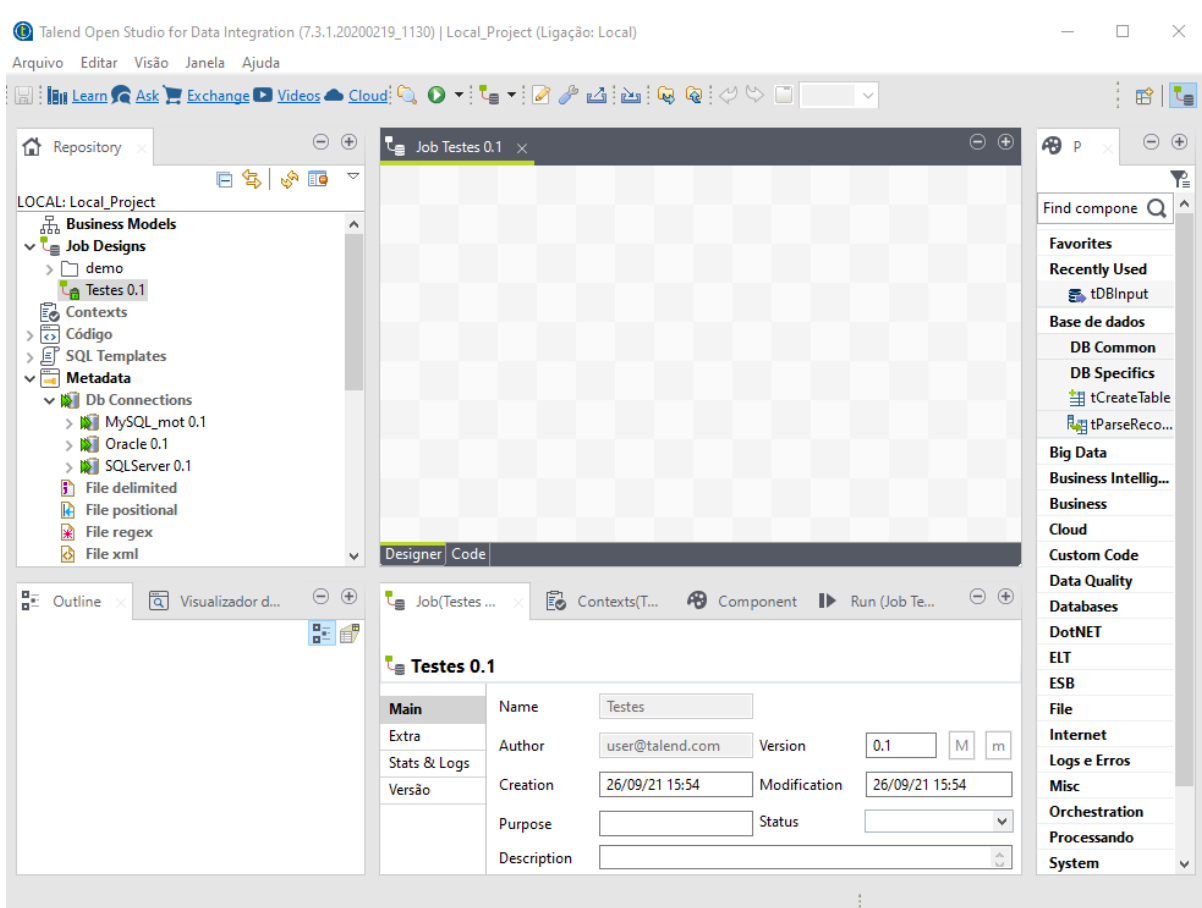


Fonte: Próprio autor.

Ao adicionar o banco de dados SQL Server, foi necessário realizar o *download* de um *plugin* mssql-jdbc.jar diretamente do site do fabricante, devido o Talend Open Studio não realizar o *download* automático.

Após o sistema estar parametrizado a execução do caso do teste 1 termina. O resultado esperado é o sistema configurado conforme a Figura 23, onde todos os bancos de dados estão conectados.

Figura 23 - Talend Open Studio parametrizado



Fonte: Próprio autor.

### 5.1.2 Caso de Teste 2: Avaliação de riscos dos dados

O caso de teste 2 (Quadro 12), consiste em testar se a ferramenta é capaz de detectar os riscos existentes e medir a probabilidade de possíveis riscos acontecerem, além de gerar um relatório destes riscos. A ferramenta Talend Open

Studio não foi capaz de produzir os resultados esperados, a mesma não foi capaz de avaliar a probabilidade de risco, muito menos de gerar relatórios de avaliação de risco. Em consulta à página oficial do fabricante, em versões pagas da ferramenta, é possível realizar essas ações, mas as mesmas ações não estão disponíveis na versão Talend Open Studio. O resultado esperado, que era obter um relatório ou visualizar os riscos encontrados pela ferramenta, não foi alcançado.

### **5.1.3 Caso de Teste 3: Segurança de Acesso**

No caso de teste 3 (Quadro 13) é verificado se a ferramenta possui mecanismos que gerenciem o acesso de usuários, assim garantindo maior segurança para o uso da ferramenta. A ferramenta não atende os requisitos para realização do mesmo, a ferramenta não possui um gerenciador de usuários ou a possibilidade de definir uma senha para a abertura do projeto, o que possibilita que alguém copie os arquivos do projeto e abra em outros dispositivos assim visualizando as tarefas ou até mesmo se conectando as fontes de dados caso tenha acesso. Em outras versões pagas do Talend, como no Talend Cloud ou Talend Data Integration, conforme a documentação disponibilizada pelo fabricante, é possível realizar esse controle. Assim sendo, o resultado esperado era que a ferramenta negasse o acesso não autorizado, o que não ocorreu.

### **5.1.4 Caso de Teste 4: Identificação de dados sensíveis**

O resumo do caso de teste 4 é se a ferramenta possibilita realizar a identificação dos dados sensíveis. O Talend Open Studio não realizou a classificação dos dados, somente em sua versão paga existe um componente chamado tClassify que realiza a classificação através de técnicas de inteligência artificial (TALEND, 2021). Não foi possível atingir o resultado esperado devido à falta de recurso pela ferramenta.



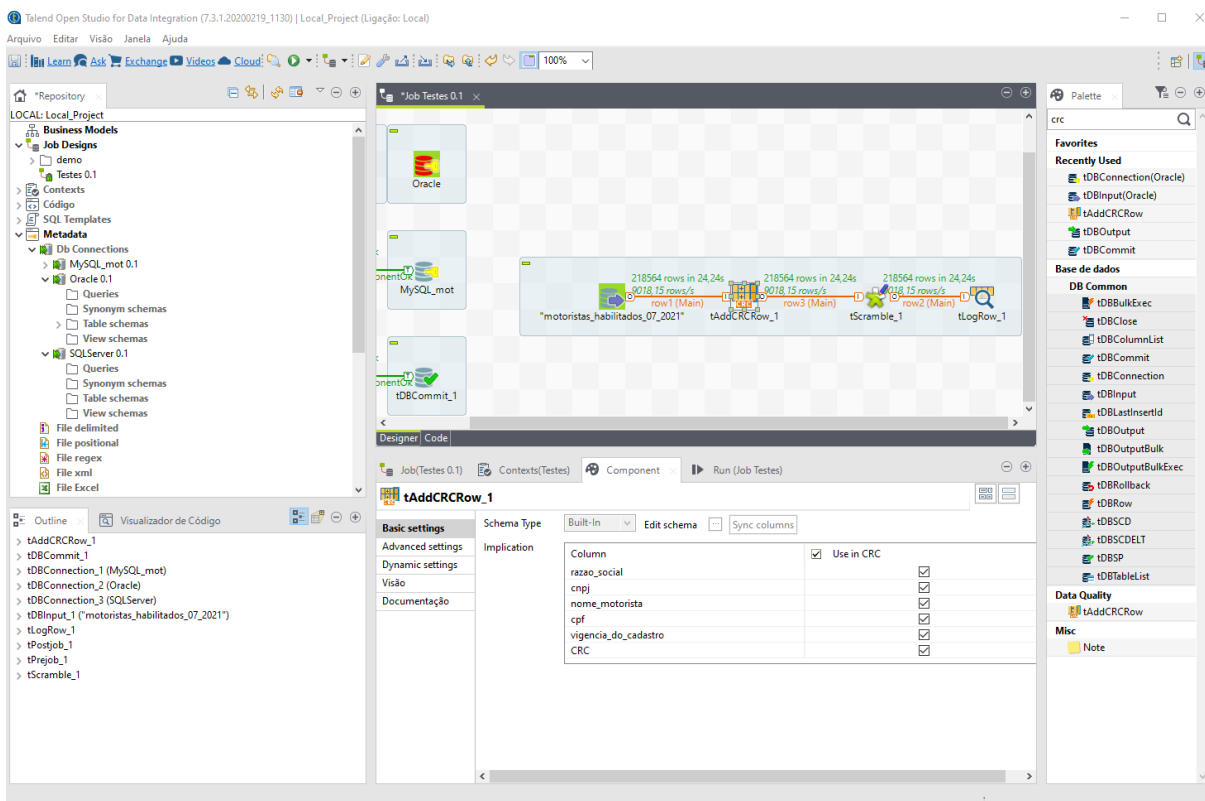
### 5.1.5 Caso de Teste 5: Rotulação dos dados

O caso de teste 5 resume-se em testar se a ferramenta possibilita correlacionar rótulos aos seus dados pessoais. O Talend Open Studio não possibilitou a inserção de rótulos, não possui nenhum componente que possibilite a descrição dos dados tratados. Deste modo o resultado esperado não pode ser atingido, e não se pode identificar os dados sensíveis apenas pela sua visualização.

### 5.1.6 Caso de Teste 6: Criptografia dos dados

O caso de teste 6 (Quadro 16) consiste em testar se a ferramenta é capaz de criptografar os dados pessoais. O Talend Open Studio não possui nenhum componente nativo capaz de realizar a criptografia dos dados. Porém, através de um componente de terceiros chamado tScramble, é possível realizar a criptografia e a descryptografia. Esse componente é disponibilizado em uma loja de componentes no site da Talend. Primeiramente foi necessário utilizar um *plugin* chamado tAddCRC, que realizou a verificação cíclica de redundância (CRC32), que é um método projetado para proteger os dados contra erros e garantir a qualidade dos mesmos durante todo processo, sendo adicionado uma coluna chamada CRC que contém *checksum* dos dados (Figura 24).

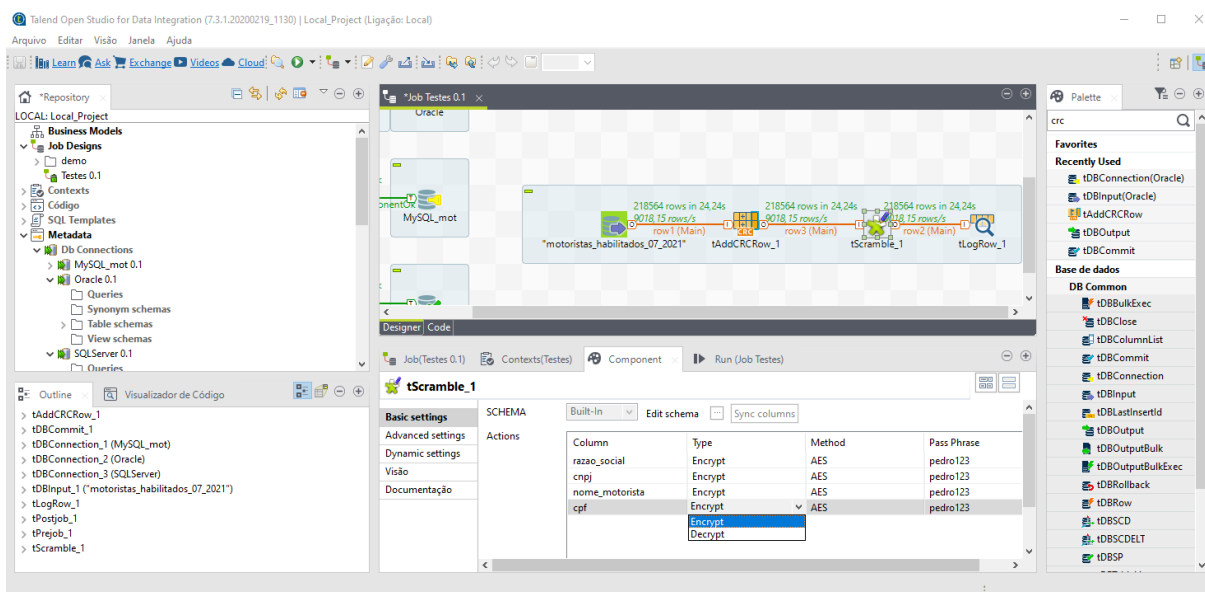
Figura 24 - Talend Open Studio aplicando CRC



Fonte: Próprio autor.

Após os dados passarem pelo processo de CRC, os mesmos são enviados para o componente tScramble, esse que aplica os métodos de criptografia AES ou DES, utilizando uma *passphrase* que funciona como uma senha. O processo de descriptografia funciona da mesma forma, apenas é necessário mudar o tipo de operação para descriptografia e fornecer a *passphrase* utilizada na criptografia (Figura 25).

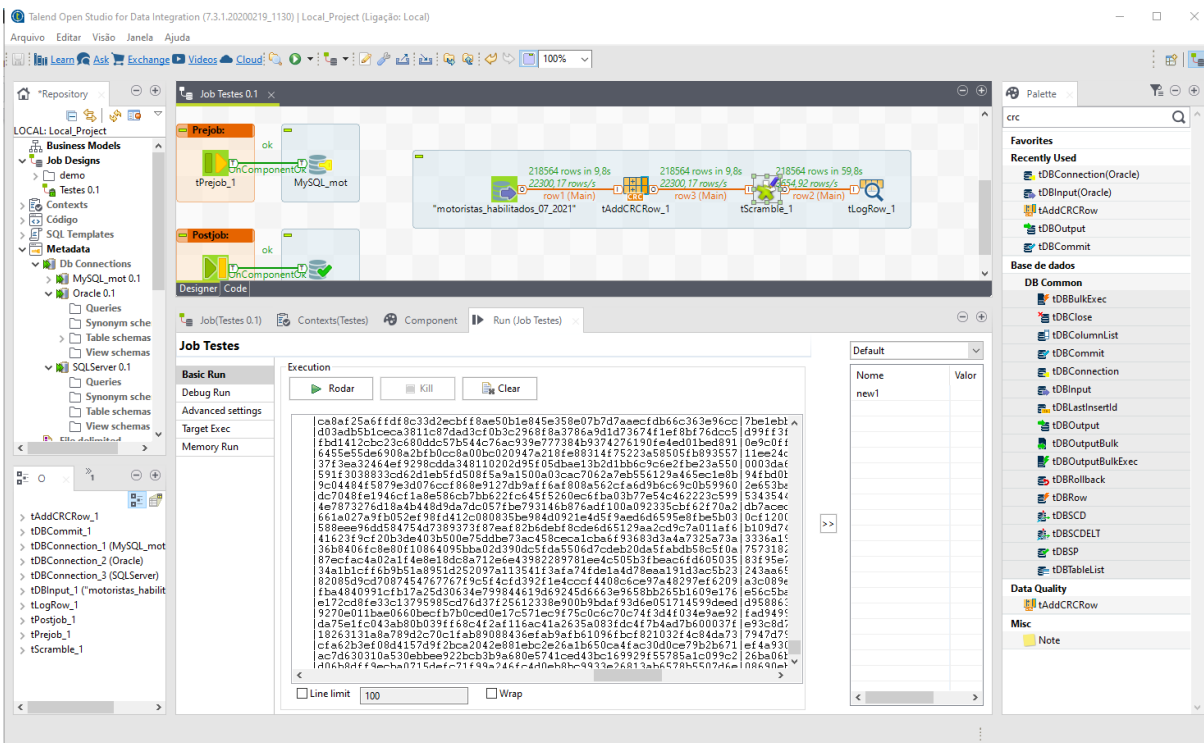
Figura 25 - Talend Open Studio configurando tScramble



Fonte: Próprio autor.

Com a execução da tarefa foi possível obter o resultado esperado, e todos os dados que foram selecionados (razao\_social, cnnpj, nome\_motorista, cpf) para passarem pelo processo foram criptografados conforme a Figura 26 e Figura 27. Alterando o tipo para descriptografia, foi possível retornar os dados aos valores originais.

Figura 26 - Talend Open Studio dados criptografados



Fonte: Próprio autor.

Figura 27 - Talend Open Studio dados criptografados

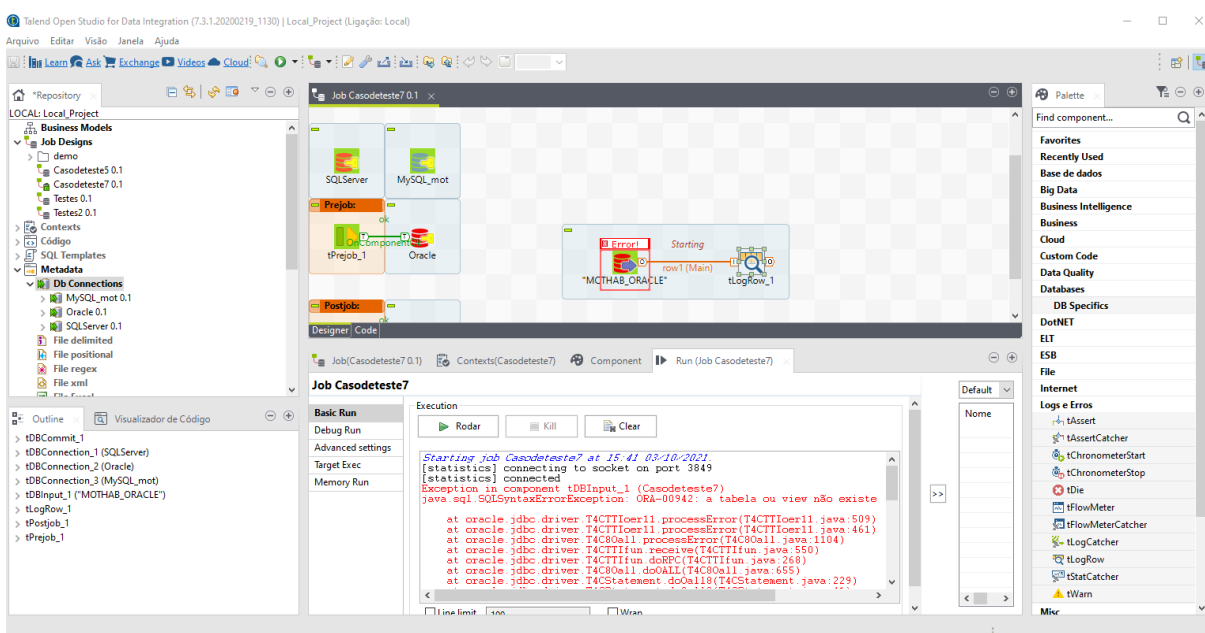


Fonte: Próprio autor.

### 5.1.7 Caso de Teste 7: Armazenamento de eventos

Testar se a ferramenta é capaz de armazenar os eventos(*logs*) ocorridos é o caso de teste 7 (Quadro 17). O Talend Open Studio possui uma série de componentes que permite verificar erros. O principal componente para este caso de teste é o tLogRow, que coleta todos os eventos realizados em um projeto. Os eventos como, alterações nos componentes da tarefa e erros na execução da tarefa, apenas são mostrados na tela e não são salvos. Esperava que a ferramenta armazenasse todos os eventos realizados na ferramenta desde alterações nos projetos como de configurações da ferramenta . A Figura 28 e 29 mostra o resultado do tLogRow onde apresenta o logs de execução.

Figura 28 - Talend Open Studio armazenamento de logs



Fonte: Próprio autor.

Figura 29 - Talend Open Studio resultado armazenamento de logs

```

Starting Job Testes at 20:24 07/10/2021
[statistics] connecting to socket on port 3472
[statistics] connected
Exception in component tFileOutputDelimited_1 (Testes)
java.io.FileNotFoundException: D:\Talend\TOS_DI-Win32-20200219_1130-V7.3.1\workspace\out.csv (O arquivo já está sendo usado por outro processo)
at java.io.FileOutputStream.open0(Native Method)
at java.io.FileOutputStream.open(Unknown Source)
at java.io.FileOutputStream.<init>(Unknown Source)
at java.io.FileOutputStream.<init>(Unknown Source)
at local_project.testes_0_1.Testes.tDBInput_1Process(Testes.java:1902)
at local_project.testes_0_1.Testes.runJobInTOS(Testes.java:3348)
at local_project.testes_0_1.Testes.main(Testes.java:3179)
[statistics] disconnects
Job Testes ended at 20:24 07/10/2021. [exit code = 1]

```

Fonte: Próprio autor.

O resultado esperado não foi atingido pela ferramenta, ela só coletou as ações realizadas durante o projeto, caso seja alguma ação realizada nas configurações do Talend por exemplo, elas não são coletadas. O principal requisito que faz a ferramenta não atingir o objetivo foram os eventos coletados não foram armazenados.

#### **5.1.8 Caso de Teste 8: Resposta a incidentes de segurança**

O caso de teste 8 (Quadro 18) corresponde a como a ferramenta identifica as consequências e o período de tempo do ocorrido de algum incidente. Através do sistema de gerenciamento de banco de dados (SGBD) foi realizada para o teste a exclusão de forma aleatória de 20 mil de dados sensíveis da coluna nome\_motorista e cpf porém a ferramenta Talend Open Studio não foi capaz de identificar que estes dados foram apagados. Retornando a fonte de dados para o estado original, foi realizada a alteração de 8 mil dados para "XXXX" de uma única vez, para simular um volume incomum da alteração, mas não foram detectadas pelo Talend que apenas mostra os dados alterados.

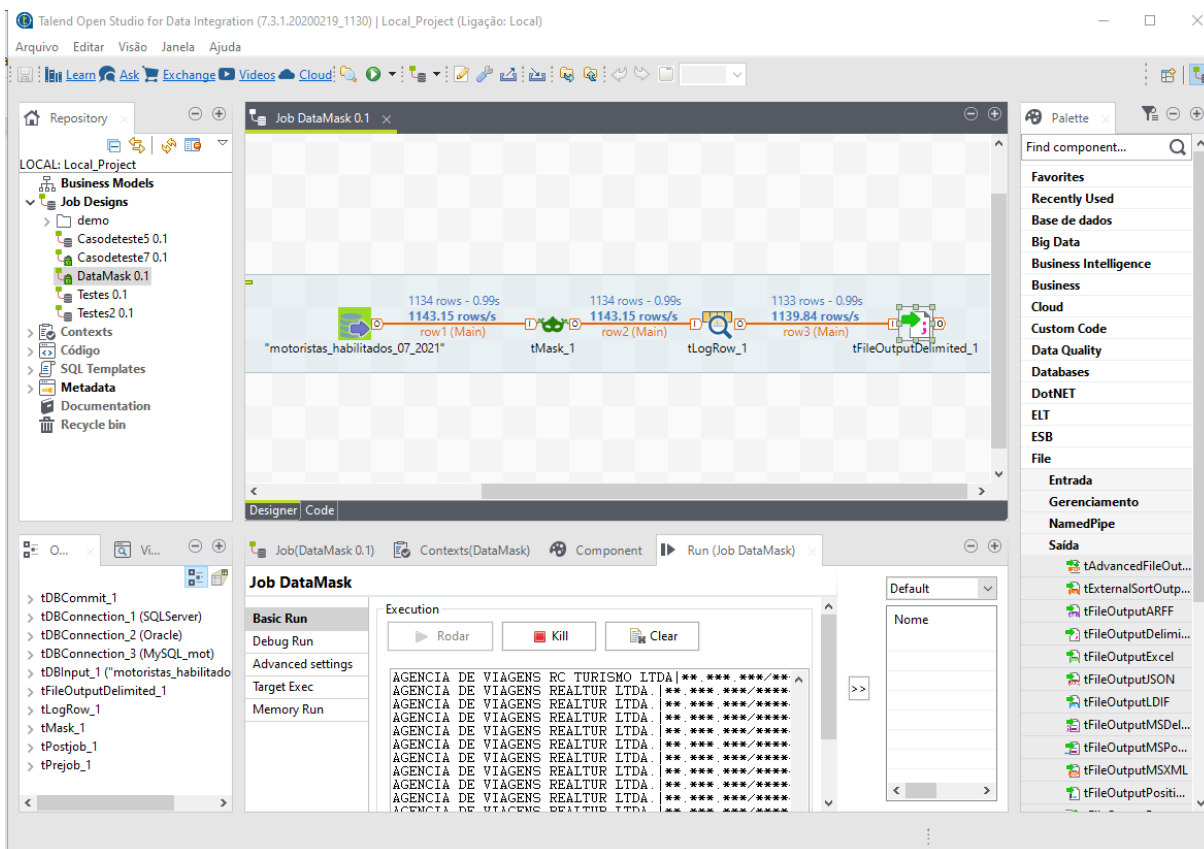
A ferramenta não pode atingir o resultado esperado, ela não detectou a ocorrência de incidentes nos dados armazenados pelas fontes de dados. Também não permitiu detectar o período de tempo da ocorrência de segurança.

#### **5.1.9 Caso de Teste 9: Análise crítica técnica do compliance**

No caso de teste 9 (Quadro 19) especifica-se que a ferramenta pode identificar qual parte do ciclo do tratamento dos dados está sendo realizada. A ferramenta Talend Open Studio não pode determinar em que fase do tratamento de dados está sendo realizada, com o seu objetivo de desenvolvimento a integração de dados não possui componentes que são voltados para a conformidade. Pode-se identificar somente as partes do tratamento de dados que está sendo feito pelo Talend. Conforme a Figura 30 é possível ver que os dados estão passando para o

componente tMask e sendo mascarados, na parte inferior é possível ver os dados em tempo real que estão sendo transformados por este componente.

Figura 30 - Talend Open Studio tratamento dos dados



Fonte: Próprio autor.

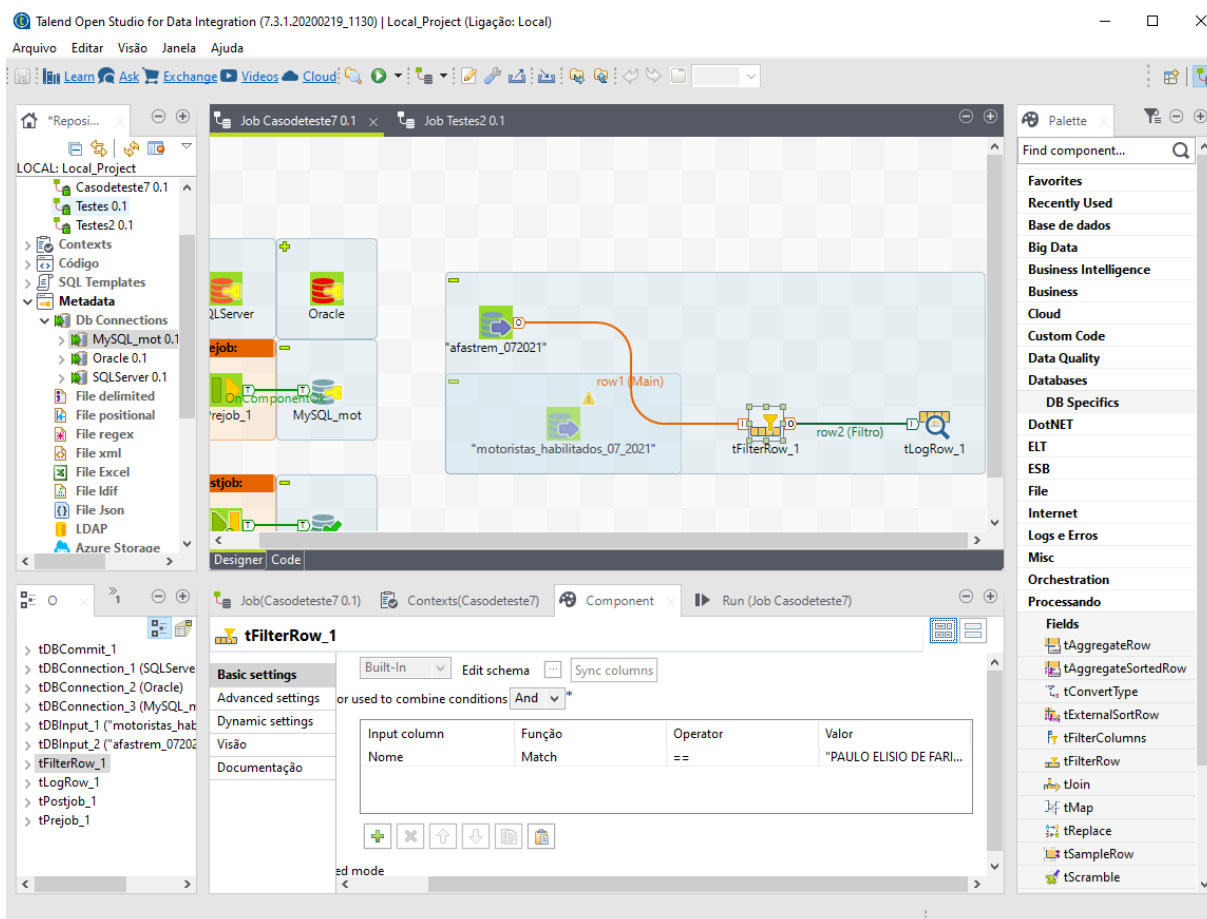
Mesmo que a ferramenta mostre o ciclo de dados em que ela faz a transformação dos dados, a ferramenta não pode atingir o resultado esperado que é a identificação do ciclo de tratamento dos dados referente a todas operações realizadas nas fontes de dados .

### 5.1.10 Caso de Teste 10: Entrega de dados ao titulares

No caso de teste 10 (Quadro 20), é especificado se a ferramenta é capaz de identificar os dados pessoais de determinado titular quando solicitado. O Talend Open Studio possui um componente para a realização de filtros chamado tFilterRow,

e é possível filtrar os dados de diversas fontes de dados. Ao determinar o nome de um titular dos dados foi possível identificar todos os dados armazenados para o mesmo (Figura 31).

Figura 31 - Talend Open Studio aplicar filtro aos dados



Fonte: Próprio autor.

O resultado atingido é o esperado, a ferramenta é capaz de apresentar todos os dados de determinado titular.

### 5.1.11 Caso de Teste 11: Anonimização

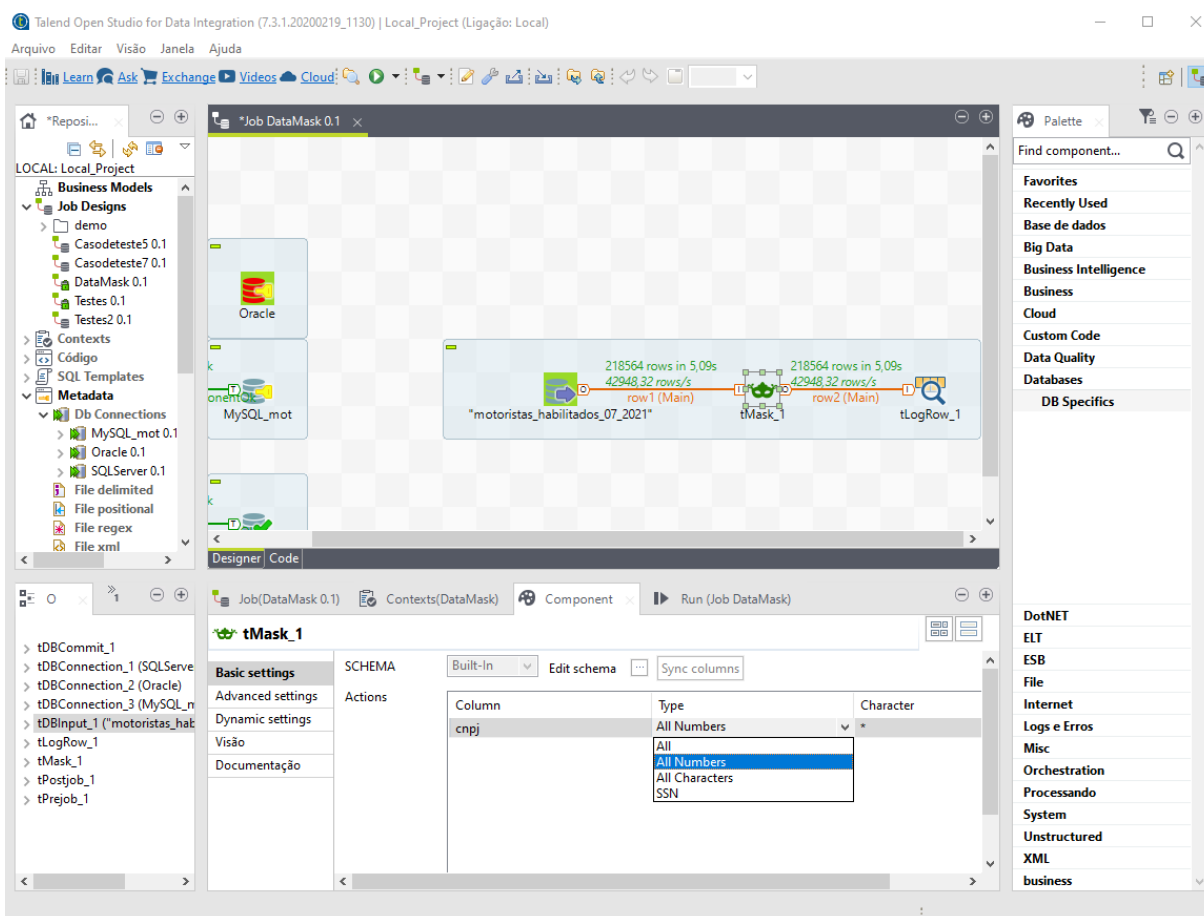
O caso de teste 11 (Quadro 21) compreende em testar se a ferramenta pode anonimizar dados pessoais e não possibilitar a sua reversão. O Talend Open Studio não possui nenhum componente nativo que realize a anonimização dos dados. Por



meio de um componente de terceiros tMask, foi possível realizar a anonimização dos dados. Este componente está disponível na loja de componentes no site da Talend.

Para a utilização do componente tMask, são selecionadas as colunas para realizar a anonimização. É possível selecionar se deseja anonimizar somente números ou todos os caracteres (Figura 32). No caso de somente os números serem anonimizados por exemplo o dado de uma placa de veículo que é IYF-6661, após o processo de anonimização seria IYF-\*\*\*\*\*.

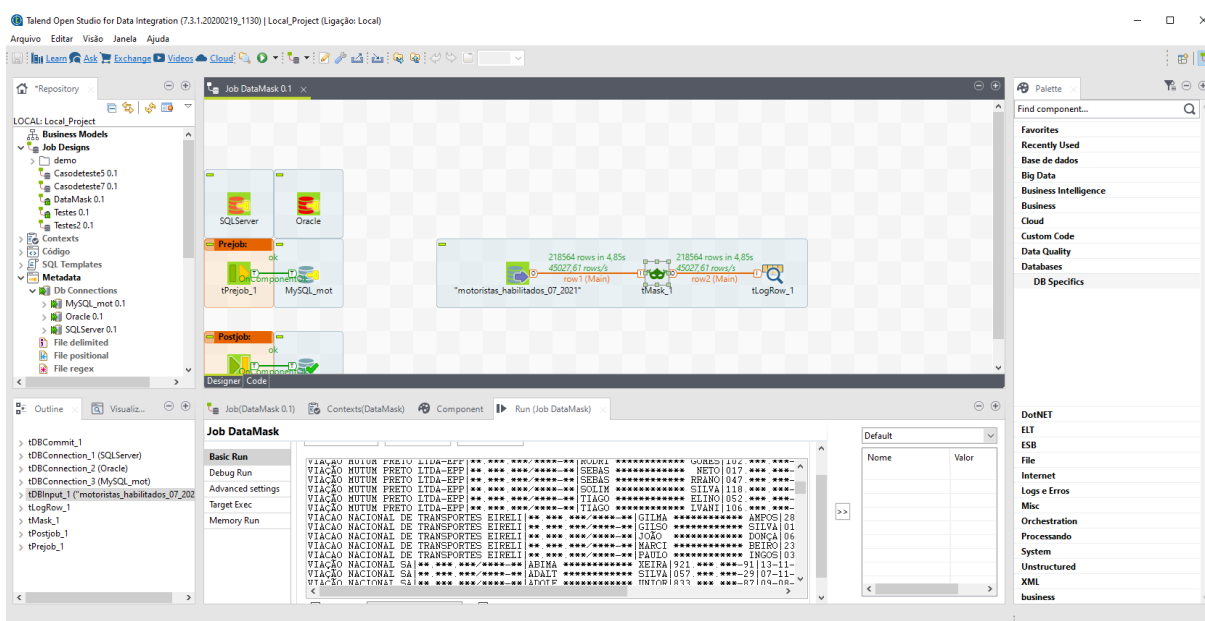
Figura 32 - Talend Open Studio configuração de anonimização



Fonte: Próprio autor.

O resultado esperado foi alcançado pela ferramenta que após o processo de anonimização da coluna cnpj, apresentou os dados como \*\*.\*\*\*.\*\*\*/\*\*\*.\* (Figura 33). A ferramenta não possibilita realizar a reversão dos dados anonimizados após sua gravação na fonte de dados.

Figura 33 - Talend Open Studio resultado da anonimização



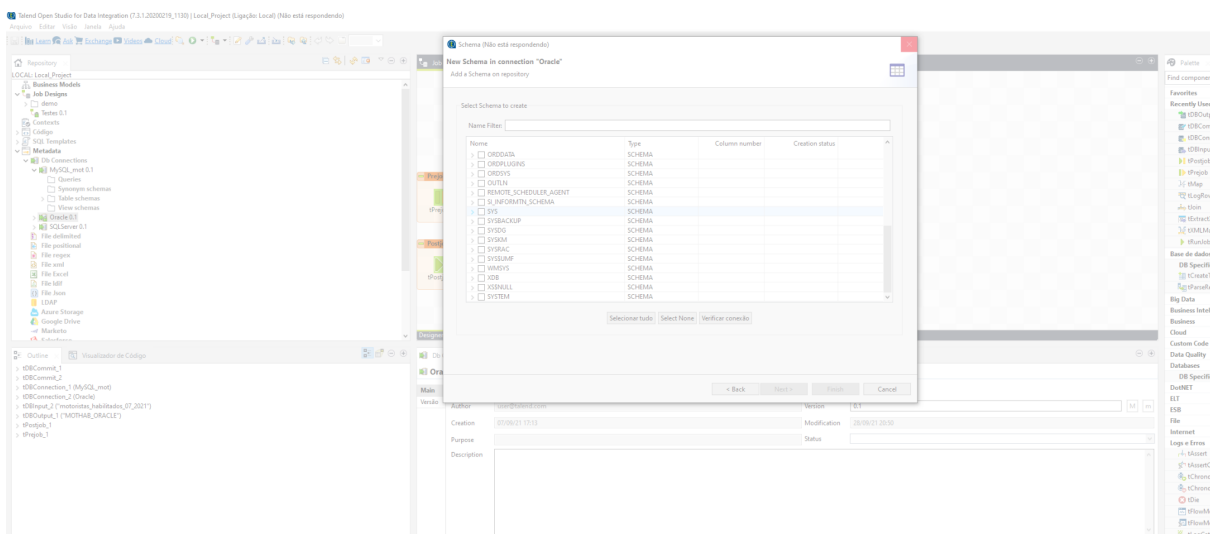
Fonte: Próprio autor.

### 5.1.12 Caso de Teste 12: Coexistência

O caso de teste 12 (Quadro 22) se faz necessário para testar se a ferramenta executa sem apresentar falhas ao operar em simultâneo com as ferramentas de banco de dados no mesmo ambiente. O teste é feito em um sistema operacional Windows 10, utilizando um processador AMD Ryzen 5 2600 e 16GB de memória RAM, na execução do teste apenas é executado tarefas próprias do sistema operacional e os bancos de dados SQL Express, Oracle XE e MySQL em conjunto com Talend Open Studio.

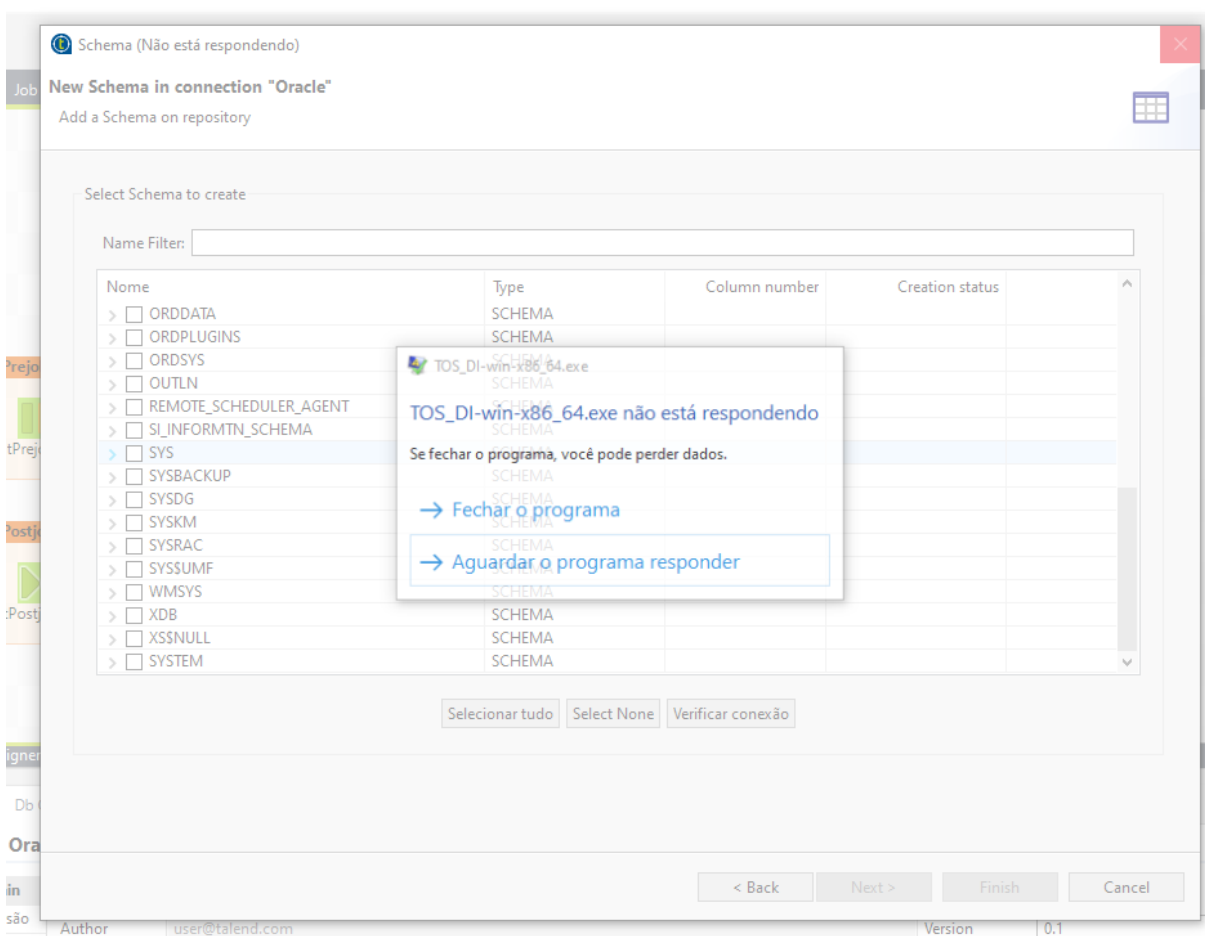
Ao conectar a ferramenta ao banco de dados do Oracle XE a ferramenta parou de responder e demorou 1 minuto e 22 segundos para se recuperar (Figura 34 e 35). Com os bancos de dados MySQL e SQL Express a ferramenta não apresentou travamentos no processo de conexão dos dados.

Figura 34 - Talend Open Studio apresenta travamentos com Oracle XE



Fonte: Próprio autor.

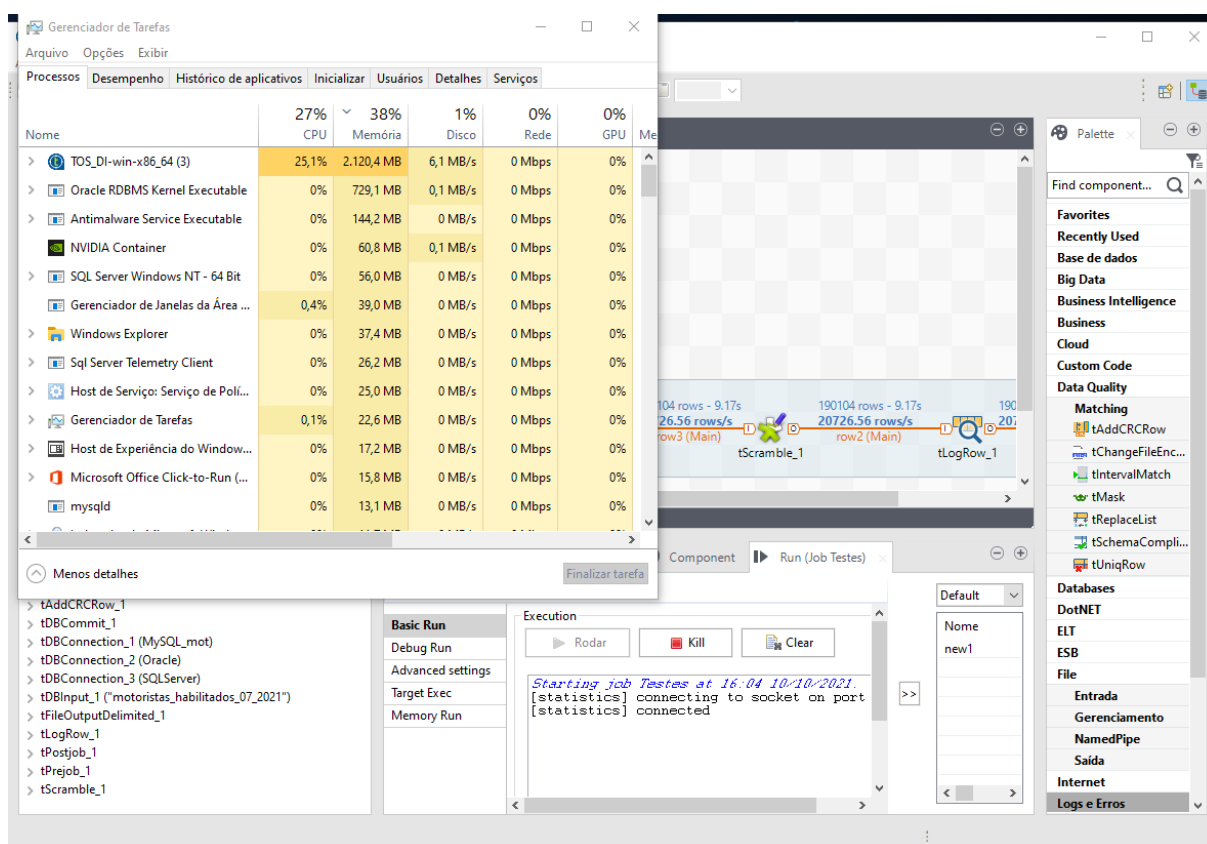
Figura 35 - Talend Open Studio apresenta travamentos com Oracle XE



Fonte: Próprio autor.

Quando monitorado o uso de recursos do computador na execução do caso de teste 6 (Seção 5.1.6), o processo do Talend Open Studio que tem o nome de TOS\_DI-win-x86\_64 apresentou uso de 2GB de memória RAM conforme Figura 36. Nenhum travamento foi apresentado pelo Talend Open Studio ou por algum dos bancos de dados.

Figura 36 - Talend Open Studio uso de recursos



Fonte: Próprio autor.

O resultado esperado foi atingido parcialmente, devido aos travamentos e lentidão que ocorreram nas operações de conexão do Talend com o Oracle XE, as mesmas operações realizadas nos demais bancos de dados não apresentaram problemas. Durante a execução dos demais casos de testes a ferramenta não apresentou travamentos.

## 5.2 PENTAHO KETTLE

O Pentaho Kettle Community Edition é uma ferramenta de código aberto para a extração, carregamento e transformação de dados. É desenvolvida na linguagem de programação Java e mantida pela Hitachi Vantara Corporation. Também existe o Pentaho Enterprise Edition, uma versão paga do Pentaho que possui maiores funcionalidades.

A Pentaho Kettle possui uma interface intuitiva na qual possibilita o usuário a arrastar e soltar os componentes para montar o fluxo de trabalho desejado. A Hitachi Vantara mantém o Pentaho Marketplace, que é possível compartilhar e obter novos *plugins* e componentes para disponibilizar mais funcionalidades ao Pentaho Kettle.

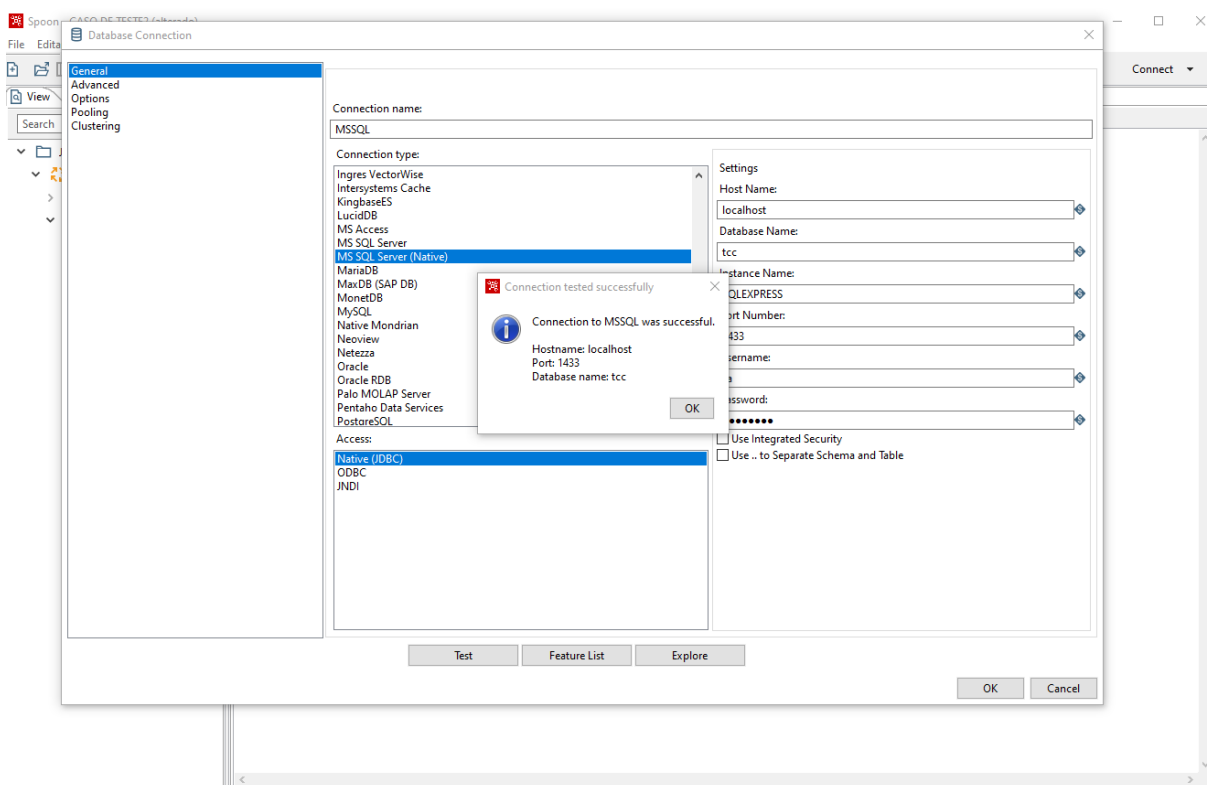
Para a instalação do Pentaho Kettle é necessário ter instalado o JDK/JRE do Java como pré-requisito. O processo de instalação é descomplicado, consiste em executar o arquivo de instalação e selecionar seu local de instalação, ao acessar a pasta escolhida deve ser executado primeiramente o Spoon.bat que depois irá executar o Pentaho Kettle.

### 5.2.1 Caso de Teste 1: Parametrização da ferramenta

O caso de teste 1 inclui parametrizar a ferramenta e se conectar às fontes de dados. É realizada a parametrização para se conectar às fontes de dados MySQL, Oracle XE e SQL Server definidas anteriormente na seção 4.5.

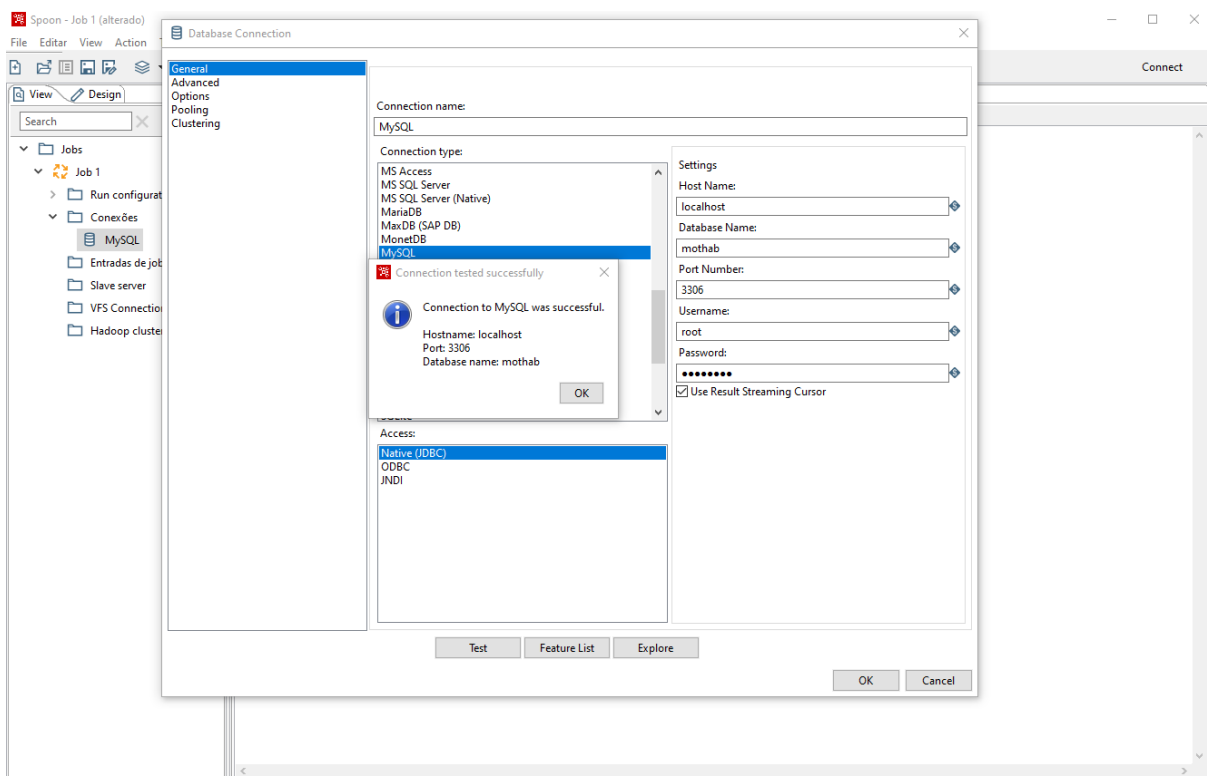
Na conexão com o banco de dados MySQL, Oracle e SQL Server a ferramenta apresentou o erro que não poderia se conectar por falta de *driver* para realizar a conexão. Foi necessário realizar o *download* do *driver* JDBC de cada banco de dados, e extraí-lo para a pasta de instalação do Pentaho. Com o drive instalado e pós preenchido os parâmetros de endereço, *login* e senha do banco de dados a conexão foi realizada com sucesso a todos os bancos de dados conectados conforme Figuras 37, 38 e 39.

Figura 37 - Pentaho Kettle conectado com sucesso ao MySQL



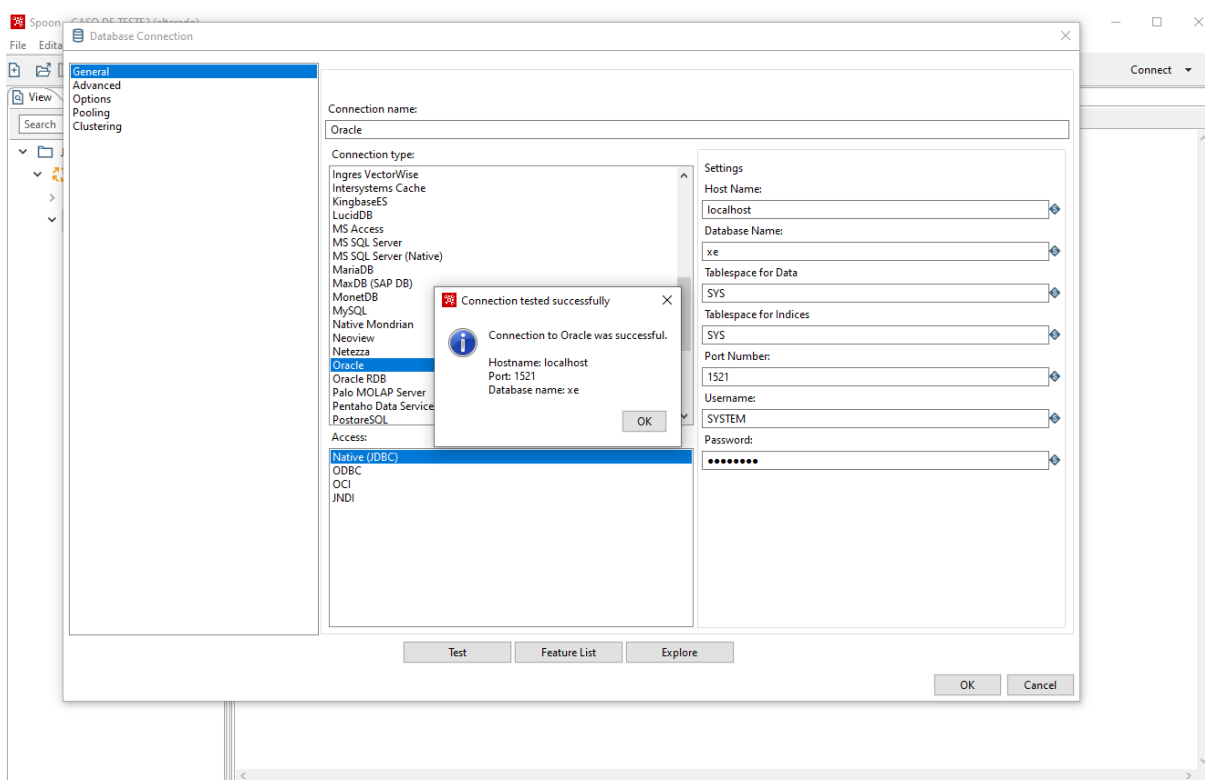
Fonte: Próprio autor.

Figura 38 - Pentaho Kettle conectado com sucesso ao SQL Server



Fonte: Próprio autor.

Figura 39 - Pentaho Kettle conectado com sucesso ao Oracle XE



Fonte: Próprio autor.

O resultado esperado foi atingido, mas houve dificuldade em localizar os *drivers* JDBC para realizar o *download*, após a instalação a conexão com as fontes de dados funcionaram com sucesso.

### 5.2.2 Caso de Teste 2: Avaliação de riscos dos dados

O caso de teste 2 (Quadro 12), corresponde em testar se a ferramenta é capaz de detectar os riscos existentes e medir a probabilidade de os riscos acontecerem, além de gerar um relatório desses riscos. Pentaho Kettle não pode produzir os resultados esperados, pois não avaliou a probabilidade de risco e muito menos gerou um relatório de avaliação desses riscos. A Hitachi Vantara possui uma ferramenta paga para o gerenciamento de dados chamada Lumada, que conforme sua documentação realiza a avaliação de riscos, porém o Pentaho Kettle não

apresenta nenhum recurso para isso. O resultado esperado que era obter um relatório ou visualizar os riscos encontrados pela ferramenta não foi alcançado.

### **5.2.3 Caso de Teste 3: Segurança de Acesso**

No caso de teste 3 (Quadro 13) é verificado se a ferramenta possui mecanismos que gerenciam o acesso de usuários, assim garantindo maior segurança para o uso da ferramenta. O Pentaho Kettle não possui um gerenciador de usuários ou a possibilidade de definir uma senha para abrir o projeto. Isso permite que alguém copie o arquivo do projeto e abra-o em outro dispositivo para ver as tarefas ou até mesmo se conectar à fonte de dados. A versão Pentaho Enterprise permite a criação de um repositório que serve como servidor, no qual os demais usuários se conectam nele, e esse sim permite o gerenciamento de usuários. O resultado esperado era que a ferramenta negasse o acesso não autorizado, o que não foi satisfeito.

### **5.2.4 Caso de Teste 4: Identificação de dados sensíveis**

O resumo do caso de teste 4 é observar se a ferramenta possibilita realizar a identificação dos dados sensíveis. O Pentaho Kettle não realizou a identificação dos dados. Não foi possível atingir o resultado esperado devido a falta de recurso pela ferramenta.

### **5.2.5 Caso de Teste 5: Rotulação dos dados**

No caso de teste 5 deve-se testar se a ferramenta possibilita adicionar rótulos aos seus dados pessoais. O Pentaho possui um componente capaz de inserir valores nas tabelas existentes, desta forma possibilitando adicionar rótulos em locais que armazenam dados sensíveis, permitindo que quem venha visualizar esses dados posteriormente identifique que são sensíveis (Figura 40).



Figura 40 - Pentaho Kettle rotulação dos dados

The screenshot shows the Pentaho Kettle Spoon interface. The main workspace displays a job design with three components: 'mot\_hab', 'Rotula', and 'Dummy (do nothing)'. The 'Rotula' component is highlighted, indicating it is the active step. Below the workspace, the 'Execution Results' window is open, showing a table of data. The table has the following columns: 'al', 'cnpj', 'nome\_motorista', 'cpf', 'vigencia\_do\_cadastro', and 'rotulo'. The 'rotulo' column contains the value 'dado\_sensivel' for all rows.

al	cnpj	nome_motorista	cpf	vigencia_do_cadastro	rotulo
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	MURIL ***** REIRA	302.***.***-88	2025/11/03 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	NILTO ***** MENES	093.***.***-29	2024/06/25 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	PAULO ***** ANELI	015.***.***-32	2025/08/12 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	REGIN ***** TOME	095.***.***-42	2025/07/16 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	REGIN ***** AVIER	103.***.***-80	2024/10/17 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	RICAR ***** RREIA	104.***.***-54	2023/09/04 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	RODRI ***** ERINE	317.***.***-80	2023/10/02 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	VAGNE ***** SILVA	191.***.***-10	2024/10/14 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	VALDI ***** IGUES	266.***.***-00	2023/12/18 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	VANDO ***** EOBET	995.***.***-20	2023/01/03 00:00:00.000000000	dado_sensivel
NEVES TRANSPORTES LTDA - ME	07.574.686/0001-73	VINIC ***** ASTRO	371.***.***-63	2024/10/15 00:00:00.000000000	dado_sensivel
Σ XIMENES TRANSPORTES LTDA	10.320.526/0001-02	FRANC ***** ALVES	113.***.***-87	2024/08/01 00:00:00.000000000	dado_sensivel
Σ XIMENES TRANSPORTES LTDA	10.320.526/0001-02	JOSE ***** LIRA	966.***.***-00	2024/08/01 00:00:00.000000000	dado_sensivel
TRANSPORTE E LOCAÇÃO DE VEICULOS LTDA	60.341.922/0001-94	ELTON ***** RTINS	289.***.***-12	2023/03/03 00:00:00.000000000	dado_sensivel
TRANSPORTE E LOCAÇÃO DE VEICULOS LTDA	60.341.922/0001-94	JAIR ***** SSOLI	716.***.***-15	2022/10/03 00:00:00.000000000	dado_sensivel
TRANSPORTE E LOCAÇÃO DE VEICULOS LTDA	60.341.922/0001-94	LUIS ***** BALHO	086.***.***-59	2022/02/06 00:00:00.000000000	dado_sensivel

Fonte: Próprio autor.

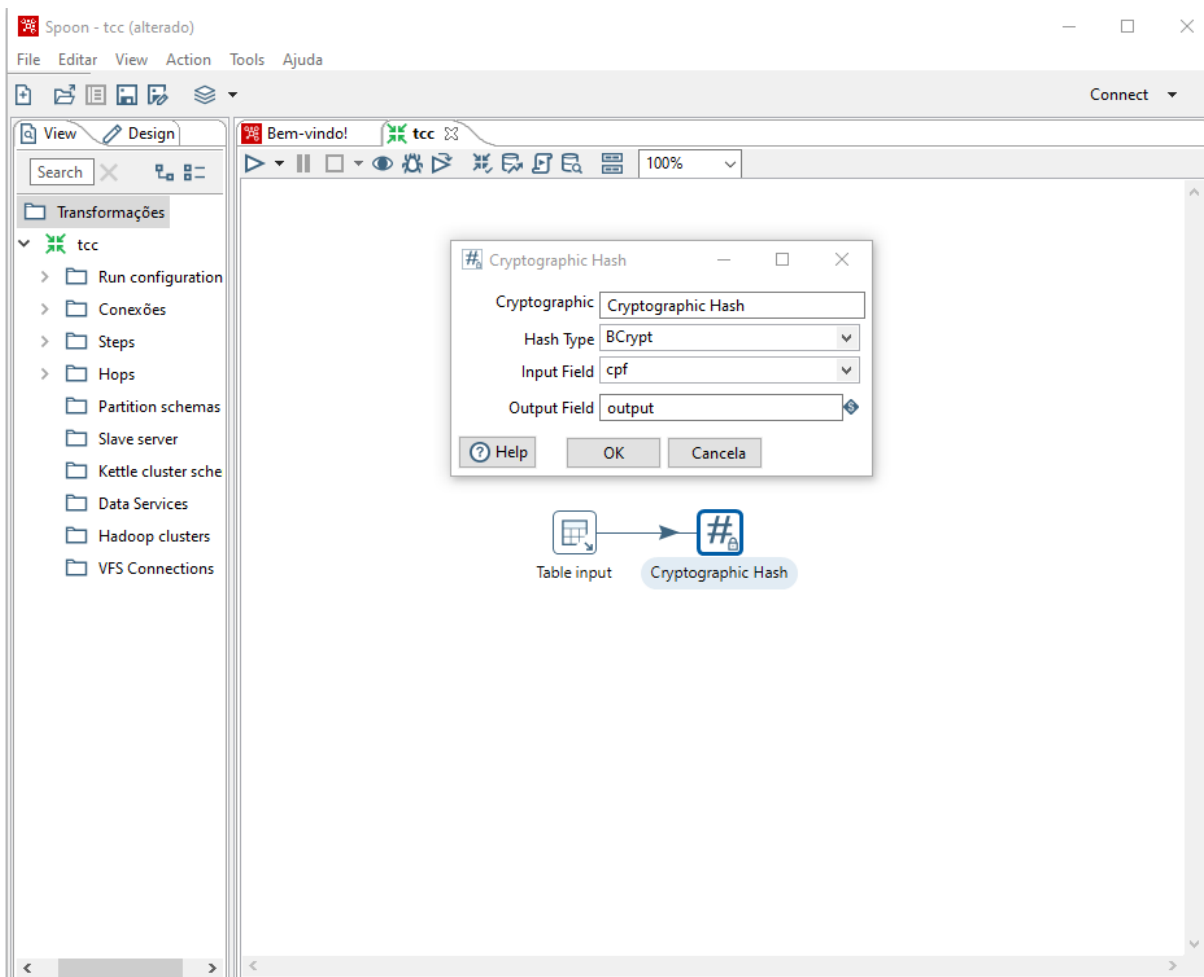
O resultado esperado é atingido pela ferramenta, a mesma permite que seja inserido rótulos nos locais que armazenam dados sensíveis. Assim facilitando que usuários que forem visualizar os dados posteriormente consigam identificar os dados através dos rótulos.

### 5.2.6 Caso de Teste 6: Criptografia dos dados

O caso de teste 6 (Quadro 16) consiste em testar se a ferramenta é capaz de criptografar os dados pessoais. O Pentaho Kettle possui um componente nativo capaz de realizar a criptografia dos dados. Através do componente *Cryptographic Hash* é possível aplicar os tipos de criptografia como BCrypt ou SHA-3. Pode ser selecionado somente uma coluna de dados por cada componente, caso necessário mais de uma coluna de dados é necessário adicionar mais componentes (Figura 41). Em *Output Field* é possível escolher em qual coluna devem ser escritos os dados

criptografados, caso selecionado a mesma coluna de origem esses dados serão sobrescritos.

Figura 41 - Pentaho Kettle criptografia de dados



Fonte: Próprio autor.

Com a aplicação da criptografia foi possível obter o resultado esperado, e todos os dados referentes a coluna “cpf” que foram selecionados para passarem pelo processo foram criptografados conforme a Figura 42. Porém, não foi possível realizar a reversão da criptografia através do Pentaho Kettle.

Figura 42 - Pentaho Kettle resultado criptografia de dados

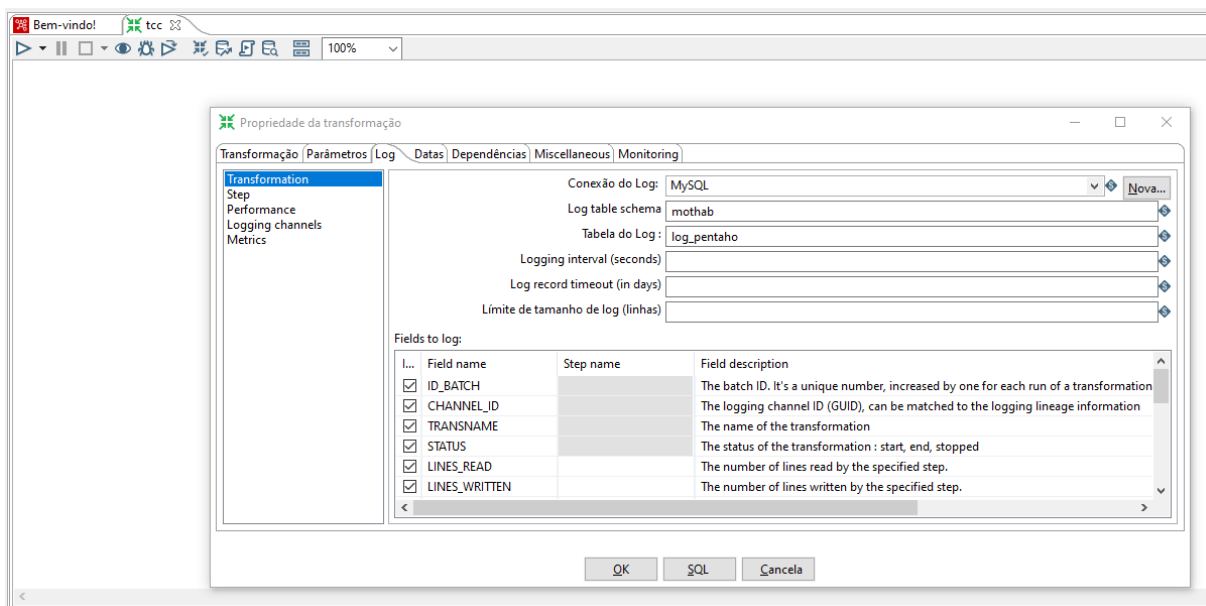
The screenshot shows the Pentaho Kettle Spoon interface. The main window displays a data transformation job named 'tcc'. The job is composed of two steps: 'Table input' and 'Cryptographic Hash'. The 'Execution Results' pane is open, showing a table with the following data:

cnpj	nome_motorista	cpf
73.500.696/0001-99	ALDOM ***** MENTO	\$2a\$12\$G3r6l.5yH.dcDW/L8A/rfO2KqEX1ikrzXlbACRYgNiFdxefuAKTm.
73.500.696/0001-99	ALESS ***** HORST	\$2a\$12\$XohLqK87qmwLZezVPjsehOOFarf15jbCVtMm3dFH.FxiW1ZBq292
73.500.696/0001-99	ALOIS ***** LMANN	\$2a\$12\$6e.TYMFsJuTuIvtQTFQ2CetnPemZhdCBNGo5BfhkwMwYpYcwx2GBO
73.500.696/0001-99	ANTON ***** DEIRA	\$2a\$12\$H/jNy7uYGHu6dFsCzP8ax.e62MZV.S197hcdS8Mz.urzVdglqcadm
73.500.696/0001-99	CEZAR ***** OSTER	\$2a\$12\$2qb5ebiwGaeFmPGttSwbhuaL4EmqCrm2wEf2FWObEc3f5SWfJEa
73.500.696/0001-99	GELSO ***** SIDRA	\$2a\$12\$1N4N6kqQ./GtKljbLyUBOINJ3XDaqj2KJly53uwu9NDhdc1kst.
73.500.696/0001-99	HARRY ***** MOBBS	\$2a\$12\$DCMS43tM/4flzBEwzlou9LTit6ypF2.slr/WaRQCpJfc56i8Qg2
73.500.696/0001-99	IRANI ***** VEIRA	\$2a\$12\$4iRIEDpgJ/8V3rtH1YU7OnfBCjjsXhQMJaPFcwzaZINE.uqibvGC
73.500.696/0001-99	JOAO ***** VEIRA	\$2a\$12\$PdX1TDCYbGjBvHS9J84zeRml3aTKlxMbczw1GsMvRy8UdVo3NOm6
73.500.696/0001-99	JUCEM ***** AJOLO	\$2a\$12\$vcukSHZmWr.k2sVT/yrAKeU2BWVgleTphJNo2mExoDIBHUCnh615K
73.500.696/0001-99	LAIRO ***** CHEER	\$2a\$12\$TdGDSQKcTjZ93Po1B.T0jerWUKSHJkw2eISIP2Dk33CCERPHOZa
73.500.696/0001-99	LUCIA ***** NUNES	\$2a\$12\$YaLpkDjR2wUEjxyKucoOV.I/YGISXNUGrM8dTFxR0bY8UwmaFtt3q
73.500.696/0001-99	LUIZ ***** ALVES	\$2a\$12\$siC46Zx.KVBw8BQwkrivLe9jcU.MxV0iilUb6levfG8lx.vmvVK.
73.500.696/0001-99	LUIZ ***** IEIRA	\$2a\$12\$4uQDKcFTSlgT.DIEJBeT.WwW6QKPrNgkwn8UFVWWkwGKX1LOVqPUm
73.500.696/0001-99	LUIZ ***** AUSEN	\$2a\$12\$BkULqVJWZo/nqYNDazVEe7/b1YYBff7oHoWpE3dfll8VmqBaUmj.
73.500.696/0001-99	MARCE ***** ANTOS	\$2a\$12\$ZAIslM95i7zi7WJ.T.eptvz4dl1h4z94217ewK6o6/KA0CLzi
73.500.696/0001-99	NELSO ***** NETTI	\$2a\$12\$fb9rxn3V3.Uzys.5hH.Nj.Bsf2tC4pxP5d2JeaZgLRyY9PX/n9Jdu

Fonte: Próprio autor.

### 5.2.7 Caso de Teste 7: Armazenamento de eventos

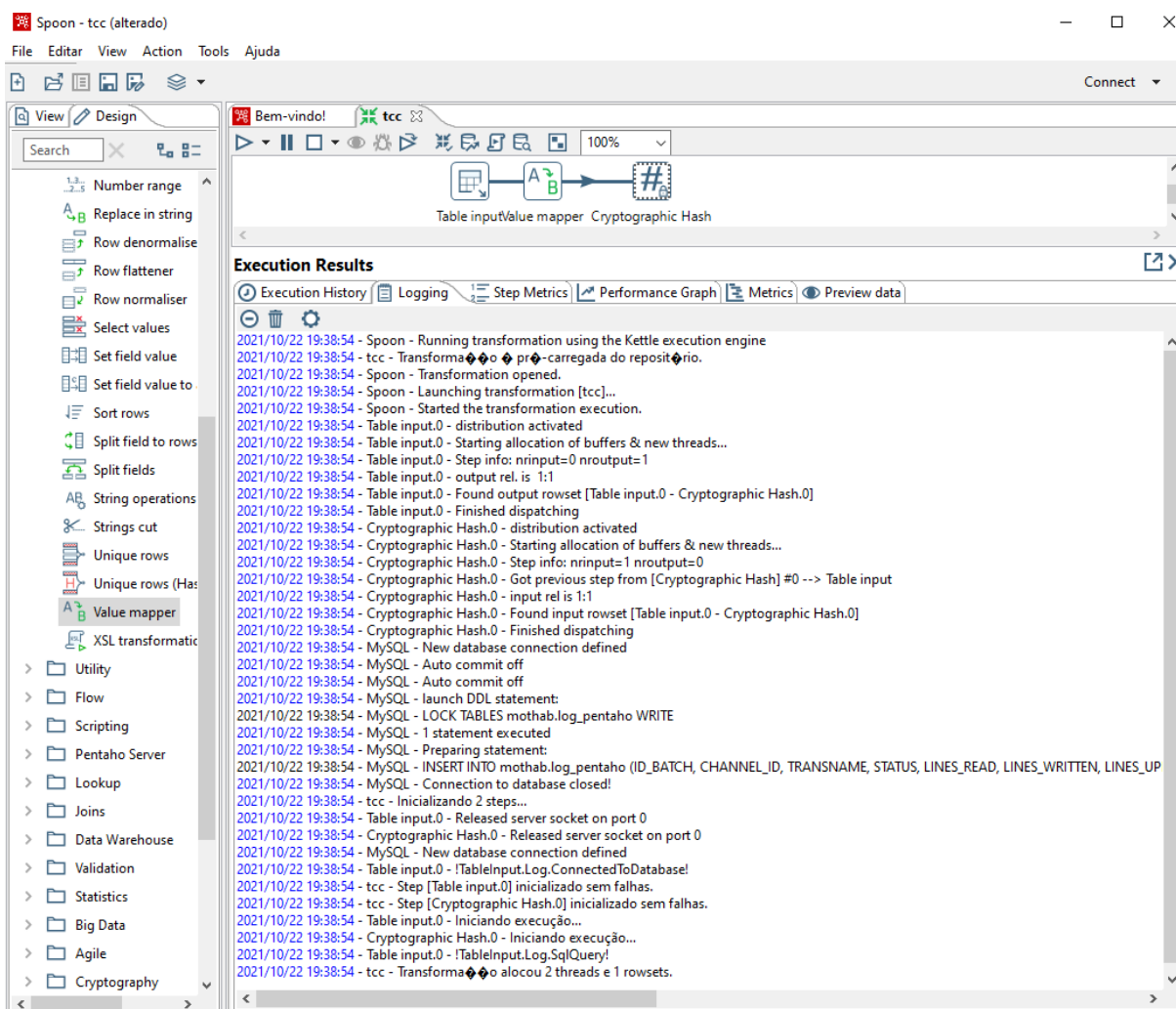
O caso de teste 7 (Quadro 17) compreende em observar se a ferramenta é capaz de armazenar os eventos (*logs*) ocorridos. A configuração de log do Pentaho exige que seja configurado uma conexão com algum banco de dados para que ele descarregue os eventos. Conforme a Figura 43 a tela de configuração possibilita escolher o banco de dados e a tabela para o armazenamento dos dados. Para o teste foi criado uma tabela chamada `log_pentaho`, ao clicar no botão SQL o próprio Pentaho realiza a criação das colunas necessárias para armazenar os *logs*.

Figura 43 - Pentaho Kettle configurar armazenamento *log*

Fonte: Próprio autor.

Quando carregado e executado uma tarefa todos os logs que o Pentaho e o usuário realizam são apresentados na aba *Logging*, e também são salvos no banco de dados selecionados para consulta posterior. A Figura 44 demonstra os eventos coletados durante a execução de um processo de criptografia, é possível verificar as ações do usuário como quando a transformação foi aberta, e também todos os eventos que o próprio Pentaho realizou.

Figura 44 - Pentaho Kettle visualizando eventos



Fonte: Próprio autor.

O resultado esperado é atingido pela ferramenta, a mesma coletou os eventos realizados pelo usuário e pela ferramenta. Os eventos ficam armazenados por tempo indeterminado e podem ser consultados a qualquer momento.

### 5.2.8 Caso de Teste 8: Resposta a incidentes de segurança

O caso de teste 8 (Quadro 18) corresponde a como a ferramenta identifica as consequências e o período de tempo do ocorrido de algum incidente. Por meio do sistema de gerenciamento de banco de dados (SGBD) foi excluído aleatoriamente

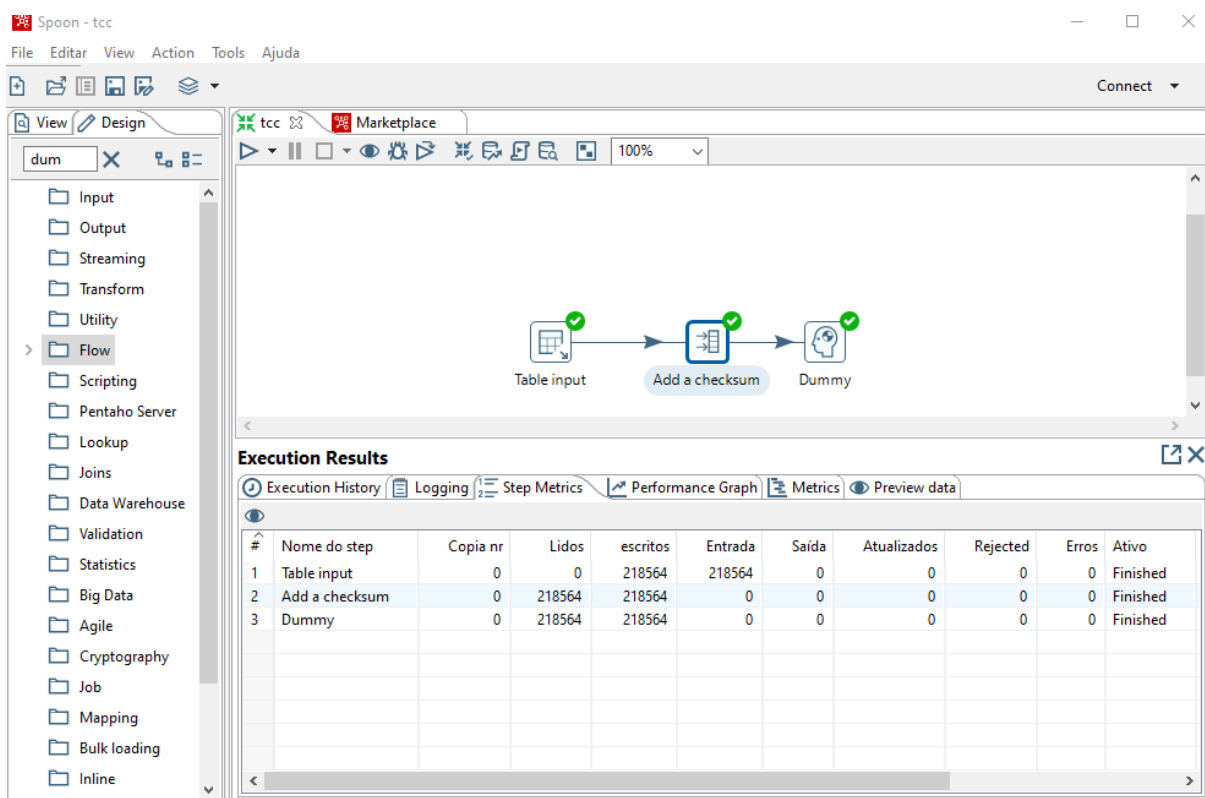
20.000 dados sensíveis das colunas “nome\_motorista e cpf” para o teste, mas a ferramenta Pentaho Kettle não conseguiu identificar que esses dados foram excluídos. Retornando a fonte de dados para o estado original, foram realizadas a alteração de 8 mil dados para “XXXX” de uma única vez, simulando um volume incomum da alteração, mas não foram detectadas pelo Pentaho Kettle que apenas mostra os dados alterados.

A ferramenta não pode atingir o resultado esperado, ela não detectou a ocorrência de incidentes nos dados armazenados pelas fontes de dados. Também não permite detectar o período de tempo da ocorrência de segurança.

### **5.2.9 Caso de Teste 9: Análise crítica técnica do compliance**

O caso de teste 9 (Quadro 19) menciona se que a ferramenta pode identificar em qual parte do ciclo do tratamento dos dados está sendo realizada. O Pentaho Kettle não pode designar em que fase do tratamento de dados está ocorrendo, como seu objetivo é a integração de dados, o mesmo não possui componentes que são voltados para a conformidade. Pode-se identificar somente os processos de tratamento de dados que são realizados pela própria ferramenta. Pode-se identificar somente os processos de tratamento de dados que foram realizados pela própria ferramenta, como exemplo, o processo de adição de checksum, através do Pentaho, o qual verificou cada etapa realizada pela ferramenta conforme a Figura 45.

Figura 45 - Pentaho Data Integration tratamento dos dados



Fonte: Próprio autor.

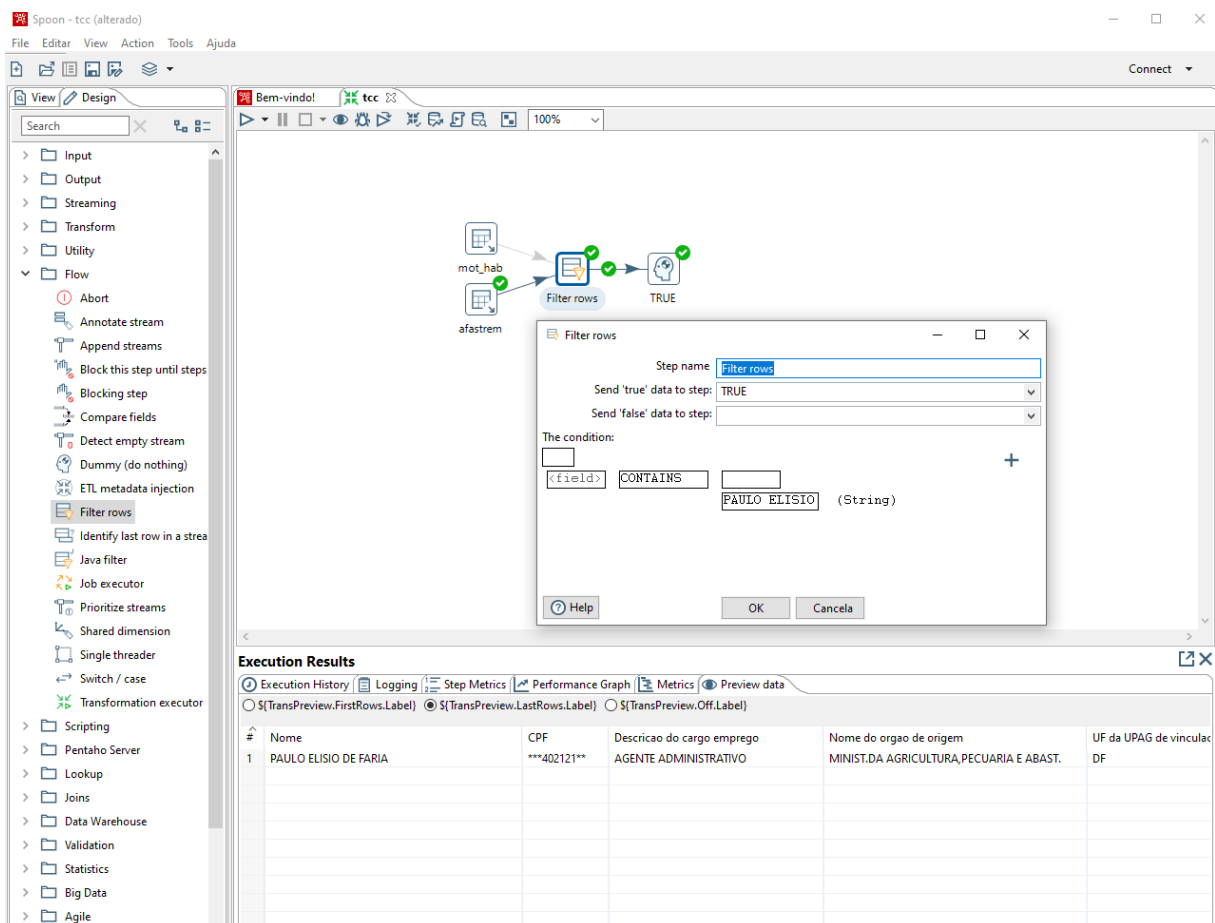
Portanto, ainda que a ferramenta demonstre o ciclo de dados no qual ela fez a essa transformação, a mesma não pode atingir o resultado esperado que era a identificação do ciclo de tratamento dos dados referente a todas operações realizadas.

### 5.2.10 Caso de Teste 10: Entrega de dados ao titulares

No caso de teste 10 (Quadro 20), é especificado se a ferramenta é capaz de identificar os dados pessoais de determinado titular quando solicitado. O Pentaho Kettle possibilita filtrar os dados de diversas fontes de dados. Para isso configurou-se um componente de filtro que realizou a pesquisa de todos os dados que continham o valor PAULO ELISIO DE FARIA, nome de um titular dos dados escolhido aleatoriamente, em ambos os conjuntos foi possível visualizar todos os

dados desse usuário conforme a Figura 46. Entretanto, mesmo o Pentaho possibilitando o encontro dos dados do titular, ele não pode determinar de qual das fontes conectadas esse dado pertencia.

Figura 46 - Pentaho Kettle aplicar filtro aos dados



The screenshot shows the Pentaho Kettle interface. On the left is a tool palette with various transformation steps. The main workspace displays a data flow diagram with three steps: 'mot\_hab', 'afastrem', and 'Filter rows'. The 'Filter rows' step is highlighted, and its configuration dialog is open. The dialog shows the step name 'Filter rows', 'Send 'true' data to step: TRUE', and 'Send 'false' data to step:'. The condition is set to '<field> CONTAINS PAULO ELISIO (String)'. Below the dialog, the 'Execution Results' section is visible, showing a table with one row of data.

#	Nome	CPF	Descricao do cargo emprego	Nome do orgao de origem	UF da UPAG de vinculac
1	PAULO ELISIO DE FARIA	***402121**	AGENTE ADMINISTRATIVO	MINIST.DA AGRICULTURA,PECUARIA E ABAST.	DF

Fonte: Próprio autor.

Mesmo assim o resultado atingido é o esperado, a ferramenta foi capaz de apresentar todos os dados de determinado titular.

### 5.2.11 Caso de Teste 11: Anonimização

O caso do teste 11 (Tabela 21) compreende-se em testar se a ferramenta pode anonimizar dados pessoais sem possibilitar a reversão. O Pentaho Kettle não



possui nenhum componente que realize a anonimização dos dados e a sua loja de plugins não possui nenhum *plugin* que possibilite o mascaramento dos dados.

Por isso o resultado esperado não pode ser alcançado pela ferramenta já que ela não possui um componente de anonimização desses dados.

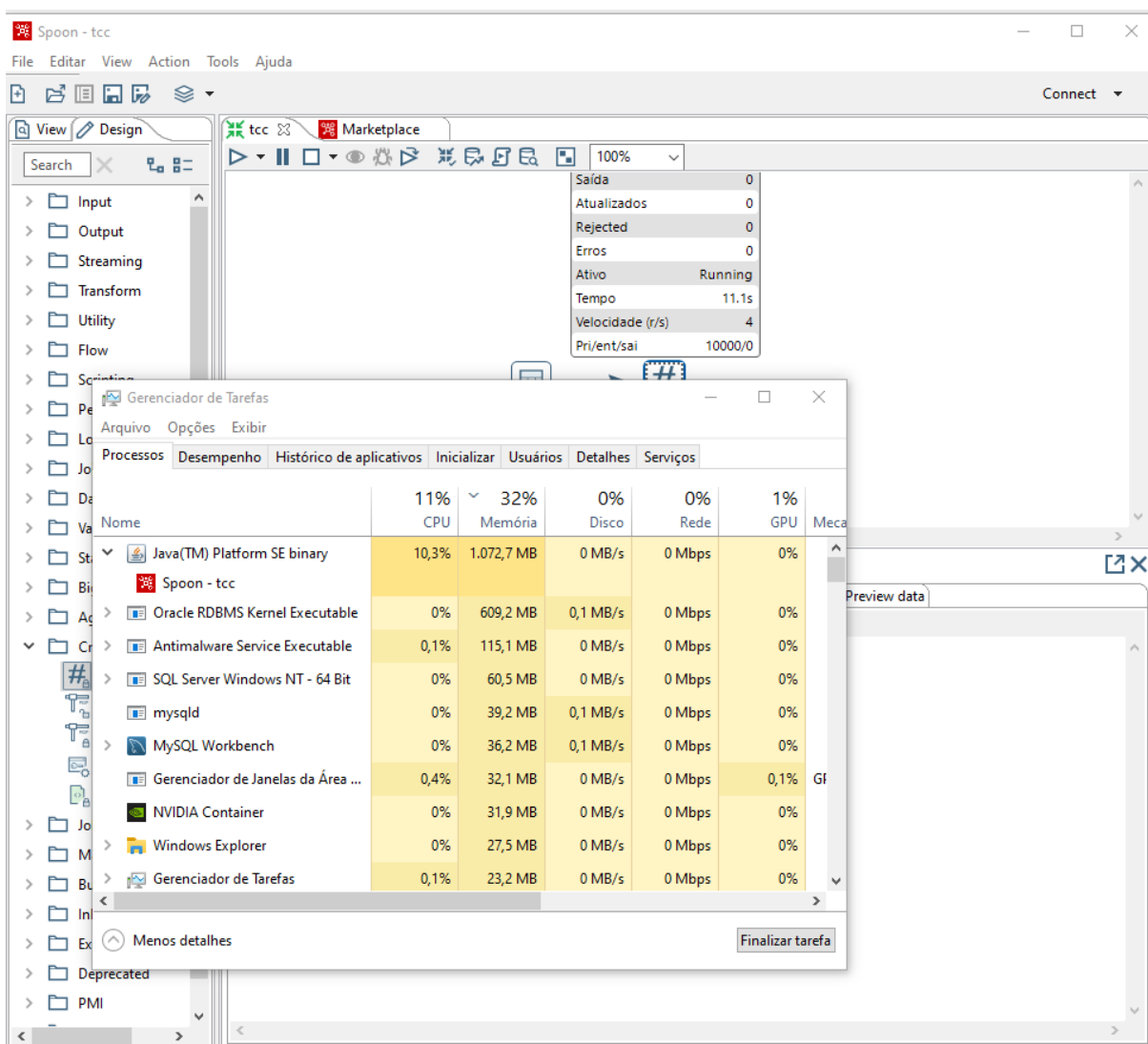
#### **5.2.12 Caso de Teste 12: Coexistência**

O caso de teste 12 (Quadro 22) consiste em testar se a ferramenta executa sem apresentar falhas ao operar em simultâneo com as ferramentas de banco de dados no mesmo ambiente. O teste é feito em um sistema operacional Windows 10, utilizando um processador AMD Ryzen 5 2600 e 16GB de memória RAM, para a execução do teste apenas é executado tarefas próprias do sistema operacional e os bancos de dados SQL Express, Oracle XE e MySQL em conjunto com Pentaho.

A ferramenta apresentou estabilidade e não apresentou nenhum erro ou travamento durante a execução dos demais casos de teste. Durante a conexão e operação dos dados, também não exibiu nenhum erro ou travamento, funcionando normalmente com todas as fontes de dados conectadas.

Após monitorado o uso de recursos do computador na execução do caso de teste 6 (Seção 5.2.6) , o processo do Pentaho Kettle que tem o nome de Spoon apresentou uso de 1,1GB de memória RAM conforme Figura 47. Nenhum travamento aconteceu pelo Pentaho Kettle ou por algum dos bancos de dados.

Figura 47 - Pentaho Kettle uso de recursos



Fonte: Próprio autor.

O resultado esperado foi atingido, a ferramenta executada com demais programas não apresentou erros e travamentos. Durante a execução dos demais casos de teste a ferramenta também não apresentou travamentos e erros.

### 5.3 CLOVERDX

O CloverDX é uma ferramenta com licenciamento pago mantido pela empresa Clover desde 2007. Seu principal objetivo é os processos de ETL e de gerenciamento de dados.

A ferramenta possui uma interface amigável e disponibiliza componentes que arrastados montam um fluxo de trabalho para a transformação dos dados. Pode ser executada diretamente em ambientes em nuvem, como AWS e Azure, e permite a integração e conexão com diversas fontes de dados externas.

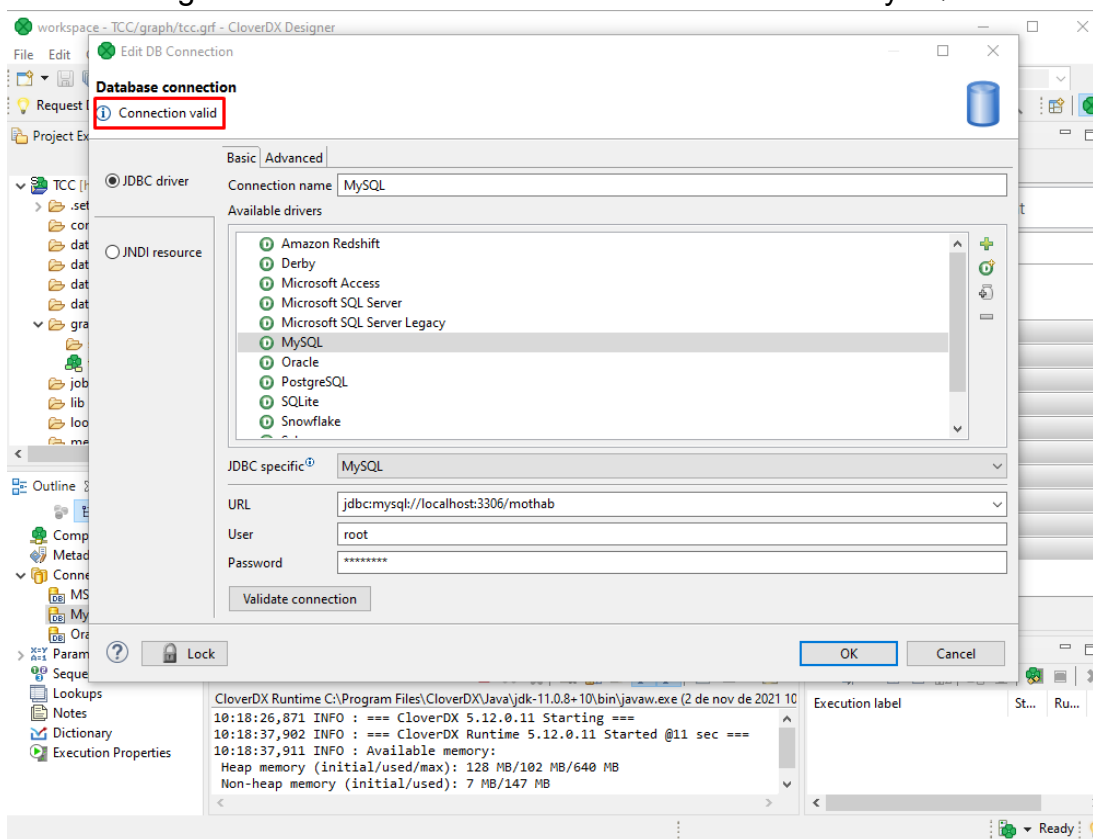
A instalação não necessita de nenhum pré-requisito, e é feita por um único pacote de instalação. Com o processo de instalação simples, é instalado o CloverDX Server que é onde ficam armazenadas as configurações, projetos, gerenciamento de usuários e de eventos e também o CloverDX Designer que é o aplicativo usado para manipular e transformar os dados.

### **5.3.1 Caso de Teste 1: Parametrização da ferramenta**

O caso de teste 1 inclui parametrizar a ferramenta e se conectar às fontes de dados. É realizada a parametrização para se conectar às fontes de dados MySQL, Oracle XE e SQL Server definidas anteriormente na seção 4.5.

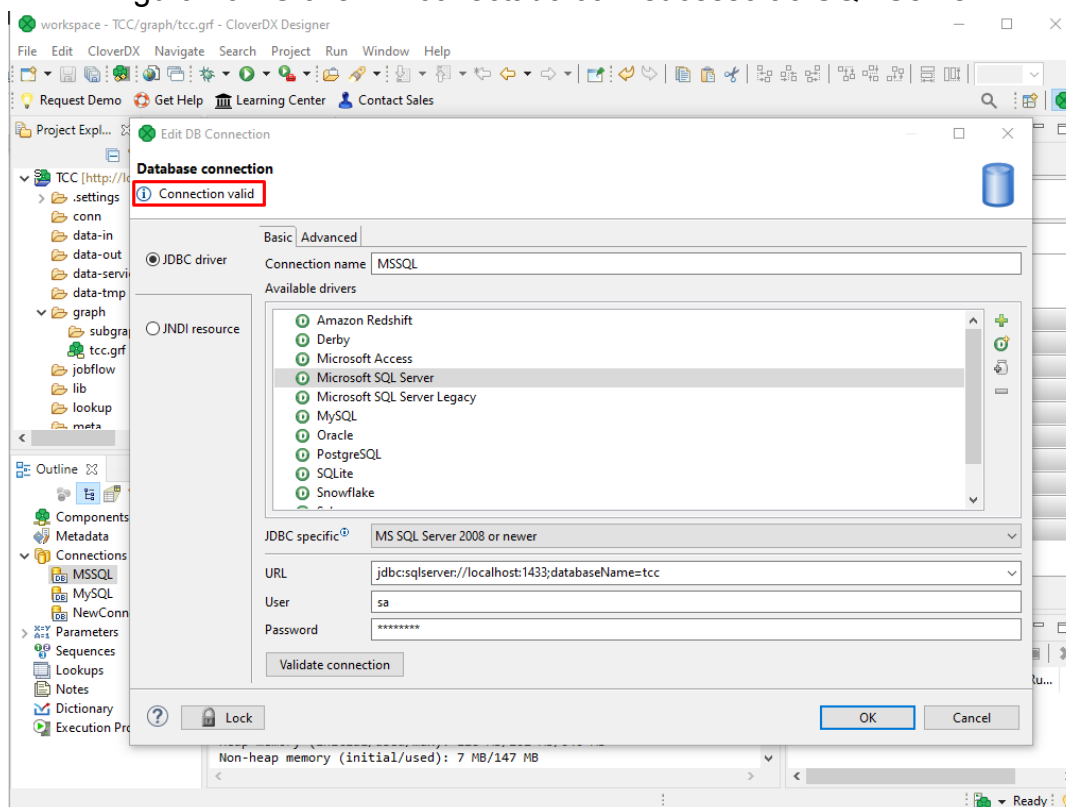
Na conexão com o banco de dados MySQL, Oracle e SQL Server a ferramenta não apresentou erros. A ferramenta disponibiliza diversos *drivers* para realizar a comunicação entre a ferramenta e as fontes de dados. Foi necessário preencher os parâmetros de endereço, *login* e senha do banco de dados, desta forma a conexão foi realizada com sucesso a todos os bancos de dados conectados conforme Figuras 48, 49 e 50.

Figura 48 - CloverDX conectado com sucesso ao MySQL



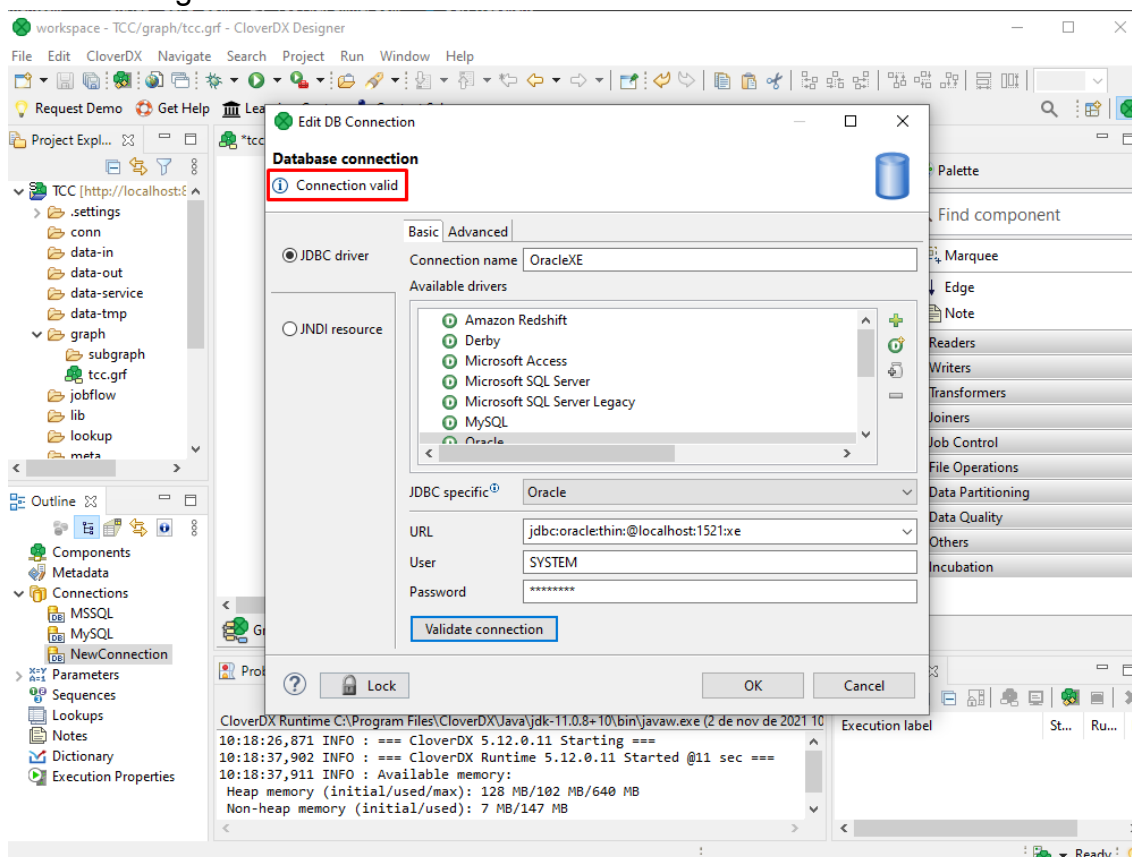
Fonte: Próprio autor.

Figura 49 - CloverDX conectado com sucesso ao SQL Server



Fonte: Próprio autor.

Figura 50 - CloverDX conectado com sucesso ao Oracle XE



Fonte: Próprio autor.

O resultado esperado foi atingido, a ferramenta apresenta facilidade para se conectar às fontes de dados, sendo necessário somente apontar as informações solicitadas para a conexão.

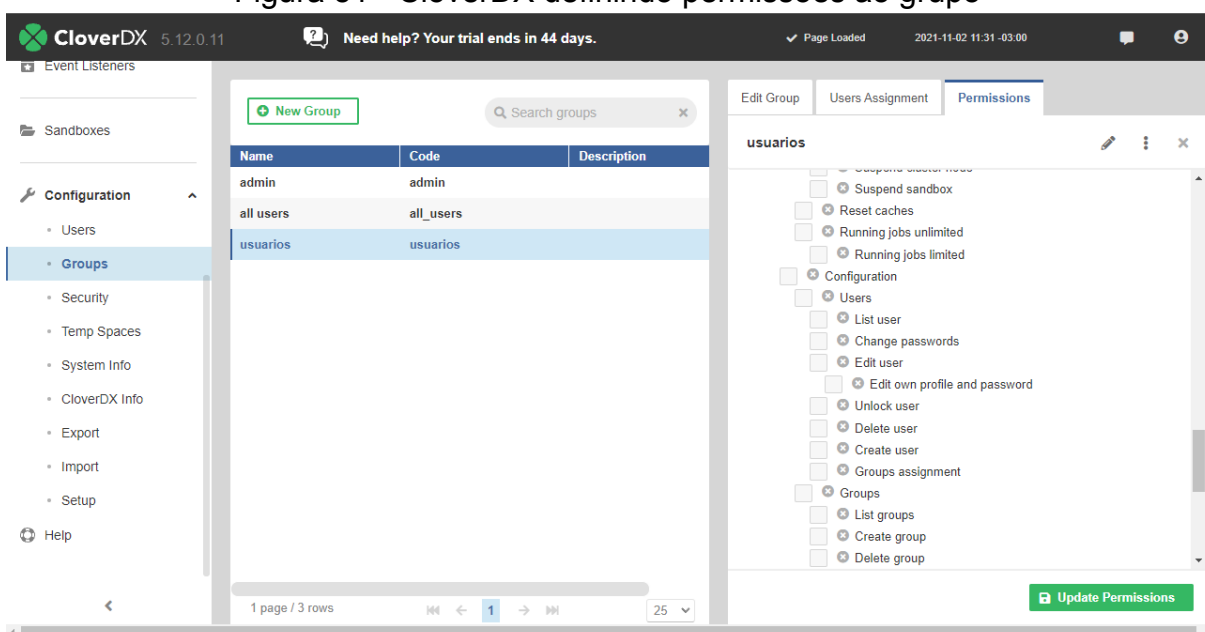
### 5.3.2 Caso de Teste 2: Avaliação de riscos dos dados

No caso de teste 2 (Quadro 12), corresponde testar se a ferramenta é capaz de detectar os riscos existentes e medir a probabilidade de os riscos acontecerem, além de gerar um relatório desses riscos. O CloverDX não pode produzir os resultados esperados, não pode avaliar a probabilidade de risco, muito menos gerar um relatório de avaliação de risco. O resultado esperado que era obter um relatório ou visualizar os riscos encontrados pela ferramenta não foi alcançado.

### 5.3.3 Caso de Teste 3: Segurança de Acesso

No caso de teste 3 (Quadro 13) é verificado se a ferramenta possui mecanismos que gerenciem o acesso de usuários, assim garantindo maior segurança para o uso da ferramenta. O CloverDX possui um gerenciador de usuários, onde é possível criar grupos e definir permissões. Através do CloverDX Server é criado um grupo chamado usuário e definindo as permissões para que não possa ser alterado o menu de configurações (Figura 51).

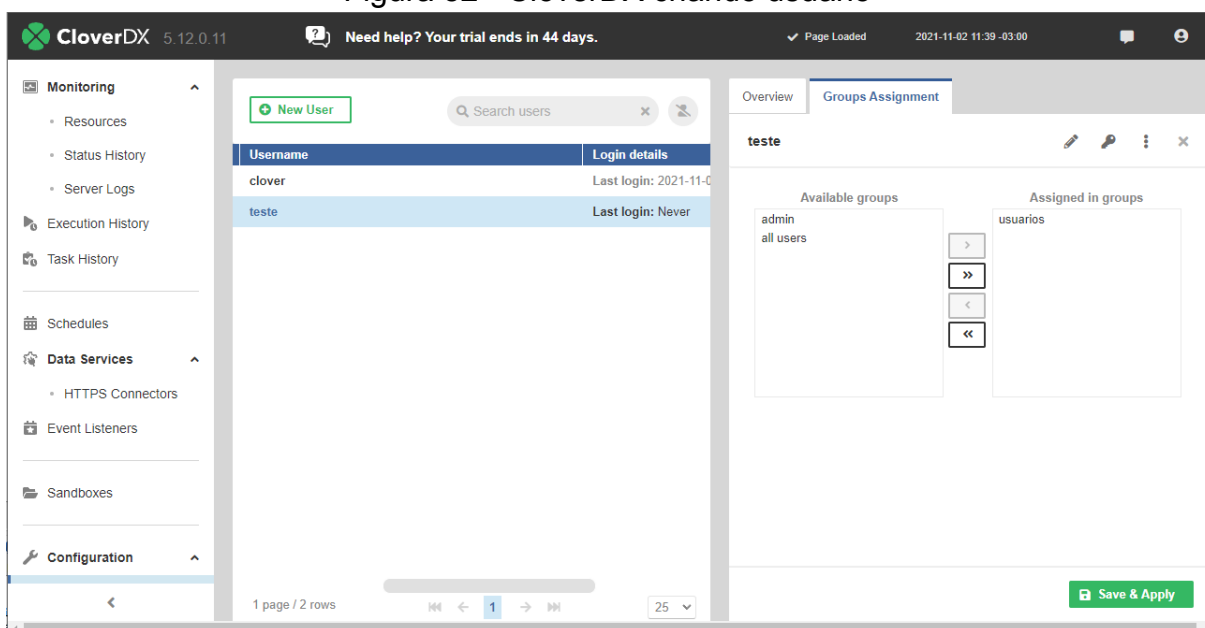
Figura 51 - CloverDX definindo permissões ao grupo



Fonte: Próprio autor.

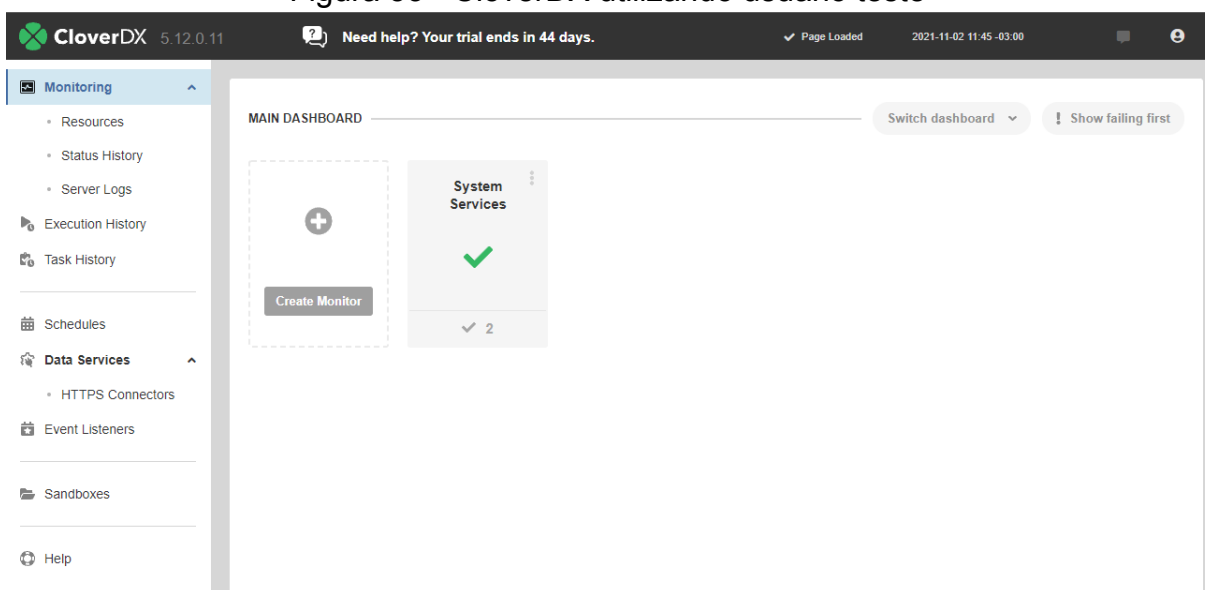
Foi criado um usuário teste e definido para o grupo criado anteriormente chamado usuários (Figura 52). Ao acessar o CloverDX o mesmo não aparece no menu de configurações que o usuário não possui acesso, e apresenta somente os itens em que ele tem permissão (Figura 53).

Figura 52 - CloverDX criando usuário



Fonte: Próprio autor.

Figura 53 - CloverDX utilizando usuario teste



Fonte: Próprio autor.

O resultado esperado é atingido com sucesso, a ferramenta possibilita o gerenciamento de usuário, e permite definir permissões de acesso para cada grupo.

#### **5.3.4 Caso de Teste 4: Identificação de dados sensíveis**

O resumo do caso de teste 4 é se a ferramenta possibilita realizar a identificação dos dados sensíveis. O CloverDX não realiza a identificação dos dados. Não é possível atingir o resultado esperado devido a falta de recurso pela ferramenta.

#### **5.3.5 Caso de Teste 5: Rotulação dos dados**

O caso de teste 5 resume-se em testar se a ferramenta possibilita correlacionar rótulos aos seus dados pessoais. O CloverDX não possibilita a inserção de rótulos, não possui nenhum componente que possibilite a descrição dos dados tratados. Deste modo o resultado esperado não pode ser atingido, e não se pode identificar os dados sensíveis apenas pela sua visualização.

#### **5.3.6 Caso de Teste 6: Criptografia dos dados**

O caso de teste 6 (Quadro 16) consiste em testar se a ferramenta é capaz de criptografar os dados pessoais. O CloverDX não possibilita a criptografia dos dados, pois não possui nenhum componente que possibilite aplicar os métodos de criptografia e descryptografia. O resultado esperado não pode ser atingido, já que a ferramenta não possibilita a realização do caso de teste.

#### **5.3.7 Caso de Teste 7: Armazenamento de eventos**

O caso de teste 7 (Quadro 17) compreende em observar se a ferramenta é capaz de armazenar os eventos(*logs*) ocorridos. O CloverDX armazena todos os eventos ocorridos de forma automática, a cada 1000 eventos coletados os eventos mais antigos começam a ser apagados. Para exemplificar é executado um processo de leitura dos dados no banco de dados MySQL, ao verificar a página da *logs* é



possível visualizar os eventos que ocorreram, e qual usuário realizou as tarefas (Figura 54).

Figura 54 - CloverDX armazenamento de eventos

Timestamp	Log Type	Message
2021-11-02 16:48:51,476[nio-8083-exec-10]	userAction	INFO logout; session=bn9KZlAXUjnyU1Q50U1SUI64fyNwUjAbMDownUjA6MDownUjEbNjE4MIKL
2021-11-02 16:48:54,507[nio-8083-exec-4]	userAction	INFO Login successful; username=clover, address=0:0:0:0:0:1
2021-11-02 16:48:54,509[nio-8083-exec-4]	LoginController	INFO Login successful, username=clover
2021-11-02 16:48:54,510[nio-8083-exec-4]	LoginController	INFO Login successful, user= User#17:clover
2021-11-02 16:56:52,691[nio-8083-exec-5]	userAction	INFO execute job; user=User#17:clover jobExecutionCommand=GraphExecutionComma
2021-11-02 16:56:52,802[ JobStarter_295]	a	INFO 295 Starting job: TCC/graph/tcc.grf
Run ID: 295		
User: clover		
Triggered by: SOAP WS API		
Executing on: [worker@node01]		
2021-11-02 16:56:52,865[ threadPool-8]	au	ERROR [worker@node01]: Tue Nov 02 16:56:52 BRT 2021 WARN: Establishing SSL co
2021-11-02 16:56:53,555[obfinalizer_295]	h	INFO 295 Job finished: TCC/graph/tcc.grf
Run ID: 295		
Status: FINISHED_OK		
Duration: 864 ms		
2021-11-02 16:56:57,970[nio-8083-exec-3]	userAction	INFO get edge debug data; user=User#17:clover options=[sandbox:TCC_graph:grap
2021-11-02 16:58:43,551[nio-8083-exec-2]	userAction	INFO execute job; user=User#17:clover jobExecutionCommand=GraphExecutionComma
2021-11-02 16:58:43,670[ JobStarter_296]	a	INFO 296 Starting job: TCC/graph/tcc.grf
Run ID: 296		
User: clover		
Triggered by: SOAP WS API		
Executing on: [worker@node01]		
2021-11-02 16:58:43,705[ threadPool-8]	au	ERROR [worker@node01]: Tue Nov 02 16:58:43 BRT 2021 WARN: Establishing SSL co
2021-11-02 16:58:44,594[obfinalizer_296]	h	INFO 296 Job finished: TCC/graph/tcc.grf
Run ID: 296		
Status: FINISHED_OK		
Duration: 1 sec		
2021-11-02 16:58:48,834[nio-8083-exec-10]	userAction	INFO get edge debug data; user=User#17:clover options=[sandbox:TCC_graph:grap

Fonte: Próprio autor.

O resultado esperado é atingido pela ferramenta, a mesma coleta os eventos realizados pelo usuário e pela ferramenta. A ferramenta permite armazenar até 10 mil eventos no máximo, após isso os mais antigos são sobrescritos. É possível que os eventos sejam exportados para um arquivo externo e salvos no disco do computador.

### 5.3.8 Caso de Teste 8: Resposta a incidentes de segurança

O caso de teste 8 (Quadro 18) corresponde a como a ferramenta identifica as consequências e o período de tempo do ocorrido de algum incidente. Por meio do sistema de gerenciamento de banco de dados (SGBD) é excluído aleatoriamente 20.000 dados sensíveis das colunas nome\_motorista e cpf para o caso teste, mas a ferramenta CloverDX não consegue identificar que esses dados foram excluídos. Retornando a fonte de dados para o estado original, foram realizadas a alteração de

8 mil dados para “XXXX” de uma única vez, simulando um volume incomum da alteração, mas não foram detectadas por meio do CloverDX.

A ferramenta não pode atingir o resultado esperado, ela não detectou a ocorrência de incidentes nos dados armazenados pelas fontes de dados. Também não permitiu detectar o período de tempo da ocorrência de segurança.

### **5.3.9 Caso de Teste 9: Análise crítica técnica do compliance**

O caso de teste 9 (Quadro 19) menciona se que a ferramenta pode identificar qual parte do ciclo do tratamento dos dados está ocorrendo. O CloverDX não pode designar em que fase do tratamento de dados está ocorrendo, como seu objetivo é a integração de dados não possui componentes que são voltados para a conformidade. Pode-se identificar somente os processos de tratamento de dados que são realizados pela própria ferramenta. Pode-se citar como exemplo quando realizado a tarefa de gravar os dados tratados em um arquivo de texto externo, através do Clover DX pode-se verificar cada etapa realizada pela ferramenta conforme a Figura 55.

Figura 55 - CloverDX tratamento dos dados

The screenshot displays the CloverDX Designer interface. The main workspace shows a data flow graph with three components: DatabaseReader, UniversalDataWriter, and Success. The UniversalDataWriter component is selected, and its output is displayed in a table below. The table has three columns, all labeled 'dado\_sensivel'. The data rows are as follows:

#	dado_sensivel	dado_sensivel	dado_sensivel
1	011 TRANSPORTES E TURISMO LTDA	12.463.028/0001-44	FERNA ***** FILHC
2	011 TRANSPORTES E TURISMO LTDA	12.463.028/0001-44	RAFAE ***** OARE:
3	1A - GRANTURISMO TRANSPORTE LTDA ME	03.865.605/0001-33	ALEX ***** ANTOS
4	1A - GRANTURISMO TRANSPORTE LTDA ME	03.865.605/0001-33	CARLO ***** VEIRA
5	1A - GRANTURISMO TRANSPORTE LTDA ME	03.865.605/0001-33	HUMBE ***** TEIRA
6	1A - GRANTURISMO TRANSPORTE LTDA ME	03.865.605/0001-33	ITALO ***** ANDIC
7	1A - GRANTURISMO TRANSPORTE LTDA ME	03.865.605/0001-33	SUZAN ***** ARIA:
8	2 MARCOS TURISMO LTDA	11.151.062/0001-10	JOAO ***** VALHC
9	2 MARCOS TURISMO LTDA	11.151.062/0001-10	MARCO ***** UNI.

The interface also shows a Palette on the right with components like Marquee, Edge, Note, Readers, Writers, Transformers, and Joins. The Execution tab at the bottom right shows the execution status for 'tcc.grf' with a status of 'Ready' and a record count of 312.

Fonte: Próprio autor.

Mesmo que a ferramenta demonstre o ciclo de dados em que ela fez a transformação dos dados, a ferramenta não pode atingir o resultado esperado que é a identificação do ciclo de tratamento dos dados referente a todas operações realizadas nos dados .

### 5.2.10 Caso de Teste 10: Entrega de dados ao titulares

No caso de teste 10 (Quadro 20), é especificado se a ferramenta é capaz de identificar os dados pessoais de determinado titular quando solicitado. O CloverDX possibilita filtrar os dados de diversas fontes de dados. Configurado um componente de filtro para realizar a pesquisa de todos os dados que contenham o valor **\*\*\*537537\*\***, que é o cpf de um titular dos dados escolhido aleatoriamente, entre ambos conjuntos de dados e foi possível visualizar todos os dados desse usuário

conforme a Figura 56. Mesmo o CloverDX possibilitando o encontro dos dados do titular ele não pode determinar de qual das fontes conectadas esse dado pertence.

Figura 56 - CloverDX aplicar filtro aos dados

The screenshot shows the CloverDX Designer interface. The main workspace displays a data flow graph with the following components and connections:

- DatabaseReader** (green box) with ID 0, connected to the **Filter** component (orange box) with ID 62 296.
- Filter** component connected to the **afastrem\_072021 <auto>** component (yellow box) with ID 0.
- The **afastrem\_072021 <auto>** component is connected to two output nodes: **Success** (yellow box with a green checkmark) and **Fail** (yellow box with a red exclamation mark).

The **Data Inspector** window is open, showing the following table:

#	Nome	CPF	Descricao do cargo emprego
1	ELOI RAMOS	***537537**	AGENTE DE INSP SANIT IND PROD ORI

Below the table, it indicates "Loaded records: 1 | All records loaded" and "Filter is not set". The **Execution** window shows the job "tcc.grf" with status "Failed" (red exclamation mark) and "Run ID: 358".

Fonte: Próprio autor.

O resultado atingido é o esperado, a ferramenta foi capaz de apresentar todos os dados de determinado titular.

### 5.3.11 Caso de Teste 11: Anonimização

No caso de teste 11 (Quadro 21) compreende em testar se a ferramenta pode anonimizar dados pessoais e não possibilitar a reversão. O CloverDX não possui nenhum componente que realize a anonimização dos dados.

O resultado esperado não pode ser alcançado pela ferramenta que não realiza o processo de anonimização dos dados.

### **5.3.12 Caso de Teste 12: Coexistência**

O caso de teste 12 (Quadro 22) consiste em testar se a ferramenta executa sem apresentar falhas ao operar em simultâneo com as ferramentas de banco de dados no mesmo ambiente. O teste é feito em um sistema operacional Windows 10, utilizando um processador AMD Ryzen 5 2600 e 16GB de memória RAM, para a execução do teste apenas é executado tarefas próprias do sistema operacional e os bancos de dados SQL Express, Oracle XE e MySQL em conjunto com o CloverDX.

A ferramenta apresentou estabilidade e não apresentou nenhum erro ou travamento durante a execução dos demais casos de teste. Na conexão e operação dos dados não exibiu nenhum erro ou travamento, e funcionou normalmente com todas as fontes de dados conectadas.

Monitorado o uso de recursos do computador na execução do caso de teste 10 (Seção 5.3.10) , o processo do Pentaho Kettle que tem o nome de Spoon apresentou uso de 900MB de memória RAM conforme Figura 57. Nenhum travamento aconteceu pelo CloverDX ou por algum dos bancos de dados.

Figura 57 - CloverDX uso de recursos

The screenshot shows the CloverDX Designer interface with a workflow diagram. The workflow consists of a 'DatabaseReader' node connected to a 'Filter' node, which then branches into 'Fail' and 'Success' nodes. An 'Execution' window shows the job 'tcc.grf' running successfully. Overlaid on this is the Windows Task Manager 'Gerenciador de Tarefas' window, showing the 'Desempenho' (Performance) tab with the following data:

Nome	CPU	Memória	Disco	Rede	GP
OpenJDK Platform binary (4)	21,3%	951,2 MB	0,2 MB/s	0 Mbps	29%
OpenJDK Platform binary (2)	3,5%	894,5 MB	0,3 MB/s	0 Mbps	
Oracle RDBMS Kernel Executable	0,1%	631,2 MB	0,1 MB/s	0 Mbps	
Antimalware Service Executable	0,6%	126,8 MB	0 MB/s	0 Mbps	
Spotify (32 bits) (6)	0,1%	123,9 MB	0,1 MB/s	0 Mbps	
SQL Server Windows NT - 64 Bit	0%	95,9 MB	0 MB/s	0 Mbps	
Steam Client WebHelper	0%	58,4 MB	0 MB/s	0 Mbps	
Microsoft OneDrive	0%	48,6 MB	0 MB/s	0 Mbps	
Windows Explorer	0,5%	43,2 MB	0 MB/s	0 Mbps	

Fonte: Próprio autor.

O resultado esperado foi atingido, a ferramenta executada com demais programas não apresentou erros e travamentos. Durante a execução dos demais casos de teste a ferramenta também não apresentou travamentos e erros.

#### 5.4 CONSIDERAÇÕES FINAIS

O capítulo aplica os casos de testes (Seção 4.4) nas ferramentas selecionadas durante o trabalho (Seção 4.1) para a realização das avaliações das ferramentas de mapeamento de dados.

Cada seção é aplicado e apresentado os resultados dos casos de teste em cada ferramenta, a seção 5.1 é apresentada como ocorreu os casos de testes no Talend Open Studio. A ferramenta foi capaz de completar 5 dos 12 casos de testes

a que foi submetida. As principais funcionalidades que o Talend Open Studio possui são a criptografia e anonimização dos dados. Também possui uma loja onde é possível instalar componentes desenvolvidos pela comunidade, o que possibilita uma maior flexibilidade da ferramenta.

A seção 5.2 demonstra os resultados dos casos de testes aplicados à ferramenta Pentaho Kettle. Dos 12 casos de teste aplicados a ferramenta pode completar 6 com sucesso. O armazenamento de eventos e a rotulação e criptografia dos dados, foram as principais funcionalidades propostas pela ferramenta. Por ser uma ferramenta mais popular, possui bastante conteúdo em sua documentação e nos fóruns da comunidade.

A ferramenta CloverDX apresentada na seção 5.3 completou de 5 a 12 casos de testes a que foi submetida. As principais funcionalidades apresentadas pela ferramenta foram o gerenciamento de acesso e armazenamento de eventos. Por ser uma ferramenta paga não possui flexibilidade para instalar componentes de terceiros, a documentação que a ferramenta apresenta é bem completa porém não se encontra conteúdo na internet, exceto nas páginas do fabricante da ferramenta.

As ferramentas não puderam concluir os casos de testes por não terem os componentes necessários para a realização dos casos de teste, e nem apresentaram em suas documentações formas de se atingir os resultados esperados pelos casos de testes.

## 6 AVALIAÇÃO DAS FERRAMENTAS

Para executar a avaliação das ferramentas foram utilizados critérios e métricas definidas no Capítulo 4. Foram determinados critérios de qualidade em uso e de qualidade externa (Quadro 8). Os critérios selecionados possuem relação às métricas (Quadros 9 e 10) que são utilizadas para a mensuração da avaliação das ferramentas. O peso de cada métrica foi atribuído na seção 4.5.

Em seguida à avaliação das ferramentas, aplicando os critérios definidos, foi realizada a avaliação individual de cada ferramenta considerando a pontuação recebida em cada critério.

### 6.1 ADEQUAÇÃO

O propósito da avaliação é identificar o quão adequadas são as funções avaliadas. É considerado o número de funções que são adequadas para a execução de tarefas especificadas, em comparação com o número de funções que foram avaliadas.

Nenhuma ferramenta atendeu todos os requisitos. As funcionalidades atendidas por cada um das ferramentas localizam-se no Quadro 26.

Quadro 26 - Funcionalidades Adequação

Funcionalidade	Satisfeitas		
	Talend	Pentaho Kettle	CloverDX
Conexão a fonte de dados MySQL	X	X	X
Conexão a fonte de dados Oracle XE	X	X	X
Conexão a fonte de dados SQL Server	X	X	X
Avaliação de Risco			
Personalização da Interface		X	
Disponível no idioma Português	X	X	
Documentação	X	X	X

Fonte: Próprio autor.

A adequação é calculada utilizando a fórmula  $X = 1 - A / B$ , onde A retrata o número de funcionalidades não atendidas e o B o número de funções avaliadas



conforme o Quadro 26. É considerado que quanto mais próximo de 1 mais apropriado é a ferramenta, Pentaho Kettle (0.85) obteve o melhor resultado (Quadro 27).

Quadro 27 - Critério Adequação

Característica	Subcaracterística	Métrica	Ferramenta	Interpretação	Resultado
Funcionalidade	Adequação	Adequação Funcional	Talend	X =1-2/7	0,71
Funcionalidade	Adequação	Adequação Funcional	Pentaho Kettle	X =1-1/7	0,85
Funcionalidade	Adequação	Adequação Funcional	CloverDX	X =1-3/7	0,57

Fonte: Próprio autor.

## 6.2 ACURÁCIA

O critério de acurácia tem o propósito de identificar com que frequência os usuários finais encontram resultados com precisão inadequada. Como exemplo disso, foram utilizados relatórios ou consultas realizadas que continham o resultado final de dados imprecisos.

Foram efetuados diversos testes objetivando identificar a imprecisão das ferramentas através dos casos de testes número 2 (Avaliação dos riscos dos dados), 4 (Classificação dos dados), 5 (Rotulação dos dados), 7 (Armazenamento de eventos), 8 (Resposta a incidentes de segurança), 10 (Entrega de dados aos titulares) determinados na seção 4.4. Para identificar a precisão dos resultados foram realizadas as tarefas do Quadro 28.

Quadro 28 - Tarefas Acurácia

Tarefas	Precisão		
	Talend	Pentaho Kettle	CloverDX
Conexão a fonte de dados MySQL	X	X	X
Conexão a fonte de dados Oracle XE		X	X
Conexão a fonte de dados SQL Server	X	X	X
Armazenamento de eventos		X	X
Criptografia dos dados	X	X	
Entrega de dados aos titulares	X	X	

Fonte: Próprio autor.

A acurácia é calculada usando a fórmula  $X = A / B$ , onde A representa o número de resultados com precisão inadequada e B o tempo em horas de operação do sistema. Considerando que o resultado mais próximo de 0 mais adequado, a ferramenta Pentaho Kettle obteve o melhor resultado (Quadro 29).

Quadro 29 - Critério Acurácia

Característica	Subcaracterística	Métrica	Ferramenta	Interpretação	Resultado
Funcionalidade	Acurácia	Precisão	Talend	X = 2 / 30	0,06
Funcionalidade	Acurácia	Precisão	Pentaho Kettle	X = 0 / 30	0
Funcionalidade	Acurácia	Precisão	CloverDX	X = 2 / 30	0,06

Fonte: Próprio autor.

### 6.3 INTEROPERABILIDADE

O critério de interoperabilidade tem a finalidade de reconhecer o uso da ferramenta interagindo com os outros sistemas. É verificada a frequência que é encontrada restrições ou falhas entre a troca de dados da ferramenta e demais sistemas.

Para isso, foram executados diversos testes da ferramenta avaliada em conjunto com a troca de dados de outros softwares. Para a realização da avaliação foram realizadas tarefas com os sistemas descritos no Quadro 30.

Quadro 30 - Tarefas Interoperabilidade

Tarefas	Interoperabilidade disponível		
	Talend	Pentaho Kettle	CloverDX
Troca de dados: MySQL	X	X	X
Troca de dados: Oracle XE	X	X	X
Troca de dados: SQL Server	X	X	X
Troca de dados: Microsoft Excel	X	X	X

Fonte: Próprio autor.

A interoperabilidade é calculada empregando a fórmula  $X = 1 - A / B$ , onde A representa o número de sistemas que apresentaram falha ao trocar dados com a

ferramenta avaliada e B número de sistemas avaliados. Considerando que o resultado mais próximo de 1 é o melhor, todas as ferramentas apresentaram o resultado máximo (Quadro 31).

Quadro 31 - Critério Interoperabilidade

Característica	Subcaracterística	Métrica	Ferramenta	Interpretação	Resultado
Funcionalidade	Interoperabilidade	Interoperabilidade disponível	Talend	X = 1 - 0 / 4	1
Funcionalidade	Interoperabilidade	Interoperabilidade disponível	Pentaho Kettle	X = 1 - 0 / 4	1
Funcionalidade	Interoperabilidade	Interoperabilidade disponível	CloverDX	X = 1 - 0 / 4	1

Fonte: Próprio autor.

#### 6.4 CONFORMIDADE

Este critério é o de maior peso na avaliação das ferramentas, porque aponta o quão compatível é a ferramenta com os regulamentos, padrões e convenções aplicáveis a este tipo de ferramenta. São contados as funcionalidades definidas na seção 4.3 com base na ABNT NBR ISO/IEC 27701 e na LGPD.

Foram testados as ferramentas através dos casos de teste de acordo com os itens de conformidade (Quadro 32).

Quadro 32 - Conformidades

Conformidades	Satisfeitas		
	Talend	Pentaho Kettle	CloverDX
Avaliação de riscos dos dados			
Segurança de acesso			X
Identificação dos dados			
Rotulação dos dados		X	
Criptografia dos dados	X	X	
Armazenamento de eventos(logs)		X	X
Resposta a incidentes de segurança			
Análise crítica técnica do compliance			
Entrega de dados aos titulares	X	X	X
Anonimização	X		

Fonte: Próprio autor.

Para realizar o cálculo da conformidade através da fórmula  $X = 1 - A / B$ , onde A representa o número de itens não conformes e B representa o número de itens totais testados. O trabalho realiza a avaliação total de 10 itens de conformidade. Considerando mais próximo de 1, mais adequada será a ferramenta, o Pentaho Kettle obteve índice de 0,40 , seguido pelo Talend Open Studio e CloverDX com índice de 0,30 (Quadro 33).

Quadro 33 - Critério Conformidade

Característica	Subcaracterística	Métrica	Ferramenta	Interpretação	Resultado
Funcionalidade	Conformidade	Conformidade de funcionalidade	Talend Open Studio	$X = 1 - 7 / 10$	0,30
Funcionalidade	Conformidade	Conformidade de funcionalidade	Pentaho Kettle	$X = 1 - 6 / 10$	0,40
Funcionalidade	Conformidade	Conformidade de funcionalidade	CloverDX	$X = 1 - 7 / 10$	0,30

Fonte: Próprio autor.

## 6.5 RECUPERABILIDADE

O critério tem a finalidade de identificar a capacidade da ferramenta em restabelecer seu nível de desempenho e recuperar os dados afetados em caso de uma falha.

Foram executados diversos testes, objetivando identificar a inconsistência dos componentes de interface das ferramentas, através dos casos de teste de número, 1 (Parametrização da ferramenta), 6 (Criptografia dos dados), 7 (Armazenamento de eventos), 10 (Entrega de dados ao titular), 11 (Anonimização), 12 (Coexistência).

A recuperabilidade foi calculada utilizando a forma  $X = A / N$  , onde A representa o número de erros recuperados e N o número de erros ocorridos na ferramenta. Portanto, considerando o resultado o quanto mais próximo de 0 mais

adequado, tanto o Pentaho Kettle (0) quanto o CloverDX (0) e o Talend Open Studio (0) obtiveram resultados perfeitos (Quadro 34).

Quadro 34 - Critério Recuperabilidade

Característica	Subcaracterística	Métrica	Ferramenta	Interpretação	Resultado
Confiabilidade	Recuperabilidade	Flexibilidade de recuperação	Talend Open Studio	$X = 1 / 2$	0,5
Confiabilidade	Recuperabilidade	Flexibilidade de recuperação	Pentaho Kettle	$X = 0 / 0$	0
Confiabilidade	Recuperabilidade	Flexibilidade de recuperação	CloverDX	$X = 0 / 0$	0

Fonte: Próprio autor.

## 6.6 OPERACIONALIDADE

Este critério tem o objetivo de identificar, por meio da observação, o quão consistente é o componente da interface do usuário.

Realizaram-se diversos testes objetivando identificar a inconsistência dos componentes de interface das ferramentas, através dos casos de teste de número, 1 (Parametrização da ferramenta), 6 (Criptografia dos dados), 7 (Armazenamento de eventos), 10 (Entrega de dados ao titular), 11 (Anonimização).

A operacionalidade foi calculada utilizando a fórmula  $X = A / B$ , onde A representa o número de operações que o usuário encontrou inaceitavelmente inconsistente com a expectativa do usuário e B o tempo em horas de operação do sistema. Portanto, considerando o resultado o quanto mais próximo de 0 mais adequado, tanto o Pentaho Kettle (0) quanto o CloverDX (0) e o Talend Open Studio (0) obtiveram resultados perfeitos (Quadro 35).

Quadro 35 - Critério Operacionalidade

Característica	Subcaracterística	Métrica	Ferramenta	Interpretação	Resultado
Usabilidade	Operacionalidade	Consistência operacional em uso	Talend Open Studio	X = 0 / 30	0
Usabilidade	Operacionalidade	Consistência operacional em uso	Pentaho Kettle	X = 0 / 30	0
Usabilidade	Operacionalidade	Consistência operacional em uso	CloverDX	X = 0 / 30	0

Fonte: Próprio autor.

## 6.7 COEXISTÊNCIA

O critério tem o propósito de determinar a frequência com que os usuários encontram limitações ou falhas inesperadas, ao executarem a ferramenta simultaneamente com outro software usado com frequência pelos usuários.

Nesse sentido, foram efetuados diversos testes com o objetivo de identificar a disponibilidade da coexistência dos softwares através dos casos de testes de número 1 (Parametrização da ferramenta) e 12 (Coexistência).

A coexistência foi calculada utilizando a fórmula  $X = A / B$ , onde A representa o número de restrições ou falhas inesperadas que o usuário enfrenta ao operar em simultâneo outro software, e B o tempo em horas de operação simultânea com outros softwares. Dessa forma, considerando que o resultado o quanto mais próximo de 0 mais adequado. Apenas a ferramenta Talend Open Studio apresentou índice maior que 0 (Quadro 36).

Quadro 36 - Critério Coexistência

Característica	Subcaracterística	Métrica	Ferramenta	Interpretação	Resultado
Portabilidade	Coexistência	Coexistência disponível	Talend Open Studio	X = 1 / 30	0,03
Portabilidade	Coexistência	Coexistência disponível	Pentaho Kettle	X = 0 / 30	0
Portabilidade	Coexistência	Coexistência disponível	CloverDX	X = 0 / 30	0

Fonte: Próprio autor.

## 6.8 EFETIVIDADE

Este critério tem a finalidade de identificar qual é a frequência de erros apresentados pelas ferramentas.

Foram realizados diversos testes visando identificar a frequência de erros das ferramentas por meio dos casos de teste (Quadro 37).

Quadro 37 - Tarefas Efetividade

Tarefas	Satisfeitas		
	Talend	Pentaho Kettle	CloverDX
Troca de dados: MySQL	X	X	X
Troca de dados: Oracle XE		X	X
Troca de dados: SQL Server	X	X	X
Troca de dados: Microsoft Excel	X	X	X
Armazenamento de eventos		X	X
Entrega de dados aos titulares	X	X	X
Exportar dados para arquivo	X	X	X
Adição e remoção de componentes	X	X	X

Fonte: Próprio autor.

O teste com a ferramenta Pentaho Kettle não apresentou nenhum erro em 9 tarefas completadas, enquanto a ferramenta Talend Open Studio encontrou problemas ao executar conexões com o Oracle XE.

A efetividade é calculada utilizando a fórmula  $X = A / B$ , onde A representa o número de erros encontrados e B número de tarefas completadas pelo usuário. Desse modo, considerando que o resultado quanto mais próximo de 0 mais adequado, a ferramenta as ferramentas Pentaho Kettle e CloverDX obtiveram um melhor resultado nesse critério (Quadro 38).

Quadro 38 - Critério Efetividade

Característica	Métrica	Ferramenta	Interpretação	Resultado
Efetividade	Frequência de erro	Talend Open Studio	$X = 2 / 9$	0,22
Efetividade	Frequência de erro	Pentaho Kettle	$X = 0 / 9$	0
Efetividade	Frequência de erro	CloverDX	$X = 0 / 9$	0

Fonte: Próprio autor.

## 6.9 PRODUTIVIDADE

Este critério tem o propósito de identificar quanto tempo demora para o usuário completar uma tarefa. É medido pelos testes com o usuário, calculando por meio da divisão do tempo ocioso do usuário pelo tempo das tarefas completadas pelo usuário, dessa forma obtendo o valor e qual o tempo da tarefa.

Foram realizados diversos testes visando identificar o tempo de tarefa das ferramentas através das tarefas(Quadro 37) definidos com base nos casos de testes (Seção 4.4).

A produtividade é calculada através da fórmula  $X = A / B$ , onde A representa o tempo ocioso do usuário e B o tempo das tarefas completadas pelo usuário. Dessa forma, considerando que o resultado quanto mais próximo de 0 mais adequado, o CloverDX obteve melhor resultado neste critério (Quadro 39).

Quadro 39 - Critério Produtividade

Característica	Métrica	Ferramenta	Interpretação	Resultado
Produtividade	Tempo da tarefa	Talend Open Studio	$X = 4 / 35$	0,11
Produtividade	Tempo da tarefa	Pentaho Kettle	$X = 1 / 35$	0,02
Produtividade	Tempo da tarefa	CloverDX	$X = 0 / 35$	0

Fonte: Próprio autor.

## 6.10 CONSIDERAÇÕES FINAIS

Após ter sido realizada a avaliação individual das ferramentas através dos critérios e métricas selecionadas, é obtido o resultado das avaliações. A análise não foi realizada de forma quantitativa, devido algumas métricas possuírem resultados melhores quando se aproximavam de 1 e outras obtinham resultados melhores quando se aproximavam de 0.

O critério de conformidade possui maior peso, devido a importância que as funcionalidades avaliadas possuem para atingir o objetivo do trabalho. Assim o Quadro 40 classifica através de números ordinais cada funcionalidade do critério de



conformidade, quando em branco a ferramenta não pode completar a funcionalidade.

Quadro 40 - Conformidades

Conformidades	Satisfeitas		
	Talend	Pentaho Kettle	CloverDX
Avaliação de riscos dos dados			
Segurança de acesso			1º
Identificação dos dados			
Rotulação dos dados		1º	
Criptografia dos dados	1º	2º	
Armazenamento de eventos(logs)		1º	2º
Resposta a incidentes de segurança			
Análise crítica técnica do compliance			
Entrega de dados aos titulares	2º	1º	3º
Anonimização	1º		

Fonte: Próprio autor.

Portanto, analisando os resultados obtidos, nenhuma das ferramentas atende integralmente aos critérios propostos, nos padrões de conformidade, que é o principal critério. Portanto, a ferramenta que demonstra as melhores características e obtém maior índice é o Pentaho Kettle.

As ferramentas Talend Open Studio e CloverDX obtiveram resultados próximos, então foi realizada a comparação e classificação através dos números ordinais dos demais critérios avaliados (Quadro 41).

Quadro 41 - Classificação dos critérios

Critérios	Talend Open Studio	Pentaho Kettle	CloverDX
Adequação	2º	1º	3º
Acurácia	3º	1º	2º
Interoperabilidade	1º	1º	1º
Conformidade	2º	1º	3º
Recuperabilidade	3º	1º	2º
Operacionalidade	1º	1º	1º
Coexistência	3º	1º	2º
Efetividade	3º	1º	2º
Produtividade	3º	2º	1º

Fonte: Próprio autor.

Levando em consideração as análises realizadas, nenhuma das ferramentas está adequada totalmente aos critérios apresentados, principalmente ao de conformidade, que é o principal. A ferramenta que portanto demonstrou melhor capacidade e obteve maior valor foi o Pentaho Kettle.

O Pentaho Kettle se mostrou mais adequado na maioria dos critérios, sendo estes a acurácia, interoperabilidade, recuperabilidade, operacionalidade, coexistência e efetividade. O índice total do Pentaho Kettle deve-se principalmente pela ferramenta ter atingido maior média no critério de conformidade, assim se mostrando maior aderente aos requisitos exigidos pela LGPD e pela norma ABNT NBR ISO/IEC 27701. Apesar da ferramenta não ter resultados satisfatórios, se mostrou mais adequada que as demais.

O Talend Open Studio e CloverDX obtiveram médias muito próximas, principalmente por possuírem o mesmo índice no critério de conformidade. Embora tenham atingido melhor índice em outros critérios, as duas ferramentas não mostraram potencial para alcançar o objetivo proposto.

## 7 CONCLUSÕES

O objetivo principal deste trabalho foi analisar e avaliar ferramentas de mapeamento de dados com o intuito de identificar qual ferramenta possui maior aderência a LGPD e a norma ABNT NBR ISO/IEC 27701. Para atingir este objetivo e executar as avaliações, foram necessários estudos de conceitos de segurança da informação e da LGPD. Conforme foi elaborando-se o desenvolvimento do trabalho, compreendeu-se que a norma ABNT NBR ISO/IEC 27701 é essencial para escolher uma ferramenta para a LGPD.

Todavia, para que uma organização atinja todos os requisitos da LGPD, ainda é importante e necessário que se implemente políticas e controles de segurança, ou seja, as ferramentas de mapeamento de dados são essenciais para facilitar essas implementações.

Nesse sentido, foram selecionadas algumas ferramentas em uma análise inicial. Com essa análise foi possível identificar dois tipos de ferramentas de mapeamento de dados. Um grupo de ferramentas possui o objetivo na extração, transformação e carregamento dos dados em diversas fontes de dados, e é muito utilizado em data warehouse. Já o outro grupo de ferramentas é mais recente, e tem o objetivo de adequar as organizações às leis de conformidade existentes.

As ferramentas do grupo de conformidade seriam mais alinhadas ao objetivo deste trabalho, não obstante, por esse nicho ser muito recente, ainda não se encontra disponível no mercado versões que sejam gratuitas. Conseqüentemente, todas as ferramentas pagas analisadas como o OneTrust, BigID, DataGrail e TrustArc não possuem uma versão de testes que possibilitem realizar a avaliação. Por isso, as empresas desenvolvedoras dessas ferramentas foram contatadas por e-mail para verificar a possibilidade de liberar uma versão teste dessas ferramentas, porém nenhuma deu retorno.

Deste modo, como nenhuma ferramenta específica pode ser testada, tornou-se necessário selecionar as ferramentas do grupo de extração, transformação e carregamento dos dados. Sendo selecionadas as ferramentas

Talend Open Studio, Pentaho Kettle e CloverDX, as duas primeiras sendo gratuitas e a terceira paga, mas possuindo um período de 45 dias para avaliação.

A fase da escolha da definição dos critérios e métricas de avaliação foram mais teóricas, porém muito importantes para o desenvolvimento do trabalho. Com base nos estudos das normas ABNT NBR ISO/IEC 25020 e 25030 que regem os padrões de qualidade de softwares, foram definidos os critérios, métricas e casos de testes para realização da avaliação das ferramentas, atrelando as funcionalidades fundamentais necessárias para atingir o atendimento das necessidades da LGPD e da ABNT NBR ISO/IEC 27701.

Conseqüentemente à realização da avaliação das ferramentas, foi possível analisar e identificar que quem se mostrou mais adequado foi o Pentaho Kettle, mas embora ele tenha tido o melhor resultado, ainda precisa ser aprimorado para atingir o objetivo proposto. Observou-se que o Pentaho pode atingir 4 requisitos de 10 sobre o critério de conformidade, que é o critério fundamental para a avaliação. A ferramenta pode completar os requisitos de rotulação dos dados, criptografia, armazenamento de eventos e entrega de dados dos titulares. As ferramentas Talend Open Studio e CloverX só conseguiram atingir 3 dos 10 requisitos sobre o critério de conformidade. O Talend Open Studio pode completar os requisitos de criptografia, entrega dos dados aos titulares e anonimização. Já o CloverDX concluiu os requisitos de segurança de acesso, armazenamento de eventos e entrega de dados aos titulares. Apesar do Pentaho Kettle obter o melhor resultado, as demais ferramentas não ficaram com um resultado muito abaixo do Pentaho.

Portanto, foi possível chegar a conclusão que nenhuma das ferramentas está totalmente adequada, já que a ferramenta que obteve o melhor desempenho sequer pode atingir a maioria dos requisitos solicitados. As ferramentas de extração, transformação e carregamento de dados, de fato, não puderam ajudar as organizações a atingirem seus objetivos de conformidade com a LGPD e a norma ABNT NBR ISO/IEC 27701.

Indubitavelmente, mesmo que possam ajudar a organização a atingir determinados requisitos, precisam ser fortemente aprimoradas para realmente favorecerem as organizações. Considerando até mesmo o CloverDX que é uma

ferramenta paga, mas também, não demonstrou maiores resultados que as demais ferramentas que são gratuitas, e acabou obtendo índice próximo ao Talend Open Studio. Deste modo, as ferramentas de extração, transformação e carregamento de dados por licenciamento pago, não são a solução para o problema proposto.

O mercado ainda apresenta poucas ferramentas de mapeamento de dados voltadas à conformidade, mesmo sendo elas, ferramentas já existentes são direcionadas a um público de grandes organizações, custando altos valores de licenciamento. É importante ressaltar que as organizações de médio e pequeno porte passam desatendidas por essas ferramentas, mesmo com as leis de privacidade de dados já sendo aplicadas em diversas partes do mundo. Outro aspecto fundamental é levar-se em conta que a transformação digital se tornou essencial para a estratégia competitiva das organizações, gerando a necessidade das mesmas se adequarem às conformidades impostas por leis de privacidade de dados. Como alternativa, grandes organizações podem optar por ferramentas do grupo de conformidade. Já médias e pequenas organizações que não têm grande poder de investimento, podem utilizar as ferramentas avaliadas para auxiliar em sua adequação.

Assim, reitera-se que: este trabalho contribui diretamente para auxiliar as organizações sobre a adesão ao uso de ferramentas de mapeamento de dados ao seu planejamento, auxiliando na adequação a LGPD e a norma ABNT NBR ISO/IEC 27701.

Por fim, sugere-se, como continuidade ou complementação da realização desse trabalho, a avaliação de ferramentas de mapeamento de dados voltadas à conformidade, assim como a inclusão de outras ferramentas de mapeamento de dados que não foram testadas aqui como OneTrust e BigID. Também recomenda-se o desenvolvimento de ferramentas de código-livre voltadas para a conformidade das organizações.

## REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 16167**: Segurança da Informação — Diretrizes para classificação, rotulação e tratamento da informação. Rio de Janeiro, 2013.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 25020:2009**: Engenharia de software - Requisitos e avaliação da qualidade de produto de software (SQuaRE) - Guia e modelo de referência para medição. Rio de Janeiro, 2009.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 25030:2008**: Engenharia de software - Requisitos e Avaliação da Qualidade de Produto de Software (SQuaRE) - Requisitos de qualidade. Rio de Janeiro, 2008.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 25000:2008**: Engenharia de software - Requisitos e avaliação da qualidade de produtos de software (SQuaRE) - Guia do SQuaRE. Rio de Janeiro, 2008.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 27001:2013**: Tecnologia da informação - Técnicas de segurança - Sistemas de gestão de segurança da informação - Requisito. Rio de Janeiro, 2013.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 27002:2005**: Tecnologia da Informação: Código de Prática para a Gestão da Segurança da Informação. Rio de Janeiro, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 29100:2020**: Tecnologia da informação — Técnicas de segurança — Estrutura de Privacidade. Rio de Janeiro, 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 27701:2019**: Técnicas de segurança — Extensão da ABNT NBR ISO/IEC 27001 e ABNT NBR ISO/IEC 27002 para gestão da privacidade da informação — Requisitos e diretrizes. Rio de Janeiro, 2019.

ALTAMIRA TECHNOLOGIES CORPORATION . **Lumify**. 2015. Disponível em: <https://github.com/lumifyio/lumify>. Acesso em: 15 nov. 2020.

AMY-VOGT, Betsy. **BigID brings a new era of data security, privacy and governance in the cloud**. 2021. Disponível em:

<https://siliconangle.com/2021/03/24/bigid-brings-a-new-era-of-data-security-privacy-and-governance-in-the-cloud-cubeoncloudawsstartups/>. Acesso em: 15 abr. 2021.

BOWEN, Jonathan. **Getting Started with Talend Open Studio for Data Integration**. Birmingham, Uk: Packt Publishing, 2012.

BRASIL. AGÊNCIA NACIONAL DE TRANSPORTES TERRESTRES - ANTT. . **Motoristas Habilitados**. 2021. Disponível em: <https://dados.gov.br/dataset/motoristas-habilitados>. Acesso em: 07 ago. 2021.

BRASIL. MINISTÉRIO DA ECONOMIA. . **Gestão de Pessoas (Executivo Federal) - Afastamentos e Licenças**. 2021. Disponível em: <https://dados.gov.br/dataset/afastamento-remunerado>. Acesso em: 07 ago. 2021.

BRASIL. LEI Nº 13.709, DE 14 DE AGOSTO DE 2018. **Lei Geral de Proteção de Dados Pessoais**, Brasília,DF, agosto 2018. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2015-2018/2018/Lei/L13709.htm](http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm). Acesso em: 24 ago. 2020.

CASTERS, Matt; BAUMAN, Roland; VAN DONGEN, Jos. **Pentaho® Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration**. Indianapolis, Indiana: Wiley Publishing, 2010.

CIO. **14 principais ferramentas para saber se você cumpre os requisitos do GDPR**. 2018. Disponível em: <https://cio.com.br/tendencias/14-principais-ferramentas-para-saber-se-voce-cumpre-os-requisitos-do-gdpr/>. Acesso em: 13 abr. 2021.

CISCO. **Maximizing the value of your data privacy investments**. 2019. Disponível em: [https://www.cisco.com/c/dam/en\\_us/about/doing\\_business/trust-center/docs/dpbs-2019.pdf](https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/dpbs-2019.pdf). Acesso em: 07 set. 2020

COMPUTERWORLD. **Com serviço de “DPO as a Service” Engineering lança oferta focada na LGPD**. 2020. Disponível em: <https://computerworld.com.br/negocios/com-servico-de-dpo-as-a-service-engineering-lanca-oferta-focada-na-lgpd/>. Acesso em: 04 abr. 2021.

COMPUTERWORLD. **2019: o ano do ecossistema digital orientado por dados**. 2018. Disponível em: <https://computerworld.com.br/2018/12/10/2019-o-ano-do-ecossistema-digital-orientado-por-dados/>. Acesso em: 24 set. 2020.

DATAGRAIL. **Simplify data privacy, create transparency**. 2021. Disponível em: [https://images.g2crowd.com/uploads/attachment/file/133103/DataGrail-1-pager.pdf?\\_hstc=171774463.38646764ccb8985a74b5fef1de63adb2.1618263729999.16185241](https://images.g2crowd.com/uploads/attachment/file/133103/DataGrail-1-pager.pdf?_hstc=171774463.38646764ccb8985a74b5fef1de63adb2.1618263729999.16185241)

52711.1618526586376.4&\_\_hssc=171774463.2.1618526586376&\_\_hsfp=4003680491. Acesso em: 15 abr. 2021.

ESPINOZA, Javier. **EU admits it has been hard to implement GDPR.** 2020. Disponível em: <https://www.irishtimes.com/business/technology/eu-admits-it-has-been-hard-to-implementation-gdpr-1.4286207>. Acesso em: 21 abr. 2021.

FERNANDES, A.; ABREU, V. **Implantando a governança de TI: da estratégia à gestão dos processos e serviços.** 2. ed. Rio de Janeiro: Brasport, 2008.

G2 . **Best Data Mapping Software.** 2020. Disponível em: <https://www.g2.com/categories/data-mapping>. Acesso em: 24 set. 2020.

GARTNER. **Gartner Magic Quadrant for Data Integration Tools.** 2020. Disponível em: <https://www.gartner.com/en/documents/3989223/magic-quadrant-for-data-integration-tools>. Acesso em: 21 ago. 2021.

HOFFMAN, Jason. **Top 4 Open Source Data Mapping Tools.** Disponível em: <https://wisdomplexus.com/blogs/open-source-data-mapping-tools/>. Acesso em: 14 set. 2020.

HOSPELHORN, Sarah. **Document Record of Processing and Sharing Activity.** 2020. Disponível em: <https://bigid.com/blog/document-record-of-processing-and-sharing-activity/>. Acesso em: 13 abr. 2021

IBM (Brasil). **Estudo da IBM mostra que contas comprometidas de funcionários levaram às violações de dados mais caras durante o ano passado.** 2020. Disponível em: <https://www.ibm.com/blogs/ibm-comunica/estudo-da-ibm-mostra-que-contas-comprometidas-de-funcionarios-levaram-as-violacoes-de-dados-mais-caras-durante-o-ano-passado/>. Acesso em: 24 ago. 2020.

IYER, Sailesh; LAKHTARIA, Kamaljit. **Practical Evaluation and Comparative Study of Big Data Analytical Tools.** 2017. Disponível em: [https://www.researchgate.net/profile/Sailesh\\_Iyer/publication/316190087\\_Practical\\_Evaluation\\_and\\_Comparative\\_Study\\_of\\_Big\\_Data\\_Analytical\\_Tools/links/58f5dc1eaca27289c21d26d5/Practical-Evaluation-and-Comparative-Study-of-Big-Data-Analytical-Tools.pdf](https://www.researchgate.net/profile/Sailesh_Iyer/publication/316190087_Practical_Evaluation_and_Comparative_Study_of_Big_Data_Analytical_Tools/links/58f5dc1eaca27289c21d26d5/Practical-Evaluation-and-Comparative-Study-of-Big-Data-Analytical-Tools.pdf). Acesso em: 15 nov. 2020.

KASPERSKY . **International Privacy Day 2020 Kaspersky Report.** 2020. Disponível em: <https://media.kasperskydaily.com/wp-content/uploads/sites/92/2020/01/27103216/International-Privacy-Day-2020-Kaspersky-report.pdf>. Acesso em: 02 set. 2020.



MATSUMOTO, Cristina Yoshie. **A IMPORTÂNCIA DO BANCO DE DADOS EM UMA ORGANIZAÇÃO.** 2006. Disponível em: <https://core.ac.uk/download/pdf/199473173.pdf>. Acesso em: 10 set. 2020

ONETRUST. **OneTrust User Guide.** 2020. Disponível em: <https://my.onetrust.com/s/topiccatalog>. Acesso em: 19 nov. 2020.

PARLAMENTO EUROPEU. REGULAMENTO 2016/679, DE 27 DE ABRIL DE 2016. **Regulamento Geral sobre a Proteção de Dados**, Bruxelas, abril 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679>. Acesso em: 08 set. 2020

PRASAD, Bakshi Rohit; AGARWAL, Sonali. Comparative Study of Big Data Computing and Storage Tools: A Review. **International Journal Of Database Theory And Application.** Australia, p. 45-66. set. 2016. Disponível em: [http://article.nadiapub.com/IJDTA/vol9\\_no1/5.pdf](http://article.nadiapub.com/IJDTA/vol9_no1/5.pdf). Acesso em: 15 nov. 2020.

SECURITYMAGAZINE. **Nearly half of privacy requests last year were to stop the sale of personal data.** 2021. Disponível em: <https://www.securitymagazine.com/articles/94913-nearly-half-of-privacy-requests-last-year-were-to-stop-the-sale-of-personal-data>. Acesso em: 15 abr. 2021

SEMIDÃO, Rafael Aparecido Moron. **dados, informação e conhecimento enquanto elementos de compreensão do universo conceitual da ciência da informação: contribuições teóricas.** 2014. Disponível em: [https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/semidao\\_ram\\_me\\_mar.pdf](https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/semidao_ram_me_mar.pdf). Acesso em: 02 set. 2020.

SÊMOLA, M. **Gestão da segurança da informação: uma visão executiva.** Rio de Janeiro: Campus, 2003.

SHAHBAZ, Qamar. **Data Mapping for Data Warehouse Design.** Waltham, Ma: Elsevier, 2016.

YUNIANTA, Arda; YUSOF, Norazah; BRAMANTORO, Arif; HAVILUDDIN, Havaluddin; OTHMAN, Mohd Shahizan; DENGGEN, Nataniel. Data mapping process to handle semantic data problem on student grading system. **International Journal Of Advances In Intelligent Informatics.** Yogyakarta, p. 157-166. nov. 2016. Disponível em: <https://doi.org/10.26555/ijain.v2i3.84>. Acesso em: 25 out. 2020.

TALEND. **TClassify.** 2021. Disponível em: <https://help.talend.com/r/en-US/7.3/machine-learning/tclassify>. Acesso em: 06 out. 2021.

TRUSTARC. **PrivacyCentral: Close the Insight-to-Action Gap**. 2021. Disponível em: [https://info.trustarc.com/Web-Resource-2021-03-02-PrivacyCentral-SB\\_LP.html](https://info.trustarc.com/Web-Resource-2021-03-02-PrivacyCentral-SB_LP.html). Acesso em: 18 abr. 2021.

YUNIANTA, Arda; BARUKAB, Omar Mohammed; YUSOF, Norazah; DENGEN, Nataniel; HAVILUDDIN; OTHMAN, Mohd Shahizan. **Semantic data mapping technology to solve semantic data problem on heterogeneity aspect**. 2017. Disponível em: [https://www.researchgate.net/publication/328048268\\_Semantic\\_data\\_mapping\\_technology\\_to\\_solve\\_semantic\\_data\\_problem\\_on\\_heterogeneity\\_aspect/fulltext/5a3fa85daca272d29451a95d/Semantic-data-mapping-technology-to-solve-semantic-data-problem-on-heterogeneity-aspect.pdf?origin=publication\\_detail](https://www.researchgate.net/publication/328048268_Semantic_data_mapping_technology_to_solve_semantic_data_problem_on_heterogeneity_aspect/fulltext/5a3fa85daca272d29451a95d/Semantic-data-mapping-technology-to-solve-semantic-data-problem-on-heterogeneity-aspect.pdf?origin=publication_detail). Acesso em: 26 out. 2020.